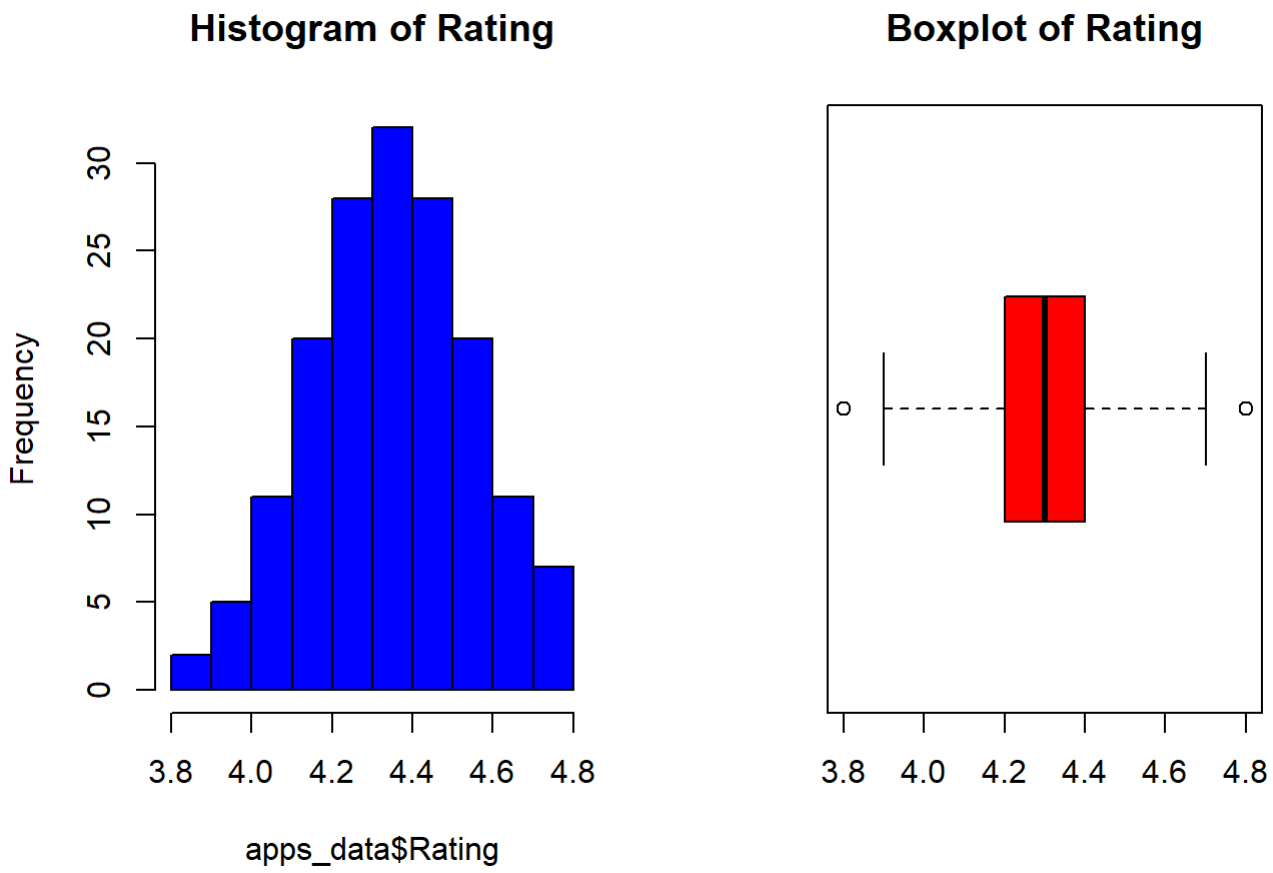
PROJECT 1

Svetlana Voda

10/11/2022

TASK1

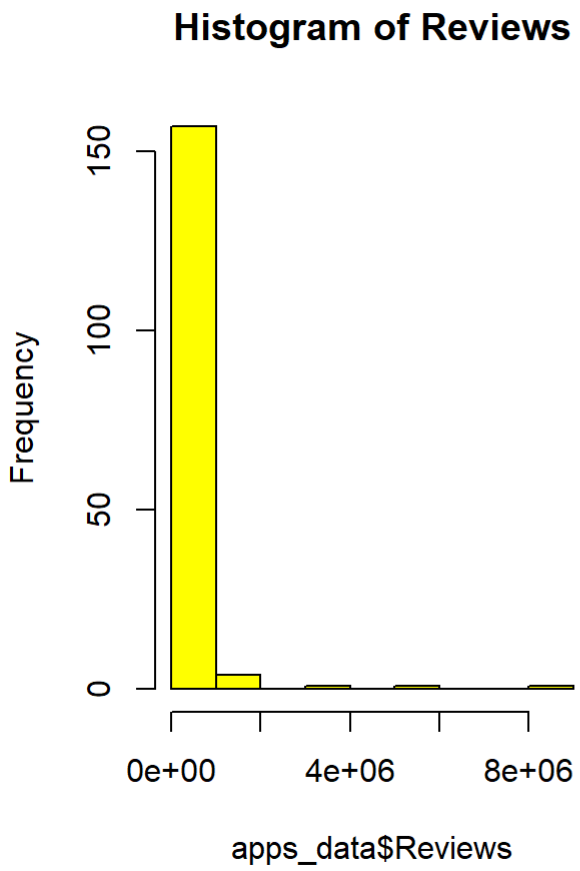
Histogram and Boxplot of Rating

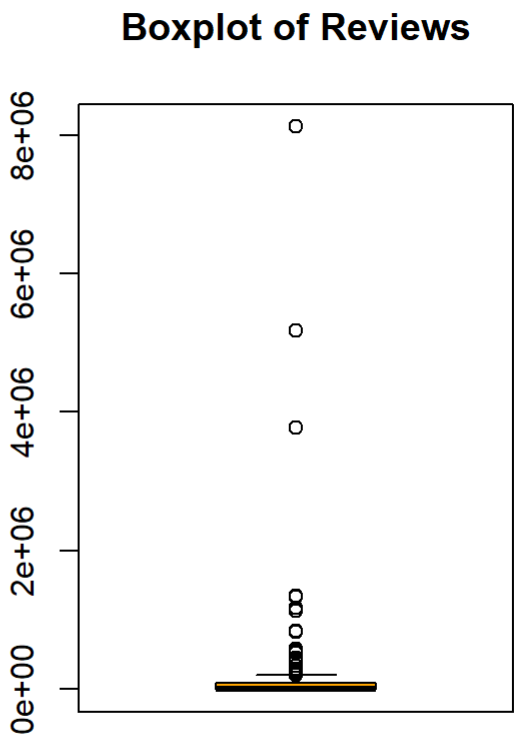


Analyzing histogram and boxplot of quantitative variable Rating we can describe our data as symmetric (bell-shaped) with one peak and two outliers, where we can conclude that data is Normal distributed.The center of the data is represented by equal mean, median, and mode.The median and mean are equal 4.3 . The standard deviation of variable Rating is 0.2033462.The range of the data is 3.8, 4.8, and the difference between range is 1. The IQR is 0.2

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     3.8     4.2     4.3     4.3     4.4     4.8
```

Histogram and Boxplot of Reviews





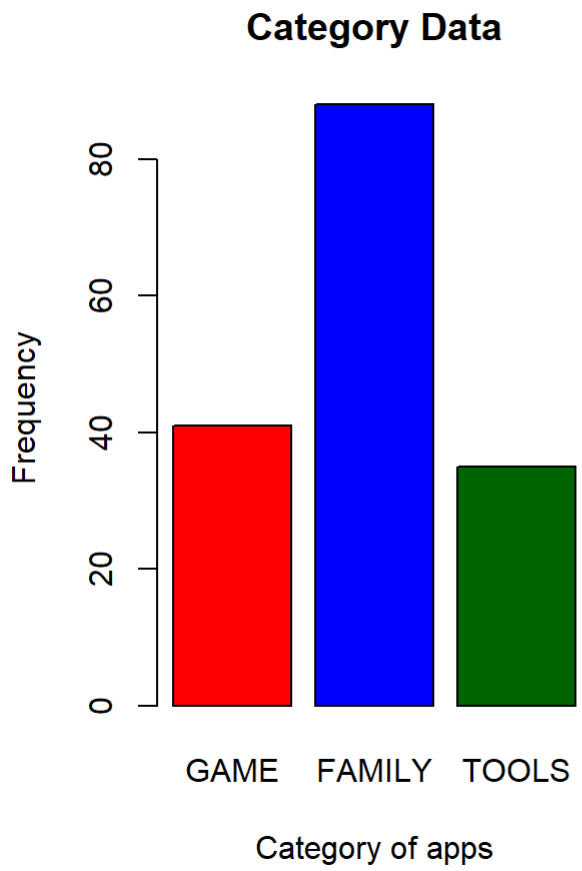
Analyzing histogram and boxplot of quantitative variable Reviews we can describe as Right-Skewed with 3 outliers. The value for the center of the data is median. So, we can conclude that data is not Normal distributed. From out descriptive statistic we can see that the median is 7306.5 and mean is 2.0411059^{5}. The standard deviation of variable Reviews is 8.2426647^{5}. The range of the data is 4 and 8118609 with difference between range 8118605. The IQR is 82895.5 .

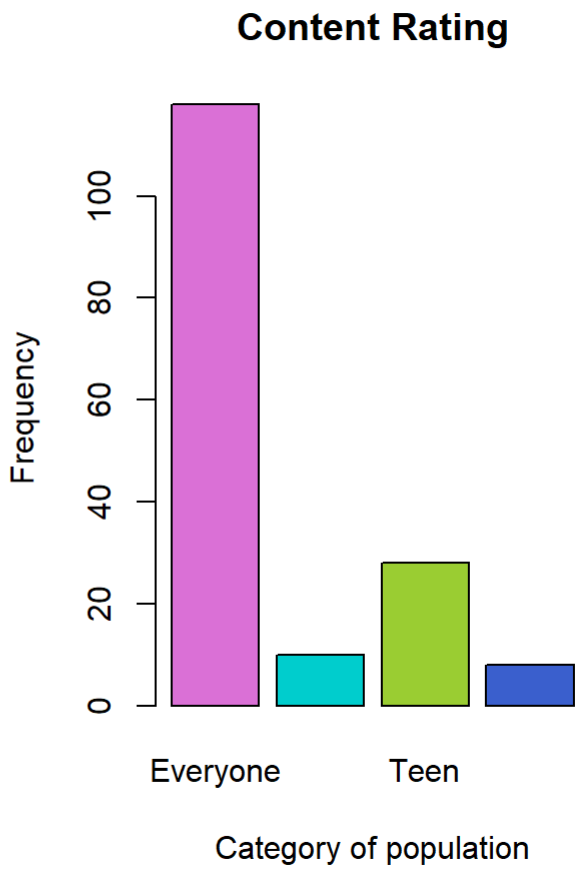
min	Q1	median	Q3	max
<int>	<dbl>	<dbl>	<dbl>	<int>
4	509	7306.5	83404.5	8118609

1 row

TASK2

Barplot of Category and Content Rating





GAME	FAMILY	TOOLS
41	88	35

	min <int>	Q1 <dbl>	median <int>	Q3 <dbl>	max <int>
TOOLS	35	35	41	88	88

1 row

Analyzing bar plot of app Category we compares the frequency of different category of apps. Family category are the most popular types of apps, followed by Game, and Tools. The frequency range from about 35 to 88. The median its come to Game department. Least popular is Tool category.

Everyone	Everyone 10+	Teen	Mature 17+
118	10	28	8

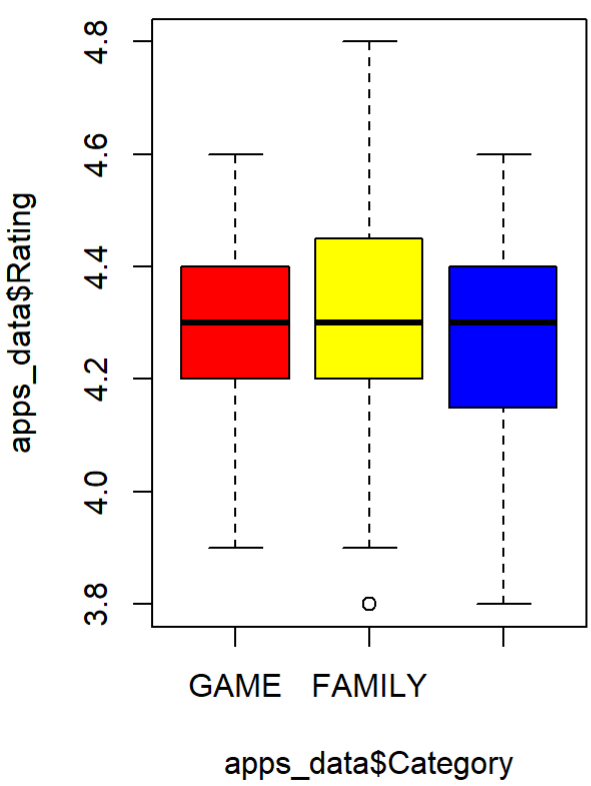
	min <int>	Q1 <dbl>	median <dbl>	Q3 <dbl>	max <int>
Mature 17+	8	9	19	73	118

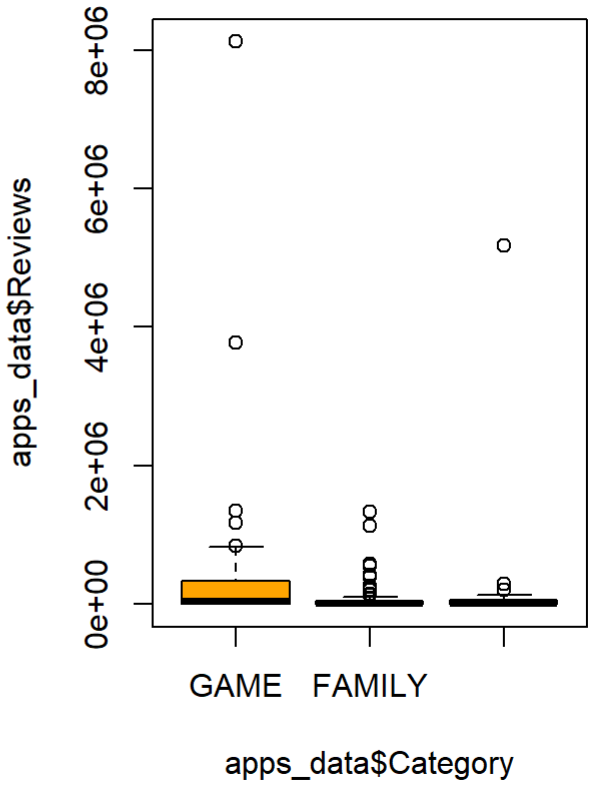
1 row

Analyzing bar plot of age Category the most common group of age using Google Play Store is Everyone followed by Teen,Everyone 10+, and Mature 17+. The frequency range from about 8 to 118. The median statistic is between Everyone 10+ and Teen.

TASK3

Boxplot of Rating and Reviews Category





Analyzing and comparing Boxplot of Rating and Reviews by Category(Game, Family and Tools) we can see that the medians on Rating is equal 4.300, the range is almost the same. Outlier is present only in Family category. In contrast to Rating, in Reviews Category we see huge discrepancy between median, mean , and presents of multiple outlier. So we can conclude that this Category are vary from each other.

Table of Rating

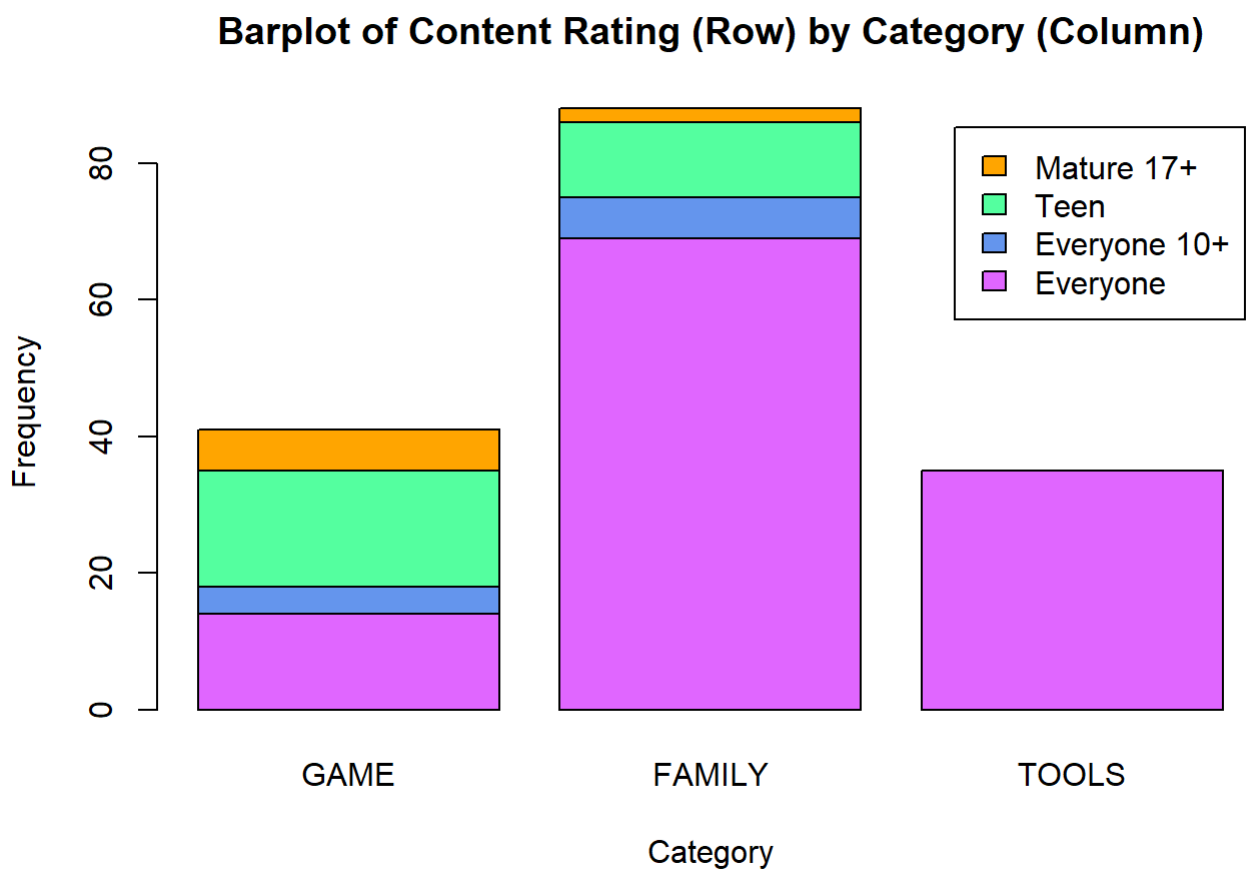
```
## $GAME
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.900   4.200   4.300   4.293   4.400   4.600
##
## $FAMILY
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.800   4.200   4.300   4.315   4.425   4.800
##
## $TOOLS
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.800   4.150   4.300   4.271   4.400   4.600
```

Table of Review

```
## $GAME
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      14    4878   43055  480221  336386 8118609
##
## $FAMILY
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##       4.0     314.5    3510.5   83313.2   39755.0 1333338.0
##
## $TOOLS
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       9     500    3446  184386   61258 5180480
```

TASK4

Barplot of Content.Rating by Category

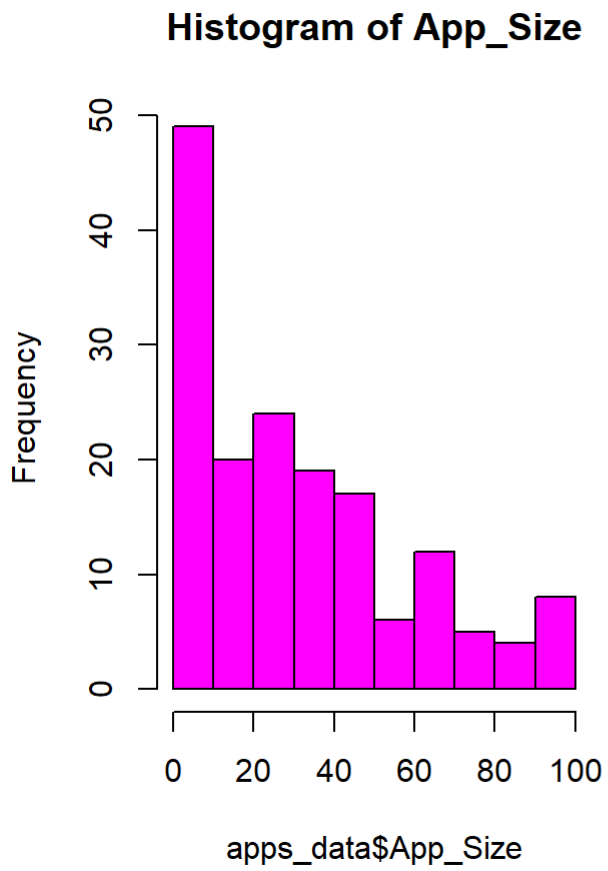


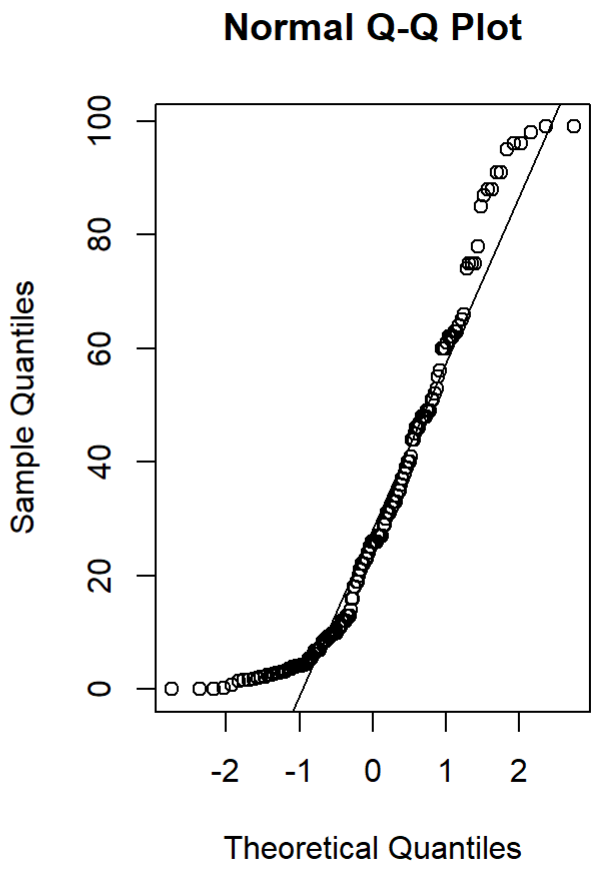
Looking at our stacked bar chart, we clearly see, for example, that Tools Category is the least choosing , and this is mainly because the age groups such as Teen, Mature 17+ , and Everyone where not interested on this Category. In contrast the stacked bar chart category Family for group Everyone is the most effective, has the most number of apps. On Category Game the most popular is for group Teen followed by Everyone.

	FAMILY	GAME	TOOLS
Everyone	69	14	35
Everyone 10+	6	4	0
Mature 17+	2	6	0
Teen	11	17	0

TASK5

Histogram and Q-Q Plot of App-Size.



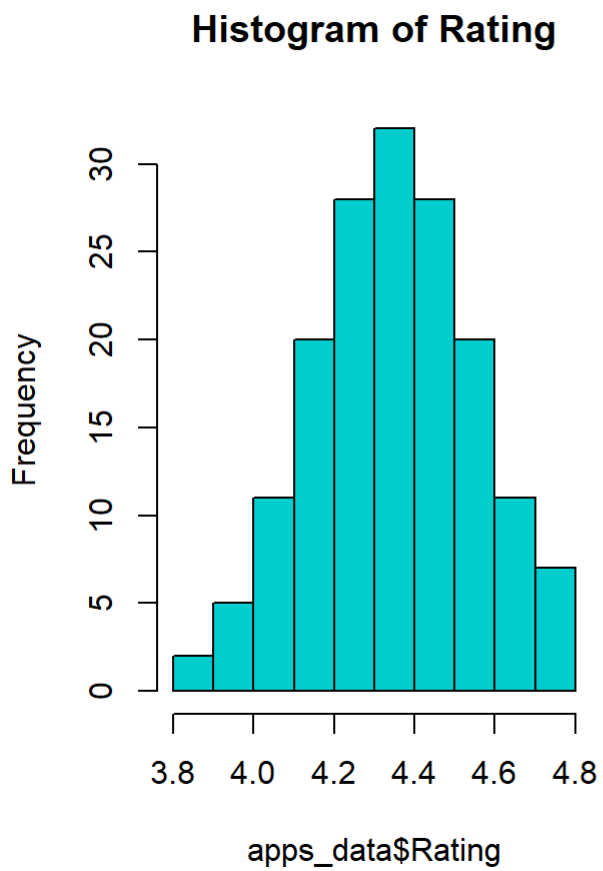


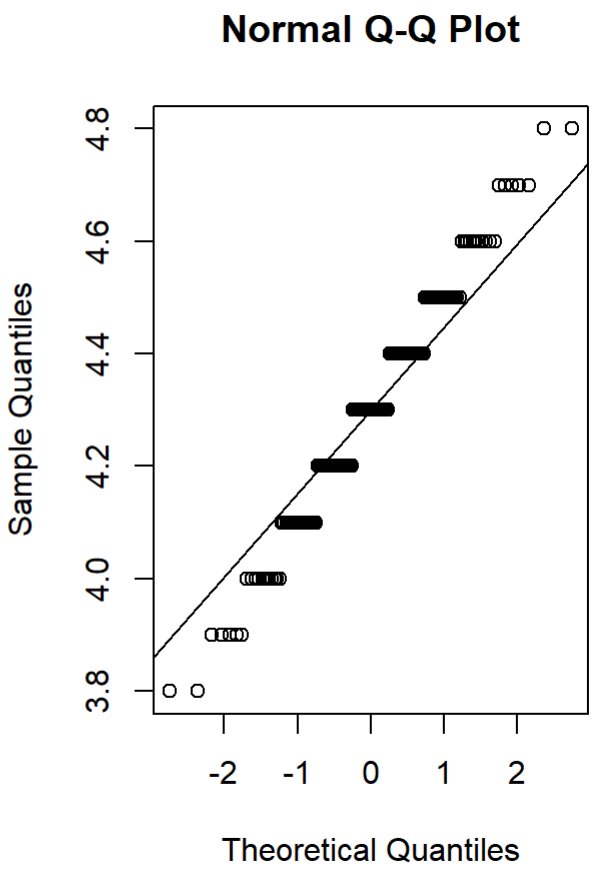
From our graphs, we can conclude that our data is not normal distributed, because histogram look Right-Skewer. There are issues with Q-Q Plot, it does not follow reference line very well.There are issues on the bottom left and top right. The skew is 0.8831487 and is far from zero. It's not symmetric. The Kurtosis is 2.8843481 and is far from 3 which mean the data is not normal.

```
##
##  Shapiro-Wilk normality test
##
## data:  apps_data$App_Size
## W = 0.89667, p-value = 2.633e-09
```

Also, from Shapiro-Wilk Test we can see that the p_value is very small and we can conclude that Normality does not seem reasonable.

Histogram and Q-Q Plot of App-Size.





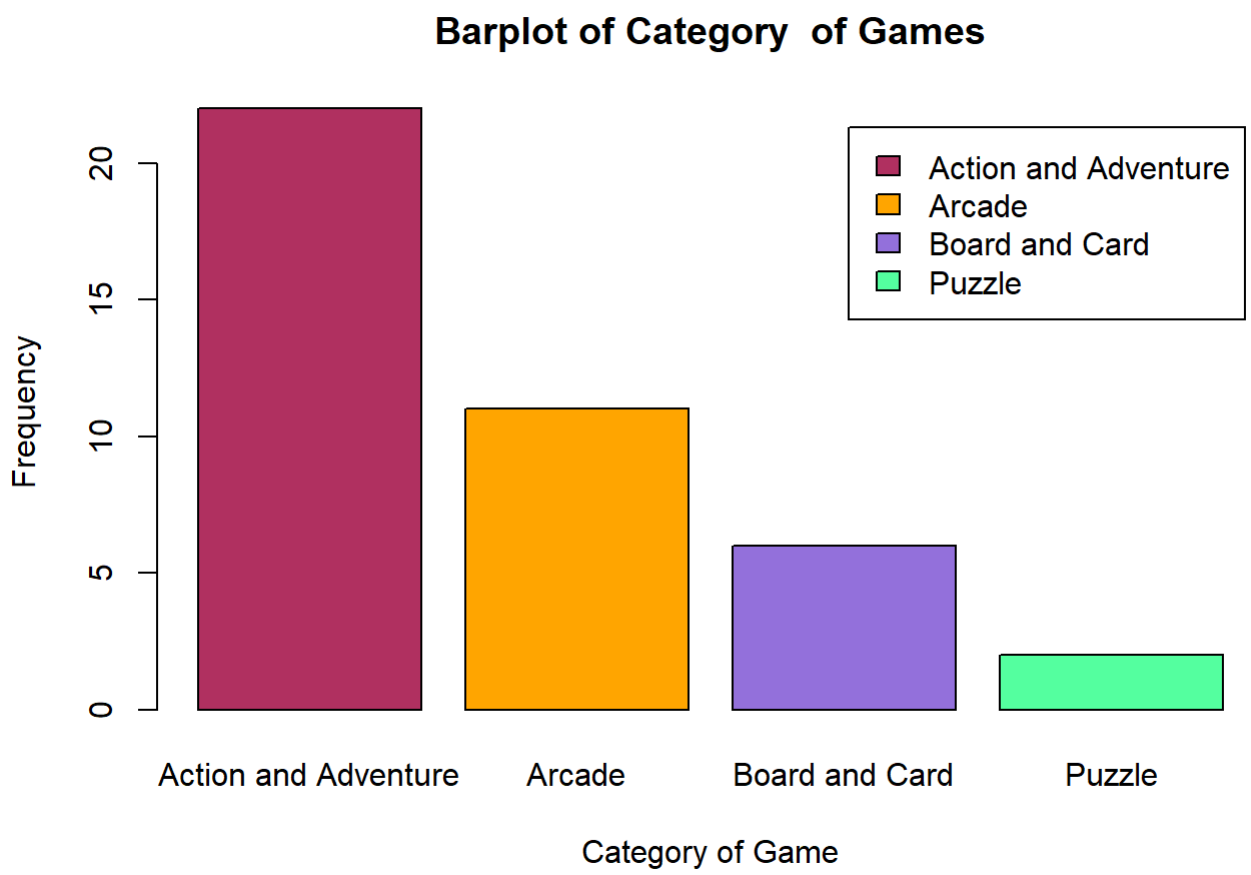
From our graphs, we can conclude that our data could potentially come from a normal distribution, because histogram is somewhat symmetric. The Q-Q Plot look ok but some potential issues at the bottom and on the top. The skew is 1.296361^{-15} and is far from zero. The Kurtosis is 2.6882379 and is far from 3 which mean the data is not normal.

```
##
##  Shapiro-Wilk normality test
##
## data:  apps_data$Rating
## W = 0.97804, p-value = 0.01041
```

Also, from Shapiro-Wilk Test we can see that the p_value is very small and we can conclude that Normality does not seem reasonable.

TASK6

Barplot of Genre of apps.

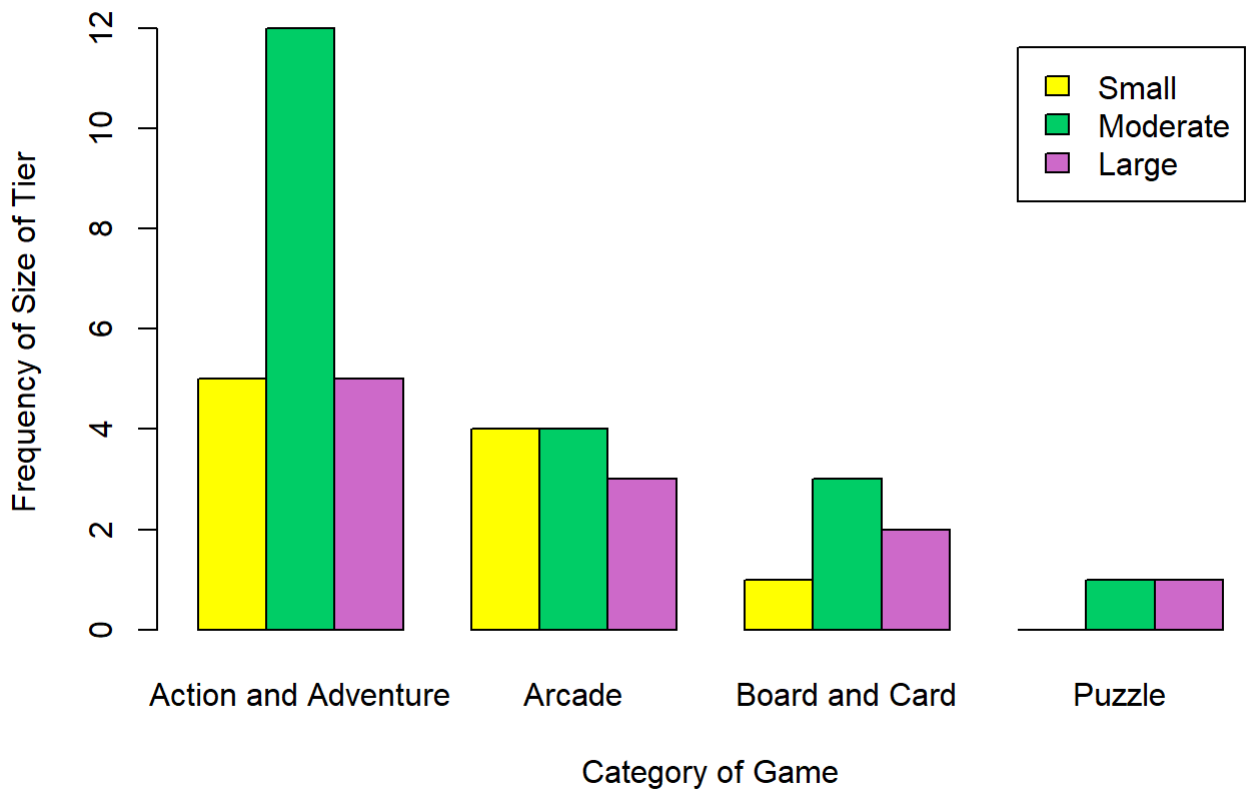



```
  Action and Adventure | Arcade       | Board and Card  | Mature 17+
-----------------------|------------- |-----------------|-----------
  22                   | 11           | 6               | 2
```

From the graph the genre of Game Category we can see that “Action and Adventure” there are most popular types of Game App in Google Play Store with frequency of 22. Followed by “Arcade and Board” with frequency 11, and “Card”. The least popular is “Puzzle” with frequency of 2.

TASK7

Barplot of Size_Tier by Genre Apps



	Action and Adventure	Arcade	Board and Card	Puzzle
Small	5	4	1	0
Moderate	12	4	3	1
Large	5	3	2	1

By Analyzing this bar plot we can conclude that Size_Tier varies by genre because on genre “Action and Adventure” the most efficient is frequency “Moderate” where the “Small” and “Large” are the same level. On “Arcade” genre the frequency “Small” and “Moderate” are the same but “Large” is smaller which differ extremely by genre Action and Adventure. While on “Puzzle” genre the frequency “Small” does not exist.