Alex Scarlatos
CSE 564
Dr. Klaus Mueller
4/11/17

Proposal:
Effects of Education on the Well-Being of Nations

Motivation

Improving education has always been a major goal of philanthropists and world organizations when attempting to aid poor nations. In first world countries, we believe that a good education is key to our way of life and high standard of living. But it is important to examine exactly what role education has in the well-being of a country. Certain aspects of a country could improve due to providing education, while others could be dependent on other factors. If organizations are to determine where education aid should be going, they need to know about strong correlations between aspects of a country's well-being and the education of its people. We can discover which countries should be targeted for education aid by examining global and localized trends, and comparing them to other cases across the world.

Data Source

I will be using the World Bank as the source for my data. They have a database of a huge variety of attributes for countries around the world over a range of years. Their data is formatted so that each country is a row, each year is a column, and each cell is the value of the variable in question. I plan to create a new dataset using these preexisting ones by making country and year into a key pair, and making a column for each variable I want to examine.

I can prepare several datasets using this method in order to compare correlations between different areas. The World Bank has groups of variables such as health, aid effectiveness and poverty. Each of these groups of variables could be made into its own dataset, and combinations of groups, for example education and poverty, could be useful when trying to focus in on specific correlations. These will be very high dimensional datasets (some of the groups have more than 50 attributes) so I will need to use judgement to make initial reductions, and then PCA or other methods to further reduce the dimensionality.

Analysis Plan

After finding the variables with the strongest PCA loadings in each dataset, I will combine them into a single large dataset with the most important variables from each relevant group. Then I'll compare the correlations between these variables and visualize them with MDS. Since the datasets provide years, I will watch the correlations evolve over time, and the way that a

variable's correlation with education changes will be revealing. It could indicate a false positive if the correlation moves away from education over time but is more likely to be relevant if it stays close over time.

I plan to do clustering on the countries to find out which ones have similar attributes. It will be useful to compare trends between countries within clusters. For example, maybe one group of countries has poor infrastructure, and within that group education benefits are different from countries with better infrastructure. Within these clusters I'll try to find the strongest correlated variables to education and create 2D visualizations across those variables. I will also identify attributes that set clusters apart from each other. If the clusters show different correlations with education, then I will be able to identify attributes that change the way that education affects or is affected by a country's status.

Some deviations may be hard to explain, so I will identify outlying countries and attempt to explain their peculiarities with data that is not found in the dataset. These may have to do with politics or internal conflict.

By the end of these analyses, I will have identified variables that are strongly correlated to education in countries under different conditions. I expect some of these correlations to vary across different regions of the world, and others to be more global.

Plan for Final Presentation

After discovering which attributes and regions are worth examining, I plan to present evidence of these trends along with future predictions and analysis.

The evidence will comprise a variety of visualizations, including 2D attribute charts, 2D artificial dimension charts (like MDS) and higher dimension charts (like parallel coordinates). I will use analysis techniques, like traveling salesman, to find the best way to present these results.

I will also use a world map to show the most relevant attributes and relations. It will be useful to display how certain attributes correlate with education differently across the world. For example, if health, poverty and infrastructure vary with their education correlations across different groups of countries, I will assign those attributes colors and fill in the map accordingly.

Based on these results, I will extrapolate the strongest trends to make predictions about the future conditions of education and well-being for countries that are going through clear changes. I will also use the results to determine which countries could benefit from education by comparing them to similar countries that have improved from it. This could be useful for organizations in determining which countries would benefit the most from education aid.