When and Where do auto thefts occur?

Analyze of the auto theft data from Toronto Police Service

Jessica Jia-Ni Xu, Shaunee Callender, Shi Tang, Jianhong Guo, 0103, 103-1

Introduction

- **Auto thefts Data** from Toronto Police Service
- **Goal:** reduce the number of Auto Thefts
- **Focus** on the location and time
- Statistical inferences: Histograms, box plots, bar plots
- Chi-Squared tests, hypothesis tesing

Objectives

To answer the following questions:

- Do some **areas** have a higher incidence of auto thefts?
- Are most auto thefts carried out at a specific **time** of day?
- Do the **area** & the **time** of incident occurred independent of each other?

Data Summary

New Variables:

- **day_night** - Groups the hours of the day: Day 06:00-18:00 & Night 18:00-06:00.

- **daytime** - Groups the hours of the day: Morning 06:00-12:00, afternoon 12:00-18:00, evening 18:00-00:00, late night 00:00-06:00.

- **Area** - Groups Hood ID into districts: East York, Etobicoke, North York, Old Toronto, Scarborough and York.

- **crime_rate** - Crime rate for a specific area in 2016.

Data Summary (cont'd)

New Data Frames:

- **new_auto_thefts** - created from *auto_thefts* with variables related to the time and location, included year 2014-2018.

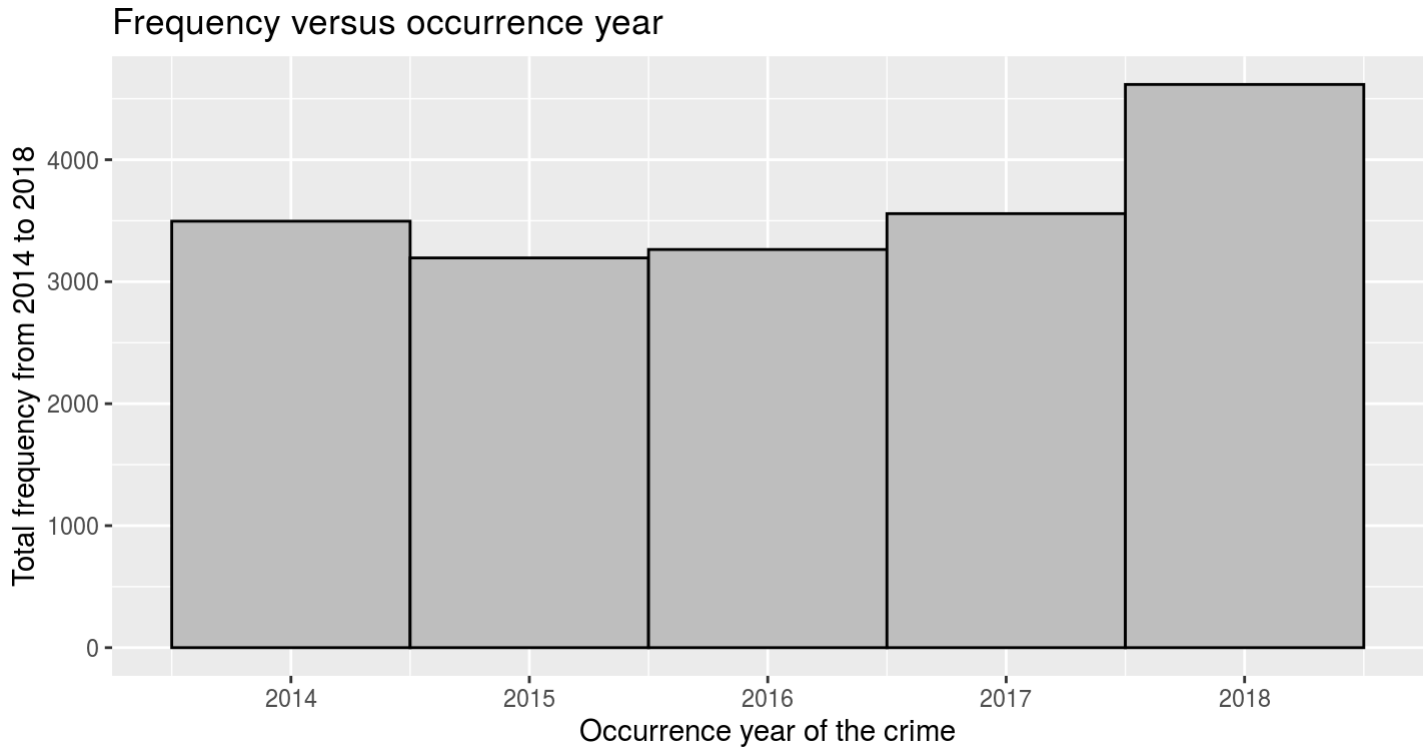
```
## Observations: 18,130
## Variables: 10
## $ Index_              <dbl> 169469, 169470, 169471, 169472, 169473, 1694…
## $ occurrenceyear      <dbl> 2018, 2018, 2018, 2018, 2017, 2017, 2017, 20…
## $ occurrencemonth     <fct> August, August, August, August, June, June, …
## $ occurrenceday       <dbl> 24, 26, 27, 24, 13, 14, 14, 15, 15, 14, 17, …
## $ occurrencedayofyear <dbl> 236, 238, 239, 236, 164, 165, 165, 166, 166,…
## $ occurrencedayofweek <fct> Friday, Sunday, Monday, Friday, Tuesday, Wed…
## $ occurrencehour      <dbl> 3, 22, 5, 18, 22, 21, 22, 23, 15, 14, 0, 12,…
## $ Hood_ID             <dbl> 130, 131, 131, 126, 28, 15, 1, 130, 47, 138,…
## $ Neighbourhood       <chr> "Milliken (130)", "Rouge (131)", "Rouge (131…
## $ premisetype         <chr> "House", "Outside", "Other", "Outside", "Hou…
```

Statistical Methods

- **Hypothesis test** Relationship between auto thefts occur in day and night
- **Chi squared test** Independence between the occurrence time and location & Independence between occurrence time and premise type.
- **Bootstrap sampling** Confidence interval of proportion of auto thefts occur in day time

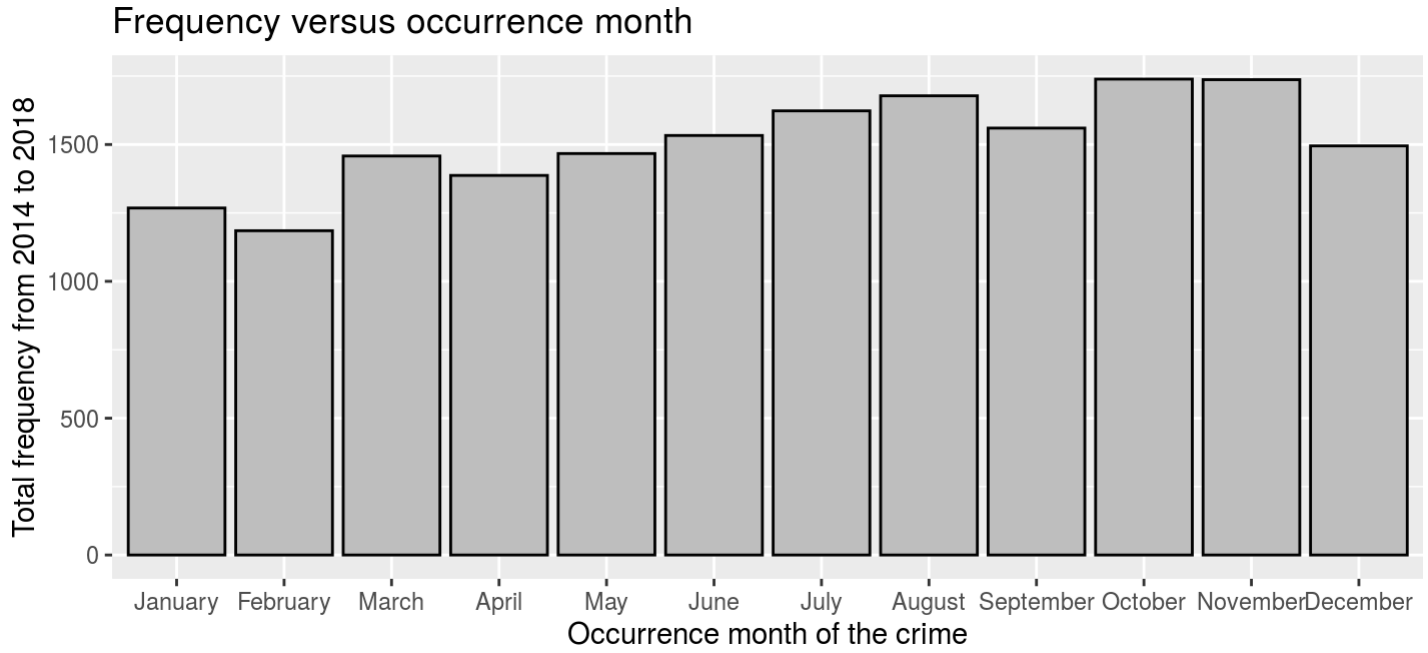
Results

Frequency versus years (2014-2018)



- Auto thefts remained fairly constant from 2014 - 2017, with an increase in 2018.

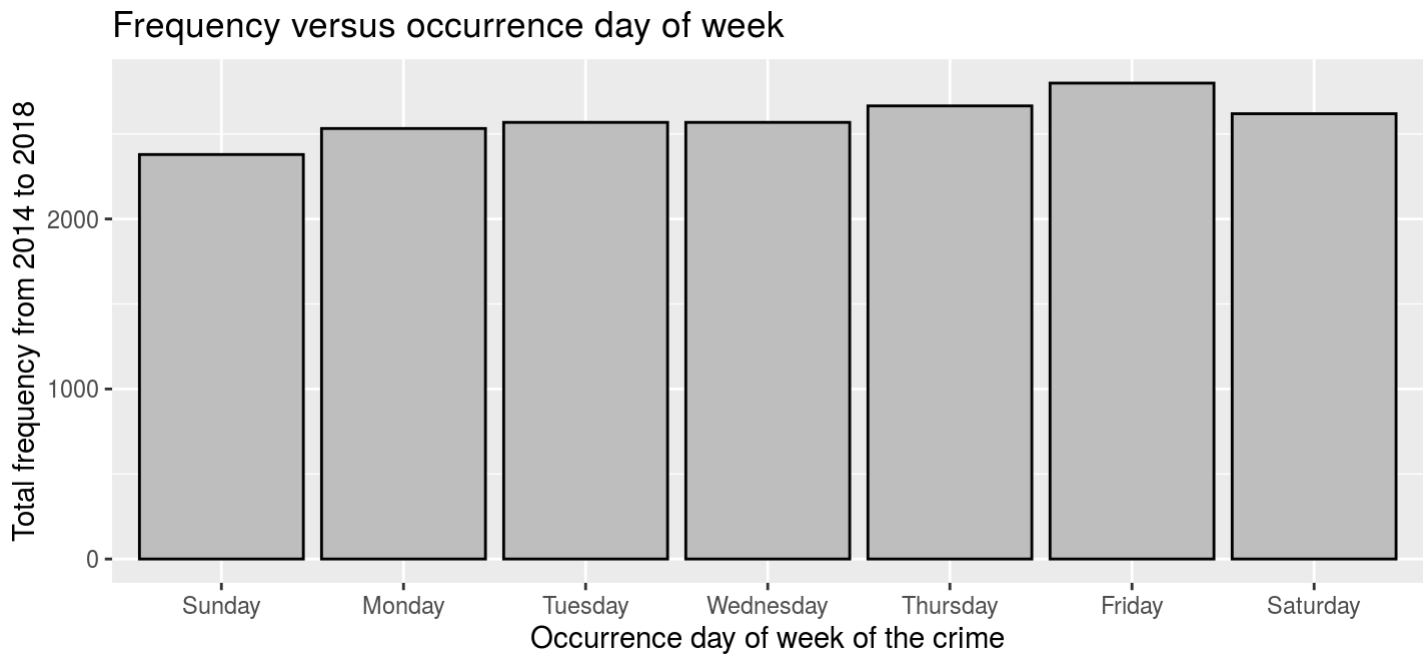
Frenquency versus months



- No obvious trends between occurrence month and number of auto thefts crimes.

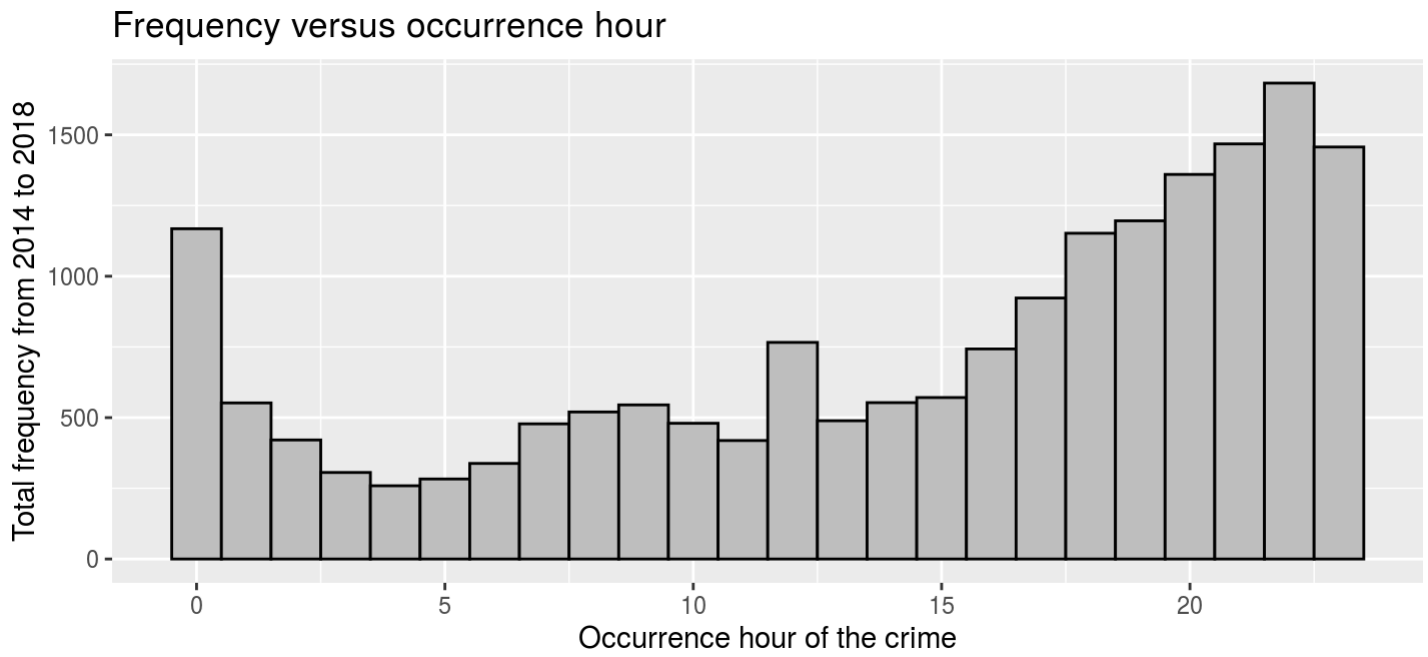
- The months with the least auto theft crimes are January and February.

Frenquency versus day of week



-The number of auto theft crimes remains fairly consistent thoughout the week.

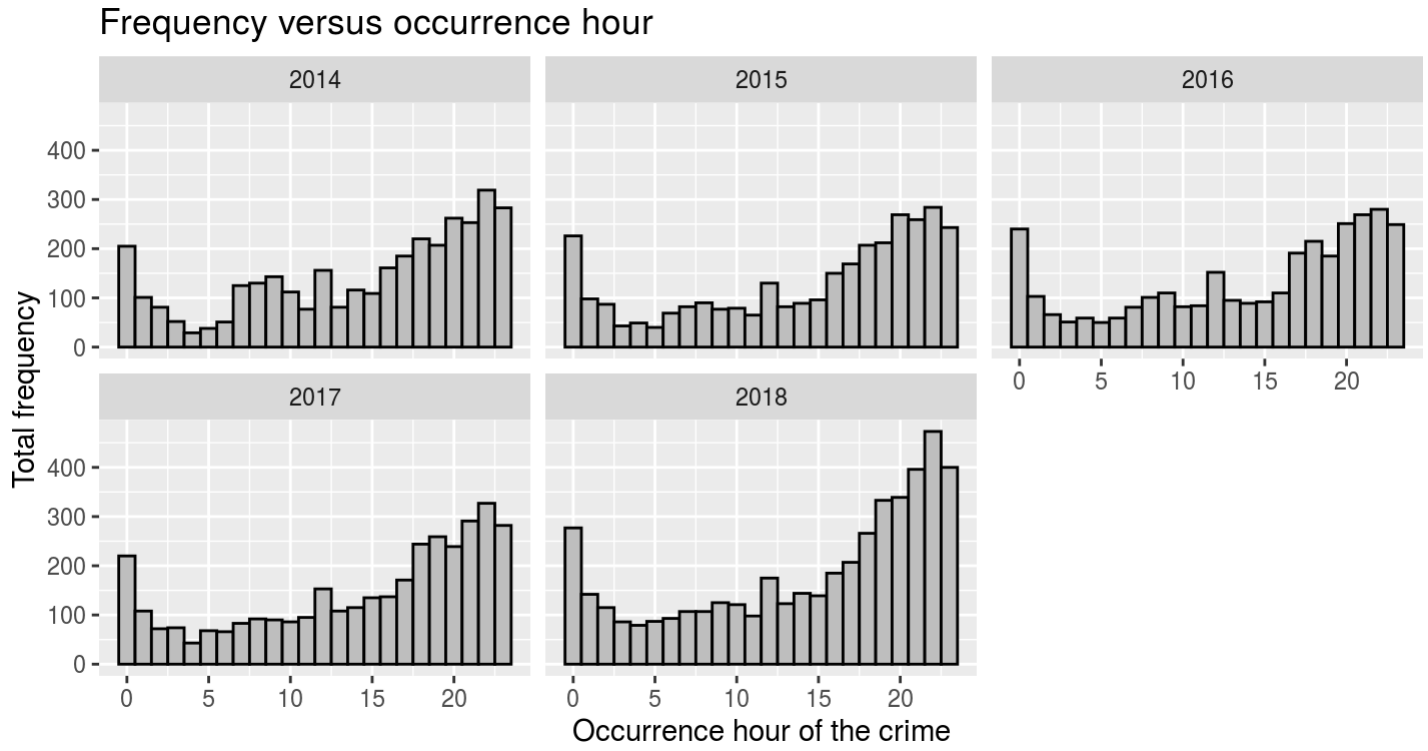
Frequency versus occurrence hour



- The frequency of auto thefts begins to steadily increase after 16 o'clock.

- There is a mode around 23 which corresponds to 11 o'clock in the evening.

Frequency versus occurrence hour (2014 to 2018)



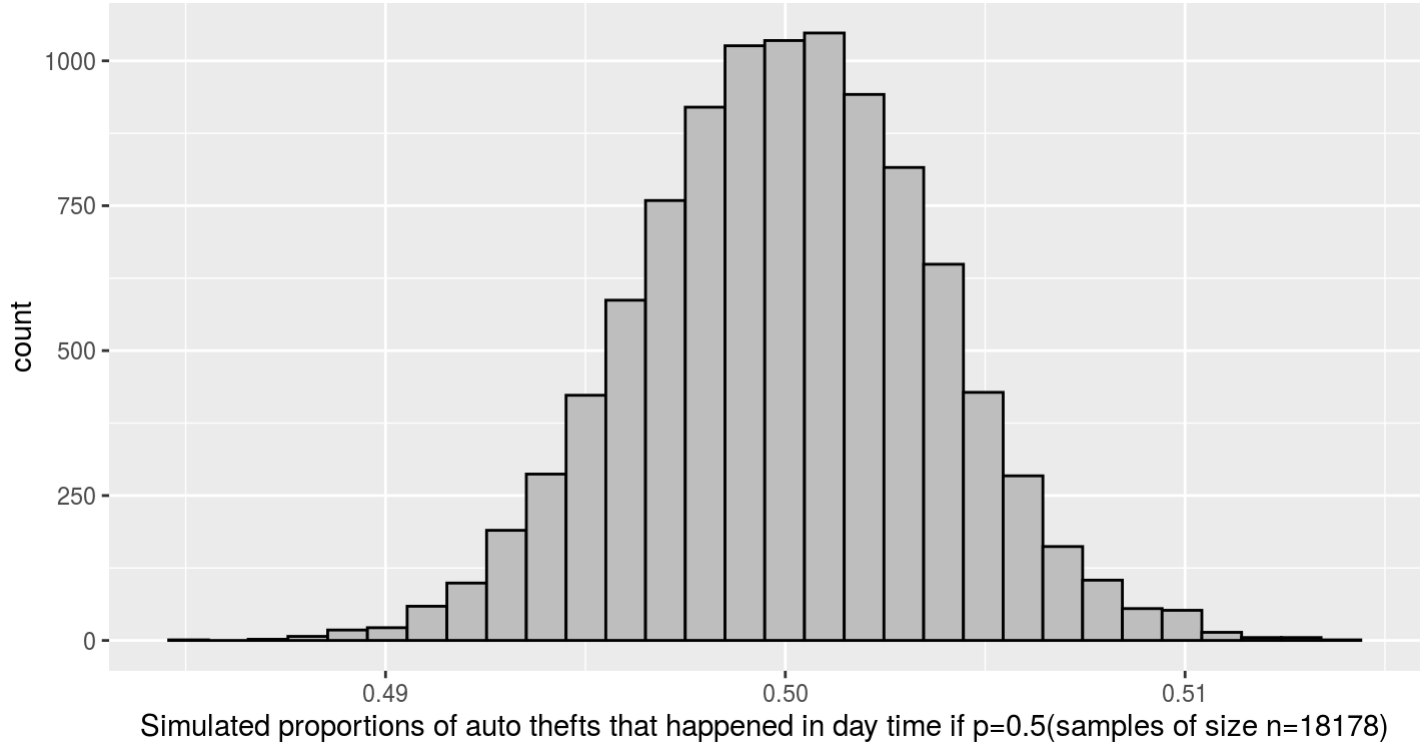
- The occurrence hour of auto theft crimes tends to remains fairly consistent throughout the years.

Summary table (proportion of auto thefts in day and night)

```
## # A tibble: 2 x 3
##   daytime     n Proportion
##   <chr>   <int>      <dbl>
## 1 Day      6825      0.376
## 2 Night   11305      0.624
```

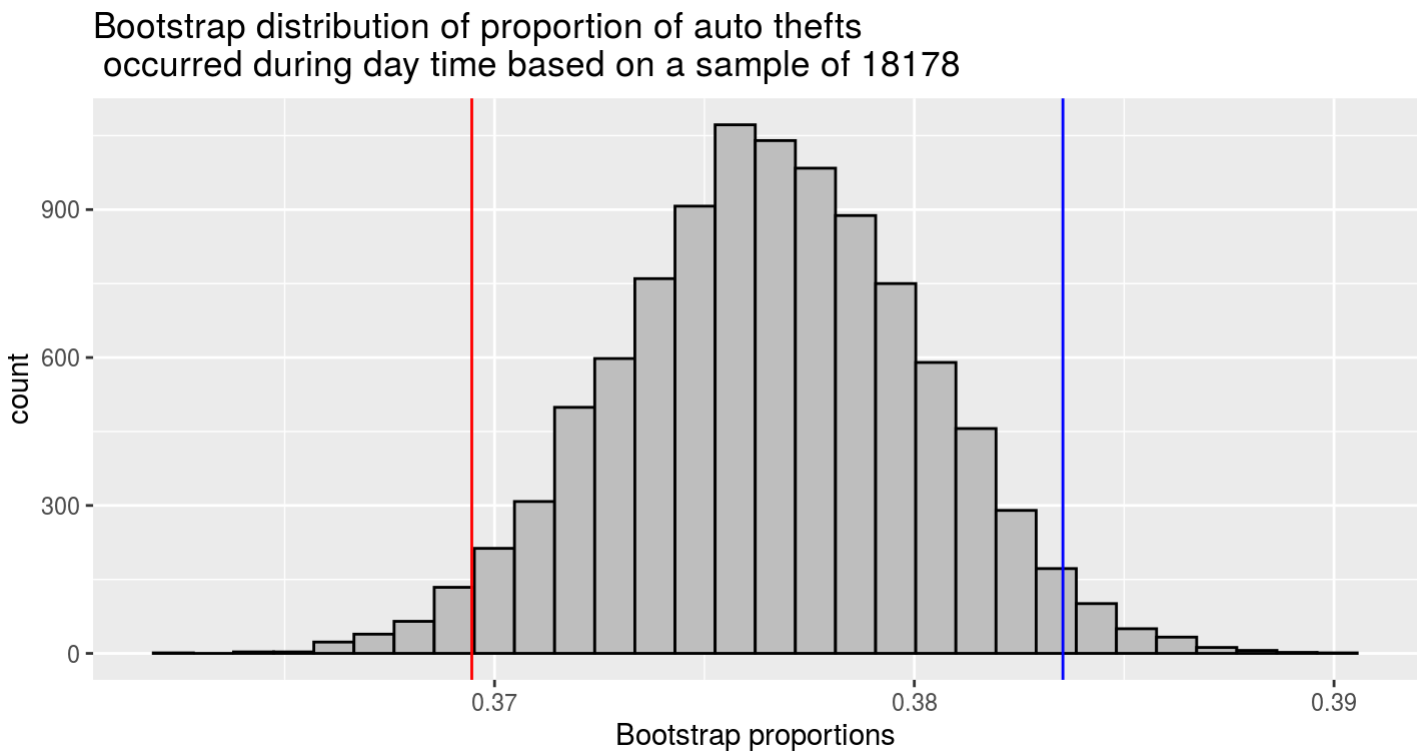
Hypothesis test

H_0 = 0.5 (assume that the proportion that auto thefts happen in day time is 0.5)



- P_value is extremely small, very close to 0.

Bootstraps and Confidence interval



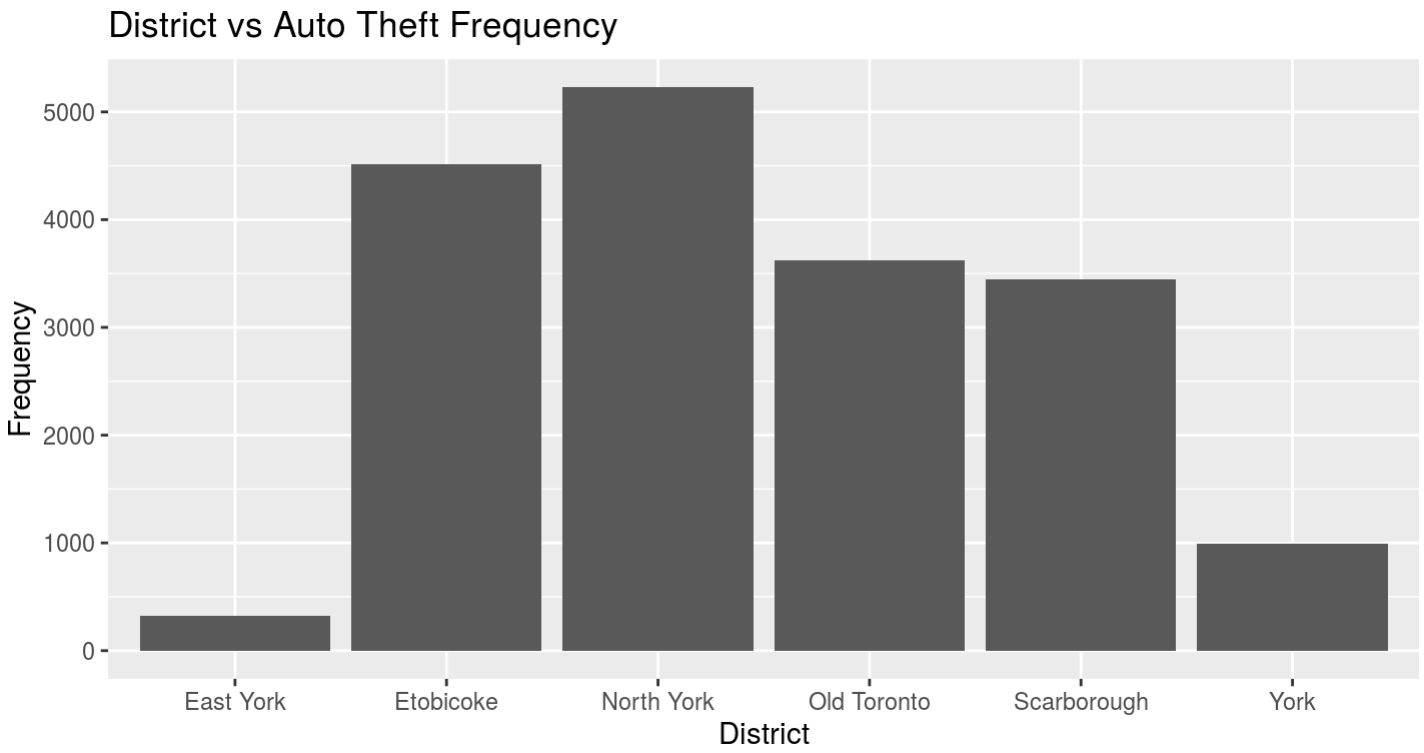

```
##      2.5%     97.5%
## 0.3694576 0.3835405
```

Location versus auto theft crimes

```
## # A tibble: 10 x 3
##    Hood_ID number_of_crimes proportion
##      <dbl>            <int>      <dbl>
##  1       1             1679     0.0905
##  2      14              565     0.0304
##  3      27              488     0.0263
##  4      26              449     0.0242
##  5      21              384     0.0207
##  6     119              305     0.0164
##  7       2              292     0.0157
##  8      31              267     0.0144
##  9     131              260     0.0140
## 10      24              258     0.0139
```

- The highest number of crimes occured in Hood_ID 1 and 14.

District and auto theft crimes.



- North York had the most auto theft crimes, while East York had the least.

District and Crime Rate

Crime Rate for 2016

Area	n_2016	population_2016	crime_rate
Etobicoke	857	365143	234.70
North York	992	672955	147.41
York	185	139238	132.87
Scarborough	557	632098	88.12
Old Toronto	624	804066	77.61
East York	49	118071	41.50

Chi-squared test for time of day and district

```
##
##                Afternoon Late night Morning Night
##   East York           73         43      46   165
##   Etobicoke          975        791     710  2041
##   North York        1120        845     825  2437
##   Old Toronto        874        542     527  1677
##   Scarborough        817        569     548  1514
##   York               186        199     124   482
```

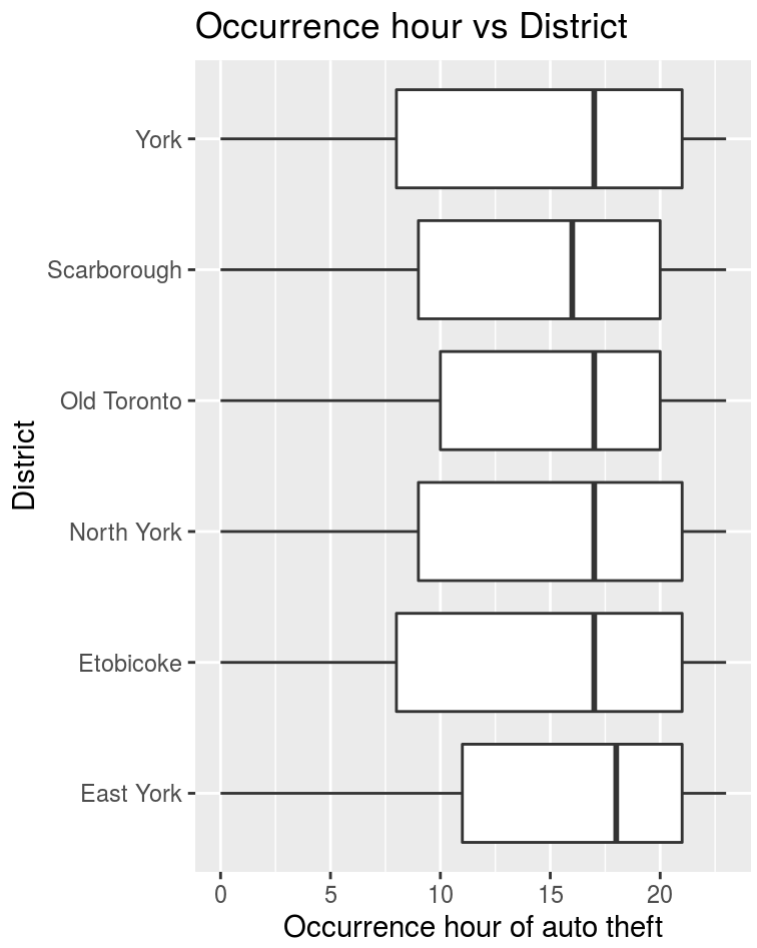
```
##
##  Pearson's Chi-squared test
##
## data:  areatimetable
## X-squared = 51.145, df = 15, p-value = 7.815e-06
```

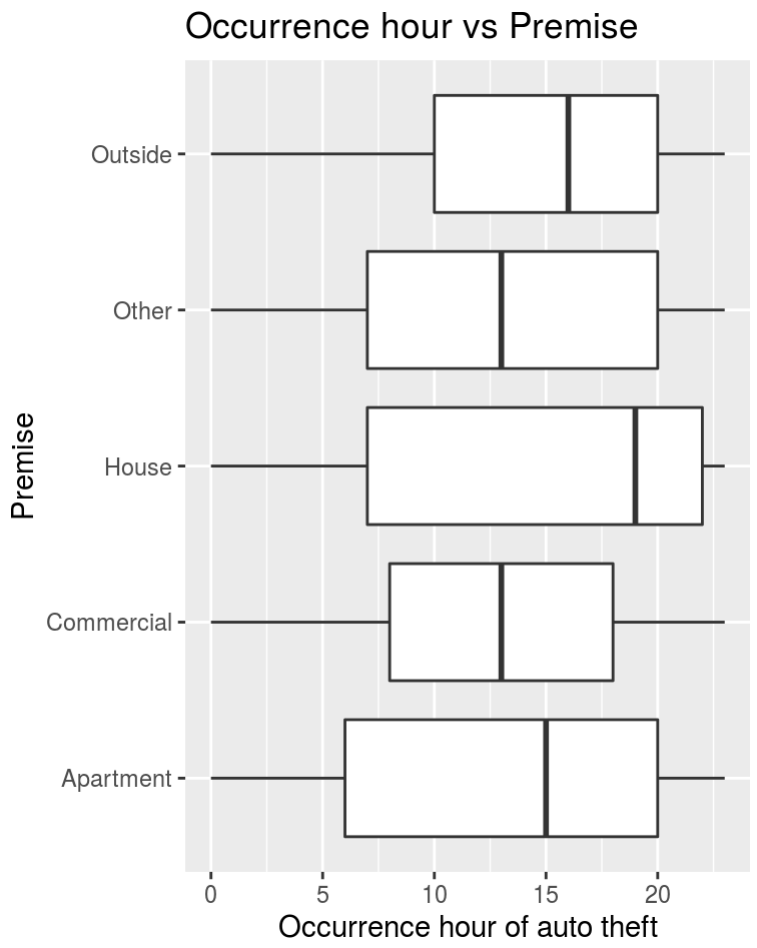
Chi-sqaured test for time of day and premise type

```
##
##              Afternoon Late night Morning Night
##   Apartment         83         97      62   151
##   Commercial       566        361     353   523
##   House            474       1078     467  2955
##   Other             90         64     149   173
##   Outside         2832       1389    1749  4514
```

```
##
##  Pearson's Chi-squared test
##
## data:  premisetimetable
## X-squared = 1287.9, df = 12, p-value < 2.2e-16
```

Distribution across location





Conclusion

- Certain areas have a higher incidence of auto thefts which is shown by the varying crime rates.
- Most auto thefts tend to be carried out at night, more specifically between evening and late night. This was confirmed by bootstrap testing and hypothesis testing.
- There is strong evidence to suggest the time of day of the incident and the location of the incident are interdependent.

Recommendations

- The Toronto Police Service should send more patrols to areas with a higher crime rate.
- Officers should be put on patrol at night because that is when more auto thefts tend to be carried out.

Limitations

- As the seasons change the sun sets at different times and this may have an impact on auto thefts.
- The location data had to be generalised and this resulted in large variations in the size of areas, which affected our analysis.
- The type of car stolen was not given. Newer cars have are built differently and could possibly be easier to break into.