

Elements of Data Visualization Course Syllabus

Dr. Philip E. Cannata

email: cannata@cs.utexas.edu

Course Description and Objectives:

This course teaches the essential and practical skills necessary to communicate information about data clearly and effectively through graphical means. Rendering data clearly and effectively with appropriate visual analytics reduces the time required to achieve understanding and helps in managing the every growing amount of available digital data. Students will learn to use various software tools including R, ggplot, Tableau 8, and Oracle APEX 5 during the class.

R is the most popular free software environment for statistical computing and graphics. ggplot2 is a data visualization package for R that that can be used to produce publication-quality graphics. This course will show how to use R and ggplot to not only produce production-quality graphics but also how to produce large multiplot images (i.e., dozens and dozens of different plots in one image) that can be used as a standardized form of analysis.

Tableau Software is one of the most popular commercial data visualization tools on the market today. In this course you will learn how to use this tool to quickly analyze, visualize and share information. Tableau can help you see and understand your data whether it's stored in flat files, databases, warehouses, cubes or in the cloud.

Oracle Application Express (APEX) is a rapid application development tool that lets you create custom, professional Web and Mobile data visualization applications using only a Web browser and limited programming.

There will be a strong emphasis in the class on the visualization of data stored in database management systems such as MySQL, Oracle and possibly NoSQL databases.

Prerequisites:

- a grade of C or better in cs313e or equivalent
- the class will work in groups of 2 or 3 students all semester, each group needs to have at least one laptop in class each class day.

Required Texts:

- *R for Everyone: Advanced Analytics and Graphics*. Jared P. Lander, Addison-Wesley Professional. (ISBN: 978-0321888037)
- *Tableau 8: The Official Guide*. George Peck, McGraw-Hill Osborne Media. (ISBN: 978-0071816786)

Tools:

- R and ggplot are open source software and free to use.
- Student licenses of Tableau 8 will be made available free of charge by Tableau Software for the duration of the class.
- Oracle APEX software is free to use and is hosted in the cloud by Oracle.

Class Discussion Tool:

Piazza will be used for this purpose.

- I will post class announcements and information to the Piazza discussion group
- post your questions about the class to the Piazza discussion group. Do not post solutions or chunks of code > 3 lines on Piazza,

Grading:

- There will be 8 quizzes over the course of the semester each worth 10 points for a total of 80 points.
- There will be 10 lab projects over the course of the semester each worth 20 points for a total of 200 points. Students will work in groups of 2 or 3 on these labs.

Groups will be expected to complete the lab projects on your own using the provided code when given.

Class examples and examples from the web can be used, but these must be documented. To be clear, each group must do a significant amount of the work on the lab projects by themselves.

- Each group will have to complete a separate data visualization project of their choosing (after consulting with Dr. Cannata) which will be worth 50 points.

Grading will be on a straight scale as follows:

A =	100-90%	(297-330 pts)
B =	89-80%	(264-296 pts)
C =	79-70%	(231-263 pts)
D =	69-60%	(198-230 pts)
F =	<60%	(below 198 pts)

Course Topics

- Connecting to various data sources
- R Basics:
 - Vectors
 - Factors

- Sequences
- Subsetting
- Matrices and Arrays
- Lists
- Data Frames
- Functions
- ggplot2:
 - graphical language
 - geoms and aesthetics
 - line and point plots
 - scatterplots
 - histograms
 - boxplots
 - density plots
 - bar charts
 - radar (star) plots
 - heat maps
 - error bars
- Tableau:
 - Customizing your view of the data
 - Chart types
 - Viewers
 - Dragging Fields
 - Shelves and Cards
 - Multidimensional Hierarchies
 - Drilling Down and Up in a Hierarchy
 - Perfect Pivoting
 - Maps
 - Dashboards
 - Advanced Analytics:
 - Actions
 - Calculations
 - Forecasting
 - Parameters
 - Background Images
 - Background Maps
 - Trend Lines and Statistics
 - Log Axes
 - Integration with R
 - Distributing and Sharing visualizations
- APEX
 - Rapid Web/Mobile application development
 - Flash and HTML 5 Charts
 - Column
 - Horizontal Bar
 - Pie & Doughnut

- Scatter
- Scatter
- Line
- Candlestick
- Gauges
- Gantt