Political Stability Presentation

Ross Pingatore

12/1/2020

Introduction: Political Stability Across the Globe

- Why does political stability vary across the globe? Are some nations innately built upon stability producing institutions while others are doomed, or can instability be triggered even within the most stable regimes? My research project will investigate the factors the produce political stability, or fail to, within the various countries across the globe.

The Data

- We will utilizes The World Bank's data set on political stability measured by the absence of violence and terrorism. The time-series data contains 213 countries and provides estimates on political stability from 1996 to 2019. The estimate of political stability ranges from -2.5 (weak stability) to 2.5 (strong stability).
- The data is combine with additional data from The World Bank the includes predictors: population, fuel exports, military expenditure, ease of conducting business, inflation rate, literacy rate, and access to electricity.

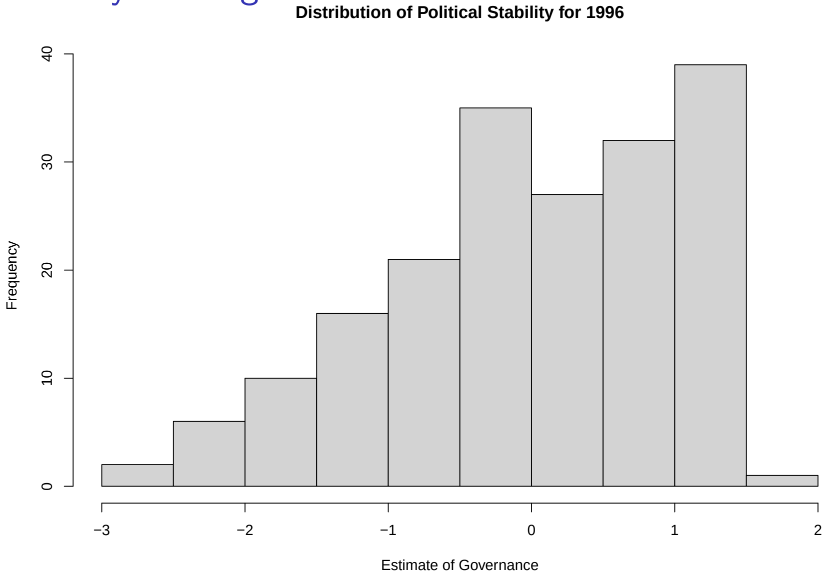
Reprocessing

- The data required a great deal of cleaning and pre-processing. Functions such as pivot-longer and dplyr's join functions were used.

```
head(clean_stability)
```

```
## # A tibble: 6 x 4
##   country code   year estimate
##   <chr>   <chr> <dbl> <chr>
## 1 Andorra ADO    1996 1.1701573133468628
## 2 Andorra ADO    1998 1.1836445331573486
## 3 Andorra ADO    2000 1.1670020818710327
## 4 Andorra ADO    2002 1.282038688659668
## 5 Andorra ADO    2003 1.4649856090545654
## 6 Andorra ADO    2004 1.4014873504638672
```

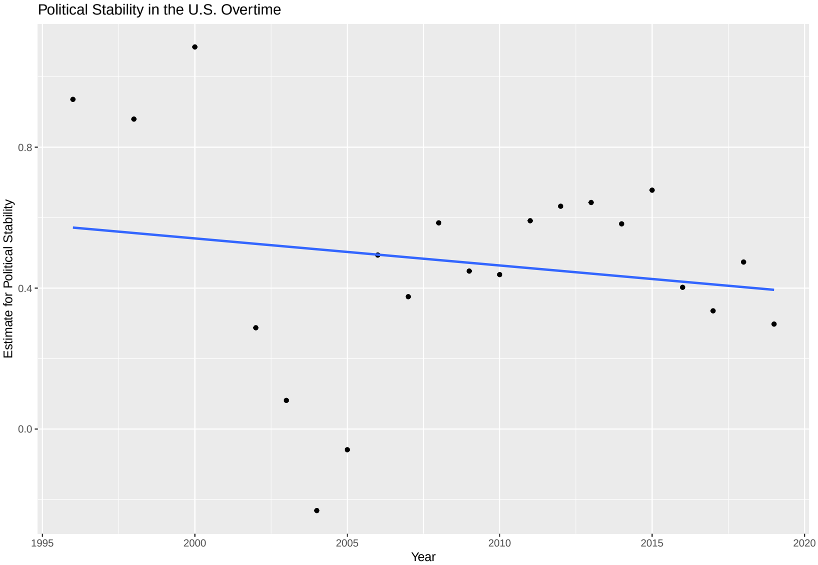
Preliminary Investigation



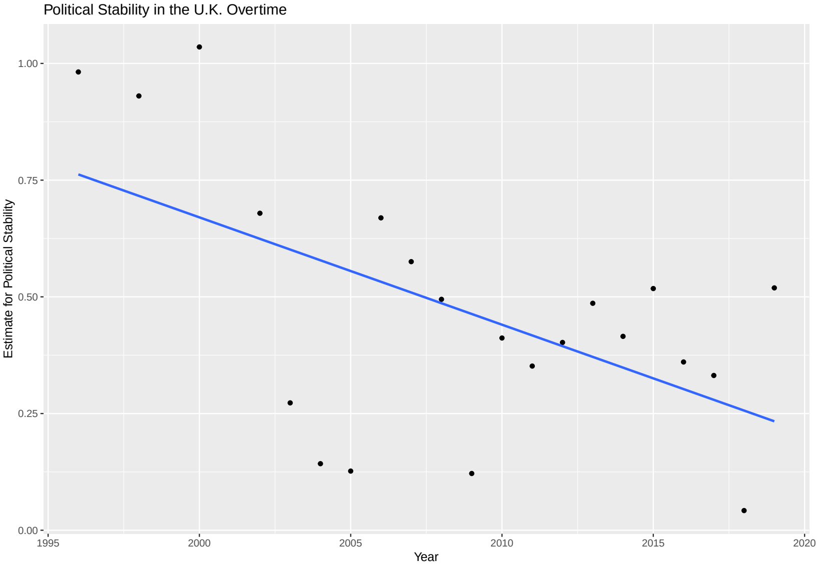
Preliminary Investigation

Distribution of Political Stability for 2019

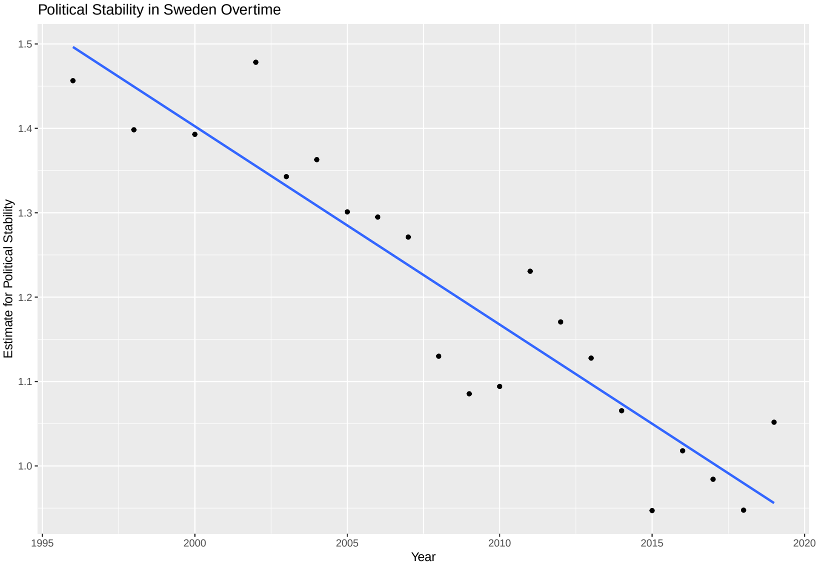
Preliminary Investigation



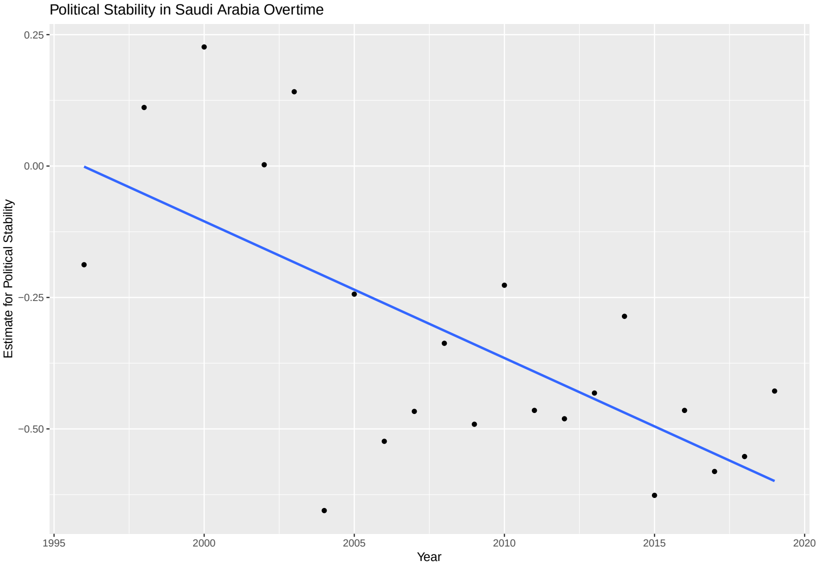
Preliminary Investigation



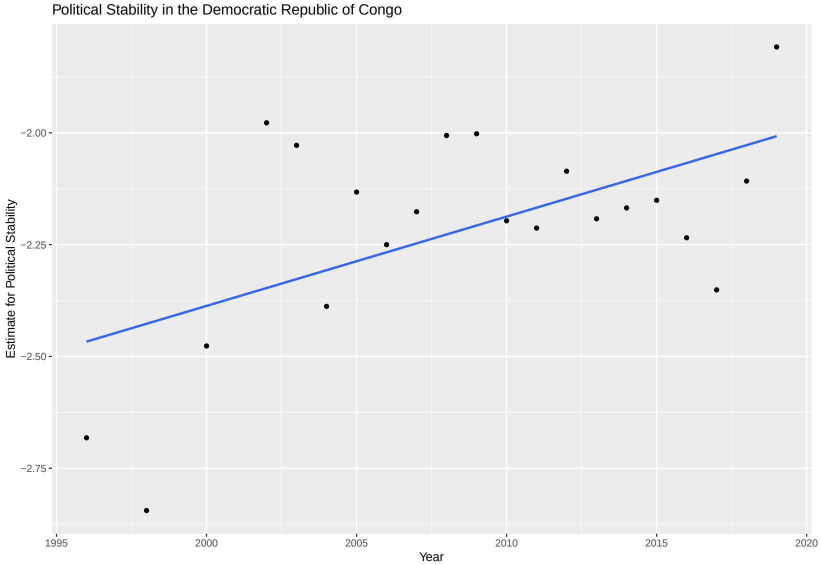
Preliminary Investigation



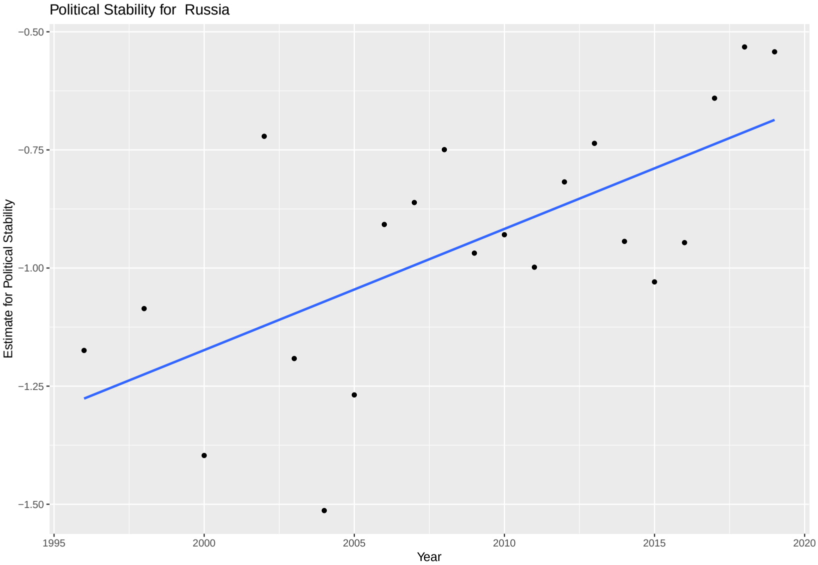
Preliminary Investigation



Preliminary Investigation



Preliminary Investigation



Preliminary Investigation

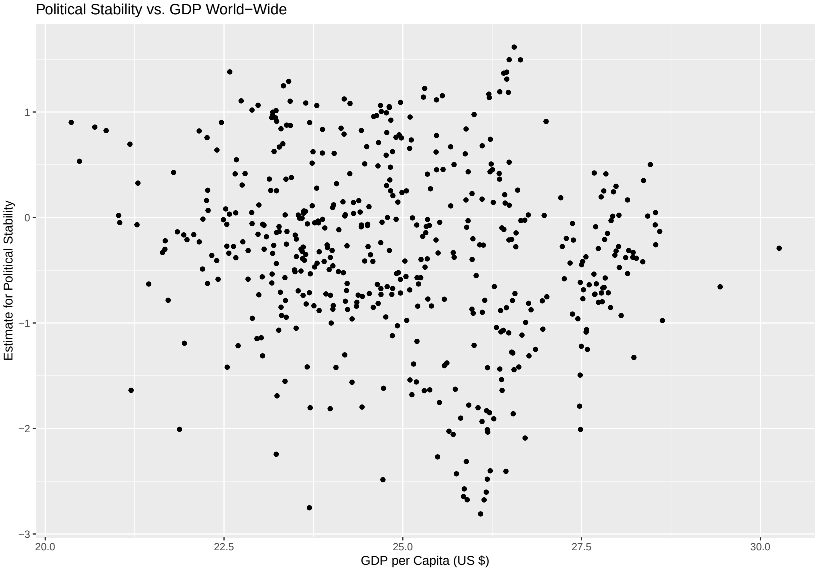
Adding More Data

```
##  [1] "country"                     "code"                        "year
##  [4] "estimate"                    "gdp"                         "popu
##  [7] "fuel_ex"                     "military_expenditure" "infl
## [10] "lit_rate"                    "electric_access"
```

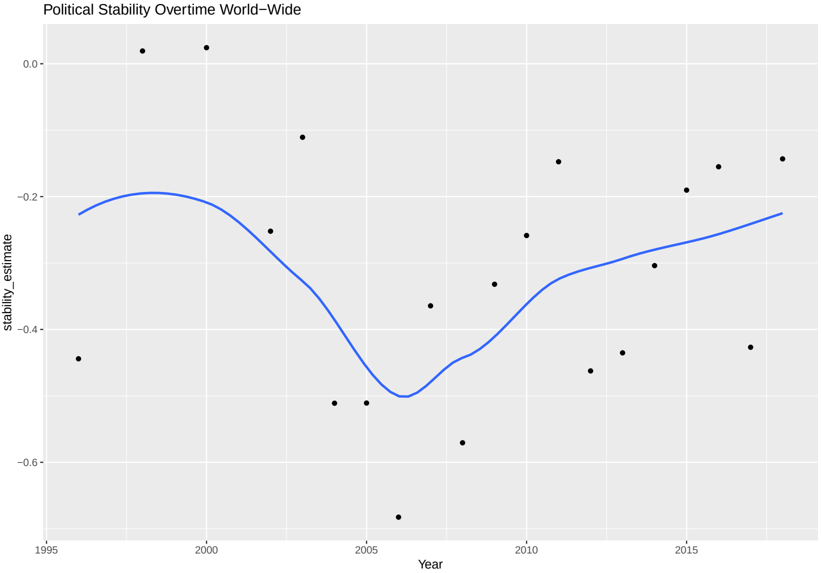
Adding More Data

```
##    country              code                year       Mi
##  Length:495         Length:495         Min.   :1996   Mi
##  Class :character   Class :character   1st Qu.:2007   1s
##  Mode  :character   Mode  :character   Median :2011   Me
##                                        Mean   :2011   Me
##                                        3rd Qu.:2015   3r
##                                        Max.   :2018   Ma
##       gdp                population            fuel_ex
##  Min.   :6.975e+08   Min.   :8.372e+04   Min.   : 0.000
##  1st Qu.:1.748e+10   1st Qu.:6.094e+06   1st Qu.: 1.374
##  Median :6.091e+10   Median :1.527e+07   Median : 6.673
##  Mean   :3.096e+11   Mean   :5.612e+07   Mean   :21.303
##  3rd Qu.:2.717e+11   3rd Qu.:4.729e+07   3rd Qu.:29.116
##  Max.   :1.389e+13   Max.   :1.393e+09   Max.   :99.986
##    inflation          lit_rate      electric_access
##  Min.   :-4.863   Min.   :12.85   Min.   :  3.696
##  1st Qu.: 2.283   1st Qu.:77.20   1st Qu.: 76.887
##  Median : 4.199   Median :92.06   Median : 98.035
```

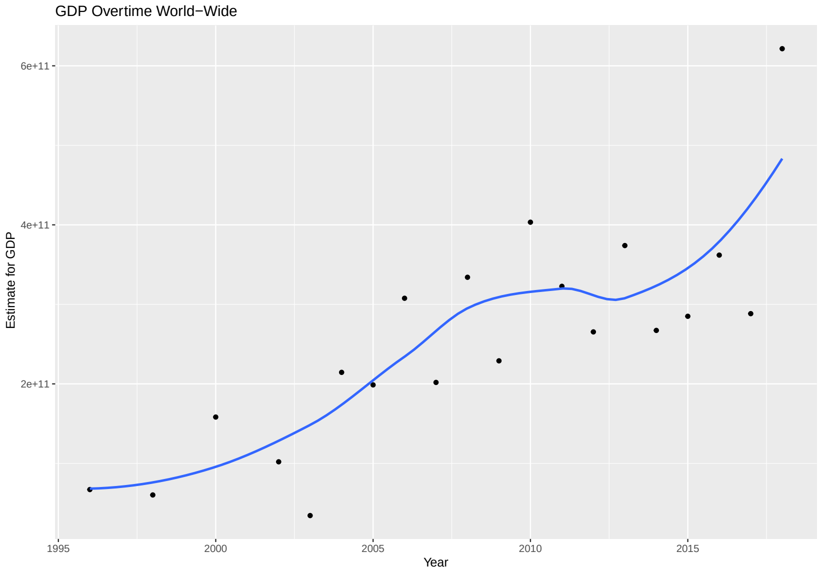
Further Investigation



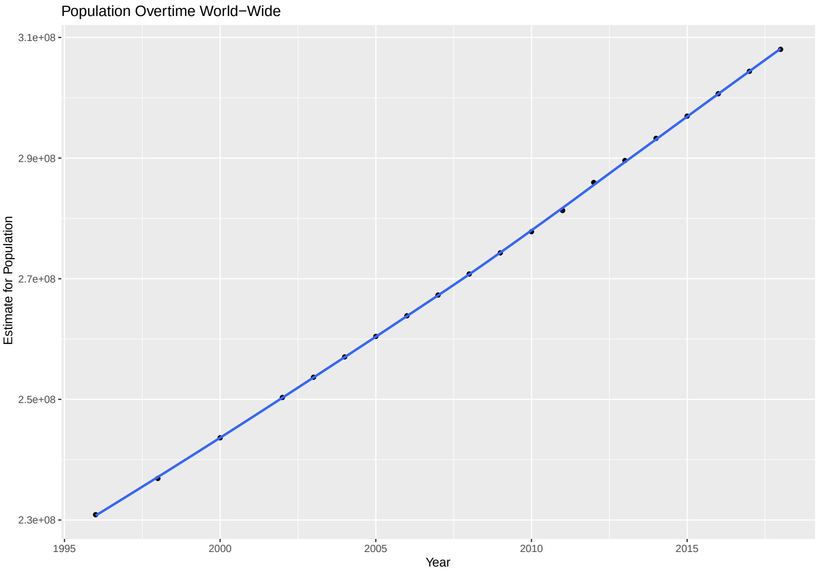
Further Investigation



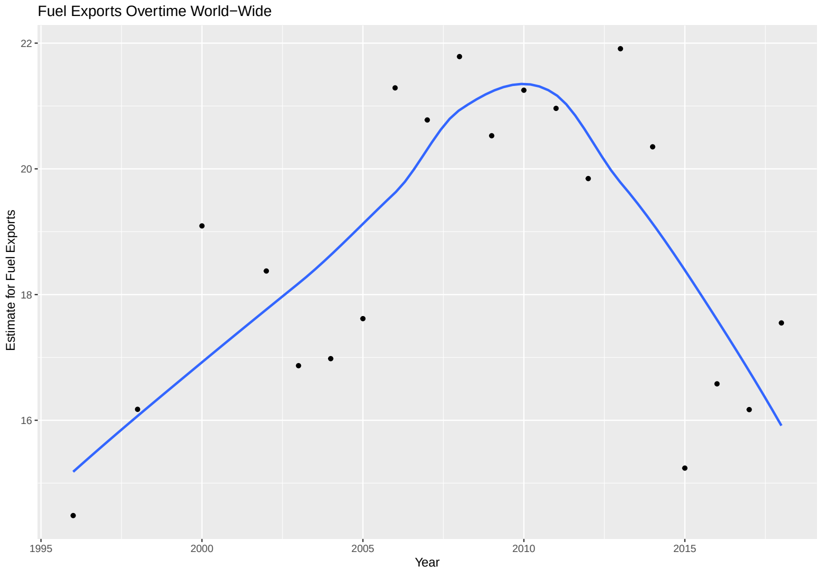
Further Investigation



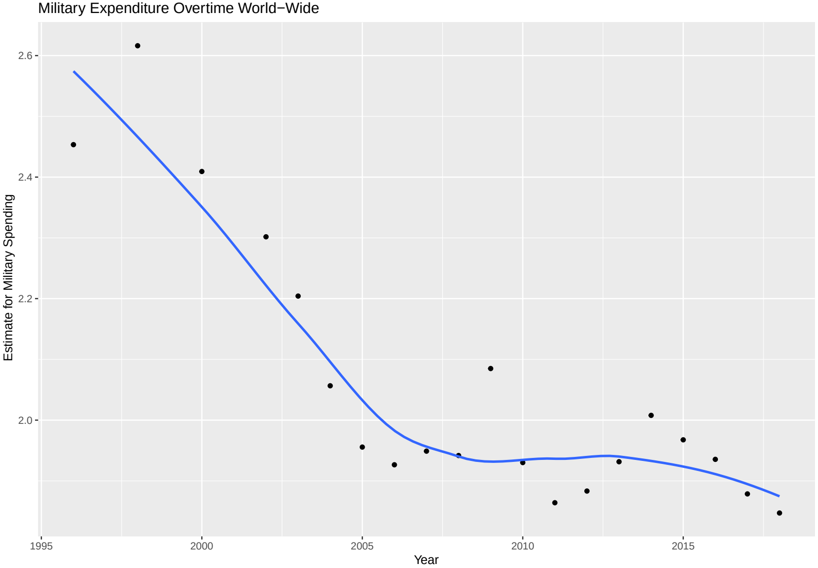
Further Investigation



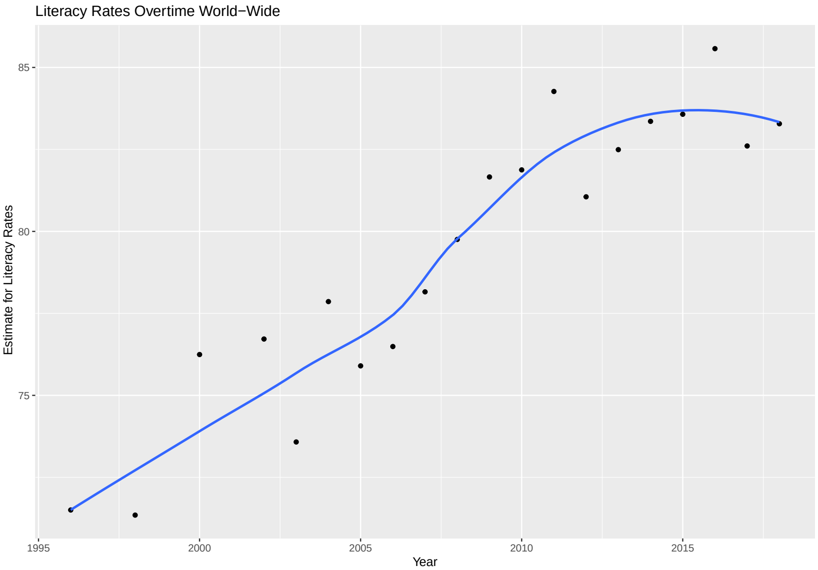
Further Investigation



Further Investigation



Further Investigation



Variable Selection and Model Building

```
##
## Call:
## lm(formula = estimate ~ gdp + population + fuel_ex + mil
##     inflation + lit_rate + electric_access, data = predi
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.98265 -0.48470  0.00826  0.51707  1.87188
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>
## (Intercept)          -1.078e+00  1.837e-01  -5.866 8.25e
## gdp                   6.298e-14  5.949e-14   1.059   0.2
## population           -1.440e-09  3.228e-10  -4.460 1.02e
## fuel_ex              -3.654e-03  1.363e-03  -2.680   0.0
## military_expenditure -1.496e-02  2.320e-02  -0.645   0.5
## inflation            -3.326e-02  6.065e-03  -5.484 6.70e
## lit_rate              1.713e-02  3.022e-03   5.668 2.47e
```

Variable Selection and Model Building

```
##                                    Best Subsets Regressi
## ------------------------------------------------------------
## Model Index    Predictors
## ------------------------------------------------------------
##      1         lit_rate
##      2         inflation lit_rate
##      3         population inflation lit_rate
##      4         population fuel_ex inflation lit_rate
##      5         population fuel_ex inflation lit_rate ele
##      6         gdp population fuel_ex inflation lit_rate
##      7         gdp population fuel_ex military_expenditu
## ------------------------------------------------------------
##
##                                                          Sub
## ------------------------------------------------------------
##                         Adj.        Pred
## Model    R-Square    R-Square    R-Square     C(p)
## ------------------------------------------------------------
```

Optimal Linear Model

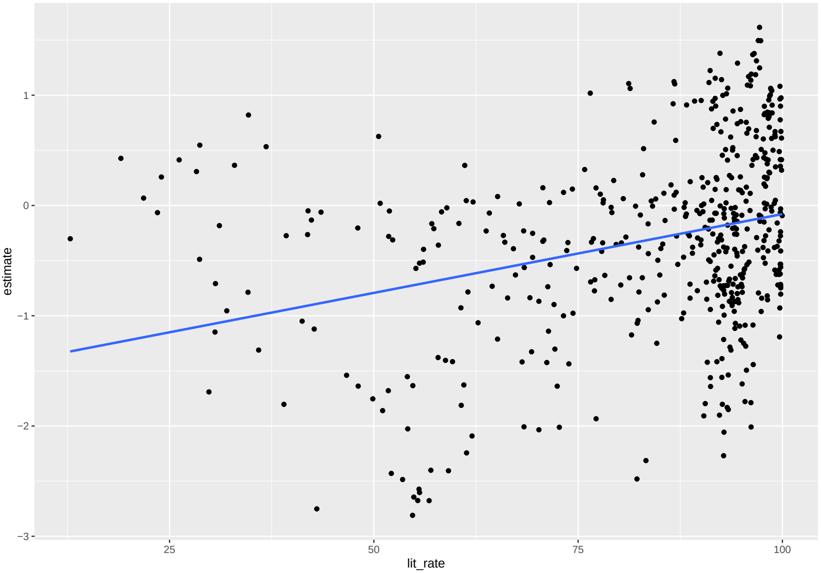
```
##
## Call:
## lm(formula = estimate ~ lit_rate, data = predictor_stab
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.14793 -0.54198  0.02002  0.57854  1.83231
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.507444   0.174928  -8.618  < 2e-16 ***
## lit_rate     0.014295   0.002037   7.016 7.57e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.
##
## Residual standard error: 0.8223 on 493 degrees of freed
## Multiple R-squared:  0.09079,    Adjusted R-squared:  0
## F-statistic: 49.23 on 1 and 493 DF,  p-value: 7.571e-12
```

Lack of Fit Test

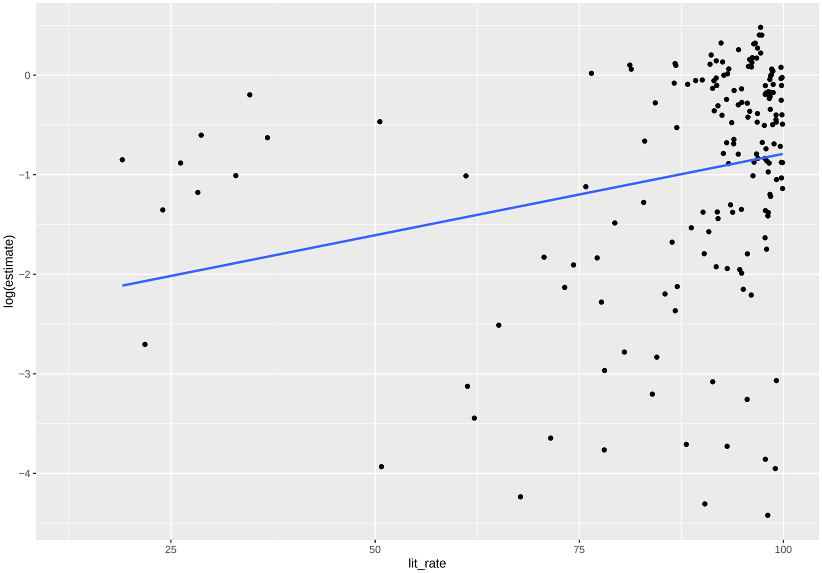
- Test assumption of linearity between political stability and literacy.
- With a small p-value we may have evidence against linearity. This need to be further investigated.

```
## Analysis of Variance Table
##
## Model 1: estimate ~ lit_rate
## Model 2: estimate ~ as.factor(lit_rate)
##   Res.Df    RSS  Df Sum of Sq      F   Pr(>F)
## 1    493 333.35
## 2      2   0.01 491    333.34 148.11 0.006729 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

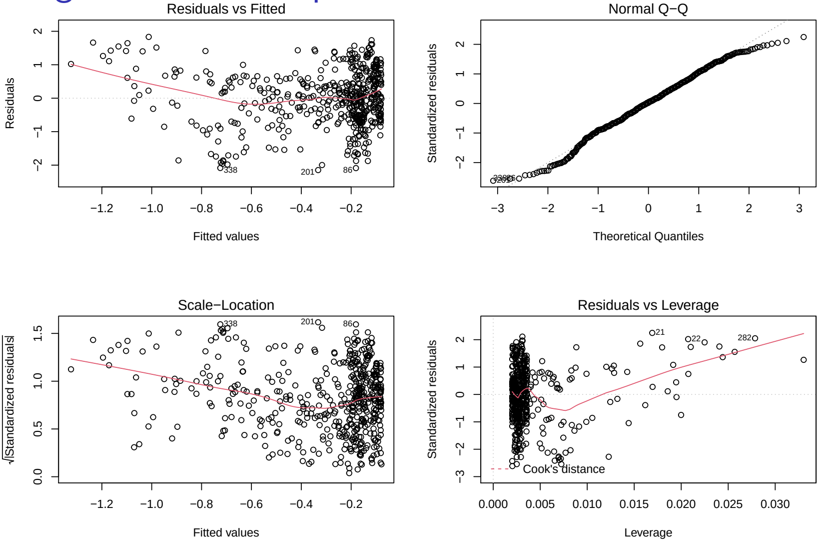
Linearity investigation



Linearity investigation



Plotting Residuals of Optimal Model



KNN

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     12.85   77.20   92.06   83.92   95.86   99.97
```

KNN

#KNN + Our first round when K = 3 gives a classification accuracy rate of 30.06%

```
##              knn.results
## Y.testing   High
##   High       117
##   Low        246
##   Standard   124
##   Very High    1
##   Very Low     1
```

```
## [1] 0.2392638
```

KNN

- A loops is then used to test our classification accuracy rate for all values of K from 1 to 20. We find that our optimal K with the highest classification accuracy rate is when K = 1 which gives a classification accuracy rate of 33.12%

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

```
## [1] 0.2392638
```

“‘