Activity 1.2

Facundo Bautista Barbera

Access the cars dataset.

```
library("tidyverse")
```

```
## -- Attaching core tidyverse packages ------------------------------------------------------------
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.2     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.1.0
## -- Conflicts --------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(nortest)
library(moments)
```

```
dataset <- cars
summary(dataset)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Create tests

Test the univariate normality of speed and stopping distance. Select at least two of the methods covered in class.

Shapiro Wilk

```
shapiro_test <- shapiro.test(dataset$speed)
print(shapiro_test)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  dataset$speed
## W = 0.97765, p-value = 0.4576
```

```r
shapiro_test <- shapiro.test(dataset$dist)
print(shapiro_test)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  dataset$dist
## W = 0.95144, p-value = 0.0391
```

Anderson Darling

```r
ad_test <- ad.test(dataset$speed)
print(ad_test)
```

```
##
##  Anderson-Darling normality test
##
## data:  dataset$speed
## A = 0.26143, p-value = 0.6927
```

```r
ad_test <- ad.test(dataset$dist)
print(ad_test)
```

```
##
##  Anderson-Darling normality test
##
## data:  dataset$dist
## A = 0.74067, p-value = 0.05021
```

Kolmogorov Smirnov

```r
lillie_test <- lillie.test(dataset$speed)
print(lillie_test)
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  dataset$speed
## D = 0.068539, p-value = 0.8068
```

```r
lillie_test <- lillie.test(dataset$dist)
print(lillie_test)
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  dataset$dist
## D = 0.12675, p-value = 0.04335
```

Jarque Bera

```r
jarque_test <- jarque.test(dataset$speed)
print(jarque_test)
```

```
##
##  Jarque-Bera Normality Test
##
## data:  dataset$speed
## JB = 0.80217, p-value = 0.6696
## alternative hypothesis: greater
```

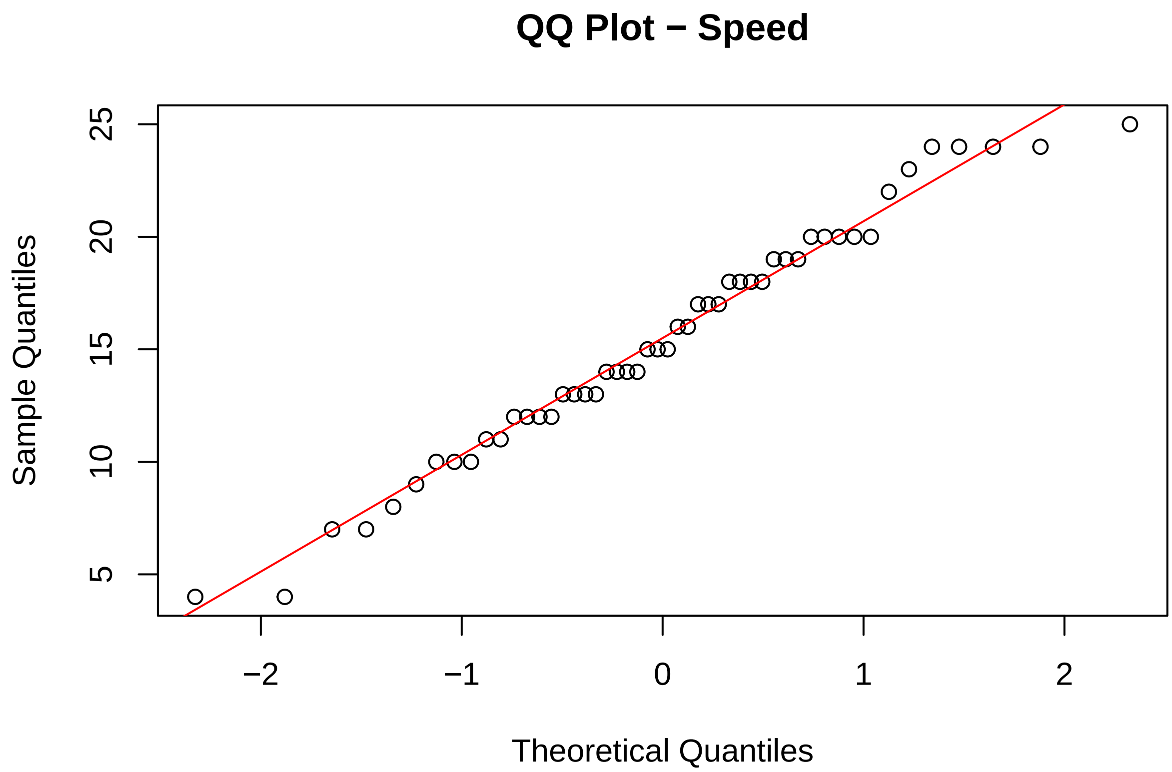
```r
jarque_test <- jarque.test(dataset$dist)
print(jarque_test)
```

```
##
##  Jarque-Bera Normality Test
##
## data:  dataset$dist
## JB = 5.2305, p-value = 0.07315
## alternative hypothesis: greater
```

QQPlot and QQLine

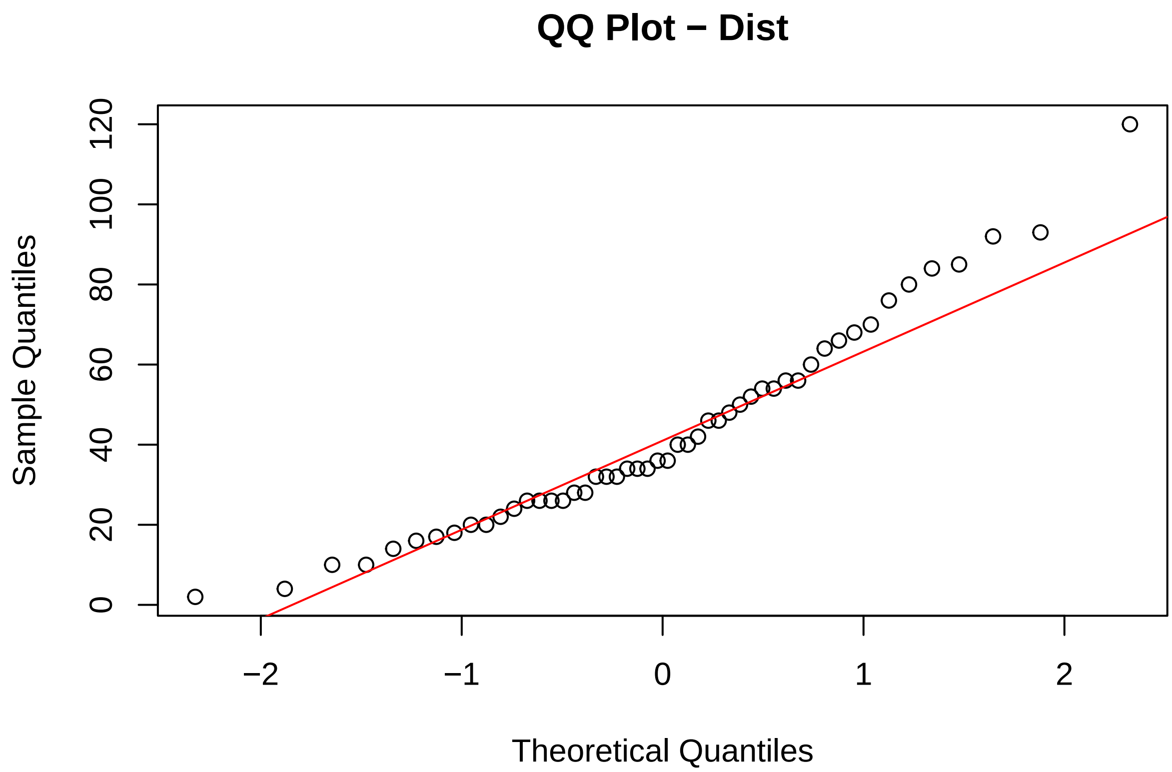
Speed

```r
qqnorm(dataset$speed, main = "QQ Plot - Speed")
qqline(dataset$speed, col = "red")
```



Dist

```
qqnorm(dataset$dist, main = "QQ Plot - Dist")
qqline(dataset$dist, col = "red")
```



Skweness and kurtosis

Speed

```r
skewness(dataset$speed)
```

```
## [1] -0.1139548
```

```r
kurtosis(dataset$speed)
```

```
## [1] 2.422853
```

Dist

```r
skewness(dataset$dist)
```

```
## [1] 0.7824835
```

```r
kurtosis(dataset$dist)
```

```
## [1] 3.248019
```

Mean, Median, Midrange

Speed

```r
mean_speed <- mean(dataset$speed)
median_speed <- median(dataset$speed)
midrange_speed <- (min(dataset$speed) + max(dataset$speed)) / 2

print(c(mean_speed, median_speed, midrange_speed))
```

```
## [1] 15.4 15.0 14.5
```

Dist

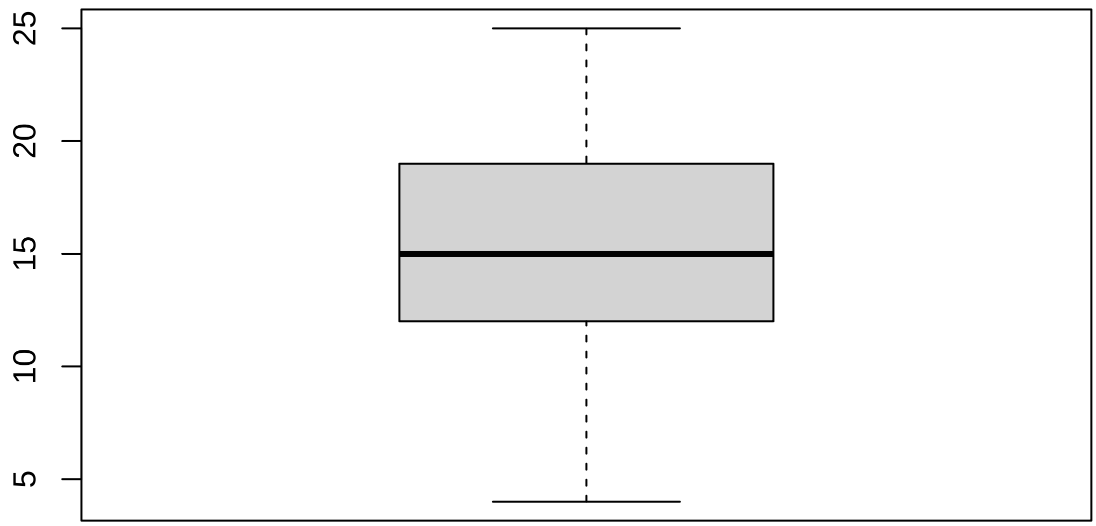
```r
mean_dist <- mean(dataset$dist)
median_dist <- median(dataset$dist)
midrange_dist <- (min(dataset$dist) + max(dataset
$dist)) / 2

print(c(mean_dist, median_dist, midrange_dist))
```

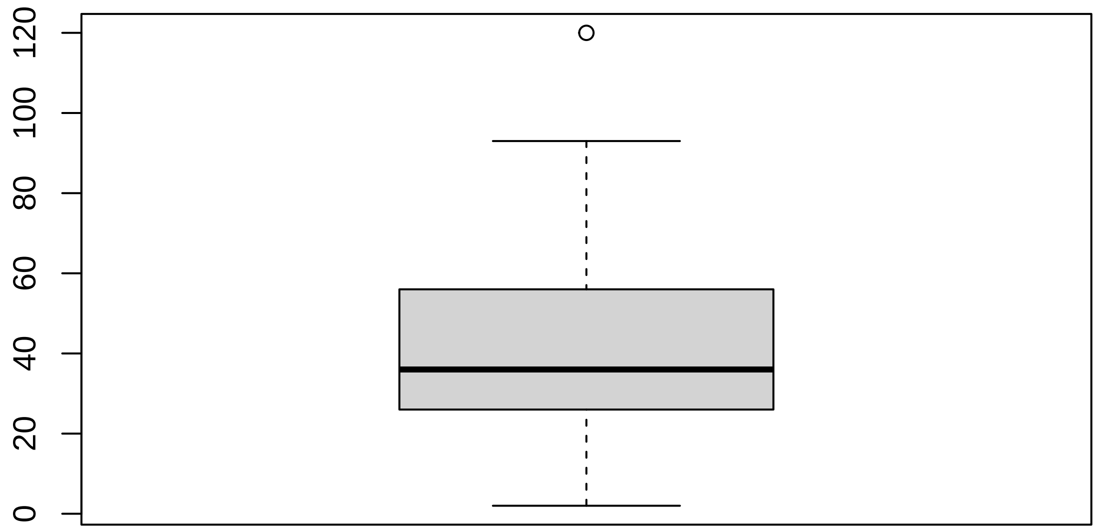
```
## [1] 42.98 36.00 61.00
```

Boxplot

```r
boxplot(dataset$speed)
```

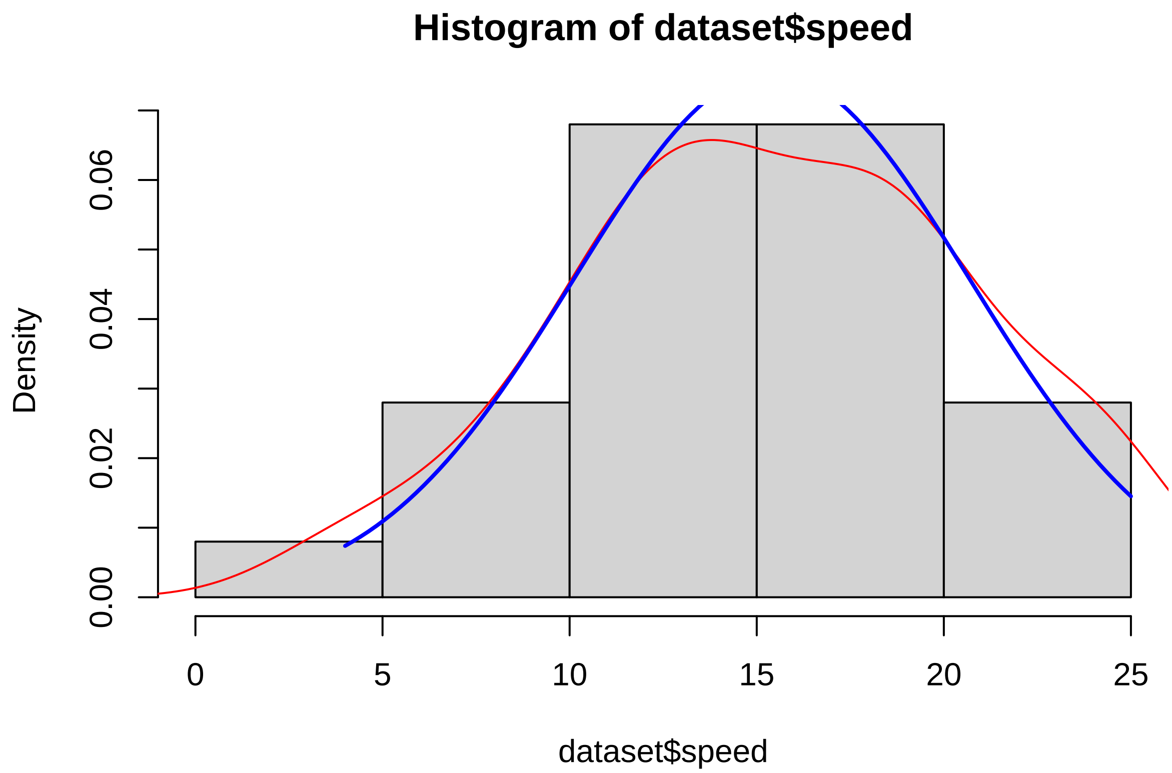



```r
boxplot(dataset$dist)
```

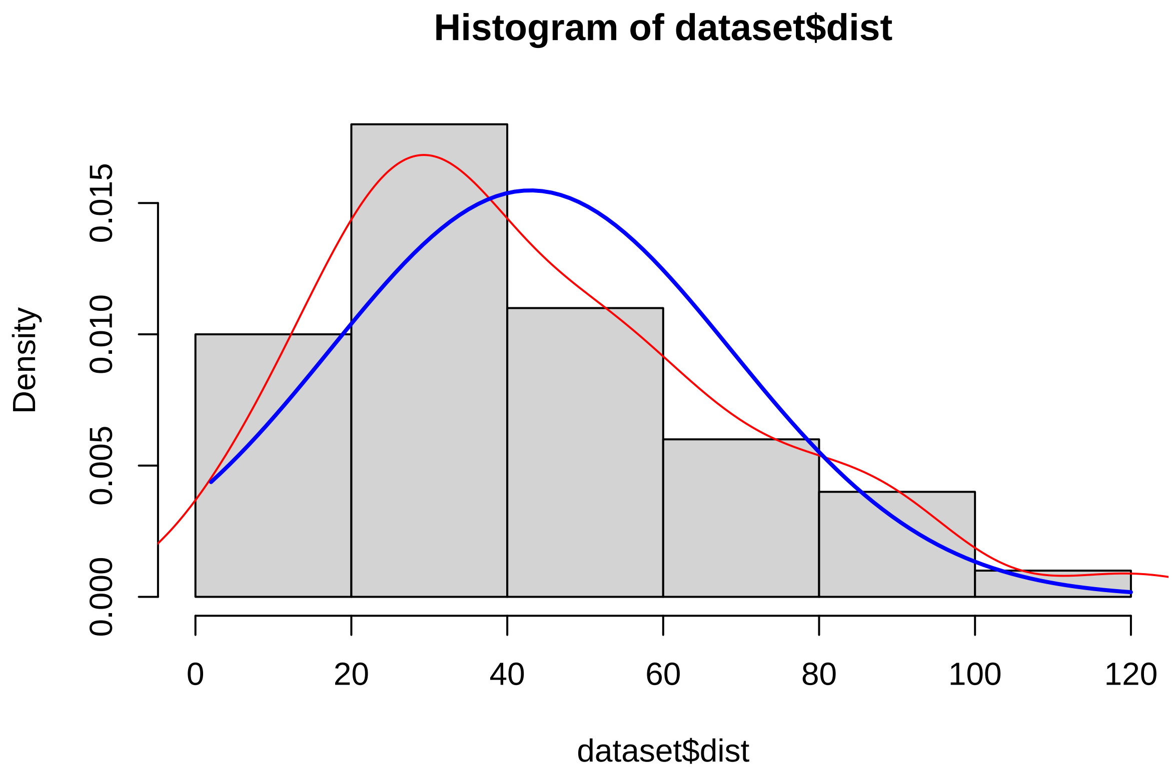


Histogram

```
hist(dataset$speed, probability = TRUE)
lines(density(dataset$speed), col = "red")
curve(dnorm(x, mean = mean(dataset$speed), sd = sd(dataset$speed)), from = min(dataset$speed), to = max
```




```r
hist(dataset$dist, probability = TRUE)
lines(density(dataset$dist), col = "red")
curve(dnorm(x, mean = mean(dataset$dist), sd = sd(dataset$dist)), from = min(dataset$dist), to = max(dat
```



The first plot looks like like a normal distribution, with a bit of skewness to the left. The second plot looks very skewed to the right, with a long tail on the right side.