Report Times of Toronto Car Thefts

During Work Hours Vs Outside Work Hours

Neeco Fabian, Charles Wang, Kevin Zhao, TUT 0209, 209-4

The Problem

- Auto thefts are a regular occurrence in Toronto
- Vehicles are often stolen to be sold for parts or used for other crimes within 48 hours [1]
- New technology used to protect vehicles from theft are frequently being bypassed by new methods

Objectives

- Analyze past auto theft data from the Toronto Police Service
- Find interesting relationship between variables
- Propose suggestions to help Toronto Police Service reduce car thefts after research

Introduction

Research based on 18,178 auto thefts in Toronto gathered from 2000 to 2018 reported to the Toronto Police Service.

We chose to investigate factors that may influence the time it takes for an auto theft to be reported.

Initial Observations
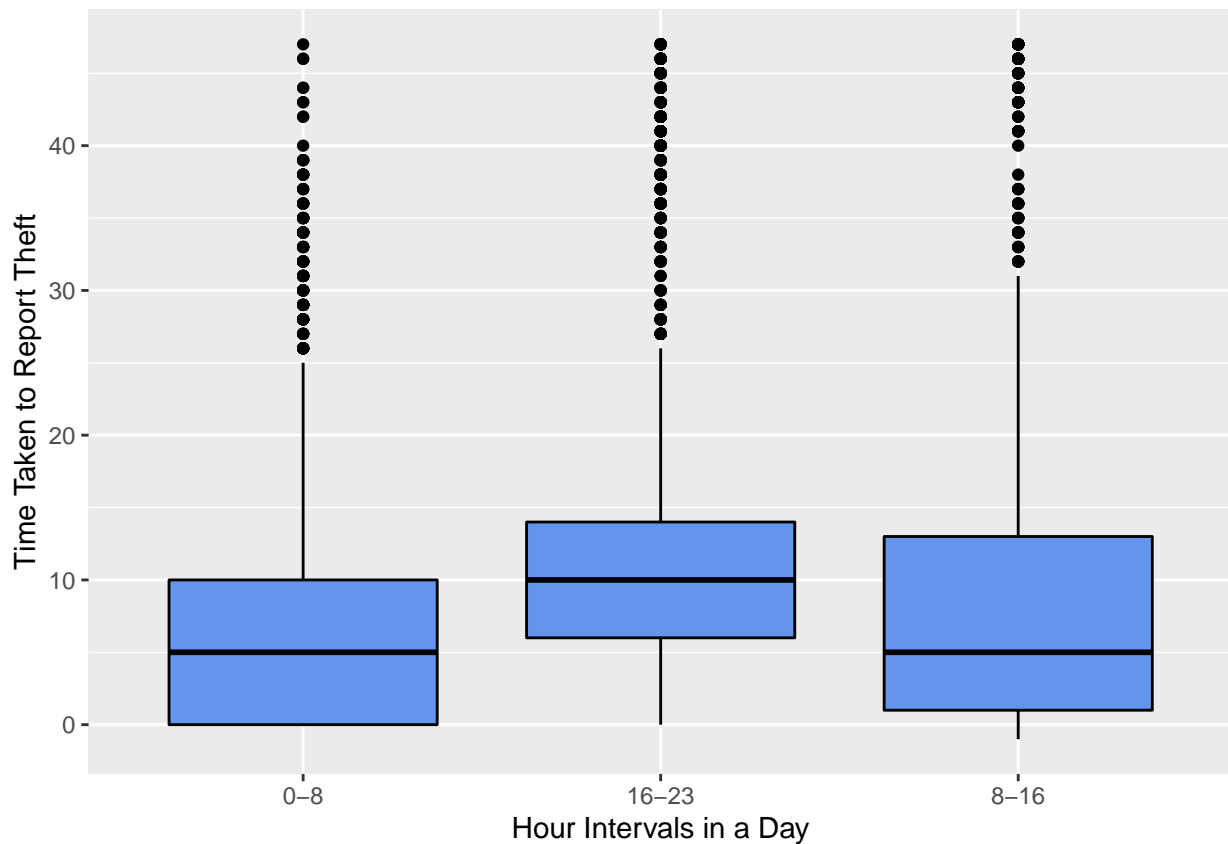
Observing if one third of the day has a longer mean report time.

The Research Question

What is the relationship between the report times of vehicles stolen during typical work hours and outside those hours?

Report Time - number of hours for an auto theft to be reported to the TPS

Working Hours - typical work hours, (9:00 - 17:00) on weekdays

Why do we want to minimize report time?

The Insurance Bureau of Canada highlights key points for car thefts[1]:

1. Car dismantling
2. Longer report times $\implies$ less likely to retrieve car
3. Stolen cars are often used for other crimes.

   These can all happen in under 48 hours after the crime.
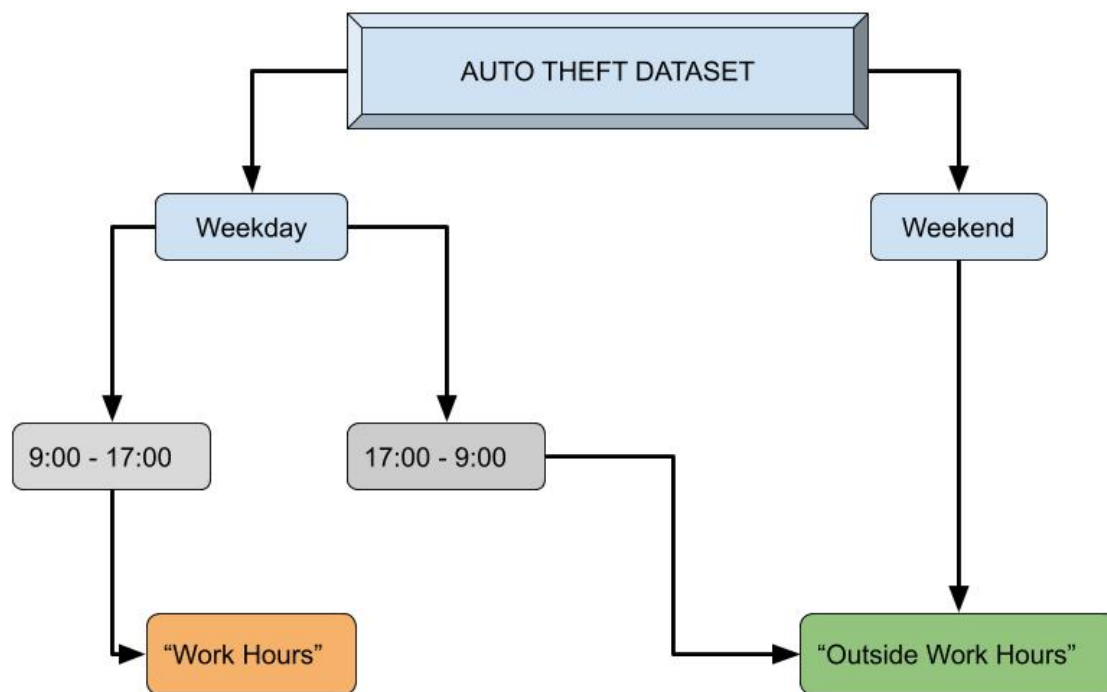   **Reduce auto thefts, reduce other crimes.**

Figure 1: Calculation of "Work Hours"

Data Wrangling I

report time = reported hour - occurred hour

- number of hours it took for each vehicle in the dataset to be reported stolen

- calculated by the difference between the hour the car was reported stolen and the hour the car was stolen.

Data Wrangling II

Split data into cars stolen during work hours or otherwise

Data Observations

```
## # A tibble: 2 x 2
##   work_category      `Mean Report Times`
##   <chr>                            <dbl>
## 1 Outside Work Hours                115.
## 2 Work Hours                        272.
```

Work Hours defined by the hours between 9-5 on weekdays, Outside Work Hours defined by hours outside 9-5 and all hours on weekends.

Statistical Methods

Step 1: Two-sided Hypothesis Test
Purpose: to analyze the difference between report times during work hours and outside work hours

Step 2: Confidence Interval
Purpose: to develop an interval at which we are 95% confident that the true difference between the mean report times during work hours and outside work hours

The Hypothesis Test

Null Hypothesis: The mean time for an auto-theft to be reported in Toronto is **equal** if the theft occurs during 9am-5pm (inclusive) or if it occurs outside 9am-5pm (exclusive).
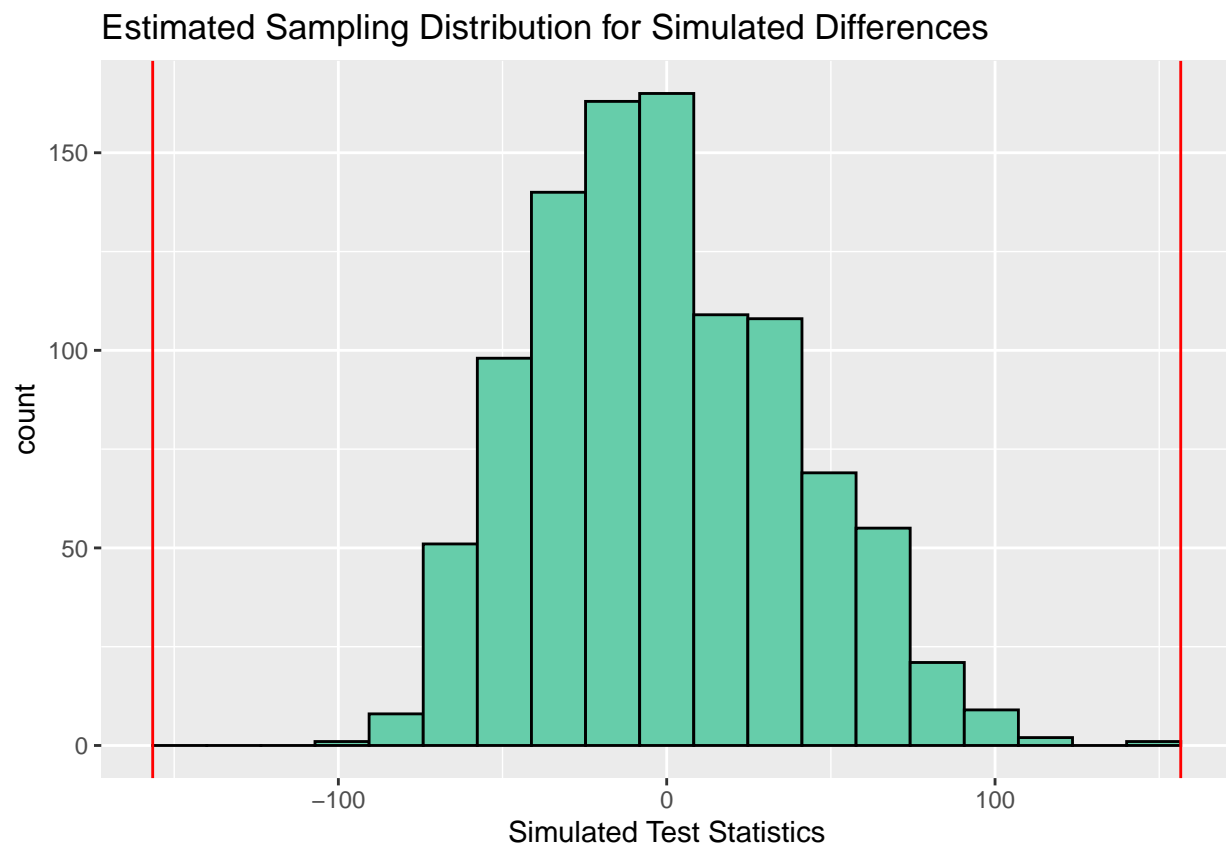
Alternative Hypothesis: The mean time for an auto-theft to be reported in Toronto is **not equal** if the theft occurs during 9am-5pm (inclusive) or if it occurs outside 9am-5pm (exclusive).

Run simultion 10,000 times assuming the null hypothesis is true.
Below is the test statistic:

```
## [1] 156.5538
```

Results of Hypothesis Test

4

P-Value Interpretation

```
## # A tibble: 1 x 1
##   pvalue
##    <dbl>
## 1      0
```

- < 0.001
- Strong evidence against the notion that the difference between the mean report times during and outside work hours is 0
- 0.001% chance of making a Type 1 Error (rejecting the null hypothesis when it is true)
- Probability of obtaining an observation as or more extreme as 156 hours
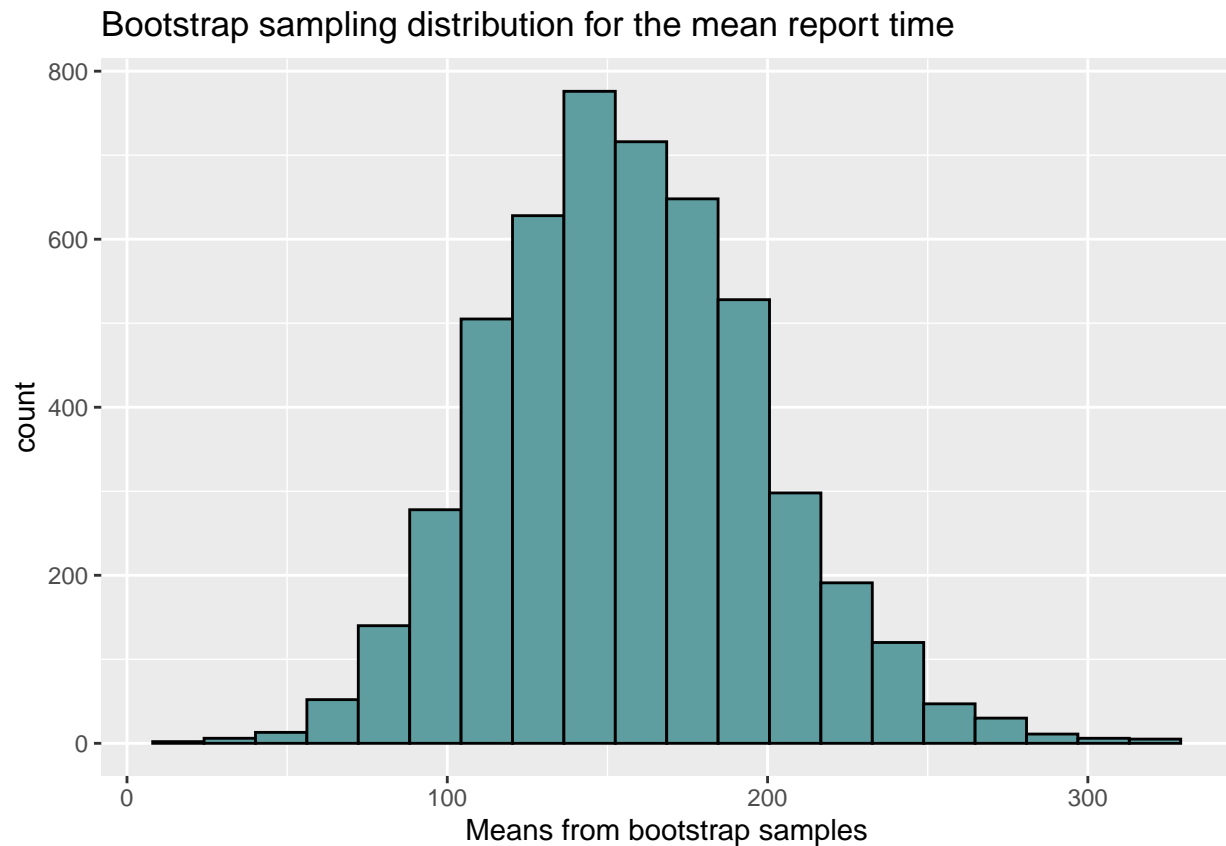
Hypothesis Test to Confidence Interval

- Strong evidence against the null hypothesis
- Opposes preconceived ideas
- Next, we conduct a confidence interval to form new insight on difference between mean report times

Confidence Interval Context & Reasoning

Goal:
* find an interval that is plausible to contain the true difference between mean report times * take a random sample of the same size as the number of our original dataset * generate random bootstrap samples * calculate the difference in mean report times during and outside work hours for each bootstrap sample

Results of Confidence Interval

Bootstrap sampling distribution for the mean report time

Confidence Interval

```
##      2.5%      97.5%
##   81.01468  243.52878
```

Confidence interval: (62.29, 209.65)

Discussion of Results

- 95% confident that our parameter lies within this interval
- since 0 is not in this interval, we have more evidence against our null hypothesis
- positive confidence interval suggests that report time during work hours is greater than report time outside work hours

Limitations

- The 95% confidence interval is arbitrary
- Assumes that work hours for everyone is 9am - 5pm - to generalize our data in order to more effectively analyze
- Only based off of the Toronto Police Service Data and may not be as accurate for other cities
- Only cars that have been reported stolen

Conclusion

Suggestions for the Toronto Police Service

- Educate the public on the importance of a quick report time Helps increase recovery chance of the stolen vehicle
- Expand neighbourhood watch programs for vehicles
- **Possibly** invest more resources during work hours

Acknowledgements and Sources

The authors would like to thank Daphne Choi and Nicholas Martin. The former for an understanding of the issue, and the latter for suggestions and comments that proved to be pivotal in the creation of the project.

Sources 1: http://www.ibc.ca/on/auto/theft/top-ten-stolen-cars

−>