

# Design in the Deep End

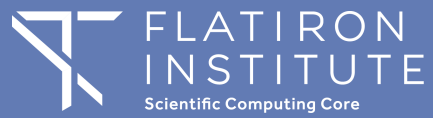
Liz Lovero · [elizabethlovero.com](http://elizabethlovero.com)



# Hi I'm Liz. 🖐️


Liz Lovero  
Data Visualization Engineer  
Scientific Computing Core  
Flatiron Institute  
Simons Foundation  
New York City 🗽

SIMONS FOUNDATION



## Our Centers

---




### Center for Computational Astrophysics

Astronomical Data  
Compact Objects  
Cosmology X Data Science  
Galaxy Formation  
Gravitational Wave  
Astronomy  
Planet Formation




### Center for Computational Biology

Biophysical Modeling  
Genomics  
Neuroscience  
Systems Biology



### Center for Computational Mathematics

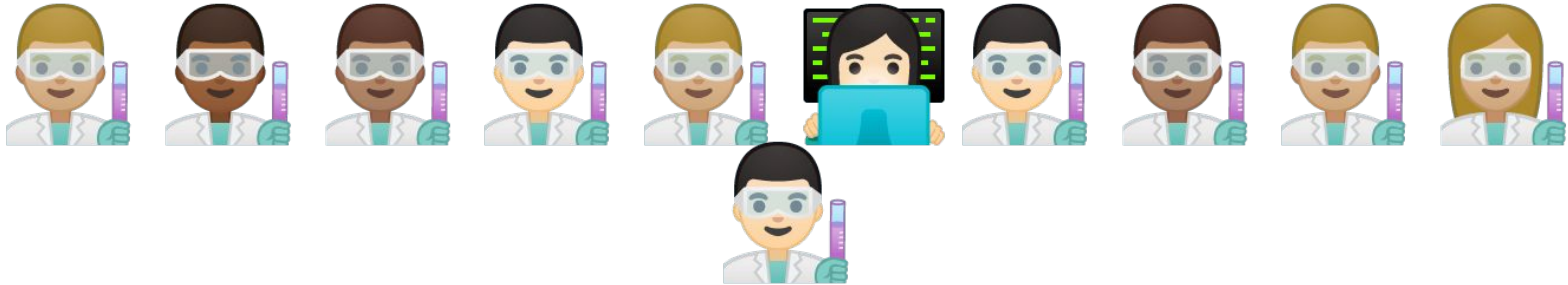
Image and Signal Processing  
Machine Learning and Data  
Analysis  
Numerical Analysis



### Center for Computational Quantum Physics

Dynamics and Control  
Quantum Materials  
Software Libraries  
Theory and Methods

How do you approach data visualization and design when you are not a subject matter expert?



# SpikeForest

Case Study

## Our Centers

---

### Center for Computational Astrophysics

Astronomical Data  
Compact Objects  
Cosmology X Data Science  
Galaxy Formation  
Gravitational Wave  
Astronomy  
Planet Formation

### Center for Computational Biology

Biophysical Modeling  
Genomics  
Neuroscience  
Systems Biology

### Center for Computational Mathematics

Image and Signal Processing  
Machine Learning and Data  
Analysis  
Numerical Analysis

### Center for Computational Quantum Physics

Dynamics and Control  
Quantum Materials  
Software Libraries  
Theory and Methods

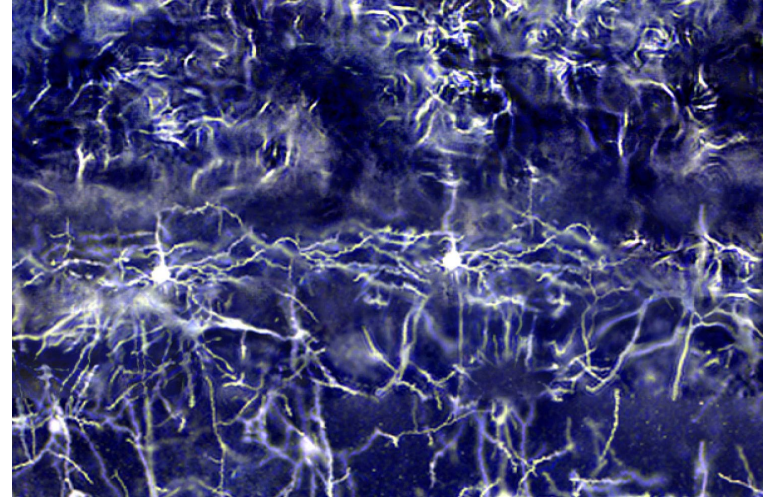
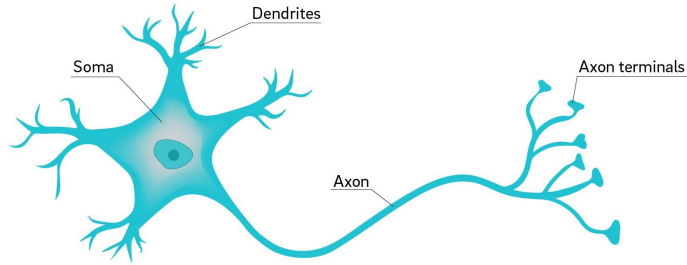


# SpikeForest

Website and open source computing framework for evaluating and comparing spike sorting algorithms for neurophysiology data analysis.

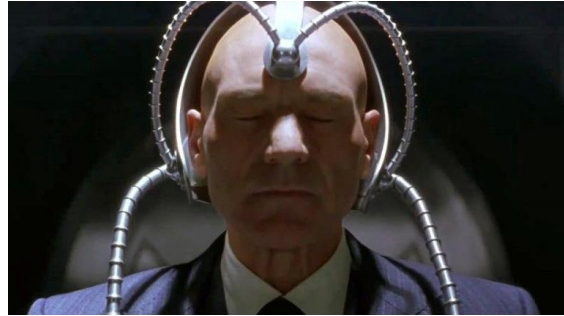
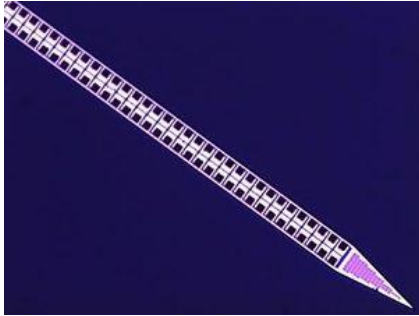
Public website launching June 2019

# Neuron



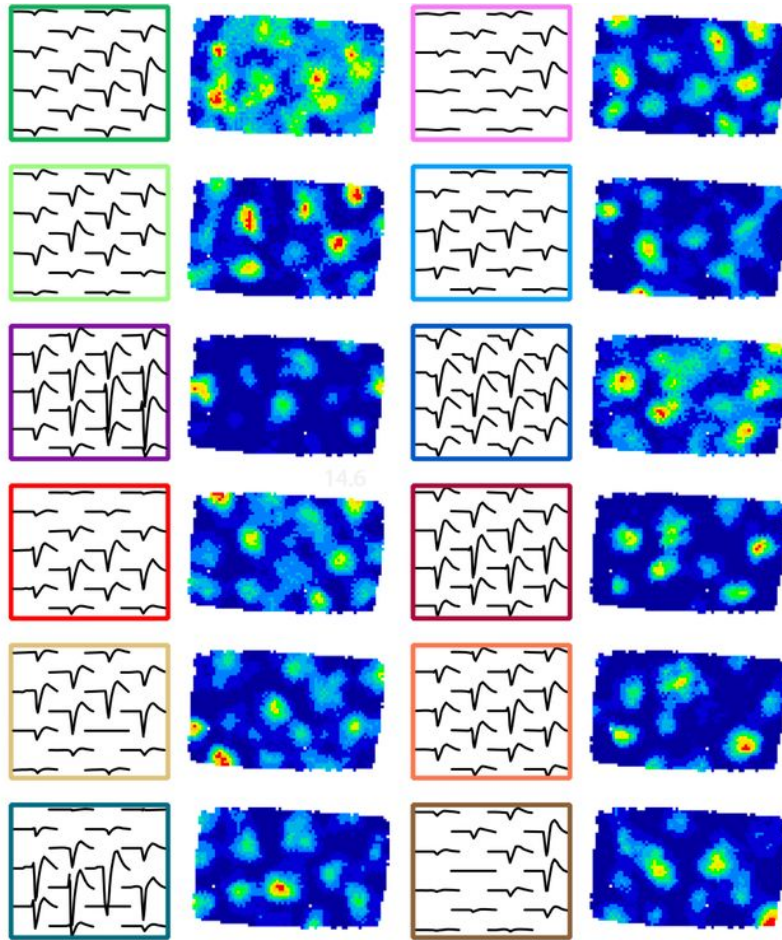
Neuron is a cell that carries electrical impulses. Neurons transmit information via these impulses to form neural pathways and circuits.

# Neurophysiology



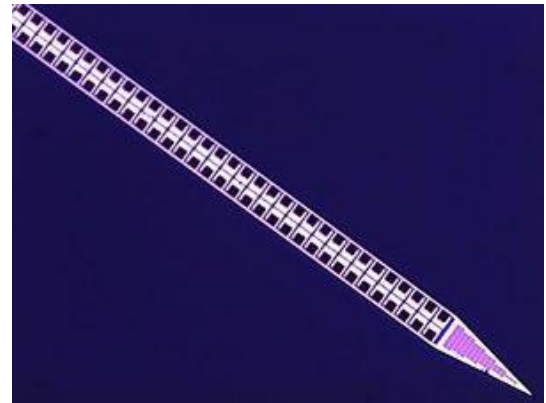
Neuroscientists record the flow of electricity, in this case in a mouse brain, in order to discern the location and behavior of individual neurons.

An early generation of the Neuropixels probe. Credit: Imec



Left: Activity patterns for 12 different cells recorded simultaneously using Neuropixels probe.

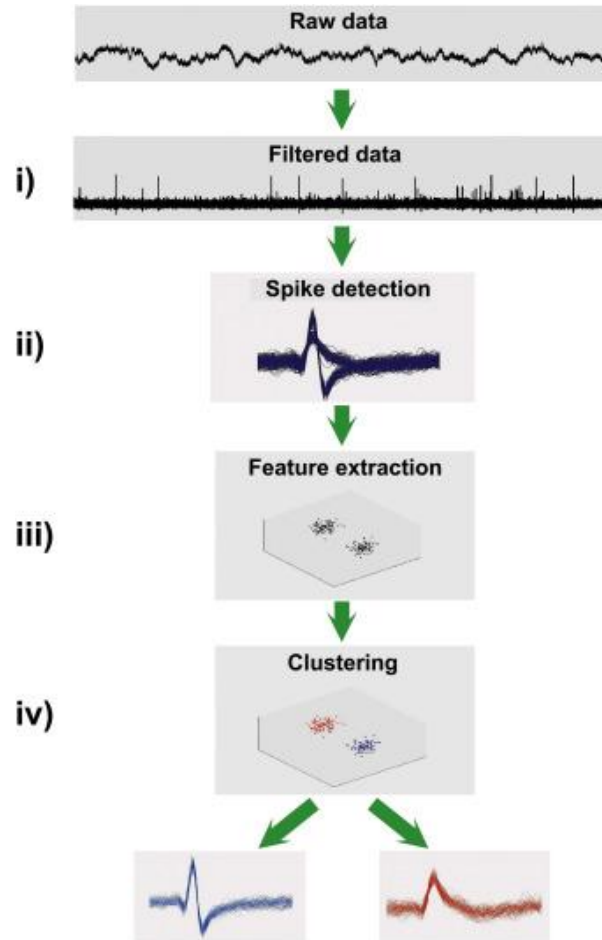
Below: Neuropixels probe



Credit: Marius Bauza

# Spike sorters

Algorithms that parse neuron signals from background noise



# Lots of Algorithms

No clear consensus on the best

UltraMegaSort 2000

JRCLUST

KiloSort

mountainsort



MClust

IronClust

Herding Spikes 2

tridesclous

YASS: Yet Another Spike Sorter





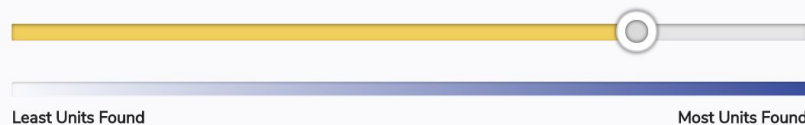
# The Initial Request

A simple heatmap to compare  
spike sorters

# Spike Sorting Results Overview

Number of groundtruth units above accuracy threshold

Minimum accuracy: 0.8



	IronClust-drift	IronClust-tetrode	KiloSort	MountainSort4-thr3	SpykingCircus
bionet_drift_32c	386		441	355	145
bionet_drift_8c	92		85	95	54
bionet_shuffle_32c	371		438	356	152
bionet_shuffle_8c	94		88	90	55
bionet_static_32c	463		499	429	339
bionet_static_8c	105		103	112	73
magland_synth_noise10_K10_C4		100	45	89	60

# Stakeholders !== Users

Get input from both sides

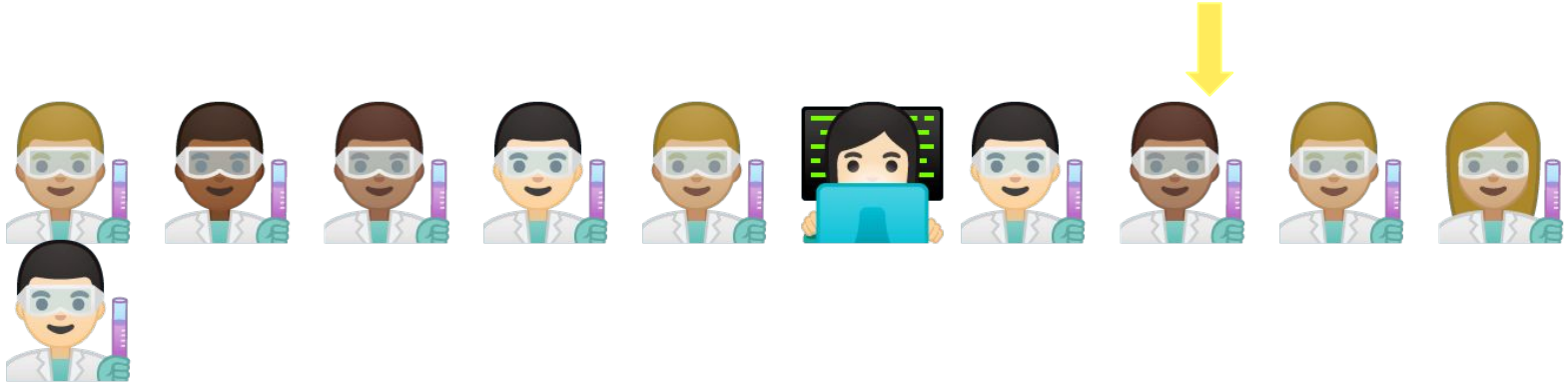
## *My stakeholders*

Mid-career spike-sorting algorithm developers

## *My users*

Early career neuroscientists with mouse labs

# Use outsider status to your advantage



# Provide accessible feedback mechanisms



Use a tool to mark an area



About

Q New query

# BRCA1



BRCA1, DNA repair associated

Tissue gene network

Tissue expression

Biological process gene network

Biological process

Disease

## Tissue-specific interactions

Functional networks that capture tissue-specific interactions from large data compendia

Enter a comment...

Genes

Tissue

Data types

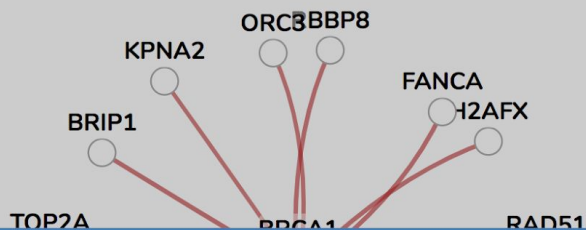
BRCA1

testis

All selected (5)



testis - A typically paired male reproductive gland that produces sperm and that in most mammals is contained within the scrotum at sexual maturity.



Genes

Process

Tissue

Detail

## Functionally related genes

Top connected genes to the query

Search

SYMBOL

NAME

- BARD1 BRCA1 associated RING domain protein
- BRIP1 BRCA1 interacting protein C-terminal domain helicase 1
- MSH2 mutS homolog 2



COMMENT



PEN



ARROW



BLACKOUT



Your e-mail

Your message

powered by Usersnap

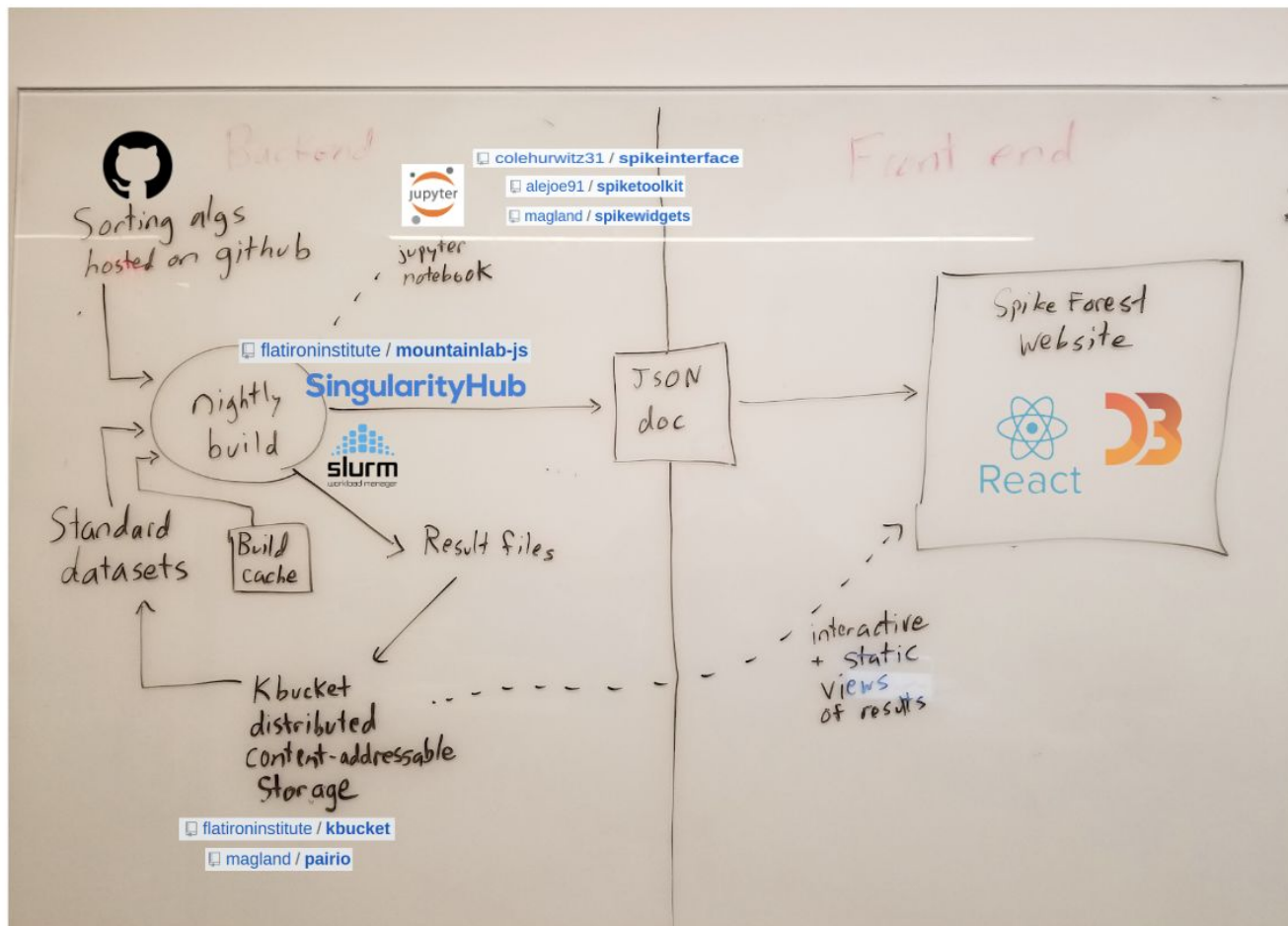
SEND

Usersnap  
usersnap.com

# Everything Is Data

Establish a specific vocabulary





# “One Single JSON”

```
▶ recordings (pin): [{...}, {...}, {...}, {...}, ...]  
  selectedRecording (pin): null  
  selectedStudy (pin): null  
▶ sorters (pin): [{...}, {...}, {...}, {...}, ...]  
▶ studies (pin): [{...}, {...}, {...}, {...}, ...]  
▶ units (pin): [{...}, {...}, {...}, {...}, ...]
```

# “One Single JSON”

```
▼ recordings (pin)
  ▼ 0 (pin)
    name (pin): "001_synth"
    study (pin): "magland_synth_noise10_K10_C4"
    directory (pin): "kbucket://15734439d8cf/groundtruth/magland_synth/datasets_noise10_K10_C4/001_synth"
    description (pin): "One of the recordings in the magland_synth_noise10_K10_C4 study"
```

# “One Single JSON”

- 5 Study Sets each containing 3-9 Studies each.
- 25 Studies total.
- Each Study has 10-12 Recordings.
- Each Recording has 10-50 True Units.
- Each True Unit is sorted by a minimum of 4 Algorithms to create at least 4 sets of Sorting Results
- Algorithms are also called Sorters.

When someone says

“Did you get my data?”

**GOBBLEDYGOOK...**



# Data Drill Down Questions

1. *How is this data generated? (Survey, experiment, synthetic)*
2. *What are the units? Are they consistent?*
3. *Are these terms of art or terms you've coined for this research?*
4. *What's the hierarchy or relationship between data types?*
5. *What is the data source? (CSV, API, public, private)*
6. *Is data security an issue? (PII, financial data)*



# Get It In Writing

Jean Dubuffet, 1966  
*Verbal Agreement: Six Figures*  
Felt-tip pen on paper  
Museum of Modern Art, NYC



# Label, Label, Label

Put your vocabulary to use

# Studies

A study is a collection of recordings. Sorting results may be aggregated over a study.

STUDY NAME ▲	NUMBER OF RECORDINGS	TOTAL FILE SIZE	CHANNELS	DURATION (IN SECONDS)	EXPERIMENT TYPE (SYNTHETIC / IN VIVO / IN VITRO)	PROBE TYPE	BRAIN REGION	GROUNDTRUTH UNITS	DESCRIPTION
--------------	----------------------	-----------------	----------	-----------------------	--	------------	--------------	-------------------	-------------

# Expose your formulas

Not all users are anglophones, scientists or literate in the same kinds of statistics.

Including formulas used, especially during early revisions, avoids simple confusions.

Sorter: MountainSort4-thr3

Study: bionet\_static\_32c

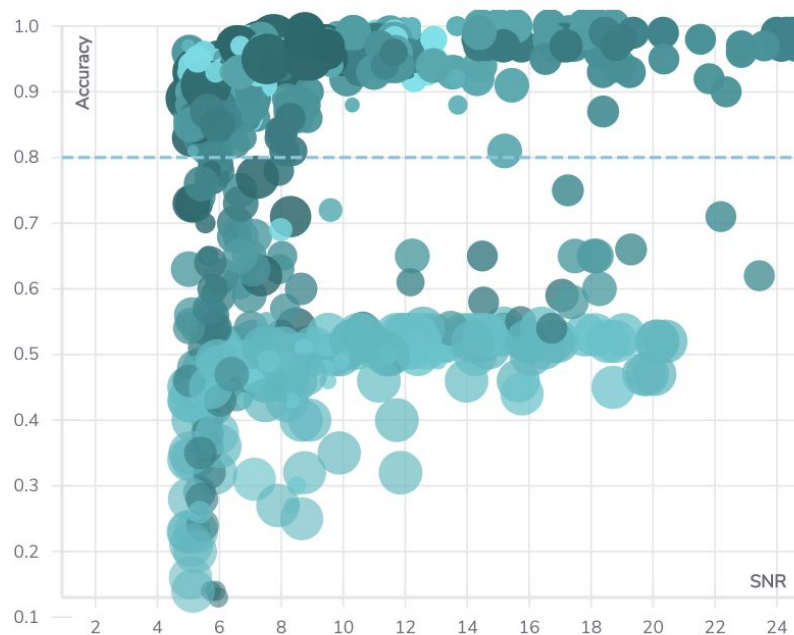
Scatterplot

Study

Sorter

Other

Items above accuracy threshold: 429



$$\text{SNR} = \frac{P_{\text{signal}}}{P_{\text{noise}}},$$

# Label for Accessibility

## **Web Accessibility Guidelines**

<https://www.w3.org/WAI/>

## **Web Accessibility Google + Udacity Course**

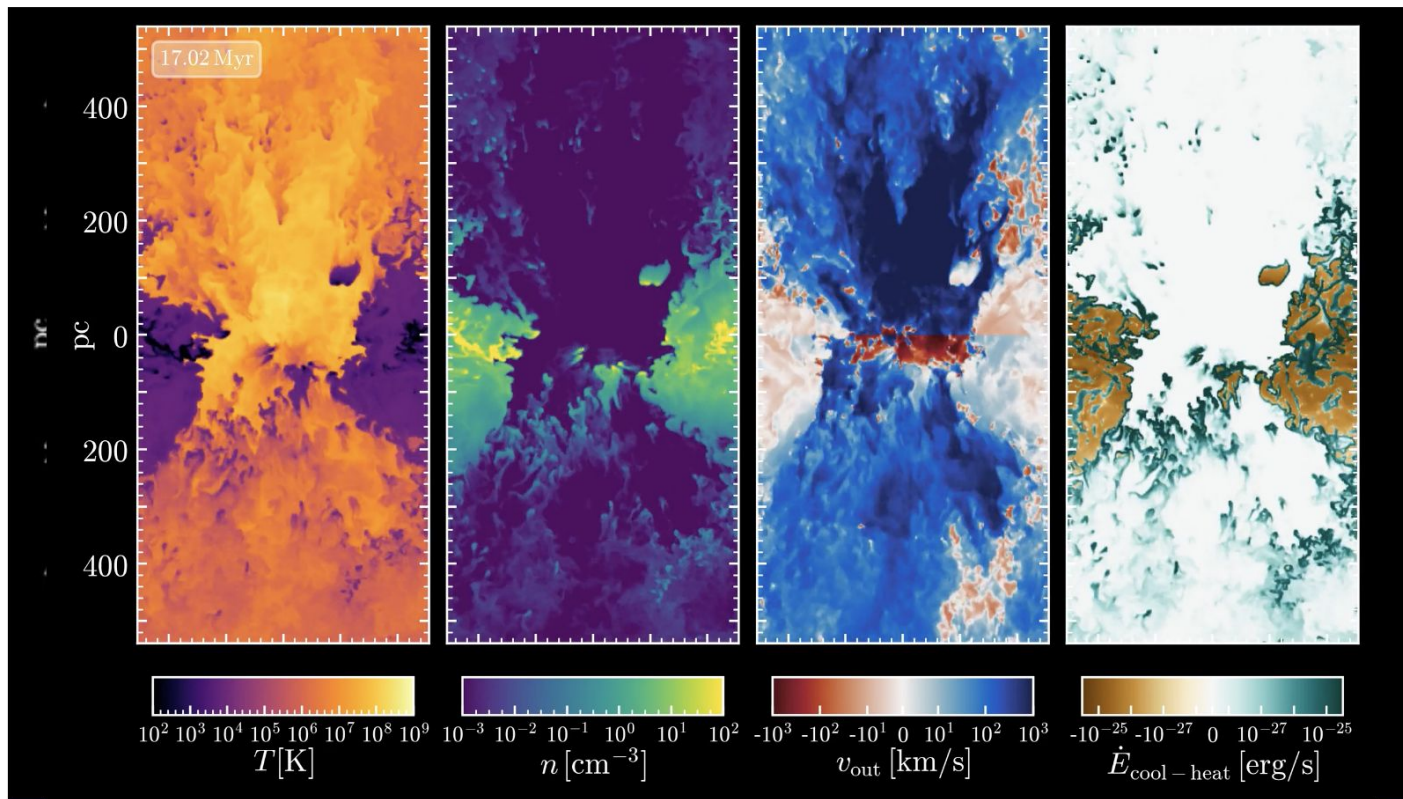
<https://www.udacity.com/course/web-accessibility--ud891>

## **React / ESLint for Accessibility (react-a11y)**

<https://github.com/evcohen/eslint-plugin-jsx-a11y>

# Significant Units

When in doubt, be explicit



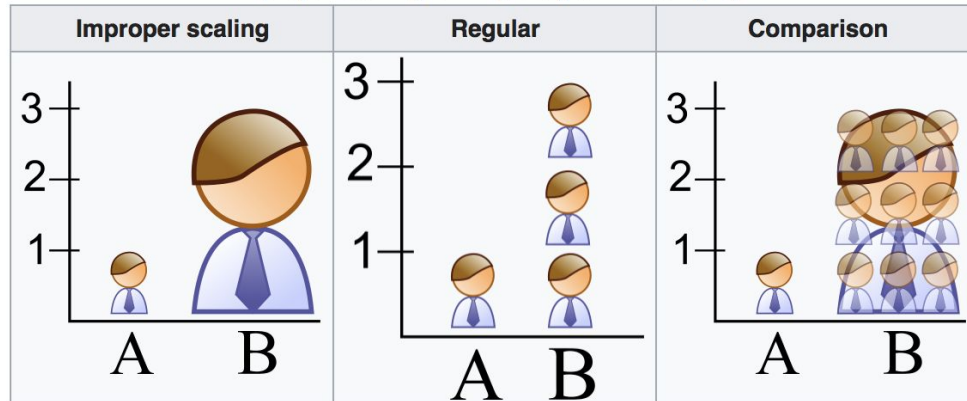
## Drummond Fielding

<https://www.simonsfoundation.org/flatiron/center-for-computational-astrophysics/galaxy-formation/visualizations/>

# Set Aside the Design Eye

Pay particular attention to scale, ratio, and visual organization

Improper scaling of 2D pictogram in bar graph



In the improperly scaled pictogram bar graph, the image for B is actually 9 times as large as A.



# You Must Learn



Read the papers. Look up the terms.

# Skill up

College 101 textbooks are great resources.

Read the summary of your colleagues papers.

Look up the terms you don't know.  
(Wikipedia is surprisingly helpful.)

# Learn Acronyms

API

Application Programming Interface

API

Annual Percent Interest





About

Q New query

# BRCA1



BRCA1, DNA repair associated ↗

Tissue gene network

Tissue expression

Biological process gene network

Biological process

Disease

## Tissue-specific interactions

Functional networks that capture tissue-specific interactions from large data compendia

Genes

Tissue

Data types

BRCA1

testis ▾

All selected (5) ▾



+ Add network

testis - A typically paired male reproductive gland that produces sperm and that in most mammals is contained within the scrotum at sexual maturity.



Genes

Process

Tissue

Detail

## Functionally related genes

Top connected genes to the query



Search



SYMBOL

NAME

SCORE ?



BARD1

BRCA1 associated RING domain 1

0.99



BRIP1

BRCA1 interacting protein C-terminal helicase 1

0.99



MSH2

mutS homolog 2

0.99



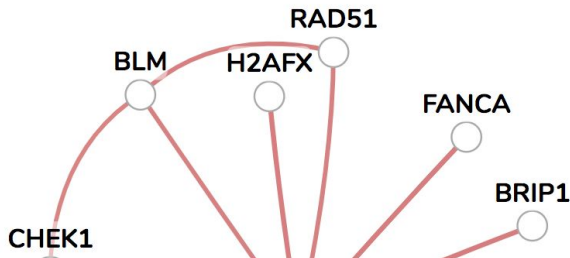
H2AFX

H2A histone family member X

0.98



CONTACT US



# Useful rabbit holes



The BRCA1 gene is located on chromosome 17q21 -- the long (q) arm of chromosome 17 at position 21.

The symbol BRCA stands for BReast CAncer.

# Reinvent No Wheels

Trust your own expertise

# KBucket

 flatironinstitute / **kbucket** 



The tool already exists and the name is already taken .



# Animate with Meaning

Only loading icons may wiggle





# Show & Tell



[github.com/flatironinstitute](https://github.com/flatironinstitute)  
[simonsfoundation.org/flatiron/software](https://simonsfoundation.org/flatiron/software)



Search or jump to...



[Pull requests](#)

[Issues](#)

[Marketplace](#)

[Explore](#)



# Flatiron Institute

@simonsfoundation

New York City

<https://www.simonsfoundation.org/flatir...>

**Repositories** 36

**People** 55

**Teams** 8

**Projects** 0

Find a repository...

Type: All ▾

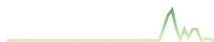
Language: All ▾

**New**

## spikeforest-docs

Miscellaneous documentation for projects related to spikeforest

Jupyter Notebook Apache-2.0 Updated 2 hours ago



## astrosims-reproto

A simple prototype for serving astronomical simulation catalog data

Jupyter Notebook 1 Apache-2.0 Updated 2 hours ago



## NoRMCorre

### Top languages

Python C++ Jupyter Notebook  
 JavaScript MATLAB

### Most used topics

[Manage](#)

[calcium-imaging](#)

[matlab](#)

[motion-correction](#)

[source-extraction](#)

# Common Cause

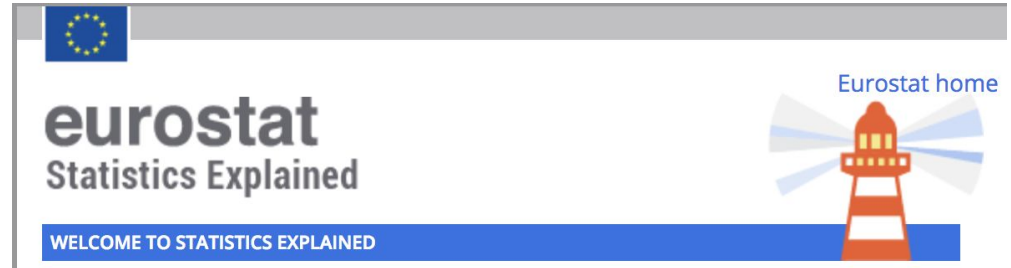


# Statistical Literacy

Not just for the status-conscious

*In an era of fake news, it is increasingly important for citizens to understand figures presented in newspapers, on the internet and elsewhere.*

- Louise Corselli-Nordblad and Britta Gauckler, Eurostat



<https://ec.europa.eu/eurostat/statistics-explained/index.php>

OECD (2017), "Education at a glance: Adult education and learning (Edition 2017)", *OECD Education Statistics* (database), <https://doi.org/10.1787/bdcfdf79-en> (accessed on 25 January 2019)

# Questions?



# Thanks!

