Topic ideas

STA 210 - Project

Ginger and Stats - Rakshita Ramakrishna, Aimi Wen, Nathan Nguyen, Bryan Pan

```
library(tidyverse)
```

Project idea 1

Introduction and data

- State the source of the data set.
  - The source of the dataset is from the Board Game Geek website through the tidy tuesday package. A larger dataset can be found on Kaggle.
- Describe when and how it was originally collected (by the original data curator, not necessarily how you found the data)
  - The data was collected through crowd-sourced ratings posted on the Board Game Geek website over the span of various years, but for our analysis we will only focus on ratings during the year 2016.
- Describe the observations and the general characteristics being measured in the data
  - The data is splitted into two datasets, one contains the technical information of the board games, and more details on the number of users owning them.

Research question

- Describe a research question you're interested in answering using this data.

Glimpse of data

- Use the `glimpse` function to provide an overview of the dataset

```
details <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/mas

ratings <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/mas

glimpse(ratings)
```

```
Rows: 21,831
Columns: 10
$ num           <dbl> 105, 189, 428, 72, 103, 191, 100, 3, 15, 35, 30, 182, 13~
$ id            <dbl> 30549, 822, 13, 68448, 36218, 9209, 178900, 167791, 1733~
$ name          <chr> "Pandemic", "Carcassonne", "Catan", "7 Wonders", "Domini~
$ year          <dbl> 2008, 2000, 1995, 2010, 2008, 2004, 2015, 2016, 2015, 20~
$ rank          <dbl> 106, 190, 429, 73, 104, 192, 101, 4, 16, 36, 31, 183, 14~
$ average       <dbl> 7.59, 7.42, 7.14, 7.74, 7.61, 7.41, 7.60, 8.42, 8.11, 7.~
$ bayes_average <dbl> 7.487, 7.309, 6.970, 7.634, 7.499, 7.305, 7.508, 8.274, ~
$ users_rated   <dbl> 108975, 108738, 108024, 89982, 81561, 76171, 74419, 7421~
$ url           <chr> "/boardgame/30549/pandemic", "/boardgame/822/carcassonne~
$ thumbnail     <chr> "https://cf.geekdo-images.com/S3ybV1LAp-8SnHIXLLjVqA__mi~
```

```
glimpse(details)
```

```
Rows: 21,631
Columns: 23
$ num                   <dbl> 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, ~
$ id                    <dbl> 30549, 822, 13, 68448, 36218, 9209, 178900, 16~
$ primary               <chr> "Pandemic", "Carcassonne", "Catan", "7 Wonders~
$ description           <chr> "In Pandemic, several virulent diseases have b~
$ yearpublished         <dbl> 2008, 2000, 1995, 2010, 2008, 2004, 2015, 2016~
$ minplayers            <dbl> 2, 2, 3, 2, 2, 2, 2, 1, 2, 1, 3, 2, 1, 2, 2, 2~
$ maxplayers            <dbl> 4, 5, 4, 7, 4, 5, 8, 5, 2, 5, 5, 4, 5, 5, 5, 4~
$ playingtime           <dbl> 45, 45, 120, 30, 30, 60, 15, 120, 30, 150, 150~
$ minplaytime           <dbl> 45, 30, 60, 30, 30, 30, 15, 120, 30, 30, 90, 3~
$ maxplaytime           <dbl> 45, 45, 120, 30, 30, 60, 15, 120, 30, 150, 150~
$ minage                <dbl> 8, 7, 10, 10, 13, 8, 14, 12, 10, 12, 12, 10, 1~
$ boardgamecategory     <chr> "['Medical']", "['City Building', 'Medieval', ~
$ boardgamemechanic     <chr> "['Action Points', 'Cooperative Game', 'Hand M~
$ boardgamefamily       <chr> "['Components: Map (Global Scale)', 'Component~
```

```
$ boardgameexpansion      <chr> "['Pandemic: Gen Con 2016 Promos - Z-Force Tea~
$ boardgameimplementation <chr> "['Pandemic Legacy: Season 0', 'Pandemic Legac~
$ boardgamedesigner       <chr> "['Matt Leacock']", "['Klaus-Jürgen Wrede']", ~
$ boardgameartist         <chr> "['Josh Cappel', 'Christian Hanisch', 'Régis M~
$ boardgamepublisher      <chr> "['Z-Man Games', 'Albi', 'Asmodee', 'Asmodee I~
$ owned                   <dbl> 168364, 161299, 167733, 120466, 106956, 105748~
$ trading                 <dbl> 2508, 1716, 2018, 1567, 2009, 930, 1110, 538, ~
$ wanting                 <dbl> 625, 582, 485, 1010, 655, 692, 340, 2011, 924,~
$ wishing                 <dbl> 9344, 7383, 5890, 12105, 8621, 6620, 5764, 192~
```

Project idea 2

Introduction and data

- State the source of the data set.
- Describe when and how it was originally collected (by the original data curator, not necessarily how you found the data)
- Describe the observations and the general characteristics being measured in the data

Research question

- Describe a research question you're interested in answering using this data.

Glimpse of data

- Use the `glimpse` function to provide an overview of the dataset

```
chocolate <- read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/
```

```
Rows: 2530 Columns: 10
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (7): company_manufacturer, company_location, country_of_bean_origin, spe...
dbl (3): ref, review_date, rating

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
glimpse(chocolate)
```

```
Rows: 2,530
Columns: 10
$ ref                            <dbl> 2454, 2458, 2454, 2542, 2546, 2546, 2~
$ company_manufacturer           <chr> "5150", "5150", "5150", "5150", "5150~
$ company_location               <chr> "U.S.A.", "U.S.A.", "U.S.A.", "U.S.A.~
$ review_date                    <dbl> 2019, 2019, 2019, 2021, 2021, 2021, 2~
$ country_of_bean_origin         <chr> "Tanzania", "Dominican Republic", "Ma~
$ specific_bean_origin_or_bar_name <chr> "Kokoa Kamili, batch 1", "Zorzal, bat~
```

```
$ cocoa_percent                   <chr> "76%", "76%", "76%", "68%", "72%", "8~
$ ingredients                     <chr> "3- B,S,C", "3- B,S,C", "3- B,S,C", "~
$ most_memorable_characteristics  <chr> "rich cocoa, fatty, bready", "cocoa, ~
$ rating                          <dbl> 3.25, 3.50, 3.75, 3.00, 3.00, 3.25, 3~
```

Project idea 3

> **!** Important
>
> Project idea 3 is optional. If you decide to submit only 2 ideas, please delete the section headings below and leave a note below stating so. If you decide to submit the 3rd idea, please delete this callout.

Introduction and data

- State the source of the data set.
- Describe when and how it was originally collected (by the original data curator, not necessarily how you found the data)
- Describe the observations and the general characteristics being measured in the data

Research question

- Describe a research question you're interested in answering using this data.

Glimpse of data

- Use the `glimpse` function to provide an overview of the dataset

```
# add code to load and glimpse data here
```