My title*

My subtitle

Ziheng Zhong

September 23, 2024

My abstract

Table of contents

1 Introduction **2**

2 Data **2**
2.1 Source 2
2.2 Measurement 2
2.3 Method 3

3 Results **5**
3.1 Data Trend 5
3.2 Distributions 5
3.3 Maps 5

4 Discussion **8**
4.1 First discussion point 8
4.2 Second discussion point 8
4.3 Third discussion point 8
4.4 Weaknesses and next steps 8

A Appendix **9**

References **10**

*Code and data are available at: https://github.com/iJustinn/Toronto_Cycling_Network

1 Introduction

You can and should cross-reference sections and sub-sections. We use R Core Team (2023).

The remainder of this paper is structured as follows.

2 Data

2.1 Source

The data used in this paper was collected by the OpenDataToronto Library (Gelfand 2020). OpenDataToronto provides a platform for the public to access various datasets related to Toronto's civic operations and urban infrastructure. The specific dataset used in this research is the 'Cycling Network' (Data 2024), which provides detailed information on Toronto's bicycle infrastructure, including dedicated lanes, multi-use trails, and shared roadways. This dataset plays a crucial role in promoting active transportation and urban sustainability efforts within the city. It is frequently updated to reflect ongoing expansions and modifications to the cycling infrastructure, aligning with Toronto's broader efforts to reduce congestion, enhance mobility, and support environmental goals. This dataset not only informs local decisions but also contributes to broader discussions on urban transportation planning and sustainability in growing cities around the world.

Data used in this paper was downloaded, cleaned and analyzed with the programming language R (R Core Team 2023). Also with support of additional packages which will be talked about in the Section 2.3.

2.2 Measurement

The 'Cycling Network' dataset from Open Data Toronto tracks various aspects of Toronto's cycling infrastructure, such as installation dates, upgrades, street names, and the type of infrastructure (e.g., sharrows, multi-use trails). Each entry represents a segment of the cycling network, where real-world phenomena, such as the construction or upgrading of cycling paths, are documented. For example, when a cycling path is installed or upgraded, the responsible authorities collect information such as the installation year (e.g., 2001) and any subsequent upgrades (e.g., 2021). This data is then digitized into structured entries in the dataset.

To go from a physical event (e.g., the installation of a bike lane) to a dataset entry, detailed records are maintained by the city's transportation department. These records are geospatially coded (column geometry in this dataset), allowing each cycling path to be mapped precisely in relation to other city infrastructure. Thus, the phenomena of constructing or updating a bike lane becomes an entry with attributes such as the street name, road class, and installation

Table 1: Raw Coordinates

[!h]

Simplified Geometry
{'type': 'MultiLineString', 'coordinates': [[[-79.4035069136297, 43.6952595244941], [-79.403
{'type': 'MultiLineString', 'coordinates': [[[-79.4036386937994, 43.6349938275931], [-79.403
{'type': 'MultiLineString', 'coordinates': [[[-79.2752236011477, 43.7415816548541], [-79.275
{'type': 'MultiLineString', 'coordinates': [[[-79.4677191149987, 43.7720763078272], [-79.467
{'type': 'MultiLineString', 'coordinates': [[[-79.5481155366534, 43.5902679719901], [-79.548
{'type': 'MultiLineString', 'coordinates': [[[-79.5430542283657, 43.58695658908], [-79.54307

Table 2: Extracted and Cleaned Coordinates

Longitude	Latitude	ID
-79.40351	43.69526	1
-79.40309	43.69535	1
-79.40240	43.69549	1
-79.40364	43.63499	2
-79.40355	43.63519	2
-79.40346	43.63535	2

history. This allows researchers to analyze trends in cycling infrastructure and its expansion over time.

2.3 Method

Data used in this paper was cleaned, processed, modeled and tested with the programming language R (R Core Team 2023). Also with support of additional packages in R: `sf` [], `readr` (Wickham, Hester, and Bryan 2023), `ggplot2` (Wickham 2016), `osmdata` [], `tidyverse` (Wickham et al. 2019), `jsonlite` [], `dplyr` (Wickham et al. 2023), `here` [], `knitr` [], `kableExtra` [], `[],` [], .

In the first data cleaning process, the focus was on extracting and cleaning geographical coordinates from the 'geometry' column, which contains location data in JSON-like format, Table 1 displays how the data was like. A custom function was created to parse the coordinates, filter out invalid entries, and extract the longitude and latitude values. The extracted coordinates were combined into a new dataset shon in Table 2, with unique IDs to match the original entries. Any rows with missing coordinate data were removed, and the cleaned data was saved into a separate file for further analysis.

Table 3: Number of Lane Upgrades by Year

Year	Number of Upgraded Lanes
2018	9
2019	14
2020	46
2021	28
2022	15
2023	12

Table 4: Number of Bikeways Installed by Year

Year	Number of Installed Bikeways
2001	539
2002	12
2003	15
2004	20
2005	65
2006	75

The second cleaning process dealt with the 'UPGRADED' column, which contains information about the years when cycling lanes were upgraded. Invalid or missing years were filtered out, and the remaining valid data was converted to a numeric format. The dataset was then grouped by the year of upgrade, with a summary created to count the number of lanes upgraded in each year. This summarized data, Table 3, was saved for use in analyzing trends in lane upgrades over time.

In the third cleaning process, the focus was on the 'INSTALLED' column, which records the years when cycling infrastructure was first installed. Similar to the 'UPGRADED' column, invalid or missing years were removed, and the data was converted to numeric format. The installation data was grouped by year to count the number of bikeways installed each year. The resulting summary, Table 4, was saved for further trend analysis on the growth of the cycling network.

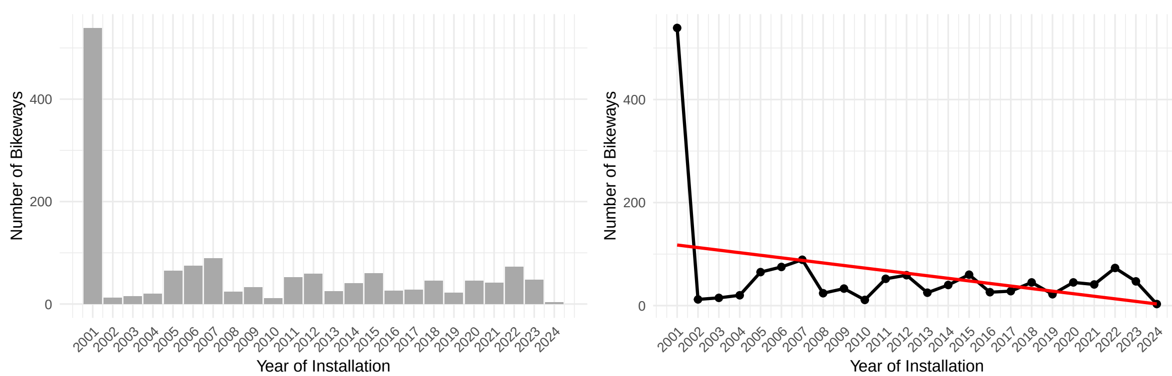
Finally, the fourth cleaning process involved classifying cycling lanes based on the type of infrastructure. Data from the 'INFRA_HIGHORDER' and 'INFRA_LOWORDER' columns was cleaned, and missing values were removed. The lanes were classified into three comfort levels: High Comfort, Moderate Comfort, and Low Comfort, depending on whether the infrastructure was protected or involved bike lanes. The classification, along with year data for installation and upgrades, was saved for analysis on the types of cycling lanes and their comfort levels. Table 5 shows how it looks like.

Table 5: Classification of Cycling Lanes by Comfort Level

Installed Year	Upgraded Year	Comfort Level
2001	2021	Low Comfort
2001	2009	High Comfort
2001	2011	High Comfort
2001	2011	High Comfort
2001	2012	High Comfort
2001	2012	High Comfort

3 Results

3.1 Data Trend



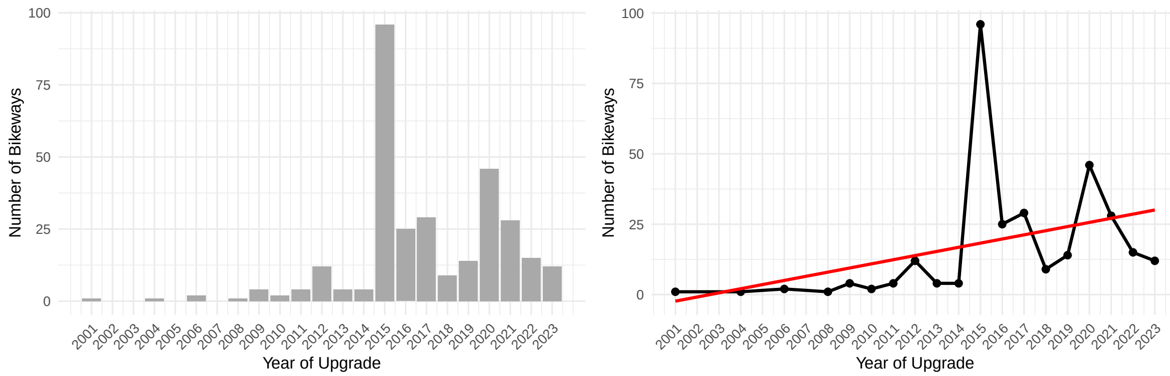
(a) Number (b) Trend

Figure 1: Toronto Bikeways Installed by Year

3.2 Distributions

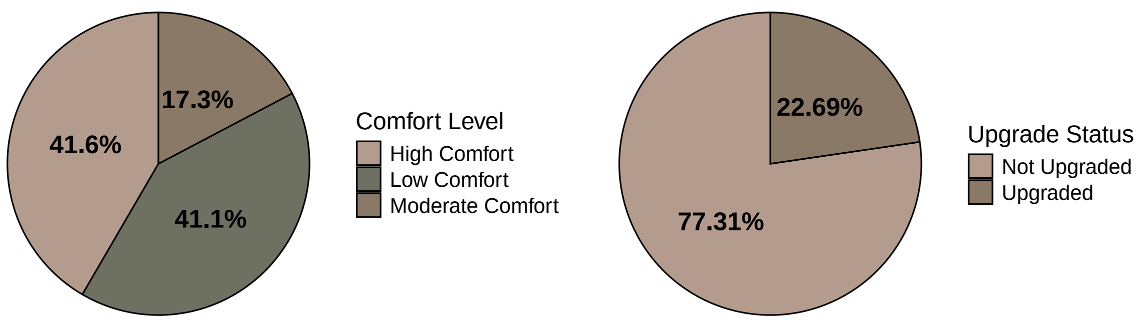
3.3 Maps

Figure 4 provides a visual representation of the bike lane infrastructure across Toronto. The lines on the map correspond to various cycling paths, including dedicated bike lanes, multi-use paths, and shared roadways, which span different areas of the city. The geographic coordinates, ranging from around 43.60°N to 43.85°N in latitude and -79.6°W to -79.3°W in longitude, encompass most of Toronto's urban region. This map offers an important spatial perspective, illustrating the extensive network of cycling routes that connect different neighborhoods and areas within the city.

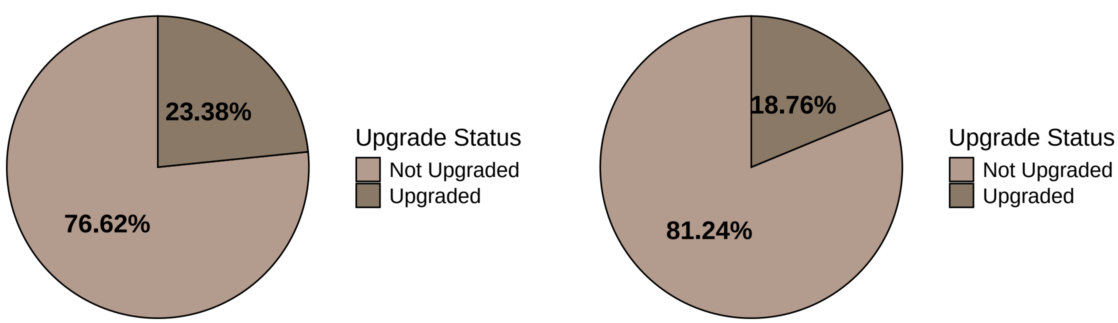


(a) Number (b) Trend

Figure 2: Toronto Bikeways Upgraded by Year



(a) All levels (b) Low Comfort



(c) Moderate Comfort (d) High Comfort

Figure 3: Toronto Bikeway Comfort Level Distribution

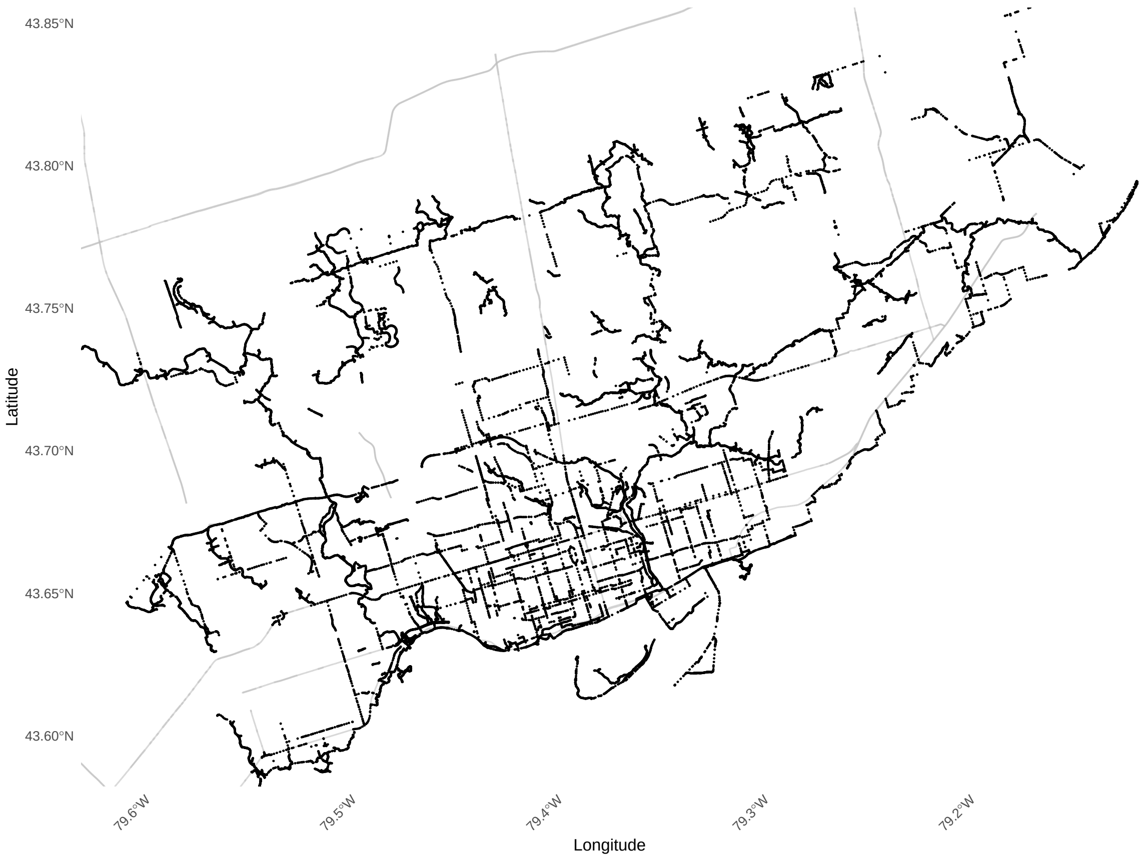


Figure 4: Map of Bike Lanes Locations in Toronto

A key observation from the map is the dense concentration of bike lanes in the downtown core and along major roadways, indicating a well-established network in central Toronto. The downtown area's grid-like structure is particularly evident, showcasing a high density of lanes that facilitate easy cycling access within the city's busiest areas. As one moves further away from the city center, the map shows fewer bike lanes, suggesting that suburban regions may have less developed cycling infrastructure. However, there are notable multi-use paths extending toward the outskirts, particularly in parks and recreational areas, providing important cycling connections for those commuting or using bikes for recreational purposes.

This map highlights the importance of cycling infrastructure in promoting sustainable urban transportation. The wide distribution of bike lanes supports Toronto's efforts to enhance active transportation options, reduce traffic congestion, and improve accessibility for cyclists. The representation also provides a foundation for analyzing areas where future infrastructure developments could be prioritized, particularly in regions where cycling connectivity appears sparse.

4 Discussion

4.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

4.2 Second discussion point

4.3 Third discussion point

4.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

A Appendix

References

Data, Toronto Open. 2024. "Cycling Network." https://open.toronto.ca/dataset/cycling-network/.

Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://cran.r-project.org/package=opendatatoronto.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2023. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.