Interdisciplinary Approach to Data Visualization

IDC 6700, Summer 2018 (3hrs credit)

Instructor   R. Paul Wiegand

Contact   wiegand@ist.ucf.edu
Research park, IST
Partnership III (PIII), Room 209
407.882.0313

Office Hours   Mondays 2p-4p

Texts   *Now You See It!* (about $35)
by Stephen Few

Visualize This (about $25)
by Nathan Yau

Various readings relating to data analysis

Class Room   Partnership III, room 233

Class Time   Thursdays, 3p-6p

Course Objectives

The objective of this course is to provide an interdisciplinary introduction to the principles and fundamentals of visualizing statistical information. We will cover limitations of common visualization methods, effective visualization methods, perception and representation issues, as well as general principles in the design of effective data visualizations. This course is a hands-on, project-based course. Labs and examples will use the open source statistical package **R**. By the end of this class, students will be expected to spot problems with existing visualizations, be able to fix these problems, as well as be able to design and implement effective data visualization for their own problem domains.

Useful Course Background

This course assumes some basic computer skills as well as at least a preliminary background in statistics. One of the following courses is required: ESI 5219, STA 5206, or DIG 5876.

Grading & Course Requirements

The grade distribution is as follows: lab assignments (50%), mid-term project (20%), final project (30%). All graded work is individual work; there are no group assignments in this course.

Lab Assignments

There will be five lab assignments designed to help students learn how to perform data analysis and visualization using the **R** software package. The assignments will be posted online via UCF's CANVAS system, and the solutions will be submitted via CANVAS, as well. Homework turned in within 24 hours after the original deadline are worth half their original points, after that it will not be graded.

Mid-Term Project

Each student will have to find two (different types of) published data visualizations, determine the flaws in the visualizations, and correct those flaws by building a new visualization for the same underlying data. The students will submit to CANVAS a short report containing these visualizations and their discussion. Also each student will give a very short presentation to the class regarding these visualizations.

Final Project

Each student will be expected to find a data source and an open question related to that data, perform data analysis involving data visualization to answer this question, and present this analysis in the form of a final report and a final oral presentation to the class. Potential topics must be approved by the instructor by the middle of term.

Letter Grades

Letter grades will be determined at the end of the semester by finding "natural" divisions between clusters of numeric grades based *roughly* on the table shown below. The $+$ and $-$ modifiers for grades will be rare and determined at my discretion, largely based on class participation.

Numeric	Letter
$[90, 100]\pm2$	A
$[80, 89)\pm2$	B
$[70, 79)\pm2$	C
$[60, 69)\pm2$	D
$[0, 59)\pm2$	F

Mixed Mode & Attendance

This is a mixed-mode class. With one or two exceptions, there will be face-to-face lectures each week. In addition, a screencast lecture will be posted on CAN-VAS. Students are encouraged to attend all face-to-face lectures and expected to attend at least half of them. In particular, students must attend the first two weeks worth of lectures during which attendance will be taken to meet UCFs financial aid requirements. If you cannot attend, please contact me directly so that we can work out an alternative arrangement.

Honor Policies

Students are referred to UCF's honor policy for review. Honor violations will result in *at minimum* an undroppable zero on whatever is being graded, a drop in a letter grade, and notification of incident to UCF office of Student Conduct. If you are concerned that some choice you may make could be construed as unethical, you are free to consult me for clarification.

NOTE: As graders, we reserve the right to compare content submitted by students to alternative content (either students' or otherwise) by any means necessary (automated or non-automated).

Course Topics

This course will include the following topics:

- Introduction to **R** and **ggplot2**
- Limitations of common visualizations
- Perception & representation
- Analysis interaction
- Simple, effective graphs
- Techniques & practices (baselines, aspect ratios, scales, axes, etc.)
- Designing visualization with purpose
- Pattern analysis & visualizing patterns over time
- Comparative analysis & spotting differences
- Regression analysis & visualization relationships
- Methods for plotting higher dimensional data
- Scientific visualization