

# Tutorial 4

## Research Methods for Political Science A

Michele McArdle

28 10 2020

### R Markdown

R Markdown allows you to combine R Code with normal text. It will automatically generate beautifully laid out documents. However, you will have to make some minor adjustments when typing. I have included a cheat sheet showing you the most common codes in R Markdown at the end of this document.

### Printing R Code

R Markdown will automatically include your R Code and print the output.

```
mtcars <- mtcars
```

You can name an R chunk for ease of navigation. These won't show up in the final file.

```
summary(mtcars)
```

```
##      mpg          cyl          disp          hp
##  Min.   :10.40   Min.    :4.000   Min.    : 71.1   Min.    : 52.0
## 1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
## Median :19.20   Median :6.000   Median :196.3   Median :123.0
## Mean   :20.09   Mean    :6.188   Mean    :230.7   Mean    :146.7
## 3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0
## Max.   :33.90   Max.    :8.000   Max.    :472.0   Max.    :335.0
##      drat          wt          qsec          vs
##  Min.   :2.760   Min.    :1.513   Min.    :14.50   Min.    :0.0000
## 1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000
## Median :3.695   Median :3.325   Median :17.71   Median :0.0000
## Mean   :3.597   Mean    :3.217   Mean    :17.85   Mean    :0.4375
## 3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
## Max.   :4.930   Max.    :5.424   Max.    :22.90   Max.    :1.0000
##      am          gear          carb
##  Min.   :0.0000   Min.    :3.000   Min.    :1.000
## 1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
## Median :0.0000   Median :4.000   Median :2.000
## Mean   :0.4062   Mean    :3.688   Mean    :2.812
## 3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
## Max.   :1.0000   Max.    :5.000   Max.    :8.000
```

If you don't want to print the R Code you can add the `echo = FALSE` parameter.

```
## [1] "Hello World"
```

If you want to hide the output you can add the `results = 'hide'` parameter.

```
mtcars
```

If you want to hide warning messages you can add the `message = FALSE` parameter. (This can be useful, especially when loading packages).

```
library("foreign")
```

## Creating the PDF

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document.

## Describing a Sample

We have looked at basic codes for data inspection in tutorials 2 & 3. This week we will focus on producing graphic and tabular inspection of our data.

In this tutorial we will be using the `mtcars` dataset, which is included in base R.

```
mtcars <- mtcars
```

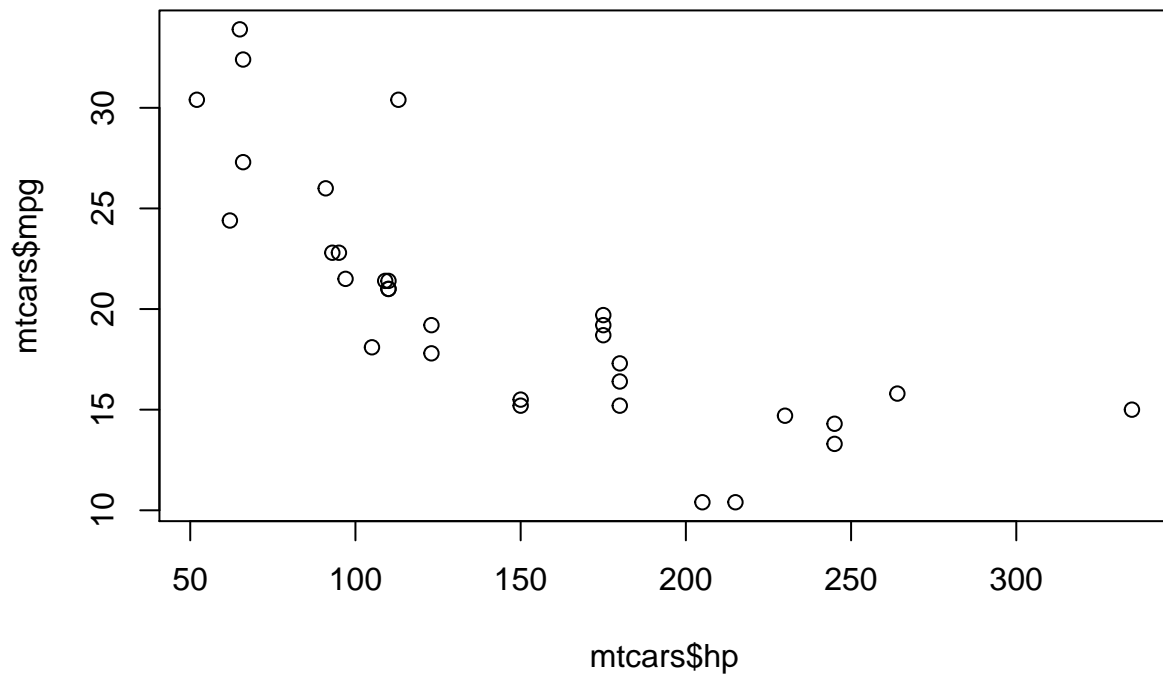
## Graphs

We will be using the plot functions from base R. However, the `ggplot2` package is also widely used and offers customization options.

Plots are especially useful to get a quick overview on the relationship of two variables.

### Scatterplot

```
plot(y = mtcars$mpg, x = mtcars$hp)
```



mpg: Miles/(US) gallon hp : Gross horsepower

### Task 1

Describe the scatterplot. What can we learn about our data ?

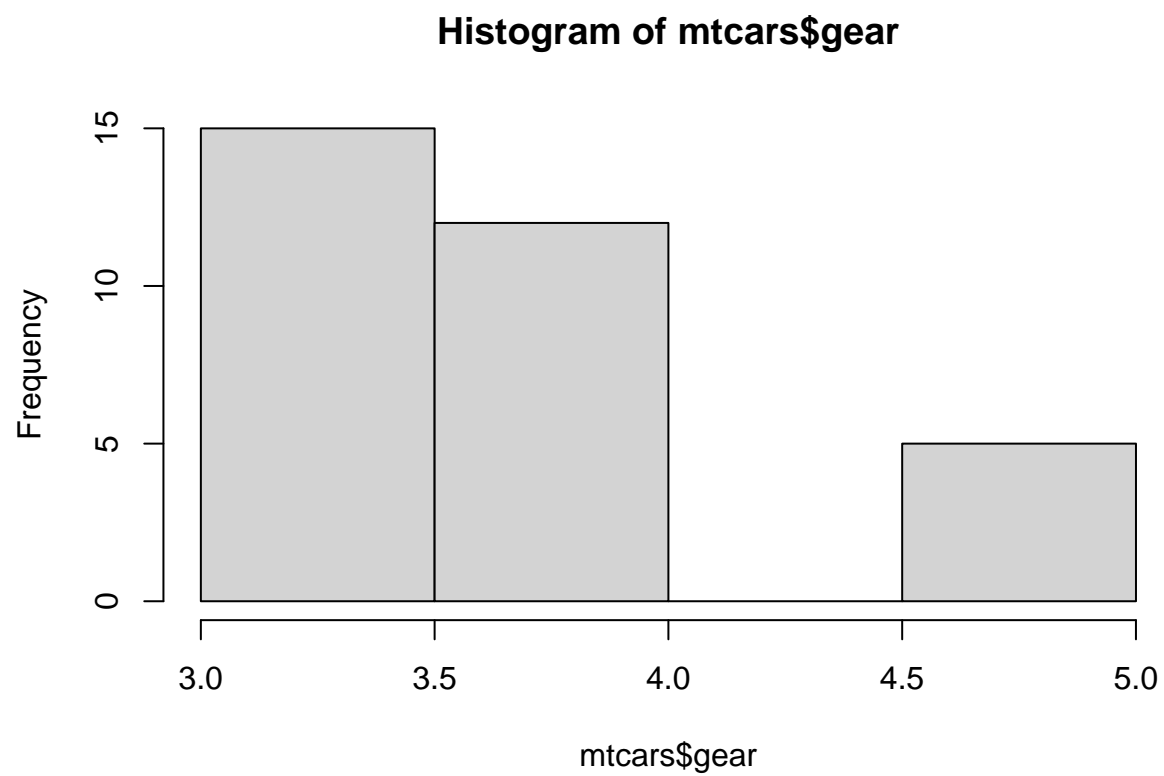
We can see the relationship between the two variables. They are negatively correlated. We can also see that the distribution of variables.

Scatterplots are more useful for continuous variables, and less so for categorical and especially binary variables.

### Histogram

Note we do not include a y variable, as the y-axis provides the frequency.

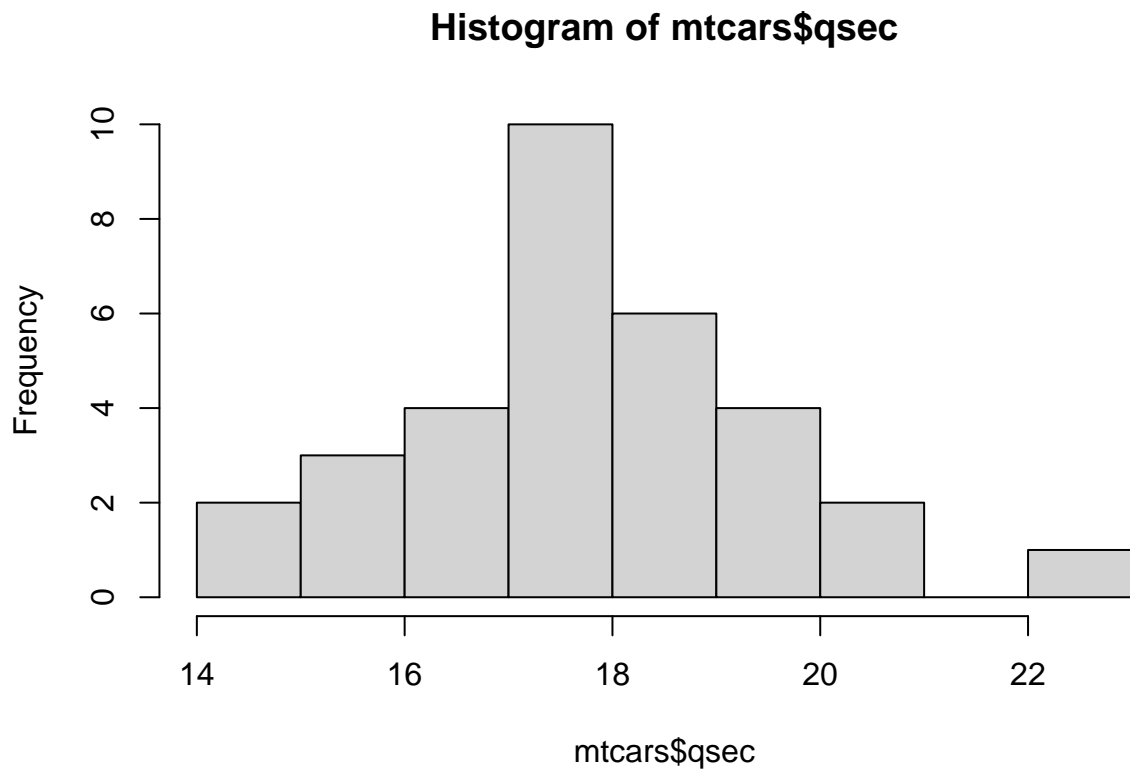
```
hist(mtcars$gear)
```



#### Task 2

Plot a histogram for qsec. (1/4 mile time in seconds). Describe the graph. What does it tell us?

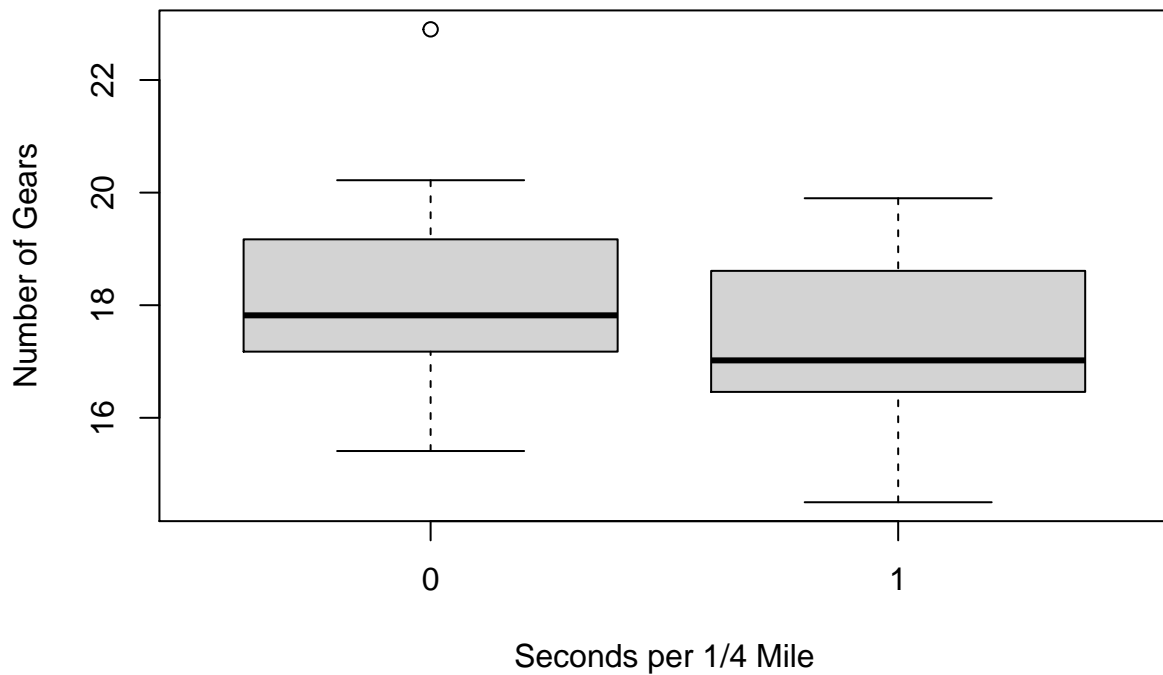
```
hist(mtcars$qsec)
```



We can see the distribution of the plotted variable. Histograms are more useful for categorical variables, as the variable is plotted in bins (the individual columns).

#### Boxplot

```
boxplot(qsec ~ am, mtcars, xlab = "Seconds per 1/4 Mile", ylab = "Number of Gears")
```



### Task 3

Describe the boxplot.

The boxplot has multiple lines. The bold, black line in the center is the median (or 50<sup>th</sup> percentile). The lower and upper limit of the box are the 25<sup>th</sup> and 75<sup>th</sup> percentile aka the first and third quartile. The distance between these two is known as the interquartile range (IQR). The top and bottom whiskers are  $Q1 - 1.5 * IQR$  and  $Q3 + 1.5 * IQR$ . The whiskers can also be defined as 2.5<sup>th</sup> and 97.5<sup>th</sup> percentile

## Tables

### Basic Tables

These types of tables provide frequencies of two (or more) variables. This can be useful in grasping a basic understanding.

```
table(mtcars$cyl, mtcars$am)
```

```
##
##      0  1
##    4  3  8
##    6  4  3
##    8 12  2
```

```
table(mtcars$cyl, mtcars$vs, mtcars$am) #row, column and split
```

```
## , , = 0
##
##
```

```
##      0  1
##    4  0  3
##    6  0  4
##    8 12  0
##
## , , = 1
##
##
##      0  1
##    4  1  7
##    6  3  0
##    8  2  0
```

am: Transmission (0 = automatic, 1 = manual) cyl: Number of cylinders gear: gear Number of forward gears

## CrossTables

Cross tables are handy as they provide multiple additional feature, compared to the basic tables. They allow us to identify the relative proportion of each cell.

```
library(gmodels)
```

```
## Warning: package 'gmodels' was built under R version 4.0.3
```

```
CrossTable(mtcars$cyl, mtcars$am) #row, column
```

```
##
##
##   Cell Contents
## |-----|
## |                      N |
## | Chi-square contribution |
## |      N / Row Total |
## |      N / Col Total |
## |      N / Table Total |
## |-----|
##
##
## Total Observations in Table:  32
##
##
##      | mtcars$am
## mtcars$cyl |      0 |      1 | Row Total |
## -----|-----|-----|-----|
##      4 |      3 |      8 |      11 |
##      |  1.909 |  2.790 |      |
##      |  0.273 |  0.727 |  0.344 |
##      |  0.158 |  0.615 |      |
##      |  0.094 |  0.250 |      |
## -----|-----|-----|-----|
##      6 |      4 |      3 |      7 |
##      |  0.006 |  0.009 |      |
##      |  0.571 |  0.429 |  0.219 |
##      |  0.211 |  0.231 |      |
##      |  0.125 |  0.094 |      |
## -----|-----|-----|-----|
```

##	8		12		2		14	
##			1.636		2.391			
##			0.857		0.143		0.438	
##			0.632		0.154			
##			0.375		0.062			
##	-----		-----		-----		-----	
##	Column Total		19		13		32	
##			0.594		0.406			
##	-----		-----		-----		-----	
##								
##								



# R Markdown Cheat Sheet

learn more at [rmarkdown.rstudio.com](http://rmarkdown.rstudio.com)

rmarkdown 0.2.50 Updated: 8/14



**1. Workflow** R Markdown is a format for writing reproducible, dynamic reports with R. Use it to embed R code and results into slideshows, pdfs, html documents, Word files and more. To make a report:

i. **Open** - Open a file that uses the .Rmd extension.

ii. **Write** - Write content with the easy to use R Markdown syntax

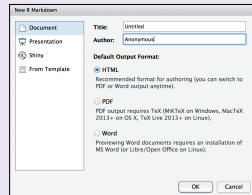
iii. **Embed** - Embed R code that creates output to include in the report

iv. **Render** - Replace R code with its output and transform the report into a slideshow, pdf, html or ms Word file.



**2. Open File** Start by saving a text file with the extension .Rmd, or open an RStudio Rmd template

- In the menu bar, click **File > New File > R Markdown...**
- A window will open. Select the class of output you would like to make with your .Rmd file
- Select the specific type of output to make with the radio buttons (you can change this later)
- Click OK



**4. Choose Output** Write a YAML header that explains what type of document to build from your R Markdown file.

## YAML

A YAML header is a set of key: value pairs at the start of your file. Begin and end the header with a line of three dashes (---)

```
---
title: "Untitled"
author: "Anonymous"
output: html_document
---
```

This is the start of my report. The above is metadata saved in a YAML header.

The RStudio template writes the YAML header for you

The output value determines which type of file R will build from your .Rmd file (in Step 6)

output: **html\_document** ..... html file (web page)  
output: **pdf\_document** ..... pdf document  
output: **word\_document** ..... Microsoft Word .docx  
output: **beamer\_presentation** ..... beamer slideshow (pdf)  
output: **ioslides\_presentation** ..... ioslides slideshow (html)



RStudio® is a trademark of RStudio, Inc. • [CC-BY](http://CC-BY) RStudio • [info@rstudio.com](mailto:info@rstudio.com) • 844-448-1212 • [rstudio.com](http://rstudio.com)

**3. Markdown** Next, write your report in plain text. Use markdown syntax to describe how to format text in the final report.

## syntax

### Plain text

End a line with two spaces to start a new paragraph.

*\*italics\** and *\_italics\_*  
**\*\*bold\*\*** and **\_\_bold\_\_**  
<sup>superscript^2</sup>  
~~--strikethrough--~~  
[\[link\] \(www.rstudio.com\)](#)

# Header 1


## Header 2

### Header 3

#### Header 4

##### Header 5

##### Header 6

endash: --  
emdash: ---  
ellipsis: ...  
inline equation:  $\$A = \pi r^2\$$   
image: 

horizontal rule (or slide break):

\*\*\*

> block quote

\* unordered list  
\* item 2  
+ sub-item 1  
+ sub-item 2

1. ordered list  
2. item 2  
+ sub-item 1  
+ sub-item 2

Table Header	Second Header
Table Cell	Cell 2
Cell 3	Cell 4

## becomes

### Plain text

End a line with two spaces to start a new paragraph.

*italics* and *italics*  
**bold** and **bold**  
<sup>superscript</sup>  
~~strikethrough~~  
[link](#)

## Header 1

## Header 2

### Header 3

#### Header 4

#### Header 5

#### Header 6

endash: --  
emdash: ---  
ellipsis: ...  
inline equation:  $A = \pi r^2$

image:   
horizontal rule (or slide break):

> block quote

\* unordered list  
\* item 2  
+ sub-item 1  
+ sub-item 2  
  
1. ordered list  
2. item 2  
+ sub-item 1  
+ sub-item 2

Table Header	Second Header
Table Cell	Cell 2
Cell 3	Cell 4