HW3_arflowers

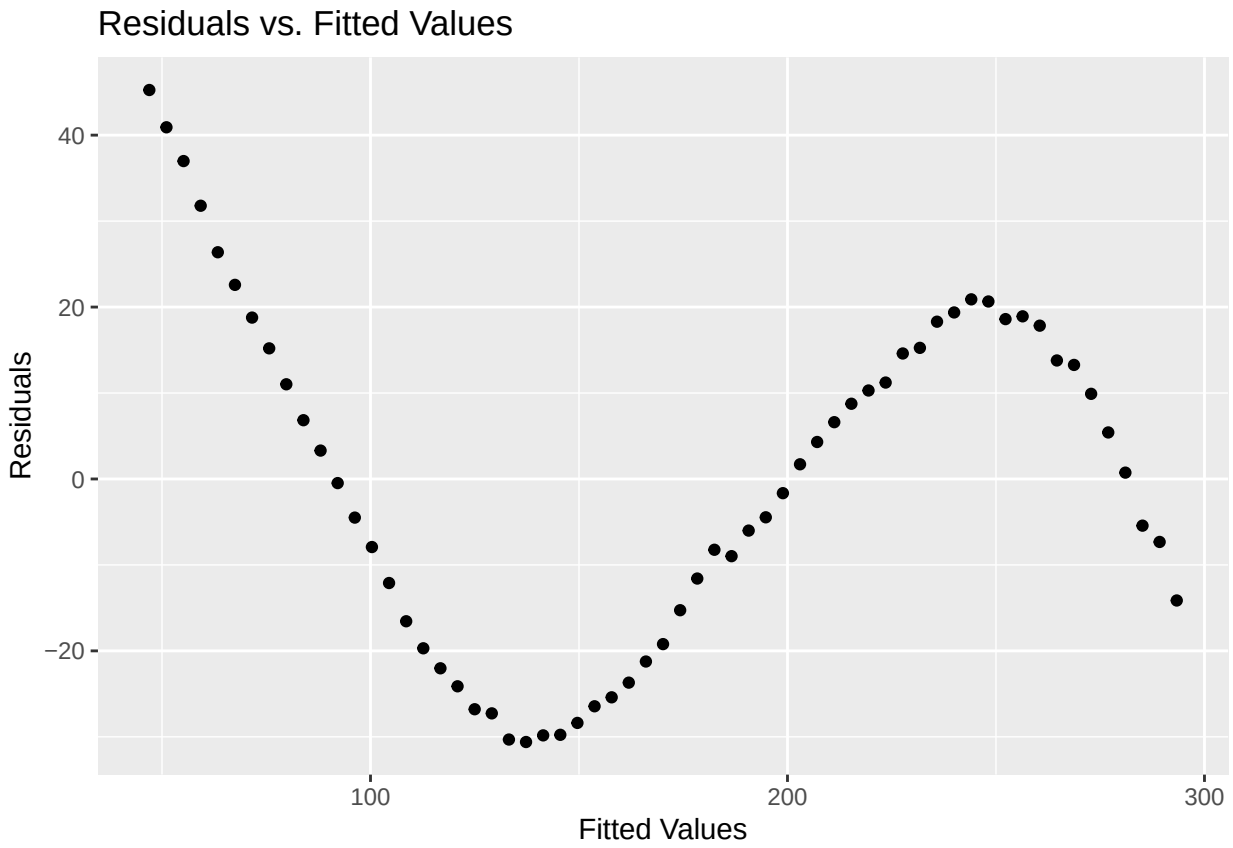
Anna Flowers

9/24/2021

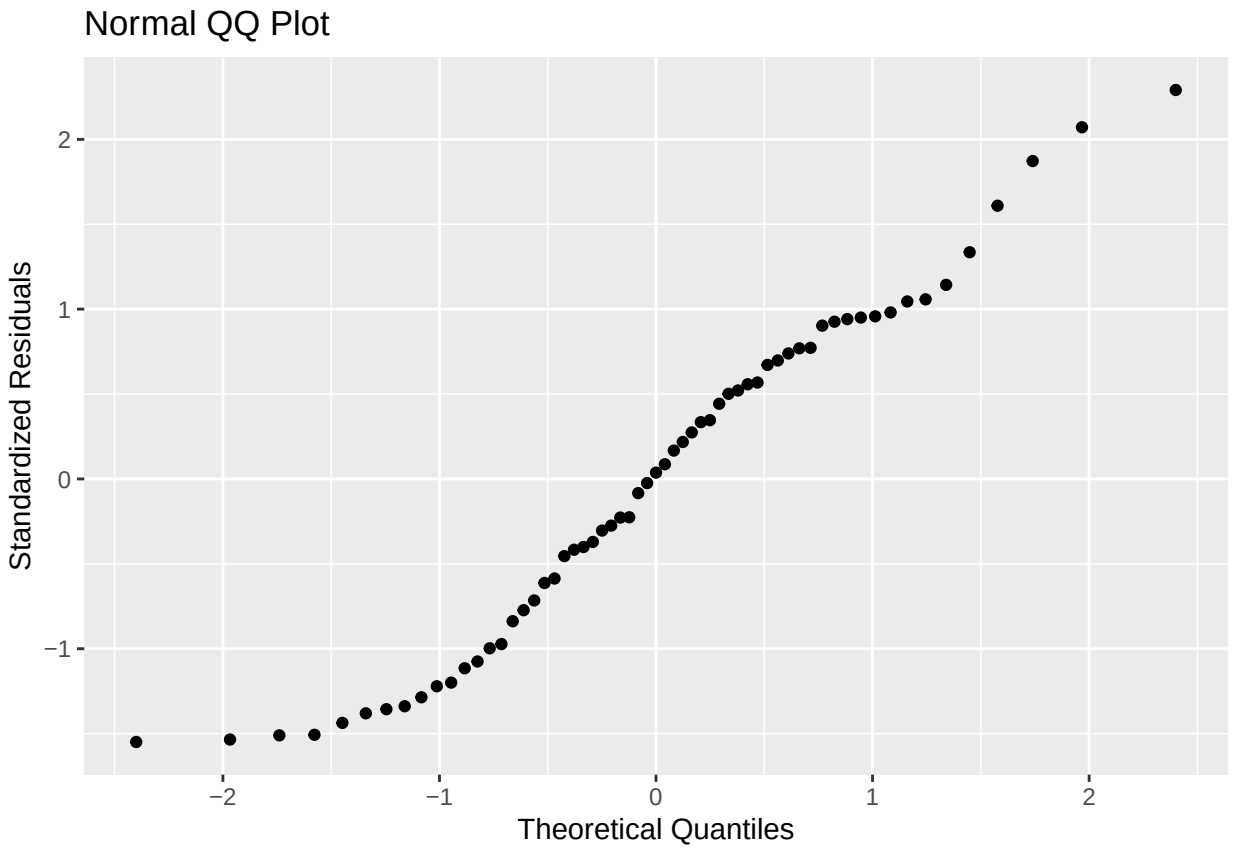
Homework 3

Problem 3

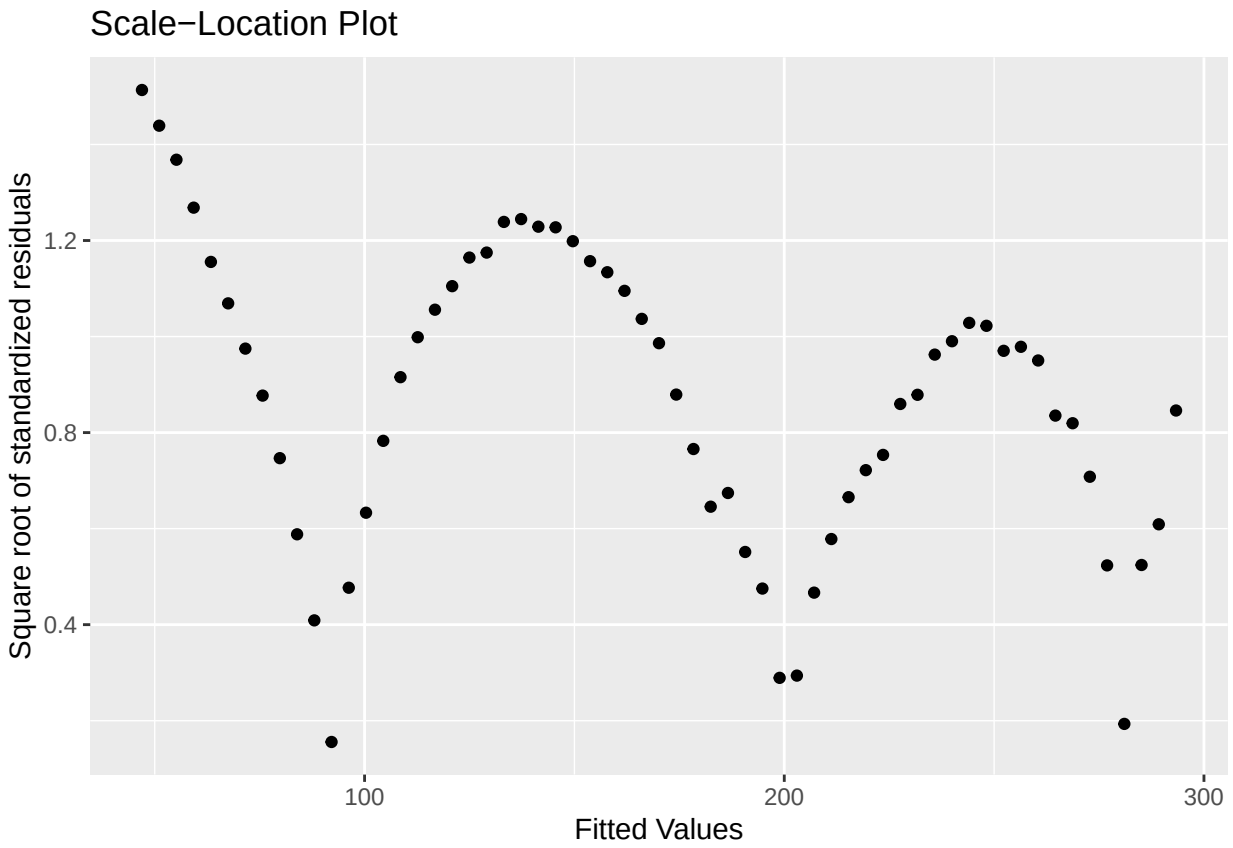
```
#install.packages('data.table')
library(data.table)
covid_raw <- fread("https://opendata.ecdc.europa.eu/covid19/casedistribution/csv")
us <- covid_raw[covid_raw$countriesAndTerritories == 'United_States_of_America',]
us_filtered <- us[us$month %in% c(6:7),]
us_filtered$index <- rev(1:dim(us_filtered)[1])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

## augment the data as previous
#install.packages("broom")
fit.diags <- broom::augment(fit)

library(ggplot2)
rsd_vs_fitted <- ggplot() + geom_point(aes(x=fit$fitted.values,y=fit$residuals)) + ggtitle("Residuals vs
rsd_vs_fitted
```



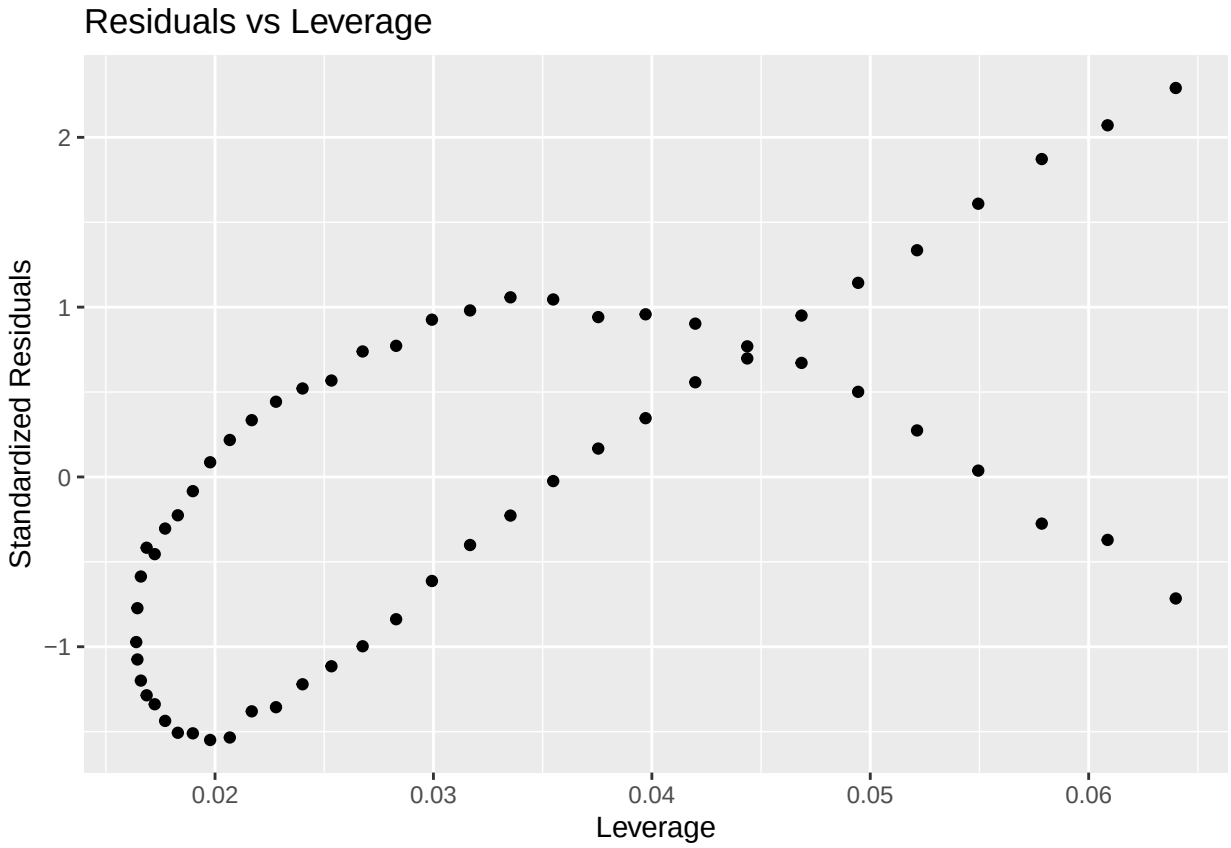

```
sorted <- sort(fit$residuals)
std_sorted <- (sorted - mean(sorted))/sd(sorted)
quantiles <- (1:length(sorted) - 0.5)/ length(sorted)
theoretical<- qnorm(quantiles)
qq<- ggplot() + geom_point(aes(x=theoretical, y= std_sorted)) + ggtitle("Normal QQ Plot") + labs(x="The
qq
```




```
std_resid2 <- (abs(fit$residuals - mean(fit$residuals)))/sd(fit$residuals)
scale_location <- ggplot() + geom_point(aes(x=fit$fitted.values,y=sqrt(std_resid2))) + ggtitle("Scale-L
scale_location
```

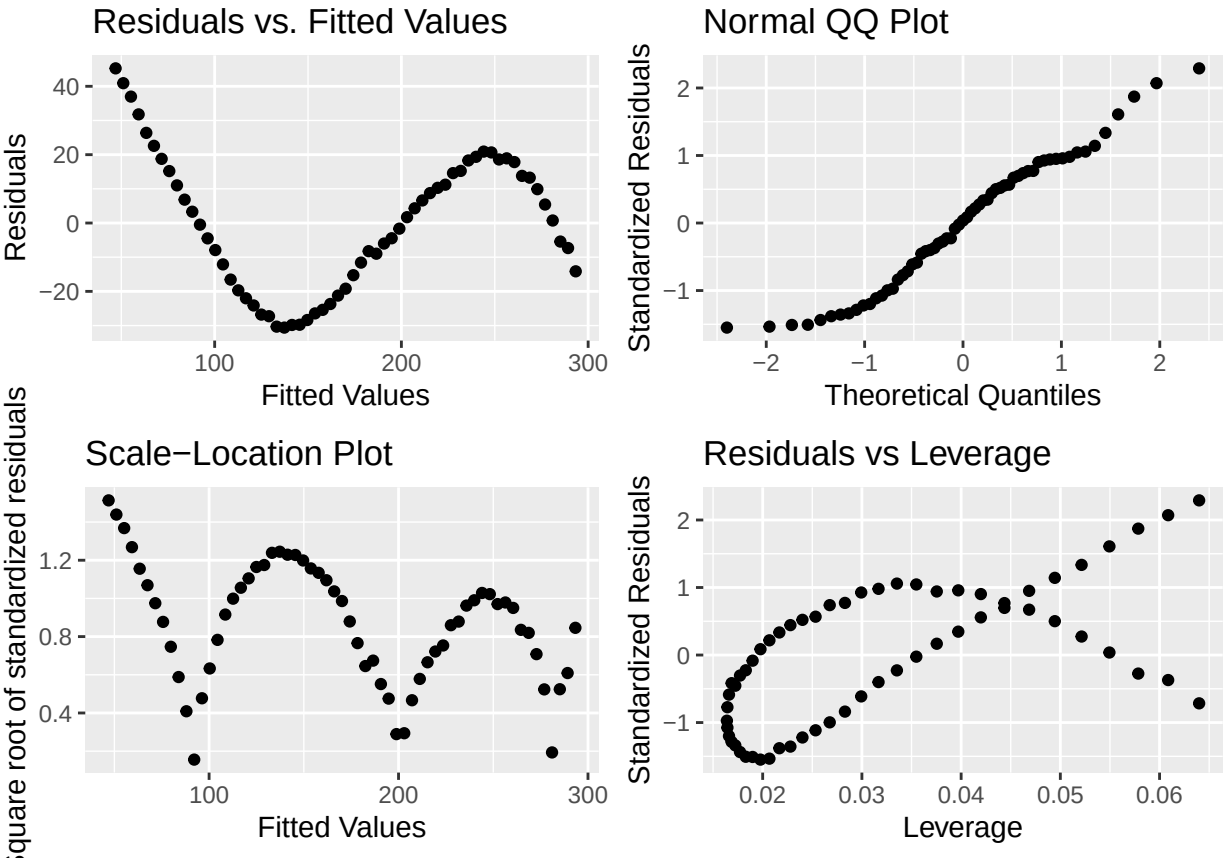



```
std_resid <- (fit$residuals - mean(fit$residuals))/sd(fit$residuals)
resid_vs_leverage <- ggplot() + geom_point(aes(x=hatvalues(fit), y= std_resid)) + ggtitle("Residuals vs
resid_vs_leverage
```



Problem 4

```
library(ggpubr)
ggarrange(rsd_vs_fitted,qq,scale_location, resid_vs_leverage, ncol=2, nrow =2)
```


Residuals vs. Fitted Values
Residuals
Fitted Values
Normal QQ Plot
Standardized Residuals
Theoretical Quantiles
Scale−Location Plot
Fitted Values
Residuals vs Leverage
Standardized Residuals
Leverage