Lab Report

Title: Accessibility Analysis Based on Cycling Infrastructure in Hennepin County
Notice: Dr. Bryan Runck
Author: Kyle Olson
Date: 12/17/2024

Project Repository: https://github.com/kolson5581/GIS5571/tree/main/Final_Project
Google Drive Link: *N/A*
Time Spent: *10 hours*

Abstract

Bike infrastructure is rapidly changing in Hennepin County as policymakers invest in new cycling facilities, and our understanding of what constitutes good infrastructure changes over time. However, cyclists of varying skill and confidence are comfortable on different kinds of cycling infrastructure. This project aims to analyze how the shape of cycling infrastructure in Hennepin County changes based on differing levels of comfort with differing facilities. How does the accessibility from various parts of the county change for cyclists who are comfortable riding on all streets, riding on basic cycling infrastructure, and only on the highest-quality infrastructure? Data from Hennepin County that differentiates not only the location, but also the category of bike facilities throughout the county is used to develop separate networks for three network comfort/quality scenarios.

Problem Statement

It is very simple to create a map of all the bike lanes in a jurisdiction and get an impression of where it is possible to get to using primarily cycling infrastructure - this is one way to understand the Bicycle Network. However, this gives an imperfect understanding of the bicycle network as it is used and understood by cyclists of varying ability. Some cyclists are confident riding on city streets, while others have higher standards of comfort for where they choose to ride and prefer facilities that provide protection or separation from vehicle traffic. In this project I develop a method to measure the differences in access to parts of the county for cyclists who are comfortable in safer and less safe cycling spaces, based on data around the type of bike facilities in different parts of the network.

Three separate networks are created as part of this process. Details on the specific attributes of each network are discussed in the Methods section, but the three networks are discussed in Table 1:

Table 1. Requirements

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	Road network	Raw input dataset	Streets geometry	Type of road	Hennepin County	Removing roads where cycling is not allowed/very rare

2	Cycling network	Bike paths, bike lanes, bridges, trails, etc	Line geometry	Type of facility	Hennepin County	Separating into two datasets depending on facility classification
3	Network Analysis tools	Arcgis Pro packages				Creating appropriate rules and connections between datasets
4	High quality network	Network representing highest comfort facilities	Cycling Network	Highest quality bike facilities; Separated bike paths, cycle tracks, and protected bike lanes	Hennepin County Bike and Pedestrian Network	Filling gaps and ensuring reasonable connections
5	Medium quality network	Network representing medium comfort facilities - all designated “bike facilities”	Cycling Network	All segments from high quality network, plus all other bike lanes designated by county	Hennepin County Bike and Pedestrian Network	Filling gaps and ensuring reasonable connections
6	Low quality network	Network representing anywhere a bicyclist could reasonably travel	Cycling network, streets centerlines	All segments from medium/high quality network, plus Local Streets	Hennepin County Bike and Pedestrian Network, Hennepin County Streets Centerline	Ensuring connectivity between line datasets, filling gaps.

Input Data

This project uses two datasets - both from Hennepin County. First is the Hennepin County Bicycle and Pedestrian Network dataset which contains line features representing pieces of the bike network, and details about the type of bicycle facility for each segment on the map. The second is Streets Centerline data, also provided by Hennepin County. This provides line features for all roadways in the county, along with attribute information about the kind of road. Using these attributes, I can filter out roads that a reasonable cyclist would be likely to use.

Table 2. Input Data

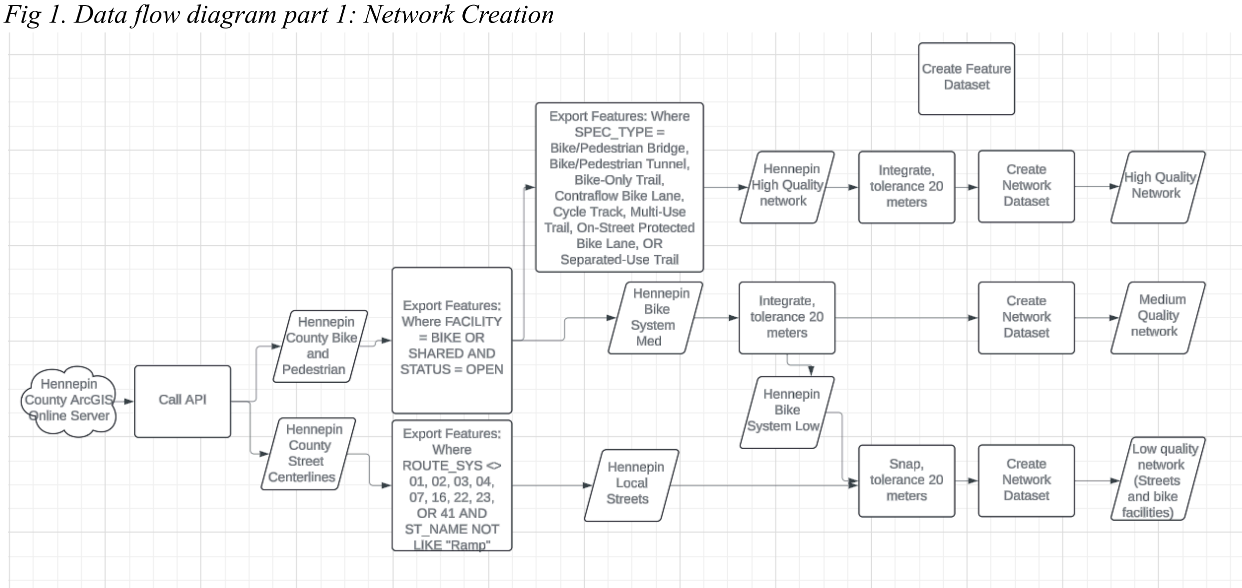
#	Title	Purpose in Analysis	Link to Source
1	Hennepin County Bike and Pedestrian System	Provides spatial data for at least two types of “bike-only” networks: medium quality and high quality	[Hennepin County GIS]()
2	Hennepin County Street Centerlines	Provides spatial data for streets network, this will be used for the “low comfort/no bicycle accommodation” scenario which will provide a comparison with other layers	[Hennepin County GIS]()

Methods

There are two parts to the construction of this project: the network creation and the analysis. While most of this can be done using Python in order to maximize reproducibility, there are several steps related to ArcGIS Pro’s Network analysis functionality that are easier to achieve using the GUI interface. Figure 1 details the data flow diagram for the Network creation stage. All commands are through ArcPy functionality unless otherwise noted.

All input data is acquired from Hennepin County via the ArcGIS Online API. The Bicycle Network and Street Centerlines are downloaded from URL and stored locally before being unzipped and brought into the ArcGIS Pro environment.

Fig 1. Data flow diagram part 1: Network Creation



After this, I need to create the components of each network. The Hennepin Bike and Pedestrian Network dataset includes data on sidewalks and other pedestrian-only facilities that are not needed for this analysis. I used Export Features to pull only “BIKE” or “SHARED” facilities, using the “FACILITY” field. This subset represents the Medium quality network inputs. Export Features is used again to create a High quality network feature class using the “SPEC_TYPES” field to narrow the Medium quality network to only higher quality segments. The types of facilities and street types can be seen in Table 3, below.

For the Low quality network, local streets were needed, as many confident cyclists will bike on city streets where necessary. Again using the Export Features tool, I used the "ROUTE_SYS" field to narrow down the types of streets to mostly local roads. The Hennepin County dataset did not have full metadata available, but points users to the Metro Regional Centerline Collaborative, where additional metadata for all field types could be found. A full accounting of types of roads included in this analysis is listed in Table 3. (Note: while I used an operation to exclude types of roads, Hennepin County does not have any streets classified in many of the MRCC classification types, so proactively selecting the roads I do want would have been simpler).

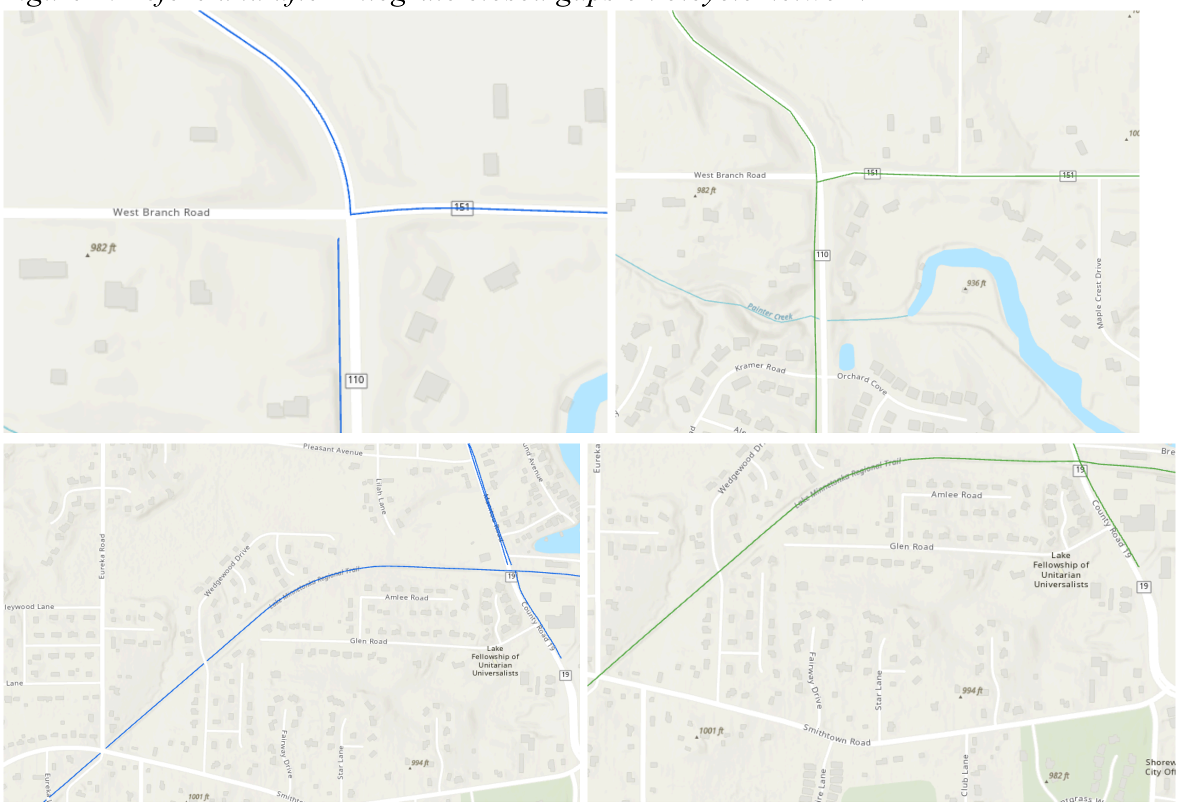
Table 3. Network construction

Network level	Input dataset	Field	Classifications
High quality	Hennepin County Bike and Pedestrian Network	SPEC_TYPE	Bike-Only Trail, Bike/Pedestrian Bridge, Contraflow Bike Lane, Cycle Track, Multi-Use Trail, On-Street Protected Bike Lane, Separated-Use Trail
Medium quality	Hennepin County Bike and Pedestrian Network	SPEC_TYPE	High quality network plus: Advisory Bike Lane, Bike Boulevard, Bikeable Shoulder, Buffered Bike Lane, Crossing, Sharrow/Shared Lane, Sidewalk, Standard Bike Lane, Unknown
Low quality network	Hennepin County Bike and Pedestrian Network	SPEC_TYPE	Same as Medium quality network
	Hennepin County Street Centerlines	ROUTE_SYS	05 Municipal-State Aid Street, 10 Municipal Street, 15 State Park Road, 21 Privately Maintained Public Access Road, 34 Non-trafficway

Once the feature classes which would go into each network were created, small gaps in the network had to be accounted for. For the High and Medium quality networks, there were many scenarios where the lines provided by Hennepin County did not cross streets, or had small gaps, or where the ends of street centerlines did not meet precisely with bikeway lines. In these scenarios I assumed a reasonable cyclist would simply make a connection, so I needed to correct for these situations before loading the feature classes into a network dataset.

For the Cycling networks, I used the Integrate tool, with a search tolerance of 20 meters. This tool searches for instances where the ends of line features are within 20 meters of one another, and moves the vertices to be in the same location. For the low quality network, I needed to ensure that the two line feature classes would connect, so I used the Snap tool with a search tolerance of 25 meters to ensure both datasets would align in the network. Examples of specific connections that were enabled by this process are shown in Figure 2.

Figure 2. Before and after Integrate closed gaps on bicycle network



Next, a Feature Dataset had to be created for all of the network datasets to go in. Then the Create Network tool was run to create three Network Datasets within the Feature dataset. From here, adjustments to the Network Datasets Costs Settings were done by hand to set the distance parameter we would use for the calculations for the next step. Because this analysis assumed average speeds, I chose to just get a distance in meters for all three datasets, and any conversions to time values would be done later.

Figure 3. Networks after classification: Low and medium quality networks include all higher quality components, too.

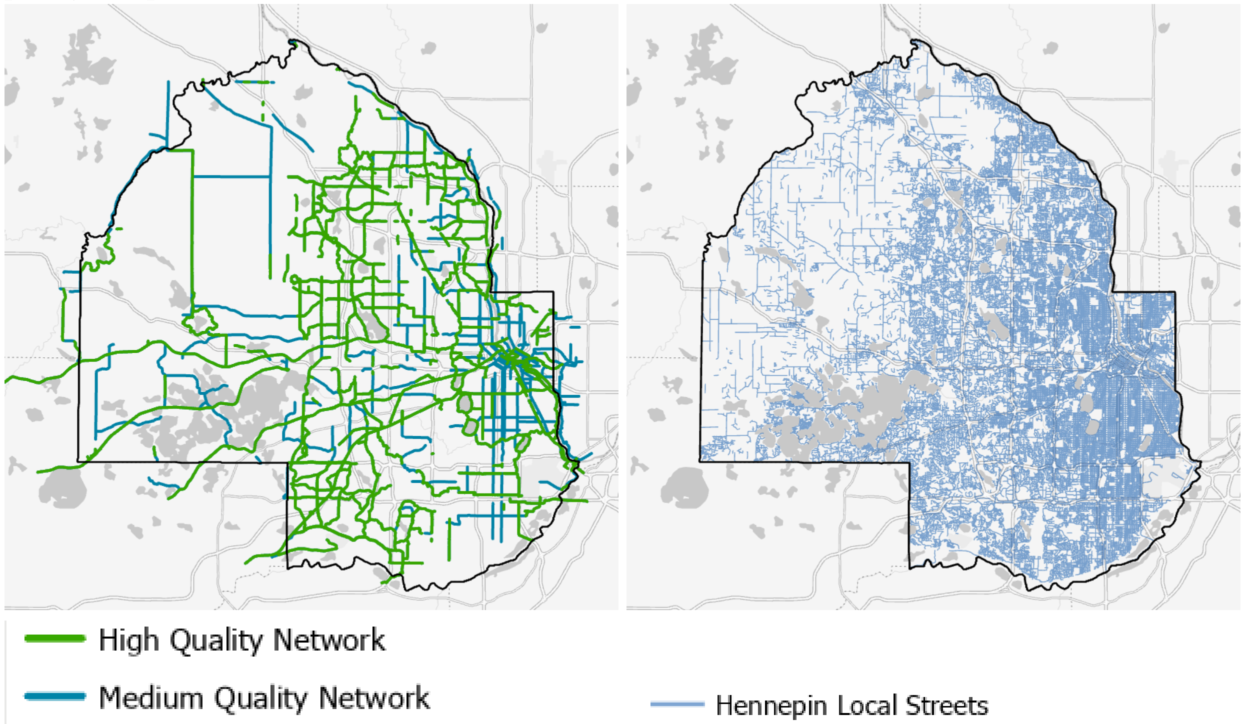
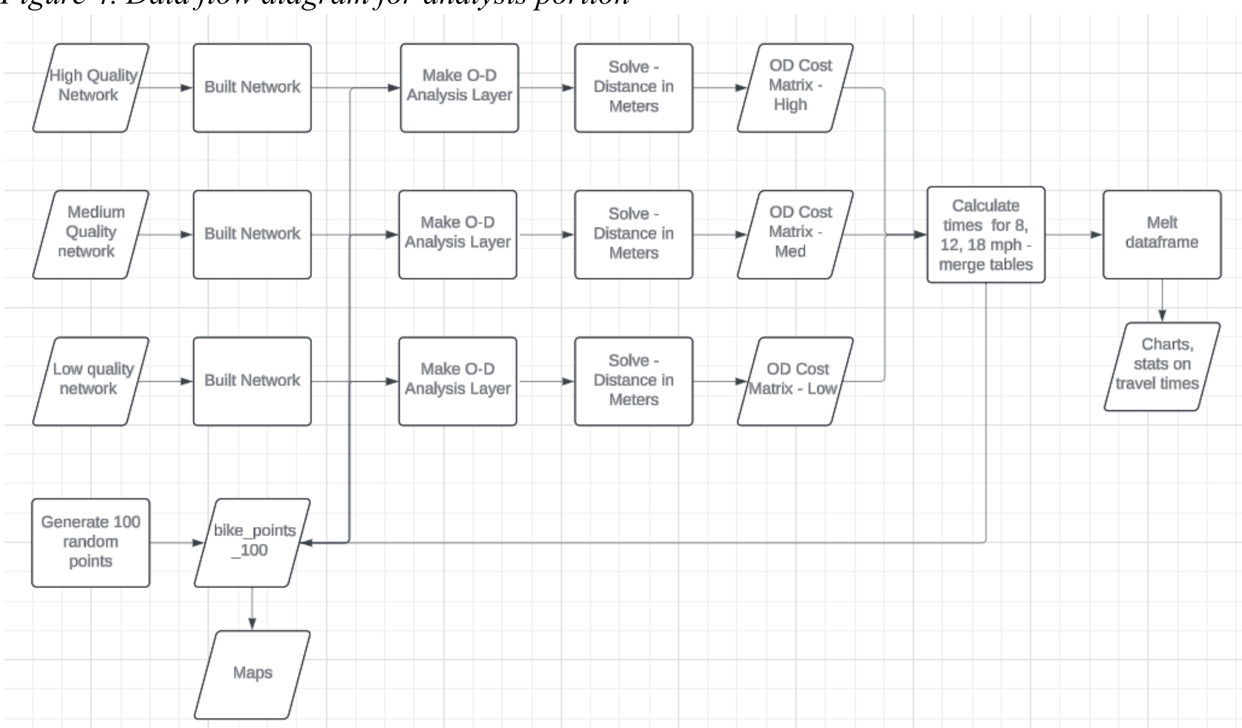


Figure 4. Data flow diagram for analysis portion



After the networks had been created, ArcGIS Pro requires the networks to be “Built” to account for cost calculations and to connect edges. This was done for each network, then all three were used to create O-D Analysis Layer. The O-D Analysis layers calculate the network distance between each Origin and Destination (if a connection can be made).

The Origins and Destinations for the O-D Analysis layers were determined by creating 100 random points within the county boundary (all points were added as Origins and destinations). They are added to the O-D Analysis Layers by hand. I set the search radius to 1000 meters for my analysis. This means that each point will search within 1000 meters for a connection to the network. While it is likely this creates some unlikely connections relative to the real world, my justification for a 1000 meter radius is that almost all cycling trips require some degree of off-network travel at the beginning or end of a trip.

Next, the O-D Analysis can be Solved. Table 4 summarizes the number of points which can connect to each network, and the number of points which connect to at least one other point.

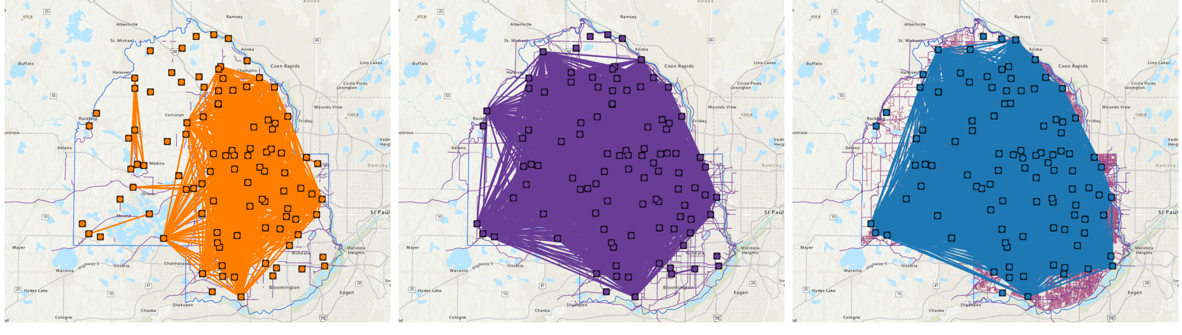
Table 4. Points connected to each network and points making connections via each network

Network	Points connected to the network	Points making at least one connection via network
High quality	81	61
Medium quality	90	81
Low quality	100	92

The distances in the tables were converted to times based on three different average speed scenarios: 8mph, 12 mph, and 18 mph. These speeds were selected to represent a wide range of cyclist speeds, and to help test to see if changes in accessibility held up across networks even under different speed scenarios.

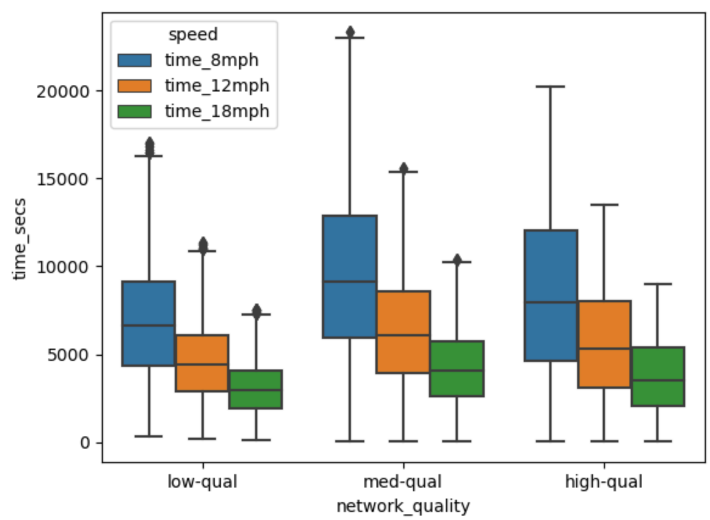
The time fields for each speed scenario were then counted for the number of connections that could be made in under 30 minutes. This represents an average commute distance, and is otherwise a reasonable measure for accessibility. Each point was given an “accessibility score”, which was equal to the number of other points that can be reached in under 30 minutes (each point had an accessibility score for each speed scenario).

Figure 4. Map of connections made under each network scenario (L-R: high, medium, low)



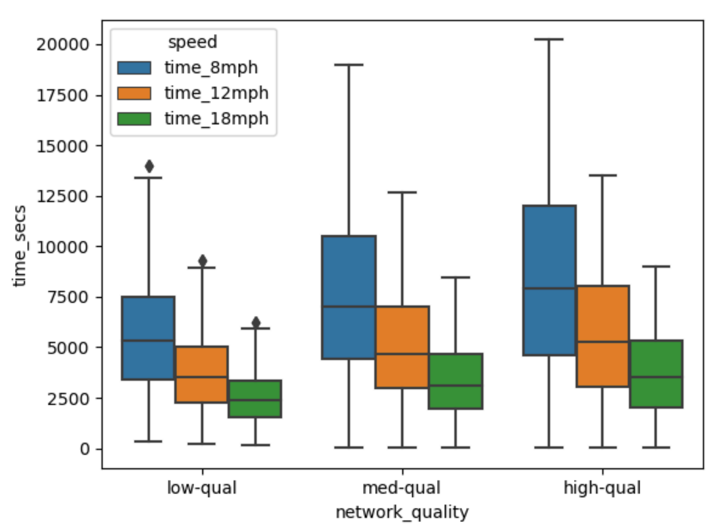
Results
First, I considered all connections (no time cutoff) and created box plots for each of the network scenarios based on the times of all connections (Figure 5). Curiously, the High quality network had lower average and quartile values than the medium quality network. I would expect that the values on the high quality network would generally be higher, as the network is more limited and fewer direct routes would be available.

Figure 5. Boxplot showing measures of central tendency for all connections



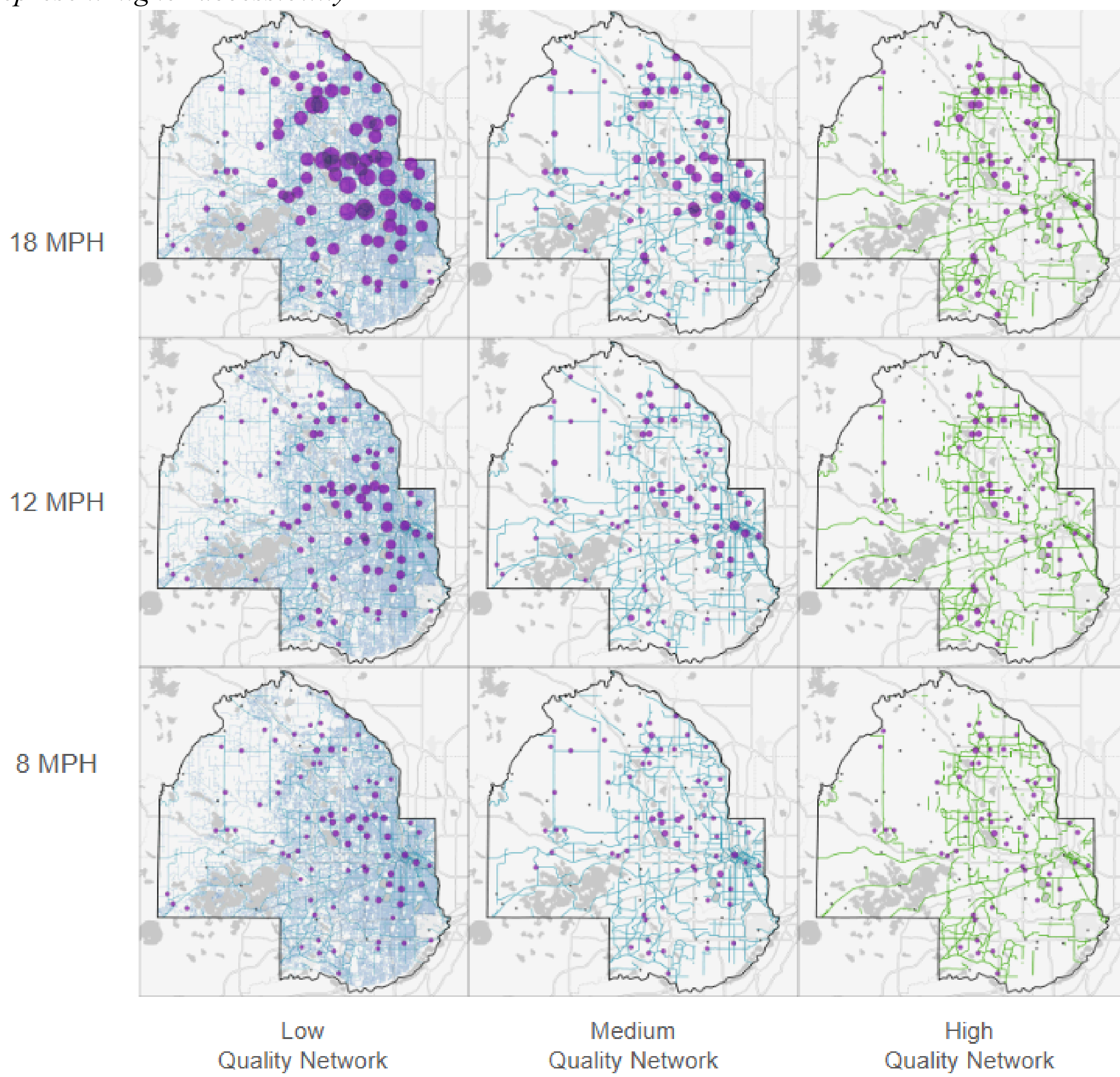
However, one thing that could be driving this is that there are simply far fewer connections being made on the High quality network, which could be skewing the results. Figure 6 shows the same measures, but only considering Origin-Destination pairs which were completed on all three networks. This removes the discrepancy, and is in line with what we would expect to see - that trip times are longer on higher quality networks relative to lower quality networks.

Figure 6. Boxplot showing measures of central tendency O-D pairs made on all three networks



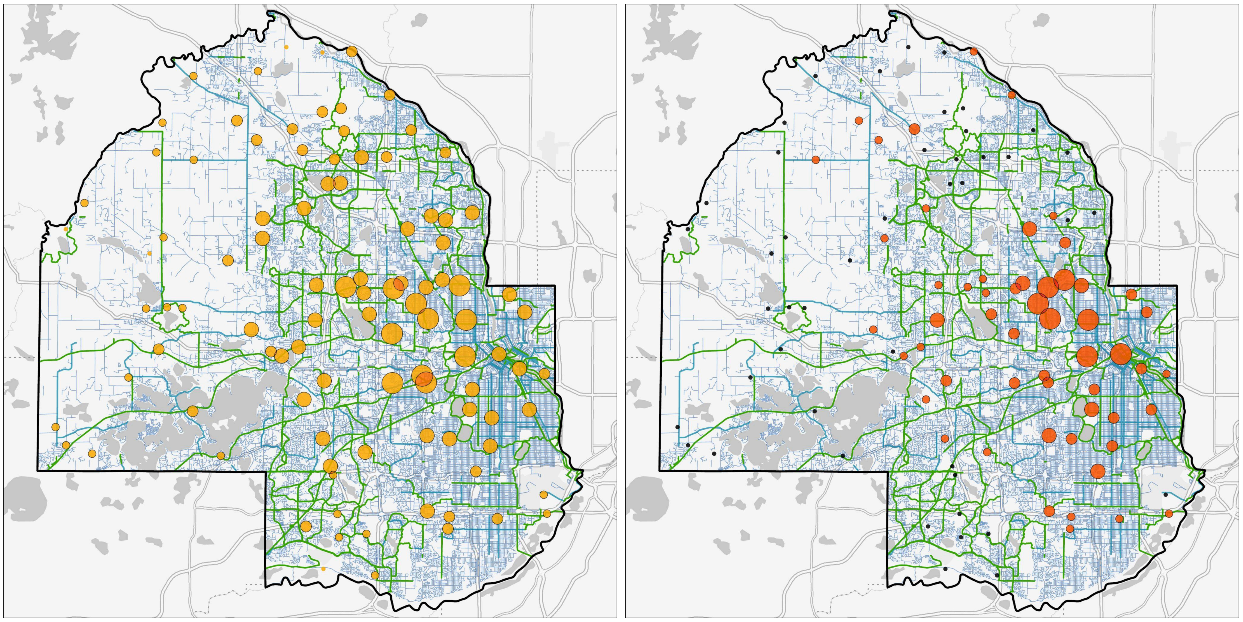
As seen in Figure 7, by symbolizing the Accessibility score for each network and speed scenario, it becomes clear that higher speeds and lower network quality allow for more connections to be made within 30 minutes. This is in line with expectations, as the low quality network is vastly more expansive than the high quality network. We can also see that the relationship between network quality and accessibility score remain negatively correlated across all the speed scenarios. This suggests that network quality is driving at least some of the changes in accessibility.

Fig 7. Accessibility scores mapped for each network and speed scenario. Larger purple dots represent higher accessibility



We can also compare the range of scores seen across the network scenarios to see where network accessibility is most reliant on network quality. Figure 8 shows a map of the total range across all speed scenarios, and the range for the networks under just one speed scenario (12 mph).

Figure 8. Right: range of accessibility score for all speeds and networks. Left: Range of accessibility scores for all networks at 12 mph.



We generally see larger dots to the north and northwest of downtown Minneapolis in both maps, suggesting that this is the area in Hennepin County for which the accessibility in network disparity is at the highest.

Results Verification

Results verification happened at several stages along the development of this analysis. The first was during network construction - I had identified several specific gaps that I noted in the line data provided by Hennepin County, and I returned to these locations after running Integrate and Snap on the feature classes to ensure the gaps were closed. Additionally, after building the networks, I used ArcGIS Pro's "Explore Network Dataset" tool to verify that connections were being made in network space before I ran any distance-based analysis.

Lastly, results are able to be verified by comparing the results to our reasonable expectations around the networks as they are constructed. I found that the high quality network generally had the lowest accessibility scores, and that the low quality network generally had the higher accessibility scores. Additionally, the fact that the relationship between network quality and accessibility was in the same direction in each speed scenario gives us confidence that the network analysis essentially worked.

Discussion and Conclusion

Overall, this analysis shows what we might intuitively expect about these three network scenarios: cyclists relying on higher quality cycling facilities have lower levels of accessibility around Hennepin County than those who are comfortable using lower quality facilities and city streets. Additionally, there are significant differences in the accessibility based on the speed at which a cyclist is able to travel via these networks. Lastly, by looking at the range of accessibility scores based on speed, we see that the near northwestern suburbs have the largest gap in accessibility between the high and low quality networks.

While the results from this project are interesting and in line with expectations, it is important to acknowledge that they are the result of many decisions I made along the way that could be considered arbitrary. I have a decent understanding of the differences between the bike facilities, but a reasonable person may decide that some of the classifications I made along the way were inaccurate or incorrect. Similarly, the decisions I made around which type of street to use for the low quality network may not have been ideal - specifically my decision to leave major local streets in, but excluding county roads. Additionally, the decisions around the Snap and Integrate steps introduced room for judgment, as well as the search tolerance for the start of trips in the Network analysis tools are all places where reasonable disagreement could prompt different decision making.

I believe these are reasonable and defensible decisions, but not perfect. Fortunately, it would be easy to modify the code to account for such differences and the analysis could be run again with ease.

In future study, it would be useful to identify specific locations, rather than random points, as the origins and destinations for study. Perhaps a popular destination-based approach would reveal differences in the results than mine based on random points. Additionally, it would be useful to develop methodology to place the points on the map according to population density, as a way to control for the amount of people likely to use the networks. Lastly, it would be helpful to explore alternative methods to closing the gaps in the network datasets that might be more consistent than the methods applied here.

Ultimately, this analysis shows that there are significant differences in the accessibility of Hennepin County by bike once network quality, cyclist confidence/comfort levels, and average speed are accounted for. We see the greatest difference in network quality to the northeast of Minneapolis, but there are differences present in all corners of the map. Importantly, differences between network quality persist under different speed scenarios, increasing the confidence that the network quality is driving some of the differences in accessibility. Further analysis of network quality could be useful for decisionmakers here in Hennepin County, but as long as detailed attribute data on facility type is available, this type of analysis should be repeatable in other jurisdictions as well.

References

Esri. Create a Network Dataset.
https://pro.arcgis.com/en/pro-app/latest/help/analysis/networks/how-to-create-a-usable-network-dataset.htm#ESRI_SECTION1_26C727D088894FA8A04C61E9963139B8

Esri. Connectivity.
https://pro.arcgis.com/en/pro-app/latest/help/analysis/networks/understanding-connectivity.htm#GUID-CC282895-4C4A-4934-B1D1-D27FC2704C77

Hennepin County. Hennepin County Bikeway and Pedestrian System.
https://gis.hennepin.us/arcgis/rest/services/HennepinData/TRANSPORTATION/MapServer/1

Hennepin County. Hennepin County Street Centerlines. https://gis.hennepin.us/arcgis/rest/services/HennepinData/TRANSPORTATION/MapServer/0

Metro GIS. Metro Regional Centerline Collaborative (MRCC). Accessed on December 1, 2024. https://metrogis.org/projects/metro-centerlines/

Self-score

Category	Description	Points Possible	Score
Structural Elements	All elements of a lab report are included **(2 points each)**: Title, Notice: Dr. Bryan Runck, Author, Project Repository, Date, Abstract, Problem Statement, Input Data w/ tables, Methods w/ Data, Flow Diagrams, Results, Results Verification, Discussion and Conclusion, References in common format, Self-score	28	28
Clarity of Content	Each element above is executed at a professional level so that someone can understand the goal, data, methods, results, and their validity and implications in a 5 minute reading at a cursory-level, and in a 30 minute meeting at a deep level **(12 points)**. There is a clear connection from data to results to discussion and conclusion **(12 points)**.	24	24
Reproducibility	Results are completely reproducible by someone with basic GIS training. There is no ambiguity in data flow or rationale for data operations. Every step is documented and justified.	28	26
Verification	Results are correct in that they have been verified in comparison to some standard. The standard is clearly stated **(10 points)**, the method of comparison is clearly stated **(5 points)**, and the result of verification is clearly stated **(5 points)**.	20	20
		100	98