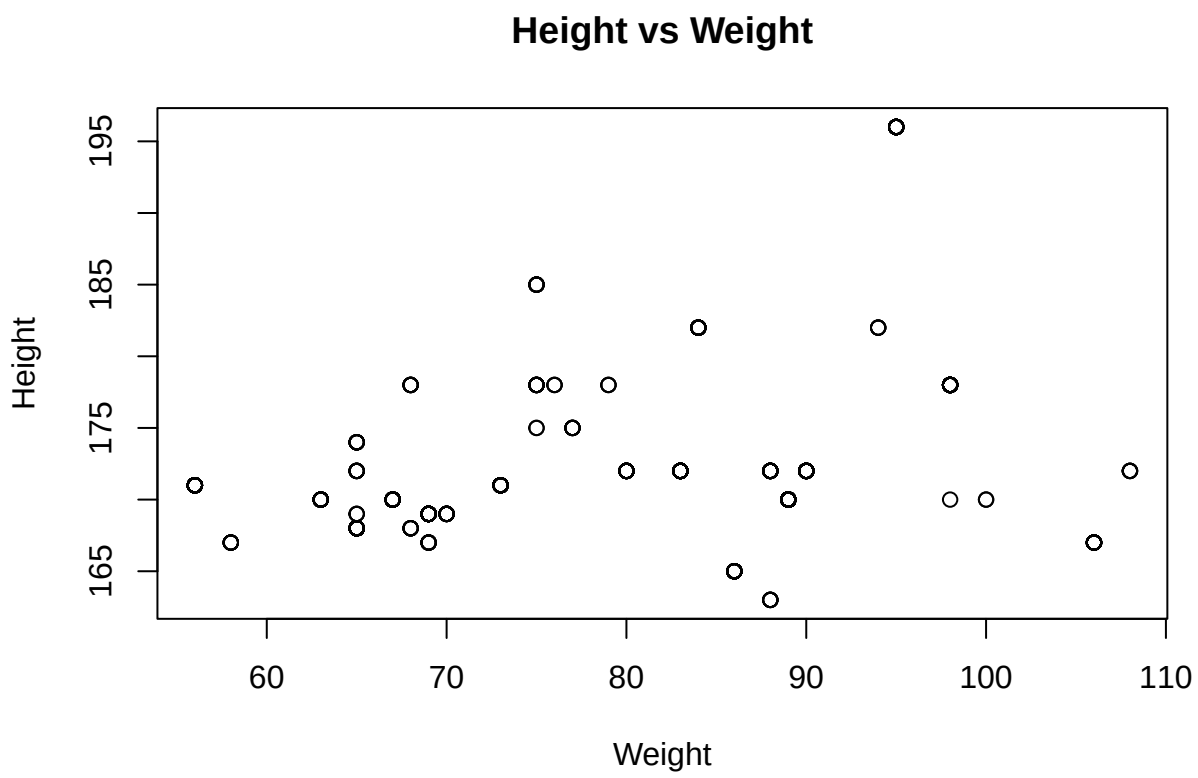
Lab 1

Taylor Saunders

2023-10-29

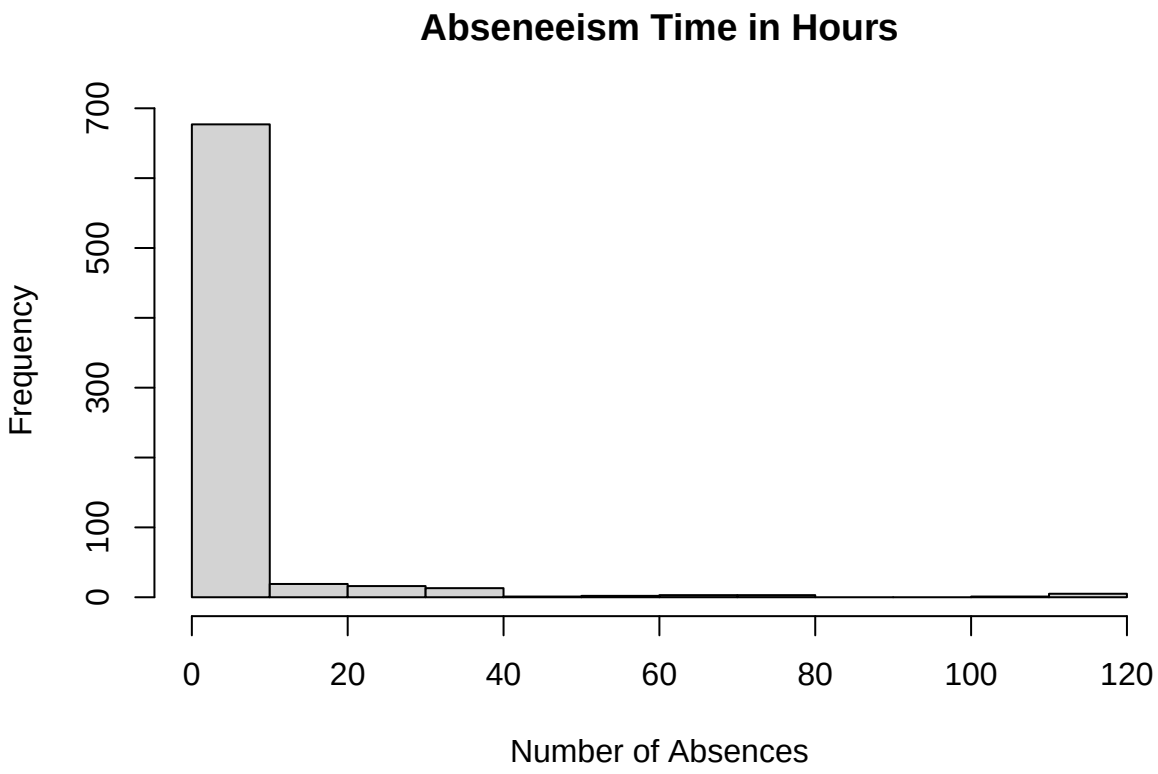
```
setwd("C:\\Users\\taylo\\Downloads\\DAT301")
df = read.csv("Absenteeism_at_work.csv",sep=";",header=TRUE)
```

```
#Problem 1
plot(df$Weight, df$Height, xlab ="Weight", ylab="Height", main = "Height vs Weight")
```



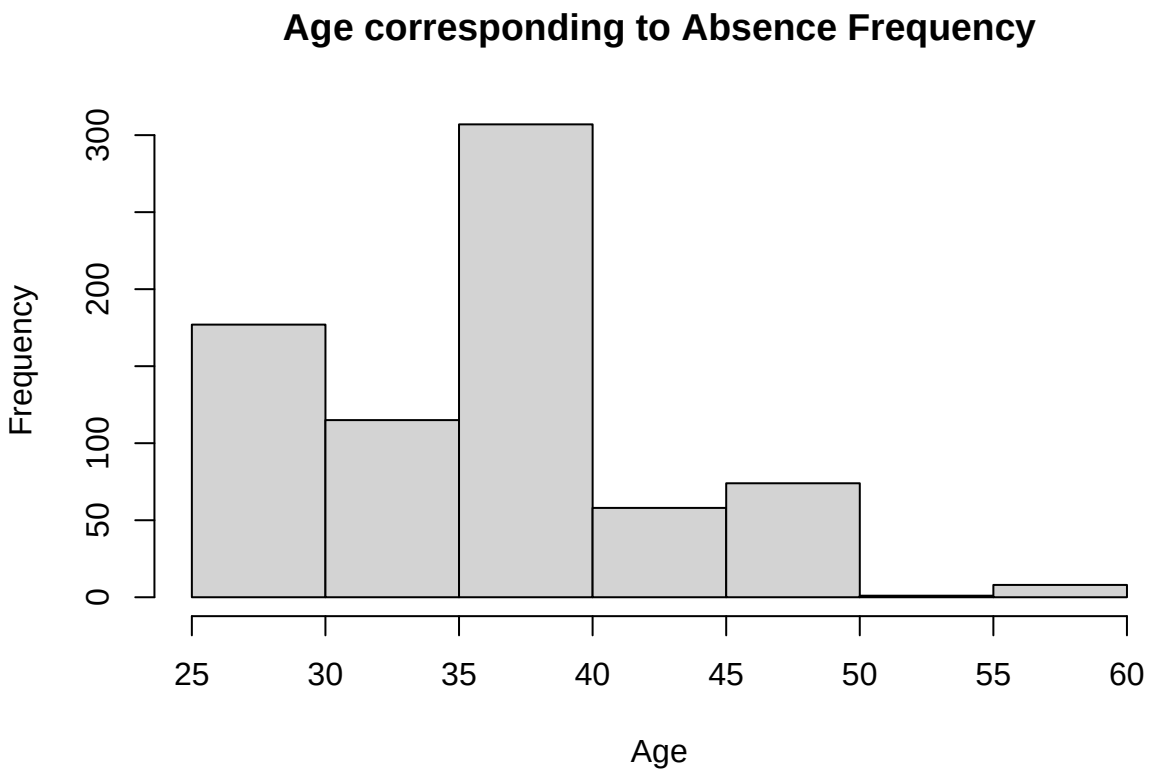
This graph suggests there is no correlation between weight and height.

```
#Problem 2
hist(df$Absenteeism.time.in.hours, xlab = "Number of Absences", main = "Abseneeism Time in Hours")
```



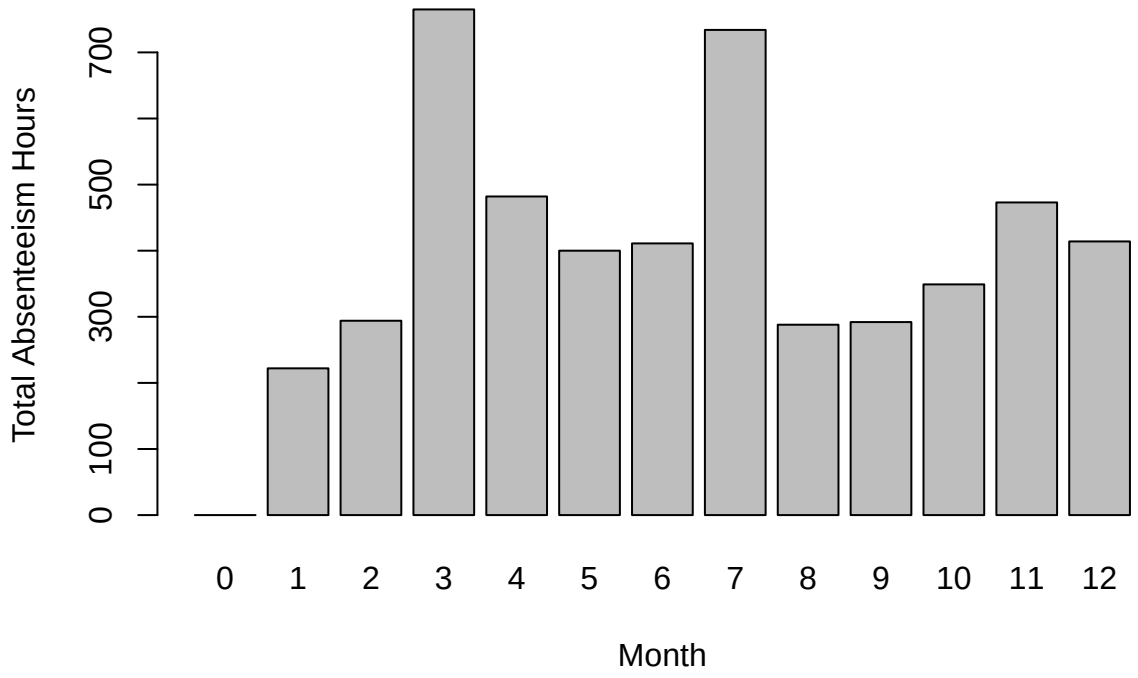
This graph shows that there is a majority of individuals with absent hour within 0-10 hours.

```
#Problem 3
#hist(df$Age)
hist(df$Age, xlab = "Age", main = "Age corresponding to Absence Frequency")
```



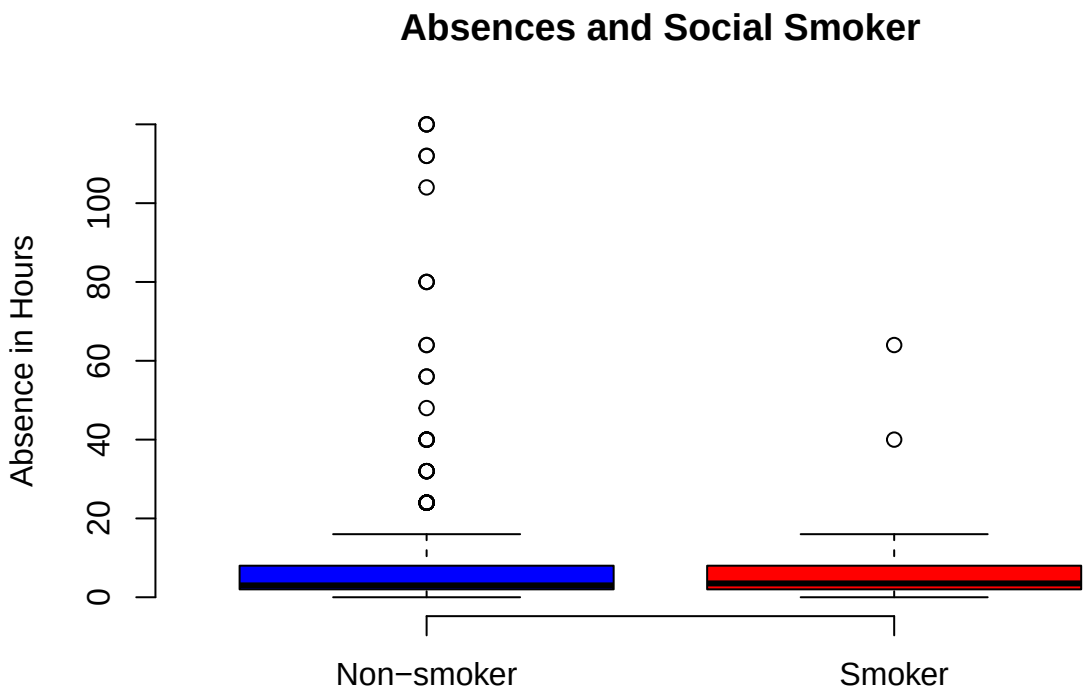
This graph suggests employees 35-40 years of age have the most absences.

```
#Problem 4
hours_by_month <- aggregate(df$Absenteeism.time.in.hours, by = list(month = df$Month), FUN = sum)

barplot(hours_by_month$x, names.arg = hours_by_month$month, xlab = "Month", ylab = "Total Absenteeism H
```



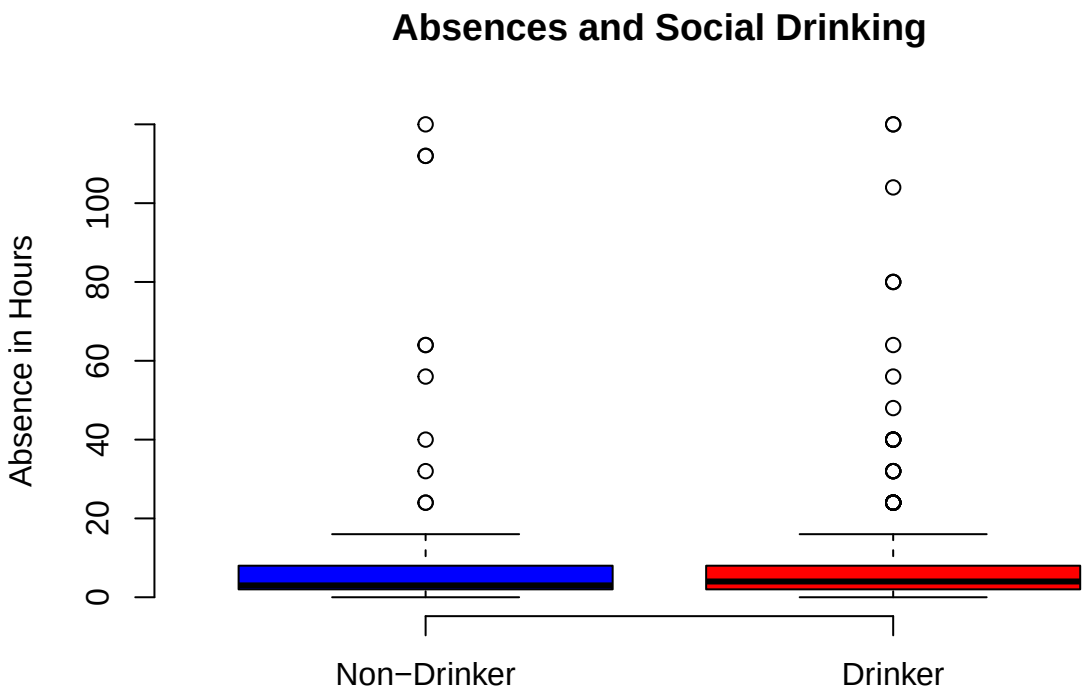
This graph suggests that during the months of March (3) and July (7) have the highest amount of absenteeism hours.

```
#Problem 5
df$Social.smoker = gsub(0, "No", df$Social.smoker)
df$Social.smoker = gsub(1, "Yes", df$Social.smoker)

boxplot(df$Absenteeism.time.in.hours~df$Social.smoker,
        main = "Absences and Social Smoker",
        xlab = "",
        names = c("Non-smoker", "Smoker"),
        ylab = "Absence in Hours",
        frame = FALSE,
        col = c('blue', 'red'))
```



This graph suggests non-smokers have a higher count of absent hours.

```
#Problem 6
df$Social.drinker = gsub(0, "No", df$Social.drinker)
df$Social.drinker = gsub(1, "Yes", df$Social.drinker)

boxplot(df$Absenteeism.time.in.hours~df$Social.drinker,
        main = "Absences and Social Drinking",
        xlab = "",
        names = c("Non-Drinker", "Drinker"),
        ylab = "Absence in Hours",
        frame = FALSE,
        col = c('blue', 'red'))
```



This graph does not display a large distinction between the absent hours of drinkers and individuals who do not drink.