EDLD 610 Final Project

Woocheol Kim[1] & Jessica Canfield[1]

[1] University of Oregon

Author Note

Jessica Canfield & Woocheol Kim are both Marketing PhD students at the University of Oregon.

Correspondence concerning this article should be addressed to Woocheol Kim, 1208 University St, Eugene, OR 97403. E-mail: wkim4@uoregon.edu

Abstract

Marketing research has frequently used the context of sports to explore one facet of consumption. Additionally, the data within the sports realm is well-documented and detailed across time which allows for analyses to be tracked across time and different locations. While the current analysis is mainly exploratory in nature the goal of this project is to familiarize ourselves with this dataset prior to using it in future marketing studies. In this project specifically we look at how the 2008 financial crisis impacts ticket price for professional sports teams. However, in the future we plan to use this data in conjunction with other datasets that have unique time and location identifiers to look more specifically at how consumers engage with sports in reaction to other events occuring simultaneously, whether that be financial crises, political uncertainty, or natural disasters.

Keywords: sports, NBA, NHL, NFL, MLB, NCAAF

Word count: X

22 EDLD 610 Final Project

```
mlb <- import(here("Data", "MLB.xlsx")) %>%
  characterize() %>%
  clean_names() %>%
  select(sport, team, year, capacity,
         attend_tot, attend_avg, games,
         ticket_price, home_wins) %>%
  as_tibble()


mlb <- mlb %>%
  mutate(capacity = as.numeric(capacity),
         attend_tot = as.numeric(attend_tot),
         attend_avg = as.numeric(attend_avg),
         games = as.numeric(games),
         ticket_price = as.numeric(ticket_price),
         home_wins = as.numeric(home_wins))


#is.character(mlb$capacity)
```

```
nba <- import(here("Data", "NBA.xlsx")) %>%
  characterize() %>%
  clean_names()%>%
  select(sport, team, year, capacity,
         attend_tot, attend_avg, games,
         ticket_price, home_win) %>%
  as_tibble() %>%
  rename(home_wins = home_win) %>%
```

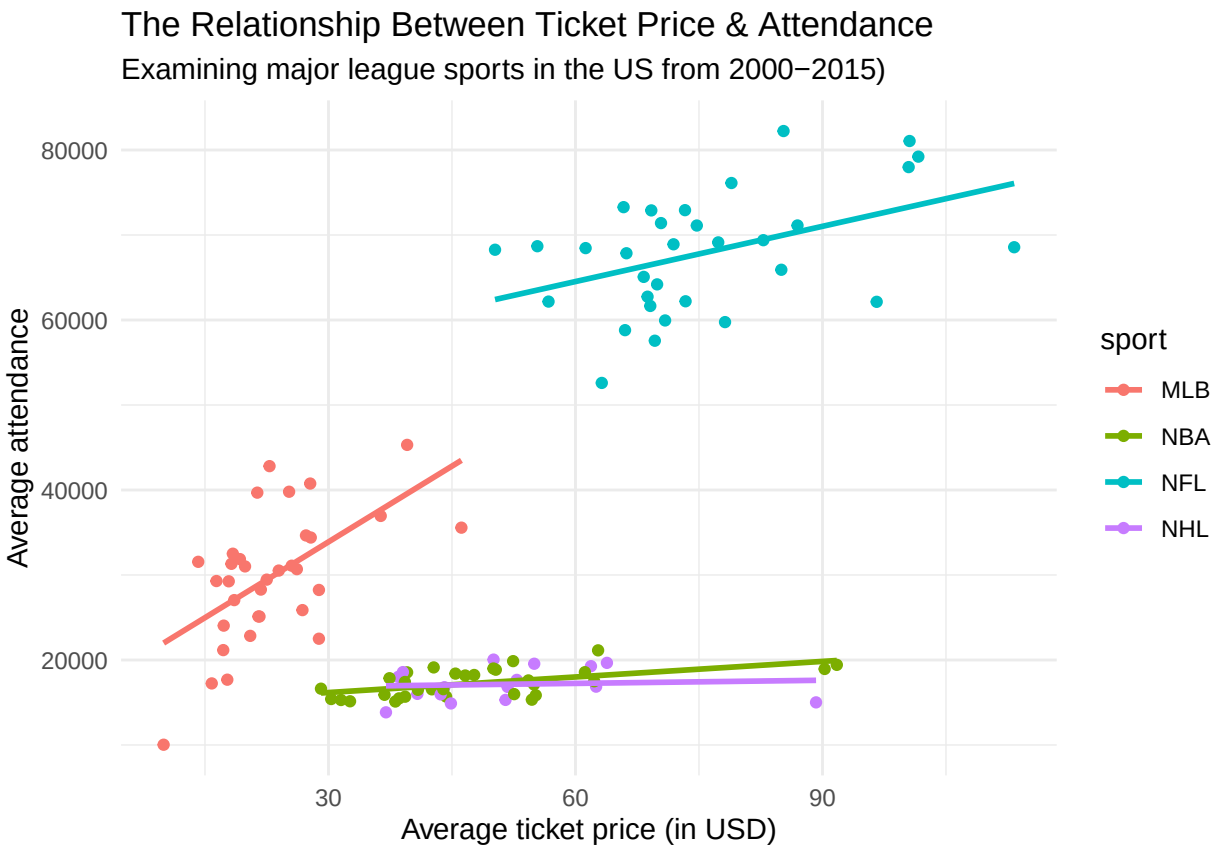
```
  as_tibble()

nba <- nba %>% mutate(capacity = as.numeric(capacity),
                      attend_tot = as.numeric(attend_tot),
                      attend_avg = as.numeric(attend_avg),
                      games = as.numeric(games),
                      ticket_price = as.numeric(ticket_price),
                      home_wins = as.numeric(home_wins))

ncaaf <- import(here("Data", "NCAAF.xlsx")) %>%
  characterize() %>%
  clean_names() %>%
  select(sport, team, year, capacity,
         attend_tot, attend_avg, games,
         ticket_price, home_wins) %>%
  as_tibble()

nfl <- import(here("Data", "NFL.xlsx")) %>%
  characterize() %>%
  clean_names()%>%
  select(sport, team, year, capacity,
         attend_tot, attend_avg, games,
         ticket_price, home_wins) %>%
  as_tibble()

nfl <- nfl %>% mutate(attend_tot = as.numeric(attend_tot),
                      attend_avg = as.numeric(attend_avg),
```

```
                      games = as.numeric(games),
                      ticket_price = as.numeric(ticket_price),
                      home_wins = as.numeric(home_wins))

nhl <- import(here("Data", "NHL.xlsx")) %>%
  characterize() %>%
  clean_names()%>%
  select(sport, team, year, capacity,
         attend_tot, attend_avg, games,
         ticket_price, home_wins) %>%
  as_tibble()

nhl <- nhl %>% mutate(attend_tot = as.numeric(attend_tot),
                      attend_avg = as.numeric(attend_avg),
                      games = as.numeric(games),
                      ticket_price = as.numeric(ticket_price),
                      home_wins = as.numeric(home_wins))

sports <- bind_rows(mlb, nba, ncaaf, nfl, nhl)%>%
  as_tibble()
```

```
sports <- sports %>%
  mutate(home_wins_pct = home_wins/games*100) %>%
  drop_na()

sports_rev <- sports %>%
  group_by(team, sport) %>%
```

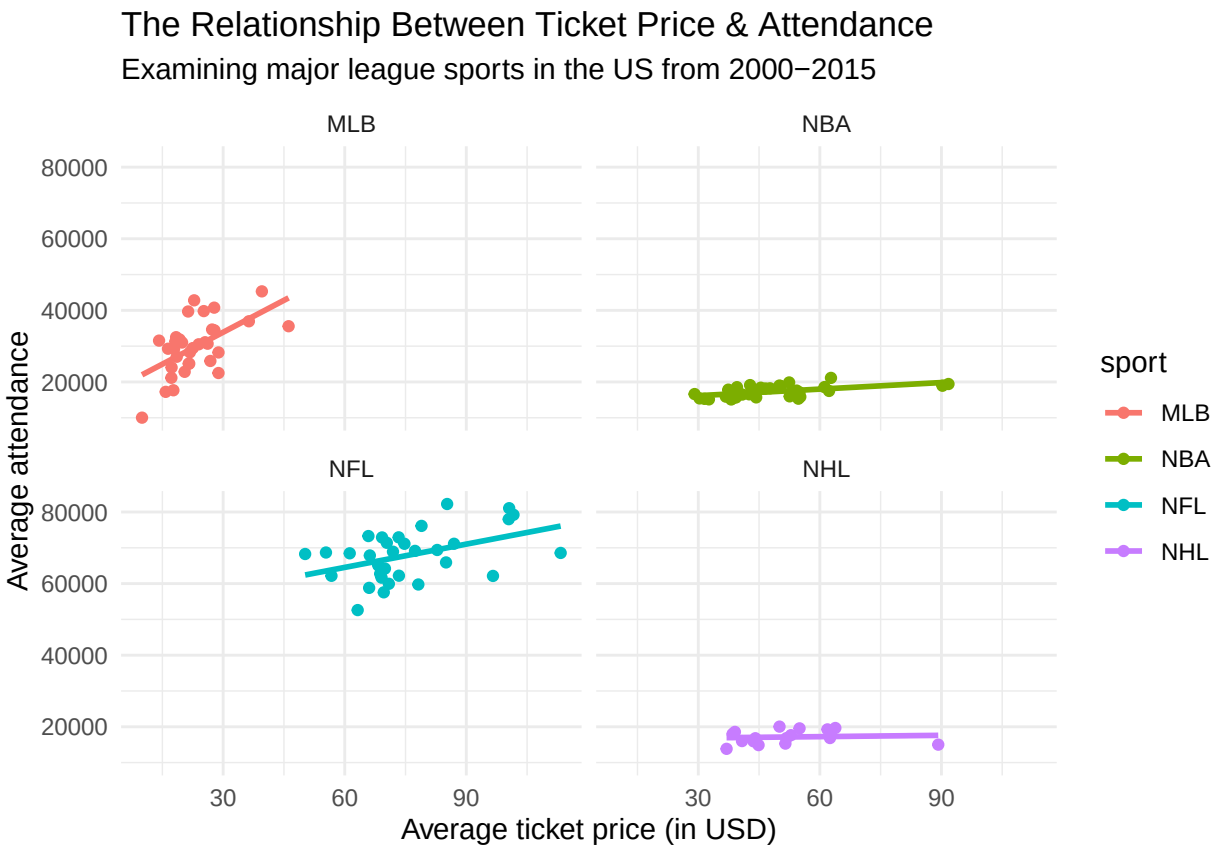
```
  summarize(avg_ticket_price = mean(ticket_price),
            avg_homewins = mean(home_wins),
            avg_attendance = mean(attend_avg),
            avg_homewinspct = mean(home_wins_pct))
```

```
sports_rev %>%
  ggplot(aes(avg_ticket_price, avg_attendance, color = sport)) +
  geom_point() +
  geom_smooth(method = lm, se = FALSE) +
    labs(x = "Average ticket price (in USD)",
       y = "Average attendance",
       title = "The Relationship Between Ticket Price & Attendance",
       subtitle = "Examining major league sports in the US from 2000-2015)") +
  theme_minimal()
```



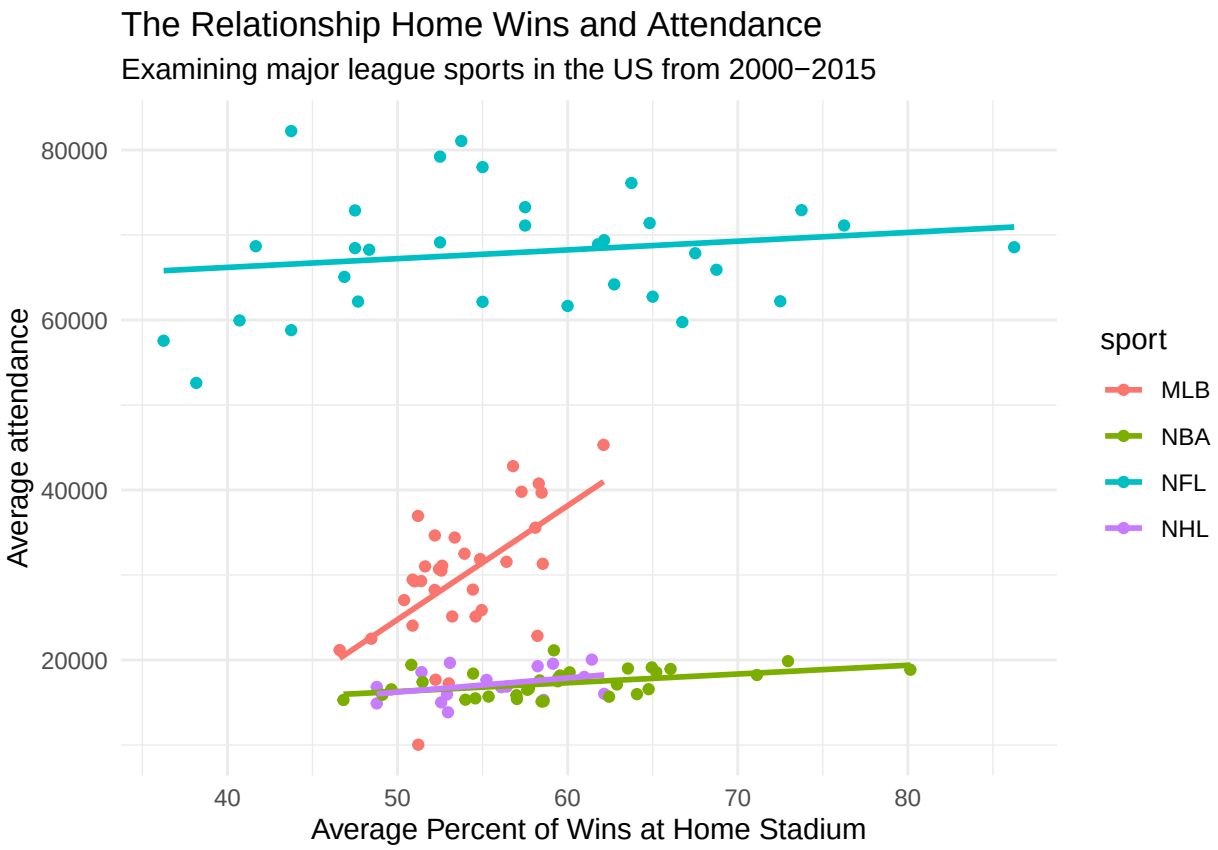
23

```
sports_rev %>%
  ggplot(aes(avg_ticket_price, avg_attendance, color = sport)) +   facet_wrap(~sport) +
  geom_point() +
  geom_smooth(method = lm, se = FALSE) +
  labs(x = "Average ticket price (in USD)",
       y = "Average attendance",
       title = "The Relationship Between Ticket Price & Attendance",
       subtitle = "Examining major league sports in the US from 2000-2015") +
  theme_minimal()
```



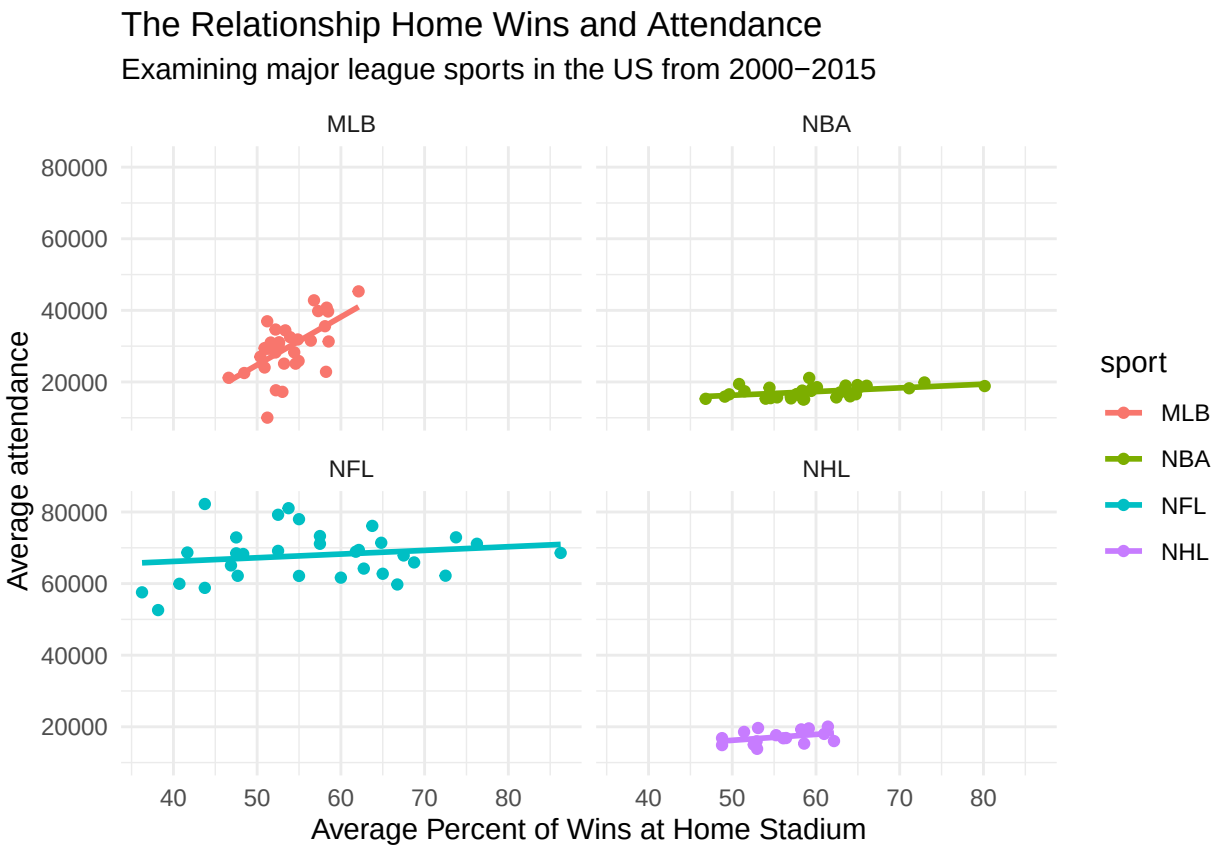
24

```
sports_rev %>%
  ggplot(aes(avg_homewinspct, avg_attendance, color = sport)) +
  geom_point() +
  geom_smooth(method = lm, se = FALSE) +
   labs(x = "Average Percent of Wins at Home Stadium",
        y = "Average attendance",
        title = "The Relationship Home Wins and Attendance",
        subtitle = "Examining major league sports in the US from 2000-2015") +
  theme_minimal()
```



25

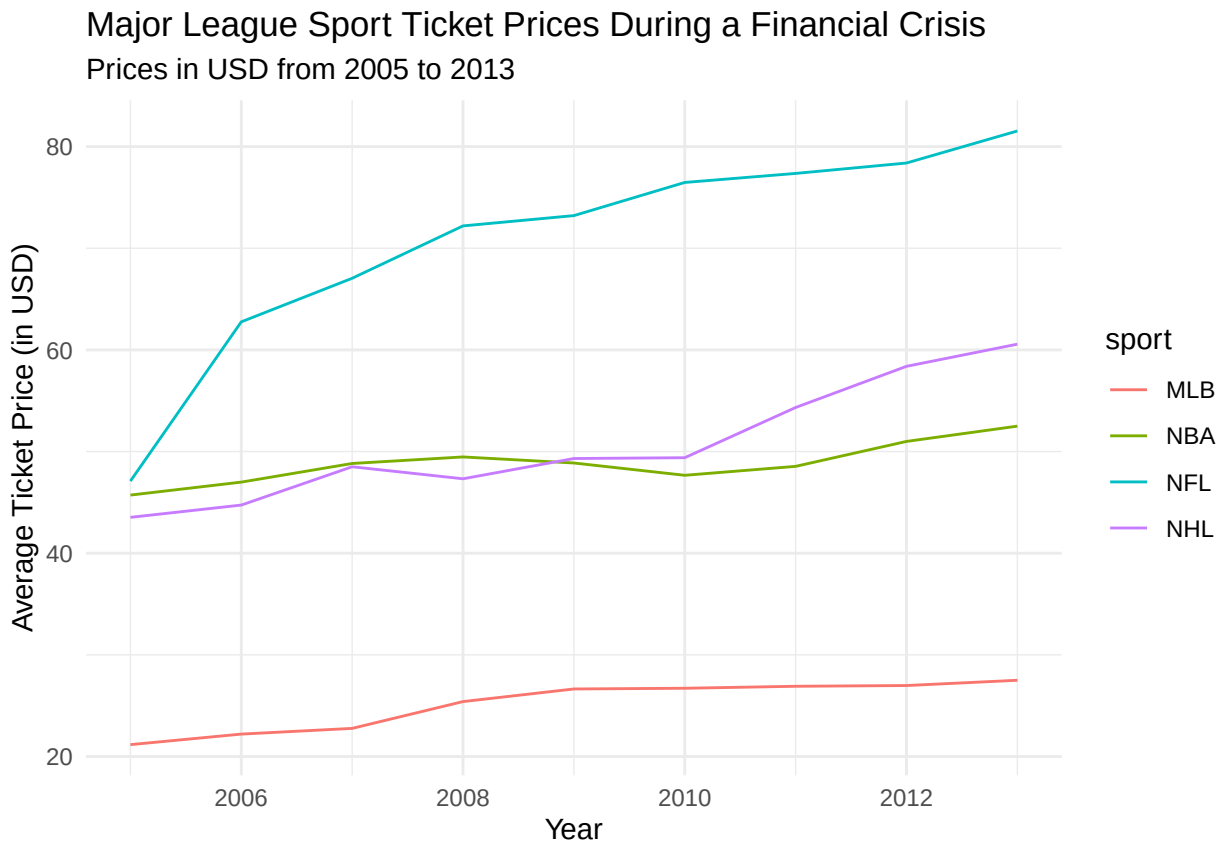
```
sports_rev %>%
  ggplot(aes(avg_homewinspct, avg_attendance, color = sport)) +
  facet_wrap(~sport)+
  geom_point() +
  geom_smooth(method = lm, se = FALSE) +
   labs(x = "Average Percent of Wins at Home Stadium",
        y = "Average attendance",
        title = "The Relationship Home Wins and Attendance",
        subtitle = "Examining major league sports in the US from 2000-2015") +
  theme_minimal()
```



26

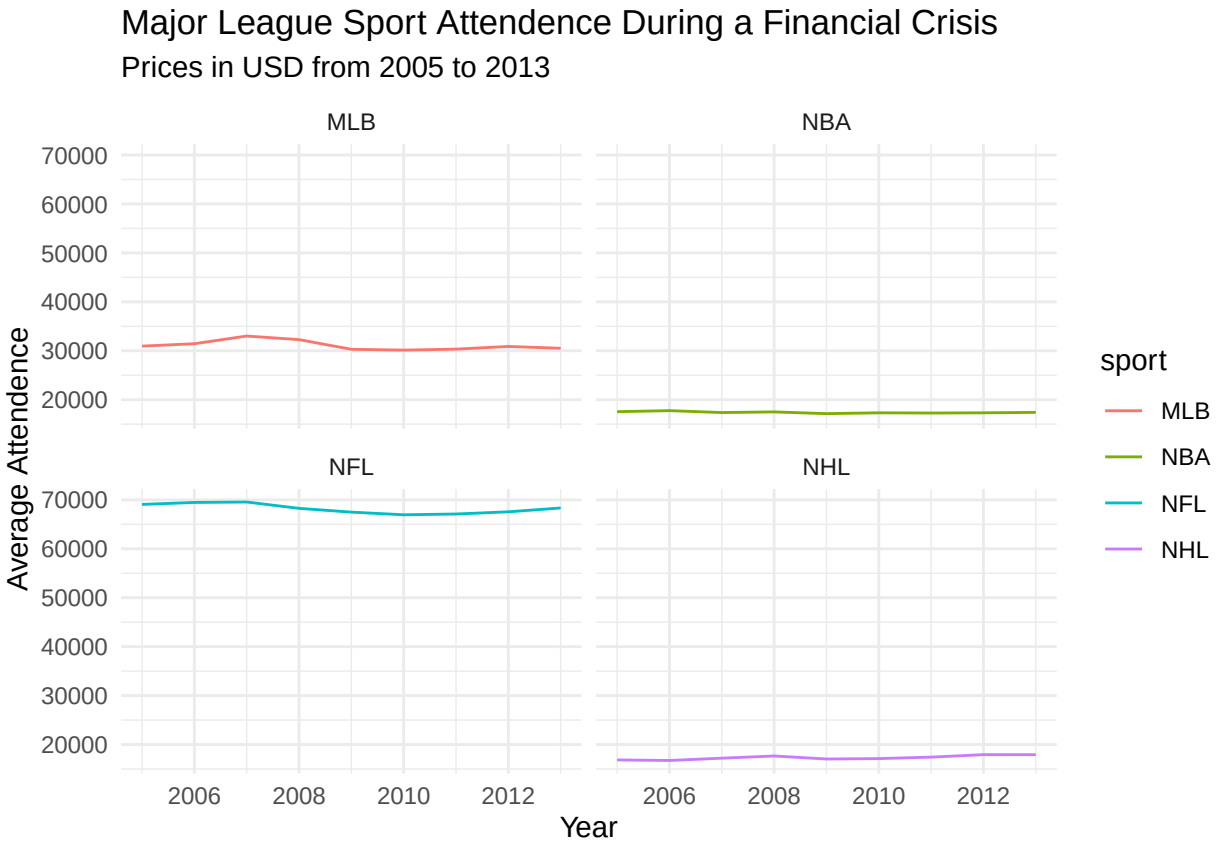
```
sports_crisis <- sports %>%
  group_by(year, sport) %>%
  summarize(avg_ticket_price = mean(ticket_price),
            avg_attendance = mean(attend_avg))


sports_crisis %>%
  filter(year >= 2005, year <= 2013) %>%
  ggplot(aes(year, avg_ticket_price, color = sport)) +
  geom_line() +
  labs (x = "Year",
        y = "Average Ticket Price (in USD)",
        title = "Major League Sport Ticket Prices During a Financial Crisis",
```

```
        subtitle = "Prices in USD from 2005 to 2013") +
  theme_minimal()
```



27

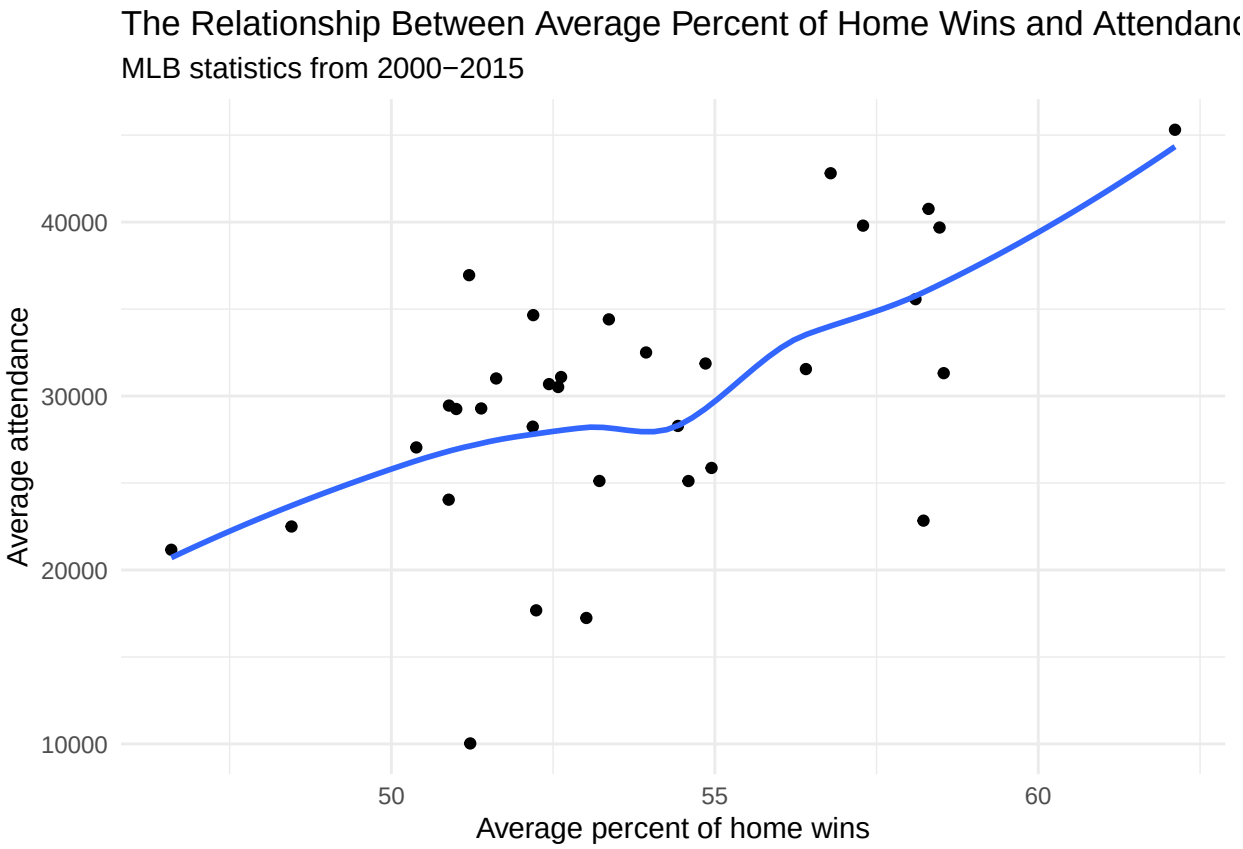
```
sports_crisis %>%
  filter(year >= 2005, year <= 2013) %>%
  ggplot(aes(year, avg_attendance, color = sport)) +
  facet_wrap(~sport) +
  geom_line() +
   labs (x = "Year",
        y = "Average Attendence",
        title = "Major League Sport Attendence During a Financial Crisis",
        subtitle = "Prices in USD from 2005 to 2013") +
  theme_minimal()
```



28

```
fit = sports %>%
  group_by(sport) %>%
  do(model = lm(attend_avg ~ ticket_price + home_wins_pct, data = .))


sports_rev %>%
  filter(sport == "MLB") %>%
  ggplot(aes(avg_homewinspct, avg_attendance)) +
  geom_point() +
  geom_smooth(se = FALSE)+
  labs(x = "Average percent of home wins",
       y = "Average attendance",
       title = "The Relationship Between Average Percent of Home Wins and Attendance",
```

```
      subtitle = "MLB statistics from 2000-2015")+
  theme_minimal()
```




```
sports_pivot <- sports %>%
  pivot_longer(home_wins, names_to = c("home", "wins"), names_sep = "_", values_to = "vi
  pivot_wider(names_from = wins, values_from = victory) %>%
  select(-c(9)) %>%
  rename(home_wins = wins)
```

Requirements

1. pivot_longer: done
2. pivot_wider: done
3. group_by: Done

4. summarize: Done
5. filter: Done
6. select: Done
7. mutate: Done
8. one table:
9. two visualization: Done
10. inline code:

Methods

The sports dataset comes from marketing professor Conor Henderson. It covers four major league sports (NBA, MLB, NFL, NHL) as well as NCAA college football. For each sport, the data spans from 2000 through 2015 and is currently in the process of being updated through present. The data was originally compiled from a number of reputable sports-focused sources including Rodney Fort's Sports League Database as well as ESPN.

Participants

Material

Procedure

Data analysis

We used R (Version 3.6.1; R Core Team, 2019) and the R-packages *dplyr* (Version 0.8.3; Wickham et al., 2019), *forcats* (Version 0.4.0; Wickham, 2019a), *ggplot2* (Version 3.2.1; Wickham, 2016), *here* (Version 0.1; Müller, 2017), *janitor* (Version 1.2.0; Firke, 2019), *kableExtra* (Version 1.1.0; Zhu, 2019), *knitr* (Version 1.25; Xie, 2015), *papaja* (Version 0.1.0.9842; Aust & Barth, 2018), *purrr* (Version 0.3.2; Henry & Wickham, 2019), *readr* (Version 1.3.1; Wickham, Hester, & Francois, 2018), *rio* (Version 0.5.16; Chan, Chan,

Leeper, & Becker, 2018), *stringr* (Version 1.4.0; Wickham, 2019b), *tibble* (Version 2.1.3; Müller & Wickham, 2019), *tidyr* (Version 1.0.0; Wickham & Henry, 2019), and *tidyverse* (Version 1.2.1; Wickham, 2017) for all our analyses.

Results

Discussion

Sports continue to play an important role in the United States. In an time when individuals are becoming increasingly isolated [Chalmers2012differences;Shachar2011brands], sports games provide a form of entertainment that can be bring people together, whether that be through watching the game at the sadium or field or on television. While the motivation to watch sports differs for individuals, the widespread appeal of watching teams compete provides a context for marketers to understand sponshorship, group marketing strategies, and targeted advertising.

References

```
r_refs(file = "r-references.bib")
```


Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown*. Retrieved from https://github.com/crsh/papaja

Chan, C.-h., Chan, G. C., Leeper, T. J., & Becker, J. (2018). *Rio: A swiss-army knife for data file i/o*.

Firke, S. (2019). *Janitor: Simple tools for examining and cleaning dirty data*. Retrieved from https://CRAN.R-project.org/package=janitor

Henry, L., & Wickham, H. (2019). *Purrr: Functional programming tools*. Retrieved from https://CRAN.R-project.org/package=purrr

Müller, K. (2017). *Here: A simpler way to find your files*. Retrieved from https://CRAN.R-project.org/package=here

Müller, K., & Wickham, H. (2019). *Tibble: Simple data frames*. Retrieved from https://CRAN.R-project.org/package=tibble

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

Wickham, H. (2017). *Tidyverse: Easily install and load the 'tidyverse'*. Retrieved from https://CRAN.R-project.org/package=tidyverse

Wickham, H. (2019a). *Forcats: Tools for working with categorical variables (factors)*. Retrieved from https://CRAN.R-project.org/package=forcats

Wickham, H. (2019b). *Stringr: Simple, consistent wrappers for common string operations*. Retrieved from https://CRAN.R-project.org/package=stringr

Wickham, H., François, R., Henry, L., & Müller, K. (2019). *Dplyr: A grammar of data manipulation*. Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham, H., & Henry, L. (2019). *Tidyr: Tidy messy data*. Retrieved from https://CRAN.R-project.org/package=tidyr

Wickham, H., Hester, J., & Francois, R. (2018). *Readr: Read rectangular text data*. Retrieved from https://CRAN.R-project.org/package=readr

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Boca Raton, Florida: Chapman; Hall/CRC. Retrieved from https://yihui.name/knitr/

Zhu, H. (2019). *KableExtra: Construct complex table with 'kable' and pipe syntax*. Retrieved from https://CRAN.R-project.org/package=kableExtra