CHRP Competition
April 12, 2020
Sean Eli
Rice University

Introduction: idea/goal

We analyze the relationship between COVID-19 rate of spread (i.e. the rate of growth of number of cases) and local temperature, in the United States. We briefly summarize our methods for estimating the rate of spread, and describe temperature considerations. We then describe a multiple testing setup/evaluation for correlation between growth rates and temperature. Our result (**as of April 16**), based on 27 U.S. cities, suggests these variables may be correlated, in a *mild*, but interesting way. As a next step, we test a predictive model for growth rate at higher temperatures (time permitting).

Quantifying the growth rate

Disease spread tends to roughly follow an exponential model (every day, the # cases is approximately 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

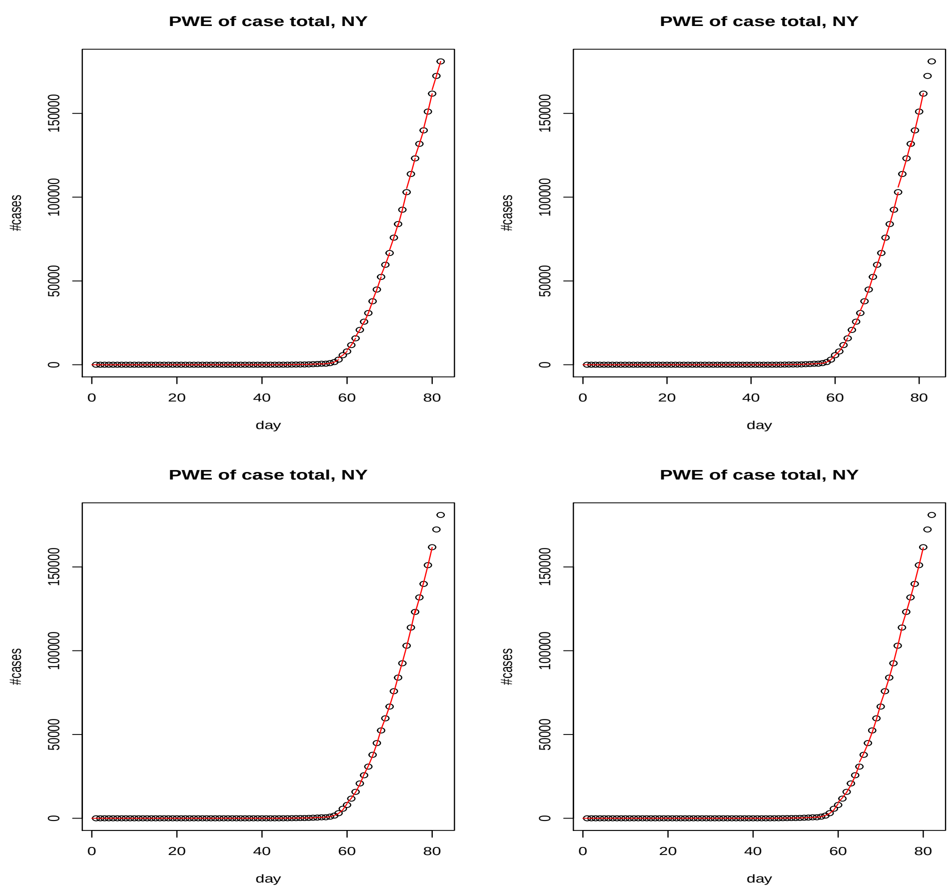
$$N(t) \propto e^{t \cdot c(t) + b(t)},$$

where $c(t)$ and $b(t)$ are time-dependent growth parameters. Keeping this in mind, here are the three ways in which we 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 $MF(t) = N(t+1)/N(t)$, starting at the first t for which $N(t) > 0$. This is the approximate daily multiplicative factor.

3. (**Piecewise Exponential (PWE)**) Given case totals $N(t)$, 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 testing for growth v. temperature

Now that we can 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 (the various) rate measurements $R(t)$ and (various) 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 correlation tests**, where we run one test for each set of paired data $(R(t), T(t-i))$, for:

1. each choice of $R(t)$ (linear difference, mult. factors, 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 maxes,mins, etc.)
4. **each choice of correlation test** (R provides Spearman's ρ, Kendall's τ, and Pearson's test. The first two are appropriate when data is not necessarily $N(0,1)$.)

Given a choice of $R(t)$, $T(t)$, and correlation test, we evaluate the resulting series of tests using both the standard rule of $p < .05$ and also the stricter **Benjamini-Hochberg procedure**, with FDR control at .05.

Remark:

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. Pooling this data together (pairs from all communities) should allow the most information about the rate/temperature relationship, while possibly limiting the effects from other variables.
The results of the following testing scheme change as we continue to add more data (growth rates from the current date). This suggests possibly running the testing scheme for data *up to time* i and then computing derivatives of correlations, or p-values with respect to i, for predictive purposes.

Preliminary Results: April 16

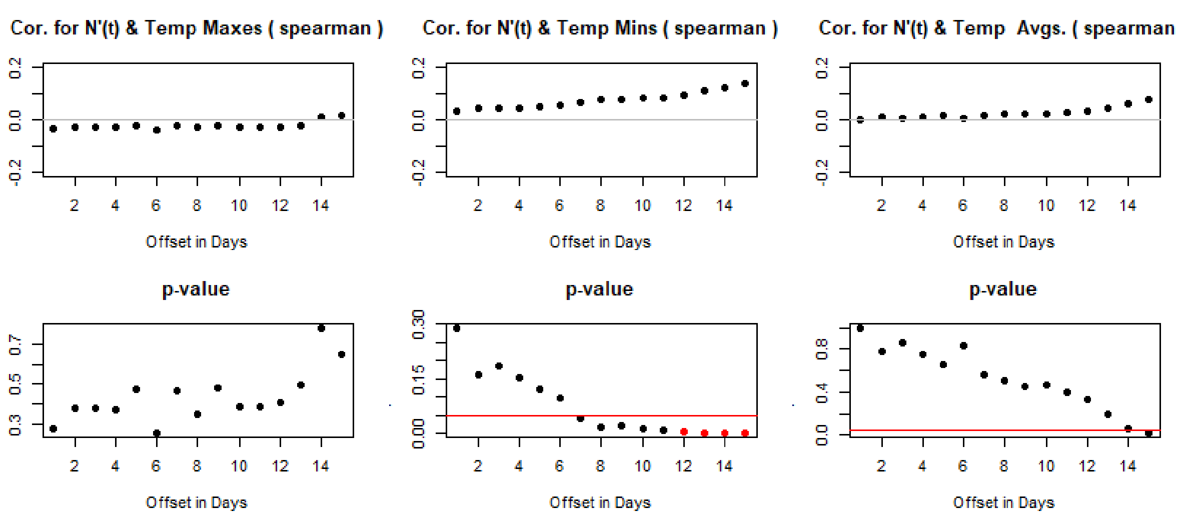
For now, we pool together pairs of temperature and rate data for the following 27 regions. This list contains 13 of the highest case totals in the U.S.

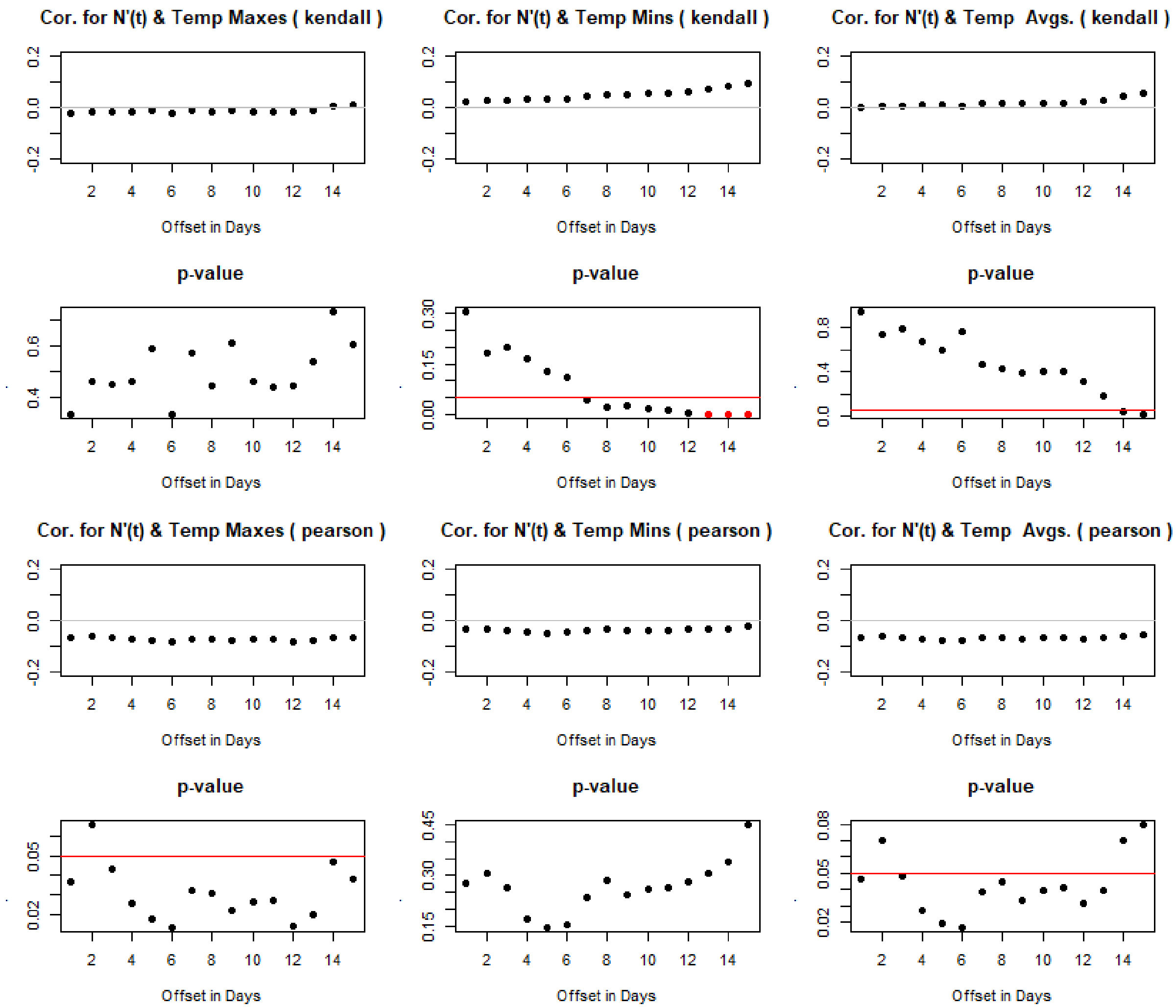
We obtain temperature data for individual counties/cities from the interactive map by the NOAA (URL at the end of this document). This is time series data, recording either the daily high, daily low, or daily average temperature, in the desired county, every day from January 1 to April 16.

Linear Difference Rule. We run 15×3×3 correlation tests, for each of the following combinations:

1. Temperature offset from 0 to 14 days.
2. Max, Min, and Average daily temperature
3. Spearman, Kendall, or Pearson correlation

To clarify, this is a sequence of nine multiple tests (one multiple test for each choice of temperature variable and correlation statistic). Below, we summarize our results with plots: the red line is at $p = 0.05$, and p-values which are highlighted red correspond to null hypotheses rejected by the Benjamini-Hochberg procedure at $q = 0.05$. (i.e. these p-values correspond to **nonzero correlations**)



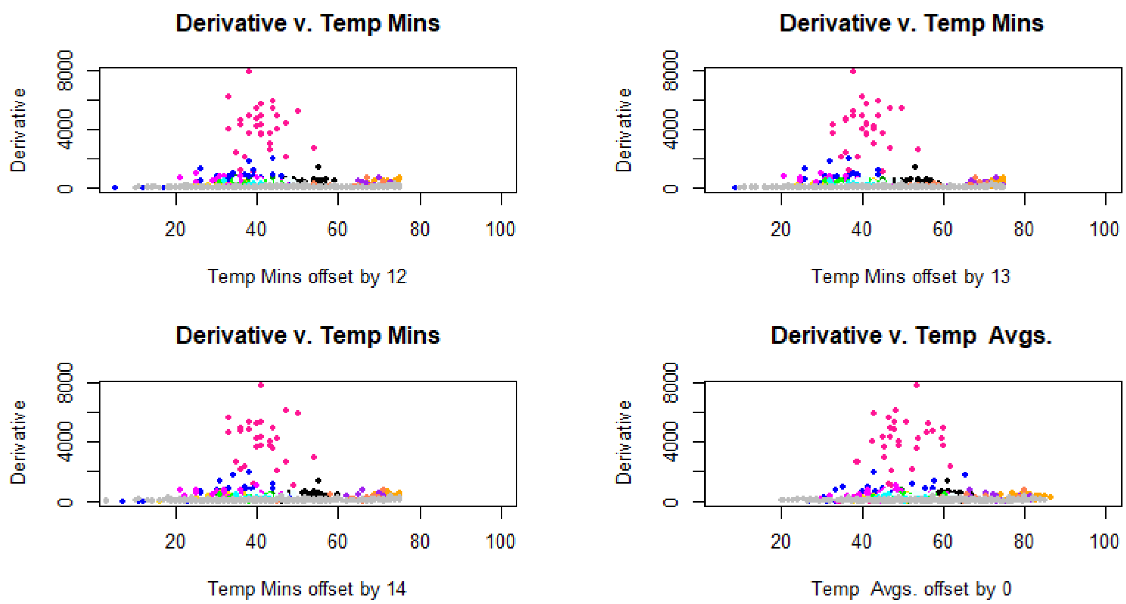


Using Pearson's correlation is questionable since these data are certainly not normally distributed (rate data is nonnegative). Both BH and the standard rule ($p < .05$) fail to reject any null hypotheses (i.e. fail to reject any hypotheses that correlations are zero).

Kendall's τ gives interesting results. We see τ is not correlated with max temperature, for any of the time offsets we checked. By the BH procedure at $q = 0.05$, **we fail to reject hypotheses that τ is not correlated with minimum temperature, offset 12-14 days**. By the BH rule, τ is not correlated with average temperature, for any offset. Spearman's ρ gives similar results as τ.

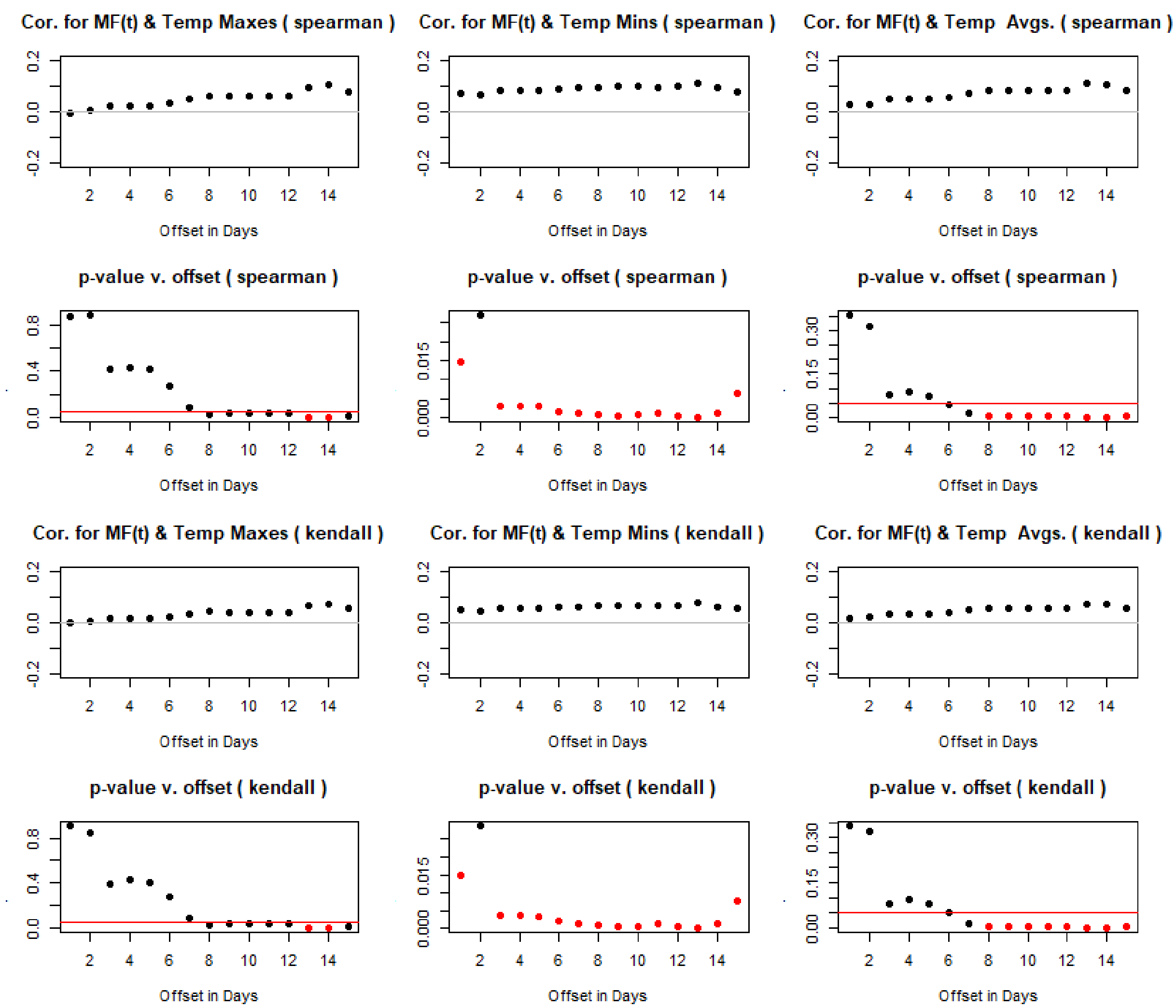
It is interesting to note that while the linear difference rate is possibly correlated with the minimum temperature from $\approx$ 2 weeks before, the correlation is still *mild* ($\tau < 0.1$). **All of the above correlations have $|\tau| < 0.15$.**

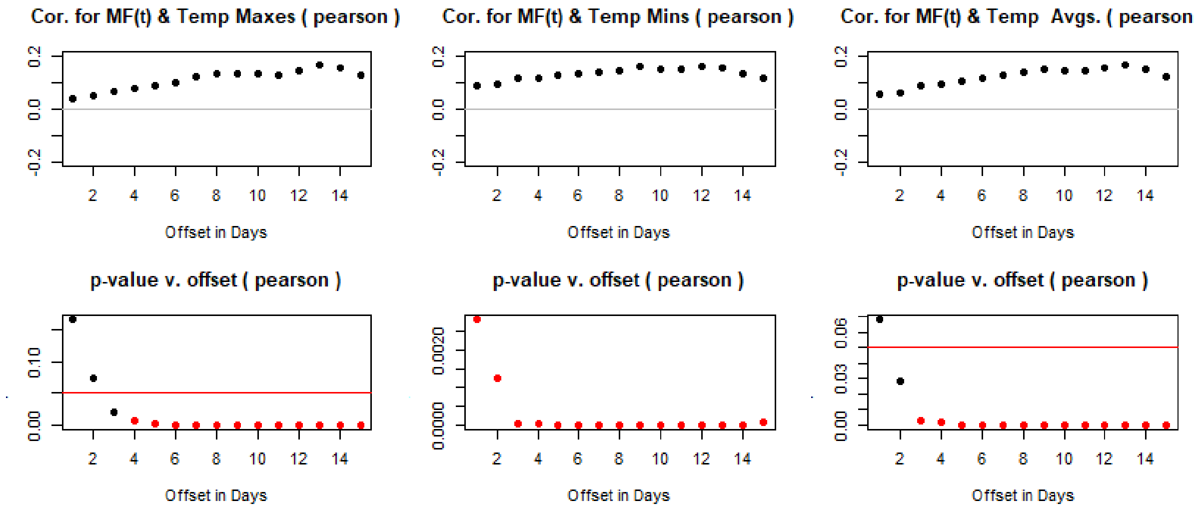
When we examine the graphs of linear difference vs. minimum temperature, offset from 12 - 14 days (these were the significant hypotheses), notice in the first 3 graphs we have several points with extremely low temperature and low growth rate. It is possible these points influence the result and are the source the significant positive correlation.



We will run the same sequence of tests with these points identified and removed (they may be extremely low temperatures in Chicago during periods when there was only one infected person) as part of our analysis.

Multiplicative Factor Rule. We run the same 15×3×3 correlation tests as above, but for the daily multiplicative rates (after the first day at which the rate was positive) instead of linear difference (rate approximation). Here are the results:

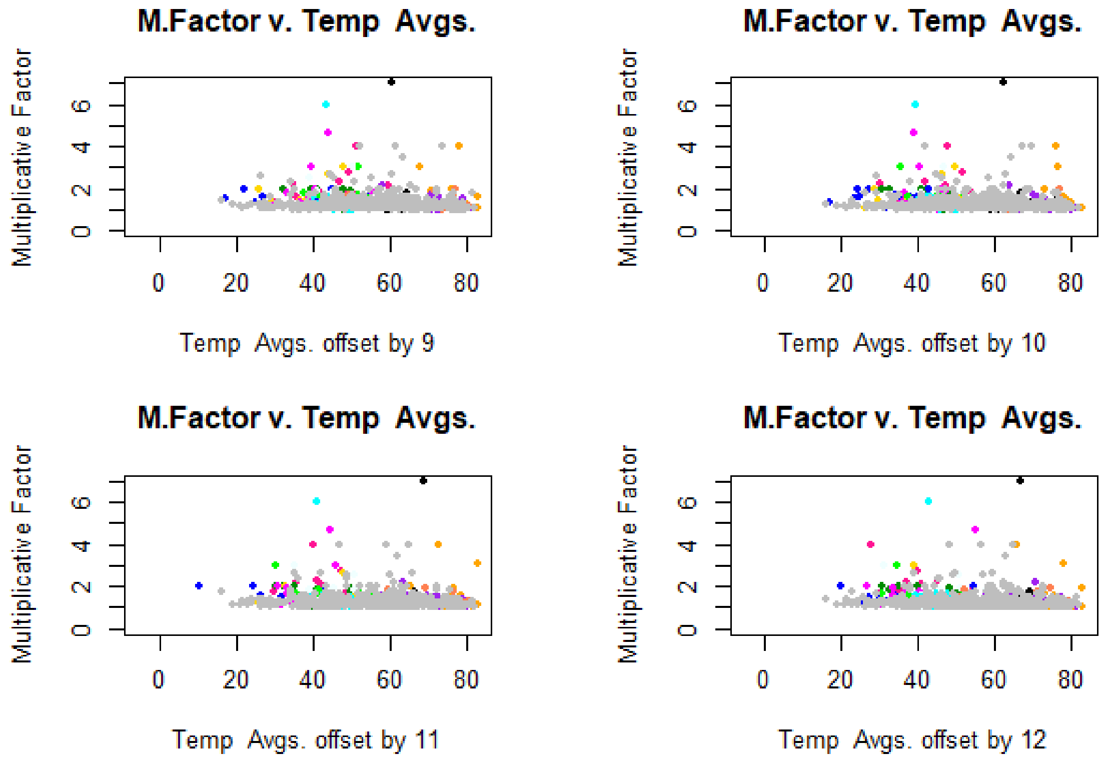




All correlations are bounded in absolute value by 0.2, so if any are true, they are *mild*. Pearson's test gives the most extreme correlations, and both our evaluation methods reject most of the nulls (uncorrelation hypotheses)

Kendall's statistic behaves similarly to Spearman's. These tests reject hypotheses that max temperature from 10-12 days offset is uncorrelated with the multiplicative factors. These also reject almost all nulls for minimum temp vs. multiplicative factor, and all nulls for average temp offset 7-14 days vs. multiplicative factor.

The respective plots of multiplicative factor v. temperature are not enlightening:



Summary of Results: April 16

We can not conclude the 1^{st} linear difference at day t is correlated with minimum temperature from 12-14 days prior, without analyzing the outlying cold points. We conclude the daily multiplicative factor $\mathrm{MF}(t) = N(t+1)/N(t)$ at day t is correlated with maximum temperature from 12-13 days prior, minimum temperature from any of 3-14 days prior, and average temperature from any of 8-14 days prior, with a positive correlation less than 0.12.

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

COVID-19 data:

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

Weather from NOAA:

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