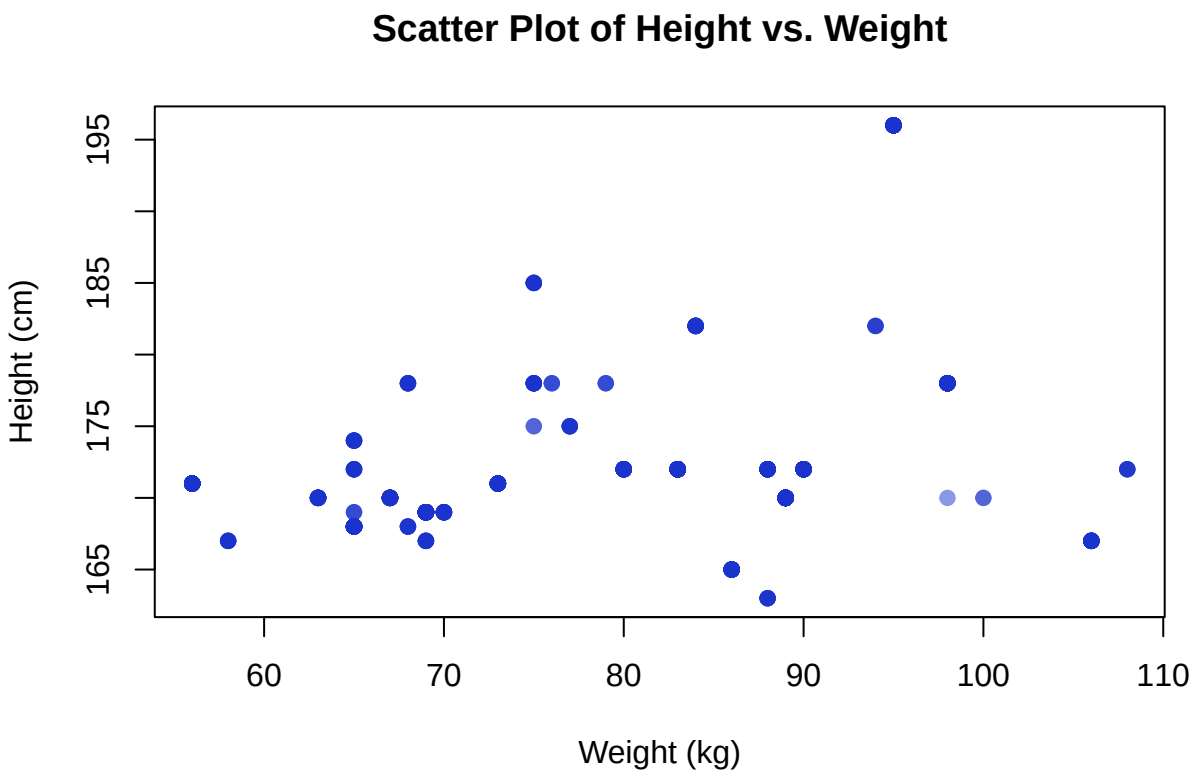
Absenteeism Analysis

Savaira Imran

2025-02-22

1. Scatter Plot of Height vs. Weight

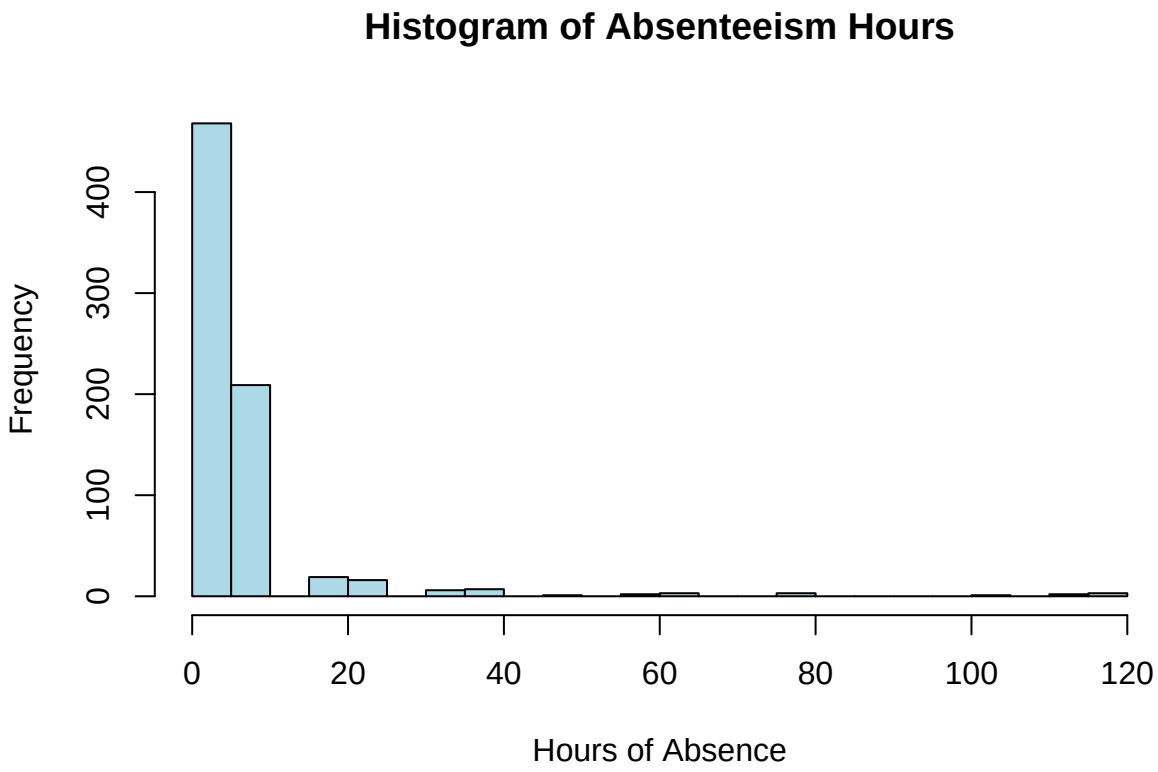
```
plot(df$Weight, df$Height, main="Scatter Plot of Height vs. Weight",
     xlab="Weight (kg)", ylab="Height (cm)", pch=19, col=rgb(0.1, 0.2, 0.8, 0.5))
```



As weight goes up, the plot points become more scattered.

2. Histogram

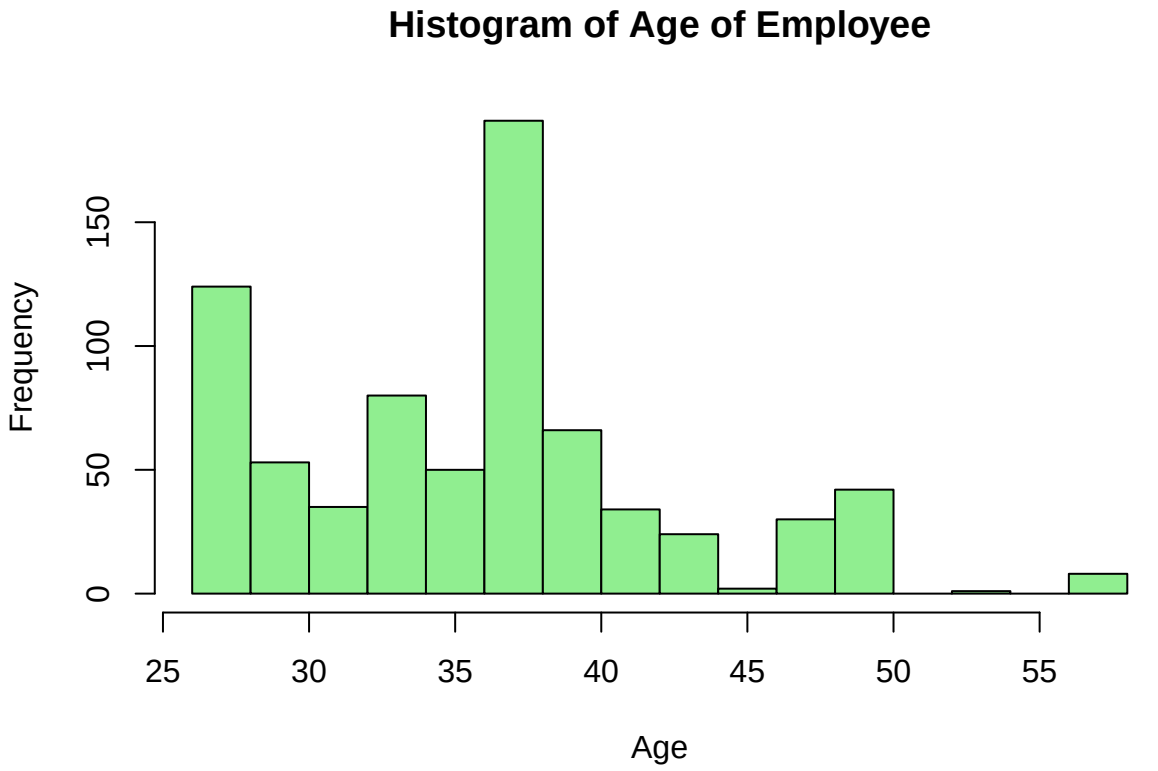
```
hist(df$Absenteeism.time.in.hours, main="Histogram of Absenteeism Hours",
     xlab="Hours of Absence", col="lightblue", border="black", breaks=20)
```



Frequency goes down as hours of absence go up.

3. Histogram of Age of Person Corresponding to each Absence

```
hist(df$Age, main="Histogram of Age of Employee",
     xlab="Age", col="lightgreen", border="black", breaks=15)
```



Between age 35-40, the frequency is highest.

4. Bar Plot of Hours by Month

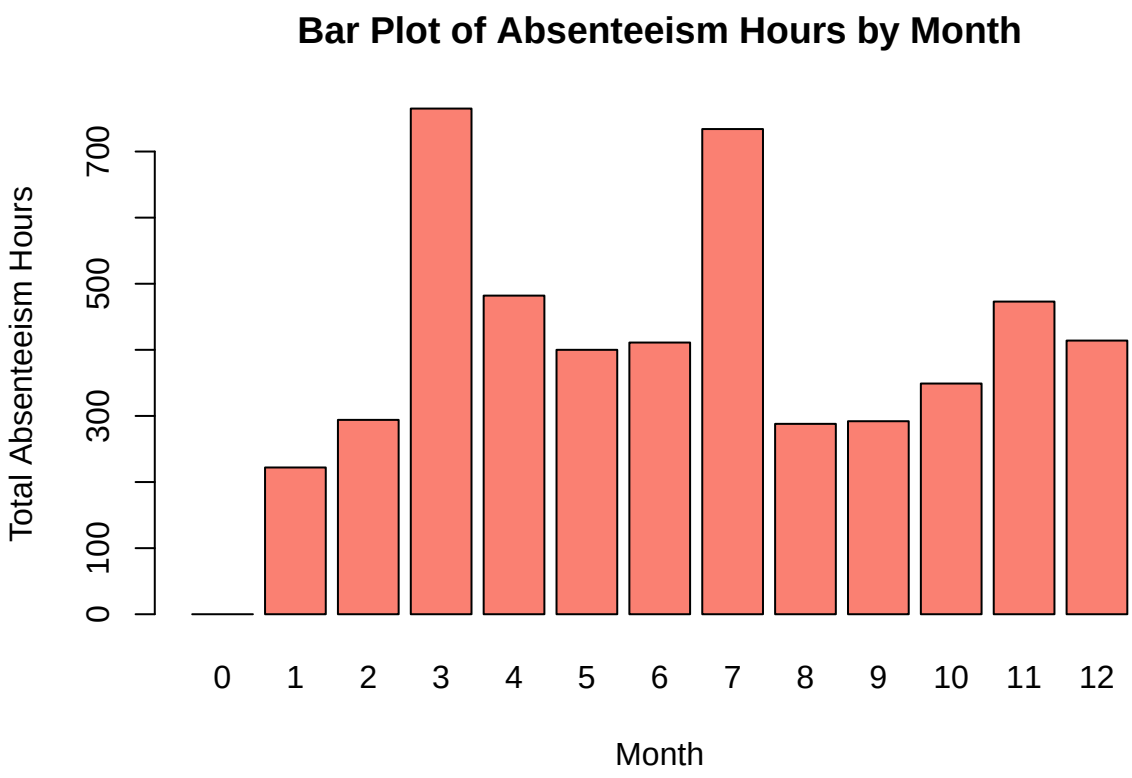
```
library(dplyr)
```

```
##
## 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
```

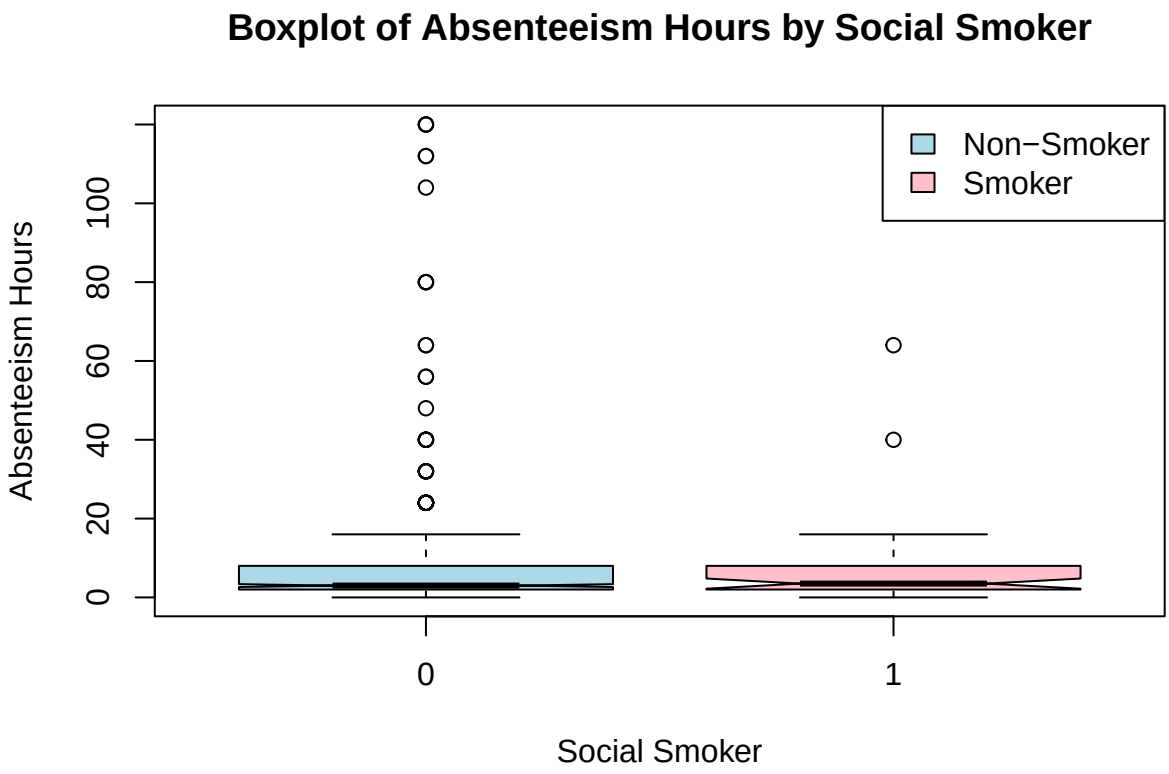
```
df_monthly <- df %>%
  group_by(Month.of.absence) %>%
  summarise(Total.absent.hours = sum(Absenteeism.time.in.hours, na.rm=TRUE))

barplot(df_monthly$Total.absent.hours, names.arg=df_monthly$Month.of.absence,
        main="Bar Plot of Absenteeism Hours by Month", xlab="Month", ylab="Total Absenteeism Hours",
        col="salmon", border="black")
```



Total hours are highest on month 3 and month 7.

5. Box plots of hours by social smoker status

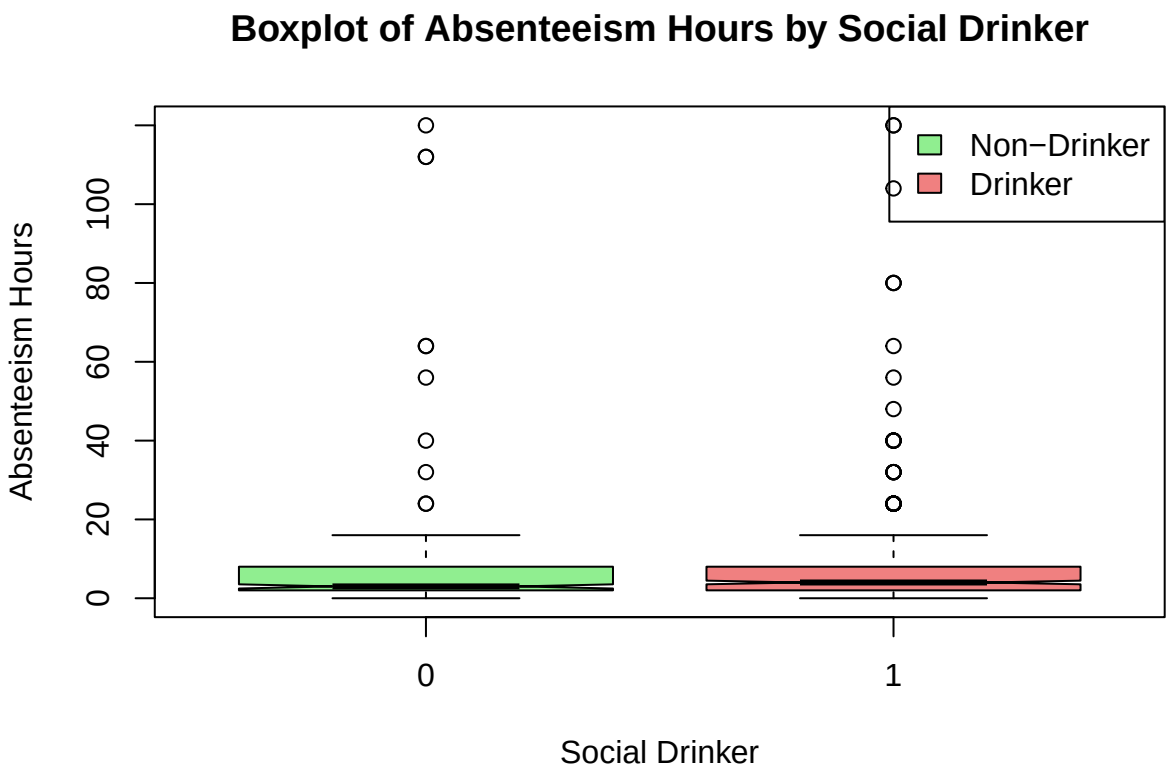
```
boxplot(Absenteeism.time.in.hours ~ Social.smoker, data=df,
        main="Boxplot of Absenteeism Hours by Social Smoker",
        xlab="Social Smoker", ylab="Absenteeism Hours",
        col=c("lightblue", "pink"), border="black", notch=TRUE)
legend("topright", legend=c("Non-Smoker", "Smoker"), fill=c("lightblue", "pink"))
```



Non-smokers have a lot of outliers.

6. Box Plots of Hours by Social Drinker Status

```
boxplot(Absenteeism.time.in.hours ~ Social.drinker, data=df,
        main="Boxplot of Absenteeism Hours by Social Drinker",
        xlab="Social Drinker", ylab="Absenteeism Hours",
        col=c("lightgreen", "lightcoral"), border="black", notch=TRUE)
legend("topright", legend=c("Non-Drinker", "Drinker"), fill=c("lightgreen", "lightcoral"))
```



The distribution of absenteeism hours for non-drinkers and drinkers are similar.