**Data+ 2018 Project 6 Women’s Spaces: Visions Sheet**

|  |
| --- |
| **Overall Goals + Products from Project**  **(What do we want to do? What will we produce?)** |
| 1. **Data Visualization** 2. **Bechdel-Type Test** 3. **Presentation/Poster**    1. Visualize data about media messages    2. Completed text analysis charts and “average” face images |
| **Steps to Complete Goals + Products**  **(\* indicates final product)** |
| **I. Data Collection and Review (Nathan)**   1. Compile all text files 2. Spell check text files - spelling software that identifies the “misspelled” words and lets us control the editing step 3. Developing a list of words that we change or respell **(All researchers remember to do this - in Google Drive)** 4. Save all text files in multiple formats on all users local discs and Google Drive and people.duke.edu/netid directory. 5. Compile all image files. 6. Convert all to jpg. 7. Save all image files in multiple formats on all users local discs and Google Drive and people.duke.edu/netid directory.   **II. Running Data Through Programs (Alexis)**   1. Text Files    1. Getting all text files into R data frames (Alexis - put in GoogleDrive)    2. “Clean text”:       1. TF-IDF and LDA does not require removing stop words       2. Topic modeling 2. Image Files    1. Run through FaceDetect and collect attributes we want    2. Find program that detects (in order of importance) skin tone, size of model, symmetry, masculinity/femininity **(Sandra)**    3. Bind together data frames from different programs in R   **III. Text Analysis Tools (Alexis and Sandra)**   1. tf-IDf 2. Sentiment analysis - choose what Sentiment dictionary fits best, may be more than one 3. LDA 4. Figure out n-gram deal   **IV. Data Visualization (Nathan)\***   1. Self-study Tableau 2. Create attractive, complex, interactive and highly comprehensive data visualizations of various variables   **V. Deep Text Analysis (All)**   1. Analyze 100 covers for what our computer model is looking for - look at existing research   **VI. Structural Equation Modeling (Nathan)**   1. Identify latent variables we want to know about (e.g. femininity, sex appeal, virginity-appeal, etc.) 2. Identify manifest and observed variables, narrowed down from what we collected and using research on the subject 3. Draw out and diagram the SEM. 4. Run it through a prediction test   **VII. Bechdel-Type Test Data (All)\***   1. Developing a test that a magazine cover passes yes/no 2. Compare passing rates to another related variable and test for statistical significant--for example, outcome   **VIII. Cross-Validate Text Mining Results (Automated) with Deep Text Analysis (Manual)**   1. Do as we go to better understand how text analysis should proceed 2. Be able to generalize final results?   **IX. Poster Presentation (All)\***   1. Create a 48” x 24” poster summarizing results with figures, visualizations and data in Microsoft Powerpoint Online, shared via DukeMail |
| **Blindspots + Strategies to Solve** |
| **Learning curve with Python; migrating all processes to R** |

**Planning Schedule**

|  |  |  |  |
| --- | --- | --- | --- |
| **Week #** | **MUST complete** | **SHOULD complete** | **ACTUALLY completed** |
| 4 (6/18-22) | All text from all 5 magazines 2010-2018.    Text needs to be cleaned – spell-checked and formatted correctly    Move all face images into Python    **Phase I** | Dictionary of synonyms or phrases that mean the same thing for text analysis    Organize our text and images | Spell Checked All Text Files  FaceDetected all Image Files  Organized text and images  Began analysis/ data exploration |
| 5 (6/25-29) | **Phase II, III**  Learn Tableau (Nathan) and make preliminary data visualizations. Explore further options for designing and showcasing data visualizations (javascript?) | **Phase IV**  Construct data visualizations  Dictionary | Constructed some preliminary data visualizations  Developed coding scheme for image analysis  Began coding images |
| 6 (7/2 – 7/6) | **II, III, IV**  Work on presentation | **V**  Fix image data  Continue with data visualization  Begin coding images for deep image analysis  Deep text analysis? |  |
| 7 (7-9 – 7/13)  **\**Presentation #2*** | **VI, VII**  **Goal #1: A new sentiment library that better relates to magazine data**   * *Filter words:* Determine which words are related to project, and which words are not * *Justify the decisions for relevant words* * *Get + read research on gendered language* | **Goal #2: Organizing data/creating a system for data**   * Divide areas of data among team members * Team members create visualizations for data and add them to the Google doc |  |
| 8 (7/16-20) |  |  |  |
| 9 (7/23-27) |  |  |  |
| 10 (7/30-8/3)  ***\*Poster Presentation*** |  |  |  |