Individual Assignment3 - Appendix

Data source:

Reference:

[1] https://www.statisticssolutions.com/assumptions-of-linear-regression

[2] "Discovering Statistics Using R" by Andy Field, Jeremy Miles, Zoe Field.

Appendix -1

```
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.0.4

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(tidyr)

## Warning: package 'tidyr' was built under R version 4.0.4

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.0.4

dataset <- read.csv(path)
```

Appendix -2

```
dataset<- dataset %>% filter(Company=="BANK OF AMERICA, NATIONAL
ASSOCIATION")
dataset <- na.omit(dataset)
```

```r
dataset<-dataset %>% separate(Date.received,
c("Split_Month","Split_Day","Split_Year"),remove=FALSE)
dataset$Split_Year<- as.integer(dataset$Split_Year)
dataset$Split_Month <- as.integer(dataset$Split_Month)
dataset$Split_Year <- (dataset$Split_Year + 2000)

all_years<-c(2011,2012,2013,2014,2015,2016,2017,2018,2019,2020,2021)
all_months<-c(1,2,3,4,5,6,7,8,9,10,11,12)
all_quaters<-c(1,2,3,4)
```

Appendix -3

```
##      year month time complains
## 1    2012     3    1       183
## 2    2012     4    2       218
## 3    2012     5    3       279
## 4    2012     6    4       217
## 5    2012     7    5       230
## 6    2012     8    6       204
## 7    2012     9    7       167
## 8    2012    10    8       225
## 9    2012    11    9       162
## 10   2012    12   10       152
## 11   2013     1   11       205
## 12   2013     2   12       171
## 13   2013     3   13       173
## 14   2013     4   14       169
## 15   2013     5   15       184
## 16   2013     6   16       175
## 17   2013     7   17       206
## 18   2013     8   18       186
## 19   2013     9   19       200
## 20   2013    10   20       209
## 21   2013    11   21       157
## 22   2013    12   22       160
## 23   2014     1   23       197
## 24   2014     2   24       205
## 25   2014     3   25       203
## 26   2014     4   26       228
## 27   2014     5   27       214
## 28   2014     6   28       195
## 29   2014     7   29       193
## 30   2014     8   30       201
## 31   2014     9   31       190
## 32   2014    10   32       219
## 33   2014    11   33       165
## 34   2014    12   34       217
## 35   2015     1   35       213
## 36   2015     2   36       178
## 37   2015     3   37       211
## 38   2015     4   38       226
```

```
## 39   2015    5   39      253
## 40   2015    6   40      283
## 41   2015    7   41      287
## 42   2015    8   42      278
## 43   2015    9   43      262
## 44   2015   10   44      258
## 45   2015   11   45      235
## 46   2015   12   46      241
## 47   2016    1   47      278
## 48   2016    2   48      212
## 49   2016    3   49      270
## 50   2016    4   50      248
## 51   2016    5   51      240
## 52   2016    6   52      255
## 53   2016    7   53      256
## 54   2016    8   54      313
## 55   2016    9   55      326
## 56   2016   10   56      292
## 57   2016   11   57      284
## 58   2016   12   58      270
## 59   2017    1   59      288
## 60   2017    2   60      279
## 61   2017    3   61      316
## 62   2017    4   62      259
## 63   2017    5   63      246
## 64   2017    6   64      268
## 65   2017    7   65      275
## 66   2017    8   66      260
## 67   2017    9   67      259
## 68   2017   10   68      253
## 69   2017   11   69      308
## 70   2017   12   70      270
## 71   2018    1   71      276
## 72   2018    2   72      272
## 73   2018    3   73      313
## 74   2018    4   74      373
## 75   2018    5   75      305
## 76   2018    6   76      291
## 77   2018    7   77      319
## 78   2018    8   78      284
## 79   2018    9   79      292
## 80   2018   10   80      278
## 81   2018   11   81      230
## 82   2018   12   82      190
## 83   2019    1   83      223
## 84   2019    2   84      186
## 85   2019    3   85      245
## 86   2019    4   86      279
## 87   2019    5   87      293
## 88   2019    6   88      287
```

```
## 89   2019    7   89       309
## 90   2019    8   90       347
## 91   2019    9   91       294
## 92   2019   10   92       295
## 93   2019   11   93       268
## 94   2019   12   94       294
## 95   2020    1   95       300
## 96   2020    2   96       261
## 97   2020    3   97       273
## 98   2020    4   98       306
## 99   2020    5   99       270
## 100 2020    6  100       290
## 101 2020    7  101       284
## 102 2020    8  102       255
## 103 2020    9  103       289
## 104 2020   10  104       317
## 105 2020   11  105       288
## 106 2020   12  106       369
## 107 2021    1  107       388
## 108 2021    2  108       338
## 109 2021    3  109        24
```

Appendix -3

```
##      year     n
## 1   2012 2037
## 2   2013 2195
## 3   2014 2427
## 4   2015 2925
## 5   2016 3244
## 6   2017 3281
## 7   2018 3423
## 8   2019 3320
## 9   2020 3502
## 10 2021  750
```

Appendix -4

one important assumption of the linear regression is that the relationship between the outcome and predictor variables is linear and additive. Here quarter and no of complains are strongly & positively related. Correlation can be given as:

```
cor( df$no_of_quarter,df$complains)*100
```

```
## [1] 75.62481
```

Appendix -5

```
##
## Call:
## lm(formula = complains ~ no_of_quarter, data = df)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -185.647  -61.632    5.355   66.851  158.355
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    568.956     30.435   18.69  < 2e-16 ***
## no_of_quarter    9.668      1.434    6.74 9.59e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 89.41 on 34 degrees of freedom
## Multiple R-squared:  0.5719, Adjusted R-squared:  0.5593
## F-statistic: 45.42 on 1 and 34 DF,  p-value: 9.589e-08
```

Appendix -6

```
##                   2.5 %    97.5 %
## (Intercept)   507.105081 630.80603
## no_of_quarter   6.752448  12.58269
```
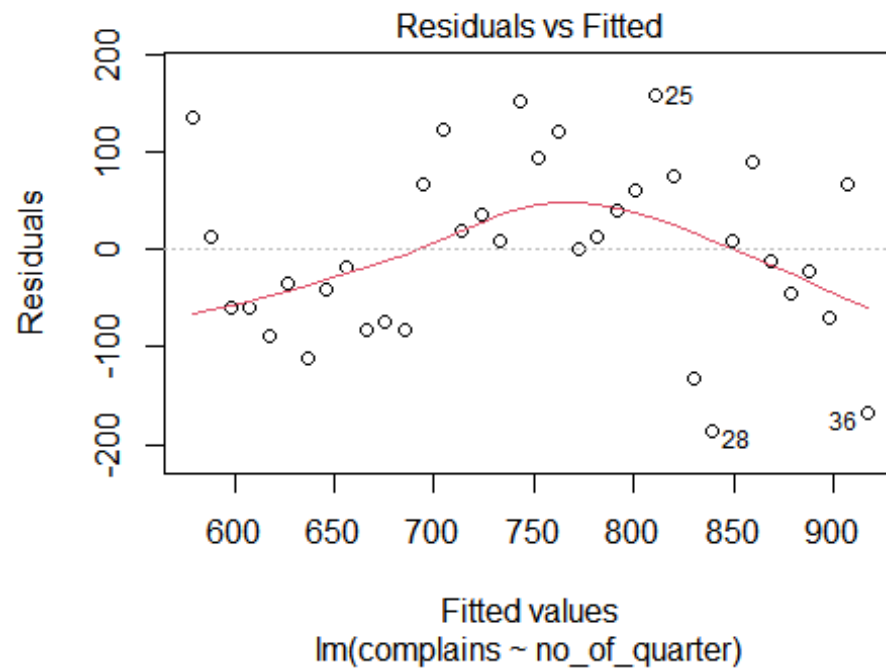
Appendix -7

```
sigma(regressor)*100/mean(df$complains)

## [1] 11.95604
```
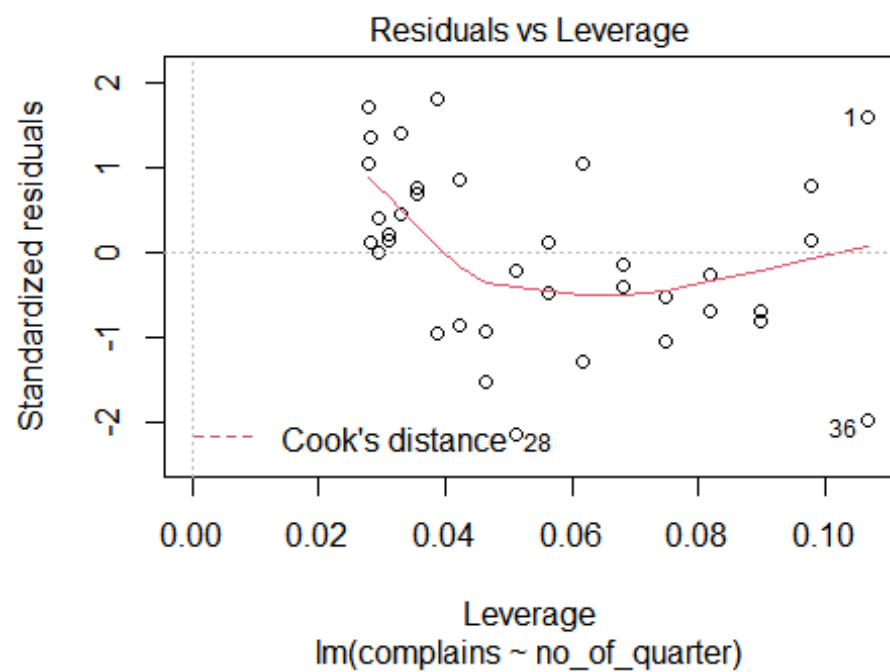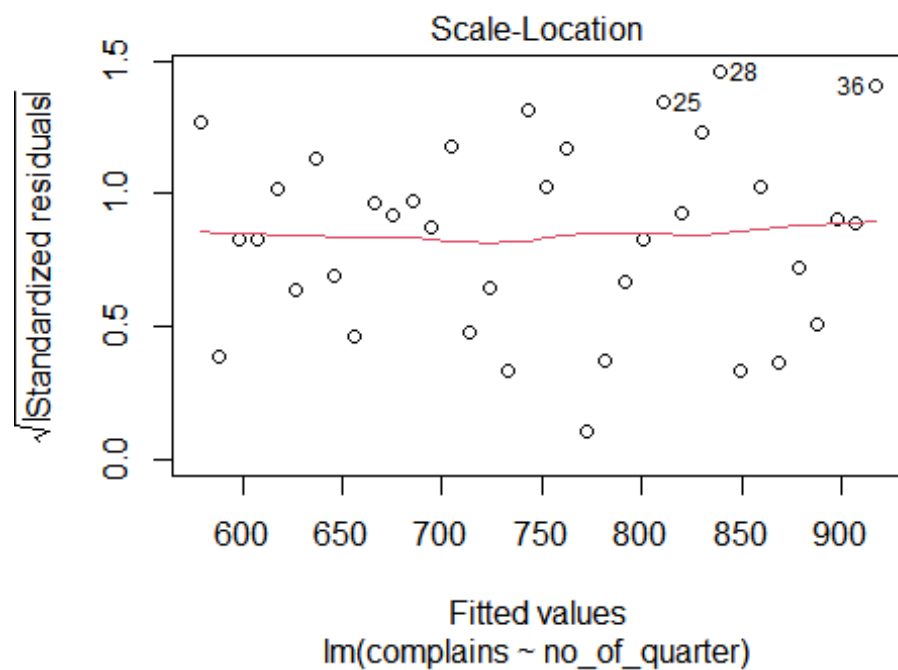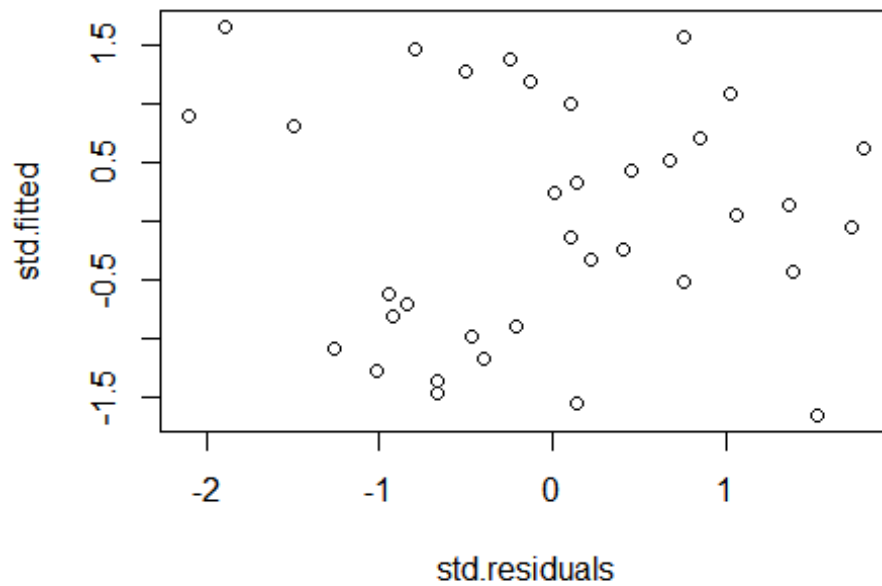
Appendix -8

complains ~ time

Appendix -9

Residuals vs Fitted

Fitted values
lm(complains ~ no_of_quarter)

Normal Q-Q

Theoretical Quantiles
lm(complains ~ no_of_quarter)

Scale-Location

√|Standardized residuals|

Fitted values
lm(complains ~ no_of_quarter)

Residuals vs Leverage

Standardized residuals

Cook's distance

Leverage
lm(complains ~ no_of_quarter)
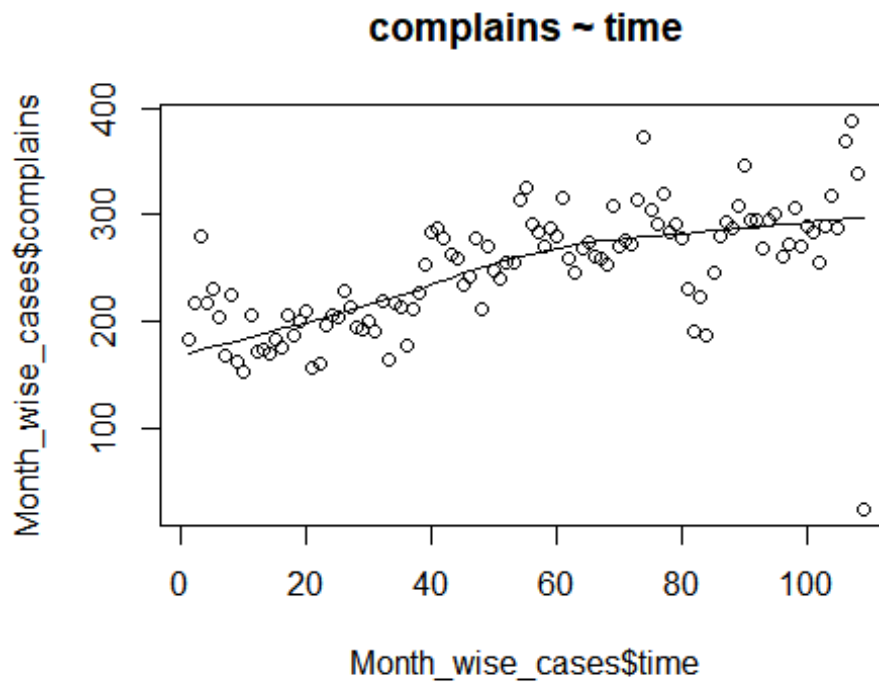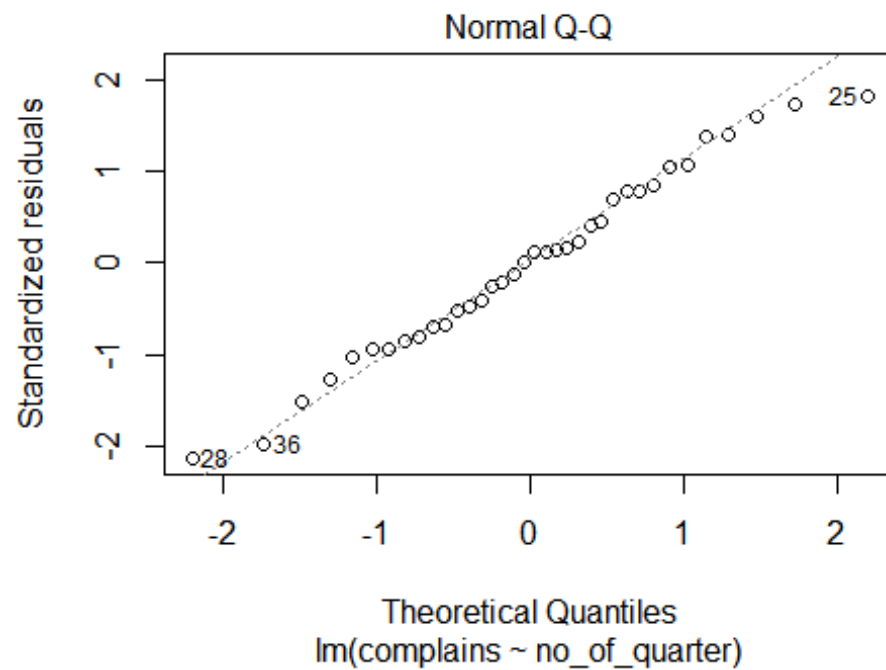
Appendix -10

Appendix -11
```
## [1] 1
```

Appendix -12
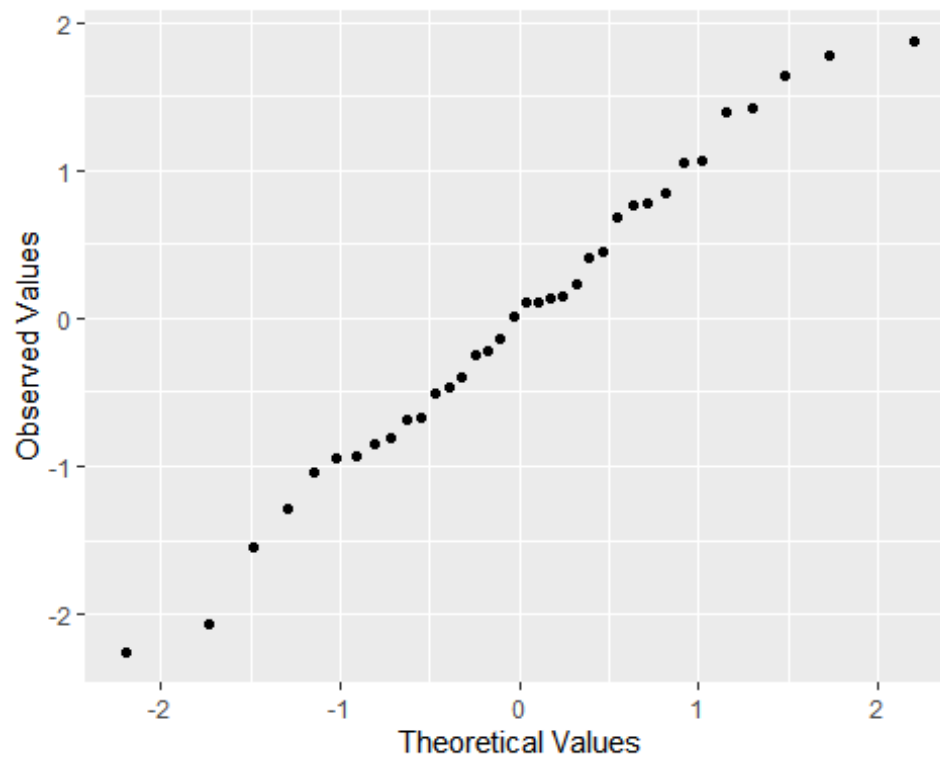
Assumption of normality

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
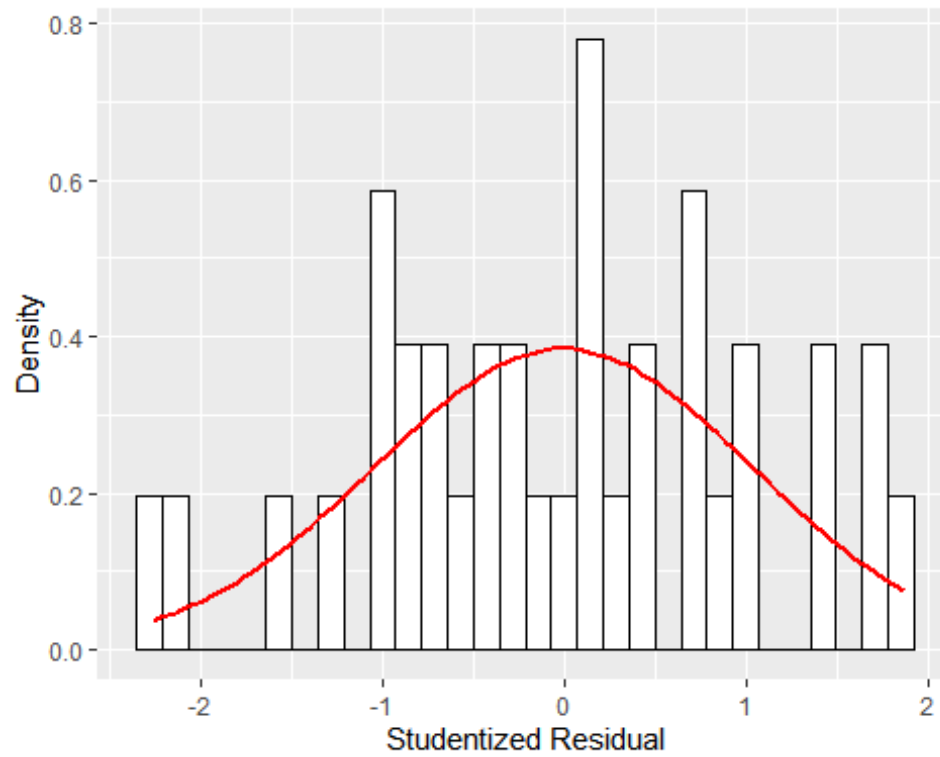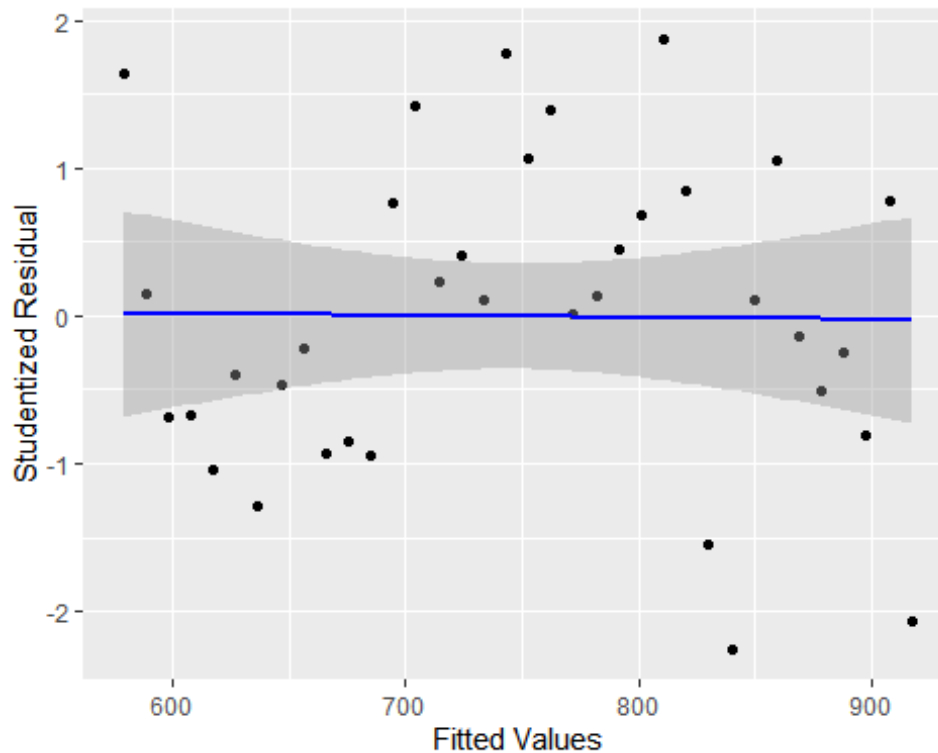
Appendix -13

Warning: `stat` is deprecated

Appendix -14

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

```
##        1        2        3        4        5        6        7
## 926.6556 936.3231 945.9907 955.6583 965.3258 974.9934 984.6610
```

```
## [1] 3880.638
```

```
## Warning: package 'car' was built under R version 4.0.4

## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode

##  lag Autocorrelation D-W Statistic p-value
##    1       0.3428544      1.144262   0.002
##  Alternative hypothesis: rho != 0
```

```
## Warning: package 'boot' was built under R version 4.0.4
```

```
##
## Attaching package: 'boot'

## The following object is masked from 'package:car':
##
##     logit

## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 2000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = results, type = "bca", index = 1)
##
## Intervals :
## Level       BCa
## 95%    (521.6, 635.2 )
## Calculations and Intervals on Original Scale

## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 2000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = results, type = "bca", index = 2)
##
## Intervals :
## Level        BCa
## 95%    ( 6.566, 12.433 )
## Calculations and Intervals on Original Scale
```