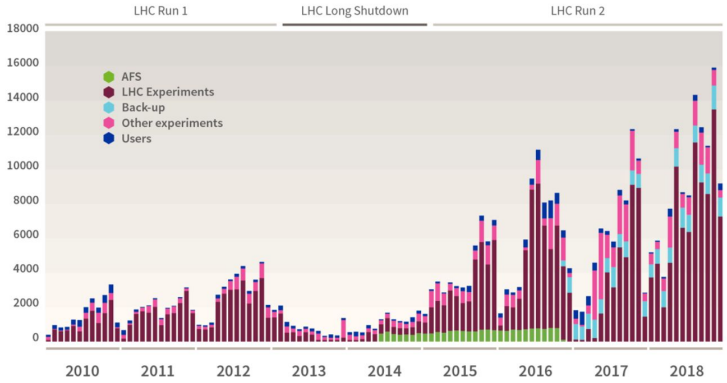
# Software Engineering for Scientists (SWE4S)

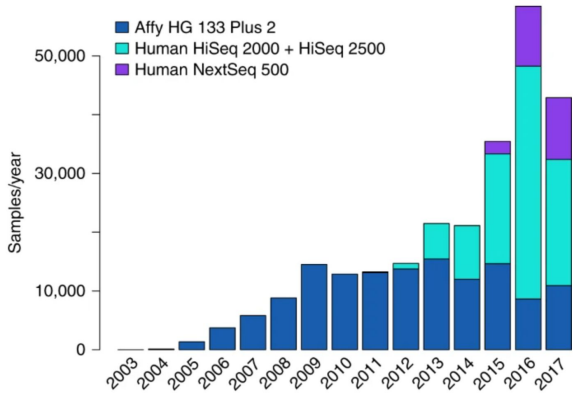
# Today’s plan/objectives

- Course intro/motivation
- Syllabus and logistics
- Howdies
- For next time…




Data (in terabytes) recorded on tape at CERN month-by-month. This plot shows the amount of data recorded on tape generated by the LHC experiments, other experiments, various back-ups and users. In 2018, over 115 PB of data in total (including about 88 PB of LHC data) were recorded on tape, with a record peak of 15.8 PB in November (Image: Esma Mobs/CERN)

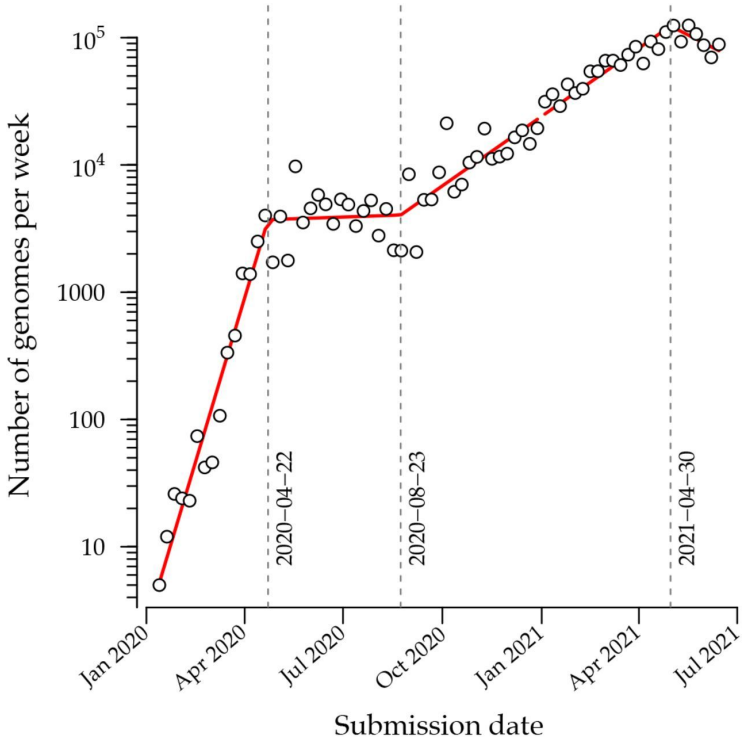
Ozcesmeci, 2019



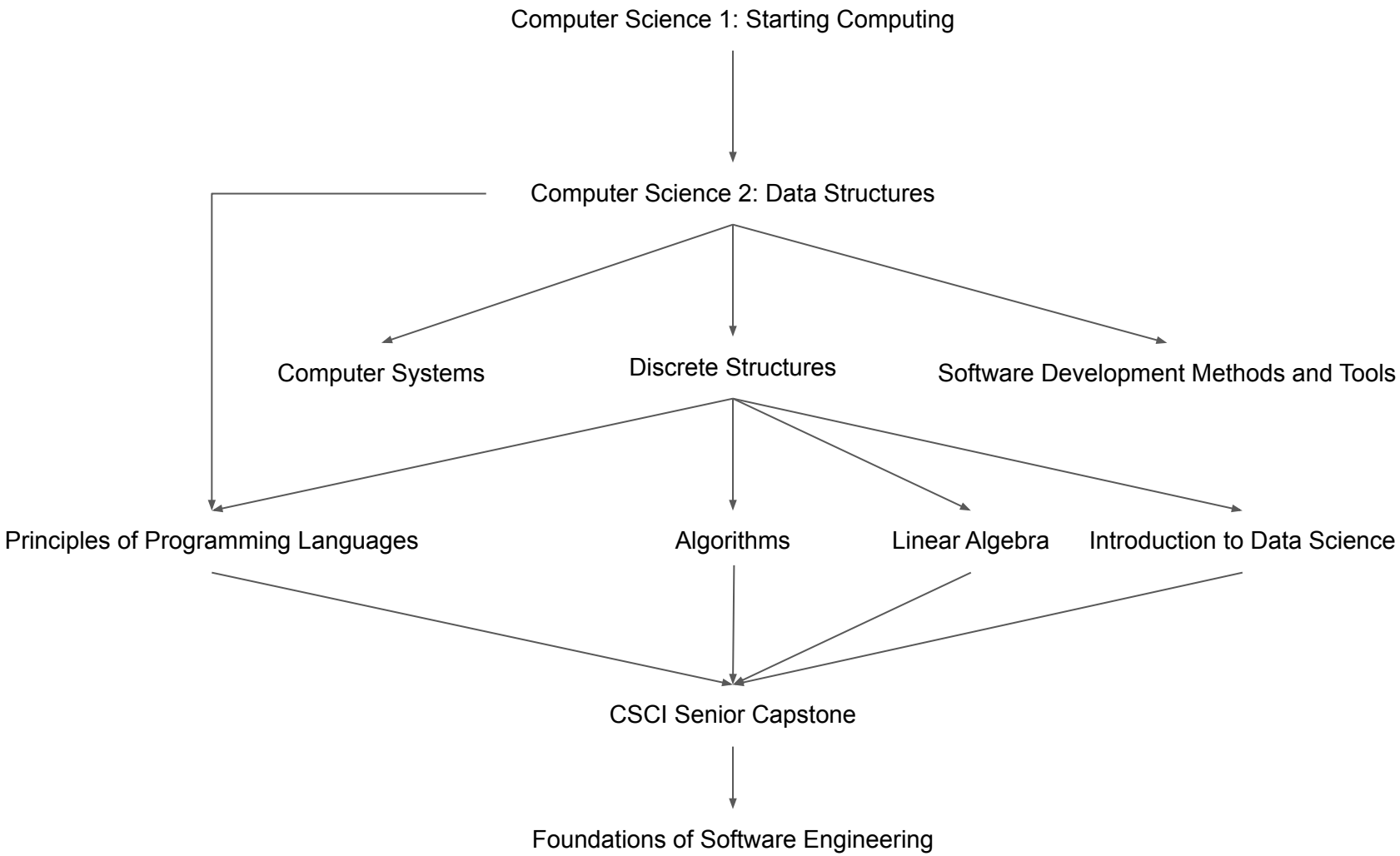

Publicly available RNA-seq samples currently available at GEO/SRA for human and mouse compared to available samples collected with the popular Affymetrix HG U133 Plus 2 platform

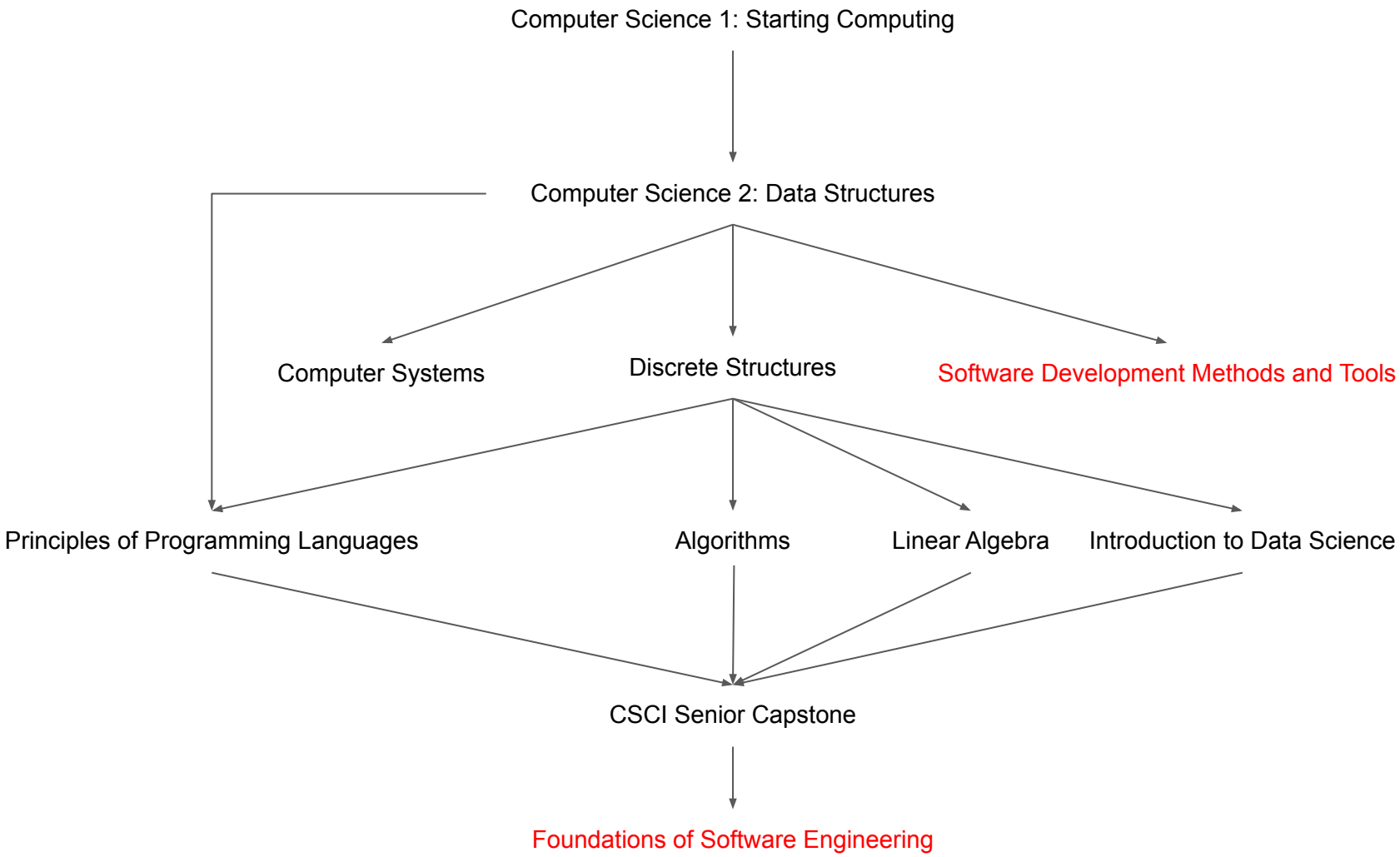
Lachmann, 2018

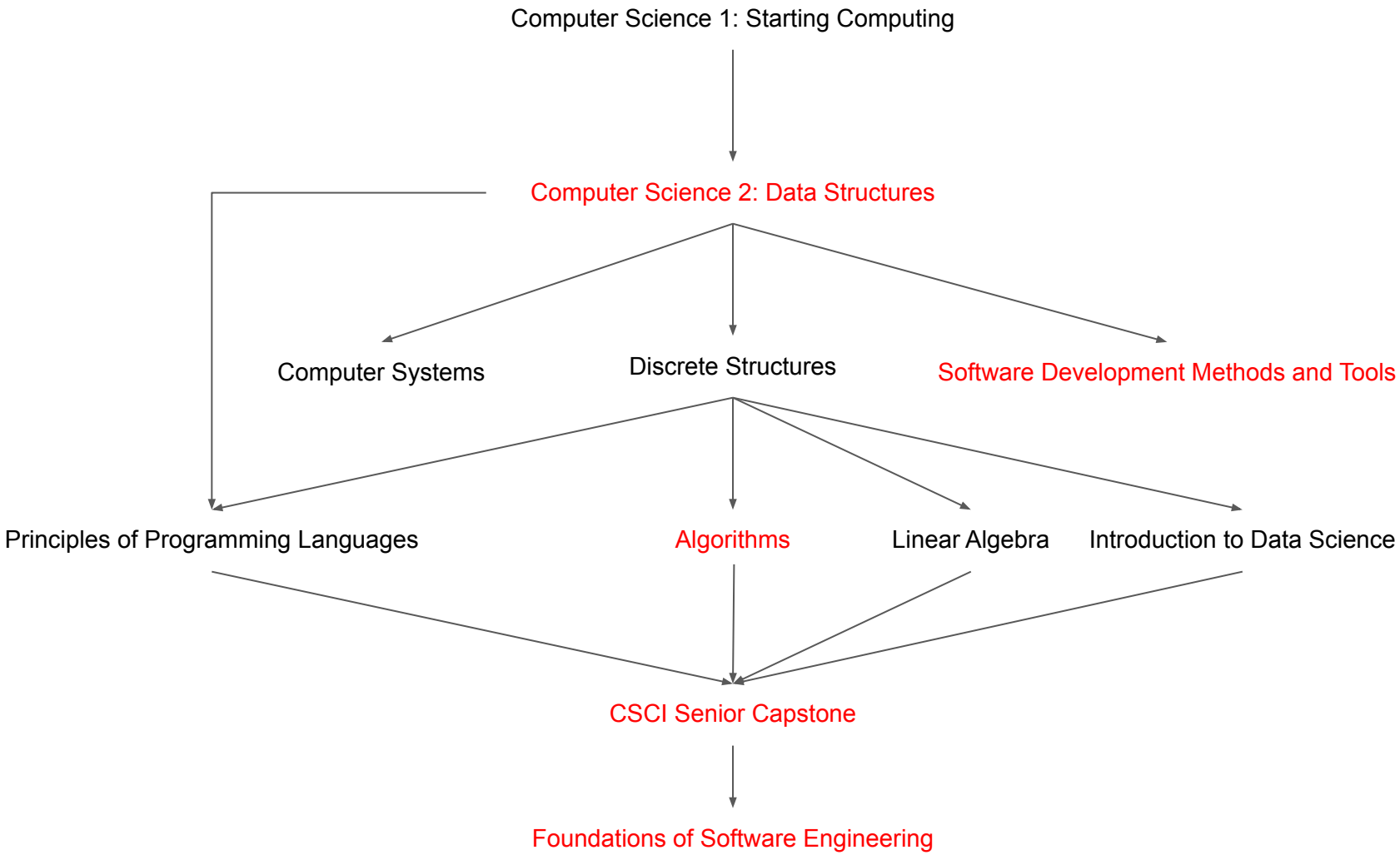
# Science is becoming data science!

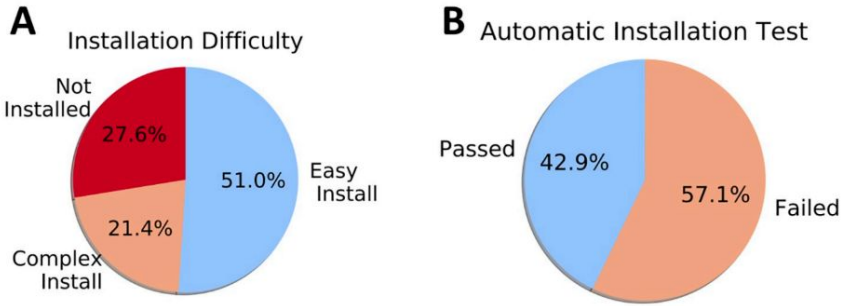



Ferreira, 2021

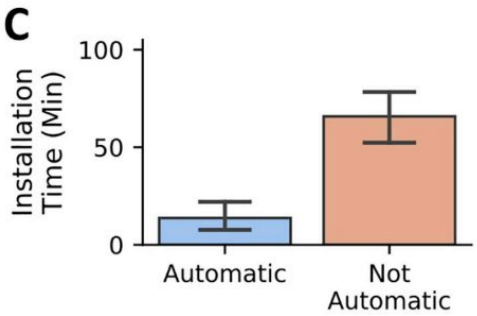

Computer Science 1: Starting Computing
Computer Science 2: Data Structures
Computer Systems
Discrete Structures
Software Development Methods and Tools
Principles of Programming Languages
Algorithms
Linear Algebra
Introduction to Data Science
CSCI Senior Capstone
Foundations of Software Engineering


Computer Science 1: Starting Computing
Computer Science 2: Data Structures
Computer Systems
Discrete Structures
Software Development Methods and Tools
Principles of Programming Languages
Algorithms
Linear Algebra
Introduction to Data Science
CSCI Senior Capstone
Foundations of Software Engineering

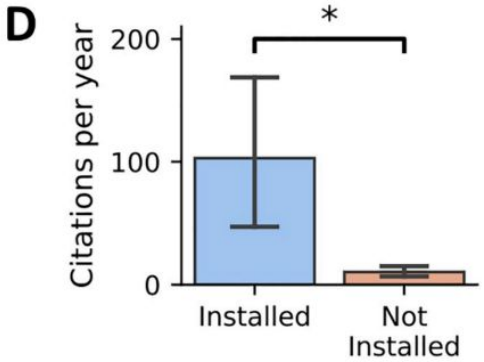

Computer Science 1: Starting Computing
Computer Science 2: Data Structures
Computer Systems
Discrete Structures
Software Development Methods and Tools
Principles of Programming Languages
Algorithms
Linear Algebra
Introduction to Data Science
CSCI Senior Capstone
Foundations of Software Engineering




Installation is hard




and slow




and it is affects impact

Mangul, 2019

# A random forest-based framework for genotyping and accuracy assessment of copy number variations

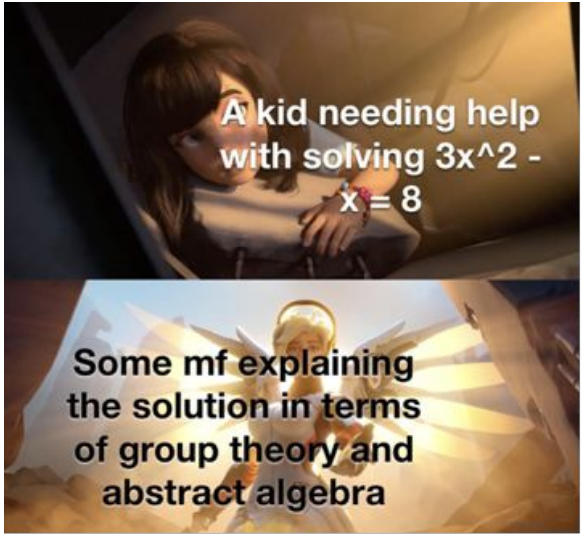
Xuehan Zhuang, Rui Ye, Man-Ting So, Wai-Yee Lam, Anwarul Karim, Michelle Yu, Ngoc Diem Ngo, Stacey S Cherny, Paul Kwong-Hang Tam, Maria-Mercè Garcia-Barcelo, Clara Sze-man Tang, Pak Chung Sham

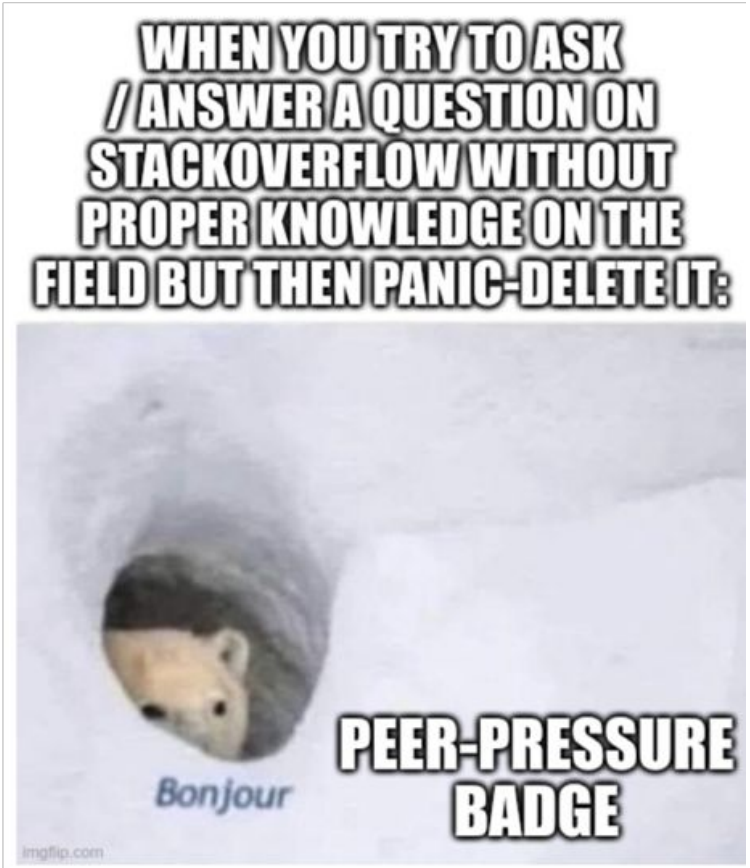
https://github.com/sunnyzxh/CNV-JACG

```
&showLog("Calculating het probability for $sample...");
`/home/yerui/miniconda2/bin/perl $home/bin/Het-prob.pl -f $ref -m $snps -r $precnv -i $bam -o $outdir/$precnvbase.het.prob`;
```

Coding best practices
Version control
Git and GitHub
Testing
Continuous integration
Code reviews
Benchmarking
Workflows
Algorithms
Data structures
Some data science
Project

SWE4S


A kid needing help with solving 3x^2 - x = 8
Some mf explaining the solution in terms of group theory and abstract algebra


WHEN YOU TRY TO ASK / ANSWER A QUESTION ON STACKOVERFLOW WITHOUT PROPER KNOWLEDGE ON THE FIELD BUT THEN PANIC-DELETE IT:
Bonjour
PEER-PRESSURE BADGE
imgflip.com

# Technologies we'll use

- Course communication and announcements: Slack and Canvas
  - Slack is better so let's try to use/lean towards Slack for communication
- Grades: Canvas
- Lecture slides, course notes, PDFs of assignments: course webpage
- Assignment submission: GitHub Classroom

COURSE WEBPAGE + SYLLABUS

# About me

- From Salt Lake City
- Physics + math undergrad
- Applied math PhD at CU, co-advised by Dan Larremore (CS + BioFrontiers) and Stephen Becker (applied math)
- ~1 year at Invitae as bioinformatician
- ~1 year at TetraScience as software and data engineer
- Back at CU teaching this class and doing a postdoc with Ryan Layer (CS + BioFrontiers) and Stephen Kessler (CS)
- This is my first time teaching this class! Feedback and any and all questions are welcome!

# Who are you!?

- Name
- Program and year
- Research interests
- Previous coding and software experience
- Fun fact
- One thing you’re hoping for from this class
- One thing you’re nervous/apprehensive about

# For next time…

- Fill out the getting-to-know-you Google Form
- Get your development environment set up
  - Refer to the notes on the course webpage
  - This might take some time!
- Make sure you can access our course webpage, Slack, Canvas, GitHub Classroom
- Let me know if you have any questions!