Guillermo Soto - Clement Godbarge - Yuhe Zheng

1. The project is a “chromatic index” for Bnf Ms Fr 640. The idea was to map and trace color throughout the text, and allow the user to use that mapping to navigate the text in a novel, fun way. We don’t know how applicable this might be to academic work on the manuscript, but we hoped this could serve as a sort of “public outreach” project by making the manuscript more accessible. Given the complexity and scope of this task, we chose to focus on the color red. There are two components to our submission: a visual mockup of what a fully realized graphical user interface version of our project might look like, and a Python program/Ipython notebook that implements most of the functionality in a command line environment. The mock-up outlines the process for how a user might interact with an idealized GUI; starting with a color wheel, they would select a primary color. Afterwards, some information about that “generic” color would be displayed, and then several categories of things falling under the umbrella of that color would be displayed; for example, materials, pigments, dyes, and color derivatives. The user could then select any of these options to get a page with similar information, including a “keyword in context” style search result, a list of the folios in which the term appears, a color code, etc. The Python program strives to model this functionality, though not presented as elegantly. It heavily utilizes functions from the Natural Language Toolkit library. It is dependent upon the plain text files of all the folios, both the massive one containing the entire transcript and each folios individual file. The program consists of a basic command-line menu, and a ‘searchFolios’ function that will take the inputted term, search through the folios for it, and then do the following: generate a dispersion plot, print the title of the folio(s) the term is found on followed by a key word in context search result, then a link to all of the matching folios on the Making and Knowing github, and then a link to a Frantext

search for the term (note: this last functionality is currently broken, as the program works off of the English translation and frantext is, well, a French resource...if the term is not the same in English as it is in French, the results will be meaningless). There are some complications to this program; namely, NLTK does not support multiple word phrases in its concordance function, so inputting a term with whitespaces results in no matches being found. For example, you could search for ‘Armenian’, or ‘bole’, but not ‘Armenian bole’. The function is generically implemented, so even though the program will tell the user to input a term from a precompiled list, any term can be searched for. ‘Chocolate’ may not appear as a term for any of the options in the menu, but a search for it will work just the same.

2. As mentioned before, the use cases for this project may not be seriously academic, but rather, aim towards making the manuscript accessible and interesting to those who have no background in early modern history or digital humanities. There is also the potential for monetization here; we discussed in our presentation perhaps extending the program to generate links to vendors of particular dyes, and collecting a percentage of those sales.
3. See additional files submitted. The ‘Chromatic Index Analysis’ folder contains all the code and file dependencies for the code, and should be copied over in its entirety (as in, the entire folder; the contents all need to be in the same folder for the program to work).
4. The most significant insight we gained as we worked on the project, in terms of course content, relates most closely to our discussion of “archive” and curation. We were faced with many decisions of curation and categorization; for example, many instances of “red” in text refer to the reddening of metal as it gets hot. This is not really a red material, or a “derivative color” of red, so should we include such instances? How, logistically, would we even go about excluding them? We decided not to, ultimately, and this led us to further broaden our scope; if the user is examining “red” in the text, should we include materials that are red themselves? What about ones that can be used to produce a red color? How do we

differentiate these from “derivative colors” in some cases-for example, brick as a material and brick as a color? With regard to the content of the manuscript itself, it was surprising to find just how often items could be multiply categorized. For example, minium (red lead) is a red material, but is also used for pigments and dyes; such was the case for many other things. The dispersion plots of certain items also potentially provide insight. One might reasonably assume that a particular material will appear in a “cluster” of folios; however, in the case of “vermillion”, the dispersion plots show that it almost only occurs at the beginning and end of the manuscript. Upon first examination, this raises a number of questions: is it just that there are a few, very distinct recipes that all happen to utilize vermillion, but are not characterized by their use of it and thus are not clumped together? Could this be indicative of a discrepancy in the ordering of the pages from their initial order?

5. The project changed considerably from proposal to final product. We at first thought to do ‘all’ the colors (not literally all, obviously), but then realized we would be better served by focusing a particular one in order to give us time to develop a more functional project. We had proposed a basic GUI at first, but later decided it made more sense to simply provide a mockup of what a more refined interface might look like while backing it up with a rudimentary implementation in Python. We also shifted from simply focusing on ‘color’, to anything that might fall under the category of a particular ‘color’; to illustrate, initially we may have focused on red and then focused only on its color derivatives, such as crimson, scarlet, etc. However, we decided it made more sense to also include materials that are red, even though they are not necessarily evocative of a “shade of red”.