



Center for
Organizational
Excellence



COE DATA INSIGHTS:
VISUALIZING TRENDS IN SAFETY & SECURITY TIME
SERIES DATA FROM FEDERAL TRANSIT ADMINISTRATION

SEPTEMBER 2021

Safety & Security Time Series Data from Federal Transit Administration

The public transportation system in the United States is severely underutilized and stigmatized outside of major cities. The majority of public transit systems are used by a tiny fraction of the population of the cities served, with the exception of New York City, San Francisco, Washington, D.C., and a few others. Following a string of high-profile accidents, the Federal Transit Administration (FTA), which is under the US Department of Transportation (DOT), was tasked with improving safety measures for public transportation. Now, FTA is in high demand to develop a data-driven plan which not only maximizes public transportation use, but also improves safety and security for passengers.



In this article, we'll use the [Safety & Security Time Series Data](#) set provided by FTA to analyze and understand the uses and safety of public transportation across the US. This data is updated monthly and is accessible to the public over [here](#). These analytics can help us to understand the uses and safety of public transportation from a high level. Some of the analytics that will be presented here include:

- Examining the uses of public transportation in different cities, transportation modes (rail/bus/light rail), and monitoring growth.
- Tracking the safety of different modes of public transportation by monitoring reported accidents.

FTA gathers annual financial, asset, and operating data from public transportation departments around the country through the National Transit Database (NTD). Agencies that file as Full Reporters must also report monthly operating and safety data in the Annual Report, which includes financial, operating, and asset statistics. In the United States, FTA uses NTD results to apportion funds to urbanized and rural areas. All agencies must adhere to uniform reporting requirements. This includes timely reporting, accurate data collection, and uniform accounting procedures. The NTD Annual Report must include information from the agency's 12-month fiscal year, which is set to end in 2020. The NTD data validation process ensures that reporting requirements are met and that the reported data is reasonable. For further information please see [National Transit Database 2020 Policy Manual](#).

Results of the Analysis

The first question that I investigated was; which major transit system in the US and which mode of transportation was used the most in 2020? To find an answer, I created a unique column combining the agency name, mode of transportation, and UZA name (for example, metro bus (MB) operated by MTA New York City Transit is represented by “MTA New York City Transit _MB_ New York-Newark, NY-NJ-CT”). I filtered the dataframe for the year 2020, which had more than 100,000,000 UTPs (number of boardings). Heavy rail (HR) from MTA New York City Transit system serves more than 2.7 billion passengers for the year 2020, followed by a metro bus (MB) of the same system serving more than 690 million passengers. A graph of the results is shown in Figure 1.



Major Transit systems

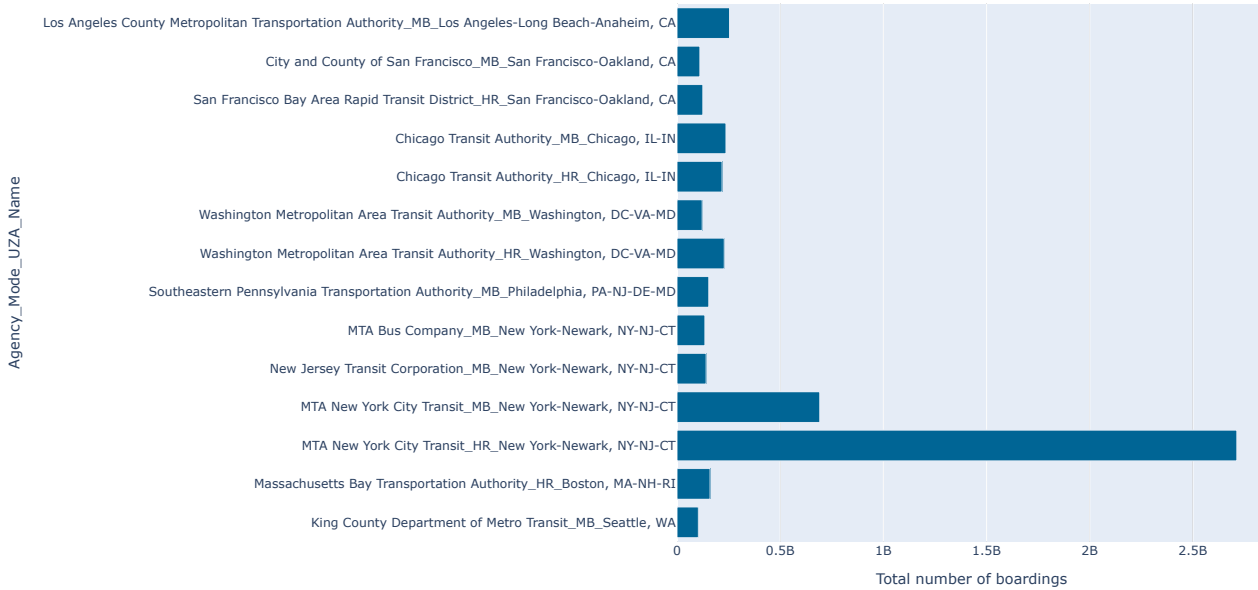


Figure 1. Major transit systems of the USA and their number of boardings for the year of 2020.

Significant growth is observed for heavy rail (HR) of MTA New York City Transit system from 1.7 billion passengers to 2.7 billion passengers from 2002 to 2020, whereas the rest of the major public transportation systems were losing their riders. See results in Figure 2 below.

Major Transit systems



Figure 2. Growth of the major public transportations systems in the USA from 2002 to 2020.

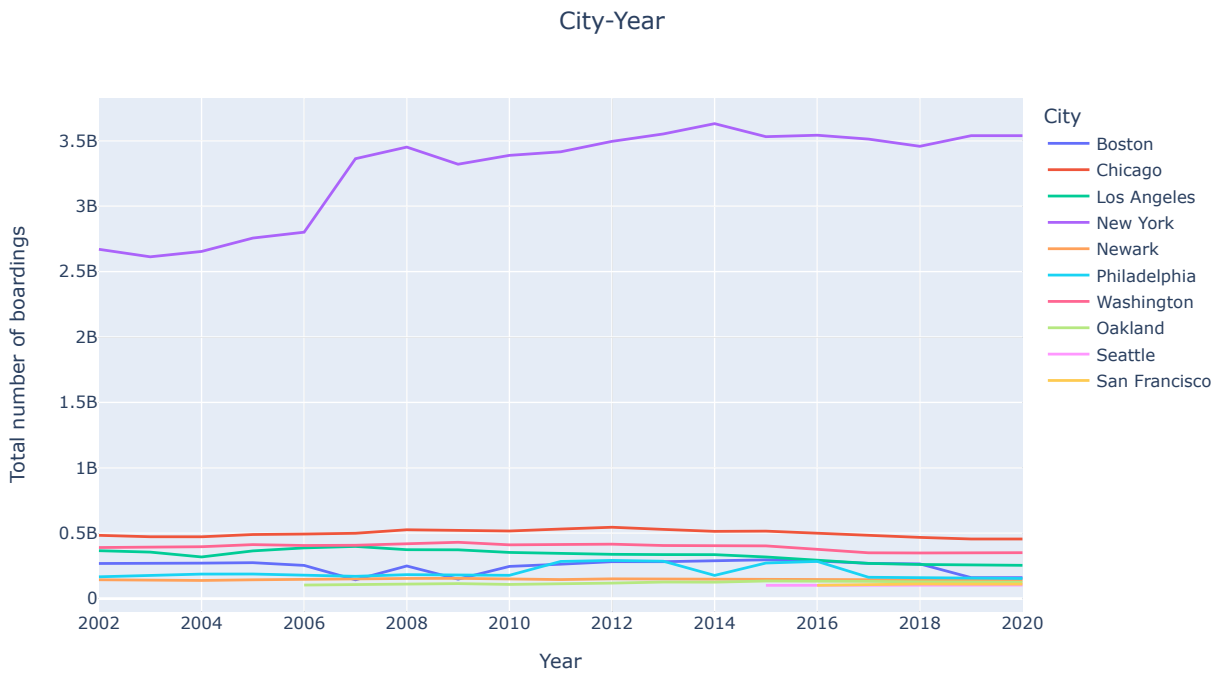


Figure 3. A city-wide comparison of uses and growth of public transportation.

A city-wide comparison of uses and growth of public transportation in absolute number is shown in Figure 3 for major cities in the US. It is observed that the use of public transportation grew only in New York, while the rest of the cities remain approximately the same.

Safety is always a debatable issue in public transportation. To dig deeper into this topic, I tried to find an answer to the question; who are the top 20 most unsafe agencies in public transportation (by mode of transportation)? From Figures 4 and 5 we can see that the highest number of unsafe events and deaths happened on the heavy rail of the MTA New York City Transit system. But we should not forget that the heavy rail of MTA New York City Transit system is also serving the highest number of passengers.



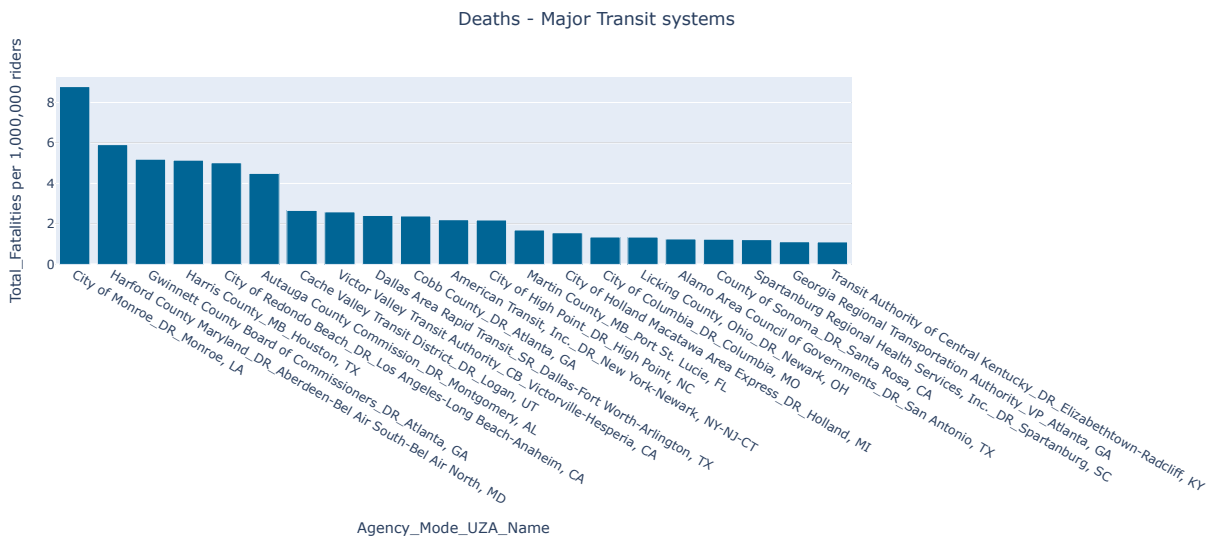


Figure 7. Top 20 Agencies by deaths.

One interesting finding stands out in Figures 6 and 7. In Figure 6, displaying the top 20 agencies by unsafe events, Demand Response (DR) services appear 19 times out of 20 times. And in Figure 7, displaying the top agencies by the number of deaths, Demand Response (DR) services appear again 16 out of 20 times. These findings indicate that the safety of Demand Response (DR) services may require further investigation.

The growth of public transportation and unsafe events are presented in Figure 8 below.

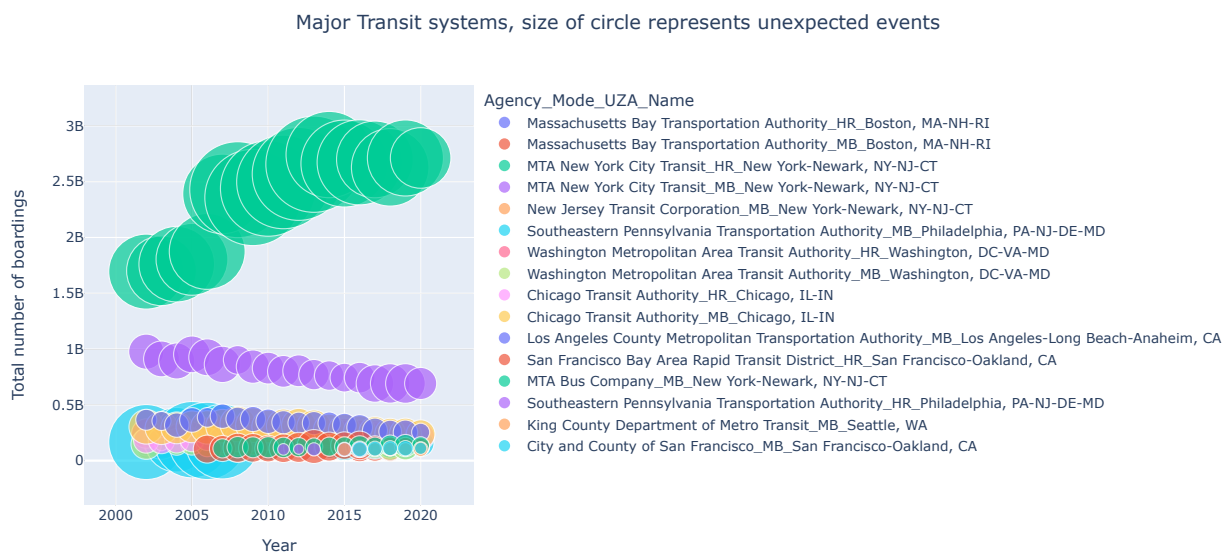


Figure 8. Growth of public transportation and unsafe events.

This is just a brief sample of some of the visualizations and findings made possible by analyzing open data from US Federal Agencies like the Federal Transit Administration (FTA). Many important and interesting findings can be obtained by analyzing the U.S. Government's open data which is a boon for the public. Here at COE, we collaborate with our customers to build thriving data ecosystems as well as for analytics that enable reliable and effective data-driven decision-making.

Additional Information on the Data Set and Analysis

The Safety & Security Time Series Data published in 2021 by urban Full Reporters to National Transit Database (NTD) combines full records from 909 agencies from 2002 to the present contains monthly-updated Safety and Security (S&S) information about the geographical area that the transit provider serves, services, events, fatalities, and injuries. The Excel file was processed with Python programming language with the help of Pandas library for DataFrame preparation, cleaning, and wrangling. Plotly express library was used for plotting. Some of the definitions used in the dataset and analytics are presented here.

Data Definitions

NTD ID	The five-digit number that is unique to each transit system reporting to the NTD.
AGENCY	The name of the system.
MODE	A two-letter code assigned to the mode of service operated. For more details read the excel file.
TOS	Type of Service: a two-letter code describing the type of service operated. “DO” = directly operated service. “PT” = purchased transportation (i.e. “contracted out”) service. “TN” = transportation network company. “TX” = taxi.
UZA NAME	The name of the Census-designated Urbanized Area in which the system primarily operates.
UPT	Unlinked Passenger Trips: The number of boardings on transit vehicles reported by the system in its own fiscal year corresponding to the year listed.
PMT	Passenger Miles Traveled: The number of miles traveled by passengers reported by the system in its own fiscal year corresponding to the year listed.
VRM	Vehicle Revenue Miles: The miles that vehicles are scheduled to or actually travel while in revenue service.
VRH	Vehicle Revenue Hours: The hours that vehicles are scheduled to or actually travel while in revenue service.
VOMS	Vehicles Operated in Annual Maximum Service: The number of revenue vehicles operated to meet the annual maximum service requirement.



AUTHORS

M. A. Saeed, Ph.D.

Data Scientist

m.saeed@center4oe.com

601.291.4670

Paul Eder, Ph.D.

Director

Paul.Eder@center4oe.com

443.637.2850

CONTACT

Center4oe.com

15204 Omega Drive, Suite 300

Rockville, Maryland 20850

301.948.1922