

Lab 2: Grid Graphics

The purpose of this lab is to practice producing data visualisations using the 'grid' package, either on its own or in combination with the 'ggplot2' package.

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()`.

```
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

Questions of Interest

We will focus on just youth crime, where we define "youth" as aged 15-19 inclusive.

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

We will look at the proportion of male versus female offenders in this age group and we will look at the trends over time in the number of incidents, broken down by type of crime.

- Does the proportion of male versus female offenders change over time?
- Can we see evidence of an increase or decrease in crime over time?
- Do the answers to these questions differ based on the type of crime?

Data Visualisations

1. The following code generates proportions of male versus female offenders. It also defines some colours that we will use in several places throughout the lab.

```
youthSex <- table(youth$SEX)/nrow(youth)
female <- "#E46C0A"
male <- "#0070C0"
```

Write R code that **only uses 'grid'** to produce a simple data visualisation of the proportion of male versus female offenders. This is a reproduction of a data visualisation from page 9 of the [Youth Justice Indicators Summary Report](#). Some important features to replicate are:

- the bars are the full width of the image and are 1cm high.
- the bars have no border and they are filled using the colours defined above.
- the text labels are 2mm in from either end of the bars.

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



2. **Write R code** that **only uses 'grid'** to draw the bars from the previous question again, but this time draw them within a 'grid' viewport that is 2cm narrower than the image (creating a 1cm margin on either side).

The code that draws the bars should be identical to what you used in the previous question (apart from code used to create the 'grid' viewport). The most efficient way to achieve that is to write a function that encapsulates the drawing code in the previous question and then just call that function again in this question.



3. The following code calculates proportions of male versus female offenders per year.

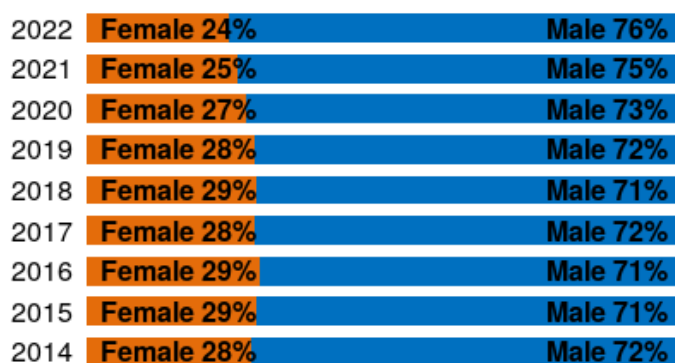
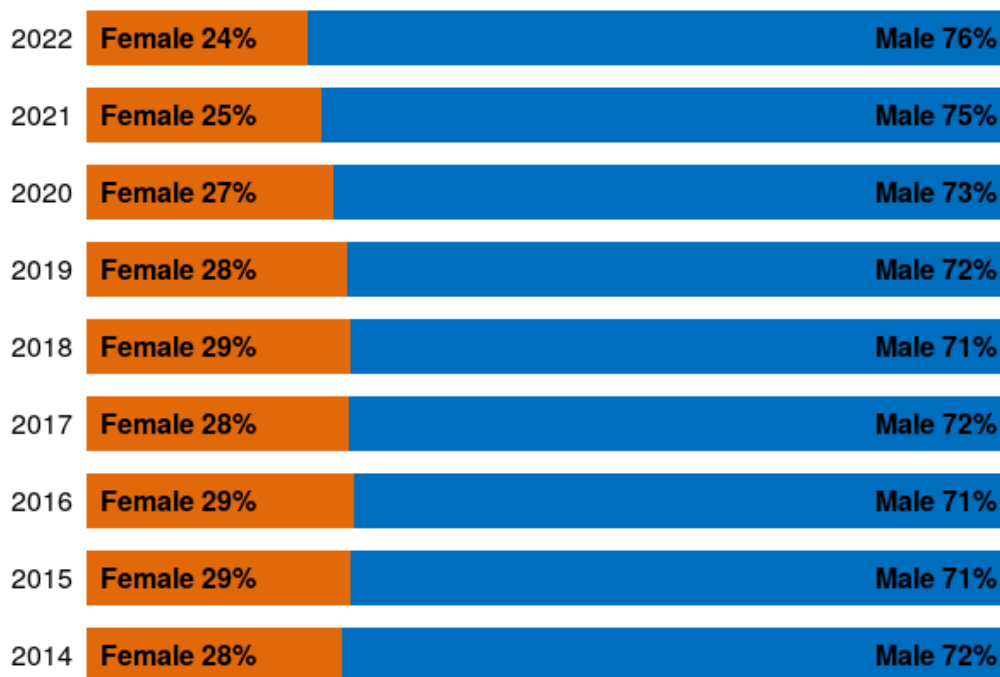
```
youthSexYear <- t(apply(table(substr(youth$Date, 1, 4), youth$SEX), 1,
  function(x) x/sum(x)))
```

Write R code that **only uses 'grid'** to draw an array of bars showing the proportion of male versus female offenders broken down by year. Your code should create at least one viewport to draw the bars within. This viewport should have a margin around the outside, including a wider one on the left for the year labels.

- The left margin is 1in (to allow for the year labels); all other margins are 1cm.
- The year labels are right-justified 2mm to the left of the bars.
- The year labels should be generated from data values (not just typed out explicitly).
- The height of the bars is 8% of the height of the viewport that they are drawn within.

You must draw **two versions** of the plot: one at the default image size and one within an image that is 5in wide and 3in high (as shown below). Again, the most efficient way to achieve this is to write a function that does the drawing and then call that function in two separate code chunks with different `fig.width` and `fig.height` settings.

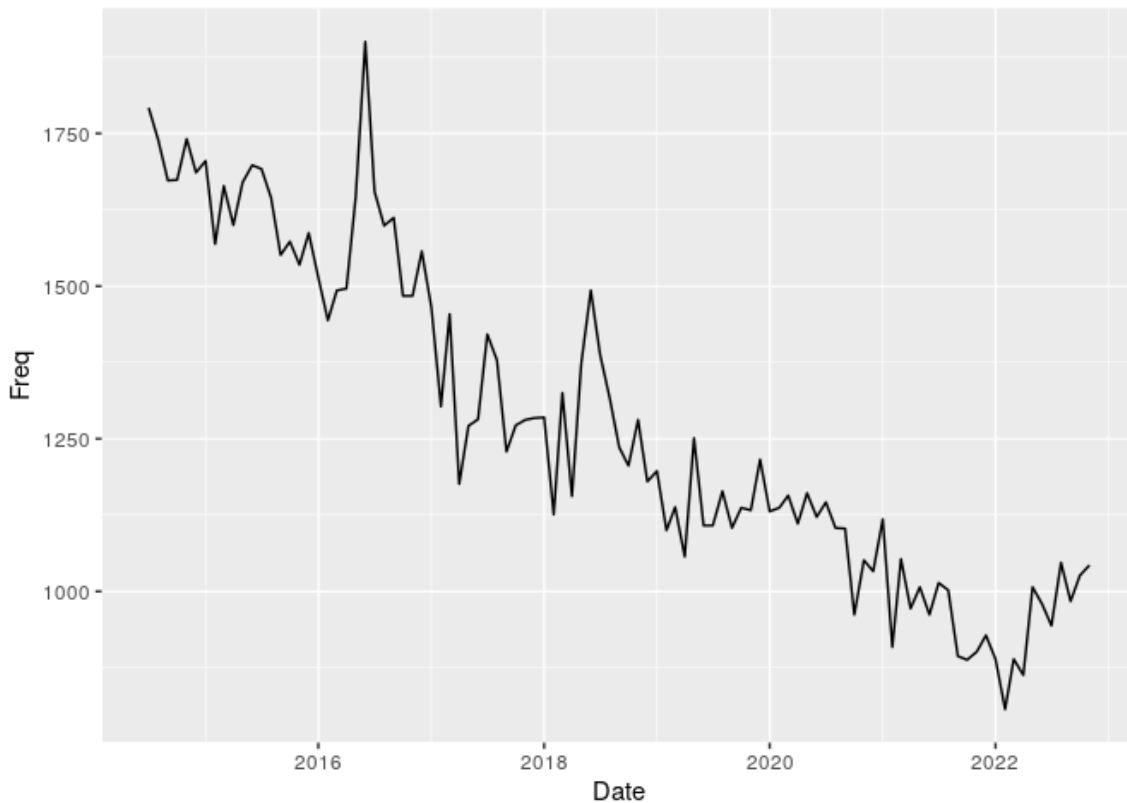
Comment on what this data visualisation tells us about the questions of interest.



Combining 'grid' with 'ggplot2'

We now turn our attention to combining 'grid' output with 'ggplot2' output. We have previously seen the plot below, which shows that the number of incidents shows a downward trend over time with a

slight suggestion of an increase since the start of 2022. We will look in more detail at these trends by breaking crimes down by type of crime.



4. The following code generates counts of incidents per month for each type of crime.

```
youth$Abbrev <-
  gsub(",", "",
        unlist(lapply(strsplit(as.character(youth$ANZSOC.Division), " "),
                        function(x) x[1])))
youthTrendType <- as.data.frame(table(youth$Date, youth$Abbrev))
youthTrendType$Date <- as.Date(youthTrendType$Var1)
## Order crime types
types <- table(youth$Abbrev)
newlevels <- names(types)[order(types, decreasing=TRUE)]
youthTrendType$Type <- factor(youthTrendType$Var2, levels=newlevels)
```

We have abbreviated the crime type labels so that the facet labels are not too long; the output below provides a “legend” to decode facet labels back to full crime type labels.

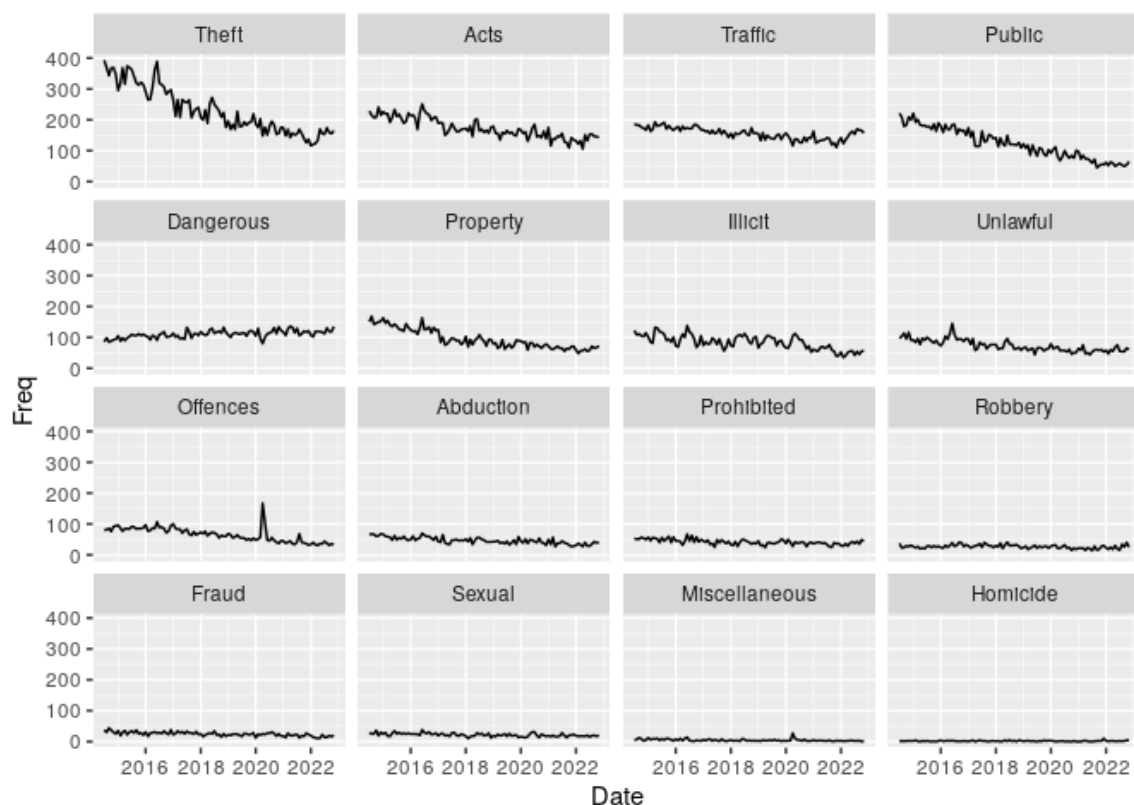
```
print(unique(youth[c("Abbrev", "ANZSOC.Division")]),
       right=FALSE, row.names=FALSE)
```

Abbrev	ANZSOC.Division
Acts	Acts Intended to Cause Injury
Dangerous	Dangerous or Negligent Acts Endangering Persons
Illicit	Illicit Drug Offences

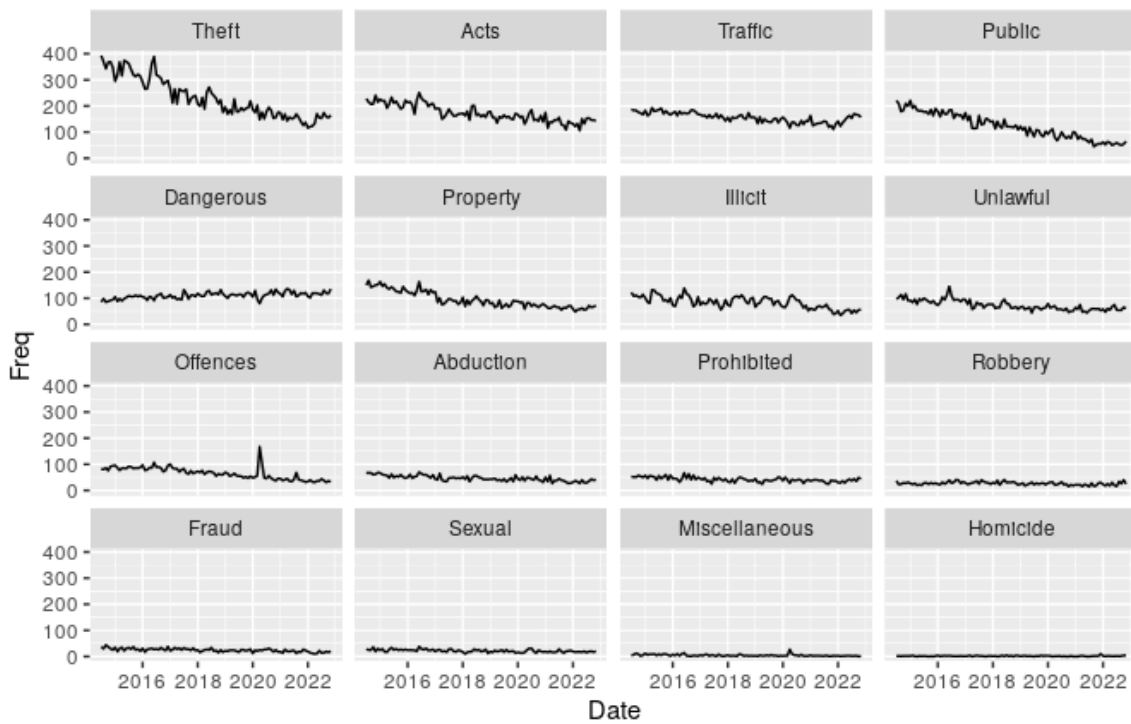
Offences	Offences Against Justice Procedures, Govt Sec and Govt Ops
Property	Property Damage and Environmental Pollution
Public	Public Order Offences
Robbery	Robbery, Extortion and Related Offences
Theft	Theft and Related Offences
Traffic	Traffic and Vehicle Regulatory Offences
Prohibited	Prohibited and Regulated Weapons and Explosives Offences
Fraud	Fraud, Deception and Related Offences
Sexual	Sexual Assault and Related Offences
Unlawful	Unlawful Entry With Intent/Burglary, Break and Enter
Abduction	Abduction, Harassment and Other Related Offences Against a Person
Miscellaneous	Miscellaneous Offences
Homicide	Homicide and Related Offences

Write R code to produce a line plot of the number of incidents over time with a facet for each type of crime.

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



5. **Write R code** that produces the plot from the previous question *combined with* the bars from the first question.

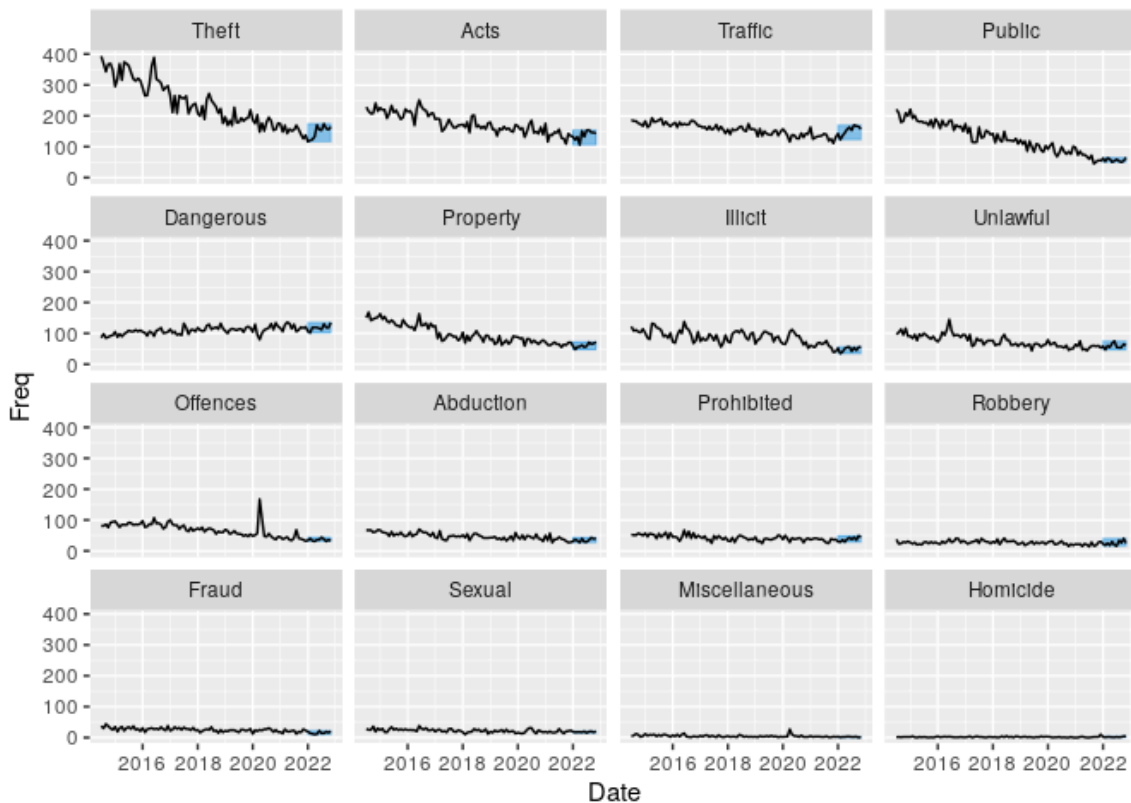
Female 28%**Male 72%**

6. **Write R code** that uses the 'gggrid' package to add a semitransparent blue rectangle to each panel that highlights the recent values. The rectangle should be the same for each panel, it should start 85% of the way across the panel and end at the right edge of the panel.



7. **Write R code** that uses 'gggrid' to add semitransparent blue rectangles to each panel, but this time, the rectangle will be different for each panel: it should be restricted to the range of x- and y- values for data from 2022-01-01 onwards.

Comment on whether this makes it easier or harder to answer the questions of interest.



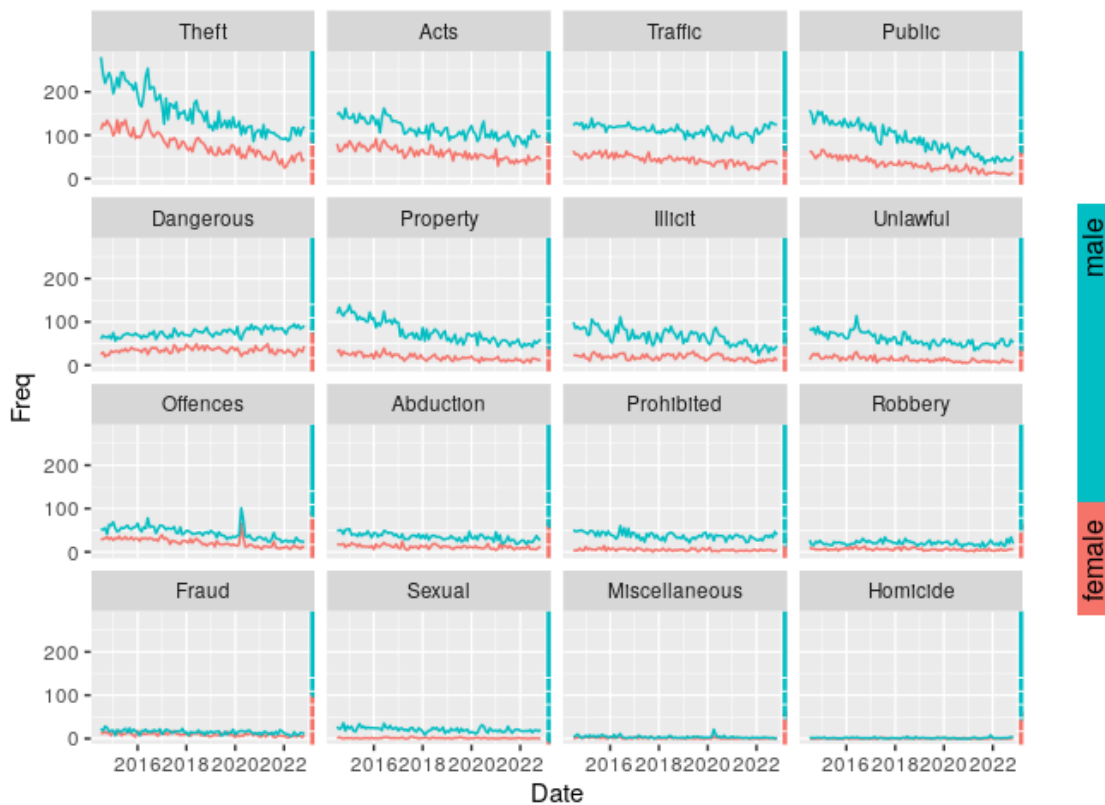
Challenge

8. No marks will be given for this question.

Can you produce the data visualisation below? This shows a line plot of the number of incidents over time, broken down by the type of crime and further broken down by the sex of the offender. A mini bar has been drawn on the right side of each panel that represents the proportion of male versus female offenders within each crime type. Horizontal lines are drawn on these bars at 0.1 to 0.5 in steps of 0.1 to assist with comparisons between panels. Finally, a custom legend has been drawn to the right of the panels.

The following data preparation may help to get you started.

```
youthTrendTypeSex <- as.data.frame(table(youth$Date, youth$Abbrev,
                                          youth$SEX))
youthTrendTypeSex$Date <- as.Date(youthTrendTypeSex$Var1)
youthTrendTypeSex$Type <- factor(youthTrendTypeSex$Var2, levels=newlevels)
```



The Report

Your submission should consist of an R Markdown document, submitted via Canvas. You should write your document so that I can process it on my computer without any manual intervention. For example, do not include any calls to `setwd()` or `file.choose()`. You should write code that assumes that data files are in the current working directory. Also submit a processed version of your R Markdown document (an HTML document) in case I cannot process your document on my computer.

Your report should include:

- A brief description of the data.
- For each data visualisation, R code AND a brief text commentary.
- A brief summary.

Marking

Marks will be lost for:

- R Markdown file is missing.
- R Markdown file does not run.
- Processed file (HTML) is missing.
- 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/).