

# Workshop 3: Instructions and example report

Group\_member\_1 and group\_member\_2

2022-10-25

- Instructions
- Workshop 3 Assignment: Colombia Motorcycle Accidents data
  - Age distribution of fatalities per sex group
    - Areas of improvement
  - Distribution of roles, per sex group

## Instructions

Welcome!

Today's exercise will be done in groups of 2 to 4. You are allowed to pick your own partners. Once you have found a group, you can go into any of the small pair rooms to work.

Each member in the group should follow the instructions below:

1. First download the course repo [here](#). (You will mostly work locally today). Unzip the downloaded folder, then click on the ".Rproj" file to open the project in RStudio. From the Files pane of RStudio, open the "week\_03" folder.
2. In the "rmd" sub-folder, the instructions for your exercise are outlined (these are the same instructions you see here).
3. **Each group should pick one of the datasets** in the "data" folder. (You can read through the document titled "00\_info\_about\_each\_dataset" to get information about these datasets.)
4. Next, each group member should **select one categorical variable** from their chosen dataset. Their task will be to **create a short R-Markdown-based HTML report showing the frequency distribution of the chosen variable across two sexes**.

For example, Jane and John pick the India TB dataset. Jane looks at the frequency distribution of the *education* variable for men and women. And John looks at the distribution of the *employment* variable for men and women.

5. You can do the initial work on their own, but the final document for submission will be a single HTML file containing a section for each chosen categorical variable.

For example, Jane and John will submit an HTML document with two sections: the first section (primarily done by Jane) on the distribution of the education variable for both sexes, and the second section (primarily done by John) covering the distribution of the employment variable for both sexes.

**6. Each section of the report must contain these four things:**

- a. A plot created with `{ggplot2}/{esquisse}`
- b. A table created with `{flextable}` (See the flextable [book](#) for tips)
- c. At least one use of *inline R code* within the Rmd.
- d. At least one possible *area of improvement* mentioned.

7. As noted above, since RStudio cloud caused some problems in the last session (specifically with `{esquisse}`), it is recommended that you first work on your local RStudio, then when it is time to combine your work, you can:

- a. both copy your Rmd code into a document in your pod folder on RStudio Cloud, and do the final render from there;
- b. copy the material onto one of the group member's computers, and perform the final render from there.

8. **To submit your work** when you are done, you should share your document online, using the [Rpubs](#) service. This can be done as described in the video [here](#). The link to this Rpubs page should be posted as a comment on our lesson page.

9. \*Around 7:20pm UTC+2, a single person from the group will present the work done so far. Your work does not have to be finished by this time. You'll simply present what you succeeded at doing and what you struggled with.

10. The final due date is **Tuesday, Nov 1 at 23:59pm UTC+2**. You are encouraged to visit one of our study halls if you need assistance with this later.\*

Finally, note that your work will be judged simply on whether you have met the four requirements mentioned above; it does not have to be an amazing document. Just follow the instructions and you'll get full marks!

The rest of this document is an example of what one section of a report might look like.

# Workshop 3 Assignment: Colombia Motorcycle Accidents data

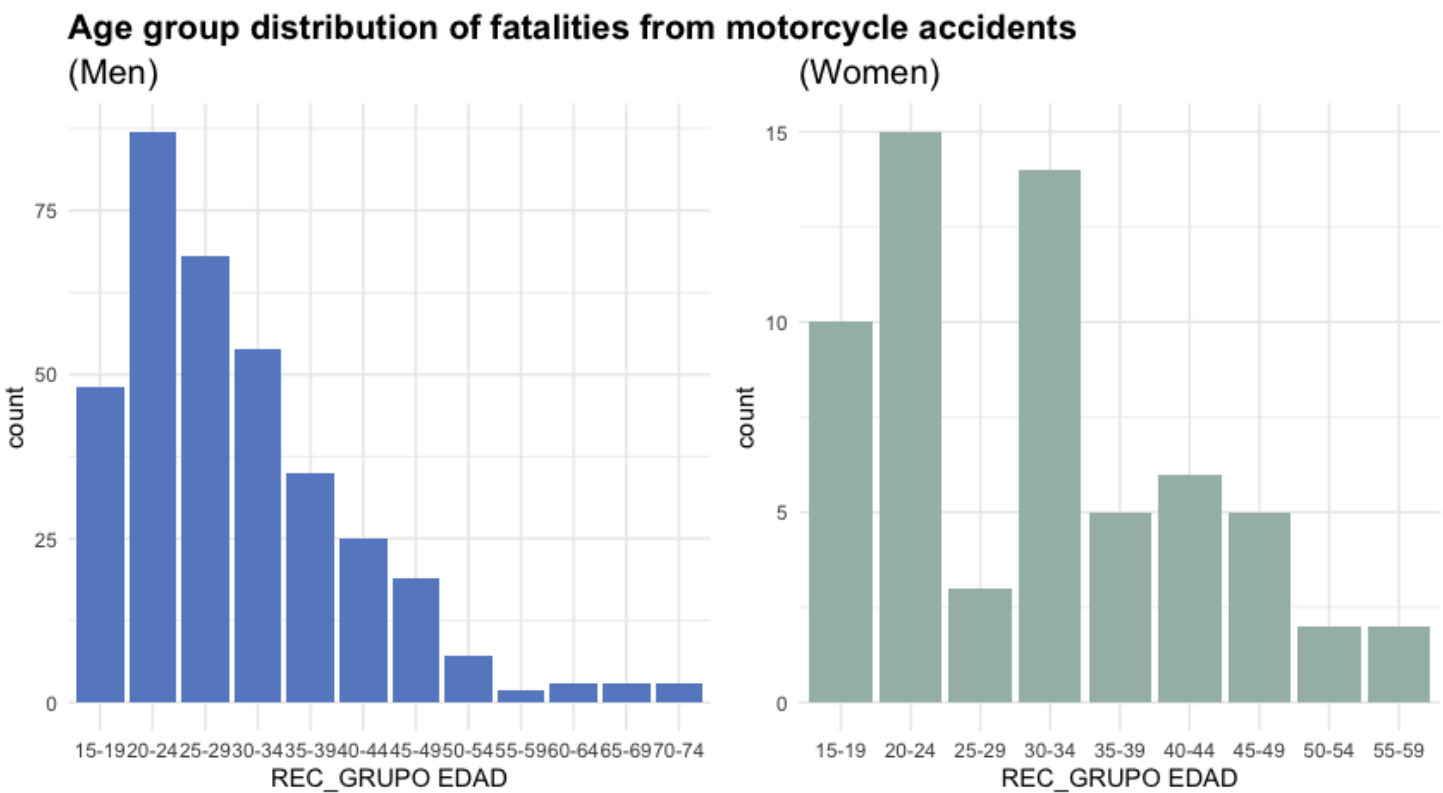
## Age distribution of fatalities per sex group

Work primarily done by group member 1, Jane Doe

The dataset we chose provides information on deaths due to motorcycle accidents in Medellín, a Colombian city, as recorded in medical and police certificates.

I chose to look at the difference in the age distribution for male and female victims.

The plot below shows the age distributions for each sex:



Teacher commentary: We have not yet learned how to create dodged/faceted bar charts, which would be best for showing differences between genders. So for now, you should use the + and / syntax from the patchwork package to combine two plots, one for each sex. This is shown in the code chunk above. (Note that if you are reading this from the HTML document, you cannot see the code chunks. You need to look at the source Rmd file for this.)

And the table below shows similar information:

Age group distribution of fatalities from motorcycle

accidents		
REC_GRUPO EDAD	Femenino	Masculino
15-19	10	48
20-24	15	87
25-29	3	68
30-34	14	54
35-39	5	35
40-44	6	25
45-49	5	19
50-54	2	7
55-59	2	2
60-64	0	3
65-69	0	3
70-74	0	3

For both sexes, the age group with the most deaths was the 20 to 24 age group. In women, there were 15 deaths in this age group, and in men there were 87 deaths.

*Teacher commentary: We have not yet learned how to extract specific slices of data from a data frame. So for now, you should use the syntax demonstrated above: within a pair of square brackets, place the row number, a comma, then the column number. (Note that if you are reading this from the HTML document, you cannot see the code chunks. You need to look at the source Rmd file for this.)*

## Areas of improvement

- The two bar charts have different numbers of bars. This is not ideal for comparisons.
- The age group labels are too horizontally compressed.
- I do not yet know how to change the variable names, so I left in the `REC_GRUPO EDAD` name in the table and plot.

## Distribution of roles, per sex group

**Work primarily done by group member 2, John Doe...**

*Teacher commentary: As already mentioned, your final document should contain all the sections from the different group members. Note that once the work has been done for one variable, it is easy to copy and paste the code to reproduce the analyses on another variable. This makes it easy to help out your group members who are struggling!*