Deep Shah

Professor Kowal

HQSS 370: Data Storytelling

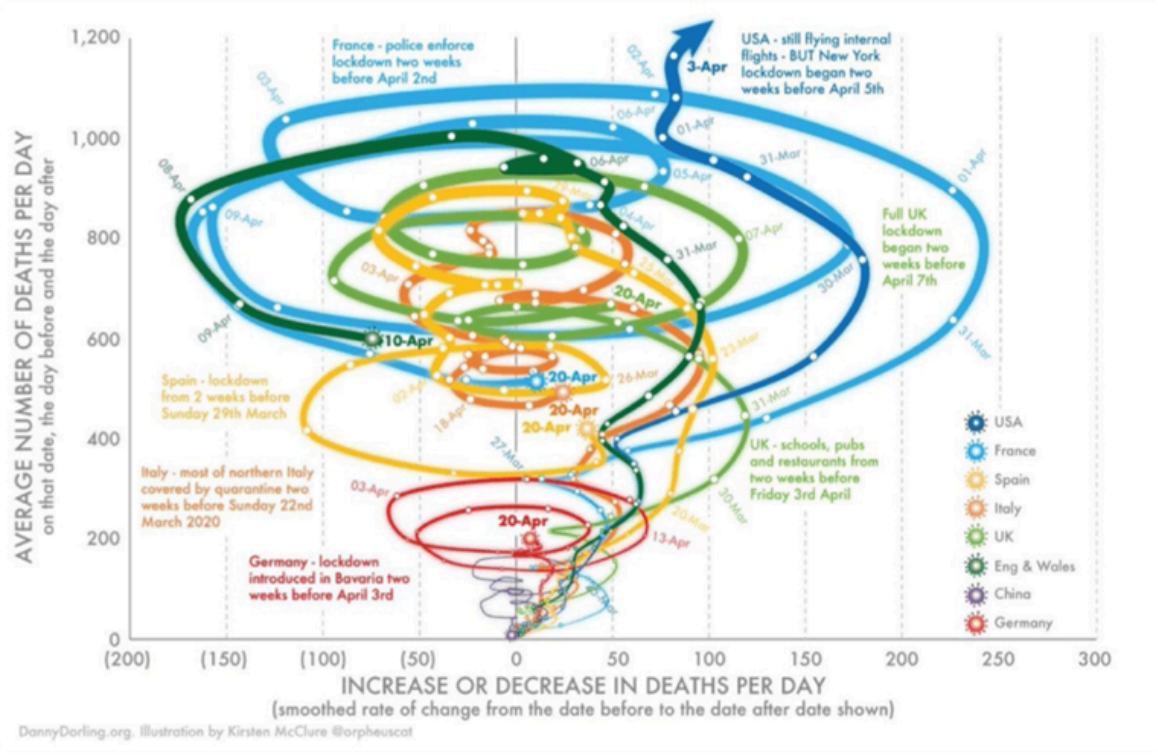
3/30/2025

"I pledge my honor that I have abided by the Stevens Honor System." - Deep Shah

Midterm Exam

For this midterm, you will have unlimited time. You can consult notes and readings from the class or do your own research. Be sure to cite any sources. This is an individual exercise, so you may not collaborate on this assignment.

1. Below you will find a data visualization of the 2020 COVID Pandemic. In one or two paragraphs, answer the following questions. Is this a good data visualization? If so (or not), why? How could you improve the readability of this graphic to better convey this information to the public?



Answer: After analyzing the data visualization of the 2020 COVID Pandemic, this is not a good

data visualization. First of all, I feel that there is a lot happening in this visualization, which can confuse the audience. In terms of the visualization, I feel that there is a lot of content written in the image. Having a lot of information can overwhelm the audience and lead to misunderstanding of the true meaning of the visualization. As I'm analyzing the image, I find it not to be visually appealing as I am not sure what the arrow on the top of the image signifies and I feel that everything is clustered together, especially the dates of when the 2020 COVID pandemic happened. In terms of the dates, I feel that this is significant to know and should be more appealing for the audience, so making the text larger could make this visualization better and more appealing. Lastly, I feel that the intersection of all the lines are also confusing me as I can't seem to piece together the true meaning of the model. To improve the readability of this graphic to better convey this information to the public, you change the type of data visualization from this to a line graph as you can represent each country and the increase or decrease in the death rate from the pandemic by months. With doing a line graph, you can make each of these lines separate and color code them, so that it can be significant in distinguishing the impact the pandemic had on different countries. All the text that is written for each country could be given in a short description below, so that it doesn't seem overwhelming for the readers. You can keep the legend, which is significant for understanding which lines correspond to the countries. Also, you can make the text larger for the dates and make some of the colors more appealing for the audience to see. Lastly, you can make it seem more balanced by making the lines the same thickness.

2. On Canvas you will find data from the 2022 Cooperative Election Study, along with codebooks (more info can be found here). Using this data and analyzing it in Excel, R, Python or the software of your choice, tell me a story about American life or politics in 2022. You should have a basic research question, and tell me why we should care about your question. You should include at least one data visualization, and should write 1-2

(2-3 for graduate students) pages telling your story. You do not need advanced stats or methods, but you should tell a compelling story.

For American life, going to college is considered an essential step to pursue for higher education and personal & professional growth, but the cost of the education is very high, which may be a concern for many individuals. Many people are passionate about learning more and want to expand their knowledge by really pursuing college, but unfortunately can't pay for it. Alternatively, people that can't afford to go to college would also take the initiative to take out student loans, where they will have the opportunity to study and pursue a degree, while also having the ability to pay back their loans after they graduate. Seeing this being the current issue that is faced in 2025, I am taking the initiative to invest in this topic further and understand the concerns with if students are currently studying and taking out loans in 2022. Specifically, the research question I am exploring is entitled: How does the location of respondents throughout the U.S. influence whether they are paying off student loans and if they are enrolling in school as a student? I am referring to the school as college because it describes a student being full or part time. Lastly, I am really curious to know how it was like in 2022 in comparison to the way it is now as it helps gauge the involvement of students in the consensus.

First of all, the respondents involved in the questionnaires were all located in the United States (U.S.), except for one respondent who is near Democratic Republic of Congo. A variety of loans can be taken out by many individuals, including house loans, car loans, and business loans, but the main one that students take out for education is student loans. To represent the locations of the respondents in relation to whether they are paying off their student loans, I created figure 1 which is displayed in the appendix section. For this map, two types of colors were used for circles, one being orange and the other being blue. The blue represents the respondents responsible for paying off a student loan, while the orange represents them not being responsible

for paying the loan. Throughout the U.S., many of the respondents are not responsible for paying off a student loan. A larger number of these respondents are mostly located on the left-most side of the map, covering parts from Minnesota all the way down to Texas. As a result, this can signify that a lot of the respondents may not be college students or could be college students that haven't taken out a loan. Additionally, one unique element that is represented is from the states of Alaska, Hawaii, and Nevada which have very few responders. This is significant as this is the one of the few states with the lowest respondents. Having these minimal respondents in these states can be represented as underrepresented since they will not have the opportunity to have a big influence on the final result. Lastly, while many respondents answered these questions, there were still individuals unsure or did not have to pay student loans, which plays a significant role in shaping the final result.

Student loans play a significant role in higher education for college students throughout the U.S. in 2022. To illustrate a visualization of the respondents' locations to their status as a student, I created figure 2 and displayed it in the appendix section. For this map, three different colors of circles were used, where one is blue, second is orange, and third is red. The blue represents full-time students, orange represents part-time students, and red is not currently enrolled in school. Throughout the U.S., many of the respondents to this survey are not currently enrolled in school. They have completed school already and could be working in the industry. Additionally, I feel that to fill out these forms, you need to be at least 18 years of age as there are questions in the survey that are mentioned in regards to voting for specific candidates and you need to be this age to at least vote, which would influence the results for the questions. The second most respondents would be the full-time students, which makes sense as this age group would be college students that would be interested in filling out this form as they have the

opportunity to vote. Some unique elements demonstrated for this are that the state of Alaska has three respondents, two not enrolled in school and one being full time. Hawaii has only two, both not enrolled in school and other states with minimal respondents include North Dakota, South Dakota, Wyoming, and Montana. Having minimal respondents in a state is an issue as it shows an underrepresentation of these states since they will not have the opportunity to have a big influence on the final outcome.

Even though both of these topics are significantly concerning individually, combining two together and comparing it to the respondents' location would provide valuable insight about respondents in terms of whether full-time students are currently paying off their loans. To demonstrate this, I created figure 3 and displayed it below in the appendix section. For this map, six different colors of circles were used, where one is light blue, second is pink, third is green, fourth is orange, fifth is dark blue, and six is yellow. Each of the colors correlates to two different types of answers. For example, light blue represents (1, 1), which means that a full time student is paying off a loan. The main purpose of combining the two is to demonstrate what types of students are taking out educational loans for studying in 2022. Based on the model, it demonstrates that most of the respondents are not enrolled in school and aren't taking out loans, implying that they are potentially working in the industry which is similar to figure 2. These results mostly came from the left-most side of the map, emphasizing how significant of an impact the left-most states on the map have. It makes sense that there aren't a lot of full-time students paying off their loans as they could still be studying and would pay it off after they graduate. Another case could be that many people may have taken out loans, but may be paying some of it off, but not all, so they indicated that they haven't paid it off yet. In 2025, I feel that it is common for people to take out loans just to pursue their education, but many full-time students

would feel embarrassed that they took out a loan as they would think that they are being categorized or they may not indicate it as their parents could be the ones paying for it. This dataset only represents 60,000 respondents from the U.S., while we currently have over 300 million living here. This is a really small portion and is not as significant. Knowing this, with only 60,000 responses, it is hard to prove the valid results. Increasing the number of the responses will be significant in increasing the validity of the results as it gives more people an opportunity to state their opinion. Lastly, the model represents one outlier as it is located outside of the U.S.. I feel that this could be interpreted as someone who visited the U.S. during this time and filled out the questionnaire, but was originally from outside of the country. Additionally, the zip code may be similar to a different part of the world and the database immediately recognized that and not the one in U.S.. In turn, the models and their corresponding descriptions prove the importance of the variety of respondents' location on the American life of student loans and enrollment in school.

Appendix

Figure 1: Respondents' Location vs Educational Loan

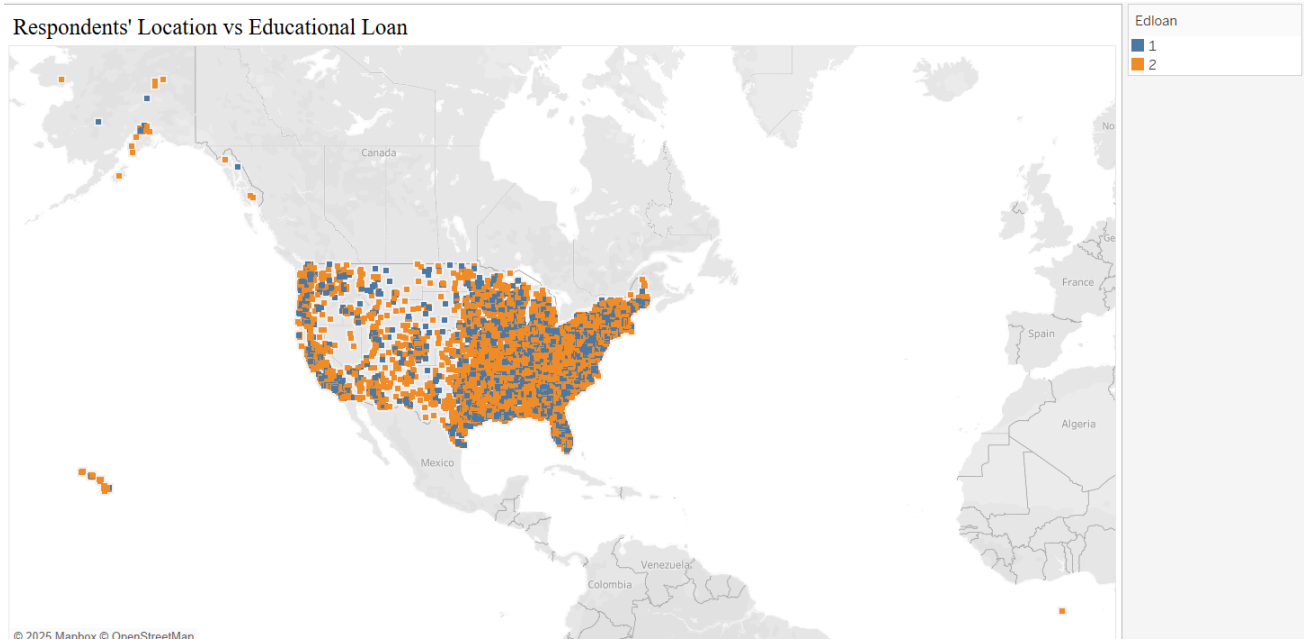


Figure 2: Respondents' Location vs Status as a Student

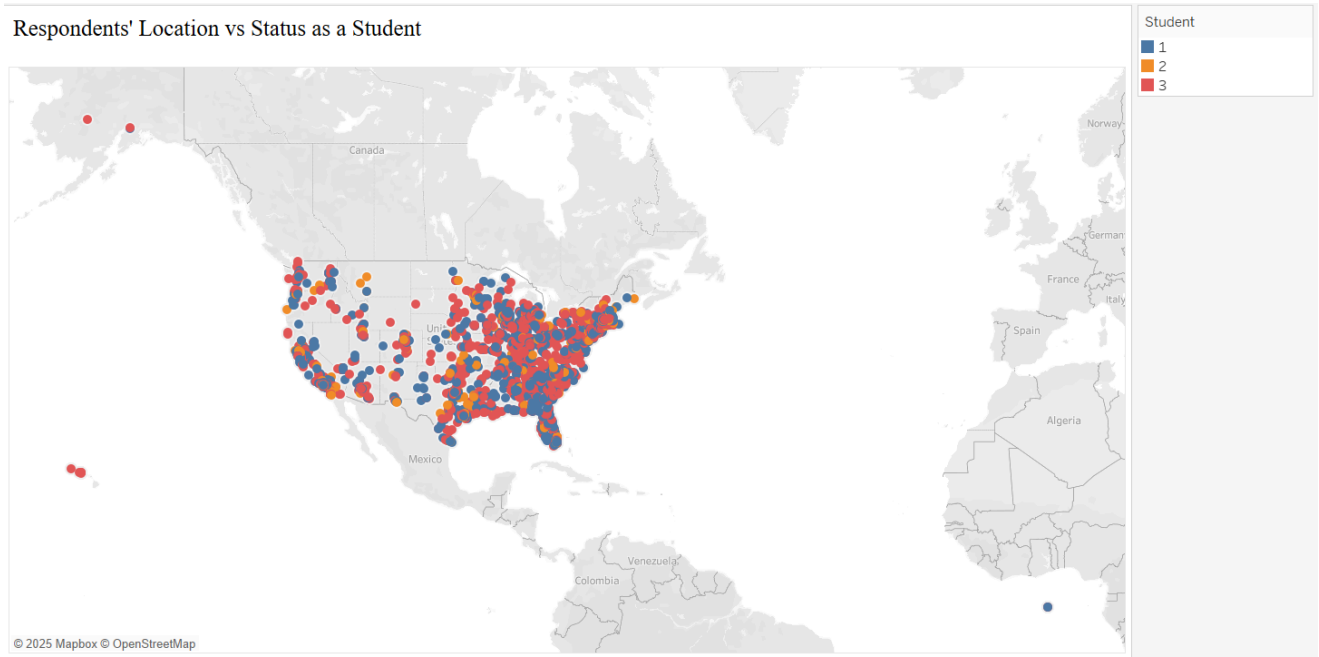


Figure 3: Respondents' Location vs. Educational Loans & Status of Students