Capstone project 1 proposal

Ross Brown

Springboard Data Science Career Track

January 31, 2018

For my first capstone project, I will investigate whether winter severity in the United States correlates with alcoholism. The hypothesis is that severe winters cans cause people to stay inside more, become isolated, become depressed, all of which may cause people to drink alcohol and potentially develop an alcohol dependency.

Data sets:

To quantify the severity of winter, I’ll use the Accumulated Winter Season Severity Index, which was developed to “objectively quantify and describe the relative severity of the winter season.” The AWSSI is calculated for U.S. cities, and is based on “max/min temperature, snowfall, and snow depth or precipitation.”

To quantify the number of people experiencing severe alcohol dependency, I will use the “Treatment Episode Data Set: Admissions” from the federal government. Dating from 1992, this data set “provides annual data on the number and characteristics of persons admitted to public and private substance abuse treatment programs that receive public funding.” Alcohol dependency is one of the patient characteristics in this data. Given that data collection is tied to programs that receive public funding, I presume the data is representative, as well as reliable and comparable across locations and time. The TEDS-A exists as separate files for separate years, and besides cities, individual cases are also identified by state and by nine regions.

There are several options for analysis and comparison, including across cities, states, or regions; and across years, or greater aggregate time periods.

To best reflect the effect of winter severity on the preponderance of alcoholism, and to minimize the effect of variance due to other causes, I propose to compare two consecutive years reflecting the greatest difference in winter severity, i.e., for example, one year a location experiences a mild winter, followed by a severe winter. I propose to compare locations with larger populations, either large cities or states that have similar winters across the entire geography (i.e., not looking at, for example, California, which due to its size, may have very different winter severities across the state). Using successive years will reduce longer-term effects on the data, i.e., changes in diagnostic criteria for alcoholism that were gradually implemented, or maybe changes in recognition of alcoholism as a treatable problem.

The TEDS-A is counted by calendar years. The AWSSI describes winter seasons that cross the calendar year. However, I believe admissions for alcoholism is a lagging indicator of winter severity, i.e., a person is not likely to develop alcoholism and seek treatment as an immediate effect of a severe winter, but rather would do so after some passage of time. Therefore I will compare the year that a winter ends, as described by the AWSSI, with that year’s admissions for treatment as reflected in TEDS-A.

Broadly, the components of this project will include

· Data wrangling

· Data cleaning

· Combining data

· Accounting for missing data

· Data analysis

· Data Visualization

February 7, 218

Analysis steps:

Review winter severity data for index data that meets the following criteria:

* Two consecutive or near consecutive winters of disparate severity,
* That has a minimum missing data,
* After 2005