Assignment: Building, Analyzing, and Presenting a Dataset

Assignment Description:
This project introduces students to the fundamentals of creating a dataset, developing data visualizations (with Palladio, Voyant, and RAWGraphs), working with a content management system (Omeka), and digital imaging, and furthers written and oral presentation skills. Students work with a very small collection of objects (10 or 30), drawn from the physical holdings of the college (archival documents, rare books, artworks) or from social media. This is essentially a proof of concept, and a chance to learn and apply new software (all cost-free). Students can choose any topic they are interested in, and often choose unconventional and non-academic topics – the focus here is on process, so the content and findings are unimportant.

This project takes place over six weeks of the twice-weekly class, with three in-class tutorials (digital imaging (half class), data visualization (Voyant, Palladio, RAWGraphs), and Omeka), three in-class work days (with instructor and peer help), and one presentation day. The remainder of work is completed outside of class, and the remainder of class time is devoted to other coursework.

The tutorials, originally developed by a former course TA, Taylor Elise Mills, are minimal, and the software is accessible. Some students quickly learn the tools, and they become very effective peer teachers. The scale of the assignment is very small and low-stakes, so students are flexible about their topics, and can make changes and corrections easily. The assignment, despite its many pieces, doesn’t actually require a lot of student time. Having assigned this three times, I see that the bulk of student time is spent deciding on a topic (despite assurances that the stakes are low). Some students get to the point of building a dataset, and then scrap the topic and start over (though I always advise against this). Even at the end of a busy semester, this flexibility remains available to them because of the small scale and emphasis on practice over product in this assignment.

Outcomes:
Working with very small datasets, students learn the fundamentals of drawing data out of arts and humanities ‘texts,’ organizing and cleaning data, and data visualization. They generally find they have to reorganize their datasets once they start attempting to create visualizations, and because they are working on such a small scale, they are able to do this easily – they can go back to the tweets or objects to draw more information if needed. Students do get frustrated with how insignificant their visualizations often are at this scale, and that they may not be able to produce good visualizations about what they find most meaningful about their objects/texts, however there are always surprises and new insights. In the final step of the project, the prospectus, students are asked to imagine how they would expand this project, or approach it differently, given unlimited time and resources, which allows them to think creatively about the limitations of their initial project. What is most satisfying about this project, from the instructor perspective, is the deeper discussions about data that develop naturally during this assignment,
as students interrogate data-gathering practices, consider their own subjectivities as researchers, and debate the merits of a data-driven approach to the sometimes data-resistant objects they are working with.

**Broader Context for Assignment**
This assignment is the focus of the second half of the first semester of a two-seminar digital humanities/liberal arts research course, which is the introductory experience in Hope College’s Mellon Scholars Program (2010-2020). The aim of this course is to introduce students to the digital humanities, to hands-on research experience (in the college’s archives, rare books, and art collections), and to provide them with tools and skills to pursue independent research projects in their later coursework.
Assignment Instructions

**Building, Analyzing, and Presenting a Dataset**
(25% of semester grade)
In this project, you will develop a computer-readable dataset based on text and/or images, with archival documents, rare books, a museum collection, or social media posts as your starting point. You will start out with a research question or two that will help determine what information to include in your datasets, and how to organize this information to best understand meaningful relationships among your items. You will use data-analysis tools (hence, ‘computer-readable’) to test your research questions, and possibly develop more questions based on any connections you draw from the analysis. The project will include digital imaging (scanning/photography), transcription or text recognition, spreadsheet organization, data visualization, and collection management via Omeka. The project will also include a prospectus and brief presentation. This project will introduce you to additional digital research methods and tools, through the creation and manipulation of a manageable dataset.

**Topic**
As in all humanities research, you will begin with texts (interpreted broadly – documents or artifacts), and convert them into text that can be analyzed digitally. Your documents will come from one of these four categories, and you should report your chosen subject/documents by November 5:
A. Archives – 10+ items
B. Rare Books – 10+ pages or a chapter from a single book; or 10+ images from one book or several
C. Museum Collection – 10+ items from the campus art museum
D. Social Media – 30+ posts by an account or related to a subject #hashtag

**You will create:**
**Spreadsheet of Dataset** (first draft ready by 11/14, due 12/5, 20% of assignment grade)
A spreadsheet is a type of document to organize data, and can be used to understand relationships between categories of information. You will make a spreadsheet with Excel (Microsoft), Sheets (Google), or Numbers (Apple) to organize your dataset. One category of data should include extensive text, so the full text of each of your items (if applicable), meaning you will create a transcription of the text. Instruction and hands-on help with your spreadsheets will be provided in class.

**Set of Digital Images**
You will collect/create digital images of your documents/artifacts with screenshots, scanning, or photography technology, and develop a naming schema to help organize the image files (jogs) in a folder on your computer (or in Google Drive). Instruction on imaging will be provided in class on 11/7. These images will be uploaded to your Omeka site (see below) so you don’t need to turn in these files separately.

**Data Visualizations** (draft due 11/19, final due 11/26, 10%)
Data visualization is a method of creating visual models from data that aids you in presenting data or determining meaningful relationships and areas for further study. With your complete
dataset, you will run visualizations using the digital tools of Voyant, Palladio, and RAWGraphs. A tutorial on these tools will be held in class on 11/14, and there will also be time in class for hands-on help. These visualizations will be uploaded to your Omeka site (see below).

**Omeka Site** (due 12/5, 30%)
Omeka is a digital database for organizing datasets, as well as a platform for curating this data and sharing it on the web. You will enter the information and images from your dataset into your own Omeka sites (tutorial 11/21 in class). Your Omeka website will include a home page with an abstract of your project, your collection of 10/30 items, an exhibit or page of data visualizations and explanations, and a final page with research questions and possible findings from your data analyses. There will be time in class for hands-on help.

**Presentation** (in class 12/3 or 12/5, 20%)
You will present your project to the class on Dec. 3 or 5, for 5-7 minutes (including time for questions), using your Omeka site as your presentation tool (in place of a Powerpoint or Slides presentation). Your presentation should include an introduction to your topic/documents/objects, a demonstration of your data visualization and explanation of its usefulness, one or more research questions, and the potential significance of this research.

**Prospectus** (due 12/9, 20%)
Imagine that you will be pursuing this project beyond this course. Create a prospectus to argue for the value of your project, following the format for the Prospectus at the end of this document. Nearly all of the intellectual work for creating this prospectus will be complete for your presentation the previous week – this is essentially a text-based summary of your project with an indication of how it might develop further in the future.

**Logistics for item types**

**A. Archives** – 10+ items
You will need to choose 10+ items from the archives, which might be related to your prior project, digitize (scan/photograph) the documents, create a spreadsheet of the metadata (essentially, the bibliographic information), and transcribe the texts (either via typing or optical character recognition (OCR)). If you choose this item type, be in touch with the archives director.

**B. Rare Books** – 10+ pages or a chapter from one book; or 10+ images from one or more books
There are two ways to approach the Rare Books, either as text or as image. In either case, once you chose a book, be in touch with the library to reserve your book. If you choose to work with the book as text, you’ll have to digitize the pages (either scanning with permission, depending on the condition of the book, using the library’s large-format book scanner; or photographing the pages with a camera (your own, or rented from the Media Center, or your phone)), create a spreadsheet of metadata (essentially bibliographic data) and any additional information you find significant to record, including a column for the text you will transcribe (either via typing or optical character recognition (OCR)). You can browse the shelves in the Rare Books room, or browse by subject via the library catalogue. To request a Rare Book, inquire at the circulation desk.
**C. Museum Collection** – 10+ items from the campus art museum
You will need to choose items from the collection, in consultation with the museum director. The collections database is not public, so you’ll have to get the data from the director or the collections manager, and include the parts of this data that are relevant to your project in your spreadsheet. Your spreadsheet should also contain a column with a brief text about each artwork, the equivalent of what you might find on a museum wall (description of the work, its significance or context). You will also need to acquire images of each artwork, either with your own photography or from the museum.

**D. Social Media** – 30+ posts by an individual or related to a subject #hashtag
You’ll need a social media account to find a set of posts to make up your dataset. You’ll want to spend some time playing with the search function to find a set of posts that are interesting and relevant to your project, and that have TEXT. While twitter is essentially text (though people also post images), and Instagram is essentially images (but people also provide text) you will want to consider how essential the visual is to your project – you will need a visual element for your Omeka site, so make screenshots as you work. If you choose to work with social media, you will need to acquire a set of posts, screenshot the posts for a visual record, and create a spreadsheet of metadata (bibliographic data and perhaps other data like location or networking stats) including a column of the actual text of the post.

**Prospectus Instructions**
A research prospectus is a preliminary plan for conducting a study. This is not a detailed, technical research proposal, but, rather, a considered analysis of the issues you are likely to confront in such a study. In essence, it is a *preliminary* proposal. A prospectus in some form is part of any research proposal or grant application. Preparing a prospectus asks you to define a subject matter for inquiry, and imagine how you might undertake research in order to complete a project. Some parts of your prospectus will be concrete, and some will be imaginary or projected. Writing the prospectus will require some research into secondary materials. See *The Craft of Research*, chapters 3-6, for help in developing research topics, questions, and problems.

1. Title
2. Abstract
   a. briefly explain the subject matter, research problem, and your intention (about 100 words)
3. Introduction
   a. Clearly define the research problem, aim, and scope of study. Why is it an exciting and important project? Be persuasive.
   b. Provide appropriate background to set the stage; this may include what led to your interest in this subject
   c. List main research questions and topics that will be addressed by this project
   d. Identify potential benefits of the study to your discipline and/or audience
   e. (optional) Note related digital humanities projects, and how your project will differ enhance/contribute/expand/contradict
4. Literature Review (not required for this assignment)
a. Discuss key relevant secondary literature, providing a list of 4 secondary sources with one (total) paragraph explaining what these sources are and how they are significant for your project. List your sources following MLA citation format. Note: you do not have to have read each of these sources in their entirety – but rather have a basic idea of their coverage.

5. Research Plan
   a. List research questions you aim to answer with this project
   b. Discuss which primary sources you will be consulting for this project (archives, collections, interview subjects, etc), and provide an annotation for each explaining what it is, why it may be significant, and how it may help answer your research questions
   c. Anticipated difficulties and pitfalls - what might go wrong? How will you try to prevent problems or minimize their impact?
   d. Logistics - how will you access your research materials (where are they physically, do you need to travel or use foreign language skills, etc), how long will it take to complete the research? Will you need financial support to do the research (travel funds, etc)
   e. Digital tools - which tools do you anticipate using, and how? Will they assist with the research itself, or with presentation? If you don’t know which tools to use yet, what kind of tools would you like to make use of?

6. Appendix (not required)
   a. Any additional materials that are relevant to your project, such as imagery or sound recordings to convince the reader of the value of your proposed project

Format
- maximum of 3 pages (excluding optional appendix)
- clearly label each numbered section of the prospectus with a heading (can just be Abstract, Appendix, etc) (you don’t need separate headings for the lettered items)
- single-space within sections
- proofread your project carefully

Grading Rubric
Each item (dataset, visualizations, Omeka site, and presentation – see specific rubric for prospectus below) will be graded based on the inclusion of all required elements, the thoughtfulness demonstrated in the organization of your data, the evidence of your engagement and experimentation with the new digital tools and methods introduced in this unit.

A – Excellent – meets the expectation of the assignment with enthusiasm and grace
B – Very good – meets the expectations of the assignment
C – Needs Improvement – attempts to meet the expectations of the assignment, but falls short
D – Fair – some expectations are met, but not satisfactorily
F – Unsatisfactory – no attempt to meet the expectations of the assignment

Rubric for Prospectus

______ Overall Concept
Prospectus presents a project of interest that has been developed through its initial stage

Analysis

Paper shows thoughtful consideration of the potential project and its challenges

Literature Review

Shows evidence of initial research and assessment of possible sources

Fits Assignment Parameters

Within 3 pages, includes all elements of assignment description

Paper Style

Paper was written following standard rules of English grammar
(e.g. no misspellings, no sentence fragments or run-ons)

Total Grade for Assignment: __________ ; __________ %