**Google Play Store Apps**

**Data Description:**

Data Analysis on GooglePlayStore which contains 13 variables where Name, Current Ver, Android Ver are not quiet used to analyze the user requirements, by excluding those variables we have 10 variables.

Those 10 variables are classified into two types, namely:

1. Continuous variables:

* Rating
* Reviews
* Size
* Price

1. Categorical variables:

* Category
* Installs
* Type
* Content Rating
* Genres
* Last Updated

**Data Analysis:**

* Read the csv file of GooglePlayStore into a data frame called “data”.
* By analyzing the numeric variables, let us analyze Rating.
* The max of Rating is 19,as we all know the Rating should be from 1 to 5, so the data with Rating max 19 is stored in a dataframe named “corrupted”.
* Removing the corrupted data from the data using subset function. Now the max is 5.
* Ggplot of rating indicates the median in 4.3 and mean in 4.1
* It is clear that most of the apps in the dataset is free, so the ggplot shows linear line at price zero and has no much variety in plot.