CHRP Competition
April 12, 2020
Sean Eli
Rice University

Introduction: idea/goal

We wish to analyze the relationship between COVID-19 rate of spread (i.e. the rate of growth of number of cases, and possibly of death rate) and local min/max/average temperature.

Quantifying the growth rate

Disease spread tends to roughly follow an exponential model (every day, the # cases is multiplied by some factor). However, it is never the case that we have a constant rate factor: the growth rate changes with time. Thus, a realistic functional approximation $N(t)$ to the number of cases per day is of the form

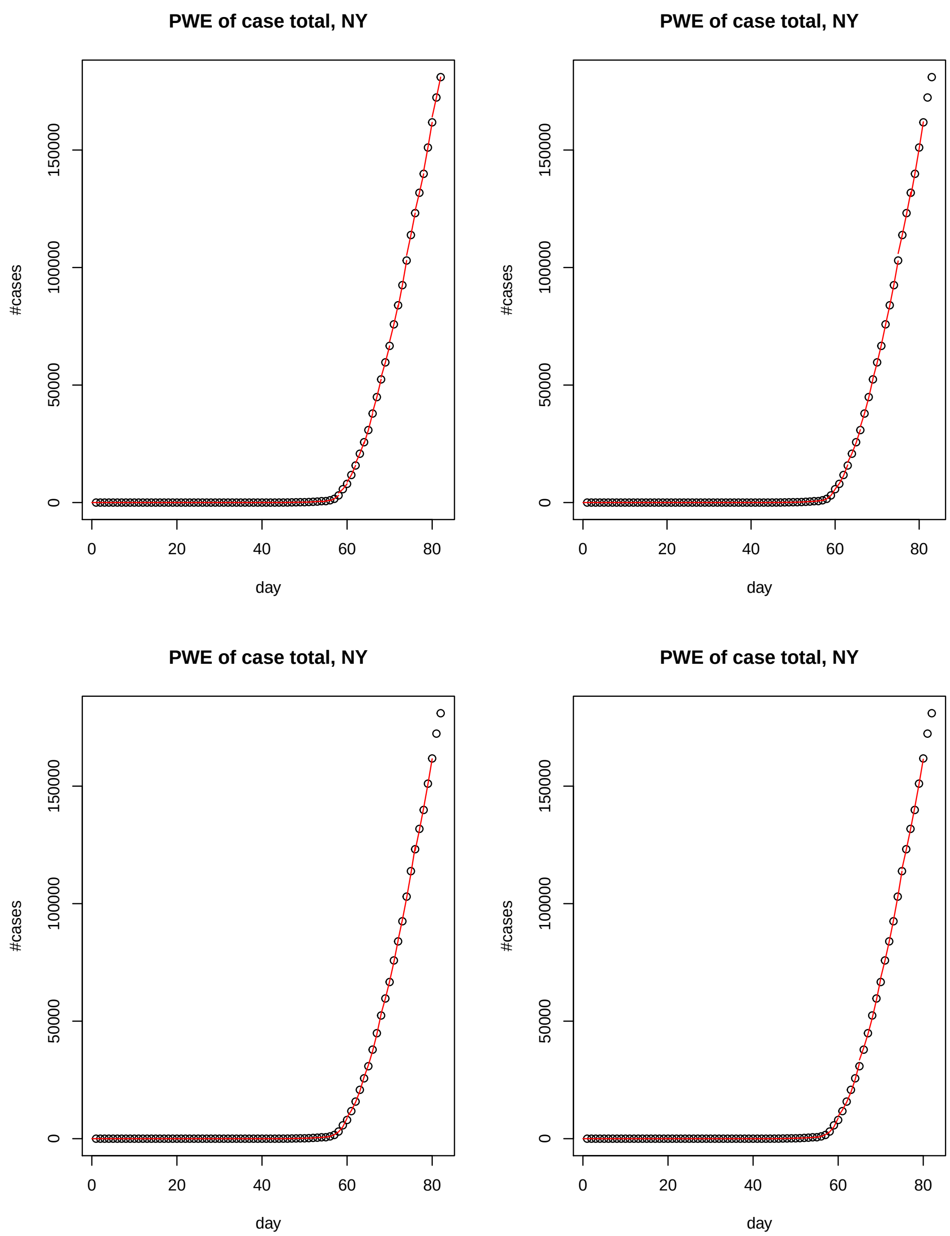
$$N(t) \propto e^{t \cdot c(t) + b(t)},$$

where $c(t)$ and $b(t)$ are time-dependent growth parameters. **Here are three ways we can quantify the COVID-19 rate of spread in a given community:**

1. **(First Difference)** In a given region, let $N(t)$ be the total number of cases at day t. Compute directly $(N(t+k) - N(t))/k$ for each t, and various k (usually $k = 1$). This is an approximation to the first derivative of $N(t)$.

2. (**Multiplicative Factors**) Given case totals $N(t)$, compute $R(t) = N(t+1)/N(t)$, starting at the first t for which $N(t) > 0$. This is the daily multiplicative factor.

3. (**Piecewise Exponential (PWE)**) Given case totals $N(t)$ for some region, partition time (days) into groups of k, and fit each group of k points with an exponential curve. This gives a PWE approximation to $N(t)$, which can be used to obtain rate information. Below are (PWE) approximations to the NY state case total (04/12/20), for $k = 2, 3, 4, 5$.



Correlation analysis of rate and temperature (in progress)

Now that we have ways to quantify COVID-19 spread rates within populations, let's look at the relationship between these rates and temperature. Our data consists of pairs $(R(t), T(t))$ for (various different) rate measurements $R(t)$ and temperature measurements $T(t)$. We face a difficulty in this analysis since COVID-19 has an incubation period between 0 days and 2 weeks. Moreover, persons showing symptoms may not be tested right away. We address this by offsetting temperature, i.e. using data pairs such as

$$(R(t), T(T-i)), \qquad \text{for } i = 1, ..., 14.$$

Our analysis takes the form of **a series of Spearman correlation tests**, where we run one test for each set of paired data $(R(t), T(t-i))$, for:

1. each choice of $R(t)$ (multiplicative, linear difference, PWE rates)
2. each temperature time offset $i = 1, ..., 14$ (and possibly higher than 2 weeks)
3. each choice of temperature $T(t)$ (daily max, daily min, daily average, k-day max,min, etc.)

To address the effects of other variables (social distancing, humidity, etc) we would ideally divide the U.S. into small communities (i.e. counties or cities, where the temperature can be assumed constant) and compute the rate information within each community. The pooled data (pairs from all communities) will contain the most information about the rate/temperature relationship, and possibly lower the effects from other variables (other variables may not be constant over the whole U.S.)

Preliminary Results

For now, we group together pairs of temperature and rate data for the following regions. We determined, as of **April 12, 2020, the counties/cities in the US with over 3000 total positive cases.** They are as follows (alphabetically by state)

1. Los Angeles, CA: 8453
2. Fairfield, CT: 5407
3. Miami-Dade, FL: 6487
4. Cook county, IL: 13417
5. Jefferson, LA: 4877
6. Orleans, LA: 5535
7. Middlesex, MA: 4872
8. Suffolk, MA: 4926
9. Macomb, MI: 3164
10. Oakland, MI: 4802
11. Wayne, MI: 10951
12. Bergen, NJ: 9362
13. Essex, NJ: 7007
14. Hudson, NJ: 6851
15. Middlesex, NJ: 5406
16. Monmouth, NJ: 3651
17. Ocean, NJ: 3403
18. Passaic, NJ: 5295
19. Union, NJ: 5865
20. Nassau, NY: 22584
21. New York, NY: 98308
22. Orange, NY: 4847
23. Rockland, NY: 7477
24. Suffolk, NY: 20321
25. Westchester, NY: 18729
26. Philadelphia, PA: 6022
27. HARRIS, TX: 3561

28. King, WA: 4262

We are currently obtaining temperature data for individual counties/cities from the interactive map by the NOAA:

`https://w2.weather.gov/climate/`

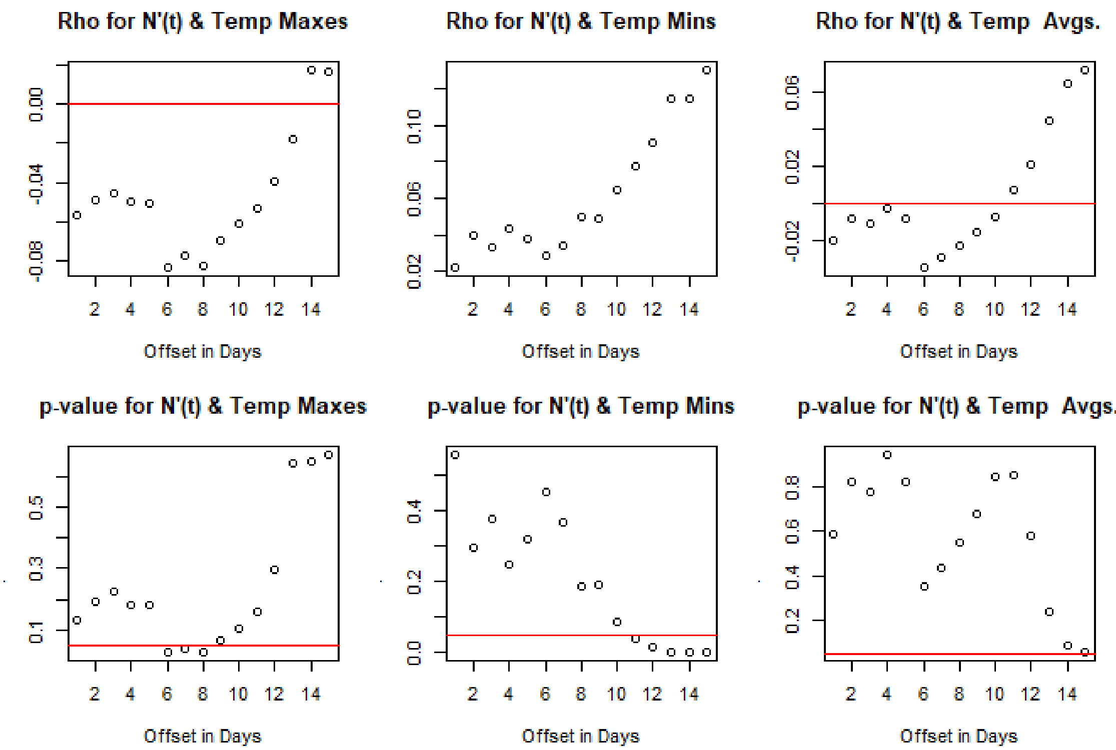
This is time series data, recording either the daily high, daily low, daily deviation from average, or daily average temperature, in the desired county, every day from January 1 to April 12 (as of right now.)

(First series of tests: April 14, 2020) We group together data from following 24 regions: Los Angeles, New York City, Bridgeport (CT), Miami, Chicago, Jefferson (New Orleans, LA), Suffolk (Boston, MA), Wayne (Detroit, MI), Middlesex (New Brunswick, NJ) , Philadelphia (PA), Harris (lovely Houston, TX), and King (Seattle, WA). We also include Phoenix, AZ, Dallas, Montgomery county, TX, Washington DC, Kansas city, Denver, Fulton county (Atlanta) GA, Las Vegas, Albuquerque, San Diego, Erie, and Boise, ID. Here are these locations on a map of the US:

Using the linear difference rule for the local COVID-19 spread rate, we run 45 Spearman correlation tests, for each of the following combinations:

1. Max, Min, and Average daily temperature
2. Temperature offset from 0,...,14 days.

Here is a summary of the correlations ρ and p-values:



The red line on the p-value graphs is $y = 0.05$. **From this, we see a significant negative correlation between $N'(t)$ and Maximum temperature, offset one week: we also see a significant positive correlation between $N'(t)$ and Minimum temperature, offset $\approx$ 2 weeks.**

Any suggestions are welcome: email me at seaneli@rice.edu or message me on slack!