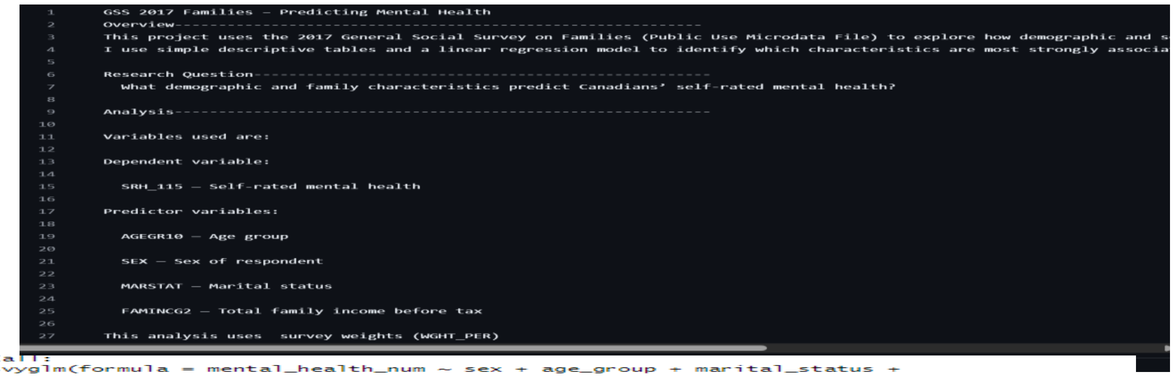
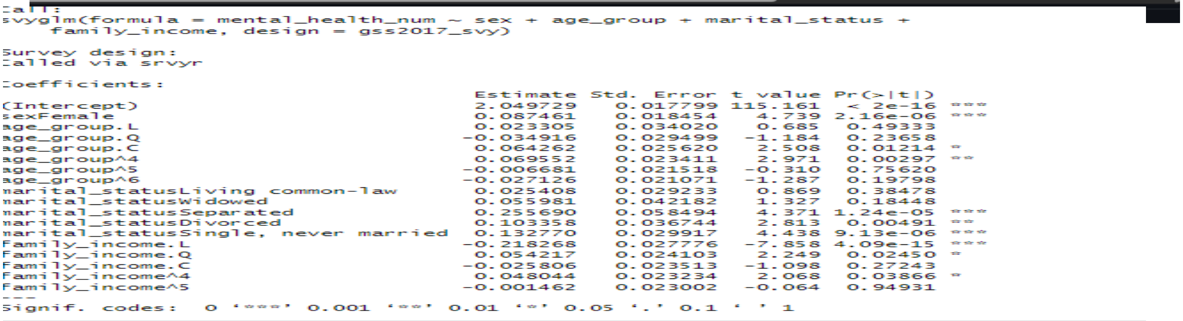
RRWM Assignment 2: Post-RRWM Report & Discussion

Poll: Were you able to reproduce the work based only on the program (not the code)?
• *Program doesn't even make a table that looks like what I was sent.*

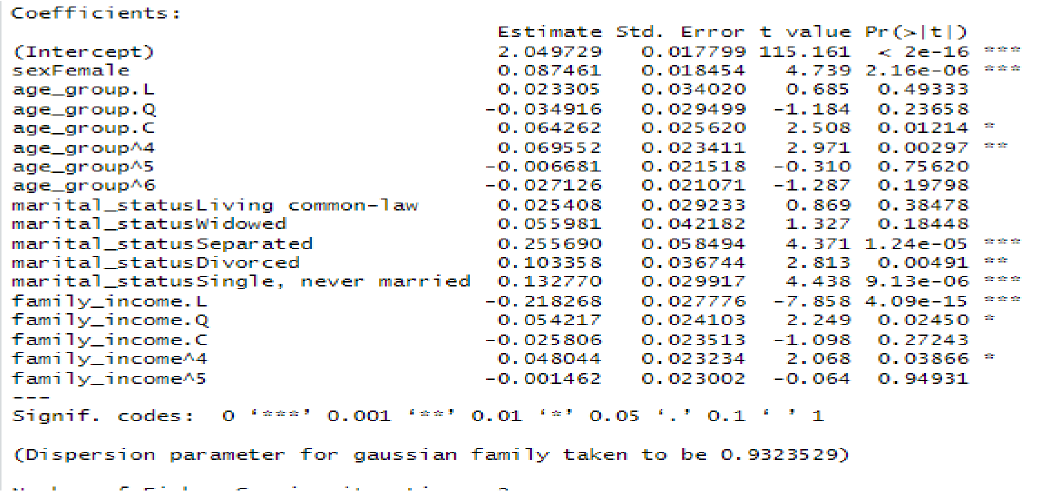
The reason for this is as follows: my partner Fabio and I did not fully understand what the exercise was about. We had instead shared files via GitHub containing the **script**, rather than a **program (in text format)** describing the different steps to follow — namely, the selection of variables, their coding and cleaning, and the generation of descriptive and regression tables. Nevertheless, using the code and the image file, we were able to easily reproduce the tables and results of these analyses (see image files below).



```
GSS 2017 Families – Predicting Mental Health
Overview-------------------------------------------------------------
This project uses the 2017 General Social Survey on Families (Public Use Microdata File) to explore how demographic and s
I use simple descriptive tables and a linear regression model to identify which characteristics are most strongly associa

Research Question----------------------------------------------------
  What demographic and family characteristics predict Canadians' self-rated mental health?

Analysis-------------------------------------------------------------

Variables used are:

Dependent variable:

  SRH_115 — Self-rated mental health

Predictor variables:

  AGEGR10 — Age group

  SEX — Sex of respondent

  MARSTAT — Marital status

  FAMINCG2 — Total family income before tax

This analysis uses  survey weights (WGHT_PER)
```



```
Call:
svyglm(formula = mental_health_num ~ sex + age_group + marital_status + 
    family_income, design = gss2017_svy)

Survey design:
Called via srvyr

Coefficients:
                                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)                          2.049729   0.017799 115.161  < 2e-16 ***
sexFemale                            0.087461   0.018454   4.739 2.16e-06 ***
age_group.L                          0.023305   0.034020   0.685  0.49333    
age_group.Q                         -0.034916   0.029499  -1.184  0.23658    
age_group.C                          0.064262   0.025620   2.508  0.01214 *  
age_group^4                          0.069552   0.023411   2.971  0.00297 ** 
age_group^5                         -0.006681   0.021518  -0.310  0.75620    
age_group^6                         -0.027126   0.021071  -1.287  0.19798    
marital_statusLiving common-law      0.025408   0.029233   0.869  0.38478    
marital_statusWidowed                0.055981   0.042182   1.327  0.18448    
marital_statusSeparated              0.255690   0.058494   4.371 1.24e-05 ***
marital_statusDivorced               0.103358   0.036744   2.813  0.00491 ** 
marital_statusSingle, never married  0.132770   0.029917   4.438 9.13e-06 ***
Family_income.L                     -0.218268   0.027776  -7.858 4.09e-15 ***
Family_income.Q                      0.054217   0.024103   2.249  0.02450 *  
Family_income.C                     -0.025806   0.023513  -1.098  0.27243    
Family_income^4                      0.048044   0.023234   2.068  0.03866 *  
Family_income^5                     -0.001462   0.023002  -0.064  0.94931    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

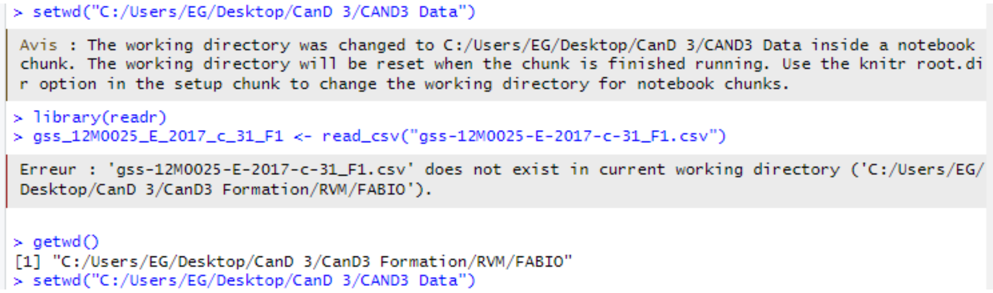


```
Coefficients:
                                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)                          2.049729   0.017799 115.161  < 2e-16 ***
sexFemale                            0.087461   0.018454   4.739 2.16e-06 ***
age_group.L                          0.023305   0.034020   0.685  0.49333    
age_group.Q                         -0.034916   0.029499  -1.184  0.23658    
age_group.C                          0.064262   0.025620   2.508  0.01214 *  
age_group^4                          0.069552   0.023411   2.971  0.00297 ** 
age_group^5                         -0.006681   0.021518  -0.310  0.75620    
age_group^6                         -0.027126   0.021071  -1.287  0.19798    
marital_statusLiving common-law      0.025408   0.029233   0.869  0.38478    
marital_statusWidowed                0.055981   0.042182   1.327  0.18448    
marital_statusSeparated              0.255690   0.058494   4.371 1.24e-05 ***
marital_statusDivorced               0.103358   0.036744   2.813  0.00491 ** 
marital_statusSingle, never married  0.132770   0.029917   4.438 9.13e-06 ***
family_income.L                     -0.218268   0.027776  -7.858 4.09e-15 ***
family_income.Q                      0.054217   0.024103   2.249  0.02450 *  
family_income.C                     -0.025806   0.023513  -1.098  0.27243    
family_income^4                      0.048044   0.023234   2.068  0.03866 *  
family_income^5                     -0.001462   0.023002  -0.064  0.94931    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.9323529)
```

Where did things go off track? Was there one error that caused everything to go wrong?

Since I based my work on the code file, here is what I noticed:

Even though I used setwd() to set my working directory and access the database, executing the command to load the dataset kept returning an error message, as shown above. It seemed that the R Markdown script forced me to work in a specific directory that was not the one I had defined. I had to open a new script and copy-paste each of the commands shown in the figure to solve the problem.



```
> setwd("C:/Users/EG/Desktop/CanD 3/CAND3 Data")
Avis : The working directory was changed to C:/Users/EG/Desktop/CanD 3/CAND3 Data inside a notebook
chunk. The working directory will be reset when the chunk is finished running. Use the knitr root.di
r option in the setup chunk to change the working directory for notebook chunks.
> library(readr)
> gss_12M0025_E_2017_c_31_F1 <- read_csv("gss-12M0025-E-2017-c-31_F1.csv")
Erreur : 'gss-12M0025-E-2017-c-31_F1.csv' does not exist in current working directory ('C:/Users/EG/
Desktop/CanD 3/CanD3 Formation/RVM/FABIO').

> getwd()
[1] "C:/Users/EG/Desktop/CanD 3/CanD3 Formation/RVM/FABIO"
> setwd("C:/Users/EG/Desktop/CanD 3/CAND3 Data")
```

What parts of the program left you uncertain about what to do?

In my case, the variable coding was clear.

Do you have any comments or reflections about the reproducibility exercise? For instance, has it changed your mind about the benefits (or costs) of open science?

This exercise is very useful. In fact, I recently submitted an article to a journal that required me to make my results reproducible by sharing the data I used via GitHub. However, I think reproducibility should be limited to sharing the **code files or scripts** developed by participants. Moreover, I believe that the words *program* and *reproducibility*, when used together, refer to writing lines of code using a software program and sharing them with the scientific community.

For those in fields where replication materials are not readily available, could an emphasis on reproducibility disadvantage researchers who use qualitative methods or restricted (e.g., administrative) data?

I don't think so. For researchers using qualitative methods, it is possible to share their "research protocol" with the scientific community, including the sampling procedure, the interview guides (individual and group), and the methods of analysis.
As for restricted data, most scientific journals require authors to state that such data are only available upon request from the data-holding institution.
Therefore, regardless of the field of research, there is always some framework that allows the reproducibility of results to be assessed — or at least approached.