

EDLD 610 Final Project

Woocheol Kim¹ & Jessica Canfield¹

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

One or two sentences providing a **basic introduction** to the field, comprehensible to a scientist in any discipline.

Two to three sentences of **more detailed background**, comprehensible to scientists in related disciplines.

One sentence clearly stating the **general problem** being addressed by this particular study.

One sentence summarizing the main result (with the words “**here we show**” or their equivalent).

Two or three sentences explaining what the **main result** reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge.

One or two sentences to put the results into a more **general context**.

Two or three sentences to provide a **broader perspective**, readily comprehensible to a scientist in any discipline.

Keywords: sports, NBA, NHL, NFL, MLB

Word count: X

26

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() %>%
  mutate(attend_tot = as.numeric(attend_tot),
         attend_avg = as.numeric(attend_avg),
         capacity = as.numeric(capacity),
         games = as.numeric(games),
         ticket_price = as.numeric(ticket_price),
         home_wins = as.numeric(home_wins)) #changed attent_tot, games, home_wins, ticket_price
str(mlb) #I added code to see the structure of each sports dataset when I was trying to

```

```

27 ## Classes 'tbl_df', 'tbl' and 'data.frame':    528 obs. of  9 variables:
28 ## $ sport      : chr  "MLB" "MLB" "MLB" "MLB" ...
29 ## $ team       : chr  "(Anaheim) Angels" "(Anaheim) Angels" "(Anaheim) Angels" "(Anaheim) Angels" ...
30 ## $ year       : num  2000 2001 2002 2003 2004 ...
31 ## $ capacity   : num  65000 65000 65000 45050 45050 ...
32 ## $ attend_tot : num  2066977 2000919 2305565 3061094 3375677 ...
33 ## $ attend_avg : num  25118 24702 28463 37791 41675 ...
34 ## $ games      : num  81 81 81 82 81 81 81 81 81 81 ...
35 ## $ ticket_price: num  13.2 13.4 11.8 16 16.6 ...
36 ## $ home_wins  : num  46 39 54 45 45 49 45 54 50 49 ...

```

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

```
37 ## [1] FALSE
```

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

nba <- nba %>% mutate(capacity = as.numeric(capacity), attend_tot = as.numeric(attend_tot))
str(nba)
```

```
38 ## Classes 'tbl_df', 'tbl' and 'data.frame':    480 obs. of  9 variables:
39 ## $ sport      : chr  "NBA" "NBA" "NBA" "NBA" ...
40 ## $ team       : chr  "Hawks" "Hawks" "Hawks" "Hawks" ...
41 ## $ year       : num  2000 2001 2002 2003 2004 ...
42 ## $ capacity   : num  19445 19445 19445 19445 19445 ...
43 ## $ attend_tot : num  560324 506110 528644 565728 586390 ...
44 ## $ attend_avg : num  13666 12344 12894 13798 14302 ...
45 ## $ games      : num  41 41 41 41 41 41 41 41 41 41 ...
46 ## $ ticket_price: num  45.9 42.8 37.5 37.5 37.7 ...
47 ## $ home_wins  : num  18 23 26 18 9 18 18 25 31 34 ...
```

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

```
48 ## Classes 'tbl_df', 'tbl' and 'data.frame':    2064 obs. of  9 variables:
49 ## $ sport      : chr  "NCAAF" "NCAAF" "NCAAF" "NCAAF" ...
50 ## $ team       : chr  "Air Force" "Air Force" "Air Force" "Air Force" ...
51 ## $ year       : num  2000 2001 2002 2003 2004 ...
52 ## $ capacity   : num  52480 52480 52480 52480 52480 ...
53 ## $ attend_tot : num  255357 230631 298993 235259 266302 ...
54 ## $ attend_avg : num  42560 38439 42713 39210 38043 ...
55 ## $ games      : num  6 6 7 6 7 ...
56 ## $ ticket_price: logi  NA NA NA NA NA NA ...
57 ## $ home_wins   : num  NA NA NA NA NA NA NA NA NA NA ...
```

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

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

```
58 ## Classes 'tbl_df', 'tbl' and 'data.frame':    528 obs. of  9 variables:
```

```

59 ## $ sport      : chr  "NFL" "NFL" "NFL" "NFL" ...
60 ## $ team       : chr  "Arizona Cardinals" "Arizona Cardinals" "Arizona Cardinals" "Ar
61 ## $ year       : num  2000 2001 2002 2003 2004 ...
62 ## $ capacity   : num  73379 73379 73379 NA 71706 ...
63 ## $ attend_tot : num  387475 307315 327272 288499 300267 ...
64 ## $ attend_avg : num  NA NA NA NA NA ...
65 ## $ games      : num  8 8 8 8 8 8 8 8 8 8 ...
66 ## $ ticket_price: num  39.6 37.6 33.7 36 39.7 ...
67 ## $ home_wins  : num  3 3 3 4 5 3 3 6 6 4 ...

```

```

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

```

```

68 ## Classes 'tbl_df', 'tbl' and 'data.frame':   480 obs. of  9 variables:
69 ## $ sport      : chr  "NHL" "NHL" "NHL" "NHL" ...
70 ## $ team       : chr  "Ducks" "Ducks" "Ducks" "Ducks" ...
71 ## $ year       : num  2000 2001 2002 2003 2004 ...
72 ## $ capacity   : num  17174 17174 17174 17174 17174 ...
73 ## $ attend_tot : chr  "553470" "492089" "573524" "614476" ...
74 ## $ attend_avg : chr  "13499" "12002" "13998" "14987" ...
75 ## $ games      : num  41 41 41 41 0 41 41 41 41 41 ...
76 ## $ ticket_price: chr  "N/A" "N/A" "N/A" "N/A" ...
77 ## $ home_wins  : num  15 15 22 19 0 26 26 28 20 25 ...

```

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

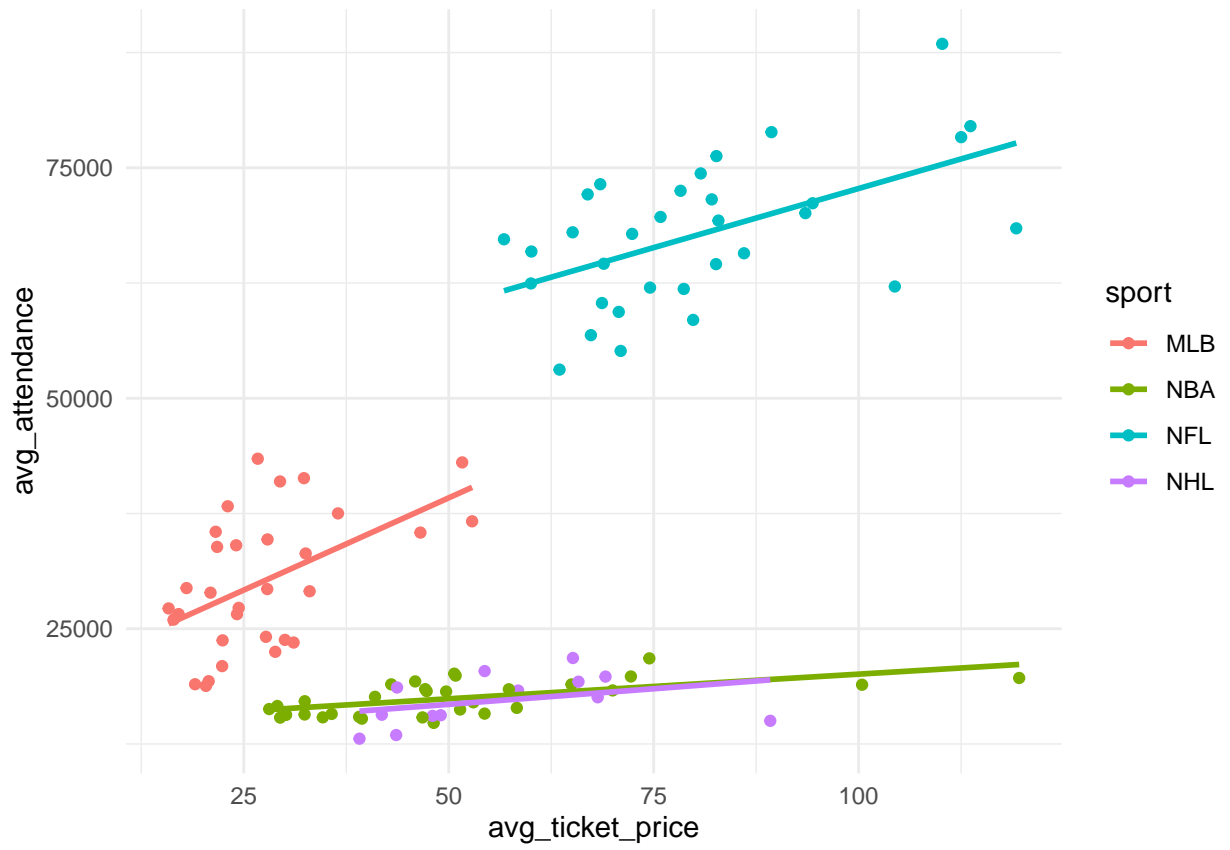
sports <- bind_rows(mlb, nba, ncaaf, nfl, nhl) %>% as_tibble()

str(mlb)
```

```
78 ## Classes 'tbl_df', 'tbl' and 'data.frame':    528 obs. of  9 variables:
79 ## $ sport      : chr  "MLB" "MLB" "MLB" "MLB" ...
80 ## $ team       : chr  "(Anaheim) Angels" "(Anaheim) Angels" "(Anaheim) Angels" "(Anaheim) Angels" ...
81 ## $ year       : num  2000 2001 2002 2003 2004 ...
82 ## $ capacity   : num  65000 65000 65000 45050 45050 ...
83 ## $ attend_tot : num  2066977 2000919 2305565 3061094 3375677 ...
84 ## $ attend_avg : num  25118 24702 28463 37791 41675 ...
85 ## $ games      : num  81 81 81 82 81 81 81 81 81 81 ...
86 ## $ ticket_price: num  13.2 13.4 11.8 16 16.6 ...
87 ## $ home_wins   : num  46 39 54 45 45 49 45 54 50 49 ...
```

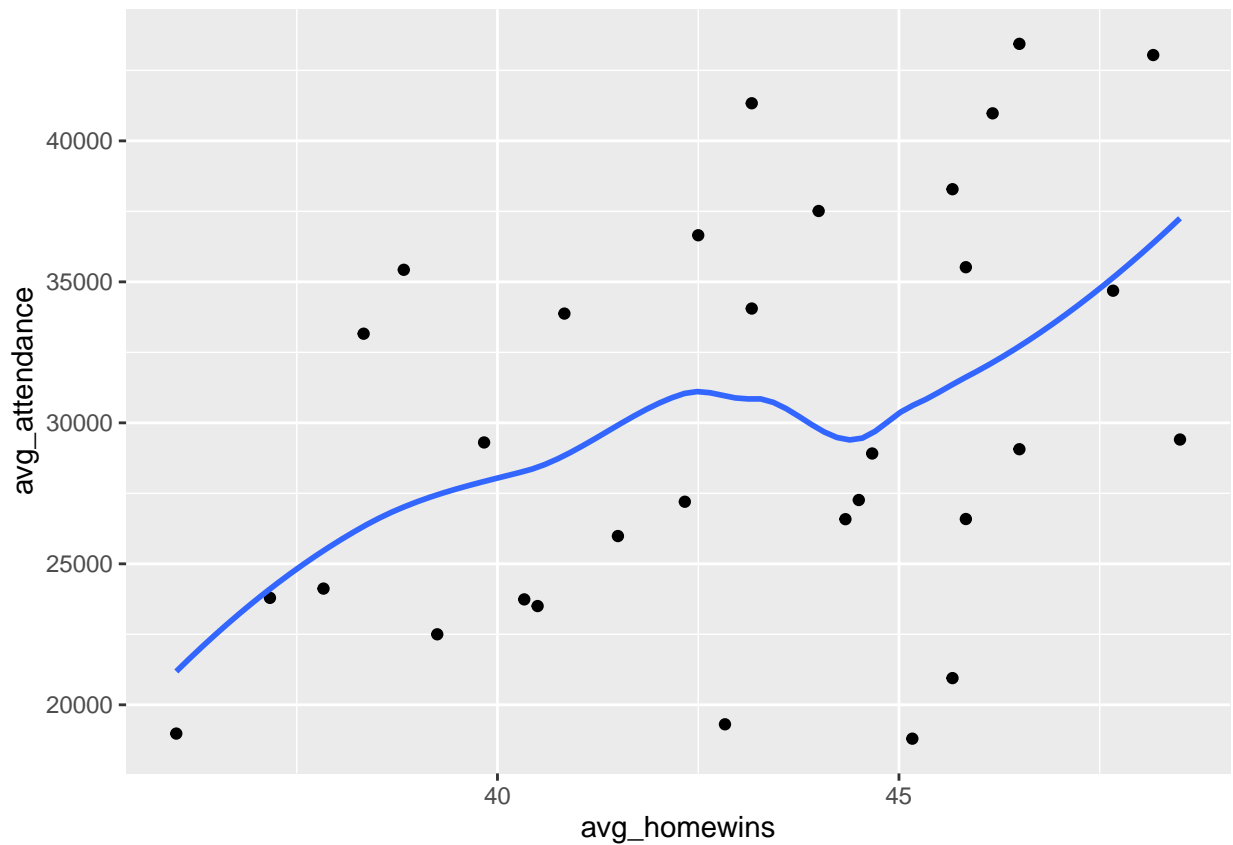
```
sports_rev <- sports %>%
  drop_na() %>%
  filter(year >= 2010) %>%
  group_by(team, sport) %>%
  summarize(avg_ticket_price = mean(ticket_price), avg_homewins = mean(home_wins), avg_attend_avg = mean(attend_avg))

sports_rev %>%
  ggplot(aes(avg_ticket_price, avg_attendance, color = sport)) +
  geom_point() +
  geom_smooth(method = lm, se = FALSE) +
  theme_minimal()
```



88

```
sports_rev %>%  
  filter(sport == "MLB") %>%  
  ggplot(aes(avg_homewins, avg_attendance)) +  
  geom_point() +  
  geom_smooth(se = FALSE)
```

89

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

90 Requirements

- 91 1. pivot_longer: Done
- 92 2. pivot_wider: Done
- 93 3. group_by: Done
- 94 4. summarize: Done
- 95 5. filter: Done

96 6. select: Done

97 7. Mutate: Done

98 **Methods**

99 We report how we determined our sample size, all data exclusions (if any), all
100 manipulations, and all measures in the study.

101 **Participants**

102 **Material**

103 **Procedure**

104 **Data analysis**

105 We used R (Version 3.6.1; R Core Team, 2019) for all our analyses.

106 **Results**

107 **Discussion**

References

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

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/>