**Road Map**

1. Check missing values and decide imputation/interpolation strategy. Which interpolation method should be selected? These are the existing approaches: linear, log-linear, Catmull-rom spline, Cardinal spline or Cubic spline? Linear is a good start.
2. Visually inspect the data for outliers and remove them before analysis.
3. I want to create a dataset called **replica**, which consists of the **PolityIV** variables and my development indicators starting from 1990. The objective is to replicate the results of another study that clearly shows how improvements in governance cause increases in income levels.
4. I want to create another dataset called **WGI**, which consists of the World Bank governance indicators and our world development indicators dataset ranging from 1996 to 2016.
5. Then I want to create yet another dataset called **Frazer**, which consists of the Frazer institute indicators dataset ranging from 2000 to 2015. It has more features.
6. We do the same thing to all three datasets and come up with as many modelling and analysis approaches as possible. To really dig into the data and come up with actionable insights.
7. I had the idea of coming up with a visualisation that has two vertical axes, one left (log GDP) and one right (spend of some sorts by the government). A scatter plot showing the relationship between log GDP and governance measures together with bars on the right vertical axis.
8. Another really interesting idea would be an area chart decomposing GDP into agricultural, industrial and service sector part over the time range, showing which part of GDP grew more than other parts visually. Perhaps do this by region, or by type of country, is it one with oil resources? Is it one that exports natural resources? We can do it for two groups of countries, (1) resource rich, (2) resource poor. And by resources we are referring to countries that export a lot of natural resources, such as oil, minerals and diamonds/gold.
9. Maybe the economic growth of African countries has just been driven by higher natural resource prices than by betterment of institutions.
10. Another important idea I have seen in Hossein Academy pertains to using an alternative measure of economic growth. Instead of using GDP increases, one might want to decompose it into agricultural production, industrial production, natural resource booms, or service sector increases individually, or even combine them.
11. When conducting Linear Regression, maybe it would be a good idea to compare np.polyfit() with sklearn's .fit() and statsmodel OLS method, so that we can see differences in the out of the box algorithms. There is also a way we can visualize what the algorithm does.
12. Maybe it would be a good idea to transform features by taking their natural logarithm in order to make it more easily interpretable, as it is interpreted as an elasticity if both target variable and feature variables are in log form.