

PSTAT 100 Project plan report

This is a guide to preparing your project plan. It functions both as a guide to the work you'll need to do and as a guide to preparing the deliverable. You can use it as a template to draft the plan report; if so, please remove the text explanations of each section.

While you may find it useful initially to follow the outline given, you do not need to adhere to it exactly -- you're free to organize your submission in the way that seems most natural to you. However, please do keep the high-level sections, so that your report includes the following headers:

1. Background
2. Data description
3. Initial explorations
4. Planned work

Your report does not need to be long. It should be about 2-4 pages, and may not be much longer than this template once you replace the guiding text with your own work.

Group information

Group members:

Contributions:

1. Member 1 studied the data documentation and prepared the data description.
2. Member 2 worked on tidying the dataset.
3. Member 3 ...

0. Background

This section should introduce your reader to the general topic you're engaging with in your project and explain any specialized knowledge that they may need to understand your dataset and why it's interesting. It doesn't need to be long, but should touch on the following points:

- Introduce the topic of your project.
- What area or areas of study are you in dialogue with for your project?
- What is your data about, broadly?
- What is the motivation for collecting the kind of data you're working with, and what sorts of things could you potentially learn?

You can look to the background sections in the homework assignments for examples. (There you can also see how to include images in your notebook.) The background sections of the homeworks are usually short and focused paragraphs intended to orient you to what you'll do in the assignment. They don't go into a lot of detail -- just enough to (hopefully) convince you that the data are interesting and explain any terminology or general information you may not know.

You may find it useful to write up the data description first, think about what the reader should know before they peek at your dataset, and then come back to the background section. I often write the background sections of your assignments last, once I have a sense of what kind of information would be most useful going into the assignment.

Start your draft here.

1. Data description

This section should introduce your dataset in detail. It should reflect your having gone through the collect/acquaint/tidy stages of the lifecycle. Below I've provided you with an outline. You do not need to adhere to this strictly -- in fact, it would be more natural to divide the items among a few short paragraphs -- but you should touch on each item in a format that suits your project.

Basic information

Help your reader understand what your data is, where it came from, and how it can be used. Provide the following.

General description: provide a one- or two-sentence description of the data right at the beginning. For instance, "The data are diatom counts sampled from evenly-spaced depths in a sediment core from the gulf of California." Nothing too complicated, just something to give your reader a sense of the 'what' right off the bat.

Source: indicate where your data came from. Provide a verbal description -- who collected it as part of what project and where -- and either a citation or a hyperlink.

Collection methods: How were the data values obtained? Provide a simple description of how measurements were taken (using scientific equipment? web scraping? surveys?).

Sampling design and scope of inference: Indicate the relevant population. If identifiable from data documentation, state the sampling frame and sampling mechanism and indicate the scope of inference. If no information is available about the sampling design, indicate this instead, and discuss the extent to which having no scope of inference is a limitation for the particular topic you're investigating.

Start your draft here.

Data semantics and structure

Units and observations: State the observational units.

Variable descriptions: Provide a table of variable descriptions. If your dataset is large and you'll only work with a subset of the total available variables, limit your attention to the variables that you'll work with. Here's a template you can work with:

Name	Variable description	Type	Units of measurement
GCLASS	Graduating class	Numeric	Calendar year

Example rows: Print a few example rows of your dataset in tidy format. Please don't include the codes you used to manipulate the raw data. Do that in a separate notebook and export the result to a .csv file -- `data.to_csv('tidy-data.csv')` -- to load directly into the cell below.

Start your draft here.

In [3]:

```
# Load tidied data and print rows
```

2. Initial explorations

At this stage, you may spend most of your effort on the computing side tidying up the data. You're not expected to complete a thorough exploratory analysis, and if your dataset was especially messy to start with, you may not even begin your exploratory analysis by the time you prepare this report. You have the option to leave exploration for the next stage of work and simply report basic properties of the dataset, but you should at minimum address the items in the 'basic properties' section below.

Basic properties of the dataset

Help the reader get acquainted with your dataset on a simple level by identifying characteristics of the dataset and variable summaries. Some amount of code is fine here, but try to use code cells sparingly.

Dimensions: state the dimensions of the data (in tidy format, of course).

Missing values: Are there missing values? If so, why are they missing?

Variable summaries: Provide simple variable summaries for the most important variables in your dataset. Preferably, you'll do this for all variables, but if you have a large number, you might need to prioritize and focus on the ones most of interest. What exactly you do is a little case-specific, but think of things like means and variances, min/max, number of levels and observation counts for categorical variables, etc.

Start your draft here.

Exploratory analysis

If you were lucky and your dataset was neat, you should aim to include a few exploratory plots or tables here -- they don't need to be polished at this stage, but you should select plots that are informative (rather than including all plots you may have looked at).

If you do include exploratory graphics or tables, please explain in a sentence or two what each one shows. Try to include a minimum of code. Consider [saving your plots as images](#) and inputting images into markdown cells instead of generating them anew via code cells.

3. Planned work

Here you should indicate your tentative ideas for your analysis. Don't worry, these aren't final -- you can always change your mind later or shift gears if they don't pan out. The objective is to have you start thinking ahead about what you'll do.

Questions

Please propose two focused questions that you plan to explore.

- 1. Question 1 here
- 2. Question 2 here

Proposed approaches

For each question, please describe an idea or two about how you might approach the question.

- 1. Approach 1 here
- 2. Approach 2 here

Submission Checklist

1. Save file to confirm all changes are on disk
2. Run *Kernel > Restart & Run All* to execute all code from top to bottom
3. Save file again to write any new output to disk
4. Select *File > Download as > HTML*.
5. Open in Google Chrome and print to PDF on A3 paper in portrait orientation.
6. Submit to Gradescope