Laura Emberson

Data Science Capstone

April 11th, 2017

Project Proposal

Access to freshwater resources is vital for human existence. Water is essential to life, but it is also essential for quality of life. Access to freshwater can increase individuals’ productivity, reduce poverty levels, and reduce the financial burden of health-care placed on governments. In 2012, the world achieved the United Nations’ Millennium Development Goal of halving the population without access to improved sources of water. But by 2014, the percentage of people with water access in some countries, like South Africa, was slipping backwards.

The project I will be working on for my capstone uses data from the Food and Agriculture Organization of the United Nation’s AQUASTAT database to determine if trends in national water and agricultural economies can predict the future social health of a country. These predictions can be used to determine which countries will have the most need for national and international aid programs in the immediate future. With this knowledge, international aid organizations could coordinate aid efforts before the infrastructure problems become social problems; thus making the preventative programs more effective.

AQUASTAT is an open source database, maintained by the FAO. The information has been curated with the explicit goals of bolstering social progress and increasing the sustainable management of natural resources. The six main topics that data is collected under are 1) geography and population, 2) water resources, 3) water use, 4) irrigation and drainage development, 5) conservation agriculture and water harvesting, and 6) environment and health. The two main indexes of social health that I will be using as indicators of general social well-being, Human Development index and Gender Inequality index, are already part of the dataset.

The approach I will take to solving this problem will be to analyses trends in the water and agriculture data and determine if they correspond changes in the indices. For example, one large and easy comparison to make is if the amount of water resources/per capita within a country corresponds to changes in social indices. Another similar comparison to investigate is whether the proportion of in different reserves (i.e. ground water, surface water, and reservoirs) has a social effect. The data is grouped into 5 year periods, so I will try to link the indices of future year ranges to previous water trends, thus giving the model predictive power.

The products of this project, will be code that can be used with the AQUASTAT dataset to make predictions about the future social health of nations. Hopefully, this code could also be applied to other similar datasets for countries that the UN does not hold data for. Slides and comments will also be provided that make predictions of national social health for the next 5-10.