Mini Project #2

Goal: practice the basic visualization tools used in visual analytics

- use data from mini project #1 (or other)
- client-server system: python for processing (server), D3 for VIS (client)
- non-CS students use plotly Dash and python or R

Task 1: basic dimension reduction and data visualization with PCA

- use PCA to compute the Eigenvectors of the data and visualize the Eigenvalues as a scree plot
- add an interaction element into the scree plot that allows the user to mark and select the intrinsic dimensionality index (d_i)
- plot the data into a PCA-based biplot

Task 2: visualize the data with a scatterplot matrix

- use the PCA components $\leq d_i$ to obtain the 4 attributes with the highest squared sum of PCA loadings and list them in a table on the webpage
- use these four attributes and construct a scatterplot matrix
- use k-means to find clusters and color the points by cluster ID

continued...

...Continued

Task 3: MDS plots (numerical data dimensions only)

- construct the data MDS plot (use the Euclidian distance) and visualize it via a scatterplot (use *metric* MDS – python sklearn.manifold.MDS)
- color the points by cluster ID (see task 3)
- construct the variables' MDS plot (use the (1-|correlation|) distance) and visualize it via a scatterplot (also here, use *metric* MDS)

Task 4: parallel coordinates plot (PCP)

- visualize the data in a parallel coordinates plot (all data dimensions, categorical and numerical)
- come up with a meaningful axes ordering by user interaction
- color the polylines by cluster ID (see task 3)

Extra Credit: find a good PCP axes ordering from correlations

- numerical values only: use the correlations observed in the variables' MDS plot to help with the axis ordering -- the user would click on points in sequence and the axes would be arranged in that sequence

Scoring and Due Dates

Each (task) bullet item carries 10 points

- an extra 10 pts for overall elegant implementation and function
- extra credit earns 10 points

Don't forget to

- label the axes and tick marks where appropriate
- show color legends where appropriate
- provide a meaningful header on each plot

Due dates

- Tasks 1 and 2 are due March 11, end of day
- Tasks 3, 4, and extra credit item are due March 18, end of day

Deliverables

By March 11, submit on Blackboard (lab 2a)

- zip file with source code for tasks 1 and 2, as well as the data
- evidence (screenshots) that you accomplished task 1 and 2
- put these screenshots in a report and submit as pdf file
- add captions so grader knows what they are

By March 18, submit on Blackboard (lab 2b)

- voice-narrated video file to show all features of your software in action
- note any interesting observations you were able to make in the data
- note the strengths & weaknesses of the various visualization methods
- 2-3 page report
  - describe interesting observations (beyond the video)
  - mention anything noteworthy about implementation (beyond the video)
- zip file with complete source code for tasks 1-4, as well as the data

Grading

Grading

- TA will pick students at random for thorough code review sessions
- you better know your code !!!
- so, please do not just copy code beyond the D3 templates
- or even worse, videotape someone else's program