NHL vs AHL 2023-2024 Attendance

Taylor Fourier

2024-09-28

```
#install.packages("ggplot2")
library(ggplot2)

#install.packages("dplyr")
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
#install.packages("tibble")
library(tibble)

# NHL 2023-2024 Average Attendance Statistics based off of "https://www.hockeydb.com/nhl-attendance/att

nhl_attendance <- c("Montreal Canadiens", "MTL", 21099,
                    "Edmonton Oilers", "EDM", 19173,
                    "Tampa Bay Lightning", "TBL", 19092,
                    "Detroit Red Wings", "DET", 18980,
                    "Chicago Blackhawks", "CHI", 18836,
                    "Vancouver Canucks", "VAN", 18826,
                    "Carolina Hurricanes", "CAR", 18798,
                    "Toronto Maple Leafs", "TOR", 18789,
                    "Florida Panthers", "FLA", 18632,
                    "Dallas Stars", "DAL", 18532,
                    "Minnesota Wild", "MIN", 18529,
                    "Philadelphia Flyers", "PHI", 18438,
                    "Vegas Golden Knights", "VGK", 18139,
                    "Colorado Avalanche", "COL", 18103,
                    "New York Islanders", "NYI", 18099,
                    "St. Louis Blues", "STL", 18084,
                    "New York Rangers", "NYR", 17983,
                    "Los Angeles Kings", "LAK", 17945,
                    "Pittsburgh Penguins", "PIT", 17909,
                    "Seattle Kraken", "SEA", 17887,
                    "Boston Bruins", "BOS", 17850,
                    "Washington Capitals", "WSH", 17841,
```

```
                      "New Jersey Devils", "NJD", 17598,
                      "Ottawa Senators", "OTT", 17580,
                      "Calgary Flames", "CGY",17501,
                      "Nashville Predators", "NSH", 17306,
                      "Columbus Blue Jackets", "CBJ", 17016,
                      "Buffalo Sabres", "BUF", 15981,
                      "Anaheim Ducks", "ANA", 15686,
                      "San Jose Sharks", "SJS", 13559,
                      "Winnipeg Jets", "WPG", 13490,
                      "Arizona Coyotes", "ARI", 4600)

nhl_attendance <- tibble(
    Team = nhl_attendance[c(TRUE, FALSE, FALSE)],
    Team_Abbreviation = nhl_attendance[c(FALSE, TRUE, FALSE)],
    Attendance = as.numeric(nhl_attendance[c(FALSE, FALSE, TRUE)])
    )

print(nhl_attendance)
```

```
## # A tibble: 32 x 3
##    Team                Team_Abbreviation Attendance
##    <chr>               <chr>                  <dbl>
##  1 Montreal Canadiens  MTL                    21099
##  2 Edmonton Oilers     EDM                    19173
##  3 Tampa Bay Lightning TBL                    19092
##  4 Detroit Red Wings   DET                    18980
##  5 Chicago Blackhawks  CHI                    18836
##  6 Vancouver Canucks   VAN                    18826
##  7 Carolina Hurricanes CAR                    18798
##  8 Toronto Maple Leafs TOR                    18789
##  9 Florida Panthers    FLA                    18632
## 10 Dallas Stars        DAL                    18532
## # i 22 more rows
```

```
# NHL 2023-2024 Average Attendance Statistics based off of "https://www.hockeydb.com/nhl-attendance/att
```

```
ahl_attendance <- c("Cleveland Monsters", "CLE", 10347,
                    "Hershey Bears", "HER", 9439,
                    "Laval Rocket", "LAV", 9256,
                    "Chicago Wolves", "CHI", 8984,
                    "Coachella Valley Firebirds","CV", 8844,
                    "Providence Bruins", "PRO", 7713,
                    "Grand Rapids Griffins", "GR", 7641,
                    "Ontario Reign", "ONT", 7469,
                    "San Diego Gulls", "SD", 7249,
                    "Charlotte Checkers", "CLT", 6979,
                    "Lehigh Valley Phantoms", "LV", 6710,
                    "Iowa Wild", "IA", 6401,
                    "Springfield Thunderbirds", "SPR", 6321,
                    "Milwaukee Admirals", "MIL", 6139,
                    "Rochester Americans", "ROC", 5994,
                    "Texas Stars", "TEX", 5962,
                    "Toronto Marlies", "TOR", 5889,
                    "Syracuse Crunch", "SYR", 5477,
```

```
                    "Hartford Wolf Pack", "HFD", 5456,
                    "Colorado Eagles", "COL", 5089,
                    "Bakersfield Condors", "BAK", 4823,
                    "Abbotsford Canucks", "ABB", 4816,
                    "Wilkes-Barre/Scranton Penguins", "WBS", 4768,
                    "Rockford IceHogs", "ROK", 4516,
                    "Henderson Silver Knights", "HSK", 4144,
                    "Tucson Roadrunners", "TUC", 4123,
                    "Calgary Wranglers", "CGY", 4101,
                    "Manitoba Moose", "MB", 3898,
                    "Bridgeport Islanders","BRI", 3167,
                    "Utica Comets", "UTC", 2954,
                    "Belleville Senators", "BEL", 2738,
                    "San Jose Barracuda", "SJ", 2116)

ahl_attendance <- tibble(
    Team = ahl_attendance[c(TRUE, FALSE, FALSE)],
    Team_Abbreviation = ahl_attendance[c(FALSE, TRUE, FALSE)],
    Attendance = as.numeric(ahl_attendance[c(FALSE, FALSE, TRUE)])
    )

print(ahl_attendance)
```

```
## # A tibble: 32 x 3
##    Team                       Team_Abbreviation Attendance
##    <chr>                      <chr>                  <dbl>
##  1 Cleveland Monsters         CLE                    10347
##  2 Hershey Bears              HER                     9439
##  3 Laval Rocket               LAV                     9256
##  4 Chicago Wolves             CHI                     8984
##  5 Coachella Valley Firebirds CV                      8844
##  6 Providence Bruins          PRO                     7713
##  7 Grand Rapids Griffins      GR                      7641
##  8 Ontario Reign              ONT                     7469
##  9 San Diego Gulls            SD                      7249
## 10 Charlotte Checkers         CLT                     6979
## # i 22 more rows
```

```
affiliate_data  <- c("Anaheim Ducks", "San Diego Gulls",
                     "Arizona Coyotes", "Tucson Roadrunners",
                     "Boston Bruins", "Providence Bruins",
                     "Buffalo Sabres", "Rochester Americans",
                     "Calgary Flames", "Calgary Wranglers",
                     "Carolina Hurricanes", "Chicago Wolves",
                     "Chicago Blackhawks", "Rockford IceHogs",
                     "Colorado Avalanche", "Colorado Eagles",
                     "Columbus Blue Jackets", "Cleveland Monsters",
                     "Dallas Stars", "Texas Stars",
                     "Detroit Red Wings", "Grand Rapids Griffins",
                     "Edmonton Oilers", "Bakersfield Condors",
                     "Florida Panthers", "Charlotte Checkers",
                     "Los Angeles Kings", "Ontario Reign",
                     "Minnesota Wild", "Iowa Wild",
                     "Montreal Canadiens", "Laval Rocket",
```

```
                    "Nashville Predators", "Milwaukee Admirals",
                    "New Jersey Devils", "Utica Comets",
                    "New York Islanders", "Bridgeport Islanders",
                    "New York Rangers", "Hartford Wolf Pack",
                    "Ottawa Senators", "Belleville Senators",
                    "Philadelphia Flyers", "Lehigh Valley Phantoms",
                    "Pittsburgh Penguins", "Wilkes-Barre/Scranton Penguins",
                    "San Jose Sharks", "San Jose Barracuda",
                    "Seattle Kraken", "Coachella Valley Firebirds",
                    "St. Louis Blues", "Springfield Thunderbirds",
                    "Tampa Bay Lightning", "Syracuse Crunch",
                    "Toronto Maple Leafs", "Toronto Marlies",
                    "Vancouver Canucks", "Abbotsford Canucks",
                    "Vegas Golden Knights", "Henderson Silver Knights",
                    "Washington Capitals", "Hershey Bears",
                    "Winnipeg Jets", "Manitoba Moose")

affiliate_data  <- tibble(
    NHL_team = affiliate_data [c(TRUE, FALSE)],
    AHL_affiliate = affiliate_data [c(FALSE, TRUE)],
    )

print(affiliate_data)
```

```
## # A tibble: 32 x 2
##    NHL_team              AHL_affiliate
##    <chr>                 <chr>
##  1 Anaheim Ducks         San Diego Gulls
##  2 Arizona Coyotes       Tucson Roadrunners
##  3 Boston Bruins         Providence Bruins
##  4 Buffalo Sabres        Rochester Americans
##  5 Calgary Flames        Calgary Wranglers
##  6 Carolina Hurricanes   Chicago Wolves
##  7 Chicago Blackhawks    Rockford IceHogs
##  8 Colorado Avalanche    Colorado Eagles
##  9 Columbus Blue Jackets Cleveland Monsters
## 10 Dallas Stars          Texas Stars
## # i 22 more rows
```

```
complete_attendancedata <- nhl_attendance %>%
    inner_join(affiliate_data, by = c("Team" = "NHL_team")) %>%
    inner_join(ahl_attendance, by = c("AHL_affiliate" = "Team"))%>%
    rename(
      NHL_Team = Team,
      NHL_Team_Abr = Team_Abbreviation.x,
      NHL_Attendance = Attendance.x,
      AHL_Affiliate = AHL_affiliate,
      AHL_Affiliate_Abr = Team_Abbreviation.y,
      AHL_Attendance = Attendance.y
    )
View(complete_attendancedata)
  print(complete_attendancedata)
```

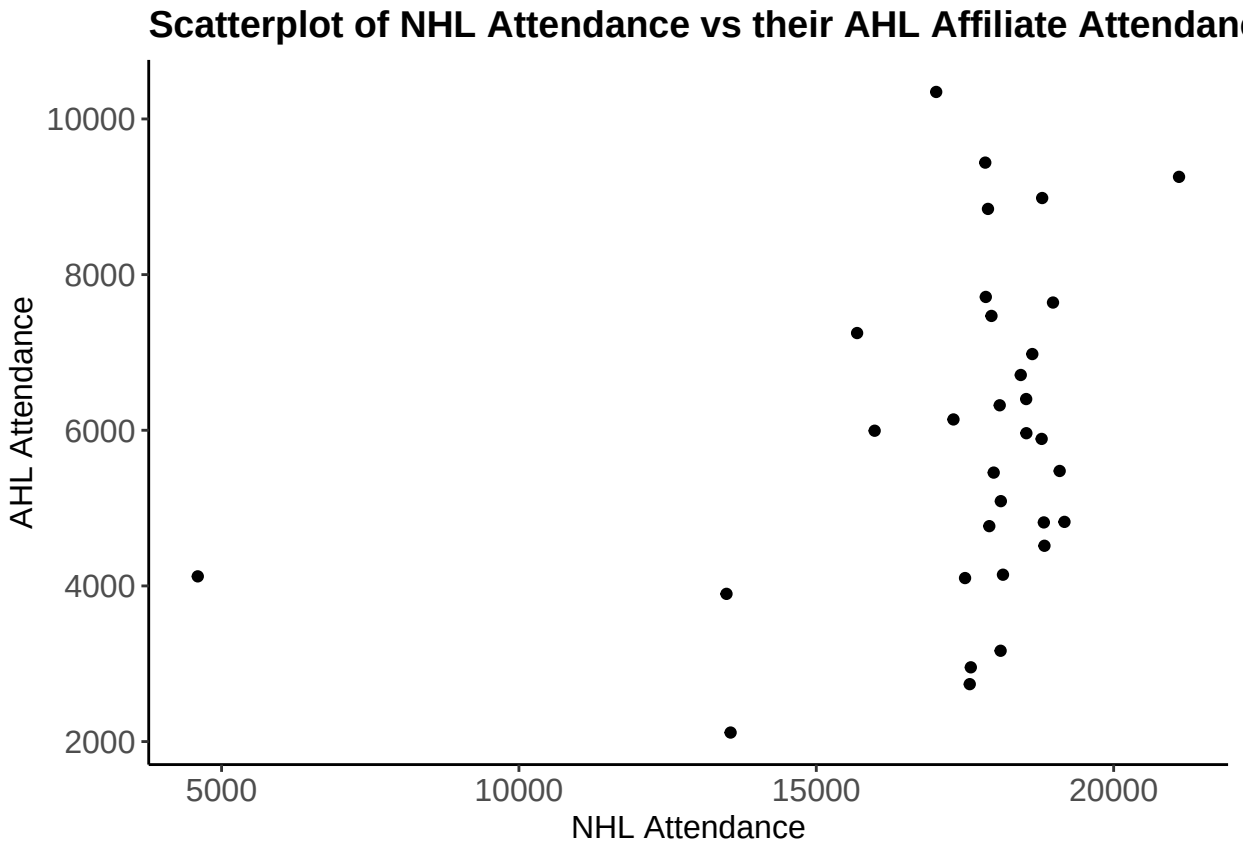
```
## # A tibble: 32 x 6
```

```
##    NHL_Team          NHL_Team_Abr NHL_Attendance AHL_Affiliate AHL_Affiliate_Abr
##    <chr>             <chr>                <dbl> <chr>         <chr>
##  1 Montreal Canadie~ MTL                  21099 Laval Rocket  LAV
##  2 Edmonton Oilers   EDM                  19173 Bakersfield ~ BAK
##  3 Tampa Bay Lightn~ TBL                  19092 Syracuse Cru~ SYR
##  4 Detroit Red Wings DET                  18980 Grand Rapids~ GR
##  5 Chicago Blackhaw~ CHI                  18836 Rockford Ice~ ROK
##  6 Vancouver Canucks VAN                  18826 Abbotsford C~ ABB
##  7 Carolina Hurrica~ CAR                  18798 Chicago Wolv~ CHI
##  8 Toronto Maple Le~ TOR                  18789 Toronto Marl~ TOR
##  9 Florida Panthers  FLA                  18632 Charlotte Ch~ CLT
## 10 Dallas Stars      DAL                  18532 Texas Stars   TEX
## # i 22 more rows
## # i 1 more variable: AHL_Attendance <dbl>
```

```r
#install.packages("tidyverse")
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v readr     2.1.5
## v lubridate 1.9.3     v stringr   1.5.1
## v purrr     1.0.2     v tidyr     1.3.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

```r
ggplot(data = complete_attendancedata,
         mapping = aes(x = NHL_Attendance, y = AHL_Attendance)) +
    geom_point() +
    labs(
      title = "Scatterplot of NHL Attendance vs their AHL Affiliate Attendance",
      x = "NHL Attendance",
      y = "AHL Attendance",
    ) +
    theme_classic() +
    theme(
      plot.title = element_text(size = 14, face = "bold"),
      axis.text = element_text(size = 12),
      axis.title = element_text(size = 12)
    )
```




```
correlation <- cor(complete_attendancedata$NHL_Attendance, y =    complete_attendancedata$AHL_Attendanc
cat("Correlation: ", correlation)
```

```
## Correlation:  0.3124337
```

```
regression <- lm(AHL_Attendance ~ NHL_Attendance, data = complete_attendancedata)
summary(regression)
```

```
##
## Call:
## lm(formula = AHL_Attendance ~ NHL_Attendance, data = complete_attendancedata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3219.0 -1453.2     0.6  1372.2  4522.8
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1816.8090  2306.8575   0.788   0.4371
## NHL_Attendance    0.2355     0.1307   1.801   0.0817 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2016 on 30 degrees of freedom
## Multiple R-squared:  0.09761,    Adjusted R-squared:  0.06754
## F-statistic: 3.245 on 1 and 30 DF,  p-value: 0.08169
```

```r
# Capacity Information NHL from https://geojango.com/pages/list-of-nhl-teams

#install.packages("rvest")
#install.packages("stringr")
library(rvest)
```

```
##
## Attaching package: 'rvest'
```

```
## The following object is masked from 'package:readr':
##
##     guess_encoding
```

```r
library(stringr)

url <- "https://geojango.com/pages/list-of-nhl-teams"
webpage <- read_html(url)

nhl_arena_data <- webpage %>%
  html_node("table") %>%
  html_table

nhl_arena_data <- nhl_arena_data[-1, ] %>%
  rename(
    NHL_Team = X1,
    Arena_Name = X2,
    Arena_Location = X3,
    Arena_Capacity = X4,
    Opened = X5
  )
View(nhl_arena_data)
print(nhl_arena_data)
```

```
## # A tibble: 32 x 5
##    NHL_Team              Arena_Name          Arena_Location Arena_Capacity Opened
##    <chr>                 <chr>               <chr>          <chr>          <chr>
##  1 Anaheim Ducks         Honda Center        Anaheim, Cali~ 17,174         1993
##  2 Arizona Coyotes       Mullett Arena       Tempe, Arizona 4,600          2022
##  3 Boston Bruins         TD Garden           Boston, Massa~ 17,850         1995
##  4 Buffalo Sabres        KeyBank Center      Buffalo, New ~ 19,070         1996
##  5 Calgary Flames        Scotiabank Saddle~  Calgary, Albe~ 19,289         1983
##  6 Carolina Hurricanes   PNC Arena           Raleigh, Nort~ 18,700         1999
##  7 Chicago Blackhawks    United Center       Chicago, Illi~ 19,717         1994
##  8 Colorado Avalanche    Ball Arena          Denver, Color~ 17,809         1999
##  9 Columbus Blue Jackets Nationwide Arena    Columbus, Ohio 18,500         2000
## 10 Dallas Stars          American Airlines~  Dallas, Texas  18,532         2001
## # i 22 more rows
```

```r
url <- "https://en.wikipedia.org/wiki/List_of_American_Hockey_League_arenas"
webpage <- read_html(url)

ahl_arena_data <- webpage %>%
  html_node("table") %>%
  html_table() %>%
  mutate(Team = str_remove(Team, "\\[\\d+\\]")) %>%
```

```
  mutate(Capacity = str_remove(Capacity, "\\[\\d+\\]"))

View(ahl_arena_data)
print(ahl_arena_data)
```

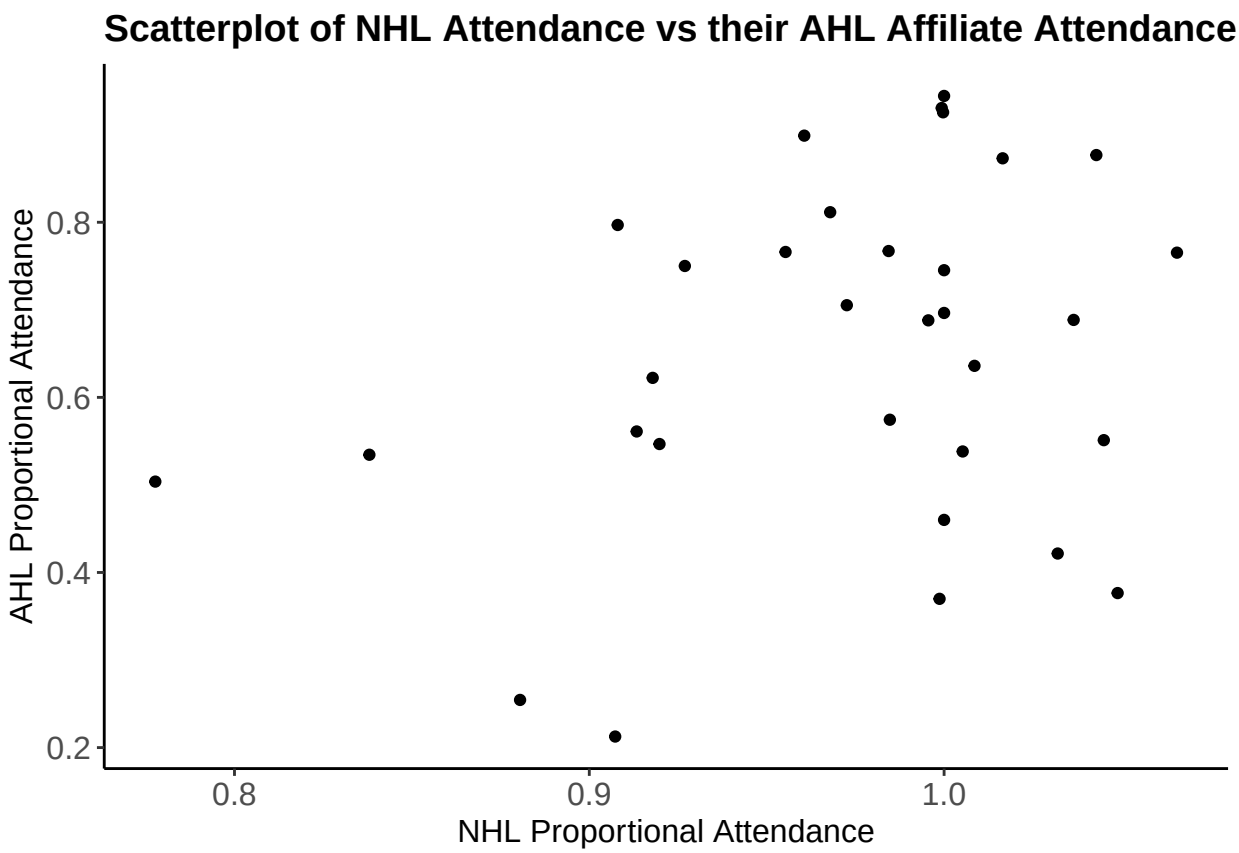
```
## # A tibble: 32 x 5
##    Arena                      Location                    Team   Capacity Opened
##    <chr>                      <chr>                       <chr>  <chr>     <int>
##  1 Abbotsford Centre          Abbotsford, British Columbia Abbot~ 7,000      2009
##  2 Adirondack Bank Center[1] Utica, New York              Utica~ 3,860      1960
##  3 Allstate Arena[2]          Rosemont, Illinois          Chica~ 16,692     1980
##  4 Acrisure Arena             Thousand Palms, California  Coach~ 10,087     2022
##  5 Amica Mutual Pavilion      Providence, Rhode Island    Provi~ 11,075     1972
##  6 Blue Arena[3]              Loveland, Colorado          Color~ 5,829      2003
##  7 Blue Cross Arena[4]        Rochester, New York         Roche~ 11,215     1955
##  8 BMO Harris Bank Center[5] Rockford, Illinois           Rockf~ 5,895      1981
##  9 Bojangles' Coliseum[6]     Charlotte, North Carolina   Charl~ 8,600      1955
## 10 CAA Arena[7]               Belleville, Ontario         Belle~ 4,400      1978
## # i 22 more rows
```

```
proportion_attendancedata <- complete_attendancedata %>%
    full_join(nhl_arena_data, by = c("NHL_Team" = "NHL_Team")) %>%
    full_join(ahl_arena_data, by = c("AHL_Affiliate" = "Team"))

proportion_attendancedata <- proportion_attendancedata %>%
  rename(NHL_Capacity = Arena_Capacity,
         AHL_Capacity = Capacity) %>%
  select(NHL_Team, NHL_Team_Abr, NHL_Attendance, NHL_Capacity,
         AHL_Affiliate, AHL_Affiliate_Abr, AHL_Attendance, AHL_Capacity) %>%
  mutate(
    NHL_Attendance = as.numeric(gsub(",", "", NHL_Attendance)),
    NHL_Capacity = as.numeric(gsub(",", "", NHL_Capacity)),
    AHL_Attendance = as.numeric(gsub(",", "", AHL_Attendance)),
    AHL_Capacity = as.numeric(gsub(",", "", AHL_Capacity)),
    NHL_Attendance_Proportion = NHL_Attendance / NHL_Capacity,
    AHL_Attendance_Proportion = AHL_Attendance / AHL_Capacity
  )

View(proportion_attendancedata)

ggplot(data = proportion_attendancedata,
         mapping = aes(x = NHL_Attendance_Proportion, y = AHL_Attendance_Proportion)) +
    geom_point() +
    labs(
      title = "Scatterplot of NHL Attendance vs their AHL Affiliate Attendance",
      x = "NHL Proportional Attendance",
      y = "AHL Proportional Attendance",
    ) +
    theme_classic() +
    theme(
      plot.title = element_text(size = 14, face = "bold"),
      axis.text = element_text(size = 12),
      axis.title = element_text(size = 12)
    )
```




```
 correlation <- cor(proportion_attendancedata$NHL_Attendance_Proportion, y = proportion_attendancedata$
  cat("Correlation: ", correlation)
```

```
## Correlation:  0.2833121
```

```
  regression <- lm(AHL_Attendance_Proportion ~ NHL_Attendance_Proportion, data = proportion_attendanceda
  summary(regression)
```

```
##
## Call:
## lm(formula = AHL_Attendance_Proportion ~ NHL_Attendance_Proportion,
##     data = proportion_attendancedata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.38097 -0.09969  0.02104  0.14599  0.27042
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -0.1925     0.5217  -0.369    0.715
## NHL_Attendance_Proportion    0.8664     0.5355   1.618    0.116
##
## Residual standard error: 0.1911 on 30 degrees of freedom
## Multiple R-squared:  0.08027,    Adjusted R-squared:  0.04961
## F-statistic: 2.618 on 1 and 30 DF,  p-value: 0.1161
```

```
plot(fitted(regression), residuals(regression),
     xlab = "Fitted Values",
```

```
     ylab = "Residuals",
     main = "Residuals vs Fitted Values")
abline(h = 0, col = "red")
```