LAB-04

THEMATIC MAPS

By

Nishant Rodrigues

2348045

5 MSc Data Science

(Reinforcement Learning)

Introduction to the Data

The dataset used in this report contains spatial and demographic information sourced from the U.S. Census Bureau, specifically focusing on the 2010 census data for the contiguous United States. The data includes various attributes for each state, such as:

- **State Name (NAME10)**: The name of each U.S. state.
- **State Code (STUSPS10)**: The two-letter postal abbreviation for each state.
- **Total Population (DP0010001)**: The total population of each state.
- **Hispanic Population (DP0100002)**: The Hispanic population of each state.
- **Area (ALAND10)**: The land area of each state in square meters.
- **Density**: The population density in each area (not explicitly available in the provided dataset but could be derived).
- **Latitude and Longitude**: Coordinates representing the geographical location of each state.

This data is vital for understanding population distribution, regional characteristics, and demographic patterns across the United States. By combining these tabular data with spatial boundary information (shapefiles), it is possible to visualize population metrics and perform spatial analysis.

Procedure for Extracting Vector Data

1. **Loading the Data into QGIS** The first step in working with the data was to load it into **QGIS**, an open-source Geographic Information System software. To begin, the shapefile containing the boundaries of U.S. states was loaded into QGIS. This was done by navigating to **Layer > Add Layer > Add Vector Layer** and selecting the appropriate shapefile.

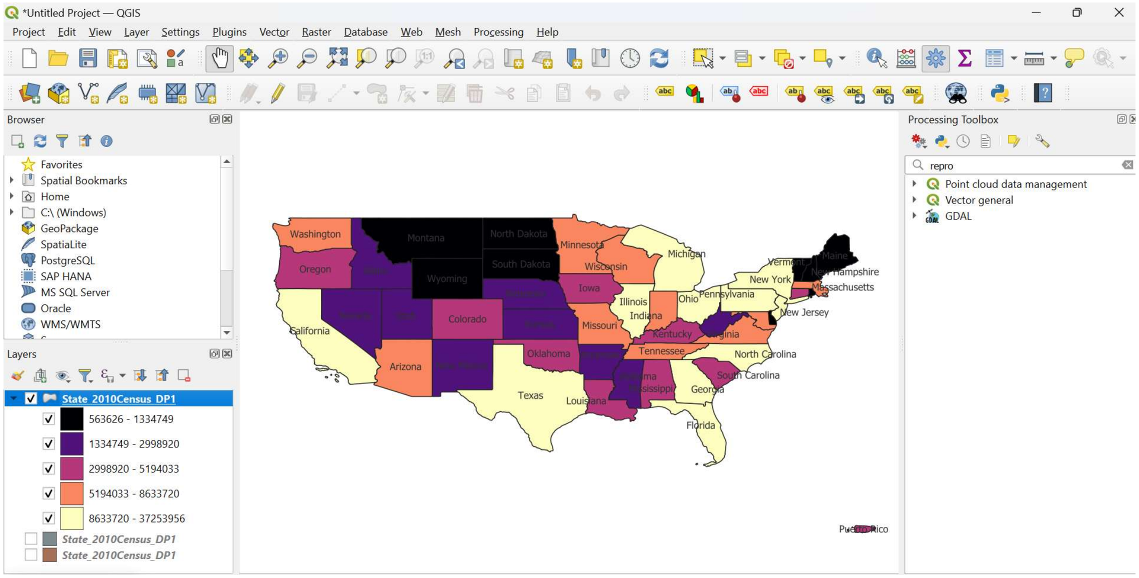
For point data (e.g., if the data were in CSV format with coordinates), the file could be loaded by choosing **Layer > Add Layer > Add Delimited Text Layer**. The columns corresponding to **Latitude** and **Longitude** were then mapped to the X and Y fields, allowing the data to be visualized as a point layer on the map.

2. **Setting the Correct Coordinate Reference System (CRS)** After loading the data, it was essential to ensure the CRS was correct. For this dataset, the appropriate CRS was **EPSG:5070**, which corresponds to NAD83 Contiguous USA Albers. To confirm or change the CRS, I right-clicked the layer and selected **Layer CRS > Set CRS** to match this projection.

3. **Exporting the Data** Once the data was loaded and CRS was verified, I exported the vector data of interest (such as the states' boundaries) for further analysis. By selecting the desired features using the **Select by Attributes** or **Select by Location** tools, I right-clicked the layer and selected **Export > Save Selected Features As** to create a new shapefile for the selected states.

Procedure for Creating Thematic Maps

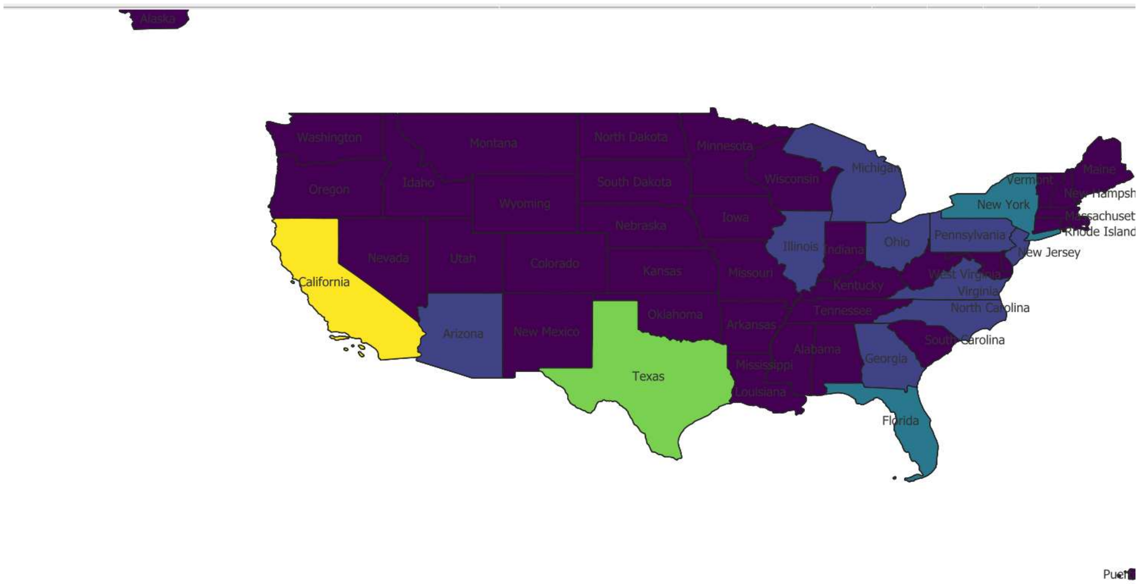
Creating thematic maps allows for better visualization and interpretation of spatial patterns and data trends. Below are the steps followed to create three types of thematic maps: **Choropleth Map**, **Proportional Symbol Map**, and **Heatmap**.



1. Creating a Choropleth Map

A **Choropleth Map** displays the spatial distribution of a specific data variable (such as population density) by shading geographic regions.

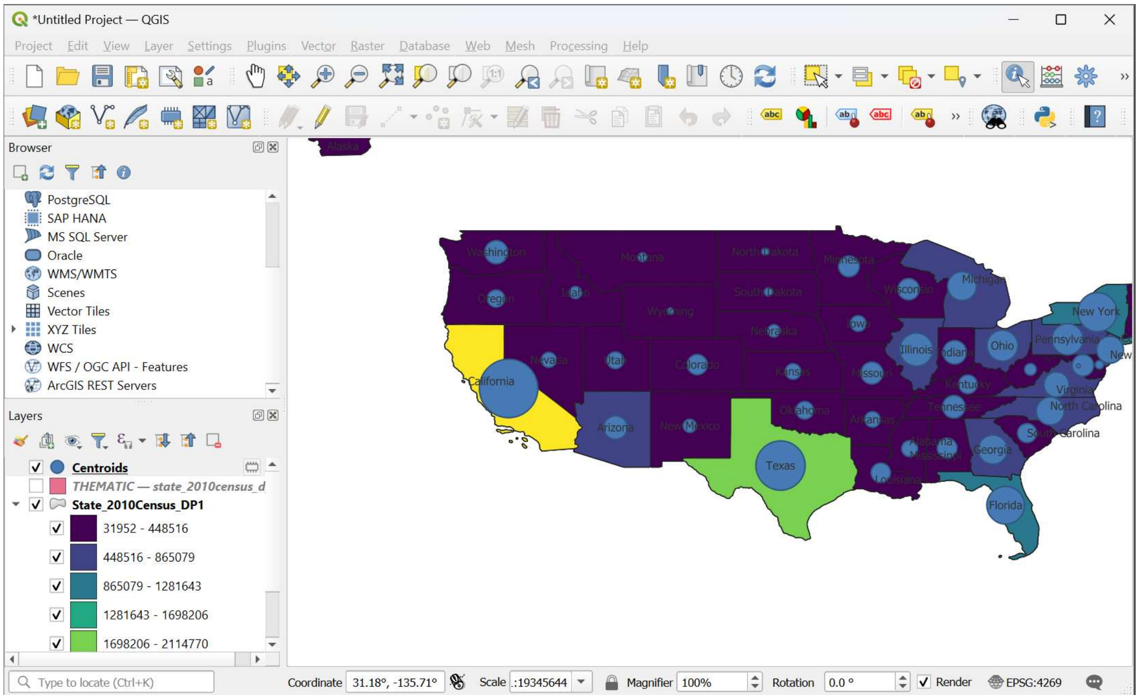
1. **Layer Properties**
   - In QGIS, I accessed the **Layer Properties** by right-clicking on the **Contiguous_States** shapefile (which contains the U.S. states' boundaries) and selecting **Properties**.
2. **Applying Graduated Colors**
   - Under the **Symbology** tab, I chose the **Graduated** option, which allows me to visualize the data in classes. I selected **DP0010001** (Total Population) as the **Value** field to display the choropleth based on total population across the states.
   - I selected a color ramp (e.g., a gradient from light to dark) and clicked **Classify** to divide the data into categories based on population size.
3. **Labeling**
   - For better readability, I added state codes as labels. Under the **Labels** tab, I selected **STUSPS10** (state code) and clicked **OK** to display state abbreviations on the map.
4. **Final Map**
   - The final choropleth map displayed each state shaded according to its total population, with larger populations represented by darker colors. This map effectively communicates the distribution of population across the U.S.



2. Creating a Proportional Symbol Map

A **Proportional Symbol Map** uses symbols (often circles) whose sizes are proportional to the value of the data variable. In this case, I used the total population of each state as the variable.

1. **Layer Properties**
   - I right-clicked on the **Contiguous_States** layer and selected **Properties**. Under the **Symbology** tab, I selected **Proportional Symbols** under the **Quantities** section.
2. **Selecting the Value Field**
   - For the **Value** field, I chose **DP0010001** (Total Population). The size of each circle representing a state was proportional to the state's population.
3. **Adjusting Symbol Size**
   - I adjusted the **symbol size** to make the map more readable, ensuring that larger states (with higher populations) were represented by larger circles.
4. **Final Map**
   - The resulting proportional symbol map used varying circle sizes to represent the total population of each state. This map helped visualize the population magnitudes across the United States.

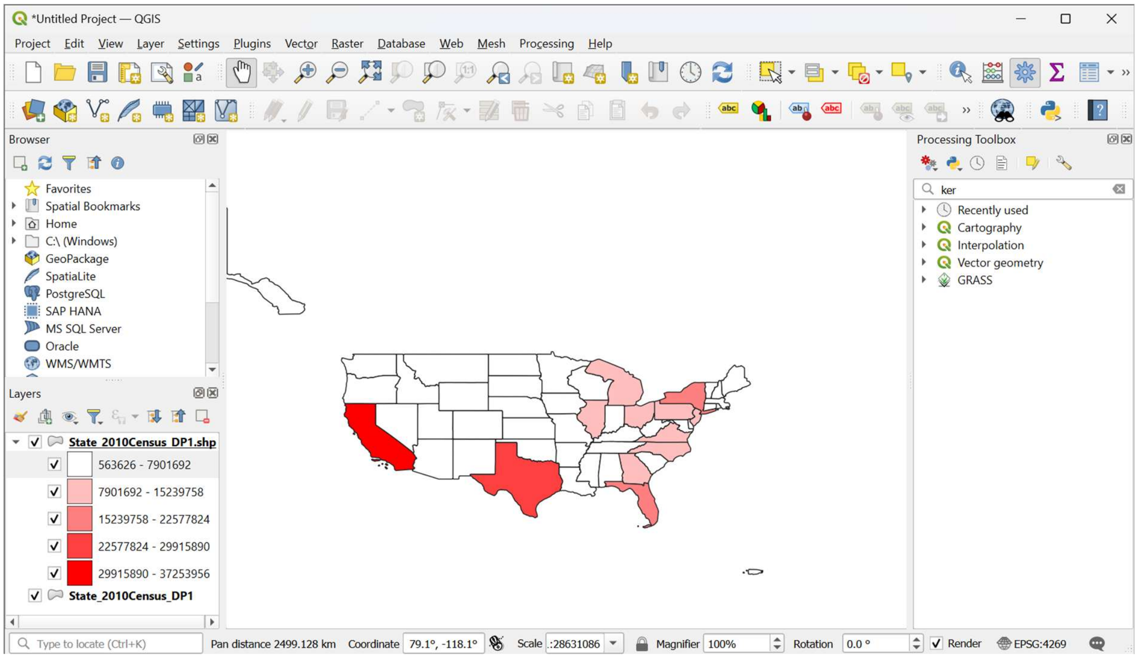


3. Creating a Heatmap

A **Heatmap** is used to visualize the intensity or concentration of point data. In this case, it would be helpful for visualizing the density of population points across geographic locations.

1. **Creating the Heatmap**
    - Using the **Raster** menu in QGIS, I selected **Heatmap** under **Analysis > Heatmap from Point Layer**. I used the **Latitude** and **Longitude** columns from the dataset as the point data source.

2. **Adjusting Heatmap Settings**
    - I fine-tuned the heatmap by adjusting the **radius** to control how concentrated the heatmap effect would be. The color ramp was also selected to highlight areas of higher intensity, with warmer colors indicating higher densities of data points.



3. **Final Map**

   - The heatmap visualized areas with higher concentrations of population or specific demographic features (e.g., Hispanic population). It provided a clear view of "hotspots" where population density was highest.

Conclusion

In this report, we demonstrated how to extract vector data from a U.S. Census dataset and create thematic maps using **QGIS**. Through the use of choropleth maps, proportional symbol maps, and heatmaps, we were able to effectively communicate geographic patterns and trends in population and demographics across the United States.

These techniques are crucial for spatial analysis and are often applied in urban planning, resource allocation, and demographic studies. The maps created provide valuable insights that can assist in making data-driven decisions based on population distribution and other geographical metrics.