QAC 251 - Data Visualization

Syllabus: Fall 2015

General Information

Instructor: Valerie L. Nazzaro  Email: vnazzaro@wesleyan.edu
Office: Allbritton 019  Drop-in Hours: M 4:30-5:30 pm, W 1-2:30pm
Classroom: Allbritton 204  Class Time: Tu/Th 2:40-4 pm

Course Description

Constructing visualizations which yield insight and understanding is one important way to aid communication within your own areas of interest. This course will provide an opportunity to learn the necessary programming skills to construct basic and advanced (interactive) graphics with the intent of applying it to your own areas of interest.

Course Materials

- Textbooks
  - Howard, Wainer, Graphic Discovery: A Trout in the Milk and Other Visual Adventures, (accessible online through Wesleyan library).
  - Teeter, Paul, R Cookbook, O’Reiley, (portions will be made available)
  - Yau, Nathan, Data Points: Visualization That Means Something, (accessible online through Wesleyan library)
  - Yau, Nathan, Visualize This: The Flowing Data Guide to Design, Visualization, and Statistics, (accessible online through Wesleyan library)

- Software: The use of the R statistical environment (downloadable from cran.rstudio.com/) with the RStudio interface (downloadable from rstudio.org) is used to perform the computations and visualizations within this course. Although this program is available on all campus computers, you may want to download both R and R studio to your personal computer.

Course Topics

The course schedule and readings will be maintained on Moodle. The following topics will be covered:

- Introduction to data visualization:
  - What is data visualization?
- What makes visualizations effective/in-effective?
- Misleading visualizations (current news examples) and underlying ethical concerns
- Discuss good practices (scales and conventions)

- **R basics**
  - Getting oriented in R studio (panes and tabs)
  - Simple calculations

- **Visualizations using lattice:** Constructing and interpreting standard plots with lattice (boxplots, histograms, density plots, scatterplots, violin plots)

- **Reproducible workflow with R markdown**

- **Visualizations using ggplot:**
  - Describing plots using grammar of graphics
  - Make plots from previous week using aesthetics and faceting

- **Preparing data in R for visualization:**
  - Scraping web data
  - Cleaning data
  - Merging data sets

- **3 variable plots**
  - Bubble plots, Contour, raster, and tile

- **Map visualizations:**
  - Shape files and spatial classes
  - Geocoding and google maps
  - Coordinate Reference System (CRS)

- **Text data visualizations:**
  - Text mining (Project Gutenberg, Google N-grams)
  - River plots, word clouds, dendrograms, map/text plots, (check for eyegraph)

- **Data as an art form**
  - Using random data in R to make art
  - Guest artist presentation

- **Advanced/interactive graphs**
  - Use tools from previous weeks to create interactive plots which allow user to select variables of interest, change scale, make specific comparisons
GRADING

- **HOMEWORK AND IN-CLASS ASSIGNMENTS (40%)**: A large portion of the class requires you to actually work with data. Homework/lab work is an important way of reinforcing techniques and tools learned in class. Assignments will be posted in Moodle and announced during class. You are welcomed and encouraged to collaborate with your classmates, but you must write up your own code and provide your own output and write-up - otherwise you will not receive credit. Please note that assignments are due before midnight on the due date. One assignment will be accepted up to 72 hours late. After that, you will not receive credit for late assignments.

- **MIDTERM EXAM (25%)**: You will have one take-home exam that will be distributed on October 30th and will be due on November 5th. This is an individual exam and you are not permitted to discuss the exam with any other student or person. All clarification questions should be directed to me. All other resources may be used.

- **COURSE PROJECT (35%)** The final project is a demonstration of your knowledge and fluency with data visualization techniques and tools. The details of the project and the breakdown of the grade is provided in the next section. All components of the project will be submitted into Moodle no later than 11:59 pm on the due dates provided.

**Course Project**

For the project, you will select a data set of your choice that we haven’t used in class (aim for larger than 1000 observations and at least 8 variables). From the data, you will develop a question (or two) that you wish to explore. You will develop a series of 5 visualizations for this project. 3-4 visualizations to answer your question(s) of interest and 1-2 visualizations through exploration. The question and the visualizations should be your own. That is, for this assignment, the objective is not to replicate someone else’s graphs. All visualizations you use should be substantially different in some way. You will display your visualizations in moodle for peer and instructor feedback and assessment. Your visualizations should “speak for themselves”. A short (5 page) paper will accompany your final visualizations that explain your visualizations and the questions they address. This paper will be used to confirm that your visualizations did indeed send the correct (intended) message to the audience. The project schedule is as follows:

- **Intention Statement (5 points): due September 25, 2015**
  This is a 1-2 paragraph statement to get you thinking about your project and for you and I to start a dialogue about what direction your project may take. You do not need to make any final decisions at this stage. You may wish to address one or more of the following questions.

3
– Do you have a data set in mind? If so, where does it come from and how does it relate to your areas of interest? Do you have any original question that arises from this data set?
– Do you have a subject matter you are looking to explore? Do you know where you can begin looking for data related to that subject? Discuss your plans moving ahead.
– Do you have a lot of areas of interest, but are not sure how to narrow things down or make a project decision?
– In what ways do you anticipate needing my support in order to move forward?

• Data Set (5 points): due October 9, 2015
Identify the data set that you plan to use and send the url or file attachment.

• Question of Interest (10 points): due October 23, 2015
Decide on an original question or two that you think will be interesting to explore with your data set. Identify the variables that you will use to answer those questions. Provide some insight as to why this question is important and include references that explore similar questions. A small literature review (1-2 pages) should be included.

• Manipulated Data Set and Code (5 points): due November 6, 2015
Send me a glyph-ready data set and the corresponding code you needed to get your data in that form. Some data sets may already be in glyph-ready format. If this is the case, submit a statement that indicates no manipulations were necessary.

• Two Preliminary Visualizations and Code (5 points): due November 20, 2015
Send me a rough draft of two visualizations with relevant code. Submit a paragraph about what each visualization is supposed to show me. This is a chance for me to give you some preliminary feedback and let you know whether your intended objective was met. (Use R-markdown).

• Final Visualizations (40 points): due December 11, 2015
Submit your final five visualizations with relevant code. (Use R-markdown).

• Final Paper (20 points): due December 18, 2015
This paper should include an introduction to your question, why the question is important or interesting, and your relevant (revised) literature review. In addition, you should include your visualizations and a paragraph about each visualization. These paragraphs should state what information each visualization is trying to communicate and any conclusions that can be drawn. Finally, mention any additional questions that arose from your findings and any future work of interest. (Use R-markdown).
• **Peer Assessment (10 points): due December 18, 2015**
  You will be asked to assess a random sample of visualizations from your peers. The feedback you provide to your peers will be anonymous. Your grade here will be based on the quality of feedback and comments you provide in addition to the feedback and assessment your peers give you.

**Honor Code**

Wesleyan University expects all students to be honest and committed to the principles of academic and intellectual integrity in their preparation and submission of course work and examinations. Students and faculty at Wesleyan University are part of an academic community defined by its commitment to scholarship, which depends on scrupulous and attentive acknowledgement of all sources of information, and honest and respectful use of college resources.

**Academic Accommodations**

Wesleyan University is committed to ensuring that all qualified students with disabilities are afforded an equal opportunity to participate in and benefit from its programs and services. To receive accommodations, a student must have a documented disability as defined by Section 504 of the Rehabilitation Act of 1973 and the ADA Amendments Act of 2008, and provide documentation of the disability. Since accommodations may require early planning and generally are not provided retroactively, please contact Disability Resources as soon as possible. If you believe that you need accommodations for a disability, please contact Dean Patey in Disability Resources, located in North College, Room 021, or call 860/685-5581 for an appointment to discuss your needs and the process for requesting accommodations.

**Final Notes**

I am looking forward to a great semester! Please know that feedback on class is always welcomed and appreciated. I hope you all find the time to visit my office at least a few times throughout the semester.