Concrete Jungle - New York State Traffic Visualization

Authors

錢承	311551096
Halvor Skribeland	312832801
李耕雨	109550055

Data Visualization and Visual Analytics, 2023

Lecturer

王昱舜

Table of Contents

Introduction

This project introduces an interactive data visualization system. The system provides a user with insight into the traffic of New York State. The main page of the system contains a map of the state of New York and the counties within it. This is an interactive map where a user can click any of the counties and get information about it.

After the user has selected a county, a sidebar appears from the right side. This page contains general information about the selected county in addition to different plots and graphs. These plots include a correlation matrix, a theme river, a line plot, and stacked bar charts. The plots convey different traffic information to the user.

The goal of this project is to create an easy-to-understand system that shows traffic information for specific counties in New York State. It should be interactive, informative, and keep the user interested.

Methodology

Version control

GitHub is used for version control in this project. This allows the authors to work on the same project at the same time. It keeps track of the code and the changes that are made.

Dataset

Three different datasets are used in this project. They contain information about the traffic in the different counties in New York State, the air quality for New York State, and general information about the counties in the state.

- **Annual Average Daily Traffic**
  https://data.ny.gov/Transportation/Annual-Average-Daily-Traffic-AADT-Beginning-1977/6amx-2pbv?fbclid=IwAR266CHyvwVTcTjJPXQUG10qtyQKu8NLxK1z9qHpjr31bcfzp8SSozCawU8
  This dataset contains a lot of traffic information and is where the amount of traffic is collected from. It contains data about every county in New York which includes average daily traffic and what types of roads the different traffic belongs to.

- **Air quality for New York State**
  https://aqicn.org/data-platform/register/?fbclid=IwAR21hrX-qF5UX-13xR7ff-1D7LUwdcz3r_2C053b5p_8P7TptgXoqCJML9M
  This data set is used to look at the air quality and pollution in New York State. The data used from this dataset is a measurement called pm25. PM25 is a measurement on air pollution where the particles are smaller than 2.5 microns in diameter. These particles are bad for humans as they are so small that they can get deep into the lungs and into the bloodstream.

- **New York State county information**
  https://en.wikipedia.org/wiki/List_of_counties_in_New_York
  The dataset is collected from wikipedia, including FIPS Code, County seat, Est., Formed from, Named for, Density, and Area.

- **Data Pre-processing**
  1. Group the data by years and the county
     - The original Annual Average Daily Traffic data is recorded on a day-by-day basis. We group this data by county and calculate the yearly average. Consequently, we obtain the data for each county on a year-by-year basis.

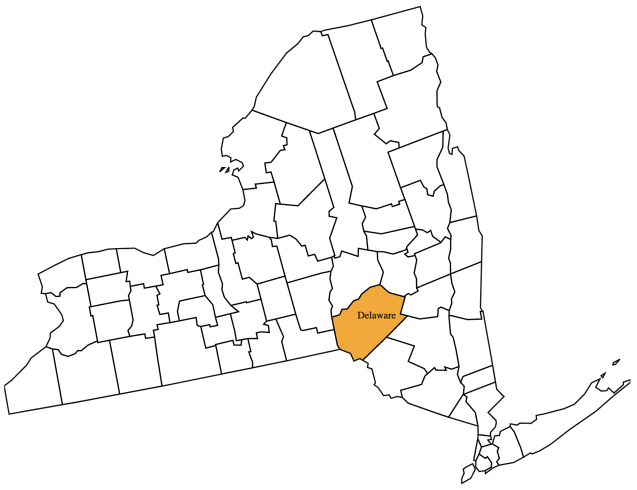
  2. Categorize traffic data by different road types (state, county, and others)

- In addition to the year-by-year traffic data for each county, we also categorize the traffic data by different road types to facilitate more detailed analysis. This approach allows us to understand traffic patterns and trends across various types of roads, such as state roads, county roads, and other roads, providing a comprehensive overview of traffic flow and usage in different areas.

3. Transform daily air quality data into annual averages
    - We transform daily air quality data into annual averages, enabling us to assess long-term trends and patterns in air quality.

Visualizations

- New York State Counties Map
    - Display county names on mouse hover.
    - Enable user interaction for detailed information on specific counties. And show more detail when users click.

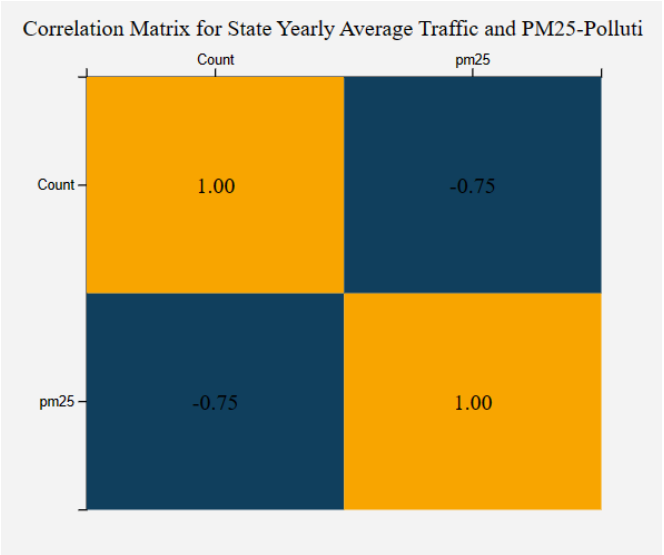


- Counties Information
    - Show the base information of the counties, including FIPS Code, County seat, Est., Formed from, Named for, Density, and Area.

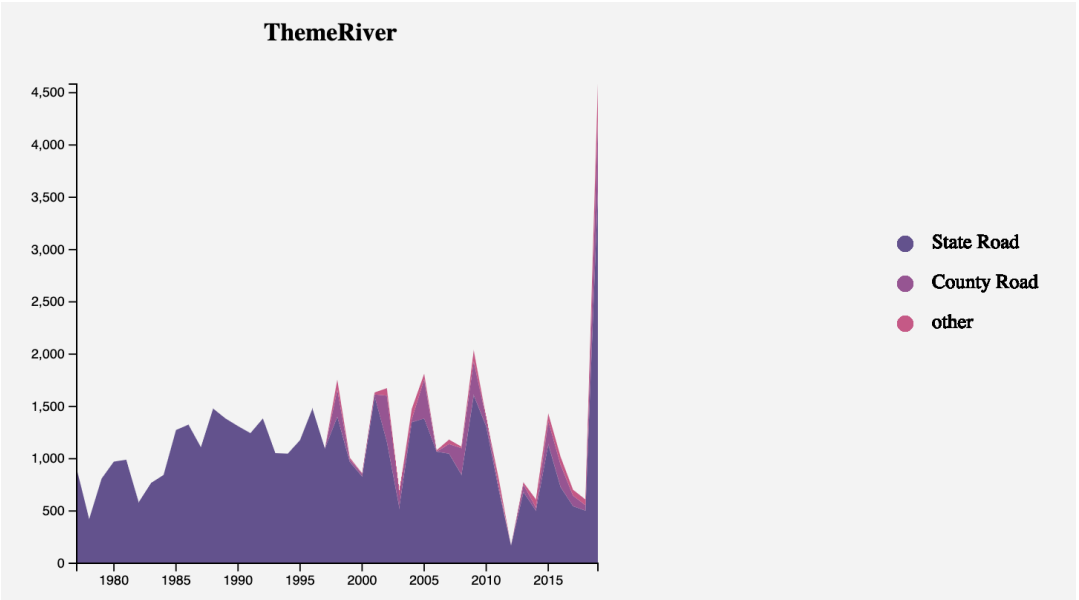
County	FIPS Code	County seat	Est.	Formed from	Named for	Density (Pop./mi2)	Pop. (2022)	Area
Delaware County	25	Delhi	March 10, 1797	Otsego County and Ulster County	Thomas West, 3rd Baron De La Warr (1577–1618), an early colonial leader in Virginia. Name applied to the bay, river, and Lenape Native Americans	30.48	44740	1,468 sq mi (3,802 km2)

- Correlation Matrix

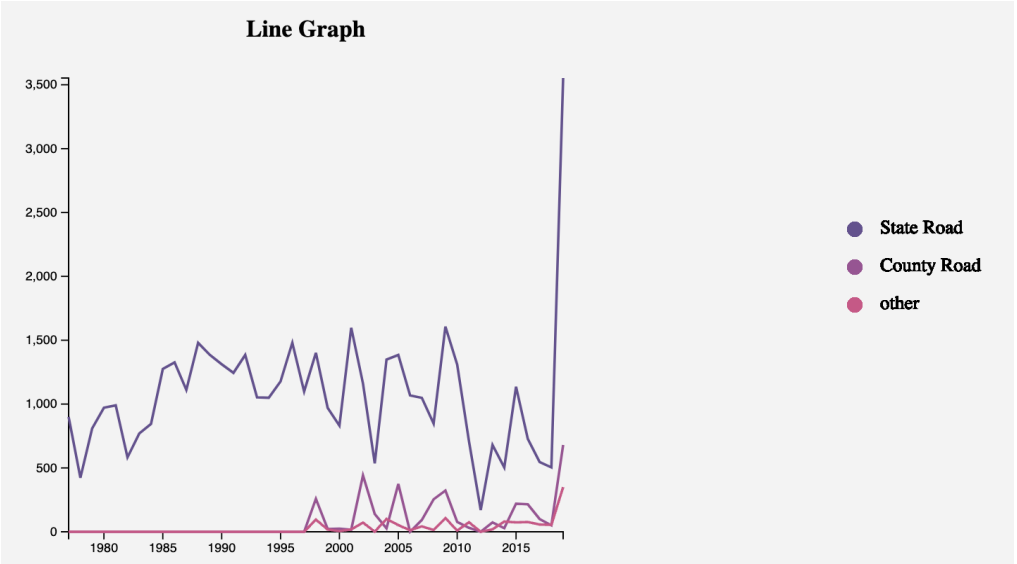
Illustrate the correlation between pollution level and traffic volume.



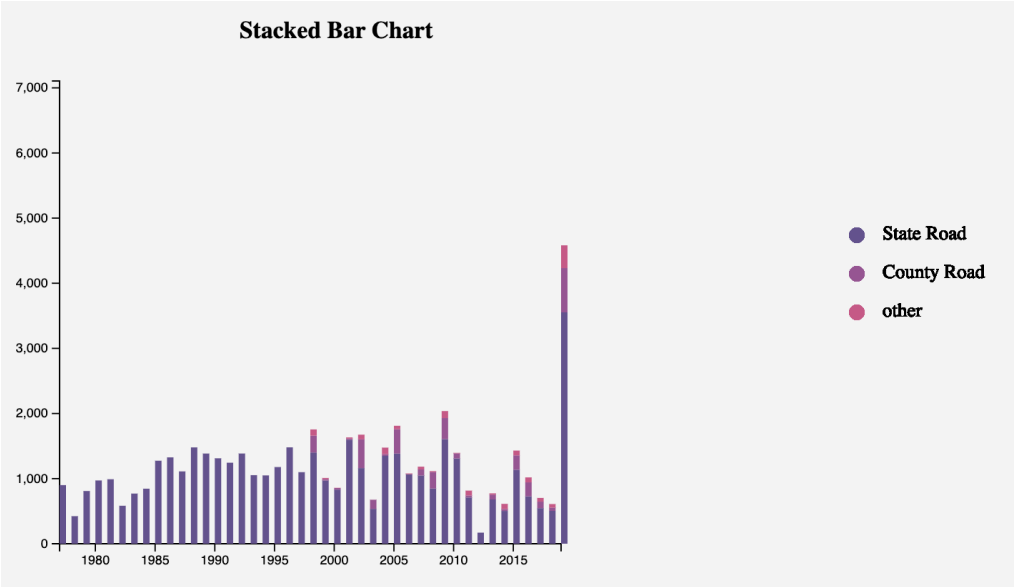
- Theme River
    - Visualize the correlation between traffic and pollution over the years.



- Line Plot
    - Present the average traffic over time.
    - Break down the average traffic by road types: state roads, county roads, and other roads.

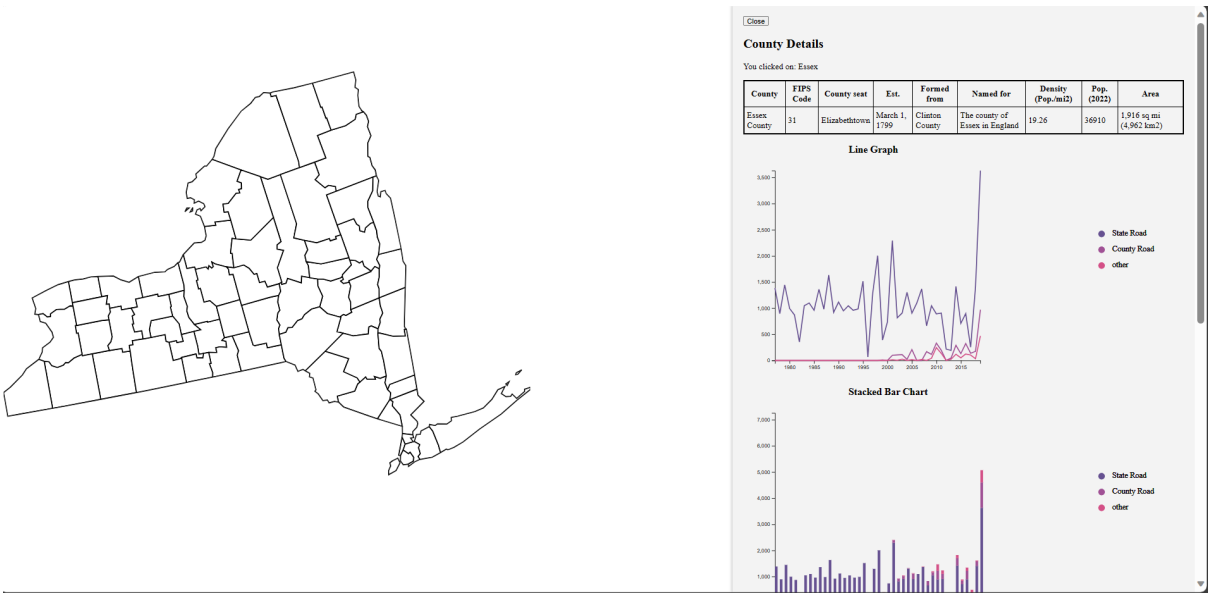


- Stacked Bar Charts
    - Show three kinds of roads (state roads, county roads, and other roads).
    - Highlight road types on mouse hover.



Result

The result from this project looks like this.



It shows information about traffic and air quality in New York State.

It can be used by some people like Researchers, curious citizens, Public Roads administration, and politicians to get information about the trend of traffic counts and the relationship between air quality and traffic counts.

The system is fast, responsive, and easy to use. The user clicks on a county they want to learn more about. A sidebar appears from the right containing a table and plots with general information about the county, and information about the traffic.

Future Work & Discussion

For future work, we think we could expand the area of this project from New York State to the whole of America.

Moreover, we could invite some users to use our project and get some feedback from them or talk with domain experts in traffic and air quality and let them analyze the reason behind the trends. It would also be very valuable to get the help from domain experts to ask them what kind of data is interesting to them, and how they prefer it to be displayed in a manner that is most useful to them.

In the air quality part, due to the lack of the dataset, we could only show the relationship between the pm2.5 and traffic counts. We hope that we could find a few more datasets regarding air quality in the future and show more comprehensive information between air quality and traffic count.

Apart from traffic counts, we can collect some data about traffic accident counts in the future, then we can visualize the relationship between air quality, traffic count, traffic accident count.

Key findings

What we have learned from the data and visualizations is that the overall daily average traffic fluctuates a bit in every county, and usually with a slight trend towards increased traffic. Another finding is that the roads with the most amount of traffic by a big margin are the State roads. By looking at the correlation matrix it is also possible to see that there is a high negative correlation between traffic count and pm25 pollution. When clicking around at the different counties it becomes clear that the population count varies a lot. A county with a high population also has a higher traffic count than a county with a small population.

Conclusion

In conclusion, this project created a fast and interactive data visualization tool that makes the complex traffic data understandable and accessible. The system uses four kinds of charts to visualize the data after some pre-processing.

By using multiple datasets for this system, it is possible to discover patterns in the traffic and pollution across the counties. While there are limitations and room for future improvement, the project can be used as a single accessible place to get a better understanding and overview of the traffic situation in the state. This information is conveyed to the user in the form of a table and four different plots.