



R Girls School Network



title: "Second R Markdown Lesson"

author: "write your name here"

date: "`r format(Sys.time(), '%d %B %Y')`"

output:

word_document: default

html_document: default

```
```{r setup, include=FALSE}
```

```
knitr::opts_chunk$set
```

```
library(tidyverse)
```

```
```
```

Before we start the lesson, type your name where it says "write your name here" on line 3. Make sure you type between the speech marks.

Knit the file. Can you see your name at the top of the output file? Close the output file by clicking on the X

Lesson Objectives

To become familiar with the layout of an R Markdown file

To practice editing the R code

To identify R commands needed to plot a bar chart

To produce an output file of my work using the 'knit' command

Success Criteria:

I can identify free text in the Markdown file

I can identify where to type the R code in the Markdown file

I can use some R commands to plot a bar chart

I can produce an output file of my work

Keywords:

- * R code chunk
- * free text
- * knit
- * variable

Practising editing the code in R Markdown

We are going to draw a bar chart showing the percentage of different blood groups in the UK population.

Then we are going to practice writing the title and the labels to explain the bar chart.

Finally, we are also going to practice changing the colour of the bars.

After making each change, click on the 'knit' command at the top of the pane. After a short wait you will see the output file of your work. Check that the changes you made to the code have come out on the output file as you wanted them. eg You may have added a title for your bar chart in the code chunk in the R Markdown file, but perhaps the title has not been printed in the output file. Go back to the R Markdown file to check you put speech marks around it.

Our data

This is the data we are going to use for the bar chart

| | | | | | |
|--------------------------|----|----|----|----|--|
| | | | | | |
| Blood Group | O | A | B | AB | |
| Percentage of population | 44 | 42 | 10 | 4 | |

: Table 1: Percentage of population with each blood group

Now we will find out how to enter the data from the above table into R.

Copy the following two lines of R code into chunk1 below. Make sure you paste the text between the top and bottom 3 back ticks.

```
blood_group <- c('O','A','B','AB')
```

```
pop_percentage <- c(44,42,10,4)
```

The first line tells R to store these letters in a variable called blood_group

The second line tells R to store the percentages in a variable called pop_percentage.

```
```${r chunk1}
```

```
```
```

To check what R has stored in the variable blood_group, type blood_group inside code chunk2 and select the arrow on the far right of the chunk. R will tell you what it has stored. It should be the four blood groups

```
```{r chunk2}
```

```
```
```

Now check what R has stored in the variable `pop_percentage` by typing `pop_percentage` in code chunk3. Click on the arrow on the far right of the chunk. R will tell you what it has stored. It should be the four percentage numbers.

```
```{r chunk3}
```

```
```
```

Let's check the percentage numbers add up to 100.

Copy `sum(pop_percentage)` into code chunk 4 and click on the little arrow on the right of the chunk. What number does R print for you? Is it 100?

```
```{r chunk4}
```

```
```
```

Now we will save the blood group and percentages data in a table (known as a tibble in R) called 'mydata'. This is so that the two variables can be combined into a single data table.

Copy the following code into R chunk5.

```
mydata <- tibble(blood_group, pop_percentage)
```

```
```{r chunk5}
```

```
```
```

Lets check that 'mydata' has the correct information in it.

Type `mydata` into code chunk6 and then click on the arrow on the right. Does the computer show you a little table with the correct information in?

```
```{r chunk6}
```

```
```
```

Drawing a bar chart using our blood group data

Now we have entered the data and checked that the data is correct we can tell R to plot a bar chart using these data.

Copy and paste the three lines of R code into chunk7 and click on the little arrow on the right.

```
ggplot(mydata) +
```

```
  aes(x=blood_group, y=pop_percentage) +
```

```
  geom_bar(stat = 'identity')
```

What does this code mean?

`ggplot` means 'draw me a graph'

`mydata` means 'use the numbers in the mydata tibble'

`aes` is short for aesthetics and means 'what I want to see is...'

`x=blood_group` means 'plot blood group on the x axis'

`y=pop_percentage` means 'plot percentage of people on the y axis'

`geom_bar` means 'the graph I want is a bar chart'

stat = 'identity' means 'make the bars as tall as the y values in the tibble'

```
```{r chunk7}  
```
```

Check your bar chart. Are the bars black?

Writing labels and title

Now lets alter the labels and add a title so that the bar chart is clearer.

Copy and paste the following R code into chunk8 then click on the little arrow on the right.

```
ggplot(mydata) +  
  aes(x=blood_group, y=pop_percentage) +  
  geom_bar(stat = 'identity')+  
  ylab('Percentage of the population')+  
  xlab('Blood group')+  
  ggtitle('UK Population Blood Groups')
```

You have added some more instructions for R. What do they mean?

ylab('Percentage of the population') means 'make this the label for the y axis'

xlab('Blood group') means 'make this the label for the x axis'

ggtitle('UK Population Blood Groups')means 'make this the title for the barchart'

Note how there is a plus sign after each command

```
```{r chunk8}  
```
```

Check your bar chart. Are the labels correct? Is the title correct? Are the bars still black?

Well done!

Changing the colour of the bars

Let's make it prettier.

Let's make the bars red.

Copy and paste the R code below into chunk9 and click on the little arrow on the right

```
ggplot(mydata) +  
  aes(x=blood_group, y=pop_percentage) +  
  geom_bar(stat = 'identity', fill = 'red')+  
  ylab('Percentage of the population')+  
  xlab('Blood group')+  
  ggtitle('UK Population Blood Groups')
```

```
```{r chunk9}
```

```
```
```

Are the bars red? Suggest what part of the code told R to make the bars red.

Let's make it even prettier. Copy and paste the following code into chunk10

```
ggplot(mydata) +  
  aes(x=blood_group, y=pop_percentage, fill = blood_group) +  
  geom_bar(stat = 'identity')+  
  ylab('Percentage of the population')+  
  xlab('Blood group')+  
  ggtitle('UK Population Blood Groups')
```

```
```{r chunk10}
```

```
```
```

What happened? Suggest what part of the code told R to colour the bars. How did R colour the bars?

Well done! You have successfully completed your second R lesson.

Now KNIT YOUR DOCUMENT for the final time. This is the version of your work that your teacher will mark.

THE END