STA288H1 Lab 0 Assignment - Introduction to RStudio & STA288H1 Labs

TYPE YOUR *NAME* HERE (type your *STUDENT NUMBER* & *LEC section* here)

2026-01-22

Lab 0 Assignment (both .Rmd & .pdf) - Due 11:59 ET, THURSDAY, JAN 22, 2026 on Quercus: https://q.utoronto.ca/courses/427449/assignments/1670278)

Graded out of *1 point*

NOTE: you must export *both* your completed R Markdown (i.e., rmd) file and your pdf file of your Lab 0 answers from U of T JupyterHub and save on your machine; then upload to Quercus.

Note that late assignments and assignments submitted other ways (e.g., over email) will not be accepted. If you have trouble knitting your assignment, submit your completed Rmd on its own by the deadline so you can earn partial credit toward this assignment.

This is an R Markdown document. Markdown is a syntax for authoring documents that are a mix of text and R code and output. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button above, a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. When you submit lab assignments in STA288, you will need to submit this R Markdown (.Rmd) file AND the pdf file you generate by clicking on **Knit > Knit to PDF** above (save that file as Lab0yourname.pdf). To save on your own machine to upload to Quercus, check the boxes next to the two files in the files tab on the right, then click **More > Export** and download to your own machine.

In this lab, you will read data from the Baker (2016) survey on reproducibility that we discussed during the first week into R and produce and interpret several data summaries. The goal of this lab is to give you hands-on experience with RStudio and our lab workflow before the first lab.

A Nature paper was published in 2016 (https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970) reporting the results of a large internet survey of scientists' opinions on the existence of a reproducibility crisis. The dataset "Reproducibility_Survey.csv" includes data on a few of the survey questions asked. The following code reads in this data and stores it in an R dataset called *surveydat* and the 'glimpse' function lists the variables stored in this dataset along with the first values of these variables.

```
surveydat<-read_csv("Reproducibility_Survey.csv")
```

```
## Rows: 1576 Columns: 8
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (6): familiar, crisis, age, discipline, discipline-other, continent
## dbl (2): responseid, reproducible
##
```

```
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
glimpse(surveydat)
```

```
## Rows: 1,576
## Columns: 8
## $ responseid         <dbl> 24, 27, 36, 107, 112, 114, 117, 145, 179, 180, 188,~
## $ familiar           <chr> "4 - Fairly familiar", "5 - Very familiar", "5 - Ve~
## $ crisis             <chr> "yes", "yes", "yes", "yes", "yes", "yes", "yes", "y~
## $ reproducible       <dbl> 90, 40, 30, 50, 70, 20, 70, 30, 80, 80, 40, 80, 70,~
## $ age                <chr> "65 or over", "25 - 34", "45 - 54", "35 - 44", "18 ~
## $ discipline         <chr> "Other", "Other", "Other", "Biology", "Biology", "O~
## $ `discipline-other` <chr> "Interferon", "communication studies", "Neurophysio~
## $ continent          <chr> "Europe", "Europe", "Europe", "Europe", "South Amer~
```

QUESTION 1 Survey respondents were asked to rate their familiarity with reproducibility concerns in science on a scale from *1-Not at all familiar* to *5-Very familiar* (stored in variable `familiar`). Are scientists familiar with reproducibility concerns in science? Produce an appropriate summary of these data and briefly interpret the results.

QUESTION 2 Respondents were asked to report the percentage of research in their fields they think is reproducible.

(a) Create an appropriate summary of the distribution of reproducible research percentages reported by survey respondents. Briefly interpret your summary.

(b) Many other questions were asked on the survey. For instance, respondents were also asked their discipline (`discipline`). Does the percentage of research thought to be reproducible vary by discipline? Produce and briefly interpret an appropriate graph to address this question.

LAB ASSIGNMENT REPRODUCIBILITY Your Lab 0 assignment file submission in the Quercus Assignment must include *both* the rmd file with your lab 0 answers that was compiled (or *knitted*) to produce a pdf file of your lab 0 answers.

Rubric:

- 0/1 point - Submitted rmd file does not produce submitted pdf file OR answers are missing.
- 0.5/1 point - The rmd or pdf file is missing OR several answers incomplete (i.e., missing relevant R code, output, or brief interpretations).
- 1/1 point - Both rmd and the corresponding pdf files submitted AND relevant R code, output, and brief interpretations included for all questions.

THIS IS THE END OF LAB ASSIGNMENT 0

```
## [1] 0.392006
```

```
## [1] "Wed Jan 14 22:38:46 2026"
```