SVD project

1. Find some data you want to analyze using SVD/PCA.
2. Formalize the question you'd like to answer. (Stop and think: can you answer this question using SVD/PCA?)
3. Decide whether you need to center your data, scale it, use a covariance matrix, or otherwise prepare your data. Chat with me if you're not sure.
4. Carry out SVD.
5. Analyze:
   (a) Determine how many components you need to explain/recover at least eighty-five percent of your data.
   (b) Plot a projection of the data to the first two dimensions (the plane spanned by the first two eigenvectors).
   (c) Do more as you choose: do you have enough information to answer your question? Would other visualizations of the data help explain the story?
6. Write up your conclusions. Explain, in particular, what the first and second primary components may mean.
7. Explain how you could test your conclusions with more data or other techniques. Explain other weaknesses in your analysis.

   Basic version of this assignment for full credit: Carry out the steps above and hand in a typed version that has no more than 3 pages. Use full English sentences to explain your questions, analysis, and conclusions.

   In your discussion of methods, you need to provide a link to the data in its original location (or the dataset itself if an original dataset), discuss how it is transformed (centering, scaling, etc.), and cite any other sources you use to help you do this project.

   There are many expansions to this project if you'd like to delve deeper into data analysis or write something nice to put on your website or LinkedIn profile. These can be longer. Make sure you can do the basics first though!

   Due: April 11; if this is a difficulty or if you want to expand the project, talk to me!