# Data Bootcamp: Project Guide

Revised: December 14, 2015

## Overview

One of our goals is for you to produce a piece of work that you can use to demonstrate your data and programming skill set to potential employers. That work will come in the form of an IPython/Jupyter notebook, a format that combines code, text, and graphics in one user-friendly document. We will add them to our GitHub repository so that you can use a link to show others what you've done. You will also be able to see what your classmates have done.

The relative absence of structure in self-directed projects like this makes them more challenging than most things you'll do in in school. But it's also an opportunity to use your creativity.

We have divided the project into a number of smaller steps to keep you on track. The intent is to make the project easier, by breaking it down into a number of small, more manageable sub-projects. The early steps are graded only on whether you do them: you get 5 points if you do them, none if you don't. We start with individual work. Partway along, we encourage you to form groups of two or three. If you'd prefer not to form a group, that's ok, too.

Specifics follow.

## Project outline

The project steps are

| Assignment | Format | Individual/Group | Points |
|---|---|---|---|
| Three Initial Project Ideas | Paper (hand-written) | Individual | 0 |
| Three Revised Project Ideas | Paper (professional) | Individual | 5 |
| Project Proposal | Paper (professional) | Group | 5 |
| Data Report | Paper (professional) | Group | 5 |
| Final Project | IPython/Jupyter notebook | Group | 85 |

The due dates for each one are posted on the course website. **Dates are firm and not open to negotiation.**

## Project specifics

The steps in the project production process are

- **Three initial project ideas.** Write down three project ideas in class. One or two sentences each is enough. Use your imagination. Be creative. Speak to others. Write down things that interest you. This will not be handed in or graded, but it will give you something to work with later on.

  Over the coming weeks, we recommend you keep track of ideas that cross your mind as you prepare for further idea development.

- **Three revised project ideas.** After giving this more thought, write down three project ideas in slightly more detail. **Include likely data sources.** All together this should be roughly one paragraph per idea and one page overall. Like all work in this course, it should be clean and professional. We will keep the copy you hand in to use in class, so keep a separate copy for yourself.

- **Project proposal.** Form a group of two or three — no more — and choose a single project from those you submitted — or perhaps some other idea if you get a sudden flash of inspiration. Flesh out the project in more detail, including the source of data and two figures you plan to produce with it. Total length should be no more than two pages.

  We stress: you should **be clear about the data** you plan to use. If you can't get the data, that can derail your whole project.

- **Data report.** Describe your data and how you accessed it in enough detail that someone else could do it. Include your Python code. Internet sources are preferred, because others can reproduce your results from your code.

- **Final project.** You should submit your IPython notebook to db3@nyu.edu by the start of the last class. The subject line should be: "bootcamp project ug" or "bootcamp project mba" depending on the section you are in. The file name should be your last names separated by dashes and a short title — something like `Jones-Smith-Zhang-India.ipynb`.

  Your project should include:

  – Title, list of authors, and summary.

  – Data entry. Short description of data sources and code to read it and reformat it as needed.

  – A series of figures that tell us something interesting. Three or four would be enough, but do what you think best.

  We'll post a template.

## Free advice

Some things to keep in mind:

- **Keep it simple.** Most project ideas turn out to be too big. You're generally well-advised to carve out a manageable subset of what you think you can do.

- **Find data.** Make sure you can get the data you need. One way to assure this is to start with data and ask what you can do with it. Ideally you want the intersection (picture a Venn diagram) of an interesting idea and good data. You can start with either one, but ideas are often easier to get than data.
- **Ask for help.** We have years of experience with this kind of thing. If you're stuck, let us know and we'll try to help. You can also post a question on the class discussion board. Perhaps one of your classmates has the answer.