Covid-19 Hackathon

Does weather influence COVID-19 infection rate?

Team: Covid Ninjas
Members: Ankita Pal, Kush Tekriwal, Mansi Rathod, Vikas Kini

Outline

- Background
- Problem Statement
- Data
- Project Reproducibility
- Methodology & Metrics
- Results
- Drawbacks & Limitation
- Possible Explanation

Background

The Covid-19 pandemic has clutched the entire world for as long as 7 months. This has caused unprecedented damages socially and economically all over the world.

With the entire force of world leaders and health workers working towards it, there still exist a lot of untested domains. In our efforts to combat this fight, we have selected one particular area and worked to generate insights.

This is the first pandemic in which we have access to real time data, and possibilities for investigating hypotheses is vast

Problem Statement

We aim to understand how weather changes (more specifically air temperature) influence the rate of infection (R0). We know other coronaviruses like SARS have better stability at lower temperatures (source), but higher temperatures can increase social contact.

For our hypothesis testing, we have considered only King County-level data. The reason for this is that our hypothesis requires localized county level case data, which is lacking.

The case numbers are heavily dependent on government interventions, so we narrowed down out study time frame to only the lockdown period.

Data

We have scrapped APIs for accessing the weather data at city level. Below is the overview of our datasets:

Source	Useful Features	Link	Dimension/ Time Frame
Seattle Weather Data	Date,"24 Hrs., Min Max temperature”	https://www.ncdc.noaa.gov/cdo-web/datasets/GHCND/stations/GHCND:USW00024233/detail	January 1st - May 31st
King County COVID-19 Dashboard	New Cases, Total Tests, Positive Percentage	https://kingcounty.gov/depts/health/covid-19/data.aspx	April 14th - May 31st

Project Reproducibility

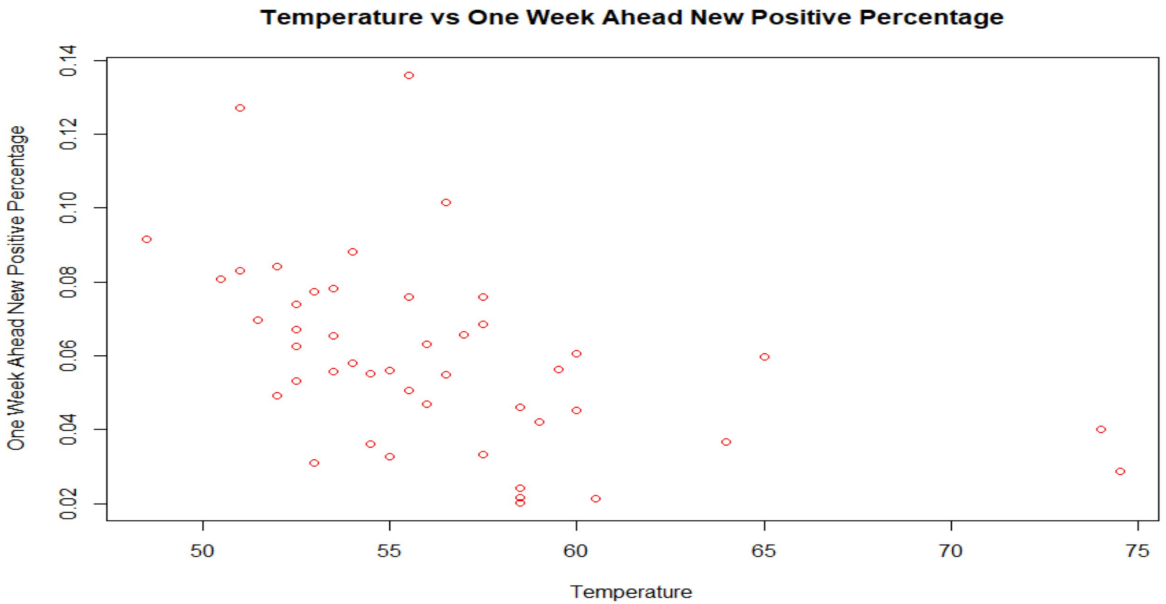
- This hypothesis should be tested for multiple counties all around the world. We would expect counties with similar geography and weather to have similar trend lines in positive percentages.
- Reproducing this test in other counties where there is relaible data will let us be more confident on which one is closest to the true effects of temperature on transmission

Methodology & Metrics

The ideal metric to gauge the growth of the epidemic in a given county would be Rt or the adjusted reproduction rate for the virus. However, access to county level reproduction rate is not accessible as of now, and even the state Rt levels are crude estimates that are not robust enough to account for highly variable testing capacities and targets.

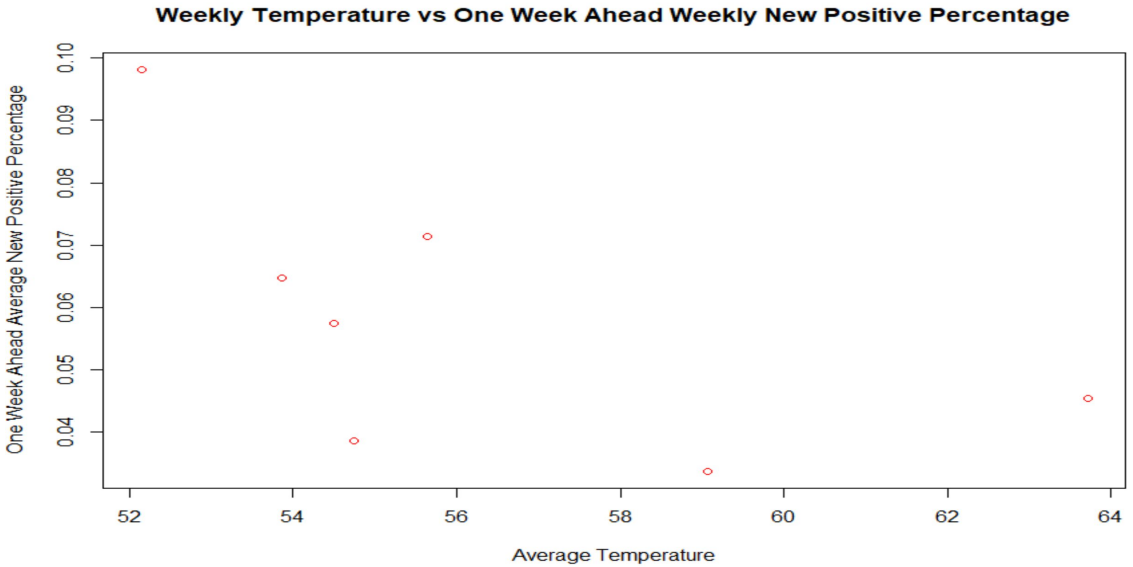
So, the next best metric would be **New Percentage Positive (NPP)** which represents the fraction of positive antigen test results from all the tests conducted on a specific date. To further insulate from noise caused by testing capacities and temperature anomalies, we used the mean daytime temperature for a given week and compared it to the mean NPP over the following week. This assumes that a newly contracted virus takes one week to show up on the King county case tally.

Results:



Pearson Coefficient = -0.4602

Results:



Pearson Coefficient = -0.6175

Drawbacks & Limitations

- Since, the range of temperatures is very small, we cannot account for year-round impacts of temperature on the spread and create a model for it
- We were able to get data only for 7 weeks. This leads to limited range of temperatures and effects of spread
- We are not taking into account any external factors like closure of schools and businesses, effects of mask-wearing and possible mutations of the virus.

Conclusion & Possible Explanations

We saw a weak correlation that showed a decrease in the NPP following a week of higher temperatures. This initially seems to confirm peer reviewed studies of the SARS outbreak. However, there is still no clear trend, which could be in part because of:

- Weather does not play a significant role in the change of positive percentage
- Weather does have a correlation but the other confounding variables play a larger role and and the weather variability is lost
- The data being generated now is not granular enough to assess the true temperature effects
- The change in temperature creates both positive and negative feedback in NPP that negate each other