WorkingWithDataAssignment

João Mota

Table of contents

/*
A few hours of trial and errors can save you a few minutes of reading the proper documentation :) https://quarto.org/docs/output-formats/pdf-basics.html
Go to terminal tab down there and type quarto install tool tinytex
NOTE TO SELF!!!! using quarto is the same as playing restart Rstudio simulator 2022 because nothing is properly recached and they have a worse garbage collecter than assembly so if you still get the same error after changing the just restart rstudio and remember to never ever ever change the initial format or add anything close to it because it will break the pdf and start generating html also please be smart and read https://quarto.org/docs/reference/formats/pdf.html for the formating
*/

```
  library(haven)
  library(tidyverse)
```

```
-- Attaching packages --------------------------------------- tidyverse 1.3.2 --
v ggplot2 3.3.6      v purrr   0.3.4
v tibble  3.1.8      v dplyr   1.0.10
v tidyr   1.2.1      v stringr 1.4.1
v readr   2.1.3      v forcats 0.5.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

```
  library(dplyr)
  library(geometry)
  #install.packages('hyperref')
  library(formatR)
```

```
## Very important documentation for the 2018 data set //it is a
## surprise toll that will help us later
technicalAnnex2018 = "https://doc.ukdataservice.ac.uk/doc/8406/mrdoc/pdf/8406_cyber_security_b

## this is the loading the first year of this level of survey data set
## after burning my entire brain, replacing it with the backup one and
## also burning that one I discovered that it is just these lines that
## aren't being formatted in pdf because they are absolutely huge but
## at least it works for the other ones #FicaADica I assume it was
## thanks to formatR ?? I won't bother to redo every single bloody step
## again, enough for a lifetime and a half
dataCyberSecuritySurvey2018 = read_spss("C:/AppliedDataScienceAndStatistics/Applied-Data-Scien

## adding the variable year because none of the data sets have any
## proper way to distinguish between the years of each survey
dataCyberSecuritySurvey2018$year = "2018"
```

1 Now we do the same for the other years before we merge them

```
## loading the second year of this level of survey data set
dataCyberSecuritySurvey2019 = read_spss("C:/AppliedDataScienceAndStatistics/Applied-Data-Scien

## adding the variable year because none of the data sets have any
## proper way to distinguish between the years of each survey
dataCyberSecuritySurvey2019$year = "2019"
```

```
## loading the third year of this level of survey data set
dataCyberSecuritySurvey2020 = read_spss("C:/AppliedDataScienceAndStatistics/Applied-Data-Scien

## adding the variable year because none of the data sets have any
## proper way to distinguish between the years of each survey
dataCyberSecuritySurvey2020$year = "2020"
```

```
## loading the forth year of this level of survey data set
dataCyberSecuritySurvey2021 = read_spss("C:/AppliedDataScienceAndStatistics/Applied-Data-Scien

## adding the variable year because none of the data sets have any
## proper way to distinguish between the years of each survey
dataCyberSecuritySurvey2021$year = "2021"
```

```
## loading the fifth and final year of this level of survey data set
dataCyberSecuritySurvey2022 = read_spss("C:/AppliedDataScienceAndStatistics/Applied-Data-Scien

## adding the variable year because none of the data sets have any
```

```r
## proper way to distinguish between the years of each survey
dataCyberSecuritySurvey2022$year = "2022"
```

```r
## Now that we have all data loaded lets start by tidying up data set
## by data set start from 2018


## for some sweet sweet documentation about the questions starting from
## page 26 TODO remove comment browseURL(technicalAnnex2018)
```

```r
## This entire code snippet is tidying up the type of organisation for
## the 2018 survey renaming the bloody variables to a more java like
## name

dataCyberSecuritySurvey2018TidyName = rename(dataCyberSecuritySurvey2018,
    isBusiness = "samptype")

## if isBusiness is 1 it is a business if it is 2 it is a charity will
## change it the 2 to 0 because what is the point of a boolean without
## boolean values

n = length(dataCyberSecuritySurvey2018TidyName$isBusiness)



for (i in 1:n) {
    if (dataCyberSecuritySurvey2018TidyName$isBusiness[i] == 2) {
        dataCyberSecuritySurvey2018TidyName$isBusiness[i] = 0
    }
}

## daily remimder that there is a boolean type but it is called logical
## Numeric -\tSet of all real numbers Integer -\tSet of all integers, Z
## Logical - -\tTRUE and FALSE Complex -\tSet of complex numbers
## Character -\t"a", "b", "c", …, "@", "#", "$", …., "1", "2", …etc

str(dataCyberSecuritySurvey2018TidyName$isBusiness)
```

```
dbl+lbl [1:2088] 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, ...
@ label        : chr "Samptype"
@ format.spss  : chr "F8.0"
@ display_width: int 10
@ labels       : Named num [1:2] 1 2
 ..- attr(*, "names")= chr [1:2] "Business" "Charity"
```

```
## it is a string so lets make it a proper boolean

dataCyberSecuritySurvey2018TidyName$isBusiness = as.integer(dataCyberSecuritySurvey2018TidyNam

dataCyberSecuritySurvey2018TidyName$isBusiness = as.logical(dataCyberSecuritySurvey2018TidyNam

str(dataCyberSecuritySurvey2018TidyName$isBusiness)
```

```
logi [1:2088] TRUE TRUE TRUE TRUE TRUE TRUE ...
```

```
## dataCyberSecuritySurvey2018TidyName never forget if R can't show all
## displayed text from a computation it breaks both the rendering and
## ##the refreshing of the rendered code for some reason ¯\_( )_/¯
## future edit anything and everything breaks for no reason at all,
## just kill it and reopen refer to the first NOTE TO SELF for more
## information

## typex is 1-2 for businesses and 3 for charities so redundant and can
## be removed

dataCyberSecuritySurvey2018TidyName = dataCyberSecuritySurvey2018TidyName %>%
    select(-typex)
```

2 Quarto

Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see https://quarto.org.

3 Running Code

When you click the **Render** button a document will be generated that includes both content and the output of embedded code. You can embed code like this:

```
1 + 1
```

```
[1] 2
```

You can add options to executable code like this

```
[1] 4
```

The `echo: false` option disables the printing of code (only output is displayed).