

### **Dataset Creation -- 15%**

- **Due Friday, March 19**
  - **Submit plan for dataset as part of response paper 4, due Friday, March 12**
- **Dataset + codebook + ~900-1200-word (~3-4 page) reflection**
- **MLA citation style**
- **Turn in via “Dataset Creation” portal on Blackboard Assignments page. You should turn in both your dataset itself and your reflection paper via this portal.**

For your dataset creation assignment, you will create your own (subset of a) dataset and you will write a 900-1200 word (~3-4 double-spaced pages) reflection on this dataset.

You may work with a partner to collect your data and create your dataset. If you work with a partner, you should each aim to collect the minimum number of observations/records (see below). You will each write your own reflection paper.

#### **Planning your dataset**

Your goal is to create a dataset with a minimum of 10 individual observations/records. Your dataset should be geared toward answering questions about what people do and/or the kinds of objects they create (i.e., it should be broadly geared toward answering questions in the humanities and social sciences). It might focus on contemporary “cultural objects” (i.e., books, films, images, tweets, news articles, etc.), or on contemporary social behaviors or actions. Your dataset should also include at least 7 metadata fields, and it should be conceptually meaningful, meaning the observations/records that comprise it should be grouped together logically and according to some explicit criteria. You should think of the ~10 entries you collect yourself as just a subset of a larger dataset (i.e., a proof of concept, NOT the entirety of the dataset itself).

How you collect this data is up to you, but you should take considerations about data collection very seriously when deciding what kind of data you want to collect. You should think hard about what kind of data it will be possible for you to collect in the time that you have to complete this assignment. When deciding what data you want to collect, consider the following criteria:

- You should be able to collect this data ethically and transparently. If collecting data about people and/or human or social behavior, you should collect this data anonymously but, if possible, the people you are collecting data from or with should be aware of your collection efforts. If collecting data about cultural objects, you should be aware of copyright restrictions.
- This dataset should not already exist.
- You should be able to collect this data in a reasonable amount of time (a few days at most) with a reasonable amount of effort.
- You should have ideas about how you would scale up data collection efforts if you had the time (and/or the money) to collect the full dataset (instead of just a subset, like you are doing for this assignment).

You will write up your plan for your dataset for response paper 4, due Friday, March 12. Your plan should include the following information:

1. If you are working with a partner, who your partner is.
2. The title of your dataset.
3. A brief description of the kind of data your dataset will contain and how you plan to collect it (~1-2 paragraphs).
4. The titles and brief descriptions of each metadata field in your dataset (~7 minimum). This may change as you work on your dataset, but you should have an initial plan.
5. The audience for your dataset. Who are you creating this dataset for?
6. 1-2 questions your dataset is designed to help this audience answer.
7. A list of 2-3 already existing related datasets and an explanation of how your dataset offers a unique contribution/how it is different from these existing datasets (~1-2 paragraphs).

### **Creating your dataset**

You can present your dataset in whatever format makes the most sense for your data, but the easiest thing is probably to present your data as an Excel/Google sheets spreadsheet. In this case, each observation/record should be 1 row of your dataset, and your metadata fields should comprise the columns of your spreadsheet. As always, please let me know if you have questions about the best format for presenting your data.

Again, your goal is to collect at least 10 individual observations/records and to fill out the metadata for each observation/record (if you are working with a partner, you should aim for at least 20). If you are working with a partner, you only need to turn in 1 version of your dataset (i.e., only one of you needs to turn in the actual completed dataset). When you turn in your dataset, you should also turn in a codebook, or a list of each metadata field and a brief description of what that field means/the kind of information it records. You can turn this in as a separate list, as a separate tab in your spreadsheet, or in whatever format suits your data the best.

### **Writing your reflection**

After collecting your dataset, you will write a 900-1200 word reflection on the process and the finished product. You may organize your reflection how you choose, but it should contain the following elements:

1. Revision of what you wrote in your plan. If, after creating your dataset, your thinking about anything you wrote about in response paper 4 (see above) has changed, discuss those revisions/changes here.
2. A reflection on 1-2 issues, problems, or larger concepts that collecting this data helped you to see or to think about more clearly. What did this process illuminate for you, either about the data you chose to collect specifically or the process of data collection more generally or the concept of data itself? You should relate this discussion to at least 1 of our readings from class so far (you should NOT discuss the reading you discussed in your dataset analysis).
3. If you worked with a partner, you should provide a brief evaluation of your partner's contributions to the assignment. What work did they do/how did they contribute to the overall success of your assignment? Did they complete their work on time? Did you split the work evenly? Did you run into any problems in working with your partner that I should be aware of? (This evaluation should comprise no more than a few sentences of your reflection.)