Homework 3

Radhakrishna Adhikari

9/29/2021

#Problem 3

First, we load the data and fit the data by linear regression as following.

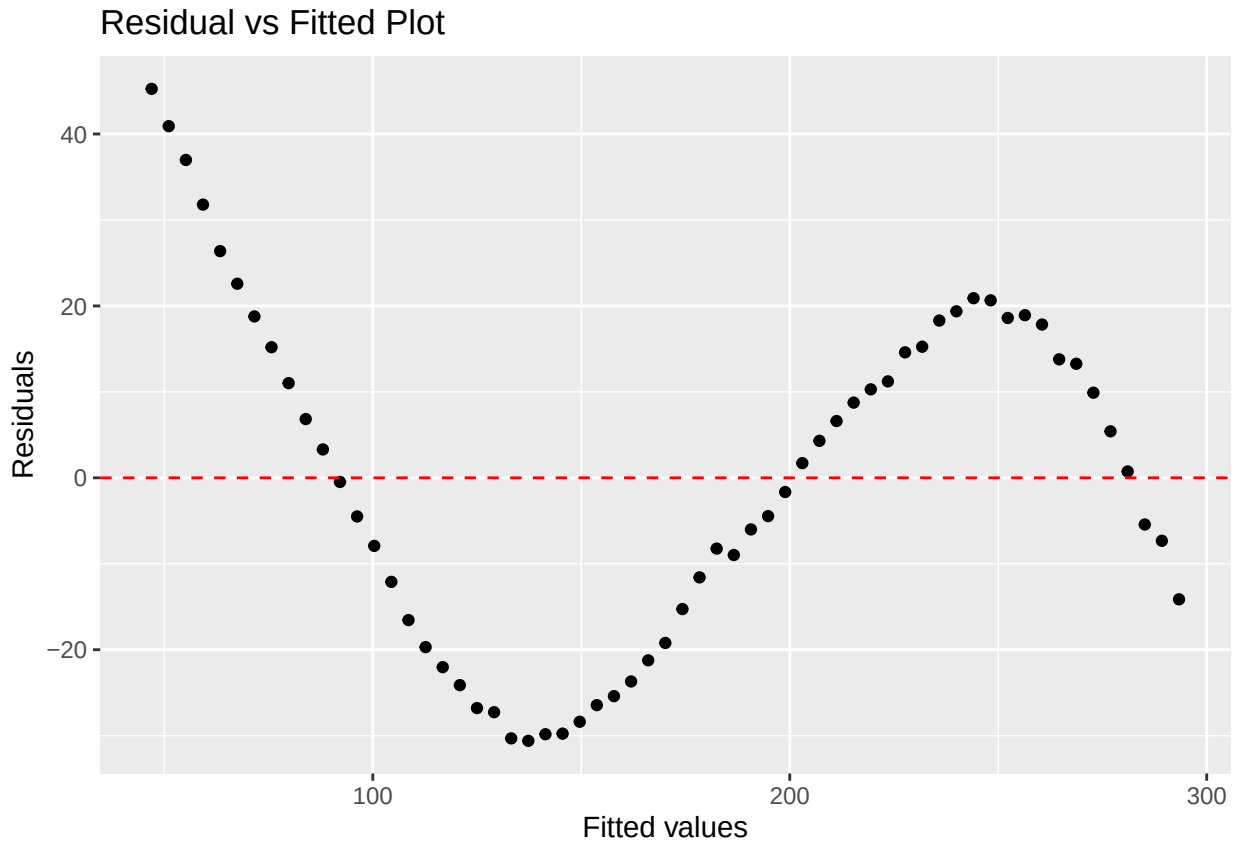
```
library(data.table)
library(broom)
covid_raw <- fread("https://opendata.ecdc.europa.eu/covid19/casedistribution/csv")
us <- covid_raw[covid_raw$countriesAndTerritories == 'United_States_of_America',]
us_filtered <- us[us$month %in% c(6:7),]
us_filtered$index <- rev(1:dim(us_filtered)[1])
fit<-lm(`Cumulative_number_for_14_days_of_COVID-19_cases_per_100000`~index, data=us_filtered)

## augment the data as previous
fit.diags <- broom::augment(fit)
```

Now, we plot the diagnostic plots as following.

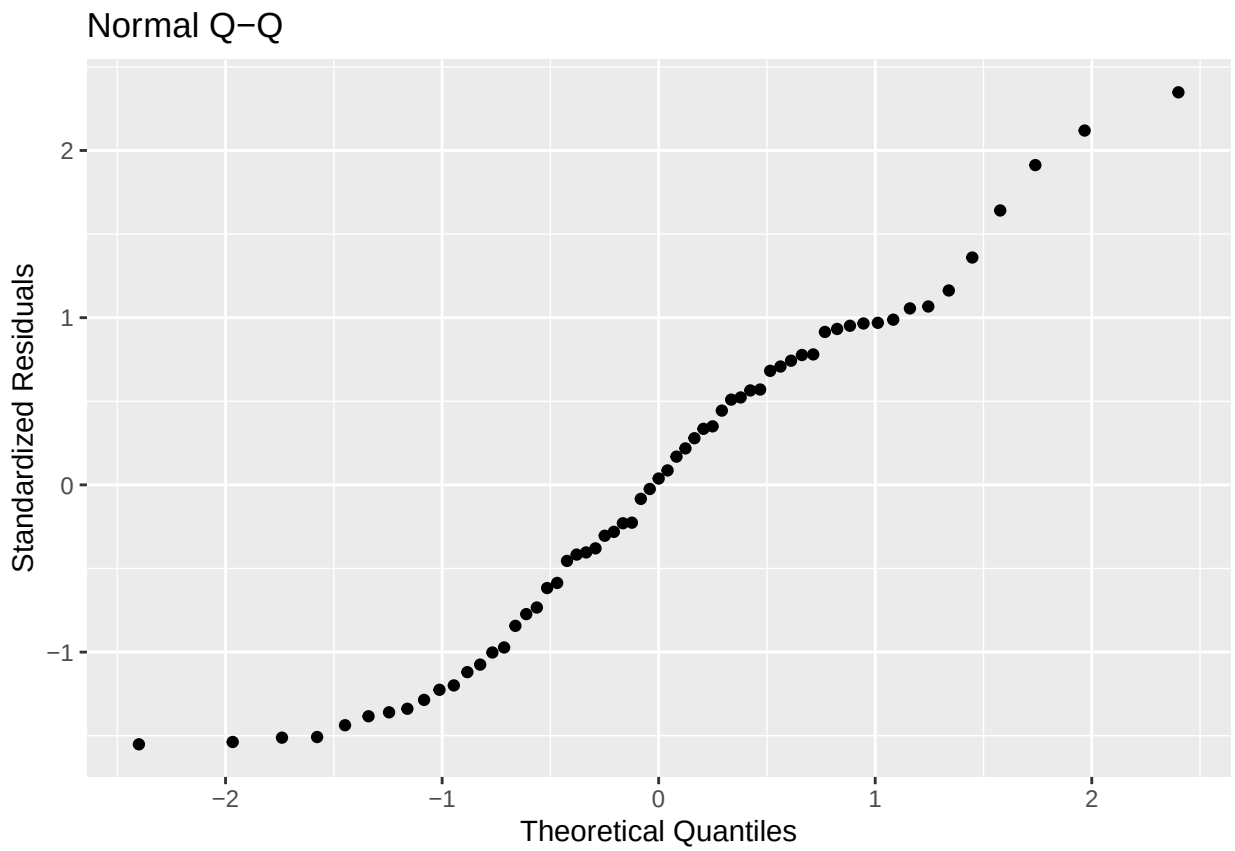
1. Here is the plot of Residual vs Fitted Plot

```
library(ggplot2)
resvsfitplot<-ggplot(fit, aes(.fitted, .resid))+geom_point()+geom_hline(yintercept=0, col="red",
    linetype="dashed")+xlab("Fitted values")+ylab("Residuals")+ggtitle("Residual vs Fitted Plot")
resvsfitplot
```



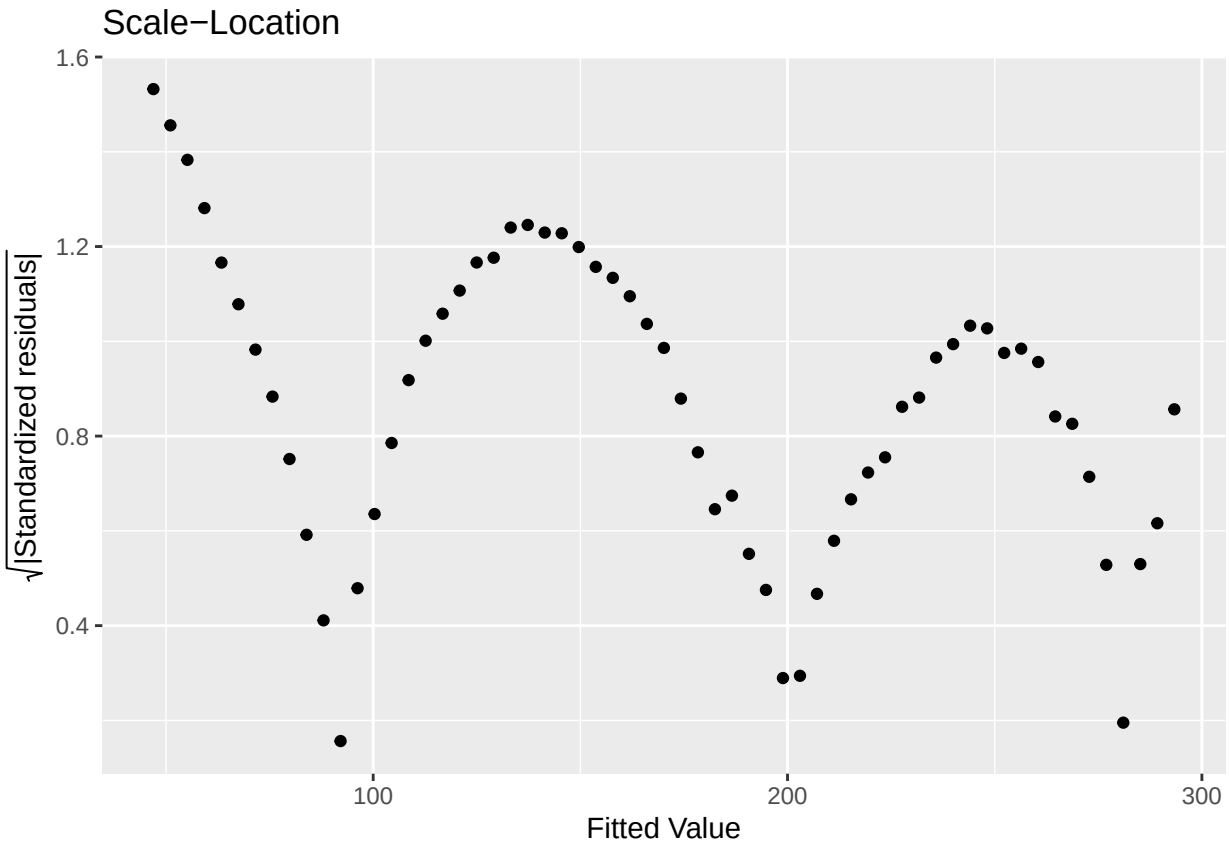
2. The plot of Normal Q-Q

```
qq<-ggplot(fit, aes(qqnorm(.stdresid)[[1]], .stdresid))+geom_point(na.rm = TRUE)+
  xlab("Theoretical Quantiles")+ylab("Standardized Residuals")+ggtitle("Normal Q-Q")
qq
```


Normal Q−Q
Standardized Residuals
Theoretical Quantiles
2
1
0
−1
−2
−1
0
1
2

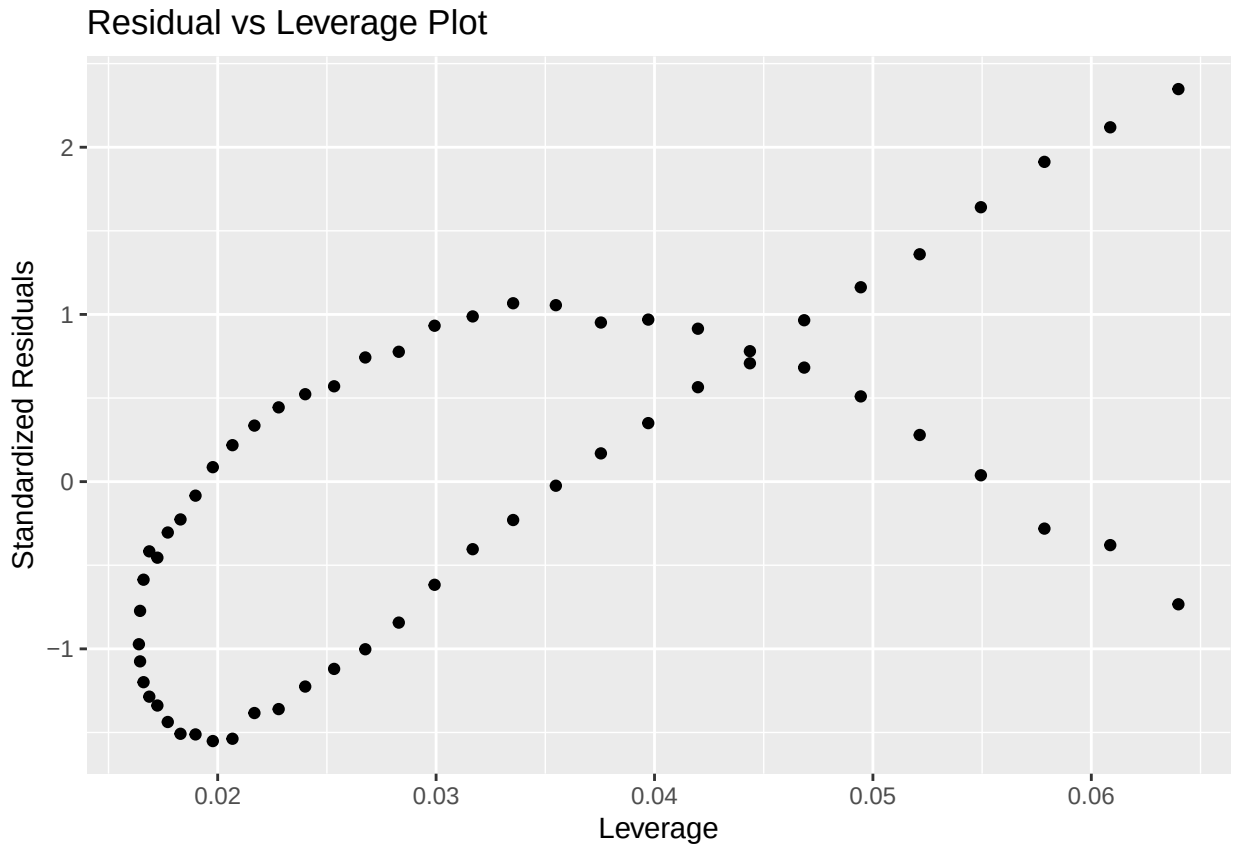
3. Scale Location plot

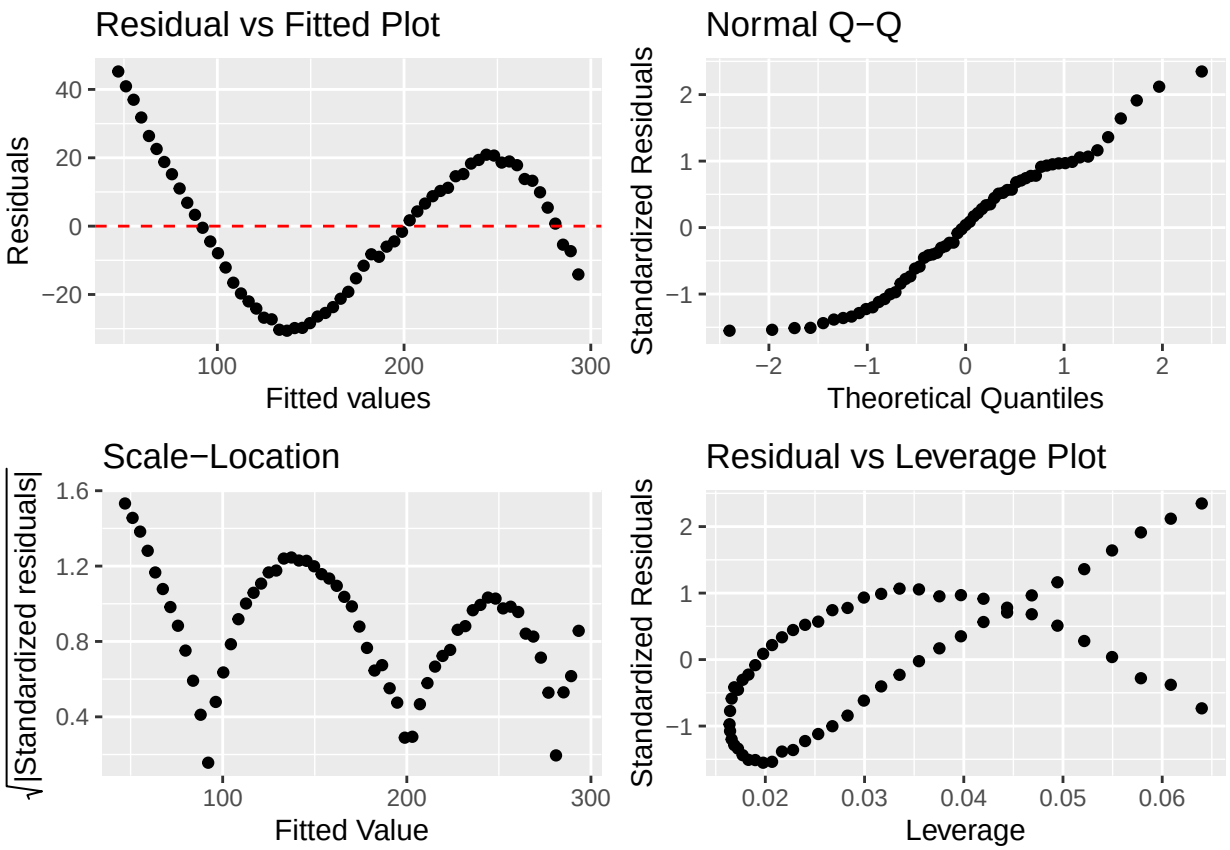
```
sl<-ggplot(fit, aes(.fitted, sqrt(abs(.stdresid))))+geom_point(na.rm=TRUE)+xlab("Fitted Value")+
    ylab(expression(sqrt("|Standardized residuals|")))+ggtitle('Scale-Location')
sl
```



4. Residual Vs Leverage Plot

```
rl<-ggplot(fit, aes(.hat, .stdresid))+geom_point(na.rm=TRUE)+
    xlab("Leverage")+ylab("Standardized Residuals")+ggtitle("Residual vs Leverage Plot")
rl
```



#Problem 4 Multipanel plot

```
library(gridExtra)
grid.arrange(resvsfitplot,qq,sl,rl)
```