

Lab 4: Visual Perception

The purpose of this lab is to apply principles of visual perception to evaluate and improve data visualisations.

The Data Set

The data set is a CSV file, `nzpolice-proceedings.csv`, which was derived from “Dataset 5” of Proceedings (offender demographics) on the [policedata.nz](https://www.policedata.nz) web site.

We can read the data into an R data frame with `read.csv()`.

```
crime <- read.csv("nzpolice-proceedings.csv")
crime$Month <- as.Date(crime$Date)
```

We will focus on youth crime (aged 15-19 inclusive).

```
youth <- subset(crime, Age.Lower == 15)
head(youth)
```

	Age.Lower	Police.District	ANZSOC.Division	SEX	Date
1	15	Tasman	Acts Intended to Cause Injury	Female	2015-12-01
5	15	Auckland City	Acts Intended to Cause Injury	Female	2015-12-01
6	15	Auckland City	Acts Intended to Cause Injury	Female	2015-12-01
22	15	Auckland City	Dangerous or Negligent Acts Endangering Persons	Female	2015-12-01
23	15	Auckland City	Dangerous or Negligent Acts Endangering Persons	Female	2015-12-01
41	15	Auckland City	Illicit Drug Offences	Female	2015-12-01

	Month
1	2015-12-01
5	2015-12-01
6	2015-12-01
22	2015-12-01
23	2015-12-01
41	2015-12-01

The following code reorders the levels of the `ANZSOC.Division` factor according to the overall counts for each type of crime. It also generates `typelabels`, which are line-wrapped versions of the `ANZSOC.Division` levels.

```
divLevels <- levels(factor(youth$ANZSOC.Division))
divLabels <- unlist(lapply(strwrap(divLevels, width=30, simplify=FALSE),
  function(x) {
    if (length(x) < 3)
      x <- c(x, rep(" ", 3 - length(x)))
    paste(x, collapse="\n")
  })))
types <- table(youth$ANZSOC.Division)
typeLevels <- names(types)[order(types, decreasing=TRUE)]
typeLabels <- unlist(lapply(strwrap(typeLevels, width=30, simplify=FALSE),
  function(x) {
    if (length(x) < 3)
      x <- c(x, rep(" ", 3 - length(x)))
    paste(x, collapse="\n")
  })))
youth$Type <- factor(youth$ANZSOC.Division, levels=typeLevels)
```

The following code generates a table of counts for the number of incidents for youth broken down by type of crime.

```
youthType <- as.data.frame(table(youth$Type))
head(youthType)
```

	Var1	Freq
1	Theft and Related Offences	23280
2	Acts Intended to Cause Injury	17456
3	Traffic and Vehicle Regulatory Offences	15737
4	Public Order Offences	12358
5	Dangerous or Negligent Acts Endangering Persons	11144
6	Property Damage and Environmental Pollution	9534

The following code reorders the levels of the `ANZSOC.Division` factor according to the *first* month count for each type of crime. It also generates `newlabels`, which are line-wrapped versions of the `ANZSOC.Division` levels.

```
monthTypes <- table(youth$ANZSOC.Division, youth$Month)[,1]
monthTypeLevels <- names(types)[order(monthTypes, decreasing=TRUE)]
monthTypeLabels <- unlist(lapply(strwrap(monthTypeLevels, width=30,
  simplify=FALSE),
  function(x) {
    if (length(x) < 3)
      x <- c(x, rep(" ", 3 - length(x)))
    paste(x, collapse="\n")
  })))
youth$monthType <- factor(youth$ANZSOC.Division, levels=monthTypeLevels)
```

The following code generates a table of counts for the number of incidents per month, broken down by type of crime.

```
youthMonthType <- as.data.frame(table(youth$Month, youth$monthType))
youthMonthType$Month <- as.Date(youthMonthType$Var1)
head(youthMonthType)
```

	Var1	Var2	Freq	Month
1	2014-07-01 Theft and Related Offences		393	2014-07-01
2	2014-08-01 Theft and Related Offences		369	2014-08-01
3	2014-09-01 Theft and Related Offences		343	2014-09-01
4	2014-10-01 Theft and Related Offences		367	2014-10-01
5	2014-11-01 Theft and Related Offences		369	2014-11-01
6	2014-12-01 Theft and Related Offences		348	2014-12-01

Questions of Interest

For youth crimes, we are interested in comparisons between different types of crime.

- What is the most common type of crime?
- What is the least common type of crime?

Amongst the three most common crimes:

- How much more common is number 1 vs number 2 vs number 3?

We will also make a specific comparison between “Public Order Offences” and “Dangerous or Negligent Acts Endangering Persons”:

- Which is more common overall?
- What is happening over time for these types of crime?

Data Visualisations

1. The data visualisation below shows a pie chart of the number of incidents for each different type of crime. This was modelled on a figure from page 11 of the [Youth Justic Indicators Summary Report](#) (i.e., a data visualisation of this sort was published in an official government report).

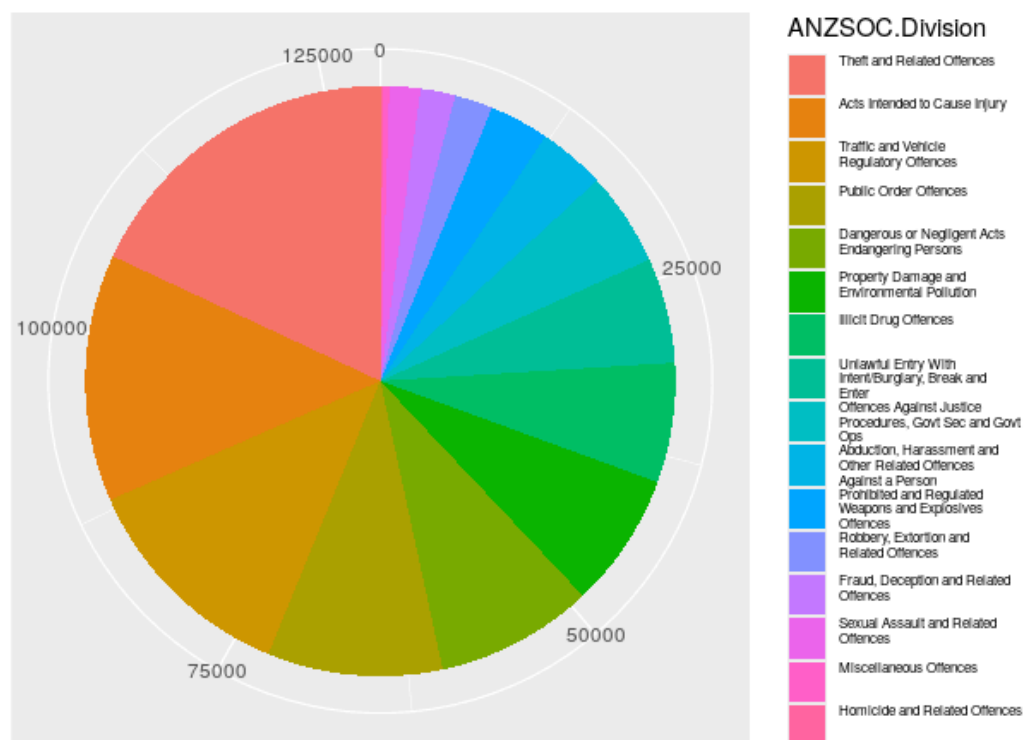
Write R code to produce this pie chart.

Comment on what this data visualisation tells us about the questions of interest.

Identify the visual channel or channels that are being used to represent the data values in this data visualisation and **comment** on whether they are appropriate channels.

Identify a perceptual problem with this data visualisation that is related to “contrast” effects. If you add a white border to the segments of the pie chart, does this fix the problem? Explain why or why not?

Overall Number of Incidents



2. **Write R code** to produce **two** different data visualisations of the data from the preceding question that make use of different visual channels: one should make use of at least one better visual channel and one should make use of at least one worse visual channel.

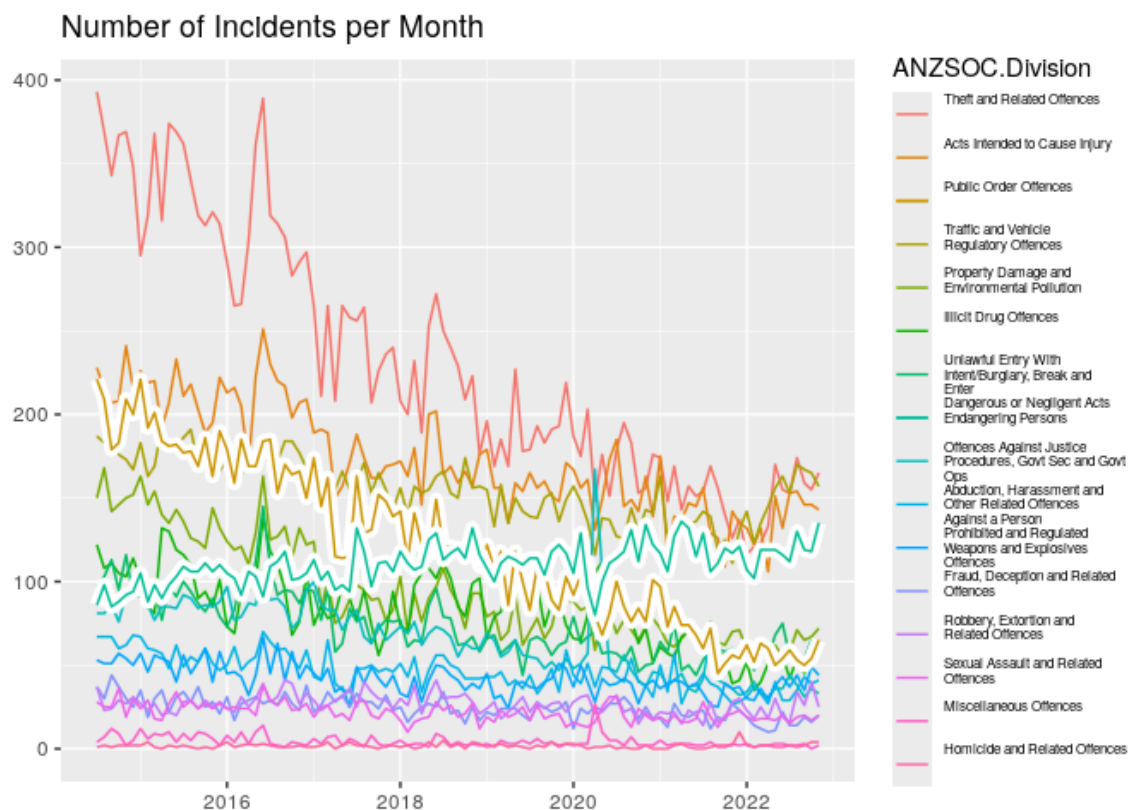
Comment on the changes that you have made and whether they make it easier or harder to answer the questions of interest.

3. The plot below shows the number of incidents per month for each type of crime, with two of the crimes highlighted by drawing a white line behind the normal (coloured) line.

Write R code to produce this data visualisation.

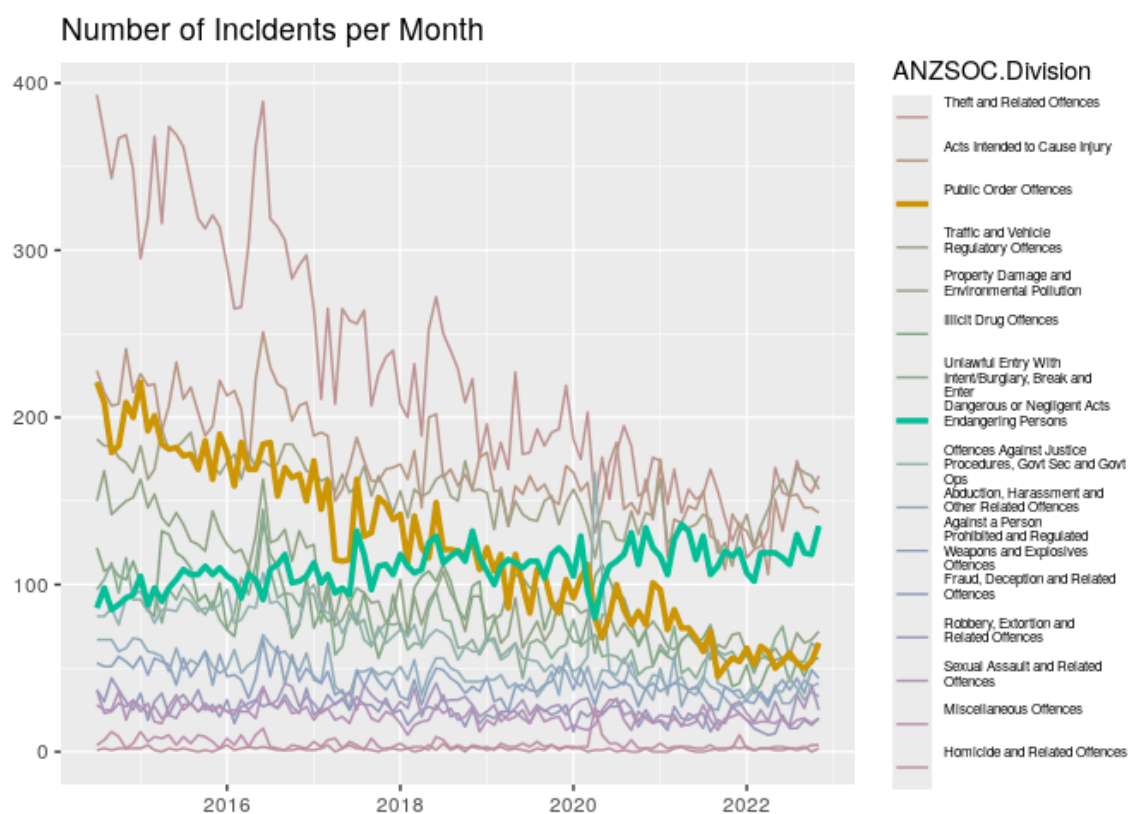
Comment on what this data visualisation tells us about the questions of interest.

Identify at least **one** example of **preattentive pop out** and at least **three** examples of **Gestalt Rules** in this data visualisation.



4. The plot below shows the number of incidents per month for each type of crime, with two of the crimes highlighted by desaturating the default colours for most lines, but retaining full saturation for the two lines (i.e., with a custom colour scale).

Write R code to produce this data visualisation.



5. **Write R code** to produce a version of the data visualisation from the previous question that shows what a viewer with deuteranomaly would see.

Comment on whether the data visualisation would be effective for a viewer with deuteranomaly.

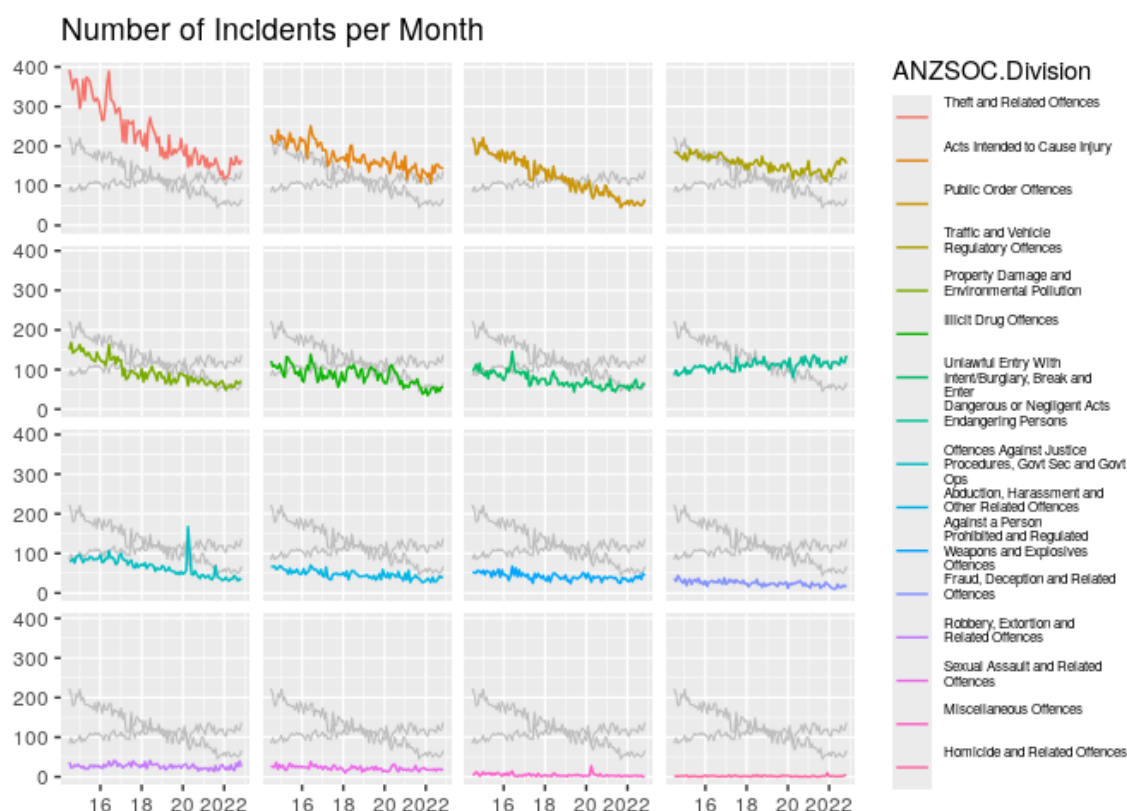
6. The plot below shows the number of incidents per month for each type of crime, with a separate panel for each type of crime.

NOTE that each panel has the Public Order and Dangerous or Negligent Acts data plotted in grey in addition to the relevant panel data (in colour). Also, the x-axis labels have been customised.

Write R code to produce this data visualisation.

Comment on what this data visualisation tells us about the questions of interest.

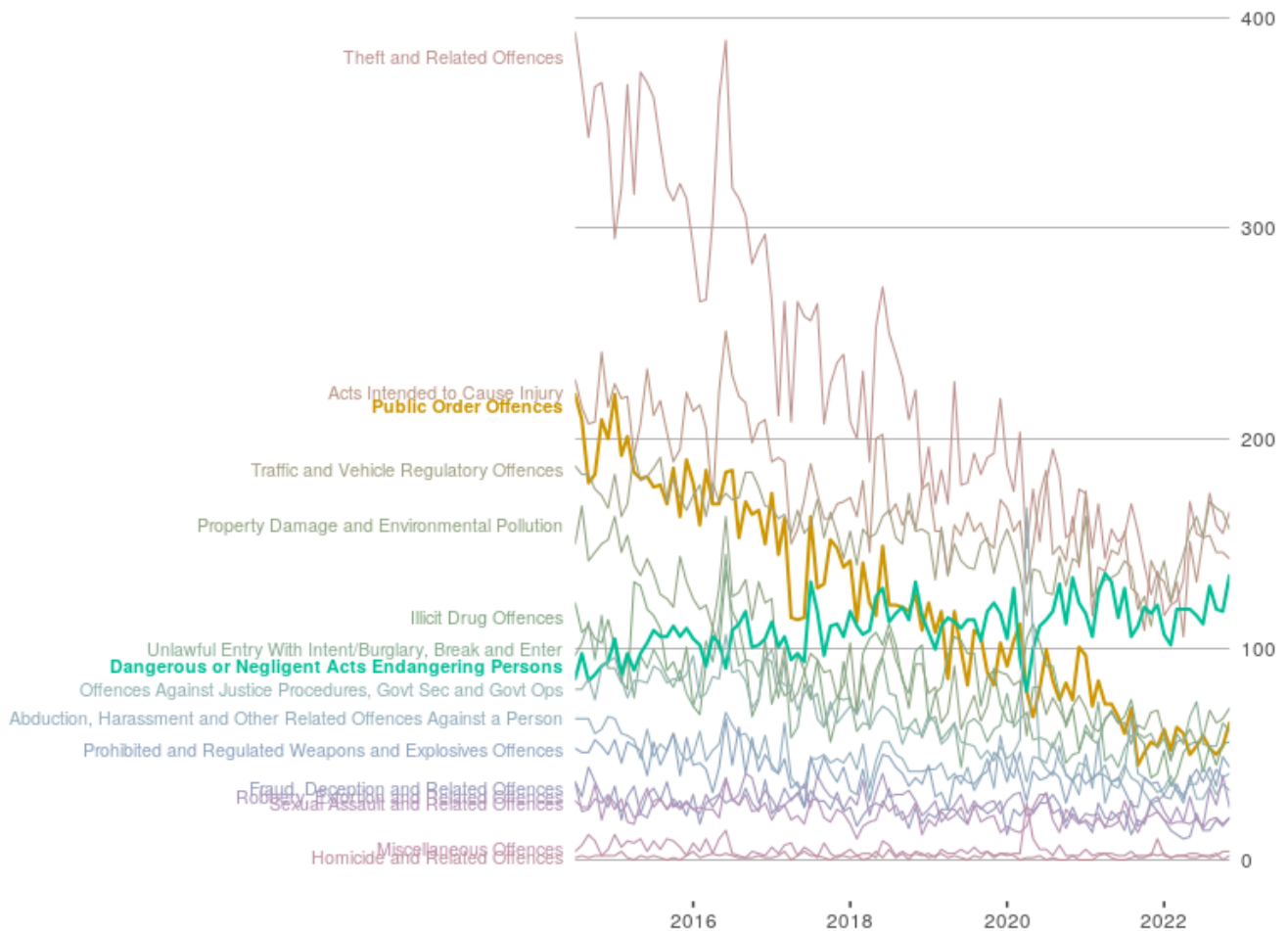
Identify the visual channel or channels that are being used to represent the data values in this data visualisation and **comment** on whether they are appropriate channels.



Challenge

7. No marks will be given for this question.

Can you produce the plot below? Does this improve on any of the previous plots in any way? What visual perception concepts are being used here?



The Report

Your submission should consist of a knitted R Markdown document, in HTML format, submitted via Canvas.

Your report should include:

- A brief description of the data and the question we are trying to answer.
- For each data visualisation, R code AND a brief text commentary.
- A brief overall summary.

Don't forget to also complete the Canvas Quiz!

Marking

Marks will be lost for:

- Plagiarism.
- Section of the report is missing.
- The summary is too short or does not make sense.
- Significantly poor R (or other) code.
- Overly verbose code, output, or commentary.



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).