Project 1 Instructions

Please use the project template R Markdown document to complete your project. The knitted R Markdown document (as a PDF) *and* the raw R Markdown file (as .Rmd) must be submitted to Canvas by 11:00pm on Mon., **Feb 22th, 2021**. These two documents will be graded jointly, so they must be consistent (as in, don’t change the R Markdown file without also updating the knitted document!).

All results presented *must* have corresponding code. **Any answers/results given without the corresponding R code that generated the result will be considered absent.** To be clear: if you do calculations by hand instead of using R and then report the results from the calculations, **you will not receive credit** for those calculations. 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.)

For this project, you will be using a dataset about Himalayan expeditions, taken from the Himalayan Database, a compilation of records for all expeditions that have climbed in the Nepal Himalaya. The dataset members contains records for all individuals who participated in expeditions from 1905 through Spring 2019 to more than 465 significant peaks in Nepal. The dataset members\_everest contains records specifically for expeditions to Mount Everest since 1960 and expedition members of known age.

Each record contains information including the name of the mountain (peak\_name), the year of the expedition (year), the season (season), the age of the expedition member (age), their citizenship (citizenship), whether they used oxygen (oxygen\_used), and whether they successfully summitted the peak (success). More information about the dataset can be found at <https://github.com/rfordatascience/tidytuesday/blob/master/data/2020/2020-09-22/readme.md> and <https://www.himalayandatabase.com/>.

This project consists of two parts. Each part should be structured as follows:

* Introduction (1–2 paragraphs)
* Approach (1–2 paragraphs)
* Analysis (2–3 code blocks, 2 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.

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.

Part 1 Instructions

In Part 1, we provide you with a specific question to answer and with specific instructions on how to answer the question.

In the Introduction section, write a brief introduction to the dataset, the question, and what parts of the dataset are necessary to answer the question. 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.

In the Approach section, describe what types of plots you are going to make to address your question. 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](https://clauswilke.com/dataviz/directory-of-visualizations.html) for guidance.) The two plots should be of different types, and at least one of the two plots needs to use either color mapping or facets.

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. 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.

Part 2 Instructions

In Part 2, you will supply the question (using either the members or the members\_everest dataset, at your option) and the approach. Your question cannot repeat the material from Part 1.

In answering your question, follow the instructions of Part 1. For the Introduction, you do not need to repeat the whole dataset description from Part 1, but you do need to describe the columns required to answer your question. **Important: Your plots for Part 2 must be of a different type than the ones you made for Part 1.**

Answer your question by interpreting your plots and identifying any trends they reveal, or do not reveal, as the case may be.