Introduction to CS109B a.k.a. STAT121B, AC209B, CSCI E-109B

# **Data Science 2: Advanced Topics in Data Science**

Pavlos Protopapas, Mark Glickman, and Chris Tanner



## INSTITUTE FOR APPLIED COMPUTATIONAL SCIENCE

AT HARVARD UNIVERSITY



- From 109A to 109B
- Who are we?
- A walk through the syllabus (i.e., the rules of the road!)

Second half of today's lecture starts the new course material!



- 1. Basic statistics, visualization
- 2. Regression with continuous outcome
- 3. Regression with a binary outcome
- 4. Ethics
- 5. PCA
- 6. Decision trees and ensemble methods
- 7. Basics of Neural Nets
- 8. Model interpretation



- 1. Smoothing, basis functions, generalized additive models
- 2. Unsupervised learning
- 3. Crash course in Bayesian statistics
- 4. Text Analysis
- 5. Neural Networks:
  - CNNs
  - RNNs
  - Autoencoders
  - Generative models
  - Reinforcement Learning 94, Protopapas, Glickman, Tanner



- 1. Segmentation Techniques
- 2. Echo State Networks
- 3. NLP
- 4. Inference in NN
- 5. Cycle GANS and other GANS techniques
- 6. Deep Reinforcement Learning



#### **Pavlos Protopapas**

Scientific Director of the Institute for Applied Computational Science (IACS). Teaches CS109(a/b), and AC295 (advanced practical data science).

**Research** in astrostatistics: machine learning, statistical learning, big data for astronomical problems.

He has picked some new hobbies besides 109 and **eating**:

Going to BSO (hopefully next fall), cross country ski (completed at Engadin skimarathon), cheese making and being a TikToker (check me out @pavlosprotopapas)



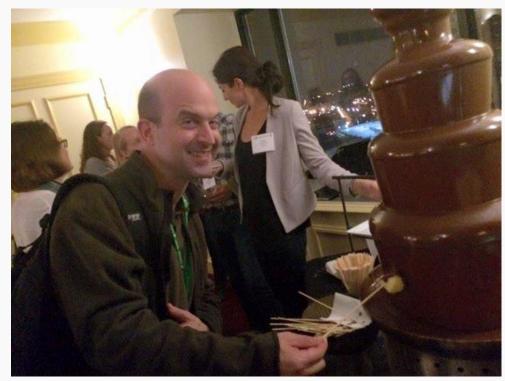


### Mark Glickman

Senior Lecturer in Statistics

Chess master, inventor of Glicko and Glicko-2 rating systems for head-to-head competition Director of the Harvard Sports Analytics Laboratory

Fellow of the American Statistical Association



Board of Directors member of the American Statistical Association (ASA).

Plays in a rock band called "Errors in Bars" with three other statisticians.



## **Chris Tanner**

Lecturer at IACS, teaching:

- CS109A/B
- AC297r (Master's Capstone)
- A new NLP course in the Fall

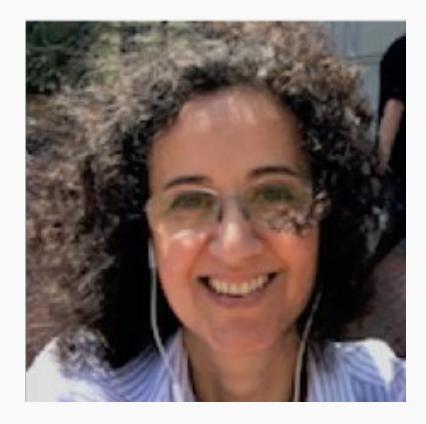
Research is within Natural Language Processing. Hobbies include ridiculous hikes, camping, woodworking, and photography.





## Eleni Angelaki Kaxiras

Eleni is assist. Director for Data Science and Computation at SEAS. For the first 3 years she was this and 109A course's Head TF, while last year she was a lab instructor. She is currently a doctoral student interested in the application of deep learning in analyzing biological signals. She owns olive trees in the island of Crete.





## Who? Head TF

#### **Chris Gumb**

Chris is currently working towards a graduate degree in Data Science from Harvard Extension School with a particular focus on NLP. His other interests and hobbies include: music theory, jazz improvisation, and film history.





Marios Matthaiakis Cecilia Garrafo Nabib Ahmed Mike Sedelmeyer Angela Garabet Ethan Cowan Patrick DeKelly Henry Jin **Tess McNulty** 

Kaela Nelson Rashmi Banthia Hayden Joy Shivas Jayaram Zhao Lyu **Javier Machin** Alex Spiride Andrew Smith



## Magical Mystery Tour of the 109B syllabus









The process to get help is:

- 1. Post the question in Ed and hopefully your peers will answer. We monitor the posts and we will respond within 8 hours from the posting time.
- 2. Go to Office Hours, this is the best way to get help.
- 3. For private matters send an email to the Helpline: <u>cs109b2021@gmail.com</u>. The Helpline is monitored by all the instructors and TFs.
- 4. For personal matters send an email to Pavlos, Mark and Chris.

#### Sundays will be slow days, so please be patient!

