

**Question 1 (R question)**

This question is to ensure you can generate an R Markdown report using R Studio.

- (a) Open the file `example_rmarkdown_to_html.Rmd` in R Studio. The file is available on Blackboard.
- (b) Click the 'Knit' button to generate a HTML version of this file. If you cannot see a 'Knit' icon, look for a 'Preview' icon. Click on the drop-down arrow next to 'Preview' and select 'Knit to HTML'.
- (c) If the scatterplot does not appear, click on the drop-down arrow on 'Run' command (top-right corner of editor window) and select 'Run All' and re-knit.
- (d) Modify the 'subtitle' field at the top of your Rmd file to have your full name and CID number.
- (e) Edit the Rmd file to fix the expression for the normal probability density function. This is an example of LaTeX code for creating equations; although you may not have seen it before, try to match the code with the output and change the numbers accordingly. After editing, click 'Knit' again to regenerate the HTML.
- (f) Edit the Rmd file to plot a histogram instead of a scatterplot.
- (g) Increase the number of observations generated from 100 to 10000. Re-knit the document. Re-knit the document.
- (h) For the histogram, set the parameter `freq` to `FALSE`, i.e. `hist(z, freq=FALSE)`, and replot the histogram (re-knit the document).
- (i) Plot the probability density function for the appropriate normal distribution over the histogram. You will need to use the `lines` command, and set the parameters for the normal distribution's mean and standard deviation inside the `dnorm` command correctly.

**Question 2 (R question)**

In this question you will create your own R Markdown document from scratch (use Question 1 as a template).

- (a) Create a new R Markdown document which creates a HTML file as output, and name it `question2.Rmd`
- (b) Download the file `shoesizes.txt` from Blackboard, and place it in the same directory as the R Markdown document `question2.Rmd` that you just created.
- (c) In `question2.Rmd`, create a section heading named ‘Question 2’
- (d) Below the section heading, write a sentence saying that you intend to plot the data contained in `shoesizes.txt`.
- (e) Create an R code block that reads in `shoesizes.txt` into a dataframe named `df`.
- (f) In the same code block, print the first few lines of the dataframe `df` using the `head` command, i.e. `head(df)`.
- (g) From the dataframe `df`, extract the column of data containing the shoe sizes to a vector named `shoesize`. You will need to use the ‘\$’ operator, and use the name of the column as printed out from the `head` command, i.e. `VECTOR_NAME <- df$COLUMN_NAME`.
- (h) From the dataframe `df`, extract the column of data containing the heights to a vector named `height`
- (i) Create a scatter plot of the heights vs the shoe sizes, i.e. a plot with the heights on the vertical axis and the shoe sizes on the horizontal axis. Use the `xlab` and `ylab` parameters in the `plot` command to label the axes ‘Height (inches)’ and ‘Shoe size (US sizes)’.
- (j) Use the `cor` function to compute the correlation of heights and shoes sizes and save this value in the variable `cor_inches`. Print this value to screen.
- (k) Create a vector `height_in_cm` which is the same as the vector `height`, but the measurements are in centimetres. Hint: 1 inch is 2.54 cm.
- (l) Compute the correlation of heights in cm and shoes sizes and save this value in the variable `cor_cm`. Print this value to screen, and compare to `cor_inches`.

**To print an R Markdown notebook to PDF**

In the R Markdown window which displays the knitted HTML notebook, click on ‘Open in Browser’ in the tab at the top. This will open the notebook in a browser such as Firefox, Chrome, etc. In the browser version of the notebook, print the notebook to PDF. Exactly how you do this will, but for most browsers selecting the ‘File’ tab of the browser and selecting ‘Print’ is an option.