Goals: Welcome to CMPS80L! The main goal of this course is to understand how visual representations can help in the analysis of and elucidation of complex controversial data from social sciences.

First, we ask data science questions. Can you trust the data? Is it true? What is the source? Is it consistent? What are the sources? Are they reliable? What is missing? Can it be interpreted in various ways? How large is the margin of error? How often does it change?

Second, what is really relevant in this data? What subset do I choose? Why do I choose this? What point do I want to communicate? To whom? To what purpose? Which statistical summaries do I choose – mean, mode, variance? Is there bias? Are variables correlated? What ranking or scales should I choose? How do I fit a straight line and should I? How does it transpire that same data can be used to create opposite messages? Is data extrapolated? What are the underlying assumptions?

For the more sophisticated, which variables matter most (factor analysis)? Are there clusters (cluster analysis)?

Third, ah, now that I trust and understand the data and know what I want to communicate and to whom, how do I communicate the data visually? Bar chart, line chart, pie chart, scatter plot, but why? Do they make sense? What colors do I choose and why? What legends, labels, axes, annotations, should I use and why?

Finally, how can I make an impact using visualization?!!!!??

There will be programming exercises that allow you to create simple visualizations. Examples will be drawn from energy, environment, economics, education, and empowerment.

Textbooks:
1. The Functional Art: An Introduction to Information Graphics and Visualization by Alberto Cairo, 2013

Both the books are required. Readings will be assigned from these books in the very first class. It is important that you have ready access to these books. These books (new) are available in $29.33 (hard copy) and $18.41 (hard copy) from Amazon as on 02/05/2014 (excluding shipping).

Additional Readings: There will be several additional weekly state-of-the-art readings, videos, websites, blogs, and visualizations to read/view and critique.

Overview (catalog description):
Social Data Analytics: Veracity, Variety, Volume, Velocity; Statistical Computation: Misuse, Bias, Dispersion, Correlation, Regression, Differential Scales, Normal Distributions, Factor and

**Prerequisites:** Although there are no prerequisites for the class, you will be a good candidate to take this class if (i) you love beautiful visualizations, (ii) you truly care about some social issues, and (iii) sometime you truly wonder which data to believe? Although statistical computation and programming will be covered in the class, I believe, motivation will be the key to success in the class.

If you are simply looking to fulfill your statistics GE requirement in the hope that you will have to do the least amount of work, this is not a class for you. I will assign lots of work and hope to keep you educated, entertained and enlightened. I hope you will enjoy and learn a lot in this class that you will cherish. So, I request that if you are only marginally interested in this class, please give somebody else a chance, because this class will get full.

If you are uncertain whether this class is for you, attempt to answer the following questions: (i) find a web link to a beautiful visualization, (ii) find a media article or a weblink to a news item related to a social topic that you care about, and (iii) find a web link to some data (that you care about) on the web that either does not ring true to you, or you have difficulty understanding/trusting. If you enjoyed doing these tasks, this class is for you.

Although this class is currently scheduled to be offered in Fall 2014 also, it may be cancelled if the enrollment stays below 75. So, this may be your only chance to take the class for a while.

**Evaluations:** This class will use an innovative tiered evaluation system. If something is unclear, ask questions in the very first class.

C (40 points): Every student has to finish a set of problems (some written assignments and some programming ones) *self-paced* in order to obtain at least a C (passing grade) in the class.

C+; B−: B (30 points): In order to obtain one of these three grades, you will have to finish three mini-projects: (i) design critique project, (ii) data curation project, and (iii) visualization design project. These projects will be graded out of 10 each for a maximum of 30 points. In addition, you can augment your points in this category by taking random quizzes in the class (quiz points do not count towards bonus points mentioned below) and engage meaningfully in the classroom. (Bonus Points) If your performance in any of these projects are “stellar”, then you may earn bonus points beyond 30 that can carry over to the next tier described below. A maximum of 10 bonus points will be allowed.

B+: A−: A (30 points): In order to earn these grades, you will have to do a final project and present it to the class on the final examination day. This final project will be graded out of 30. Extra bonus points will be provided for exceptional effort and work.

A+: your final project will blow my mind! I will bow to you and will remember you for ever! In every class that I have taught in last 20 years, I have had stellar students in my class who taught me more than I know.

Posted Date: 02/25/2014