QAC 251 - Data Visualization

Syllabus: Summer 2022

General Information

Instructor: Valerie L. Nazzaro **Email**: vnazzaro@wesleyan.edu
Live Sessions: MT/ThF 10:50am-12:05pm
Office Hours: W 10:50am-12:05pm

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

- **Readings**

    - 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 and Schedule

Week	Day	Date	Topic	Assignment Due
1	Thursday	May 26th	Introduction to Data Visualization, Technology Check	
	Friday	May 27th	Topic 1: Elements of Data Graphics and Introduction to R Topic 2: Standard Plots	
2	Monday	May 30th	Topic 3: Facets and R Color Brewer	
	Tuesday	May 31st	Topic 4: Data Manipulation	
	Wednesday	June 1st	*Office Hours*	Homework 1
	Thursday	June 2nd	Lab 1	Intention Statement
	Friday	June 3rd	Topic 5: Area, Dumbbell, and Tile	Lab 1
3	Monday	June 6th	Topic 6: Spatial Analysis	Homework 2
	Tuesday	June 7th	Lab 2	
	Wednesday	June 8th	*Office Hours*	Project Data Set
	Thursday	June 9th	Topic 7: Riverplots and Chord Diagrams	Lab 2
	Friday	June 10th	Topic 8: Alluvial Plots	Homework 3
4	Monday	June 13th	Exam	Exam
	Tuesday	June 14th	Topic 9: Shiny	
	Wednesday	June 15th	*Office Hours*	
	Thursday	June 16th	Lab 3	Preliminary Visualization
	Friday	June 17th	Topic 10: Plotly and Topic 11: Leaflet	Lab 3
5	Monday	June 20th	Topic 12: Statebins and GeoFacet & Topic 13: Hierarchical	
	Tuesday	June 21st	Work on Project	Homework 4
	Wednesday	June 22nd	*Office Hours*	
	Thursday	June 23rd		Final Project

Course notes and corresponding videos will be provided for all topics. Please go over the notes and videos carefully prior to the live session where the topic is listed. Live sessions are devoted to work on the 'Exercises to Try' section of the corresponding topic notes or for Lab work. We will break you off into small groups to work on these exercises. The instructor and a TA will be available to you for help during this time. It is encouraged that you attend these sessions regularly to ensure that you are not falling behind! The material moves quickly and you will make better use of your time if you are practicing and discussing your code with others.

GRADING

The following three elements are used to compute your course average. Your numeric average will be assigned to a course grade as follows: A 94-100; A- 90-94; B+ 87-90; B 84-87; B- 80-84; C+ 77-80; C 74-77; C- 70-74; etc. If you are taking the course as CR/U, you must earn an average of 70 or higher to earn credit. Please note - you must take the course graded A-F if you would like it to count towards the data science minor or towards the applied data analysis certificate.

- *HOMEWORK AND LABS (30%)*: 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. One assignment will be accepted up to 72 hours late. After that, you will not receive credit for late assignments.

- *EXAM (35 %)*: You will have one exam for this course. It will be distributed on Sunday, June 12th and **due on June 13th**. This is an individual exam and you are not permitted to discuss the exam with any other student or person (online or otherwise). All clarification questions should be directed to me. Course notes 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.

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 visualiztions for this project. 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. The work you do on your final project should be yours. Any assistance you get from tutors or teaching assistants for your project must be approved by me and clearly stated. You will present your visualizations 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 (3 points)** 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.

  - 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 (2 points)** Identify the data set that you plan to use and send the url or file attachment. You can take a snapshot of any large file (the purpose is for me to see the form the data is in).

- **One Preliminary Visualization (10 points)**

  - Be sure your data can be read into R – for some this step will be a lot of work – start early!

  - Perform necessary steps to get data into the format required to make graphs. Check structure of data – are quantitative variables being read as numeric? Do you want to add new variables to the data set? Other data steps you are taking?

  - Construct a preliminary visualization to answer one of your main questions of interest or through a separate exploration. Submit a paragraph about what the visualization is supposed to show me. Do you think your visualization is effective? Are you having trouble accomplishing any specific task with this? This is a chance for me to give you some preliminary feedback and let you know whether your intended objective was met.

- **Final Visualizations (85 points)** Your final product should include an introduction to your exploration (why the question is important or interesting, any background information about the data, and any relevant literature - if necessary/appropriate). 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. Your final product can be submitted as a paper with images inserted, as a shiny app, or as an html. Also submit your .R file that includes all of your code.

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 for an appointment to discuss your needs and the process for requesting accommodations.

Final Notes

I am looking forward to a great session with you all! Please know that feedback on class is always welcomed and appreciated.