

Intro to Data Science  
CSCI 183, Spring 2015

Instructor: George Mohler  
office: O'Connor 04  
phone: 554-4544  
e-mail: gmohler@scu.edu

**Course overview:**

Data manipulation, analysis, and visualization. Statistical modeling, dimension reduction and techniques of supervised and unsupervised learning. Big data software technologies. Prerequisites: The ability to program in some scientific language, and MATH 122, or permission of the instructor.

**Course topics:**

- Data “munging” and exploratory analysis
- Classification: logistic regression, decision trees, cross validation
- Regression: least squares, random forests, boosting
- Dimension reduction: PCA, visualization of high dimensional data, feature selection, LASSO
- Neural networks and applications to computer vision
- Big data: nosql, Hadoop, EC2

**Text:** Doing data science by Cathy O’Neil and Rachel Schutt

**Software:** Homework assignments will require the use of R, Kaggle in-class, and Github. Other open-source software may be required for some assignments and further details will be given in class.

**Homework:** Homework assignments will be assigned weekly. Because the class will be partially flipped, homework may involve assigned reading or videos in addition to programming assignments.

**In class Kaggle competitions:** There will be three in class Kaggle competitions. The deadlines will be the end of 4th week, the end of 7th week, and the end of 10th week. More details on these will be given the first day of class.

**Blog assignment** Instead of a final paper or exam, you will each need to complete a 1500 word blog post on the topic of your choosing. This post should have quantitative analysis and visualizations in the vain of fivethirtyeight.com. You can post the blog (or a rough draft) and send me a link anytime after 7th week and before the last day of finals.

### **Grading:**

Homework 50%    In-class Kaggle competitions 30%    Blog assignment 20%

The grading scale will depend on the will not be lower than the standard scale: A 90-100, B 80-89, C 70-79, D 60-69 and may be curved.

**Academic Integrity:** The penalty for cheating is a failing grade for the course, and the University may take further disciplinary action. All of the work that you turn in should be your own, and not that of a classmate or copied from another source. Please see the following webpage for further information: <http://www.scu.edu/studentlife/resources/integrity.cfm>

**Disability Accommodation Policy:** To request academic accommodations for a disability, students must contact the Disability Resources Office located in Benson room 216, (408) 554-4111; TTY (408) 554-5445. See <http://www.scu.edu/advising/learning/disabilities/> for further information.