Disability Rate Comparisons Across Demographics Between Buffalo and Rochester NY - 606 Final

Lu Beyer

05/08/2024

Abstract

This research is intended to view and compare disability rates across demographics in Buffalo and Rochester New York Metro Areas. The Metro Areas of both Buffalo and Rochester are similarly sized in population, and are based near each other in the same Western New York Region. Given this information, one might not expect a significant difference between the two cities. However, if a statistically significant difference existed, we could begin looking into the factors that may contribute to any differences in disability rates between the two cities.

Data was sourced from the US Census, specifically the 2022 American Community Survey. Comparisons were performed using chi-square tests, comparing the population who reported identifying as disabled between the two cities. After running chi-square tests across multiple demographics, the data showed that Rochester NY had a statistically higher disability rate compared to Buffalo across the general population, as well as most selected general demographics.

Research Questions

Null Hypothesis: There is no statistically significant difference in disability rate between Buffalo and Rochester NY Alternate Hypothesis: There is a significant statistical difference in disability rate between Buffalo and Rochester NY

Data

Data was pulled from US Census, through the 2022 American Community Survey - https://data.census.gov/table/ACSST1Y2022.S1810?q=disability&g=310XX00US15380,40380&moe=false&tp=false

We are looking at Buffalo and Rochester Metro Areas, rather than strictly city limits, as they are closer in terms of total population for comparison.

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
```

```
## v purrr     1.0.2
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(infer)

data <- read.csv("Buf_Roch_2022.csv") %>%
  rename(Category = 1,
         Buf_Population = 2,
         Buf_Population_Disability = 3,
         Buf_Disability_Percentage = 4,
         Roch_Population = 5,
         Roch_Population_Disability = 6,
         Roch_Disability_Percentage = 7) %>%
  mutate(across(where(is.character), ~str_replace_all(., "Â", ""))) %>%
  mutate(Category = str_trim(Category)) %>%
  mutate(across(where(is.character), ~str_replace_all(., "^N$", as.character(NA)))) %>%
  mutate(across(c(Buf_Population, Buf_Population_Disability), ~as.numeric(str_replace_all(., ",", "")))
  mutate(across(c(Roch_Population, Roch_Population_Disability), ~as.numeric(str_replace_all(., ",", "")
  mutate(Buf_Disability_Percentage = as.numeric(str_remove(Buf_Disability_Percentage, "%")) / 100) %>%
  mutate(Roch_Disability_Percentage = as.numeric(str_remove(Roch_Disability_Percentage, "%")) / 100) %>
  mutate(Buf_Population_Nondisabled = Buf_Population - Buf_Population_Disability,
         Roch_Population_Nondisabled = Roch_Population - Roch_Population_Disability) %>%
  select(Category, Buf_Population, Buf_Population_Disability, Buf_Population_Nondisabled, Buf_Disability
         Roch_Population, Roch_Population_Disability, Roch_Population_Nondisabled, Roch_Disability_Perc
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `across(...)`.
## Caused by warning:
## ! NAs introduced by coercion
## There was 1 warning in `mutate()`.
## i In argument: `across(...)`.
## Caused by warning:
## ! NAs introduced by coercion
```

```
sex <- data %>%
  slice(3:4)

race <- data %>%
  slice(6:14)

age <- data %>%
  slice(16:21)

general <- data %>%
  slice(1, 3:4, 6:14, 16:21) %>%
  drop_na()

head(general, 5)
```

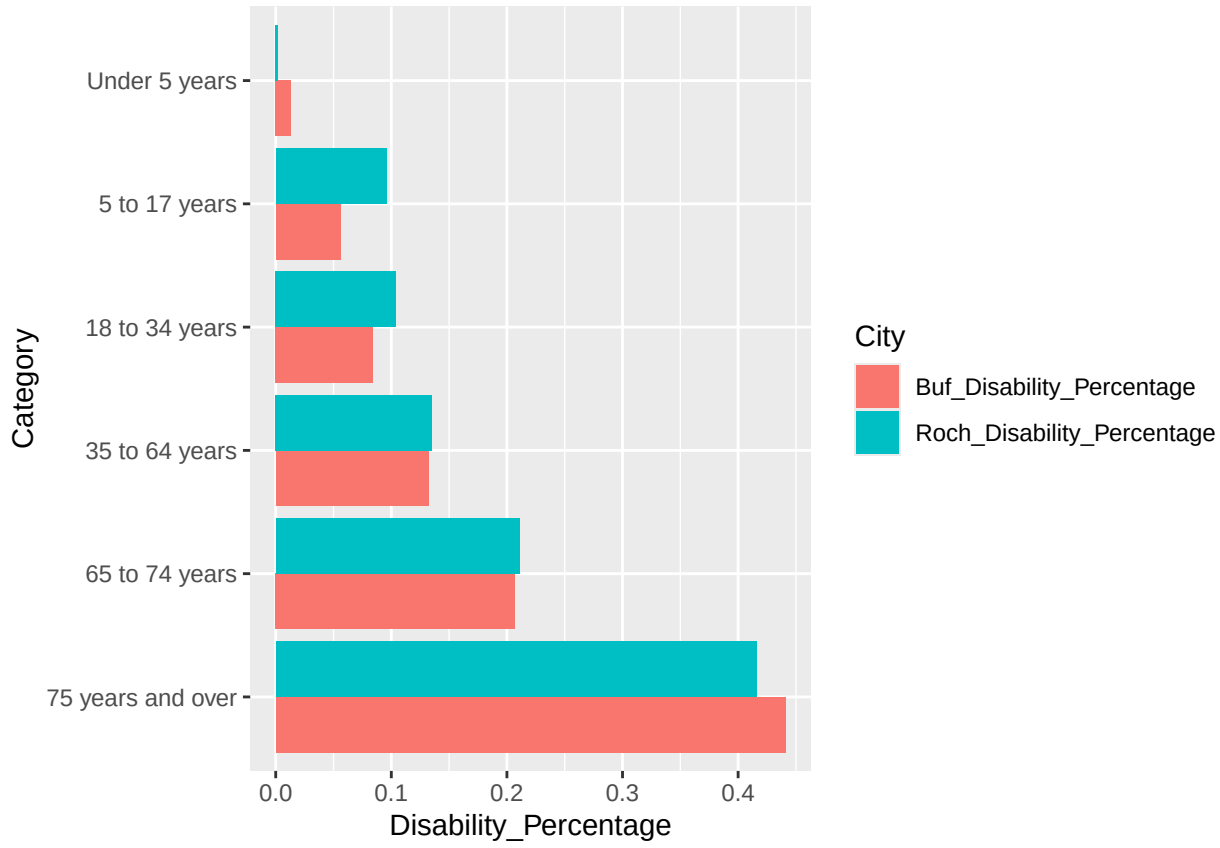
```
##                                     Category Buf_Population
```

```
## 1 Total civilian noninstitutionalized population        1151880
## 2                                         Male         563533
## 3                                       Female         588347
## 4                                White alone         867525
## 5            Black or African American alone         135420
##   Buf_Population_Disability Buf_Population_Nondisabled
## 1                    157331                     994549
## 2                     74424                     489109
## 3                     82907                     505440
## 4                    123577                     743948
## 5                     19264                     116156
##   Buf_Disability_Percentage Roch_Population Roch_Population_Disability
## 1                     0.137         1069601                     156570
## 2                     0.132          522980                      75662
## 3                     0.141          546621                      80908
## 4                     0.142          804283                     111706
## 5                     0.142          112916                      19648
##   Roch_Population_Nondisabled Roch_Disability_Percentage
## 1                      913031                      0.146
## 2                      447318                      0.145
## 3                      465713                      0.148
## 4                      692577                      0.139
## 5                       93268                      0.174
```

Summary Statistics

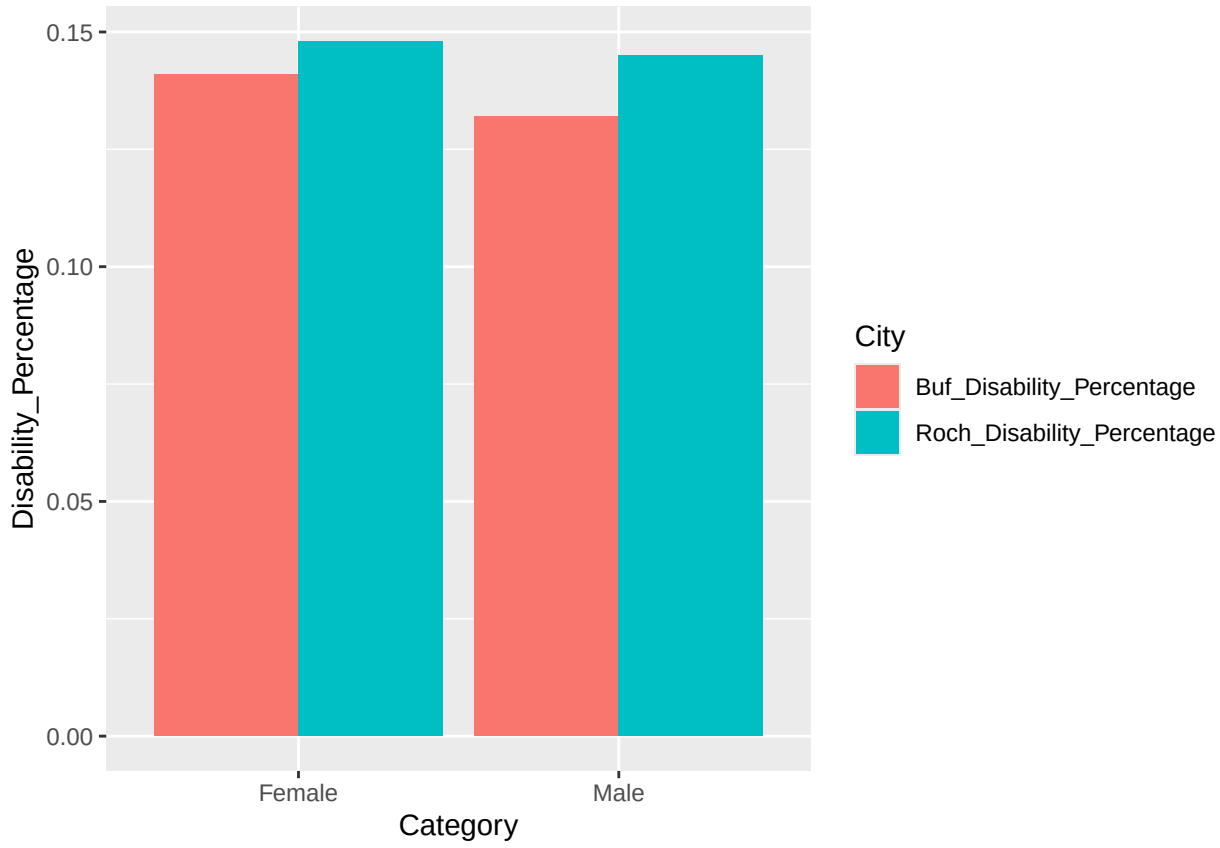
Disabilty Rates by Age

```
cat <- c("Under 5 years", "5 to 17 years", "18 to 34 years", "35 to 64 years", "65 to 74 years", "75 ye

age_long <- age %>%
  pivot_longer(
    cols = c(Buf_Disability_Percentage, Roch_Disability_Percentage),
    names_to = "City",
    values_to = "Disability_Percentage"
  ) %>%
  mutate(Category = factor(Category, levels = rev(cat)))

ggplot(data = age_long, aes(x = Category, y = Disability_Percentage, fill = City)) +
  geom_bar(stat = "identity", position = "dodge") +
  coord_flip()
```



Disability Rates by Sex

```
sex_long <- sex %>%
  pivot_longer(
    cols = c(Buf_Disability_Percentage, Roch_Disability_Percentage),
    names_to = "City",
    values_to = "Disability_Percentage"
  )

ggplot(data = sex_long, aes(x = Category, y = Disability_Percentage, fill = City)) +
  geom_bar(stat = "identity", position = "dodge")
```



Disabiliy Rates by Race

```
race_long <- race %>%
  pivot_longer(
    cols = c(Buf_Disability_Percentage, Roch_Disability_Percentage),
    names_to = "City",
    values_to = "Disability_Percentage"
  )

ggplot(data = race_long, aes(x = Category, y = Disability_Percentage, fill = City)) +
  geom_bar(stat = "identity", position = "dodge") +
  coord_flip()
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range
## (`geom_bar()`).
```

Looking at the summary statistic charts, I've identified the following demographics as the most likely to have a statistically significant difference in disability rates, due to their notably large differences in rates

Age: 0 - 5 5 - 17

Race: Two or more races Some other race Hispanic American Indian and Alaska Native

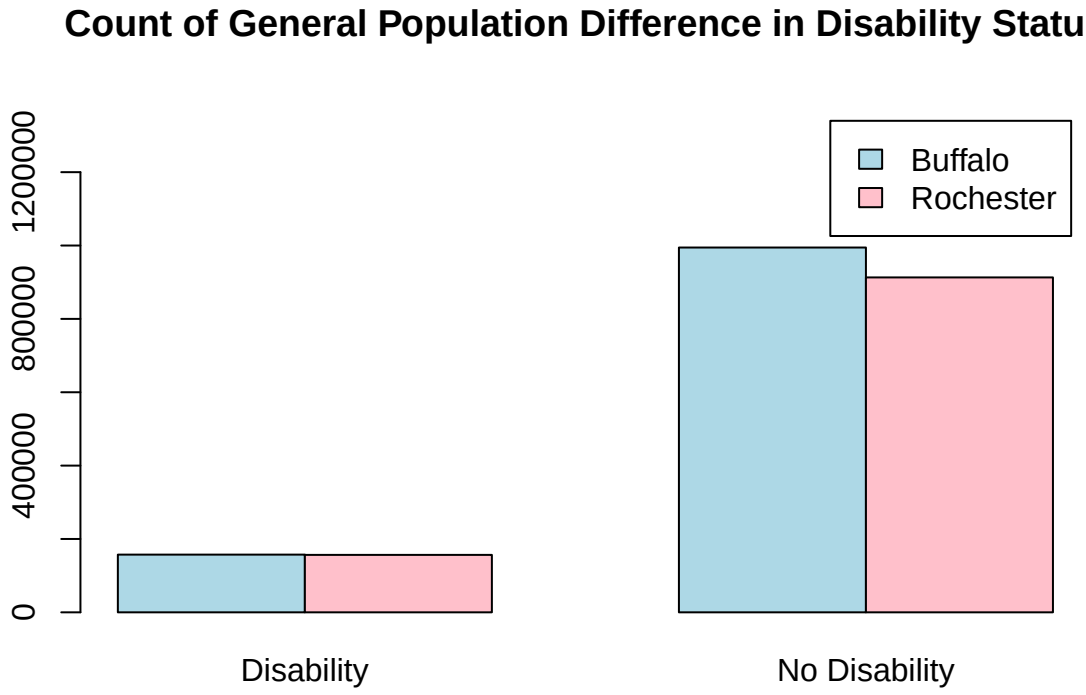
Chi-Square Tests

General Population

```
general_table <- matrix(c(general$Buf_Population_Disability[1], general$Buf_Population_Nondisabled[1],
                          general$Roch_Population_Disability[1], general$Roch_Population_Nondisabled[1]),
                        nrow = 2, byrow = TRUE,
                        dimnames = list(c("Buffalo", "Rochester"), c("Disability", "No Disability")))

chisq_result_general <- chisq.test(general_table)
print(chisq_result_general)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  general_table
## X-squared = 438.49, df = 1, p-value < 2.2e-16
```

```r
barplot(general_table,
        beside = TRUE,
        col = c("lightblue", "pink"),
        ylim = c(0, max(general_table) * 1.4),
        legend = rownames(general_table),
        main = "Count of General Population Difference in Disability Status")
```




```r
print(chisq_result_general$residuals)
```

```
##           Disability No Disability
## Buffalo    -13.46575      5.462424
## Rochester   13.97408     -5.668631
```

We can see that for the general population, the p-value is below 2.2e-16 which is under the p-value limit of 0.05, and can say that there is significant statistical difference in the disability rates between Buffalo and Rochester. Looking futher at the chi-squared residuals, we can see that the data skews to show Buffalo having a lower disability rate compared to Rochester.

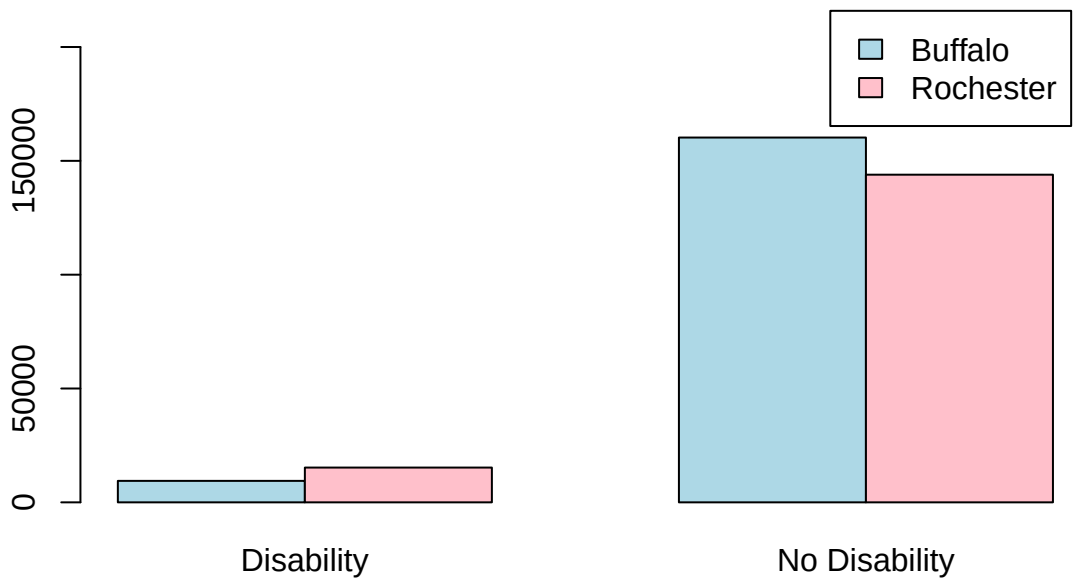
Population - Between 5 to 17 Years

```r
#5 to 17 Years
age_table <- matrix(c(age$Buf_Population_Disability[2], age$Buf_Population_Nondisabled[2],
```

```
                        age$Roch_Population_Disability[2], age$Roch_Population_Nondisabled[2]),
                      nrow = 2, byrow = TRUE,
                      dimnames = list(c("Buffalo", "Rochester"), c("Disability", "No Disability")))

chisq_result_age <- chisq.test(age_table)
print(chisq_result_age)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  age_table
## X-squared = 1926.4, df = 1, p-value < 2.2e-16
```

```
barplot(age_table,
        beside = TRUE,
        col = c("lightblue", "pink"),
        ylim = c(0, max(age_table) * 1.4),
        legend = rownames(age_table),
        main = "Difference in Disability Status, Age 5 to 17 Years")
```




```
chisq_result_age$stdres
```

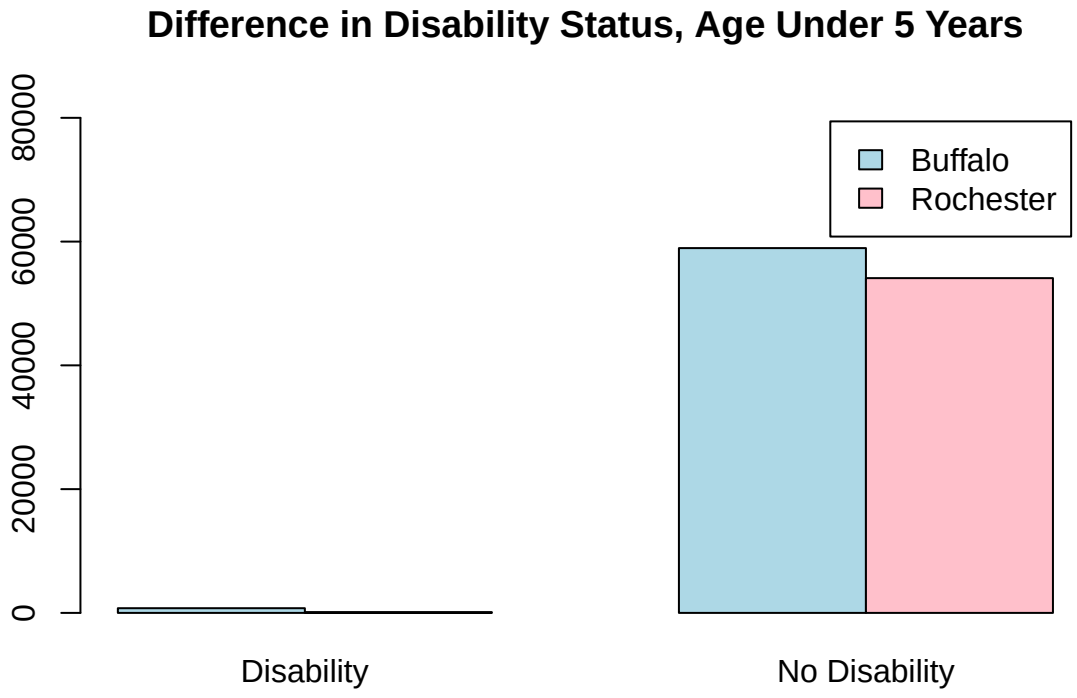
```
##           Disability No Disability
## Buffalo      -43.897        43.897
## Rochester     43.897       -43.897
```

Population - Under 5

```
#Under 5
age_table <- matrix(c(age$Buf_Population_Disability[1], age$Buf_Population_Nondisabled[1],
                      age$Roch_Population_Disability[1], age$Roch_Population_Nondisabled[1]),
                    nrow = 2, byrow = TRUE,
                    dimnames = list(c("Buffalo", "Rochester"), c("Disability", "No Disability")))

chisq_result_age <- chisq.test(age_table)
print(chisq_result_age)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  age_table
## X-squared = 413.41, df = 1, p-value < 2.2e-16
```

```
barplot(age_table,
        beside = TRUE,
        col = c("lightblue", "pink"),
        ylim = c(0, max(age_table) * 1.4),
        legend = rownames(age_table),
        main = "Difference in Disability Status, Age Under 5 Years")
```



```
chisq_result_age$stdres
```

```
##           Disability No Disability
## Buffalo      20.3664      -20.3664
## Rochester   -20.3664       20.3664
```

Here the disability rates for 0 - 5 is significant, with Rochester having lower rates compared to Buffalo for this demographic for the first time in our tests. However, looking at the data, it seems like the reported population for ages 0 - 5 in Rochester is suspiciously low. When organizing the primary demographics by rate, we can see the Rochester 0 - 5 age demographic has a rate of 0.002, with the next lowest having a rate of 0.09, 45 times smaller than the next lowest.

```
general %>%
  arrange(Roch_Disability_Percentage) %>%
  select(Category, Roch_Population_Disability, Roch_Population_Nondisabled, Roch_Disability_Percentage)
  head(3)
```

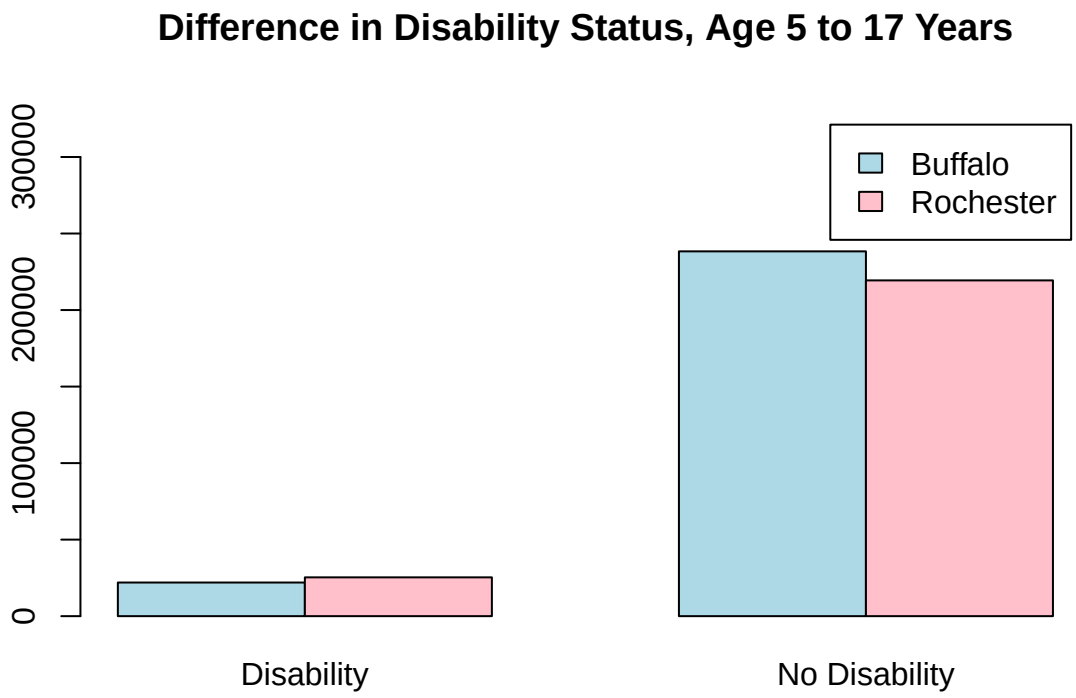
```
##        Category Roch_Population_Disability Roch_Population_Nondisabled
## 1 Under 5 years                       117                        54099
## 2   Asian alone                      2725                        27507
## 3 5 to 17 years                     15278                       143919
##   Roch_Disability_Percentage
## 1                      0.002
## 2                      0.090
## 3                      0.096
```

Population - Between 18 and 34 Years

```
#18 to 34 Years
age_table <- matrix(c(age$Buf_Population_Disability[3], age$Buf_Population_Nondisabled[3],
                      age$Roch_Population_Disability[3], age$Roch_Population_Nondisabled[3]),
                    nrow = 2, byrow = TRUE,
                    dimnames = list(c("Buffalo", "Rochester"), c("Disability", "No Disability")))

chisq_result_age <- chisq.test(age_table)
print(chisq_result_age)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  age_table
## X-squared = 547.06, df = 1, p-value < 2.2e-16
```

```
barplot(age_table,
        beside = TRUE,
        col = c("lightblue", "pink"),
        ylim = c(0, max(age_table) * 1.4),
        legend = rownames(age_table),
        main = "Difference in Disability Status, Age 5 to 17 Years")
```




```
chisq_result_age$stdres
```

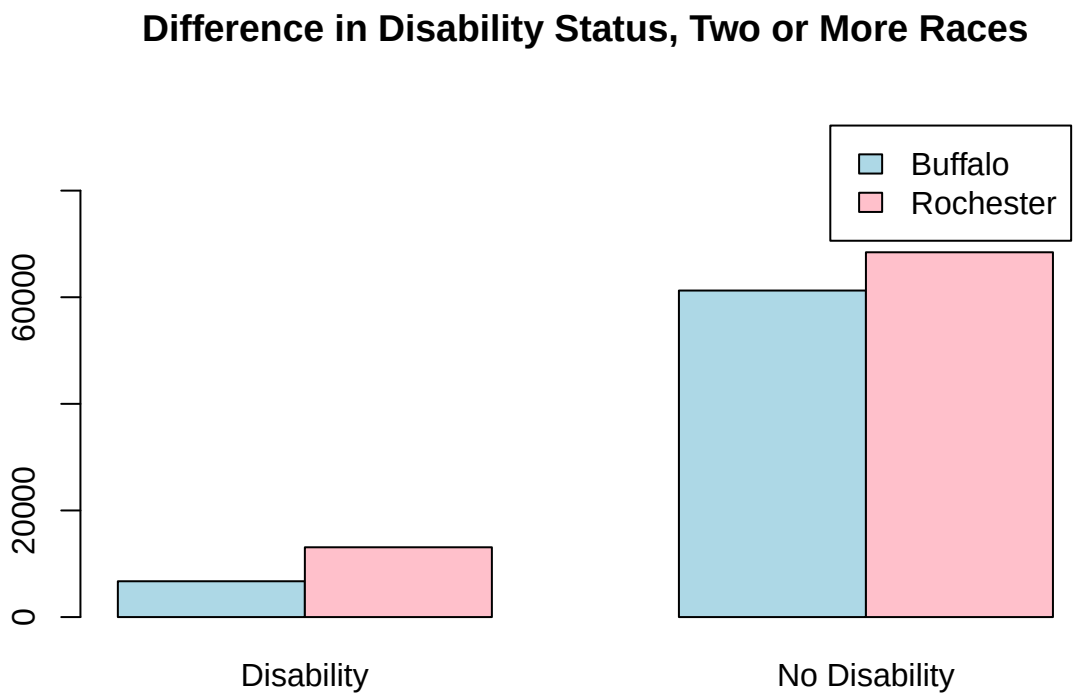
```
##           Disability No Disability
## Buffalo    -23.39419      23.39419
## Rochester   23.39419     -23.39419
```

Population - Two or More Races

```
#Two or More Races
race_table <- matrix(c(race$Buf_Population_Disability[7], race$Buf_Population_Nondisabled[7],
                       race$Roch_Population_Disability[7], race$Roch_Population_Nondisabled[7]),
                     nrow = 2, byrow = TRUE,
                     dimnames = list(c("Buffalo", "Rochester"), c("Disability", "No Disability")))

chisq_result_race <- chisq.test(race_table)
print(chisq_result_race)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  race_table
## X-squared = 1229.2, df = 1, p-value < 2.2e-16
```

```r
barplot(race_table,
        beside = TRUE,
        col = c("lightblue", "pink"),
        ylim = c(0, max(race_table) * 1.4),
        legend = rownames(race_table),
        main = "Difference in Disability Status, Two or More Races")
```




```r
chisq_result_race$stdres
```

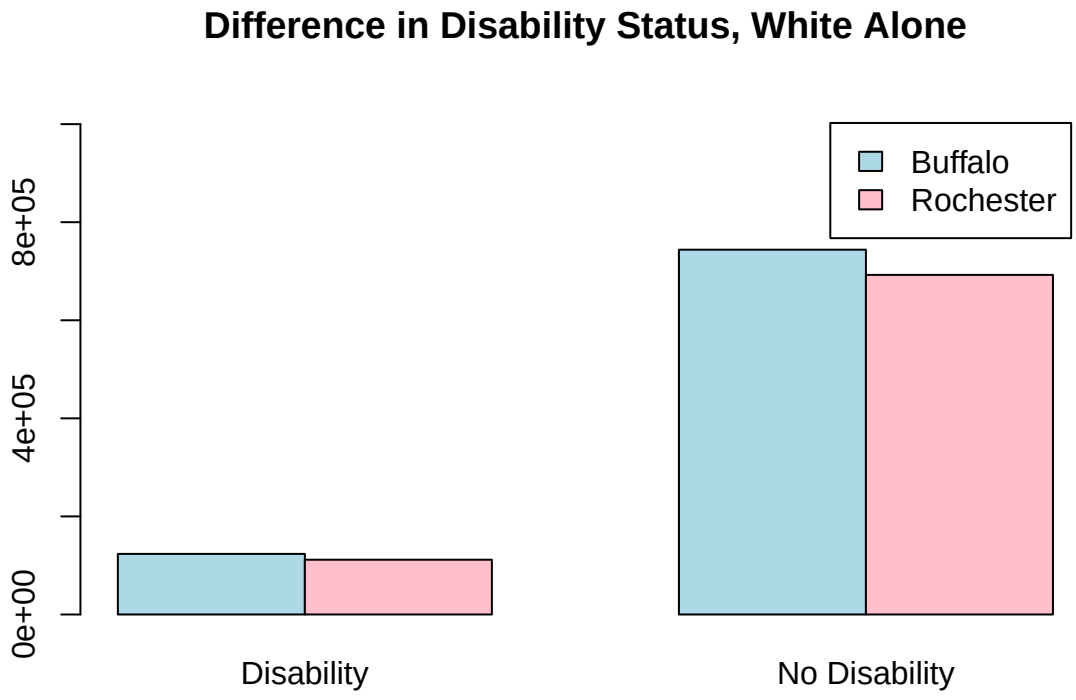
```
##           Disability No Disability
## Buffalo    -35.06721      35.06721
## Rochester   35.06721     -35.06721
```

Population - White Alone

```r
#White
race_table <- matrix(c(race$Buf_Population_Disability[1], race$Buf_Population_Nondisabled[1],
                       race$Roch_Population_Disability[1], race$Roch_Population_Nondisabled[1]),
                     nrow = 2, byrow = TRUE,
                     dimnames = list(c("Buffalo", "Rochester"), c("Disability", "No Disability")))

chisq_result_race <- chisq.test(race_table)
print(chisq_result_race)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  race_table
## X-squared = 43.682, df = 1, p-value = 3.864e-11
```

```
barplot(race_table,
        beside = TRUE,
        col = c("lightblue", "pink"),
        ylim = c(0, max(race_table) * 1.4),
        legend = rownames(race_table),
        main = "Difference in Disability Status, White Alone")
```




```
chisq_result_race$stdres
```

```
##           Disability No Disability
## Buffalo     6.611436     -6.611436
## Rochester  -6.611436      6.611436
```

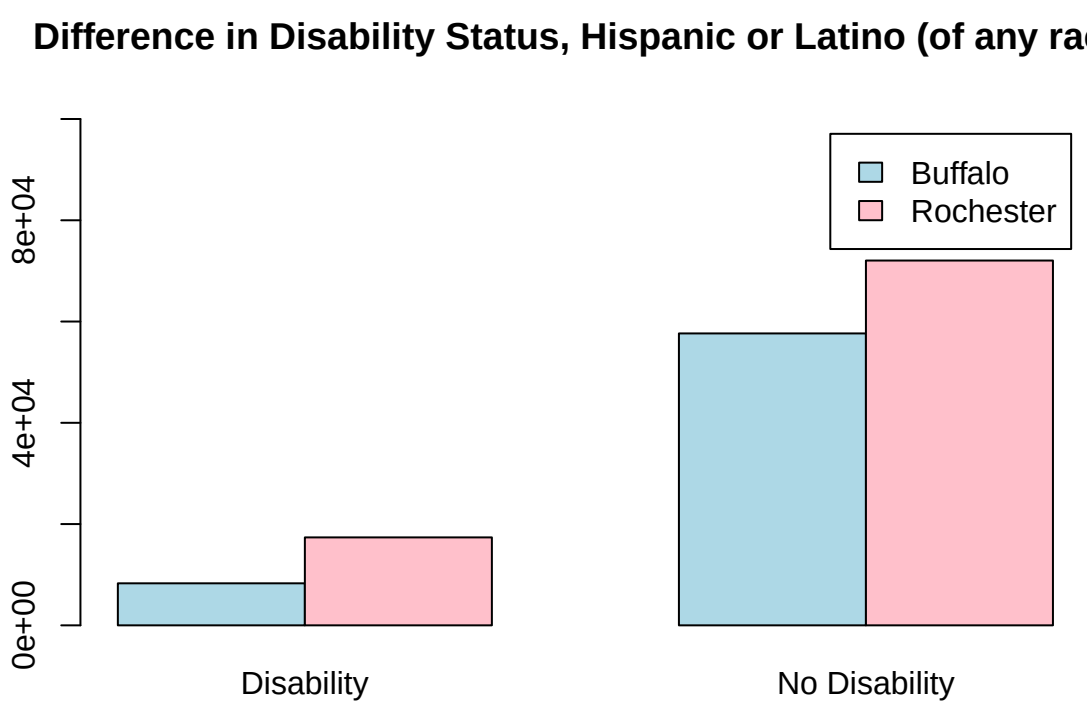
Population - Hispanic or Latino

```
#Hispanic or Latino
race_table <- matrix(c(race$Buf_Population_Disability[9], race$Buf_Population_Nondisabled[9],
```

```
                        race$Roch_Population_Disability[9], race$Roch_Population_Nondisabled[9]),
                      nrow = 2, byrow = TRUE,
                      dimnames = list(c("Buffalo", "Rochester"), c("Disability", "No Disability")))

chisq_result_race <- chisq.test(race_table)
print(chisq_result_race)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  race_table
## X-squared = 1287.7, df = 1, p-value < 2.2e-16
```

```
barplot(race_table,
        beside = TRUE,
        col = c("lightblue", "pink"),
        ylim = c(0, max(race_table) * 1.4),
        legend = rownames(race_table),
        main = "Difference in Disability Status, Hispanic or Latino (of any race)")
```




```
chisq_result_race$stdres
```

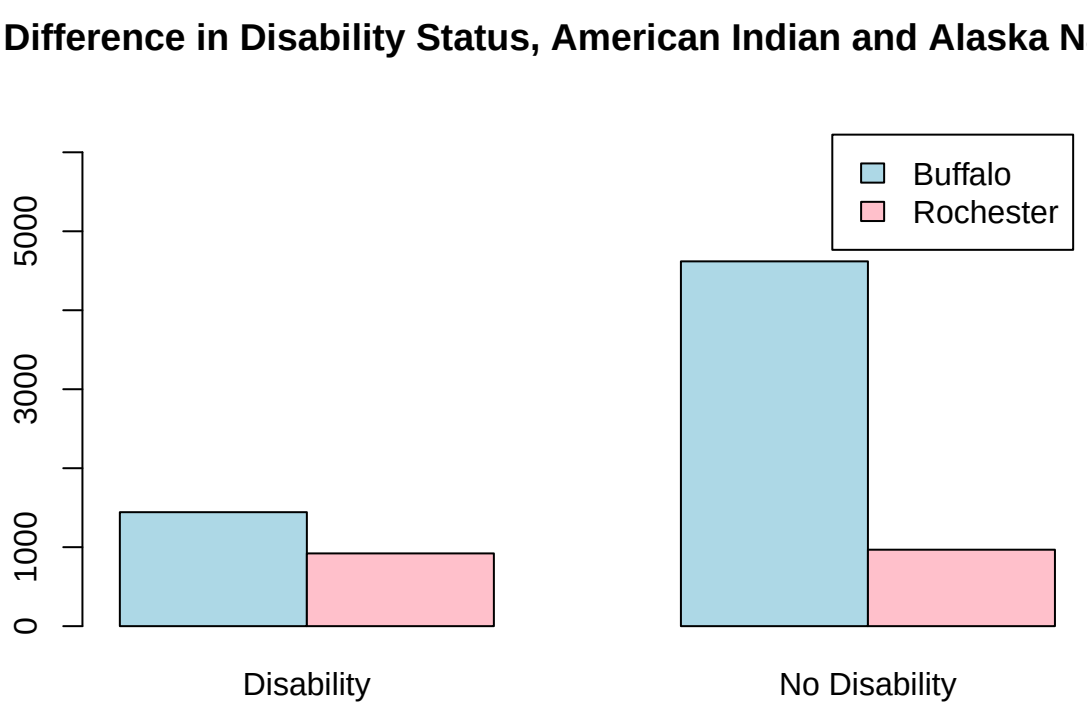
```
##           Disability No Disability
## Buffalo    -35.89181      35.89181
## Rochester   35.89181     -35.89181
```

Population - American Indian and Alaska Native Alone

```
#American Indian and Alaska Native alone

race_table <- matrix(c(race$Buf_Population_Disability[3], race$Buf_Population_Nondisabled[3],
                       race$Roch_Population_Disability[3], race$Roch_Population_Nondisabled[3]),
                     nrow = 2, byrow = TRUE,
                     dimnames = list(c("Buffalo", "Rochester"), c("Disability", "No Disability")))

chisq_result_race <- chisq.test(race_table)
print(chisq_result_race)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  race_table
## X-squared = 428, df = 1, p-value < 2.2e-16
```

```
barplot(race_table,
        beside = TRUE,
        col = c("lightblue", "pink"),
        legend = rownames(race_table),
        ylim = c(0, max(race_table) * 1.4),
        main = "Difference in Disability Status, American Indian and Alaska Native")
```



```
chisq_result_race$stdres
```

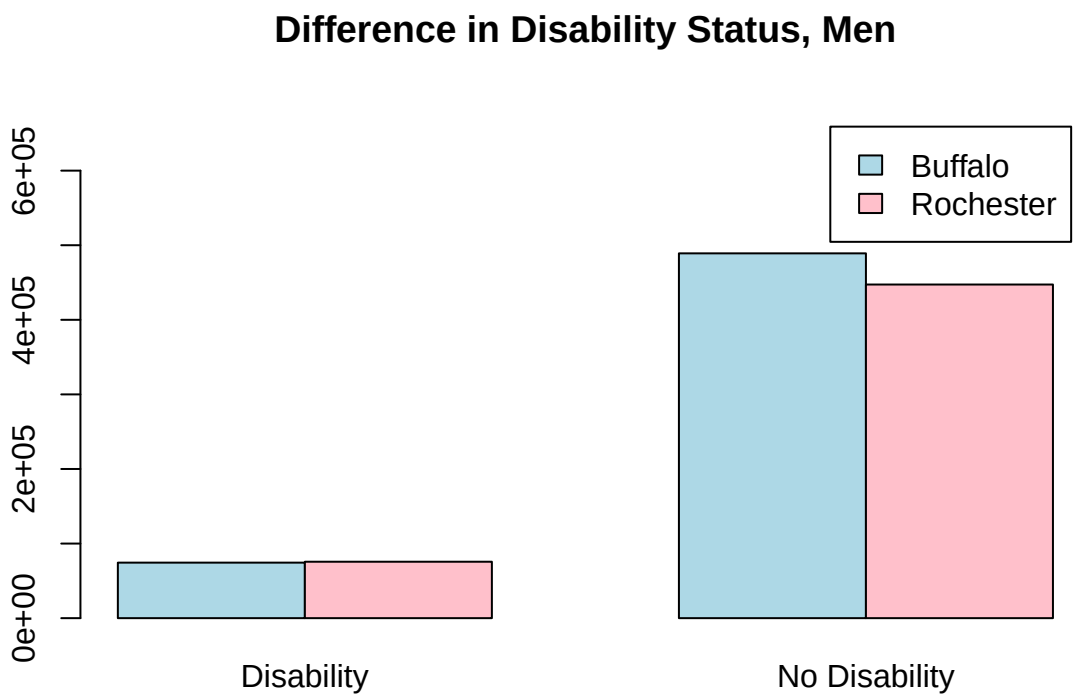
```
##           Disability No Disability
## Buffalo    -20.71695      20.71695
## Rochester   20.71695     -20.71695
```

Population - Men

```
males_table <- matrix(c(sex$Buf_Population_Disability[1], sex$Buf_Population_Nondisabled[1],
                        sex$Roch_Population_Disability[1], sex$Roch_Population_Nondisabled[1]),
                      nrow = 2, byrow = TRUE,
                      dimnames = list(c("Buffalo", "Rochester"), c("Disability", "No Disability")))

chisq_result_males <- chisq.test(males_table)
print(chisq_result_males)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  males_table
## X-squared = 362.07, df = 1, p-value < 2.2e-16
```

```
barplot(males_table,
        beside = TRUE,
        col = c("lightblue", "pink"),
        legend = rownames(males_table),
        ylim = c(0, max(males_table) * 1.4),
        main = "Difference in Disability Status, Men")
```




```
chisq_result_males$stdres
```

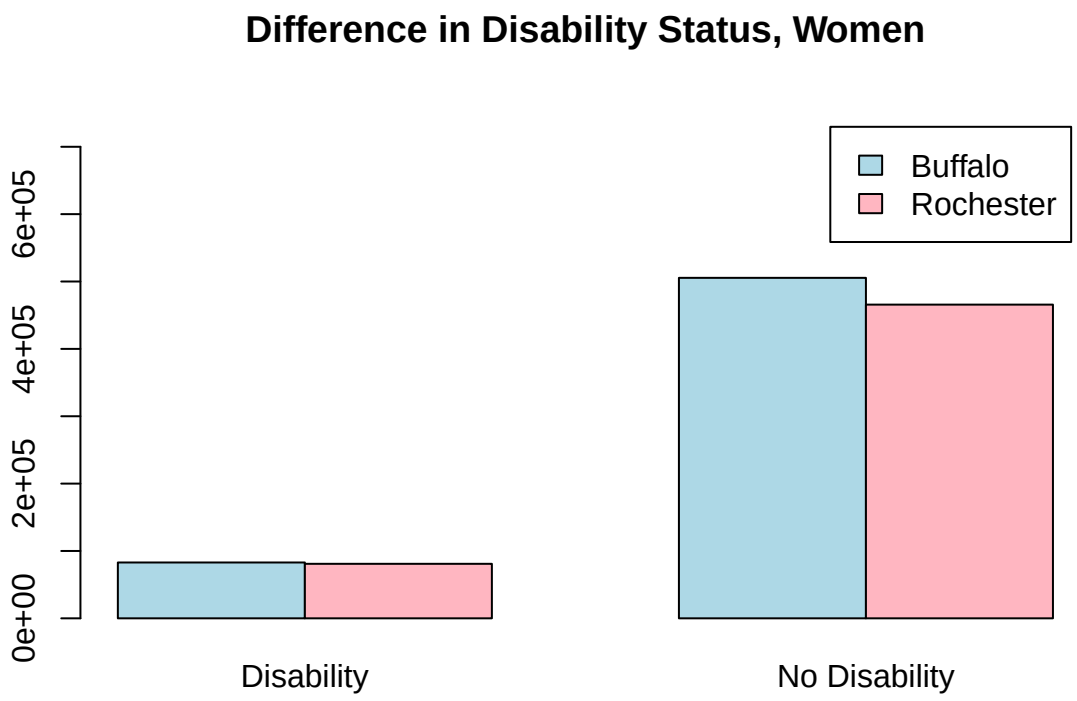
```
##            Disability No Disability
## Buffalo     -19.0308       19.0308
## Rochester    19.0308      -19.0308
```

Population - Women

```
females_table <- matrix(c(sex$Buf_Population_Disability[2], sex$Buf_Population_Nondisabled[2],
                          sex$Roch_Population_Disability[2], sex$Roch_Population_Nondisabled[2]),
                        nrow = 2, byrow = TRUE,
                        dimnames = list(c("Buffalo", "Rochester"), c("Disability", "No Disability")))

chisq_result_females <- chisq.test(females_table)
print(chisq_result_females)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  females_table
## X-squared = 115.59, df = 1, p-value < 2.2e-16
```

```r
barplot(females_table,
        beside = TRUE,
        col = c("lightblue", "lightpink"),
        legend = rownames(females_table),
        ylim = c(0, max(females_table) * 1.5),
        main = "Difference in Disability Status, Women")
```




```r
chisq_result_females$stdres
```

```
##            Disability No Disability
## Buffalo     -10.75397      10.75397
## Rochester    10.75397     -10.75397
```

We can note that across all racial demographics, Rochester reported statistically higher disability rates compared to Buffalo, with the exception of those who identify as White, who were linked to higher disability rates in Buffalo.

Conclusion

After analyzing two comparably sized populations in a shared geographic area, it has been established that there is a clear statistical difference in the disability rates between these two cities. Given this information, we can begin to look at what other factors differ between the two populations that may contribute to this

rate difference. Establishing evidence of this disparity is an important first step in addressing potential inequity.

A significant limitation of this data was difficult to look at intersecting demographics. Due to the default aggregation of counts to each demographic group, it was more difficult to analyze trends across multiple overlapping demographics. Additionally, while we can see that there is a statistical difference in disability rates, the data does not necessarily show us what is causing these differences, nor the context of each city's material conditions, only that specific demographics may be more significantly impacted by the conditions that do exist.