Story-Telling Assignment

Ask the following questions and look for the answers using code and plots:

Hypothesis:
My hypothesis is that the data will show people above the age of 50 are most likely to die from infection.

Can you count something interesting?

Yes, I can count the types and numbers of Coronavirus victims and the spread of the virus by

locations. I can also see what the relationships are between those who die after contracting the virus

and their other variables (age, weight, time of infection, etc.)

Can you find trends (e.g. high, low, increasing, decreasing, anomalies)?

Yes, I can see that the trends are that older people are more likely to die, as well as males.

Can you make a bar plot or a histogram?

Yes, I can make a bar plot of the infected, and a histogram of the results of those infected

(recovered, still infected, deceased).

Can you compare two related quantities?

Yes. I can compare the quantities of infected people, the numbers of dead, and numbers of

recovered. This is more than two, but still doable.

Can you make a scatterplot?

Yes, I can make a scatterplot of the data.

Can you make a time-series plot?

Yes, I can make a time-series plot for the rates of infected by date and city.

Looking at the plots, what are some insights you can make? Do you see any correlations?
 Is there a hypothesis you'd like to investigate further? What other questions do the insights lead you to ask?

I can make insights about possible gateways, connections between hosts and the virus' spread over time. This would lead me to ask which countries the virus spread to first, and can show the level of traffic of the infected to foreign countries from China.

Now that you've asked questions, hopefully you've found some interesting insights. Is there a narrative or a way of presenting the insights using text and plots that tells a compelling story? What are some other trends/relationships you think will make the story more complete?

Yes, I can use the text and plots to tell a story about how many people are different types of hosts throughout the world. I can also use a neural network to build a network mapping, where it can guess which types of hosts are around which part of the world as a new infected shows up every 15 minutes, and associate the relationships between hosts to guess which type of vulnerability scanners would be the best fit for those hosts. This would require significantly more information from datasets that may or may not be public due to HIPPA laws. I can also just use the data to create a neural network that works based off of the given data to estimate the numbers by city and avoid mapping people's interrelational networks altogether.