

# Data Dictionary For Board Games Data

Lauren Grove, Crystal Kim, Nate Ren

```
board_games_data <- read.csv("./board_games.csv")
```

## Data Overview

This is data from BoardGameGeek used in TidyTuesday in 2019. This data contains every board game released from 1950 to 2016 that have at least 50 ratings in order to follow along the FiveThirtyEight article about board games which mentions the golden age of board games started after 1950. This data frame has 10,532 rows with 22 variables.

## Research Question

### Research Question 1

Is there a relationship between average\_rating and average\_playtime?


### Variables Needed

- average\_rating
- average\_playtime


### Methodology

We need to create a new variable average\_playtime using mutate with max\_playtime and min\_playtime and then we can create a scatter plot to visualize this data.


## Research Question 2

Between, playtime, number of users rated, and max players, what has the most impact on average rating. 

### Variables Needed


- ~~max\_playtime~~
- ~~min\_playtime~~
- users Rated 
- max\_players
- min\_players
- average\_rating

### Methodology


Utilizing ggplot and faceting function to create box plots to plot the relationship between all the variables and average\_rating. 

## Data Cleaning Required


### New Variable using pivot\_wider to get game duration range

- Create a new variable game duration by finding the range of playing time and then use pivot\_wider to have separate columns for each duration range. 


### Split Category data by the commas to remove secondary categories

- Separate category data by splitting with commas to get the primary category 

### Combining two columns


- Combine artist and designer into one column and make a distinction with / 

## Removing Unnesccary columns


- Filter out image and thumbnail variables 

## Clean Variable Names and Values


### Variable names

- Change variable names to camel case or some other case 

### Clean NA Values

- Variables artist, category, compilation, designer, expansion, family, mechanic, and publisher have NA values.
  - Remove rows with NA values in category and publisher 

### Change Variable Values

- Recode compilation and expansion values to Yes or No or Yes expansion or No expansion
- Recode max\_players of 999 for game\_id 2922 to something more reasonable, the game is for 1 or more players 

## Factorizing a Column

- Use string replace to remove secondary mechanics or categories to get only the primary mechanic or category for each game
- Change separated category variable from character to a factor



## Variable List

- game\_id = unique board game ID Number
  - Data Type = Integer
- description = short description of the game
  - Data Type = String

- max\_players = the largest amount of players that can play the board game at a certain time
  - Data Type - Integer
- max\_playtime = the longest estimated amount of time that it will take for people to finish playing the game. This is written in minutes
  - Data Type = Integer
- min\_age = the minimum age a person needs to be in order to play the board game
  - Data Type = Integer
- min\_players = the minimum amount of people that can play the board game at a certain time
  - Data Type = Integer
- min\_playtime = the shortest estimated amount of time that it will take for people to finish playing the game. This is written in minutes.
  - Data Type = Integer
- name = the name of the board game
  - Data Type = String
- playing\_time = the estimated average amount of time that it will take for people to finish playing the game. This is written in minutes
  - Data Type = Integer
- year\_published = the year that the game was published
  - Data Type = Integer
- category = this is a string that says what categories each board game belongs to. If a board game has multiple categories all of them will be written down and will be separated by commas
  - Data Type = String
- compilation = if the game is part of a multi-game compilation is lists what compilation it is apart of
  - Data Type = String

- designer = this is the name of the person or people that designed the game. If there are multiple people that designed a game their names are split by commas in between each name
  - Data Type = String
- expansion = if the game has an expansion pack this variable has the name of th expansion pack
  - Data Type = String
- mechanic = this explains how the game is played and what the game uses to work. If a game uses multiple elements to run the elements are split by commas
  - Data Type = String
- publisher = this is the company or people that published the game. If a game has multiple publishers then the publishers are split by commas
  - Data Type = String
- average\_rating = this is the average rating of the game on Board Game Geek. The rating system runs on a scale from 1 to 10
  - Data Type = Double
- users Rated = This is the number of user that rated the game on Board Game Geek.
  - Data Type = Integer