Case Study 2 R Supplement, Stat 230

Helen Moses

1(Visualize the Data).

```r
#Scatterplot of all pairwise combinations in my full model
wb |>
  ggpairs(columns = c("gdp_growth", "inflation", "internet", "barter", "devel_assist", "pop_growth"))
```

```
## Warning: Removed 2 rows containing non-finite values (stat_density).
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 2 rows containing missing values
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 10 rows containing missing values
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 2 rows containing missing values
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 3 rows containing missing values
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 2 rows containing missing values
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```

```
## Warning: Removed 2 rows containing non-finite values (stat_density).
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 10 rows containing missing values
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 2 rows containing missing values
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 3 rows containing missing values
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 2 rows containing missing values
```

```
## Warning: Removed 10 rows containing missing values (geom_point).
## Removed 10 rows containing missing values (geom_point).
```

```
## Warning: Removed 9 rows containing non-finite values (stat_density).
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 10 rows containing missing values
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 10 rows containing missing values
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 9 rows containing missing values
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
## Removed 2 rows containing missing values (geom_point).
```

```
## Warning: Removed 10 rows containing missing values (geom_point).
```

```
## Warning: Removed 1 rows containing non-finite values (stat_density).
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 2 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 2 rows containing missing values

## Warning: Removed 3 rows containing missing values (geom_point).
## Removed 3 rows containing missing values (geom_point).

## Warning: Removed 10 rows containing missing values (geom_point).

## Warning: Removed 2 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing non-finite values (stat_density).

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 2 rows containing missing values

## Warning: Removed 2 rows containing missing values (geom_point).
## Removed 2 rows containing missing values (geom_point).

## Warning: Removed 9 rows containing missing values (geom_point).

## Warning: Removed 2 rows containing missing values (geom_point).
## Removed 2 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing non-finite values (stat_density).
```
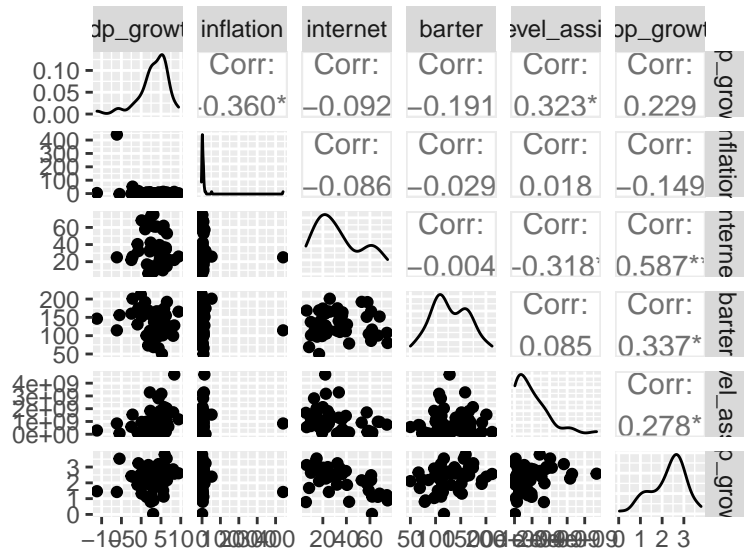

To get a good picture of what was happening in my full model, I wanted to plot a scatterplot of all of the pairwise combinations of the data. Before evaluating the residuals, it is obvious that the inflation column needs a transformation as well as development assistance column.

```
#Adding a column to the data set that is the log of the inflation data
wb <- wb |>
  mutate(loginflation = log(inflation))
```

```
## Warning in log(inflation): NaNs produced
```

```
#Adding a column to the data set that is the log of the development assistance
wb <- wb |>
  mutate(logdevel_assist = log(devel_assist))
#Plotting the new scatterplot with the transformed data
```
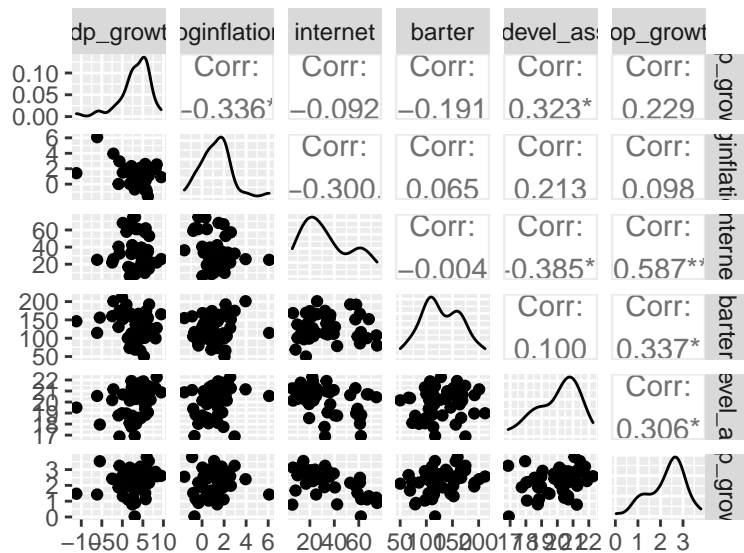
```
wb |>
  ggpairs(columns = c("gdp_growth", "loginflation", "internet", "barter", "logdevel_assist", "pop_growth
```

Warning: Removed 2 rows containing non-finite values (stat_density).

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 8 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 10 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 2 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 3 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 2 rows containing missing values

Warning: Removed 8 rows containing missing values (geom_point).

Warning: Removed 8 rows containing non-finite values (stat_density).

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 14 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 8 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 9 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 8 rows containing missing values

Warning: Removed 10 rows containing missing values (geom_point).

Warning: Removed 14 rows containing missing values (geom_point).

Warning: Removed 9 rows containing non-finite values (stat_density).

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 10 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 10 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 9 rows containing missing values

Warning: Removed 2 rows containing missing values (geom_point).

Warning: Removed 8 rows containing missing values (geom_point).

Warning: Removed 10 rows containing missing values (geom_point).

Warning: Removed 1 rows containing non-finite values (stat_density).

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 2 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 2 rows containing missing values

```
## Warning: Removed 3 rows containing missing values (geom_point).

## Warning: Removed 9 rows containing missing values (geom_point).

## Warning: Removed 10 rows containing missing values (geom_point).

## Warning: Removed 2 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing non-finite values (stat_density).

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 2 rows containing missing values

## Warning: Removed 2 rows containing missing values (geom_point).

## Warning: Removed 8 rows containing missing values (geom_point).

## Warning: Removed 9 rows containing missing values (geom_point).

## Warning: Removed 2 rows containing missing values (geom_point).
## Removed 2 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing non-finite values (stat_density).
```


Here all of the data looks much better, and there do not appear to be any obvious signs of multicollinearity. It is now possible to assess the model assumptions more in depth. *It is important to note

2 (Are Model Assumptions Satisfied).

```r
#Fitting the full model
wb_mlr <- lm(gdp_growth ~ loginflation + internet + barter + logdevel_assist + pop_growth, data = wb)
summary(wb_mlr)
```

```
##
## Call:
## lm(formula = gdp_growth ~ loginflation + internet + barter +
##     logdevel_assist + pop_growth, data = wb)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5349 -1.5381 -0.0007  1.6840  5.2611
##
```

```
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -10.34886    6.91480  -1.497  0.14399
## loginflation     -1.06509    0.30455  -3.497  0.00137 **
## internet          0.02671    0.03089   0.865  0.39345
## barter           -0.01715    0.01336  -1.284  0.20814
## logdevel_assist   0.65029    0.33407   1.947  0.06014 .
## pop_growth        1.56529    0.81003   1.932  0.06193 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.444 on 33 degrees of freedom
##   (15 observations deleted due to missingness)
## Multiple R-squared:  0.4078, Adjusted R-squared:  0.3181
## F-statistic: 4.545 on 5 and 33 DF,  p-value: 0.00291
```

The standard deviation for the gdp growth is a fair amount larger than the others.

```
#Adding a column to the data set that is the log of gdp_growth
wb <- wb |>
  mutate(loggdp_growth = log(gdp_growth))
```

```
## Warning in log(gdp_growth): NaNs produced
```
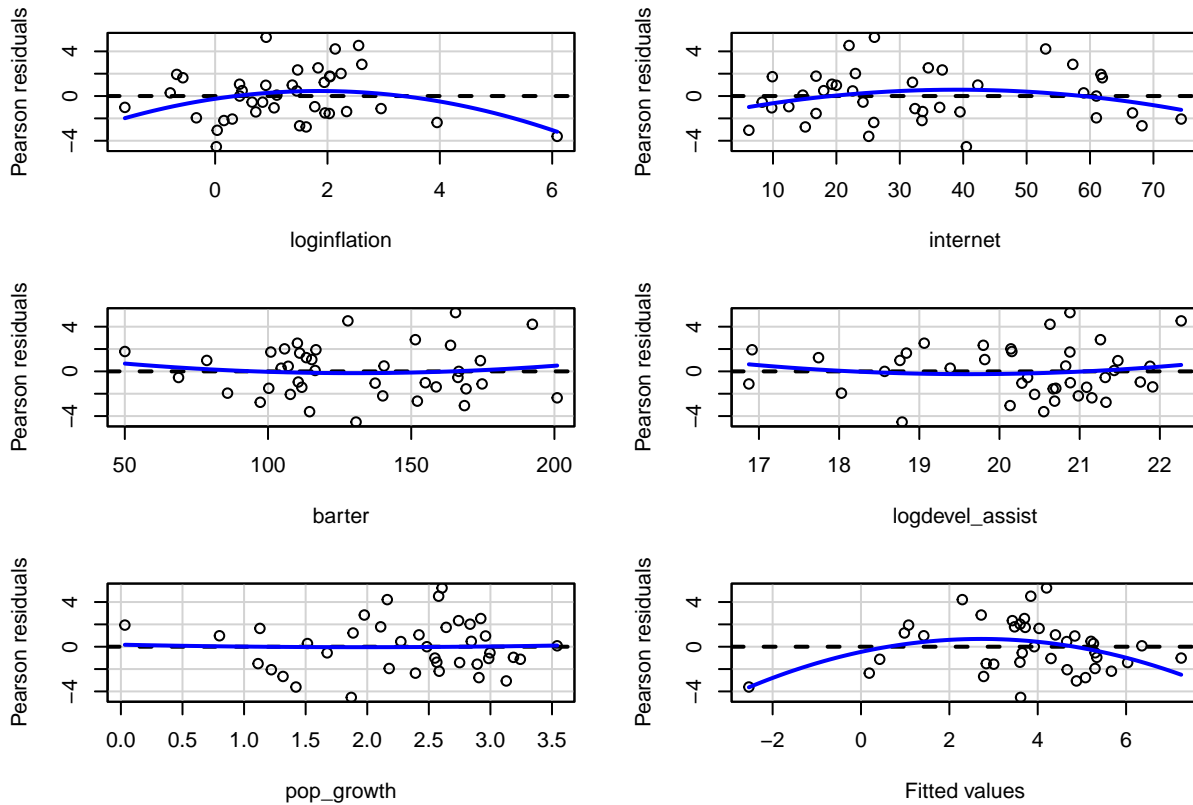
```
#Testing the new linear model
wb_mlr1 <- lm(loggdp_growth ~ loginflation + internet + barter + logdevel_assist + pop_growth, data = wb
summary(wb_mlr1)
```
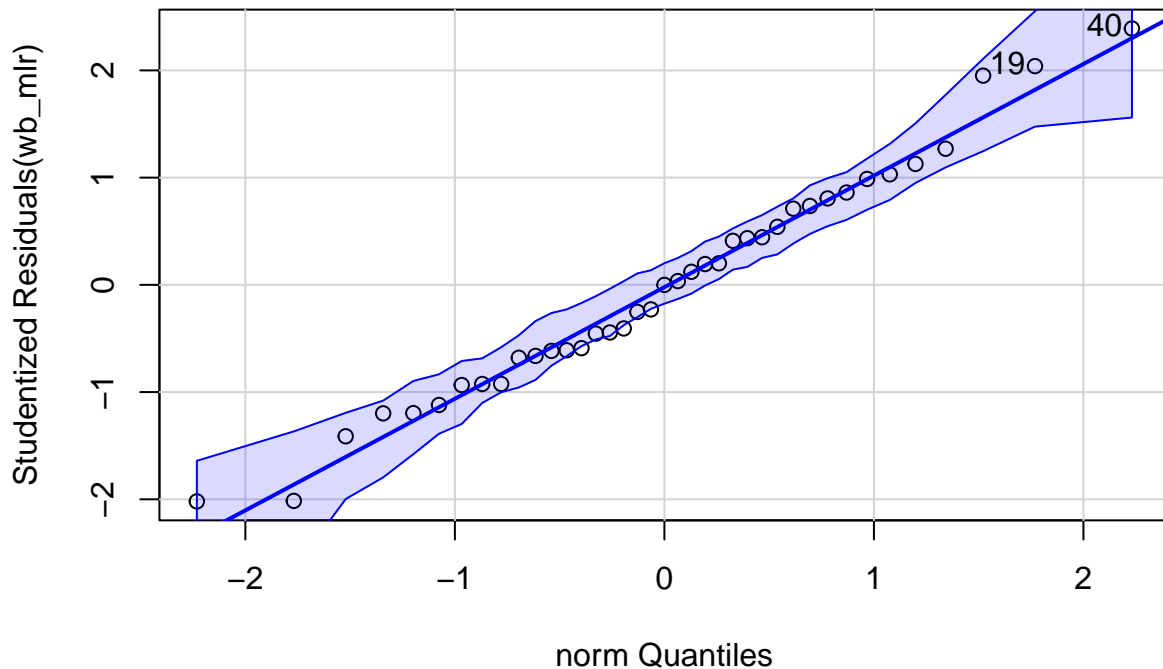
```
##
## Call:
## lm(formula = loggdp_growth ~ loginflation + internet + barter +
##     logdevel_assist + pop_growth, data = wb)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8368 -0.2607  0.0612  0.3367  1.0645
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.615572   2.564070   0.630    0.534
## loginflation    -0.137892   0.137191  -1.005    0.323
## internet        -0.002052   0.009967  -0.206    0.838
## barter          -0.003715   0.004618  -0.804    0.428
## logdevel_assist -0.028942   0.133417  -0.217    0.830
## pop_growth       0.419795   0.285979   1.468    0.153
##
## Residual standard error: 0.7629 on 29 degrees of freedom
##   (19 observations deleted due to missingness)
## Multiple R-squared:  0.1745, Adjusted R-squared:  0.03221
## F-statistic: 1.226 on 5 and 29 DF,  p-value: 0.3221
```

The standard error for the gdp growth is lower but now the standard error for internet and barter are not consistent with the data, so I think it is best to go with the original data but make a note of the disparity of the standard error in the report.

```
#Checking error asumptions
residualPlots(wb_mlr,  tests = FALSE)
```

```r
qqPlot(wb_mlr, distribution = "norm")
```



```
## [1] 19 40
```

The qqplot indicates the residuals are relatively normally distributed. In addition, the residuals vs explanatory variable plots do not appear to show any clear patterns, which suggest the residuals all have mostly equal variance.
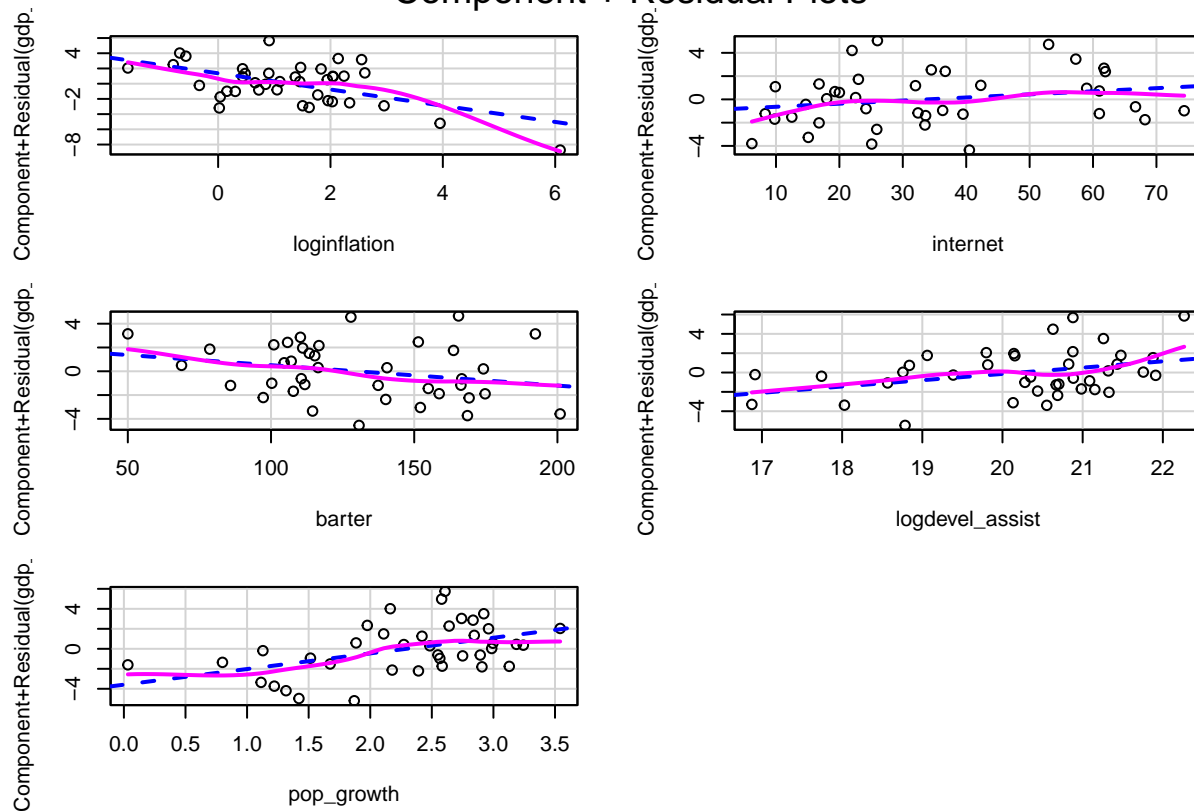
```
#Confirming no multicolinearity
vif(wb_mlr)
```

```
##    loginflation        internet          barter logdevel_assist      pop_growth
##        1.146813        2.286415        1.371598        1.267682        2.477099
```

All vif values are well below 5, so I do not see any concerns of multicollinearity!

```
#Component + Residual plots
crPlots(wb_mlr)
```
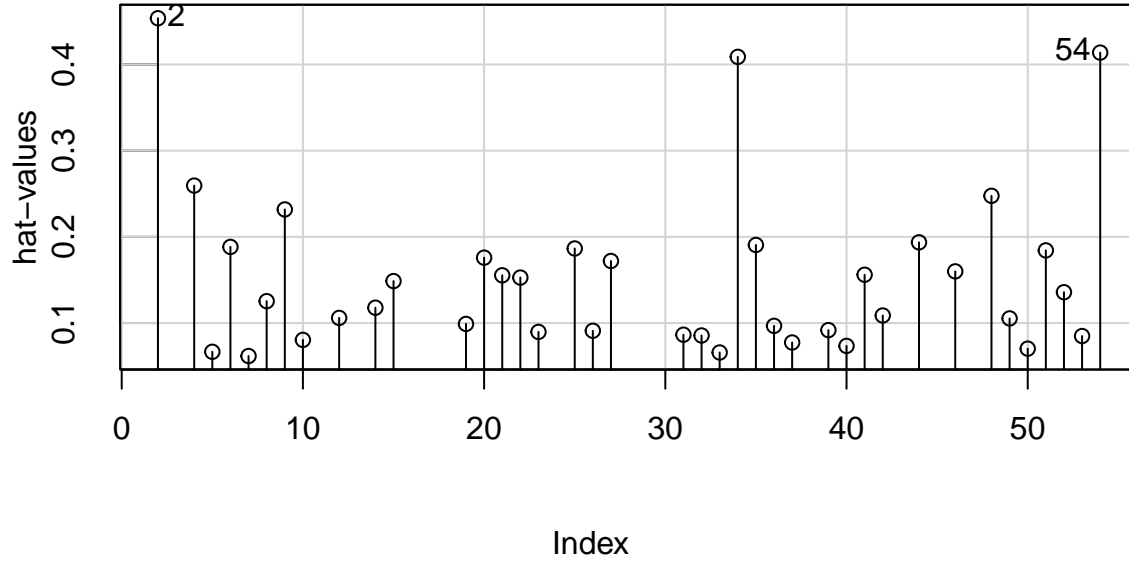

The Component and residual plots do not reveal any major non-linearity problems.
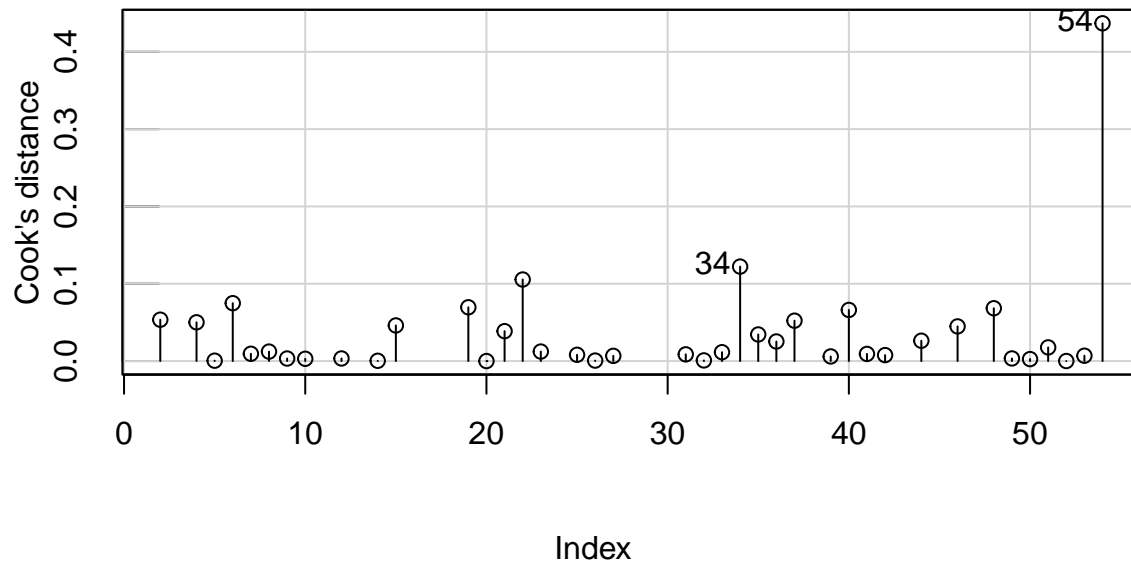
```
#Looking for outliers
infIndexPlot(wb_mlr, vars = "hat")
```

Diagnostic Plots


```
infIndexPlot(wb_mlr, vars = "Cook")
```

Diagnostic Plots

No Cook's distances close to one, so it would appear as though there are no outliers.

3(Transformations).

It appears as though other than the initial loging of inflation and development assistance, there are no other transformations that need to be made.

4(Interpreting Results).

```
#Constructing a 95% confidence interval
confint(wb_mlr, level = 0.95)
```

```
##                          2.5 %       97.5 %
## (Intercept)     -24.41712239  3.71940569
## loginflation     -1.68470048 -0.44547583
## internet         -0.03613448  0.08955348
## barter           -0.04431766  0.01002569
## logdevel_assist  -0.02938120  1.32997119
## pop_growth       -0.08271982  3.21330825
```

CI: (-.0827, 3.3133 * For the coefficient of the population growth variable.