

# How Bike Theft Trends in Toronto Have Changed Over Time

And How Cyclists Can Leverage This Information to Keep Their Bikes Safe

Ricky Yuen, Xinyi Zhang

3/5/2021

- bike purchase
- risk assessment
- if and how they changed over time
- bike features and risk factors associated with bike theft story
- meaningful title
- introduction: goal
- clear sections with meaningful section headers
- for each visualization: what question the visualization addresses
  - what the reader should see that is useful/interesting for understanding the story
  - how its analysis leads to new question and the visualization that follows
- summarize main finding and actionable insights

```
library(ggmap)
```

```
## Warning: package 'ggmap' was built under R version 4.0.4
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 4.0.3
```

```
## Google's Terms of Service: https://cloud.google.com/maps-platform/terms/.
```

```
## Please cite ggmap if you use it! See citation("ggmap") for details.
```

```
library(ggplot2)
```

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## v tibble 3.0.3      v dplyr 1.0.2
```

```
## v tidyr 1.1.2      v stringr 1.4.0
```

```
## v readr 1.3.1      v forcats 0.5.0
```

```
## v purrr 0.3.4
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
library(latticeExtra)
```

```
## Loading required package: lattice
```

```
##
```

```
## Attaching package: 'latticeExtra'
```

```
## The following object is masked from 'package:ggplot2':
```

```
##
```

```
## layer
```

```
data <- read.csv("bicycle-thefts-2014-2019.csv")
```

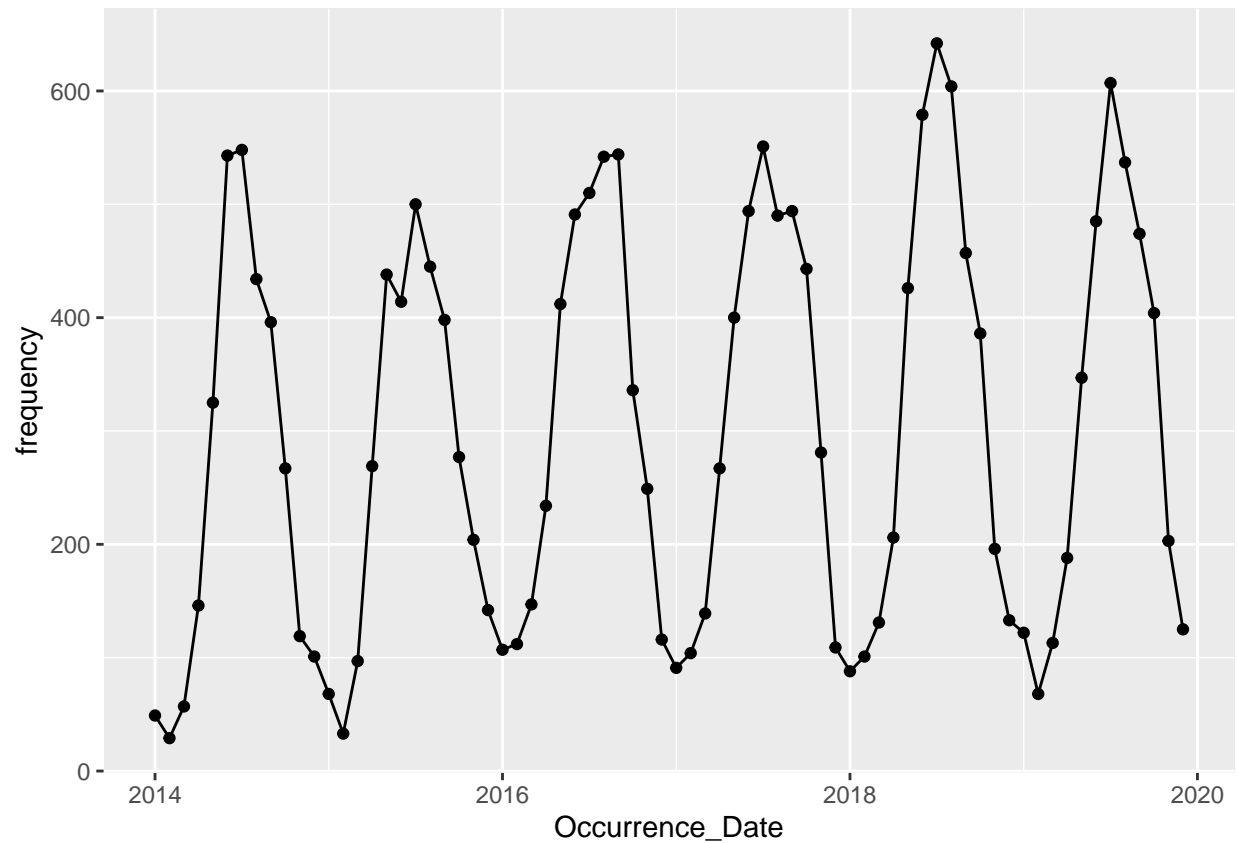
- trends over months and years
- costs over time

```
data$Cost_of_Bike[is.na(data$Cost_of_Bike)] <- 0  
data_month <- data %>%  
  group_by(Occurrence_Year, Occurrence_Month) %>%  
  summarise(frequency=n(), cost=sum(Cost_of_Bike))
```

```
## 'summarise()' regrouping output by 'Occurrence_Year' (override with '.groups' argument)
```

```
data_month$Occurrence_Date <- zoo::as.yearmon(paste(data_month$Occurrence_Year, data_month$Occurrence_Month))
```

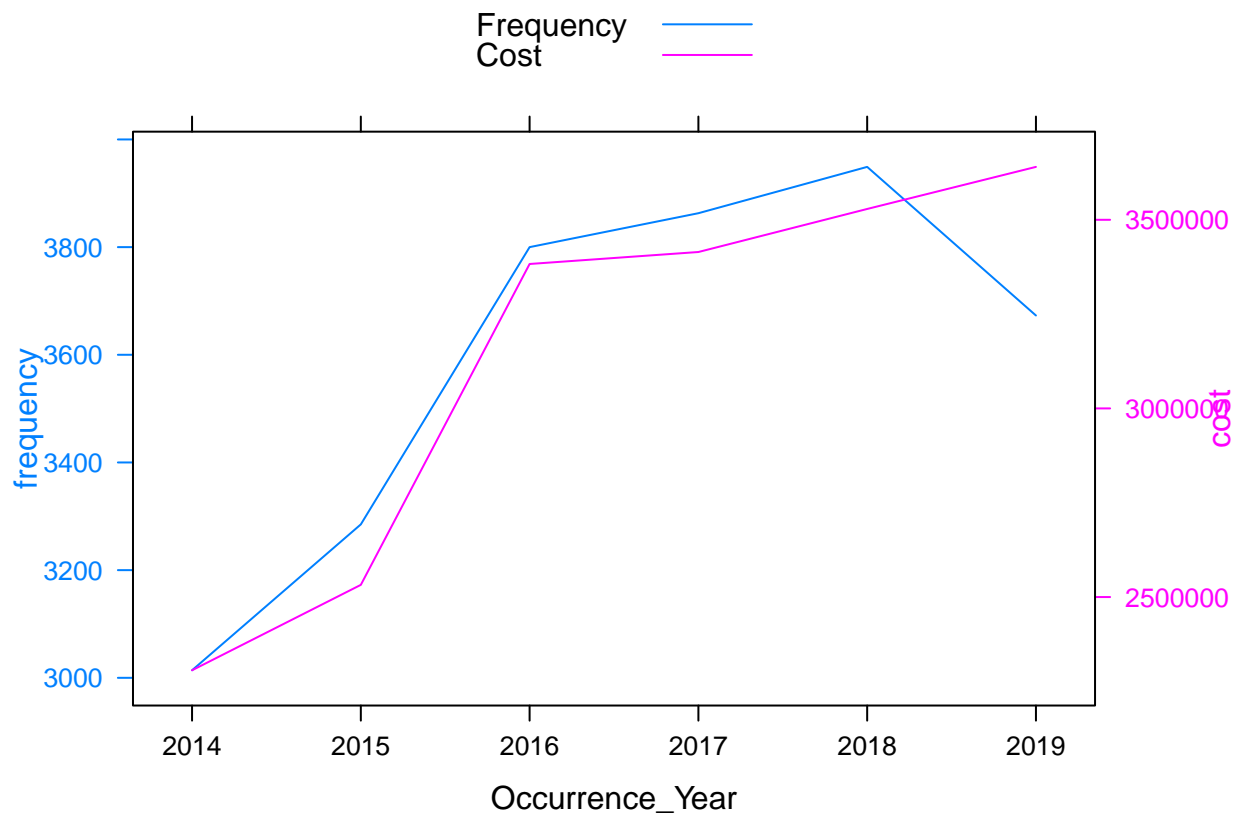
```
ggplot(data=data_month) +  
  geom_line(aes(x=Occurrence_Date, y=frequency)) +  
  geom_point(aes(x=Occurrence_Date, y=frequency))
```



```
data_year <- data_month %>%
  group_by(Occurrence_Year) %>%
  summarise(frequency=sum(frequency),
            cost=sum(cost))
```

```
## 'summarise()' ungrouping output (override with '.groups' argument)
```

```
obj1 <- xyplot(frequency ~ Occurrence_Year, data_year, type="l")
obj2 <- xyplot(cost ~ Occurrence_Year, data_year, type="l")
doubleYScale(obj1, obj2, text=c("Frequency", "Cost"), add.ylab2 = TRUE)
```



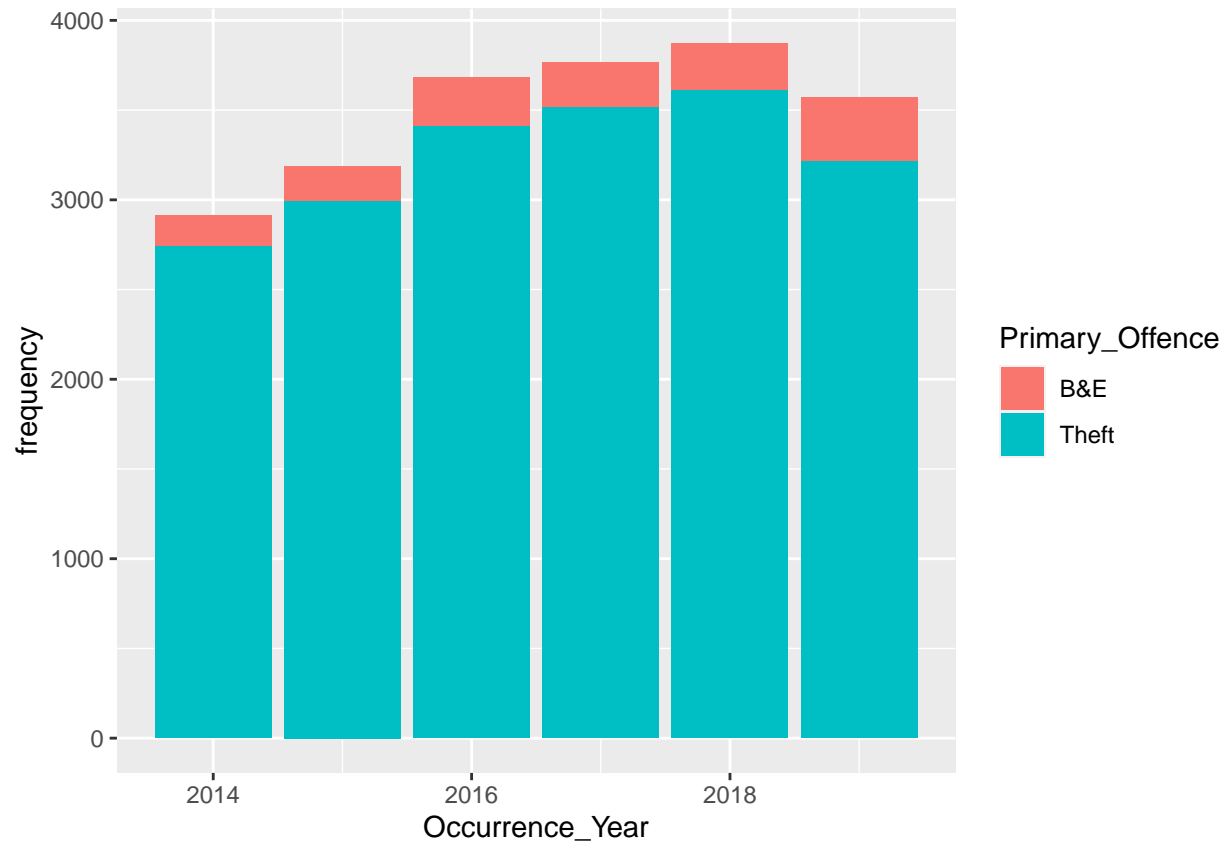
- primary offense over time - make, speed, color - location of theft - location of most thefts - time of day

```
data_offense <- data %>%
  filter(str_detect(Primary_Offence, "THEFT") | str_detect(Primary_Offence, "B&E")) %>%
  mutate(Primary_Offence=str_replace(Primary_Offence, "THEFT.*", "Theft")) %>%
  mutate(Primary_Offence=str_replace(Primary_Offence, "B&E.*", "B&E"))
```

```
data_offense <- data_offense %>%
  group_by(Occurrence_Year, Primary_Offence) %>%
  summarise(frequency=n())
```

## 'summarise()' regrouping output by 'Occurrence\_Year' (override with '.groups' argument)

```
data_offense %>% ggplot(aes(x=Occurrence_Year, y=frequency, fill=Primary_Offence)) +
  geom_bar(stat="identity")
```



## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

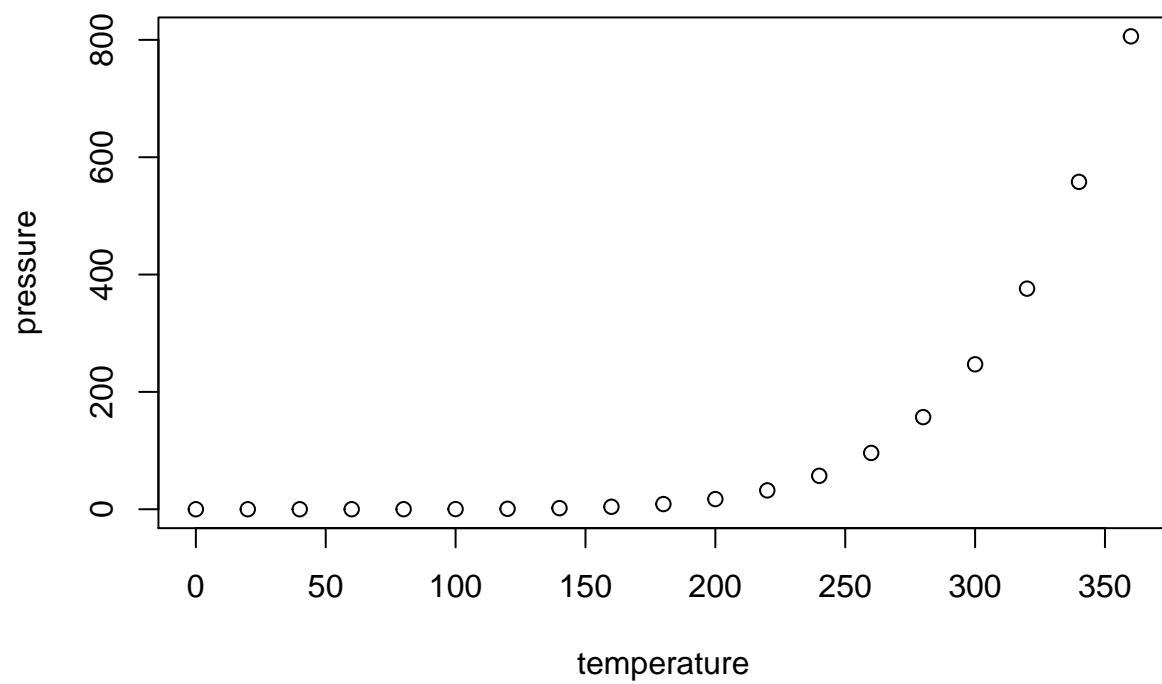
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   :  2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##  Mean   :15.4    Mean   : 42.98
##  3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.   :25.0    Max.   :120.00
```

## Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.