

# Independent Project - Assignment 4 - Original data

[BioSCI 1120 |Independent Research Project

- xx Points (y% of total points for paper)
- Assigned:
- Due:

## Introduction

The goal of this assignment is confirm that you have obtained a copy of your raw data and to provide basic descriptions of what needs to be done to this data in order for it to be moved in to R. This assignment should also help you identify if you need to reformat the data so that you can do your analysis in R.

## Tasks

Share a copy of the raw data or at least a screen grab of the spreadsheet. Then answer the following questions about your data set. Some of these questions might have very short answers (“No” or “does not apply”). Feel free to say “I don’t know” but follow up on this with your best guess.

1. **Merging:** Is all of the data on a single worksheet, or is it spread across multiple worksheets/spreadsheets.
2. **Layout:** What is the layout of the data: wide, long, or unknown? What layout should it be in for analysis in R?
  - Is any data encoded in the column names?
  - Does each row contain data on a single observation, eg data for a single animal that provides a single data point?
3. **Filtering:** Do you need all of the columns in the dataset or are some not needed?
4. **Subsetting:** Do you need all of the rows in the dataset or are some not needed?
5. **Reshaping:** What needs to be done to reshape the data to be used in R (such as switch from long to wide format)? Do you know how to make these changes or are you unsure?

If you are unsure about your answers to any of these we can discuss them. You should also consult these papers (We can talk first, but I’ll direct you to them also)

Broman, KW and K Woo. 2018. Data organization in spreadsheet. The American Statistician. <https://doi.org/10.1080/00031305.2017.1375989>

Goodman et al. 2014. Ten Simple Rules for the Care and Feeding of Scientific Data. PLoS Computational Biology. <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003542>

Ellis, SE and JT Leek. 2018. How to share data for collaboration. The American Statistician. <https://doi.org/10.1080/00031305.2017.1375987>