Appendix C: R Code

Attached is all of the R code written for the essay and analysis. Not all of it is included in the essay, for example the coding of mean party divergence, but I have included everything since it informed all of my work.

At each step it should be clear where any additional data is sourced.

```
#Loading packages
library(stargazer)
library(tidyverse)
library(ggplot2)
library(repmis)
library(readxl)

#Importing datasets
data <- read.csv("http://andy.egge.rs/data/L.csv")
qog_data <- read.csv("http://www.qogdata.pol.gu.se/data/qog_std_ts_jan20.csv")
```

```
## Warning in scan(file = file, what = what, sep = sep, quote = quote, dec = dec, :
## EOF within quoted string
```

```
## Warning in scan(file = file, what = what, sep = sep, quote = quote, dec = dec, :
## embedded nul(s) found in input
```

```
disproportionality_data <- repmis::source_data(url = "http://bit.ly/Ss6zDO")
newdemocracy_dimension_data <- read_excel("/Users/gussmith/Desktop/notes/Politics/Political Analysis/Ess

#Remove scientific numbers
options(scipen=999)
```

I began by replicating Lijphart's initial analysis from Patterns of Democracy.

```
#Lijphart excludes Israel and Uruguay from inflation as outliers.
data$cpi_1981_2009[18] #Israel is in the dataset
data$cpi_1981_2009[18] <- NA #Israel has now been excluded
data$cpi_1981_2009[35] #Uruguay returns the value 'NA'; no need to exclude.

#Regression of Exec-Parties Dimension on CPI
multiregression1 <- lm(data$cpi_1981_2009 ~ data$exec_parties_1981_2010
                       + data$hdi_2010 + log(data$pop_in_thousands_2009))
#Regression of Exec-Parties Dimension on Unemployment
multiregression2 <- lm(data$unemployment_1981_2009 ~ data$exec_parties_1981_2010 +
                         data$hdi_2010 + log(data$pop_in_thousands_2009))

#Building a regression table.
stargazer(multiregression1, multiregression2,
          title="Lijphart's Analysis",
          covariate.labels = c("Executive-Parties Dimension (1981-2010)", "HDI 2010",
                               "Population 2009 (logged, 1000s)"),
          column.labels = c("CPI (1981-2009)", "Unemployment (1981-2009)"),
          model.numbers = FALSE, dep.var.labels.include = FALSE,
          font.size = "small",
          omit.stat = c("f", "ser"),
          column.sep.width = "-15pt", out="regressiontable4.tex")
```

I reconstructed a new Executive-Parties dimension, removing interest group pluralism and re-defining executive dominance with a non-arbitrary measure so that presidential systems could be properly included.

As I outline in the code, I follow the example of the dataset and not the book in coding more majoritarian democracies as having negative values on the executive-parties dimension.

```
#Fixing executive dominance; want to create a variable: 'number of governing parties'.
#Source: Detailed in Appendix A.
meangoverningparties <- c(1.6, 1.5, 1.9, 1, 1, 4.5, 1, 1.2, 1.6, 3.5, 3.9,
                          2.6, 2.2, 1.2, 2.4, 1.5, 2.7, 5.4, 5.1, 1, 2.1, 1.8, 2, 1,
                          2.2, 2.7, 2.9, 1.8, 1.4, 1.9, 2.7, 4, 1.2, 1, 1.2, 1.7)

#Here - set pluralism = NA
data$index_of_interest_group_pluralism_1981_2010 = NA

#Lijphart (2012, p. 243) states that higher values are more majoritarian states,
#Here, more majoritarian states have negative values on the e-p dimension.
#I have followed the dataset's example and therefore reversed minimal winning cabinets
#and disproportionality.

effparties2 <- scale(data$eff_num_parl_parties_1981_2010)
minimalcabinets2 <- scale(-data$pct_minimal_winning_one_party_cabinet_1981_2010)
execdominance2 <- scale(meangoverningparties)
disproportionality2 <- scale(-data$index_of_disproportionality_1981_2010)

newepdimension <- ((effparties2 + minimalcabinets2 +
                      execdominance2 + disproportionality2) / 4)

#We can test to see if this closely aligns with Lijphart's dimension.
cor.test(data$exec_parties_1981_2010, newepdimension)
#This gives us that r = 0.96

#I also test for Cronbach's alpha on the new EP dimension
epconstituentsdataframe <- cbind(effparties2, minimalcabinets2,
                                 execdominance2, disproportionality2)
#Testing for Cronbach's alpha
psych::alpha(epconstituentsdataframe) #This returns that alpha = 0.89

#Testing for the alpha of Lijphart's E-P dimension
lijphartepdataframe <- cbind(data$eff_num_parl_parties_1981_2010,
                             data$index_of_disproportionality_1981_2010,
                             data$pct_minimal_winning_one_party_cabinet_1981_2010,
                             data$index_of_exec_dominance_1981_2010,
                             data$index_of_interest_group_pluralism_1981_2010)
psych::alpha(lijphartepdataframe, check.keys = TRUE) #Alpha = 0.88
```

```
## Warning in psych::alpha(lijphartepdataframe, check.keys = TRUE): Item = had no
## variance and was deleted
```

```
## Warning in psych::alpha(lijphartepdataframe, check.keys = TRUE): Some items were negatively correlat
##  This is indicated by a negative sign for the variable name.
```

I created a mean party divergence variable, thinking that the difference in policy position between parties might impact upon economic variables. In the end it only reduced the adjusted R-squared of my regressions, so I did not include it.

```
#Then - want to set up party divergence variable.
#Data for 29 countries, inputting mean divergence over the period. Not included in essay.
meanpartydivergence_1980_2010 <- c(1.457, 3.284, 1.890, NA, NA, 2.173, 0, 2.234, NA, 4.020, 1.977,
```

```
                                    2.404, 2.963, 2.983, 2.966, 3.963, 1.235, 2.434, 2.737, NA, 2.169,
                                    1.187, 1.394, 0.677, NA, 1.22, 2.872, 3.372, 1.546, 2.013, 4.574,
                                    5.515, NA, 3.404, 5.475, 2.869)
```

Given Lijphart's arbitrary inclusion of outliers, I re-ran quartile tests for outliers in the inflation and unemployment data from 1981-2010.

```
#Want to test for CPI first.
boxplot(data$cpi_1981_2009, main = "Figure 2: CPI Outliers",
        xlab = "CPI Inflation (1981-2009)",
        horizontal = TRUE)
```

```
#Now - need to calculate which countries are the exception
quantile(data$cpi_1981_2009, na.rm = TRUE)
#IQR = 4.21075, Q3 + 1.5IQR = 13.89987
#Now, need to identify which are the values over 13.9.
plot(data$cpi_1981_2009, data$country, type = "n")
text(data$cpi_1981_2009, data$country, labels = data$country, cex = 0.7)
```

```
#So - Costa Rica and Jamaica are outliers with regards to inflation.
#So is Israel, but I have already excluded it.

#Second variable - unemployment
boxplot(data$unemployment_1981_2009, main = "Figure 3: Unemployment Outliers",
        xlab = "Unemployment % (1981-2009)", horizontal = TRUE)
```

```
#2 outliers again.
quantile(data$unemployment_1981_2009, na.rm = TRUE)
#IQR = 3.59525, Q3 + 1.5IQR = 15.07737
plot(data$unemployment_1981_2009, data$country, type = "n")
text(data$unemployment_1981_2009, data$country, labels = data$country, cex = 0.7)
```

```
#The states are Spain and Jamaica with unemployment.
#Inflation - excluded Israel and Uruguay already.
data$cpi_1981_2009[9] <- NA #Costa Rica
data$cpi_1981_2009[20] <- NA #Jamaica
data$unemployment_1981_2009[30] <- NA #Spain
data$unemployment_1981_2009[20] <- NA #Jamaica
```

I re-ran the regressions on CPI inflation and unemployment with the new executive-parties dimension and outliers removed. The regressions controlling for party divergence are included.

```
#with old controls of HDI and population logged
multiregression3 <- lm(data$cpi_1981_2009 ~ newepdimension +
                         data$hdi_2010 + log(data$pop_in_thousands_2009))

multiregression4 <- lm(data$unemployment_1981_2009 ~ newepdimension +
                         data$hdi_2010 + log(data$pop_in_thousands_2009))

#Now - produce a table for this first set of multiregression analysis.
stargazer(multiregression3, multiregression4, title="Refining Lijphart's Analysis",
          covariate.labels = c("Updated Executive-Parties Dimension (1981-2010)",
                               "HDI 2010",
                               "Population 2009 (logged, 1000s)"),
          column.labels = c("CPI (1981-2009)", "Unemployment (1981-2009)"),
          model.numbers = FALSE, dep.var.labels.include = FALSE,
```

```
          font.size = "small",
          omit.stat = c("f", "ser"),
          column.sep.width = "-15pt", out="regressiontable5.tex")

#Next, we want to add the party divergence variable as an extra control.
#Not included in essay.
multiregression5 <- lm(data$cpi_1981_2009 ~ newepdimension +
                         data$hdi_2010 + log(data$pop_in_thousands_2009) +
                         meanpartydivergence_1980_2010)
multiregression6 <- lm(data$unemployment_1981_2009 ~ newepdimension + data$hdi_2010 +
                         log(data$pop_in_thousands_2009) + meanpartydivergence_1980_2010)

#Produce a table for these two regressions. Not included in the essay
stargazer(multiregression5, multiregression6,
          title="Table 3: Considering Party Divergence",
          covariate.labels = c("Updated Executive-Parties Dimension (1981-2010)",
                               "HDI 2010",
                               "Population 2009 ((1000s,) logged)",
                               "Mean Party Divergence (1980-2010)"),
          column.labels = c("CPI (1981-2009)", "Unemployment (1981-2009)"),
          model.numbers = FALSE, dep.var.labels.include = FALSE,
          out="regressiontable6.html")
```

I then performed regressions on a variety of economic indicators relevant to the theoretical argument - inflation volatility, investment, and R&D spending.

```
#Research and Development Spending as a % of GDP
#Source: World Bank [https://data.worldbank.org/indicator/GB.XPD.RSDV.GD.ZS]
mean_research_spending_percent_gdp_1980_2010 <- c(0.44777, 1.91622, 2.1427, NA, NA,
                                                  1.87435, 0.53286, 1.86676, 0.38073,
                                                  2.3988, 3.21814, 2.1181, 2.41962,
                                                  0.56665, 2.51366, 0.75914, 1.23902,
                                                  3.78194, 1.07031, 0.05503, 3.04645,
                                                  2.58704, 1.60422, 0.47602, 0.32447,
                                                  1.75547, 1.58709, 1.11415, 0.89638,
                                                  1.03846, 4.33717, 2.53986, 0.09488,
                                                  1.60445, 0.30927, 2.59394)
#Mean Real Interest Rate
#Source: World Bank [https://data.worldbank.org/indicator/FR.INR.RINR]
mean_realir_1980_2010 <- c(5.73775, 5.15559, NA, 2.87593, 5.25683, NA, 3.74883,
                           5.0055, 6.76317, NA, NA, NA, NA, NA, 5.82981,
                           6.18931, NA, 6.91291, 5.63949, 6.38277, 3.75266,
                           5.23501, NA, 3.78474, 8.73435, 1.31556, NA, 5.04796,
                           NA, NA, 5.67907, 2.02603, 8.04526, 3.12076, 27.11934, 5.26575)
#Investment [Gross Capital Formation] as % of gdp
#Source: World Bank [https://data.worldbank.org/indicator/NE.GDI.FTOT.ZS]
investment_percent_gdp_1980_2010 <- c(18.34842, 26.57779, 25.13376, 25.7561,
                                      16.91716, 22.2687, 30.49037, 21.4423,
                                      22.34538, 20.88851, 24.62634, 22.14893,
                                      22.96917, 25.55618, 23.40727, 28.21113,
                                      22.47671, 22.06105, 21.57884, 23.91849,
                                      28.7914, 34.02656, 21.56247, 24.32699,
                                      25.2084, 22.04612, 25.2008, 23.10037,
                                      26.09765, 24.72911, 23.96684, 26.67047, NA,
```

```
                                    19.50643, 16.00303, 22.19966)
#Mean Inflation Volatility ]
#Source: Appendix B
cpi_volatility_1980_2010 <- c(NA, 0.46695, 0.45838, 0.45936, 0.87289, 0.3917,
                              0.67234, 0.45604, 1.02759, 0.43351, 0.45318, 0.34204,
                              0.38846, 1.01622, 0.90186, 0.96441, 0.59202, 1.1558,
                              0.35332, 1.26056, 0.4725, 0.71989, 0.42493,
                              0.70044, 0.9075, 0.39054, 0.51668, 0.56954,
                              0.68027, 0.55606, 0.58243,
                              0.46025, 0.7629, 0.53995, 1.30292, 0.43073)
#Trade (% GDP)
#Source: World Bank [https://data.worldbank.org/indicator/NE.TRD.GNFS.ZS]
trade_percent_gdp_1980_2010 <- c(23.81338, 36.60527, 77.00931, 98.22535, 99.21468,
                                 126.9815, 101.49023, 62.22604, 78.83972,
                                 75.99064, 63.38736, 47.77438, 54.1415, 45.15598,
                                 71.49991, 25.07988, 132.40828, 69.75524, 43.82693,
                                 96.00476, 22.34597, 63.84618, 226.05526, 200.71947,
                                 199.02762, 111.54283, 71.44611, 58.13953,
                                 61.83021, 45.55861, 69.6311, 91.9803,
                                 NA, 51.04469, 44.35633, 21.81968)

#First Analysis; inflation volatility
boxplot(cpi_volatility_1980_2010) #No outliers
```

```
#Without controlling for economic openness
multiregression7 <- lm(cpi_volatility_1980_2010 ~ newepdimension + data$hdi_2010 +
                         log(data$pop_in_thousands_2009))
#Controlling for economic openness
multiregression8 <- lm(cpi_volatility_1980_2010 ~ newepdimension + data$hdi_2010 +
                         log(data$pop_in_thousands_2009) + trade_percent_gdp_1980_2010)

#Producing the regression table for the essay.
stargazer(multiregression7, multiregression8, title="Inflation Volatility",
          covariate.labels = c("Updated Executive-Parties Dimension", "HDI 2010",
                               "Population 2009 (logged, 1000s)",
                               "Economic Openness (1980-2010)"),
          dep.var.labels = "CPI Inflation Volatility (1980-2010)",
          font.size = "small",
          omit.stat = c("f", "ser"),
          column.sep.width = "-15pt",
          out="regressiontable7.tex")
```

```
#Second Analysis: R&D as a % of GDP
boxplot(mean_research_spending_percent_gdp_1980_2010) #No outliers
```

```
multiregression9 <- lm(mean_research_spending_percent_gdp_1980_2010 ~ newepdimension +
                         data$hdi_2010 + log(data$pop_in_thousands_2009))

#Third Analysis: Investment as % of GDP
boxplot(investment_percent_gdp_1980_2010, main = "Figure 4: Investment Outliers",
        xlab = "Gross Capital Formation 1980-2010 (% of GDP", horizontal = TRUE)
```

```
#Quartiles?
quantile(investment_percent_gdp_1980_2010, na.rm = TRUE)
```

```
#IQR = 3.2871
#Lower Outlier Bound - 17.06052
#Higher Outlier Bound - 30.37535

#Plot to identify the outliers
plot(investment_percent_gdp_1980_2010, data$country, type="n")
text(investment_percent_gdp_1980_2010, data$country, labels=data$country, cex=0.7)
```

```
#Outliers are Botswana, Korea, Uruguay, Barbados
#Clean up the outliers
investment_percent_gdp_1980_2010[5] <- NA #Barbados
investment_percent_gdp_1980_2010[7] <- NA #Botswana
investment_percent_gdp_1980_2010[22] <- NA #Korea
investment_percent_gdp_1980_2010[35] <- NA #Uruguay

#Run first regression
multiregression10 <- lm(investment_percent_gdp_1980_2010 ~ newepdimension +
                          data$hdi_2010 + log(data$pop_in_thousands_2009))
#Controlling for real IR
multiregression11 <- lm(investment_percent_gdp_1980_2010 ~ newepdimension +
                        data$hdi_2010 + log(data$pop_in_thousands_2009) +
                          mean_realir_1980_2010)
```

```
#Producing Stargazer tables for R&D and Investment
stargazer(multiregression9, title="R&D Spending (1980-2010)",
          covariate.labels = c("Updated Executive-Parties Dimension",
                               "HDI 2010",
                               "Population 2009 (logged, 1000s)"),
          dep.var.labels = "Research and Development Spending as % of GDP (1980-2010)",
          font.size = "small",
          omit.stat = c("f", "ser"),
          column.sep.width = "-15pt",
          out = "regressiontable8.tex")

stargazer(multiregression10, multiregression11,
          title="Gross Capital Formation (1980-2010)",
          covariate.labels = c("Updated Executive-Parties Dimension", "HDI 2010",
                               "Population 2009 (logged, 100s)",
                               "Mean Real Interest Rate (1980-2010)"),
          dep.var.labels = "Investment as % of GDP (1980-2010)",
          font.size = "small",
          omit.stat = c("f", "ser"),
          column.sep.width = "-15pt",
          out="regressiontable9.tex")
```

```
#Checking R&D measures of central tendency and dispersion
mean(mean_research_spending_percent_gdp_1980_2010, na.rm = TRUE)
sd(mean_research_spending_percent_gdp_1980_2010, na.rm = TRUE)
```

I then expand the number of cases to reflect democracies which fulfill Lijphart's criteria as of 2019; 20 years of continous democracy judged by 'Free' on Freedom House rankings and having a population of over 250,000. Those are Belize, Bulgaria, Cape Verde, Chile, Cyprus, Estonia, Latvia, Lithuania, Mongolia, Namibia, Panama, Poland, Romania, Slovenia, South Africa, Taiwan, Vanuatu. First of all I formulated an executive-parties dimension for the countries.

```
#Data for all QoG data is:
#[https://qog.pol.gu.se/data/datadownloads/qogstandarddata]
#calculating effective number of parties
qog_data %>%
  group_by(cname) %>%
  filter(ccodealp == "BLZ"| ccodealp == "BGR" | ccodealp == "CPV" | ccodealp == "CHL" |
           ccodealp == "CYP" | ccodealp == "EST" | ccodealp == "LVA" | ccodealp == "LTU" |
           ccodealp == "MNG" | ccodealp == "NAM" | ccodealp == "PAN" | ccodealp == "POL" |
           ccodealp == "ROU" | ccodealp == "SVN" | ccodealp == "ZAF" | ccodealp == "TWN" |
           ccodealp == "VUT") %>%
  summarise(gol_enep = gol_enep[which(!is.na(gol_enep))[1]])

#SA; elections from 1994 onwards. Source: African Elections Database.
#[http://africanelections.tripod.com/]
south_africa_epp = mean((1/(0.62^2 + 0.2039^2 + 0.104^2 + 0.022^2 + 0.07^2 + 0.015^2)),
                        (1/(0.66^2 + 0.096^2 + 0.086^2 + 0.069^2 + 0.034^2 + 0.14^2)),
                        (1/(0.69^2 + 0.124^2 + 0.0697^2 + 0.023^2 + 0.0173^2 + 0.0165^2 + 0.16^2)),
                        (1/(0.659^2+0.1666^2+0.0742^2+0.0455^2)))
#= 2.260397 = 2.26

#Namibia: elections from 1994 onwards. Data from African Elections Database.
namibia_epp = mean((1/(0.739^2+0.28^2+0.0272^2)), (1/(0.7615^2+0.0994^2+0.0948^2+0.00293^2)),
                   (1/(0.7611^2+0.0729^2+0.0511^2+0.0415^2+0.036^2+0.0196^2)),
                   (1/(0.7527^2+0.1131^2+0.0317^2+0.0305^2+0.0243^2+0.0135^2)))
#=1.599333 = 1.6

mean_eff_parties_new_democracies <- c(2.06, 4.12, 1.81, 7.05, 3.62, 8.72, 6.21, 4.58,
                                      2.71, 1.6, 4.71, 13.9, 2.21, 8.34, 2.26, 3.14, 2.56)

#Data from Gallagher Electoral Disproportionality Data, 1945-2014, v2, Gandrud
#[http://christophergandrud.github.io/Disproportionality_Data/]
disproportionality_data %>%
  group_by(country) %>%
  filter(iso2c == "BZ"| iso2c == "BG" | iso2c == "CV" | iso2c == "CL" | iso2c == "CY" |
           iso2c == "EE" | iso2c == "LV" | iso2c == "LT" | iso2c == "MN" | iso2c == "NA" |
           iso2c == "PA" | iso2c == "PL" | iso2c == "RO" | iso2c == "SI" | iso2c == "ZA" |
           iso2c == "TW" | country == "Vanuatu") %>%
  summarise(mean_disproportionality = disproportionality[which(!is.na(disproportionality))[1]])

#No disproportionality data for Vanuatu from this dataset. Source - Wikipedia.
#Calculating Vanuatu disproportionality from 1991-2012, without data for 2004.
vanuatu_disproportionality = mean(sqrt(0.5 * ((30.6-41)^2 + (22.6-21.7)^2 +
                                              (20.4-10/46)^2 +
                                              (15.4-4/46)^2 + (4.6-1/46)^2 +
                                              (2.9-1/46)^2 + (1.9-1/46)^2)),
                                  sqrt(0.5 * ((31.4-20/50)^2 + (27.4-17/50)^2 +
                                              (23.4-9/50)^2 +
                                              (2.7-1/50)^2 + (2.3-0)^2 +
                                              (1.8-1/50)^2 + (1.6-0)^2)),
                                  sqrt(0.5 * ((21 - 18/52)^2 + (20.1-12/52)^2 +
                                              (15.9-11/52)^2 +
                                              (14-6/52)^2 + (7.6-1/52)^2 +
                                              (2.2-2/52)^2 + (1.1-0)^2)),
```

```
                                  sqrt(0.5 * ((17-14/52)^2 + (15.1-15/52)^2 +
                                                (13.6-8/52)^2 +
                                                (7.1-3/52)^2 + (6.2-3/52)^2 +
                                                (5.1-1/52)^2 + (4.7-2/52)^2),
                                  sqrt(0.5 * ((24.23-11/52)^2 + (15.66-8/52)^2 +
                                                (13.26-7/52)^2 +
                                                (11.63 -7/52)^2 + (4.92-4/52)^2 +
                                                (3.44 - 2/52)^2 +
                                                (3.44 - 1/52)^2)),
                                  sqrt(0.5 * ((12.19-5/52)^2 + (11.29-8/52)^2 +
                                                (8 - 6/52)^2 +
                                                (6.2-4/52)^2 + (6.02-4/52)^2 +
                                                (5.75-3/52)^2 +
                                                (4.23 - 3/52)^2))))
#19.8

mean_disproportionality_new_democracies <- c(19.7, 5.37, 7, 6.3, 17.1, 7.23, 4.14,
                                             9.61, 27.4, 0.93, 7.87, 3.62, 0.9, 2.94,
                                             0.36, 7.05, 19.8)


#Next variable is minimal winning cabinets. The code used is long; available on request.
#I just use dplyr repeatedly to search QoG dataset.

#No QoG data for Belize, Cape Verde, Chile, Mongolia, Namibia,
#Panama, South Africa, Taiwan, Vanuatu.
cabinet_new_democracy <- c(NA, 30.36, NA, NA, 100, 44.23, 18, 36.54, NA,
                           NA, NA, 40.74, 39.29, 34, NA, NA, NA)

#Data from African Elections and Carribbean Elections and Wikipedia for the rest.
cabinet_new_democracy <- c(100, 30.36, 100, 21, 100, 44.23, 18, 36.54, 83,
                           100, 20, 40.74, 39.29, 34, 100, 86, 29)

#Finally, we find the executive dominance measure.
#Source: PGDS, African Elections, Carribbean Elections, Wikipedia
newdemocracy_dimension_data %>%
  group_by(Country) %>%
  select(exec_dominance) %>%
  summarise(mean = mean(exec_dominance))

dominance_new_democracy <- c(1, 2, 1, 4.71, 2.25, 2.67, 4.06, 2.58, 1.5, 1, 3, 3.47, 3,
                             3.44, 1, 1.57, 3.71)

#Now standardise the variables:
standardised_newdem_epp <- scale(mean_eff_parties_new_democracies)
standardised_newdem_disproportionality <- scale(-mean_disproportionality_new_democracies)
standardised_newdem_cabinet <- scale(-cabinet_new_democracy)
standardised_newdem_dominance <- scale(dominance_new_democracy)
newdemocracy_ep_dimension <- ((standardised_newdem_epp +
                                 standardised_newdem_disproportionality +
                                 standardised_newdem_cabinet +
                                 standardised_newdem_dominance) / 4)
```

```r
#Checking the internal validity
newdemocracy_ep_dataframe <- cbind(standardised_newdem_epp,
                                   standardised_newdem_disproportionality,
                                   standardised_newdem_cabinet,
                                   standardised_newdem_dominance)

psych::alpha(newdemocracy_ep_dataframe)
#Alpha = 0.76
```

Having defined an executive-parties dimension for the new democracies, I then regress them against the economic indicators again.

```r
#Countries are: Belize, Bulgaria, Cape Verde, Chile, Cyprus, Estonia, Latvia, Lithuania,
#Mongolia, Namibia, Panama, Poland, Romania, Slovenia, South Africa, Taiwan, Vanuatu

#International Financial Statistics
#[http://data.imf.org/?sk=388DFA60-1D26-4ADE-B505-A05A558D9A42]
cpi_volatility_newdem_1980_2010 <- c(NA, 2.26931, 1.16605, 0.82212, 0.84151, 1.34902,
                                     1.65866, 1.59255, 2.03858, 0.72881, 0.44063, 1.66167,
                                     1.75869, 1.78236, 0.67919, NA, 0.82473)
#World Bank Data - all same source as above.
mean_realir_newdem_1980_2010 <- c(12.83, 6.75, 6.57, 6.63, NA, NA, NA, NA, 16.47, 6.72,
                                  7.47, NA, 0.08, NA, 4.35, NA, 8.2)
#World Bank Data
trade_percent_gdp_newdem_1980_2010 <- c(106.58, 86.6, 90, 58.6, 114.79, 135.33, 89.94,
                                        101.97, 103.73, 98.94, 129.29, 64.85, 56.73,
                                        112.96, 47, NA, 98.27)
#World Bank Data
investment_percent_gdp_newdem_1980_2010 <- c(22.6, 25.41, 47.44, 22.51, 25.82, 30.18,
                                             27.49, 22.38, 38.95, 19.28, 29.46, 21.99,
                                             25.06, 25.84, 20.46, NA, 25.39)
#World Bank Data
mean_research_spending_percent_gdp_newdem_1980_2010 <- c(NA, 0.47, NA, 0.34, 0.33,
                                                         0.94, 0.46, 0.68, 0.23,
                                                         0.14, 0.26, 0.61, 0.36, 1.46,
                                                         0.8, NA, NA)

#World Bank Data
population_newdem_data <- read_csv("/Users/gussmith/Desktop/Notes/Politics/Political Analysis/Essay/Data

population_newdem_data %>%
  filter(Country == "Belize" | Country == "Bulgaria" | Country == "Cabo Verde" |
           Country == "Chile" |  Country == "Cyprus" | Country == "Estonia" |
           Country == "Latvia" | Country == "Lithuania" | Country == "Mongolia" |
           Country == "Namibia" | Country == "Panama" | Country == "Poland" |
           Country == "Romania" | Country == "Slovenia" | Country == "South Africa" |
           Country == "Taiwan" |  Country == "Vanuatu") %>%
  mutate(pop_in_thousands = Population/1000)

pop_newdem_thousands_2009 <- c(315, 7444, 487, 16886, 1098, 1335, 2142, 3163, 2674, 2081,
                               3579, 38152, 20367, 2040, 50477, NA, 236)
#Vanuatu pop is >250k now, but measuring as of 2009 since that is
#limit of economic variables and same as Lijphart.
```

```r
#From the United Nations Human Development Reports
#[http://hdr.undp.org/en/data]
hdi_newdem_data <- read_excel("/Users/gussmith/Desktop/Notes/Politics/Political Analysis/Essay/Data/New

hdi_newdem_data %>%
  group_by(Country) %>%
  filter(Country == "Belize" | Country == "Bulgaria" | Country == "Cabo Verde" |
           Country == "Chile" | Country == "Cyprus" | Country == "Estonia" |
           Country == "Latvia" | Country == "Lithuania" | Country == "Mongolia" |
           Country == "Namibia" | Country == "Panama" | Country == "Poland" |
           Country == "Romania" | Country == "Slovenia" | Country == "South Africa" |
           Country == "Taiwan" | Country == "Vanuatu")

hdi_newdem_2010 <- c(0.693, 0.779, 0.626, 0.8, 0.85, 0.844, 0.817, 0.824, 0.697, 0.588,
                     0.758, 0.835, 0.797, 0.881, 0.662, NA, 0.585)

#Combining to get a full data frame of 53 democracies
newdemcountrylist = c("Belize", "Bulgaria", "Cape Verde", "Chile", "Cyprus",
                      "Estonia",  "Latvia",  "Lithuania", "Mongolia", "Namibia",
                      "Panama", "Poland", "Romania", "Slovenia",
                      "South Africa", "Taiwan", "Vanuatu")

total_36_data <- data.frame(country = data$country, epdimension=newepdimension,
                            pop_2009=data$pop_in_thousands_2009,
                            hdi_2010=data$hdi_2010,
                            investment_mean=investment_percent_gdp_1980_2010,
                            realir_mean=mean_realir_1980_2010,
                            rand_mean=mean_research_spending_percent_gdp_1980_2010,
                            cpi_mean=cpi_volatility_1980_2010,
                            trade_mean=trade_percent_gdp_1980_2010)

total_new_data <- data.frame(country = newdemcountrylist,
                             epdimension=newdemocracy_ep_dimension,
                             pop_2009=pop_newdem_thousands_2009,
                             hdi_2010=hdi_newdem_2010,
                             investment_mean=investment_percent_gdp_newdem_1980_2010,
                             realir_mean=mean_realir_newdem_1980_2010,
                             rand_mean=mean_research_spending_percent_gdp_newdem_1980_2010,
                             cpi_mean=cpi_volatility_newdem_1980_2010,
                             trade_mean=trade_percent_gdp_newdem_1980_2010)

combined_dataset <- rbind(total_36_data, total_new_data)

#Now, before running regressions, I return the values previously removed as outliers.
combined_dataset$investment_mean[5] <- 16.91716
combined_dataset$investment_mean[7] <- 30.49037
combined_dataset$investment_mean[22] <- 34.02656
combined_dataset$investment_mean[35] <- 16.00303

#Investment
boxplot(combined_dataset$investment_mean)
```

```r
quantile(combined_dataset$investment_mean, na.rm = TRUE)
22.10499 - 1.5 * (25.83-22.10499) #16.5
```

```
#So, outliers are 22, 35, 39, 45
#Exclude the outliers:
combined_dataset$investment_mean[22] <- NA #KOREA
combined_dataset$investment_mean[35] <- NA #URUGUAY
combined_dataset$investment_mean[39] <- NA #CAPE VERDE
combined_dataset$investment_mean[45] <- NA #MONGOLIA

#Run the regression
multiregression21 <- lm(combined_dataset$investment_mean ~ combined_dataset$epdimension +
                          combined_dataset$hdi_2010 + log(combined_dataset$pop_2009))
multiregression22 <- lm(combined_dataset$investment_mean ~ combined_dataset$epdimension +
                          combined_dataset$hdi_2010 + log(combined_dataset$pop_2009) +
                          combined_dataset$realir_mean)
#R&D
boxplot(combined_dataset$rand_mean)
```

```
#Run the regression
multiregression23 <- lm(combined_dataset$rand_mean ~ combined_dataset$epdimension +
                          combined_dataset$hdi_2010 + log(combined_dataset$pop_2009))

#Inflation Volatility
boxplot(combined_dataset$cpi_mean)
```

```
combined_dataset$cpi_mean
#Outliers are 38 Bulgaria, 45 Mongolia
combined_dataset$cpi_mean[38] <- NA #Bulgaria
combined_dataset$cpi_mean[45] <- NA #Mongolia

#Run the regression
multiregression24 <- lm(combined_dataset$cpi_mean ~ combined_dataset$epdimension +
                          combined_dataset$hdi_2010 + log(combined_dataset$pop_2009))
multiregression25 <- lm(combined_dataset$cpi_mean ~ combined_dataset$epdimension +
                          combined_dataset$hdi_2010 + log(combined_dataset$pop_2009) +
                          combined_dataset$trade_mean)

#Investment
stargazer(multiregression21, multiregression22,
          title="Fifty-Three Democracies' Gross Capital Formation",
          covariate.labels = c("Executive-Parties Dimension", "HDI 2010",
                               "Population 2009 (logged, 100s)",
                               "Mean Real Interest Rate (1980-2010)"),
          dep.var.labels = "Investment as % of GDP (1980-2010)",
          font.size = "small",
          omit.stat = c("f", "ser"),
          column.sep.width = "-15pt",
          out="regressiontablenew1.tex")

#R&D
stargazer(multiregression23, title="Fifty-Three Democracies' R&D Spending",
          covariate.labels = c("Executive-Parties Dimension", "HDI 2010",
                               "Population 2009 (logged, 1000s)"),
          dep.var.labels = "Research and Development Spending as % of GDP (1980-2010)",
          font.size = "small",
          omit.stat = c("f", "ser"),
```

```
          column.sep.width = "-15pt",
          out = "regressiontablenew2.tex")

#Inflation Volatility
stargazer(multiregression24, multiregression25,
          title="Fifty-Three Democracies' Inflation Volatility",
          covariate.labels = c("Executive-Parties Dimension", "HDI 2010",
                               "Population 2009 (logged, 1000s)",
                               "Economic Openness (1980-2010)"),
          dep.var.labels = "CPI Inflation Volatility (1980-2010)",
          font.size = "small",
          omit.stat = c("f", "ser"),
          column.sep.width = "-15pt",
          out = "regressiontablenew3.tex")
```

Finally I reworked the plotting of the original figure displaying the correlation between Lijphart's exec-parties dimension and mine, this time using ggplot.

```
dataupdated <- cbind(data, newepdimension)
#Figure is plotting old E-P dimension against new EP dimension
#ggplot(data=dataupdated, aes(x=exec_parties_1981_2010, y=newepdimension)) +
#  geom_smooth(method="lm", colour="red") +
#  labs(title="Figure 1: Executive-Parties Dimension Correlation",
# subtitle="Excluding Pluralism and new Executive Dominance Measure",
# x="Lijphart's Executive-Parties Dimension (1981-2010)",
# y="Refined Executive-Parties Dimension (1981-2010)") +
#  geom_text(aes(label=country), size=3, colour="black",
# check_overlap = TRUE, family="Times New Roman") +
#  theme_minimal()
```