Dylan DeWolfe, Andrew Bauer

Professor German

CPSC 408

23 May 2020

Coronavirus vs Temperature Database

Synopsis:

Given the current worldwide pandemic on our hands, it is the responsibility of those who can help to help in whatever way they can. Given our software engineering experience, and our data analytics experience, we decided the best way to help was by creating a database to analyze correlations of data about coronavirus. There is the notion that Covid-19 spreads easier in colder climates and slows in warmer climates. Our project is designed to pull data from online repositories and populate our own database in order to create comparisons. Once populated we performed various operations to visualize the data, to analyze the correlation between average temperature by state and the number of confirmed cases. We found that the correlation between temperature and number of cases does in fact exist. We were therefore able to prove our hypothesis that Coronavirus does not do as well in hot climates.

Introduction:

We will be using the terms COVID-19 and coronavirus, and although they are respectively different, they are similar enough where we will be using the terms interchangeably. While this disease is affecting nations worldwide, we chose to analyze the data in the US as we

knew the US has been up to date with testing, tracking the weather accurately, and wanted to use a smaller scope for our project.

Problem:

If you didn't already know, there's currently a worldwide pandemic happening dealing with the disease known as COVID-19 (CO - corona | VI - virus | D - disease | 19 - 2019) caused by a new coronavirus. Due to this new disease, stay-at-home orders have been put in place with 11 states in full lock down and 26 in partial lock down. Currently, the disease has taken the lives of 97,000 and infected 1,600,000. Due to the fast growth and 6% mortality rate of this virus, fears have consumed the nation causing the stock market (Dow Jones) to drop 36% in a matter of a week and unemployment to rise to 14.7%. These fears of a fast growing disease have also perpetuated a lot of fake news in the media and thus quick believers due to this chaos.

Goals:

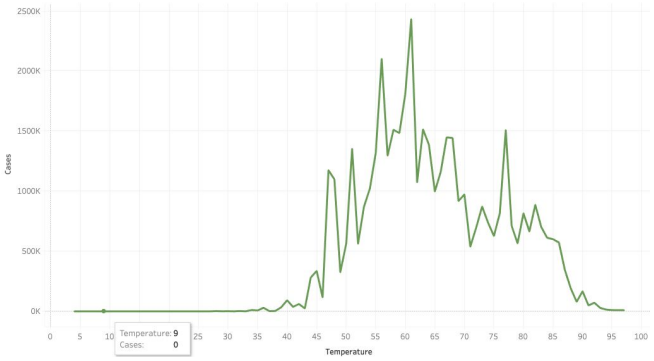
Our main goal is to prove whether or not high temperatures slow the spread of coronavirus. Our secondary goals include the prevention of fake news on this topic, aiding researchers through our developments, and slowing down the spread of coronavirus through new information. These secondary goals will all be completed by completing our primary goal.

Methodology:

We began our project by sourcing data from multiple online repositories. These repositories were in various formats. We used Python's Pandas library to clean the data as well as

import the data into our MySql database. Our repositories included Google Cloud's Big Query NSOD weather database (provided average weather from each state on a day to day basis for 2020) and John Hopkins publicly available coronavirus Github (provided all information on number of cases). From there we went about constructing tables in a normalized form (3NF) to organize our data as a relational database. We used the state and date as foreign to sort and query our data into different tables. We honestly would have normally kept these all together, but per project specifications, we needed three different tables to join. We have our main table containing all the useful information mapped into one spot, a states key table, and a dates key table. We created indexes on columns that have semi-unique values such as states, dates, and temperature. Temperature index is a little debatable but we decided to go ahead with it as the large dataset we have contains many correlating numbers of temperature. Our database contains a collection of 5900 rows and a total of 29,500 data points.

We utilized Google Cloud for our cloud database and imported the data through Python. We inevitably used Python and the Tkinter toolkit to build our interface. From here we added all the specifications including insertion, updating, soft deletion (done through a boolean), searching and much more. We utilized the application Tableau to visualize our data in a graphical interface (as seen below) in order to easily understand whether or not there was a correlation.



Errors:

There was a lot that could have messed up our analysis and data. Our data could have been inaccurate due to people being unreported who are asymptomatic, weather could be affecting the behaviors and interactions of people, as well as the day that states initialized quarantine could affect our data. We had quite a bit of personal errors throughout our process. Some included our inability to get live data working (making us having to throw away about half our data and restart with a historical repository), we had laptops die while working on this (thus also losing us valuable work), lots of problems building a GUI as neither of us had experience building one, and lastly we very recently figured out that our data was massive containing 350,000 ROWS of data with 9 columns (I don't even wanna do the math to figure out how much data we were wasting).

Conclusions:

Luckily for us the world seems to be healing and the spread of the disease seems to be decreasing thanks to social distancing and the work of our wonderful medical staff. Based on our results, our hypothesis that coronavirus spread would weaken as temperature increases was demonstrated. Although we can't necessarily be certain that correlation equals causation, we can hopefully distribute this data to scientists who will hopefully use our findings to conduct tests on either killing or slowing the rate of spread. This project was definitely a long and hard one, one that we're happy with our conclusions, but frustrated with the amount of work we took to get there due to all our numerous mistakes. All in all, make sure to stay in, stay safe, and wait for your steak and lobster dinner once quarantine is over.

Sources:

https://www.cdc.gov/coronavirus/2019-ncov/faq.html

https://www.wsj.com/articles/a-state-by-state-guide-to-coronavirus-lockdowns-11584749351

https://news.google.com/covid19/map?hl=en-US&gl=US&ceid=US%3Aen&mid=%2Fm%2F09c7w0

https://www.bbc.com/news/business-51706225

https://console.cloud.google.com/marketplace/details/noaa-public/gsod

https://github.com/CSSEGISandData/COVID-19