ENV 710: Lab 4

Jiahuan Li

Feb 14, 2023

Problem 1

myqqplot function

```
#my qqplot function

myqqplot <- function(c){
  d <- sort(c)
  n <- length(d)

  p <- numeric(n)
  qZ <- numeric(n)

  for (i in 1:n){
    p[i] <- (i - 0.5)/n
    qZ[i] <- qnorm(p[i], mean = 0, sd = 1)
  }

  i1 <- 0.25 * n + 0.5
  i2 <- 0.75 * n + 0.5
  x1 <- qZ[i1]
  x2 <- qZ[i2]
  y1 <- d[i1]
  y2 <- d[i2]

  m <- (y2-y1)/(x2-x1)
  b <- y2 - m*x2


  fig <- plot (qZ, d, xlab = "Theoretical quantiles", ylab = "Sample values") + abline(b, m)
  return(fig)
}
```

(a) `qqplot` generated using `myqqplot` function:

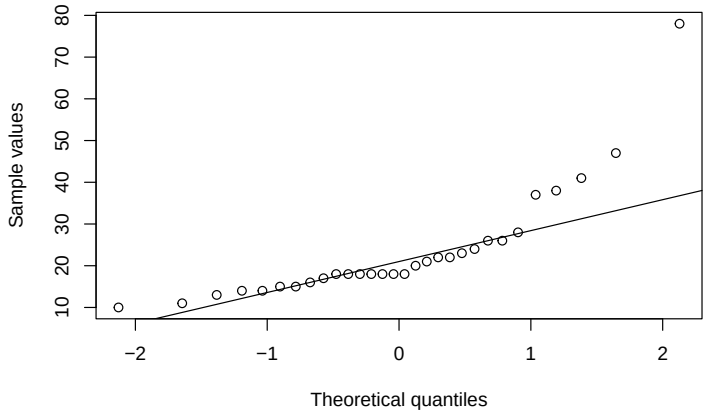

Figure 1: Problem 1a

```
## integer(0)
```

(b) `qqplot` generated using `qqnorm` and `qqline`:

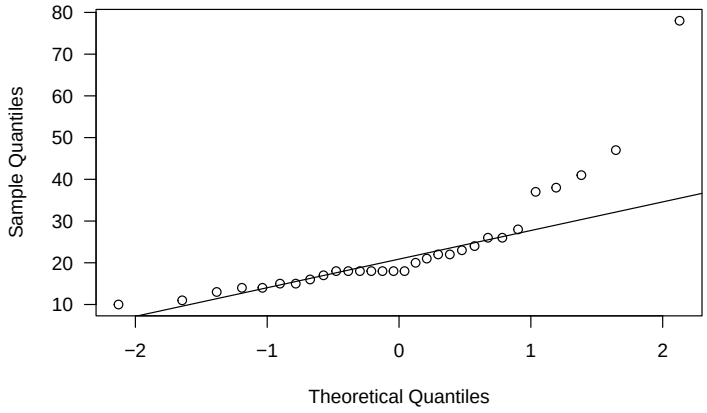

Figure 2: Problem 1b

Processing summary of my qqplot: at first we should understand the meaning of x and y axis. If we do not want to calculate Z scores of the original data to standardize the graph (making the y values also distributed in the interval around (-2,2)), the y values should just be the original observation values.

As for the calculation of x values, we should first find their linkages to y values, which is that they share the same percentile value. Thus, we should find the percentile of each observation by sorting and using the formula (i - 0.5)/n. Then, according to the corresponding percentile, we are able to calculate theoretical

quantiles as the x values for the standardized normal distribution. And the -0.5 in the formula is applied to avoid the mismatch of 100% percentile which is infinity in the normal distribution.

Problem 2

(a)

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
```

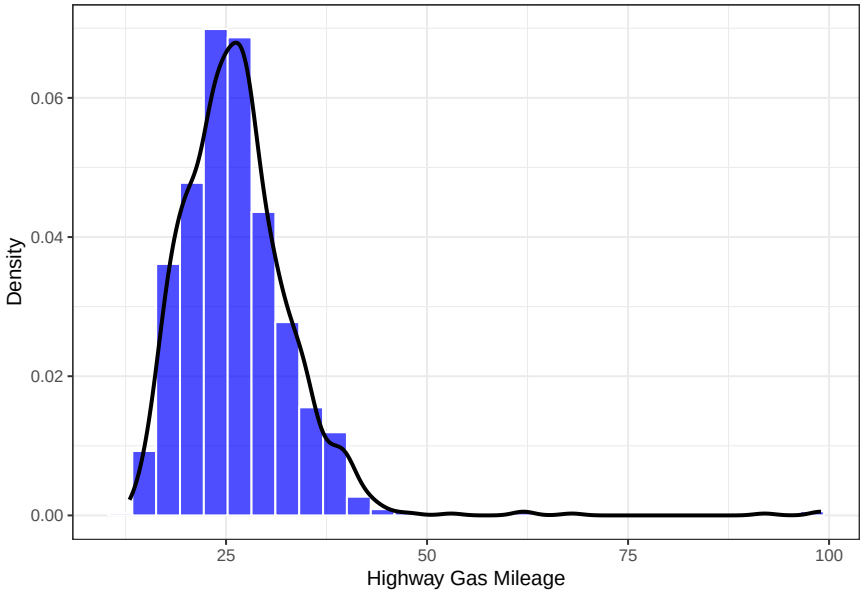

Figure 3: Density Histogram

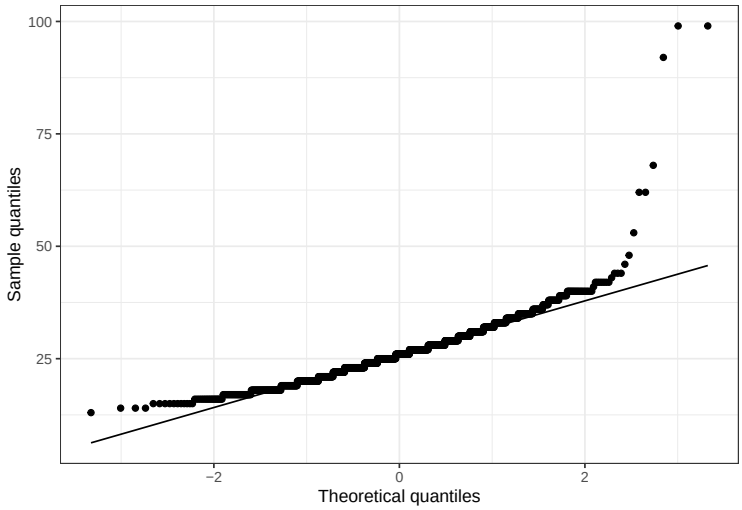

Figure 4: Q-Q plot

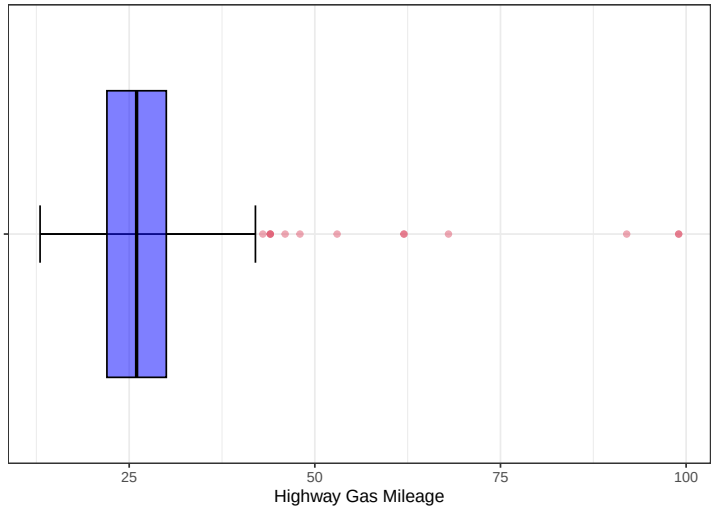

Figure 5: boxplot

```
## [1] "The skewness is 2.98725974253604"
```

```
## [1] "The kurtosis is 23.689899073668"
```

(b)

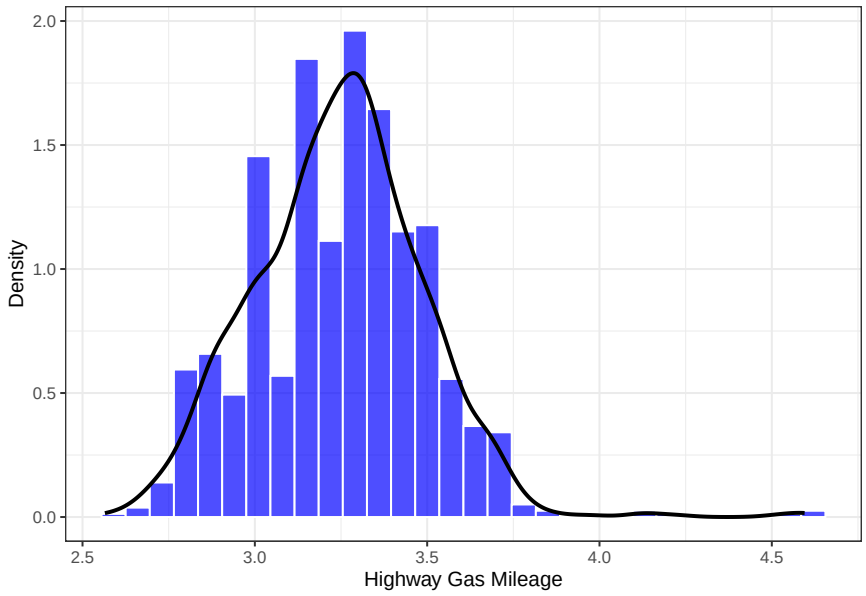

Figure 6: Logarithmic Dencity Histogram

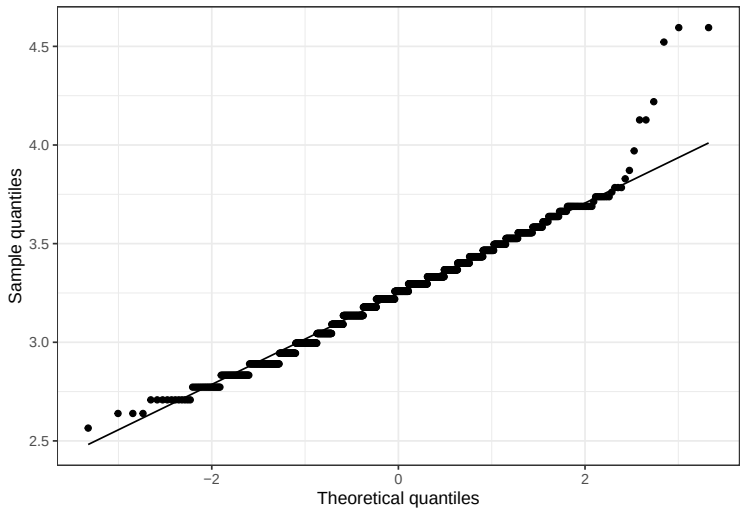

Figure 7: Logarithmic Q-Q Plot

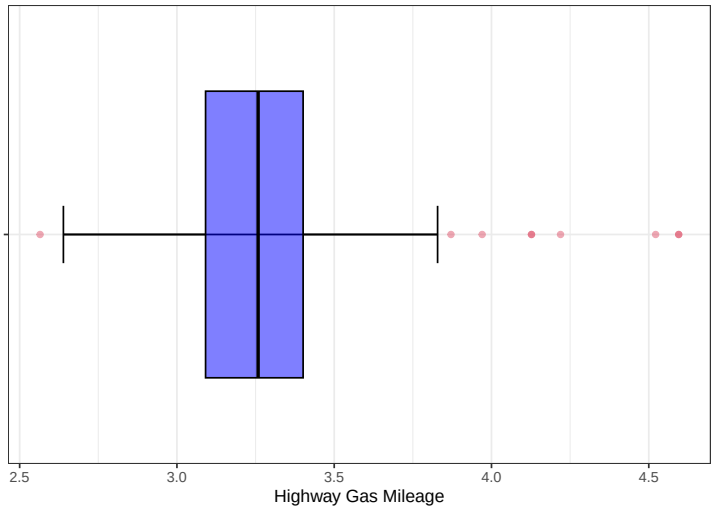

Figure 8: Logarithmic Boxplot

```
## [1] "The skewness is 0.446321227129158"
```

```
## [1] "The kurtosis is 1.97555651611997"
```

(c) Conclusion

The log-transformed data looks more normally distributed

```
## [1] "The mean is 3.24364314265321"
```

```
## [1] "THe CIs is ( 3.22940781560999 , 3.25787846969642 )"
```