Lab 4 Key

Install and load the package *Lahman*, which will give you access to the dataset *Teams*

```
install.packages("Lahman")
library(Lahman)
```

- Produce a subset of the data that has the following characteristics:
  - Only one team (your choice)
  - data from 1980 to present (or as present as the dataset gets)
  - Includes 5 columns: name, yearID, W, L, R, RA

(The variables above correspond to the team name, the year, wins, losses, runs scored, and runs allowed)

- Make sure you select a team that is currently still around, or it probably won't be interesting (see a list of current at http://www.espn.com/mlb/teams).
- Create a new variable corresponding to the winning percentage for the team you chose over time

$$w_{pct} = \frac{wins}{wins + losses}$$

- Order by winning percentage: Least to greatest
- Order by winning percentage: greatest to least
- Compute the mean and standard deviation of winning percentage
- With the full dataset
  - compute the average and standard deviation of winning percentage for each team.
  - Order by highest winning percentage
- Use the full data to reproduce the plot below

```
library(Lahman)
library(tidyverse)

teams <- Teams %>%
  janitor::clean_names()

cubs <- teams %>%
    filter(name == "Chicago Cubs" & year_id >= 1980) %>%
    select(name, year_id, w, l, r, ra)

cubs <- cubs %>%
    mutate(w_pct = w / (w + l))

cubs %>%
  arrange(w_pct)

cubs %>%
  arrange(desc(w_pct))

cubs %>%
  summarize(mean_winning_pct = mean(w_pct),
            sd_winning_pct = sd(w_pct))

teams %>%
  mutate(w_pct = w / (w + l)) %>%
```

```
  group_by(name) %>%
  summarize(n = n(),
            mean_winning_pct = mean(w_pct, na.rm = TRUE),
            sd_winning_pct = sd(w_pct, na.rm = TRUE)) %>%
  arrange(desc(mean_winning_pct))
```

```
library(Lahman)
library(tidyverse)
teams <- Teams %>%
  janitor::clean_names()

teams %>%
  tbl_df() %>%
  mutate(w_pct = w / (w + l)) %>%
  filter(name == "New York Yankees" |
         name == "Detroit Tigers" |
         name == "San Diego Padres") %>%
  ggplot(aes(year_id, w_pct)) +
    geom_line(aes(color = name)) +
    ggthemes::theme_hc()
```