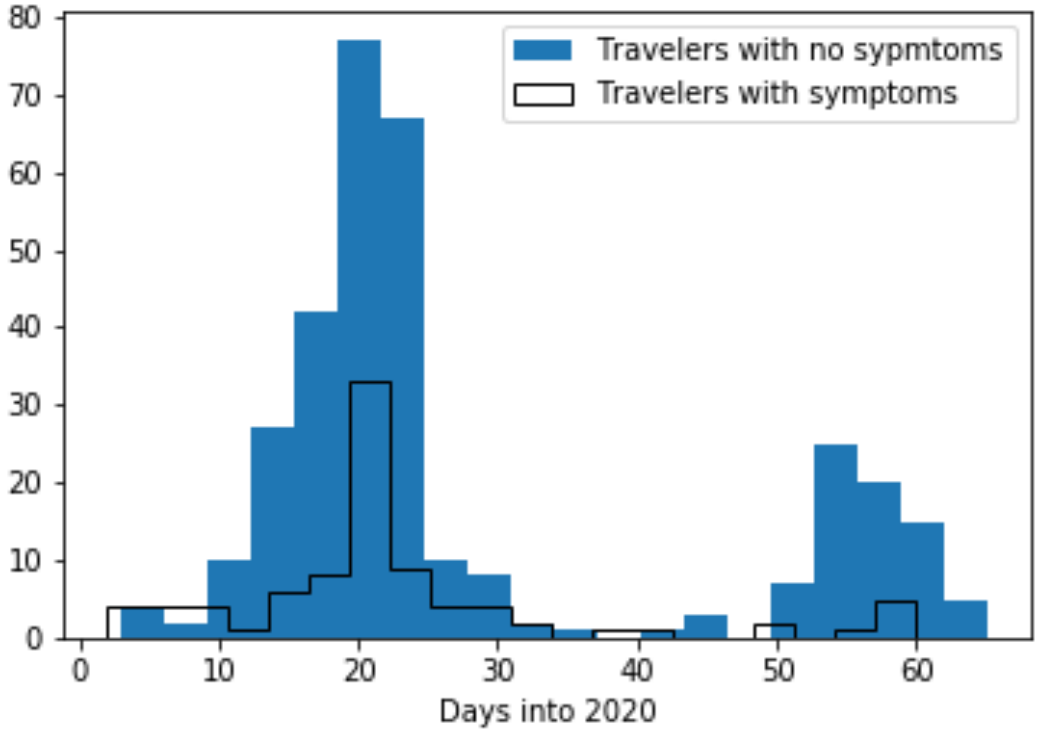
Sick Travelers

I'm going to explore some of the COVID-19 cases recorded by Xu et al. in their Github repository at https://github.com/beoutbreakprepared/nCoV2019/tree/master/latest_data, with details in https://www.nature.com/articles/s41597-020-0448-0.pdf . I chose to focus on the patients who had traveled while sick, and whose travels were well enough documented to analyze. The idea was appealing because of the realization that only when someone first carried coronavirus to a new area was it possible to trace how it was moving around the world (Once the virus was spreading within a community, it was too hard to determine how it was moving, without someone doing a lot of detective work and publishing the findings). By looking at the dates of travel, the dates of onset of symptoms, and the dates of hospitalization, we can get not only a nice partial visualization of how a pandemic occurs, but also some statistical proof of the way people's actions accelerate the rate of contagion.

The first part of this exploration involves 316 patients and 415 segments of travel, between Jan.2 and Mar.5. About 20% of the travelers already had symptoms, or at least thought they did, and the other 80% had already been infected but hadn't shown any symptoms yet, or at least claimed not. Here's the distribution:



The mean day of travel for people who already had symptoms was about Jan.23, and for still-asymptomatic people, Jan.28. So that seems initially encouraging—People were maybe becoming more aware as news reports came out, and they were either choosing to travel less once they noticed symptoms or maybe they were being prevented from doing so by somebody.

But let’s see what’s going on with the first, highest peak, for both types of travelers, just after Jan.20. Its overall shape arises from the fact that the virus was spreading rapidly in China in mid-January, so much so that China shut down travel to and from Wuhan on January 22nd, leading to the sudden dropoff in sick travelers on the chart. But the chart also suggests that people caught wind of the imminent Wuhan shutdown and made a last-ditch effort, knowingly sick or not, to flee to their hometowns before it was too late. The evidence for that lies in the ratio of symptomatic travelers to asymptomatic ones in the prior week, where the ratio appears to be less than 1 to 4, compared to the ratio on 20-21 Jan, where it spikes up to maybe 3 to 7. Let's see the actual data:

Proportion of travelers with symptoms Jan.13-19	0.13
Proportion of travelers with symptoms Jan.20-21	0.28

Taking 10,000 random permutations of the travels for these 9 days, the observed difference here is about 3 standard deviations above the mean sampled difference of 0. In other words, this is not just a random distribution, but rather is the result of people traveling at any cost in order to escape being cordoned off from the world.

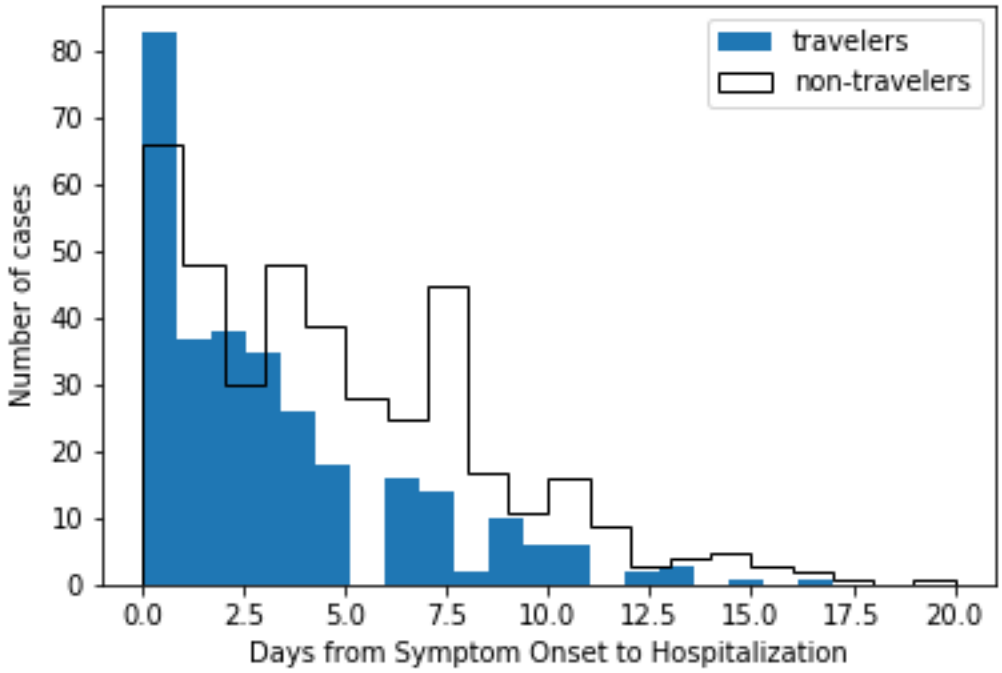
The part of this analysis that's somewhat questionable is that it smacks of the "Texas Sharpshooter" phenomenon, since I began by focusing on a specific part of the data that seemed interesting, and then I drew a circle around the perfect dates that perhaps minimized p-values for the null hypothesis. What I would say in my defense is that the dates before the selected ones contain very few cases, and indeed the symptomatic travelers at that point most likely had no idea their symptoms might be from COVID-19 (The Chinese government only announced the existence of the new coronavirus on Jan.7). Also in my defense, the most likely underlying cause for the data I circled--the Wuhan shutdown--draws its own endpoint in time, on Jan.22. And finally, as I pointed out, this analyzed trend runs counter to the overall trend, which makes it inherently less likely and more worth exploring.

My next question is whether sick travelers are reporting their dates of onset of symptoms differently from sick non-travelers. The idea is to compare the time between STATED onset of symptoms until hospital admission, for people who traveled vs people who didn't. Here are some potential underlying reasons: People who don't travel could be more inclined to list their symptom onset sooner, and they also might wait longer to go to the hospital, being in the safety of their homes and knowing the hospital was nearby if absolutely necessary. Either one of those proving to be true, the difference in dates should be longer for them vs. for travelers. Travelers might not want to believe or admit they have symptoms, thus stating the max possible date for that, but their hospital visit might go either way: No one wants to be stuck in a foreign hospital, to miss the balance of their vacation, or to deal with the financial costs of rescheduling their return home, so they're more likely to ignore warning signs while traveling. On the other hand, if they started feeling a little sick BEFORE their travels started, they might have decided to plow onward optimistically, only to find themselves in dire straits on vacation, and perhaps paranoid about what they're hearing about coronavirus, to the extent that they visit a doctor while traveling.

For this part of the analysis I start with all of the the cases reported as of now, April 20. The ~430K rows that are listed, after removing duplicates and constraining the patients to only those whose dates of travel, symptom onset, and hospitalization are listed, boil down to just

around 300 travelers and 400 non-travelers. Sure enough, the mean number of days for that period for non-travelers is considerably higher:

Days of symptoms before hospital, non-travelers	4.43
Days of symptoms before hospital, travelers	3.13



I checked 10,000 random permutations of the data, and only 1 showed a higher difference, so there clearly is a reason we see this effect.

The more you look at these sick traveler cases, the more you realize just how dangerous they are:

1) People are spreading the disease between communities, increasing the odds of a pandemic.

2) They're claiming over 1 day less of being symptomatic before going to the hospital than are non-travelers. Unless you want to believe that they're just a more cautious group of people, or are more worried about their symptoms just by virtue of having traveled, the most likely explanation is that they want to report their symptom onset later than it actually happened, perhaps because they know how dangerous their actions may be perceived.

3) The standard modes of transportation bring many travelers together in close quarters, making airborne and surface-borne diseases especially contagious. And the part of this disease's lifecycle when the patient is most contagious is at or near this point when symptoms first appear. Link to source article about contagiousness for COVID-19

I'm not claiming travel is bad just because it's dangerous, and I'm also not claiming people are bad for trying to believe their symptoms actually showed up the second their flight home landed. Besides the fact that people are only human, there's the fact that "When did you first have symptoms?" is often very hard to answer accurately. So all I can do is point out the obvious fact that we need better screening of travelers, preferably before they board planes, trains, buses, and boats. Much better screening than testing their forehead temperature after their trip is over, when they might have taken antipyretics to reduce their fevers. Or maybe we need to make it easier for people to get money back when they cancel travel plans at the last second, thinking they may possibly be sick. We've seen the alternative, and it's more costly than any preventative measures like these.

At the very least, the data explored here provide statistical support for what you probably already suspected: People will place their society at greater risk, if doing so makes themselves feel more secure. And scientific findings that rely upon people stating when they first experienced symptoms are built on shaky ground, especially if a large proportion of the people are travelers, as in the case of this pandemic. The source article linked to, above, studies the question of when a COVID-19 patient is most infectious. The question has a huge bearing on how we go about dealing with the pandemic, and the article's researchers conclude that a patient is likely most infectious just before his own symptom onset. Yet they end their argument by pointing out that if people tend to state onset dates that are later than actual onset dates, the point of maximum contagiousness moves later in time. Considering the huge ramifications and the data presented here, we should assume that this is the case.