

ECON 334 Week 9 Assignment

Name: TAK SHING POON

Student Number: 301389544

```
library(tidyverse)
library(lubridate)
library(vtable)
library(modelr)
library(broom)
```

Part 1

With “Vancouver crime data”

```
#download the data
data1 <- read_csv("crimedata_csv_AllNeighbourhoods_AllYears.csv", show_col_types = FALSE)

crimes <- data1 %>%
  filter (YEAR >=2006 & YEAR <= 2021)%>% #2006-2021 observations
  mutate (date= make_date(YEAR,MONTH,DAY))%>% # create date var.
  group_by(date,TYPE)%>% #summary on total incident of each type on each date
  summarize(count=n())%>%
  pivot_wider(names_from = TYPE,values_from = count)%>% replace(is.na(.),0)#tidy up the data
```

Summary statistics of daily number of crimes of each type

```
sumtable(crimes)
```

Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Break and Enter Commercial	5844	6.021	3.771	0	4	8	178
Break and Enter Residential/Other	5844	8.773	4.539	0	6	11	32
Mischief	5844	13.558	6.872	2	10	16	376
Offence Against a Person	5844	10.052	4.262	0	7	12	51

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Other Theft	5844	28.072	7.534	6	23	33	62
Theft from Vehicle	5844	30.31	11.592	4	22	38	86
Theft of Bicycle	5844	5.257	4.013	0	2	7	30
Theft of Vehicle	5844	4.429	3.244	0	2	6	24
Vehicle Collision or Pedestrian Struck (with Injury)	5844	3.7	2.209	0	2	5	15
Vehicle Collision or Pedestrian Struck (with Fatality)	5844	0.041	0.203	0	0	0	2
Homicide	5844	0.04	0.197	0	0	0	2

Annual average of the daily count of each type of crime for each year between 2006-2021

```
#create the "crimes" dataset but with years variable
crimesYear <- data1 %>%
  filter (YEAR >=2006 & YEAR <= 2021)%>%
  group_by(YEAR,MONTH,DAY,TYPE)%>%summarize(count=n())%>%
  pivot_wider(names_from = TYPE,values_from = count)%>% replace(is.na(.),0)

#colnames(Avg2[2:11]) #this shows me the exact column names
```

Averages of the column's observations

```
Avg2 <- crimesYear %>%
  group_by(YEAR) %>%
  summarize(`B and E Commercial` = mean(`Break and Enter Commercial`),
    `B and E Residential/Other` = mean(`Break and Enter Residential/Other`),
    Mischief = mean(Mischief),
    `Offence Against a Person` = mean(`Offence Against a Person`),
    `Other Theft` = mean(`Other Theft`),
    `Theft from Vehicle` = mean(`Theft from Vehicle`),
    `Theft of Bicycle` = mean(`Theft of Bicycle`),
    `Theft of Vehicle` = mean(`Theft of Vehicle`),
    `Vehicle Collison(w/Injury)` = mean(`Vehicle Collision or Pedestrian Struck (with
Injury)`),
    `Vehicle Collison(w/Fatality)` = mean(`Vehicle Collision or Pedestrian Struck (with
Fatality)`),
    Homicide = mean(Homicide))
Avg2 %>% head()
```

```
## # A tibble: 6 × 12
##   YEAR B and ...1 B and...2 Misch...3 Offen...4 Other...5 Theft...6 Theft...7 Theft...8 Vehic...9
##   <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
## 1  2006      7.79      15.5      14.2      11.9      28.2      40.4      4.02      10.1      3.80
## 2  2007      6.67      13.7      13.2      12.1      26.6      33.5      3.30      9.05      3.39
## 3  2008      6.08      12.1      14.4      11.5      25.9      30.9      3.21      6.61      3.24
## 4  2009      5.09      9.58      12.1      10.6      25.9      27.4      4.50      5.15      3.50
## 5  2010      4.55      8.95      12.4      10.3      25.7      23.6      4.57      4.02      3.64
## 6  2011      4.8       8.85      13.3      10.6      25.7      20.4      4.16      3.00      3.45
## # ... with 2 more variables: `Vehicle Collison(w/Fatality)` <dbl>,
## #   Homicide <dbl>, and abbreviated variable names 1`B and E Commercial`,
## #   2`B and E Residential/Other`, 3Mischief, 4`Offence Against a Person`,
## #   5`Other Theft`, 6`Theft from Vehicle`, 7`Theft of Bicycle`,
## #   8`Theft of Vehicle`, 9`Vehicle Collison(w/Injury)`
```

Each crime's average per each year (long format)

```
AvgCrimeByType2 <- Avg2 %>%
  pivot_longer(cols = c("B and E Commercial",
                        "B and E Residential/Other",
                        "Mischief","Offence Against a Person",
                        "Other Theft","Theft from Vehicle",
                        "Theft of Bicycle","Theft of Vehicle",
                        "Vehicle Collison(w/Injury)",
                        "Vehicle Collison(w/Fatality)",
                        "Homicide"),
    names_to = "Types of Crime",
    values_to = "avg2")
AvgCrimeByType2 %>% head()
```

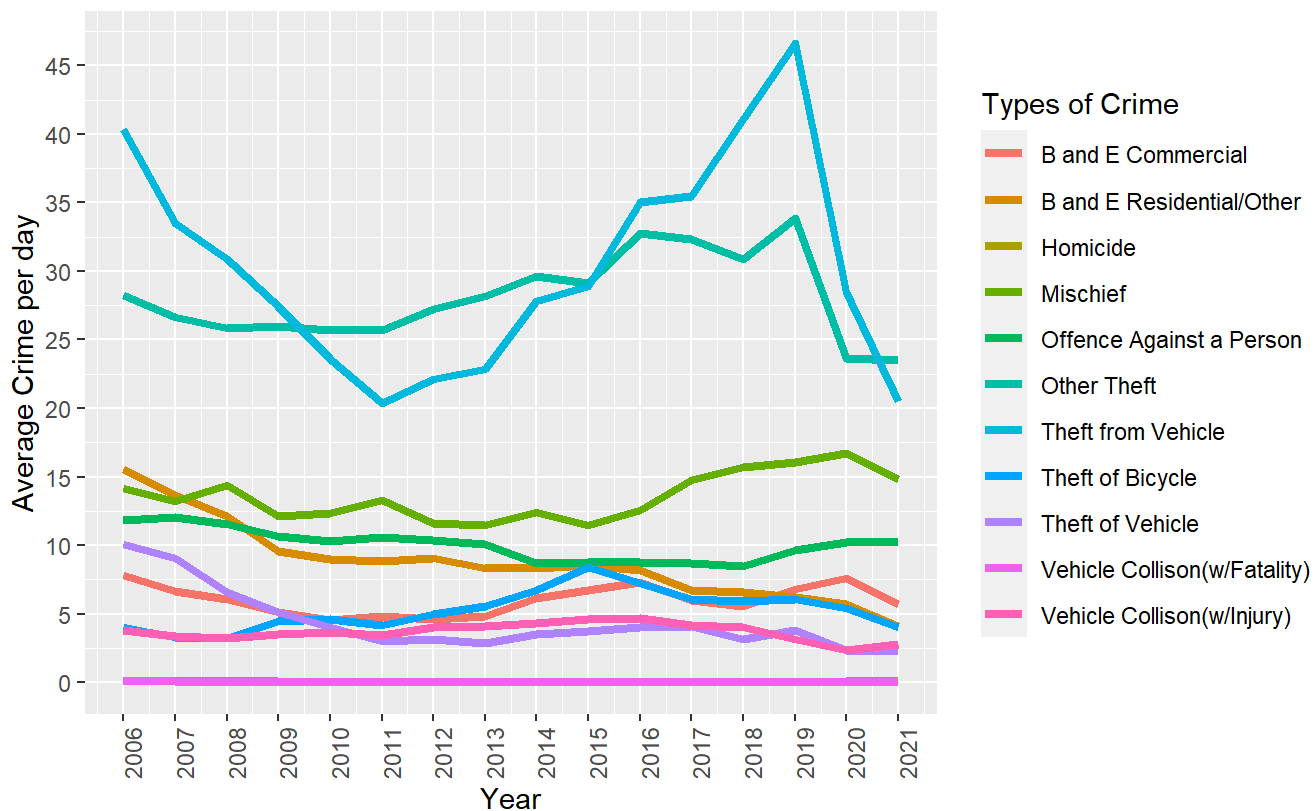
```
## # A tibble: 6 × 3
##   YEAR `Types of Crime`      avg2
##   <dbl> <chr>              <dbl>
## 1  2006 B and E Commercial      7.79
## 2  2006 B and E Residential/Other 15.5
## 3  2006 Mischief                14.2
## 4  2006 Offence Against a Person 11.9
## 5  2006 Other Theft              28.2
## 6  2006 Theft from Vehicle      40.4
```

Plot the average

```
AvgCrimeByType2 %>%
  ggplot(aes(x = YEAR, y = avg2 , color = `Types of Crime`)) +
  geom_line(lwd=1.5) +
  scale_x_continuous(breaks = c(2006:2021)) +
  scale_y_continuous(breaks = seq(0,55,5))+
  theme(axis.text.x = element_text(angle = 90))+
  labs(x = "Year",
       y = "Average Crime per day",
       title = "Annual average daily count of each type of crime",
       subtitle = "For each year between 2006 to 2021",
       caption = "Source: Statistic Canada"
  )
```

Annual average daily count of each type of crime

For each year between 2006 to 2021



Source: Statistic Canada

- Observations: I found that any crimes with death is significantly low compares to other, close to zero. There is a significant spike of offence against a person and other theft crimes from 2016 to 2019. And rest of the crimes remain pretty stagnate also with a slight increase during 2016-2019 period. Overall, crimes rate decreases starting from 2016 and begin rising starting around 2010 and 2011 and decreased after 2019.

Part 2

With “BCHolidays” and “weatherstats_vancouver_daily” dataset

```
BCHoliday <- read_csv("BCHolidays.csv",show_col_types = FALSE)%>% #List of BC holidays 2003-2021
  filter(YEAR >= 2006& YEAR <= 2021)%>%
  mutate (date= make_date(YEAR,MONTH,DAY))%>%
  select(date,holidayname,holiday)
vanwStat <- read_csv("weatherstats_vancouver_daily.csv",show_col_types = FALSE) #daily weather i
nfo 1930s through November 2022
```

Join the crimes data with holiday data

```
#join BCHoliday to Crimes set Logical holiday
crimPLUSHoli <- crimes %>%
  left_join(BCHoliday, by = "date")%>%
  mutate(holidaylogi = if_else(is.na(holidayname),FALSE,TRUE),
    DayOfWeek = wday(date, label = TRUE),
    Month = month(date, label = TRUE))
```

Create a filter for the weather dataset and create a joint dataset with Crime, holiday, and weather. w/ Date variable.

```
filterSTAT <- vanwStat %>%
  select(date,max_temperature,min_temperature,avg_temperature,precipitation)

weathandHOLI <- crimPLUSHoli %>%
  left_join(filterSTAT,by="date")%>%
  mutate(DayOfWeek = wday(date, label = TRUE),
    Month = month(date, label = TRUE))
```

Summary Statistics

```
#only holiday
onlyHoli <- crimPLUSHoli %>%
  filter(holidaylogi== "TRUE") %>%
  select(-(holidayname:Month)) #this makes the sumtable cleaner and shorter
sumtable(onlyHoli)
```

Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Break and Enter Commercial	154	6.286	2.903	0	4	8	18

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Break and Enter Residential/Other	154	9.182	4.43	1	6	12	23
Mischief	154	13.961	4.944	3	11	18	27
Offence Against a Person	154	12.714	9.854	1	7	13	51
Other Theft	154	26.292	8.566	6	21	33	48
Theft from Vehicle	154	31.708	12.259	13	22	39	79
Theft of Bicycle	154	5.416	4.356	0	2	8	25
Theft of Vehicle	154	4.818	3.174	0	2.25	7	16
Vehicle Collision or Pedestrian Struck (with Injury)	154	2.909	1.873	0	2	4	10
Vehicle Collision or Pedestrian Struck (with Fatality)	154	0.039	0.194	0	0	0	1
Homicide	154	0.026	0.16	0	0	0	1

- Observations: Average crimes occurred per day are higher for the holiday only dataset than the dataset for all days. However, most of them are negligible with less than one increase. With Offence Against a Person , Other Theft and Theft from Vehicle increase more than one. However, Homicide decreased from 0.04 (all day) to 0.026(holiday only).

Part 3

Plot weather relationship with crime “Offence Against a Person”

```
#outliers of y > 30 removed
p1 <- weathandHOLI%>%
  ggplot(aes(x = max_temperature,y = `Offence Against a Person`))+ geom_point()+ylim(0, 30)

p2 <- weathandHOLI%>%
  ggplot(aes(x = min_temperature,y = `Offence Against a Person`))+ geom_point()+ylim(0, 30)

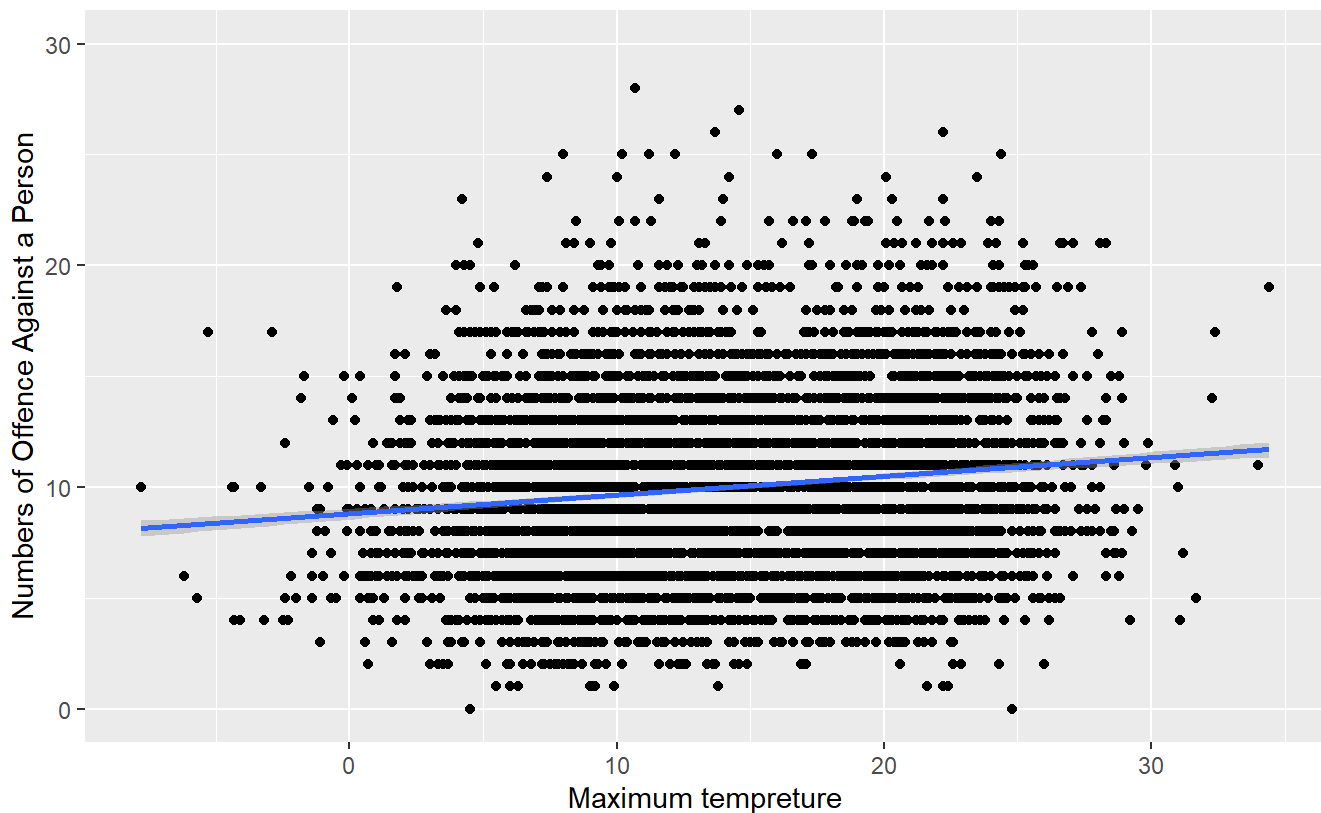
p3 <- weathandHOLI%>%
  ggplot(aes(x = avg_temperature,y = `Offence Against a Person`))+ geom_point()+ylim(0, 30)

p4 <- weathandHOLI%>%
  ggplot(aes(x = precipitation,y = `Offence Against a Person`))+ geom_point()+ylim(0, 30)

p1+geom_smooth(method = "lm")+
  labs(x = "Maximum temperture",
       y = "Numbers of Offence Against a Person",
       title = "Relationship of Offence Against a Person and Maximum temperture" ,
       subtitle = "For each year between 2006 to 2021",
       caption = "Source: BCHolidays & weatherstats_vancouver_daily dataset"
  )
```

Relationship of Offence Against a Person and Maximum temperture

For each year between 2006 to 2021

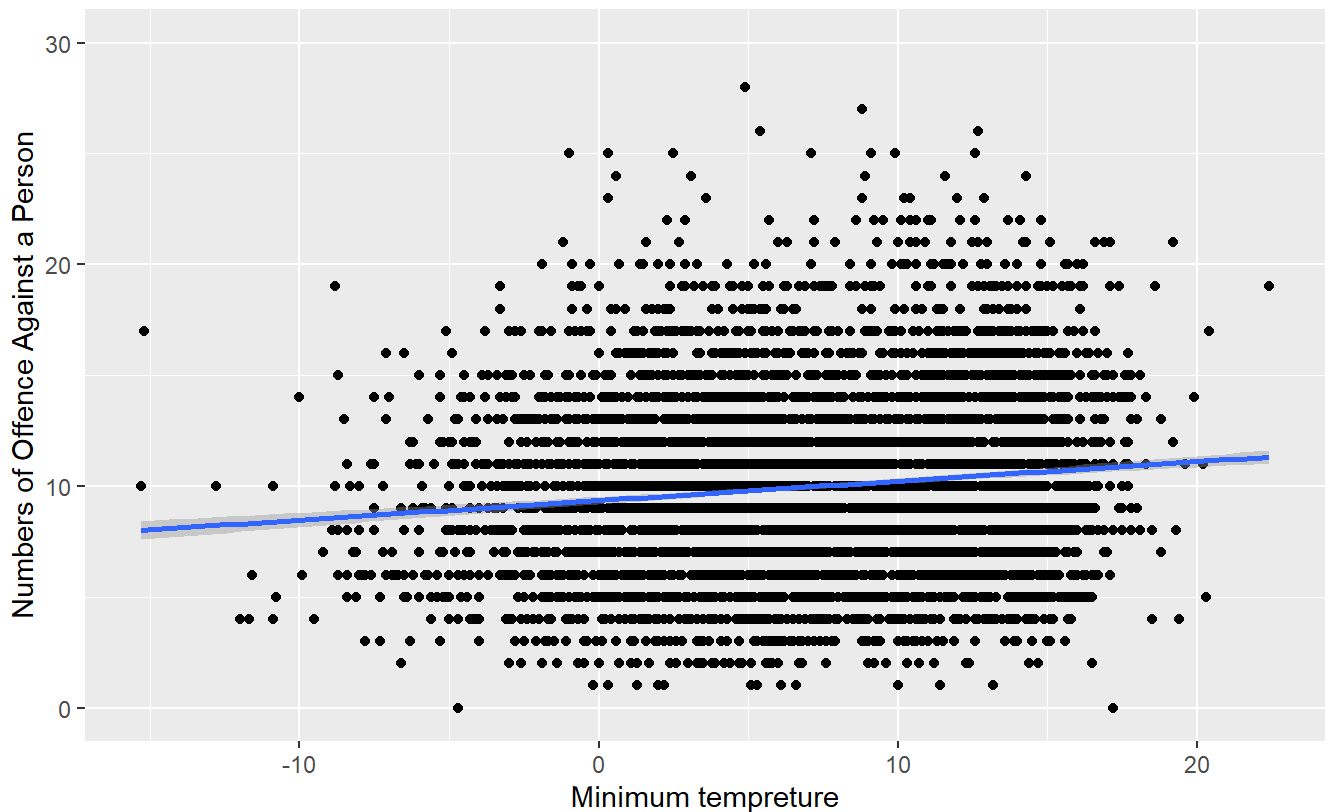


Source: BCHolidays & weatherstats_vancouver_daily dataset

```
p2+geom_smooth(method = "lm")+
  labs(x = "Minimum temperture",
       y = "Numbers of Offence Against a Person",
       title = "Relationship of Offence Against a Person and Minimum temperture" ,
       subtitle = "For each year between 2006 to 2021",
       caption = "Source: BCHolidays & weatherstats_vancouver_daily dataset"
  )
```

Relationship of Offence Against a Person and Minimum temperture

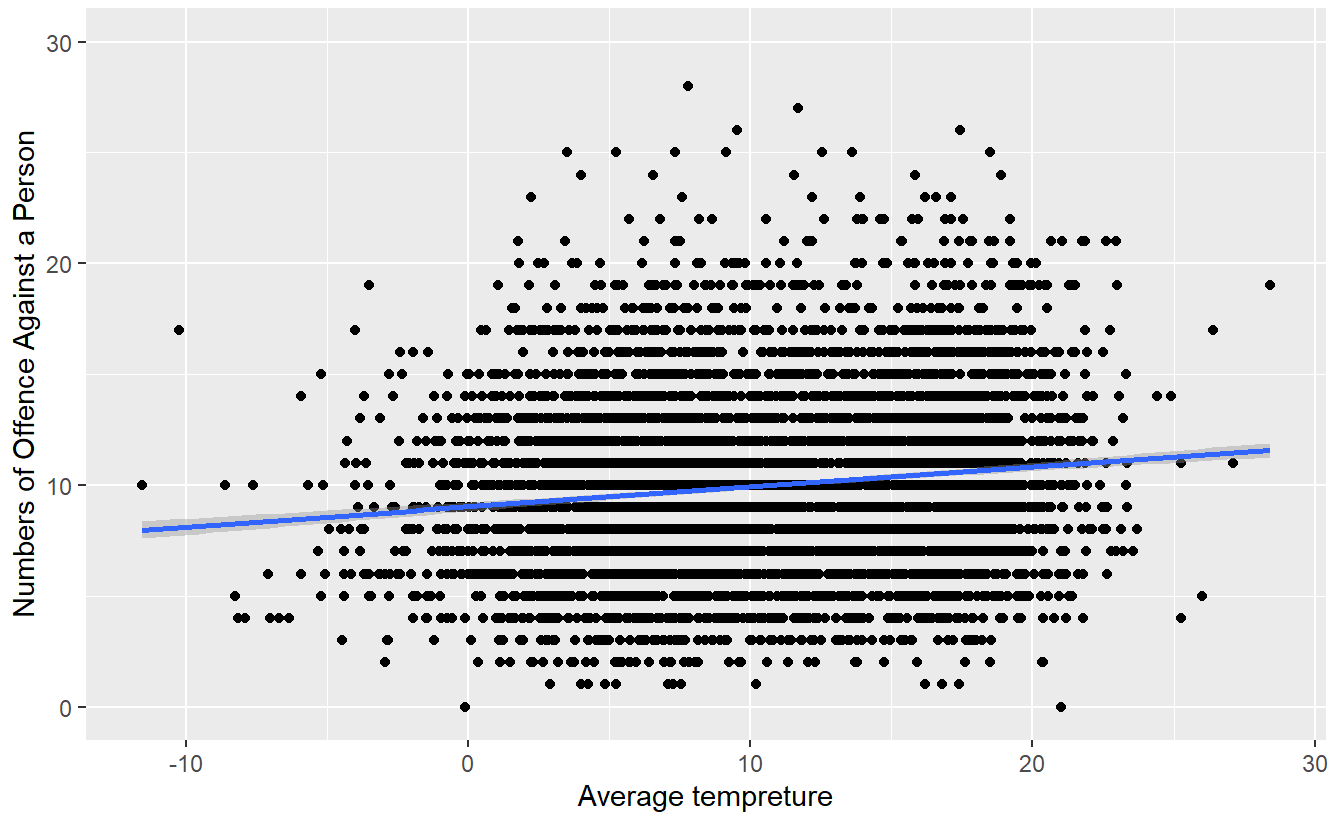
For each year between 2006 to 2021



```
p3+geom_smooth(method = "lm")+
  labs(x = "Average temperture",
       y = "Numbers of Offence Against a Person",
       title = "Relationship of Offence Against a Person and Average temperture" ,
       subtitle = "For each year between 2006 to 2021",
       caption = "Source: BCHolidays & weatherstats_vancouver_daily dataset"
  )
```


Relationship of Offence Against a Person and Average temprature

For each year between 2006 to 2021

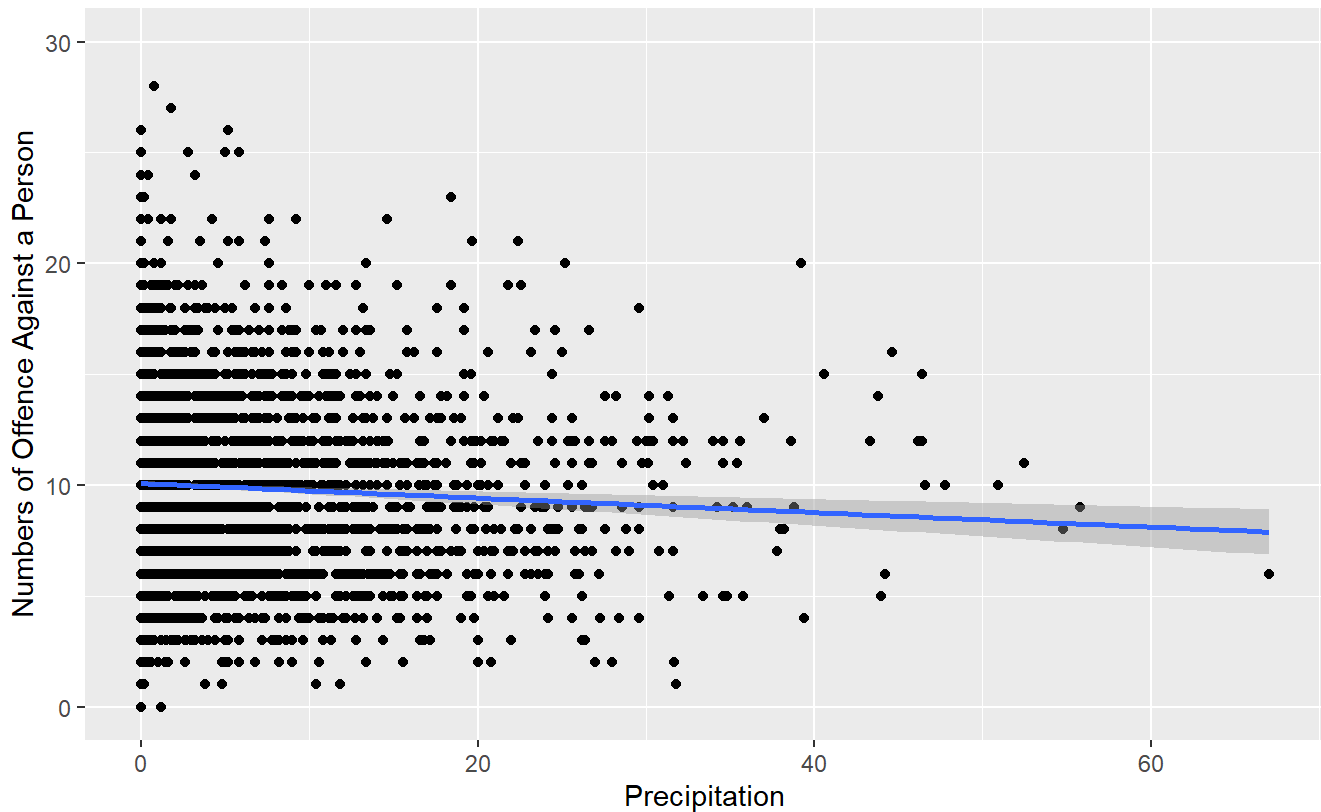


Source: BCHolidays & weatherstats_vancouver_daily dataset

```
p4+geom_smooth(method = "lm")+
  labs(x = "Precipitation",
       y = "Numbers of Offence Against a Person",
       title = "Relationship of Offence Against a Person and Maximum temprature" ,
       subtitle = "For each year between 2006 to 2021",
       caption = "Source: BCHolidays & weatherstats_vancouver_daily dataset"
  )
```

Relationship of Offence Against a Person and Maximum temperature

For each year between 2006 to 2021



Source: BCHolidays & weatherstats_vancouver_daily dataset

- Observations: All temperature variables suggest positive correlation except precipitation weather variable with little negative correlation with the Offence Against a Person variable. The first three plots and regression are very similar because they all suggest a relationship with temperature. They all suggest most crime committed for all years and highest average crime committed is at around 10 degrees C. Whereas, the 4th plot of precipitation and crime suggest that most crime committed for all years are with no precipitation, vice versa. This can suggest that criminals are most comfortable to commit a crime at around 5 ~ 15 degrees with no rain.

Estimate regression 1

```
out1 <- lm(`Offence Against a Person` ~ max_temperature, weathandHOLI)
summary(out1)
```

```
##
## Call:
## lm(formula = `Offence Against a Person` ~ max_temperature, data = weathandHOLI)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.796  -2.839  -0.482   2.391  41.246
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    9.085009    0.131642   69.01 < 2e-16 ***
## max_temperature 0.068975    0.008515    8.10 6.61e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.238 on 5842 degrees of freedom
## Multiple R-squared:  0.01111,    Adjusted R-squared:  0.01094
## F-statistic: 65.62 on 1 and 5842 DF,  p-value: 6.61e-16
```

Estimate resgreession 2

```
weathandHOLI <- weathandHOLI %>%
  mutate(DayOfWeek = factor(DayOfWeek, ordered = FALSE),
         Month = factor(Month, ordered = FALSE))

out2 <- lm(`Offence Against a Person` ~ max_temperature+ DayOfWeek + Month + holidaylogi,data= w
eathandHOLI)
summary(out2)
```

```
##
## Call:
## lm(formula = `Offence Against a Person` ~ max_temperature + DayOfWeek +
##      Month + holidaylogi, data = weathandHOLI)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.172  -2.698  -0.375   2.318  37.081
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    10.96364    0.25785  42.519 < 2e-16 ***
## max_temperature  0.04919    0.01878   2.620 0.008822 **
## DayOfWeekMon    -2.51839    0.20079 -12.543 < 2e-16 ***
## DayOfWeekTue    -2.25849    0.19807 -11.402 < 2e-16 ***
## DayOfWeekWed    -1.81737    0.19808  -9.175 < 2e-16 ***
## DayOfWeekThu    -1.72302    0.19808  -8.699 < 2e-16 ***
## DayOfWeekFri    -1.02966    0.19813  -5.197 2.10e-07 ***
## DayOfWeekSat     0.95143    0.19814   4.802 1.61e-06 ***
## MonthFeb        -0.83063    0.26343  -3.153 0.001623 **
## MonthMar        -0.63499    0.26353  -2.410 0.016004 *
## MonthApr        -1.00406    0.28302  -3.548 0.000392 ***
## MonthMay        -0.38207    0.32117  -1.190 0.234241
## MonthJun        -0.54564    0.35605  -1.532 0.125464
## MonthJul        -0.17084    0.39338  -0.434 0.664091
## MonthAug         0.02067    0.39036   0.053 0.957774
## MonthSep        -0.42452    0.34420  -1.233 0.217483
## MonthOct        -0.48630    0.28480  -1.707 0.087783 .
## MonthNov        -0.70337    0.26346  -2.670 0.007612 **
## MonthDec        -0.81760    0.25741  -3.176 0.001500 **
## holidaylogiTRUE  3.50797    0.33935  10.337 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.047 on 5824 degrees of freedom
## Multiple R-squared:  0.1011, Adjusted R-squared:  0.09819
## F-statistic: 34.48 on 19 and 5824 DF, p-value: < 2.2e-16
```

```
regresion_models <- ungroup(weathandHOLI) %>%
  add_residuals(model=out1, var = "m1residuals") %>%
  add_residuals(model=out2, var = "m2residuals")

msfe <- regresion_models %>%
  summarize(MSFE1 = mean(m1residuals^2),
            MSFE2 = mean(m2residuals^2))
msfe
```

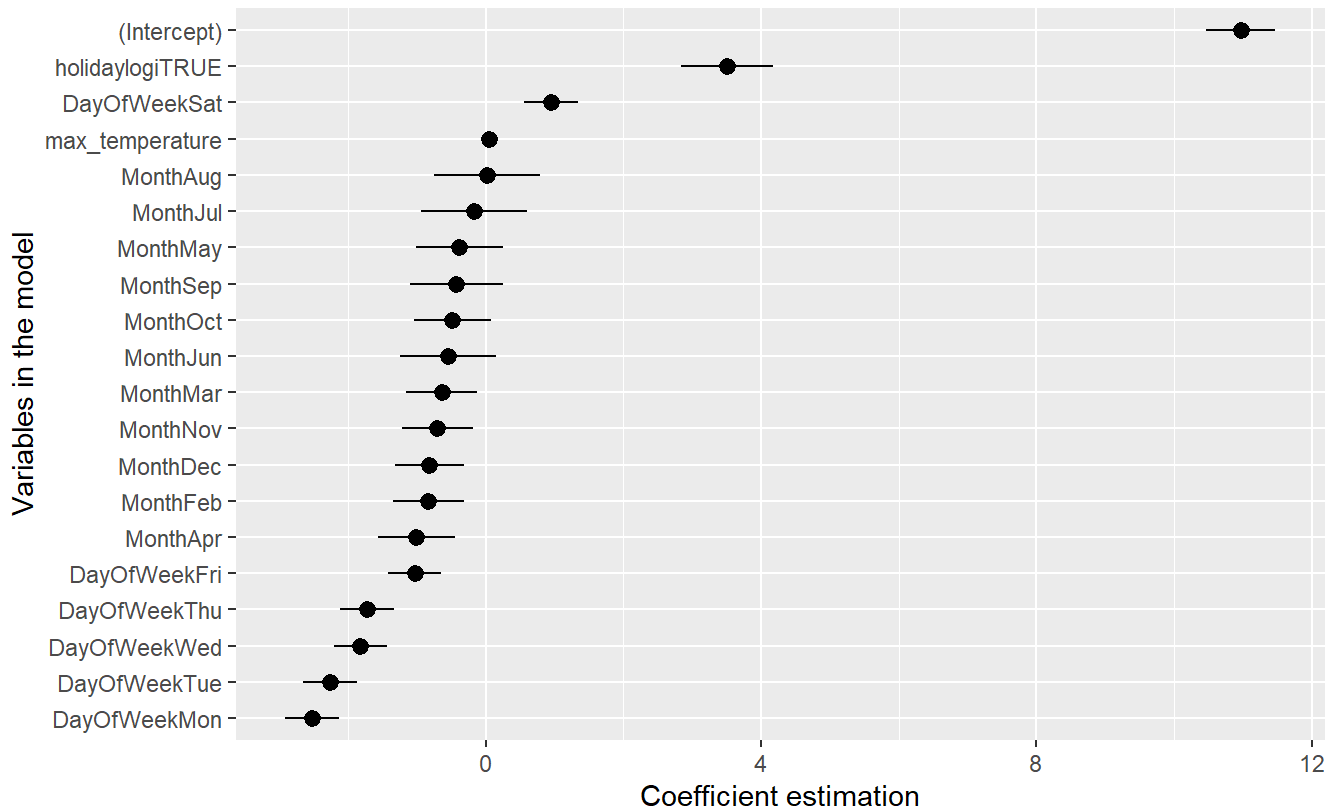
```
## # A tibble: 1 × 2
##   MSFE1 MSFE2
##   <dbl> <dbl>
## 1  18.0  16.3
```

- Comment: The second model has the smallest MSFE. This suggest that as we include more variable, it improves the model fit by reducing the error. The residual error decreased, which suggest that some or at least one variable that improves the fit of the model.

```
p <- tidy(out2, conf.int = TRUE) %>%
  ggplot(aes(x = reorder(term,estimate), y =estimate, ymin=conf.low,ymax=conf.high))
p+geom_pointrange()+coord_flip()+
  labs(y = "Coefficient estimation",
       x = "Variables in the model",
       title = "Confidence interval for regression estimations",
       subtitle = "With model 'out2'",
       caption = "Source: From the `weathandHOLI` dataset"
  )
```

Confidence interval for regression estimations

With model 'out2'



Source: From the `weathandHOLI` dataset

- Comment: This graph suggest that max-temperature and month(Aug, Jul, May , Sep, Oct, Jun) is not significant in estimating the Offence Against a Person value because their confidence intervals include a zero. Confidence intervals include a zero means that its estimate coefficient value could potential be a zero. And as suggested the coefficient estimation is the effects of 1 unit change of itself have on the y variable Offence Against a Person . The rest of the variables with $\Pr(>|t|)$ small than 0.05 suggest the estimated

value is significant. Saturday has the highest day of the week highest incidence. March has the highest incidence which is significant. August has the highest incidence which is potentially insignificant.

MADE BY_TAK SHING POON(301389544)