Combating Auto Thefts

Investigating attritubes of areas prone to auto thefts

Shiyuan Zhou, Xinpeng Shan, Xinyi Chen, Tala ELkukhun, TUT0206, Group3

Introduction

- Given two data sets on auto thefts in numerous neighbourhoods in Toronto for over 11,000 vehicles between 2014 and 2018
- Thefts of vehicles are a major concern
- Hoping to analyze the public data critically and identify both trends and anomalies
- Suggest functional ways on how to prevent auto theft

Objectives

- Is there a significant difference between the number of auto theft occurrences for each month from 2014-2018?
- What characteristics do the top 20 neighbourhoods with the highest number of auto theft occurrences have in common?

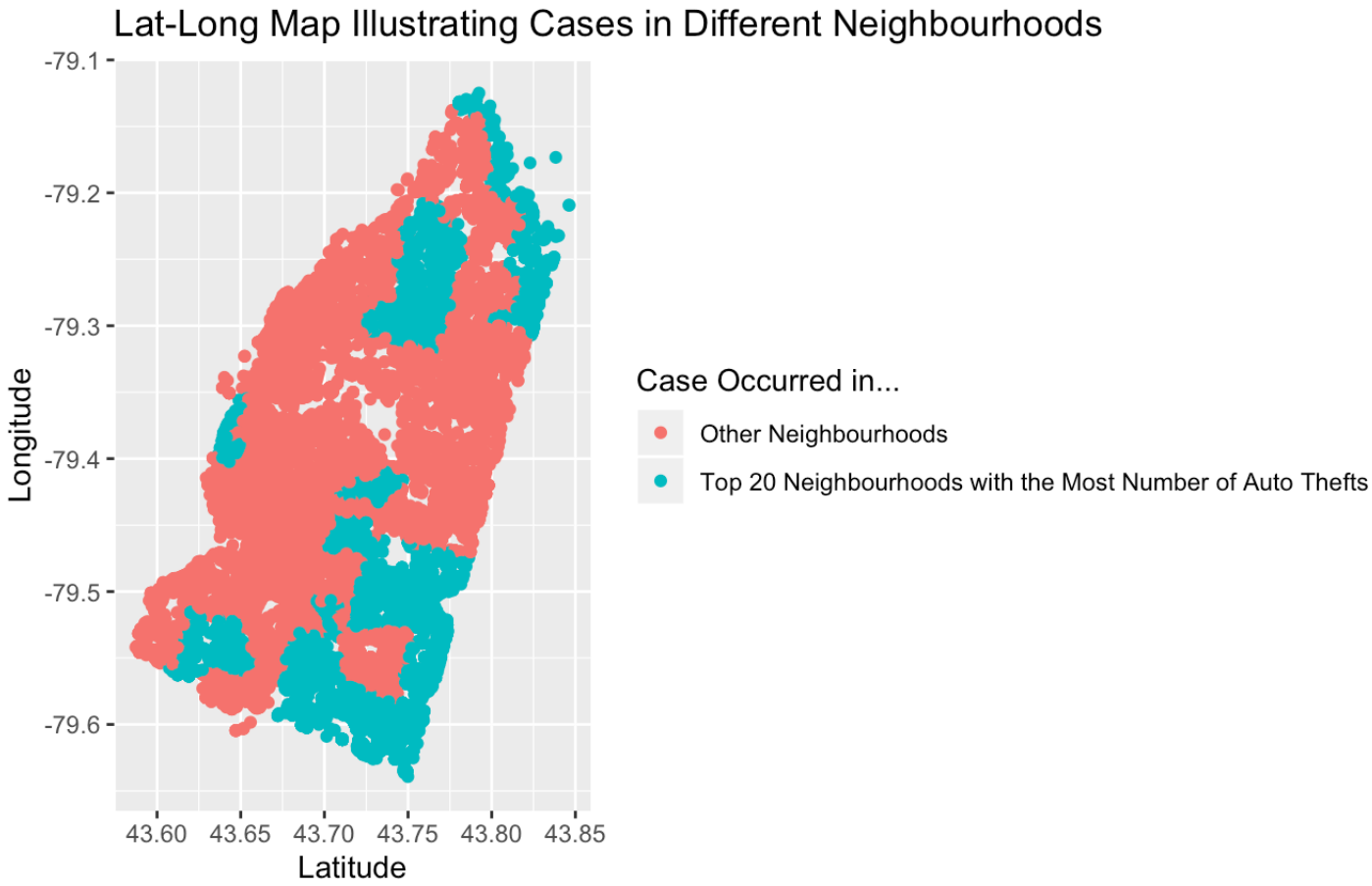
Data Summary (Data Wrangling)

- Joined the auto_thefts dataset, the neighbourhood_profiles_2016 dataset, and the table we created with the number of occurrences in each neighbourhood
- Imported another public dataset from TPS that has information about the number of cases for all crime types in neighbourhoods
- Joined the imported dataset with our existing joint data set

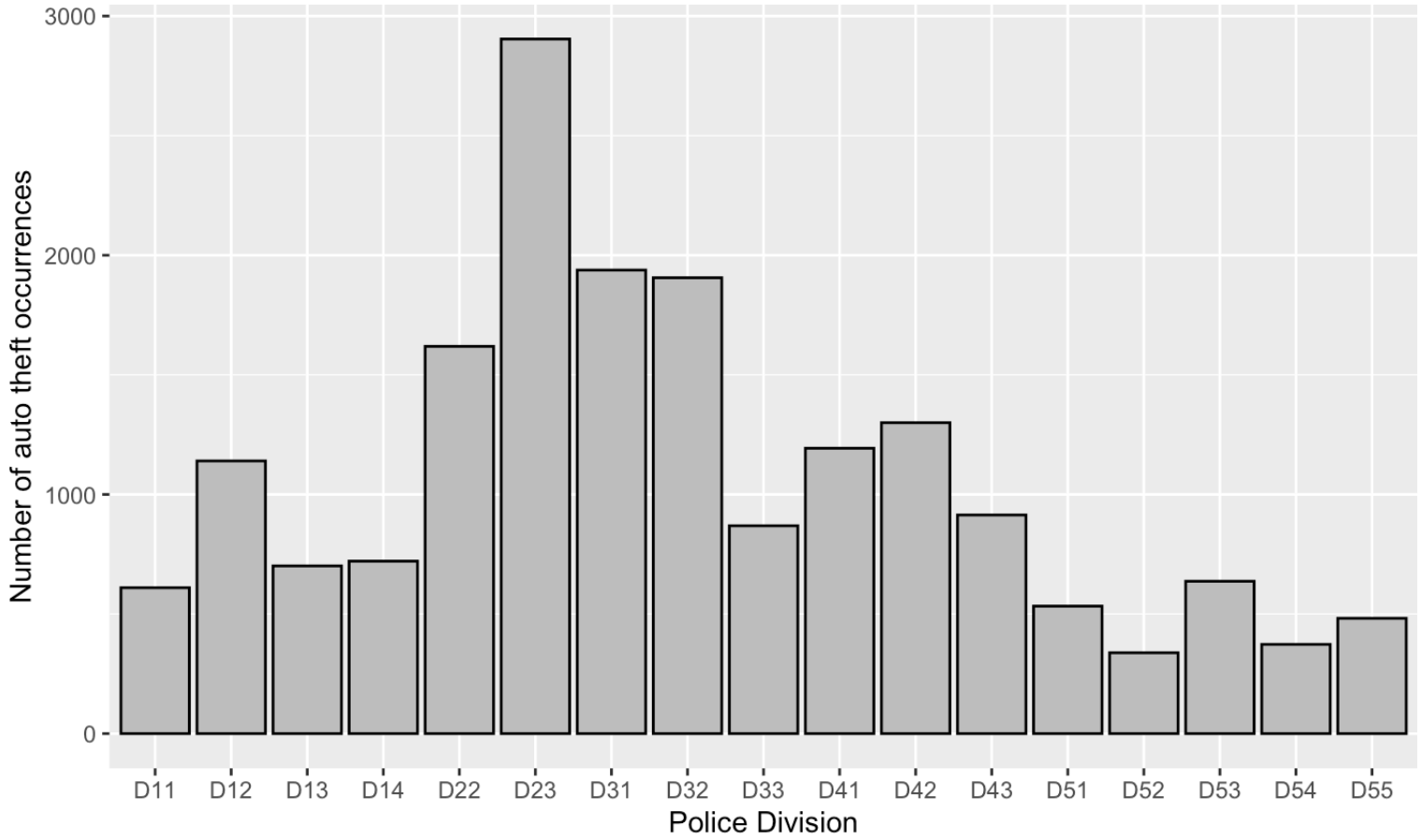
We created three new variables:

- number of auto theft occurrences per neighbourhood
- occurrence category for the number of occurrences (ex. high)
- total number of other theft cases over 5 years

What characteristics do the top 20 theft prone neighbourhoods have?



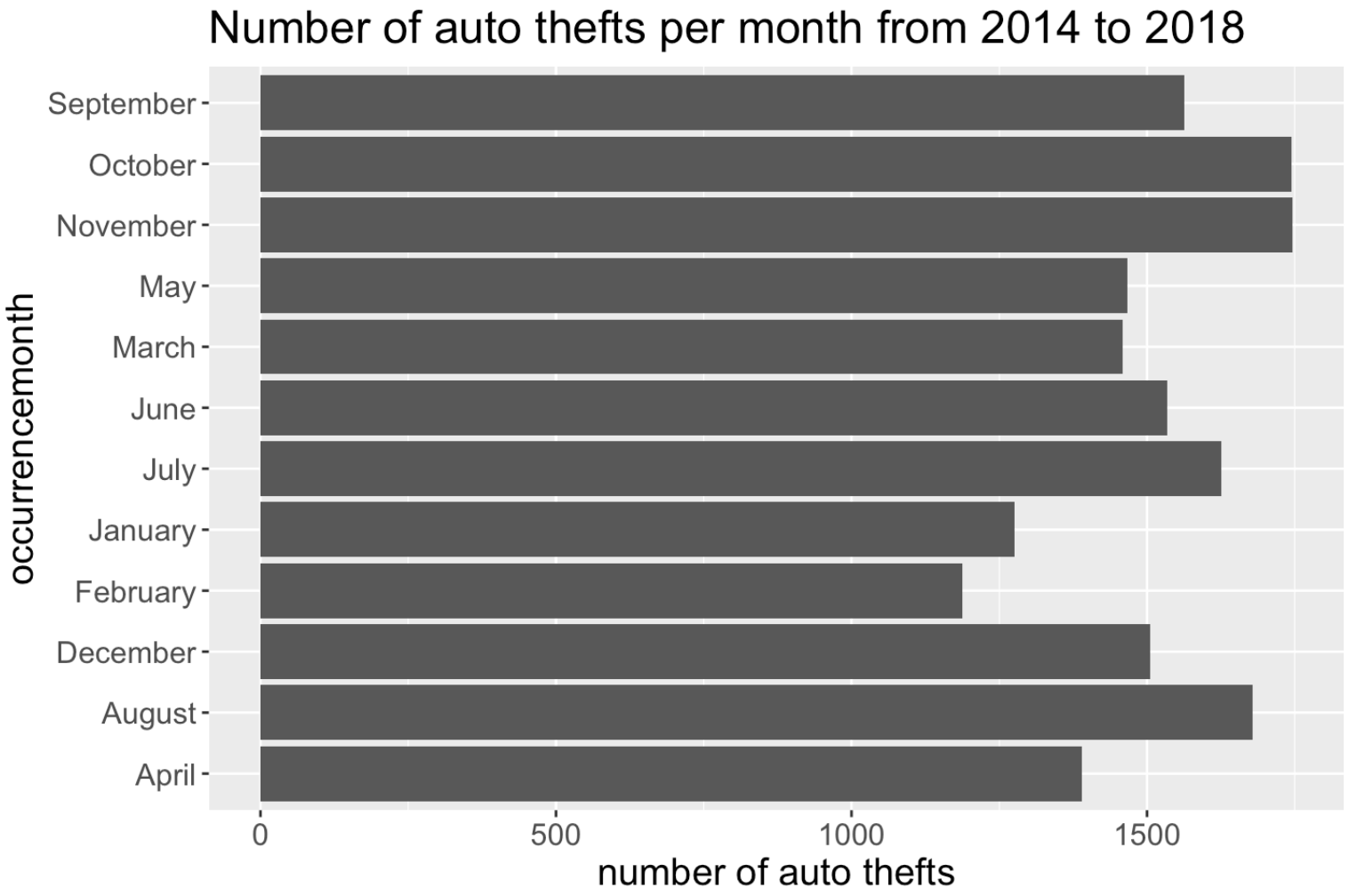
Relationship between police division and auto theft occurrences



When does autotheft happen most frequently?

- Summary table showing number of autotheft per month from 2014 to 2018

```
## # A tibble: 12 x 2
##    occurrencemonth number_autotheft
##    <chr>                      <int>
##  1 November                    1746
##  2 October                     1745
##  3 August                      1679
##  4 July                        1625
##  5 September                   1563
##  6 June                        1534
##  7 December                    1506
##  8 May                         1467
##  9 March                       1458
## 10 April                       1389
## 11 January                     1275
## 12 February                    1188
```

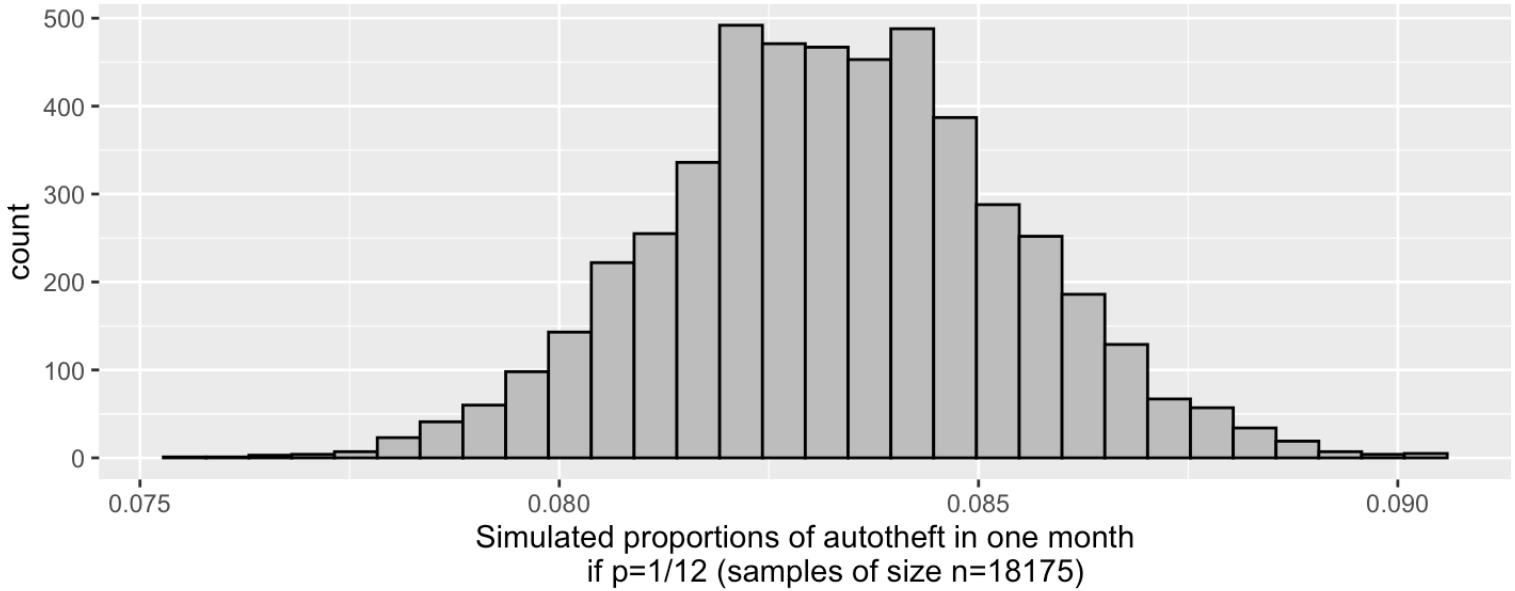

Number of auto thefts per month from 2014 to 2018
September
October
November
May
March
June
July
January
February
December
August
April
occurrencemonth
0
500
1000
1500
number of auto thefts

Statistical Methods

- Ran a hypothesis test to determine if there is a significant difference between the number of occurrences of each month over all five years
- Created a linear regression between pop density per square km. Produced the R^2 value that provides information on how well our model captured the variability in the data

Hypothesis Testing

- Null hypothesis(H_0):P_m = 1/12
- Alternative hypothesis(H_A):$P_m \neq$ 1/12 where P_m is the proportion of crime occurence in each month



p-value of 12 months ranked according to amount of auto theft ocurrence form highest month to lowest

```
## [1] 0
```

```
## [1] 0
```

```
## [1] 0
```

```
## [1] 0.0038
```

```
## [1] 0.2038
```

```
## [1] 0.6104
```

```
## [1] 0.8134
```

```
## [1] 0.2072
```

```
## [1] 0.133
```

```
## [1] 0.001
```

```
## [1] 0
```

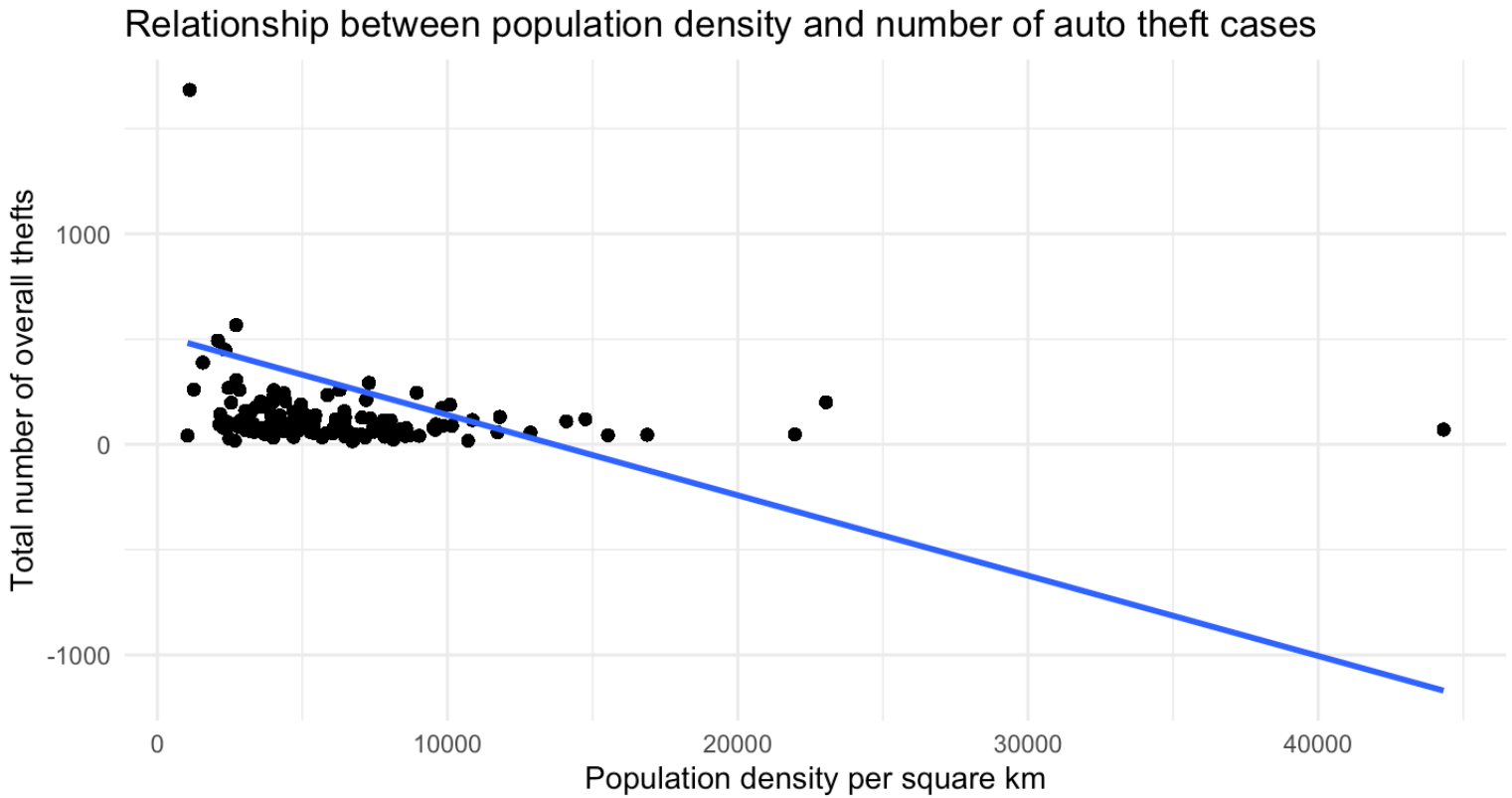
```
## [1] 0
```

Linear Regression between Population Density and Number of Auto Thefts

Estimating the Coeffcients of the Model

```
##                                Estimate   Std. Error   t value Pr(>|t|)
## (Intercept)                521.2244590 4.8168107197 108.20945        0
## pop_density_per_square_km  -0.0381599 0.0007196968 -53.02219        0
```

Adding the fitted line to the plot…



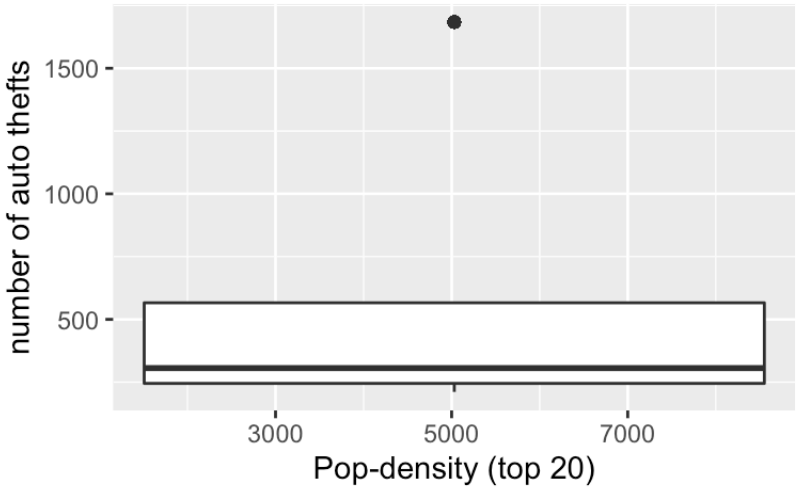
How accurate is our model? (R^2)

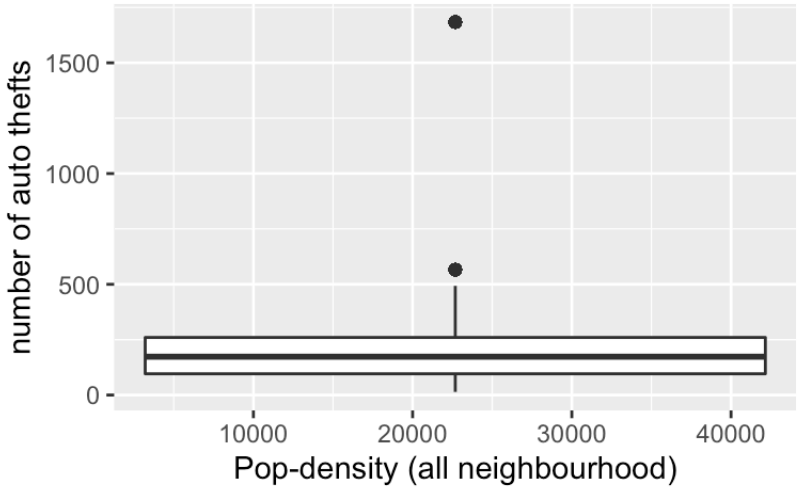
```
## [1] 0.1339546
```

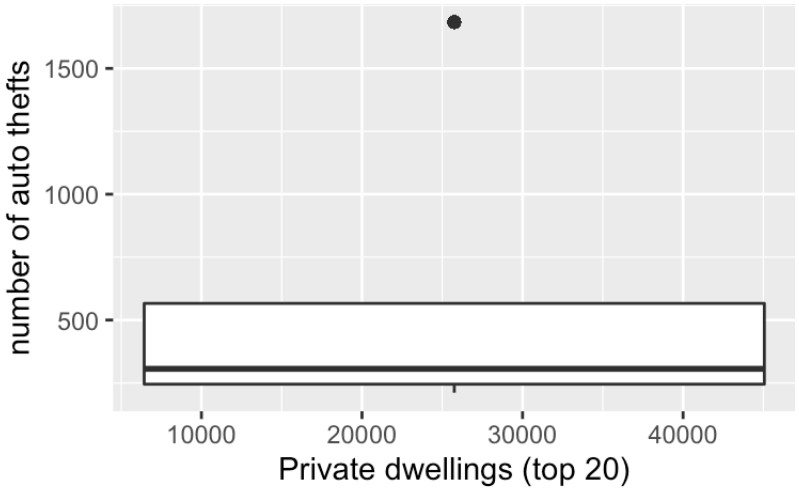
Results

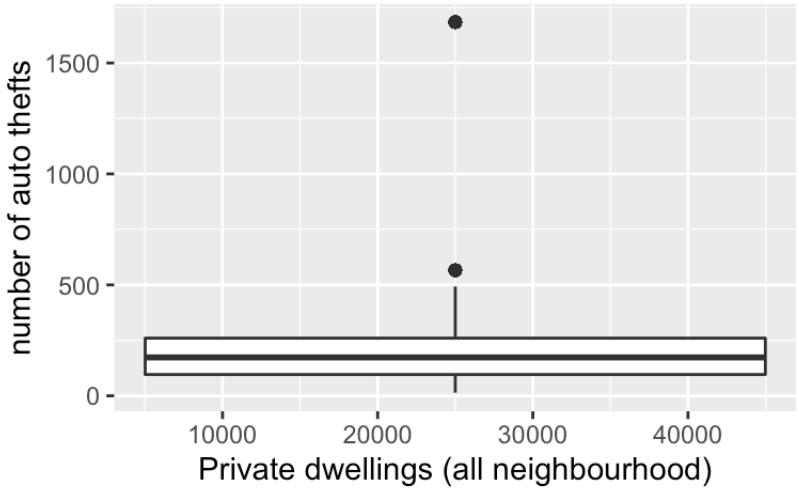
Is there a significant difference of auto thefts in each month?

```
## # A tibble: 5 x 2
##   Conclusion                                      Months
##   <chr>                                           <chr>
## 1 No evidence against null hypothesis             3, 9, 5, 6, 12
## 2 Weak evidence against null hypothesis           <NA>
## 3 Moderate evidence against null hypothesis       <NA>
## 4 Strong evidence against null hypothesis         7
## 5 Very strong evidence against null hypothesis 1, 2, 4, 8, 10, 11
```

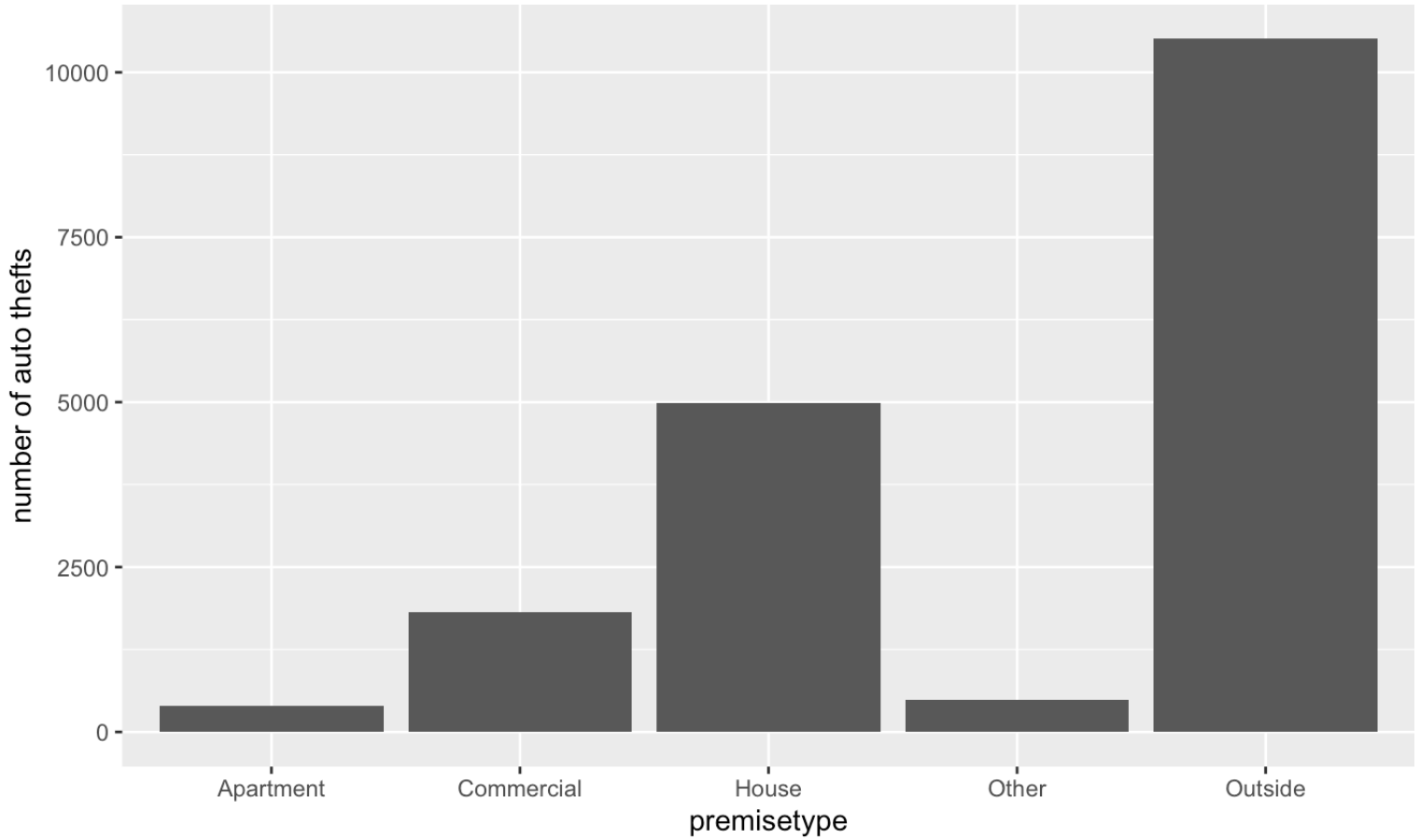

number of auto thefts
1500
1000
500
3000
5000
7000
Pop-density (top 20)


number of auto thefts
1500
1000
500
0
10000
20000
30000
40000
Pop-density (all neighbourhood)

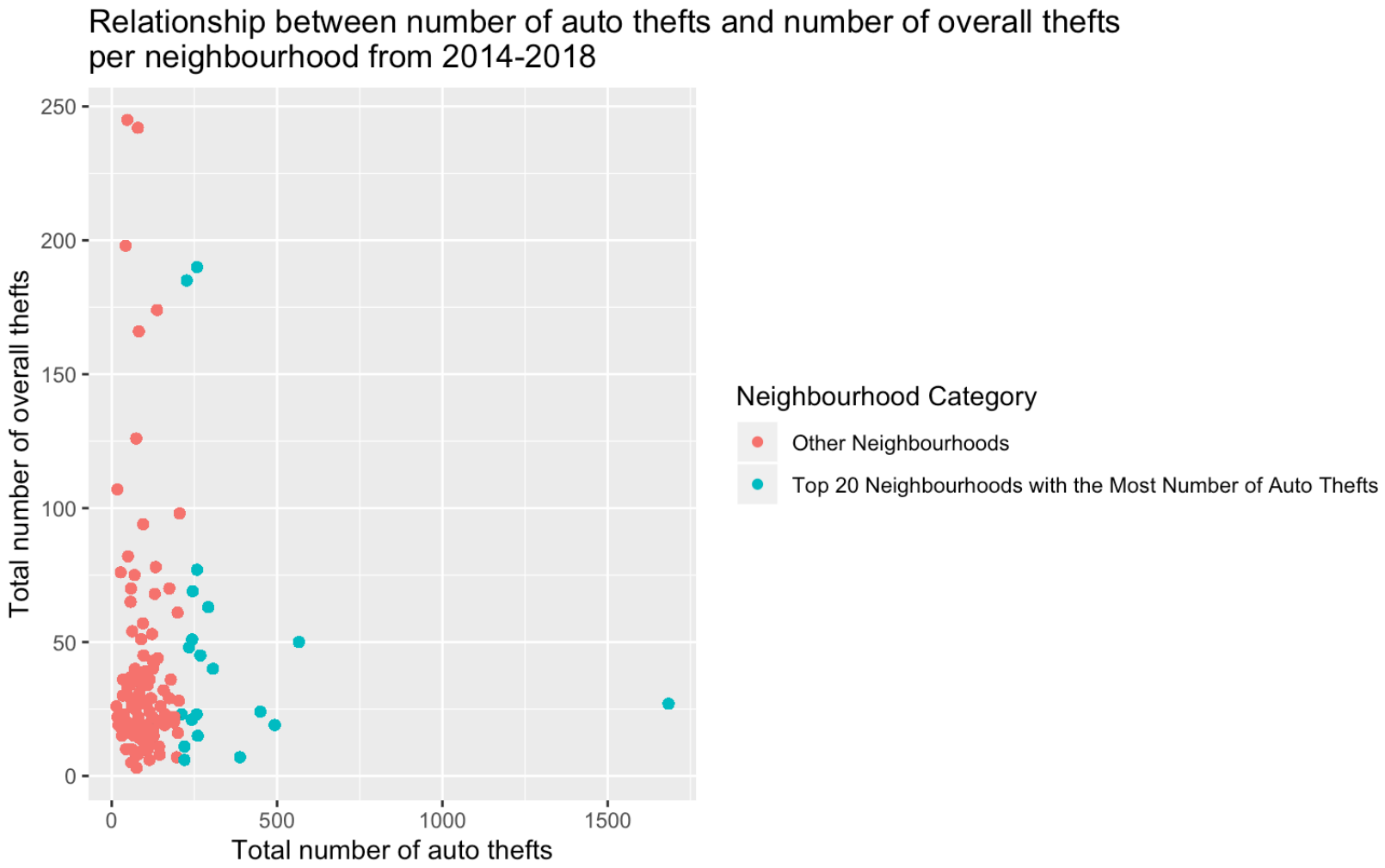

number of auto thefts
1500
1000
500
10000
20000
30000
40000
Private dwellings (top 20)


number of auto thefts
1500
1000
500
0
10000
20000
30000
40000
Private dwellings (all neighbourhood)

Which premise type has the most amount of auto thefts?



What about the other type of thefts?



Neighbourhoods with highest number of other thefts

```
## # A tibble: 6 x 3
##   total_num_other_theft Neighbourhood                    Hood_ID
##                   <dbl> <chr>                              <dbl>
## 1                   245 Ionview                              125
## 2                   242 Cliffcrest                           123
## 3                   198 Danforth East York                    59
## 4                   190 Black Creek                           24
## 5                   185 Willowridge-Martingrove-Richview       7
## 6                   174 Rexdale-Kipling                        4
```

Neighbourhoods with highest number of auto thefts

```
## # A tibble: 11 x 2
##    num_occurrences Hood_ID
##              <int>   <dbl>
##  1            1684       1
##  2             566      14
##  3             493      27
##  4             449      26
##  5             388      21
##  6             306     119
##  7             292       2
##  8             268      31
##  9             260     131
## 10             258      24
## 11             258     130
```

Conclusion

- We could **increase** police force for November, October and August and direct some of police forces on other crimes for July, April, February and January.
- We also discovered that the premise type with the most auto theft occurrences was **outside**, the TPS could establish some concentrated parking areas in order to allow for supervision for outside parks and streets.
- We thought that more populated neighbourhoods would have a higher number of occurrences but the **opposite was revealed**, so the TPS should be aware that areas with more people do not necessarily mean that they are prone to auto thefts.

- We found that the top 20 neighbourhoods with the highest number of auto thefts had a **higher average population density** and a **higher average number of private dwellings** compared to those of all neighbourhoods, the TPS could notify residence areas in these areas to be cautious and promote the usage of digital video recorder.

- **Black creek and Willowridge-Martingrove-Richview** which were ranked in the top 20 of all neighbourhoods for highest auto theft occurrences and also had a high rate of other thefts, the TPS could focus on reinforcing and educating the areas about theft in general.

- **limitation:** Most of our answers were based off of the top 20 neighbourhoods with the highest number of auto thefts. This is a limitation as we did not account for the average number of auto thefts for all years. Therefore, all of our conclusions are not always beyond the doubt.