Arise Citizen Scientist 2020: Ask Questions

by M.A.Tucker 27APR20

**Motivation**

A favorite satirist and punster, Alexandra Petri (Twitter: @petridishes), told a story of how Newton, fleeing to the countryside during a plague, decided to use the time to work out the math leading to the gravitational mechanics we still use today. She encouraged us to do something different during this time, something equivalent to Newton's effort. I was inspired. I had just such a thought in early February as our plague picked up steam. Above the TODO list items such as "paint bathroom ceiling" or "fix leaky faucet", I had placed "discover anti-gravity". It turns out that anti-gravity is a REALLY HARD PROBLEM. So as an alternative, perhaps slightly more achievable goal, I decided to try to better understand our COVID travails. By asking and answering questions, I hoped to better judge the veracity of often conflicting pronouncements.

**Method**

Citizen Scientists have an important role to play in understanding our current COVID travails. There are hundreds of thousands of Citizen Scientists contributing worldwide. The following leverages their extensive knowledge.

We are blessed in the USA to have ready access to a vast amount of data and powerful computer resources (much of it free). Data capturing various aspects of the plague are maintained and updated daily by the CDC, Kaggle.com, Johns Hopkins and many other organizations. However, huge streams of numbers purportedly describing something become rather mind numbing. My goal as a budding Citizen Scientist is to understand the data and capture the essence in some visualizations (e.g. graphs, tables). In short, convert "data" into "facts". How hard could that be?

By looking at smaller sized areas, you may discern aspects not apparent if all of the USA or world is considered. Creating a Kaggle notebook at www.kaggle.com/imdevskp/corona-virus-report allows you to easily use the latest USA-County-Wise data. The answers below are based on a snapshot of this data as of April 27th, 2020. It changes by the hour with updates posted daily.

Here is a link to a Johns Hopkins dashboard focusing on the same USA county wise data (<https://coronavirus.jhu.edu/us-map>). If you reference the Johns Hopkins dashboard tomorrow, you'll likely see different results. It will have been updated while the data used for this post will be dated.

**Questions**

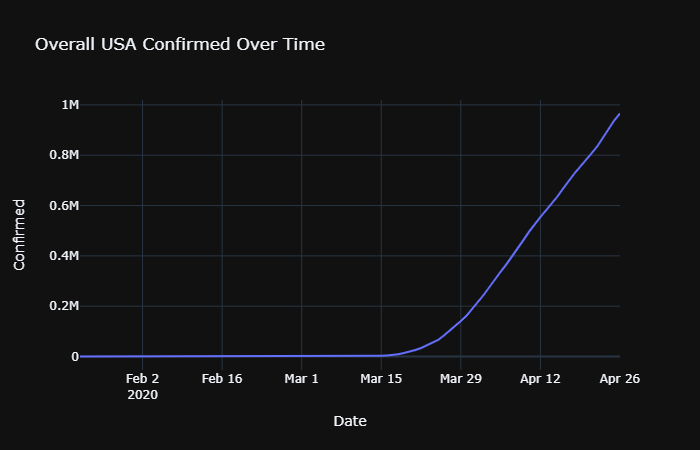
The following questions and some "answers" are based on our best (but undoubtedly flawed) available data (based on data gathered through April 27th, 2020).

Is it a hoax?

No, it's a modern day plague.

Is it almost over?

No, it's spreading rapidly with increasing numbers of confirmed cases & deaths. If the number of daily new cases begins to decline, we'll see the "curve flatten". In some areas (e.g. Pittsburgh, PA), there were no deaths reported yesterday after a steady drumbeat. Also, the "doubling rate" has extended from 2 days to 2 weeks in some areas. This is very good news. However, it's still doubling.



Do various US county sized areas "look alike"?

All areas are experiencing increasing numbers of confirmed cases. The meaningful differences relate to infection rates (confirmed cases given the population size of the area) and mortality rates (deaths given the number of infections). Choosing a mix of mid-sized cities and suburban areas vs. a baseline (first reported cases in USA) of King County, WA, leads to more questions than answers.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **region** | **population** | **confirmed cases** | **infection rate** | **deaths** | **mortality rate** |
| King County, WA | 2,253,000 | 5863 | 0.26% | 408 | 6.96% |
| Pittsburgh, PA | 1,216,000 | 1211 | 0.10% | 73 | 6.03% |
| Kansas City, KS | 890,089 | 1043 | 0.12% | 67 | 6.42% |
| Youngstown, OH | 228,683 | 737 | 0.32% | 66 | 8.96% |
|  |  |  |  |  |  |
| Montgomery County, MD | 1,051,000 | 3645 | 0.35% | 180 | 4.94% |
| Snohomish, WA | 822,000 | 2311 | 0.28% | 104 | 4.50% |
|  |  |  |  |  |  |
| Columbus, OH | 1,683,751 | 2199 | 0.13% | 58 | 2.64% |
| Virginia Beach+, VA | 910,385 | 732 | 0.08% | 19 | 2.60% |
| Prince William, VA | 470,335 | 1265 | 0.27% | 19 | 1.50% |

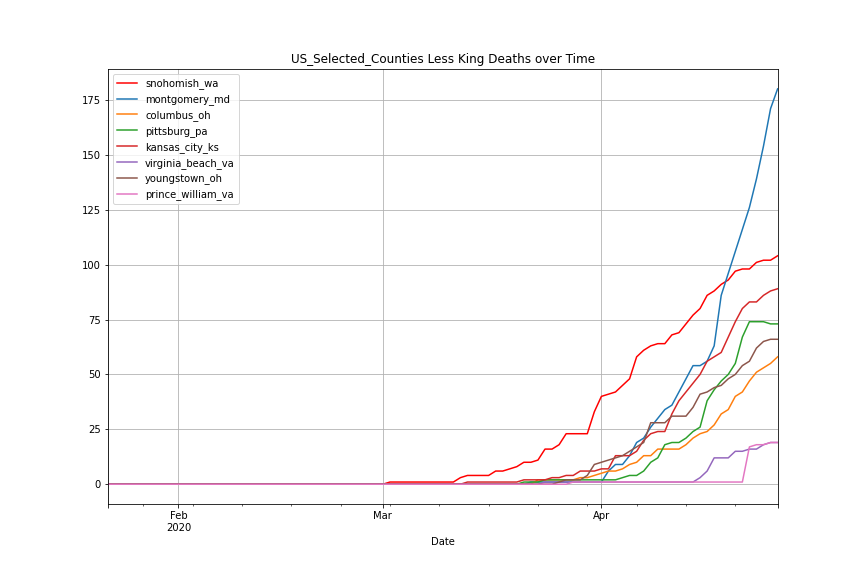
These areas are labeled by "city" but the results are for the relevant county. Several are a combination of counties including:

|  |  |
| --- | --- |
| **region** | **county enumeration** |
| Columbus, OH | Franklin, Delaware, Fairfield |
| Kansas City, KS | Wyandotte, Douglas, Johnson |
| Virginia Beach+, VA | Virginia Beach, Chesapeake, Norfolk |

King County, Pittsburgh, Kansas City and Youngstown, OH all have a very high mortality rate (while varying in population from 2.2 million to a couple hundred thousand). Why? Due to limited testing, the actual number of infections is likely much higher. This could lead to the mortality rates being inflated. Many cited mortality rates are much lower (~2%) due to modeling of infection rates in the population rather than based on the known data.

Montgomery County, MD and Snohomish, WA have similar mid-range mortality rates. Montgomery County, MD recently surged past Snohomish, WA. Since February, Snohomish, WA had been near the top of the leaderboard. Why is Montgomery County surging? It should be noted that this is not a contest you want to win. Montgomery County, MD (my home) does not appear to be leveling off like many areas. Why? We've closed restaurants, bars, venues, "non-essential" businesses since early on. We're wearing masks, social-distancing, etc. Are these measures inadequate? Or are we recognizing infections and deaths "better" than areas with low rates?

Columbus, OH, greater Virginia Beach, VA and Prince William, VA are all remarkably low in both infection rates and mortality rates. Are they "behind the curve" and due to "catch up" or will their lucky state persist?



What's next? Can we predict well?

We don't know what's next. No, we can't predict well. General, wide-spread testing is required for any level of precision in our predictions (rather than the current rationed testing when severe symptoms present).

**Your Data Must Be Wrong!**

While I have cross-checked much of the data, I could very well have made mistakes leading to errors. Also, this analysis is obviously very basic and barely scratches the surface. I'd welcome the opportunity to share the Kaggle notebook that generated this & many more charts. (Kaggle.com is a free cloud based org that provides the computer resources that would allow you to ask your own questions). Moreover, you might bring unique insights to the problem. If you have a bit of time and the inclination, it's a fascinating & timely time to study. Kids from 6-60+ can dig in and discover answers. Arise Citizen Scientists - Ask Questions!