Transit Corridor Study

Interim Report

By:
Zack Aemmer
Borna Arabkhedri
Don MacKenzie (Principal Investigator)
Andisheh Ranjbari (Co-Investigator)

Department of Civil & Environmental Engineering
University of Washington

Prepared for Stakeholder Review
June 2021

This study is sponsored by Amazon, Bellevue Transportation department, Challenge Seattle, King County Metro, Seattle Department of Transportation, Sound Transit, and Uber, with support from the Mobility Innovation Center at UW CoMotion.

Table of Contents

Executive Summary

Population and extended economic growth in many Seattle neighborhoods are driving increased demand for private car travel along with transportation services such as ridehailing and on-demand delivery. Together, these trends are adding to existing demand for loading and unloading operations throughout the city, and exacerbating traffic congestion. Anecdotal evidence indicates that passenger/delivery vehicle stops at or next to transit stops can interfere with bus operations, causing longer or more volatile delays. The increased travel times and reduced reliability further erode the attractiveness of transit to travelers. Thus, it is important to understand how transit, ridehailing, and goods delivery vehicles interact in terms of both operations and travel demand.

This project focuses on the analysis of open-source transit data to screen for locations with slow and/or unreliable bus travel times, and couple that data with interference observation, environmental, and traffic-related data to potentially predict the likely causes. We have developed tools to identify transit corridors with high levels of interference from other road users, including passenger cars, ridehailing vehicles and goods delivery vehicles. These tools are applied to transit corridors in Seattle and Bellevue, and methods have been developed to identify likely sources of interference from available data.

We drew on multiple data sources for identifying high-interference corridors in the region, including:

- a virtual workshop with participants from beneficiary agencies and stakeholders to solicit input;
- an online crowdsourcing survey to engage the community and gather feedback from all road users;
- route-level ridership data from King County Metro; and
- aggregated pick-up/drop-off data on ridehailing activities from SharedStreets.

Data was consolidated and 10 corridors were selected based on their likelihood of containing interference between buses and other road users, transit ridership levels, and stakeholder and community feedback.

In addition, we have developed a tool for identifying corridors with slow and/or unreliable bus travel times from open-source real-time transit data. We implemented a pipeline for ingesting and analyzing King County Metro's real-time Generalized Transit Feed Specification data (GTFS-RT) at 10-second intervals. Using this pipeline, active bus coordinate and schedule adherence data has been scraped and stored to an Amazon Web Services (AWS) server since September 2020. We developed efficient methods to aggregate tracked bus locations and assign them to roadway segments, and quantified delays in terms of schedule deviation and ratio of median to free-flow speeds, among other metrics. We have developed a web based visualization tool to display this data, and it is being updated daily with aggregated performance metrics from our database.

To collect ground truth validation data along selected corridors, we implemented an online data collection tool for field observations, and recruited research assistants to observe bus operations along the study corridors and record information on bus traversals and instances of interference. This dataset is analyzed alongside the GTFS-RT data, environmental, and traffic related data to identify instances of delay and predict the likely causes.

Field data was collected for three weeks along eight of the selected corridors in March 2021, but was later paused due to depressed levels of transportation activity during the COVID-19 pandemic and the current unstable condition of travel choices and city traffic (and thus interferences). Preliminary analysis on the collected data revealed that there is not a substantial effect shown in the GTFS-RT data when a bus is interfered with; however, there were not a lot of interference observations in the collected field data. So, it remains to be seen whether the lack of an identifiable effect is due to the lack of ground truth data, lack of precision in the automatic vehicle location system, or the relatively low impact of an interference when compared to the effects of general traffic congestion, signals, and other roadway conditions. A linear regression model was also generated to determine the extent to which roadway characteristics can predict segment performance, which produced mildly predictive results.

As businesses and transit services continue to reopen, there will likely be an increase in the amount of transit interference experienced between buses and other roadway users, which will potentially allow for the gathering of more ground truth validation data. Field observations will resume in late Summer/early Fall 2021 and will continue until enough data is collected to either (1) model connections between observed interference and bus delays in the GTFS-RT data; or (2) determine whether significant delays cannot be linked to observed instances of interference in the study corridors. The GTFS-RT data scraping will continue daily, and summarized in the developed interactive visualization tool.

The major anticipated benefits of the project can be summarized as follow:

- This work will help identify network-wide road and route segments with slow and/or unreliable bus travel times. We may also be able to identify main causes of delay in the study corridors.
- Moreover, we expect that this work will generate reusable analytical tools that can be applied by local agencies on an ongoing basis, and by other researchers and transportation agencies in their own jurisdictions.
- The outcomes of this work will enable identifying corridors with slow and/or unreliable bus travel times as candidates for specific countermeasures to increase transit performance, such as increased enforcement, modified curb use rules, or preferential bus or street use treatments. Targeting such countermeasures towards priority locations will result in faster and more reliable bus operations, and a more efficient transportation network at a lower cost to transit agencies.

Introduction

Population and extended economic growth in many Seattle neighborhoods are driving increased demand for private car travel along with emerging services including ridehailing and on-demand delivery. Together, these trends are adding to existing demand for loading and unloading operations throughout the city, and exacerbating traffic congestion. Anecdotal evidence indicates that passenger/delivery vehicle stops at or next to transit stops can interfere with bus operations, causing longer or more volatile delays. The increased travel times and reduced reliability further erode the attractiveness of transit to travelers. To improve the performance of transit services, it is important to understand whether other road users such as ridehailing and goods delivery vehicles interfere with transit vehicles in urban areas, and deploy measures to mitigate the causes of delay to transit operations.

This project focuses on the analysis of open-source transit data to screen for locations with slow and unreliable bus travel times, and couple that data with additional environmental and traffic-related data to potentially predict the likely causes. To accomplish this, we identified transit corridors in Seattle and Bellevue with high levels of interference from other road users which may include passenger cars, ridehailing vehicles and goods delivery vehicles. *Figure 1* shows an overview of the project stages and tasks.

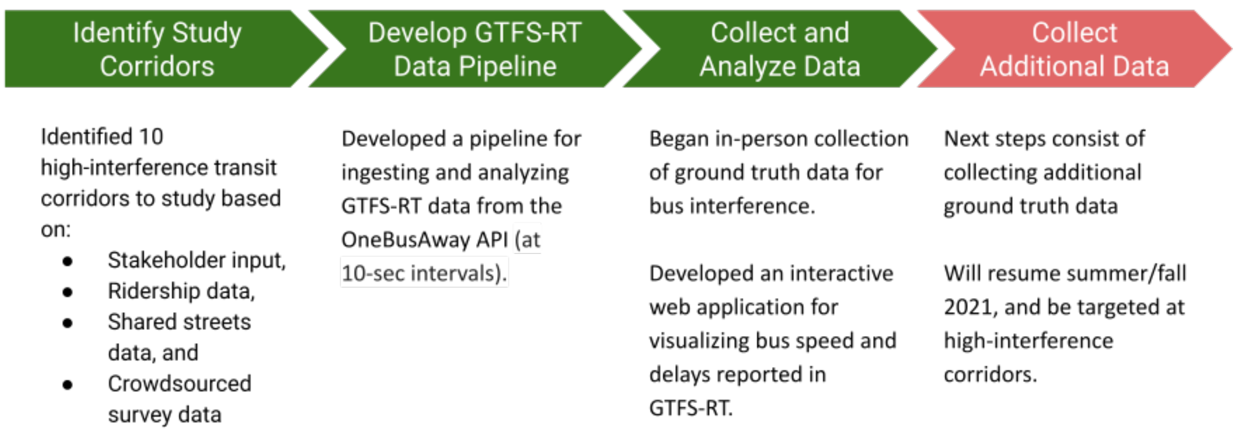


***Figure 1*: Overview of project stages and tasks.**

In screening for transit interference, we draw on multiple data sources for identifying interference, including the King County Metro Generalized Transit Feed Specification (GTFS) data and its real time component (GTFS-RT). Once locations with significant interference are located, we identify the likely causes by incorporating additional data (e.g. surrounding land use, number of lanes, speed limit, traffic volume, ridehailing pick-up drop-off, loading zone occupancy). Ground truth data for validation will be collected by field observers along pre-screened corridors throughout the Seattle-Bellevue region that are anticipated to have high levels of transit interference with other roadway users.

The GTFS data has provided a fully generalizable framework by which agencies and planners can collect and share transit scheduling data. More importantly, it has led to the development and proliferation of the GTFS-RT (real-time) standard by which *actual* bus locations and stop times can be shared in a standardized format, once collected by various Automatic Vehicle Location (AVL) systems. The real-time aspect of this data allows for both the quantification of transit delays, and potential classification of them. That is to say that if a bus is blocked or slowed during its trip, not only can we quantify the amount of time that it has been put behind schedule, but also observe with relative precision where that delay initially occurred, and derive or estimate additional information about what might have caused it. As part of this project, we have developed a tool to visualize several performance metrics derived from variables in the GTFS-RT standard. We have implemented this tool based on data scraped and stored from the OneBusAway API. The OneBusAway API is a programming interface which allows a user to request specific information (e.g. which vehicle ID is currently serving a specific trip, or where its position was last updated) about vehicles in the system. The organization which manages OneBusAway is the Open Transit Software Foundation, which handles the code-base and also provides a mobile application for users to easily view real-time bus data in various supported cities. The OneBusAway API contains real-time and scheduling data for the King County Metro transit system, which covers the entire study area.

Literature Review

There has been extensive work attempting to predict the quantity of delay, and overall travel time variability experienced by transit vehicles, at the segment and trip levels. In general, successful approaches utilize trip-level characteristics such as traffic volumes, number of stops, and the length of a trip to predict travel time variability. Various models have been applied to these predictors, the simplest of them being univariate linear regressions and arguably the most complex being artificial neural networks. Other works have approached the problem in other ways, such as measuring the dissipation of delay, or using proxies like headway deviation. But ultimately the unifying question seems to be; “what is the likely quantity of delay induced by a given corridor or trip?”. This question is relevant to endeavors regarding schedule padding and delay accommodation; however, given the proliferation of GPS systems and widespread connectivity, bus arrival times can be predicted on a case-by-case basis in real-time. This then begs the question, “what specifically causes the bus delay induced by a given segment?”. This question has received significantly less attention.

Most approaches to delay prediction use regression models, and draw from predictors that are not necessarily specific to transit delays, but are indicative of all vehicles using a segment. This approach has been used to develop models on a segment basis and apply predictors such as the number of buses, schedule adherence, number of stops, and traffic density as well as roadway characteristics such as left/right/through lane densities (1–3). The findings tend to suggest that variables related to moving time/travel time ratio, segment length, and number of bus stops are the most predictive variables; indicating that perhaps a combination of bus related variables (e.g. number of stops) and general traffic related variables (e.g. moving time/travel time, segment length) are most capable of quantifying transit delay for a given roadway segment. This distinction begins to suggest the opportunity for a classification model by which delay can be separated into quantities caused by bus activity for a segment (e.g. boarding and alighting), and quantities that are inherent to the segment itself (e.g. congestion). Another study explored this through variables drawn entirely from automated passenger counter (APC) data, such as specific quantities of boarding/alighting, lift use, headway deviation, and average passenger load (4). They were able to develop a fairly predictive model by using these transit-specific predictors. Additionally, several models were tested in which they used various measures of transit delay including run time, run time deviation, and headway deviation. The most success was found with run time as the dependent variable, in which nearly all variables in the model were significant.

Other works have gone on to apply increasingly exotic models to similar predictors. One study approached the problem of predicting transit delay through a series of multivariate regressions, followed by an artificial neural network (ANN) model (3). They defined travel time between each of the stops as the dependent variables and used arrival time, dwell time, and schedule adherence at each stop along a corridor as independent variables. Their findings supported the use of the more complex model, indicating that compared to the regression model, the ANN was able to predict arrival time (and by extension accumulated delays) more precisely for each stop on a single route. This is in congruence with another study that tested linear regression, ANN, and support vector machine (SVM) models and found linear regression to be the least accurate arrival time model (5). This suggests a more complicated relationship between travel time and characteristics of the route. As a matter of fact, ANNs have been applied frequently to attempt to predict transit delay (3,5–8). So much so that work has been performed to document many successful cases, from which one new study built itself by creating separate link- and stop-based models, then validated those models against a regional traffic simulation (7). Their results suggested that there were fundamental differences in the ANNs developed on stop-based characteristics than in those on link-based characteristics, specifically that stop-based models perform better when predicting segments with multiple intersections. This seems to suggest that there is a relationship between bus stops and intersections that could cause unscheduled delay, and that there is value in separating and accounting for segment and stop delay in different ways. They conclude by advocating for a hybrid model that predicts overall delay using characteristics of both stop and segment ANNs.

In the realm of delay classification, much less work has been performed. There are works which have broken down a single component of delay (e.g. transit delay due to an intersection) (9) to predict with relative ease the delays due to that single component. However, these papers tend to make extensive use of AVL and APC data, while few, if any, have based their models on GTFS-RT. This could be due to the still limited adoption of GTFS-RT, or more likely, due to it being a less comprehensive data source than AVL/APC. Part of the work that this study proposes is to lay the groundwork for a predictive model, which can identify delay characteristics easily obtained from APC data such as stop length, by using the positioning data from GTFS-RT. The work most similar to ours classified delays using AVL data according to their phase of operation (slowdown in traffic, stopped at a signal, boarding passengers, and free flow) (10). Separately, they developed linear regressions attempting to predict travel times for each segment in a transit network, but were not able to explain much of the travel time variance (10-15%). However, this work relied heavily on door-open and bus stopped datasets, which are typically only available in AVL datasets and not found in GTFS-RT.

Overall, there are strong use-cases for GTFS-RTdata, for example, retrospective analysis can determine hotspots of delay in transit networks, or trip segments with tendencies toward high variability (11,12). Both of these analyses can be applied by planners and agencies to adjust transit schedules and direct resources toward corridors, stops, or segments that need it most. In some cases, it is not delay itself that is modeled and analyzed, but the rate at which it increases or decreases across various land-use and roadway characteristics (13). In a conceptually similar way, we measure the individual occurrences of delay which contribute to these increases and decreases. Lastly, some studies have gone a step further and applied real-time prediction methods to determine delays in active systems, and used that information to provide better scheduling information to travelers. This was found to lead to lower overall wait times for travelers, and a more positive perception of transit reliability (14). This is to say that prior works have not only documented extensively our ability to predict delay, but the value in doing so.

Corridor Selection

We selected 10 corridors for observing and studying interference between buses and other road users (passenger cars, ridehailing vehicles, delivery vehicles, etc.). We obtained data from various sources for the purpose of systematically determining these locations, and have utilized several strategies to process this data and to identify corridors, which are listed below.

Input from Stakeholders

We solicited feedback from all project stakeholders prior to the project kickoff meeting, and discussed recommendations during the meeting. This list of corridors was the primary decision factor in our corridor selection process. We then extended or altered the proposed locations to best capture areas with high ridership, survey responses, or ridehailing activity.

Crowdsourced Interference Data

We designed a survey to solicit input from the community and road users, and to understand their feedback on busy transit corridors with high interference. All the survey questions were phrased to avoid requests for any personal information, instead asking for general input regarding observation location and type. The questions were designed to capture locations of interference, the time at which they were observed, the primary cause of the interference, and the bus routes that were affected.

We published this survey through UW news on March 17, 2020 and allowed responses for 10 days. We received a total of 78 responses from the public. Almost all of the locations were mentioned in a descriptive and accurate way; but inevitably, some effort was required to recode responses into a standard address format. Furthermore, the survey contained a small number of irrelevant, duplicate, and unclear responses. To address these deficiencies, we performed a data cleaning process in which all the locations were rewritten in standard address format, and unclear responses were re-coded appropriately or discarded. A map of locations solicited through the survey is shown in *Figure 2*. Further details on the data cleaning process and the results of the survey can be found in Appendix A.

Bus Ridership Data

King County Metro (KCM) provided a dataset containing Fall 2019 weekday ridership data and summary ridership levels in the form of total counts, which helped us identify corridors with high ridership levels. *Figure 3* shows corridors with more than 3000 riders during the combined daily peak hour periods (6-9AM and 3-7PM). This was calculated by determining the peak hour frequency for each route from KCM GTFS data, and multiplying it by the average number of riders on board during peak hour observations provided by the KCM ridership data.

To determine average ridership, we first filtered the provided observations to only include those observed during peak hour. We then averaged the number of riders onboard during each of those observations. The significance of this is that it does not account for ridership levels that fluctuate heavily across different stops for a route; it is a route-level estimate of the average number of riders that can be expected onboard a peak hour coach for a given route. This data was then joined based on route number to a publicly available shapefile displaying route paths for the KCM network.

Once the network was complete and overlaid with the proposed corridors, total peak hour bus ridership values were determined by using the trip frequency in the KCM GTFS, and summing the total ridership across all routes for each proposed corridor. These ridership values are reported in Section 3 along with their respective corridors.

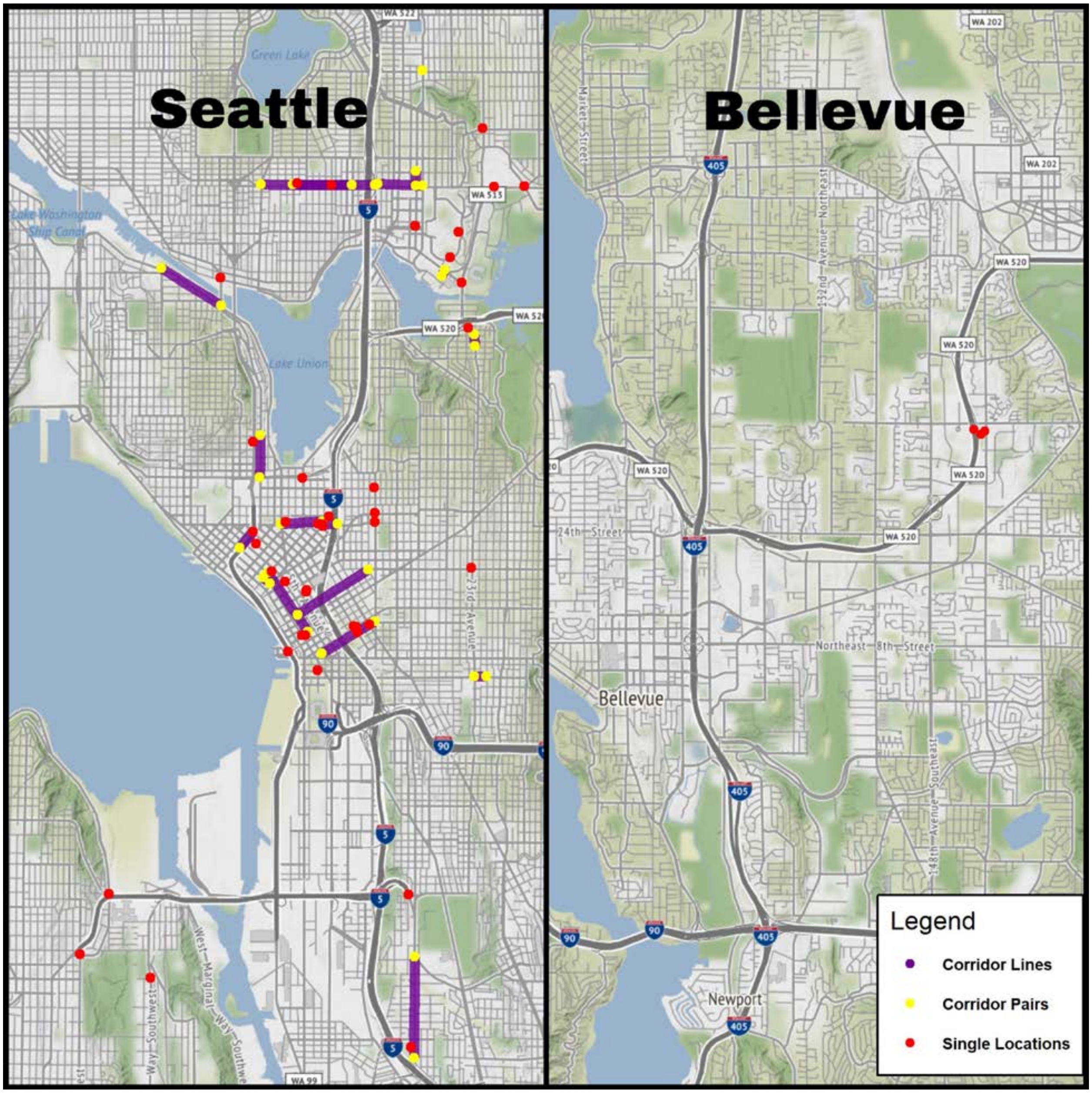


Figure 2: **Map of survey response locations. Red points show single-location responses such as an intersection or a bus stop (e.g. "Bell St & 5th Ave"); Yellow points, which are connected by purple lines, indicate locations where the respondent specified a corridor rather than a single location (e.g. "University Way NE between 45th and 47th St").**

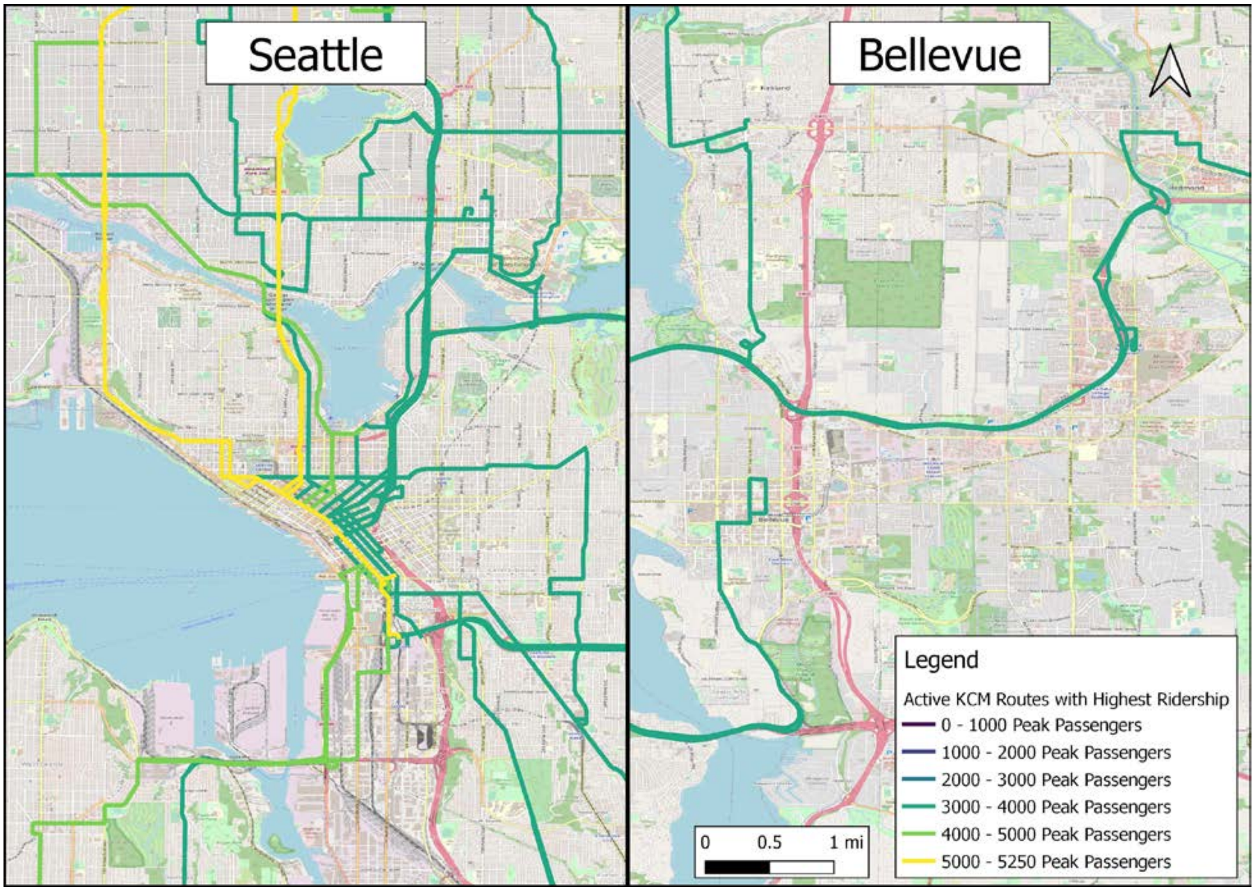


Figure 3: **Transit routes with peak period ridership greater than 3000 riders.**

Ridehailing Pick-up/Drop-off Data

Pick-up and Drop-off (PUDO) data was provided by SharedStreets, an initiative developed by the Open Transport Partnership nonprofit group, which develops software intended to aid city planners in decision-making tasks. One component of the SharedStreets project is an open-data partnership between Transportation Network Companies (TNCs) and cities, which provides PUDO data in a convenient user interface, and allows for visualization of areas with high TNC activity through a heatmap. As a project partner, SharedStreets provided us with access to their database, which contained average hourly counts for PUDO activity (aggregated over Uber and Lyft data) across nine weeks spread out across 2018-2019. For the purpose of this study, we only considered data for peak hours (all days of the week from 6-9AM and 3-7PM), and treated pick-up and drop-off activity the same.

The point-value based data was interpolated using an Inverse Distance Weighting (IDW) tool in GIS, which allowed for the generation of heat maps and revealed areas near our study corridors with tendencies towards high PUDO activity. Some artifacts occurred in this process due to individual locations in the dataset having very low average PUDO values, despite being surrounded by points with very high average PUDO values. The IDW process tends to leave small maximum and minimum values at the location of point values. This is manifested in many small “points” of low activity surrounded by areas with heavy PUDO activity, so it is important to judge the maps holistically and not on a meter-by-meter basis.

This data did not drive our initial corridor selection process, but was instead used to adjust, extend, or retract boundaries of our proposed corridors, such that areas of high average PUDO activity were included. It was the initial step in the process of refinement, the results of which are discussed in the next section.

Selected Corridors

Figure 4 shows an overview of the selected study corridors, and a list of corridor locations with additional information is provided in Table 1. A closer look into each individual corridor is provided in Appendix B.

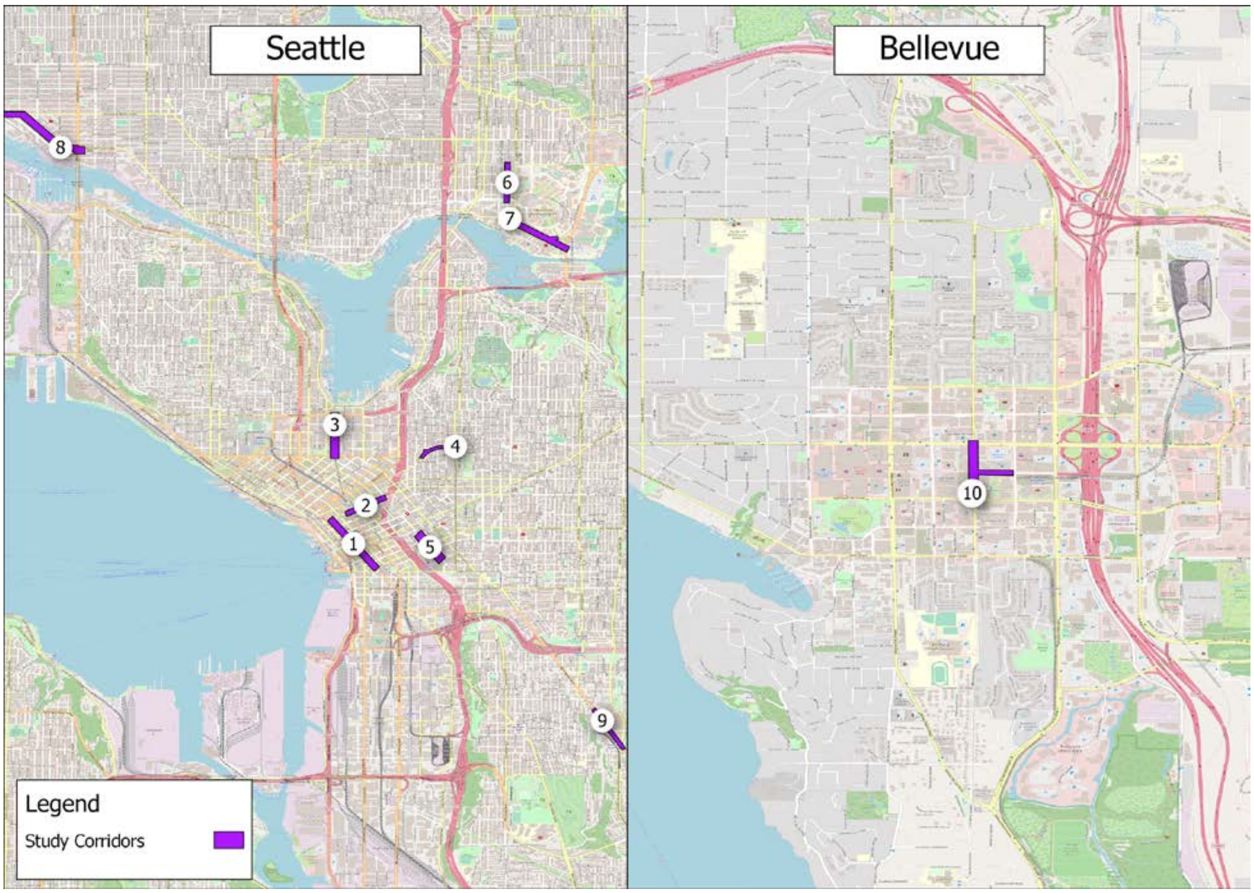


Figure 4: **Overview of the selected corridor**

Table 1: Summary statistics for selected corridors

Corridor ID	Location	Number of Blocks	Primary Bus Routes (Not All Routes)	Recommended by	Total Bus Ridership (1000 Peak Hour Riders)	Number of Nearby Survey Responses	Notes
1	2nd Ave; Pike to James	9	7, 10, 44, 45, 36, 522, 550, 577, 578, 590, 592, 594, 595	KCM, SDOT	23	2	-Downtown -Near transit-only corridor -High commercial and TNC activity -Large number of routes/ridership
2	Pike; 3rd to 9th	6	10, 11, 47, 9	KCM, SDOT	22	0	-Downtown -Near convention center -High commercial and TNC activity
3	Westlake; Denny to Mercer	5	40	SDOT	9	1	-South Lake Union -Near Amazon -High TNC activity
4	E Olive / E John; Denny to 10th	6	10, 43, 8	KCM, SDOT	5	1	-Capitol Hill -Near light rail station
5	9th Ave; Alder to Columbia	4	303, 13, 2, 4, 3, 60, 193	None	10	3	-First Hill -Near medical center -Cluster of survey responses
6	University Way; Campus Pkwy to 45th	5	71, 73, 373, 45	KCM	4	2	-University District -Near UW campus and dorms -High commercial and pedestrian activity
7	Pacific; 15th to Montlake	5	71, 73, 373, 45, 65, 980, 982, 986, 78, 541, 542, 586	SDOT	17	2	-University District -Near UW campus and medical center
8	Northwest Market / Leary; 24th to 15th	4	40, 29, 18, 17	KCM, SDOT	6	0	-Ballard -Convergence point for Ballard bridge -High ridership
9	Rainier; S Bayview to MLK	4	106, 14, 7, 48, 8, 987, 9	SDOT	13	0	-South Seattle -Near Mt Baker transit center
10	108th Ave; 4th to 12th	4	75, 77, 172, 270, 532, 535, 550, 556, 566	City of Bellevue, Amazon	10	0	-Bellevue -Near Microsoft campus -Near Overlake transit center -Construction interaction

Quantifying and Visualizing Transit Performance

In addition to the corridor selection and in-person data collection process, a pipeline for ingesting and analyzing GTFS-RT data from the OneBusAway API has been developed and put in place for determining transit performance as reported by the KCM AVL systems. Using this pipeline, active bus coordinate and schedule adherence data has been scraped and stored to an Amazon Web Services (AWS) server since September 2020. Efficient methods to aggregate tracked bus locations and assign them to roadway segments have also been developed, and delays have been quantified in terms of schedule deviation and ratio of median to free-flow speeds, among other metrics. A web based visualization tool has been developed to display this data, and is currently updated daily with aggregated performance metrics from this database. In this section we detail the process by which GTFS-RT data was collected and analyzed, following the general steps outlined in *Figure 5*.

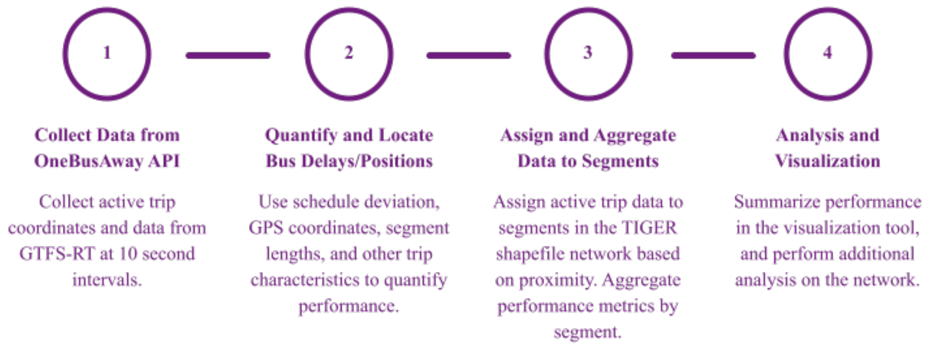


***Figure 5*: Overview of the framework proposed for measuring and classifying transit delays through GTFS-RT data.**

Collect API Data

The key variables that are extracted from active trips in the GTFS-RT system are:

- Trip ID
- Vehicle ID
- Last known GPS Coordinates
- Closest Stop
- Schedule Deviation
- Coordinate Update Timestamp

The only variable added to the data as it is stored is:

- Data Collection Timestamp

Several API endpoints are available in the OneBusAway API, and each provides a unique look at the state of the system. For example, certain endpoints provide static, up-to-date schedule data such as schedule-for-stop or route-ids-for-agency. Others provide location data such as trip-for-vehicle or trips-for-location. The endpoint used in this study is the vehicles-for-agency endpoint, which provides a set of all vehicles in the KCM system, as well as the location and trip information for those which are currently active and traversing a route. The variables gathered from this endpoint, and a high-level overview of the GTFS-RT data collection pipeline is shown in *Figure 6*.

Queries are made at the highest resolution feasible; different buses may update the API at different frequencies, and the system itself may only update every so often. The higher frequency that the data is collected, the more precisely delays can be quantified and located. This of course comes at heavier computational costs during the analysis, and more calls to the API. We chose to query the API at 10-second intervals, from 6am to 9pm each day.

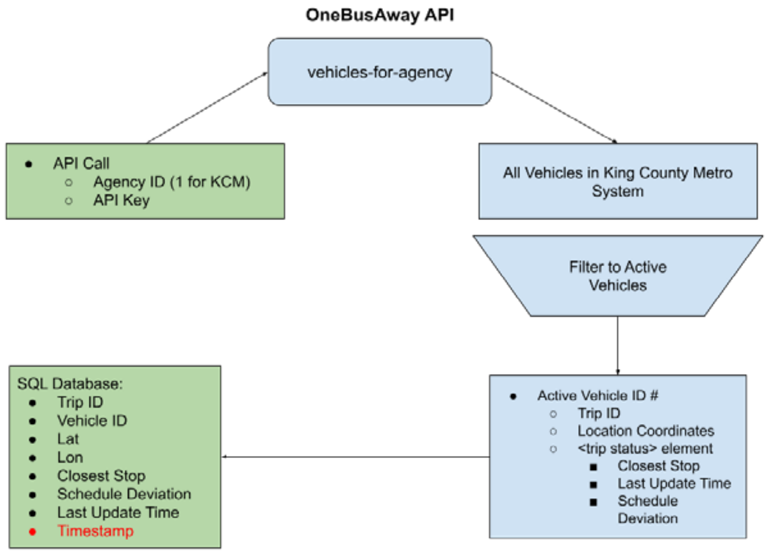


***Figure 6*: Input and output structure for data collection used with the "vehicles-for-agency" endpoint of the OneBusAway API. All active vehicles are queried at 10 second intervals, timestamped, and added to a SQL database. The timestamp parameter is added to the GTFS-RT response when it is added to the database, and is different from the time corresponding to the last coordinate update.**

Quantify Bus Metrics and Locate their Positions

Once bus trip updates have been recorded during collection and stored in the database, performance metrics are calculated between consecutively timestamped trip IDs for each unique day in the data. Thus, a particular measurement for a performance metric does not necessarily represent a point-observation, but a period of time between tracked locations. For the sake of analysis, we treat these measurements as point-observations and assign them the coordinates of the latter of the two tracked locations. The following performance metrics are calculated for every trip in the GTFS-RT data:

- Speed at bus position *j* consecutive to position *i* was calculated using the trip distance (TD) and location timestamp (T) variables from GTFS-RT:

$$Speed_j\ (m/s) = \frac{TD_j - TD_i}{T_j - T_i}$$

- Pace is then calculated as the inverse of speed to provide a more direct measure of delay incurred on a given segment:

$$Pace_j\ (s/m) = \frac{1}{Speed_j}$$

- Delay is determined using the schedule deviation (SD) variable that is available for all buses broadcasting real-time arrival information, and is defined as a cumulative measure of all delays and speedups experienced by a vehicle during a given trip. In other words, the schedule deviation states how far ahead, or behind schedule a bus currently is. Therefore, we identify delay occurrences as locations where there is a change in the value for schedule deviation between consecutive locations:

$$Delay_j\ (sec) = SD_j - SD_i$$

It is important to note that schedule deviation can be negative (ahead of schedule) or positive (behind schedule). Any decrease in schedule deviation is treated as a negative delay, i.e. the bus sped up due to lack of congestion or other factors. Any increase in schedule deviation is treated as a positive delay, i.e. the bus slowed down due to delaying factors (*Figure 7*).

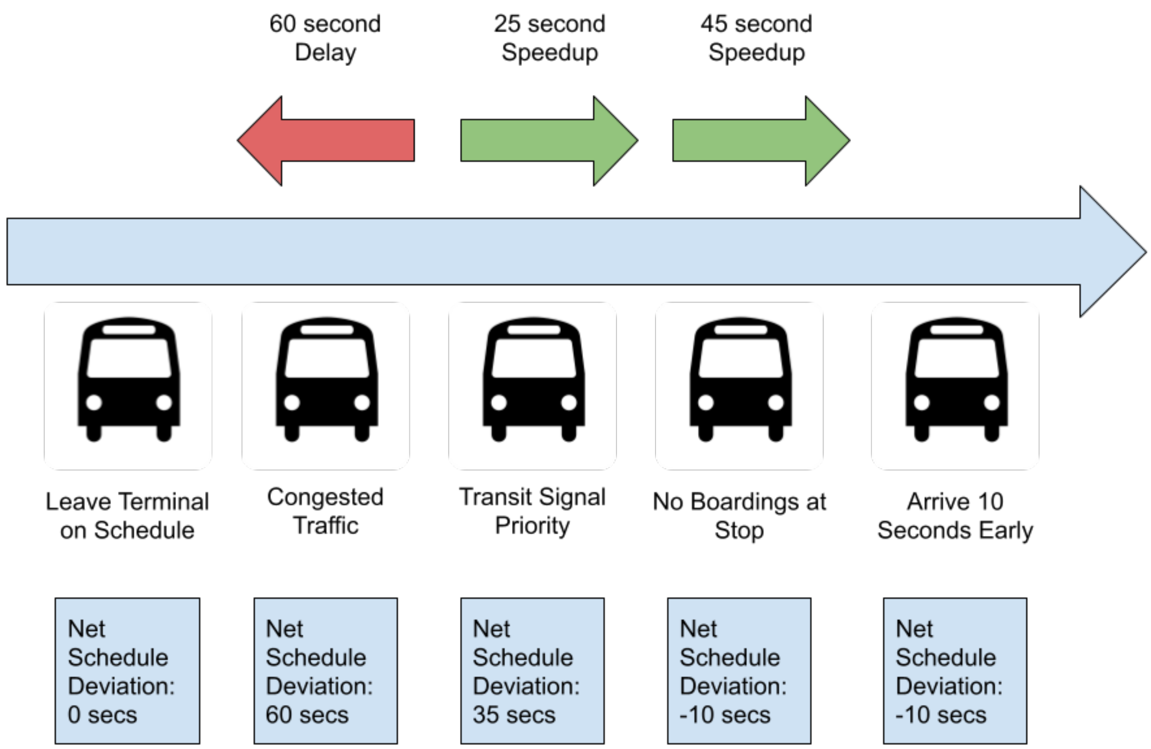


***Figure 7*: The process of determining instances of delay from the cumulative "schedule deviation" measure in GTFS-RT.**

Assign Segments and Aggregate Metrics

To analyze bus performance at the street segment level, we utilized the TIGER road shapefiles provided by the American Community Survey (ACS) as a link representation of the Seattle-Bellevue street network. These roadway shapefiles must first be decomposed into individual line segments, then re-joined into suitable streets. This pre-processing step is performed using GIS software to "explode" the shapefile into segments at every line intersection, then rejoin segments that are below a certain length threshold with their neighbors. When matching bus location data to these segments, the processing time can be significantly reduced if it is already known which segments a particular route will traverse. To this end, a route-to-segment mapping is also created using GIS software by buffering a shapefile containing all routes in the KCM network, then performing a spatial join with the "contains" predicate on our street segments dataset. This attaches route information to each segment, and allows us to greatly reduce the number of potential candidates when matching bus locations to street segments.

After the OneBusAwayAPI data is processed to determine all instances of tracked performance metrics, and each metric has been assigned a location based on the bus location of its second timestamp, these observations are overlaid with street segments and nearest-neighbor matching is used to assign each observation to the closest street segment. To facilitate this, a Ball-Tree spatial index and the aforementioned route-to-segment mapping are used to speed up processing time. The resulting dataset contains one row for each consecutively tracked bus location, with identifiers for the trip, route, vehicle, and roadway segment in question, as well as performance metrics calculated from the previous location to the current. These observations are then aggregated at the segment-level for analysis of the network as a whole. Aggregating all observations on a segment allows for analyzing the performance of each segment regardless of transit routes going through that segment.

Interactive Visualization

To interactively examine descriptive results of this collection and aggregation process, a visualization tool, called TransitVis, has been developed which is capable of displaying metrics for all roadway segments in the network for the most recently collected day's data. Currently, the metrics for each segment calculated from the database are summarized once daily, and uploaded to the site. The user interface is hosted at TransitVis.com (under active development), and a snapshot of that is shown in *Figure 8*.

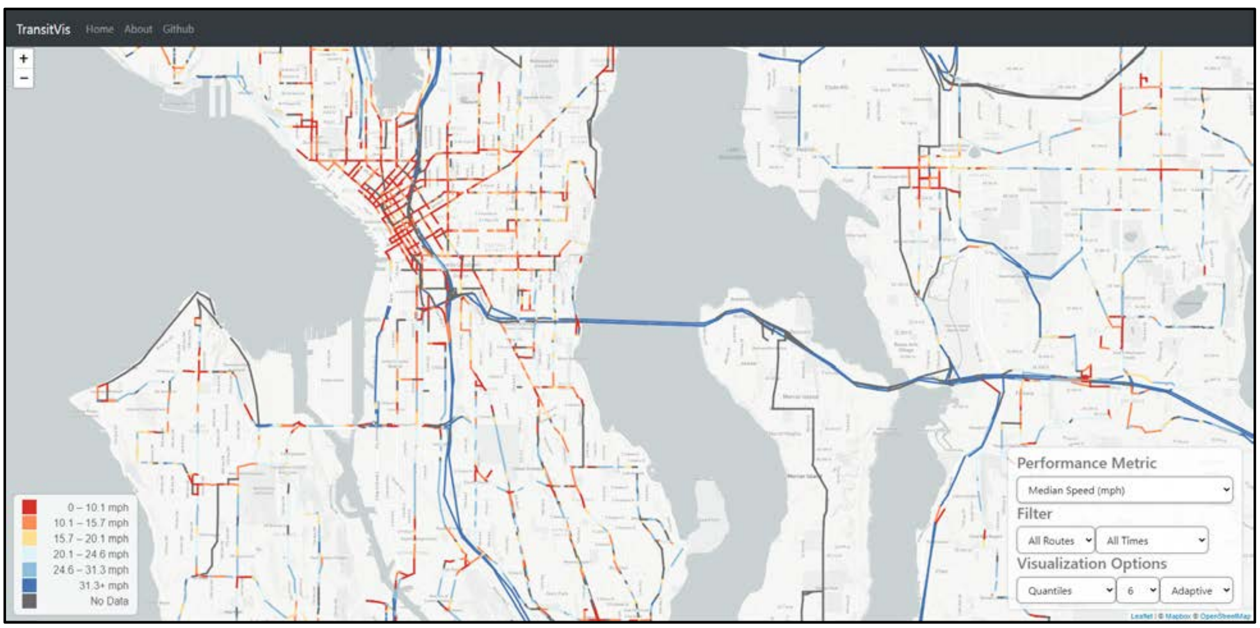


***Figure 8*: User interface for TransitVis, the developed tool for online interactive visualization.**

The TransitVis tool provides interactive ways for users to choose the metric displayed, as well as filter the data by route and time period, and change parameters related to the visualization scale for coloring roadway segments. Selecting a roadway segment on the map provides detailed information about that segment, such as the name of the street, the metric value, and the time that the segment was last updated. The user may interact through panning and zooming on the map, and perform transformations on the data and visualization through a set of widgets in the bottom right corner of the display (*Figure 9*).

This tool enables the user to explore the network for segments with uncharacteristically high delays by observing their performance metric of interest on a map, or to examine the details of a suspected route, corridor, or segment for more information on its delays or the delays of its neighboring segments.

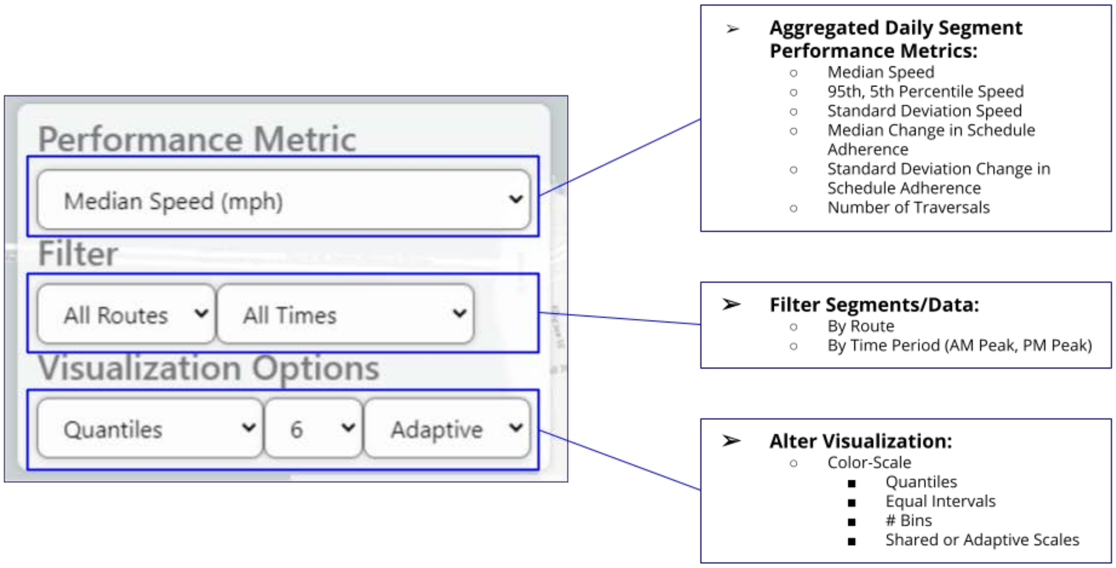


***Figure 9*: Widgets available for adjusting the data shown in the TransitVis tool, and the visualization parameters. Adaptive scale allows metrics with the same units to use different bins.**

Descriptive Delay Analysis

In addition to the one-day summary statistics provided in the visualization tool, a more robust analysis of the ingested GTFS-RT data was performed based on a full month of data collected during March 2021.

As mentioned previously, API queries were made starting at 6AM, and ending at 7PM. To obtain information on only the active vehicles, trip statuses listed as "canceled", or trips with a null identifier were removed. Additionally, any active trips without posted GPS coordinates were removed, on the assumption that these vehicles were not equipped with functioning real-time tracking. After preliminary cleaning, this dataset consisted of ~8,400,000 tracked locations belonging to 1,314 unique vehicle ids across 12,308 street segments. Approximately 50% of tracked delays/speedups (places where there was a change in schedule deviation) occurred at a transit stop. Summary statistics are provided for all delays and collection periods in *Figure 10*.

Spatial distributions of each performance metric are shown in *Figures 11-14*. In general, downtown areas create higher delays and perform worse than highway segments (e.g. I5, I90). An exception to this is the SR520 bridge, for which observed performance metrics were poor. The SR520 bridge has HOV/transit lanes and is tolled; however, there are several on- and off-ramps at the west end of the bridge. The segment representing SR520 is also fairly large in the street network shapefile, and so delays could accumulate at this location, showing a low performance for the entire bridge. When it comes to schedule deviation, highway segments deviate further from the mean and are less predictable than other segments. From a network-wide perspective, areas of low performance generally occur as clusters of streets rather than corridors. This is perhaps embodied best by *Figure 14*, which shows the ratio of median to free flow speeds for all segments. That being said, *Figure 13*, which shows the pace of buses traversing network segments, is slightly less noisy than other metrics and reveals some key locations where transit performance could be improved relative to nearby segments.

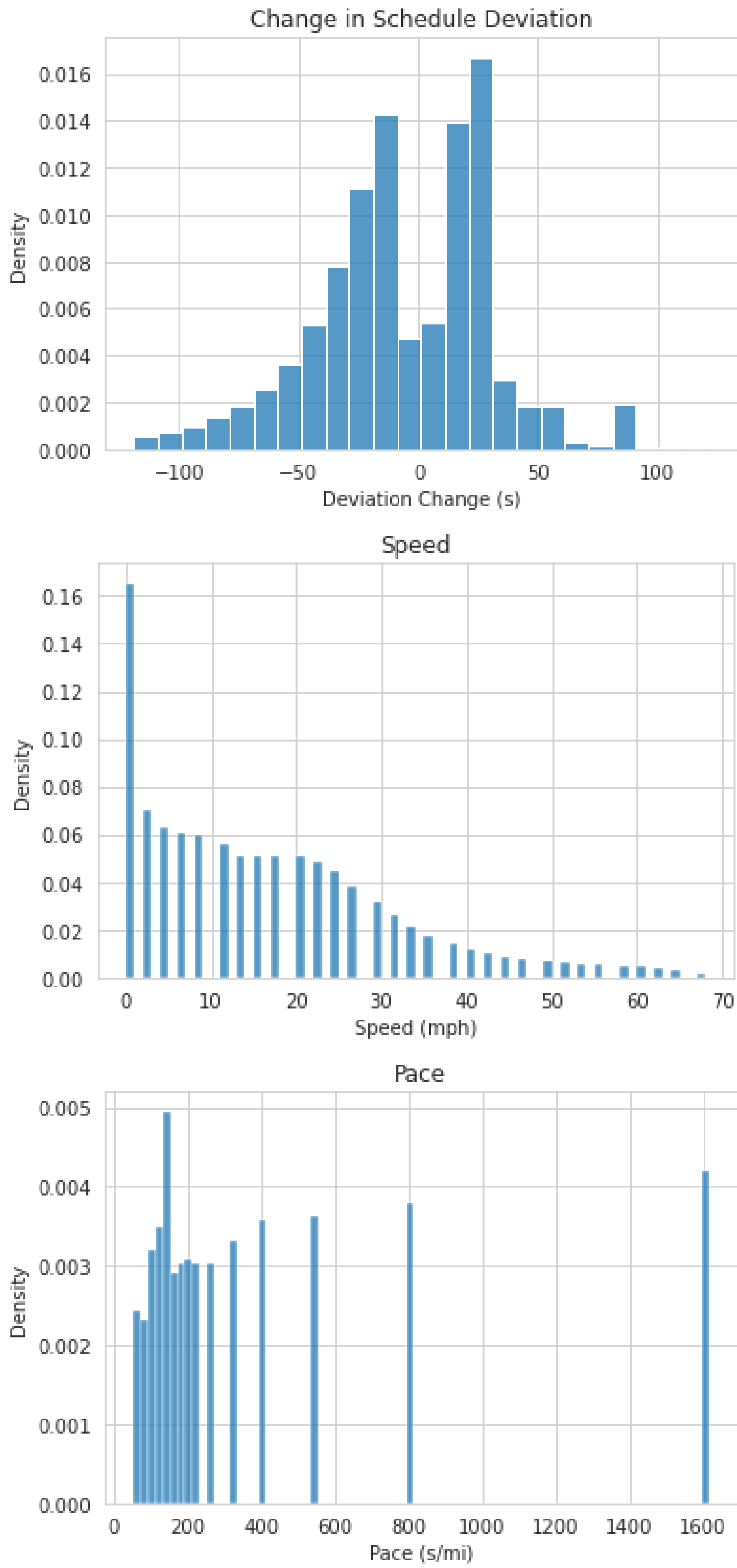


***Figure 10*: Network-wide univariate distributions of performance metrics based on data collected during March 2021.**

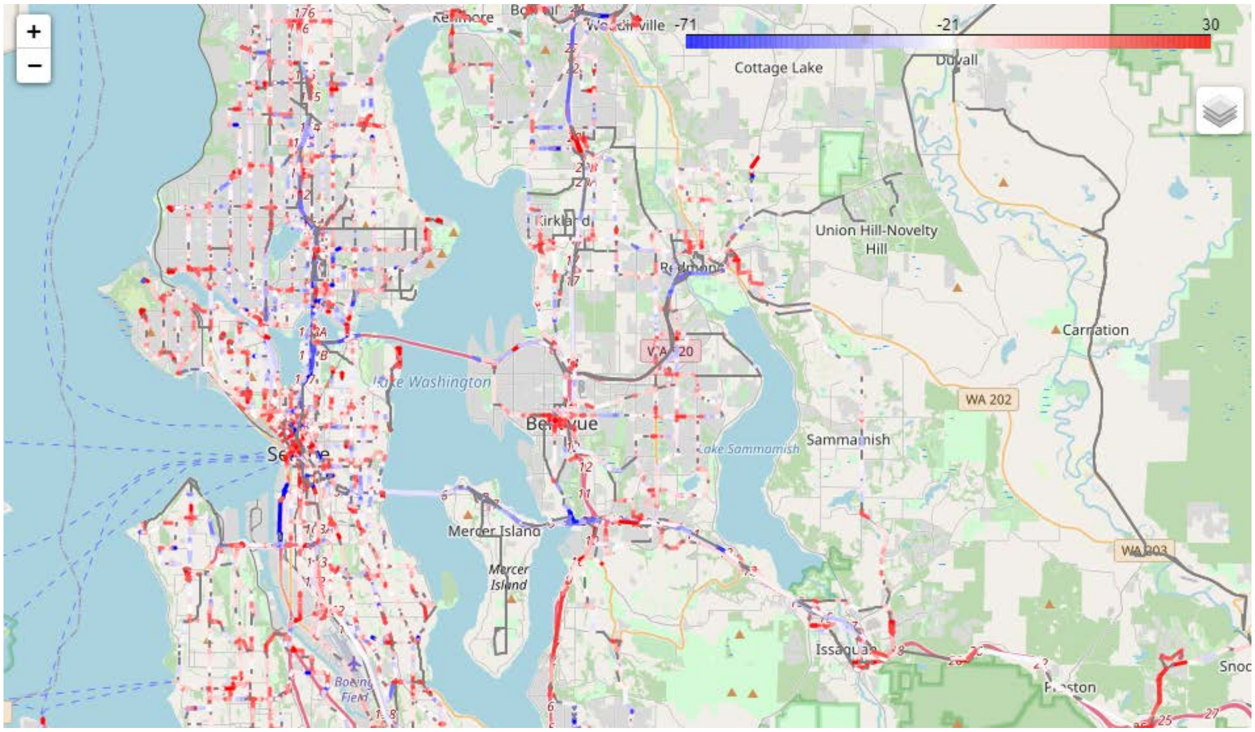


Figure 11: **Median transit delays (sec) for all segments in the network.**

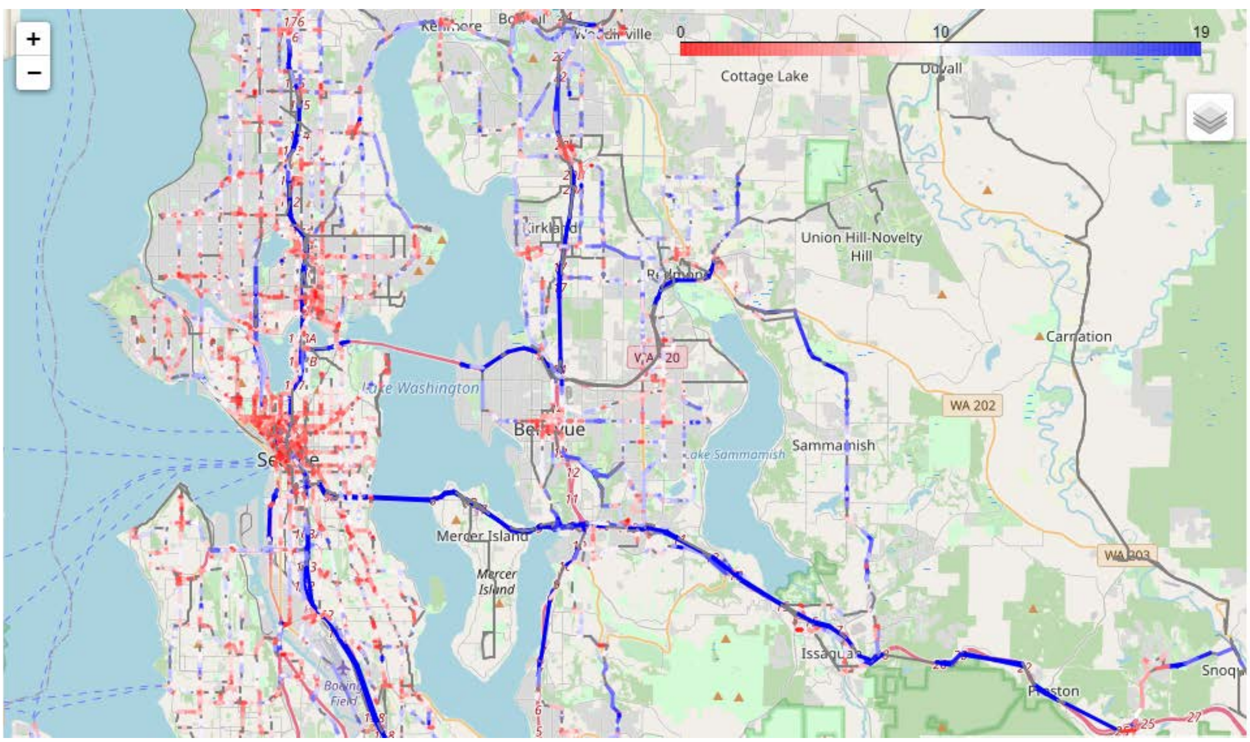


Figure 12: **Median transit speeds (mph) for all segments in the network.**

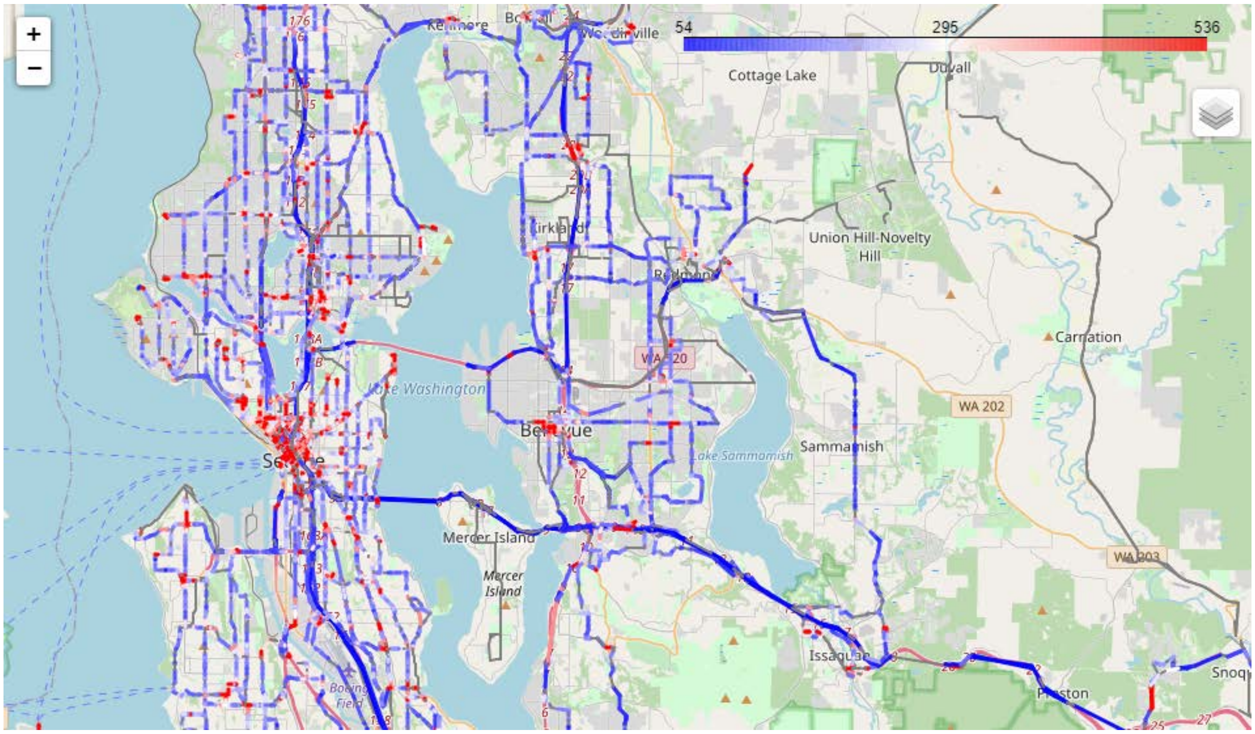


***Figure 13*: Median pace (sec/mi) for all segments in the network.**

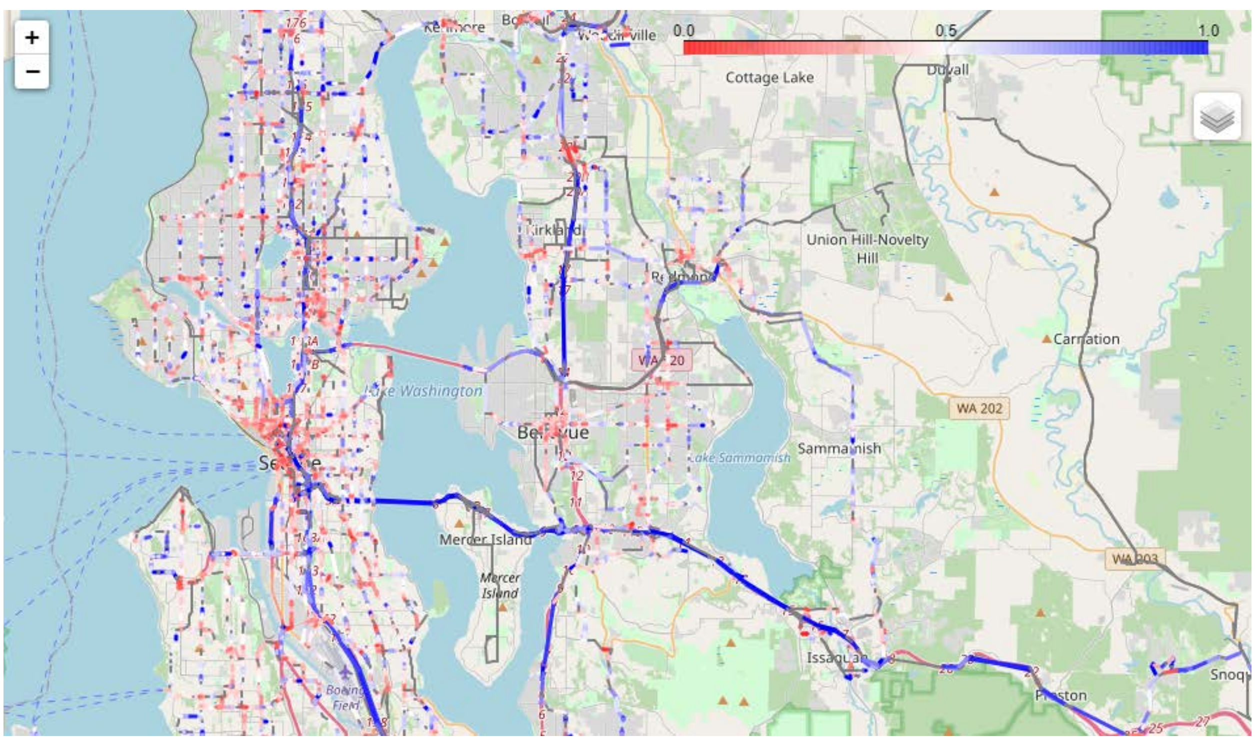


***Figure 14*: Transit performance (median/freeflow speed) for all segments in the network.**

Figures 15-19 display the pace of roadway segments near selected study corridors. In most cases, the corridors capture the worst performing segments in their respective regions. In the case of the downtown corridors (*Figure 15*), all corridors show poor overall performance; this is consistent with the surrounding streets, and most segments in the downtown area fall above the network median pace, as would be expected. The Fremont corridor (*Figure 16*) is perhaps the only corridor that could be changed in future studies by aligning with Northwest Market Street that has a slower pace in the surrounding areas.

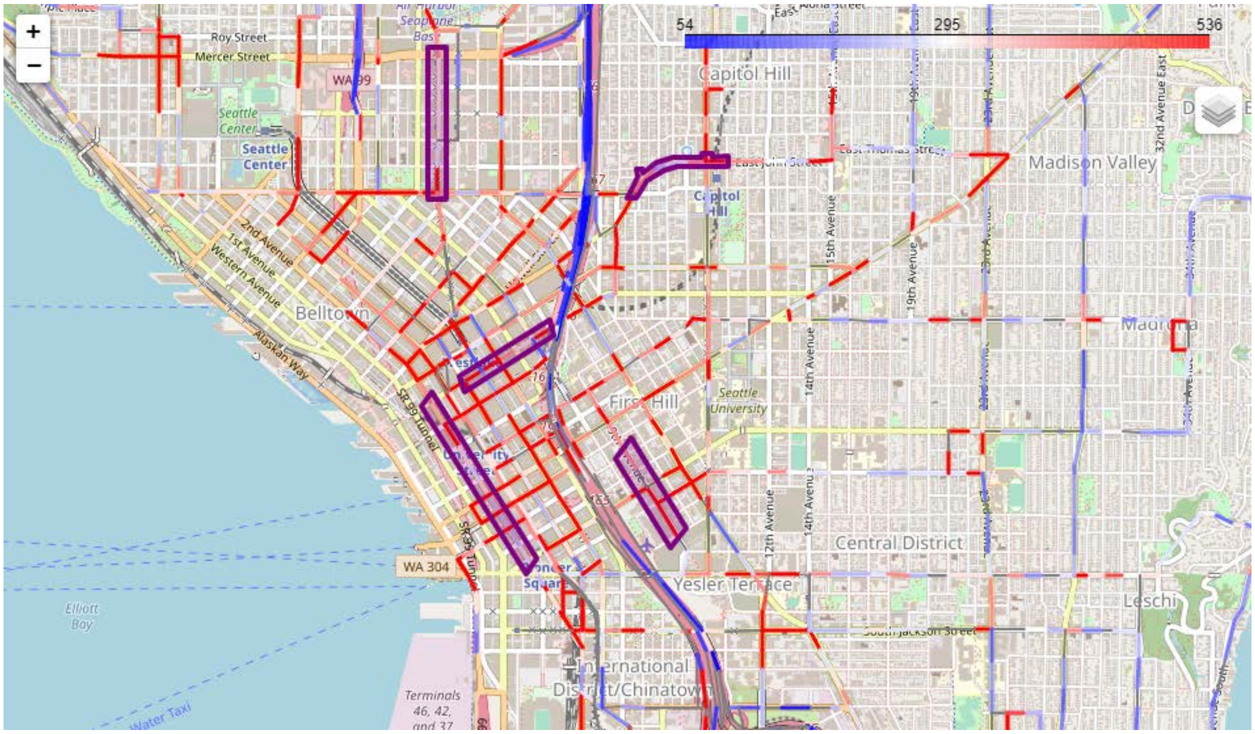


***Figure 15*: Segment pace for downtown study corridors and surrounding streets.**

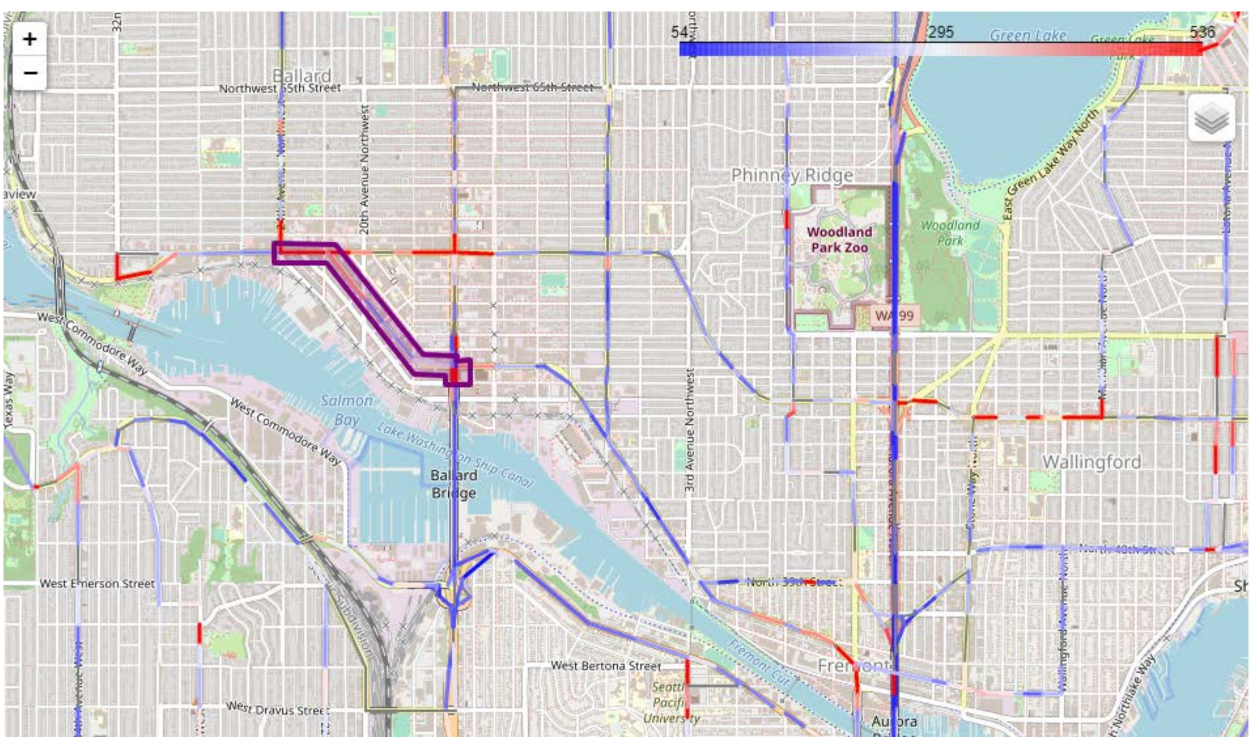


***Figure 16*: Segment pace for Fremont study corridor and surrounding streets.**

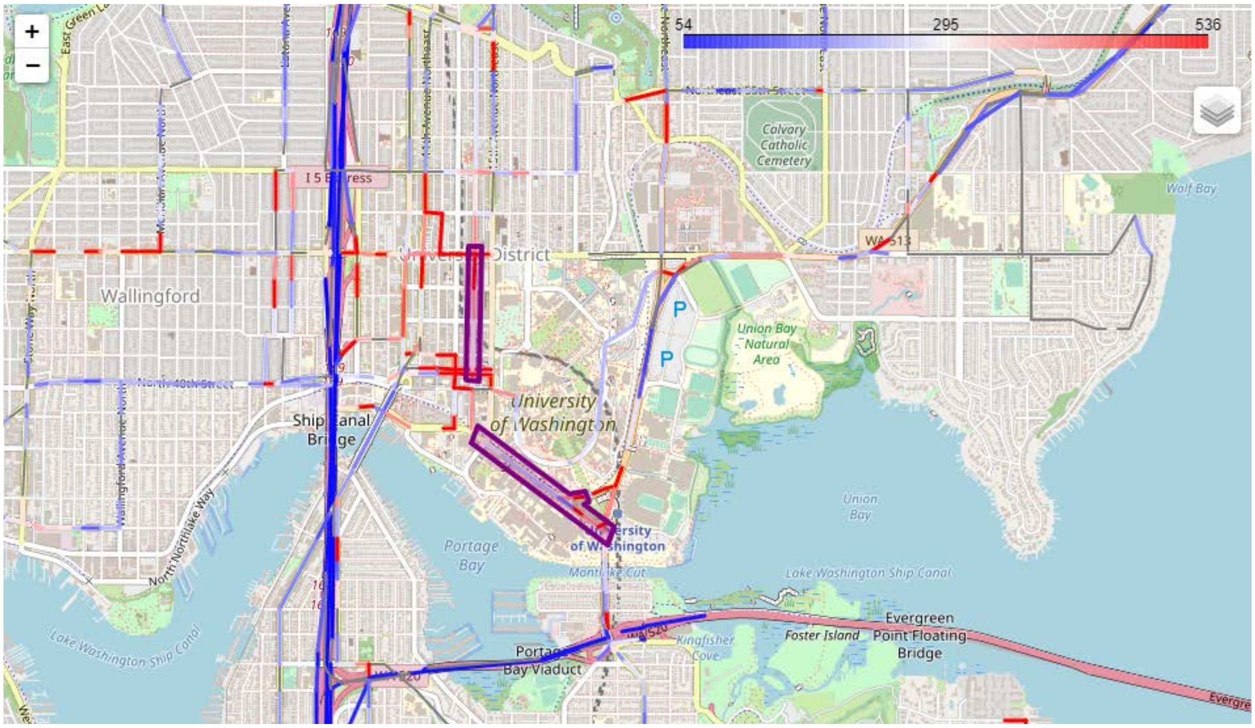


Figure 17: **Segment pace for University District study corridors and surrounding streets.**

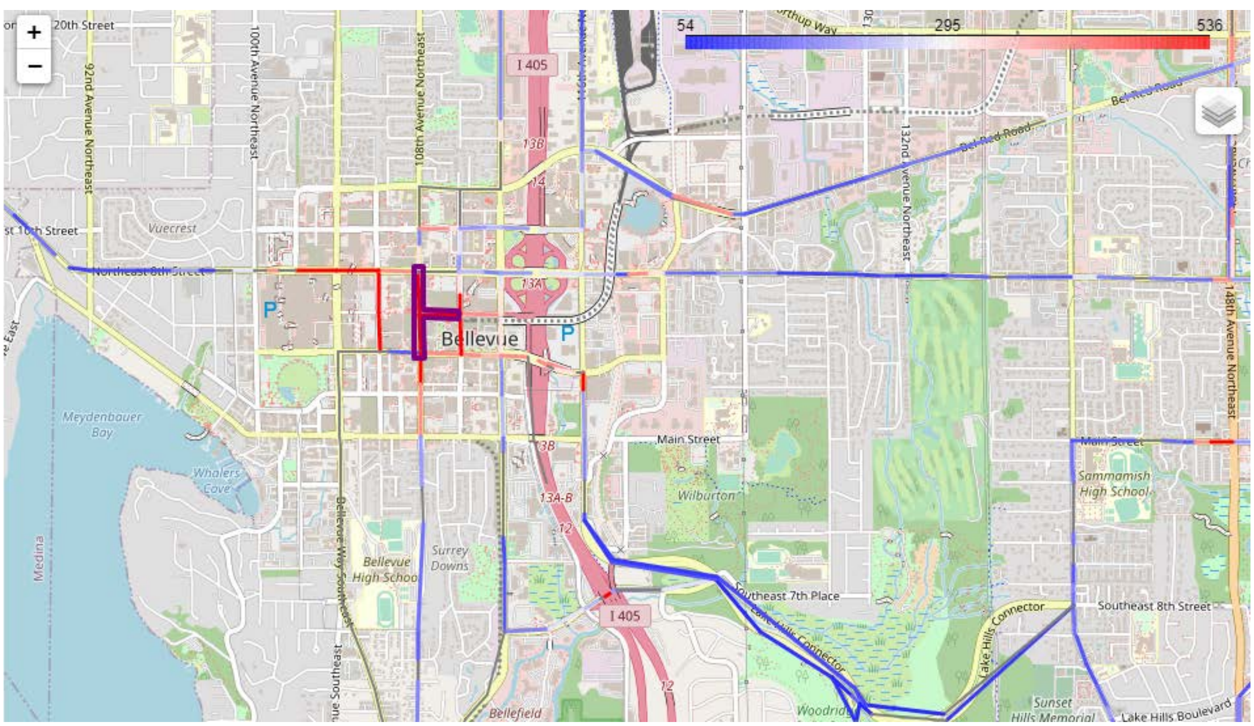


Figure 18: **Segment pace for Downtown Bellevue study corridor and surrounding streets.**

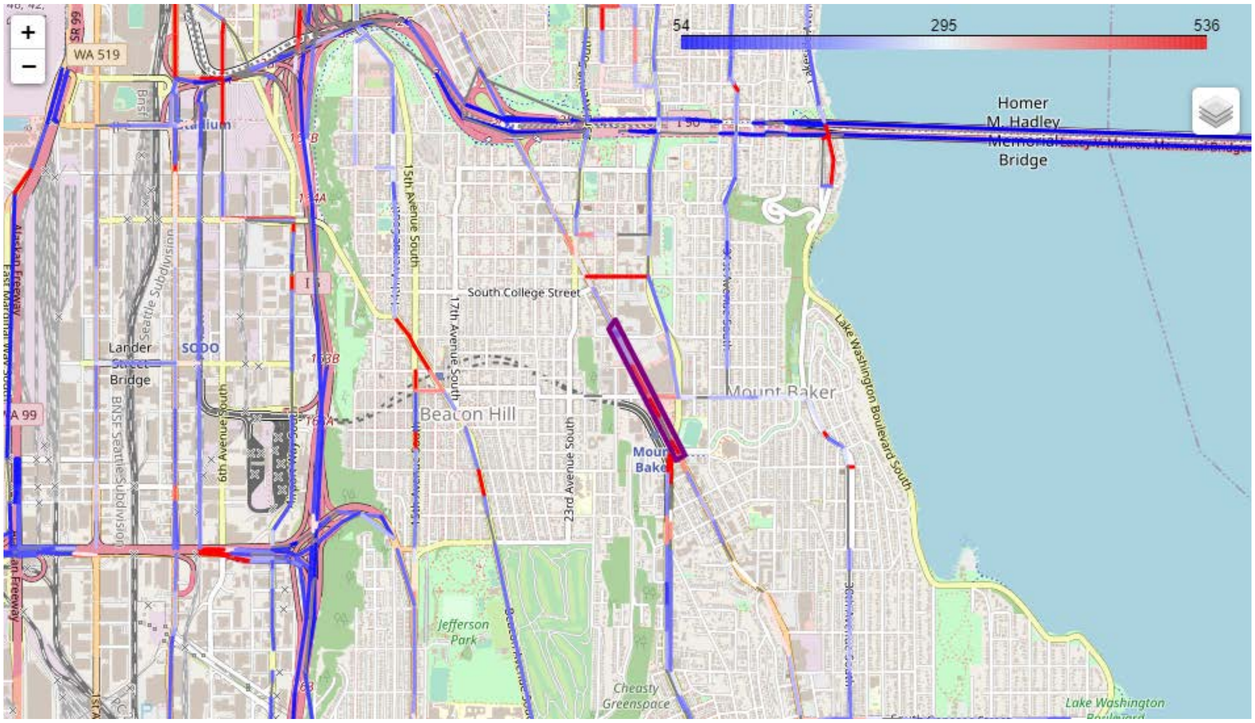


***Figure 19*: Segment pace for South Seattle study corridor and surrounding streets.**

Video Data Collection

One of the objectives of this project was to collect real-world data on instances of interference, as well as bus arrivals and departures at bus stops, to locate sources of delay in the selected transit corridors. The original plan was to achieve these ends through video data from bus-mounted cameras on KCM buses.

Most KCM buses are equipped with multiple cameras facing different directions. We received three video samples from KCM (averaging 20 minutes each) for different times of day and weather conditions to explore the feasibility of using video data. A screenshot of the user interface of the KCM video software can be seen in *Figures 20-21*. Double-clicking on a particular camera feed will expand it to the whole screen (see *Figure 22*).

Video data has multiple benefits and can effectively be used for identifying interference instances. Some of the benefits of using video data are:

1) Video data reducers can pause or play the video at any given moment and have the ability to rewind videos to review something they missed;
2) Access to multiple cameras with various directions allows seeing an interference incident from different angles. The list of cameras includes forward facing, front door, exit/rear door, mid door (if available), curbside view for front and rear, side facing view for front and rear, inside front to rear, inside mid to rear, etc.;
3) Ability to zoom in on a particular camera by double clicking its box to see an interference or event in more detail. ;
4) Field data collection would have been avoided;
5) The cameras function well in rainy weather and darker conditions with little to no performance drop compared to normal conditions.

A data request was made to KCM for obtaining several weeks of video data from a number of buses going through the selected corridors for the AM and PM peak periods (i.e., 7:30-9:00 AM & 3:30-6:00 PM). In order to process the request and provide the requested videos, KCM needed vehicle IDs and time windows. To address that, we developed a method for identifying the start and end time of vehicle IDs traversing study corridors using the GTFS-RT data. This was accomplished by matching tracked vehicle locations recorded in the GTFS-RT database to roadway segments in the street network, and filtering out locations that were assigned to segments not in a study corridor. After this, tracked locations were grouped by trip ID and date to get unique corridor traversals. The first and last tracked locations were then the start and end time for a particular vehicle ID traversing a given corridor.

Covering all buses that go through the selected corridors in peak periods would result in around 550 different buses and 24 hours of video, which would have been hard for KCM to extract. To reduce the size of the data request, we decided to request a sample of corridor traversals in a particular period of time. We tested two sampling methods: 1) choose the top 10 vehicles that traverse our corridors and request video from those buses during the time period in question; 2) choose 10 randomly selected vehicles (which could traverse areas outside our selected corridors), and extract whatever number of hours of video footage that relates to our corridors across the time period.

On average, we found that Method 1 would give us around 2 hours of video data in our corridors for the two weeks; while Method 2 would give us around 1 hour of data. However, Method 2 would be easier for KCM to handle, because it did not involve any processing, and all they needed to do was to pull the hard drives installed on buses and transfer the video footage.

Ultimately, the video data collection was disregarded since due to the pandemic KCM was experiencing some staff shortage and was not able to fulfill our data request. Instead, we collected data through field observations as explained in the following section.

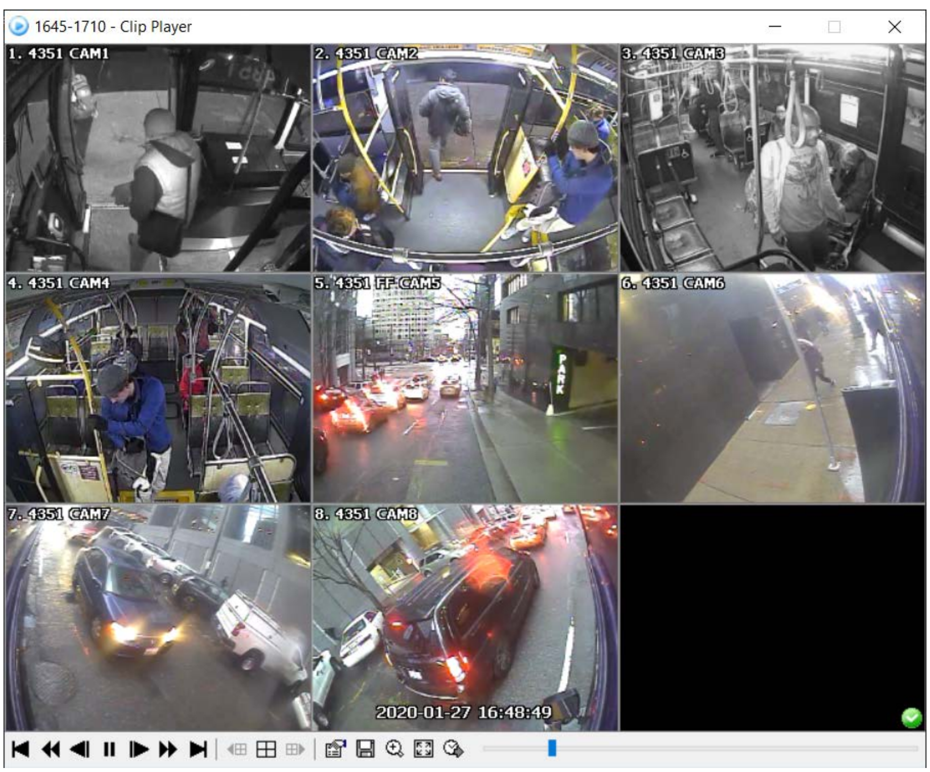


***Figure 20*: Passengers getting off the bus at a stop in Downtown Seattle in rainy weather**

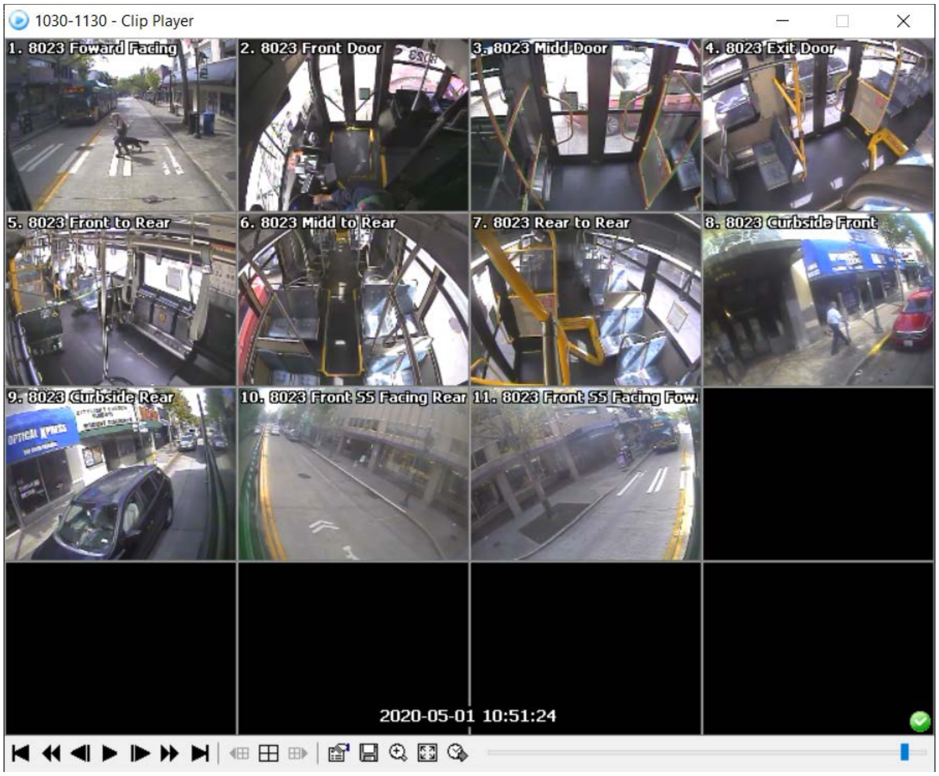


***Figure 21*: Bus coming to a full stop and waiting for a pedestrian crossing the street at a signed crosswalk on University Way NE**

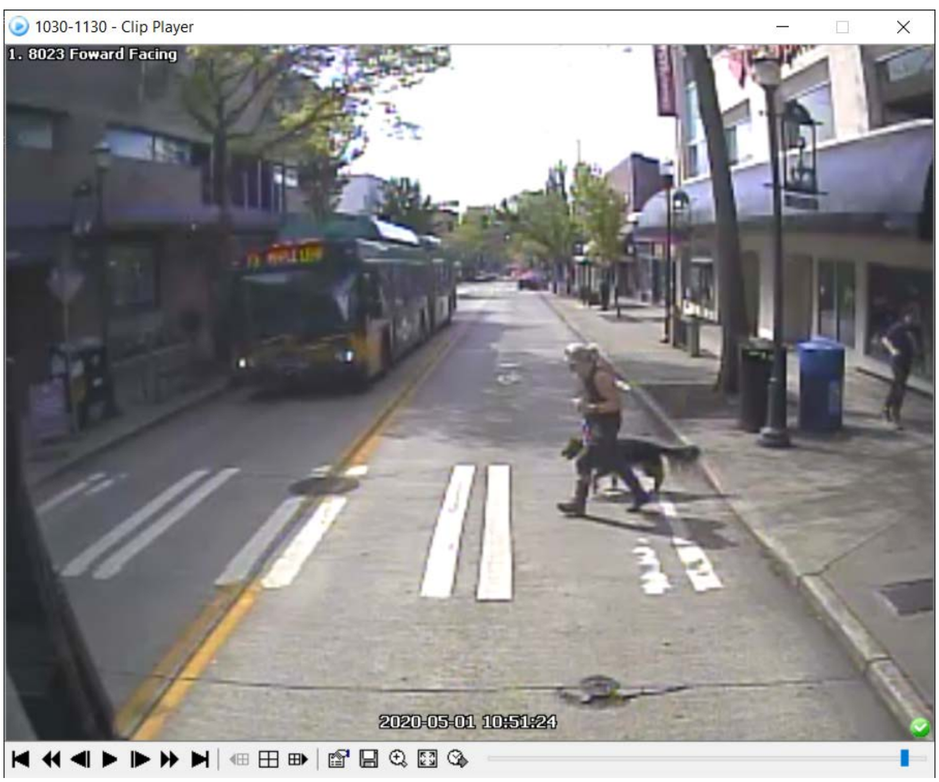


Figure 22*: Expanded Footage from the forward facing camera showing the same scene as *Figure 21

Field Data Collection

There are two types of data that we plan to collect from buses in the field to better understand sources of delay: 1) interference by other modes with buses (timestamp, interfering mode, duration, etc.), and 2) bus arrival and departure information (bus arrival timestamp, door open time, whether bike rack was used or not, whether flip-out wheelchair ramp was used or not, bus departure timestamp, etc.). This data will be used to 1) model interference and delay by establishing a relation between modes of interference attributed to delay and identifying significant variables, and 2) understand what happens at individual bus stops that lead to additional delay for buses, such as the use of bike racks and flip-out ramps. The bus arrival/departure data could also add additional value as they could be used to validate the accuracy of GTFS-RT data by comparing its timestamps to the timestamps collected in the field.

Variables of interest for the data collection are listed as follows:

- Interference information:
    - Interference start time (moment when another mode interferes with the bus's movement by reducing its speed or making it come to a stop)
    - Interference stop time (moment when the other mode stops the interference)
    - Interfering mode (e.g., pedestrian, ridehailing vehicle, delivery vehicle, etc.)
    - Location of interference (latitude and longitude)
    - Location category: 1) at a bus stop; 2) in a dedicated bus lane; 3) in a driving lane; and 4) other
    - Description of interference
    - Bus ID (four-digit number seen from all sides of the bus)

- Bus arrival and departure information:
    - Time of bus arrival at a bus stop (moment when the bus comes to a full halt at the bus stop)
    - Time of door opening (first door that opens)
    - Time of passenger flow stop (time when last passenger gets in or out)
    - Time of door closing (last door that closes)
    - Time of bus departure (moment when the bus starts accelerating to leave the bus stop)
    - Was the flip-out access ramp deployed?
    - How many people used the flip-out access ramp at the bus stop?
    - Was the bike rack used during the stop or not?
    - Did the use of the bike rack cause any delay for the bus?
    - Delay (in seconds) that the use of the bike rack caused.
    - Bus ID (four-digit number seen from all sides of the bus)

We established a framework to capture all variables of interest efficiently and cost effectively through field observations.

Two options were studied for the field data collection: 1) ride check, in which research assistants (RAs) board the bus and log information about bus operation from inside the vehicle, and 2) point check, in which data is collected by RAs standing on the sidewalk either at a bus stop or at a certain point along a transit corridor. Ride check data is more similar to video data reduction, because a field observer will stay inside the vehicle for an extended period of time, observing interferences and other desired variables (just like how the cameras also move with the bus). In the point check data, unlike the ride check and video data, the field observer collects data from a single point while different buses approach and leave. In order to cover the whole corridor through the point check method, multiple observers are required across the corridor, covering the entire length and avoiding blind spots which might lead to missing an interference. The pros and cons for each method are mentioned in the following.

Ride Check

- Advantages:
  - Observers can capture an interference regardless of where in the transit corridor it happens since they are moving with the bus (i.e., not having blind spots).
- Disadvantages:
  - One disadvantage, compared to video data collection, is that there are not multiple cameras facing different directions; to get that same view of the surroundings and all doors there might have been a need for two data collectors on the bus.
  - Due to the pandemic and the lack of vaccination at the time of the data collection it would not have been the safest option since it was indoors.
  - We calculated the approximate number of reports per hour from the route data collection method and it was found that it is lower compared to the point data collection meaning that it was less cost-effective.

Point Check

- Advantages:
  - More cost effective per our calculations.
  - Students get more familiar with the point they are assigned to over time leading to better efficiency in data collection.
  - Standing in outdoor space and more than six feet away from crowds to minimize risk of COVID-19.
  - More reliable data due to having the field observers' locations in advance. Field observers were assigned to a particular bus stop or location along the corridor in advance and had to stay there for the whole duration of their shift; thus, their collected data for the whole shift was linked to one bus stop only. For route data collection, we would have relied on their phones' GPS to understand which stop they were collecting data from when the bus arrived there, which was less reliable.
- Disadvantages:
  - No matter how hard we try, there still may not be complete coverage of the corridor compared to the route data collection option.

Due to the pandemic and to avoid any health risk posed to the data collectors, we proceeded with the point check method for field data collection.

Field Observer Recruitment

A job description was posted on the UW's job posting website and was also pushed through UW Civil and Environmental Engineering (CEE) undergraduate mailing list to hire RAs for field data collection. A total of nine RAs were hired for the data collection. A data collection protocol was sent out to students and they were asked to attend an online training session to prepare for the field work. To make sure the planned work met all health and safety requirements associated with the COVID-19 restrictions, a fieldwork health and safety plan was submitted to and approved by the CEE department.

To identify the optimal point check locations and number of posts needed to cover each corridor, two graduate RAs walked and investigated all the selected corridors prior to assigning posts to field data collectors. Based on the length and typology of corridors, the number of people needed for field data collection ranged between 4 to 10 students per corridor.

Data collection tool development

To make data collection more efficient and to minimize the errors associated with the process, we used an online data collection tool, called KoboToolbox, which data collectors could access through their smartphones. KoboToolbox has three main features:

- The **form builder** is utilized to design forms using an intuitive UI, allowing more than 20 different question types, such as multiple choice, select all that apply, number answer, text answer, date and time stamp, location, image, etc. The form also has skip logic for skipping certain questions based on the answer provided for previous questions as well as validation for answers. Finally, it allows importing and exporting of XLS forms to easily share and edit a form with other XLS-based applications.
- The **data collection tool** uses a web-based application called **Enketo** which works on any modern browser on a phone or tablet to collect data online and offline. Namely, the OS of the phone does not matter, and the app is supported on both Android and iOS smartphones with Safari or Chrome. The data is synchronized via SSL to make sure it cannot be read by a third party. Data is available immediately after it is collected.
- The **data management and analysis** system which allows exporting of data at any time in formats such as XLSX, CSV, ZIP (for media), etc. and accessing data through their API. It also has features for creating summary reports with graphs and tables and to visualize collected data on a map (we did not utilize this feature as we did our post-processing in Python and R).

Our data collection form was designed in a simple fashion that would follow real-world events in the order that is expected to occur. *Figure 23* shows the first page once the app is opened. The form asks for the bus ID or vehicle ID at the very first step. This is a four-digit number which can be clearly seen from the front, rear, and sides of all buses. The four-digit ID uniquely identifies the bus. If the data collector does not see the vehicle ID at first glance, they can attend to the other fields first and return to the vehicle ID field as the bus approaches them. Next, the form allows the user to choose between reporting 1) an interference, 2) an arrival at a bus stop, and 3) both (i.e., if an interference happens during the bus arrival or departure at a bus stop or when the bus is waiting at a stop. This option is enabled in order to mark both events concurrently). *Figures 24(a)* and *24(b)* respectively show the forms when the user selects “Interference” (i.e., interference form) or “Bus Arrival” events (i.e., bus arrival form). The developed tool is available here.

Details of the bus arrival and interference forms, as well as another form used by the RAs to report on their shifts (called shifts form) are provided in Appendix C. Information regarding the observation posts and locations is provided in Appendix D.

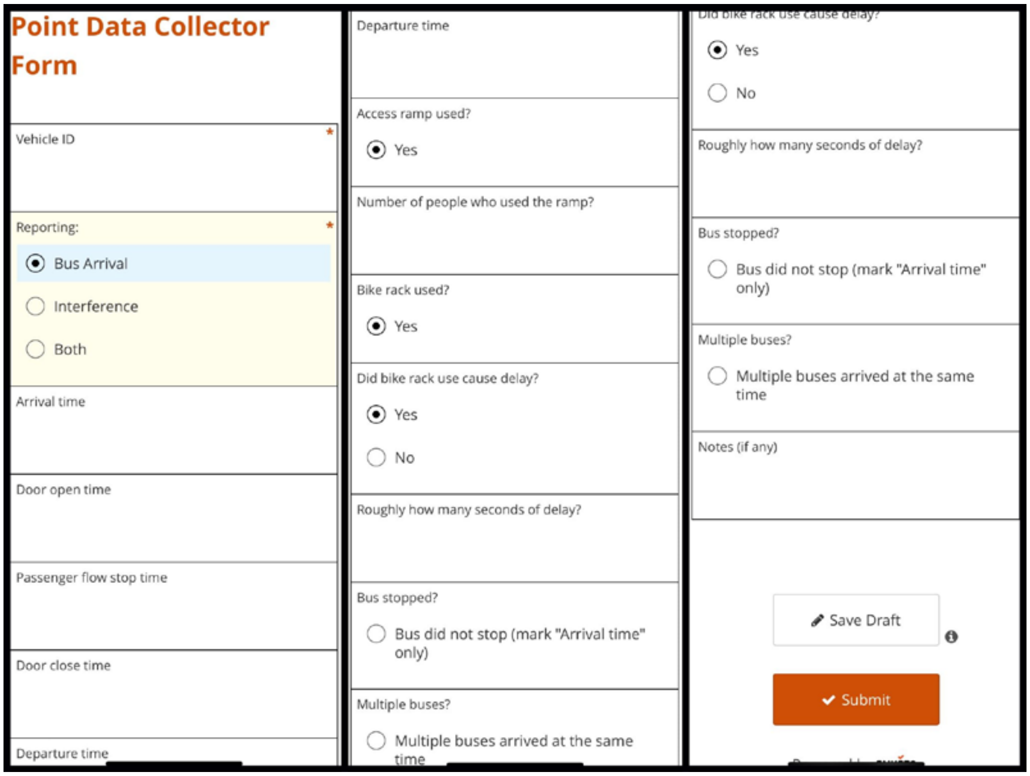
Figure 23: **Point data collector form first page**

Summary statistics

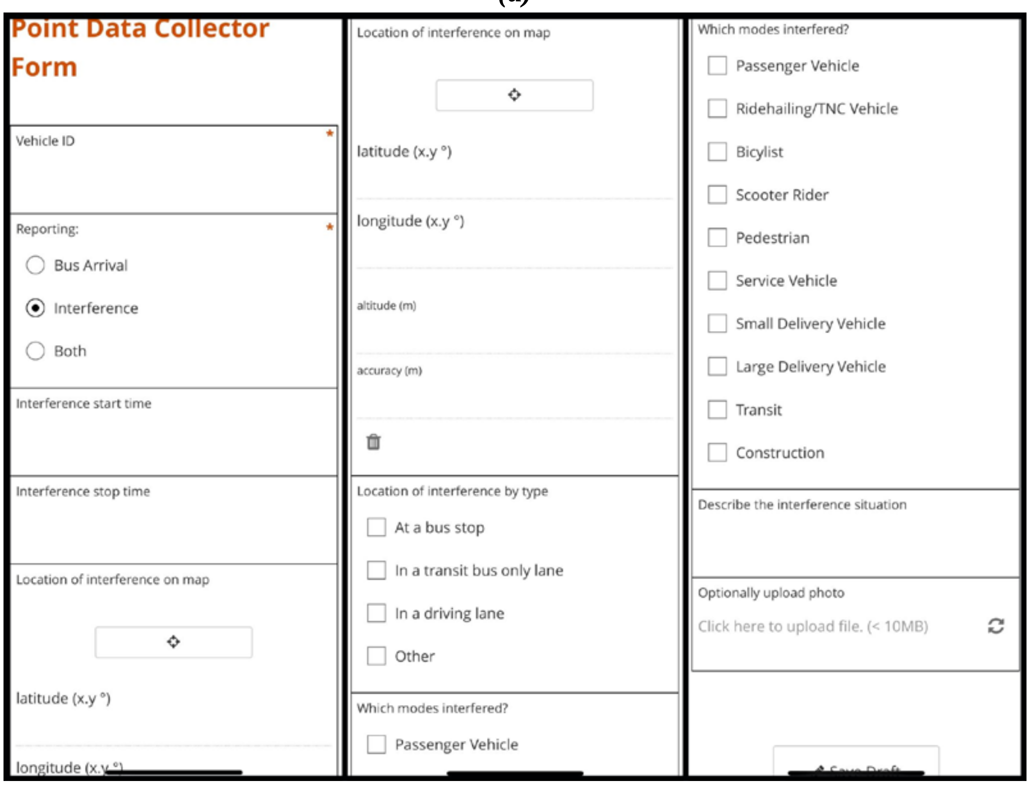
The data collection was done in 12 2.5-hour shifts (9 AM peak and 3 PM peak shifts) between March 3rd to19th, 2021. Out of our 10 transit corridors, we only collected data from eight of them during the three weeks of data collection. The two corridors from which data was not collected are 2nd Avenue between Pike Street and James Street, and Northwest Market/Leary from 24th Avenue to 15th Avenue Northwest. Table 2 has information on the shifts, dates, and the corridors that the data was collected from.

After the data was collected from the field, we cleaned and organized it and created some summary reports. The cleaning steps included reviewing the descriptions of the interference instances, ensuring it matches the selected interfering mode, and correcting the selected mode if needed. We also checked the bus IDs to ensure that the interference instances are unique and not repeated. There was only one instance where the same interference was observed by two different RAs, in which case the two observations were consolidated into one.

Figures 25-30 show some of the summary statistics. A total of 1639 reports were submitted, with 1609 being bus arrival reports at stops, 22 from interferences with other modes, and 8 under the "both" category meaning that an interference happened at the same time the bus arrived at the stop. *Figure 25* shows the number of reports separated by type of report for each transit corridor, and *Figure 26* shows the same data in relative ratios.



(a)



(b)

Figure 24: **Point Data Collector Forms for (a) Bus arrival and (b) Interference events**

Table 2: Dates and periods that data was collected from the corridors

No.	Weekday and Date	Shift Time	Corridor ID	Corridor Name	Total Shifts
1	Wednesday (3/3/21)	7:00-9:30 AM	6	University Way; Campus to 45th	2
2	Wednesday (3/3/21)	3:30-6:00 PM	7	Pacific; 15th to Montlake	2
3	Friday (3/5/21)	7:00-9:30 AM	2	Pike; 3rd to 9th	2
4	Monday (3/8/21)	7:00-9:30 AM	6	University Way; Campus to 45th	2
5	Monday (3/8/21)	3:30-6:00 PM	4	E Olive/E John; Denny to 10th	2
6	Tuesday (3/9/21)	7:00-9:30 AM	7	Pacific; 15th to Montlake	2
7	Wednesday (3/10/21)	7:00-9:30 AM	5	9th Ave; Alder to Columbia	1
8	Wedneday (3/10/21)	3:30-6:00 PM	4	E Olive/E John; Denny to 10th	2
9	Friday (3/12/21)	7:00-9:30 AM	2	Pike; 3rd to 9th	2
10	Tuesday (3/16/21)	7:00-9:30 AM	3	Westlake; Denny to Mercer	1
11	Wednesday (3/17/21)	7:00-9:30 AM	9	Rainier; S Bayview to MLK	1
12	Friday (3/19/21)	7:00-9:30 AM	10	108th Ave; 4th to 12th (Bellevue)	1

Figure 27 shows the frequency of all reports and bus arrival reports submitted per hour, frequency of unique bus IDs traversing the corridor per hour, and frequency of unique interferences per hour per corridor. *Figures 28 and 29* use the same data to show only the frequency of unique bus traversals per hour per corridor and frequency of unique interferences per hour per corridor, respectively. Corridors with the highest amounts of bus traversals had the lowest amount of interference (e.g., Pacific St. corridor and Westlake Ave corridor), whereas those with lower bus traversals generally showed a higher frequency of interference (e.g., Denny Way corridor and 9th Ave corridor).

Figure 30 shows the frequency of interferences by mode per hour per corridor. Pedestrians have the highest frequency of interference; with some of those caused by passengers who boarded or alighted the bus. Construction, transit, passenger vehicles, and service and delivery vehicles also seem to have caused some interference; however, no instance of a ridehailing vehicle interference was observed. Table 8 in Appendix E shows a breakdown of interferences by mode for each corridor along with the description of each interference. The average number of interferences per shift (per 2.5 hours) for morning and evening periods were about 2.1 and 3.7, respectively, meaning that the rate of interference in the evening is about two times that in the morning.

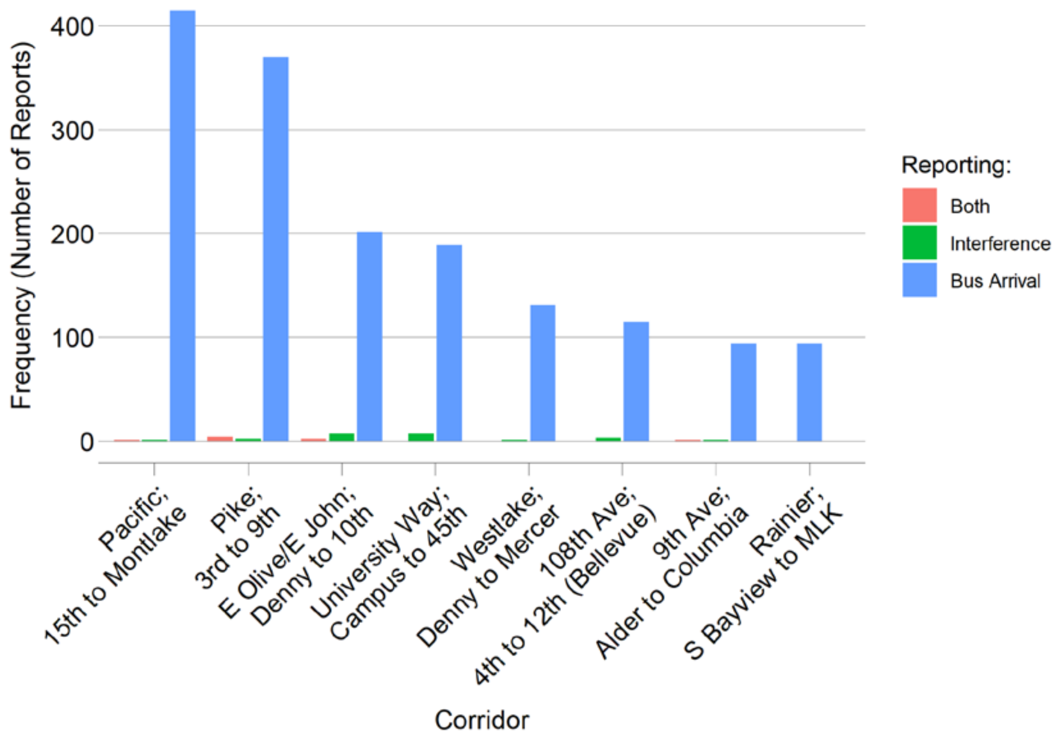


***Figure 25*: Number of reports separated by bus arrival, interference, and both, per corridor.**

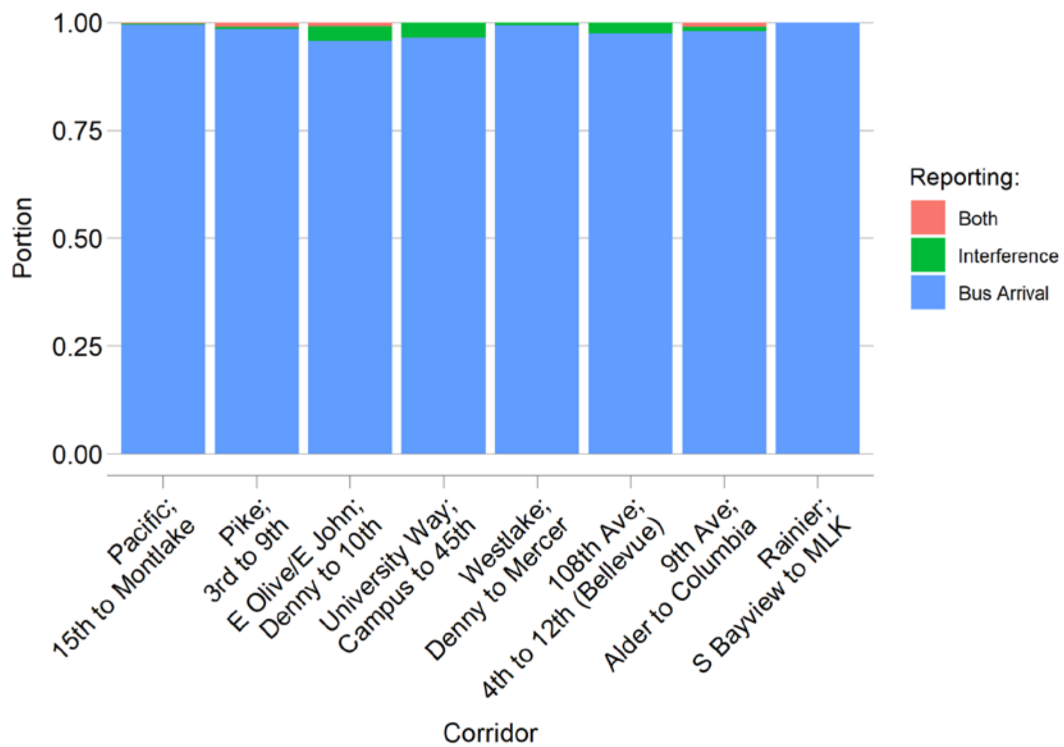


***Figure 26*: Portion of reports by bus arrival, interference, and both, per corridor.**

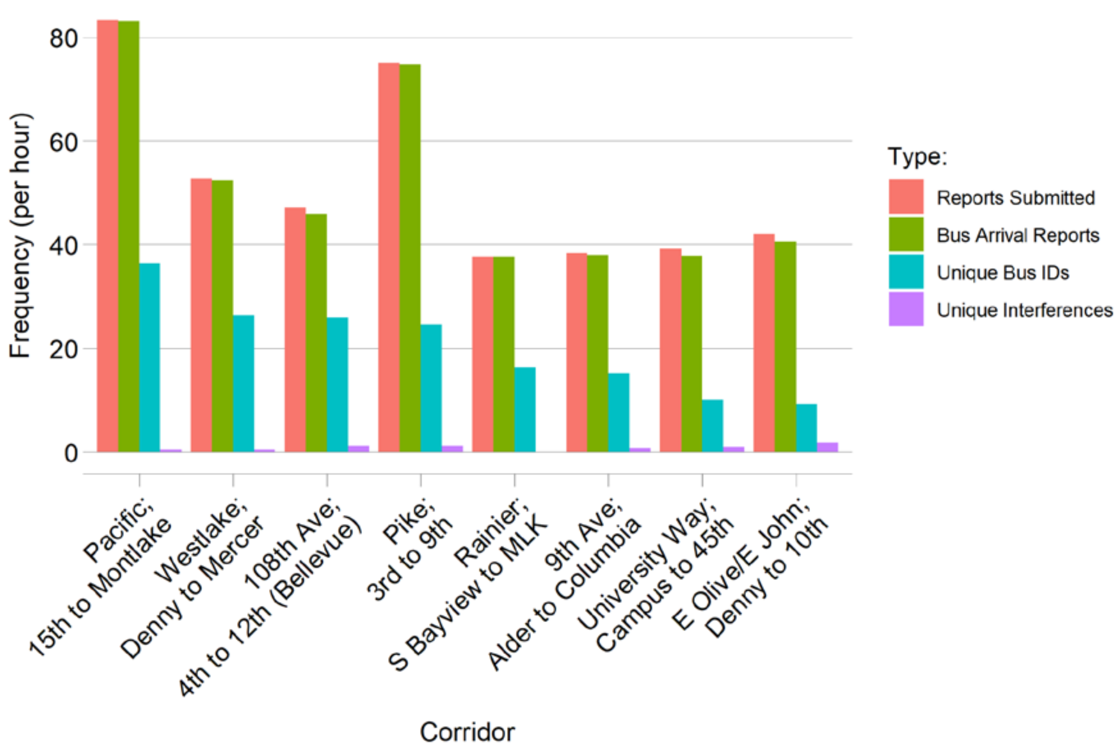


***Figure 27*: Frequency of reports submitted per hour, bus arrival reports, unique bus IDs traversing the corridor, and unique interferences per corridor (sorted by unique bus IDs from high to low).**

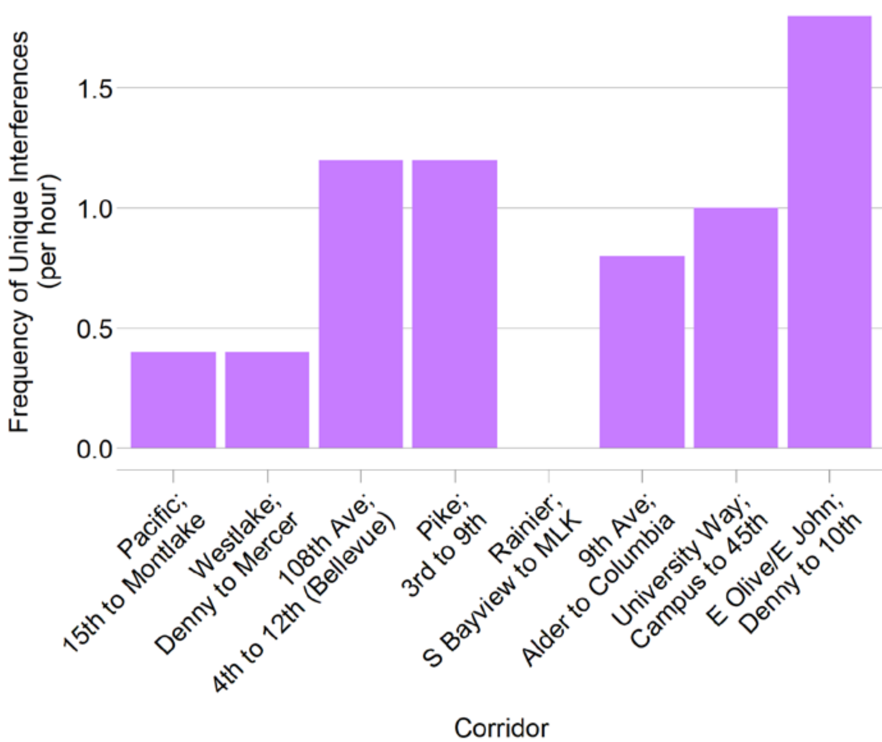


***Figure 28*: Frequency of unique interferences per corridor per hour (sorted by unique bus IDs).**

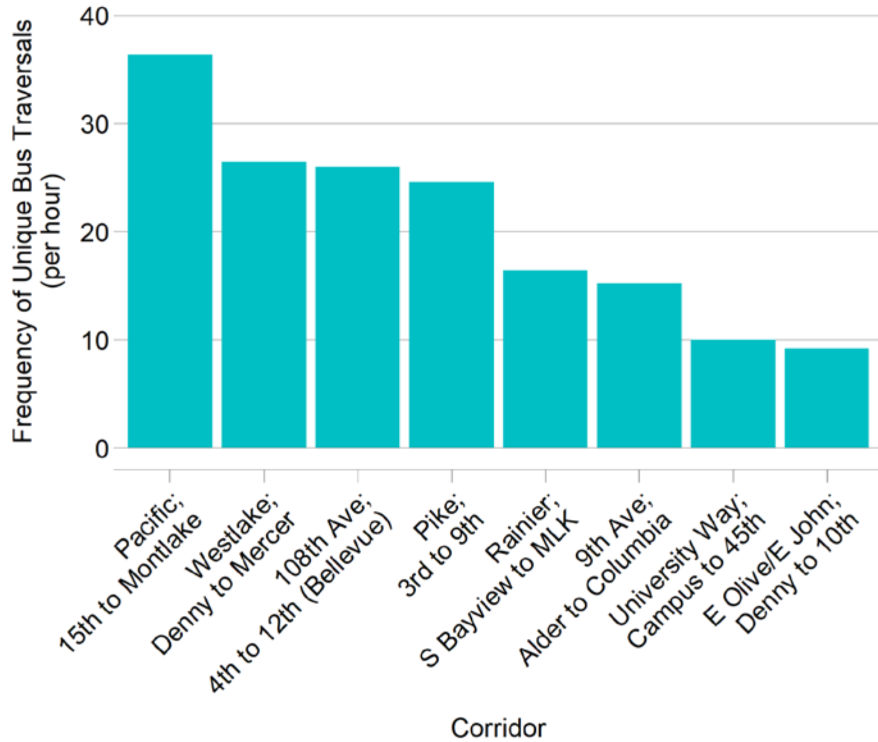


***Figure 29*: Frequency of unique bus ID traversal per corridor per hour (sorted by unique bus IDs).**

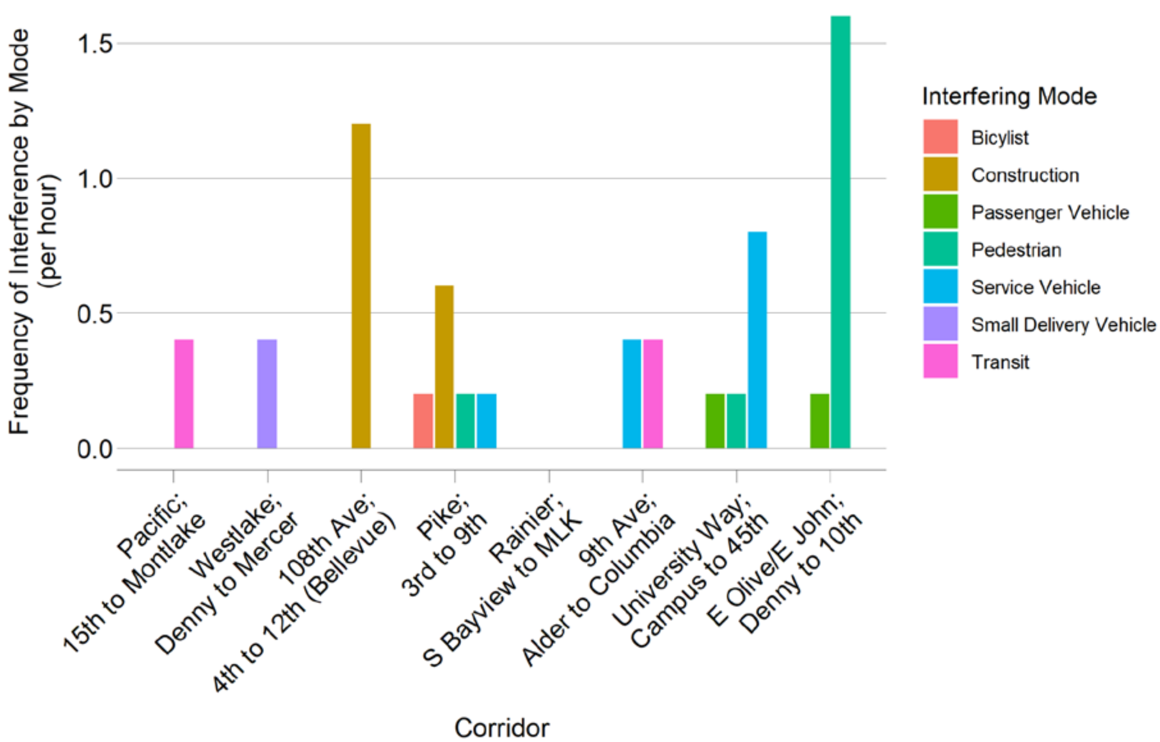


***Figure 30*: Frequency of different interferences by mode per corridor.**

Interference Modeling

Once data collection was performed, and instances of interference were identified from the collected field data, the scraped GTFS-RT data was analyzed to determine whether interference could be identified in it directly. Out of the 30 instances of interference, transit was the interfering mode in three of the cases, and those were excluded for modeling purposes, leaving 27 instances.

One goal of this analysis was to ensure that a connection between in-person data collection and tracked vehicle locations could be rectified despite the somewhat low (~30 second interval) time resolution of the GTFS-RT data and potential loss of precision between reality and the AVL data. In this endeavor, we were able to match all tracked interference that had corresponding real-time data, and so each observation’s location, route, and vehicle ID were aligned with the location and identifiers found in the GTFS-RT data. In some cases, where the interference affected a vehicle not equipped with real-time tracking technology, the interference could not be located in the GTFS-RT data. Out of the 27 instances of interference, nine were not matched to real-time data, leaving 18 instances. The other objective of this analysis was to train a model capable of identifying bus interference using only the GTFS-RT data, by using the in-person data collection as ground truth validation data. This work is still in progress and requires a larger volume of ground truth data.

Preliminary analysis revealed that there is not a substantial effect shown in the GTFS-RT data when a bus is interfered with; however, there were not a lot of interference observations in the collected field data. So, it remains to be seen whether the lack of an identifiable effect is due to the lack of ground truth data, lack of precision in the AVL collection system, or the relatively low impact of an interference when compared to the effects of general traffic congestion, signals, and other roadway conditions. A follow-up linear regression model was generated to determine the extent to which roadway characteristics can predict segment performance, with mildly predictive results.

Effects of Interference Observed in Real-Time Performance Data

To determine whether the identified interferences created a noticeable delay for vehicles traversing the study corridor, distributions of vehicle speed, schedule deviation, and travel time for the delayed vehicle traversing the study corridor within +/-2 minutes of the point of interference were generated and presented in *Figures 31-33*. The distributions show both the median performance metric during the specified time period (blue bar) as well as the value of the metric at the point in time where the interference occurred (red bar). If the interference can be identified using the GTFS-RT data, we would expect the red bar to be consistently to the left of the blue bar in *Figure 31* (the vehicle speed metric) and to the right of the blue bar in *Figure 32-33* (the schedule deviation and travel time metrics); indicating a negative influence on the vehicle’s performance at the point of interference. However, that is not what is observed in the figures.

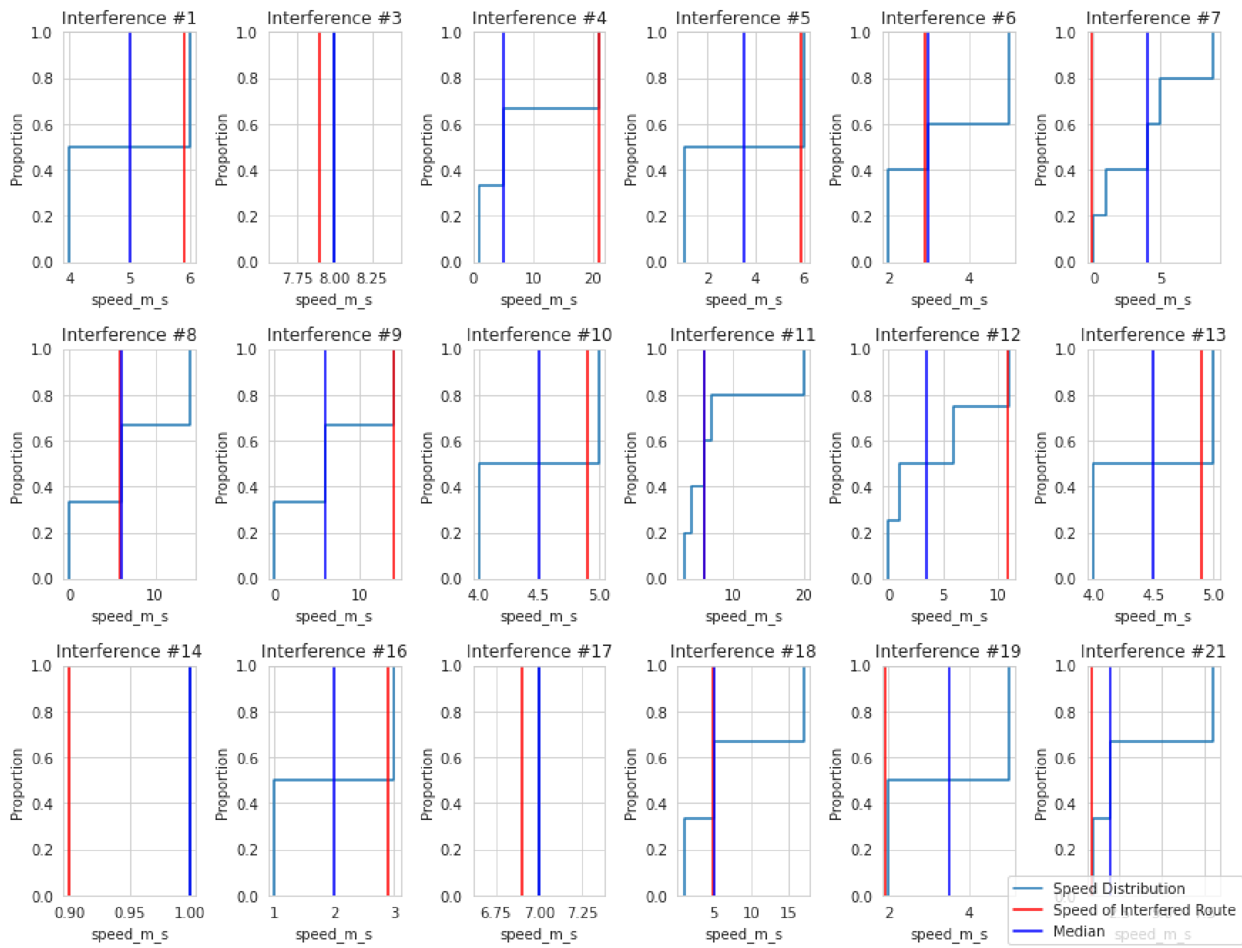


***Figure 31*: Effects of interference on vehicle speed. Some interferences were not matched to a vehicle with real-time tracking.**

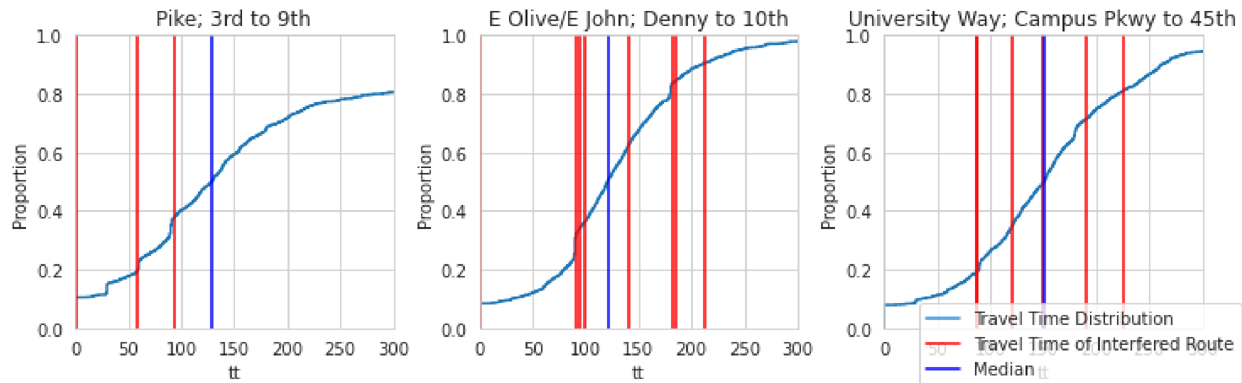


***Figure 32*: Distribution of travel times for all vehicles traversing the study corridors during the days of field data collection. Only corridors with recorded instances of field interference are shown. Some interferences were not matched to a vehicle with real-time tracking.**

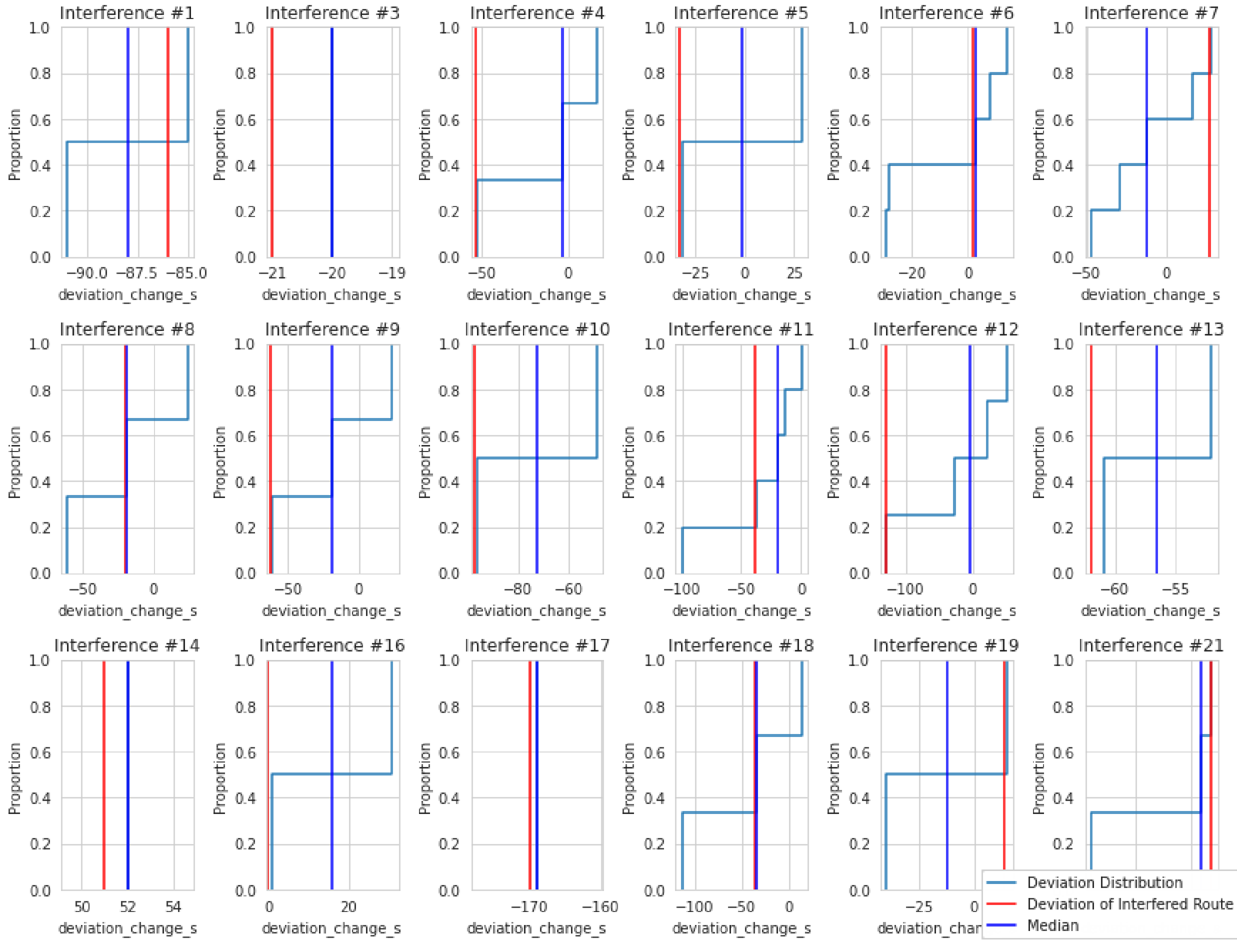


***Figure 33*: Effects of interference on vehicle schedule deviation. Some interferences were not matched to a vehicle with real-time tracking.**

In addition to the performance distributions for individual vehicles, we also aggregated the interferences to see whether across multiple interferences and transit vehicles, there would be a discontinuity in the time-series data for the schedule deviation metric. This would indicate that an unplanned delay occurred, given that schedule deviation should account for congestion or other predictable effects through schedule padding. The results of this are shown in Figure 34. Each line represents the trend in schedule deviation for a vehicle +/-2 minutes from the point of interference. The black line indicates a LOESS curve fit to all of the combined trajectories. The left portion of the curve is fit to data from before the interference, and the right portion is fit to data after the interference. If the instances of interference had an impact on the GTFS-RT data, it would be expected that the curve would shift upward at its discontinuity, indicating an increase in schedule deviation as the vehicles fall behind schedule, which is not observed in the preliminary data.

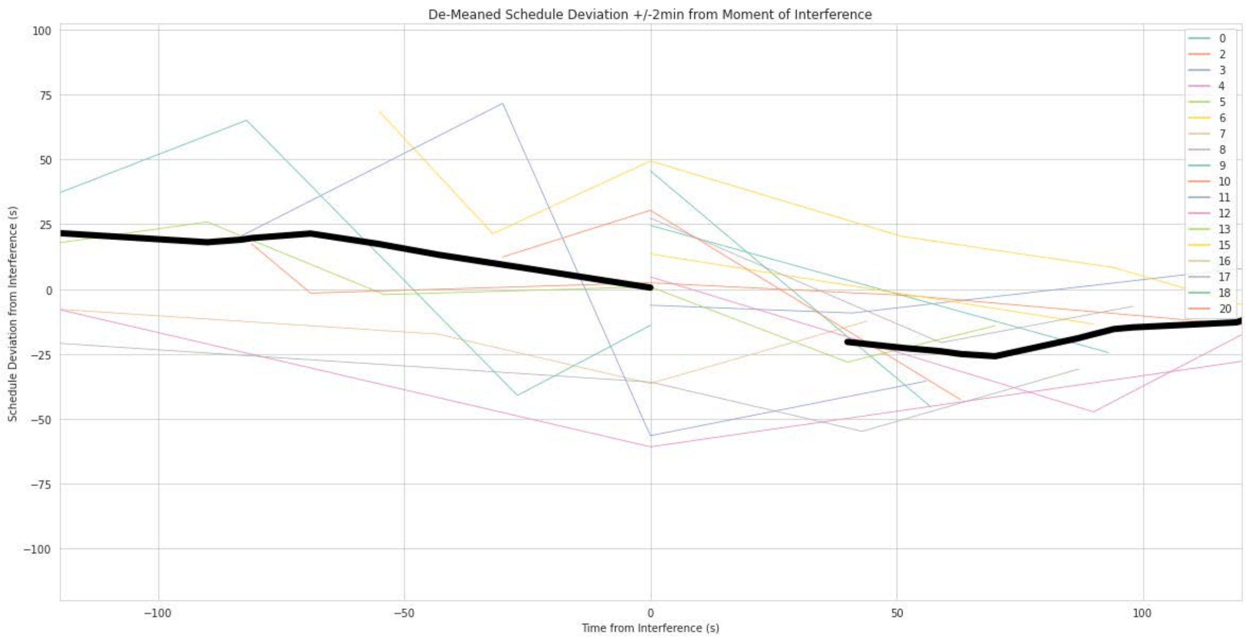


***Figure 34*: Trajectories of schedule deviation for all interfered vehicles +/-2 minutes from the point of interference. Black lines show fitted LOESS curves pre- and post-interference. The zero value on the x-axis represents the time of the tracked location of a vehicle immediately prior to its interference.**

Exploratory Modeling of Segment Performance with Urban Characteristics

We also built models to determine whether the performance of transit segments could be predicted using urban characteristics of the segment and its surrounding land use. Bus route, stop, intersection, and roadway segment locations and characteristics were accessed through the City of Seattle Open Data Portal. This resource provides shapefiles and other datasets that can be spatially linked to segment performance using GIS software.

The roadway segments which form the most disaggregate units of analysis for this work were drawn from the "Seattle Streets" dataset, which can be linked to other available datasets such as Average Weekday Daily Traffic (AWDT) and existing bike facilities using a shared "compkey" identifier. Each segment in the dataset is approximately one block long and covers all lanes traveling in one direction of a given roadway. Due to limited availability of roadway characteristic data, only segments in Seattle and its immediate vicinity were analyzed. However, these models can be easily extended to other areas if additional roadway data for them is available.

A mix of continuous and categorical variables were used in a linear regression model, attempting to predict the performance of each transit segment. Performance was quantified as the ratio of median to 95th percentile (free flow) speeds as observed across one month of GTFS-RT data. To model categorical variables, a reference class is identified and the effects of the other classes are measured with respect to that class. In Table 3, the reference class for each categorical variable is the first class listed. All variable data was gathered from the City of Seattle's Open Data Portal.

Table 3: Urban Characteristics

Variable Type	Variable Name	Classes	Source
Continuous	Speed Limit	N/A	https://data.seattle.gov/dataset/Seattle-Streets/b856-55i2
Continuous	Segment Length	N/A	https://data.seattle.gov/dataset/Seattle-Streets/b856-55i2
Continuous	Combined Lane Width	N/A	https://data.seattle.gov/dataset/Seattle-Streets/b856-55i2
Continuous	Segment Slope	N/A	https://data.seattle.gov/dataset/Seattle-Streets/b856-55i2
Continuous	Average Weekday Daily Traffic	N/A	https://data.seattle.gov/dataset/2018-Traffic-Flow-Counts/mrnk-ws8x
Continuous	Number of Commercial Vehicle Load Zones	N/A	https://data-seattlecitygis.opendata.arcgis.com/datasets/SeattleCityGIS::blockface/about
Continuous	Number of Parking Zones	N/A	https://data-seattlecitygis.opendata.arcgis.com/datasets/SeattleCityGIS::blockface/about
Continuous	Hourly Combined Passenger Pickup and Dropoff Rate	N/A	SharedStreets
Binary	One Way	-One Way -Not One Way	https://data.seattle.gov/dataset/Seattle-Streets/b856-55i2
Categorical	Arterial Classification	-Minor Arterial -Collector Arterial -Principal Arterial	https://data.seattle.gov/dataset/Seattle-Streets/b856-55i2
Categorical	Street Classification	-Neighborhood Corridor -Downtown -Downtown Neighborhood -Industrial Access -Urban Center Connector -Urban Village Main -Urban Village Neighborhood -Urban Village Neighborhood Access	https://data.seattle.gov/dataset/Seattle-Streets/b856-55i2
Categorical	Transit Classification	-Not Designated -Local Transit Route -Major Transit Route -Minor Restricted Transit Route -Minor Transit Route -Principal Transit Route -Temporary Transit Route	https://data.seattle.gov/dataset/Seattle-Streets/b856-55i2
Categorical	Bike Facility Type	-No Facility -In Street, Major Separation -In Street, Minor Separation -Multi-Use Trail -Sharrow	https://data.seattle.gov/dataset/Existing-Bike-Facilities/pth5-3a9v

Continuous variables (both dependent and independent) were examined for outliers (*Figure 35*) and pairwise scatterplots were constructed for each combination of variables to look for cases of strong multicollinearity (*Figure 36*). The presence of extreme outliers was not found; however, we observed a relationship between street width and AWDT, which makes sense as streets with more lanes naturally support larger traffic flows and built in areas with more traffic.

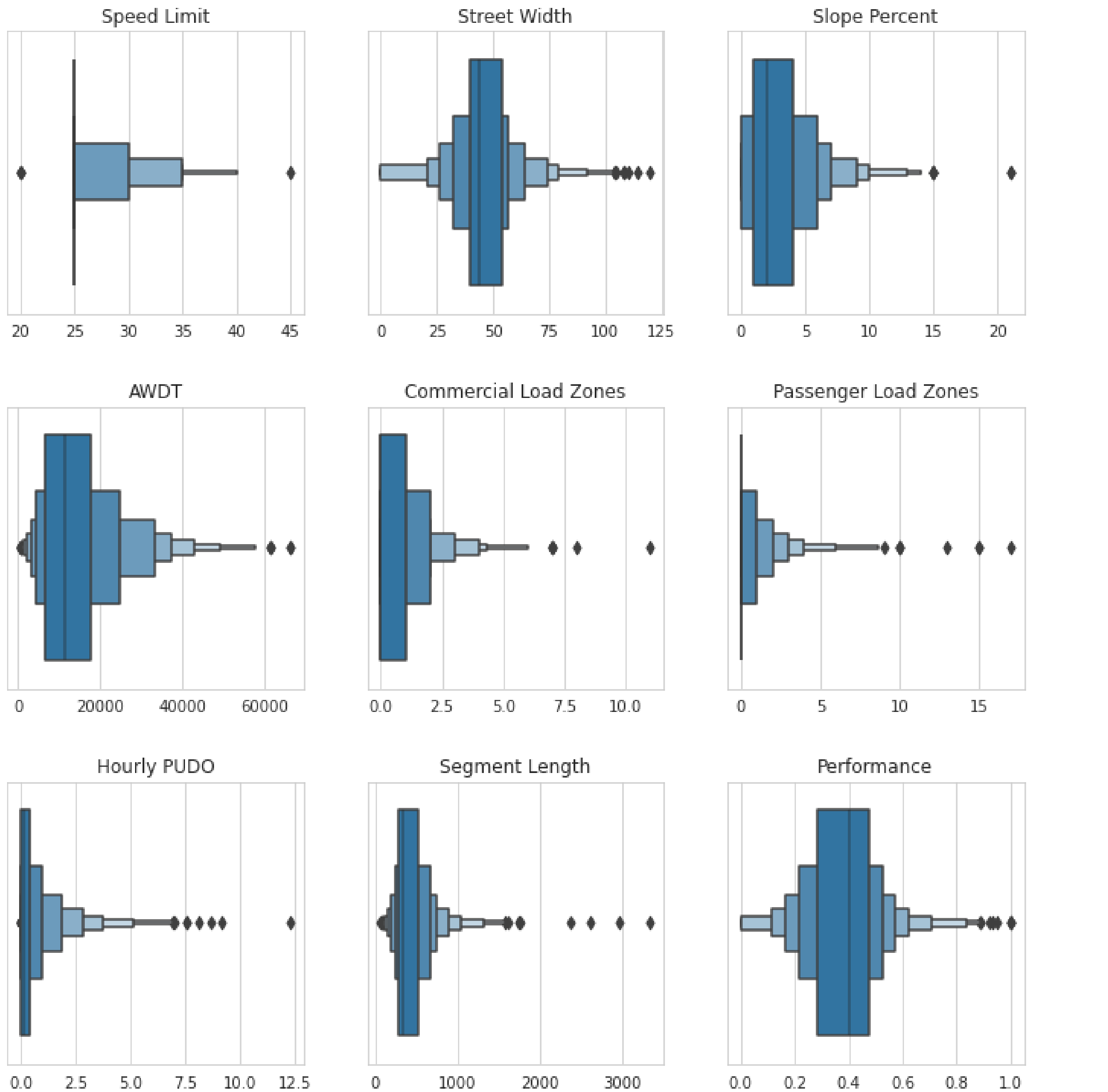


***Figure 35*: Checking for outlier segments across continuous independent variables and the dependent variable.**

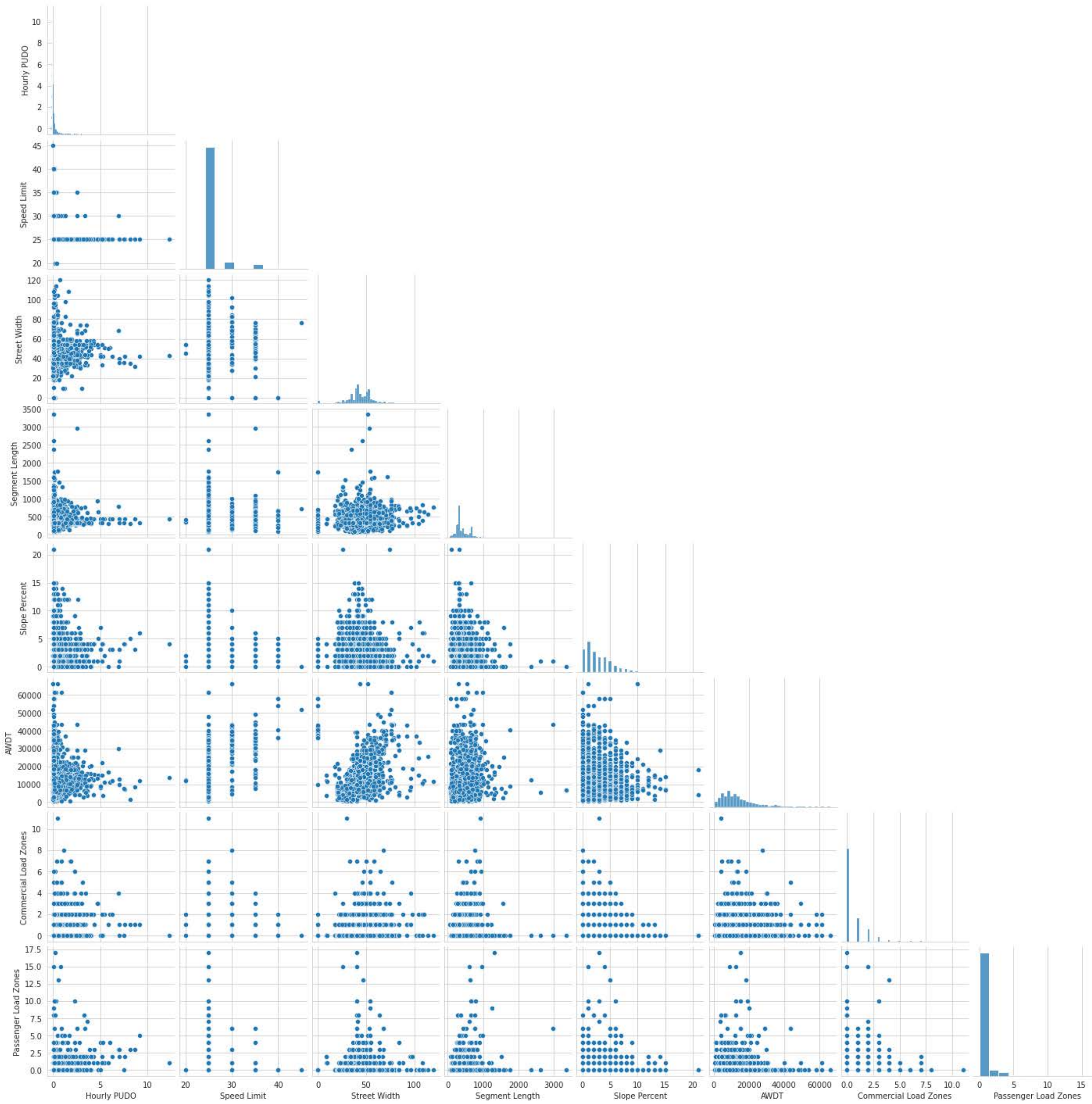


***Figure 36*: Checking for correlation between independent variables.**

In total, 1,878 segments were analyzed, and the results of a regression on their characteristics are shown in Table 4. Ordinary Least Squares (OLS) was used to fit the linear regression, and a number of input variables were found to be significant, such as street type, transit description, bike facility type, speed limit and AWDT. Perhaps most interesting is the finding that in-street bike facilities increase the performance of a segment relative to no facility. However, the existence of a shared facility was not a statistically significant detriment to transit performance. Overall, with an adjusted R-squared value of 0.126, the model based on urban characteristics was found to be only mildly predictive of the performance for a given roadway segment.

Table 4: OLS Analysis Results (N = 1878, Adjusted R-Squared = .126)

Variable	Reference	Class	Coefficient	P-Value
Intercept			0.2654	**0.0000**
Arterial Description	Minor Arterial	Collector Arterial	0.0112	0.3170
		Principal Arterial	-0.0137	0.3280
One Way	No	Yes	-0.0033	0.7820
Street Type	Neighborhood Corridor	Downtown	-0.0979	**0.0000**
		Downtown Neighborhood	-0.0947	**0.0000**
		Industrial Access	-0.0180	0.3750
		Urban Center Connector	0.0148	0.2990
		Urban Village Main	-0.0600	**0.0010**
		Urban Village Neighborhood	-0.0566	**0.0000**
		Urban Village Neighborhood Access	-0.1249	0.3630
Transit Description	Not Designated	Local Transit Route	-0.0887	0.2670
		Major Transit Route	0.0396	**0.0230**
		Minor Restricted Transit Route	0.0205	0.7980
		Minor Transit Route	0.0331	**0.0490**
		Principal Transit Route	0.0168	0.4330
		Temporary Transit Route	-0.1013	0.3050
Existing Bike Facility Type	No Facility	In Street, Major Separation	0.0428	**0.0010**
		In Street, Minor Separation	0.0553	**0.0000**
		Multi-use Trail	-0.0025	0.9850
		Sharrow	0.0071	0.4410
Speed Limit			0.0043	**0.0090**
Segment Length			-0.00001	0.4250
Street Width			-0.0005	0.0870
Slope Percent			0.0013	0.2970
AWDT			0.0000015	**0.0060**
Commercial Load Zones			0.0040	0.2310
Passenger Load Zones			-0.0011	0.6790
Hourly PUDO			-0.0026	0.5220

Next Steps

We have thus far put together a framework for screening and analyzing transit delays on segments in the transit network, but found little overlap with preliminary interference field data collection. This may be due to depressed levels of transportation activity during the COVID-19 pandemic. Given the current unstable condition of travel choices and city traffic (and thus interferences), we decided to pause the field data collection until late Summer or early Fall 2021. As businesses and transit services continue to reopen, there will likely be an increase in the amount of transit interference experienced between buses and other roadway users, which will potentially allow for the gathering of more ground truth validation data. Over a larger dataset, we may see the methods outlined in this report produce more consistent coordination between transit delays in the GTFS-RT and field data collection. We plan to resume field data collection in late summer or early Fall of 2021. The GTFS-RT data scraping will continue daily, and summarized in the developed interactive visualization tool.

References

1. Abdelfattah AM, Khan AM. Models for Predicting Bus Delays. Transportation Research Record. 1998 Jan 1;1623(1):8–15.
2. Alfa AS, Menzies WB, Purcha J, McPherson R. A regression model for bus running times in suburban areas of winnipeg. Journal of Advanced Transportation. 1988;21(3):227–37.
3. Jeong R, Rilett R. Bus arrival time prediction using artificial neural network model. In: Proceedings The 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat No04TH8749). 2004. p. 988–93.
4. El-Geneidy AM, Horning J, Krizek KJ. Analyzing transit service reliability using detailed data from automatic vehicular locator systems. Journal of Advanced Transportation. 2011;45(1):66–79.
5. Yu B, Lam WHK, Tam ML. Bus arrival time prediction at bus stop with multiple routes. Transportation Research Part C: Emerging Technologies. 2011 Dec 1;19(6):1157–70.
6. Kalaputapu JR. Modeling Schedule. Deviations of Buses Using Automatic Vehicle-Location Data and Artificial Neural Networks. TRANSPORTATION RESEARCH RECORD. :9.
7. Chien SI-J, Ding Y, Wei C. Dynamic Bus Arrival Time Prediction with Artificial Neural Networks. Journal of Transportation Engineering. 2002 Sep 1;128(5):429–38.
8. Amita J, Sukhvir Singh J, Pradeep Kumar G. PREDICTION OF BUS TRAVEL TIME USING ARTIFICIAL NEURAL NETWORK. IJTTE. 2015 Dec;5(4):410–24.
9. Hellinga B, Yang F, Hart-Bishop J. Estimating Signalized Intersection Delays to Transit Vehicles: Using Archived Data from Automatic Vehicle Location and Passenger Counting Systems. Transportation Research Record. 2011 Jan 1;2259(1):158–67.
10. Coghlan C, Dabiri S, Mayer B, Wagner M, Williamson E, Eichler M, et al. Assigning Bus Delay and Predicting Travel Times using Automated Vehicle Location Data. Transportation Research Record. 2019 Mar 1;2673(3):624–36.
11. Raghothama J, Shreenath VM, Meijer S. Analytics on public transport delays with spatial big data. In: Proceedings of the 5th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data [Internet]. Burlingame, California: Association for Computing Machinery; 2016 [cited 2020 Jul 13]. p. 28–33. (BigSpatial '16). Available from: https://doi.org/10.1145/3006386.3006387
12. Wessel N, Widener MJ. Discovering the space–time dimensions of schedule padding and delay from GTFS and real-time transit data. J Geogr Syst. 2017 Jan 1;19(1):93–107.
13. Lin W-H, Bertini RL. Modeling schedule recovery processes in transit operations for bus arrival time prediction. Journal of Advanced Transportation. 2004;38(3):347–65.
14. Watkins KE, Ferris B, Borning A, Rutherford GS, Layton D. Where Is My Bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders. Transportation Research Part A: Policy and Practice. 2011 Oct;45(8):839–48.

Appendix A: Details of Crowdsourced Survey Data

This section discusses the results of our survey and how we performed data cleaning to achieve those results. The four questions of the survey were:

- Where did the observed interference occur? (e.g. 3rd St between Virginia and Yesler)
- Which bus routes were affected by this interference? (e.g. 42, 71, RapidRide E)
- What was the primary cause of this interference?
    - Options: 1) Pedestrians; 2) Bicyclists; 3) Delivery Trucks/Commercial Vehicles; 4) Passenger Vehicles; 5) Ridehailing Vehicles (Taxis, Lyft, Uber, etc.)
- What time of day did the interference occur?

The survey received 78 responses. The most reported cause of interference was overwhelmingly passenger vehicles, with ridehailing services and other being the next most common responses. This could in part be a result of respondents not being able to distinguish between TNCs and private vehicles. It may also be a result of misconstruing “interference” to simply mean “traffic”. *Figure 37* shows the results for the causes of interference.

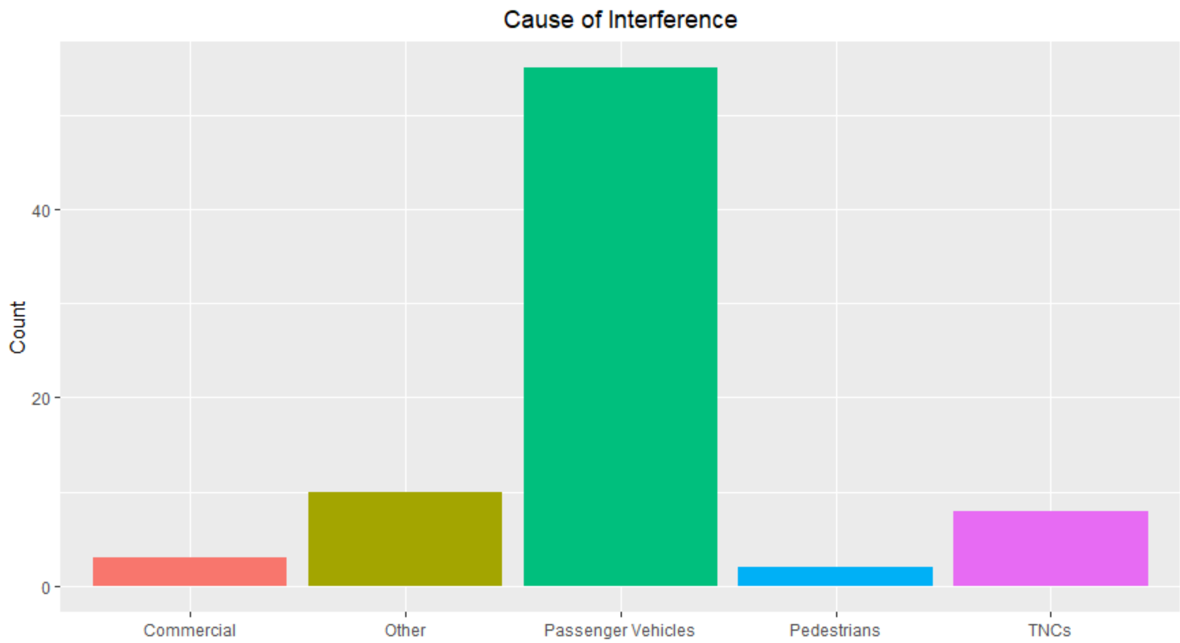


Figure 37: **Causes of interference according to survey respondents.**

Respondents also observed two clear peak periods for transit interference. We see a less severe but slightly more sustained morning peak between 6-10AM, with a more intense afternoon peak between approximately 3-7PM. This coincides fairly well with standard peak hours, suggesting that most interference occurs (or is observed to occur) while respondents are commuting. *Figure 38* shows a histogram of the interference time distribution provided in the survey responses, and *Figure 39* illustrates a density estimation for the time of interference.

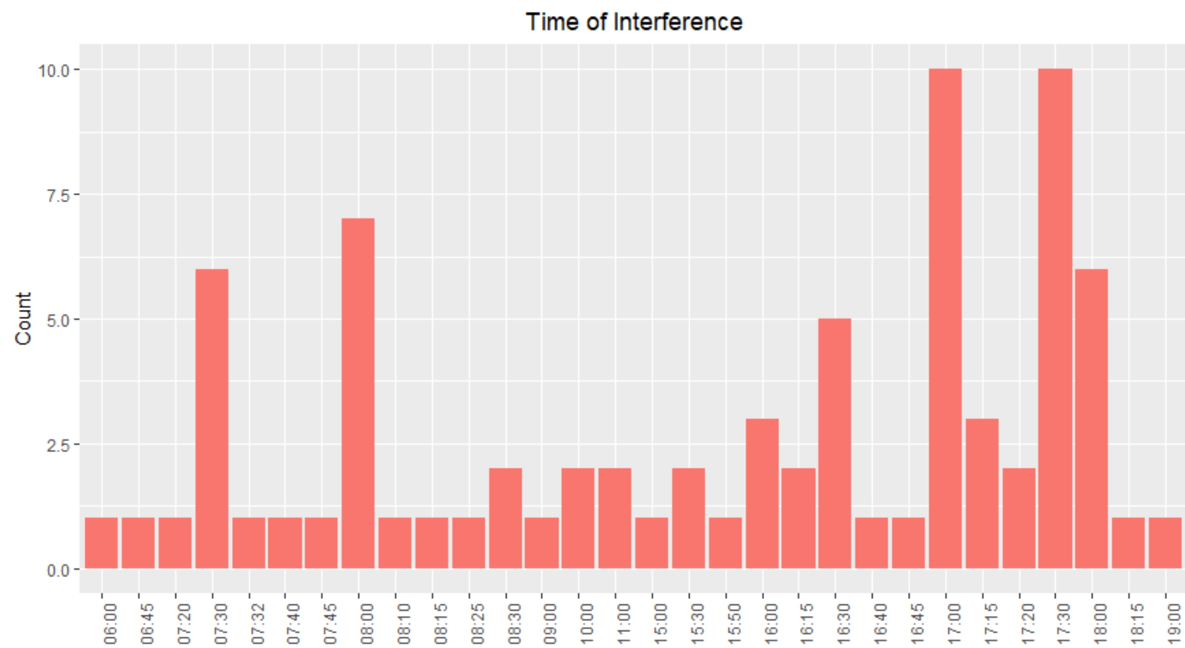


Figure 38: **Most frequent times of interference according to survey respondents.**

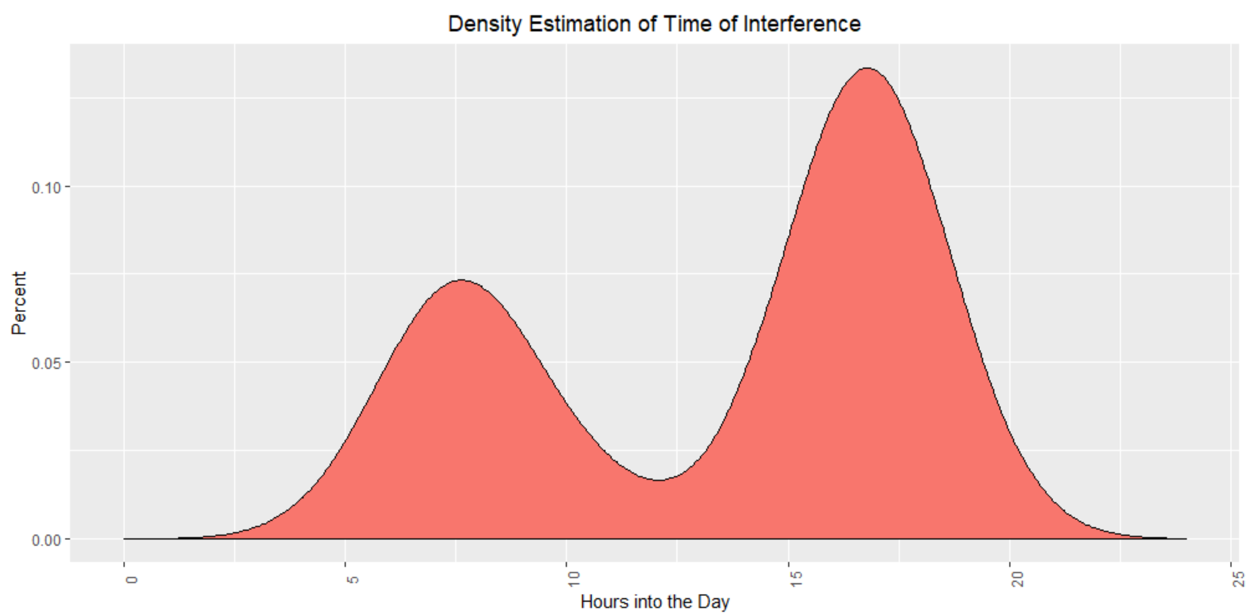


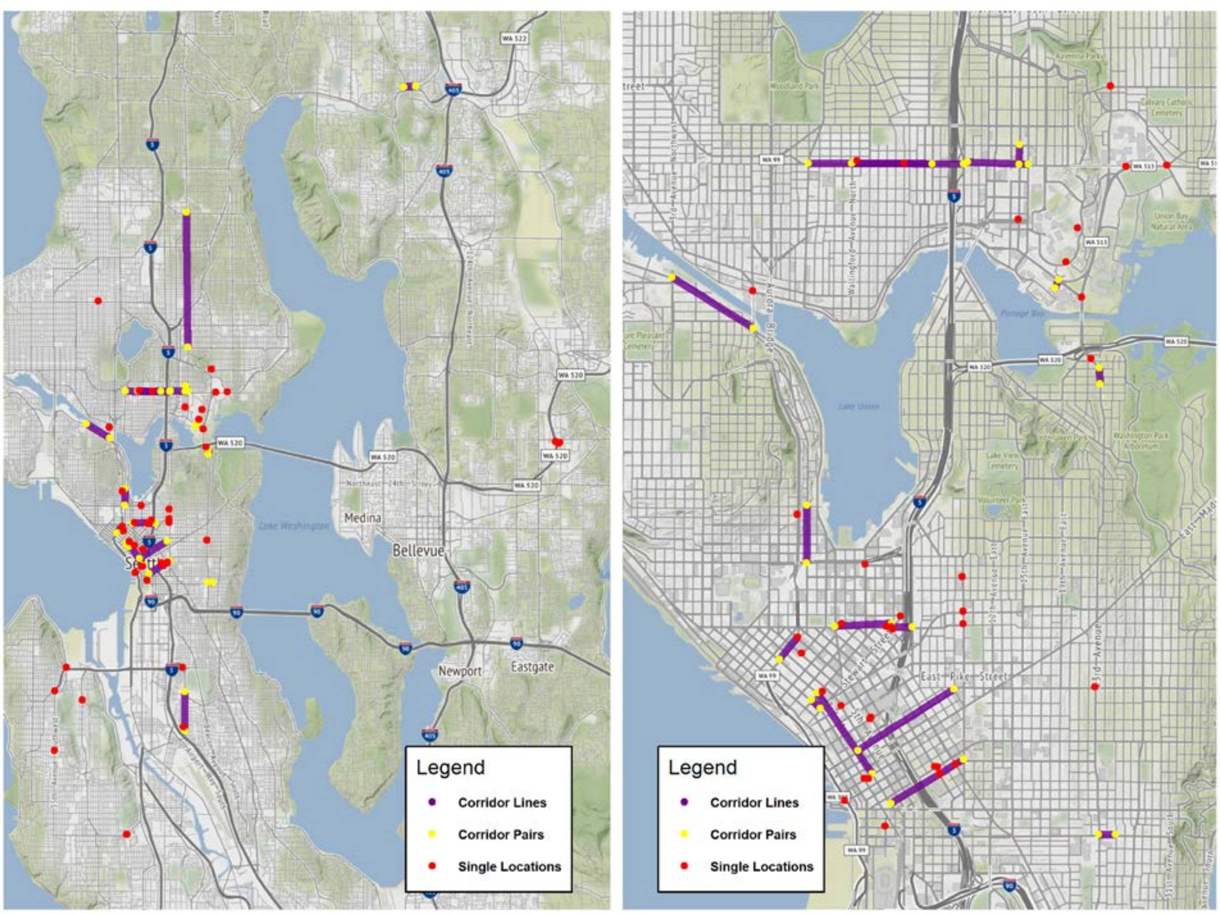
Figure 39: **Estimated distribution of interference times based on survey responses.**

To identify transit corridors with high interference, we cleaned responses to the first question (i.e., where did the observed interference occur?) as follows. We first removed 12 irrelevant or inaccurate responses that: (1) was too broad and did not specified a specific location (e.g. U District); (2) was irrelevant (e.g. the whole ride, all over the place, etc.); or (3) had an unclear or very long corridor (e.g. 405 North between Renton and Newcastle, which is a highway piece longer than 6 miles). After removing these locations, a total of 66 valid responses remained.

Then, the remaining data was categorized into two groups. The first group included responses specifying a single location such as a certain bus stop or an intersection (e.g. “Harbor Ave SW & SW Spokane St”). The second group included responses specifying a corridor or block (e.g. “Jefferson St between 3rd Ave and Broadway”). Out of the 66 remaining valid entries, 47 were single-location responses and 19 were corridors. Responses of the second group were then translated into a pair of two points (the two bounds of the specified corridor). All points were written in a standard address format as follows: *street address* or *two crossing streets* + *city* + *state* + *zip code*.

The cleaned surveyed locations are plotted in *Figure 40*. *Figure 40(a)* shows all responses in the Greater Seattle area including a small number of responses in Redmond, Bothell, and White Center. *Figure 40(b)* focuses on areas with the highest number of responses mostly being in Downtown Seattle, South Lake Union, and areas in North and Northeast Seattle. Red points represent single-location responses, and yellow points represent the two ends of a corridor response, which is shown by a purple line. This data was then combined with other data sources to identify the study corridors, which are mentioned in Appendix B.

The cleaned survey response addresses are listed in Tables 5 and 6. Table 5 presents the responses which specified single locations and Table 6 lists the two ends of the responses which specified corridors.



(a) *(b)*

Figure 40: (a) **All survey responses in the Greater Seattle area; *(b)* Areas with higher number of responses (e.g. Downtown, SLU, North, and Northeast Seattle)**

Table 5: Cleaned survey results for single-location responses

No.	Single-Location Addresses
1	156th Ave NE & Overlake Transit Center, Redmond, WA 98052
2	2nd Ave S & S Main St, Seattle, WA 98104
3	35th Ave SW & SW Avalon Way, Seattle, WA 98126
4	4001 E Stevens Way NE, Seattle, WA 98195
5	6th Ave & Union St, Seattle, WA 98101
6	523 Union St, Seattle, WA 98101
7	9th Ave & James St, Seattle, WA 98104
8	901 James St, Seattle, WA 98104
9	Alaskan Way & Columbia St, Seattle, WA 98104
10	Aurora Ave N & Prospect St, Seattle, WA 98109
11	Bell St & 5th Ave, Seattle, WA 98121
12	Borealis Ave & 6th Ave, Seattle, WA 98121
13	Broadway E & E Republican St, Seattle, WA 98102
14	5511 15th Ave S, Seattle, WA 98108
15	Delridge Way SW & SW Alaska St, Seattle, WA 98106
16	964 Denny Way, Seattle, WA 98121
17	E Denny Way & Broadway E, Seattle, WA 98102
18	E John St & Broadway E, Seattle, WA 98102
19	E Union St & 23rd Ave, Seattle, WA 98122
20	East Montlake Pl E & E Roanoke St, Seattle, WA 98112
21	Fairview Ave N & Mercer St, Seattle, WA 98109
22	Jefferson St & 9th Ave, Seattle, WA 98104
23	Jefferson St & Boren Ave, Seattle, WA 98104
24	Marion St & 2nd Ave, Seattle, WA 98174
25	205 Marion St, Seattle, WA 98104
26	N 45th St & Eastern Ave N, Seattle, WA 98103
27	1820 N 45th St, Seattle, WA 98103
28	Montlake Blvd NE & NE 45th St, Seattle, WA 98105
29	NE 45th St & Union Bay Pl NE, Seattle, WA 98105
30	NE 55th St & 25th Ave NE, Seattle, WA 98105
31	N 85th St & Greenwood Ave N, Seattle, WA 98103
32	NE Campus Pkwy & University Way NE, Seattle, WA 98105
33	NE Pacific St & Montlake Blvd NE, Seattle, WA 98105
34	Overlake Transit Center, Redmond, WA 98052
35	Pine St & 4th Ave, Seattle, WA 98181

Table 5 (cont'd)

No.	Single-Location Addresses
36	SW Roxbury St & 5th Ave SW, White Center, WA 98106
37	S Columbian Way & S Spokane St, Seattle, WA 98144
38	SR 520 & NE 40th St, Redmond, WA 98052
39	Stewart St & Denny Way, Seattle, WA 98109
40	1828 Yale Ave, Seattle, WA 98101
41	Stewart St & John St, Seattle, WA 98109
42	SW Avalon Way & SW Spokane St, Seattle, WA 98126
43	SW Morgan St & 35th Ave SW, Seattle, WA 98126
44	1930 3rd Ave, Seattle, WA 98101
45	W Stevens Way NE & Rainier Vista, Seattle, WA 98195
46	Yale Ave & Denny Way, Seattle, WA 98109
47	Fremont Ave N & N 34th St, Seattle, WA 98103

Table 6: Cleaned survey results for corridor responses. The two addresses listed for each corridor represent the two ends of a suggested corridor in the survey.

No.	Corridor Addresses
1	15th Ave S & S Orcas St, Seattle, WA 98108
	15th Ave S & S Columbian Way, Seattle, WA 98144
2	2nd Ave & Stewart St, Seattle, WA 98101
	2nd Ave & Virginia St, Seattle, WA 98101
3	3rd Ave & Marion St, Seattle, WA 98104
	3rd Ave & Virginia St, Seattle, WA 98101
4	Battery St & 3rd Ave, Seattle, WA 98121
	Battery St & 6th Ave, Seattle, WA 98121
5	Denny Way & Westlake Ave, Seattle, WA 98109
	1304 Stewart St, Seattle, WA 98109
6	Dexter Ave N & Highland Dr, Seattle, WA 98109
	Dexter Ave N & Mercer St, Seattle, WA 98109
7	E Louisa St & E Montlake Pl E, Seattle, WA 98112
	24th Ave E & E Calhoun St, Seattle, WA 98112
8	Jefferson St & 3rd Ave, Seattle, WA 98104
	Jefferson St & Broadway, Seattle, WA 98104
9	Main St & 103rd Ave NE, Bothell, WA 98011
	Main St & Bothell Way NE, Bothell, WA 98011
10	N 45th St & Wallingford Ave N, Seattle, WA 98103
	NE 45th St & 7th Ave NE, Seattle, WA 98105
11	706 NE 45th St, Seattle, WA 98105
	NE 45th St & 15th Ave NE, Seattle, WA 98105
12	N 45th St & Stone Way N, Seattle, WA 98103
	NE 45th St & Latona Ave NE, Seattle, WA 98105
13	NE 65th St & 15th Ave NE, Seattle, WA 98125
	NE 125th St & 15th Ave NE, Seattle, WA 98125
14	S Jackson St & 23rd Ave S, Seattle, WA 98144
	S Jackson St & 25th Ave S, Seattle, WA 98144
15	University Way NE & NE 45th St, Seattle, WA 98105
	University Way NE & NE 47th Street, Seattle, WA 98105
16	1959 NE Pacific St, Seattle, WA 98195
	NE Pacific Pl & NE Pacific St, Seattle, WA 98195
17	Harvard Ave & Seneca St, Seattle, WA 98122
	Seneca St & 3rd Ave, Seattle, WA 98101
18	4th Ave N & Nickerson St, Seattle, WA 98109
	W Nickerson St & 3rd Ave W, Seattle, WA 98119
19	E Denny Way & Melrose Ave, Seattle, WA 98102
	Denny Way & Stewart St, Seattle, WA 98109

Appendix B: Details of Selected Corridors

Corridor 1: 2nd Avenue, between Pike Street and James Street

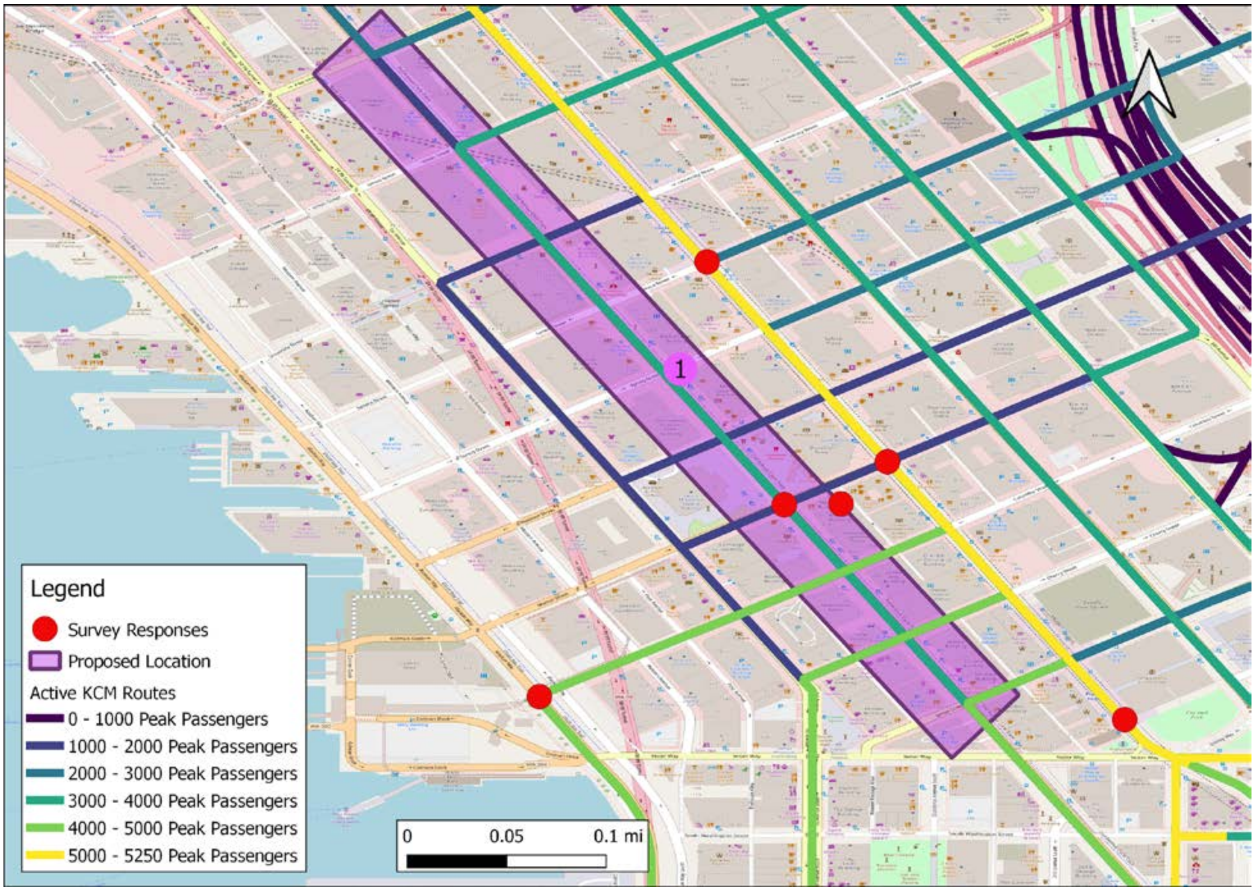
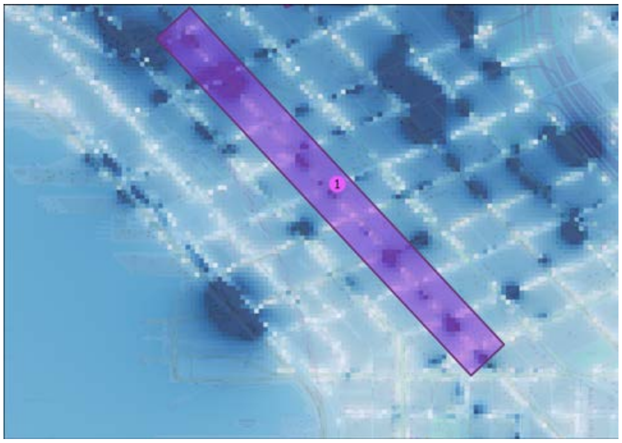


Figure 41: **Corridor 1 ridership and nearby survey responses.**



This downtown corridor is 9 blocks long and serves the largest number of transit routes of all proposed corridors. A few survey responses reported interaction near or within the corridor, and peak hour ridership for all routes is around 22,849 people/day. PUDO activity is extremely high throughout the area. A Google street view of the corridor is available here.

Figure 42: **Heat map of PUDO activity. Darker areas indicate higher peak hour activity.**

Corridor 2: Pike Street, between 3rd Avenue and 9th Avenue

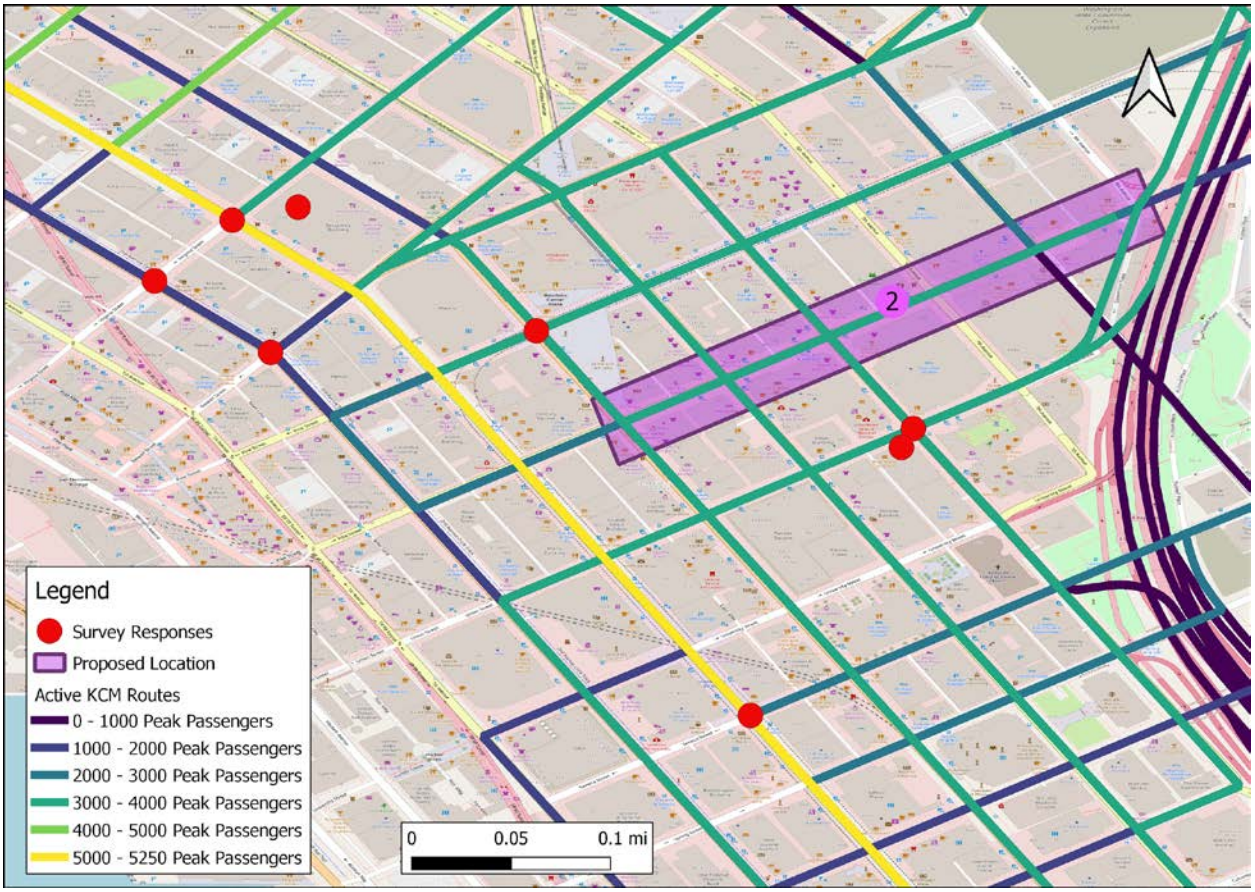
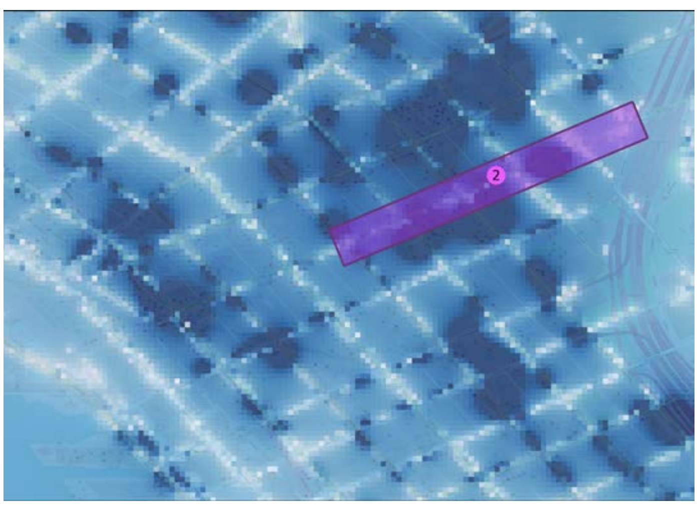


Figure 43: **Corridor 2 ridership and nearby survey responses.**



This downtown corridor is 6 blocks long and serves routes 10, 11, 47, and 9 among others. Peak hour ridership for all routes combined is around 22,157 people/day. This corridor has many commercial attractions, and captures potential transit interactions near the convention center. Union street is a similar alternative that serves more routes. PUDO activity is extremely high throughout the area. A Google street view of the corridor is available here.

Figure 44: **Heat map of PUDO activity. Darker areas indicate higher peak hour activity.**

Corridor 3: Westlake Avenue, between Denny Way and Mercer Street

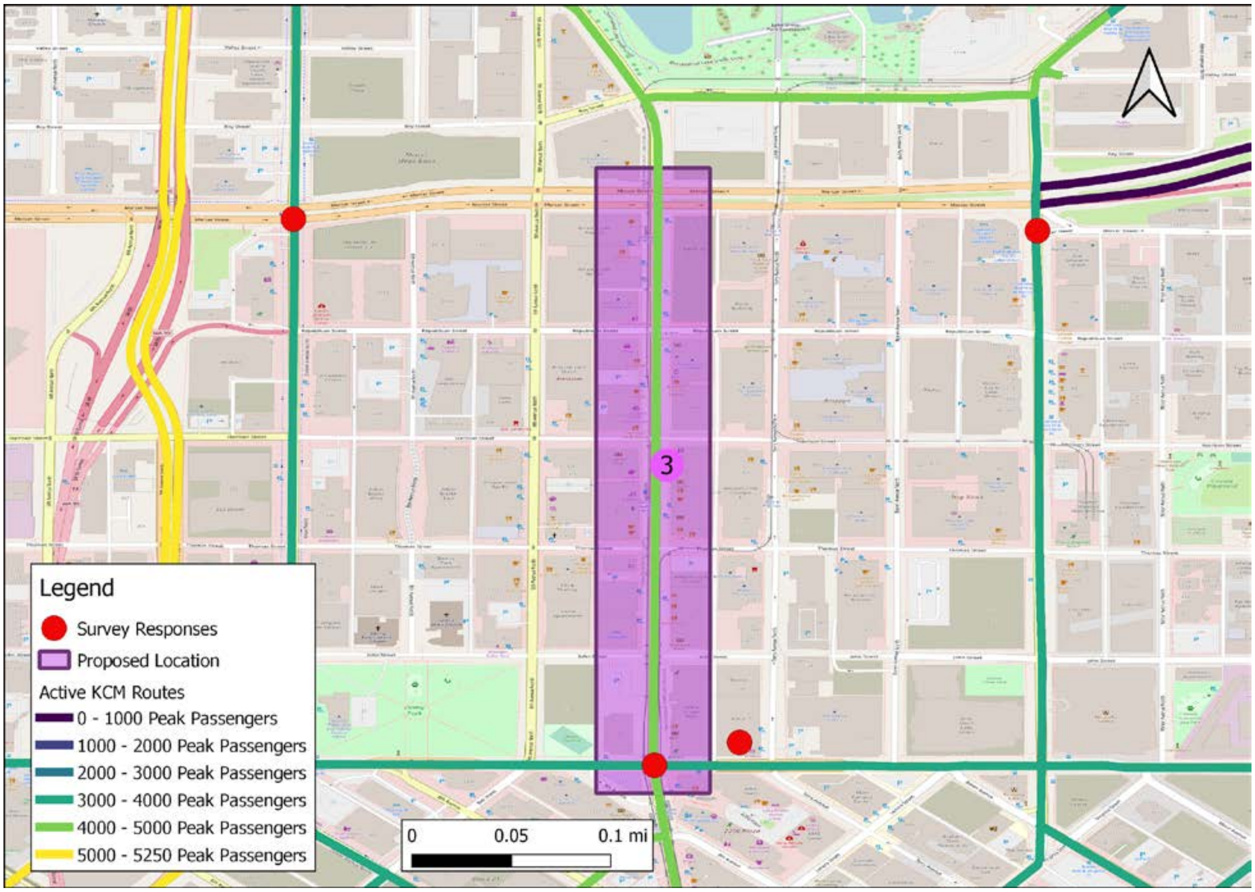
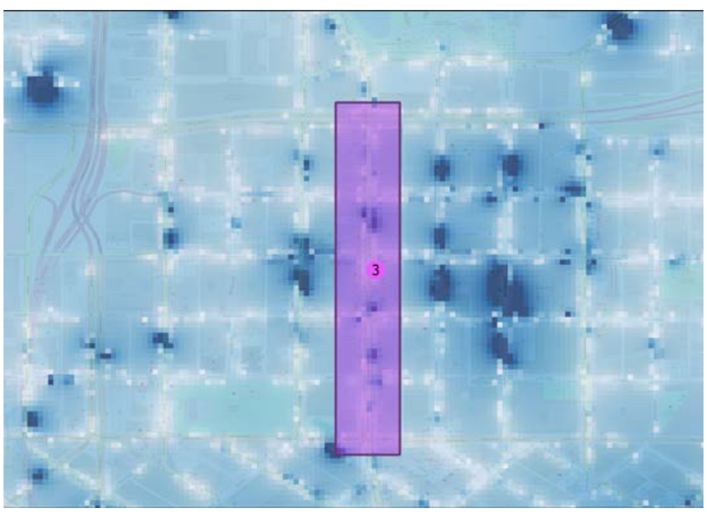


Figure 45: **Corridor 3 ridership and nearby survey responses.**



This South Lake Union corridor is 5 blocks long and serves route 40. One survey response reported interaction near the corridor, and peak hour ridership for all routes combined is around 8,893 people/day. This corridor has potential for interactions with employer-shuttles, as well as the Seattle Streetcar. PUDO data reveals higher activity to the east near Amazon, but this corridor captures the closest transit routes. A Google street view of the corridor is available here.

Figure 46: **Heat map of PUDO activity. Darker areas indicate higher peak hour activity.**

Corridor 4: E Olive Way, between E Denny Way and E Broadway

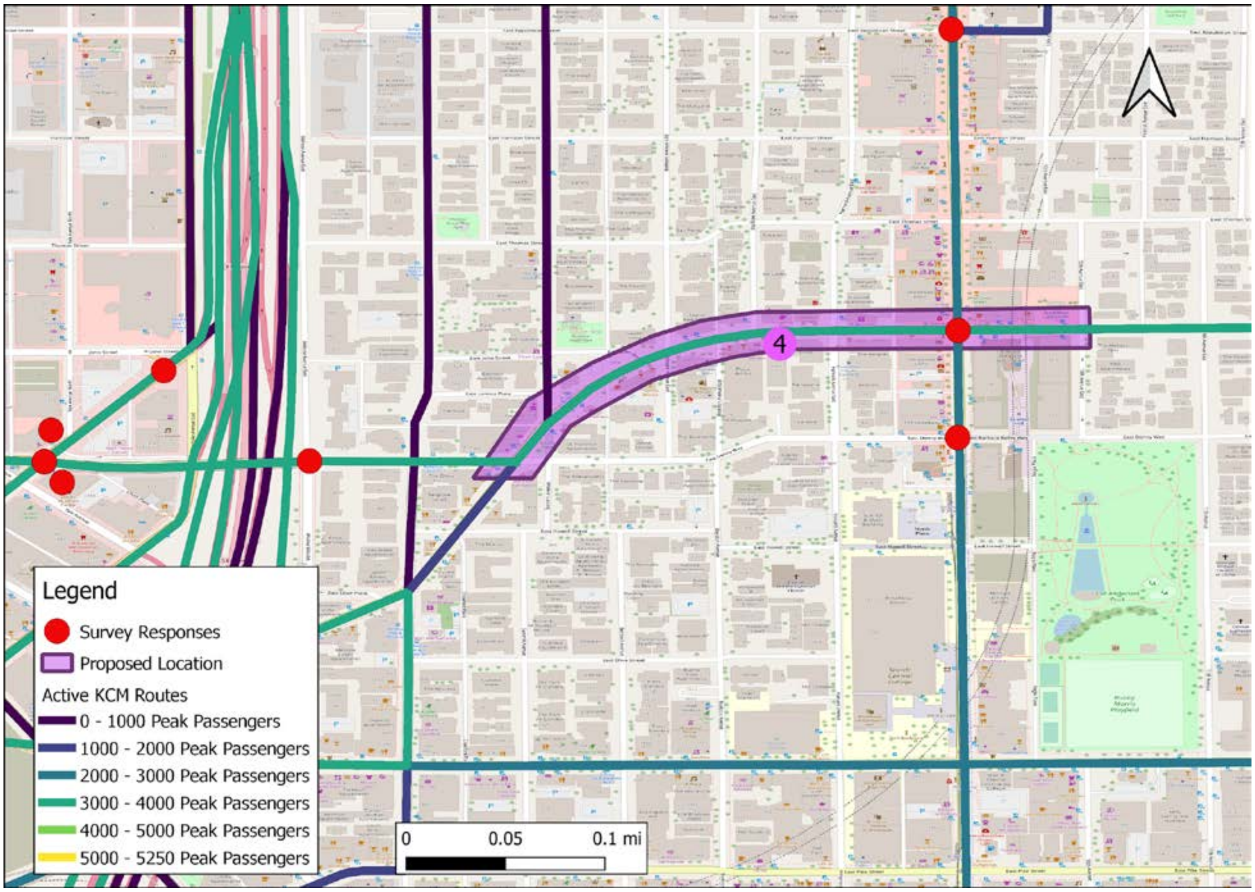
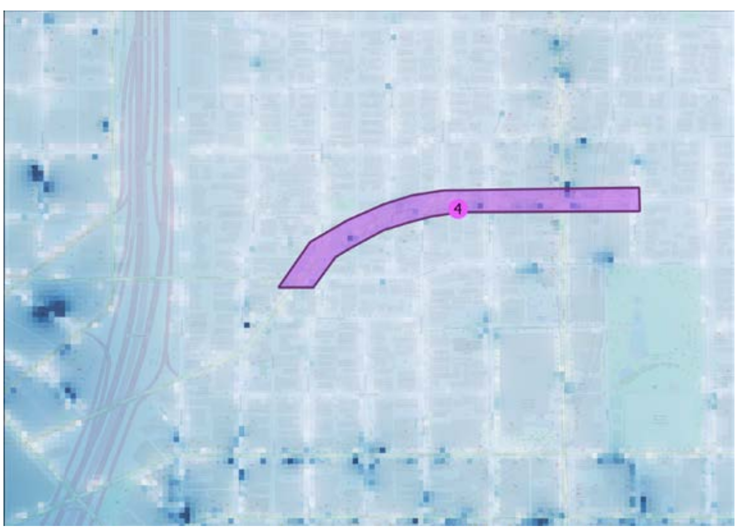


Figure 47: **Corridor 4 ridership and nearby survey responses.**



This Capitol Hill corridor is 6 blocks long and serves routes 10, 43, and 8. One survey response reported interaction in the corridor, and peak hour ridership for all routes combined is around 5,163 people/day. This corridor has potential for interaction with nearby commercial activity, as well as traffic accessing the light rail station on Broadway. A Google street view of the corridor is available here.

Figure 48: **Heat map of PUDO activity. Darker areas indicate higher peak hour activity.**

Corridor 5: 9th Avenue, between Alder Street and Columbia Street

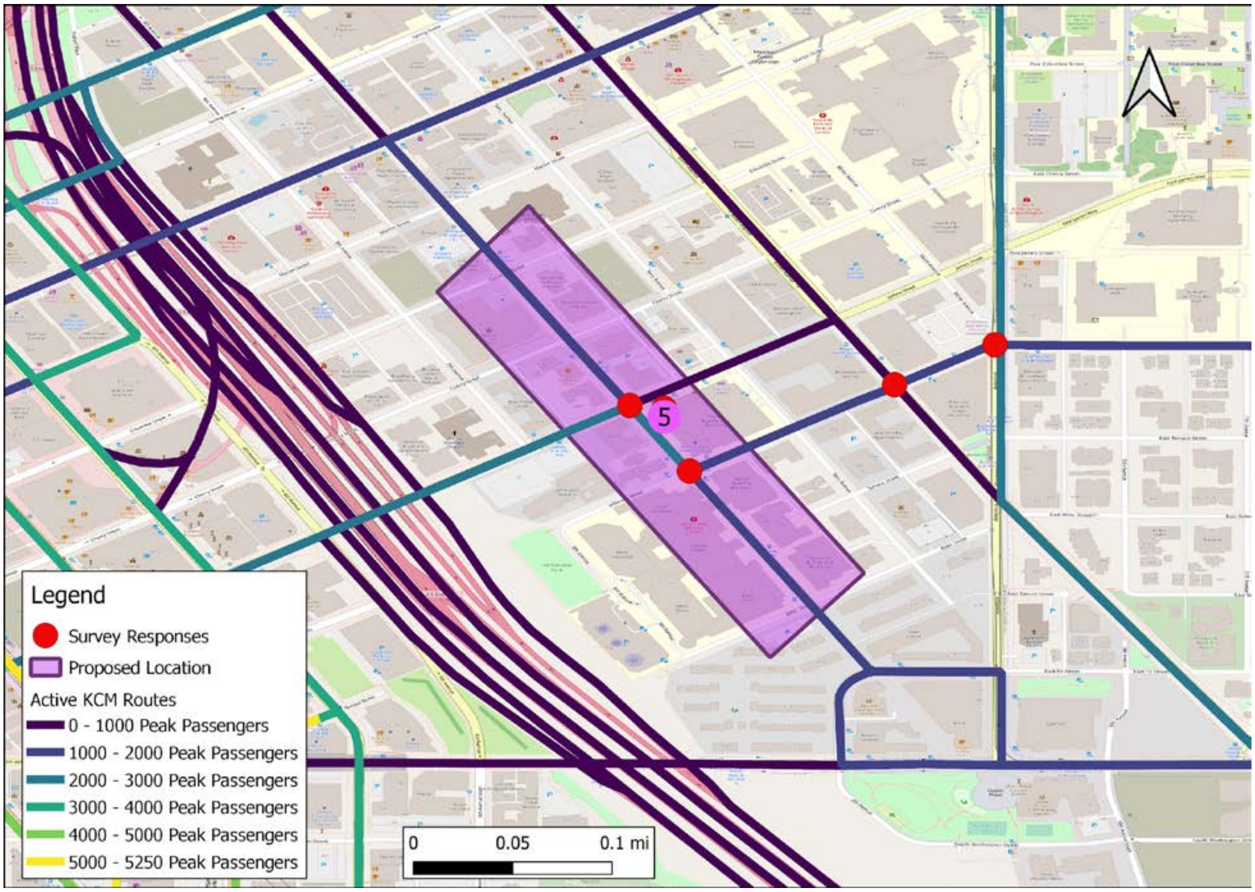
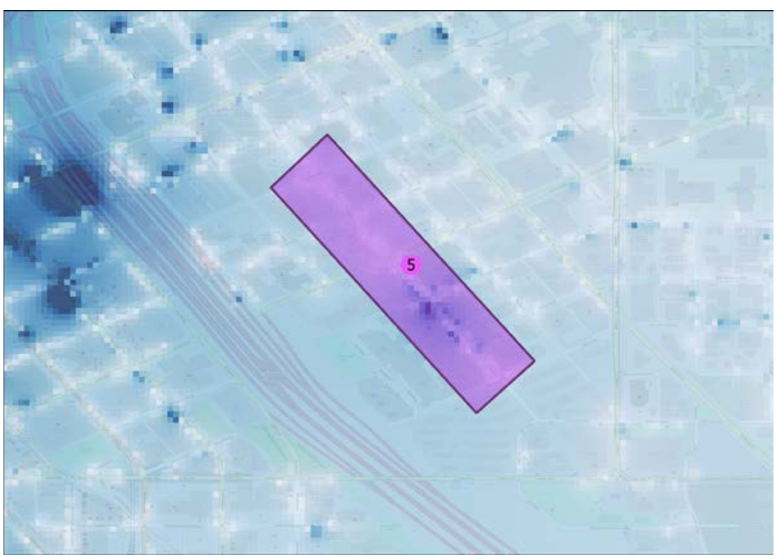


Figure 49: **Corridor 5 ridership and nearby survey responses.**



This First Hill corridor is 4 blocks long and serves routes 303, 13, 2, 4, 3, 60, and 193. Three survey responses reported interaction in the corridor, with another two responses nearby. The peak hour ridership for all routes combined is around 9,504 people/day. This corridor likely has interactions with the nearby medical center. A Google street view of the corridor is available here.

Figure 50: **Heat map of PUDO activity. Darker areas indicate higher peak hour activity.**

Corridor 6: University Way, between Campus Parkway and NE 45th Street

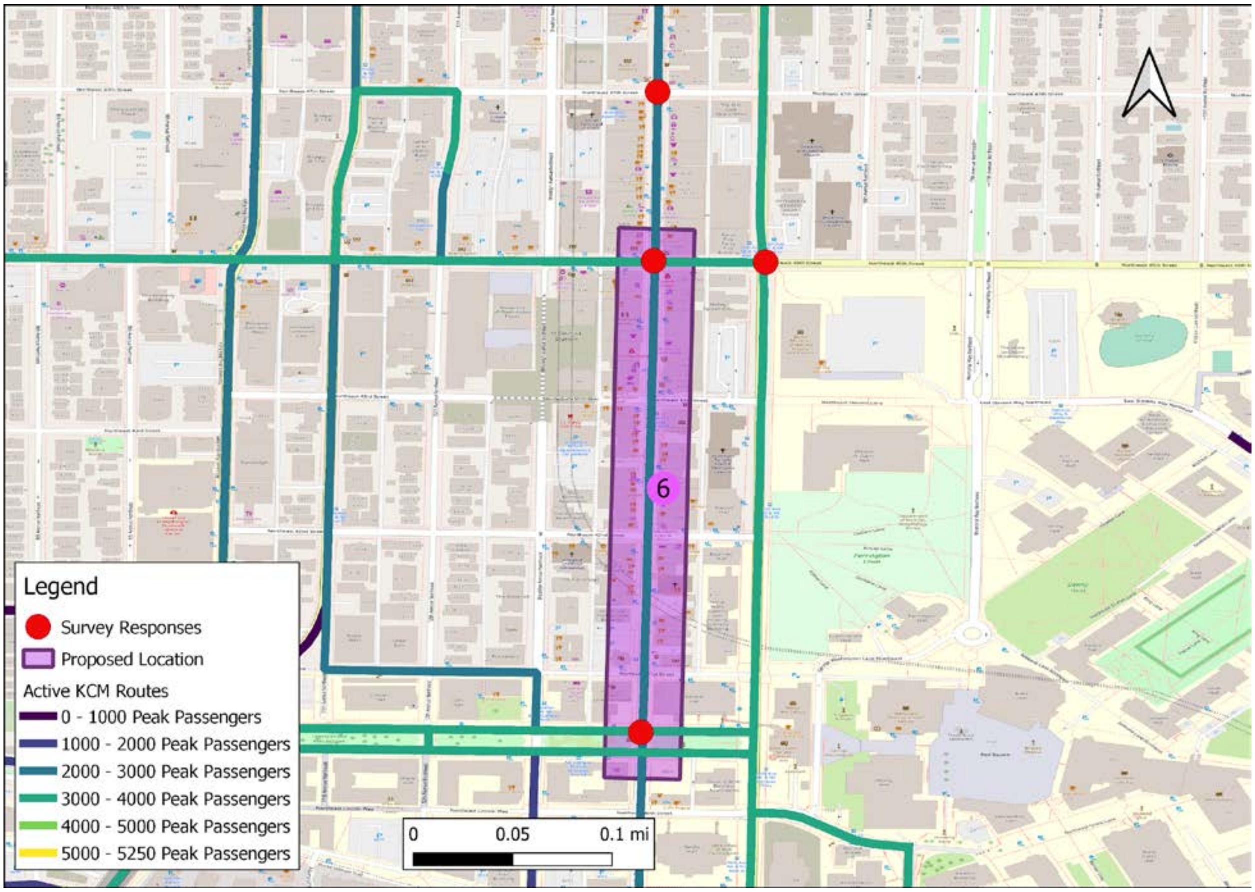
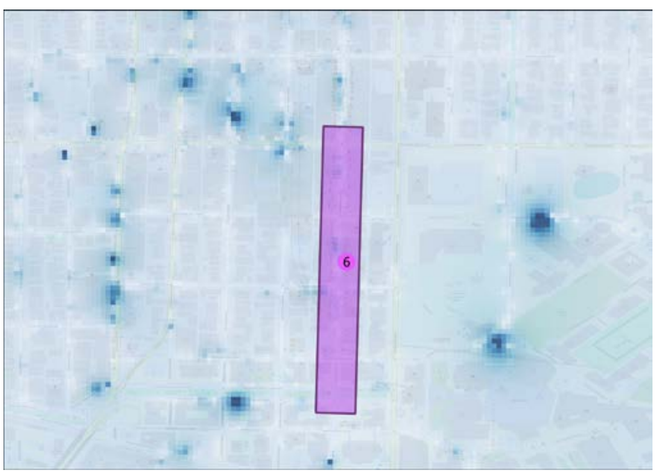


Figure 51: **Corridor 6 ridership and nearby survey responses.**



This University District corridor is 5 blocks long and serves routes 71, 73, 373, and 45. Two survey responses reported interaction in the corridor, and peak hour ridership for all routes combined is around 3,847 people/day. This corridor has a large amount of commercial interaction and potential interaction with the UW campus/dorms. A Google street view of the corridor is available here.

Figure 52: **Heat map of PUDO activity. Darker areas indicate higher peak hour activity.**

Corridor 7: NE Pacific Street, between 15th Avenue NE and Montlake Boulevard NE

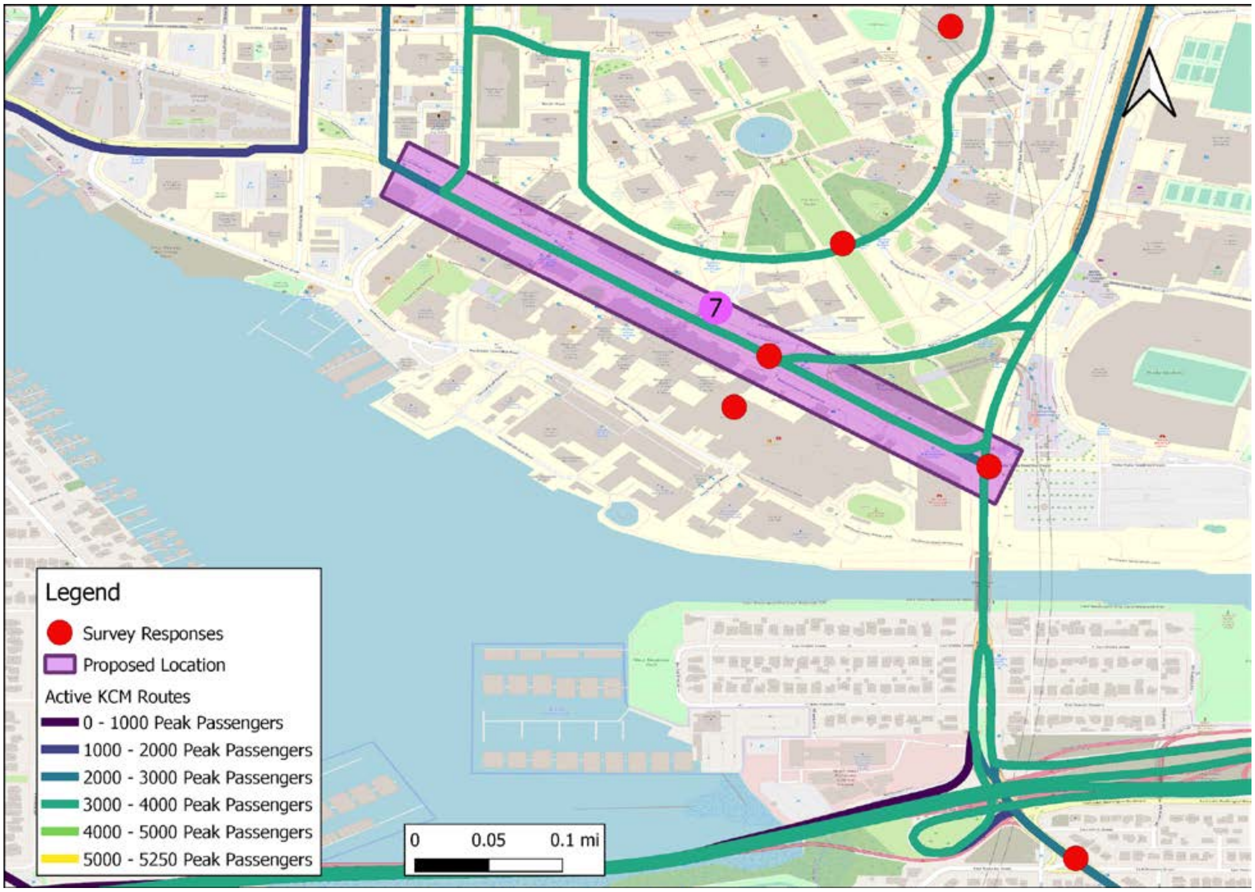


Figure 53: **Corridor 7 ridership and nearby survey responses.**

This University District corridor is roughly 5 blocks long and serves routes 41, 48, 71, 78, 542 and many others. Two survey responses reported interaction in the corridor, and peak hour ridership for all routes combined is around 16,816 people/day. This corridor has some potential for interference with traffic accessing the nearby UW facilities including the Medical Center. A Google street view of the corridor is available here.

Figure 54: **Heat map of PUDO activity. Darker areas indicate higher peak hour activity.**

Corridor 8: NW Market Street/Leary Ave, between 24th Avenue NW and 15th Avenue NW

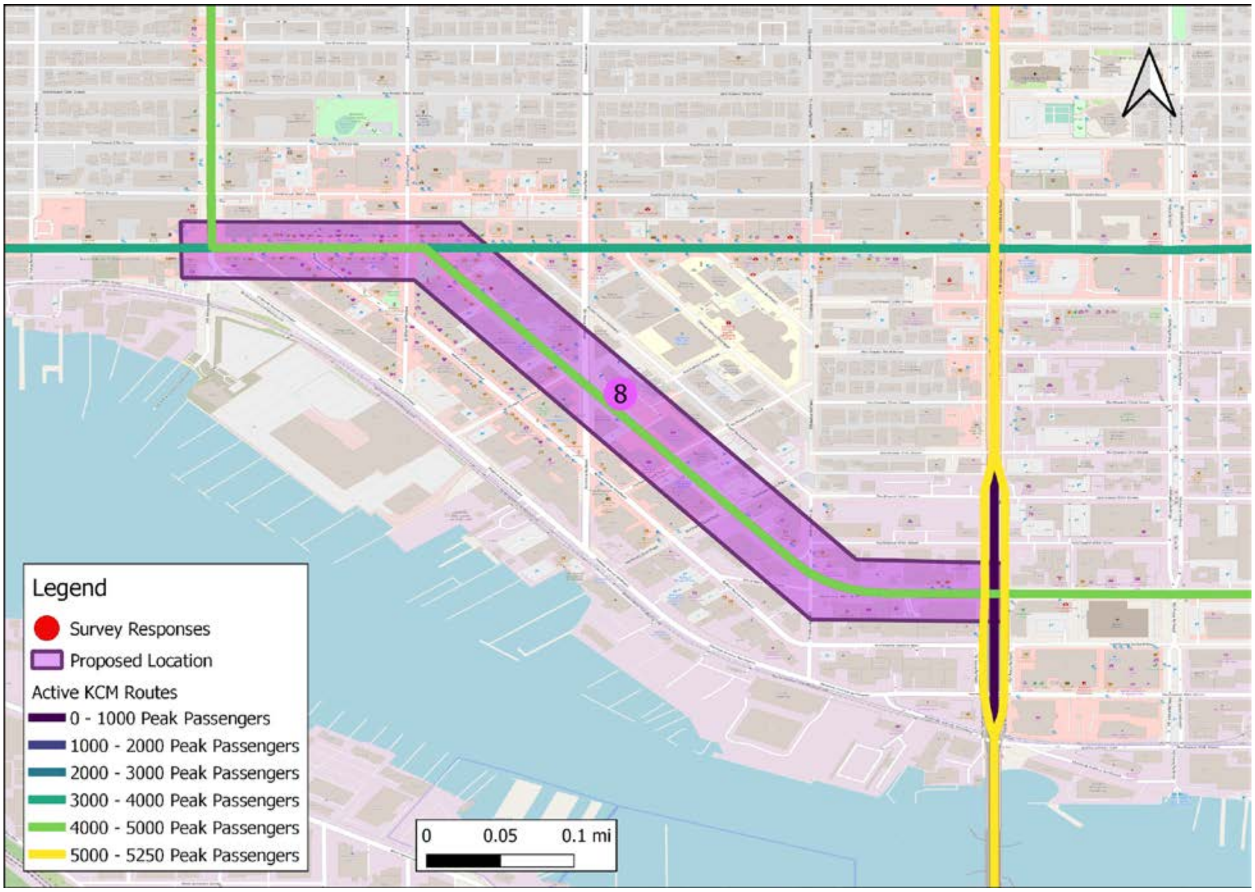


Figure 55: **Corridor 8 ridership and nearby survey responses.**

This Ballard corridor is 4 blocks long and serves route 40, 29, 18, and 17. There were no survey responses reporting interaction in the corridor, and peak hour ridership for all routes combined is around 6,068 people/day. This corridor serves as a convergence point to the Ballard bridge, and has many commercial attractions as well as for potential interaction with the nearby medical center. A Google street view of the corridor is available here.

Figure 56: **Heat map of PUDO activity. Darker areas indicate higher peak hour activity.**

Corridor 9: Rainier Avenue S, between S Bayview Street and MLK Way S

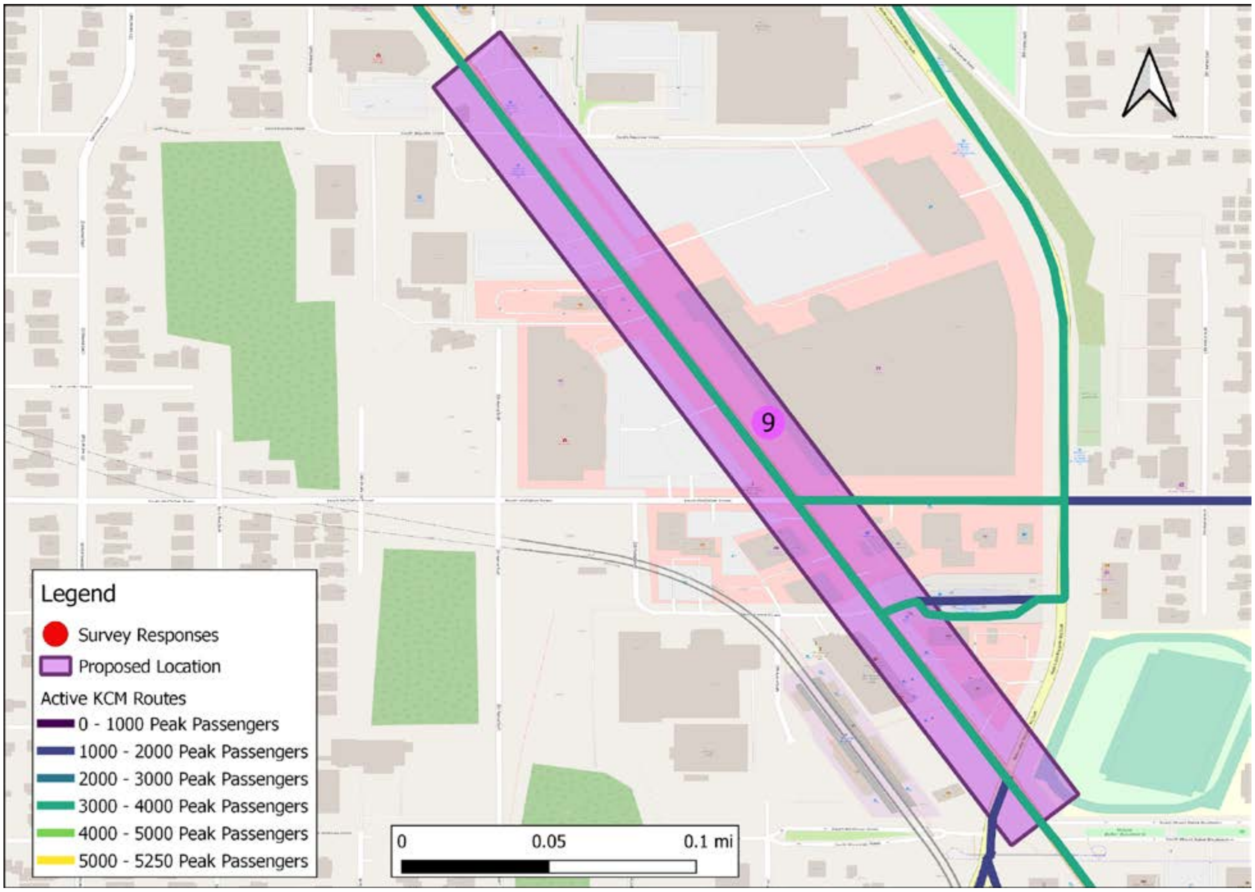


Figure 57: **Corridor 9 ridership and nearby survey responses.**

This Beacon Hill corridor is 4 blocks long and serves routes 106, 14, 7, 48, 8, 987, and 9. There were no survey responses reporting interaction in the corridor, and peak hour ridership for all routes combined is around 12,566 people/day. This corridor has potential for interaction with the Mt. Baker Transit Center, and there are a number of converging routes here. A Google street view of the corridor is available here.

Figure 58: **Heat map of PUDO activity. Darker areas indicate higher peak hour activity.**

Corridor 10: 108th Avenue NE, between NE 4th Street and NE 12th Street

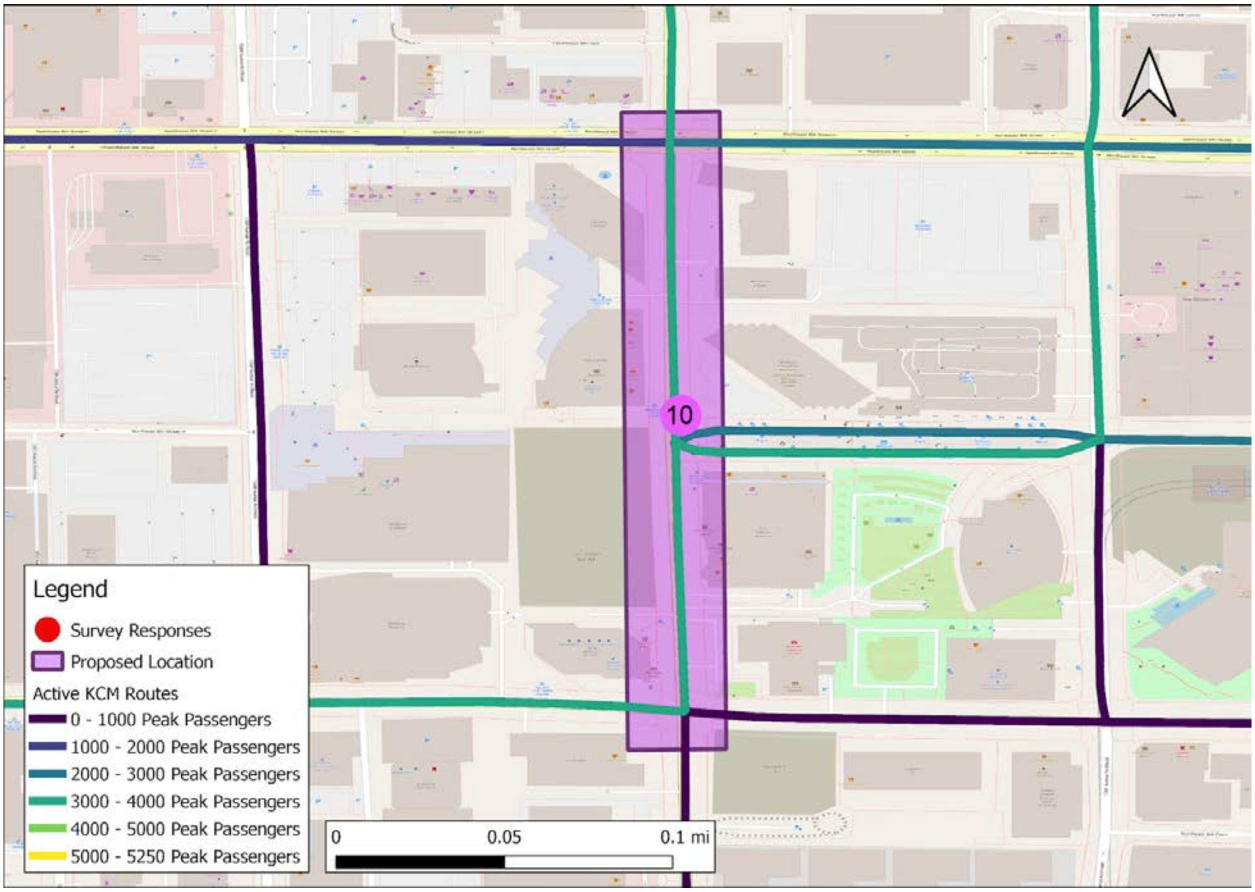
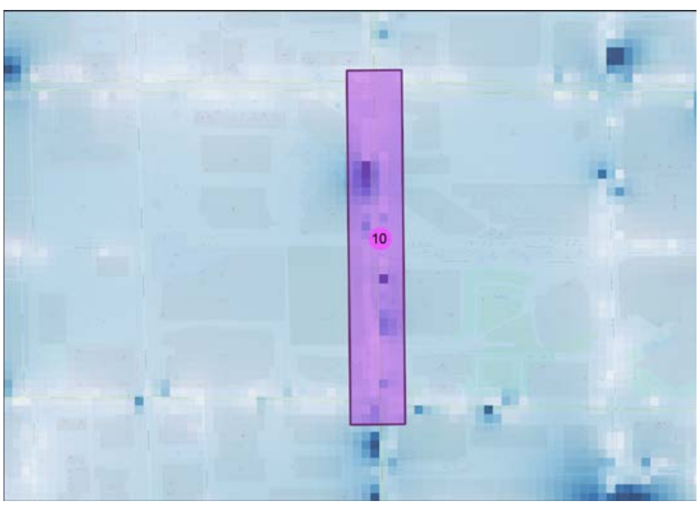


Figure 59: **Corridor 10 ridership and nearby survey responses.**



This downtown Bellevue corridor is 4 blocks long and serves routes 75, 77, 172, and 270 among many others. There were no survey responses reporting interaction in the corridor. Peak hour ridership for all routes combined is around 10,159 people/day. This corridor has potential for interactions with the nearby Amazon campus, and Bellevue Transit Center. A Google street view of the corridor is available here.

Figure 60: Heat map of PUDO activity. Darker areas indicate higher peak hour activity.

Appendix C: Field Data Collection Forms

Bus arrival form

The “Bus Arrival” form starts with questions on: 1) the bus arrival time, 2) door opening time, 3) passenger flow stop time, 4) door closing time, and 5) bus departure time. All of these fields collect a timestamp (date and time). When the data collector clicks on a field, a window pops up with the current timestamp, as shown in *Figure 61*. This automatic timestamp selection allows faster responses compared to the data collector having to type the date and time every time.

The next question is whether the access ramp was deployed. If a ramp is not deployed, this question can be skipped; if the data collector presses “Yes”, an additional field is added to the form, asking for how many people used the ramp.

The next question asks whether the bike rack in front of the bus was used. If the data collector responds yes, an additional question is shown asking if the use of the bike rack caused any delay for the bus. What is meant by bike rack delay was explained in the data collection protocol and is as follows. A delay caused by a bike rack user is defined as a delay that occurs when all other passengers have boarded/alighted the bus before the user gets their bike off or puts their bike on. In other words, if all passengers had not yet boarded/alighted the bus by the time the bike rack user gets their bike on/off the rack, it means that the bus had to wait anyway for other passengers, and so no delay was caused by the bike rack user. If the user reports that a delay was caused, another question will be added, which asks roughly how many seconds of delay was caused by the bike rack user.

The next question asks whether the bus stopped at the bus stop, and aims to capture instances when the bus does not stop at the bus stop (e.g. because there are no passengers to board or alight at that stop). There is only one answer choice for this question which is “Bus did not stop (mark arrival time only)”, and the question can be skipped if the bus stops to pick up/drop off passengers. If the bus did not stop, the data collectors were only asked to put in the arrival timestamp of the bus, marking the moment when the bus passed the bus stop, and leave all other fields (e.g., door open time) empty. This question was included to collect data for all buses passing through a corridor, even if they did not stop, so that they can later be used to verify GTFS-RT timestamps.

The next question asks for multiple buses arriving at the same time. This is an additional measure put in place in case, at a busy station, two or more buses arrive together and the data collector is not able to submit all timestamps for multiple buses simultaneously. In such a case, they would select “Multiple buses arrived at the same time”, and leave additional information in the notes sections, such as the other bus vehicle IDs.

Data collectors were also asked to add notes to provide more context on what was going on at their duty post; for example, a delayed movement because the bus wanted to merge into traffic, or a traffic collision nearby.

Interference form

For the interference form, the questions asked are interference start and stop time, location of interference, location category, the interfering mode.

The interference start time is the timestamp at which the interfering mode starts affecting the bus operation, by causing the bus to move slower or come to a full stop. The interference stop time is the timestamp at which the bus starts moving away from the location of interference by either starting to move after a complete stop or starting to accelerate.

To record the location of interference, the user can simply click a button and the latitude, longitude, and altitude fields get automatically filled through the GPS of the smartphone (see *Figure 62*). The bus interference may be happening up to 200 feet away from where the data collector stands; however, we decided that the rough location of the interference would suffice.

The next field is the location category of where the interference is occurring. The options include the following location categories:

1. **At a bus stop**: Anywhere alongside the curb within close proximity of a bus stop (or wherever the bus turnout and curb transitioning starts/stops).
2. **In a bus only lane**: Anywhere in a lane which is dedicated to buses either shown by road signs, pavement markings, or separation from the rest of the road.
3. **In a driving lane**: Any other travel lane that is not a dedicated bus lane or in the proximity of a bus stop.
4. **Other**: If the location does not fall within the above categories. In this case, data collectors are asked to enter their response in the textbox space which will show after they select this field.

Next, data collectors are asked to specify what mode(s) interfered with the bus. This question is in the *select all that apply* format which allows selecting multiple modes. The available options are passenger vehicle, ridehailing vehicle, bicyclist, scooter rider, pedestrian, service vehicle, small delivery vehicle, large delivery vehicle, transit, and construction. Each option was further defined in the data collection protocol as follows:

- **Passenger Vehicle**: A motor vehicle with the body type sedan, hatchback, minivan, SUV, or other similar class, with at least four wheels and comprising no more than eight seats (including the driver's seat), with the main purpose of transporting passengers.
- **Ridehailing/TNC vehicle**: A passenger-car-sized vehicle in which a customer hires a driver to take them to their destination. This vehicle can be a taxi or a vehicle hailed virtually through an app offered by a Transportation Network Company (TNC, e.g. Uber and Lyft).
  For data collection, a ridehailing vehicle was defined as a passenger-car-sized vehicle which is either colored/painted with taxi patterns (e.g. yellow/green color with taxi top sign) or has a detectable Uber or Lyft sign on the front window, and is picking up or dropping off one or more passengers or waiting to do so.
  If a TNC sign is not detectable on the car, but a passenger is seen stepping out/in through the backdoor of the vehicle, it can be identified as a ridehailing vehicle. For other cases, data collectors were asked to apply their own judgment of the situation to either classify the vehicle as a passenger vehicle or ridehailing vehicle.
- **Bicyclist**: Any person riding a privately owned bicycle or a bike-share service (e.g. JUMP).
- **Scooter Rider**: Any person riding a privately owned scooter or a scooter-share service (e.g. Lime, Wheels, or Link).

- **Pedestrian**: Any person walking on the road in a manner which is obstructing the bus by either causing delay in movement, blocking access to right-of-way, or otherwise causing an interference.
- **Service vehicle**: A motor vehicle that: 1) bears commercial plates; 2) is primarily used for a commercial/other service; and 3) displays the registrant's name and address permanently affixed in characters on both sides of the vehicle; which is used primarily for either the transportation of property or for the provision of commercial services (such as plumbing, electrician, waste management, etc.).
- **Small Delivery Vehicle**: A passenger-car-sized vehicle vehicle (not a truck or a van) which is making/receiving deliveries to and from businesses. Examples of these vehicles are those used by Instacart, UberEats, DoorDash, etc.
- **Large Delivery Vehicle**: Vans or trucks which are used for the transportation of property, goods, mail, parcels, and packages. Examples of this include UPS, USPS, FedEx, Amazon Prime vans, etc. Other examples could also be unmarked local delivery trucks delivering goods to a supermarket (observers were asked to describe the vehicle in the next field).
- **Transit**: In a situation where another bus has stopped at the bus stop (either boarding/alighting passengers or waiting to merge to traffic), which does not allow the arriving bus to stop at the curb and open/close doors to pick up and drop off passengers, the stopped bus is causing delay to the arriving bus. This counts as interference from another transit mode. Cases where a bus is fully occupying the bus stop, yet the arriving bus is able to pull over to the curb and open/close its doors without any delay do not count as interference.
- **Construction**: In corridors where there is construction going on, if buses are forced to stop by flaggers or other workers with a stop sign, this counts as interference.

Data collectors are asked to provide a brief description of the interference describing how the other mode(s) interfered with the bus movement or further description of the interfering mode. They are also allowed to upload a photo if they felt like it would help with the description. Once selected, the app would prompt their phone's camera to open and allow them to take and upload a photo of the situation.

Shifts form

The data collectors would also fill out a "shifts form" (shown in *Figure 63*) before their shifts start, stating their first and last name, the point_id (an ID associated with their post location, which was given to them in advance) and stop_id (an ID associated with the stop, if any, they are assigned to, which was given to them in advance), date and time of when they filled the form, and whether their shift was for the morning or afternoon period.

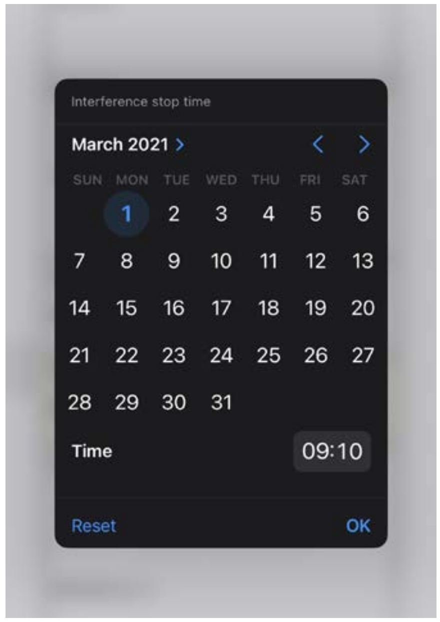
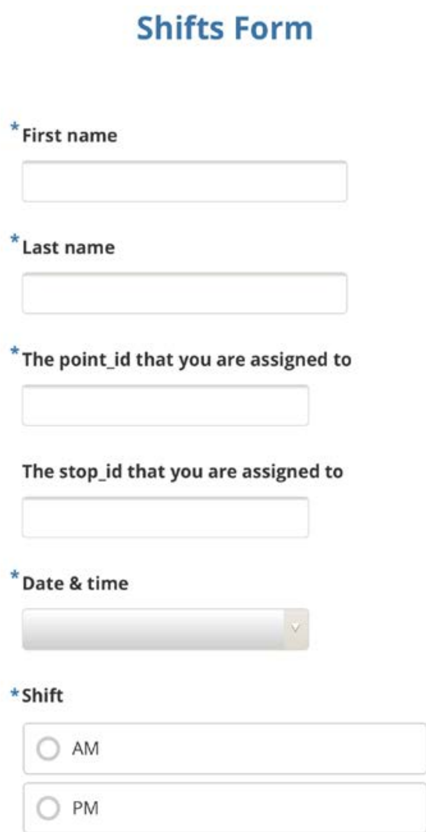


Figure 61: **This window pops up when the data collector clicks on a timestamp field. The exact date and time of that moment are displayed. The data collector simply has to press the "OK" button in the bottom right. The date is shown in a calendar format, and the time is shown in hours and minutes. The "seconds" are also recorded and uploaded in the backend even though they are not displayed in the user interface.**

Figure 62: **Location example: once you press the "Find my location" button (shaped like a target) it automatically loads the latitude, longitude, altitude, and location accuracy.**



Shifts Form

*First name

*Last name

*The point_id that you are assigned to

The stop_id that you are assigned to

*Date & time

*Shift

AM

PM

***Figure 63*: Shifts form asking for first and last name, the point_id and stop_id that the RA was assigned to, date and time of when they filled the form, and whether their shift was for the morning rush hour or the evening rush hour shift.**

Appendix D: Observer Posts in the Field

Table 7 has information regarding the field observers' post locations. The information in the table includes the corridor ID and name, a point ID (that we assigned to each post), whether it is a bus stop, the stop name, latitude and longitude, and boundaries that the observer should monitor.

Table 7: Observation posts

Corridor ID	Corridor Name	Point ID	Bus Stop	Stop ID	Stop Name	Latitude	Longitude	Boundaries
1	2nd Ave; Pike to James	100	✓	300	2nd Ave & Pike St	47.60865	-122.33840	Pike St to Union St
1	2nd Ave; Pike to James	101	✗	N/A	N/A	47.60732	-122.33732	Union St to University St
1	2nd Ave; Pike to James	102	✓	320	2nd Ave & Seneca St	47.60621	-122.33620	University St to Seneca St
1	2nd Ave; Pike to James	103	✗	N/A	N/A	47.60525	-122.33538	Spring St to Madison St
1	2nd Ave; Pike to James	104	✓	340	2nd Ave & Marion St	47.60447	-122.33460	Madison St to Columbia St
1	2nd Ave; Pike to James	105	✓	360	2nd Ave & Cherry St	47.60302	-122.33330	Columbia St to Cherry St
1	2nd Ave; Pike to James	106	✓	361	2nd Ave & James St	47.60248	-122.33280	Cherry St to James St
2	Pike; 3rd to 9th	200	✓	1180	Pike St & 4th Ave	47.61007	-122.33680	3rd Ave to 4th Ave
2	Pike; 3rd to 9th	201	✗	N/A	N/A	47.61054	-122.33579	4th Ave to 5th Ave
2	Pike; 3rd to 9th	202	✓	1190	Pike St & 6th Ave	47.61103	-122.33450	5th Ave to 6th Ave
2	Pike; 3rd to 9th	203	✓	1195	Pike St & 7th Ave	47.61181	-122.33270	6th Ave to 8th Ave
2	Pike; 3rd to 9th	204	✓	11130	Pike St & Convention Pl	47.61237	-122.33140	8th Ave to 9th Ave
3	Westlake; Denny to Mercer	300	✓	2255	Denny Way & Westlake Ave	47.61847	-122.33820	Denny Way to John St
3	Westlake; Denny to Mercer	301	✗	N/A	N/A	47.62030	-122.33830	John St to Thomas St
3	Westlake; Denny to Mercer	302	✓	26645	Westlake Ave N & Harrison St	47.62138	-122.33860	Harrison St to Thomas St (SB)
3	Westlake; Denny to Mercer	303	✓	26715	Westlake Ave N & Harrison St	47.62185	-122.33840	Thomas St to Harrison St (NB)
3	Westlake; Denny to Mercer	304	✗	N/A	N/A	47.62271	-122.33860	Harrison St to Republican St
3	Westlake; Denny to Mercer	305	✓	26641	Westlake Ave N & Mercer St	47.62415	-122.33850	Republican St to Mercer St
3	Westlake; Denny to Mercer	306	✓	26730	Westlake Ave N & Mercer St	47.62550	-122.33840	Mercer St to Valley St
4	E Olive/E John; Denny to 10th	400	✓	29266	E Olive Way & Summit Ave E	47.61915	-122.32510	Belmont Ave E to Summit Ave E
4	E Olive/E John; Denny to 10th	401	✓	29268	E Olive Way & Summit Ave E	47.61913	-122.32490	E Denny Way to Summit Ave E
4	E Olive/E John; Denny to 10th	402	✗	N/A	N/A	47.61996	-122.32298	Belmont Ave E to Harvard Ave E
4	E Olive/E John; Denny to 10th	403	✗	N/A	N/A	47.61997	-122.32213	Harvard Ave E to Broadway E
4	E Olive/E John; Denny to 10th	404	✓	29270	E John St & Broadway E	47.61985	-122.32050	Broadway E to 10th Ave E
4	E Olive/E John; Denny to 10th	405	✓	29262	E John St & 10th Ave E	47.61998	-122.32000	10th Ave E to Broadway E
5	9th Ave; Alder to Columbia	500	✓	41930	9th Ave & Alder St	47.60404	-122.32320	Alder St to Jefferson St (NB)
5	9th Ave; Alder to Columbia	501	✓	42010	9th Ave & Alder St	47.60347	-122.32300	Jefferson St to Alder St (SB)
5	9th Ave; Alder to Columbia	502	✓	12910	9th Ave & Jefferson St	47.60487	-122.32420	Jefferson St to James St (NB)
5	9th Ave; Alder to Columbia	503	✓	12880	Jefferson St & 9th Ave	47.60493	-122.32410	Terry Ave to 9th Ave
5	9th Ave; Alder to Columbia	504	✓	41940	9th Ave & James St	47.60602	-122.32510	James St to Cherry St
5	9th Ave; Alder to Columbia	505	✓	42000	9th Ave & Cherry St	47.60609	-122.32540	Cherry St to Columbia St

Corridor ID	Corridor Name	Point ID	Bus Stop	Stop ID	Stop Name	Latitude	Longitude	Boundaries
6	University Way; Campus Pkwy to 45th	600	✓	9142	University Way NE & NE 41st St	47.65695	-122.31330	Adjacent stop to NE Campus Pkwy (SB)
6	University Way; Campus Pkwy to 45th	601	✓	9581	University Way NE & NE 41st St	47.65739	-122.31310	NE 41st St to NE 42nd St (NB)
6	University Way; Campus Pkwy to 45th	602	✗	N/A	N/A	47.65887	-122.31312	NE 42nd St to Chipotle
6	University Way; Campus Pkwy to 45th	603	✓	9582	University Way NE & NE 43rd St	47.65965	-122.31311	Chipotle to NE 43rd St
6	University Way; Campus Pkwy to 45th	604	✓	9134	University Way NE & NE 43rd St	47.66015	-122.31325	NE 45th St to NE 43rd St
7	Pacific; 15th to Montlake	700	✓	29240	NE Pacific St & 15th Ave NE	47.65208	-122.31090	15th Ave NE to Stop
7	Pacific; 15th to Montlake	701	✓	29420	NE Pacific St & 15th Ave NE	47.65235	-122.31110	Magnuson Health Sciences to Stop
7	Pacific; 15th to Montlake	702	✓	29247	NE Pacific St & Montlake Blvd NE - Bay 1	47.64914	-122.30500	Montlake Blvd NE to NE Pacific Pl (WB)
7	Pacific; 15th to Montlake	703	✓	29405	NE Pacific St & Montlake Blvd NE - Bay 2	47.64973	-122.30580	NE Pacific Pl to Montlake Blvd NE (EB)
8	Northwest Market/Leary; 24th to 15th	800	✓	18120	NW Market St & Ballard Ave NW	47.66861	-122.38550	24th Ave NW to Leary Ave NW (WB)
8	Northwest Market/Leary; 24th to 15th	801	✓	18740	NW Market St & Ballard Ave NW	47.66875	-122.38610	Leary Ave NW to 24th Ave NW (EB)
8	Northwest Market/Leary; 24th to 15th	802	✗	N/A	N/A	47.66791	-122.38377	Diamond Parking to 24th Ave NW
8	Northwest Market/Leary; 24th to 15th	803	✓	18145	Leary Ave NW & NW Vernon Pl	47.66732	-122.38300	Diamond Parking to 20th Ave NW
8	Northwest Market/Leary; 24th to 15th	804	✓	18720	Leary Ave NW & NW Vernon Pl	47.66721	-122.38260	Stop to 20th Ave NW
8	Northwest Market/Leary; 24th to 15th	805	✓	18150	Leary Ave NW & NW Ione Pl	47.66536	-122.38070	20th Ave NW to NW Dock Pl (SB)
8	Northwest Market/Leary; 24th to 15th	806	✓	18706	Leary Ave NW & NW Ione Pl	47.66553	-122.38060	NW Dock Pl to 20th Ave NW (NB)
8	Northwest Market/Leary; 24th to 15th	807	✗	N/A	N/A	47.66392	-122.37907	NW Dock Pl to NW 48th St
8	Northwest Market/Leary; 24th to 15th	808	✓	18696	NW Leary Way & 15th Ave NW	47.66373	-122.37710	15th Ave NW to NW 48th St
8	Northwest Market/Leary; 24th to 15th	809	✓	18165	NW Leary Way & 15th Ave NW	47.66359	-122.37560	15th Ave NW to Stop
9	Rainier; S Bayview to MLK	900	✓	8681	Rainier Ave S & S Forest St - Bay 4	47.57683	-122.29710	S McClellan St to MLK Way (SB)
9	Rainier; S Bayview to MLK	901	✓	8401	Rainier Ave S & Mount Baker Transit Center - Bay 1 - Bay 1	47.57751	-122.29740	MLK Way to S McClellan St (NB)
9	Rainier; S Bayview to MLK	902	✗	N/A	N/A	47.57921	-122.29899	Wendy's to S McClellan St
9	Rainier; S Bayview to MLK	903	✓	8660	Rainier Ave S & S Bayview St	47.58079	-122.30010	S Bayview St to Wendy's
9	Rainier; S Bayview to MLK	904	✓	8429	Rainier Ave S & S Bayview St	47.58120	-122.30020	Wendy's to S Bayview St
10	108th Ave; 4th to 12th (Bellevue)	1000	✗	N/A	N/A	47.62014	-122.19624	NE 12th St to NE 10th St
10	108th Ave; 4th to 12th (Bellevue)	1001	✗	N/A	N/A	47.61841	-122.19648	NE 10th St to NE 8th St
10	108th Ave; 4th to 12th (Bellevue)	1002	✗	N/A	N/A	47.61625	-122.19643	NE 8th St to Transit Center Rd
10	108th Ave; 4th to 12th (Bellevue)	1003	✗	N/A	N/A	47.61547	-122.19643	Transit Center Rd to HomeStreet Bank
10	108th Ave; 4th to 12th (Bellevue)	1004	✗	N/A	N/A	47.61466	-122.19613	HomeStreet Bank to NE 4th St

Appendix E: List of Observed Interference Instances

Table 8 illustrates a list of all interferences by mode and corridor, accompanied by time and description of interference.

Table 8: List of observed interference instance

No.	Interfering Mode	Corridor	Time and Date	Description
1	Bicylist	Pike; 3rd to 9th	3/12/2021 8:26	Bicylist was moving slowly in the bus lane
2	Construction	Pike; 3rd to 9th	3/5/2021 7:19	Construction vehicle blocked the turning lane of the bus.
3	Construction	Pike; 3rd to 9th	3/5/2021 7:36	Cement truck with hazards on partially blocked lane for bus to pass in
4	Construction	Pike; 3rd to 9th	3/5/2021 7:39	Truck partially blocked lane for bus to travel in.
5	Construction	108th Ave; 4th to 12th (Bellevue)	3/19/2021 8:00	Bus was moving slowly and cautiously due to construction. About 3sec
6	Construction	108th Ave; 4th to 12th (Bellevue)	3/19/2021 8:35	Bus was moving slowly and cautiously due to construction. Bus didn't stop
7	Construction	108th Ave; 4th to 12th (Bellevue)	3/19/2021 9:21	Bus was moving slowly and cautiously due to construction. Bus did not stop.
8	Passenger Vehicle	University Way; Campus Pkwy to 45th	3/8/2021 9:15	Car was parallel parking
9	Passenger Vehicle	University Way; Campus Pkwy to 45th	3/8/2021 9:15	Car was parking.
10	Passenger Vehicle	E Olive/E John; Denny to 10th	3/8/2021 16:10	Car had stopped for pedestrian pushing a stroller which, therefore, stopped the bus.
11	Pedestrian	University Way; Campus Pkwy to 45th	3/8/2021 8:04	Passenger walks across bus stop
12	Pedestrian	E Olive/E John; Denny to 10th	3/8/2021 15:42	Pedestrian was crossing the street and bus slowed significantly, but did not stop completely.
13	Pedestrian	E Olive/E John; Denny to 10th	3/8/2021 16:37	Bus driver yield to pedestrian intending to cross the street with a hand signal.
14	Pedestrian	E Olive/E John; Denny to 10th	3/8/2021 16:48	Bus stopped twice, once for a pedestrian crossing a cross walk and then again for a pedestrian jaywalking.
15	Pedestrian	E Olive/E John; Denny to 10th	3/8/2021 16:53	Bus stopped for pedestrian at crosswalk.
16	Pedestrian	E Olive/E John; Denny to 10th	3/10/2021 15:29	Crossing the street
17	Pedestrian	E Olive/E John; Denny to 10th	3/10/2021 15:32	Pedestrian crossed street but their child would not come with so bus was halted briefly.
18	Pedestrian	E Olive/E John; Denny to 10th	3/10/2021 16:14	Cross walk
19	Pedestrian	E Olive/E John; Denny to 10th	3/10/2021 17:11	Passenger dropped a bunch of their property they were carrying on the sidewalk when getting off the bus. Caused bus to stay there longer than needed.
20	Pedestrian	Pike; 3rd to 9th	3/12/2021 7:19	Pedestrian jaywalked ahead of bus, making it slow slightly.
21	Service Vehicle	University Way; Campus Pkwy to 45th	3/8/2021 7:35	Seattle city light truck parallel parked on side of road is too big so the bus slows down to pass it.
22	Service Vehicle	University Way; Campus Pkwy to 45th	3/8/2021 7:48	Same yellow seattle city light truck parked on street. Bus has to slow down and go around.
23	Service Vehicle	University Way; Campus Pkwy to 45th	3/8/2021 8:05	Seattle city light truck parked on side of road. Bus has to slow down to go arohnd
24	Service Vehicle	University Way; Campus Pkwy to 45th	3/8/2021 8:30	Yellow Seattle city light truck parked on side. Bus has to slow down to go around.
25	Service Vehicle	9th Ave; Alder to Columbia	3/10/2021 9:14	Ambulance (siren on) weaved through traffic to cross the intersection, causing the bus to stop moving
26	Service Vehicle	Pike; 3rd to 9th	3/12/2021 7:04	A service vehicle was standing on the way
27	Small Delivery Vehicle	Westlake; Denny to Mercer	3/16/2021 7:43	A streetcar was stopped by a delivery vehicle attempting to parallel park; streetcar honked numerous times. Interference lasted approximately 45 seconds but times recorded may be slightly off as I recorded this form at a later time.
28	Transit	Pacific; 15th to Montlake	3/3/2021 16:18	Front bus had ramp down for prolonged period of time and back bus attempted to pass it. Bus was so large, it was hard to navigate out of lane.
29	Transit	Pacific; 15th to Montlake	3/3/2021 16:32	Bus got in the way of this bus.
30	Transit	9th Ave; Alder to Columbia	3/10/2021 9:01	Shuttle bus blocked bus while trying to lane change.