



Google Play Store Analysis in R



Dr. Shilpa Balan

CIS 5250 – Visual Analytics

College of Business and
Economics

California State
University, Los Angeles



By:
Ninelia
Talverdi



Fall
2020

Contents

Introduction.....	3
Data Description.....	4
Data Cleaning.....	6
Analysis and Visualizations.....	14
Statistical Summary and Functions.....	29
Conclusion.....	38

Introduction

Google Play, which was formerly known as Android Market, is a digital distribution service that is both operated and developed by Google. It serves as the official app store for the Android operating system, allowing users to browse and download applications developed with the Android software development kit (SDK) and published through Google. Google Play also serves as a digital media store, offering music, books, movies, and television programs.¹

The goal of this project is to provide insights about Android applications to understand users' expectations better and thus help developers improve the applications. In my analysis, I tried to find out the success factors for the apps and visualize the data for the following questions:

- What is the app distribution by type?
- What are the top 10 installed apps?
- What are the top 5 reviewed categories?
- Which age category has rated the apps and what is the association between content rating and app rating?

¹ MITAOE, Developer Students Club. "GOOGLE PLAY Google Play (Previously Android Market) Is a Digital Distribution Service Operated and..." Medium, Medium, 29 Apr. 2019, medium.com/@dsc_3204/google-play-google-play-previously-android-market-is-a-digital-distribution-service-operated-and-7e8b3e7aaeec.

Data Description

The dataset used for this analysis is “*googleplaystore.csv*” dataset which is downloaded from Kaggle.

Dataset URL: <https://www.kaggle.com/lava18/google-play-store-apps>

It consists of in total of 10841 observations (number of apps) and 13 variables (number of characteristics for each app). The Google Play Store data includes information such as app, category, ratings, reviews, type, price, version and etc.

The below table describes the dataset in details:

Field Name	Type	Description
App	char	Name of the application.
Category	char	Category the app belongs to.
Rating	numeric	Overall user rating of the app. (Out of 5)
Reviews	char	Number of user reviews for the app.
Size	char	Size of the app.
Installs	char	Number of users that download/install the app.
Type	char	Type of the app. Paid or Free
Price	char	Cost of the app.
Content Rating	char	Age group the app is targeted at - Children / Mature 21+ / Adult.
Genres	char	Genre of the app. An app can belong to multiple genres (apart from its main category). e.g. a musical family game will belong to Music, Game, and Family genres.
Last Updated	char	Date when the app was last updated on Play Store.

Current Ver	char	Current version of the app available on Play Store.
Android Ver	char	Minimum required Android version.

Below is a sample screenshot of the dataset in RStudio:

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content.Rating	Genres	Last.Updated	Current.Ver	Android.Ver
1 Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
2 Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up
3 U Launcher Lite â€ FREE Live Cool Themes, Hide Apps	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up
4 Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Varies with device	4.2 and up
5 Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4 and up
6 Paper flowers instructions	ART_AND_DESIGN	4.4	167	5.6M	50,000+	Free	0	Everyone	Art & Design	March 26, 2017	1.0	2.3 and up
7 Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	178	19M	50,000+	Free	0	Everyone	Art & Design	April 26, 2018	1.1	4.0.3 and up
8 Infinite Painter	ART_AND_DESIGN	4.1	36815	29M	1,000,000+	Free	0	Everyone	Art & Design	June 14, 2018	6.1.61.1	4.2 and up
9 Garden Coloring Book	ART_AND_DESIGN	4.4	13791	33M	1,000,000+	Free	0	Everyone	Art & Design	September 20, 2017	2.9.2	3.0 and up
10 Kids Paint Free - Drawing Fun	ART_AND_DESIGN	4.7	121	3.1M	10,000+	Free	0	Everyone	Art & Design;Creativity	July 3, 2018	2.8	4.0.3 and up

```
> setwd("C:/Users/Ninel/Desktop/School/CSULA/CIS 5250/R Labs/DATA")
> ds_play_store<-read.csv("googleplaystore.csv",header = T, sep = ",")
> View(ds_play_store)
> dim(ds_play_store)
[1] 10841 13
> head(ds_play_store)
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content.Rating	Genres	Last.Updated	Current.Ver	Android.Ver
1	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
2	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up
3	U Launcher Lite â€ FREE Live Cool Themes, Hide Apps	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up
4	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Varies with device	4.2 and up
5	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4 and up
6	Paper flowers instructions	ART_AND_DESIGN	4.4	167	5.6M	50,000+	Free	0	Everyone	Art & Design	March 26, 2017	1.0	2.3 and up

```
> str(ds_play_store)
'data.frame': 10841 obs. of 13 variables:
 $ App : chr "Photo Editor & Candy Camera & Grid & ScrapBook" "Coloring book moana" "U Launcher Lite â€200" FREE Live Cool Theme
s, Hide Apps" "Sketch - Draw & Paint" ...
 $ Category : chr "ART_AND_DESIGN" "ART_AND_DESIGN" "ART_AND_DESIGN" "ART_AND_DESIGN" ...
 $ Rating : num 4.1 3.9 4.7 4.5 4.3 4.4 3.8 4.1 4.4 4.7 ...
 $ Reviews : chr "159" "967" "87510" "215644" ...
 $ Size : chr "19M" "14M" "8.7M" "25M" ...
 $ Installs : chr "10,000+" "500,000+" "5,000,000+" "50,000,000+" ...
 $ Type : chr "Free" "Free" "Free" "Free" ...
 $ Price : chr "0" "0" "0" "0" ...
 $ Content.Rating: chr "Everyone" "Everyone" "Everyone" "Teen" ...
 $ Genres : chr "Art & Design" "Art & Design;Pretend Play" "Art & Design" "Art & Design" ...
 $ Last.Updated : chr "January 7, 2018" "January 15, 2018" "August 1, 2018" "June 8, 2018" ...
 $ Current.Ver : chr "1.0.0" "2.0.0" "1.2.4" "Varies with device" ...
 $ Android.Ver : chr "4.0.3 and up" "4.0.3 and up" "4.0.3 and up" "4.2 and up" ...
```

```
> setwd("C:/Users/Ninel/Desktop/School/CSULA/CIS 5250/R Labs/DATA")
> ds_play_store<-read.csv("googleplaystore.csv",header = T, sep = ",")
> View(ds_play_store)
> dim(ds_play_store)
[1] 10841 13
> head(ds_play_store)
> str(ds_play_store)
```

Note: I changed the data type of some of the variables during the “Date Cleaning” process.

Data Cleaning

After observing the raw data, I found out there are almost 1000 missing (i.e. “NAN”, blank) and invalid values for different variables. Also, I realized that there are few rows which are duplicates of other. So, I cleaned the data using the following steps in RStudio:

1. Dropping out of scope data

Since our goal was to analyze the factors that affect app success, I decided to remove “Current Ver” and “Android Ver” columns from the dataset since they were not giving valuable information.

Before cleaning:

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content.Rating	Genres	Last.Updated	Current.Ver	Android.Ver
1 Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
2 Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up
3 U Launcher Lite â€” FREE Live Cool Themes, Hide Apps	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up
4 Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Varies with device	4.2 and up
5 Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4 and up
6 Paper flowers instructions	ART_AND_DESIGN	4.4	167	5.6M	50,000+	Free	0	Everyone	Art & Design	March 26, 2017	1.0	2.3 and up
7 Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	178	19M	50,000+	Free	0	Everyone	Art & Design	April 26, 2018	1.1	4.0.3 and up
8 Infinite Painter	ART_AND_DESIGN	4.1	36815	29M	1,000,000+	Free	0	Everyone	Art & Design	June 14, 2018	6.1.61.1	4.2 and up
9 Garden Coloring Book	ART_AND_DESIGN	4.4	13791	33M	1,000,000+	Free	0	Everyone	Art & Design	September 20, 2017	2.9.2	3.0 and up
10 Kids Paint Free - Drawing Fun	ART_AND_DESIGN	4.7	121	3.1M	10,000+	Free	0	Everyone	Art & Design;Creativity	July 3, 2018	2.8	4.0.3 and up
11 Text on Photo - Fontee	ART_AND_DESIGN	4.4	13880	28M	1,000,000+	Free	0	Everyone	Art & Design	October 27, 2017	1.0.4	4.1 and up
12 Name Art Photo Editor - Focus n Filters	ART_AND_DESIGN	4.4	8788	12M	1,000,000+	Free	0	Everyone	Art & Design	July 31, 2018	1.0.15	4.0 and up
13 Tattoo Name On My Photo Editor	ART_AND_DESIGN	4.2	44829	20M	10,000,000+	Free	0	Teen	Art & Design	April 2, 2018	3.8	4.1 and up
14 Mandala Coloring Book	ART_AND_DESIGN	4.6	4326	21M	100,000+	Free	0	Everyone	Art & Design	June 26, 2018	1.0.4	4.4 and up
15 3D Color Pixel by Number - Sandbox Art Coloring	ART_AND_DESIGN	4.4	1518	37M	100,000+	Free	0	Everyone	Art & Design	August 3, 2018	1.2.3	2.3 and up
16 Learn To Draw Kawaii Characters	ART_AND_DESIGN	3.2	55	2.7M	5,000+	Free	0	Everyone	Art & Design	June 6, 2018	NaN	4.2 and up
17 Photo Designer - Write your name with shapes	ART_AND_DESIGN	4.7	3632	5.5M	500,000+	Free	0	Everyone	Art & Design	July 31, 2018	3.1	4.1 and up
18 350 Diy Room Decor Ideas	ART_AND_DESIGN	4.5	27	17M	10,000+	Free	0	Everyone	Art & Design	November 7, 2017	1.0	2.3 and up
19 FlipaClip - Cartoon animation	ART_AND_DESIGN	4.3	194216	39M	5,000,000+	Free	0	Everyone	Art & Design	August 3, 2018	2.2.5	4.0.3 and up
20 ibis Paint X	ART_AND_DESIGN	4.6	224399	31M	10,000,000+	Free	0	Everyone	Art & Design	July 30, 2018	5.5.4	4.1 and up

```
> ds_play_store<-ds_play_store[,-(12:-13)]
> View(ds_play_store)
> dim(ds_play_store)
[1] 10841 11
> head(ds_play_store)
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content.Rating	Genres
1	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design
2	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play
3	U Launcher Lite â€” FREE Live Cool Themes, Hide Apps	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design
4	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design
5	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity
6	Paper flowers instructions	ART_AND_DESIGN	4.4	167	5.6M	50,000+	Free	0	Everyone	Art & Design

```
Last.Updated
1 January 7, 2018
2 January 15, 2018
3 August 1, 2018
4 June 8, 2018
5 June 20, 2018
6 March 26, 2017
```

```
> ds_play_store<-ds_play_store[,-(12:-13)]
```

```
> View(ds_play_store)
```

```
> dim(ds_play_store)
```

```
[1] 10841 11
```

The “Current Ver” and “Android Ver” variables were deleted from the dataset.

After cleaning:

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content.Rating	Genres	Last.Updated
1 Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018
2 Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018
3 U Launcher Lite #€° FREE Live Cool Themes, Hide Apps	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018
4 Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018
5 Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018
6 Paper flowers instructions	ART_AND_DESIGN	4.4	167	5.6M	50,000+	Free	0	Everyone	Art & Design	March 26, 2017
7 Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	178	19M	50,000+	Free	0	Everyone	Art & Design	April 26, 2018
8 Infinite Painter	ART_AND_DESIGN	4.1	36815	29M	1,000,000+	Free	0	Everyone	Art & Design	June 14, 2018
9 Garden Coloring Book	ART_AND_DESIGN	4.4	13791	33M	1,000,000+	Free	0	Everyone	Art & Design	September 20, 2017
10 Kids Paint Free - Drawing Fun	ART_AND_DESIGN	4.7	121	3.1M	10,000+	Free	0	Everyone	Art & Design;Creativity	July 3, 2018

2. Removing “NA”/ “NAN”, blank and invalid data

1) There were some “NAN” values for “Rating” and “Type” variables. To create consistency in data for analysis, I decided to remove them using “omit” function.

First I checked the “NA” values:

```
> is.na(ds_play_store)
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content.Rating	Genres	Last.Updated
[1,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[2,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[3,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[4,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[5,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[6,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[7,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[8,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[9,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[10,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[11,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[12,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[13,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[14,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[15,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[16,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[17,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[18,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[19,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[20,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[21,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[22,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[23,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[24,]	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[25,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE

```
> is.na(ds_play_store)
```

Before cleaning:

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content.Rating	Genres	Last.Updated
1021 Gametime - Tickets to Sports, Concerts, Theater	EVENTS	4.5	8800	24M	1,000,000+	Free	0	Everyone	Events	July 17, 2018
1022 PTI Flex Maker, Photo Frame Editor & Songs 2018	EVENTS	4.6	99	13M	10,000+	Free	0	Everyone	Events	July 6, 2018
1023 Wipe out	EVENTS	4.7	30	Varies with device	500+	Free	0	Teen	Events	May 23, 2018
1024 Birthdays & Other Events	EVENTS	4.3	456	2.8M	50,000+	Free	0	Everyone	Events	June 21, 2018
1025 Sarajevo Film Festival - Official	EVENTS	NaN	6	9.7M	100+	Free	0	Everyone 10+	Events	July 31, 2018
1026 Summer Madness	EVENTS	NaN	4	1.4M	100+	Free	0	Everyone	Events	July 26, 2018
1027 Events High - Meet Your City!	EVENTS	4.2	3200	Varies with device	100,000+	Free	0	Everyone 10+	Events	June 23, 2018
1028 vide-greniers.org	EVENTS	4.3	5839	3.5M	100,000+	Free	0	Everyone	Events	November 8, 2016
1029 SUMMER SONIC app	EVENTS	5.0	4	61M	500+	Free	0	Everyone	Events	July 24, 2018
1030 Quake & Volcanoes: 3D Globe of Volcanic Eruptions	EVENTS	4.4	663	6.0M	100,000+	Free	0	Everyone	Events	October 14, 2017
1031 Prosperity	EVENTS	5.0	16	2.3M	100+	Free	0	Everyone	Events	July 9, 2018
1032 Picktrainer: India's largest photo contest app	EVENTS	4.5	1065	9.0M	100,000+	Free	0	Everyone	Events	May 31, 2018
1033 AAS-IN-ASIA 2018	EVENTS	NaN	3	11M	100+	Free	0	Everyone	Events	July 4, 2018
1034 Moot Peru 2018	EVENTS	NaN	13	27M	1,000+	Free	0	Everyone	Events	June 28, 2018
1035 FM News	EVENTS	NaN	46	3.6M	5,000+	Free	0	Everyone	Events	July 4, 2018
1036 Goldstar: Live Event Tickets	EVENTS	4.5	1953	Varies with device	100,000+	Free	0	Teen	Events	August 1, 2018
1037 Emmabodafestivalen	EVENTS	4.8	12	29M	1,000+	Free	0	Everyone 10+	Events	June 28, 2018
1038 KudaGo - things to do in NY	EVENTS	4.4	4298	4.4M	100,000+	Free	0	Everyone	Events	April 27, 2017
1039 Mindvalley U Tallinn 2018	EVENTS	5.0	1	21M	100+	Free	0	Everyone	Events	July 3, 2018
1040 Rockmaraton Info	EVENTS	NaN	49	13M	1,000+	Free	0	Everyone	Events	July 8, 2018
1041 Freitas Auctioneer Official	EVENTS	3.7	100	27M	50,000+	Free	0	Everyone	Events	December 15, 2017
1042 mobLee Events	EVENTS	NaN	11	11M	5,000+	Free	0	Everyone	Events	May 30, 2018
1043 Xceed - Clubs, DJs, Festivals & Tickets	EVENTS	4.1	399	24M	100,000+	Free	0	Everyone	Events	August 3, 2018
1044 AMM Events & CPD	EVENTS	NaN	5	6.1M	100+	Free	0	Everyone	Events	August 4, 2018
1045 Arab Halls - For Wedding & Events	EVENTS	4.8	28	6.3M	1,000+	Free	0	Everyone	Events	May 20, 2018

```
> ds_play_store<-na.omit(ds_play_store)
> sum(is.na(ds_play_store))
[1] 0
> dim(ds_play_store)
[1] 9367 11
> View(ds_play_store)
```

```
> ds_play_store<-na.omit(ds_play_store)
> sum(is.na(ds_play_store))
[1] 0
> dim(ds_play_store)
[1] 9367 11
> View(ds_play_store)
```

The “NAN” and blank values were removed.

After cleaning:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content.Rating	Genres	Last.Updated
1021	Gametime - Tickets to Sports, Concerts, Theater	EVENTS	4.5	8800	24M	1,000,000+	Free	0	Everyone	Events	July 17, 2018
1022	PTI Flex Maker, Photo Frame Editor & Songs 2018	EVENTS	4.6	99	13M	10,000+	Free	0	Everyone	Events	July 6, 2018
1023	Wipe out	EVENTS	4.7	30	Varies with device	500+	Free	0	Teen	Events	May 23, 2018
1024	Birthdays & Other Events	EVENTS	4.3	456	2.8M	50,000+	Free	0	Everyone	Events	June 21, 2018
1027	Events High - Meet Your City!	EVENTS	4.2	3200	Varies with device	100,000+	Free	0	Everyone 10+	Events	June 23, 2018
1028	vide-greniers.org	EVENTS	4.3	5839	3.5M	100,000+	Free	0	Everyone	Events	November 8, 2016
1029	SUMMER SONIC app	EVENTS	5.0	4	61M	500+	Free	0	Everyone	Events	July 24, 2018
1030	Quake & Volcanoes: 3D Globe of Volcanic Eruptions	EVENTS	4.4	663	6.0M	100,000+	Free	0	Everyone	Events	October 14, 2017
1031	Prosperity	EVENTS	5.0	16	2.3M	100+	Free	0	Everyone	Events	July 9, 2018
1032	Picktrainer: India's largest photo contest app	EVENTS	4.5	1065	9.0M	100,000+	Free	0	Everyone	Events	May 31, 2018
1036	Goldstar: Live Event Tickets	EVENTS	4.5	1953	Varies with device	100,000+	Free	0	Teen	Events	August 1, 2018
1037	Emmabodafestivalen	EVENTS	4.8	12	29M	1,000+	Free	0	Everyone 10+	Events	June 28, 2018
1038	KudaGo - things to do in NY	EVENTS	4.4	4298	4.4M	100,000+	Free	0	Everyone	Events	April 27, 2017
1039	Mindvalley U Tallinn 2018	EVENTS	5.0	1	21M	100+	Free	0	Everyone	Events	July 3, 2018
1041	Freitas Auctioneer Official	EVENTS	3.7	100	27M	50,000+	Free	0	Everyone	Events	December 15, 2017
1043	Xceed - Clubs, DJs, Festivals & Tickets	EVENTS	4.1	399	24M	100,000+	Free	0	Everyone	Events	August 3, 2018
1045	Arab Halls - For Wedding & Events	EVENTS	4.8	28	6.3M	1,000+	Free	0	Everyone	Events	May 20, 2018

2) There was a row (row #10473) that had blank value for “Content Rating”, invalid values for “Type”, “Rating” and “Genres” variables. So, I decided to delete that row.

Before cleaning:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content.Rating	Genres	Last.Updated
10473	Life Made Wi-Fi Touchscreen Photo Frame	1.9	19.0	3.0M	1,000+	Free	0	Everyone		February 11, 2018	1.0.19

```
> ds_play_store<-ds_play_store[!(ds_play_store$Type=="0"),]
> dim(ds_play_store)
[1] 9366 11
> View(ds_play_store)
```

```
>ds_play_store<-ds_play_store[!(ds_play_store$Type=="0"),]
> dim(ds_play_store)
[1] 9366 11
> View(ds_play_store)
```

After cleaning:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content.Rating	Genres	Last.Updated
10470	TownWiFi Wi-Fi Everywhere	COMMUNICATION	3.9	2372	58M	500,000+	Free	0	Everyone	Communication	August 2, 2018
10471	Jazz Wi-Fi	COMMUNICATION	3.4	49	4.0M	10,000+	Free	0	Everyone	Communication	February 10, 2017
10472	Xposed Wi-Fi-Pwd	PERSONALIZATION	3.5	1042	404k	100,000+	Free	0	Everyone	Personalization	August 5, 2014
10474	osmino Wi-Fi free WiFi	TOOLS	4.2	134203	4.1M	10,000,000+	Free	0	Everyone	Tools	August 7, 2018
10475	Sat-Fi Voice	COMMUNICATION	3.4	37	14M	1,000+	Free	0	Everyone	Communication	November 21, 2014
10476	Wi-Fi Visualizer	TOOLS	3.9	132	2.6M	50,000+	Free	0	Everyone	Tools	May 17, 2017
10477	Lennox iComfort Wi-Fi	LIFESTYLE	3.0	552	7.6M	50,000+	Free	0	Everyone	Lifestyle	March 22, 2017
10478	Sci-Fi Sounds and Ringtones	PERSONALIZATION	3.6	128	11M	10,000+	Free	0	Everyone	Personalization	September 27, 2017
10479	Sci Fi Sounds	FAMILY	3.2	4	8.0M	1,000+	Free	0	Everyone	Entertainment	November 2, 2017
10480	Free Wi-fi HotspoT	COMMUNICATION	4.1	382	2.3M	50,000+	Free	0	Everyone	Communication	July 20, 2018

3. Deleting duplicate data

I noticed there were duplicate values in the dataset, so I decided to get rid of them by using “unique” function.

Before cleaning:

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content.Rating	Genres	Last.Updated
10253 1. FC KÄfln App	SPORTS	4.6	2019	41M	100,000+	Free	0	Everyone	Sports	July 20, 2018
1394 10 Best Foods for You	HEALTH_AND_FITNESS	4.0	2490	3.8M	500,000+	Free	0	Everyone 10+	Health & Fitness	February 17, 2017
1408 10 Best Foods for You	HEALTH_AND_FITNESS	4.0	2490	3.8M	500,000+	Free	0	Everyone 10+	Health & Fitness	February 17, 2017
8484 10 Minutes a Day Times Tables	FAMILY	4.1	681	48M	100,000+	Free	0	Everyone	Education	July 3, 2014
7739 10 WPM Amateur ham radio CW Morse code trainer	COMMUNICATION	3.5	10	3.8M	100+	Paid	\$1.49	Everyone	Communication	May 12, 2018
8220 10,000 Quotes DB (Premium)	BOOKS_AND_REFERENCE	4.1	70	3.5M	500+	Paid	\$0.99	Everyone	Books & Reference	August 30, 2013
9402 100 Doors of Revenge	FAMILY	4.1	105766	48M	10,000,000+	Free	0	Teen	Puzzle	May 24, 2018
4031 100+ C Programs	FAMILY	4.2	20	1.6M	5,000+	Free	0	Everyone	Education	January 3, 2017
8704 100000+ Messages - DP, Status, Jokes & GIF 2018	LIFESTYLE	3.7	121	3.8M	10,000+	Free	0	Mature 17+	Lifestyle	December 29, 2017
4027 101 C Programming Problems	FAMILY	4.6	498	5.0M	50,000+	Free	0	Everyone	Education	October 15, 2017
221 104 Looking for a job - looking for a job, looking for a job, L...	BUSINESS	4.4	74359	25M	1,000,000+	Free	0	Everyone	Business	July 27, 2018
2695 11st	SHOPPING	3.8	48732	20M	10,000,000+	Free	0	Everyone	Shopping	July 31, 2018
5345 12 Step Meditations & Sober Prayers AA NA AL-ANON	LIFESTYLE	4.7	759	15M	50,000+	Free	0	Everyone 10+	Lifestyle	January 22, 2018
6761 17th Edition Cable Sizer	BOOKS_AND_REFERENCE	4.4	47	1.4M	1,000+	Paid	\$3.08	Everyone	Books & Reference	May 27, 2016
2323 1800 Contacts - Lens Store	MEDICAL	4.7	23160	26M	1,000,000+	Free	0	Everyone	Medical	July 27, 2018
2544 1800 Contacts - Lens Store	MEDICAL	4.7	23160	26M	1,000,000+	Free	0	Everyone	Medical	July 27, 2018

```
> ds_play_store<-unique(ds_play_store)
> dim(ds_play_store)
[1] 8892 11
> View(ds_play_store)
```

```
> ds_play_store<-unique(ds_play_store)

> dim(ds_play_store)

[1] 8892 11

> View(ds_play_store)
```

After cleaning:

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content.Rating	Genres	Last.Updated
10253 1. FC KÄfln App	SPORTS	4.6	2019	41M	100,000+	Free	0	Everyone	Sports	July 20, 2018
1394 10 Best Foods for You	HEALTH_AND_FITNESS	4.0	2490	3.8M	500,000+	Free	0	Everyone 10+	Health & Fitness	February 17, 2017
8484 10 Minutes a Day Times Tables	FAMILY	4.1	681	48M	100,000+	Free	0	Everyone	Education	July 3, 2014
7739 10 WPM Amateur ham radio CW Morse code trainer	COMMUNICATION	3.5	10	3.8M	100+	Paid	\$1.49	Everyone	Communication	May 12, 2018
8220 10,000 Quotes DB (Premium)	BOOKS_AND_REFERENCE	4.1	70	3.5M	500+	Paid	\$0.99	Everyone	Books & Reference	August 30, 2013
9402 100 Doors of Revenge	FAMILY	4.1	105766	48M	10,000,000+	Free	0	Teen	Puzzle	May 24, 2018
4031 100+ C Programs	FAMILY	4.2	20	1.6M	5,000+	Free	0	Everyone	Education	January 3, 2017
8704 100000+ Messages - DP, Status, Jokes & GIF 2018	LIFESTYLE	3.7	121	3.8M	10,000+	Free	0	Mature 17+	Lifestyle	December 29, 2017
4027 101 C Programming Problems	FAMILY	4.6	498	5.0M	50,000+	Free	0	Everyone	Education	October 15, 2017
221 104 Looking for a job - looking for a job, looking for a job, L...	BUSINESS	4.4	74359	25M	1,000,000+	Free	0	Everyone	Business	July 27, 2018
2695 11st	SHOPPING	3.8	48732	20M	10,000,000+	Free	0	Everyone	Shopping	July 31, 2018
5345 12 Step Meditations & Sober Prayers AA NA AL-ANON	LIFESTYLE	4.7	759	15M	50,000+	Free	0	Everyone 10+	Lifestyle	January 22, 2018
6761 17th Edition Cable Sizer	BOOKS_AND_REFERENCE	4.4	47	1.4M	1,000+	Paid	\$3.08	Everyone	Books & Reference	May 27, 2016
2323 1800 Contacts - Lens Store	MEDICAL	4.7	23160	26M	1,000,000+	Free	0	Everyone	Medical	July 27, 2018
1998 1Line â€” One Line with One Touch	GAME	4.6	214878	10M	10,000,000+	Free	0	Everyone	Board	August 3, 2018

4. Removing symbols

Since I wanted to analyze the statistics of data, I had to change the type of some of the variables. Before doing that, I decided to remove the unnecessary alphabets or symbols from the values.

1) I removed the “\$” from the values of “price” column using “gsub” function.

Before cleaning:

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content.Rating	Genres	Last.Updated
1829 Chapters: Interactive Stories	GAME	4.5	73539	96M	1,000,000+	Free	0	Mature 17+	Role Playing	August 2, 2018
1830 Honkai Impact 3rd	GAME	4.7	59017	82M	1,000,000+	Free	0	Teen	Action	July 3, 2018
1831 Master of Eternity(MOE)	GAME	4.2	5829	70M	100,000+	Free	0	Teen	Strategy	July 26, 2018
1832 The Game of Life	GAME	4.4	18621	63M	100,000+	Paid	\$2.99	Everyone	Board	July 4, 2018
1833 Clue	GAME	4.6	19922	35M	100,000+	Paid	\$1.99	Everyone 10+	Board	July 30, 2018
1834 The Room: Old Sins	GAME	4.9	21119	48M	100,000+	Paid	\$4.99	Everyone	Puzzle	April 18, 2018
1835 The Escapists	GAME	4.4	7412	84M	100,000+	Paid	\$4.99	Teen	Strategy	April 26, 2018
1836 Farming Simulator 18	GAME	4.5	18125	15M	100,000+	Paid	\$4.99	Everyone	Simulation:Education	July 9, 2018
1837 RollerCoaster TycoonÂ® Classic	GAME	4.6	10795	69M	100,000+	Paid	\$5.99	Everyone	Simulation	December 21, 2017
1838 Call of Duty:Black Ops Zombies	GAME	4.2	13004	46M	100,000+	Paid	\$6.99	Teen	Action	December 20, 2016
1839 Star Wars â„ƒ: DIRTY	GAME	4.5	38207	15M	100,000+	Paid	\$9.99	Teen	Role Playing	October 19, 2015
1840 Monument Valley 2	GAME	4.6	9394	33M	100,000+	Paid	\$4.99	Everyone	Puzzle	November 30, 2017
1841 DRAGON BALL LEGENDS	GAME	4.6	337913	48M	5,000,000+	Free	0	Teen	Action	August 3, 2018
1842 ROBLOX	GAME	4.5	4449882	67M	100,000,000+	Free	0	Everyone 10+	Adventure:Action & Adventure	July 31, 2018
1843 Candy Crush Saga	GAME	4.4	22429716	74M	500,000,000+	Free	0	Everyone	Casual	July 5, 2018

```
> ds_play_store$Price=gsub("\\$", "",ds_play_store$Price)
> View(ds_play_store)
```

```
> ds_play_store$Price=gsub("\\$", "",ds_play_store$Price)
> View(ds_play_store)
```

After cleaning:

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content.Rating	Genres	Last.Updated
1829 Chapters: Interactive Stories	GAME	4.5	73539	96M	1,000,000+	Free	0	Mature 17+	Role Playing	August 2, 2018
1830 Honkai Impact 3rd	GAME	4.7	59017	82M	1,000,000+	Free	0	Teen	Action	July 3, 2018
1831 Master of Eternity(MOE)	GAME	4.2	5829	70M	100,000+	Free	0	Teen	Strategy	July 26, 2018
1832 The Game of Life	GAME	4.4	18621	63M	100,000+	Paid	2.99	Everyone	Board	July 4, 2018
1833 Clue	GAME	4.6	19922	35M	100,000+	Paid	1.99	Everyone 10+	Board	July 30, 2018
1834 The Room: Old Sins	GAME	4.9	21119	48M	100,000+	Paid	4.99	Everyone	Puzzle	April 18, 2018
1835 The Escapists	GAME	4.4	7412	84M	100,000+	Paid	4.99	Teen	Strategy	April 26, 2018
1836 Farming Simulator 18	GAME	4.5	18125	15M	100,000+	Paid	4.99	Everyone	Simulation:Education	July 9, 2018
1837 RollerCoaster TycoonÂ® Classic	GAME	4.6	10795	69M	100,000+	Paid	5.99	Everyone	Simulation	December 21, 2017
1838 Call of Duty:Black Ops Zombies	GAME	4.2	13004	46M	100,000+	Paid	6.99	Teen	Action	December 20, 2016
1839 Star Wars â„ƒ: DIRTY	GAME	4.5	38207	15M	100,000+	Paid	9.99	Teen	Role Playing	October 19, 2015
1840 Monument Valley 2	GAME	4.6	9394	33M	100,000+	Paid	4.99	Everyone	Puzzle	November 30, 2017
1841 DRAGON BALL LEGENDS	GAME	4.6	337913	48M	5,000,000+	Free	0	Teen	Action	August 3, 2018
1842 ROBLOX	GAME	4.5	4449882	67M	100,000,000+	Free	0	Everyone 10+	Adventure:Action & Adventure	July 31, 2018
1843 Candy Crush Saga	GAME	4.4	22429716	74M	500,000,000+	Free	0	Everyone	Casual	July 5, 2018

2) I also removed unnecessary symbols like ‘+’ and ‘,’ from “installs” column using “gsub” function.

Before cleaning:

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content.Rating	Genres	Last.Updated
1 Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10.000+	Free	0	Everyone	Art & Design	January 7, 2018
2 Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500.000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018
3 U Launcher Lite à€ FREE Live Cool Themes. Hide Apps	ART_AND_DESIGN	4.7	87510	8.7M	5.000.000+	Free	0	Everyone	Art & Design	August 1, 2018
4 Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50.000.000+	Free	0	Teen	Art & Design	June 8, 2018
5 Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100.000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018
6 Paper flowers instructions	ART_AND_DESIGN	4.4	167	5.6M	50.000+	Free	0	Everyone	Art & Design	March 26, 2017
7 Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	178	19M	50.000+	Free	0	Everyone	Art & Design	April 26, 2018
8 Infinite Painter	ART_AND_DESIGN	4.1	36815	29M	1.000.000+	Free	0	Everyone	Art & Design	June 14, 2018
9 Garden Coloring Book	ART_AND_DESIGN	4.4	13791	33M	1.000.000+	Free	0	Everyone	Art & Design	September 20, 2017
10 Kids Paint Free - Drawing Fun	ART_AND_DESIGN	4.7	121	3.1M	10.000+	Free	0	Everyone	Art & Design;Creativity	July 3, 2018
11 Text on Photo - Fontee	ART_AND_DESIGN	4.4	13880	28M	1.000.000+	Free	0	Everyone	Art & Design	October 27, 2017
12 Name Art Photo Editor - Focus n Filters	ART_AND_DESIGN	4.4	8788	12M	1.000.000+	Free	0	Everyone	Art & Design	July 31, 2018
13 Tattoo Name On My Photo Editor	ART_AND_DESIGN	4.2	44829	20M	10.000.000+	Free	0	Teen	Art & Design	April 2, 2018
14 Mandala Coloring Book	ART_AND_DESIGN	4.6	4326	21M	100.000+	Free	0	Everyone	Art & Design	June 26, 2018
15 3D Color Pixel by Number - Sandbox Art Coloring	ART_AND_DESIGN	4.4	1518	37M	100.000+	Free	0	Everyone	Art & Design	August 3, 2018

```
> ds_play_store$Installs=gsub("\\\\+", "",ds_play_store$ Installs)
> ds_play_store$Installs=gsub("\\\\,", "",ds_play_store$ Installs)
> View(ds_play_store)
```

```
> ds_play_store$Installs=gsub("\\\\+", "",ds_play_store$ Installs)
> ds_play_store$Installs=gsub("\\\\,", "",ds_play_store$ Installs)
> View(ds_play_store)
```

After cleaning:

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content.Rating	Genres	Last.Updated
1 Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10000	Free	0	Everyone	Art & Design	January 7, 2018
2 Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500000	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018
3 U Launcher Lite à€ FREE Live Cool Themes. Hide Apps	ART_AND_DESIGN	4.7	87510	8.7M	5000000	Free	0	Everyone	Art & Design	August 1, 2018
4 Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50000000	Free	0	Teen	Art & Design	June 8, 2018
5 Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100000	Free	0	Everyone	Art & Design;Creativity	June 20, 2018
6 Paper flowers instructions	ART_AND_DESIGN	4.4	167	5.6M	50000	Free	0	Everyone	Art & Design	March 26, 2017
7 Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	178	19M	50000	Free	0	Everyone	Art & Design	April 26, 2018
8 Infinite Painter	ART_AND_DESIGN	4.1	36815	29M	1000000	Free	0	Everyone	Art & Design	June 14, 2018
9 Garden Coloring Book	ART_AND_DESIGN	4.4	13791	33M	1000000	Free	0	Everyone	Art & Design	September 20, 2017
10 Kids Paint Free - Drawing Fun	ART_AND_DESIGN	4.7	121	3.1M	10000	Free	0	Everyone	Art & Design;Creativity	July 3, 2018
11 Text on Photo - Fontee	ART_AND_DESIGN	4.4	13880	28M	1000000	Free	0	Everyone	Art & Design	October 27, 2017
12 Name Art Photo Editor - Focus n Filters	ART_AND_DESIGN	4.4	8788	12M	1000000	Free	0	Everyone	Art & Design	July 31, 2018
13 Tattoo Name On My Photo Editor	ART_AND_DESIGN	4.2	44829	20M	10000000	Free	0	Teen	Art & Design	April 2, 2018
14 Mandala Coloring Book	ART_AND_DESIGN	4.6	4326	21M	100000	Free	0	Everyone	Art & Design	June 26, 2018
15 3D Color Pixel by Number - Sandbox Art Coloring	ART_AND_DESIGN	4.4	1518	37M	100000	Free	0	Everyone	Art & Design	August 3, 2018

5. Fixing data types

I had to change the type of some of the variables (“Rating”, “Reviews”, “Installs”, “Price” and “Last Updated”) before analyzing the statistics of data and visualization.

```
> str(ds_play_store)
'data.frame': 8892 obs. of 11 variables:
 $ App      : chr "Photo Editor & Candy Camera & Grid & ScrapBook" "Coloring book moana" "U Launcher Lite â\200" FREE Live Cool Themes, Hide Apps" "Sketch
- Draw & Paint" ...
 $ Category : chr "ART_AND_DESIGN" "ART_AND_DESIGN" "ART_AND_DESIGN" "ART_AND_DESIGN" ...
 $ Rating   : num 4.1 3.9 4.7 4.5 4.3 4.4 3.8 4.1 4.4 4.7 ...
 $ Reviews  : chr "159" "967" "87510" "215644" ...
 $ Size     : chr "19M" "14M" "8.7M" "25M" ...
 $ Installs : chr "10000" "500000" "5000000" "50000000" ...
 $ Type     : chr "Free" "Free" "Free" "Free" ...
 $ Price    : chr "0" "0" "0" "0" ...
 $ Content.Rating: chr "Everyone" "Everyone" "Everyone" "Teen" ...
 $ Genres   : chr "Art & Design" "Art & Design;Pretend Play" "Art & Design" "Art & Design" ...
 $ Last.Updated : chr "January 7, 2018" "January 15, 2018" "August 1, 2018" "June 8, 2018" ...
- attr(*, "na.action")= 'omit' Named int [1:1474] 24 114 124 127 130 131 135 164 181 186 ...
.. attr(*, "names")= chr [1:1474] "24" "114" "124" "127" ...
> ds_play_store$App<-as.character(ds_play_store$App)
> ds_play_store$Category<-as.character(ds_play_store$Category)
> ds_play_store$Rating<-as.numeric(ds_play_store$Rating)
> ds_play_store$Reviews<-as.numeric(ds_play_store$Reviews)
> ds_play_store$Size<-as.character(ds_play_store$Size)
> ds_play_store$Installs<-as.integer(ds_play_store$Installs)
> ds_play_store$Type<-as.character(ds_play_store$Type)
> ds_play_store$Price<-as.numeric(ds_play_store$Price)
> ds_play_store$Content.Rating<-as.character(ds_play_store$Content.Rating)
> ds_play_store$Genres<-as.character(ds_play_store$Genres)
> ds_play_store$Last.Updated <- as.Date(ds_play_store$Last.Updated,format="%B %d, %Y",tryFormats = "%m %d %Y")
> str(ds_play_store)
'data.frame': 8892 obs. of 11 variables:
 $ App      : chr "Photo Editor & Candy Camera & Grid & ScrapBook" "Coloring book moana" "U Launcher Lite â\200" FREE Live Cool Themes, Hide Apps" "Sketch
- Draw & Paint" ...
 $ Category : chr "ART_AND_DESIGN" "ART_AND_DESIGN" "ART_AND_DESIGN" "ART_AND_DESIGN" ...
 $ Rating   : num 4.1 3.9 4.7 4.5 4.3 4.4 3.8 4.1 4.4 4.7 ...
 $ Reviews  : num 159 967 87510 215644 967 ...
 $ Size     : chr "19M" "14M" "8.7M" "25M" ...
 $ Installs : int 10000 500000 5000000 50000000 100000 50000 50000 1000000 1000000 10000 ...
 $ Type     : chr "Free" "Free" "Free" "Free" ...
 $ Price    : num 0 0 0 0 0 0 0 0 0 0 ...
 $ Content.Rating: chr "Everyone" "Everyone" "Everyone" "Teen" ...
 $ Genres   : chr "Art & Design" "Art & Design;Pretend Play" "Art & Design" "Art & Design" ...
 $ Last.Updated : Date, format: "2018-01-07" "2018-01-15" "2018-08-01" "2018-06-08" ...
- attr(*, "na.action")= 'omit' Named int [1:1474] 24 114 124 127 130 131 135 164 181 186 ...
.. attr(*, "names")= chr [1:1474] "24" "114" "124" "127" ...
```

```
> str(ds_play_store) ...

> ds_play_store$App<-as.character(ds_play_store$App)

> ds_play_store$Category<-as.character(ds_play_store$Category)

> ds_play_store$Rating<-as.numeric(ds_play_store$Rating)

> ds_play_store$Reviews<-as.numeric(ds_play_store$Reviews)

> ds_play_store$Size<-as.character(ds_play_store$Size)

> ds_play_store$Installs<-as.integer(ds_play_store$Installs)

> ds_play_store$Type<-as.character(ds_play_store$Type)

> ds_play_store$Price<-as.numeric(ds_play_store$Price)

> ds_play_store$Content.Rating<-as.character(ds_play_store$Content.Rating)

> ds_play_store$Genres<-as.character(ds_play_store$Genres)

> ds_play_store$Last.Updated <- as.Date(ds_play_store$Last.Updated,format="%B %d,
%Y",tryFormats = "%m %d %Y")

> str(ds_play_store) ...
```

6. Splitting a column into multiple columns

Since the “Last Updated” variable was containing a complete value of date, I decided to split it into separate columns to be able to use the values separately if needed for my analysis.

First, I changed the type of “Last Updated” variable from *date* to *character* because I wanted to use “separate” function.

```
> ds_play_store$Last.Updated<-as.character(ds_play_store$Last.Updated)
```

Next, I used the “separate” function to split “Last Updated” column into separate columns.

Before cleaning:

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content.Rating	Genres	Last.Updated
1 Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10000	Free	0	Everyone	Art & Design	2018-01-07
2 Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500000	Free	0	Everyone	Art & Design;Pretend Play	2018-01-15
3 U Launcher Lite æ FREE Live Cool Themes, Hide Apps	ART_AND_DESIGN	4.7	87510	8.7M	5000000	Free	0	Everyone	Art & Design	2018-08-01
4 Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50000000	Free	0	Teen	Art & Design	2018-06-08
5 Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100000	Free	0	Everyone	Art & Design;Creativity	2018-06-20
6 Paper flowers instructions	ART_AND_DESIGN	4.4	167	5.6M	50000	Free	0	Everyone	Art & Design	2017-03-26
7 Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	178	19M	50000	Free	0	Everyone	Art & Design	2018-04-26
8 Infinite Painter	ART_AND_DESIGN	4.1	36815	29M	1000000	Free	0	Everyone	Art & Design	2018-06-14
9 Garden Coloring Book	ART_AND_DESIGN	4.4	13791	33M	1000000	Free	0	Everyone	Art & Design	2017-09-20
10 Kids Paint Free - Drawing Fun	ART_AND_DESIGN	4.7	121	3.1M	10000	Free	0	Everyone	Art & Design;Creativity	2018-07-03
11 Text on Photo - Fontee	ART_AND_DESIGN	4.4	13880	28M	1000000	Free	0	Everyone	Art & Design	2017-10-27
12 Name Art Photo Editor - Focus n Filters	ART_AND_DESIGN	4.4	8788	12M	1000000	Free	0	Everyone	Art & Design	2018-07-31
13 Tattoo Name On My Photo Editor	ART_AND_DESIGN	4.2	44829	20M	10000000	Free	0	Teen	Art & Design	2018-04-02
14 Mandala Coloring Book	ART_AND_DESIGN	4.6	4326	21M	100000	Free	0	Everyone	Art & Design	2018-06-26
15 3D Color Pixel by Number - Sandbox Art Coloring	ART_AND_DESIGN	4.4	1518	37M	100000	Free	0	Everyone	Art & Design	2018-08-03

```
> ds_play_store<-separate(ds_play_store,Last.Updated,c("Last.Updated.Year","Last.Updated.Month","Last.Updated.Day"),sep="-")
> View(ds_play_store)
> head(ds_play_store)
```

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content.Rating	Genres	Last.Updated.Year
1 Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10000	Free	0	Everyone	Art & Design	2018
2 Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500000	Free	0	Everyone	Art & Design;Pretend Play	2018
3 U Launcher Lite æ FREE Live Cool Themes, Hide Apps	ART_AND_DESIGN	4.7	87510	8.7M	5000000	Free	0	Everyone	Art & Design	2018
4 Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50000000	Free	0	Teen	Art & Design	2018
5 Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100000	Free	0	Everyone	Art & Design;Creativity	2018
6 Paper flowers instructions	ART_AND_DESIGN	4.4	167	5.6M	50000	Free	0	Everyone	Art & Design	2017
Last.Updated.Month	Last.Updated.Day									
01	07									
02	15									
08	01									
06	08									
06	20									
03	26									

```
ds_play_store<-
separate(ds_play_store,Last.Updated,c("Last.Updated.Year","Last.Updated.Month","Last.Updated.Day"),sep="-")
```

```
View(ds_play_store)
```

```
head(ds_play_store)
```

```
...
```

After cleaning:

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content.Rating	Genres	Last.Updated.Year	Last.Updated.Month	Last.Updated.Day
1 Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10000	Free	0	Everyone	Art & Design	2018	01	07
2 Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500000	Free	0	Everyone	Art & Design;Pretend Play	2018	01	15
3 U Launcher Lite ¼ FREE Live Cool Themes.Hide Apps	ART_AND_DESIGN	4.7	87510	8.7M	5000000	Free	0	Everyone	Art & Design	2018	08	01
4 Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50000000	Free	0	Teen	Art & Design	2018	06	08
5 Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100000	Free	0	Everyone	Art & Design;Creativity	2018	06	20
6 Paper flowers instructions	ART_AND_DESIGN	4.4	167	5.6M	50000	Free	0	Everyone	Art & Design	2017	03	26
7 Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	178	19M	50000	Free	0	Everyone	Art & Design	2018	04	26
8 Infinite Painter	ART_AND_DESIGN	4.1	36815	29M	1000000	Free	0	Everyone	Art & Design	2018	06	14
9 Garden Coloring Book	ART_AND_DESIGN	4.4	13791	33M	1000000	Free	0	Everyone	Art & Design	2017	09	20
10 Kids Paint Free - Drawing Fun	ART_AND_DESIGN	4.7	121	3.1M	10000	Free	0	Everyone	Art & Design;Creativity	2018	07	03
11 Text on Photo - Fontee	ART_AND_DESIGN	4.4	13880	28M	1000000	Free	0	Everyone	Art & Design	2017	10	27
12 Name Art Photo Editor - Focus n Filters	ART_AND_DESIGN	4.4	8788	12M	1000000	Free	0	Everyone	Art & Design	2018	07	31
13 Tattoo Name On My Photo Editor	ART_AND_DESIGN	4.2	44829	20M	10000000	Free	0	Teen	Art & Design	2018	04	02
14 Mandala Coloring Book	ART_AND_DESIGN	4.6	4326	21M	100000	Free	0	Everyone	Art & Design	2018	06	26
15 3D Color Pixel by Number - Sandbox Art Coloring	ART_AND_DESIGN	4.4	1518	37M	100000	Free	0	Everyone	Art & Design	2018	08	03

Let's take a look at the final clean data using "glimpse" function from "dplyr" library.

```
> glimpse(ds_play_store)
Rows: 8,892
Columns: 13
$ App           <chr> "Photo Editor & Candy Camera & Grid & ScrapBook", "Coloring book moana", "U Launcher Lite â\200...
$ Category      <chr> "ART_AND_DESIGN", "ART_AND_DESIGN", "ART_AND_DESIGN", "ART_AND_DESIGN", "ART_AND_DESIGN", "ART_...
$ Rating        <dbl> 4.1, 3.9, 4.7, 4.5, 4.3, 4.4, 3.8, 4.1, 4.4, 4.7, 4.4, 4.4, 4.2, 4.6, 4.4, 3.2, 4.7, 4.5, 4.3, ...
$ Reviews       <dbl> 159, 967, 87510, 215644, 967, 167, 178, 36815, 13791, 121, 13880, 8788, 44829, 4326, 1518, 55, ...
$ Size         <chr> "19M", "14M", "8.7M", "25M", "2.8M", "5.6M", "19M", "29M", "33M", "3.1M", "28M", "12M", "20M", ...
$ Installs      <int> 10000, 500000, 5000000, 50000000, 100000, 50000, 50000, 1000000, 1000000, 10000, 1000000, 10000...
$ Type          <chr> "Free", "Free", "Free", "Free", "Free", "Free", "Free", "Free", "Free", "Free", "Free", "Free", ...
$ Price         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
$ Content.Rating <chr> "Everyone", "Everyone", "Everyone", "Teen", "Everyone", "Everyone", "Everyone", "Everyone", "Ev...
$ Genres        <chr> "Art & Design", "Art & Design;Pretend Play", "Art & Design", "Art & Design", "Art & Design;Crea...
$ Last.Updated.Year <chr> "2018", "2018", "2018", "2018", "2018", "2017", "2018", "2018", "2017", "2018", "2017", "2018", ...
$ Last.Updated.Month <chr> "01", "01", "08", "06", "06", "03", "04", "06", "09", "07", "10", "07", "04", "06", "08", "06", ...
$ Last.Updated.Day <chr> "07", "15", "01", "08", "20", "26", "26", "14", "20", "03", "27", "31", "02", "26", "03", "06", ...
```

Analysis and Visualizations

After cleaning the data, I did the following data analysis and visualizations:

Note: Multiple package such as “tidyverse”, “janitor”, “scales”, etc. were installed to be able to use different libraries.

1. What is the app distribution by type?

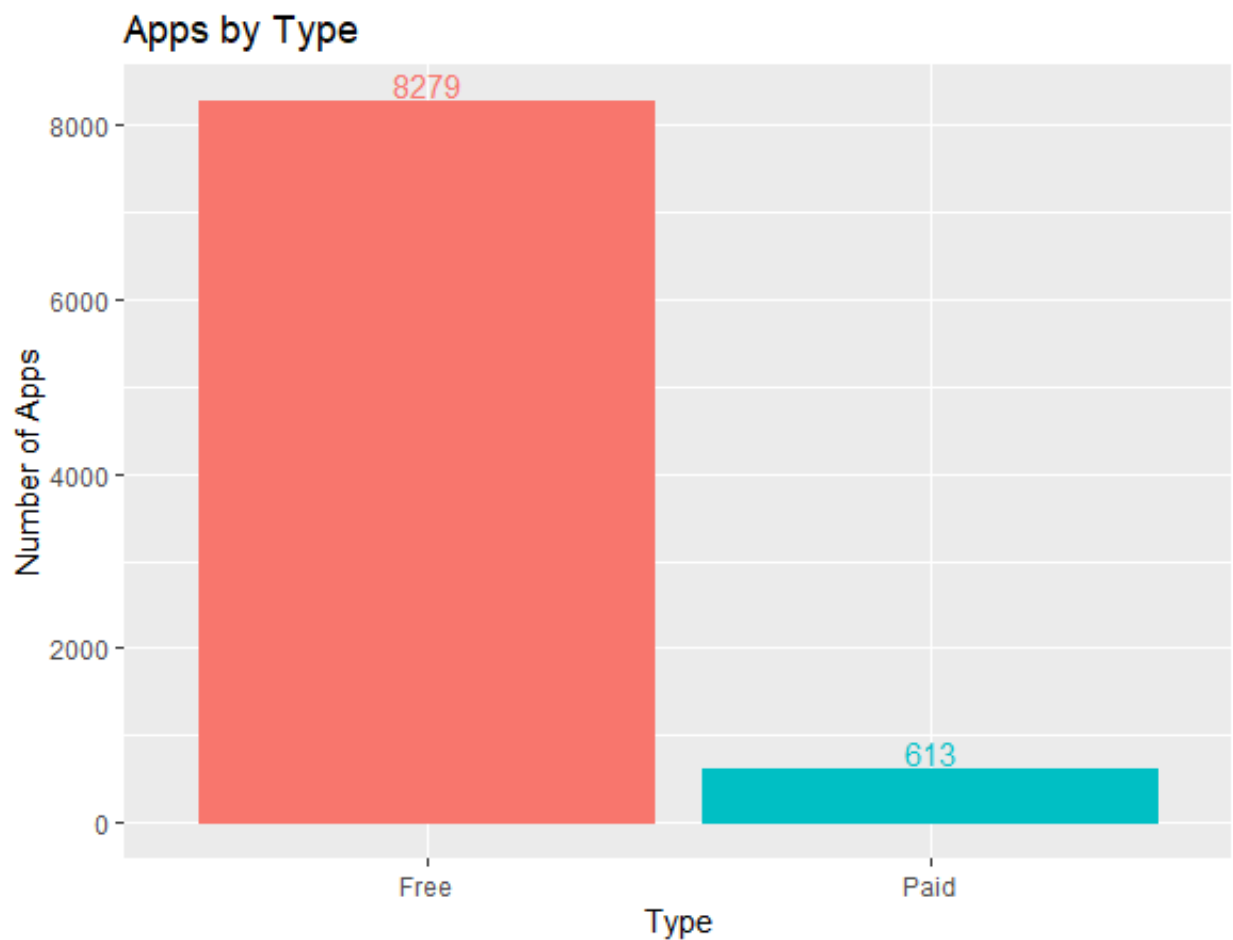


Figure 1 Vertical Bar Chart showing app distribution by type

R features:

- **Plot Type:** Bar plot (vertical)
- **Functions:** `myplot`, `aesthetics(aes)`, `factor`, `geom_bar`, `labs`, `theme`, `geom_text`, `guides`
- **Libraries:** `ggplot2`

A user-defined function named “**myplot**” was used to create this plot.

```
library(ggplot2)

myplot <- function(myds, myxcol, myycol, mytitle, myx, myy) {

  ggplot(myds, aes(x=factor(myxcol), color=factor(myycol), fill = myycol))+
    geom_bar(position="dodge")+
    labs(title= mytitle, x= myx, y = myy)+ theme (legend.position = "none") +
    geom_text(aes(label=..count..), stat='count', position=position_dodge(0.9), vjust=-0.2) + guides(color = FALSE)
}
```

The workspace was pointed to the appropriate directory:

```
> setwd("C:/Users/Ninel/Desktop/School/CSULA/CIS 5250/R Labs/R Scripts")
```

Next, the source function was used to load the “script file and then the “**myplot**” function was called to create the plot.

```
> source('apps_by_type_script.R')
> myplot(ds_play_store, ds_play_store$Type, ds_play_store$Type, "Apps by Type", "Type", "Number of Apps")
```

```
library(ggplot2)

myplot <- function(myds, myxcol, myycol, mytitle, myx, myy) {

  ggplot(myds, aes(x=factor(myxcol), color=factor(myycol), fill = myycol))+
    geom_bar(position="dodge")+
    labs(title= mytitle, x= myx, y = myy)+ theme (legend.position = "none") +
    geom_text(aes(label=..count..), stat='count', position=position_dodge(0.9), vjust=-0.2) + guides(color
= FALSE)
}
```

```
> source('apps_by_type_script.R')
```

```
> myplot(ds_play_store, ds_play_store$Type, ds_play_store$Type, "Apps by Type", "Type", "Number of
Apps")
```

Insights:

My initial approach to analyze the data was to first see what is the distribution of free apps and paid apps. As we can see, almost 8k of all Android applications were available for free and only 613 apps were paid. This means that most apps are free on Google play store. The reason is that free apps generally receive more downloads than paid apps. So, it easier to get tons of more users. Also, users tend to have lower expectations with free apps. There can be users who download an app, use it for a little while but then they stop using it since it isn't helping them in their everyday life. On the other side, there can be thousands of people that download the app and find it useful. So, they can recommend it to others or even give positive reviews. That is why many apps are distributed for free through app stores.

2. What are the top 10 installed apps?

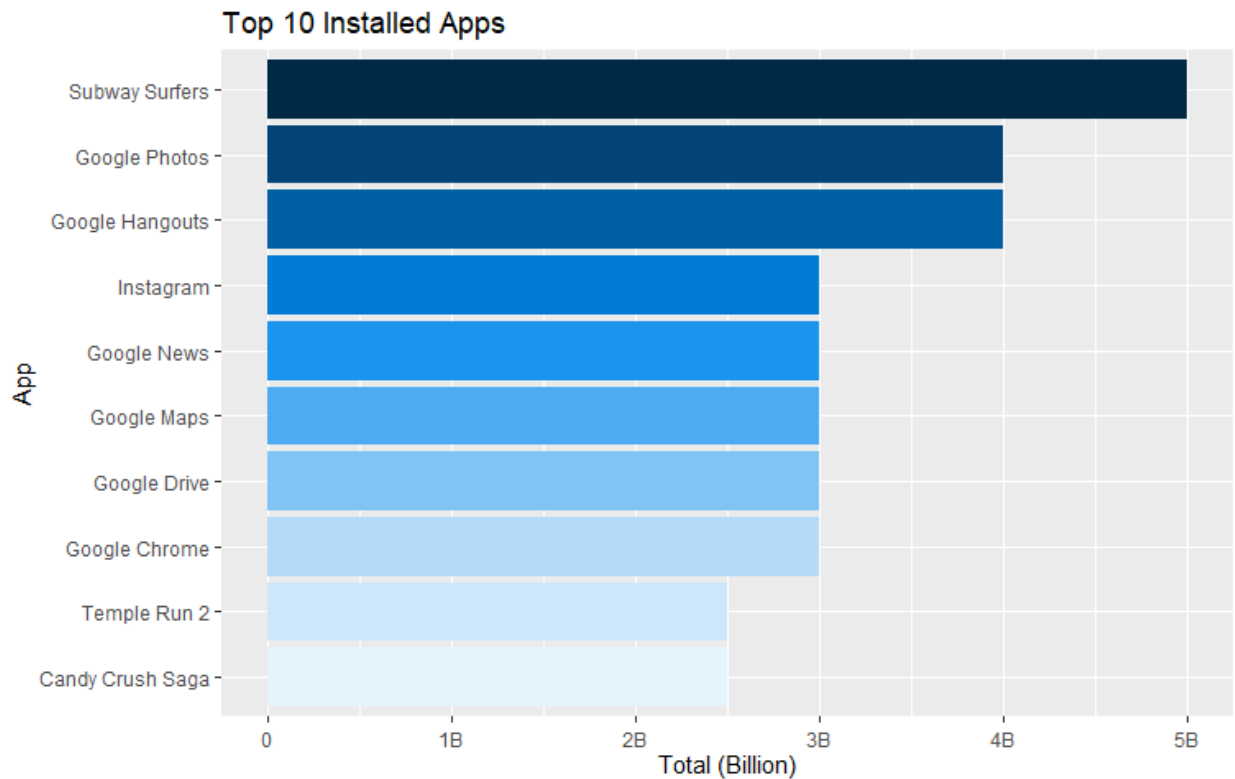


Figure 2 Horizontal Bar Chart showing top 10 installed apps

R features:

- **Plot Type:** Bar chart (horizontal)
- **Functions:** `group_by`, `summarize`, `arrange`, `head`, `aesthetics(aes)`, `coord_flip`, `geom_bar`, `labs`, `theme`, `scale_fill_manual`, `scale_y_continuous`
- **Libraries:** `ggplot2`, `dplyr`, `scales`

Since the total number of installs was a large number, it was displayed in Scientific format in exponential notation so I used **`options(scipen = 999)`** function to turn off the Scientific notation before creating the plot.

```

> library(ggplot2)
> library(dplyr)
> library(scales)
> ds_play_store%>%
+   group_by(App) %>%
+   summarize(Total = (sum(Installs))) %>%
+   arrange(desc(Total)) %>%
+   head(10) %>%
+   ggplot(aes(x = reorder(App,Total), y = Total, fill = App)) +
+   coord_flip() +
+   geom_bar(stat="identity") +
+   labs(title= "Top 10 Installed Apps", x = "App" , y = "Total (Billion)" ) +
+   theme(axis.text.x = element_text(angle = 360, vjust=0.5), legend.position = "none") +
+   scale_fill_manual(values=c("#E5F3FD", "#B3DBF9", "#80C4F5", "#015FA4", "#4DACF1" , "#1B94ED", "#014475", "#017BD3", "#002946", "#CCE7FB")) +
+   scale_y_continuous(labels = scales::label_number_si())

```

```

> library(ggplot2)

> library(dplyr)

> library(scales)

> ds_play_store%>%

group_by(App) %>%

summarize(Total = (sum(Installs))) %>%

arrange(desc(Total)) %>%

head(10) %>%

ggplot(aes(x = reorder(App,Total), y = Total, fill = App)) +

coord_flip() +

geom_bar(stat="identity") +

labs(title= "Top 10 Installed Apps", x = "App" , y = "Total (Billion)" ) +

theme(axis.text.x = element_text(angle = 360, vjust=0.5), legend.position = "none") +

scale_fill_manual(values=c("#E5F3FD", "#B3DBF9", "#80C4F5", "#015FA4", "#4DACF1"
, "#1B94ED", "#014475", "#017BD3", "#002946", "#CCE7FB")) +

scale_y_continuous(labels = scales::label_number_si())

```

Insights:

My next thought was to find what are the top 10 apps that are most popular among users. For this analysis, I have considered popularity in terms of the number of installs the users have made.

All the top 10 apps had over a billion downloads. As we can see most of the apps are from Google because most of these apps are pre-installed. However, “Subway Surfers” mobile game had the largest number of installs with almost 5 billion times among 8892 apps. Based on my research, there are many reasons behind the popularity of this game. What is worth mentioning here is the charm that this game carries to attract people by adding adventures while keeping it simple and cool.² Google Hangouts and Google Photos had the second greatest number of installs with almost 4 billion installs. Besides some Google apps, we can also see 2 more games among the top installed apps.

² admin. “Subway Surfers - Reasons Behind Higher Revenues and Downloads.” Ommzi Solutions - Game Development Company, 7 May 2018, www.ommzi.com/subway-surfers-reasons-behind-higher-revenues-downloads/.

3. What are the top 5 reviewed categories?

Top 5 Reviewed Categories

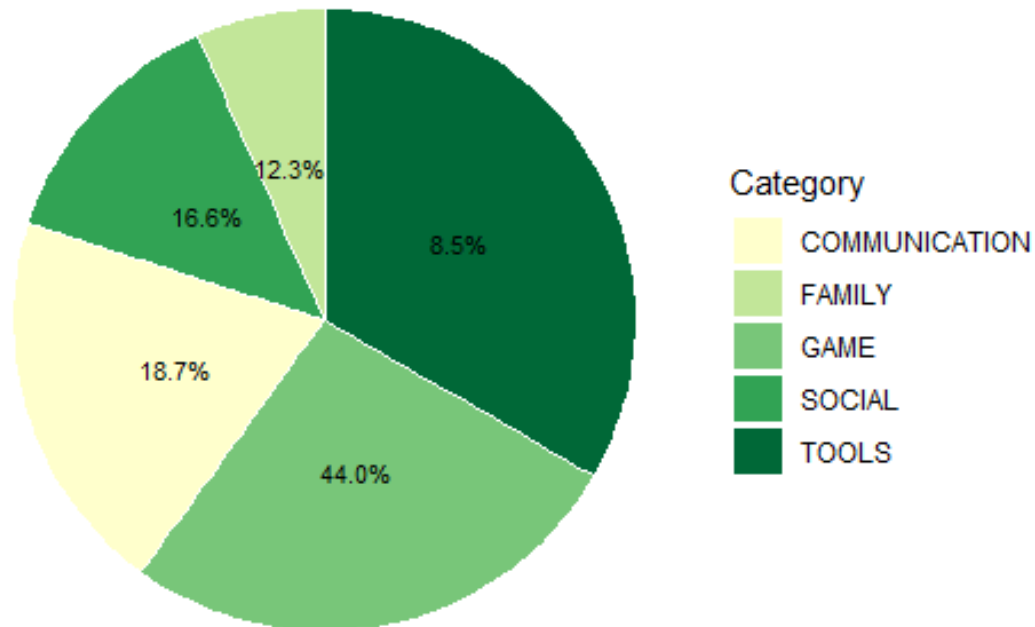


Figure 3 Pie Chart showing top 5 reviewed categories

R features:

- **Plot Type:** Pie chart
- **Functions:** `group_by`, `summarize`, `arrange`, `head`, `select`, `aesthetics(aes)`, `coord_polar`, `geom_text`, `labs`, `theme`, `scale_fill_brewer`, `theme_void`
- **Libraries:** `ggplot2`, `dplyr`, `scales`, `RColorBrewer`

In order to get the top 5 reviewed categories, I calculated the total reviews of top 5 categories and added the results into a smaller dataframe by doing the following:

```

> ds_top5<-ds_play_store%>%
+   group_by(Category) %>%
+   summarize(Total = (sum(Reviews))) %>%
+   arrange(desc(Total)) %>%
+   head(5) %>%
+   select (Category, Total)

```

```

> library(dplyr)

> ds_top5<-ds_play_store%>%

group_by(Category) %>%

summarize(Total = (sum(Reviews))) %>%

arrange(desc(Total)) %>%

head(5) %>%

select (Category, Total)

```

	Category	Total
1	GAME	1415533878
2	COMMUNICATION	601273091
3	SOCIAL	533576498
4	FAMILY	396768720
5	TOOLS	273184168

Then I calculated the percentage of the total reviewed and inserted it into a new column called “prop” to the smaller dataframe “ds_top5”.

```

> ds_top5 %>%
+   arrange(desc(Total)) %>%
+   mutate(prop = percent(Total / sum(Total))) -> ds_top5

```

```

> library(dplyr)

> ds_top5 %>%

arrange(desc(Total)) %>%

mutate(prop = percent(Total / sum(Total))) -> ds_top5

```

	Category	Total	prop
1	GAME	1415533878	44.0%
2	COMMUNICATION	601273091	18.7%
3	SOCIAL	533576498	16.6%
4	FAMILY	396768720	12.3%
5	TOOLS	273184168	8.5%

Then I created the plot.

```

> ggplot(ds_top5, aes(x="", y=prop, fill=Category)) +
+   geom_bar(stat="identity", width=1, color = "white") +
+   coord_polar("y", start=0) +
+   geom_text(aes(label = paste0(prop)), position = position_stack(vjust=0.5), size=3) +
+   labs(x = NULL, y = NULL, title = "Top 5 Reviewed Categories") +
+   theme(axis.line = element_blank(), axis.text = element_blank(), axis.ticks = element_blank()) +
+   scale_fill_brewer(palette="YlGn") + theme_void()

```



```

> library(ggplot2)
> library(scales)
> library(RColorBrewer)
> ggplot(ds_top5, aes(x="", y=prop, fill=Category)) +
  geom_bar(stat="identity", width=1, color = "white") +
  coord_polar("y", start=0) +
  geom_text(aes(label = paste0(prop)), position = position_stack(vjust=0.5), size=3) +
  labs(x = NULL, y = NULL, title = "Top 5 Reviewed Categories") +
  theme(axis.line = element_blank(), axis.text = element_blank(), axis.ticks = element_blank())
  +
  scale_fill_brewer(palette="YlGn") + theme_void()

```

Insights:

My next motive was to find out the top 5 reviewed categories. The above pie chart shows the percentage of top 5 reviewed categories. As we can see, “Game” category had the maximum reviewed rate with 44% compared to other categories among 33 categories. This was expected since we had 3 apps (Subway Surfers, Temple Run 2, Candy Crush Saga) from “Game” category in our top 10 installed apps. There are many reasons about the popularity of mobile games. Perhaps one of the easiest reasons is because they are easily accessible. Practically everyone has a smartphone nowadays, so anyone can download a game on his or her phone and start gaming right away. Another reason can be that mobile games are more user friendly and cheap or free. “Communication” category had the second greatest number of reviewed with almost 18% reviews. For example, we had “Google Hangouts” and “Google Chrome” from this category among the top 10 installed apps. And “Social” comes in at third place with 16.6% reviews. In general, we can say the more an app is installed, the more it gets reviews.

4. Which age category has rated the apps and what is the association between content rating and app rating?

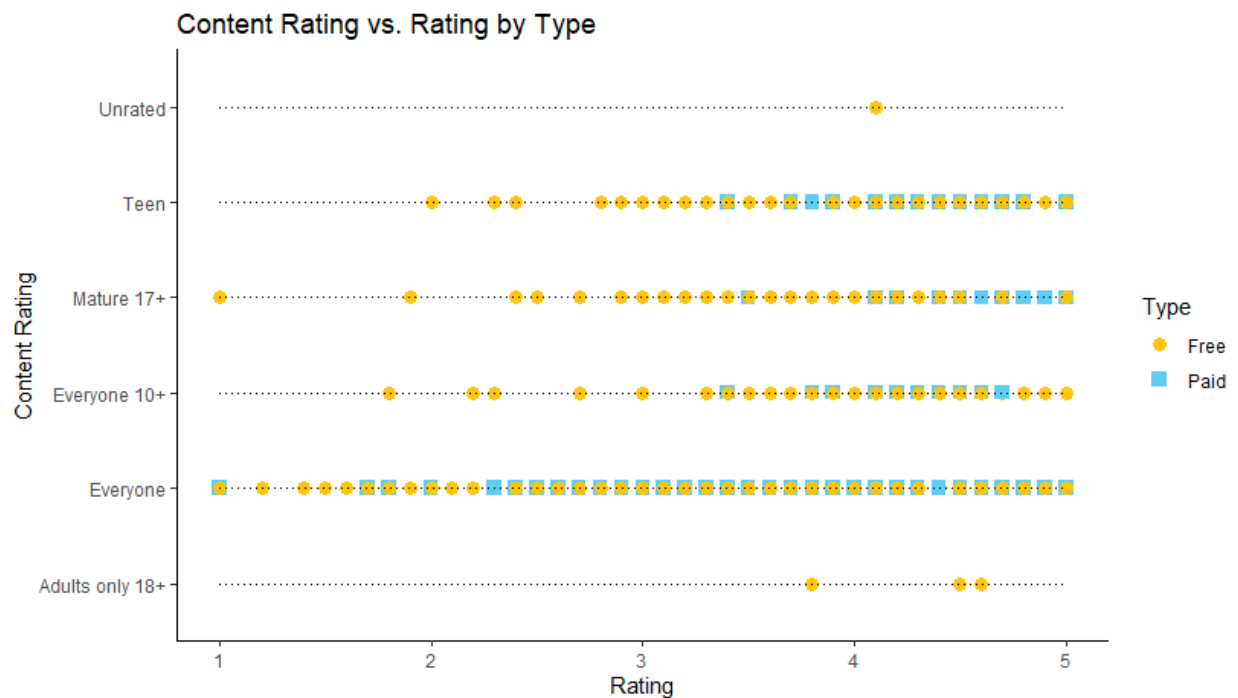


Figure 4 Dot plot showing relationship between content rating and app rating

R features:

- **Plot Type:** Dot plot (or Scatterplot)
- **Functions:** mydotplot, aesthetics(aes), geom_point, geom_segment, coord_flip, labs, theme, scale_shape_manual, scale_color_manual
- **Libraries:** ggplot2, scales

A user-defined function named “**mydotplot**” was used to create this plot.

```
library(ggplot2)
library(scales)

mydotplot <- function(myds, myxcol, myycol, mygroup, mytitle, myx, myy, myshape1, myshape2, mycolor1, mycolor2) {

  ggplot(myds, aes(x=myxcol, y=myycol, group=mygroup)) +
    geom_point(size=3, aes(shape=mygroup, color=mygroup)) +
    geom_segment(aes(x=myxcol, xend=myxcol, y=min(myycol), yend=max(myycol)), linetype="dotted", size=0.1) +
    labs(title=mytitle, x = myx, y = myy) +
    coord_flip() +
    scale_shape_manual(values=c(myshape1, myshape2)) +
    scale_color_manual(values=c(mycolor1, mycolor2)) +
    theme(legend.position="right")
}
```

The workspace was pointed to the appropriate directory:

```
> setwd("C:/Users/Ninel/Desktop/School/CSULA/CIS 5250/R Labs/R Scripts")
```

Next, the source function was used to load the “script file and then the “**mydotplot**” function was called to create the plot.

```
> source('apps_rating_by_type_script.R')
> mydotplot(ds_play_store, ds_play_store$Content.Rating, ds_play_store$Rating, ds_play_store$Type, "Content Rating vs. Rating by type", "Content Rating", "Rating", 16, 15, '#FFC40C', '#5ECCF3')
```

```
library(ggplot2)

library(scales)

mydotplot <- function(myds, myxcol, myycol, mygroup, mytitle, myx, myy, myshape1, myshape2, mycolor1, mycolor2) {

  ggplot(myds, aes(x=myxcol, y=myycol, group=mygroup)) +

    geom_point(size=3, aes(shape=mygroup, color=mygroup)) +

    geom_segment(aes(x=myxcol, xend=myxcol, y=min(myycol), yend=max(myycol)), linetype="dotted", size=0.1) +

    labs(title=mytitle, x = myx, y = myy) +

    coord_flip() +

    scale_shape_manual(values=c(myshape1, myshape2)) +

    scale_color_manual(values=c(mycolor1, mycolor2)) +

    theme(legend.position="right")

}
```

```
> source('apps_rating_by_type_script.R')
```

```
> mydotplot(ds_play_store, ds_play_store$Content.Rating, ds_play_store$Rating, ds_play_store$Type, "Content Rating vs. Rating by Type", "Content Rating", "Rating", 16, 15, '#FFC40C', '#5ECCF3')
```

Insights:

My last thought was to find out which age category has rated the apps the most and if there is any relationship between the content rating and app rating. We can see that most of the rated apps are from Everyone category while Unrated category has the lowest number. An interesting point was that “Adults Only 18+” also had a low number of rating. In the contrast, “Mature 17+” category had been active rating the apps. In general, free apps are more likely to receive positive reviews and we can infer that the paid apps are also the among the successful apps with ratings 4 or higher.

Statistical Summary and Functions

Statistical Summary

Below we see the statistical summary of our complete dataset:

```
> summary(ds_play_store)
  App          Category      Rating      Reviews
Length:8892   Length:8892   Min.    :1.000   Min.    :    1
Class :character Class :character 1st Qu.:4.000   1st Qu.:   164
Mode  :character Mode  :character Median :4.300   Median :   4714
                        Mean  :4.188   Mean  :  472776
                        3rd Qu.:4.500   3rd Qu.:  71267
                        Max.   :5.000   Max.   :78158306

  Size          Installs      Type      Price
Length:8892     Min.    :    1   Length:8892   Min.    : 0.0000
Class :character 1st Qu.:  10000   Class :character 1st Qu.: 0.0000
Mode  :character Median : 500000   Mode  :character Median : 0.0000
                        Mean  : 16489648   Mean  : 0.9632
                        3rd Qu.: 5000000   3rd Qu.: 0.0000
                        Max.   :1000000000   Max.   :400.0000

  Content.Rating      Genres  Last.Updated.Year  Last.Updated.Month
Length:8892          Length:8892   Length:8892   Length:8892
Class :character     Class :character   Class :character   Class :character
Mode  :character     Mode  :character   Mode  :character   Mode  :character

  Last.Updated.Day
Length:8892
Class :character
Mode  :character
```

➤ Statistical Summary of Rating

```
> summary(ds_play_store$Rating)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000  4.000   4.300   4.188  4.500   5.000

> min(ds_play_store$Rating)
[1] 1

> max(ds_play_store$Rating)
[1] 5

> mean(ds_play_store$Rating)
[1] 4.187877

> median(ds_play_store$Rating)
[1] 4.3

> sd(ds_play_store$Rating)
[1] 0.5223767
```

```

> summary(ds_play_store$Rating)

  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
 1.000  4.000  4.300  4.188  4.500  5.000

> min(ds_play_store$Rating)

[1] 1

> max(ds_play_store$Rating)

[1] 5

> mean(ds_play_store$Rating)

[1] 4.187877

> median(ds_play_store$Rating)

[1] 4.3

> sd(ds_play_store$Rating)

[1] 0.5223767

```

Based on the statistical summary, the min value of rating for an app is 1 and the max value is 5.

Only 16 apps had a rating value of 1 and 271 apps had a rating value of 5.

```

> ds_rating_5 <-filter(ds_play_store,Rating==5)
> View(ds_rating_5)
> dim(ds_rating_5)
[1] 271  13
> ds_rating_1 <-filter(ds_play_store,Rating==1)
> View(ds_rating_1)
> dim(ds_rating_1)
[1] 16 13

```

```
> ds_rating_5 <-filter(ds_play_store,Rating==5)
> View(ds_rating_5)
> dim(ds_rating_5)
[1] 271 13
> ds_rating_1 <-filter(ds_play_store,Rating==1)
> View(ds_rating_1)
> dim(ds_rating_1)
[1] 16 13
```

The mean rating is 4.188 which means lots of apps had high ratings. As we can see 6408 apps had ratings higher than 4.0 among 8892 apps.

```
> ds_rating_more_4 <-filter(ds_play_store,Rating>4)
> View(ds_rating_more_4)
> dim(ds_rating_more_4)
[1] 6408 13
```

```
> ds_rating_more_4 <-filter(ds_play_store,Rating>4)
> View(ds_rating_more_4)
> dim(ds_rating_more_4)
[1] 6408 13
```

Lastly the standard deviation is 0.5 which is a low value. This means that our price values are close to the mean of our set.

➤ Statistical Summary of Reviews

```
> summary(ds_play_store$Reviews)
      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
      1       164      4714    472776    71267   78158306
> min(ds_play_store$Reviews)
[1] 1
> max(ds_play_store$Reviews)
[1] 78158306
> mean(ds_play_store$Reviews)
[1] 472776.4
> median(ds_play_store$Reviews)
[1] 4714.5
> sd(ds_play_store$Reviews)
[1] 2905052
```

```
> summary(ds_play_store$Reviews)
      Min. 1st Qu.  Median   Mean 3rd Qu.   Max.
      1    164   4714  472776  71267 78158306
> min(ds_play_store$Reviews)
[1] 1
> max(ds_play_store$Reviews)
[1] 78158306
> mean(ds_play_store$Reviews)
[1] 472776.4
> median(ds_play_store$Reviews)
[1] 4714.5
> sd(ds_play_store$Reviews)
[1] 2905052
```

Based on the statistical summary, the least reviewed value for an app was 1. Based on the dataset, many apps have been reviewed only once. On the other hand, the most reviewed for an app is

78,158,306 which belonged to Facebook app. The mean of the reviewed value was 472,776 which means users do not leave reviews that much. Also, as we can see the datapoint for 1st Qu. is farther away from the median than 3rd Qu. is from the median, so we can say that there is a greater dispersion among the smaller values of the dataset than among the larger values.

Lastly the standard deviation is 2,905,052 which is a high value. This indicates that our values in Reviews variable are spread out over a wider range.

➤ **Statistical Summary of Price**

```
> summary(ds_play_store$Price)
   Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
 0.0000   0.0000   0.0000   0.9632   0.0000  400.0000
> min(ds_play_store$Price)
[1] 0
> max(ds_play_store$Price)
[1] 400
> mean(ds_play_store$Price)
[1] 0.9631545
> median(ds_play_store$Price)
[1] 0
> sd(ds_play_store$Price)
[1] 16.18934
```

```

> summary(ds_play_store$Price)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0000  0.0000  0.0000  0.9632  0.0000 400.0000

> min(ds_play_store$Price)
[1] 0

> max(ds_play_store$Price)
[1] 400

> mean(ds_play_store$Price)
[1] 0.9631545

> median(ds_play_store$Price)
[1] 0

> sd(ds_play_store$Price)
[1] 16.18934

```

Based on the statistical summary of Price, the min, Q1 and Q3 have a value of 0 and max value is 400. The mean value for the price is 0.96 which means a vast majority of Google Play Apps are Freely available.

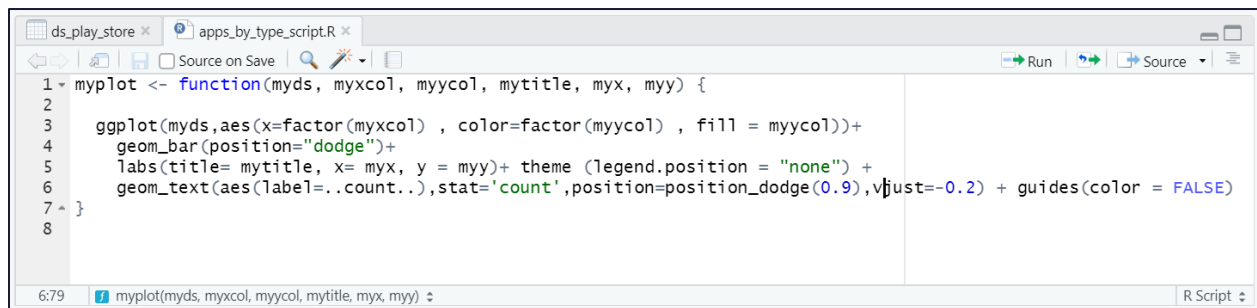
Lastly the standard deviation is 16.18 which is greater than the mean. So more of our data is clustered about the mean.

User-defined Function and Scripts

I have created 2 functions and used them in my analysis. Both functions are created and saved in script files.

1. The function below is used to create a bar plot for the first analysis - **app distribution by type**.

It is a function with input arguments of the data frame name, the x column, the y column, the graph title and the column names.



```
1 myplot <- function(myds, myxcol, myycol, mytitle, myx, myy) {  
2  
3   ggplot(myds,aes(x=factor(myxcol) , color=factor(myycol) , fill = myycol))+  
4     geom_bar(position="dodge")+  
5     labs(title= mytitle, x= myx, y = myy)+ theme (legend.position = "none") +  
6     geom_text(aes(label=..count..),stat='count',position=position_dodge(0.9),vjust=-0.2) + guides(color = FALSE)  
7 }  
8
```

```
myplot <- function(myds, myxcol, myycol, mytitle, myx, myy) {  
  
  ggplot(myds,aes(x=factor(myxcol) , color=factor(myycol) , fill = myycol))+  
  
  geom_bar(position="dodge")+  
  
  labs(title= mytitle, x= myx, y = myy)+ theme (legend.position = "none") +  
  
  geom_text(aes(label=..count..),stat='count',position=position_dodge(0.9),vjust=-0.2) +  
  guides(color = FALSE)  
  
}
```

The workspace was pointed to the appropriate directory:

```
> setwd("C:/Users/Ninel/Desktop/School/CSULA/CIS 5250/R Labs/R Scripts")
```

Next, the source function was used to load the “script file and then the “**myplot**” function was called to create the plot.

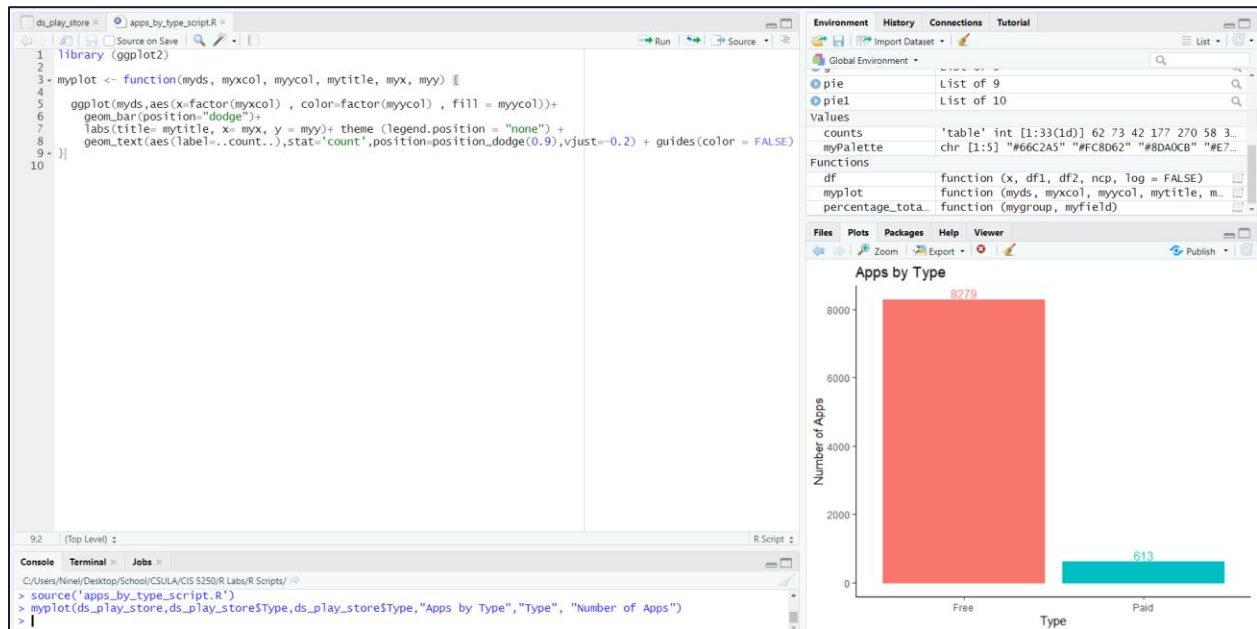


```
> source('apps_by_type_script.R')  
> myplot(ds_play_store,ds_play_store$Type,ds_play_store$Type,"Apps by Type","Type", "Number of Apps")  
>
```

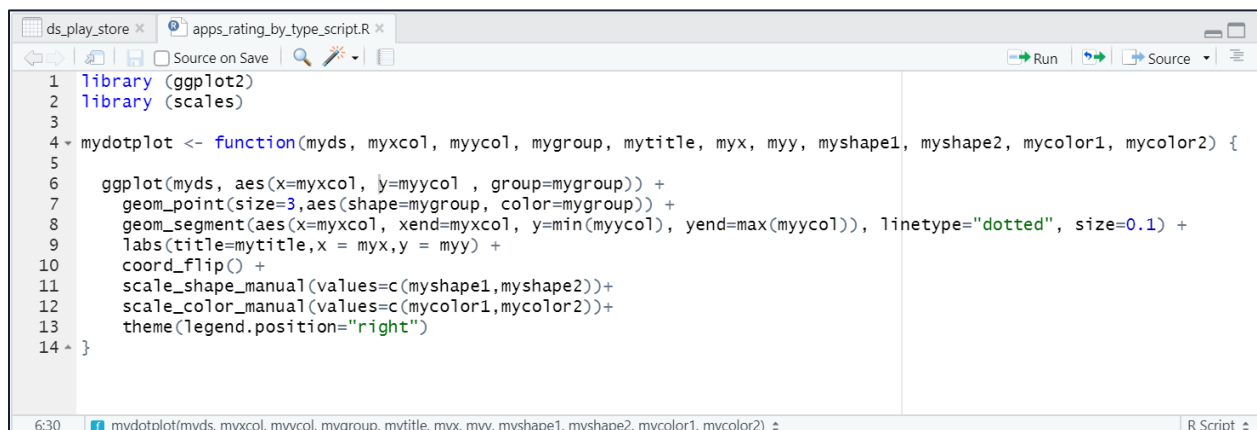
```
> source('apps_by_type_script.R')

> myplot(ds_play_store,ds_play_store$Type,ds_play_store$Type,"Apps by Type","Type",
"Number of Apps")
```

If we call the function *myplot()* with these arguments, it will create a bar plot as shown below:



- The function below is used to create a dot plot for the fourth analysis - **content rating vs. rating by type**. It is a function with input arguments of the data frame name, the x column, the y column, the group, the graph title, the column names, the shapes and the colors.



```

library (ggplot2)

library (scales)

mydotplot <- function(myds, myxcol, myycol, mygroup, mytitle, myx, myy, myshape1,
myshape2, mycolor1, mycolor2) {

  ggplot(myds, aes(x=myxcol, y=myycol , group=mygroup)) +

    geom_point(size=3,aes(shape=mygroup, color=mygroup)) +

    geom_segment(aes(x=myxcol, xend=myxcol, y=min(myycol), yend=max(myycol)),
linetype="dotted", size=0.1) +

    labs(title=mytitle,x = myx,y = myy) +

    coord_flip() +

    scale_shape_manual(values=c(myshape1,myshape2))+

    scale_color_manual(values=c(mycolor1,mycolor2))+

    theme(legend.position="right")

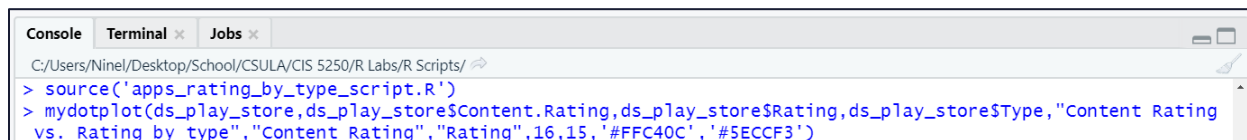
}

```

The workspace was pointed to the appropriate directory:

```
> setwd("C:/Users/Ninel/Desktop/School/CSULA/CIS 5250/R Labs/R Scripts")
```

Next, the source function was used to load the “script file and then the “**mydotplot**” function was called to create the plot.



```

Console Terminal Jobs
C:/Users/Ninel/Desktop/School/CSULA/CIS 5250/R Labs/R Scripts/
> source('apps_rating_by_type_script.R')
> mydotplot(ds_play_store,ds_play_store$Content.Rating,ds_play_store$Rating,ds_play_store$Type,"Content Rating
vs. Rating by type","Content Rating","Rating",16,15,'#FFC40C','#5ECCF3')

```

```

> source('apps_rating_by_type_script.R')

>mydotplot(ds_play_store,ds_play_store$Content.Rating,ds_play_store$Rating,ds_play_store$Type,"Content Rating vs. Rating by type","Content Rating","Rating",16,15,'#FFC40C','#5ECCF3')

```

Conclusion

My goal was to find out main factors that affects user's decision to download and install an app.

Although the analysis done above was not very deep, I concluded the following:

- Google Play Store contains apps that are mostly free to use.
- Play Store has a high number of apps from Game and Communication categories with highest number of downloads and reviews.
- Average number Reviews on an app is below 500,000.
- The most expensive app is 400 dollars.
- More than 6000 apps have ratings higher than 4.0.