

# Data Science exam: Question XX

Erik Valentin Schulte<sup>a</sup>

<sup>a</sup>*Stellenbosch University, South Africa*

---

## Abstract

x.

---

## 1. Introduction

References are to be made as follows: [Fama & French \(1997: 33\)](#) and [Grinold & Kahn \(2000\)](#) Such authors could also be referenced in brackets ([Grinold & Kahn, 2000](#)) and together [Grinold & Kahn \(2000\)](#). Source the reference code from scholar.google.com by clicking on “cite” below article name. Then select BibTeX at the bottom of the Cite window, and proceed to copy and paste this code into your ref.bib file, located in the directory’s Tex folder. Open this file in Rstudio for ease of management, else open it in your preferred Tex environment. Add and manage your article details here for simplicity - once saved, it will self-adjust in your paper.

I suggest renaming the top line after @article, as done in the template ref.bib file, to something more intuitive for you to remember. Do not change the rest of the code. Also, be mindful of the fact that bib references from google scholar may at times be incorrect. Reference Latex forums for correct bibtex notation.

To reference a section, you have to set a label using “\label” in R, and then reference it in-text as e.g. referencing a later section, Section [3](#).

Writing in Rmarkdown is surprisingly easy - see [this website](#) cheatsheet for a summary on writing Rmd writing tips.

---

\*Corresponding author: Erik Valentin Schulte  
Email address: 26802325@sun.ac.su (Erik Valentin Schulte)

---

## Data

Notice how I used the curly brackets and dash to remove the numbering of the data section.

Discussion of data should be thorough with a table of statistics and ideally a figure.

In your tempalte folder, you will find a Data and a Code folder. In order to keep your data files neat, store all of them in your Data folder. Also, I strongly suggest keeping this Rmd file for writing and executing commands, not writing out long pieces of data-wrangling. In the example below, I simply create a ggplot template for scatter plot consistency. I suggest keeping all your data in a data folder.

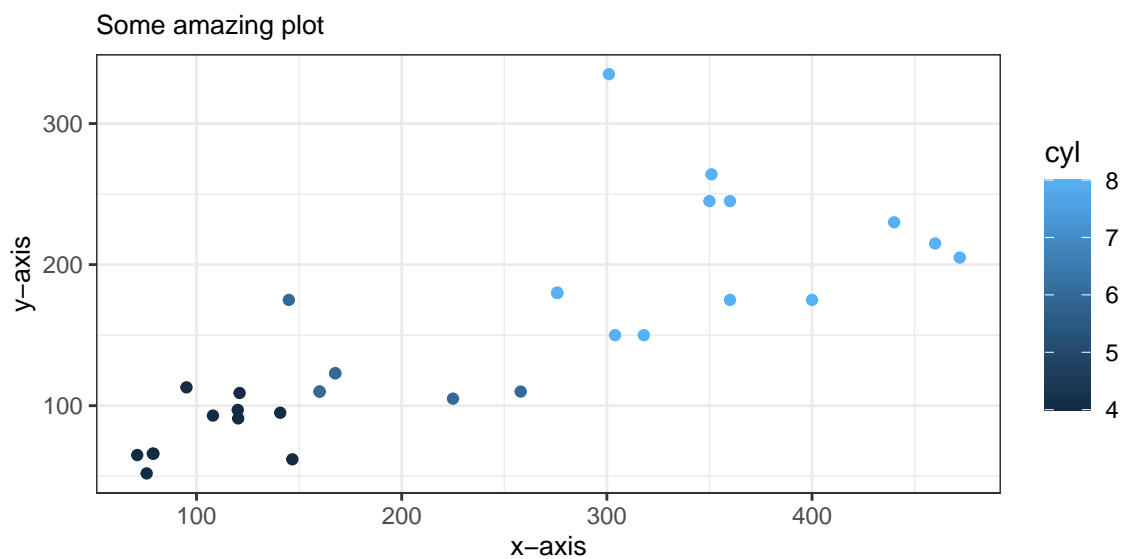


Figure 1.1: Caption Here

To make your graphs look extra nice in latex world, you could use Tikz device. Replace dev - 'png' with 'tikz' in the chunk below. Notice this makes the build time longer and produces extra tex files - so if you are comfortable with this, set your device to Tikz and try it out:

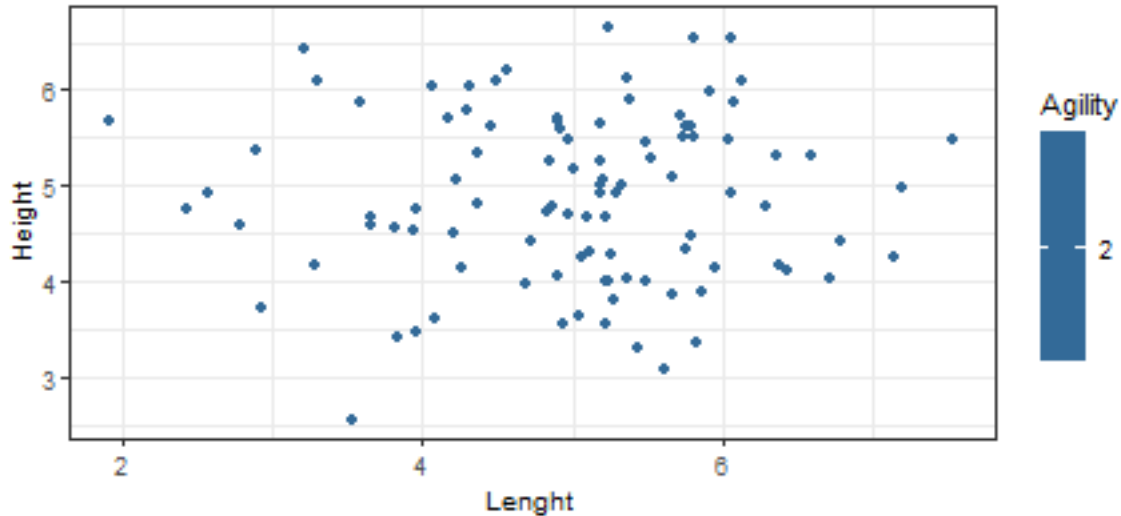


Figure 1.2: Caption Here

To reference the plot above, add a “\label” after the caption in the chunk heading, as done above. Then reference the plot as such: As can be seen, Figures [1.1](#) and [1.2](#) are excellent, with Figure [1.2](#) being particularly aesthetically pleasing due to its device setting of Tikz. The nice thing now is that it correctly numbers all your figures (and sections or tables) and will update if it moves. The links are also dynamic.

I very strongly suggest using ggplot2 (ideally in combination with dplyr) using the ggtheme package to change the themes of your figures.

Also note the information that I have placed above the chunks in the code chunks for the figures. You can edit any of these easily - visit the Rmarkdown webpage for more information.

## 2. Splitting a page

You can also very easily split a page using built-in Pandoc formatting. I comment this out in the code (as this has caused issues building the pdf for some users - which I presume to be a Pandoc issue), but you are welcome to try it out yourself by commenting out the following section in your Rmd file.

---

### 3. Methodology

#### 3.1. Subsection

Ideally do not overuse subsections. It equates to bad writing.<sup>1</sup>

#### 3.2. Math section

Equations should be written as such:

$$\beta = \sum_{i=1}^{\infty} \frac{\alpha^2}{\sigma_{t-1}^2} \int_{x=1}^{\infty} x_i = 1 \quad (3.1)$$

If you would like to see the equations as you type in Rmarkdown, use \$ symbols instead (see this for yourself by adjusted the equation):

$$\beta = \sum_{i=1}^{\infty} \frac{\alpha^2}{\sigma_{t-1}^2} \int_{x=1}^{\infty} x_i = 1$$

Note again the reference to equation 3.1. Writing nice math requires practice. Note I used a forward slashes to make a space in the equations. I can also align equations using &, and set to numbering only the first line. Now I will have to type “begin equation” which is a native L<sup>A</sup>T<sub>E</sub>X command. Here follows a more complicated equation:

$$\begin{aligned} y_t &= c + B(L)y_{t-1} + e_t \\ e_t &= H_t^{1/2} z_t; \quad z_t \sim N(0, I_N) \quad \& \quad H_t = D_t R_t D_t \\ D_t^2 &= \sigma_{1,t}, \dots, \sigma_{N,t} \\ \sigma_{i,t}^2 &= \gamma_i + \kappa_{i,t} v_{i,t-1}^2 + \eta_i \sigma_{i,t-1}^2, \quad \forall i \\ R_{t,i,j} &= \text{diag}(Q_{t,i,j}^{-1}) \cdot Q_{t,i,j} \cdot \text{diag}(Q_{t,i,j}^{-1}) \\ Q_{t,i,j} &= (1 - \alpha - \beta) \bar{Q} + \alpha z_t z_t' + \beta Q_{t,i,j} \end{aligned} \quad (3.2)$$

---

<sup>1</sup>This is an example of a footnote by the way. Something that should also not be overused.

Note that in 3.2 I have aligned the equations by the equal signs. I also want only one tag, and I create spaces using “quads’ ’.

See if you can figure out how to do complex math using the two examples provided in 3.1 and 3.2.

## 4. Results

Tables can be included as follows. Use the *xtable* (or *kable*) package for tables. Table placement = H implies Latex tries to place the table Here, and not on a new page (there are, however, very many ways to skin this cat. Luckily there are many forums online!).

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
1	21.00	6.00	160.00	110.00	3.90	2.62	16.46	0.00	1.00	4.00	4.00
2	21.00	6.00	160.00	110.00	3.90	2.88	17.02	0.00	1.00	4.00	4.00
3	22.80	4.00	108.00	93.00	3.85	2.32	18.61	1.00	1.00	4.00	1.00
4	21.40	6.00	258.00	110.00	3.08	3.21	19.44	1.00	0.00	3.00	1.00
5	18.70	8.00	360.00	175.00	3.15	3.44	17.02	0.00	0.00	3.00	2.00

Table 4.1: Short Table Example

To reference calculations **in text**, *do this*: From table 4.1 we see the average value of mpg is 20.98.

Including tables that span across pages, use the following (note that I add below the table: “continue on the next page’ ’). This is a neat way of splitting your table across a page.

Use the following default settings to build your own possibly long tables. Note that the following will fit on one page if it can, but cleanly spreads over multiple pages:

Table 4.2: Long Table Example

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
	21.00	6.00	160.00	110.00	3.90	2.62	16.46	0.00	1.00	4.00	4.00
	21.00	6.00	160.00	110.00	3.90	2.88	17.02	0.00	1.00	4.00	4.00
	22.80	4.00	108.00	93.00	3.85	2.32	18.61	1.00	1.00	4.00	1.00
	21.40	6.00	258.00	110.00	3.08	3.21	19.44	1.00	0.00	3.00	1.00
	18.70	8.00	360.00	175.00	3.15	3.44	17.02	0.00	0.00	3.00	2.00
	18.10	6.00	225.00	105.00	2.76	3.46	20.22	1.00	0.00	3.00	1.00
	14.30	8.00	360.00	245.00	3.21	3.57	15.84	0.00	0.00	3.00	4.00
	24.40	4.00	146.70	62.00	3.69	3.19	20.00	1.00	0.00	4.00	2.00

Continued on next page

Table 4.2: Long Table Example

mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
22.80	4.00	140.80	95.00	3.92	3.15	22.90	1.00	0.00	4.00	2.00
19.20	6.00	167.60	123.00	3.92	3.44	18.30	1.00	0.00	4.00	4.00
17.80	6.00	167.60	123.00	3.92	3.44	18.90	1.00	0.00	4.00	4.00
16.40	8.00	275.80	180.00	3.07	4.07	17.40	0.00	0.00	3.00	3.00
17.30	8.00	275.80	180.00	3.07	3.73	17.60	0.00	0.00	3.00	3.00
15.20	8.00	275.80	180.00	3.07	3.78	18.00	0.00	0.00	3.00	3.00
10.40	8.00	472.00	205.00	2.93	5.25	17.98	0.00	0.00	3.00	4.00
10.40	8.00	460.00	215.00	3.00	5.42	17.82	0.00	0.00	3.00	4.00
14.70	8.00	440.00	230.00	3.23	5.34	17.42	0.00	0.00	3.00	4.00
32.40	4.00	78.70	66.00	4.08	2.20	19.47	1.00	1.00	4.00	1.00
30.40	4.00	75.70	52.00	4.93	1.61	18.52	1.00	1.00	4.00	2.00
33.90	4.00	71.10	65.00	4.22	1.83	19.90	1.00	1.00	4.00	1.00
21.50	4.00	120.10	97.00	3.70	2.46	20.01	1.00	0.00	3.00	1.00
15.50	8.00	318.00	150.00	2.76	3.52	16.87	0.00	0.00	3.00	2.00
15.20	8.00	304.00	150.00	3.15	3.44	17.30	0.00	0.00	3.00	2.00
13.30	8.00	350.00	245.00	3.73	3.84	15.41	0.00	0.00	3.00	4.00
19.20	8.00	400.00	175.00	3.08	3.85	17.05	0.00	0.00	3.00	2.00
27.30	4.00	79.00	66.00	4.08	1.94	18.90	1.00	1.00	4.00	1.00
26.00	4.00	120.30	91.00	4.43	2.14	16.70	0.00	1.00	5.00	2.00
30.40	4.00	95.10	113.00	3.77	1.51	16.90	1.00	1.00	5.00	2.00
15.80	8.00	351.00	264.00	4.22	3.17	14.50	0.00	1.00	5.00	4.00
19.70	6.00	145.00	175.00	3.62	2.77	15.50	0.00	1.00	5.00	6.00
15.00	8.00	301.00	335.00	3.54	3.57	14.60	0.00	1.00	5.00	8.00
21.40	4.00	121.00	109.00	4.11	2.78	18.60	1.00	1.00	4.00	2.00

#### 4.1. Huxtable

Huxtable is a very nice package for making working with tables between Rmarkdown and Tex easier.

This cost some adjustment to the Tex templates to make it work, but it now works nicely.

See documentation for this package [here](#). A particularly nice addition of this package is for making the printing of regression results a joy (see [here](#)). Here follows an example:

If you are eager to use huxtable, comment out the Huxtable table in the Rmd template, and uncomment

---

the colortbl package in your Rmd's root.

Note that I do not include this in the ordinary template, as some latex users have complained it breaks when they build their Rmds (especially those using tidytex - I don't have this problem as I have the full Miktex installed on mine). Up to you, but I strongly recommend installing the package manually and using huxtable. To make this work, uncomment the *Adding additional latex packages* part in yaml at the top of the Rmd file. Then comment out the huxtable example in the template below this line. Reknit, and enjoy.

Table 4.3: Regression Output

	Reg1	Reg2	Reg3
(Intercept)	-2256.361 *** (13.055)	5763.668 *** (740.556)	4045.333 *** (286.205)
carat	7756.426 *** (14.067)		7765.141 *** (14.009)
depth		-29.650 * (11.990)	-102.165 *** (4.635)
N	53940	53940	53940
R2	0.849	0.000	0.851

\*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$ .

FYI - R also recently introduced the gt package, which is worthwhile exploring too.

## 5. Lists

To add lists, simply using the following notation

- This is really simple
  - Just note the spaces here - writing in R you have to sometimes be pedantic about spaces...
- Note that Rmarkdown notation removes the pain of defining  $\text{\LaTeX}$  environments!

---

## 6. Conclusion

I hope you find this template useful. Remember, stackoverflow is your friend - use it to find answers to questions. Feel free to write me a mail if you have any questions regarding the use of this package. To cite this package, simply type citation("Texevier") in Rstudio to get the citation for [Katzke \(2017\)](#) (Note that uncited references in your bibtex file will not be included in References).



---

## References

10 Fama, E.F. & French, K.R. 1997. Industry costs of equity. *Journal of financial economics*. 43(2):153–193.

Grinold, R.C. & Kahn, R.N. 2000. Active portfolio management.

Katzke, N.F. 2017. *Texevier: Package to create elsevier templates for rmarkdown*. Stellenbosch, South Africa: Bureau for Economic Research.

## Appendix

### *Appendix A*

Some appendix information here

### *Appendix B*