

# Modern workflows in data science

## Final assignment

Alexandru Cernat

In this assignment we will work again with the EVS data. So you can use the data cleaning from the second assignment. Now you will have to develop a Shiny app that explores that data and that creates dynamic reports. The submission will once again be a GitHub repo.

In the Shiny app we will be using the same data as last time but in a different way. Your app will have three sections:

1. **Overview** where you describe the aims of the app and how to navigate it.
2. **Exploration** where you will do a graph to describe: the outcome variable, and the three controls: age, education and sex.
3. **Regression** where you will have a table showing the regression coefficients and a scatter plot showing the predicted versus the residuals from the regression. By default the regression has just one outcome (whichever you choose as the default) and one predictor, age.

You will also have four inputs:

- a. **Country:** here the can select what country all the tables and figures will refer to. The default should be the entire sample (I called it “overall”).
- b. **Outcome:** here the user can select the two outcomes we used in the previous assignment as well: if the child suffers when the mother works and if the job should be given to a national. The selection of this button will influence the first graph in the “Exploration” section as well as the table and graph from “Regression”
- c. **Controls:** This will be a multiple choice input where the user can select sex and education. If these are selected they should be included in the regression model
- d. **Age polynomial:** This is a numeric input. The default and minimum are 1 and the maximum 5. Increasing the number should include multiple polynomials of age in the regression (e.g., selecting 3 -> “age + age<sup>2</sup> + age<sup>3</sup>” should be in the regression)

In addition these inputs you should include a button that creates a html report that includes the information (text, figures and tables) from Shiny. This should also be dynamic and take as parameters the inputs in Shiny.

Grading:

- Overview section (10%)
- Exploration section (20%)
- Regression section (30%)
- Dynamic report (20%)
- Repo presentation (10%)
- Publish Shiny app (10%) (include the link to the app in the readme file)

Thank you for taking this course. I really hope it was useful.

Good luck!

## Top tips

- To include a polynomial in a regression a useful command is: “poly()”. For example “poly(age, 4)” would include 4 polynomials of age
- to create the regression you can treat the formula as a string and you can concatenate things depending on the inputs chosen by the user
- a easy way to extract the coefficients from a regression is the `tidy()` command from the **broom** package. This creates a table that you can display in Shiny.
- for creating dynamic reports a nice little example can be found here: <https://shiny.rstudio.com/articles/generating-reports.html>
- for the dynamic report you want to recreate the sections and outputs from shiny in markdown. You can give inputs from Shiny as parameters for the report. Look at assignment 2 for some inspiration
- It is easy to publish a shiny app from rstudio. After you run the app just press the “publish” button. You will need to make an account on: <https://www.shinyapps.io/>
- If is fine it the dynamic creating of the report works on your computer but not on the shinyapps.io (it can be tricky to set-up on the server).