

stat 5014 Homework 3

Mengkun Chen (mengkun21@vt.edu)

Problem 3

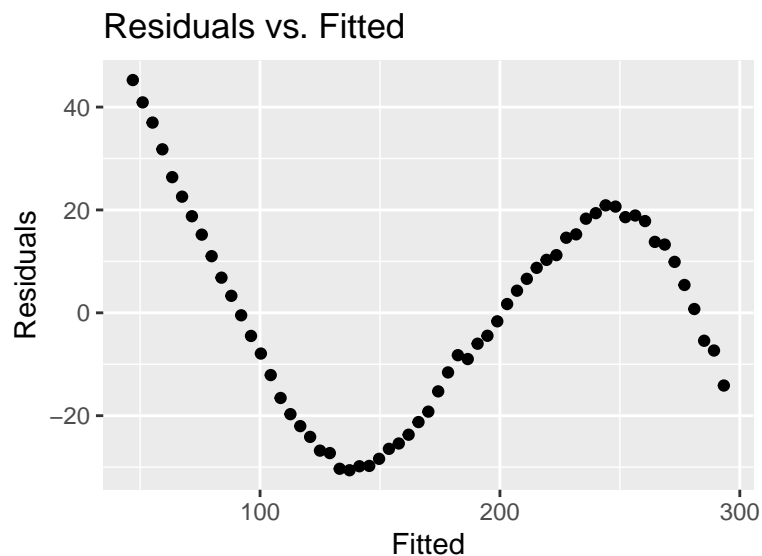
```
library(data.table)
library(broom)

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)

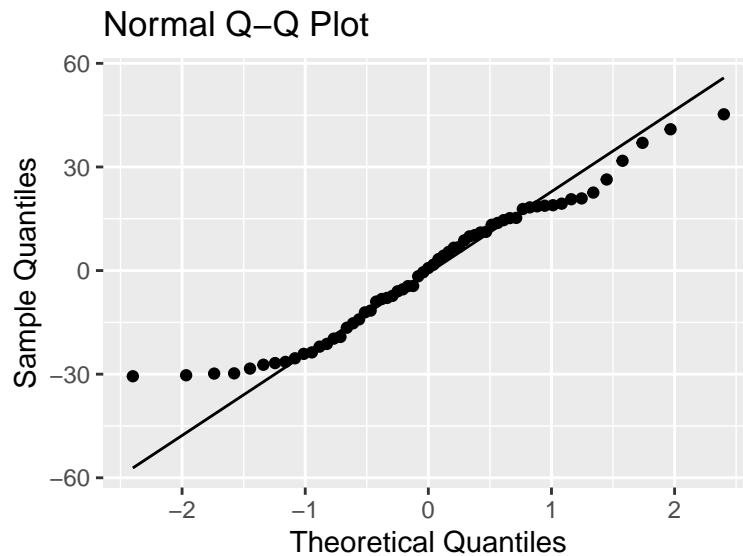
## augment the data
fit.diags<-augment(fit)
```

```
library(ggplot2)

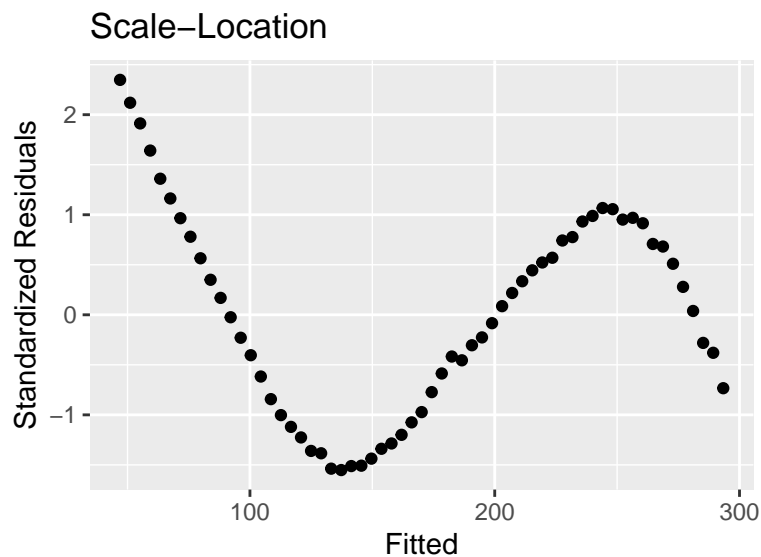
## residuals vs. fitted
ggplot(data = fit.diags, mapping = aes(x = .fitted, y = .resid)) +
  geom_point() +
  labs(title = "Residuals vs. Fitted", y = "Residuals", x = "Fitted")
```



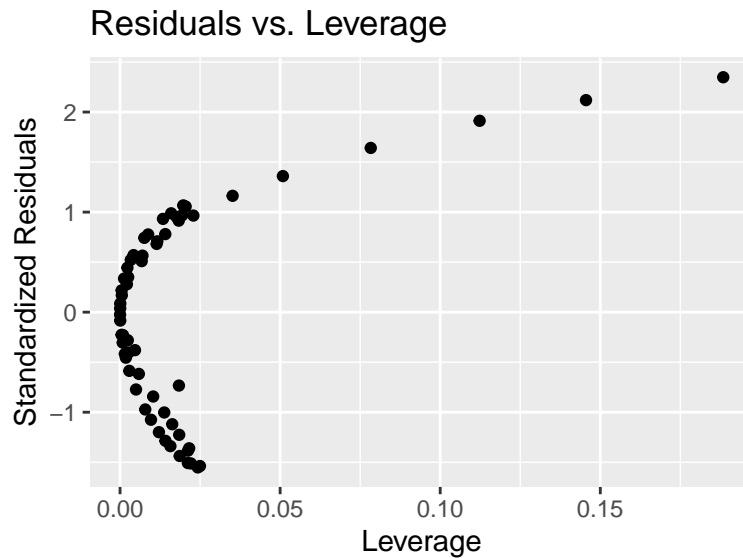
```
## normal Q-Q
ggplot(data = fit.diags, mapping = aes(sample = .resid)) +
  geom_qq() +
  geom_qq_line() +
  labs(title = "Normal Q-Q Plot", y = "Sample Quantiles", x = "Theoretical Quantiles")
```



```
## scale-location
ggplot(data = fit.diags, mapping = aes(x = .fitted, y = .std.resid)) +
  geom_point() +
  labs(title = "Scale-Location", y = "Standardized Residuals", x = "Fitted")
```



```
## residuals vs. leverage
ggplot(data = fit.diags, mapping = aes(x = .cooks, y = .std.resid)) +
  geom_point() +
  labs(title = "Residuals vs. Leverage", x = "Leverage", y = "Standardized Residuals")
```



Problem 4

```
library(ggpubr)

## residuals vs. fitted
rf<-ggplot(data = fit.diags, mapping = aes(x = .fitted, y = .resid)) +
  geom_point() +
  labs(title = "Residuals vs. Fitted", y = "Residuals", x = "Fitted")

## normal Q-Q
qq<-ggplot(data = fit.diags, mapping = aes(sample = .resid)) +
  geom_qq() +
  geom_qq_line() +
  labs(title = "Normal Q-Q Plot", y = "Sample Quantiles", x = "Theoretical Quantiles")

## scale-location
sl<-ggplot(data = fit.diags, mapping = aes(x = .fitted, y = .std.resid)) +
  geom_point() +
  labs(title = "Scale-Location", y = "Standardized Residuals", x = "Fitted")

## residuals vs. leverage
rl<-ggplot(data = fit.diags, mapping = aes(x = .cooks, y = .std.resid)) +
  geom_point() +
  labs(title = "Residuals vs. Leverage", x = "Leverage", y = "Standardized Residuals")

ggarrange(rf, qq, sl, rl, ncol = 2, nrow = 2)
```

