

# Lab 3:Data Visualization

1854116  
Mingzhi Zhu

Base on Google Play Store Dataset  
June 22, 2021

## Contents

|          |                                     |          |
|----------|-------------------------------------|----------|
| <b>1</b> | <b>Data Analysis Task</b>           | <b>2</b> |
| 1.1      | Dataset Overview . . . . .          | 2        |
| 1.2      | Objectives . . . . .                | 2        |
| 1.3      | Characteristics . . . . .           | 3        |
| 1.3.1    | Category . . . . .                  | 3        |
| 1.3.2    | Relationship . . . . .              | 4        |
| 1.3.3    | Content Rating . . . . .            | 4        |
| 1.3.4    | Price . . . . .                     | 5        |
| <b>2</b> | <b>Layout of Designed Dashboard</b> | <b>6</b> |
| 2.1      | Interface Design . . . . .          | 6        |
| 2.2      | Scatter Chart . . . . .             | 6        |
| 2.3      | Radiation Chart . . . . .           | 7        |
| 2.4      | Line Column Mix-up Chart . . . . .  | 8        |
| 2.5      | Deployment . . . . .                | 9        |

# 1 Data Analysis Task

In this section, report mainly describes a data analysis task for the chosen dataset Google-Play-Store-Dataset.

## 1.1 Dataset Overview

Google App Store has hundreds of thousands of software on the shelves, and the number of downloads has exceeded 2 billion times. For the majority of Android mobile phone users, it provides a very wide range of application options, which is very popular with users. More than 20000 pieces of music, free and paid, more than 450000 Android applications and games, the largest e-book market in the world. You can rent thousands of movies, including the latest movies and HD movies. Therefore, the app data in Google play store is more typical and comprehensive. Through the analysis of APP related data, it is helpful for domestic companies in the production and operation of app. The structure of the data in the table is as follows in figure1.

| A  | B                     | C        | D      | E       | F    | G        | H    | I     | J          | K          | L          | M           | N            |
|----|-----------------------|----------|--------|---------|------|----------|------|-------|------------|------------|------------|-------------|--------------|
| 1  | App                   | Category | Rating | Reviews | Size | Installs | Type | Price | Content Ra | Genres     | Last Updat | Current Ver | Android Ver  |
| 2  | Photo Edit ART_AND_   | 4.1      | 159    | 19M     |      | 10000    | Free | 0     | Everyone   | Art & Desi | 7-Jan-18   | 1.0.0       | 4.0.3 and up |
| 3  | Coloring b ART_AND_   | 3.9      | 967    | 14M     |      | 500000   | Free | 0     | Everyone   | Art & Desi | #####      | 2.0.0       | 4.0.3 and up |
| 4  | U Launche ART_AND_    | 4.7      | 87510  | 8.7M    |      | 5000000  | Free | 0     | Everyone   | Art & Desi | 1-Aug-18   | 1.2.4       | 4.0.3 and up |
| 5  | Sketch - D ART_AND_   | 4.5      | 215644 | 25M     |      | 50000000 | Free | 0     | Teen       | Art & Desi | 8-Jun-18   | Varies with | 4.2 and up   |
| 6  | Pixel Draw ART_AND_   | 4.3      | 967    | 2.8M    |      | 100000   | Free | 0     | Everyone   | Art & Desi | #####      | 1.1         | 4.4 and up   |
| 7  | Paper flow ART_AND_   | 4.4      | 167    | 5.6M    |      | 50000    | Free | 0     | Everyone   | Art & Desi | #####      | 1.2         | 3 and up     |
| 8  | Smoke Eff ART_AND_    | 3.8      | 178    | 19M     |      | 50000    | Free | 0     | Everyone   | Art & Desi | #####      | 1.1         | 4.0.3 and up |
| 9  | Infinite Pai ART_AND_ | 4.1      | 36815  | 29M     |      | 1000000  | Free | 0     | Everyone   | Art & Desi | #####      | 6.1.61.1    | 4.2 and up   |
| 10 | Garden C ART_AND_     | 4.4      | 13791  | 33M     |      | 1000000  | Free | 0     | Everyone   | Art & Desi | #####      | 2.9.2       | 3.0 and up   |
| 11 | Kids Paint ART_AND_   | 4.7      | 121    | 3.1M    |      | 10000    | Free | 0     | Everyone   | Art & Desi | 3-Jul-18   | 2.8         | 4.0.3 and up |
| 12 | Text on Ph ART_AND_   | 4.4      | 13880  | 28M     |      | 1000000  | Free | 0     | Everyone   | Art & Desi | #####      | 10.4        | 4.1 and up   |
| 13 | Name Art ART_AND_     | 4.4      | 8788   | 12M     |      | 1000000  | Free | 0     | Everyone   | Art & Desi | 31-Jul-18  | 1.0.15      | 4.0 and up   |
| 14 | Tattoo Nai ART_AND_   | 4.2      | 44829  | 20M     |      | 10000000 | Free | 0     | Teen       | Art & Desi | 2-Apr-18   | 3.8         | 4.1 and up   |
| 15 | Mandala CART_AND_     | 4.6      | 4326   | 21M     |      | 100000   | Free | 0     | Everyone   | Art & Desi | #####      | 10.4        | 4.4 and up   |
| 16 | 3D Color FART_AND_    | 4.4      | 1518   | 37M     |      | 100000   | Free | 0     | Everyone   | Art & Desi | 3-Aug-18   | 1.2.3       | 2.3 and up   |
| 17 | Photo Des ART_AND_    | 4.7      | 3632   | 5.5M    |      | 500000   | Free | 0     | Everyone   | Art & Desi | 31-Jul-18  | 3.1         | 4.1 and up   |
| 18 | 350 Diy RcART_AND_    | 4.5      | 27     | 17M     |      | 10000    | Free | 0     | Everyone   | Art & Desi | 7-Nov-17   | 1.2.3       | 3 and up     |
| 19 | FlipaClip - ART_AND_  | 4.3      | 194216 | 39M     |      | 5000000  | Free | 0     | Everyone   | Art & Desi | 3-Aug-18   | 2.2.5       | 4.0.3 and up |
| 20 | ibis Paint XART_AND_  | 4.6      | 224399 | 31M     |      | 10000000 | Free | 0     | Everyone   | Art & Desi | 30-Jul-18  | 5.5.4       | 4.1 and up   |
| 21 | Icon Mania ART_AND_   |          |        |         |      | 100000   | Free | 0     | Everyone   | Art & Desi | #####      | 1.1.1       | and up       |

Figure 1: Data Structure

The first step in data analysis is to have an overview of the dataset. This dataset records all the apps in Google App Store at that time and has 10840 rows with 13 columns. Thirteen attribute columns record the name, category, rating, application size, downloads and other information of these apps. After overviewing all the dataset, I find there are some missing values in the dataframe. The relationship between missing values and columns is shown in the table1 below. For the convenience of the final presentation, I have deleted all the rows that contain missing values. After the data preprocessing, the dataset still has 9359 rows.

## 1.2 Objectives

I choose Google-Play-Store dataset as the data source of my data visualization experiment. In this dataset I want to know the relation between the **Rating** and **Reviews**, the **Size of the App** and **Installs**, the **Price of the App** and **Total Number of the Apps** which the price is below this price.

I also want to know the relationship between the data in different app category. For example, I want to know the characteristics of the Apps which are belong to BUSINESS category. By dividing apps into several subclasses by category attributes, I can hold a global view of the information in the Google Play Store.

Therefor, I picked out the following columns from the original dataset: App, Category, Rating, Reviews, Size, Installs, Price, Content Rating.

Table 1: Relationship between Columns and Missing Values

| Column Name   | Missing Data Number |
|---------------|---------------------|
| App           | 0                   |
| Category      | 0                   |
| Rating        | 1474                |
| Reviews       | 0                   |
| Size          | 0                   |
| Installs      | 0                   |
| Type          | 1                   |
| Price         | 0                   |
| ContentRating | 0                   |
| Genres        | 1                   |
| LastUpdated   | 0                   |
| CurrentVer    | 8                   |
| AndroidVer    | 2                   |

## 1.3 Characteristics

The first thing I have to do is that I should know the characteristic of each attributes and the basic relationship between different attributes before plotting my data on dashboard. The dataprocess.py is the script which I write to fetch the attributes I need and use to process data aim to fit the frontend and can show in the dashboard. The characteristics of all the attributes I used are list here:

### 1.3.1 Category

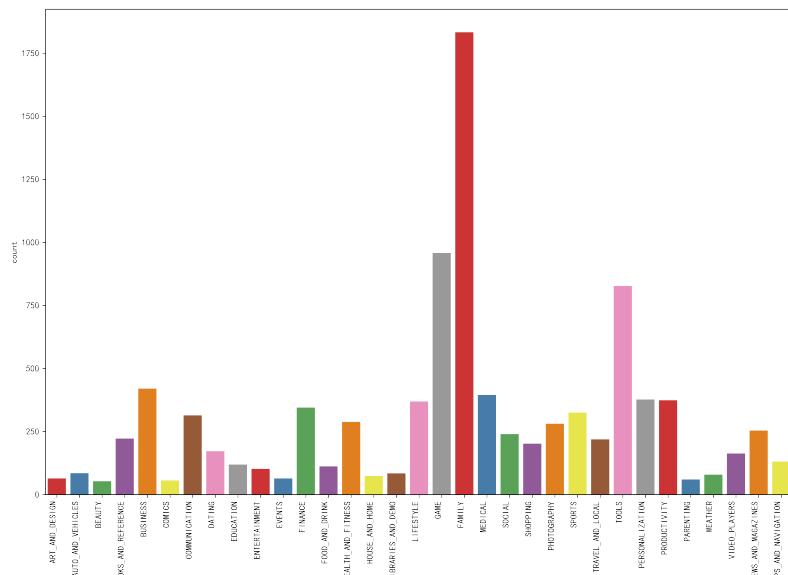


Figure 2: Count of App in Each Category

Category has 33 different values, and it can be seen that family and game have the largest number of apps.

### 1.3.2 Relationship

The following figure shows each attribute (Rating;Size;Installs;Reviews;Price) scatter plot of each other:

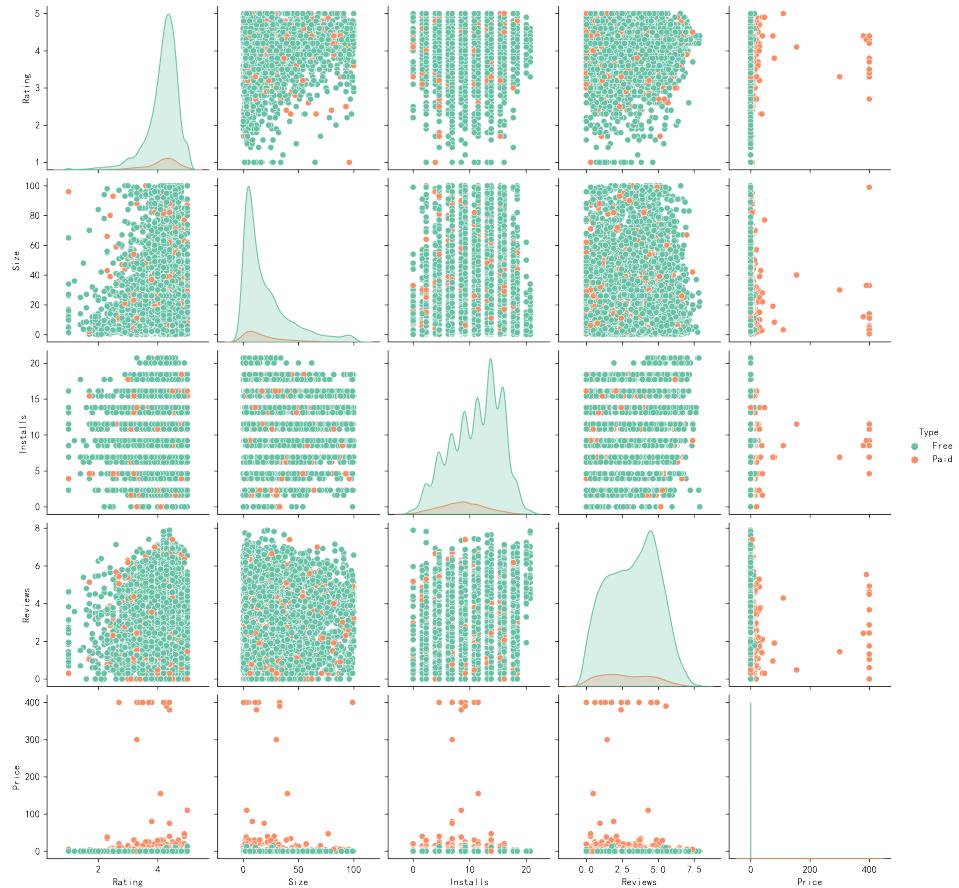


Figure 3: Relationship

### 1.3.3 Content Rating

Box chart, also known as box whisker chart, box chart or box line chart, is a statistical chart used to display a group of data dispersion. It is named for its shape like a box. It is also often used in various fields, common in quality management. It is mainly used to reflect the distribution characteristics of the original data, and can also compare the distribution characteristics of multiple groups of data. The drawing method of boxplot is: first, find out the maximum, minimum, median and two quartiles of a group of data; Then, connect the two quartiles and draw the box; Then connect the maximum and minimum with the box, and the median is in the middle of the box.

I can see a lot of information from the box diagram. Taking content rating as the classification standard and the rating as the ordinate, the free and paid box charts are drawn respectively, and the rating chart is as follows:

- It have six different values in attribute.
- I use pie graph to present the relation of content rating and other attributes and information in this section.

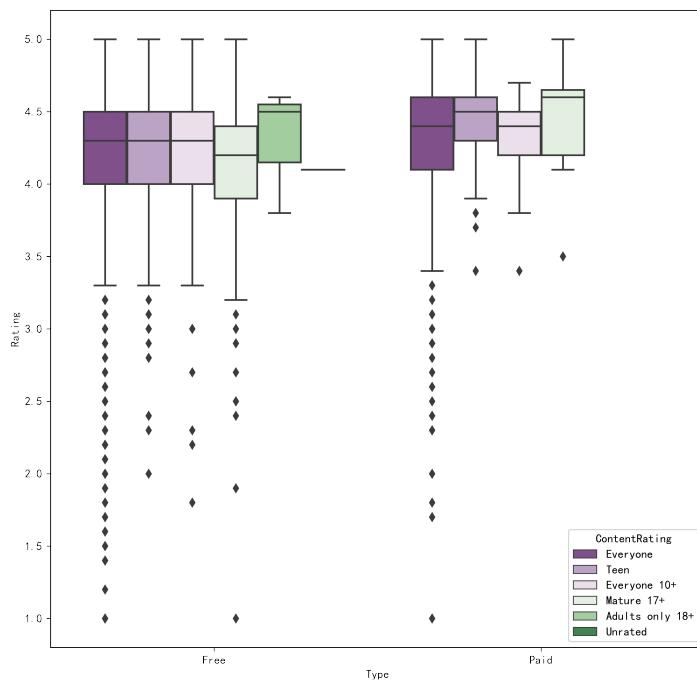


Figure 4: Content Rating with Box Chart

### 1.3.4 Price

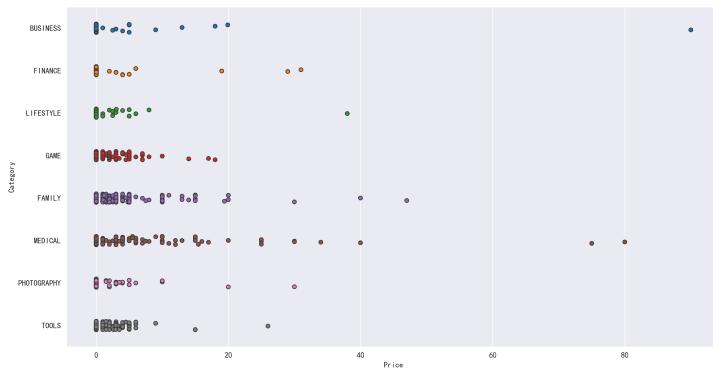


Figure 5: App Price Distribution

- 0, which means that the App is free for download. And this kind of value is specially.
- The price of pharmaceutical apps is very high, and some even reach \$80. What is novel is that the price of game apps is not high, almost no more than \$20.

## 2 Layout of Designed Dashboard

After I had a general understanding of the data, I started to design my dashboard. I use the sugar data visualization framework provided by Baidu to build my big data screen. I've just used three interactive charts to show my data: scatter, radiation and line column mix-up charts.

### 2.1 Interface Design

The interface overview of my data visualization platform is as show follows:

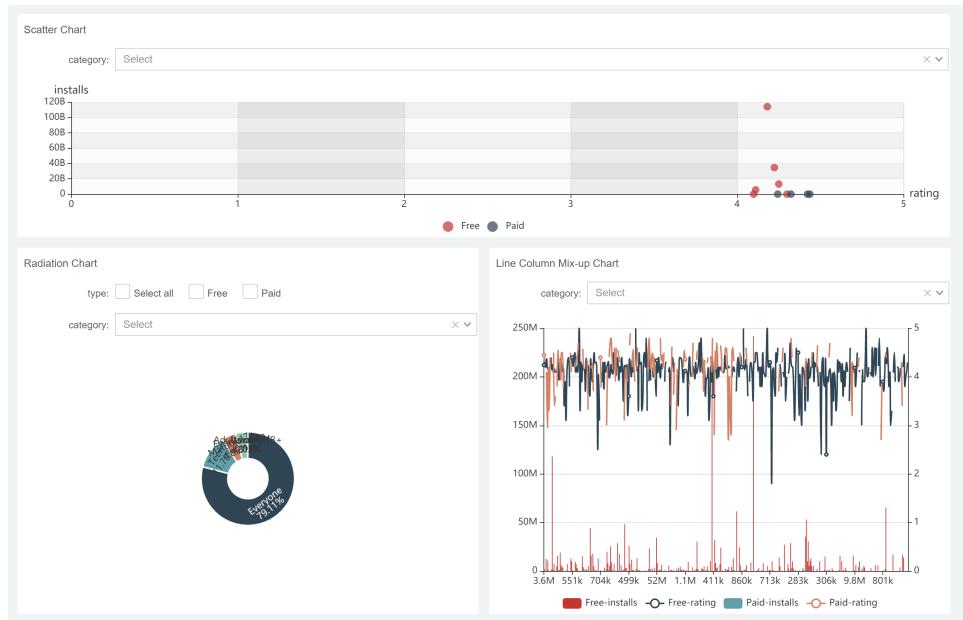


Figure 6: Data Dashboard Interface

The three charts are arranged as a triangle.

### 2.2 Scatter Chart

The abscissa of the scatter plot is the average of user ratings, the ordinate is the cumulative sum of user downloads, and the filter is the category of app. The two color tags represent paid apps and free apps.

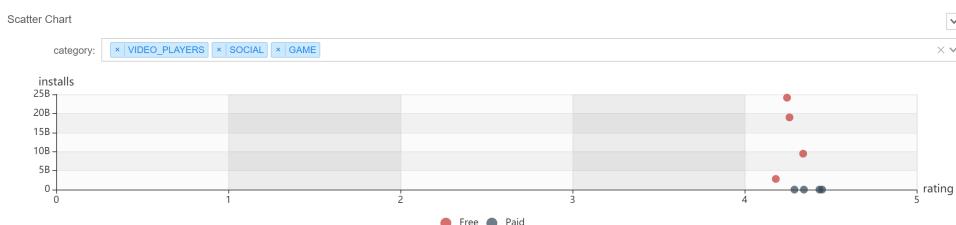


Figure 7: Scatter Chart

You can click the down button to start the drawer to select category of app you want.

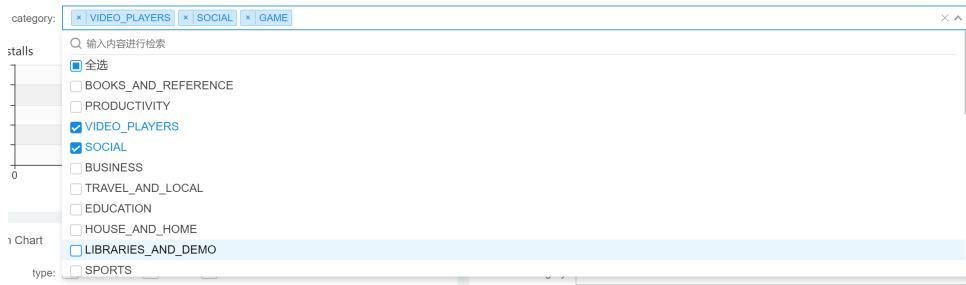


Figure 8: Use Filter

### 2.3 Radiation Chart

Radiation chart reflects the distribution of various content rating apps in different app category.

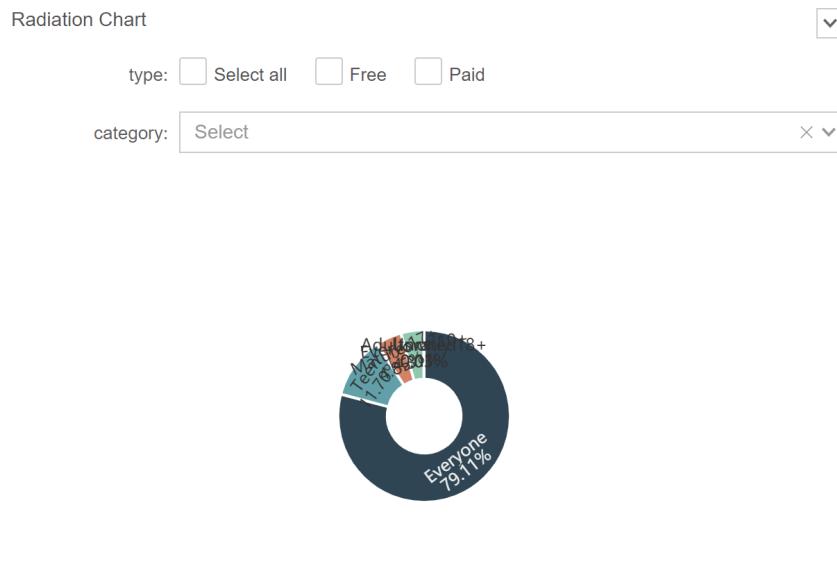


Figure 9: Radiation Chart

Similar to the scatter plot, you can use the drop-down drawer component to select filters and can use the check box to select app type.

The chart is interactive. You can move the mouse over a color block in the chart, and then the chart will display the details of the color block.



Figure 10: Interactive Function

## 2.4 Line Column Mix-up Chart

The x-axis of this mixed chart is the size of the app, and the vertical axis is the download amount and rating of the app. Similar to the scatter plot, you can use the drop-down drawer component to select filters.

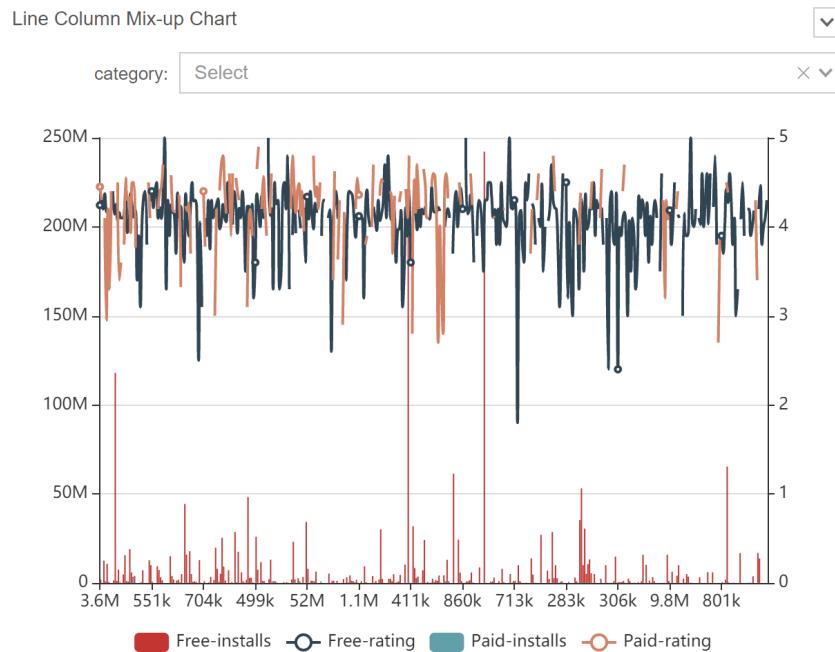


Figure 11: Line Column Mix-up Chart

Move the mouse over the chart to display the selected data.

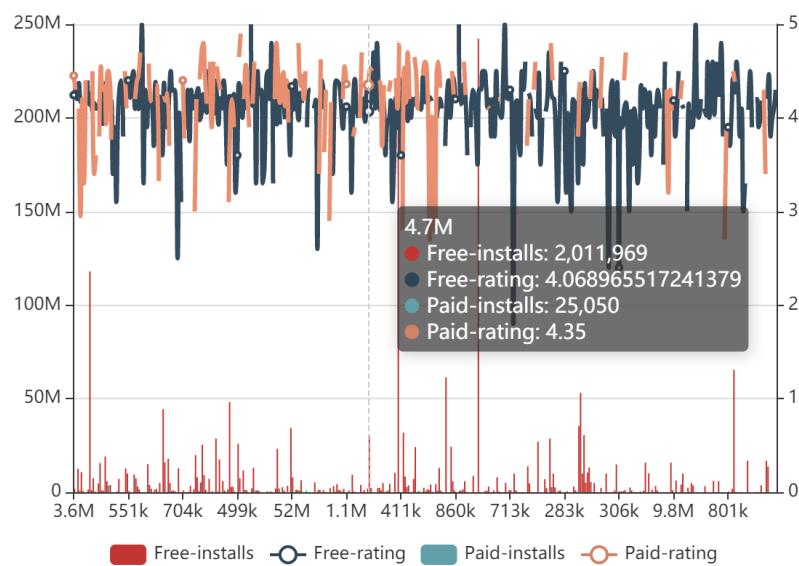


Figure 12: Interactive Function

## 2.5 Deployment

The data dashboard is deployed on a cloud server, and you can access it by scanning the QR code below.



Figure 13: QR Code

## References

- [1] Designing the User Interface: Strategies for Effective Human-Computer Interaction, 6th edition,Ben Shneiderman,Catherine Plaisant,Maxine Cohen
- [2] Baidu Data Visualization Sugar, <https://cloud.baidu.com/doc/SUGAR/index.html>