CMU ONLINE
Carnegie Mellon University

Syllabus Overview Sample: Telling Impactful Stories with Data Visualization

Program: Applied Data Science

Faculty: Dr. Ron Yurko, Assistant Teaching Professor, Dietrich College

Course Number: 36642 **Units:** 6

Zoom Class: Weekday Evening, ET

Course Description

This course introduces students to the most common forms of graphical displays and their (mis)uses. Students will learn how to create well-designed graphs and understand them from a statistical perspective, while working with increasingly common, complex data structures (temporal, spatial, and text data). All assignments will be completed in R. Throughout the course, communication skills will play an important role.

Course Learning Objectives

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

- Apply principles of data visualization to evaluate visualization approaches.
- Identify use cases for different kinds of visualizations for various types of data.
- Create visualizations to explore datasets and make statistical claims.
- Communicate conclusions with data visualizations.

Faculty Introduction



Dr. Yurko is an Assistant Teaching Professor in the Department of Statistics & Data Science at Carnegie Mellon University. His research focuses on developing methods at the interface of inference and machine learning, oriented towards problems in natural language processing, statistical genetics, and sports analytics.

He received his PhD in Statistics at Carnegie Mellon University where his thesis work focused on selective inference approaches for augmenting genetic association studies with data-guided hypothesis testing.

He also actively publishes research on statistics in sports, and has developed multiple R packages to enable easy access of publicly available data. Additionally, he co-organizes the annual Carnegie Mellon Sports Analytics Conference, which promotes the latest sports analytics research and features a reproducible research competition.

Course Topics

Topic	Learning Objectives
Intro and Visualizations for Categorical Data	● Describe the role of data visualization. ● Describe the fundamental principles of data visualization techniques. ● Build visualizations using tidy data. ● Use the Grammar of Graphics for building data visualizations. ● Display categorical data with area plots.
Visualizations for 1D Quantitative Data and Across Categories	● Visualize 1D quantitative data to inspect center, spread, and shape. ● Describe the pros and cons of boxplots, histograms, and ECDFs. ● Perform one-sample KS test to assess whether data follows assumed distribution. ● Visualize and test differences between empirical distributions. ● Create different types of density curve visualizations for comparison of conditional distributions.
Visualizations for 2D Data + Pairs Plots	● Use scatterplots for visualizing 2D quantitative data. ● Visualize relationships between two quantitative variables with regression lines. ● Visualize 2D joint distributions with contours, heat maps, and hex bins. ● Explain the role of pairs plots for visualizing several variables at once.
Visualizing High-Dimensional Data	● Explain the role of dimension reduction in visualizing high-dimensional data. ● Use multi-dimensional scaling to preserve high-dimensional distances in lower-dimensional visualizations. ● Visualize structure among variables based on correlations. ● Describe how principal component analysis (PCA) retains variance in data. ● Evaluate and interpret the output of PCA.
Spatial Data	● Describe how spatial data are typically stored.

	● Describe different types of spatial data and appropriate visualizations. ● Explain how to perform spatial inference with regression. ● Explain the role of choropleths for visualizing areal data.
Text Data	● Describe text as data using the bag of words representation. ● Use word clouds and comparison clouds for visualizing word frequencies. ● Explain the use of TF-IDF in measuring uniqueness. ● Measure text sentiment with a dictionary-based approach.
Trends, Animations, and Interactive Visualizations	● Describe text as data using the bag of words representation. ● Use word clouds and comparison clouds for visualizing word frequencies. ● Explain the use of TF-IDF in measuring uniqueness. ● Measure text sentiment with a dictionary-based approach.

NOTE: This is a sample designed to provide an overview of topics covered in the course based on previous course syllabi. All students will be provided an updated and final syllabus at the start of the course. Faculty, content, and schedule are all subject to change.