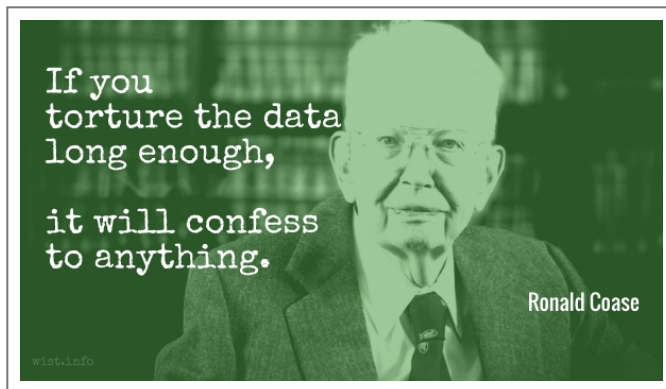# INF 250: Introduction to Data Informatics

Dr. Saty Raghavachary

M,W 3:00-4:50pm, THH B10

Preview ["Welcome to INF 250!"]

Blurb

What this course is about - here is a short description (https://classes.usc.edu/term-20183/course/inf-250/), from the USC catalog.

Overview

In this course, we are going to do a thorough overview of 'data science', which is a brand new, exciting and extremely relevant discipline. We will cover data sources, collection, storage, analysis, and interpretation (including visualization); we will also look at intersections with related areas such as Big Data, Data Mining and Machine Learning, and examine a variety of real-world applications. 'Data' is the new 'oil', and this course will explain 'why' and 'how'.

Pre-requisites

As stated in the university catalog, you don't need a lot of prep.

Note that this course involves data analysis - so there will be a bit of math, and a small amount of coding (in R, Python) - but, this is not a programming course, or a math-heavy theory course.

People

Prof

Dr. Saty Raghavachary

Office hours: 12-2, on Mon, Wed (in SAL 346)

Email: saty@usc.edu, satychary@gmail.com (I'm also on Facebook, Linkedin and Twitter - several ways to be in touch!)

I work in, teach, write about, lecture, consult, code and do informal research on computer graphics (CG). Here (http://www.viterbi.usc.edu/academics/faculty/faculty-directory/profile.php?faculty=raghavachary_saty.xml) is my brief bio.

TA, Grader

TBD.

Policies

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 a.m. - 5:00 p.m., Monday through Friday. The phone number for DSP is (213) 740-0776.

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, contains the Student Conduct Code and also the recommended sanctions for violating academic integrity. The PDF is here (https://studentaffairs.usc.edu/files/2015/11/SCampus-Final-112015_v8.pdf).

Students will be referred to the Office of Student Judicial Affairs and Community Standards (SJACS) for further review, should there be any suspicion of academic dishonesty. The Review process can be found here (https://sjacs.usc.edu/).

Copying (and then modification) of any portion of code from Internet sources or fellow students is prohibited unless cleared with the instructor. In addition, working together on a programming assignment is also a gray area. If your code resembles that of your friend's too closely, that will be considered a VIOLATION.

Here is my policy: if I come to find out that your submission resembles another submission (from a fellow student in your section or another section, a roommate's

submission from an earlier class, submission from someone else who took the class earlier, code found on GitHub etc..), **I WILL REPORT YOU TO SJACS, AND GIVE YOU A 0 FOR YOUR SUBMISSION**. You can then deal with SJACS to convince them if you need to, that you did not cheat. Please take this **very seriously**, and BELIEVE it - if you decide to flout the rules, you do so at your own risk.

Translation/bottom line: **DO NOT CHEAT under ANY circumstance(s)!!**

Emergency preparedness/course continuity in a crisis

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.

Syllabus, schedule, notes

Below, you'll find links to lecture notes - this additionally serves as the syllabus as well as the course schedule.

The lecture notes are in the form of browser-based slides [not PDF or ppt], but you can create PDFs out of them if you want, by printing to PDF using Chrome.

Week# (start date)	Topic[s]
1 (8/20)	Introduction(s) (Intro/slides.html) Data Science - an overview (Overview/slides.html)
2 (8/27)	Data sources, representation (SourcesRepr/slides.html) Data storage, formats (StorageFormats/slides.html)
3 (9/3)	EDA [Exploratory Data Analysis] (EDA/slides.html) [includes a review of statistics]
4 (9/10)	Big Data (BigData/slides.html)
5 (9/17)	Data Mining (DM/slides.html)
6 (9/24)	Machine Learning (ML/slides.html)
7 (10/1)	Ethics, privacy issues (Governance/slides.html) [includes metadata, "governance"]
8 (10/8)	Spatial data (Spatial/slides.html) **10/10: Midterm exam**
9 (10/15)	Visualization (Viz/slides.html)
10 (10/22)	Applications: recommender systems, NLP (Appl1/slides.html)
11 (10/29)	Applications: voice assistants, chatbots (Appl2/slides.html)
12 (11/5)	Applications: social media (Appl3/slides.html)
13 (11/12)	Applications: medical data (Appl4/slides.html)
14 (11/19)	Applications: more (Appl5/slides.html)

Week# (start date)	Topic[s]
15 (11/26)	What's next? (More/slides.html) [for you, for data science]
17 (12/10)	**12/10: FINAL exam, 2-4pm**

There will be four homeworks (worth 15% each), and two exams (midterm and final, each worth 20%).

Textbook - optional

Our book is 'The Data Science Design MANUAL'
(https://www.amazon.com/Science-Design-Manual-Texts-
Computer/dp/3319554433), by Steven S. Skiena.

Supplementary material

As we progress through the course, there might be additional
content (PDFs, links to sites, etc) that will be put up here.