The goal of this assessment is to capture what you have been learning in the lectures 2-4 and workshops 1-3. Use the notes from the lecture and apply the new coding techniques that you have learned to your project data. The R folder with your project, .qmd file (i.e. Quarto document), and project data should be set up to run on your DESKTOP. The **file and your project should be named: ‘STUDENT\_FIRST\_LAST\_NAME\_STUDENT#’**. For example, my file structure would look like:



What you will need in this folder is an R project, a Quarto document, a Word document that renders from the Quarto file, and a data folder that contains your project data that you’ll be cleaning. Your file (e.g Kristoffer\_Wild\_S007341) that contains all of these documents should be zipped ([how to compress a file using windows or mac](https://www.youtube.com/watch?v=aw7awCfBngQ)) and turned in on CANVAS. You will be marked on the following:

Your Quarto document will contain the following ‘main headers’ THIS ORDER, and with each heading accounting for a percentage of your total mark of 20%:

1. **Libraries (1%)**: Importing the appropriate libraries needed for the data wrangling. For example:

Library(stringr) # package used to deal with strings

Library(lubridate) # package used for cleaning up dates and times

1. **Lecture notes (5%)**: Here you will **identify 10 functions covered in the lectures 2-4**, with an example of their use. Note this is not with your project data. In the Quarto document, ABOVE the code chunk of your example, you will document what the function is doing and how it works. Be sure each example is clearly numbered. Here is how this section would look:

A screenshot of a computer

Description automatically generated

1. **Project cleaning protocol (10%):** You must provide 4 examples of how you cleaned your project data using dplyr or other relevant R functions. At least 3 examples must use pipes (%>%) to combine multiple functions within a single code chunk.

Each example should address one of the six data exploration steps outlined in class from Zuur et al. (2010), [*A protocol for data exploration to avoid common statistical problems*](https://besjournals.onlinelibrary.wiley.com/doi/10.1111/j.2041-210X.2009.00001.x). These steps include:

* 1. Outlier detection
  2. Homogeneity of variance
  3. Normality of the data
  4. Zero inflation or excessive zeros
  5. Collinearity among covariates
  6. Relationships between Y and X variables

At least one of your examples must clearly indicate which of these six steps your cleaning process addresses. This should be mentioned explicitly in part (a).

For each example, include a complete sentence above the code chunk covering:

1. What was wrong with the data initially (or the check you are doing), and which of the six Zuur et al. (2010) steps this corresponds to.
2. What function(s) you used and a brief description of what they do
3. What the entire code chunk accomplishes from start to finish
4. Use a basic visual check (e.g. head(), plot(), boxplot(), or hist()) to show how the final data is cleaner or easier to work with

On the next page is an example of how a response should look for a project cleaning and test protocol once you render the file. Note that I have clearly answered points A-C above the code chunk. Notice once the file renders, the code is shown, and the final object is shown (there are options in the .rmd file to be sure this information shows! Finally, notice there are no messages or warnings shown - be sure to turn these options off. Use this example as a guide for your own data.



1. **File formatting and how it is saved (3%):** We are looking for you to correctly set up your project structure using the structure covered in class and the example above. **Your project should be able to run on anyone’s desktop**, i.e. I can unzip the file, open your project, open the Quarto document, and it should **render into a Word document**. Go through the following checklist
   1. Is the file and project name correct?
   2. Is the file saved as a zipped document?
   3. Are the data in the right folder?
   4. Do the paths point to the data?
   5. Does it render into a Word document?
   6. Do you have the correct headings: **1) Libraries; 2) Lecture notes; 3) Cleaning protocol**
   7. Do the outputs for the code chunks show the loaded libraries, functions you’re explaining and your project cleaning examples when the file is rendered to a Word document?
2. **Spelling, grammar, and English expression are being used correctly (1%)**