# DATA 115: Personal Dataset Final Submissions

Final submissions for the personal dataset project are due December 6. You will need to submit two things through blackboard (more details about the expectations for these components are presented below):

- A link to your final repository (something like [https:github.com/drdeford/DATA115_PDP1](https:github.com/drdeford/DATA115_PDP1))

- An .mp4 file of your (at most) three minute presentation of your project

## 1 GitHub Repository

While the data is the central feature of the project, your descriptions of the motivation, cleaning process, and analysis are also important. The following information should be found in the ReadMe of your repository:

- **Motivation:** You should describe the question you set out to answer at the start of the project or what led you to the specific dataset that you selected.

- **Data Process:** The centerpiece of your writeup should be a comprehensive discussion of your process for sourcing, collecting, cleaning, and transforming the raw data into your final dataset. The instructions should be clear enough that someone else could reproduce your process based on just the ReadMe file to obtain the same final results. Note that this doesn't necessarily have to be more than a couple of sentences but it should be complete.

- **Visualization:** The ReadMe should contain at least one visualization summarizing the data. The visualizations should follow the principles of good figure design and be sufficiently documented to explain what they are demonstrating.

- **Analysis:** Finally, the ReadMe should describe the results of at least one method from this course that you applied to the dataset and what you learned from it. This can be anything from a simple EFA technique like identifying outliers, making a box plot of a column, evaluating the summary statistics (Tukey's 5 numbers), to one of the regression models from supervised learning, or one of the clustering methods from unsupervised learning. This is not an exhaustive list of possibilities and as with the previous components doesn't need to be long or drawn out.

## 2 Data Presentation

Your presentation must be at least 90 seconds long and no more than three minutes. Since this isn't an enormous amount of time, you will have to decide which aspects of the project you are most excited to communicate. You can record your presentation using Zoom or Panopto, which WSU has licences for, or other software like OBS or the default app associated to your webcam. You will be evaluated on organization, delivery skills, clarity, and your materials. Here are some potential topics that you might decide to discuss in your presentation:

- Describe the basic properties dataset and why you selected it for this project.

- Describe any processing problems you identified with the data and how you overcame those issues.

- Describe the process you followed to design your visualization and any alternatives you considered.

- Describe what you learned from your data by applying techniques you learned in this class.

- Describe any additional analyses that you would have liked to carry out and any additional data that would have been needed in order to extend your analysis.