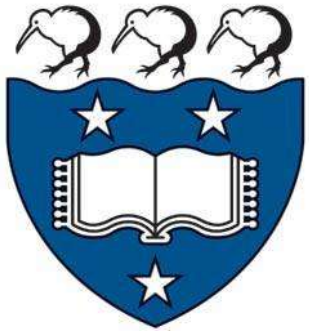
Real Time Bus Headway Estimation in Auckland New Zealand



Colby Roy Carrillo

Supervisor: Thomas Lumley

Department of Statistics and Computer Science

Date: 28th October 2019

A dissertation submitted in partial fulfillment of the requirements for a degree of Masters of Professional Studies in Data Science, The University of Auckland, 2019

Abstract

With on-time performance of buses within Auckland, New Zealand, already being studied by colleagues, this research is intended to complement their work by monitoring bus headway, a different metric for bus route performance. For frequent buses, customers will often go directly to the bus stop without checking when the next scheduled bus will arrive. Having buses each fifteen minutes late is better than having sequential buses on time and fifteen minutes late. These large variations in bus headway can positively or negatively impact public transport (PT) customer experience. This study looks to actively illustrate the current state of bus headway in Auckland, for frequent routes, and evaluate three estimator's performance at predicting bus headway. This is completed by using the Auckland Transport (AT) public application protocol interface (API), which allows us to capture real-time transit feed (GTFS) data in frequent intervals. Using the collected data, we are able to calculate the headway estimators throughout the day and provide the average waiting time of customers for routes with headway under sixteen minutes. We found that the three estimators of headway were not able to accurately predict the next buses headway, with correlations of the predicted headway to observed headway for a weeks' worth of data ranging from 16 to 27 percent. Even with these undesirable results we can visualize how headway changes throughout the day by publishing visualizations, showing the observed average waiting time, which allows for insight into Auckland Transport performance and for future work in the area to prosper. Information on the location of these visualizations or source code used for this project can be found at https://github.com/ColbyCarrillo/BusHeadway.

Table of Contents

Table of Figures

1

Introduction

1.1 Background

The background of this paper aims to give the reader basic information of expected knowledge to understand this paper. This will be done by covering Headway, Perception of Headway, Google's Transit API, Real-Time System vs Scheduled, and finally Auckland Transport.

1.1.1 Headway

Headway as defined by Merriam-Webster is "the time interval between two vehicles traveling in the same direction on the same route" [6]. This difference in time can be any possible increment of fathomable time, but in most research situations it refers to intervals which can be used as a gauge of performance or expectation. When we refer to headway and its relation to buses, we are speaking of the amount of time between buses, belonging to the same route, at a specific location, also known as a stop. From the customers perspective, it would be the time they would have to wait, if just barely missing the bus, for the next bus to arrive.

For frequent routes, those with headway under or around fifteen minutes, researchers are more concerned with the headway than whether the bus is on time. That is, if every bus is five minutes behind, the headway will remain even and customers average waiting time for buses would not increase. While if buses were tiered five minutes behind and five minutes early, there would be an increase in average waiting time for customers. For these frequent routes, it is not uncommon for passengers to arrive at the stop without checking when the next scheduled bus will arrive. Meanwhile for infrequent routes, ones with headway above approximately 15 minutes, users are more likely to check when the next scheduled bus will arrive before heading to the bus stop. For infrequent routes, researchers are more concerned with the bus being on-time than there being even headway between the buses.

More advanced PT systems will attempt to alter headway in real-time to handle fluctuations of demand, typically in the morning and afternoon peak times, but such fluctuations can be caused by numerous reasons, in example a sporting or music event held in the city. This alteration can be done by dispatching buses manually to keep even headway or even holding buses at stops. In terms of my research I am directly concerned with making average waiting time public and analyzing which of three estimators of headway, Observed, Mixed, and Current, best predicts the next observed headway of frequent routes.

1.1.2 Perception of Headway

The difference in time between the expectation, or scheduled headway, and the observed headway can heavily alter a PT customers' perception of the quality or reliability of the service they are receiving [17]. On the positive end, if the bus shows up early the customer already waiting for the bus will be delighted that they are already on their way to their destination. While the customer who arrives on-time for the buses scheduled arrival will be disappointed about missing the bus, and even more displeased if the following bus is running late, increasing the waiting time of that user.

In recent decades, the perception of waiting time was altered when the introduction of technology providing real-time updates of buses, whether via mobile or electronic signs was introduced. According to [17], the perceived waiting time of a customer at a bus stop is reduced by 20% due to the introduction of real-time information.

At the end of the day, the relationship between the public transport and its users is directly related to that of a service provider and consumer. The goal of the PT is to provide a service that is efficient, cost-effective, and retains customers, hopefully through high customer satisfaction and value for their money. If the patronage of the PT service drops too low, the cost becomes too high for a community to support and inevitably will require more community investment, tax dollars, to keep the service afloat. In most large cities, PT is a fundamental part to helping citizens complete their daily tasks in an efficient manner, as individual transportation is not feasible for every user, financially or infrastructurally. Without it, large cities are likely to see experiences of high congestion of roads as residents use individual transportation to travel to desired destinations. It is of utmost importance for the PT to not only provide cost efficient service, but to have positive perception from its customers.

1.1.3 Google's GTFS and Realtime GTFS

Google's GTFS and real-time GTFS are the system that Auckland Transport uses to make their data available to Google Maps, as well as the public. Google's Transit API's is a generic schema that has the goal of "making public transit data universally accessible". [1] When a city decides to participate in Google Transit, they must provide their static data, in example stops, routes, and trips information. This data must meet the required specification or schema as specified in [5], where it then can be integrated with Google Maps [4], which in turn allows for users of one of the most popular map services to see PT information for that city. For more integrated connectivity into the maps, a PT is eligible to submit real-time data providing geographic information and trip updates to Google Maps as well. [5]

The General Transit Feed Specification (GTFS) initial version was released on September 25th, 2006, and later GTFS Realtime initial version on August 22nd, 2011. [2][3] Since then, there are cities participating from all of the six inhabited continents, excluding Antarctica, of the world. [7]

By adhering to this schema, that Google uses to integrate into its map application, there are many positive byproducts. One being, a "standard" format of data storage for PT agencies, which in turn makes for easier analysis of every agency who participates. Another being, given the PT opens their data to developers as well, the open access allows for third-party developers and data enthusiast to gather, transform, and analyze PT performance. Each with their own goal which could be led by financial gain, insights, awareness, or even purely research.

1.1.4 Real Time System vs Scheduled

A Real Time System is referring to a system that provides updates, ranging in frequency from one to several minutes, of the current position and future estimated arrival of buses. [14,15,16] Real-Time updates increase positive perception, reduce waiting times of customers at stops, and are a cost-effective way to reduce the adverse effects of uncertainty of bus arrivals [17, 18, 19, 20]. These updates can be provided to customers via signs, mobile phone, internet, or even SMS [13].

A Scheduled System is referring to a system that only posts static timetables, where customers are only able to utilize what was scheduled and have no way of knowing the current state of the buses [13]. This means that if there are any large delays or cancellations the customer would not be informed of it and could potentially suffer from a poor experience of waiting for a bus that isn't coming. These scheduled systems are less common in recent years with how integrated technology has become in society but are not unheard of in some parts of the world.

1.1.5 Auckland Transport (AT)

In 2010, Auckland Transport (AT) was formed, during the creation of the Auckland council, as one of the council-controlled organizations [8]. Its main responsibilities were to provide Auckland's public land transportation needs but excluded the motorways. It had replaced the previous transportation service, Auckland Regional Transport Authority (ARTA), which had only served Auckland for six year [8].

AT's main responsibilities are, but not limited to, maintaining and running all the buses, trains, and ferries in the city. According to [9], this consists of 92 million individual rides dispersed amongst all the modes of transportation. Of this, the buses comprise 66.2 million of the rides or 71% of the total PT uses in Auckland. In peak times, there were a total of 1152 active buses in 2018, which has surely grown by the writing of this paper. Among these millions of trips, they provided 301,800 total weekly rides to students to and from school.

With AT embracing new technology, they manage all interactions with their service via "AT Hop" cards which are used to monitor customer usage and balances. Along with real-time tracking via Google's Transit API's, it has allowed for data to monitor and drive enhancements in the services that AT provides to the citizens and visitors of Auckland.

1.2 Motivation

The motivation of this research is to provide real-time visual graphs of average waiting time to AT users. This will provide the current condition of Auckland's bus PT and insights to its current headway performance. Along with these visualizations, the goal is to find the most accurate of three estimators for predicting the next headway. While monitoring the real-time data, we will be able to see how AT is performing in comparison to the service levels they are proposing in the RPTP [9].

With the release of Google Transit, the door has opened for integration of new technology with cities public transit data. Auckland Transport adopted the technology years ago, allowing public access to their transit data via developer API's. This data contains not only static information such as scheduling, but real-time updates and geographical location of every active bus in the city in at most 30 second intervals.

Accessing this data at frequent intervals, every 20 seconds, throughout the day allows for the creation of a slightly delayed real-time picture of the current state of the city's bus PT. This data can be as detailed as following a single bus on its trip through the city, monitoring when it reaches each stop, to seeing how observed time differs from the scheduled for every active bus on a given route, and as broad as monitoring whole divisions of the city.

According to [24], with inherent flexibility and low capital costs, transit is one of the vital services of the present and the future, with many citizens relying on some form of a transit system in their daily lives.

With growth in population of the Auckland by 120,000 from 2015 to 2018 and an anticipated population of two million by 2028, Auckland desires to use the public transport (PT) as a cornerstone tool to support growth and reduce congestion. [9] Large investments are being made into their PT system, $28 billion allocated for 2018 to 2028, Auckland is "all in" on providing an efficient and eco-friendly infrastructure to support life for years to come. [9]

With all these recent changes and investment into the city's infrastructure, I aim to provide a more public insight into the current state of frequent buses in Auckland. By monitoring bus routes that are "frequent", averaging under sixteen minutes between buses at each stop collectively for a day, I will be able to provide a real-time frequently updated image of how these routes are performing.

1.3 Research Question

For frequent buses, most customers will go directly to the pick-up location without checking when the next scheduled bus will arrive. Having the buses each fifteen minutes late is better than having buses sequential buses on time and fifteen minutes late. These large variations in bus headway can positively or negatively impact public transport customer experience. The goal is to visualize the current average waiting time within Auckland for these routes in real-time and analyze which of three estimators of headway is best at predicting the next headway.

This real-time analysis will shed light into how AT is performing in providing a desired "turn up and go" transportation service in the city of Auckland. As stated in [9], as of 2018 their rapid transit network (RTN) and frequent transit network (FTN) have a minimum headway of fifteen minutes for City Center Services, 6am-11pm seven days a week, as well as Non-City Center Services, during peak hours 7am – 7 PM seven days a week.

1.4 Scope

The primary objective of this project was to create a system to capture, analyze, and visualize AT data on headway in near real time. Along with this, we will use the captured data to calculate three estimators to try and predict the next buses headway. Finding which of the three estimators best predicts the future headway in different situations, such as adverse weather or peak travel times, is also of interest but not mandatory and can be used in future work. No other data sources are utilized in collaboration with this data source to help estimators make more accurate estimations.

1.5 Structure of Paper

Section one of this work contains an introduction covering the background, motivation, research question, and scope of the headway project. Section two will be on the related work which will discuss

research in the same field, specifically in topics of scheduling headway, maintaining headway, customer happiness and perception, and useful metrics to headway. Section three is about the System Design and Methodology which covers the logical layout, headway estimators, visualization, and methodology used in computing metrics. Section four will go into the results and output seen in our research. Finally, section five will go into the conclusion which also covers future work and is then followed by the references.

2

Related Work

2.1 Scheduling Headway

In [10] ,the authors provide a model to allocate bus companies resources to maximize net social benefit. They treat the problem as a constrained resource-allocation problem and find that this can be done by utilizing buses in-between scheduled routes and is dependent on total subsidy provided, fleet size, and levels of loadings on the vehicles.

In a later paper, [11] they consider two new service criteria in determining the scheduled headway: discomfort resulting from a vehicle servicing too many passengers at once, and its corresponding distance traveled (Crowding-over-distance), and the frequency with which a customer is rejected for the bus being too full (Probability-of-failure). This is tested using simulation and time-dependent Markov chains that are “inhomogeneous” respectively for COD and POF.

According to [12], at a basic level, when deciding the frequency of a route service to its users, a scheduler must try to balance providing an adequate service quality while trying to minimize the number of required assets to provide that service. This procedure must be constantly reevaluated using sensitivity analysis as what worked for yesterday may not be true for today. These evaluation tools utilize data referred to as “point check” and “ride check” which capture metrics at the maximum load point in its journey and metrics along the whole journey respectively. Using the metrics, [12] proposes four methods amongst two categories Max load methods and Load profile methods which use “point check” and “ride check” respectively. This all operates under the assumption that the observed metrics used for required headway calculation are based on a uniform passenger-arrival rate.

2.2 Maintaining Headway

There have been numerous research projects regarding how to best maintain the headway of a PT system over the past decades. This can be broken into monitoring the bus load, which helps manage customer demand with and in turn maintaining headway, and control systems which interactively help maintain even headway.

2.2.1 Bus Load & Headway

In a paper presented in March of 2013, [21], analyze creating scheduled with multiple vehicle sizes, focuses on two objectives: i.) Minimize the difference of observed headway and scheduled headway, ii.) Minimize the difference of observed load levels to the desired even-load levels. They hypothesized that with object i.) would increase the attractiveness of the PT service by reducing the

expected waiting time of a randomly arriving user and ii.) would increase the reliability of the PT for handling changing demands.

The algorithms developed were then applied to a bus line in Auckland, NZ, and found that with even headway their approach "can reduce the observed 38% discrepancy, from a desired load at the max-load point, to a discrepancy between only 0%-15% when utilizing properly available sizes of buses". It was noted that this could not be implemented on a "wide bus system with multi and interlining routes" without reviewing vehicle scheduling and cost. This was especially significant in routes that moves morning peak users into the Central Business District (CBD).

2.2.2 Control Systems

Control Systems is an important and heavily researched topic to try and maintain even headway for a given route. With one of the earliest research papers in 1972, [22] which analyzed whether to hold or release buses immediately at a stop in order to maintain even headway.

2.2.2.1 Dynamic Bus Dispatching

In a more recent study, [23] proposes using the Internet of Things (IoT) to dynamically dispatch buses to maintain the required quality of service. IoT, being a network of numerous objects, which are uniquely identified and able to interact with one another. In example, this method which incorporates, waiting queue at stops, GPS tracking, traffic light information, traffic density of roads, etc. to provide more data for a more holistic view of the current state of the PT. With this information the bus controlling center can adjust their solution and send instructions to the PT system on how to adapt. They found using simulation that the total passenger waiting time decreased by 13.30% , which showed their method was meaningful and the strategy can satisfy customer demand compared to static scheduling.

In [24], they present a multiagent negotiation system for a distributed control approach to maintaining even headway. That is, communication between the devices on buses and stops to determine dispatching at various stops in the route. In order to compare results, they compared their negotiation strategy against a No-wait strategy, Headway-based (even-headway) strategy, and Schedule-based (on-schedule) strategy by using simulations. They found that their negotiation strategy can find local optimal bus dispatching times in order to maintain even headway. They state that approaches such as on-schedule or even-headway work for some stationary transit situations, but do not suffice for nonstationary situations such as amusement parks, airport, or schools where their negotiation strategy has significantly better results.

2.3 Customer Happiness and Perception

In [27] and reiterated in [12], reported that passengers' perception of their local bus services was interpreted and ranked, out of 100, in the following manner: Reliability 34%, Frequency 17%, Vehicles 14%, Driver behavior 12%, Routes 11%, Fares 7%, and information 5%. For routes of concern in this paper, that means keep even headway between buses as that users do not have large waiting times between buses, as most users do not check the schedule before heading to the bus stop. Some metrics that attribute to reliability for passengers, according to [12], are Waiting Time, Boarding time, Seat availability, In-vehicle time, Alighting time, Total travel time, Transfer Time, Missed Connections, Pre-trip information time, and Pre-trip time required for changes in access path. Of course, the importance of each is subjective to each individual user of the system.

2.4 Useful Metrics to Headway

Below will cover a few different metrics that are of interest when it comes to analyzing headway. These metrics covered below are only a sample of the comprehensive list of metrics that concern PT networks and headway.

2.4.1 Waiting Time

The waiting time calculation for a passenger waiting for their bus varies in simplicity between reviewed papers. As presented in [25], that according to traditional models wait time is equal to half the headway for headways under 30 minutes, and equal to the square root of the headway in minutes for routes over 30 minutes. Since these traditional models there has been additional research done into how to calculate the waiting time between buses.

There is an issue of length biased sampling when considering headway alone, which makes the estimated waiting time important to bus headway. In example, in a given hour if buses are dispersed unevenly, their average headway may remain the same, even if their headways are not even. By analyzing waiting time, which incorporates the variance in the headway as seen above, the waiting time will increase even with no changes in the average headway, meaning that a user will be waiting longer for buses. The introduction of the variance penalizes the difference of actual headway from expected and will result in longer waiting times.

In [12], the waiting-time dilemma takes in the consideration of in-vehicle travel time, time already waited, by incorporating the possibility that the passenger will miss bus A to wait for another bus B that will have a lower overall travel time to their destination. With two assumptions: (a) buses never deny passengers for being full and (b) passenger arrival rate is independent of the vehicle-departure process and remains constant over the period of observed time. They provide a formula for average passenger waiting time under the above assumptions found in figure X:

$$E(w) = \frac{E(H)}{2}[1 + \frac{VarH}{E^2(H)}]$$

where w is the waiting time; E(w) is the average waiting time; E(H) and VarH are the mean and variance of the time headway H between vehicles respectively. The authors take into consideration many other factors which can be researched further in Chapter 12. [12]

In [26], which primary goal is how to compute average wait time of frequent stops while dealing with missing data. They define frequent routes as ones where users will randomly arrive at the bus stop instead of checking the next scheduled time. With similar assumptions to the above, the arrival rate uniform and implies that the demand for buses does not vary over the time of interest. While in real life, this is not practical as it is most likely to vary throughout the day, in example the morning and evening peak hours when users are commuting out of the central business district (CBD). They provide two formulas of for the average waiting time found in figure X and Y.

$$AWT = \frac{1}{2}\frac{\sum_{i=1}^{N-1} h_i^2}{\sum_{i=1}^{N-1} h_i}$$

where h_i are the individual bus headways for N buses

$$AWT = \frac{1}{2}\frac{\sum_{i \in F} h_i^2}{\sum_{i \in F} h_i}$$

where F is the set of indices, i, for which h_i is known

The second formula in figure 2, takes into consideration the potential missing headway values throughout the day. These could be cause by scheduled buses not running or issues with technology used to capture the real-time information of each bus. They conclude that of the six methods they used for dealing with missing data in calculating average waiting time that simply removing the data point from the equation worked best. Stating that, assuming the gaps in data could be identified, it was shown that discarding these gaps and estimating using the remaining data resulted in no substantial bias and performed well even when as much as 20% of the data was missing.

2.4.2 Excess Passenger Waiting Time

Presented in [26], was an interesting concept of excess passenger waiting time (EWT). In summary, it is a way of tracking how much the AWT differs from what the scheduled wait time should be. In an ideal situation the value of EWT is zero which means that the observed headway matches the scheduled perfectly, not late nor early. They present the formula in figure X below, which uses the scheduled headways for the buses used in AWT:

$$EWT = \frac{1}{2}\frac{\sum_{i=1}^{N-1} h_i^2}{\sum_{i=1}^{N-1} h_i} - \frac{1}{2}\frac{\sum_{i=1}^{N-1} h_{Si}^2}{\sum_{i=1}^{N-1} h_{Si}}$$

where h_i are the individual bus headways for N buses and h_{Si} are the scheduled bus headways for those buses.

$$EWT = \frac{1}{2}\frac{\sum_{i\in F} h_i^2}{\sum_{i\in F} h_i} - \frac{1}{2}\frac{\sum_{i\in F} h_{Si}^2}{\sum_{i\in F} h_{Si}}$$

where F is the set of indices, i, for which h_i is known, where $h_{Si}$are the scheduled headways, which can be summed over the entire time period, as long as they do not very during the time period of interest, otherwise the summation should also be taken using the set F of indices only.

3

System Design / Methodology

The design and implementation of the application to capture and analyze AT data was where most of the projects work took place. By creating a system that can pull, analyze, and report the real-time data, we are able analyze estimates of headway and provide the current condition of headway in Auckland.

The sections to follow are organized into: System Architecture which covers in detail the components of the system and ETL process, Headway Estimators which takes a detailed look into the headways captured and summarization process, Visualization which shows examples of our output, and finally a section on the Methodology which will cover our calculations and assumptions.

3.1 System Architecture

System Architecture section aims to create a simple illustration and explanation of the underlying application that was created for this research. The logical layout will cover the structure of the application which allows it to maintain static data while harvesting real-time data. This data, which is only maintained for an hour, is then utilized to compute the headway estimators and create visualizations of average waiting time. R [29], Python [39], and SQLite [33] were the primary tools used to create the application and were fundamental in the success of its creation.

3.1.1 Logical Layout

In the logical layout we will cover the structure of the application by its different logical components. These high-level components will be covered by Admin, Real-Time Data Ingestion and Storage, Headway Estimator Computation, Summarization, and finally the Visualizations. Each section will have its own sub-categories that comprise that component of the application.

3.1.1.1 Admin

Before we can jump into the pulling of the real-time data, we must cover the setup required to foster a location for the data to be stored.

The admin portion of the system contains all “down time” preprocessing tasks that are required so that the system can function as quickly as possible. These operations are usually completed an hour to several hours before the data starts being pulled from the real-time API which is scheduled for 6:00 NZDT. These parts are referenced in the logical diagram below in Figure 1 as ‘Static Data: Scrape Website / Download Zip’, ‘R Pull Scheduled Trips & Calculate Routes Headways’, and ‘Cleaning Process’ respectively.

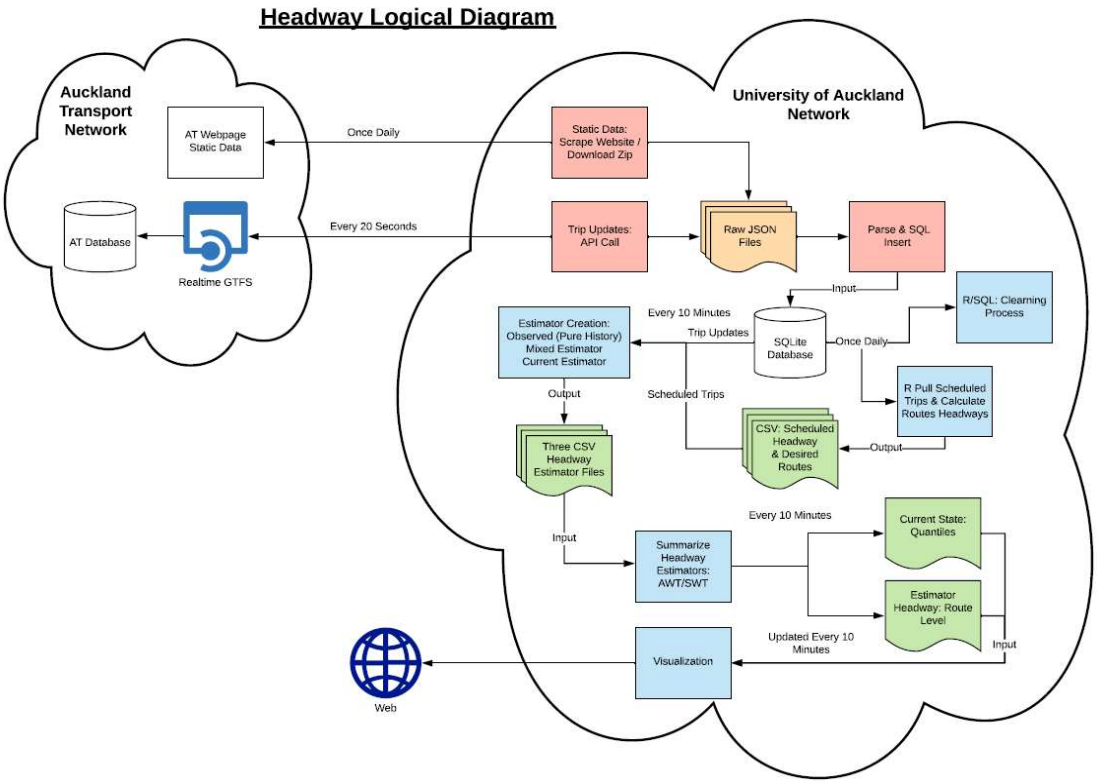


Figure 1: Logical Diagram of Application

3.1.1.2 Scrape Website/GTFS Static Data

The GTFS Static Data includes all the data that does not vary during the day, but instead is scheduled by AT beforehand for future operations. This data includes the information stored in the Calendars, Trips, Routes, StopTimes, and Stops tables of the database.

This data is acquired by calling a specific AT website page that holds the desired static data in a zip file. By checking daily if the “last updated date” text on the website is a more recent date than our last extraction date, we can keep the static data up to date without having to query the API directly. If we find that the data has been updated, we can download, extract, parse the data from JSON into SQL insert statements, and then insert the data into our database; all from the zip file located on the same webpage as referenced above.

3.1.1.3 Desired Routes

We define desired routes as those who have an average headway under the specified target parameter. For the sake of our initial study this has been set to sixteen minutes in order to capture FTN routes as well as those just on the edge of frequent. These “desired routes” are routes that contain customers who, as referenced before, are less likely to check the next scheduled arrival before heading to the bus stop, which justifies a random arrival rate.

To find these routes in the scheduled data, the difference in headway between each trip on a route, excluding the first trip, for every stop on the route, is calculated and averaged using a weighting of how many trips it contains minus one for the first trip. This average is then rolled up to become the route average which will be tested against our target parameter, sixteen minutes, to determine if this

route will be used in further computations and then saved to a csv file. This extract is then utilized when calculating headway estimators by reducing the input of the real-time data to data only on the routes that are of interest.

3.1.1.4 Cleaning Process

In order to keep a lean database, no real-time data is kept that is over two hours old, and no expired static data is kept past its labeled end date (found in Calendars table). For static data, this is executed at the start of every day, to remove data first from Calendars which then causes a trickle-down effect within Trips, Routes, StopTimes, and then Stops sequentially. For the real-time data, this requires for a process to be running during live hours, 6:00 to 22:00, that removes any real-time data with a timestamp over one hour from current time and is executed hourly.

3.1.2 Real-Time Data Ingestion and Storage

The real-time ingestion and storage of data covers the finer details of the API we use to extract the trip updates, along with the structure of the SQLite database that we use to host all the projects data.

3.1.2.1 Trip Updates API

The TripUpdates API provides real-time information on buses as they progress in their stops along the route. This is scheduled to be executed every 20 seconds from 6:00 to 22:00 NZDT. The API includes the below attributes in each record of a trip update, and many trip updates belong to a single json file returned by the API:

1. trip id: the id given to an individual trip or time slot of a route on a given day.
2. route id: the id of the route of stops and direction that a bus will travel to on a given day.
3. vehicle id: the id of the vehicle associated with the trip.
4. stop sequence: the stop in the route the update was at starting at 1.
5. stop id: the id of the stop the update was at.
6. delay: how far behind the bus is from scheduled, according to AT.
7. arrival: A binary flag, being 1 if the update was an arrival of a bus to a stop.
8. departure: A binary flag, being 1 if the update was a departure of a bus from a stop.
9. timestamp: The timestamp that the trip was updated.
10. api_timestamp: The timestamp that the api was called, this will be the same for all updates for a given call.

Each of these data points is ingested in a JSON form and parsed into a SQL statement which is then inserted into the TripUpdates table in the database. Due to our referential integrity constraints on trip id, route id, and timestamp, placed on the table, we do not have to worry about duplicates entries if the trip has not been updated in 20 seconds as it will fail to insert. After working with the data, it is not uncommon for a bus to send multiple updates at the same stop which presents multiple entries in our table. It is also not uncommon to see timestamps change throughout the day, sometimes presenting unreasonable times. We deal with this shortcoming later when calculating the estimators by only taking the latest of updates for a given stop, whether it is an arrival or departure timestamp, but sometimes can still provide bad estimates if the latest timestamp is unrealistic or made in error.

3.1.2.2 Database

An SQLite [33] database was used in this research for its ease of use and minimal setup required. It is to note that due to the requirement of concurrent transactions between the headway analysis and entry of new trip updates, a specific WAL paradigm had to be enabled for the application to work correctly. By updating this paradigm, we can have concurrent access to the same tables as earlier investigation found that queries or inserts would error, without this option adjusted, as the table would be “locked”.

Below, in Figure 2, is the database diagram showing the relationships between the tables held in the SQLite database.

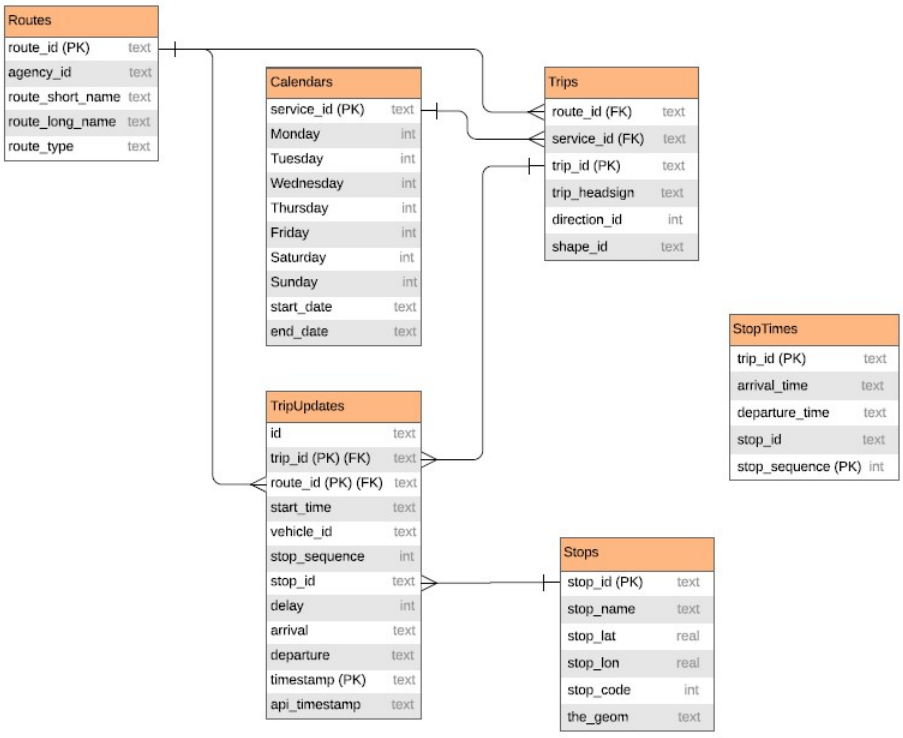


Figure 2: Database Schema

The most important relationship to calculating our estimators is from the TripUpdates table, which hosts all the real-time data, and the four connecting static tables, which hold referential integrity on the data. Note that StopTimes does not have any referential integrity requirements on any of its columns as I wanted everything from the data download to be entered regardless if it existed in Trips yet. This was to stop any data issues if AT uploaded a new StopTimes before releasing the Trip data to correlate with it. This StopTimes integrity is then later enforced, when querying the data, by joining StopTimes with related tables on its respective PK and FK.

3.1.2.3 Queries/Views

There are two main queries, which were created as views, that are used for the data extraction. One that extracts the trip updates data, which is pulled every time the headway estimators are calculated, and another that extracts the scheduled trips data, which is used to calculate the scheduled headway for each estimator estimate.

3.1.2.4 Indexes

There were several indexes created to increase the query time of the application. The biggest increase in performance was seen around indexing columns involved in querying the StopTimes table which contains a bulk of the data in the database. This includes indexing join fields on Calendars, Trips, and StopTimes tables used in our frequent queries. On top of this, SQLite optimizer also creates its own indexes to provide the best performance of the query in question. In layman's terms, indexes allow for the database to find records faster by looking at a small set of bookmarks in the data sequentially.

3.2 Headway Estimators

There are three headway estimators that are calculated from the real time data found in TripUpdates. Each has its own unique trait and value to our analysis that we later analyze, not in real-time, to find out which estimator is most accurate. We check the accuracy of the estimators by using the observed estimator, which will be referenced as the actual headway when joined back to other estimators, matched back up to our three estimators, using the trip id as an anchor.

The three headway estimators following are the Pure History headway (Observed Headway Estimator), Mixed Headway Estimator, and Current Headway Estimator. All estimators are calculated in 10-minute intervals starting at 07:10 and are saved to CSV files dedicated to the estimator and the day. The CSV files are the main output of our code and are later used for visualization and analysis including the summarization. The summarization, which is briefly covered in this section, is a roll up aggregation which allows us to calculate metrics, such as average and scheduled waiting time, at the route and stop levels for each estimator.

3.2.1 Pure History Headway (Observed Headway Estimator)

This estimator contains observed data and provide the "what the customer experienced" information. This is calculated for every route by looking at every stop for that route observed in the trip updates data. We then look at the last two timestamps for that specific route, at the stop, and take the difference of the timestamp of the updates. This provides us with a numeric minute value of how long it took for two buses of that route to reach that stop also known as the headway.

If we see that there have not been two buses that have visited the given stop, in example the first bus trip of the route, then there will be no headway and a NA will be produced. NA's can potentially be seen at the start of a routes daily activity when only the first bus has run or later in the day if no buses have run in the past two hours and have been removed from our database.

Then what the scheduled headway should have been is captured, by finding the earlier of the two last buses, and finding its associated data in the scheduled data. The difference in time between it and the next scheduled bus is then calculated, providing us the scheduled headway for that specific sequence of buses. A simplified visualization of this calculation can be seen in Figure 3 below. You can see that when the calculation runs at Time 1, we would produce the difference in time between Trip 456 and 123 for Stop 10.

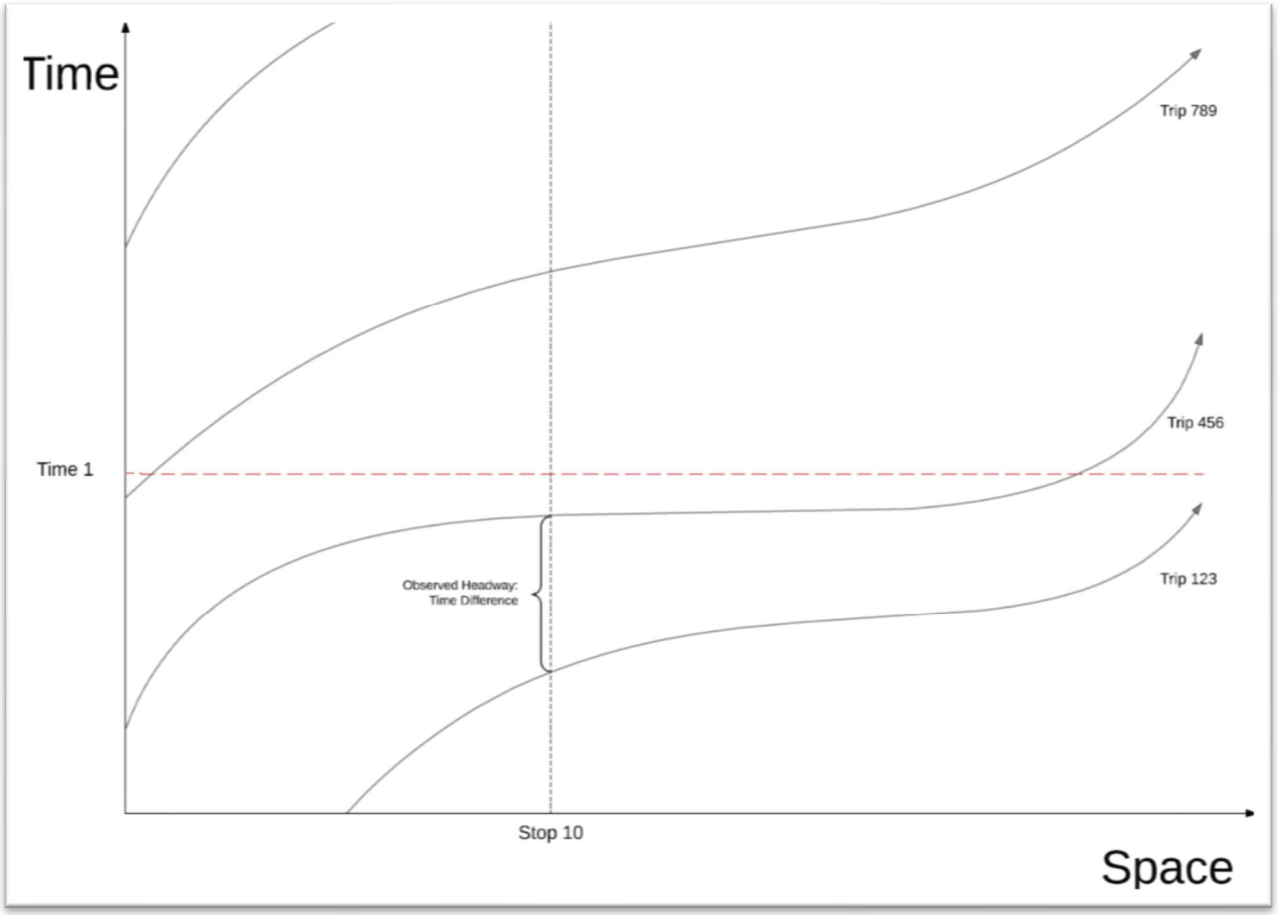


Figure 3: Observed Estimator Calculation

3.2.2 Mixed Headway Estimator

The Mixed Estimator utilizes both the observed and estimated travel times, derived from the schedule, to give us an estimator that used both historic and expectation to estimate headway. This estimator is calculated for every route and its respective stops that were observed in the TripUpdates data, similar to the Observed Estimator. For every stop, we first look at the amount of time since the last bus was at this stop by taking the difference of the current time, and timestamp of that bus at that stop. Then the next trip to arrive at this stop is found to estimate the amount of time that it takes to travel from its current last updated stop location to the stop in question. This expected travel time is derived by finding the difference in scheduled arrival times at all the stops in the route from the scheduled data, only utilizing its current stop to arriving at the stop in question. By combining the time since last bus here and expected travel time of the next bus to arrive at this stop, we are provided with a mixed historic and scheduled estimator.

In the case that there is not another active bus on this route currently traveling to this stop, the process will look at the next bus in the schedule that comes to this stop. It then combines the time since the last bus was at this stop, time until the next scheduled bus will start, and the estimated travel time until that scheduled bus will arrive at this stop, given that a bus is scheduled to start in the next hour. Otherwise in the situation that another bus does not start for the next hour, NA's will be produced.

Similar in all the estimators, we also capture what the scheduled headway should have been by using the last bus at this stops data in the scheduled data as an anchor. We then take the difference between this scheduled time and the next bus to arrive in the schedule to derive the scheduled headway for the route at the stop at a given time.

A visual representation of the Mixed Estimator is provided below in Figure 4. Note that there are two situations in this visual, Time 1 and Time 2. For the situation in Time 1, the estimator value will be A + B or the time since the bus was at this stop added to the estimated time for the next bus to travel here. While in Time 2, there is currently not another bus running, so this adds in another component to the estimate, the scheduled time until the next bus starts C.

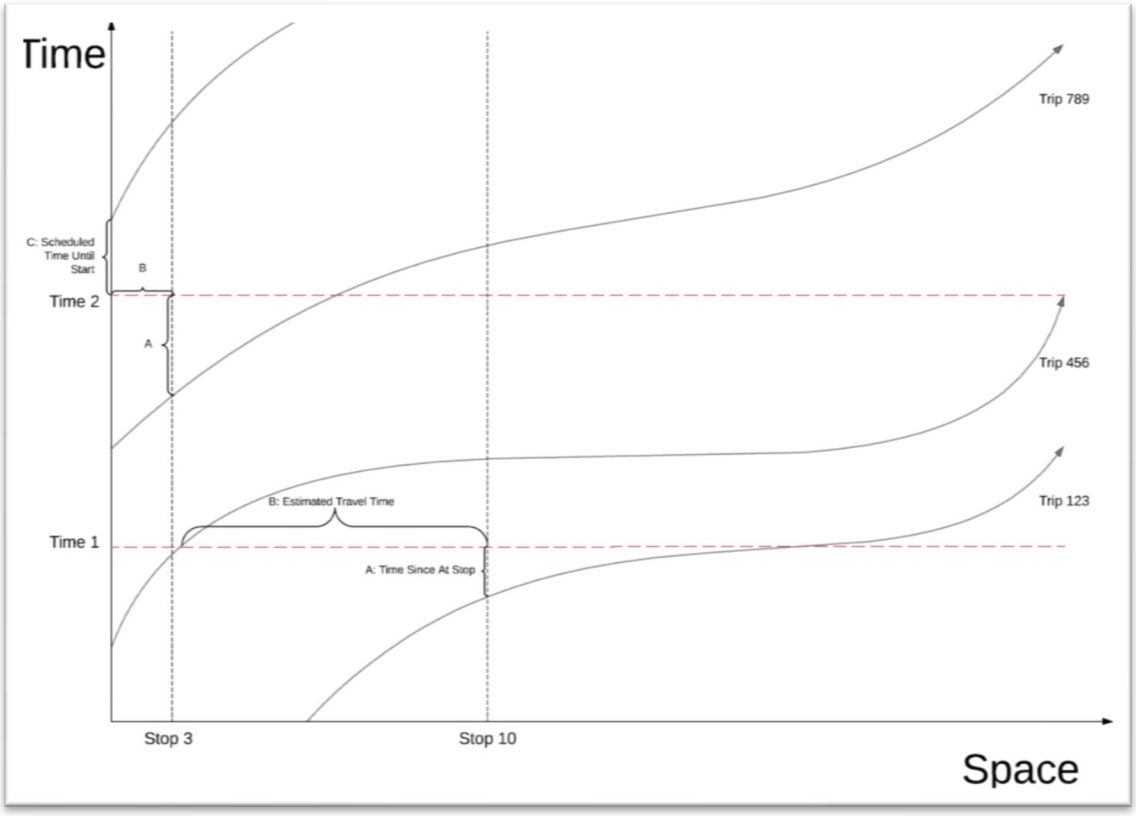


Figure 4: Mixed Estimator Calculation

3.2.3 Current Headway Estimator

The Current Headway Estimator is different than the other two estimators in that it is only calculated for every active bus and does not calculate for each individual stop on the given route. It utilizes only scheduled data, by calculating the expected travel time, the same as it was explained in the Mixed Estimator, of the current bus and the bus before it. We do not have to derive an estimate for every stop as this difference between buses at a given time will output the same value regardless of the perspective it is calculated. In example, the difference between bus A and B will always be 12 minutes regardless if you look at it of the perspective of stop 11 where it is six minutes for B to arrive here and six minutes since A was last here, or 18 where it is 8 minutes for B to arrive here and 4 minutes since A was last here.

If there is no active bus before the bus in question, the estimator will find the next scheduled bus and use its time until start and expected travel time to the current stop as the value of the estimator.

NA's can be produced when there is no bus that is scheduled to run in the next hour. As well, it is common to see values of 0 for the current headway if the live buses are at the same stop, also known as tailgating. It is not uncommon to see bus bunching as it is a common side effect to current operations.

Similar to the other two estimators, what the scheduled headway should have been is captured again by using the scheduled data for the current bus in question. We then find the difference in time between this stop in the schedule and the next arriving bus to this stop to give us what the headway should have been if it was following the schedule.

A visual representation of the Current Estimator is provided below in Figure 5. Note that there are two situations, Time 1 which has another live bus running behind it so the estimate returned will be the estimated time of travel between the two buses. While in Time 2, there is currently not another bus running, so it incorporates the time it takes for the next scheduled bus to start (B).

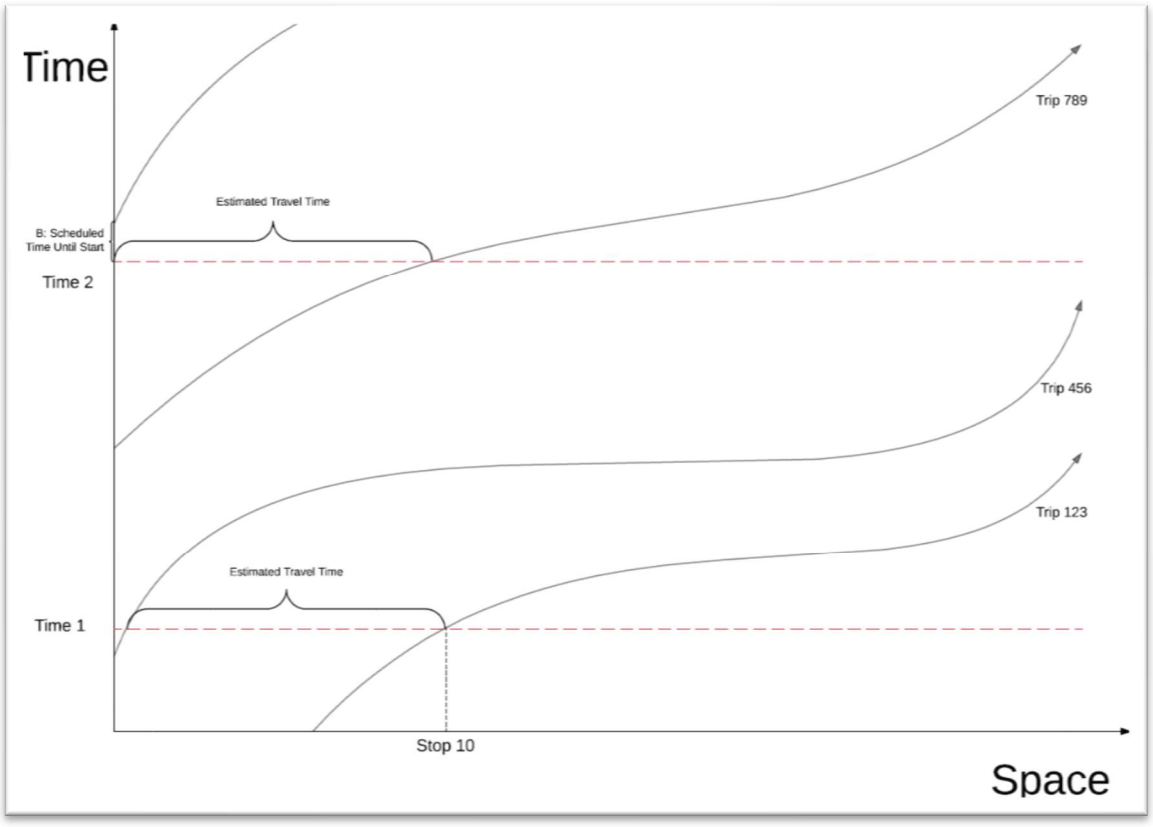


Figure 5: Current Estimator Calculation

3.2.4 Summarization

One way that the estimator's output is used is by creating summaries of the estimators, using metrics of interest, which is run directly after the estimators are calculated. The main objective of these

summaries is to provide Average Waiting Time, Scheduled Waiting Time, and Comparison of the two equations.

These summaries are run at the route and stop level of aggregation for the latest calculated data in each of the estimator files and are covered briefly below.

3.2.4.1 Route Summary

For the route level summary, we ingest all three of the estimator csv files which include not only their estimations but the respective scheduled headway for that estimation. We first reduce any NA's from either the estimate or scheduled values as learned in reference to [26]. We then can easily iterate over every unique route found in the data, calculating the AWT, SWT, Comparisons, and Quantiles of each calculation. The calculations are then used to create visualizations of how each route's respective calculations change over the day from 07:10 to 22:00 NZDT.

3.2.4.2 Stop Summary

While similar to the Route Summary, the main difference in Stop Summary is that we compute calculations from the perspective of the stop, and each route that services that stop. We only want to use complete cases of records that contain both the estimate and scheduled values as NA's will skew our output. For each unique stop we calculate each route that services that stop's AWT, SWT, Comparisons, and Quantiles of these calculations. The calculations are then used to create visualizations of the state of each stop in Auckland city, utilizing geographic location to be shown in map form.

The reason for this separate summary is for the case that a user is not worried about the headway of the route they want to catch, but instead the headway of that route at the stop they use. The route level calculations are averaged amongst the stops and therefore are not the same at every leg in the route. This extra level of granularity provides a more specific calculation to customers interested in particular stops and not the performance of the route as a whole.

3.3 Visualization

There are two types of visualizations that were created to represent the route summaries which can be broken into Metric Visualizations, such as scatter plots covering AWT, and Map Visualizations. Though there was the potential to have a Metric Visualization for every estimator and potentially every metric, the goal was to have output that could be displayed on a single web page. This meant that we needed to find the most accurate estimator and useful metric to display the current state of Auckland's Transport. While the situation was similar for the Map Visualizations, since the Current Estimator does not incorporate stops it is not a contender to be displayed.

3.3.1 Metric Visualizations

For our Metric Visualizations, interactive plots were created using plotly [34] which look at how the metric changes over the day, created at ten-minute intervals. This was created for the AWT and the difference between AWT and SWT which I have labeled as "lateness". This value can take positive or negative values depending on whether the AWT was greater than or less than the SWT. In the below figures I have shown examples of the lateness of the Observed Estimator as well as the Actual Average Waiting Time (Observed Estimator) for a given day.

One can note that the red line in diagram for Observed Estimator lateness, Figure 6, below is arbitrary in its placement and is just used a visual placeholder for where 10 minutes both positive a

negative is in respect to the point. I choose ten minutes as all the routes that we are analyzing have average headway under sixteen minutes for the day, so if the route is more than ten minutes late or early this is a large deviation in its headway and therefore average waiting time. Similarly, the line in the Actual Average Waiting Time graph, Figure 7 below, is placed at sixteen minutes, which was the cutoff for routes to be considered. Both examples are for the same day, Friday October 18th, 2019.

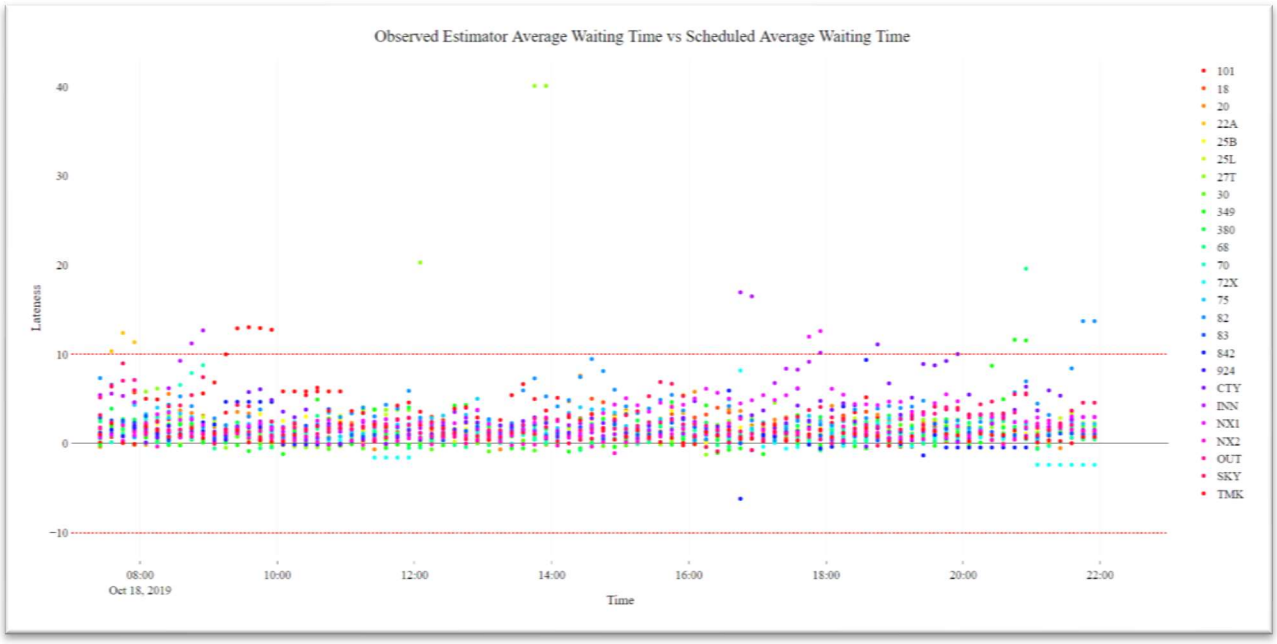


Figure 6: Observed Estimator Visualization of AWT

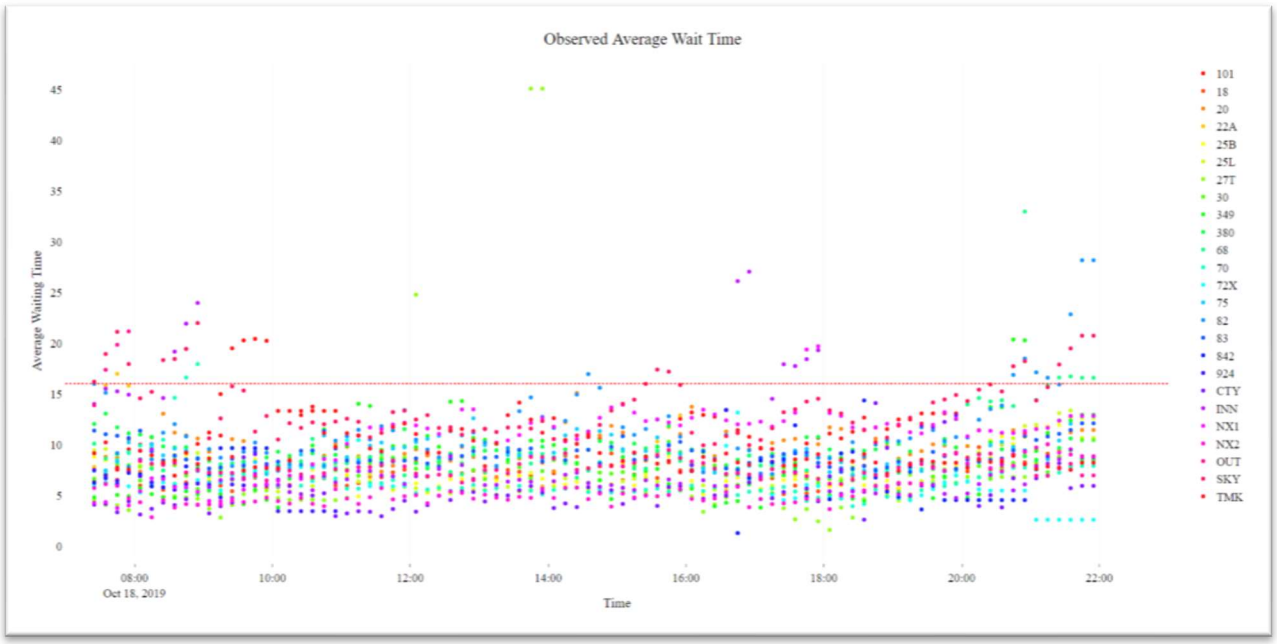


Figure 7: Observed Average Wait Time

3.3.2 Map Visualizations

The Map Visualization attempts to intertwine our metrics of interest and geographic locations of stops into a single view. The dots display only the latest output of the stop summarization, which means that stops can appear and disappear throughout the day if they are not in use. A user interacts with the map by looking at all routes or only the route they are interested in, which can be changed in the top right-hand corner and view the lateness of that route at their respective stop. By clicking on the individual circles, a popup will display showing the details of that stop. The circles are also color coded by their value of lateness, which will turn yellow when lateness is greater than five minutes positive or negative and turn red when greater than ten minutes positive or negative. A small example of the Inner Link, Outer Link, and City Link is displayed below in Figure 8.

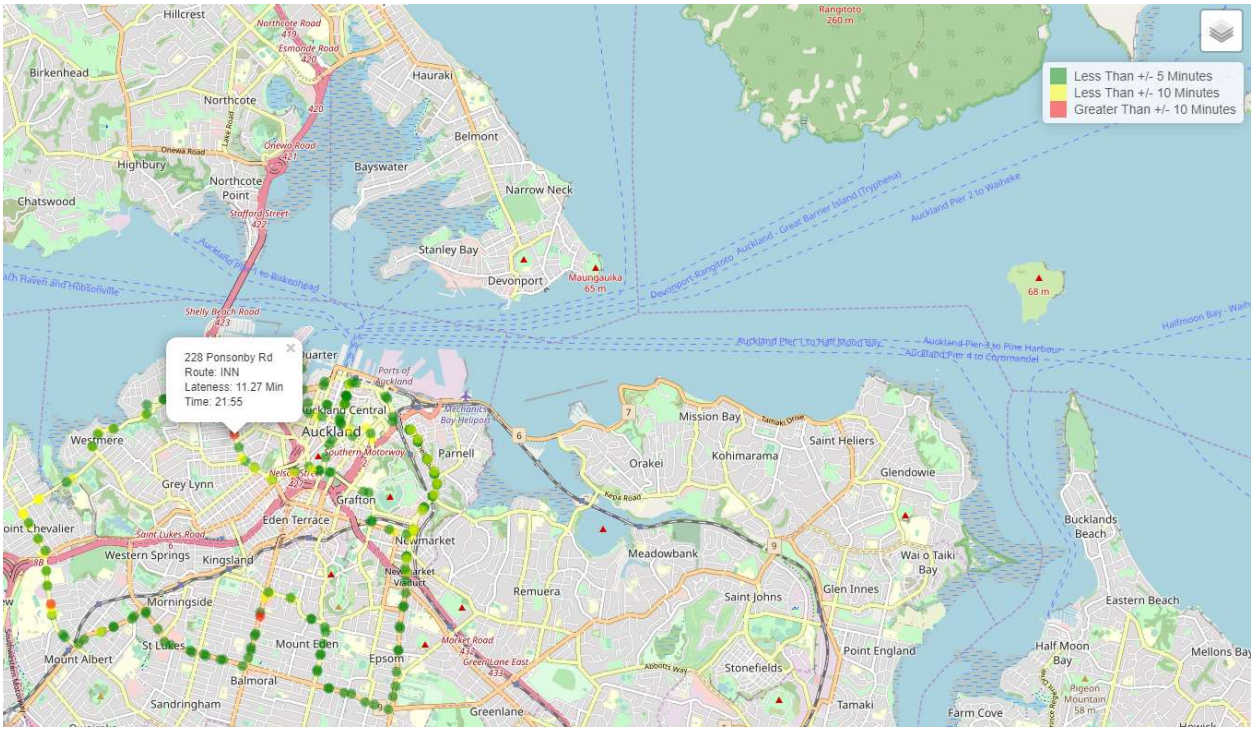


Figure 8: Stop Level AWT Visualization

3.4 Methodology

In the Methodology section, the sections are broken into an explanation of the equations used for the metrics, the evaluation metrics for each estimator, and finally the assumptions of the project and the data. Due to the length biased sampling, average waiting time is more important than the average headway for a given route. Although estimators will be evaluated based on their ability to determine the next buses headway, average waiting time provides useful information that customers are concerned with.

3.4.1 Different Calculations/Metrics Used

In our summarization process we used several metrics which we will cover in depth now. These metrics include Average Wait Time (AWT), Scheduled Wait Time (SWT), Comparison Ratio, and Lateness. Some of these equations are seen in related studies while others were provided by my supervisor [37]

3.4.1.1 Average Waiting Time

Average Waiting Time is the time a PT customer would have wait for their bus, given that they arrive in a Poisson Process, or that their arrivals are independent of the last buses arrival or how many other passengers are waiting. Though this is covered in many different papers in related work, in our work we must anticipate for the potential of missing data, since the data is collected systematically without any human intervention. Therefore, we use the equation found in [26] which allows us to calculate AWT for records that have both the scheduled estimator time as well as an actual estimator value. We require both as we do not look at AWT alone but instead incorporate it with SWT in the Comparison Ratio. For this equation to be accurate we take our all records that do not contain NA values so that the sample used for both metrics is the same.

$$AWT = \frac{1}{2}\frac{\sum_{i\in F} h_i^2}{\sum_{i\in F} h_i}$$

where F is the set of indices, i, for which h_i is known

3.4.1.2 Scheduled Waiting Time

The Scheduled Wait Time is an equation that is used to find what the Waiting Time of a PT customer should have been, given that the routes headway was perfectly on schedule and evenly spaced. This equation was derived from my supervisor [37] and looks only at the mean of the scheduled headways multiplied by one half.

$$SWT = \frac{1}{2}\frac{\sum_{n=1}^{n} h_n}{n}$$

where n is the number of scheduled headways found in our AWT calculation, given both are not NA. h_n represents the scheduled headway for trip n

3.4.1.3 Comparison Ratio

The Comparison Ratio is simply the ratio of the AWT to the SWT and identifies how late or early a route is. In the perfect scenario when the actual headway is identical to the schedule headway we will find this ratio to have a value of one. When the actual headway is delayed the ratio grows above one and vice versa for when the actual headway is quicker than the scheduled headway.

$$Ratio = \frac{AWT}{SWT}$$

We would expect to see that over long durations of headway that the ratio will gravitate closer to the value of one. As well, if we see the ratio to swing in one direction away from one, we would expect it to swing back in the opposite direction as more buses run. That is, given that there is a predetermined number of trips that will run for a route, if a sequence of buses is all early and the ratio is less than one, we expect when later buses run to see it swing in the other direction to being late. We

expect this as if the latest buses are running early, it means that the waiting time for the later buses, given it is running on time, will increase.

Due to the metrics gravitation towards the value one, we calculate all these metrics for only the most recent data. By capturing this in ten-minute intervals we aim to catch the swings in the metrics as they begin to balance out between being early and late.

3.4.1.4 Lateness

Lateness is the calculation of the estimator's headway minus the scheduled headway for the given record. We use this value in our visualizations as a comparison of the calculated headway from what the scheduled expects. This value is negative when the buses are closer together than expected and positive when the buses are further apart than expected. In a perfect scenario, the lateness would be equal to 0, meaning that the estimated and scheduled headways are equal.

3.4.1.5 Mean Squared Error

Mean Squared Error is a common and basic evaluation method for checking how close a prediction is from an actual. It utilizes the squared difference between the predicted and observed value and averages over the total number of observations. The formal equation is below which was referenced from [28].

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

3.4.1.6 Mean Absolute Error

Mean Absolute Error is similar to MSE but instead of squaring the difference between the observed and the predicted, it takes the absolute value. This is useful for dealing with outliers in the data that we should not be concerned about. The formal equation is which was also referenced from [28].

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

3.4.2 Assumptions and Limitations

Since this data is collected in real time there are many weird data situations that can happen that are hard to explain. In summary, trips not collecting data, trips starting and then never progressing past the first stop, trips running at the incorrect time, trip timestamps updating for a given stop well after its last captured timestamp at that stop, and trips providing updates for stop sequences beyond their scheduled last stop.

On top of this, there is always the possibility that AT could reconstruct how they currently work with their bus data, including: overhauls to their versioning techniques, date formation (currently YYYYMMDD), or even their schema. Below we will cover some of the assumptions and limitations that I have thought of for potential changes or deviations in the data collected. Given these errors, readers should note that these estimators are subject to errors if there are errors in the underlying data. This can be caused by bus operators entering incorrect trip or route id information as well as issues with real-time data not being captured or updated. There are several situations where I make assumptions as to

the integrity of the data collected, as the time permitted to complete this project does not allow me to code for every one-off situation that can happen. This list is not all inclusive and bugs in the code and/or data can happen at any point in the future, the task for maintaining this application is an ongoing process.

I have placed some catches for capturing faulty data when calculating the scheduled headway. This will capture the situations where the trip is running at the incorrect time, if and only if the trip is running at a two-hour difference then what it should be. This is done by taking the time of the observation and pairing it with the closest scheduled trip for that stop and route.

When calculating headway for routes that have the first and last stop sequence at the same stop location, the headway is calculated normally. There is no solution coded for the situation where a user arrives at the last/first stop, they must wait for the bus leaving stop sequence 1 and not the last to leave on the bus.

Any active buses that are shown at the first and last stop of a route are removed for the Mixed and Current estimators, meaning that the first active bus not at the first stop will look into the schedule for the next bus. This was due to multiple buses being active at the first stop and never moving past this stop.

We do not consider buses that operate before 06:00 or after 23:00 in calculating our desired routes, as we are mainly concerned about frequent routes during the popular parts of the day.

We do not compensate for routes that use many of the same stops during the middle of their route, that a user could get onto any of the buses that arrive, if they only needed to go between those shared stops. Examples in Auckland's Transport would be any route with 22 as a prefix which share many of the same stops but start and end in different areas of the city. Each routes headway is considered independently of each other.

When calculating the Average Wait Time and Scheduled Average Wait Time, we only consider trips that have both values not NA.

Though these errors can occur we assume that these situations make up only a small number of the total trips that will happen and will not hurt the overall evaluation. These are noted for readers to be conscious of this when using the numbers for any further research.

4

Results

In our results we will aim to cover the evaluation of the three different estimators' predictions compared to the actual headway observed. This is done by joining the Observed Estimator back with itself and the other estimators in order to see how close the next actual observed headway compares to the value we estimated. All references to Actual Headway are referencing to the Observed Estimator in a future state of itself, or in other terms the headway observed after the estimators were run.

We look at the correlation, distribution, and errors of the estimators performed from the perspective of all of Auckland, the route level, and time of day in order to provide a more in-depth analysis than just evaluation over the full day. To complement this, we plan to investigate the results of how the AWT, and Lateness change in different situations, such as emergency events. Due to the time constraints, seven days' worth of data will be used in the analysis, but further analysis can be done at any point as the data collection process is ongoing.

4.1 Estimators Evaluation

When analyzing how well the estimators are predicting the actual headway, I utilize both MSE and MAE as well as plotting the correlation between the predicted and actual headways. Some of the different situations I looked at besides using the full days data for the evaluation was by route and by time interval containing morning, mid-day, and afternoon. The idea was that some estimators may be better at predicting different types of routes as well as may be able to capture the peaks in the morning and afternoon better. These time intervals were broken into, morning 07:00-11:00, mid-day 11:00:01-14:59:59, and afternoon 15:00-19:00.

We first look at the correlation and found that the relationship between each of the estimators predicted time and the actual headway was low, ranging from 16% to 28% overall and stooped lower depending on the day. Below, in Figures 9, 10, and 11, we see the correlation between each of the three estimators and the actual headway over all of the data, with the number of datapoints being 394,336, 485,738, and 53,354 respectively.

One interesting point in the graphs below are the clusters that are formed on either side of the red line for the Observed and Mixed Estimators. These clusters are a result of frequent uneven headway within the PT system, that being they frequently have the same sequence of uneven headways. These delays in headway seem to be normal or expected over an extended period time of a week and may be due to traffic or peak travel times.

The predictions and observations that are seen along or close to the 0 axis are due to buses tailgating. This is a phenomenon when a bus starts and picks up PT customers frequently or get stuck in

some sort of delay. The next bus to leave may find that it does not have to stop at many stops as there is no passengers waiting, allowing for it to continually drive, catching up to the bus in front of it. This situation will result in the headway between the two buses to be very short while the time until next bus arrival will be very long, given that they all start at the time expected. The estimators will then predict that the next bus will be like the distance in time between the two observed buses, resulting in a poor prediction. We do not see these in the current estimator as the nature of the estimator only captures a few samples, one for every live bus at the time the estimator was calculated.

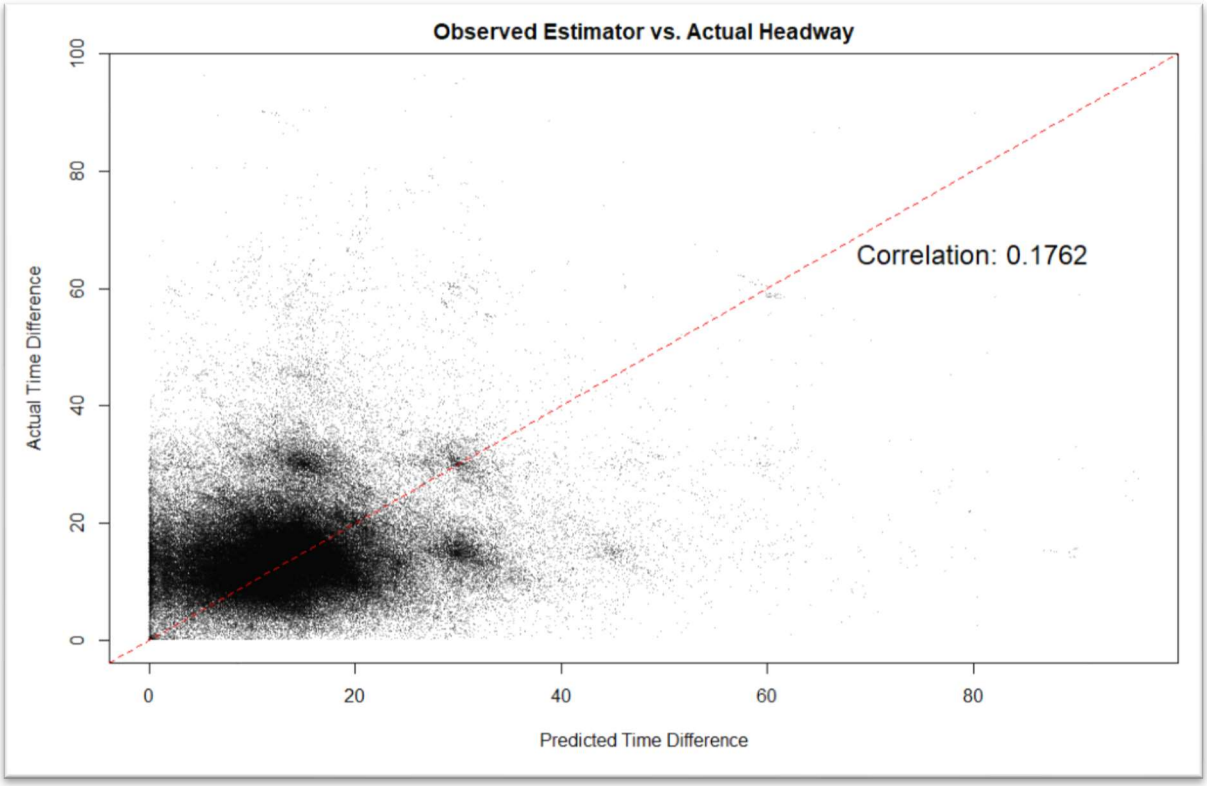


Figure 9: Correlation of Observed Estimator Prediction to Actual Headway

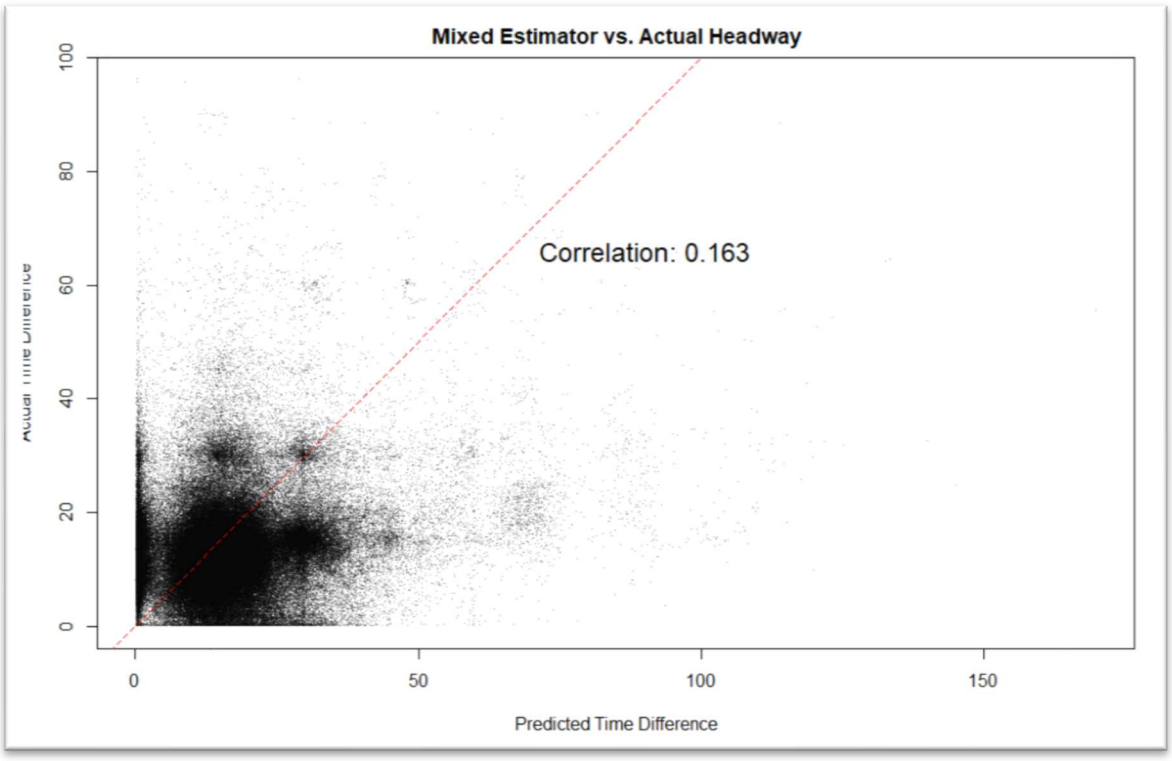


Figure 10: Correlation of Mixed Estimator Prediction to Actual Headway

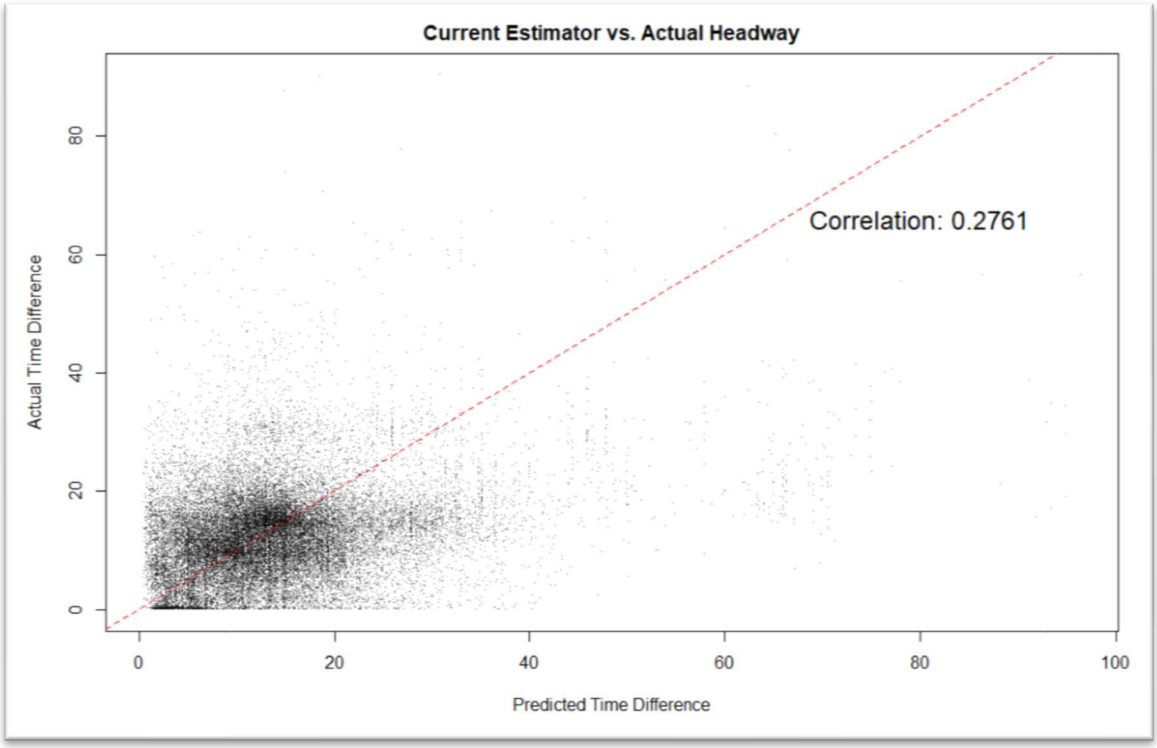


Figure 11: Correlation of Current Estimator Prediction to Actual Headway

Following analyzing the correlation over the full week, we look at the correlation of the estimators at different time intervals, shown below in Figures 12, 13, and 14. The reader should note that the morning set has 105,398, 118,853, 13,148 records respectively, mid-day has 111,082, 140,715, 15,157 records respectively, and afternoon set has 111,630, 141,204, 17,306 records respectively for Observed, Mixed, and Current estimators. Looking at the figures, we do not see as much of the clustering phenomena above in the morning time interval, compared to mid-day and afternoon where it is more evident. This may be due to heavier traffic and or passenger demand in these time intervals.

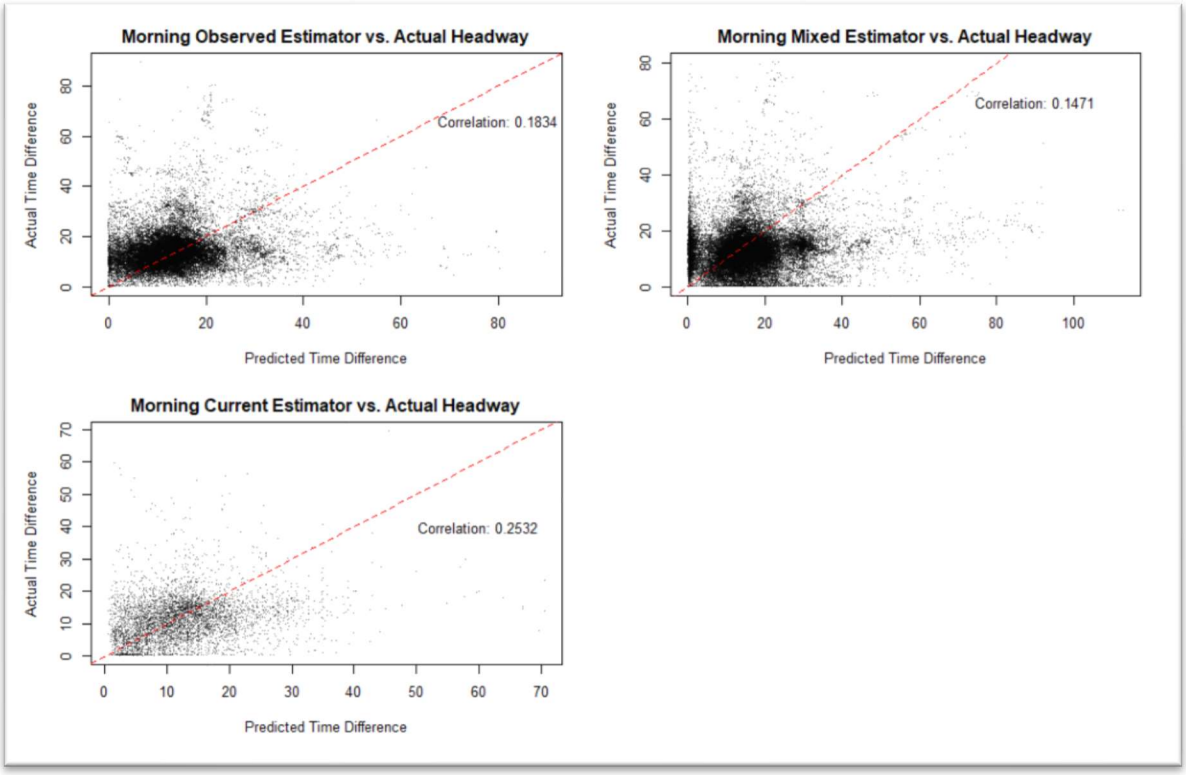


Figure 12: Correlation of Estimators Predictions to Actual Headway in Morning

We also note that there seems to be much larger delays in the Mid-Day and Afternoon segments of the data, with the axis of all the estimators ranging much further. This could also be due to the number of buses that are running in each of these four-hour intervals.

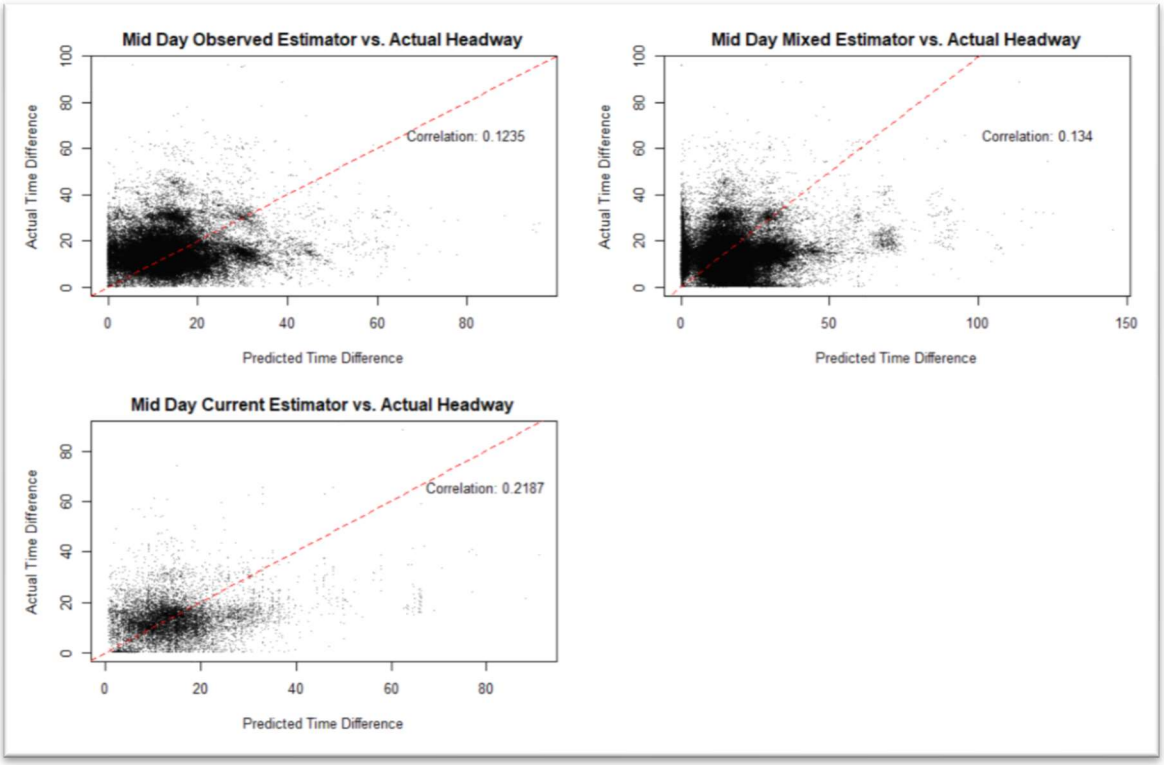


Figure 13: Correlation of Estimators Predictions to Actual Headway in Mid-Day

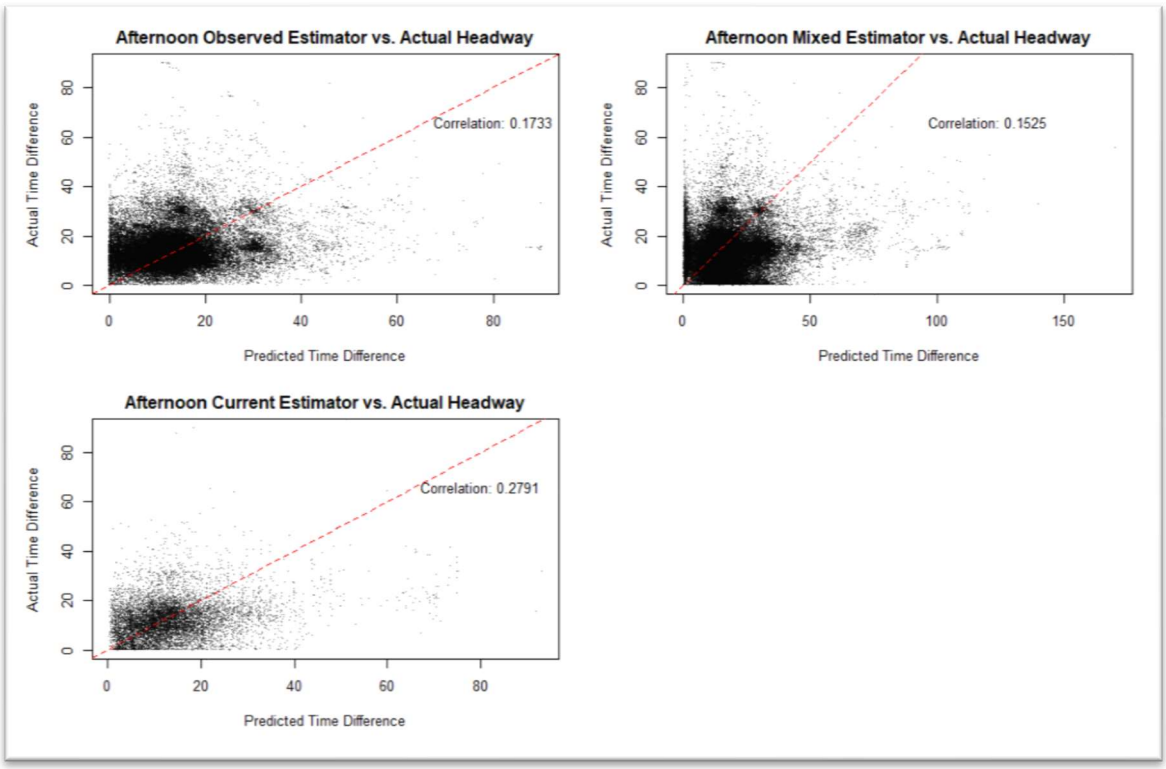


Figure 14: Correlation of Estimators Predictions to Actual Headway in Afternoon

The final analysis in terms of Correlation between the estimator's predictions and the actual headway is to look at how they performed at the route level over the seven days. The values observed vary dramatically between positive and negative correlation with one of the best seen in route NX1 seen in the Appendix. We only show one example in Figure 15 below for route 70 while and all other frequent routes plots that shown in Appendix A.1.

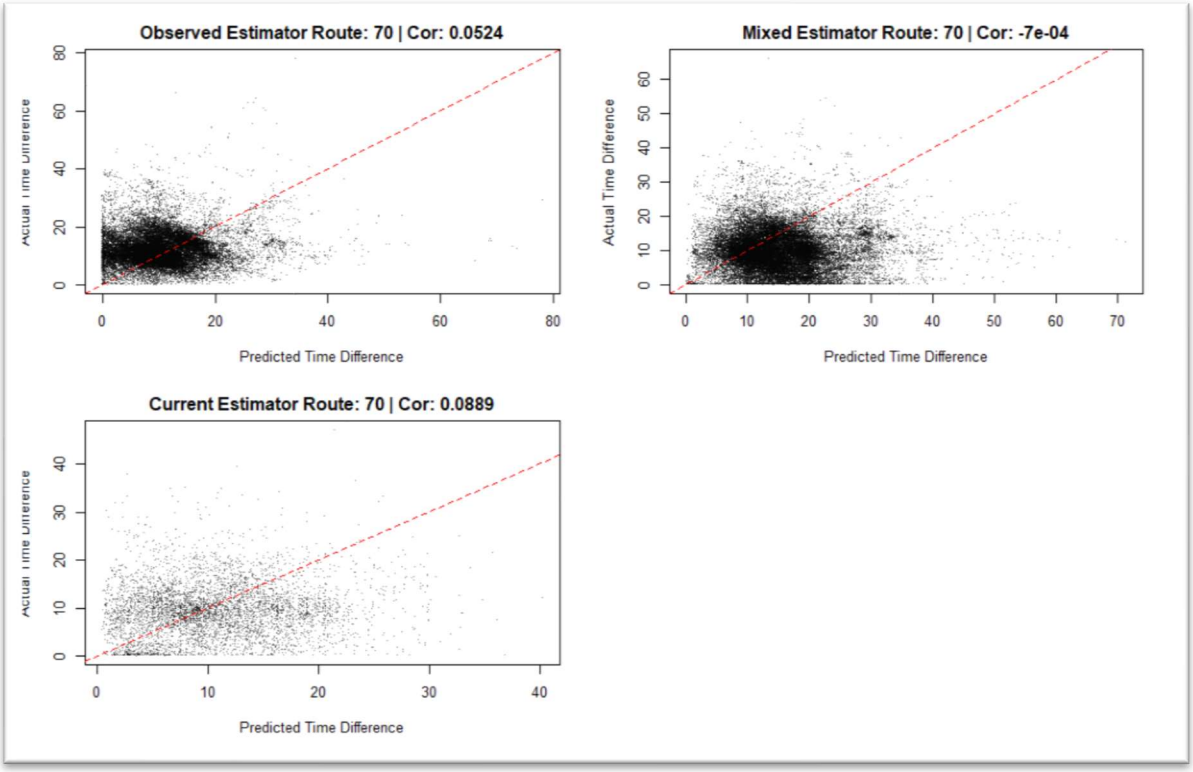


Figure 15: Correlation of Estimators Predictions to Actual Headway for Route 70

Due to the lack of correlation between the estimators, I was advised to make sure that the distributions of the estimators are the same at a given sampled point in time, with this QQ plots shown in 16. The idea behind this is that even if the estimators are not good at determining the next headway for a given stop or bus, that they still are able to illustrate the current state across all of Auckland. Since we have our Observed Estimator which is the actual headway delayed by one-time interval, we compare the other two estimators to this estimator for a sampled point in time. Ten randomly sampled times from the data were collected and Q-Q plots were made, one is shown below in Figure 16 and the rest of the nine other sampled plots are shown in the Appendix A.1. We see that the distributions between the observed estimator and the mixed and current are within reason the same and does justify them being used to predict the overall current state of Auckland.

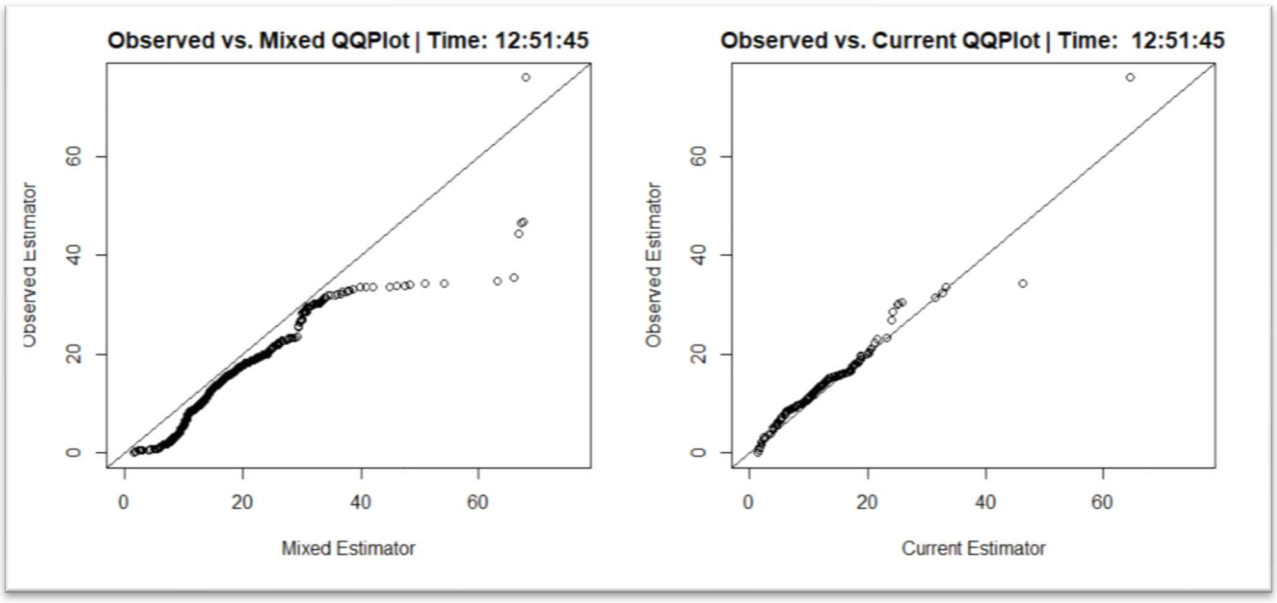


Figure 16: Q-Q Plot of Observed Estimator (Actual Headway) to Different Estimators

In order to complement the analysis of Correlation we take a deeper diver into the MSE and MAE for the full data, each route, and same time intervals as observed in the correlation analysis. First we look at which of the three estimators has the least overall error shown below in Figure 17. We see that for both MSE and MAE the Observed Estimator has the lowest amount of error over all the data which is interesting as it did not have the highest correlation as seen above.

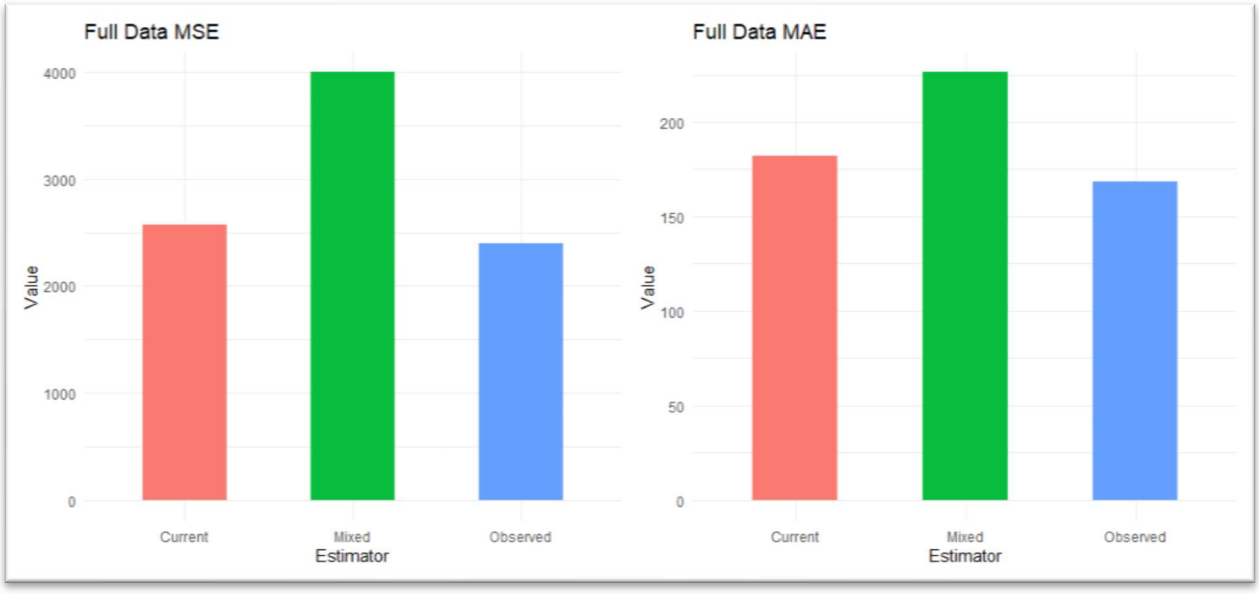


Figure 17: MSE and MAE for Estimators Over Full Data

We then look how the MSE and MAE is distributed amongst the different routes, with a visualization shown in Figure 18 below. We see that some routes are heavy contributors to the overall MSE and MAE such as the 22A and SKY bus routes. It is sensible for the SKY to have higher MSE and MAE

as it is the route that picks up all over the city and travels to the airport. Since it travels more of the city than the normal route it is more likely to be affected by traffic or other incidents that cause delays in headway.

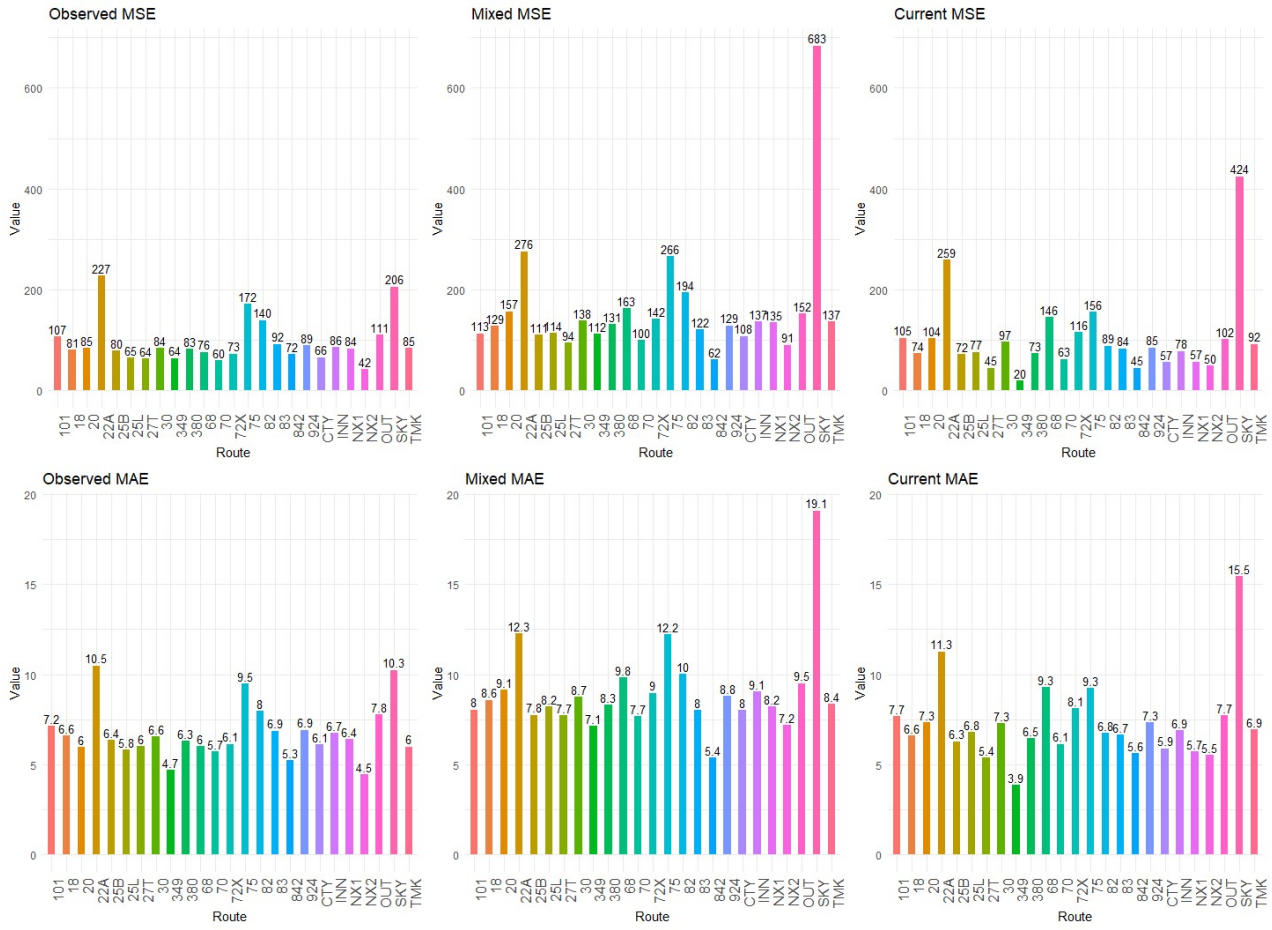


Figure 18: MSE and MAE for Estimators Aggregated at Route Level

Following the analysis by route level, we look at how these estimators change over different intervals in the day, as seen in Figure 19 below. Overall we see that the Morning time interval is the best performer with lower values for all the estimators in both MSE and MAE. This is interesting to see as earlier in our analysis of correlation we saw that the clustering phenomena of frequent uneven headways was not observed in the morning interval. It is also important to note that the observed estimator is the best performing in five of the six tests shown, this is in contrary to it not having the highest correlation above.

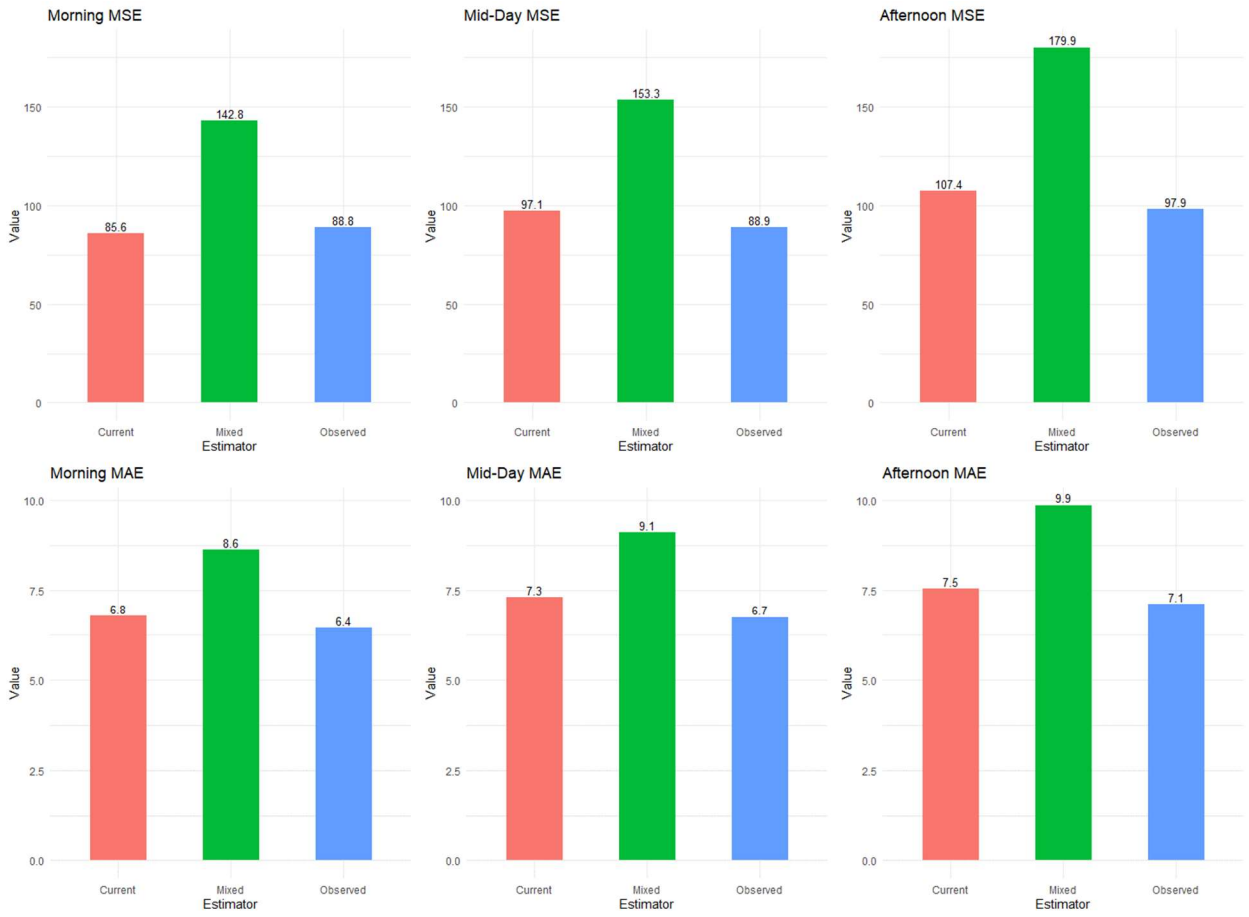


Figure 19: MSE and MAE of Estimators at Different Time Intervals

Due to results of low correlation of all the estimators, future real-time visualizations from this work will be based on the observed estimator. This is because the Observed Estimator is capturing what the actual headway was just one time or bus interval prior what it currently is now.

4.2 Interesting Event

In order to highlight different visualizations that were created, I show how the AWT and Lateness change throughout the day given an unusual event in the city. During the end of my dissertation, there was a large fire, lasting for roughly two days (22-10-2019 to 24-10-2019), in the Central Business District which caused a strain on PT in the CBD area. Below I compare the emergency events AWT and Lateness visualization of all estimators compared to data a week prior. I then show how the observed headway altered for these routes for days of the emergency compared to a week prior. In order to clear the noise of other routes that was not affected by the event, I only focus on the routes that service near the fire, which included City Link, Inner Link, Outer Link, 75, and 18.

We observe from the figures 20 and 21 which look at the Actual Average Waiting Time of the routes, that there was a clear spike in the data, which started only slightly delayed around 4:00 PM

when the fire started approximately at 2:00 PM. This delay in the metrics may have been due to the amount of buses active during this time which masked the large headway for buses stuck near the fire

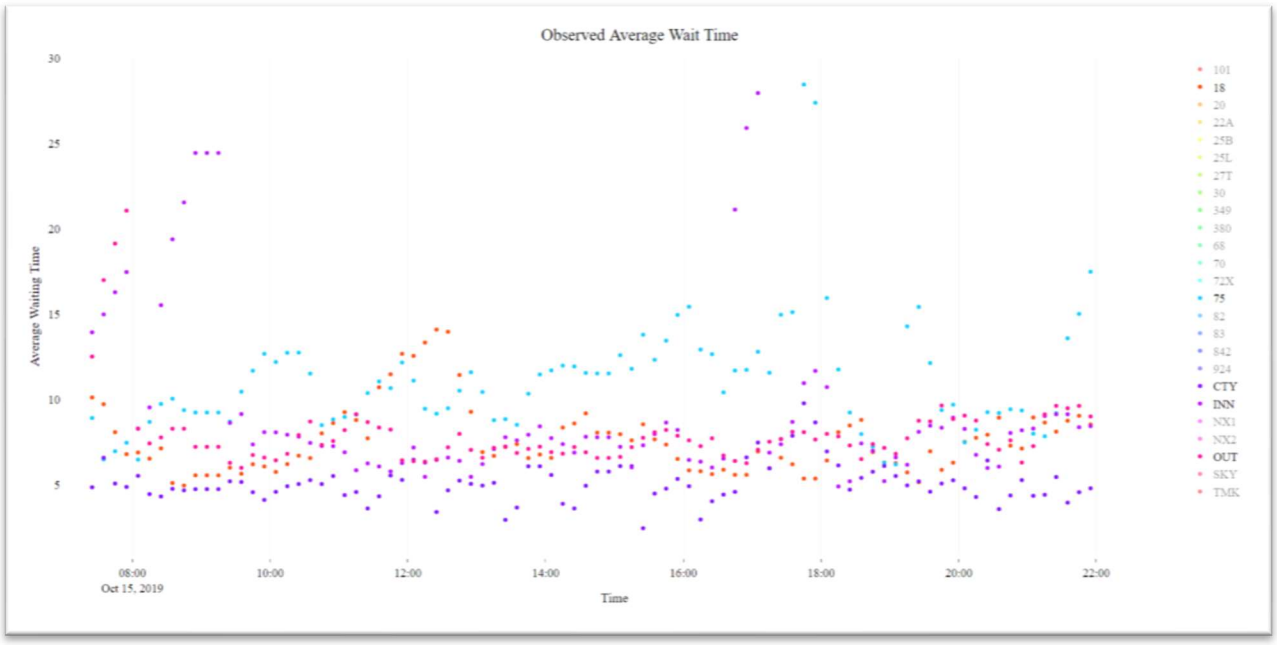


Figure 20: Actual Average Waiting Time Tuesday Prior

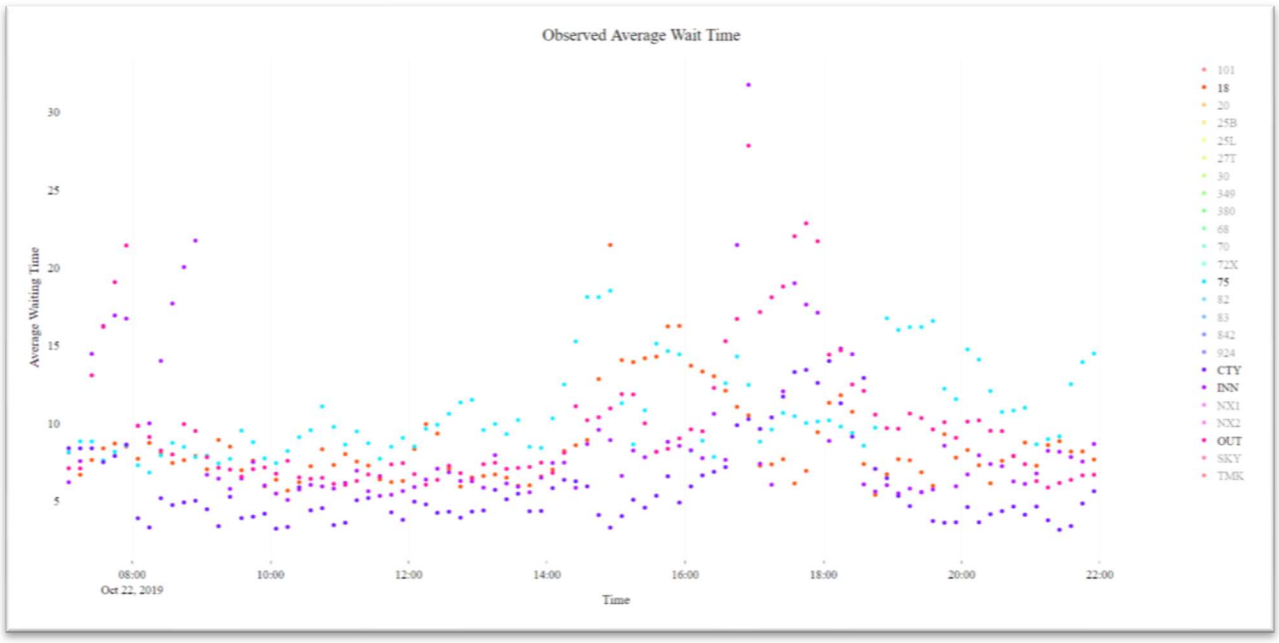


Figure 21: Actual Average Waiting Time Day of Fire (Tuesday)

In Figures 22 and 23, we look at the Observed Estimators Lateness for the Tuesday prior to the fire and the day of the emergency. We see a similar trend to that seen in the Actual Average Waiting

Time as the input to these two visualizations are similar with the former taking away scheduled headway from the AWT.

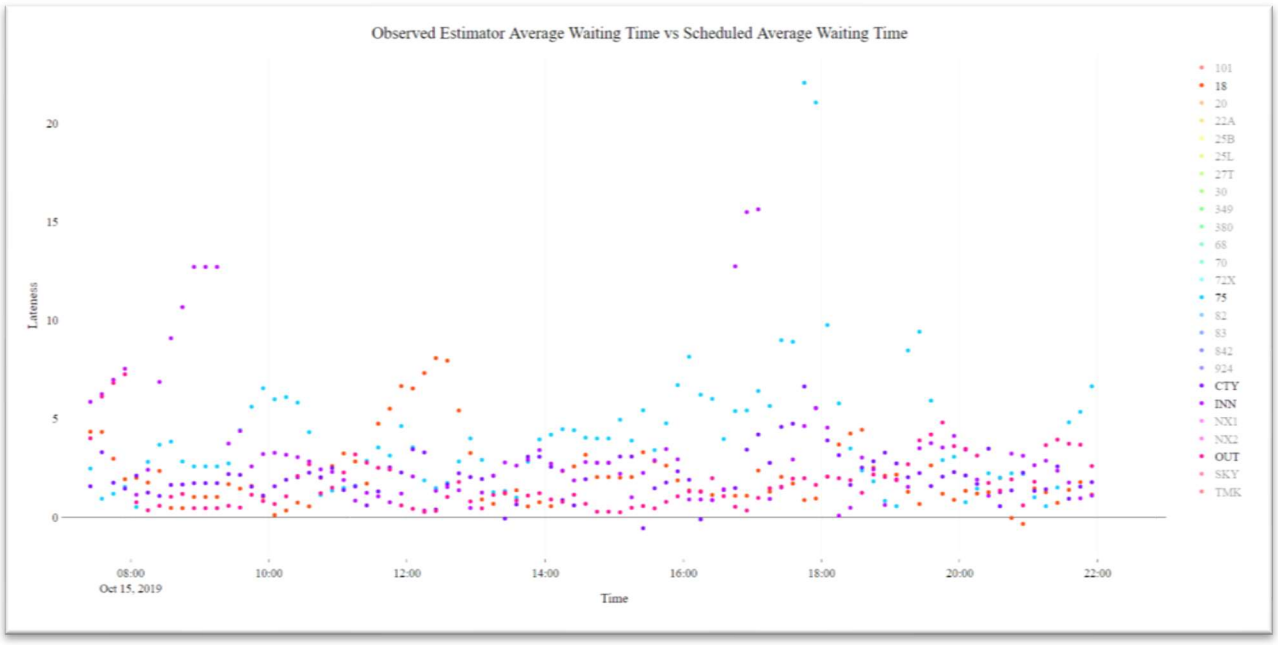


Figure 22: Observed Estimator Lateness Tuesday Prior

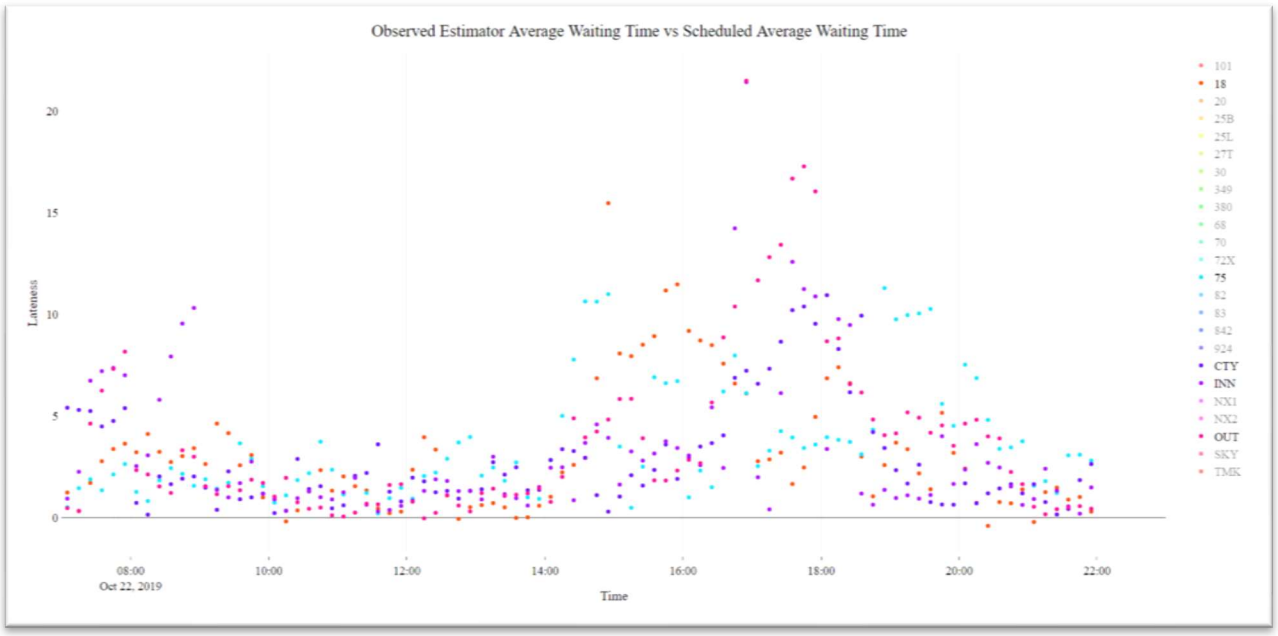


Figure 23: Observed Estimator Lateness Day of Fire (Tuesday)

In Figures 24 and 25 below, we look at the Current Estimators Lateness for the Tuesday prior to the fire and the day of the event. We see only slight increases in the Lateness of the routes, specifically in the Outer link. Since the Current Estimator calculates the Headway based on the difference in

estimated time of only live buses, which results in less predictions. That being, that the Current Estimator would not pick up that there should be longer traveling time between buses, only that they are spreading further apart. We see that it is much less affected as the Observed Estimator with a spike not being as clear.

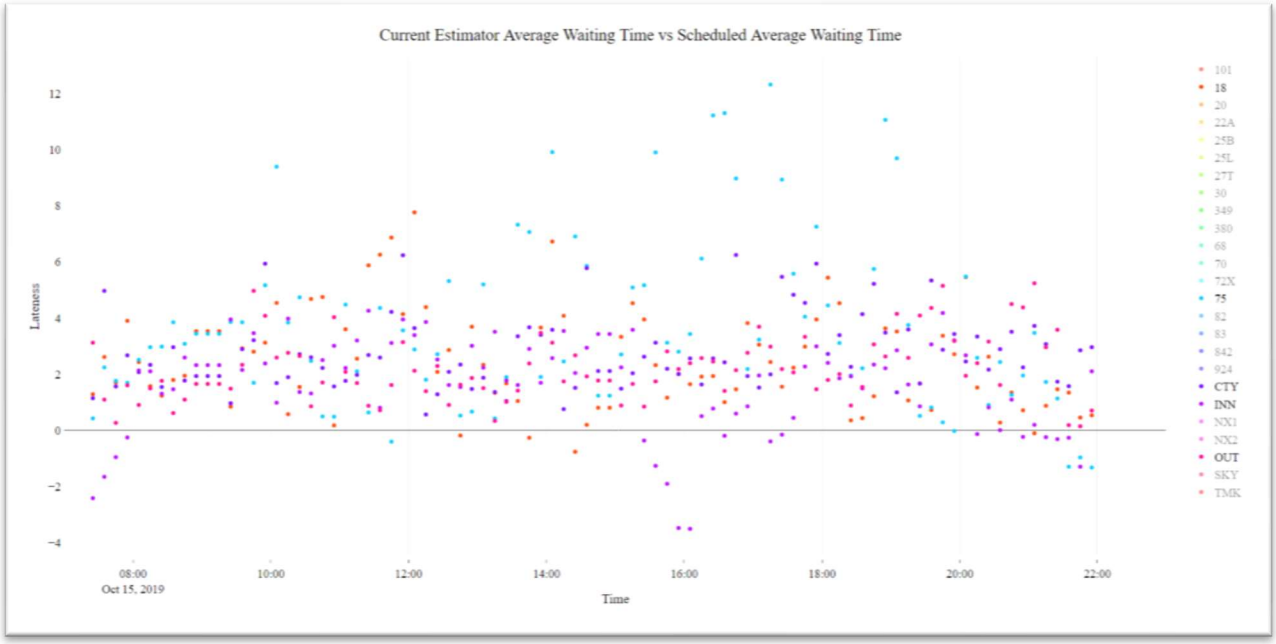


Figure 24: Current Estimators Lateness Tuesday Prior

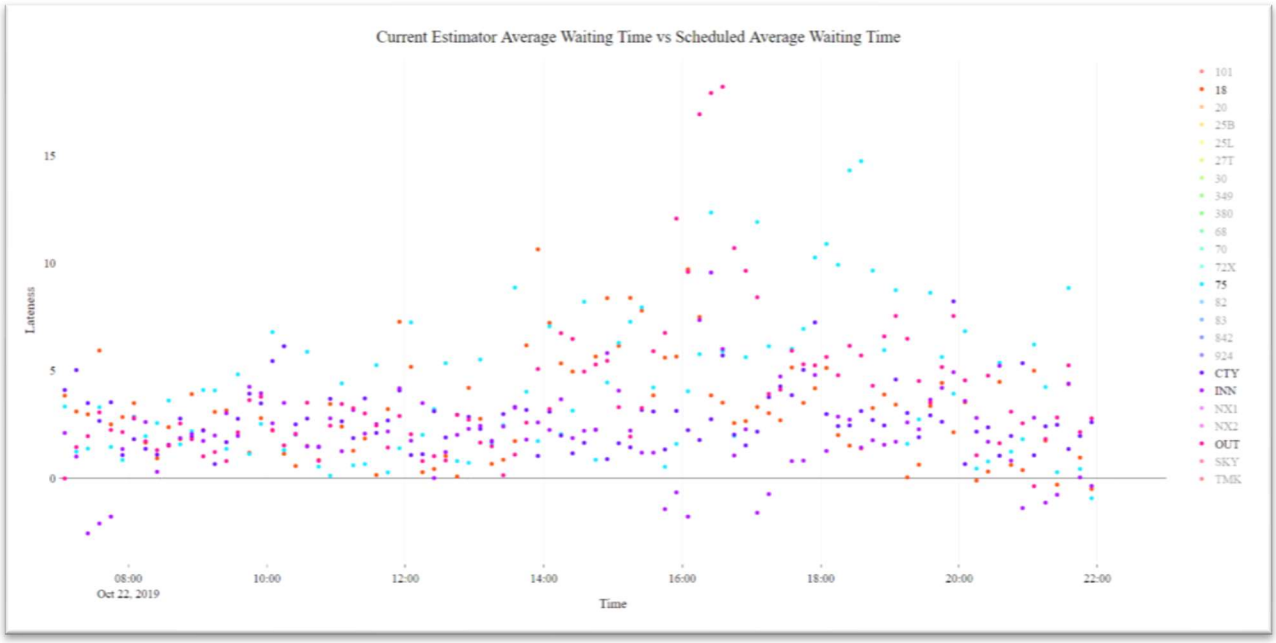


Figure 25: Current Estimators Lateness Day of Fire (Tuesday)

In Figures 26 and 27 below, we look at the Mixed Estimators Lateness for the Tuesday prior to the fire and the day of the event. At around 2:00 PM, we do see an increase of Lateness and variation that seems to die off around the same time the other estimators do. There is an increase in the AWT during the time of the fire, but it is not clear that this is the cause of the variation. It seems that there is variation throughout the day, which may be caused by other non-related incidents or faulty data.

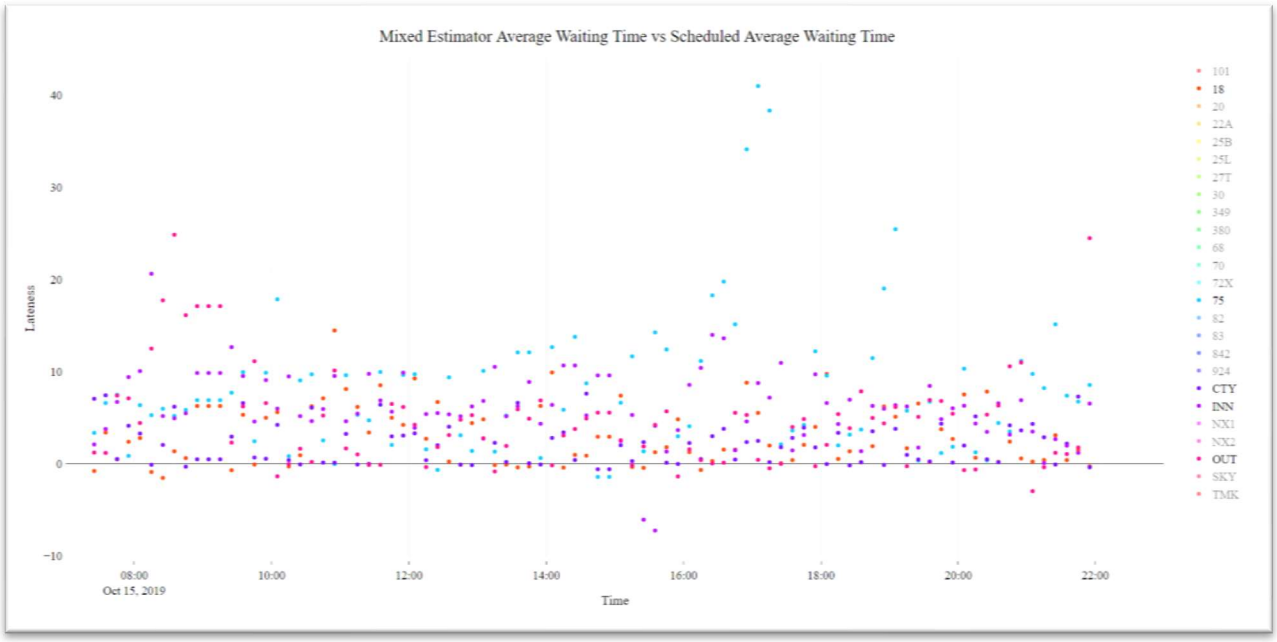


Figure 26: Mixed Estimator Lateness Tuesday Prior

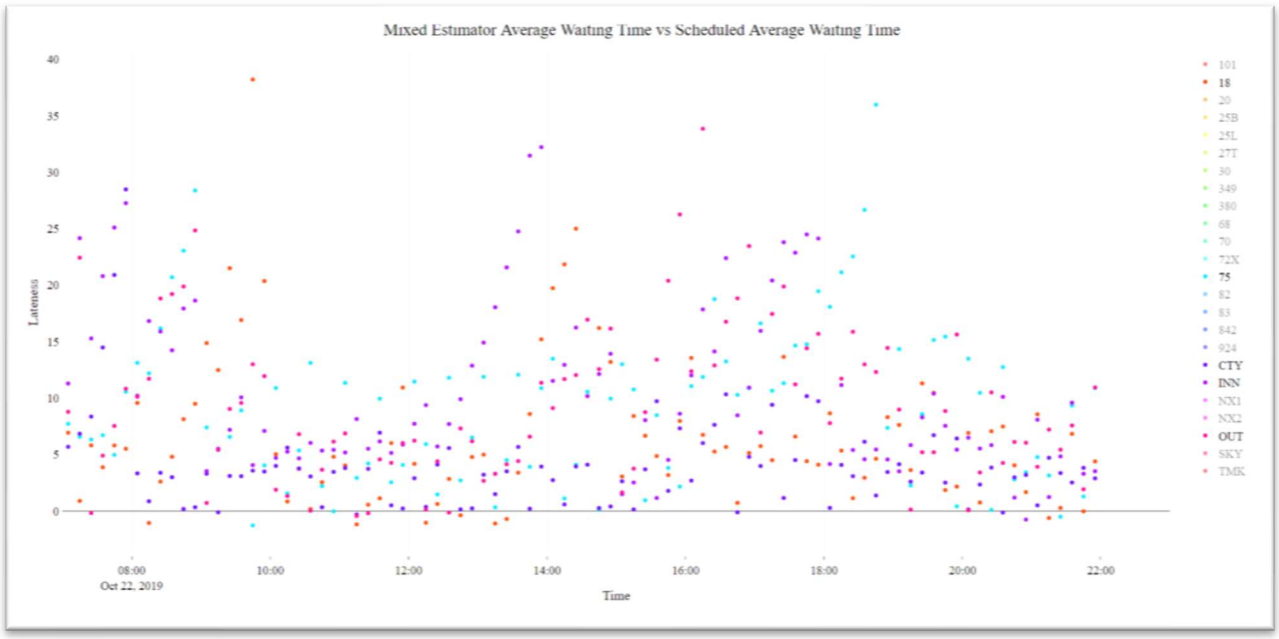


Figure 27: Mixed Estimator Lateness Day of Fire (Tuesday)

Since AWT is a calculation that incorporates many predictions of headway into consideration, it takes many observations to be delayed for the metric to be moved drastically. In our example, most of the buses in the area were most likely gridlocked in traffic, when entering and exiting areas near the fire. We do not see as clear trend in the Mixed and Current Estimators potentially because both are calculated based off of the scheduled travel time from previous bus to current bus stop to measure headway.

To complement this analysis of AWT and Lateness, we will look solely at the time difference of the Actual Headway for the two days of the fire and the week prior for the same routes. We have reduced the time scale of the plots for the day the fire started and the day a week before the fire started to only include times between 13:00 and 22:00 to better illustrate the delay. Alternatively, the plots for the day after the fire started and the correlating date a week prior will show the full days' worth of data as there is not a clear effect on the headway.

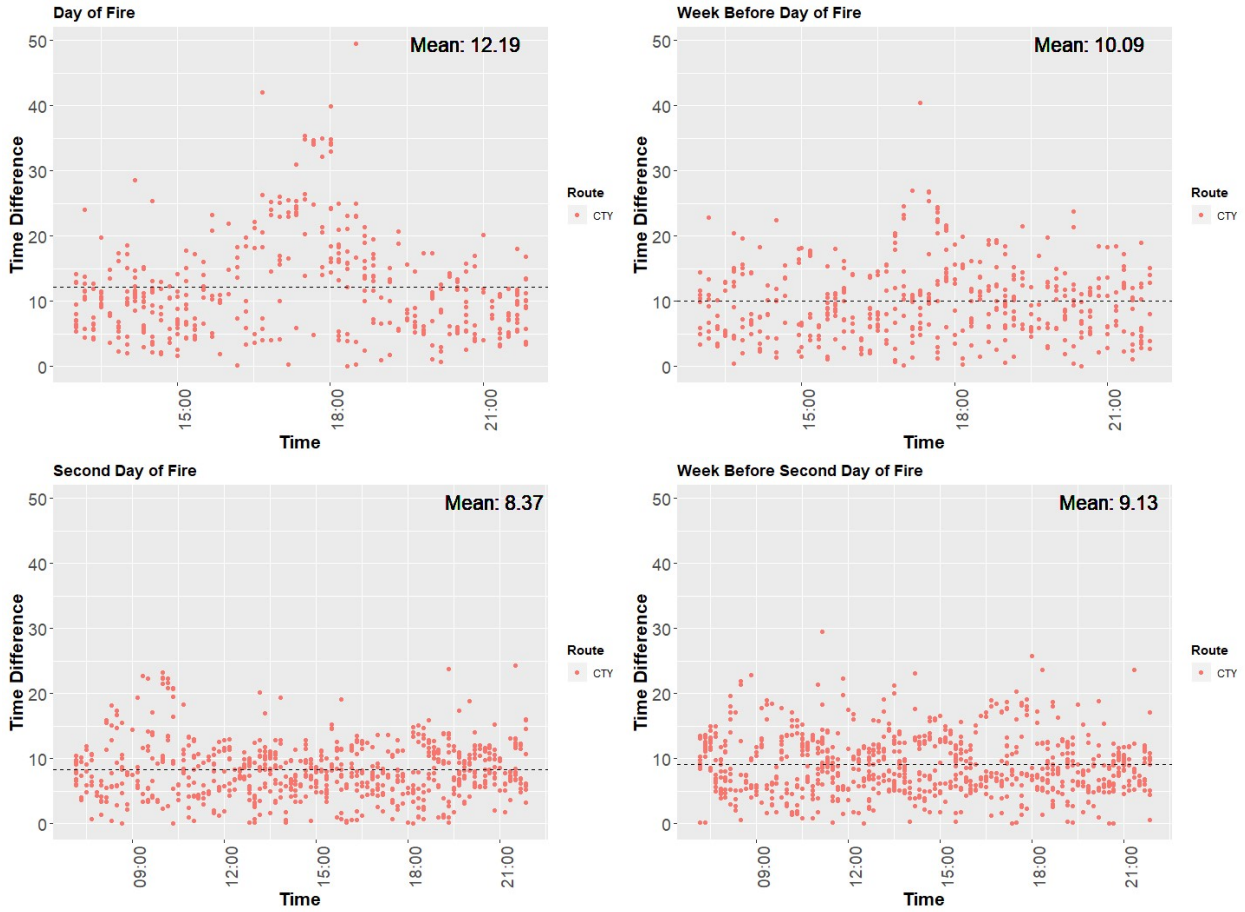


Figure 28: City Link Fire Analysis

Looking above in Figure 28, we see the data only for the City Link route, which runs in the heart of the CBD. On the day of the fire, there is a clear delay that started around 15:00 and lasted until approximately 19:00. We see that the overall mean of the time frame in question had risen by approximately 2 minutes. Although, there was little to no effect on the second day of the fire compared to the weeks prior data.

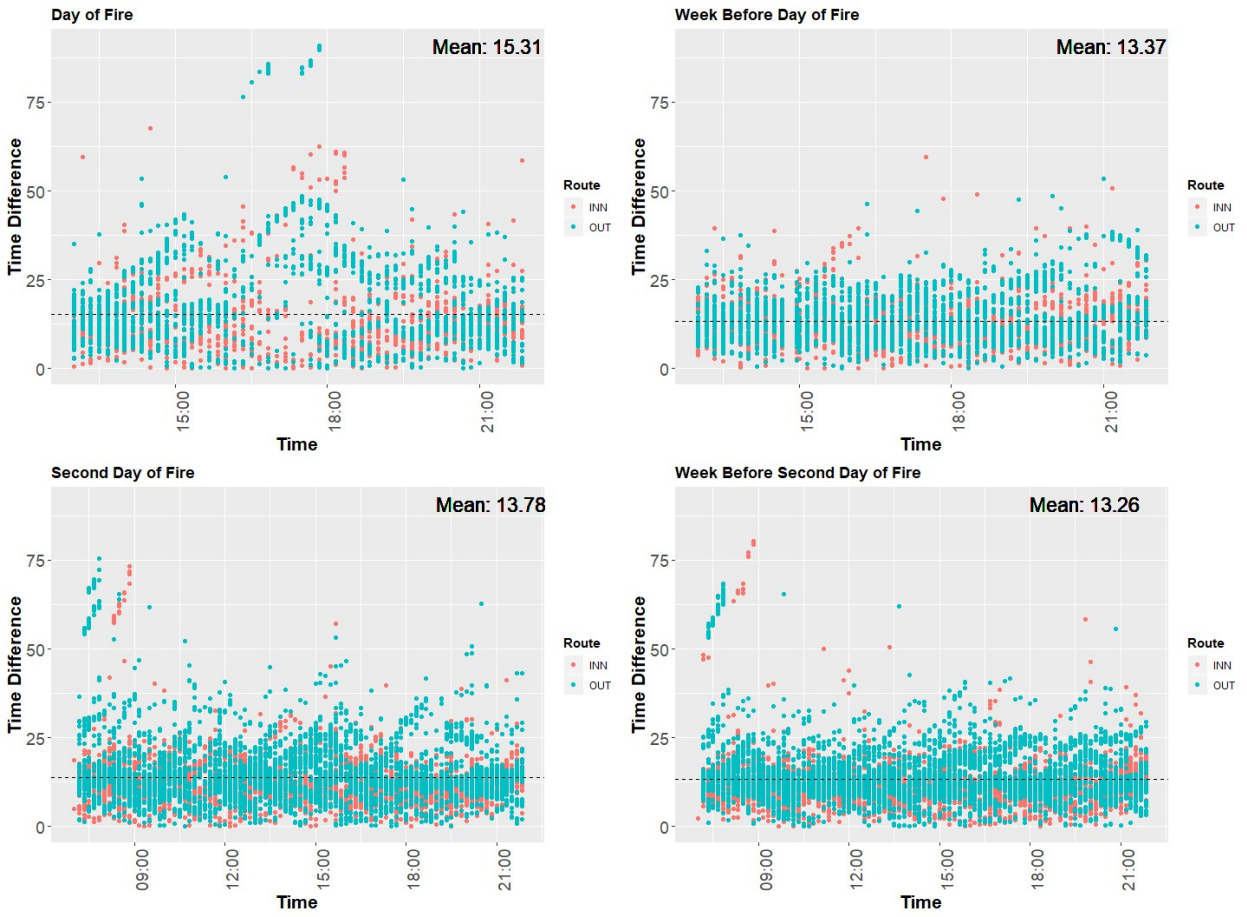


Figure 29: Inner Link and Outer Link Fire Analysis

Next we look at the data for the Inner and Outer Link above in Figure 29. There is a similar pattern for the two bus routes as seen for the City Link, we see that there is a clear spike in the headway time the day of the fire compared to the week prior, starting slightly later than the City Link at around 16:00 and lasting until approximately 18:30. The day following the fire, we do see a spike in the morning buses but believe this is due to normal fluctuations in the data, potentially from when the application starts gathering data. We deduce that the severity and duration of the spike was not as severe as the City Link as these two routes service a much larger space of the city.

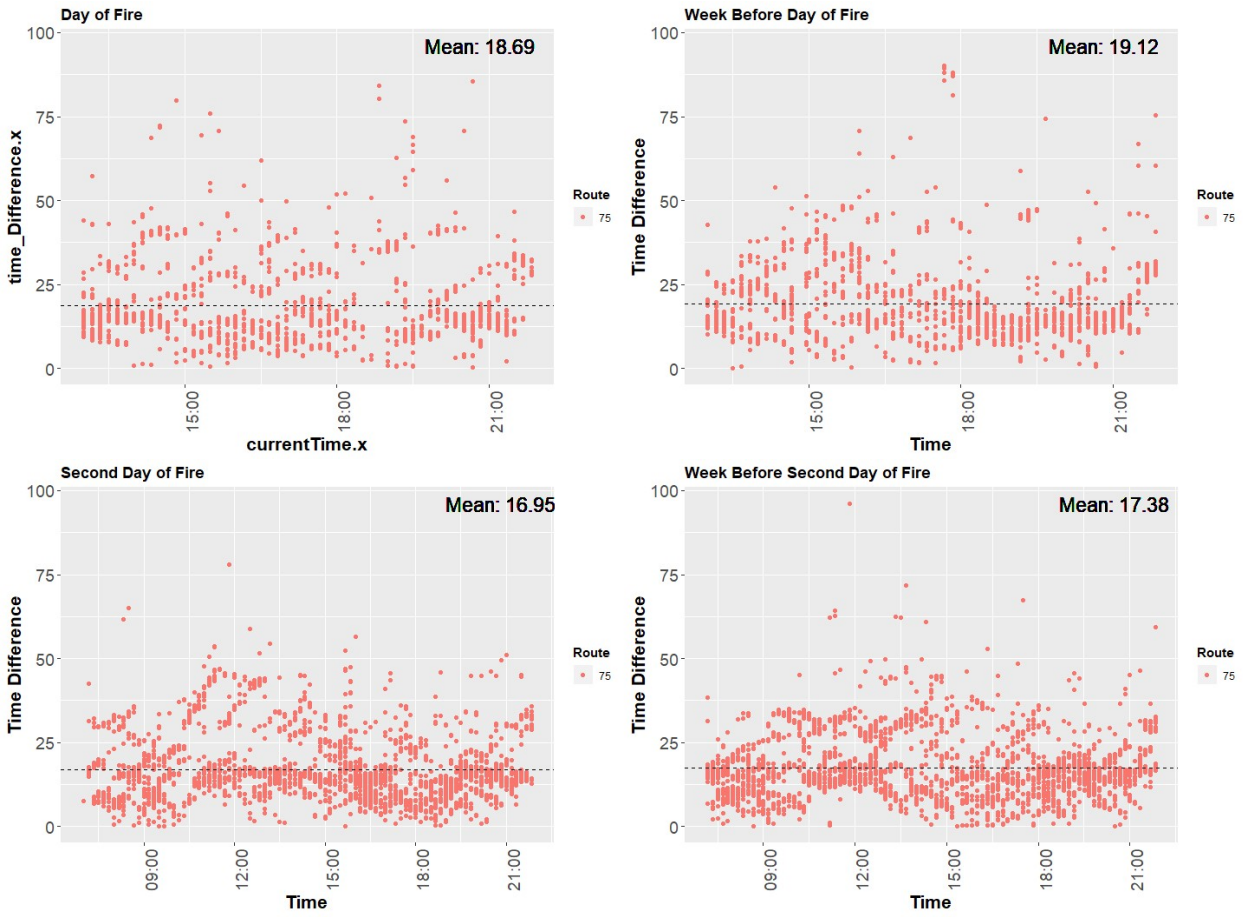


Figure 30: 75 Fire Analysis

When analyzing the data for the 75 bus route, seen above in Figure 30, it is hard to tell if there was an effect from the fire, as the route is naturally high in variability. This is most likely due to only a short segment of the bus route traveling near the location of the fire, and most likely was only slightly delayed, if at all.

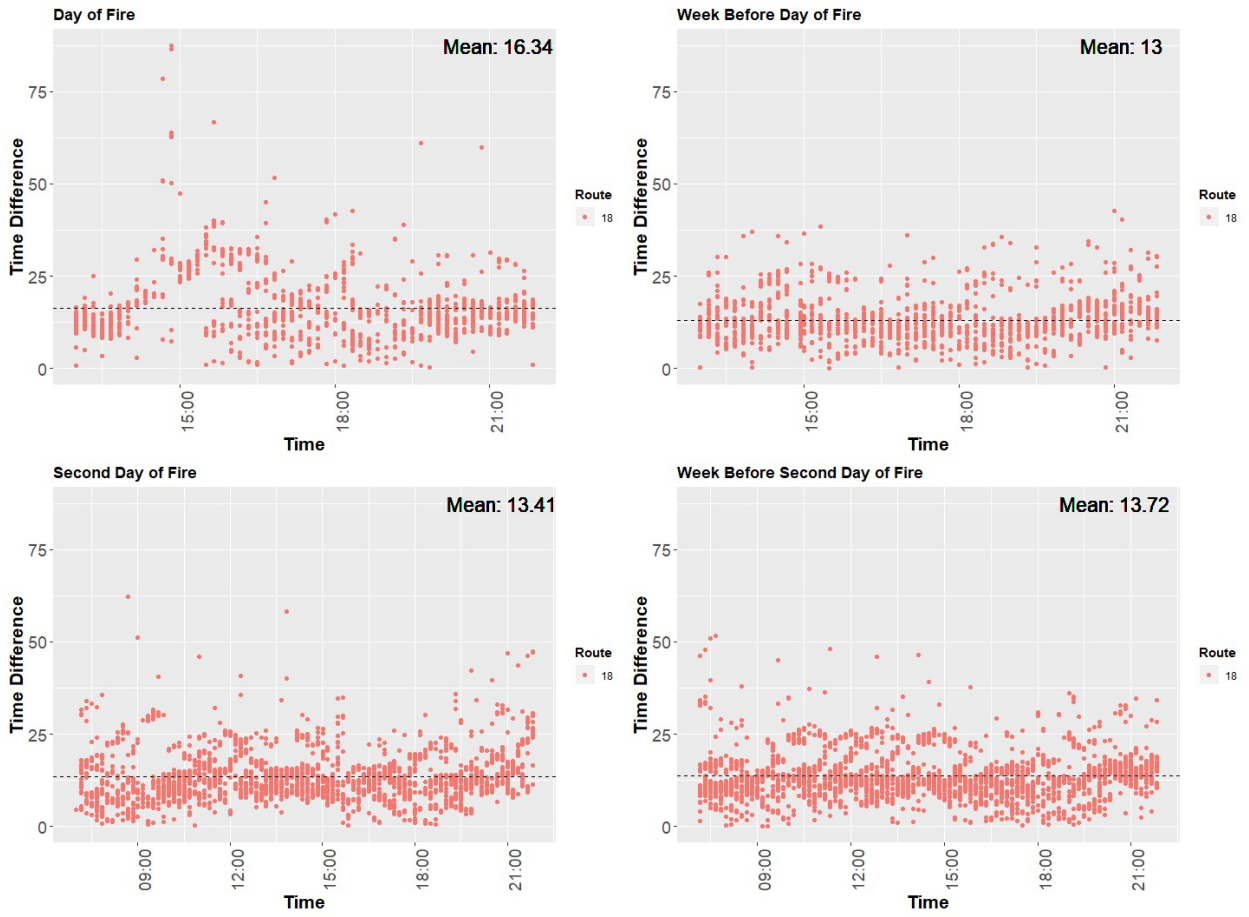


Figure 31: 18 Fire Analysis

Our final route that travels through the CBD area near the fire is the 18 bus route with its analysis shown above in Figure 31. We see that there was a potential impact due to the fire, seen earlier than the City, Inner, and Outer Link, with delays starting approximately at 14:00 and ending at approximately 17:00. Similar to the other routes analyzed, there was no clear evidence of a delay on the second day of the fire.

In summary, amongst the five routes that were affected by the fire in the city center, it is only evident that the City Link, Inner Link, Outer Link, and 18's headways were delayed. We see that the means of the headways for City Link increased to 12.19 minutes from 10.09 minutes, Inner Link and Outer Link combined increased to 15.31 minutes from 13.37 minutes, and the 18 increased to 16.34 minutes from 13 minutes on the day of the fire and week before the fire respectively. We see little to no effect on the second day of the fire on any of the routes as the city responded and adapted accordingly to the event.

5

Conclusion

5.1 Future Work

There are many areas that can be improved and researched in depth when it comes to the Headway and AWT in Auckland. We hope that this project and application can serve as a tool that enables easier monitoring and evaluation of AT data. Any research that intends to test how their research affects headway or estimates headway could use this research and application as a tool to cut out a lot of the coding overhead required to work with this Auckland's PT data. That being said, Google Transit provides a schema that is used across many different PT in the world, which means that the code and findings within this research could be applied to many other cities.

As far as the research directly done in relation to this project, I would like to fix the assumptions that are stated above, many of which could be coded for. There were far too many weird situations in the data that could arise to code for all of them within the time permitted for this project. Along with coding catches for weird situations, verifying the data quality that is being collected from AT is of high interest, as concerns have crippled progress in this project. Ideally partnering with employees of AT to better understand the underlying data and why it acts the way it does, would help in the research to provide better and cleaner output of the ingested data. Finally, I would like to analyze situations where the headway was not as expected and dig deeper to why this caused issues in providing accurate estimation.

Finding what determines headway and producing an estimator that could predict with high accuracy would be of interest, as it was a side goal of this current project. This could be done by capturing other data sources to derive useful features to predicting headway, including weather or traffic data.

Ultimately, I desire to find a way to use this information to provide a better PT experience for users that use these frequent buses. Whether that be by setting their expectations of waiting time at their stop through their AT application, or helping AT provide more consistently even headways.

5.2 Conclusion

In this research, we capture and visualize the current state of headway of frequent routes in Auckland's Public Transportation. During this process, we analyze three estimators, Observed, Mixed, and Current, to determine which of the three is best at predicting what the next headway will be, only using the previous headway as input.

This research finds that the three estimators tested were not good predictors of the next bus headway at the stop level. We find that, after checking for similar distributions, each of the estimators could be used as estimators for the current state of bus transportation in the city. Of the three estimators, Current has the highest correlation over the full data, along with most of the subsequent tests; while the Observed Estimator has the lowest MSE and MAE for a majority of the tests.

One interesting event happened during the year duration of the project which was a large fire that heavily constrained the transportation within the Central business District. The AWT and Lateness for all Estimators was captured and analyzed, showing that the Observed Estimators reflected the fluctuations most clearly, which was likely due to the other estimators using estimated travel times as their input. Analyzing the headway for the five routes that travel through the area, we see that four of the five were delayed by the event.

Our results show over a weeks' worth of data that the highest correlation seen is .2761, which for a predictor is quite low. None of the three estimators should be used as predictors of future data without further research and alterations to the estimators. Potential data quality issues of AT data stream or code used in this project may be the cause of the low ability to predict headway, but further research is required to verify.

The output of this research project will still be utilized as a visualization and exploratory tool for public transport customers and researchers alike. It will complement other work already been done at University of Auckland as referenced in [36][38].

Future use of the application and visualizations of Auckland's headway will be displayed using the Observed estimator, as it represents that actual headway, just slightly delayed. More information on these real-time visualizations and source code can be found at: https://github.com/ColbyCarrillo/BusHeadway.

References

[1] "Transit | Google Developers," Google. [Online]. Available: https://developers.google.com/transit/. [Accessed: 02-Sep-2019].

[2] "Revision History | Static Transit | Google Developers," *Google*. [Online]. Available: https://developers.google.com/transit/gtfs/guides/revision-history. [Accessed: 02-Sep-2019].

[3] "Revision history | Realtime Transit | Google Developers," *Google*. [Online]. Available: https://developers.google.com/transit/gtfs-realtime/guides/revision-history. [Accessed: 02-Sep-2019].

[4] "About Google Transit - Transit Partners Help," *Google*. [Online]. Available: https://support.google.com/transitpartners/answer/1111471. [Accessed: 02-Sep-2019].

[5] "Get started with Google Transit - Transit Partners Help," *Google*. [Online]. Available: https://support.google.com/transitpartners/answer/1111481?hl=en&ref_topic=3521043. [Accessed: 02-Sep-2019].

[6] "Headway," *Merriam-Webster*. [Online]. Available: https://www.merriam-webster.com/dictionary/headway. [Accessed: 03-Sep-2019].

[7] "Transit – Google Maps," *Google*. [Online]. Available: https://maps.google.com/landing/transit/cities/. [Accessed: 02-Sep-2019].

[8] "Key Event's In Auckland's Transport History". (2014). *Key Event's In Auckland's Transport History*. Available: https://at.govt.nz/media/400193/Transport-timeline.pdf

[9] "Regional Public Transport Plan (2018-2028)", *Regional Public Transport Plan (2018-2028)* (2018). Available: https://at.govt.nz/media/1979652/rptp-full-doc-final.pdf

[10] P. Furth, N. Wilson, "Setting Frequencies on Bus Routes: Theory and Practice". *Transportation Research Record*. Available: http://onlinepubs.trb.org/Onlinepubs/trr/1981/818/818-001.pdf

[11] B. Grosfeld-Nir, J.H. Bookbinder. "The planning of headways in urban public transit". *Annals of Operations Research*. December 1995. Available: https://doi.org/10.1007/BF02031944

[12] A. Ceder. *Public Transit Planning and Operation: Theory, Modeling and Practice*, [E-Book] Available: http://imentaraddod.com/wp-content/uploads/2017/07/337-Public-Transit-Planning-and-Operation-Theory-Modeling-and-Practice-Avishai-Ceder-075066166.pdf

[13] M. Rahman, L. Kattan, S. C. Wirasinghe, "Trade-offs between headway, fare, and real-time bus information under different weather conditions" *Public Transport*, Vol. 10, Issue 2, pages 217-240, August 2018. Available: https://doi.org/10.1007/s12469-018-0176-4

[14] M. Rahman, S. C. Wirasinghe, L. Kattan "The effect of time interval of bus location data on real-time bus arrival estimations". *Transportmetrica A: Transport Science*, Vol. 12, Issue 8, 2016. Available: https://doi.org/10.1080/23249935.2016.1166159

[15] D. Sun, H. Luo, L. Fu, W. Liu, X. Liao, M. Zhao "Predicting Bus arrival time on the basis of global positioning system data". *Transportation Research Record Journal of the Transportation Research Board*. December 2007. Available: https://www.researchgate.net/publication/245562763_Predicting_Bus_Arrival_Time_on_the_Basis_of_Global_Positioning_System_Data

[16] S.H. Park, Y.J. Jeong, T.J. Kim "Transit travel time forecasts for location-based queries: implementation and evaluation". *Journal East Asia Society Transportation Studies.* Vol. 7, 2007. Available: https://www.jstage.jst.go.jp/article/easts/7/0/7_0_1859/_pdf

[17] K. Dziekan, A. Vermeulen. “Psychological effects of and design preferences for real-time information displays”. *Journal Public Transportation*, February 2006. Available: https://www.researchgate.net/publication/241085370_Psychological_Effects_of_and_Design_Preferences_for_Real-Time_Information_Displays

[18] K.E. Watkins, B. Ferris, A. Borning, G.S. Rutherford, D. Layton “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 Practice.* Vol. 45, Issue 8, pages 839-848, October 2011. Available: https://doi.org/10.1016/j.tra.2011.06.010

[19] M. Rahman, S.C. Wirasinghe, L. Kattan. “Users’ views on current and future real-time bus information systems”. *Journal of Advanced Transportation*. April, 2013. Available: http://dx.doi.org/10.1002/atr.1206

[20] T. Litman. “Valuing transit service quality improvements”. *Journal of Public Transportation.* January 2007. Available: https://www.researchgate.net/publication/252745569_Valuing_Transit_Service_Quality_Improvements

[21] A. Ceder, S. Hassold, B. Dano. “Approaching even-load and even-headway transit timetables using different bus sizes”. *Public Transportation*, Vol. 5, Issue 3, pages 193-217. October 2013, Available: https://doi.org/10.1007/s12469-013-0062-z

[22] E. E. Osuna, G. F. Newell, “Control Strategies for an Idealized Public Transportation System”. *Transportation Science*, February 1972 Available: https://doi.org/10.1287/trsc.6.1.52

[23] X. Luo, X. Zhao, L. Sun, K. Ma and J. Tang, "Dynamic bus dispatching under the environment of Internet of things," *Proceeding of the 11th World Congress on Intelligent Control and Automation*, Shenyang, 2014, pp. 449-454.
Available: 10.1109/WCICA.2014.7052755

[24] Jiamin Zhao, S. Bukkapatnam and M. M. Dessouky, "Distributed architecture for real-time coordination of bus holding in transit networks," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 4, no. 1, pp. 43-51, March 2003.
doi: 10.1109/TITS.2003.809769

[25] W. Fan, R. Machemehl. “Characterizing bus transit passenger waiting times”. *Proceedings, Annual Conference - Canadian Society for Civil Engineering*. January 2002. Available: https://www.researchgate.net/publication/266573131_Characterizing_bus_transit_passenger_waiting_times

[26] B. McLeod, “Estimating bus passenger waiting times from incomplete bus arrivals data”. *Journal of the Operational Research Society.* Vol. 58, Issue 11, December 2007. Available: https://doi.org/10.1057/palgrave.jors.2602298

[27] R. Balcombe, R. Mackett, N. Paulley, J. Preston, J. Shires, H. Titheridge, M. Wardman, P. White. “The Demand for Public Transport: A Practical Guide”. *TRL Report*, TRL593, TRL Limited. Available: https://trl.co.uk/sites/default/files/TRL593%20-%20The%20Demand%20for%20Public%20Transport.pdf

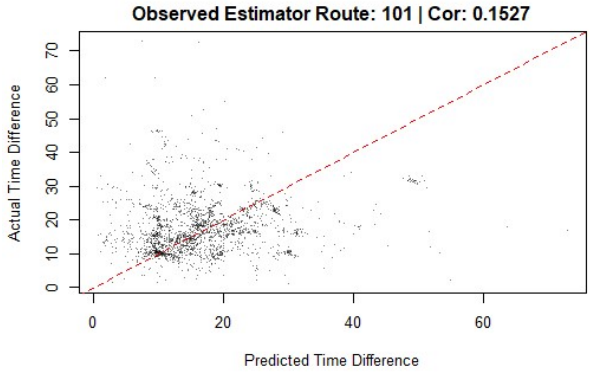
[28] G. Drakos, “How to select the Right Evaluation Metric for Machine Learning Models: Part 1 Regression Metrics,” *Medium*, 05-Dec-2018. [Online]. Available: https://towardsdatascience.com/how-to-select-the-right-evaluation-metric-for-machine-learning-models-part-1-regrression-metrics-3606e25beae0. [Accessed: 15-Sep-2019].

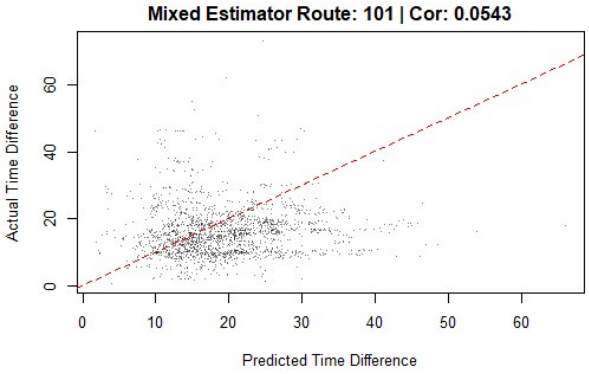
[29] R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
[30] Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2019). dplyr: A Grammar of Data Manipulation. R package version 0.8.3. https://CRAN.R-project.org/package=dplyr
[31] Hadley Wickham (2019). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0. https://CRAN.R-project.org/package=stringr
[32] Kirill Müller, Hadley Wickham, David A. James and Seth Falcon (2019). RSQLite: 'SQLite' Interface for R. R package version 2.1.2. https://CRAN.R-project.org/package=RSQLite
[33] D. Richard Hipp, D. Kennedy, J. Mistachkin, *SQLite* (Version 3.29.0) [Computer Software]. SQLite Development Team. Retrieved [Feb 11. 2019]. Available from https://www.sqlite.org/download.html.
[34] Carson Sievert (2018) plotly for R. https://plotly-r.com
[35] Joe Cheng, Bhaskar Karambelkar and Yihui Xie (2018). leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library. R package version 2.0.2. https://CRAN.R-project.org/package=leaflet
[36] T. Lumley, "tūreiti" twitter.com. [Online]. Available: https://twitter.com/tuureiti?lang=en [Accessed 20 Oct. 2019]
[37] T. Lumley. Weekly Meetings, Topic: "Bus Headway Dissertation Project". School of Statistics, University of Auckland, Auckland, New Zealand. 11 Feb. 2019 Through 22 Oct. 2019.
[38] T Elliott, Personal Communication and Auckland Bus On-Time Application
[39] Python Software Foundation. Python Language Reference, version 3.7. Available at http://www.python.org

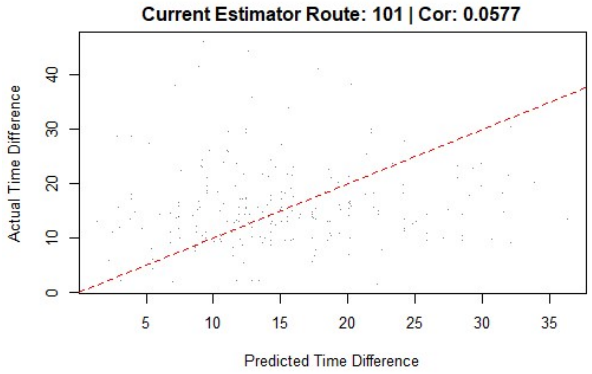
Appendix A

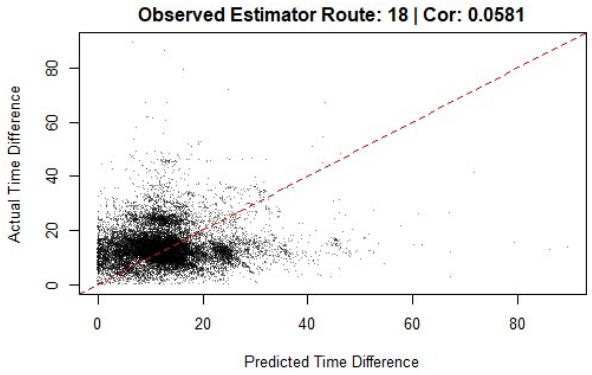
A.1 Chapter 4:

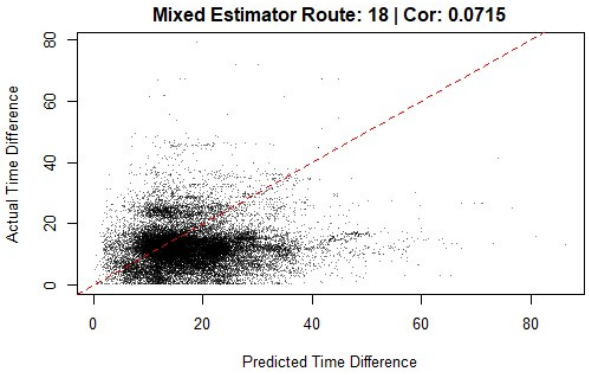
Correlation by Route Level:

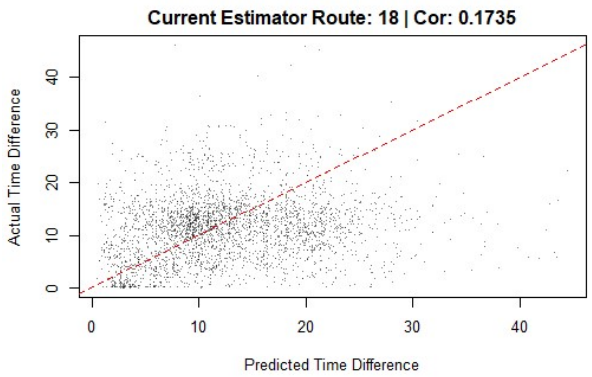


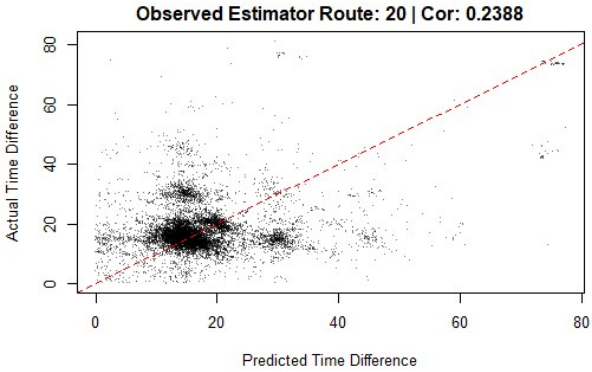


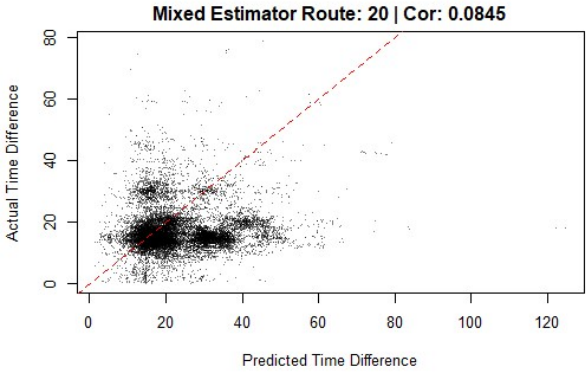


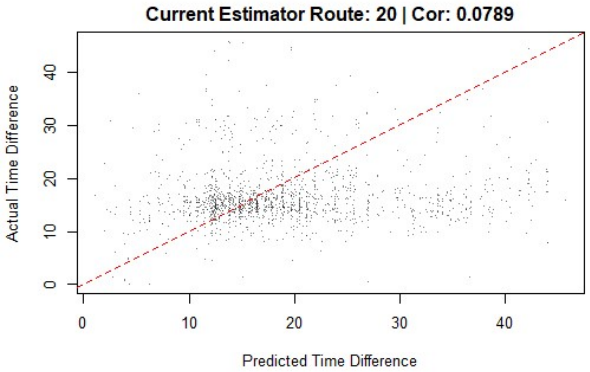


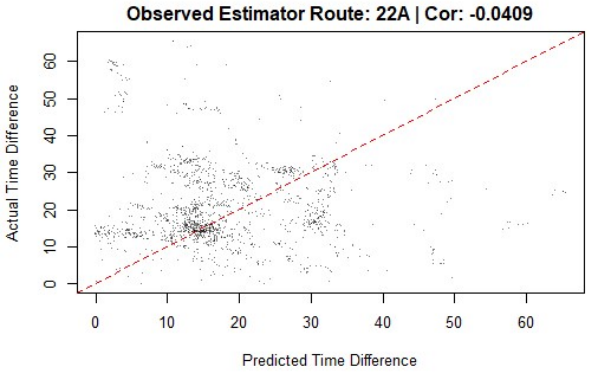


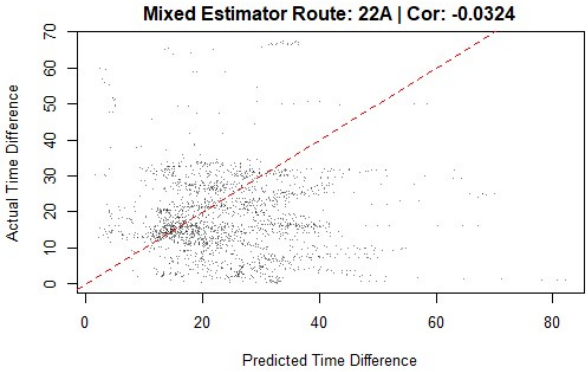


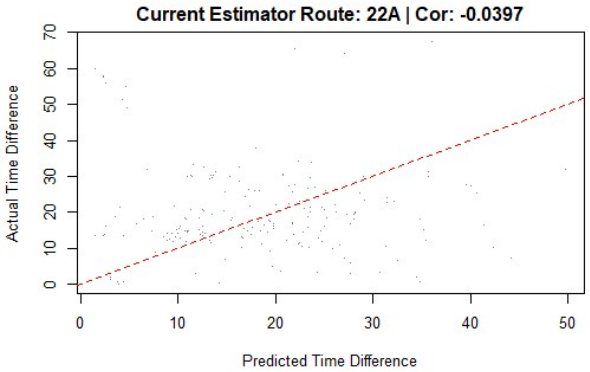

Observed Estimator Route: 20 | Cor: 0.2388
Actual Time Difference
Predicted Time Difference
0
20
40
60
80

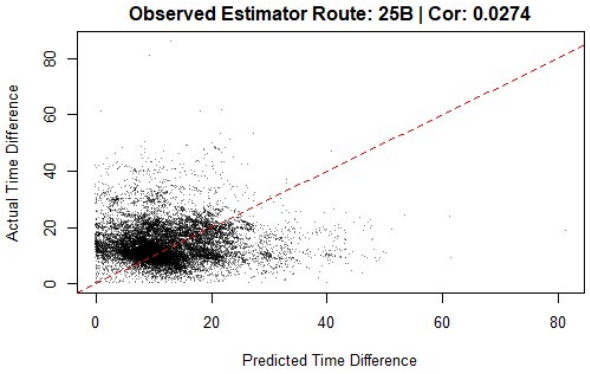

Mixed Estimator Route: 20 | Cor: 0.0845
Actual Time Difference
Predicted Time Difference
0
20
40
60
80
100
120

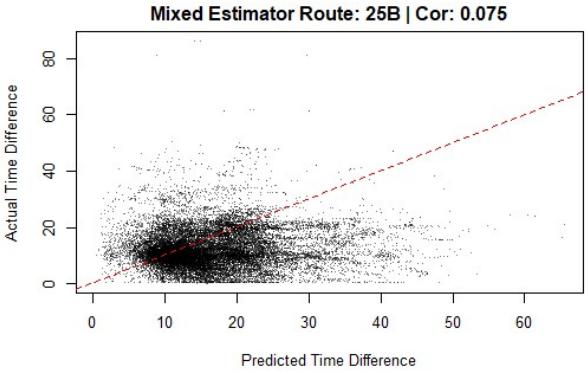

Current Estimator Route: 20 | Cor: 0.0789
Actual Time Difference
Predicted Time Difference
0
10
20
30
40

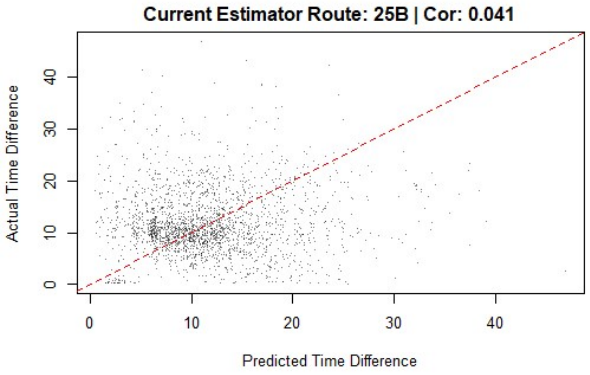

Observed Estimator Route: 22A | Cor: -0.0409
Actual Time Difference
Predicted Time Difference
0
10
20
30
40
50
60

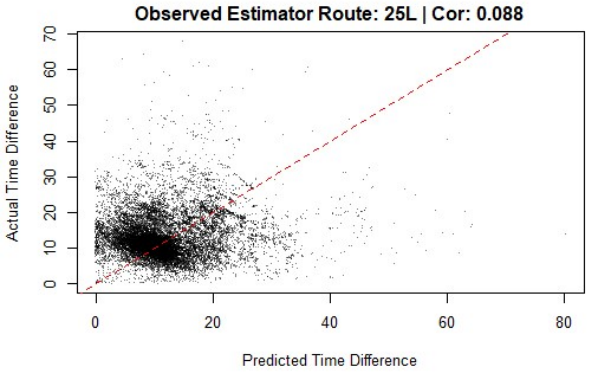

Mixed Estimator Route: 22A | Cor: -0.0324
Actual Time Difference
Predicted Time Difference
0
10
20
30
40
50
60
70

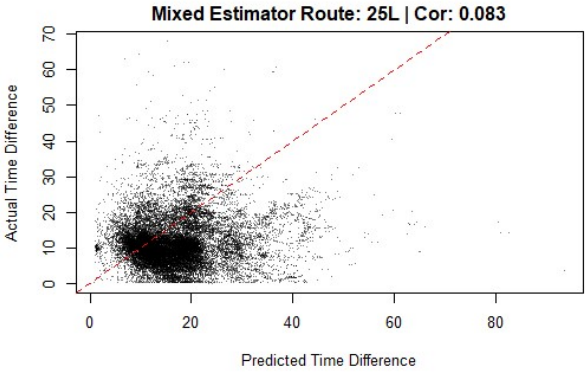

Current Estimator Route: 22A | Cor: -0.0397
Actual Time Difference
Predicted Time Difference
0
10
20
30
40
50
60
70

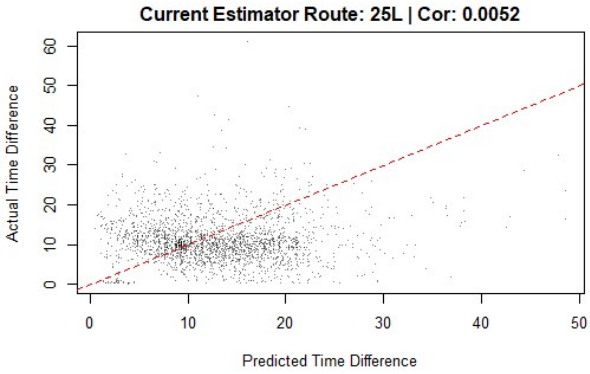

Observed Estimator Route: 25B | Cor: 0.0274
Actual Time Difference
Predicted Time Difference
0
20
40
60
80

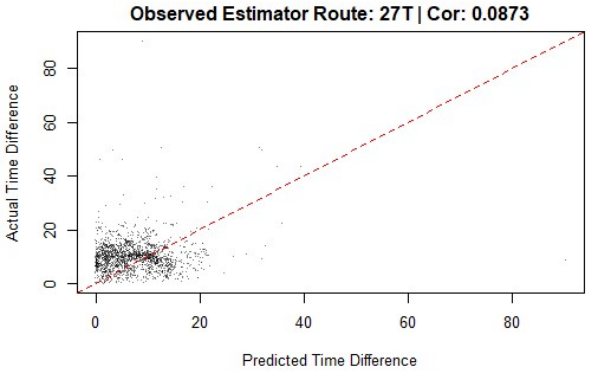

Mixed Estimator Route: 25B | Cor: 0.075
Actual Time Difference
Predicted Time Difference
0
10
20
30
40
50
60

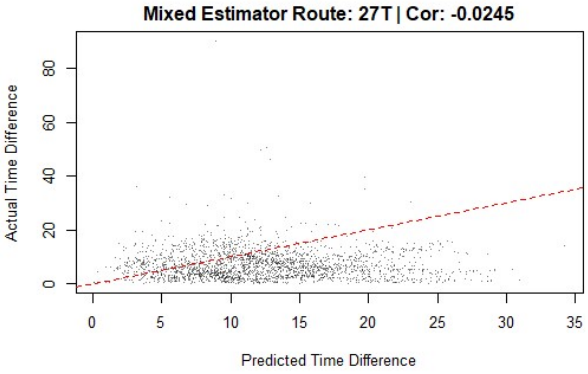

Current Estimator Route: 25B | Cor: 0.041
Actual Time Difference
Predicted Time Difference
0
10
20
30
40

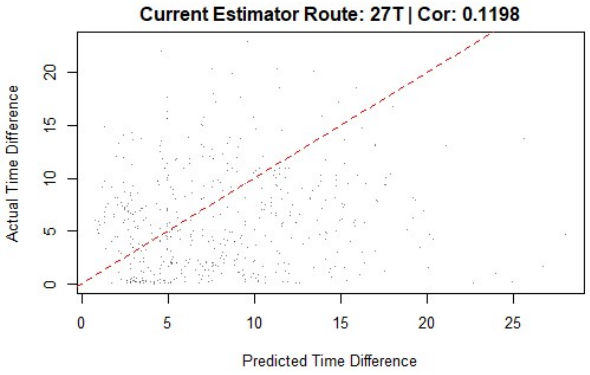

Observed Estimator Route: 25L | Cor: 0.088
Actual Time Difference
Predicted Time Difference
0
20
40
60
80

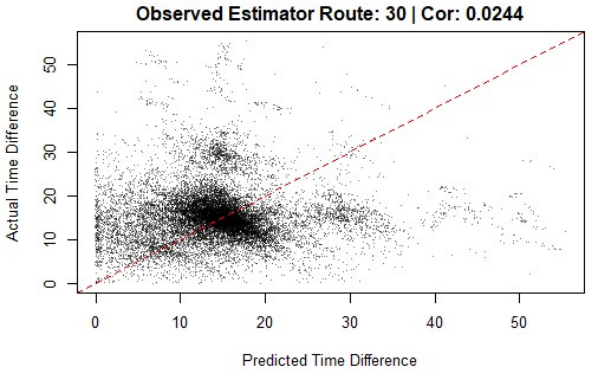

Mixed Estimator Route: 25L | Cor: 0.083
Actual Time Difference
Predicted Time Difference
0
20
40
60
80

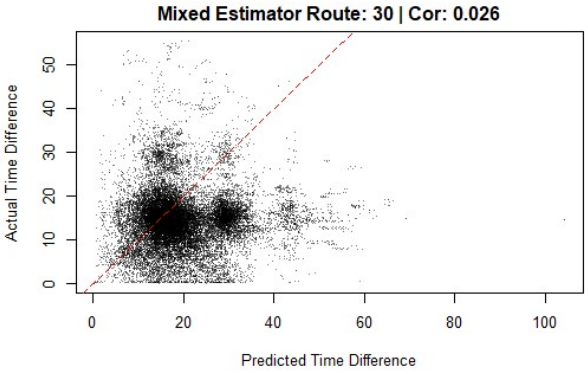

Current Estimator Route: 25L | Cor: 0.0052
Actual Time Difference
Predicted Time Difference
0
10
20
30
40
50

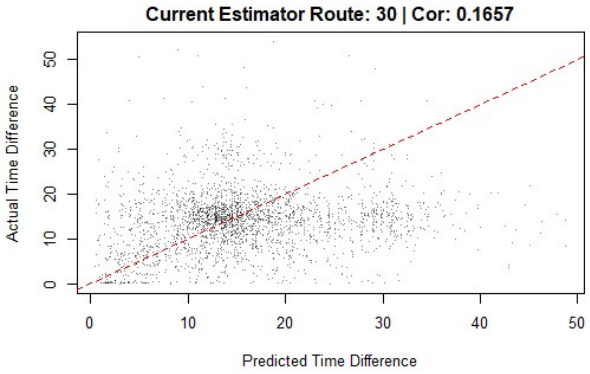

Observed Estimator Route: 27T | Cor: 0.0873
Actual Time Difference
Predicted Time Difference

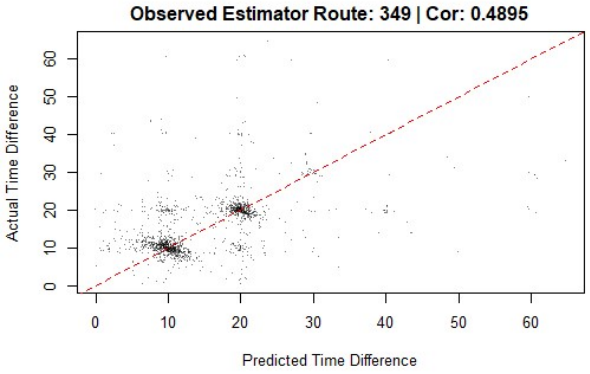

Mixed Estimator Route: 27T | Cor: -0.0245
Actual Time Difference
Predicted Time Difference

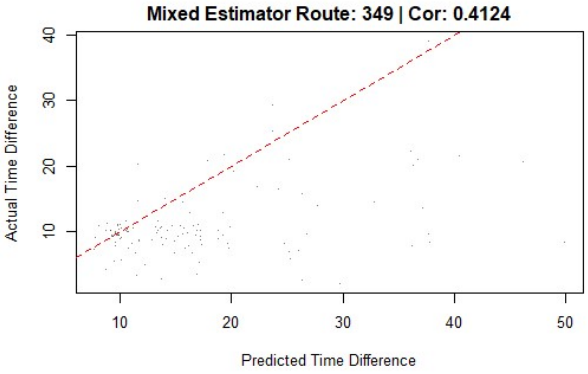

Current Estimator Route: 27T | Cor: 0.1198
Actual Time Difference
Predicted Time Difference

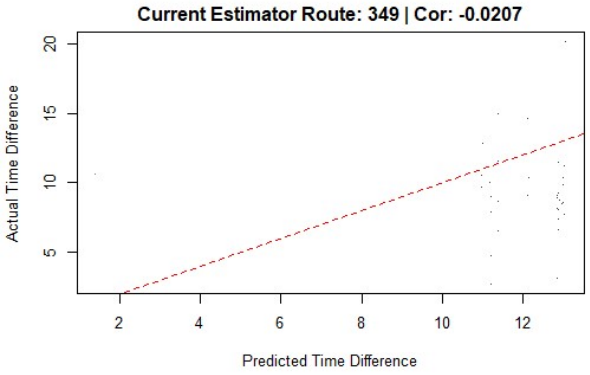

Observed Estimator Route: 30 | Cor: 0.0244
Actual Time Difference
Predicted Time Difference

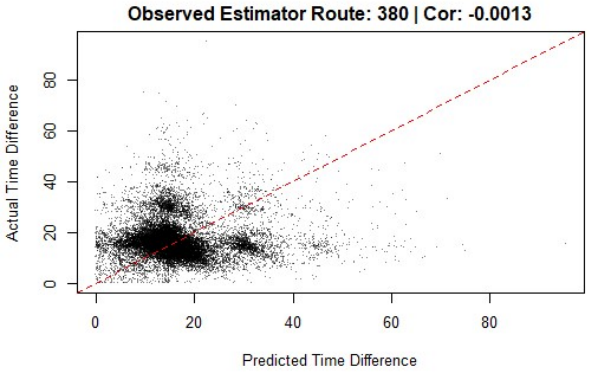

Mixed Estimator Route: 30 | Cor: 0.026
Actual Time Difference
Predicted Time Difference

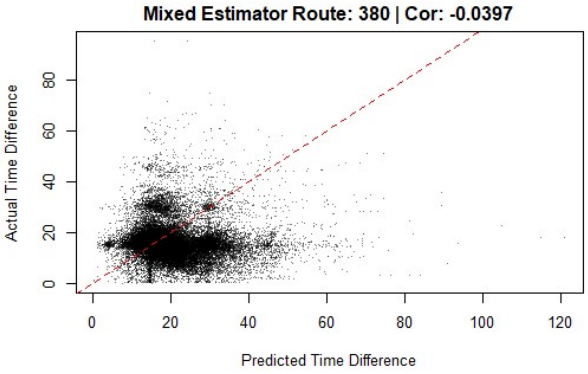

Current Estimator Route: 30 | Cor: 0.1657
Actual Time Difference
Predicted Time Difference

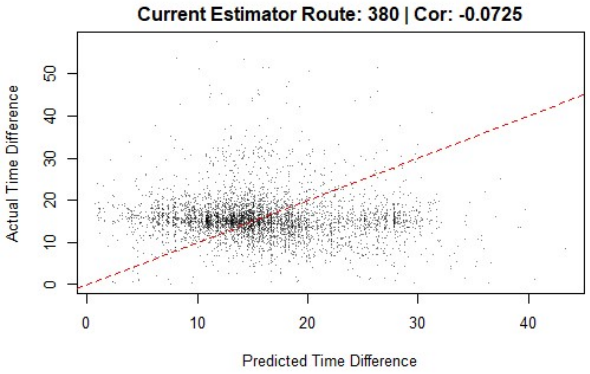

Observed Estimator Route: 349 | Cor: 0.4895
Actual Time Difference
Predicted Time Difference

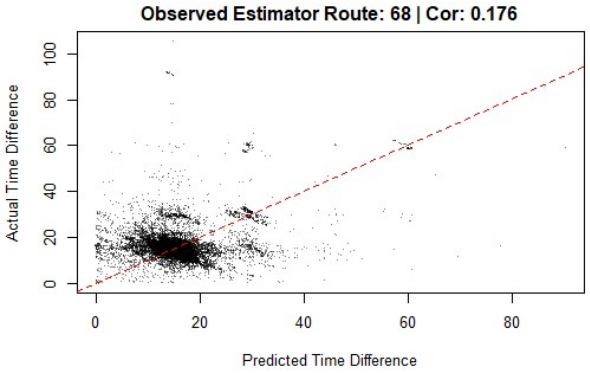

Mixed Estimator Route: 349 | Cor: 0.4124
Actual Time Difference
Predicted Time Difference

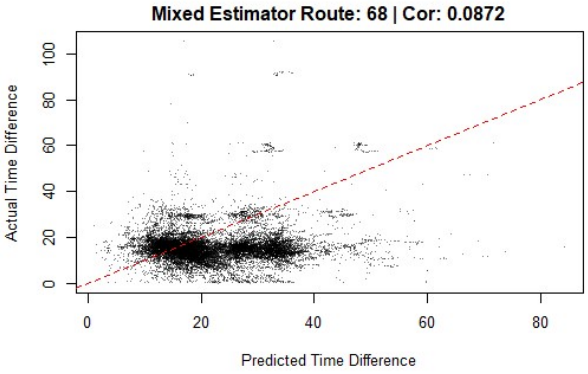

Current Estimator Route: 349 | Cor: -0.0207
Actual Time Difference
Predicted Time Difference

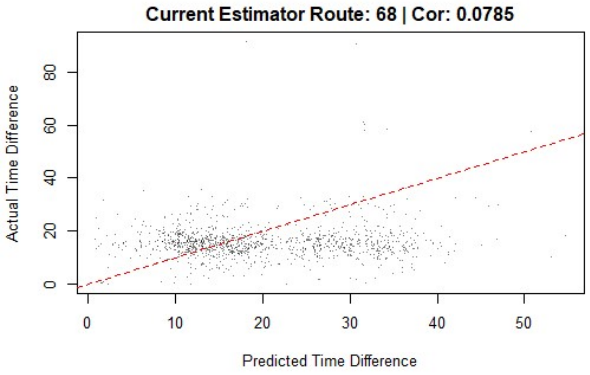

Observed Estimator Route: 380 | Cor: -0.0013
Actual Time Difference
Predicted Time Difference

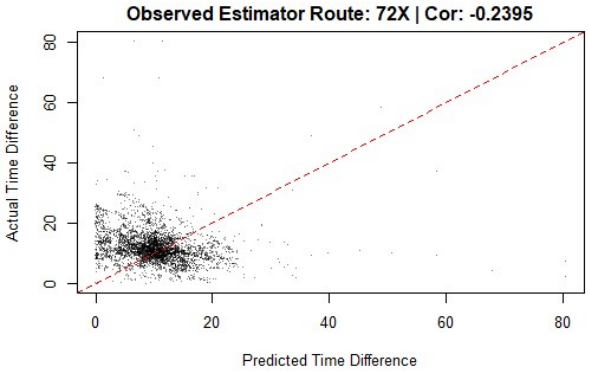

Mixed Estimator Route: 380 | Cor: -0.0397
Actual Time Difference
Predicted Time Difference

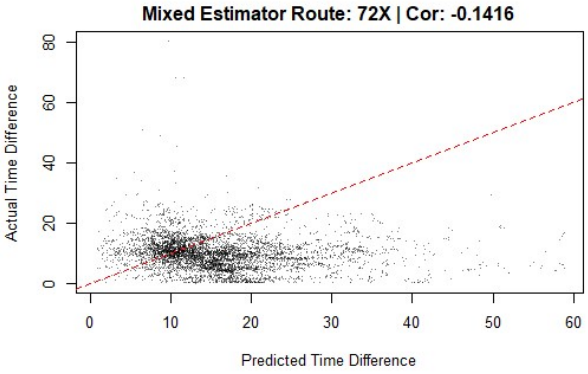

Current Estimator Route: 380 | Cor: -0.0725
Actual Time Difference
Predicted Time Difference

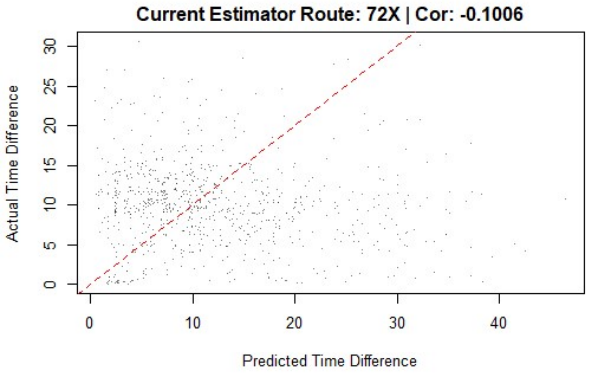

Observed Estimator Route: 68 | Cor: 0.176
Actual Time Difference
Predicted Time Difference

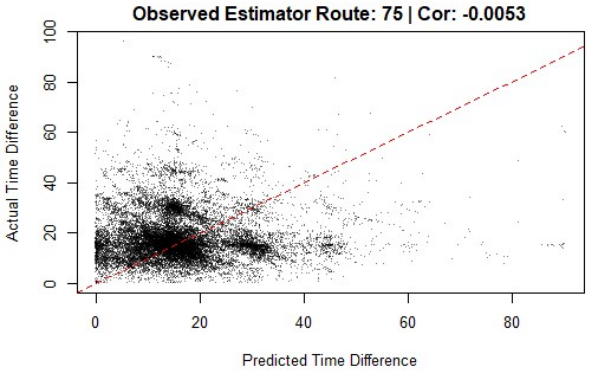

Mixed Estimator Route: 68 | Cor: 0.0872
Actual Time Difference
Predicted Time Difference

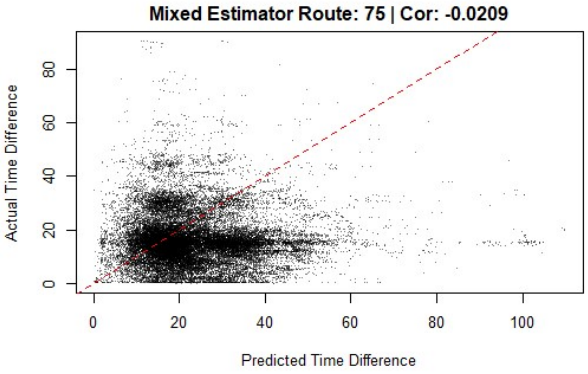

Current Estimator Route: 68 | Cor: 0.0785
Actual Time Difference
Predicted Time Difference

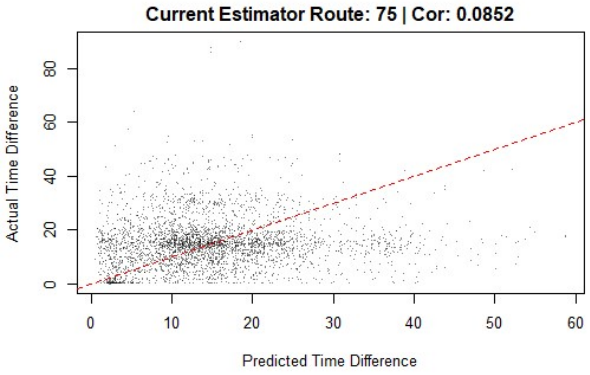

Observed Estimator Route: 72X | Cor: -0.2395
Actual Time Difference
Predicted Time Difference

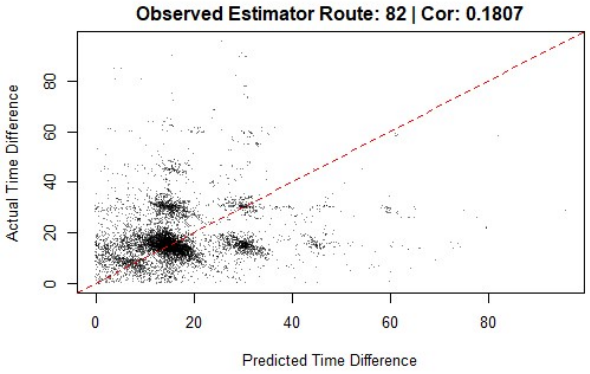

Mixed Estimator Route: 72X | Cor: -0.1416
Actual Time Difference
Predicted Time Difference

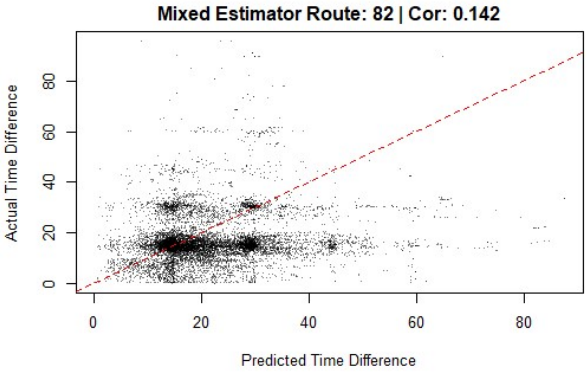

Current Estimator Route: 72X | Cor: -0.1006
Actual Time Difference
Predicted Time Difference

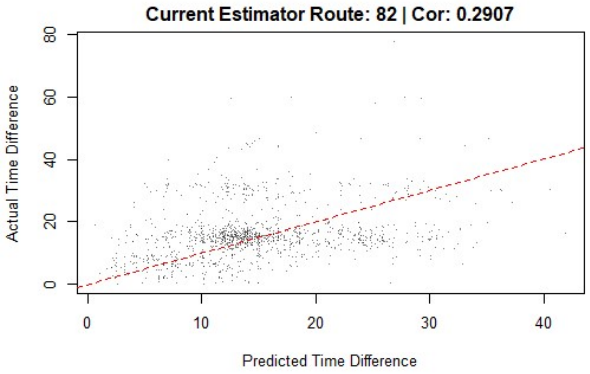

Observed Estimator Route: 75 | Cor: -0.0053
Actual Time Difference
Predicted Time Difference
0
20
40
60
80
100

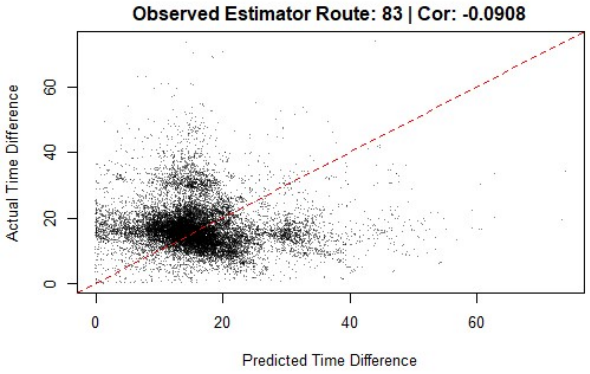

Mixed Estimator Route: 75 | Cor: -0.0209
Actual Time Difference
Predicted Time Difference
0
20
40
60
80
100

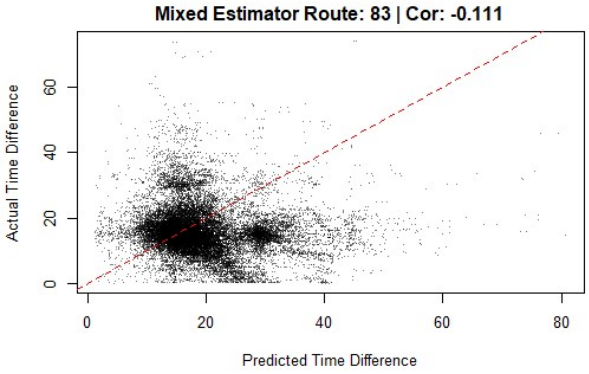

Current Estimator Route: 75 | Cor: 0.0852
Actual Time Difference
Predicted Time Difference
0
10
20
30
40
50
60

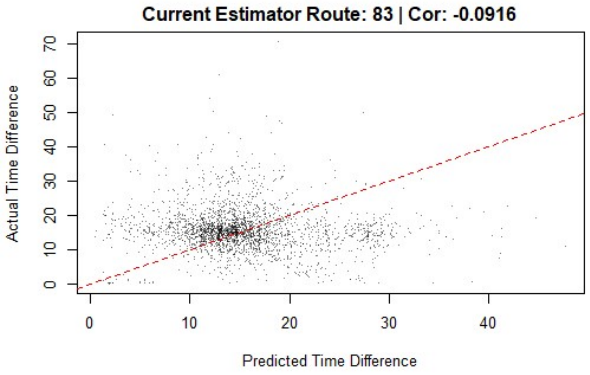

Observed Estimator Route: 82 | Cor: 0.1807
Actual Time Difference
Predicted Time Difference
0
20
40
60
80

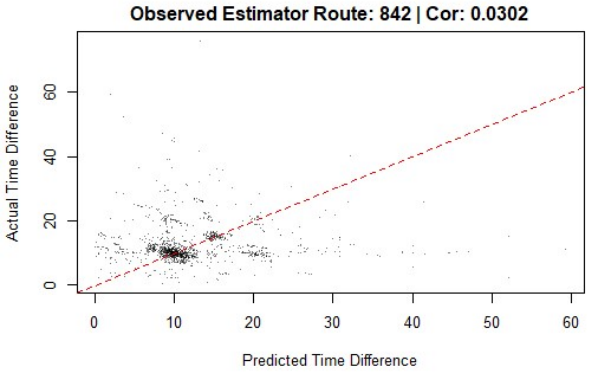

Mixed Estimator Route: 82 | Cor: 0.142
Actual Time Difference
Predicted Time Difference
0
20
40
60
80

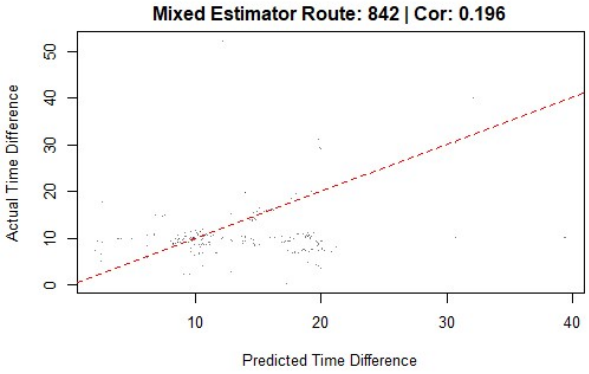

Current Estimator Route: 82 | Cor: 0.2907
Actual Time Difference
Predicted Time Difference
0
10
20
30
40

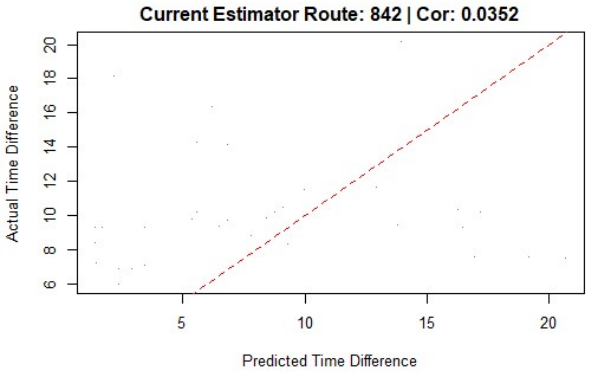

Observed Estimator Route: 83 | Cor: -0.0908
Actual Time Difference
Predicted Time Difference

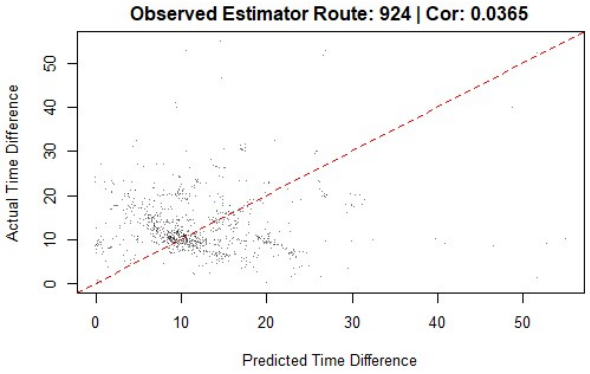

Mixed Estimator Route: 83 | Cor: -0.111
Actual Time Difference
Predicted Time Difference

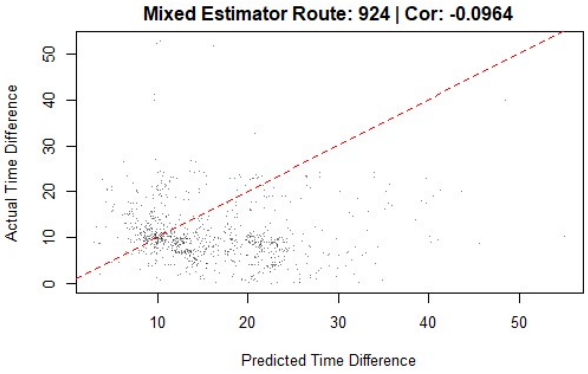

Current Estimator Route: 83 | Cor: -0.0916
Actual Time Difference
Predicted Time Difference

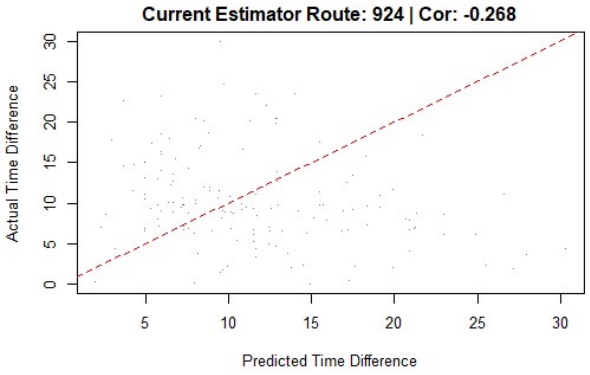

Observed Estimator Route: 842 | Cor: 0.0302
Actual Time Difference
Predicted Time Difference

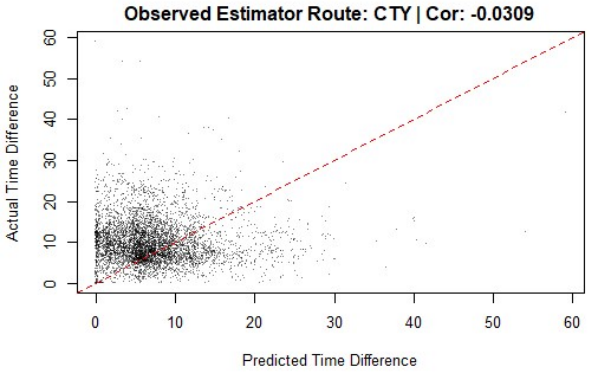

Mixed Estimator Route: 842 | Cor: 0.196
Actual Time Difference
Predicted Time Difference

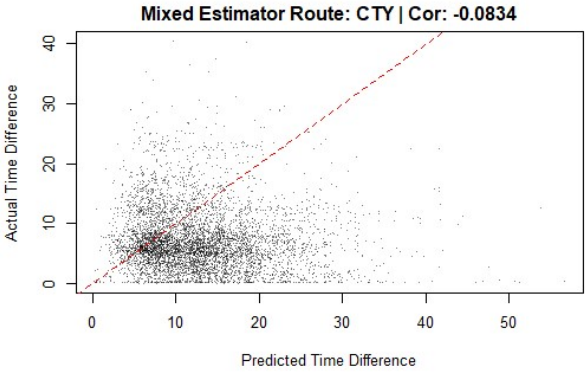

Current Estimator Route: 842 | Cor: 0.0352
Actual Time Difference
Predicted Time Difference

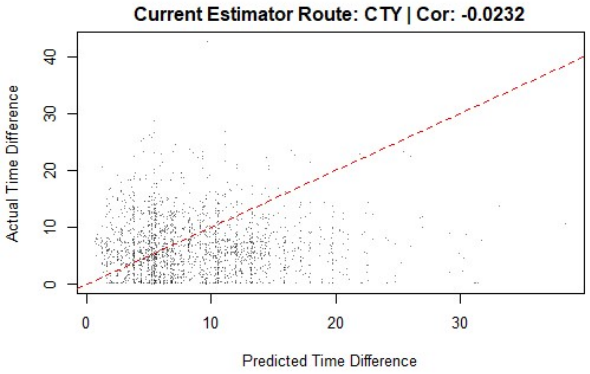

Observed Estimator Route: 924 | Cor: 0.0365
Actual Time Difference
Predicted Time Difference

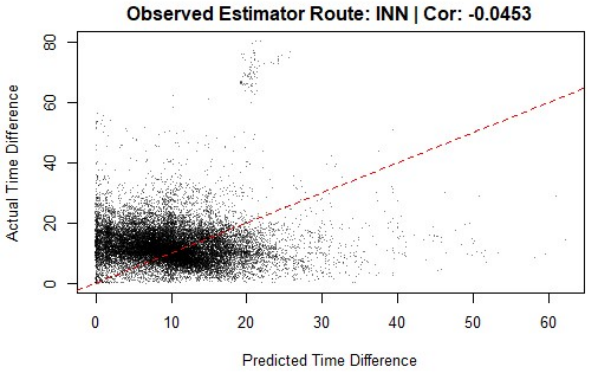

Mixed Estimator Route: 924 | Cor: -0.0964
Actual Time Difference
Predicted Time Difference

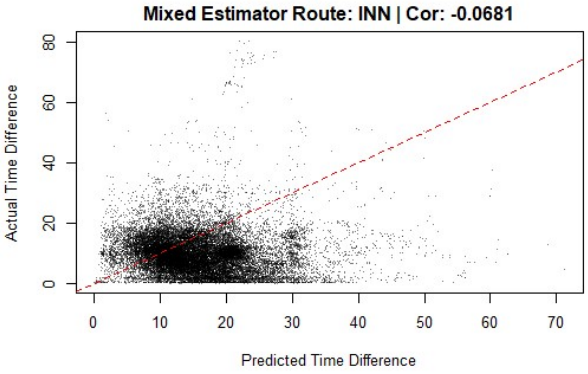

Current Estimator Route: 924 | Cor: -0.268
Actual Time Difference
Predicted Time Difference

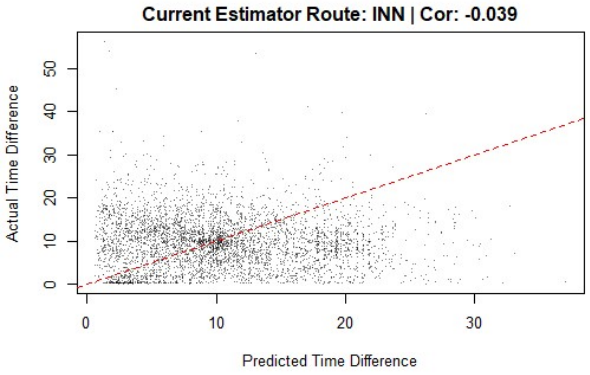

Observed Estimator Route: CTY | Cor: -0.0309
Actual Time Difference
Predicted Time Difference

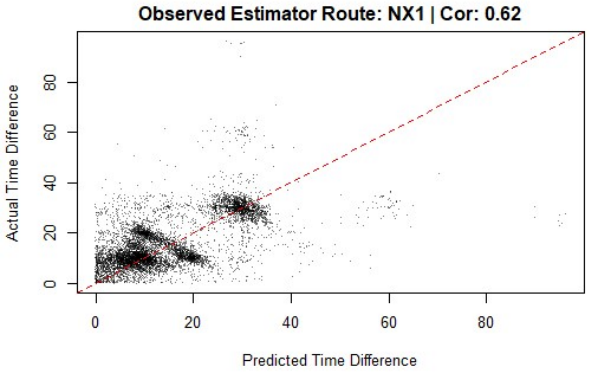

Mixed Estimator Route: CTY | Cor: -0.0834
Actual Time Difference
Predicted Time Difference

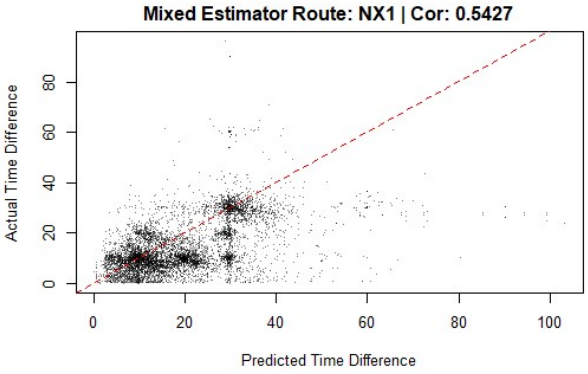

Current Estimator Route: CTY | Cor: -0.0232
Actual Time Difference
Predicted Time Difference

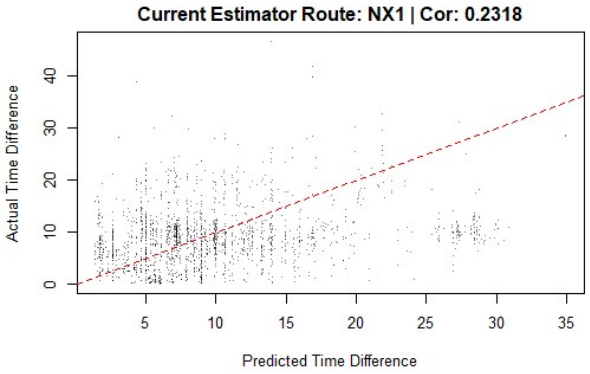

Observed Estimator Route: INN | Cor: -0.0453
Actual Time Difference
Predicted Time Difference

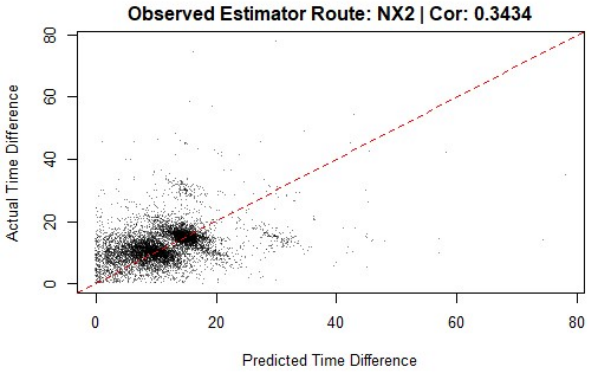

Mixed Estimator Route: INN | Cor: -0.0681
Actual Time Difference
Predicted Time Difference

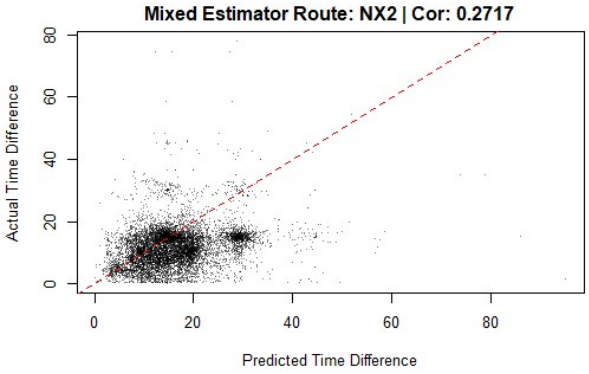

Current Estimator Route: INN | Cor: -0.039
Actual Time Difference
Predicted Time Difference

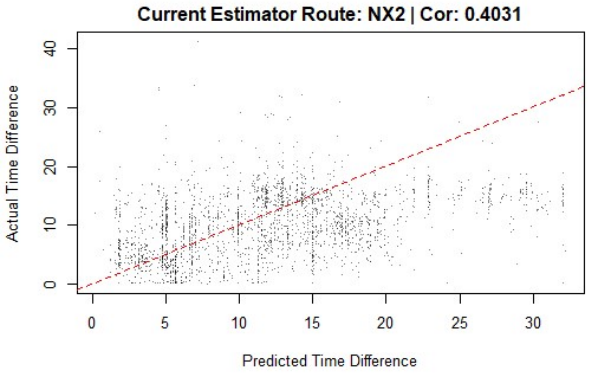

Observed Estimator Route: NX1 | Cor: 0.62
Actual Time Difference
Predicted Time Difference

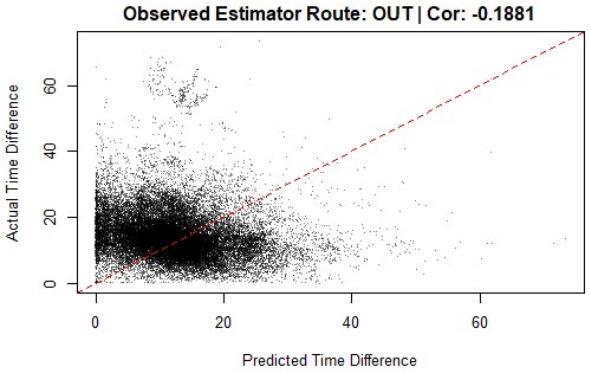

Mixed Estimator Route: NX1 | Cor: 0.5427
Actual Time Difference
Predicted Time Difference

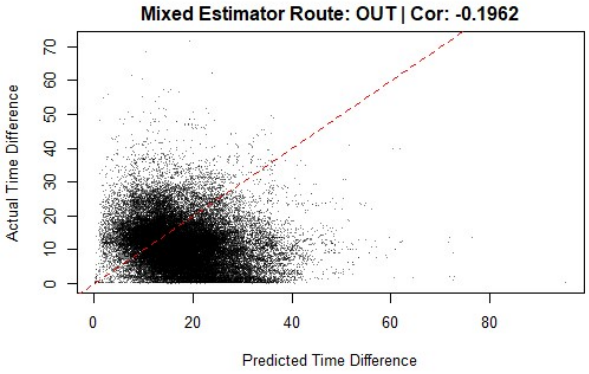

Current Estimator Route: NX1 | Cor: 0.2318
Actual Time Difference
Predicted Time Difference

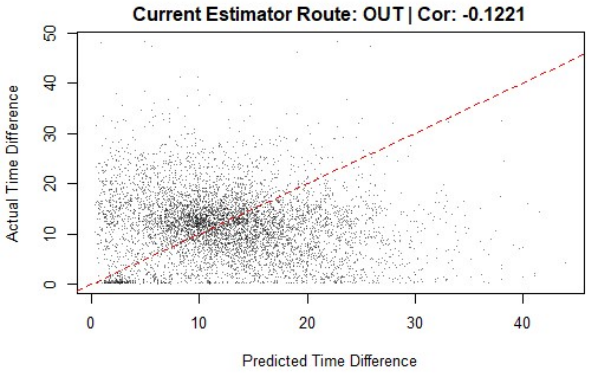

Observed Estimator Route: NX2 | Cor: 0.3434
Actual Time Difference
Predicted Time Difference
0
20
40
60
80

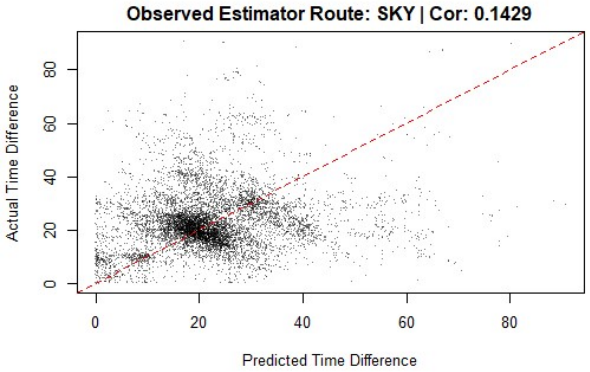

Mixed Estimator Route: NX2 | Cor: 0.2717
Actual Time Difference
Predicted Time Difference
0
20
40
60
80

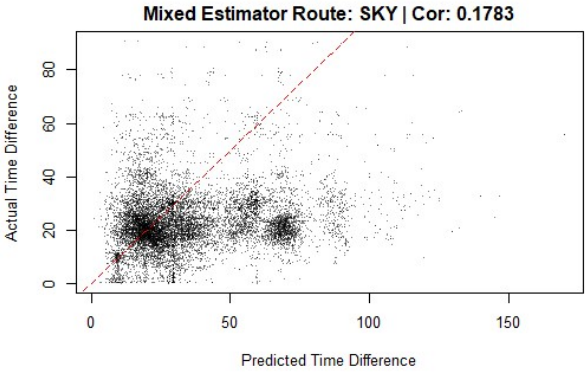

Current Estimator Route: NX2 | Cor: 0.4031
Actual Time Difference
Predicted Time Difference
0
5
10
15
20
25
30
40

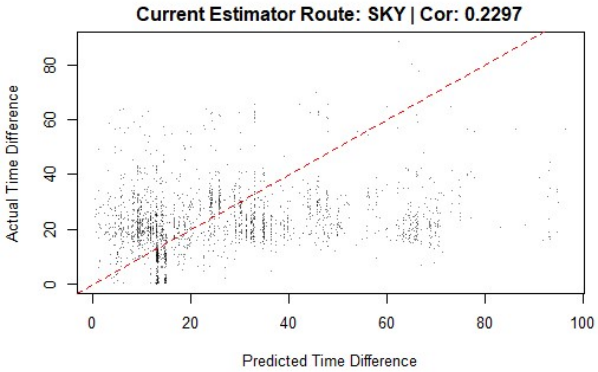

Observed Estimator Route: OUT | Cor: -0.1881
Actual Time Difference
Predicted Time Difference
0
20
40
60

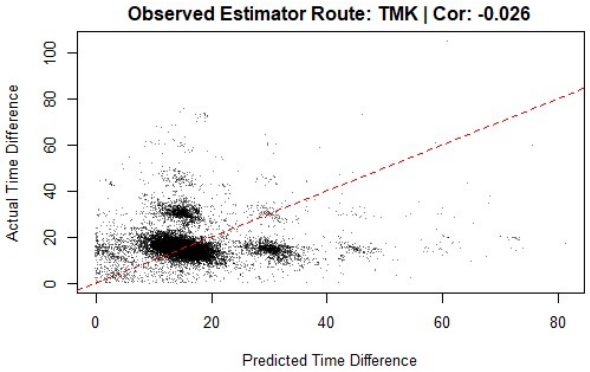

Mixed Estimator Route: OUT | Cor: -0.1962
Actual Time Difference
Predicted Time Difference
0
10
20
30
40
50
60
70
80

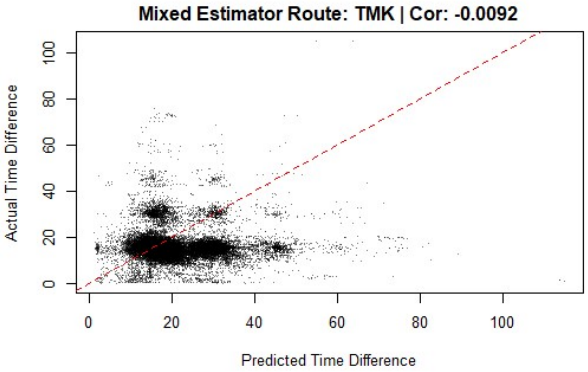

Current Estimator Route: OUT | Cor: -0.1221
Actual Time Difference
Predicted Time Difference
0
10
20
30
40
50

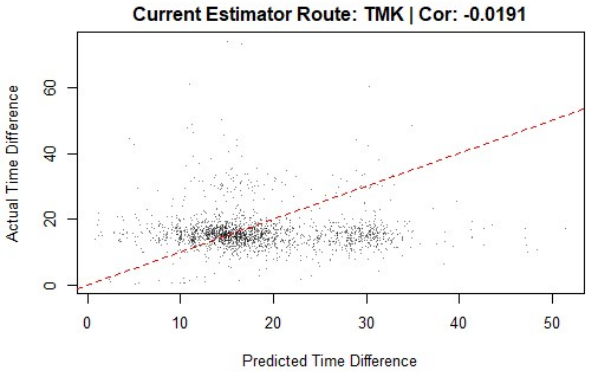

Observed Estimator Route: SKY | Cor: 0.1429
Actual Time Difference
Predicted Time Difference
0
20
40
60
80


Mixed Estimator Route: SKY | Cor: 0.1783
Actual Time Difference
Predicted Time Difference
0
50
100
150

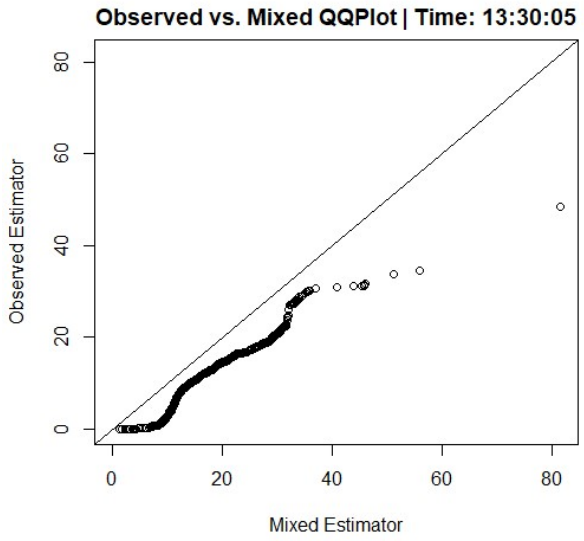

Current Estimator Route: SKY | Cor: 0.2297
Actual Time Difference
Predicted Time Difference
0
20
40
60
80
100

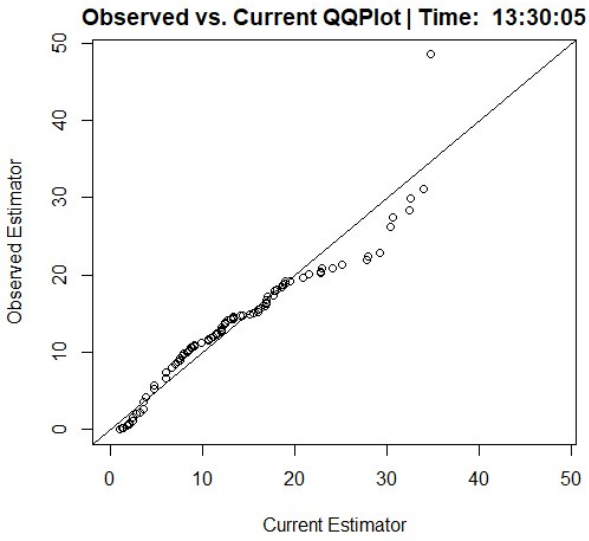

Observed Estimator Route: TMK | Cor: -0.026
Actual Time Difference
Predicted Time Difference
0
20
40
60
80
100

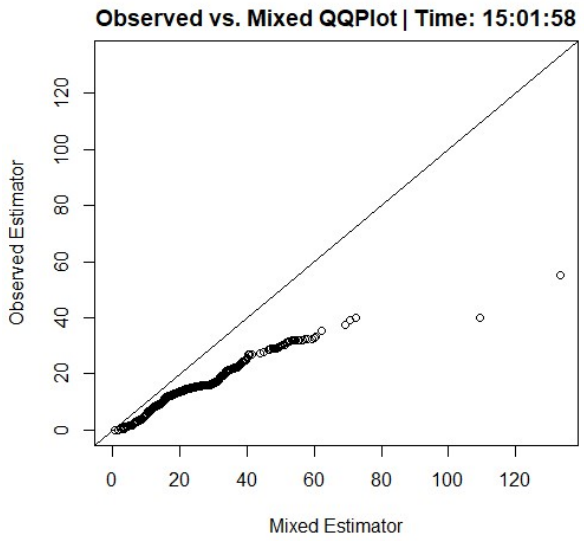

Mixed Estimator Route: TMK | Cor: -0.0092
Actual Time Difference
Predicted Time Difference
0
20
40
60
80
100

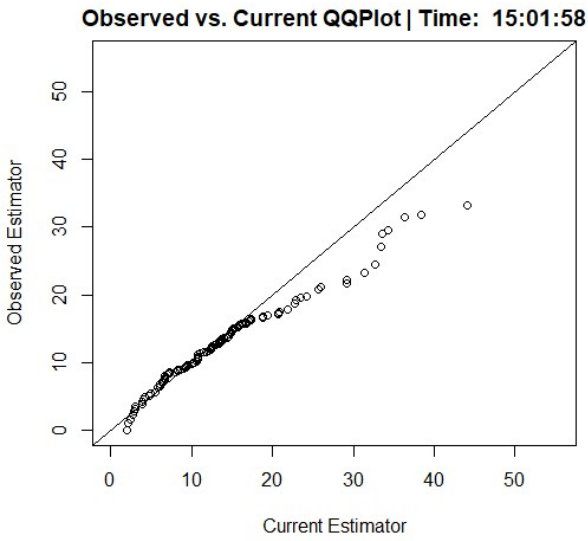

Current Estimator Route: TMK | Cor: -0.0191
Actual Time Difference
Predicted Time Difference
0
10
20
30
40
50
60

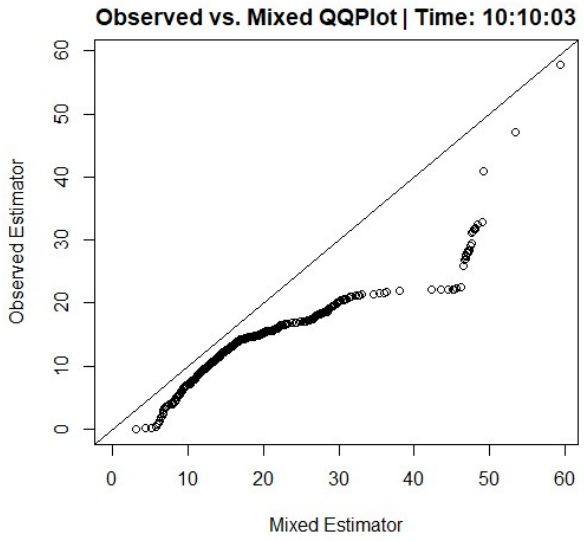
Sample QQ Plot Tests:

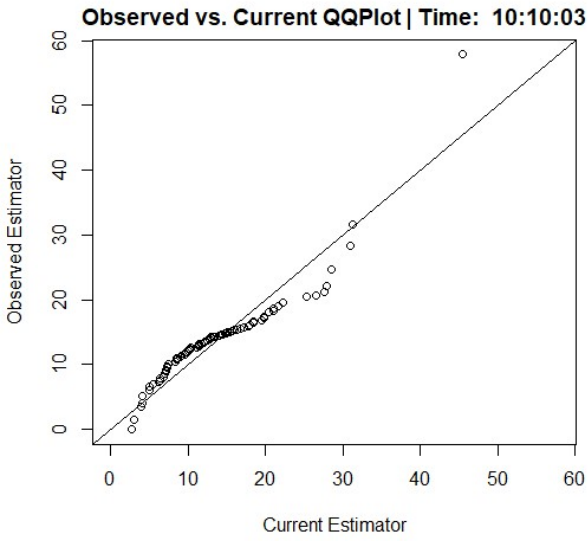


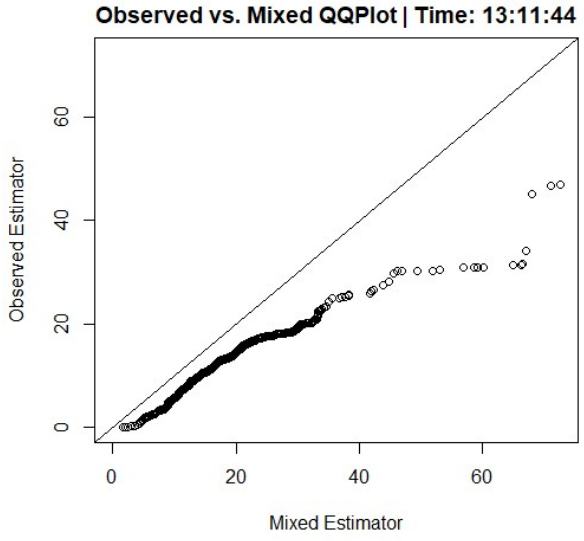


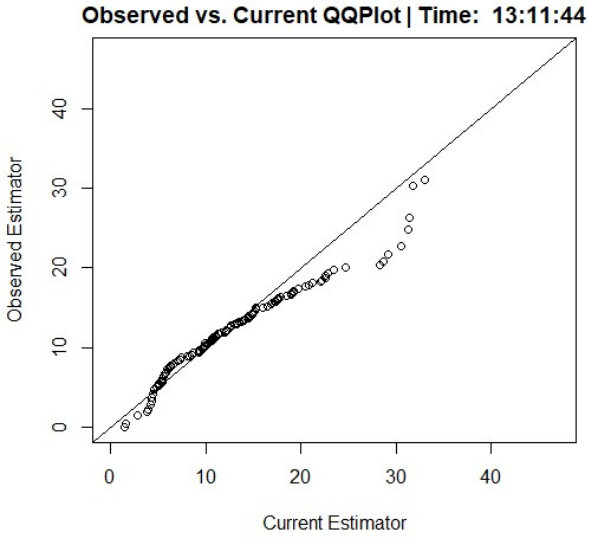


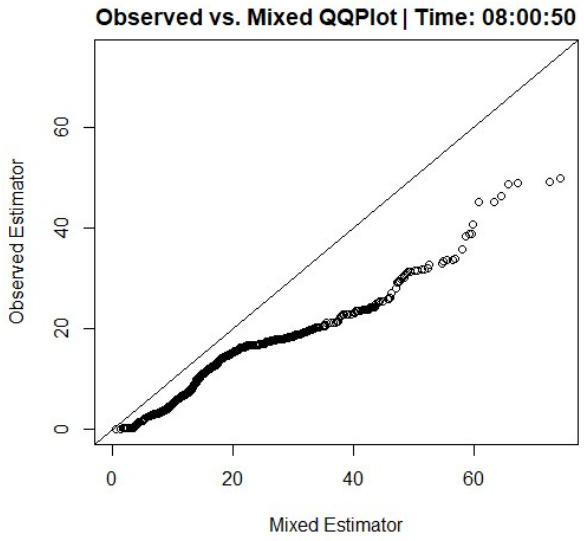


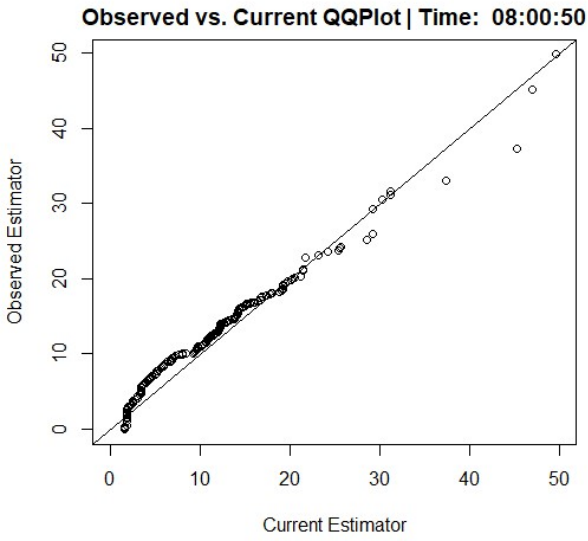

Observed vs. Mixed QQPlot | Time: 10:10:03
Observed Estimator
Mixed Estimator
0
10
20
30
40
50
60

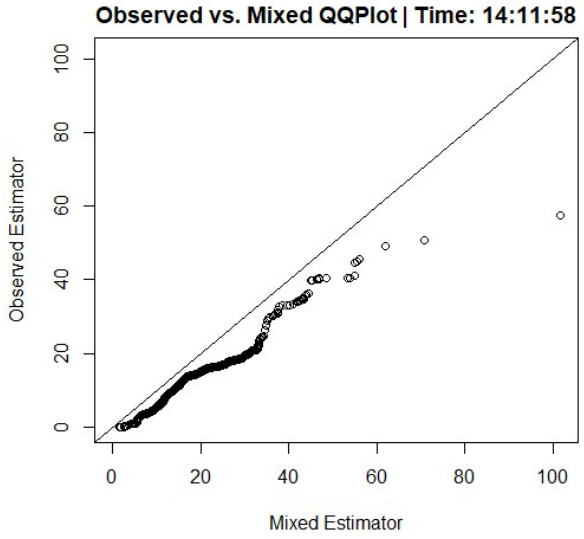

Observed vs. Current QQPlot | Time: 10:10:03
Observed Estimator
Current Estimator
0
10
20
30
40
50
60

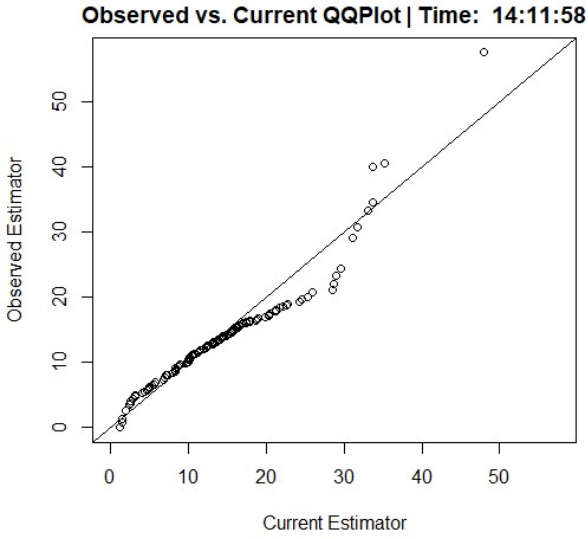

Observed vs. Mixed QQPlot | Time: 13:11:44
Observed Estimator
Mixed Estimator
0
20
40
60

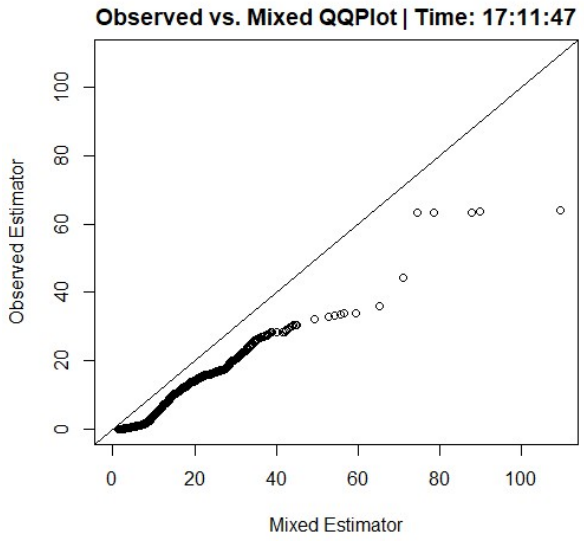

Observed vs. Current QQPlot | Time: 13:11:44
Observed Estimator
Current Estimator
0
10
20
30
40

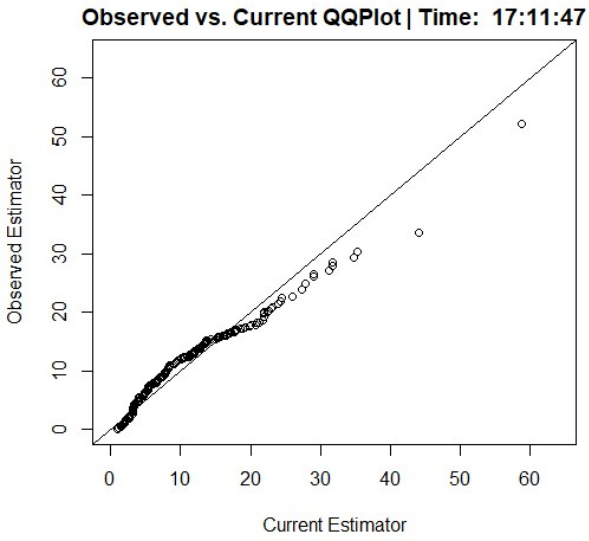

Observed vs. Mixed QQPlot | Time: 08:00:50
Observed Estimator
Mixed Estimator
0
20
40
60

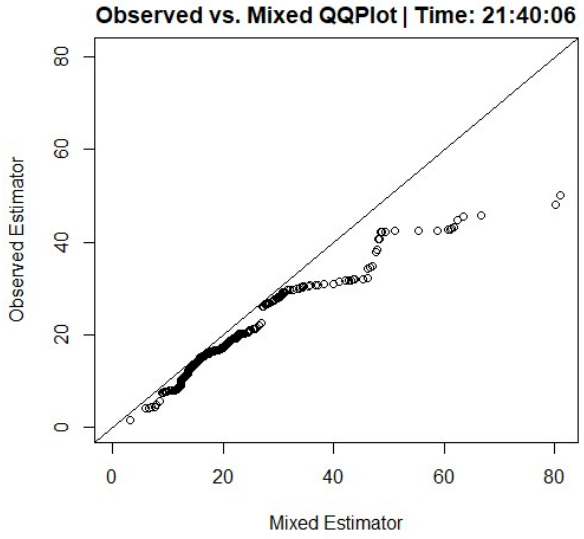

Observed vs. Current QQPlot | Time: 08:00:50
Observed Estimator
Current Estimator
0
10
20
30
40
50

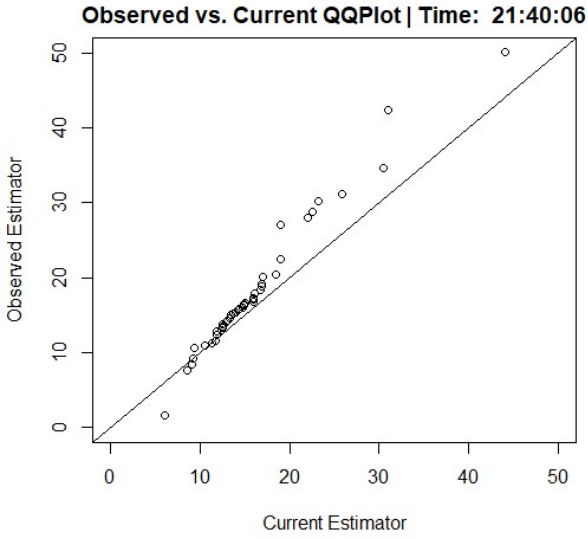

Observed vs. Mixed QQPlot | Time: 14:11:58
Observed Estimator
Mixed Estimator
0
20
40
60
80
100

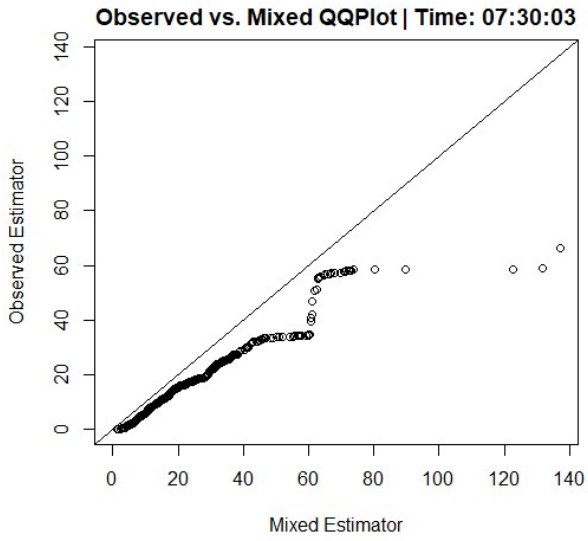

Observed vs. Current QQPlot | Time: 14:11:58
Observed Estimator
Current Estimator
0
10
20
30
40
50

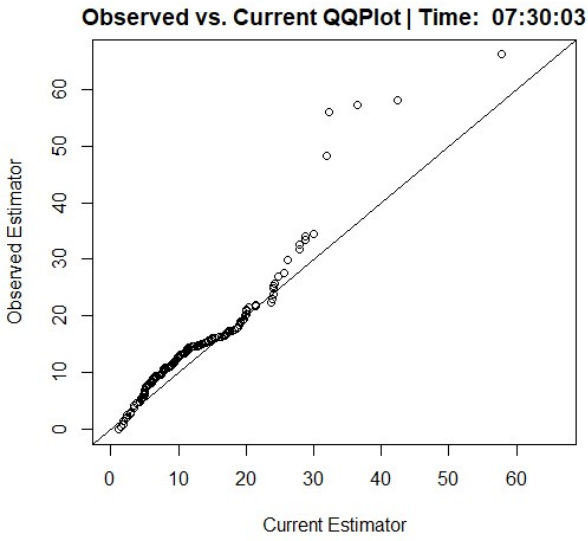

Observed vs. Mixed QQPlot | Time: 17:11:47
Observed Estimator
Mixed Estimator
0
20
40
60
80
100


Observed vs. Current QQPlot | Time: 17:11:47
Observed Estimator
Current Estimator
0
10
20
30
40
50
60


Observed vs. Mixed QQPlot | Time: 21:40:06
Observed Estimator
Mixed Estimator
0
20
40
60
80


Observed vs. Current QQPlot | Time: 21:40:06
Observed Estimator
Current Estimator
0
10
20
30
40
50


Observed vs. Mixed QQPlot | Time: 07:30:03
Observed Estimator
Mixed Estimator
0
20
40
60
80
100
120
140


Observed vs. Current QQPlot | Time: 07:30:03
Observed Estimator
Current Estimator
0
10
20
30
40
50
60