How much did work from home help? - Research Proposal

Satheesh Joseph, Paco Valdez, Yi Zhang

Background

The Covid-19 pandemic has taken more than 2.5 million lives worldwide. In the U.S. alone, more than half a million people died because of it.

COVID-19 is thought to spread mainly through close contact from person to person, including between people who are physically near each other (within about 6 feet).[1]

As the virus spread, many cities and other areas have implemented a shutdown policy that significantly restricted the mobility of the population. Specifically, many company have opted for a complete work-from-home policy.

As a team of Data Scientists, we're interested in assessing the effectiveness of restricting mobility in controlling the spread of Covid-19.

Research Question

Rather than investigating the effect of social distancing on a micro, personal interaction level, we are interested in its effect on a policy level.

Specifically the Research Question we're asking is:

`How much does stop travelling to work help with the spread of Covid-19 cases?`

Operationalization

To unpack the high level question, there are a few components to this.

Firstly, we'll use the New York Times Covid-19 data for the number of Covid cases per State. And we will use the COVID-19 Community Mobility Report for mobility score, specifically the workplace mobility score.

Secondly, because mobility may have drastically different effect on the spread of the virus depending on the population density, we think having that variable alone as the causal factor can be very misleading. Therefore we think population density should be mandatory for all models. The data will come from the COVID-19 US State Policy Database.

And Finally, we noticed that the Covid cases metric is a time-series data. Given that Covid symptoms typically start between 2 to 14 days after exposure[1], we would build in a time delay between the time of the mobility score and the Covid cases..

We would break the time from March 2020 to March 2021 into 52 weeks, and stagger the mobility score and the number of cases by 1 week. Therefore the simplest linear model would be:

$$total\_cases\_in\_week_{i+1} = \beta_0 + \beta_1 average\_mobility\_in\_week_i + \beta_2 population\_density$$

For example, each data point would be:

```
x1 = average mobility score of a state between 2020-03-01 and 2020-03-06
x2 = population density of the same state
y = total number of cases between 2020-03-07 and 2020-03-13 of the same state
```

For more advanced models, we will use more variables as control, for example the "Stay at Home" order, or the face mask policy, etc.

References:

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