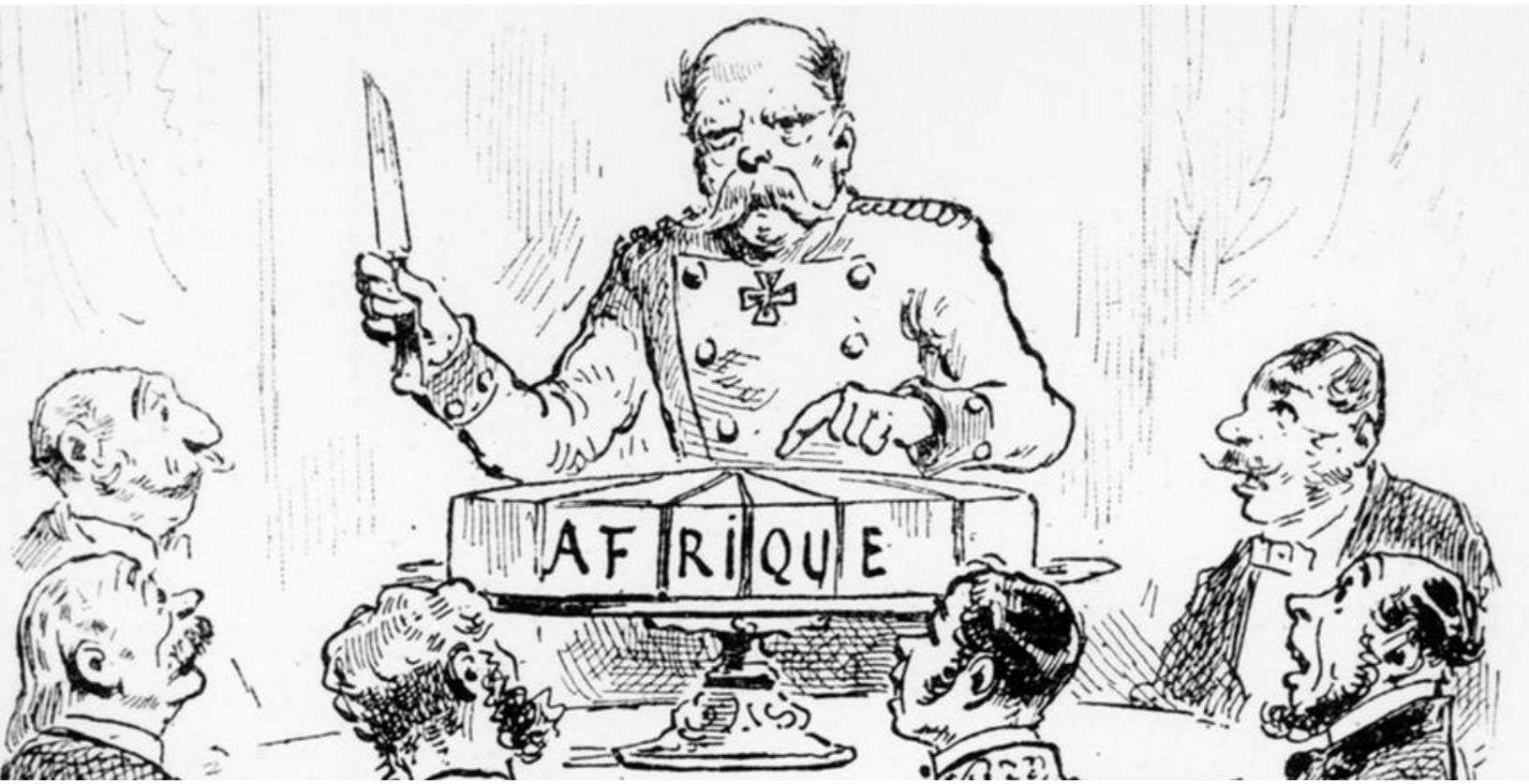


THE LINGERING EFFECT OF COLONIALISM: AN ANALYSIS OF CONTEMPORARY ECONOMIC DEVELOPMENT IN THE INTERNATIONAL WORLD

Introduction

Globalization – the flow of resources, technology, information, and employment across borders and cultures – is a historical phenomenon beginning in the 17th century with complex and politically charged effects [1]. The immense inequality observed today is the path-dependent outcome of historical processes developed in the effort to expand international trade. European colonialism is the most prevalent example of globalization's capacity to generate economic growth at the expense of indigenous populations. The lingering effects of colonialism impact both the economic potential of colonizing countries and colonized societies, leaving various institutional legacies with divergent consequences for economic development across the world [2]. The value of this research is found in the need to remain aware of our past and its effects, especially because of our distance from that era [3]. This need leads us to our main research question: *Does the effect of colonialism exist in the modern day as reflected in a country's economic performance?* As an extension, our three focus questions were: *Do countries that gained independence after 1930 have a lower GDP per capita? Do countries which exist in regions of the world that contain higher numbers of colonies (experienced higher amounts of colonialism) have a lower GDP per capita? Do countries allied with democratic states (adopted the popular western mode of government) have a higher GDP per capita?*

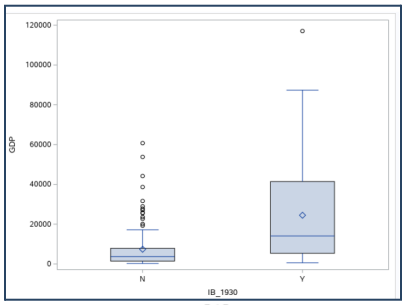
Data Summary

Information was gathered from the Central Intelligence Agency (CIA), World Bank, North Atlantic Treaty Organization (NATO), and the International Institute for Democracy and Electoral Assistance (IDEA). Our research data is constructed around a sample population of all 194 countries as recognized by the CIA using information for the most recent full year the data is available, 2020 [3]. Although some countries may have some missing data entries, the report remains robust because of their infrequency and the sizable number of complete observations [Appendix A].

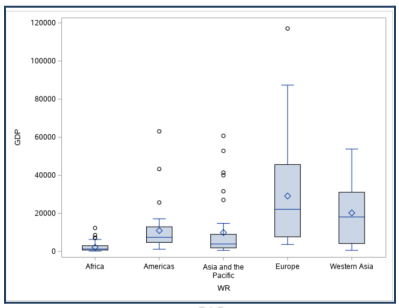
Methods and Analysis

Exploratory Data Analysis (EDA)

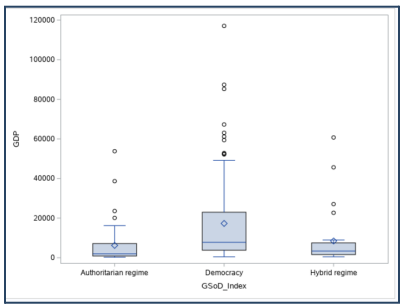
For the EDA, we explored each variable's correlations and relationship with the response variable prior to model building. To highlight the most important revelations, below are the boxplots illustrating our qualitative variables of relevance to our research questions.



[1]



[2]



[3]

Founded on the assumptions that colonialism prevents early independence, occurs outside of Europe, and often leads to a state of authoritarianism, the boxplots demonstrate a possible relationship between a history of colonialism and GDP. The large spread of the GDP values of "Yes" from Plot 1, "Europe" from Plot 2, and "Democracy" from Plot 3 show that the higher GDPs often align with countries without a history of being colonized. Regarding the quantitative variables, the EDA includes testing the Pearson correlation coefficients [App. B]. The variables Total Population and Unemployment Rate had low correlation values with GDP, while Life Expectancy and Average Population Age had high values making the latter likely candidates for inclusion in the model.

The EDA concluded with transforming the response variable in order to fix the original right-skew and normalize the distribution so that the regression assumptions could be met [App. B]. After trying a logarithmic, inverse, and square root transformation, the logarithmic transformation was most suitable. As a result, our final model is now in terms of *log(GDP)*.

Multiple Linear Regression

After EDA, model building commences. We first tested our variables for multicollinearity. No VIF was greater than 10, yet the average VIF of 2.09 causes a slight concern and necessitates variable screening [App. C]. After moving through Stage 1 (quantitative with variable screening) and Stage 2 (qualitative with t-tests and nested t-tests) testing, the model is as follows:

$$E(lgdp) = \beta_0 + \beta_1 APA + \beta_2 LE + \beta_3 dummyexport + \beta_4 dummynato + \beta_5 dummyasia + \beta_6 dummyafrica + \beta_7 dummyameri + \beta_8 dummyeruo.$$

We did not find it necessary to include any interactions and therefore did not test for them within the model. The removal of Population Density/Unemployment Rate and inclusion of Average Population Age/Life Expectancy was not surprising because of the observations about their Pearson correlation coefficients. The exclusion of the Global State of Democracy and Independence before 1930 variables was surprising due to our expectations derived from the history and political structure of our countries, but illustrated their relative insignificance predicting *log(GDP)* and thus GDP.

Using this model, all the regression assumptions were met. The residuals were spread out, centered around 0, normally distributed, and uncorrelated [App. D]. Next, we conducted an influential observation analysis resulting in the removal of 7 outliers: Bahamas, Equatorial Guinea, Ireland, Lebanon, Luxembourg, Nigeria, and Singapore. There was no clear trend amongst the outliers, yet the majority of them are smaller countries with exceptionally large *log(*GDP*)*s.

The final added technique in the model building process, External Model Validation, assessed how the fitted regression model would perform in practice when applied to new data. The technique used for the purpose of this project was data-splitting (cross validation) since collecting new data is impossible and the current observations use the most relevant international information. 50% of the original sample data was split into a training dataset and the rest 50% was split into a testing dataset. The estimation and prediction data sets are of relatively equal size and meet the qualification requirements: the entire sample should consist of at least n = 2k + 25 observations (41), where k is the number of beta parameters in the model. Using cross validation, the relatively high predictive power of the model is confirmed, concluding the analysis with an R-squared value of 0.84.

Results

Our final regression equation is as follows:

$$\hat{lGDP} = 0.842 + 0.082APA + 0.090LE - 0.460Dummyexport - 0.389Dummynato - 0.880Dummyasia - 0.498Dummyafrica - 0.729Dummyameri - 1.113Dummyeuro$$

The final model had a p-value of <0.0001, meaning there was sufficient evidence to claim that our model is adequate in predicting the *log(GDP)*. With an adjusted R-squared of 0.839, we can claim that 84% of the variability in *log(GDP)* is explained by our regression line.

The model appears to express that membership in NATO and geographical location decrease GDP per capita. However, most NATO members and European countries do not actually have

comparatively low *log(*GDPs*)* per capita, as the negativity of the coefficients may indicate at first glance. Instead, the coefficients imply that if there were two countries, one in Europe and one in Africa with otherwise identical values for the other variables, the African country is expected to have the higher *log(GDP)* per capita.

Weight: REWEIGHT

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	272.71285	34.08911	110.03	<.0001
Error	159	49.26155	0.30982		
Corrected Total	167	321.97440			

Root MSE	0.55662	R-Square	0.8470
Dependent Mean	8.53169	Adj R-Sq	0.8393
Coeff Var	6.52410		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.84218	0.82278	1.02	0.3076
dummyexport		1	-0.45954	0.09161	-5.02	<.0001
dummynato		1	-0.38854	0.15879	-2.45	0.0155
dummyasia		1	-0.87978	0.20353	-4.32	<.0001
dummyafrica		1	-0.49785	0.22097	-2.25	0.0256
dummyameri		1	-0.72907	0.20488	-3.56	0.0005
dummyeuro		1	-1.11334	0.24925	-4.47	<.0001
Average_Population_Age	Average_Population_Age	1	0.08191	0.01161	7.05	<.0001
Life_Expectancy	Life_Expectancy	1	0.09008	0.01224	7.36	<.0001

We applied our MLR equation twice in order to test its accuracy and predict the *GDP* per capita of two countries. To find the GDP, we had to raise the original answer to the e and reverse the log transformation. The first country, Samoa, was randomly selected. Our regression equation predicted a *GDP* per capita of 3928.135, which has a residual value of 139.68 when compared to its actual GDP per capita of 4067.815. For the second country, we selected Ghana since it relates directly to our research question as a formerly colonized country. We predicted a GDP per capita of 1785.404, which has a residual value of 440.102 when compared to Ghana's actual GDP per capita of 2225.506. Both examples prove that our model prediction is relatively accurate and competitive.

Conclusion

In many ways, our research and final model mirror the scope and variety of the lasting effects of colonialism. Our formulation of a statistically significant model for *log(*GDP*)* per capita using basic national statistics shows the disadvantages that formerly colonized countries face. In terms of our research question, whether the lingering effects of colonialism are directly evident in these statistics is a harder question to answer. A more accurate interpretation of the model would include the exponentiation of our coefficients to reverse the effect of the log transformation on the response variable. While this process was undertaken in the prediction examples to generate real GDP instead of *log(*GDP*)* values, the process in the analysis of model variables is more complex. The sign of the coefficients would remain the same with a transformation – therefore our conclusions above remain true – but the amount of change in GDP would be altered through this process.

We also recognize limitations in our research. First, altering the measurement of variables currently in our dataset could lead to more accurate conclusions. For example, the power of each country's passport could be a more accurate measure of citizens' freedom than the nation's form of government. Additionally, as alluded to above, our current model is limited in its ability to answer our research questions since the final model didn't include two of the three variables of interest (Independence before 1930 and mode of government, represented by Global State of Democracy). Further analysis could be conducted with our explanatory variables in isolation, in order to measure how closely they mirror former imperialism. Finally, there are a number of possibilities for further research pertaining to our research question. One lens through which to view future research is a study of the natural resources of countries and their effect on GDP per capita. Although countries with greater natural resources will typically have a higher GDP per capita, countries that were exploited through colonization may be exceptions to this trend.

Through this and future research, we can ultimately alleviate the current global wealth disparity by orienting solutions towards negating the specific factors of colonialism. As countries continue to seek new avenues of encouraging economic growth, we hope our findings evidence the need for more equitable systems of development.

References

[1] “What Is Globalization?” *PIIE*, 17 Aug. 2022, https://www.piie.com/microsites/globalization/what-is-globalization.

[2] Acemoğlu, Daron, and James Robinson. “The Economic Impact of Colonialism.” *CEPR*, 30 Jan. 2017, https://new.cepr.org/voxeu/columns/economic-impact-colonialism.

[3] Stinson, Paul. “Texas Senate Votes to Remove Required Lessons on Civil Rights.” *Bloomberg Law*, 17 July 2021, https://news.bloomberglaw.com/social-justice/texas-senate-votes-to-remove-required-lessons-on-civil-rights.

Dataset Sources Not Referenced in Report

[4] “Independence.” *Central Intelligence Agency*, Central Intelligence Agency, https://www.cia.gov/the-world-factbook/field/independence/.

[5]“Median Age.” *Central Intelligence Agency*, Central Intelligence Agency, https://www.cia.gov/the-world-factbook/field/median-age/country-comparison.

[6] “Unemployment, Youth Ages 15-24.” *Central Intelligence Agency*, Central Intelligence Agency, https://www.cia.gov/the-world-factbook/field/unemployment-youth-ages-15-24.

[7]“GDP per Capita (Current US$).” *Data*, https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?end=2020&start=1960.

[8]“Population, Total.” *Data*, https://data.worldbank.org/indicator/SP.POP.TOTL.

[9]“Life Expectancy at Birth, Total (Years).” *Data*, https://data.worldbank.org/indicator/SP.DYN.LE00.IN.

[10]“Top Exporters and Importers.” *Top Exporters and Importers by Country and Region 2020 | WITS Data*, https://wits.worldbank.org/CountryProfile/en/Country/WLD/Year/LTST/TradeFlow/EXPIMP.

[11]Nato. “Member Countries.” NATO, https://www.nato.int/cps/en/natohq/nato_countries.htm.

[12]“Data Set and Resources.” International IDEA, https://www.idea.int/gsod-indices/dataset-resources.

[13] *Databank*, https://databank.worldbank.org/metadataglossary/statistical-capacity-indicators/series/5.51.01.10.gdp#:~:text=GDP%20per%20capita%20is%20the,divided%20by%20mid%2Dyear%20population.

[14] “Finance & Development.” *Finance & Development | F&D*, https://www.imf.org/external/pubs/ft/fandd/basics/gdp.htm.

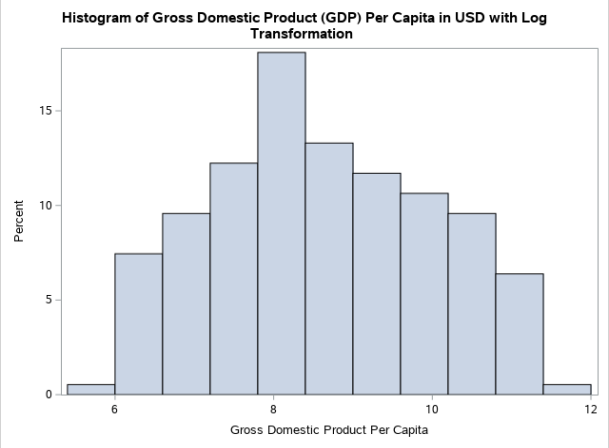
[15] Sen, Kunal. *Unu-Wider : Blog : Why Should I Care about Economic Growth?* Jan. 2021, https://www.wider.unu.edu/publication/why-should-i-care-about-economic-growth.

Appendix

A. Below is the complete codebook for the *project.countrygdp* data file, including where each variable of data was found.

Name (Name in Code)	Description	Units/Levels	Source
GDP per capita (GDP)	Gross domestic product is a valuation of the total goods and services produced by a country in a single year (2020) divided by the country's population.	USD ($)	World Bank
World Region (WR)	Categorical geographical locations of countries in the world.	Asia and the Pacific, Europe, Africa, Americas, Western Asia	NATO
Total Population (Pop_2020)	The total number of people residing in a country (2020).	People	World Bank
Average Population Age (APA)	Average age of a citizen of each country (2020).	Years	CIA factbook
Life Expectancy (LE)	Based on national mortality rates, average lifespan of a person born in each country (2020).	Years	World Bank
Independence before 1930 (IB_1930)	Whether each country was self-governing before 1930.	Y: independent before 1930 N: not independent before 1930	CIA factbook
Membership in NATO (NATO)	Whether each country is a member of the North American Trade Organization (2020).	Y: NATO member N: NATO non-member	NATO
Global State of Democracy Index (GSoD_Index)	Generalized description of the form of government for each nation.	Authoritarian Regime, Democracy, Hybrid	IDEA
Major Importer or Exporter (MIE)	Whether the value of each country's imports in 2020 is greater than its exports, or vice versa. *(solved by using intermediate data: Imports - Exports ; if > 0, then major importer, if <0, then major exporter)*	Major Exporter: more exports than imports Major importer: The country imports more than it exports	World Bank **intermediate data was found here*
Unemployment Rate (UR)	Percentage of the population that is unemployed as of 2020, standard measurement of unemployment.	Percent of total population	CIA factbook

B. Exploratory Data Analysis (EDA) Figures



The CORR Procedure

1 With Variables:	GDP
4 Variables:	Pop_2020 APA LE UR

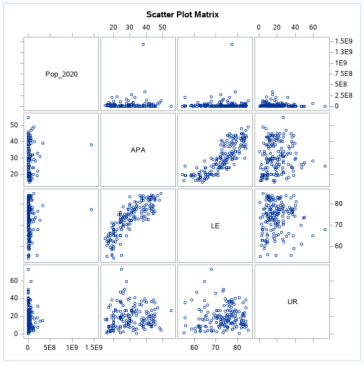
Pearson Correlation Coefficients
Number of Observations

	Pop_2020	APA	LE	UR
GDP	-0.00698 186	0.61772 182	0.64323 180	-0.12909 168

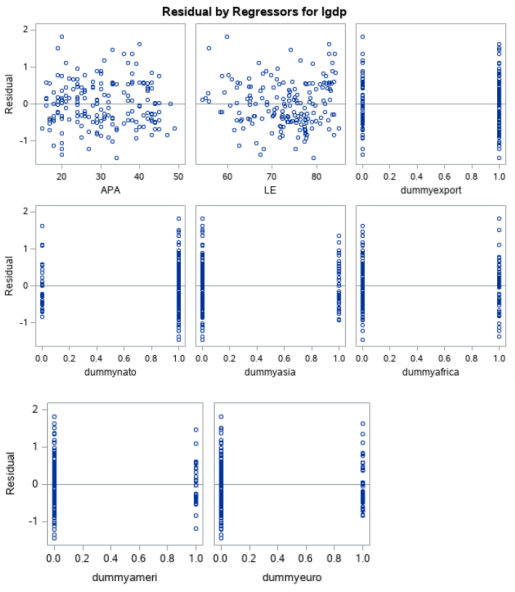
C. Multicollinearity/VIF Test

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-0.67448	0.73185	-0.92	0.3582	0
APA	1	0.06516	0.01021	6.38	<.0001	3.18374
LE	1	0.09994	0.01298	7.70	<.0001	3.17117
Pop_2020	1	2.54234E-10	1.038783E-9	0.24	0.8070	1.01861
UR	1	-0.00225	0.00431	-0.52	0.6022	1.02660



D. Regression Assumptions



The REG Procedure
Model: MODEL1
Dependent Variable: lgdp

Durbin-Watson D	2.228
Pr < DW	0.9344
Pr > DW	0.0656
Number of Observations	175
1st Order Autocorrelation	-0.116