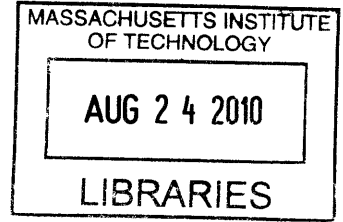


ScratchStats: A Site for Visualizing and Understanding Scratch Usage Data

by
Rita Chen



Submitted to the Department of Electrical Engineering and Computer Science

in partial fulfillment of the requirements for the degree of

ARCHIVES

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2010

© Massachusetts Institute of Technology 2010. All rights reserved.

Author
Department of Electrical Engineering and Computer Science
May 21, 2010

Certified by
Mitchel Resnick
LEGO Papert Professor of Learning Research, MIT Media Lab
Thesis Supervisor

Accepted by
15
Dr. Christopher J. Terman
Chairman, Department Committee on Graduate Theses

ScratchStats: A Site for Visualizing and Understanding Scratch Usage Data

by

Rita Chen

Submitted to the
Department of Electrical Engineering and Computer Science

May 21, 2010

In partial fulfillment of the requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

This thesis introduces ScratchStats, an extension to the Scratch website where users can view and understand Scratch usage data through a series of interactive visualizations. Scratch is a visual programming language that makes it easy to create interactive stories, games, and artwork. Accompanying the Scratch application is the Scratch Online Community (<http://scratch.mit.edu>), a website that allows users to upload and share their creations. The visualizations created in this thesis describe community, personal, and network statistics. ScratchStats aims to provide answers to questions about Scratch usage, promote reflection and introspective learning, and aid in teaching data literacy.

Thesis Supervisor: Mitchel Resnick

Title: LEGO Papert Professor of Learning Research, MIT Media Lab

Acknowledgments

First, I would like to thank Mitchel Resnick for being a fantastic advisor and always making sure I had abundant support. He constantly ensured that I was excited to be working on my project and that it was personally meaningful. He has been a great source of inspiration and guidance.

I would also like to thank Andrés Monroy-Hernández for being my mentor this year. He has been there every step of the way to help me troubleshoot, brainstorm, and develop the work that has gone into this thesis. He has taught me as much about systems as I have learned in many of my classes. His patience, knowledge, and support were key ingredients for the success of this project.

Special thanks also go to John Maloney and Evelyn Eastmond for being my mentors when I first started working with the group as a UROP last year. They introduced me to this incredible project and have continuously provided me with encouragement and support.

I would also like to thank the other graduate students in the group: Amon Millner, Eric Rosenbaum, Karen Brennan, and Jay Silver for their motivation, innovative ideas, and advice.

Lastly, I would like to thank the other members of the Lifelong Kindergarten group for always being available to chat and assist in any way possible. Special thanks go to Amos Blanton, Chris Garrity, Stephanie Gayle, Natalie Rusk, and Claudia Urrea.

Contents

1	Introduction	13
1.1	The Inspiration	13
1.2	Children and Data Literacy	14
1.3	Children and Learning	15
1.4	Thesis Overview	16
2	Current Statistics on the Scratch Website	17
2.1	Community Statistics	17
2.2	Personal Statistics	19
2.3	Evaluation of Current Approach	20
3	Introducing ScratchStats	23
3.1	Goals	23
3.2	Community Statistics	24
3.2.1	Who We Are	24
3.2.2	What We Do	30
3.2.3	Who Does What	33
3.3	Personal Statistics	34
3.3.1	Scratch Blocks Usage	35
3.3.2	Monthly Activity	37
3.4	Network Statistics	37
3.4.1	Remix Tree	37
3.5	Summary	39

4	Implementation	41
4.1	Data Collection	41
4.2	Visualization Development	43
4.2.1	Data Generation	43
4.2.2	Interactive Visualization	45
4.3	Website Design and Development	45
5	User Evaluation	49
5.1	Community Statistics	49
5.1.1	Reactions From The Community	49
5.1.2	Patterns Observed By Researchers	50
5.2	Personal Statistics	52
5.2.1	Reactions From The Community	52
5.2.2	Patterns Observed By Researchers	53
5.3	Remix Visualization	54
5.3.1	Reactions From The Community	54
6	Conclusion	57
6.1	Future Works	57
6.1.1	Maintenance and New Features	58
6.1.2	New Visualizations	58
6.1.3	Better Integration	59
6.2	Final Words	59
A	Database Tables	61

List of Figures

2-1	The banner found at the top of the Scratch homepage discloses the number of projects that have been uploaded to the website to date.	17
2-2	In the right sidebar of the Scratch homepage, a section presents raw statistics about the community.	18
2-3	Google Trends provides this chart to show how popular the joint search terms ‘MIT Scratch’ are on any given month. It was posted in the Scratch forums by a user in April 2009.	18
2-4	This is an example of a project page. Statistics and information about the particular project can be found underneath the Scratch player as well as in the sidebar under ‘Download this project!’	19
2-5	This is a list of all remixes of a particular project and is accessible via a link on the original project’s page.	20
3-1	The current ScratchStats homepage. The homepage displays all the visualizations available for users to explore in three categories: ‘Who We Are’, ‘What We Do’, and ‘Who Does What’	25
3-2	The <i>Daily Visitors</i> visualization describes how many people visit the Scratch website each day. The visualization is provided by Quantcast Analytics.	26
3-3	The <i>Location of Visitors</i> visualization displays the cities that visitors to the Scratch website came during the current month. The visualization is provided by Google Analytics.	27

3-4	This visualization displays a summary of the monthly activity on the Scratch website. It is the visualization displayed for both the <i>New Users Per Month</i> and the <i>Activity Per Month</i> pages on the ScratchStats website.	27
3-5	This visualization displays how registered Scratch users and their activity on the Scratch website breaks down by age. It is the visualization displayed for both the <i>Age of Users</i> and the <i>Scratch Activity by Age</i> pages on the ScratchStats website. This bar chart plots age against several variables including registered users, projects shared, comments posted, users that have posted comments, and users that have shared projects.	28
3-6	This figure displays the age visualization with the added variable of gender mapped onto the same chart. Users can see the gender breakdown and filter based on gender.	29
3-7	This figure shows the third tab of the age visualization. This tab allows users the freedom to customize the grouping of ages. Users can specify the range of ages to display as well as the how many ages should be grouped and displayed in each bar.	29
3-8	This figure displays the history slider that appears in many of the ScratchStats visualizations. Using this slider, users can see how the visualization looked during that particular month in history.	29
3-9	This visualization displays how frequently Scratch users use the different Scratch blocks. This data can be viewed for all time or for just the current month by using the toggle button at the upper right corner.	31
3-10	The bars in the <i>Scratch Blocks Usage</i> visualization are color coded based on the categories that appear in the Scratch application. Users can toggle the visibility of these categories by using the checkboxes provided.	31
3-11	The Project Creators visualization displays how many users have shared projects each month.	32

3-12	This visualization describes the project sharing behaviors of Scratch users. It aims to show how many projects each user creates and shares.	34
3-13	This visualization describes the comment posting behaviors of Scratch users. It aims to show how many comments each user posts.	35
3-14	This visualization displays how a particular user is interacting with the Scratch online community. The ‘What I Do’ tab shows how the user is contributing to the community – projects shared, remixes shared, galleries created, and comments posted. The ‘Community Feedback’ tab shows how the community is responding to the user’s work – how many love its received, downloads of projects, remixes of projects, and comments received.	36
3-15	This is a remix visualization of a project shared on the Scratch website. It shows how different projects are related through remixing.	38
4-1	This diagram displays how many of the community statistics visualizations are built and provided to the user. It displays the client and server architecture used to collect and store the data, build the visualization, and display the page to the user.	42
4-2	This diagram describes how the Remix Tree visualization is built and integrated into the Scratch user experience. It displays the client and server architecture used to collect the data, build the visualization, and display the page to the user.	43
4-3	The required JSON structure specified by the JavaScript InfoVis Toolkit	44
4-4	This is the entry point for personal user statistics and visualizations.	46
5-1	This is the <i>Personal Scratch Blocks Usage</i> visualization for a popular Scratch user that predominately creates animations and artwork. The average count of blocks per project is shown.	53
5-2	This is the <i>Personal Scratch Blocks Usage</i> visualization for a popular Scratch user that predominately creates games. The average count of blocks per project is shown.	54

Chapter 1

Introduction

1.1 The Inspiration

Scratch, created by the Lifelong Kindergarten Group at the MIT Media Lab, is a graphical programming language designed for novice programmers [1]. Users create interactive content by snapping together puzzle shaped blocks representing programming functions. After creating these projects, they can share them on the Scratch website (<http://scratch.mit.edu>) [2][3], which serves as a medium for self-expression and inspiration. The Scratch website allows users to present their projects, browse projects shared by others, remix other projects, and provide and receive feedback [4].

There are currently over 505,000 registered members of all ages on the Scratch website and over 1,000,000 projects uploaded. Researchers often create charts about the activity on the site when questions arise, but these charts usually represent only a snapshot of the current activity and tend to only be useful in answering the specific question. However, visualizing the data graphically often inspires many followup questions that can only be answered by looking at even more data and creating additional charts.

With Scratch 2.0 development imminent, the Scratch Team is asking even more questions in order to make informed decisions. For instance, we are interested in knowing which Scratch blocks are used most frequently, which instruments are least used, and which costumes are the most popular.

We started building data visualizations as a response to these inquiries. The goal was to make the wealth of data that we have more accessible so that researchers can answer questions more readily and be able to find solutions to followup questions without querying a database.

After seeing how interesting and informative the data can be via visualizations, we decided that it would be useful for both researchers *and* Scratch users if these tools were provided in a centralized and publicly accessible location. Thus inspired the idea of ScratchStats - a website that contains data visualizations for users to explore statistics about Scratch.

This thesis discusses ScratchStats, which aims to present data in a meaningful way through interactive visualizations. Visualizations provided describe both the status of the community as a whole as well as that of individual Scratch users. They allow users to explore data in a variety of dimensions, to customize what is being displayed, and to compare historical data alongside current data. This freedom helps us satisfy the goals of creating the ScratchStats website - to provide answers to a multitude of questions, to promote reflection and introspective learning, and to aid in teaching data literacy.

1.2 Children and Data Literacy

In this information era there is an abundance of data and a movement toward data-driven decision making. Statistical analysis using large data sets has led to key advances in fields such as health care, finance, and public policy. It is quickly becoming evident that the ability to understand and utilize data is an essential 21st century skill [5].

With the emergence and popularity of visualization blogs, we notice that data not only aids in decision making, but can paint pictures and tell stories. In these ways, interacting with data can become personally meaningful.

Software can help present data in a more easily digestible format by providing the tools to create data visualizations and thereby make statistical thinking more

accessible [6]. Well-constructed visualizations allow children to discover correlations and relationships among data sets, as well as to make well-formed statistical claims and data-driven decisions. Additionally, visualizations can also promote a hands-on approach to building data literacy, especially when the data is about their own creations. Children can explore data by experimenting with these visualizations and can discover personally relevant conclusions based on the evidence provided. These conclusions can provide insights for reflection as part of the naturally iterative learning process.

1.3 Children and Learning

Promoting data literacy in this way is well-aligned with the philosophies of the Lifelong Kindergarten group, led by Mitchel Resnick. Our group focuses on developing technologies with the goal of helping people develop as creative thinkers [7]. We are inspired by Seymour Papert’s theory of Constructionism, which suggests that people learn best when actively designing, constructing, and sharing personally meaningful artifacts [8].

The group uses these Constructionist beliefs to support a kindergarten approach to learning, which Resnick describes as “a spiraling learning process in which [people] imagine what they want to do, create a project based on their ideas, share their ideas and creations with others, and reflect on their experiences” [9]. This approach values the importance of *tinkering*, the process of learning where one discovers ideas naturally by engaging in hands-on activities.

Scratch and the Scratch website are products of these ideals. Scratch was developed with inspirations by Papert’s LOGO [10] and Alan Kay’s Etoys [11] and similarly aims to teach children mathematical, computational, and problem solving skills by experimenting and creating.

1.4 Thesis Overview

This thesis presents ScratchStats, an extension to the Scratch website where users can view and understand Scratch usage data through a series of interactive visualizations. Chapter 1 of this document presents some motivations behind building these data visualizations and providing them to the public. Chapter 2 explores the statistics that are currently available for the Scratch community to explore. Chapters 3 and 4 discuss the design and implementation details of the visualizations and the website that comprises ScratchStats. Chapter 5 discusses how these new additions have been received by the community. Finally, Chapter 6 presents the conclusions and recommendation for future efforts.

Chapter 2

Current Statistics on the Scratch Website

This chapter presents the current statistics available to the public about Scratch usage. We look at two categories of statistics: statistics about the community as a whole and statistics about individual users and projects.

2.1 Community Statistics

Community statistics provide a high level view of the entire Scratch online community. These general statistics can be found directly on the homepage of the Scratch website.

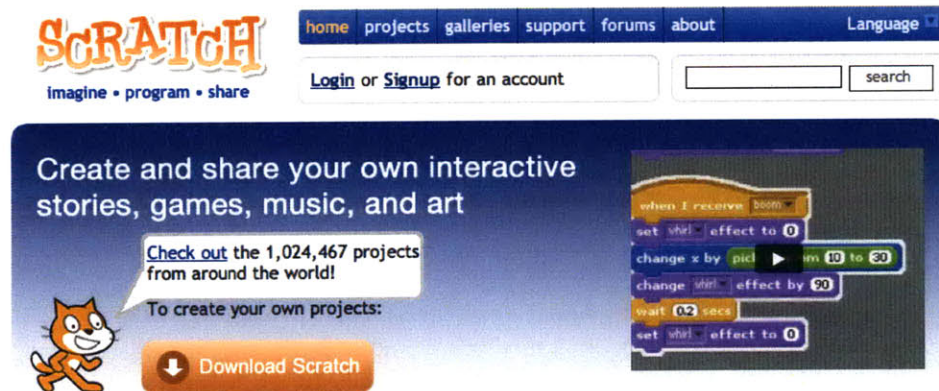


Figure 2-1: The banner found at the top of the Scratch homepage discloses the number of projects that have been uploaded to the website to date.

Community Stats
503,474 registered members,
125,447 project creators,
1,024,467 projects uploaded,
19,085,240 scripts
and 6,196,255 sprites.
That's a lot of Scratch-ing!

Figure 2-2: In the right sidebar of the Scratch homepage, a section presents raw statistics about the community.

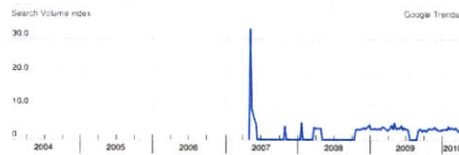


Figure 2-3: Google Trends provides this chart to show how popular the joint search terms 'MIT Scratch' are on any given month. It was posted in the Scratch forums by a user in April 2009.

Displayed prominently in a banner across the top of the page is the number of projects that have been uploaded to the website to date (see Figure 2-1). Within the sidebar the current number of registered members, project creators, projects uploaded, scripts within all projects and sprites within all projects are disclosed (see Figure 2-2).

In addition to the numbers on the front page, statistics can also be found within the Scratch forums. These generally are not provided by the Scratch Team. Users often post about statistics that they themselves have calculated or have found by other means. For example, in April 2009 a user posted a visualization provided by Google Trends¹, displayed in Figure 2-3, that illustrates how many people search on Google for the joint terms 'MIT' and 'Scratch'.

Occasionally, Scratchers also mine and compute data themselves. A user in January 2010 created a survey for other users to anonymously provide data about themselves and their usage patterns. He then created a Scratch project to visualize the data that he collected and continuously updated the project as new data emerged. Although his sample size was only a small fraction of the community, his results were quite similar to what we actually see throughout the community.

¹Google Trends, <http://www.google.com/trends?q=MIT+Scratch>

The screenshot shows the Scratch website interface. At the top, there's a navigation bar with links for 'home', 'projects', 'galleries', 'support', 'forums', and 'about'. Below this is a search bar and a 'Login or Signup for an account' link. The main content area displays a project titled '1 PacMan'. The project player shows a Pac-Man character in a maze. Below the player, the user 'SampleProjectsTeam' is credited, along with statistics: 2699 views, 2 taggers, 16 people love it, 1055 remixes by 803 people, and 594 downloads in 7 galleries. The sidebar on the right contains a 'Download this project!' button and 'Project Notes' which provide instructions on how to play the game and how to modify the maze.

Figure 2-4: This is an example of a project page. Statistics and information about the particular project can be found underneath the Scratch player as well as in the sidebar under ‘Download this project!’

2.2 Personal Statistics

Personal statistics refers to data specific to a single user: usage patterns, contributions to the community, and the feedback received from the community. On the Scratch website, each user has a ‘My Stuff’ page that collects together all of the user’s previously shared projects, galleries created, and friends. These items are publicly visible to all visitors of the Scratch website. Users can also visit their ‘My Stuff’ page to see how many total comments they have posted and to view all of these comments in a single location. These items are kept private from other Scratch users.

Each Scratch project has its own page where the project and information about the project is showcased. An example of a project page is shown in Figure 2-4. Statistics are provided on these pages about the views, downloads, tags, and love-its²

²Love-its refers to the ability for any user to express interest and excitement about a project by

Total Remixes

[yoshidude56's project](#)
[yoshi555's project](#)
[baseball09's project](#)
[bowserblade555's project](#)
[alexzades's project](#)
[TheLegoTroll's project](#)
[da123's project](#)
[reis27's project](#)
[madman5's project](#)
[Sun12855's project](#)

Figure 2-5: This is a list of all remixes of a particular project and is accessible via a link on the original project's page.

the project has received, as well as how many times the project has been remixed. Several of these statistics contain links where more information can be explored. For instance, as shown in Figure 2-5, a list of remixes is displayed when the associated link is clicked. The number of sprites and scripts the project contains is also displayed on the project page.

2.3 Evaluation of Current Approach

The data that is publicly disclosed has been received well by Scratch users as evident in forum discussions. These discussions also suggest that users are interested in seeing even more statistics about the community. After the release of the statistics on the front page, a user calculated averages using these numbers and shared these results in the forums, which sparked further discussions about the statistics. In September 2007, another Scratcher wondered about the gender distribution of Scratch users.

The current statistics provided are only snapshots in time. Because we do not disclose historical data, users cannot easily see trends that emerge over time. Additionally, all statistics provided are raw numbers, which are often difficult to interpret. Instead, a more visual approach could be more comprehensible, especially to a younger audience.

Users are continuously interacting in the Scratch community allowing natural networks to emerge such as those of friendships and collaborations. Users are connected as a web of interlinked friends. They create remixes of each other's projects, which selecting a "Love it?" button. This can be used as a measure of popularity.

can then be further remixed themselves. We currently show some of these relationships described above. For example, for each project we display the list of remixes of the project (see Figure 2-5). However, we only show one level of this natural tree structure. Therefore, it is not easy to determine how other projects may be related. We could provide tools to better explore these connections.

Our current approach allows users to see the current state of the community. However, it does not provided ways for users to achieve a deeper understanding of how the community evolves over time or how users are interacting with each other.

Chapter 3

Introducing ScratchStats

ScratchStats is a companion website to the current Scratch website where people can explore usage data through interactive visualizations. The interactive nature allows users to tinker with and manipulate the data in order to answer various questions that they pose. This chapter discusses the goals of creating the ScratchStats website and introduces what has been released at this point. The visualizations released fall into three categories: community statistics, personal statistics, and network statistics.

3.1 Goals

The ScratchStats website was created with a few goals in mind:

- To provide answers to questions about Scratch usage and inspire new ones
- To promote reflection and introspective learning
- To aid in teaching data literacy

The main goal for this project is to provide data in an easily comprehensible manner such that both researchers and Scratch users alike can better understand how the Scratch application and the online community are being utilized. These visualizations are meant to assist in answering questions about Scratch usage and to

inspire new questions. For the Scratch Team, these answers can help define how to shape the next version of Scratch.

A second goal is to help users understand their own activities with Scratch in order to learn and grow. The intuition is that if they can see and analyze what they are doing and what they are not doing, then they will be able to better reflect about their activities and modify their current behaviors to produce desired results.

Lastly, a goal is also to promote data literacy. We want to allow even young users to discover correlations and relationships among data sets, and to further use these discoveries to fuel statistical claims and data-driven decision making.

3.2 Community Statistics

The main page of the current Scratch website contains some general statistics about the Scratch community (see Figure 2-2). A ‘Learn more’ link has been added under these statistics to refer users to the new ScratchStats website, where they can explore these statistics in greater depth. A screenshot of the ScratchStats homepage is shown in Figure 3-1.

The homepage classifies all the available visualizations into three categories: ‘Who We Are’, ‘What We Do’, and ‘Who Does What’. These categories were selected to help young users navigate between the different types of data displayed. Each of these categories is discussed in detail below.

3.2.1 Who We Are

Visualizations in the ‘Who We Are’ category describe the demographics of Scratch users. There are currently 4 visualizations in this category.

Daily Visitors

The *Daily Visitors* visualization is a stacked area chart that displays how many people visit the Scratch website each day. It is provided by Quantcast Analytics¹

¹Quantcast Analytics, <http://www.quantcast.com/wd:edu.mit.scratch>

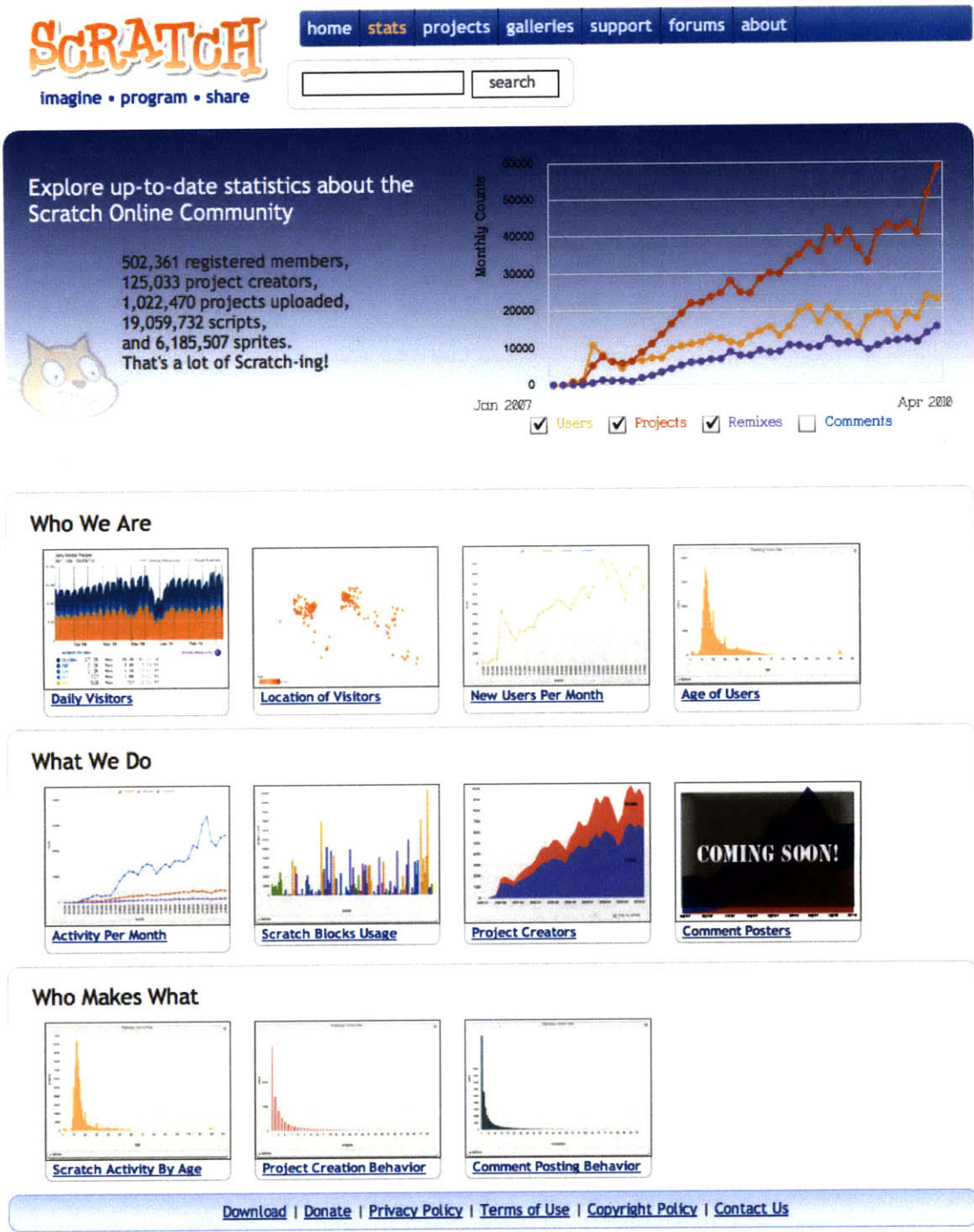


Figure 3-1: The current ScratchStats homepage. The homepage displays all the visualizations available for users to explore in three categories: ‘Who We Are’, ‘What We Do’, and ‘Who Does What’

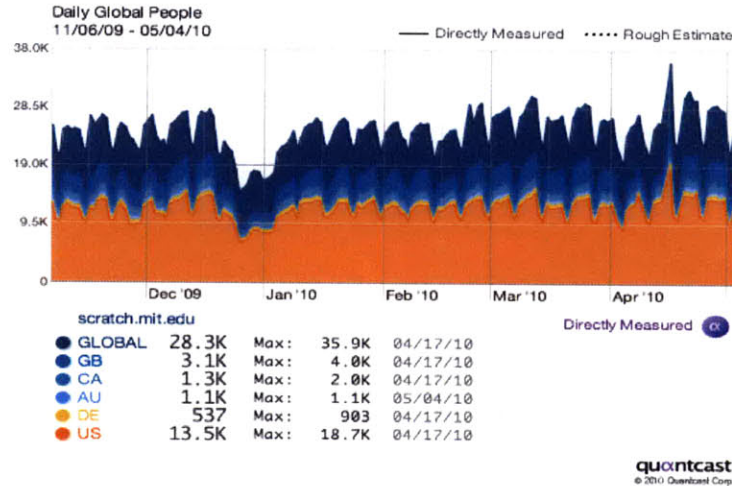


Figure 3-2: The *Daily Visitors* visualization describes how many people visit the Scratch website each day. The visualization is provided by Quantcast Analytics.

and also shows how the number of visits breaks down by country. The visualization links to the Quantcast page for the Scratch website, where users can continue to explore information about traffic, geography and demographics. A snapshot of the visualization is displayed in Figure 3-2.

Location of Visitors

The *Location of Visitors* visualization is a world map that displays the cities that visitors to the Scratch website came from during the current month. This is illustrated by circles that overlay the map and that are shaded differently to represent the frequency of visits. The visualization is provided by Google Analytics² and is displayed in Figure 3-3.

New Users Per Month

This visualization is a line chart that displays how many new users register, how many projects are shared, how many remixes are created, and how many comments are posted to the Scratch website each month since the release of the website in 2007. Each series can be turned on or off using checkboxes above the chart. Only the new

²Google Analytics, <http://www.google.com/analytics/>



Figure 3-3: The *Location of Visitors* visualization displays the cities that visitors to the Scratch website came during the current month. The visualization is provided by Google Analytics.

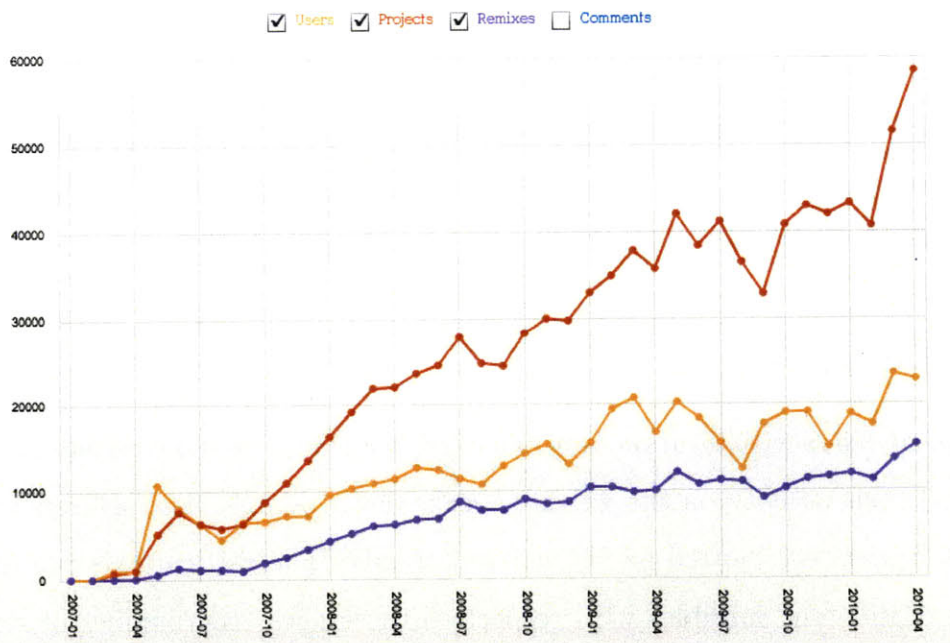


Figure 3-4: This visualization displays a summary of the monthly activity on the Scratch website. It is the visualization displayed for both the *New Users Per Month* and the *Activity Per Month* pages on the ScratchStats website.



Figure 3-5: This visualization displays how registered Scratch users and their activity on the Scratch website breaks down by age. It is the visualization displayed for both the *Age of Users* and the *Scratch Activity by Age* pages on the ScratchStats website. This bar chart plots age against several variables including registered users, projects shared, comments posted, users that have posted comments, and users that have shared projects.

users series is turned on by default since this visualization appears in the ‘Who We Are’ category. However, the other series are available as a means of comparison. The visualization is displayed in Figure 3-4.

Age of Users

The *Age of Users* visualization is a bar chart that displays how registered Scratch users and their activity on the website breaks down by age. The self-reported ages of Scratch users are plotted on the horizontal axis. The vertical axis can display a number of different variables with respect to these ages: the number of registered users, the number of projects shared, the number of comments posted, the number of users that have posted comments, and the number of users that have shared projects. Users can animate between these different variables with the click of a button. The visualization is displayed in Figure 3-5.

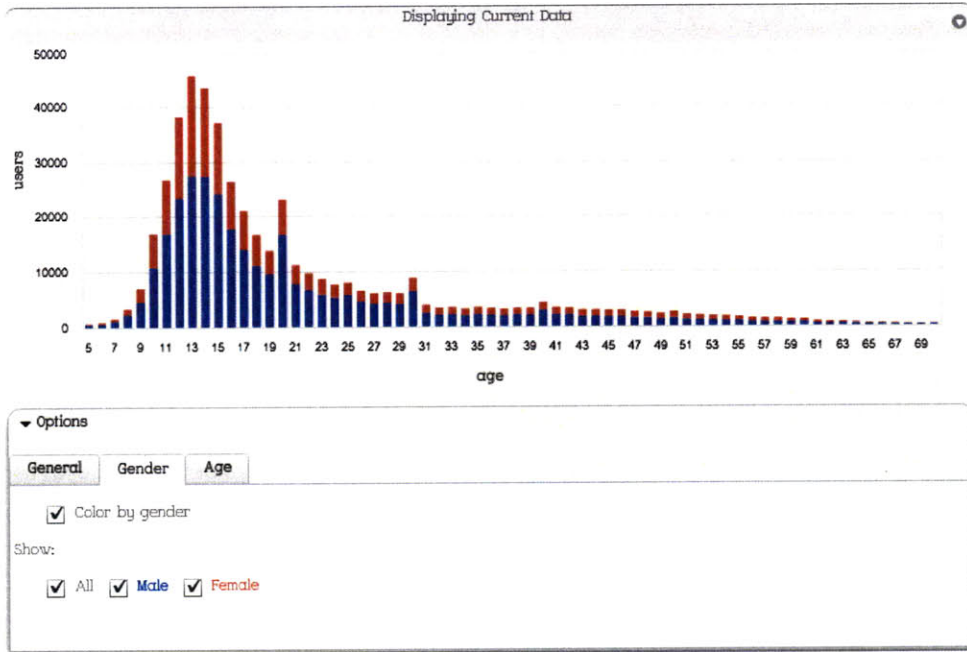


Figure 3-6: This figure displays the age visualization with the added variable of gender mapped onto the same chart. Users can see the gender breakdown and filter based on gender.

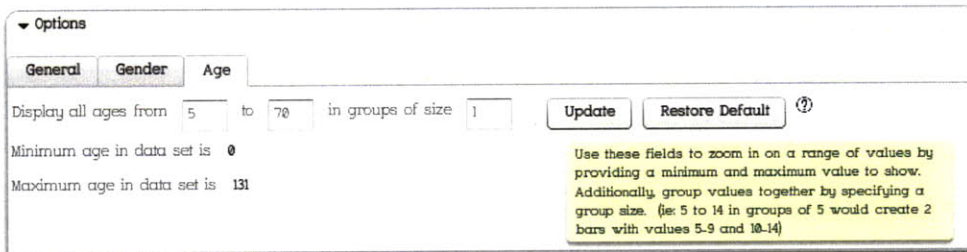


Figure 3-7: This figure shows the third tab of the age visualization. This tab allows users the freedom to customize the grouping of ages. Users can specify the range of ages to display as well as the how many ages should be grouped and displayed in each bar.



Figure 3-8: This figure displays the history slider that appears in many of the Scratch-Stats visualizations. Using this slider, users can see how the visualization looked during that particular month in history.

A third variable can also be viewed on the same chart through the use of color. The motivation behind having all these variables collected in a single visualization is so users have the freedom to manipulate the data in order to look for trends and analyze how different variables, or collection of variables, influence the data. We use color to display the gender distribution. Users can turn on and off this coloring option as well as filter the data based on gender directly from the visualization. These options are shown in Figure 3-6.

In order to provide even more freedom to the user, we also allow customization of how the independent variable (age) is grouped and displayed (see Figure 3-7). Users can specify the range of ages to display and how to group this range. For example, we can zoom in on the data by looking only at ages 8 to 17 in groups of 2. This would display 5 bars each representing 2 years of age.

We also introduce time as another dimension. The data is no longer just a snapshot in time. Users can access historical data from this visualization, as shown in Figure 3-8, and explore how the data has changed over time.

3.2.2 What We Do

Visualizations in the ‘What We Do’ category describe how the Scratch application and the Scratch website are being utilized. There are 3 visualization in this category.

Activity Per Month

The *Activity Per Month* visualization is the same visualization that is displayed on the *New Users Per Month* page (see Figure 3-4), but with a different set of series turned on by default. We have two separate pages for these visualizations because they fit into both the ‘Who We Are’ and the ‘What We Do’ categories. We also allow users the option of seeing all the other series so that they have the means to compare and find correlations between different variables.



Figure 3-9: This visualization displays how frequently Scratch users use the different Scratch blocks. This data can be viewed for all time or for just the current month by using the toggle button at the upper right corner.

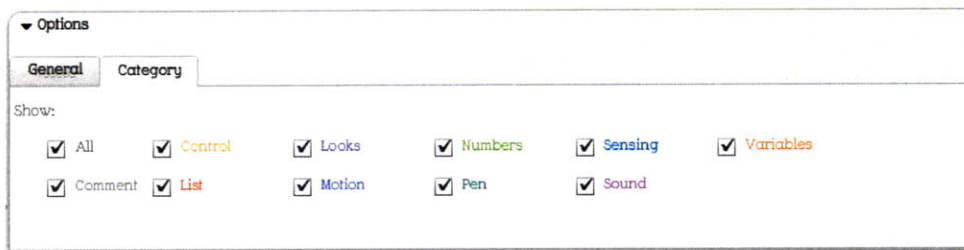


Figure 3-10: The bars in the *Scratch Blocks Usage* visualization are color coded based on the categories that appear in the Scratch application. Users can toggle the visibility of these categories by using the checkboxes provided.

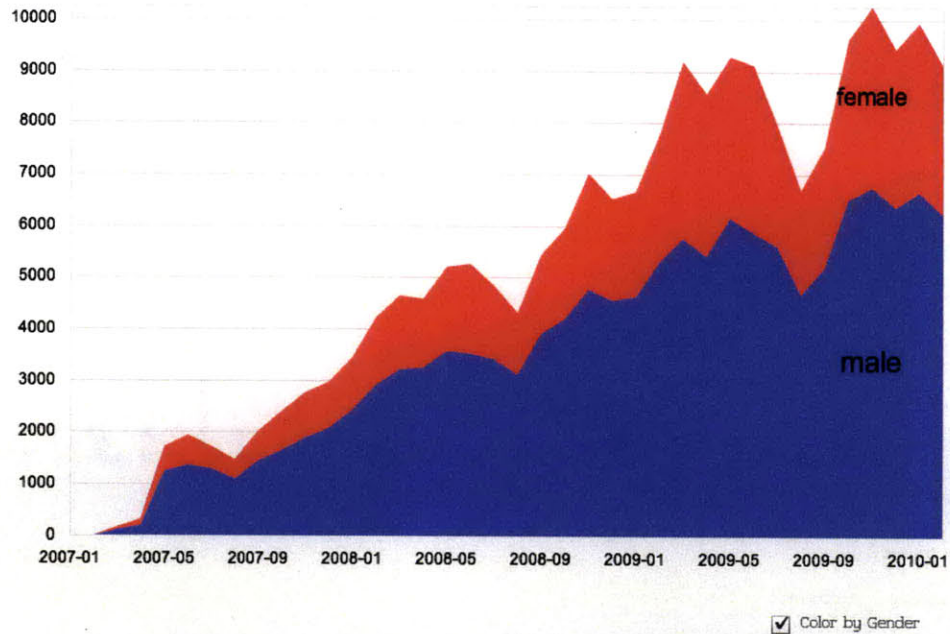


Figure 3-11: The Project Creators visualization displays how many users have shared projects each month.

Scratch Blocks Usage

The *Scratch Blocks Usage* visualization is a bar chart that displays the frequency of use for each of the different Scratch blocks. The visualization can toggle between displaying data for all time or just for the current month. The frequency can also be displayed a number of ways: the *project count* is a count of all the projects that use each block and the *total count* is a count of the total times each block is used within all projects.

The bars are color coded using the same categories that appear in the Scratch application and users can show or hide any of these categories to focus on particular blocks (see Figure 3-10). Users can further highlight particular blocks by searching for blocks by name.

Project Creators

This visualization is an area chart that displays how many Scratchers have created projects each month. The area chart can be colored by gender. Furthermore users

can filter by gender by clicking on the different stacked areas. The visualization is displayed in Figure 3-11.

This chart is not normalized. Although there are generally more male project creators than female project creators each month, this can be due to the fact that there is a higher percentage of registered male users to female users.

3.2.3 Who Does What

Visualizations in the ‘Who Does What’ category show how users are participating in the community. This generally shows combined information about users and their activities. There are 3 visualizations in this category.

Scratch Activity By Age

The *Scratch Activity By Age* visualization is the same visualization that appears on the *Users By Age* page (see Figure 3-5), but shows number of projects shared vs age by default (as opposed to the number of users vs age). We have two separate pages for these visualizations because they fit into both the ‘Who We Are’ and the ‘Who Does What’ categories depending on what data is displayed. For a more detailed description about this visualization, see the *Users By Age* section above.

Project Creation Behavior

This visualization is a histogram describing project creation behaviors. It helps illustrate how many projects each user shares on the Scratch website. On the horizontal axis is the number of projects and on the vertical axis is the number of users that have shared that many projects (see Figure 3-12).

A third variable of gender can also be displayed in this visualization with the use of color much in the same way as the age visualization described above (see Figure 3-6). Users can also range limit and customize the grouping of the independent variable (see Figure 3-7), as well as view historical data (see Figure 3-8).

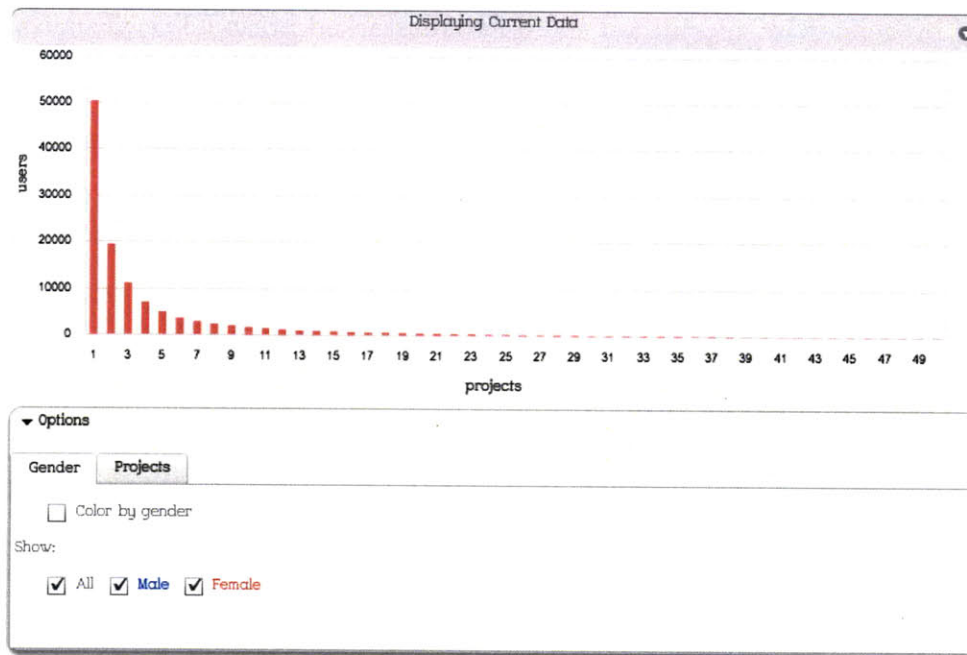


Figure 3-12: This visualization describes the project sharing behaviors of Scratch users. It aims to show how many projects each user creates and shares.

Comment Posting Behavior

This visualization is a histogram describing comment posting behaviors. It shows how many comments each user posts on project pages and gallery pages. On the horizontal axis is the number of comments and on the vertical axis is the number of users that have posted that many comments (see Figure 3-13).

Like the *Project Creation Behavior* visualization described above, users can also view the gender breakdown of the data, customize the range and grouping of the independent variable, as well as see historical data.

3.3 Personal Statistics

Providing personal statistics to Scratch users allows them to see how other members of the community are responding to their projects and to see their own creation patterns. We have created a couple of visualizations to allow users to do just that, in hopes that they can better understand, reflect about, and learn from their previous behaviors.

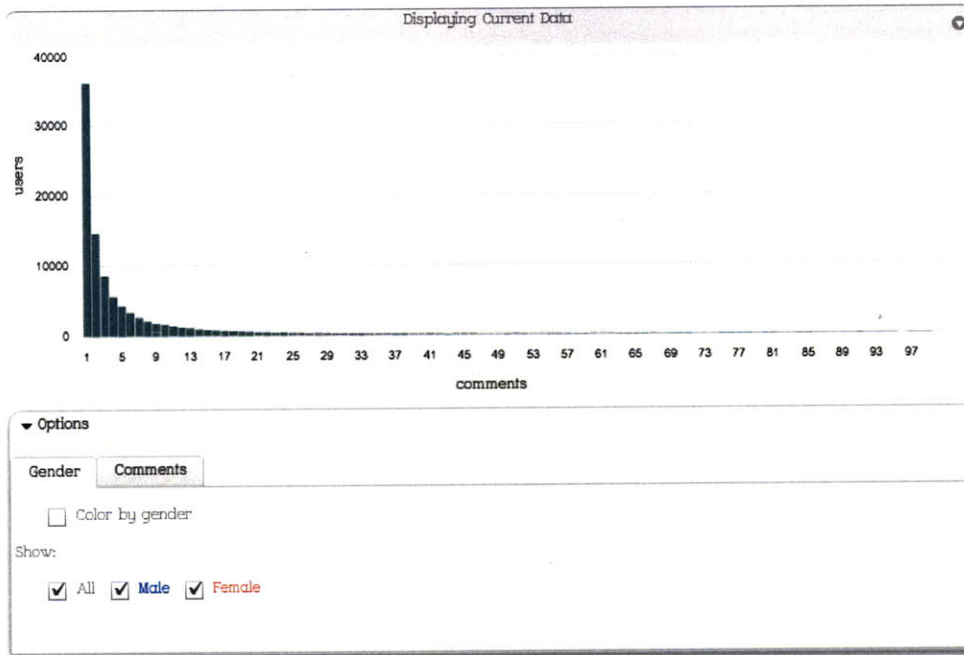


Figure 3-13: This visualization describes the comment posting behaviors of Scratch users. It aims to show how many comments each user posts.

3.3.1 Scratch Blocks Usage

The first personalized visualization created is a version of the community *Scratch Blocks Usage* visualization (see Figure 3-9) that only contains data about a single user. It displays how frequently a particular user uses each of the different Scratch blocks within shared projects. Like the community visualization, this frequency can be displayed as a *project count* or a *total count*. A new way of displaying this frequency, an *average count*, is also available. This shows the average number of times that the user used each particular block in each project. This view is particularly useful for personal statistics because it allows the user to easily observe which blocks or types of blocks they tend to prefer.

A motivation for providing this visualization to users is so they can see and reflect about how they are using the language. Perhaps they can identify blocks that they have rarely used and begin experimenting with those.

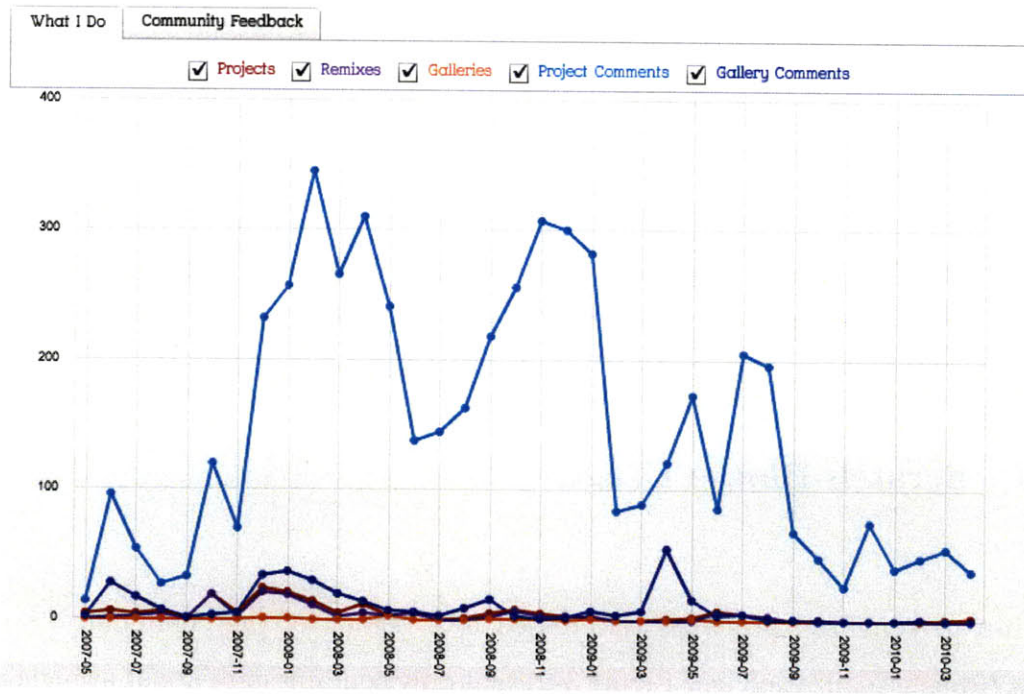


Figure 3-14: This visualization displays how a particular user is interacting with the Scratch online community. The ‘What I Do’ tab shows how the user is contributing to the community – projects shared, remixes shared, galleries created, and comments posted. The ‘Community Feedback’ tab shows how the community is responding to the user’s work – how many love its received, downloads of projects, remixes of projects, and comments received.

3.3.2 Monthly Activity

The personalized Monthly Activity visualization is a line chart much like the *Activity By Month* chart for community statistics. This personal visualization is shown in Figure 3-14. There are two sets of trend lines that the user can display on the chart. The first set is under the ‘What I Do’ tab and shows how the user is contributing to the community, including the number of projects shared, remixes shared, galleries created and comments posted. The second set is under the ‘Community Feedback’ tab and shows how the community is responding to the user’s work, including how many love-its received, downloads of projects, remixes of projects and comments received.

A motivation for providing such a visualization is for the user to explore how he or she is contributing to the community and how the community is receiving his or her work. The reason for providing the option of viewing multiple series simultaneously is so the user can discover correlations among different data sets and make statistical claims. For example, a user may discover the law of reciprocity in the community; to receive more comments, the user should provide more comments.

3.4 Network Statistics

As an online community, Scratch has a large networking component. Users become friends with other users, comment on each other’s projects, and even remix each other’s projects. This category of visualizations aims to provide users with a way of seeing and exploring these interconnections. We have added one such visualization to the Scratch website.

3.4.1 Remix Tree

Remixing is a way users are collaborating on the website and thus a way they are connected to each other. A remix refers to a project that is downloaded, modified and shared as a new project. Remixes can further be remixed themselves. Since remixes have a natural tree structure, we have created a visualization that displays remixes in

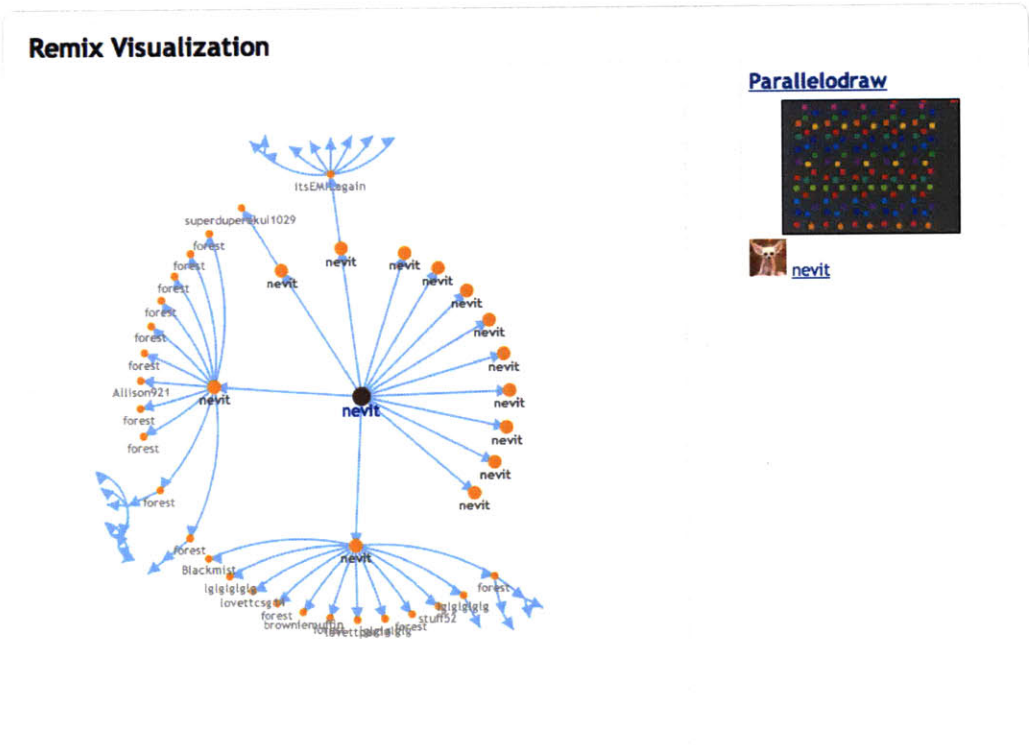


Figure 3-15: This is a remix visualization of a project shared on the Scratch website. It shows how different projects are related through remixing.

the form of a directed hyperbolic tree. An example is displayed in Figure 3-15. This visualization aims to show the relationships between projects that share a common ancestry.

The visualization is fully interactive to promote exploration. Clicking on a node in the tree will center that node on the page, effectively zooming in on that portion of the tree. The right side panel also updates with the project information that the centered node represents, including the project name, the author's username, and the project that this one was based on. The tree can also be navigated using the arrows keys since there are essentially four possible directions to move in a tree of this structure. The up arrow highlights a child of the currently selected node. The left and right arrows highlight different siblings of the selected node. Finally, the down arrow highlights the parent of the currently selected node.

Remix trees can show the evolution of a project and contributors at different levels of development. They can also assist in helping users create popular chain remixes

by providing an easy way to see which project was the most recently added to the chain.

3.5 Summary

We have described the various visualizations that were created and released to the Scratch community with the goals to provide answers to questions about Scratch usage, to promote reflection and introspective learning, and to aid in teaching data literacy. These fall into three categories: community statistics, personal statistics, and network statistics. For each visualization we have discussed what meaningful data is presented and how a user can interact with the visualization.

Chapter 4

Implementation

This chapter discusses the implementation efforts of building the ScratchStats website. Implementing ScratchStats was a three step process involving data collection, visualization development, and website design and development.

4.1 Data Collection

The Scratch website handles a wealth of data. In order to share a project via the website, a user must create an account and provide some general information about him or herself, such as age, gender, and location [3]. Therefore, user information can be found for every shared project. Once shared on the website, projects begin to generate project specific data, such as the number of times viewed, loved, and downloaded. However, the scripts and media elements that encompass a project are never parsed and stored in easily accessible forms.

To begin this project we first developed a Java application to collect the internal project data that we were lacking. Scratch projects are saved in a Scratch specific file format called sb files. The Scratch website plays these files using a Java player. We utilized the Java sb parsing code used by the player to collect the relevant project data. The parsing code creates a table of objects containing all elements of the Scratch project. There are three kinds of records that are contained in this objects table: fixed-format class records, object reference, and user-class records. The general

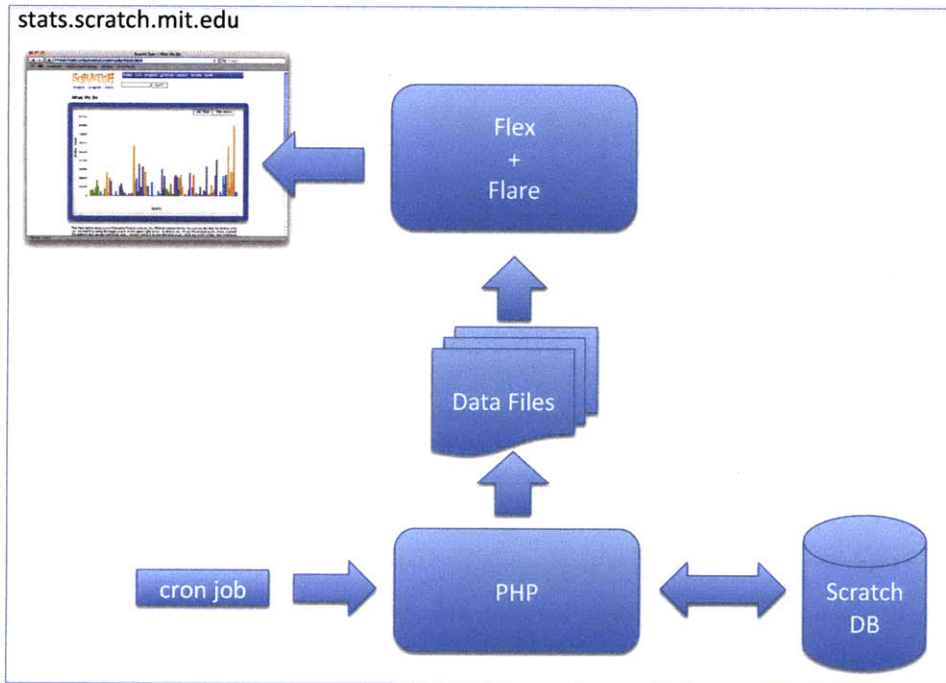


Figure 4-1: This diagram displays how many of the community statistics visualizations are built and provided to the user. It displays the client and server architecture used to collect and store the data, build the visualization, and display the page to the user.

process of reading this objects table is to first read all the objects, next build Java image and sound objects, and finally to dereference object references in the fields list of user-classes and in fixed-format collection objects. In addition to this objects table, the parsing code also creates an info table consisting of the file's metadata, such as the save and share history, author, comments, scratch-version, and language used.

The Java application extracts relevant information from the objects table and the info table and stores it in a MySQL database to be explored at a later time. There are 11 tables in this database containing information about project metadata, media usage, blocks usage, and user generated strings (see Appendix A for a description of these tables). The application was added to the upload script so that these data elements could be extracted as each project is shared on the Scratch website. We also ran a bulk analysis over all the projects that were uploaded before the script was added to the upload process in order to build a complete database.

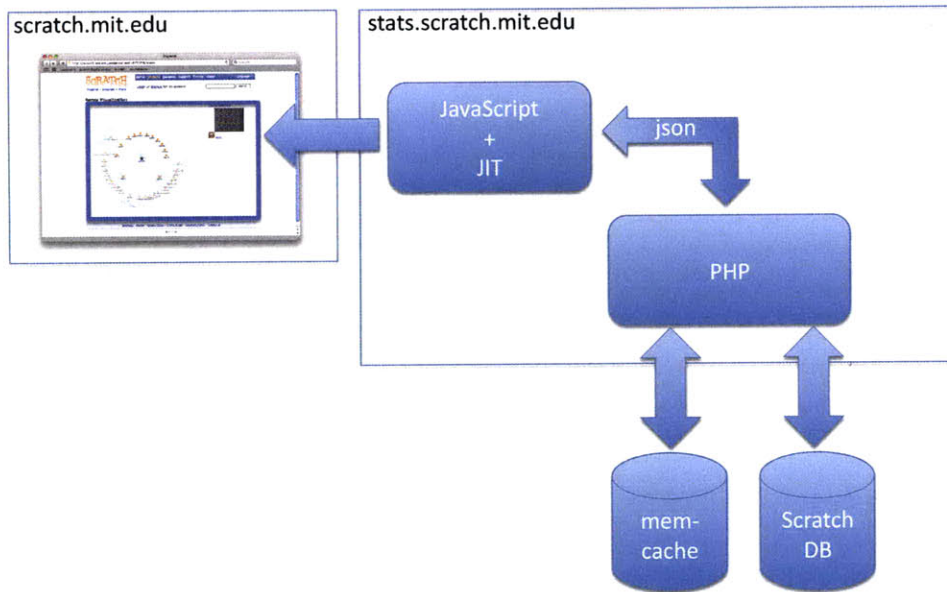


Figure 4-2: This diagram describes how the Remix Tree visualization is built and integrated into the Scratch user experience. It displays the client and server architecture used to collect the data, build the visualization, and display the page to the user.

4.2 Visualization Development

Once we had the data that we were interested in, we started building visualizations to explore this data. To develop each visualization, we needed to write a script to collect the data for the visualization and to build the interactive visualization that graphically displays this data. Refer to Figure 4-1 and Figure 4-2 for the architecture diagrams of these systems.

4.2.1 Data Generation

We first wrote php scripts that queried the database in order to obtain the necessary data and then stored that data in tab delimited text files. Text files are generated to optimize the loading time of the visualizations and to reduce the load on the databases.

Since we can expect the community data to look fairly similar between short periods of time, scripts for community statistics run on monthly cron jobs. Therefore,

```

var json = {
  "id": "aUniqueIdentifier",
  "name": "usually a nodes name",
  "data": {
    "some key": "some value",
    "some other key": "some other value"
  },
  "children": [ 'other nodes or empty' ]
};

```

Figure 4-3: The required JSON structure specified by the JavaScript InfoVis Toolkit

the community data is updated once a month.

Data for personal statistics is not updated at fixed intervals, but is instead updated on demand by the user at most once a week. When a user visits his or her visualization page and the data is more than a week old, the data will automatically refresh. However, if it is less than a week old, the data will not be refreshed.

Remix tree network visualizations work differently (see Figure 4-2). The server script does not produce text files, but instead returns a JSON representation of the remix tree. The particular JSON format follows the structure specified by the JavaScript InfoVis Toolkit¹ and is shown in Figure 4-3. Although the script does not store text files, it caches the JSON tree using memcache to improve performance and reduce database load. Therefore, remix tree data is always up to date with exception of a short caching period.

One complication we ran into was that the remix visualization causes the browser to timeout when displaying very large remix trees with thousands of children per generation. A modification was therefore made to the server script to limit each generation to a set maximum number of nodes. This approach allowed the visualization accessed by the user to be more responsive, but the tree displayed is now incomplete. The nodes are selected based on the number of remixes they have; we prefer the projects with higher numbers of remixes to those with lower ones. This is so we do not lose the depth of the tree.

¹The JavaScript InfoVis Toolkit, <http://thejit.org>

4.2.2 Interactive Visualization

The visualizations for the community and personal statistics are built in Adobe Flex with the help of the Flare², an ActionScript library for creating visualizations developed by the UC Berkeley Visualization Lab. Building these as Flash applications allow them to be easily embedded into webpages and viewed on Flash enabled browsers. These visualizations retrieve the tab delimited text files generated with the server scripts and then build the interactive charts.

The remix tree visualization is built in JavaScript with the help of the JavaScript InfoVis Toolkit, developed by Nicolas Garcia Belmonte. The visualization is a modified version of the hyperbolic tree provided by the toolkit. The original toolkit was not equipped with directed edges for hypertrees so we added this functionality in order to show the directionality of remixes. We use the URL to determine which project's tree is being requested and send a request to the server for that project's JSON tree representation. Once the JSON data structure is returned, the visualization draws the hypertree based on the information specified.

4.3 Website Design and Development

After building the collection of visualizations, we designed and developed a website to display them. Although located on a separate subdomain, we designed the ScratchStats website to adopt the look and feel of the original Scratch website so that it would be friendly to young users and older users alike. The visualizations are separated into categories on the ScratchStats homepage, much like the categories on the Scratch homepage.

An important element we were sure to include for visualization pages is a feedback mechanism so that we can collect and analyze user feedback to help improve our design. For the community statistics pages, we wanted to not only collect feedback, but to promote discussion about the data. Therefore, we opted to use an external

²Flare, <http://flare.prefuse.org>

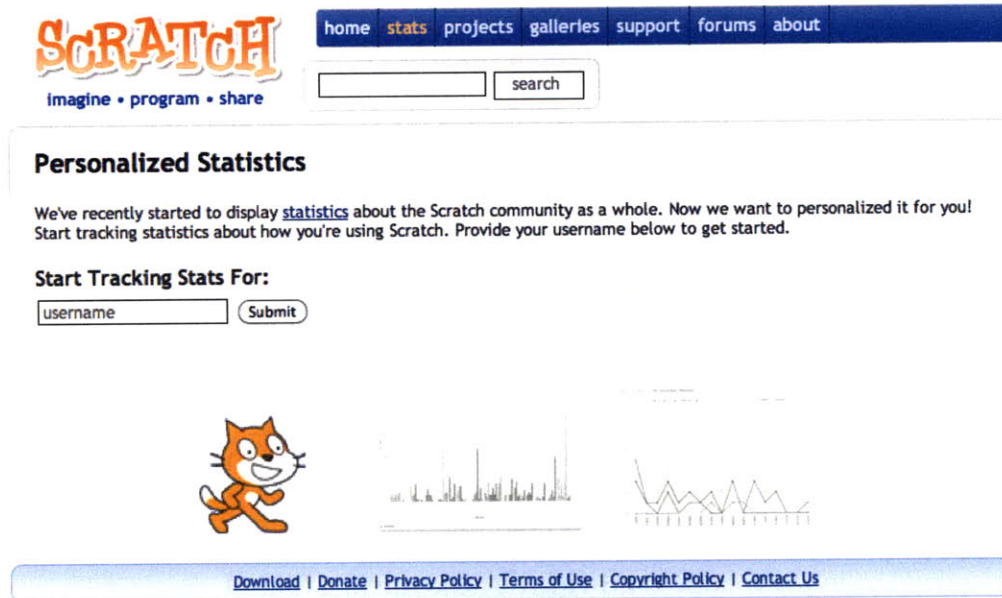


Figure 4-4: This is the entry point for personal user statistics and visualizations.

service, Disqus³, to power these discussions. Disqus provides us with a JavaScript widget that we can install easily onto each page and also provides us with moderation tools. For the personal statistics pages, we elected to use Google Docs⁴ Forms in order to collect anonymous and private feedback.

The personal statistics are not accessible from the main ScratchStats page, therefore we needed an entry point for users to access these visualizations. In the future we would like this to be integrated with the ‘My Stuff’ page on the Scratch website. However, before further engineering effort, we wished to gauge how users would feel about having these statistics tracked. For this initial test run, we built a page where users can input their username and start tracking their activity (see Figure 4-4). When a username is submitted, a php script automatically runs generating the data files needed for this user’s visualizations.

The remix tree visualization page was designed to fit seamlessly into the Scratch website and replace the original Total Remixes page (see Figure 2-5). One of the main challenges in redesigning the Total Remixes page was the cross-domain scripting issue. The actual page is hosted on scratch.mit.edu, however the JavaScript that generates

³Disqus, <http://disqus.com/>

⁴Google Docs, <http://docs.google.com/>

the visualization is hosted on `stats.scratch.mit.edu`. This was preferred so that all statistics and visualization resources would be collected together. The issue was resolved by using JSON with padding (JSONP), which allows the client to provide a function call using a query argument to the url. The server then wraps the JSON response with this padding and thus the browser receives the response in the form of a script to be executed instead of raw data.

Chapter 5

User Evaluation

We introduced statistics and visualizations to the community as interventions to the Scratch website. We as researchers also analyzed the statistics and noticed some interesting data patterns. This chapter explains how Scratch users responded to the initial interventions and the data patterns that researchers have identified using these interventions.

5.1 Community Statistics

5.1.1 Reactions From The Community

We released the new ScratchStats website for community statistics quietly via the forums and a new link in the sidebar of the Scratch homepage. Each visualization page contains a commenting widget to receive feedback and elicit discussions about the statistics shown. Through these widgets and the forum thread we received numerous feedback. The types of feedback received were that of users identifying trends, making personal connections, or suggesting new features and uses.

Users immediately began identifying trends in the data sets and making their own assumptions fueled by the data available. For example, several users identified that the best represented age is 13. Some users also happily disclosed their initial hypotheses about Scratch blocks usage and their new findings. For instance, a user

stated that seeing the visualization helped him confirm his original hypothesis that the `forever` and the `when green flag clicked` blocks were among the most frequently used, but he expressed surprise in how seldom list blocks were used. Another user attempted to correlate blocks usage with the types of projects being created, suggesting that the high frequency of the `wait` and `switch to costume` blocks may be a result of animation projects.

Many users also made personal connections with the data. A user stated “WOW! I never knew that I was one of the first scratchers” and another that “when I joined it was the second most populated.” That particular user joined Scratch in March 2009 and was referring to the large number of user registrations that occurred during that particular month, one of the highest of all time.

Still others suggested new features we could implement or new ways we could use these statistics. One user suggested that we could use the *Location of Visitors* visualization (see Figure 3-3) to “analyze why there are a lot of visitors on one city than the other cities and maybe create a plan to get more traffic from those cities that is lacking from visits.” Many users also suggested implementing a zooming mechanism so that they can explore the blocks data more thoroughly since it is hard to see bars that are significantly lower than others. Other users would just like to see more data in general for example having total views, remixes, and love-its displayed for all projects. The feature requests suggest that users are interested in seeing the data more clearly and more in depth.

5.1.2 Patterns Observed By Researchers

Daily Visitors

The *Daily Visitors* visualization (see Figure 3-2) shows that there exists a cyclic trend with traffic to the website. Traffic tends to be notably less on weekends as compared to weekdays. Traffic is also less during the summer months (June, July, August) and the last weeks of December as compared to other times of the year. This cycle correlates with the attendance periods of primary and secondary schools.

Age of Users

By looking at historical data of the *Age of Users* visualization (see Figure 3-5), we can see how the data has changed over time. We notice that the percentage of older users to all users was much higher when the website was first launched in 2007 than it is today. In May 2007, 33.5% of registered users reported an age of 25 or above. In contrast, only 26.5% of users are above the age of 25 in May 2010. Although the percentage of older users has decreased, the mean age of users has not changed as significantly over time. In May 2007 the mean age was 23.8 and in May 2010 the mean age is 22.2.

Activity Per Month

From Figure 3-4, we observe that there is a general upward trend of new users that register each month. However, there tends to be a sharp decline in new user registration during the months of August and December, which are months when schools are not in session. The general positive slope of this series suggests that the number of users on the website increases quadratically with time.

Likewise, we can see from the same visualizations that the number of projects that are shared, remixes that are created, and comments that are posted to the Scratch website also exhibit similar upward trends. These increases are expected as the number of users on the website increases.

Project Creation Behavior

We see from the *Project Creation Behavior* visualization in Figure 3-12 that the vast majority of users have shared just a few projects. The maximum number of projects shared by any user is 4,815 projects by a female user. The next highest number of projects shared is 2,193, also by a female user. The highest number of project shared by a male user is 1,990, which is the third highest number of projects shared by any one user.

Comment Posting Behavior

We see from the *Comment Posting Behavior* visualization in Figure 3-13 that the majority of users have posted just a few comments. However the maximum number of comments posted by any user, as of May 2010, is 38,245 comments posted by a female user. The next highest number of comments posted by any user is 28,628 and is again by a female user. The most comments posted by a male users is 19,629, the ninth highest number of comments posted by any one user.

5.2 Personal Statistics

5.2.1 Reactions From The Community

We released the personal statistics trial run through a forum post. Each personal visualization page has a feedback widget where feedback can be provided privately and anonymously. Through these mediums we received a wealth of positive feedback.

Since a goal for providing these visualizations was to help users reflect about and learn from their previous Scratch usage patterns, we were happy to see that users did indeed achieved this goal. A user stated, “I can finally see how I grow, Scratchwise.” Another that “I have found so much about my projects. I use a lot of variables, and lots of control. I don’t use sound very often. I think I should do more comments.”

We were also happy to see that users are discovering correlations and making statistical claims using these visualizations. A question that is often asked by new Scratchers is how to achieve more popularity. A user analyzed the data and pointed out how to do just that, “I did some searching on a few other people’s statistics, and concluded that the more you comment, the more comments you get.”

Many users also provided external knowledge to explain phenomena shown in the visualizations. As an example, a user stated that “[this] clearly depicts my history on Scratch - from left to right, it shows how my friends viewed my stuff, how I got featured, how I got curated, and lots of little bits in between. I can also see clearly the rate at how fast I release projects. It all makes sense - it’s what I remember.”



Figure 5-1: This is the *Personal Scratch Blocks Usage* visualization for a popular Scratch user that predominately creates animations and artwork. The average count of blocks per project is shown.

In the forum thread there was also a discussion about privacy, whether it would be a problem if data about individual users were publicly available to all. Privacy concerns are very important and challenging issues that people today are often concerned about. It is also an important factor to understand when learning how to collect and use data. It was refreshing to see the issue brought up and discussed by young members of the Scratch community.

Users also continuously request new features with these visualizations suggesting that they wished more data was available to explore.

5.2.2 Patterns Observed By Researchers

We noticed that there may be a distinct difference in blocks usage for different types of Scratchers. Figure 5-1 shows the *Scratch Blocks Usage* visualization for a Scratch user that predominately creates animations and artwork. The visualization shows a strong preference for Control and Looks blocks. Some Sensing and Sound blocks are also used occasionally, but blocks from other categories are very rarely used. In contrast,



Figure 5-2: This is the *Personal Scratch Blocks Usage* visualization for a popular Scratch user that predominately creates games. The average count of blocks per project is shown.

Figure 5-2 shows the same visualization for a user that predominately creates games. We first notice that the average count of blocks used per project is much higher. We then see that this user employs a wider variety of blocks in projects. Additionally, this visualization shows that the user has a strong preference for Variable, Control, and Numbers blocks.

5.3 Remix Visualization

5.3.1 Reactions From The Community

The remix tree visualization was released without announcement by replacing the remix list seen in Figure 2-5 with the interactive visualization. This was done to collect initial reactions from users without any prompts. The official announcement occurred a couple of weeks later in the forums, by which time multiple threads had already pointed out the new addition to the website.

Users seemed to receive the remix visualization well and welcomed it as a replace-

ment to the list format. They also discussed how the visualization could be useful in their Scratch activities. For example, a user described how “it’s really helpful to see who remixed who, especially on projects with a bunch of remixes where before, you had to navigate through a confusing page of links.” And another stated that “its so great because it helps scratchers find the latest remix available.” This last comment refers to popular chain remix projects, where each remixer adds an element to a project, which then gets passed on for the next remixer to do the same.

After seeing this visualization, many users also requested a visualization of a similar nature in order to see friendship connections. It suggests that users are interested in seeing how Scratchers are interconnected, since these relationships are not easy to explore currently.

Chapter 6

Conclusion

The ScratchStats website was created and released in Spring of 2010. The website displays a series of interactive visualizations for users to experiment with and understand Scratch usage data. The motivations behind launching this website are to provide answers to questions about Scratch usage, to promote reflection and learning, and to promote data literacy.

The website release was well received by researchers and Scratch users. Researchers used the visualizations to successfully answer questions about ongoing activities of the community. Scratch users explored data about their own behaviors and noticed modifications they could make to enhance their experiences. They have also noticed and shared trends about the data that they have independently discovered.

There are still many areas for improvement of the user experience. These are detailed below.

6.1 Future Works

There are three areas that would benefit from future work: maintenance of the current visualizations, new visualization additions, and better integration with the Scratch website.

6.1.1 Maintenance and New Features

Since this is the first release of a statistics exploration site for Scratch, there are many improvements to be made and software bugs to fix. Many users have alluded to some areas that could be improved and some features that would be useful to have. Additionally, with the commenting tools on the visualization pages, more of these requests can be expected in the future. Future efforts should work to improve the site's usability based on feedback collected.

To make these visualizations more valuable to the user, we can also provide customized suggestions based on what the data shows. For example, for the personalized blocks usage visualization, we can analyze which blocks were seldom used and provide sample projects that use these blocks. These customized suggestions would encourage a user to explore and grow as a developer.

6.1.2 New Visualizations

Only a small set of visualizations have been created and released to date. This set can and should be extended in the future, especially in the personal statistics category.

Currently, there are two personalized visualizations available for users to explore their behaviors. Several other visualizations can be developed to provide users with more tools for exploration. For example, a visualization that shows how statistics are broken down by project can help users pinpoint which of their projects are the most popular, which projects took the most effort, or even which projects contain the most sprites, scripts, and costumes. Another possible visualization is a text cloud of all the comments a user has posted or received. This would help users see how the community is qualitatively responding to their work. A similar visualization would be a text cloud of all the tags of a user's projects to identify what types of projects the user is more inclined to create. Many users also proposed a friendship visualization in order to explore how different Scratchers are connected to each other.

When we have a collection of these personal visualizations, we should consider a way to better present them. A possible approach is a dashboard model where users

can track their most important visualizations on a single page.

6.1.3 Better Integration

In the future we should also consider how to better integrate the ScratchStats site with the Scratch website. For instance, we currently use a separate discussion widget on the ScratchStats site, instead of Scratch comments. This requires commenters to input their name and (optionally) an email address to associate with each post. We also require users to input their username to see their personal statistics instead of detecting which user is currently logged in. Scratchers have pointed out that this means all personal statistics are actually publicly available to all and can be a privacy concern. The entry point to personal statistics should be through the ‘My Stuff’ page of the Scratch website since that is where all a user’s personal artifacts are collected.

6.2 Final Words

This thesis provides evidence to suggest that children can be interested in and capable of understanding data, especially when presented in a visual manner. Using interactive visualizations users can answer questions to satisfy their curiosity and gain experience with data analysis. We also attempt to stimulate a user’s iterative learning process by providing tools to help users reflect about their creations and activities. By continuously improving the ScratchStats website based on feedback, we can further increase its benefits to users.

Appendix A

Database Tables

project_info The `project_info` table stores the metadata for each project as well as the number of sprites included in the project. The metadata includes the authors name, comments about the project, which project it is based on, the language settings when the project was saved, the version of scratch used to create the project, as well as the text representation of the save and share history.

project_save_history The `project_save_history` table stores the save entries for each project. This information represents each time the user saved the particular sb file. It includes the date the file was saved, which filename it was saved as, and the local and shared name used at the time.

project_share_history The `project_share_history` table stores the share entries for each project. This information represents each time the user has shared his or her project. Like the `project_save_history` table, it also includes the date the file was shared, the filename (project name), and the local and shared username used. This table is particularly useful in determining the remix history of a project as the share history defines the origins of any project.

project_media The `project_media` table includes all media (sounds and images) used in projects as costumes, backgrounds, and sounds. It also records whether

or not the media may appear in the libraries included with the Scratch application. However, it tries to define this characteristic only by comparing media names. It would generate false negatives if the user renames the media upon importing into the project, and would generate false positives if the user names a custom created media the same name as one in the library.

project_drums The `project_drums` table records all usages of the drum block. It records which drum number was used or what other parameter was used in place of a drum number. This could range from the random block to a variable value.

project_midi_instruments The `project_midi_instruments` table records all usages of the instrument block recording either the instrument number or parameter used in its place.

project_sprites The `project_sprites` table provides summary information about each sprite in a project. It lists the number of scripts, sounds, and images found for each sprite, as well as the name of the sprite.

project_sprite_blocks_stack The `project_sprite_blocks_stack` table stores all the block stacks (scripts) contained in a project. These stacks are stored in two ways: a human readable form and a lisp-style form that is easily parsed.

project_blocks_count The `project_blocks_count` table contains a tally of blocks used per project. It contains fields for every block (including obsolete blocks) to simplify data analysis. For each project, the number of times each block is used in the project is stored. Disconnected blocks (blocks that are not part of a stack that starts with a hat block) are discounted from the totals.

project_sprite_disconnected_blocks The `project_sprite_disconnected_blocks` table stores blocks that are inaccessible when in presentation mode. These blocks are not part of stacks that begin with a hat block. The motivation behind storing these blocks is to gain insight on how blocks are experimented with. The

assumption here is that users may drag out blocks to try even if they do not include it in the final project itself.

project_user_generated_strings The `project_user_generated_strings` table is a table of strings a user can produce and incorporate in a project. These include code comments, say blocks, think blocks, broadcast messages, variable names and values, and list names and values. Other strings users can generate such as costume, sound, and sprite names can be readily found from other tables.

References

- [1] Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., Millner, A., Rosenbaum, E., Silver, J., Silverman, B., and Kafai, Y. (2009). Scratch: Programming for All. *Communications of the ACM*, vol. 52, no. 11, pp. 60-67 (Nov. 2009).
- [2] Monroy-Hernández, A. and Resnick, M. (2008). Empowering kids to create and share programmable media. *Interactions*, March-April 2008.
- [3] Monroy-Hernández, A. (2007). ScratchR: sharing user-generated programmable media. Proceedings of the 6th International Conference on Interaction Design and Children.
- [4] Monroy-Hernández, A. (2009). Designing a website for creative learning. In: Proceedings of the WebSci'09: Society On-Line, 18-20 March 2009, Athens, Greece. (In Press)
- [5] Ayres, I. (2007). *Super Crunchers: Why Thinking-by-Numbers Is the New Way to Be Smart*. Bantam, 1 edition.
- [6] Rubin, A. (2002). Interactive visualizations of statistical relationships: What do we gain? In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics, Cape Town, South Africa*.
- [7] Resnick, M. (2007). Sowing the Seeds for a More Creative Society. *Learning and Leading with Technology*.

- [8] Papert, S. (1987). Constructionism: A New Opportunity for Elementary Science Education. Proposal to National Science Foundation.
- [9] Resnick, M. (2009). Dispatches: Kindergarten For Life. *Edutopia*. June 2009.
- [10] Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. Basic Books. New York.
- [11] Kay, A. (1991). Computers, Networks, and Education. *Scientific American*, vol. 265, no. 3, pp. 100-107.