# midterm_project

*Ningze Zu, Jiahao Xu, Andrew Zhang, Albert Ding*

```
library(readr)
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
```

```
library(ggplot2)
library(knitr)
library(scales)
```

```
##
## Attaching package: 'scales'
```

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

```
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
```

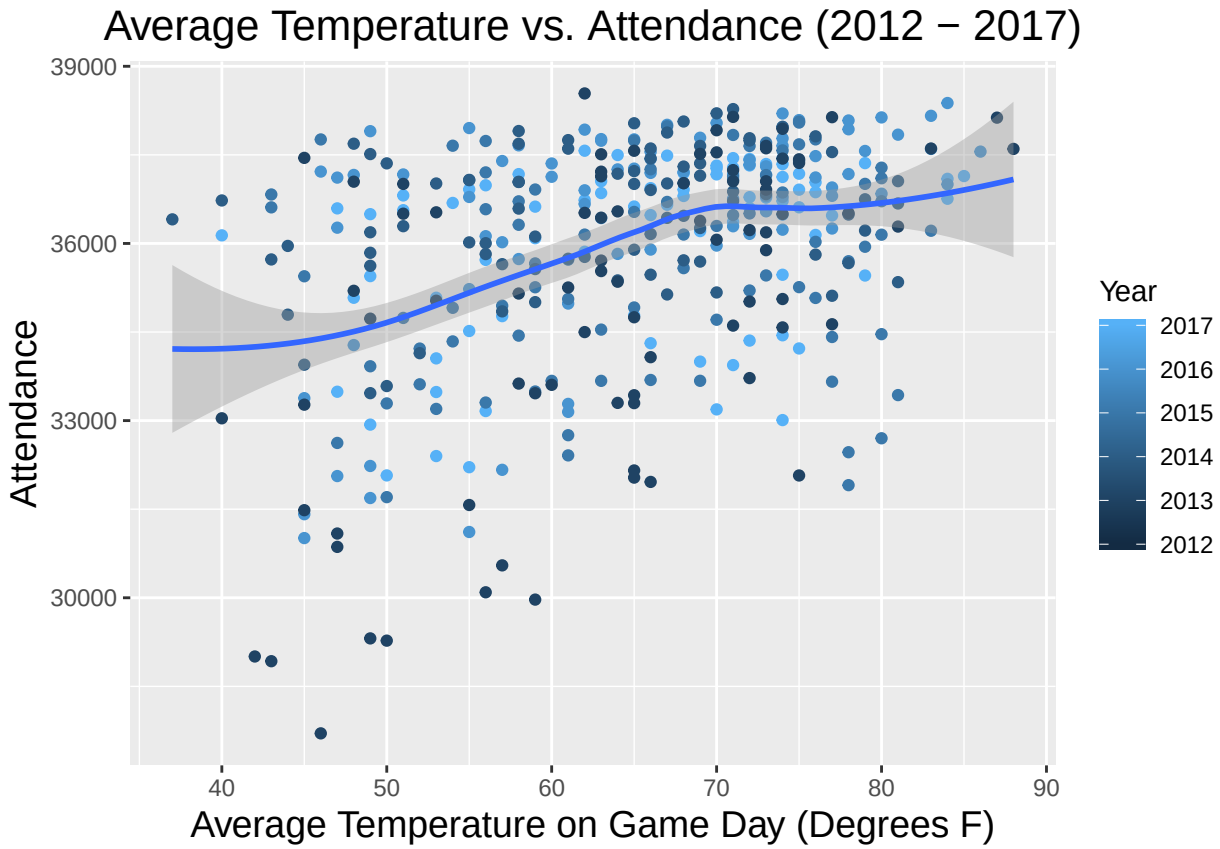
```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
# We write the scraping code for the Celtics attendance, which is in MA615 Celtics Webscrape.Rmd. And i
baseball <- read.csv("baseball_weather.csv", header = T) %>% filter(X.1 != "@")
basketball <- read.csv("basketball_weather.csv", header = T)
baseball11 <- read.csv("baseball_weather.csv", header = T) %>% filter(X.1 != "@")
baseball <- select(baseball, Gm., Year, DATE, X, Tm, Opp, W.L, Win, Loss, Save, Time, D.N, Attendance, 
write.csv(baseball, file = "baseball00.csv")
```

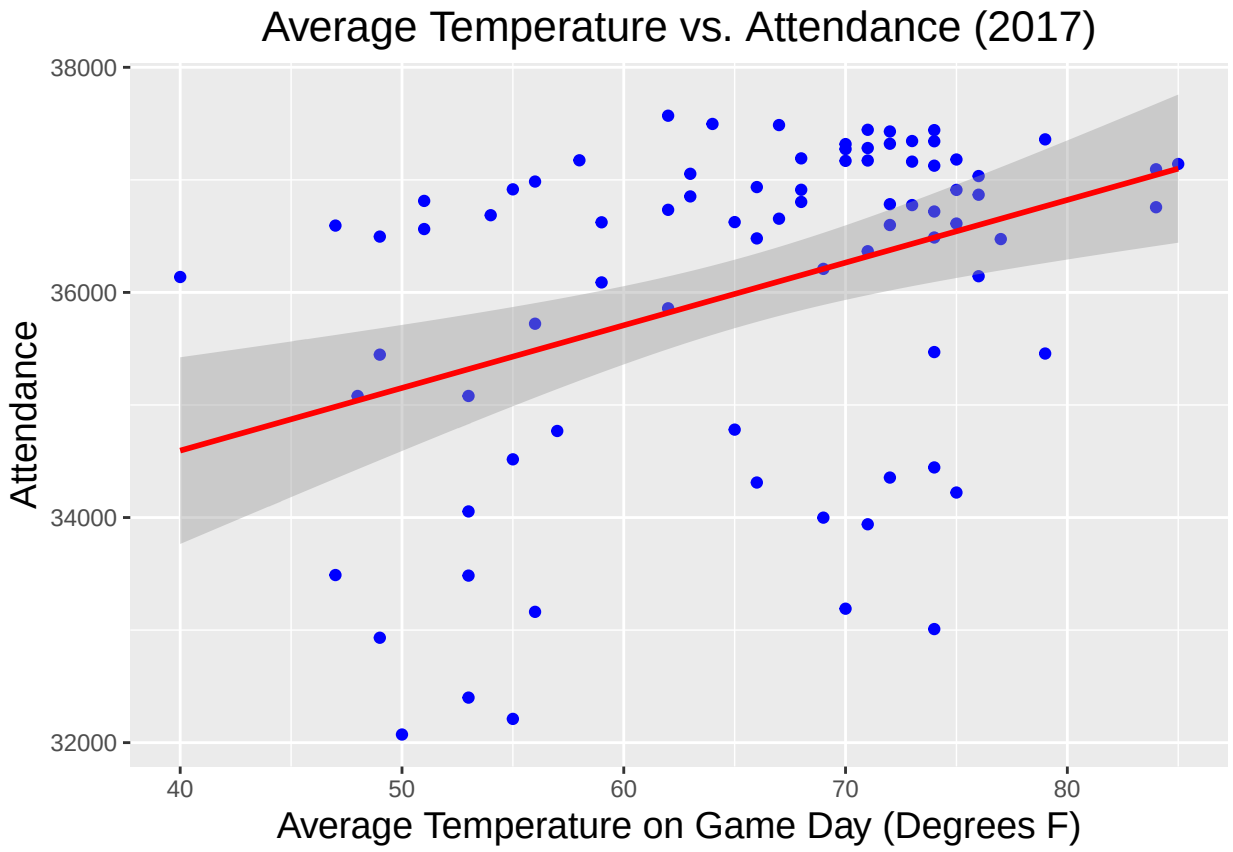
```
# There appears to be a weak but slightly positive relationship between  average temperature and attend
ggplot(baseball, mapping = aes(x = TAVG, y = Attendance)) +
  geom_point(mapping = aes(color = Year)) +
  geom_smooth() +
  ggtitle("Average Temperature vs. Attendance (2012 - 2017)") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Average Temperature on Game Day (Degrees F)") +
  ylab("Attendance")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

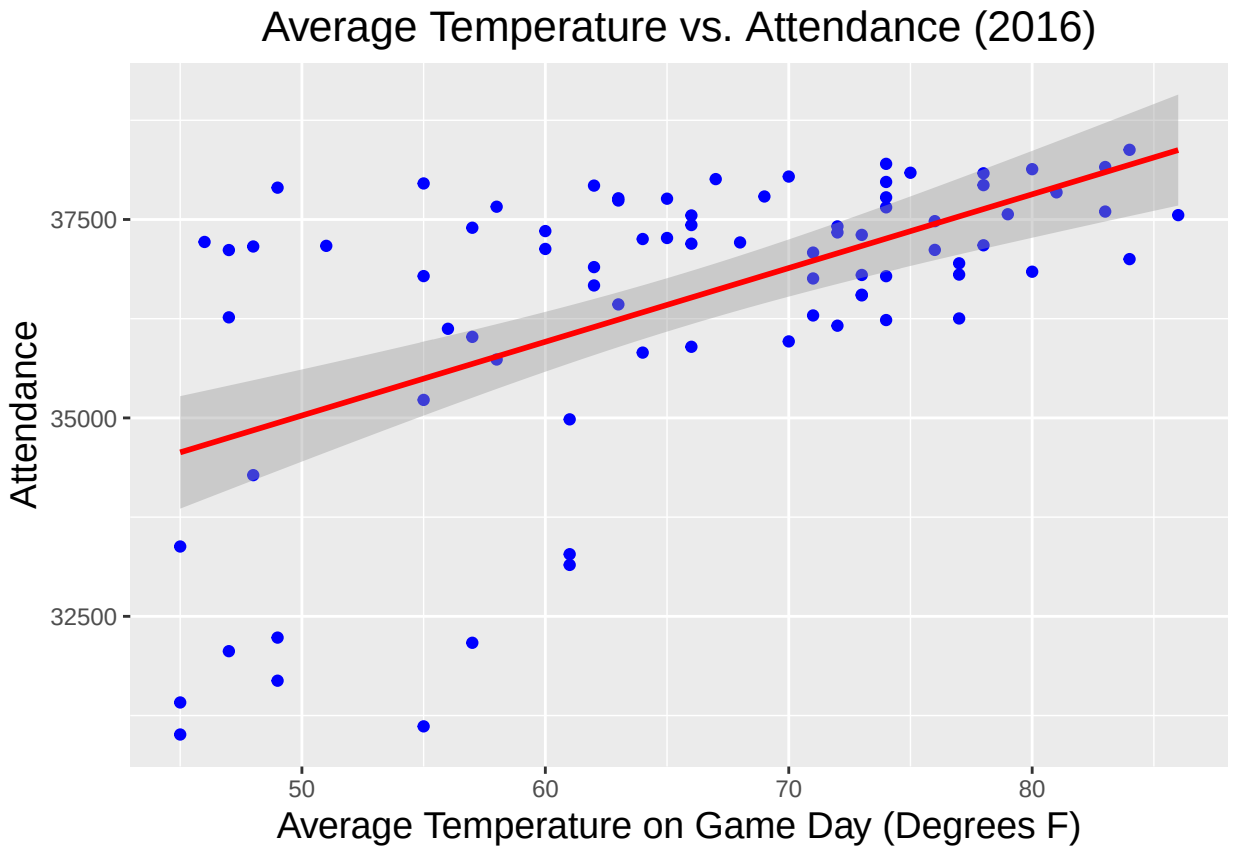
```
## Warning: Removed 81 rows containing non-finite values (stat_smooth).
## Warning: Removed 81 rows containing missing values (geom_point).
```



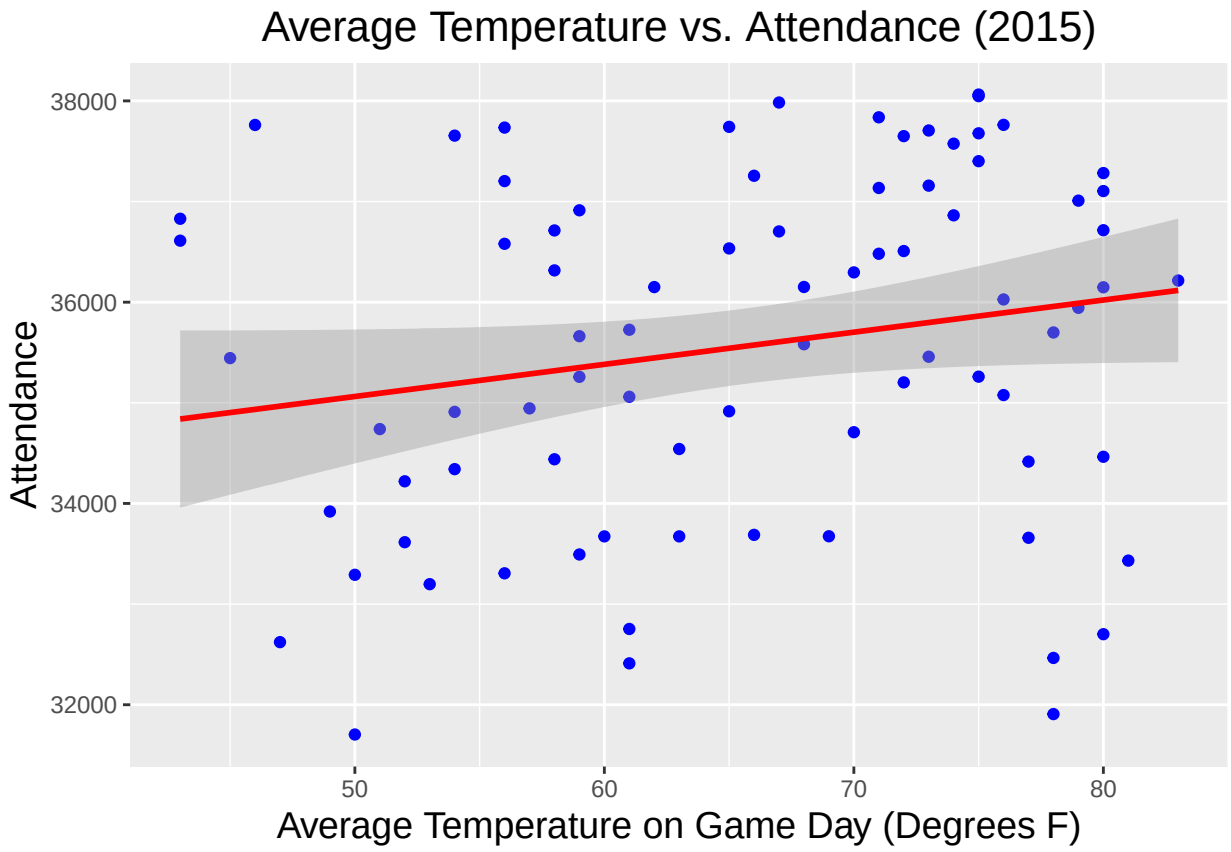

```
# 2017 Season
baseball_2017 <- baseball %>% filter(Year == 2017)
# Relationship between average temperature with attendance of season 2017
ggplot(baseball_2017, aes(TAVG, Attendance)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", color = "red") +
  ggtitle("Average Temperature vs. Attendance (2017)") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Average Temperature on Game Day (Degrees F)") +
  ylab("Attendance")
```



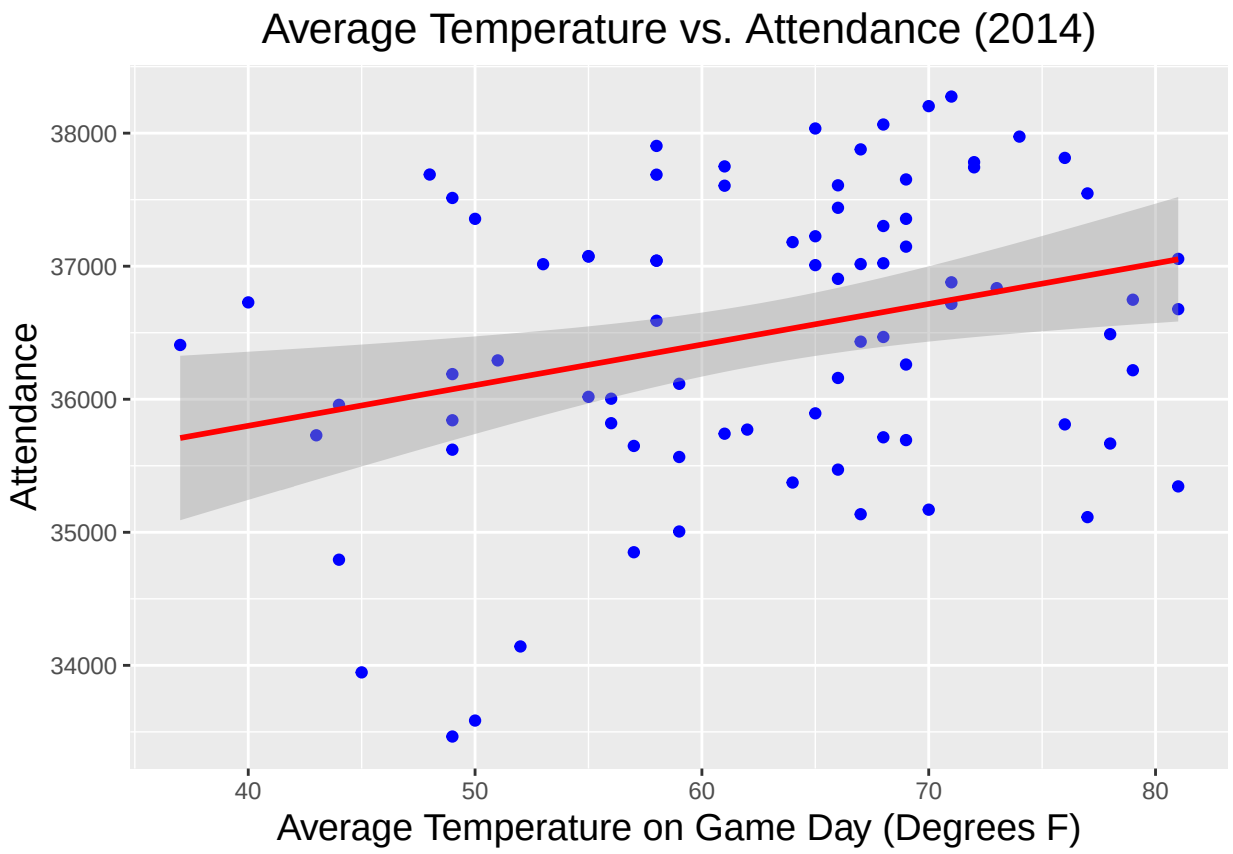

```
# 2016 Season
baseball_2016 <- baseball %>% filter(Year == 2016)
# Relationship between average temperature with attendance of season 2016
ggplot(baseball_2016, aes(TAVG, Attendance)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", color = "red") +
  ggtitle("Average Temperature vs. Attendance (2016)") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Average Temperature on Game Day (Degrees F)") +
  ylab("Attendance")
```



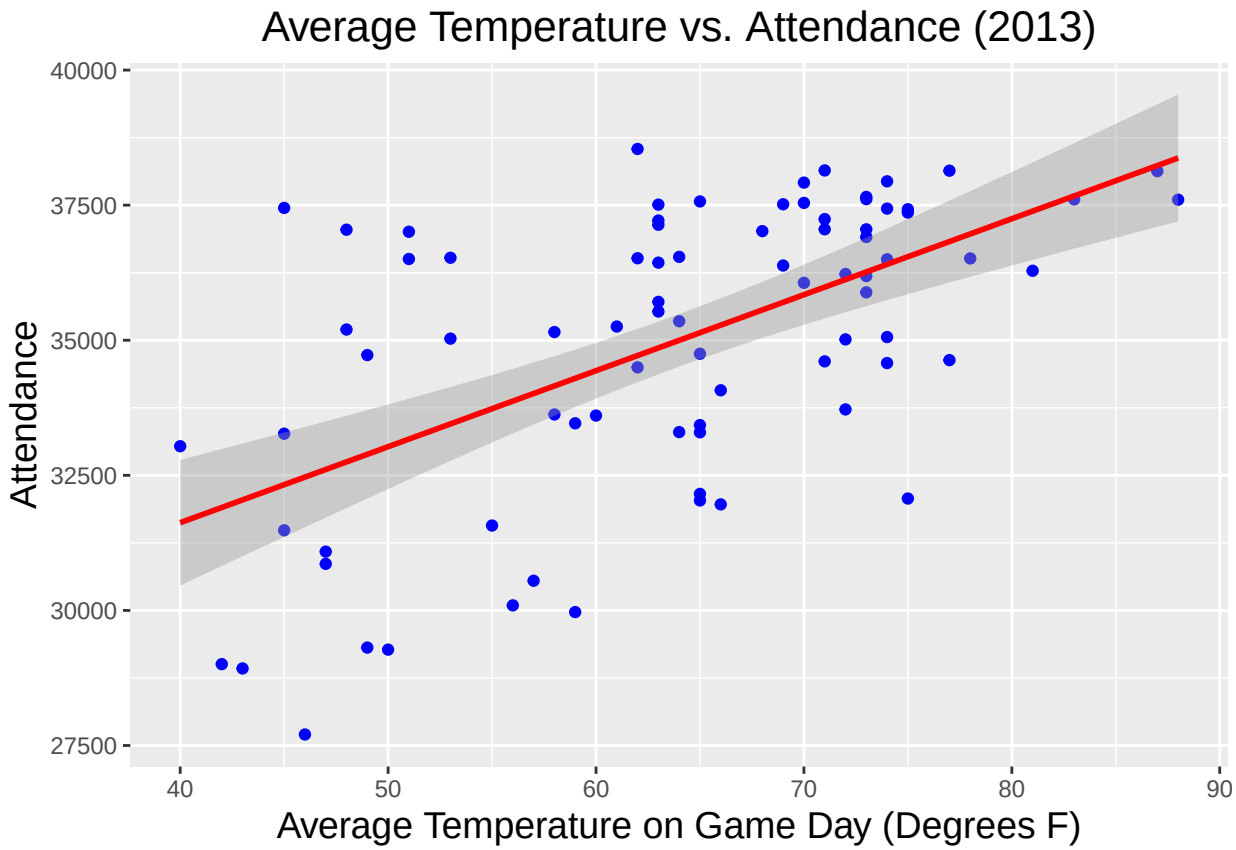

```
# 2015 Season
baseball_2015 <- baseball %>% filter(Year == 2015)
# Relationship between average temperature with attendance of season 2015
ggplot(baseball_2015, aes(TAVG, Attendance)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", color = "red") +
  ggtitle("Average Temperature vs. Attendance (2015)") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Average Temperature on Game Day (Degrees F)") +
  ylab("Attendance")
```




```
# 2014 Season
baseball_2014 <- baseball %>% filter(Year == 2014)
# Relationship between average temperature with attendance of season 2014
ggplot(baseball_2014, aes(TAVG, Attendance)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", color = "red") +
  ggtitle("Average Temperature vs. Attendance (2014)") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Average Temperature on Game Day (Degrees F)") +
  ylab("Attendance")
```

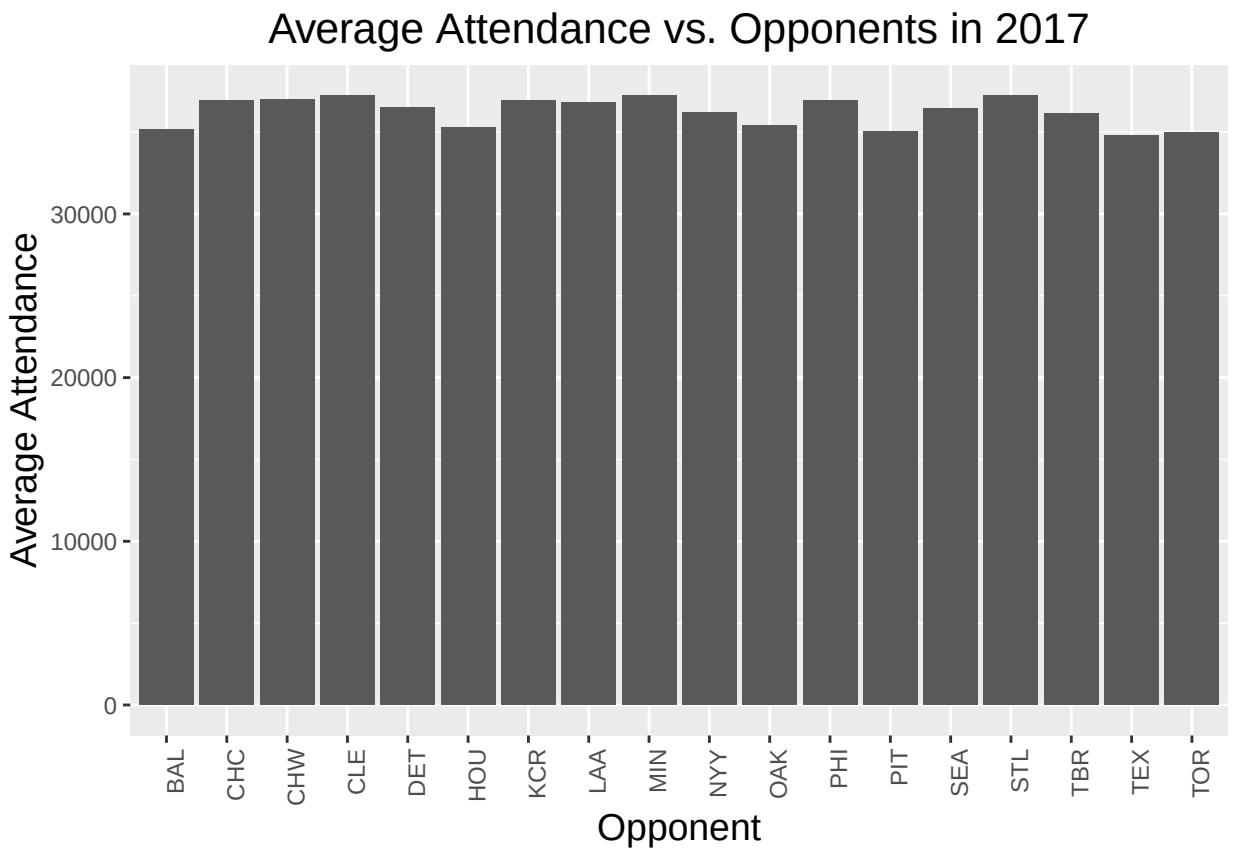



```r
# 2013 Season
baseball_2013 <- baseball %>% filter(Year == 2013)
# Relationship between average temperature with attendance of season 2013
ggplot(baseball_2013, aes(TAVG, Attendance)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", color = "red") +
  ggtitle("Average Temperature vs. Attendance (2013)") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Average Temperature on Game Day (Degrees F)") +
  ylab("Attendance")
```




Conclusion: There appears to be a weak but slightly positive relationship between average temperature and attendance over the past six seasons.

```
# 2017
# Summary of attendance in seanson 2017 with different opponent.
# Group by different opponents and arrange the attendance from high to low
baseball_opp17 <- baseball_2017 %>% group_by(Opp) %>% summarise(avg_attendance = round(mean(Attendance)
baseball_opp17 <- arrange(baseball_opp17, desc(avg_attendance))

ggplot(baseball_opp17, aes(Opp, avg_attendance)) +
  geom_bar(stat = "identity") +
  ggtitle("Average Attendance vs. Opponents in 2017") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Opponent") +
  ylab("Average Attendance")
```

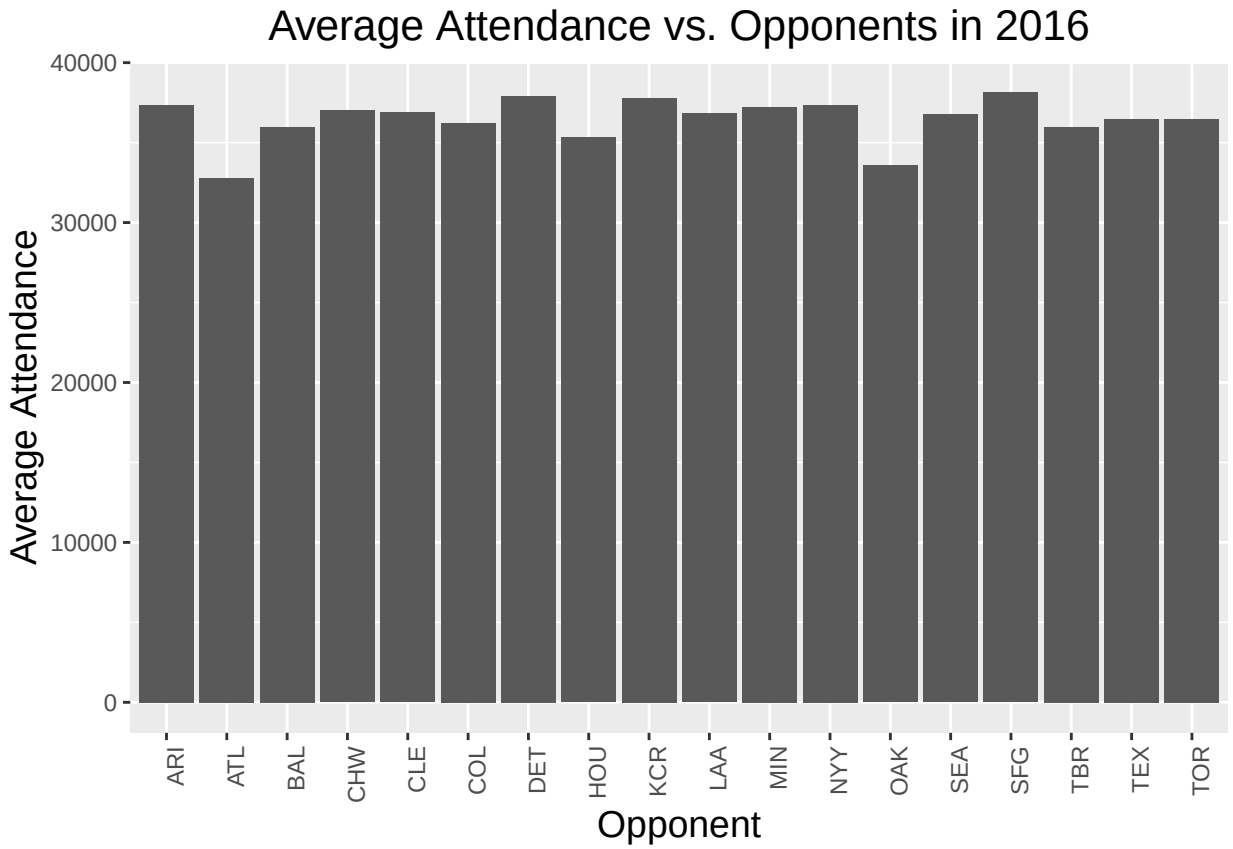



```
kable(baseball_opp17)
```

| Opp | avg_attendance |
|---|---|
| STL | 37263 |
| CLE | 37242 |
| MIN | 37227 |
| CHW | 36984 |
| PHI | 36949 |
| CHC | 36915 |
| KCR | 36904 |
| LAA | 36792 |
| DET | 36491 |
| SEA | 36413 |
| NYY | 36185 |
| TBR | 36142 |
| OAK | 35397 |
| HOU | 35271 |
| BAL | 35147 |
| PIT | 35044 |
| TOR | 34998 |
| TEX | 34781 |

```
# 2016
# Summary of attendance in seanson 2016 with different opponent.
# Group by different opponents and arrange the attendance from high to low
baseball_opp16 <- baseball_2016 %>% group_by(Opp) %>% summarise(avg_attendance = round(mean(Attendance)
```

```
baseball_opp16 <- arrange(baseball_opp16, desc(avg_attendance))

ggplot(baseball_opp16, aes(Opp, avg_attendance)) +
  geom_bar(stat = "identity") +
  ggtitle("Average Attendance vs. Opponents in 2016") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Opponent") +
  ylab("Average Attendance")
```

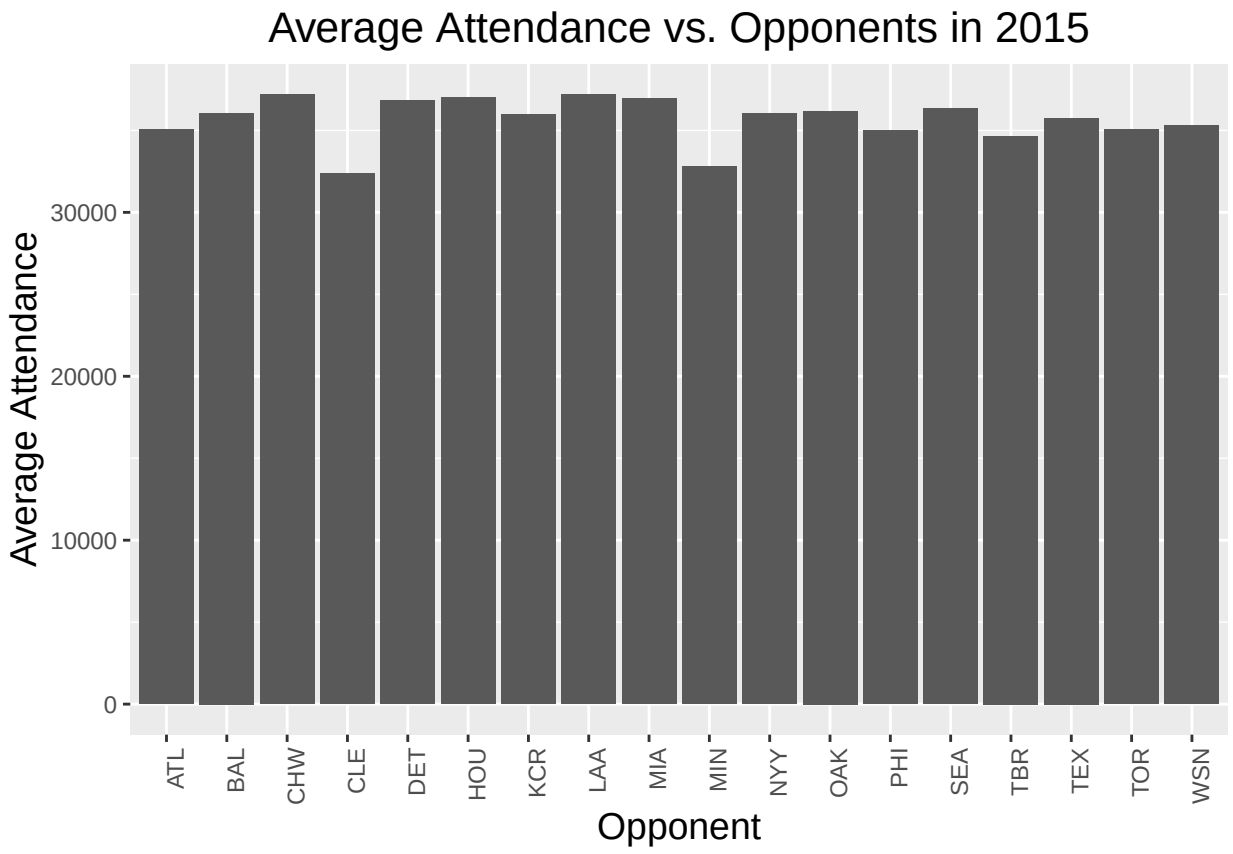



```
kable(baseball_opp16)
```

| Opp | avg_attendance |
|---|---|
| SFG | 38142 |
| DET | 37900 |
| KCR | 37801 |
| NYY | 37359 |
| ARI | 37350 |
| MIN | 37243 |
| CHW | 37010 |
| CLE | 36876 |
| LAA | 36823 |
| SEA | 36767 |
| TEX | 36464 |

| Opp | avg_attendance |
|---|---|
| TOR | 36462 |
| COL | 36238 |
| BAL | 35983 |
| TBR | 35981 |
| HOU | 35324 |
| OAK | 33559 |
| ATL | 32806 |

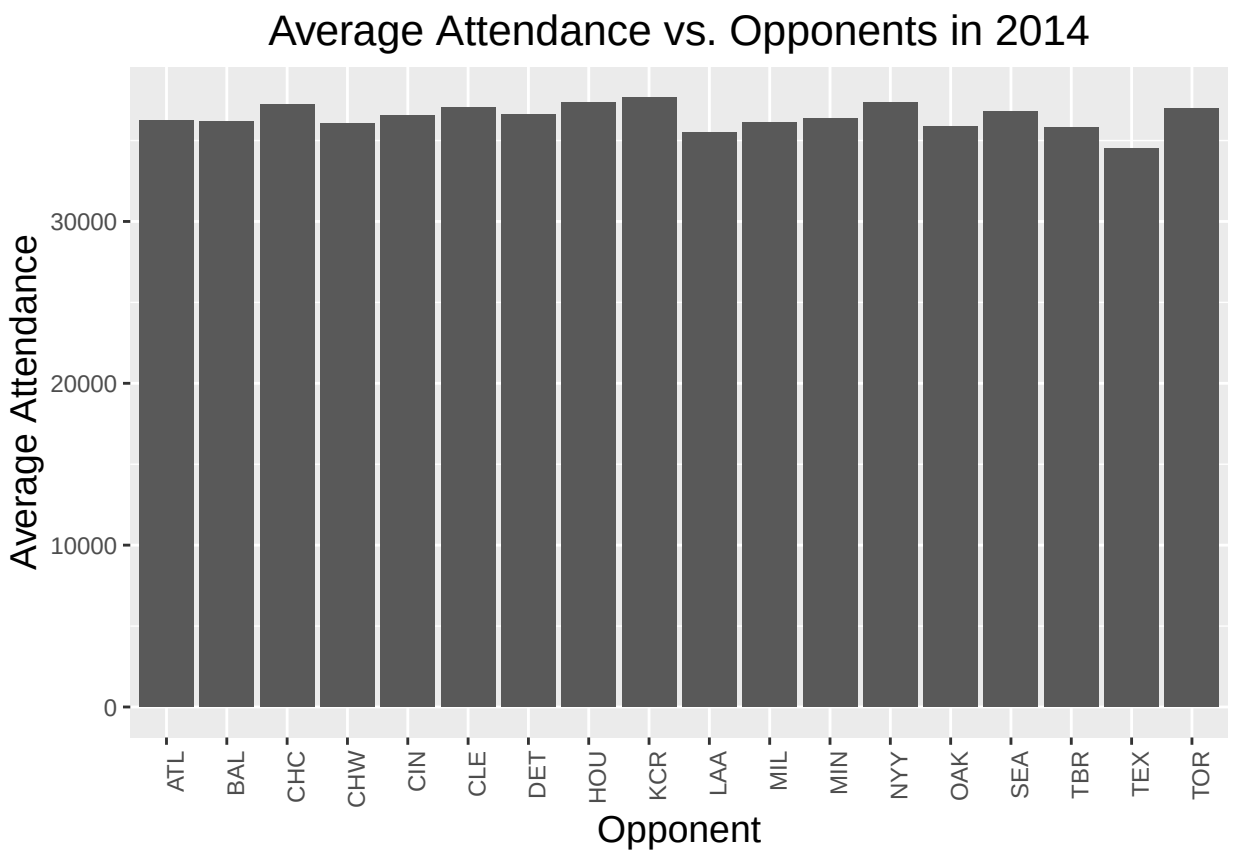
```
# 2015
# Summary of attendance in seanson 2016 with different opponent.
# Group by different opponents and arrange the attendance from high to low
baseball_opp15 <- baseball_2015 %>% group_by(Opp) %>% summarise(avg_attendance = round(mean(Attendance)
baseball_opp15 <- arrange(baseball_opp15, desc(avg_attendance))

ggplot(baseball_opp15, aes(Opp, avg_attendance)) +
  geom_bar(stat = "identity") +
  ggtitle("Average Attendance vs. Opponents in 2015") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Opponent") +
  ylab("Average Attendance")
```



```r
kable(baseball_opp15)
```

| Opp | avg_attendance |
|---|---|
| LAA | 37209 |
| CHW | 37196 |
| HOU | 37007 |
| MIA | 36936 |
| DET | 36829 |
| SEA | 36322 |
| OAK | 36179 |
| BAL | 36060 |
| NYY | 36023 |
| KCR | 35987 |
| TEX | 35750 |
| WSN | 35318 |
| TOR | 35051 |
| ATL | 35050 |
| PHI | 34972 |
| TBR | 34666 |
| MIN | 32808 |
| CLE | 32358 |

```r
# 2014
# Summary of attendance in seanson 2016 with different opponent.
# Group by different opponents and arrange the attendance from high to low
baseball_opp14 <- baseball_2014 %>% group_by(Opp) %>% summarise(avg_attendance = round(mean(Attendance)
baseball_opp14 <- arrange(baseball_opp14, desc(avg_attendance))

ggplot(baseball_opp14, aes(Opp, avg_attendance)) +
  geom_bar(stat = "identity") +
  ggtitle("Average Attendance vs. Opponents in 2014") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Opponent") +
  ylab("Average Attendance")
```

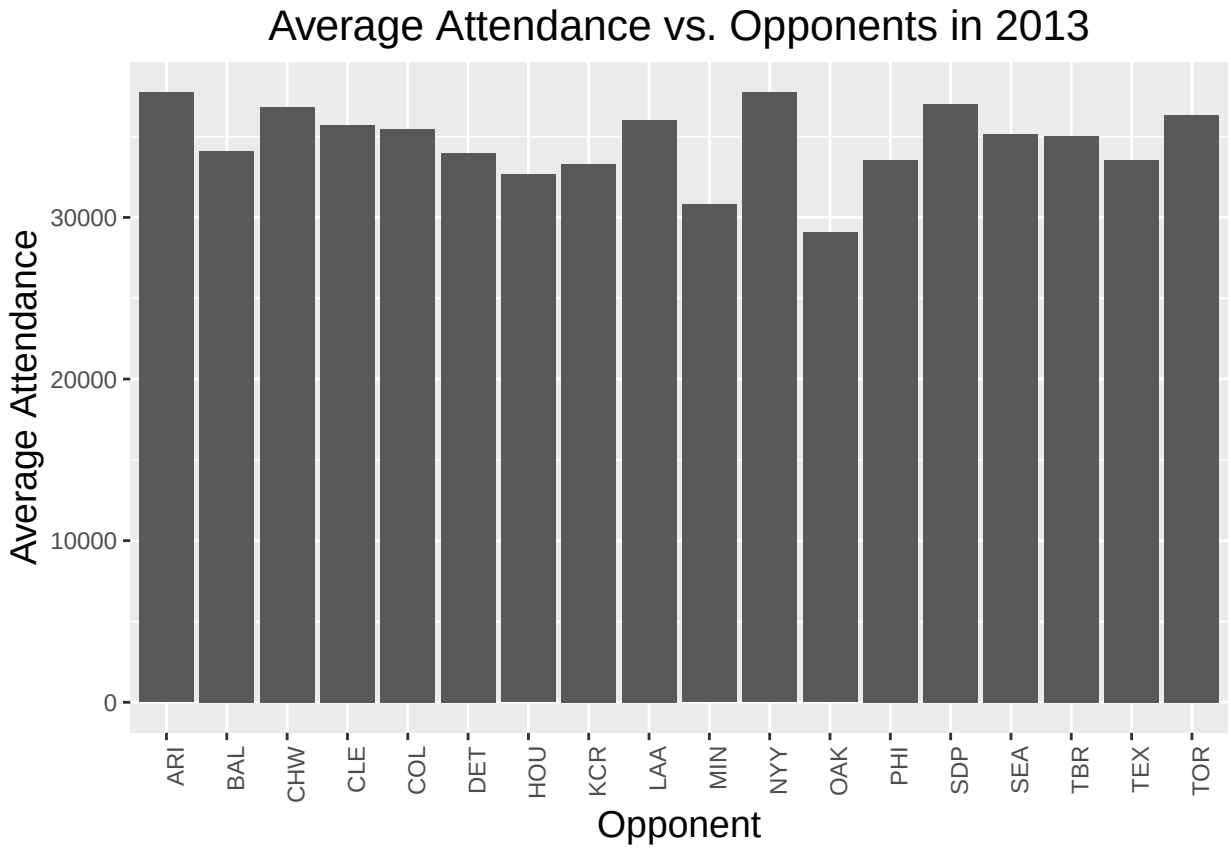



```
kable(baseball_opp14)
```

| Opp | avg_attendance |
|---|---|
| KCR | 37687 |
| HOU | 37362 |
| NYY | 37351 |
| CHC | 37206 |
| CLE | 37015 |
| TOR | 37011 |
| SEA | 36787 |
| DET | 36613 |
| CIN | 36538 |
| MIN | 36339 |
| ATL | 36240 |
| BAL | 36183 |
| MIL | 36138 |
| CHW | 36056 |
| OAK | 35847 |
| TBR | 35818 |
| LAA | 35484 |
| TEX | 34523 |

```
# 2013
# Summary of attendance in seanson 2013 with different opponent.
# Group by different opponents and arrange the attendance from high to low
baseball_opp13 <- baseball_2013 %>% group_by(Opp) %>% summarise(avg_attendance = round(mean(Attendance)
```

```
baseball_opp13 <- arrange(baseball_opp13, desc(avg_attendance))

ggplot(baseball_opp13, aes(Opp, avg_attendance)) +
  geom_bar(stat = "identity") +
  ggtitle("Average Attendance vs. Opponents in 2013")+
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Opponent") +
  ylab("Average Attendance")
```

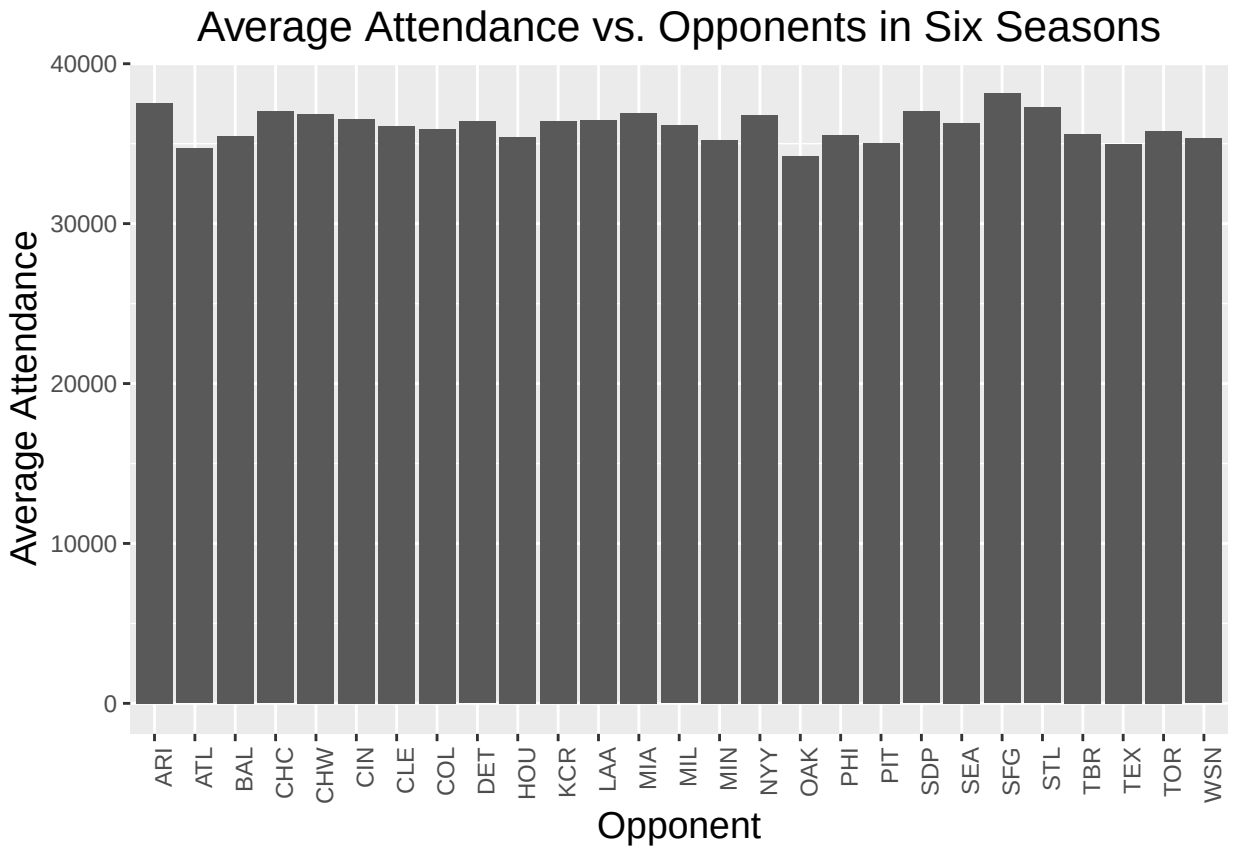



```
kable(baseball_opp13)
```

| Opp | avg_attendance |
|---|---|
| NYY | 37737 |
| ARI | 37735 |
| SDP | 37005 |
| CHW | 36826 |
| TOR | 36348 |
| LAA | 36024 |
| CLE | 35720 |
| COL | 35459 |
| SEA | 35174 |
| TBR | 35036 |
| BAL | 34119 |

| Opp | avg_attendance |
|---|---|
| DET | 33993 |
| TEX | 33561 |
| PHI | 33545 |
| KCR | 33302 |
| HOU | 32664 |
| MIN | 30794 |
| OAK | 29069 |

```
# Summary of attendance in 6 seansons with different opponent.
# Group by different opponents and arrange the attendance from high to low
baseball_opp <- baseball %>% group_by(Opp) %>% summarise(avg_attendance = round(mean(Attendance),digits
baseball_opp <- arrange(baseball_opp, desc(avg_attendance))

ggplot(baseball_opp, aes(Opp, avg_attendance)) +
  geom_bar(stat = "identity") + ggtitle("Average Attendance vs. Opponents in Six Seasons") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Opponent") +
  ylab("Average Attendance")
```



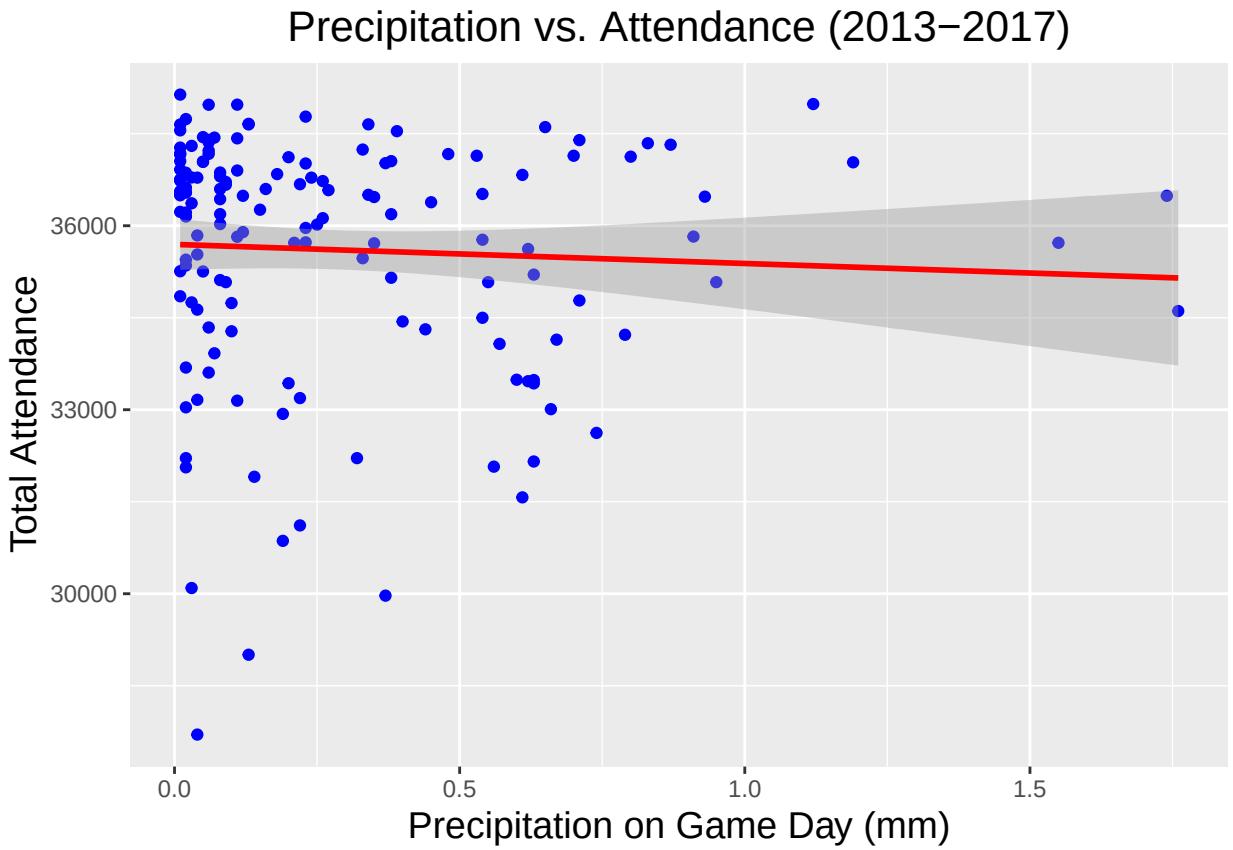

```
kable(baseball_opp)
```

| Opp | avg_attendance |
|---|---|
| SFG | 38142 |
| ARI | 37542 |
| STL | 37263 |
| CHC | 37012 |
| SDP | 37005 |
| MIA | 36936 |
| CHW | 36843 |
| NYY | 36817 |
| CIN | 36538 |
| LAA | 36466 |
| KCR | 36407 |
| DET | 36386 |
| SEA | 36313 |
| MIL | 36138 |
| CLE | 36104 |
| COL | 35927 |
| TOR | 35804 |
| TBR | 35622 |
| PHI | 35534 |
| BAL | 35464 |
| HOU | 35417 |
| WSN | 35318 |
| MIN | 35227 |
| PIT | 35044 |
| TEX | 34977 |
| ATL | 34699 |
| OAK | 34241 |

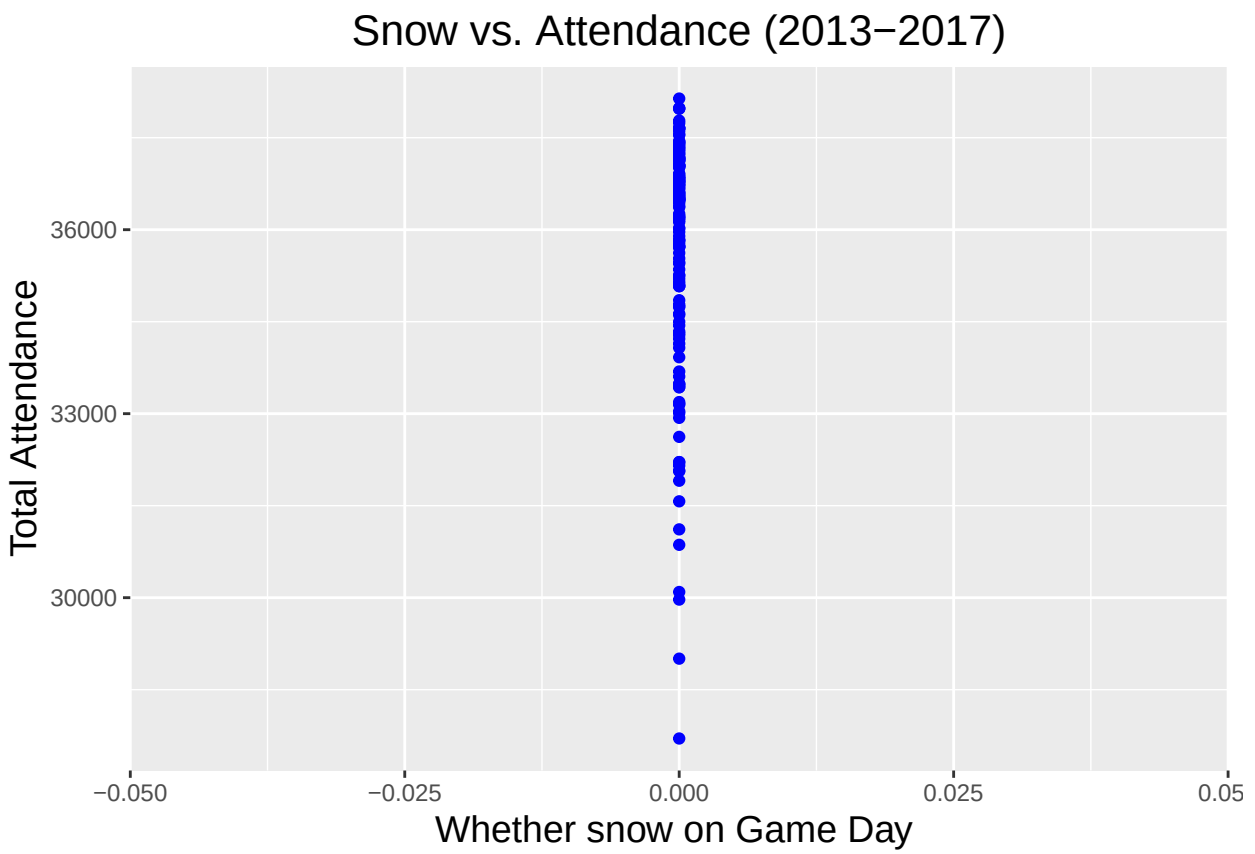
Conclusion: Based on the bar plots, different opponents obviously affect the average attendance of the baseball game.

```
# How the rain/snowy/windy days will affect the attendance of Red Sox, since base game is hypaethral.
baseball_SpecialWeather<- baseball11 %>% select(Gm.,Year,DATE, Tm, Time, Attendance,SNOW,PRCP )%>% filt

# The rain influences on Red Sox attendence
plot1<-ggplot(baseball_SpecialWeather, aes(PRCP, Attendance)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", color = "red") +
  ggtitle("Precipitation vs. Attendance (2013-2017)") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Precipitation on Game Day (mm)") +
  ylab("Total Attendance")

# The snow influences on Red Sox attendence
plot2<-ggplot(baseball_SpecialWeather, aes(SNOW, Attendance)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", color = "red") +
  ggtitle("Snow vs. Attendance (2013-2017)") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
```

```
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Whether snow on Game Day ") +
  ylab("Total Attendance")

plot1
```



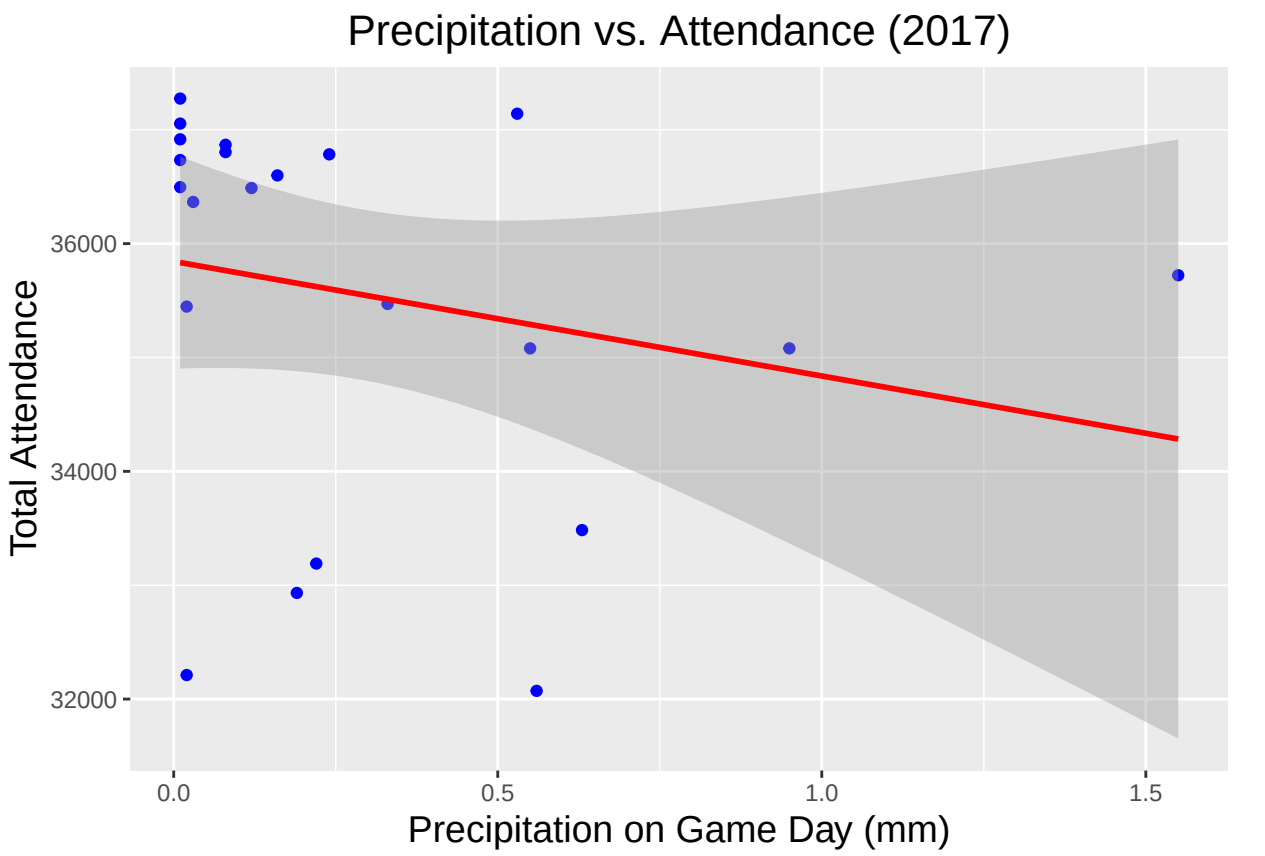

```
plot2
```



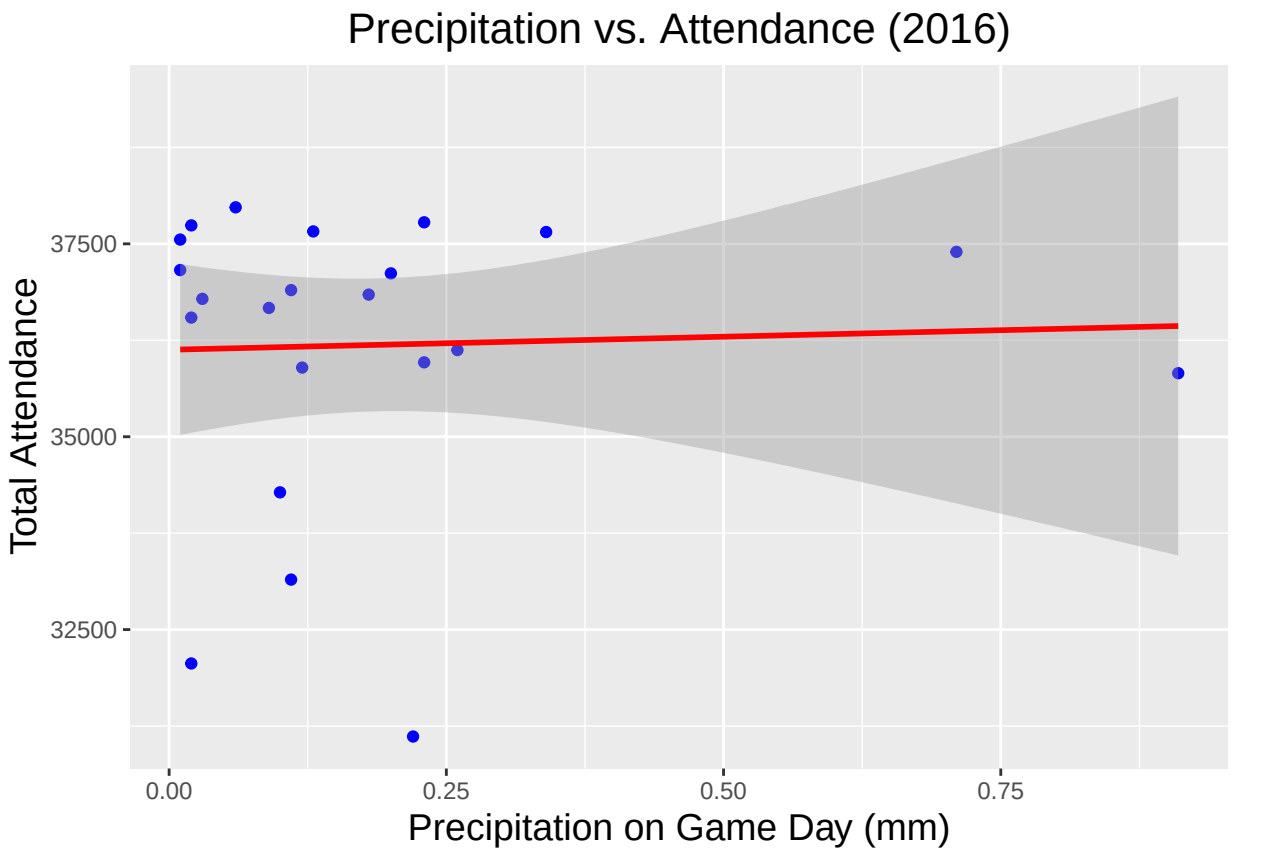

Conclusion: Rain has little influcences on Red Sox attendence. But snow does not have obvious influcences on attendence ,because most baseball games occur before the winter.

```
baseball_SpecialWeather17 <- baseball_SpecialWeather %>% filter(Year == 2017)
baseball_SpecialWeather16 <- baseball_SpecialWeather %>% filter(Year == 2016)
baseball_SpecialWeather15 <- baseball_SpecialWeather %>% filter(Year == 2015)
baseball_SpecialWeather14 <- baseball_SpecialWeather %>% filter(Year == 2014)
baseball_SpecialWeather13 <- baseball_SpecialWeather %>% filter(Year == 2013)

# Rain influences on Red Sox attendence in 2017:
plot17<-ggplot(baseball_SpecialWeather17, aes(PRCP, Attendance)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", color = "red") +
  ggtitle("Precipitation vs. Attendance (2017)") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Precipitation on Game Day (mm)") +
  ylab("Total Attendance")

# The rain influences on Red Sox attendence in 2016:
plot16<-ggplot(baseball_SpecialWeather16, aes(PRCP, Attendance)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", color = "red") +
  ggtitle("Precipitation vs. Attendance (2016)") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
```

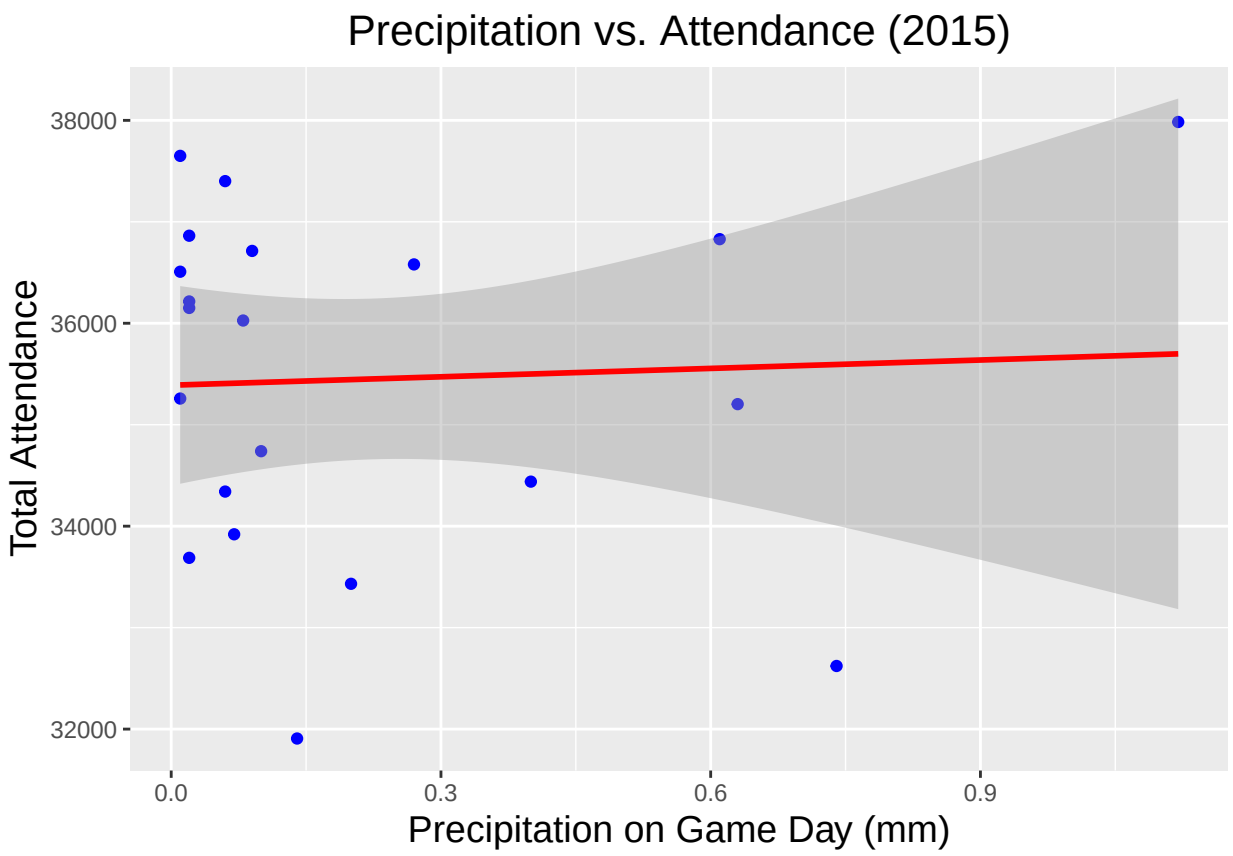
```
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Precipitation on Game Day (mm)") +
  ylab("Total Attendance")

# The rain influences on Red Sox attendence in 2015:
plot15<-ggplot(baseball_SpecialWeather15, aes(PRCP, Attendance)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", color = "red") +
  ggtitle("Precipitation vs. Attendance (2015)") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Precipitation on Game Day (mm)") +
  ylab("Total Attendance")

# The rain influences on Red Sox attendence in 2014:
plot14<-ggplot(baseball_SpecialWeather14, aes(PRCP, Attendance)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", color = "red") +
  ggtitle("Precipitation vs. Attendance (2014)") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Precipitation on Game Day (mm)") +
  ylab("Total Attendance")

# The rain influences on Red Sox attendence in 2013:
plot13<-ggplot(baseball_SpecialWeather13, aes(PRCP, Attendance)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", color = "red") +
  ggtitle("Precipitation vs. Attendance (2013)") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Precipitation on Game Day (mm)") +
  ylab("Total Attendance")

plot17
```


Precipitation vs. Attendance (2017)
Total Attendance
36000
34000
32000
0.0
0.5
1.0
1.5
Precipitation on Game Day (mm)

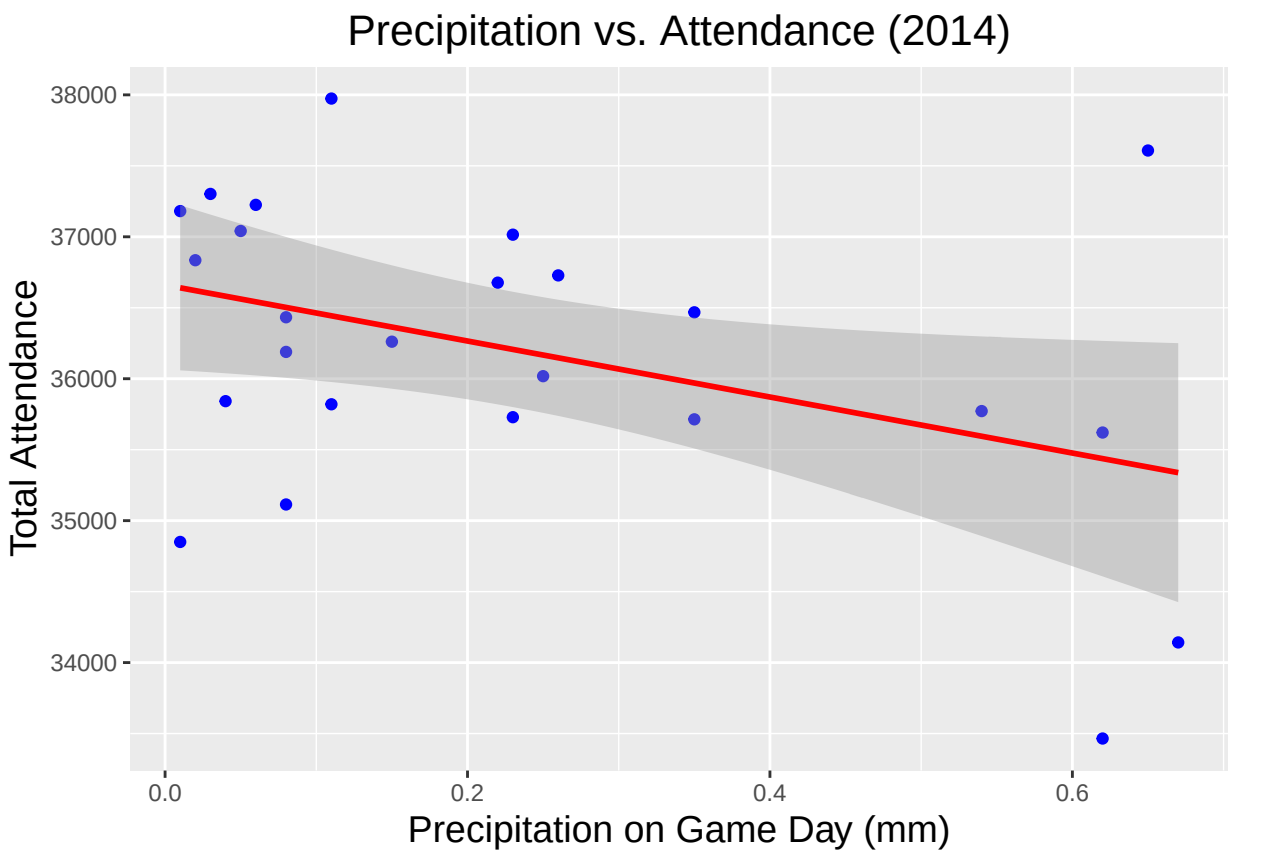
```
plot16
```


Precipitation vs. Attendance (2016)
Total Attendance
37500
35000
32500
0.00
0.25
0.50
0.75
Precipitation on Game Day (mm)

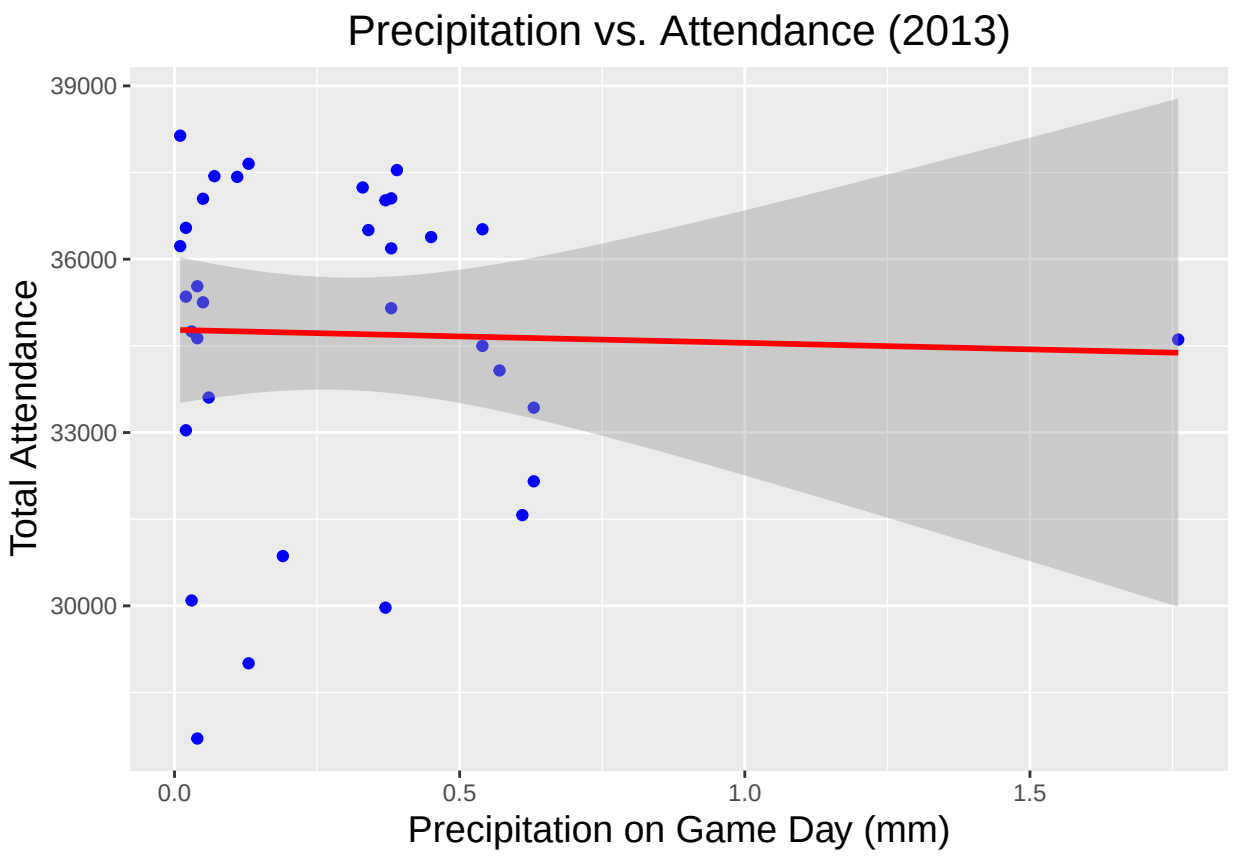
```
plot15
```


Precipitation vs. Attendance (2015)
Total Attendance
38000
36000
34000
32000
0.0
0.3
0.6
0.9
Precipitation on Game Day (mm)

```
plot14
```


Precipitation vs. Attendance (2014)
Total Attendance
38000
37000
36000
35000
34000
0.0
0.2
0.4
0.6
Precipitation on Game Day (mm)

```
plot13
```



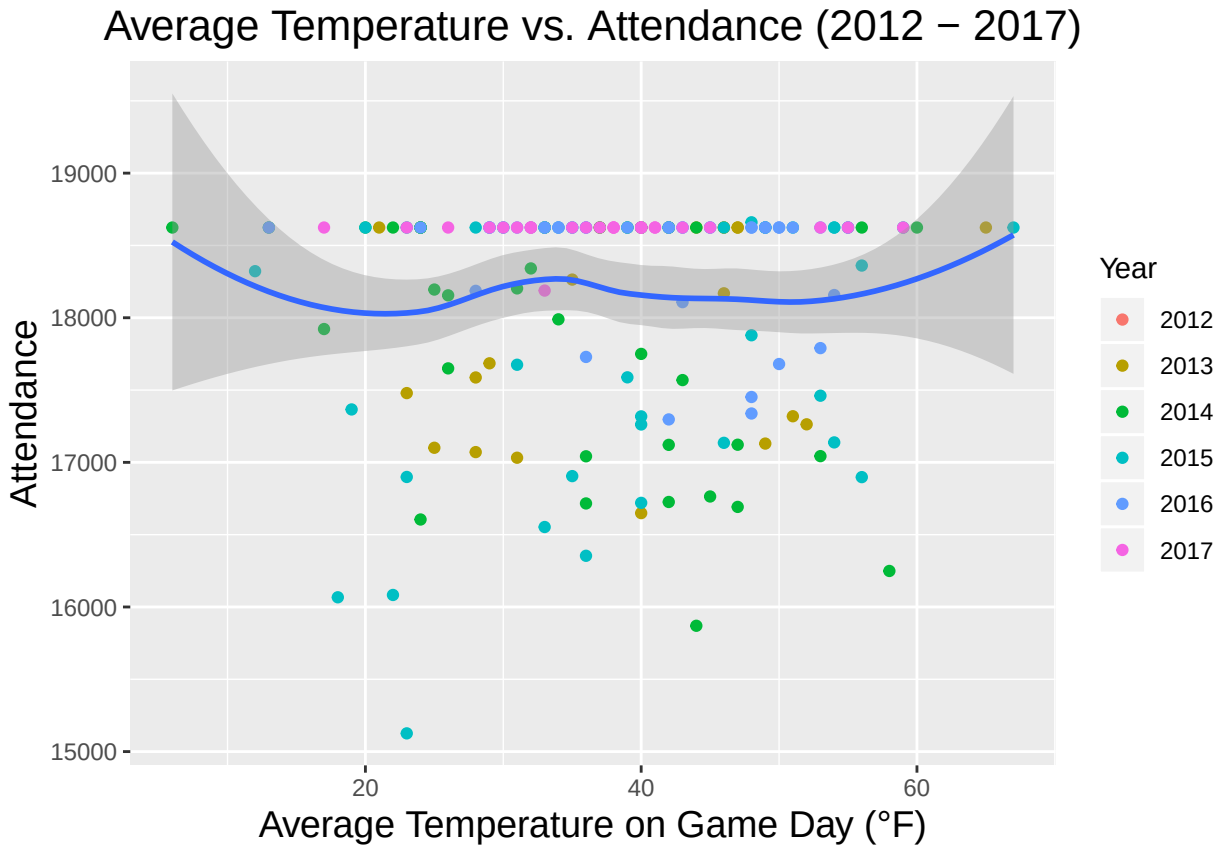

### Basketball

```
bball12 <- read.csv("attend2012.csv")
bball13 <- read.csv("attend2013.csv")
bball14 <- read.csv("attend2014.csv")
bball15 <- read.csv("attend2015.csv")
bball16 <- read.csv("attend2016.csv")

bball <- rbind(bball12, bball13, bball14, bball15, bball16)
bball$Year <- factor(bball$Year)
colnames(bball)[3] <- "Opp"
bball$Opp <- factor(bball$Opp)
write.csv(bball, file = "bball.csv")
#View(bball)
```

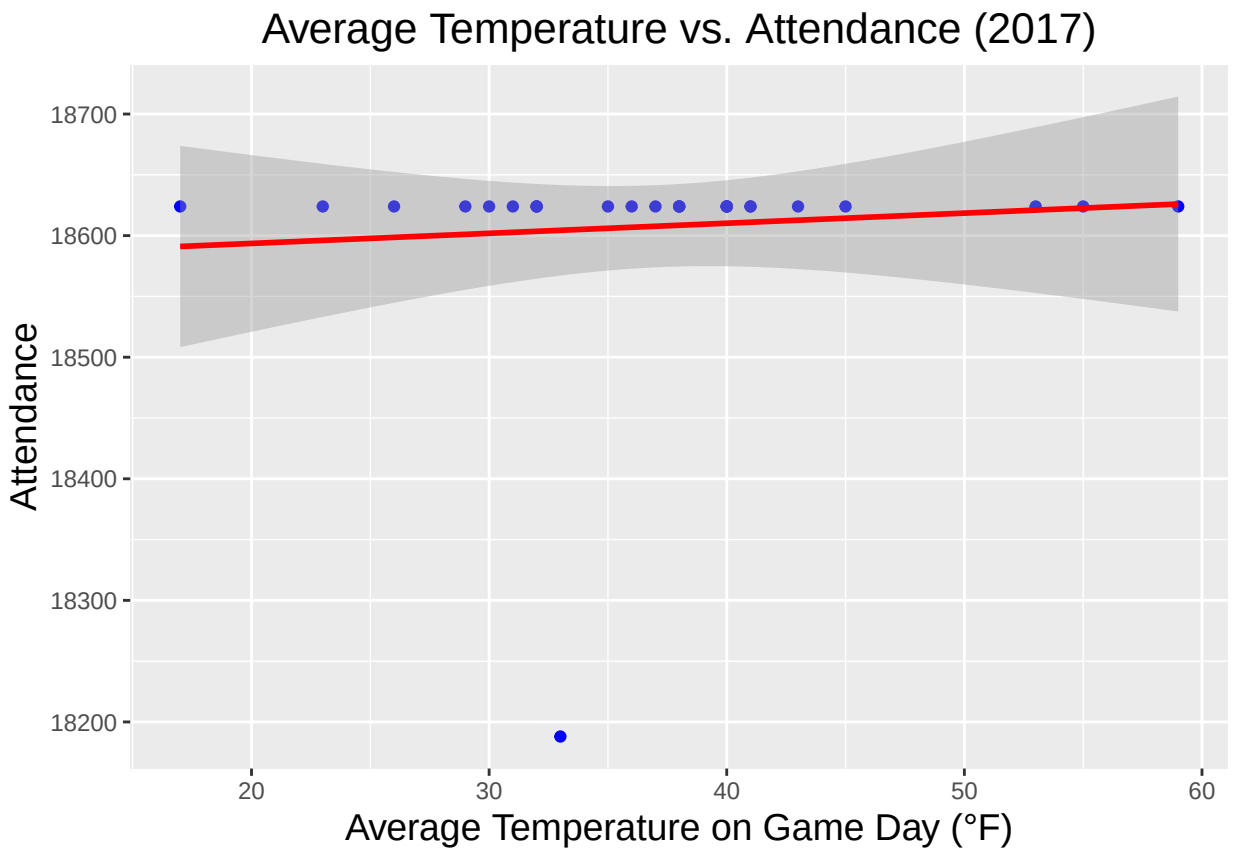
```
ggplot(bball, mapping = aes(x = TAVG, y = Attendance)) +
  geom_point(mapping = aes(color = Year)) +
  geom_smooth() +
  ggtitle("Average Temperature vs. Attendance (2012 - 2017)") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Average Temperature on Game Day (°F)") +
  ylab("Attendance")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

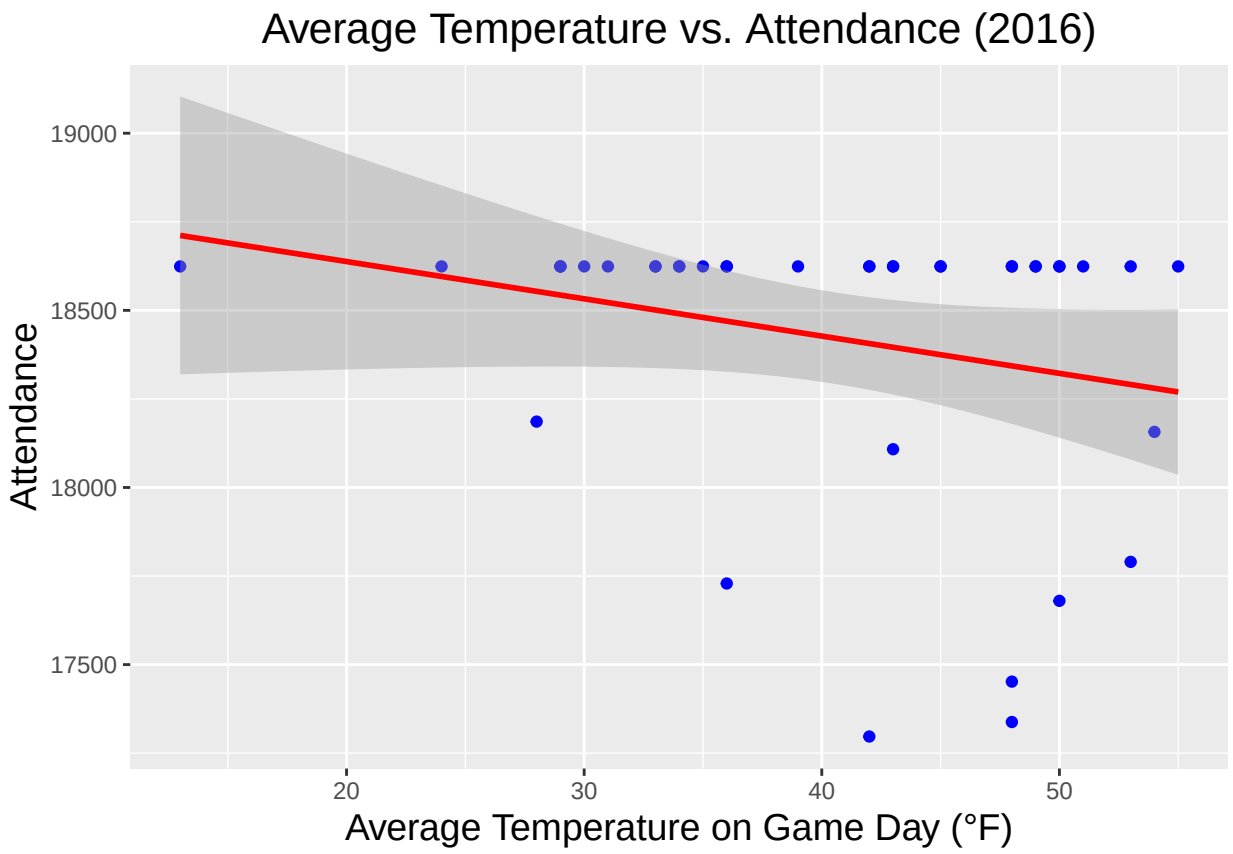
```
## Warning: Removed 36 rows containing non-finite values (stat_smooth).
## Warning: Removed 36 rows containing missing values (geom_point).
```



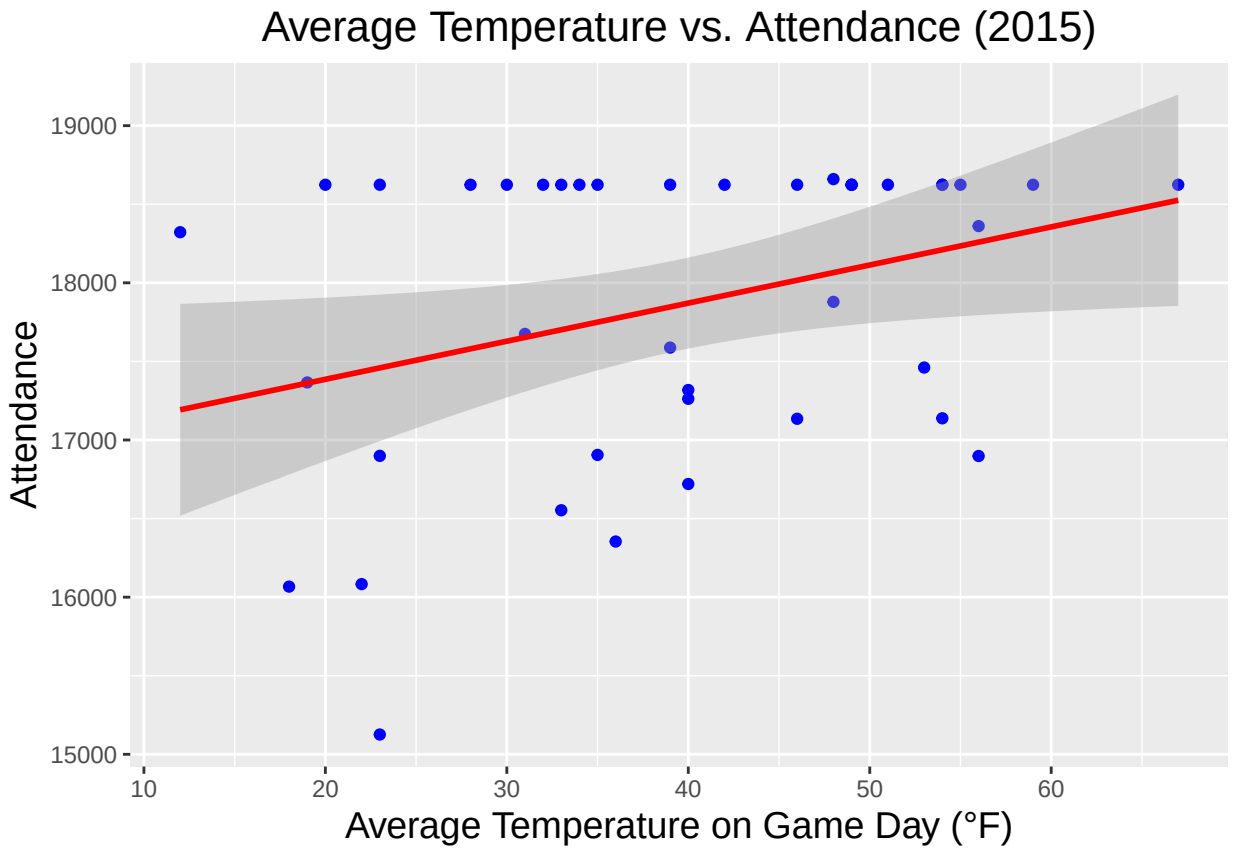

```
# 2017 Season
bball_2017 <- bball %>% filter(Year == 2017)
# Relationship between average temperature with attendance of season 2017
ggplot(bball_2017, aes(TAVG, Attendance)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", color = "red") +
  ggtitle("Average Temperature vs. Attendance (2017)") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Average Temperature on Game Day (°F)") +
  ylab("Attendance")
```



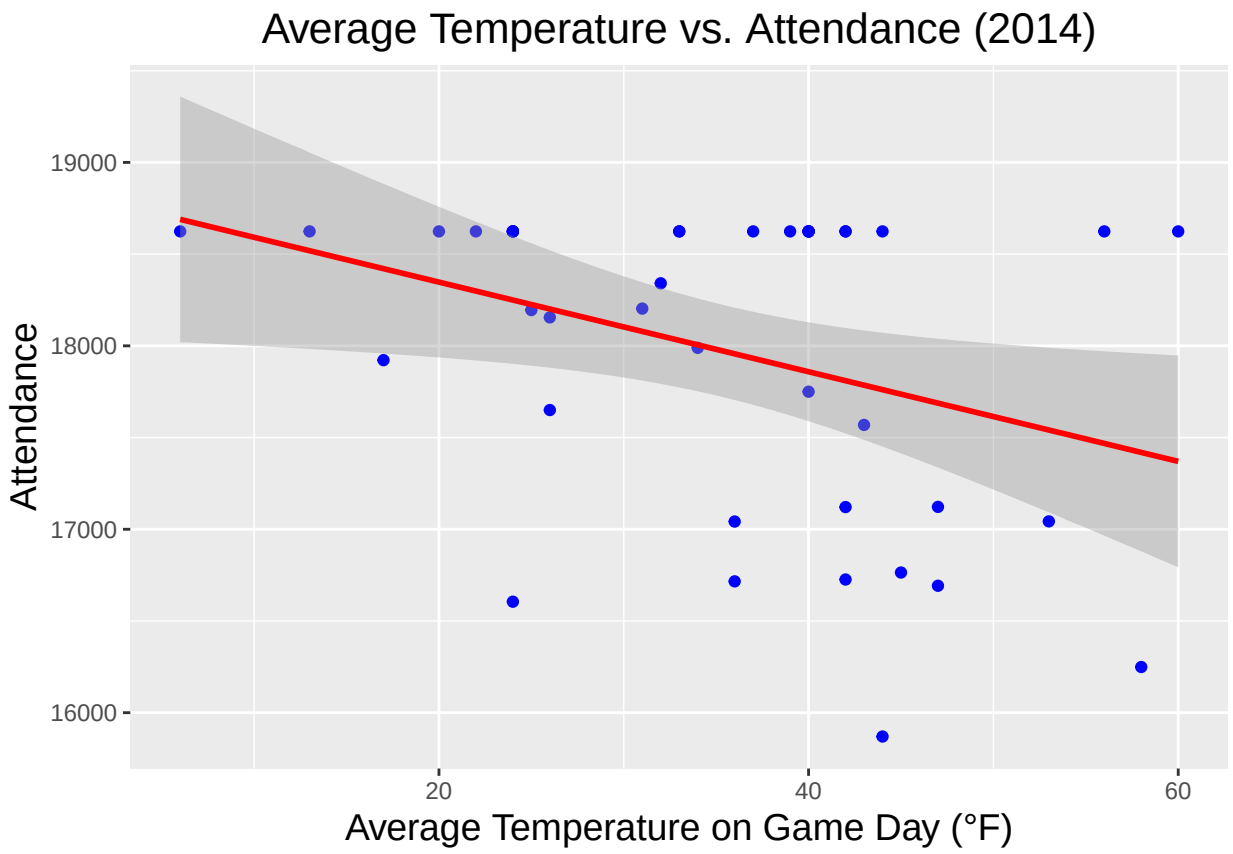

```
# 2016 Season
bball_2016 <- bball %>% filter(Year == 2016)
# Relationship between average temperature with attendance of season 2016
ggplot(bball_2016, aes(TAVG, Attendance)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", color = "red") +
  ggtitle("Average Temperature vs. Attendance (2016)") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Average Temperature on Game Day (°F)") +
  ylab("Attendance")
```



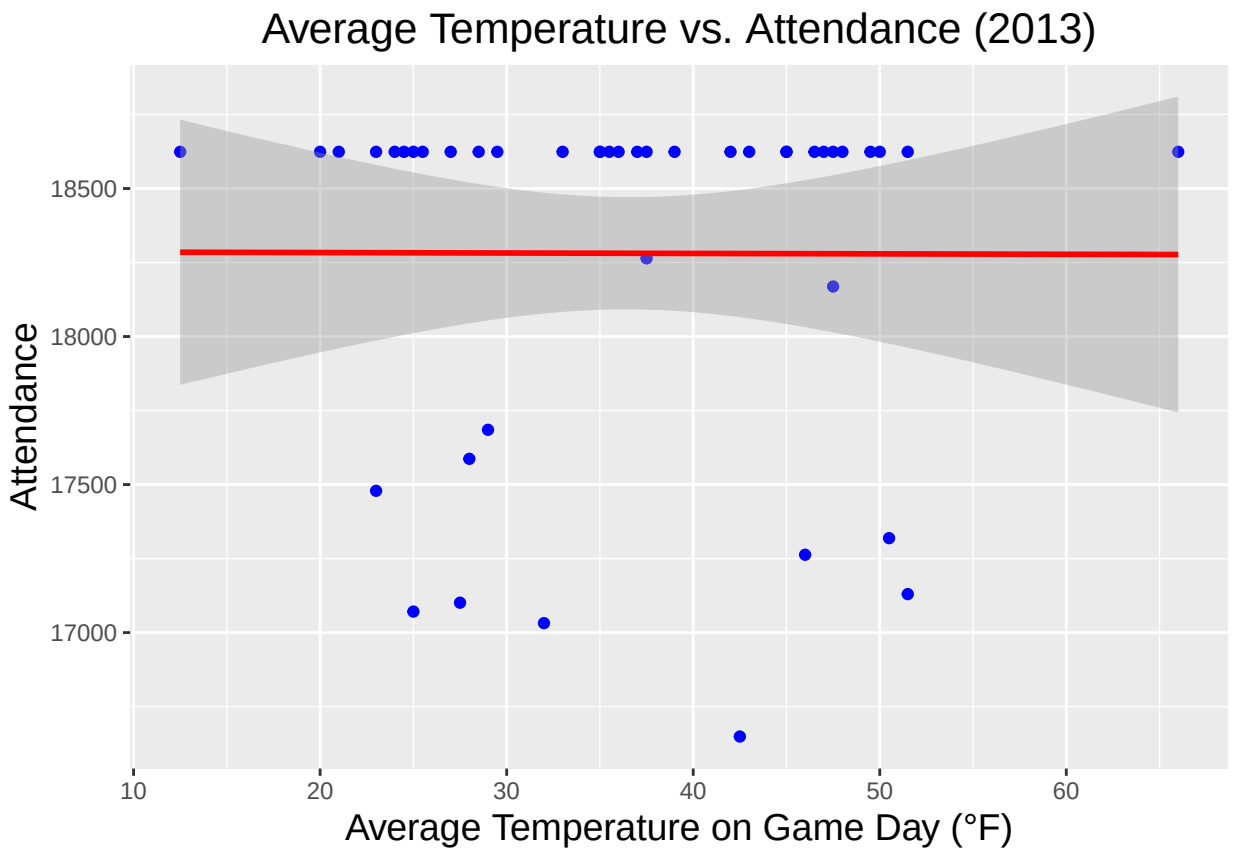

```r
# 2015 Season
bball_2015 <- bball %>% filter(Year == 2015)
# Relationship between average temperature with attendance of season 2015
ggplot(bball_2015, aes(TAVG, Attendance)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", color = "red") +
  ggtitle("Average Temperature vs. Attendance (2015)") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Average Temperature on Game Day (°F)") +
  ylab("Attendance")
```




```r
# 2014 Season
bball_2014 <- bball %>% filter(Year == 2014)
# Relationship between average temperature with attendance of season 2014
ggplot(bball_2014, aes(TAVG, Attendance)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", color = "red") +
  ggtitle("Average Temperature vs. Attendance (2014)") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Average Temperature on Game Day (°F)") +
  ylab("Attendance")
```

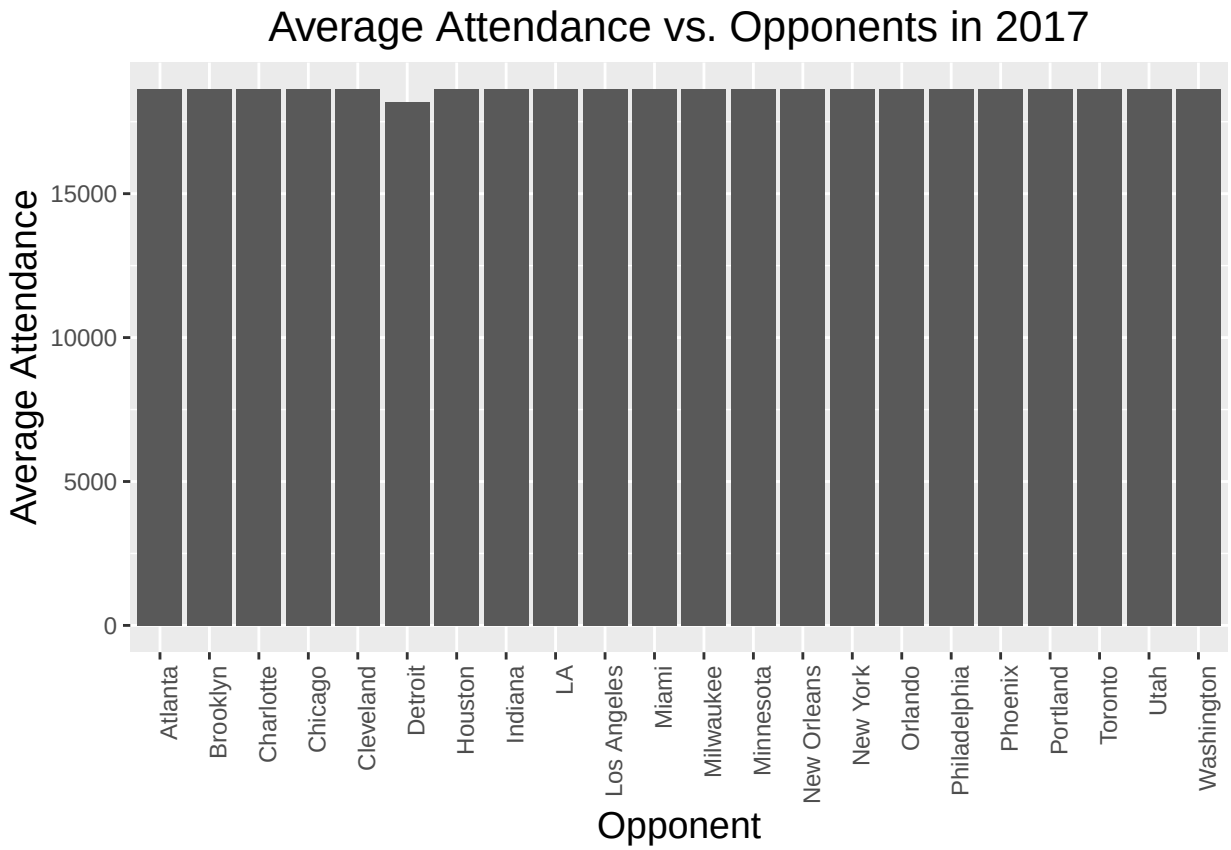



```
# 2013 Season
bball_2013 <- bball %>% filter(Year == 2013)
# Relationship between average temperature with attendance of season 2013

bball_2013$TAVG <- (bball_2013$TMAX + bball_2013$TMIN)/2
ggplot(bball_2013, aes(TAVG, Attendance)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", color = "red") +
  ggtitle("Average Temperature vs. Attendance (2013)") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Average Temperature on Game Day (°F)") +
  ylab("Attendance")
```




Conclusion: There are no evident relationship between the average temperature and the Celtics attendace of these six seasons game.

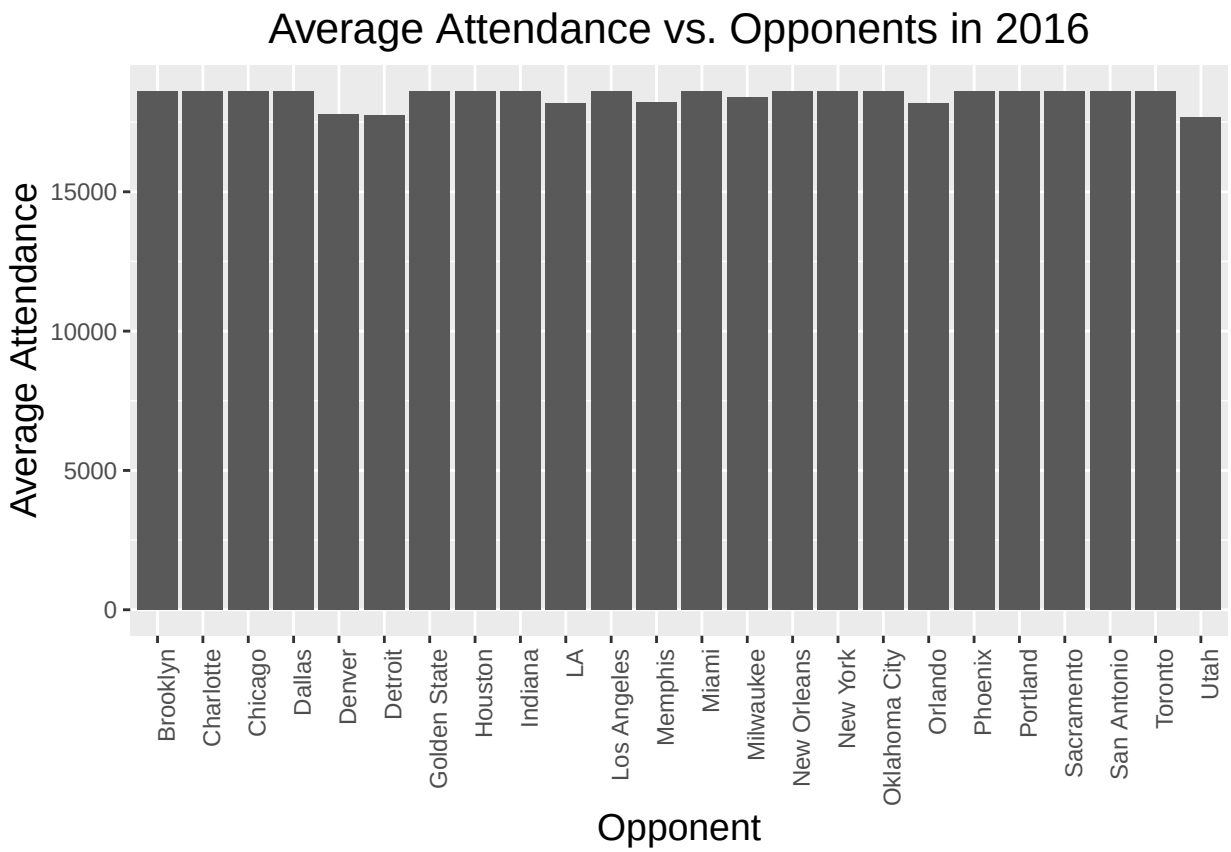
```
# 2017
# Summary of attendance in seanson 2017 with different opponent.
# Group by different opponents and arrange the attendance from high to low
bball_opp17 <- bball_2017 %>% group_by(Opp) %>% summarise(avg_attendance = round(mean(Attendance),,digi
bball_opp17 <- arrange(bball_opp17, desc(avg_attendance))

ggplot(bball_opp17, aes(Opp, avg_attendance)) +
  geom_bar(stat = "identity") +
  ggtitle("Average Attendance vs. Opponents in 2017") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Opponent") +
  ylab("Average Attendance")
```




```
knitr::kable(bball_opp17)
```

| Opp | avg_attendance |
|---|---|
| Atlanta | 18624 |
| Brooklyn | 18624 |
| Charlotte | 18624 |
| Chicago | 18624 |
| Cleveland | 18624 |
| Houston | 18624 |
| Indiana | 18624 |
| LA | 18624 |
| Los Angeles | 18624 |
| Miami | 18624 |
| Milwaukee | 18624 |
| Minnesota | 18624 |
| New Orleans | 18624 |
| New York | 18624 |
| Orlando | 18624 |
| Philadelphia | 18624 |
| Phoenix | 18624 |
| Portland | 18624 |
| Toronto | 18624 |
| Utah | 18624 |
| Washington | 18624 |
| Detroit | 18188 |

```
# 2016
# Summary of attendance in seanson 2016 with different opponent.
# Group by different opponents and arrange the attendance from high to low
bball_opp16 <- bball_2016 %>% group_by(Opp) %>% summarise(avg_attendance = round(mean(Attendance),digits
bball_opp16 <- arrange(bball_opp16, desc(avg_attendance))

ggplot(bball_opp16, aes(Opp, avg_attendance)) +
  geom_bar(stat = "identity") +
  ggtitle("Average Attendance vs. Opponents in 2016") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Opponent") +
  ylab("Average Attendance")
```

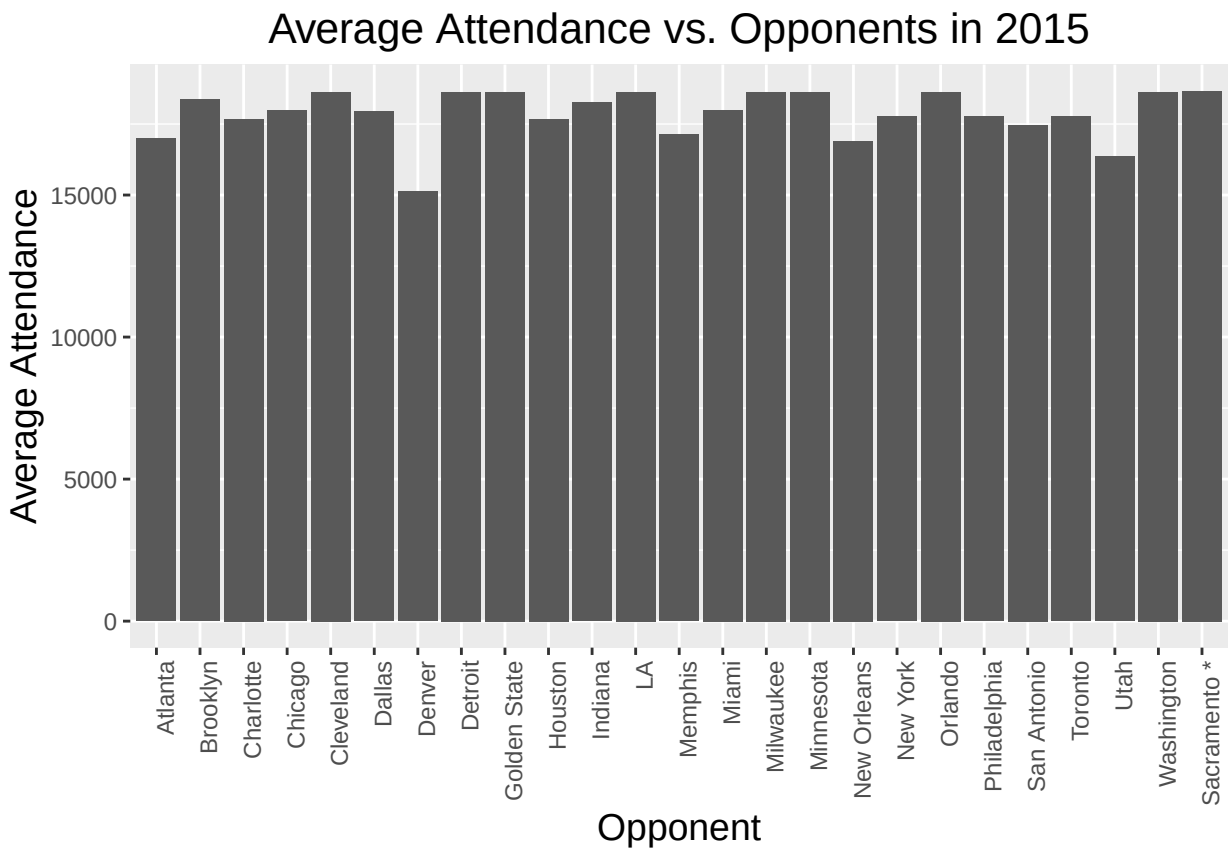



```
knitr::kable(bball_opp16)
```

| Opp | avg_attendance |
|---|---|
| Brooklyn | 18624 |
| Charlotte | 18624 |
| Chicago | 18624 |
| Dallas | 18624 |
| Golden State | 18624 |
| Houston | 18624 |
| Indiana | 18624 |

| Opp | avg_attendance |
|---|---|
| Los Angeles | 18624 |
| Miami | 18624 |
| New Orleans | 18624 |
| New York | 18624 |
| Oklahoma City | 18624 |
| Phoenix | 18624 |
| Portland | 18624 |
| Sacramento | 18624 |
| San Antonio | 18624 |
| Toronto | 18624 |
| Milwaukee | 18390 |
| Memphis | 18207 |
| LA | 18186 |
| Orlando | 18176 |
| Denver | 17780 |
| Detroit | 17753 |
| Utah | 17680 |

```
# 2015
# Summary of attendance in seanson 2016 with different opponent.
# Group by different opponents and arrange the attendance from high to low
bball_opp15 <- bball_2015 %>% group_by(Opp) %>% summarise(avg_attendance = round(mean(Attendance),digits
bball_opp15 <- arrange(bball_opp15, desc(avg_attendance))

ggplot(bball_opp15, aes(Opp, avg_attendance)) +
  geom_bar(stat = "identity") +
  ggtitle("Average Attendance vs. Opponents in 2015") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Opponent") +
  ylab("Average Attendance")
```

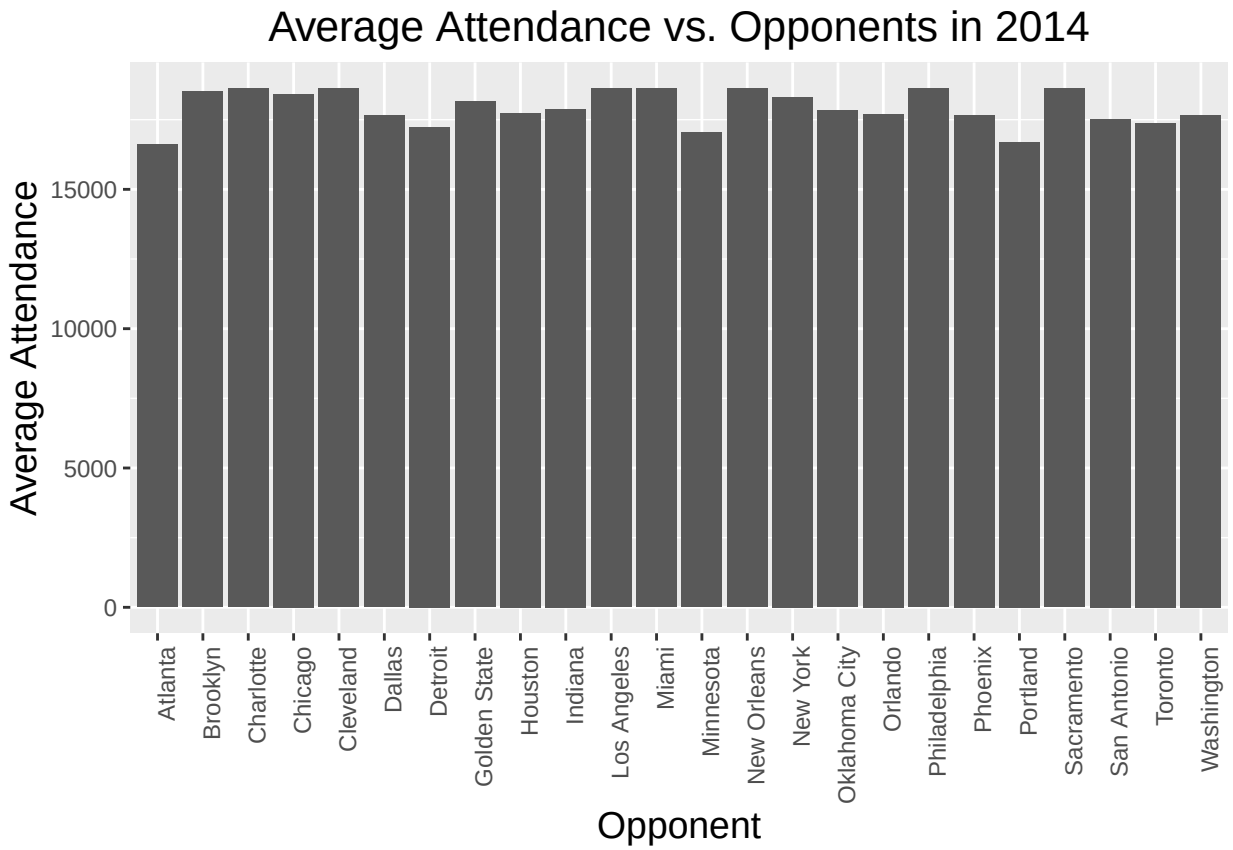



```
knitr::kable(bball_opp15)
```

| Opp | avg_attendance |
|---|---|
| Sacramento * | 18660 |
| Cleveland | 18624 |
| Detroit | 18624 |
| Golden State | 18624 |
| LA | 18624 |
| Milwaukee | 18624 |
| Minnesota | 18624 |
| Orlando | 18624 |
| Washington | 18624 |
| Brooklyn | 18361 |
| Indiana | 18252 |
| Miami | 17995 |
| Chicago | 17971 |
| Dallas | 17943 |
| Philadelphia | 17772 |
| New York | 17762 |
| Toronto | 17761 |
| Houston | 17675 |
| Charlotte | 17672 |
| San Antonio | 17461 |
| Memphis | 17135 |
| Atlanta | 16978 |
| New Orleans | 16905 |
| Utah | 16354 |

| Opp | avg_attendance |
|---|---|
| Denver | 15126 |

```
# 2014
# Summary of attendance in seanson 2016 with different opponent.
# Group by different opponents and arrange the attendance from high to low
bball_opp14 <- bball_2014 %>% group_by(Opp) %>% summarise(avg_attendance = round(mean(Attendance),,digi
bball_opp14 <- arrange(bball_opp14, desc(avg_attendance))

ggplot(bball_opp14, aes(Opp, avg_attendance)) +
  geom_bar(stat = "identity") +
  ggtitle("Average Attendance vs. Opponents in 2014") +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Opponent") +
  ylab("Average Attendance")
```

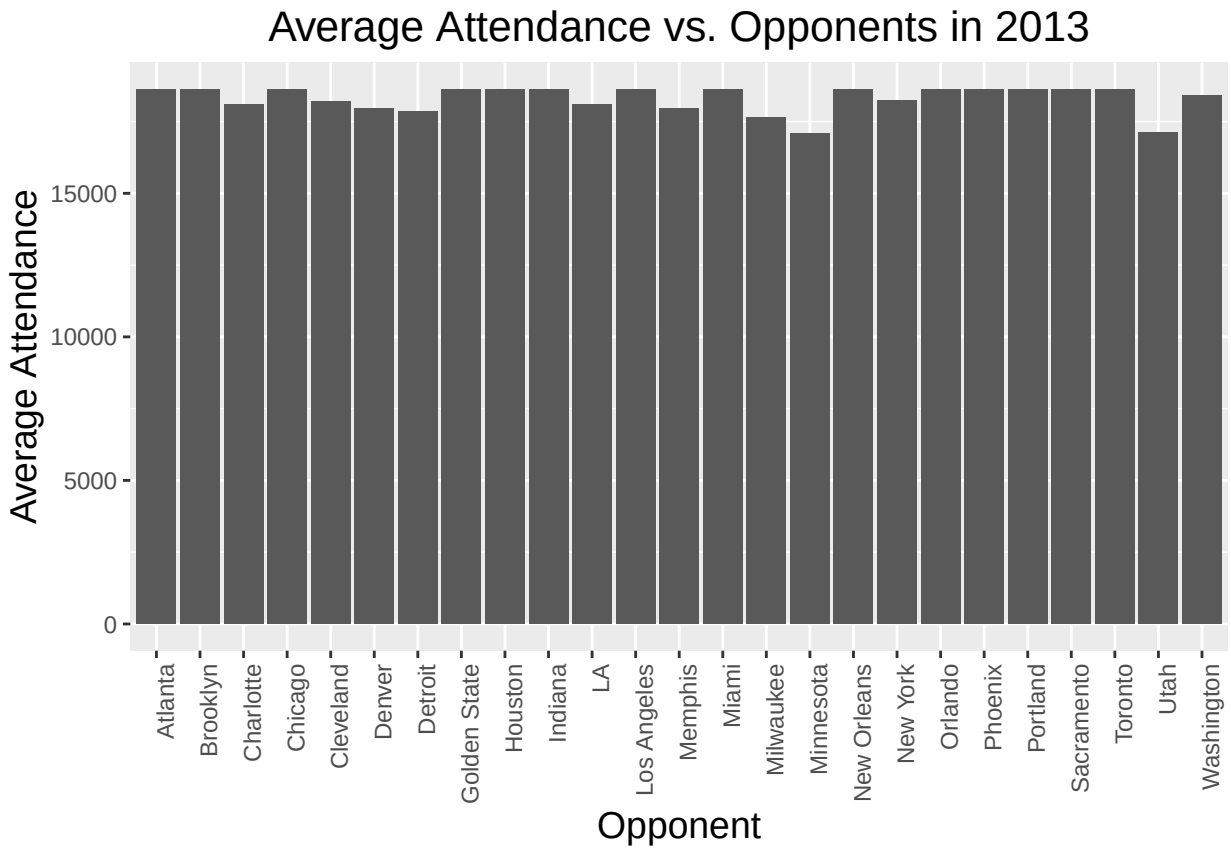



```
knitr::kable(bball_opp14)
```

| Opp | avg_attendance |
|---|---|
| Charlotte | 18624 |
| Cleveland | 18624 |
| Los Angeles | 18624 |
| Miami | 18624 |

| Opp | avg_attendance |
|---|---|
| New Orleans | 18624 |
| Philadelphia | 18624 |
| Sacramento | 18624 |
| Brooklyn | 18517 |
| Chicago | 18414 |
| New York | 18306 |
| Golden State | 18155 |
| Indiana | 17873 |
| Oklahoma City | 17834 |
| Houston | 17750 |
| Orlando | 17694 |
| Phoenix | 17675 |
| Washington | 17670 |
| Dallas | 17650 |
| San Antonio | 17522 |
| Toronto | 17386 |
| Detroit | 17247 |
| Minnesota | 17042 |
| Portland | 16692 |
| Atlanta | 16605 |

```
# 2013
# Summary of attendance in seanson 2013 with different opponent.
# Group by different opponents and arrange the attendance from high to low
bball_opp13 <- bball_2013 %>% group_by(Opp) %>% summarise(avg_attendance = round(mean(Attendance),digits
bball_opp13 <- arrange(bball_opp13, desc(avg_attendance))

ggplot(bball_opp13, aes(Opp, avg_attendance)) +
  geom_bar(stat = "identity") +
  ggtitle("Average Attendance vs. Opponents in 2013")+
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Opponent") +
  ylab("Average Attendance")
```

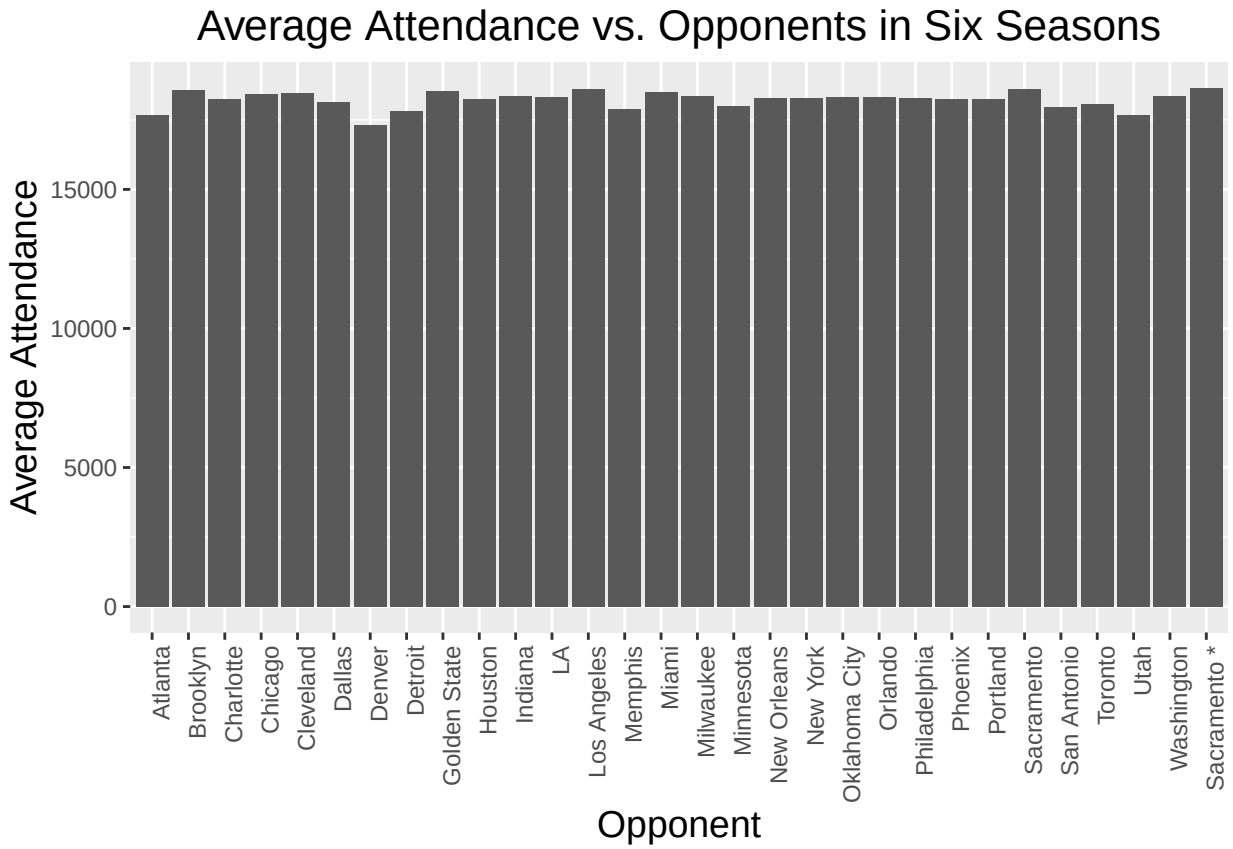



```
knitr::kable(bball_opp13)
```

| Opp | avg_attendance |
|---|---|
| Atlanta | 18624 |
| Brooklyn | 18624 |
| Chicago | 18624 |
| Golden State | 18624 |
| Houston | 18624 |
| Indiana | 18624 |
| Los Angeles | 18624 |
| Miami | 18624 |
| New Orleans | 18624 |
| Orlando | 18624 |
| Phoenix | 18624 |
| Portland | 18624 |
| Sacramento | 18624 |
| Toronto | 18624 |
| Washington | 18396 |
| New York | 18242 |
| Cleveland | 18191 |
| LA | 18106 |
| Charlotte | 18093 |
| Memphis | 17972 |
| Denver | 17944 |
| Detroit | 17862 |
| Milwaukee | 17636 |
| Utah | 17130 |

| Opp | avg_attendance |
|---|---|
| Minnesota | 17071 |

```
# Summary of attendance in 6 seansons with different opponent.
# Group by different opponents and arrange the attendance from high to low
bball_opp <- bball %>% group_by(Opp) %>% summarise(avg_attendance = round(mean(Attendance),digits = 0))
bball_opp <- arrange(bball_opp, desc(avg_attendance))

ggplot(bball_opp, aes(Opp, avg_attendance)) +
  geom_bar(stat = "identity") + ggtitle("Average Attendance vs. Opponents in Six Seasons") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  theme(axis.title.x = element_text(size = 14)) +
  theme(axis.title.y = element_text(size = 14)) +
  theme(plot.title = element_text(hjust = 0.5, size = 16)) +
  xlab("Opponent") +
  ylab("Average Attendance")
```



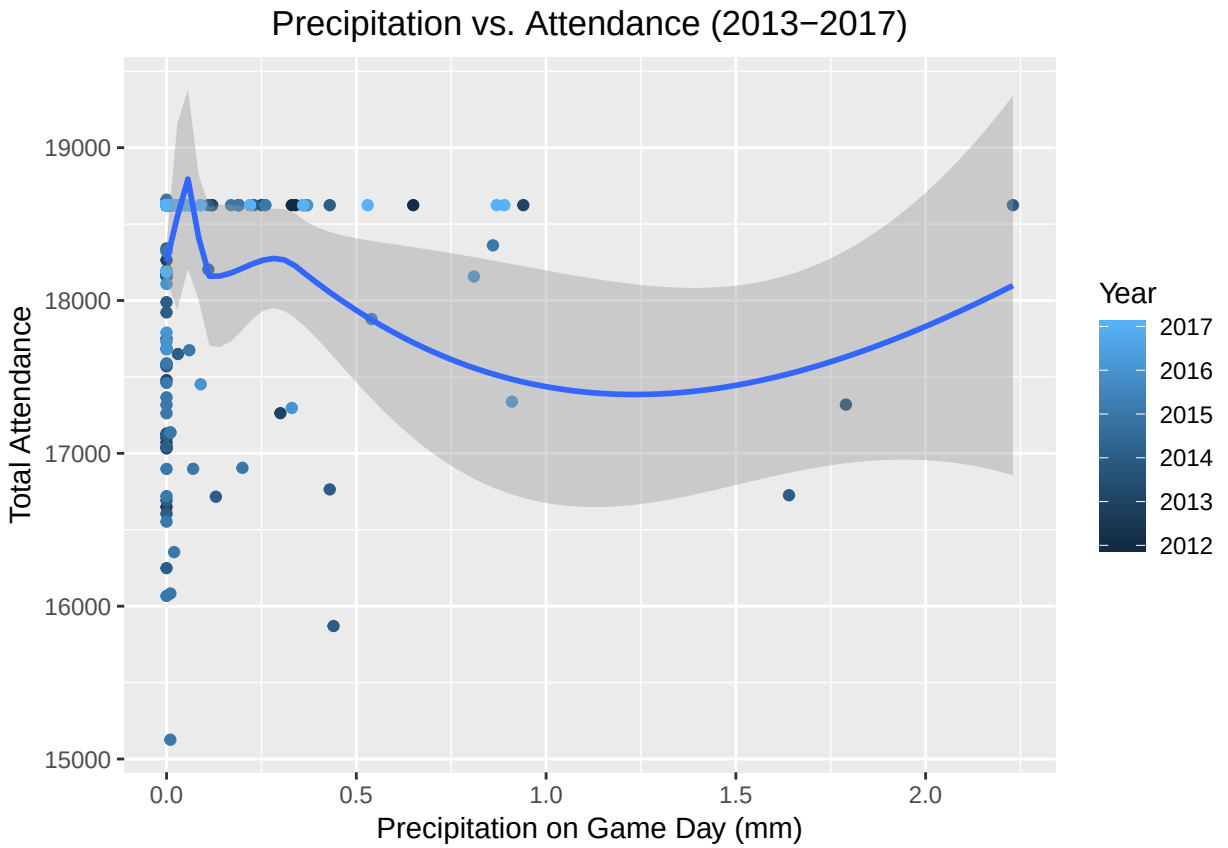

```
knitr::kable(bball_opp)
```

| Opp |
|---|
| Sacramento * |
| Los Angeles |
| Sacramento |
| Brooklyn |
| Golden State |
| Miami |

| Opp |
|---|
| Cleveland |
| Chicago |
| Washington |
| Milwaukee |
| Indiana |
| LA |
| Orlando |
| Oklahoma City |
| Philadelphia |
| New Orleans |
| New York |
| Houston |
| Phoenix |
| Portland |
| Charlotte |
| Dallas |
| Toronto |
| Minnesota |
| San Antonio |
| Memphis |
| Detroit |
| Utah |
| Atlanta |
| Denver |
| Conclusion: Oppo nents have slight but not very significant influences on Celtics Attendance. Some weak opponents, like |

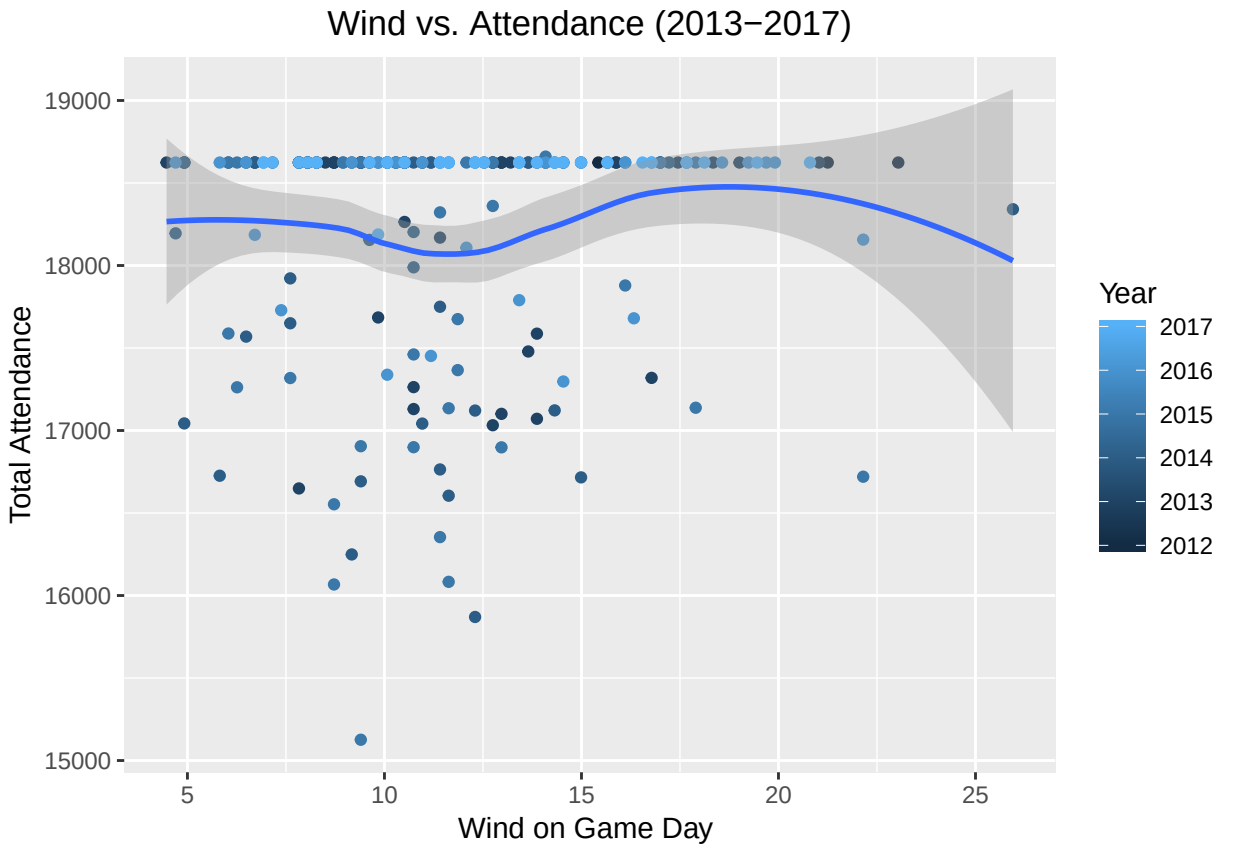
```
basketball_WeatherAll<-read.csv("bball.csv")
basketball_SpecialWeather<- basketball_WeatherAll %>% select(Year,Date, Attendance,AWND,PRCP )

# The snow influences on Celtics attendence
plot1<-ggplot(basketball_SpecialWeather, aes(PRCP, Attendance)) +
  geom_point(mapping=aes(color = Year)) +
  geom_smooth() +
  ggtitle("Precipitation vs. Attendance (2013-2017)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  xlab("Precipitation on Game Day (mm)") +
  ylab("Total Attendance")
# The wind influences on Celtics attendence
plot2<-ggplot(basketball_SpecialWeather, aes(AWND,Attendance)) +
  geom_point(mapping=aes(color = Year)) +
  geom_smooth() +
  ggtitle("Wind vs. Attendance (2013-2017)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  xlab("Wind on Game Day ") +
  ylab("Total Attendance")
plot1
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
plot2
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




Conclusion: Rainy and windy weathers have slight influences on Celtics Attendence in all 6 six seasons. Since NBA games are totally in door. But, during season 2013-2016, the influences were more obvious and significant.