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

Luyu Liu

1. Introduction

A novel coronavirus disease (COVID-19) has evolved into a global outbreak for its high contagiousness, high fatality rate, and lack of known treatments.

1. Background
2. Analysis

In this section, we first introduce our data sources of the transit demand and COVID-19 daily cases statistics. Then, we introduce our theory about the dynamics of transit demand and the linkage between the two curves.

* 1. **Data sources**

**Transit demand change.** We collected the transit demand change data via the web interface maintained by Transit app (Transit app, 2020). Transit is a popular mobile app providing real-time public transit data and trip planning. The app covers over 200 metropolitan areas around the world with more than 5 million download on Android platform. Therefore, although the actual ridership data is still largely unknown and inaccessible, the usage of the app can be regarded as a rough measure for the transit demand of these app users (Transit app, 2020). Surely, the difference between the behaviors of apps users and other users, the accurate ratio of apps users among all transit users, and other factors affecting the representativeness of the usage statistics are still largely unknown. However, we can still use it as a rough but reasonable proxy to capture the changes of transit demand and ridership caused by the COVID-19 pandemic due to the large user group and popularity of real-time transit apps.

As COVID-19 hit the world and the transit demand decreased consequently, the app website kept updating on the changes of transit demand on an everyday basis. The change values are a set of percentage, calculated by comparing actual usage of the app to projected use of the app based on last year’s numbers. The projected numbers are also adjusted for annual growth (Transit app, 2020). In this way, the values represent the difference between expected and the actual transit demand.

The data includes 182 public transit systems across the United States, Canada, Australia, New Zealand, and France. We select all 121 transit systems in 71 metro areas, 55 counties, and 30 states across the United States and conduct analyses based on these areas.

**COVID-19 case numbers.** We collected the daily case numbers from the COVID-19 Surveillance Dashboard produced by University of Virginia (Biocomplexity Institute, 2020). The data includes all counties 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 multiple counties if cross-counties systems.

* 1. **Logistic model of transit demand change**

We first plot the curve of each system as shown in Figure 1. For all systems at the start, the demand generally oscillate around the base line. As the epidemic and quarantines progressed, the demand gradually decreased, until reached a very low level and then stabilized.

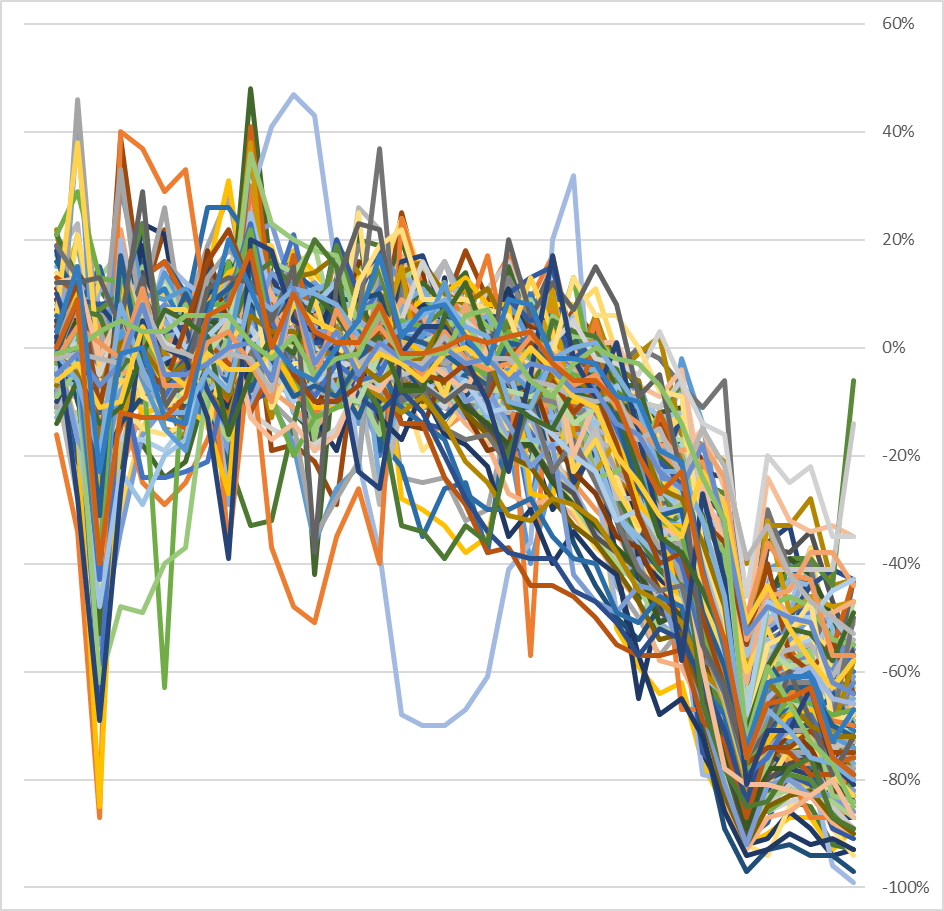


Figure 1: the cluster of demand change for all systems

The shape of the curves can be easily connected with logistic function; therefore, to generalize the change of transit demand, we will use logistic function to fit the data. A logistic or sigmoid function can be expressed into the following form:

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

Where: is the curves’ minimum value, k is the logistic growth rate or steepness of the curve, is the x value when the function reaches the midpoint. We will fit each transit system’s demand data using logistic model individually and calculate these three parameters. In the next section, we discussed the trend of three parameters for each system.

**L – Background value.** L represents the limit of the logistic curve. The curve will decrease rapidly and then stabilize around the L value, therefore we name it *background value.* In the practical sense, it represents the ratio of public transit users in this system that still will not or cannot stop using it regardless of the pandemic. For different systems in different metro areas, the background value is vastly different due to their different social and economic status.



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

The first factor that can have a significant impact on the background value is the non-physical occupation composition; or more specifically, the ratio of population that can work from home. If an area has higher ratio of non-physical jobs, workers can easily work from home thus the transit demand will decrease further.

The second factor that can have a significant impact is the dependency on transit, which can be measured with the ratio of public transit to work or transit commuter rate. Supposing two areas with same non-physical jobs ratio, the area with higher transit commuter rate can witness less demand decrease since more people still rely on the transit system to commute.

The regression analysis shows that:

**5 percentile date.** To measure the initial of the demand decrease, we can calculate the *5 percentile date* when the curve declines by 5% compared with the baseline and background value. 5 percentile date represents the first daywhen the demand curve began to diverge from the normality:

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

The 5 percentile date is not directly calculated from the raw observed data, instead, we obtained the measure from the smoothened curve to remove the stochastic noise. 5 percentile date can be a stable indicator to measure the start of the demand decrease.

**k – Decay rate.** The parameter k represents the curve’s decreasing speed from the baseline value to the background value. Therefore, we name it *decay rate*. Decay rate can be an important indicator for the actual response speed of urban residents/public transit users to the pandemic. Figure 3 shows the distribution of the decay rate.

Although the decay rate works well, the rate value per se does not have any physical meaning. Instead, we would like to use a temporal indicator to represent the speed and duration of the decline. From the logistic model, X0 represents the time/days from the day zero 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 5 percentile value, and the half-duration of decay. Therefore, we introduce half-life:

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

Half-life is only determined by the decay rate. Half-life indicates the duration of the duration of the demand decrease.

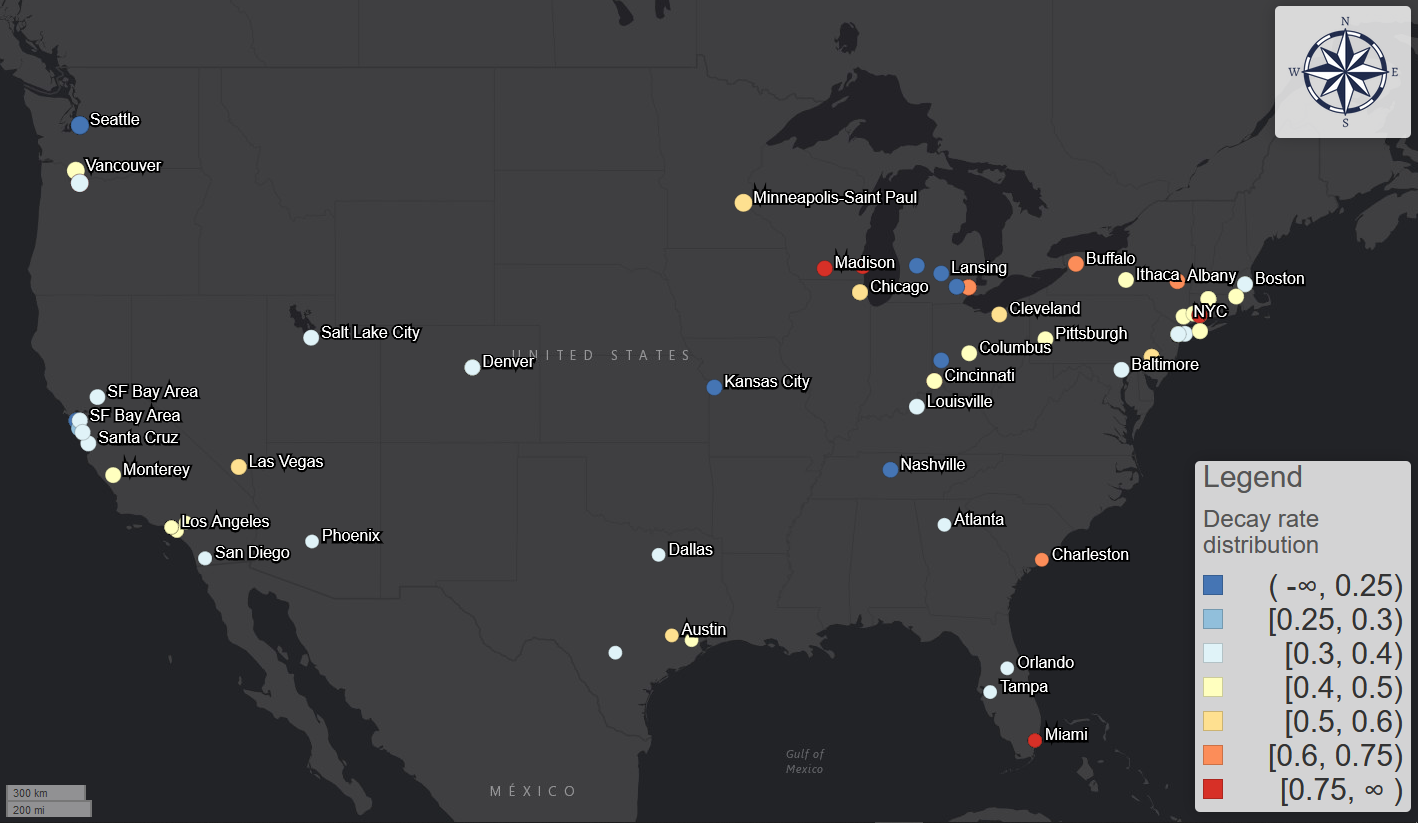


Figure 3: the distribution of decay rate across the United States

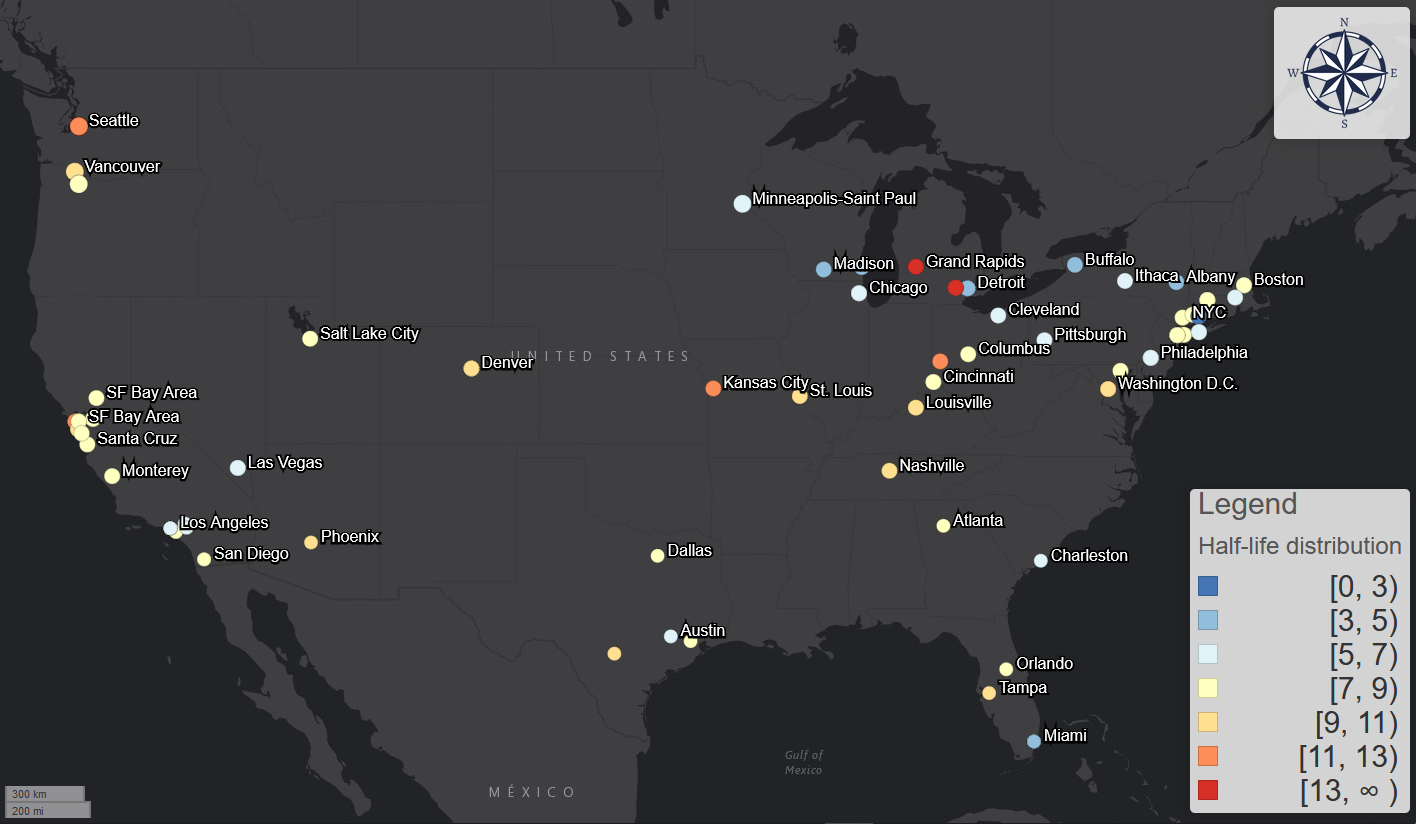


Figure 4: the distribution of half-life

* 1. **Delay between demand decrease and epidemic.**

The process of transit demand decrease is not synchronous with the development of the pandemic.

The

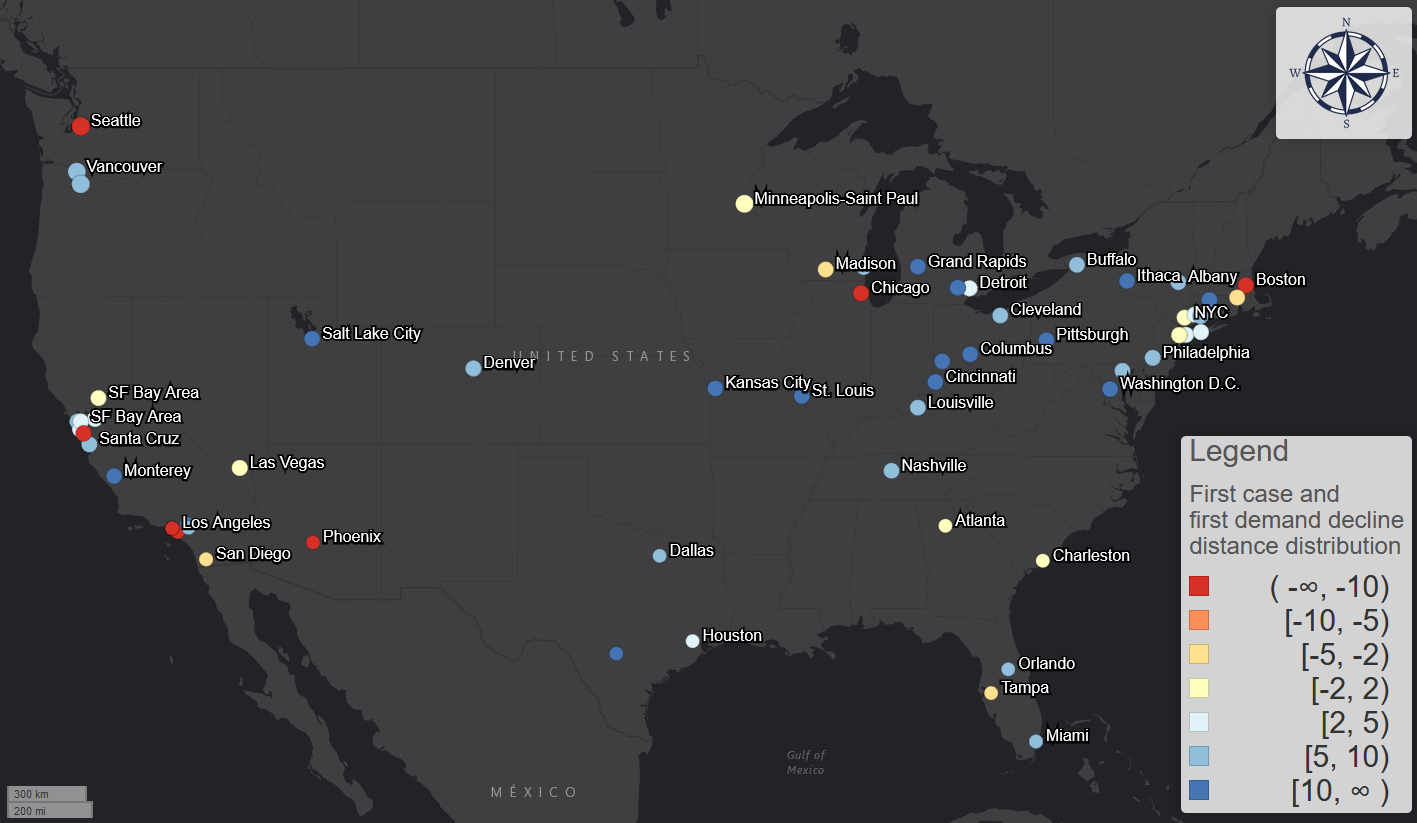


Figure 5: the distance between first confirmed case and 5 percentile date.