

This assignment will get you going with researching your topic and finding examples in the literature so that you can write your literature / background / exposition section of the paper.

First though, we want to keep ourselves organized, so we have a bit of setup to do in our personal repos.

Setup and Revisions:

From the class repo, copy over the Readme.md from the **project** folder in the course content repo into your personal repo **project** folder (remember you are supposed to keep the personal repos organized!). Edit the readme file, adding in your project title and synopsis.

If you need to make minor revisions to your synopsis (from the feedback on the proposal), this is a good time and place to do it. Prof. Wagaman will review these when checking the repos for reasonable organization during Thanksgiving Break.

The timeline for the project is provided as a reference in the readme. As you submit assignments, you could remove them from the timeline so you continually see what is coming up next as the top item.

The annotated bibliography document will be created in your repo (and a copy submitted to Gradescope for on-time points / completion points). It should remain in your project folder where you can edit it as needed as the project proceeds.

For organization, you may want to create a “Literature” or “References” or “Resources” folder within your **project** folder to put .pdfs of papers, etc. that you find. Alternatively, you can maintain (stable) links to the materials in a separate file or the bibliography file itself. Be sure you have a system that you are comfortable with.

What is an annotated bibliography?

https://owl.purdue.edu/owl/general_writing/common_writing_assignments/annotated_bibliographies/index.html

Basically, it's a bibliography (list of sources with appropriate citations) with notes (annotated) / comments about the source. This could range from what material it covers to how you think it might be incorporated into your work. The notes / comments mean that you are going a step beyond just finding the sources and getting the citations. It shows you've started to evaluate them and understand how each could be useful to you. The Purdue site describes it as starting to think more critically about the work and what you are getting out of each source. This helps you identify gaps where you need to find more sources of information.

What citation style should I use?

Your choice, but be consistent. I'm including a reference to the ASA's (American Statistical Association) style guide if you want to use that citation style (mix of MLA and APA), or either MLA or APA would be appropriate. My example on the next page uses APA. Feel free to use Zotero or another program to organize your references (you'll still need a separate annotated bibliography file to submit). The

template that will be provided for the final paper (if you want to use it) will require a bibtex file for the references.

What goes in an annotated bibliography entry?

First, you want a complete citation for the work in the appropriate format. (Pick a style and stick with it).

Next, after reviewing the work (you don't have to read fully, though that can be useful, but you need to skim / look at key parts, etc.), you'll need to write the **annotation** that you want to go along with the text. This could include any of the following pieces of information based on the work and what you got out of it:

- A summary of the relevant information in the work (e.g. this work is a great summary of the development of topic Y, covering ...)
- A description of where useful information is in the book/source (e.g. great chapter 2 that introduces topic X, example 5 on page 12 was really useful to understanding Y)
- An assessment of the material (e.g. The paper has a nice intro to the topic, but it's not presented clearly, or the source here doesn't seem to be as reliable as my other sources, so I'm not likely to use this.)
- How you expect / intend to use the information from the work (e.g. Figure 3 is a great visual that I want to use to illustrate the topic, or the applied example is one I want to feature in my literature review, or this is great historical background for my exposition)

You want to record thoughts about each source that are going to help you figure out how to use the information it contains in your write-up (primarily the exposition part). As you work, you may find that you can easily find examples but don't have enough background/introductory material. Keeping track of what you have (instead of only the list of the sources) will help you fill in gaps for what you need to construct the literature / background / exposition.

What is the structure of the annotated bibliography?

Whatever you want as long as you keep both the citation and annotation for each source/work together. The Purdue website has examples, and an example using two papers related to methods from class (we only read the first one) is below. This example uses APA format for the citations.

1) Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.

This is the original paper introducing the LASSO. The paper is written at a level that an advanced undergraduate could tackle. Figure 2 is useful for describing how the LASSO enables coefficients to be set to 0. The description of the data set in the example could be improved. Has useful notation and background motivation for the LASSO that would be good to use in an introduction to the topic.

2) Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123-140.

This is the original paper introducing the concept of Bagging (in the context of classification and regression trees). The paper has a reasonable introduction to the concept, listing previous sources that discussing tree “averaging” instead of bagging. It applies bagging to many example data sets and demonstrates the improvement that it has over classification trees, in particular. It has examples for both medium and large data sets, and includes some for regression (but focus is classification).

Workflow

Even though the annotated bibliography has a structure of citation then annotation, that doesn’t mean that you will only look up one source at a time, stop, review/skim/read it, and craft its annotation, then move to the next source. Although, you could do that.

You need to identify a workflow that works for you. Here is an example (which may or may not work for you – adjust as needed):

1. Identify several potential sources through a literature search. You may grab their citations in a document or something similar (like copy over results from a search).
2. Download their pdfs (if appropriate) and put in your reference folder. Or sketch an entry with a link in your annotated bibliography.
3. Review/skim/read each in turn, getting their citation and annotation into your annotated bibliography. (This may be in a different work session!)
4. Look for “holes” in material you are missing. Tailor a search to find relevant articles/sources to fill the holes, and go back to step 1.

Note: If you look at a source and decide you absolutely will not be using it, you should still keep it in the annotated bibliography, and put that as your annotation! That tells you what you’ve already looked at, and acknowledges the work you put in reviewing it already. You can re-organize entries in the annotated bibliography as you see fit.

What do I need in my annotated bibliography for our class assignment?

(Remember you can still add to the annotated bibliography after submitting our class assignment!)

Ideally, I’d suggest you look for sources in the following categories for the project assignment, as I think these will be the most helpful for you:

- At least one textbook that covers the topic (if the method is not too new)
- The original paper(s) (or an early paper) that introduces the topic (even if it’s at a high level), then find the earliest one you can at a level you can follow (or mostly follow, getting started)
- At least one review paper (in a field of your choice) – e.g. overview paper / introduction to the topic in field X
- At least 3 research papers with applications of the topic (at least one being fairly recent)
- (Assuming there is an R / python package for the method) A paper discussing the software for implementing the method (such as Journal of Statistical software papers)
- At least one blog entry where the topic is implemented

I'm expecting a minimum of 7 sources and their associated annotations in your submission for class.

The resources you submit do not need to follow the suggestions I outlined exactly. I'm trying to help with guidance that could assist in producing a well-rounded submission. You should adjust as you evaluate what sources you need to address gaps in the information you have.

P.S. Asking for a minimum of 7 for this assignment does not mean your final project submission should have exactly 7 references – this is a starting point.

How do I find these sources?

This will vary. Some of you picked topics from CASI so you already have a textbook source, and could see what sources it references and go from there.

You can search via Google Scholar, or the databases provided by the College, such as JSTOR and the Web of Science. Google Scholar and the Web of Science in particular are nice for citation searches (put in one search, find results, from there, see what cited that work, expand out).

Other Resources

Use the resources the library has available. <https://www.amherst.edu/library/find>

For example, there are a variety of citation systems that you can investigate. Zotero is one that the library recommends and runs workshops on. You won't "need" one for our project, but if you expect to be doing more research, it may help to learn a system.

<https://libguides.amherst.edu/citation/zotero>

JSTOR and the Web of Science are available through **A-Z Databases** from the library. There are many other databases as well.

You could meet with a librarian (probably requires scheduling an appointment) or a Peer Research Assistant for assistance too.

The Peer Research Assistants (through the Library) are offering drop-in appointments on Sundays from 3 pm – 11pm, and Monday – Thursday 5-9 pm at the Frost Library reference desk. Find more information [here](#).

And as usual, you can see me for assistance. I know more about some of these topics than others, but I'm happy to assist you and learn along with you.