States of Care: Investigating U.S. Maternal Mortality

I. Data Documentation

There are two major sources of data utilized for this project. The first comes from the CDC WONDER Database, and the second is a collection of data from the Kaiser Family Foundation (KFF) on state profiles for women's health.

CDC WONDER (CDC WONDER)

Data related to maternal mortality was not difficult to locate. However, finding data that was aggregated correctly, and consistent throughout multiple years, was difficult. At the state-level, the CDC data allowed us to look at the interaction between maternal mortality rate, race, and education, but many rows had missing values, so we decided it would be best to create a predictive model based on regional statistics. For this data, we filtered on ten-year-age groups, race, and level of education, and obtained percent of population death for these groups. Someone picking up this project for the first time needs to understand the lack of data availability for certain racial groups, regions, education level, etc. This makes it harder for our model to have a high level of accuracy.

This data informed the logistic regression model. In order to run this regression, in the data we added a binary column for the maternal mortality rate, with rates below 1% being 0 and rates above 1% being 1. This threshold becomes important in our predictive model, because it's the threshold we use to determine the probability of someone's maternal mortality rate being high. The 'low' maternal mortality rate ranges from 0% to 1%, and a 'high' maternal mortality rate ranges from 1% to 5.5%.

Kaiser Family Foundation (U.S. Maternal & Infant Health Data)

Relevant data is scraped from the *Maternal and Infant Health*, *Abortion Policies*, *Demographics*, and *Coverage* sections of the webpage. A challenge with this collection of data, and something someone picking up this project should understand, is the use of different time ranges. For some statistics, there is one year used, and those years may differ from category to category, while for other statistics, multiple years of data are aggregated and used. This inconsistency causes challenges with interpretation and comparison due to the differing time scales, which makes explanation of visualizations and analyses outputs not as straightforward. However, we account for this with clear descriptions of what the output analyses and visualizations convey.

This data is used for the map and scatter plot visualizations, where we needed to identify the variables of interest, and merge the data columns based on the key "state" to enable comparison of maternal mortality by these variables of interest (e.g., percent uninsured, women's weekly earnings).

II. Project Structure

Purpose	Functions		Files
Data Collection	Storage for keeping uncleaned data and cleaned data	- Uncleaned Data	data/scrape_data data/downloaded_data
		- Clean Data	data/
Data Preprocessing	Create web scraping script for obtaining data from KFF		mortality/scrapers
Data Cleaning	Download regional data from the CDC WONDER Database page and clean		mortality/regional_clean.py
	Use KFF scraped files, and merge them using "state" as the unique identifier.		mortality/kff_merger.py
Map Visualization	Create the state-level map visualization and scatter plot		mortality/map_viz.py
	Create Dash that allows interaction with maps, scatter plot and data exploration visualization components		mortality/map_viz.py
Data Analysis	Create predictive logistic model		mortality/predicitve_model.py
	Create Dash that allows interaction with model and data exploration visualization components		
Program Execution	Combine the Data Analysis and Map Visualization to the program on command line for user-friendly		mortality/__main__.py
Testing	Test scripts for testing the functionality of the program		test/
Archive	Placeholder for unused scripts and data		archive/

III. Team Responsibilities

Team Member	Model	Task	Files

Alexandrea Harriott	Data Collection	- Explored the KFF data on state health profiles and identified the variables of interest along with team
		- Scraped the KFF web pages
		  - Investigated page sources and network tabs to find the data
		  - Wrote scraping files
		- Helped in framing the regression model and interpreting results
		- User interface design by making
			(*all files in*) mortality/scrapers
			
			(*all files in*) data/scrape_data
			
			tests/test_web_scraping.py

		aesthetic modifications	
Dorothy Wongkarnta	Interactive Predictive Model	- Wrote the logistic regression model - Create the Dash that contain interactive model and data visualization containing - Create test functions for predictive model - Create the user-friendly command lines format	mortality/predicitve_model.py tests/tests_predictive_model.py mortality/__main__.py
Elizabeth Ronan	Interactive Map Visualization	- Imported KFF abortion legislation data and merged dataset, assisted with cleaning of data for visualization - Created data dashboard consisting of maps and tables for each dataset with interactivity to switch between views - Created scatter plot with merged dataset and the ability to change the x-axis to display different characteristics - Created command line interface for map - Collaborated on overall structure of project, finding data sources, design of dashboards	mortality/map_viz.py mortality/__main__.py mortality/kff_merger.py
Madelin De Jesus Martinez	Data Collection and Data Cleaning	- Gathered and analyzed different types of data sets, to see which fit our predictive model - Converted and cleaned regional dataset - Merged/cleaned KFF data - Created test function for the regional and KFF data - User interface design by making aesthetic/interpretive modifications	tests/tests_kff_merger.py test/test_regional_clean.py mortality/regional_clean.py mortality/kff_merger.py mortality/predicitve_model.py mortality/map_viz.py

IV. Final Thoughts

In today's changing political landscape, reproductive justice is more important than ever. Through our goal of exploring the implications of policy changes on women's health, we see that women are not affected equally by these issues: the intersections of location, race, education, age, access to insurance, poverty, and more determine health outcomes. Despite challenges with the data due to scarcity and inconsistency across sources, the analyses conducted through this project still allowed for these key insights to be conveyed. This analysis will continue to be of importance as reproductive health laws change across our country and women's access to lifesaving healthcare is limited. We hope to continue this analysis beyond the scope of this course.

V. Appendix - Project Structure Diagram