

126.153 *PGH Bus Times*

Introduction

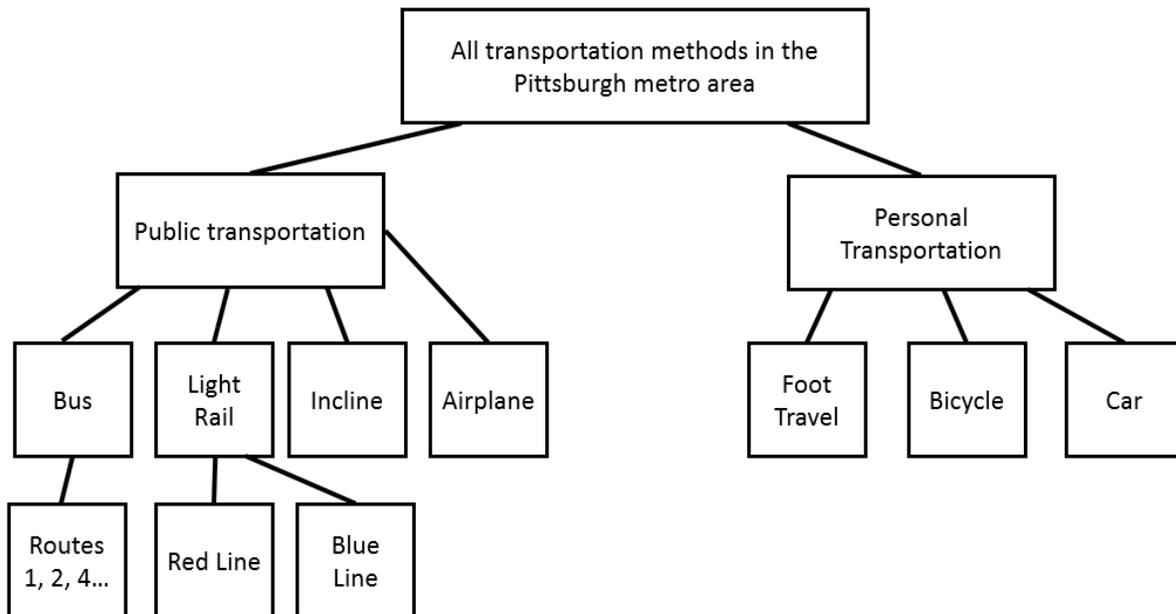
Pittsburgh has 98 bus routes serving the metro area on a daily basis, as well as 3 light rail and 1 incline. Each bus route carries a certain number of riders each day, with variations between weekdays, Saturdays and Sundays. However, the Port Authority only has a limited amount of buses and drivers to service all those routes.

Data is available from the Western Pennsylvania Regional Data Center for all 98 bus routes, on a monthly basis from January 2017 through October 2019 (34 months).

Inquiry questions include:

- How does ridership vary based on time of week?
- Which buses are the most popular on weekdays vs. Saturdays vs. Sundays?
- Which buses are the most delayed on weekends vs. Saturdays vs. Sundays?
- Which routes should the Port Authority dedicate additional resources to serving with offerings such as Bus Rapid Transit, and on which days of the week?

Classification Tree



Populations

The sub-population selected is:

1. Bus routes (excluding light rail and inclines), as light rail and inclines serve different population segments due to their limited routes and do not face the same types of scheduling considerations

Variables

Each row in the dataset corresponds to a particular bus route, month, and day or period of the week during the course of that month. The two key variables measured are:

2. Average daily ridership during the measured period
3. Percentage of on-time arrivals (no more than 1 minute early or 5 minutes late at each measured point on its journey)

Sample Set and Procedure

1. Determine the sampling axis to use. Across the 62 bus routes and 34 months in the data set, there are two primary potential sampling axes:
 - a. Sampling by route (e.g. selecting every 5th route)
 - b. Sampling by month (e.g. selecting every 3rd month)For this study, as the intent is to assess which routes should receive additional resources, sampling by month was conducted.
2. Collect data. For this study, data was collected as a downloaded datasets from the Western Pennsylvania Regional Data Center at <http://www.wprdc.org/>.
 - a. Both variables were present in the available data, across two different datasets.
3. Combine the datasets, ensuring that the on time data and ridership data are both available for study.
4. Remove superfluous route types, including light rail and inclines.
 - a. No variables were expected to produce anomalous responses, as the data was reported from official Port Authority tracking information.
 - b. Several bus routes have multiple paths, including 61 (A,B,C,D), 71 (A,B,C,D) and others. The analyst must determine whether these should be considered as one bus line or several. As the goal of this study was to examine specific routes that buses take, the overlap in line numbering was set aside and each path was considered a separate route, in alignment with how they were categorized by the Port Authority.
5. For this study, as the intent is to assess which routes should receive additional resources, sampling by month was conducted. Given 34 months in total, it was also important to ensure that a sufficiently sized sample was studied.
 - a. Randomize between every 1-4 months; in this case, every 3rd month was selected.
 - b. Resulting sample size was 12 months sampled (12/34), including:
 - i. 2017: January, April, July, October
 - ii. 2018: January, April, July, October
 - iii. 2019: January, April, July, October

Relevant Literature

- A 2018 report by the Congressional Research Service found that transit ridership declined by 7% between 2008 and 2018. However, it offered no explanation for the decline, indicating a need for additional research to study where public transit such as buses are most often used in order to emphasize service on key days and in key locations.
 - <https://fas.org/sgp/crs/misc/R45144.pdf>
- A 2017 article into the impact of Bus Rapid Transit (BRT) services explored the possibility of BRT increasing ridership through increasing the rider experience and reducing transit times. This study, conducted in Washington State from 2010-2014, found a strong causal relationship between BRT and as much as 35% increased ridership. This supports the theory that routes in

Pittsburgh with high ridership and low on-time performance could be good candidates for BRT services.

- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5627619/>

Hypothesis

The routes with the highest ridership on any given day – weekday, Saturday or Sunday – are likely to also be the routes with the lowest on-time performance. If this is true, it could help identify opportunities to provide enhanced service on those routes (such as Bus Rapid Transit lines) to serve these populations better and increase ridership.

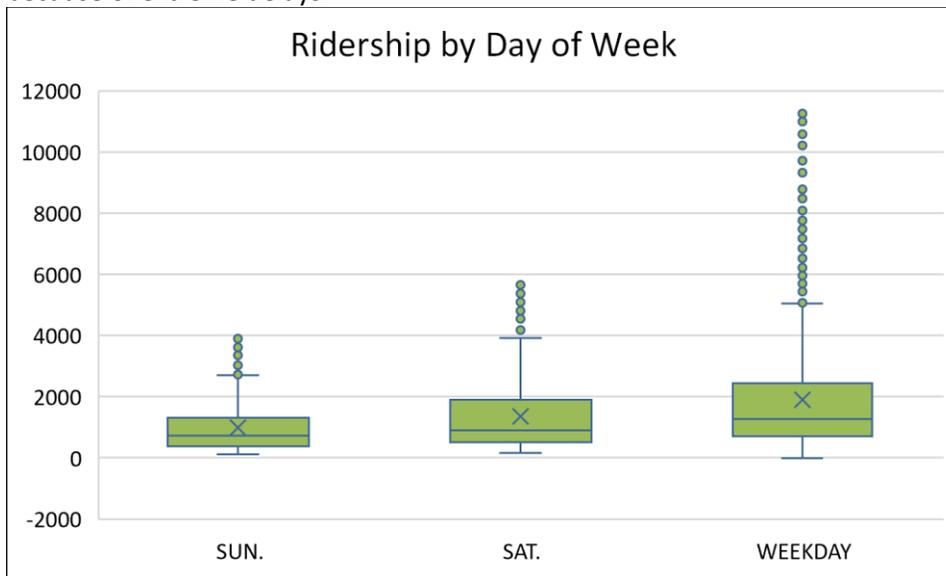
High ridership is likely to be associated with low on-time performance because:

- More riders getting on and off means increased logistical delays at each stop for people to pay the fare, climb in and out, etc.
- More riders means that there is likely to be an event. For example, a Steelers game would likely result in both many riders on the bus and many delays for that bus due to automobile traffic and traffic control measures.

Data Analysis

Ridership by Day of Week

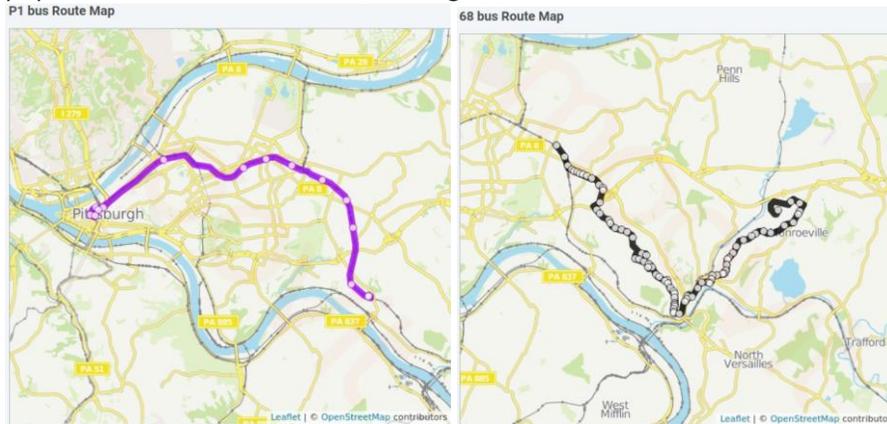
An overview of all bus routes in Pittsburgh found that ridership was highest on average on weekdays, with a median of 1,274 riders across all bus lines, a Q1 of 708 and a Q3 of 2450. The maximum value was also by far the highest, at 11,266 riders on the P1 route. In fact, the Pittsburgh Port Authority has recently (April 2019) [taken steps to limit boarding on the popular downtown Pittsburgh P1 route](#) because of extreme delays.



In comparison, Saturday travel had a median of 898 riders and Sunday had a median of 734 riders. As anticipated, weekday travel was more common, but there was significant overlap between the days.

There was also the greatest variation between weekday routes, with a standard deviation of 1,821 riders, compared to 1,174 riders on Saturdays and 819 on Sundays. This speaks to the highly varying popularity of different commuter routes. While on weekends most people do errands, visit friends or attend events regardless of where they live – leading to more comparable use across all bus lines – on

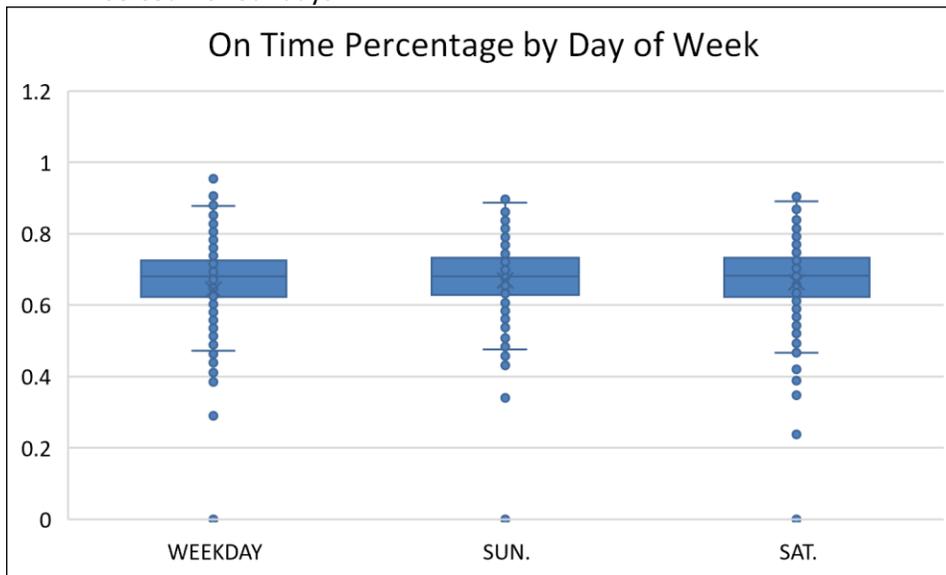
weekdays there were extreme variations between the highly popular routes like P1 and the much less popular ones like the 68 route, serving the eastern suburbs such as Monroeville.



On Time Percentage by Day of Week

No matter the day, the Port Authority maintained substantial consistency in its on-time arrivals. Barring a few outliers in all days (e.g. 0% on-time arrivals for one line that does not appear to have had accurate data reporting in the sample), the medians were quite similar:

- 67.96% for weekdays
- 68.14% for Saturdays
- 68.09% for Sundays



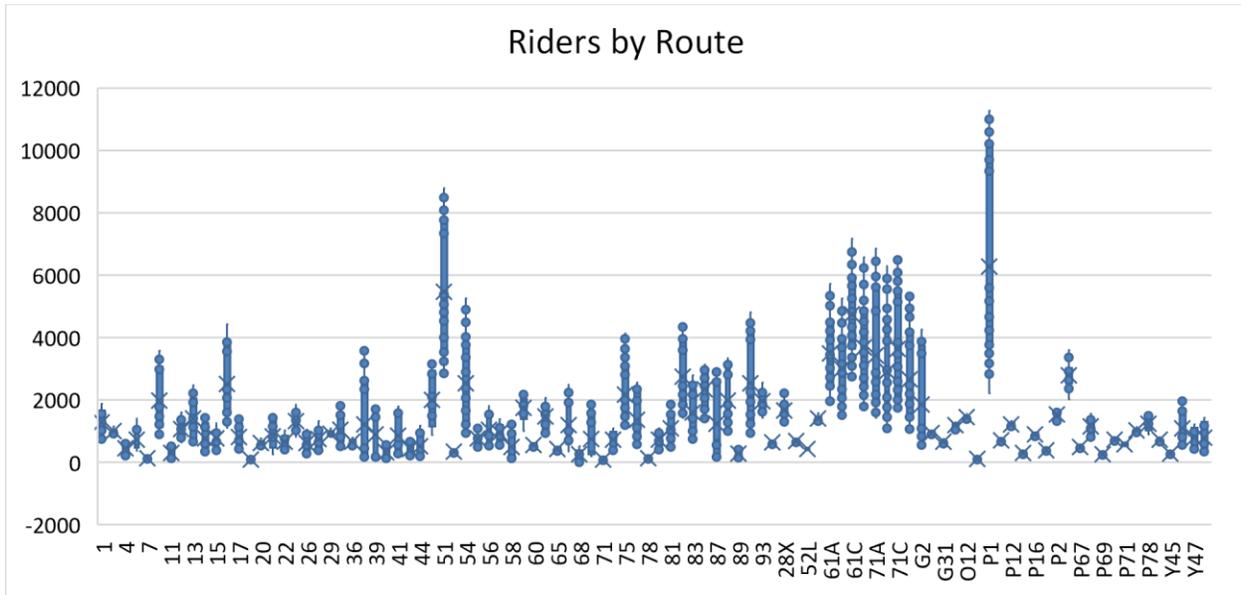
The weekday maximums were highest, at 97.19% compared to 90.65% on Saturday and 89.83% on Sunday. In addition, the standard deviation demonstrated consistency, at 12.3% on weekdays, 13.49% on Saturdays and 17.45% on Sundays.

Overall, Sundays were the least consistent days, but overall on-time performance saw that, within one standard deviation of the median, on-time performance ranged between 73.21% and 62.35%. While this shows significant room for improvement, it appears to indicate a standard that the Port Authority is working towards and consistently achieving.

Ridership by Route

Although each route had a limited set of data points (just 12 months measured for each route), it is still possible to see some interesting trends in the data, including that:

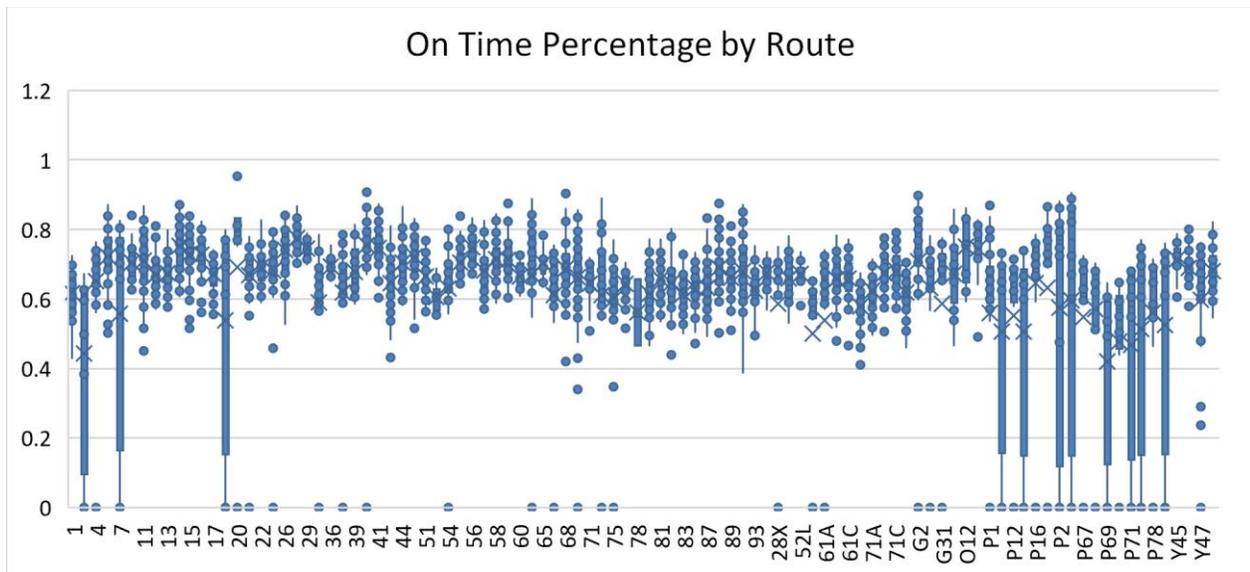
- Routes 51 and P1 have notably higher mean ridership than any other routes analyzed.
- 16 routes have mean riderships between 2,000 and 5,000 (from most to least: 61C, 61D, 71C, 61A, 71A, 61B, 71B, P3, 82, 71D, 54, 91, 16, 86, 75, 48)
- The remaining 80 routes have less than 2,000 mean riders.



On Time Percentage by Route

Overall, the on time percentages by route were quite clustered, as expected given the overall on time clustering. Exceptions included:

- 8 routes had less than 60% median on-time arrivals (from most to least: 79, 61D, P78, 53L, P71, P7, 78, P69)
- 11 routes had more than 75% median on-time arrivals (from most to least: 20, P2, O5, 40, 27, G2, P3, O12, 14, 26, 41)
- The two most popular routes, P1 and 51, were slightly below the median for on-time arrivals, at 62.49% and 66.01% respectively, but well within a single standard deviation of the overall median on-time arrivals.



Conclusions

- High ridership was not correlated with on-time performance. Of the 18 most popular routes in terms of number of riders, only one (61D) overlapped with the 8 worst-performing routes in terms of on-time arrivals.
- Weekday buses demonstrated substantially greater variation in average ridership between routes, with differences between major urban center and commuter routes and other less-traveled routes
- The Port Authority appears to have already put into place systems to compensate for increased ridership on different routes and days of the week, as nearly all routes fell within one standard deviation of the median on-time arrival percentage.
- However, on-time percentages were still low, with a median around 67-68%, [compared to the Port Authority's target of 73%](#). This does demonstrate a potential need for Bus Rapid Transit or other methods of reducing commuter delays, concentrated on popular routes to impact the greatest number of riders.

Documentation and Sharing

In gathering the data, I encountered several unexpected elements, including:

- Routes with limited data (perhaps due to being newer or being closed and re-opened over the sampling period)
- Routes reporting outlier data such as 0 passengers or -1 passengers, likely data artifacts from erroneous tracking systems (as opposed to months where data was simply not available)

Potential sources of error include:

- The accuracy of reporting may have changed over time as newer tracking devices were installed on buses
- While samples were taken from all four seasons, seasonality may still have impacted on-time performance (e.g. winter travel is more likely to be slower)

In the future, I would recommend focusing in on a smaller subset of routes and comparing downtown vs. suburban routes, routes on a specific day of the week, or bus vs. light rail performance.

Key Resources

Datasets were collected and sampled from the Western Pennsylvania Regional Data Center at

<https://data.wprdc.org/dataset?organization=port-authority-of-alleghey-county>, including:

<https://data.wprdc.org/dataset/port-authority-monthly-average-on-time-performance-by-route>

<https://data.wprdc.org/dataset/port-authority-monthly-average-ridership-by-route>