Google Play Store Data set

Andrew Wang

03/09/2020

```r
require(dplyr)
```

```
## Loading required package: dplyr

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
# Read in data
data = read.csv("googleplaystore.csv")
```

Cleaning the data

```r
# Remove + in Installs
data$Installs = gsub("[+,]", replacement = "", x = data$Installs)

# Remove app in row 10473 because data is in wrong columns (and is an insignificant app)
bad_dat = data[10473,]
data = data[-10473,]

# Making reviews column numeric
data$Reviews = as.numeric(as.character(data$Reviews))

# Making Installs column numeric
data$Installs = as.numeric(data$Installs)

# Making Installs column unit 1000
data$Installs = data$Installs/1000

# Formatting the date
```

```r
data$Last.Updated = as.Date(as.POSIXct(strptime(data$Last.Updated, "%B %d, %Y", tz="")))

# Find out why apps have NA for a rating
NA_apps = data[is.na(data$Rating),]
```

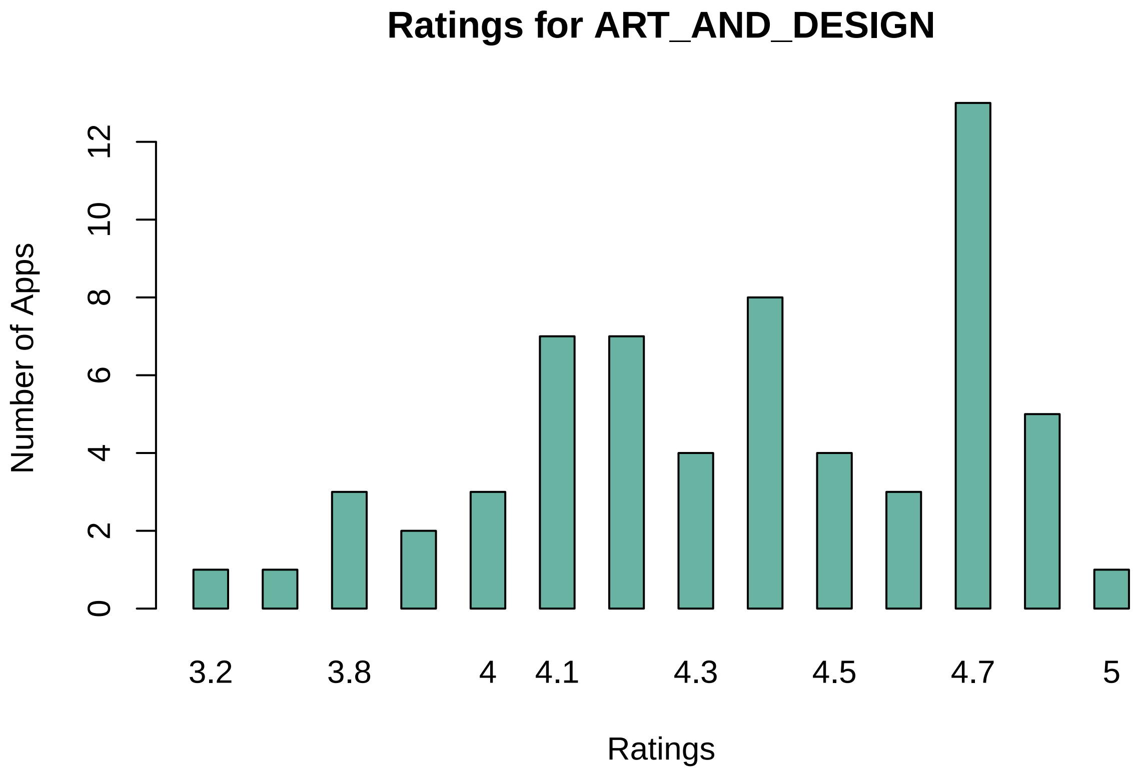
```r
# Type in category of interest
category = "ART_AND_DESIGN"
data_subset = data[data$Category == category,]

# Show number of apps in this category
num_apps = sum(data$Category == category)
sprintf("Number of apps in this category is %d", num_apps)
```

```
## [1] "Number of apps in this category is 65"
```

```r
# Filter out apps that have not gotten an update within the last 3 months
latestupdates = as.Date("2018-05-07")
active_apps = data_subset[data_subset$Last.Updated > latestupdates,]
```

```r
# Create a plot of the category by rating e.g. how many apps in each rating category...
ratings = na.omit(data_subset) %>% count(Rating)
bar_p1 = barplot(ratings$n,
                 names = ratings$Rating,
                 width = 5,
                 space = 1,
                 xlab = "Ratings",
                 ylab = "Number of Apps",
                 main = sprintf("Ratings for %s",category),
                 col="#69b3a2")
```

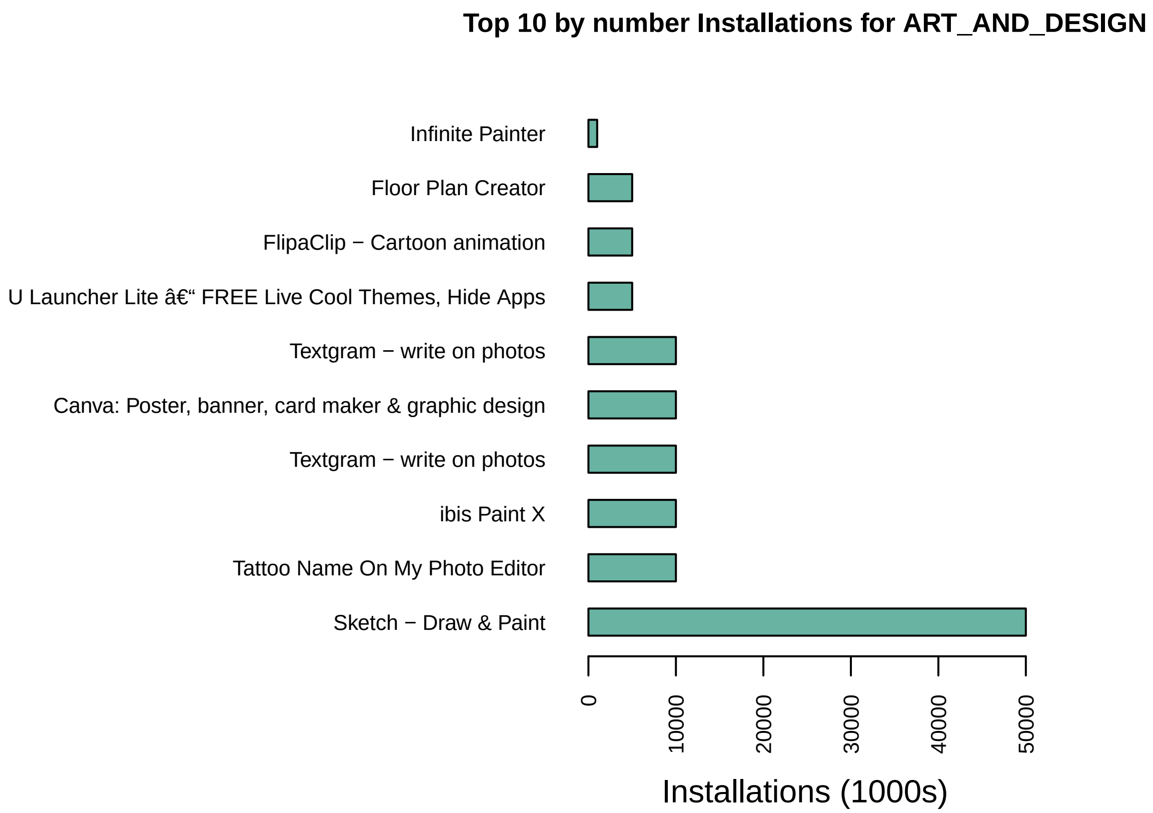


```r
# Filter out apps with less than 200 reviews and plot category by rating e.g. reliable apps with a good
reliable_apps = data_subset[data_subset$Reviews > 200 & data_subset$Rating >= 4.5,]
reliable_apps = reliable_apps[order(reliable_apps$Rating, decreasing = TRUE),]

# Get the top 10 reliable apps (by rating and only using apps with over 200 reviews)
top_10 = reliable_apps[1:10,]
top_10
```

```
##                                                        App       Category Rating
## 3983                                    Cardi B Wallpaper ART_AND_DESIGN    4.8
## 4760    X Launcher Pro - IOS Style Theme & Control Center ART_AND_DESIGN    4.8
## 3    U Launcher Lite â\200" FREE Live Cool Themes, Hide Apps ART_AND_DESIGN    4.7
## 17           Photo Designer - Write your name with shapes ART_AND_DESIGN    4.7
## 23                Superheroes Wallpapers | 4K Backgrounds ART_AND_DESIGN    4.7
## 27                          Colorfit - Drawing & Coloring ART_AND_DESIGN    4.7
## 35                                        I Creative Idea ART_AND_DESIGN    4.7
## 37         UNICORN - Color By Number & Pixel Art Coloring ART_AND_DESIGN    4.7
## 46     Canva: Poster, banner, card maker & graphic design ART_AND_DESIGN    4.7
## 4750   X Launcher: With OS11 Style Theme & Control Center ART_AND_DESIGN    4.7
##      Reviews Size Installs Type Price Content.Rating                  Genres
## 3983     253 3.7M       50 Free     0       Everyone            Art & Design
## 4760    1216 8.6M       10 Paid $1.99       Everyone            Art & Design
## 3      87510 8.7M     5000 Free     0       Everyone            Art & Design
## 17      3632 5.5M      500 Free     0       Everyone            Art & Design
## 23      7699 4.2M      500 Free     0   Everyone 10+            Art & Design
## 27     20260  25M      500 Free     0       Everyone Art & Design;Creativity
## 35       353 4.2M       10 Free     0           Teen            Art & Design
## 37      8145  24M      500 Free     0       Everyone Art & Design;Creativity
## 46    174531  24M    10000 Free     0       Everyone            Art & Design
## 4750    5754 4.4M      100 Free     0       Everyone            Art & Design
##      Last.Updated Current.Ver  Android.Ver
## 3983   2017-10-31       1.0.0   4.0 and up
## 4760   2018-06-25       1.0.0   4.1 and up
## 3      2018-07-31       1.2.4 4.0.3 and up
## 17     2018-07-30         3.1   4.1 and up
## 23     2018-07-11     2.2.6.2 4.0.3 and up
## 27     2017-10-10       1.0.8 4.0.3 and up
## 35     2018-04-26         1.6   4.1 and up
## 37     2018-08-01       1.0.9   4.4 and up
## 46     2018-07-30       1.6.1   4.1 and up
## 4750   2018-07-29       2.1.2   4.1 and up
```

```
# Get the top 10 apps by number of installations
sorted_by_installations = data_subset[order(data_subset$Installs, decreasing = TRUE),]
par(mar=c(4,17,4,4))
bar_p2 = barplot(sorted_by_installations$Installs[1:10],
                 names = sorted_by_installations$App[1:10],
                 width = 5,
                 space = 1,
                 xlab = "Installations (1000s)",
                 main = sprintf("Top 10 by number Installations for %s",category),
                 las = 2, # Makes x axis labels turn 90 degrees
                 horiz = T,
                 col="#69b3a2",
                 cex.names = 0.7,
                 cex.axis = 0.7,
                 cex.main = 0.8)
```


Top 10 by number Installations for ART_AND_DESIGN
Infinite Painter
Floor Plan Creator
FlipaClip − Cartoon animation
U Launcher Lite â€“ FREE Live Cool Themes, Hide Apps
Textgram − write on photos
Canva: Poster, banner, card maker & graphic design
Textgram − write on photos
ibis Paint X
Tattoo Name On My Photo Editor
Sketch − Draw & Paint
0
10000
20000
30000
40000
50000
Installations (1000s)