Homework 3

Samuel Soon

9/23/2021

Setup

```
knitr::opts_chunk$set(echo = TRUE)
library(data.table)
```

```
## Warning: package 'data.table' was built under R version 4.0.5
```

```
library(broom)
```

```
## Warning: package 'broom' was built under R version 4.0.5
```

```
library(gridExtra)
```

```
## Warning: package 'gridExtra' was built under R version 4.0.5
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
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)

fit.diags <- broom::augment(fit)
```

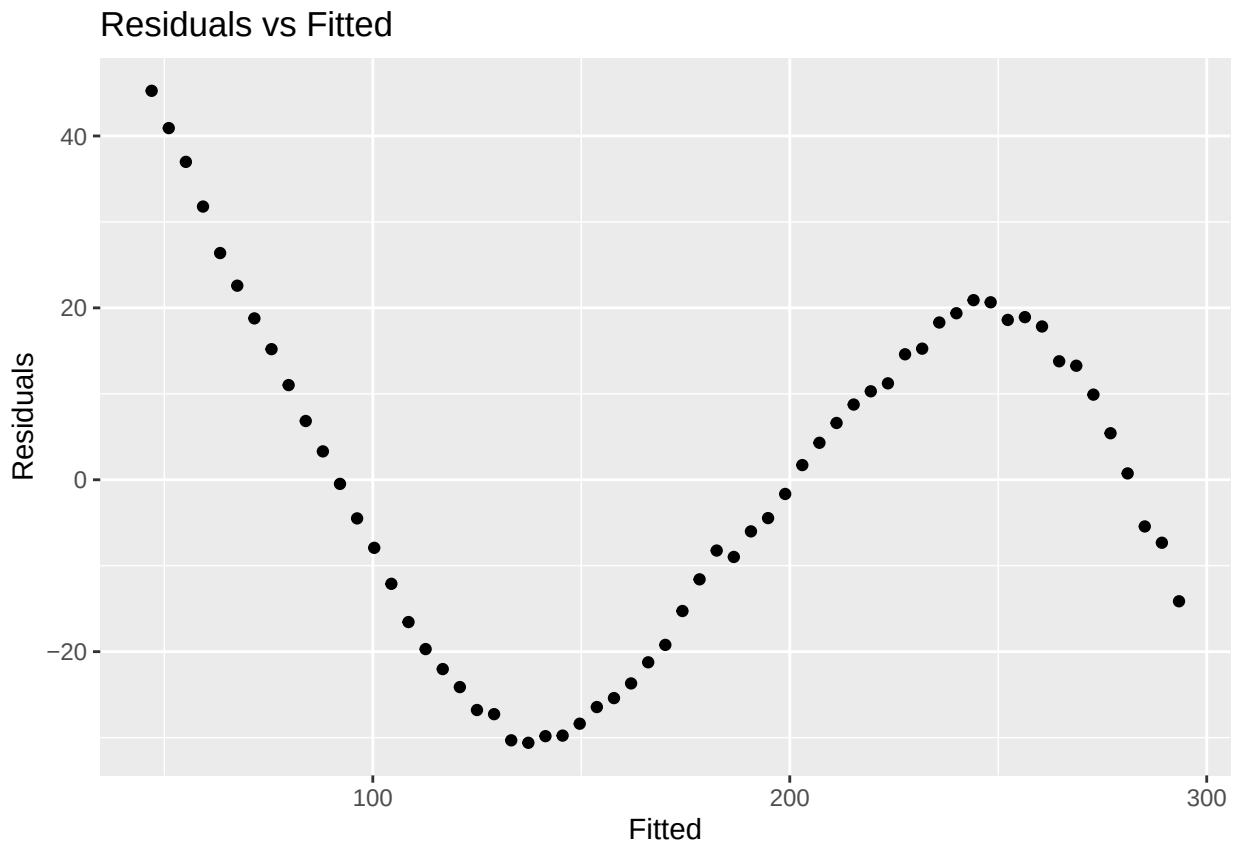
Problem 3

```
head(fit.diags)
```

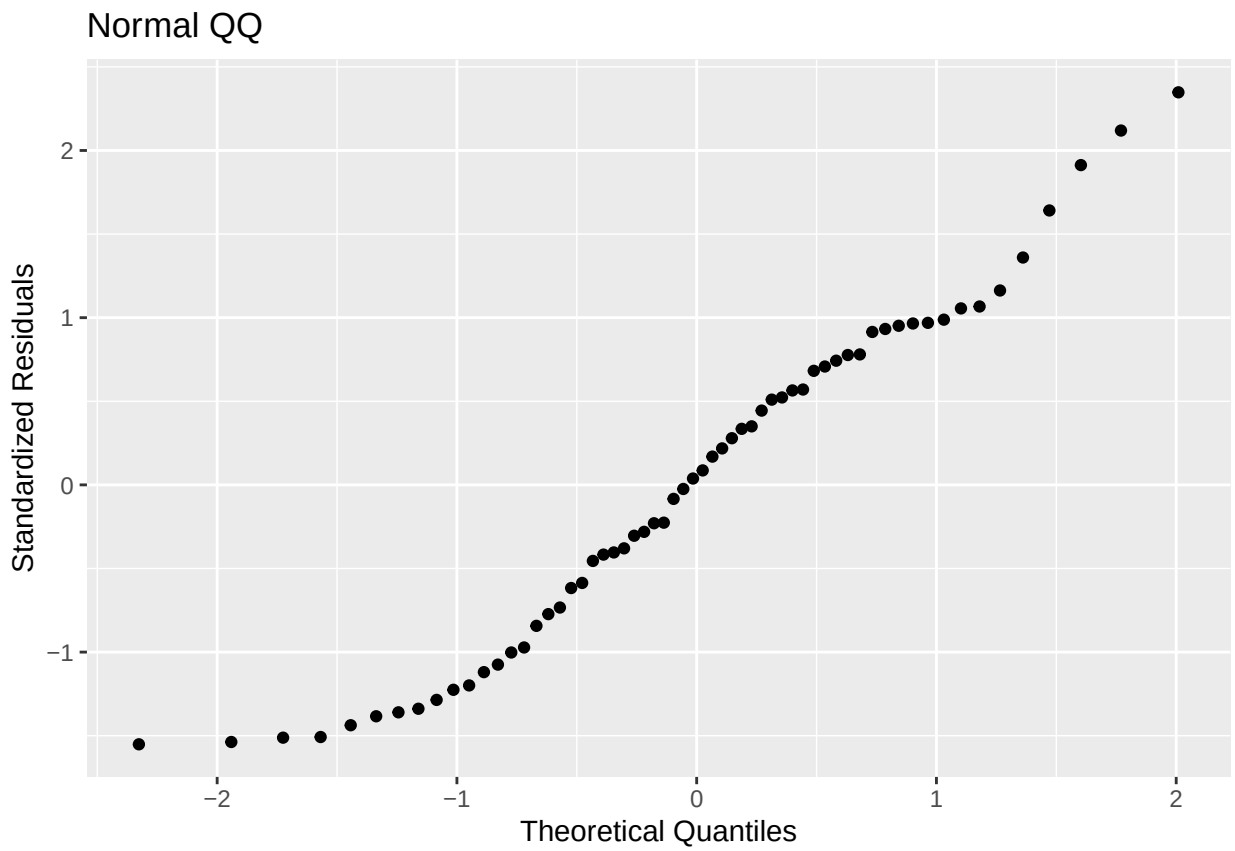
```
## # A tibble: 6 x 8
##   `Cumulative_number_for~ index .fitted  .resid   .hat .sigma .cooksd .std.resid
##                     <dbl> <int>   <dbl>   <dbl>  <dbl>  <dbl>   <dbl>      <dbl>
## 1                    279.    61    293. -14.1   0.0640   20.0 1.84e-2    -0.734
## 2                    282.    60    289.  -7.33  0.0609   20.1 4.67e-3    -0.380
## 3                    280.    59    285.  -5.43  0.0579   20.1 2.42e-3    -0.281
## 4                    282.    58    281.   0.738 0.0549   20.1 4.22e-5     0.0381
## 5                    282.    57    277.   5.41  0.0521   20.1 2.14e-3     0.279
## 6                    283.    56    273.   9.90  0.0494   20.0 6.76e-3     0.510
```

```
rvf <- ggplot(fit.diags, aes(x=.fitted, y=.resid))+
  geom_point() +
```

```
  labs(x="Fitted", y="Residuals", title="Residuals vs Fitted")
rvf
```

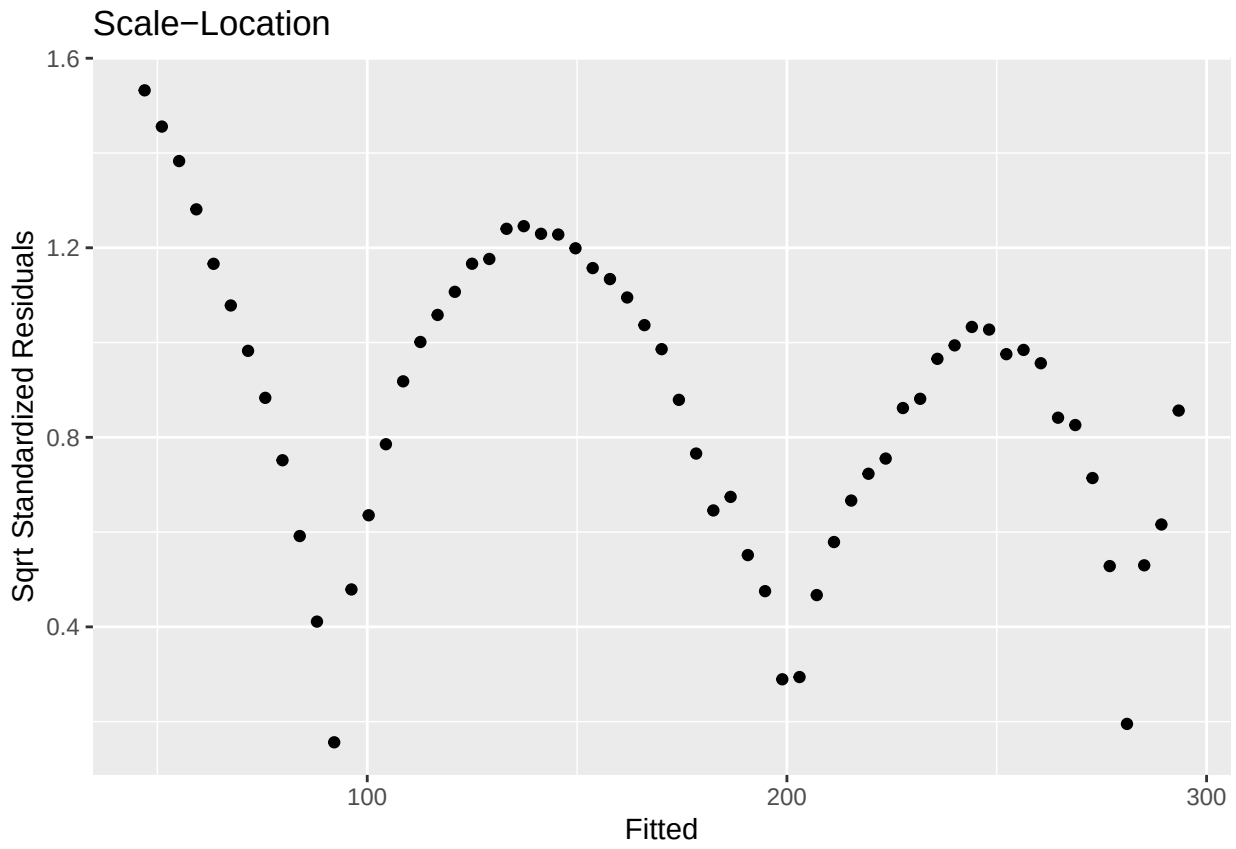



```
qq <- ggplot(fit.diags, aes(x=qnorm(seq(0.01,0.99,1/(nrow(fit.diags) + 1))),
                            y=.std.resid[order(.std.resid)])) +
  geom_point() +
  labs(x="Theoretical Quantiles", y="Standardized Residuals", title="Normal QQ")
qq
```

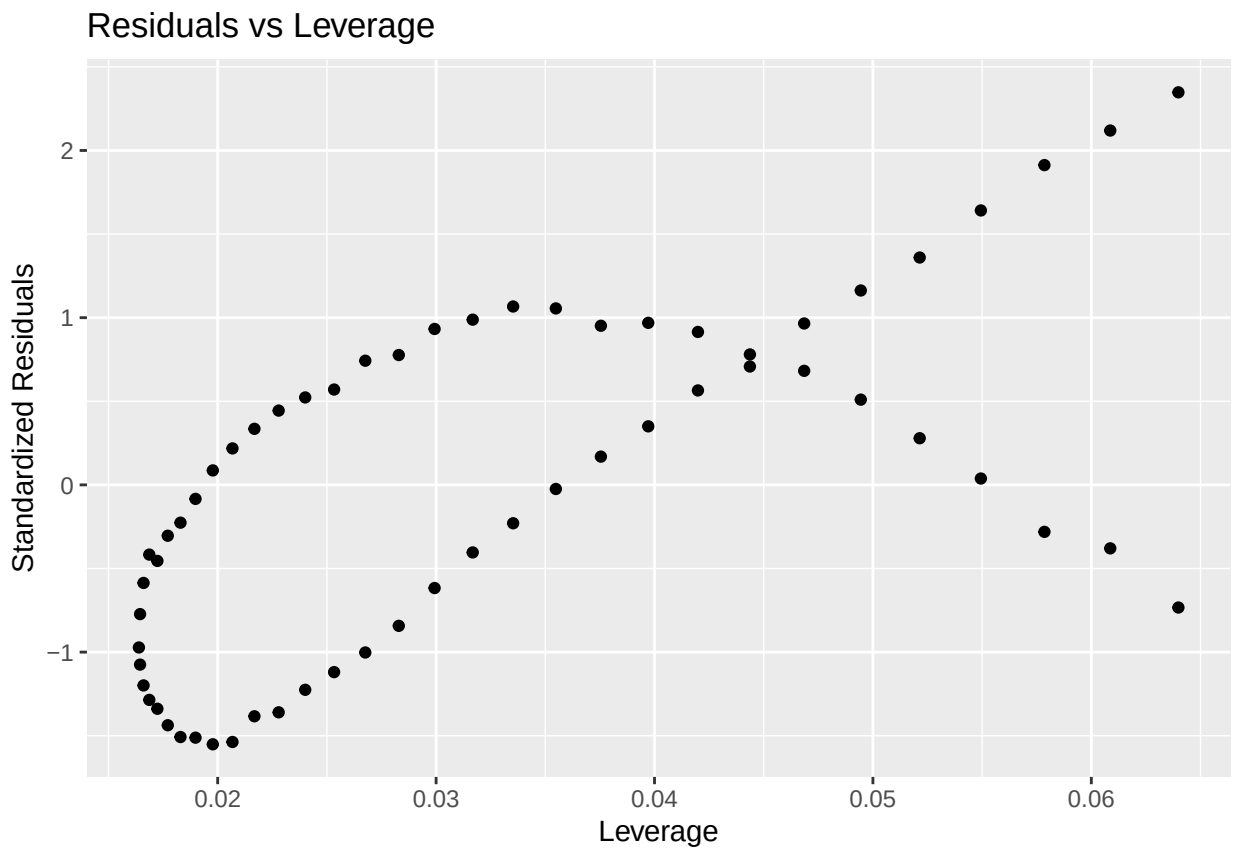



```
sl <- ggplot(fit.diags, aes(x=.fitted, y=sqrt(abs(.std.resid))))+
  geom_point() +
  labs(x="Fitted", y="Sqrt Standardized Residuals", title="Scale-Location")

sl
```

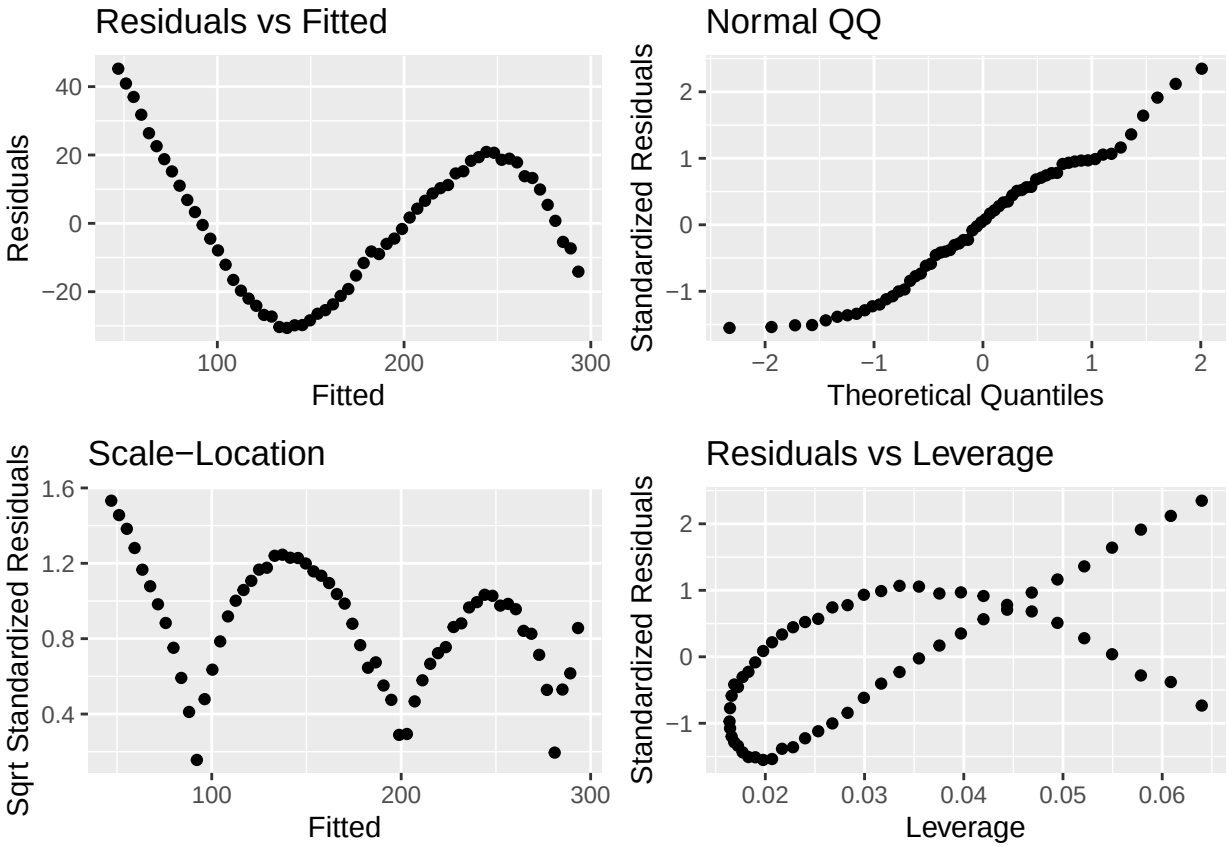



```
rvl <- ggplot(fit.diags, aes(x=.hat, y=.std.resid))+
  geom_point() +
  labs(x="Leverage", y="Standardized Residuals", title="Residuals vs Leverage")
rvl
```



Problem 4

```
grid.arrange(rvf, qq, sl, rvl, ncol = 2, nrow = 2)
```


Residuals vs Fitted
Residuals
Fitted
Normal QQ
Standardized Residuals
Theoretical Quantiles
Scale−Location
Sqrt Standardized Residuals
Fitted
Residuals vs Leverage
Standardized Residuals
Leverage