*Hello friends*,

My name is Ravuri Maheshbabu, In this notebook, I'm gonna analyze Google Play Store data. While I was analyzing the data, I used R.

*Let’s starts the data analysis,*

**CONTENT**

1. [Introduction to data](https://www.kaggle.com/ecemboluk/google-play-store-analysis#1)
2. [Cleaning Data](https://www.kaggle.com/ecemboluk/google-play-store-analysis#2)   
        A. [Category](https://www.kaggle.com/ecemboluk/google-play-store-analysis#a)   
        B. [Rating](https://www.kaggle.com/ecemboluk/google-play-store-analysis#b)   
        C. [Reviews](https://www.kaggle.com/ecemboluk/google-play-store-analysis#c)   
        D. [Size](https://www.kaggle.com/ecemboluk/google-play-store-analysis#d)   
        E. [Installs](https://www.kaggle.com/ecemboluk/google-play-store-analysis#e)   
        F. [Price](https://www.kaggle.com/ecemboluk/google-play-store-analysis#f)   
        G. [Last Updated](https://www.kaggle.com/ecemboluk/google-play-store-analysis#g)
3. [Exploratory Data Analysis](https://www.kaggle.com/ecemboluk/google-play-store-analysis#3)

# 1. INTRODUCTION TO DATA

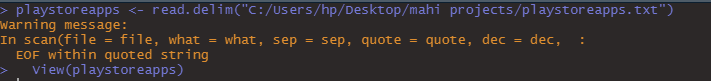
Before, get into the first step of the process , we need install and call the all required libraries.

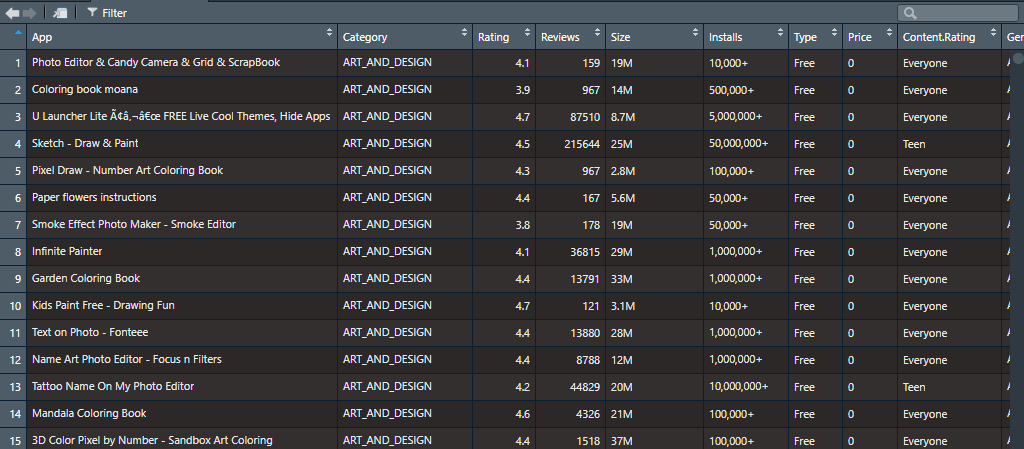
1. Install.packages(“dplyr”) , library(dplyr)
2. Install.packages(“ggplot2”) , library(ggplot2)
3. Install.packages(“corrgram”), library(corrgram)
4. Install.packages(“corrplot”), library(corrplot)

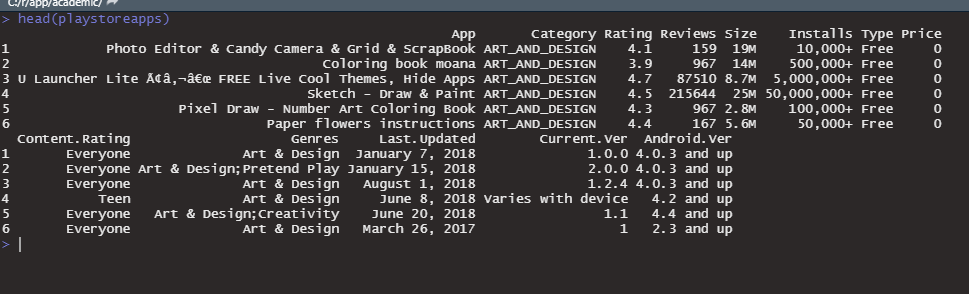
Let’s start the first step of this process, Firstly let's get to know data. While I was analyzing the data, I used Rstudio.

* Summary() : To know about the total summary of the data table.
* Head() : It returns the top six rows from data table.
* Tail(): It returns the below six rows from data table.
* Dim(): it returns the dimension of the data frame.
* nrows(): it gives number of rows in data table
* ncol(): it gives number of columns in data table.

Firstly, we need to import the dataset to Rstudio. Use below command to import the dataset





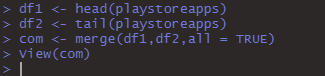


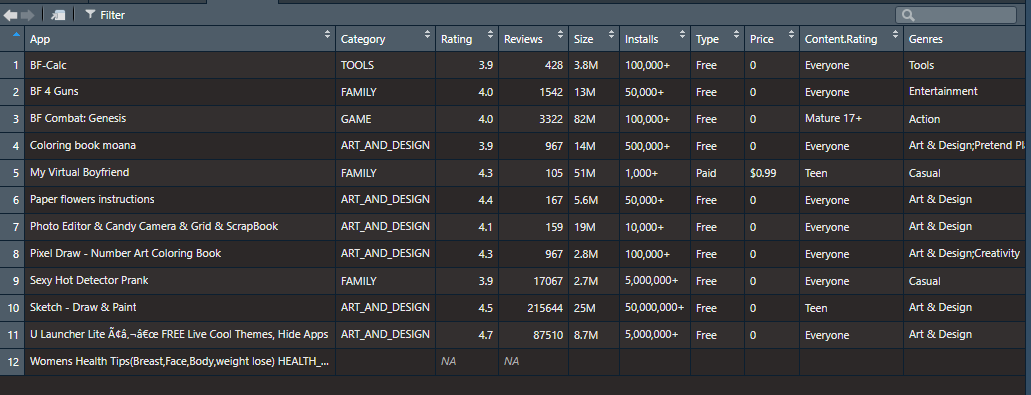






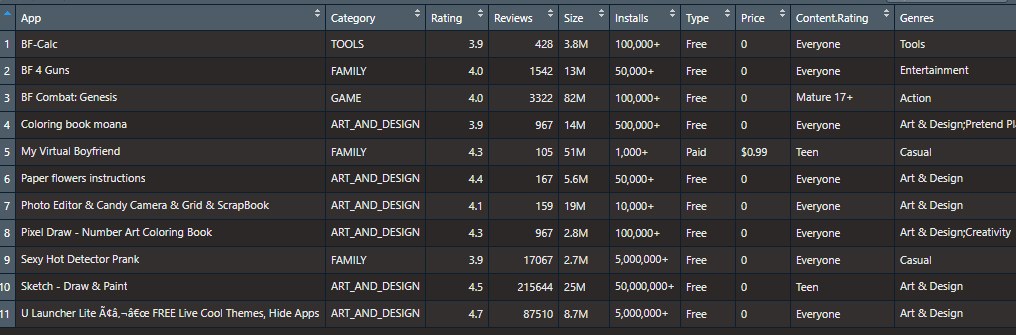
Let’s combine the head and tail datasets





The last row contains Na values,so, we can remove the row. Use below command.



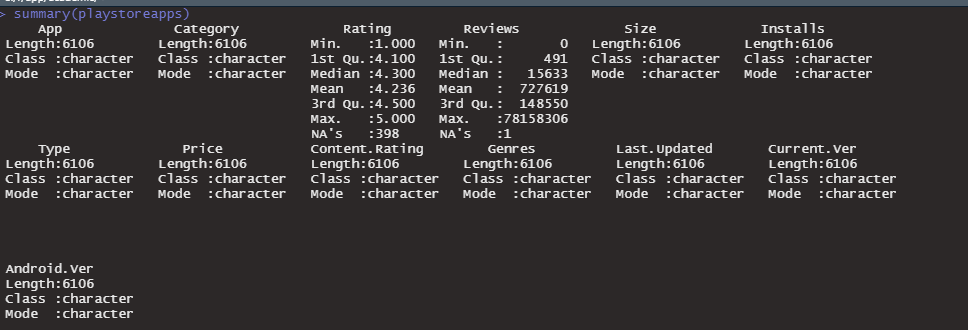


# 2. Cleaning Data

This is the second step to analyze the particular dataset,

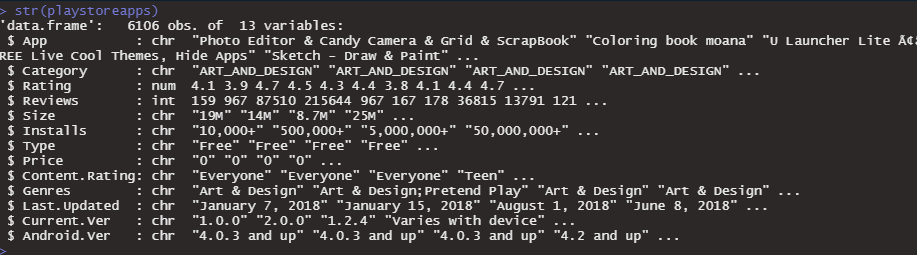
Let’s have some fun;

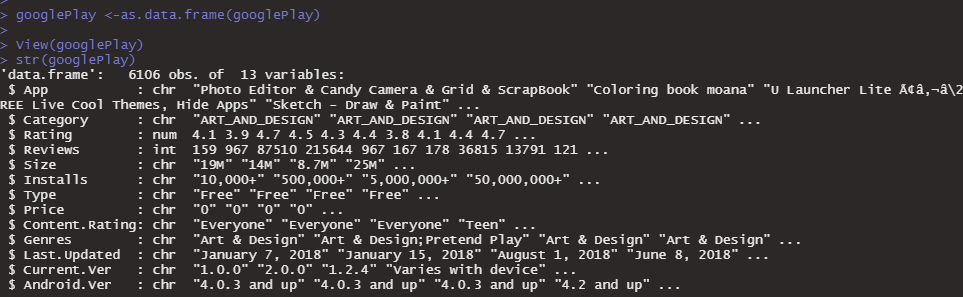
Firstly, lest’s look into the summary of dataframe

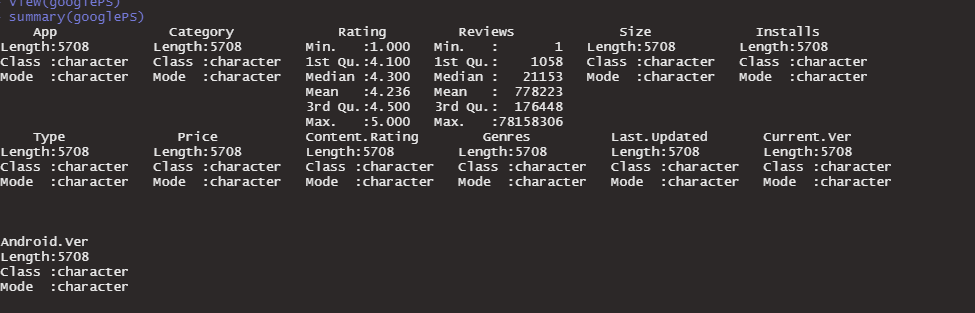


In this, we have “NA” values and some irrelevant data.so, let’s start the data cleaning for better visualization.

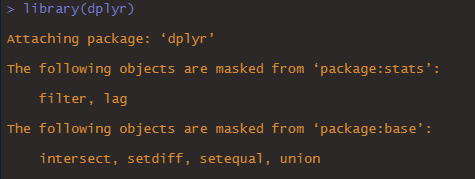
Firstly, we remove the “NA” values in Rating and Review.







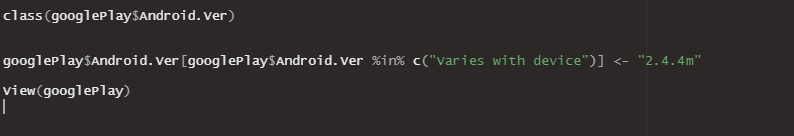
We removed the “NA” values from our dataframe. Let’s clean the irrelevant data in Size,Current.ver and Andriod.ver columns.





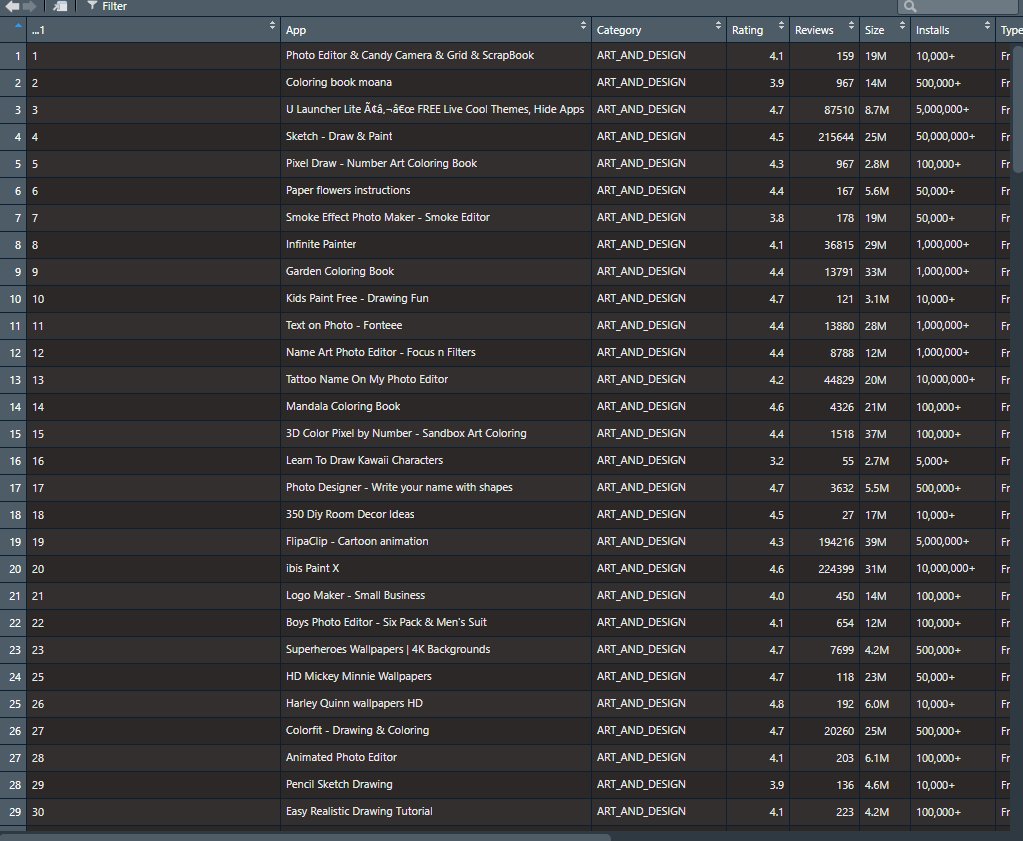






Apply above all commands, our dataset is ready to analyze. Now once look into dataframe . I rename the name of the dataframe as mahiexel instead of googleplay.

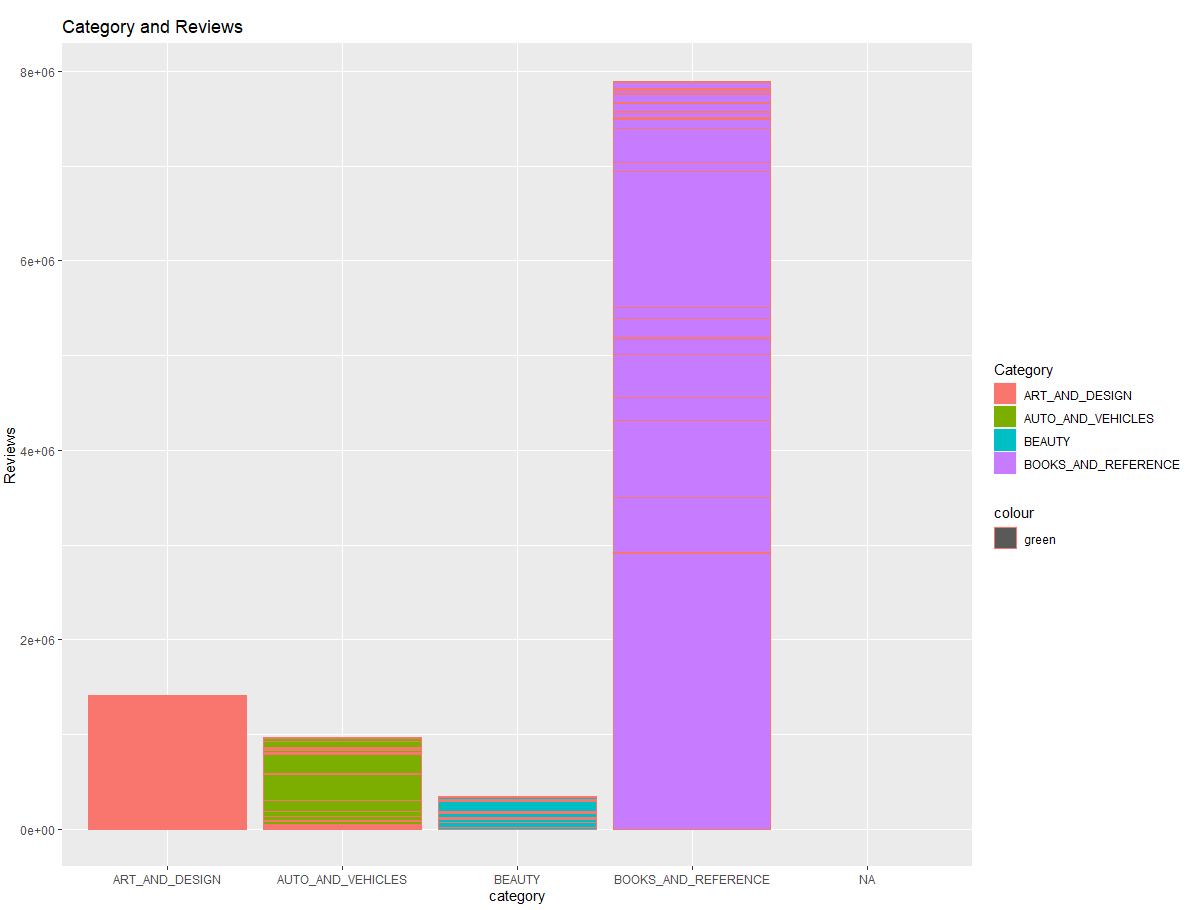




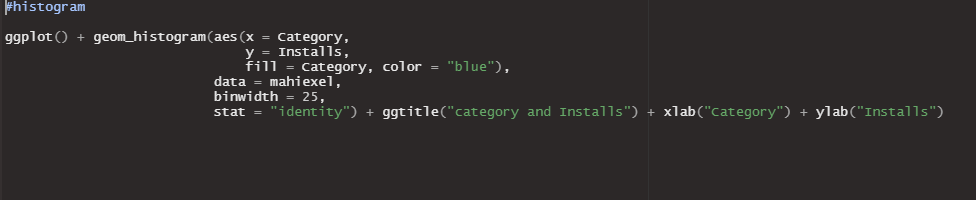
# 3.Exploratory Data Analysis

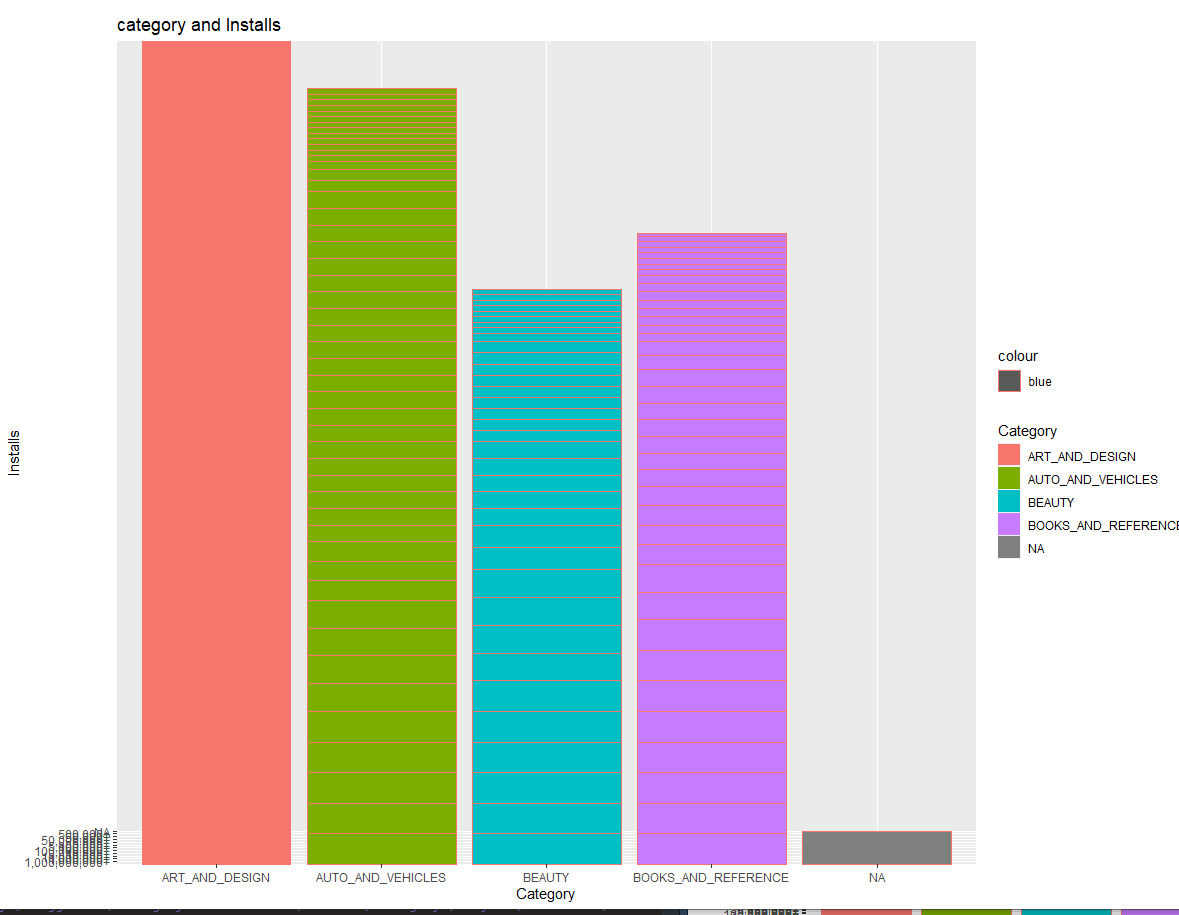
## Category and Reviews





## Category and Installs

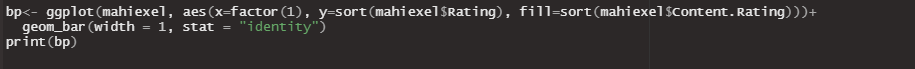


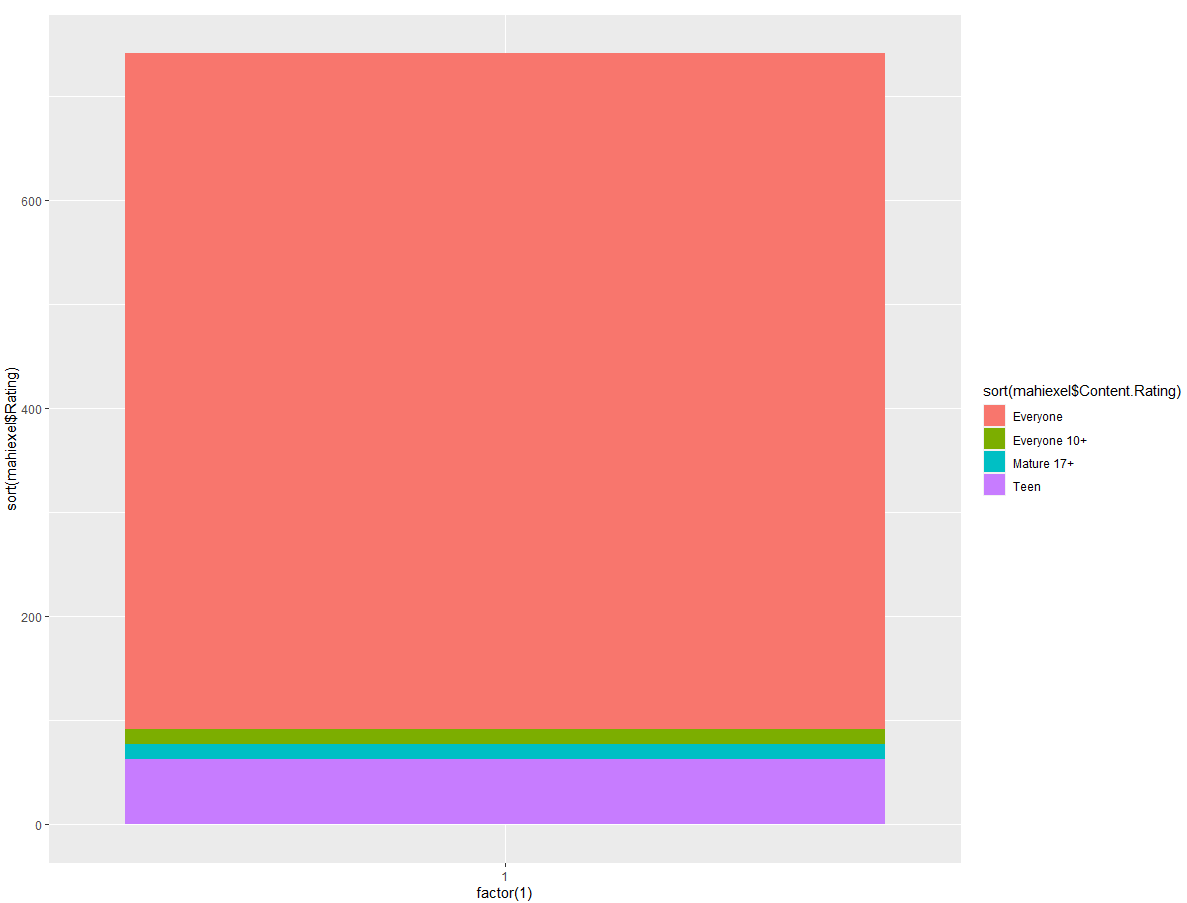


## Content Rating

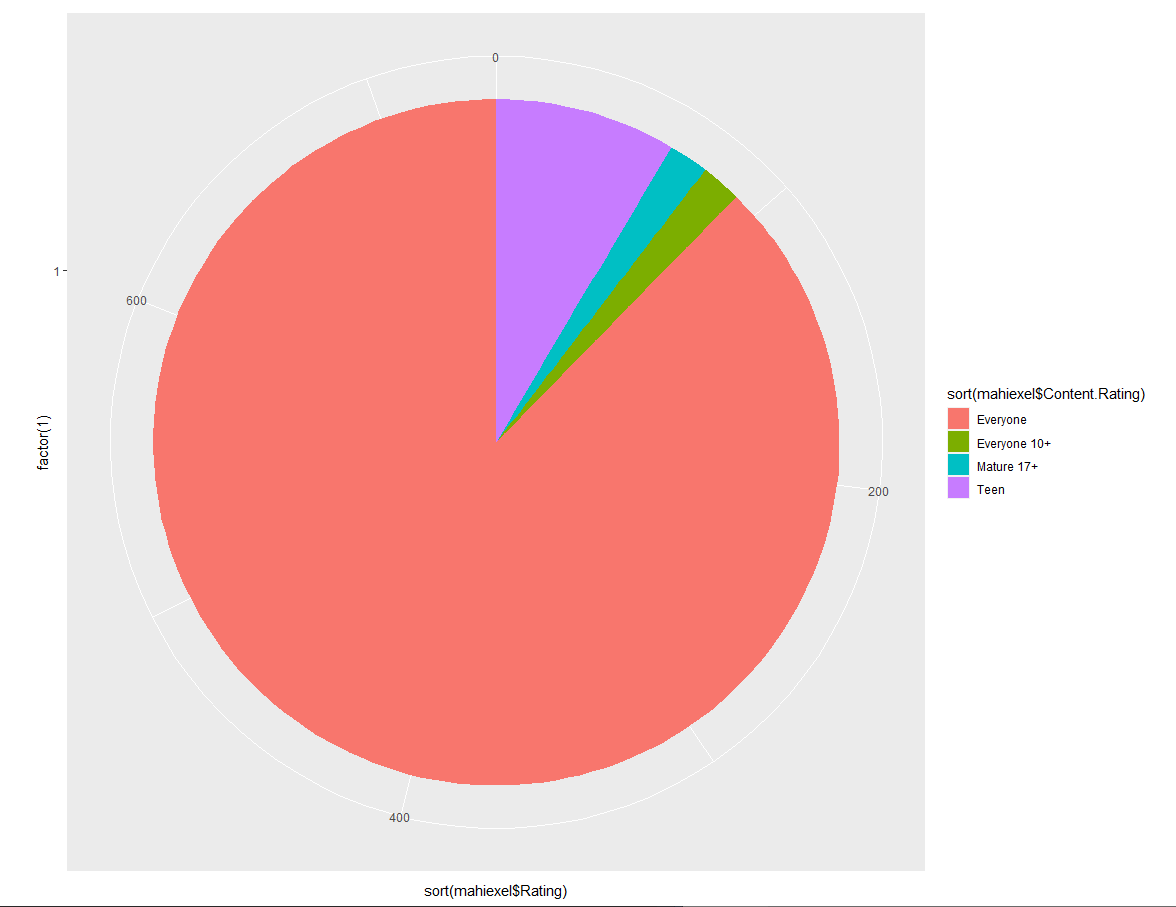
#pie chart

To make a pie chart, first we need to create a stacked bar plot. Then transfer the bar plot to pie chart.

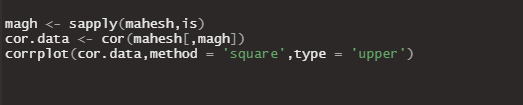


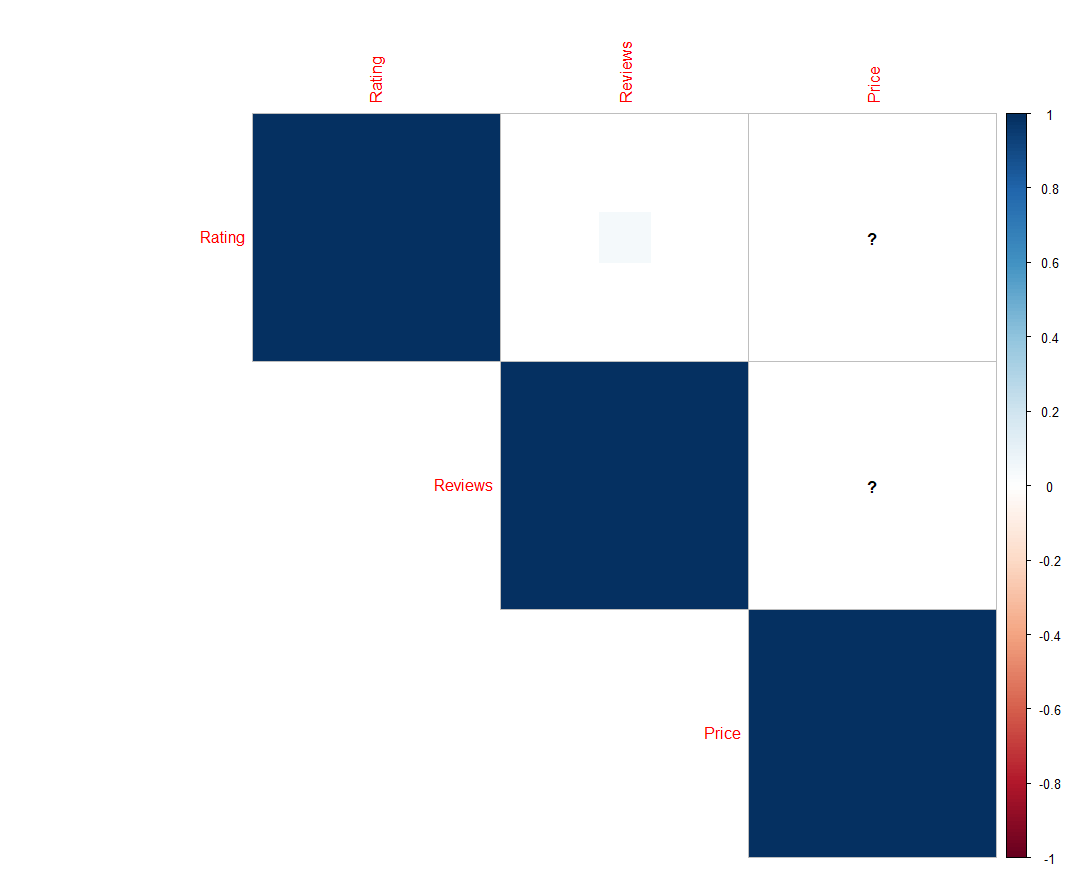






#corrplot graph





# Conclusion

This is the end of the story. You can visualize with a lot of different models. It's important for ML models. Above three steps are the key process of data analysis.

**THANK YOU ALL!**