Using web data to explore the relationship between economic growth, population, and democracy

World Bank API and Polity data case

Arrue, J. and Batas, R.

2019-12-11

Contents

1. Introduction

The study of economic growth has tremendous impact towards a country's policies and regulation. Different governments around the world invest in economic research to help them guide in making decisions. One way that these projects are being conducted is by looking at the relationship between variables that are related to the economy and its growth. These studies often include demographic variables such as race, income, gender, and educational level, to name a few.

It is known that demographic variables are important determinants of economic growth especially when using Gross Domestic Product (**GDP**) as a measure of growth. One of these demographic variables that have been studied extensively is population growth. However, researchers have debated whether this affects a country's economy positively and promotes economic growth or if it actually constraints its growth.

Thomas Malthus a scholar in the field of political economy has developed a theory about the relationship between economic growth and population. He believed that a continuous growth in population will eventually lead to humans outgrowing available resources and stump economic growth (Malthus 1993). This theory has been the focus of several research in the past years questioning the nature of the relationship between the population and economic growth rather than the existence of the relationship itself. To name one example, researchers from the University of Pretoria found different effects of economic growth towards population (Chang et al. 2014). This study included 21 countries where a one-way Bootstrap Panel Granger Causality test was conducted. They have identified five countries where a relationship between population development and economic growth has been established: Canada, Norway, Germany, Japan, and Switzerland. Out of these five countries, Canada, Norway, and Switzerland showed a positive relationship between the two variables being studied. However, they have found the opposite results for Germany and Japan where economic growth has been found to negatively impact population growth.

With the existence of the relationship between the two variables supported by many scholars, it is then important to have a look at this subject. However, when discussing economic growth, one must take into careful consideration the potential effects of confounding variables. One research which has conducted a meta-analysis of published papers showed that democracy has no negative effects on economic growth and has no direct impact to it (Doucouliagos and Ulubasoglu 2007). This implies that when predicting a country's economic growth, democracy is a potential variable to control for. Several papers have controlled for sundry of variables when looking at the relationship between population and economic growth, however, there's a lack for researches controlling for a country's democratic level.

The absence of specific literature on the relationship between population and economic growth and how democracy may affect this relationship merits the conduct of this paper. To do so, this project aims to scrape recent data from the web in order to look at the relationship between population and economic growth, to observe whether there is indeed a correlation between the two, and to look whether a country's democracy would affect impact this relationship as well.

2. Objective, hypothesis and methods

2.1 Objective

The objective of this research is to explore the relationship between GDP, population, and democracy. It is known that the relationship within and between these three variables has never been clear. Recent economic trends have shown that a country's type of regime is not a determinant of economic growth (e.g., China and Russia's latest GDP increase in the past half-century compared to U.S. or western Europe countries). In addition to that, population has not been able to predict economic development as well (e.g., Nigeria, with a population approximately less than 200 million people doesn't have the same GDP as Japan, with a population of approximately 130 million people).

In order to address this, we will extract indicators from two different sources namely: World Bank (**WB**) Indicators, which are available through the World Bank API, will account for GDP, population, and governance indicators, which are related to GDP increase; and PolityIV Project (**Polity**) to determine whether a type of regime is more or less democratic. This last source will account for the democratic index, a measure that determines the type of regime a country has.

2.2 Hypothesis

As an exploratory analysis, the central hypothesis is to test whether there is a relationship between GDP, population, and democracy. In order to test for this, the following questions should be addressed

1. Is there a relationship between GDP and democracy, and how is it?
2. Is there a relationship between GDP and population, and how is it?
3. Can democracy check for significant differences in the relationship between GDP and population?
4. Is governance related to GDP, and how can democracy check for this?

With the literature review, the following hypotheses were developed:

1. Given the latest trends regarding the world's largest economies, the effect of democracy on GDP, which once was considerably high, has diminished.
2. Considering the relationship between GDP and expenditure, we suspect that there is not only a positive relationship between both variables but also a significant one.
3. Even with a small effect for GDP, democracy can check for significant changes in the relationship between GDP and other variables, such as population.
4. Given that good governance is associated with consistent economic growth, we consider that its relationship with GDP is not only positive but also significant. Democracy, as with population, can check for significative differences between this relationship.

2.3 Methods

2.3.1 Data manipulation with Polity

Polity was obtained through https://ourworldindata.org/democracy, an online scientific publication that gathers information on many topics around the world and whose research team is based at Oxford University. In order to obtain Polity data on political regimes, available information was downloaded in a *comma separated value* file (.csv). Political regimes ranges on a scale of -10 *(full autocracy)* to +10 *(full democracy)*. In between values consist of anocracies. In addition, the scale provides information for colonies *(with a value of -20)*. The information on political regimes includes country name, country code, POLITY score, and year. This last variable allows researchers to check for political regimes changes in time. Table 1 provides an example of how POLITY dataset looks like.

Table 1: Polity data frame

Entity	Code	Year	Political.Regime..OWID.based.on.Polity.IV.and.Wimmer...Min...Score.
Afghanistan	AFG	1816	-6
Afghanistan	AFG	1817	-6
Afghanistan	AFG	1818	-6
Afghanistan	AFG	1819	-6
Afghanistan	AFG	1820	-6
Afghanistan	AFG	1821	-6
Afghanistan	AFG	1822	-6
Afghanistan	AFG	1823	-6

In order to merge POLITY with WB data, data manipulation was applied. Country names were deleted, and variables names were modified. In addition, the political regime label given by POLITY was added as well. Table 2 shows how data looks like after reorganization.

Table 2: Polity data frame modified

iso3c	date	INDEX	INDEX_TYPE
AFG	1960	1	Autocracy
AFG	1961	1	Autocracy
AFG	1962	1	Autocracy
AFG	1963	1	Autocracy
AFG	1964	4	Autocracy
AFG	1965	4	Autocracy
AFG	1966	4	Autocracy
AFG	1967	4	Autocracy

2.3.2 Data manipulation with World Bank API

The WB API provides different data frames from which researchers can obtain information. Two of which entails important details relevant to a country, while the rest offers an extensive amount of development indicators and variables. Unfortunately, the indicators information is too extensive to work within R. For instance, indicators have confusing codes, and the labels are often complicated. This makes extracting the indicators data frame into an Excel file (.xls) as the most efficient means. This provides a friendlier environment to search for the desired indicators. Another data frame that was considered was the one with more information on countries. As mentioned before, this data frame provides country names and ISO code, which is used by POLITY. This code, alongside the variable related to year, was utilized to merge WB data with POLITY data. Table 3 gives an example of the information extracted from the World Bank API.

Table 3: World Bank data frame

iso3c	date	iso2c	country	NY.GDP.MKTP.CD	SP.POP.TOTL
ABW	1960	AW	Aruba	NA	54211
ABW	1961	AW	Aruba	NA	55438
ABW	1962	AW	Aruba	NA	56225
ABW	1963	AW	Aruba	NA	56695
ABW	1964	AW	Aruba	NA	57032
ABW	1965	AW	Aruba	NA	57360
ABW	1966	AW	Aruba	NA	57715
ABW	1967	AW	Aruba	NA	58055

As with POLITY, WB extracted data needs to be rearranged and cleaned prior to merging with POLITY. The following table depicts how the final dataset looks like.

Table 4: Polity with World Bank data frames merged

iso3c	region	date	country	GDP_USD	POP	INDEX	INDEX_TYPE
AFG	South Asia	1960	Afghanistan	537777811	8996973	1	Autocracy
AFG	South Asia	1961	Afghanistan	548888896	9169410	1	Autocracy
AFG	South Asia	1962	Afghanistan	546666678	9351441	1	Autocracy
AFG	South Asia	1963	Afghanistan	751111191	9543205	1	Autocracy
AFG	South Asia	1964	Afghanistan	800000044	9744781	4	Autocracy
AFG	South Asia	1965	Afghanistan	1006666638	9956320	4	Autocracy
AFG	South Asia	1966	Afghanistan	1399999967	10174836	4	Autocracy
AFG	South Asia	1967	Afghanistan	1673333418	10399926	4	Autocracy

3. Predicting economic growth by the democracy index

In order to test the first hypothesis, the relationship between economic growth and the type of regime a country has was taken into account. According to literature and actual trends regarding the most significant economies around the world, it was suspected that there was a weak relationship between the two, especially for recent years. For this purpose, information from WB data was utilized to provide information regarding GDP as a measure of economic growth. On the other hand, POLITY data was used to provide information regarding political regimes in the form of an index. To conduct linear regressions, we transformed the ranges from -10 to 10 to ranges from 0 to 21. And to begin with the analyses, the trend for GDP across time was plotted. The trends and changes over time highlighted the presence of some outliers and behavior. Boxplots were used where years were divided by decades for visual presentation.
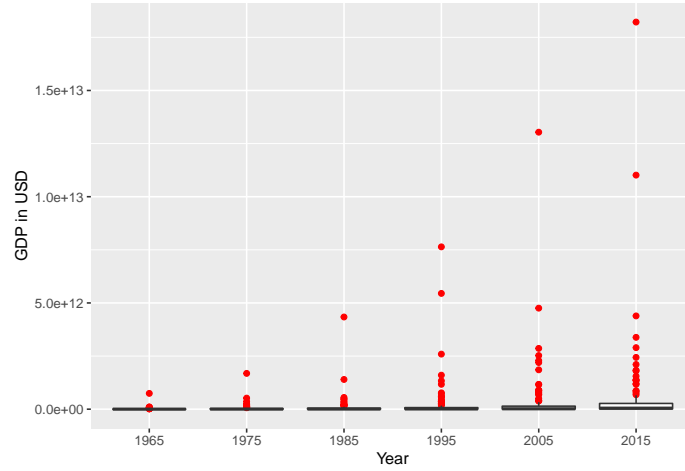
Figure 1: Boxplots: GDP in USD through the years

As seen above, GDP presents issues for visual presentation. For instance, the outliers make distribution of GDP too skewed. Log_{10} transformation was utilized to remedy this and normalize the GDP variable.
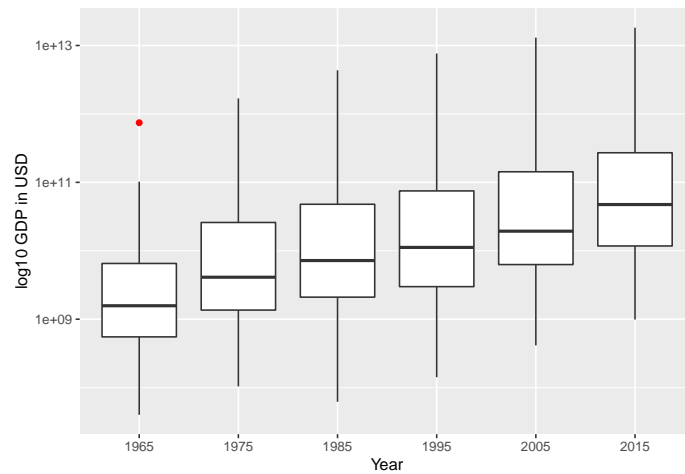

Figure 2: Boxplots: log10 GDP in USD through the years

As Figure 2 suggests, there has been a constant increase in countries' GDP over the years. On the other hand, the GDP variable transformation normalized data and allows the use of a linear regression model, using the democratic index as the predictor and $\log_{10}$ GDP in USD as the dependent variable.
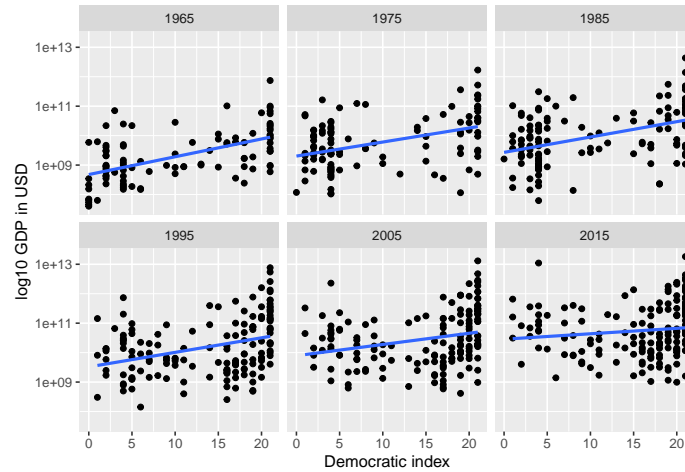
Figure 3: Linear regression model: log10 GDP by democracy index

Summary 1: Linear regression model parameter change through the years. Log_{10} GDP by democratic index

```
## $`1965`
##               Estimate  Std. Error  t value     Pr(>|t|)
## (Intercept) 8.67845203 0.116982428 74.18595 6.485333e-88
## INDEX       0.06038143 0.008992531  6.71462 1.238261e-09
##
## $`1975`
##               Estimate Std. Error  t value     Pr(>|t|)
## (Intercept) 9.30537493  0.1228984 75.715998 2.009680e-95
## INDEX       0.04766865  0.0100670  4.735139 6.687633e-06
##
## $`1985`
##               Estimate  Std. Error  t value      Pr(>|t|)
## (Intercept) 9.43149896 0.124916902 75.502184 6.770844e-105
## INDEX       0.05211178 0.009808588  5.312873  4.881977e-07
##
## $`1995`
##               Estimate Std. Error  t value      Pr(>|t|)
## (Intercept) 9.51456456 0.16884215 56.351833 7.768664e-103
## INDEX       0.04940528 0.01102594  4.480823  1.466178e-05
##
## $`2005`
##               Estimate Std. Error  t value      Pr(>|t|)
## (Intercept) 9.89507372 0.18302529 54.06397 4.706568e-102
## INDEX       0.03811853 0.01139268  3.34588  1.030531e-03
##
## $`2015`
##                Estimate Std. Error  t value      Pr(>|t|)
## (Intercept) 10.45541099 0.19835341 52.711023 7.023343e-101
## INDEX        0.01843161 0.01195618  1.541597  1.252105e-01
```

Table 5: R squared value through the years: GDP by Democratic Index

Year	R squared
	x
1965	0.3150978
1975	0.1719160
1985	0.1866511
1995	0.1180506
2005	0.0677679
2015	0.0151009

As hypothesized, there is a weak relationship between GDP and how democratic a country is. The democracy index effect is significant for all decades ($p<0.05$), except for the last decade. On the other hand, its power to explain GDP variation decreased dramatically through time. When fitting a better line for the regression model relationship between the two can be seen better.
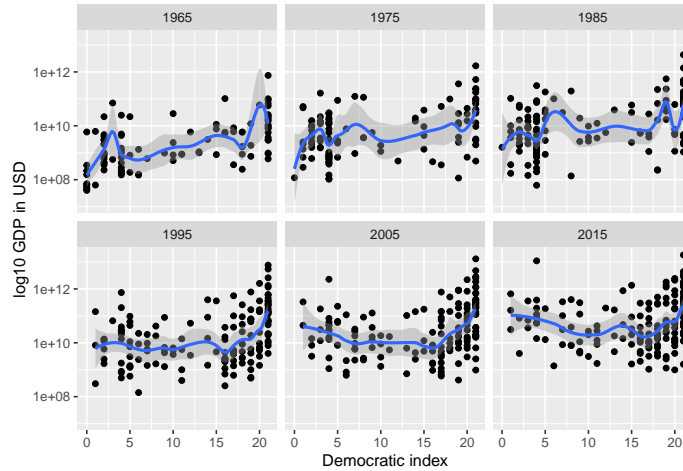

Figure 4: Linear regression model: log10 GDP in USD by democracy index. Fitted relation

In response to the first hypothesis, it can be noted that there is an existing relationship between economic growth and how democratic the country is. However, the relationship is getting weaker over time, or virtually non-existent, especially in the last few years. In addition, the nature of the democratic index might not be useful for a linear regression. For instance, although we transform the democracy index to fit a linear regression model, the index resembles a discrete variable, which might be useful for other types of regression models.

4. Democracy working as a fixed effect model: GDP by population

Given the nature of the democratic index and the results from the simple linear regression model, it can be concluded that democracy is better used to test relationships between other variables. It is highly possible that, as trends have changed with economic growth and political regimes, the latter has lost its power to explain GDP variation on its own. As a second hypothesis, it is suspected that the type or regime a country has can actually account for differences in the relationship between GDP and other variables in this particular test, with population. In order to test this, another simple linear regression model was ran utilizing population as an economic growth predictor. The theory regarding the relationship between these two variables is also unclear. Although having more population can lead to more inner consumption, this is only possible if the population has sufficient incomes. In consequence, even with a huge population, this might not necessarily imply an increase in GDP. To test the relationship between population and GDP, we decided to see whether different types of regimes can account for significant changes in the aforementioned relationship. Given that

the democratic index has up to 21 different values, broader categories were utilized, which can be summarized into 5 different categories. Each one of them, according to the fixed-effect model, would be used as a dummy variable to test if there are any significant differences between population and GDP when the type of regime intervenes. Same as with GDP, the nature of population as variable was also explored.
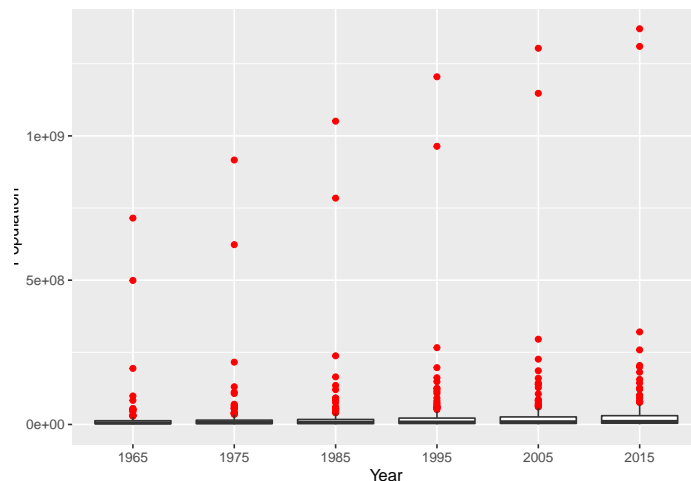

Figure 5: Boxplots: Population through the years

Figure 5 does not only show the presence of multiple outliers through time, but that data for population is very skewed. We proceed to transform the population variable with $\log_{10}$, similar to what was done with GDP. This can be seen in Figure 6.
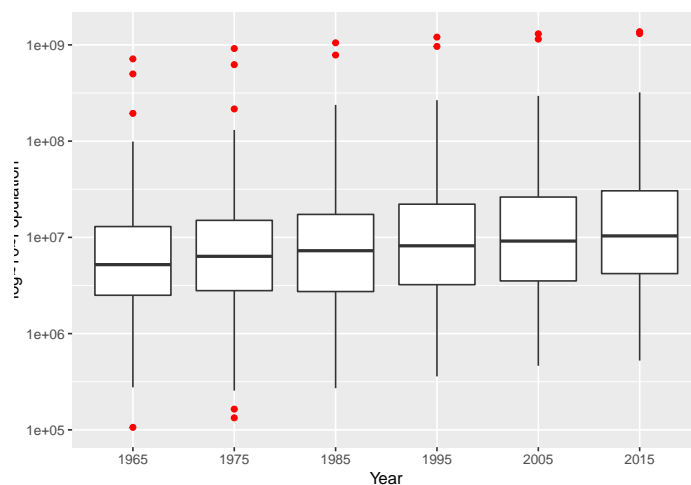

Figure 6: Boxplots log 10 Population through the years

Even though outliers were identified, these were not removed since there were barely any outliers when checking for GDP. It is suspected that this might not have an effect when running a simple regression model between the two variables. If removed, it is suspected that this would have an effect on the determination coefficient (R^2) that might prejudice the regression interpretation. Now we proceed to run our linear regression model, where population will be the predictor and GDP the dependent variable.
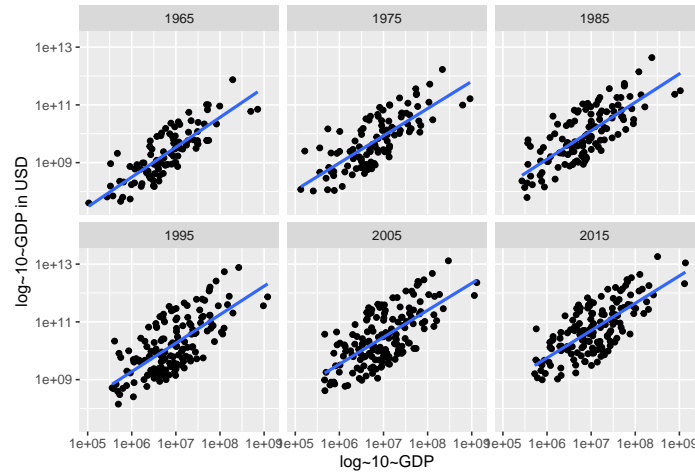
8

Figure 7: Linear regression model: log10 GDP in USD by log10 Population through the years

Summary 2: Linear regression model parameter change through the years. Log_{10} GDP by democratic index

```
## $`1965`
##              Estimate Std. Error    t value       Pr(>|t|)
## (Intercept) 2.290521 0.49550546   4.622596 1.156686e-05
## log10(POP)  1.034672 0.07279962  14.212606 1.514565e-25
##
## $`1975`
##               Estimate Std. Error    t value       Pr(>|t|)
## (Intercept) 3.3355932 0.53189311   6.271172 7.556310e-09
## log10(POP)  0.9420826 0.07761025  12.138637 6.616618e-22
##
## $`1985`
##               Estimate Std. Error    t value       Pr(>|t|)
## (Intercept) 3.2065112 0.52181551   6.144914 1.021482e-08
## log10(POP)  0.9853313 0.07573175  13.010809 7.155288e-25
##
## $`1995`
##               Estimate Std. Error    t value       Pr(>|t|)
## (Intercept) 3.3617327 0.57768463   5.819322 3.452987e-08
## log10(POP)  0.9861138 0.08296879  11.885358 2.253739e-23
##
## $`2005`
##               Estimate Std. Error    t value       Pr(>|t|)
## (Intercept) 3.9314462 0.58356786   6.736914 3.037249e-10
## log10(POP)  0.9338678 0.08313277  11.233449 9.173850e-22
##
## $`2015`
##               Estimate Std. Error    t value       Pr(>|t|)
## (Intercept) 4.0728017 0.52796147   7.714203 1.384411e-12
## log10(POP)  0.9466262 0.07461309  12.687133 9.522845e-26
```

Table 6: R squared value through the years: GDP by Population

Year	R squared
	x
1965	0.6733313
1975	0.5770453
1985	0.5791723
1995	0.4849993
2005	0.4503737
2015	0.5094368

Unlike the linear regression model to explore the relationship between GDP and type of regime, the relationship between GDP by population is more powerful. Not only graphically, but also when interpreting the parameters from the model ($p<0.001$ for all the explored years). On the other hand, the R^2, although diminishing through time, presents a less dramatically dropdown. With that, the relationship was tested for significant differences when the relationship is checked by how democratic the country is. In order to do this, a linear regression model was utilized to look at several years. This time, the effect of the political regime will be highlighted.
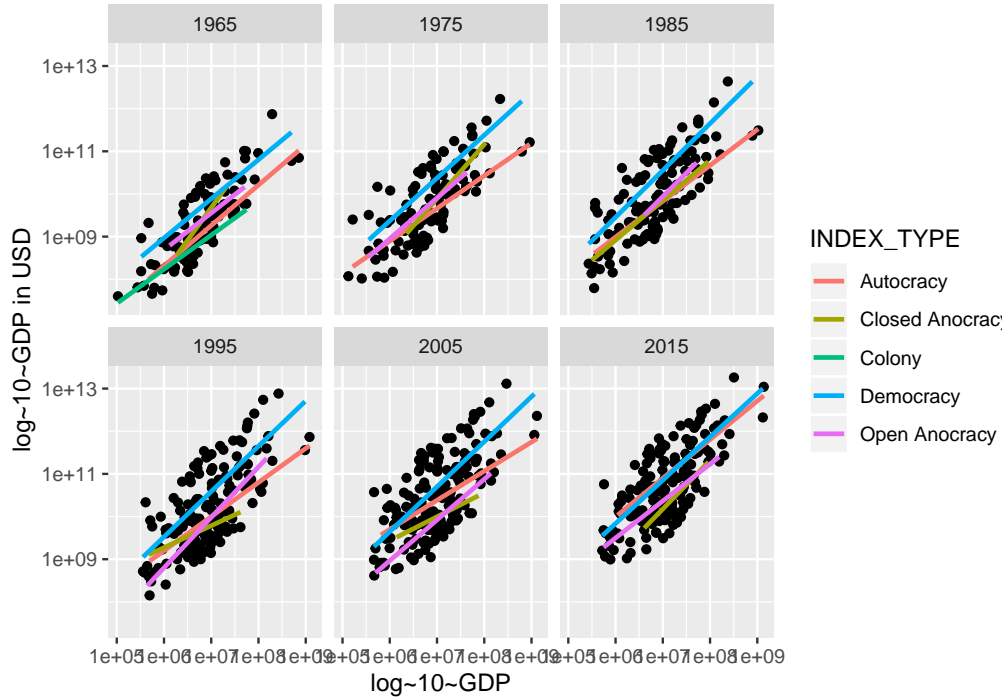

Figure 8: Linear regression model: log10 GDP by log10 Population - Checked by type of regime

As seen above, the relationship between GDP and population is consistent even when controlled by the type of regime. Nonetheless, depending on the type of regime, the effect of population on GDP will be more or less pronounced. To better look at this, a fixed-effect model analysis was utilized.

Summary 3: Fixed effect model: log10 GDP by log10 population, controlled by type of regime. Year 1965

```
##
## Call:
## lm(formula = log10(GDP_USD) ~ log10(POP) + factor(INDEX_TYPE) -
##     1, data = FULL_1965)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.02544 -0.28566  0.00051  0.25628  1.29156
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## log10(POP)                       0.92390    0.06808  13.571  < 2e-16 ***
## factor(INDEX_TYPE)Autocracy      2.78796    0.46469   6.000 3.67e-08 ***
## factor(INDEX_TYPE)Closed Anocracy 2.99920   0.47598   6.301 9.46e-09 ***
## factor(INDEX_TYPE)Colony         2.70147    0.42599   6.342 7.87e-09 ***
## factor(INDEX_TYPE)Democracy      3.41909    0.47810   7.151 1.84e-10 ***
## factor(INDEX_TYPE)Open Anocracy  3.07827    0.50299   6.120 2.14e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4177 on 94 degrees of freedom
## Multiple R-squared:  0.9981, Adjusted R-squared:  0.998
## F-statistic:  8311 on 6 and 94 DF,  p-value: < 2.2e-16
```

Summary 4: Fixed effect model: log10 GDP by log10 population, controlled by type of regime. Year 2015

```
##
## Call:
## lm(formula = log10(GDP_USD) ~ log10(POP) + factor(INDEX_TYPE) -
##     1, data = FULL_2015)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.28878 -0.37328 -0.03397  0.40938  1.53763
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## log10(POP)                       0.97911    0.06853  14.287  < 2e-16 ***
## factor(INDEX_TYPE)Autocracy      4.04479    0.49737   8.132 1.38e-13 ***
## factor(INDEX_TYPE)Closed Anocracy 3.34868   0.50595   6.619 5.85e-10 ***
## factor(INDEX_TYPE)Democracy      3.99864    0.48422   8.258 6.70e-14 ***
## factor(INDEX_TYPE)Open Anocracy  3.49832    0.50170   6.973 8.89e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5732 on 152 degrees of freedom
## Multiple R-squared:  0.9973, Adjusted R-squared:  0.9972
## F-statistic: 1.107e+04 on 5 and 152 DF,  p-value: < 2.2e-16
```

As seen in the results for the year 1965, being a democracy has a significant difference among other types of regimes. However, based on the 2015 results, it can be seen that the effect of democracy has is slightly smaller than autocracies. This is consistent with the linear regression model graphs presented before. With

this analysis, it can be concluded that when using the political regime as a discrete variable, differences between two different variables can be accounted for. In this particular case, for GDP and population.

5. Testing democracy in a multiple linear regression model

Given that our results showed that the different types of regimes could account for significant differences regarding the relationship between GDP and population, testing the effect using a multiple linear regression model would be ideal. For this, GDP was used as the predicted variable since the main scope of our work is to account for the effect of democracy on GDP. On the other hand, variables related to governance as predictors were utilized. As mentioned earlier, one key element for GDP development is government effectiveness, which guarantees a structure for constant economic growth by providing the minimal assurance to promote investment, consumption, and law enforcement. For this purpose, WB data offers a series of variables which will come in handy:

1. Voice and Accountability (**VA**)
2. Political Stability and Absence of Violence (**RV**)
3. Government Effectiveness (**GE**)
4. Regulatory Quality (**RQ**)
5. Rule of Law (**RL**)
6. Control of Corruption (**CC**)

It can be expected that governance variables will fit well in the multiple linear regression model and will have a significant effect on GDP increase. In addition, controlling the relationship by type of regime will highlight significant differences in the aforementioned relationship. For this purpose, data manipulation was applied with POLITY and WB data once again. Table 7 shows how data looked like.

Table 7: Polity with World Bank data frames merged, including governance variables

	iso3c	region	date	country	CC	GE	GDP_USD	PV	RL	RQ	POP	VA	INDEX	INDEX_TYPE
1	AFG	South Asia	2014	Afghanistan	-1.354784	-1.331396	20484873230	-2.411068	-1.445648	-1.1181820	33370794	-1.135445	10	Closed Anocracy
2	AFG	South Asia	2015	Afghanistan	-1.341994	-1.322968	19907111419	-2.571222	-1.504040	-0.9973711	34413603	-1.117563	10	Closed Anocracy
3	AGO	Sub-Saharan Africa	1996	Angola	-1.167702	-0.859940	7526446606	-2.057286	-1.630469	-1.4153450	14400719	-1.578164	9	Closed Anocracy
5	AGO	Sub-Saharan Africa	1998	Angola	-1.405836	-1.341783	6506229607	-2.313072	-1.700753	-1.6896590	15359601	-1.411879	8	Closed Anocracy
7	AGO	Sub-Saharan Africa	2000	Angola	-1.522685	-1.462310	9129594819	-2.035583	-1.663307	-1.7993310	16395473	-1.458065	8	Closed Anocracy
9	AGO	Sub-Saharan Africa	2002	Angola	-1.183629	-1.238857	15285594828	-1.577046	-1.617329	-1.4631600	17519417	-1.238779	9	Closed Anocracy
10	AGO	Sub-Saharan Africa	2003	Angola	-1.320665	-1.156531	17812705294	-1.006538	-1.521410	-1.2121440	18121479	-1.272615	9	Closed Anocracy
11	AGO	Sub-Saharan Africa	2004	Angola	-1.313559	-1.314578	23552052408	-1.062798	-1.496279	-1.2515390	18758145	-1.285480	9	Closed Anocracy

After preparing the dataset, a multiple linear regression model was run. Since the information for this dataset is only available since the year 1996, four models were run, each one for the years 2000, 2005, 2010 and 2015.

Summary 5: Multiple linear regression model parameter change through the years. Log_{10} GDP by governance indicators

```
## $`2000`
##              Estimate Std. Error    t value      Pr(>|t|)
## (Intercept) 10.26012612  0.0595288 172.3556767 1.600646e-171
## CC          -0.55016235  0.1891093  -2.9092296  4.186448e-03
## GE           1.32179656  0.2470309   5.3507334  3.290924e-07
## PV          -0.33770335  0.1012033  -3.3368795  1.073103e-03
## RL           0.28916484  0.2392486   1.2086378  2.287426e-01
## RQ          -0.16403829  0.1771936  -0.9257575  3.560893e-01
## VA          -0.03599897  0.1224115  -0.2940815  7.691104e-01
##
## $`2005`
##             Estimate Std. Error   t value      Pr(>|t|)
## (Intercept) 10.4535229 0.05430624 192.492126 1.623590e-180
## CC          -0.3507497 0.20010066  -1.752866  8.168167e-02
## GE           1.3203004 0.26332268   5.014002  1.494970e-06
## PV          -0.3370274 0.09365778  -3.598499  4.349320e-04
## RL          -0.3035066 0.25942198  -1.169934  2.438956e-01
## RQ           0.3380498 0.20378140   1.658884  9.924266e-02
## VA          -0.1883828 0.10894275  -1.729190  8.584615e-02
##
## $`2010`
##              Estimate Std. Error    t value      Pr(>|t|)
## (Intercept) 10.67714791 0.05833299 183.0378904 3.422260e-179
## CC          -0.54278040 0.19638753  -2.7638231  6.424336e-03
## GE           1.01471410 0.22394141   4.5311588  1.184866e-05
## PV          -0.35809624 0.09099612  -3.9352913  1.264168e-04
## RL           0.45894073 0.28306882   1.6213044  1.070387e-01
## RQ          -0.14344892 0.17661419  -0.8122163  4.179462e-01
## VA          -0.07171273 0.09995467  -0.7174525  4.742034e-01
##
## $`2015`
##              Estimate Std. Error    t value      Pr(>|t|)
## (Intercept) 10.67255510 0.04885147 218.4694754 1.022980e-189
## CC          -0.45704253 0.18343013  -2.4916438  1.380371e-02
## GE           1.46153045 0.17263596   8.4659677  2.142903e-14
## PV          -0.49204495 0.07839356  -6.2765994  3.531341e-09
## RL           0.12080275 0.27698136   0.4361404  6.633619e-01
## RQ          -0.32536765 0.16017138  -2.0313720  4.398293e-02
## VA           0.02743739 0.08299930   0.3305738  7.414277e-01
```

It can be noticed that for the years explored, the effect of governance variables is not the one we expected. It can be drawn that:

1. Not all of them have positive effects on GDP.
2. Not all of the parameters for each predictor variable is actually significant for the model.
3. Government Effectiveness seems to have a more significative effect when compared to all other governance variables, which is consistent and significant for all years.

This is in contrast to recent literature where it is suggested that certain conditions, like good governance, actually foster an increase in GDP and not necessarily a decrease in it. R^2 was also calculated to see how much of the variance of GDP can be explained by governance variables.

Table 8: R squared value through the years: GDP by governance variables

Year	R squared
	x
1996	0.4834428
1998	0.4848506
2000	0.4863922
2002	0.4925589
2003	0.5201595
2004	0.5619809
2005	0.5425662
2006	0.4946142
2007	0.4852471
2008	0.4607376
2009	0.4681741
2010	0.4583025
2011	0.4308168
2012	0.4302861
2013	0.4713283
2014	0.5465282
2015	0.5812316

The R^2 is actually high. However, it is not consistent through the years, which indicates that the variance on GDP through time is not necessarily due to or in accordance with governance-related indicators. Although results suggest that a different model has to be utilized by now, regime type was still used as a fixed effect in order to see any significant differences in the relationship between governance and GDP.

Summary 6: Fixed effect model: log10 GDP in USD by log10 population, controlled by type of regime

```
##
## Call:
## lm(formula = log10(GDP_USD) ~ CC + GE + PV + RL + RQ + VA + factor(INDEX_TYPE) -
##     1, data = FULLGOV)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.41735 -0.45587 -0.02837  0.45537  2.04448
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## CC                               -0.45376    0.04651  -9.756  < 2e-16 ***
## GE                                1.11156    0.05533  20.090  < 2e-16 ***
## PV                               -0.39921    0.02353 -16.969  < 2e-16 ***
## RL                                0.21822    0.06401   3.409 0.000661 ***
## RQ                               -0.03665    0.04450  -0.824 0.410213
## VA                               -0.03965    0.04644  -0.854 0.393266
## factor(INDEX_TYPE)Autocracy      10.69624    0.05928 180.424  < 2e-16 ***
## factor(INDEX_TYPE)Closed Anocracy 10.33684   0.04374 236.337  < 2e-16 ***
## factor(INDEX_TYPE)Democracy      10.54330    0.02495 422.550  < 2e-16 ***
## factor(INDEX_TYPE)Open Anocracy  10.40679    0.04035 257.943  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6894 on 2644 degrees of freedom
## Multiple R-squared:  0.9958, Adjusted R-squared:  0.9958
## F-statistic: 6.222e+04 on 10 and 2644 DF,  p-value: < 2.2e-16
```

Results on the fixed-effect model for years 1996 through 2015 showed two interesting findings:

1. When accounting for the governance variables, only two of them are not significant.
2. The effect of regime type as a fixed-effect is showing no significant differences in the relationship between GDP and governance.

In conclusion, we can say that regime type is not suitable to address the differences between governance variables and GDP. We suspect that this does not necessarily occur due to the variable nature of the type of regime but by the possible high correlation between governance and democracy, causing a collinearity effect that was not anticipated.

6. Discussion and conclusions

The study scope was merely exploratory. However, as suspected, democracy is not suitable to explain GDP variance. In the past, its effect was already small and has continued to decrease. There has been a significant shift when it comes to the world's largest economies, where countries such as China, Russia, Saudi Arabia, among others, are showing that the relationship between democracy and economic growth is no longer viable. It would have been more interesting to account for the differences these economies can show if selected for a regression model.

On the other hand, it was fascinating to see how the type of regime a country has can account for differences in the relationship between different variables, one of them being GDP. Regarding population and GDP, the type of regime showed that there were some significant changes depending on how democratic the country is. Consistent with the shift, every year, it is losing its power as an intervening variable, especially for explaining differences regarding full democracies against autocracies

Regarding governance, it seems that variables regarding how much a country can guarantee the stability for economic growth are not significant for our multiple regression model or may not be not providing the desired positive effect. Although possible collinearity problem regarding governance variables and type of regime has been accounted for, the latter didn't prove to be useful when used to check for differences in the relationship between governance and GDP.

7. Limitations

It is essential to acknowledge the WB API as a source of invaluable information, which is continuously updated. However, it is also important to take into account that much information is missing, not only through time but also for certain countries. When using its data to compare different countries in general, some bias may arise. On the other hand, the WB indicator data set contained more than 17,000 variables. It would have been more beneficial to run a component analysis or an exploratory factor analysis in order to consider more variables that were related to GDP instead of only the ones noted in literature.

In addition, the process of merging two different data sets ended up being troublesome than beneficial, mostly because of the decision to take one of them as the base data set from where all the merging would happen. This meant losing even more information, and since governance variables as predictors were an interest, cases that only contained both governance and GDP were intentionally selected, disregarding several GDP single cases, as well as several governance single cases. This could have affected our multiple linear regression model.

Finally, the lack of statistical skills when it comes to logistic regression by the researchers has to be recognized. For instance, due to the nature of the governance variables, as well as the type of regime variable, it would have been better to run a logistic regression model rather than a linear one. Nonetheless, due to R's usefulness, it will be possible to re-run the model once a better understanding of a more appropriate and powerful statistical method has been obtained by the researchers.

8. References

Chang, Tsangyao, Hsiao-Ping Chu, Frederick Deale, Rangan Gupta, and Stephen Miller. 2014. "The Relationship Between Population Growth and Standard-of-Living Growth over 1870ñ2013: Evidence from a Bootstrapped Panel Granger Causality Test." *Empirica* 44 (February). https://doi.org/10.1007/s10663-016-9315-9.

Doucouliagos, Chris, and Mehmet Ulubasoglu. 2007. "Democracy and Economic Growth: A Meta-Analysis." *PSN: Econometrics, Polimetrics, & Statistics (Topic)* 52 (September). https://doi.org/10.2139/ssrn.1014333.

Malthus, T. R. 1993. *Essay on the Principle of Population.* https://ebookcentral.proquest.com/lib/umichigan/detail.action?docID=3008464.