Lab1

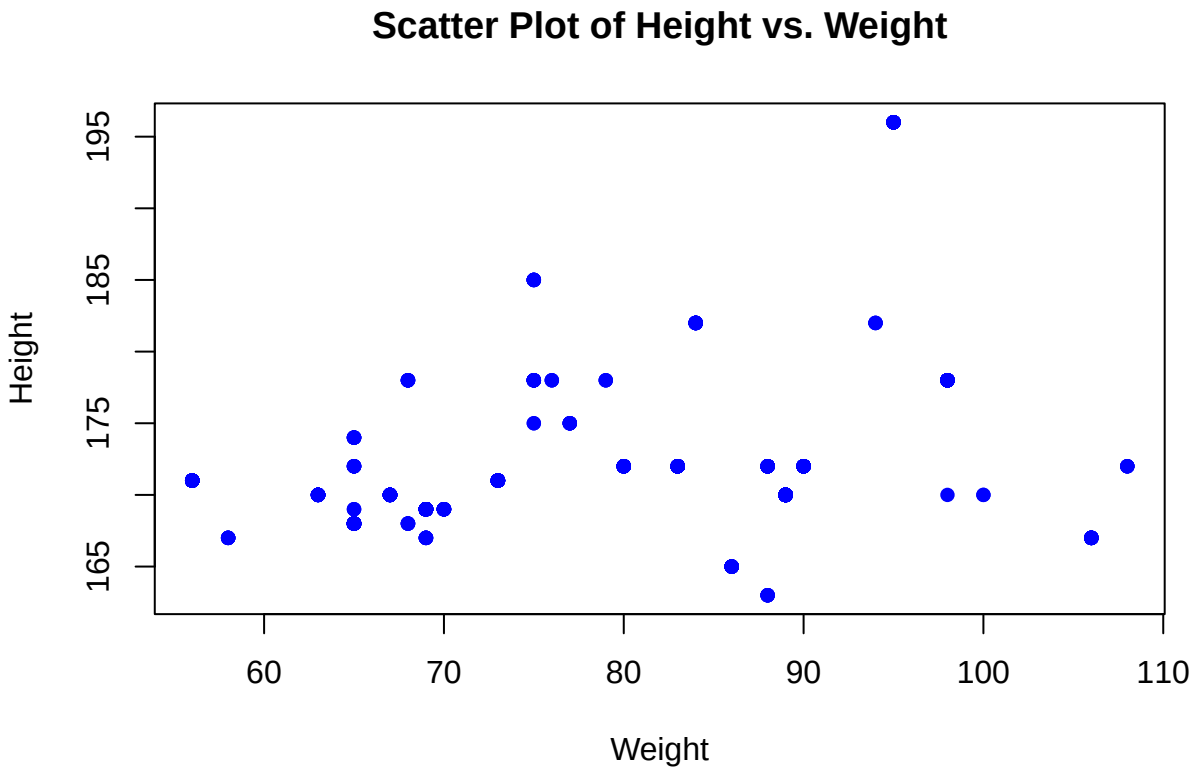
2024-09-08

Bhavya Patel

```
abs = read.csv("~/Documents/Fall\ 24/Dat\ 301/Lab1/Absenteeism_at_work.csv",
               sep=";",header=TRUE)
```

Q1

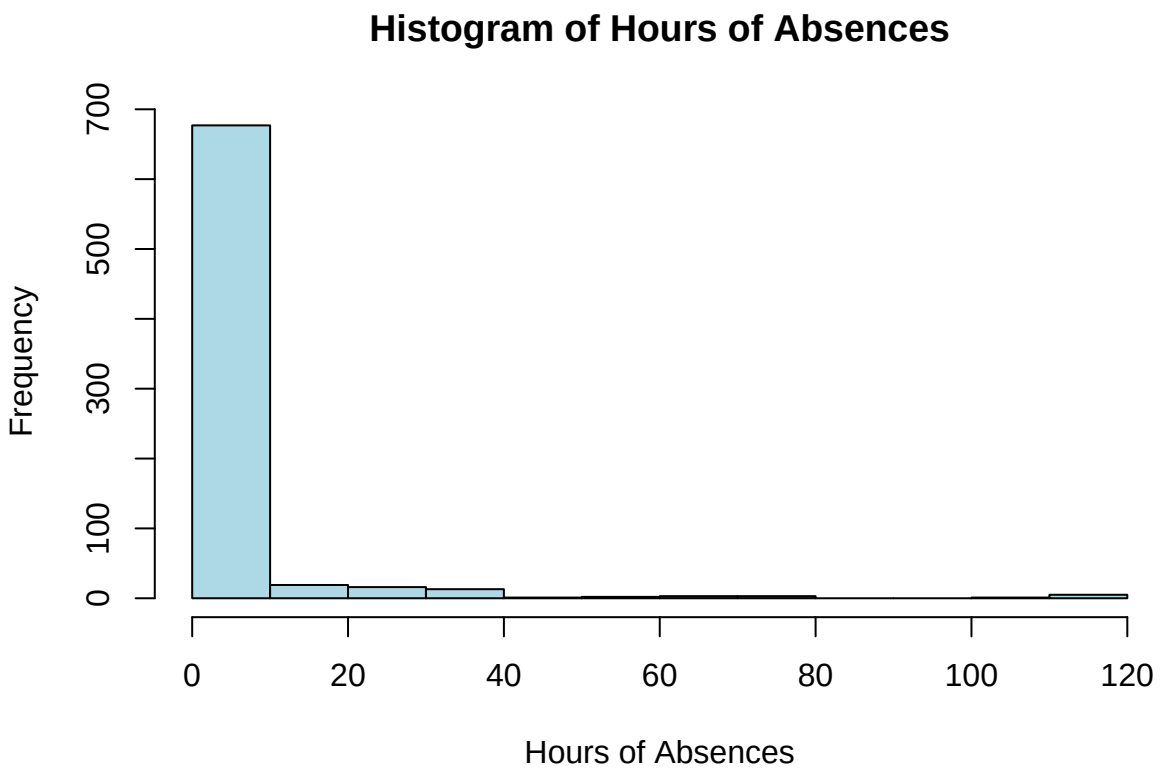
```
plot(x=abs$Weight, y=abs$Height, xlab="Weight", ylab="Height",
     main="Scatter Plot of Height vs. Weight", pch=16, col="blue")
```



From the above scatter plot, based on wider spread of points this indicates larger standard deviation. This means bigger the range of height-to-weight ratios among employees.

Q2

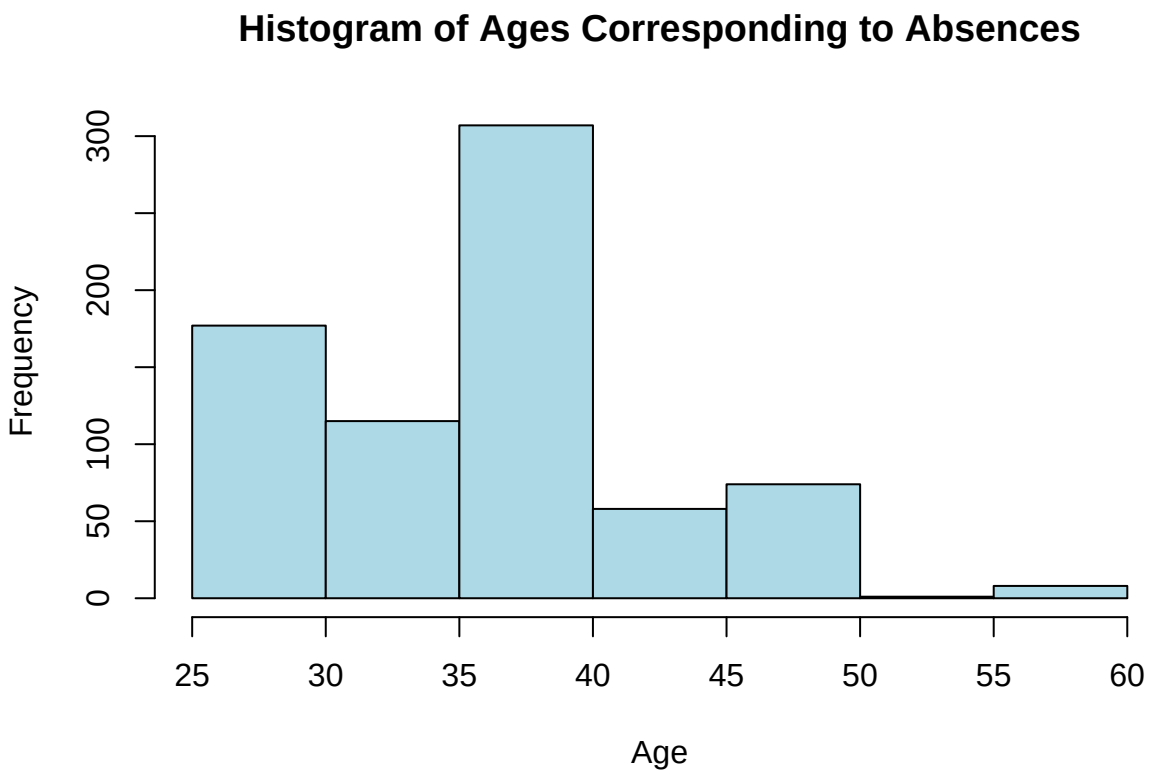
```
hist(abs$Absenteeism.time.in.hours, xlab="Hours of Absences",
     main="Histogram of Hours of Absences",
     border = "black", col="lightblue")
```



This histogram is skewed to the right. Also the most employees have hours of absence between 0 to 20 only. But there a few employees with more than 40 hours of absences.

Q3

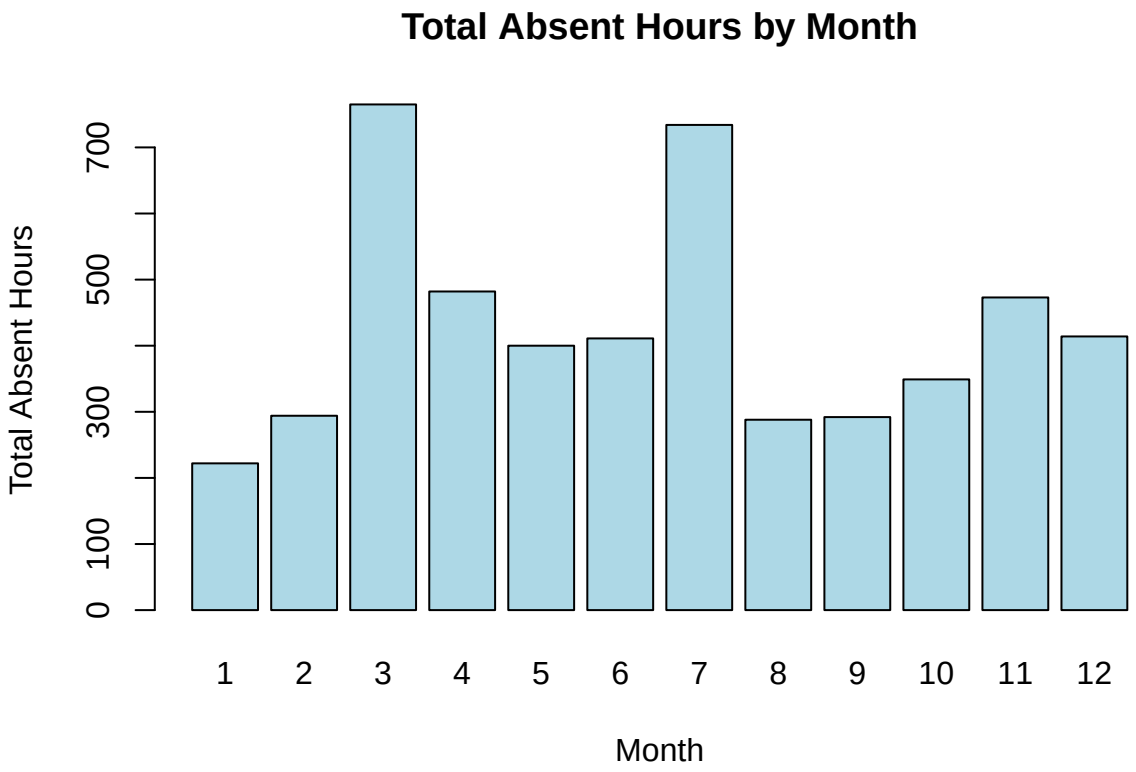
```
hist(abs$Age, main = "Histogram of Ages Corresponding to Absences",
     xlab = "Age", border = "black", col="lightblue")
```



This histogram is skewed to the right. Also this histogram indicates younger employees have more absences. The age group of 35-40 years have the highest frequency of absences followed by the youngest age group of 25-30.

Q4

```
uni <- abs[abs$Month.of.absence!=0,]
total_hours_by_month <- tapply(uni$Absenteeism.time.in.hours, uni$Month.of.absence, sum)
barplot(total_hours_by_month, main = "Total Absent Hours by Month",
        xlab = "Month", ylab = "Total Absent Hours",
        col = "lightblue", border = "black")
```

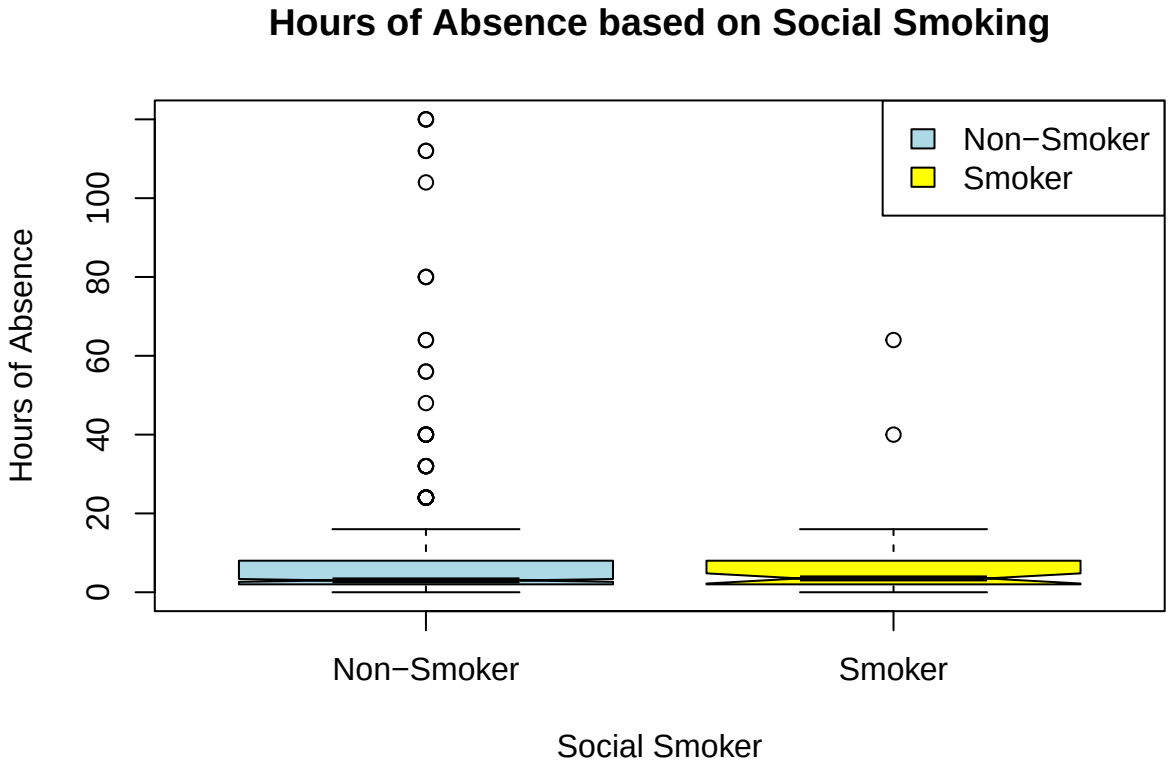


This bar graph represents Absences categorized based on Months. This indicates most absences were in the month of March followed by July.

Q5

```
boxplot(abs$Absenteeism.time.in.hours ~ abs$Social.smoker,
        main = "Hours of Absence based on Social Smoking",
        notch=1, notchwidth=0.2,
        xlab = "Social Smoker",
        ylab = "Hours of Absence",
        col = c("lightblue", "yellow"),
        names = c("Non-Smoker", "Smoker"),
        border = "black")

legend("topright", legend = c("Non-Smoker", "Smoker"),
       fill = c("lightblue", "yellow"), border = "black")
```

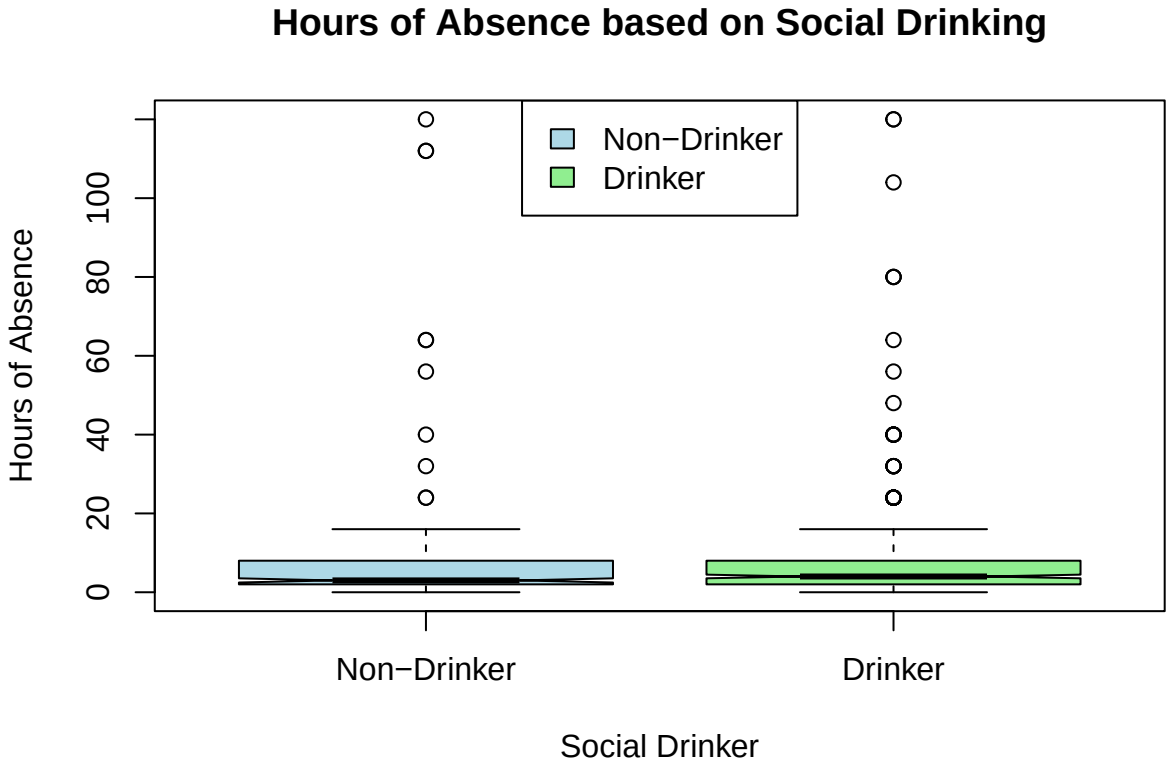


This box plot shows Non-Smokers seem to have more outliers with a higher number of absence hours compared to Smokers, indicating that a few individuals in the Non-Smoker group may take significantly more time off work.

Q6

```
boxplot(abs$Absenteeism.time.in.hours ~ abs$Social.drinker,
        main = "Hours of Absence based on Social Drinking",
        notch=1, notchwidth=0.2,
        xlab = "Social Drinker",
        ylab = "Hours of Absence",
        col = c("lightblue", "lightgreen"),
        names = c("Non-Drinker", "Drinker"),
        border = "black", )

legend("top", legend = c("Non-Drinker", "Drinker"),
       fill = c("lightblue", "lightgreen"))
```



This box plot shows both Drinkers and Non-Drinkers have similar hours of absence patterns, with the majority of individuals in both groups having relatively low hours of absence.