USC Marshall
School of Business

DSO 545: Statistical Computing and Data Visualization
Fall 2019
3 units

Instructor: *Abbass Sharif, PhD*
Office: *BRI 308 F*
Office Hours: *Thursday 10am-10:50am or by appointment*
Email: *asharif@marshall.usc.edu*

Sections: *16234 (Tuesday 11:00am-1:50pm, JKP 204)*
16235 (Thursday 11:00am-1:50pm, JKP 204)

I. Course Description

Statistical computing and data visualization are increasingly important and necessary aspects of a data analyst job. Whether they are dealing with small, big, structured or unstructured data, data analysts are expected to be able to access data from various sources, apply the latest statistical methodologies, and communicate their findings to others in novel visualizations.

In this course, students will learn how to make sense of data, and not the specifics of modeling. The course starts with statistical computing, and students will gain experience with a programming language called **Python**. They will learn the practice of data cleaning, reshaping of data, basic tabulations, and aggregations in order to be able to produce high quality visualizations.

Then, the course proceeds with graphics that are critical elements of modern data analysis and presentation. From initial exploration of a data set to the final presentation of results to the end user, statistical graphics play a very vital role in shaping our understanding of our data. Through proper use of graphics, we can make critical discoveries, and communicate them clearly. Conversely, poor use or misuse of graphics can seriously mislead (by accident or design).

II. Learning Objectives

In this course, the students will gain explicit experience with programming language concepts such as variables, assignments, flow control, functions, parameters, data structures, input and output, error handling, debugging, and so on. In addition, students will learn how to write a computer program in **Python** to create a graph from scratch and manipulate its different attributes (axis, title, orientation, color, etc.), identify appropriate data visualization techniques given particular requirements imposed by the data, analyze and criticize examples of visualizations from newspapers, scientific papers, business reports, and journals. In addition, we will learn how to work with large datasets that are usually hard to work with in a regular way.

III. Prerequisites

I expect basic "operational" knowledge from an introductory stats course. Prior R knowledge is not required, but you will rather learn the basics of R in this course.

IV. Software

This is a hand-on course and it is computationally intensive. We will primarily be using **Python,** and towards the end of the semester we will introduce **R.** The main IDE we will be utilizing in the course for both **R** and **Python** will be **Rstudio. Students are expected to bring their laptops to class during all class sessions.**

Python:

1. Python (Windows)
https://repo.anaconda.com/archive/Anaconda3-2019.07-Windows-x86_64.exe
2. Python (Mac)
https://repo.anaconda.com/archive/Anaconda3-2019.07-MacOSX-x86_64.pkg

R:

1. R (Windows)
https://cran.r-project.org/bin/windows/base/R-3.6.1-win.exe
2. R (Mac)
https://cran.r-project.org/bin/macosx/R-3.6.1.pkg

Rstudio:

1. Rstudio (Windows)
https://download1.rstudio.org/desktop/windows/RStudio-1.2.1335.exe
2. Rstudio (Mac)
https://download1.rstudio.org/desktop/macos/RStudio-1.2.1335.dmg

V. Methods

This class is taught through a combination of lectures, computer labs, hands on computing tasks in homework, and group project and presentations. The projects are key because they integrate multiple computational topics in the content of a modern data problem. Students gain hand on experience with statistical concepts flowing from contextual problem solving with data, and they make their own discoveries by posing and answering questions rather than solely fitting models or using "this week's lecture's methodology" as a computing exercise.

VI. Required Books

- Case in Point: Graph Analysis for Consulting and Case Interviews 2nd Edition by Marc P. Cosentino and Mukund Jain

VII. Suggested Books

- Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython 2nd Edition by Wes McKinney (online version is available for **free** through USC libraries: https://libraries.usc.edu)
- R for Data Science by Hadley Wickham: http://r4ds.had.co.nz (**free**)
- Storytelling with Data: a data visualization guide for business professionals by Cole Nussbaumer Knaflic (online version is available for **free** through USC libraries: https://libraries.usc.edu)

VIII. Assessment

- Homework
  To do well in this course you will need to spend 4-5 hours a week (outside of class!), and the homework are designed to encourage you do that. For each homework assignment, you will need to revise the week's work, as well synthesize some new information, from the help pages or the web.
- Midterm
  There will be two midterms in this class. You can use the notes provided by the professor on blackboard.
- Final project
  For your final project, you'll be expected to find your own dataset. As well as writing a report, you'll present at a formal poster session. It is a group project (groups of 5).

Assessment	% of Grade
Homework	25%
Project	20%
Case in Point Presentations	5%
Midterm I	25%
Midterm II	25%

IX. Schedule

Assessment	Date	Place
Midterm I	TBD	TBD
Project Proposal	TBD	TBD
Project Submission	TBD	TBD
Midterm II	TBD	TBD

X. Students with Disabilities

Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to me as early in the semester as possible. DSP is located in STU 301 and is open 8:30 am - 5:00 pm, Monday through Friday. The phone number for DSP is 213 740-0776.

XI. Topics:

- Data Cleaning
- Exploratory Data Analysis
- Data Visualization
- Data Science Tools: Markdown, Rstudio, and Github
- Principles of good data display
- Spatial Data Visualization
- Time series Data Visualization
- Dashboards
- Web scraping
- Regular Expressions
- Other topics: Data Analysis Using R (ggplot2 and dplyr)

Statement for Students with Disabilities

Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to me (or to your TA) as early in the semester as possible. DSP is located in STU 301 and is open 8:30 a.m.–5:00 p.m., Monday through Friday. The phone number for DSP is (213) 740-0776. For more information visit www.usc.edu/disability .

Statement on Academic Integrity

USC seeks to maintain an optimal learning environment. General principles of academic honesty include the concept of respect for the intellectual property of others, the expectation that individual work will be submitted unless otherwise allowed by an instructor, and the obligations both to protect one's own academic work from misuse by others as well as to avoid using another's work as one's own. All students are expected to understand and abide by these principles. *SCampus*, the Student Guidebook, (www.usc.edu/scampus or http://scampus.usc.edu) contains the University Student Conduct Code (see University Governance, Section 11.00), while the recommended sanctions are located in Appendix A.

Students will be referred to the Office of Student Judicial Affairs and Community Standards for further review, should there be any suspicion of academic dishonesty. The Review process can be found at: http://www.usc.edu/student-affairs/SJACS/ . Failure to adhere to the academic conduct standards set forth by these guidelines and our programs will not be tolerated by the USC Marshall community and can lead to dismissal.

Emergency Preparedness/Course Continuity

In case of a declared emergency if travel to campus is not feasible, USC executive leadership will announce an electronic way for instructors to teach students in their residence halls or homes using a combination of Blackboard, teleconferencing, and other technologies.

Please activate your course in Blackboard with access to the course syllabus. Whether or not you use Blackboard regularly, these preparations will be crucial in an emergency. USC's Blackboard learning management system and support information is available at blackboard.usc.edu.