MiWay Transit Access for UTM Students: District-Level Analysis

Felix Gao
Zhuoran Zhu
Feiyi Dai

STA304H5 – Data Analysis Project

November 13, 2025

Abstract

Mississauga's MiWay transit system plays a major role in how University of Toronto Mississauga (UTM) students commute to campus. In this project, we use the City of Mississauga's MiWay Bus Stops dataset to compare transit accessibility across districts, with a focus on direct UTM service and GO Train connectivity. We analyze bus stops across the city and summarize district-level differences in accessibility, shelter coverage, route diversity, direct UTM access, and GO connectivity. For each research question, we compute conditional proportions, create visual summaries, and conduct chi-square tests of association to assess whether these characteristics vary by district. We find strong evidence that accessibility, shelter availability, UTM service, and GO connectivity all differ substantially across Mississauga. Districts along Dundas Street and Mississauga Road provide the strongest direct access to UTM, while districts containing GO stations have more GO-serving bus routes. These findings highlight important transit inequalities that affect where UTM students may choose to live.

Introduction

Public transit availability can strongly influence how easily University of Toronto Mississauga (UTM) students can travel between home, campus, and regional transportation hubs. MiWay provides local and express routes across the city, including several routes that serve UTM directly and others that connect to GO Train stations. Understanding how these routes and stops are distributed geographically is useful for students deciding where to live.

This project uses the **City of Mississauga MiWay Bus Stops dataset**, which includes one row per bus stop and information on district, accessibility, shelter availability, and routes. We cleaned the dataset by removing missing or "UNKNOWN" district labels (primarily GO platforms or boundary stops) and separated the route list into a long format.

We address four research questions:

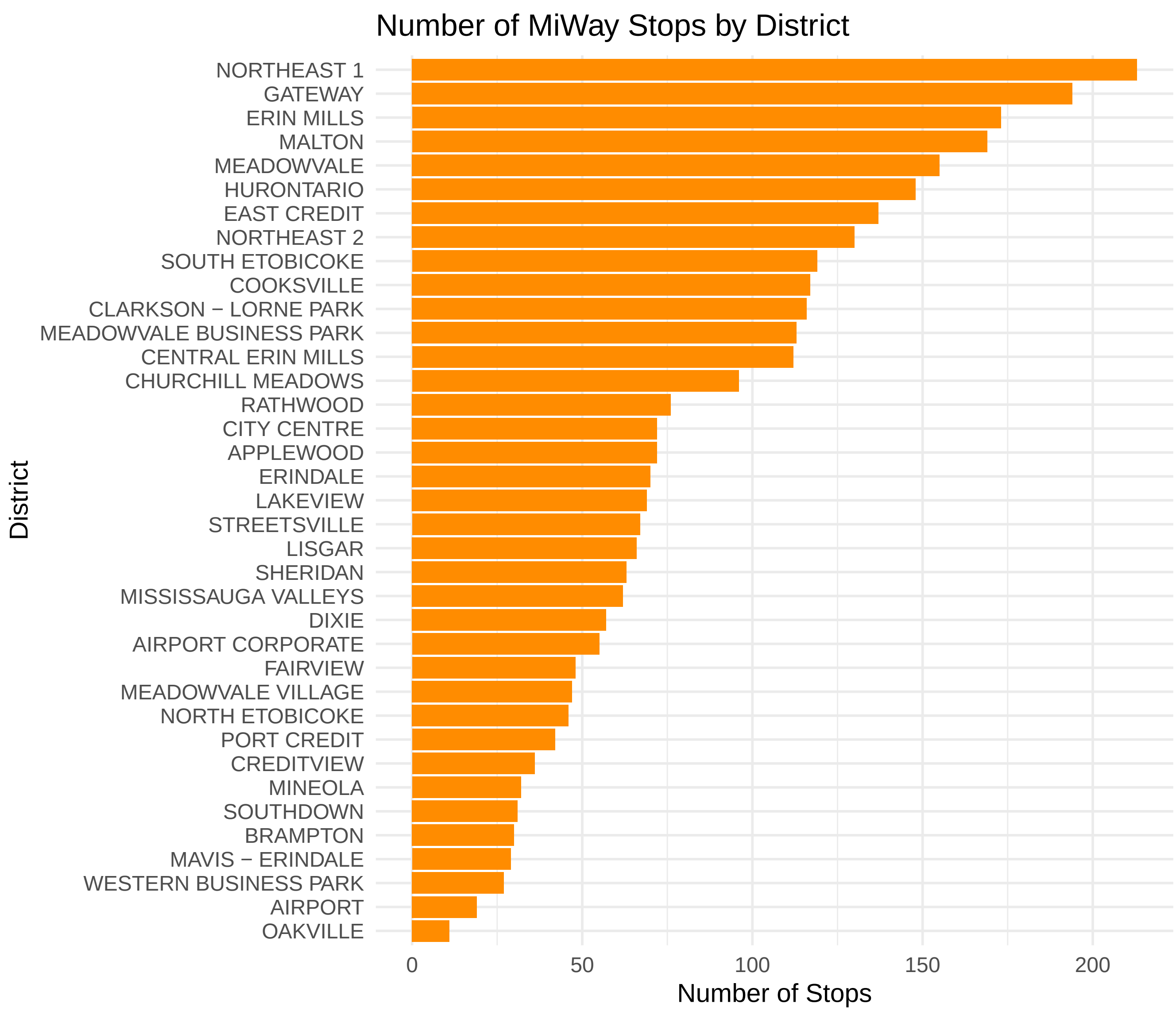
1. **How does MiWay stop coverage vary across districts?**
2. **Does accessibility and shelter coverage differ across districts?**
3. **Which districts have the strongest direct connections to UTM?**
4. **Which districts have the best connectivity to GO Train stations?**

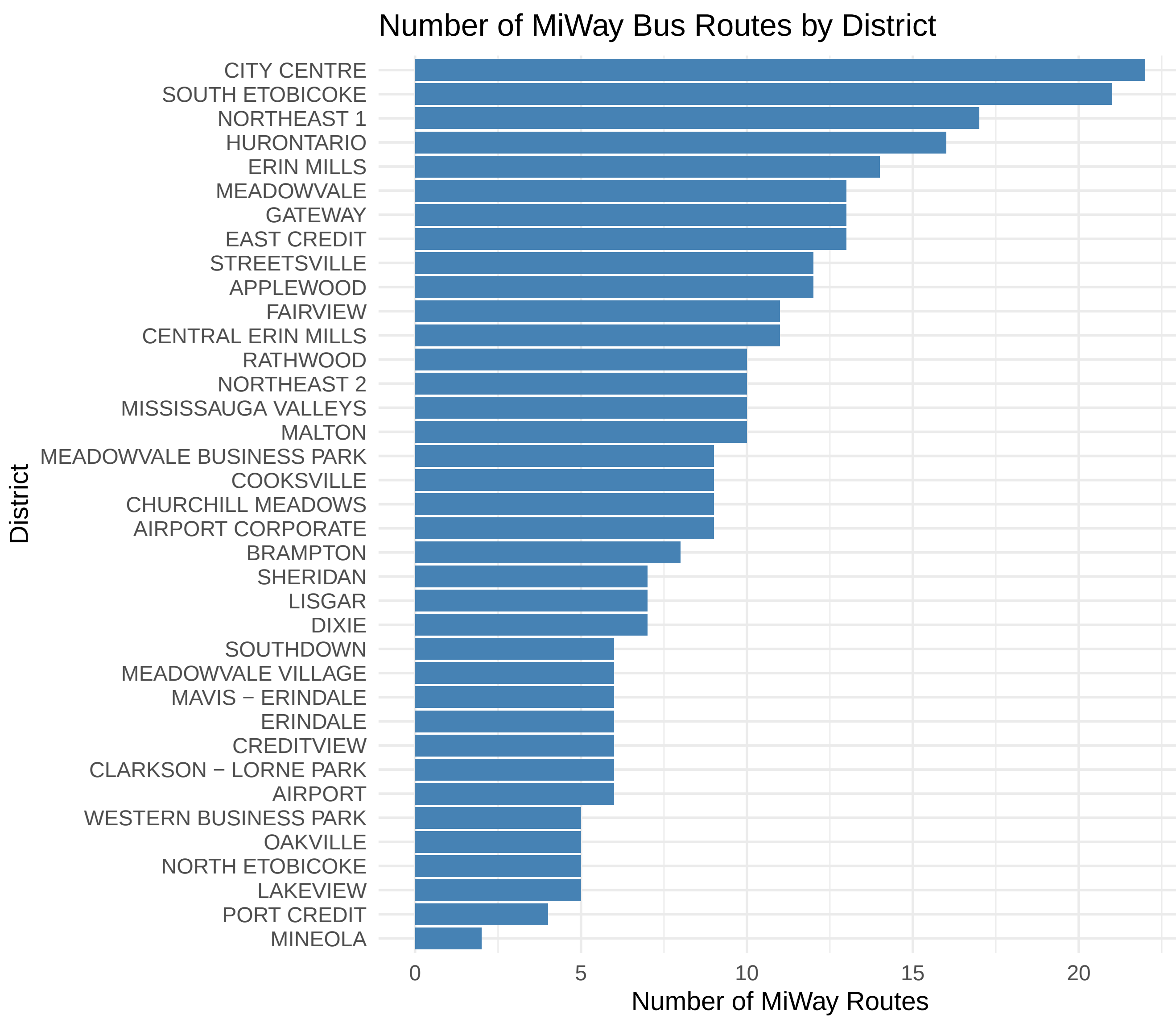
Each question is answered using summary statistics, bar charts, and chi-square tests.

Results

1. MiWay Coverage Across Districts

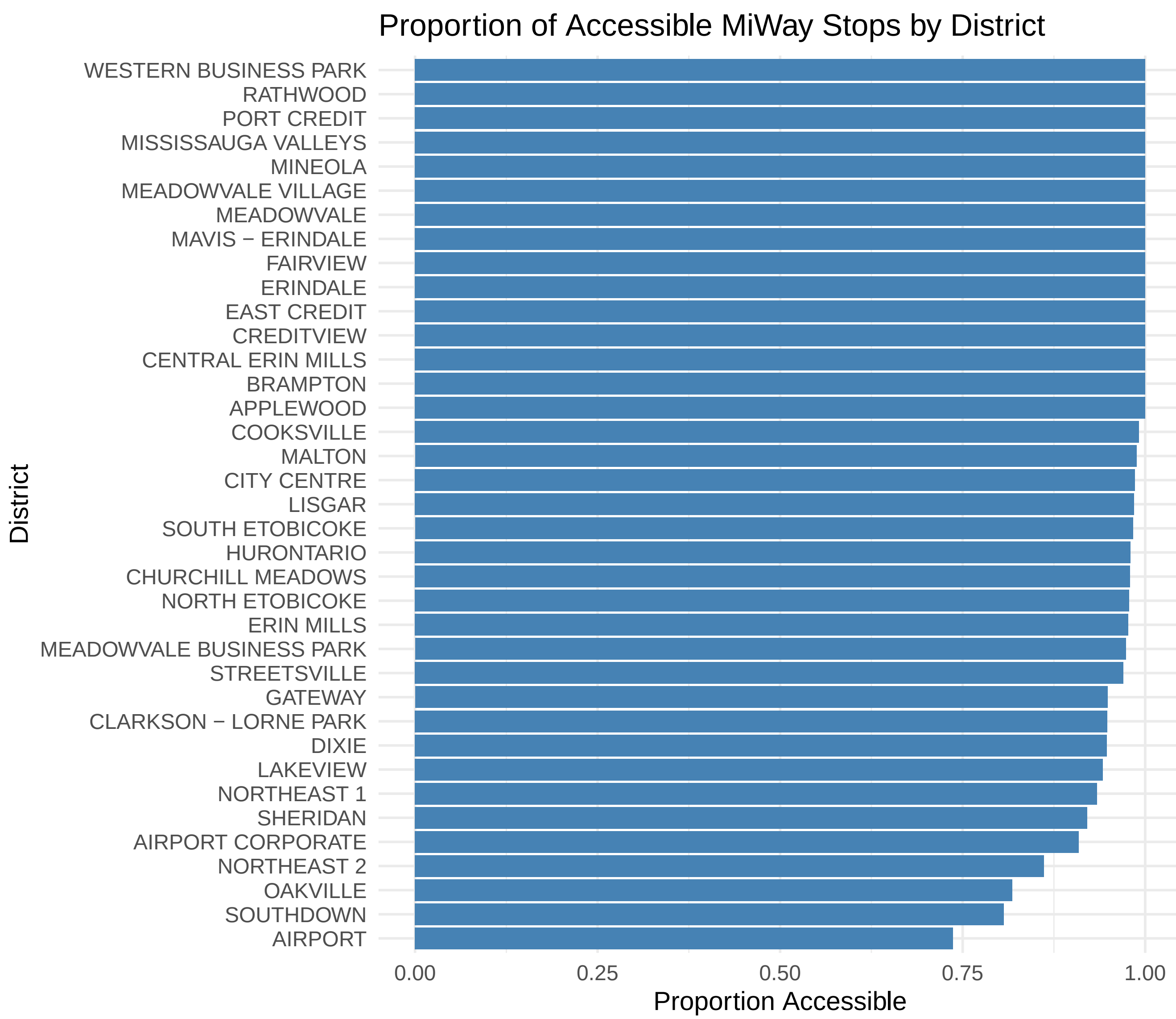
Districts differ widely in the number of MiWay bus stops they contain. Some districts along major corridors have significantly more stops, while others have fewer. Route diversity also varies, with some districts having many distinct routes and others offering fewer options.





2. Accessibility and Shelter Coverage

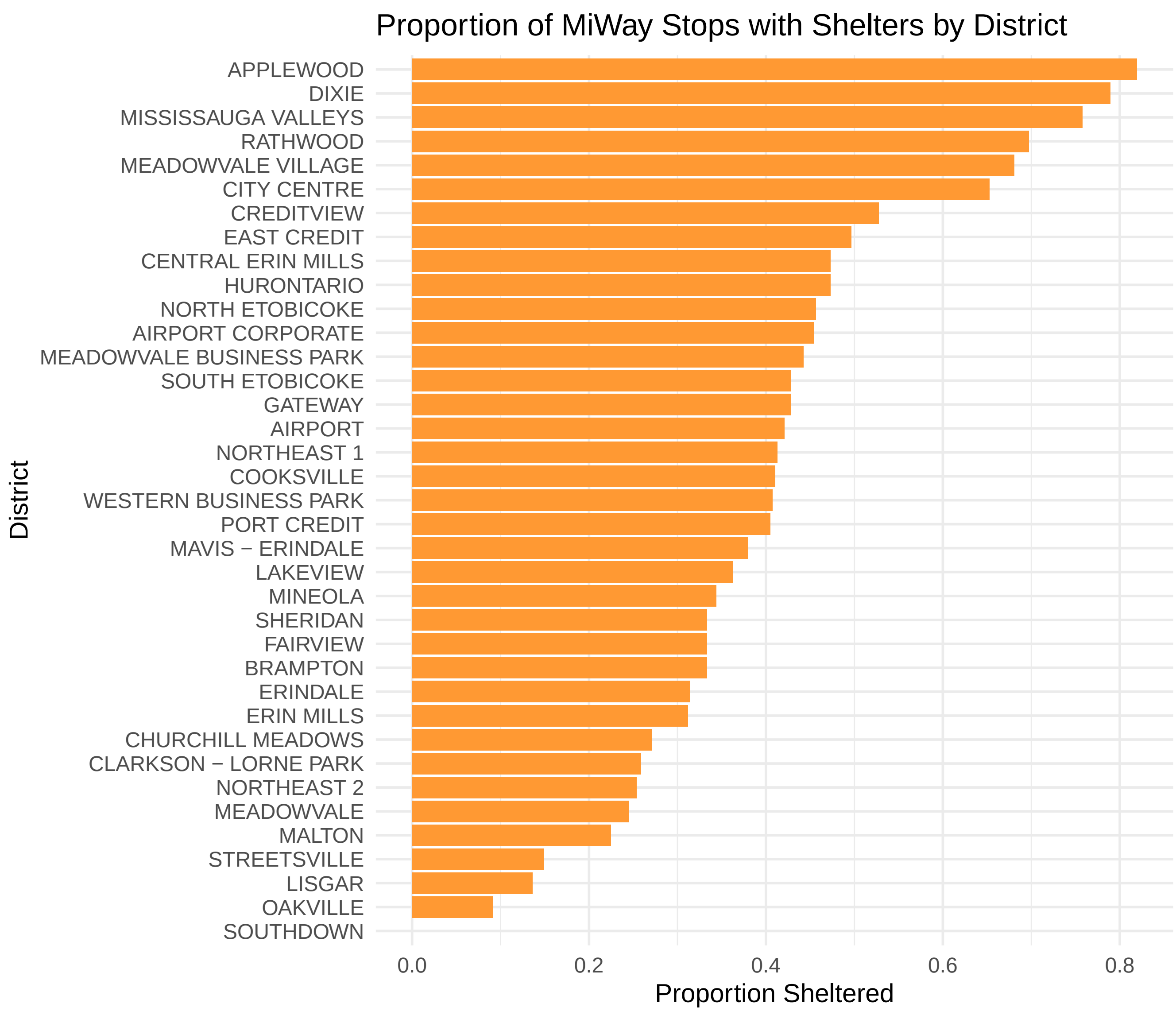
Most MiWay stops are designed so people with mobility challenges can board the bus more easily. Because of this city-wide design standard, almost every district has very high accessibility rates.




```
##
##  Pearson's Chi-squared test
##
## data:  tab_access
## X-squared = 179.32, df = 36, p-value < 2.2e-16
```

The chi-square test shows that accessibility is not evenly distributed across districts ($p < 0.001$). Although most stops in Mississauga are accessible, some districts have slightly higher or lower proportions than others, meaning accessibility is associated with district location.

Shelter coverage varies more widely:

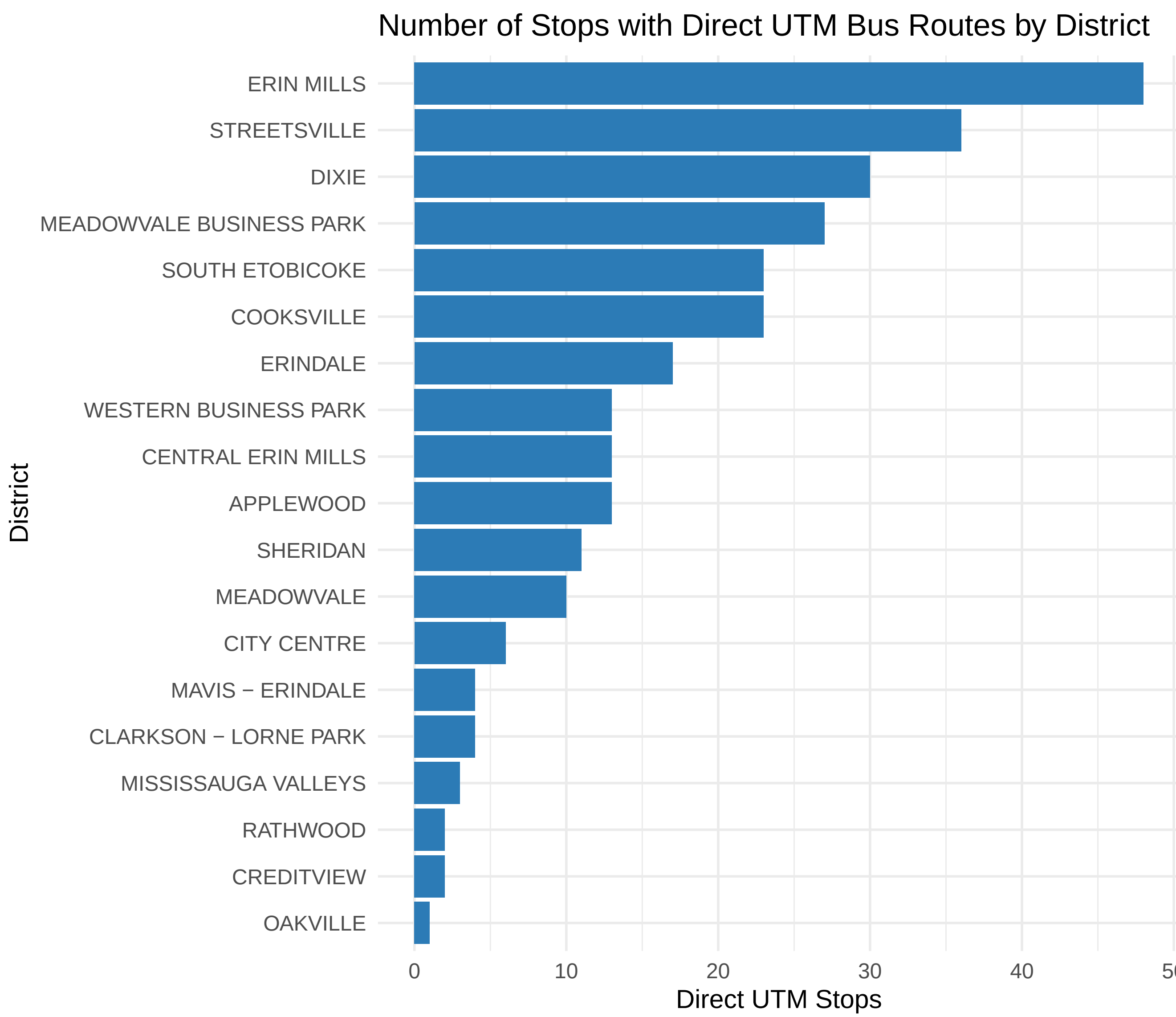



```
##
##  Pearson's Chi-squared test
##
## data:  tab_shelter
## X-squared = 339.18, df = 36, p-value < 2.2e-16
```

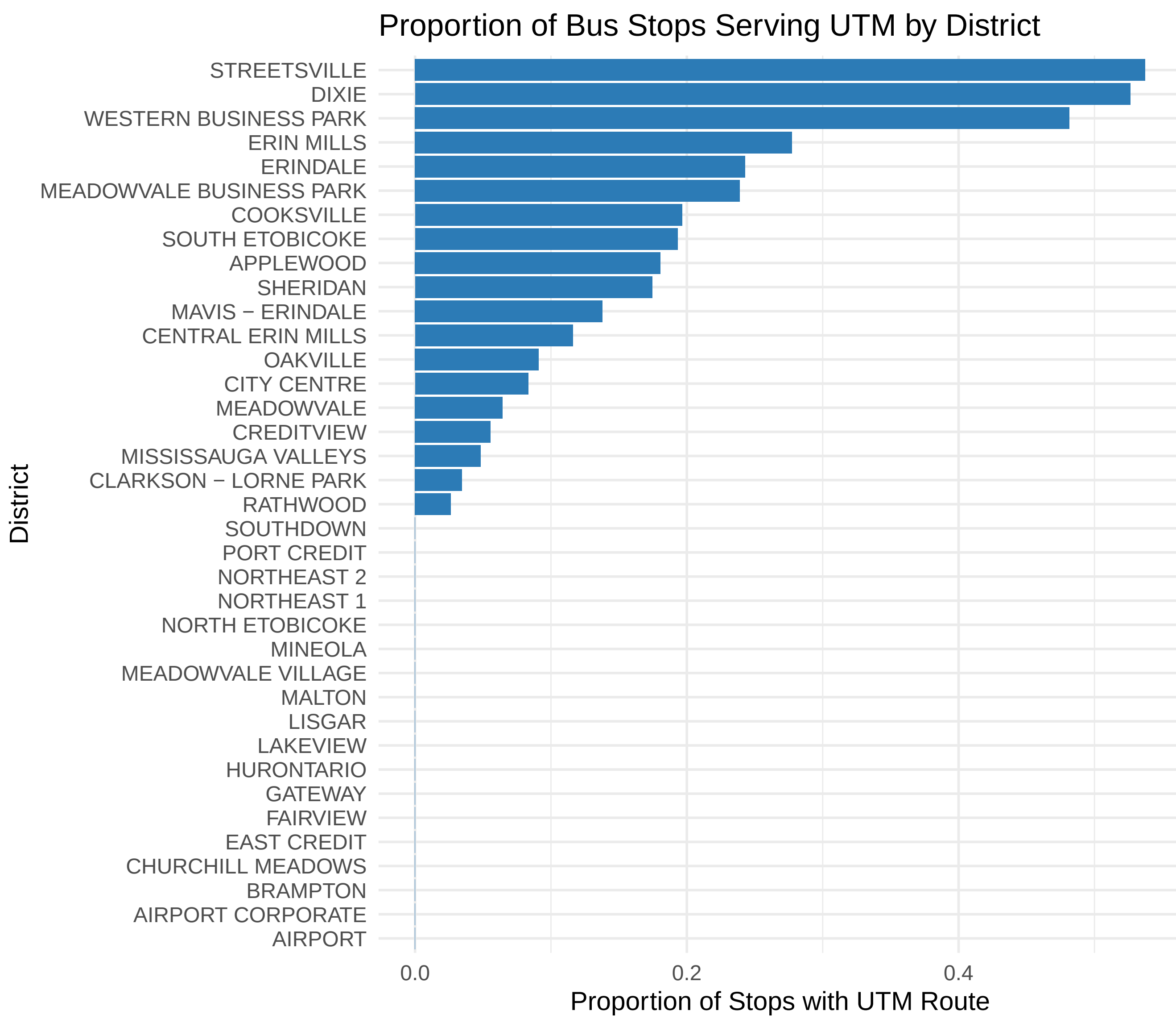
The chi-square test indicates that shelter coverage differs significantly across districts ($p < 0.001$). Some districts provide many more sheltered stops than others, showing that weather protection is not evenly available across the city.

3. Direct UTM Access

Direct UTM routes (1, 44, 101, 110, 126) concentrate in districts along Dundas and Mississauga Road.



Normalized rates provide a fairer comparison:

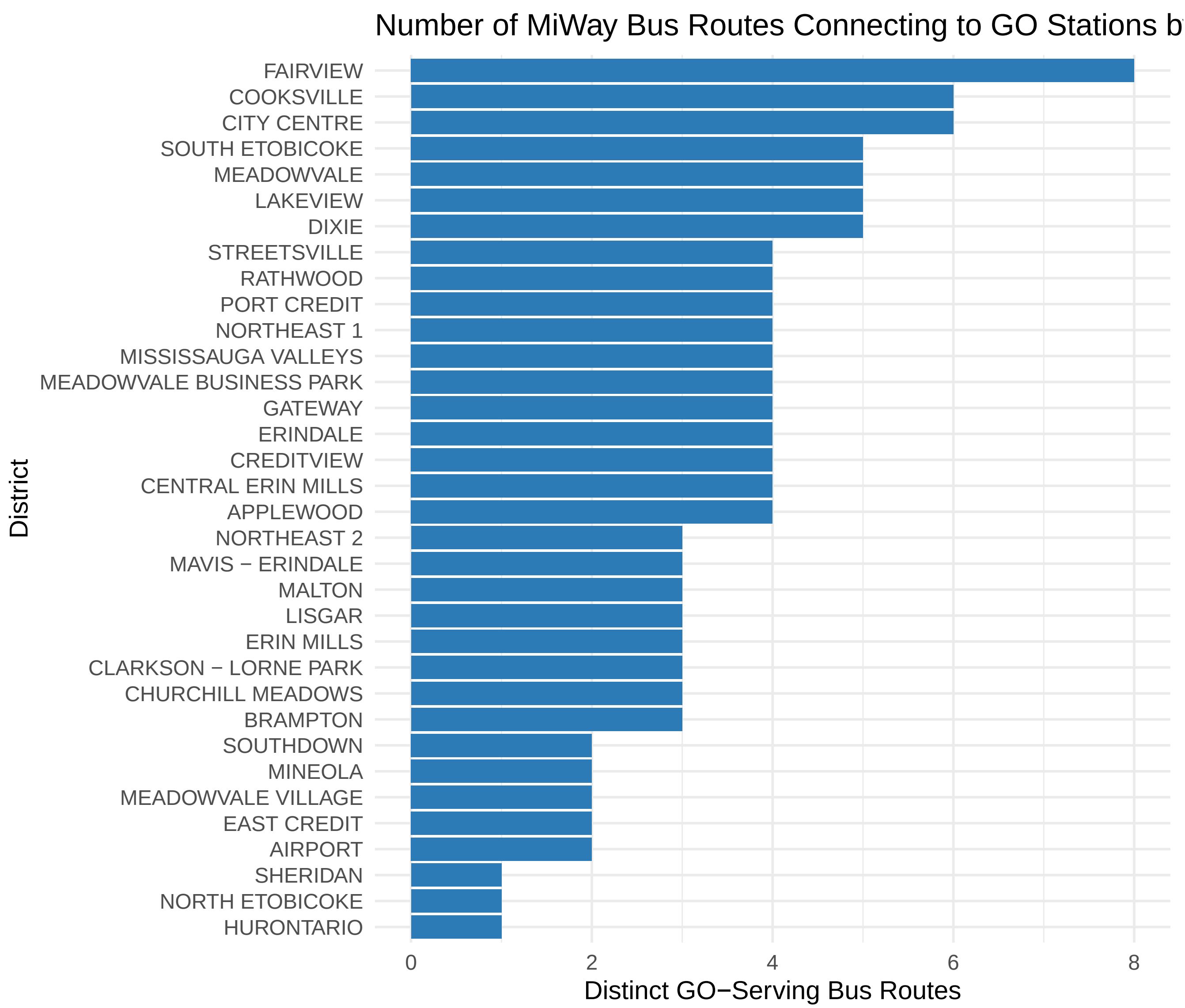



```
##
##  Pearson's Chi-squared test
##
## data:  tab_utm
## X-squared = 672.88, df = 36, p-value < 2.2e-16
```

The chi-square test shows a strong association between district and whether a stop serves a UTM route ($p < 0.001$). Districts along Dundas Street and Mississauga Road have far more UTM-serving stops than other districts, meaning direct access to campus varies considerably by location.

4. GO Train Connectivity

Districts near Cooksville, Port Credit, and Meadowvale GO stations have more GO-serving routes.




```
##
##  Pearson's Chi-squared test
##
## data:  tab_go
## X-squared = 139.69, df = 36, p-value = 3.843e-14
```

The chi-square test finds a significant relationship between district and whether a stop is located at a GO station ($p < 0.001$). Only a few districts contain GO stations, so GO-connected routes are concentrated in those districts rather than being evenly distributed.

Conclusion

Our analysis shows clear and statistically significant differences in MiWay transit access across Mississauga's districts. The chi-square tests for accessibility, shelter coverage, direct UTM service, and GO connectivity all provided very small p-values, indicating that these features are not evenly distributed across the city. Some districts, especially those along Dundas Street and Mississauga Road, have more stops that directly serve UTM, while districts containing GO stations have a greater number of routes that connect to regional rail. In contrast, other districts have fewer sheltered stops or fewer direct UTM and GO connections, which may make commuting less convenient or comfortable for students living there.

These patterns matter in practice because UTM students often choose where to live based on a mix of rent, safety, and transit access. Our results suggest that students who live in districts with strong UTM and GO connectivity may have more reliable and flexible commute options, while those in less-served areas may face longer or less convenient trips. At the same time, accessibility proportions are high across most districts, reflecting AODA requirements and an aging population, so the main differences lie in comfort (shelters) and connectivity (UTM and GO routes) rather than basic physical access.

There are several important limitations. First, our analysis uses only stop locations and route presence, not schedule or frequency information, so we cannot distinguish between a route that runs every few minutes and one that runs infrequently. Second, we do not incorporate travel time or crowding, which also affect how attractive a commute is. Finally, we work with district-level summaries, which may hide variation within districts.

Future work could address these gaps by incorporating GTFS schedule data to study service frequency and expected travel times, constructing a composite "transit access score" that combines shelters, accessibility, route diversity, UTM direct service, and GO connectivity, and linking MiWay data with student survey data on actual commute experiences. Together, these extensions would provide an even clearer picture of how transit infrastructure shapes the daily lives and choices of UTM students.

Appendix

```r
# packages
library(dplyr)
library(ggplot2)
library(tidyr)

# read data
df <- read.csv("./Datasets/MiWay_Bus_Stops_-4490514804018385747.csv")

# clean districts
df <- df %>%
  mutate(District = trimws(District)) %>%
  filter(District != "", District != "UNKNOWN")

# UTM routes (MiWay routes that go to UTM)
utm_routes <- c("1", "44", "101", "110", "126")

# flag GO station stops
df$is_GO_stop <- grepl("GO STATION", df$Stop_Description, ignore.case = TRUE)

# one row per stop-route
routes_long <- df %>%
  separate_rows(Routes_Service, sep = "/") %>%
  mutate(Routes_Service = trimws(Routes_Service)) %>%
  filter(Routes_Service != "")

# stop-level data: whether each stop has a UTM route
stop_level <- routes_long %>%
  group_by(Stop_Number, District) %>%
  summarise(
    has_utm = any(Routes_Service %in% utm_routes),
    .groups = "drop"
  )

# routes that touch a GO station at least once
go_routes <- df %>%
  filter(is_GO_stop) %>%
  separate_rows(Routes_Service, sep = "/") %>%
  mutate(Routes_Service = trimws(Routes_Service)) %>%
  filter(Routes_Service != "") %>%
  pull(Routes_Service) %>%
  unique()

# summaries by district
counts <- df %>%
  group_by(District) %>%
  summarise(n = n()) %>%
  arrange(desc(n))

access_summary <- df %>%
  group_by(District) %>%
  summarise(
```

```
    accessible = sum(Accessible),
    total      = n(),
    proportion = accessible / total
  ) %>%
  arrange(desc(proportion))

shelter_summary <- df %>%
  group_by(District) %>%
  summarise(
    shelters             = sum(Shelter),
    total                = n(),
    proportion_sheltered = shelters / total
  ) %>%
  arrange(desc(proportion_sheltered))

routes_by_district <- routes_long %>%
  group_by(District) %>%
  summarise(n_routes = n_distinct(Routes_Service)) %>%
  arrange(desc(n_routes))

utm_long <- routes_long %>%
  filter(Routes_Service %in% utm_routes)

district_utm <- utm_long %>%
  group_by(District) %>%
  summarise(
    utm_stops           = n_distinct(Stop_Number),
    utm_bus_occurrences = n(),
    utm_distinct_routes = n_distinct(Routes_Service),
    .groups = "drop"
  ) %>%
  arrange(desc(utm_stops))

utm_normalized <- stop_level %>%
  group_by(District) %>%
  summarise(
    total_stops = n(),
    utm_stops   = sum(has_utm),
    utm_rate    = utm_stops / total_stops,
    .groups = "drop"
  ) %>%
  arrange(desc(utm_rate))

go_by_district <- routes_long %>%
  filter(Routes_Service %in% go_routes) %>%
  group_by(District) %>%
  summarise(
    distinct_go_routes = n_distinct(Routes_Service),
    total_go_stops     = n_distinct(Stop_Number),
    .groups = "drop"
  ) %>%
  arrange(desc(distinct_go_routes))
```

```r
# chi-square tests
tab_access   <- table(df$District, df$Accessible)
chisq_access <- chisq.test(tab_access)

tab_shelter   <- table(df$District, df$Shelter)
chisq_shelter <- chisq.test(tab_shelter)

tab_utm   <- table(stop_level$District, stop_level$has_utm)
chisq_utm <- chisq.test(tab_utm)

tab_go   <- table(df$District, df$is_GO_stop)
chisq_go <- chisq.test(tab_go)
```