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 temperature.** Disease growth tends to follow a roughly 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

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

where $c(t)$ is the time-dependent growth rate.

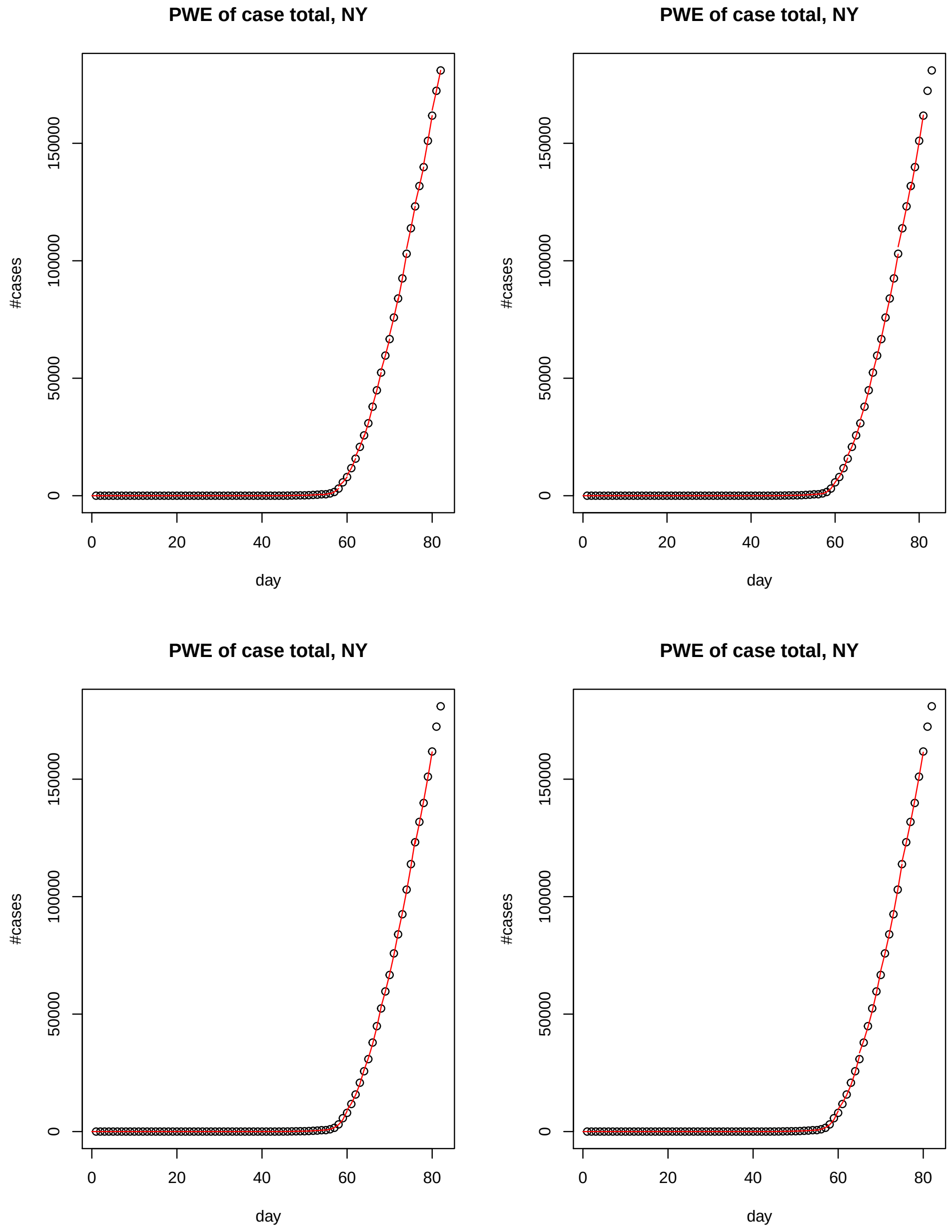
## Quantifying the growth rate

**We quantify/approximate the growth rate $c(t)$ two ways:**

1. (**Direct/Naive**) Given case totals $T_t$ for some region (say, L.A.), where $t$ is the date, compute $R_t = T_{t+1}/T_t$, starting at the first $t$ for which $T_t > 0$. This is the daily multiplicative factor.

2. (**Piecewise Exponential (PWE)**) Given case totals $T_t$ for some region, to account for randomness of the exponential rate parameter, we partition the time series data into groups of $k$, ($k \geq 1$) and fit each group of $k$ points with an exponential curve (through a log transform). This is a technical process which we describe in detail later and it captures local rate information well. Our code outputs a sequence of exponential rates for different times, which can then be used to plot a piecewise curve:

## Piecewise Exponential (PWE) approximations To NY case total (04/12/20), k = 2,3,4,5



## Correlation analysis of rate and temperature (in progress)

Now that we have a couple of ways to quantify case growth rates, let's look at the relationship between growth rate and temperature. **We face a major difficulty in this analysis since it appears COVID-19 has an incubation period between 0 days and 2 weeks**, and also since we can not be sure that the case total of one day reflects all of the people who started showing symptoms on that day (e.g. they could have waited a while, thinking it was a cold).

Thus, **to obtain the most possible information from the data we have**, we pair rate observations (either naive/direct rates, or the PWE rates) with temperature observations (daily temperature for direct rates, and averaged/maxed temperature for PWE rates), **with offset**, for offsets between 0 days up to 3 weeks. We test correlation between rate and temperature with the **Spearman correlation test** on the resulting paired data, where we group together several pairs.

## Difficulties

How to group together pairs from different regions? We can try to have a variety of locations with a variety of climates. It seems the only way to do this analysis without skewing the correlation result would be to form pairs for every county/region in the US on which we have data.

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

These large case totals should give more data on the way the growth rate changes over time.

To perform analysis on correlation with temperature, we need temperature data. We are currently obtaining data on the above counties from the **fantastic interactive map** by the NOAA:

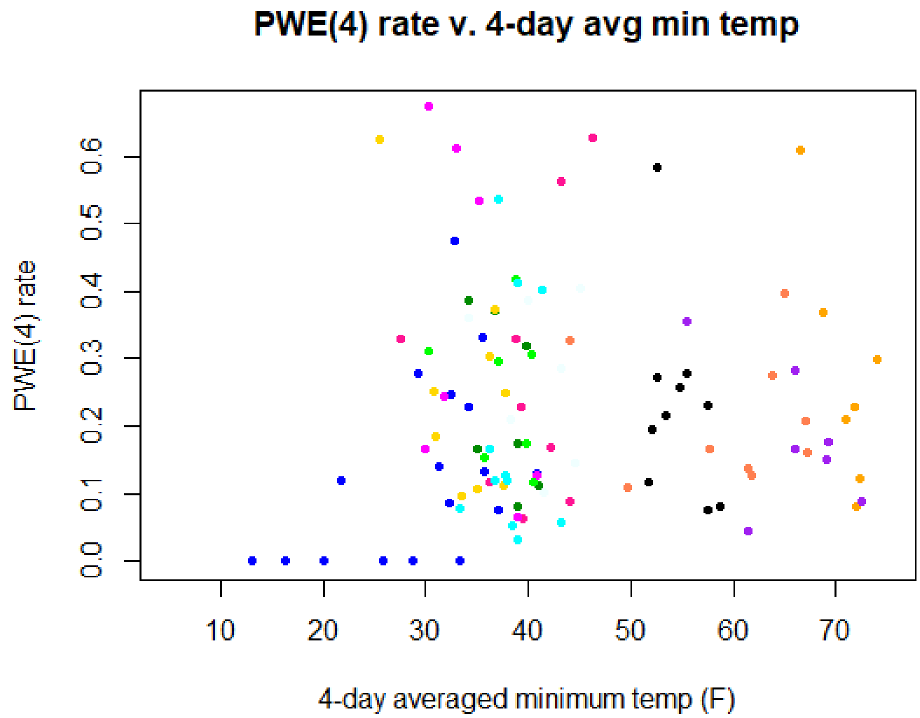
`https://w2.weather.gov/climate/`

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.)

## Preliminary Results

(April 12, 2020) We group together data from one state each out of the above list: for the following, we used the following 12 regions: Los Angeles (CA), New York City (NY), Fairfield (Bridgeport, CT), Miami (FL), Cook (Chicago, IL), Orleans (New Orleans, LA), Suffolk (Boston, MA), Wayne (Detroit, MI), Middlesex (New Brunswick, NJ) , Philadelphia (PA), Harris (lovely Houston, TX), and King (Seattly, WA).

We compute the PWE(4) rate (4- day piecewise fitted exponential rate) for each of the above regions, with the 4-day averaged minimum temperature, offset by one week (count data is from 1-22 to 4-11, so we take the daily minimum temperature from 1-15 to 4-4, and find its sequence of 4 day averages. We do this so the data matches our PWE rates. We can also incorporate temperature by using the 4-day overall minimum, max, etc.) We only consider times after the first nonzero PWE rate has happened (once the virus has started to spread.)




From this data, it is hard to see an apparent relationship between rate and temperature.

To test whether there is a significant correlation between these PWE(4) rates and temperature, we perform Spearman's rank correlation test. We obtain a sample estimate of the Spearman rank correlation

$$\rho = .07868, \qquad \text{p-value} = 0.4183.$$

**This seems to suggest these data are not correlated, or very mildly correlated.**

**We will continue** by using different $k$ for PWE($k$) and also looking at the naive/direct rate, and testing correlation. We will add more locations, not only in the US.

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