Arise Citizen Scientist 2020: Know Your Facts

These sections are intended to stand by themselves so do skip to those that interest you rather than feel like you have to wade through each in order:

* **Motivation**
* **Method**
* **Questions**
* **Findings**
* **Accuracy**
* **Follow-on Questions**
* **Resources**
* **Broader Questions**
* **Easier Problems**
* **Acknowledgements**

**Motivation**

Reading a favorite satirist and punster, Alexandra Petri (Twitter: @petridishes), sometimes brings chuckles, often groans but a recent column near the end of March was nothing short of profoundly inspiring. She 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 gravitational mechanics. She encouraged us to do something different during this time, something equivalent to Newton's effort. Then, of course, she careened off into the weeds with groan-worthy puns, similes and analogies.

However, I was inspired. I had just such a thought in early February as our plague picked up steam. So above the list items such as "paint bathroom ceiling" or "fix leaky faucet", I 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 become a Citizen Scientist. Here's the story so far.

The plague news is disturbingly short on science. The pronouncements, updates and prognostications are understandably filled with emotionally charged reporting. Stories about people facing fear, heartbreak, loneliness, destitution, powerlessness. All reminders of the old adage: "There, but for the grace of God, go I…". Many experts offered dramatically differing opinions ("it's a hoax!", "millions may die!", "it's the flu", "it's a death sentence..."). But I want to see if we could use science to better understand patterns, judge credibility and perhaps even support the adoption of new behaviors.

~~So I adopted a maxim from the old arms control negotiations - "Trust but verify".~~

We are blessed in the US to have ready access to a lot of data. Data capturing various aspects of the plague are maintained and updated daily by the CDC, Kaggle.com, Johns Hopkins, other orgs. However, huge streams of numbers purportedly describing something become rather mind numbing. My goal is to understand the data and capture the essence in some visualization (e.g. a graph). In short, convert "data" into "facts". How hard could that be?

I first decided looking at data from all the United States would likely "wash out" emerging trends and particularly, regions of interest. For example, confirmed cases in the US are ~600,000 while data from Seattle and New York with thousands of cases would mask trends in different regions of interest such as my area with "mere" 1000 cases. Johns Hopkins (<https://coronavirus.jhu.edu/us-map>) just launched a site showing US county wise data with a similar motivation.

**Method**

The initial task is to extract the data for various areas of interest:

* King County, WA
* Snohomish, WA
* Montgomery County, MD
* Columbus, OH
* Pittsburg, PA
* Kansas City, KS
* Virginia Beach/Norfolk/Chesapeake, VA
* Youngstown, OH
* Prince William, VA

These areas are selected for several reasons. King County is the first area impacted by the plague. It serves as a reliable "baseline". The other areas are mid-sized or suburban cities where I know, love and talk to people. I have at least anecdotal evidence of the timing various mitigation methods (or lack thereof). Questions are asked for each of these areas.

**Accuracy**

Wherever possible, the findings are crossed checked with credible sources. However, there are distinct possibilities that the findings are WRONG. The data may be incomplete or inaccurate. For example, with limited testing available, deaths may be mis-attributed to pneumonia or other respiratory ailments. Also (unavoidably), the budding Citizen Scientist may have inadvertently made mistakes.

Moreover, generalizations will be made that may gloss over nuances.

**Questions**

For each area:

1. Do confirmed cases and deaths "look similar" to King County, WA?
2. Do confirmed cases and deaths "look similar" to each other?
3. Does population affect the magnitude of confirmed cases and deaths.
4. Is there a fixed timeframe for mitigation methods to take effect?

**Findings**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **region** | **population** | **confirmed cases** | **infection rate** | **deaths** | **mortality rate** |
| **King County, WA** | **2,253,000** | **4426** | **0.20%** | **292** | **6.60%** |
| **Youngstown, OH** | **228,683** | **417** | **0.18%** | **31** | **7.43%** |
| **Kansas City, KS** | **890,089** | **675** | **0.08%** | **38** | **5.63%** |
|  |  |  |  |  |  |
| **Snohomish, WA** | **822,000** | **1867** | **0.23%** | **69** | **3.70%** |
| **Montgomery**  **County, MD** | **1,051,000** | **1631** | **0.16%** | **42** | **2.58%** |
|  |  |  |  |  |  |
| **Pittsburg, PA** | **1,216,000** | **857** | **0.07%** | **19** | **2.22%** |
| **Columbus, OH** | **1,683,751** | **1118** | **0.07%** | **16** | **1.43%** |
|  |  |  |  |  |  |
| **Virginia Beach+,VA** | **910,385** | **454** | **0.05%** | **1** | **0.22%** |
| **Prince William, VA** | **470,335** | **389** | **0.08%** | **1** | **0.26%** |

1. Do confirmed cases and deaths "look similar" to King County, WA?

Kansas City, KS and Youngstown, OH "look similar" to King County: high mortality rates (~5-7%) vs the mortality rates in other areas (~0-3%).

2. Do confirmed cases and deaths "look similar" to each other?

**Follow-on Questions**

**Resources**

**Broader Questions**

**Easier Problems**

**Acknowledgements**

Acknowledgements:

Alexandra Petri Twitter: @petridishes

CDC

Kaggle.com

Johns Hopkins CSSE