Lab 1: Grammar of Graphics

The purpose of this lab is to practice producing data visualisations using the ‘ggplot2’ package and its Grammar of Graphics concepts.

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

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

```r
crime <- read.csv("nzpolice-proceedings.csv")
head(crime)
```

```
  Age.Lower Police.District                                                   ANZSOC.Division
1        15          Tasman                                     Acts Intended to Cause Injury
2        20   Auckland City Abduction, Harassment and Other Related Offences Against a Person
3        40   Auckland City Abduction, Harassment and Other Related Offences Against a Person
4        10   Auckland City                                     Acts Intended to Cause Injury
5        15   Auckland City                                     Acts Intended to Cause Injury
6        15   Auckland City                                     Acts Intended to Cause Injury
     SEX       Date
1 Female 2015-12-01
2 Female 2015-12-01
3 Female 2015-12-01
4 Female 2015-12-01
5 Female 2015-12-01
6 Female 2015-12-01
```

Each row contains information on a single incident that the Police handled. The `Age.Lower` column gives the lower bound of a 5-year age band of the offender, the `SEX` column gives the sex of the offender, and the `Date` column gives the year and month of the incident (all incidents are marked as occurring on the first day of the month). There are over 800,000 incidents recorded between 2014 and 2022.

Questions of Interest

Our main interest is in **trends over time in youth offending** (up to age 19), particularly at the end of 2021 and the beginning of 2022.

We are also interested in the a comparison of **youth offending versus adult offending** and any differences between **males and females**.

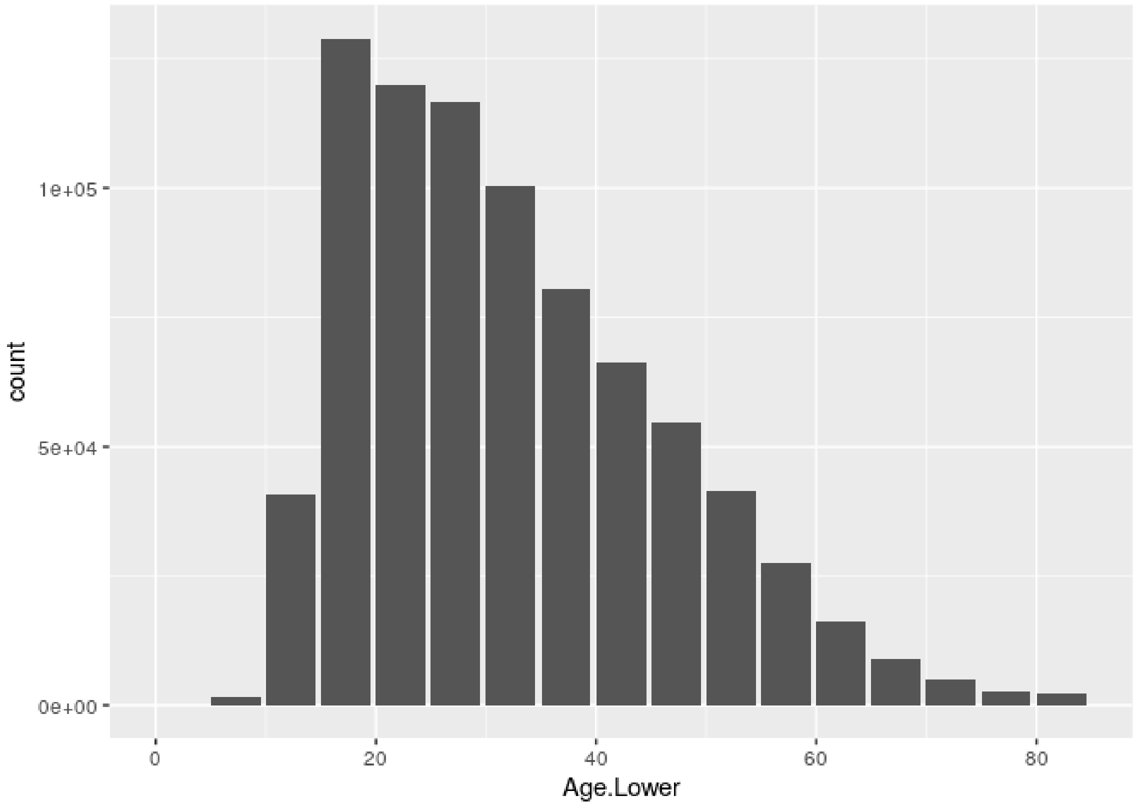
Data Visualisations

1. **Write R code** to produce a bar plot of the number of incidents in each age group.

   Identify the geoms and stats and aesthetic mappings that you are using in this plot.

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

   Note the detail that each bar is left-aligned with the lower bound of the age band. Reading the help page `?geom_bar` should reveal an argument that will help you to do that.



2. The following code creates a table of counts from the data.

```
crimeTab <- as.data.frame(table(crime$Age.Lower))
```

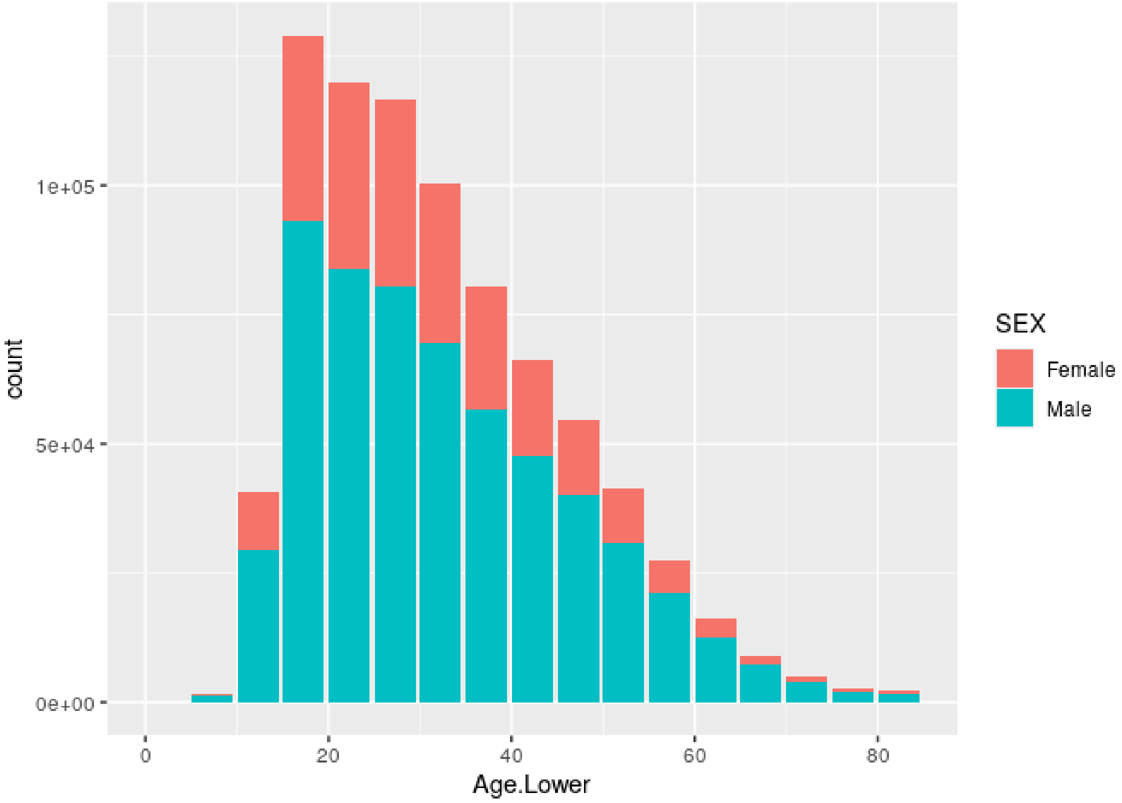
   Write R code that uses this `crimeTab` data frame to produce the same plot as in the previous question.

   Identify the geoms and stats and aesthetic mappings that you are using in this plot.

3. **Write R code** to produce a stacked bar plot of the number of incidents in each age group broken down by the sex of the offender. We are back to using the `crime` data frame in this question.

   Identify the geoms and stats and aesthetic mappings that you are using in this plot.

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



4. **Write R code** to produce three variations on the bar plot from the previous questions that use “dodge”, “identity” and “fill” positioning of the bars.

   Comment on what these data visualisations tell us about the questions of interest.

5. The following code creates a data frame containing the number of incidents in each month.

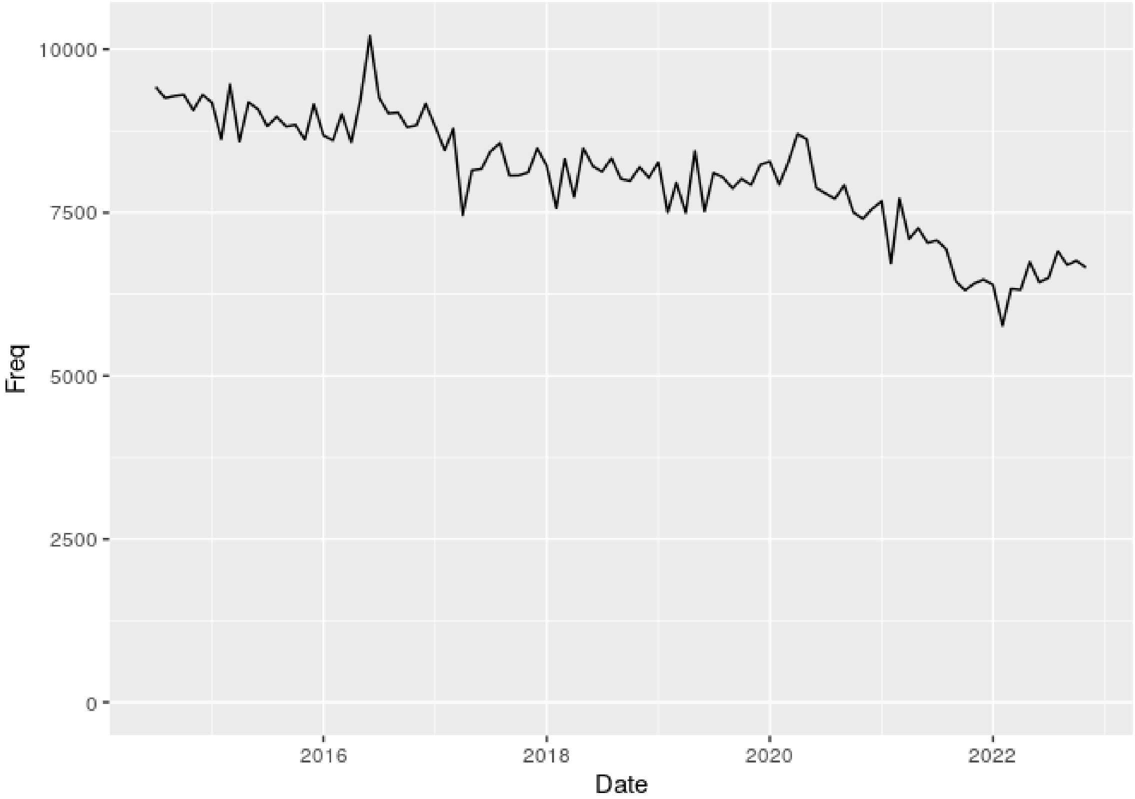
```
crimeTrend <- as.data.frame(table(crime$Date))
crimeTrend$Date <- as.Date(crimeTrend$Var1)
```

   Write R code to produce a plot of the number of incidents per month.

   Identify the geoms and stats and aesthetic mappings that you are using in this plot.

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

   Note the scale on the y-axis.



6. The following code creates a data frame containing the number of incidents per month broken down by the sex of the offender.

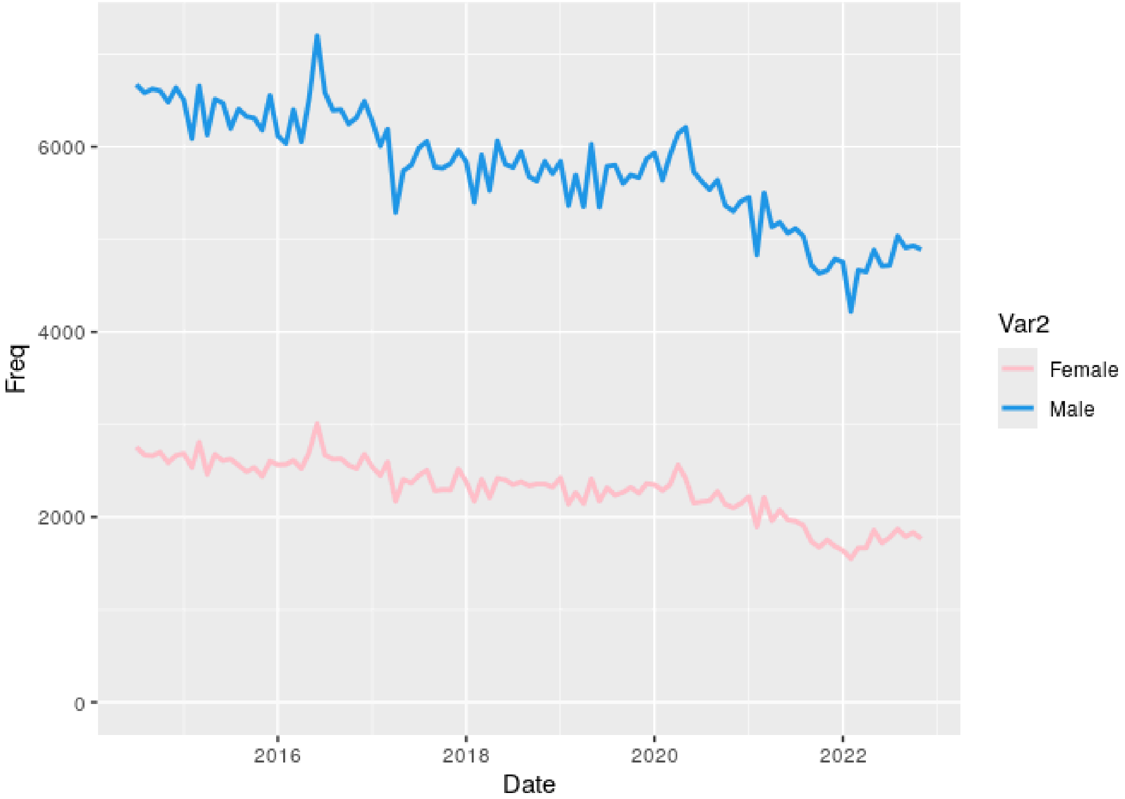
```
crimeTrendSex <- as.data.frame(table(crime$Date, crime$SEX))
crimeTrendSex$Date <- as.Date(crimeTrendSex$Var1)
```

   Write R code to produce a plot showing the number of incidents per month with separate lines for males and females.

   Identify the geoms and stats and aesthetic mappings that you are using in this plot.

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

   Note the colours of the lines and the thickness of the lines. The blue is the colour `4` in R and the pink is the colour `"pink"` in R.



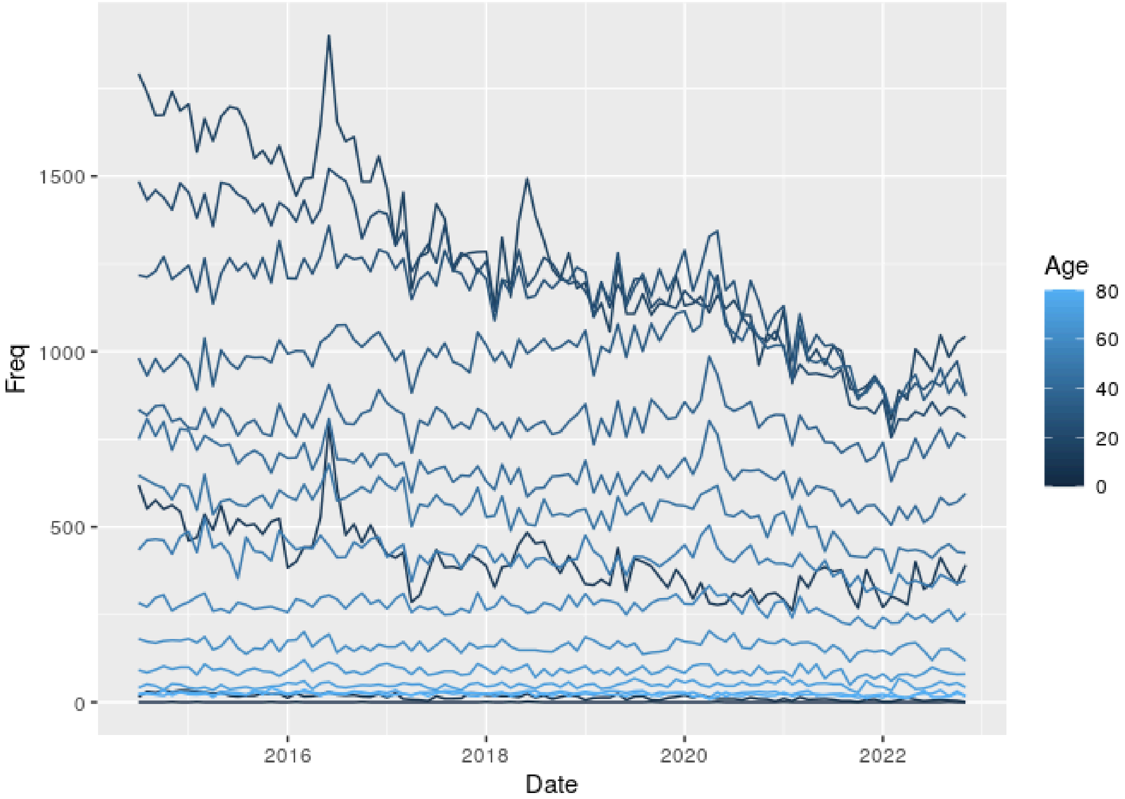
7. The following code creates a data frame containing the number of incidents per month for each age group.

```r
crimeTrendAge <- as.data.frame(table(crime$Date, crime$Age.Lower))
crimeTrendAge$Date <- as.Date(crimeTrendAge$Var1)
crimeTrendAge$Age <- as.numeric(as.character(crimeTrendAge$Var2))
```

Write R code that produces a plot of the number of incidents per month with a separate line for each age group.

Identify the geoms and stats and aesthetic mappings that you are using in this plot.

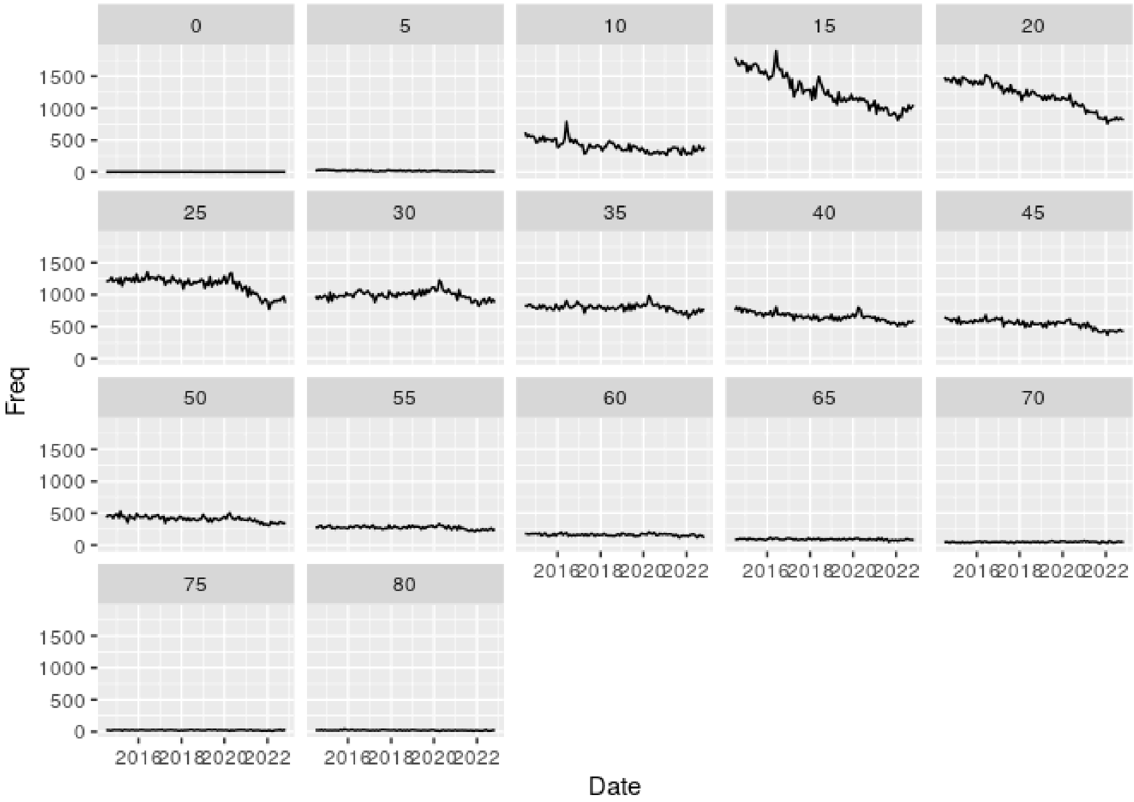
Comment on what this plot tells us about the questions of interest.



8. **Write R code** to produce a plot of the number of incidents per month, with a different facet for each age group.

   Identify the geoms and stats and aesthetic mappings that you are using in this plot.

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



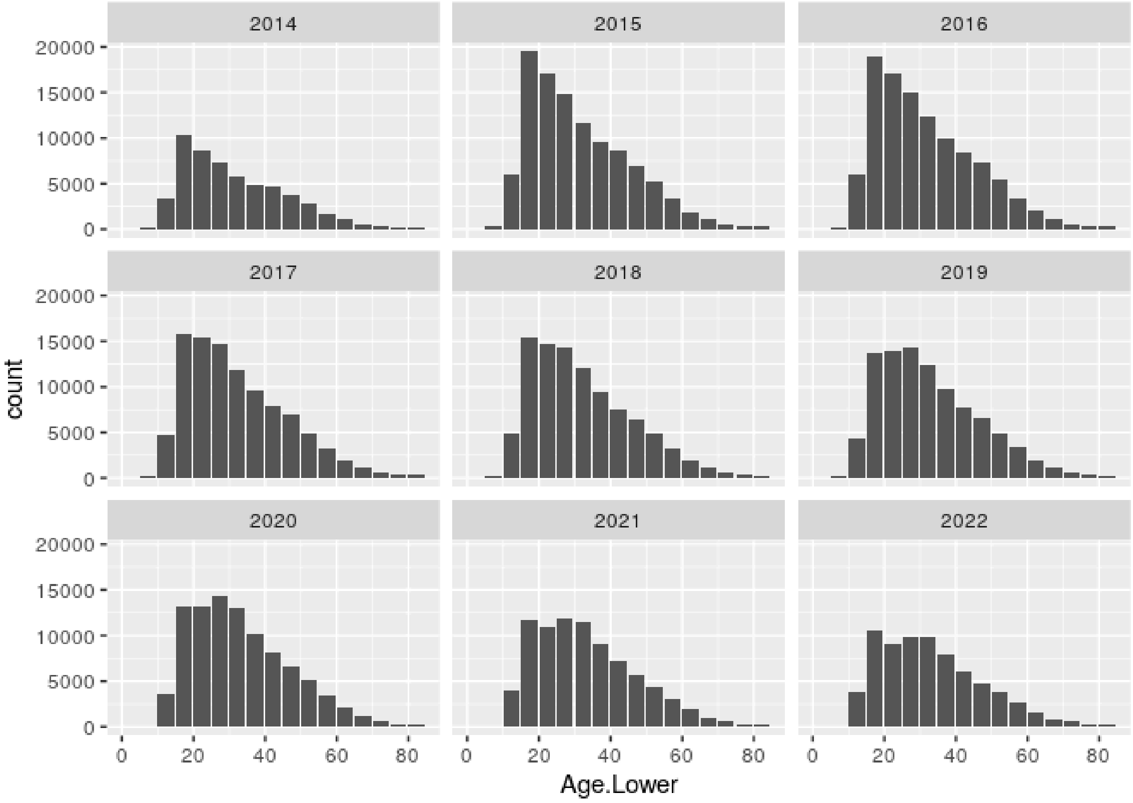
9. The following code adds a `Year` column to the original `crime` data frame.

```
crime$Year <- as.POSIXlt(crime$Date)$year + 1900
```

Write R code to produce a bar plot of the number of incidents in each age group, with a different facet for each year.

Identify the geoms and stats and aesthetic mappings that you are using in this plot.

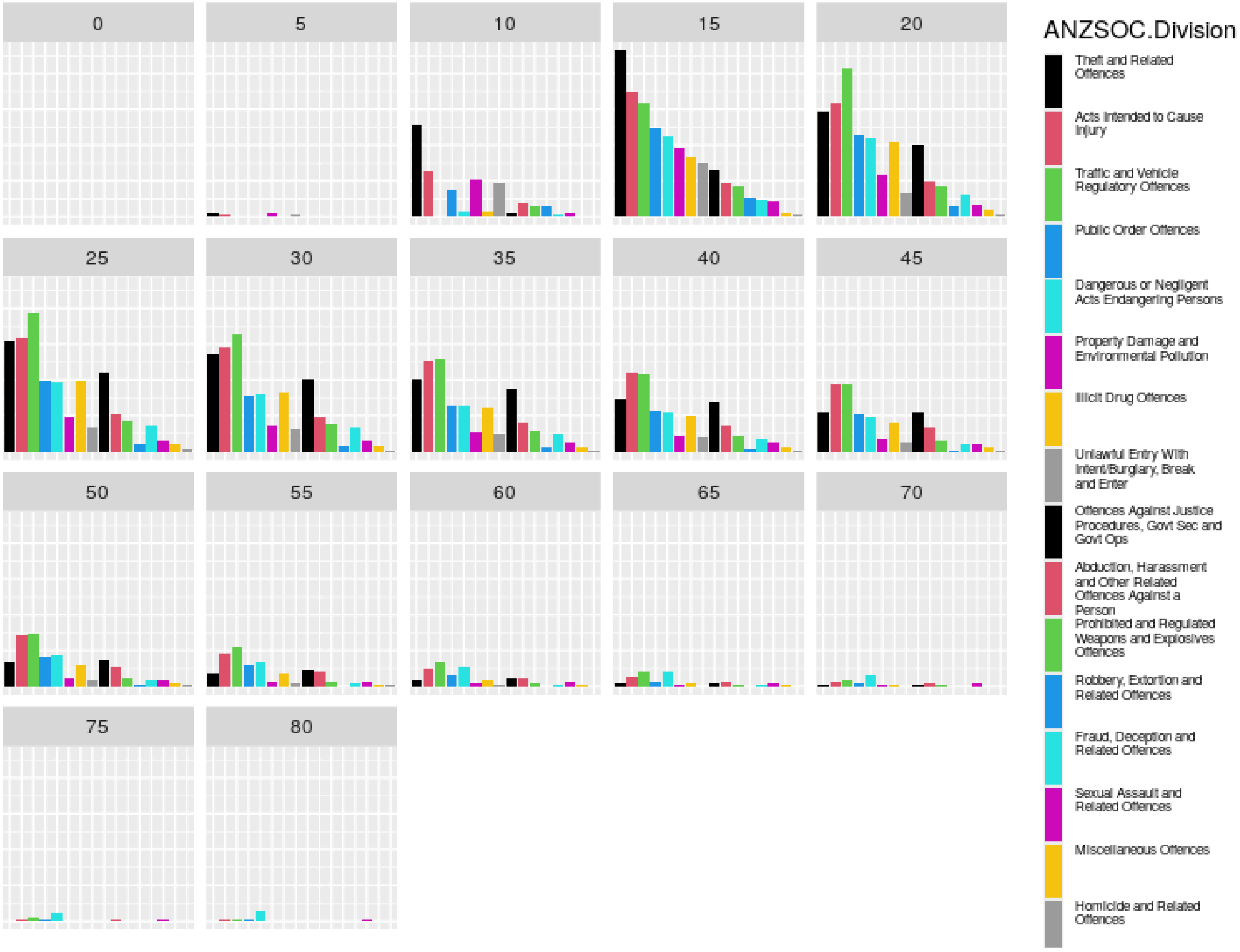
Comment on what this plot tells us about the questions of interest.



Challenge

10. **No marks will be given for this question.**

    Can you produce the data visualisation below? This shows the number of incidents for different types of crime within each age group. Can you see any interesting features?



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.