36-721	Fall 2015
Statistical Graphics and Visualization	6.0 units

Homework 2: Visual Perception
Due Sat 9/12/15, 5pm

This HW is inspired by a *Scientific American* article entitled "U.S. Science Degrees Are Up." This article includes a nice infographic, but also makes some claims that are not shown in that graphic. For example, they state "More women are entering college, which in turn is changing the relative popularity of disciplines," but their graphic does not show the degree breakdown by gender.

Using a similar dataset, I have made an **intentionally-bad** graphic for you to critique and remake: HW2_CritiqueThis.pdf.

Your critique of my graphic should explain what could be improved. Justify your criticism using ideas from visual perception research (hint: use this rubric).

Your remake should consist of **one or two graphs** that help answer these questions: In what fields are more women entering college? How is each field's gender balance changing?

- You may use my **simplified dataset** degreeTotals.Rdata, which has totals by year, gender, and STEM vs non-STEM.

- Or, if you're curious, you may (but don't have to) use the **full dataset** degrees.csv, which has a more detailed breakdown by academic discipline.

(These datasets contain the number of bachelor's degrees "conferred by degree-granting institutions," tabulated from the 2010 Digest of Education Statistics, tables 308 to 330.)

Please submit

- a **stand-alone image file** of your remade graph,

- a **two-paragraph writeup** (a critique of why my original graph is poor, and a summary of what we can learn from your remake), and

- **your code** or written instructions;

or a single combined PDF or HTML file, if using knitr and RMarkdown or similar.

See rubric on next page.

1

Component	Competent	Not yet competent
Quantitative Comparisons	Quantitative variables use visual encodings high on the Cleveland-McGill ordering. Encodings are used sensibly (bars start at 0; hues are ordered intuitively; etc.). Elements to be compared are as near each other as possible.	Quantitative variables use visual encodings low on the ordering. Encodings are implemented poorly (bars not anchored at 0; arbitrary hues assigned to quantitative/ordinal variable). Elements to be compared are distant.
Grouping and Search	Gestalt and preattentive processing features are chosen to ease task (find important groups, follow lines, etc.) Elements to be compared are aligned, as much as possible. Distinct variables are mapped to separable dimensions. Choice of colors, shapes, etc. is easy to discriminate.	Difficult to find groups, follow lines, etc. Elements to be compared are not aligned. Distinct variables are mapped to integral dimensions (e.g. point width and height). Distinct elements cannot be discriminated.
Cognition	Differences, proportions, or other important derived variables are plotted directly. Items are ranked by variables on which comparisons are to be made.	User must compute differences, etc. mentally. Ranking is arbitrary or unhelpful for analysis (e.g. alphabetical).
Consistency	Meaning of graphical elements is consistent across small multiples. Changes in design are purely data-driven. Visual variables are used only when mapped to data. Semantic associations are used, if possible (e.g. blue = cold, red = hot). More means more (larger size or deeper hue maps to larger value of the variable).	Small multiples are not consistent. Design changes are stylistic or arbitrary (e.g. new colors for the same categories). Superfluous visual variables are shown (3D, shadow, other variables not mapped to data). Semantics are mangled (e.g. 'orange' and 'blue' crab species are not mapped to orange and blue colors). More (stronger encoding) is mapped to less (lower value of data variable).
Critique	Critique of the given graphic points out major flaws according to this Visual Perception framework and shows understanding of these principles.	Critique of given graphic is incomplete or does not show understanding of this framework.
Older Skills	Also meets all requirements for HW1: Legible, Comprehensible, Reproducible, and Informative, and Administrivia format.	Does not meet some of Legible, Comprehensible, Reproducible, Informative, or Administrivia requirements.