Assignment # 2: EDA with Data Studio
Course: CAP 6610 Applied Machine Learning
Professor: Dr. Oge Marques
Student Number: Z23596812
Student Name: Renee Raven
Term: Fall 2022

Part 1: Introduction to Data Studio course

Figure 1. PDF of Data Studio course completion

This brief course offered a nice overview of the interface and a few walk-through exercises to help the user become familiar with the basic workflow. It appears easy to create a basic presentation or report with a template, but Data Studio also allows customization of the visualization.

The “read further” sections provided helpful information on charts, coloring, and themes via links to targeted articles.

- Chart references in the Data Studio Help Center
- Data coloring overview
- Data coloring techniques
- Themes

Part 2: Selection from the Data Studio Gallery

I selected the Fructose Checker from the community section of the Data Studio gallery because it is useful, communicates a clear purpose, has a github (Fig. 3) with

documentation and references, and looks well organized. The initial screen shows the nutrition information for all the foods referenced from the Australian Food Composition Database.

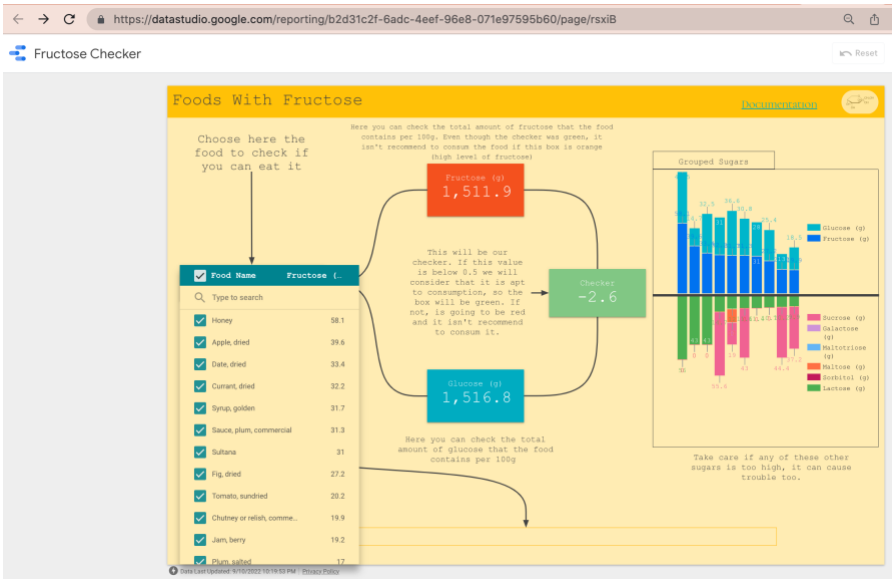

Figure 2. Screen shot of the Fructose Checker Data Studio dashboard

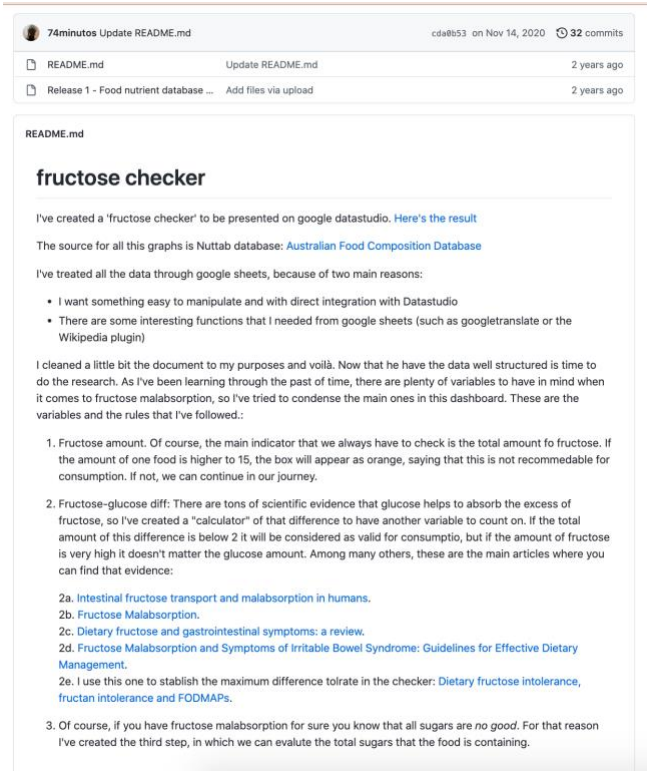

Figure 3. Github site that provides explanation for the Fructose Checker

The interface is fairly intuitive and easy to use. One types a food into the “Food Name” box and click on the “only” if that is the only food you want.

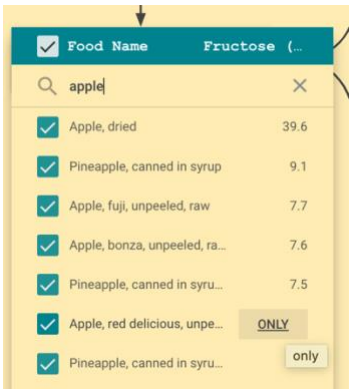

Figure 4. Text entry of "apple"

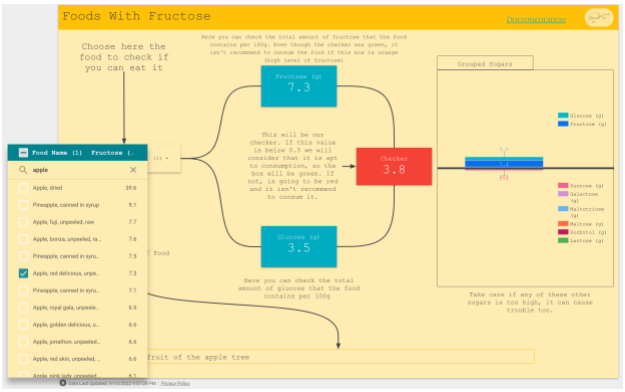

Figure 5. Nutritional results for selected apple, red delicious

However, the product also allows a running nutritional profile when adding multiple items by clicking the check boxes to the left of the items. For instance, a bowl of berries would yield a prohibitive fructose score.

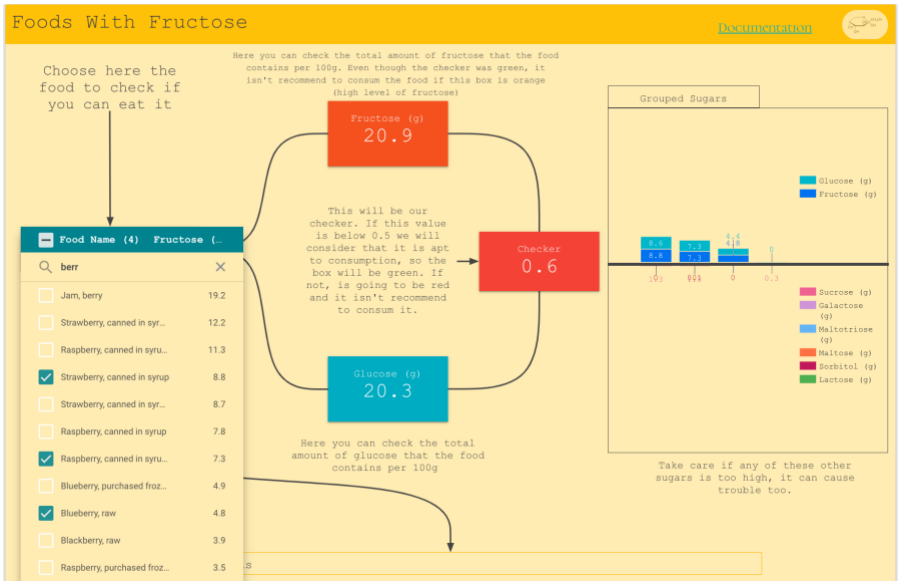

Figure 6. Nutritional profile of combined berries

Improvements for this project might include:

- Adding more items to the database (for instance “cheesecake,” “halibut,” and “mole” were not found)
- Creating a section that lists all the selected items so the user knows what has been added (for example bread, turkey, mayo, lettuce for a sandwich)
- Remove the “Wiki Definition of Food” section. When “date” is selected for food, the section displays “particular day represented within a calendar system.”

Part 3: Creation of new Data Studio report (URL)

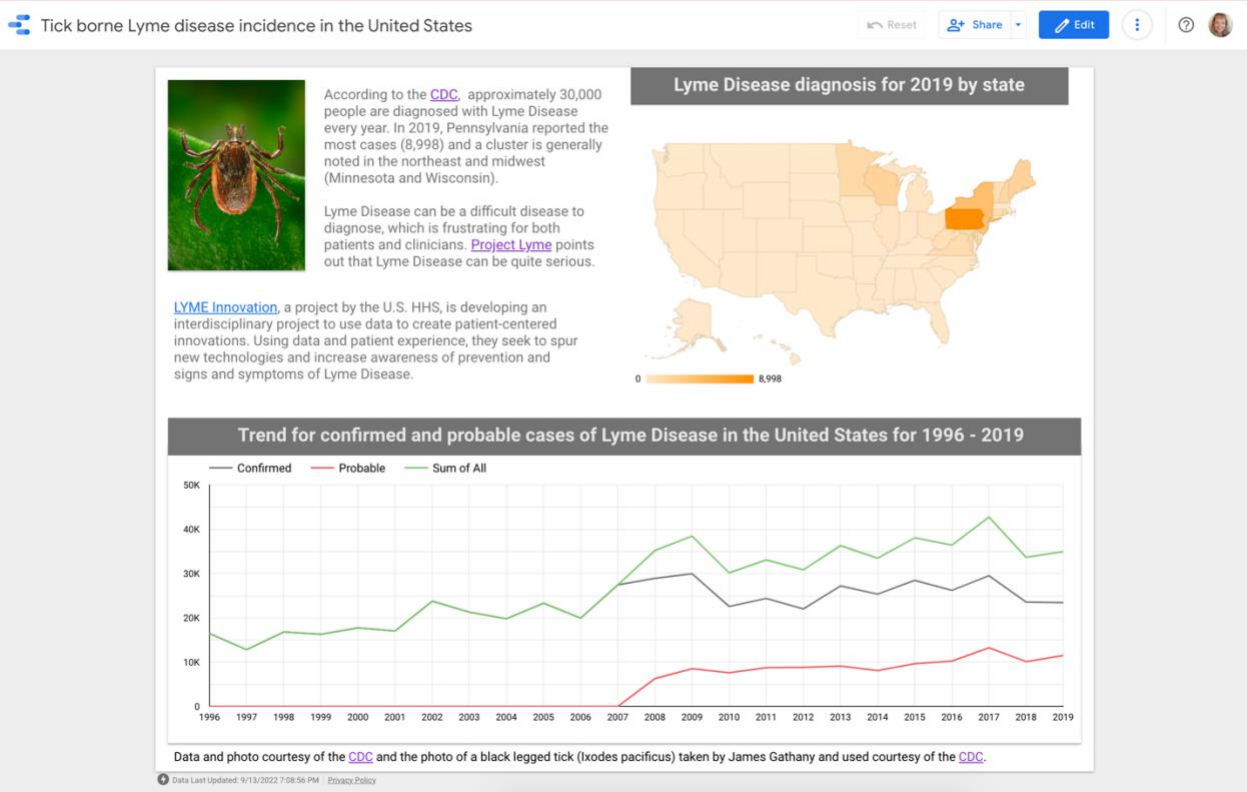

Figure 7. Screen shot of Lyme Disease Report created with Data Studio

I am primarily interested in medicine and healthcare, so I searched for open source files through government agencies. I selected 2 csv files from the CDC that provided information about Lyme Disease between 1996 and 2019. One of the files also focused on reported cases by state, which allowed me to work with the map options in Data Studio.

- Lyme Disease Cases by State or Locality [XLS – 3 KB]
- Lyme Disease Overall Cases by Year, United States [XLS – 416 B]

I decided not to use a template. Instead, I used the snap to grid function and tried a few chart styles.

Data Studio Help proved invaluable, but I also used a youtube video to help with maps. I wanted to limit the map to the United States.

The process involved a lot of trial and error. Loading the csv files was easy, but I did need to change the data types of several files from text to number or text to geo location for the map. After familiarizing myself with the grid, some of the basic chart types, and (most importantly) the data, I was able to create a cohesive visual snap shot of Lyme Disease prevalence in the U.S. over time (1996 to 2019) and focused on locations for 2019.

Key techniques I learned while working with Data Studio:

- How to connect data to individual visualizations (I used one for the map, one for the line graph)
- How to use the grid and the snap to grid option to keep things aligned
- How to import photos
- How to resize and position items
- How to add tooltips
- How to limit the geographical area displayed on a map
- How to work with themes, styles, and text boxes
- How to add hyperlinks

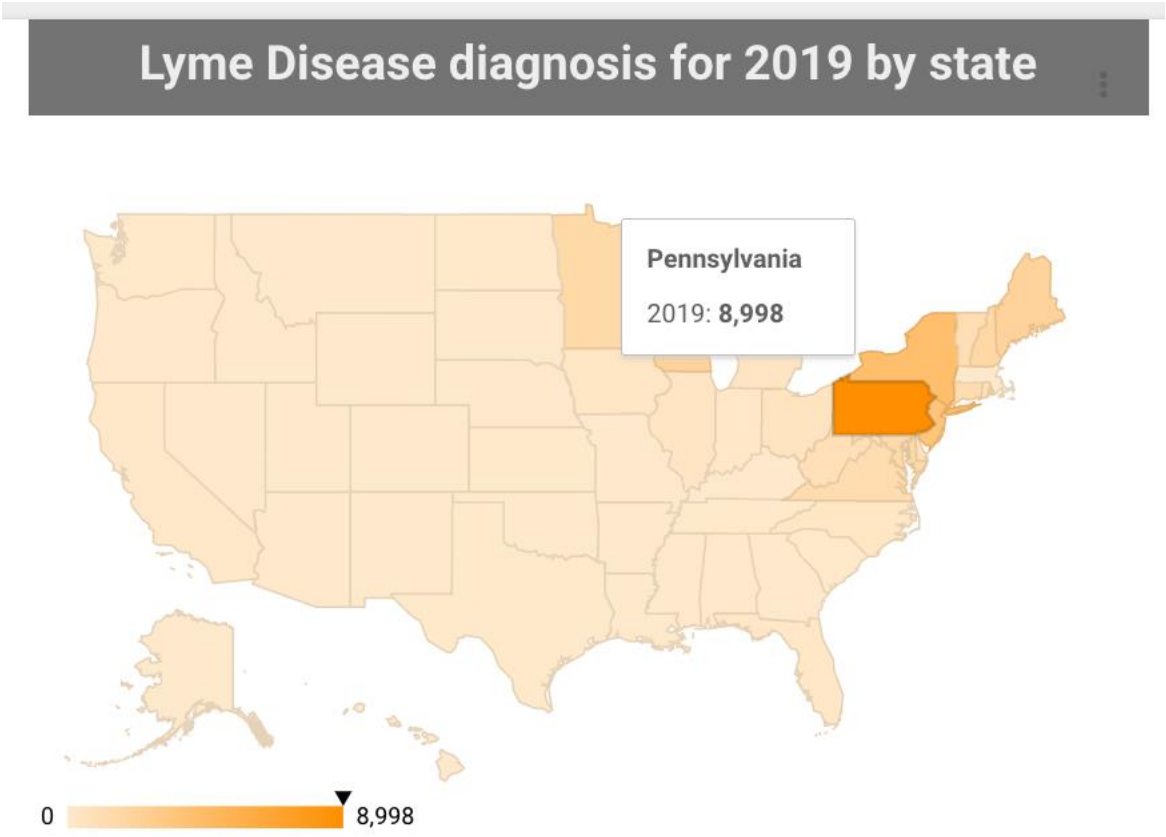

Figure 8. Tooltip with Lyme Disease case count appears with mouse hoover

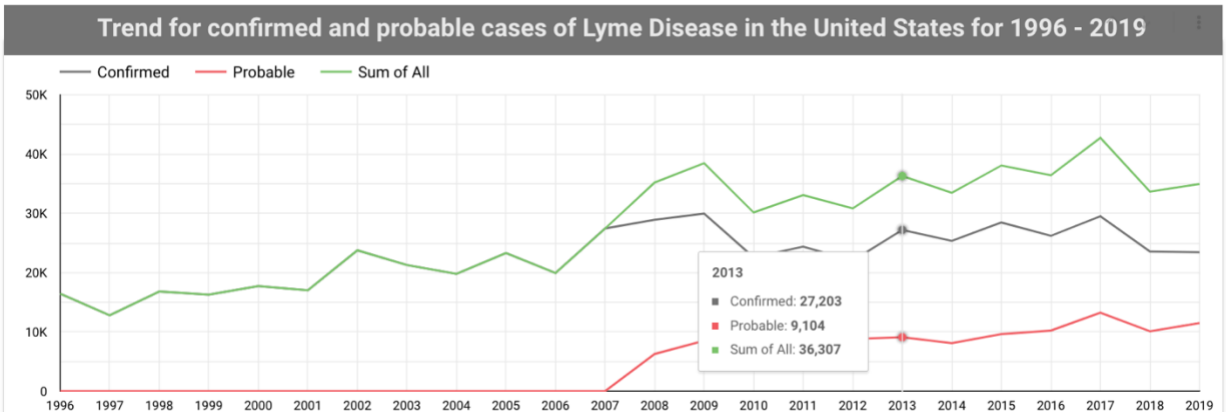

Figure 9. Example of tooltip information

I believe this information would be useful for the average person, my target audience. I wanted the report to convey, in just a few seconds, that there are certain geographical hotspots and that the incidence of cases is increasing. I provided links for further information (CDC, Project Lyme, and LYME innovation).

I would like to research the way Data Studio can blend data sources so one can create a graphic from multiple datasets. If I were to use Data Studio in the future, I would definitely want to watch some tutorials about calculated fields. I attempted to use CASE per the help page, but things didn't work quite the way I expected.

I submitted my site to the "Community" area of the Data Studio. Overall, it was a quick process which required verifying ownership of the data and giving google the ability to edit the report/data.

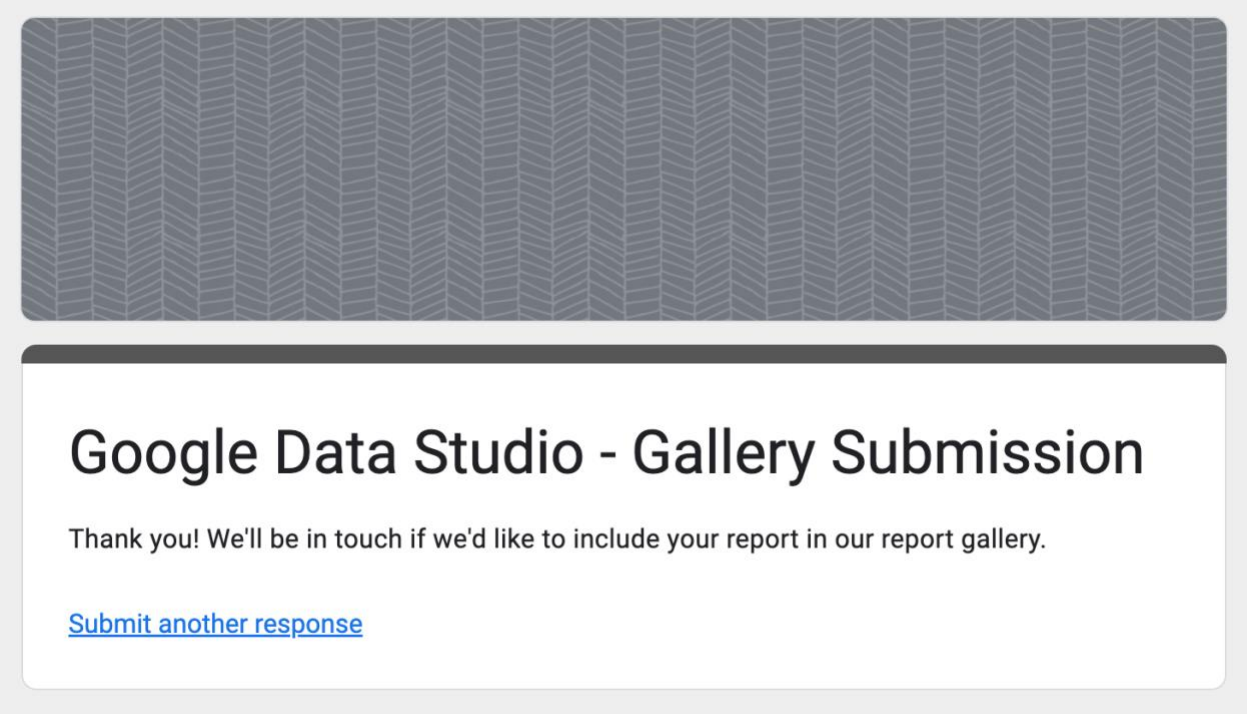

Figure 10. Submission to Data Studio community

Conclusion

Data Studio is ideal if one is working with google ads or youtube analytics. There are many templates and the information updates automatically. I took on the challenge on creating a health issue related report because I am more interested in medicine than business. Still, I think it served my purpose fairly well and it is more intuitive than some of the other dashboard style tools I've tried.