

STAA57 W20 Final Project Report

Group 14: Chui Ling Mok, Cassandra Julianne Mojica, and Pablo Jose Araujo Camacho.

Introduction

The main question that is being addressed in this report mostly concerns the geographical locations of drug-related incidents and of the number of drug prevention centres stationed in the near-general area of such incidents. Incidents are focused on arrests/crimes.

Other questions this project will be investigated on are based on the original 211 Caller Data spreadsheet provided for this project which include how many calls are made for different centres at different times, and the general age groups and genders for each caller.

With this data, we are expecting to see general trends of drug-related incidents and drug-related calls compared with the location of centres that aim towards preventing drugs, helping people who are addicted to drugs, and rehabilitation.

Data

The data used for this project mainly focuses on answering the questions as stated in the Introduction. We have used datasets on drug-related crimes and drug-related arrests, which included data on valuable information such as the frequency of such incidents per neighbourhood in Toronto.

We have also made use of the 211 Caller Data, and mostly analyzed with the variables such as postal area, level 4 name, date of call, if need was met, gender, age category, and city name.

All the gatherings were used to create an analysis and view all the trends of incidents vs. available help. If there was any area where there were more incidents than available help, a recommendation will be made. If any area had more events in which the need was not met, and recommendation will be made especially if it is a high-incident area.

All the data and variables used in each dataset for this project is considered trustworthy as they are open public data provided by the government and other reputable sources, so there are easily collected and verified.

- The 211 Caller Data: “211 Caller Data - Toronto - Needs - 2017-2019.csv” (211Ontario.ca).
 - This data source is the one provided by 211.
 - Some of the variables we analyze include *PostalArea* (to find the caller locations), *Level4Name* (to filter for most common reason for people calling), *NeedWasMet* (to filter by whether the call met their needs or not), and *Gender*, *Age Category*, and *CityName* (for finding demographic data and for visualization).
 - The *PostalArea*, *Gender*, and *Age Category* variables are not going to be completely accurate since they contain a lot of *NULL* values from people who did not want to give their personal information.
- Wellbeing Toronto Safety Data: “RawData-Ref Period 2008.csv” and “RawData-Ref Period 2011.csv” (Toronto Open Data Catalog).

- This data source is provided by the City of Toronto.
- One of the files is for data recorded in 2008 and another for data recorded in 2011.
- The two main variables we analyze are *Drug Arrests* and *Neighbourhood* to find the neighbourhoods that have had the most drug arrests and to find the neighbourhoods where drug activity has increased over the two years.
- The data is a few years old now so there might be discrepancies when compared to reality.
- Ministry of Health Service Provider Locations: “Ministry_of_Health_Service_Provider_Locations.csv” (Ontario GeoHub).
 - This data source is provided by the government of Ontario.
 - The main variables we use are *SERVICE_TYPE_DETAIL* (to filter for “Addiction Services” and “Addiction Services and Mental Health Services”), *COMMUNITY* (to filter for services located in the Toronto area), and *X* and *Y* (to plot the services’ location on the map).
 - The data is reliable since it is provided by the provincial government and was last updated last year.
- Neighbourhoods: “Neighbourhoods.csv” (Toronto Open Data Catalog).
 - This data source is provided by the City of Toronto.
 - The main variables we use are *AREA_SHORT_CODE*, *LONGITUDE*, and *LATITUDE* to plot the location of neighbourhoods on the map.
 - The data is reliable since it is provided by the government and was last updated last month.

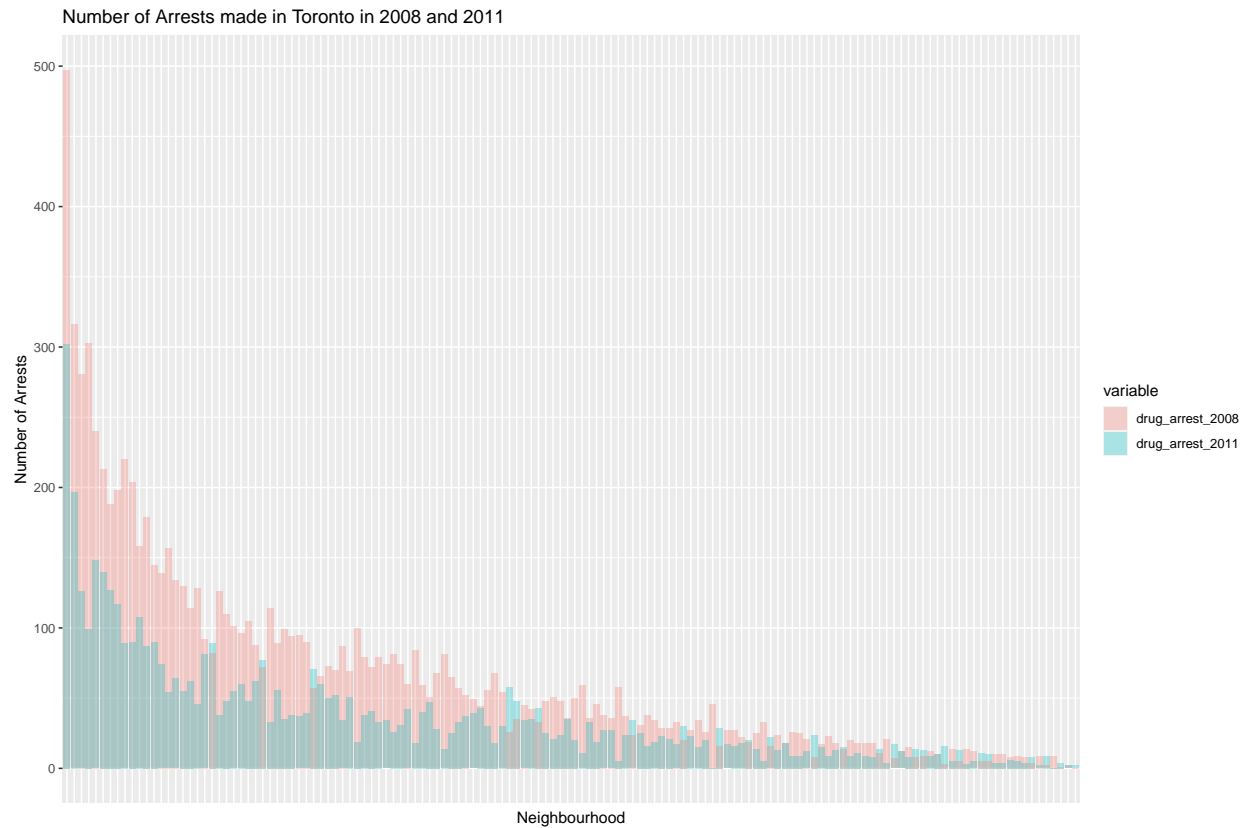
Analysis

The first thing we did when starting the project was figure out what part of the 211 data we wanted to focus on.

After we saw that **most of the calls whose needs weren’t met were calls regarding Detoxification** (shown above), we decided to make that the focus of our project. Within detoxification we decided to focus on drugs. From there we began to research which areas in Toronto had the most drug usage and which of those areas lacked services to aid citizens in the matter.

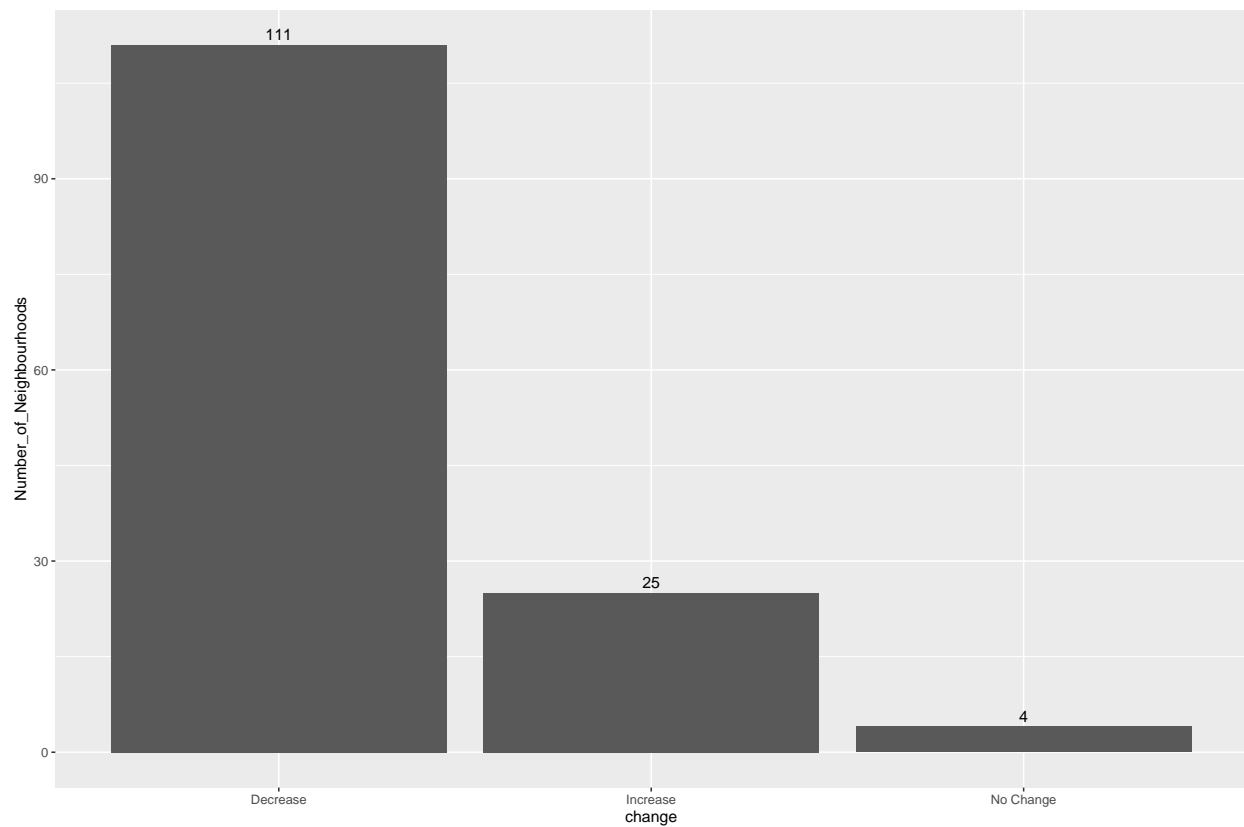
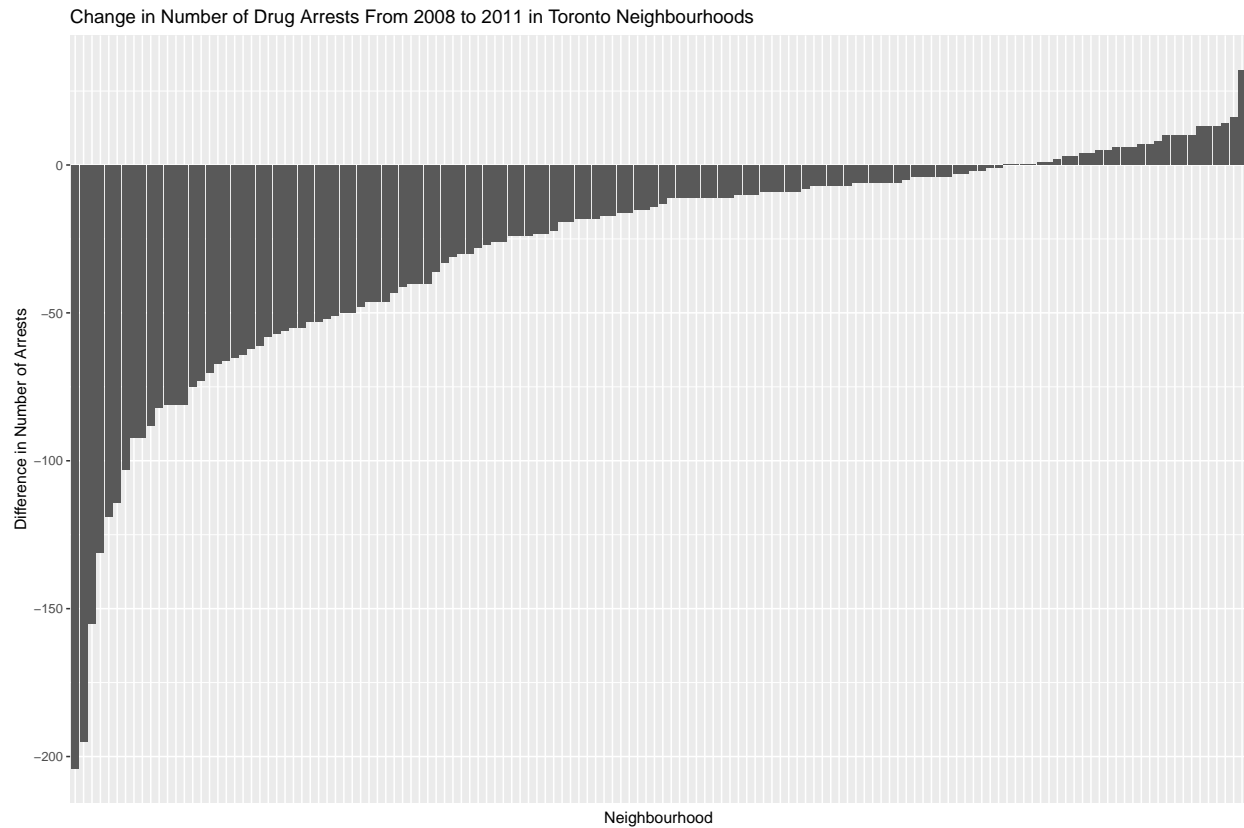
Neighbourhood Drug Incidents

Incidents Per Neighbourhood We first looked at how many incidents occurred in each neighbourhood:



The above graph is arranged such that the neighbourhoods with the most number of drug arrest made in 2008 is on the left to the least number of arrests on the right. The red bars represent the number of drug arrest made in 2008 and the blue bars represent the number of arrest made in 2011. From this graph, we can see that in general, the number of arrests have declined within a 4 year period (2008 - 2011), specially in neighbourhoods with the most drug arrests.

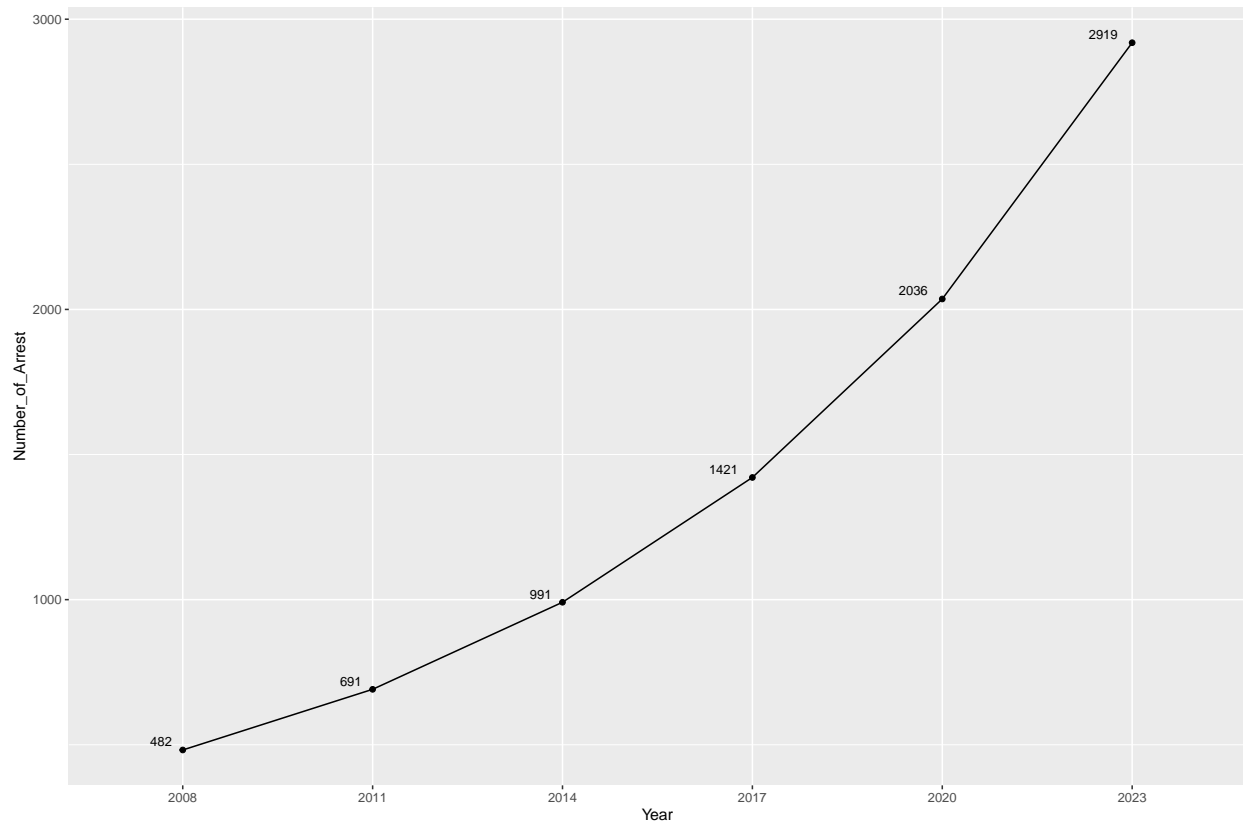
Rate of Change of Incidents The following graphs and tables gives are better visual of how the number of arrests have changed within 4 years:



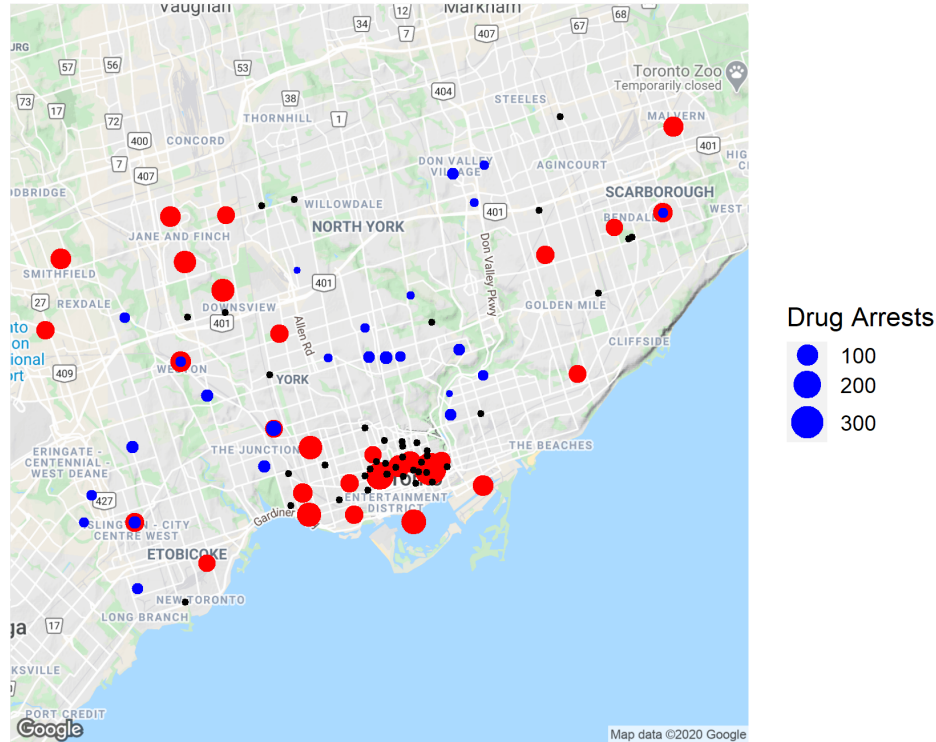
There's up to 111 neighborhoods in Toronto where the number of arrests has decreased within a 4 year

period from 2008 to 2011 and 25 neighborhoods where there's an increase. Even with the overall number of arrests declining, this does not mean that drug problems are not still a serious thing; there's still a large number of arrests made around Toronto dealing with drug abuse.

With this in mind, we decided that since most of the more problematic neighbourhoods are seeing a decline in drug incidents, we would focus our efforts on the neighbourhoods that are seeing an increase in incidents as well as some of the neighbourhoods that have the most arrests. The reason we want to focus on neighbourhoods that have an increase in arrests is so that they might not become even worse like in the estimates in the graph below:



Service Provider Locations

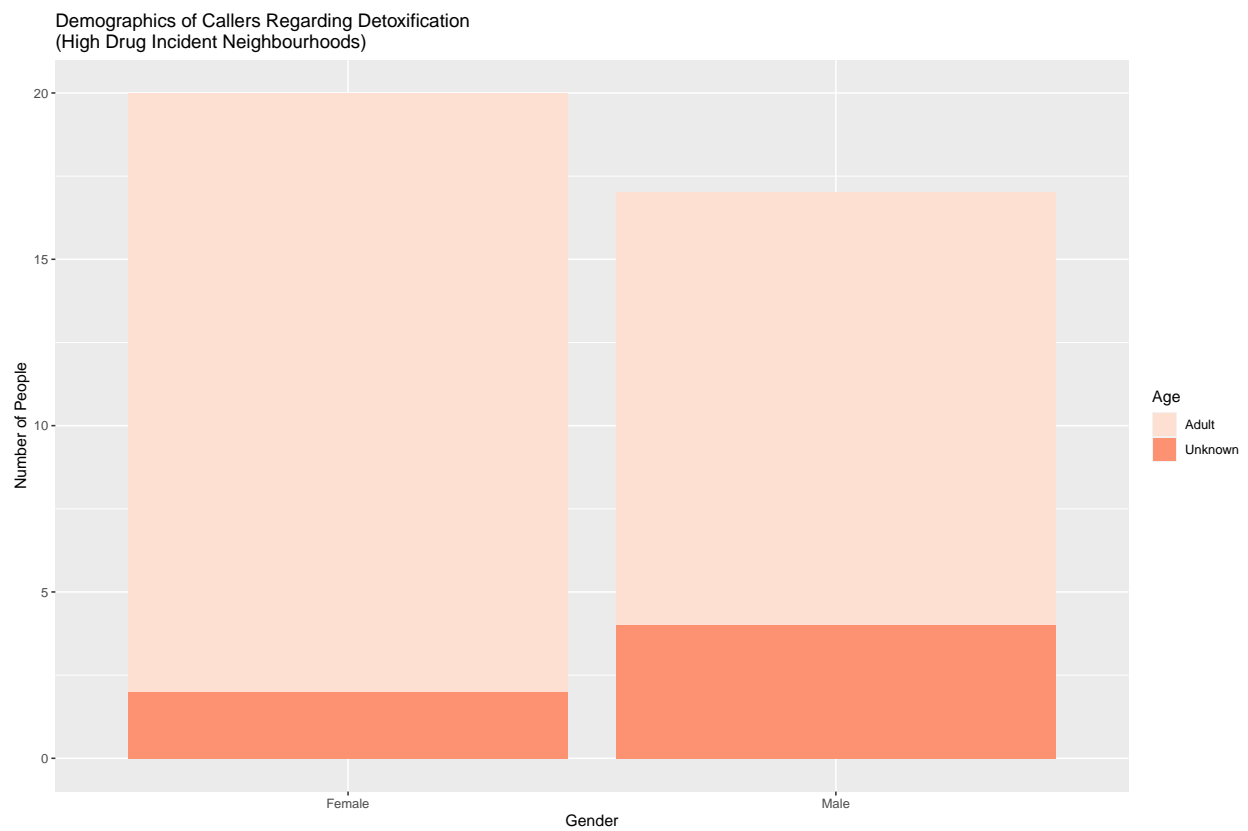
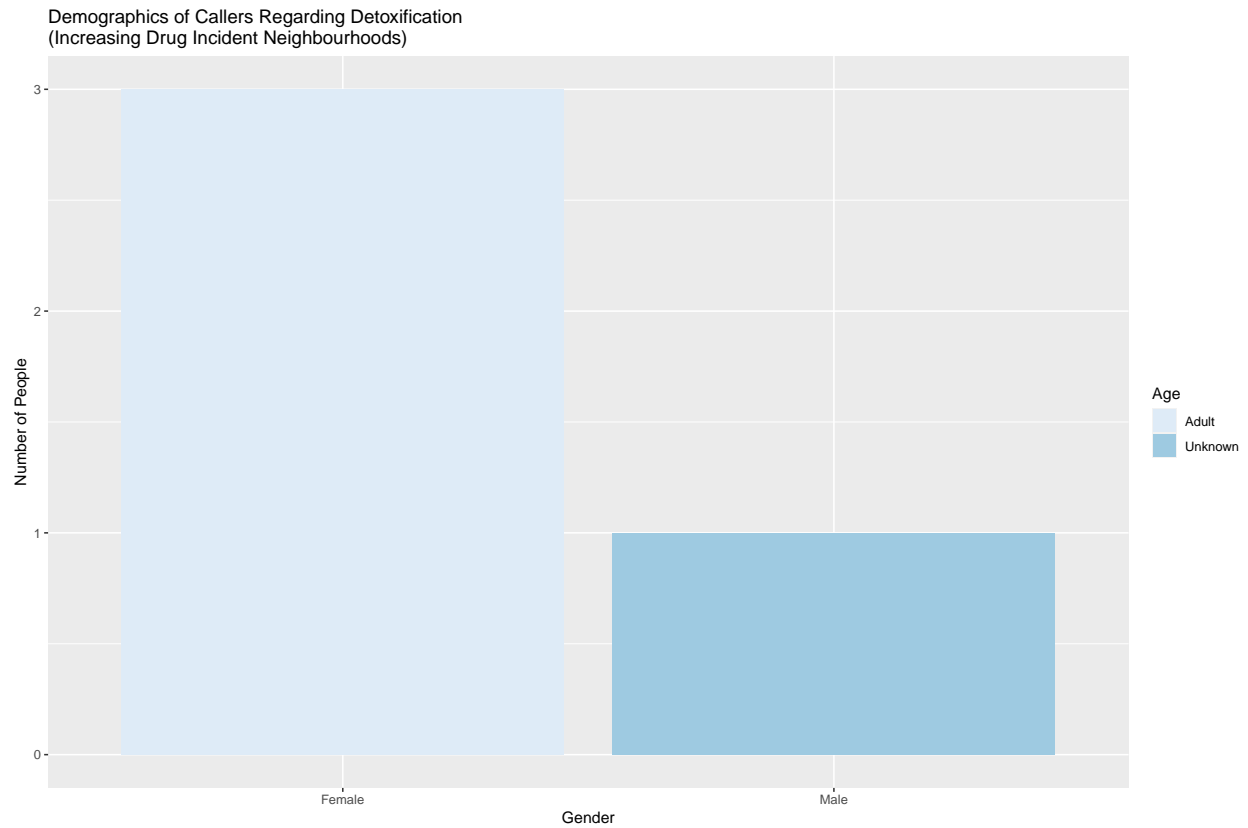


The above map shows the locations of health services (black), the location of neighbourhoods with increasing drug related incidents (blue), and the location of neighbourhoods that have over 50 drug related incidents. As you can see, **there are a number of neighbourhoods that don't have a health service nearby that citizens can be referred to for help** (such as the areas West of and North of the Entertainment District), which are also some of the neighbourhoods with the biggest differences.

This clearly shows that new health centers need to be built closer to these neighbourhoods so that more people can have access to and be referred to services that can help them. Some of these neighbourhoods could include *Islington-City Centre West, West Humber-Clairville, Humbermede, Humber Summit, Mount Pleasant West, Yonge-Eglinton*, and their nearby areas.

The Best Way To Help

Now that we know what neighbourhoods require more health services, we want to find out what demographic these neighbourhoods should expect to target so that they can provide the best service possible to the citizens.



The above graphs show the demographic of callers from our target neighbourhoods calling about Detoxifi-

cation. As you can see, very few people who fit into this category have called seeking help. Because of this, we no longer need to find a demographic which the new health centers should target, and a new strategy is required all together:

Due to the lack of callers in these neighbourhoods, we believe that in order to make the new health centers successful **marketing has to be done in the area to promote the 211 helpline and the new services being offered so that more people are aware of it and reach out for help.**

Summary

After analyzing the 211 Caller Data spreadsheet, we found out that most of the calls that failed to meet needs were calls related to detoxification, so our main topic for this report was to find a way in order to improve this area of service. We have gathered information through different datasets from reliable sources related to drug arrests, location of health centers, demographics of the callers, and the location of callers. With this information, we were able to create clear visualizations showing that health services are not present where most of the callers for detoxifications are located. **Because of this we suggest that new health centers should be built around these areas (such as the *Islington-City Centre West*, *West Humber-Clairville*, *Humbermede*, *Humber Summit*, *Mount Pleasant West*, and *Yonge-Eglinton* neighbourhoods).** We also suggest advertising and making the service more known to potential callers.

Appendix

```
caller_data %>%
  group_by(Level4Name) %>%
  filter(NeedWasMet == FALSE) %>%
  summarize(n_types = n()) %>%
  filter(Level4Name != "NULL") %>%
  arrange(-n_types)
```

The above code creates a table showing the number of times people called for each reason and didn't have their needs met.

The above code is for gathering the 2008 and 2011 Wellbeing Toronto Safety Data and finding the difference between the number of arrests reported between the years for each neighbourhood.

```
# We load the reshape library here since it overrides some of R's default functions and only want it loaded
library(reshape)

melted = difference_graph %>%
  arrange(Difference) %>%
  select(Neighbourhood, drug_arrest_2008, drug_arrest_2011) %>%
  melt(id="Neighbourhood") %>%
  arrange(desc(value))

ggplot(melted, aes(x=reorder(Neighbourhood, -value), y=value, fill=variable)) + geom_bar(stat="identity", position="dodge")

# Unload the reshape library
detach("package:reshape", unload=TRUE)
```

The above code creates a graph showing the number of arrests made in each neighbourhood in both 2008 and 2011.


```

full_graph = ggplot(difference_graph, aes(x = reorder(Neighbourhood, Difference), y = Difference))+geom_bar()

#Changing the values
change_var = difference_graph[difference_graph$change == 1, "change"] <- "Increase"
change_var = difference_graph[difference_graph$change == 0, "change"] <- "No Change"
change_var = difference_graph[difference_graph$change == -1, "change"] <- "Decrease"

full_graph

change_table = difference_graph %>%
  group_by(change) %>%
  count(change)%>%
  rename(Number_of_Neighbourhoods = n)

change_table %>%
  ggplot(aes(x = change, y = Number_of_Neighbourhoods)) + geom_bar(stat = "identity") + geom_text(aes(l

```

The above code creates two graphs, both for visualizing the change in number of drug arrests from 2008 to 2011 in Toronto neighbourhoods.

```

merged = merge(col1,col2) %>%
  mutate(difference= drug_arrest_2011-drug_arrest_2008)%>%
  mutate(change = sign(difference)) %>%
  filter(change == 1) %>%
  select(Neighbourhood, drug_arrest_2008, drug_arrest_2011, difference)%>%
  arrange(desc(difference))

total_2008 = sum(merged$drug_arrest_2008)
total_2011 = sum(merged$drug_arrest_2011)
average = (total_2011/total_2008)

Year <- c("2008", "2011", "2014", "2017", "2020", "2023")
Number_of_Arrest <- c(total_2008,
                      total_2011,
                      ceiling(total_2011*average),
                      ceiling(total_2011*average*average),
                      ceiling(total_2011*average*average*average),
                      ceiling(total_2011*average*average*average*average))
arrests <- data.frame(Year,Number_of_Arrest)

ggplot(arrests, aes(x=Year, y=Number_of_Arrest, group=1)) +
  geom_line()+
  geom_point() +
  geom_text(aes(label = Number_of_Arrest), size = 3, hjust = 1.5, vjust = -0.4)

```

The above code creates a line graph that shows an estimate of how high drug related arrests could get in neighbourhoods where those arrests are currently increasing if nothing is done about the situation.

```

joint <- neighbourhoods %>% rename(`Neighbourhood Id` = AREA_SHORT_CODE) %>% inner_join(x = ., y = arrests)

diff = difference_graph %>% filter(`Difference` > 0)

diff_joint <- neighbourhoods %>% rename(`Neighbourhood Id` = AREA_SHORT_CODE) %>% inner_join(x = ., y =

```

```
# For security, please provide your own key to use the Google Maps Static API.
# Here is a link on how to do this: https://www.littlemissdata.com/blog/maps-register-google(key = "[YOUR KEY GOES HERE]")

TorontoMap <- ggmap(get_googlemap(center = c(-79.398, 43.698), zoom = 11, scale = 4, maptype = 'terrain',
  theme(axis.title.x=element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank(),
        axis.title.y=element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank(),
        legend.key.size = unit(0.3, "cm"))

TorontoMap + geom_point(data = joint, aes(x = LONGITUDE, y = LATITUDE, size = `Drug Arrests`, color =
```

The above code creates a map showing the neighbourhoods in Toronto that have an increase in drug arrests and a high number of drug arrests with the locations of health services overlapping.

```
caller_data %>%  
  filter(Level4Name == "Detoxification" & (PostalArea == "M9B" | PostalArea == "M9C" | PostalArea == "M9D")  
  ggplot(aes(fill = `Demographics of Inquirer - Age Category`, x = `Demographics of Inquirer - Gender`))  
  
caller_data %>%  
  filter(Level4Name == "Detoxification" & (PostalArea == "M3J" | PostalArea == "M6B" | PostalArea == "M6C")  
  ggplot(aes(fill = `Demographics of Inquirer - Age Category`, x = `Demographics of Inquirer - Gender`))
```

The above code creates a graph that shows the demographics of callers in certain neighbourhoods that have called regarding detoxification.