

**Introduction** Please answer these introductory questions for your case study in a few sentences.

- 1) Who are you and what is your research field? Include your name, affiliation, discipline, and the background or context of your overall research that is necessary specifically to introduce your specific case study.
- 2) Define what the term “reproducibility” means to you generally and/or in the particular context of your case study.

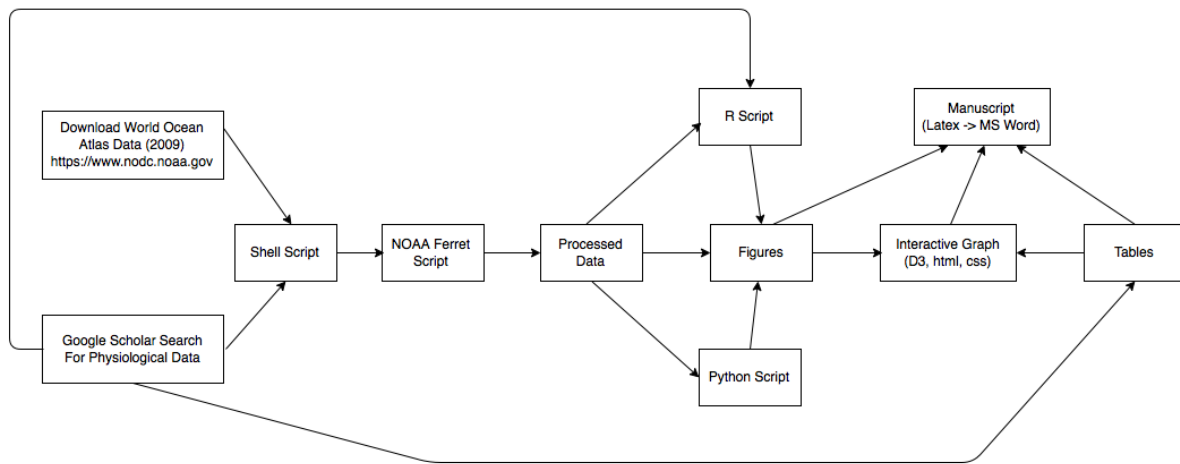


Figure 1: Allison\_Workflow

## Workflow diagram

**Workflow narrative** World Ocean Atlas 2009 (WOA09) data is publicly available through the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information. The WOA09 data is on a geographic grid and available in several different file formats. I downloaded the netCDF (.nc) file format because the netCDF files are relatively small in size and tools are available to read and manipulate the files. NOAA Ferret is publicly available software for visualization and analysis and has built-in functions designed for efficient processing of gridded data. One of the objectives of the analysis was to vary two input parameters and determine the effect on a geographic characteristics. I automated the process of creating ferret scripts with different input parameters by writing a shell script that generated and processed ferret scripts (.jnl files). The processed files were saved as netCDF files.

R and Python were used to create the figures although it would have been possible to create all the figures in one or the other of these software packages. I learned R before Python so I tend to do most of my analysis in R because I am most familiar with it. However, I like the tools for making geographic plots in Python so I used Python to make the geographic plots.

The physiological data was gleaned from published studies found through Google Scholar searches for key words. As the data was extracted from the papers, it was put into a spreadsheet. The data of interest was often available in tables and the text of the papers. In some cases, the data was only available in scatter plots. The ruler tool in Adobe Photoshop was used to manually extract the data from plots. Once all the information from the studies was entered into the spreadsheet, the data from the spreadsheet was saved as a

tab-delimited text file. Additionally, the information from the studies was plotted in a scatter plot in R and each point in the plot had a number associated with it. Information about the individual points was put into two tables. The information is available, but inefficient to obtain by looking between the plot and the two tables. Therefore, an interactive plot that includes the information from the tables was created in D3.

**Pain points** *Describe in detail the steps of a reproducible workflow which you consider to be particularly painful. How do you handle these? How do you avoid them? (200-400 words)*

**Key benefits** *Discuss one or several sections of your workflow that you feel makes your approach better than the “normal” non-reproducible workflow that others might use in your field. What does your workflow do better than the one used by your lesser-skilled colleagues and students, and why? What would you want them to learn from your example? (200-400 words)*

**Key tools** *If applicable, provide a detailed description of a particular specialized tool that plays a key role in making your workflow reproducible, if you think that the tool might be of broader interest or relevance to a general audience. (200-400 words)*

**General questions about reproducibility** *Please provide short answers (a few sentences each) to these general questions about reproducibility and scientific research. Rough ideas are appropriate here, as these will not be published with the case study. Please feel free to answer all or only some of these questions.*

- 1) Why do you think that reproducibility in your domain is important?
- 2) How or where did you learn the reproducible practices described in your case study? Mentors, classes, workshops, etc.
- 3) What do you see as the major pitfalls to doing reproducible research in your domain, and do you have any suggestions for working around these? Examples could include legal, logistical, human, or technical challenges.
- 4) What do you view as the major incentives for doing reproducible research?
- 5) Are there any broad reproducibility best practices that you’d recommend for researchers in your field?
- 6) Would you recommend any specific websites, training courses, or books for learning more about reproducibility?