CS-199 Big Data Intro

D.A. Forsyth

Key ideas

- Understanding data is important to all engineers
  - a bit like, say, calculus
- By the end, you should
  - have some experience with data at scale
    - including finding, downloading, cleanup
  - know a bit about, and/or have used, tools for
    - data lifecycle
    - visualization
    - clustering
    - classification
    - regression
  - have some experience with environments for understanding data
  - have worked with data of different types, in different contexts

Course Staff

- David Forsyth, CS, (lead)
  - any problems, it's my fault
- Mani Golparvar-Fard, CEE
- Farzad Kamalabadi, ECE
- Romit Roy Choudhury, ECE
- Richard Sowers, ISE
- Matthew West, ME
- ChengXiang Zhai, CS

Simple Admin stuff

- We meet once a week
  - Tue, 16h00-17h15, Siebel 1131
- It's a pilot
  - We're figuring out how to do this properly
  - You're here to tell us what works
  - Not everything will work or be at the right level - don't panic
  - Participate, do the work, and you'll get a nice grade
- Work
  - practical homeworks, discussion in class
  - we'll try regulate workload to be consistent with two hours; keep us posted

Simple Admin stuff

- Class web page:
  - google David Forsyth,
    - choose the right one, go to web page,
    - look on “Teaching” tab
  - slides, reading material, etc.
- Lectures will cover reading material rather superficially
  - you’ll need to read stuff on webpage

Recommended reading

- Notes on web page
- Doing Data Science
  - Look on web page for amazon URL, etc.
  - http://www.amazon.com/Doing-Data-Science-Straight-Frontline/dp/1449358659/ref=sr_1_1?ie=UTF8&qid=1390246614&sr=8-1&keywords=doing+data+science

Assignment 1

- REGISTER (so I have your email)
- Download and install R, and try some of the things I'll demonstrate
  - google is your friend
- Find a neat web dataset at some scale
- Suggest questions to be attacked by visualization
- Email suggestions to daf@uiuc.edu by Friday
  - subject: 199 Dataset
- I'll choose one, circulate URL via email, make some plots, and we can discuss