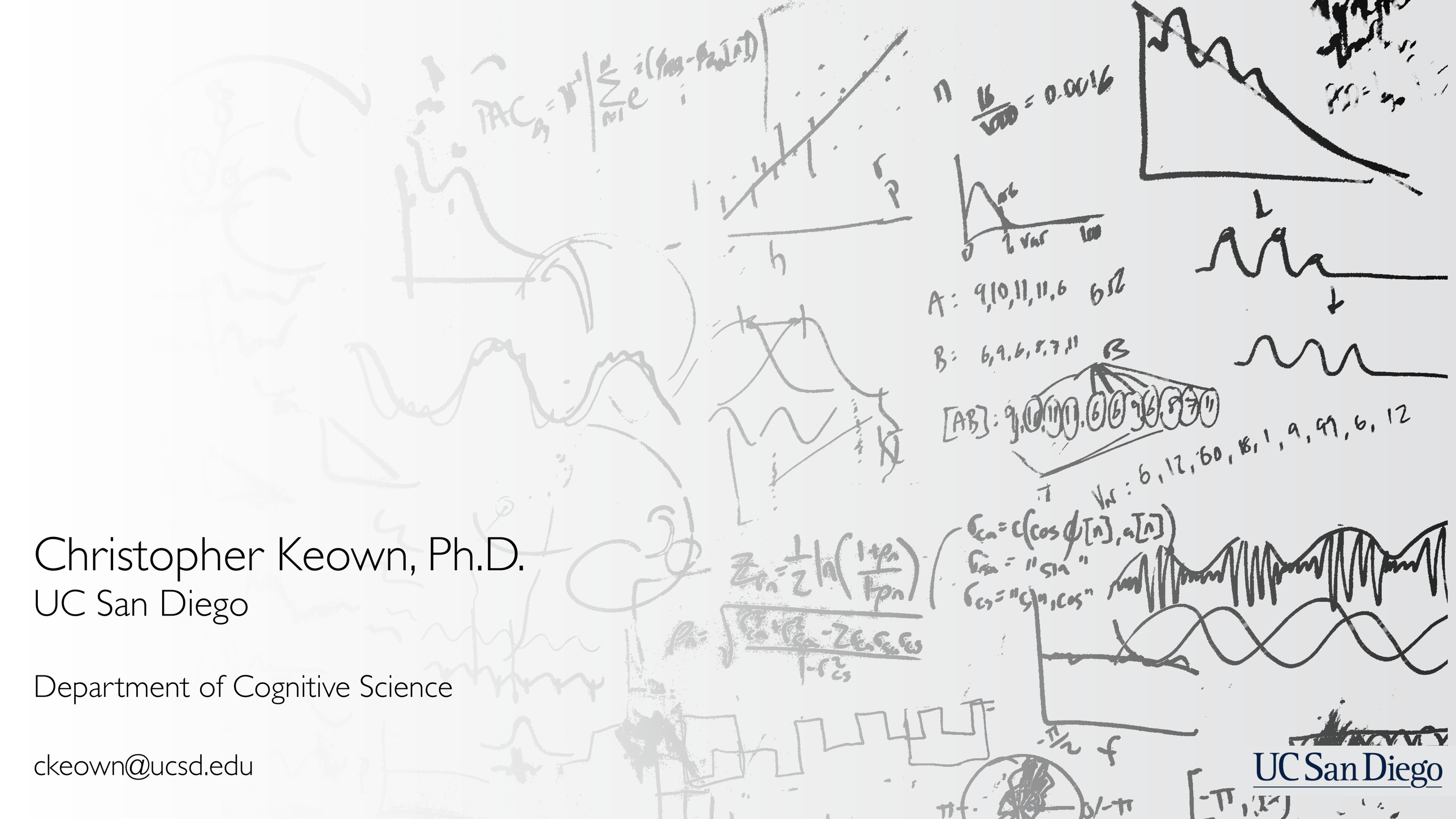


Christopher Keown, Ph.D.

UC San Diego

Department of Cognitive Science

ckeown@ucsd.edu

UC San Diego

Syllabus Clarification

- Five assignments (12% each)
- Final Project (35%)
- Participation (5%)
- Section attendance does NOT count toward your grade and is NOT mandatory.
- First guest lecture is on January 24th.

Other Questions

- Piazza: Not listed there? Let me know!
- Groups: ONLY for Final Project, *not* home works!
  - 3-6 people; you choose your own groups

Office Hours

- TBD

Sections and homeworks

- Sections are *not* mandatory
- But bring a computer if you can
- Python, Jupyter, and git *before sections*! See the git tutorials section.
- Sections: intend to go to the section for which you're registered, but you can attend another if needed
- Sections are a great place for assignment guidance
- TA office hours will be in the computer lab, giving you more time to work on assignments
- Assignments will be released by Monday mornings
- Assignments will be due Sunday nights
- Assignment solutions will be posted one week after deadline
- Late assignments graded at 50% penalty; no late assignments allowed once solutions are posted

WAITLIST

- Still ~140 students long
- Waitlist exists for the first two weeks of the quarter—wait and see what happens.
- I can't manipulate the list to let anyone in, even if you're a graduating senior, etc.
- Two sections offered in spring.
- Need this course ASAP? Talk to your academic advisor (e.g., Thanh Maxwell for CogSci).

How to get a job—without a PhD?

All of the applications look the same.
How do you stand out?

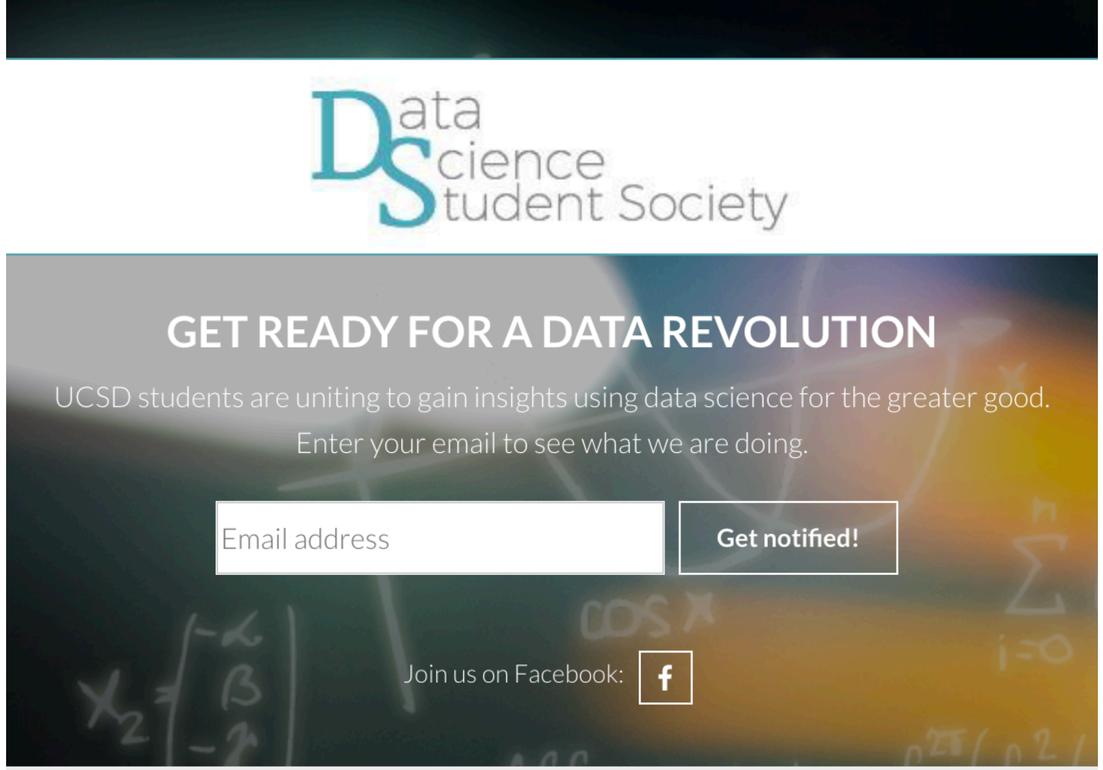
How to get a job—without a PhD?

- Know someone
- DO something that stands out

UC San Diego

Get involved—Learn more and network!

- HDSI events page: https://datascience.ucsd.edu/news-and-events/events.html
- Event Brite: https://www.eventbrite.com/d/ca--san-diego/data-science/
- Research



http://ds3.ucsd.edu/

About

The Data Science Student Society at UCSD is an interdisciplinary academic organization designed to immerse the community in the diverse and growing facets of Data Science: Machine Learning, Computational Statistics, Data Mining, Visualization, Predictive Analytics, and any new emerging relevant fields of study. With practical hands-on data projects, a professional portfolio-building approach, and fun outreach activities, the Data Science Student Society at UCSD strives to enrich the academic life of the student community by strengthening them for success in their current and future pursuits of Data Science related fields.

San Diego Machine Learning

12
JAN

Saturday, January 12, 2019

Kaggle Competition - New Location, New Contest!

Hosted by Reza and 2 others
From San Diego Machine Learning
Public group

You're going 8 people going

Share:

Details

The best way we can learn is by doing, and what better way than to participate in a Kaggle contest! Bring your laptops and come prepared to work :-)

Organizer tools

Saturday, January 12, 2019
12:00 PM to 4:00 PM
Every week on Saturday
Add to calendar

The Sandbox
3770 Tansy Street Suite 101 · San Diego, ca

What does a professor do?

Teaching

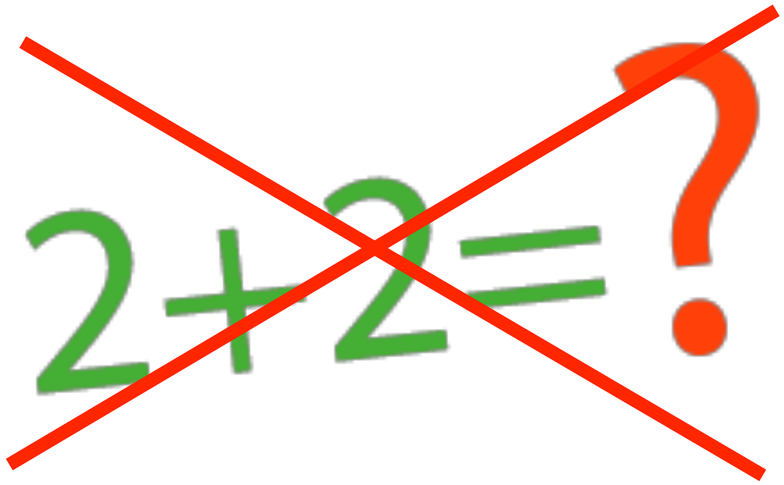
Teaching

Research

What does research look like?

History of knowledge

Didn't have access to a lot of data.
Problems could be solved through logic.



Collect data and make sense of it!

Deep brain stimulation

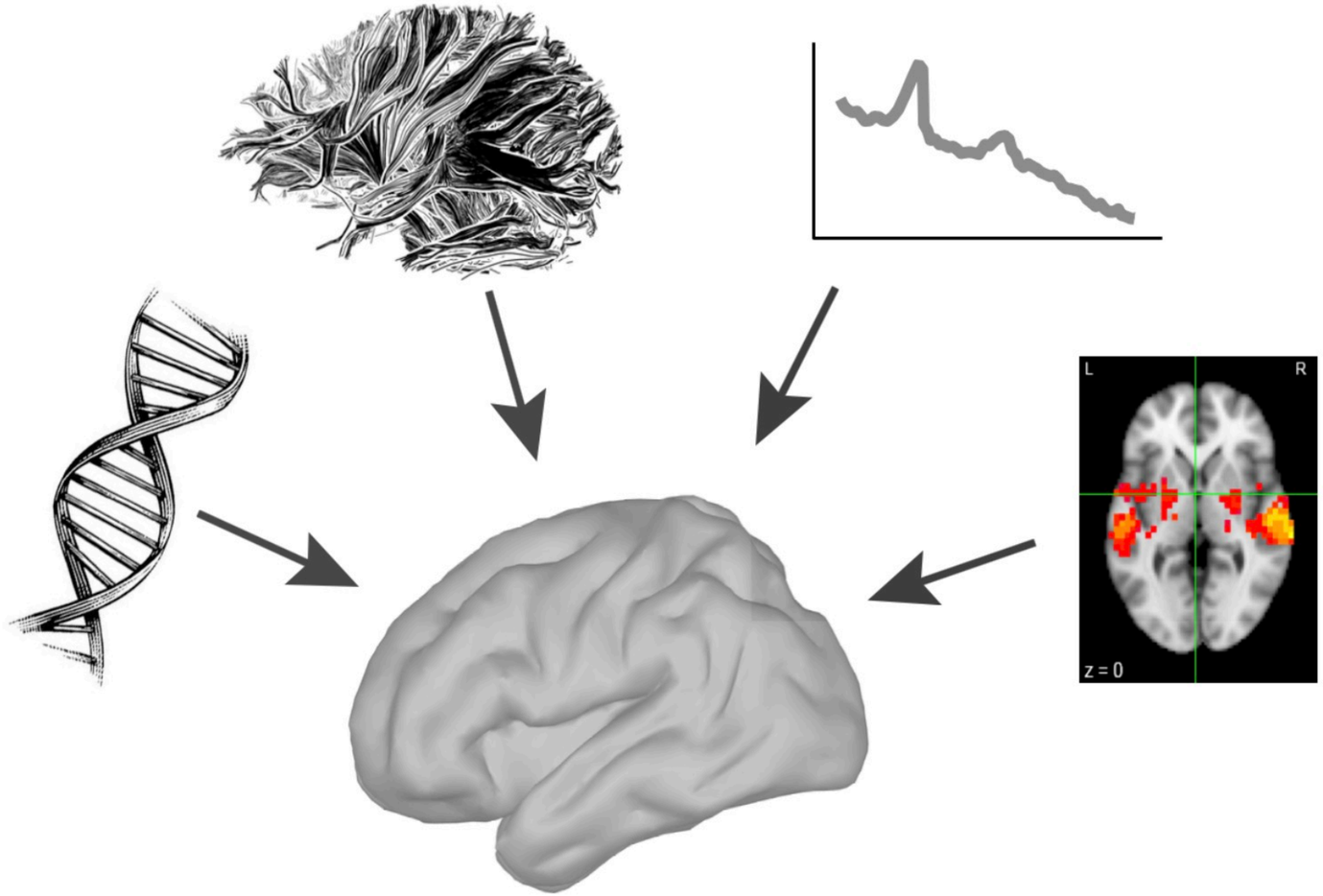
Deep brain stimulation

WHY DOES DBS WORK?

WARNING

Electrocorticography

Multimodal comparisons



Source: Donoghue *et al.* (in preparation)

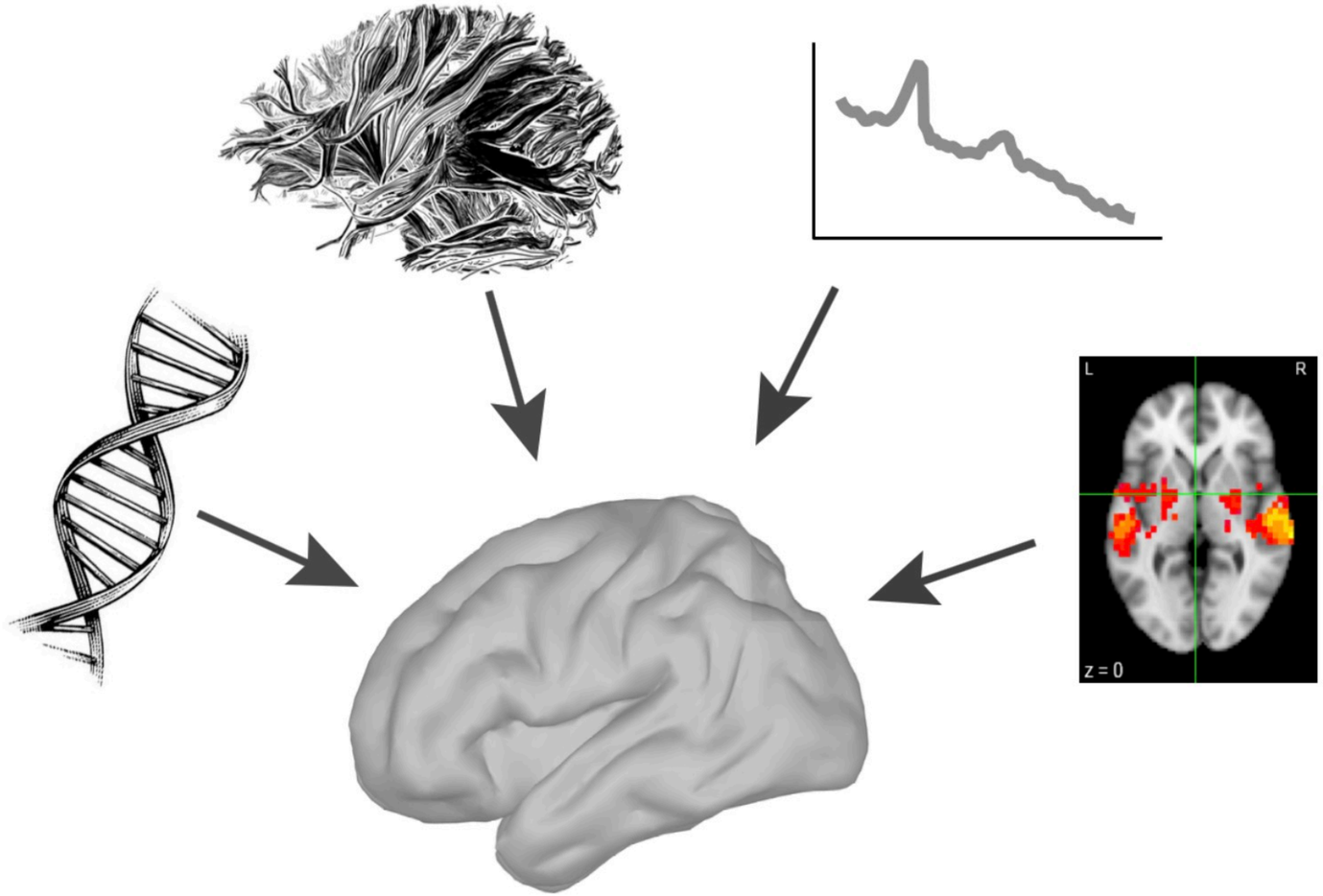
Genomics


TTGGCCAGCTGGACGAGGGCGATGAC
TTGGCCAGCTGGATGAGGGCGATGAC
Control
Autism

Mouse gene knockout

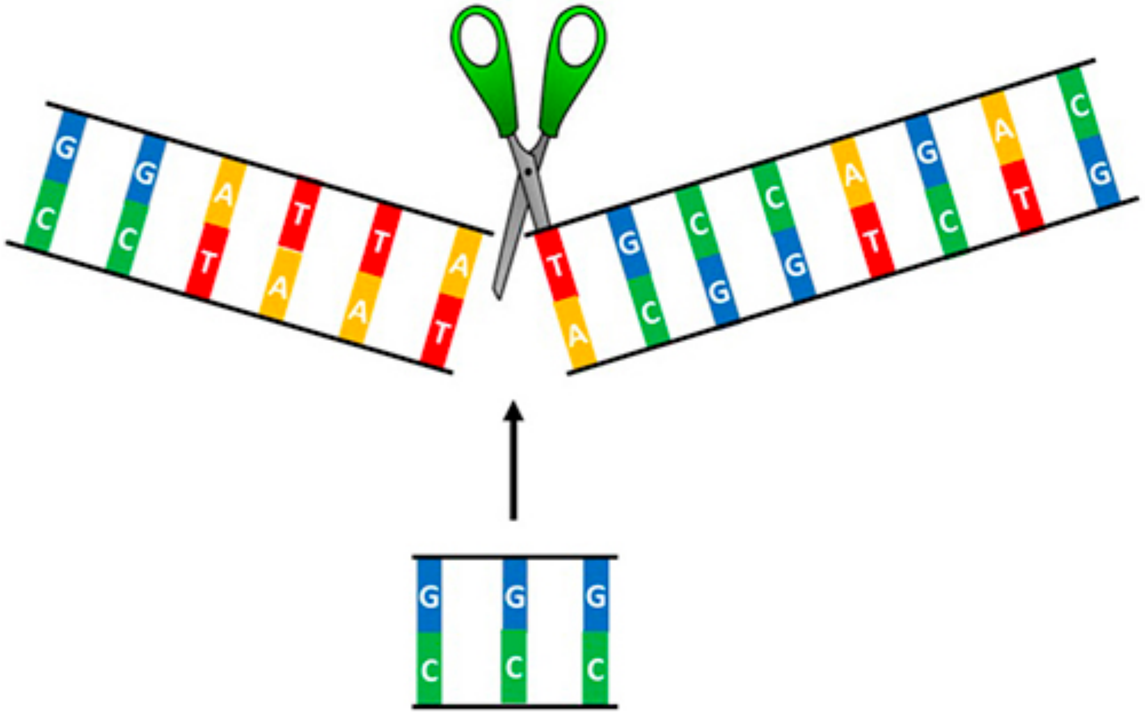
Source: Donoghue *et al.* (in preparation)

Multimodal comparisons



Source: Donoghue *et al.* (in preparation)

The Future: CRISPR—repair the gene!



Source: Donoghue *et al.* (in preparation)

COGS 108
Data Science in Practice

Why and how do we analyze data?

UC San Diego

Proposed course order

1. Introduction: Why data analysis? (prediction and classification)
2. Python!
3. Data Science in Python (jupyter, pandas, numpy, scipy, scikit-learn, etc.)
4. Data Science in Python, Part II
5. Data gathering, wrangling, and cleaning (How do you find and clean data? (JSON, CSV, XML, SQL, APIs)
6. Jan. 24: Guest lecture – Ryan Chesler to discuss Kaggle
7. Basic data visualization
8. Data privacy, ethics, and HIPAA (anonymization)
9. Data intuition and the "small n" Fermi estimation; distributions and outliers (histograms, CDF, PDFs)
10. Geospatial analysis
11. Linear modeling or non-parametric?
12. OLS (optimization) and multiple linear regression and collinearities
13. Model validation (bootstrapping, resampling, k-fold, leave-p-out, train/test)
14. Dimensionality reduction (PCA); clustering and classification (k-means, knn, SVM)
15. Feature selection
16. NLP and text-mining (bag of words, tf-idf, sentiment analysis)

YOU CAN LITERALLY TAKE AN ENTIRE CLASS ON EACH OF THESE

UC San Diego

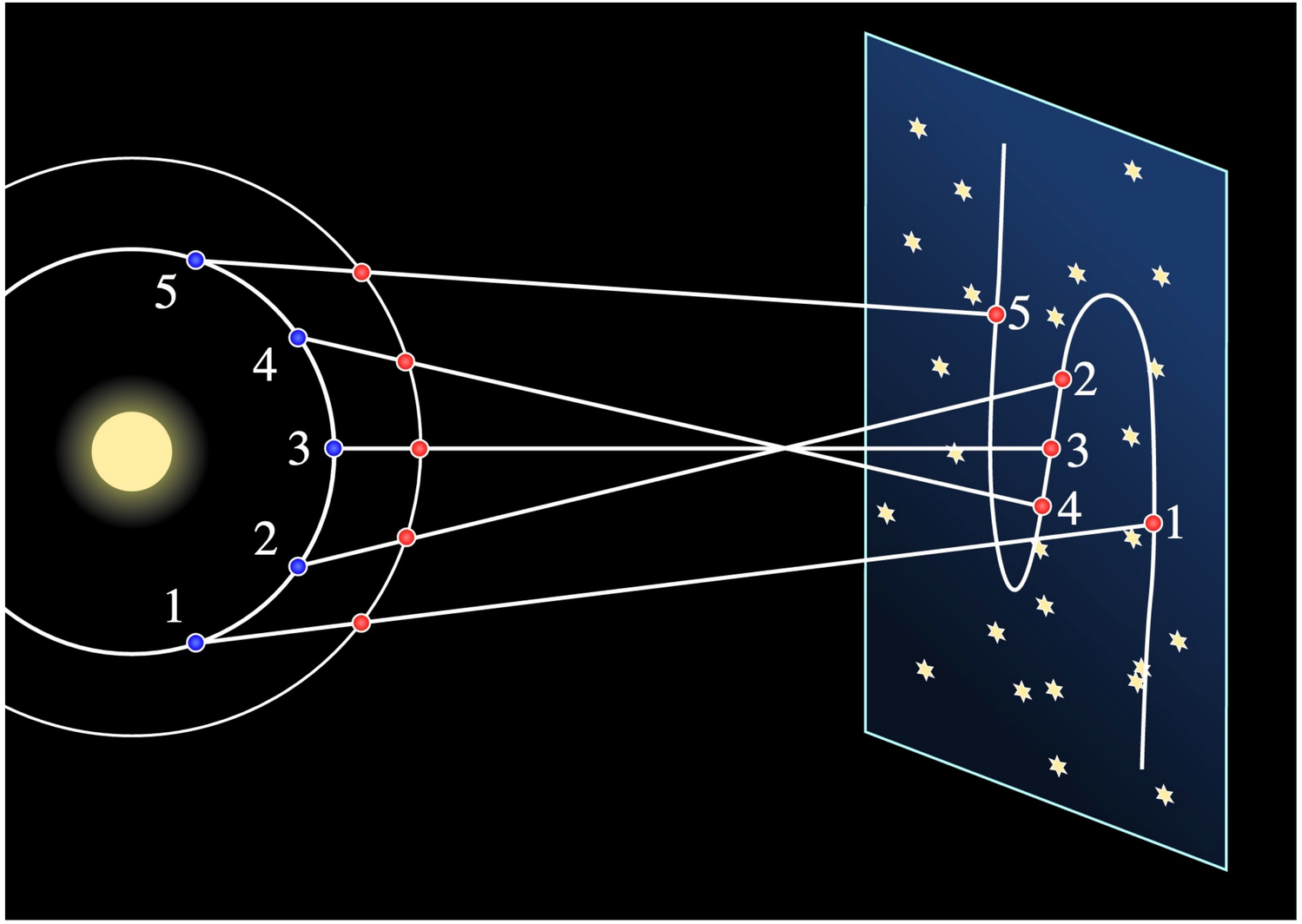
What is the point of data analysis?

- Prediction
- Classification
- Knowledge discovery?

What is the point of data analysis?

- Prediction
- Classification
- Knowledge discovery?
- **Doing useful stuff!**

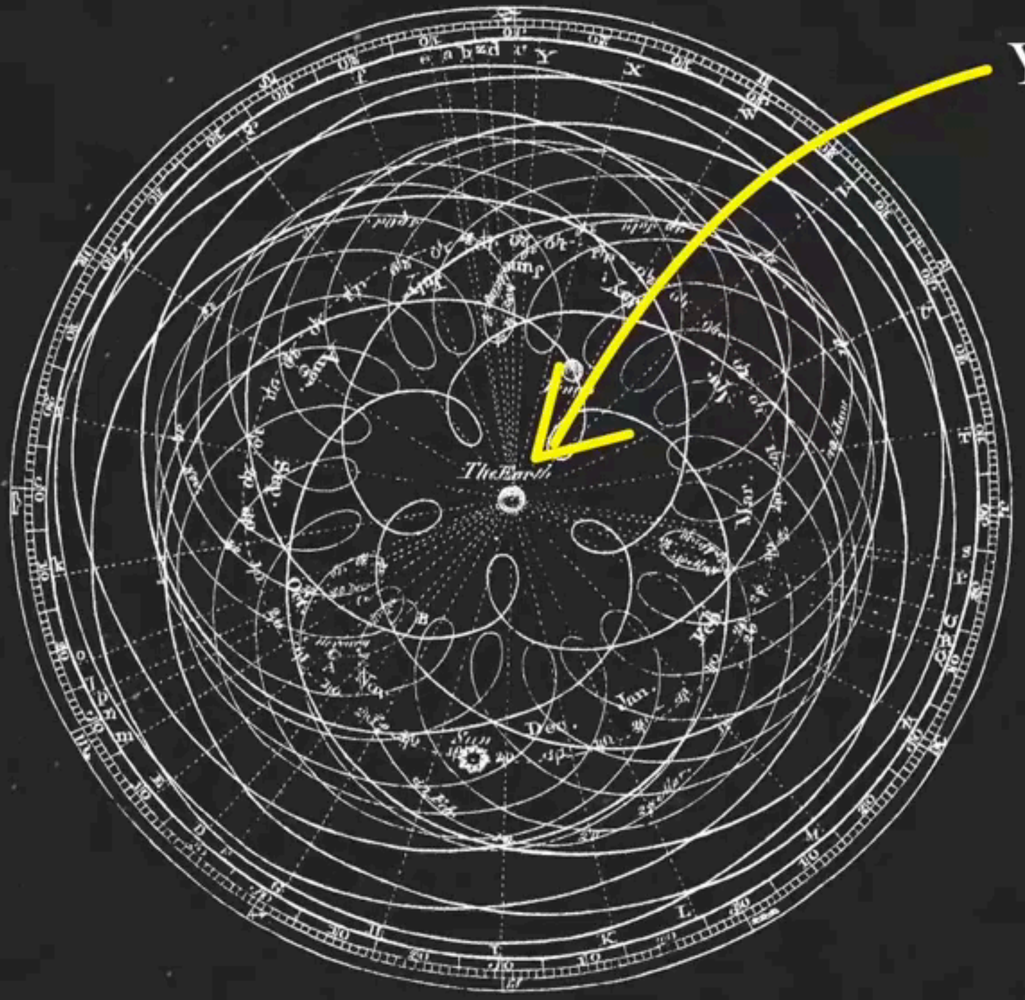
Prediction



Source: Wikipedia

Vox

OBSERVATORY


YOU ARE HERE
The Earth

Models

2.3 Parsimony

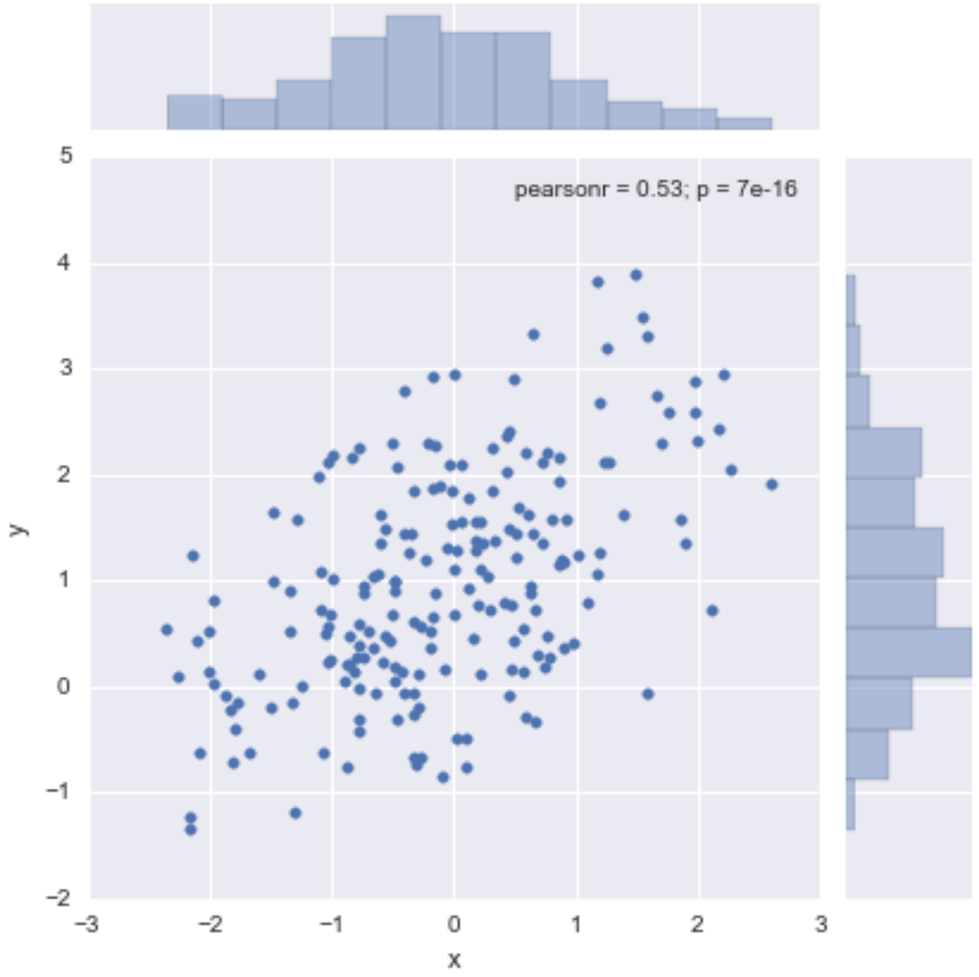
Since all models are wrong the scientist cannot obtain a "correct" one by excessive elaboration. On the contrary following William of Occam he should seek an economical description of natural phenomena. Just as the ability to devise simple but evocative models is the signature of the great scientist so overelaboration and overparameterization is often the mark of mediocrity.

2.4 Worrying Selectively

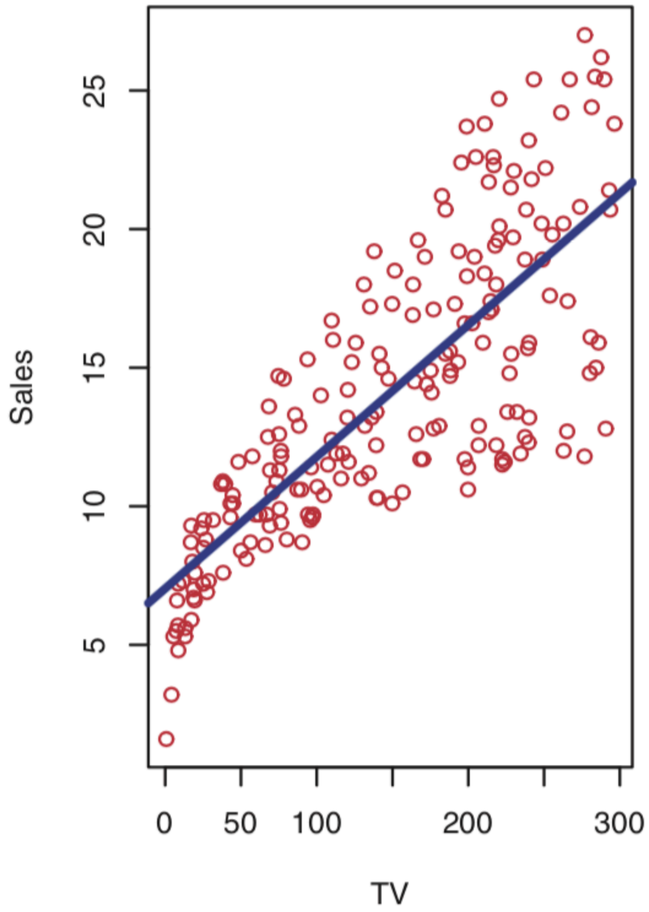
Since all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad.

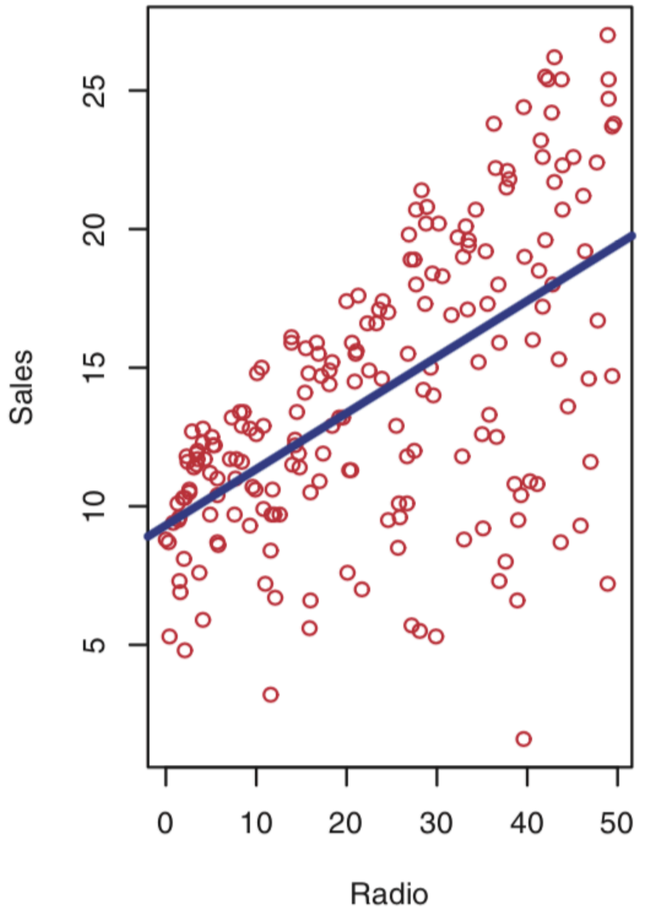
Source: Box GEP, *J Am Stat Assoc*, 1976

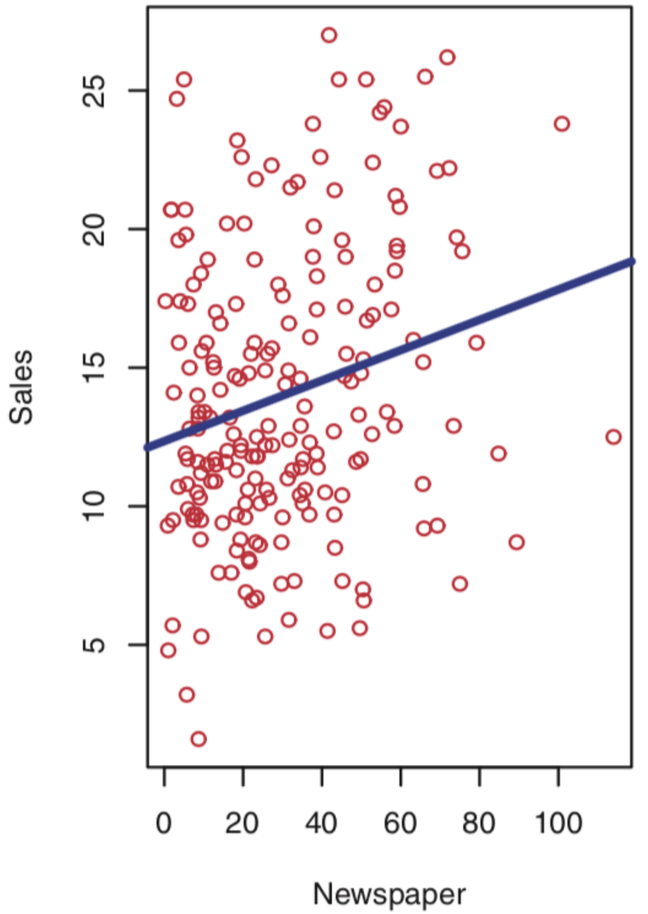
How can we make predictions?



Prediction

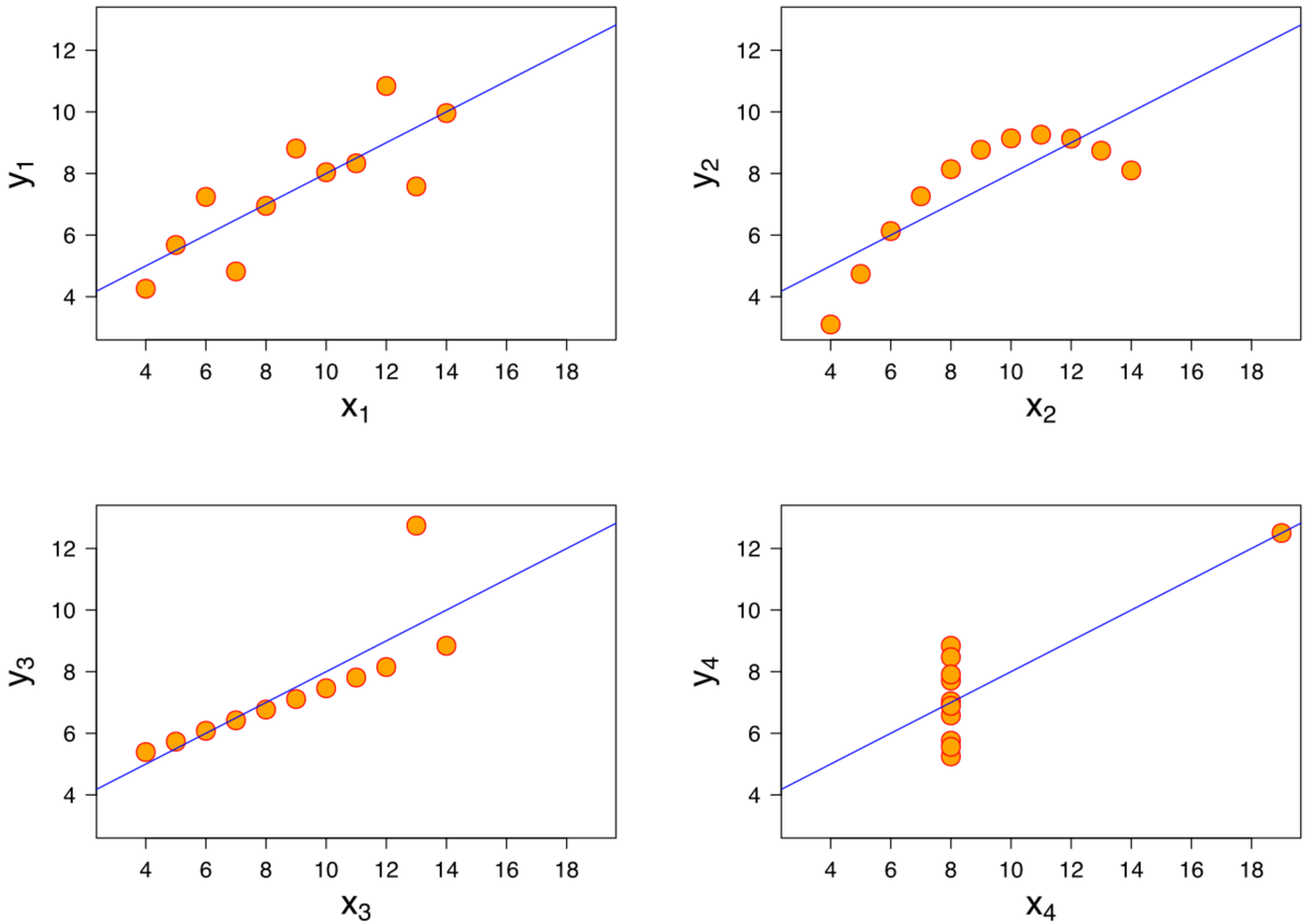






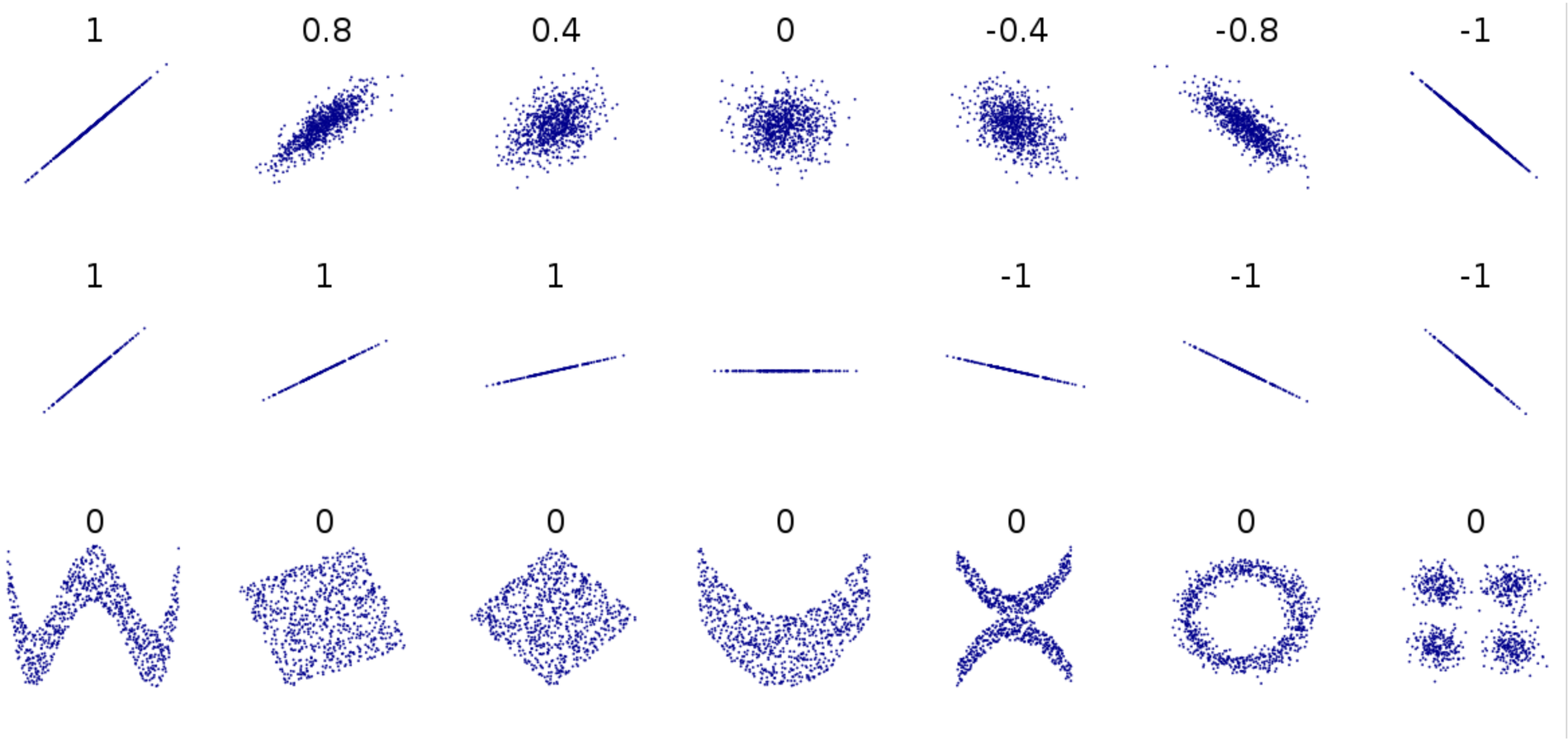
Source: An Introduction to Statistical Learning with Applications in R

Prediction - Anscombe's Quartet

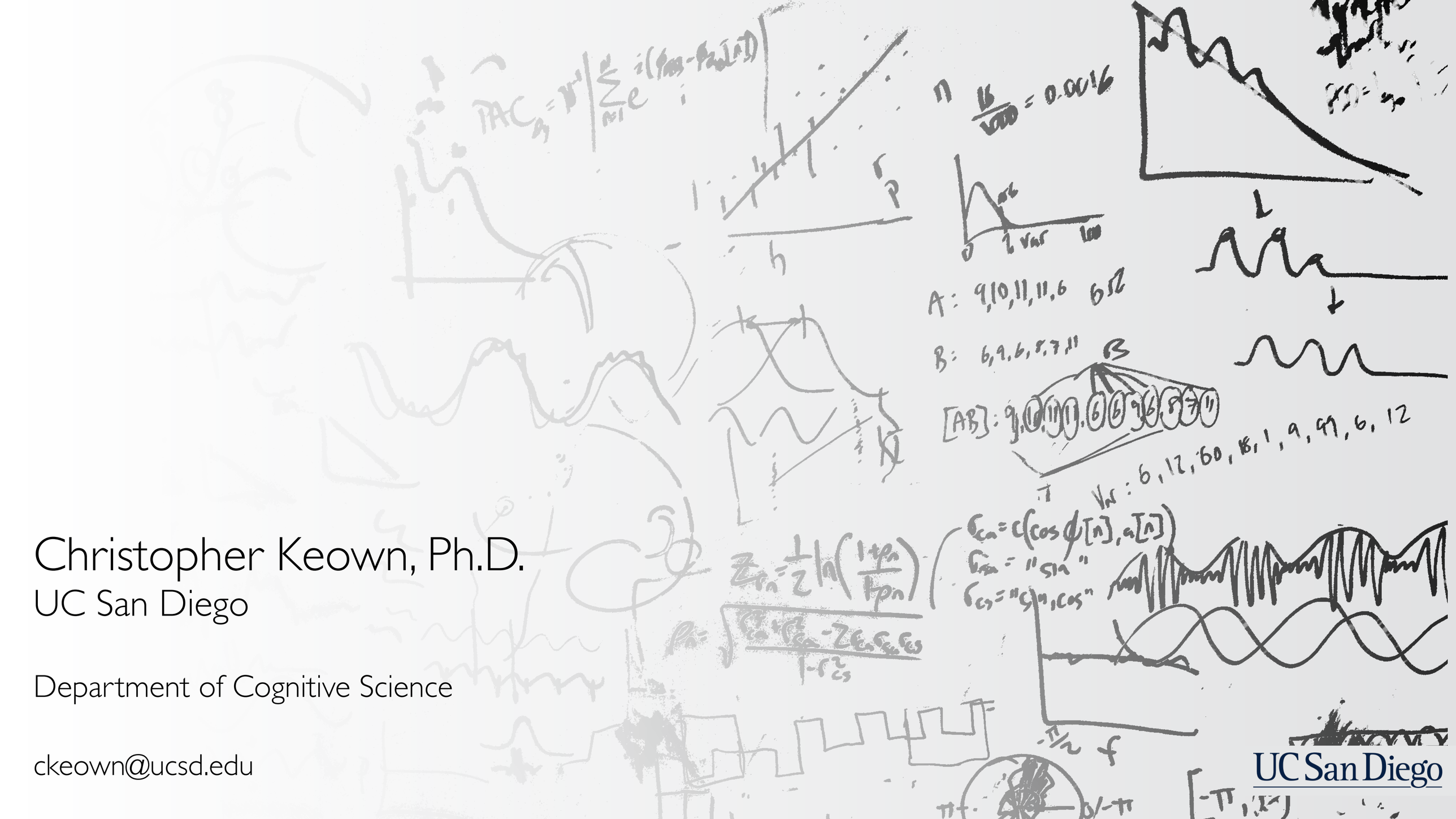


Source: Wikipedia

Prediction



Source: Wikipedia



Christopher Keown, Ph.D.

UC San Diego

Department of Cognitive Science

ckeown@ucsd.edu

UC San Diego