**ALY 6070 Data Communication and Visualization**

Effects of Remixing Scratch Projects

By

Gift Ameh, Kaushik Burra, Pratik jitendra Gala, Savannah (Ya-Hsuan Lin), Yuanchang Li

### **Table of Contents**

[Introduction](#_nkt1v54en3ng)

[Research Questions](#_e4agjc9kj0um)

[Data Source](#_6iy5j1jpaudk)

[Data Cleaning](#_cyu1nd3511iv)

[Methods](#_sb3f0j2hv5la)

[Analysis Results](#_zfk3reqt1mna)

[Conclusion](#_iqxy9xks0w90)

[Appendices](#_au6qx5q0k175)

[References](#_nn1w8cj4n9eg)

### 

### 

### **Introduction**

The scratch community is a place where people from all ages and backgrounds come together to collaborate, share, mentor, and develop skills that are relevant in our today’s data driven society. Scratch is specifically catered with young children in mind, however, people from various backgrounds and ages use the platform to educate, collaborate, and inspire each other. The Scratch platform enables users to create interactive stories, games, and animations in the form of projects. The platform also enables users to share their projects with the Scratch online community [[9](#lk41m6b3qcbk)].

According to an online article we read on the Scratch website, about 26.5% of all recently shared projects are project remixes [[6](#d3kxg2x9ym01)]. There have been many controversies about whether remixing projects is an innovative way of inspiring or enhancing educational development. Based on some of the articles we have read, it’s our understanding that some users are unhappy because other users are remixing their projects without properly acknowledging the original creator. We agree that it is unfair when some users remix other users’ work and take full credit for the work that original creators have done. Some articles pointed out that even though remixing projects can motivate users’ behavior to be more active in the Scratch community, it may cause plagiarism when students forget to refer to an original project [[3](#x9eiejerqke7)].

On the hand, some articles highlight the creativity of remixed projects in Scratch, claiming that remixed projects not only educates users, but also inspires them to think outside the box and innovate [[7](#9t3s62iov3tv)]. These controversies piqued our curiosity regarding the effectiveness of remixed projects. Hence, our goal for this research is to understand these trends, drive to a conclusion, and provide recommendations based on our data driven results.

### **Research Questions**

For our research, we want to know whether remixed projects do more harm than good to the Scratch community. Some research questions that we figured would be useful for our analyses are as follows:

1. How does the number of projects, the number of views, and the number of downloads influence the behavior of a user clicking the *Love It* button on a Scratch project?
2. Based on the variable(s) that has the highest correlation with *Love It*, can we determine what types of projects users are mostly drawn to? Do they prefer remixed or original projects?
3. How does the trend between the highest correlated variables and the numbers of *Love It* differ from 2007 to 2011?

### **Definitions**

* **Is Remix**: When the original project is remixed for the very first time.
  + Here the parent\_user\_id and the user\_id are different.
* **Is Self-Remix**: When the project is remixed by the original creator of the project.
* **Is Remixed**: When the project has already been remixed and it is being remixed again.

### 

### **Data Source**

This dataset was given to us by our instructor, Professor Alice Mello, and it is our understanding that the data was originally collected from the Lifelong Kindergarten group (LLK) at the MIT Media Lab or the Harvard Dataverse Network. The data sets consists of 2,213 observations and 43 variables. The variables used for our analysis includes: **Love It**, **Views**, **Project**, **Downloads**, **Scripts**, and **Sprites.**

### 

### **Data Cleaning**

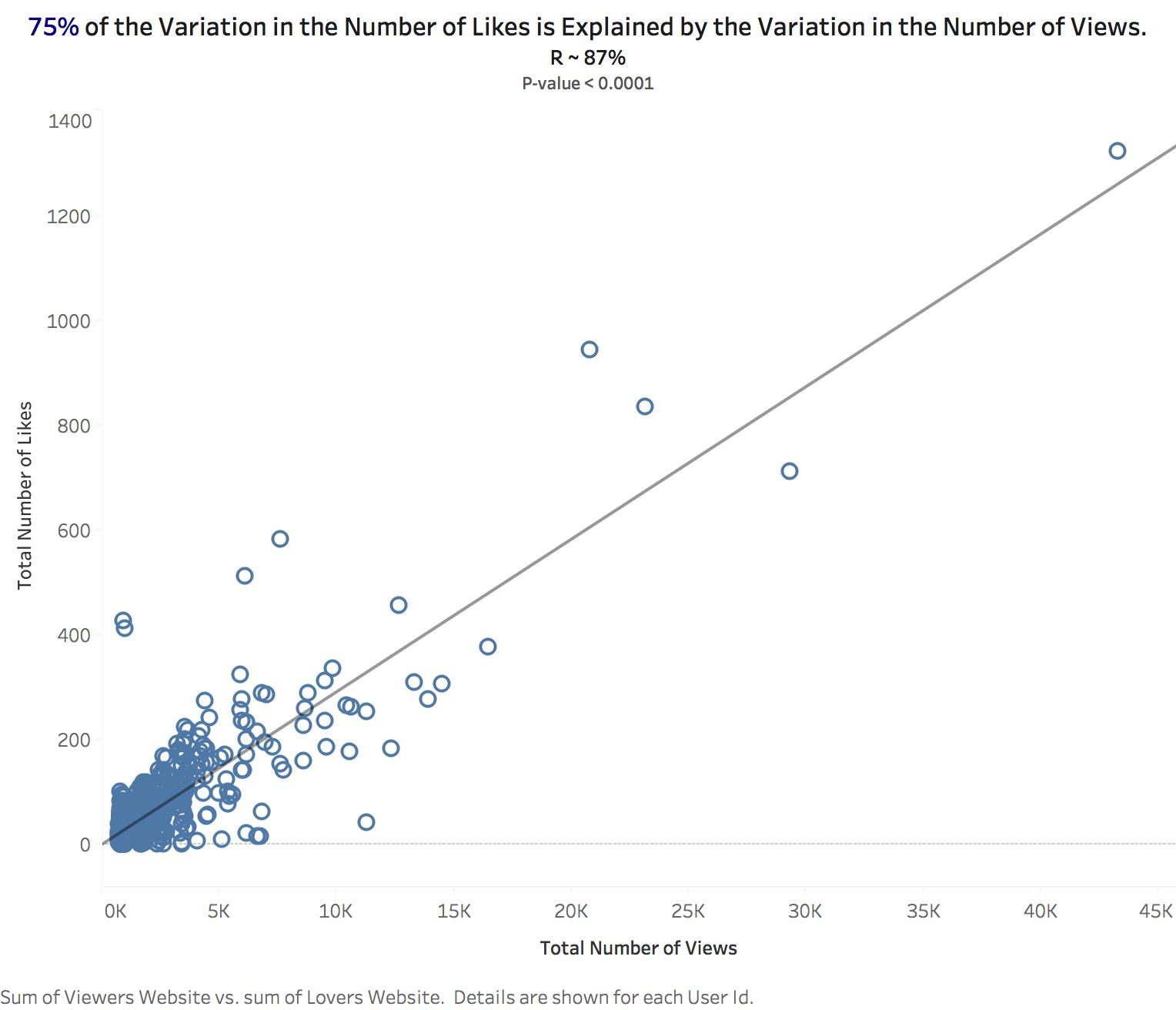
This dataset was given to us in a clean format except for some missing values. The dataset did not require much cleaning. We excluded null values during our analyses.

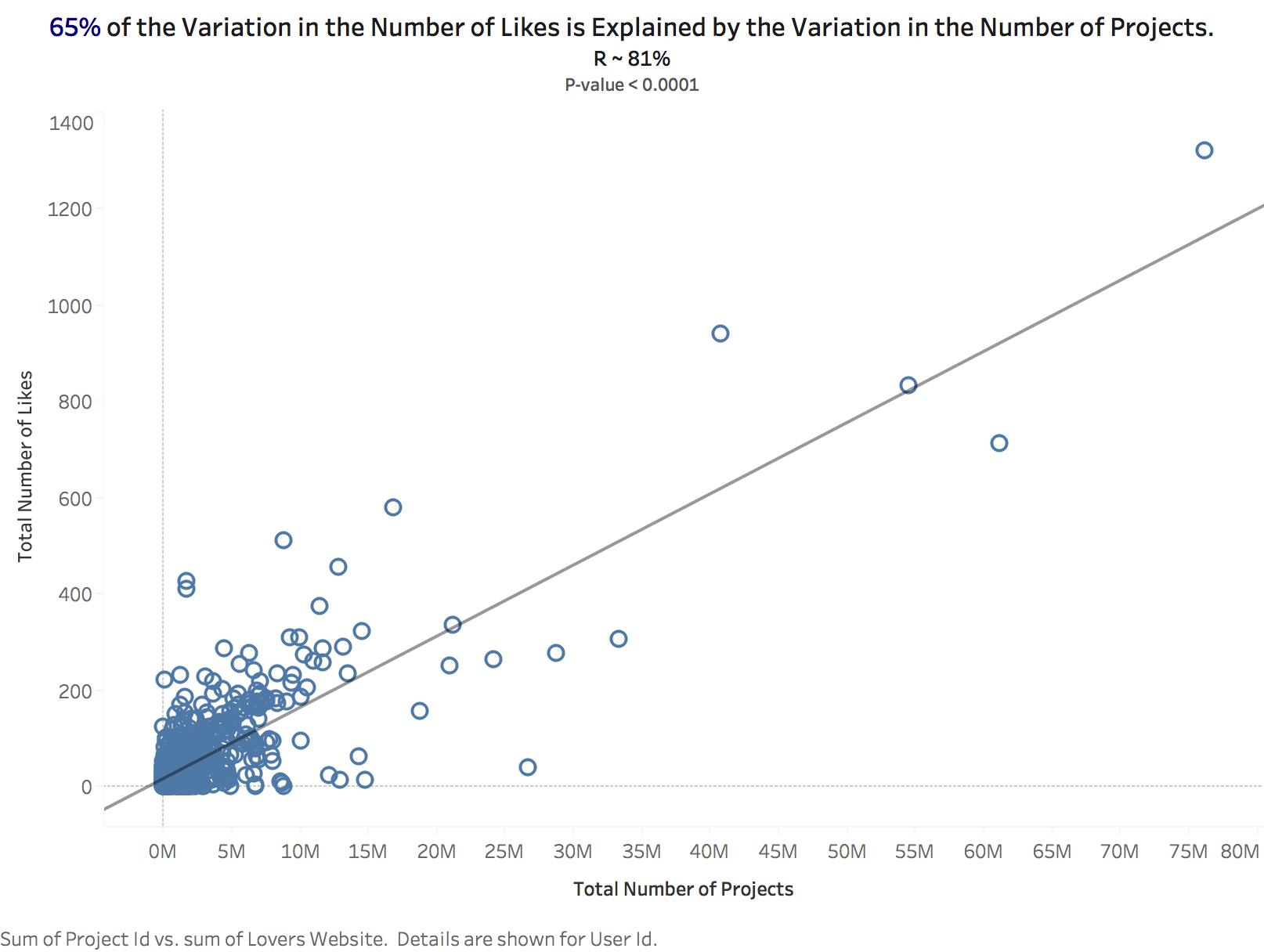
### **Methods**

We began by exploring the dataset, then we created several dashboards in Tableau to understand the distribution of our dataset. Furthermore, we performed a multilinear regression between the number of (*Views*, *Downloads*, *Scripts*, *Sprites* for each user) and the number of (*Love* for each user). We chose *love It* as our target variable because we believe that the number of likes a user gets explains how impactful or creative that project is. After performing a Multilinear regression, we began to explore further to understand the relationship between our variables.

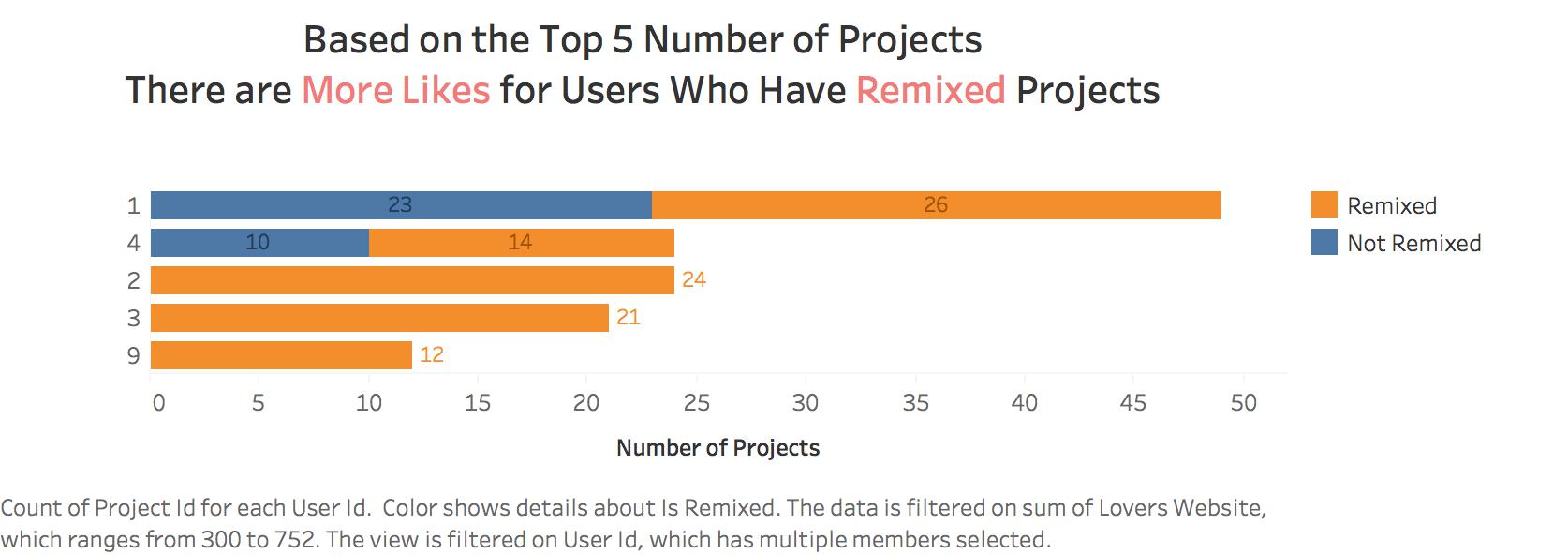
### **Analysis Results**

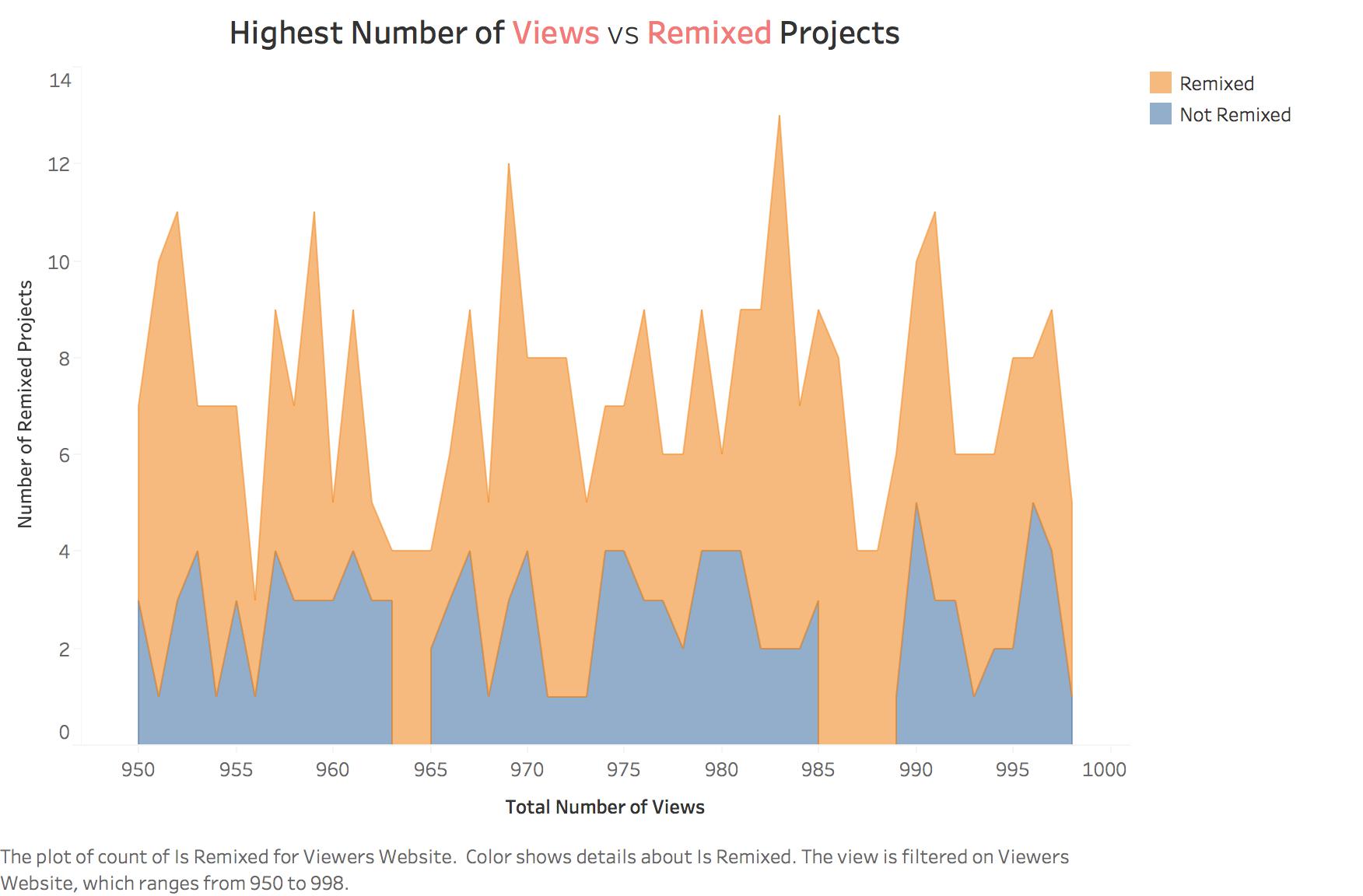
From our analyses, we discovered that the number of views and the number of projects are highly correlated with the number of likes for each user. Also, we determined that there exists a positive correlation between our independent and dependent variables. Hence, we can infer that as the number of views or project increases, so does the number of likes for each user. Visit [Appendix 1](#kkhhs06io1l2) for more graphs.



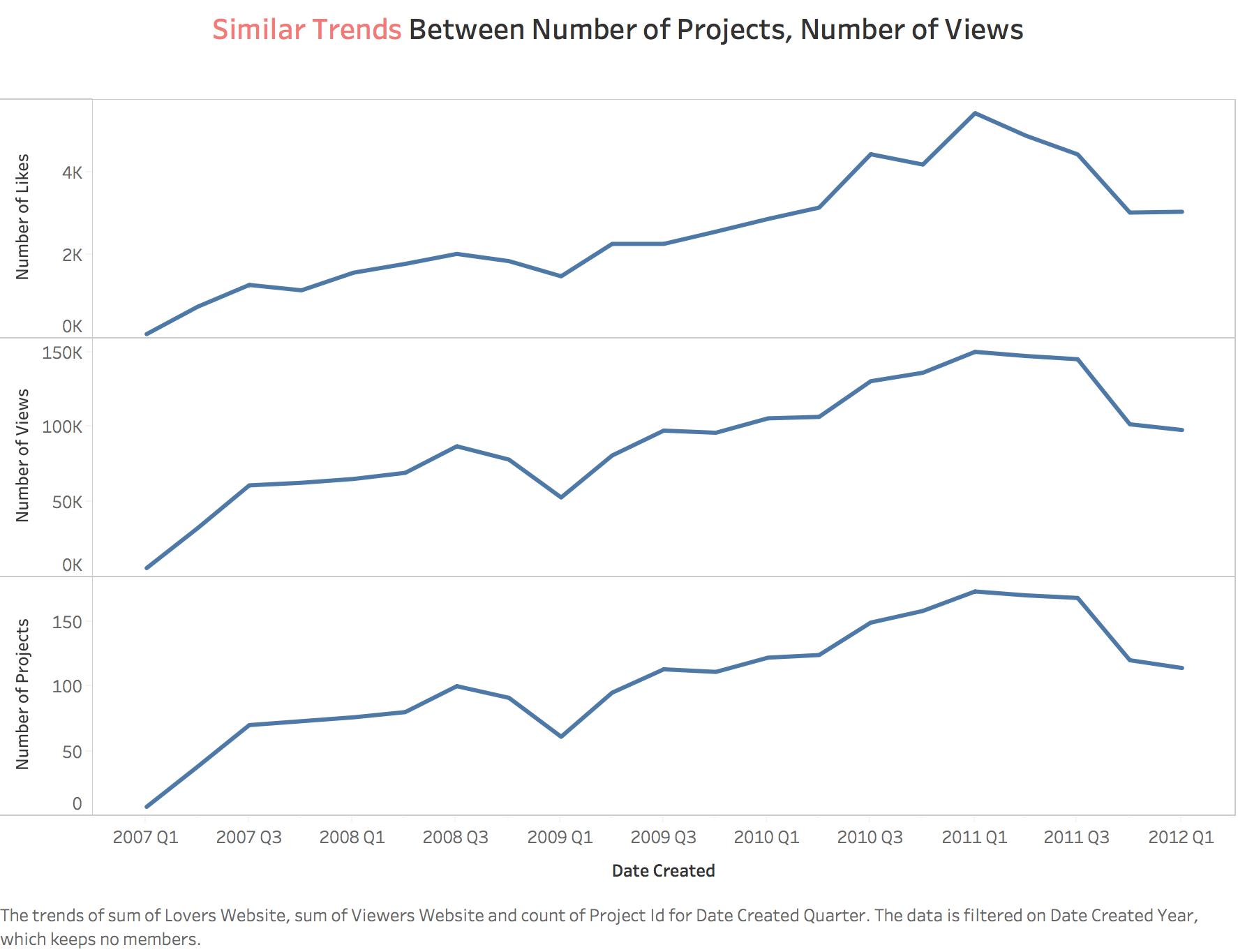


Why is this information useful? Well, we want to understand what variable(s) has the highest correlation with our dependent variable *Love It*, and explore from there. Hence, we want to determine whether users with the highest number of views/projects have mostly remixed projects. Our analyses results showed that indeed users with the highest number of project/likes are creating mostly remixed projects.





Lastly, we observed the trend from 2007 to 2011 between the number of projects, number of views, and number of likes. Based on the chart shown below, we can see that the trends are similar, hence, supporting our correlation results. Visit [Appendix 2](#3it0l9s8z6gp) for more graphs.



### Conclusion

In this white paper, we demonstrated that number of number of views and number of projects a user has influences the behavior of a user clicking the *Love It* button. In Fact, there exists a high positive correlation between the number of views for each user and the number of likes, as well as the number of projects for each user and the number of likes. Furthermore, we demonstrated that majority of the users are viewing remixed projects, and the users with the highest number of projects/likes are creating mostly remixed projects.

What does this mean for the Scratch community? Well, despite the controversies, our analyses proves that majority of the Scratch community value the idea of remixing scratch projects. Since remixed projects are the most liked for users who create many projects, we can infer that the Scratch community is inspired by the creativity of remixed projects, and we can now confidently agree that remixing projects does enhance collaboration and educational development.

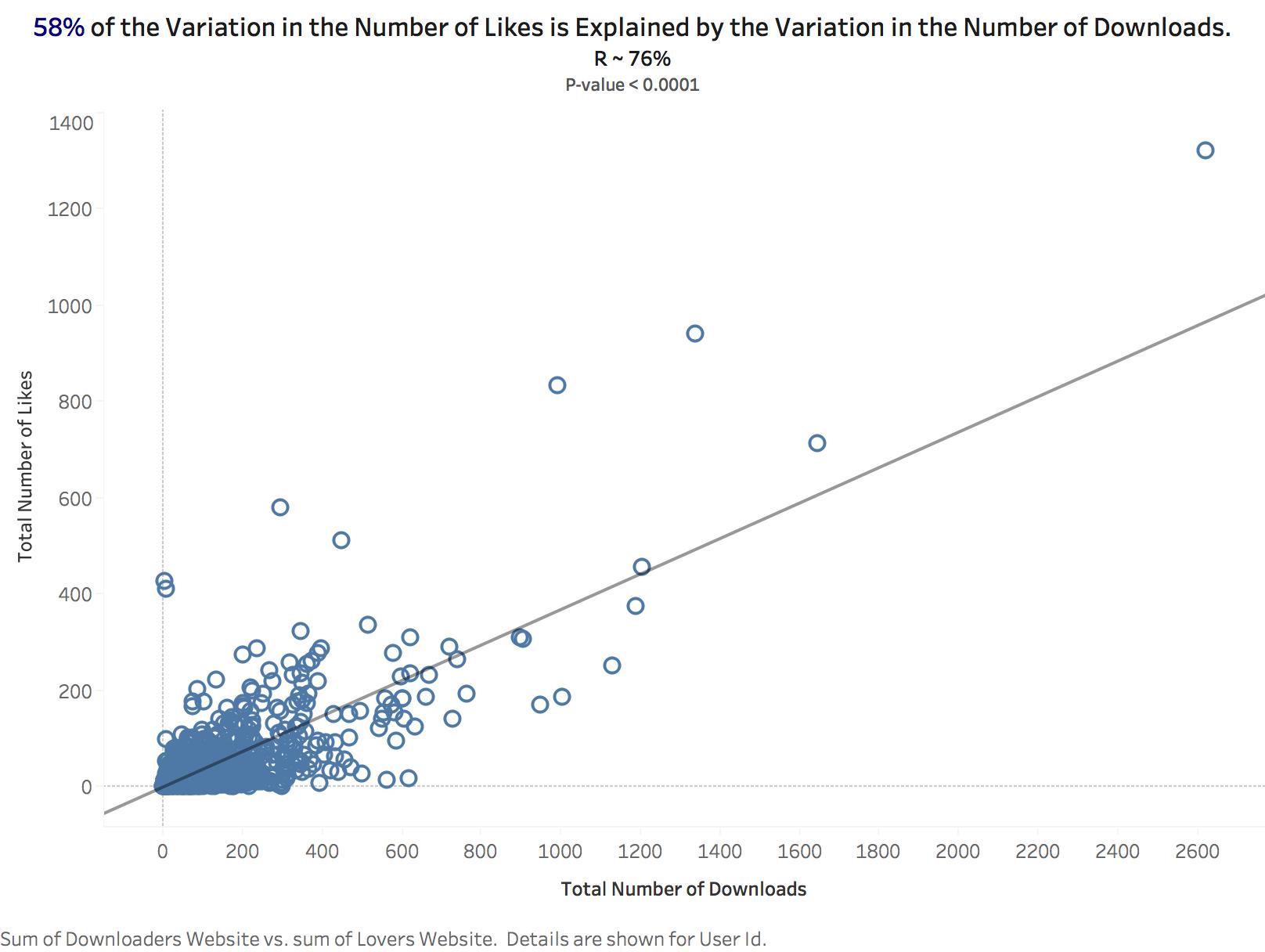
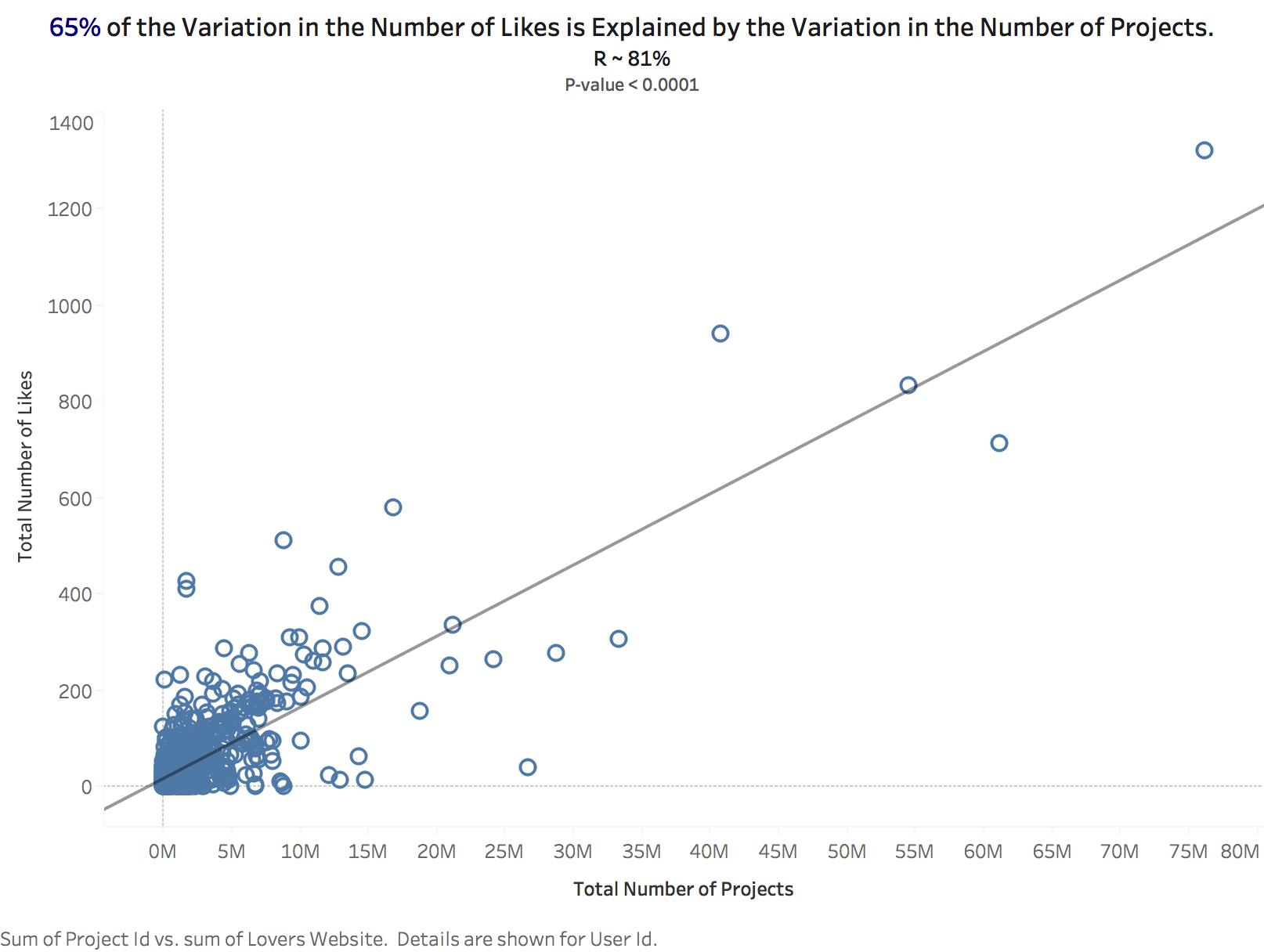
In addition, we do recommend that the Scratch inventors continue to encourage remixed projects to boost collaboration, creativity, and educational development while giving due credit to the original creators to improve user satisfaction. In addition, it would be helpful to educate users about plagiarism by encouraging them to take self-efficacy assessments before using the Scratch platform. Self-efficacy assessment can also motivate students to have a sense of belonging and become active producers in the Scratch community [[5](#l1l4uyay6m0v)][[10](#aatupok451yz)].

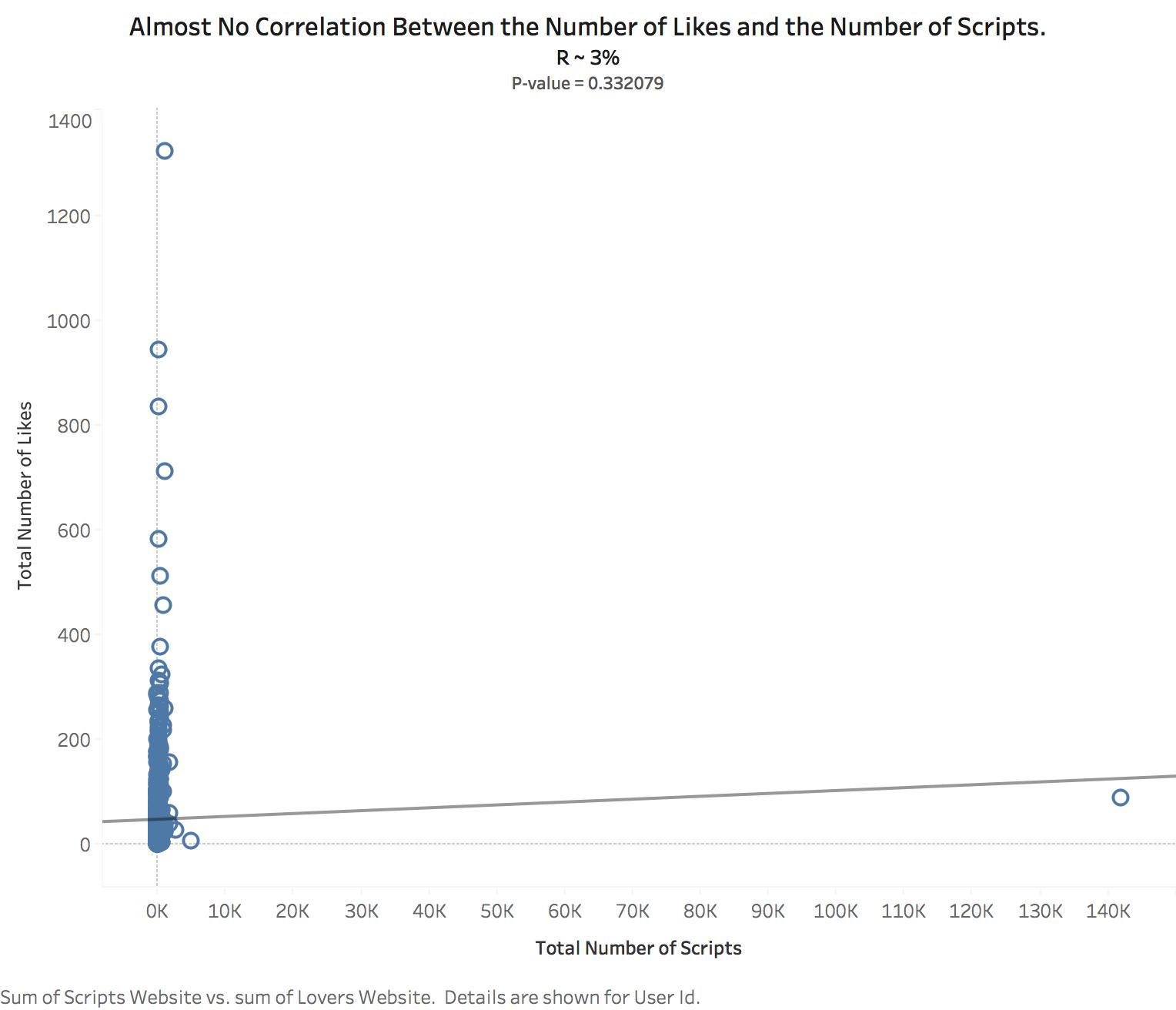
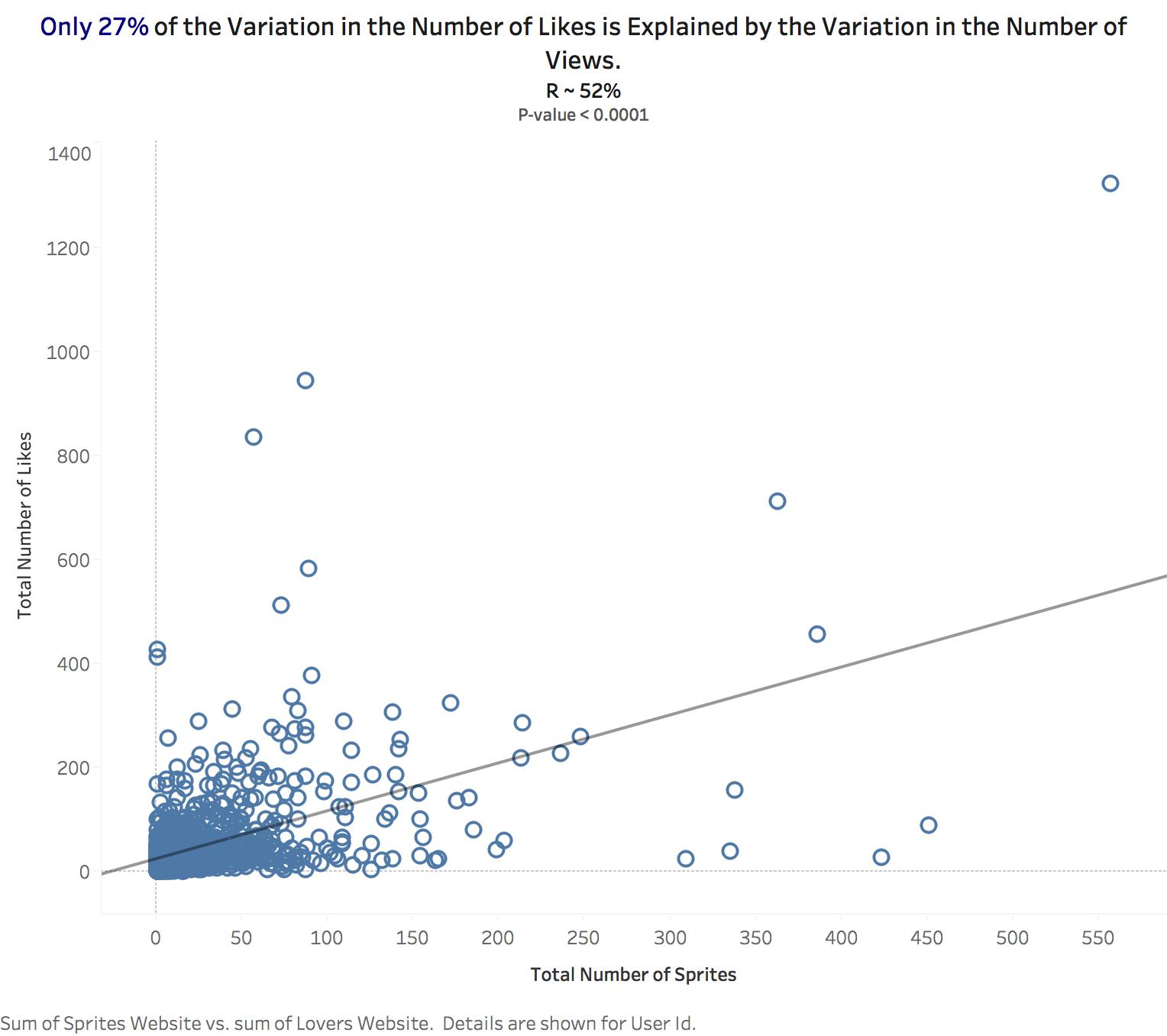
### 

### 

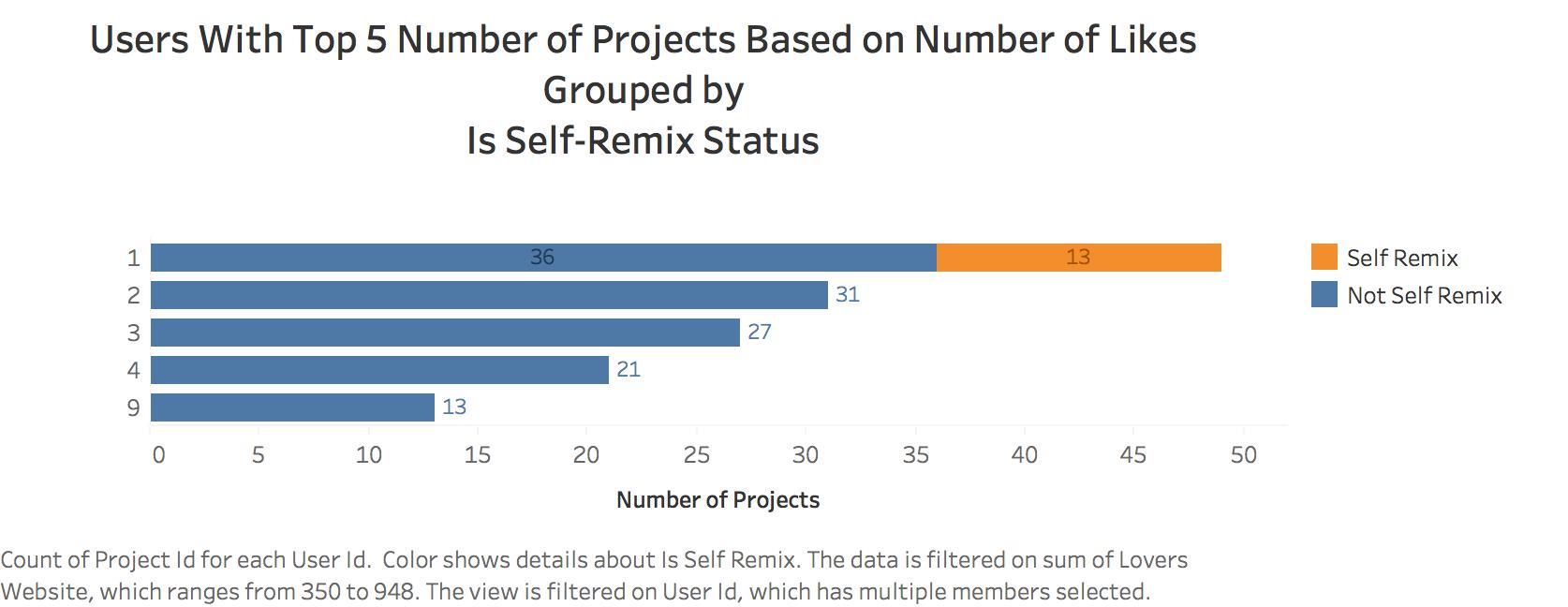
### Appendices

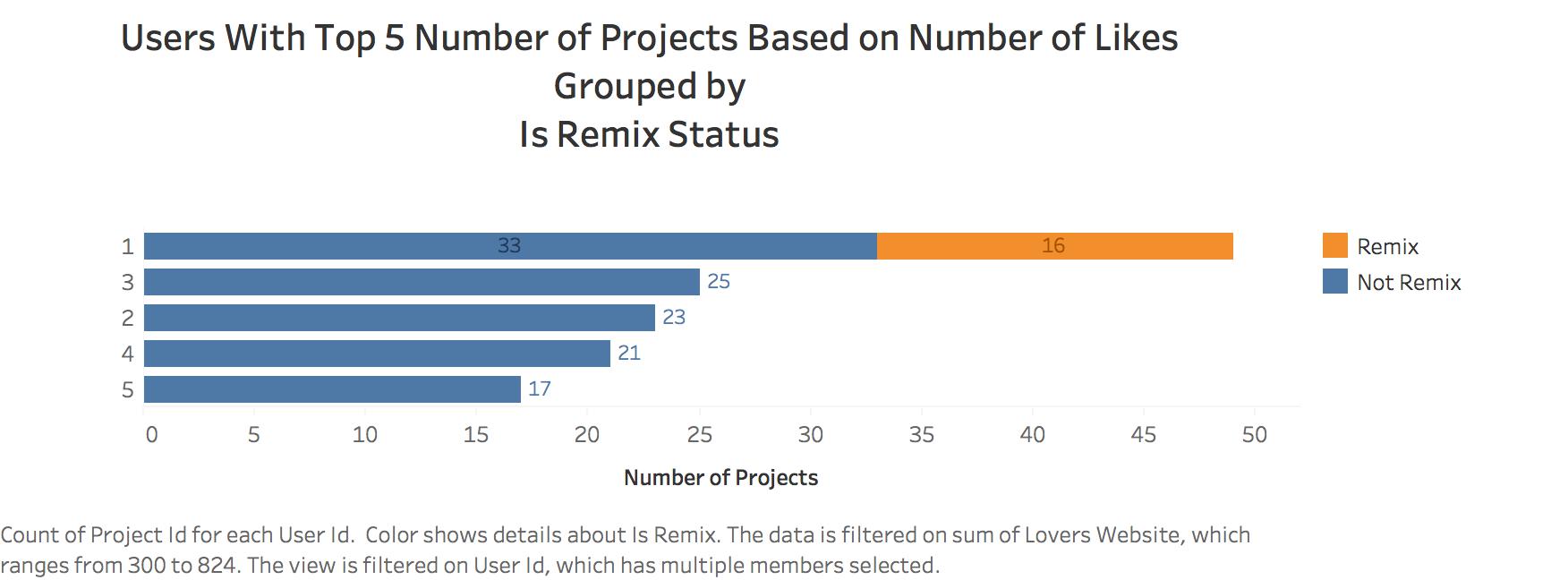
Appendix 1

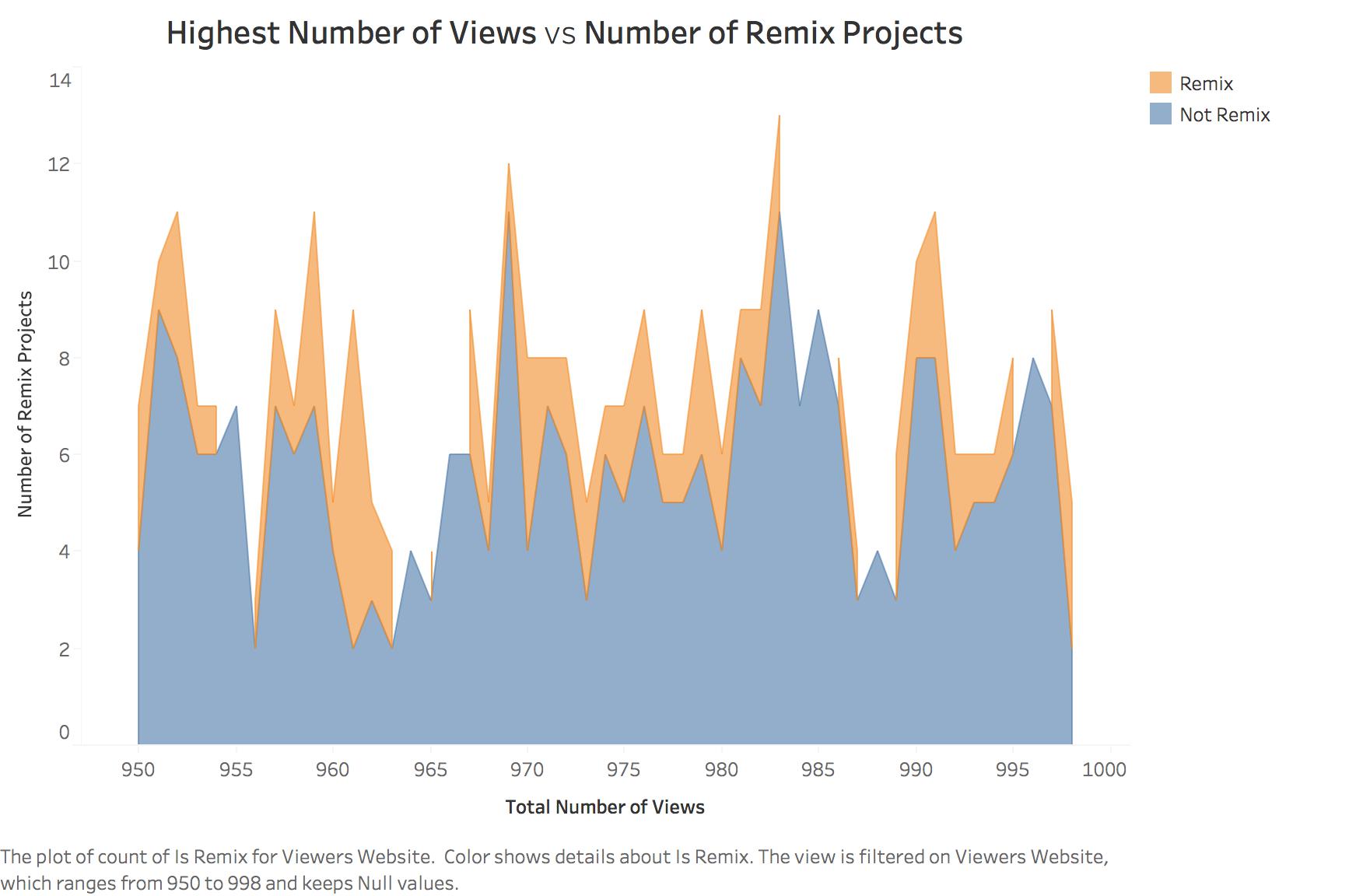
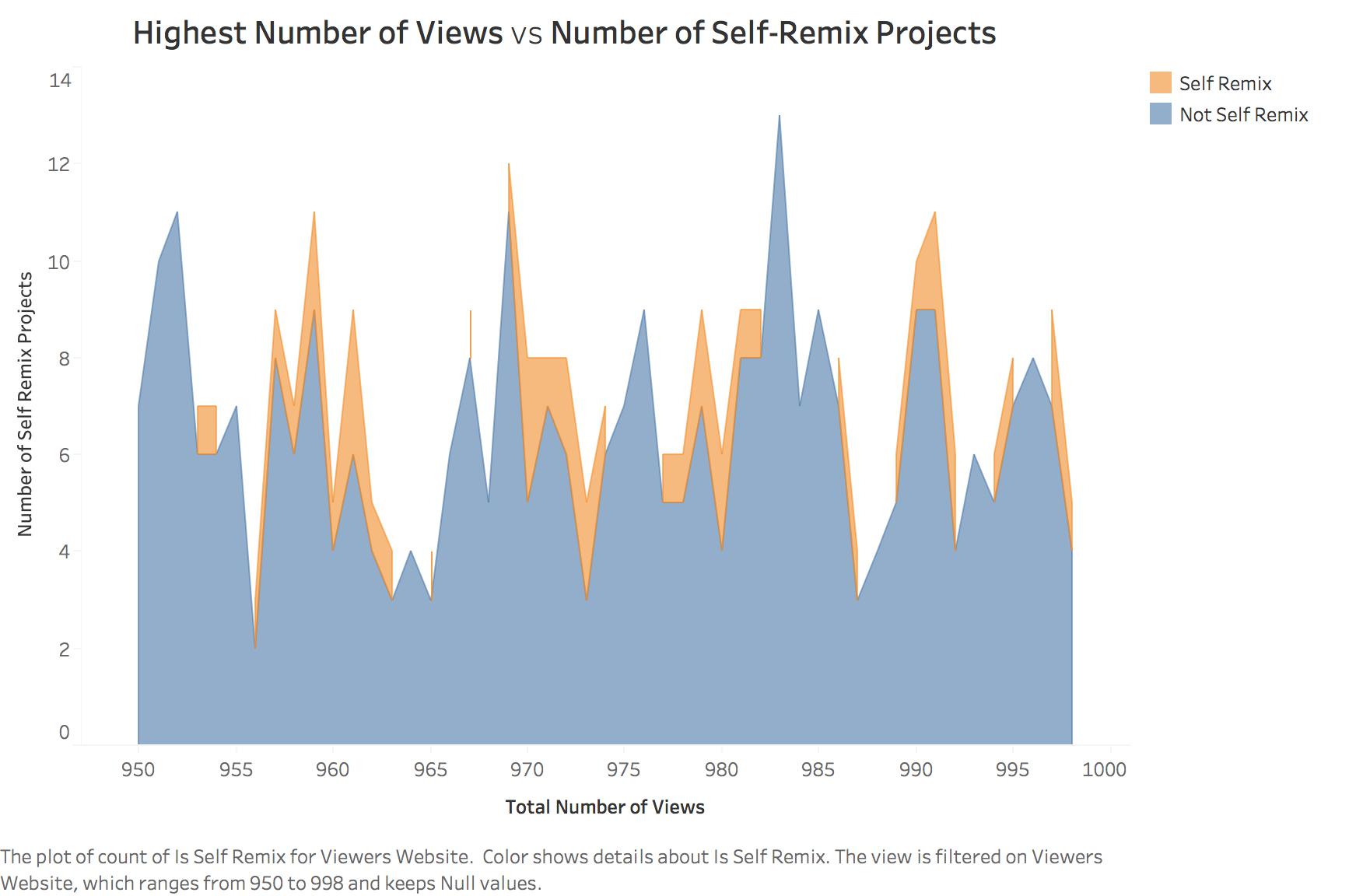




Appendix 2







#### **References**

1. Anderson, T. (2008). *The theory and practice of online learning*: Athabasca University Press.

1. Cambre, J., & Klemmer, S. (2017). *Standing on the Shoulders of Peers: Tournament-Style Remixing in Project Courses*. Paper presented at the Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, Portland, Oregon, USA.

1. Dasgupta, S., Hale, W., Monroy-Hernández, A., & Hill, B. M. (2016). *Remixing as a pathway to computational thinking.* Paper presented at the Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing.

1. Kirkwood, A., & Price, L. (2014). Technology-enhanced learning and teaching in higher education: what is ‘enhanced’ and how do we know? A critical literature review. *Learning, media and technology, 39*(1), 6-36.

1. Monroy-Hernández, A. and Resnick, M. (2008). *Empowering kids to create and share programmable media*. interactions 15, 2 (March 2008), 50-53.

1. Remix. (2016, February). Retrieved January 28, 2018, from https://wiki.scratch.mit.edu/wiki/Remix#cite\_note-1

1. Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., Millner, A., Rosenbaum, E., Silver, J., Silverman, B., Kafai, Y. (2009). Scratch: *Programming for All*. Communications of the ACM, November 2009.

1. Remix culture. (2018, January 15). Retrieved January 28, 2018, from https://en.wikipedia.org/wiki/Remix\_culture

1. Scratch - About. (n.d.). Retrieved January 28, 2018, from https://scratch.mit.edu/about

1. Stern, P. (2017). To scratch or not to scratch: American Association for the Advancement of Science.

1. Teaching students about the creative commons “http://www.learnnc.org/lp/pages/6443s” :LEARN NC, a program of the [UNC School of Education](http://soe.unc.edu), finds the most innovative and successful practices in K-12 education.

1. Yamagata-Lynch, L. C. (2010). *Activity systems analysis methods: Understanding complex learning environments*: Springer Science & Business Media.