



Portland Traffic Analysis: Final Report

Xu, Suru Nelson, Austen
suru@pdx.edu ajn6@pdx.edu

March 19, 2025

Contents

1 Executive Summary	2
2 Introduction and Background	2
3 Data	2
3.1 Data Quality	2
4 Methods	3
4.1 Generalized Additive Model (GAM)	3
4.2 Maximum Mean Discrepancy (MMD)	4
4.2.1 Data Structuring Strategy	4
4.2.2 MMD Model for Traffic Pattern Comparison	5
4.2.3 Statistical Inference Using Permutation Testing	5
5 Results	5
5.1 Generalized Additive Model	5
5.1.1 Weekday ‘Morning Rush’	5
5.1.2 Weekday ‘Afternoon Rush’	5
5.1.3 Weekend	6
5.2 Maximum Mean Discrepancy	6
5.2.1 Key Observations and Implications	6
5.2.2 Summary of Observations	6
6 Conclusions	6
6.1 Code Availability	7
A Median Peak Time and Volumes	8
B MMD Groupings	9
C MMD P-Values	14
D GAM Figures	15
E Data Availability	22
F Data Issues Deep Dive by Locations	23



1 Executive Summary

The COVID-19 pandemic significantly altered traffic patterns in Portland, reshaping commuter behavior and roadway demand. This report analyzes historical traffic count data to assess how peak traffic periods and overall volumes have shifted on two major highway crossings: the I-5 and Glenn Jackson bridges.

Our analysis identifies key changes in traffic demand across different time periods and weekdays, highlighting how commuter habits have evolved post-pandemic. Findings indicate a notable flattening of morning rush-hour peaks, a slight reduction in afternoon traffic volumes, and relatively stable weekend travel patterns. Additionally, distinct variations between the two bridges suggest localized differences in travel behavior and infrastructure utilization.

These insights provide valuable guidance for transportation planning, congestion management, and infrastructure investments. As commuting patterns continue to adapt, future studies incorporating real-time data sources could further refine our understanding of long-term traffic trends in the region.

2 Introduction and Background

The COVID-19 pandemic brought significant shifts in mobility patterns, altering daily commutes and reshaping traffic flow on major highways [4, 2]. This study investigates these changes by analyzing historical traffic count data, with a focus on two critical transportation corridors in the Portland area: the I-5 and Glenn Jackson bridges. Understanding how traffic demand and peak periods have evolved post-pandemic is essential for transportation planning, congestion management, and infrastructure development. These insights can inform policy decisions on commuting patterns, public transit investments, and roadway capacity planning.

To address these needs, the Oregon Department of Transportation (ODOT) seeks a comprehensive, data-driven analysis using historical traffic data from Automatic Traffic Recorders (ATR). To capture the impact of the pandemic on traffic behavior, this report explores the following key questions:

- How have peak traffic periods changed in terms of structure (flatter, sharper, later, or earlier throughout the day) at these key locations?
- How has daily traffic demand shifted across different days of the week at these bridges?
- Are there distinct trends between these two major highway crossings?

3 Data

ATR data provides 24/7 traffic volume at specific highway locations across Oregon and has been curated by ODOT's monitoring unit for high reliability. The data consists of vehicle counts over hourly time periods at 12 different highway locations, each consisting of 2 opposite directions (either NB/SB or EB/WB). Some locations have many more years available than others, but in total there are about 3.25 million hours of data recorded over the 24 location and direction combinations.

Preliminary analysis of median values reveals notable trends and location-specific patterns in the data. Table 1 presents an approximation to weekday peak-time and peak-volume based on cubic periodic spline interpolation of hourly medians.

General visible trends seem to be a slightly later peak traffic time in the mornings and location dependent changes in afternoon peak times. The peak volume is lower post-Covid all across the board, with very few exceptions outside of data availability issues.

3.1 Data Quality

Overall, the provided dataset is high quality and quite complete. For a visual representation of the time coverage for each location / direction combination, see figure 33. The only issues with data quality we discovered in our analysis fall into two main categories:

1. Daylight Savings (DST) Inconsistencies

The dataset is recorded in Coordinated Universal Time (UTC), which does not account for Daylight Savings Time (DST) adjustments. This led to occasional timestamp misalignments, particularly at midnight during the spring transitions (April before 2006 and March in later years). These discrepancies suggest that some timestamps may have been incorrectly converted or recorded, potentially affecting time-sensitive traffic analyses. A visual example where this happens is shown in figure 34, where counts from March through August are clearly shifted exactly one hour forward during the early morning hours when the traffic is most consistent.



While these inconsistencies do not significantly impact most analyses, they can introduce errors in identifying precise peak-hour shifts. In particular, Table 1 shows instances where peak times appear to shift by nearly an hour (e.g., Vista Ridge EB 2018, Wilsonville SB 2024, and Interstate Bridge NB 2024). Further adjustments to correct for these errors may be beneficial for future studies.

2. Under-Reported and Malformed Data

Sensor malfunctions and estimation techniques introduced inconsistencies in some data points, as noted in the reliability report provided by the client. Certain locations and time periods exhibited artificially low traffic volumes due to under-reporting or data imputation. One example of this problematic data is I5 bridge SB for the year 2018. During this time there was a construction project so traffic could potentially be lower, but reported (potentially estimated) values are so low that the period is not useful for analysis.

Another example of malformed data is periods of time with many zero values. In the reliability report it is stated that the 2024 Vista Ridge Tunnel data is reliable, but from 10-07 to 12-23 there are 1440 hourly entries with 0 and 1872 non-zero hourly entries.

To enhance data reliability, future datasets could incorporate an additional column flagging estimated or low-confidence entries. This would allow for selective filtering, ensuring that models account for data reliability and providing transparency in trend analysis.

A complete listing of potential issues we found in the analysis can be found in appendix F.

4 Methods

4.1 Generalized Additive Model (GAM)

Generalized Additive Models (GAMs) are an extension of generalized linear models that allow for flexible modeling of non-linear relationships by representing the target variable as an additive combination of smooth functions of the explanatory variables [3, 6, 7]. This flexibility makes GAMs particularly suitable for capturing complex patterns in data without imposing strict parametric forms.

In our analysis, we focus on modeling hourly traffic counts (y) as a function of variables such as the hour of the day (h), month of the year (m), and a binary indicator for the period relative to the COVID-19 pandemic (p), where $p = 0$ denotes pre-pandemic and $p = 1$ denotes post-pandemic periods. The GAM framework allows us to express this relationship as:

$$y = s_1(h) + s_2(ph) + s_3(m) + \epsilon$$

Here, s_1 , s_2 , and s_3 are smooth functions estimated from the data, and ϵ represents the error term. The inclusion of the interaction term ph enables the model to capture differential effects of the pandemic across different hours of the day.

A key advantage of GAMs is their interpretability; the additive components s_i provide insights into how each explanatory variable individually influences the target variable. To estimate these smooth functions, we employ spline functions, which are piecewise polynomials joined smoothly at specific points known as knots. Splines offer greater flexibility than global polynomials and help mitigate issues such as the Runge phenomenon—a problem where high-degree polynomial interpolation can lead to large oscillations at the boundaries of the interpolation interval (classic German reference: [5]).

For implementation, we utilize the `scikit-learn` library's tools for spline fitting. The library provides functionalities to construct spline basis functions and integrate them into a regression framework, facilitating the estimation of the smooth components in the GAM [1]. This approach allows us to model the non-linear relationships in the traffic data effectively while maintaining computational efficiency and interpretability.

After some experimentation we found the periodic cubic splines with knot positions:

- 15 uniform knots for s_1
- knots located at (0, 4, 6, 8, 10, 15, 17, 19, 24) for s_2
- 4 uniform knots for s_3

generalized well for our data. The knots for s_3 were chosen because we are primarily interested in the effect on peak traffic periods, so we choose the knots to be more densely located during these times to give the model more ability to express differences in the interaction while limiting negative effects of over-paramaterizing the model.



4.2 Maximum Mean Discrepancy (MMD)

Traditional statistical methods often rely on parametric assumptions, which may fail to capture the complexity and high-dimensional nature of real-world traffic patterns. To address these limitations, Maximum Mean Discrepancy provides a robust, non-parametric alternative for detecting shifts in traffic distributions. MMD is a kernel-based statistical test that quantifies differences between probability distributions without imposing parametric constraints. By embedding distributions in a reproducing kernel Hilbert space (RKHS), MMD enables a flexible two-sample comparison by mapping traffic data into a high-dimensional space and evaluating the divergence in their mean embeddings. This approach is particularly well-suited for assessing shifts in traffic patterns before and after the pandemic, offering a data-driven framework for capturing complex, non-linear changes in distribution.

4.2.1 Data Structuring Strategy

To ensure a systematic comparison of pre- and post-pandemic traffic patterns, the dataset is structured into two key dimensions:

1. Time segmentation through monthly groups
2. Traffic behavior segmentation by day-type classification

This structure allows for temporal consistency and prevents confounding effects from seasonal or weekday variations.

Monthly Grouping for Temporal Consistency To account for variation in traffic demand across different times of the year, we divide the dataset into five distinct monthly groups instead of using traditional seasonal definitions. This prevents seasonal confounders from affecting our analysis.

- **January–February:** Typically, the lowest traffic counts due to winter weather conditions.
- **March:** Contains Spring break week, influencing travel demand.
- **April–May & September–October:** Stable school/work schedules, making them strong baseline months for comparison.
- **June–July–August:** Summer months with a different traffic pattern, characterized by fewer school-related trips and more recreational travel.
- **November–December:** Affected by winter conditions and holiday travel, introducing variability in demand.

This grouping strategy ensures that traffic trends are analyzed under similar external conditions before and after the pandemic.

Day-Type Classification for Behavioral Segmentation Traffic patterns vary based on the day of the week, requiring further classification to capture meaningful behavioral trends. We categorize traffic data into four primary day types:

- **Monday & Friday:** Often show distinct travel behaviors, with Friday frequently exhibiting early departures and heavier congestion.
- **Tuesday–Wednesday–Thursday (T/W/T):** Assumed to have relatively consistent traffic patterns, grouped together for statistical efficiency.
- **Saturday & Sunday (Weekends):** Typically show non-commuting, recreational travel behaviors, distinct from weekday traffic trends.

By applying this dual-layered structuring approach, we ensure that comparisons between pre- and post-pandemic traffic data are made under logically consistent conditions, minimizing external sources of variation.



4.2.2 MMD Model for Traffic Pattern Comparison

After structuring the dataset, we use the Maximum Mean Discrepancy statistic to measure shifts in traffic distributions. Given two sets of traffic data:

$$X = \{x_{i,j} \mid i = 1, \dots, N; j = 1, \dots, 24\}, \quad Y = \{y_{i,j} \mid i = 1, \dots, M; j = 1, \dots, 24\}$$

where:

- $x_{i,j}$ and $y_{i,j}$ are traffic counts at hour j on day i for pre- and post-pandemic periods, respectively.
- X and Y represent hourly traffic counts for each day type within a specific monthly group.

The MMD statistic is calculated as:

$$\text{MMD}^2(X, Y) = \mathbb{E}_{x,x'}[k(x, x')] + \mathbb{E}_{y,y'}[k(y, y')] - 2\mathbb{E}_{x,y}[k(x, y)]$$

where $k(\cdot, \cdot)$ is a Gaussian kernel function. This provides a non-parametric measure of distributional shift in traffic behavior.

4.2.3 Statistical Inference Using Permutation Testing

Statistical significance is evaluated using a permutation test. Traffic data is pooled and randomly reassigned into new pre- and post-pandemic groups, maintaining the same sample sizes as the original datasets. The MMD statistic is then recomputed for each permutation, generating an empirical distribution under the null hypothesis. The p-value is estimated as the proportion of permuted MMD values that exceed or equal the observed MMD, ensuring statistical robustness in assessing distributional differences.

The empirical p-value is calculated as:

$$p = \frac{\sum_{b=1}^B \mathbb{1}(\text{MMD}_b^2 \geq \text{MMD}_{\text{obs}}^2)}{B}$$

where B is the number of permutations, $\text{MMD}_{\text{obs}}^2$ is the observed MMD statistic, and $\mathbb{1}(\cdot)$ is an indicator function.

This methodology allows us to determine whether the observed shift in traffic patterns is statistically significant or could have arisen due to random fluctuations.

5 Results

5.1 Generalized Additive Model

Visualizations of the GAMs for each ATR location are found in the appendix in section D. The primary findings can be classified into three main categories:

5.1.1 Weekday ‘Morning Rush’

Examples of ATR locations with strong weekday ‘morning rush’ characteristics (usually heading into the metro area) are Vista Ridge Tunnel EB, Interstate Bridges SB, and Iowa Street NB. The general visible change in the morning rush at these locations is a reduction in peak volume and often a slight shift to later in the day. This reduction in peak volume is seen as a ‘flatter’ morning peak where the most intense traffic in the morning is spread out over a longer period of time. A clear outlier to these patterns is Wilsonville NB which has almost identical morning traffic patterns.

The observed reduction in morning peak volume could be attributed to decreased demand, increased congestion, or a combination of both; however, this dataset alone does not allow for a definitive conclusion. Only for some locations is this reduction compensated by increases in volume later throughout the mid-day.

5.1.2 Weekday ‘Afternoon Rush’

ATR locations with ‘afternoon rush’ characteristics are often the directional pairs to those with ‘morning rush’ ones. The afternoon rush structure is much different, generally less concentrated and spread out over a longer period. The changes in afternoon traffic are less pronounced than the morning ones and the peak volume times are mostly consistent before and after Covid-19. The most obvious change is a slight earlier shift in when the traffic volume begins to taper off and a slight reduction in peak volume counts.



5.1.3 Weekend

The visible changes in weekend patterns is much less pronounced than the weekday ones, which is confirmed quantitatively with the statistically testing using MMD discussed in section 5.2. Although less pronounced, there is still a visible reduction in volume at most locations.

5.2 Maximum Mean Discrepancy

To evaluate the statistical significance of traffic pattern shifts, we applied the Maximum Mean Discrepancy test along with a permutation test. The results are visualized in heatmaps, where each cell represents the p-value associated with traffic distribution changes for a specific combination of day type and month group.

5.2.1 Key Observations and Implications

The findings indicate that the most notable adjustments in traffic patterns occur midweek (Tuesday–Thursday), where consistent shifts are observed across multiple month groups. Mondays and Fridays display a more variable pattern, with changes appearing more context-dependent. Weekend traffic remains relatively stable, with some month-specific variations.

These observations suggest that changes in travel behavior post-pandemic are more evident in structured weekday commuting patterns rather than in discretionary weekend travel. Additionally, differences between I-5 and I-205 may be influenced by location-specific factors such as infrastructure, urbanization, or regional travel demand.

5.2.2 Summary of Observations

Based on the analysis, the following key patterns emerge:

1. Tuesday–Thursday traffic shows consistent changes across all month groups, suggesting adjustments in midweek commuting behavior.
2. Monday and Friday display variability in changes, with certain months exhibiting stability and others showing shifts.
3. Weekend travel remains largely stable, particularly on I-5, while I-205 exhibits some month-dependent variations.
4. I-5 demonstrates broader consistency in traffic changes, while I-205 presents greater fluctuations across different month groups.

These findings highlight how post-pandemic travel behavior varies across different day types, with midweek commuting experiencing the most pronounced shifts, while weekend travel remains relatively stable.

6 Conclusions

This report analyzed traffic patterns on key highway crossings in Portland, focusing on pre- and post-pandemic trends using ATR data. Our investigation employed both Generalized Additive Models (GAM) and Maximum Mean Discrepancy (MMD) testing to capture changes in peak traffic times, overall volume, and distributional shifts across different time periods and day classifications. The following key insights emerged from our study:

1. **Morning Rush Flattening:** Peak morning traffic volumes have decreased post-pandemic, often accompanied by a later shift in peak times. This suggests changes in commuter behavior, possibly due to increased telecommuting or flexible work schedules.
2. **Afternoon Stability:** Unlike the morning rush, the afternoon peak period remains relatively stable in timing and structure. However, a slight reduction in volume is observed, and the peak tapering off occurs earlier in some locations.
3. **Weekend Traffic Resilience:** Weekend traffic patterns have shown less significant shifts compared to weekdays. Volume reductions are present but less pronounced, indicating that discretionary travel has been less affected by post-pandemic behavioral changes.
4. **Variability by Location:** While general trends hold across the studied locations, site-specific differences highlight the impact of infrastructure, urban development, and local travel needs. The Interstate Bridge and Glenn Jackson Bridge exhibit distinct post-pandemic traffic shifts, suggesting that regional travel demand and alternate route usage may have evolved.



6.1 Code Availability

5. **Data Quality Considerations:** Our analysis identified some inconsistencies in the ATR dataset, including potential daylight savings time discrepancies and under-reported or estimated data. Addressing these concerns by incorporating reliability indicators in the dataset could improve future modeling efforts.

These findings align with previous studies that have documented significant alterations in urban mobility and emissions during the pandemic [4]. Our findings suggest that long-term shifts in commuting behavior post-pandemic may continue to impact transportation planning. The observed flattening of morning peaks and reduction in overall volume could inform future infrastructure development, congestion management strategies, and transit planning. Further studies incorporating traffic flow models and real-time congestion data could provide deeper insights into the mechanisms driving these changes and their long-term implications for urban mobility.

6.1 Code Availability

The code used for analysis and generating this report can all be found in the GitHub repo and is free to use and distribute under the MIT license.

References

- [1] Polynomial and spline interpolation — scikit-learn 1.6.1 documentation. Accessed: 2025-03-18.
- [2] Bill Eisele and David Schrank. Now you see it; now you don't: The traffic hiatus from covid-19 was historic and brief. *When Transportation Becomes Personal*, 57(3):8–10, 2020.
- [3] Trevor Hastie and Robert Tibshirani. Generalized Additive Models. *Statistical Science*, 1(3):297 – 310, 1986.
- [4] Duo Li and Joan Lasenby. Investigating impacts of covid-19 on urban mobility and emissions. *Cities*, 135:104246, 2023.
- [5] Carl Runge. Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten. *Zeitschrift für Mathematik und Physik*, 46:224–243, 1901.
- [6] Sujeet Sahoo, Changati Sudha, and Prasanta kumar Bhuyan. Generalized additive models to assess truck driver's perceived los of signalized intersections. *Transportation Research Procedia*, 82:2232–2244, 2025. World Conference on Transport Research - WCTR 2023 Montreal 17-21 July 2023.
- [7] Yuanchang Xie and Yunlong Zhang. Crash frequency analysis with generalized additive models. *Transportation Research Record*, 2061(1):39–45, 2008.



A Median Peak Time and Volumes

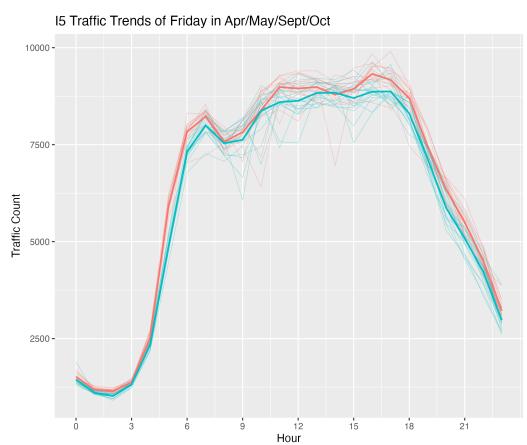
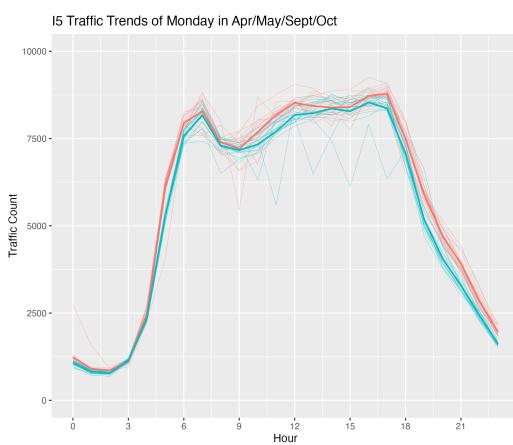
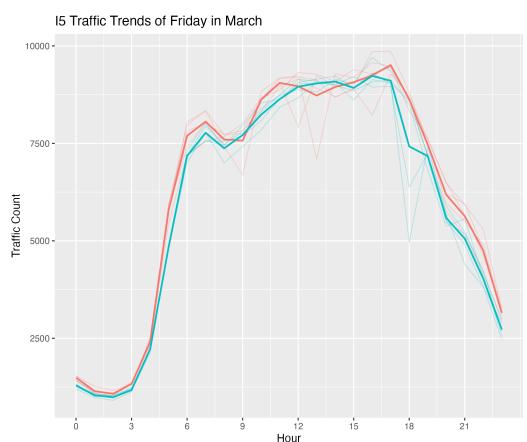
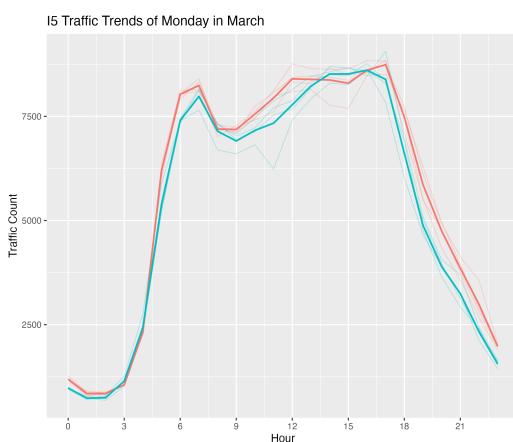
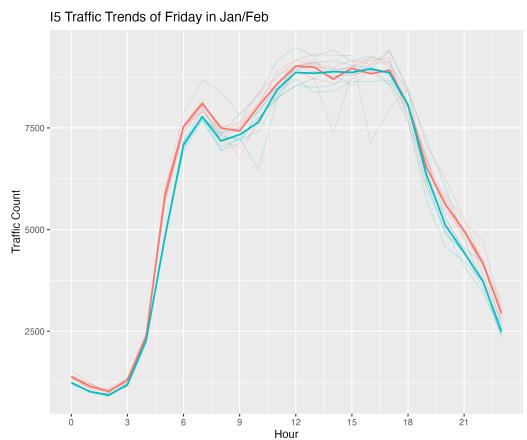
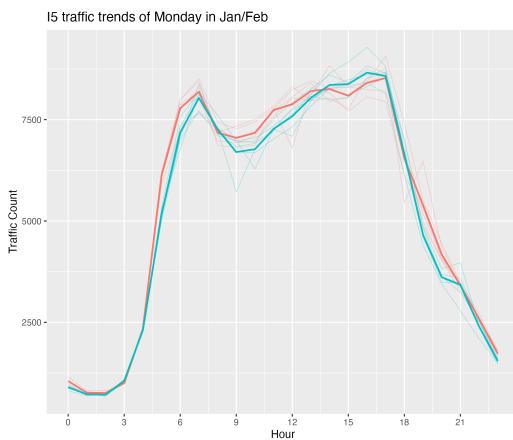
Table 1: Weekday Peak Time Table

Location Info			Median Peak Time				Median Peak Volume			
ID	Name	Dir	2018	2019	2023	2024	2018	2019	2023	2024
34007	North Plains	EB	06:35	06:38	06:49	06:52	918	939	797	811
		WB	16:06	16:03	15:49	15:55	1145	1166	1121	1167
26022	Lents	SB	15:09	16:32	16:29	16:32	5521	5542	5250	5187
		NB	06:23	06:26	07:01	06:58	5055	5060	5152	5111
26002	Vista Ridge Tunne	EB	06:38	07:38	07:21	07:27	5047	4953	4695	4602
		WB	07:15	07:18	15:34	07:24	5623	5612	5169	5173
26001	Troutdale	WB	14:34	14:19	14:51	14:48	1229	1311	1266	1294
		EB	10:26	10:40	10:11	10:08	1286	1368	1279	1331
26004	Interstate Bridge	SB	06:29	05:34	06:09	06:06	1908	5609	5224	4939
		NB	16:32	16:09	16:06	16:44	4619	4805	4625	4380
26028	Fairview	EB	16:29	16:35	16:18	16:29	4493	4532	4391	4369
		WB	06:29	06:35	06:41	06:35	4034	4064	3850	3761
26024	Glenn Jackson Bridge	NB	15:06	15:06	15:06	14:45	7168	7137	6728	6614
		SB	06:15	06:18	06:44	06:20	7411	7354	6188	6159
34010	Beaverton-Bethany	EB	16:03	16:03	06:49	06:49	4521	4464	4191	4211
		WB	07:15	07:15	16:21	16:24	4975	5152	4500	4531
3011	Wilsonville	SB	14:57	14:57	14:37	15:40	3778	3791	3625	3640
		NB	05:57	05:54	06:06	06:03	3832	3837	3756	3871
26014	Hoyt	WB	05:40	05:37	06:03	06:03	5850	5809	5344	5363
		EB	13:56	13:51	13:45	13:42	5587	5569	5283	5247
3016	Stafford	NB	13:59	14:05	14:05	14:05	3267	3411	3186	3283
		SB	06:23	06:23	06:55	06:55	3353	3367	3222	3309
26016	Iowa Street	SB	16:32	16:29	16:29	16:26	5374	5339	4786	4758
		NB	06:29	06:32	07:12	07:09	5687	5671	5303	5231

Hourly median traffic volume for each entry (location, direction, year) for weekdays is interpolated with a cubic spline to find approximate peak traffic time.

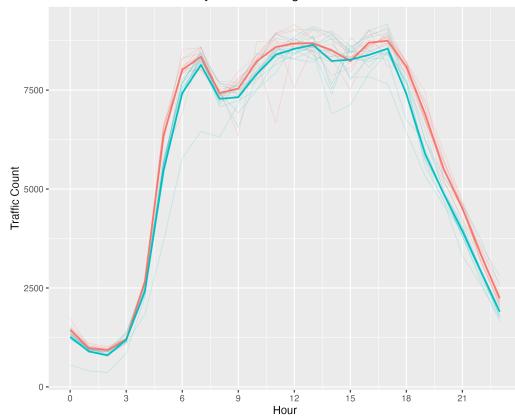


B MMD Groupings

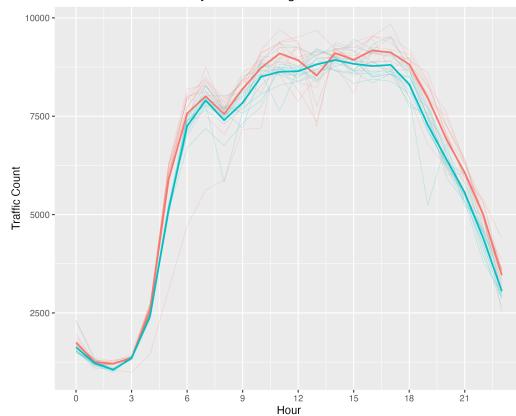




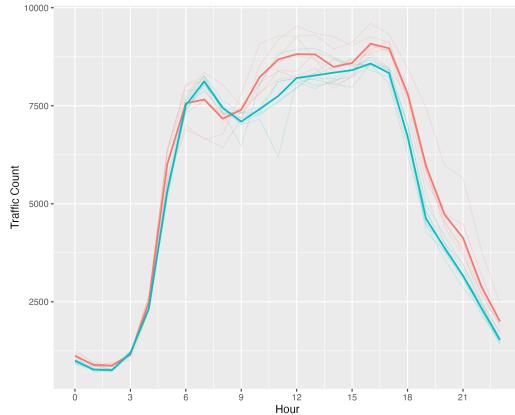
I5 Traffic Trends of Monday in Jun/Jul/Aug



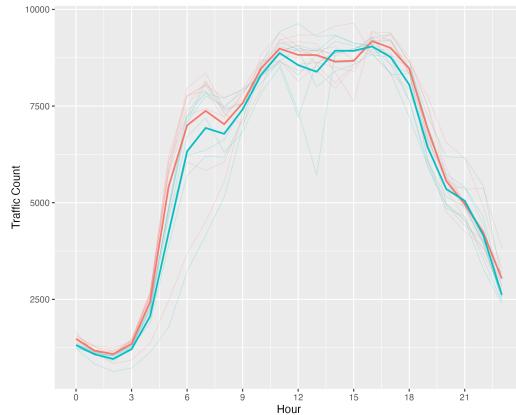
I5 Traffic Trends of Friday in Jun/Jul/Aug



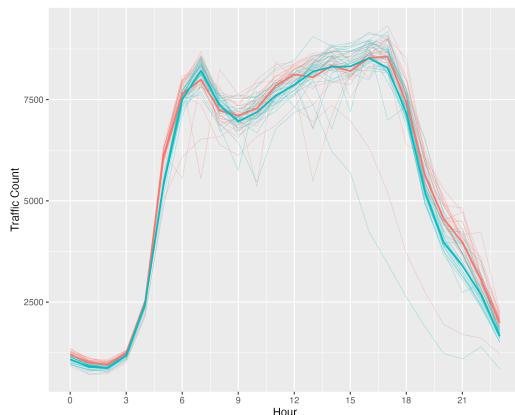
I5 Traffic Trends of Monday in Nov/Dec



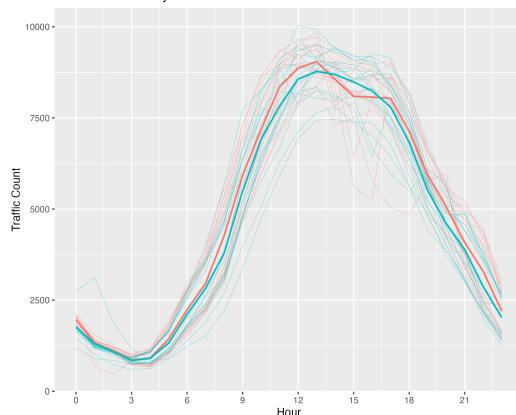
I5 Traffic Trends of Friday in Nov/Dec



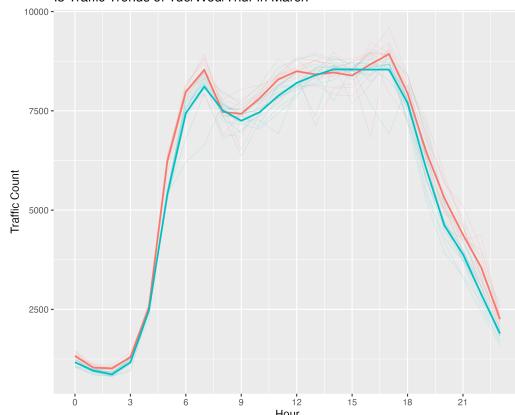
I5 Traffic Trends of Tue/Wed/Thur in Jan/Feb



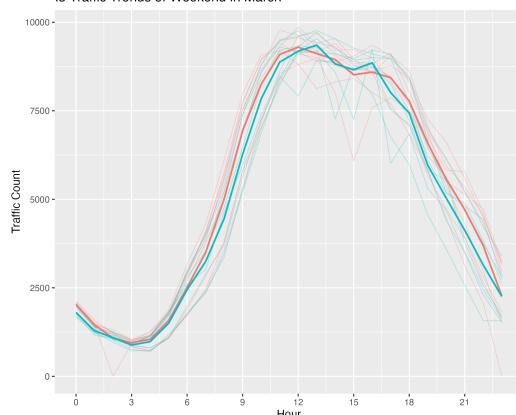
I5 Traffic Trends by Weekend Jan/Feb

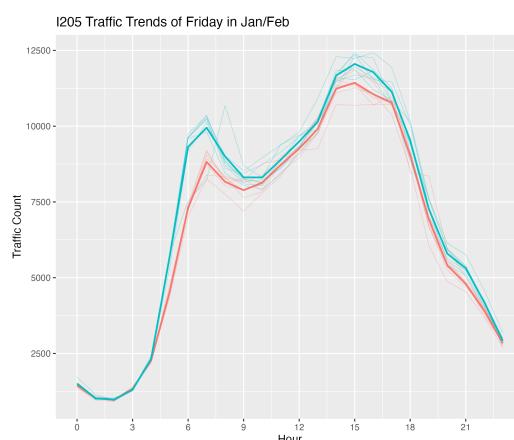
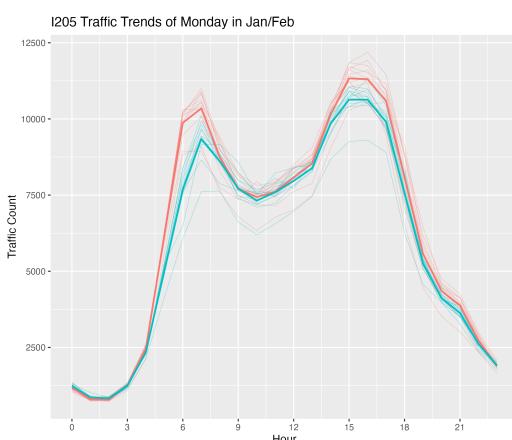
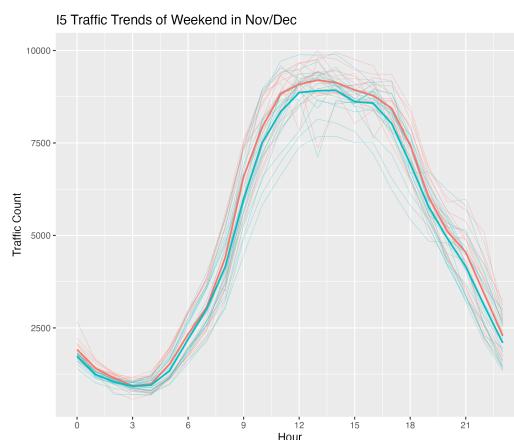
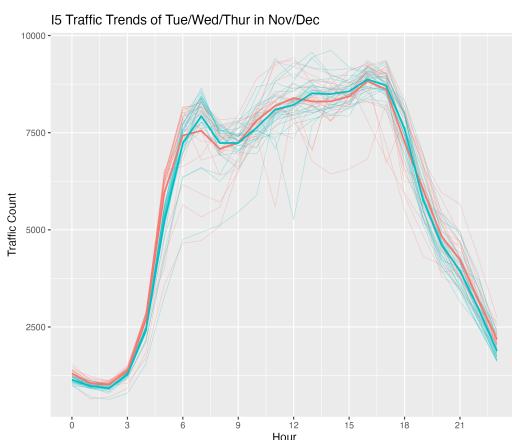
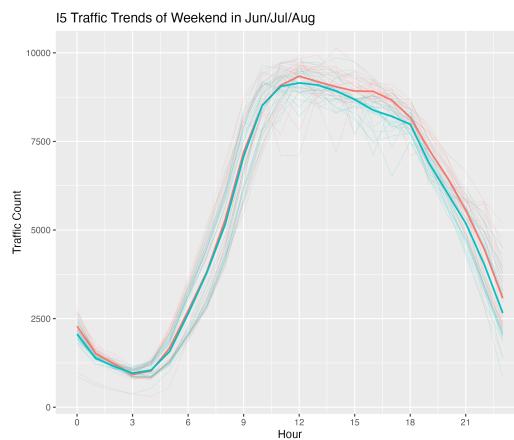
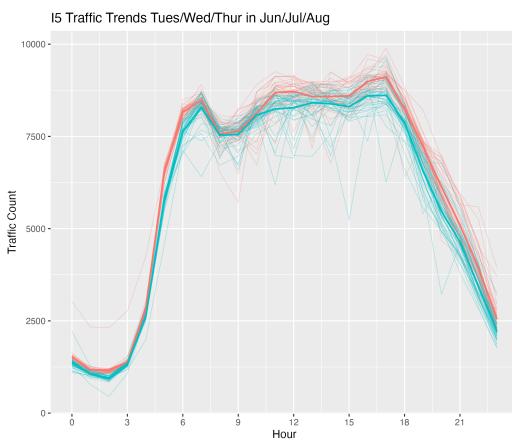
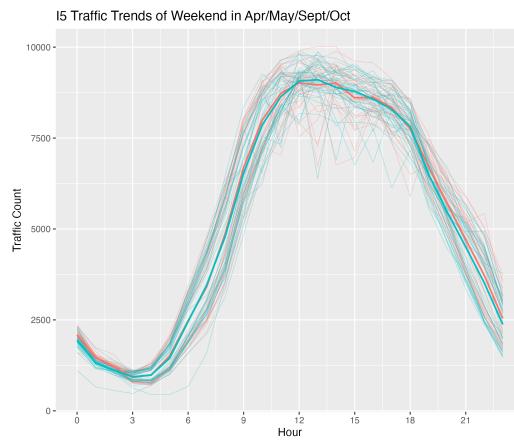
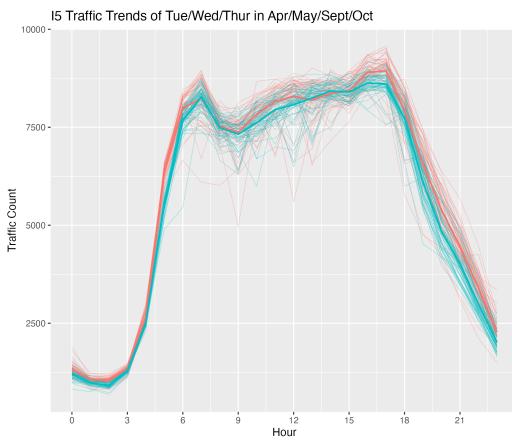


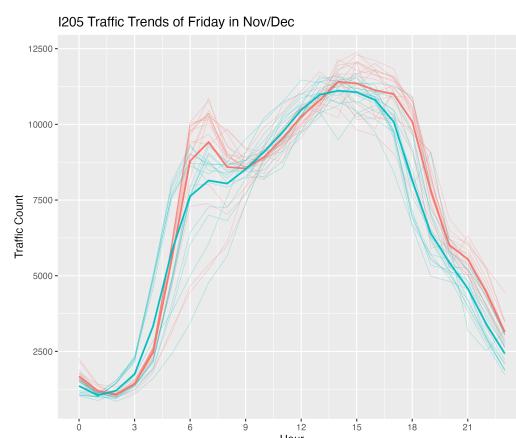
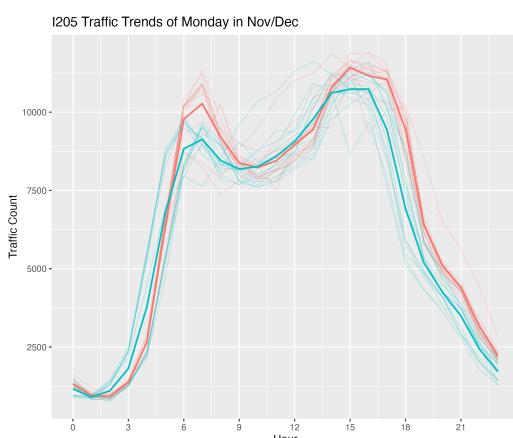
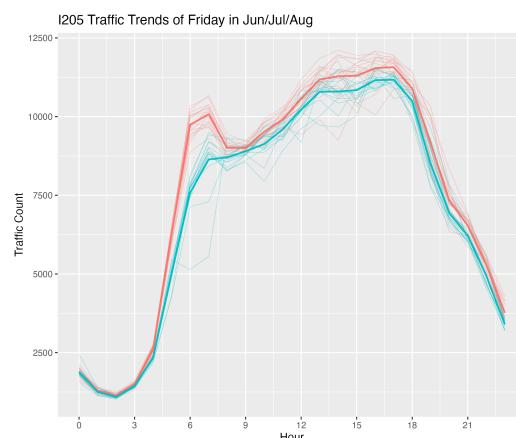
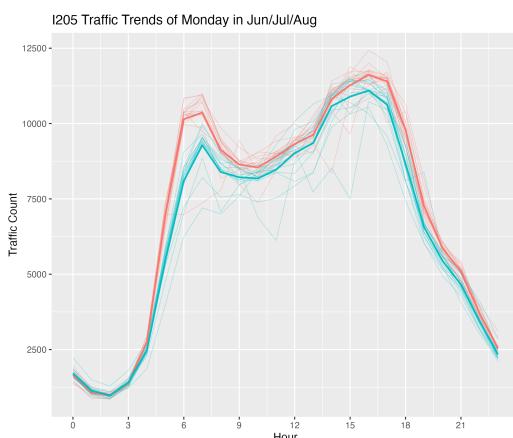
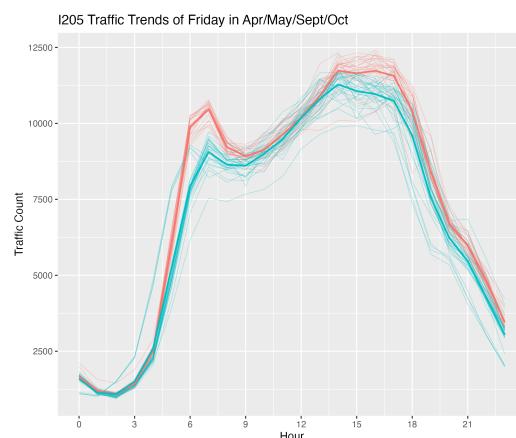
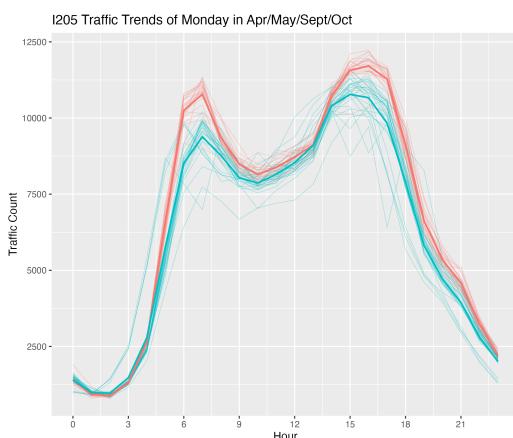
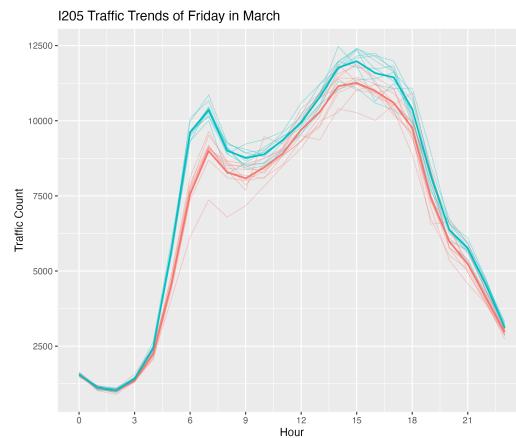
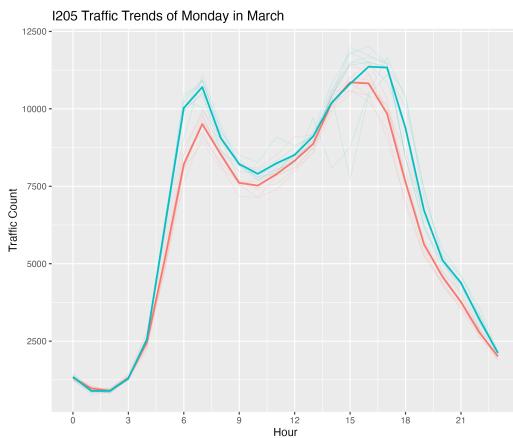
I5 Traffic Trends of Tue/Wed/Thur in March

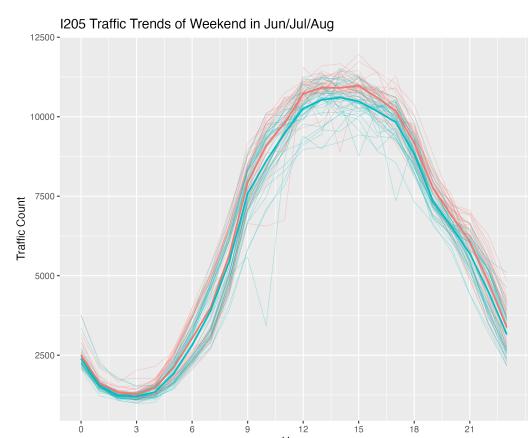
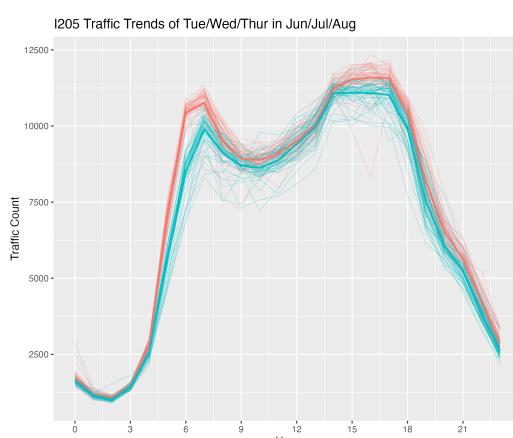
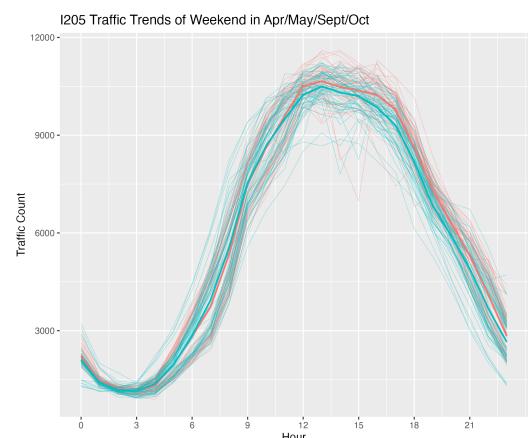
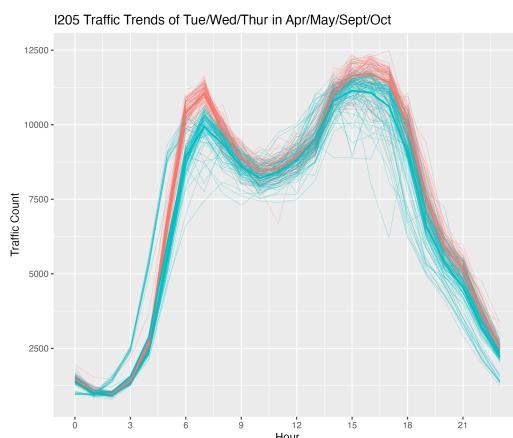
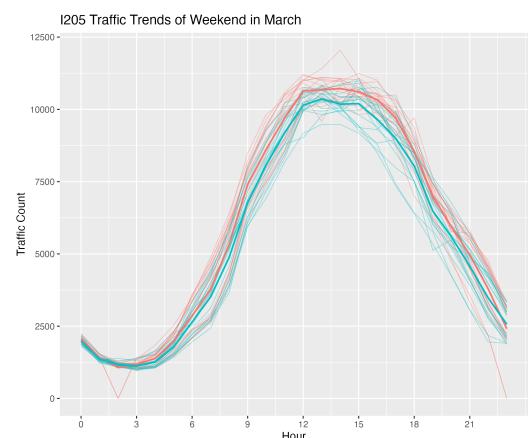
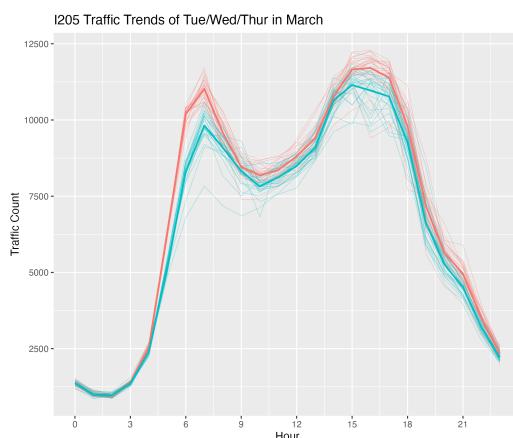
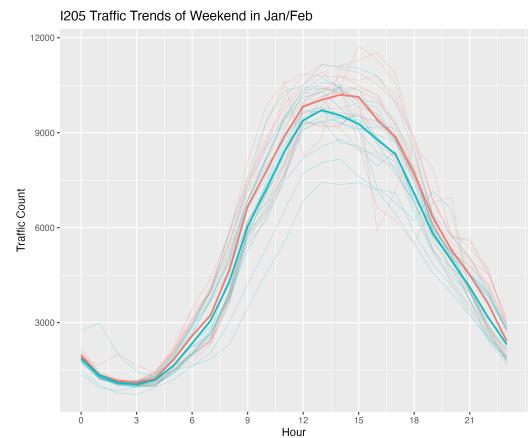
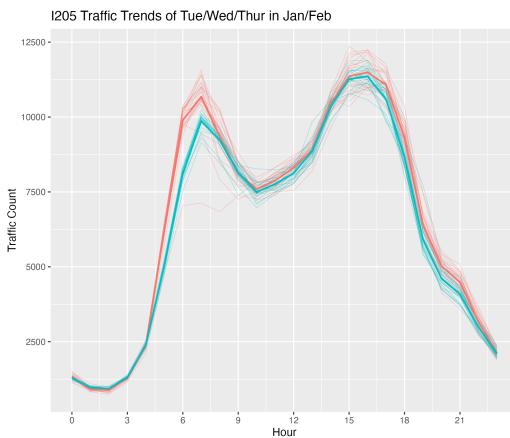


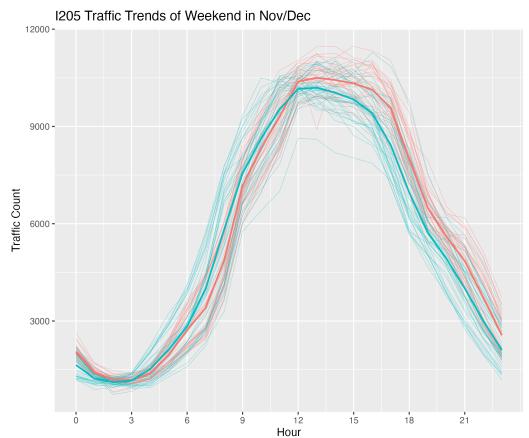
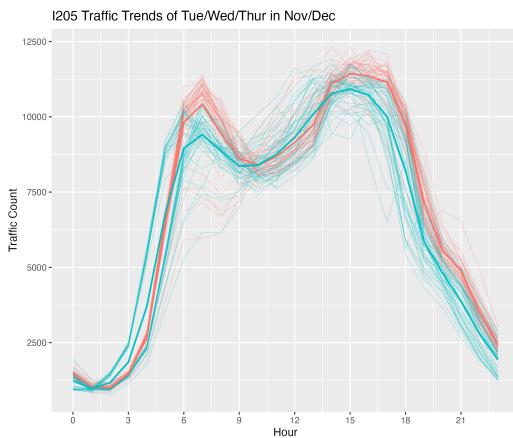
I5 Traffic Trends of Weekend in March



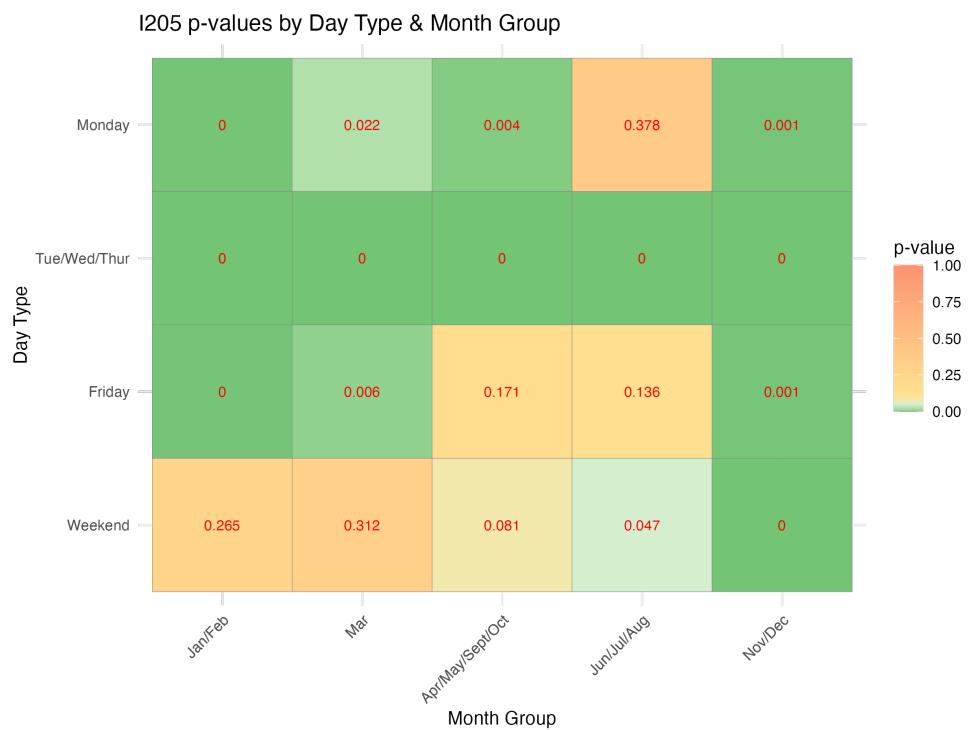






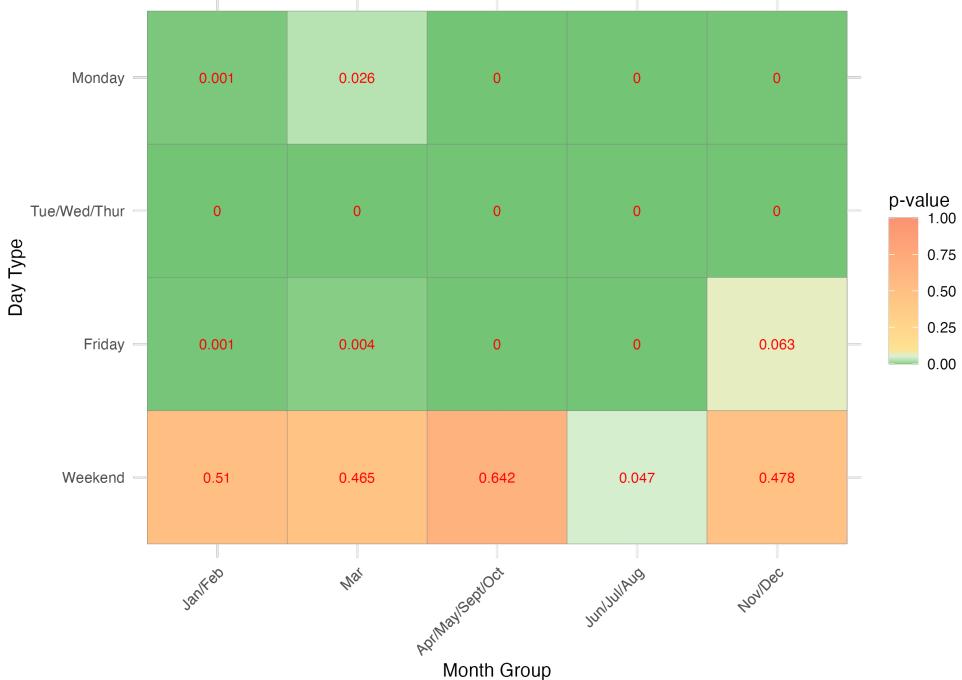


C MMD P-Values

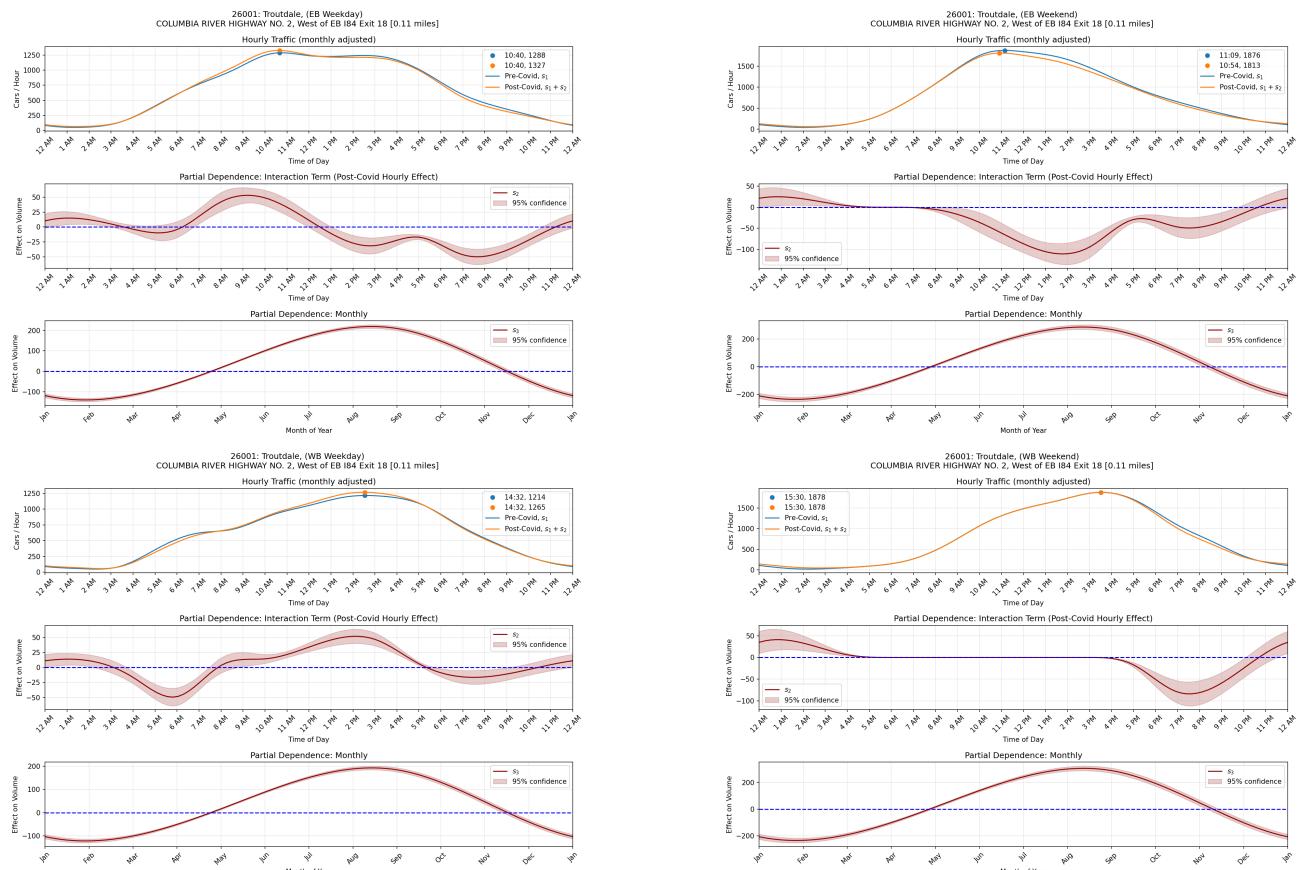


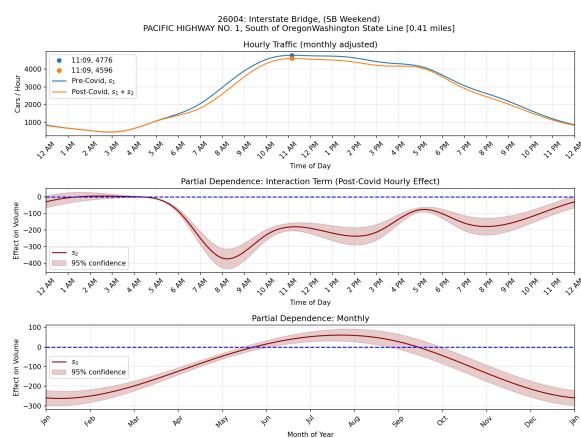
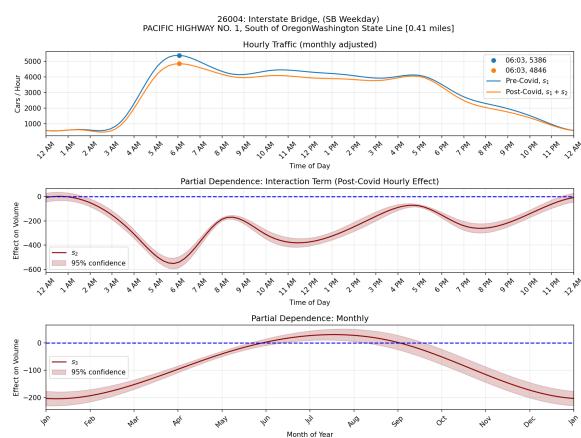
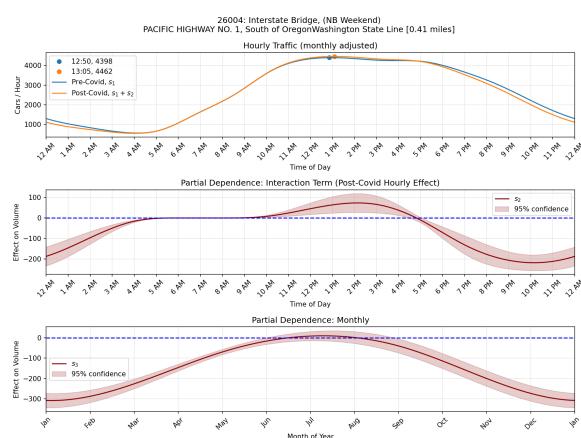
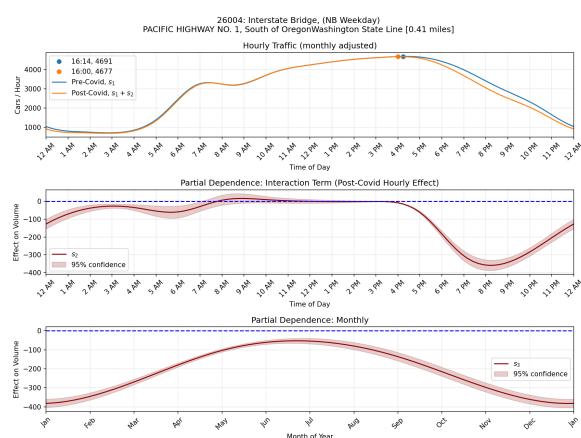
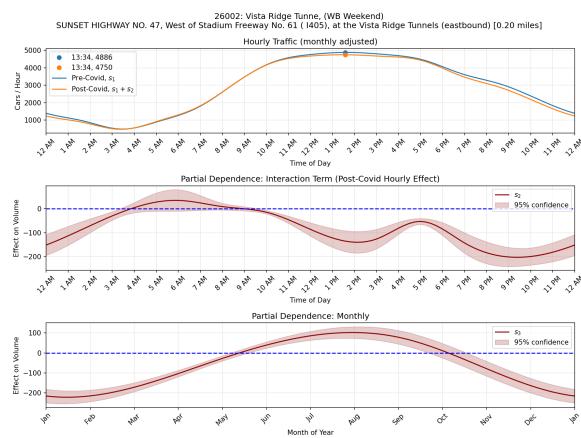
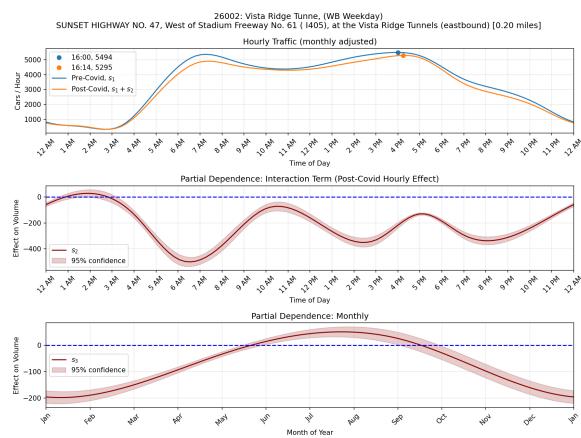
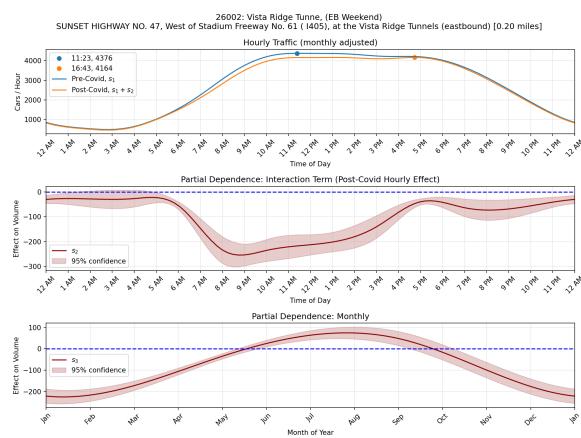
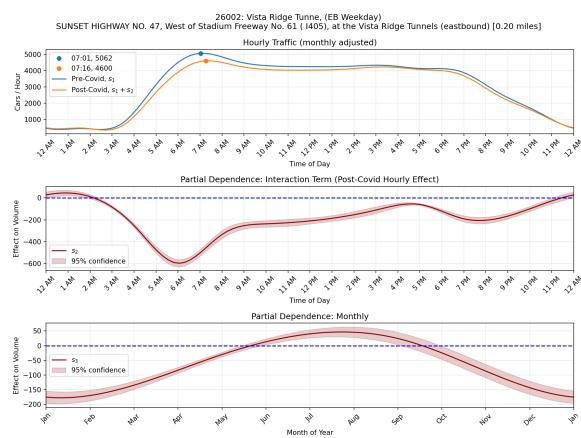


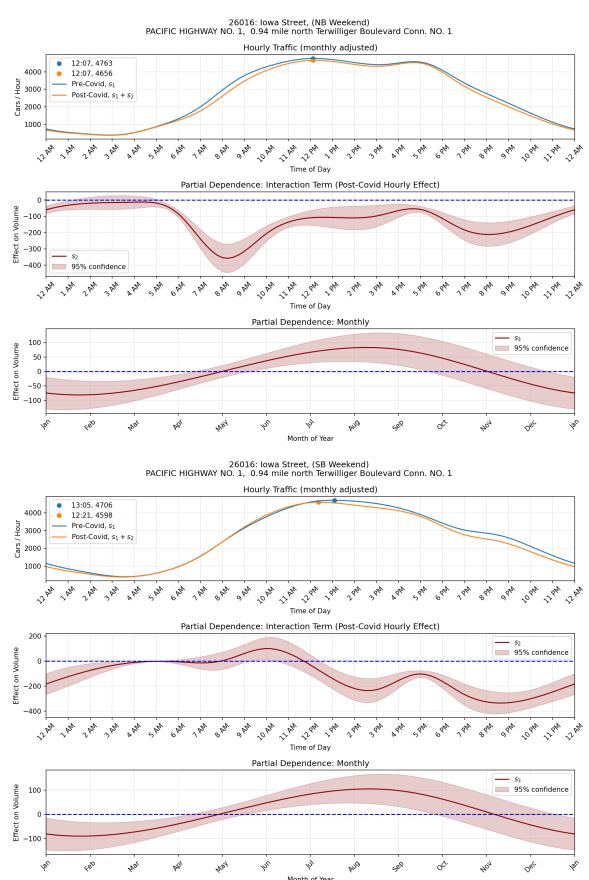
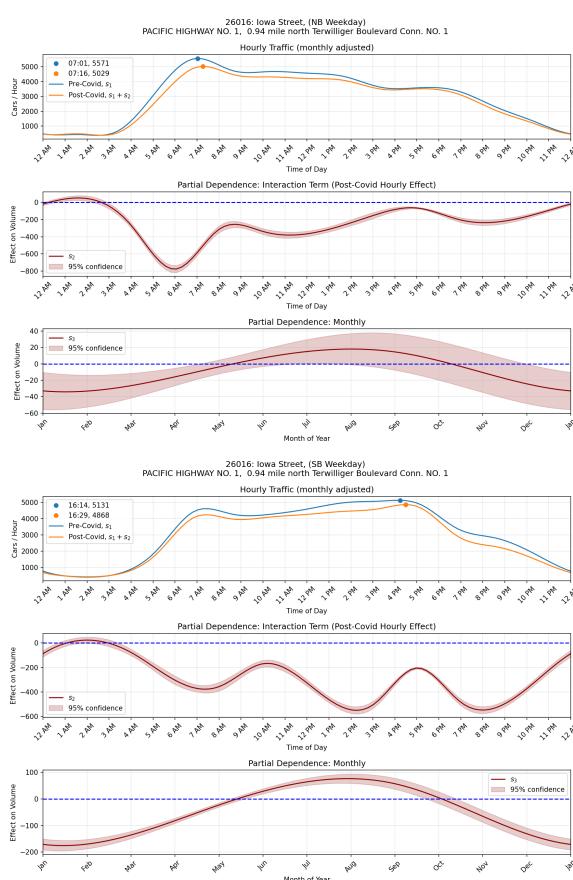
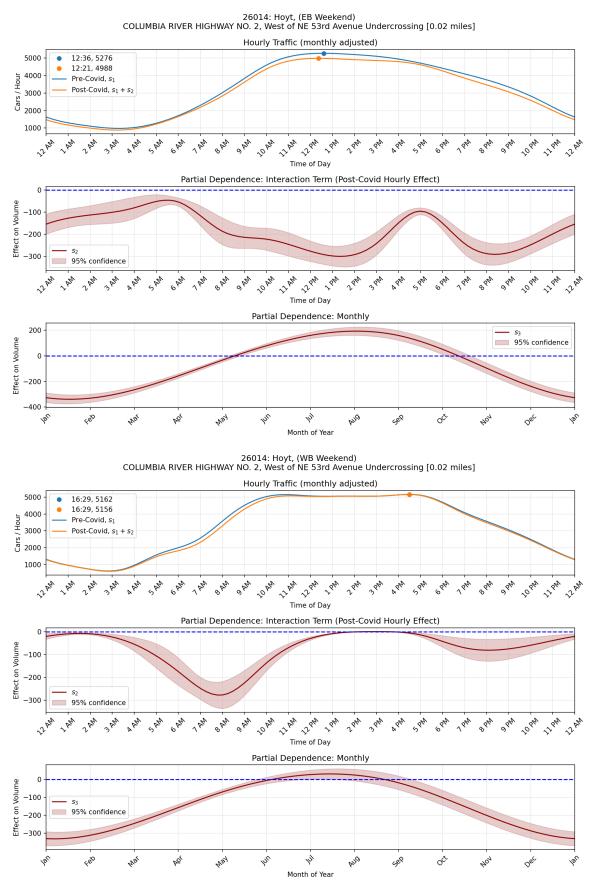
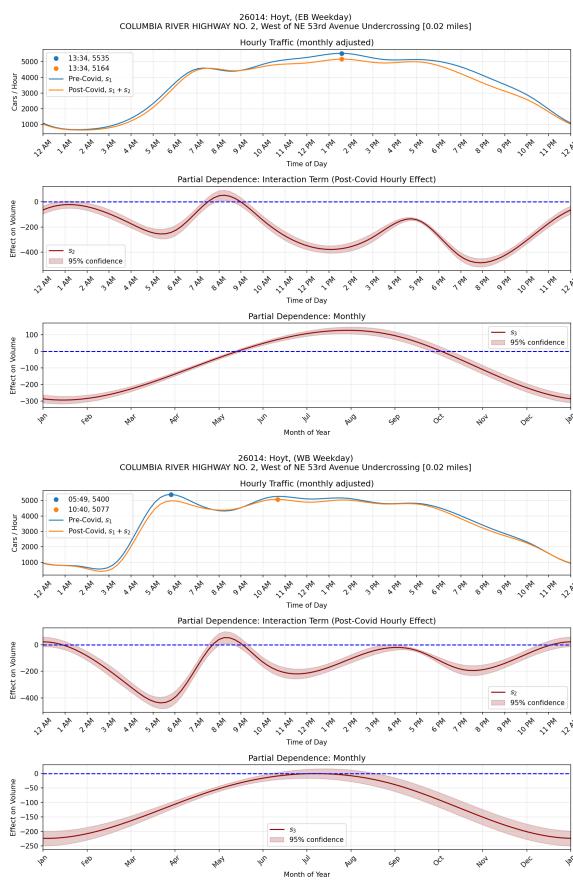
I5 of p-values by Day Type & Month Group

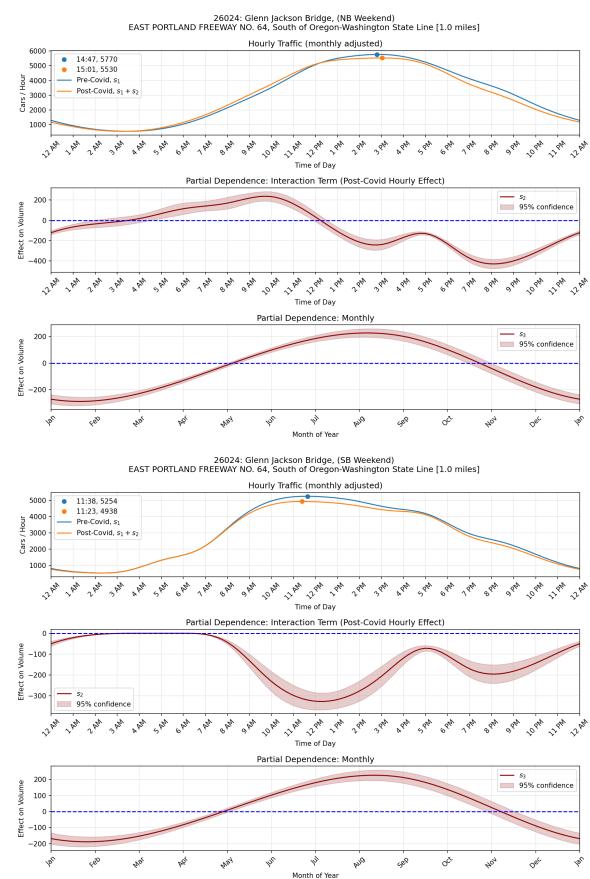
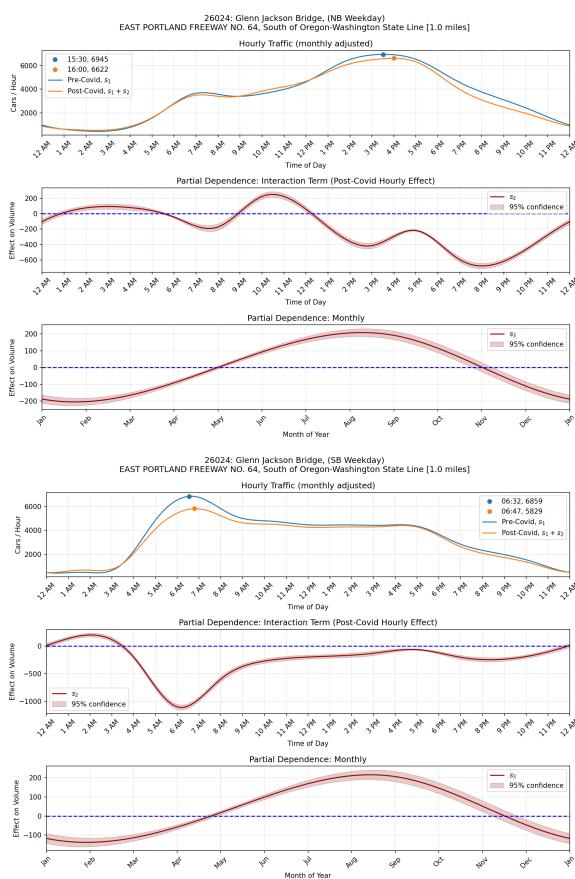
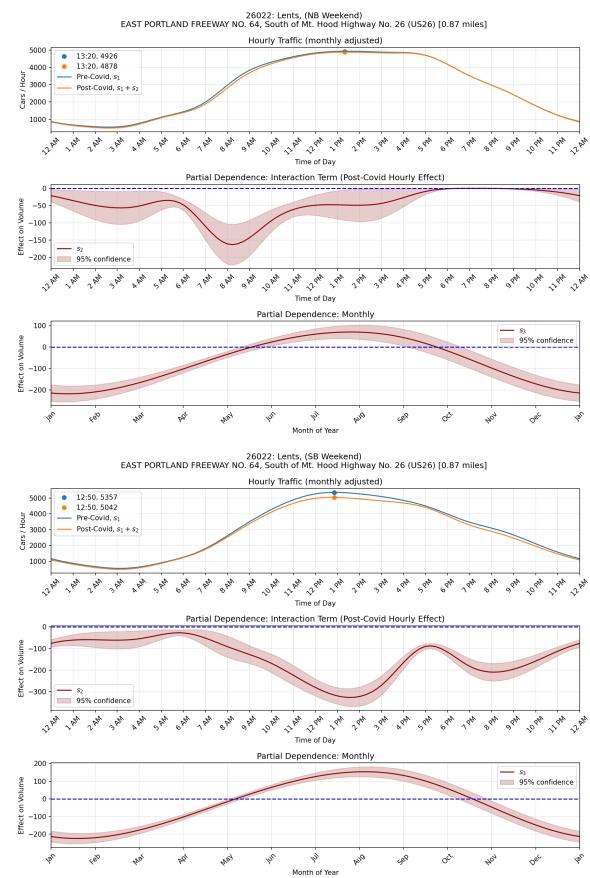
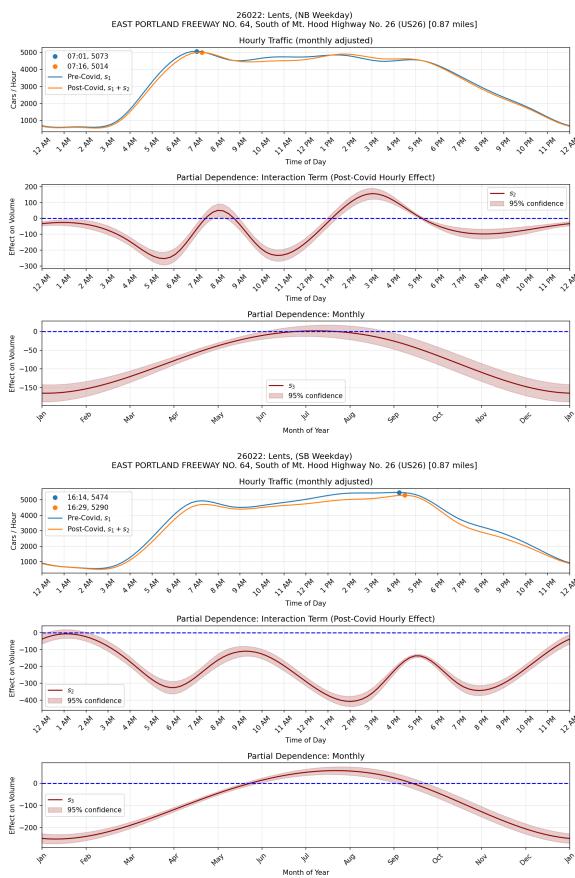


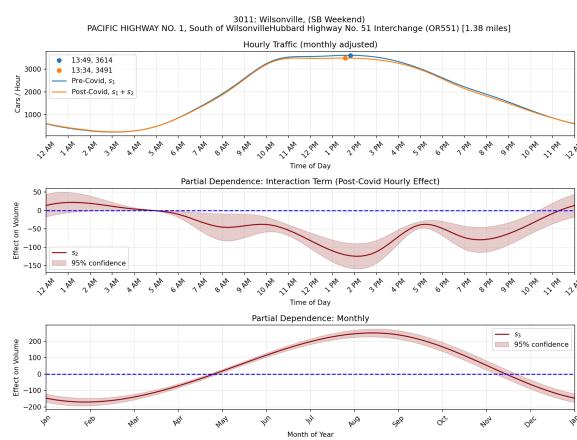
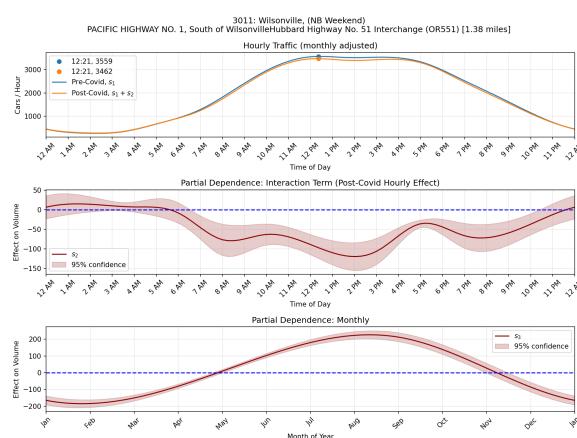
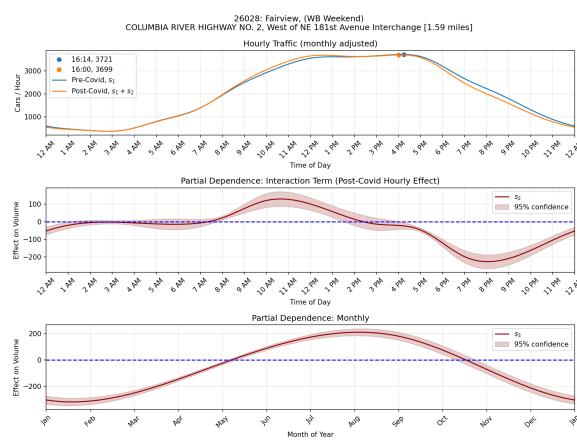
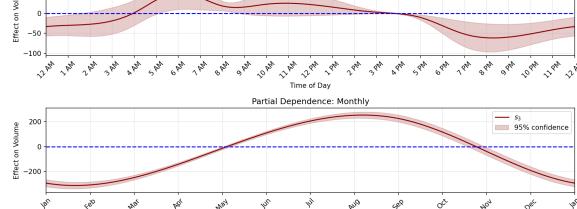
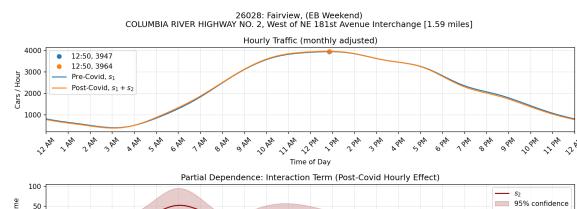
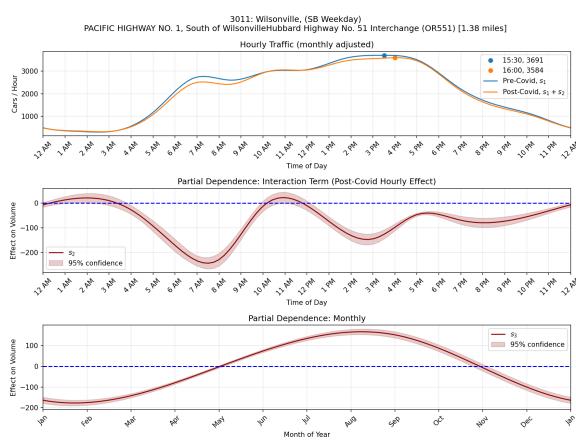
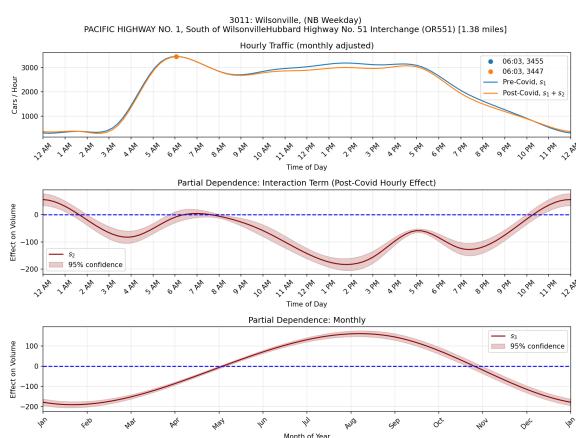
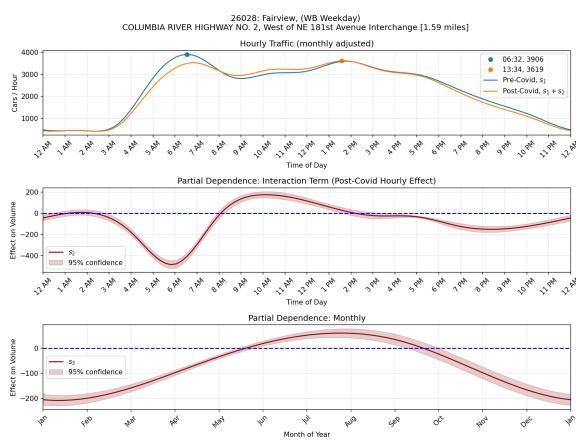
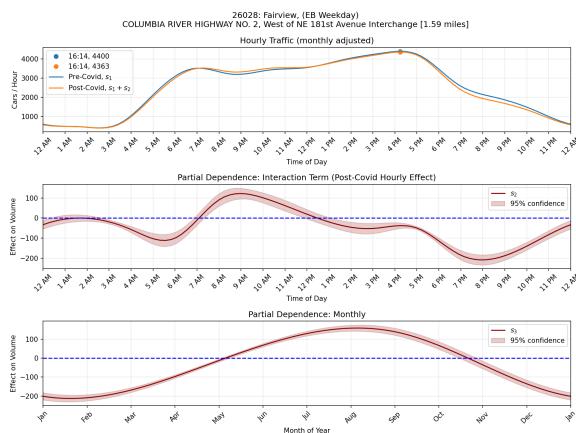
D GAM Figures

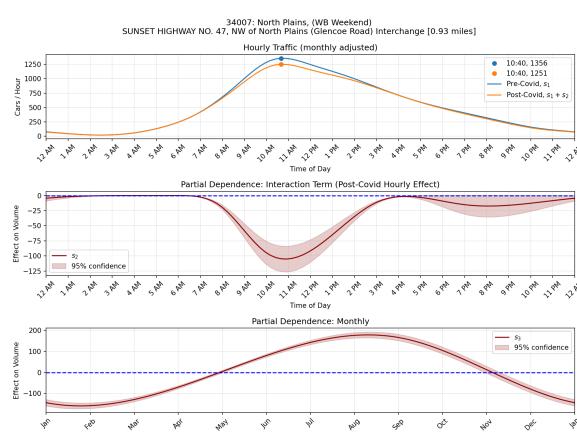
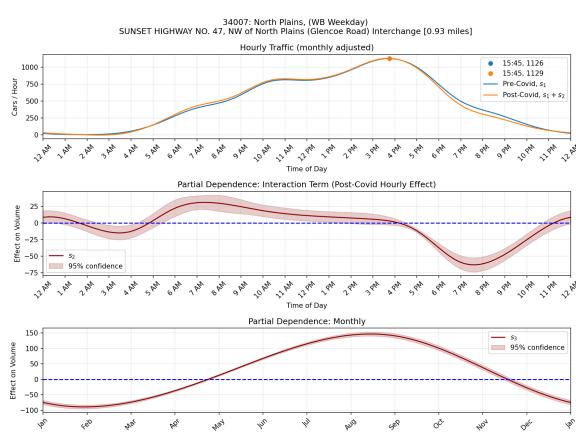
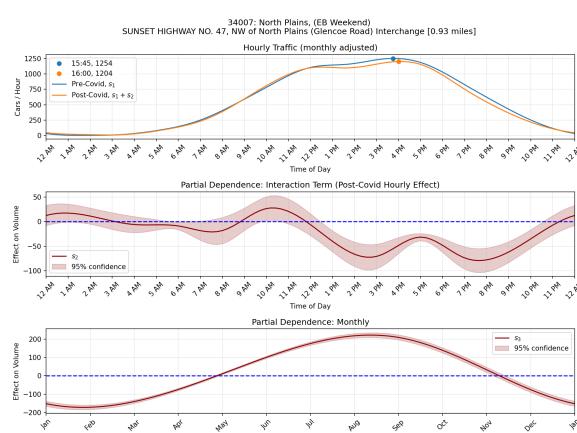
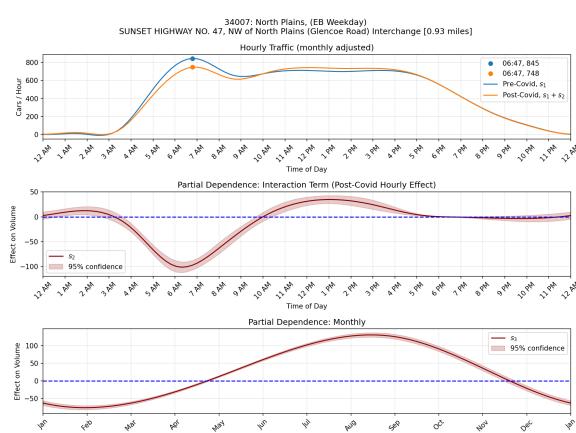
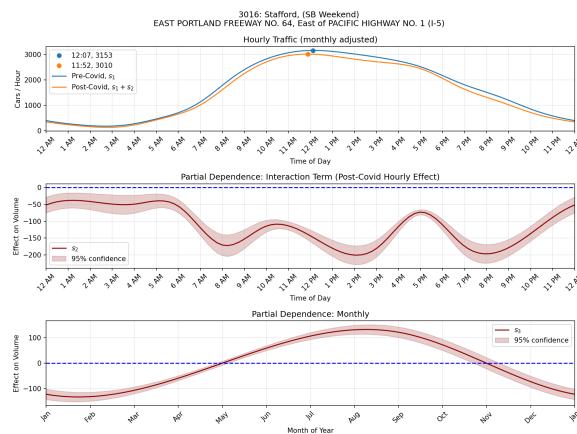
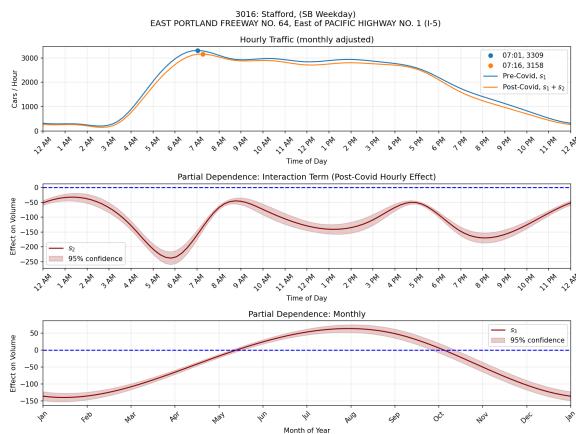
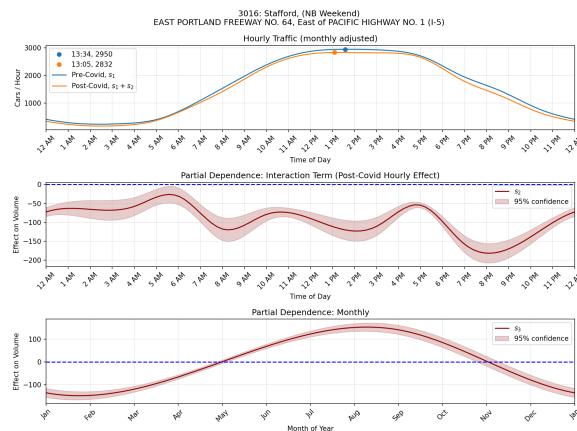
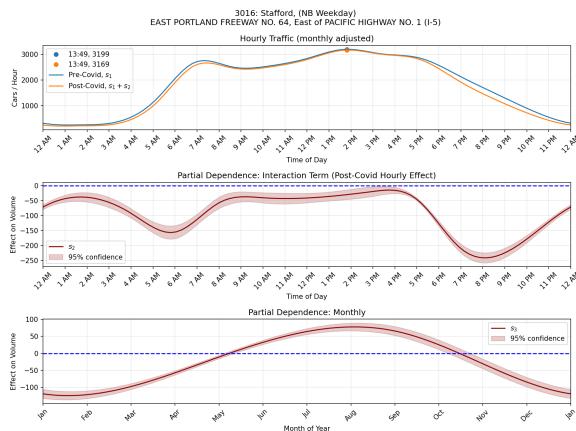


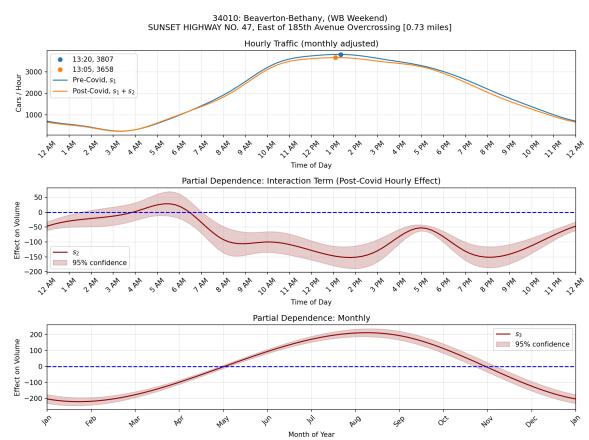
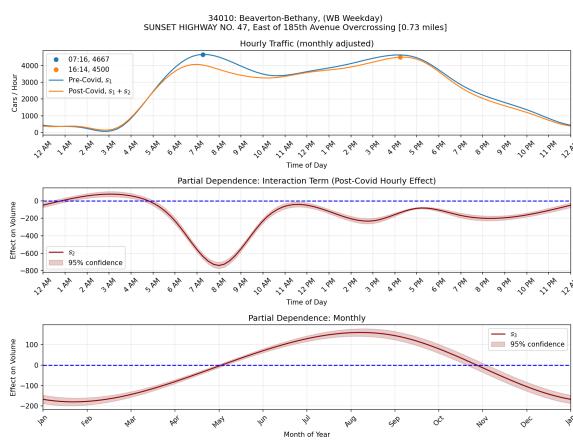
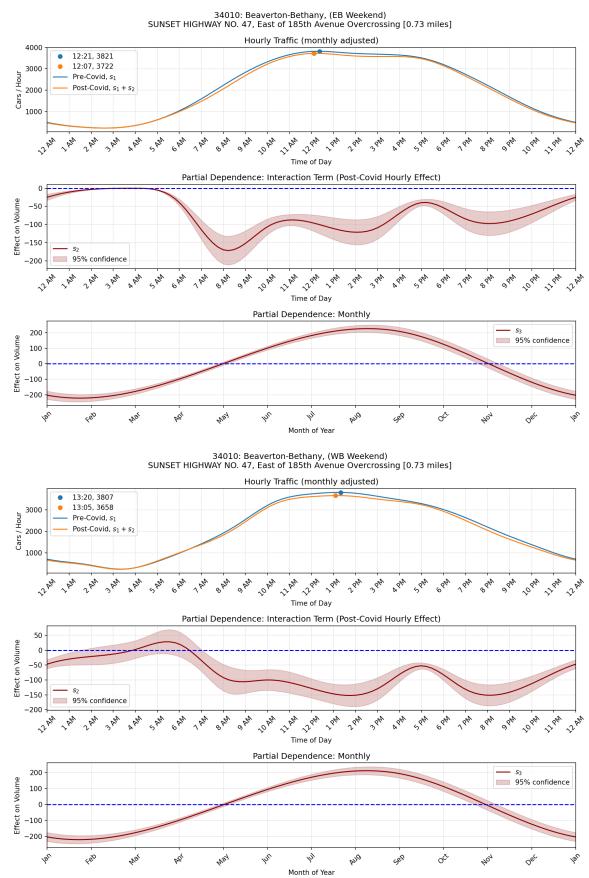
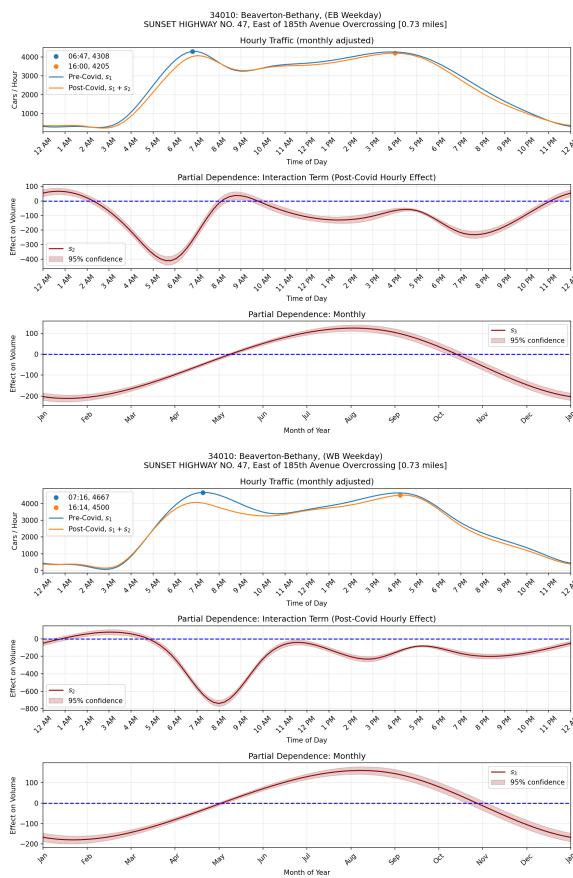














E Data Availability

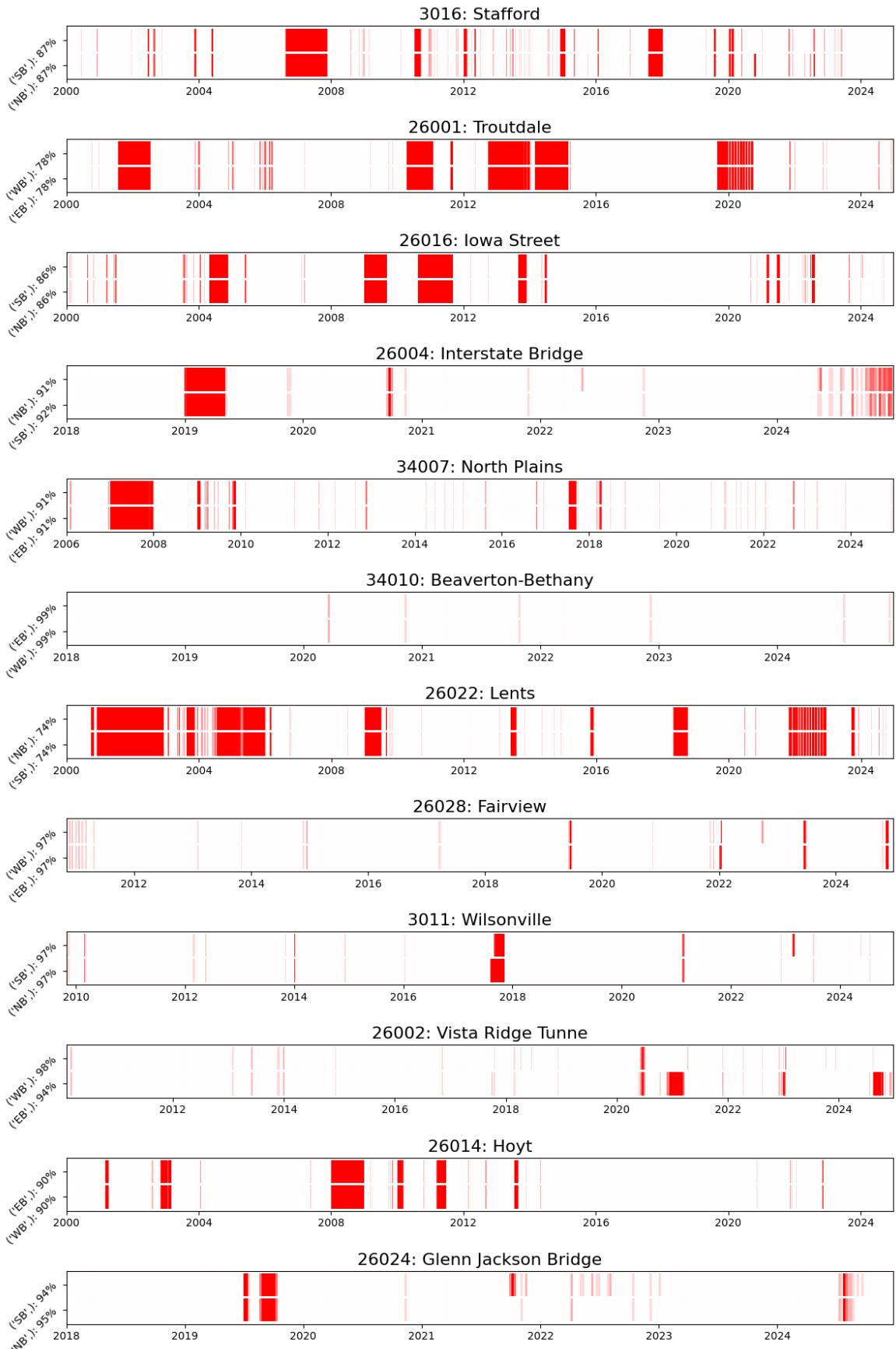


Figure 33: Red indicates missing data, opacity indicates how much of that week is missing. This doesn't account for "0" data which is technically missing and misrepresented.



F Data Issues Deep Dive by Locations

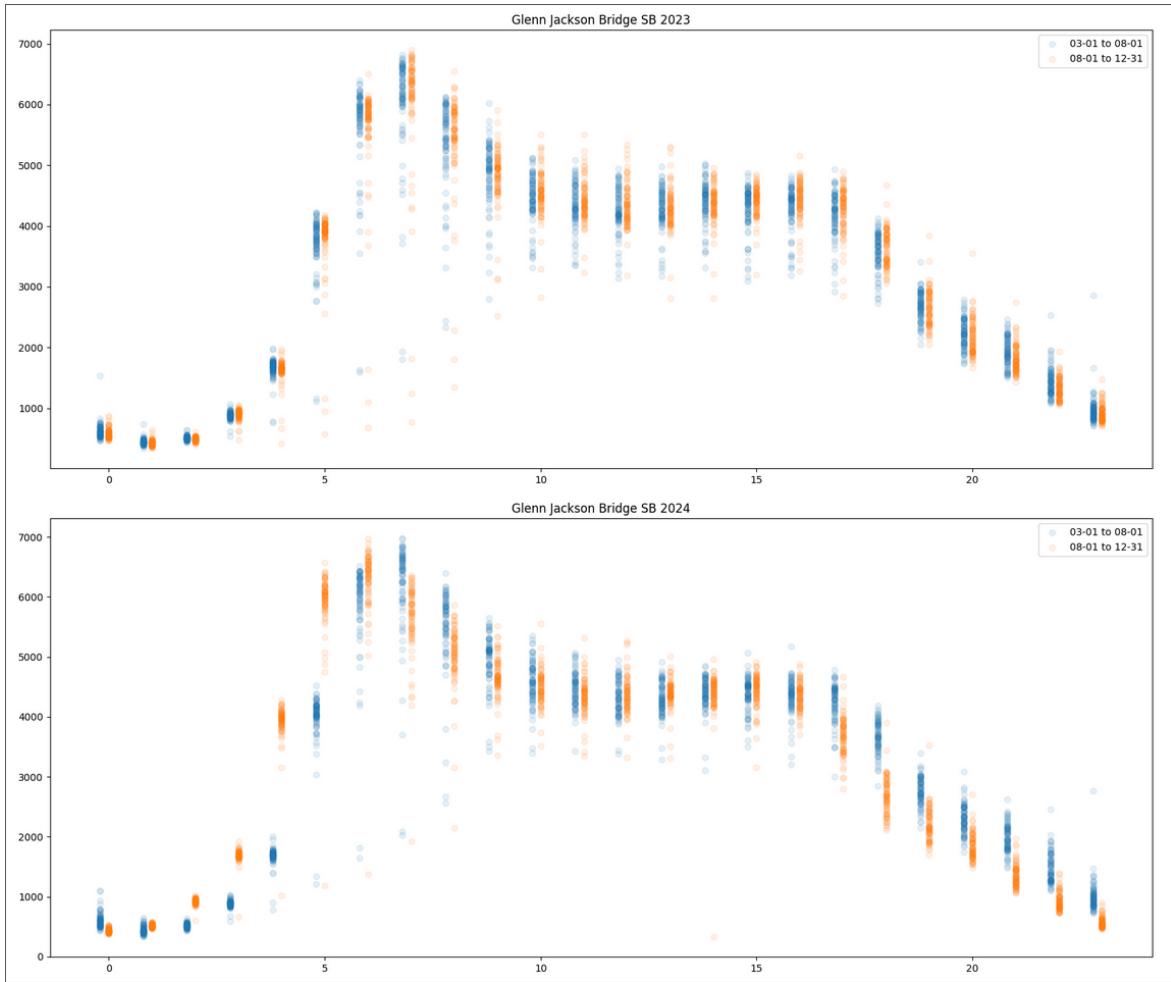


Figure 34: An example of when the data seems to be shifted by exactly an hour indicating issues with how DST was handled. Interestingly, the date ranges don't match exactly with when the time changes happen.

These plots provide a visual for similar data presented in table 1 but includes every year available in the dataset with quantile shading. Each line is the periodic cubic spline interpolation of the hourly medians/quantiles for each (location, direction, weekday/weekend) tuple. Some strange data issues become apparent in these plots, but they are almost entirely only in years we aren't concerned with in this analysis.

