Does Google Really Help? Measuring Real-time Transit Application’s Impact on Public Transit Trip Waiting Time Using High-resolution Real-time data

Luyu Liu

The Ohio State University

1. Introduction

Mobile technologies are changing people’s life in different ways, also for transportation especially public transit. Real-time information (RTI) and real-time transit application (RTA) are reshaping our way to take public transit (PT) system. With all the benefits of PT system, many people still felt reluctant to take public transit. A part of topmost reasons why people do not take PT are: long travel time especially long waiting time, lack of comfort, lack of control/certainty, and unreliability (Beirão & Cabral, 2007). Correspondingly, RTA can improve PT user experience in three factors: strengthening control over the timetable, optimizing waiting time, and adapting unreliability.

RTA will provide users ability to comprehend the sophisticated timetable in a PT system, for both scheduled and real-time timetable. Paper timetable are inflexible and limited to schedule, and traditional phone call and text service are also limited to scheduled static timetable. RTA provide both scheduled and real-time support for PT users with smart phone through user-friendly interface.

RTA will decrease the waiting time. Waiting time is a critical factor affecting people’s preference of transportation (Beirão & Cabral, 2007). RTA can plan best time for users to leave for the public transit based on the walking time and PT timetable. Ideally, RTA can diminish the waiting time to 0, which means as soon as users arrive at the stop, the bus arrives. Many studies investigated the waiting time reduction by real-time information: 91% percent of RTI users self-reported spending less waiting time in Seattle, 2010 (Ferris, Watkins, & Borning, 2010), and RTI users can save 2.4 minutes in Seattle, 2011 (Watkins, Ferris, Borning, Rutherford, & Layton, 2011) and 1.79 minutes in Tampa, 2014 (Brakewood, Barbeau, & Watkins, 2014) according to self-reported survey.

RTA will adapt PT system’s unreliability when optimizing waiting time (Brakewood et al., 2014). For all PT system, delay is inevitable. When delay happens, real-time transit application (RTA) users can use their smart phone to get access to the real-time status of buses and plan their trip accordingly. With the help of RTA, users can have an anticipation about the on-time transit’s delay, thus postponing the departure time to reduce the waiting time at the stop. Many RTAs provide real-time trip planner to help users to schedule a best time to take public transit, such as Google Map and Transit.

However, besides the benefit RTA produces, the smart trip strategies can lead to several undesirable results. The optimization of waiting time also comes with high risk of missing a bus. Figure 1 shows the changing trend of waiting time along with user’s arrival time at the stop. The optimal point, where waiting time is 0, is exactly the point with highest risk of missing the scheduled bus. Meanwhile, with the walking time increases, the instability of the PT trip’s real-time performance also increases. During the time when users are walking from home to the stop, a delayed bus can speed up to catch up the former delay, which can result in missing the bus for RTA users. Therefore, we would like to calculate and compare the missing rate for both non-RTA and RTA users.

Figure 1 Waiting time's changing pattern along with user’s arrival time.

Nevertheless, even with more and more technology breakthroughs in this area, there are still economic and intellectual barriers for certain population to get access to the real-time public transit information. People who do not use smart phone application or do not own a smart phone cannot know the real-time status. These ordinary users have to plan their trip according to the schedule, their daily experience, or even random. Based on this fact, we would like to assess the average waiting time difference between non-RTA and RTA users and social justice issue behind the difference.

In conclusion, the project is to assess and optimize the waiting time reduction and consequent missing risk by consulting real-time bus feed compared to ordinary users. Based on the results, we can prove that whether it is statistically useful for users to follow different possible trip planning strategies derived by RTA’s trip planner, and how reliable different RTA strategies are. We will adopt several reasonable assumptions to build a mathematical model and implement the model with GTFS (General Transit Feed Specification, real-time transit data standard) data provided by COTA (Central Ohio Transit Agency) in Columbus, Ohio.

1. Literature review

The idea of measuring real-time information’s impact on public transit system and waiting time was first introduced even before the internet era. In 1995, Reed (1995) investigated signage and telephone’s real-time information’s impact on passengers’ waiting time (Reed, 1995). After the widespread application of smart personal devices, numerous studies examined new technologies’ influence on users’ behaviors. Recently, real-time information is becoming more accessible due to less expensive automated vehicle location system and the open data policy. Correspondingly, the body of literature is steadily growing and there are numerous studies investigating real-time information’s impact on public transit users (Brakewood et al., 2014; Brakewood, Macfarlane, & Watkins, 2015; Cats & Gkioulou, 2017; Ferris et al., 2010; Fries, Dunning, & Chowdhury, 2011; Liu, Shi, & Jian, 2017; Papangelis, Nelson, Sripada, & Beecroft, 2016; Watkins et al., 2011).

Brakewood and Watkins (2018) systematically reviewed and categorized most research studying real-time transit information impact on the public transit system and users (Brakewood & Watkins, 2018), including waiting time, path choice, and ridership. The paper classifies all studies according to five dimensions: information media, mode of transit, location, methodology, and findings. Similarly, this literature review will inspect studies that investigated RTA’s impact on the waiting time, based on their real-time information media and method.

* 1. Real-time information media

We can categorize the research according to their information media, including static signage, telephone and text services, and smart phone application. Signage and at-stop displays can provide PT users useful information and reduce actual and perceived wait time (Dziekan & Vermeulen, 2006; Reed, 1995). Moreover, at-stop displays’ psychological effect is even more important: systems showing the next train or bus’s departure time can greatly release anxiety (Dziekan & Vermeulen, 2006). However, due to its static nature, user can only get the information at the stop, which limits its actual effectiveness. Telephone service and text services are more useful than signage (Reed, 1995), since for the first time, people can get access to the real-time data out of the station.

However, it is after the introduction and wide application of the real-time transit information that people’s transit experience is actually changed by smart phone and real-time transit applications.

* 1. Method

Brakewood and Watkins (2018) categorized all studies based on the methods used: Survey-based methods, simulation models, and aggregate-level econometric analysis (Brakewood & Watkins, 2018). Survey-based methods include on-board surveys (Fan, Guthrie, & Levinson, 2016), before-after surveys (Chow, Block-Schachter, & Hickey, 2014), web-based surveys (Ferris et al., 2010), in-person surveys (Watkins et al., 2011), interviews and observations (Papangelis et al., 2016), and stated preference surveys (Liu et al., 2017).

[Survey-based method]

Survey-based method is definitely the majority among all RTI impact studies. Surveys sample respondents with the same questions; they measure many variables, test hypotheses, and conclude temporal sequence from questions about past behavior, experiences, or characteristics (Neuman & Robson, 2014). Without the support of automatic real-time data, passenger surveys are the most direct methods to assess PT system use, especially for user experience and perceptions. Meanwhile, survey data can quantitatively assess different attributes using self-reported data (Neuman & Robson, 2014), which partially guarantees generalizability and authenticity. Besides, in contrast to the automatic generated data, the surveys’ data also point to users, instead of vehicles. The human-centered nature of survey data also guarantees its direct and close connection with human per se.

However, survey methods have their problems: Some survey-based methods rely on user’s self-reported information, which lacks precision and authenticity especially for noncognitive value. Surveys measure the user’s perceptual estimation of the assessed value. Compared with value obtained by physical sensors, the self-reported information may be obfuscated and biased by threatening questions (Bradburn et al., 1979) and its self-evident nature (Goyder, 1986).

Survey’s sampling strategy and measurement is often questionable. First, the survey’s size, especially for some traditional data collection methods, is critically small due to high cost of data collection (Goyder, 1986). Second, some methods, especially IT (Information technology)-based methods, are often biased and dubious. Mail survey (Rossi, Wright, & Anderson, 2013), text/phone call survey, and internet-based survey (Wright, 2005) can significantly reduce the time and economic cost of the survey. However, these methods face a same problem: it is hard to access a representative sample (Rossi et al., 2013; Wright, 2005). For public transport system, not all users can get access to these services while the survey will only sample these specific people.

[Simulation]

Mathematical simulation is often used to investigate and solve problems that are too difficult or costly to measure directly. For example, Cats and Gkioulou (2014) adopted an agent-based model to simulate the influence of PT reliability and real-time information on waiting time uncertainty (Cats & Gkioulou, 2017). Agent-based model simulation usually adopts several assumptions and represents the simultaneous actions and interactions of various agents. The simulation tries to imitate and predict the performance of a complex system such as PT system (Gkioulou, 2013). However, the effectiveness of the model is debatable, and the adopted assumptions could also be inconsistent with the reality.

Although the overall impact of RTI on waiting time is well explored, few studies investigated the variance (Brakewood & Watkins, 2018), including spatial and temporal patterns. Meanwhile, ironically, there are no studies using actual real-time transit data source due to the lack of these data and corresponding theory support. We would like to address these gaps of waiting time problem in the real-time transit data context and implement it using actual real-time transit data.

1. Methodology
   1. Data source

Introduced by Google first in 2006, GTFS is a collection of two data types: GTFS static and GTFS real-time expansion. GTFS static indicates the schedule data of a PT system in several separate tables (Google Developers, 2016). GTFS static is the current *de facto* standard for PT system schedules and PT geographic information (Google Developers, 2016). As a standard for open data, it is easy to share and access for the public, open-source programmer, and researchers. For this paper, we collected and organized all history schedule data in MongoDB and Python environment from Application Programming Interface (API) provided by Central Ohio Transit Authority (COTA) bus system in Columbus, Ohio.

Besides schedule data, GTFS real-time data provides vehicles’ geographic data with high temporal resolution. GTFS real-time includes two parts: buses’ location real-time data and the trip updates, which specifies the fluctuations in the real-time timetable (Google Developers, 2018). Unlike many traditional transit data, GTFS real-time data have large volume, high resolution, normalized standard, and full system coverage. For this study, we collected GTFS real-time data from API from COTA bus system with the frequency of 1 minute for more than 1 year.

* 1. Synchronization

Taking a bus could be conceptualized as a synchronization process between the walking trip to the target stop and the target bus’s *trip sequence array*. Trip sequence array is defined as the collection of trips running on the same route as the target bus in the direction of target bus.

Depending on user’s arrival time at the stop *t*, the actual bus that user will take can be different from the scheduled one. We use the same concept in the transfer synchronization process: *desynchronization degree* (DD), to measure the desynchronization between the bus and user at the stop. DD is an integer indicator that represent how many buses the user loses in the trip sequence array.

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

If the user arrives at the stop between bus DD = n – 1 and bus DD = n, then the user will take bus DD = n.

The process of walking is linear: the users can strictly control the walking time by selecting their home departure time (HDT). For a user, the relationship between HDT and arrival time is linear.

Nevertheless, the actual real-time performance of the bus is non-linear: first, the users cannot directly control the boarding time by selecting their HDT, and the relationship between HDT and user’s boarding time is non-linear; second, the bus will not run at a fixed velocity and the expected time of arrival of bus at the stop is constantly changing. If the bus is delayed when the user departs home, during the walking time, the bus may catch up a part of delay by accelerating.

We therefore define the concept of *reclaimed delay (RD).* It is the time difference between the actual time of arrival (ATA) and the expected time of arrival (ETA) at the stop.

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Where: S is the collection of stops on the route, i is the target stop, t is the current time.

The reclaimed delay could be tiny but critical: consequently, the user will lose the bus and suffer waiting time penalty for a relatively long time. Thus, the synchronization of these two processes is highly unstable. Figure 3 shows a space-time diagram of the synchronization process.

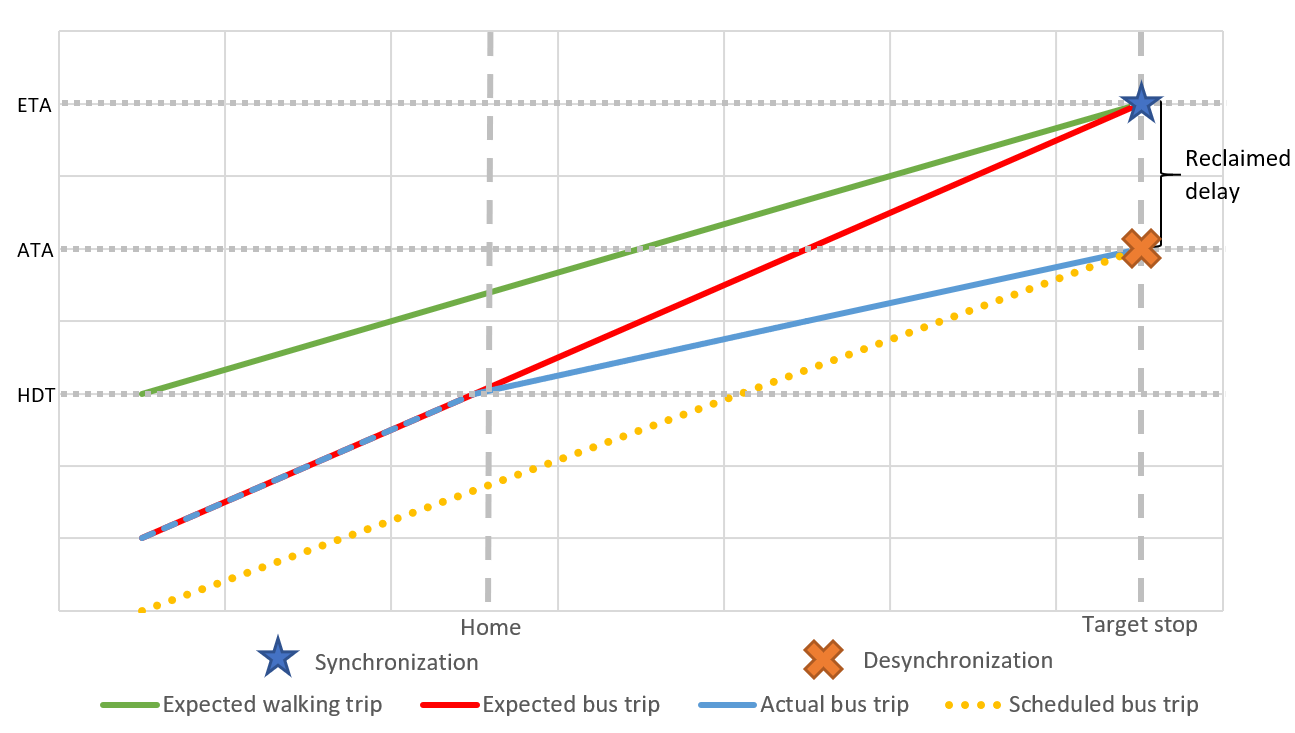


Figure 2 Space-time diagram of the synchronization and desynchronization

* 1. Volunteered optimization

Previous research concentrated on optimization in the stage of planning and operation. Due to the lack of authority and information access, ordinary users were rarely considered as a part of optimization process. However, with RTI, although users still cannot directly improve the real-time systematic performance (delay, ridership) of the system per se, optimization can be conducted in the individual level to reduce waiting time. Despite PT systems’ instability and uncertainty, users with RTA can adapt and optimize each trip according to the delay and real-time information. With the waiting time reduction in the individual level, the overall waiting penalty will also be diminished.

To reduce waiting time, the only controllable factor for all public transit users is when to leave home for the transit. RTA relaxes the fixed timetable in a frequently delayed PT system, thus saving waiting time for RTA users. Depending on how to determine the leaving time, there are different strategies for both RTA and non-RTA users and their different purposes.

[**Optimal relaxation (OR)**]

Supposing an omniscient public transit user who are always aware of the real-time and future status, they can adjust their home departure time (HDT) accordingly. The user will always catch the desired bus in time without any waiting, regardless of his home’s distance from the stop. Practically, nobody can achieve real global optima. OR is a good benchmark for all strategies, which represents the best strategy for all possible waiting time reduction strategies.

[**Null relaxation (NR)**]

If a user has urgent affairs, such as work with strict timetable and medical emergency, he prefers earlier final arrival time than convenience. Under this circumstance, he will follow the scheduled timetable of the PT system regardless of waiting time. Consequently, the RTA user will not benefit from waiting time reduction. However, since no bus/train will leave earlier than the scheduled time, NR minimizes the missing risk. NR is another benchmark for waiting time reduction strategies, which has the lowest missing risk.

[**Radical relaxation (RR)**]

If a user encounters extreme weather events, especially cold weather and heavy precipitation, he may want to reduce the waiting time as much as possible. In practice, an RR user will only leave if RTA tells him so based on the real-time data, when the bus’s ETA at the stop is equal to walking time plus current time:

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Where: is the user’s home departure time (HDT), is the walking time from user’s home to the stop, is the scheduled bus’s ETA at the stop given by RTA and real-time data, and is the current time when .

This strategy can achieve temporary optima. However, due to the instability of PT system, the missing risk of RR is also the highest. Due to the possible reclaimed delay, the bus will likely arrive earlier than ETA.

[**Prudent relaxation (PR)**]

If a user would like to save waiting time and keep some degree of missing risk, he may leave several minutes earlier than RR. This short time buffer, which is defined as insurance buffer (IB), trades some time to reduce missing risk, thus decreasing instability.

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

Where: IB is the short time buffer.

PR and RR can be categorized as a PR family, for RR is a special case of PR with IB = 0. With different IBs, each PR strategy can vary in actual waiting time. We would like to optimize IB and find the best PR strategy with max average waiting time reduction.

[**Empirical relaxation (ER)**]

If a user can get access to the historical information, either from a database or daily experience, the user can relax the DTH based on the empirical average time without any real-time help. This is a common non-RTA strategy. We will also calculate the waiting time difference between ER and NR to confirm its efficiency.

[**Arbitrary relaxation (AR)**]

Before the time of smart phone, text, and public information, under many circumstances, PT users are not particularly planning their trips. Normally, they just walk to the stop and catch the next bus arbitrarily. Intuitively, this strategy is not very efficient. To prove it, we will also calculate the waiting time difference between AR and NR.

|  |  |  |
| --- | --- | --- |
| Strategy | Non-RTA users | RTA users |
| Optimal relaxation | X | X |
| Null relaxation | ✓ | ✓ |
| Radical relaxation | X | ✓ |
| Prudent relaxation | X | ✓ |
| Empirical relaxation | ✓ | ✓ |
| Arbitrary relaxation | ✓ | ✓ |

Table 1 Different strategies' availability for non-RTA and RTA users

* 1. Measures and optimization

We would like to measure the difference of waiting time and risk of missing bus/train between the RTA users and non-RTA users. Therefore, we present the definitions of two indexes: missing risk (MR) and average waiting time (AWT).

[**Missing risk**]

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

Where: n is total number of trips;

[**Waiting time**]

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

Where: is the actual waiting time, T(t) is user’s boarding time, and t is user’s arrival time at the stop. is user’s HDT, and is user’s walking time from home to the stop.

So, the user boarding time T(t) is:

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

Thus:

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

(7) proves that the only factor that user can control and can affect waiting time is .

Besides single trip’s waiting time, we can also calculate the mean of waiting time’s distribution of all trips:

|  |  |  |
| --- | --- | --- |
|  |  | (8) |

Where: n is total number of trips.

[**PR optimal**]

For PR or RR strategy, the users will plan their HDT according to the bus’s ETA instead of the real arrival time. Here we define expected waiting time:

|  |  |  |
| --- | --- | --- |
|  |  | (9) |

Where: is the ETA, t is the user’s arrival time at the stop, is PR strategy’s HDT, and is the walking time.

Combining (2) and (9), we have:

|  |  |  |
| --- | --- | --- |
|  |  | (10) |

For PR family, the expected waiting time is equal to the insurance buffer.

Figure 2 shows the theoretical relationship between user’s HDT and expected waiting time, missing risk, and waiting time.

Figure 3 Theoretical relationship between user HDT and expected waiting time, risk and waiting time.

To find the , we constitute the optimization problem in the following formula.

Subject to:

Where: is the expected departure time of the scheduled bus, t is the user’s expected arrival time with radical relaxation, Tr is the collection of all trips and i is a trip in this collection, is the user’s walking time within a designated range.

The optimization process of depends on IB, therefore, there are different restrictions that can be applied to relax the conditions. First, we can minimize over for each implemented trip, which will have a different IB for each day and each trip. However, this could be an overkill because implemented trip is too specific. Second, we can minimize over a same for all trips and days. It is easy to be calculated and applied by users since it has only one parameter. However, this optimization result does not consider variations between different stops and routes. Third, we can minimize over for each trip’s average. This is a reasonable aggregate level, guaranteeing flexibility and variation. In practical, we will calculate the AWT for every stop for different IB, such as 10s, 20s, 30s, 1 minute. After plotting the AWT – HDT graph, we will find the optimal one.

[**Waiting time difference**]

For non-RTA users who has no access to the real-time data, the most rational and practical strategy is NR or AR. And for RTA users, the most beneficial strategy is PR with an optimal insurance buffer. We introduce the difference between NR's waiting time and ’s waiting time:

|  |  |  |
| --- | --- | --- |
|  |  | (9) |

In practice, we will calculate for RR,, ER, AR, and compare these strategies’ efficiency.

* 1. Implementation and analysis

We collected GTFS real-time trip update data and corresponding GTFS static schedule data of the COTA (Central Ohio Transit Agency) bus system in Columbus, Ohio from January 2018. The data is stored in a MongoDB database. The GTFS trip update contains the ETA and ATA for every stop and every minute. We will develop the algorithm in Python environment and maintain our smart transit database.

Our analysis includes:

1. Find the PR optimal in the PR family;
2. Calculate waiting time difference for RR, PR optimal, ER, AR;
3. Visualize the waiting time difference for different strategy, based on different stops and different walking time;
4. Visualize the temporal pattern.
5. Analysis

We first calculate the waiting time difference for walking time from 0 to 10 minutes between PR (IB = 2 minutes) and NR users, which represents RTA and non-RTA users. Figure X visualizes the difference on COTA bus route No. 10.

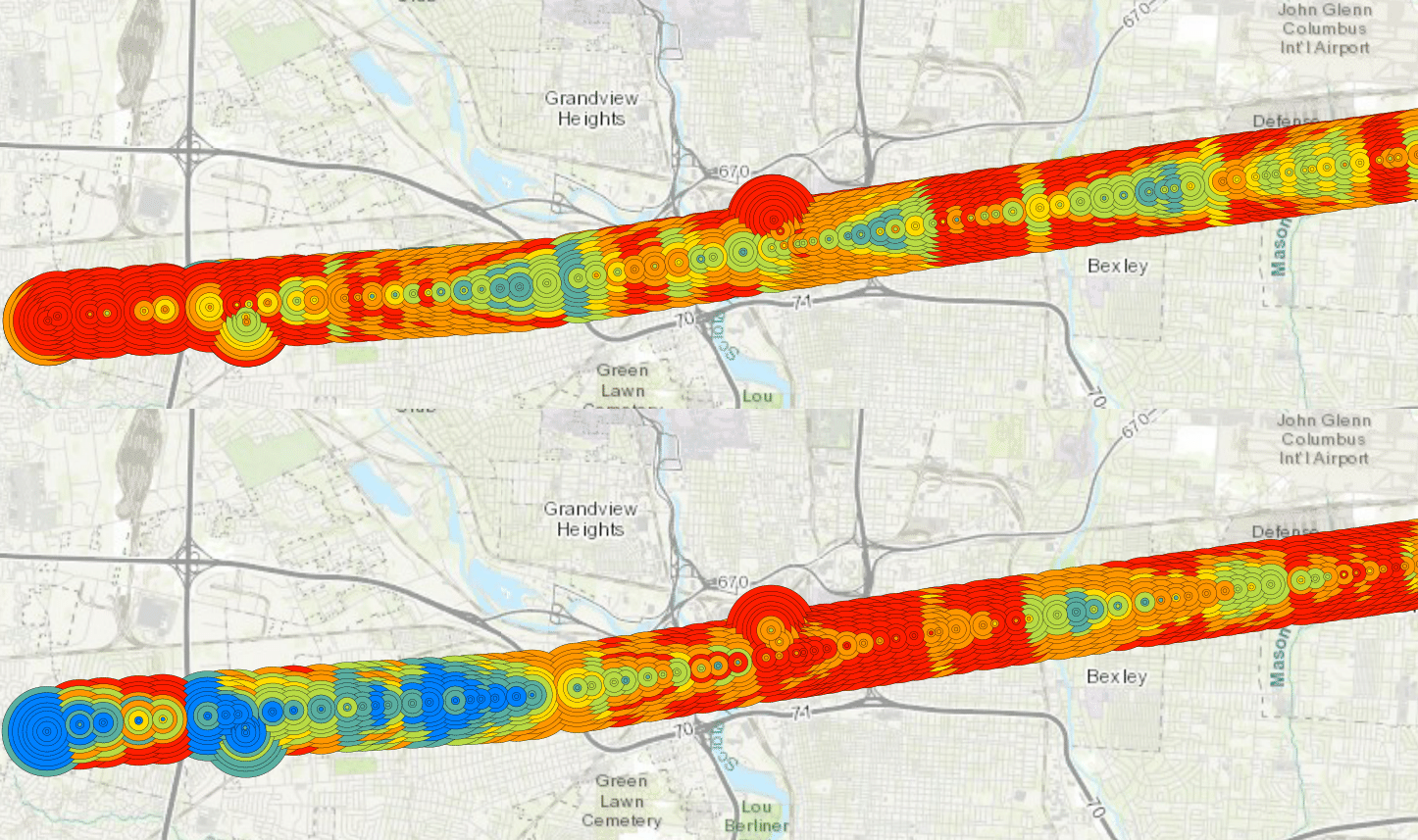


Figure 4 The average waiting time difference between RTA (PR IB=2min) and non-RTA users (NR) on COTA bus route No. 10.

Figure 4 shows a cl

1. Conclusion

Reference:

Beirão, G., & Cabral, J. A. S. (2007). Understanding attitudes towards public transport and private car: A qualitative study. *Transport Policy*, *14*(6), 478–489.

Bradburn, N. M., Sudman, S., Blair, E., Locander, W., Miles, C., Singer, E., & Stocking, C. (1979). *Improving interview method and questionnaire design: Response effects to threatening questions in survey research*. Jossey-Bass San Francisco.

Brakewood, C., Barbeau, S., & Watkins, K. (2014). An experiment evaluating the impacts of real-time transit information on bus riders in Tampa, Florida. *Transportation Research Part A: Policy and Practice*, *69*, 409–422.

Brakewood, C., Macfarlane, G. S., & Watkins, K. (2015). The impact of real-time information on bus ridership in New York City. *Transportation Research Part C: Emerging Technologies*, *53*, 59–75.

Brakewood, C., & Watkins, K. (2018). A literature review of the passenger benefits of real-time transit information. *Transport Reviews*, 1–30.

Cats, O., & Gkioulou, Z. (2017). Modeling the impacts of public transport reliability and travel information on passengers’ waiting-time uncertainty. *EURO Journal on Transportation and Logistics*, *6*(3), 247–270.

Chow, W., Block-Schachter, D., & Hickey, S. (2014). Impacts of real-time passenger information signs in rail stations at the Massachusetts Bay Transportation Authority. *Transportation Research Record*, *2419*(1), 1–10.

Dziekan, K., & Vermeulen, A. (2006). Psychological effects of and design preferences for real-time information displays. *Journal of Public Transportation*, *9*(1), 1.

Fan, Y., Guthrie, A., & Levinson, D. (2016). Waiting time perceptions at transit stops and stations: Effects of basic amenities, gender, and security. *Transportation Research Part A: Policy and Practice*, *88*, 251–264.

Ferris, B., Watkins, K., & Borning, A. (2010). OneBusAway: results from providing real-time arrival information for public transit. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1807–1816). ACM.

Fries, R. N., Dunning, A. E., & Chowdhury, M. A. (2011). University traveler value of potential real-time transit information. *Journal of Public Transportation*, *14*(2), 2.

Gkioulou, Z. (2013). Evaluating the impact of waiting time uncertainty on passengers´ decisions.

Google Developers. (2016). GTFS Static Overview | Static Transit | Google Developers. Retrieved March 8, 2018, from https://developers.google.com/transit/gtfs/

Google Developers. (2018). Trip Updates. Retrieved April 8, 2019, from https://developers.google.com/transit/gtfs-realtime/guides/trip-updates

Goyder, J. (1986). Surveys on surveys: Limitations and potentialities. *Public Opinion Quarterly*, *50*(1), 27–41.

Liu, Y., Shi, J., & Jian, M. (2017). Understanding visitors’ responses to intelligent transportation system in a tourist city with a mixed ranked logit model. *Journal of Advanced Transportation*, *2017*.

Neuman, W. L., & Robson, K. (2014). *Basics of social research*. Pearson Canada Toronto.

Papangelis, K., Nelson, J. D., Sripada, S., & Beecroft, M. (2016). The effects of mobile real-time information on rural passengers. *Transportation Planning and Technology*, *39*(1), 97–114.

Reed, T. B. (1995). Reduction in the burden of waiting for public transit due to real-time schedule information: a conjoint analysis study. In *Pacific Rim TransTech Conference. 1995 Vehicle Navigation and Information Systems Conference Proceedings. 6th International VNIS. A Ride into the Future* (pp. 83–89). IEEE.

Rossi, P. H., Wright, J. D., & Anderson, A. B. (2013). *Handbook of survey research*. Academic Press.

Watkins, K. E., Ferris, B., Borning, A., Rutherford, G. S., & Layton, D. (2011). Where Is My Bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders. *Transportation Research Part A: Policy and Practice*, *45*(8), 839–848.

Wright, K. B. (2005). Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication*, *10*(3), JCMC1034.