**Canadian Climate Change**

Dave Sullivan (500955153)

2020-04-05

https://github.com/davexsullivan/climate

**Introduction**

Climate science is a vastly complex field of study with massive amounts of existing historical data as well as an ever-increasing amount of daily data being generated. Despite a consensus on the trends towards more extreme temperatures and weather patterns, there are still skeptics and naysayers who declare doubt that mankind has any sort of accountability in the cause of these potential outcomes. 97% or more of published climatologists and related experts agree that climate warming trends are caused by anthropogenic means.

The research question for this project is to determine if it’s possible to arrive at the same conclusions with readily available open-source data and statistical tools. The sources of data I will use will involve Canadian weather stations aggregation data (daily and monthly) and global rates of both atmospheric carbon dioxide and solar irradiance.

**Literature Review**

The publications from NASA were, by far, the largest source of information on the science and mechanics behind weather data collection. The correlation between earths temperature and solar radiation is naturally linked, but NASA has shown that annual fluctuations and the gradual increases we are experiencing are more strongly linked to the concentration of greenhouse gases such as CO2 in our atmosphere.

Sources:

Hansen, J., Ruedy, R., Glascoe, J., & Sato, M. (1999). GISS analysis of surface temperature change. *Journal of Geophysical Research: Atmospheres, 104*(D24), 30997–31022. doi: 10.1029/1999jd900835

Hansen, J., Sato, M., & Ruedy, R. (1997). Radiative forcing and climate response. *Journal of Geophysical Research: Atmospheres, 102*(D6), 6831–6864. doi: 10.1029/96jd03436

Lacis, A. A., Schmidt, G. A., Rind, D., & Ruedy, R. A. (2010). Atmospheric CO2: Principal Control Knob Governing Earths Temperature. *Science*, *330*(6002), 356–359. doi: 10.1126/science.1190653

Online sources:

<https://aeronet.gsfc.nasa.gov/>

<https://solrad-net.gsfc.nasa.gov>

<https://climatedataguide.ucar.edu>

<https://www.lumigrow.com/learning-center/blogs>

**Dataset**

The main source of climate data used for this project can be found at http://climate.weather.gc.ca/. This government portal only allows the download of a single stations data, one month at a time. To circumvent this limitation, the R package WeatherCan was utilized. (https://cran.r-project.org/src/contrib/Archive/weathercan). It contains several useful functions for searching and downloading much larger ranges of the data.

Atmospheric CO2 annual levels data was downloaded from: <https://www.esrl.noaa.gov/gmd/ccgg/trends/data.html>

Satellite solar radiation readings were collected for 1984 to 2003

As well as monthly solar radiance levels for the Toronto area from:

<https://aeronet.gsfc.nasa.gov/cgi-bin/webtool_aod_v3>

The CO2 and solar data sets were available in flat text format only and were converted into CSV for use.

**Approach**

The approach taken is shown in the graph below with descriptions following after.

5.Exploratory Analysis and Correlations

1. Download bulk data in R via WeatherCan and supplemental data from other sources

4.Export aggregated data into R for analysis

3.Clean and aggregate data with SQL

2.Export data as CSVs from R for insertion into SQL database

Step 1. Download data via R with WeatherCan

The weathercan package is quite robust, allowing the full historical download of data from multiple stations. A search feature is built in allowing the user to filter stations by the recording frequency, by province, by the date it starts and ends, by latitude & longitude, or even radial distance from a specified set of coordinates. I chose to focus on several different data sets. I downloaded the daily data for all stations with consistent recordings from at 1980 to 2018 as well as a set for all stations within 150km of Toronto, Ontario for the period of at least 1984 -2018 to match the date range of the local solar radiation data. At first the larger data sets were encountering memory limitations and triggering errors, prompting an upgrade to the latest 64-bit version of R. This caused all downloads to fail, but success was achieved after reverting the version of R installed on the machine used (from 3.6.3 to 3.5.3 64-bit).

The global monthly CO2 levels were downloaded from <https://www.esrl.noaa.gov/gmd/ccgg/trends/> and reformatted for use. Ground solar radiation data was downloaded for several

Step 2-3. Export from R as CSV and import into SQL as tables and clean and aggregate with SQL

I used SQL to import all the datasets, replaced NA values with NULLs and manually inspected the data for errors or anomalies. The satellite solar radiation data had several recordings of -99.99 which were removed from the final set. I created new set of aggregates from the weathercan daily data into monthly and annual averages with min and max temps, and avg summer temperature data, joined to the corresponding CO2 levels and appropriate solar data.

Step 4-5. Import aggregated data into R for analysis and final results

I brought the data back into R and performed correlation and regression analysis on the temperature, solar, and CO2 variables.