Don't Break a Leg! Road Safety in the City of Toronto

STA2453 - Project II Draft

Sunwoo (Angela) Kang, Sergio E. Betancourt, Jing Li, and Jiahui (Eddy) Du

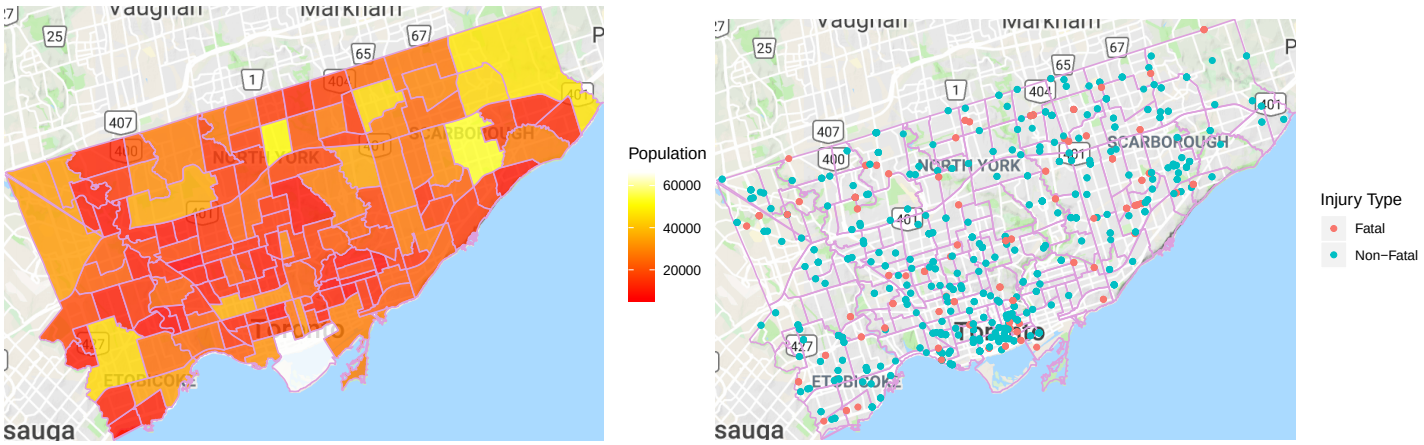
2019-03-08

Introduction

Road traffic safety is a crucial component of urban planning and development. Nowadays governments (and sometimes the private sector) dedicate significant resources to providing ample and sufficient infrastructure to accommodate diverse modes of transportation, thereby increasing the productivity of any given urban area. In this project we examine road safety in the City of Toronto from 2007 to 2017 and explore the areas with highest risk of a traffic incident, controlling for different factors.

Methods

We define the City of Toronto as per the these guidelines (https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/neighbourhood-profiles/). Below are the neighborhood limits and the 2016 population estimates:



(a) Population by neighborhood in the census year 2016

(b) Fatal and Non-Fatal Collisions in 2016

Figure 1: EDA with regards to the City of Toronto

Primary Questions

The analysis focuses on answering two main questions:

1. Given a collision occurred which areas in Toronto are the most deadly controlling for other factors?
2. Which factors are related to the collision safety of neighbourhoods?

Data Collection

For our analysis we employed data from the Toronto Police Service, the City of Toronto, and Environment Canada. Each of these datasets contains different levels of granularity and information, and were therefore combined to obtain the following variables of interest outlined in **Appendix: Dataset Variables and Definitions**.

Data Preparation

The following table provides an overview of the merged data.

Accident_ID	Fatal	Date	Neighborhood	Population	Max_Temp
5002235651	1	2015-12-30	Greenwood-Coxwell	7072	4.7
5000995174	1	2015-06-13	Annex	26703	22.3
5000995174	1	2015-06-13	Annex	26703	22.3
1249781	0	2011-08-04	Bay Street Corridor	19348	26.4

Traffic incident information provided by Toronto Police served as a base for the data used for this analysis. Each of the 12,557 entries represent a party involved in a traffic collision event where a person was either killed or seriously injured. Other features such as the location of the collision (intersection, neighborhood, ward), road condition (visibility, road precipitation), driver action (e.g. speeding, involved alcohol), and type of vehicles (e.g. automobile, pedestrian, cyclist) involved were also used.

Population counts for 2011 and 2016 are available through the national census for each neighborhood. The populations for the dates not provided by the census were extrapolated using a linear growth model.

Historical weather data collected from the station in University of Toronto was also merged based on the day the accident occurred.

Exploratory Analysis

By summing up counts from 2007 to 2017, West Humber-Clairville appears to be the deadliest intersection followed by South Parkdale, then Wexford/Maryvale. Thankfully, the fatalities appear to be quite low compared to the total number of collisions reported by the Toronto Police.

Neighbourhood	Total Fatalities	Total Collisions
West Humber-Clairville	22	426
South Parkdale	21	197
Wexford/Maryvale	15	225
Clairlea-Birchmount	14	193
Waterfront Communities-The Island	14	492
Glenfield-Jane Heights	11	105

West Humber-Clairville, and Wexford/Maryvale appear again as a dangerous neighborhood even when focussing on pedestrian or cyclist fatalities.

Neighbourhood	Total Pedestrian Fatalities	Total Pedestrian Collisions
Clairlea-Birchmount	11	39
Wexford/Maryvale	10	45
Moss Park	9	35
West Humber-Clairville	9	43
Newtonbrook West	8	27
Waterfront Communities-The Island	8	87

Neighbourhood	Total Cyclist Fatalities	Total Cyclist Collisions
South Parkdale	3	9
Dovercourt-Wallace Emerson-Junction	2	10
Kensington-Chinatown	2	19
Wexford/Maryvale	2	4
Annex	1	14
Bay Street Corridor	1	25

Neighbourhood	Total Other Fatalities	Total Other Collisions
South Parkdale	14	163
West Humber-Clairville	12	376
Islington-City Centre West	8	198
Glenfield-Jane Heights	6	82
Don Valley Village	5	73
Downsview-Roding-CFB	5	150

Modeling

We model our outcome of interest (fatal collision) using a generalized mixed effects model (to be expanded to spatial in the following iteration), clustered by neighborhood. We estimate the odds of experiencing a fatal accident with respect to experiencing a non-fatal one, accross neighborhoods in Toronto, controlling for each day's total precipitation and minimum temperature. We will continue exploring model complexity and structure for the next iteration.

Results

Our model indicates that a one milimeter increment of total precipitation for any neighborhood in the timeframe in question leads to an increment of 1.2% in the odds of suffering a fatal accident.

Conclusions and Discussion

One of the biggest limitations in our project has been data quality and granularity. The data made available by Geotab does not include large areas of the City of Toronto. Moreover, there are plenty missing observations. We also acknowledge the fact that the collision information we procured from the Toronto Police Service may not describe perfectly the actual number of incidents, as there are many of these that are non-fatal or go unreported.

Appendix: Dataset Variables and Definitions

Feature	Description	Source
YEAR	Year in range (2007-2017) inclusive	Automobile (Toronto Police)
MONTH	Month in range 1-12 inclusive	Automobile (Toronto Police)
Ward_ID	Ward in range (1-44) inclusive	Automobile (Toronto Police)
IncidentsTotal_TP	Total number of incidents	Automobile (Toronto Police)
Dark	Count accidents ocurred on dark conditions	Automobile (Toronto Police)
Dawn	Count accidents ocurred on dawn conditions	Automobile (Toronto Police)
Daylight	Count accidents ocurred on daylight conditions	Automobile (Toronto Police)
Dusk	Count accidents ocurred on dusk conditions	Automobile (Toronto Police)
Inv_PED	Count accidents involved pedstrains	Automobile (Toronto Police)
Inv_CYC	Count accidents involved cyclists	Automobile (Toronto Police)
Inv_AM	Count accidents involved automobiles	Automobile (Toronto Police)
Inv_MC	Count accidents involved motorcycles	Automobile (Toronto Police)
Inv_TC	Count accidents involved trucks	Automobile (Toronto Police)
Speeding	Count accidents ocurred on speeding condition	Automobile (Toronto Police)
Ag_Driv	Count accidents ocurred on angry driving condition	Automobile (Toronto Police)
Redlight	Count accidents ocurred with redlight	Automobile (Toronto Police)
Alcohol	Count accidents ocurred with driver with alcohol	Automobile (Toronto Police)
Disability	Count accidents ocurred with driver with disability	Automobile (Toronto Police)
SeverityScore	Average Score of Severitylevel (harsh brake)	HDA(Geotab)
IncidentsTotal_Geotab	Monthly average of total number of incidents	HDA(Geotab)
AvgAcceleration	Monthly average acceleration	RI(Geotab)
PercentOfVehicles	Monthly average on percentage of vehicles	RI(Geotab)
AvgMonthlyVolume	Monthly average on vehicle volumes	RI(Geotab)
PercentCar	Monthly average on car percentage	RI(Geotab)
PercentMPV	Monthly average on MPV percentage	RI(Geotab)
PercentLDT	Monthly average on LDT percentage	RI(Geotab)
PercentMDT	Monthly average on MDT percentage	RI(Geotab)
PercentHDT	Monthly average on HDT percentage	RI(Geotab)
PercentOther	Monthly average on other vehicle percentage	RI(Geotab)
Daily_dif	Monthly average on daily Weather change (in celsus)	Weather
Max_Temp	Monthly max on highest daily Weather degree (celsus)	Weather
Min_Temp	Monthly min on lowest daily Weather degree (celsus)	Weather
Ave_Temp	Monthly average on daily average Weather (in celsus)	Weather
Rain_vol	Monthly average on daily rain volumn	Weather
Snow_vol	Monthly average on daily snow volumn	Weather

Appendix: Neighborhoods of Toronto

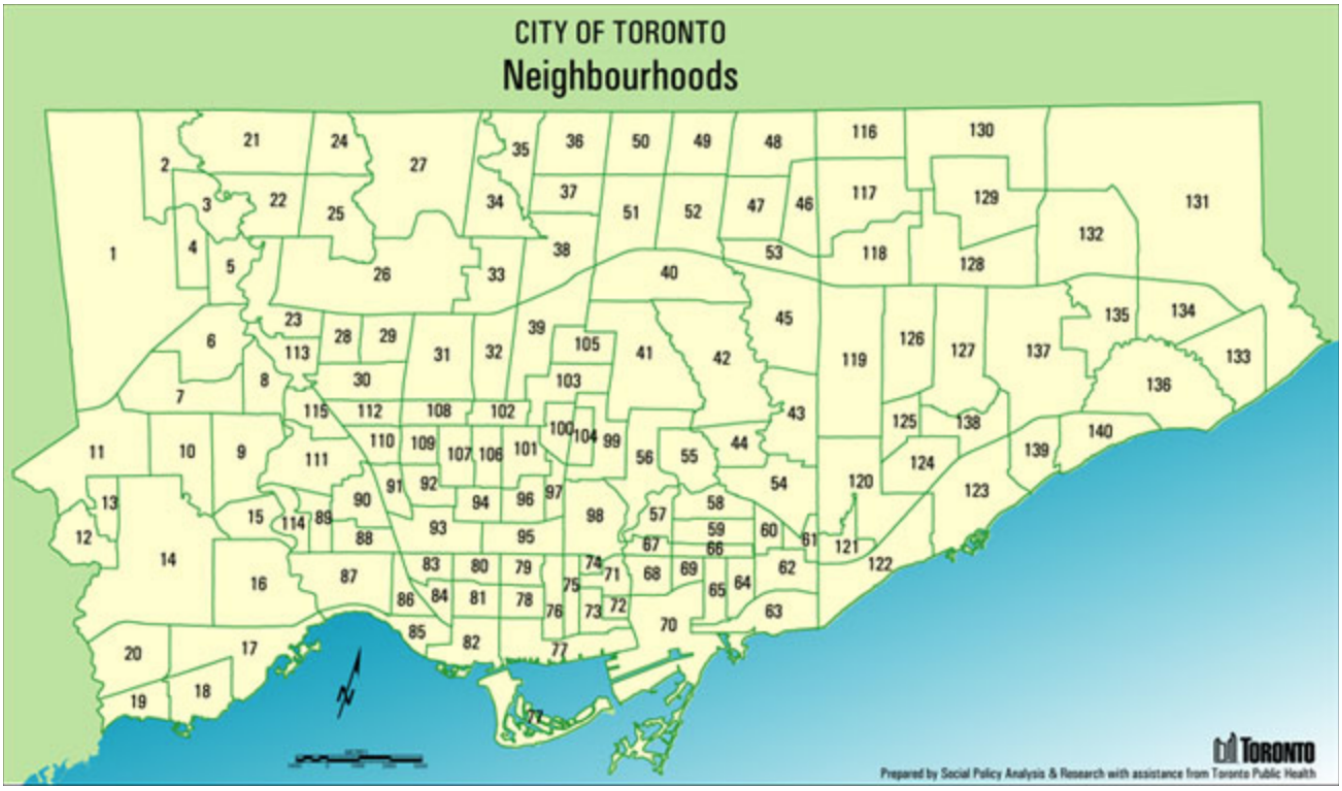


Figure 2: Official City of Toronto Neighborhoods

Refer to the **City of Toronto** for the neighborhood names matching the indeces above.

Appendix: Code

```
library(MASS); library(lmtest); library(knitr); library(ggmap); library(kableExtra); library(nleqslv);
library(Pmisc); library(extrafont); library(VGAM); library(INLA); library(MEMSS); library(rgdal); library(ggp
library(nlme); library(ciTools); library(sf); library(tibble); library(sp); library(plyr);library(dplyr);
knitr::opts_chunk$set(fig.pos = 'H');
# Loading polygon and population data from the City of Toronto
population <- read.csv("https://raw.githubusercontent.com/sergiosonline/data_sci_geo/master/data/neighbourhoo

require(sf)
shape <- read_sf(dsn = "https://raw.githubusercontent.com/sergiosonline/data_sci_geo/master/data/neighbourhoo

neighborhoods <- shape

# Adding populaation info to neighborhood polygon
neighborhoods <- add_column(neighborhoods, '2016pop'=NA, 'x_coords' = NA, 'y_coords' = NA)

# Separating X and Y coordinates from polygon
for (hood in neighborhoods$AREA_NAME) {
  ## Adding population
  pop = as.numeric(neighborhoods[neighborhoods$AREA_NAME == hood,][["AREA_S_CD"]])
  neighborhoods[neighborhoods$AREA_NAME == hood,]$'2016pop' =
    population[population$HoodID == pop,]$Pop2016
  ## Adding x-y
  temp = unlist(subset(neighborhoods,AREA_NAME == hood)$geometry[[1]])
  ll = length(temp)
  x_coord = list(temp[1:(ll/2)])
  y_coord = list(temp[((ll/2)+1):ll])
  neighborhoods[neighborhoods$AREA_NAME == hood,]$x_coords = x_coord
  neighborhoods[neighborhoods$AREA_NAME == hood,]$y_coords = y_coord
}

st_write(neighborhoods,"~/Desktop/Grad_School/COURSEWORK/Spring 2019/Data Science/rough work/neighbourhoods_p
         , delete_layer = TRUE)

neighborhoods <- read_sf(dsn = "~/Desktop/Grad_School/COURSEWORK/Spring 2019/Data Science/rough work/neighbou

###ALTERNATIVE VISUALIZATION
neighborhoods = rgdal::readOGR(dsn = "~/Desktop/Grad_School/COURSEWORK/Spring 2019/Data Science/rough work/ne
accidents <- read.csv("https://raw.githubusercontent.com/sergiosonline/data_sci_geo/master/data/accidents.csv


# Set up df
neighborhoods@data$id = rownames(neighborhoods@data)
neighborhoods.points = fortify(neighborhoods, region="id")
neighborhoods.df = join(neighborhoods.points, neighborhoods@data, by = "id")

# Plotting command - basic

#ggplot(neighborhoods.df) + aes(long,lat,group=group,fill=X2016pop)+ geom_polygon() +
#+   geom_path(color="black") + coord_equal()

# Adding points

#sum_accidents <- accidents %>%
#  group_by(Neighbourhood, YEAR) %>%
#  summarize(`Total Fatalities` = sum(INJURY == "Fatal", na.rm = T),
 #           `Total Collisions` = n()) %>%
```

```r
#  arrange(desc(`Total Fatalities`))

cbPalette <- c("#999999", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7")

#To use for fills, add
#scale_fill_manual(values=cbPalette)

# To use for line and point colors, add
#scale_colour_manual(values=cbPalette)


ggmap::register_google(key = "AIzaSyB13QyZy3PLnR5BYGtwezYWFaSq_pjrNjA")


#####
p0 <- ggmap(get_googlemap(center = c(lon = -79.384293, lat = 43.71),
                    zoom = 10, scale = 2,
                    maptype ='terrain',
                    color = 'color'), maprange=T,extent = "normal") +
    labs(x = "", y = "") +
    scale_x_continuous(limits = c(-79.63926, -79.11524), expand = c(0, 0)) +
scale_y_continuous(limits = c(43.581, 43.85546), expand = c(0, 0)) +
  theme(legend.position = "right",
        panel.background = element_blank(),
        axis.line = element_blank(),
        axis.text = element_blank(),
        axis.ticks = element_blank(),
        plot.margin = unit(c(0, 0, -1, -1), 'lines')) +
  xlab('') +
  ylab('')

p2 <- p0 + geom_polygon(aes(long,lat,group=group,fill=NA,color="white"),color="plum",fill=NA,data=neighborhoo
                        breaks=c("Fatal", "Non-Fatal Injury"),
                        labels=c("Fatal", "Non-Fatal"))

p1 <- p0 + geom_polygon(data=neighborhoods.df, aes(long,lat,group=group, fill=X2016pop),alpha = 0.8,color="pl

p1

p2

# Visualization of fatal vehicular incidents in the City of Toronto 2010-2016
collisiondat <- read.csv("https://raw.githubusercontent.com/sergiosonline/data_sci_geo/master/data/Fatal_Coll

coordinates(collisiondat) <- ~LONGITUDE+LATITUDE
#4326 - WGS84 std
proj4string(collisiondat) <- "+init=epsg:3034" #"+init=epsg:4326"
data_L93 <- spTransform(collisiondat, CRS("+proj=lcc +lat_1=44 +lat_2=49 +lat_0=46.5 +lon_0=3 +x_0=490000 +y_
#x_0/y_0 = 0.1060606


url1 <- "https://raw.githubusercontent.com/sergiosonline/data_sci_geo/master/reports/draft/STA2453-Toronto-20
download.file(url = url1,
          destfile = "toronto_incidents.png",
          mode = 'wb')

knitr::include_graphics(path="Toronto-2016.png")
```

```r
#spTransform() #Transform polygon or raster into Euclidian object - 3026 is Google std

data.frame(Accident_ID = c(5002235651, 5000995174, 5000995174, 1249781),
           Fatal = c(1, 1, 1, 0),
           Date = c("2015-12-30", "2015-06-13", "2015-06-13", "2011-08-04"),
           Neighborhood = c("Greenwood-Coxwell", "Annex", "Annex", "Bay Street Corridor"),
           Population = c(7072, 26703, 26703, 19348),
           `Max_Temp` = c(4.7, 22.3, 22.3, 26.4)) %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped"))
accidents <- read.csv("https://raw.githubusercontent.com/sergiosonline/data_sci_geo/master/data/accidents.csv
                      check.names = F)

accidents %>% group_by(Neighbourhood) %>%
  summarize(`Total Fatalities` = sum(INJURY == "Fatal", na.rm = T),
            `Total Collisions` = n()) %>%
  arrange(desc(`Total Fatalities`)) %>%
  head() %>%
  kable()%>%
  kable_styling(bootstrap_options = c("striped"))


accidents %>% mutate(Pedestrian = INVTYPE == "Pedestrian",
                     Cyclist = INVTYPE == "Cyclist",
                     Other = INVTYPE != "Pedestrian" & INVTYPE != "Cyclist") %>%
  group_by(Neighbourhood) %>%
  summarize(`Total Pedestrian Fatalities` = sum(INJURY == "Fatal" & Pedestrian == 1, na.rm = T),
            `Total Pedestrian Collisions` = sum(Pedestrian == 1, na.rm = T)) %>%
  arrange(desc(`Total Pedestrian Fatalities`)) %>%
  head() %>%
  kable()%>%
  kable_styling(bootstrap_options = c("striped"))

accidents %>% mutate(Pedestrian = INVTYPE == "Pedestrian",
                     Cyclist = INVTYPE == "Cyclist",
                     Other = INVTYPE != "Pedestrian" & INVTYPE != "Cyclist") %>%
  group_by(Neighbourhood) %>%
  summarize(`Total Cyclist Fatalities` = sum(INJURY == "Fatal" & Cyclist == 1, na.rm = T),
            `Total Cyclist Collisions` = sum(Cyclist == 1, na.rm = T)) %>%
  arrange(desc(`Total Cyclist Fatalities`)) %>%
  head()%>%
  kable()%>%
  kable_styling(bootstrap_options = c("striped"))

accidents %>% mutate(Pedestrian = INVTYPE == "Pedestrian",
                     Cyclist = INVTYPE == "Cyclist",
                     Other = INVTYPE != "Pedestrian" & INVTYPE != "Cyclist") %>%
  group_by(Neighbourhood) %>%
  summarize(`Total Other Fatalities` = sum(INJURY == "Fatal" & Other == 1, na.rm = T),
            `Total Other Collisions` = sum(Other == 1, na.rm =T)) %>%
  arrange(desc(`Total Other Fatalities`)) %>%
  head()%>%
  kable()%>%
  kable_styling(bootstrap_options = c("striped"))
# Loading final monthly incident data, by neighborhood
incidentdata <- read.csv("https://raw.githubusercontent.com/sergiosonline/data_sci_geo/master/data/accidents.

#incidentdata$Population2 <- incidentdata$Population/1000
```

```r
#incidentdata$Days_since_start2 <- incidentdata$Days_since_start/100
#incidentdata <- filter(incidentdata, ACCLASS != "Property Damage Only")

#population <- read.csv("https://raw.githubusercontent.com/sergiosonline/data_sci_geo/master/data/toronto_hoo

#Adding neighborhood area
#incidentdata_test <- incidentdata %>%
#  left_join(dplyr::select(population, HoodID, area_sqkm), by = c("Hood_ID" = "HoodID")) #%>% mutate(density

#write.csv(incidentdata_test, "~/Desktop/Grad_School/COURSEWORK/Spring 2019/Data Science/rough work/accidents

freqmod1 <- glmer(as.factor(ACCLASS) ~ Days_since_start2 + Tot_precip + Min_temp + (1 + Days_since_start2 |Ne
                  control=glmerControl(optimizer= "Nelder_Mead"))
var_def <- read.csv("https://raw.githubusercontent.com/sergiosonline/data_sci_geo/master/reports/draft/variab

knitr::kable(var_def, format="latex", booktab=T, linesep = "") #escape=F,
## Visualizing neighborhoods of Toronto for reference
url7 <- "https://raw.githubusercontent.com/sergiosonline/data_sci_geo/master/reports/draft/toronto-hoods.png"
download.file(url = url7,
          destfile = "toronto-hoods.png",
          mode = 'wb')

knitr::include_graphics(path="toronto-hoods.png")
```