Emory College of Arts and Sciences

MATH 385W: PIC Math

February 29 2020

A Comparative Effectiveness Study: Optimal Route to Washington High School Utilizing Atlanta Beltline with GIS Analysis

Authors:

Tony H. Chen[a]

Mina Lee[b]

Hazel Sima[c]

Seobin Yoon[d]

Mentor:

Bree Ettinger, PhD

[a]Dept of Biology, Dept of Rehab Medicine, Dept of Math, Dept of QTM
[b]Dept of Math, Dept of QTM, Dept of RELAC
[c]Dept of Math, Dept of Env Science
[d]Dept of Math, Dept of QTM

EMORY | COLLEGE OF ARTS AND SCIENCES

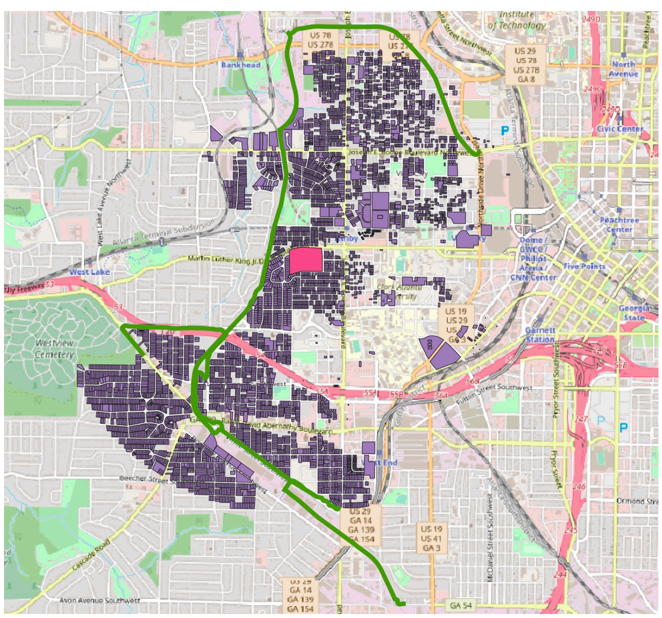
Atlanta BeltLine®

I. Introduction

This project is a joint effort between PIC Math at **Emory University Department of Mathematics** and **Atlanta Beltline Organization** ®. Atlanta Beltline is one of the largest and comprehensive urban redevelopment and revitalization program (scale-wise) in the United States. It aims to construct (and reconstruct) city transportation infrastructures for pedestrians & non-motor vehicle users, and public recreational facilities to make urban communities more prosperous, safer, and healthier. In particular, Beltline project reconstructs the old railroad corridors that encompass the city of Atlanta into pedestrian-friendly trails, along with comprehensive safety facilities, street lights, and urban arts demonstrations.

Booker T. Washington High School (short: Washington high school) is a member school of the Atlanta Public School (APS) system that was at the west of Downtown Atlanta, a designated school for the Washington school zone of Atlanta. The Washington school zone is a slanted geographic region that extends in the southwest-northeastern direction. Washington school zone engulfs 11,522 residential units (homes) of which resident students will receive school designation to Washington high school. The APS schools have school bus services on the stipulation that only students reside beyond the 1.5 miles radius of the campus location; otherwise, the student will not be eligible to set up a school bus pick-up stop at student's home location.

Fig. 1 An illustration of households of Washington school zone within 1.5 miles radius of Washington High School



Note: green (Beltline trails), purple (households), pink (Washington High School).

Intuitively, the 1.5-miles-radius condition is likely to pose commuting difficulties for many students who do not qualified for school bus service, especially those who live within but close to the edge of the border of bus-qualifying radius, considering Washington school zone is densely populated.

Safe Routes to School (SRTS) Program is an approach that was initiated by the U.S. Department of Transportation that encourages walking or bicycling to school by improvement of city transportation infrastructures and safety education to students. Atlanta Beltline trails are constructed with the principle that it will be a convenient, wide-covering, safe, and continuous traveling option for urban residents. Therefore, **it aligns with SRTS to propose a solution to this situation by utilizing the Beltline trails to commute to school, in contrast to traverse through neighborhoods** for those students who have to resort to walking or biking. In a general scheme, students will travel from their neighborhood to the closest entry point of the Beltline, get out from the closest Beltline entry point after traversing through the Beltline trails, and finally walk to school from there.

The nature of this project is *Comparative Effectiveness Analysis*. We constructed general optimal commuting strategies to school without any restriction, and another parallel optimal commuting strategies with maximized utilization of the Beltline. We then investigated the following two metrics:

1) Time consumed to commute to school for Beltline-usage maximized routes and for general optimal routes.
2) High injury intersections traversed along the Beltline-maximized routes and along general optimal routes.

We performed statistical analysis on these two metrics in a pseudo-experimental block design setting. Subsequently, we intended to provide Atlanta Beltline Organization with policy-supporting evidence of potential enhancement in efficacy and safety for school commuting Beltline-usage maximized routes.

II. Methodology & Analysis

A. Overview of Methodology

Provided by Atlanta Beltline organization, the data we analyzed for this project include Atlanta Beltline trails shapefiles (geographical data), Beltline entry points, high-injury intersections, grids of street and sidewalks, and residential household blocks. We approached the solution to the problem in following **three general phases**:

1) Classify households within Washington school zones into neighborhoods based on their homogeneous Beltline accessibility, such that all households within the same *neighborhood* will share a same commuting strategies to

travel to Washington high school.

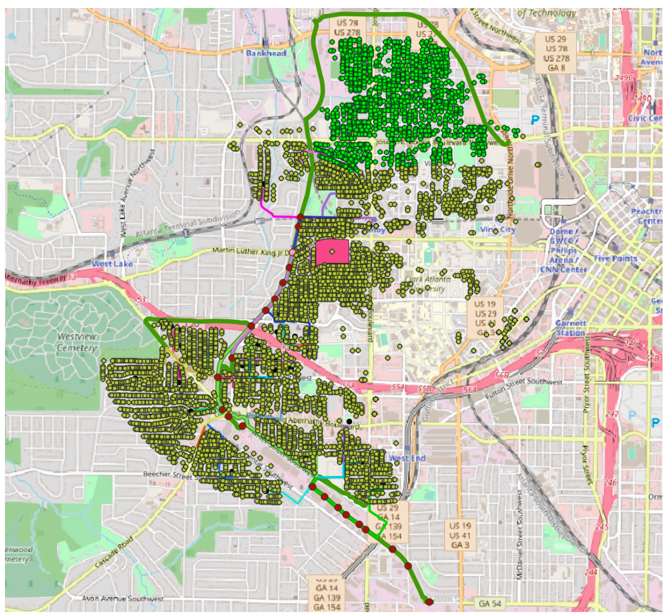
2) For each neighborhood, determine a starting location, and derive an optimal route to Washington high school with maximized utilization of Beltline, and another with general optimal routing algorithm on `QGIS` software.

3) Obtain time consumption to school based on two commuting strategies for each neighborhood. Perform comparative effectiveness analysis in `R` with block design to make inference on the time efficiency and safety of Beltline utilization maximized commuting strategy.

B. Phase I: Categorize Household Parcels

We first substituted every household parcels (house/building) with a centroid for clarity because centroids are area-less dots recording the geo-locations of households.

Fig. 2 Exclusion criteria applied to household parcels in Washington School Zone



Note: Yellowish dots represent residential parcels (centroids) included for this study. Green dots represent parcels that are excluded from this study (beyond 10 mins walking-distance from closest Beltline entry). Red dots are Beltline entries; Green line is the Beltline trail.

We excluded households centroids beyond the 1.5 miles radius since students from those households are qualified for school bus. We further applied another criterion to includes centroids to the north of Martin Luther King Jr Dr. *and* also within 10 minutes walking distance from the northern most planned Beltline entry with `ORS Tools` in `QGIS` (Fig. 2). The motivation for the second criterion was that northern extension of the Beltline is currently under construction, and doesn't not have determined entry points yet, so it is beyond the purpose of this project. We thus have a final household centroid sample size $N = 4039$, for further analysis.

Per the instruction of the Beltline Organization, we are expected to conceive a common starting point for households within a yet-to-defined neighborhood based on geographical vicinity. Additionally, it would be computationally cumbersome and excessive to design individualized commuting path from all households to Washington HS exhaustively. Therefore, we classified residential household centroids into neighborhoods, and then determined a common location as starting point for commuting to school for each neighborhood.

To implement this classification of neighborhoods, we first use the `Distance to Nearest Hub` in `QGIS` to roughly classify households based common *closest* Beltline entry, which implies that those households are in geographical vicinity. We initially ended up with 22 primary neighborhood divisions in accordance to total number of Beltline entries in an automated fashion.

Afterwards, we scrutinized and manually toggled to refine the preliminary classification of households, breaking some large neighborhoods into smaller ones. Our rationale was to further group households into **smaller** parcels based on locating on same street, cul-de-sac, and mostly street block. Doing this allows us to now have an unified starting point for routes to school for all households in the same neighborhoo with `Mean Coordinate` in `QGIS` on each neighborhood. See an illustration below (Fig. 3).

Therefore, we were able to have in total **N = 43** centroids as starting points for neighborhoods when conducting the optimal route analysis (Table 1).

Table 1 Summary statistics for total households, total starting points, and total Beltline entries.

Total households	Total start points (Neighborhoods)	Total Beltline entries
4039	43	22

C. Phase II: Optimal routing strategy for each neighborhood under Beltline-utilization maximization condition and unrestricted condition

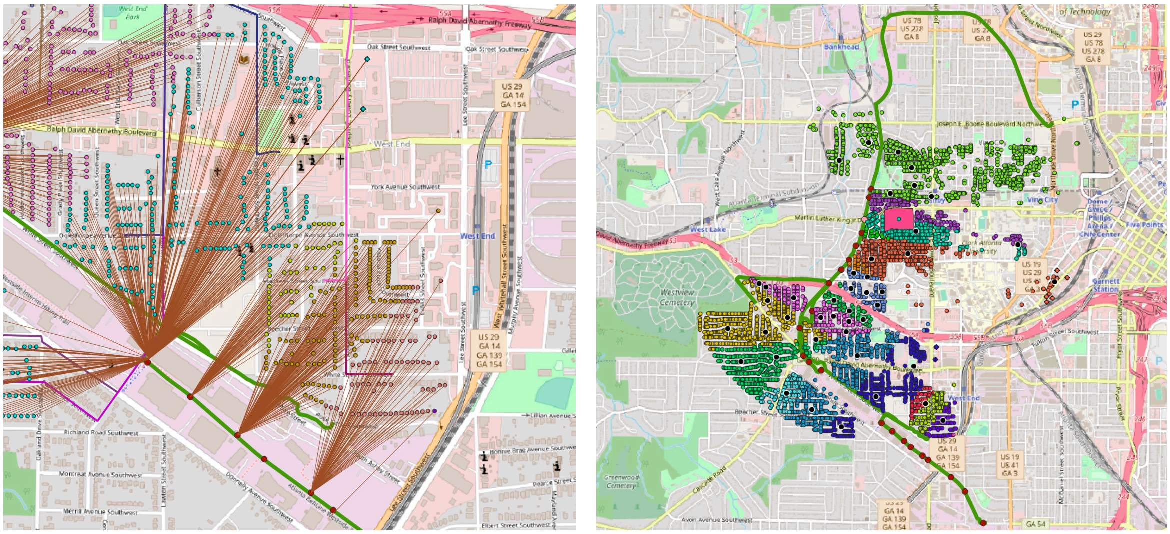
WITH the starting points for each neighborhood, indirectly for each household parcels, defined, we attempted to find the optimal routing strategies for each start point under **two constrains**:

1) Shortest distance from start point to school unconditional of choice of routes.

2) Shortest distance from start point to school while maximize the usage of Beltline.

Before routing analysis, we first imposed general intersections and high injury intersections on the map. Thus,

Fig. 3 Classification of residential household parcels into neighborhoods based on shortest Euclidean distance to common Beltline entries and final resulting neighborhood after human manipulation. Left panel: preliminary neighborhood classification example; Right panel: final neighborhood classification with finalized starting point for route to school for each neighborhood.

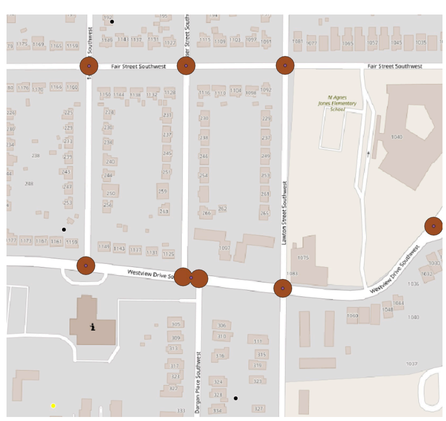


(a) An example of residential household parcels roughly classification based on common nearest Beltline entry point with QGIS algorithm. Households are connecting to its closest Beltline entry with a brown line.

(b) Overview of neighborhoods classification results. Residential households with same color assignment are within the same neighborhood. Hovering black dots are the assigned starting point for each neighborhood

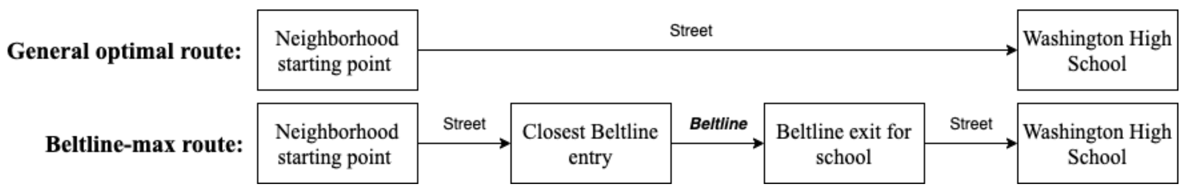
for general intersections, we applied `Analysis Tool::Line Intersection` in `QGIS` on street network to generate intersections. We then further removed redundant intersections: create a buffer (circular area) with radius of 30 feet using `Buffer Tool`; for those intersections that are so close that their buffers intersect, we averaged the coordinates of those intersections (≥ 2) with `Mean Coordinate` (Fig. 4). After defining intersections, we can then proceed to find optimal routes

Fig. 4 Example of defining intersections based on street network



We present a heuristic flow chart for two schemes of routing strategy to school (Fig. 5). The first condition is equivalent to find the shortest-distance optimal route from start points to school on existing street network. Thus, we implemented this **general optimal route** for every single starting point with `Network Analysis::Shortest Path Analysis::Layer to Point` in `QGIS`, and we set `Topology Tolerance = 45` feet to make routing more flexible. We obtained 43 general optimal routes. After we obtained the routes, we create a buffer around optimal routes, again with `Buffer Tool` and count general/high injury intersections with `Count Points in Polygon` for each routes (Fig. 6).

Fig. 5 Heuristic of two commuting strategies to Washington High School (Beltline/General)



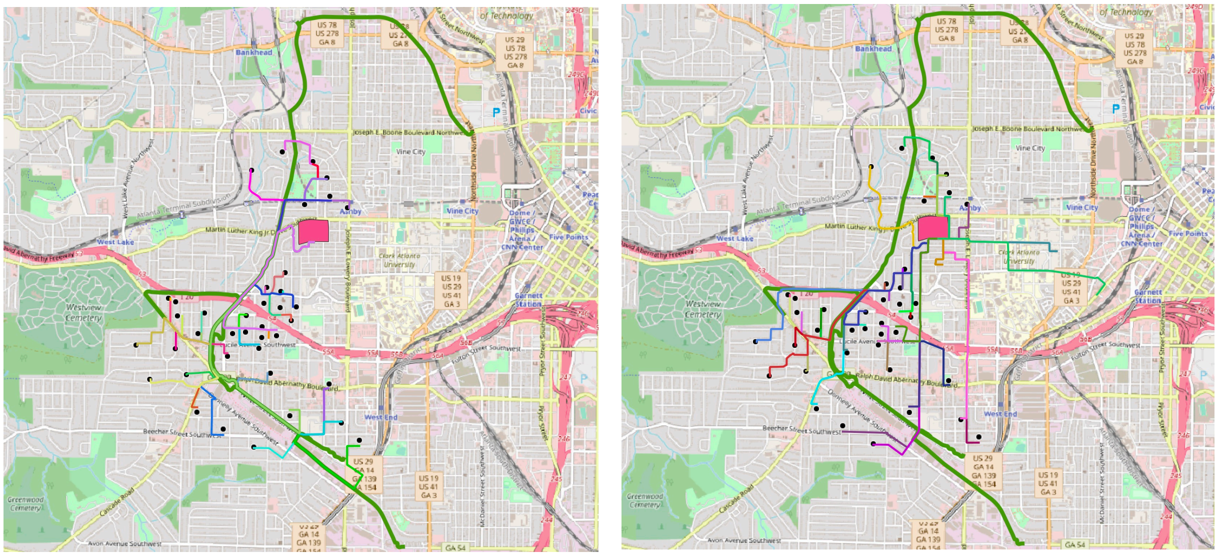
To force the maximized utilization, one would go from starting point to the closest Beltline entry via street, then traversing through Beltline to the exit closest to school before getting off from there, and finally travel to Washington High School using city street network (Fig. 5).

To translate these into `QGIS`, it's necessary to point out the following:

1) During the first stage, we had to separately find the closest Beltline entry for each of 43 starting points, and obtain the route to the closest entry from starting points. Thus, 43 routes would be produced from first stage.
2) During the second stage, we need to find the routes from various Beltline entries to the Beltline exit for school *along the Beltline*. Besides the Beltline exit for school, there are 21 Beltline entries, which implies there would be 21 ways to traveling along the Beltline.
3) During the last stage, we only had to compute one route from Beltline exit to school.

Based on these three corollaries, we found Beltline-max routes and implemented `Network Analysis::Shortest Path Analysis::Layer to Point` with `Topology Tolerance = 45` for more flexible routing. In first stage, we first temporarily categorized neighborhood starting points based on common closest Beltline entry, and save starting points from same categories to same `Layer` in `QGIS`. We operated `Network Analysis::Shortest Path Analysis::Layer to Point` on each layer for all start points in it, and obtained first-stage shortest routes from neighborhood to closest entry, for all 43 neighborhoods. For second and third-stage routes, it's fairly straightforward so we merely repetitively applied `Network Analysis::Shortest Path Analysis::Point to Point` function

Fig. 6 Illustration of General optimal route versus Beltline-max route for all 43 neighborhood starting points.



(a) Beltline-utilization maximized routes (b) General optimal routes

with entries to exit and exit to school, respectively. Routes from three stages were saved into three `Layer`'s.

To merge routes produced from three stages into a complete Beltline-max routes for 43 neighborhoods, we adopted a novel approach. With `Merge Vector Layer`, we merged routes from all three layers, along with Beltline, resulting in a linear network (union of all routes traversed via Beltline-max routes for all start points). Hence, we treated this merged network as a underlying street network and imposed all start points, and then applied `Network Analysis::Shortest Path Analysis::Layer to Point` to find the optimal route to school **on this network forged from all segments of optimal routes they were supposed to traverse**. By doing this, we conveniently obtained the complete Beltline-max route from each start point to school (Fig. 6). Similarly, we recorded the number of general/high injury intersections along the Beltline-max routes.

D. Phase III: Experimental block design & statistical inference

At this point, we had *number of intersections, number of high injury intersections, route distance (general optimal route), route distance (Beltline-max route)* from 43 neighborhoods. We viewed this as if we were doing a block design experiment, where two treatments (Beltline-max route vs. general optimal route) were administered on the same group of individuals (43 neighborhood starting points). We could make the inference about the difference of the average outcome measures (response) of these starting points between two treatments. The outcome measures were two metrics we obtained, **time consumption to travel to school** and **number of high injury intersections**

encountered. In a formal mathematical language for treatment literature:

$$\tau \in \{Beltline\ (b),\ General\ (g)\}$$

$$Y_1 = time,\ Y_2 = intersections$$

$$\mu_{d1} = \frac{1}{n}\sum_{i=1}^{n}(Y_{1bi} - Y_{1gi}), \text{ in sample } \bar{d}_1 = \frac{1}{n}\sum_{i=1}^{n}(y_{1bi} - y_{1gi})$$

$$\mu_{d2} = \frac{1}{n}\sum_{j=1}^{n}(Y_{2bj} - Y_{2gj}), \text{ in sample } \bar{d}_2 = \frac{1}{n}\sum_{j=1}^{n}(y_{2bj} - y_{2gj})$$

$$S_{d1} = \frac{1}{n-1}\sum_{i=1}^{n}\left[(y_{1bi} - y_{1gi}) - \bar{d}_1\right]^2$$

$$S_{d2} = \frac{1}{n-1}\sum_{j=1}^{n}\left[(y_{2bj} - y_{2gj}) - \bar{d}_2\right]^2$$

$$\text{Let } D_1 = Y_{1b} - Y_{1g},\ D_1 \sim \mathcal{N}(0, \sigma_1^2)$$

$$\text{Let } D_2 = Y_{2b} - Y_{2g},\ D_2 \sim \mathcal{N}(0, \sigma_2^2)$$

First we tested the **difference of time consumption** between Beltline-max and general optimal route. The time consumption for two routing schemes was computed with: $Y_1 = Time_i^{\{g,d\}} = tI_i + D_i/v$, where I_i is the number of general intersections for a given route; $t = 2$ mins is expected waiting time at a crossing in Atlanta; D_i is the total distance (mi) for a given route; v is set to 3.5 mi/h the average teenager walking speed.

$$H_0 : \mu_{d1} = 0$$

$$H_1 : \mu_{d1} \neq 0$$

$$\text{with } T_{n-1} = \frac{\bar{d}_1 - 0}{\frac{S_{d1}}{\sqrt{n}}}$$

We then tested the **difference of high injury intersections encountered** between Beltline-max and general optimal route.

$$H_0 : \mu_{d2} = 0$$

$$H_1 : \mu_{d2} \neq 0$$

$$\text{with } T_{n-1} = \frac{\bar{d}_2 - 0}{\frac{S_{d2}}{\sqrt{n}}}$$

For statistical analysis, we **further excluded 4 neighborhoods** (39) because they were either so close to school that it's counterintuitive to use Beltline, or they in fact would arrive at school first before getting to the closest Beltline entry.

Table 2 Summary statistics for general optimal routes and Beltline-max routes.

	Avg. Time (mins)	Avg. Intersections	Avg. High Injury Intersections
General optimal route	51.01 (±19)	15.28 (±6)	6.87 (±10)
Beltline-max route	47.95 (±13)	9.62 (±2)	3.36 (±4)

III. Results

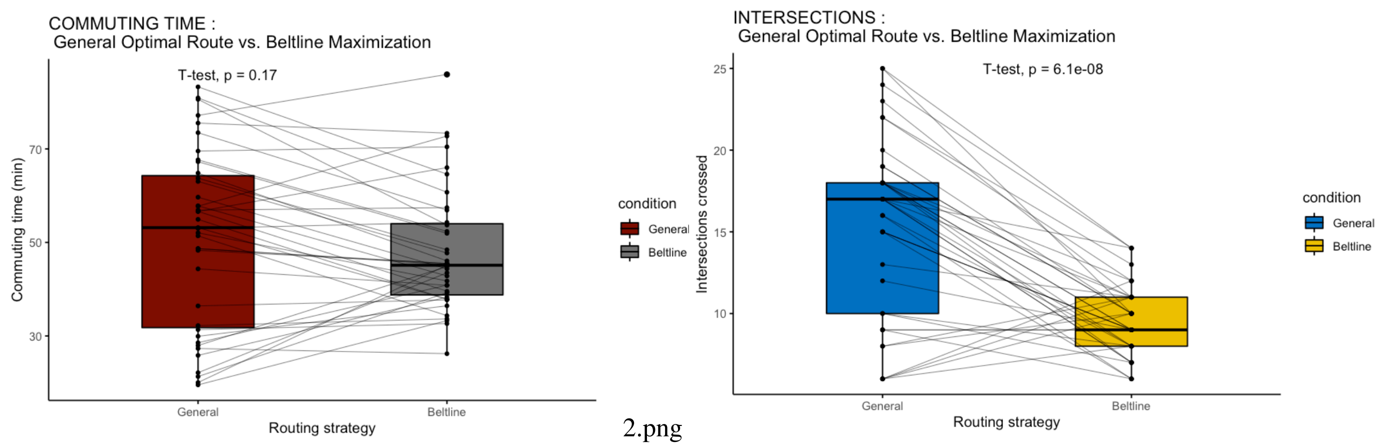
We were able to extract the general optimal route and Beltline-max route for all 43 neighborhoods associated with relavant statistics, and after final exclusion criterion, we analyzed N = 39 observations. We presented the summary table (Table 2). After inspecting the distribution of the metrics, including *time, intersections, & high injury intersections*, we assumed approximate normality for the difference in time and difference in intersections. We then proceeded to test our two main hypotheses for difference in time efficiency and in high injury intersections between general optimal routes and Beltline-max routes; we also tested a sub-hypothesis on the difference of general intersections encountered between general optimal routes and Beltline-max route. The comprehensive result of the paired-T test with reported P-value along with metrics visualization is presented (Fig. 7).

In terms of time efficiency, we found no statistically significant ($p = 0.17$) difference in time consumption between using the general optimal routes and using Beltline-max route, on average, although Beltline-max route tends to make detour (Fig 7. (a)). This is in part due to a statistically significant increase ($p < 0.0001$) in the number of general intersections traversed if one uses the general optimal route, on average, which in turned increases the time consumption because of crossing waiting time (Fig 8. (b)).

As for high injury intersections, we also observed a statistically significant decrease ($p = 0.0068$) in passing high injury intersections if one travels through the Beltline-max route to school. This is in principal due to the fact that traveling through the Beltline basically could avoid encountering intersections of any kind, since Beltline was treated as a smooth and non-intersected pathway.

To conclude, we successfully categorized residential households into neighborhoods with common commuting start points, and an Beltline-max routes to school from each start point was also obtained. Additionally, we found that commuting to school using the Beltline-max route avoids encountering high injury intersections and has a time efficiency that rivals general optimal route after considering waiting time at each intersection, even though detour is common for Beltline-max route.

Fig. 7 Difference in commuting time & general intersections encountered between general optimal & Beltline-max routes.

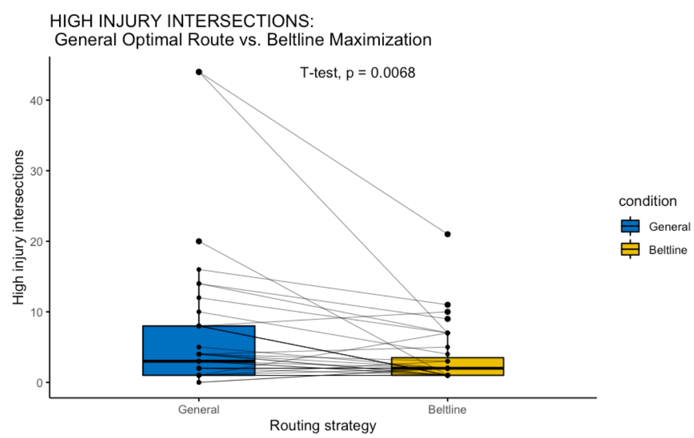


2.png

(a) Commuting time between general optimal route and Beltline-max route. Lines connect same observation between two treatment scenario for comparison. $p = 0.17$

(b) General intersections between general optimal route & Beltline-max route. Lines connect same observation between two treatment scenario for comparison. $p < 0.0001$

Fig. 8 Heuristic of two commuting strategies to Washington High School (Beltline/General)



Significant decrease in high injury intersection if travel with Beltline-max route. $p = 0.0068$.

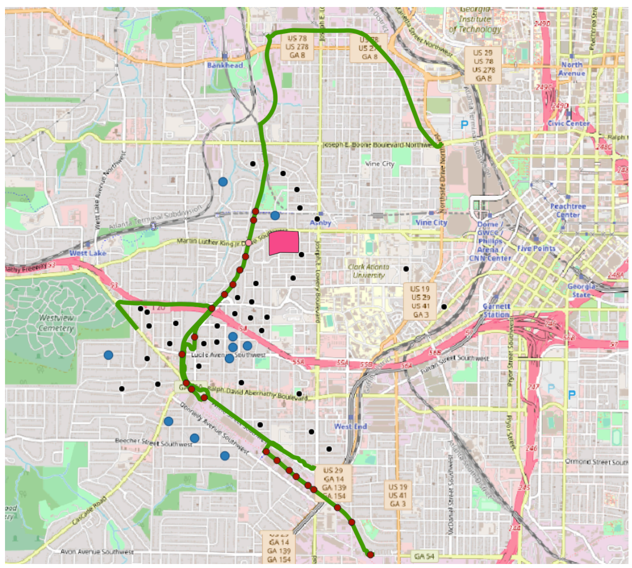
IV. Discussion

In our project, we provided evidence that by using the Beltline-utilization maximization route to school, safety, in terms of high injury intersections passed, can be enhanced, while did not suffer from sacrifice in time efficiency on average, compared to use general optimized route. Also, by passing significantly less general intersections through communities, commuting to school with Beltline-max route can in fact virtually avoid much less crime hot zones, which due to time limitation, we were not able to implement geographically and test on statistically. Therefore, our results can serve as evidence for grant application and proof for effectiveness of urban planning decisions of Atlanta Beltline Partnership. With our Beltline-max optimal routes results, the ABP could broadcast and recommend a specialized

Table 3 Summary matrix of neighborhood outcome under Beltline-max route.

Beltline-max Route Outcome (N = 39)	**Decrease Time**	**Increased Time**
Decrease High Injury Intersections	10	11
Unchanged High Injury Intersections	10	1
Increased High Injury Intersection	2	5

Fig. 9 Example for recommended neighborhoods benefit the most from Beltline-max route to school.



Note: Blue dots (10/39) are neighborhoods who will have a shorter time and pass less high injury intersection when traveling through the Beltline. Black dots are rest of the neighborhoods.

commuting path to Washington High School to potential school-year families in each neighborhood, perhaps in a joint effort with Board of Education/School.

Furthermore, based on summary statistics of Beltline-max routes, we advise that ABP could especially target to recommend students/families from 31 out of total 39 neighborhoods to use the Beltline-max route, since they will benefit from Beltline-max route in either spending shorter time or passing less high injury intersections, or both (Table 3). In particular, 10 our of 39 neighborhoods will benefit the most, with both decreased time and decreased high injury intersection, from using Beltline-max route to commute (Fig. 9).

Yet our study is limited by the fact that the GIS analysis is entirely based on computational methods and digital geographical shapefiles, with a little known about actual walkability and commutability of streets/sidewalks. This was reflected on our decision to increase `Topology Tolerance` in `QGIS` when conducting shortest route analysis, where we were allowing extra flexibility for routing decision to jump between streets with minor disconnection. We were not able to confirm in reality if the `Topology Tolerance` we set is viable for all routes. Lastly, we also manually connect some street disconnections on the network using to ensure route continuity, although having confirmed with ABP.

Reflecting on our project, we proposed three future directions. First recommendation, relating to our limitation, is to conduct a field survey of feasibility of routes in reality, and finalize best route guideline for each neighborhood. Second, human-powered vehicles, like bicycles, should also be considered in the routing framework for their different speed, as they are also popular choices among teenagers. Last, our study consider the Beltline usage exclusively under its maximized context, and yet it will be even more realistic to explore more flexible utilization (rather than exclusive maximization) of the Beltline in an algorithmic fashion, from which new results might yield.

V. References

Atlanta Beltline, “Atlanta Beltline Corridor Trails”, 31 Jan. 2017. Atlanta, GA. Atlanta Beltline, “Beltline Access Points”, 12 Jan. 2020. Atlanta, GA. Atlanta Beltline, “Grids of Street and Sidewalks”, 12 Jan. 2020. Atlanta, GA. Atlanta Beltline, “High-injury Intersections”, 16 May 2018. Atlanta, GA. Atlanta Beltline, “Parcels$_{W}ashington_{Z}one$], 12Jan.2020.$Atlanta$, $GA.AtlantaBeltline$, “$Washington_{C}ampus$],

Herald, Paige Fieldsted - Daily. “Red, Yellow, Green: The Science behind Traffic Lights.” Daily Herald, 6 May 2013, www.heraldextra.com/news/local/red-yellow-green-the-science-behind-traffic-lights/article$_{d}488e07e-abbf-545e-b23e-8c22c223e6fd.html?fbcli$
$IwAR3NuxxSA3iboTpu1PPMSSkRtrGo86zXDbAHap2x1AIfCTJW9Oxgeplxyh I$.

Toor, Amrit, et al. “Real World Walking Speeds of Young Pedestrians.” SAE Technical Paper Series, May 2001, doi:10.4271/2001-01-0897.

VI. Acknowledgments

PIC Math is a program of the Mathematical Association of the America (MAA) Support for this MAA program is provided by the national Science Foundation (NSF grant DMS-1722275) and the National Security Agency (NSA). **We give our earnest special thanks to Dr. Bree Ettinger (especially from Tony Chen). We were inspired so much from the conversation with you during our darkest time and we weren’t be able to come up with a statistical design if it weren’t for you.**

VII. Appendix

`R` code will be provided upon request for comparative effectiveness analysis.