

Jonathan Nease

PA3_3

PA3_1

I used the World Bank API from <https://data.worldbank.org/>

This is a very large dataset with multiple data sources and over 14 hundred data series between them, covering every country (with regional and global averages) from 1960-present.

I used the python wbgapi package to query, explore, and conduct data calls from the API.

Initially I tried to pull a list of several series of data across a 50 year timeframe, but this resulted in both long delays (several minutes before cancelling) or 400-errors. I tried breaking this up into smaller datasets but was having difficulty organizing the output in a way that would make it easy to merge and then build visualizations from. So I reduced the calls for the list to one year increments (from 2010-2020) adding each to its own dataframe, resulting in 11 data calls from the API. Then I added an additional 'Year' column to each by-year dataframe corresponding to the year of the pull. Finally, since I organized the output for each dataframe in a way that would make merging simple, I concatenated each dataframe into a single dataframe that I then exported to csv.

PA3_2

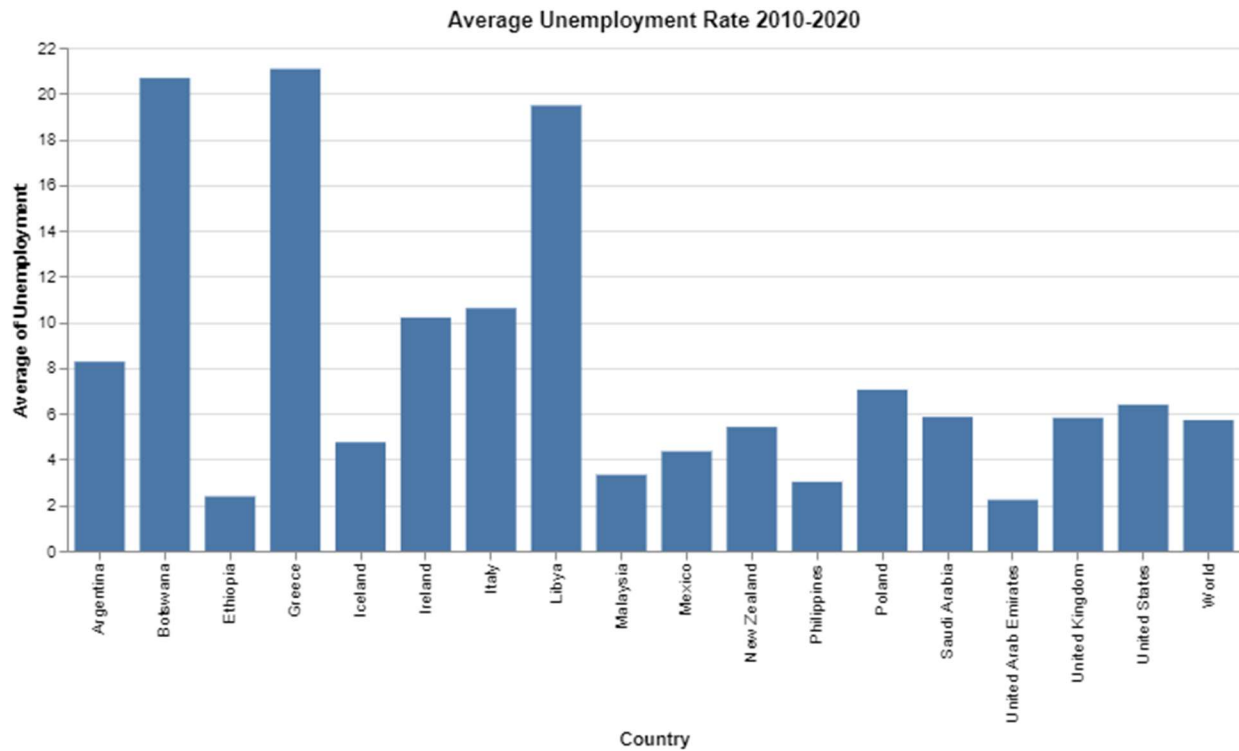
For the next segment of the assignment, I was running into difficulty working with such a large dataset, so I removed many of the country/regional entries (still leave nearly 200 rows to work with), to leave several demonstrative countries (Argentina, Botswana, Ethiopia, Greece, Iceland, Ireland, Italy, Libya, Malaysia, Mexico, New Zealand, Philippines, Poland, Saudi Arabia, UAE, UK, USA) and a global average.

Then to convert the necessary amount of columns for the assignment, since the type of data was correctly assessed by Altair from the pandas conversion, I decided to build several new ratios of columns that were pulled from the API. These were all done by building a new column based on an input formula from two existing columns:

- 1) Rate of slum dwelling % of national population : national unemployment rate – this might provide an indication whether or not some countries are more susceptible to poor urban living conditions based on their unemployment rate, regardless of any additional factor.
- 2) Inflation as a % of GDP (Inflation/Per Capita GDP) – provided an additional metric for measuring economic factors and comparing the humanitarian indicators that were pulled from the API.
- 3) Adjusting per capita GDP as applied to only employed persons (unemployment rate/per capita GDP)/1000 – this provided a slightly different GDP rate for only those employed in the population to possibly compare future metrics in the group project that measure social services
- 4) Ratio of infant mortality : live birth rate – ratio might have proven interesting to see if a correlation existed between these two factors and if so under what circumstances.
- 5) Rate of bribery/inequality – ratio might have provided insight into a correlation between these two factors, regardless of other economic or social metrics.

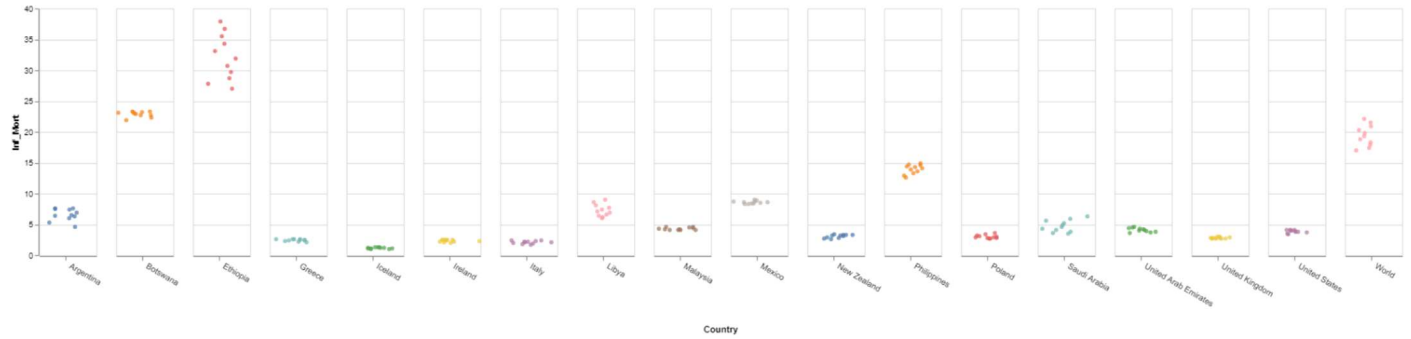
Visualization 1:
Histogram of Average Unemployment by Country from 2010-2020

This histogram took each country and the world average from the input CSV, and averaged out the annual unemployment rate. This chart show how several countries averaged very high rates of unemployment for the time period (Botswana, Greece, Libya) with most coming within 4% of the global average.



Visualization 2: Strip Plot of Sample Countries' Infant Mortality Rates 2010-2020

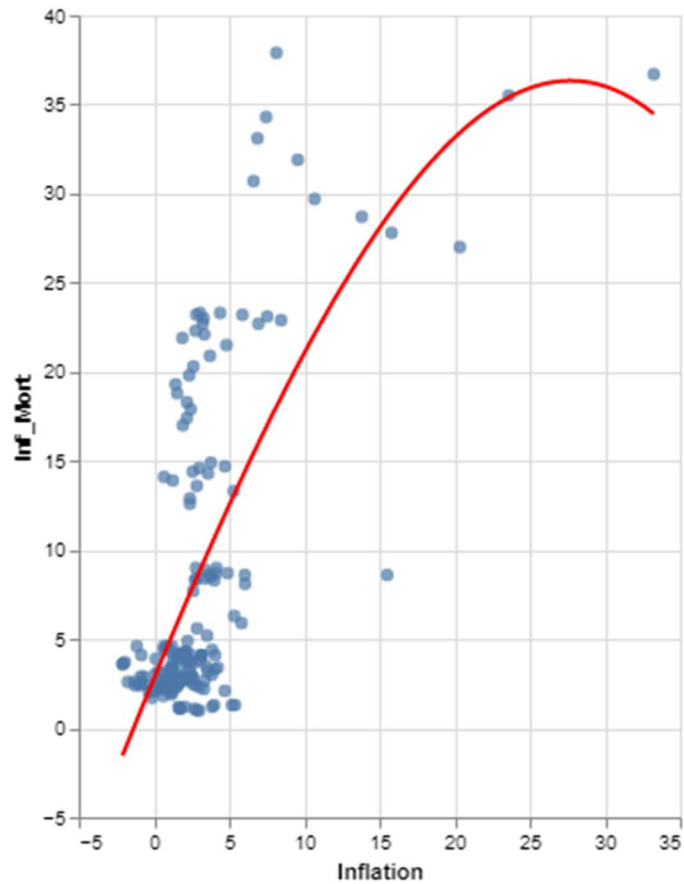
This strip plot indicates the infant mortality rates for the sample time period, and the global average. Each year with available data is depicted as a dot in the chart. The chart mostly shows a lower, below-global average rate (of <5/1000). A few countries scored higher, but still below average – Philippines, Libya, Mexico, Argentina. Botswana and Ethiopia were higher than average. This chart also shows was not a lot of rate variance between years for most countries.



Visualization 3:

Scatterplot with Regression Line of Infant Mortality Rate Compared to Inflation Rate

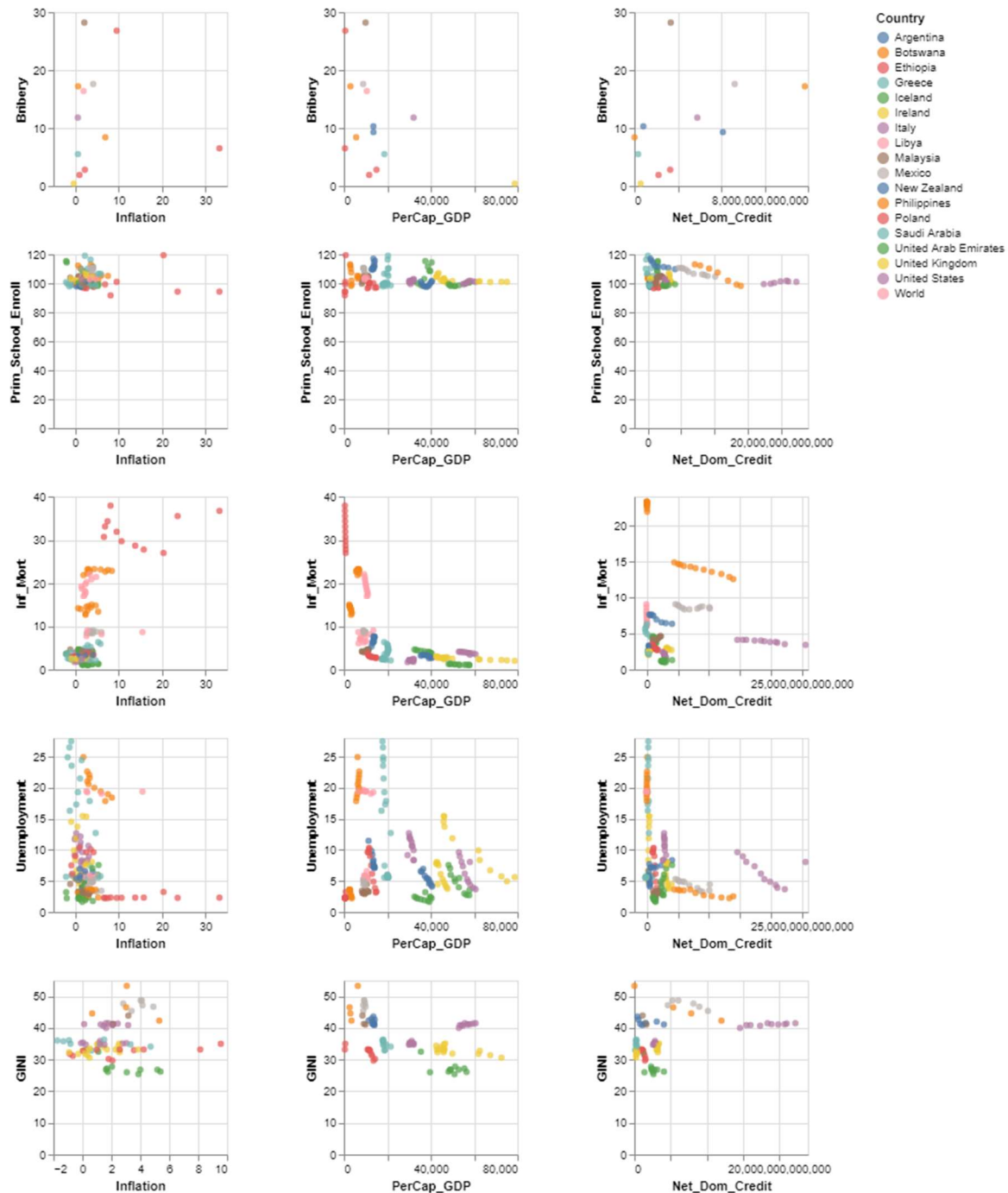
This chart forms the basis for much of the work I will be doing in my group project – comparing inflation rates to humanitarian indicators. In this case, while far less data exists for higher inflation rates during this time period, there is a clear correlation between higher infant mortality rates and higher inflation rates.



Visualization 4:

Small Multiples of Various Economic Measurements and Humanitarian Indicators

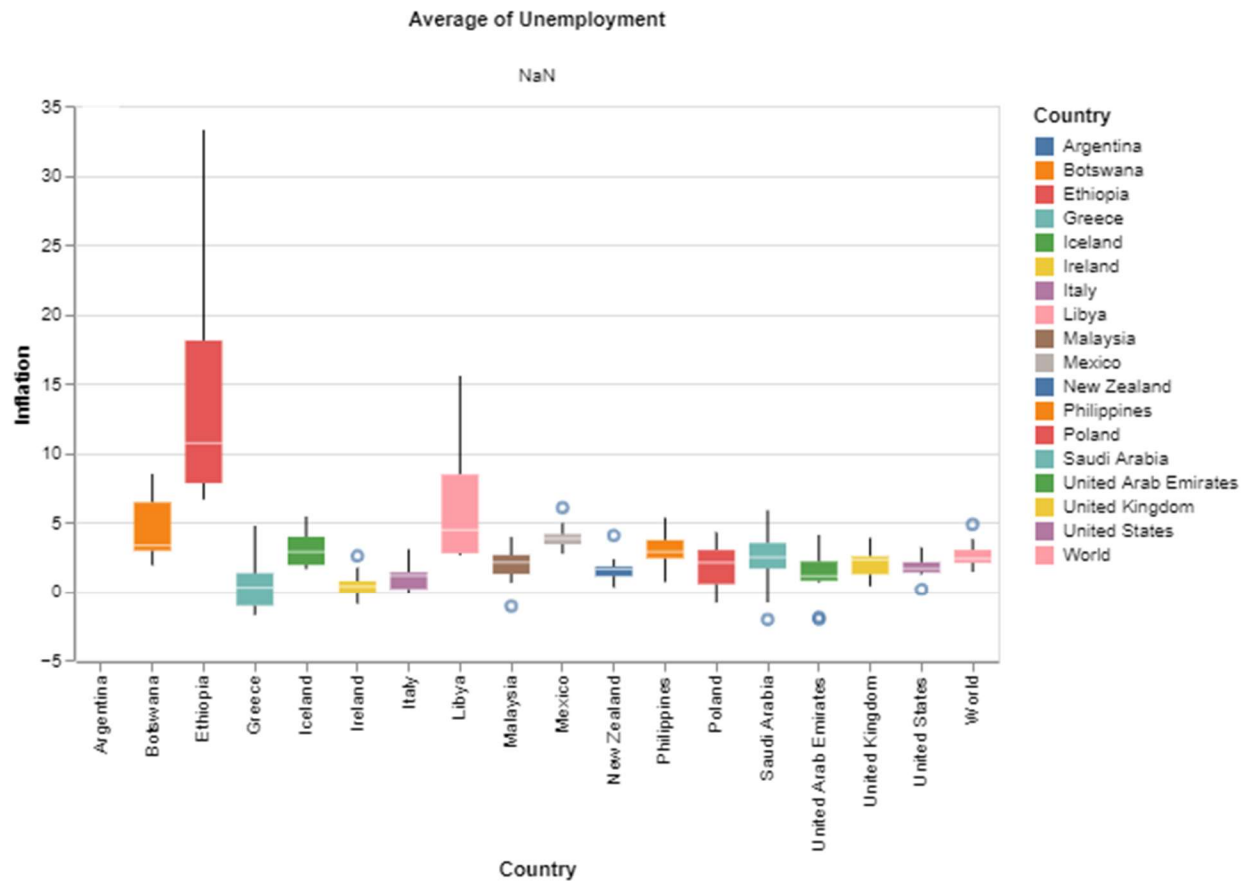
This chart series was built in order to check if any other economic measurement indices besides inflation also had correlations to humanitarian indicators. In this case per capita GDP also appeared to have correlations. The humanitarian indicators which appeared to present a significant relationship with economic indicators are infant mortality, unemployment, and possibly inequality (GINI Coefficient).



Visualization 5:

Box Plot Comparing Average Unemployment to Inflation Across Sample Countries

This chart compared each country (and also the global average) inflation rates, to the average unemployment from 2010-2020. The results show that most countries varied little in their rates (most ~2-3%) but Libya and to a greater extent Ethiopia plotted a large variance of inflation and average unemployment.



1) A one-factor analysis of variance for Infant Mortality rate by Country Analysis of variance has shown that there is a significant difference between the categorical variable economy and the variable Inf_Mort $F = 704.25$, $p = <.001$

F	df1	df2	p
20.19	17	180	<.001

2) A two-factor analysis of variance with repeated measures between primary school enrollment and inflation showed that there is significant difference between the groups of the first factor " Inflation and Prim_School_Enroll " in relation to the dependent variable, $p=aN$, significant difference between the groups of the first factor economy in relation to the dependent variable, $p=aN$, interaction between the two variables economy and " Inflation and Prim_School_Enroll " in relation to the dependent variable, $p=aN$.

3) A Pearson correlation was performed to test whether there was a association between Inflation and Unemployment. The result of the Pearson correlation showed that there was no significant association between Inflation und Unemployment, $r(178) = -0.13$, $p = .079$.

There is a low, negative correlation between the variables Inflation and Unemployment with $r = -0.13$. Thus, there is a low, negative association between Inflation and Unemployment in this sample.

	r	p (2-tailed)
Inflation and Unemployment	-0.13	.079