

Project 1 Instructions

Please use the project template Quarto document to complete your project. The rendered document (as a pdf) *and* the raw Quarto file (as .qmd) must be submitted to Canvas by 11:00pm on Thurs., **Feb 13, 2025**. These two documents will be graded jointly, so they must be consistent (as in, don't change the Quarto file without also updating the pdf document!).

All results presented *must* have corresponding code, and the code should be visible in the final generated pdf for ease of grading. **Any answers/results given without the corresponding R code that generated the result will be considered absent.** All code reported in your final project document should work properly. Please do not include any extraneous code or code which produces error messages. (Code which produces warnings is acceptable, as long as you understand what the warnings mean and explain this.)

For this project, you will be using an Olympic Games dataset, which is a compilation of records for athletes that have competed in the Olympics from Athens 1896 to Rio 2016.

Each record contains information including the name of the athlete (name), their sex, their age, their height, their weight, their team, their nationality (noc), the games at which they played, the year, the olympic season, the city where the olympics took place, the sport, the name of the event (event), the decade during which the Olympics took place (decade), whether or not the athlete won a gold medal (gold), whether or not the athlete won any medal (medalist) and if the athlete won "Gold", "Silver", "Bronze" or received "no medal" (medal). More information about the dataset can be found at <https://github.com/rfordatascience/tidytuesday/blob/master/data/2021/2021-07-27/readme.md>

We will provide you with specific questions to answer and specific instructions on how to answer the questions. The project should be structured as follows:

- Introduction (1–2 paragraphs)
- Approach (2–3 paragraphs)
- Analysis (3–4 code blocks, 3 figures, text/code comments as needed)
- Discussion (1–3 paragraphs)

We encourage you to be concise. A paragraph should typically not be longer than 5 sentences. (But don't count sentences, if you need a few more that's fine!)

You are **not** required to perform any statistical tests in this project, but you may do so if you find it helpful to answer your question.

Instructions

In the Introduction section, write a brief introduction to the dataset, the questions, and what parts of the dataset are necessary to answer the questions. You may repeat some of the information about the dataset provided above, paraphrasing on your own terms. Imagine that your project is

a standalone document and the grader has no prior knowledge of the dataset. You do not need to describe variables that are never used in your analysis.

In the Approach section, describe what types of plots you are going to make to address your questions. For each plot, provide a clear explanation as to why this plot (e.g. boxplot, barplot, histogram, etc.) is best for providing the information you are asking about. (You can draw on the materials provided here for guidance.) **All plots should be of different types, and all should use either color mapping or faceting or both.**

In the Analysis section, provide the code that generates your plots. Use scale functions to provide nice axis labels and guides. You are welcome to use theme functions to customize the appearance of your plot, but you are not required to do so for this project. **All plots must be made with ggplot2.** Do not use base R plotting functions.

In the Discussion section, interpret the results of your analysis. Identify any trends revealed (or not revealed) by the plots. Speculate about why the data looks the way it does.