Will Friedrichs

CPLN 680

30 April 2022

Mid-Review Write-up

**Introduction/Motivation:**

SEPTA bus operations have a variety of issues, but this report focuses specifically on two of them: i) buses arriving to and departing from stops early, and ii) buses that are cancelled in real-time, within hours of their original scheduled departure times. The scope of both issues may be examined by tracking data associated with all early bus departures and cancellations.

**Importance of Mapping Early and Cancelled Buses**

Early buses are troublesome because unlike cases of late buses, on-time riders likely miss their buses entirely. This is specifically detrimental in Philadelphia, where there is a proliferation of low-frequency routes especially outside Center City. Cancellations are similarly disruptive, but are also detrimentally not reflected by Google Maps Transit, a service used by thousands of Philadelphians every day to map their transit use. Currently, buses cancelled in real time are shown as scheduled to arrive on-time on Google Maps, treated similarly to an untracked yet still expected bus.

Early departures and real-time cancellations negatively affect the quality of the SEPTA bus experience for existing riders. Service quality has been repeatedly shown to be the most common complaint from regular bus riders across Philadelphia and other American urban public transit systems, outpacing other concerns like fare rate, and dependability of network services is a key tenant of service quality. Low-income urban residents are especially vulnerable to economic shocks associated with missed buses, from lost jobs, to even child custody issues being exacerbated by the shortcomings of public transit. Service issues like early departures and untracked cancellations also discourage ridership growth by tainting potential regular riders’ first impressions of the system. Riders that see the bus network as one option but do have potential other ways to complete their trips have a lower toleration for poor service quality, and are the most likely to ride less frequently as a result of these issues. Generating more trips from higher income potential riders that fall into this category is also crucial to establishing a resilient, equitable transit system.

**Gaps in the Existing Data**

While early bus data is publicly accessible through Philadelphia’s real-time transit output feed, real-time non-GTFS Philadelphia bus cancellation data is not as accessible and has yet to be explored in a meaningful way. This report looks at both issues to uncover potential relationships between them, and to highlight the need for improvement. Additionally, while widely available public data streams have provided a steady flow of information about Philadelphia’s transit performance to the public, no public database has tracked performance over time in a public-facing way. This project aims to fill this void with a cumulative examination of early buses and cancellation data such that performance over time is tracked and recent issues are logged for public display.

**Brief Literature Review:**

Kenneth A. Small conducted some of the first extensive rider surveys studying San Francisco area bus riders’ attitudes towards delays in bus networks in an era when American commuter delay research had been predominantly concerned with auto transit (Small 1982). While the ultimate negative results coming from trip cancellation and delay have long been apparent to researchers, more recent research has suggested early departures to be especially detrimental where buses are infrequent due to the greater costs of missing a bus in those circumstances (Currie et al. 2011). Given the comparatively high number of bus routes in Philadelphia with frequencies of 20 minutes or more, early buses seem to be an especially relevant area of analysis for Philadelphia. Ibarra-Rojas et al. (2015) in their review of bus transport systems noted that early departures in bus networks are an impediment to building systems that depend on commuters transferring between buses. Recent SEPTA fare policies and planned changes to the network point to SEPTA’s encouragement of bus transfers, a policy that depends on buses leaving no earlier than planned. This project will fill a gap in the research by looking at the pervasiveness of early buses in SEPTA’s system.

**Question Definition:**

In tracking SEPTA bus early arrivals and real-time cancellations across routes, stops, and times over the course of five days, which routes and stops are the most adversely affected? What are the spatial patterns? How much do cancellations and early buses likely contribute to probably missed buses, respectively?

**Results:**

The final dataset used for this general analysis of SEPTA’s early and cancelled buses consists of data collected from the five following full days in 2022: April 11, April 12, April 17, April 18, and April 19. This preliminary analysis may be expanded in the future when coding issues are resolved such that additional full days of code may be compiled and cleaned without issue. Across the five previously mentioned days, though, a program downloaded SEPTA’s trip updates every five seconds, with each download consisting of up to 1,000 pieces of information regarding a bus’s position relative to its route, its next stop, its timeliness, and other details. For each stop on each bus trip, only one set of information was recorded. Even with downloads at this rapid pace, extrapolation for stops missed was necessary to fill in gaps and get a clearer estimate of the number of buses cancelled and the number of stops missed.

Over the five-day analysis period, of the more than 1.5 million data points collected associated with stops made by SEPTA buses during the five-day period (not including extrapolation), 31.2% of all bus stops were reached earlier than the scheduled time. As illustrated by figure 1, this measure remained steady across days for which data was collected. Meanwhile, less than 0.15% of stops were listed by SEPTA as cancelled. This number likely does not reflect the extent of the problem, however—once a bus has been cancelled, its updates are usually no longer recorded in the SEPTA real-time updates feed.

Figure 1

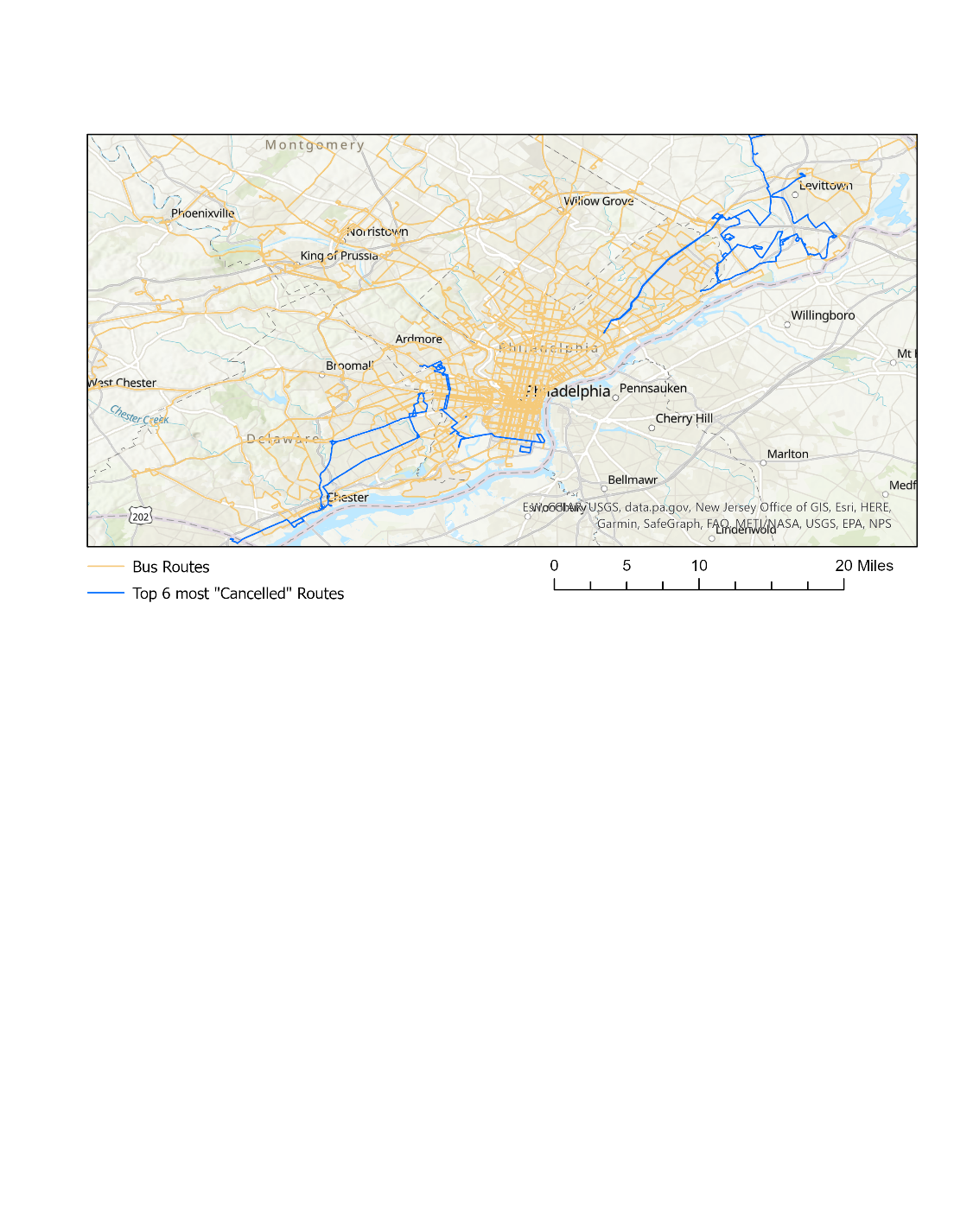
 Despite the inaccurate percentages of stops being labelled as cancelled, the measures are useful for comparing where cancellations are disproportionally frequent in the network. The figure below displays the six most cancelled routes, assuming that undercounting affected routes similarly. These are far and away the most cancelled in the network, with the sixth most affected by cancellations having more than twice that of the seventh most affected. Spatially, these routes are located away from the city center, and as a result may be viewed as “less important” in the case of short-term service cuts. It is also possible that the short period of time during which data was collected, a specific event or short-term local driver shortage contributed to this situation. Conversely, SEPTA’s most pervasively early buses do not have much of a distinct spatial pattern. They are scattered more evenly throughout the city. See the Appendix for the map of the five most “Early” routes (figure 3).

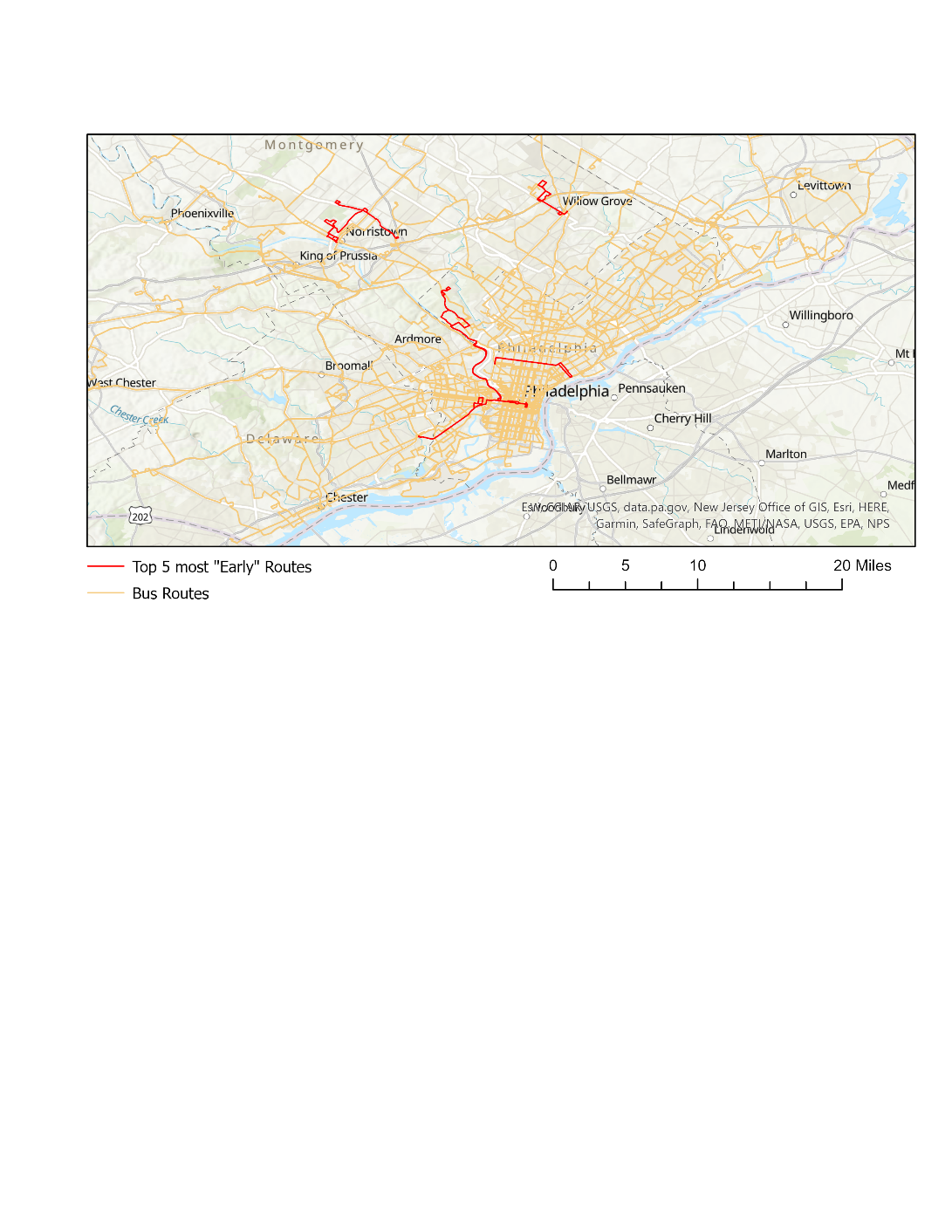
Figure 2

An algorithm to extrapolate non-observed stops based on timestamps, route information, and SEPTA scheduling information produced estimates of the actual total numbers of stops affected by bus cancellations and early arrivals. Additional Python scripts used these variables to estimate the time cost to a potential bus rider who theoretically showed up five minutes ahead of the supposed departure time (as per SEPTA’s recommendation) for any given early arrival o more than five minutes, as well as any bus cancellation.

“Hours” is a measure quantifying the impact of missed buses (though it does not factor in ridership). For every instance of a cancelled or early bus, the amount of time until another bus comes by associated with that route is counted towards the “hours” total of that stop or route. This variable is displayed spatially in the appendix (figures 4, 5). Early buses contribute to this measure significantly more to the total than cancellations do.

**Next Steps:**

Regardless of whether it falls within the scope of the capstone as it is submitted, I am eager to eventually get the Shiny Dashboard working. Currently it is still non-functioning, although the structure is there. Additionally, expanding extrapolation capabilities would make the code used in this report into a more effective tool.

**Appendix**

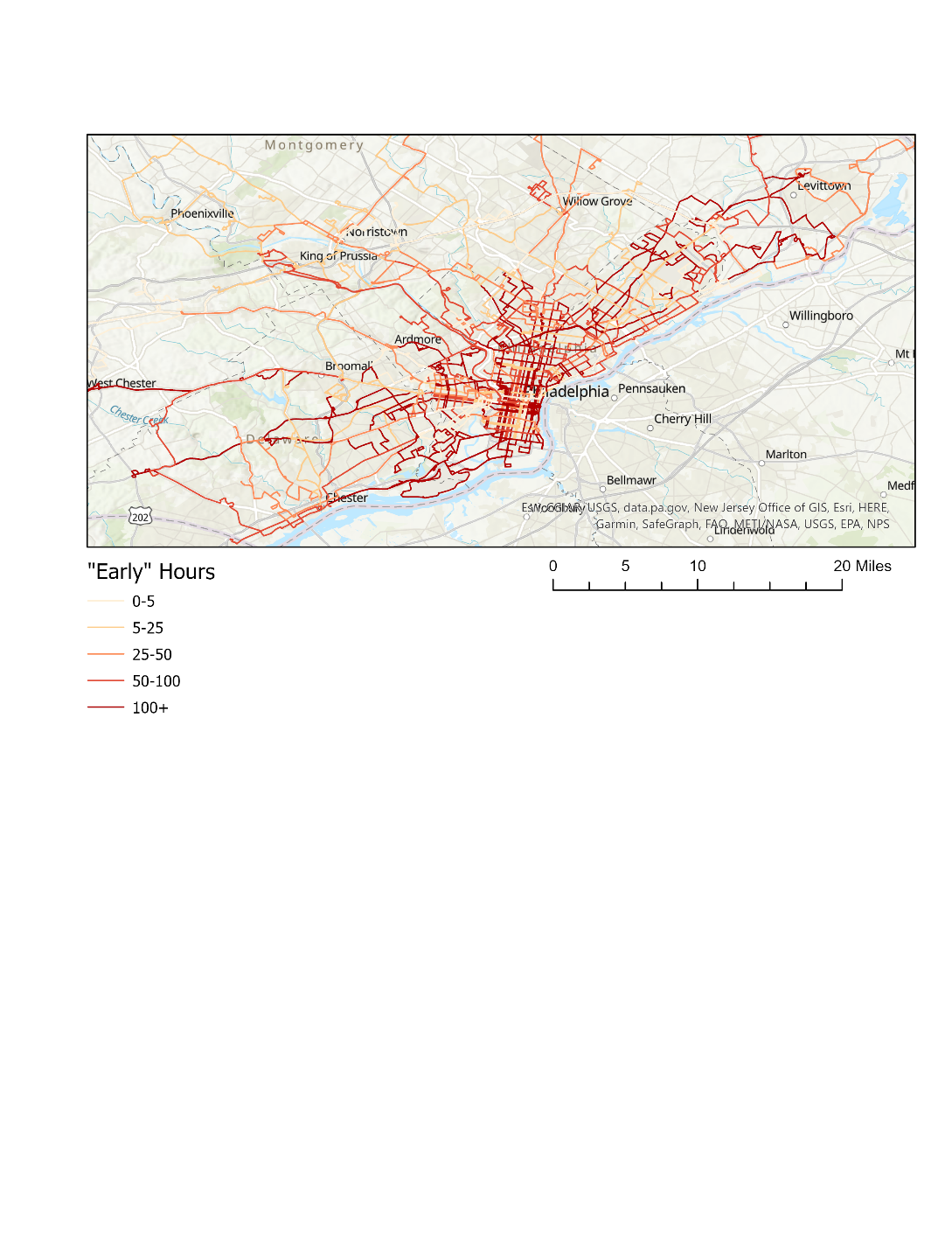


Figure 3

Figure 4

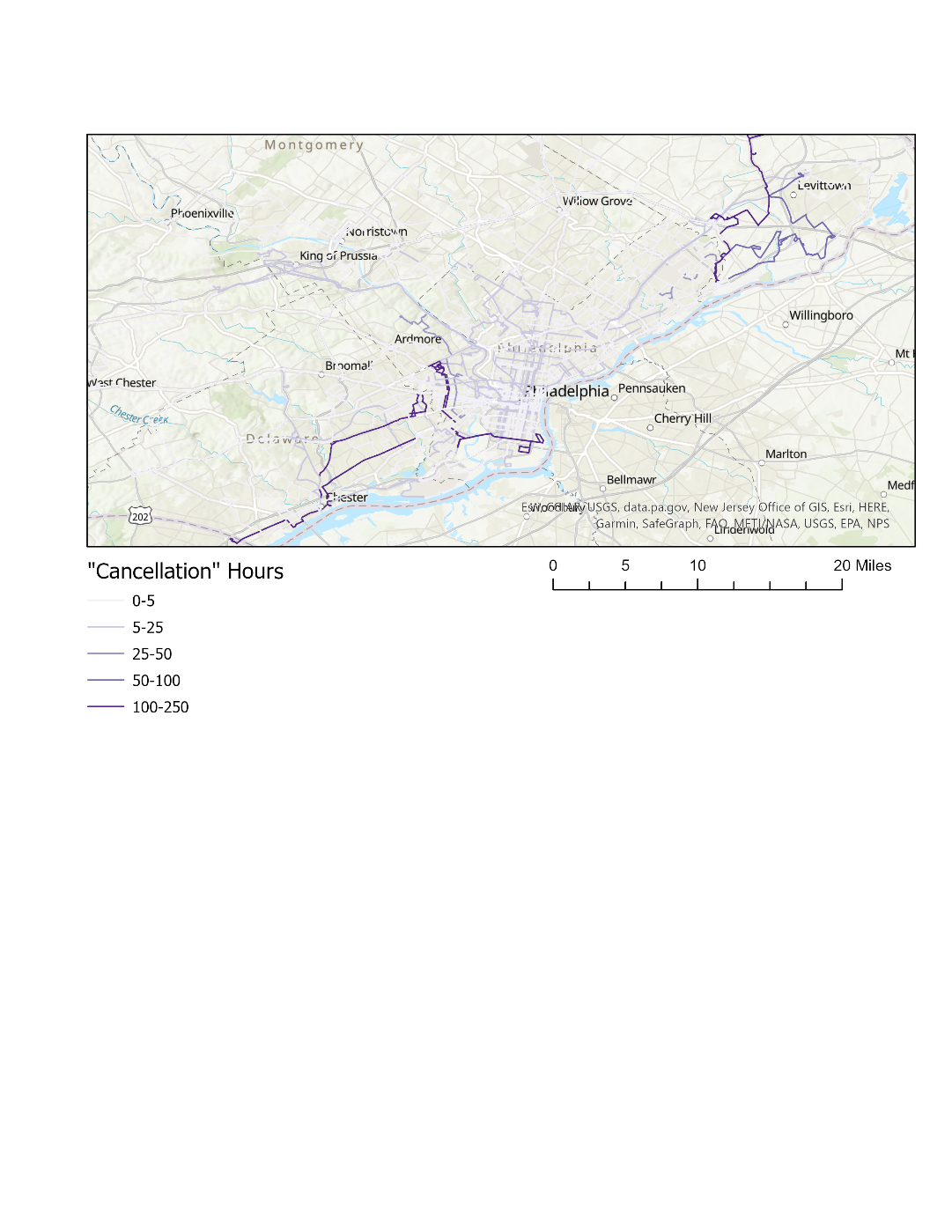


Figure 5

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Route | Percent of Recorded Stops Early | Percent of Recorded Stops Cancelled | Sum of "Early" hours | Some of "Cancellation" Hours |
| 1 | 33.742% | 0.000% | 41.73 | 0.00 |
| 2 | 33.544% | 0.086% | 82.43 | 4.50 |
| 3 | 39.486% | 0.047% | 58.47 | 4.33 |
| 4 | 33.756% | 0.037% | 198.90 | 2.60 |
| 5 | 27.376% | 0.000% | 39.05 | 0.00 |
| 6 | 29.714% | 0.231% | 54.33 | 21.00 |
| 7 | 31.961% | 0.092% | 129.57 | 10.82 |
| 8 | 22.563% | 0.000% | 21.55 | 0.00 |
| 9 | 33.162% | 0.010% | 83.60 | 0.43 |
| 12 | 35.360% | 0.134% | 49.73 | 4.67 |
| 14 | 26.759% | 0.064% | 358.18 | 13.60 |
| 16 | 34.832% | 0.131% | 169.22 | 8.95 |
| 17 | 39.201% | 0.082% | 172.02 | 5.50 |
| 18 | 36.734% | 0.004% | 220.70 | 0.33 |
| 19 | 23.548% | 0.000% | 9.47 | 0.00 |
| 20 | 36.749% | 0.115% | 373.27 | 27.53 |
| 21 | 32.890% | 0.059% | 86.82 | 3.73 |
| 22 | 33.477% | 0.000% | 44.67 | 0.00 |
| 23 | 38.747% | 0.073% | 223.62 | 8.83 |
| 24 | 37.951% | 0.093% | 21.30 | 2.50 |
| 25 | 32.655% | 0.000% | 156.02 | 0.00 |
| 26 | 21.524% | 0.009% | 91.83 | 0.98 |
| 27 | 35.610% | 0.007% | 93.85 | 0.00 |
| 28 | 28.737% | 0.053% | 59.40 | 2.10 |
| 29 | 31.076% | 0.037% | 20.45 | 0.93 |
| 30 | 28.246% | 0.000% | 20.97 | 0.00 |
| 31 | 37.728% | 0.047% | 46.07 | 1.50 |
| 32 | 26.301% | 0.044% | 45.00 | 1.50 |
| 33 | 32.364% | 0.105% | 74.20 | 12.32 |
| 35 | 44.089% | 0.057% | 0.60 | 0.00 |
| 37 | 31.390% | 0.137% | 156.38 | 24.40 |
| 38 | 39.330% | 0.116% | 134.63 | 5.28 |
| 39 | 15.317% | 0.000% | 9.87 | 0.00 |
| 40 | 35.629% | 0.068% | 146.52 | 5.27 |
| 42 | 24.925% | 0.000% | 107.25 | 0.00 |
| 43 | 16.035% | 0.000% | 23.07 | 0.00 |
| 44 | 34.695% | 0.082% | 155.07 | 4.15 |
| 45 | 28.747% | 0.072% | 12.80 | 1.80 |
| 46 | 33.414% | 0.000% | 45.35 | 0.00 |
| 47 | 29.351% | 0.041% | 423.00 | 14.37 |
| 47M | 20.996% | 0.000% | 4.40 | 0.00 |
| 48 | 39.564% | 0.019% | 230.13 | 2.03 |
| 49 | 33.939% | 0.015% | 128.47 | 0.00 |
| 50 | 31.290% | 0.034% | 4.50 | 1.75 |
| 52 | 33.835% | 0.099% | 94.57 | 33.50 |
| 53 | 16.424% | 0.000% | 7.97 | 0.00 |
| 54 | 43.076% | 0.242% | 146.65 | 13.85 |
| 55 | 23.848% | 0.030% | 31.78 | 2.37 |
| 56 | 39.943% | 0.050% | 161.08 | 10.37 |
| 57 | 34.240% | 0.033% | 211.08 | 4.00 |
| 58 | 35.903% | 0.018% | 117.02 | 2.00 |
| 59 | 25.524% | 0.000% | 100.73 | 0.00 |
| 60 | 36.026% | 0.216% | 67.63 | 15.60 |
| 61 | 28.276% | 0.100% | 46.27 | 8.00 |
| 62 | 49.811% | 0.126% | 50.15 | 0.00 |
| 64 | 33.408% | 0.036% | 49.02 | 1.40 |
| 65 | 28.988% | 0.318% | 70.28 | 21.40 |
| 66 | 33.358% | 0.168% | 114.65 | 19.97 |
| 67 | 27.407% | 0.005% | 120.57 | 0.00 |
| 68 | 33.013% | 0.000% | 119.77 | 0.00 |
| 70 | 28.524% | 0.048% | 28.57 | 2.40 |
| 73 | 17.081% | 0.012% | 54.30 | 0.00 |
| 75 | 26.041% | 0.296% | 16.32 | 9.25 |
| 77 | 35.315% | 0.000% | 6.58 | 0.00 |
| 78 | 8.000% | 0.000% | 2.00 | 0.00 |
| 79 | 23.816% | 0.000% | 61.40 | 0.00 |
| 80 | 21.591% | 0.000% | 2.70 | 0.00 |
| 84 | 30.599% | 0.000% | 55.27 | 0.00 |
| 88 | 22.865% | 0.000% | 21.10 | 0.00 |
| 89 | 26.700% | 0.000% | 80.42 | 0.00 |
| BLVDDIR | 28.910% | 0.759% | 6.32 | 8.75 |
| G | 28.200% | 0.718% | 197.77 | 139.18 |
| H | 30.577% | 0.317% | 5.25 | 23.52 |
| J | 28.537% | 0.000% | 41.18 | 0.00 |
| K | 33.196% | 0.111% | 83.45 | 5.97 |
| L | 25.212% | 0.034% | 42.47 | 5.87 |
| R | 37.325% | 0.000% | 277.08 | 0.00 |
| XH | 41.285% | 0.088% | 163.37 | 3.03 |
| 10 | 32.737% | 0.297% | 46.07 | 13.73 |
| 15 | 38.067% | 0.130% | 231.78 | 15.53 |
| 204 | 11.398% | 0.000% | 4.33 | 0.00 |
| 310 | 25.819% | 0.000% | 54.58 | 0.00 |
| 311 | 51.571% | 0.000% | 70.83 | 0.00 |
| 312 | 0.000% | 0.000% | 0.00 | 0.00 |
| LUCYGO | 10.052% | 0.000% | 14.50 | 0.00 |
| LUCYGR | 11.939% | 0.000% | 0.00 | 0.00 |
| 11 | 69.802% | 0.389% | 0.00 | 0.00 |
| 13 | 32.718% | 0.180% | 0.00 | 0.00 |
| 34 | 31.743% | 0.316% | 0.00 | 0.00 |
| 36 | 36.674% | 0.130% | 0.00 | 0.00 |
| 90 | 42.363% | 0.037% | 8.78 | 0.20 |
| 92 | 11.003% | 0.000% | 0.00 | 0.00 |
| 93 | 23.247% | 0.000% | 19.60 | 0.00 |
| 94 | 20.024% | 0.000% | 5.37 | 0.00 |
| 95 | 28.179% | 0.000% | 15.65 | 0.00 |
| 96 | 27.669% | 0.097% | 36.02 | 3.90 |
| 97 | 32.517% | 0.000% | 40.95 | 0.00 |
| 98 | 20.409% | 0.056% | 4.82 | 0.25 |
| 99 | 26.951% | 0.015% | 9.92 | 0.50 |
| 124 | 22.947% | 0.359% | 11.10 | 12.98 |
| 127 | 23.387% | 0.119% | 27.33 | 2.92 |
| 128 | 9.109% | 0.000% | 26.93 | 0.00 |
| 129 | 26.979% | 1.466% | 117.75 | 61.67 |
| 130 | 18.856% | 4.437% | 70.17 | 170.00 |
| 131 | 24.378% | 0.111% | 29.10 | 1.40 |
| 132 | 20.523% | 0.000% | 82.68 | 0.00 |
| 133 | 12.401% | 0.000% | 1.10 | 0.00 |
| 135 | 9.838% | 0.000% | 27.50 | 0.00 |
| 139 | 29.574% | 0.118% | 6.48 | 2.65 |
| 150 | 1.111% | 0.000% | 0.00 | 0.00 |
| 201 | 5.108% | 0.000% | 0.00 | 0.00 |
| 206 | 34.582% | 0.000% | 2.47 | 0.00 |
| 101 | 20.473% | 0.128% | 15.73 | 5.20 |
| 102 | 28.830% | 0.324% | 13.93 | 6.62 |
| 103 | 23.039% | 0.033% | 3.60 | 0.68 |
| 104 | 38.798% | 0.026% | 444.55 | 1.55 |
| 105 | 16.123% | 0.000% | 0.00 | 0.00 |
| 106 | 21.309% | 0.000% | 29.57 | 0.00 |
| 107 | 36.350% | 0.050% | 140.68 | 0.82 |
| 108 | 20.899% | 0.161% | 145.27 | 19.35 |
| 109 | 19.312% | 1.207% | 42.80 | 211.10 |
| 110 | 30.193% | 0.088% | 121.53 | 2.60 |
| 111 | 28.645% | 0.069% | 105.00 | 1.60 |
| 112 | 36.492% | 0.069% | 100.80 | 2.23 |
| 113 | 17.128% | 1.190% | 66.02 | 108.45 |
| 114 | 21.055% | 0.051% | 44.03 | 3.83 |
| 115 | 31.540% | 0.000% | 334.07 | 0.00 |
| 117 | 27.454% | 0.024% | 35.12 | 0.00 |
| 118 | 17.413% | 0.383% | 7.50 | 5.42 |
| 119 | 16.333% | 0.000% | 90.60 | 0.00 |
| 120 | 18.973% | 0.083% | 56.08 | 2.70 |
| 123 | 9.718% | 0.000% | 6.07 | 0.00 |
| 125 | 22.355% | 0.171% | 63.32 | 6.28 |
| 126 | 10.400% | 0.000% | 5.57 | 0.00 |