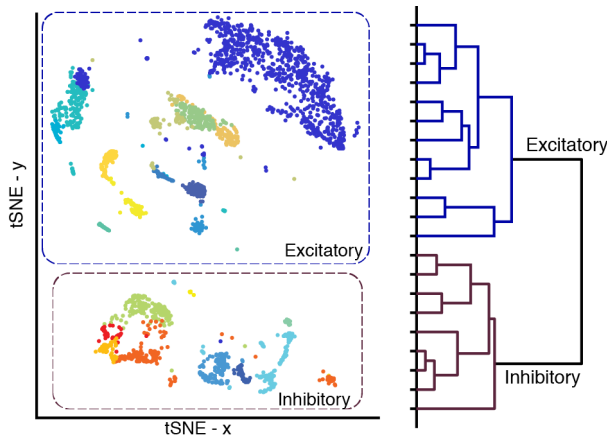


## Cogs 109, Fall 2018: Modeling and Data Analysis

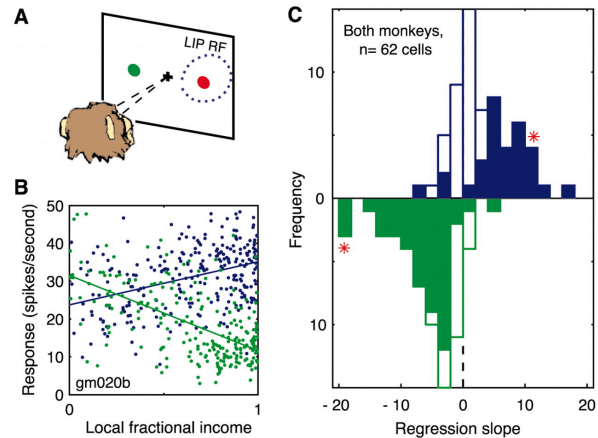
Data surrounds us. Scientists use data to build and test our understanding of the natural and physical world. Governments, civic organizations and businesses use data to improve products and services, and athletes use data to enhance performance. Cognitive scientists have a double interest in data: We use experimental data to understand the brain and behavior; at the same time, we seek to understand how our brains analyze a constant stream of sensory data. In this course, you will learn how the fundamental techniques of data modeling and analysis can be used to derive insights across a broad range of domains. These techniques are sometimes inspired by, and in turn they inspire, new understanding of our own cognitive abilities. Data is the starting point for all of our empirical (experience-derived) knowledge of the world, and this course introduces the mathematical, statistical and computational ways of thinking that can unlock the insights hidden in the numbers.

Specific topics will include: Data visualization and exploratory analysis, Prediction, Linear models, Statistical inference, Clustering, Dimensionality reduction and principal component analysis, Clustering



Luo, Keown et al., *Science* (2017); <http://brainome.org>

*DNA methylation data from single human brain cells allows clustering of neuron cell types.*



*Modeling neuronal control of behavioral choices in a monkey.*

Sugrue, Corrado, Newsome, *Science* (2004)

**Prerequisites:** Cognitive Science 14B and Mathematics 20F and CSE 7 or CSE 11 or CSE 8A or consent of instructor

**Important dates:**

Midterm exam in class in Week 5

Final exam: Take-home, Thursday 12/13 (you will have 24 hours to complete it)

**Time and location:** TTh 2:00 – 3:30 PM, Solis 104

**Instructors:**

Prof. Eran Mukamel mukamel@ucsd.edu Office: SSRB 255 Office hour: Th 3:30-4:30 pm, or by appointment	TA: Connie Guan cguan@eng.ucsd.edu Office hour: Th 10-11 am, CSB 114 Sections: F 11-11:50, CSB 115 F 12-12:50, CSB 115	TA: Justin Lazarow jlazarow@eng.ucsd.edu Office hours: Tu 1-2 pm, CSB 114 Sections: M 1-1:50 pm, CSB 115 W 3-3:50 pm, CSB 115	Instructional assistants (office hours in CSB 114) Qi Leng - <a href="mailto:qleng@ucsd.edu">qleng@ucsd.edu</a> , OH: W 1-2 pm Severine Soltani – <a href="mailto:psayar@ucsd.edu">psayar@ucsd.edu</a> OH: M 12-1 pm Yiwei Zhao - <a href="mailto:yiz486@ucsd.edu">yiz486@ucsd.edu</a> OH: Tu 4-5 pm
---	---	--	--

**Learning objectives:** Students in Cognitive Science 109 will learn to:

- Understand core concepts in analyzing and interpreting data, such as prediction, inference, model complexity, and data dimensionality
- Use and critically evaluate fundamental data analysis techniques such as regression, clustering, and principal component analysis
- Create informative data visualizations to aid hypothesis generation
- Use statistical methods to test associations in data sets

**Evaluation and grading**

- 30% - Weekly problem sets due before the start of class, beginning week 1 (Thursday 10/4). Turn in your assignments electronically at gradescope.com. Each student's lowest homework grade will be dropped; no late assignments will be accepted without a valid official document (e.g. doctor's note).

- 20% - Midterm exam
- 20% - Final project
- 20% - Final exam
- 10% - Participation in section and lectures. Attend 6/9 sections to receive full credit.
- Extra credit: Participate as a SONA research subject, up to 2%. (Info: <http://www.psychology.ucsd.edu/undergraduate-program/undergraduate-resources/sona>)

**Academic integrity:** Instructors and students are expected to honor the UCSD Policy on Integrity of Scholarship (linked from the course website).

### Course Materials

Website: <https://triton.ed.ucsd.edu>

Discussion board for questions: [piazza.com/ucsd/fall2018/cogs109](https://piazza.com/ucsd/fall2018/cogs109)

**Textbook:** *Introduction to Statistical Learning with applications in R (ISLR)* by James, Witten, Hastie and Tibshirani. Although the book includes examples in R, students in Cogs 109 may complete assignments in either R, MATLAB or Python. The book, as well as related data sets and code, is available as a free PDF download at: <http://www.bcf.usc.edu/~gareth/ISL/>

Additional references (optional):

1. *Applied Predictive Modeling*, Kuhn and Johnson:  
Similar to ISLR, but covering a broader range of topics in the style of a handbook/reference manual.
2. *The Signal and the Noise*, Nate Silver  
A popular book on data and statistics, with applications in fields from poker to baseball to climate.
3. *Pattern Recognition and Machine Learning*, Bishop  
A key introductory text for advanced concepts and methods in machine learning.

Date	Topics	Chapter(s) in ISLR, assignments
<b>Week 1</b> <b>9/27-10/4</b>	Introduction: Data Exploration, Modeling and Statistical Learning	Chapters 1, 2 HW1 due Thursday 10/4
<b>Week 2</b> <b>10/9-10/11</b>	Linear and multiple-linear regression	Ch. 3
<b>Week 3</b> <b>10/16-10/18</b>	Classification	Ch. 4
<b>Week 4</b> <b>10/23-10/25</b>	Resampling and cross-validation	Ch. 5
<b>Week 5</b> <b>10/30-11/1</b>	Model selection and regularization	<b>Tuesday 10/30: Midterm exam in class</b> Ch. 6.1, 6.2
<b>Week 6</b> <b>11/6-11/8</b>	Regularization and dimensional reduction; Non-linear regression <i>Election Day 11/6 – remember to vote!</i> <a href="http://sovac.ucsd.edu/register/">http://sovac.ucsd.edu/register/</a>	Ch. 6.3, 6.4, Ch. 7.1
<b>Week 7</b> <b>11/13-11/15</b>	Non-linear regression	Ch. 7.2-5
<b>Week 8</b> <b>11/20</b>	Support vector machines	<b>Thursday 11/22: Happy Thanksgiving! No class.</b> Ch. 9.1-9.3
<b>Week 9</b> <b>11/27-11/29</b>	Unsupervised learning and clustering	Ch. 10
<b>Week 10</b> <b>12/4-12/6</b>	Review and project presentations	
<b>Final exam</b> <b>12/13, 9 am</b>	Take home exam (24 hours)	