Problem Set 04

WRITE YOUR NAME HERE

2021-09-18

Learning goals

- Using more data visualization techniques: changing colors and adding trend lines
- Baby's first data wrangling exercise!

Setup

Load necessary packages:

```
library(ggplot2)
library(dplyr)
library(babynames)
```

Question 1: Honor code

For this problem set I worked with (please indicate even if with no one):

Question 2

In this exercise, you're going to recreate the figure from Practice Midterm I Question 4 (see `#midterms` channel in Slack), allowing us to visualize the extent to which the names "Casey" and "Riley" were used for babies of both sex male and female.

Part a)

Perform the data wrangling necessary to transform the `babynames` data frame included in the `babynames` package into a new data frame called `babynames_riley_casey` that will allow us to create the visualization.

Hint: I recommend you first draw on a piece of paper what the data frame should look like; that way you'll know what your target looks like and when you've hit it.

Part b)

Recreate the above visualization *exactly* including the capitalization of all label text.

Part c)

Once again, recreate the above figure *exactly*, however this time add an appropriately chosen trend lines. For clarity's sake, do NOT include the standard error bars.

Bonus

In this exercise, you're going to recreate the figure from Practice Midterm I Question 4 (see `#midterms` channel in Slack). This time however, you're going to limit it to years 1960 and later:

Part a)

Perform the data wrangling necessary to transform the `babynames` data frame included in the `babynames` package into a new data frame `babynames_riley_casey_1960_later` that only has data for 1960 or later.

Part b)

Recreate the above visualization *exactly* including the capitalization of all label text, and with `"forestgreen"` and `"orange"` lines for male and female respectively. This time, the x-axis should only be for years 1960 and later, as saved in the `babynames_riley_casey_1960_later` data frame.