Week12\_FinalStep3

RamizuddinMohammedShabuddin

2/25/2021

## Loading required package: tidyverse

## Warning: package 'tidyverse' was built under R version 4.0.3

## -- Attaching packages --------------------------------------------------------------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2 v purrr 0.3.4  
## v tibble 3.0.3 v dplyr 1.0.1  
## v tidyr 1.1.1 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.5.0

## -- Conflicts ------------------------------------------------------------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

## Loading required package: GGally

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

## Loading required package: RSQLite

## Loading required package: scales

##   
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':  
##   
## discard

## The following object is masked from 'package:readr':  
##   
## col\_factor

## Loading required package: wesanderson

## Warning: package 'wesanderson' was built under R version 4.0.3

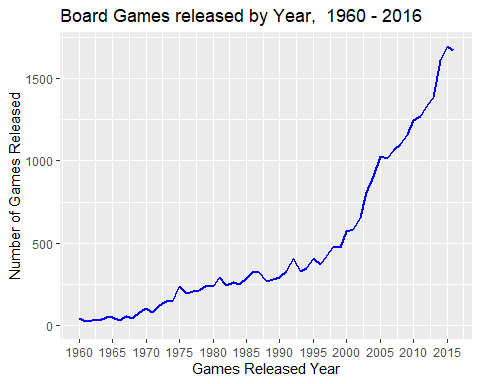
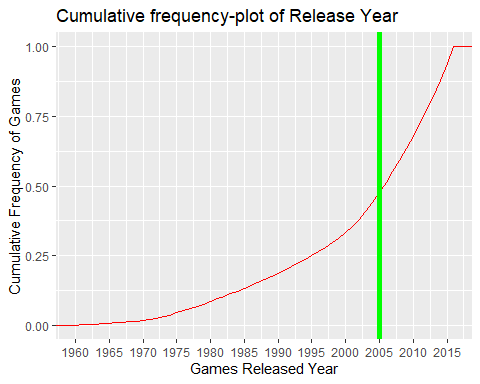
With being a parent of 2 kids, I think I spend lot of money on board games and video games. The data set which was found interesting because of the children in the family. Where we spend some dollars in picking the board games. There are multiple board games in the market. Using the data set available let’s try to find “What constitute the perfect Board Game?” As said before, It is very broad question and cannot be answered without a deep knowledge of the data itself. Because of the Eurogame Revolution, the Board Game industry is becoming interesting. There are young and talented Engineers from Europe introduced new design which has highly improved the quality of games.

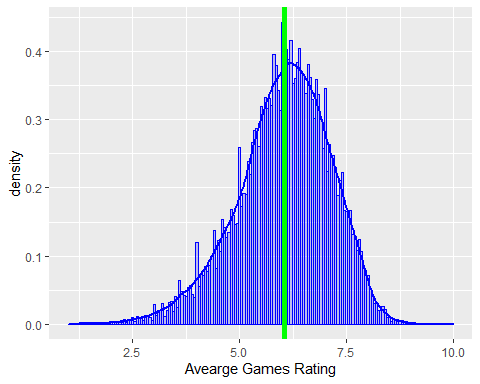
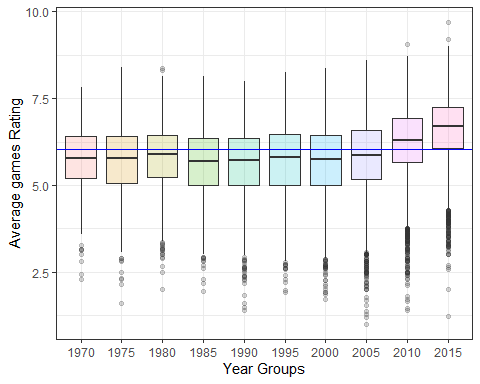
## Importing and cleaning the data

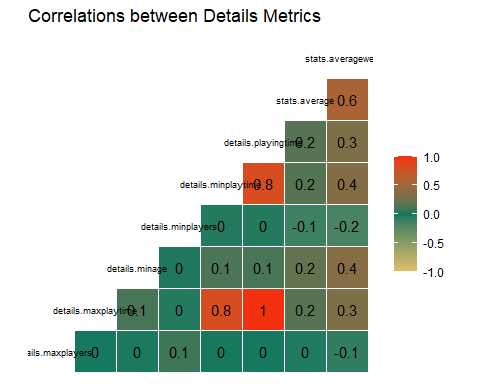
1. There are parameter which we are not planning to study. So, we can remove those parameters.
2. There are some columns which are with zero values. I am thinking of making then to N/A. So that will help me with ignoring the values when needed.
3. Then, We can remove some games which were before 1960s. Also, it would be great to remove the games which are not yet released.

As we were planning, lets see how the gaming industries have improved over the years. With the gaming data available, let plot a graph to see how the number of games released per year. The graphs below shows the number of games were very less until year 2000, after that the rapid growth was noticed.

## `summarise()` ungrouping output (override with `.groups` argument)

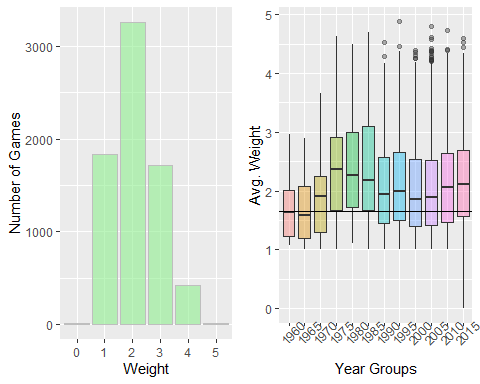
 The gaming marked exploded to a greater number. If we check for the median, we noticed, its is the year-2005, where the gaming industry was growing drastically. 

The graph below shows that the average gaming rating follows the bell curve. The rating is distributed normal from 1 to 10.  The graph below box graph shows that, the new games started to be emerged around 2010. The box plot with the line graph shows the comparison between the years group with the average.  Here, we are trying to make the filter of user rated and to build a simple correlation plot below

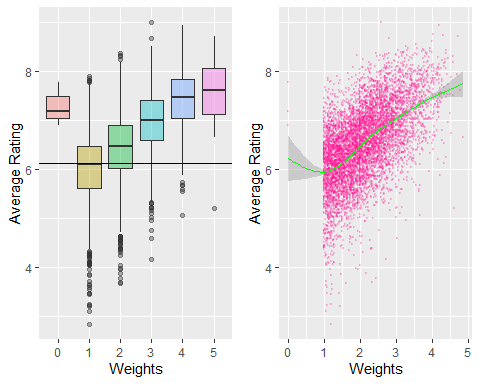
 using the correlation above, we see the Playingtime, maxplaytime and minplaytime are strongly related.

Another empirical knowledge about Board Games says that, if we exclude Party Games, games that involves many players, tend to last longer. There’s no correlation between maxplayers and maxplayingtime.

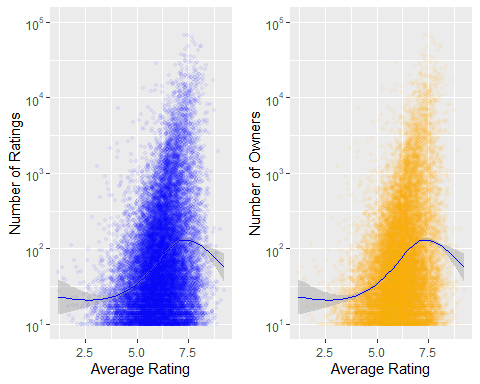
The below histograph shows that, most common games recorded in the database are definitively simple (Weight = 2).We can assess the quality and the popularity of those games.For the second col, we are trying to find the average weight using the year group.



## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

 Popularity and Ratings: To check if the BoardGameGeek has the reputation of being a competent community. we can find the correlation between game rating and game popularity. We have filtered the games with a rating of less than 10. We will consider the games as non-popular.

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'  
## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'



The graphs above shows that, there is a positive correlation between the Number of Rating and the number of Owners. If there are more better rating for the games, people tend to own it.

## Discuss the limitations of your analysis and how you, or someone else, could improve or build on it.

With the large number of data columns we could have attempted to research more on the other columns. Which was not done in this project step.