Climate Data Visualisation in R - Worksheet

Year 9-10 | Digital Technologies | Unit: Organise, visualise and analyse | Topic: Data

Australian Curriculum Code: Analyse and visualise data to create information and address complex problems, and model processes, entities and their relationships using structured data **(ACTDIP037)**. This resource is suitable for use by teachers who have chosen to use R to help teach the Digital Technologies curriculum.

Brief introduction to this worksheet

The worksheet consists of two sections, **Section 1** and **Section 2**.

Section 1: Droughts, and intense rainfall are killing Australian trees and plants. **TERN**'s on-the-ground ecosystem surveys help map out where they're dying, and eventually spot changes to ecosystem diversity. In this section, students will interpret graphs of *mean annual precipitation*, *mean annual temperature*, *growth forms* of the most abundant Australian native plants, and whether or not they have been found *alive*. Students will infer from the graphs local climate conditions where certain types of plants are more likely to die.

- Mean annual precipitation is the average amount of rain in mm a location receives in a year based on the 1976 - 2005 reference period.
- Mean annual temperature is the mean daily temperature recorded at a location, averaged for all days over 1976-2005.
- **TERN** is Australia's land ecosystem observatory. It measures key terrestrial ecosystem attributes over time from hundreds of representative locations. Ecosystem field surveys provide valuable data on the environment and how it is changing over time. Scientists analyse ecosystem survey data to detect and interpret changes in land ecosystems and, in turn, inform policy and management.

Section 2: Students will use the data visualisation package **ggplot2** in **R** to visualise **TERN**'s ecosystem survey data. Specifically, they will use different types of graphs to summarise the species and growth forms of various Australian native plants encountered on a surveyed plot.

Prior knowledge

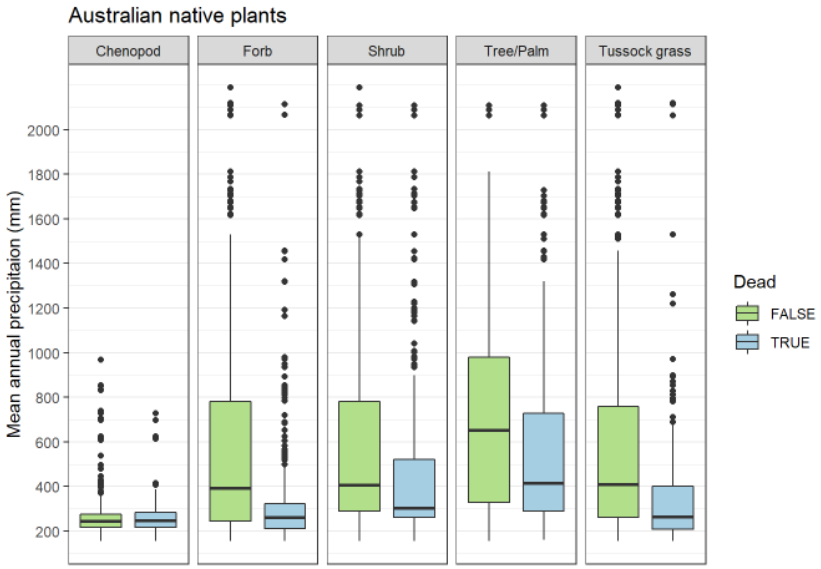
Students are expected to have worked before in the **RStudio** environment, and to be familiar with the **R** package **ggplot2** and the basic arguments of **ggplot()** functions.

Contact Sanaa Hobeichi at s.hobeichi@unsw.edu.au to request the answers to these questions

Section 1: Analysing graphs

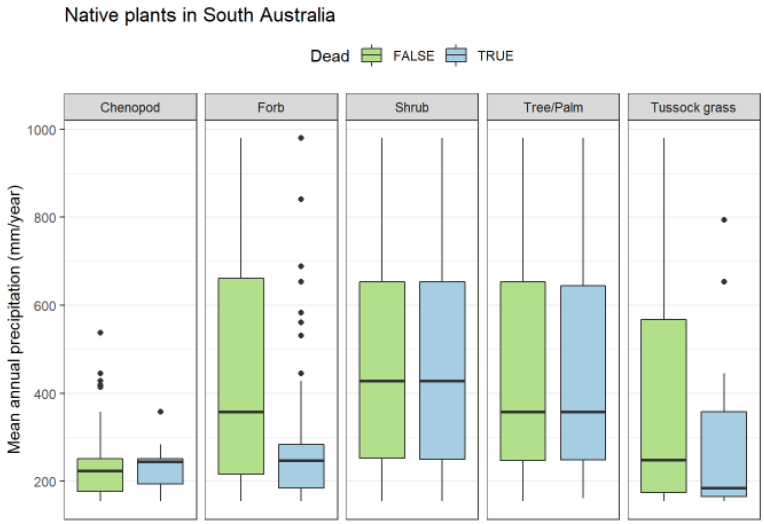
Q1. This graph summarises information about abundant plants' growth forms from more than 700 survey plots across Australia. Plants growth forms include:

Chenopod, i.e. salt- and drought- tolerant shrublands;
Forb, i.e. flowering, nongrassy, herbaceous plants;
Tussock grass, i.e. is grass that grows in clumps or tufts;
Shrub, i.e. quick-growing and drought-tolerant plants, some look like trees but have multiple stems; and
Palms and *trees*.



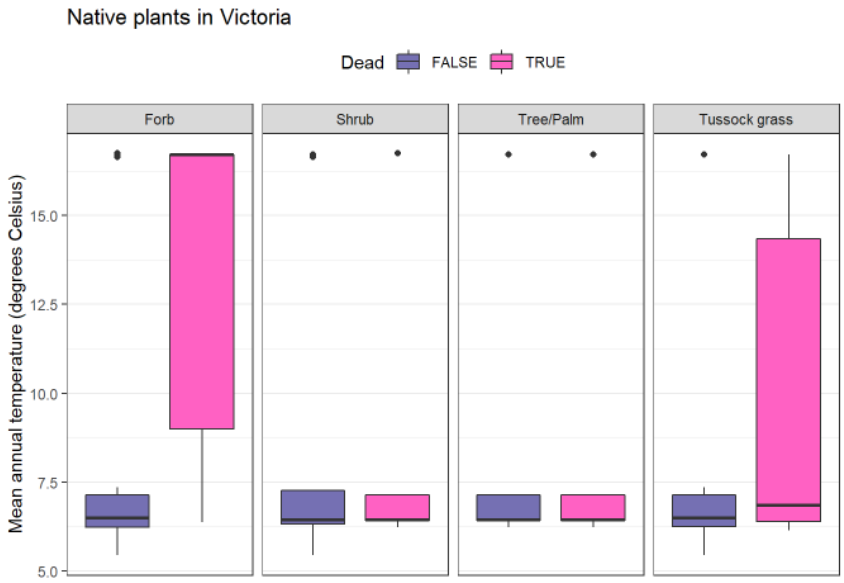
Which of the following statements is correct? There may be multiple correct answers.
A. Half of the dead Forbs are observed in locations that receive less than 250 mm of rain per year
B. Half of the dead Shrubs are observed in locations that receive less than 300 mm of rain per year
C. Tussock grass is likely to die if it receives less 250 mm of rain per year

Q2. This graph summarises information from plots surveyed in South Australia



If less than 300 mm of rain is received, what plant form(s) is(are) likely to die?
A. Forb
B. Tussock grass
C. Forb, Shrub, Tree/Palm and Tussock grass

Q3. This graph summarises information from plots surveyed in Victoria.



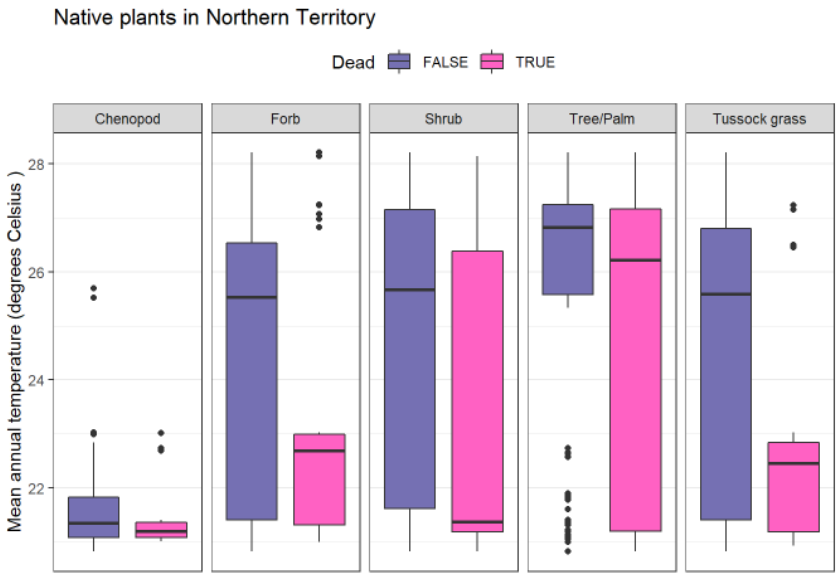
Which plant form(s) is (are) found dead on plots where the mean annual temperature exceeds 7.5 degrees Celsius?

A. Shurb and Tree/Palm

B. Forb and Tussock grass

C. Forb

Q4. This graph summarises information from plots surveyed in Northern Territory.



Which of the following statements is correct? There may be multiple correct answers.

A. Most chenopods grow in regions where mean annual temperature is below 23 degrees Celsius

B. Forbs do not tolerate temperatures above 23 degrees Celsius

C. The surveyed plants are more likely to die in locations where the mean annual temperature is low than in locations where it is high

Section 2: Data Visualisation
Prepare the data

1. Download TERN's ecosystem data

TERN's ecosystem data is available from the webpage https://www.tern.org.au/educational-resources/. Click on ***Data spreadsheet*** to download the data **TERN_Data_for_Students_FINAL-1.xlsx**.

The Excel file contains several spreadsheets, each described in the document https://www.tern.org.au/wp-content/uploads/TERN_data_for_students.pdf. In this worksheet, you will use the **Point intercept data** sheet which contains observations on the vegetation and soil cover across over 700 surveillance plots from various Australian natural environments and vegetation communities. In **Section 1**, you will analyse graphs that summarise the local climate (temperature and precipitation) and the condition (dead/not dead) of 5 most abundant growth forms of plants including Chenopod, Forb, Shrub, Tree/Palm and Tussock, while in **Section 2**, you will summarise the encountered species of plants on a single survey plot and their growth forms.

2. Setup your working directory

- Create a sub-directory named **Data** in your working directory .
- Copy the downloaded file **TERN_Data_for_Students_FINAL-1.xlsx** in folder **Data/**.
- In **RStudio**, use the menu to change your working directory under **Session > Set Working Directory > Choose Directory ...**, and navigate to your working directory .

3. Load the data

```
#Library for reading Excel files
library("readxl")

#Read spreadsheet 'Point intercept data' from TERN_Data_for_Students_FINAL-1.xlsx in
data frame 'df_plant'.
df_plant <- read_excel("Data/TERN_Data_for_Students_FINAL-1.xlsx", sheet = "Point
intercept data")

#Display the names of the columns of data frame df_plant
colnames(df_plant)
```

```
##  [1] "plot_location_name"             "visit_start_date"
##  [3] "herbarium_determination"        "CountOfherbarium_determination"
##  [5] "Vernacular Name"                "growth_form"
##  [7] "dead"                           "in_canopy_sky"
##  [9] "CountOfpoint_number1"           "Introduced species"
```

Narrow *df_plant* on survey data from a single plot named **NSABHC0016**, and from 3 variables of interest: **herbarium_determination**, **CountOfherbarium_determination**, and **growth_form**. Note that column **herbarium_determination** contains the name of identified plant species, **CountOfherbarium_determination** is the number of observed plants, and **growth_form** is the plants' form of growth.

To make sense of the data, print a subset of the dataframe, and the unique growth forms encountered on this plot.

```
df_plant_nsw <- df_plant[df_plant$plot_location_name == "NSABHC0016",
c("herbarium_determination", "CountOfherbarium_determination", "growth_form")]

## Display a subset of the dataframe.

head(df_plant_nsw)
```

```
## # A tibble: 6 × 3
##   herbarium_determination CountOfherbarium_determination growth_form
##   <chr>                                            <dbl> <chr>
## 1 Sida intricata                                      21 Forb
## 2 Maireana turbinata                                  18 Chenopod
## 3 Maireana astrotricha                                17 Chenopod
## 4 Sclerolaena lanicuspis                              15 Chenopod
## 5 Sclerolaena divaricata                              13 Chenopod
## 6 Sclerolaena ventricosa                              10 Chenopod
```

```
## The growth forms encountered in this plot
unique(df_plant_nsw$growth_form)
```

```
## [1] "Forb"          "Chenopod"      "Tussock grass"
```

Let's see how much you know about the data visualization library **ggplot2** by answering questions Q1 through Q10

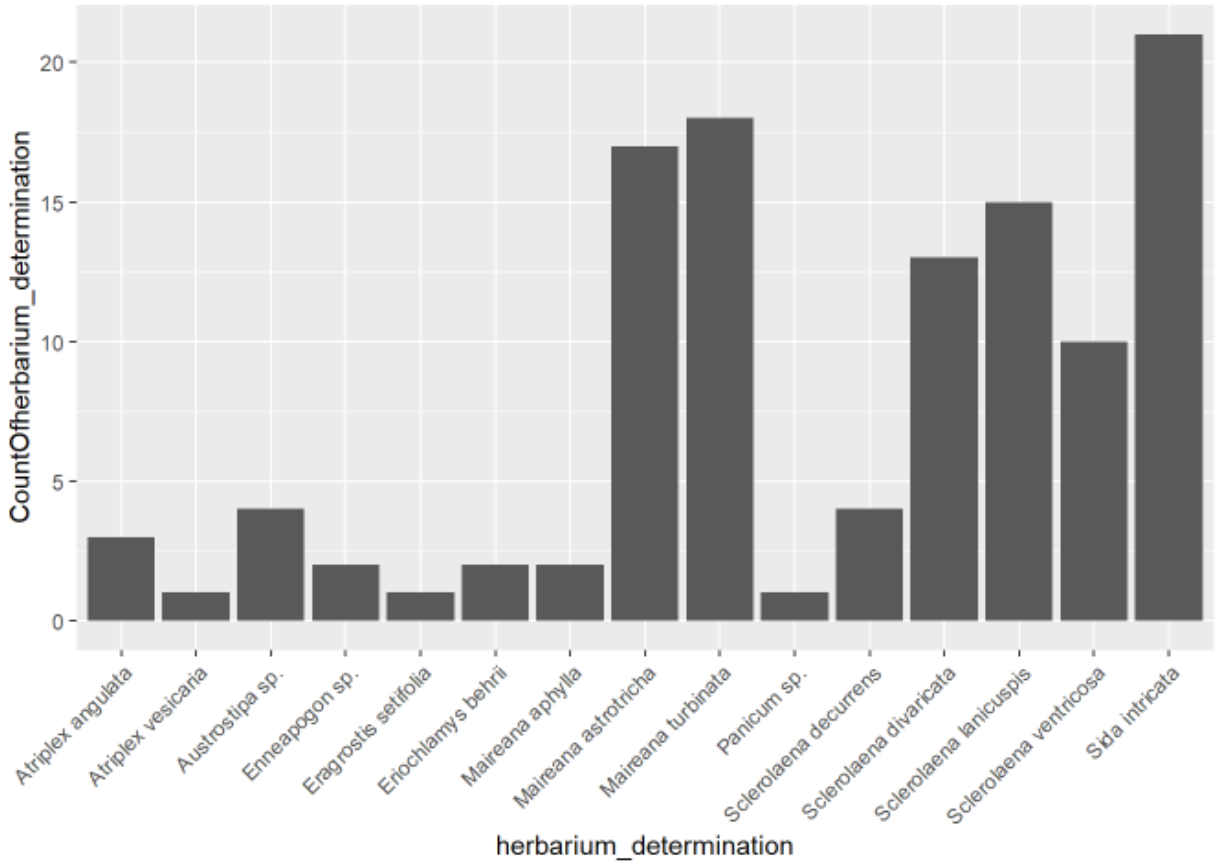
Q1. The script below plots a bar graph that shows the number of identified plants for each species. What happens if the command **p <- p + theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1))** is removed?

```
library("ggplot2")

p <- ggplot(data=df_plant_nsw, aes(x=herbarium_determination,
y=CountOfherbarium_determination))+
  geom_bar(stat="identity")

p <- p + theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1))

p
```



Q2. What will happen to the graph above if you add this line: **p <- p + coord_flip()**?

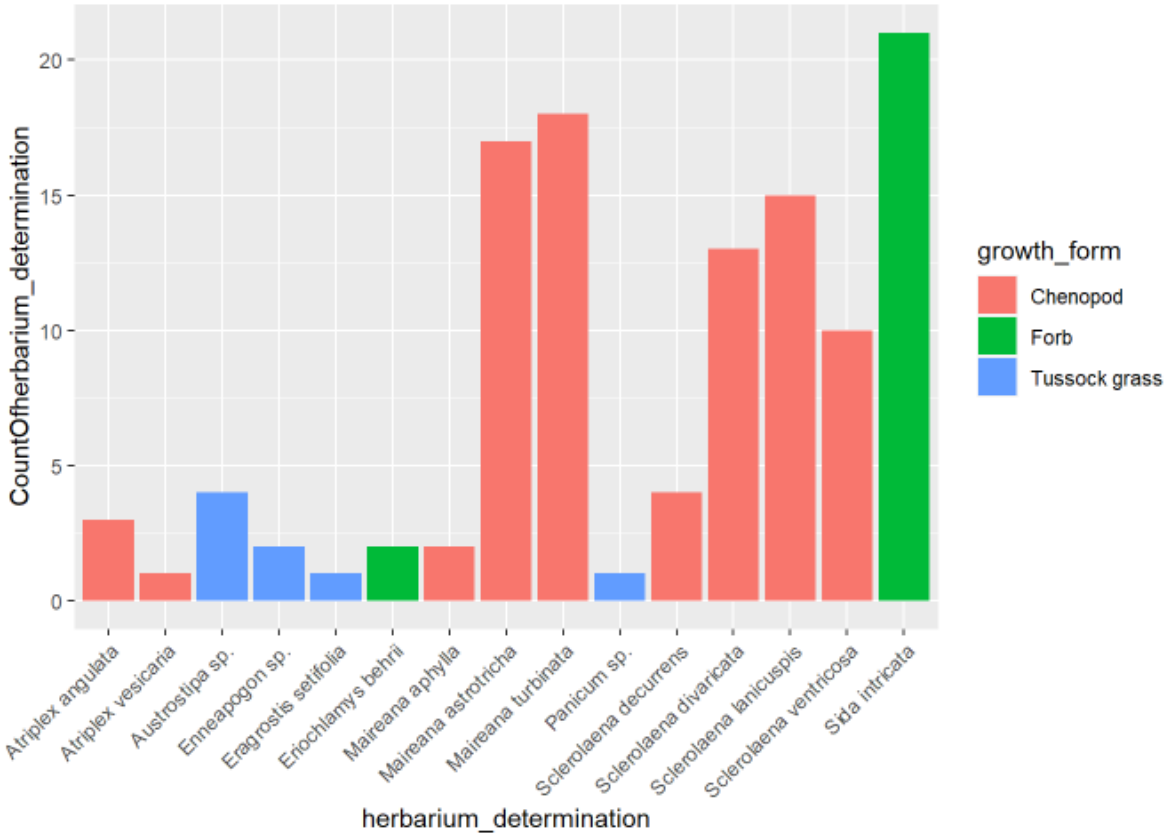
A. We will see horizontal bars instead of vertical bars

B. The bars will be displayed in order from highest to lowest

C. The y axis will be reversed

Q3. What is this script missing to produce the graph below?

```
ggplot(data=df_plant_nsw, aes(x=herbarium_determination,
y=CountOfherbarium_determination)) +
  geom_bar(stat="identity")+
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1))
```



A. Add *fill = growth_form* as an argument of the **ggplot()** function as shown in the script below

```
ggplot(data=df_plant_nsw, fill = growth_form,  aes(x=herbarium_determination,
y=CountOfherbarium_determination)) +
  geom_bar(stat="identity")+
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1))
```

B. Add *fill = growth_form* as an argument of **aes()** in the **ggplot()** function as shown in the script below

```
ggplot(data=df_plant_nsw, aes(x=herbarium_determination,
y=CountOfherbarium_determination, fill = growth_form )) +
  geom_bar(stat="identity", fill = growth_form)+
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1))
```

C. Add *fill = growth_form* as an argument of the **geom_bar()** function as shown in the script below

```
ggplot(data=df_plant_nsw, aes(x=herbarium_determination,
y=CountOfherbarium_determination, fill = growth_form )) +
  geom_bar(stat="identity", fill = growth_form)+
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1))
```

Q4. Which line of code does **not** add a title to the graph?

A.

```
p + guides(fill=guide_legend(title="NSW plot"))
```

B.

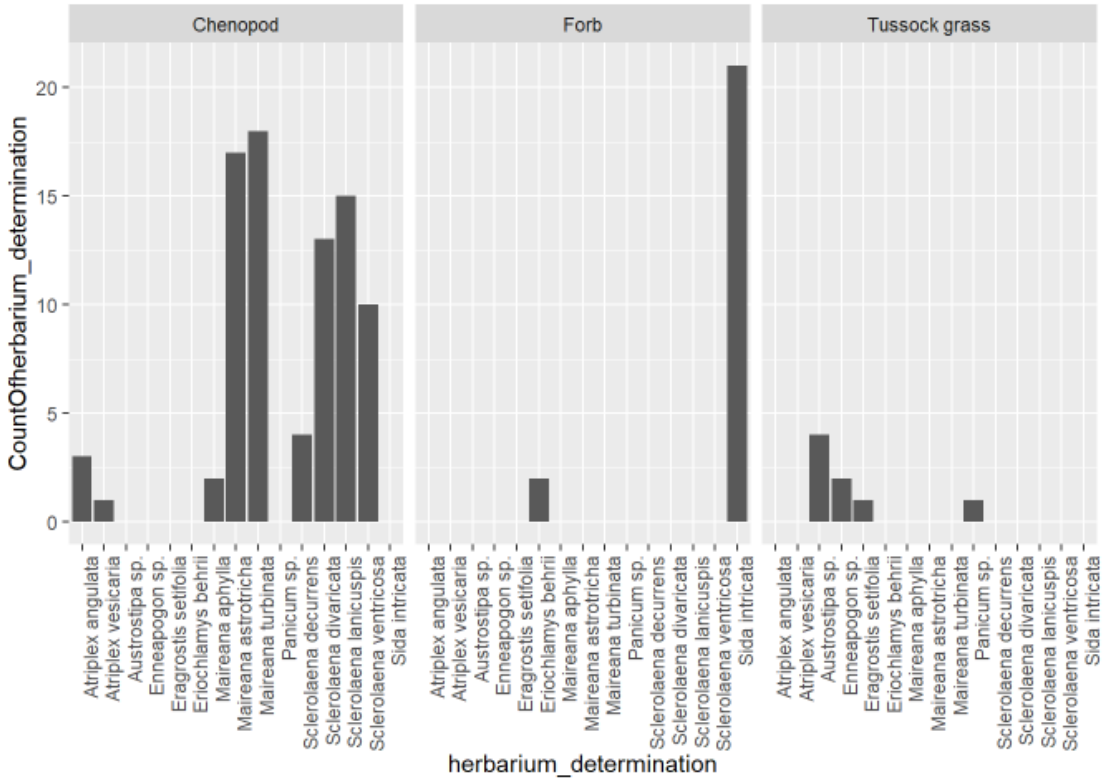
```
p + ggtitle("NSW plot")
```

C.

```
p + labs(title = "NSW plot")
```

Q5. What is this script missing to produce the graph below?

```
p = ggplot(data=df_plant_nsw, aes(x=herbarium_determination,
y=CountOfherbarium_determination)) +
  geom_bar(stat="identity")+
  theme(axis.text.x = element_text(angle = 90, vjust = 1, hjust = 1))
p
```



A.

```
p + facet_wrap(~growth_form )
```

B.

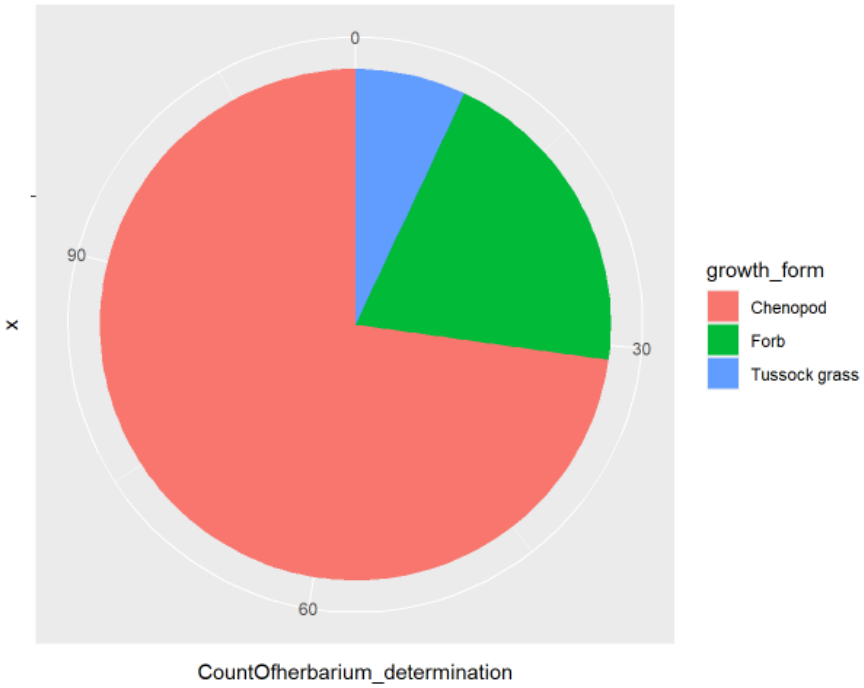
```
p + facet_wrap(~growth_form, ncol = 1 )
```

C.

```
p + facet_wrap(~CountOfherbarium_determination, nrow = 1 )
```

Q6. This script produces a pie chart that shows the number of plants *(CountOfherbarium_determination)* in each growth form *(growth_form)*.

```
ggplot(data=df_plant_nsw, aes(x = "", y = CountOfherbarium_determination, fill =
growth_form)) +
  geom_col() +
  coord_polar(theta = "y")
```



Which of the following new scripts will change the colors of *Chenopod*, *Forb* and *Tussock* grass to **"brown"**, **"darkorchid"**, and **"darkcyan"** respectively?

A.

```
ggplot(data=df_plant_nsw, aes(x = "", y = CountOfherbarium_determination, fill =
growth_form, color = c("brown", "darkorchid", "darkcyan") )) +
  geom_col() +
  coord_polar(theta = "y")
```

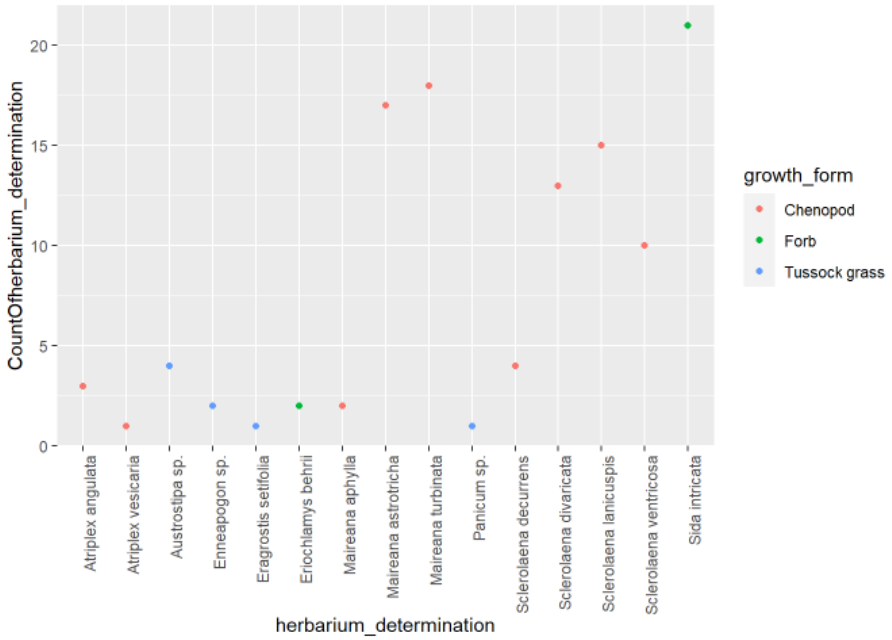
B.

```
ggplot(data=df_plant_nsw, aes(x = "", y = CountOfherbarium_determination, fill =
growth_form)) +
  geom_col(color = c("brown", "darkorchid", "darkcyan")) +
  coord_polar(theta = "y")
```

C.

```
ggplot(data=df_plant_nsw, aes(x = "", y = CountOfherbarium_determination, fill =
growth_form)) +
  geom_col() +
  coord_polar(theta = "y")+
  scale_fill_manual(values = c("brown", "darkorchid", "darkcyan"))
```

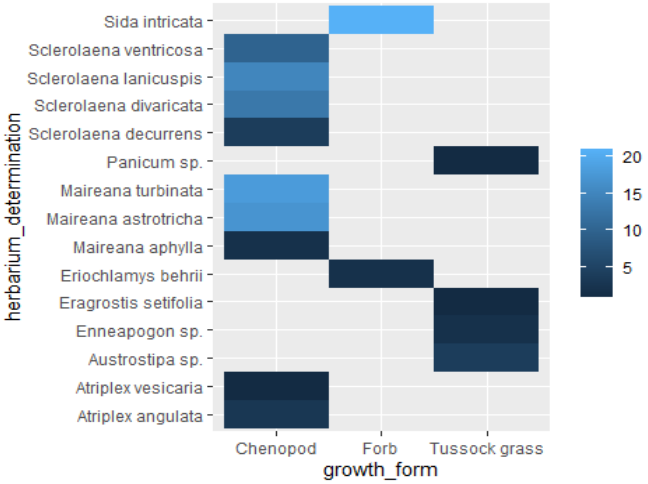
Q7. Write a script that produces the graph below



Q8. Which line of code removes the legends of a plot?
A. theme(legend.title = element_blank())
B. theme(legend.position="none")
C. labs(colour = NULL)

Q9. Which geometric layer should be used to create line graphs in ggplot2?
A. geom_segment()
B. geom_hline()
C. geom_line()

Q10. Which geometric layer was used to create this plot?
A. geom_boxplot()
B. geom_tile()
C. geom_density()



Acknowledgment

This is a teaching resource developed by the ARC Centre of Excellence for Climate Extremes (CLEX) and the Monash Climate Change Communication Research Hub (MCCCRH) with contributions from Dr Sanaa Hobeichi (CLEX); Amelia Pearson and Dr David Holmes (MCCCRH); and Cedric Le Bescont (Pymble Ladies' College).