Visualizing Birth Rates in Wisconsin

Group 8: Julia Paciorek, Nathan Arbogash, Yalei Zhao, Christopher Gottwaldt, Maxwell Sonnenburg, Keval Patel

Link to Shiny App

Box Links to Birth Data, Geojson Data, and Statistics Data

GitHub Link to Code

[Introduction]

We would like to analyze how various demographic factors are correlated with 2022 birth rates in Wisconsin. We decided to implement a few different types of visualizations for our purpose of bridging the demographic inequalities associated with birth rate statistics. One of our main findings upon conducting the literature review was the usefulness of interactive visualizations for illustrating findings to the reader. This is why each of our visualizations has some level of interactivity, whether it's as simple as hovering over a row in a table or more complex, such as re-rendering a graph depending on the user's selection. We also tried to focus on the differences between counties as opposed to just merely analyzing demographic information about the mothers to make the analysis more well-rounded in the context of representing Wisconsin as a whole. We believe that our visual representations of the data will enable others, such as researchers or those in higher education, to do further analysis to determine what is causing county-wide disparities (e.g.: access to education, healthcare, income, etc.) to help highlight and address these inequalities.

[Literature Review]

In our literature review, we found a few useful sources to aid in making our visualization more effective.

1. **Heatmap Perception Study:**
    a. The paper explores factors that enhance the perception of heatmaps, focusing on how visual design elements help cognitive processing. The article identifies the most effective strategies to be:
        i. Color Palette: using color schemes that provide clear distinction between values without causing misinterpretation of gradients between colors, basically trying to avoid any visual strain on the user (Middel Soria, C., 2022).
        ii. Dimensionality and Cell Size: emphasizing the importance of balancing the heatmap's dimensions and cell sizes to ensure that individual data points are discernible and the overall pattern is recognizable without overwhelming the user (Middel Soria, C., 2022).
        iii. Interactivity: If the data is complex, adding zooming in and filtering features to the visualization will greatly improve the perceptiveness of the heatmap (Middel Soria, C., 2022).

iv. Legends: Including legends, and descriptive titles, and practicing good naming conventions all add to the readability of the heat map (Middel Soria, C., 2022).

In this Heatmap Perception Study, they mention the main aspects of a heatmap that help readers understand the graph, including a clearly differentiated color palette, large enough cell size, and interactivity (Middel Soria, C., 2022). We made sure to strike a balance with these aspects in our design, making the map interactive with clearly differentiated colors and large county borders representing the graph cells.

2. **Association between socioeconomic status and fertility among adolescents aged 15 to 19: an analysis of the 2013/2014 Zambia Demographic Health Survey (ZDHS):**
    a. This paper analyzes the reproduction rates of adolescents based on their socioeconomic status - primarily education level and family income.
    b. The paper found that higher family income was negatively correlated with birth rates, and education level was also negatively correlated with birth rates. Meaning more education and higher income tend to reduce the amount of kids one has, or reduce the chance of having one at any point (Munakampe, M.N., et al. 2021).
    c. While this paper does not include any visualizations, it does detail various regressions, and how they compared regressions to each other, comparing effectiveness and reliability.

This paper helped us understand what we might expect to find in our Wisconsin dataset. I think it was fairly evident that using visualizations to help understand the data would be helpful for future studies like this, so we were hoping to provide that visual component to be used by researchers in the future. They also showed that there was a strong correlation between education and birth rates, so we made sure to include education as a key variable in our analysis.

3. **MatrixQCvis: shiny-based interactive data quality exploration for omics data**
    a. This paper shows an example of a Shiny-based application and makes an argument for why it is useful to create interactive visualizations: creating a UI of visualizations enables the visualizations to be more tailored to the reader's interests instead of having a long document of individual graphs, a similar point made in the heatmap study (Naake, T., et al., 2021).

Although this paper explored a different dataset and research topic, we used their methodology for advice on producing a higher-quality visualization that helps convey to the reader information visually.

[Design]

In order to address our main goal of shedding light on the demographic inequalities regarding birth rate statistics, we researched effective methods of communicating data to an audience and found that interactive visualizations and heatmaps are very effective and interesting modes of doing so. We created a shiny app that allows users to interact extensively with our data, with a map graph showing the WI counties according to the selected statistic, and multiple tabs displaying more detailed information using tables and bar graphs. There is a lot of information to take into consideration on each pane of the app, so we made sure to make the heatmap palette very clear and distinct in differentiating the different levels in measurements with large cell sizes according to the info in the first study in our literature review (Middel Soria, C., 2022). You can see the map visualization in Figure 1 where upon selecting a county, its information will be displayed in a pop-up in the upper right corner. This map also has options to adjust the zoom and to pan around for additional exploration options.

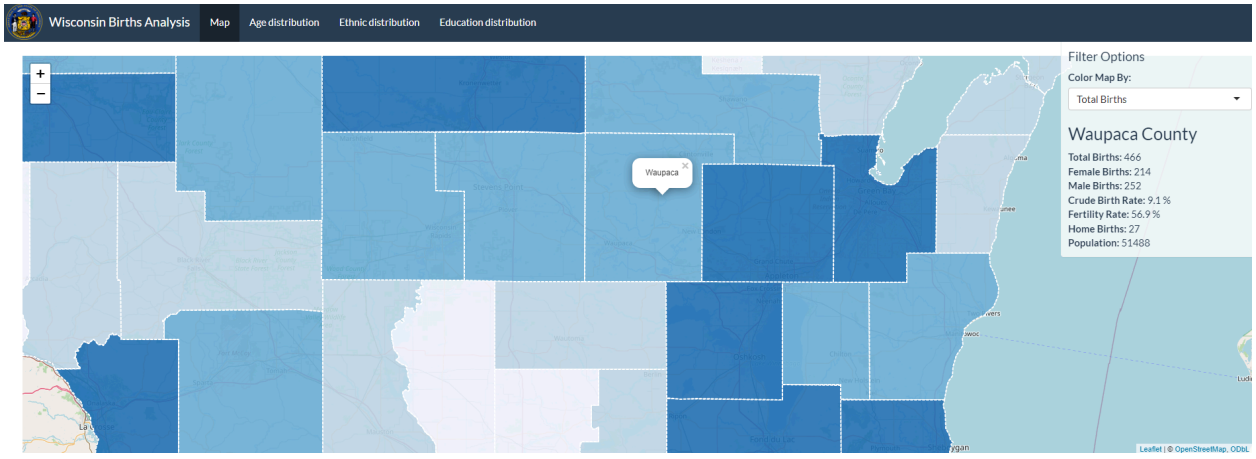

Figure 1

The shiny app also has additional tabs for deeper analysis of specific contributing factors. We have an Age distribution tab, an Ethnic distribution tab, and an Education distribution tab that each illustrate the distribution of the data according to their respective variables. For example, Figure 2 demonstrates the top 10 Wisconsin counties by number of births with a table, and has a bar graph illustrating the distribution of the count of mothers in Dane county across different age groups. For the table, you can choose the age range you'd like to view, and you can select the county you'd like to view for the bar graph distribution.

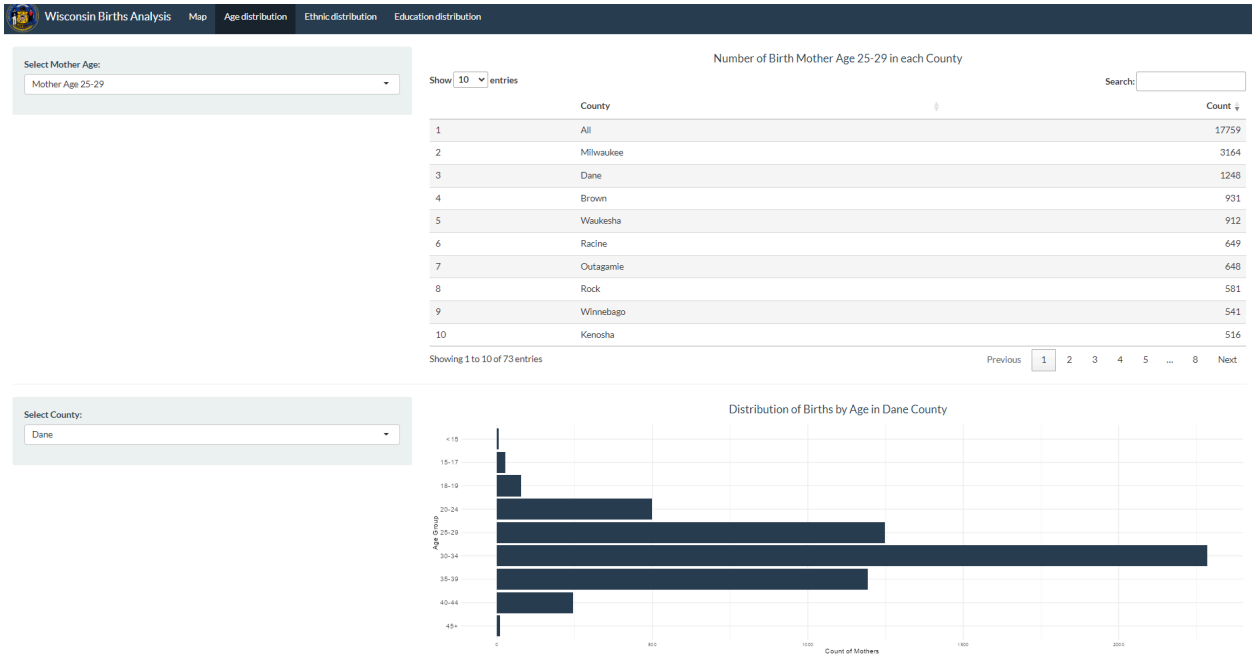

Figure 2

In Figure 3 you can see how most of the births are from White mothers, followed by Hispanic and Black or African American mothers. This was surprising to us because a lot of us weren't originally from Wisconsin and therefore were unaware of the ethnic makeup of the state.

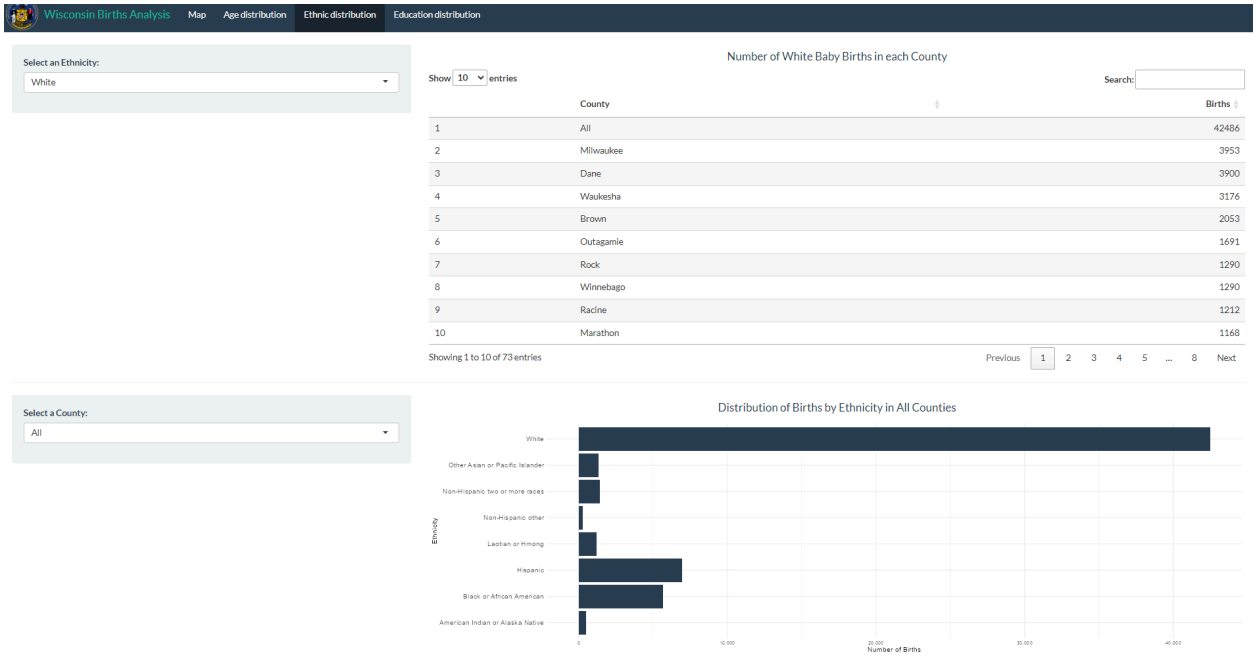

Figure 3

Finally, we had a tab for analyzing how education's impact on birth rates. Figure 4 helps convey that most of the mothers in 2022 were either at a high school or college education

level, with few having large amounts of education (Doctorate or Professional Degree) and also relatively few with low amounts of education (less than a high school diploma). Also, there weren't many with associate degrees or master's degrees, potentially because these degrees are less common than a bachelor's or high school diploma. Of course, you are able to select the county you'd like to view for the bar chart and select the education level of the mother for the table data. As with the others, you can limit the number of entries, search, or sort the data based on birth count or the county name alphabetically.

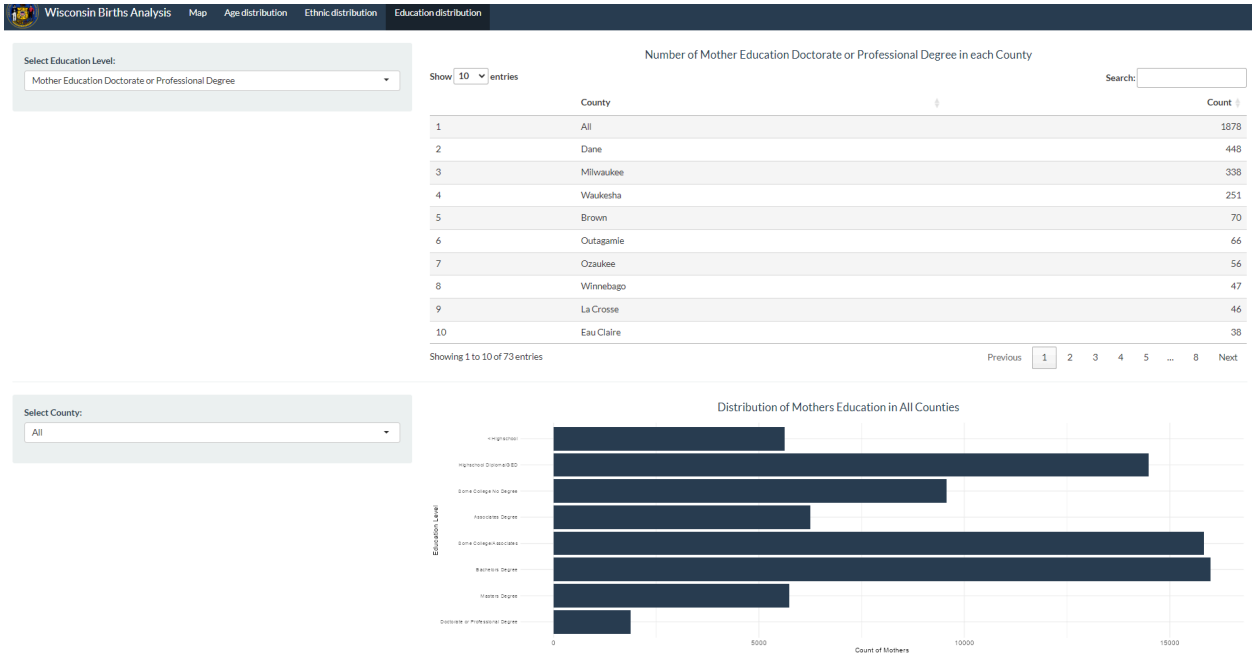

Figure 4

[Synthesis]

A challenging task in analyzing Wisconsin birth rates is trying to retain the geospatial aspect of the data; We were analyzing the differences between counties, so we wanted to illustrate the differences in a clear way with GeoJSON in an interactive thematic heatmap (shown in Figure 1). Another challenging task for this topic is that there are a lot of factors at play in determining birth rates, so we have a separate tab to explore the effects of education, age, and ethnic factors, with tables and charts to help understand their effects. We wanted to keep a somewhat uniformity to our visualizations for each tab where possible to make checking out different tabs more straightforward to the user and allow them to delve deeper into the data than you get with the fun interactive map.

[Conclusion]

There are a multitude of factors at play in the demographic inequalities regarding birth rates. We hope that our app can assist those in higher education and research by enabling them to explore this data visually across our different metrics. This could help researchers identify

areas in which there are high rates of e.g. premature pregnancies or help with determining where to most effectively provide assistance to mothers struggling with the costs of motherhood.

Works Cited

Middel Soria, C. (2022). Heatmap Perception Study. Bachelor's thesis, Universitat Politècnica de Catalunya. https://upcommons.upc.edu/bitstream/handle/2117/367508/164344.pdf?sequence=1&isAllowed=y

Munakampe, M. N., Fwemba, I., Zulu, J. M., et al. (2021). Association between socioeconomic status and fertility among adolescents aged 15 to 19: an analysis of the 2013/2014 Zambia Demographic Health Survey (ZDHS). Reproductive Health, 18(182). https://doi.org/10.1186/s12978-021-01230-8

Naake, T., & Huber, W. (2021). MatrixQCvis: shiny-based interactive data quality exploration for omics data. Bioinformatics, 38(4), 1181–1182. https://doi.org/10.1093/bioinformatics/btab748