The Constant Atlas: Mapping Public Data for Individuals and Their Cities

By

Jia Zhang

B.F.A. Rhode Island School of Design
M.F.A. Parsons
M.S. MIT

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning, in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Media Arts and Sciences
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2018

Author Signature redacted

Jia Zhang
Program in Media Arts and Sciences
May 4, 2018

Certified by Signature redacted

Ethan Zuckerman
Associate Professor of the Practice in Media Arts and Sciences
Program in Media Arts and Sciences
Thesis Supervisor

Signature redacted

Accepted by

Todd Machover
Academic Head
Program in Media Arts and Sciences

The Constant Atlas: Mapping Public Data for Individuals and Their Cities

By

Jia Zhang

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning on May 11, 2018 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Media Arts and Sciences

Abstract

Over the past ten years the ability of institutions and businesses to capture, aggregate, and process an individual's data has grown significantly as digital technology has increasingly integrated into our daily lives. In the urban informatics context and in computational social science, projects use data collected about our behavior in the urban environment to solve problems including traffic congestion and public safety, the creation of targeted advertising, and the development of entire neighborhoods. Some projects using aggregate data may ultimately benefit individuals by making improvements to their environment at large. Although individuals are the source of aggregate information, an individual citizen often does not directly engage with the data collected about them.

The research contained in this dissertation explores a series of visualization experiments concerning direct engagement between citizens and public datasets such as the U.S.Census. In order for such visualizations to be effective, they not only have to efficiently communicate data, but must also be intuitive, evocative, and utilize narratives presented from the user's perspective. In this dissertation I address the question: How can we design visualizations which inform daily interaction between individuals and public data about their environment?

To answer this question, the dissertation introduces 4 sets of maps: (1) the Powers Map and Scopes Map contextualizes Census data(American Community Survey) by invoking changes in scale, (2) the Sightline Map and Cross Section Map use a person's physical experiences to orient Census data, (3) the Filtered Satellite Maps give qualitative comparisons of conditions described by Census tables, and (4) the Personal History Map leverages an individual's geospatial history to filter Census data. These 4 map groups share the goal of allowing us, as individuals, to use public data to design our own experiences within our environments and to make use of public data directly on our own behalf.

Thesis Supervisor: Ethan Zuckerman

Title: Associate Professor of the Practice in Media Arts and Sciences

The Constant Atlas: Mapping Public Data for Individuals and Their Cities

By

Jia Zhang

The following people served as members of the dissertation committee:

Signature redacted

Thesis Reader _ ____________________

Sarah Williams
Associate Professor of Technology and Urban Planning
MIT, Cambridge

The following people served as members of the dissertation committee:

Signature redacted

Thesis Reader ___ ______________________

Sepandar Kamvar

Acknowledgements

This process would not be possible without the three inspiring and supportive mentors on my committee. I am deeply grateful for the time and effort you have given to this project.

Thank you to Ethan for being a tireless advocate for my work, and for being generous and insightful with your thoughts. You ability to crystallize my thoughts by drawing connections between ideas is endlessly helpful and never fails to amaze me. Thank you for including me in the civic family. Most importantly, thank you for being an incredibly kind advisor and friend.

Thank you to my reader Sarah for being a critical, honest, knowledgeable, funny, and warm presence throughout this process. I have learned and still have a great deal to learn from the inspiring work you do.

Thank you to my reader Sep for creating the beautiful works of art that introduced me to data visualization, and for giving me this opportunity to be a part of the Media Lab.

Thank you to my dear friend Elyse Graham, for being my meticulous editor and unofficial English teacher. Thank you for your expertise, your stories, and for having a sense of humor about how I write.

Over the past few years, I have been lucky to have the chance to work with members of the Civic Media and Social Computing groups at the Media Lab and also to share this experience with the incredible women in my dissertation working group. Jacqueline, Laura, Jen, and Cindy, thank you for offering your valuable critique, and camaraderie. Thank you to Yonatan Cohen, Neil Freeman, Darius Kazemi, and Allison Parrish for creating work that have inspired me, and for being so generous with your time and ideas.

I could not have completed this process without the support of Keira and Linda from the MAS office, your patience and knowledge were central to my experience.

Thank you Eric Ross Finkelstein for always being honest, working for what you believe in and for inspiring me to do the same. Thank you for being my partner in this wonderful life you have built for us so that I can follow through on this experience.

Finally, words cannot express how grateful I am for my mom, dad, and grandparents. It is my greatest fortune to have grown up surrounded by the most thoughtful and hardworking people I know.

Table of Contents

Chapter 1 Introduction

The distribution of the U.S. Census and mandatory participation in the Census became law upon the 1787 signing of the U.S. Constitution. Beginning in 1790 the Census has been collected - with few exceptions - every ten years[1]. As the backbone of the Federal Statistical System, the Census has the goal of counting, at regular intervals, every single person residing in the United States. When it was first conceived, the Census served two main functions: (1) to provide a statistical foundation for governance in a representative democracy in which the population of each state has accurate representation in the Federal Government, and (2) to enable the federal government to tax states in proportion to the general population[2].

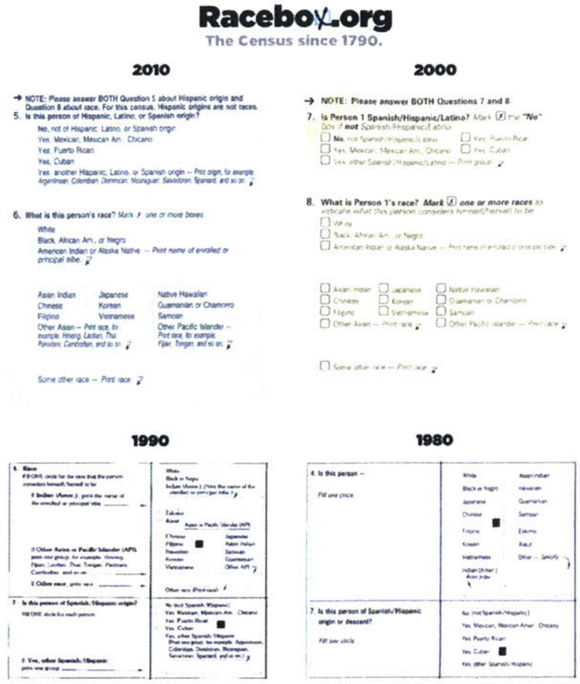
Over its 220 year history, the questions and categories that the Census contains have evolved continually to reflect changing cultural and political climates. For the example, the definition of race on Census forms has changed almost every time the Census has circulated. The artist Josh Begley's "Race Box" is a visualization of this change[3]. The project consists of a series of snapshots of the racial category question as it has appeared on the Census from 1790 to today (2010 is the date of the most recent decennial Census). In this work, Begley displays the cut-out question box about "Race" from each census form to show how the government has historically recognized race.

---

[1] Anderson, The American Census

[2] From the census.gov site

[3] racebox.org



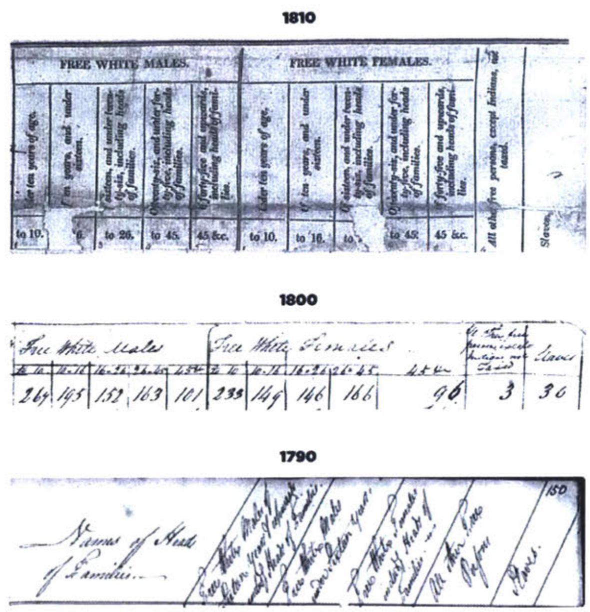


Figure 1: (left) The “Race” category on the four most recent decennial Census forms(years 1980, 1990, 2000, 2010).
Figure 2: (right) The “Race” category on the three earliest decennial Census forms(years 1790,1800,1810).

Until 1970, the Census was conducted by a team of Census takers that visited each home and filled out Census forms by asking questions of the residents. In 1970 the Census became, for the first time, self-enumerated: the new forms asked residents to fill out and mail in the forms themselves. When many Americans set a pen to the Census forms for the first time, some of them struggled to categorize themselves. For example, one of the questions that the form asked was “Relationship to Head of Household”. The question gave the resident options to define his or her relationship as “Wife of head”, “Son or daughter of head”, “Other relative”, “Roomer, boarder, lodger”, “Patient or inmate”, or “Other not related to head”. Married women discovered, for the first time, that they were not able to self-identify as the heads of their own households. In the subsequent Census of 1980, likely due to public feedback, the "Wife of Head" option was changed instead to "Husband/Wife".[4]

Amid all the other changes that occurred between one Census and another, the rewording of this category may seem like an insignificant detail. However, this is an example that tells the story of the difference between how we are counted and how we count ourselves.

Our ability to self-identify, as distinguished from *being* identified, in aggregate data is an especially important concept today in the context of so many projects that leverage big data to make decisions for us. Though more and more automated methods of data collection aggregate from our devices, our online behavior, and our consumer patterns arise, we are no longer asked to identify *ourselves* in the data collected from us. Yet how datasets make us available to be seen determines many aspects of our daily lives, from what advertising we see to what types of businesses are placed where we live. As a result, it is now more important than ever to understand how we can engage with aggregated data in thoughtful ways in order to better understand how particular datasets establish lenses through which we are seen, and also how we can make use of public datasets for our personal benefit.

I argue that we can derive two major benefits from enabling individuals to make use of data that is aggregated from them. First, doing so allows individuals to better understand how forms of data collection that are omnipresent in their social environment frame their physical environment. Second, doing so will demonstrate the value of understanding large public datasets, such as the Census' American Community Survey, in the context of everyday life. But in order to enable individuals to make use of this data, we must design tools that value

---

[4] Anderson, The American Census

individual concerns and perspectives, and that makes a primary goal of enhancing awareness of our environmental context.

In this dissertation, I examine how we can build relationships between individuals and aggregate data such as the American Community Survey (Census). I do so by creating a series of interactive maps that focus on creating representations of Census data that are qualitative, experiential, and personal. In this series of maps, quantitative data from the Census is re-contextualized by juxtapositions of satellite imagery, scaled through animation, and re-presented through the physical experiences of place. The goal of these maps is to develop a foundation for a future in which every individual has ability to use Census data - or other large, aggregated data sets - to be a useful tool in his or her everyday life.

Chapter two provides background to my dissertation. It discusses the works that I have made in preparation for this dissertation, and introduces concepts derived from my mapmaking experience that are central to my current research. The first section of this chapter reviews the youarehere project[5], explains the methodologies and source materials used, and provides an assessment of the project and the lessons learned. The second section introduces concepts in critical cartography and data humanism, a emerging concept within data visualization which focuses on the narrative design and personalization aspects of data, and explores how these concepts can be applied to future visualization and mapmaking practices. The set of practices embedded in critical cartography and humanistic data visualization, as well as the lessons learned from the youarehere project, inform the works that represent the bulk of this dissertation.

[5] youarehere.cc

In chapter three I introduce two maps that I created to deal with variations of scale within Census data. In them, I contextualize data of a location by visualizing its relationship with its immediate surroundings, as well as to the whole country.

Chapter four discusses two ways of using experiences to orient the data of the city: by movement and sight. For each type of orientation, the chapter discusses relevant work and then demonstrate a prototype that addresses the importance of visualizing data in the context of daily experiences in the city such as seeing and walking.

Chapter five discusses projects that use data to search for and make comparisons between places. These projects are organized into four broad categories according to the data they use: markets and demographics, images, physical characteristics, and the combining of datasets in location intelligence. It concludes with a pair of maps in which public data is juxtaposed with satellite images, and a discussion of the implications of using data to search for similar places.

Chapter six contains the final set of maps which deals directly with an individual's geospatial patterns - the paths they personally take in everyday life - and how these can be visualized through Census data. In addition, this chapter discusses projects and ideas that relate to the practice of self-quantification. It presents a series of experiments that use Census data for self-quantification.

The concluding chapter discusses design principles emerging from the series of maps produced for this dissertation and also future directions for work that involves the production of an individually driven public data atlas.

Chapter 2

You are here

> *"It is a truism to say that how we design cities depends on how we understand them."*
>
> \- Bill Hillier *Space is the Machine*

Figure 3: Home page of youarehere.cc introducing the project.

In early 2014, the Social Computing Group initiated the "youarehere" project with the intention of creating one map of a one city everyday over the course of a year. Our team at the Social Computing Group was diverse and included computer scientists, architects, and designers.[6] Despite our different disciplines, we agreed that in order to build future interventions that improved the environment around ourselves and others, we had to first understand the diversity of factors that made each place unique. Our collective goal was to use this prescribed daily mapping exercise to open up new lines of inquiry in how data can be used to describe places that we lived, and places that we hoped to learn about. We theorized that mapping places with relevant data would allow us to see patterns in which we and those living in the mapped communities could intervene to remedy the issues that the data revealed. For example, by creating a map of bicycle accidents, and also representing the conditions under which they occurred, we could then identify area in which bike lanes would be the most beneficial[7]. Through this endeavor, I came to understand the work of mapmaking in way that did not fit into traditional cartographic education, yet was intuitive at the same time.

There are many different ways for a person to know a place, with or without data. Geo-epistemology, or how people know the things they know about the physical environment, is made of both first hand and data-centered knowledge.[8] Every tool people use to navigate places, whether it is human memory or a GPS device, shapes how they interact with those places. For our "youarehere" team, web-based maps that communicated various datasets related to everyday life was a way to bridge two ways of knowing: the way residents already understood the places they lived, and the broader context offered through introducing a data set. In the instance of the coffee shop map, local residents know where the coffee shops are,

---

[6] Initial youarehere team: Yonatan Cohen, Wesam Manassra, Pranav Ramkrishnan, Steve Rife.

[7] Bike Crashes Map by Pranav Ramkrishnan

[8] Rankin, After the Map

but mapping the "watershed" of those shops help people think about whether they gravitate to what's closest, what's enroute to work, what allows them to make choices based on functionality or taste.

There are also many different ways that planners and theorists make sense of a place. In “What is a City?”, sociologist Lewis Mumford saw the city as a social institution where the organization of physical features must be made not only to satisfy the housing and transportation needs of the population, but also serve to strengthen the dynamic relationships found within neighborhoods and communities.[9] The view that a city is a social system and should be treated as such is not unique to Mumford.

When Kevin Lynch conducted his landmark study in 1960 of how citizens perceive Boston, he formulated a language of nodes, links, and edges with which planners and residents alike can use to articulate and frame their perception of neighborhoods.[10] Through this system where nodes are places of interest, links are connections, and edges meant barriers, Lynch's language for the city highlighted the close relationship between physical barriers and conceptual ones. Today, using data about the time it takes to get from one point to another (transit efficiency), we could see what Lynch described, that the physical connectivity of places drives our idea of neighborhood boundaries. Places that are geographically close together but difficult to travel between feel like they are in separated places.

For Jane Jacobs, liveliness, a central condition for livability, was driven by physical, social, and temporal diversity.[11] The opening and closing hours of a bookstore could be the anchor to many complementary businesses, and might indicate how active an area remains at night. The mixed

---

[9] Mumford, “What is a City?”

[10] Lynch, The Image of the City

[11] Jacobs, The Life and Death of Great American Cities

uses for a public park throughout the day could be a reflection of the diversity of nearby building uses, or the diversity of uses for an area could be an indicator for new urban interventions such as public parklets (the where to place map).[12]

Both Lynch and Jacobs, mentioned above, framed their descriptions and studies of the city as a living system. In mapping data to communicate a data-centered understanding of cities, our maps were design for those who lived in or might want to visit neighborhoods rather than those who built them. Therefore, we realized that when making maps for those living in this system, the task of mapping data for citizens is to clarify the roles they take on in that system as individual participants in the places they live.

Drawing on the experiences and ideas of both theorists and our own roles as residents, we took datasets that would either express particular aspects of our daily lives (such as the safety of commuting by bike), or provided opportunities for comparisons between places or time periods within the city's history (such as the diagnosis for diseases for a city's hospitals and the registration of businesses). Over the course of the year, we were able to codify the making of these maps into more formal methodologies for using data to describe cities. The following section describes the maps we produced, the data we used, and the lessons we learned.

Maps

The maps we made covered topics such access to foods, the location of noise complaints, and the opening hours of businesses, as well as the physical structure of cities. Below is an overview of the themes and data sources we took on.

---

[12] The Case for Parklets

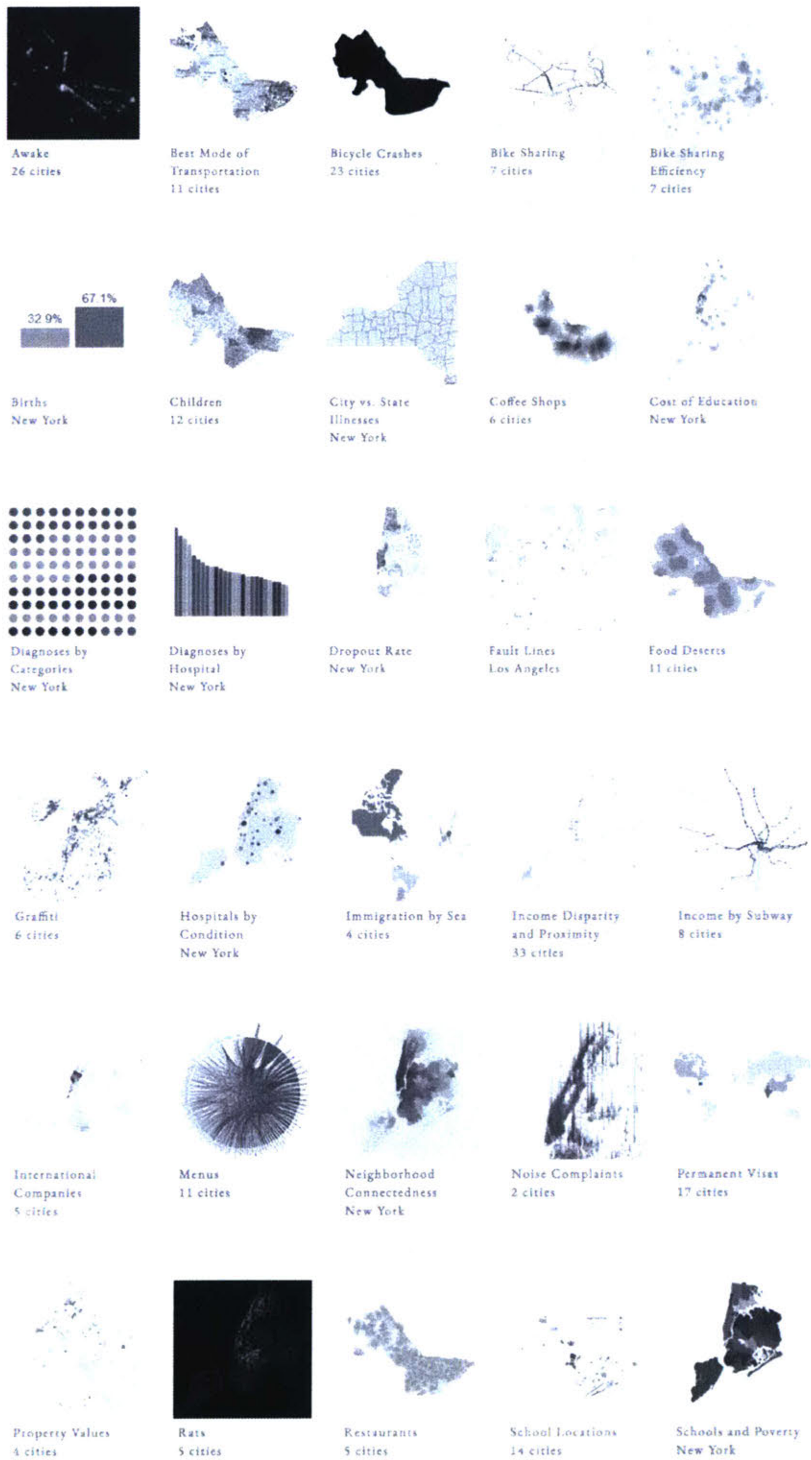


Figure 4: The map thumbnails from the youarehere.cc homepage shows the diverse forms that our maps took.

Themes

Among the many themes our team addressed through youarehere, I worked specifically with two groups of topics. The first considered impacts on our daily quality of life, which included food, noise, and our commutes. The second group included topics that described relationships between areas, for example the differences between areas as seen through income, or the growth of cities as seen through business registration records. Both the daily impact and the systematic view of public datasets continue to be central in the maps I have continued to make for this dissertation.

Two particularly relevant examples of the types of maps that I pursued independently are the "Noise Complaint" map and the "Fault Lines" map. The "Noise Complaints" map is a visualization of New York City's noise complaints categorized by human and machine noise. New York City is full both of noise and also complaints that report noise (image below). In mapping noise, I saw that it was not enough to simply plot the locations of complaints, but also to have the complaints reflect something of the city. By color coding the complaints by whether they were caused by machines or humans, we can see how different Manhattan is from the other boroughs in the city. Additionally, we see the geographic concentration of machine noise in blue versus the more persistent human noises in red.

In the "Fault Lines" map, I highlighted a city's internal borders by visualizing the distinct differences between adjacent block groups in Los Angeles and other cities (image below). Diversity and geographical proximity make cities complex places where differences are constantly being negotiated. The map is a directory of liminal spaces within the city; places that mark considerable statistical difference between adjacent urban areas in the categories of education, race, transportation, income, and occupancy. Through this visualization, we can see how some areas – specifically Census block groups - exist in isolation from neighboring block groups in terms of certain aspects of the data, and also where more homogeneous areas are within a city.

Figure 5: The Fault Lines Map highlighting neighboring areas with the biggest differences in educational attainment.

Figure 6: "Noise Complaints Map" of New York City.

Data

For the purposes of "youarehere" and projects going forward in this dissertation, all the data used come from sources and platforms that are visible to the public (see appendix of data sources). *Public data* refers to datasets that are directly open for use and analysis with no restrictions, or information that is in the public domain. In the context of this project, examples are datasets found on open data portals hosted by a branch of the government, as well as PDF files that are published by the government. *Private data* refers to data that have been obtained through APIs or through publicly visible websites. Relevant examples here are datasets on the travel time between two points obtained from the Google Directions API, and listings in the classifieds website Craigslist. While both the quantity and quality of publically available geospatial data has improved in recent years with the development of open mapping data networks such as Open Street Maps community, our devices still utilize largely proprietary data: the Google and Apple mapping platforms.

The openness of datasets was crucial to our project. As we experienced first hand, both public and private data have varying degrees of openness. In the case of public data, we considered openness as defined by the five star scale from Tim Berners-Lee, includes the online availability and licensing, structure (is it structured to be machine readable?), format (is it structured in a proprietary format?), the use of URIs (Uniform Resource Identifiers for easy reference as metadata), and context(for linking to other data).[13][14] These factors contributed to our selection of datasets and map subjects. For example, we were able to map the health data of New York

---

[13] The Open Knowledge Foundation's definition of Open Data

[14] The 5 Stars are from Berners-Lee's 5stardata.info

City and State because of a recent publication of well documented records. The accessibility of the datasets at times determined what we prioritized in mapping and allow more open datasets to be easily and readily mapped. It was an important lesson that the openness of data is not an end, but rather a crucial part of its quality, and the determining factor on its power.

For APIs and website data that are private, a similar scale as above could be used to assess their openness. In our context, the openness of private APIs were expressed in the limitations in the terms of use (availability, cost and licensing), the accessibility of data (structure), the flexibility of methods/programing languages allowed to access the data (format), the use of codes and entity names that are transferable between datasets, and the way data is linked to other relevant information. These limitations often determined the extent to which we are able to map a subject. For example, the granularity of locations (the grid) we can use to map connections between places is dependent on the number of queries we can make over a reasonable period of time (i.e., how many queries we were able to make using a free API key).

Public Data Sources

We use three main sources of data, the U.S. Census, the open data portals hosted by individual states and cities, and the enigma.io public data platform. For Census data, which we used to get population density, age and other general demographic factors, we used the Social Explorer exporter as well as the government's own American Fact Finder. We were able to obtain different types of public datasets for New York City from the city's open data portal. Taking advantage of the comprehensiveness of the city's open data project, we used the available datasets in the portal as a guide to search for and request similar datasets from other cities.[15]

---

[15] opendata.cityofnewyork.us/data/

The enigma.io platform was useful in obtaining datasets that are otherwise scattered across different government agencies, such as business registrations, immigration, and visa data.

Private Data Sources

Private data sources are acquired using either APIs or web scrapers.

APIs included:

Google Addresses

Used to label locations with their street names to make the map metadata more easily readable. Incidents such as bike crashes or 311 calls which had latitude and longitude that required this process.

URL: https://developers.google.com/maps/documentation/geolocation/intro

Google Streetview

Streetview images were used to look at colors, parking, trees on the street, as well as to visualize marathon routes along streets.

URL: https://developers.google.com/maps/documentation/streetview/intro

Factual and Google Places

Factual is a location intelligence company that provided us with business listings and opening hours. We also used Google Places to determine some opening hours.

URL: https://my.factual.com/data/t/places, https://cloud.google.com/maps-platform/

Google Directions

Google directions was used extensively to determine the time-distance, best mode of transportation, and connectivity between places.

URL: https://developers.google.com/maps/documentation/directions/intro

Web Scrapers (written by me):

Craigslist

Used to obtain listings for 20 major cities to compare the jobs, housing, dating, and for sale sections.

Menupages

Used to obtain menus for restaurants in the city to visualize what dishes and ingredients were found in each city.

Yelp
Used to obtain reviews to determine whether restaurants were reviewed by local or visiting users.

Units of Measurement and Their Representations

A significant outcome of our mapping project was knowledge of how cities can be represented through various datasets and where existing datasets fail us. Cartographers are familiar with the modifiable areal unit problem (MAUP). MAUP is a common issue encountered by cartographers and statisticians in the determination of proper scale of representation and tabulation for a particular dataset.[16] When data is combined into larger units in the process of aggregation, the ways in which discrete points are grouped has the potential to distort the final representation of that data. A common example that those who do not work with cartography regularly would be familiar with is gerrymandering, in which boundaries with particular voting demographics are manipulated by grouping them in a highly specific way so that the resulting areas create advantages for a particular political party. As mapmakers with varying backgrounds outside of cartography, we also encountered the same issues of statistical bias inevitably caused by spatial aggregation in resolution (or scale) and zoning.

We explored this issue in the context of mapping cities and found a comparative and critical approach to discussing units of measurement and their representations. As a first step, we created multiple preliminary representations for the same data for initial assessment. The visualizations we tested included markers, sheds (heatmaps), areas (choropleths, dot density, heatmap grids), and edges (boundaries of administrative areas and physical streets).

---

[16] Wong, The Modifiable Areal Unit Problem

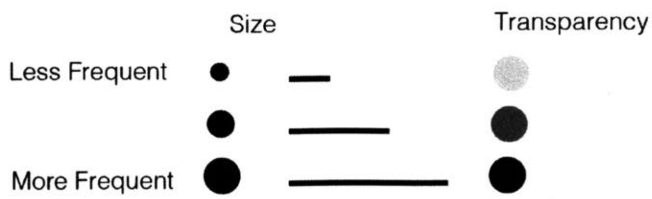


Figure 7: Diagram of the markers we used for maps.

Geometric markers were used when a entry in the data described a specific coordinate (location of a single bicycle crash) rather than an area (the number of crashes in a neighborhood). [17]

For data where geographical areas such as zip codes, census blocks, or neighborhoods are clearly defined by the data and assigned values, we used choropleth maps. We also used dot density representations when data is of a nature where it is appropriate, such as in the case of populations and households (but not in the case of bicycle crashes). To aggregate data that are not aligned with administrative boundaries (bicycle crashes), we employed hexagon or square grids to create heatmaps. When data is available, shading was added to area maps to create sheds that indicate an additional dimension. For example, within an area that is serviced by the same subway station, the area is shaded according to the amount of time it took to reach that station.

[17] Slocum, McMaster, Kessler, Howard, Thematic Cartography and Geographic Visualization, P.63

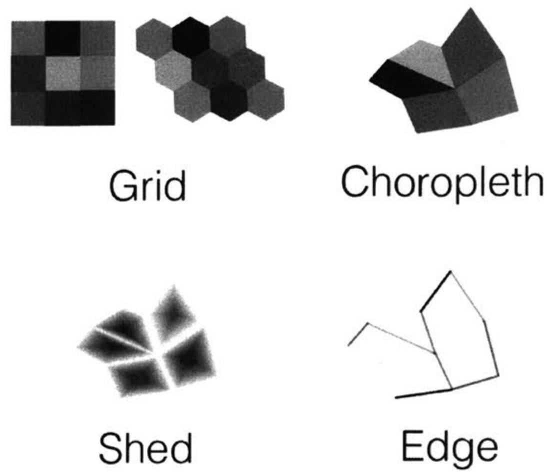


Figure 8: Diagram of the area visualizations we used for maps.

Edges were used in two ways. The first is to highlight differences between two areas by drawing attention to their shared boundary (income discrepancy map). The second is to use physical features such as streets to summarize data. For example, we created a histogram to provide an overview of safety by street by mapping the bicycle crashes to specific streets and highlighting the most dangerous streets.

In the process of visualizing the same dataset in a variety of ways, we found that communicating a dataset clearly often required a hybrid interface design where data can be view at different scales and both qualitatively and quantitatively. Using the bicycle crashes maps (below) again as an example, we create three particular views of data to compose the map. The first plots the dataset itself to locate the crashes (left), the second to provide an overview by summarizing the number of crashes by street (center), and the third to give a qualitative view of the crashes by adding a window of the street view of the crash location (right).

Figure 9: The Bicycle Crashes Map contains three different views of data, geospatial, summary, and image.

From mapping each dataset, we came away with the understanding of the various units of measurement imposed on places. Among the units of measurement we encountered within our data, below are ones we found particularly useful in capturing the experience of cities in the process of applying data to maps.

Spatial dimensions of Measurement

Technology has historically changed our perceptions of distance.[18] Quotidian technologies, from social networks to news coverage to ridesharing, constantly alter our perception of proximity and the level of complexity we encounter in a place. Understanding the differing demarcations and measurements on space made by these technologies is a significant part of studying how we make use of places.

- **Digital neighborhoods** are areas in a city that have concentrations of geo-located social media activity. Studies have found that common culture and language connects different parts of the city that are otherwise physically distant. [19]

[18] Virilio, Speed and Politics

[19] Anseline and Williams, Digital Neighborhoods

- **Historical units** are formed around common heritage such as architectural character or cultural legacy. These are important in many cases, such as the historical connection of a neighborhood that has been fractured by the building of a highway, or the impacts legacy of zoning on a particular development.[20]
- **Real estate units** are based on the naming conventions dérived from real estate trends of areas that have seen population and price increases and demographic shifts. These naming conventions are subjective, constructed primarily for commercial purposes, and are often highly contentious.[21]
- **Time-distance based units** are determined by the walking, transit, and driving ranges using time distance rather than physical distance. These take into consideration the infrastructure of roads and transportation, which are unlikely to change in short terms. Time-distance, in conjunction with cost-distance, describe the ease with which one is able to travel between two neighborhoods, cities, and even countries. The social, cultural, or financial connection between two places often defy the physical distance between them. The walkability of an area often determines the connectivity of amenities within that area and ultimately contribute to the liveliness of its streets.
- **Amenities- determined units** are formed by a division of different entities such as coffee shops, supermarkets, and houses of worship. The proportion of different types of amenities are a highly visible way to characterize an area.[22]

---

[20] Spaan, NYPL Space and Time Directory

[21] Woodruff and Wallace, Bostonography

[22] Hidalgo and Castaner, The Amenity Space and the Evolution of Neighborhoods

Temporal Dimensions of Measurement

Some datasets have temporal dimensions that are important to the characterization of places. Temporal units can be described by cycles of activity and hibernation as well as by extended periods of change. Ephemeral units can be no less important than ones with more permanence. For example, sides of streets can be divided as separate units because of how places behave differently according to the quality of light they receive. Our team discovered this first hand when walking down Newbury Street in Boston on a sunny fall day and saw that pedestrians and shoppers congregated on one side and shifted with the sun. [23]

In hourly, daily cycles and weekly cycles, the flow of people and quality of light change the immediate landscape of the city, while in seasonal cycles, the physical qualities of the city change in terms of weather, greenery, tourism and annual events. In even longer periods of gentrification and renewal, demographic shifts occur along with the economics and physical growth and decline of the city. Infrastructures for transit transforms neighborhoods but require many years for the transformation to be complete. While a small public garden is planted in a day and changes as the flowers bloom, its impact on the traffic in a neighborhood can be seen over months or years. The transit infrastructure drastically alters the image of the city, while the garden only the moods of those who walk by.

These episodic and rhythmic cycles are driven by both nature and human activity.[24] In some cases, changes to our environment occur at such a gradual pace that we do not readily sense it

[23] We later found relevant technical work that deals with quantifying sunlight density. Mapping the Shadows of New York City, published by Bus and White of the New York Times, with research conducted by Silva and Doraiswamy from Tandon School of Engineering at New York University

[24] Lynch, What Time is this Place? P.168

without mediation.[25] Data allows us to map both the trajectory of business growth in an area over a long period of time, and also map smaller cycles of activity reflected by the opening and closing hours of the businesses that populate an area throughout the day. The health of a place as described by the longer trajectory of progress and its daily cycles of activity are interdependent and equally crucial to our understanding of a place.[26] These temporal and often ephemeral qualities in data brings new dimensions to the traditionally static form. These temporal units of measurement also provide groundwork for the visualizations for individuals over time.

Measurements Summary

Temporal and spatial units of measurement can be combined to provide a more evocative picture of a place. Utilizing the relationship between datasets, scales, and perspectives is an important aspect in describing a place with layers and nuance. Qualities of place are dependent as much on the activity of its inhabitants as it is on physical structures. This intersection of individual will and infrastructure is a unique opportunity for the filtering of aggregate datasets through unique perspectives and stories.

---

[25] Lynch, What Time is this Place? P.184-187

[26] Jacobs, The Rise and Fall of Great American Cities, P.150

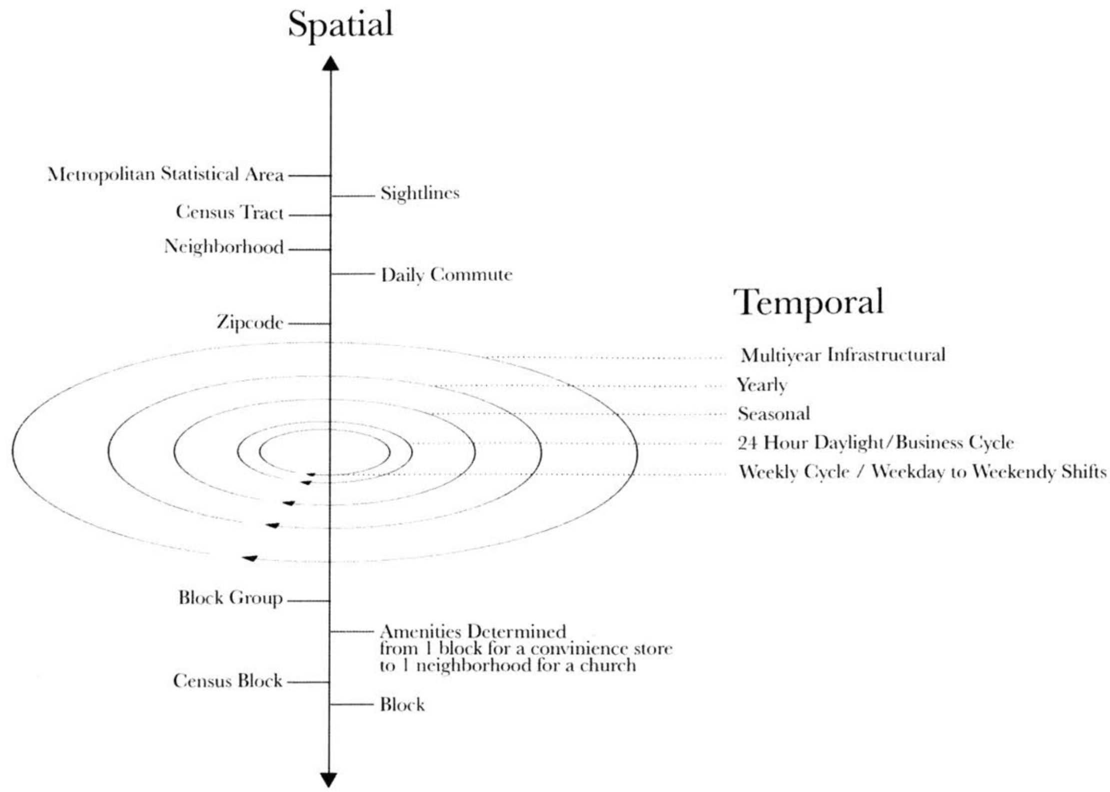


Figure 10: A overview of spatial and temporal units as they are matched to experience

Assessment

Over the course of the year, we distilled what we learned firsthand and tested established design principles to construct a design pattern with which we can build future visualizations. Five initial and central principles that I continue to work with are explained below.

1. *It should be easy to go from the micro to the macro, from the entire city view to the view of the individual home/place and back again.*

We started with the popular mantra "Overview, zoom, details on demand" from Ben Shneiderman's landmark 1996 paper on visualization tasks.[27] When working with data of the city, we needed to go between different scales to better understand the data in the visualization

[27] Shneiderman, The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations

process itself. We confirmed the significance of Shneiderman's mantra by observing our users in the lab when we demonstrated new maps. From these user observations, we saw that interactions with scales are often nonlinear, and users often go back and forth between overviews and small details. To make this type of interaction possible, we established the second and third principles.

2. *Information elements should be navigation elements and vice/versa.*

To aid the smooth transition between scales and views of data, it was crucial that almost every visual element on screen had an interactive function. Most commonly seen on our maps are the use of histograms, markers, and other parts of the information display as filters.

3. *The visualization should give several different views into the data: e.g. the map, a ranking/histogram, and a qualitative view when possible.*

Demonstrated previously in the Bike Crashes map, we presented multiple views of the same data within each map. In most cases, this meant having a direct translation of the data onto maps, a summary view to communicate patterns and trends, as well as a qualitative view that gave the user an idea of what areas look like physically.

4. *The visualization should aesthetically evoke the underlying data.*

We struggled initially with visual design. It is difficult to build visual coherence out of many mapmakers and varying subject matters. After discussions and design reviews of early maps, we arrived at the notion that visualizations should give some indication for the subject matter through its aesthetic. For example, maps of trees were green, and maps on migration included motion. We used the subject matter as a primary visual design constraint and chose to simplify all other elements in design.

5. *Every map should have a clear story that it tells.*

Storytelling is at the heart of our project to communicate a sense of place through daily mapping. Although each visualization implicates a specific perspective, it is not always the case that the framing and editorial choices are made clear to the viewer. We knew that the best way to make our data structuring and visualization processes transparent was to integrate elements of them into the maps. We felt that how we arrived at the final map was a big part of the story we wanted to tell. For example, when possible we used histograms we relied on during data cleaning as a part of the final interface. However, the level of transparency we were able to achieve visually did not always reach our standards. For all maps, we included a text explanation linked from the subtitle of the map. The text for each map closely follows a template that answered three questions: What the map shows, why we made it, and how we made it. This template was our way to demonstrate that our motivations for every map were closely tied to our design decisions.

More telling were perhaps the principles we believed to be valuable, but struggled to realize. There were two main goals that we were unable to fully achieve reach with our maps. These challenges are central to the questions I have taken on what I hope to address in the dissertation research.

1. Make each map tell a story by re-orienting the viewer's knowledge of the city.

We wanted each map to be able to answer a question specific to the user. We wanted to turn the reading of our maps into an active encounter where the maps can serve as tools for the user to find data. We learned that interactivity doesn't always guarantee it can be useful as a tool to navigate data Reframing maps from artifacts with positions of authority into objects to critique and think with was a difficult challenge.

2. Use the cumulative power of interactions with maps over time to build relationships between individuals and their environment through data.

We were able to cover a wide range of topics with the roughly 150 maps we published over the course of this project. Internally, the process of mapmaking generated new ideas. For example, the connectivity of neighborhoods led us to think more deeply about transportation and explore the best modes of transportation between neighborhoods. However, we did not communicate these connections between maps to our users. We used a simple database structure to hold our maps and listed maps by city or by theme. This straightforward organization was oriented towards searching and referencing, but not storytelling across maps.

The subsequent work is partially inspired by turning these last two principles into goals. Both of these goals indicated to me that rather than considering a map as a snapshot in the system of measurements we had framed to describe the city, I would instead be considering maps based on the experience of an individual user or resident.

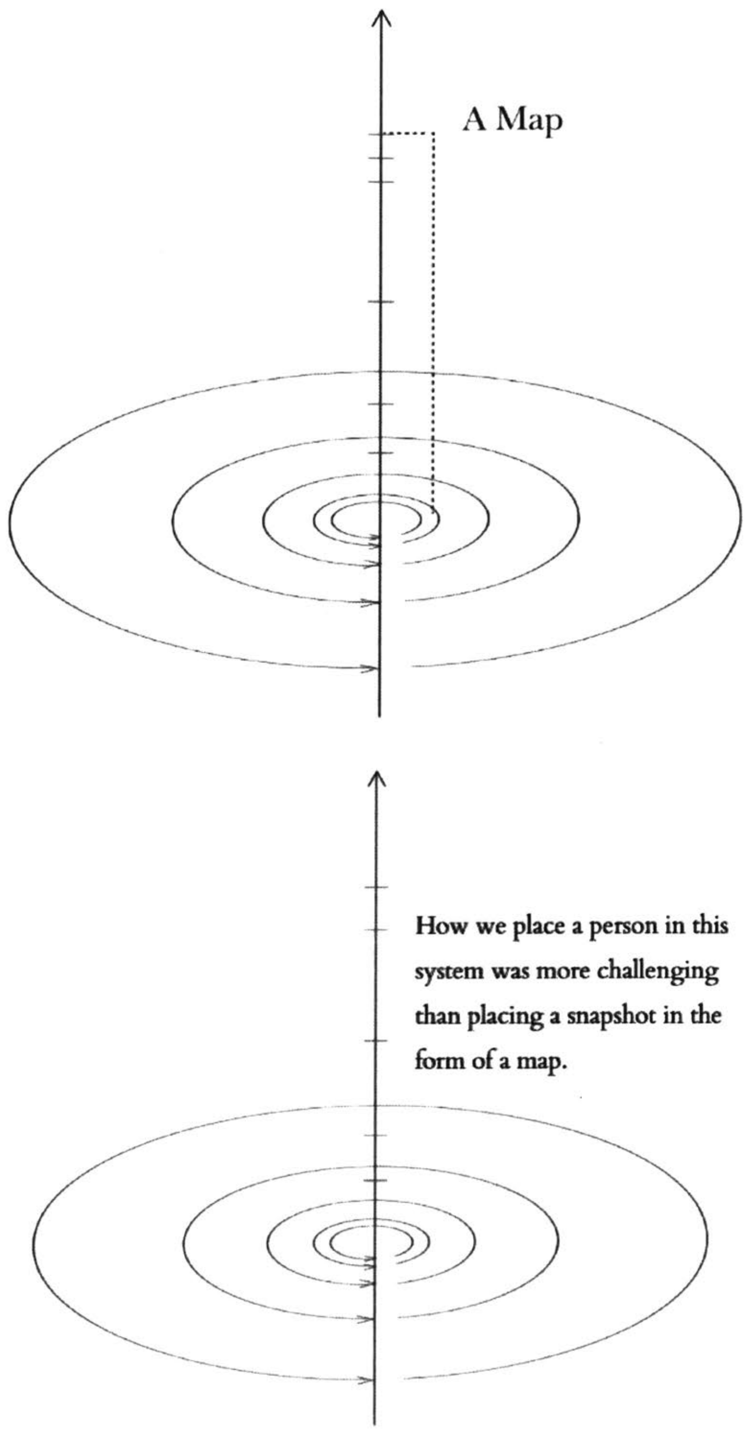


Figure 11 - 12: We initially used the system of measurements we framed to create maps that described the city at different spatial and temporal levels. Going forward, I will be using the similar framework to discuss the experiences of an individual rather than a map.

The Census

The maps at the center of my dissertation are based on data from the US Census. Throughout this dissertation, I have made use of Census data from three main outputs: the decennial Census, the American Community Survey, and the Public Use Microdata Sample (PUMS) of the American Community Survey.

The decennial Census is required of every person living in the United States and includes basic questions such as age, race, and gender. The American Community Survey, which is not comprehensive of the population, surveys a sample of 3.5 million Americans on a yearly basis. The survey contains detailed questions on both persons living in a household, and questions on characteristics of housing. The survey asks questions about a person's quality of life including a person's profession, income, education, commuting time, marital status, and health insurance status. Data from the American Community Survey is used as a primary dataset in all of the maps contained in this dissertation.

The PUMS is an untabulated sample of the American Community Survey. This sample preserves individual responses and is not summarized by geography. The PUMS dataset is only used in the @censusAmericans Twitter bot.

The Census is the foundation for many private data enterprises. According to the 2017 Hamilton Project report on government data, public (government collected) data has always been used to ground private data collected by business or non-governmental organizations. For example, the American Community Survey has been used by businesses in the process of relocating or planning expansions to look at skill markets, age groups, and economic indicators associated with potential locations. In the age of big data, the collection and analysis of data are constantly ongoing. Scalability, the ability to predict and target data insights with precision, and the speed of the turnaround are prized characteristics of data collection and processing. Within this context, the Census is distinguished by being comprehensive, consistent, and credible.

We can see, in the distribution and subsequent wider use of Census data, as well as its development of the American Community Survey, how the Census has changed technologically over the years, as well as the impact of those changes. The development of the Topologically Integrated Geographic Encoding and Referencing(TIGER) system and the use of GIS software in the 1980s allowed the Census Bureau to administer, tabulate, and distribute results using uniform maps. In the 1990s, Census data began to be distributed on the internet via the American Factfinder, a web interface for retrieving Census data that is still in operation today. As digital dissemination became the default method of accessing Census data, the potential of digital aggregation to improve methods of collection gave rise to plans for a rolling Census. The result, the American Community Survey, is a detailed survey that is sent out on a rolling basis every month to gather information on 3.5 million Americans every year. It was established as a country-wide program only in 2005.

As problematic as self-enumeration can be, there is also great comfort in being counted by our own definitions and in our own terms. Many private datasets do not afford us with such agency, collecting data instead from the automated output of our devices, transactions, and behavior online and in the physical world. The ability to define explicitly how we are counted and included as a part of the American public is important. This distinction makes the Census a unique public good.

Public goods have two defining characteristics, non-rivalry and non-excludability. Non-rivalry means that use by one person does not diminish the ability of others to share the same benefits of the public good. Non-excludability means that its utilization cannot be limited to specific parties. Census data, which is made publically available for use by many parties at the same time, and distributed as a free resource should - in theory- fulfill criteria of the definition. However, as it is currently utilized - by businesses and institutions rather than individuals - the Census falls short of its potential to provide contextual knowledge of places in the United States to everyone with the same level of accessibility.

As a public good, the Census could potentially benefit individuals by enabling them to provide data that grounds decision-making for local governments and businesses. Census data is presented to users as datasets, and also as maps in the Census atlas. The first official Census map gave data for the year 1850; the most recent gave data for the year 2010. The information design in these maps has largely remained the same in the official Census atlases throughout this time. Below left is a Census of wealth distribution from the 1870 atlas, and on the right is the "Median Household Income" map of 2010's atlas.

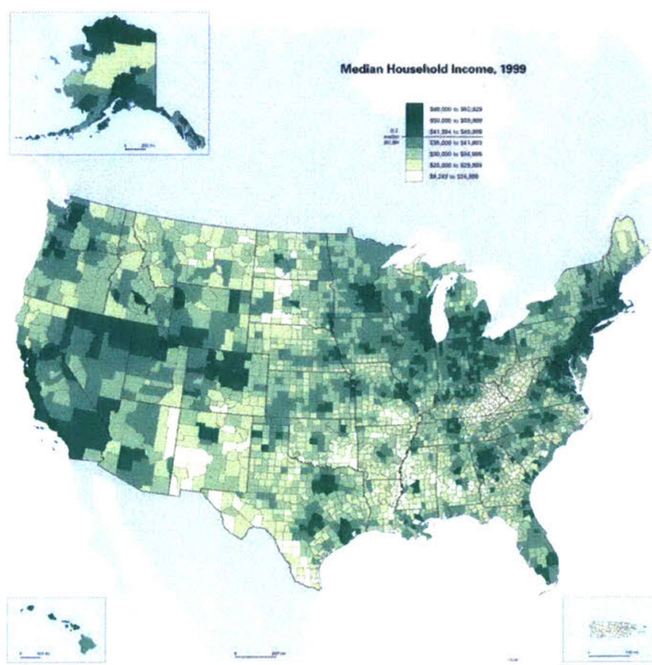


Figure 13: **(**Left) Census map of wealth distribution, 1870.
Figure 14: **(**Left) Census map of "Median Household Income", 2010.

Innovative and humanistic ways to communicating Census data can often be found in work produced with specific users in mind. For investigative journalists, there is the Census Reporter project, which "helps journalists navigate and understand information from the U.S. Census bureau." For New Yorkers, the local public radio station WNYC uses the Census as the lens to discusses the changes occurring in different parts of the city in the "Your Anecdotal Census" series, covering a different county in the New York area (ncluding parts of New Jersey and Connecticut that are adjacent.)

As a part of the "Your Anecdotal Census" radio show, local leaders, residents and demographic experts discuss changes in where they live based on their differing knowledge and expertise. A common conversation on the show starts with a listener calling in to describe a particular phenomenon that he/she has witness first hand while living in the county, and Census data experts responding to the listener's observations by contextualizing it with data and trends in found in Census data. The conversations are engaging because they originate from a local resident's unique perspective and become ways to ground the conversation on what larger trends in data show.

Design principles of maps for people

As the medium of our information consumption shifts from paper and TV to computer, tablet, and smartphone, interactive visualization has become an increasing part of the way we consume that information. Visualizations of such data as stock market trends and weather forecasts have long been a part of our daily diet of information, but they are increasingly ubiquitous.

During the past election cycle, visualizations of forecast data in the form of confidence barometers and electoral maps were also added to the front pages of major publications. Front and center on election night 2016 was the New York Time's forecasting needle, which swung wildly as votes were counted. As the night went on, the needle twitched its way across the screen from a 80% Clinton victory to a 95% Trump one (image below). The needle graphic, highly criticized by readers and dubbed “a realtime panic meter”, became known for its ability both to reflect and to cause anxiety in voters. However, by the time it returned a year later for senate races, the needle itself had become an election night staple and the subject of discussion for readers.[28]

Figure 15: The election needle that was first used on the night of the presidential election in 2016.

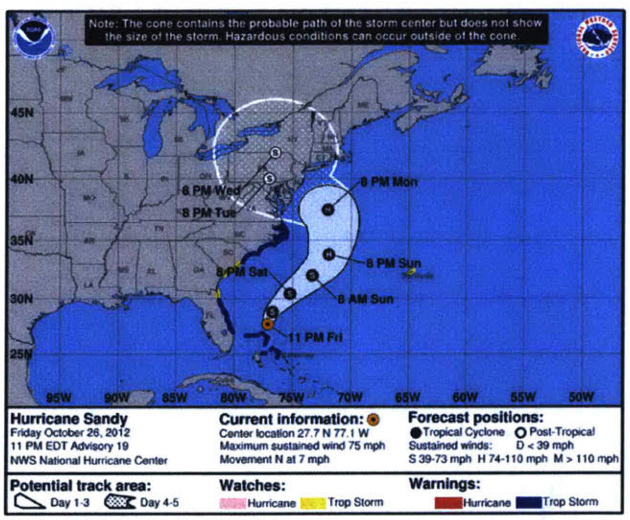


Figure 16: The Hurricane forecasting cone used by the National Hurricane Center to forecast the paths of storms. [29]

The needle is significant because it symbolizes a turn in the way we represent data with graphics. The reason the needle shifted back and forth so dramatically and frequently over the course of the night was the result an important editorial decision made on the part of the Times. The Times purposefully designed their graphic to directly reflect the uncertainty of the real-time

[28] The New York Times needle was first published in 2016, and reused in December 2017

[29] The cone of uncertainty describes the evolution of the amount of uncertainty over the course of an event. Definition from the National Hurricane Center.

data. In their design discussion, the editors referenced a more familiar representation of uncertainty, weather forecasts. They point out that the cone of uncertainty, commonly used in describing the paths of hurricanes is an indispensible part of our visual language. The needle is, in essence, a graphical interpretation of the same cone. Representing nuance, uncertainty, and subverting the objectivity traditionally attributed to charts and graphs is a new and necessary challenge in the visualization of data outside of our weather system and elections as well.

Data humanism and Critical Practices in Cartography

Data humanism, a term that the visualization expert Georgia Lupi has written and spoken at length about, encompasses some crucial new values emerging in the field of information visualization.

> *"Every act of perception is necessarily a highly directed and selective affair, whether the guiding principles are conscious or inadvertent."* - Martin Kemp

Data humanism describes the reorientation of the data visualization field from its main task of simply communicating data to connecting knowledge to behaviors and people. The term can best be summarized as the drawing by Lupi below.

Figure 17: A drawing accompanies Lupi's original article on the concept of Data Humanism. [30]

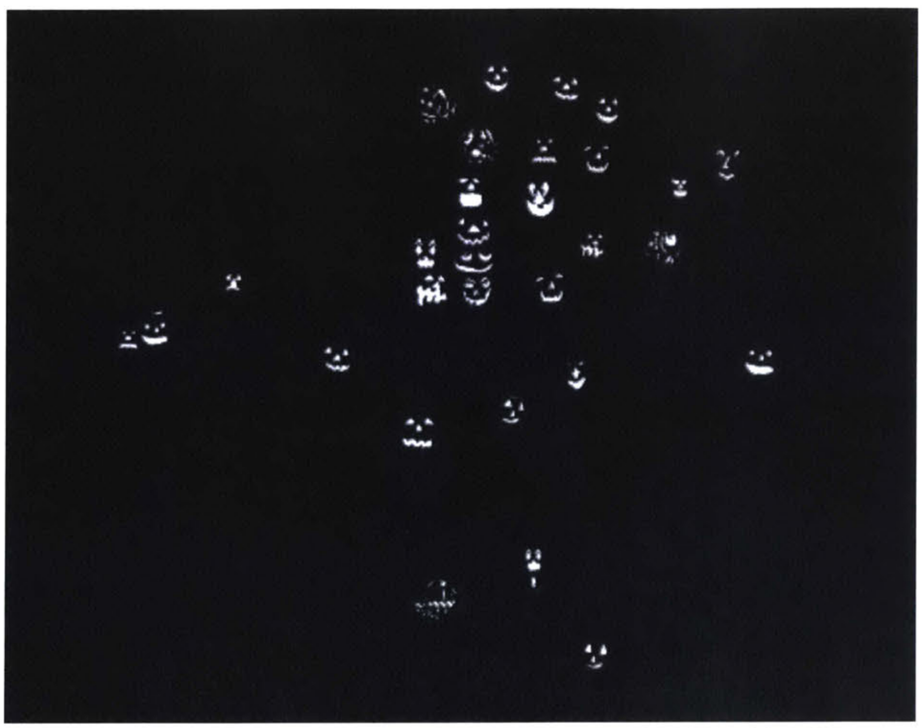


Figure 18: Wood's map of pumpkins on front porches in Everything Sings, a atlas of his Boylan Heights neighborhood in Raleigh, North Carolina

The ideas found in Lupi's list are similar to those explored in critical cartography and experimental geography. Both data humanism and critical cartography present highly personal and subjective ways to deal with data. Experimental geography deals with the notion that

[30] From Lupi's Medium article: Data Humanism. the Revolution will be Visualized.

cultural production is inextricably tied to the physical production of space.[31] By building experiences (films, performances, music, instructions), spaces (observatories, ditches, walls, windows, etc.) or tools (augmented reality, maps, GPS applications) that alter our perspective of the built environment, works in this field have the potential to bring forth conditions (preconceived notions and behaviors) that otherwise go unquestioned. Particularly relevant to the work at hand are the maps of critical cartographer Denis Wood, the idea of "geographical intimacy", and the Situationist concept of the dérive.

The themes of small, qualitative, imperfect, and human data are especially resonant in Wood's work. Wood's atlas *Everything Sings* demonstrates the highly selective nature of maps by using formal cartographic language to depict his hometown with idiosyncratic data. In doing so, the top down viewpoint that give traditional maps so much power is subverted by the content. More important to my research, Wood's topics are vernacular, the maps evocative and memorable because they are driven by a slowly formed familiarity with place that is based on first hand experience. Wood's *Everything Sings* and Lupi's data humanism definition use mapping or visualization to provide a critical lens for people to view their highly personal relationship with their experiential and conceptual environment. [32]

In mapping and visualizations, the critical lens also highlights the fact that data never speaks for itself. It speaks through a prescribed narrative and from a specific perspective. Both are dictated by the author of the visualization or map. The following sections will further explore perspective (and personalization) and narrative as digital texts shift from being determined by authors to being shaped by readers.

---

[31] Thompson, Kastner, and Paglen in Experimental Geography

[32] Posavec and Lupi, Dear Data

Personalization of aggregate data

It's been said time and again that we have unprecedented access to public data as citizens. However, how well we digest such data is still a point of contention. In this dissertation, I propose that one way for users to understand aggregate data is to create ways of connecting large datasets to personal experience. This means both allowing users to find themselves in the data, to make use of the data by posing their unique questions to it, and having repeated encounters with the dataset over time. The idea of the dérive, which stems from experimental geography practices, further illustrates the power of making personal connections to large phenomenon.

For the Situationists, experiencing the city through daily acts such as walking can be a way to change one's mindset and one's perspective of a familiar environment. The practice of a dérive, developed by Guy Debord and the Situationists,[33] asks urban residents to become conceptually untethered to daily purpose by physically becoming lost. It asks its participants to take long unplanned walks and be directed by physical and social encounters as they occur rather than to fulfill specific functions in their day. The practice is a purposeful disruption of the behaviors that are imposed on residents of a city through the city's physical infrastructure. In the context of interacting with data, a dérive among a dataset can be a series of links that takes the user along an unexpected journey.

The discovery that a traveler makes in the process of a dérive is, above all, self discovery, as in theory she will end in the place she is most interested in finding. The dérive can be made both through a space on foot, or through a dataset online. For example, a online dérive for me

---

[33] Ideas that also draw on Benjamin's Arcade Project and his description of the French tradition of the Flaneur.

through Wikipedia might start at at Wayne County, Tennessee, and travel through Roman Catholic Women priests, before landing inevitably in Anna Wintour.[34] At a glance, we could see wikipedia as its 5,608,851 English articles.[35] The path through the 20 or so articles between Wayne County and Anna Wintour does not give the user an overview of the size and scope of Wikipedia as simple set of numbers would. Nonetheless, it informs the reader in a memorable way of the significance of what is contained and the reasons for its being. The significance of Wikipedia in the abstract, and its subjective experience to a given reader are better expressed by a walk on the ground than the view from 30,000 feet.

Current efforts to make personalizations a reality include two major groups of examples. The first is the personalization of web articles for readers that are made by digital publications such as the New York Times, a feat that was not possible in print newspapers. The other is the suite of tools and visualizations associated with the "quantified self" movement, a collective pursuit of self measurement takes forms ranging from wearable devices to daily drawing exercises.

Personalized Interfaces for Readers

In recent graphical output of several major news institutions, personalization aspects of interactive graphics are growing more prevalent. These are interactive graphics that are designed to change depending on the input or characteristics of individual readers. These graphics use information solicited from the user or the user's computer as a lens through which users can understand either complex data or very general data. They also aim to increase the

[34] An exercise for those who would like to discover their true interests: Use the "random article" feature of Wikipedia as the starting point. Click on outlinks from the article that interests you and within 20 clicks you will land on something you are not only extremely interested in, but also deeply familiar with.

[35] Across all languages, Wikipedia consist of 27 billion words, 40 million articles, in 293 languages.

memorability of data through engaging personal beliefs of the reader. There are three types of interactions that reflect an aspect of the reader to increase engagement.

The first type of user input places the reader physically at the center of national stories by using IP address geolocation to filter the story's data and create content for the reader's location. It is the most automated, least visible, and perhaps most intrusive of the three. Geolocation is used when comprehensive datasets with national coverage may only be of interests in local slices. For example, in "The Best and Worst Places to Grow Up: How Your Area Compares", instead of providing a national ranking of best and worst places, the county you are located in is the starting point of the story. In the Outrider Foundation's interactive visualization effects of nuclear weapons, the user's location is used as the center of the blast, making the affected areas and population counts immediately relatable and more intuitively understandable.

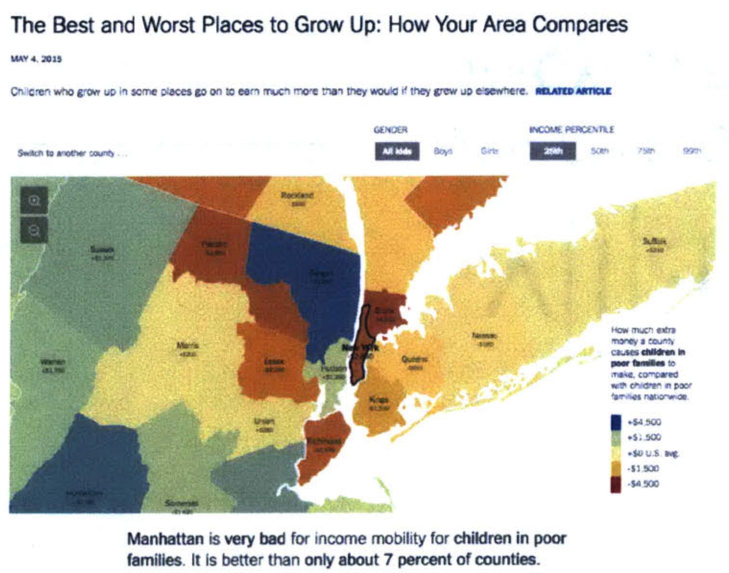


Figure 19: "The Best and Worst Places to Grow Up: How Your Area Compares" geolocates the user as a starting point for the story.

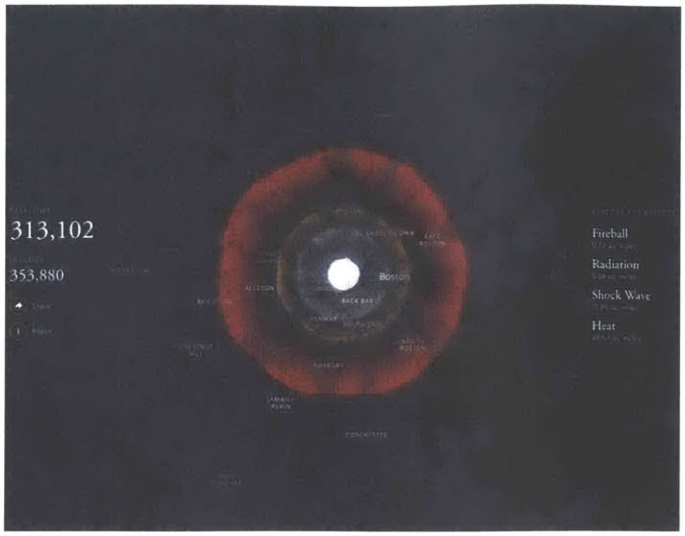


Figure 20: Outrider Foundation's interactive visualization effects of nuclear weapons.

A second type of interactive visualization solicits input from readers based on their knowledge of the story topic before presenting readers with the correct information. In the most popular case, the New York Times' "You Draw It: How family Income Predicts Children's College Chances" (figures below) prompted readers to draw a trend line of how income correlates with the probability of college attendance on the top of the story page, before being shown the actual data. What Times editors found was that this not only increased the amount of engagement with the story itself, the response to the content was also stronger. Follow-up articles reported that readers were able to remember how important a factor income is in the attainment of higher education because many were surprised by the difference between their guesses and the actual trendline. The format of having the user guess at a phenomenon before informing them of the facts has since been repeatedly used on other subjects, as shown below.

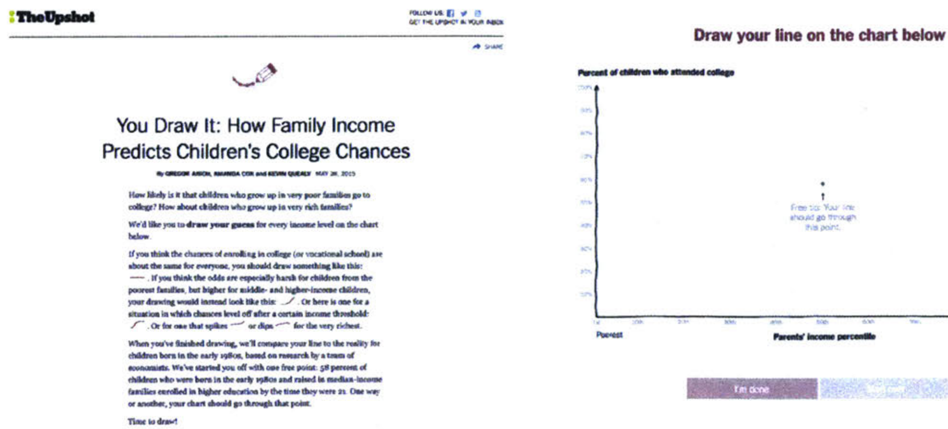
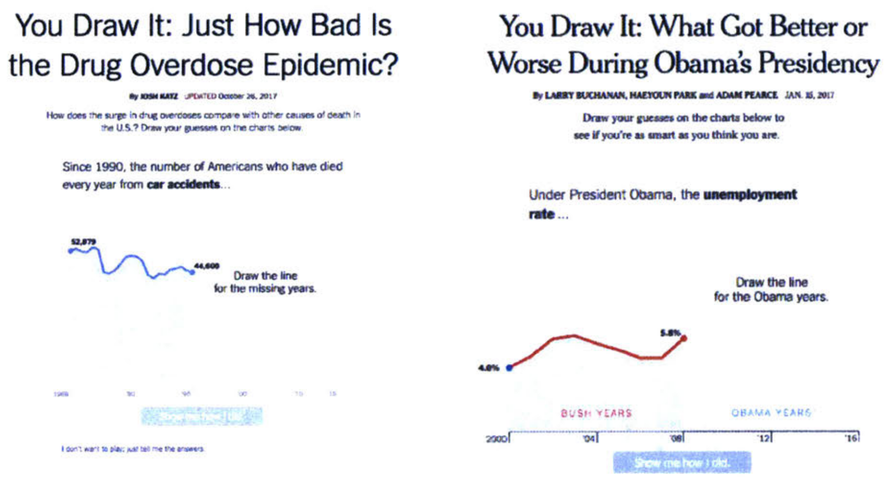


Figure 21 - 22: The original "You Draw It" Article.[36]



Figures 23 - 24: subsequent "You Draw It" Articles on different topics.[37][38]

The third type of interactive stories that incorporates personalized input include searchable or filterable interfaces such as quizzes and calculators. Instead of depicting larger trends, the reader is able to view data that concerns them in a specific way. These stories answer questions for readers directly, such as "Is It Better to Rent or Buy?" or "Here is where you should live to find your perfect match." The rental calculator uses real estate data to project outcomes based

[36] https://www.nytimes.com/interactive/2015/05/28/upshot /you-draw-it-how-family-income-affects-childrens-college-chances.html

[37]https://www.nytimes.com/interactive/2017/04/14/upshot/drug-overdose-epidemic-you-draw-it.html

[38]https://www.nytimes.com/interactive/2017/01/15/us/politics/you-draw-obama-legacy.html

on your real costs. The match finder uses Census data to give probabilities of finding a person fitting your criteria in each city.

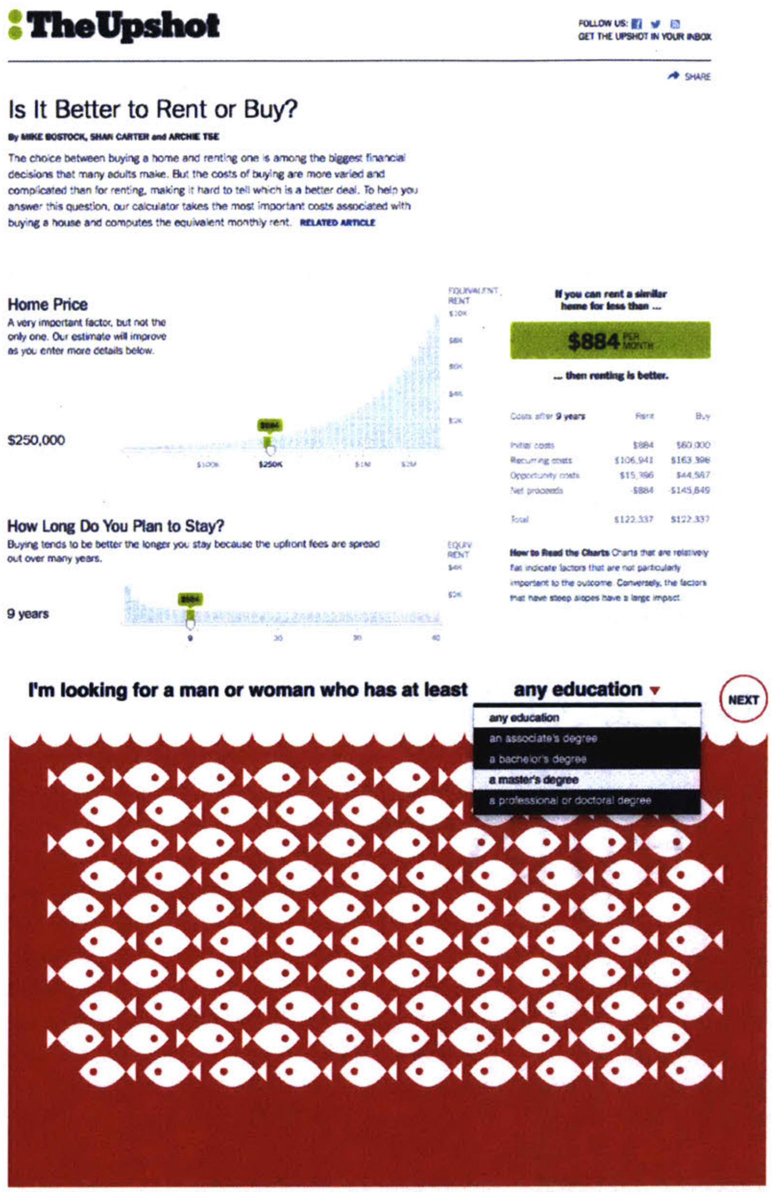


Figure 25-26: The "Rent or Buy" calculator and the "Where to Find Your Perfect Match" calculator.

These participatory interactive visualizations that expand a reader's view of a dataset from the personal to the national are a significant shift from 20 years ago. Additionally, these projects also center on the power of personal relatability as a key in the communication of data. They represent an alternative to the established notion of visualization in which simplification, the attribution of objectivity, and the power of the overview is prized. As design patterns, they challenge the visualization design mantra - "Overview first, zoom and filter, then details-on-demand." to include alternative paths through data and a more flexible progression of scale in views that visualizations present.

Quantifying the Self

Although this thesis addresses personalization in the context of aggregate data (personal filtering of external data) and the quantified self is traditionally focused on externalizing internal data, it is important to note the works and tools that come out of this centuries-old practice and its importance in building personal experiences out of data.

The Quantified Self movement promotes "self knowledge through numbers", and uses quantification of daily actions as a site for self assessment and reflection.[39] In the movement, the act of self quantification is a process of externalizing internal states through data. The use of self data collection to generate insight for self knowledge and assessment is not a new idea. The regular recording of a person's internal and external states have been around since the Middle Ages. However, the extent to which we are now able to record our actions with increasing ease and automation has allowed anyone to create extremely detailed portraits of themselves.[40]

There are growing cases in which we as individuals use data directly for self monitoring in our daily lives. Devices and applications that measure physical and financial health are common.[41] [42]Our actions can be tracked and reviewed in the steps we take, the books and music we

[39] Wolf and Kelly, What is the Quantified Self?

[40] Young, The Virtual Self

[41] Fitbit, a fitness tracking device, has sold over 20 million devices as of its IPO in 2015.

[42] In measuring financial health, mint.com is used to monitor spending and has over 20 million users.

consume, the food we eat, the quality and length of sleep we manage each night, the proportion of spending we devote to different aspects of our lives, and in the physical characteristics we exhibit such as heartbeat and skin conductance.

Geographers and data visualization designers have also explored the landscape of self quantification. Nicholas Feltron's annual reports can be read as a practice of comprehensive self knowledge. Bill Rankin's mappings of cities are a form of self portrait through portraying the environment. F the hand drawn visualizations series Dear Data by Stefanie Posavac an Giorgia Lupi is an example of using data visualization to build a personal relationship.

The Feltron Report

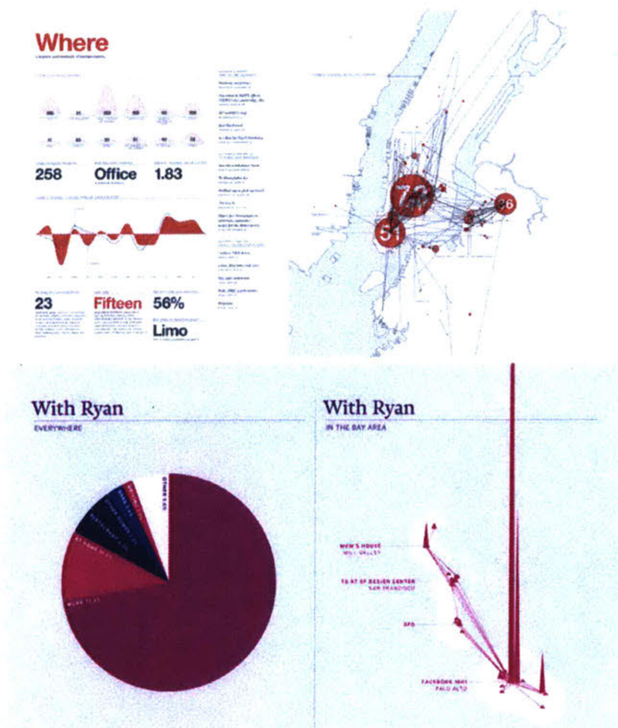


Figure 27 - 28: Pages from the Felton Report http://feltron.com/FAR11.html

The Feltron reports (2005 - 2014) where Nicholas Felton compiles personal annual reports on many aspects of his life are a notable example self quantification. In 2005 when the first report began, tracking one's habits was a technical and logistical challenge.[43] By 2014, when the last Feltron report was compiled, one's means of transportation, location, eating, sleeping, spending, and heart rates could be recorded through many consumer devices and applications, and made ready for analytic deep dives and soul searching by any self quantifier.

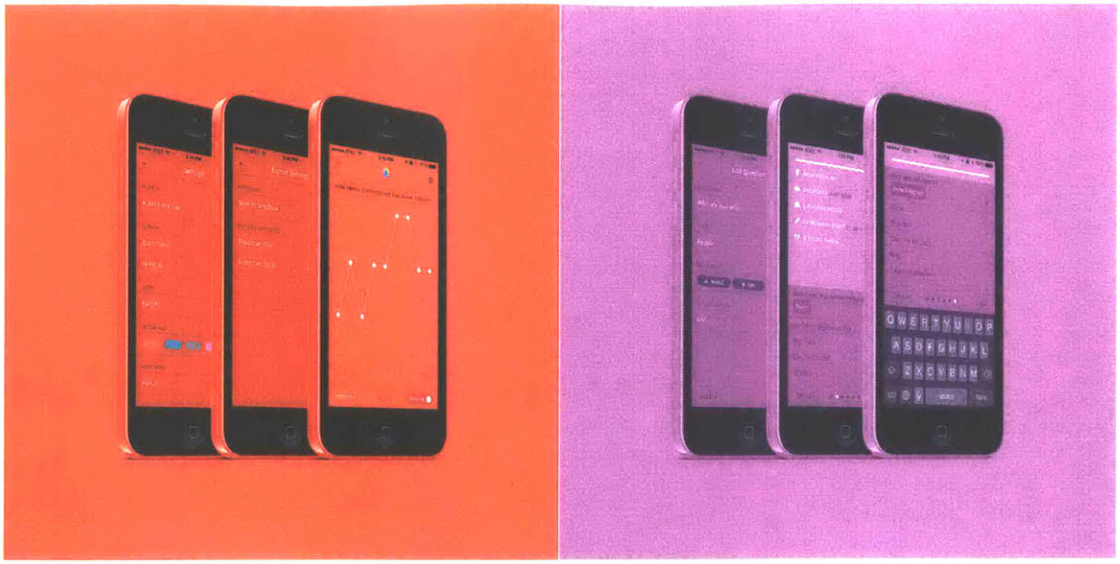
As daily users of digital devices, we are aware - and complacent to a degree - that some level of surveillance is implied by the use of digital tools. Not only does the digitization of our interactions with people and things imply recording, digital tools have also become more constant companions in our daily lives as they physically evolve into more convenient objects, making such surveillance more effective.[44]

As the Feltron reports become more dependent on automated apps for data collection, the designer himself produced an application for anyone to make Feltron-style reports for themselves. The Reporter App below is a minimal and flexible application that prompts at random intervals for a user to answer a set of questions. Over time, it creates simple charts to summarize the questions the user has answered. There are several predetermined/ recommended questions to begin with, such as if the person is working or how much coffee they have had. Ultimately it is up to the user to think about what they would like to learn about themselves. For me it was an opportunity to turn my gaze outward towards my environment.

---

[43] Felton spoke at Eyeo Festival in 2013

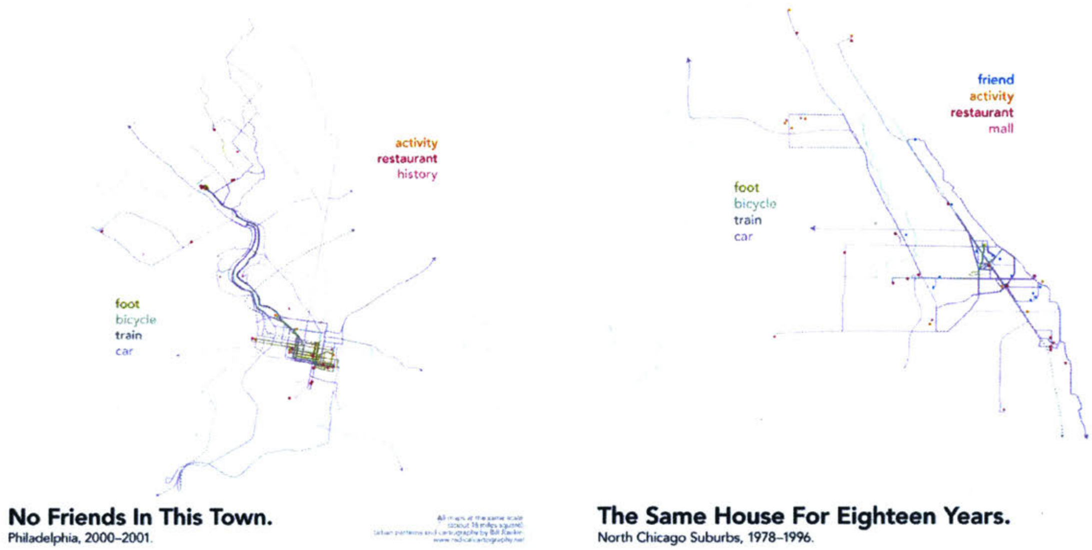
[44] Young, The Virtual Self, P. 26



Figures 29 - 30: From the Reporter Application designed by Felton.

Using the Reporter app proved to be a thought provoking experience. As I am immersed in Census data, I set up my Reporter app to self-report on the census categories of gender, race, and age to measure at any given time how many people are in my immediate presence that fit into each group. As I stood in line for coffee at 6am in my own neighborhood the morning after setting up my app, I received my first prompt and I saw that in the early morning that doorways and bus shelters were occupied by more than 30 sleeping bodies. The significance of this simple exercise to look at the external environment as a part of self tracking was immediately apparent. Over the course of the week, I continued to be surprised by my familiar surroundings. On the subway car, I am sharing space with between 5 and more than 100 people whose makeup shifts with every stop along the route. I was experiencing the city by quantifying proximate strangers, and this put my everyday movements in a new light.

Tracking the relationship between a person and their experience of a city is a slightly different kind of self quantification. Monitoring one's environment has the potential to place the person in a new context. While spaces and places are shared, experiences and perspectives do not have to be. In reality they are often unique due to both personal idiosycracies and the multiple functions shared spaces serve (see example Figure 8). Moreover, experiences often cannot be wholly shared due to limitations of time, knowledge, interests, and economics. Highlighting this difference in experiences between each resident of the city is important because it aids our understanding of our place in relation to the whole.



Figures 31 - 32: Rankin's maps of cities from his perspective.

Cartographer and historian William Rankin's maps of Chicago and Philadelphia were made to represent his unique personal experiences of the two cities. For Rankin, a city that he calls home is depicted differently from a city without a single familiar face. You may have experienced both of these cities. If told to draw a map of each city using the exact same parameters, you would still have made something that is possibly unrecognizable to Rankin. Even as our paths cross, our experiences are not necessarily in common.

A final example of self quantification demonstrates the diversity in data, representations, and methods of communication that is possible in quantifying one's everyday life. "Dear Data" is a set of correspondences using data visualizations between two new friends: Stefanie Posavec, and Giorgia Lupi. After meeting at a conference in 2015, the two data visualizers spent a year sending each other postcards of visualized data on a shared topic. From to do lists, to doorways, to swearing and drinking, the subjects covered in their 52 weekly exchanges are uniquely mundane. Overtime, the two get to know each other through what they measure and how they measure, demonstrating that you can get to know someone(or yourself) through their data in a personal and nuanced way.

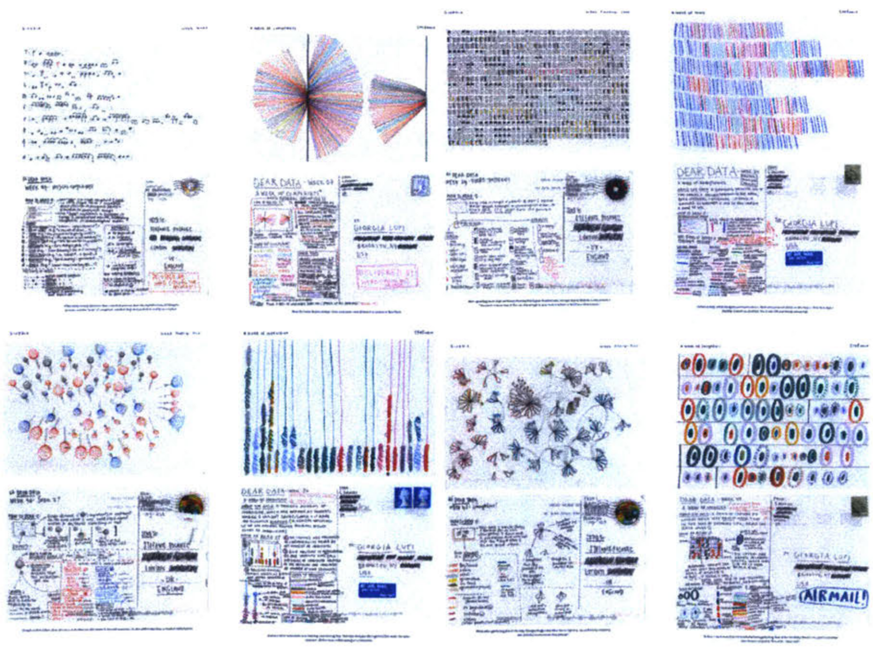


Figure 33: http://www.dear-data.com/theproject

The personalization of data can take many forms. It can be a search box, a IP address, a set of filters that locates the reader for the moment. It can also be a prolonged self examination using data recorded by hand, by phone, or in an automated way. All these forms are meaningful ways for us to engage with our environment through data .

Narratives

Narratives are another important aspect of humanistic data visualization. As data visualizations grow more technically, experientially, and structurally sophisticated, they begin to resemble well crafted creative non-fiction. Creative nonfiction is simply explained as "true stories, well told." [45]

The richness of interactive visuals would indicate a lot of potential for creative storytelling. However, as Heer and Segel's 2010 paper on the subject argued, visualizations of certain sophistication were still more often found in exploratory data analysis rather than storytelling. Their paper provided guidance on how narratives could be used in different visualizations. Within "Narrative Visualization: Telling Stories with Data", Segel and Heer discuss narrative structures in three ways, through genre, narrative tactics, and narrative structure.

Narrative tactics include design elements such as layouts, transitions, and interactivity that allow a reader to follow objects and themes within the story from one view of the data to the next. A subset of these tactics are widely applied in many interactive stories today.[46] Among them, the most commonly found are the use of object constancy, consistent visual frameworks, and transformation and transition guidances.

In order for representations of particular data points to be visualized in different contexts, **object constancy** is required when switching between views of data. This means that the representation of a piece of data or an object in the story must be easily followed from one view of the data to the next. The same consistency of representation also applies to **visual frameworks**, where layouts and control elements of pages are designed to orient the reader. In many cases, the visual framework is evident in the consistent location of filters or menus.

---

[45] Definition from creativenonfiction.org

[46] From the Segel and Heer paper

Transformations, in the form of judicious use of animation, can allow users to connect objects in different contexts. Other **transition guidance** methods also allow readers to navigate between views while focusing on the same object. Finally, **tacit tutorials** give a tour of the available interactions of a visualization by presenting an initial animation using these components before inviting readers to explore on their own. The standardization, proliferation, and improvement of these tactics make today's complex narratives possible.

In terms of narrative structure, Segel and Heer characterize visualizations as falling on a spectrum between author driven and reader driven. An author driven structure is linear and has limited interactivity, while a reader driven one has no prescribed order and is highly interactive. Almost all of the cases they studied were some combination of both. One popular structure that visualizers know as the "martini glass" is a linear narrative that opens into a interactive at the very end,  forming the shape of a martini glass on its side. Placing interactive visualizations within this spectrum is a helpful way to think about the purpose of the visualization. In many cases of interactive news graphics, the interactives are author driven, while in experiments dealing with simulations or maps, the interactives are more open to exploration and driven by the interests of readers.

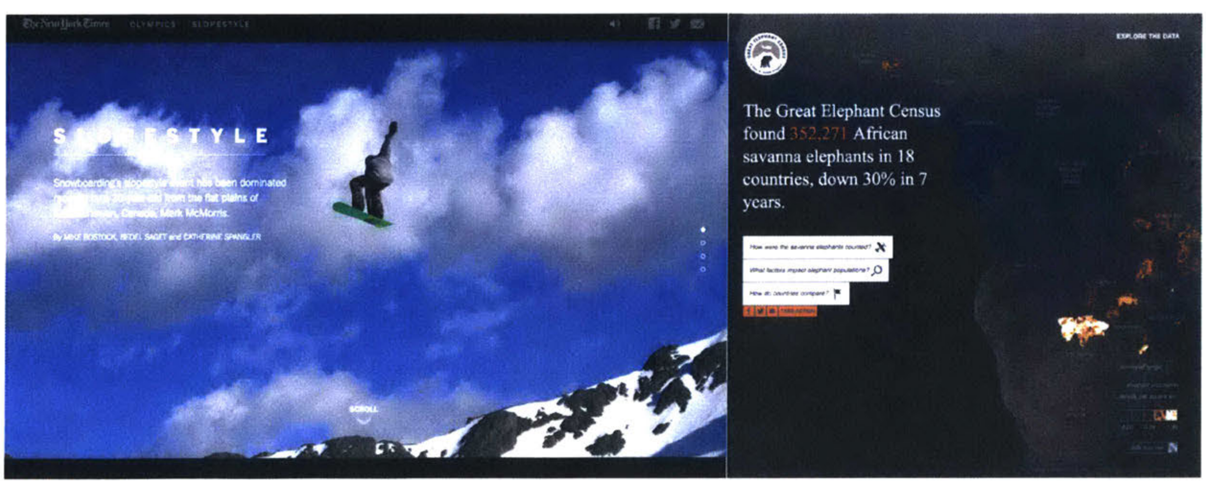
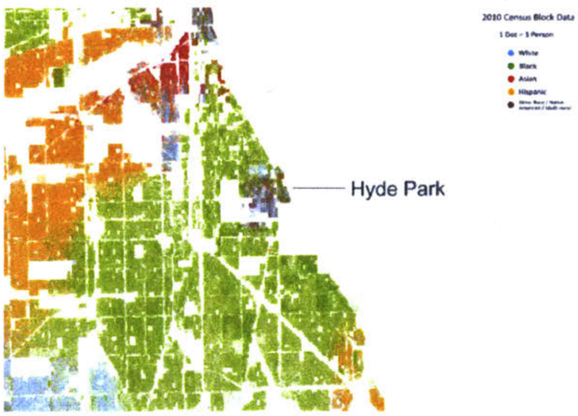
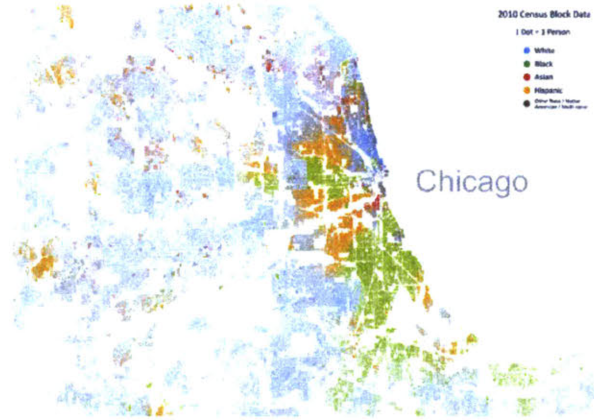


Figure 34 - 35: Author and reader driven interactive stories.

Above right is a reader driven example, the elephant atlas which shows the population of elephants in a open interactive structure. On the left is a author driven example from the New York Times Olympics coverage in which a timeline with markers orients the reader to move forward. In the author driven example, a guided linear narrative is indicated by the navigation bar to the right of the screen, which clearly marks different views and chapters in the story. In the case of the Elephant Census, open questions are posed, and no linear visual guidance in the form of menus or navigation is given.

Below is a hybrid user driven and author driven example. This FiveThirtyEight article on diversity and segregation in cities follows the traditional martini glass structure. The article explains the data before allowing users to find their own cities. On the left are the first visual elements in the article, and on the right is the open ended searchable chart of the data explained by the article.





Search

CITY	CITYWIDE DIVERSITY INDEX	NEIGHBORHOOD DIVERSITY INDEX	INTEGRATION/ SEGREGATION INDEX
Chicago, IL	70.3%	35.7%	-18.6%
Atlanta, GA	56.8%	30.7%	-14.5%
Milwaukee, WI	66.4%	38.6%	-13.1%
Philadelphia, PA	65.6%	38.6%	-12.6%
St. Louis, MO	56.7%	33.8%	-11.3%
Washington, DC	60.6%	36.7%	-11.1%
Baltimore, MD	50.4%	29.5%	-11.1%
Baton Rouge, LA	55.4%	33.2%	-11.0%
Cleveland, OH	58.7%	35.6%	-10.9%
New	[illegible]	[illegible]	[illegible]

Figure 36 - 38: Silver's article on diversity and segregation.

Thanks in part to Segel and Heer's paper providing the groundwork for narratives in visualizations, the lack of sophistication in interactive narratives they observed is no longer as true. In particular, data journalism has been largely successful in using the narrative power of the interactive visualization form to explore many subject areas.

In the 2017 OpenVis Conference keynote, New York Times "Upshot" editor Amanda Cox demonstrated the importance of writing narratives in data visualization in a series of works that allow for inference, prediction, juxtaposition, and the creative use of data. One of her examples demonstrating the power of relating to a reader's assumed prior knowledge was a graphic that communicated a new finding about the Tyrannosaurus' size by showing it standing next to a bus for scale. Cox described it as "made up", or as a type of embellishment. Everyone embellishes, but not all embellishments are created equal. Some, like the devices Cox utilizes (juxtaposing a T-Rex and a bus, the election needle), are a sign of artistry on the part of the storyteller and also important for pushing a larger narrative forward.

Narratives are particularly important in addressing complex subjects that include not only multiple datasets and views, but also multiple voices and storylines. The Marshall Project, a journalism platform that focuses on criminal justice is an apt example of a multidimensional long form creative nonfiction work. The journalistic goal of the Marshall Project is to inform the public of the state of criminal justice system through the methods of biography, event driven reporting, curation, and simulation. Criminal justice as a subject is not only complex and highly politicized, it also requires combining a substantial amount of background (sometimes archival) work with breaking news. The project combines first person narratives from the incarcerated, their families, and correction officers with discussions of policy changes, presentations of data, and contextual history.

There are several creative uses of data narratives in the Marshall project. A evocative image centered section called "Next to Die" turns a spreadsheet of names of the incarcerated into a live clock counting down to the next scheduled execution. This section highlights - down to the minute – the brief gap between the present and the next scheduled execution along with a descriptive account of the background of the next person to be executed. Interactive simulations such as an explainer on predictive sentencing allow readers to adjust factors that go into sentencing and watch the outcomes to understand the populations that are the most affected by changes in the policy. Other projects include both traditional reporting as events take place, and first person perspectives that document individuals whose lives have been affected by the criminal justice system.

These components work together to provide readers with comprehensive coverage of a single complex topic. The website works as a data collection project and at the same time present stories in highly stylized and readable formats with multiple scales and perspectives. With its combination of old and new, readers are invited back as events occur to use it as contextual reference, to seek out the opinion of domain experts, but also to view the site as a memorial as the next execution approaches.

From Narrative to Atlas

This hybrid of perspectives, methods, and representations will be increasingly influential in the future of interactive visualizations. The combination of many smaller narratives from different viewpoints and methods into a larger subject as demonstrated by the Marshall Project also provides a good model for interactive data atlases. A digital atlas should build a coherent logic through its intentional structure that evoke an immersive world around the facts that are rendered.[47]

An atlas assembles a collection of maps with shared geography, timeframes, or themes into one place.[48] Not merely a reference tool, an atlas requires juxtaposition of individual maps in a highly directed manner from which narratives often emerge.[49]

Analog atlases varied in size: a survey of early printed atlases showed volumes with sizes varying between 16 and 1371 pages.[50] In the digital realm, the size of something is more complex to measure because a simple count of pages will not suffice, nor will adding up the size of the data involved. The actual size of an interactive work is further complicated by the views and possible outcomes that differ between readers. Under these conditions, it is more useful to describe the scope of an atlas. The scope of early atlases were as much determine by their materiality as they are by editorial intent.[51] In the digital context, although the constraints of print can be translated to that of data, the scope is more in the hands of the cartographer/editor than ever before.

The production of digital atlases, although in a different medium, maintain some overlapping practices with traditional, analog atlases. Digital atlases still require a careful editing process. Atlas structure, coherence, and standardization of design are inherited from their predecessors as well. At the same time, some characteristics of the traditional atlas are being challenged by

---

[47] Akerman, Editing Early and Historical Atlases

[48] This definition of Atlas comes from the Oxford English Dictionary.

[49] Wood, Pleasure in the Idea/The Atlas as Narrative Form.

[50] Akerman, From Books with Maps to Books as Maps: The Editor in the Creation of the Atlas Idea

[51] Dean, Atlas Structures and Their Influence

the shift to digital. For example, pre-existing debates on the inevitability of an atlas to communicate narratives are complicated by the disruption of linearity in interactive projects.

In recent years, mapmaking practices have been democratized rapidly. Making more precise and functionally sophisticated maps online now requires less expertise and time.[52] At the same time, large collections of knowledge online have grown rapidly, as exemplified by projects such as Wikipedia or, in the field of cartography, Open Street Maps. Amateurs and professional contributors alike continuously input information into these projects that range from personal histories to global perspectives. In web cartography specifically, their contributions range in scale from first-person confirmation of land features via personal GPS devices, to the hand-labeling of satellite imagery, to the unknowing generation of personal location data that is automatically tracked.

The very growth of these large collections of knowledge online - both available data and the increasing ability to provide infrastructures to organize and coordinate that data - has resulted in the proliferation of open-ended and open-access platforms. These platforms can be government sponsored (data.gov), privately held (enigma.io), or produced as a part of the knowledge common. The production of open-access geospatial data can be seen in long term projects like Open Street Maps which requires careful orchestration to achieve the common goal of providing detailed and accurate public data for every place on the globe.[53] These aggregation platforms provide cartographers with a unique opportunity to reconsider the role and functions of the atlas today. In this context, data driven atlases - and other data-driven works that are equal parts

---

[52] Gerlach, Vernacular Mapping, and the Ethics of What Comes Next

[53] From notes taken while attending State of the Map, the Open Street Maps conference in 2016.

reference and narrative - are especially important because of their ability to assemble from immense collections of data, lexia that is compelling and digestible.[54]

With a traditional paper atlas, the role of the author or editor of an atlas is, by convention, separate from the role of the actual creator of the maps. Each role requires a different set of skills. In his 1995 essay on the how the role of the atlas editors evolved over time, historian James R. Akerman described the role of the atlas author in the following way - "An atlas is a map of maps, and its editor a meta-cartographer". The mapmaker applies an implicit theory within a map while the editor renders that theory explicit in prose by creating the overarching narrative with a collection of maps.

In the digital context, the distinction between the editor of an atlas and the creator of the maps is not as clear. In early 2017, I had the opportunity to work on the visualization of a digital atlas that was a collaborative project between MIT's Civic Data Design Lab and Philips Lighting. The Atlas of Light uses a combination of quantitative and qualitative data to discuss the impact of lighting infrastructure in five major American cities. In my own experience working on the Atlas of Light, I have found that the creation of a digital atlas can be iterative: technological opportunities as well as conceptual opportunities push the project forward, while giving the creator(s) room for trial and error. This process gives rise to new ways to approach atlas production. Team members work in multiple roles at the same time, which allows everyone to be the cartographer and the editor (the creator who works at the meta-level, thinking in terms of the "book" rather than the individual map) in turns.

On the surface, digital atlases seem to reflect less of an editorial presence and to need no editor at all. Map order is less obviously predetermined than that of a book. However, for digital

[54] Manovich, The Language of New Media

atlases, the role of the editor is in fact more important than ever. Narrative structure is embedded in website architecture, navigational elements, the narrative text surrounding and within visualizations and information that appears on rollovers. Visual hierarchies as well as visual tactics found in Segel and Heer inform the structure of digital atlases. With digital atlases, the editorial hand may be less obvious, but it is no less necessary.

Summary

This chapter outlined design principles originating from the youarehere project by incorporating the ideas of critical cartography and humanistic data visualizations. It shows that useful constraints come in the form of the differing scales, perspectives, and limitations embedded in each dataset, and solutions to working within these constraints in the context of visualization. It clarifies the central role of personalization and narrative structure in visualization through examples in cartography, news, and art. The next chapters will take the reader through four series of cartographic experiments which further address the themes introduced here. These are the relationship of entities and scale, using physical experience as orienting concepts for data, using imagery to understand different areas, and using geospatial self quantification to contextualize aggregated datasets.

Chapter 3

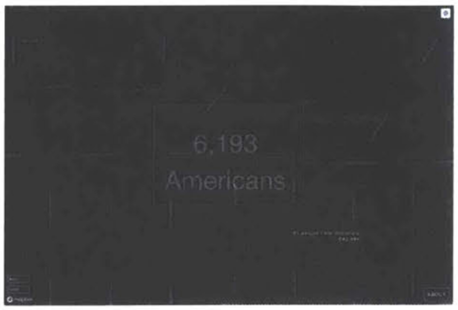
Powers Map

"Powers of Ten" (1977), a data-driven short film by Charles and Ray Eames, has a powerful narrative design that allows the film to memorably contextualize an atom in terms of the size of the universe. As the film zooms from a man's hand to the outer boundaries of the known universe, we see through the animated sequence the relationships that are embedded within each scale.

Figure 39 - 41: Scenes from the original "Powers of Ten" film, 1977.

I created the Power Map as a "Powers of Ten"-inspired visualization that applies the narrative construct of scale and powers to Census data in order to give human meaning to the relative scale of the Census' geographies and demographics. The map is a continuous animation that zooms in and out of different randomly selected locations in the United States. On its trips across scales, the map highlights specific data points based on what portion of the country is in view. For example, the map might highlight the number of high school dropouts in the portion of the map that is on screen at a specific zoom level.

In the first frame of the animation, the placement of human figures is random. The number of human figures that appears in this frame is based on (i) the calculation of land area displayed within your browser window and (ii) the population density of that area.



As the map zooms out, the central figure of total population grows, and different data points are displayed. At this level, 6,193 people live in the area within the area visible in the browser window, and "22 people have Doctorate Degrees"

"2879 people dropped out of high school"

"847,117 people are unemployed."

"12,780 people work in construction."

Figures 42 -46 : the map at different zoom levels showing the total population in view as the central statistic, and an additional statistic about the population in smaller text.

The map uses the geographical features that are in the land area displayed on the screen (block groups, tracts, and counties) to filter demographic information from the Census (occupation, age, education, income) and represent it on the screen. (Calculations of areas are approximate because they use geographical polygons that appear on screen. For example, if any portion of a Census tract area appears on the screen, then the entire tract is included in the calculations of data.) The current population of the United States is around 323 million; the population shown on this map does not include Alaska (0.74m), Hawaii (1.4m), or Puerto Rico (3.4m).

In the Powers Map, scale is visualized through movement, with the goal of portraying small areas in context with the whole country. Although the visual language of the browser window is omnipresent on the internet, it is has largely become invisible to the user because of its ubiquity. This map uses the dimensions of browser window itself to filter the mapview and display data that is visible only in that view. As the zoom level within the window changes, the data that is included in the visualization also changes.

To evaluate viewer impressions of the Powers Map, I made three separate presentations to small groups of five to seven people each by playing through the animations of zooming in and out before opening the discussion to the impressions of my viewers. The viewers' responses on the aesthetics and animation style of the map were positive. Viewers were interested in the way the map was able to connect smaller geographies to the overview of the country in a continuous way. After its initial showing at a meeting of the Civic Media Group at the Media Lab, several suggestions were made to change the display of Census data for clarity. These suggestions included the placement of text, their duration, and frequency.

In its second iteration made after the initial discussion and presented again at a meeting in the Civic Media Group, the text placement was redesigned to be placed in a circular radius rather than its original random configuration(figure below). The categories of data that the text came from were reduced so that data categories would appear with more repetition at different scales. For example, the number of “High School Dropouts” would be visible at different levels more frequently. The duration of screen time that each data point appeared on screen was increased from 1 seconds to 5 seconds, and each data point appeared alone only after the previous text has faded from view. Additionally, the initial view of stick figures were added to the loading screen to make the scale of the map more apparent on start.

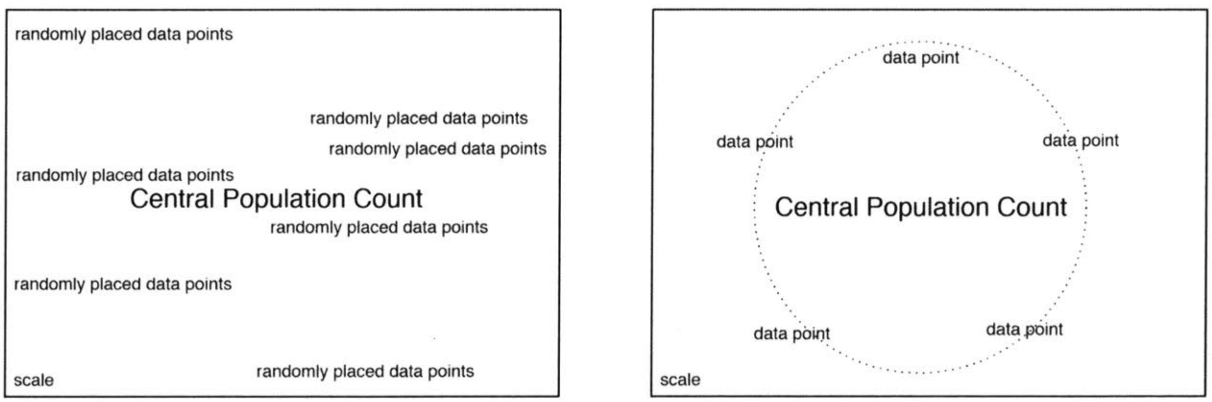


Figure 47 - 48: Left is the original configuration, and right is the revised version after feedback.

In all three presentations of the Powers Map, viewers expressed interest in the way the animation oriented them between small towns and the entire country through zooming. However, viewers also questioned the utility of the map because of the limited interactions they had. Viewers especially wanted to zoom out from their current locations or their homes. A button to geolocate the user was added for the second iteration. In addition to these design changes, I used the idea of concentric scales as filters to data in creating a second map, the Scope Map which allowed users to have more control over places they examined. The Scope Map imposes the organizational form of concentric circles on a location to filter data, making where the user focuses the conceptual center of the data and the map.

Scope Map

The Scope Map uses American Community Survey data to compare data associated with different-sized areas that are centered on the same given point. When a user places the circle on a coordinate point, they are given three data points for each dataset. The data points each describe the population within a one-mile, two-mile, and five-mile radius. From comparisons of different-sized areas, we can see, for example, that when a coordinate point is centered in Central Park in New York City, the inner radius has the highest income, highest education levels, and lowest percentage of houses with mortgages(below top). By contrast, in Elizabeth, New Jersey, income increases as one moves to the outer regions of the concentric circles (below bottom).

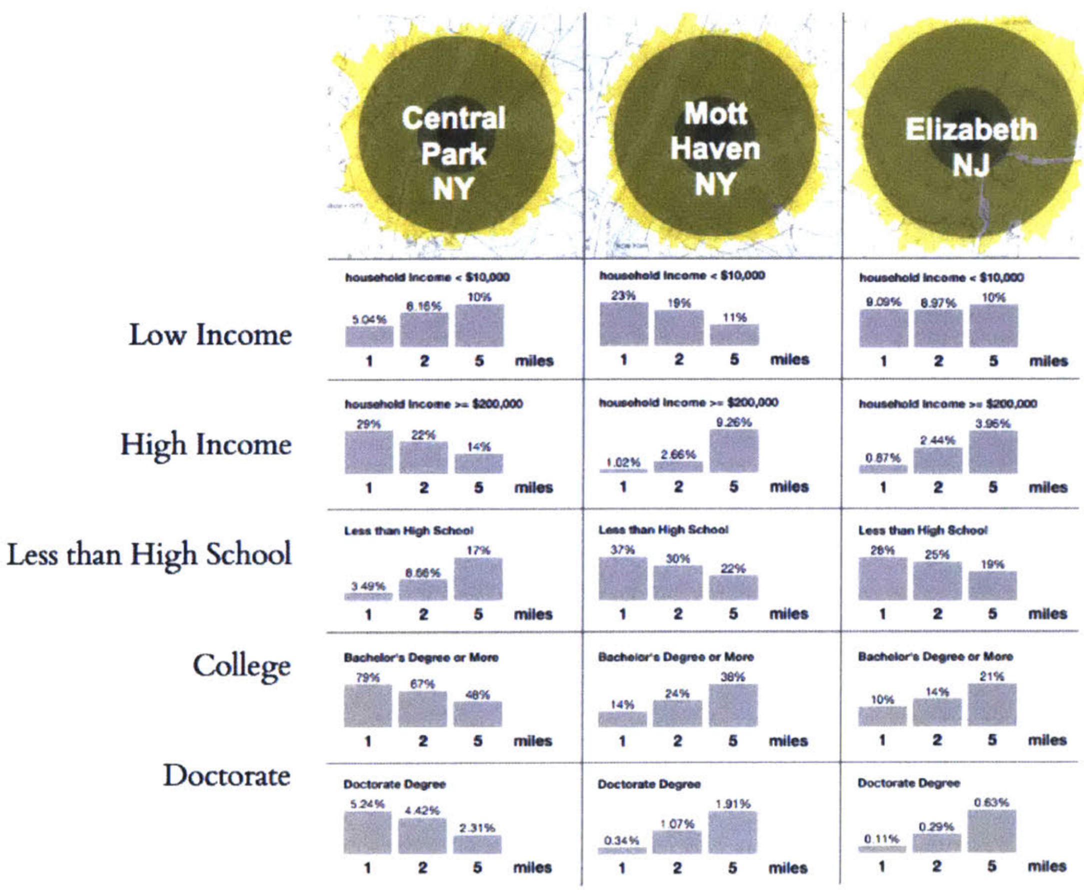


Figure 49: Data for three locations show how the concentric circles can create the increasing or decreasing profiles in data of increasingly larger areas.

The Scope Map similarly uses Mapbox GL to load and query the geographical features (in this case block groups) that are within each radius. It then uses a unique geographical identifier - the block group number of each feature - to compile the data from the American Community Survey. The data for the five-mile radius area includes both the two-mile and one-mile data; likewise the two-mile data includes the smaller one-mile area.

I tested the Scope Map by observing three separate users operating the map and discussing their reflections after each interaction. Each user started their interaction by placing the circles over a location with which they were familiar without any prompting. For two of the users, the starting location was their home, for a third, the neighborhood where friends lived which she had visited the previous weekend. For each user, they observed the income and educational attainment indicators as the most clearly understood, in that they understood the surroundings of their homes (and place of interest) as fitting into the economic profile created by the circles. When prompted, users explained that they expected particular profiles in income to occur based on their knowledge of the cities they looked at through the map, and the map was able to confirm their assumptions. However, they had not previously viewed those places as centers of a radiating pattern, and would be interested in examining places further in through this lens.

All three of my users requested that the radii be adjustable, and found the one, two, and five mile settings to be limiting for their interactions with the map. They stated that making the radii adjustable would allow users to perform more in-depth interactions to view particular areas, especially smaller areas. It may also be helpful to investigate other concentric shapes that are drawn depending on different units of measurement, such as time-distance(making 5, 10, 20 minute radii) rather than Cartesian distance.

It is challenging to visualize the dynamics of places by portraying it in the context of near and distant surroundings. Both of these maps attempt to do that through imposing scales with visual frames. Although both are small steps in attempting to use a location's surroundings to describe it, I believe there are more potential uses for the interfaces I have created in this chapter. Specifically, the more purposeful use of the boundaries of superimposed shapes and the edges of the browser window itself to interact with data over a map opens possibilities for zooming and scaling to be more than a navigational element. The maps frame data in novel ways by placing the user at the center of the data and the map.

Chapter 4

In this chapter, quotidian experiences that are a part of our daily routine are used as an orienting principle in the visualization of data. I visualize American Community Survey data via the two experiences of traveling through and looking in the city. The resulting maps - Cross Sections and Sightlines - show how we can visualize Census data from highly specific perspectives that track individuals' varied paths through the city.

The following examples, and the prototype from which they are drawn, explore the idea of visualizing physical routes that we travel along (such as a commute to work, or a trip to the coffeeshop) as cross sections of our environment. In exploring data from the census about the different people we cohabitate with in cities, this map filters data by creating cross sections of the city created by a person's movement.

Two maps of that use physical routes to organize data inspired the Cross Section Map. Both use Census demographic data. The first map uses the lens of the New York City Marathon Route to illustrate shifts in the population along the route and in the city at large. The second map uses the paths of subway lines in the city to describe the income dynamics of the city.

Figure 50: Official route map of New York City Marathon

The marathon in New York City was originally conceived in 1970 as a loop around Central Park. In its seventh year, as a celebration of the City's bicentennial, the marathon's organizers redesigned the course to cut through the city itself. The marathon has retained that route ever since. The course deliberately takes runners (and spectators) through all five boroughs of New York City, and along a varied path that reflects the physical and social makeup of the city. For this reason, it seems especially apt to use this route to sample the city's demographic. In 2011, the New York Times mapped the shifting characteristics of the neighborhoods along the route of the New York City Marathon. The animation they created (image below) uses a histogram visualization overlayed on the geographical route of the marathon in a 3D map(below).

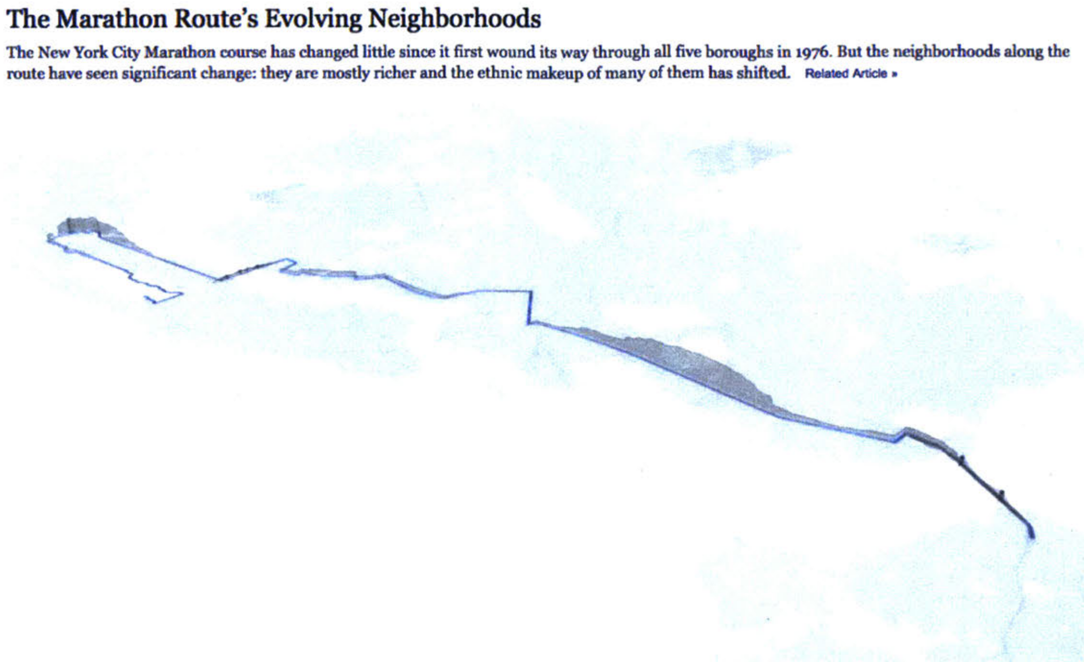


Figure 51: The New York Times Marathon Map By Graham Roberts, Alan Mclean, Archie Tse, Lisa Waananen, Timothy Wallace, Xaquin G.V., Joe Burgess And Joe Ward

A few years later, the New Yorker Magazine used physical routes inspired by the original Times' marathon map, to create an interactive map focusing on the city's subway system. For many people who live in New York City, the subway is an indispensable method of travel.[55] The New Yorker's interactive feature, "Inequality and New York's Subway" is a visualization of median household income along the city's subway infrastructure. After clicking on a specific subway line, the user sees a display showing how median household income is distributed along that line.

In the example to the below left, we can see the L train line cut across Manhattan at 14th Street and continues to Brooklyn. The median income displayed in the visualization peaks near the 6th Avenue stop and has a nadir at Canarsie to the right. On the right, the number 2 train, which cuts the city in a near vertical, encounters the highest incomes at the financial district in Lower Manhattan, and tapers in the Bronx and Brooklyn.

---

[55] According to mta.gov, there are an average of 5,655,755 riders per day on a weekday.

Figure 52 - 53: The New Yorker Subway Income Inequality Map by Larry Buchanan

Both of these examples use established physical routes to sample and then portray a slice of the city. They illuminate social and economic differences within the city by following the physical paths of familiar infrastructures, such as the marathon and the subway system, that famously cross different socioeconomic strata of the city. The scope that these visualizations use, reduces the data without diminishing it; they use a smaller subset of data tailor to the view of the commuter and the marathon participant that speaks to a first person experience. They also emphasize the juxtapositions inherent in a city, where people living radically different lives in close physical proximity encounter each other as part of a larger whole, whether through a daily subway commute or the annual celebration of a running race.

The Cross Section Map

The prototype that I present here similarly filters data according to physical experience. The Cross Section map allows users to create a physical route through the city or non-urban place of their choice and visualize the Census data associated with the places along this route. Users create routes by clicking on the map or by entering addresses in the search function. They are then given a route generated by the Mapbox Direction API. The side panel displays data along the route. It provides median data points for income, density, age, rent, and building age. The user can draw additional routes by clicking on new locations or by entering addresses in the search box.

The Census geographical unit used depends on the scale of the route drawn. For trips across a city, block group level data are used, while for cross country trips, data for counties are used.

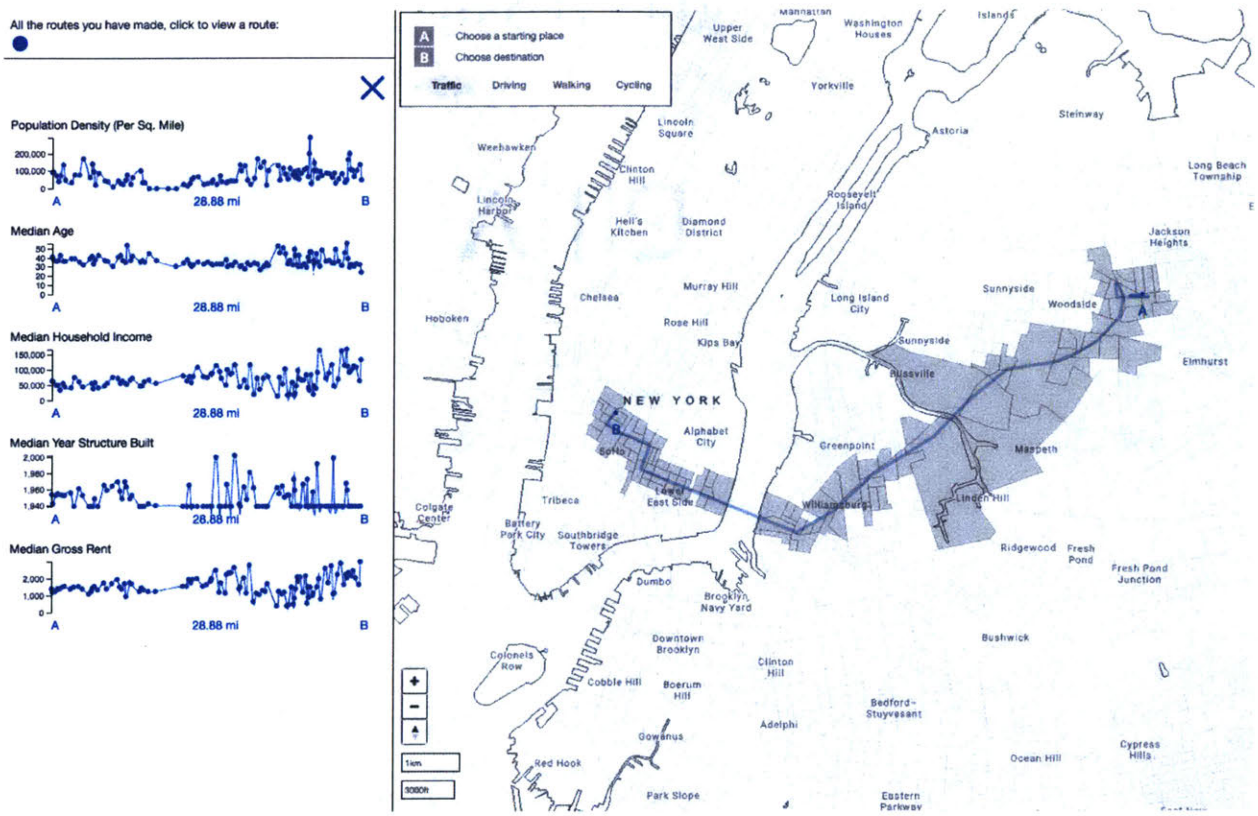


Figure 54: Screenshot of my commute visualized through the Cross Sections Map

The Cross Sections Map was tested in two ways. For the initial designs, I discussed the intentions of the map with Larry Buchanan, the creator of the original New Yorker Subway interactive on the possibility of turning his subway map into a tool for users to draw their own commutes. While he agreed that personalizing the routes may allow data to be sliced in more ways than there are subway lines, Buchanan also raised the point that people's familiarity with the subway infrastructure in New York City was important for grounding the nontraditional visualization.

After creating the initial working prototype for the map. I asked Buchanan to test its functionality and clarity by drawing his own commute. He encountered some technical issues with loading time and the layout of the graphs. Specifically, Buchanan suggested that the charts sit on separate panels rather than on top of the map in a transparent panel. He also suggested making significant improvements to the loading speed to make the map more functional. I took both of these considerations into account in creating a second iteration by repositioning the graph and map panels on the webpage. I then restructured the dataset from a single file containing all geographies into a file system where each geography(block group, tract, county) is only loaded when a user has requested it through his or her route selections. This greatly reduced both the initial loading time, and the speed at which the routes can be drawn.

The second iteration of the Cross Sections Map was then tested by asking six users to interact with the map by drawing a route of their choosing, and answering a short survey on their experience. When prompted for what each user learned through their interaction, users answered generally expressed surprise at a certain aspect of their commute. For example, one user learned that a cycling route could be designed to pass through more middle class neighborhoods that the user had not visited, and another realized that the places they passed all had the similar median age of 30.

In terms of functionality, each user had suggestions for adding additional interactivity to the map, such as more flexible route drawing via dragging, or the ability to compare cycling and public transportation routes. In terms of the data content, users also asked to see additional geospatial datasets that were not covered by the American Community Survey, such as food deserts and the per student spending of public schools. Despite these suggestions for improvement, users found the map helpful and five out of my six users responded that they would live to use a similar tool to plan future commutes.

Sightlines Map

> *Seeing Manhattan from the 110th floor of the World Trade Center. Beneath the haze stirred up by the winds, the urban island, a sea in the middle of the sea, lifts up the skyscrapers over Wall Street, sinks down at Greenwich, then rises again to the crests of Midtown, quietly passes over Central Park and finally undulates off into the distance beyond Harlem.*
>
> - De Certeau, "Walking in the City" *The Practice of Everyday Life*

The second map in this series juxtaposes viewshed analysis with Census data to create views of data from the perspective of looking outward from a location point in the city. A viewshed is a type of geospatial analysis and visualization describes the area that is visible from a specific point location.

Viewshed analysis is a widely available in GIS platforms such as Arc and Quantum, as well as Google Earth from digital elevation files. Uses for viewshed analysis include understanding the impact of real estate development, calculation of solar radiation and conducting visual analysis for landscape planning or archeological research. The Google Earth platform performs the calculations with their existing 3d model, but did not allow the export of KML data generated by its viewshed analysis tool.

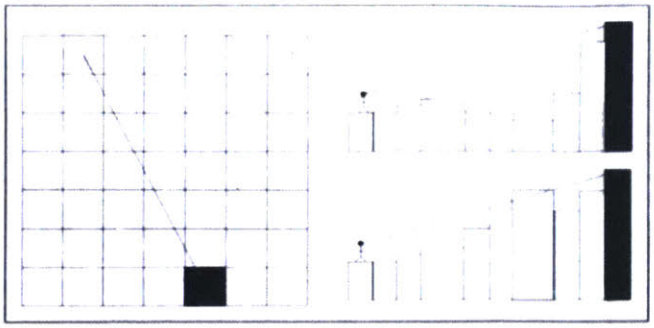


Figure 55: Diagram of direct sightline calculation. Fisher, Extending the Applicability of Viewsheds in Landscape Planning

The algorithm used to calculate line of sight is mathematically straightforward. As a result, in the case of my prototype, I have replicated the straight line analysis of viewshed calculations by sampling elevation points of building tops based on its footprint and applying simple geometric calculations in Python. I created the Python scripts from the Math and shape analysis library Shapely. I then overlaid Google Earth viewshed image outputs from a screenshot with my results to confirm them.

Figure 56 - 57: Public domain image of the Statue of Liberty with traditional view[56]. Right: The view of the statue from the 4th Avenue and 9th Street subway stop in Brooklyn.

The Sightlines Map is a viewshed map I created that is layered with demographic information. Views of the Statue of Liberty, as well as views from the Statue of Liberty, are iconic. They are perhaps the most readily available images from the world's repertoire of images representing

[56] Image from an article highlighting the Greenwood Cemetery area of Brooklyn, NY. www.green-wood.com

the city. Knowing this image is very different from knowing many aspects of the city that it represents. Nonetheless, describing the city from the perspective such a globally familiar place as the statue can yield fascinating results.

The Sightlines map uses the Statue of Liberty as the starting point for exploration; it asks: from what sites in the city can you see the statue, and what does the statue see of the city? Because in abstract situations (as in data visualizations) sightlines are reciprocal, the same map can be drawn of the view from the Statue of Liberty as of the places that can view the statue.

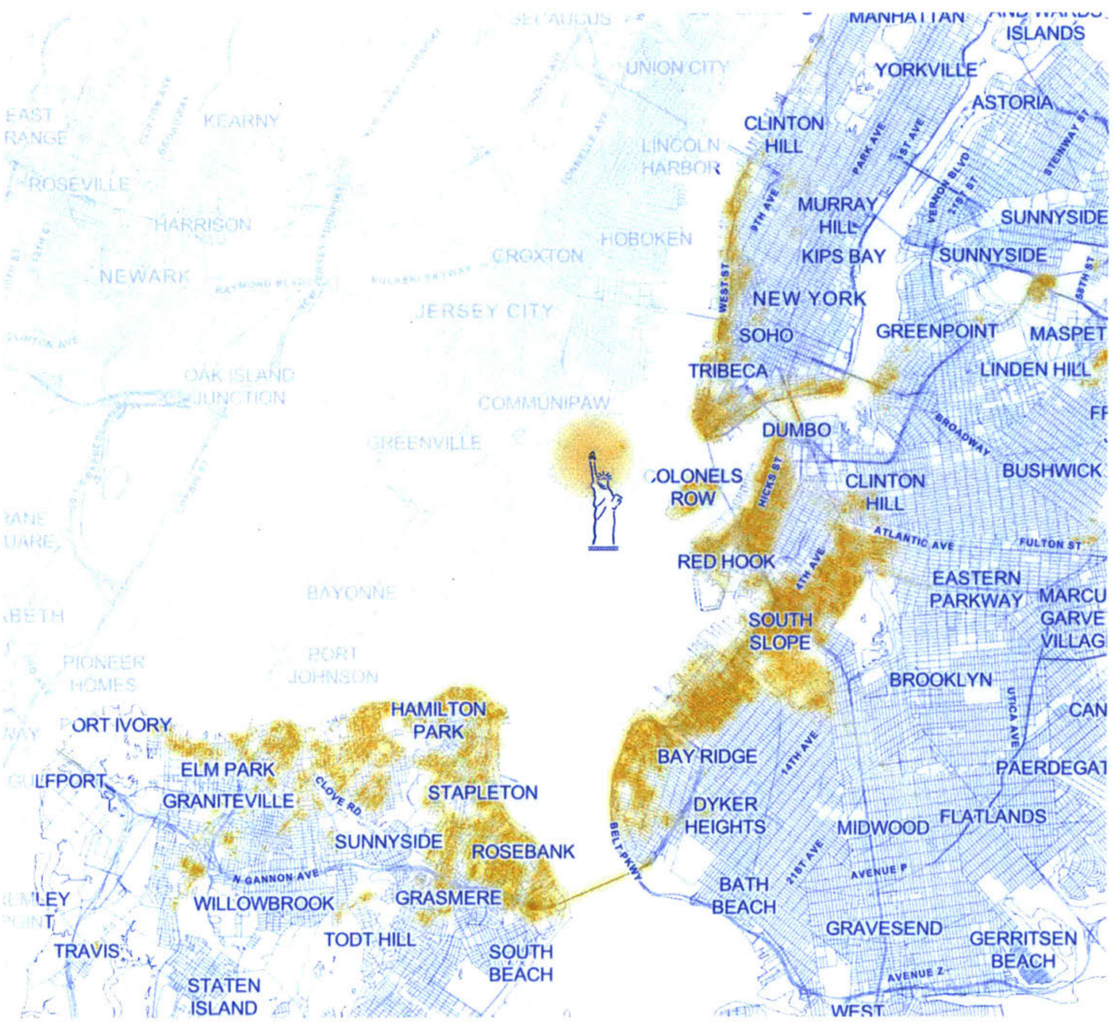


Figure 58: Screenshot of Sightlines Map.

In New York City, we are fortunate to have dataset of footprints of buildings and elevation data stored in the city's open data portal, updated yearly. Using building coordinates and the sum of elevations and building heights to determine a point in three dimensional space, we can quickly estimate if each rooftop has a direct sightline to the statue or is blocked by another building. Using the same elevation data, we can then apply the same method to the outdoor areas, including streets and parks.

After identifying elevation points with direct sightlines to the statue, census block group polygons are layered over the same map and areas containing visible places. This resulted in the comparison between visible places and the rest of New York City (shown below). The results show that visible areas have higher concentrations of populations that are young (ages 25 to 34), White (57% versus 43% citywide), higher income (15% of household have income higher than $200,000 versus 8% citywide), and working in specific professions (Business, Financial operations, as well as Professional fields).

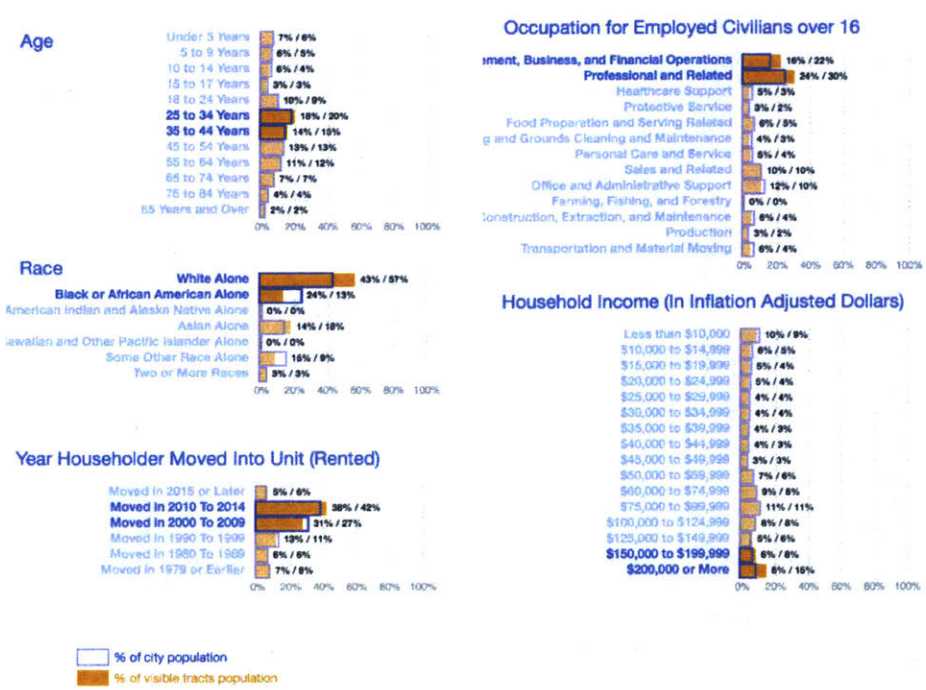


Figure 59: Charts comparing visible areas with rest of the city.

The printed version of this map was presented as a poster the 2017 North American Cartographic Information Society annual meeting, where it won the Student Research Prize. The discussion that arose in response to this project included as topics possible uses for viewshed maps for public events such as fireworks, parades, as well as views from other iconic places. These functional suggestions on new areas to analyze helps to show the value of understanding viewsheds within an everyday context where data can be applied.

However, the most interesting topics that came up through discussions were the hidden politics of sightlines in the city. My viewers suggested that while my selection of the Statue of Liberty was symbolic and expressed the idea behind analyzing viewsheds, more contentious locations could also benefit from such analysis. Additionally, some discussion participants suggested that it may be a powerful way to look at sites that the city intentionally isolates such as Riker's Island (New York City's main jail complex), the newly placed waste management facility in Brooklyn, or the housing complexes that are built for lower income populations in the City. These suggestions to further investigate what is seen and unseen indicated that viewers actively thought about issues of sightlines but have not yet applied data to visibility analysis.

The discussion then extended to the visualization of other senses such as "sound-sheds" and "smell-sheds". One viewer suggested that filtering the demographic data through soundscapes radiating from the elevated roads and trains that traversed the city could tell a very different story then what I have presented with the Statue of Liberty. Although the specific methods of producing such visualizations of more ephemeral senses were not discussed, my viewers agreed with me that the use of experiential measures such as sight, sound, and smell to filter otherwise static datasets has tremendous potential to highlight the often biased way the city is designed and used.

Through the two ways to filter Census data using physical aspects of experience presented in this chapter, I believe that employing methods of reduction and abstraction based on first person perspectives can make data visualizations more humanistic. I intent to explore these approaches further in work beyond this dissertation.

Chapter 5

As demonstrated in chapter 2, locations from whole cities to individual neighborhoods can be visualized using a multitude of private and public data sources. There are just as many ways to make comparisons between different places. Comparisons between cities appear in daily conversations, community blogs, the news, location intelligence services for businesses, and real estate search platforms aimed at individuals, families, and investors.

Comparisons are driven by multiple objectives. Some comparisons suggest similar or dissimilar cities that one might move to, or examine the strengths and weaknesses of particular communities. Some maps attempt to answer complex but practical questions such as "Where should I live?", "In which neighborhood would an innovation district thrive?" or "Where should Amazon open its next headquarters?" Individuals and companies alike have a wealth of data to aid our search for places that match our preferred conditions - these comparisons are very different if they're reacting to what we like about a city, rather than just what general data distinguishes one community from another.

Another common motivation for the comparison of disparate places is as an exercise in empathy and understanding and as an opportunity to gain perspective. With increasingly nuanced and accurate data portraits of places, we can, as users of public data start to imagine the experiences of other places, and perhaps even other people, through comparing places we know and places that are foreign to us.

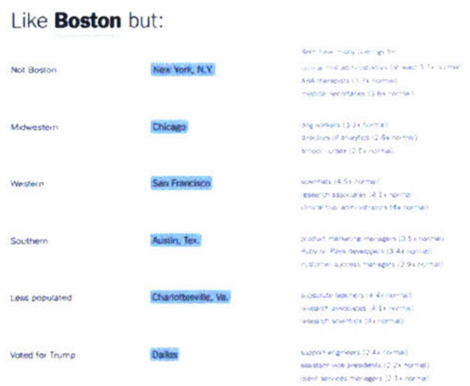
This section will briefly describe projects that match places by similar characteristics, including markets and demographics, imagery, physical characteristics, as well as hybrid methods of comparison. The section will then introduce a pair of visualizations produced for this dissertation based on the novel methods of using Census demographic data and imagery to make meaningful comparisons between places.

Ways of finding and comparing places

Markets and Demographics

What places are compatible depends highly on what we choose to measure. For example, we can compare places according to how far one's income will go, the quality of school districts, the diversity of nightlife, accessibility to cultural institutions, proximity to the ocean, walkability, safety, or even the politics of its residents.

In "What is Your City's Twin?", Kolko and Katz use the mixture of job postings found on a job search platform to establish which cities are compatible. Using Boston as the origin for example, the algorithm returns Chicago as a Midwestern version of Boston, Austin as a southern one, and Dallas as Boston for Trump voters. The implication of these comparisons is that if you have a particular mix of preferences such as political learning, climate, and profession, you could consider building your life in one of the places that are compatible with your skills, preferences, and beliefs.



Metros most similar to **Boston**

		SIMILARITY	DISTANCE (MI.)
1.	New York, N.Y.	82	187
2.	Philadelphia	82	274
3.	Chicago	81	859
4.	San Francisco	81	2,682
5.	San Diego	80	2,571
6.	Seattle	79	2,482
7.	Hartford	79	93
8.	Bridgeport, Conn.	79	144
9.	Austin, Tex.	78	1,693
10.	Los Angeles	78	2,588

Figure 60: Results for Boston's comparisons from "What is Your City's Twin?"

Other example matches preferences with places based on the people living in them. The Time Labs "Here's Where You Should Live to Find Your Perfect Match" calculator is a playful take on searching for places. The interactive matches the user's preferences for partners with the demographic of particular cities. The criteria used are age, education, income, marital status (divorced, widowed, or never married), and whether you prefer children. Using Census data, the calculator finds where populations with particular characteristics you select are more prominent, and is able to suggest the cities where you are the most likely to find your ideal partner. As rich as these datasets and methods are in covering different aspects of what it is like to live in a particular place, they leave out as much as they leave in.

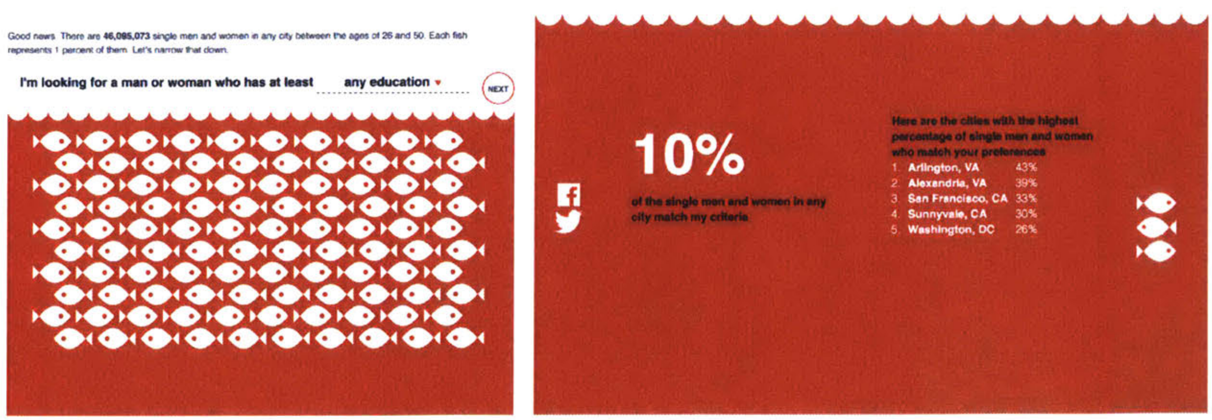


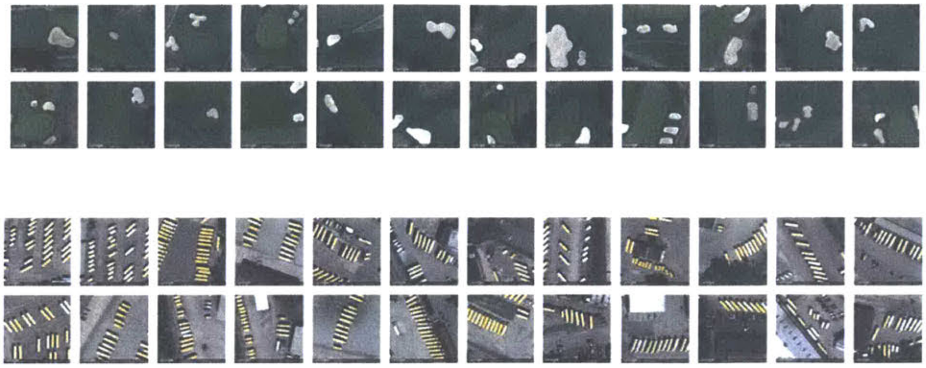
Figure 61 - 62: The Time Labs find your match calculator.

Images

Both Jacobs and Lynch spoke of the visual patterns a city creates and the importance of deriving meaning from them.[57] Both used ethnographic methods of interview and observation to understand the relationship between the form and function of places and how it can be made visible and communicated to urban planners and a place's own residents. When Jacobs and Lynch made their observations, they did not have automated image collection systems such as satellite photographhjy and Google Street Views in mind. In fact, the data they based their conceptual model of the city on was personal, anecdotal and drastically different in terms of dimension and detail. However, the idea of making sense of a place through the intrinsic connection between how a place looks and how it operated is still incredibly relevant today. Many projects utilize the image(s) of a city in the form of the troves of photographs from satellite, to Google Street View, to social media that we have at our disposal.

In the Terrapattern project from Carnegie Mellon, a neural network is trained on pre-labeled satellite images to create a visual search application. The user chooses a satellite image to originate the visual search ("query by example"), and is given similar images of other nearby areas. Some particularly striking examples can be found in golf courses (below top left), school bus depots (below bottom left), and shipping containers. The application then shows on a map how these places are distributed across the search area (image below center).

[57] In addition to Jacobs and Lynch, visual patterns as indicators of function can also be found in Venturi and Scott Brown's Learning from Las Vegas.



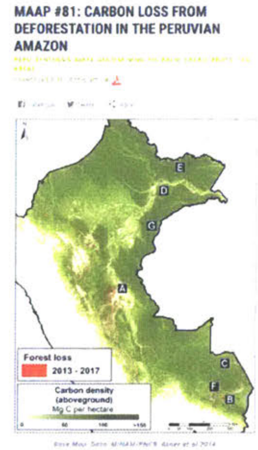


Figure 63: (left) Two examples of Terrapattern output.
Figure 64: (right) The mapping of the Amazon habitat loss using satellite imagery.

Terrapattern operates on the belief that deriving intelligence from satellite imagery will be increasingly important. It asserts that conceptual guidance for what kind of intelligence to seek out from this increasingly democratic dataset cannot be trusted to government or the multinational corporations that currently control the use of this technology. Several environmental monitoring efforts have used satellite imagery with great effectiveness, analysis of the Amazonian rainforest (image top right), documentation of elephant migration (example from chapter 2), and documentation of human rights abuses (cite https://www.eurekalert.org/pub_releases/2006-05/cp-sio053006.php) among them. Despite those efforts, the impetus for the project came when researchers found that the analysis of satellite imagery still fell in the hands of select corporations and government agencies with the exclusive goals of providing "actionable intelligence". For example, the Remote Sensing Metrics company uses satellite images to plan commercial real estate by monitoring density of cars in the parking lots of retail spaces.

Satellite imagery may be clinical, and the god's eye view reminds us that these technologies originate from surveillance, but as images they can also be deeply evocative. Two particularly relevant uses of satellite imagery (aside from the above Terrapattern example) use Google Earth satellite views to achieve different objectives. Jenny Odell's "Satellite Collections" makes collages from objects found in satellite views. It uses the repetitive patterns of human traces on land to convey a sense of connection between individual endeavors. When seen as a collection, we can marvel at the familial resemblance of the marks we make to alter our various environments.

> *The view from a satellite is not a human one, nor is it one we were ever really meant to see. But it is precisely from this inhuman point of view that we are able to read our own humanity, in all of its tiny, repetitive marks upon the face of the earth. "*
>
> *- Satellite Collections by Jenny Odell*

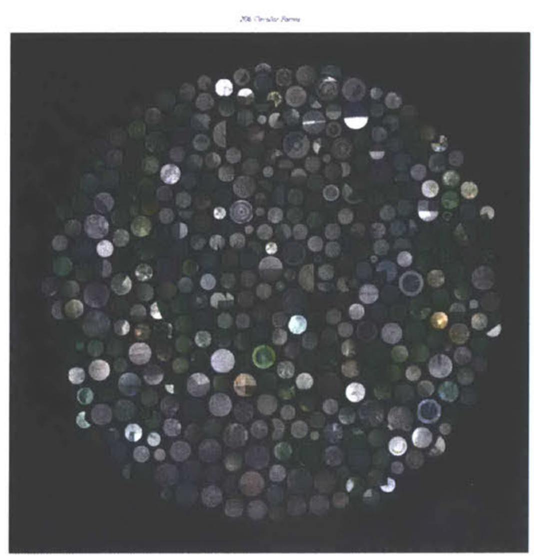


Figure 65: Image of swimming pools from Satellite Collections by Jenny Odell.
Figure 66: Image of circular farms from Satellite Collections by Jenny Odell.

Other times, we use satellite images to communicate a specific physical perspective, which in some cases can convey scale and magnitude more effectively than a list of numbers. In "Best of Luck with the Wall" (images from film below), Josh Begley stitches together 20,000 satellite images to constitute the entire U.S. Mexico border to show the vast and varying terrain that sits between 2 countries and the absurdity of the notion of walls in such a landscape.

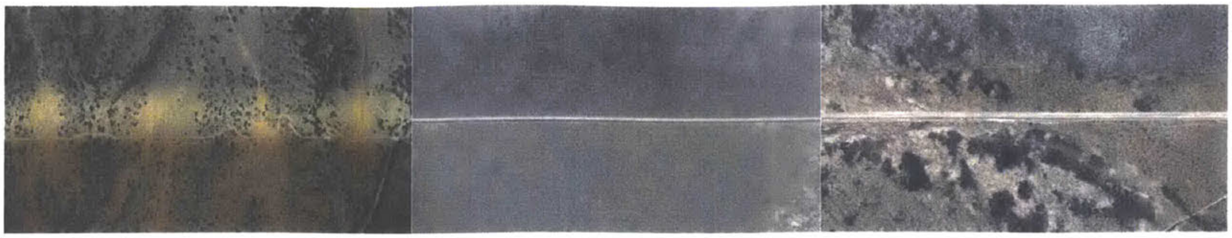


Figure 67 - 69: screenshots from the film "Best of Luck with the Wall"

Other collections of images that visualize place are found in Google Street View and social media. Image collections as datasets can be processed through massive human participation (such as Amazon Mechanical Turk), machine learning, or a combination of both to quantify their contents and characteristics.

Place Pulse by the Macro Connections group at the Media Lab is a platform that allows visitors to compare side by side Google Street View images of different locations in the world. In order to solicit qualitative assessments of images at scale, visitors to the Place Pulse site are prompted to choose between a pair of images based on a specific criteria (image below). Since 2013, the site has collected over 1.5 million clicks on images for 56 cities around the world. Over time, the project plans to use the participatory dataset to make image assessment algorithms for places, and scale the ability to algorithmically define areas of a city by terms such as "safe", "lively", "unique", or "family friendly". Both the potential uses and the potential biases that stems from this method of evaluating places are large. The labeling of places through a qualitative assessment of a image includes many unspoken value judgement from the participants of such an exercise, thus the quality of results are dependent on the participation of a diverse pool. However, this project opens the possibility of massively participatory qualitative assessment of images, making use of Google Street View which is otherwise reserved for navigational purposes.

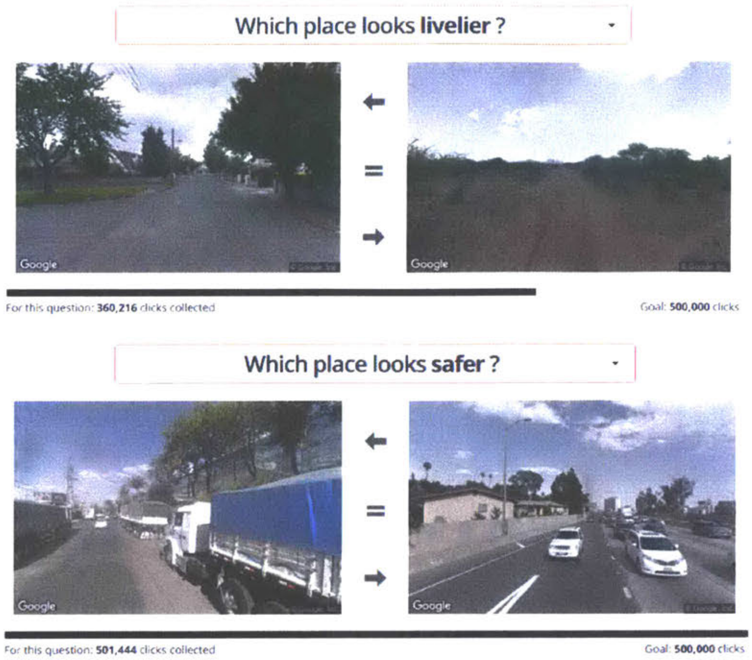


Figure 70 - 71: Two screens on the Place Pulse site.

Figure 72: Map from the Atlas of Light showing image tags.

The Atlas of Light utilizes an existing algorithm for tagging images. In the case of the Atlas, the imagery is "seen" through the Google Vision API, which applies key terms to each image processed to detect objects and faces. The terms can then be used to filter previously uncategorized images, allowing viewers to gain insight by viewing the geospatial pattern of particular types of images through the filtering of its subjects and objects. The inclusion of these terms allows users to filter the map according to image content, and also vice versa, making for a richer experience by connecting qualitative imagery with the other quantitative datasets in the atlas.

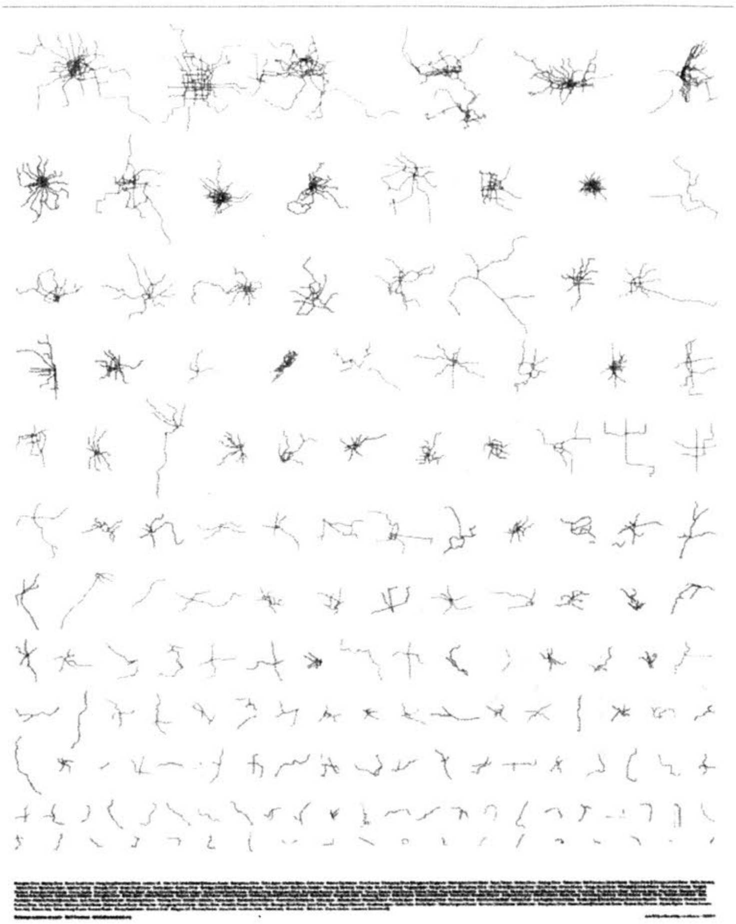
Both processes result in turning collections of images of places into descriptors that can be easily quantified, summarized, and visualized. Connecting geolocated data and behavior to how places look is not only important to those who study cities. It is also helpful to those who live in them.

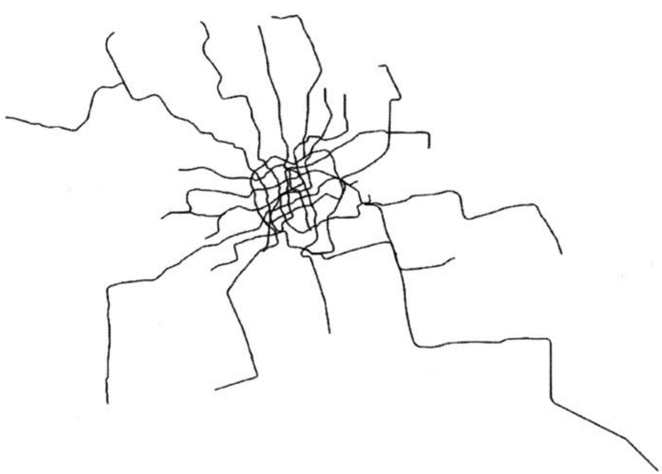
Physical Characteristics

Physical characteristics such as the average length of blocks, architectural style, and building age are datasets that do not include images but describes visual and experiential aspects of place. The block lengths, architectural style and age, as well as the lens through which its experienced (such as mode of transportation, length of stay) can all alter how a city is seen. Comparing the physical structures and physical characteristics of different cities such as the length and orientations of streets and transportation systems allow us to understand how each city differs in scale and in organization.

Neil Freeman's experimental cartographic images address the scale and orientation of our streets and transportation systems by placing them in context with each other. Below in the image of urban train systems (subway, light rail, and commuter rail) in order of size(below left), we can see the reach of each city as compared to others beginning with the largest, Shanghai(center), and ending with the smallest, Lausanne(right).

The significance of these comparisons only became obvious I was able to place myself in a city I know well, Beijing. I had been familiar with its subways system, and aware of its scale through first hand experience. When placed next to the Beijing system map, the transit systems from other cities become instantly more readable. Although my reading of these comparisons stemmed from first hand and personal experiences, one can also start to think about quantitative comparisons that are made visual in this way. For example, the fact that the Lausanne Metro spans roughly the same distance as one of the inner city lines at the center of the largest system Shanghai gives a quantitative comparison between the sizes of the two cities.





Figures 73 - 75: Transit Systems of cities by Neil Freeman. Closeups of the train systems in Shanghai and Lausanne.

Location Intelligence

From job availability, visual patterns, and demographics illustrated above, to climate, architecture, the length of streets, density of light, social media activity, and even the market for used cars, places can be compared and contrasted in a multitude of ways. Many different companies use their own algorithms to distill multiple datasets into a set of metrics to be uniformly applied to different places. These data models often use the term "location intelligence", a term popular with startups focused on suitability analysis in geography. They are a part of the multi-criteria decision analysis process, which uses multiple spatial datasets to inform decisions on where and what to place in a particular space.

A case in which multiple factors are presented for evaluation in location based decision making is real estate search where neighborhoods within the same city are compared. Real estate sites that offer services to analyze potential for investment and financing provides basic comparisons between places using a combination of housing rental, sales history and public dataset such as demographic data from the census or crime data from the local government. For example, the real estate search site Trulia offers rankings and ratings of schools, crime, commute, amenities, and general demographics alongside listings for housing. Within the demographics categories, the platform focuses on home ownership, tenure, education, marital status, and age. All this data emerges from a single source: the Census.

Figure 76: Trulia dashboard for prospective buyers to "Get a feel for the community" through demographic and other data.

The above set of datasets and methods provide an overview of how we currently find and compare places. They paint a rich picture in financial, demographic, physical, and visual terms.

Projects for direct consumption by the public such as news articles and real estate platforms often answer specific questions with data backed certainty. The questions they answer seldom address how individuals can get to know places through data without seeking to purchase goods and services. These comparisons are important ways for us to get to know where we live, and how it is similar and different from the rest of the country. The following maps attempt to serve that purpose.

The Satellite Comparison Maps

We can use any of the above described datasets and methods to answer the question of how places that are alike in data can be seen as a group. A choropleth map can provide a summary of places according to a specific criteria, for example population density, is visualized below on the left. However, the choropleth map only gives the geographical location and the sense of place that is specific to the criteria it visualizes. The shared goal of the maps produced in this dissertation is to use public data to alter the clarity and legibility of places for its citizens through experimenting with novel ways of filtering and representing data that is traditionally visualized through maps such as the choropleth. The maps below serve to shift the lens through which users view places: from the location of places fitting specific data points, to one in which they view visual characteristics of place alongside data points. These qualitative views that depend on both quantitative data and qualitative visual assessment may contribute to their understanding of both what specific demographics are describing and the reading of visual characteristics of place. Connecting image to data is a crucial part of that experience. Below right is a different representation of population density using satellite imagery to visualize census tracts with the highest and lowest densities.

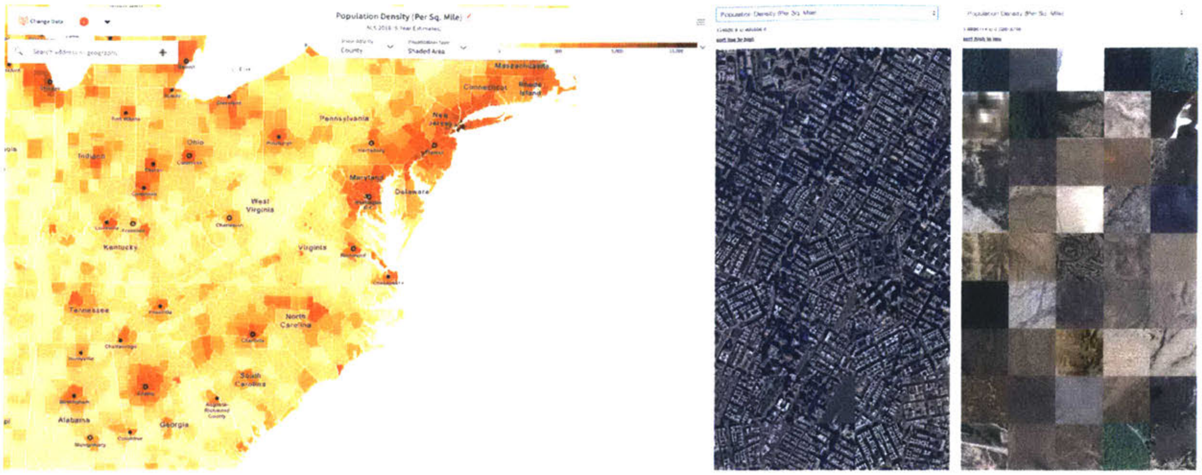


Figure 77 - 78: left shows a traditional map of population density, the right is my representation through satellite imagery.

In my use of satellite imagery, I take both functional and artistic examples as influences. I believe that imagery used in conjunction with public datasets can effectively communicate a sense of place and at the same time give a sense of the collective quality of a particular condition in a dataset.

The Census, an established form of public data that describes the conditions and people of the country in a consistent and specific way, is a lens through which users can understand their environment. In producing maps that uses satellite imagery as its main method of communication, I hope to demonstrate that not only can satellite imagery be expressive of certain demographic and quality of living characteristics, but that the images will add to our understanding of what the American Community Survey describes. Specifically, that the search for visual patterns through sorting and filtering of demographic data encourage users to not only think in terms of the data, but also in terms of the lives that this data describes.

The first Satellite Comparison map ranks and sorts images of Census tracts according to different criteria in the American Community Survey to convey a visual sense of place for where different demographic groups reside in the United States. There are two main components in this map. The left is a comparative view where two different criteria can be selected for the two panels to display image rankings side by side. The user can currently choose from 52 census tables, covering population density, age, race, education, industry, job sector, income, housing, health insurance, and transportation. Users are able to view a list of places sorted according to the selected table. On the right is space to view detailed information of tracts that the user has clicked on, including a larger interactive map, the location of the tract in the country, and the data for that tract.

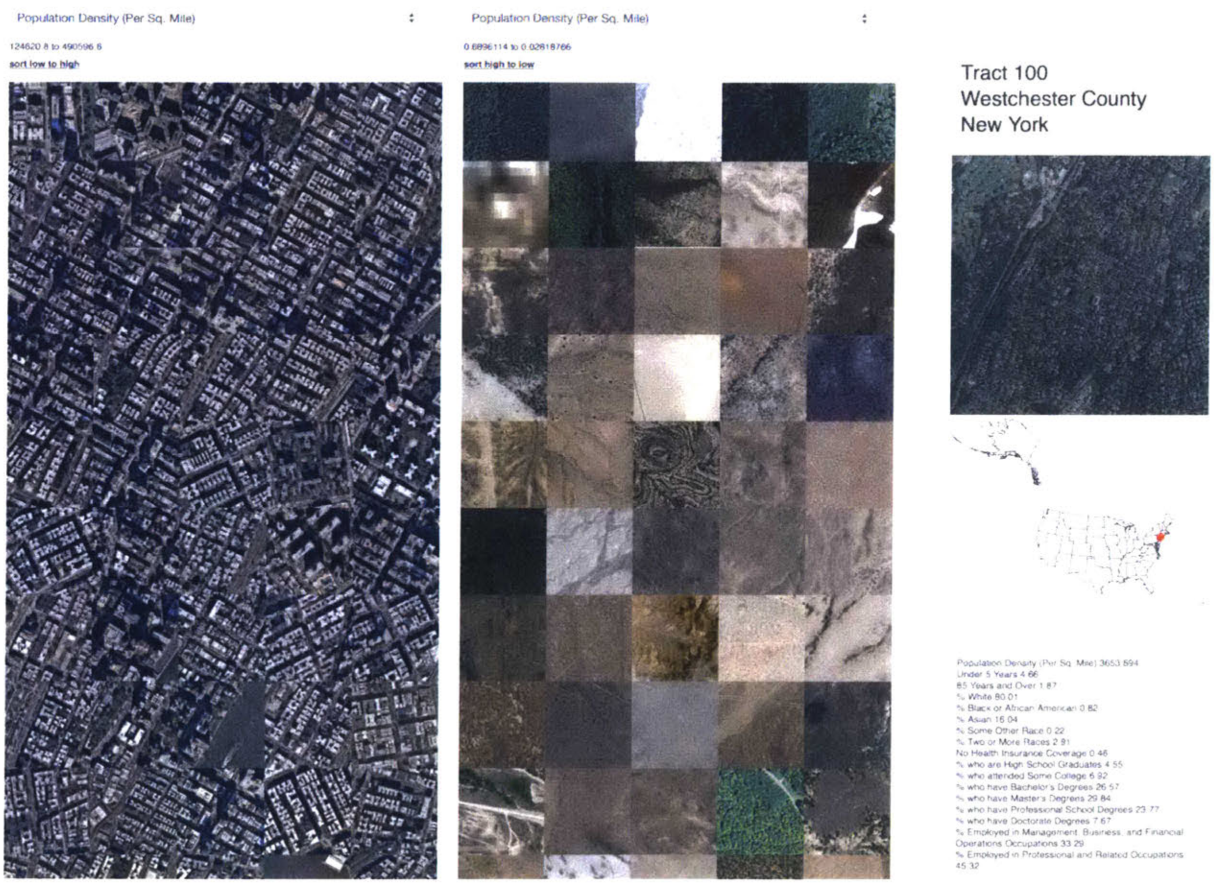


Figure 79: The third column on the right is shown when a user clicks on an image to provide additional information.

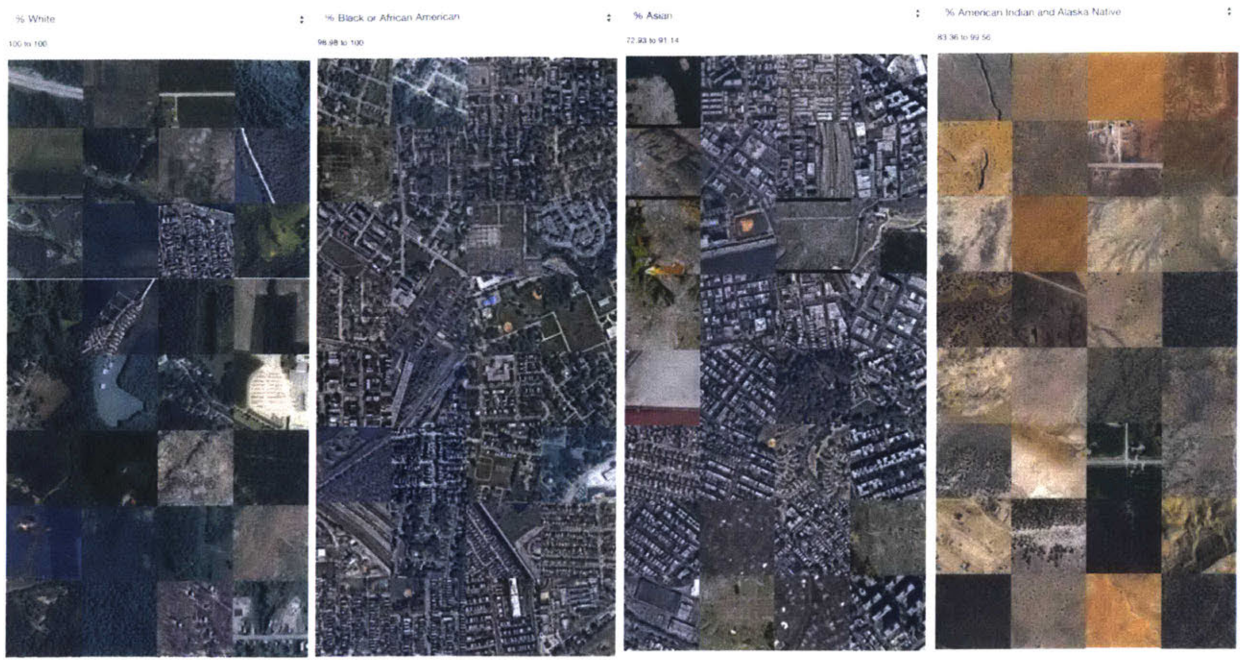
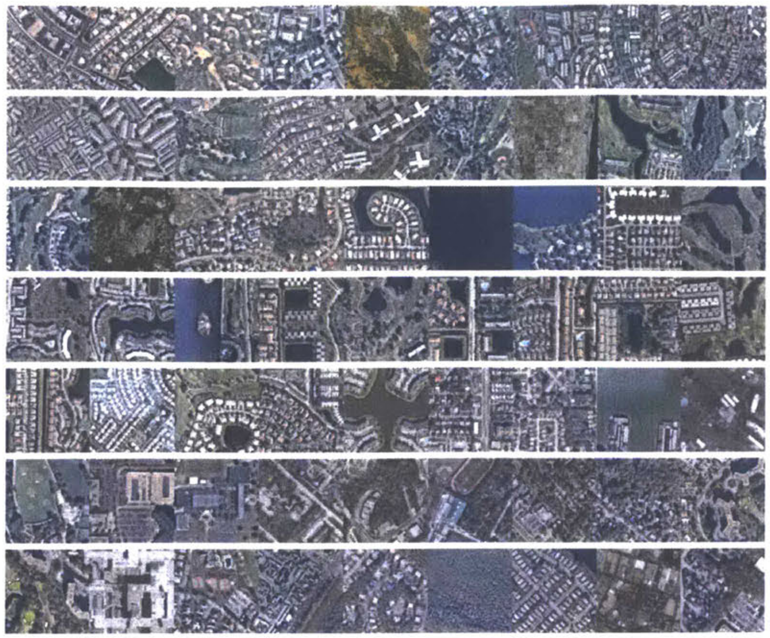


Figure 80: Image of tracts with the highest percent of each racial group identified by the Census. From left to right: White, Black or African American, Asian, and American Indian and Alaska Native.

A second map allows the user to use a combination of data filters to select census areas and view the resulting group of areas as a grid of satellite images. Similar to the previous map, the user can manipulate the dataset for visual results, this time by filtering ranges for multiple criteria in the same Census dataset. The user can then also click on any of the images together further details of the selected area.

Below, both groups of areas have a high percentages of residents over the age of 85. However the areas on the right also contain high percentages of residents with public assistance income. We can see the difference qualitatively of how those with higher incomes within a specific age group lives compared with those with lower incomes.

High % of Residents over 85



High % of Residents over 85 & High % of Public Assistance Income

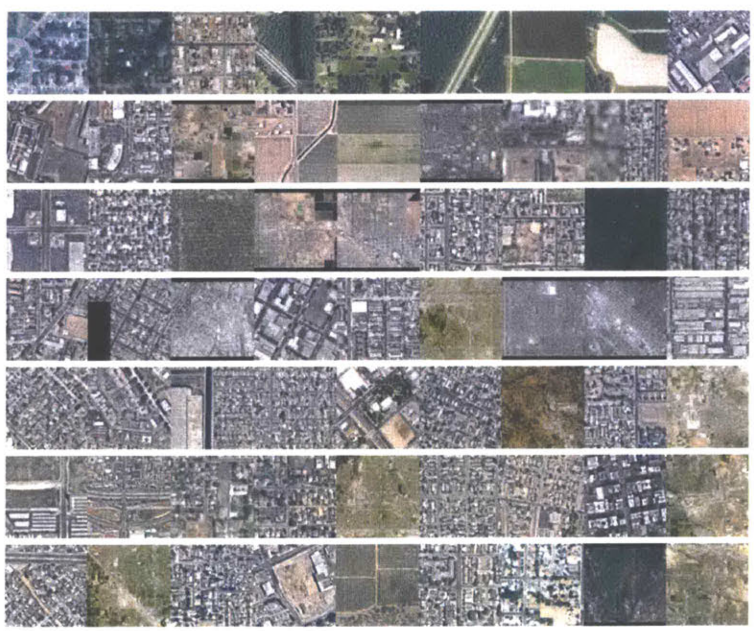


Figure 81: (top) Tracts with high percentage of residents over 85 years of age.
Figure 82: (bottom) Tracts with high percentage of residents over 85 years of age and high percentage of public assistance income.

The above two maps use the same underlying dataset of the American Community Survey, the US Census' annual survey that includes richer and finer grained details than the decennial census conducted each decade. . Satellite images are acquired through Mapbox using a custom client side downloader I made for this project. The downloader uses Leaflet library to load and the html2canvas javascript library to capture the satellite images. The visualizations are made with a combination of JQuery, D3, and Crossfilter Javascript libraries, while the data is processed through Python. For the full documentation, code, and data used, please visit this map grouping's repository at https://github.com/jjjiia/satellite_comparisons.

The Satellite Comparison Maps elicited the most immediate response out of the group. I tested these maps on several groups of people by presenting them with the tracts sorted sided by side for areas with high percentages of four different races. After initial feedback on design, I then surveyed a group of twelve users on their reactions to the maps. The users immediately noticed the "stark differences in racial and economic distributions", the relationship between density and race, as well as the vegetation and environment qualitatively. One user noted that the difference in the landscapes that white residents and Native American and Alaska Natives are concentrated describe a traumatic history of forced migration that is an essential component of the American experience.

The Implications of Finding that Exact Place

> *"Americans now had unprecedented choice about where and how they wanted to live. They had incredible physical and economic mobility - but these freedoms seemed only to have increased segregation, not lessened it. Why?"* - Bill Bishop in *The Big Sort*

Where we choose to live or to visit shape the nation's public life. Rather than out of necessity in response to economic hardship and opportunity, populations are increasingly shifting according to their values, tastes, and beliefs. Much like homophily and the filter bubble in the digital realm, like-minded people are searching out the same physical places to build their lives and careers. These highly personal and individual preferences and biases shape communities and the physical public sphere at large.

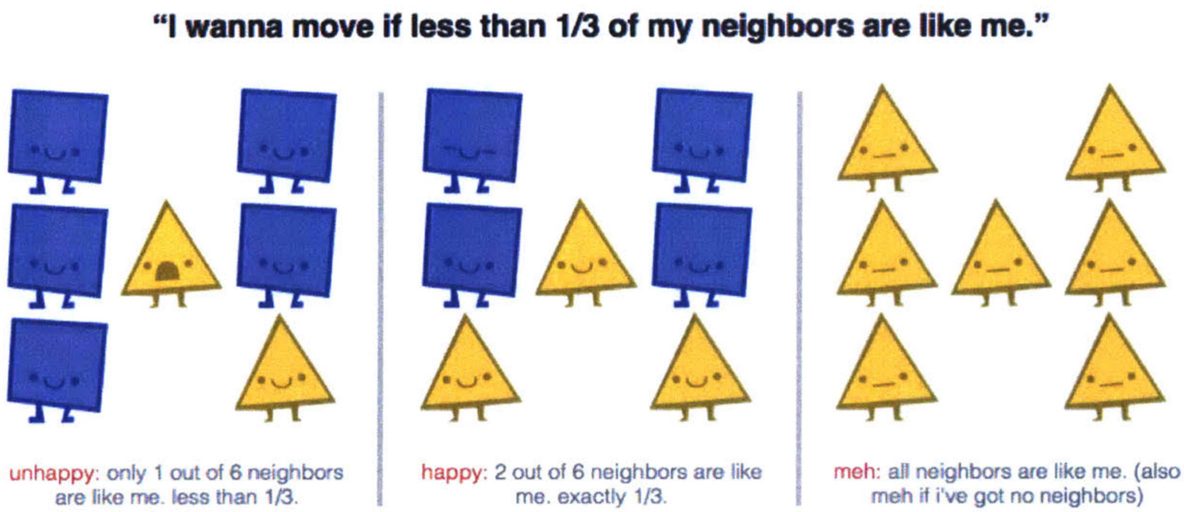


Figure 83: Introduction screen of Parable of Polygons.

"The Parable of Polygons" is a powerful abstract interactive visualization created by artist Nicky Case, representing the phenomenon of progressive segregation based on individual decision making. In it, "shapist" triangle and squares play out their own simulation of a segregated society. Each shape is "happy" when surrounded by the same shape. The reader is asked to adjust the placement of shapes to maximize their happiness, and quickly finds that they are unwittingly separating the shape and creating segregated areas(before, left and after, right).

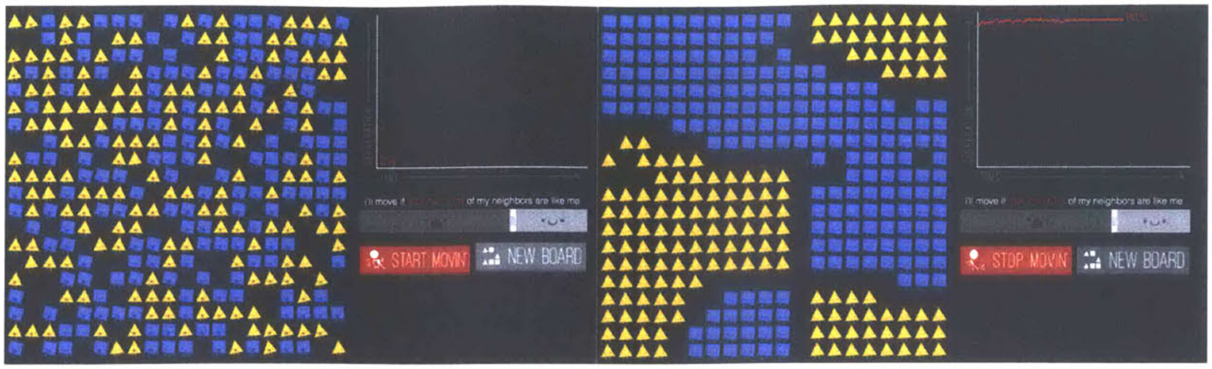


Figure 84 - 85: The before(left) and after(right) images of self sorting.

> *"When in Rome, we begin to do as the Romans do - often without even being conscious of it."*
>
> *- Alex Pentland, "To rescue democracy, go outside"*

The communities we choose to live in are now better predictors of our political attitudes than our friends or families. Seeking out our similarities rather than our differences when choosing our own environment further deepens mutually held beliefs. Alex Pentland, who has long been studying behaviors and attitudes in the context of physical proximity, suggests that our realm of ideas are shaped by the people we physically encounter, and the wealth of ideas is dependent on more exposure to faces that are unlike our own. In order to broaden your own mind set, Pentland suggests as Debord might have, going outside for a walk.

In "Talking to Strangers", Danielle S. Allen goes further, stating that to truly counter cultural self segregation, we must not only talk to strangers, but trust them as well. The most diverse cities in America are also the most segregated. Allen, who taught at University of Chicago at the time, was at the center of one of the most segregated places in America. She was well aware of the enormous differences between her Hyde Park neighborhood and those just a short walk away in Washington Park and Englewood. Using herself as a starting point, she devises a plan to "form political friendships" with the 99,999 people that live in her immediate vicinity. Talking to strangers and being with strangers that share differences is a start and one of the most self empowering actions we can take as citizens.

In this dissertation, I have been experimenting with maps that allow citizens to place themselves in the context of their surroundings through the use of census data. The last series of experiments in the following chapter uses self quantification through public data to increase both the legibility of the data and the environment it describes.

Chapter 6

Self-quantification

As noted in chapter 2, humans have practiced self-quantification since at least the Middle Ages, although the term itself is quite new. Present-day practitioners of the art use digital devices to document and track various aspects of their own daily lives.

In some cases, practitioners develop self-quantification projects by hand. Famously, Nicholas Felton meticulously documented his life between 2005 and 2014 in the annual Feltron Report which he released in hard copy and digital form, which data scientists celebrated as witty and beautiful exemplars of their art.

Quantification also makes use of automated data collection. Fitness applications are increasingly ubiquitous, which means that any user of a smartphone or smart watch can track their own heart rate, eating habits, and activity. Users can also monitor their habits of media consumption by tracking their own computer usage and reading habits.

What these projects of self-documentation do not cover is how we can situate ourselves and how we are situated in our communities. Where we are, and what other people occupy the spaces we occupy, are part of our expression of ourselves. For example, each resident of a city has a different geospatial pattern which contributes to a different perspective of the city. To understand the places we occupy, we must look outside of ourselves from time to time; because of this, the spaces we occupy and the people we share them with can be an important part of self-quantification.

I argue that Census data can provide a unique opportunity to track the strangers in our lives. In the following series of mapping experiments, I use my own travel patterns as a basis for visualizing Census data in order to better understand the places I occupy. The experiments start with small sporadic interactions in which I and others "check-in" to see how the Census describes our current locations. I then developed these short interactions into longer periods of monitoring. Finally, I apply both the methods that I have summarized from this series and methods I have applied in previous chapters to a final project; the visualization of a geospatial diet for a single test user.

In my initial test of visualizing one's personal geospatial context with Census data, I created three working web applications to show the user Census data for their current location. Each application generates Census geography from the user's current location using the FCC's Census Block Conversions API to turn latitude and longitude into a block group geographical identifier.

Figure 86: The first application displays data in a dynamically generated paragraph.

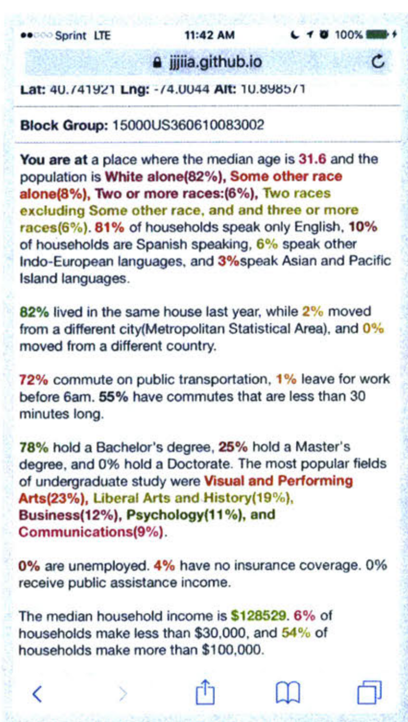


Figure 87: The second application displays the adjacent Block Groups in a graph.

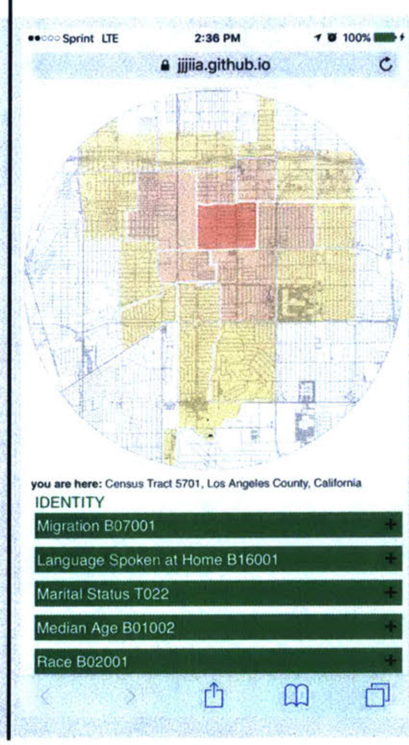
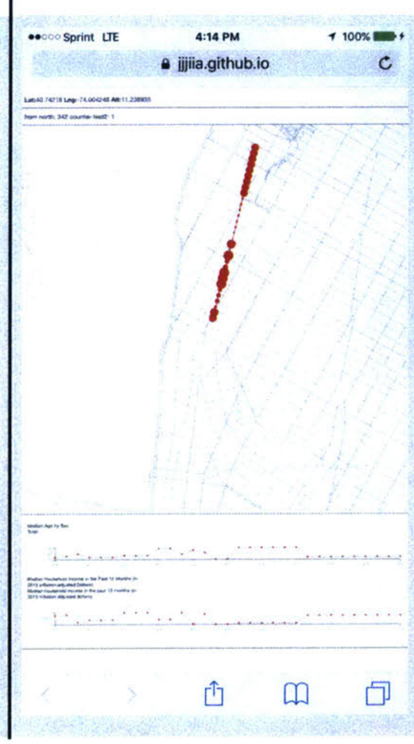


Figure 88: The third application allows the user to point in a direction and marks a compass over the map with data points.



These initial prototypes were tested with several users on their smartphones, I spoke briefly with each user to gauge their reactions on the data presented. I was able to confirm, by asking if the visualizations were understandable by the user, that both the data and the API were functional and that data can be communicated through the interface of bar graphs, maps, and dynamically generated text.

Living with Data

Data, a little bit at a time, over a long period

In 2007, the Twitter bot @everyword started to tweet every single word in the English language. After 109,229 tweets and seven years, the bot finished its run by tweeting, with great fanfare in late 2014, the word “étui”(a handheld case for makeup or sewing supplies). I heard the bot’s author, Allison Parrish, speak after its run, discussing the changing public impact of a dataset when public engagement is prolonged, as well as the importance of Twitter’s ability to intersperse her (automated) tweets with those in any follower’s feed, thus juxtaposing automated tweets with the follower’s own news headlines and personal exchanges.

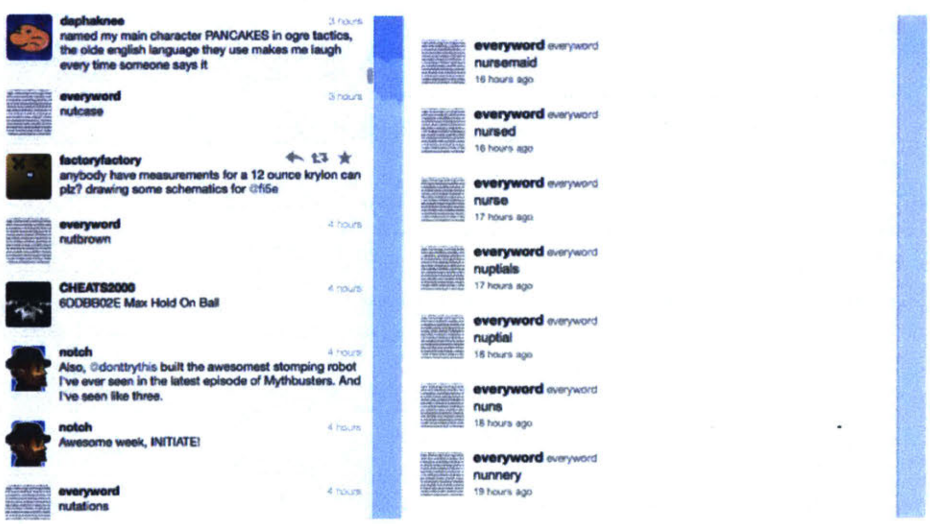


Figure 89 - 90: (left) external and (right) internal juxtapositions in @everyword.

Parrish speaks about the juxtaposition as being two-fold. First, Parrish was interested in the connection between the tweets contained within the @everyword feed(image right), which find grammatical connections between the words. Because words in the feed are sequenced alphabetically, they often share roots with the words they appear above or below in the feed. The second is the feed's interactions with external tweets that changed for every follower(image left). For example, at moments the words seem to provide commentary on the tweets that it follows.

Both of these forms of juxtaposition, internal and external, fascinated me. Following Parrish's example, I used the American Community Survey to make a similar Twitter bot. Every hour, the Twitter bot @censusAmericans tweets a mini biography of a real American from the PUMS dataset of the American Community Survey. Now in its second year of running, it has tweeted over 23 thousand people's biographies from the Census.

Although @censusAmericans did contain both types of juxtaposition that Parrish had highlighted from her experience, the Census data I used also gave different results. In the case of internal juxtapositions, @censusAmericans allows the reader to make comparisons between descriptions of people rather than words. Externally, people's lives were inserted at intervals in much the same way as @everyword. In the end, I learned that prolonged exposure to a dataset changes not only our understanding of the dataset but also our understanding of the underlying motivations of those who collect the data. Not only does the public's understanding of the Census as whole and what it does change through prolonged exposure to the dataset, but the public also gains a new closer perspective on the people the Census catalogs.

Although the tone of the Twitter bot was designed to be neutral, the cumulative effect of the tweets was sad. This shows how prolonged engagement could *complicate* our understanding of the data. Unemployment, divorce, and disability stood out in tweet after tweet. I chose to include every row of data (every person), but in order to accommodate tweet length, I constructed the biographies out of the most rarely occurring values in each row - that is, what I thought made each person special. A few thousand tweets later, it became clear that in the eyes of the Census, what makes people special are the ways in which they would require funding of services and programs of the government. For example, the questions regarding disability, unemployment, and change in marital status are used assess needs in different areas in order to plan for financial assistance services, assess laws and regulations on discrimination, evaluate existing and plan future programs. The capacity of underlying beliefs and intentions to color the data became clearer to me after this experience than it had been when I produced hundreds of visualizations the year before. Living with a dataset over time was an important aspect of understanding its character.

The second thing I learned is that data does not have to be comprehensive or complete in order to be evocative. @censusAmericans is limited to 140 characters, and it randomly chooses three to seven factoids out of each person's data. Smaller data points, or even just one datapoint, can communicate meaning if framed correctly.

Figure 91: Sample Tweets from @censusAmericans


SCIENCE

Continents Split Up at the Same Speed Finger Nails Grow. And That's Fast.

Trilobites

By NICHOLAS ST. FLEUR JULY 22, 2016

North America lurching away from Africa some 195 million years ago after spending tens of millions of years only slowly breaking apart.

Figure 92: Article explaining the speed of continents splitting up.

Small subjective representations of data can be evocative. The experimental geography concept of "geographical Intimacy" captures the ability of some artifacts to contextualize our larger environment. The feeling evoked by the concept is best expressed by facts or stories that have the power, allegorically and metonymically, to connect the awareness of a small detail to larger phenomenon. For example, the factoid that your fingernails grow at the same speed as tectonic plates move has the power to reduce large global phenomenon to human scale. In the context of interacting with data, methods and artifacts that make the abstract tangible are elusive, but important ones to seek out. Making public data more accessible is not just a matter of lowering technical barriers; it is also a matter of building relationships between datasets and a person's physical and emotional experience. I call this creating *data intimacy*.

censusPlaces @censusPlaces · 14 Mar 2017
J spent the last 6mins at a place where 0.86% work in Construction, Extraction, and Maintenance, Median Year Structures are Built is 19...

censusPlaces @censusPlaces · 14 Mar 2017
J spent the last 6mins at a place where 22.54% are Unemployed, 4.41% have a commute of Less than 10 Minutes, and 0% Housing Units use S...

censusPlaces @censusPlaces · 14 Mar 2017
J spent the last 4mins at a place where 0% drive to work, 0% work in Healthcare Support, and 0% Occupied Housing Units are Owner Occupied.

censusPlaces @censusPlaces · 14 Mar 2017
J spent the last 4mins at a place where 0.69% have a Doctorate Degree, 6.24% work in Office and Administrative Support, and 5.1% are Un...

censusPlaces @censusPlaces · 14 Mar 2017
J spent the last 6mins at a place where 0.72% are Black or African American, 7.39% are Asian, and 24.8% Housing Units use Electric heat.

censusPlaces @censusPlaces · 14 Mar 2017
J spent the last 34mins at a place where 9.17% are Unemployed, 2.6% have a Professional School Degree or More, and 0% work in Production.

censusPlaces @censusPlaces · 14 Mar 2017
J spent the last 4mins at a place where 0% are 85 or older, 57.84% have a Bachelor's Degree or More, and 0% Housing Units use Coal, Cok...

Figure 93: Sample tweets from @censusPlaces

Inspired by the concepts of longevity and small targeted data, I experimented with a second Twitter bot aimed only at myself. The @censusPlaces Twitter bot geolocates my phone every few minutes and alerts me via a tweet when I have moved to a different Census block group. The tweets include basic information about the location I have just left, including how long I spent there. Despite the similarities of this Twitter bot to the original @censusAmericans account, this bot's data did not have the cumulative power I intended. Over the course of three days, I was unable to make connections between the different places I visited or to see any trends or insights emerge from reading the tweets. Each Census block group was interesting in itself, and the experience did make me more aware of how unalike block groups in the same city can be. However, as the tweets accumulated, they became less readable as a dataset because the text-stream format of tweets did not lend itself to comparisons and summaries as a graph of the same series of places might have. I realized that while tracking places through the Census was an interesting concept, the method of communicating the data would have to very different, and most likely visual.

Following the initial Twitter experiment, I started tracking myself to generate interactive reports of where I had been and the places and data I had collected. The first iteration of this report was generated after one month of data collection and showed basic bar graphs of the block groups I had visited as they compared to the surrounding areas. These initial standard bar graphs served as the foundation for making a personalized atlas for one user.

User-Generated Atlas Report

In the months leading up to the experiment, I asked a group of five users to record their movements continuously for ten days on the Moves application. I then chose my user for this test case based on his interest in data visualization and in the amount of data he was able to collect. The data the user recorded using the Moves application included these dimensions: location (latitude, longitude), start time, end time, and duration (seconds between start time and end time). Over the recording period, which was ultimately twelve days, my user's application recorded 85 entries of 26 unique locations.

I then matched each recorded coordinate pair with its block group level Census geography. Using the American Community Survey, I created a "user places" dataset by filtering Census data for every place they traveled to. I created three more datasets that included block groups that filtered Census data by (within) quarter mile, half mile, and one mile radii of the places they traveled. In addition to these four location-filtered Census datasets, I also used state-wide user activity was contained in one state), and country-wide Census data for comparison.

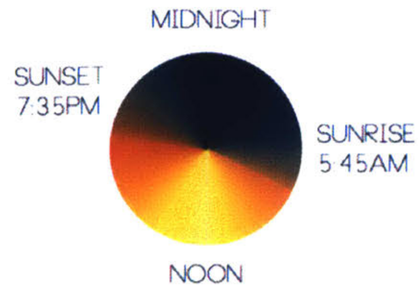
Ten visualizations containing a combination of the above graphical formats were shown to the user and followed by a long discussion of each.

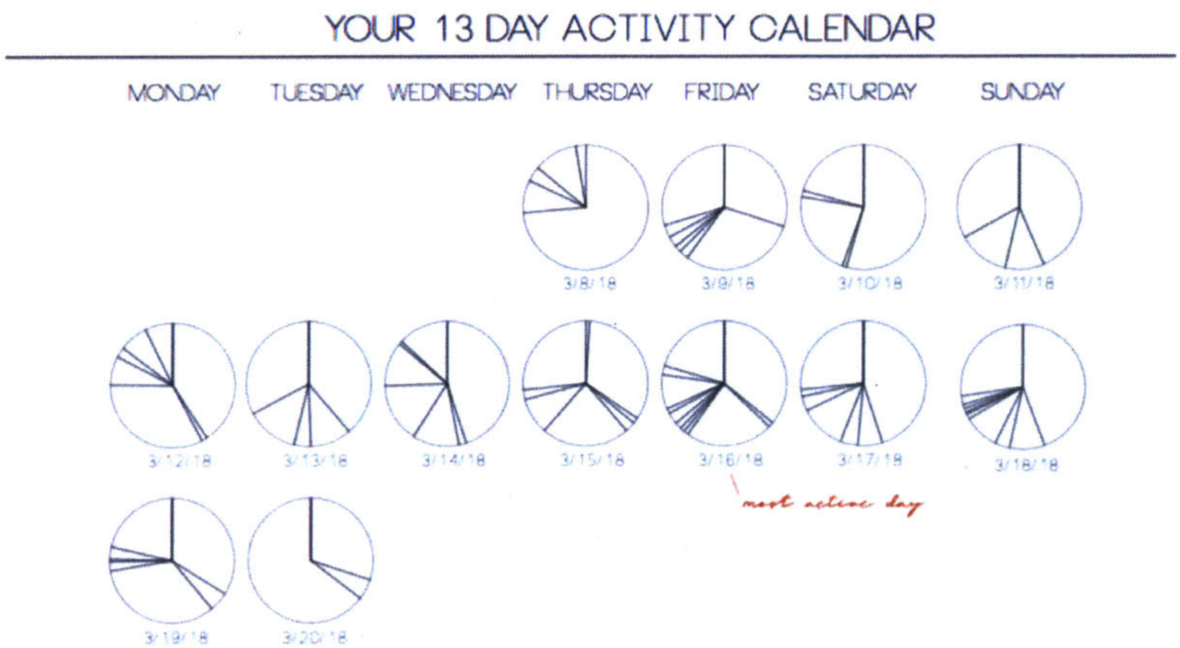
Below is the series of maps I showed to my user in the order I presented him with annotations of the discussion as well as the user's responses. (Figures 94 - 104)

I started by creating an overview of the user's activity to visualize the time of day, distance, and time spent at each location.

The user traveled several places only once. On the left is a heatmap of the Census Block Groups that the user traveled to in the test period. The longest time that the user in a single geography was 8 days, 6 hours, and 59 minutes - the darker triangle area in the upper center of the map.

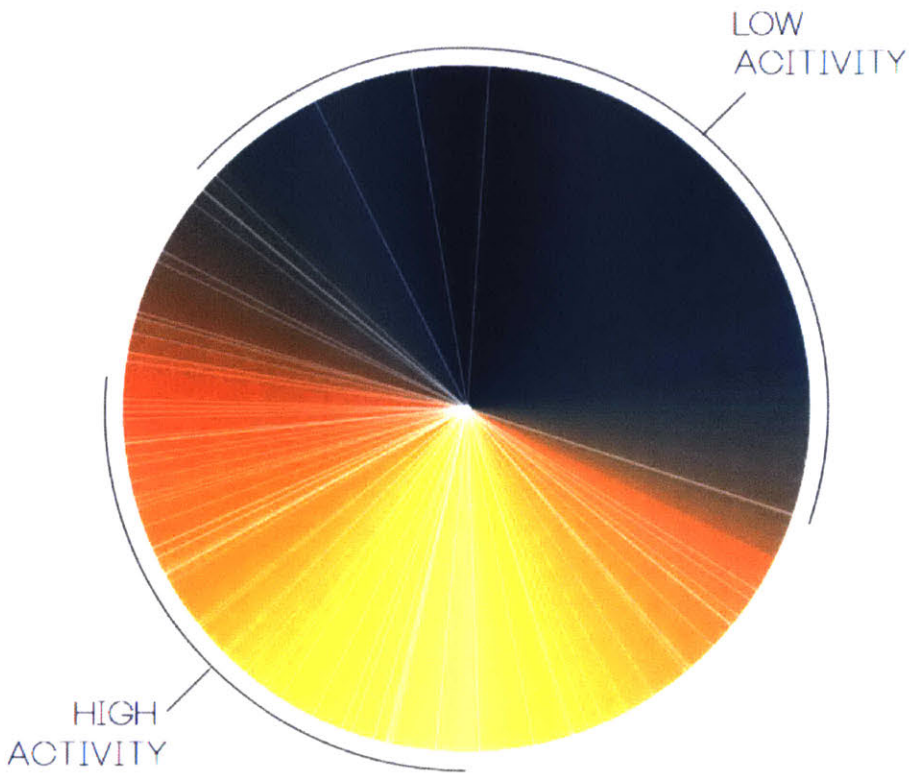
The user described the week in which the data was recorded as particularly active. For a start, he was able to confirm on the map above the areas he traveled to and locate their home and work. My user was consistently active for most of the hours between 9am and 9pm, and only made three trips over the test period outside of these hours. I represented the activity in a 24 hour clock visualization. On the right is the clock face.





Above is a calendar view of the times that the user made each trip mapped onto the clock face for each day. My user responded to this map by describing how long their days are, and how the calendar illustrates his weekly patterns of travel, such as the two days per week in which he picks up children. Their remarked that the long periods of inactivity were due to long uninterrupted hours of work where he often eats at his desk. They see the concentrated period of afternoon activity as a time period and opportunity to explore new locations.

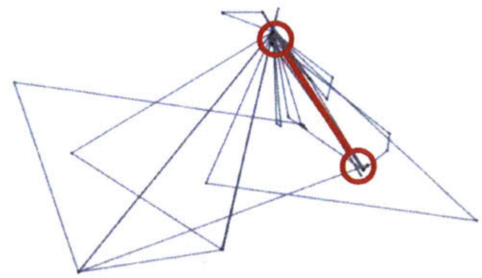
YOUR 24HOUR ACTIVITY CLOCK

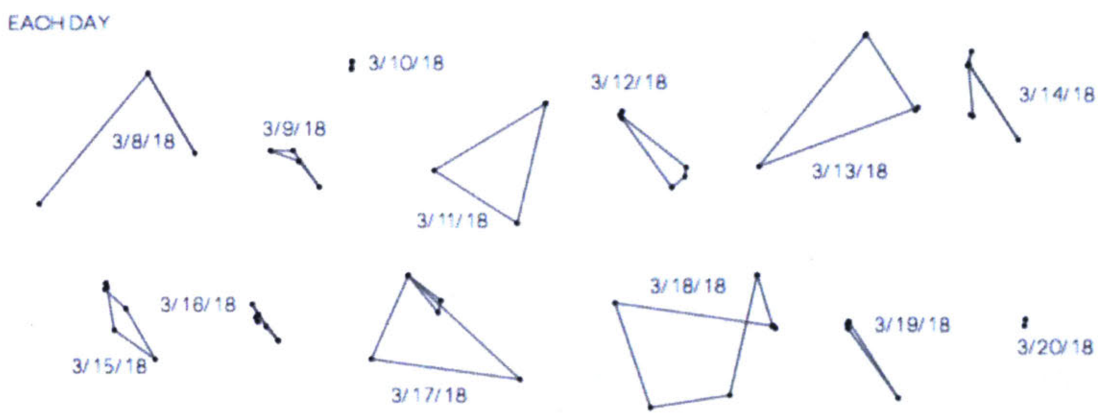


By overlaying all the days onto the same clock face, we can see concentrated activity over certain hours of the day above.

The user's paths(shown without the basemap on right) shows a pattern of traveling between two locations repeatedly. The testing period was a time that the user had a more diverse pattern than usual, resulting in little daily repetition in each day's pattern. Based on the number and timing of these trips, these locations are their daily commuting pattern between home and work. The edge formed between home and work can still be seen in red in the image to the right. Below are the trips separated by day. In this particular time period, the user made several trips to visit family and friends. His daily path form triangles on half of the of days in the dataset. While viewing this map, he observed that his usual routine is a triangle and that is "what his life looks like".

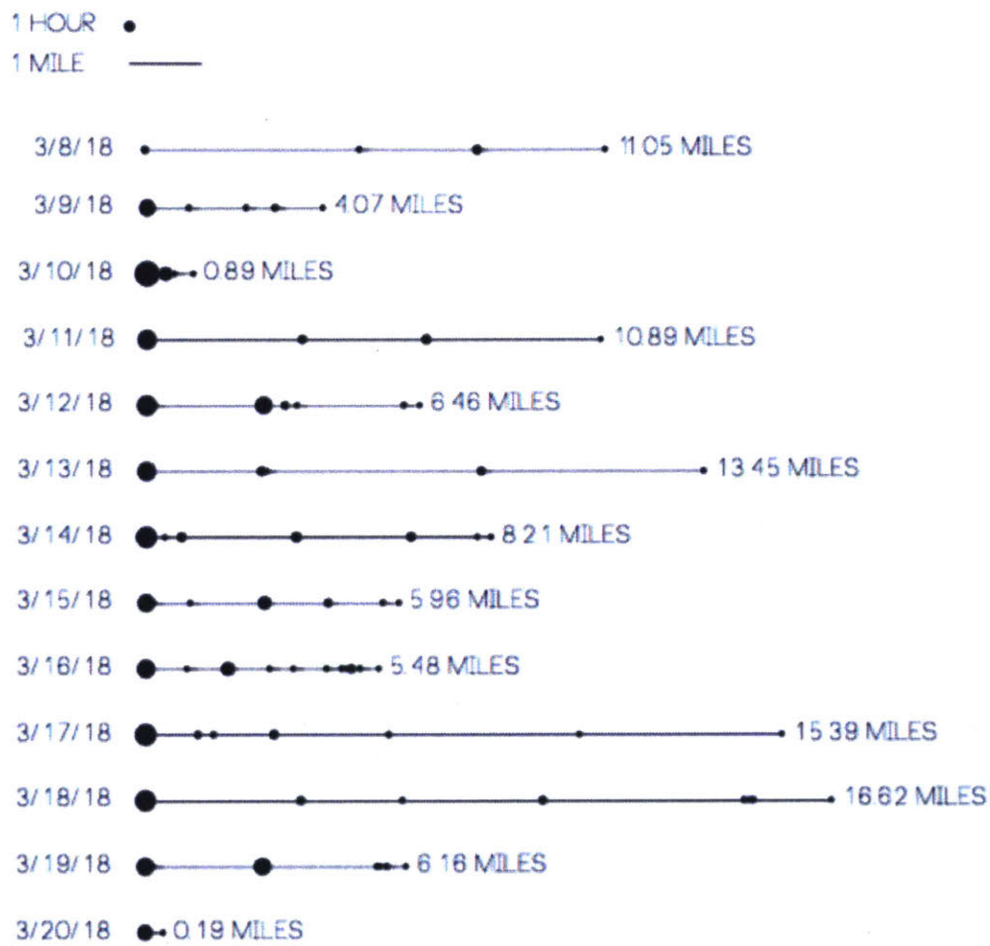
PATHS



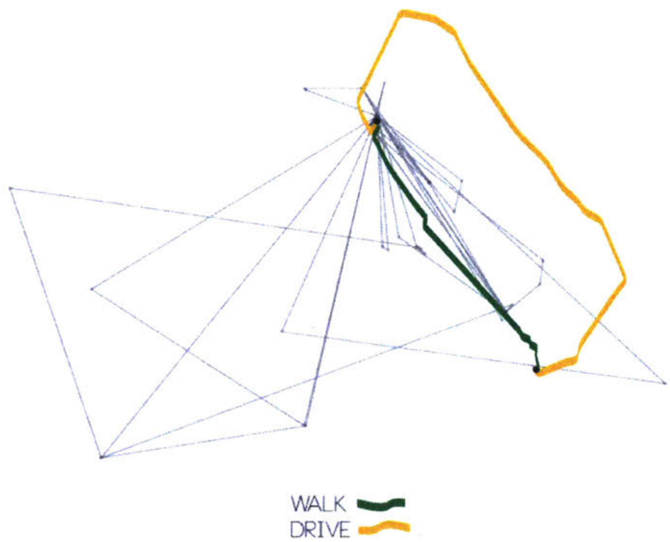


Using the distance between locations for the length of lines, and the time spent at each location as the size of the circles to mark each trip. Each day is drawn as a single line below. I then chose the day with the longest distance covered(the longest line below), the 18th of March to be a test date for further exploration.

DISTANCES/DURATION

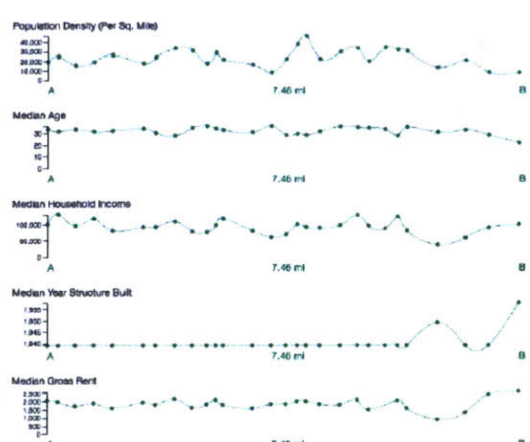


COMMUTE

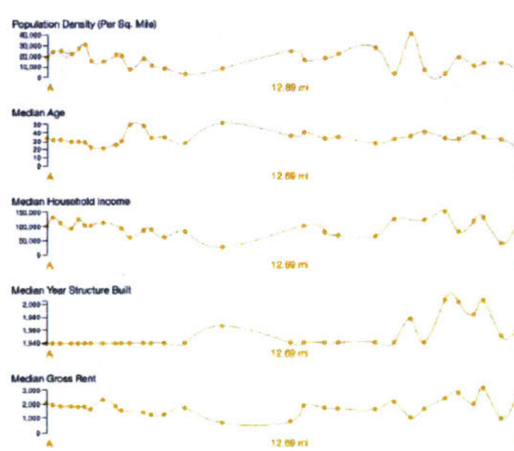


The locations tracked on the application are discrete points, and do not describe the route of travel between them. Using the directions API from Mapbox, I routed the pair of locations identified as home and work to make a turn by turn commuting route. The two routes describe driving and walking. The user commutes by public transportation, which in this case is the same route as walking between the points(the green line on the map). The income and building age graphs were the most descriptive and reflective of what the user thought of his commute. He commuted from an older residential neighborhood to a newly developed one, he also passes areas of wealth along the route that is reflected in the graph.

WALK

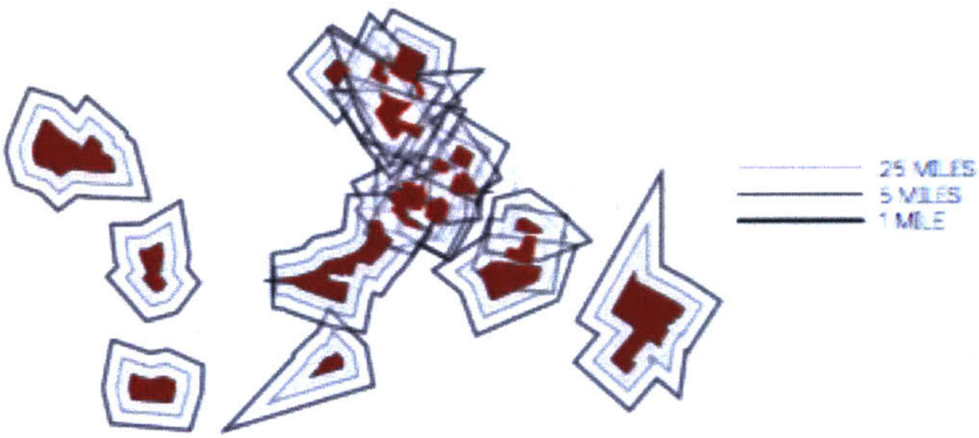


DRIVE

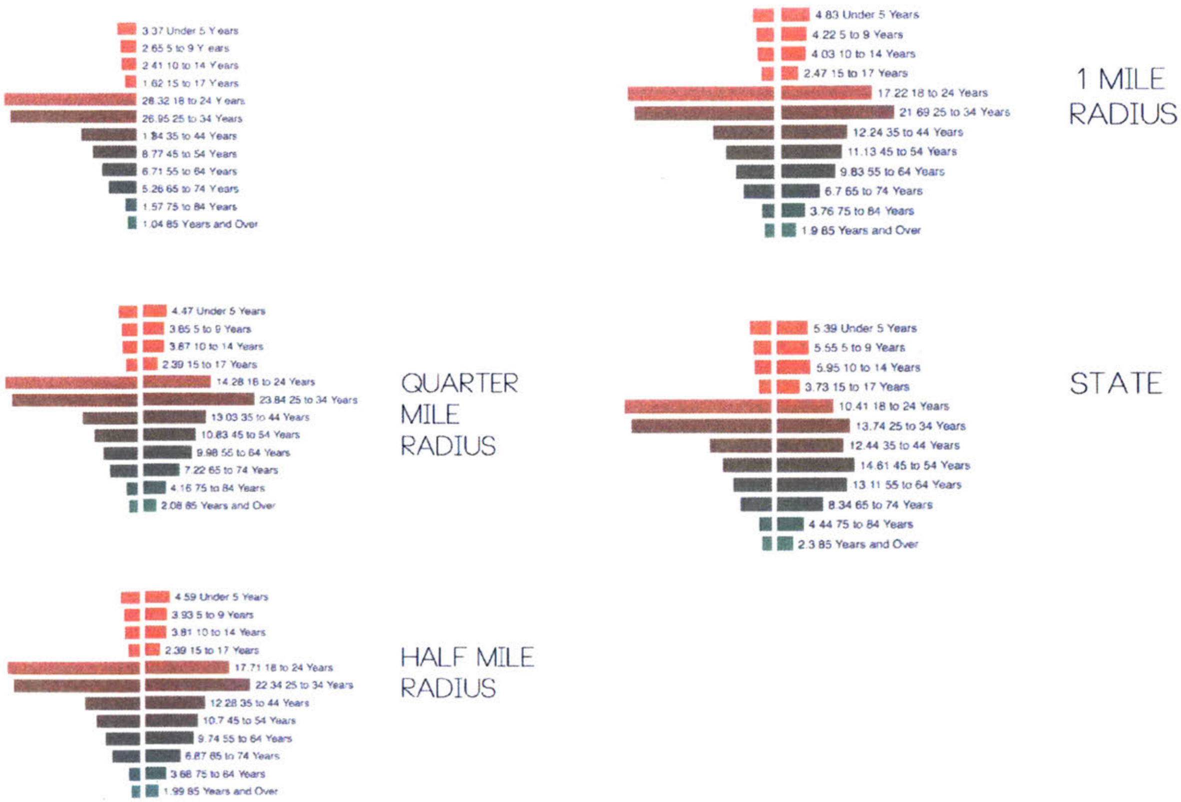


Using the radii, I then provided an overview of the demographics between the block groups the user visited, block groups within the three different radii. I presented these differences as a ranked list to provide an overview and also a bar graph for each category.

Here on the left are the block groups visited by the user in red. And the three different radii areas. This comparison was surprising for the user. He saw that areas he frequents have characteristics of being young reflected by the age, marital status, and percentage of renters.



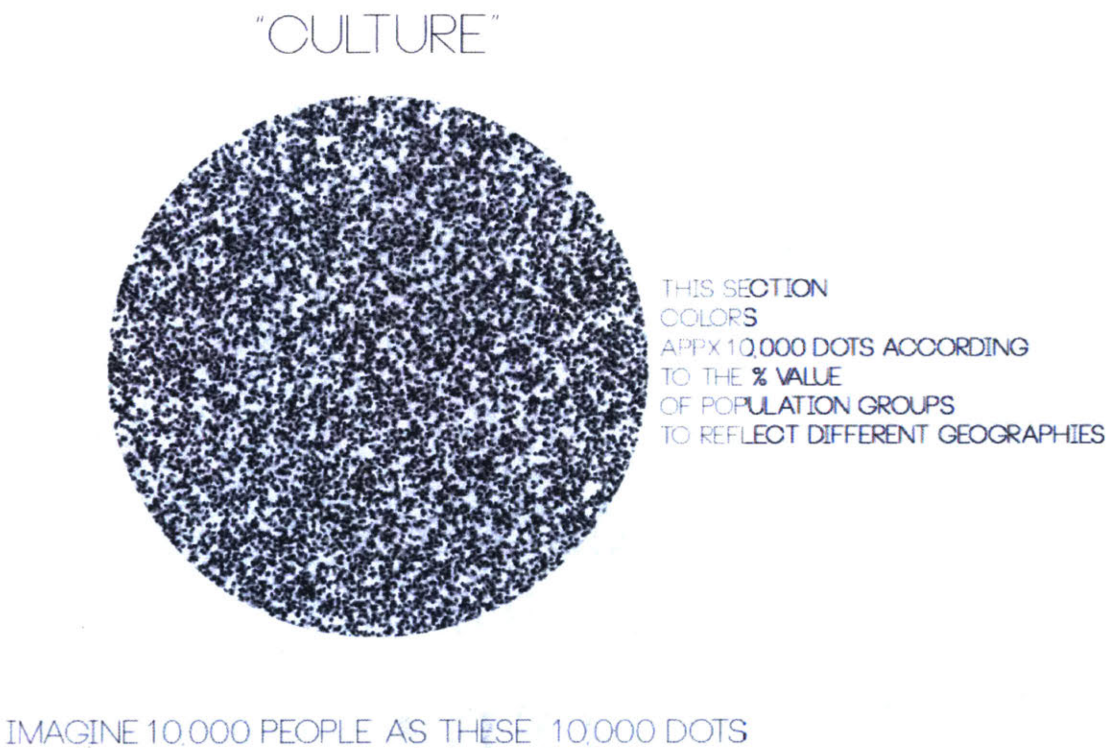
I used bar graphs to create comparisons between each of the geographies below. The top graph is the distribution of ages in the block groups that the user visited. The bottom two show the quarter and half mile radii areas on the right side, they are paired with original user's graph to make the comparisons more clear visually. The same format was also used for income, education, and race categories.



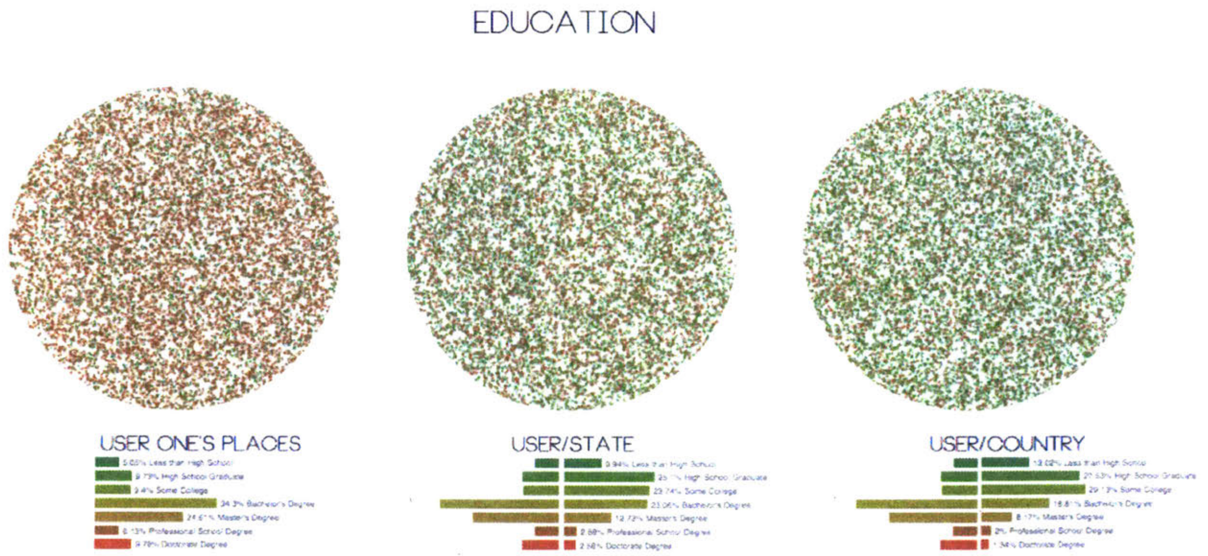
For the previous series of visualizations, the user was very interested his habits and especially aspects of his experience that he had previously thought about, and also saw reflected in the data. For example, the user is aware of the relatively higher percentage of "Hispanic or Latino" population in the area around his children's school which he also saw reflected in the data.

While the user expressed that he is interested in exploring new places that would expose him to different demographics, he also mentioned the difficulty of finding experiences and events that are in places and communities that are unknown to him. Overall, this first series of maps was able to ground our conversation on what the daily experience of demographics are, and how useful recommendations along those lines in future iterations might be helpful.

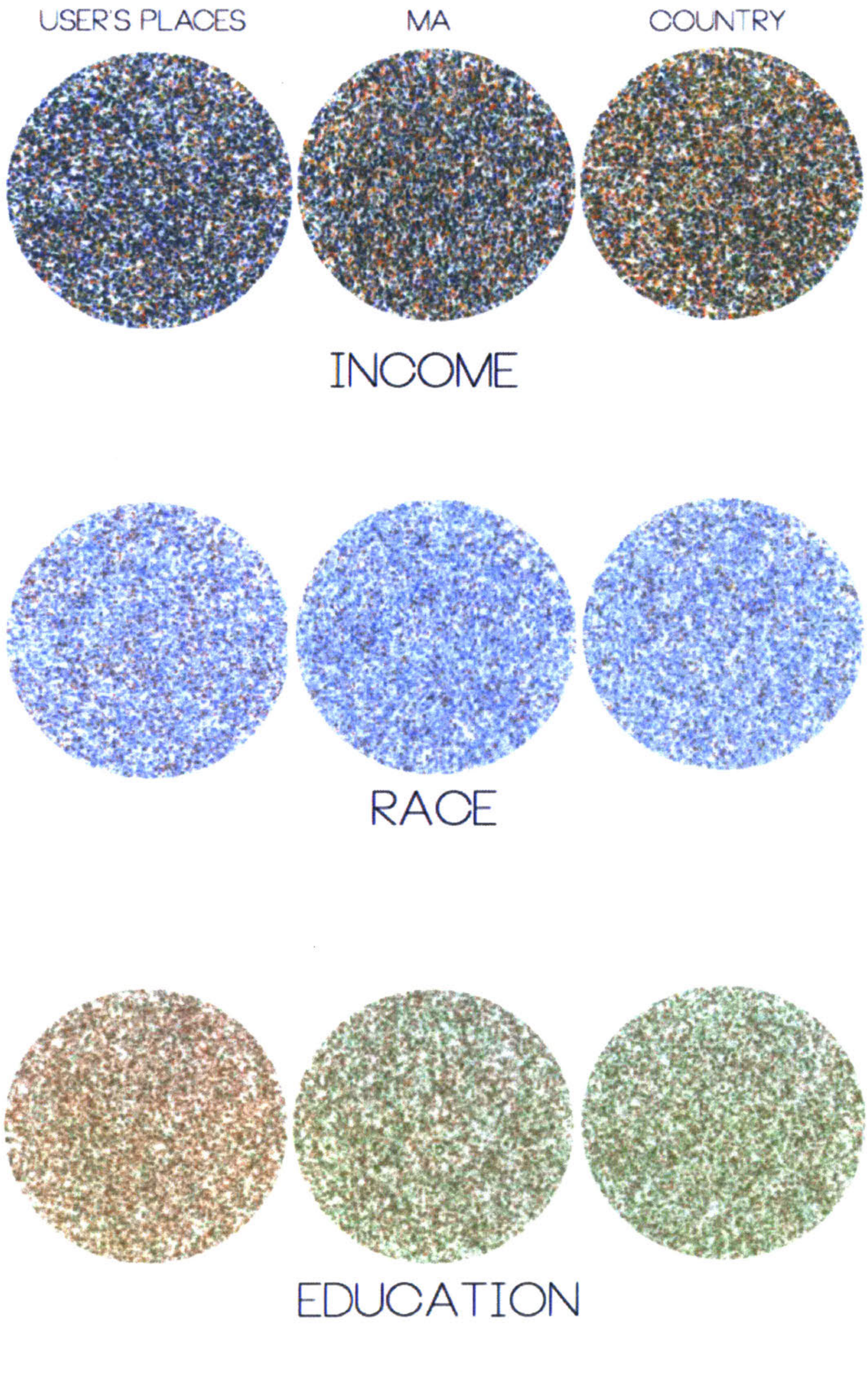
For a second design, I experimented with the idea of using dot density visualization methods for an alternative visual output that is meant to evoke petri dishes and cultures. I paired these dot density images with the previous bar graphs for clarity. I created a circular area with 10,000 dots in it, and used different demographic categories of different geographies to color these dots.



The user expressed interest in this type of visualization, especially the possibility of having interactive versions that gave a qualitative view of the data. Below are the dots colored according to the education level of three different geographies, the percentages of highest attained education level in the user's places, the statewide percentages, and the country-wide (from left to right). The user saw the different qualitatively, and was more interested in comparisons between his places and the state rather than the country level data.



is last set the visualizations allowed me to better understand what types of information might be useful over time. I believe that both single interaction and repeated interaction projects covered have potential to make Census data more functional through allowing it to be more personal. I was able to communicate to my last user the demographic context of the user's geospatial patterns and compare them with other geographies. However, within the scope of this experiment I was unable to provide my user with recommendations of places to alter the demographic context of their geospatial patterns. I believe that in future iterations of work that explores Census data as a self quantification dataset, recommendations based on the user's patterns should be included after the initial visualization phase that I have done above.


USER'S PLACES
MA
COUNTRY
INCOME
RACE
EDUCATION

Chapter 7 Conclusion

This chapter summarizes the maps produced for this dissertation and discusses their potential future development.

I began this thesis by asking if visualizations and maps can be designed to engage users with public aggregated datasets (such as the American Community Survey) in a way that users will find valuable in decision making as a part of their everyday personal lives. Specifically, I wanted to know if I could help users see personal utility in Census data by encouraging them to think about the contexts and implications of places, whether those places are locations they frequent, locations just outside of the habitual geospatial patterns they create, or locations that are new and unrelated to their daily travels.

What I found by making and testing iterative prototypes of maps, as well as by having conversations around these maps, is that people are genuinely interested not only in the contexts and implications of the places they inhabit, but also in how public data can be used to help them understand their environment at large.

Throughout this process, I have had the repeated experience of showing or explaining a map to a group (either at a conference, in a Meetup, in the lab, or in conversation) and receiving responses that were filled with curiosity about places people already occupied and also places that were occupied by those who led significantly different lives. Questions that people raised ranged from issues of fairness (how concentrations of elevator placement in a local subway system matched demographics) to personal utility (the design of a commute which gave most access to different neighborhoods) to issues of knowledge and connection (as, for example, when people expressed surprise that they could see such different lifestyles portrayed through data). These questions and conversations revealed to me that people already navigate spaces with purposes that are outside of the pure function of getting from one place to another, but also that their explorations are largely, as yet, unguided by data.

Towards a personal Census data

My experiments with using Census data, through various interactive maps and other digital tools, to communicate to individuals a sense of place and demographics, has led me to believe that making use of public data that foster personal understandings has the potential to deepen our engagement with the places we inhabit. The maps in this thesis serve as the foundation for future work in developing a Census Atlas for everyday use by citizens who wish to better understand themselves and their environment.

Four primary design principles emerged from the maps within this dissertation: (1) the communicative power of qualitative imagery, (2) the use of experiences to orient data, (3) the importance of long term interactions with data, and finally, (4) the value of creating personal atlases out of public aggregate data. These principles inform the next stage of the Census atlas, which is also outlined below.

Using Qualitative Data for Communication of the Quantitative

In the Satellite Comparison map, I explore the use of qualitative satellite images to provide additional context for the quantitative dataset of the American Community Survey. Rather than directly visualizing data such as concentrations of particular populations, the maps I produced use satellite images of the areas that contain that population to represent the data. In doing so, I was able to give a more relatable view of the data to which my users responded with great interest and often surprise. Although satellite images have been used in other projects as data, the interface of the Satellite Comparison maps are novel in their use of how quantitative data is used to filter images to created landscapes. I believe that many quantitative mapping endeavors will benefit from similar usage of images. As satellite images become more accessible as a dataset for use by individuals and small groups alike, the purposes they can serve will outgrow its current limited research and commercial applications to become a means to communicate conditions and qualities in a rich and informative manner.

Experiential Orientation

In the Sightlines map and Cross Sections map, I explore how the juxtaposition of physical experiences can be used as a filter for data. Experiences can be used to highlight different aspects of a place. These explorations led me to believe that communicating data in the moment to users can enrich their understanding of a place, whether it is the hidden power dynamics of the seen and unseen, or the profiles users cut across cities with their own movement.

Long-term Engagement

Since creating the @censusAmericans twitter bot and my year of daily mapping with youarehere, I have been interested in the cumulative power of communicating small bits of data over an extended period of time. More recently, asking a single user to track his location over time and visualizing that data was also a revealing experience for both the user and myself. Within the scope of the dissertation, I was not able to provide personal tracking over longer periods of time and communicate results to the user intermittently. However, the experiences above indicated that our relationship with datasets change over time as we interact with them, and that longer and more regular interactions would be informative. Throughout the series of maps I created, and especially in the last maps on tracking that I present in this dissertation, I used myself and others to gauge the usefulness of tracking the spaces we occupy through Census data. Personal context can be a powerful way to understand aggregate datasets. For this reason, I believe that the public impact of Census data can be improved by giving people tools to use Census data for self-quantification; conversely, people can use Census data to improve their own self-understanding. As a result, I believe that in future work, the format of a mobile application as the delivery method of personalized aggregated public data is worthy of further exploration.

Personal Context and the Atlas

Maps have traditionally implied authority of the territory they depict. Critical cartographic practices have produced works that subvert the authoritative voice and the perceived objectivity often attributed to the top-down view of the map. From Wood's atlas of his hometown that turns mapping conventions on its head by purposefully mapping the idiosyncratic, to practices like the dérive that attempt to reimagine the structures that city streets impose on our personal lives, works in this field have long challenged how space and spatial data should be visualized and utilized. I believe that the process of personalization, through our own movement, of inherently impersonal aggregate public datasets will extend the concepts within the critical cartography framework. Specifically, the use of cartography to mark territory and impose authority of what is mapped can be challenged by the act of individuals building their own territories and maps for introspection rather than outward expansion.

The idea that self quantification is the measurement of internal states to improve self knowledge can be complicated by the act of using external factors and strangers in our environments for the same purpose. By framing places we travel to as a culture from which we draw nourishment, and our exposure to strangers as a part of our geospatial diet, self quantification becomes less about how we behave alone, and more about how we contribute to a larger community.

Finally, as an extension of the idea that atlases require unifying themes and narratives, the construction of an atlas entirely driven by personal filtering of aggregate data places the user as the author of the atlas narrative, and the life of the user as the unifying theme of the maps displayed. In this case, the purpose of the atlas is not to describe the world with a specific perspective, but to mark the territories that constructs the highly personal worldview of a single user.

In addition to the four themes described above, two questions were raised during my defense presentation that are important to note. The first is the use of the maps and data contained within them to address less urban environments. The second is the use of the maps by populations that may live as minorities inside otherwise homogeneous places.

Although countrywide data drive the maps that I have built within this dissertation, the examples I have focused on have used urban settings to illustrate each map's use. I believe that using the categories of race, income, education, profession, and age to visualize similarities and differences between places where population is less dense is an interesting challenge. The satellite comparisons maps showed that certain very dense urban settings and rural settings shared certain characteristics, such as having both a high percentage of high school dropouts and longer average commuting times. It is possible that rural areas have as much as if not more differences than populous places, but those differences are not made visible through the visualizations I have provided because the visualizations and topics for comparison originate from examinations of urban settings. In order to find these differences, I would use existing maps to identify areas with lower population density and test varying degrees of thresholds for differences between areas in order to find the correct magnitude of disparity to measure.

An early assumption I made while building the maps was that individuals could use the American Community Survey to understand aspects of their everyday experiences in terms of lacking in diversity and seek out diversity where possible. What I did not take into account were users who live within homogeneous environments that belonged to minority groups and could also use such data to seek out communities of people with whom they have more in common. Although I did not implement designs with this goal in mind, I believe that recommendations of places to visit and the addition of search functions could adapt the maps for this purpose.

Taking these six summary points into consideration, some future tasks comes to mind that will contribute to the building of a personalized Census atlas, work which I will pursue in the next few years as I continue this research. Specifically, the ability for users to identify, over time, what is lacking in their geospatial diet and to locate communities that they would like to engage with through visualized public aggregate data are the primary goals of the next phase of my research.

Census data, as an example of public data, describes our environment in a consistent but highly specific way. Through it we can better understand aspects of our daily behavior in the context of where we live and where we visit. It is my hope that when we have data and the means to locate ourselves in the larger system of our environment, we also become more responsible for how we contribute to it.

References

1. Aitken, Stuart, and James Craine. "Guest Editorial: Affective Geovisualizations." *Directions Magazine* 7(2006)
2. Akerman, James, R.'From books with maps to books as maps: the editor in the creation of the atlas idea', Editing Early and Historical Atlases: Papers Given at the Twenty-ninth Annual Conference on Editorial Problems, ed. by Winearls, J., Toronto: University of Toronto Press, 1995. 3-48.
3. Alexander, Christopher, and Sara Ishikawa. A Pattern Language: Towns, Buildings, Construction. New York: Oxford University Press, 1977.
4. Anselin, Luc., Y.-W. Kim, and I. Syabri.. "Web-Based Analytical Tools for the Exploration of Spatial Data." *Journal of Geographical Systems* 6 (2004): 197–218.
5. Anselin, Luc, Ibnu Syabri, and Youngihn Kho. "GeoDa: An Introduction to Spatial Data Analysis." *Handbook of Applied Spatial Analysis* (2009): 73-89.
6. Bedö, Viktor. "A Visual Approach to Locative Urban Information." *Handbook of Research on Urban Informatics: The Practice and Promise of the Real-time City*. By Marcus Foth. Hershey, PA: Information Science Reference, 2009. 219-229.
7. Bertin, Jacques. Semiology of Graphics. Madison, WI: U of Wisconsin, 1983.
8. Biermann, Kai. "Data Protection: Betrayed by our own data." *ZEIT ONLINE*. ZEIT ONLINE Nachrichten auf ZEIT ONLINE, 10 Mar. 2011. Web. 18 Mar. 2017.
9. Board, Christopher. "Cartographic Communication." *Cartographica: The International Journal for Geographic Information and Geovisualization* 18.2 (1981): 42-78.
10. Borkin, Michelle A., Zoya Bylinskii, Nam Wook Kim, Constance May Bainbridge, Chelsea S. Yeh, Daniel Borkin, Hanspeter Pfister, and Aude Oliva. "Beyond Memorability: Visualization Recognition and Recall." IEEE Trans. Visual. Comput. Graphics IEEE Transactions on Visualization and Computer Graphics 22.1 (2016): 519-28.
11. Bostock, Michael, and Jason Davies. "Code as Cartography." *The Cartographic Journal* 50.2 (2013): 129-35.
12. Bostock, Mike, V. Ogievetsky, and J. Heer. "D³ Data-Driven Documents." IEEE Trans. Visual. Comput. Graphics IEEE Transactions on Visualization and Computer Graphics 17.12 (2011): 2301-309.
13. Burns, Ryan, and André Skupin. "Towards Qualitative Geovisual Analytics: A Case Study Involving Places, People, and Mediated Experience." Cartographica: The International Journal for Geographic Information and Geovisualization 48.3 (2013): 157-76.
14. Calabrese, Francesco, Kristian Kloeckl, and Carlo Ratti. "WikiCity: Real-Time Location-Sensitive Tools for the City." Handbook of Research on Urban Informatics: The Practice and Promise of the Real-time City. By Marcus Foth. Hershey, PA: Information Science Reference, 2009. 390-413.
15. Caquard, Sébastien. "Cartography I: Mapping Narrative Cartography." *Progress in Human Geography* 37.1 (2011): 135-44.
16. Card, S.K., and J. Mackinlay. "The Structure of the Information Visualization Design Space." Proceedings of VIZ '97: Visualization Conference, Information Visualization Symposium and Parallel Rendering Symposium (1997).
17. Cattoor, Bieke, and Chris Perkins. "Re-cartographies of Landscape: New Narratives in Architectural Atlases." *The Cartographic Journal* 51.2 (2014): 166-78.

18. Certeau, Michel De., and Steven Rendall. "Walking in the City." The Practice of Everyday Life. Berkeley: U of California, 1984. 91-110.
19. Chen, Chaomei, and Mary P. Czerwinski. "Empirical Evaluation of Information Visualizations: An Introduction." International Journal of Human-Computer Studies 53.5 (2000): 631-35.
20. Cleveland, William S., and Robert Mcgill. "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods." Journal of the American Statistical Association 79.387 (1984): 531-54.
21. Cleveland, William S. *Visualizing Data*. Murray Hill, NJ: At & T Bell Laboratories, 1993.
22. Ciuccarelli, Paolo, Giorgia Lupi, and Luca Simeone. Visualizing the Data City Social Media as a Source of Knowledge for Urban Planning and Management. Cham: Springer International, 2014. Print.
23. Corner, James. "The Agency of Mapping: Speculation, Critique and Invention." *Mappings*. Ed. Denis Cosgrove. London: Reaktion, 1999. 213-252.
24. Crampton, Jeremy W., and John Krygier. "An Introduction to Critical Cartography." ACME: An International E-Journal for Critical Geographies 4.1 (2005): 11-33. 4 Nov. 2015.
25. Currid, Elizabeth., and Sarah. Williams. "Two Cities, Five Industries: Similarities and Differences within and between Cultural Industries in New York and Los Angeles." Journal of Planning Education and Research 29.3 (2010): 322-35.
26. Dalton, Craig M. "Sovereigns, Spooks, and Hackers: An Early History of Google Geo Services and Map Mashups." *Cartographica: The International Journal for Geographic Information and Geovisualization* 48.4 (2013): 261-74.
27. Debord, Guy (1956). "Theory of the Derive". *Situationist International Online*. Translated by Ken Knabb.
28. Denil, Mark. "Cartographic Design: Rhetoric and Persuasion." *Cartographic Perspectives* 45 (2003): 8-67.
29. Dodge, Martin, and Rob Kitchin. "'Outlines of a World Coming into Existence': Pervasive Computing and the Ethics of Forgetting." *Environment and Planning B: Planning and Design*34.3 (2007): 431-45. Web.
30. Dodge, Martin, Rob Kitchin, and C. R. Perkins. "Chapter 1 Thinking about Maps." *Rethinking Maps: New Frontiers in Cartographic Theory*. London: Routledge, 2009. 1-25
31. Dubey, Abhimanyu, Nikhil Naik, Devi Parikh, Ramesh Raskar, César A. Hidalgo. “Deep Learning the City : Quantifying Urban Perception At A Global Scale” 2016
32. Eccles, Ryan, Thomas Kapler, Robert Harper, and William Wright. "Stories in GeoTime." *2007 IEEE Symposium on Visual Analytics Science and Technology* (2007): 19–26.
33. Engelhardt, Yuri. "Syntactic Structures in Graphics." *Computational Visualistics and Picture Morphology* 5.1 (2007): 23-35.
34. Farman, Jason. Mobile Interface Theory: Embodied Space and Locative Media. New York: Routledge, 2012.
35. Farman, J. "Mapping the Digital Empire: Google Earth and the Process of Postmodern Cartography." *New Media & Society* 12.6 (2010): 869-88.
36. Fitbit IPO, https://www.sec.gov/Archives/edgar/data/1447599/000119312515176980/d875679ds1.htm
37. Fonseca, Margarida, flowingcity.com
38. Ford, Simon. The Situationist International: a User's Guide. Black Dog Publishing, 2005.
39. Galloway, Anne. "Intimations of Everyday Life: Ubiquitous Computing and the City." *Cultural Studies* 18.2-3 (2004): 384-408.Gerlach, Joe. "Vernacular Mapping, and the Ethics of What

Comes Next." Cartographica: The International Journal for Geographic Information and Geovisualization 45.3 (2010): 165-68.
40. Goldsmith, Stephen A., and Lynne Elizabeth. What We See: Advancing the Observations of Jane Jacobs. Oakland, CA: New Village, 2010. Print.
41. Goodchild, Michael F. "Citizens as Sensors: The World of Volunteered Geography." *GeoJournal* 69.4 (2007): 211-21.
42. Greco, Kael,. "Seeing the City through Data/Seeing Data through the City." *Decoding the City*. By Carlo Ratti and Dietmar Offenhuber. Basel: Birkhäuser Verlag GmbH, 2014.125-143
43. Haklay, Muki, Alex Singleton, and Chris Parker. "Web Mapping 2.0: The Neogeography of the GeoWeb." *Geography Compass* 2.6 (2008): 2011-039.
44. Harley, J. B. "Deconstructing the Map." *Theories of Mapping Practice and Cartographic Representation The Map Reader* (2011): 56-64.
45. Heer, Jeffrey, and Ben Shneiderman. "Interactive Dynamics for Visual Analysis." *Queue* 10.2 (2012): 30.
46. Heer, Jeffrey, and Michael Bostock. "Crowdsourcing Graphical Perception." Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10 (2010)
47. Heer, Jeffrey, and Michael Bostock. "Declarative Language Design for Interactive Visualization." *IEEE Trans. Visual. Comput. Graphics IEEE Transactions on Visualization and Computer Graphics* 16.6 (2010): 1149-156.
48. Heer, Jeffrey, Michael Bostock, and Vadim Ogievetsky. "A Tour through the Visualization Zoo." Communications of the ACM Commun. ACM 53.6 (2010): 59. Web.
49. Heyen, Niles. B. (2016). Self-Tracking as Knowledge Production: Quantified Self between Prosumption and Citizen Science. *Lifelogging,*283-301. doi:10.1007/978-3-658-13137-1_16
50. Hidalgo, Cesar, Elisa Castaner, "Do we need another coffee house? The amenity space and the evolution of neighborhoods", Macro Connections, The MIT Media Lab
51. Jacobs, Jane. The Death and Life of Great American Cities: Orig. Publ. 1961. New York: Vintage, 1992.
52. Jakobsen, Mikkel Rønne, and Kasper Hornbæk. "Sizing up Visualizations." *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI '11* (2011): 1451-460
53. Jenks, George F. "Generalization In Statistical Mapping." *Annals of the Association of American Geographers* 53.1 (1963): 15-26.
54. Kapler, Thomas, and William Wright. "GeoTime Information Visualization." *Inf Vis Information Visualization* 4.2 (2005): 136-46.
55. Kloeckl, Kristian. "The City as a Digital Public Space - Notes for the Design of Live Urban Data Platforms." *Decoding the City*. By Carlo Ratti and Dietmar Offenhuber. Basel: Birkhäuser Verlag GmbH, 2014. 82-95.
56. Krambeck, Holly, World Bank Open Traffic Project, http://www.worldbank.org/en/news/press-release/2016/12/19/the-world-bank-launches-new-open-transport-partnership-to-improve-transportation-through-open-data
57. Kunkel, Thomas. "The Master Writer of the City." New York Review of Books. N.p., 23 Apr. 2015. Web. 2017.
58. Kurgan, Laura. Close up at a Distance: Mapping, Technology, and Politics. Cambridge: MIT, 2013

59. Lazer, David., A. Pentland, L. Adamic, S. Aral, A.-L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne. "SOCIAL SCIENCE: Computational Social Science."Science 323.5915 (2009): 721-23. Web.
60. LeGates, Richard T. "How to Study Cities." The City Reader. Ed. Richard T. LeGates and Frederic Stout. 5th ed. New York: Routledge, 2011. 7-12.
61. Lin, Wen. "Tracing the Map in the Age of Web 2.0." *Cartographica: The International Journal for Geographic Information and Geovisualization* 50.1 (2015): 41-44.
62. Lloyd, Robert. "Understanding and Learning Maps." Ed. Scott Freundschuh. *Cognitive Mapping: Past, Present and Future.* Ed. Rob Kitchin. London: Routledge, 2000. 84-
63. Low, Setha. "Spatializing Culture: An Engaged Anthropological Approach to Space and Place." *The People, Place, and Space Reader.* Ed. Jen Jack, Setha Low, and Susan Saegert. New York: Routledge, 2014. 34-38.
64. Lynch, Kevin. The Image of the City. Cambridge, MA: MIT, 1960.
65. Lynch, Kevin. *What TIme is this Place?*. Cambridge, MA: MIT, 1972.
66. Maantay, Juliana, and John Ziegler. *GIS for the Urban Environment.* Redlands, CA: ESRI, 2006
67. Maceachren, Alan M., and Menno-Jan Kraak. "Exploratory Cartographic Visualization: Advancing the Agenda." *Computers & Geosciences* 23.4 (1997): 335-43.
68. Malczewski, Jacek, *GIS-based multicriteria decision analysis: a survey of the literature,* International Journal of Geographical Information Science, 20:7, 703-726, 2006
69. Matless, David. "The Uses of Cartographic Literacy: Mapping, Survey and Citizenship in Twentieth-Century Britain." *Mappings.* Ed. Denis Cosgrove. London: Reaktion, 1999. 193-212.
70. McCloud, Scott. Understanding Comics:. New York: HarperPerennial, 1994.
71. Milgram, Stanley. 1970. "The Experience of Living in Cities." *Science* 167(3924):
72. Mint by the Numbers: Which User Are You? (2016, April 26). Retrieved March 30, 2017, from https://blog.mint.com/credit/mint-by-the-numbers-which-user-are-you-040616/
73. Monmonier, Mark S. *How to Lie with Maps.* Chicago: U of Chicago, 1991.
74. Muehlenhaus, Ian. "Going Viral: The Look of Online Persuasive Maps." *Cartographica: The International Journal for Geographic Information and Geovisualization* 49.1 (2014): 18-34.
75. Naik, Nikhil, Jade Philipoom, Ramesh Raskar, César Hidalgo, "Street Score: Predicting the perceived Safety of One Million Streetscapes" 2014
76. Opach, Tomasz, Izabela Gołębiowska, and Sara Irina Fabrikant. "How Do People View Multi-Component Animated Maps?" *The Cartographic Journal Cartogr.* J. 51.4 (2014): 330-42.
77. Pallasmaa, Juhani, "Toward an architecture of humility: on the value of experience." *The People, Place, and Space Reader.* Ed. Jen Jack, Setha Low, and Susan Saegert. New York: Routledge, 2014. 330 - 333.
78. Pickles, John. *Ground Truth: The Social Implications of Geographic Information Systems.* New York: Guilford, 1995
79. Rankin, William. After the Map Cartography, Navigation, and the Transformation of Territory in the Twentieth Century. Chicago: U of Chicago, 2016. Print..
80. Robinson, Arthur H, and Barbara B. Petchenik. *The Nature of Maps: Essays Toward Understanding Maps and Mapping.* Chicago: University of Chicago Press, 1976.
81. Rivera-Pelayo, Verónica, Valentin Zacharias, Lars Müller, and Simone Braun. "Applying quantified self approaches to support reflective learning." *Proceedings of the 2nd*

International Conference on Learning Analytics and Knowledge - LAK '12(2012): n. pag. Web.
82. Satyanarayan, Arvind, and Jeffrey Heer. "Lyra: An Interactive Visualization Design Environment." *Computer Graphics Forum* 33.3 (2014): 351-60.
83. Segel, Edward., and Jeffery. Heer. "Narrative Visualization: Telling Stories with Data." IEEE Trans. Visual. Comput. Graphics IEEE Transactions on Visualization and Computer Graphics 16.6 (2010): 1139-148.
84. "Share Your Vision. Shape Our City." Imagine Boston 2030. N.p., n.d. Web. 05 Mar. 2017.
85. Shneiderman, B. "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations." Proceedings 1996 IEEE Symposium on Visual Languages (1996).
86. Slocum, Terry A., and Terry A. Slocum. *Thematic Cartography and Geographic Visualization*. Upper Saddle River, NJ: Pearson/Prentice Hall, 2005
87. St. Fleur, Nicholas, July 22, 2016, New York Times Science Section, https://www.nytimes.com/2016/07/23/science/continental-drift-tectonic-plates.html
88. Taylor, D.r. Fraser. "The Concept of Cybercartography." *Maps and the Internet* (2003): 405-20.
89. Thompson, Nato, Jeffrey Kastner, and Trevor Paglen. Experimental geography. Brooklyn, NY: Melville House, 2008. Print.
90. Tobler, Waldo R. "Automation and Cartography." *Geographical Review* 49.4 (1959): 526.
91. Tobler, Waldo. R. "Analytical Cartography." *Cartogr Geogr Inf Sci Cartography and Geographic Information Science* 3.1 (1976): 21-31.Turnbull, David. "Mapping encounters and (en)countering maps: A critical examination of cartographic resistance," *Knowledge and Society* 11 (1998): 15–44.
92. Ware, Chris, Information visualization: perception for design. Morgan Kaufmann, San Francisco, CA, 2004.
93. Wehrend, S., and C. Lewis. "A Problem-oriented Classification of Visualization Techniques." *Proceedings of the First IEEE Conference on Visualization: Visualization `90* (1990).
94. Whitson, Jennifer. R. (2013). Gaming the Quantified Self. *Surveillance & Society, 11*(1/2), 163-176.
95. Wilson, Chris, and Alexander Ho. "Time Labs: Open-source Data Visualization from Time.com." Time. Time, 11 Feb. 2015. Web. 05 Mar. 2017.
96. Wolf, G., & Kelly, K. (n.d.). Self Knowledge Through Numbers. Retrieved March 30, 2017, from http://quantifiedself.com/
97. Wood, Denis. "The Anthropology of Cartography." *Mapping Cultures* (2012): 280-303.
98. Wood, Denis. "How Maps Work." Cartographica: The International Journal for Geographic Information and Geovisualization. 29.3-4 (1992): 66-74.
99. Wood, Denis. "Mapping Deeply." Humanities 4.3 (2015): 304-18.
100. Wood, Denis, Ira Glass, Blake Butler, Ander Monson, and Albert Mobilio. Everything Sings: Maps for a Narrative Atlas. Los Angeles, CA: Siglio, 2013. Print.
101. Wood, Denis. "Pleasure In The Idea/the Atlas As Narrative Form." *Cartographica: The International Journal for Geographic Information and Geovisualization* 24.1 (1987): 24-46.
102. Wood, Denis, Ward L. Kaiser, and Bob Abramms. *Seeing through Maps: Many Ways to See the World*. Oxford: New Internationalist, 2006.
103. Woodruff, Andy, and Tim Wallace, "Map your neighborhood! ", 2015, http://bostonography.com/hoods/

104. Young, Nora. (2013). *The virtual self: how our digital lives are altering the world around us.* Toronto, Ontario: McClelland & Stewart.
105. Zhang, J., Wang, F., Wang, K., Lin, W., Xu, X., & Chen, C. (2011). Data-Driven Intelligent Transportation Systems: A Survey. *IEEE Transactions on Intelligent Transportation Systems, 12*(4), 1624-1639. doi:10.1109/tits.2011.2158001

Figures