



Analysis of Google Play Store Data

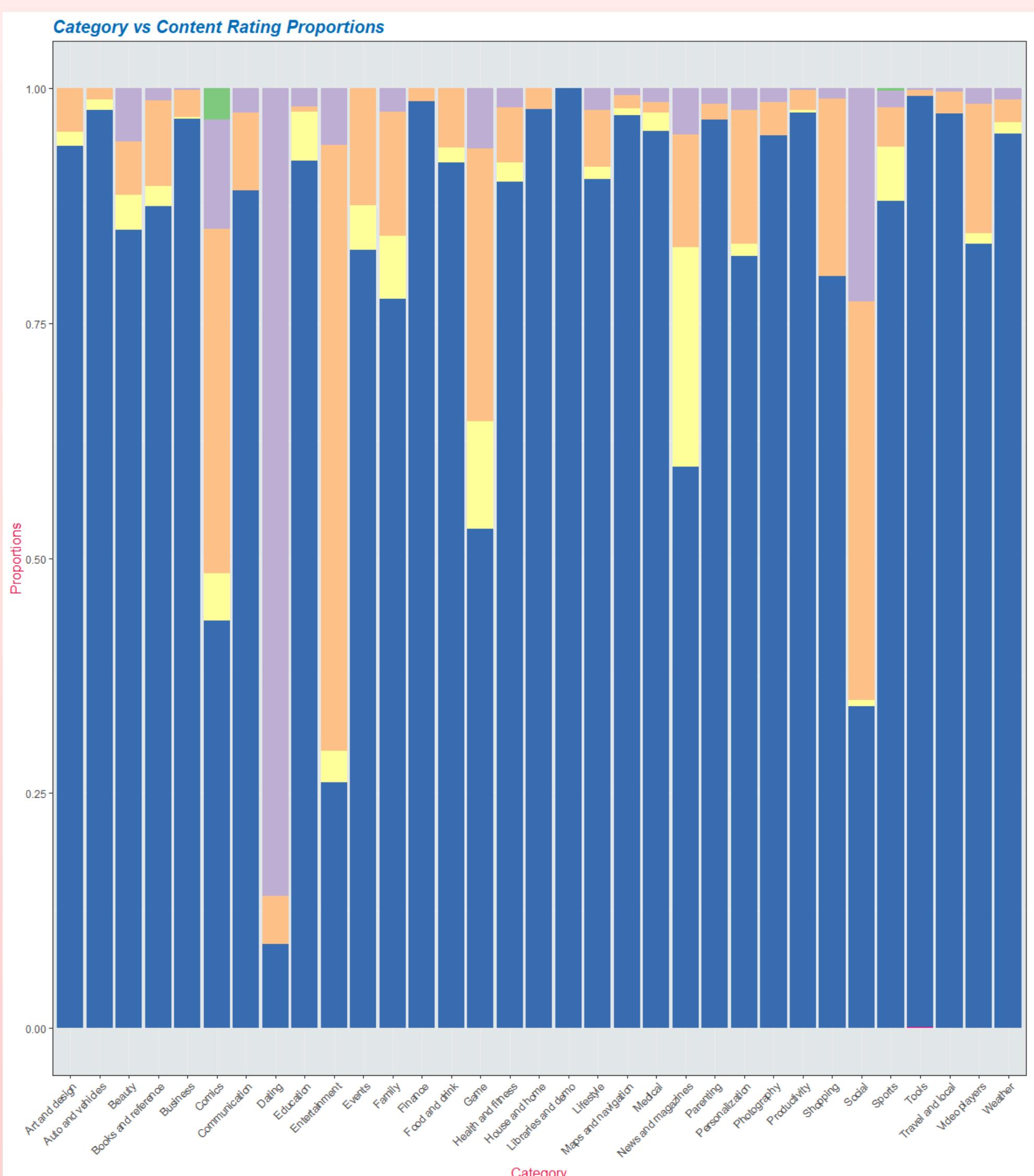
Scott Steinberg, George Brown, Jared Jory-Geiger, Kendal Krivinko
Department of Statistics, Carnegie Mellon University, Pittsburgh, PA



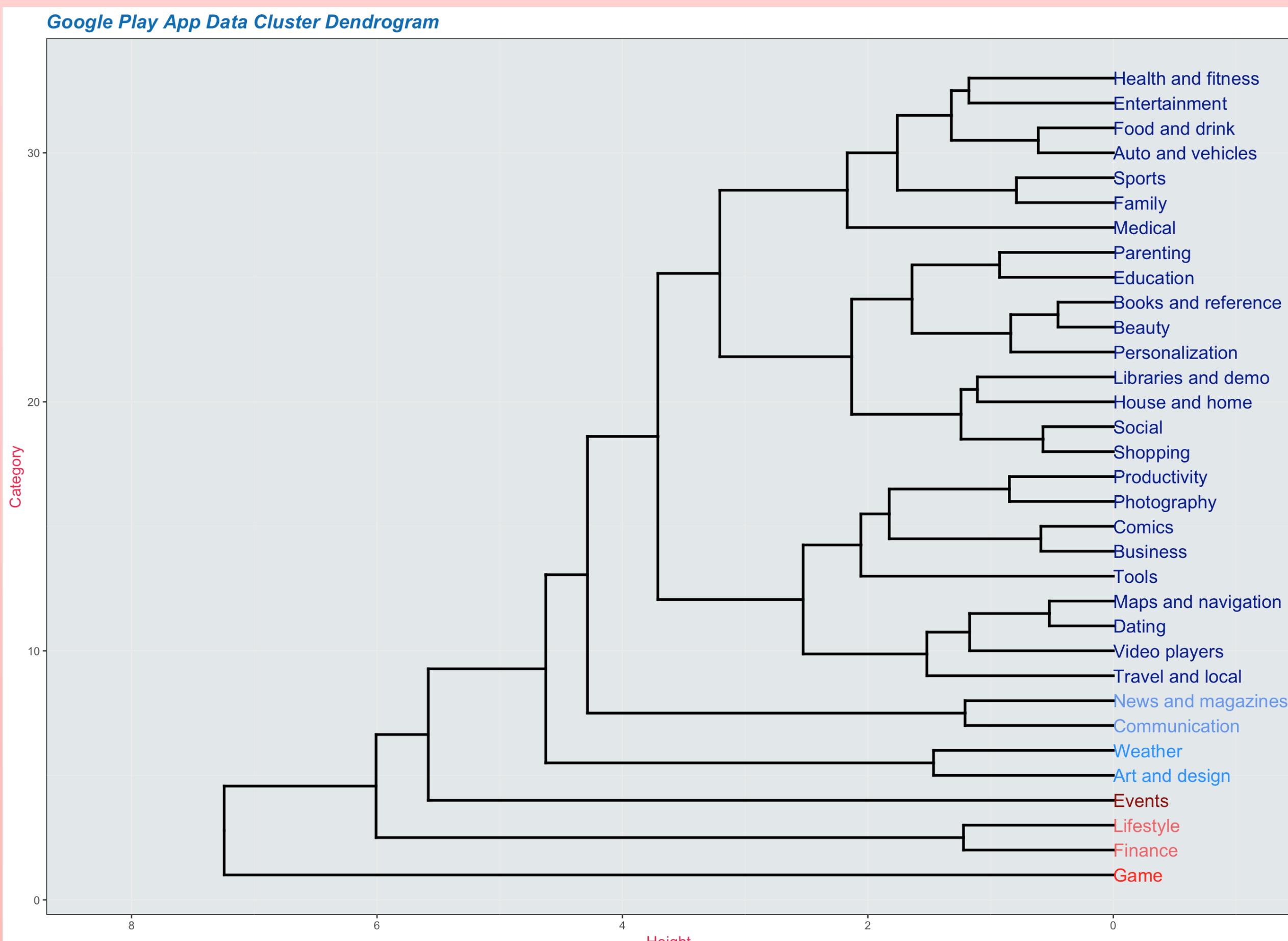
INTRODUCTION

Why this Data?

"Google Play" is a digital distribution service operated and developed by Google. "Google Play" is primarily used by Android users, and is used to buy and download apps, games, movies, and e-books. Our group was particularly interested in the categorical variables of "Type" (paid or free), "Genres," and Android "Version," as well as the continuous variables of "Price," "Rating," and "Reviews."



- Dating is the only category with a significant non Everyone rating
- Comics is the only category with a significant Adults only rating

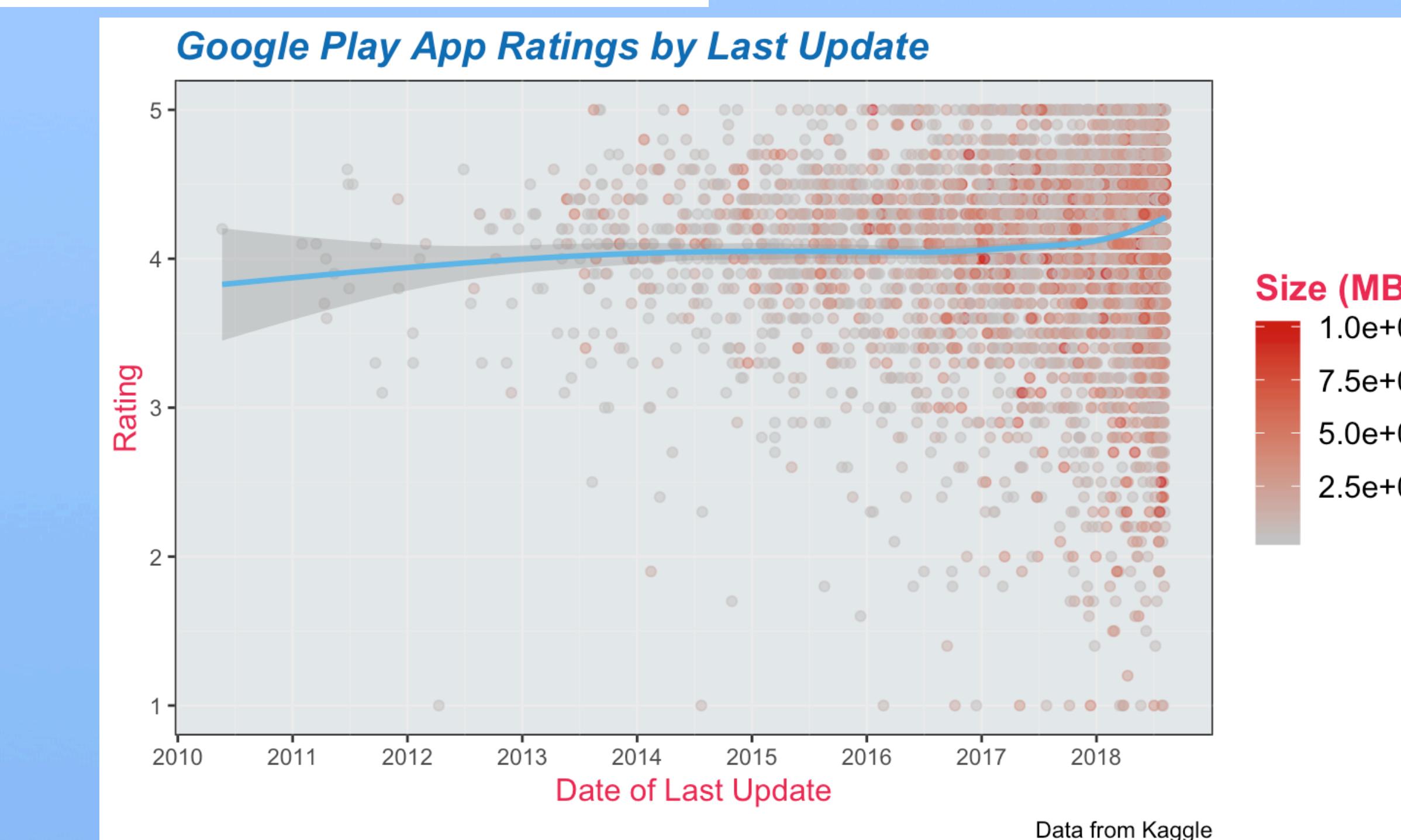


- Games are in a category of their own compared to other variables
- Social, communication, and dating apps are all in differing clusters
- Parenting and education are in the same hierarchy

Top 50 Words in App Titles

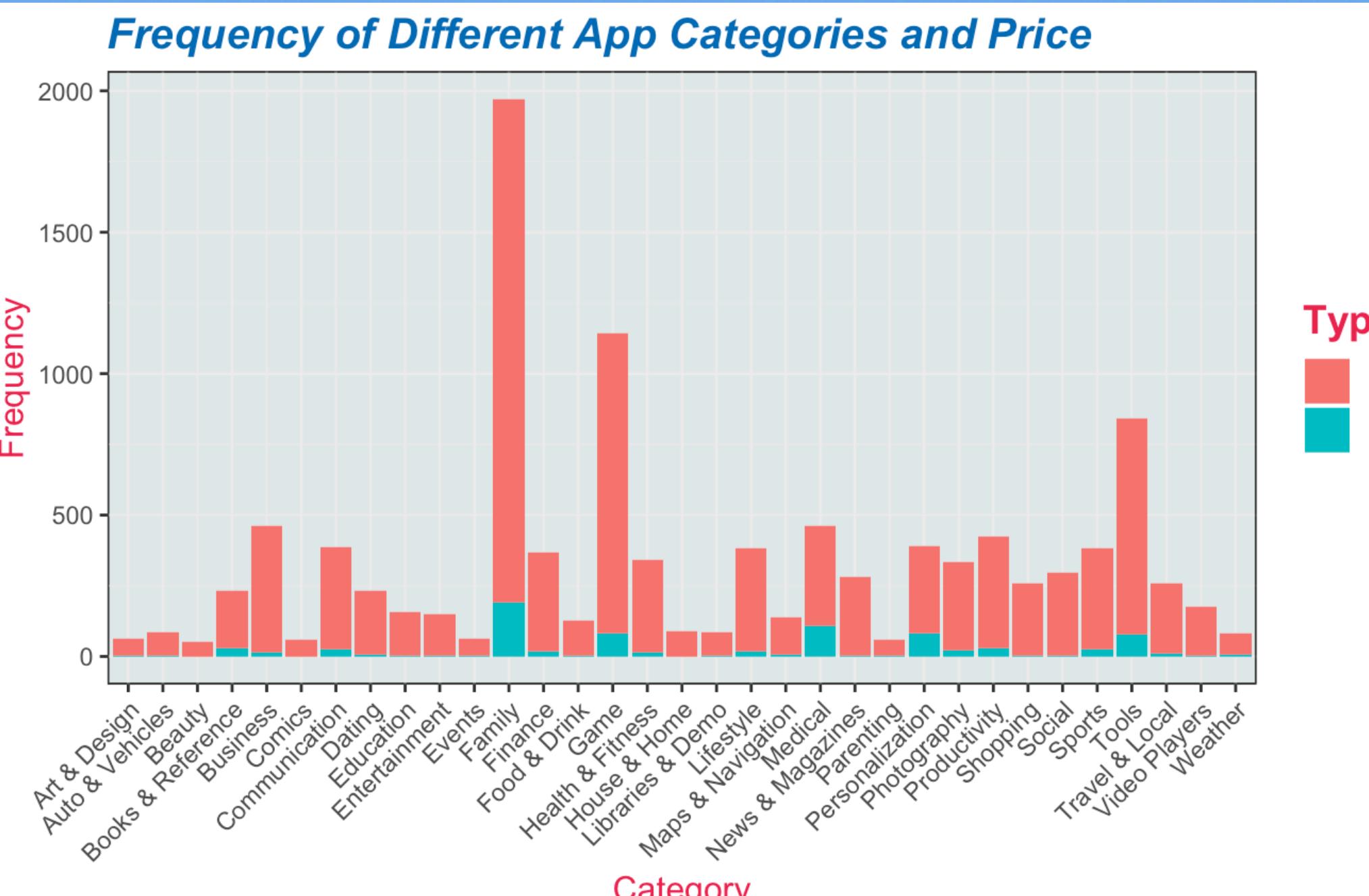
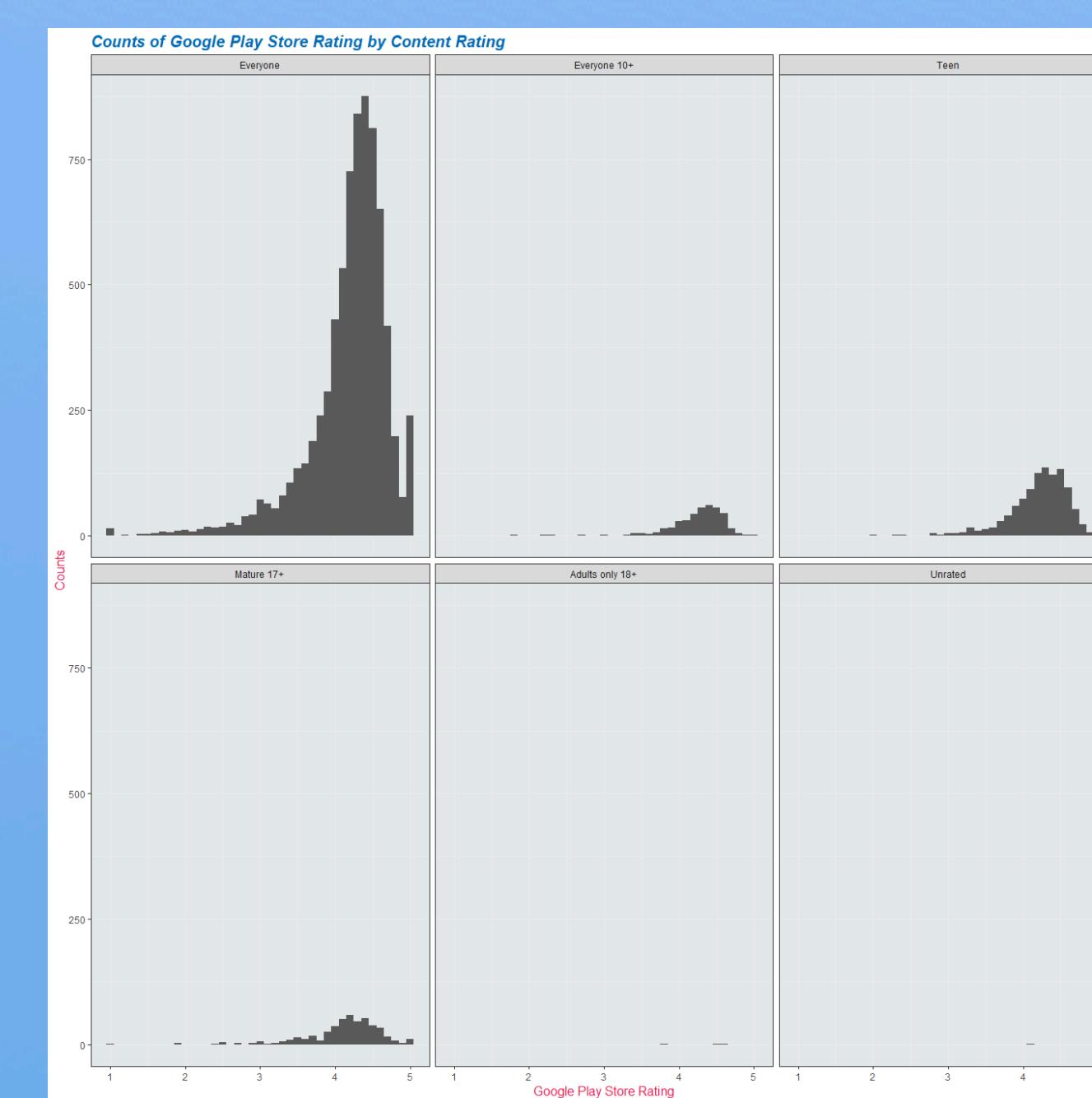


- The most common word is "Free," indicating a strong user preference for free apps
- Interestingly, redundant words like "app" and "mobile" are the next most popular
- Also, "2018" is a common word, indicating a tendency to use the current year in titles to have the appearance of modern design



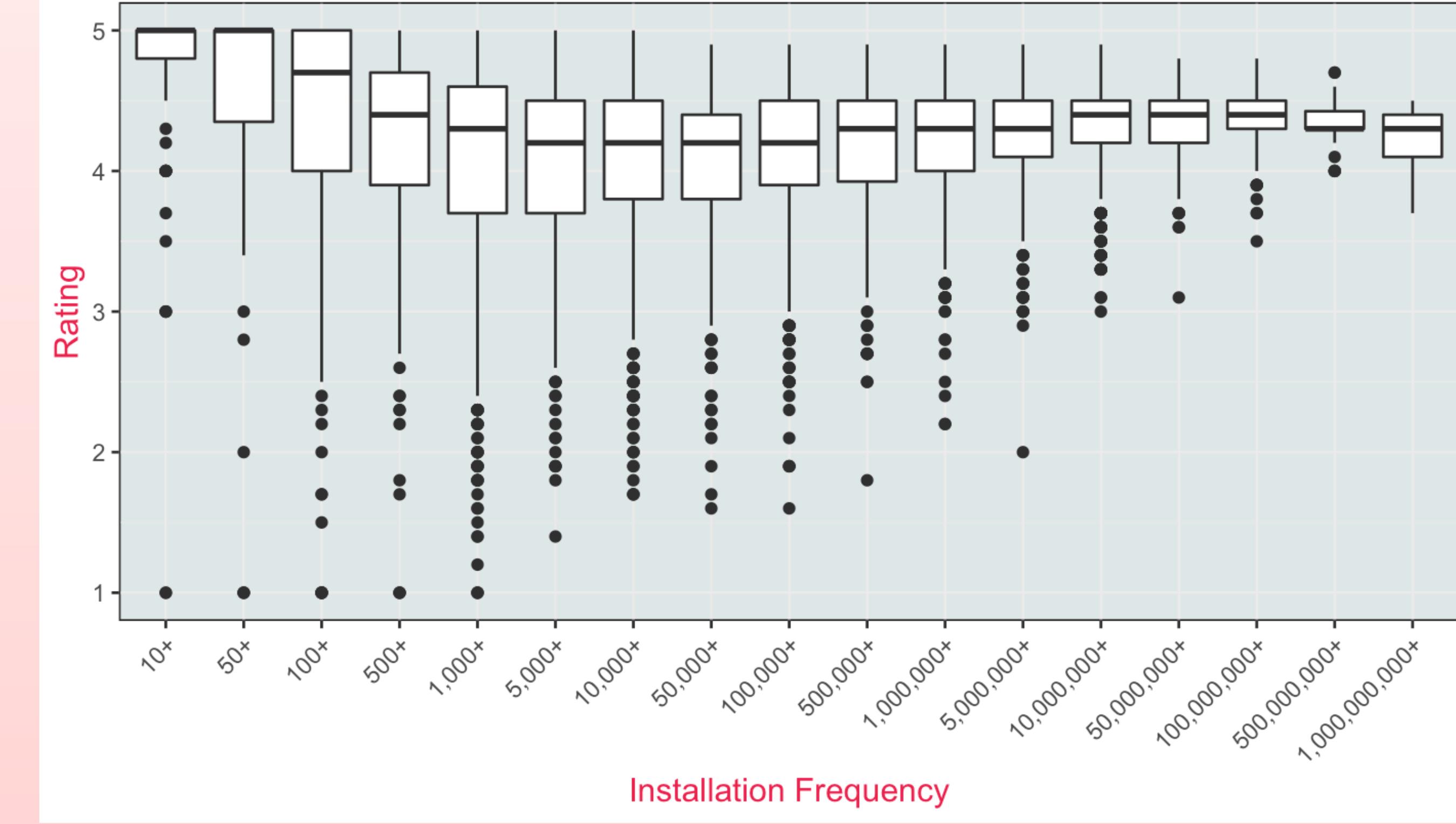
- Larger proportion for recently updated apps with lower ratings
- Apps updated in 2018 specifically tend to have higher ratings

- Teen has the highest average rating

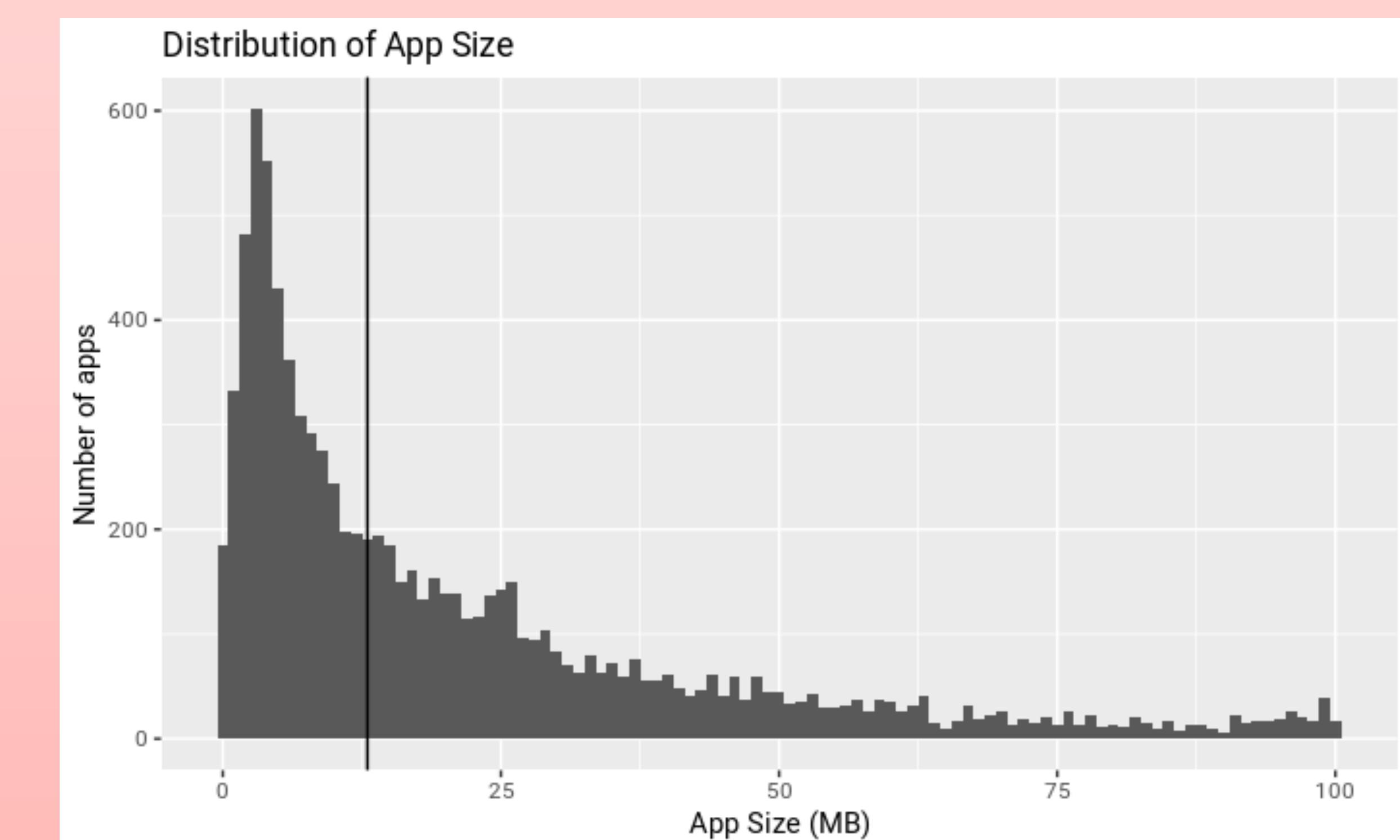


- The most frequent app category is "Family"
- Only a very small proportion of each app Category contains "Paid" apps

App Ratings vs. Installation Frequency



- The median and lower bound of the second quartile of "Rating" seems to decrease and then increase after 5,000+ ratings



- The median app size is 13 MB, and the largest app size is 100 MB
- The distribution of sizes forms a strongly right-tailed distribution
- This seems to indicate market preference for smaller-sized apps

CONCLUSION

From these graphs, we can see a few trends. Based on the word cloud, time series, and the histogram of app size, apps are generally preferred to be small, new, and free. Entertainment-based categories like Games, Entertainment, and Sports are the most common types of apps, and this is reflected in the dendrogram. On average, entertainment-based apps tend to have higher ratings and higher-maturity content ratings.

ACKNOWLEDGMENTS

The assistance and kindness of Dr. Matey Neykov and Ms. Robin Dunn should be noted in reference to this presentation. Their suggestions and edits were of great influence.