CS 257: Database Project Proposal

Yemi Shin, Rudra Subramanian, Rebecca Fox, Gracie Little

April 28, 2020

Project Proposal:

Working Title: Spring 2020 U.S Covid Cases (s20USCC)

General Site Summary:

This site allows users to browse through COVID data by active, confirmed, death, recovered cases, as well as by state and per capita in order to study trends in the spread and course of the disease in the United States.

Overview of Data:

External source: (https://nssac.bii.virginia.edu/covid-19/dashboard/)

Authors: University of Virginia Biocomplexity Institute and Initiative;
Network Systems Science and Advanced Computing (NSSAC)

Dataset summary:
Data shows active, confirmed, and recovered cases of COVID-19 as well as deaths due to COVID-19 all around the world. Data by states/regions in multiple countries including U.S.A, Italy, and India. Data spans from jan 22nd 2020 to april 28th 2020 (as of april 28th 2020) and contains the last update date and time of each country. The data is split into separate CSV files by date.

Definitions of data categorizations, as stated by NSSAC COVID-19 Surveillance Dashboard:

- "Confirmed": Number of positive cases as identified through COVID-19 testing (includes presumptive positive cases)
- "Deaths": Number of people who have died from COVID-19
- "Recovered": Number of people who have recovered from COVID 19
- "Active": Number of people currently infected with COVID-19 (confirmed - (deaths + recovered)

Terms of Use:

UVA provides this data for academic or research purposes. Data is open for the public to access and download provided that the data is cited with a link to the dashboard, users respect that the data is provided under a CC BY-NC license and that the data will be used in a non-commercial product that the public is free to use.

Audience:

A general overview of key users of s20USCC . Personas and additional user stories can be found in the attached appendices.

Potential Users:

1. *People who are closely affected by COVID*

This could include loved ones or family members, who may not have lots of experience working with statistical models but want some context for their experiences. With this site, these users might expect to:

- *View their region and understand the numbers of cases that are presented*, potentially for their own personal understanding of the bigger pictures
- *Discern what overall trends in cases and recoveries are without strong statistical backgrounds*, again to inform their own understandings or to communicate the larger situation to others.

  "As the spouse of someone in the hospital due to a positive Covid diagnosis, I want to be able to show my kids the extent of coronavirus in our state to help them understand the larger situation and why staying home from school helps other families stay out of the hospital."

2. *Medical staff putting together future treatment plans and preparing for spikes in cases*

This could include anyone involved in the operations of health care centers, working to either respond to existing cases or prepare for future increases. Given the context, understanding of the disease and its history and trajectory on national and regional levels could be useful within the medical staff to best inform responses. With this site, medical personnel might be able to:

- *Quickly discern trends on a state level*, potentially to compare strategies with other locations experiencing similar sections of the curve or get a sense of the rate at which they might expect cases to increase

- *Determine regional areas most strongly affected*, in the case that medical workers are being furloughed due to drops in clinic/office/hospital visits but who would be willing to volunteer/find work in a state with a more urgent situation.

  "As a nurse in a clinic which has essentially emptied out since the cease of elective surgery and primary care visits, I want to know if staying with my sister's family on the east coast and volunteering or getting hired at a hospital out there might be a better use of my time if they have more per capita cases."

3. *Social science academics or people involved in education*

This could include users from college professors to elementary school teachers trying to give lessons on basic graphs. Educators are focused on conveying information to their students, who could have any range of experiences with data modeling and statistics. These users might want to:

- *Compare per capita rates between states*, potentially to look at regional patterns in health care practices or disparities in public health infrastructure
- *View simple graphs that communicate the basic trends of the disease on national and state levels*, perhaps to talk about exponential growth or demonstrate a real-life example of how people use graphs.

  "As a biology teacher, I want to teach my students about exponential population growth using real-world examples, like the spread of the coronavirus in the United States."

4. *Public officials and non-profit workers*

These users include politicians, for-profit and nonprofit organizations, and community groups who have to respond to the social fallouts of the pandemic. These users again may not have a strong background in data science or complicated statistical analyses but need access to accurate information in order to coordinate government/ community responses. These users might:

- *Examine recovery rates*, potentially to inform future decisions about the safety of certain re-opening decisions
- *Examine the increases in active cases by region*, perhaps to try to prepare for future outbreaks and allocate resources accordingly
- *Look at outbreaks in neighboring states*, in case there are certain inter-state activities that might have to be adapted to stop spreads across state lines.

"As an officer of a national non-profit, I know which states are being hit the hardest, and use this information to help guide how we allocate resources to local chapters of our organization"

User Stories and Acceptance Criteria:

- As a *social scientist*, I will be able to view data per capita for each state in any of the four categories.
    - Test with each of the categories
    - Test with different states
- As a *public official*, I will be able to specify a region (state level) for which I want to view data, to see which places are hit the hardest.
    - Test with each continent (error checking for others than North America)
    - Test with different states
    - Test with all different categories.
- As a *medical professional* that's preparing for future spikes, I will be able to view COVID data over time.
    - Test with current day
    - Test with past days
    - Test with days before the outbreak in the United States (error handling and robustness of site)
- As a *person concerned about the rate of COVID*, I will be able to compare the different data sets (recovered/active etc.)
    - Test with all the datasets at once
    - Test with different combinations.
- As an *academic*, I will be able to easily download the data in preferred format to use the data in my classes.
    - Convert data (or have all data types) from csv to xlsx and other standard formats
    - Have download of data easily available
    - (Choose which specific data sets they want to download?)
- As a *student* who needs to cite my sources, I should be able to check the source of the data
    - Put source of data on homepage
    - Add disclaimer saying the data may not be 100% accurate

General Plan for Team Collaboration:

We are planning to have regular check-in meetings once a week, which we will use to allocate tasks and divide work, as well as discuss progress, speedbumps, project direction, etc... We will also use Slack to communicate outside of those meetings to plan and help each other with tasks as needed. We are using Asana's Kan Ban board function as a tool to assign and keep track of all of our tasks throughout the course of the project.

Appendix I: Personas (2)

Personas can also be found in the Personas/directory of our team repository

Persona 1: Fiona Smart, caretaker

Name: Fiona Smart

Descriptor: A woman working from her parent's house and caring for her mother who has contracted COVID-19 and is currently hospitalized.

Who is it? Fiona is an only child with two aging parents, and she returned from Fargo, ND where she attends university to her parents' house in Ann Arbor, MI when quarantine started. Her mother is in good health and doesn't present any associated underlying conditions, however, she is over 60 and may have to be put on a ventilator soon, something her care team is considering carefully, given the level of intrusion of the procedure.

What attitudes? She wants to know what the possible chances there are for her mom to survive and is mainly interested in recovery rates. She sees the site as a clear, easily-usable site for people without a strong background in statistics or data science who want more information about the COVID-19 situation in the United States.

Which behavior? Fiona has been feeling overwhelmed by the amount of data and data visualization models that have come up since the beginning of the pandemic. She's most interested in data that would be relevant to her mom's case and is trying to stay optimistic even though her mom's case is touch and go. Because she's trying to stay positive, Fiona is not only interested in case rates in Michigan and North Dakota (which would also affect her being able to return to school) but also in instances of recovery specifically. In efforts to feel productive, she makes hand-made face-covering masks and would like to donate them

to local hospitals and therefore is also checking nearby hospitalization rates to see which hospitals might benefit the most.

Persona 2: Dr. Sarah Washington, an epidemiologist at the MN Department of Health.

Name: Sarah Washington

Descriptor: Dr. Sarah Washington, MD, MPH is an epidemiologist who has worked with the MN Dept. of Health for almost ten years. She is working remotely and living in Minneapolis, and her work is primarily focused on supporting rural areas in the state.

What attitudes? Even though metropolitan centers like New York City, New Orleans, and other big cities were hit hard and quickly early on in the virus, recent evidence suggests rising per capita rates among rural populations across the country. Given lower access to ventilators, hospital beds, as well as potentially less motivation to take preventive measures due to lower initial incident rates, when the virus does reach rural counties they are more likely to be hit harder from a per-capita standpoint. Therefore Dr. Washington is interested in looking at data from other states with largely rural populations. Although Dr. Washington uses far more sophisticated models in her work, a clear interface that feels intuitive for users who don't normally work with data or computer modelling is useful when communicating with rural health centers working on community outreach.

What behavior? Again because Dr. Washington is very literate in the large number of models circulating around Covid-19, she is familiar with the numbers and the trends generally. However, she wants a site that is easily accessible by others who are not as familiar with the data. Therefore, in testing out this site she is looking for a clean and simple interface that presents the data in a way that is understandable but still accurate, and provides any necessary background information without oversimplifying numbers, trends or condescending to users. Additionally, she would like to break down the numbers by state, in order to see the trajectories of other states with large rural populations in order to examine the relative effectiveness of various public health measures.

Appendix II: Summary of User Roles and User Stories

User roles can also be found in UserRoles/directory in our team repository

User Stories

- "As the spouse of someone in the hospital due to a positive Covid diagnosis, I want to be able to show my kids the extent of coronavirus in our state to help them

understand the larger situation and why staying home from school helps other families stay out of the hospital."

- "As a nurse in a clinic which has essentially emptied out since the cease of elective surgery and primary care visits, I want to know if staying with my sister's family on the east coast and volunteering or getting hired at a hospital out there might be a better use of my time if they have more per capita cases."
- "As a biology teacher, I want to teach my students about exponential population growth using real-world examples, like the spread of the coronavirus in the United States."
- "As an officer of a national non-profit, I know which states are being hit the hardest, and use this information to help guide how we allocate resources to local chapters of our organization"
- As a social scientist, I will be able to view data per capita for each state in any of the four categories.
- As a public official, I will be able to specify a region (state level) for which I want to view data, to see which places are hit the hardest.
- As a medical professional that's preparing for future spikes, I will be able to view COVID data over time.
- As a person concerned about the rate of COVID, I will be able to compare the different data sets (recovered/active etc.)
- As an academic, I will be able to easily download the data in the preferred format to use the data in my classes.
- As a student who needs to cite my sources, I should be able to check the source of the data