Assessing the associations between unemployment and cost of living on criminal activity: A city based comparison within New Zealand

Team: Ben Johnston, Sreekala Pramod, Lin Huang, Xiruo Song and Bhanupratap Vanga

Group Project: INFO634

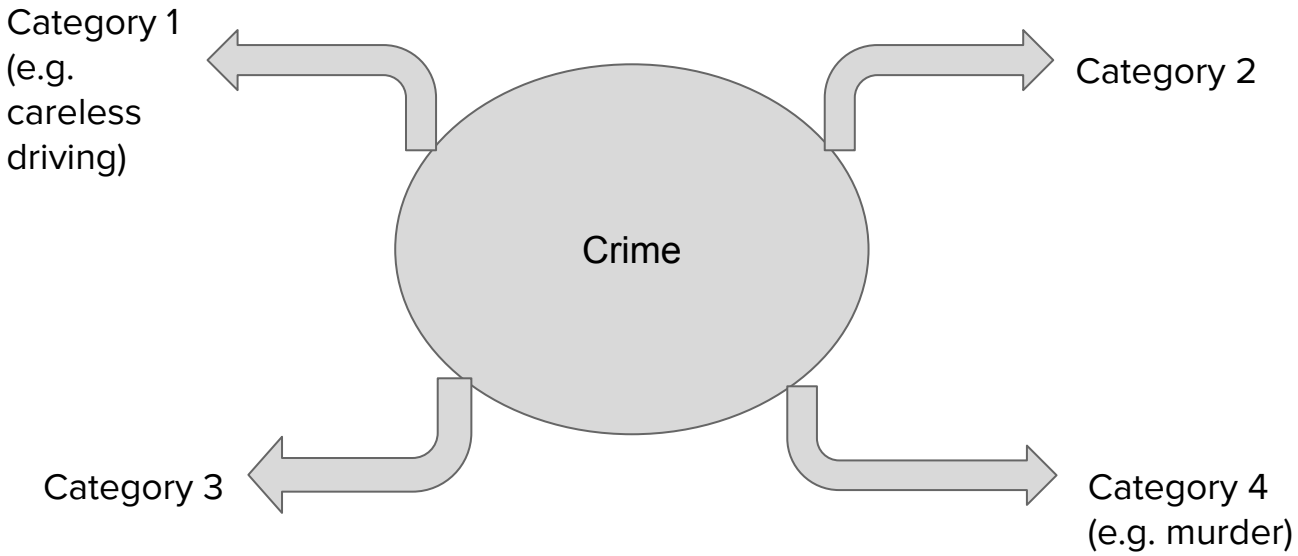
New Zealand's 'image'

- Many people come to our shores to visit or live

- New Zealand is not a country associated with crime or being a dangerous place in general

- In recent times many outlets have reported on increased rates of criminal activity in New Zealand cities

What do we mean by crime?



We are looking at 3 types:

- Robbery
- Theft of property
- Unlawful entry

Factors investigated plus other important points

1) Unemployment rate

2) Household cost of living index

3) Investigated the impacts of these variables over a 5 year timeframe

4) This investigation looks at before, during and after covid

5) An inter city comparison between Christchurch and Wellington

Hypotheses

Many international studies show that Covid-19 significantly impact Crime rates.

Crime Number

Therefore, we make our first hypothesis

Hypothesis 1: Unemployment Rate

Then Hypothesis 2: Cost of living Rate

Based on the hypotheses and research data sets were aggregated from various sources for all **three factors**, as well as the two Crime indicators(Type Theft and Related Offences,Sexual Assault and Related Offences cases).

Data pre-processing

- 3 datasets : crime dataset, cost of living dataset, unemployment dataset.

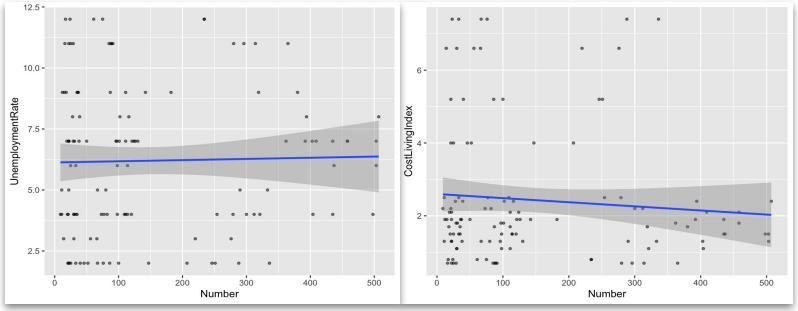
- Clean and structured
- Result

	District	Time	Type	Number	UnemploymentRate	CostLivingIndex
1	Canterbury	2017Q3	Unlawful Entry With Intent/Burglary, Break and Enter	37	4.6	1.9
2	Canterbury	2017Q4	Unlawful Entry With Intent/Burglary, Break and Enter	101	4.4	1.8
3	Canterbury	2018Q1	Unlawful Entry With Intent/Burglary, Break and Enter	111	4.6	1.7
4	Canterbury	2018Q2	Unlawful Entry With Intent/Burglary, Break and Enter	126	4.4	1.9
5	Canterbury	2018Q3	Unlawful Entry With Intent/Burglary, Break and Enter	73	3.9	2.2
6	Canterbury	2018Q4	Unlawful Entry With Intent/Burglary, Break and Enter	124	4.4	2.1
7	Canterbury	2019Q1	Unlawful Entry With Intent/Burglary, Break and Enter	130	4.4	1.3
8	Canterbury	2019Q2	Unlawful Entry With Intent/Burglary, Break and Enter	114	3.9	1.5

Data Analysis : Pearson's correlation coefficient analysis

Five conditions

- Condition 1: continuous variables
- Condition 2: The continuous variables should be paired
- Condition 3: There is a linear relationship



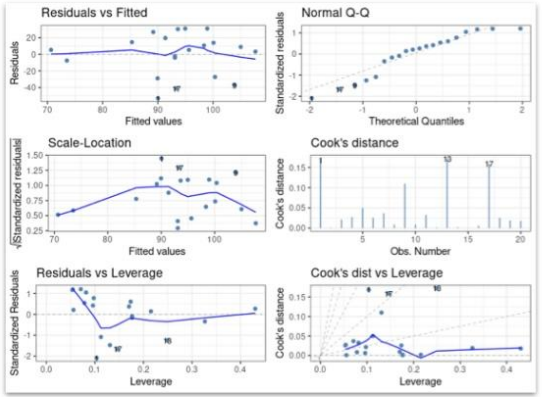
- Condition 4: no obvious outliers

```
> summary(Crime_Offenders$Number)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   9.00   30.75   90.00  149.83  259.50  507.00
> summary(Crime_Offenders$UnemploymentRate)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   2.00    4.00    6.50    6.20    8.25   12.00
> summary(Crime_Offenders$CostLivingIndex)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.700   1.300   1.850   2.430   2.425   7.400
```

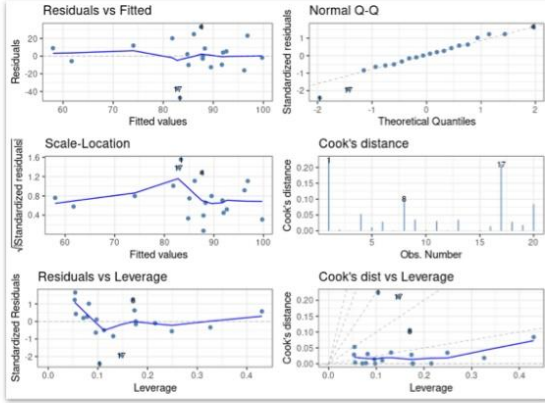
```
> is.na(Crime_Offenders$Number)
  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [17] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [33] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [65] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [81] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [97] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[113] FALSE FALSE FALSE FALSE FALSE FALSE FALSE
> is.na(Crime_Offenders$UnemploymentRate)
  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [17] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [33] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [65] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [81] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [97] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[113] FALSE FALSE FALSE FALSE FALSE FALSE FALSE
> is.na(Crime_Offenders$CostLivingIndex)
  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [17] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [33] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [65] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [81] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [97] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[113] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
>
```

➢ Condition 5: normal distribution

```
autoplot(regCanterburyBurlary, which = 1:6, ncol = 2,
label.size = 3,colour = "steelblue") + theme_bw()
```




```
autoplot(regWellingtonBurlary, which = 1:6, ncol = 2,
label.size = 3,colour = "steelblue") + theme_bw()
```



Q–Q plot

```
shapiro.test(Crime_Offenders$Number)
shapiro.test(Crime_Offenders$UnemploymentRate)
shapiro.test(Crime_Offenders$CostLivingIndex)
```

Shapiro–Wilk test function

Data Analysis : Pearson's correlation coefficient analysis

```
> cor.test(Crime_Offenders$Number,Crime_Offenders$UnemploymentRate)

        Pearson's product-moment correlation

data:  Crime_Offenders$Number and Crime_Offenders$UnemploymentRate
t = 0.25008, df = 118, p-value = 0.803
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.1568725  0.2014264
sample estimates:
       cor
0.02301603

> cor.test(Crime_Offenders$Number,Crime_Offenders$CostLivingIndex)

        Pearson's product-moment correlation

data:  Crime_Offenders$Number and Crime_Offenders$CostLivingIndex
t = -0.97043, df = 118, p-value = 0.3338
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.26401132  0.09172354
sample estimates:
       cor
-0.0889807
```

- Cor
  - 0.0230
  - 0.0889

0</r/<0.2

- P-value
  - 0.803
  - 0.338

>0.05

Very weak correlation or no correlation !

Data Analysis : Linear Regression an

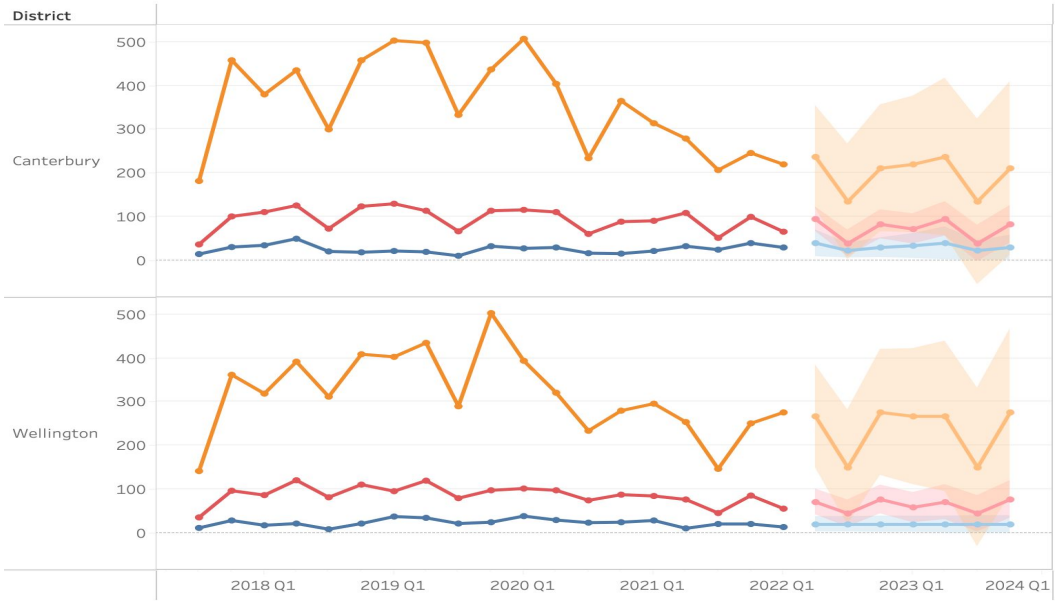
statistically non-significant !

```
> regCanterburyBurlary<-lm(formula=Crime_Offenders_Canterbury_Burglary$Number~Crime_Offenders_Canterbury_B
urglary$CostLivingIndex +Crime_Offenders_Canterbury_Burglary$UnemploymentRate)
> summary(regCanterburyBurlary)

Call:
lm(formula = Crime_Offenders_Canterbury_Burglary$Number ~ Crime_Offenders_Canterbury_Burglary$CostLivingIn
dex +
    Crime_Offenders_Canterbury_Burglary$UnemploymentRate)

Residuals:
    Min      1Q  Median      3Q     Max
-53.027 -12.477   5.361  15.931  31.047

Coefficients:
                                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                                       124.867     26.097   4.785 0.000172 ***
Crime_Offenders_Canterbury_Burglary$CostLivingIndex    -6.660      4.441  -1.500 0.152044
Crime_Offenders_Canterbury_Burglary$UnemploymentRate   -2.465      2.702  -0.912 0.374428
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 26.76 on 17 degrees of freedom
Multiple R-squared:  0.118,     Adjusted R-squared:  0.01425
F-statistic: 1.137 on 2 and 17 DF,  p-value: 0.3439
```

>0.05

Time series analysis

<Proceedings of three type crimes for Canterbury and Wellington>



The trend of sum of Proceedings (actual & forecast) for Time Quarter broken down by District. Color shows details about Crime Type and Forecast indicator.

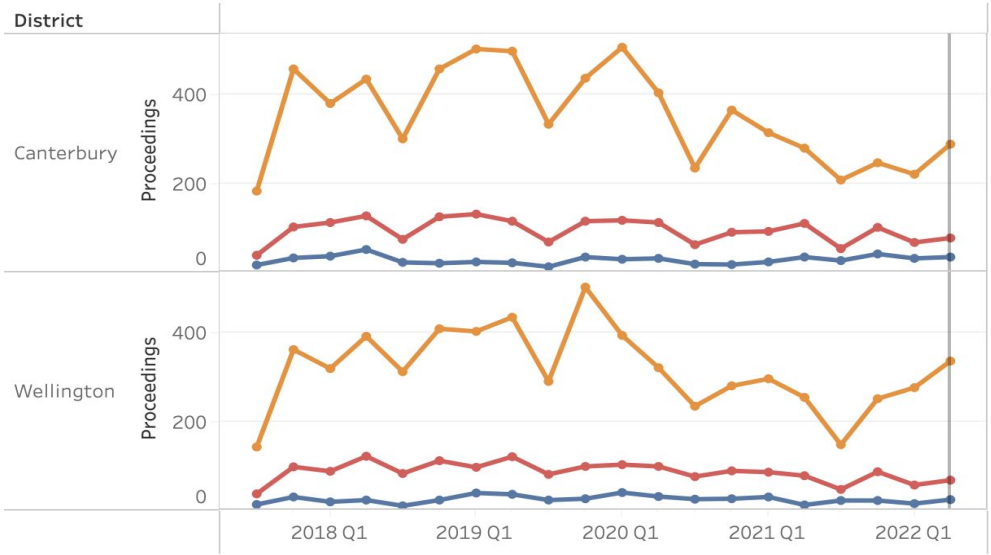
Crime Type, Forecast indicator

- Theft and Related Offences, Actual
- Theft and Related Offences, Estimate
- Unlawful Entry With Intent/Burglary, Break and Enter, Actual
- Unlawful Entry With Intent/Burglary, Break and Enter, Estimate
- Robbery, Extortion and Related Offences, Actual
- Robbery, Extortion and Related Offences, Estimate

- Time series analysis is a specific way of analyzing a sequence of data points collected over a period of time
- It shows how variables change over time and allow us to identify any trends or patterns which can be used for predicting future data based.

Change point analysis

<Crime proceedings for Canterbury and Wellington>



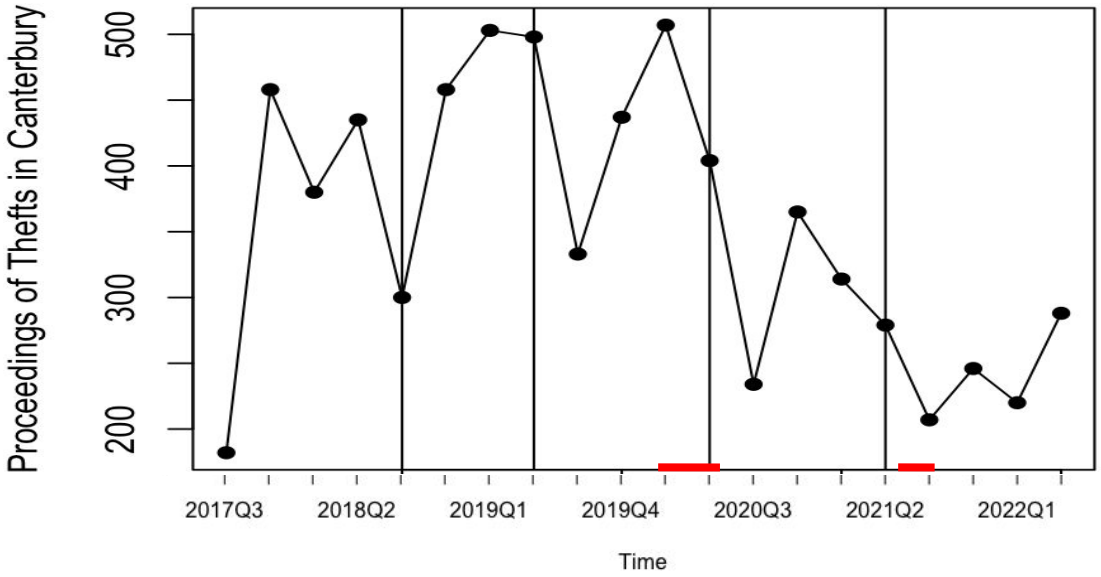
The trend of sum of Proceedings for Time Quarter broken down by District. Color shows details about ANZSOC. The view is filtered on District and Time Quarter. The District filter keeps Canterbury and Wellington. The Time Quarter filter keeps non-Null values only.

ANZSOC
- Theft and Related Offences
- Unlawful Entry With Intent/Burglary, Break and Enter
- Robbery, Extortion and Related Offences

- Change point analysis is used to identify the sudden variations in time series data
- The detection of change points is useful in modeling and prediction of time series

Change point analysis

Using R package strucchange



R result:

Optimal 5-segment partition:

Call:

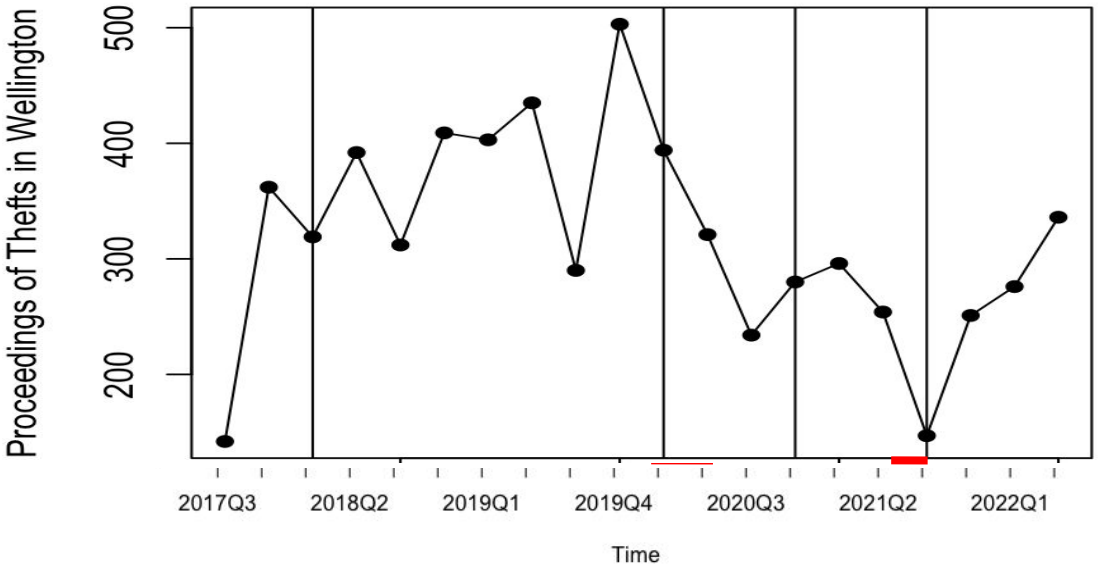
breakpoints.breakpointsfull(obj = bp, breaks = 4)

Breakpoints at observation number:

5 8 12 16

Change point analysis

Using R package strucchange



R result:

Optimal 5-segment partition:

Call:

breakpoints.breakpointsfull(obj = bp, breaks = 4)

Breakpoints at observation number:

3 11 14 17