1   EDLD 610 Final Project

2   Woocheol Kim[1] & Jessica Canfield[1]

3   [1] University of Oregon

4   Author Note

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

21   EDLD 610 Final Project

22   **Introduction and Overview**

23   Humphreys (2010) explores the impact of the global financial crisis on sport in North

24   America. He finds that while attendance and franchise values declined slightly, and a few

25   teams experienced financial problems, the nature of the sport product and institutional

26   factors associated with the sports industry have, so far, insulted professional sport from

27   significant negative impacts. Coates and Humphreys (2007) investigate the demand for

28   attendance at professional sporting events using a data set that includes ticket prices and a

29   price index reflecting prices for ancillary goods associated with attendance. Both

30   mathematical modeling and empirical methodology are used in their research. (see Coates

31   & Humphreys, 2007).

```r
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))
```

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

```r
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()
```

```r
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),
```
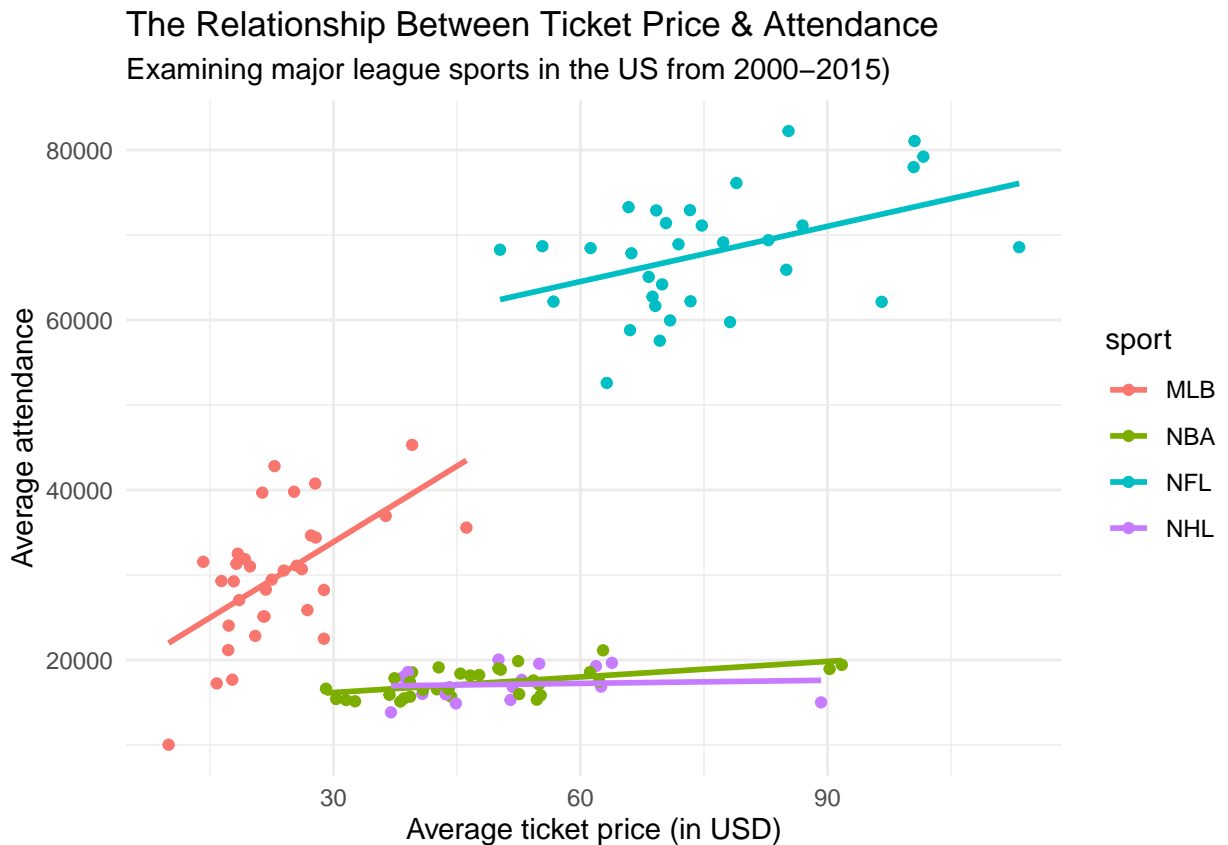
```r
                        home_wins = as.numeric(home_wins))


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

  as_tibble()
```

```r
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))
```
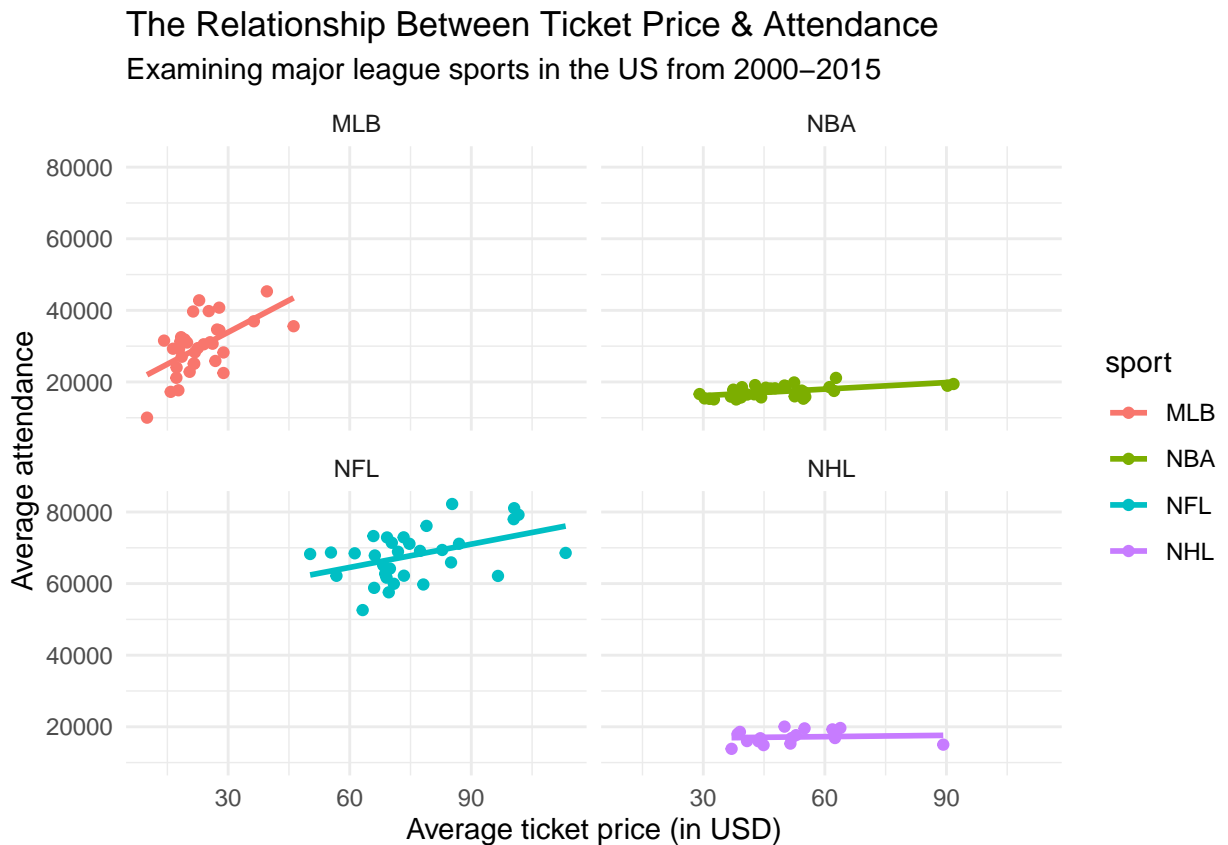
```r
sports_rev %>%

  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()
```

The Relationship Between Ticket Price & Attendance
Examining major league sports in the US from 2000–2015)

32

```r
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()
```
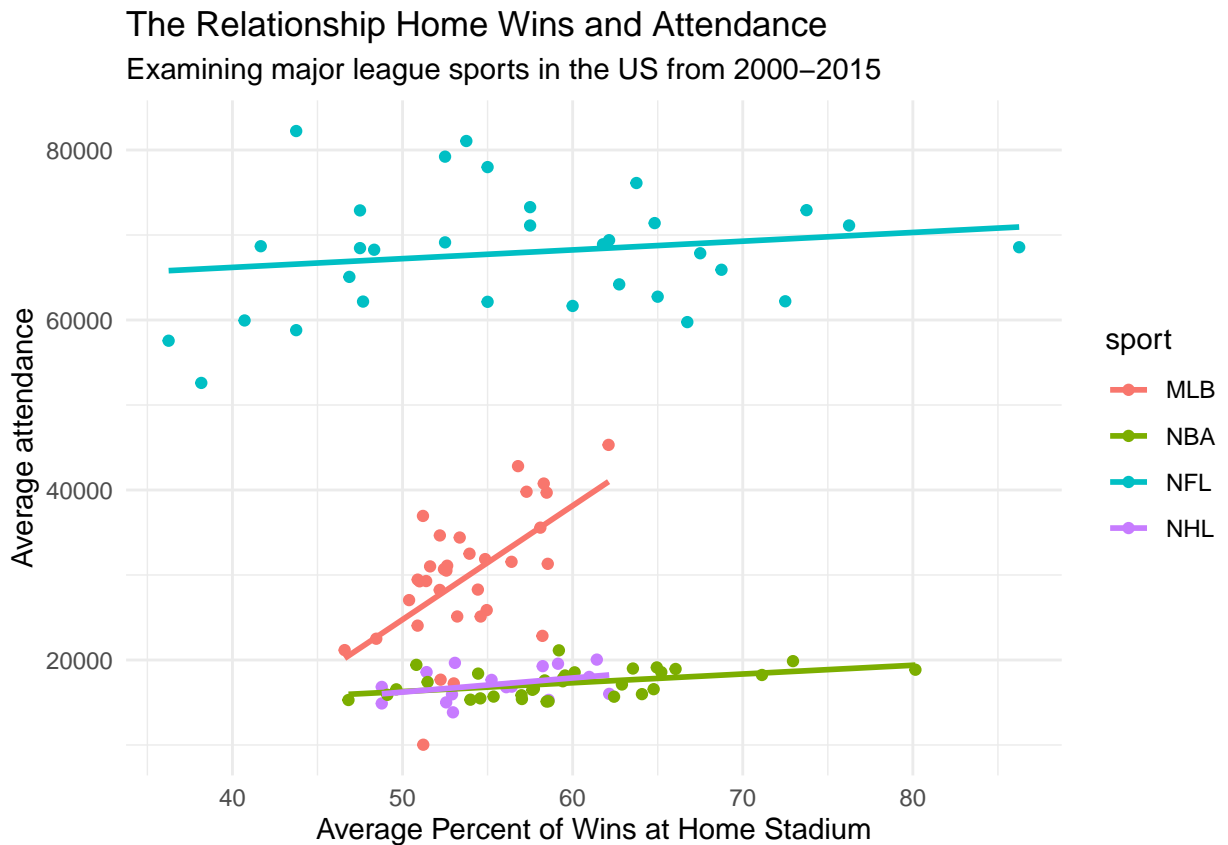
The Relationship Between Ticket Price & Attendance
Examining major league sports in the US from 2000–2015

33
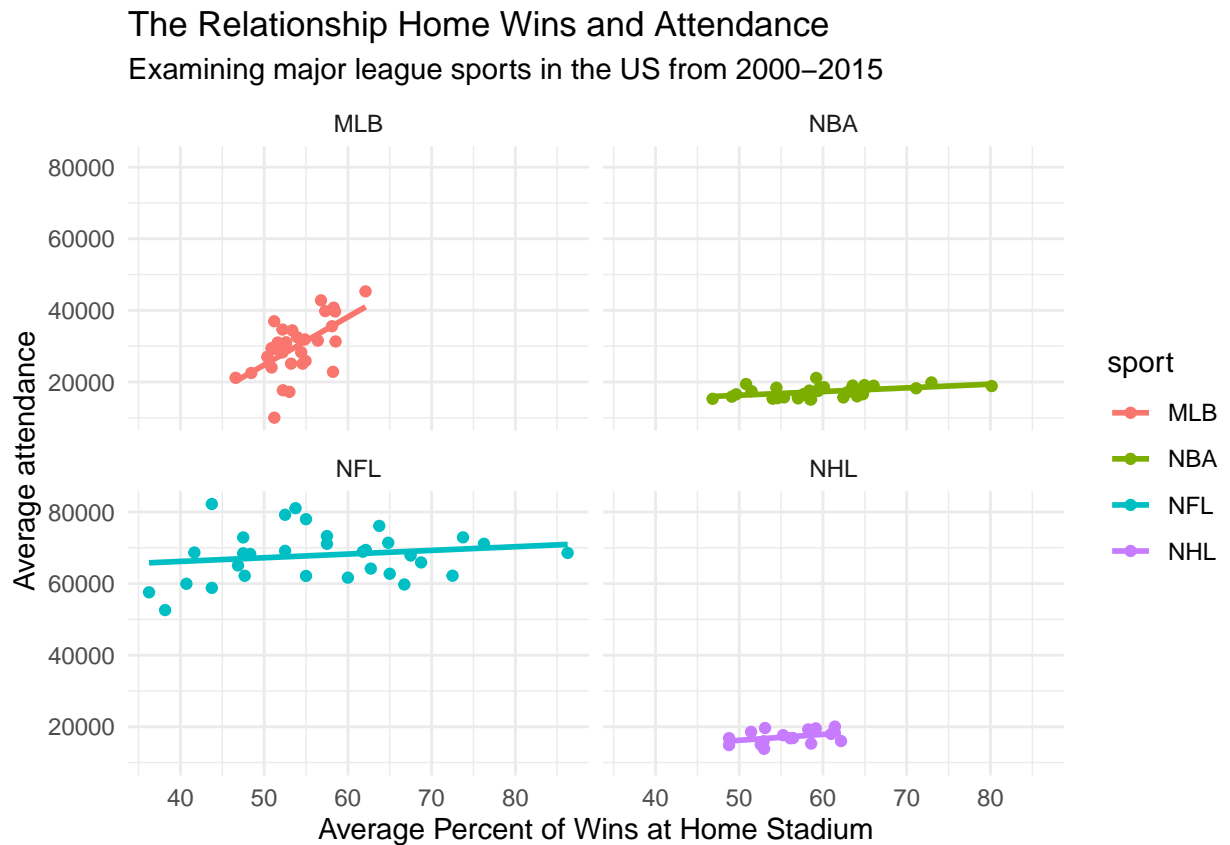
```
sports_rev %>%

  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()
```

The Relationship Home Wins and Attendance
Examining major league sports in the US from 2000–2015


```
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()
```

The Relationship Home Wins and Attendance
Examining major league sports in the US from 2000–2015

35

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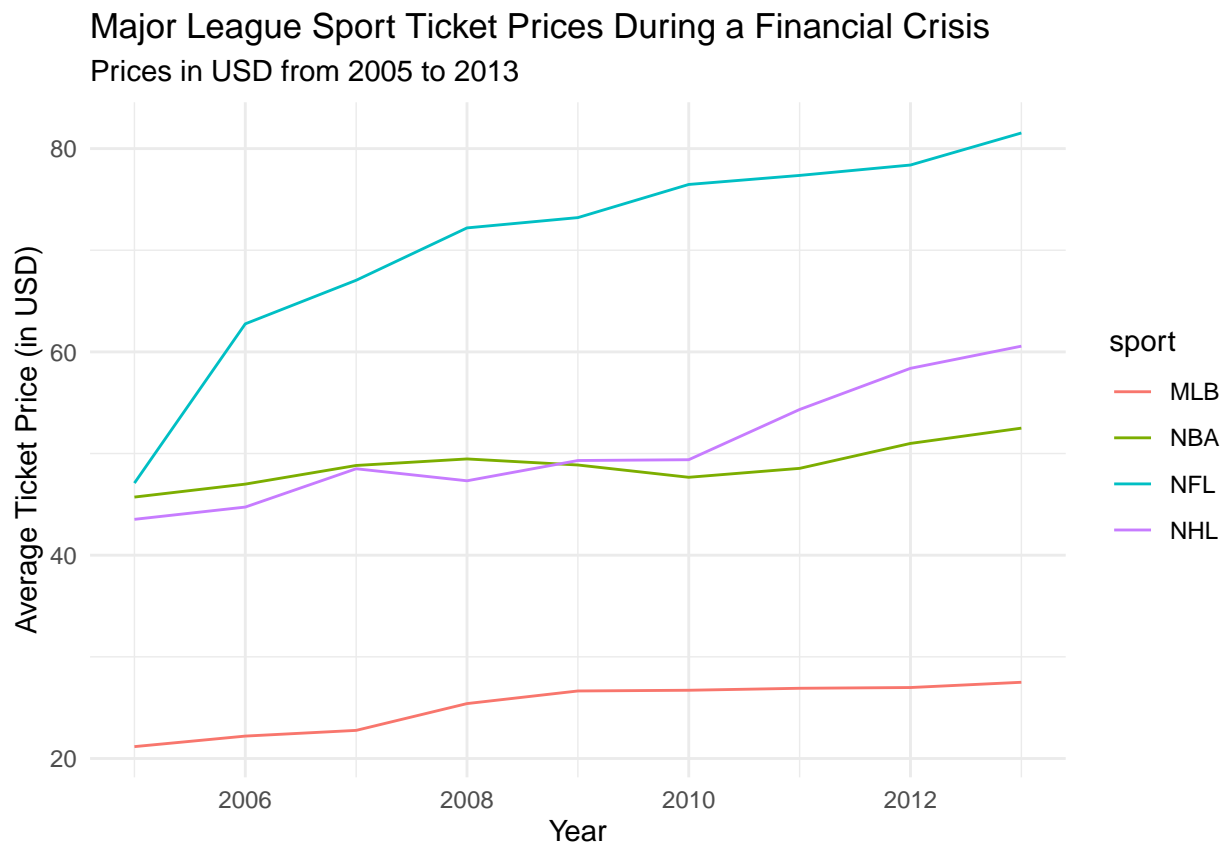
```
                subtitle = "Prices in USD from 2005 to 2013") +

    theme_minimal()
```

Major League Sport Ticket Prices During a Financial Crisis
Prices in USD from 2005 to 2013

36

```
sports_crisis %>%

  filter(year >= 2005, year <= 2013) %>%

  ggplot(aes(year, avg_attendance, color = sport)) +

  facet_wrap(~sport) +

  geom_line() +

   labs (x = "Year",

        y = "Average Attendance",

        title = "Major League Sport Attendence During a Financial Crisis",

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

  theme_minimal()
```

Major League Sport Attendence During a Financial Crisis
Prices in USD from 2005 to 2013
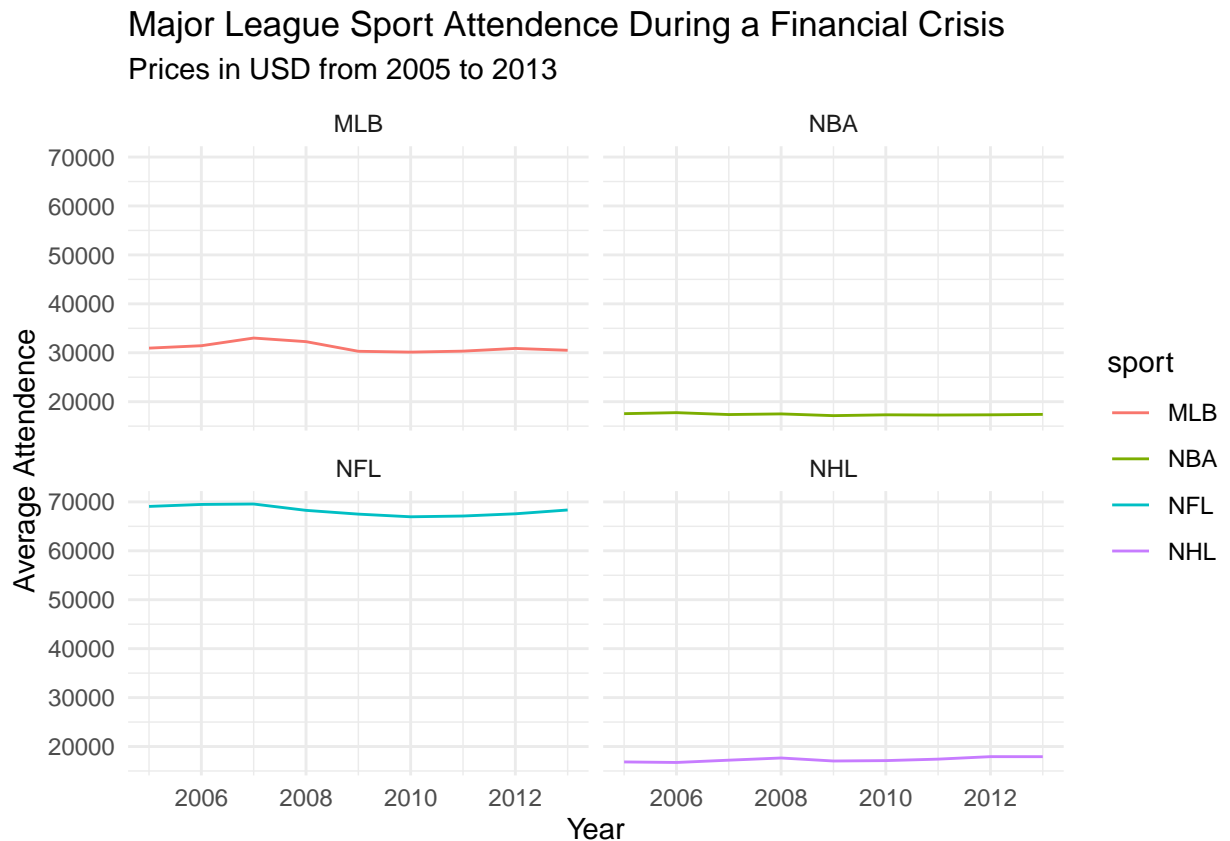
37

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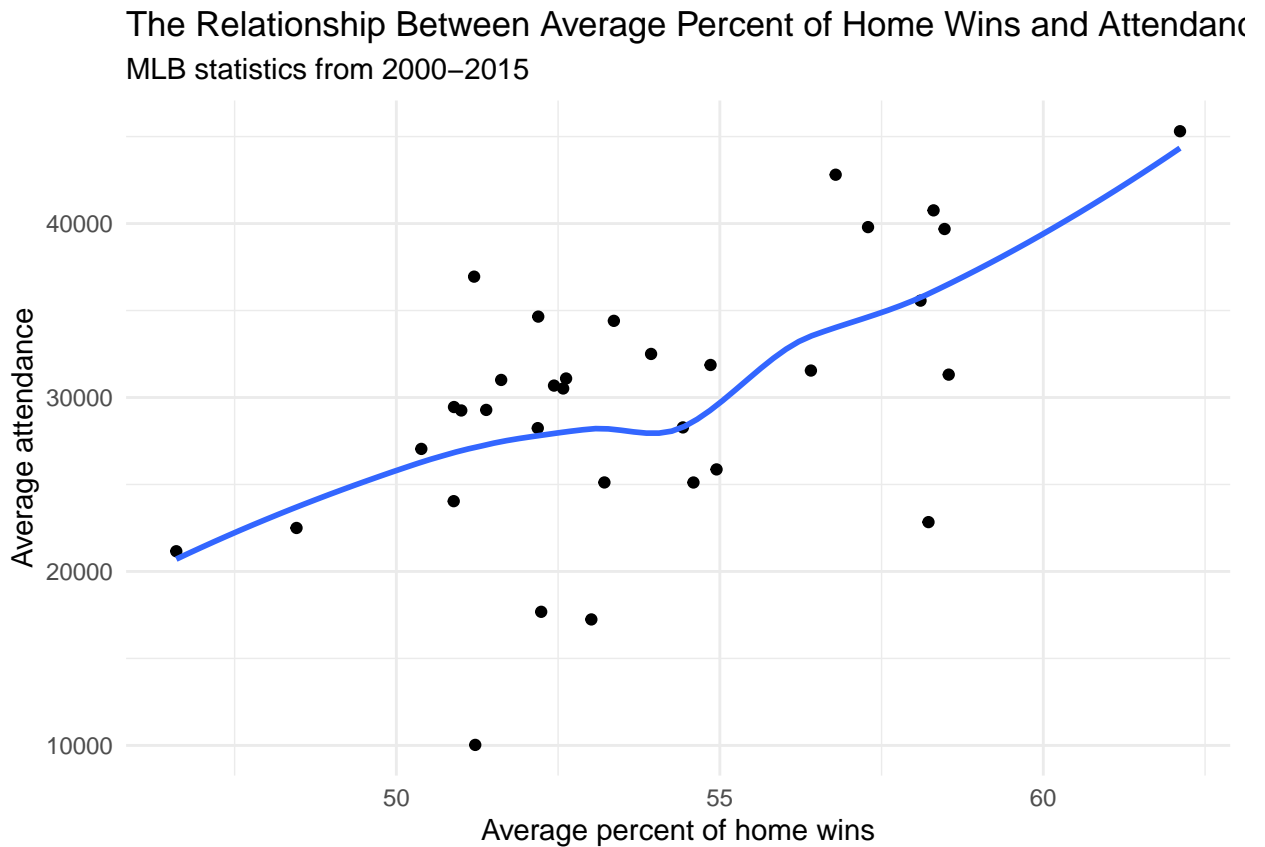
```
         subtitle = "MLB statistics from 2000-2015")+

  theme_minimal()
```

The Relationship Between Average Percent of Home Wins and Attendanc
MLB statistics from 2000–2015

38

39 **Average Home Attendance, Average Home Ticket Price and Average Home**

40 **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

```r
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 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 one for 3 years after **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.

```r
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)
```

⁴⁷ **Methods**

⁴⁸ The sports dataset comes from marketing professor Conor Henderson. It covers four

⁴⁹ major league sports (NBA, MLB, NFL, NHL) as well as NCAA college football. For each

⁵⁰ sport, the data spans from 2000 through 2015 and is currently in the process of being

⁵¹ updated through present. The data was originally compiled from a number of reputable

⁵² sports-focused sources including Rodney Fort's Sports League Database as well as ESPN.

⁵³ In the final dataset that combines all the sports we have 1398 observations across 15 years

⁵⁴ and 10 different variables. The 10 variables we selected were: sport, team, year, stadium

⁵⁵ capacity, total attendance, average attendance, number of games, ticket price (in USD),

⁵⁶ and the number of home wins.

⁵⁷ **Data analysis**

⁵⁸ We used R (Version 3.6.1; R Core Team, 2019) and the R-packages *dplyr* (Version

⁵⁹ 0.8.3; Wickham et al., 2019), *forcats* (Version 0.4.0; Wickham, 2019a), *ggplot2* (Version

⁶⁰ 3.2.1; Wickham, 2016), *here* (Version 0.1; Müller, 2017), *janitor* (Version 1.2.0; Firke,

⁶¹ 2019), *kableExtra* (Version 1.1.0; Zhu, 2019), *knitr* (Version 1.25; Xie, 2015), *lme4* (Version

⁶² 1.1.21; Bates, Mächler, Bolker, & Walker, 2015), *Matrix* (Version 1.2.17; Bates & Maechler,

⁶³ 2019), *papaja* (Version 0.1.0.9842; Aust & Barth, 2018), *purrr* (Version 0.3.2; Henry &

⁶⁴ Wickham, 2019), *readr* (Version 1.3.1; Wickham, Hester, & Francois, 2018), *rio* (Version

⁶⁵ 0.5.16; Chan, Chan, Leeper, & Becker, 2018), *stringr* (Version 1.4.0; Wickham, 2019b),

⁶⁶ *tibble* (Version 2.1.3; Müller & Wickham, 2019), *tidyr* (Version 1.0.0; Wickham & Henry,

⁶⁷ 2019), and *tidyverse* (Version 1.2.1; Wickham, 2017) for all our analyses.

⁶⁸ **Results**

⁶⁹ In all four leagues, it turns out that average ticket price and average home win rate

⁷⁰ have a positive relationship with average home attendance even though NFL fans seem to

⁷¹ be not as sensitive to win as other three leagues' fans are. In addition, the 2008 financial

⁷² crisis did not result in crisis in the US professional sports leagues. While some of the

73 leagues experienced slight decrease in home attendance, all four leagues got through

74 economic downturn as if there was nothing happened. Actually, their home game average

75 revenue went up after financial crisis. We assume this counterintuitive outcome is

76 attributed to the facts that sports were kind of immuned to the crisis for some reason or

77 even during recession people are willing to spend money for watching games.

Discussion

79   Sports continue to play an important role in the United States. In an time when

80 individuals are becoming increasingly isolated

81 [Chalmers2012differences;Shachar2011brands], sports games provide a form of

82 entertainment that can be bring people together, whether that be through watching the

83 game at the sadium or field or on television. While the motivation to watch sports differs

84 for individuals, the widespread appeal of watching teams compete provides a context for

85 marketers to understand sponshorship, group marketing strategies, and targeted

86 advertising. The current exploratory study provides inital insight into how major league

87 attendance varries over time both in regard to attedance as well as ticket prices. Through

88 the analysis, it is clear that each of the major league sports operates very differently from

89 eachother inregard to the variables of interest isolated for the purposes of this research. As

90 this dataset it used going forward, it will be important to identify more clearly the

91 differences between each of the sports to understand if they can indeed be collapsed into an

92 overarching category of "major league sports attendance" across all four major league

93 sports (MLB, NBA, NFL, NHL). Another aspect that was not taken into account in the

94 current research is team-specific factors including how long the team has been in a city as

95 well as how many time a team has moved.

96   **References**

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

97   Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown.*

98   Retrieved from https://github.com/crsh/papaja

99   Bates, D., & Maechler, M. (2019). *Matrix: Sparse and dense matrix classes and methods.*

100   Retrieved from https://CRAN.R-project.org/package=Matrix

101 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models

102   using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

103 Chan, C.-h., Chan, G. C., Leeper, T. J., & Becker, J. (2018). *Rio: A swiss-army knife for*

104   *data file i/o.*

105 Coates, D., & Humphreys, B. R. (2007). Ticket prices, concessions and attendance at

106   professional sporting events. *International Journal of Sport Finance*, *2*(3), 161.

107 Firke, S. (2019). *Janitor: Simple tools for examining and cleaning dirty data.* Retrieved

108   from https://CRAN.R-project.org/package=janitor

109 Henry, L., & Wickham, H. (2019). *Purrr: Functional programming tools.* Retrieved from

110   https://CRAN.R-project.org/package=purrr

111 Humphreys, B. R. (2010). The impact of the global financial crisis on sport in north

112   america. In *Optimal strategies in sports economics and management* (pp. 39–57).

113   Springer.

114 Müller, K. (2017). *Here: A simpler way to find your files.* Retrieved from

115   https://CRAN.R-project.org/package=here

116 Müller, K., & Wickham, H. (2019). *Tibble: Simple data frames.* Retrieved from

117   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