The Anatomy of a Soup-to-Nuts Exercise

This document is intended to help applicants to the January 2020 [Curriculum Writing Retreat](#) understand the nature and purposes of soup-to-nuts exercises.

If the description of soup-to-nuts exercises presented here strikes you as a bit long or complicated, please don't let that discourage you from applying to participate in the retreat!

We provide this information to be sure applicants know what they should be aiming for as they develop ideas for an exercise, but we do not expect the exercise descriptions that candidates submit with their applications to embody or address all of the properties described below.

The idea of soup-to-nuts exercises

[Soup-to-nuts](#) exercises are lab projects or homework assignments that guide students through the key steps of data processing and analysis involved in writing an original research paper: from the very beginning when they first access the source data for a project, through cleaning and processing the data to prepare them for analysis, to the analysis that generates the results presented in a written report. Their distinguishing characteristics are that:

- They emphasize quantitative research methods that ensure transparency and reproducibility.
- They can be completed in a short time—as little as a single lab or class period, and no more than about two weeks.

Soup-to-nuts exercises are designed primarily for instructors who wish to introduce their students to essential principles and practices of reproducible research, but face time or resource constraints that make it impossible to assign full-blown, semester-long research projects.

In addition, for classes in which students do write independent research papers, assigning one or two soup-to-nuts exercises at the beginning of the semester is an effective way of introducing them to the reproducible methods they will be following as they conduct research for their own projects over the course of the semester.

A prototype soup-to-nuts exercise, which we refer to as the Alcohol Exercise, provides a complete example.

Common features of soup-to-nuts exercises

Some features of the Alcohol Exercise that would typically (though not necessarily) be present in a soup-to-nuts exercise are as follows:

- Students access primary research data through a reliable data provider.
- They engage with the documentation provided by the provider to learn about the structure and content of the data, and do some initial exploration to familiarize themselves with the data.
- They write scripts that process the data as necessary to prepare them for analysis.
- They construct a Data Appendix—a codebook for the analysis data—and save scripts that generate the computational output presented in the Data Appendix.
- They write scripts that execute the analysis that generates the results of the exercise.
- They present the results in a report or answers to questions about the data and analysis.

A crucial feature of soup-to-nuts exercises is that they emphasize the construction of comprehensive documentation that ensures transparency and reproducibility. This documentation typically includes:

- Copies of the original data files, as well as the documentation supplied by the data provider.
- Copies of the scripts that carry out the data processing, generate the computational output shown in the Data Appendix, and execute the analysis that generates the results.
- Output files that capture the results as they are produced.
- A read me file that provides an overview of all the above documentation, and gives instructions for using that documentation to reproduce the results of the exercise.

This documentation should typically be stored on a server of file sharing platform (e.g., Dropbox, Box, Google Drive, the Open Science Framework (OSF), or GitHub) where both the students completing the exercise and the instructor can access them.

The exercise should guide the students through the preparation of all this documentation incrementally as they work on the exercise; creating the documentation should not be a task that students put off until they have finished everything else.

The workflow for doing the exercise and the completed documentation that students submit at the end should embody three related features:

- At the beginning of the project, students construct a hierarchy of folders in which they keep their files while working on the project, and preserve the complete documentation when they have finished the exercise. The structure of this hierarchy should be established at the outset, and then kept fixed.

- When students write scripts, they should be conscious of which folder should be designated as the working directory whenever a user runs the script. There should be a comment at the top of each script indicating what the working directory should be. We suggest that one folder be chosen as the one that is set as the working directory when any of the scripts is run, and that this folder should remain the working directory at all times (i.e., commands that change the working directory should be avoided).

- In the scripts, when writing commands that open up files that are not stored in the working directory or saving files in folders other than the working directory, folder locations should be specified using relative directory paths (so that it is not necessary to change the working directory).

Flexibility and variety in soup-to-nuts exercises

The conventions described above place a fair amount of structure on the exercises, but within this general structure there is great deal of flexibility in many dimensions. For example:

- Exercises can represent any quantitative discipline or field of study, any topic or subject matter, any kind of data, and any analytical methods

- They may be written for any kind of programmable software, using either a "copy and paste" workflow or a "dynamic documents" (e.g., Markdown) workflow.

- They may involve elementary and/or advanced methods of coding and statistical analysis

- They may give detailed, step-by-step instructions for students to follow (the Alcohol Exercise is an example of this style), or the instructions for working with the data may be more conceptual, leaving it up to students to figure out the details of implementation.

In general, provided the key principles and practices related to transparency and reproducibility are maintained, instructors should feel free to design the exercises in any way they believe they will be effective for some particular contexts or purposes.

Components of a soup-to-nuts exercise

In the repository of soup-to-nuts exercises that Project TIER is developing, each exercise will consist of several components:

- The exercise (i.e., the instructions given to students)
- A completely worked out "solution" showing an example of the report students would turn in along with all the reproduction documentation
- Notes to instructors
- A cover sheet with bibliographic information

The Alcohol Exercise provides illustrations of all these components.