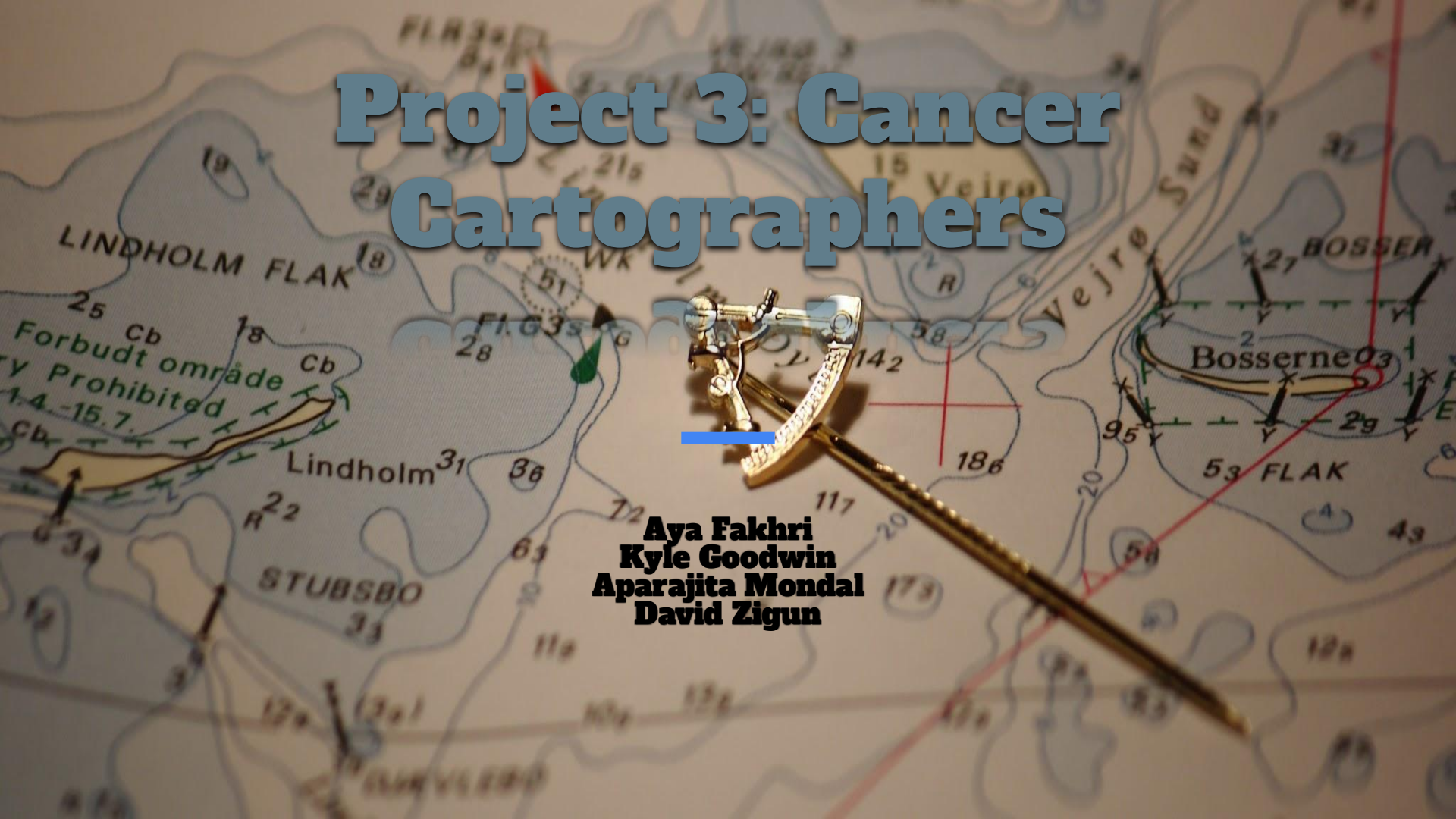
Project 3: Cancer Cartographers



Aya Fakhri
Kyle Goodwin
Aparajita Mondal
David Zigun

Introduction

Initial idea: compare cancer incidence rates by geographic proximity to nuclear power plants***

Goal: Create an interactive map of the US, with toggling layers, all cancer incidence rates by county, breast cancer incidence rates by county, and geographic locations of nuclear power plants.

Data Sources:

- GeoNuclear Data of Nuclear Power Plants (International Data filtered for US locations via Kaggle/World Nuclear Association Reactor DB)
- Cancer rates collected from NIH/National Cancer Institute (2017-2021 averages by county)

NOTE

This project does NOT attempt to prove correlation OR strength of potential correlation between mapped variables

Methods

Cleaning of Cancer Rate CSV files

Creation of SQL Database

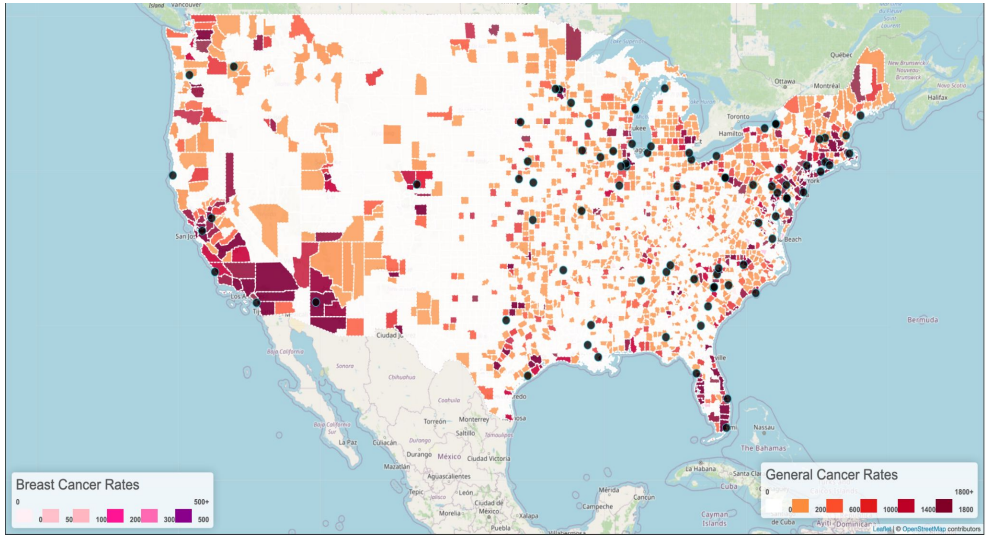
Linked GeoJson to populate map

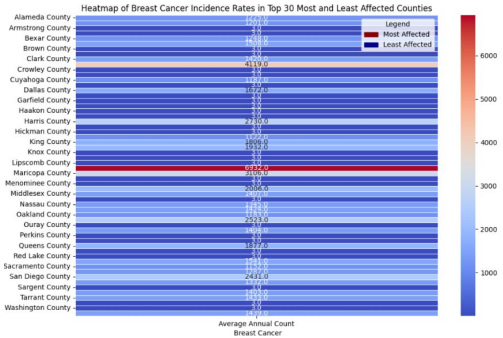
Creation of Leaflet Map

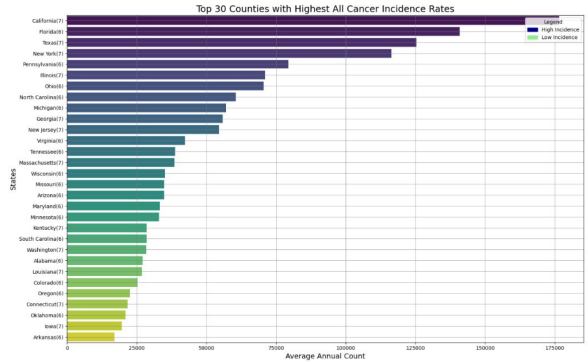
Creation of Incidence Rate Graphs

Population of Leaflet Map

Hosting of Map and Graphs via HTML



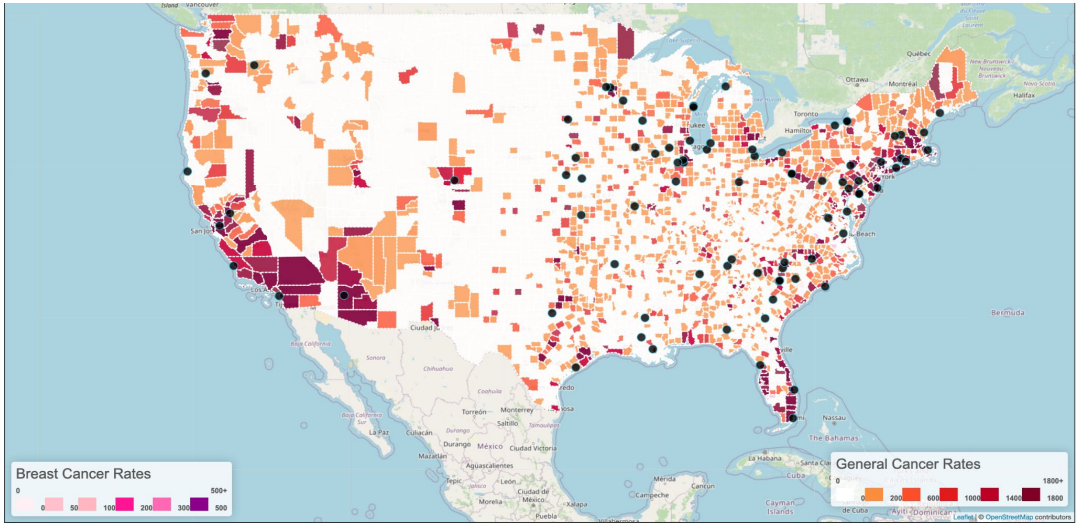


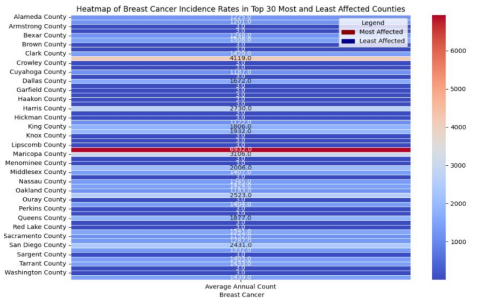


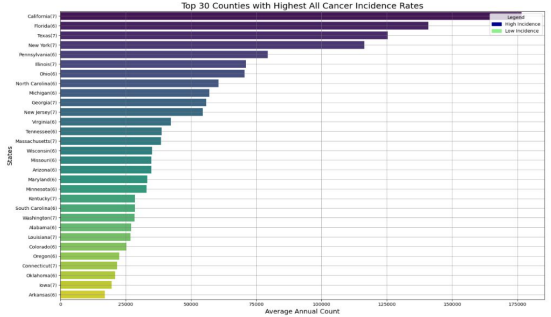
Visualisation for Nuclear plants and cancer Rates in USA 2017-2021

Map showing the cancer rates for all counties with locations of nuclear power plants.

Charts showing the counties with the highest rates of cancer.







All Cancer Data Visual Representation

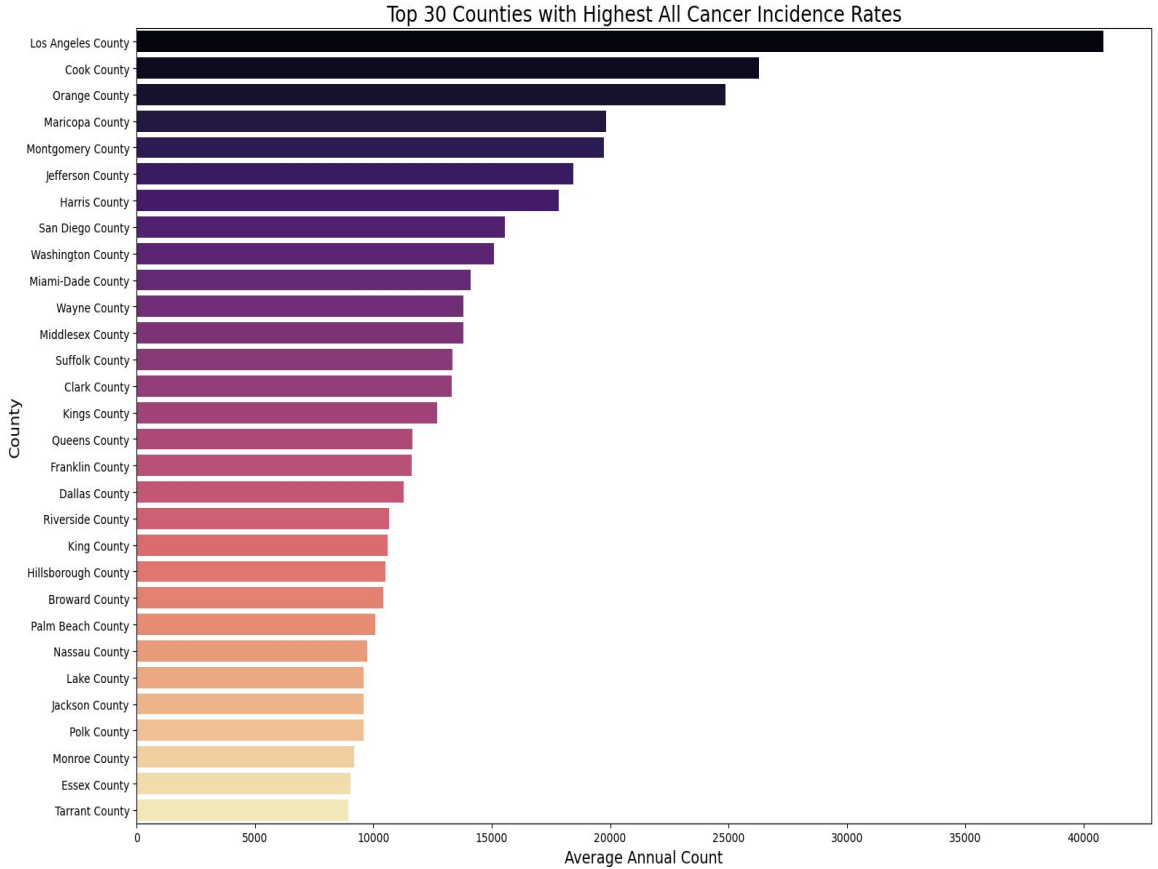
Data Cleaning and Preparation:

- Loaded cancer data using Pandas.
- Cleaned and split data fields (County, State).
- Handled missing values and filtered irrelevant rows.
- Converted necessary fields to numeric format.
- Saved the cleaned data for analysis and visualization.

Visualization: Top 30 Counties with Highest Cancer Rates

•The bar chart displays the 30 counties with the highest average annual cancer counts.

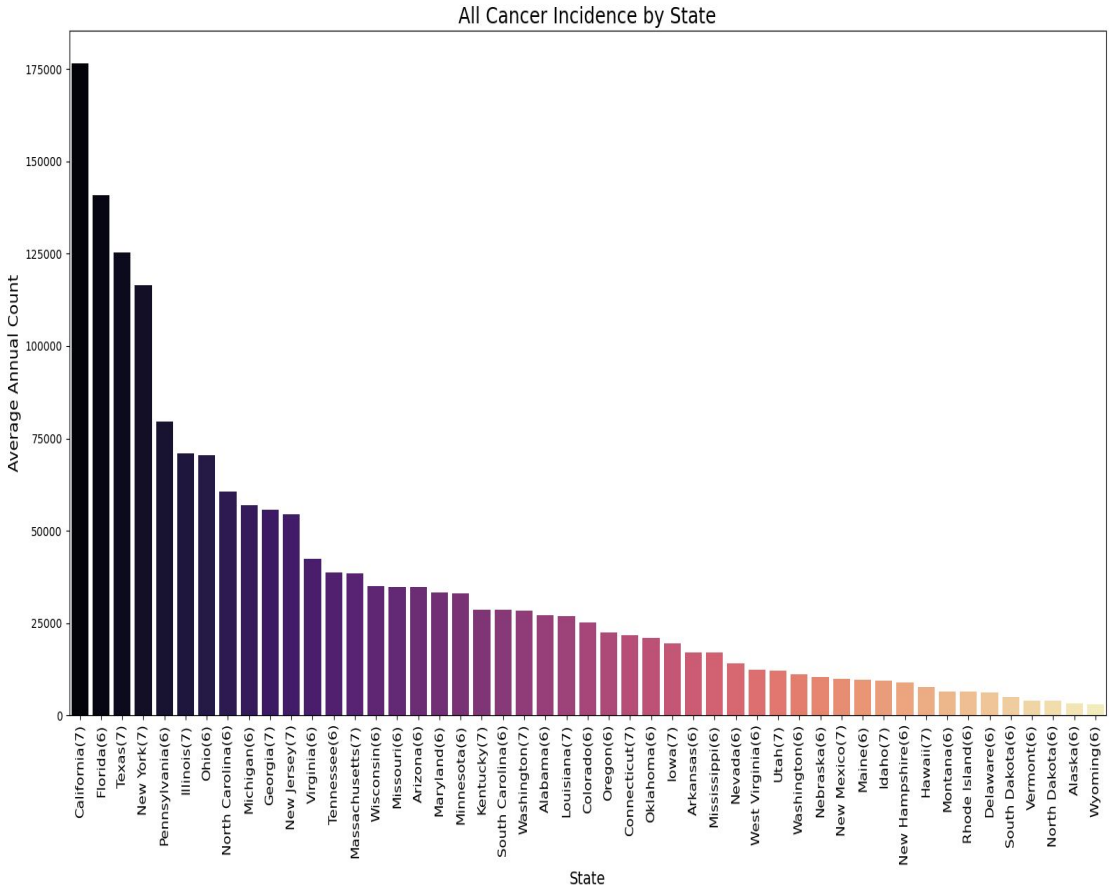
•Key Counties: Los Angeles, Cook, and Orange County rank among the top.



Visualization: All Cancer Incidence by State

•The bar chart shows cancer incidence across U.S. states based on the average annual count.

•California, Florida, and Texas have the highest cancer counts, reflecting their large populations.



The Overall Impact of Nuclear Plants on Cancer Rates

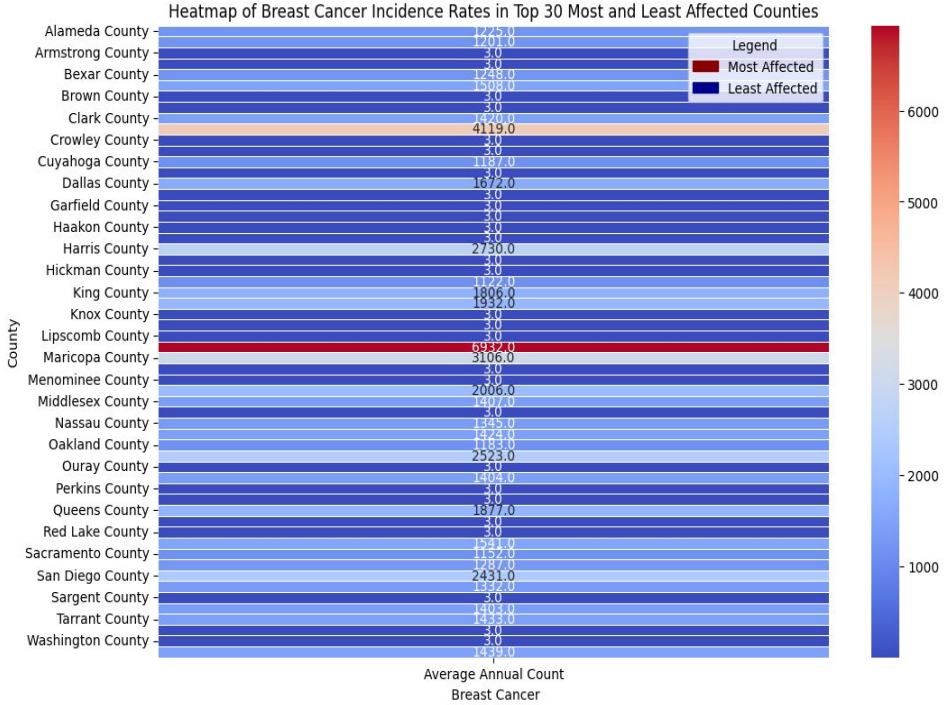
Breast Cancer Awareness

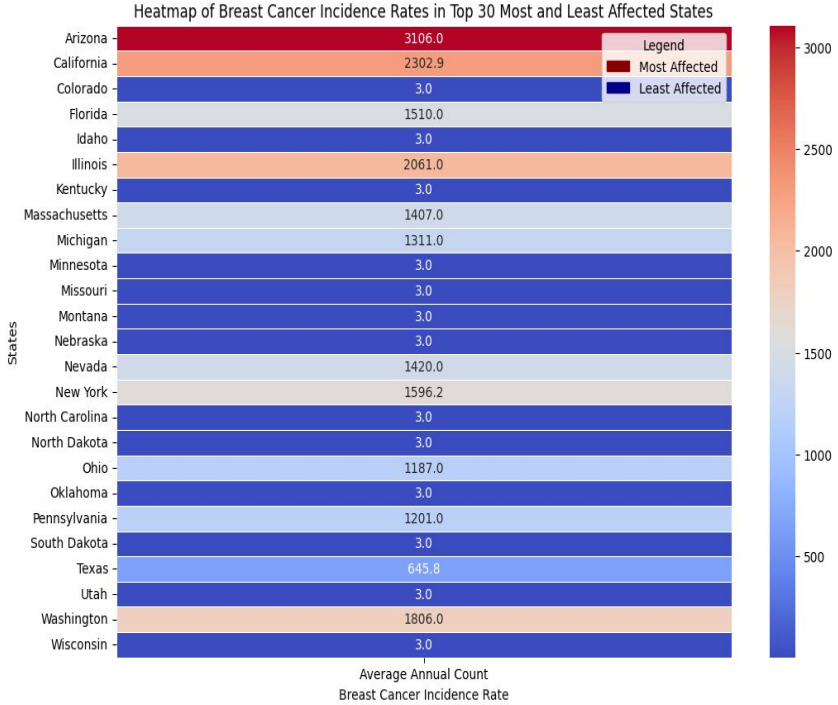
For more information and resources, please visit the link provided here.

Together, let's raise awareness and foster understanding. 💖

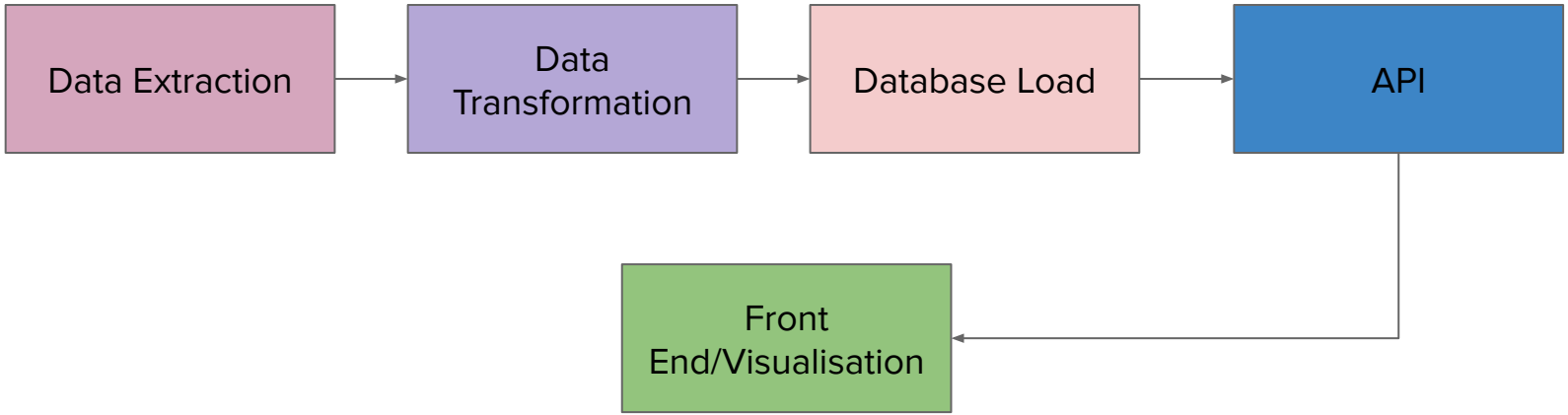
https://www.nationalbreastcancer.org/breast-cancer-awareness-month/

Breast Cancer Data Visual Representation

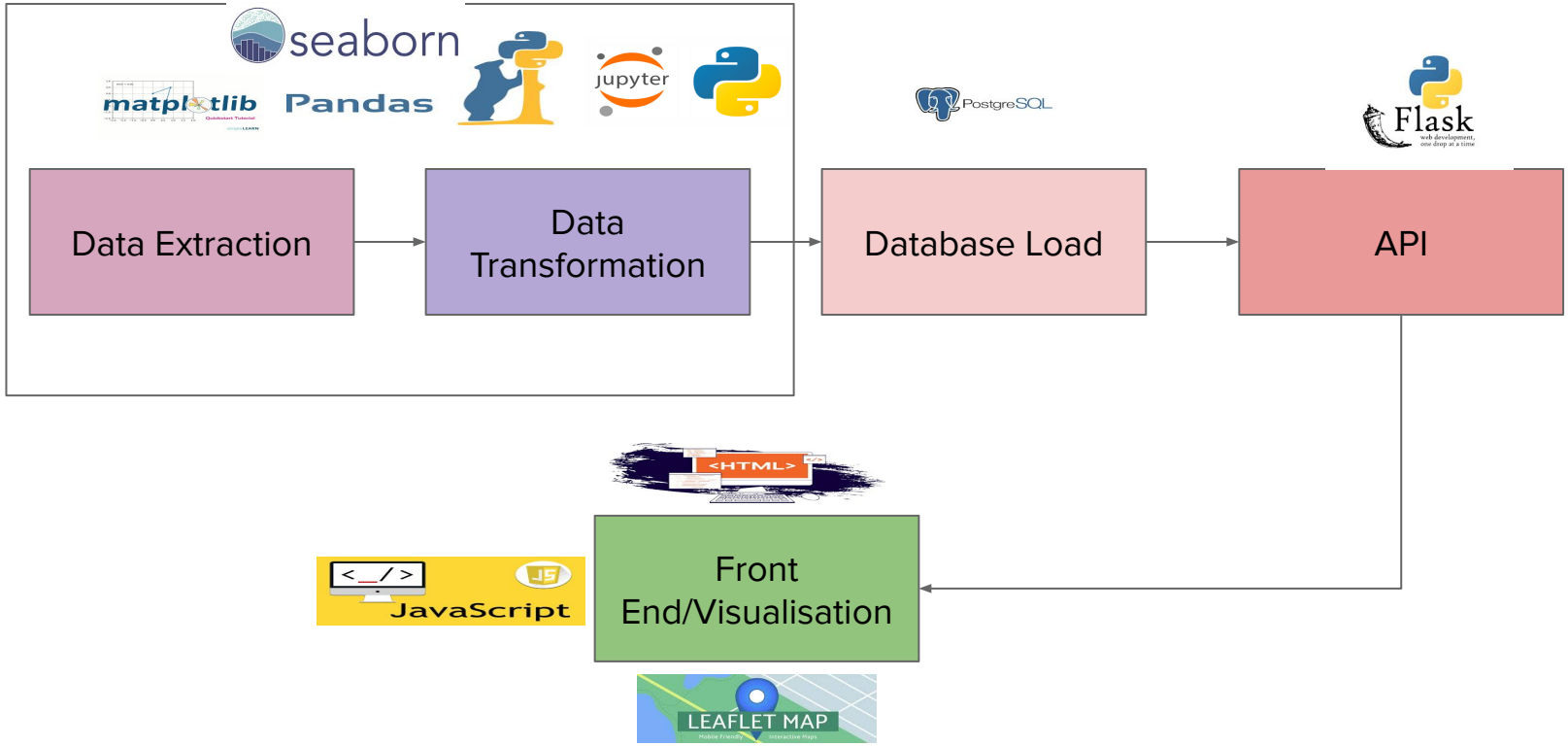




High Level Process Flow/Design



Low Level Process Flow/Design



Challenges

- Understanding the end to end flow
- collaboration outside class as we all are busy
- Working with new library and understanding the syntaxes and debug issues
- Some of the errors we got related to the data
- Matching counties with the GeoJson file

Learnings

- Creating database in sqlite
- Code integration- we pulled each others codes and able to execute
- We learned how to execute a project End to end within a tight timeline.

Thank you