Programming and p-values (pt2)

Michael Schatz

Sept 13, 2017 – Lecture 4

EN.601.452 Computational Biomedical Research

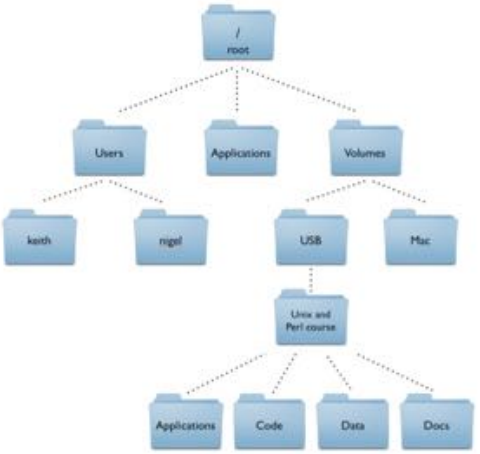
AS.020.415 Advanced Biomedical Research

What is a computer?

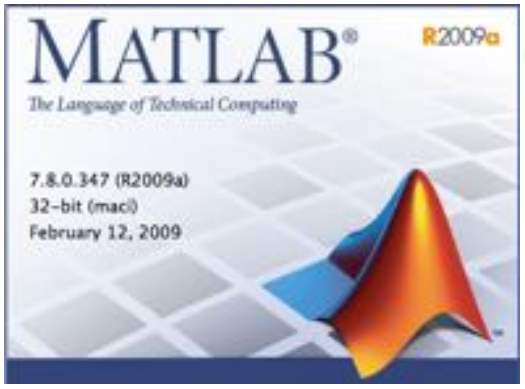
[software]

Office Applications
Presentations, Documents
Simple statistics and plots

Operating System
Mission Control
Windows, Mac, Unix, iOS



Files / Data
Papers, sequences,
measurements



Scientific Applications
Specialized Analysis
Commercial

Code / Scripts
Research Applications
Academic

What is a computer?

[hardware]

Hard Drive

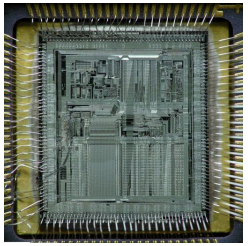
Permanent Storage – 1TB

(big, slow, cheap)

RAM

Working Storage – 8 GB

(small, fast, expensive)



Processor

Arithmetic, logic

cores, clock speed

Display

Human Interface

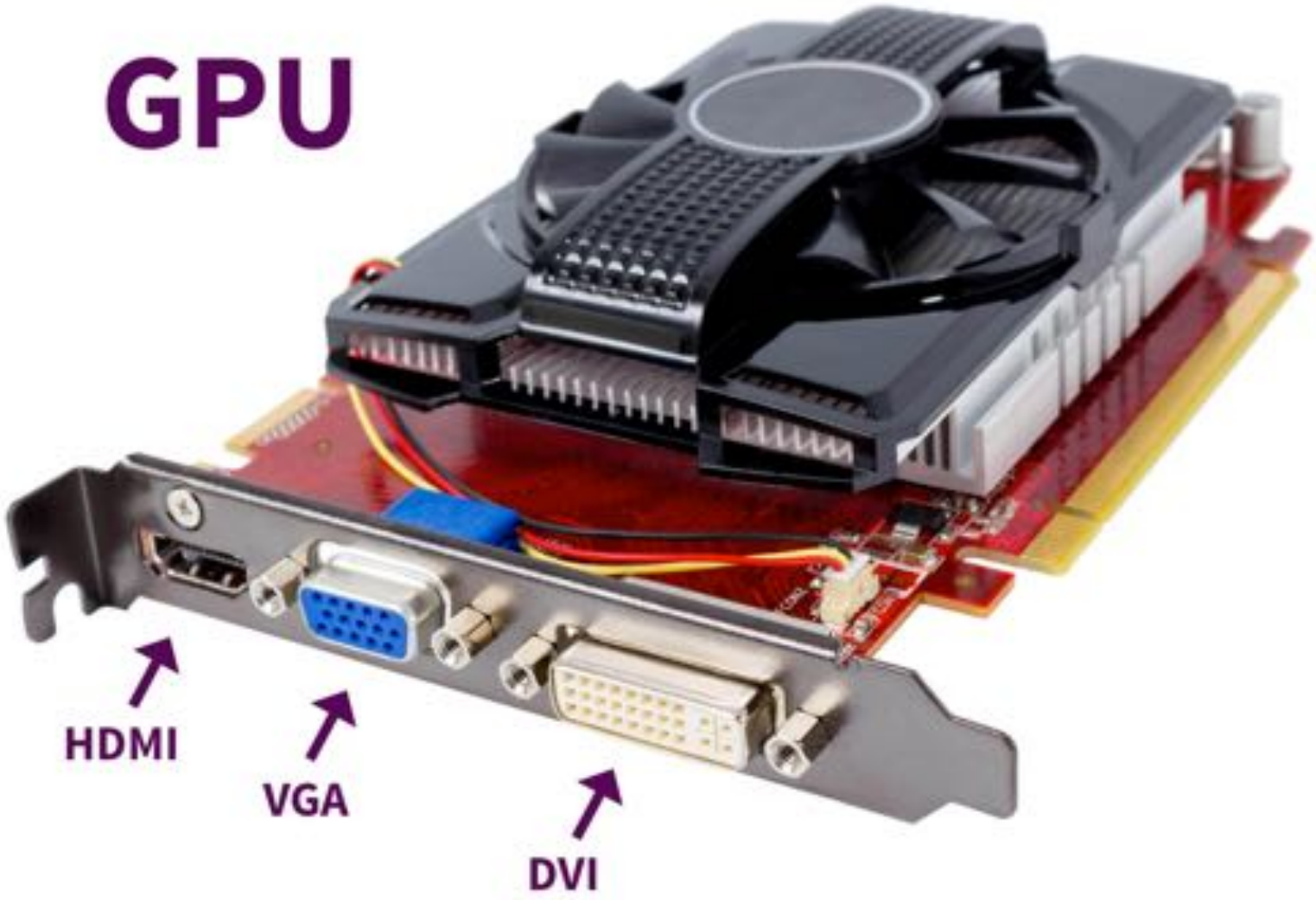
Network

Computer Interface

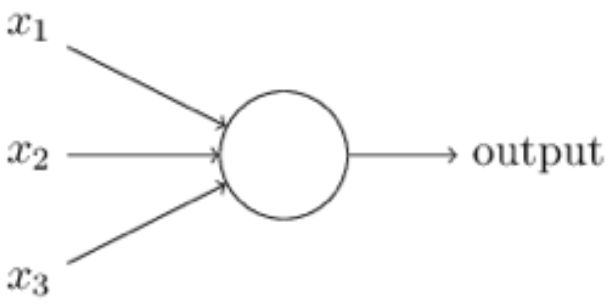
Home: 10Mb/s, JHU: 1Gb/s

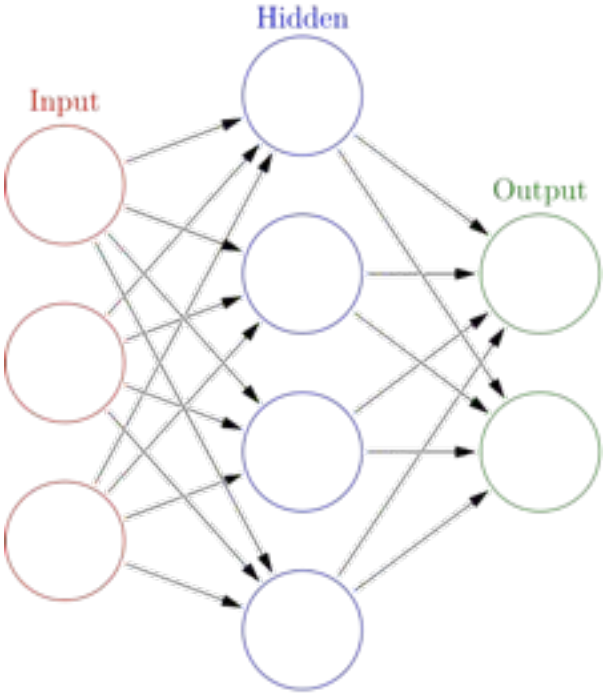
What is a computer?

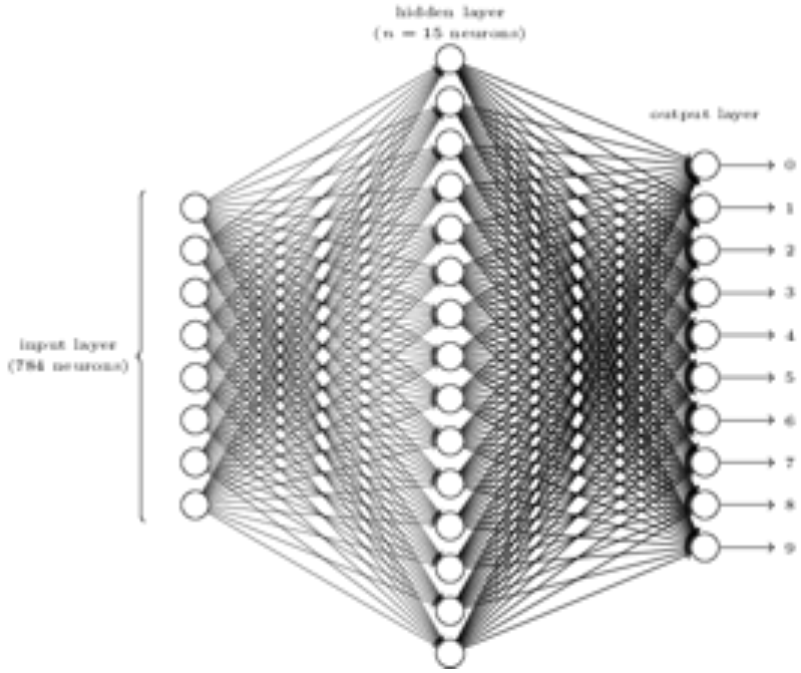
[hardware]



Artificial Neural Network







Deep Learning

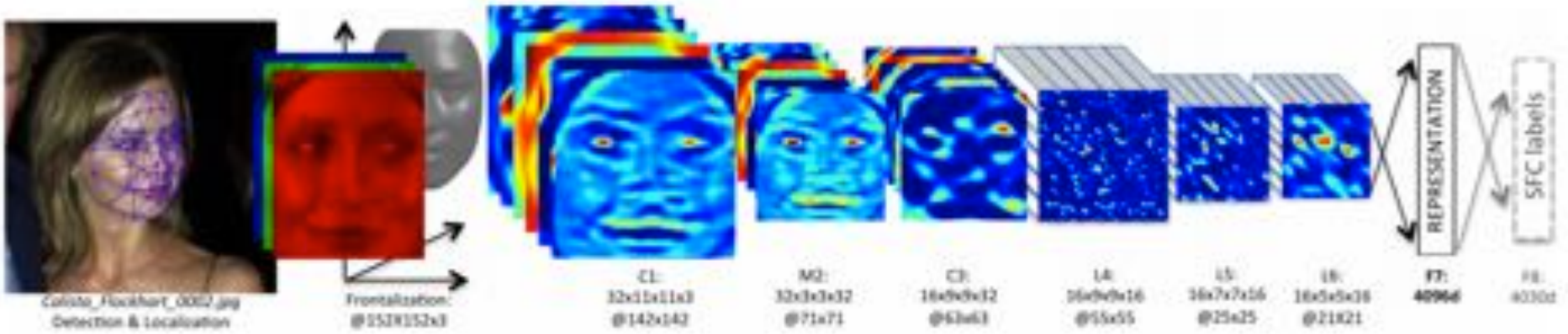
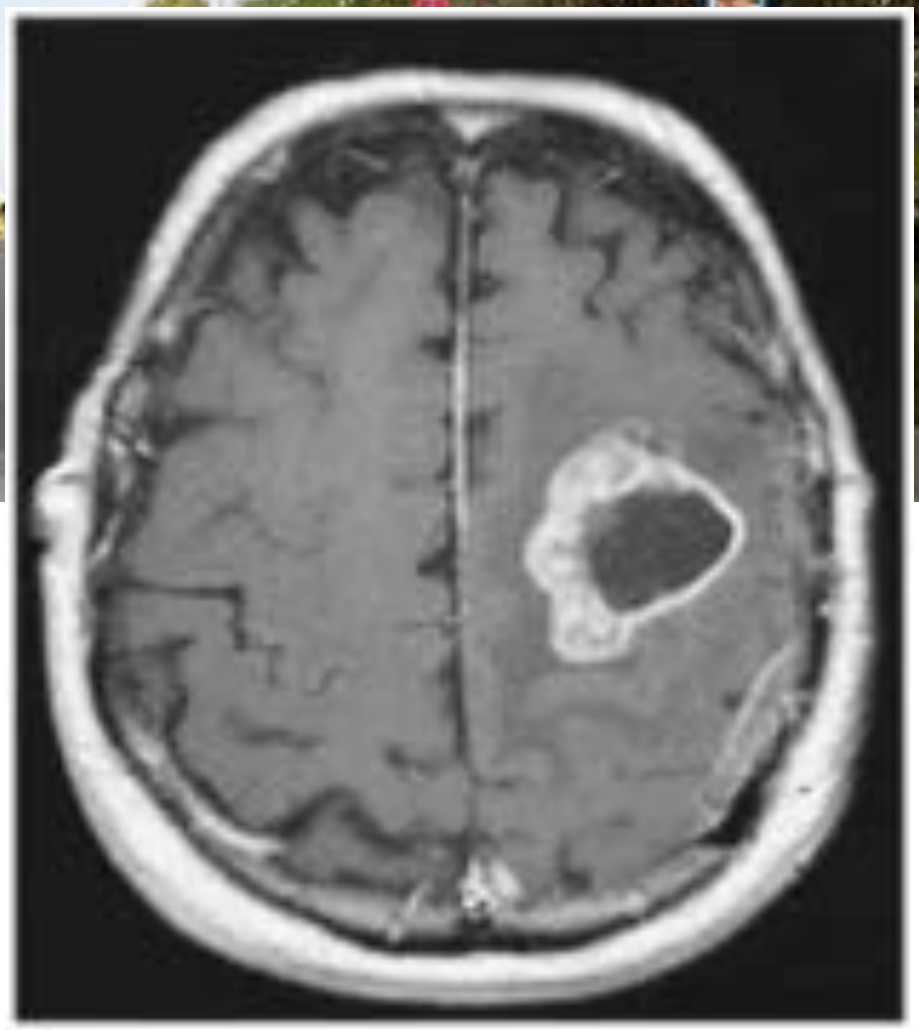


Figure 2. **Outline of the *DeepFace* architecture**. A front-end of a single convolution-pooling-convolution filtering on the rectified input, followed by three locally-connected layers and two fully-connected layers. Colors illustrate feature maps produced at each layer. The net includes more than 120 million parameters, where more than 95% come from the local and fully connected layers.

"… derive a face representation from a nine-layer deep neural network. This deep network involves more than 120 million parameters … Our method reaches an accuracy of 97.35% on the Labeled Faces in the Wild (LFW) dataset, reducing the error of the current state of the art by more than 27%, closely approaching human-level performance."

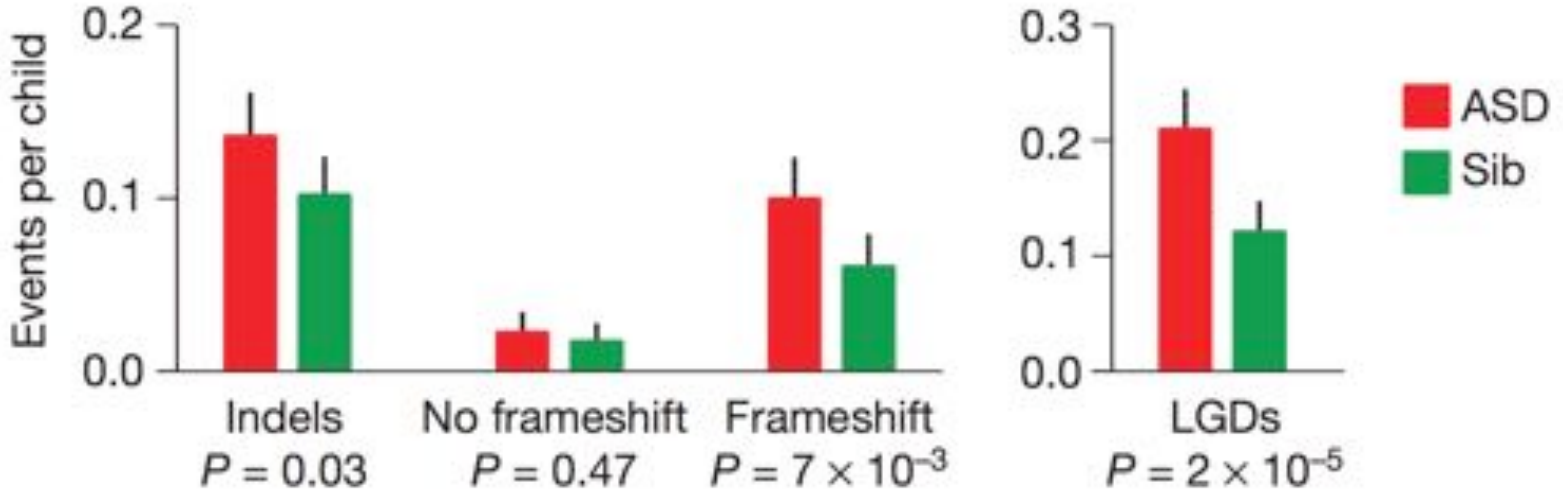
DeepFace: Closing the Gap to Human-Level Performance in Face Verification
Taigman et al (2014) Conference on Computer Vision and Pattern Recognition (CVPR)

Deep Learning Applications



P-values pt 2

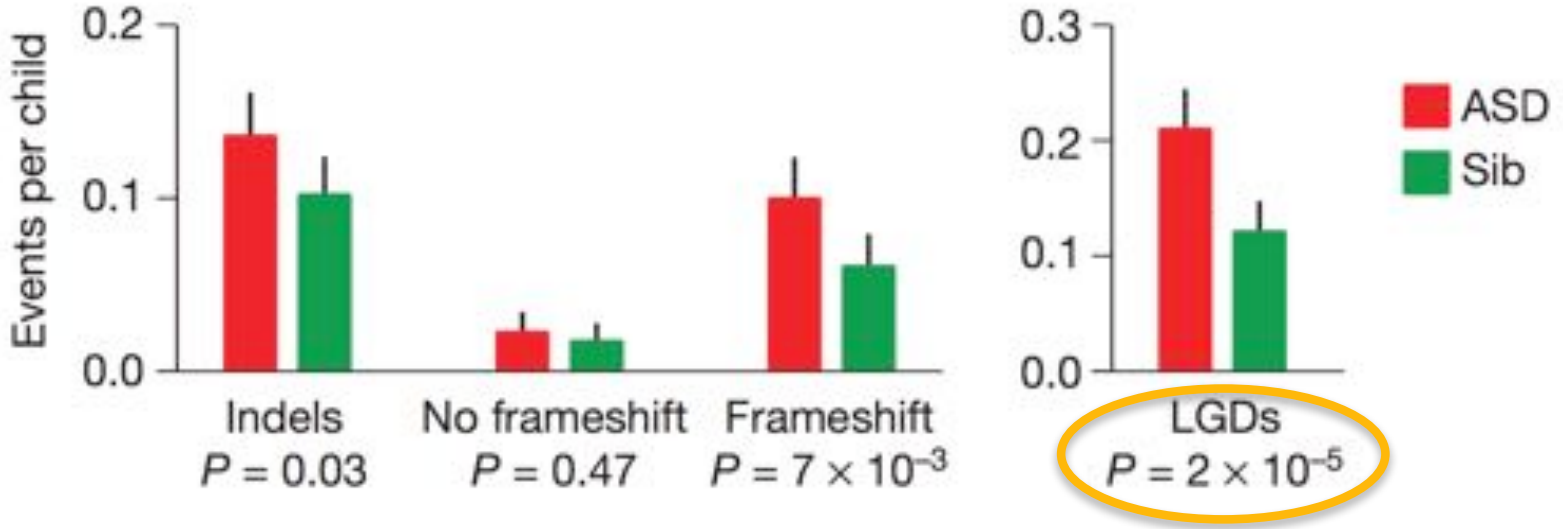
De novo Genetics of Autism



- In 2,500 family quads we see significant enrichment in de novo ***likely gene disruptions (LGDs)*** in the autistic kids
  - Overall rate basically 1:1
  - 2:1 enrichment in frameshift indels
  - Contributed dozens of new autism candidate genes, many associated with neuron development or chromatin formation

The burden of de novo coding mutations in autism spectrum disorders.
Iossifov *et al* (2014) *Nature*. doi:10.1038/nature13908

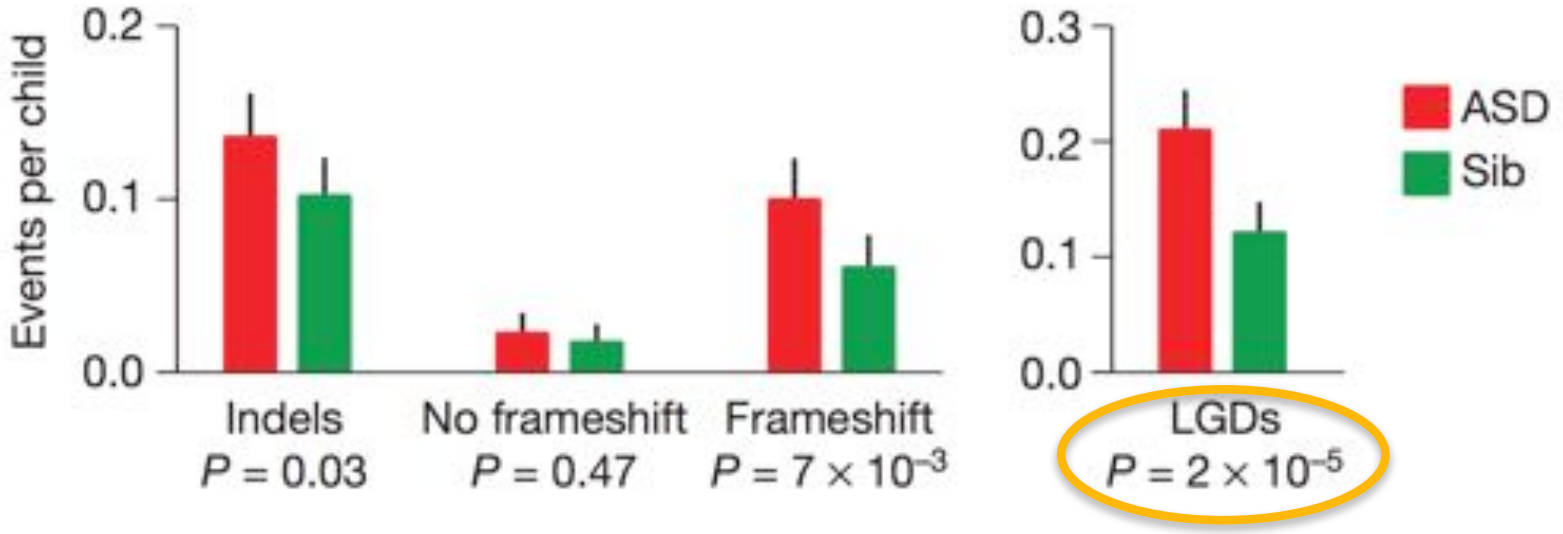
De novo Genetics of Autism



- In 2,500 family quads we see significant enrichment in de novo ***likely gene disruptions (LGDs)*** in the autistic kids
  - Overall rate basically 1:1
  - 2:1 enrichment in frameshift indels
  - Contributed dozens of new autism candidate genes, many associated with neuron development or chromatin formation

The burden of de novo coding mutations in autism spectrum disorders.
Iossifov *et al* (2014) *Nature*. doi:10.1038/nature13908

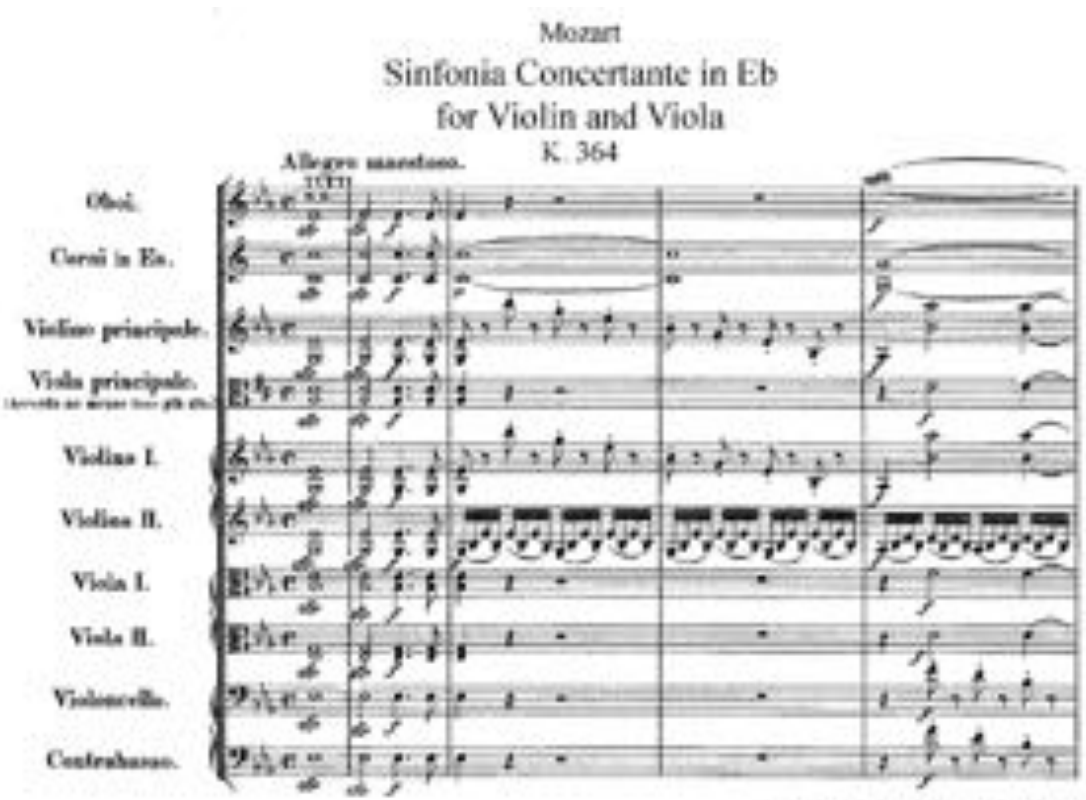
P-value



- The "p-value" is the probability of observing a difference with the same or larger magnitude as observed but completely by chance (under the null hypothesis)
  - Maybe kids with ASD genuinely have a larger number of gene disrupting mutations, or maybe we just got a slightly skewed sample?

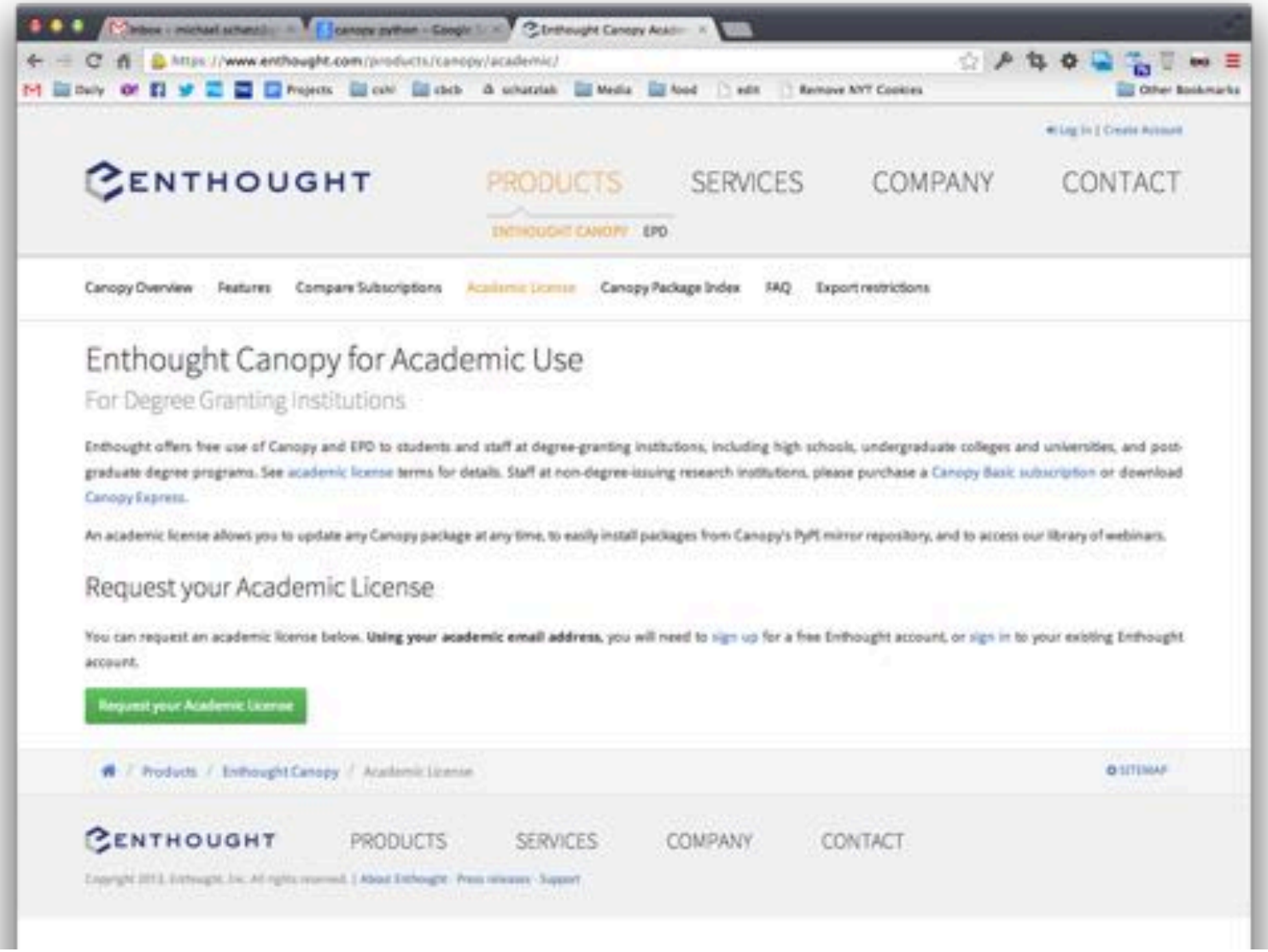
- If I flip a coin 100 times, I expect 50 heads and 50 tails, but Im not surprised if I get 49 heads and 51 tails.
  - On the other hand Im extremely surprised if I get 1 heads and 99 tails!
  - What about 25 heads? or 15 heads? Or 5?
  - Im more surprised when the probability is smaller and smaller (exponent is more and more negative)

Programming 101

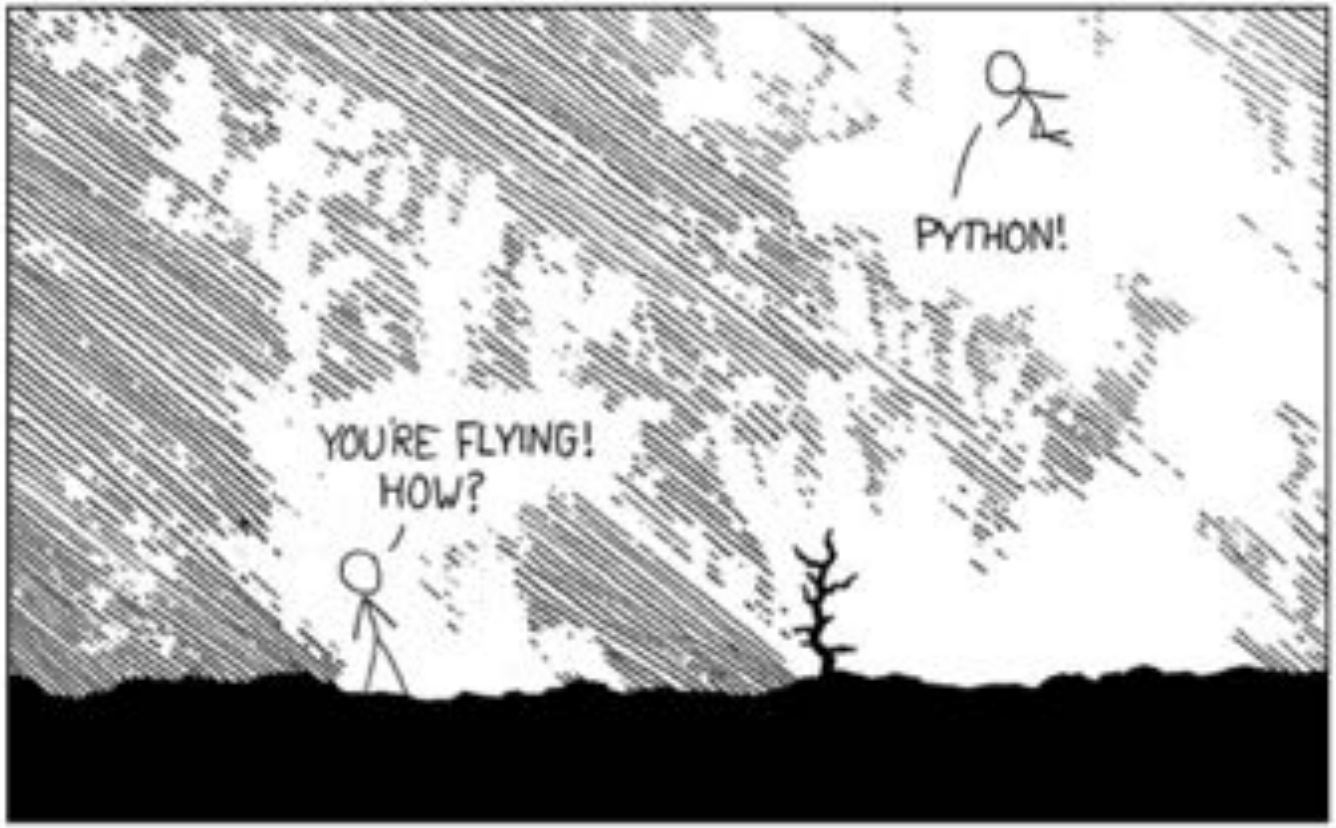


A software program is like sheet music for the orchestra inside your computer
Static, written representations of an active process

Programming with Python

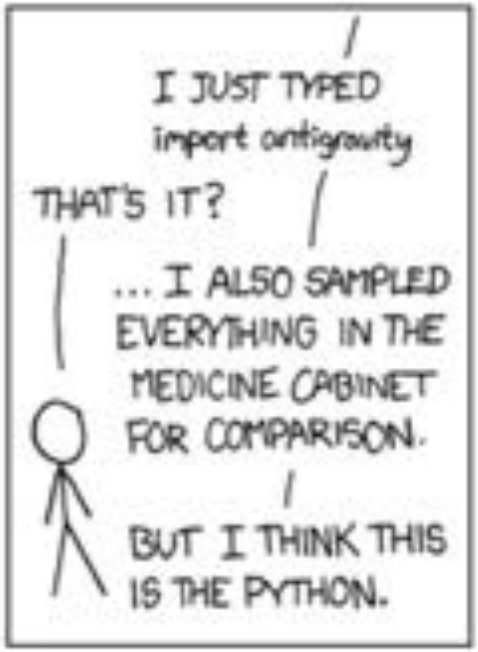


https://www.enthought.com/products/canopy/academic/
http://www.codecademy.com/tracks/python


PYTHON!
YOU'RE FLYING! HOW?


I LEARNED IT LAST NIGHT! EVERYTHING IS SO SIMPLE!
HELLO WORLD IS JUST print "Hello, world!"


I DUNNO... DYNAMIC TYPING? WHITESPACE?
COME JOIN US! PROGRAMMING IS FUN AGAIN! IT'S A WHOLE NEW WORLD UP HERE!
BUT HOW ARE YOU FLYING?


I JUST TYPED import antigravity
THAT'S IT?
... I ALSO SAMPLED EVERYTHING IN THE MEDICINE CABINET FOR COMPARISON.
BUT I THINK THIS IS THE PYTHON.