Applied Data Science – Final Project Proposal

Peter Annable, Spring 2018

The research question I would like to examine is one I submitted as part of Assignment 1. I would like to study the correlation of global climate change to local weather conditions. In particular, I happen to own a cottage on Lake Matinenda, in northern Ontario, Canada. Over the past 30 years, I've noticed a large number of seemingly unusual weather there during the winters. I'd like to understand if there truly has been a warming trend in local winters, and how they might correlate to regional and global changes. Sometimes this is known as "weather vs. climate". That is, local variation in daily or seasonal weather is incorrectly attributed to climate change. I would like to understand if my personal observation has a statistical correlation to broader climate trends.

Data Required

To examine these changes, I will need the following weather data:

- Historical Actual temperatures and averages, available at https://elliotlake.weatherstats.ca/
- Number of times winter temperatures exceeded normal by +2 standard deviations for Elliot Lake. This will be calculated from the above data.
- Global average temperatures available at: https://climate.nasa.gov/vital-signs/global-temperature/
- Average surface temperatures for the great lakes: https://coastwatch.glerl.noaa.gov/statistic/avg-sst.php
- Ice-out dates for Lake Matinenda: http://www.matinenda.ca/wp-content/uploads/2016/02/Matinenda-Ice-Out-Dates-82-Yr-Record.pdf
- Regional Northeast Ontario Data: Snow Cover Extent, http://www.statcan.gc.ca/pub/16-002-x/2012001/ct028-eng.htm

Data Exploration and Analytic Techniques

The first challenge will be to clean and harmonize data sources. For example, all temperatures need to be in Celsius, with common date formats, and common location designations. Some sources may not be available in excel format, and may require conversion.

Analysis techniques I expect to use are: Linear Correlation, Rolling Averages, Scatterplot inspection to understand if data is normally distributed, or is skewed. Anomaly Detection may also be useful to discard potentially erroneous observations. I will also investigate the use of Linear Regression models to determine if the ice-out date on our lake can be predicted based on weather trends of that year.

R will be used for technical analysis work. Tableau will be used for the final presentation.

Expected Charts and Graphics

Line Charts, Bar Charts, and Scatterplots will be the primary graph types, since I will be analyzing time-series data. Bubble charts may also be useful to understand the relationship between different data types such as ice-off dates, snow cover extent and average temperatures. I expect the final output to be a dashboard published using Tableau Online that will help to visualize the story of the data, trends found, and any conclusions that can be reached.