Assignment 1

Brian Yim

2025-09-05

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

# 1. Download dataset from the web

# I downloaded the data from Kaggle. <https://www.kaggle.com/datasets/rashadferguson/college-sports-attendance?select=Attendance.xlsx>

# There were categorial variables: University, Conference, Sport, Season. Also there were quantitative variables: Home\_Games, Total\_Attendance, Average\_Attendance. I read the data as below.

attendance <- read.csv("C:/Users/hyim/OneDrive - Kent State University/Desktop/New folder/attendance.csv")

# Quick peek at data  
head(attendance)

## University Conference Sport Season Home\_Games Total\_Attendance  
## 1 Alabama SEC Men's Basketball 2020-21 15 63750  
## 2 Alabama SEC Men's Basketball 2021-22 17 230860  
## 3 Alabama SEC Men's Basketball 2022-23 18 270000  
## 4 Alabama SEC Men's Basketball 2023-24 17 258400  
## 5 Alabama SEC Women's Basketball 2020-21 15 22500  
## 6 Alabama SEC Women's Basketball 2021-22 16 84800  
## Average\_Attendance  
## 1 4250  
## 2 13580  
## 3 15000  
## 4 15200  
## 5 1500  
## 6 5300

# 2. Descriptive statistics

# Quantitative variable summaries

summary(attendance$Home\_Games)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 13.00 15.00 16.00 17.67 18.50 30.00

summary(attendance$Total\_Attendance)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 13650 34700 58800 84807 88125 306000

summary(attendance$Average\_Attendance)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 800 2000 2800 5009 5525 17000

# Categorical variables

table(attendance$Conference)

##   
## ACC Big 12 Big Ten SEC   
## 48 40 80 88

table(attendance$Sport)

##   
## Men's Basketball Softball Women's Basketball Women's Volleyball   
## 64 64 64 64

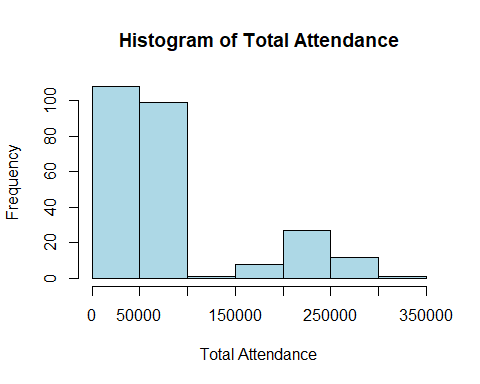
# 3. Transform a variable

# I transformed the attendance per game variable to attendance per game in thousands

attendance$Avg\_Attendance\_K <- attendance$Average\_Attendance / 1000

# 4. Plot quantitative variable

hist(attendance$Total\_Attendance,  
 main = "Histogram of Total Attendance",  
 xlab = "Total Attendance",  
 col = "lightblue")

 # 5. Scatterplot: Home Games vs Total Attendance

plot(attendance$Home\_Games, attendance$Total\_Attendance,  
 xlab = "Home Games",  
 ylab = "Total Attendance",  
 main = "Home Games vs Total Attendance",  
 col = "darkred", pch = 19)

