

# **Analysis of Break and Enter (B&E) Crimes in Toronto**

**STA130 Final Project**

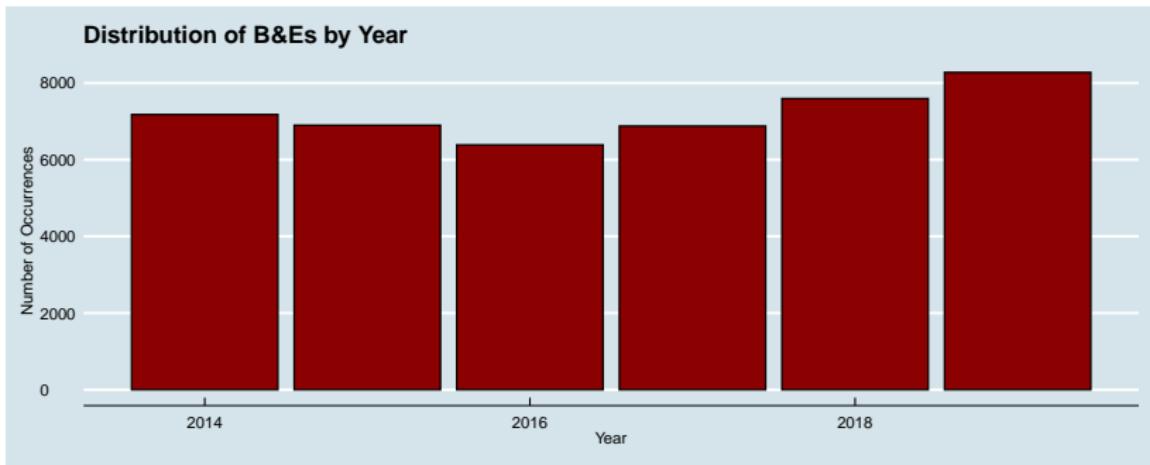
Eric Yin, Tales Scopinho, Fadi Maamarbachi, Audric Ghosh  
TUT204(1)

March 29, 2020

## Context and Introduction

- In this project, we analyze B&Es reported in the Greater Toronto Area in a dataset provided to us by Toronto Police Services (TPS)
- Our goal is to analyze trends in this data in order to make meaningful recommendations to TPS to minimize B&E occurrences in Toronto
- The predominant trend we discovered was that B&Es tended to occur in the absence of people
  - To examine this trend more closely, we decided to focus on the following research questions:
  - Are there any temporal trends (e.g. month, season, time of day) which highlight greater risk for B&Es?
  - Are there any spatial trends (e.g. division, premisetype, neighbourhood) which highlight greater risk for B&Es?

# Why is this analysis needed?

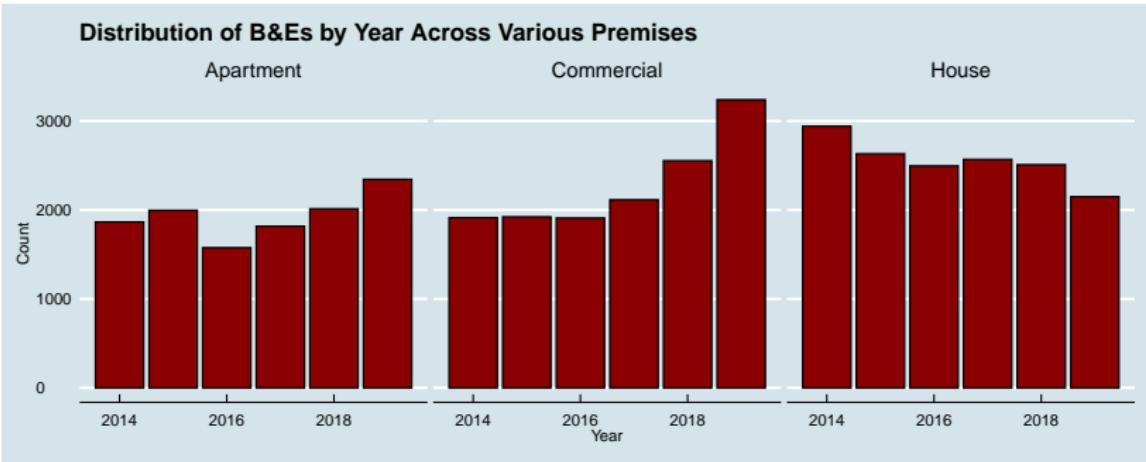


- We can see that crime has been increasing over the last few years
  - By analyzing the data, we hope to find trends and effectively make recommendations to slow this increase

## Statistical Methods

- The data on B&Es before 2014 is sparse and inconsistent; we filtered out 87 observations that occurred from 2013 and earlier
- Leap years (i.e. years with 366 days) had their last day omitted in order to ensure all years had equal length
- “Other” and “Outside” premisetype data were removed from certain tests in order to focus on Commercial/Residential analysis
- Linear regression modeling, hypothesis tests, barplots, histograms and scatterplots were used to examine the following:
  - The average distribution of B&Es throughout the year
  - Days of the week with the highest/lowest B&E frequency
  - Increases in overall B&Es over the last years
  - Business hour to non-business hours B&E proportions by seasonality
  - Business hour to non-business hours B&E proportions by division

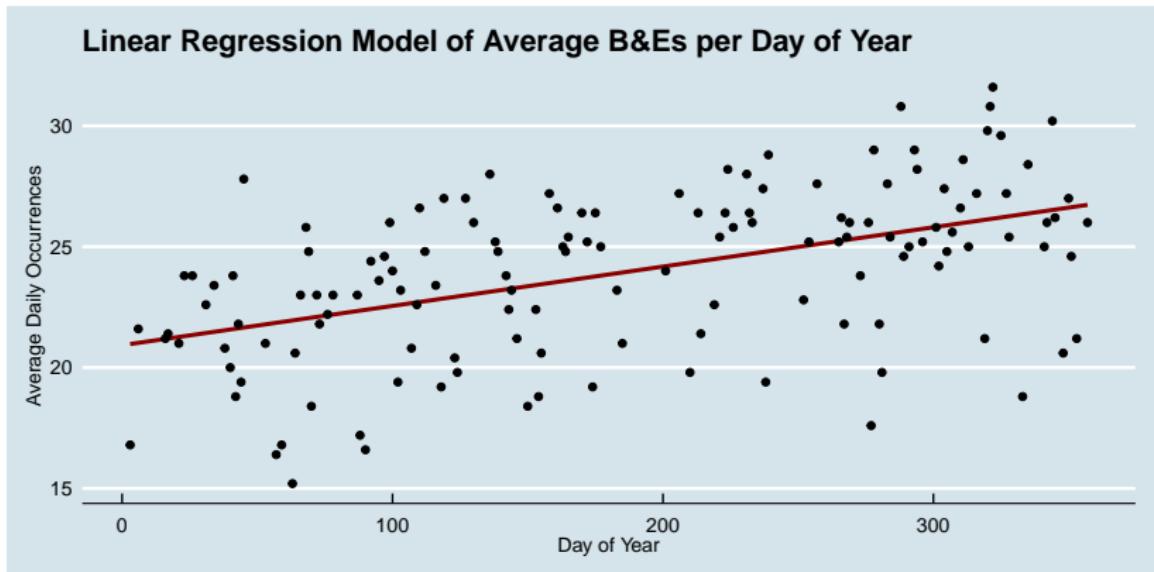
# B&Es Across Premise Type and Year



- Commercial buildings have experienced an increase in B&Es in the last 3 years
- Households have experienced a decrease in B&Es over the last years
- Apartment B&E rates have remained relatively consistent throughout the years

## How are B&Es distributed throughout the year? (Model Cont.)

- The y-axis measures variable “avgdailyoccurrences” which is defined as the average of all occurrences on this day over 5 years



## How are B&Es distributed throughout the year? (Equation)

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	20.9250509	0.5165818	40.506752	0
occurrencedayofyear	0.0162632	0.0024216	6.715975	0

- The regression equation is:  $\hat{y} = \beta_0 + \beta_1 * (\text{occurrencedayofyear})$ 
  - Where  $\hat{y}$  refers to the average daily occurrences for that particular day,  $\beta_0$  is the intercept (20.925) and  $\beta_1$  is the coefficient for the indicator variable (0.016)
- The association between Day of the Year and average number of crimes is weak ( $r^2 \approx 0.249$ ) and positively correlated
- The intercept can be interpreted as approximately 21 B&Es occurring on the first day of the year
- The slope of the line shows that for every new day of the year, there is an estimated increase of 0.016 B&Es
- In addition, the null p-value indicates that we have strong evidence that there is an association between the two variables

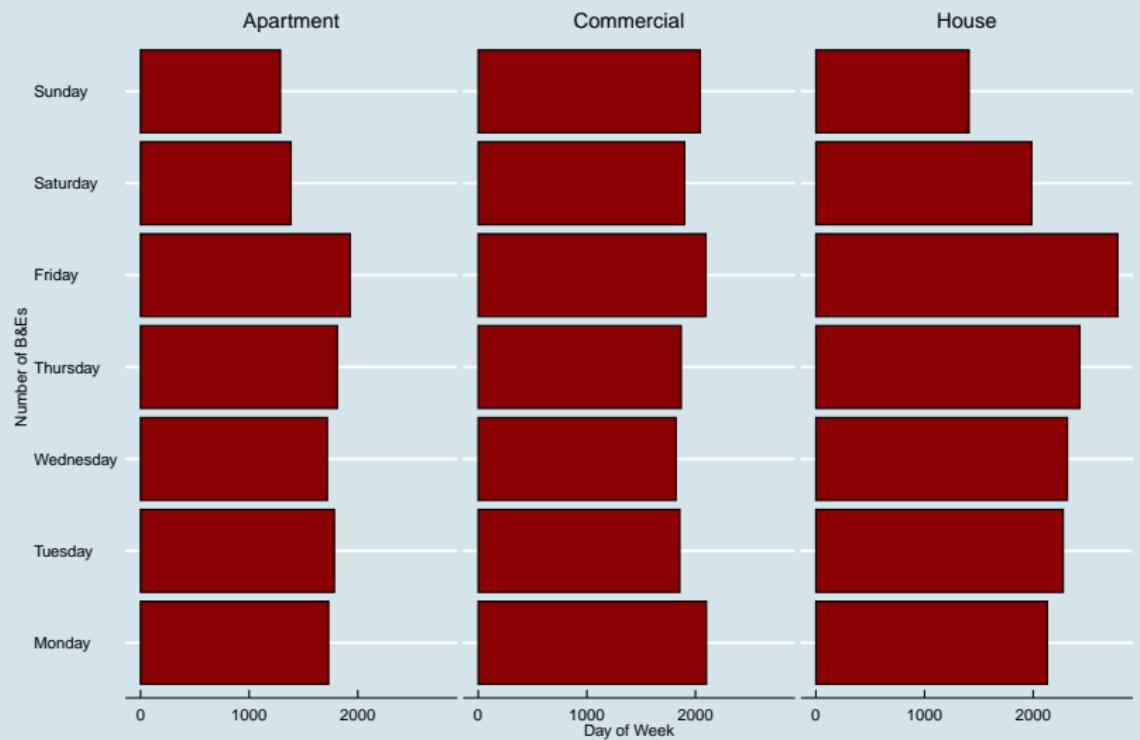
## How are B&Es distributed throughout the year? (RMSE)

```
## # A tibble: 1 x 3
##   RMSE_test RMSE_train ratio_of_RMSEs
##       <dbl>      <dbl>          <dbl>
## 1     2.87      2.95          1.03
```

- We calculated the Root Mean Squared Error (RMSE) on 80% (training data) and 20% (testing data) of our data in order to determine the accuracy of our model
  - An RMSE of  $\approx 3$  (for both the training and the testing data) indicates that the average difference between our model and the actual average occurrence number for that day is about 3 occurrences
  - A ratio of 1.0274 is very close to 1 and indicates little variation between our testing data error and training data error

# Which Day of the Week has the highest B&E Rate?

B&E Distribution Across Days of Week for Different Premises



## Which Day of the Week has the highest B&E Rate? (Cont.)

- We observed the following trends:
  - Commercial buildings experience very little variation in B&E frequency throughout the week
  - Apartments and Households experience relatively high B&E frequencies on Fridays
  - Apartments and Households experience relatively low B&E frequencies on Sundays
- New variable “Premise”
  - “Commercial” premises are defined as “Commercial” Premises
  - “Non-Commercial” premises are defined as “House” or “Apartment” premisetype

## Is there a difference in Friday B&E rates between premises?

Premise	Total	Friday	Not_Friday	Proportion
Commercial	13682	1876	11806	0.1371145
Non_Commercial	26970	4208	22762	0.1560252

- “Proportion” is defined as Friday B&Es / Total B&Es for each premise type
- We can see a difference in proportions of B&Es on Fridays between the two premises
  - A hypothesis test was conducted to determine if there was a difference between the two proportions
    - 5000 simulations were ran and a p-value of 0 was produced
    - This indicates that we have very strong evidence that there **is** a difference between Commercial and Non-Commercial Friday B&E proportions
  - Non-Commercial buildings are more at risk for B&Es on Fridays relative to Commercial buildings

## Is there a difference in Sunday B&E rates between premises?

Premise	Total	Sunday	Not_Sunday	Proportion
Commercial	13682	1748	11934	0.1277591
Non_Commercial	26970	2896	24074	0.1073786

- “Proportion” is defined as Sunday B&Es / Total B&Es for each premise type
- We can see a difference in proportions of B&Es on Sundays between the two premises
  - Another hypothesis test was conducted to determine if there is a difference between the two proportions
    - 5000 simulations were ran and a p-value of 0 was produced
    - This indicates that we have very strong evidence that there **is** a difference between Commercial and Non-Commercial Sunday B&E proportions
  - Commercial buildings are more at risk for B&Es on Sundays relative to Non-Commercial buildings

## Business Hour B&Es vs. Non-Business Hour B&Es

- New variable “time” was created
  - “Business\_Hours” was defined as 8:00am to 7:59pm
  - “Non\_Business\_Hours” was defined as 8:00pm to 7:59am
- The following summary table show the total number of B&Es, Business Hour B&Es and Non Business Hour B&Es across all data.
- “Proportion” is defined as the proportion of Business Hour B&Es [Business Hour B&Es / Total B&Es]
- The overall proportion across all the data is about 0.5193; this will act as a baseline measurement for future comparisons.
  - This shows that Business Hour B&Es are slightly more common than Non Business Hour B&Es

Total	B_Hours	NB_Hours	Proportion
43302	22487	20815	0.5193063

## How do proportions differ across seasonality?

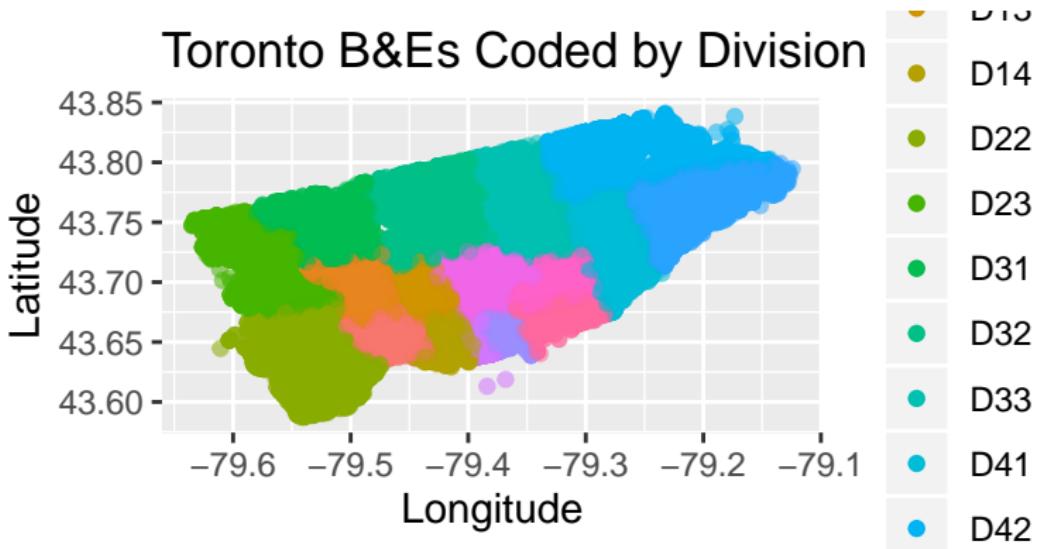
- New variable “Season” was created and months were divided accordingly

Season	Total	B_Hours	NB_Hours	Proportion
Fall	11549	6163	5386	0.5336393
Spring	10340	5376	4964	0.5199226
Summer	11088	5315	5773	0.4793470
Winter	10324	5632	4692	0.5455250

- Summer tends to experience more Non-Business Hour B&Es relative to the other seasons
- The other seasons tend experience more Business Hour B&Es than Non-Business Hour B&Es
- Fall has the largest number of B&Es across all seasons

# Toronto B&Es Business Hours and Non Business Hour Trends

- This scatterplot shows Toronto divided by division and will be referenced later to examine spatial trends across Toronto



## Business Hour vs. Non-Business Hour Risk by Division

- Proportion calculations were conducted for each division
- This table shows the first 5/17 entries for our calculated dataframe

Division	Total	B_Hours	NB_Hours	Proportion
D11	1919	916	1003	0.4773319
D12	1395	715	680	0.5125448
D13	1771	884	887	0.4991530
D14	3619	1584	2035	0.4376900
D22	2895	1452	1443	0.5015544

## Peculiar Divisions

- We have strong evidence that suggests the following divisions have disproportionate Business Hour to Non-Business Hour B&Es.
- We conducted a hypothesis test for each division and calculated a unique pvalue based on the number of observations in that division.
- By filtering out p values smaller than (0.001) we narrowed our dataset to divisions with peculiar proportions

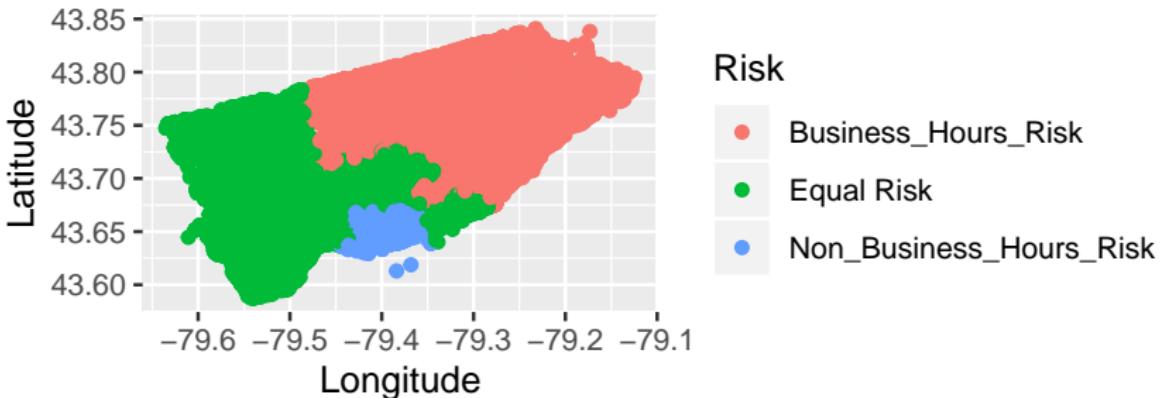
Division	pvalues	Proportion
D14	0	0.4376900
D32	0	0.5668489
D33	0	0.6780822

- These are 3/8 divisions that we found to have disproportionate B&E proportions.
- Some divisions have higher Business Hour risk and some have higher Non-Business Hour risk.

# Toronto divided by Business Hour Risk

- We can plot the districts with disproportionate p-values and examine spatial trends

Toronto Colour-Coded by Risk



- “Equal Risk” refers to divisions without peculiar proportions
- Risk areas are grouped together spatially

## Business Hour vs. Non Business Hour Trends Across Premise Type

- We looked to see how proportion varies across premise type
- New variable “Premise” used once again

Premise	Total	B_Hours	NB_Hours	Proportion
Commercial	13682	3551	10131	0.2595381
Non_Commercial	26970	17752	9218	0.6582128

- We can see a stark difference between the two proportions.
- A hypothesis test was conducted to determine if there was a difference between the two proportions
  - 5000 simulations were ran
  - The p value calculated for **this randomization test** was 0.0
  - We have strong evidence that suggests that there **is** a difference between the two proportions

# Summary of Results

- **Temporal Trends**
  - B&Es have been increasing overall, but mainly for Commercial premises
  - B&Es tend to increase as the year progresses
  - Commercial premises are at most risk on Sundays while Non-Commercial Premises are at most risk on Fridays
  - Summer has the highest rate of Non-Business Hour B&Es
  - All other seasons have higher Business Hour B&E proportions
- **Spatial Trends**
  - We can see that the North-Eastern Section of Toronto is subject to more Business Hour B&Es than Non-Business Hour B&Es
  - We can see that an area of 3 districts near the waterfront is more subject Non-Business Hour B&Es than Business Hour B&Es
  - Risk areas are grouped together spatially as well
  - Commercial premises are at a much higher risk of Non-Business Hour B&Es
  - Non-Commercial premises are at a much higher risk of Business Hour B&Es

## Conclusion & Discussion

- The majority of trends we discovered provide evidence that B&Es tend to occur when people are absent
  - 74% of Commercial B&Es occur at Non-Business Hours (i.e. when the workers have left)
  - 66% of Non-Commercial B&Es occur during Business Hours (i.e. when people are working/not at home)
  - Commercial premises are at risk on Sundays while Non-Commercial premises have their lowest B&E rates
    - This can be attributed to the majority of people staying home on Sundays and Commercial buildings being vacated
  - The Non-Business Hour B&Es primarily occurring in the 3 districts near the Harbourfront can possibly be attributed to the Financial sector being located here
    - The Financial sector is composed of many Commercial buildings

# Recommendations

- **Recommendations for Citizens**
  - Try to prevent people knowing when your premise is vacated
    - If you leave for vacation or have a regular work schedule, try not to reveal this information to strangers
- **Recommendations for Police**
  - Seasons have differing Day/Night B&E proportions, consider readjusting resources to accommodate for this temporal shift
  - Divisions in the North-Eastern Sector of Toronto have greater Business Hour proportions, potentially increase surveillance during the day
    - Of these Business Hour B&Es, Non-Commercial Premises are targeted most frequently
  - Divisions in the Southern sector of Toronto have greater Non-Business Hour Proportions, potentially increase surveillance during the night
    - Of these Non-Business Hour B&Es, Commercial Premises are targeted most frequently

## Limitations and Acknowledgements

- There are hundreds of variables which can influence B&E trends across the city; we were not provided with financial or demographic data
  - None of our correlations imply causality; confounding variables potentially influence the associations we saw
- Time constraints prevented further analysis on several trends (e.g. Does seasonality proportions differ across premises?)
- These trends are derived from data that is *reported*; B&Es that are misreported or not reported occur on a daily basis and make our assumptions less valid
- **We'd like to thank Toronto Police Services for the data provided and the STA130 Team for the support throughout the term**