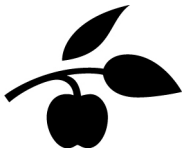

University of Southern Denmark

Teknisk Fakultet

Mærsk Mc-Kinney Møller Instituttet

NYC Traffic Enforcement:

A Data Visualization Representation

Authors:		**Exam nr.**
Sigurd Vind	sivin20@student.sdu.dk	493836
Andreas Gade	angad20@student.sdu.dk	494221
Lauge Solvang	lasol21@student.sdu.dk	4110095
Asma Khatun	askha21@student.sdu.dk	4104409

Advisor:
Parisa Niloofar parni@mmmi.sdu.dk

Semester: E23
Course code: T520040102
Hand-in: 28-12-2023

1 Background, Motivation, and Dataset Selection

Our primary objective is to uncover significant patterns and trends within the extensive **New York City Parking Tickets dataset**. Through meticulous data analysis, we aim to extract insights that extend beyond numerical values, offering a nuanced understanding of parking violations dynamics within the city. The selection of this dataset is deliberate and rooted in its real-world relevance, inherent complexity, and alignment with our academic course objectives, serving as a conduit for students to bridge theoretical knowledge with practical application.

1.1 Project Overview

Our project, focused on visualizing parking violations data, stems from a blend of interests and the ambition to explore a dataset often overlooked in the data visualization realm. By delving into parking violations, we aim to move away from commonly used datasets, offering new perspectives and fresh insights.

1.2 Dynamics of Parking Violations

Parking violations present a compelling mix of legal, societal, and urban dynamics, offering a rich area for exploration. The dataset's inclusion of temporal and geographical variables adds depth to our project. The temporal dimension allows us to create dynamic visualizations that reveal patterns and trends across different times of day, days of the week, and even seasonal variations. Similarly, the geographical aspect enables spatial analysis, aiding in identifying violation hotspots and understanding geographical distributions. The dataset's comprehensive categorical variables also offer a plethora of opportunities for diverse visualizations.

1.3 Challenges and Learning Opportunities

Furthermore, the technical challenge of processing and visualizing a complex dataset with variables like time and location is an attractive aspect of this project. It offers us an opportunity to hone our data processing and visualization skills, contributing to our growth in this field.

1.4 Academic and Educational Significance

The academic significance of our analytical pursuits extends beyond routine data exploration. This endeavor seamlessly integrates theoretical concepts with practical application, providing a holistic view of the subject matter. The outcomes of our analysis provide cross-disciplinary insights into urban studies, transportation management, and policy implications.

1.5 Data Visualization and Educational Impact

Moreover, the incorporation of data visualization into our analytical toolkit elevates the educational value of this report. It transcends traditional boundaries by providing a visually immersive experience. The utilization of various visualization tools not only enhances the comprehension of complex data structures but also cultivates essential skills in communication. As students navigate the intricate landscape of visual representation, they are not merely analyzing data but crafting narratives that bridge the gap between raw information and actionable insights.

1.6 Preparing for Data-Driven Environments

This approach, rooted in the fusion of theoretical knowledge with hands-on experience, underscores the paramount importance of data literacy in the context of future data-driven environments. By fostering an interactive and collaborative learning experience, we aim not only to support multi-perspective analysis but also to instill a passion for data-driven inquiry among students. In essence, this report signifies a journey that extends beyond the confines of academia—a holistic educational experience equipping students with the skills and mindset necessary for the complex landscapes of data analysis and visualization in the real world.

2 Project Objective

The Project visualizes graphical demonstration of important analysis and relationships among the variables of Parking violation dataset. The aim of the project is to find useful information from the dataset and apply some visualization tools to represent graphical views which can be used in future.

2.1 Questions

By analyzing the dataset, a set of questions has been found which convey major information about New York parking violations of the year of 2016.

- What is the distribution of violations in New York, and where are tickets most issued?
- When do parking violations mostly occur?
- Which types of parking violations are issued the most?
- Are there any seasonal or weekly patterns in the issuing of parking violations?
- What were the top five violation counties? What were the top five vehicle types with the highest percentages of violation for each county?
- Which brands of vehicles got the most parking tickets?
- What were the top ten parking violations in New York city in 2016?
- How did the distribution of parking violations change over the course of the year of 2016?

2.2 Objective

The objective of the project is learning and demonstrating the valuable insights of Parking violation dataset. The project provides a better understanding of geographical distributions of parking violations, pick hour of getting parking tickets, categorizing the violation types, vehicle types, vehicle body types, vehicle making brands and so on. Here, we come up with all the useful details of the dataset and represent as graphs.

3 Data

3.1 Data Source

Our data was sourced from a comprehensive dataset available on Kaggle, specifically the "NYC Parking Tickets" dataset for the fiscal year 2016. This dataset, titled *Parking_Violations_Issued_-_Fiscal_Year_2016.csv*, can be accessed at Kaggle: NYC Parking Tickets Dataset.

3.2 Data Description

We use 9 different variables that offer insight into parking violations across New York City for the specified period. These variables include:

- **Violation.Code**: A numeric identifier for each type of parking violation, such as parking in a no-parking zone or at an expired meter.
- **Violation.Time**: The specific time the parking violation was recorded, formatted in a 12-hour clock system.
- **Issue.Date**: The date of issuance for the parking violation ticket, following a month/day/year format.
- **Type**: Classification of the vehicle's plate, indicating whether it is for passenger, commercial use, etc.
- **VehicleMake**: Brand or manufacturer of the vehicle that was issued the parking violation, such as "TOYOT" for Toyota or "HONDA" for Honda.
- **VehicleBodyType**: The type of vehicle body, categorized as "SUBN" for SUV/Suburban, "4DSD" for a four-door sedan, "VAN" for a van, etc.
- **County**: The abbreviated name of the county where the violation occurred, such as "Q" for Queens, "K" for Brooklyn, or "NY" for Manhattan.
- **Violation.Precinct**: The precinct number of the police department responsible for the area where the violation took place.

The dataset is extensive, encompassing 10 million records and occupying a size of 2.15 GB.

3.3 Data Processing

The initial dataset comprising 10 million records was sampled down to 200,000 entries using a custom Python script. This random sampling was necessary to manage the dataset's size for more efficient analysis. The processing stage required minimal cleanup, predominantly involving the correction of a few records with incorrect date and time formats. The majority of the dataset was used as-is, without the need for extensive cleaning or reformatting.

4 Visualization

4.1 Heatmap

The heatmap in Figure 1 serves as an effective visualization for displaying parking violations data over time. It maps the frequency of violations across weeks on the x-axis and days of the week on the y-axis, using color intensity to indicate the number of violations. This design allows for easy identification of temporal trends, such as weekly cycles and seasonal changes. The use of the viridis color palette ensures the map is accessible and interpretable, with darker colors signaling more violations. Interactivity, enabled by plotly, allows users to hover over cells for detailed information, enhancing the utility of the heatmap. Special dates are also annotated, providing context for any deviations in the pattern due to holidays or weather events. Overall, the heatmap offers a clear and interactive way to analyze the distribution of parking violations over time.

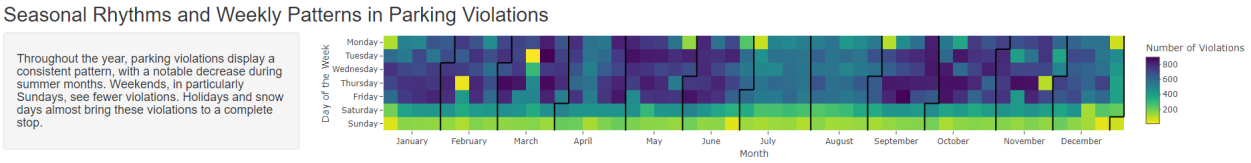

Figure 1: Heatmap Visualizing Parking Violations by Day and Week

4.2 Categories by Hour

The bar chart visualization in Figure 2 represents parking violations over the course of a day by category. The x-axis represents different times of the day, rounded to the nearest half-hour, and the y-axis shows either the count or the percentage of violations, based on user selection. Users can choose to view absolute numbers or proportional data, and they can filter by violation categories through a dropdown menu. Each bar is color-coded to represent different categories of violations, providing a clear visual distinction.

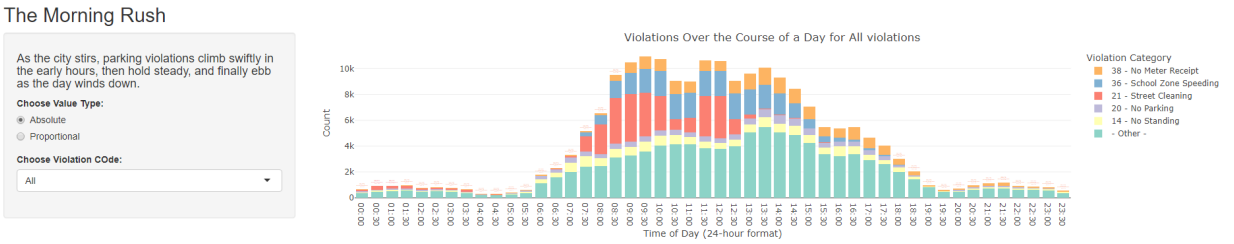

Figure 2: Hourly Distribution of Parking Violations by Category

The chart is interactive, with hover functionality that reveals more details about each time slot and violation category. This enables users to glean specific insights about the frequency and distribution of parking violations throughout the day. By stacking the bars, the chart also allows for a comparison of the relative proportions of each violation category at each time point.

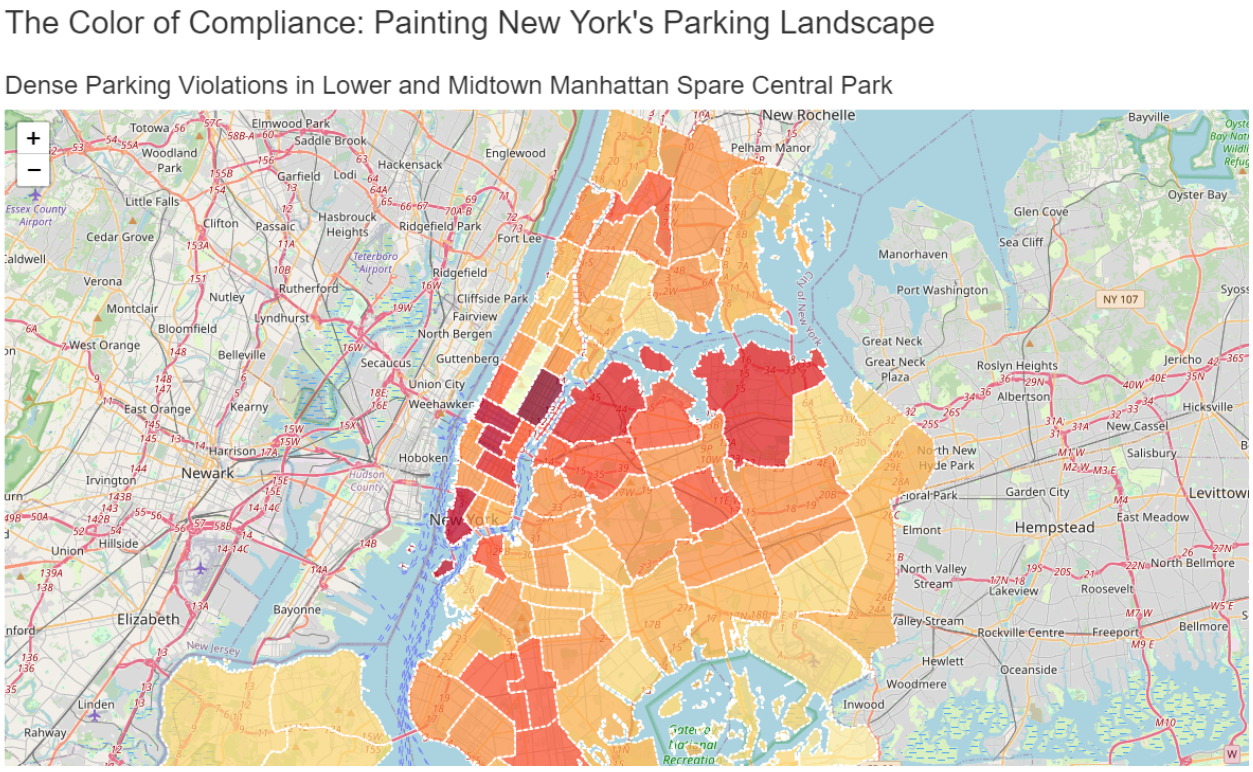

Figure 3: Map of Parking Violations by Precinct

4.3 Violations Map

The interactive map in Figure 3, created using the leaflet library, presents a geospatial visualization of parking violations across New York precincts. The map is a choropleth that assigns colors to precincts based on the number of parking violations, with the color intensity reflecting the magnitude of violations. This visual representation employs the viridis color palette.

A key feature of the map is its interactivity; users can click on individual precincts to receive information about the number of parking violations in that precinct. Highlighting functionality enhances this interaction by outlining the selected precinct, making the user experience more intuitive.

To facilitate clearer distinctions between precincts with varying violation counts, the application uses a root transformation for the color scale breaks. This approach linearizes the distribution of violation counts, allowing for more nuanced visual differentiation between areas with small differences in violations.

4.4 Violations by Vehicle Plate Type

The group bar chart shown in figure 4 demonstrates five violation counties where the highest number of violations occurred and the comparison among top five vehicle plate types across the top counties. The figure exhibits the top five vehicle plate types for each county and makes comparisons among them. Here, vehicle Plate type indicates which kind of vehicle it is. For instance, passenger type, commercial type or any other types. The users can see percentages of each plate type by hovering

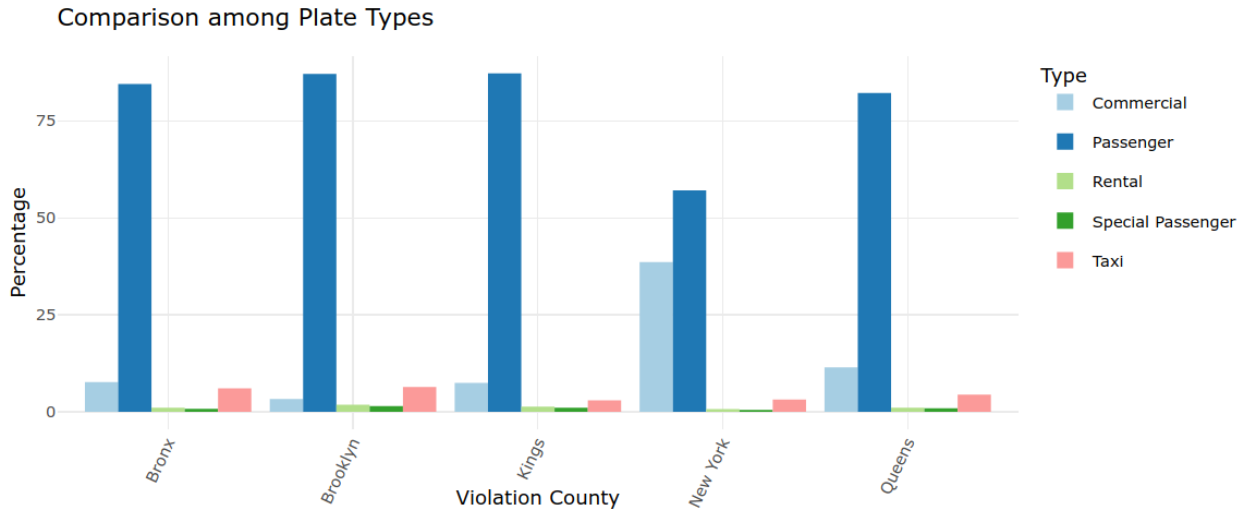

Figure 4: Comparison among Plate type of top five counties

across the bar chart. Furthermore, it has a feature of omitting one or more vehicle types. For example, if a user wants to see the comparison of passenger vehicle types across the top five counties, he or she can click on the remaining four vehicle types to do so. It will temporarily remove the bars of clicked plate types and present the only passenger type bar.

4.5 Violations by Vehicle Make

The bar chart visualized in Figure 5 shows the distribution of parking violations by vehicle make. The bar is interactive, allowing the user to choose whether to see it ascending or descending, which increases the readability of the graph. It was made with plotly, and also features a hover functionality which allows you to see the exact percentage that that brand of vehicles makes out of the total amount of vehicles that got parking tickets in 2016.

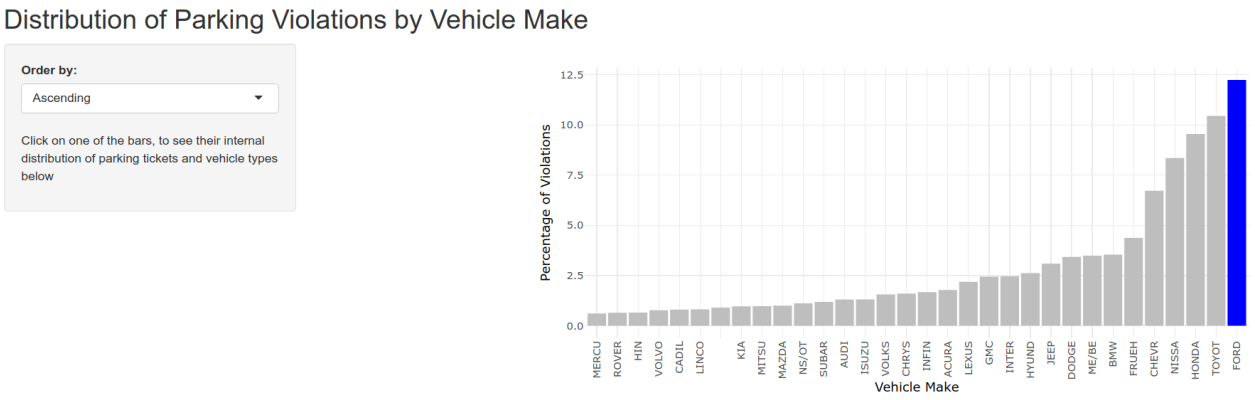

Figure 5: Distribution of parking violations by vehicle make

Additionally the bars on the graph are clickable which changes the data of the pie chart and the donut chart that is placed below, as seen in Figures 7 & 6. The bar graph featuers a popout, as

clicking on e.g. "TOYOT" changes the color of the corresponding bar in the bar graph, and then lets you see the body type distribution and most likely parking tickets for that brand of vehicles.

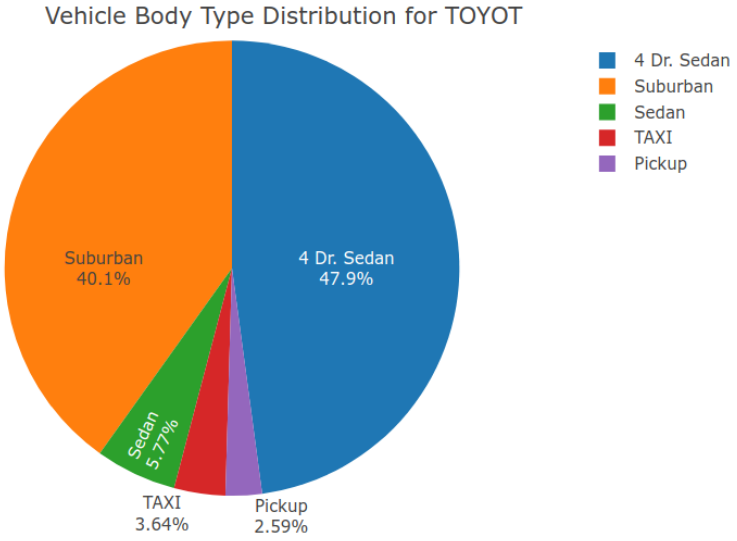

Figure 6: Vehicle body type distribution

The pie chart as previously mentioned and seen in Figure 6, shows the distribution of vehicle body types. This distribution has been translated from shorthand abbreviations to real text, to further increase the readability of the data. I.e. the abbreviation "4DSD" means "4 Door Sedan", and as an example is responsible for about 48% of parking violations for Toyotas in New York

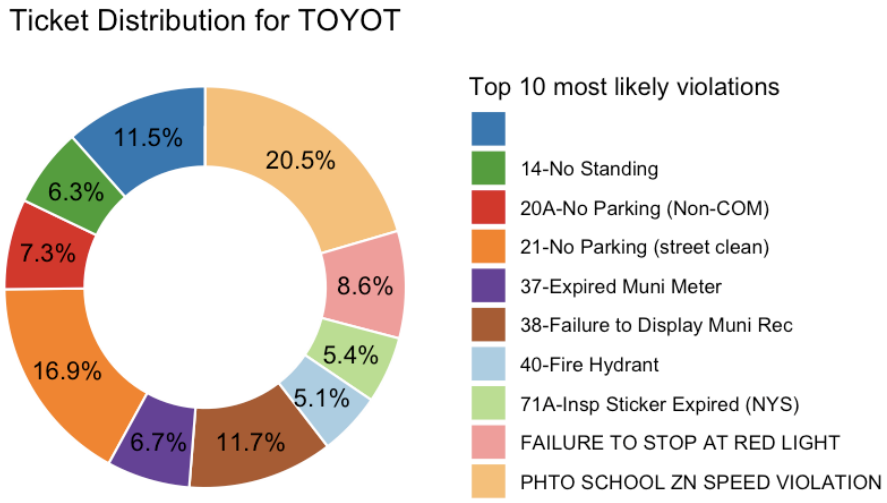

Figure 7: Donut chart for violations by vehicle make

The donut chart, as previously mentioned and seen in Figure 7, shows what ticket a given vehicle make is most likely to get based on the dataset used. Here it uses the violation codes combined with the violation description to be more descriptive in which violation was given.

As mentioned earlier in section 4.5, it is reactive and changes based on which vehicle make is chosen from Figure 5.

4.6 Violation Trends over time

The use of an animated graph, as seen in Figure 8 adds an interactive and dynamic element to the visualization, making it more engaging and visually appealing for the users. It allows them to see the trend in violations over time in a more dynamic and intuitive way, enabling them to easily identify patterns, fluctuations, and any significant changes in the data.

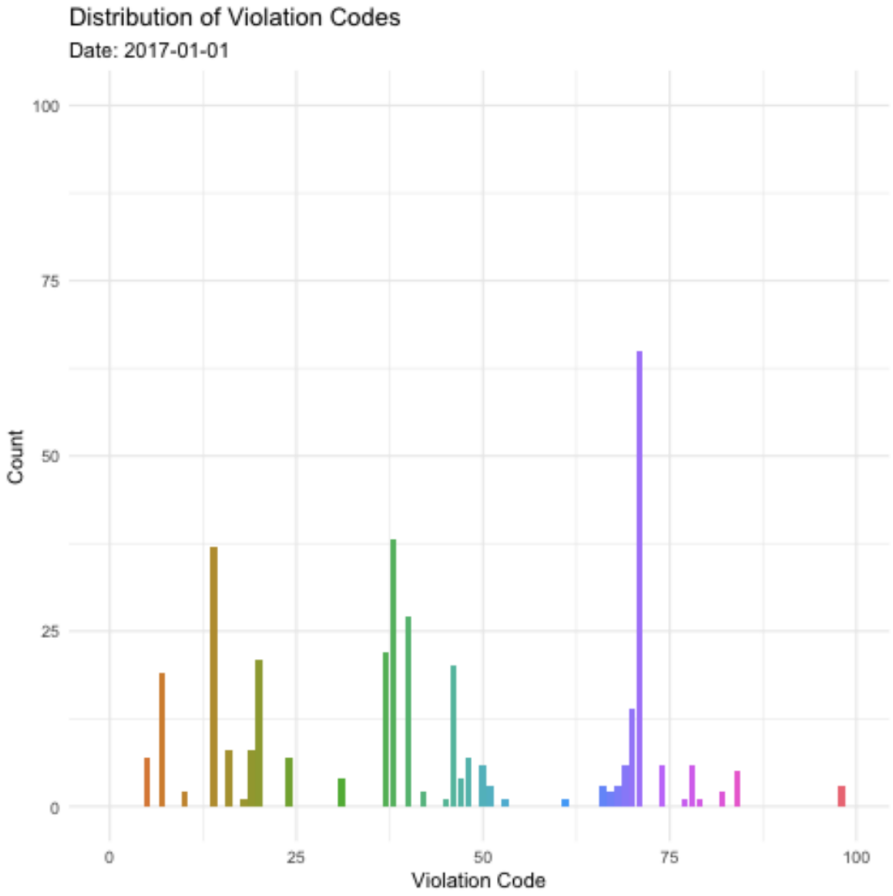

Figure 8: Animated Bar chart for Violation Trends

It was created by mapping violations to a bar chart that reflects the amount of tickets issued for a certain violation code per day. A temporal factor was then appended onto the bar chart to allow for a visual representation of the magnitude of violations over the time span that the dataset spans on.

5 Story & Results

This section aims to answer the questions presented in the in Section 2: Project Objective and suggest further improvements to the visualization of the dataset.

5.1 What is the distribution of violations in New York, and where are tickets most issued?

The interactive map clearly shows the distribution of parking violations around New York. Manhatten is shown as the place with the most parking violations, which we expected as Manhatten is also the most densely populated area of New York. Interestingly, specifically the area east of Central Park called Park Avenue has significantly more parking violations than the rest of the precincts in New York. We expected that the distribution was more even throughout Manhatten, and that the further from Manhatten you got, the less parking violations were present. This somewhat shows in the map, however Staten Island has less than the average amount of parking violations, which follows the trend of the population, as it has a much smaller amount of residents compared to Manhatten, Brooklyn, Queens, etc.

5.2 When do parking violations mostly occur, and which types of parking violations are issued the most?

Generally parking violations mostly happens betweem 0600 and 1900. The trend also shows that parking violations climb quickly during the morning, which might be an indication of the New York morning rush, where parents have to drop off their children at institutions and people have to hurry to work. It then falls a bit again from 10-11:30, which seems to be because Street Cleaning tickets are less issued. This might be because fewer street cleanings are scheduled in this timeslot.

Another big contributer to why there are more parking tickets between 08 and 15 is School Zone Speeding, which is something we expected as the schools are open in this timespan. The distribution of "No Parking", "No Standing" and "No Meter Receipt" tickets are somewhat evenly distributed in the hours that people are awake, which is also something we expected.

Another interesting fact is that you can easily connect outliers with either weather events like snowstorms or holidays like Memorial Day, Independence Day, etc. This is most likely because people are off work these days, meaning they do not drive to work or school.

5.3 Are there any seasonal or weekly patterns in the issuing of parking violations?

The heatmap of parking violations clearly shows seasonal and weekly rhythms in the issuance of parking violations. Saturdays and sundays have a significantly lower amount of parking violations than the rest of the weekdays, which is expected as people are not working or at school during the weekends, but are much more likely to stay at home.

Additionally the data also shows that parking violations were much less present during the summer months of july and august, as well as the last days of december. This might be an indication of people going on holidays, which would mean that a lot more people travel out of New York for summer vacations or go home to their parents during Christmas. The summer months are always the months where most tourists travel to New York, but as the majority of tourists most likely

travel to New York for leisure, many of them are more likely to take a taxi than to have brought their own cars or rented cars.

5.4 What were the top five violation counties? What were the top five vehicle types with the highest percentages of violation for each county?

To get the answer of the questions, two variables from the dataset had been considered e.g. violation county and plate type. In the result Bronx, Brooklyn, Kings, New York and Queens were the top five counties where the highest number of parking violations occurred. The violations were conducted by several vehicle types. The top five vehicle types for each county are passenger, commercial, rental, special passenger and Taxi. The order of plate types for every county is the same. Among these, passenger type vehicles got more parking tickets whereas special passenger type vehicles got least parking tickets. It can be seen that, passenger type vehicles got similar amounts of parking tickets in most counties with the exception of New York. In New York, commercial type vehicles got the nearest number of tickets of passenger types while in other counties, commercial type vehicles got far fewer tickets.

5.5 Which brands of vehicles got the most parking tickets?

The distribution of which vehicle makes got the most parking violations in 2016 somewhat follows the expectations of the group to the dataset. Brands like Ford, Toyota, Honda, Nissan, Chevrolet were among the brands which were responsbile for most parking violations, which follows the trend of most popular car brands in america. Sadly no data from 2016 was able to be procurred, but as shown in the article by Carlogos [1], the top 5 vehicle makes in 2021 included exactly the 5 aforementioned brands. Interestingly Fruehauf is the 6th most responsible for parking violations, which is a company that sells both semi-trailers and trailers for trucks. By clicking on it, and viewing the distribution of vehicle body types, you can see that these are mostly Delivery trucks and Vans, which makes sense as they are also susceptible to break the parking laws.

5.6 How did the distribution of parking violations change over the course of the year of 2016?

The distribution over the course of 2016 is pretty consistent with the highest concentration of tickets ranging in violation codes from 30-40. This means that, according to the dataset, most people got tickets due to parking at an expired muni meter (violation code 37), parking without a valid ticket displayed in the front windshield (violation code 38) and parking too close to a fire hydrant (violation code 40).

There is a couple of outliers, as seen in Figure 8, where there is a spike in violations ranging from 65-75, where the violation code is 71 which means the vehicle has been parked without showing a current NY Inspection sticker. Looking at the date for that ticket may tell us more as to why that happened, as it is the first day of the year.

5.7 Further improvements

To enhance our NYC parking violations study, we intend to adopt geocoding technology. This approach will transform street addresses from our dataset into precise geographic coordinates. This shift is crucial for overcoming the limitations imposed by the varying sizes of precincts, which

currently affect our analysis. By utilizing exact geographic coordinates, we can refine our mapping techniques, moving beyond precinct borders as our primary unit of measurement. This allows for a granularity that pinpoints violations down to individual houses. Such precision will enable us to more effectively identify and visualize hotspots, providing a clearer understanding of areas with high and low frequencies of parking violations. Consequently, instead of relying on choropleth maps, which offer a broader, area-based view, we plan to use heatmaps. Heatmaps will offer a more detailed and accurate representation of violation distribution, highlighting the intensity of violations in specific locations

Another improvement would be to normalize the distribution of parking violationg by vehicle make as shown in Figure 5. Normalizing it by comparing the data to all the registered vehicles in New York would make it easier to spot whether people who drives brands like Ford get statistically more parking tickets than others, or if they are responsible for approximately 12% of parking tickets simply because they make up 12% of cars registered in New York.

6 Discussion & Conclusion

Working with dates in R, was challenging due to the variety of formats and the precision required in manipulation. This became particularly evident when incorporating date-related data into the graphs, where accurate and intuitive date handling was essential for effective data representation.

The process of converting the static plots we made in R to Rshiny were where most of the problems occurred. Scaling and displaying graphs appropriately within RShiny proved to be a significant hurdle. Although creating plots in R was relatively straightforward, integrating these plots into RShiny brought forth issues of detail loss and scaling. Graphs that looked clear and well-scaled in R often encountered display problems when transitioned into the RShiny environment. The complexity further increased when these plots, initially static in R, were to be integrated into RShiny with added interactive elements. This transition required a deeper understanding of both R programming and the functionalities of the RShiny framework. The challenge was not only in making static visualizations interactive but also in ensuring that these interactive elements worked seamlessly and enhanced the user's experience without compromising the accuracy and clarity of the data presented.

Our exploration of the New York City Parking Tickets dataset exemplifies the transformative power of data visualization in understanding complex data sets. Our approach utilized various data visualization techniques, each highlighting different aspects of the dataset:

6.1 Heatmap for Temporal Trends

We utilized heatmaps to depict the frequency of parking violations across different times and days. The heatmaps unveiled patterns such as peak violation hours and lower incidences on weekends, providing valuable insights into the rhythm of urban life and its impact on parking behavior.

6.2 Bar Charts for Category Analysis

Bar charts served as an effective tool for analyzing the distribution of parking violations across different vehicle categories and times of the day. They offered a clear understanding of which types of vehicles were most commonly cited and the times when these violations occurred most frequently.

6.3 Interactive Map for Geographical Distribution

The implementation of an interactive chloropleth map allowed us to showcase the geographical spread of parking violations across New York City. This visualization highlighted areas with high violation frequencies, such as Manhattan, especially near Park Avenue, emphasizing the intricate relationship between urban geography and parking violations.

6.4 Dynamic Graphs for Trend Visualization

Dynamic graphs were seamlessly integrated to demonstrate the fluctuating trends of parking violations over the year. This dynamic representation aided in identifying periods with high violation rates, revealing correlations with seasonal variations and special events.

6.5 Summary of further improvements

To enhance our NYC parking violations study, we plan to adopt geocoding technology to convert street addresses into precise geographic coordinates. This will overcome current limitations related to varying precinct sizes, allowing us to pinpoint violations down to individual houses for more accurate hotspot identification.

Additionally, we aim to normalize the distribution of parking violations by vehicle make. This involves comparing the data to all registered vehicles in New York to identify statistically significant patterns, such as whether certain brands receive more tickets proportionate to their representation in the overall registered vehicle population.

This project not only deepened our understanding of the dynamics of parking violations in a major urban center but also taught us the pivotal role of data visualization in the analysis and communication of complex datasets.

References

[1] "Popular car brands." [Online]. Available: https://www.carlogos.org/popular-car-brands/

A Group contributions

The group contributed evenly to the project and report, as shown in Table 1 below.

Student	**Percent of Contribution**
Sigurd Vind	100%
Andreas Gade	100%
Lauge Solvang	100%
Asma Khatun	100%

Table 1: Table of contribution

B Link to dashboard

The dashboard can be found at the link below: https://sigurdsv.shinyapps.io/ParkingViolations/