

Advanced Quantitative Methods: 07 Problem Set

```
#####  
#  
# 07 PROBLEM SET  
#  
# DUE:          April 3, 2020  
#  
# CREATED BY: Lauren Sullivan  
# FILES:        "time_series_2019-ncov-Confirmed.csv"  
#  
#####
```

In order to more fully understand autocorrelated data, we will be exploring time-series data on confirmed Covid-19 cases in countries across the globe. This dataset can be found at this website, and it is compiled by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). I'm going to let you all have a bit more freedom in your homework data analysis this week. Your job for this problem set is to explore the Covid-19 dataset. You will need to do the following things:

1. Decide if you want to look at just the US data (by state) or look at the global data (by country) - whichever way you choose, you will have to think about how to deal with the fact that there are multiple observations (rows) per US state and country.
2. Find another dataset that you can merge to this one, that has a factor that you think might influence the number of Covid-19 cases (per country/US state). Some examples include: GDP, mean annual temperature, total population, etc. Don't worry if your results end up being non-significant, just try something!
3. Analyze your data to find out how $\#Covid-19_cases \sim your_variable + time$. Some things to think about include random effects, distribution of your data, etc. Make sure you try a model with and without autocorrelation (AR1 or ARMA) and report what differences you find between the two models. *Hint: check out `lubridate()` here and here. This will help you parse the data.*

You will work in your groups to complete this assignment. On April 3, you will have to present (~5 min) on your analysis in class. Please prepare a presentation that includes: 1) What extra variable you chose to look at, and how/why you cleaned your data? 2) What statistical analysis you chose and why? What was the difference when you used an autocorrelation structure vs when you did not? 3) Provide a figure to demonstrate your results. 4) Tell us at least one problem you ran into? Were you able to solve it? If so, how?

If this ends up being too difficult to do in remote groups, let me know and we can come up with a different plan. But give it a try!