Presentation Summary

Title: Traffic Data Analysis and Feature Construction

---

1. Introduction

- **Objective**: This project focuses on analyzing traffic data to identify patterns, correlations, and trends over time. The analysis includes both univariate and multivariate techniques to understand the data better.
- **Dataset**: The dataset includes traffic data with various features such as `DateTime`, `Junction`, `Vehicles`, etc.

---

2. Data Exploration

- **Initial Data Examination**:
  - The dataset is loaded and the first few rows are inspected using `data.head(5)`.
  - Column names and data types are checked.
  - Summary statistics are generated to get an overview of the dataset.
  - The dataset contains missing values and duplicates, which were counted using `data.isnull().sum()` and `data.duplicated().sum()` respectively.
- **Numeric Features Correlation**:
  - Correlation among numeric columns was calculated to identify any significant relationships using `data.select_dtypes(include="int").corr()`.

---

3. Feature Construction

- **DateTime Features**:
  - The `DateTime` column is converted to a datetime object to extract useful components such as `day`, `month`, `year`, `time`, and `weekday`.
  - The `weekday` feature is constructed to map the day of the week to a numerical value (Monday = 0, Sunday = 6).
  - The `DateTime` column was dropped after extracting these features for simplicity.

---

4. Univariate Analysis

- **Visualizing Distributions**:
  - Count plots for categorical features like `Junction`, `day`, `month`, `year`, and `weekday` were generated to understand the distribution of data across these categories.

    - Distribution plots were also generated to observe the spread of data within these features.

---

5. Multivariate Analysis

- **Time-Based Labeling**:
    - A function was created to label each hour of the day into ranges such as "00:01 - 00:59", allowing for a more granular analysis of time-based patterns.
    - This labeling was applied to the `time` feature to create a new column, `time_label`.
- **Correlation and Visualization**:
    - A violin plot was created to compare the distribution of vehicle counts across different junctions.
    - A heatmap was generated to visualize correlations among numeric features.
- **Advanced Visualizations**:
    - A box plot was created to visualize the distribution of `Vehicles` across different `time_label` categories.
    - A point plot was used to represent the mean values of `Vehicles` by year, with a confidence interval.

---

6. Statistical Analysis

- **ANOVA Test**:
    - An ANOVA test was performed to analyze whether there are statistically significant differences in vehicle counts across different years.
    - The F-value and p-value were reported, providing insights into the significance of these differences.

---

7. Conclusion

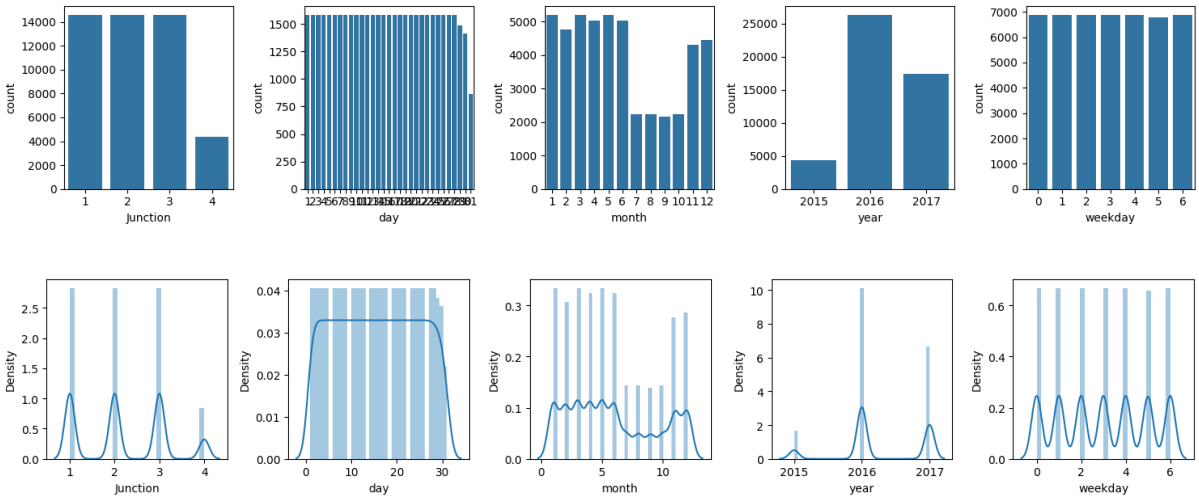
- The analysis provided insights into traffic patterns, including how vehicle counts vary by time of day, day of the week, and junction.
- The statistical test (ANOVA) helped in understanding the variation in vehicle counts over the years.
- This analysis could inform decision-making processes related to traffic management and urban planning.

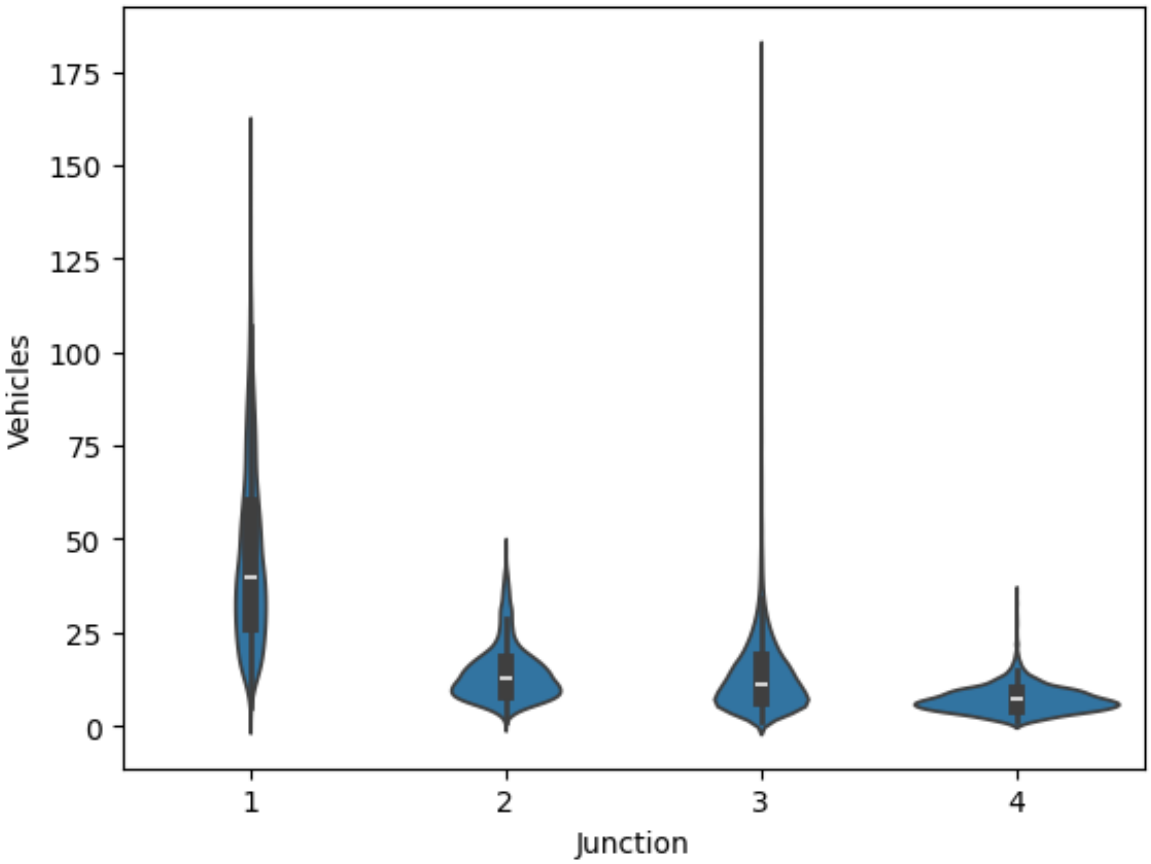
Next Steps

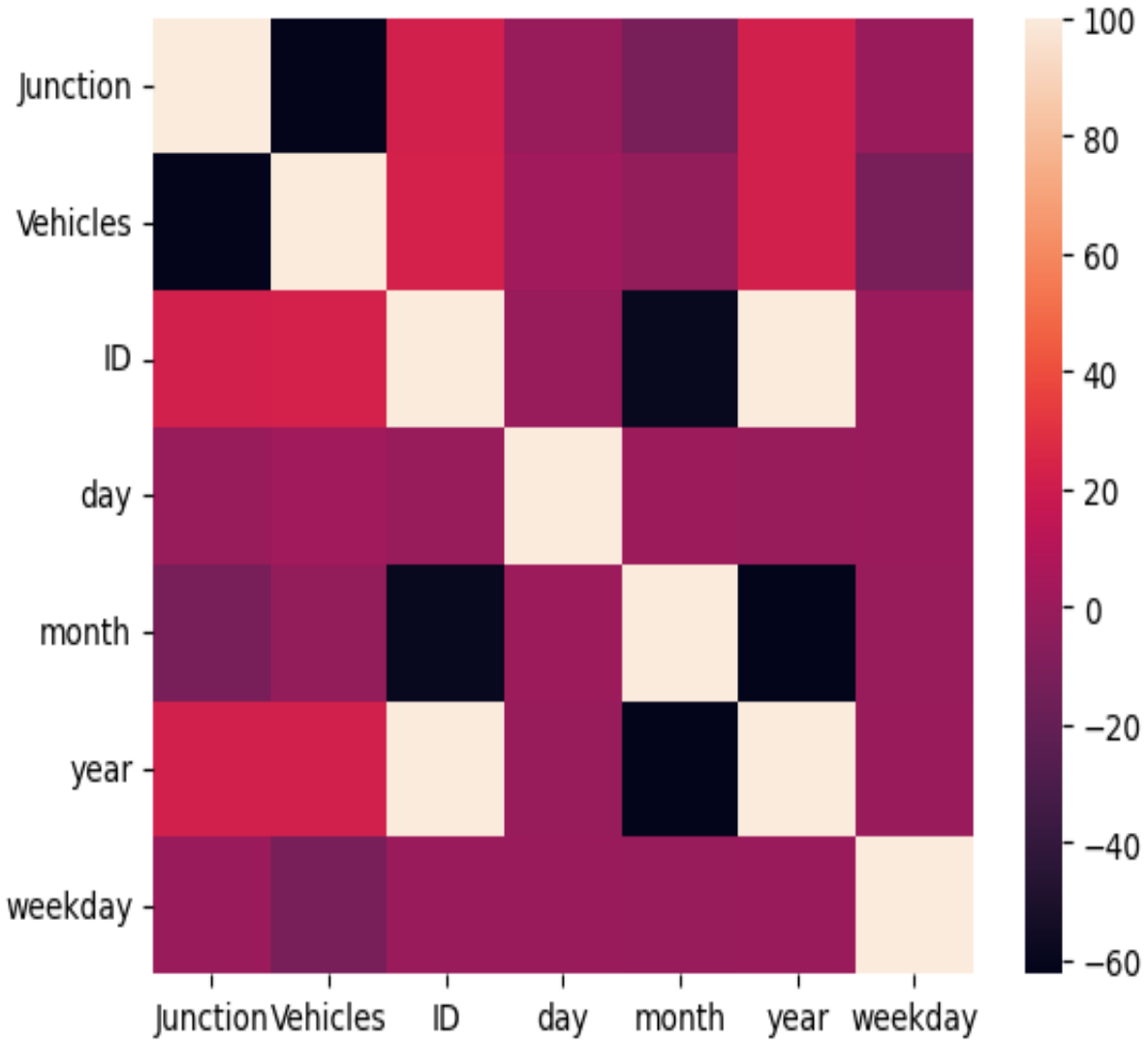
- Further analysis could explore the impact of external factors like weather or public events on traffic patterns.

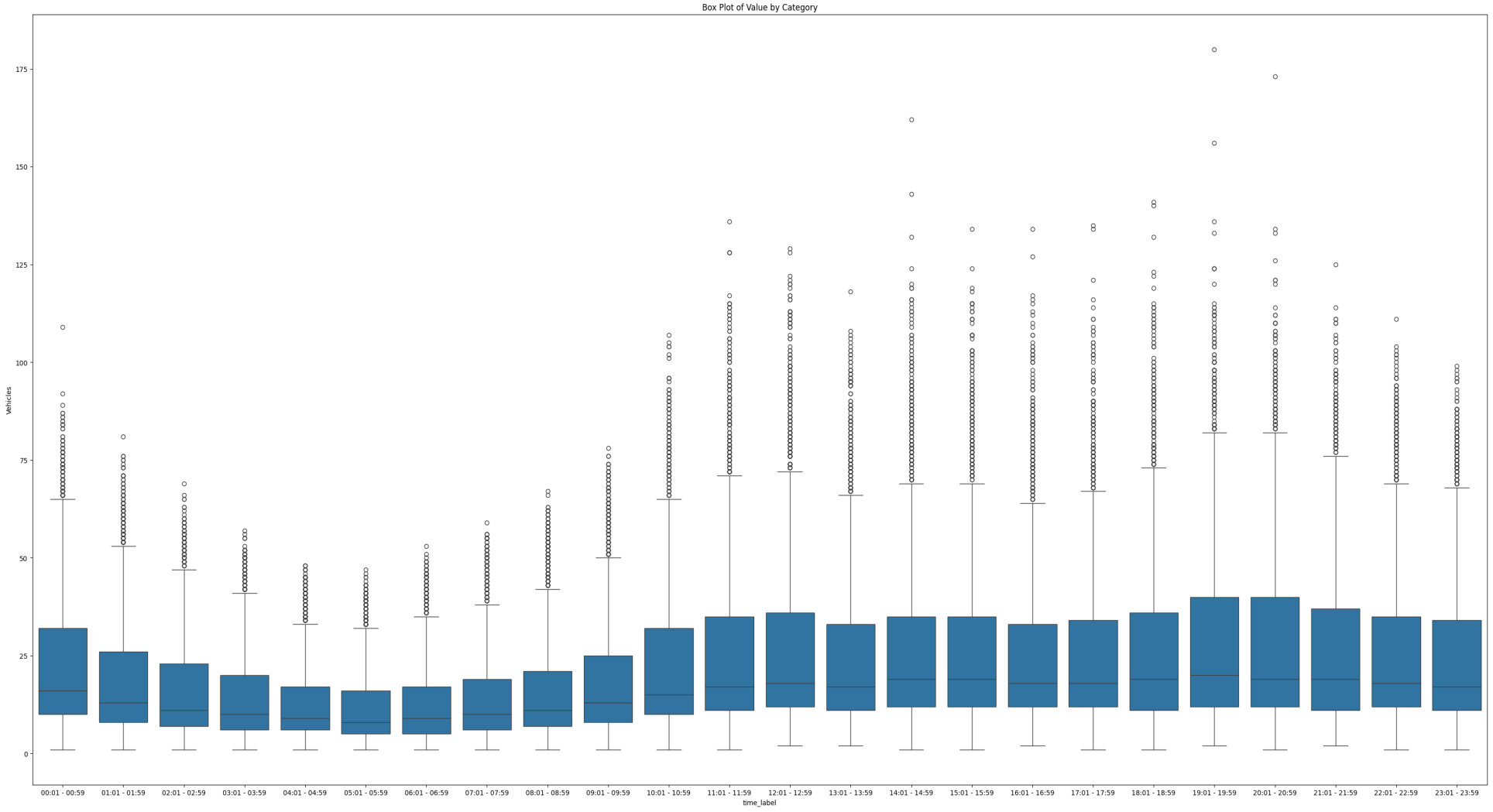
- Implementing predictive models to forecast traffic conditions based on historical data could be the next phase of this project.

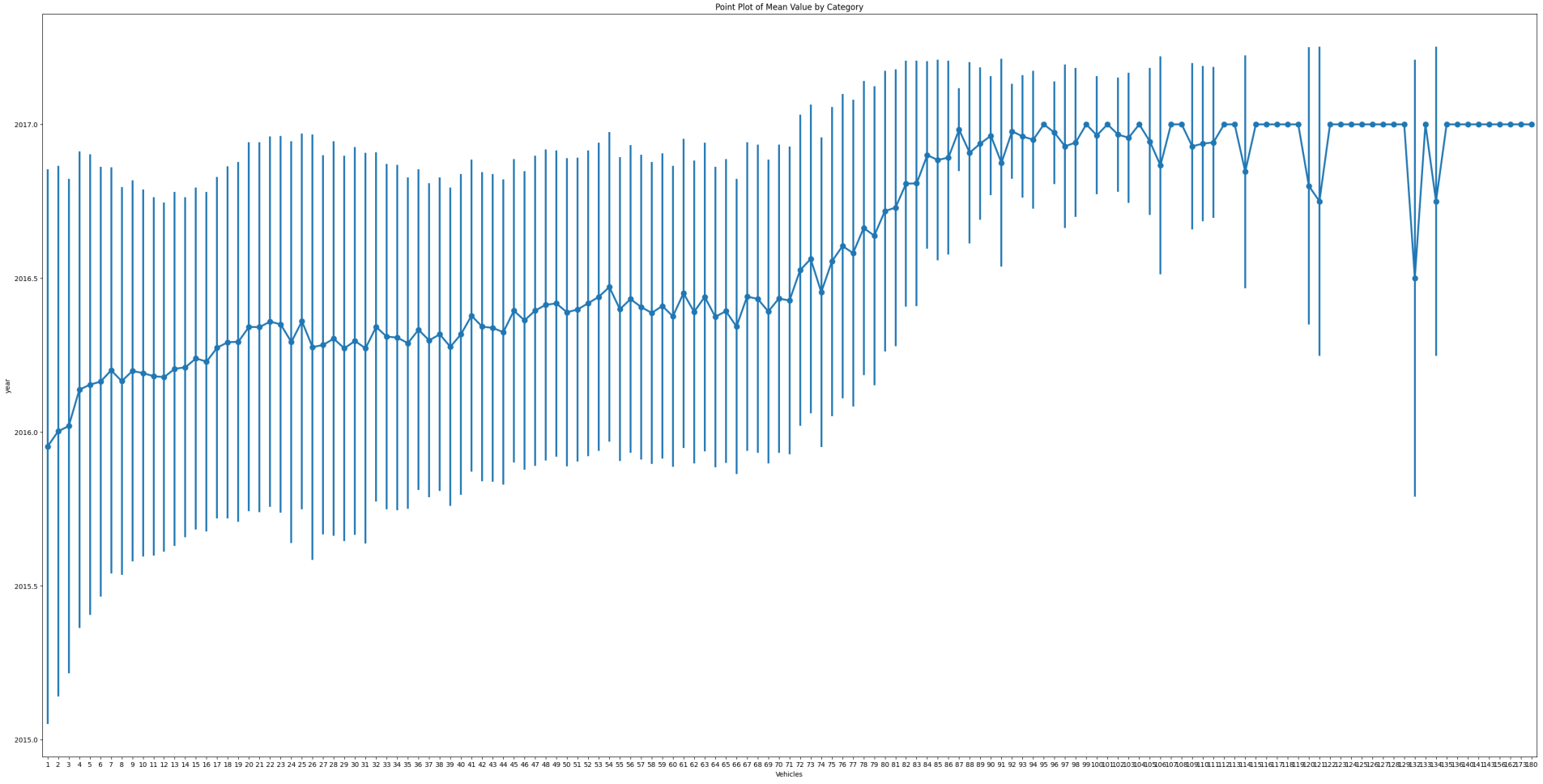
Finding:






Junction
Vehicles
ID
day
month
year
weekday
Junction
Vehicles
ID
day
month
year
weekday
100
80
60
40
20
0
−20
−40
−60


Box Plot of Value by Category
Vehicles
0
25
50
75
100
125
150
175
00:01 - 00:59
01:01 - 01:59
02:01 - 02:59
03:01 - 03:59
04:01 - 04:59
05:01 - 05:59
06:01 - 06:59
07:01 - 07:59
08:01 - 08:59
09:01 - 09:59
10:01 - 10:59
11:01 - 11:59
12:01 - 12:59
13:01 - 13:59
14:01 - 14:59
15:01 - 15:59
16:01 - 16:59
17:01 - 17:59
18:01 - 18:59
19:01 - 19:59
20:01 - 20:59
21:01 - 21:59
22:01 - 22:59
23:01 - 23:59
time_label


Point Plot of Mean Value by Category
year
2017.0
2016.5
2016.0
2015.5
2015.0
Vehicles