DATA 252 / DATA 551: Homework 8

- This homework is due by April 6, 2020 at the beginning of class. **You need to submit your answers on Moodle in a pdf document**. In addition, there will be a short quiz at the beginning of class, which might contain contents from this homework (including the assigned video), in addition to contents from the most recent lecture.

1. This week, top U.S. government scientists have estimated that coronavirus could kill 100,000 to 240,000 Americans (https://www.nytimes.com/2020/03/31/us/politics/coronavirus-death-toll-united-states.html). One of the models cited is http://covid19.healthdata.org/projections. Play with this website and answer the following questions (since the website is updated daily, you might get slightly different answers depending on which day you are doing this question.)

   (1) For the number of deaths per day, what is the projected peak date?
   2644

   (2) Which date has the most uncertainty in the estimate of deaths per day? What is the range of the estimated deaths on that day?
   24th April. 759-4323 deaths per day.

   (3) Browse through the FAQ page (http://www.healthdata.org/covid/faqs). What does the model assume in terms of social distancing? How is the shape of the model decided?
   Assumptions related social distancing:

   - Model assumes continuous social distancing throughout the modeled period.
   - Followed New Zealand Government alert system Level 4. And assumed if some location has fewer than 3 measures then they will follow level 4 measures within 7 days.

   The shape is based on other Covid-19 outbreaks throughout the world.

2. Read this article https://fivethirtyeight.com/features/why-its-so-freaking-hard-to-make-a-good-covid-19-In your own words (do not just copy and paste from the article), list five reasons that make COVID19 difficult to model.

   - Virus effected differently for different demographical people.
   - People have different underlying health conditions (comorbidities)
   - No cohesive way of gathering data. Uncollected data (could add omitted variable bias) or be misleading data.
   - Asymptomatic people are hard to identify. That makes hard to calculate true fatality rate.
   - Virus transmission is variable based on social behavior, local environment, political decisions, living situations, jobs. Considering all of them makes modeling difficult.

3. Watch this very cool video on simulating an epidemic: https://www.youtube.com/watch? v=gxAaO2rsdIs (this is the same one that Prof. Kouh mentioned in class) and answer the following questions.

   (1) You might have seen the term "SIR model" in many places these days. What does each letter in SIR stand for?

   Suspectable

   Infectious

   Recovered/Removed

   (2) Modeling an epidemic is hard (for all the reasons you've listed in Q2), and that's why simulation can be very useful: we don't have to estimate all the parameters; we can play with different parameters to see what would happen. For instance, *infection radius* (which depends on the level of engagement with other people) is one of the simulation parameters investigated in the video. What are some other simulation parameters in the video? List at list three.
   Travel distance
   Social distancing
   Probability of infection

   (3) According to this simulation, what would be the most effective measure to contain the epidemic? Do you think this measure would be difficult to implement? Briefly explain.
   Social distancing is one of the most effective solutions. However, it is hard to implement as some people cheat which can cause significant increase in the spread. Furthermore, even after the pandemic is controlled, if the people go back to normal social interactions that could lead to a second wave.

   (4) Discuss something you find interesting or surprising from this video.

Visualization of simulation was cool.