Lab1

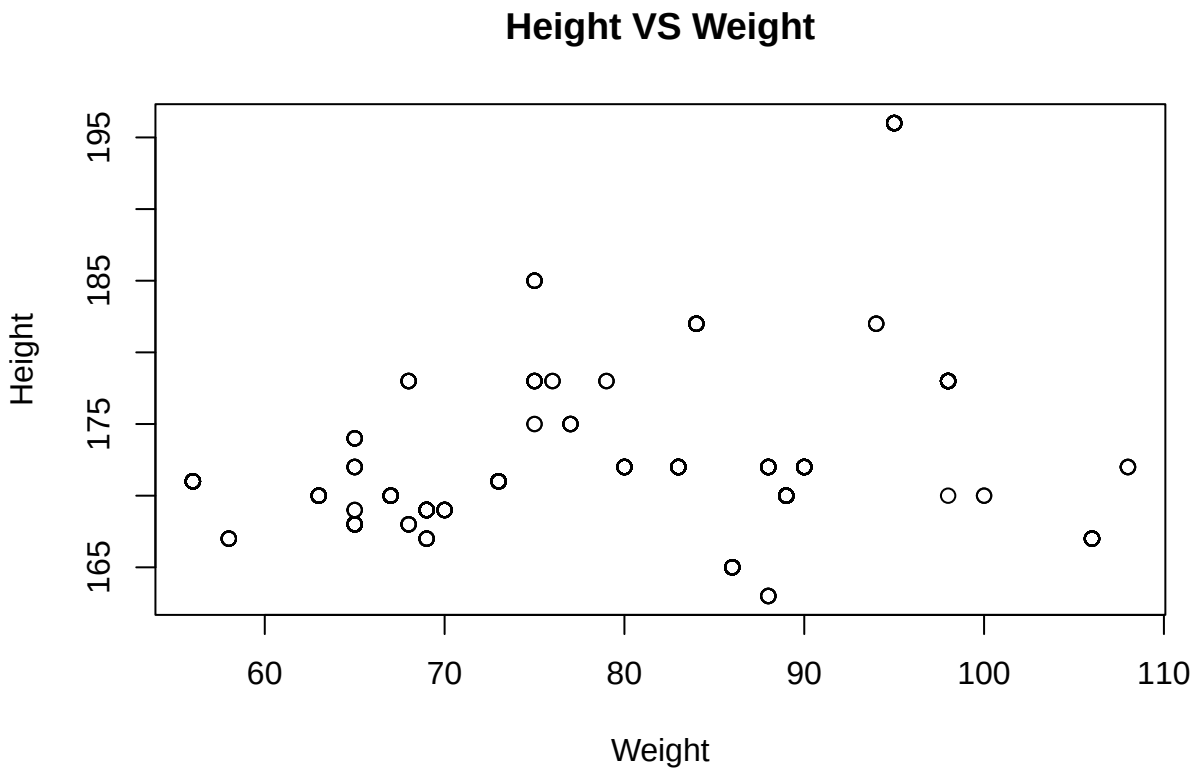
Tomohiro Justin Wada

2023-02-10

```
df = read.csv("Absenteeism_at_work.csv", sep = ";", header = TRUE)
attach(df)
```

Question 1

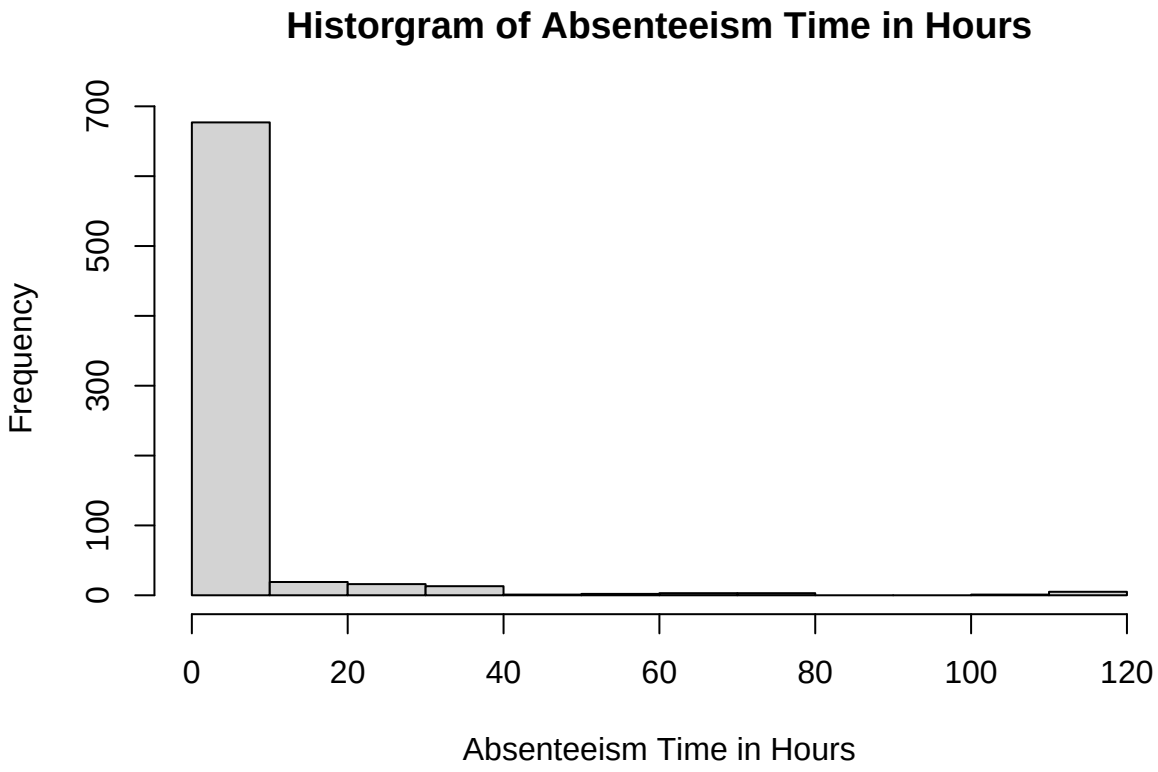
```
plot(Weight, Height, main = "Height VS Weight")
```



This is a scatter plot of Height vs Weight to depict if there is a correlation of height with weight. The plot does not depict of any significance to suggest anything.

Question 2

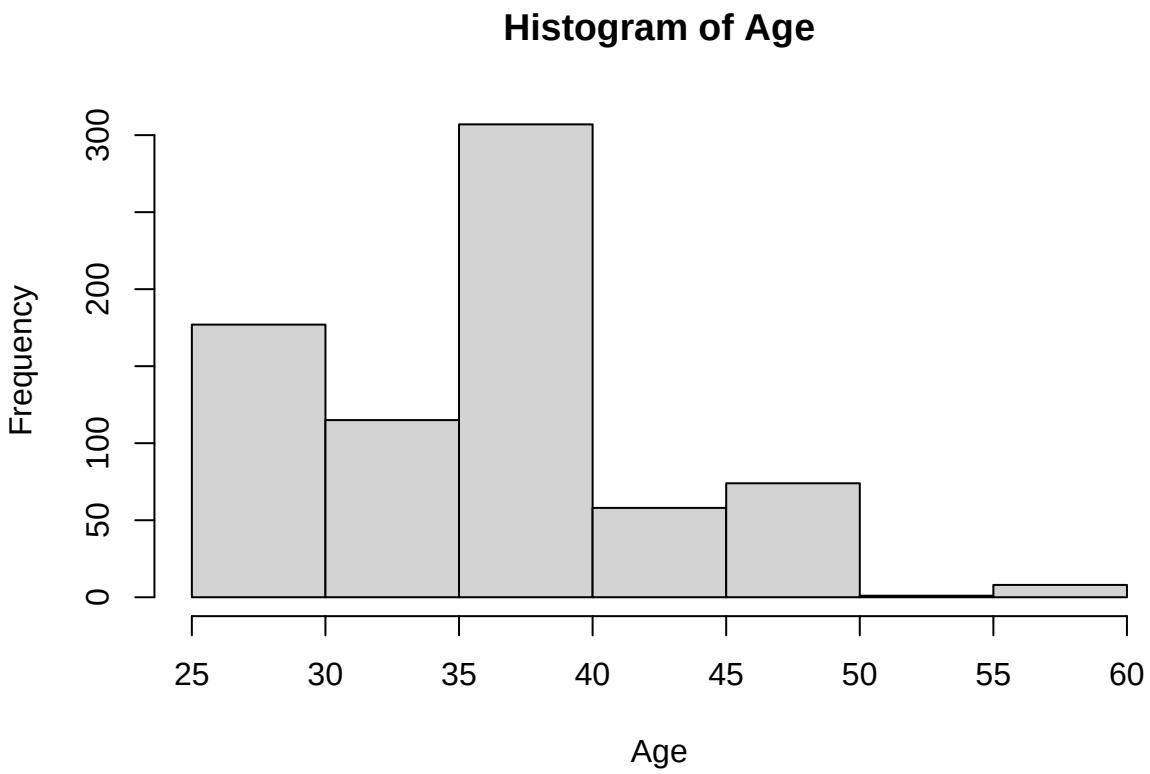
```
hist(Absenteeism.time.in.hours, xlab = "Absenteeism Time in Hours",
     main = "Historgram of Absenteeism Time in Hours")
```



This is a histogram representing absenteeism time in hours. It displays a large skew on the histogram not being too useful to decipher any useful information.

Question 3

```
hist(Age)
```



This is a histogram displaying the age and the total number of each age group. The histogram is useful to depict useful information regarding what is the main age demographic in this data.

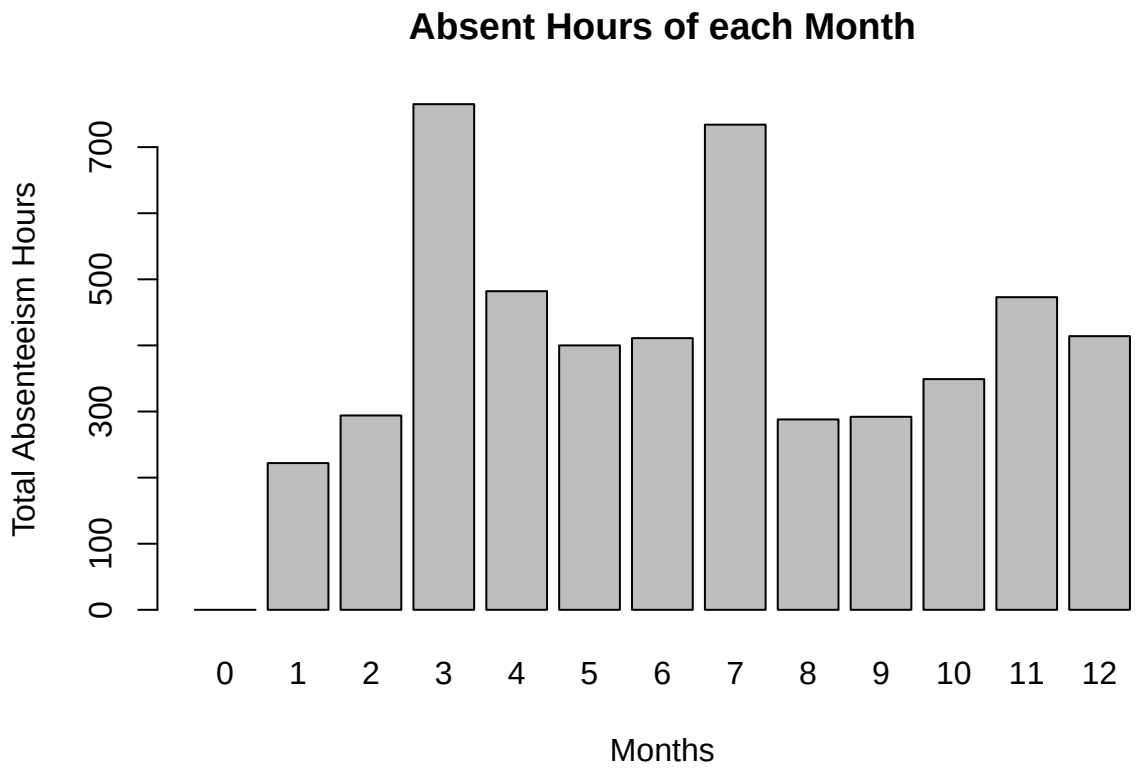
Quesiton 4

```
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
```

```
df_summary <- df %>%
  group_by(Month.of.absence) %>%
  summarise(total_absenteeism_hours = sum(Absenteeism.time.in.hours))

barplot(df_summary$total_absenteeism_hours,
```

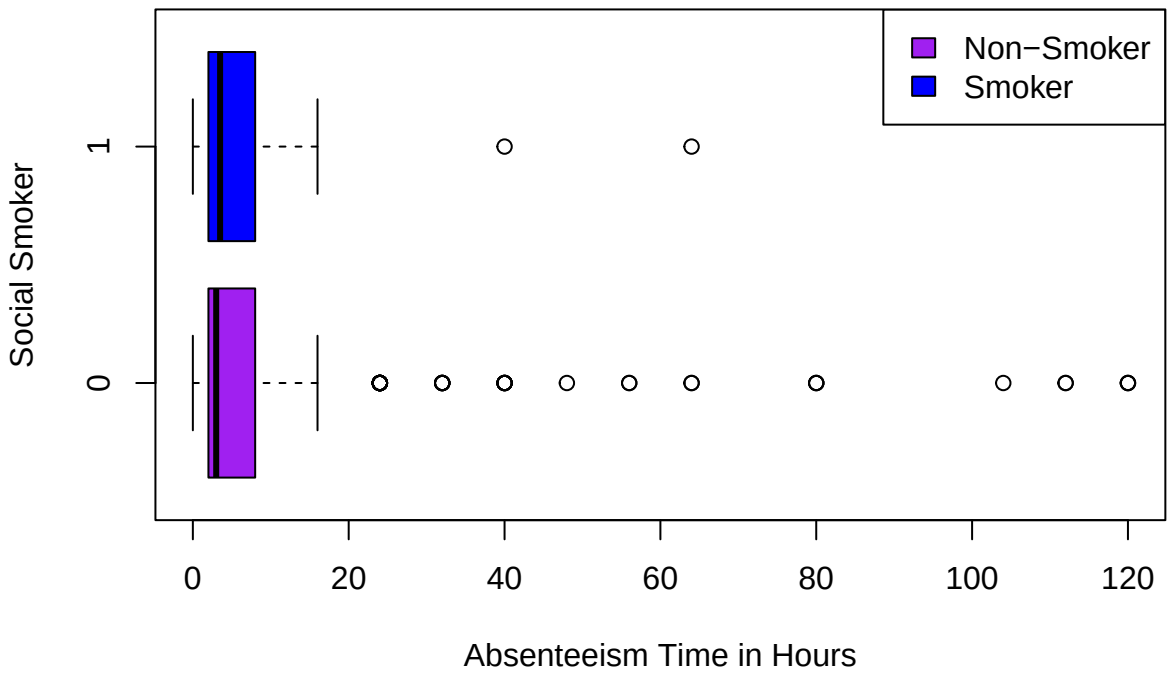
```
        names.arg = df_summary$Month.of.absence,
        xlab = "Months",
        ylab = "Total Absenteeism Hours",
        main = "Absent Hours of each Month")
```



This is a barplot displaying the total absenteeism hours for each month. This plot is useful to depict useful information regarding which month has the most hours absent.

Question 5

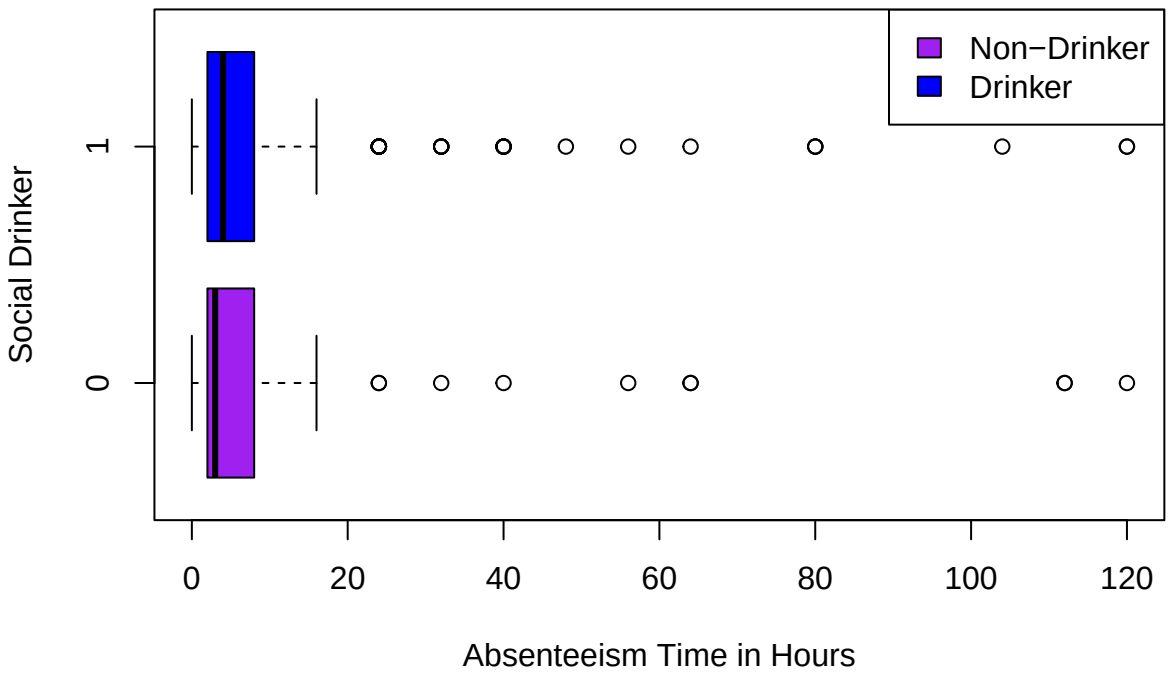
```
boxplot(Absenteeism.time.in.hours~Social.smoker, data=df, horizontal = T,
                                  col = c("purple", "blue"),
        xlab = "Absenteeism Time in Hours",
        ylab = "Social Smoker")

legend("topright", legend = c('Non-Smoker', 'Smoker'),
                    fil = c("purple", "blue"))
```



This is a boxplot displaying the absenteeism time in hours between smokers and non-smokers. This plot is not very useful to depict any information as the outlier points skews the boxplot.

Question 6

```
boxplot(Absenteeism.time.in.hours~Social.drinker, data=df, horizontal = T,
                                    col = c("purple", "blue"),
        xlab = "Absenteeism Time in Hours",
        ylab = "Social Drinker")

legend("topright", legend = c('Non-Drinker', 'Drinker'),
                    fil = c("purple", "blue"))
```



This is a boxplot displaying the absenteeism time in hours between drinkers and non-drinkers. This plot is not very useful to depict any information as the outlier points skews the boxplot.