Trends in Major League Sports in the U.S. from 2000-2015

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

Trends in Major League Sports in the U.S. from 2000-2015

Introduction

Humphreys (2010) explores the impact of the global financial crisis on sport in North America. He finds that while attendance and franchise values declined slightly, and a few teams experienced notable financial problems, the nature of sports as a consumer product in addition to institutional factors associated with the sports industry have, so far, insulted professional sports from significant negative shocks as the result of economic uncertainty. Coates and Humphreys (2007) investigate the demand for attendance at professional sporting events using a data set that includes ticket prices and a price index reflecting prices for ancillary goods associated with attendance. Both mathematical modeling and empirical methodology are used in their research (see Coates & Humphreys, 2007).

```
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))
```

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

```r
#creating a new variable, the percent of home wins for a team in a given year
sports <- sports %>%
  mutate(home_wins_pct = home_wins/games*100) %>%
  drop_na()


#creating new variables that are the averages of ticket price, home wins, average atte
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))
```

```r
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()
```

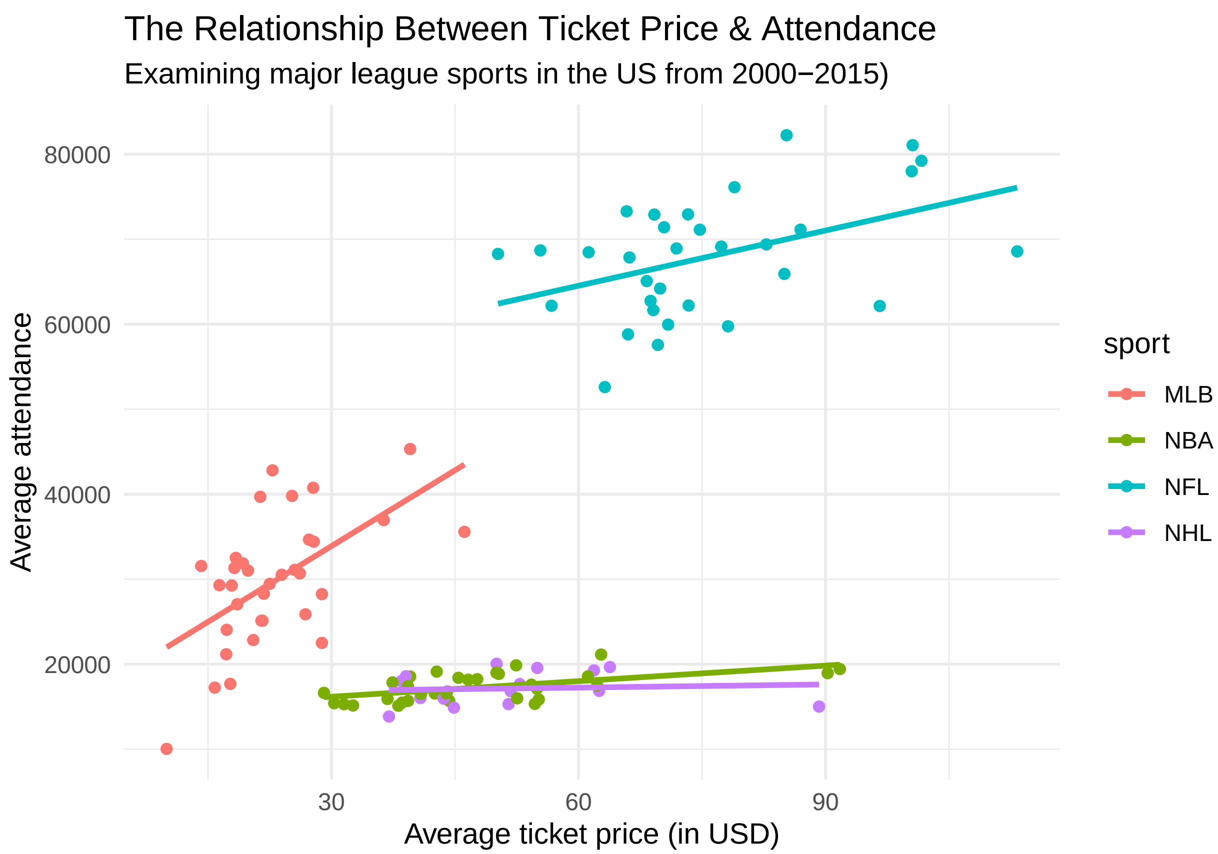

Figure 1

```
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() +
  theme(legend.position = "none")
```

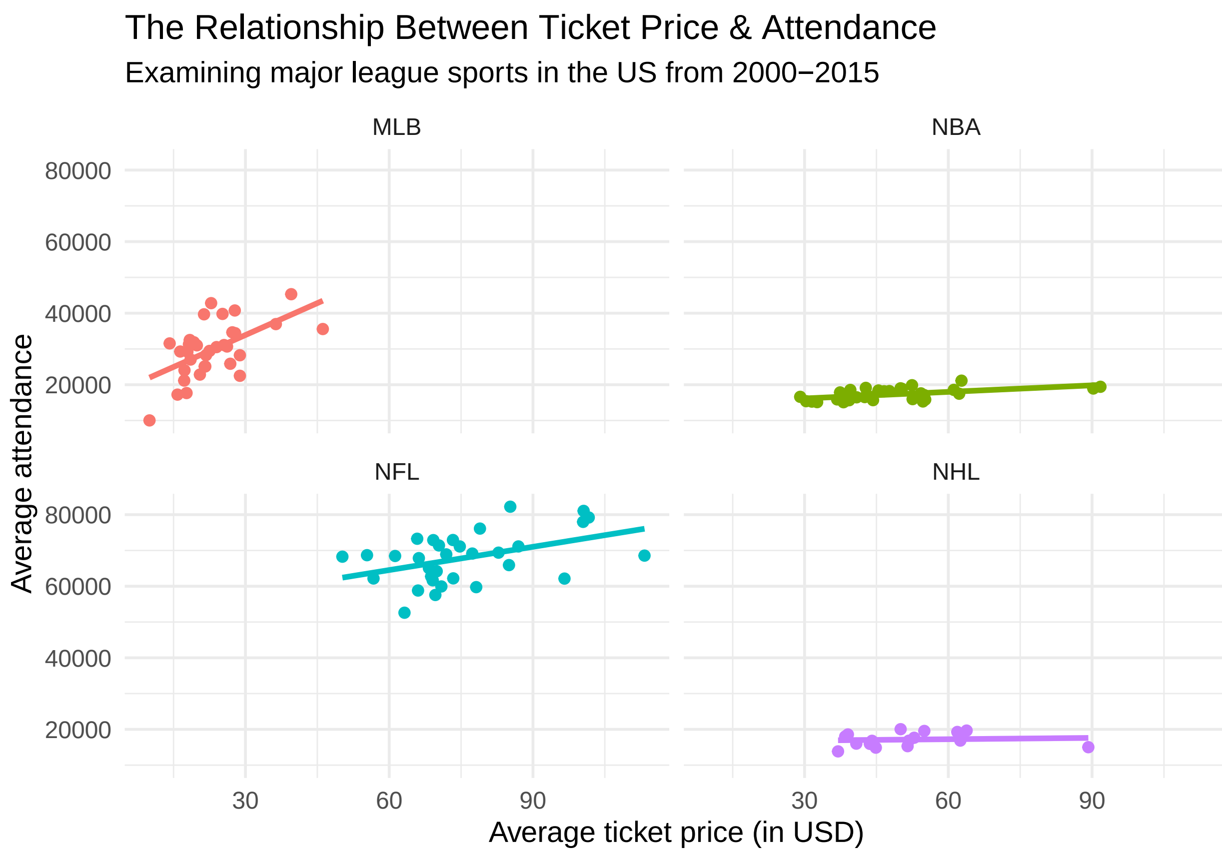

Figure 2

```
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()
```

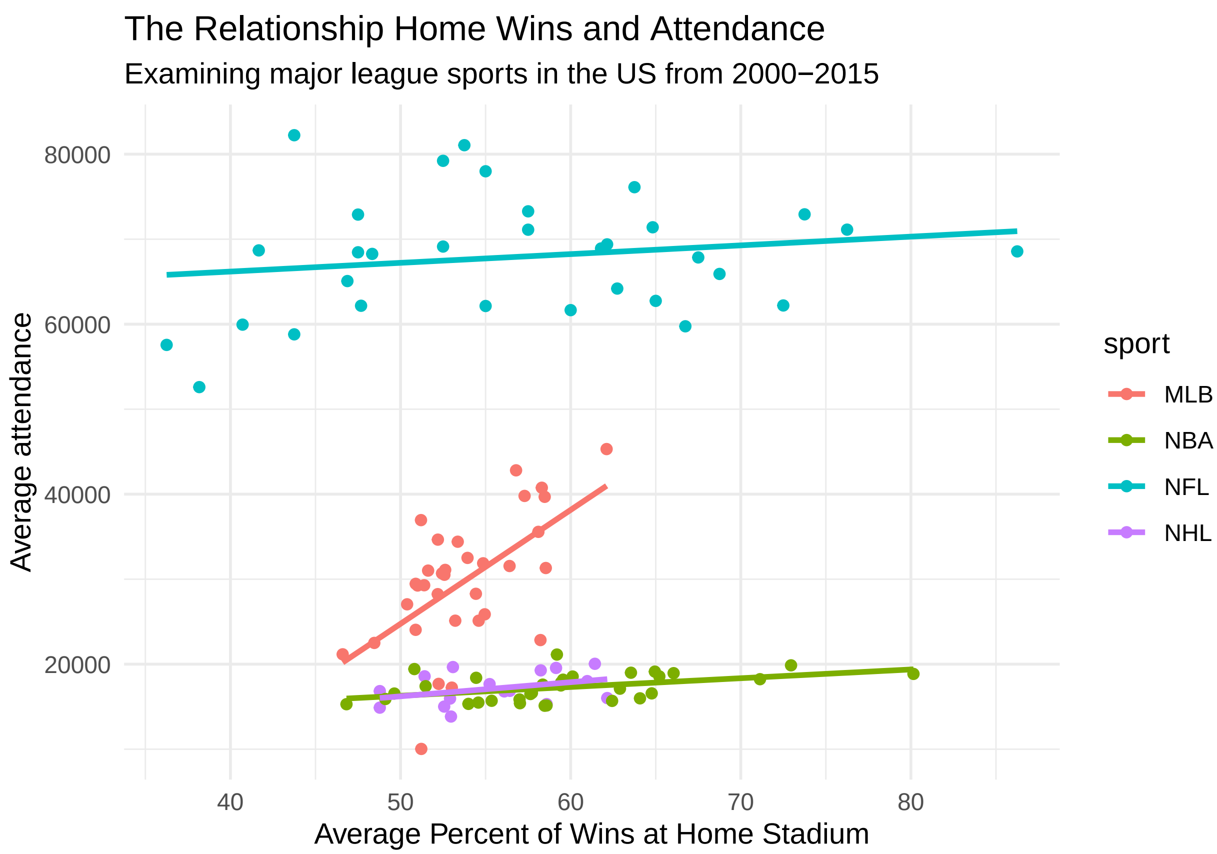

Figure 3

```r
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()+
  theme(legend.position = "none")
```

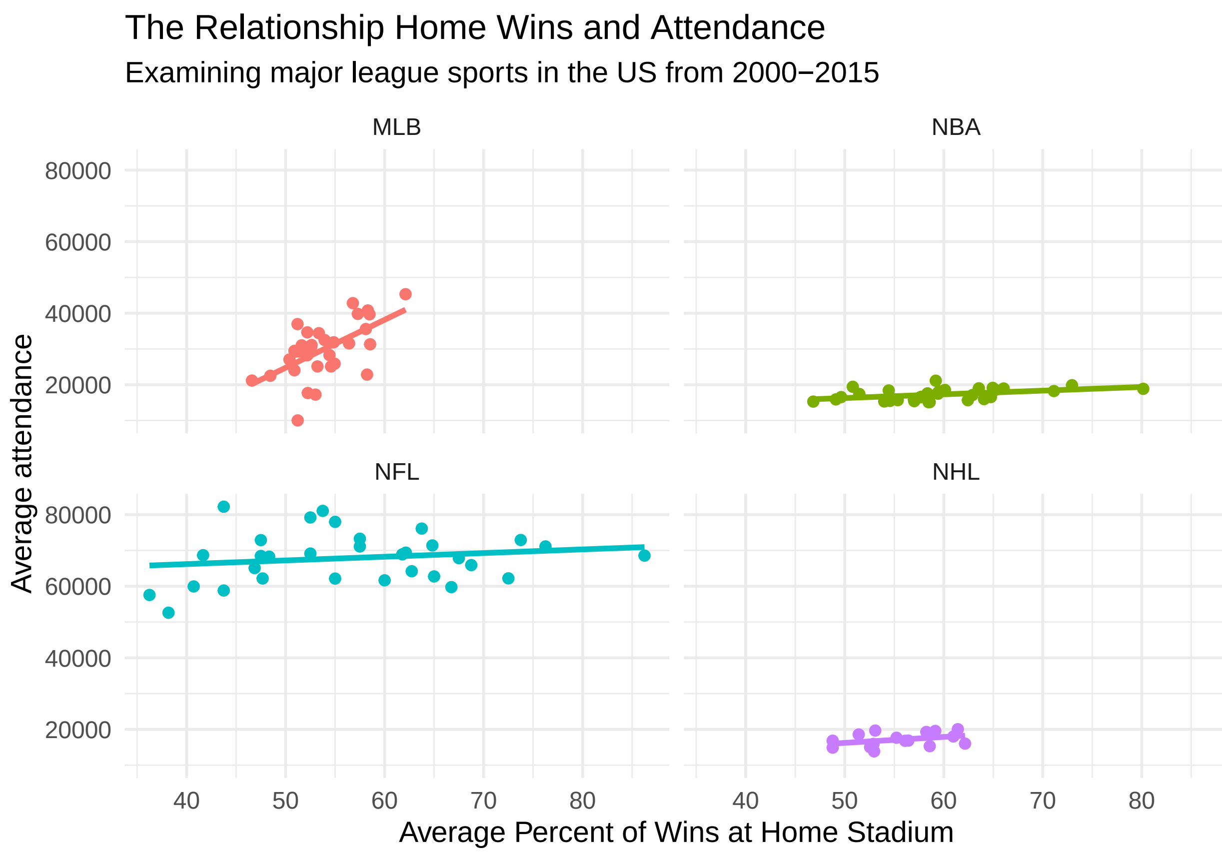

Figure 4

```
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()
```

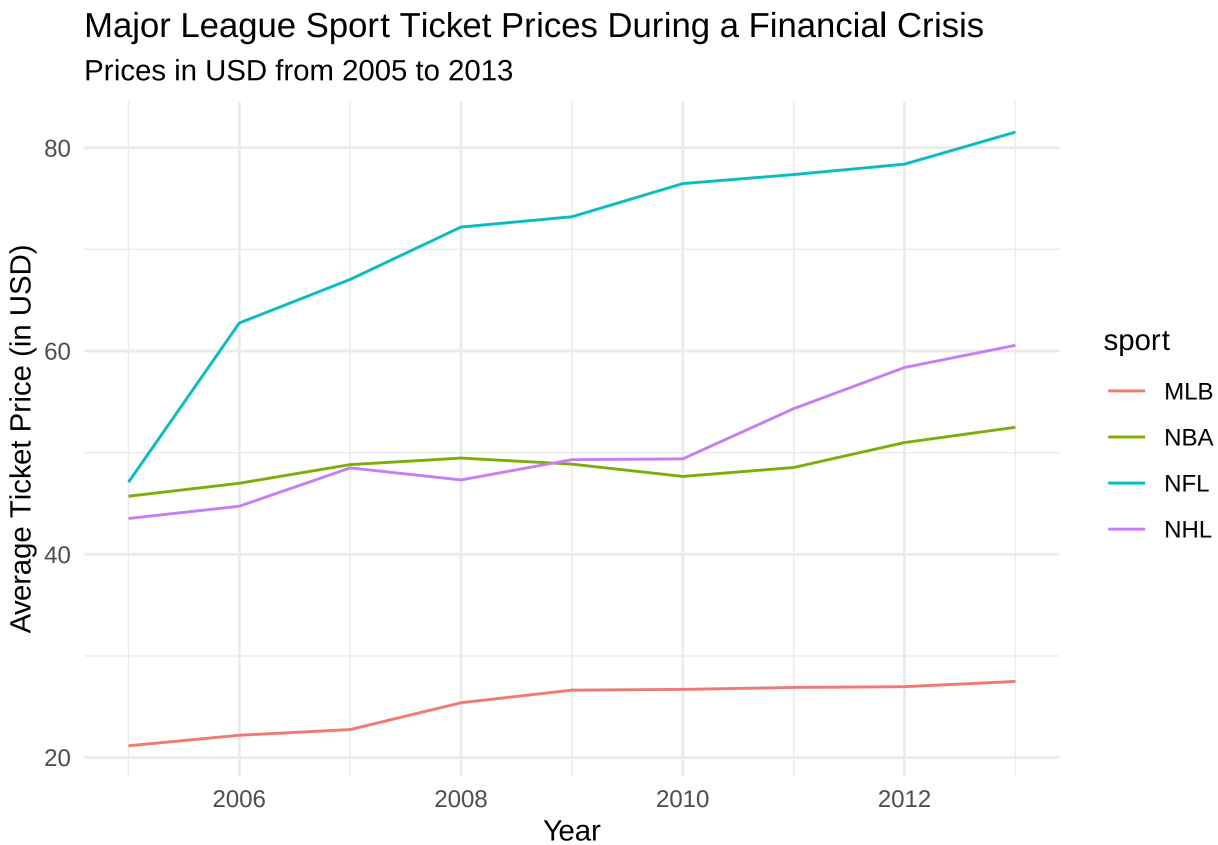

Figure 5

```
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()+
  theme(legend.position = "none")
```

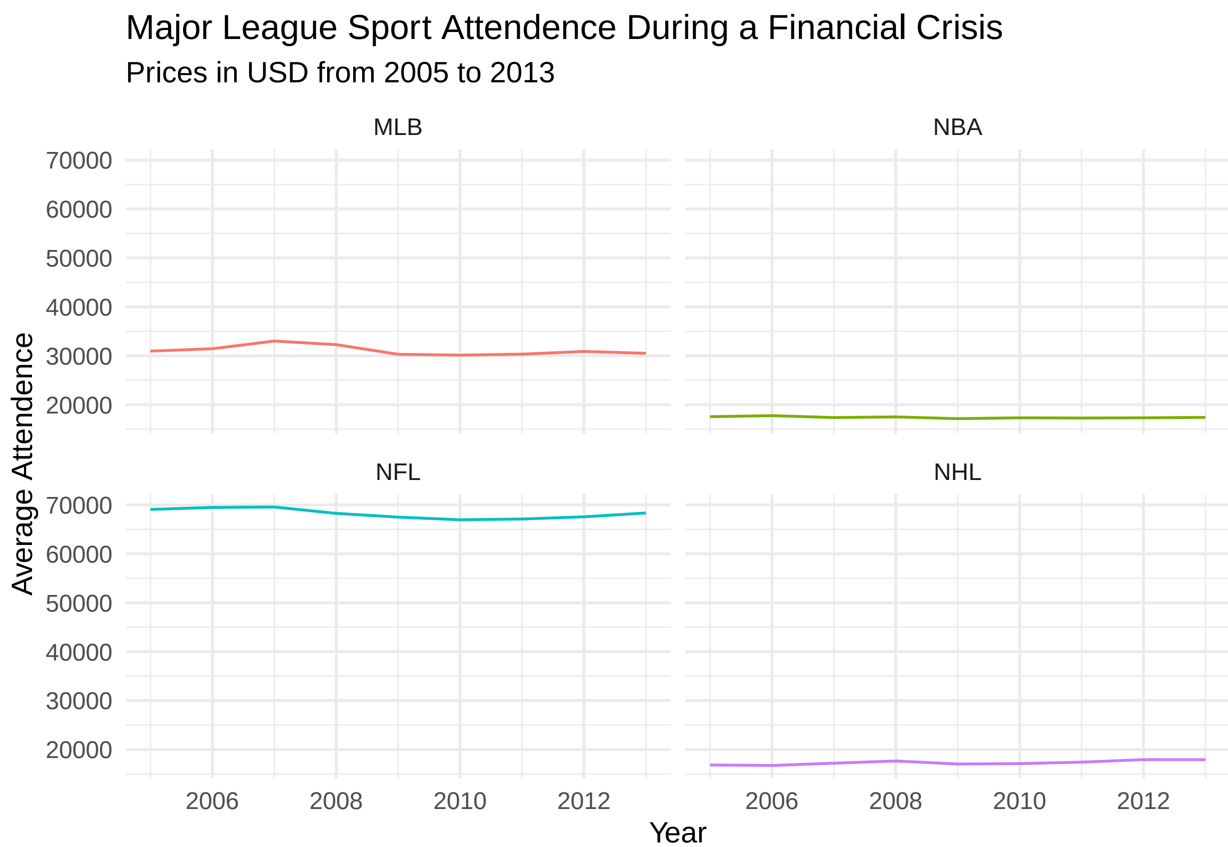

Figure 6

```
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()
```

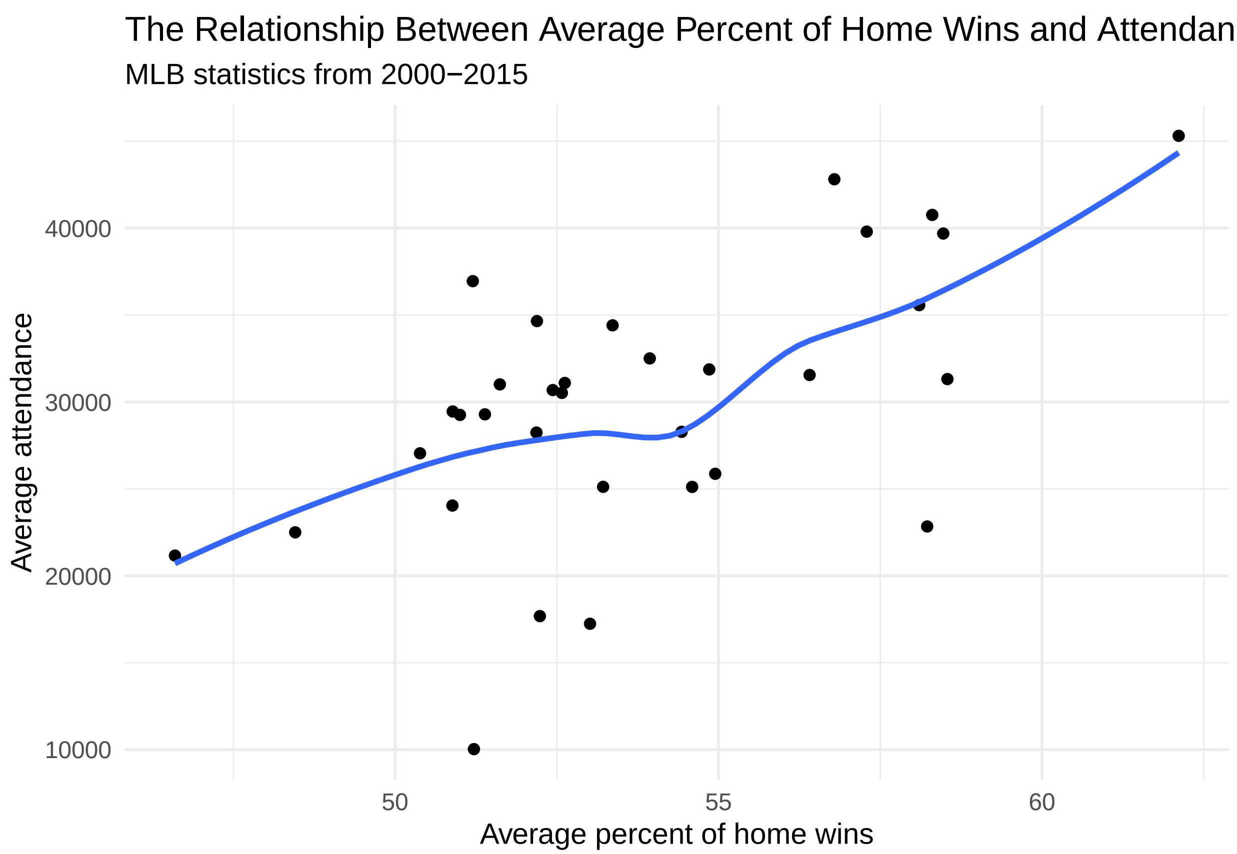

Figure 7

32 **Average Home Attendance, Average Home Ticket Price and Average Home**
33 **Win pct by Sports (2000-2015)**

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

```
            avg_homewinspct = mean(home_wins_pct)) %>%
  kable(digits = 2,
        col.names = c("sport", "attendance_mean", "ticketprice_mean", "homewinpct_mean"))
```

sport	attendance_mean	ticketprice_mean	homewinpct_mean
MLB	30420.90	23.51	53.88
NBA	17335.68	49.04	59.70
NFL	68039.56	75.29	56.86
NHL	17342.74	51.76	55.94

```
sports_crisis_before <- sports %>%
  filter(sport == "MLB", year >= 2007, year <= 2009)

sports_crisis_after <- sports %>%
  filter(sport == "MLB",
         year >= 2010,
         year <= 2012)

attendance_before <- mean(sports_crisis_before$attend_avg, na.rm = TRUE)
price_before <- mean(sports_crisis_before$ticket_price, na.rm = TRUE)
attendance_after <- mean(sports_crisis_after$attend_avg, na.rm = TRUE)
price_after <- mean(sports_crisis_after$ticket_price, na.rm = TRUE)
```

Average revenue per homegame for Major league Baseball (MLB) teams spanning from 2007 and 2009 is $794,445.22 while the one for three years after **financial crisis** is $818,114.18. Major League Baseball seems that it was not affected by recesesion in terms of *revenue* and it actually made more than before the crisis. However, to understand how the recession impacted MLB in greater detail we would need to account for other variables.

Average Home Game Revenue before and after economic recession

Average revenue per homegame of MLB spanning from 2007 and 2009 is $794,445.22 while the revenue 3 years after the **financial crisis** is $818,114.18. Major League Baseball seems that it was not affected by depression in terms of *revenue* and it actually made more than before the crisis.

```
#demonstrating the use of both pivot_longer and pivot_wider
#our data was very clean to begin with so this was for the sake of practice and demons

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)
```

```
## Project 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: Done
#9. two visualization: Done
#10. inline code: Done
```

Methods

The sports dataset was collected by marketing professor Conor Henderson. It covers four major league sports (NBA, MLB, NFL, NHL) as well as NCAA college football (NCAAF). 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. In the final dataset that combines all the sports we have 1398 observations across 15 years and 10 different variables. The 10 variables we selected were: sport, team, year, stadium capacity, total attendance, average attendance, number of games, ticket price (in USD), and the number of home wins.

Data analysis

We used R (Version 3.6.1; R Core Team, 2019) and the R-packages *dplyr* (Version 0.8.3; Wickham, François, Henry, & Müller, 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), *lme4* (Bates, Mächler, Bolker, & Walker, 2015), *Matrix* (Version 1.2.17; Bates & Maechler, 2019), *papaja* (Version 0.1.0.9842; Aust & Barth, 2018), *purrr* (Version 0.3.3; Henry & Wickham, 2019), *readr* (Version 1.3.1; Wickham, Hester, & Francois, 2018), *rio* (Version 0.5.16; C.-h. 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

In all four leagues, it turns out that average ticket price and average rate of home wins is positively associated with average home attendance even though NFL fans seem they are not as sensitive to wins as are the fans in the three other major league sports. This provides empirical evidence for a finding that is relativley intuitive in the sense that as teams win more, demand for tickets likely increases which would drive prices up. Ultimatley, people enjoy watching their home team win and as a result, are willing to pay more when their team is doing well in a given season. However, this is likely correlated with the outcomes of previous seasons as well. In all four leagues, it turns out that average ticket price and average home win rate have a positive relationship with average home attendance even though NFL fans seem to be not as sensitive to win as other three leagues' fans are. In addition, the 2008 financial crisis did not result in crisis in the US professional sports leagues. While some of the leagues experienced slight decrease in home attendance, all four leagues got through economic downturn as if there was nothing happened. Actually, their home game average revenue went up after financial crisis. We assume this counterintuitive outcome is attributed to the facts that sports were kind of immuned to the crisis for some reason or even during recession people are willing to spend money for watching games.

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 from 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. The current exploratory study provides inital insight into how major league attendance varries over time both in regard to attedance as well as ticket prices. Through the analysis, it is clear that each of the major league sports operates very differently from eachother inregard to the variables of interest isolated for the purposes of this research. As this dataset it used going forward, it will be important to identify more clearly the differences between each of the sports to understand if they can indeed be collapsed into an overarching category of "major league sports attendance" across all four major league sports (MLB, NBA, NFL, NHL). Another aspect that was not taken into account in the current research is team-specific factors including how long the team has been in a city as well as how many time a team has moved.

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

Bates, D., & Maechler, M. (2019). *Matrix: Sparse and dense matrix classes and methods.* Retrieved from https://CRAN.R-project.org/package=Matrix

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

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

Coates, D., & Humphreys, B. R. (2007). Ticket prices, concessions and attendance at professional sporting events. *International Journal of Sport Finance*, *2*(3), 161.

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

Humphreys, B. R. (2010). The impact of the global financial crisis on sport in north america. In *Optimal strategies in sports economics and management* (pp. 39–57). Springer.

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

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., 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