Nathan Walk
10/30/20

Exercise 1:

1. Date of data: 10/22/20
2. What columns have missing data:
   fips: 30, deaths: 79, confirmed_cases: 747, confirmed_deaths: 1218, probable_cases: 1930, probable_deaths: 2411
3. No FIPS code is related a city's sprawl over multiple counties and the adjusted coding they use for this
4. Properties that make outliers of the counties with <2000 confirmed_cases:
   There is a lot of incomplete data here - which I infer. For instance, RI has 3400 cases, but only 25 deaths, and no confirmed_cases.
5. workaround for UnicodeDecodeError: I used pd.read_csv(url, engine="python")
6. Column 3 indicates the state, column 4 the county, and column 18 the 2019 population estimate
7. Not all of the values in the county column indicate counties or county equivalents. So this was tricky..eventually I had to do a lot of reading on how these were set up. The simple method I ended up using was simply to read the county column - for instance, Alaska, which has no counties - these are listed as boroughs, etc. The data seems pretty well coded to account for actual counties, and metropolitan areas.
8. Data cleaning - this is mostly documented in the code I've submitted. It took me some time to get the Counties matched sufficiently for a join.

Exercise 2

1. This is not a clean api - it is not a "readable url that intuitively represents the underlying resource".
2. Yes, it is a RESTful API. It is stateless - meaning that the request we send contains all information necessary to respond to a request.
3. I estimate the gradient by taking 2 sets of points very close to each other and calculating the slope at that point. Stop criteria - I used a range function. Reasonable?...I believe so, after multiple iterations and looking at the data.
4. Global minimum: 0.59, 0.29 - 1.088, local minimum: 0.84, 0.04 - 1.099
   I'll confess to being uncertain my program works. Mixing up a and b, and putting in a bunch of numbers permits a lot of error estimates close to these with a and b far away. For instance, a = .2, b = .7, = 1.100

Exercise 3

1. This problem was challenging. I could get the data grouped and visualized at one point, but I stupidly forgot to save a copy while I continued to fiddle with it. I was not able to put it on the map. So my code doesn't work, currently.

Exercise 4

defer - we can't present the data, as its PHI. Also, we're still trying to nail down what our clinical colleagues need