Data pre-processing

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

CrimeQuarter.csv
Household-living-costs-price-indexesMain.csv
UnemploymentRatebyAgebySex.csv

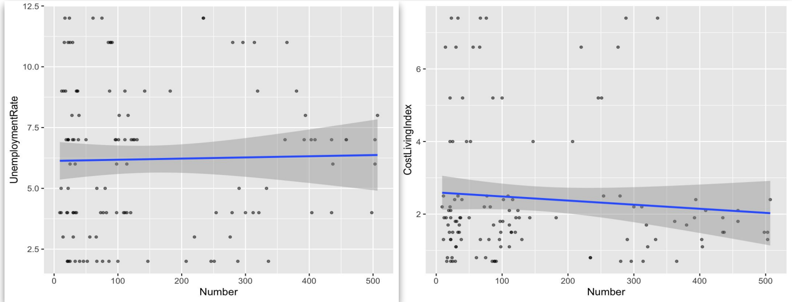
- Clean and structured
- Result

	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



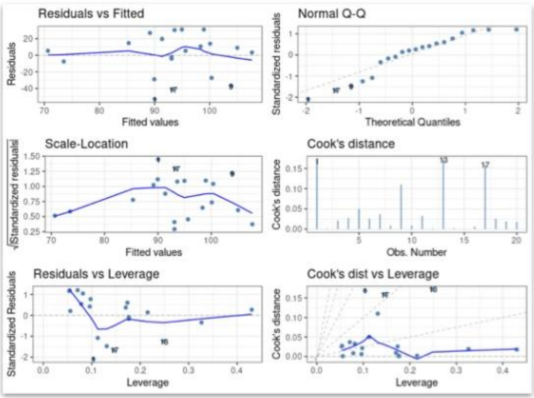
- ➢ Condition 4: no obvious outliers

```
> 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
```

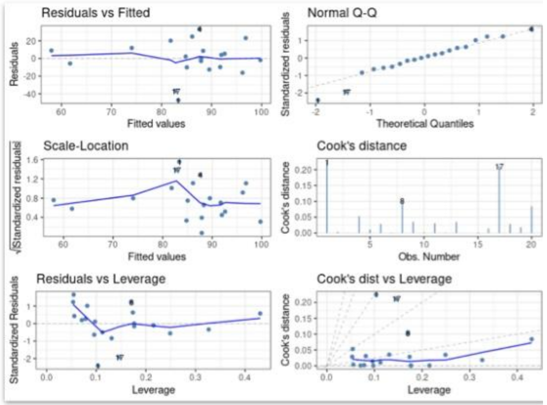
```
> is.na(Crime_Offenders$Number)
  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [17] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [33] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE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

```
> 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
```

statistically non-significant !

>0.05