group6_project630

Sagar Soneji, Roheet Bakare, Atish Panday

2022-12-06

1) Load the data and check for outliers

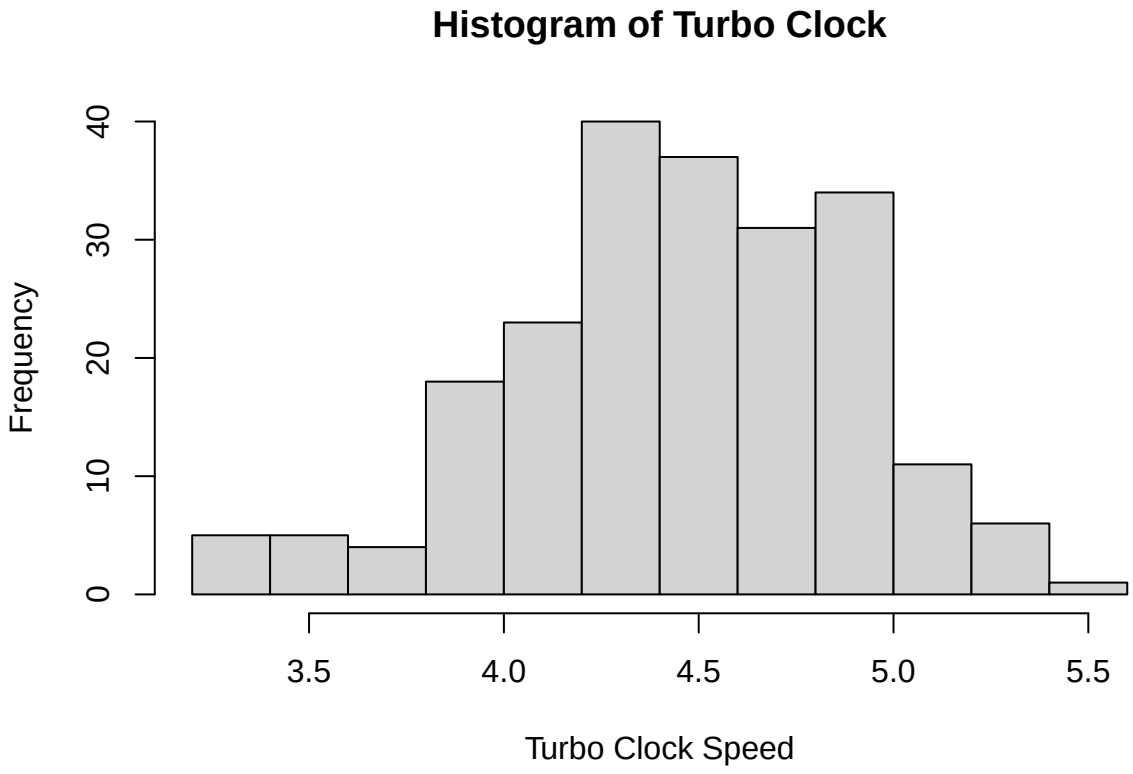
```
project <- read.csv("/Users/atishpanday/Downloads/CPU_r23_v2.csv")
project$manufacturer <- as.factor(project$manufacturer)
summary(project)
```

```
##  manufacturer   cpuName           singleScore     multiScore
##  AMD  : 86    Length:215         Min.   : 903   Min.   : 5423
##  Apple:  4    Class :character   1st Qu.:1207   1st Qu.: 8186
##  Intel:125    Mode  :character   Median :1312   Median :10890
##                                  Mean   :1368   Mean   :12980
##                                  3rd Qu.:1534   3rd Qu.:14394
##                                  Max.   :2082   Max.   :75671
##      cores           threads         baseClock       turboClock
##  Min.   : 4.000   Min.   :  6.00   Min.   :1.100   Min.   :3.200
##  1st Qu.: 6.000   1st Qu.: 12.00   1st Qu.:2.500   1st Qu.:4.200
##  Median : 8.000   Median : 16.00   Median :3.200   Median :4.500
##  Mean   : 9.367   Mean   : 17.74   Mean   :3.054   Mean   :4.515
##  3rd Qu.: 9.000   3rd Qu.: 16.00   3rd Qu.:3.600   3rd Qu.:4.800
##  Max.   :64.000   Max.   :128.00   Max.   :4.200   Max.   :5.500
##      type
##  Length:215
##  Class :character
##  Mode  :character
##
##
##
```

2) Linear regression model

```
mod <- lm(singleScore ~ turboClock, data = project)
hist(project$singleScore, xlab = "Single Score", ylab = "Frequency", main = "Histogram of Single Score")
```


```
hist(project$turboClock, xlab = "Turbo Clock Speed", ylab = "Frequency", main = "Histogram of Turbo Clo
```

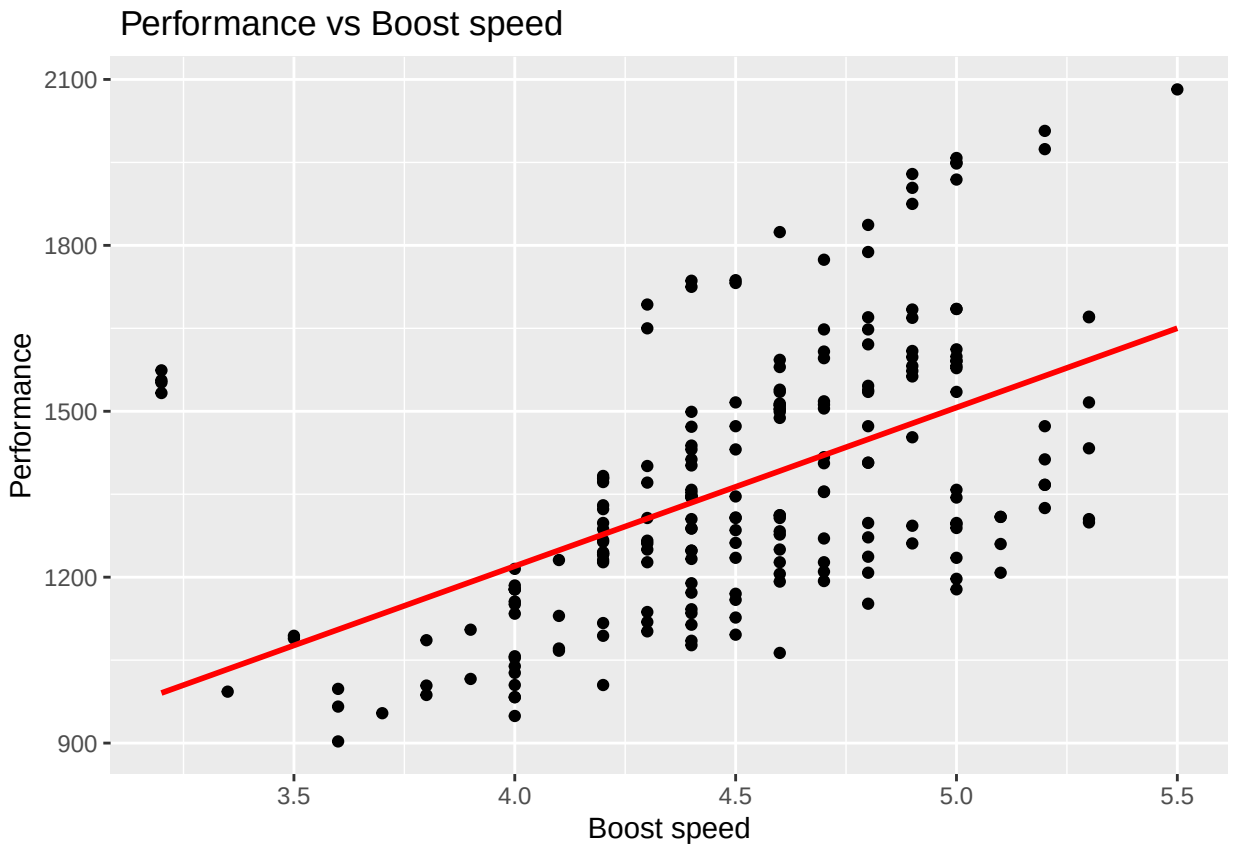


3) Check Conditions

Linearity check

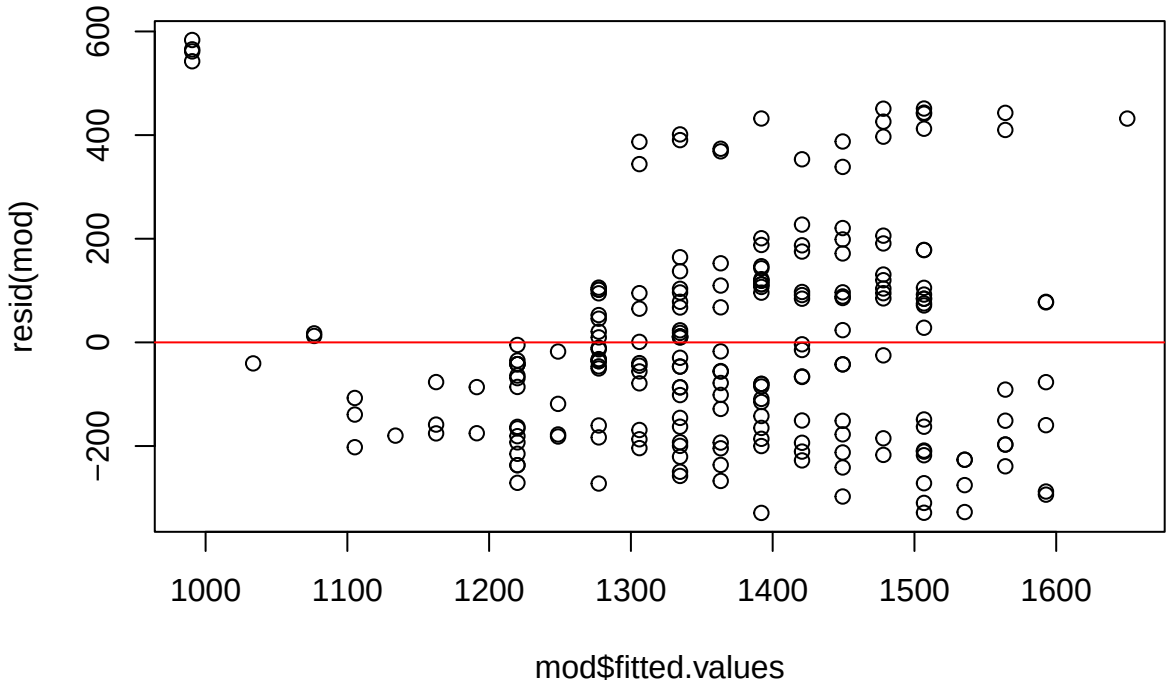
```
library(ggplot2)
ggplot(project, aes(x = turboClock, y = singleScore)) +
  geom_point() +
  geom_smooth(method = "lm",
              se = FALSE,
              col = "red") +
  ggtitle(" Performance vs Boost speed") +
  labs(x = "Boost speed", y = "Performance")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



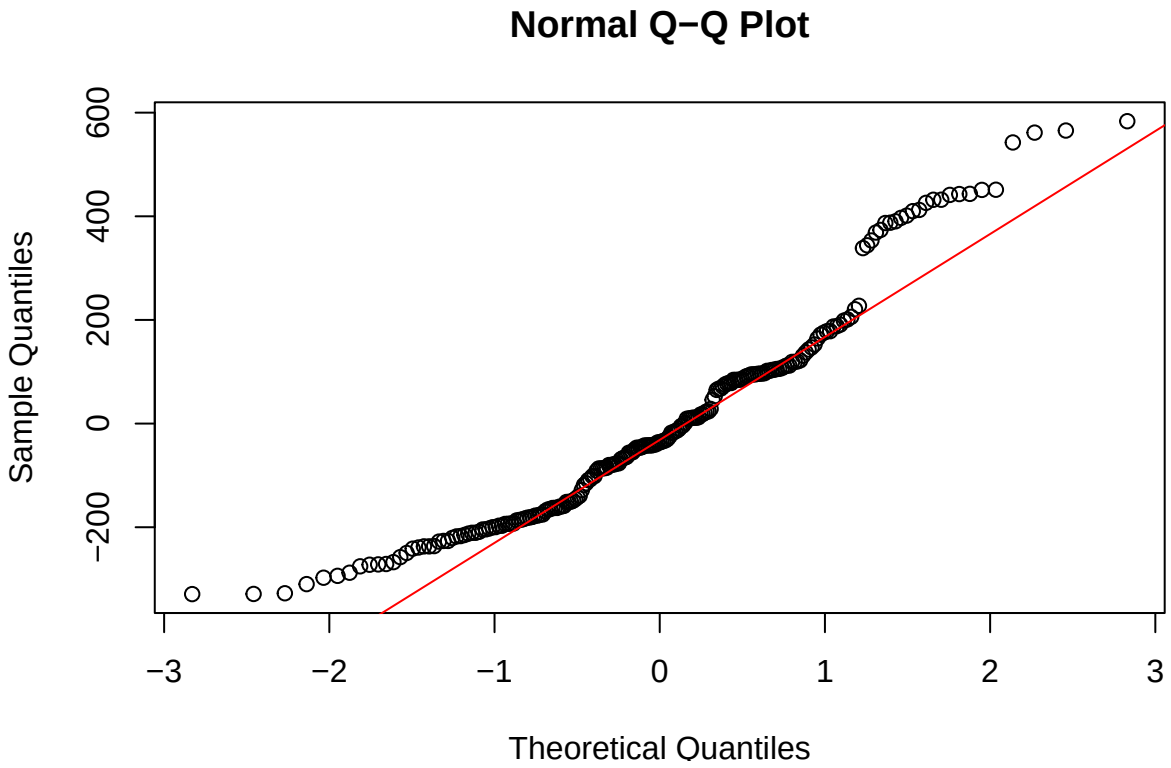
Constant variance check

```
mod <- lm(singleScore ~ turboClock, data = project)
plot(resid(mod) ~ mod$fitted.values)
abline(h = 0, col = "red")
```



Normality check

```
qqnorm(resid(mod))
qqline(resid(mod), col = "red")
```



Secondary Discovery

```
ggplot(project, aes(x = turboClock, y = singleScore, col = manufacturer)) +
  geom_point() +
```

```
ggtitle(" Performance vs Boost speed") +
labs(x = "Boost speed", y = "Performance") + scale_color_manual(breaks = c("AMD", "Apple", "Intel"),va
```