Measuring the Impacts of COVID-19 Pandemic on Public Transit Demand in US

Luyu Liu

Abstract: TBD

Keywords: TBD

1. **Introduction**

COVID-19, a novel coronavirus disease, emerged in December 2019 to become a global health crisis due to its high contagiousness, fatality rate, and lack of known vaccines or treatments. To deal with the pandemic, from early February 2020 the U.S. Centers for Disease Control and Prevention (CDC) recommended social distancing, self-quarantine, and working from home to stop the spread of the virus; states and cities followed these and similar guidelines, closing schools and businesses, and issuing calls to stay-at-home. These sudden and unprecedented shutdowns led to declines in travel demands at all geographic scales and all modes (1,2).

Public transit systems in the US experienced dramatics declines in ridership due to COVID-19. In Washington DC, Metrorail ridership declined by 90% and bus ridership declined by 75% by the end of March 2020 (3). Smaller transit system also experienced major declines; for example, such as El Dorado Transit (California) experienced a ridership decline of 75% (4). The consequent drop in fare box revenue may lead to subsequent cuts in services, particularly since cash-strapped local governments may not have the ability to increase their support. The decline in ridership is unequal across social dimensions since many information, managerial, tech and knowledge workers can telecommute while people with jobs that demand physical presence still need to travel to work (5). The remaining public transit users during a pandemic such as COVID-19 are likely captive (6) riders who depend on public transit for mobility and accessibility to jobs, health care and services. Since only essential businesses and services were open during this period, these captive riders were also likely performing necessary activities for themselves or society, highlighting the nature of public transit as a critical infrastructure (7).

In addition to the closing of business and telework, another factor affecting the decline of public transit demand during a pandemic is fear. According to an online survey, about 48% of Americans and 40% of Canadians feel that riding transit poses a high health risk due to the coronavirus (8). In an analysis of public transit ridership in Taipei during the 2003 SARS pandemic, Wang found an immediate loss of 1200 ridership and demonstrated that almost 50% of daily ridership was lost during the peak of the SARS pandemic (9). The analysis modeled the fear of using transit into two parts: *fresh fear,* which was generated by the increasing daily confirmed patient’s number; and *residual fear,* which gradually decays to the following days. Kim et al. analyzes Seoul transit system smart card transaction data during the 2015 MERS outbreak (10). They note variations in the decline in trip frequencies for different public transit mode, different populations, and different traffic analysis zone in Seoul. They find that while fear of the pandemic significantly influenced travel behavior, there were social differences in the ability for people to change their daily routine, measured by neighborhood land value.

COVID-19 provides an unfortunate but imperative opportunity to understand the differential impacts of a major shock such as a pandemic on public transit travel demand. We use the data obtained from Transit, a widely used mobile phone-based transit planning app, and conduct comprehensive analyses of the impacts of COVID-19 on US public transit systems. We fit logistic curves to describe the decline in daily transit demand across public transit systems, extracting key parameters: i) *floor value*, the apparent minimal level of demand; ii) *decay rate* and *decay duration*, representing the speed and temporal extent of the demand decline; and iii) *divergent and convergent points*, representing the initial date when transit demand began and the final date when decline decreased., respectively. We conduct regression and correlation analyses relating the floor values and decay rates to social-economic and demographic factors in each community. We also compare the distance between the divergent/convergent points and the first date of local community spread to measure whether public transit users in different metro areas reacted at different speeds to the unfolding pandemic. Finally, we use hourly transit demand data to capture COVID-19’s impact on daily patterns of transit demand; we measure the similarity of hourly demand profile during the COVID-19 pandemic compared to one year earlier. In the end, we conclude that COVID-19 had major impacts on the transit system in different dimensions and demonstrate the social equity issue of transit usage during the pandemic; we propose some future directions for transit studies in the context of COVID-19.

1. **Data and methods**
   1. **Data sources**

**Transit demand.** Since it is difficult to obtain comprehensive ridership data is difficult to obtain, we use data from the Transit mobile phone app (transitapp.com) as an indicator of changes in daily and hourly transit demand. Transit is a popular mobile phone app providing real-time public transit data and trip planning. The app covers over 200 cities around the world with more than 5 million download on Android platform (11). We treat app usage as a measure of real-time demand and an approximation of general public transit demand (12); we also examine this empirically (below). The data provided by Transit via their daily updated webpage are change values expressed as a set of percentage of app usage relative to the same date last year, adjusted for annual growth (12).

To validate the authenticity of the Transit app usage data as a measure of transit demand, we compare ridership decrease reports derived from transit systems' websites and local news outlets. Most transit system do not release estimates for every date; instead, many report single estimates for a given date. We compare these ridership decrease reports with the corresponding estimates from the Transit app data on the same date for 40 transit systems that we could trace the actual ridership decrease value. The average difference between the two estimates is 3.7%; a paired T-test indicates that we cannot reject the null hypothesis of the two means being equal results (p = 0.14 > 0.05). However, it is worth noting that the standard deviation is 15.96%; this may be due to the varying definitions of normal ridership level among agencies. Although the test suggests the Transit app data is a good overall approximation of public transit demand overall, we view the transit system-level analysis in this paper as tentative and worthy of focused investigation using ridership data It is also worth noting that we do not know the sampling frame of Transit app users: this user base will not include individuals who cannot afford a smart phone and data plan, cannot use the app due to different abilities, or choose not to use it However, these disadvantages are compensated by the large Transit app user base that allows comparison across transit systems.

The daily Transit data includes demand decreases estimates for 182 public transit systems across the United States, Canada, Australia, New Zealand, and France. We select 113 county-level transit systems in 63 metro areas, 52 counties, and 28 states across the United States. We exclude 7 state-level or cross-counties systems if their ridership could draw from large and geographically diverse areas, such as Pacific Surfliner, which extended to the whole South California coast and Metro-North Railroad, which crosses multiple counties and states. Daily data’s time period is from February 15th to May 17th. We also use hourly demand decrease for 93 public transit systems across the United States. Hourly data’s time period is from March 16th to My 17th.

**COVID-19 case numbers.** We collected the daily case numbers for each county from the COVID-19 Surveillance Dashboard produced by University of Virginia (13), COVID-19 Dashboard produced by John Hopkins University (14), and COVID-19 maps and county-level dataset produced by USAFacts (15). The data includes all counties’ confirmed cases in the US for every day. To find the linkage between the case numbers and the demand decrease, we geocoded each transit system to the corresponding county or county equivalent unit.

* 1. **Logistic model for daily transit demand change.**

Based on a visual examination of the data, we note a pattern of stable demand before the COVID-19 crisis, a period of decline, followed by re-stabilization at a lower demand level. This is a pattern described well as a logistic (anti-) growth process, expressed using a logistic or sigmoid function:

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

where is the minimum value for the decline, is the pre-COVID baseline value, *k* is the decline rate; *t* is time (day) and is the time when the function reaches the midpoint. We fit each transit system’s demand data using logistic model individually and calculate key parameters that describe the process for each transit system

**Floor value.** The parameter *B* represents how far the curve decreases from the baseline as demand re-stabilizes at a lower level; we call this the *floor value.* This represents the ratio of public transit users in the system that still would not or cannot stop needing it regardless of the pandemic. This demand level is not necessarily a persistent state: demand may destabilize and grow again due to external factors, such as re-opening of businesses or stay-at-home fatigue. The floor value represents a base level from the initial shock to the system.

We examine relationships between the estimated floor values and social-economic factors using linear regression analysis between the floor values for different systems and the social-economic data from it the corresponding US county/county-equivalent. The county-level social-economic data are from the latest American Community Survey (ACS) 5-year estimate table (2014 - 2018). We derive several socio-economic indicators. First is the *ratio of population with non-physical occupations*. Similar to *life fixity* (10), this measures the population’s degree of freedom to change the routine of their daily life: it represents how many people can work from home thus avoid regular transit commuting to reduce contagion risk. If a community has higher ratio of non-physical jobs, more workers may work from home, meaning that transit demand will decrease more. We use the occupations statistics for employed civilian population 16 years and over from the ACS data. The table contains the number of employed population in different industries. According to the statistics released by US Bureau of Labor Statistics, Information, Financial activities, and professional and business service have the highest percent who can work from home (5,16). Among these occupation categories in the ACS table, we assign “management, business, and financial operations occupations”, “professional and related occupations”, and “office and administrative support occupations” as the occupations with which people can work from home remotely.

*Income* is another social-economic factors that relates to job composition. Also, transit users tend to skew toward lower incomes in the United States. We use the median income data from ACS.

A third set of indicators is the *ratios of minority and female population*, including African American and Hispanic populations. Many studies has demonstrated the disadvantaged status of African American, female, and Hispanic population in mobility and job accessibility across different metro areas in the United States (17–19). Therefore, it is necessary to investigate the relationship between floor value and each vulnerable group’s ratio. We derive the sex, race, and Hispanic population data from the ACS data.

A fourth socioeconomic indicator is community *age structure*. Older individuals are more at risk of hospitalization and death from COVID-19 (20). We use 45 years old as a threshold to identify high-risk populations. We measure the ratio of people with age over 45 from the ACS 5-year estimates (2014 - 2018) data.

Finally, we use measures of *awareness*. If local residents are more aware of the COVID-19 and its risk, the floor value may be lower because more people will try to avoid public transit trips. To test this, we use the Google search trend index to represent the awareness of the local people (21–23). We collected the average Google search trend data for different designated market area that each transit system locates in for 90 days from January 18th to April 17th 2020 (24), the latter being the latest day we witnessed any system experiencing further decline. We select “Coronavirus” over “COVID19” as the search keyword for its popularity among the public (25).

To supplement the analysis, we also refer to the user survey results conducted by Transit app about the demography of the passengers during the pandemic. The survey was conducted in early April, 2020 across the United States (n = 15000) and Canada (n = 10000) via the Transit app interface. The survey investigated the age, race (including Spanish speakers), gender, trip purpose, occupation composition of the passengers who stick to transit systems during the pandemic (12,26,27). The user surveys provide first-hand proofs about the demography of the essential passengers and are a very good complement to the correlation conclusions.

**Cliff point and floor point.** To measure when the demand started to decrease and eventually re-stabilized at a lower level, we introduce two measures: the *cliff point*  and *floor point* . We apply confidence interval theory to derive these measures by first construct the probability density function of the normalized logistic function F(x):

|  |  |  |
| --- | --- | --- |
|  |  | (2) |
|  | , and | (3) |

where *P* is the probability density function of the normalized logistic function; to normalize the logistic function, we subtract the baseline and divide the result by *B* to construct the *P* function so that . is the confidence level. From equations (2) and (3), we can see that ; α = 0.05 ensures that the cliff and floor points demarcate the 95% of the decline. From the formula, we give the direct definition of cliff and floor point:

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

By defining the two measures with parameters in the logistic model as in formula (2), (3), and (4), they are not directly calculated from the observed data but the smoothed curve. They can be a stable indicator to measure the start and end of the demand decrease.

Beyond its statistical meaning, cliff point represents the first daywhen the demand curve began to diverge from the normality. It also represents when the transit users start to act to avoid the transit trips. The floor point is the day when decline slows and transit demand has re-stabilized.

**Decay rate and duration.** The decay rate*k* represents the rate of transit demand decline. This can indicate the speed of response from users who have the ability to stay at home or not use public transit. This by itself does not have any physical meaning; instead, we use a temporal indicator to represent the speed and duration of the decline. From the logistic model, represents the time to the curve reaches the midpoint of the curve. It reflects two major temporal factors for the pandemic: the initial date of epidemic outbreak, which is measured by the divergent point, and the duration of decay. Therefore, we introduce decay duration:

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

Decay duration is determined only by the decay rate.

Similarly, we conduct linear regression analysis and correlation analyses between the decay rate and different social-economic factors, divergent point, and people’s awareness.

* 1. **Response intervals with incubation lags**

The transit demand dynamics of decline and re-stabilization may not be in synch with the dynamics of pandemic spread in a community, in particular, the initiation of this spread. We measure the differences between the date of the first confirmed case and the cliff point/floor points. The reason of using both cliff and floor points can be justified by an analogy: public transit systems during the pandemic ensemble different houses on fire. When the fire department investigates whether different house units can timely react to the fire, they first need to compare when the fire happened (first community spread) and the time when the very first few people began to evacuate due to the fire alarm (cliff point). Moreover, it is also necessary to find out when the very last few people managed to evacuate (floor point) when the fire happened. The most ideal scenario is the floor point is before the local spread began, just like most people already evacuated the house when the fire happened; the worst scenario is the cliff point is after the local spread began, just like no people haven’t started to evacuate when the fire happened.

We also consider the incubation period for the disease due to the implication for community spread. The median of incubation period is 5 days and can be as long as 14 days (28); the virus can spread asymptomatically (29–31). Therefore, the actual initial date when the virus began to spread in the local community can be traced back to 4 - 14 days ago. The incubation lag can be even longer considering the lack of testing kits and slow response for the local authority (32,33). Since the exact incubation period is unclear, we introduce an incubation lag parameter into response intervals relative to the cliff and floor points:

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

where: is the date of first confirmed case in the county of the transit system; l is the incubation lag; and are the cliff and floor point. Positive response interval means that the transit users responded earlier than the epidemic spread; the larger the value is, the less risk the transit users are exposed to the virus. For the two versions, the divergent version focuses on the transit users’ awareness while the convergent version focuses on both the awareness and the response speed.

* 1. **Change in daily transit demand dynamics**

The analyses based on the average daily transit demand shows the coarse-grained temporal variation among different transit system. It is also possible for the demand dynamics within each day to change because of a pandemic. For example, Figure 1 below shows changes in transit demand by hour before (blue) and during COVID (orange) for the New York City subway. The normal curve is a typical US daily travel demand pattern, with morning and afternoon peak demand periods corresponding to commuting to and from work, respectively. In contrast, the COVID demand curve has relatively modest morning and afternoon peaks.

We use two methods to analyze these changes, namely, shape similarity and peak analysis.

Figure 1: MTA New York City Subway hourly transit demand curves (blue: normal curve; orange: COVID curve)

We utilize the shape analysis technique of *ordinary Procrustes analysis* to measure differences between two curves, for example, hourly travel demand during and before the COVID pandemic. This involves superimposing the two curves: in traditional Procrustes analysis for arbitrary shapes, the superimposition process includes panning, scaling, and rotating to make the shapes fit (34). We only use scaling in our application since these are one-dimensional curves. We use a stretch factor as a multiplier on one curve to fit the other curve so that their Procrustes distance is minimized:

|  |  |  |
| --- | --- | --- |
|  | Minimize: | (7) |

where: is the Procrustes distance between the two curves, is the stretch factor, is the number of data points in the dataset, and are the two curves’ value at time . The solution to this optimization problem is:

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



The optimal Procrustes distance is a measure of the difference in the shapes of the curves: a larger Procrustes distance means bigger differences in the shape of analysis between the two curves.

In our analysis, we compare during and before COVID hourly demand pattern to produce the Procrustes distance for a specified day and calculate the average from March 16th to May 17th. We also calculate the average Procrustes distance between Wednesdays (representing typical weekday) and Saturdays (representing typical weekend) demand profiles in each week. Under normal travel demand patterns in the US, weekday and weekend hourly demand profiles are different, with no sharp demand peaks on weekend days. We wish to see if weekday and weekend public transit demand profiles have converged during COVID.

1. **Results**
   1. **Floor values**

Figure 2 maps floor values for the US public transit systems in our study. We can see clear geographic differences: cities in the Deep South and Midwest have higher floor values. Meanwhile, high tech cities such as the San Francisco Bay area and university cities such as Ithaca, Ann Arbor, and Madison generally have a very low floor value.

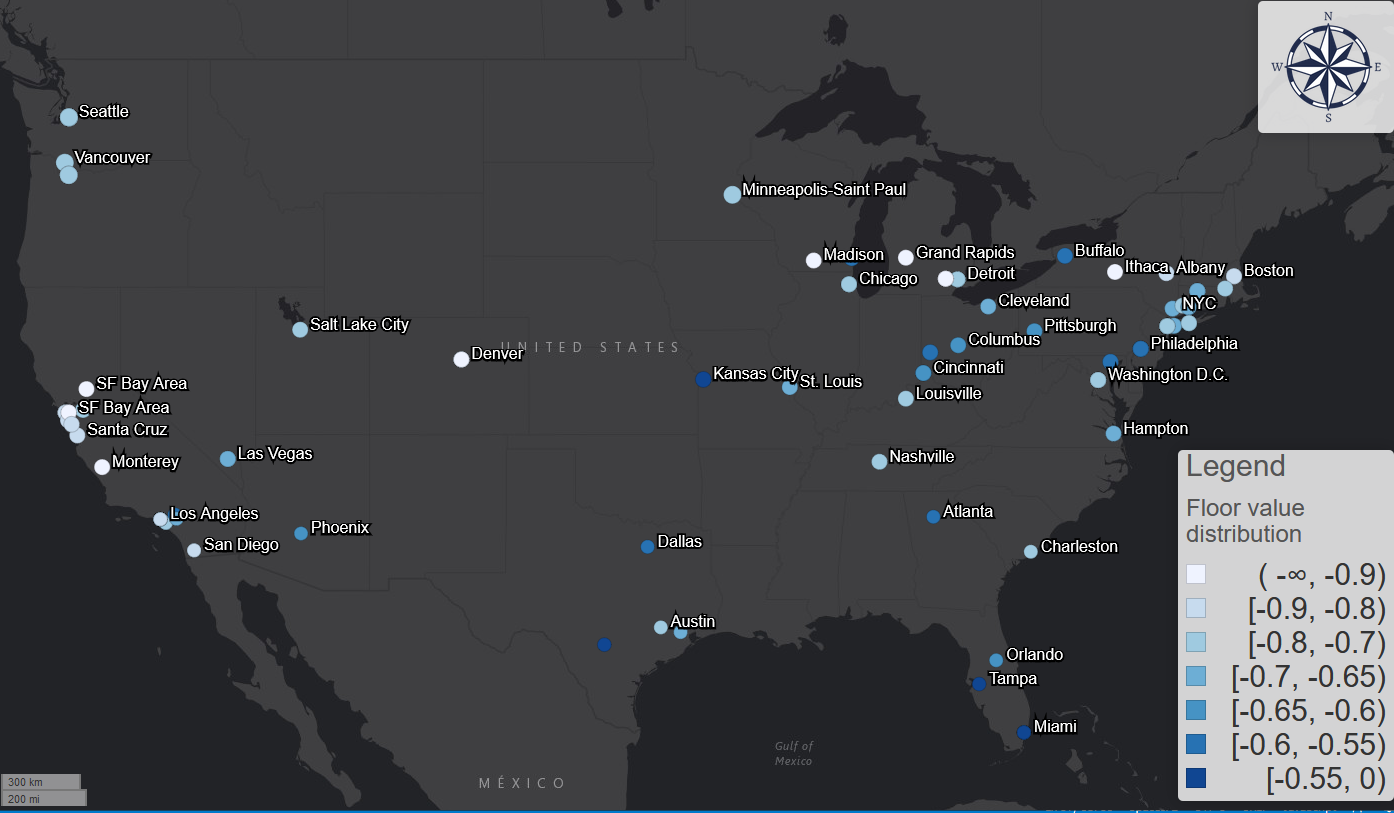


Figure 2: The distribution of floor value across the United States.

Table 1 provides results from the regression analysis related the floor values across transit systems with socio-economic and awareness indicators in each community. Four indicators are significant with p-value smaller than 0.05; the F test also shows the model is significant with p-value of 1.41e-10. We did not include Hispanic population ratio or median income in the final model because of multicollinearity with the ratio of people with non-physical occupations. The R-squared value is 0.38. A residuals assessment shows that the residuals are normally distributed and there are no lingering multicollinearity effect and no leverage effect.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Estimate coefficient | Standard Error | t value | Pr(>|t|) | Variance Inflation Factor | R-squared increment |
| Intercept | -0.52997 | 0.169042 | -3.135 | 0.00221 | - | - |
| Ratio of people with non-physical occupation | -0.4269 | 0.129248 | -3.303 | 0.0013 | 1.21 | 0.063 |
| Ratio of African American population | 0.412221 | 0.075941 | 5.428 | 3.53E-07 | 1.08 | 0.17 |
| Ratio of population over 45 years old | 0.856343 | 0.264103 | 3.242 | 0.00158 | 1.05 | 0.061 |
| Google search trend index | -0.00502 | 0.00197 | -2.55 | 0.01217 | 1.21 | 0.037 |

Table 1: Results from regression analysis of floor values with socio-economic and awareness indicators

**Population with non-physical occupations**. The results confirm the hypothesis that greater transit demand decreases associate with higher percentage of people with non-physical occupations. People who can work at home avoid public transit; people who cannot work at home and rely on public transit continue to use it.

Although we did not include the Hispanic population indicator due to multicollinearity, a very significant negative correlation between Hispanic population and population with non-physical occupations suggests the vulnerable of Hispanic populations during this health crisis: if a city has a higher Hispanic population, it is likely for the city to have a higher floor value, which means more people will not work from home during the pandemic. This is also consistent with the occupation statistics: according to the labor force characteristics survey made by US Bureau of Labor Statistics, Hispanic population has the lowest percent (22%) of management, professional, and related occupation compared with White (41%), African American (31%), and Asian people (54%) in 2018 (35).

These results are also proven by the user surveys conducted by Transit app in April 2020. According to the survey results, 92% of all the surveyed users reported that they still use public transit to commute to work (26). Meanwhile, the top-5 occupation categories that are most likely to work from home are computer and mathematical; life, physical and social science; education, training and library; architecture and engineering; and legal. Although the categorizations of the survey and the ACS data are different, this is generally consistent with the non-physical occupation categories we derived from that data.

The Transit survey also indicates that Spanish speakers are more likely to continue using the Transit app for trip planning purposes: English-language users dropped 71% from early February while Spanish-language users dropped by 50% over the same time period (27). The income correlation is also confirmed by the survey results: compared with the survey results conducted by American Public Transportation Association (APTA) in 2017, active users skew towards lower income brackets during the pandemic, especially for those whose annual income is less than $15000 (27). The survey results provide a first-hand proof and reaffirm the correlation results about the vulnerability of Hispanic population and low-income population.

**Age.** The ratio of the population over 45 years old also associates with higher floor values. This is ironic and alarming since senior people are the most vulnerable. This results is also supported by the Transit user survey. By comparing the users age composition survey conducted in September 2019 and April 2020, young people under 18 and between 25 to 44 years old tend to stay in quarantine across the US; meanwhile, the relative ratio of people between 45 to 64 years old has doubled (27).

**African American.** The regression analyses also suggest the vulnerability of African American population. It the most significant among all the other factors in Table 1 based on R-squared increment. There is a strong correlation between higher ratio of African American population and higher floor value. These results are also consistent with the results of the user survey. During the pandemic, African American people have the greatest share (>35%) of riders compared with other races in the US, while Caucasian were the majority (>40%) of the rider before the pandemic according to the APTA survey in 2017 (26). The disproportionate decrease of African American population’s transit demand supports the conclusion that cities with more African American population are more likely to have a higher floor value.

Higher floor values are also highly correlated with larger ratio of female population, however, we also do not include it due to multicollinearity with African American population ratio. Higher ratio of female population is also correlated with lower income and lower ratio of people with non-physical occupations. These results indicates the disadvantage of female population in different social-economic aspects. The Transit user survey supports these results in a dramatic manner. Among all the US users surveyed, the male and female proportions were roughly equal before the COVID-19 pandemic; during the pandemic, 56% are females while only 40% are males (26). For some cities such as Philadelphia, more than 68% of riders are women. Meanwhile, Transit app users of color are also more likely to be females during the pandemic; more than 70% of the African-American riders during the pandemic are female (26).

**Awareness.** Google search trend index is significantly associated with the floor value; cities with higher search index tend to have lower floor value, as higher search index means COVID19 has higher ratio among all the things they search. This indicates some effects of people’s awareness on their behaviors: with people cares more or more people care about the evolution and risk of the disease, more people will try to follow the stay-at-home order and avoid unessential public transit trips. However, according to the R-squared increase table in Table 1, Google search trend index has the lowest R-squared increase. It shows that people’s race, job, and age composition outweigh the awareness or preference when it comes to whether people will stay at home.

* 1. **Response intervals with incubation lags**

Figure 3 shows the distribution of the response interval measures in the US relative to the cliff and floor time points given incubation lags of 0, 5, 14 days. For response intervals from the cliff point: with an incubation lag of zero, the pattern is highly polarized. In some cities with international airports, such as Seattle where the first US COVID-19 cases were found, people still used the transit even after the first case emerged. Meanwhile in other cities, such as most cities in Midwest with the exception of Chicago, people started avoiding transit trips in advance of confirmed community spread. This may be due to Seattle's precedence in COVID-19 spread in America: the media began to report the severity of this disease and CDC made the prediction that the community spread is inevitable near the end of February 2020 (36).

The response interval patterns with zero incubation lag suggest that initial declines in public transit usage may have limited the spread of the disease in some communities. However, the picture is less sanguine after we factor in incubation periods. With lag = 0, 5 of 13 systems have positive response intervals, suggesting declines in transit usage in advance of community spread. With lag = 5, all of 13 transit systems have negative response intervals, meaning the virus could have been spreading in the community before any appreciable decline in transit demand. In contrast, most transit systems in the Midwest such as Missouri, Ohio, Michigan, and Kentucky still have positive response intervals with an incubation lag of five days. This is supported by the cellphone location data: those Midwest states above had known stay-at-home orders before March 27th and the measured trips are significantly less (37). For a more conservative scenario of an incubation lag = 14 days, most transit systems have negative response intervals.

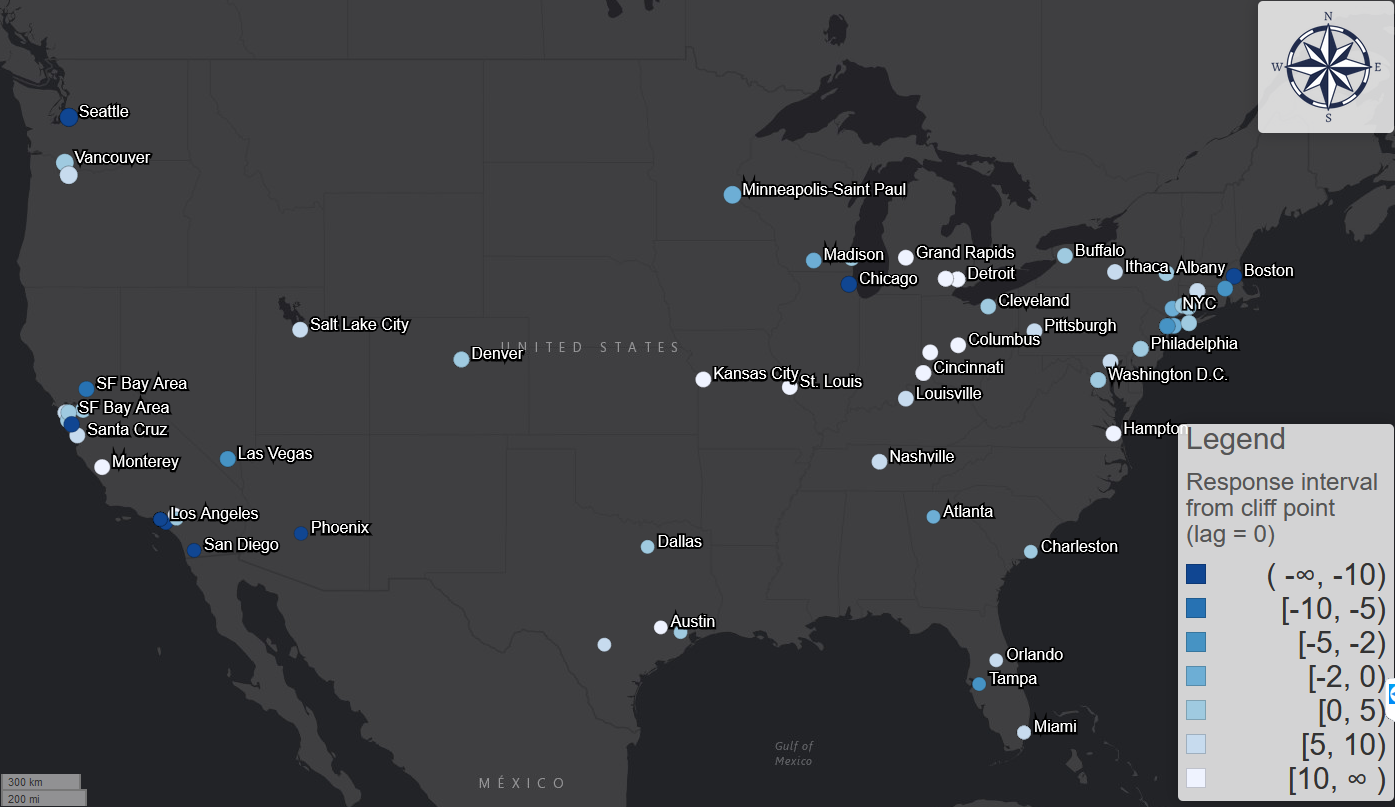


Figure 3: the geographic pattern of response interval with incubation lag =0 from cliff point

However, the geographic pattern is highly homogeneous for the response intervals from convergent point, which represent how earlier is each transit system’s users finished the stay-at-home process compared to the community spread. For scenario of lag = 0, only Capital Metro in Austin, Texas has positive response interval. For the case of Austin, the demand decrease started at March 5th and finished at March 23nd; the first case was confirmed at March 25th. However, long before the first confirmed case, the city and county authority declared the local state of emergency in March 6th (38), which is one of the earliest places to take actions in the South, even in the United States. The divergent point is also the same as the date of local state of emergency, which suggests the effectiveness of the executive order. This can be one reason for the relatively fast and earlier reaction of transit users.

Figure 4 shows the trend of ratio of systems with positive response interval with respect to different incubation lag for both cliff and floor point. As the incubation lag increases, the ratio of systems with positive responses interval from cliff point will decrease from 61% (lag=0) to 33% (lag=5) and then to 6.5% (lag=14). the curve of response interval from floor point is generally a flat line of 0, suggesting that no cities will finish the decline process when the community spread happened.

Figure 4: Trend of the ratio of transit system with positive response interval for different incubation lag

* 1. **Decay duration**

Figure 5 shows the geographic pattern of decay duration. Cities in Northeastern US have smaller decay curation and Midwest have longer duration.

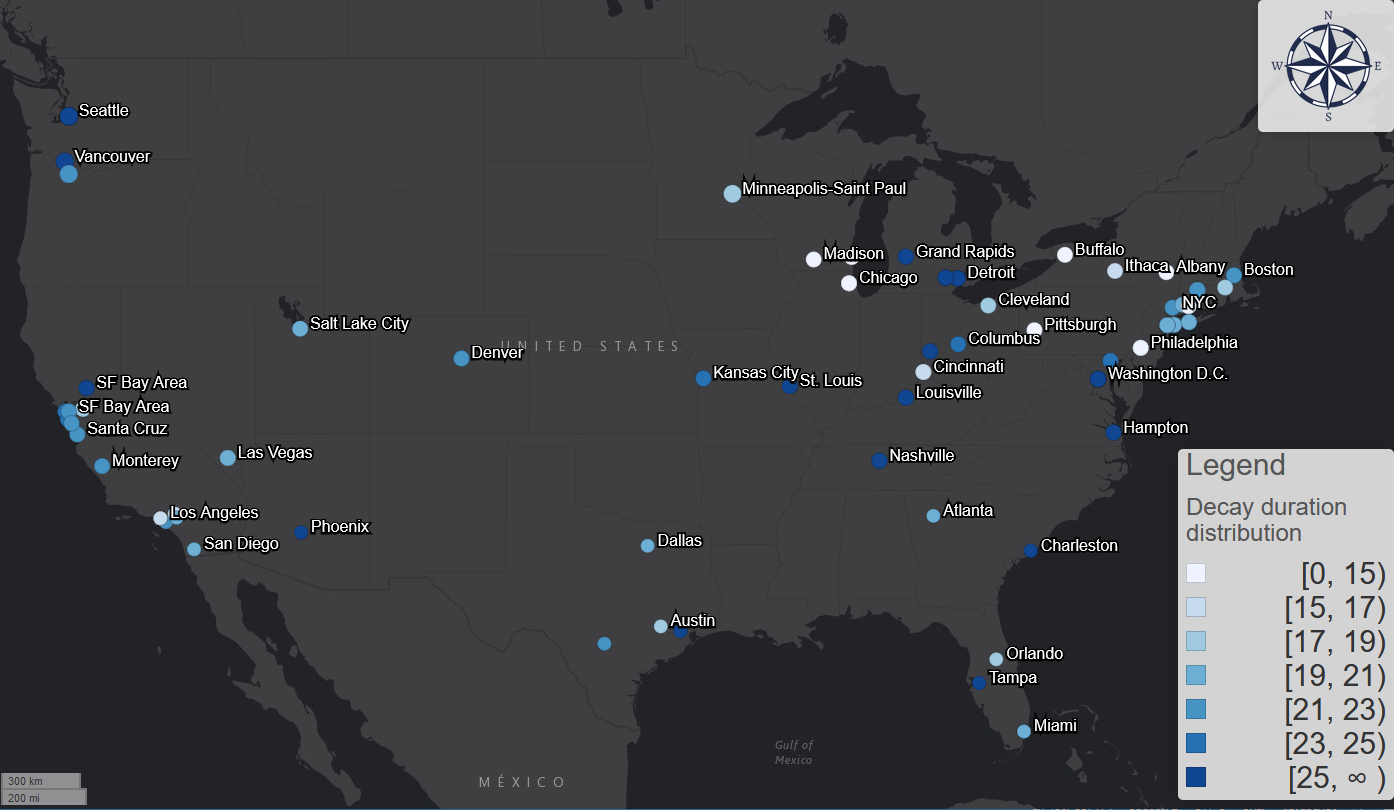


Figure 5: Geographic distribution of decay duration

Figure 5 shows the decay rate and the divergent point have a positive cubic correlation. This indicates that the later the demand decrease happened, the faster it is. This could be because the general transit passengers may be more aware of the risk of COVID-19 when more cases are reported nationally; the perceived fear grows higher as the time passed thus driving local population to act faster.

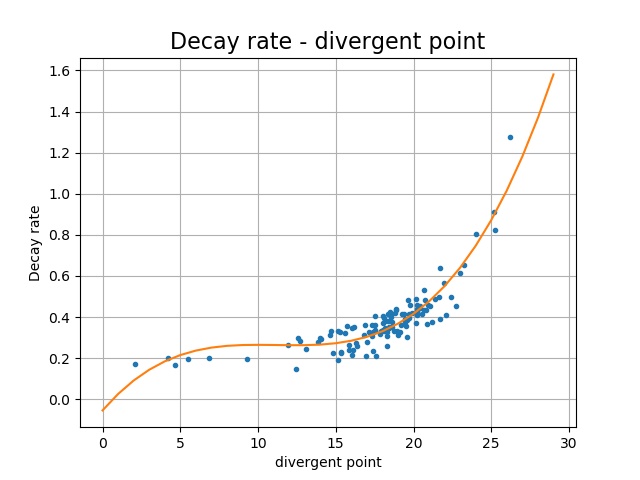


Figure 6: the relationship and a cubic polynomial fitting between decay rate and divergent point.

* 1. **Change in daily transit demand dynamics**

Figure 7 shows the geographic distribution of each transit system’s average Procrustes distance between its normal and pandemic hourly demand curves. This suggests dramatic shift in weekday demand patterns in non-physical occupation dominating areas like San Francisco Bay area, and University cities like Ithaca, Madison, and Ann Arbor. The Procrustes distance of traditional metro areas is low, meaning the hourly profile before and during the pandemic did not change much. These geographic patterns are similar to the geographic patterns of floor value. Figure 7 confirms the strong correlation between the Procrustes distance and floor values.

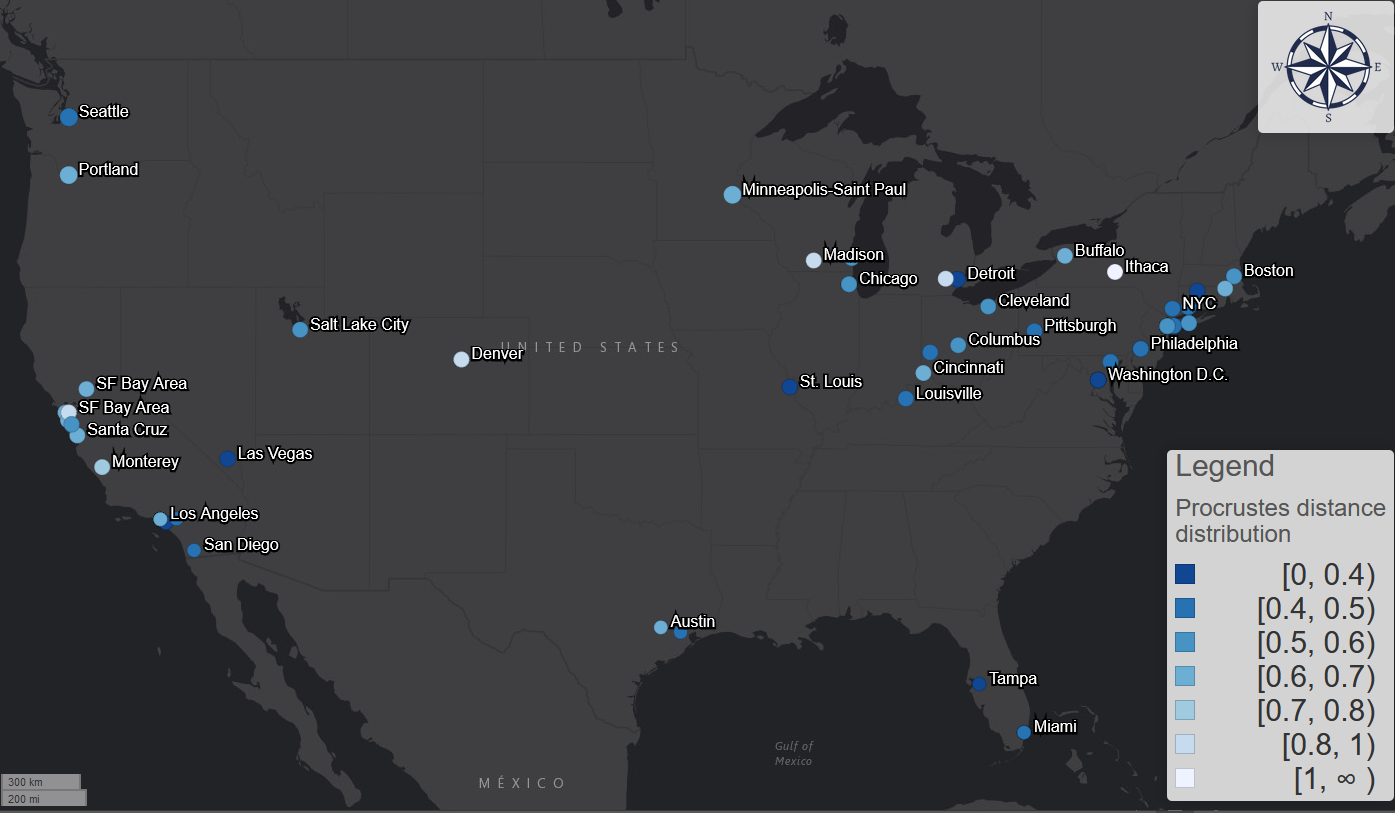


Figure 7: geographic distribution of each transit system's average Procrustes distance between normal and pandemic curves.

Figure 8: Relationship between average Procrustes distance and floor value.

Figure 9 shows the temporal distribution of all the transit systems’ average Procrustes distance (left side) and the stretch factor (right side) between its normal and actual hourly demand curves. The distance between current and expected demand steadily increased, which means the hourly transit dynamics was gradually diverging from normality.

The Procrustes distance value also shows a very regular periodical pattern for each week and the distance is much higher for weekdays than weekends, which means the shape diverged more for weekdays. This is also due to the halt of commuting activities. In this sense, during the pandemic, weekdays became more like weekends. All transit system’s Procrustes distance between weekdays and weekends also decreased during the pandemic, which moreover suggest that weekdays are more like weekends after the outbreak of the COVID-19 in terms of hourly demand dynamics. This converging process could be powered by two factors. First one is the disproportional sudden decrease of the morning and afternoon commuting activities in the weekdays. This change will generally flatten the peaks and diminish the contrast between normal hours and rush hours. This process is essentially driven by the privileged population with non-physical occupations: their leave made weekdays more like weekends. On the other hand, the reduction of unessential activities, such as leisure and shopping trips, also make commuting-relevant trips more prominent during the weekends. This effect is especially obvious in the New York City for its population highly relies on public transit and the non-physical occupation’s rate is not high. For example, for the Metropolitan Transportation Authority (MTA) systems, the curves of Sundays usually have one peak during 2 – 4 pm; however, the shape of the Sunday curves during the pandemic had two peaks, which was similar to the weekdays’ commuting pattern. This process is meanwhile driven by the underprivileged population that still have to work during the weekends: their stay made weekends more like weekdays. These two factors homogenized each day of week and make the boundary between weekends and weekdays less obvious.

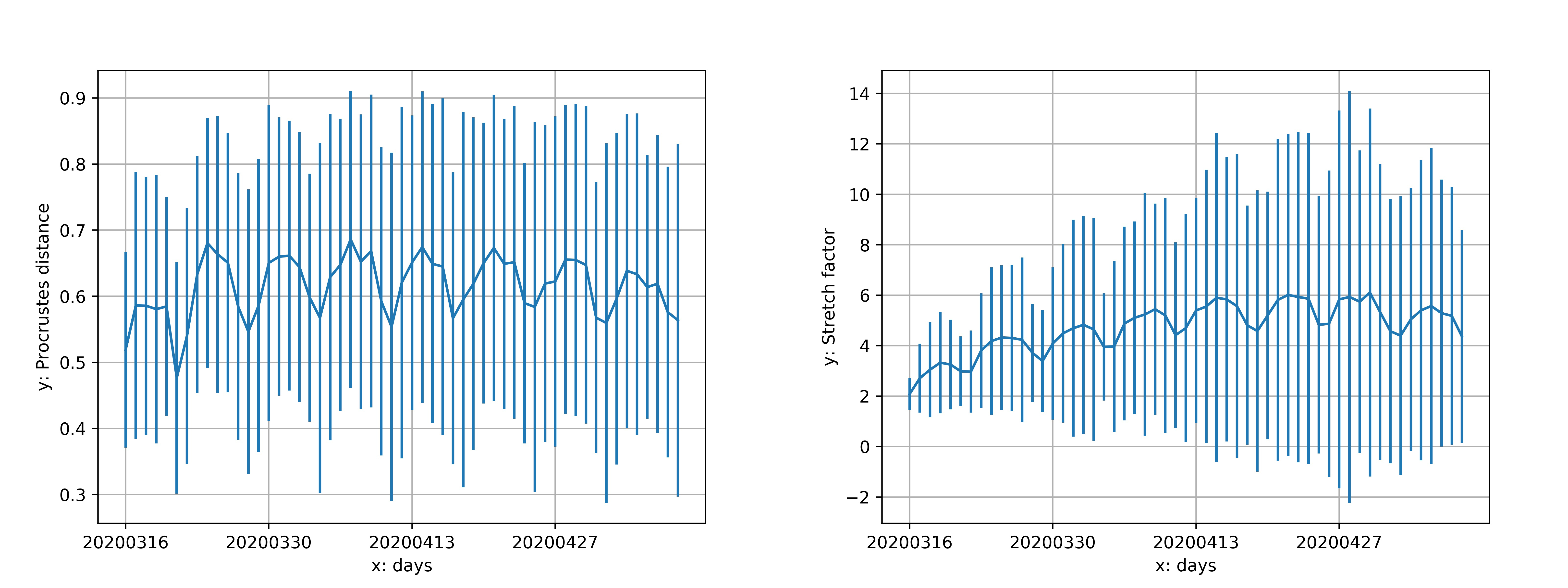


Figure 9: temporal distribution of all transit system's average Procrustes distance and stretch factor between normal and pandemic curves.

We also compare temporal shifts of the transit demand peaks before and during the pandemic. Figure 10 shows the geographic pattern of changes in morning peak (top) and afternoon peak (bottom). Figure 10 shows a polarized geographic distribution of the morning rush hour shift for all weekdays from March 16th to May 10th. The majority of US transit systems (59 systems out of 93) had a later morning peak. However, many transit systems such as the systems in Washington, New Jersey, Los Angeles, and New York City experienced an earlier morning peak. We find an average morning shift is -0.05 hours (3 minutes). For comparison, we calculated the corresponding shifts for other countries: this is -1 hours for UK, -1.8 hours for France, and -1.2 hours for Canada. Australia and New Zealand’s morning rush hours shifted 1.23 and 0.9 hours later, which makes US the only country that did not witness a significant change in morning peak during the pandemic.

The afternoon rush hour generally shifted earlier during the pandemic. Figure 10 (bottom) indicates a homogeneous geographic distribution of the morning rush hour shift for all weekdays from March 16th to May 10th. 76 out of 93 systems witnessed an earlier shift and the average shift is -0.55 hours. We observe similar pattern for most countries: while Canada’s shift is -0.55 hours, France’s shift is -1.78 hours, Australia’s shift is -0.55 hours, and New Zealand’s shift is -0.3 hours. UK is an exception with shift of 0.25 hours.

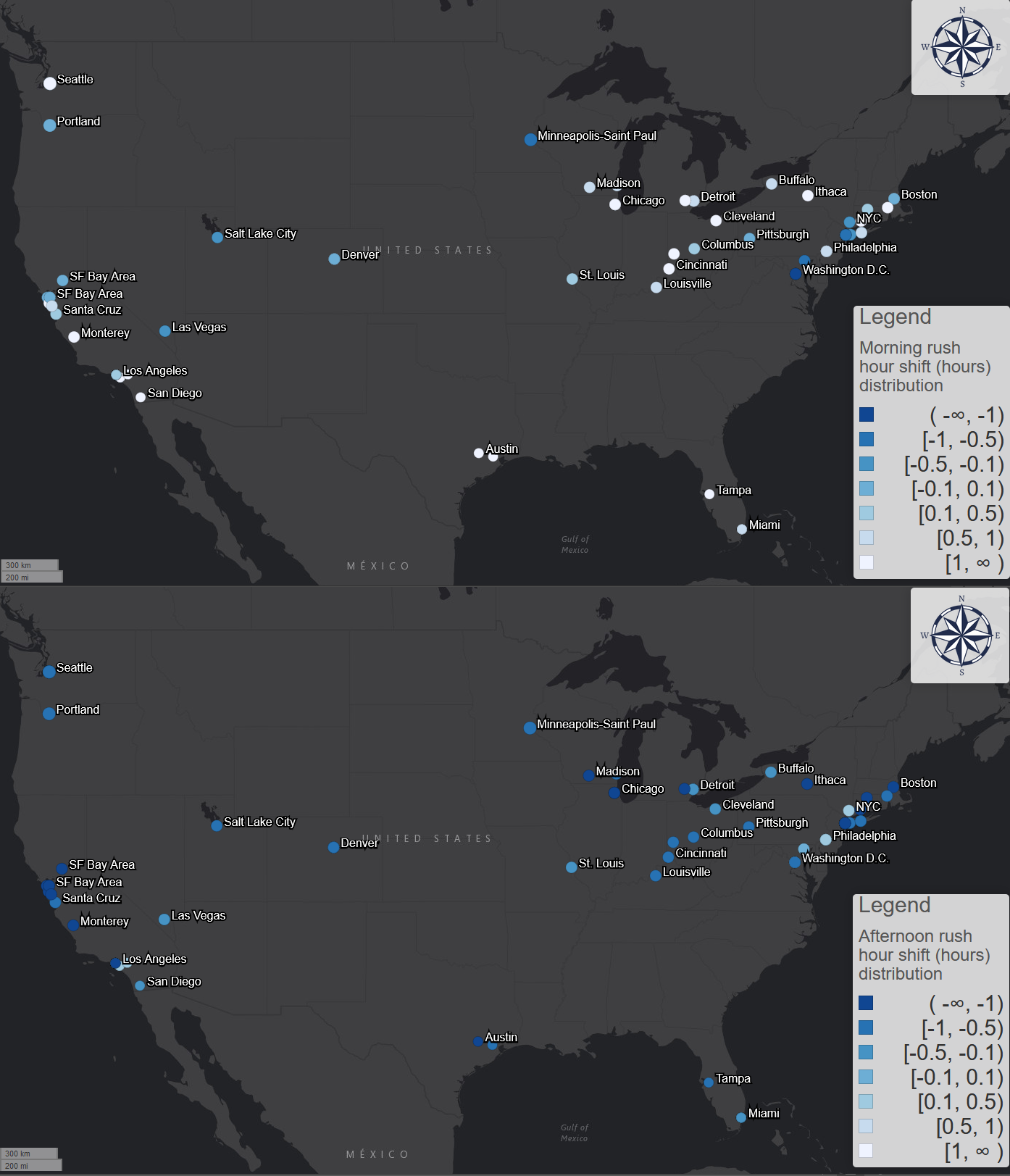


Figure 10: Shifts in morning (top) and afternoon (bottom) transit demand peaks or all weekdays

1. Conclusion

In this paper, we use smartphone app activity data to measure transit demand for 113 transit systems across the United States; We use logistic model to fit the daily changing trends of each transit system’s usage and introduce three important parameters to measure the three dimensions of the transit decline due to COVID-19: extent, start/end date, and speed of the decline.

The minimum value of the transit demand curve, *floor value*, shows the ratio of “essential passengers” who continue to use transit system regardless of the pandemic. Geographic patterns shows information industry dominating areas and university cities have lower floor value. Further linear regression with floor value and different social-economic and demographic factors shows that: cities with less non-physical occupations ratio generally have higher floor value; cities with larger ratio of minority population (African American, Female, Hispanic) tends to have higher floor value; cities with more middle-age and senior people over 45 years old tends to have higher floor value; and last, cities with lower “Coronavirus” relevant search trend index tend to have higher floor value. These conclusions are also supported by the user demographic survey results conducted by Transit app. These conclusions all affirm an alarming and gloomy fact: **cities with more essential workers and more vulnerable population tends to maintain higher transit usage rate during the pandemic**. This moreover suggests the necessity of the transit system even during the COVID-19 period when transit systems lost their most ridership: people still need transit systems and essential workers need them especially. It is a good lesson for transit planners and policy makers to rethink the role of transit systems not as a business, but as a social warfare to protect and serve the most urgent and the most vulnerable.

The start and end date of transit demand curve, *divergent point* and *convergent point*, shows when the people started and finished acting to avoid transit trips due to COVID-19. We moreover compare divergent/convergent point with the first day of local community spread and the results shows that people’s response time is not synchronized with the development of the disease. If not considering incubation period, 54% of all transit systems’ passengers can start to react to the pandemic. If considering the reported median incubation period of 5 days, the number drops to 24%; and if using the reported maximum incubation period of 14 days, the number moreover drops to 5%. Meanwhile, if comparing the convergent point with the first day of local community spread, almost no transit systems can act faster to escape most exposure to the spreading pandemic. This suggests that the reacting speed of the transit systems is much slower than the ideal scenario that most passengers can already avoid transit trips before the community spread started.

The speed of the decline, *decay rate and decay duration*, shows how fast and how long the decline process lasted. The correlation reveals that the later the decline happened, the faster the process would be. This could be because of growing awareness thus fear as time passed to make people to act faster.

The commuting analyses based on hourly transit demand data show that active essential passengers’ commuting routine during the pandemic is very different from the normal routine. We used ordinary Procrustes analysis to measure these differences between hourly curves in two dimensions: the impact on the hourly curves’ shape, measured by Procrustes distance, and the average decrease amount, measured by stretch factor. The pattern of Procrustes distance is geographically polarized and highly correlated with the floor value: information industries dominating areas and university cities shows significantly larger change during the pandemic. The impact on hourly curves’ shape was increasing as the pandemic developed. The pandemic also made the weekdays and weekends less different. Weekdays are more like weekends because disproportional decrease of the morning and afternoon commuting activities made the difference between rush hours and normal hours less obvious; meanwhile weekdays are more like weekends in some places like New York because the cessation of unessential businesses made the weekends trips become commuting-dominating.

Moreover, the pandemic shifted the morning and afternoon rush hours. Morning rush hour shifted in two directions but the afternoon rush hour shifted later for 1 hour in average. The interval between two rush hours generally became shorter, which could indicate the significant recession of the economy activities during the pandemic. However, in cities like the epicenter of New York, the interval between two rush hours became longer, which can be because of prolonged working hours of medical workers and essential workers.

The paper is a good start to investigate the impact of COVID-19 on transit systems across the United States and the world. However, there are still several limitations that we did or cannot cover in this paper:

* **Representativeness of the transit demand data for actual ridership.** Although we provided a bias test between some official ridership data and the transit demand data and the results show highly relevance, correlation, and the unbiasedness of the dataset, we still cannot argue that transit demand measures are exactly the same as the official ridership data for each day and each system. In the future, it would be better to use the official ridership data to measure the exact impact on the ridership.
* **Missing hourly systems.** In the hourly data, there are several missing transit systems especially in the Deep South, such as Atlanta.
* **The homogenizing of different cities.** The minimum resolution of the data is each system and its corresponding county-equivalent. This can be too vague for system-level analysis. To see more about how the pandemic impacted the local transit system, future studies could use surveys, local transit ridership data, and surveillance data with higher resolutions to detect more detailed patterns.

Reference:

1. Weber H. Air traffic data shows less crowded skies since the coronavirus spread [Internet]. 2020 [cited 2020 Apr 2]. Available from: https://www.fastcompany.com/90485186/its-time-to-cancel-rent

2. Bliss L. When the World Stops Moving [Internet]. 2020 [cited 2020 May 20]. Available from: https://www.citylab.com/transportation/2020/03/coronavirus-impact-public-transit-street-traffic-data-trains/607915/

3. WMATA. Metro and Covid-19: Steps we’ve taken [Internet]. 2020 [cited 2020 Apr 2]. Available from: https://www.wmata.com/service/status/details/COVID-19.cfm

4. Christensen K. COVID-19 drives down ridership as El Dorado Transit adapts [Internet]. 2020 [cited 2020 Feb 4]. Available from: https://www.mtdemocrat.com/news/covid-19-drives-down-ridership-as-el-dorado-transit-adapts/

5. Tan S, Fowers A, And DK, Tierney L. Amid the pandemic, public transit is highlighting inequalities in cities [Internet]. Washington Post. 2020 [cited 2020 May 16]. Available from: https://www.washingtonpost.com/nation/2020/05/15/amid-pandemic-public-transit-is-highlighting-inequalities-cities/?arc404=true

6. Zhao J, Webb V, Shah P. Customer loyalty differences between captive and choice transit riders. Transp Res Rec. 2014;2415(1):80–8.

7. Zimmerman R. Mass transit infrastructure and urban health. J Urban Heal. 2005;82(1):21–32.

8. Yellin J. North Americans Turning Away From Public Transit As Direct Result Of COVID-19 [Internet]. 2020 [cited 2020 Feb 4]. Available from: https://www.northstarhub.com/posts/north-americans-turning-away-from-public-transit-as-direct-result-of-covid-19

9. Wang K-Y. How change of public transportation usage reveals fear of the SARS virus in a city. PLoS One. 2014;9(3).

10. Kim C, Cheon SH, Choi K, Joh C-H, Lee H-J. Exposure to fear: Changes in travel behavior during MERS outbreak in Seoul. KSCE J Civ Eng. 2017;21(7):2888–95.

11. Transit app. Transit • Bus & Subway Times [Internet]. 2020 [cited 2020 Feb 4]. Available from: https://play.google.com/store/apps/details?id=com.thetransitapp.droid&hl=en\_US

12. Transit app. How coronavirus is disrupting public transit [Internet]. 2020. Available from: https://transitapp.com/coronavirus#monitor

13. Biocomplexity Institute U of V. COVID-19 Surveillance Dashboard [Internet]. 2020. Available from: https://nssac.bii.virginia.edu/covid-19/dashboard/

14. JHU CSSE. 2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository by Johns Hopkins CSSE [Internet]. 2020 [cited 2020 May 15]. Available from: https://github.com/CSSEGISandData/COVID-19/

15. USAFacts. Coronavirus Locations: COVID-19 Map by County and State [Internet]. 2020 [cited 2020 May 15]. Available from: https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/

16. U.S. Bureau of Labor Statistics. Table 1. Workers who could work at home, did work at home, and were paid for work at home, by selected characteristics, averages for the period 2017-2018 [Internet]. 2019 [cited 2020 May 16]. Available from: https://www.bls.gov/news.release/flex2.t01.htm

17. Cooke TJ. Geographic access to job opportunities and labor-force participation among women and African Americans in the greater Boston metropolitan area. Urban Geogr. 1997;18(3):213–27.

18. Golub A, Marcantonio RA, Sanchez TW. Race, space, and struggles for mobility: transportation impacts on African Americans in Oakland and the East Bay. Urban Geogr. 2013;34(5):699–728.

19. Iseki H, Taylor BD. The demographics of public transit subsidies: a case study of Los Angeles. 2010;

20. Centers for Disease Control and Prevention. Coronavirus 2019 (COVID-19) Surveillance [Internet]. 2020 [cited 2020 Apr 21]. Available from: https://www.cdc.gov/nchs/nvss/vsrr/covid19/index.htm

21. Lin Y-H, Liu C-H, Chiu Y-C. Google searches for the keywords of “wash hands” predict the speed of national spread of COVID-19 outbreak among 21 countries. Brain Behav Immun. 2020;

22. Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Eurosurveillance. 2020;25(10):2000199.

23. Yuan GX, Di L, Gu Y, Qian G, Qian X. The Prediction for the Outbreak of COVID-19 for 15 States in USA by Using Turning Phase Concepts as of April 10, 2020. Available SSRN 3575002. 2020;

24. Google. Google Trends - COVID19 [Internet]. 2020 [cited 2020 Dec 5]. Available from: https://trends.google.com/trends/explore?date=today 3-m&geo=US&q=COVID19

25. Google. Google Trend Compare (COVID19, Coronavirus) [Internet]. 2020 [cited 2020 May 21]. Available from: https://trends.google.com/trends/explore?geo=US&q=COVID19,Coronavirus

26. Transit app. Who’s left riding public transit? Hint: it’s not white people. [Internet]. 2020 [cited 2020 May 15]. Available from: https://medium.com/transit-app/whos-left-riding-public-transit-hint-it-s-not-white-people-d43695b3974a

27. Transit app. Who’s Onboard? Surveying Transit Riders During the Coronavirus Pandemic [Internet]. 2020 [cited 2020 May 15]. Available from: https://www.youtube.com/watch?v=qkT9XQtd1o4

28. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Ann Intern Med. 2020;

29. Cheng H-Y, Jian S-W, Liu D-P, Ng T-C, Huang W-T, Lin H-H. High transmissibility of COVID-19 near symptom onset. medRxiv. 2020;

30. Pan X, Chen D, Xia Y, Wu X, Li T, Ou X, et al. Asymptomatic cases in a family cluster with SARS-CoV-2 infection. Lancet Infect Dis. 2020;20(4):410–1.

31. Dong Y, Mo X, Hu Y, Qi X, Jiang F, Jiang Z, et al. Epidemiological characteristics of 2143 pediatric patients with 2019 coronavirus disease in China. Pediatrics. 2020;

32. Achenbach J, Mettler K, Sun LH, Guarino B. Coronavirus may have spread undetected for weeks in Washington state, which reported first two deaths in U.S. [Internet]. Washington Post. 2020 [cited 2020 Mar 4]. Available from: https://www.washingtonpost.com/health/coronavirus-may-have-spread-undetected-for-weeks-in-washington-state/2020/03/01/0f292336-5bcc-11ea-9055-5fa12981bbbf\_story.html

33. Popovich N. How U.S. Coronavirus Diagnoses Are Lagging Behind the Outbreak [Internet]. 2020 [cited 2020 Mar 4]. Available from: https://www.nytimes.com/interactive/2020/04/01/us/coronavirus-covid-19-symptoms-data.html

34. Mitteroecker P, Gunz P, Windhager S, Schaefer K. A brief review of shape, form, and allometry in geometric morphometrics, with applications to human facial morphology. Hystrix, Ital J Mammal. 2013;24(1):59–66.

35. U.S. Bureau of Labor Statistics. Labor force characteristics by race and ethnicity, 2018 [Internet]. 2018 [cited 2020 Apr 27]. Available from: https://www.bls.gov/opub/reports/race-and-ethnicity/2018/home.htm

36. McLauphlin EC. CDC official warns Americans it’s not a question of if coronavirus will spread, but when [Internet]. 2020 [cited 2020 Feb 4]. Available from: https://www.cnn.com/2020/02/25/health/coronavirus-us-american-cases/index.html

37. Glanz J, Carey B, Holder J, Watkins D, Valentino-DeVries J, Rojas R, et al. Where America Didn’t Stay Home Even as the Virus Spread [Internet]. The New York Times. 2020 [cited 2020 Apr 4]. Available from: https://www.nytimes.com/interactive/2020/04/02/us/coronavirus-social-distancing.html?referringSource=articleShare

38. Evans E. City, county leaders declare local state of disaster due to coronavirus concerns [Internet]. 2020 [cited 2020 Apr 4]. Available from: https://www.fox7austin.com/news/city-county-leaders-declare-local-state-of-disaster-due-to-coronavirus-concerns