

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.

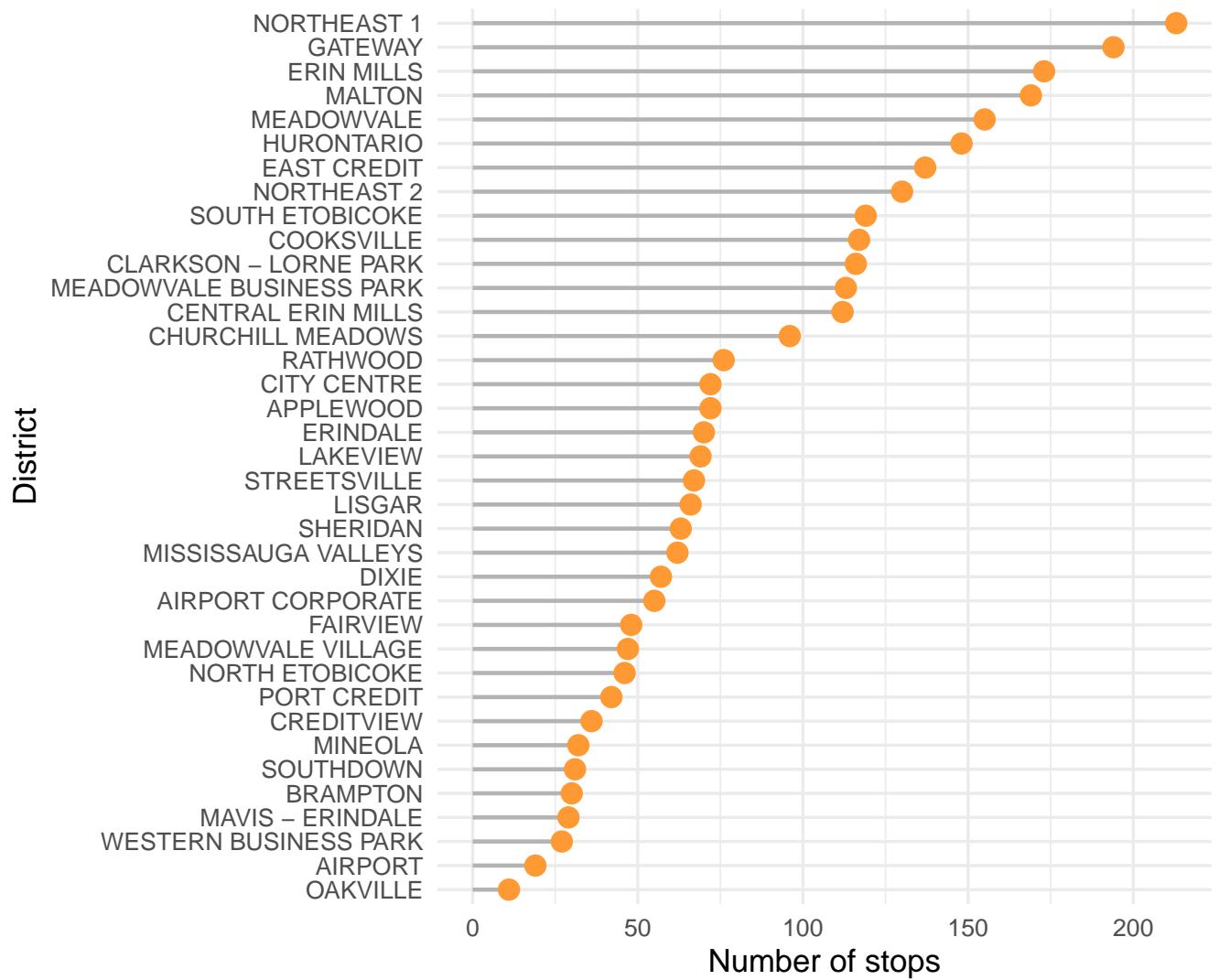
Results

In this section we move from a broad picture of MiWay coverage to more specific features that matter for UTM students. We first describe how many stops and routes each district has. We then compare accessibility and shelter coverage, before turning to direct UTM service and GO Train connectivity. We finish with an overall view that combines these features into a simple transit profile for each district.

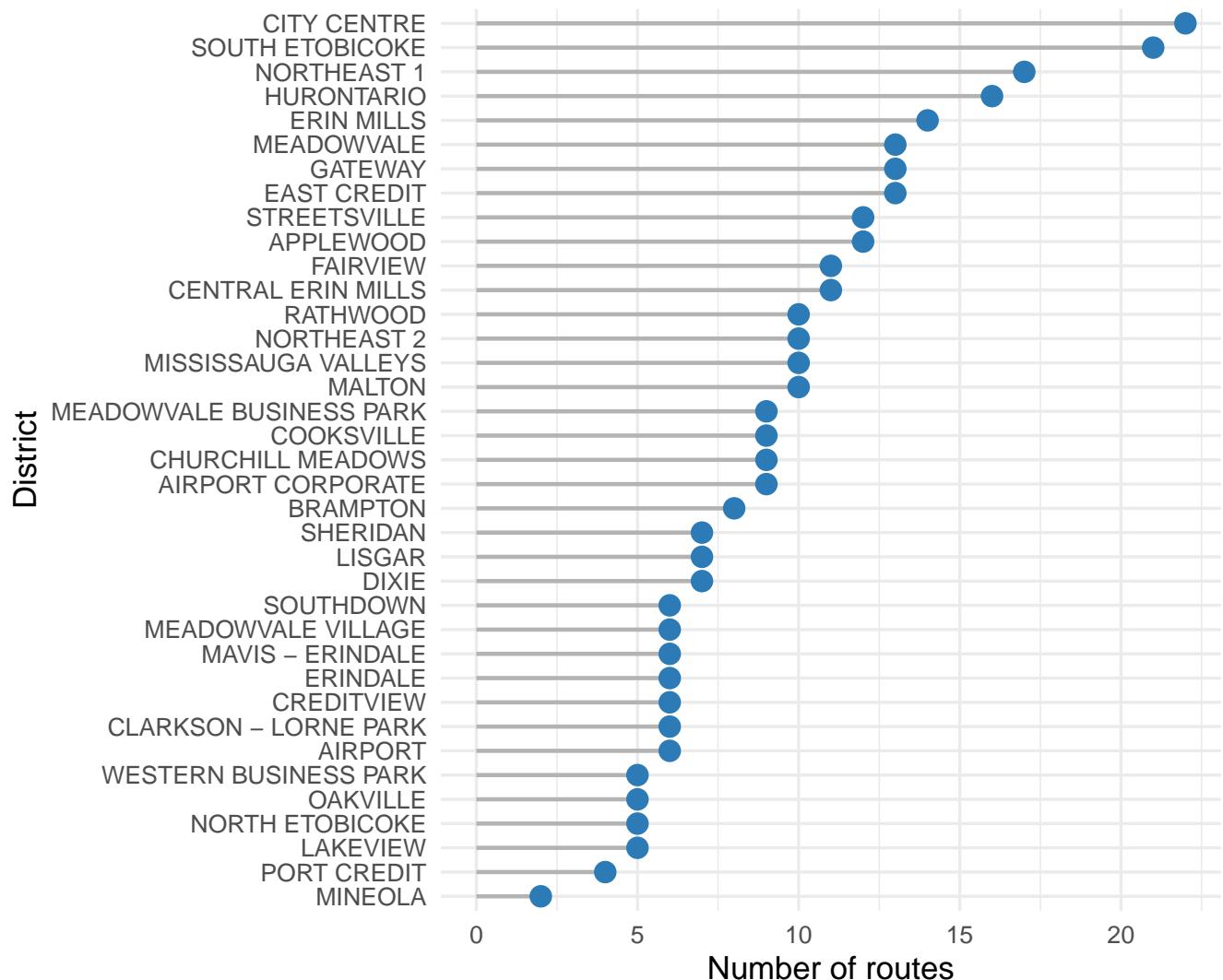
1. MiWay Coverage Across Districts

The first step is to see how many MiWay stops and routes each district contains. The lollipop plot of stop counts shows large differences across the city. Districts along major corridors and near the city centre have many more stops, while some outlying districts have far fewer. This means that the basic opportunity to catch a bus is not the same everywhere. The second lollipop plot shows the number of distinct MiWay routes serving each district. Districts with many stops also tend to have many routes, so riders there can choose between several paths and destinations. A few districts stand out as having many stops but a more limited set of routes, which suggests these stops mostly serve a small number of lines. The scatterplot of routes versus stops makes this relationship clearer. Each point represents a district, and the fitted regression line slopes upward. Districts with more stops generally have more routes, which is expected, but the spread around the line shows that some districts get more route variety than others even after accounting for how many stops they have. For UTM students, this matters because a district with many stops but little route variety may still offer fewer choices for reaching campus or a GO station.

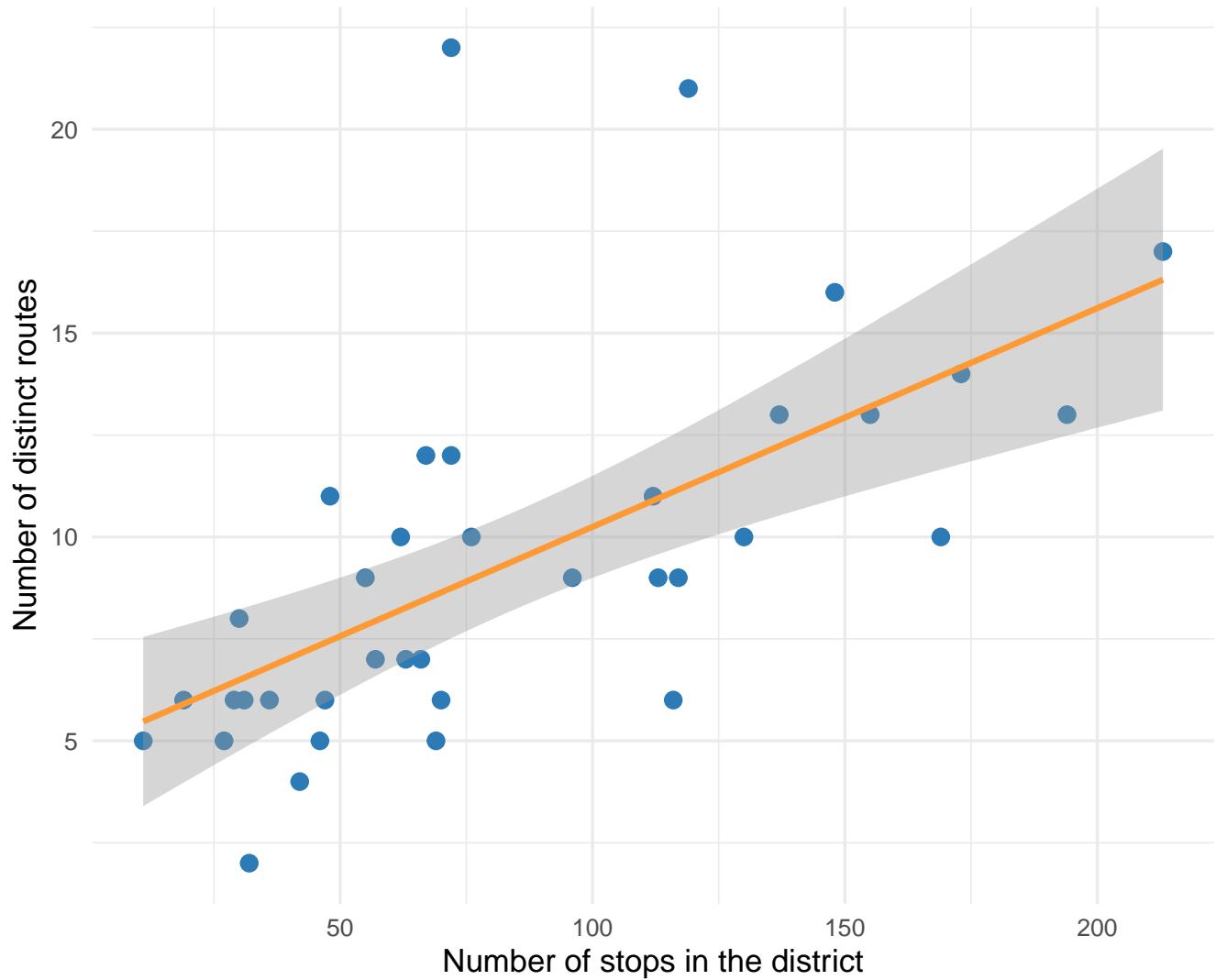
Number of MiWay Stops by District



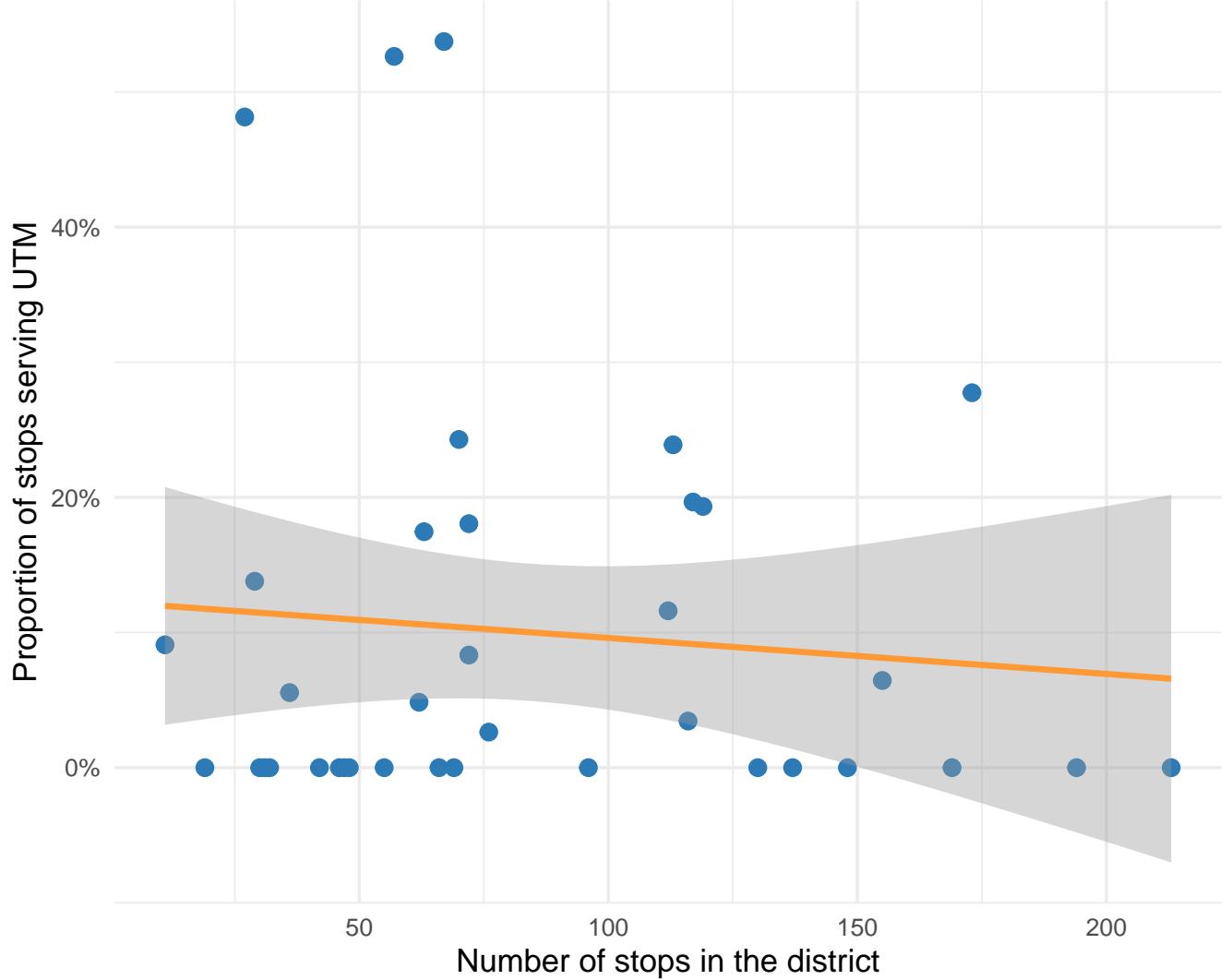
Number of MiWay Routes by District



Relationship Between Number of Stops and Routes by District



Direct UTM Access versus Number of Stops by District

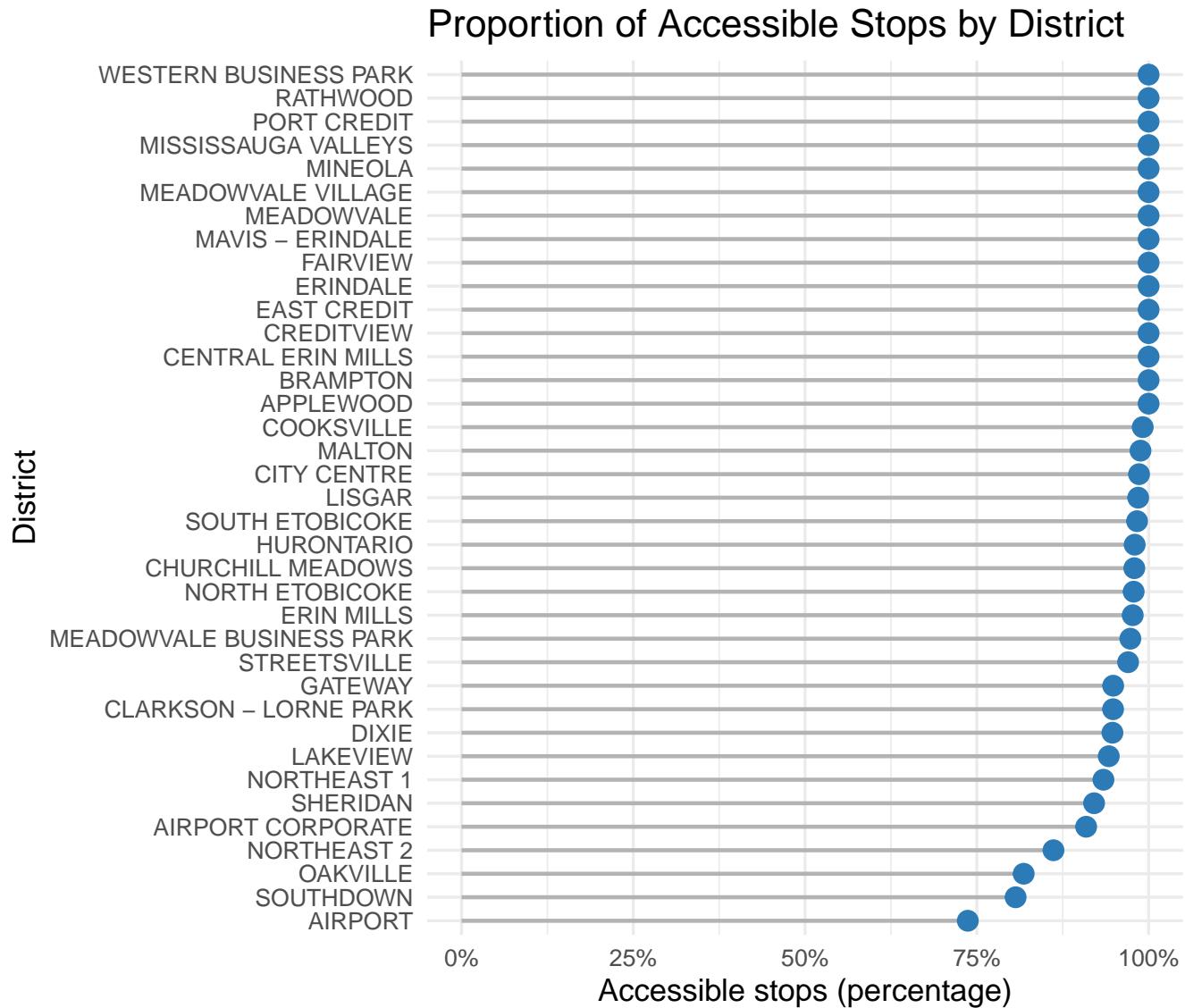


We also fitted a regression of the share of stops serving UTM on the number of stops in the district. The fitted line is fairly flat, which suggests that simply having more stops does not guarantee that a large share of them serve UTM. Some districts with modest stop counts devote a large proportion of those stops to UTM routes, while others with many stops offer relatively little direct UTM service.

2. Accessibility and Shelter Coverage

We next look at features that affect how comfortable and usable stops are for riders. The accessibility lollipop plot shows that almost all districts have very high proportions of accessible stops. This reflects city design standards and accessibility legislation. The chi square test confirms that the proportion is not exactly the same in every district, but the differences are small compared with the overall high level of accessibility. From the perspective of an individual student, this means that basic physical access to buses is fairly consistent across the city. Shelter coverage tells a different story. The second lollipop plot shows the proportion of stops with shelters in each district, and the range is much wider. Some districts provide shelters at more than half of their stops, while others provide them at only a small fraction. The chi square test indicates a strong association between district and shelter presence, so where a student lives strongly

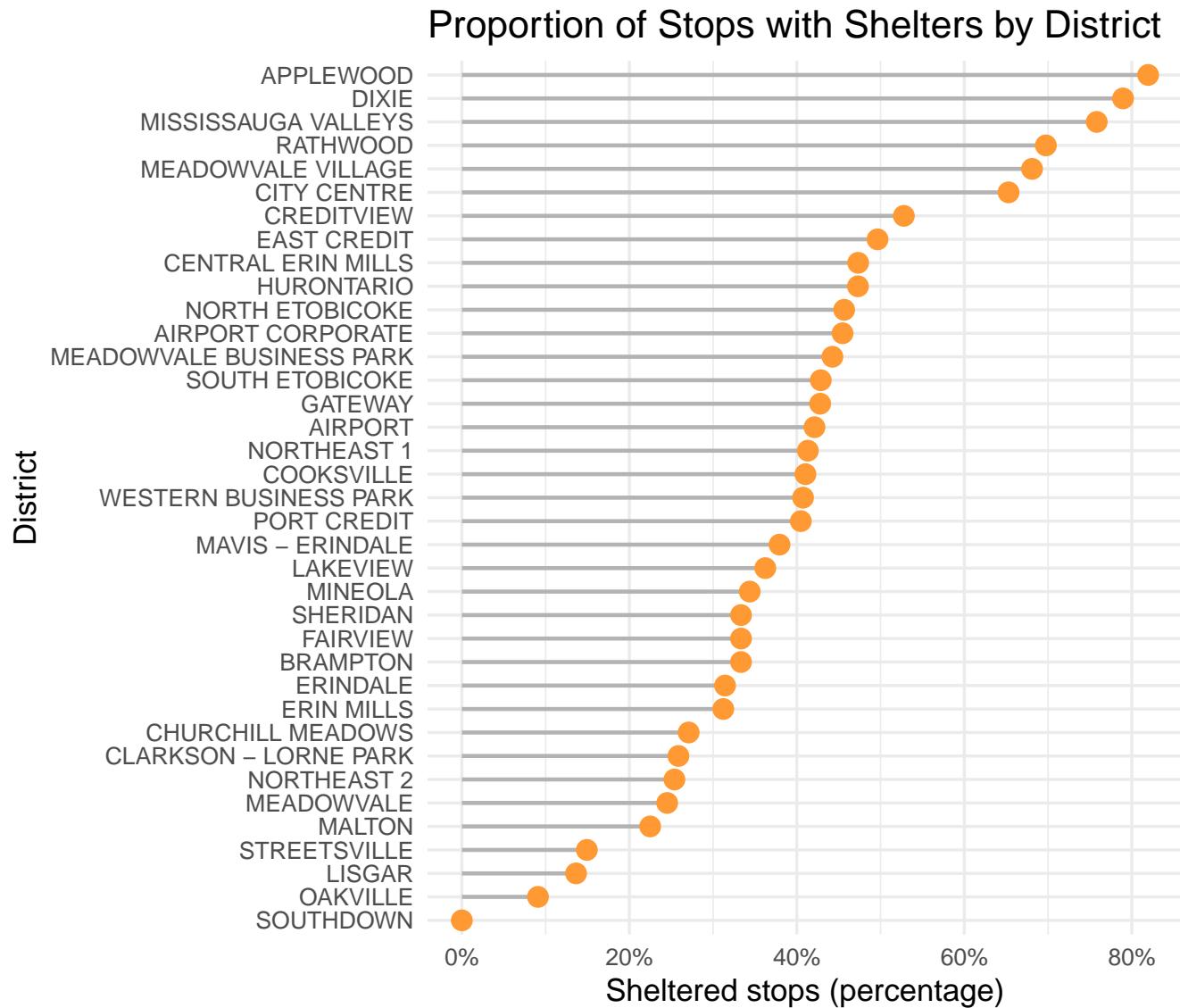
influences how often they can wait under cover during bad weather. To compare these two features directly, the joint plot of accessible and sheltered stops shows that accessibility is clustered near the top end for every district, while shelter coverage scatters more widely. Many districts achieve high accessibility with relatively low shelter coverage. This suggests that the main inequity lies not in basic physical access but in the comfort and protection offered to riders while they wait.



```
##
## 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.

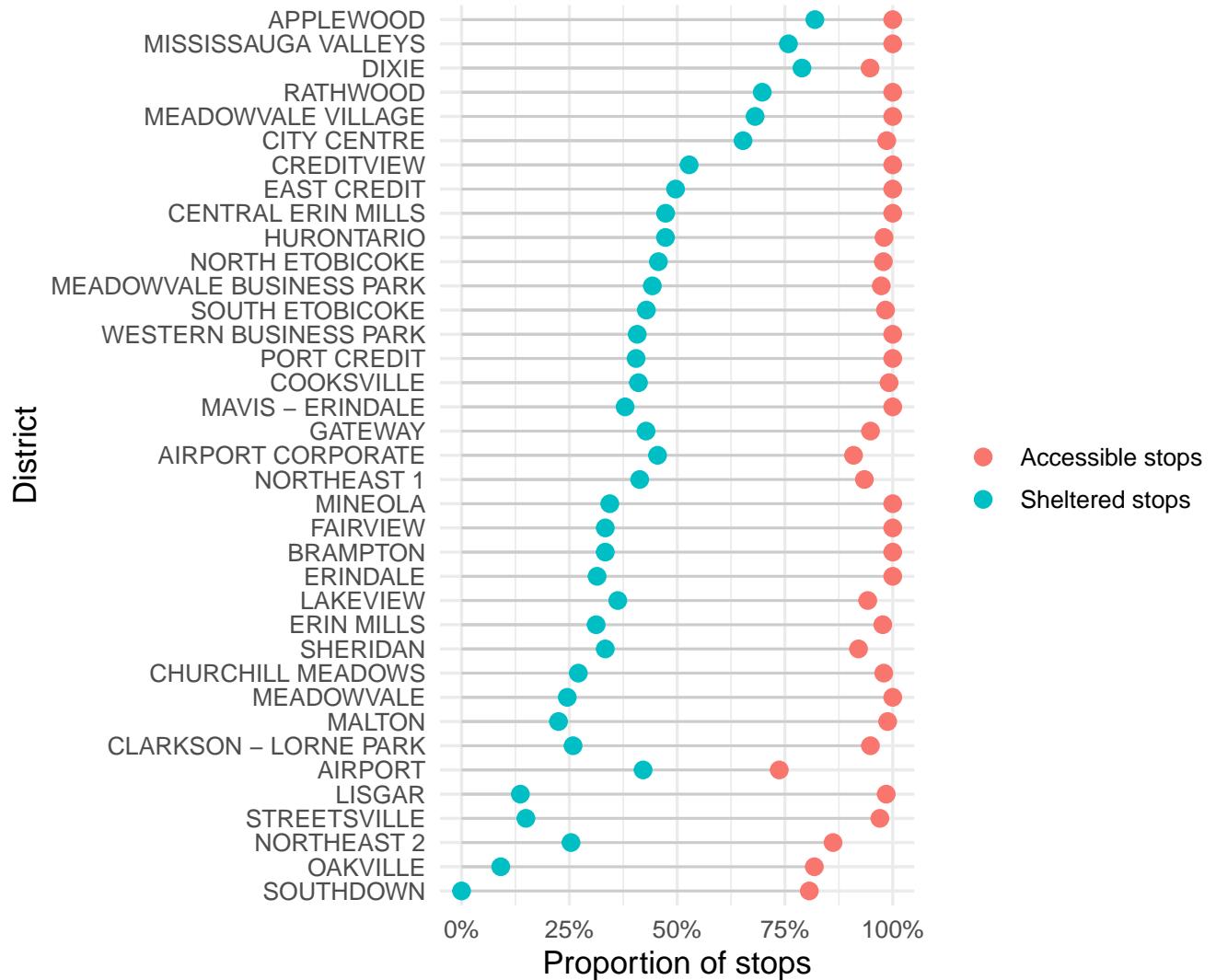
Shelter coverage varies more widely:



```
##  
## 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.

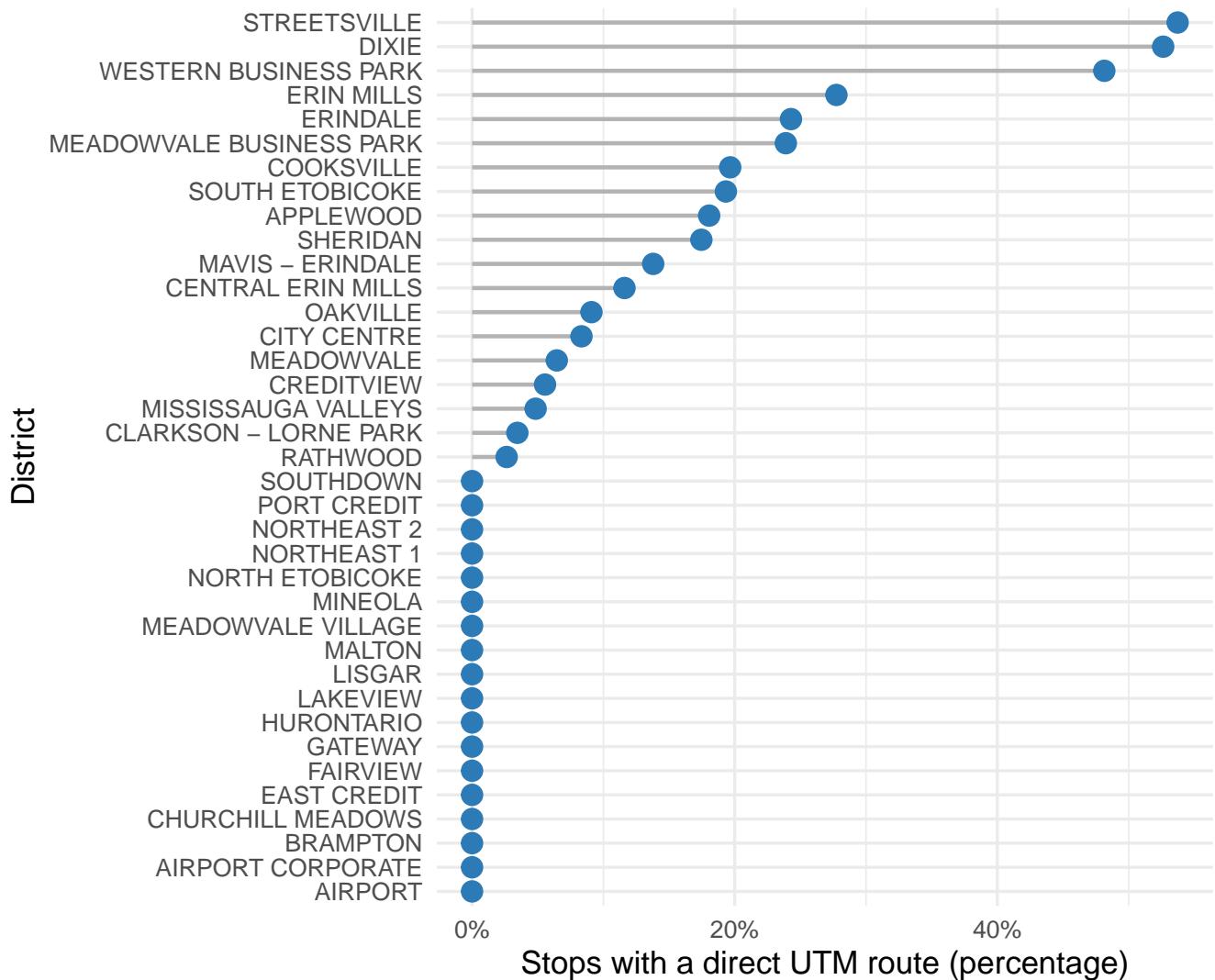
Accessibility and Shelter Coverage by District



3. Direct UTM Access

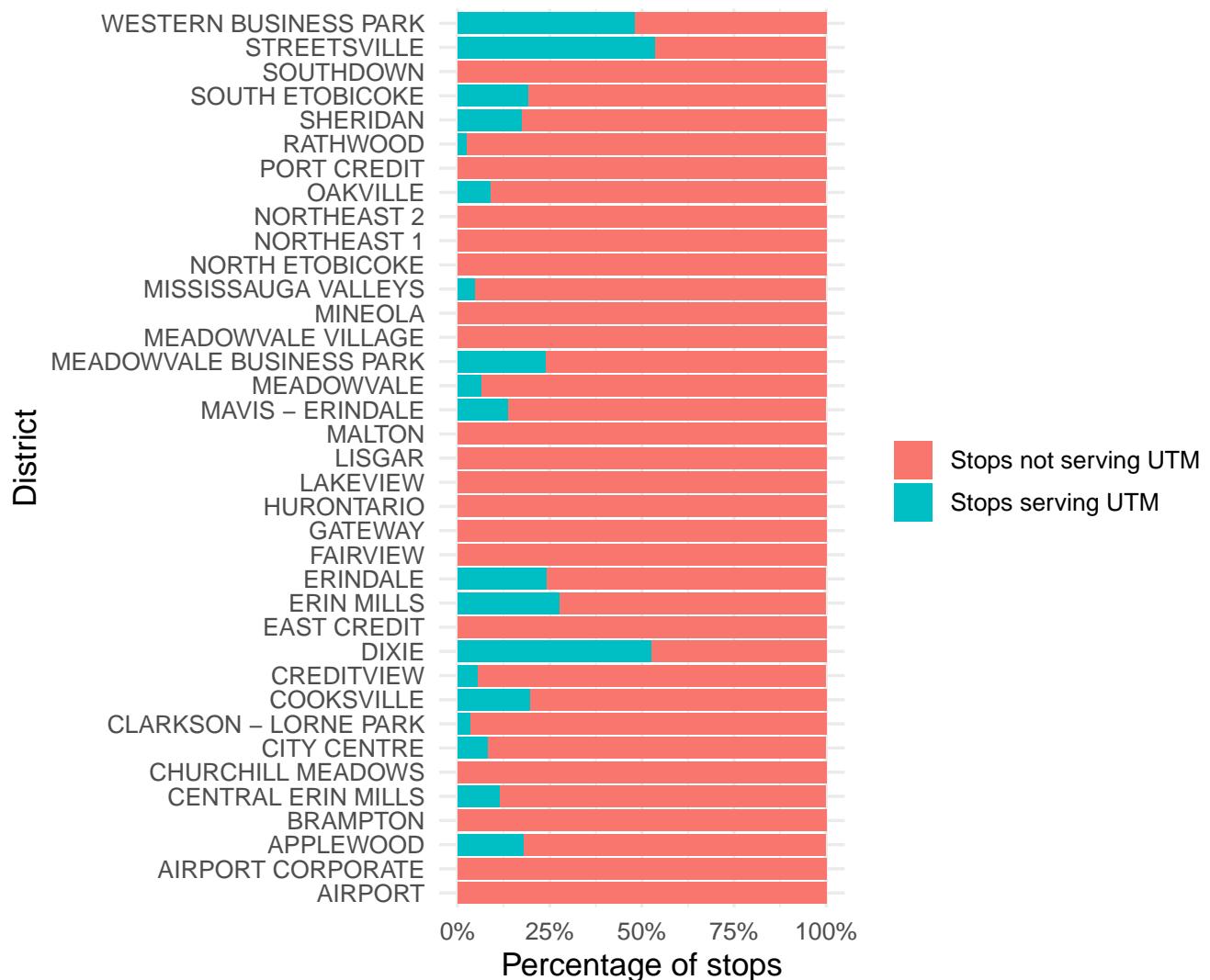
Direct MiWay routes to UTM are especially important for students who commute daily. The lollipop plot of UTM serving stop proportions shows that these routes are heavily concentrated in a subset of districts, mainly those along Dundas Street and Mississauga Road. In some districts, a large share of stops connects directly to UTM, while in many others very few stops do so. The stacked plot of UTM versus non UTM stops puts this into perspective. In high access districts, most stops serve at least one UTM route, so students can board near home and travel to campus without a transfer. In low access districts, almost all stops do not serve UTM, so students must transfer through a major corridor or terminal. The chi square test confirms that this pattern is not random. The probability that a given stop serves UTM depends strongly on the district. The scatterplot of UTM access versus shelter coverage ties these themes together. Districts in the upper right corner enjoy both a high share of UTM stops and good shelter coverage, offering both convenient and comfortable commutes. Districts in the lower left have fewer UTM serving stops and fewer shelters, which is a much less favourable situation for students. The regression line has a slight positive slope, which suggests that districts that invest more in direct UTM service also tend, on average, to provide better shelter coverage, though there is still considerable variation.

Proportion of Stops Serving UTM by District

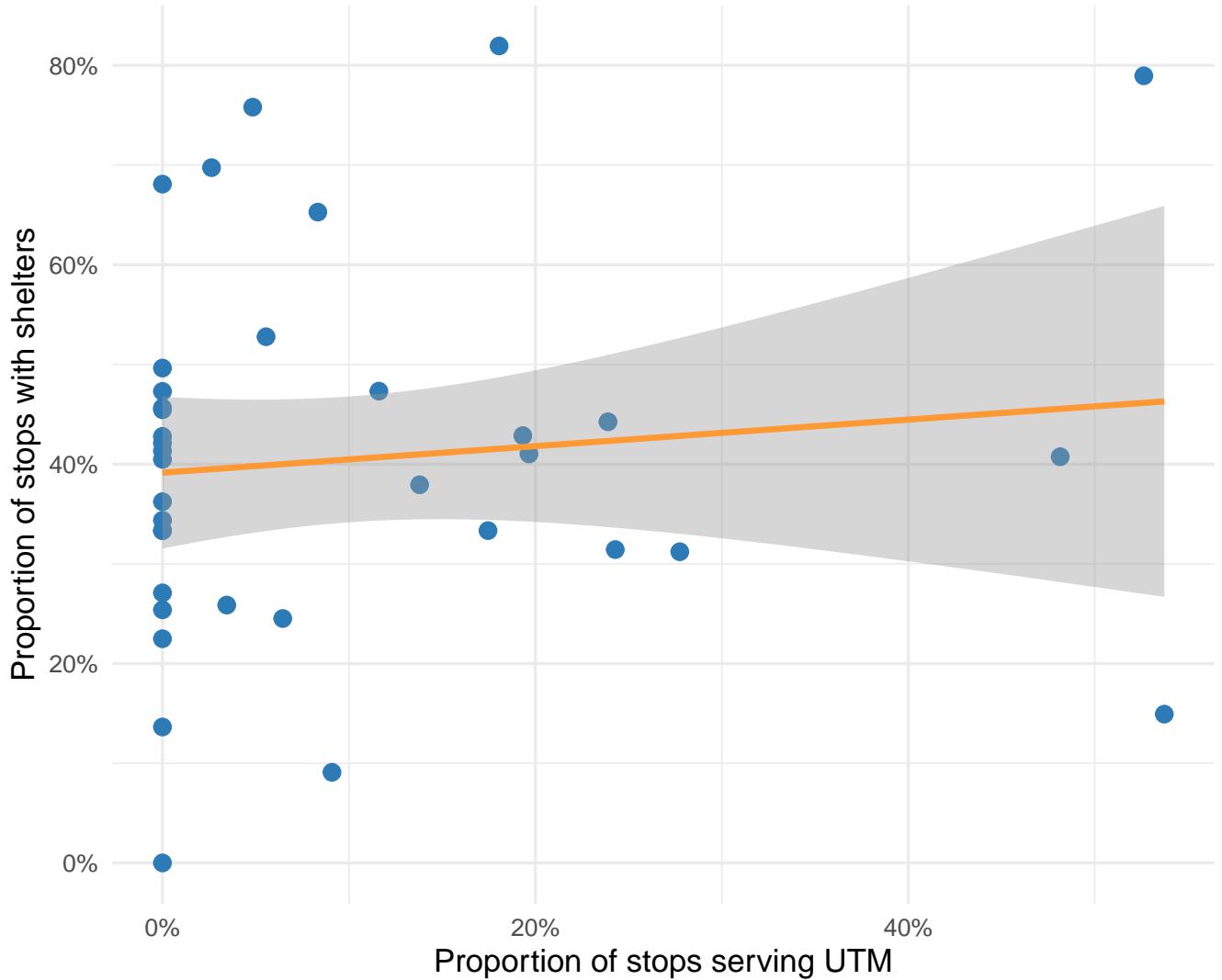


Normalized rates provide a fairer comparison:

Share of Stops Serving UTM by District



Shelter Coverage versus Direct UTM Access by District



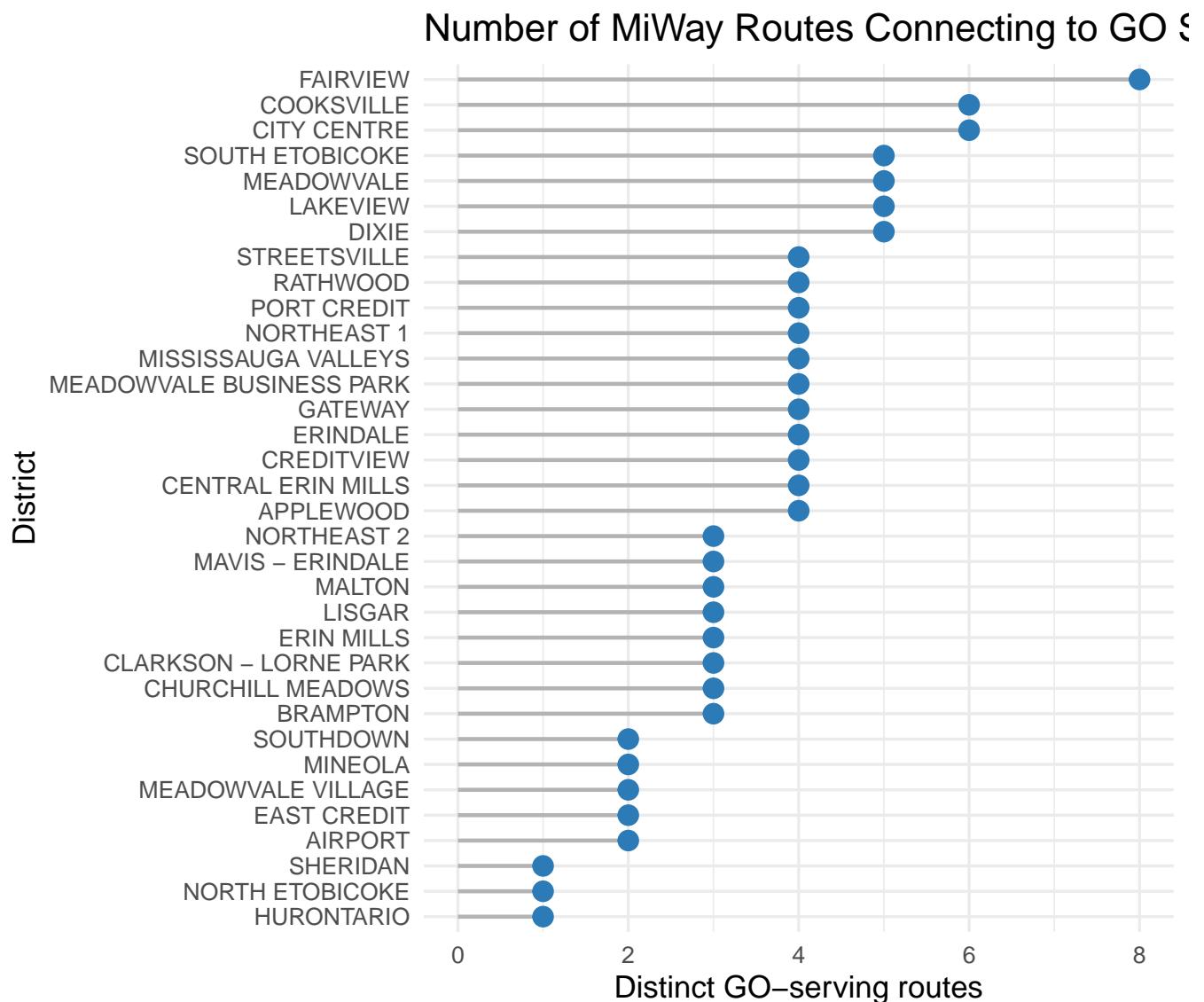
```
##
##  Pearson's Chi-squared test
##
## data: tab_utm
## X-squared = 672.88, df = 36, p-value < 2.2e-16
```

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.

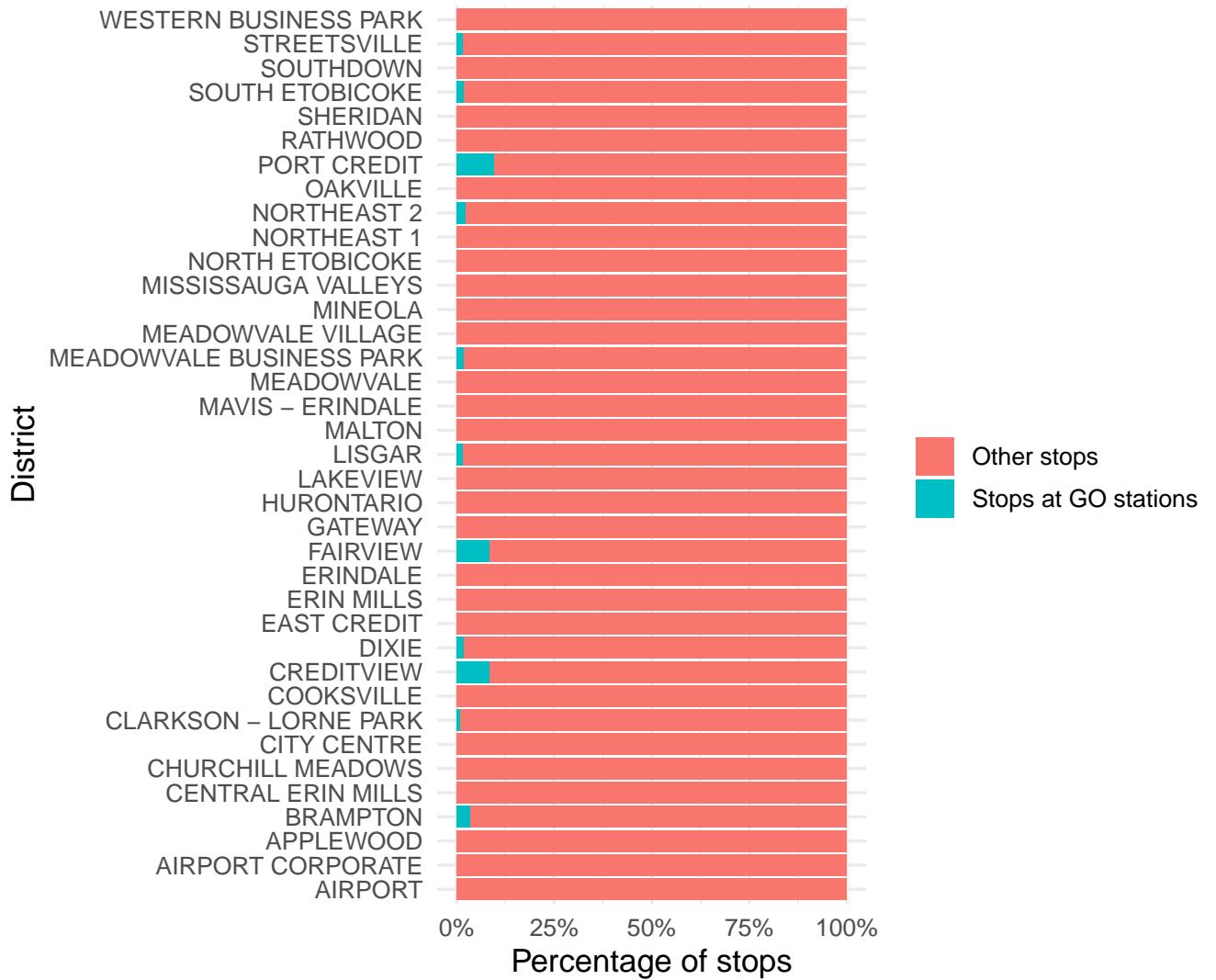
4. GO Train Connectivity

GO Train stations connect MiWay riders to the wider Greater Toronto Area. The lollipop plot of GO serving routes by district shows that only a small number of districts have many routes that serve a GO station. These districts include the areas that contain Port Credit, Cooksville, and Meadowvale GO stations. Most other districts have very few or no routes that directly access a GO platform. The share plot of stops at GO

stations versus other stops makes this concentration even more obvious. In most districts, almost all stops are regular MiWay stops that do not serve a GO station. In the few districts that contain GO terminals, a meaningful minority of stops are GO platforms or immediately adjacent stops. The chi square test confirms that being a GO stop is highly associated with district. For UTM students, this means that living in a district with a GO station can significantly reduce the time and uncertainty involved in regional travel, since transferring between MiWay and GO happens in one place. Students in districts without GO terminals often need an additional transfer or a longer MiWay ride to reach regional rail.



Share of Stops at GO Stations by District



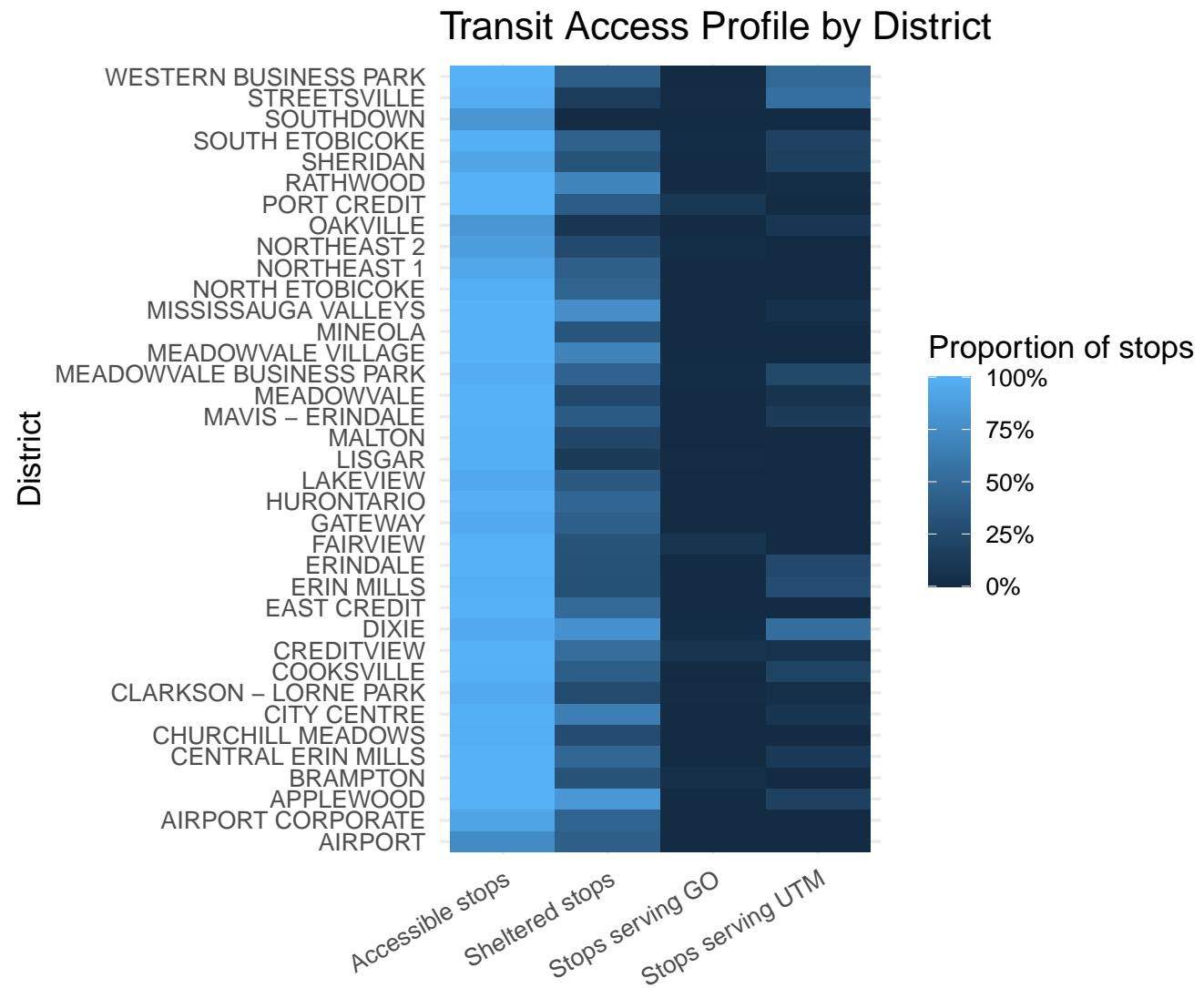
```
##
##  Pearson's Chi-squared test
##
## data: tab_go
## X-squared = 139.69, df = 36, p-value = 3.843e-14
```

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.

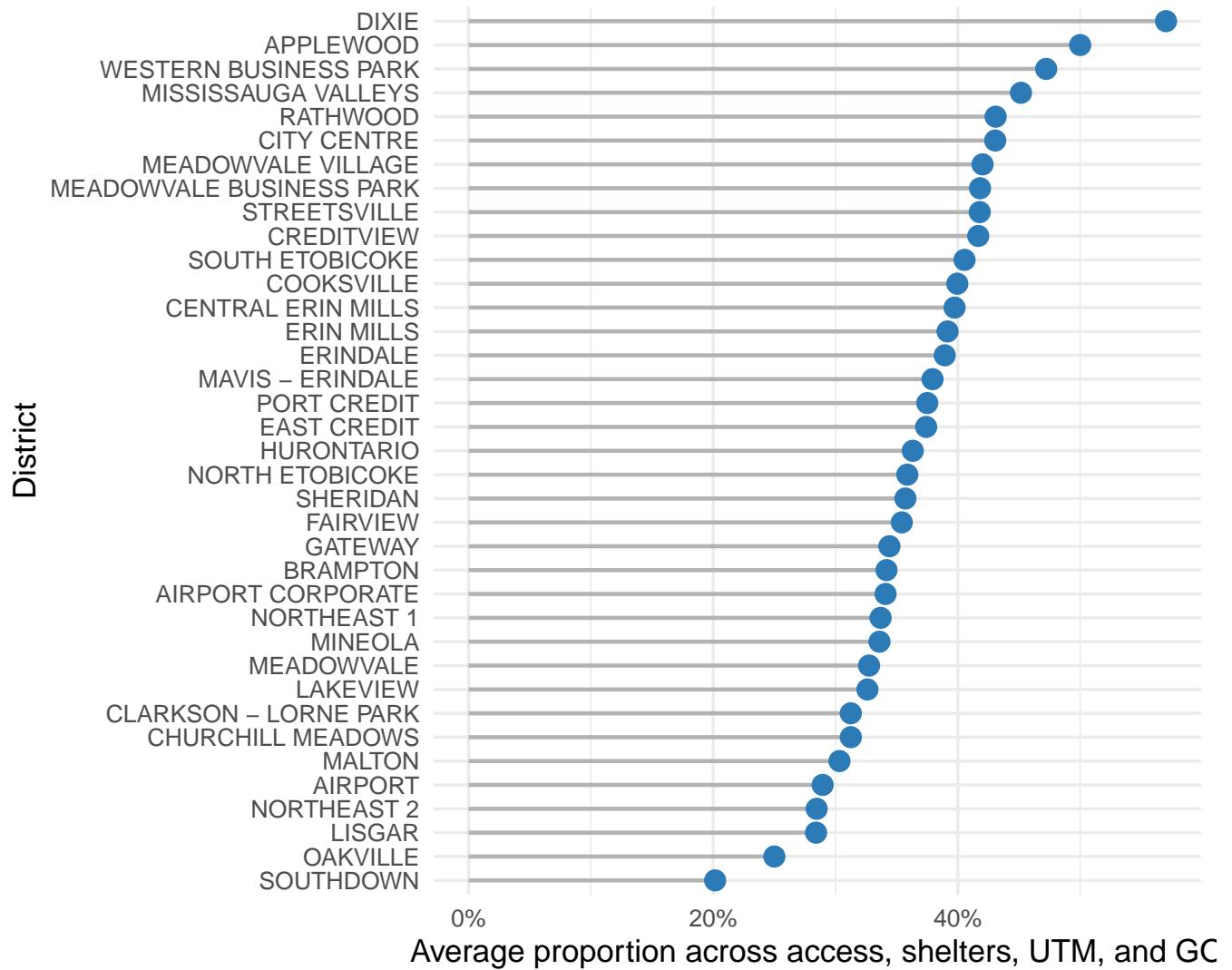
5. Overall Transit Profile by District

Finally, we combine several features into an overall transit profile for each district. The heatmap shows, for every district, the proportion of stops that are accessible, sheltered, serving UTM, and serving GO. Darker shading indicates higher coverage. This figure makes it easy to spot districts that perform consistently well across all four indicators and those that are strong on some dimensions but weak on others. For example,

some districts have high accessibility and shelter coverage but low direct UTM or GO access, while others are the opposite. To summarise these four measures with a single index, we construct a composite score that averages the proportions of accessible, sheltered, UTM serving, and GO serving stops in each district. The composite lollipop plot ranks districts from lowest to highest on this index. Districts at the top have a combination of good basic access, weather protection, direct UTM routes, and GO connectivity, while those at the bottom fall behind on one or more of these components. Taken together, these summaries show that transit opportunities for UTM students depend strongly on where they live. Some districts offer many stops, diverse routes, high shelter coverage, and strong UTM and GO connections. Others require students to rely on fewer stops, more transfers, or unsheltered waits, even though the overall city network appears dense on a map.



Composite Transit Access Score by District



The composite score highlights a small set of districts that perform well on several indicators at once, while others lag behind on one or more dimensions of access.

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

```
# packages

library(dplyr)
library(ggplot2)
library(tidyr)
library(scales)

# 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))

utm_share <- stop_level %>%
  group_by(District) %>%
  summarise(
    utm      = sum(has_utm),
    non_utm  = sum(!has_utm),
    .groups = "drop"
  ) %>%
  pivot_longer(
    cols = c(utm, non_utm),
    names_to = "utm_type",
    values_to = "n"
  ) %>%
  group_by(District) %>%
  mutate(prop = n / sum(n)) %>%
  ungroup() %>%
  mutate(
    utm_type = recode(
      utm_type,
      utm = "Stops serving UTM",
      non_utm = "Stops not serving UTM"
    )
  )

go_share <- df %>%
  mutate(
    GO_type = if_else(is_GO_stop, "Stops at GO stations", "Other stops")
  ) %>%
  group_by(District, GO_type) %>%
  summarise(n = n(), .groups = "drop") %>%
  group_by(District) %>%
  mutate(prop = n / sum(n)) %>%
  ungroup()

district_profile <- df %>%
  group_by(District) %>%
  summarise(
    prop_accessible = mean(Accessible, na.rm = TRUE),
    prop_shelter    = mean(Shelter,    na.rm = TRUE),
    prop_utm        = mean(Stop_Number %in% stop_level$Stop_Number[stop_level$has_utm], na.rm = TRUE),
    prop_go         = mean(is_GO_stop, na.rm = TRUE),
    .groups = "drop"
  )

```

```

district_profile_long <- district_profile %>%
pivot_longer(
cols = starts_with("prop_"),
names_to = "feature",
values_to = "value"
) %>%
mutate(
feature = recode(
feature,
prop_accessible = "Accessible stops",
prop_shelter = "Sheltered stops",
prop_utm = "Stops serving UTM",
prop_go = "Stops serving GO"
)
)

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)

chisq_access

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

chisq_shelter

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

chisq_utm

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

```

```
chisq_go
```

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