INFO 526 Data Visualization and Analysis

Spring 2022

Adriana Picoral, PhD adrianaps@arizona.edu

Course Description

This course provides an overview of the various concepts and skills required for effective data visualization. It presents principles of graphic design, programming skills, and statistical knowledge required to build compelling visualizations that communicate effectively to target audiences. Visualization skills addressed in this course include choosing appropriate colors, shapes, variable mappings, and interactivity based on principles of color perception, pre-attentive processing, and accessibility.

Course Prerequisites

Strong quantitative, statistical, and analytical reasoning abilities are needed to succeed in this course:

1. Students must have taken at least a one-semester introductory statistics course. The recommended course to satisfy this pre-requisite is SIE 530: Engineering Statistics OR SIE 533: Fundamentals of Data Science for Engineers or equivalent.
2. Students should have some basic experience using a programming language (e.g., R, Python).

Course Format and Teaching Methods

The instructor will supply the course content to the students in various formats, including slides, reading materials, and explanatory videos by module. The general course structure will be organized as follows:

- Module materials: weekly reading and audio-visual materials provided by the instructor and organized in modules
- Assessment:
  - Data Visualization challenges: weekly collaborative programming assignment – given a real-world data set, create a visualization, explain decisions
  - Content quizzes: weekly quiz to assess content of the week
  - Visualization critiques: weekly visualization critique by students – find a visualization, evaluate it, respond to two other critiques (check mark for this, no grade)
- Weekly summary and overview video for course content (provided by instructor)

Tech skills: programming in R / Python

To succeed in this course, 2-3 hours of study time per hour of credit unit are required (in addition to the course content). This means that approximately 12-13 hours of study time per module are needed in order to meet course expectations. These hours should be spent on reading additional texts, working on your data challenges and final project, researching for new information, or thinking about course content.

Course Objectives

In this course students will:

1) Collaborate in hands-on programming exercises and assignments to build different types of plots (e.g., histograms, line plots, maps) using real-world data sets
2) Reflect on what makes a good visualization, taking into account principles of color perception, pre-attentive processing, and accessibility
3) Practice communicating complex data analysis to different types of audiences

Expected Learning Outcomes

At the end of this course, students will be able to:

1) Articulate what aspects make good visualizations based on principles of color perception, pre-attentive processing, and accessibility
2) Apply basic visualization concepts and a broad range of visualization techniques for data analysis communication taking into account target audiences and purposes
3) Implement data visualizations that communicate answers to data questions effectively in a commonly used programming language (e.g., R, Python)
4) Demonstrate the ability to collaborate with data science and domain experts to advance data science approaches.

Program Competencies

This course prepares students to meet all program competencies, with a stronger focus on the following:

- C2: Data manipulation, analysis, and interpretation through the different materials and hands-on practice on data wrangling and visualization in R. This course focuses on the following sub-competencies:
  - C2.A: Students will be able to identify specific types of data for different analytical methods
  - C2.B: Students will be able to use/develop efficient computational methods to clean, format, transfer, and store data.
  - C2.C: Students will be able to apply appropriate statistical, machine learning, visual analytics, and other techniques to identify patterns and make sound predictions with given data.
  - C2.D: Students will be able to develop methods to align and integrate data from multiple sources.
- C3: Communication and teamwork: Students will acquire skills to work with others within and across disciplines and be effective communicators through the collaborative hands-on programming exercises and assignment (including final project). This course focuses on the following sub-competencies:
  - C3.B: Students will be able to effectively articulate various evidence supporting a solution and to communicate the results of their work, using appropriate graphics, visualizations, multi-media vehicles, or artistic performance.

Required Texts and Materials

All required reading materials will be made available on D2L. In addition to scientific papers, and other reading materials, course readings will be assigned from:

- Healy, K. (2018). Data visualization: a practical introduction. Princeton University Press.

Course Schedule

The tentative course schedule with students learning outcomes by module is provided below:

Module 1

- Topics:
  - Overview of Visualization
    - Language/terminology of encoding
      - Marks, channels
    - Basic principles & Pitfalls
      - Color
      - Expressivity & Effectiveness
      - Pre-attentive Processing
      - Accessibility
    - Coding tools intro
  - Working/ getting started examples:
    - Bar charts
    - Scatterplots
- Student learning outcomes – by the end of week 1 students will be able to:
  - Use appropriate language/terminology for describing visualizations
  - Identify and apply the basic principles of visualization, including color, expressivity, effectiveness, and pre-attentive processing
  - Find and correct visualization errors/pitfalls
  - Produce effective bar charts and scatterplots

Module 2

- Topics:
  - Review and reinforcement of:
    - Time series
    - Histograms
    - Pie Charts & Stacked Bars
    - Boxplots
    - Heatmaps
    - Geographic maps
- Student learning outcomes – by the end of week 2 students will be able to:
  - Match visualizations with specific data questions
  - Discriminate among different types of visualization and argue for why a visualization-data question match is more appropriate than others

Module 3

- Topics:
  - Composing multiple views
  - Small multiples
  - Faceting

  - Arranging axes
  - Maintaining consistent scales
- Learning goals – by the end of week 3 students will be able to:
  - Arrange visualizations by composing multiple views
  - Apply multiple views to produce more effective visualizations

Module 4

- Topics:
  - Dimensionality reduction for data visualization (2D and 3D projections): what it is, and why to use it
  - Types:
    - Principal Component Analysis (PCA)
    - T-distributed stochastic neighbor embedding (t-SNE)
    - Uniform Manifold Approximation and Projection (UMAP)
  - Trade offs of dimensionality reduction
    - advantages: data compression, noise reduction
    - drawbacks: outlier influence, non-linear dependencies
- Student learning outcomes – by the end of week 4 students will be able to:
  - Reduce the number of features in a data set by applying dimensionality reduction
  - Identify advantages and drawbacks of dimensionality reduction for data visualization

Module 5

- Topics:
  - Visualizations for Different Audiences and Purposes:
    - Storytelling
    - Persuasion
    - Decision Makers
    - Exploration
    - Context of visuals: who, what and how
- Student learning outcomes – by the end of week 5 students will be able to:
  - Distinguish among different visualization audiences and purposes
  - Anticipate visualization needs by audience and purpose
  - Construct a visualization story

Module 6

- Topics:
  - Adding interactivity to visualizations:
    - Basic linking
    - Highlighting
    - Filtering
- Student learning outcomes – by the end of week 6 students will be able to:
  - Assess the accessibility of different visualizations
  - Produce interactive visualizations

Module 7

- Topics:
  - Basic principles of bad visualization:
    - Common pitfalls of bad visualization
    - Gestalt principles of visual perception
    - Pre-attentive attributes in creating visuals
- Student learning outcomes – by the end of week 6 students will be able to:
  - Assess the quality of different types of visualizations
  - Advice on how to fix problems with bad visualizations

Assessment and Grade Distribution

Assessment Category	Unit Percentage	Percentage of Final Grade
Data Viz Challenge	6 challenges x 5	30
Weekly Quiz	6 quizzes x 5	30
Visualization Critique	2 critiques x 4	8
Discussion Board	8 topics x .25	2
Final Project	Project proposal 10	
	Final write-up 20	30

Final Project

There is a final project in place of a final exam for this class. In groups, you will work collaboratively to find a dataset that interests you and that helps answer a question that you are interested in. You will submit a project proposal by the end of week 3. You will be graded on the completed R or Python code files as well as your final data visualization.

Grading Scale and Policies

All work must be turned in on the due dates by midnight (11:59pm) Arizona time unless it is announced otherwise. No late submissions will be accepted unless there is a valid excuse.

The following standard UA grading scale will be used for all students:

Letter Grade	Numeric Grade	Interpretation
A	90+	Excellent
B	80-89	Good
C	70-79	Satisfactory
D	60-69	Poor
E	00-59	Failure

Requests for incomplete (I) or withdrawal (W) must be made in accordance with University policies, available at http://catalog.arizona.edu/policy/grades-and-grading-system#incomplete and http://catalog.arizona.edu/policy/grades-and-grading-system#Withdrawal respectively.

Course communication

All course materials and announcements will be made through D2L. Urgent or private communication will be made along with the official University of Arizona email addresses of students with @email.arizona.edu extensions.

Nondiscrimination and Anti-harassment Policy

The University of Arizona is committed to creating and maintaining an environment free of discrimination. In support of this commitment, the University prohibits discrimination, including harassment and retaliation, based on a protected classification, including race, color, religion, sex, national origin, age, disability, veteran status, sexual orientation, gender identity, or genetic information. For more information, including how to report a concern, please see: http://policy.arizona.edu/human-resources/nondiscrimination-and-anti-harassment-policy

University Policies

All university policies related to a syllabus are available at: https://academicaffairs.arizona.edu/syllabus-policies. By placing this link in your syllabus, you no longer need to have each individual policy included in your syllabus.

Subject to Change Notice

Information contained in the course syllabus, other than the grade and absence policies, may be subject to change with reasonable advance notice, as deemed appropriate by the instructor of this course.

Graduate Student Resources

For a list of programs and resources for students please access the University of Arizona's Basic Needs Resources page: http://basicneeds.arizona.edu/index.html