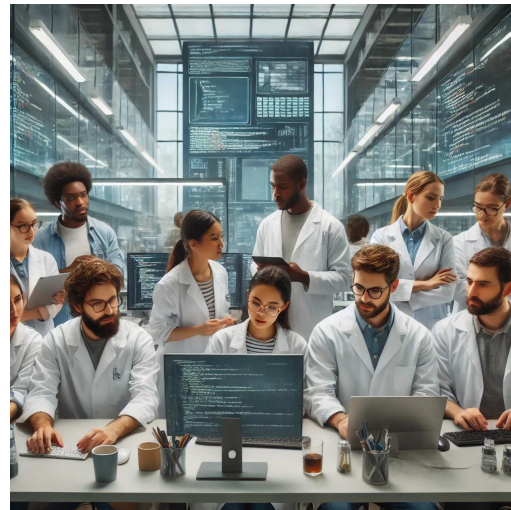
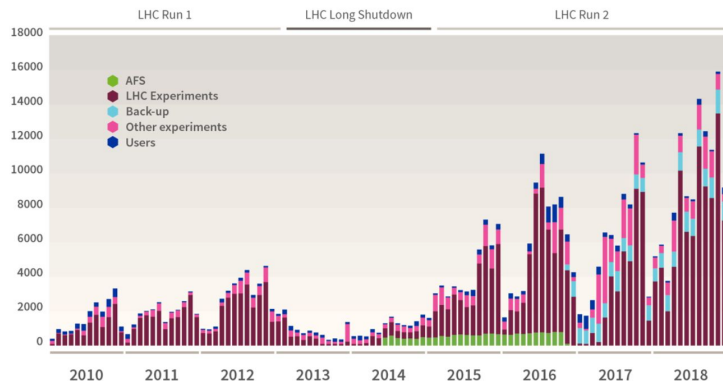


# Software Engineering for Scientists (SWE4S)



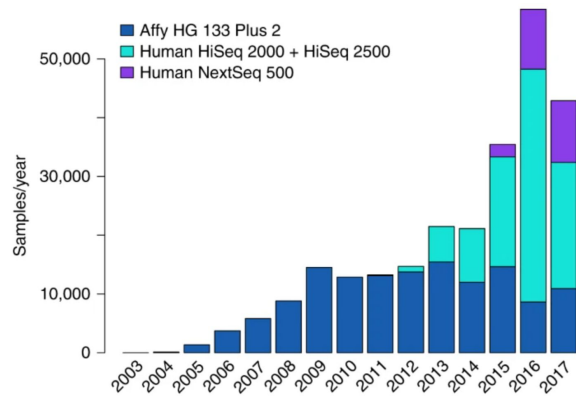
# Today's plan/objectives

- 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)

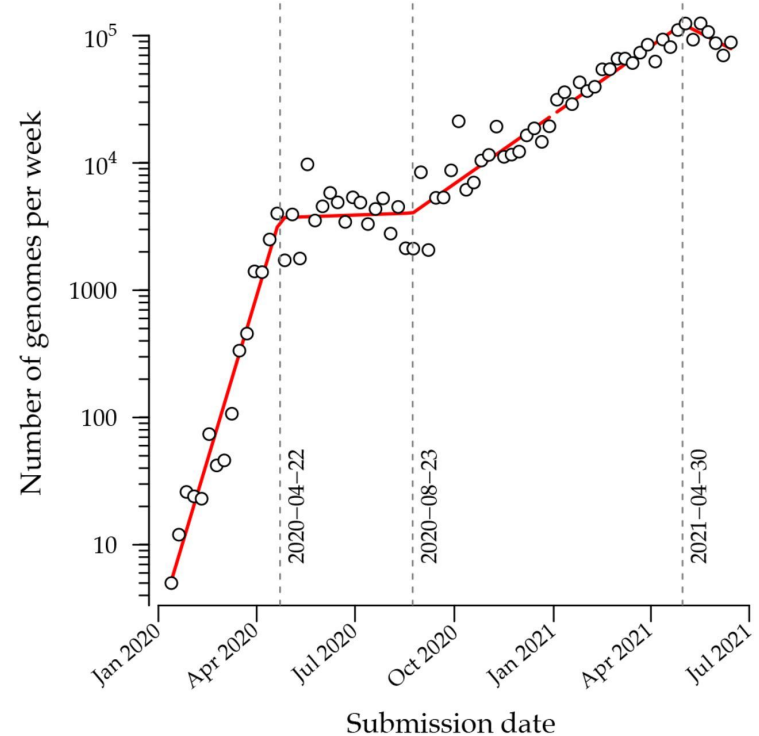
## Ozcesmeci, 2019



Publicly available RNA-seq samples currently available at GEO/SRA for human and mouse compared to available samples collected with the popular Affymetrix HG UI33 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

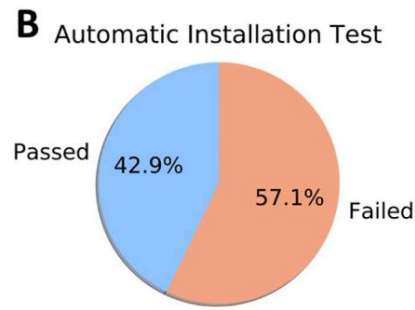
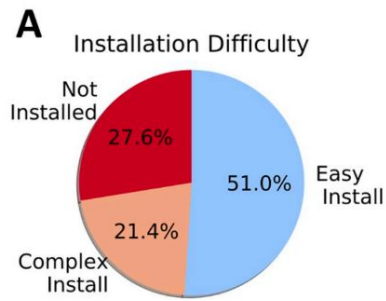
Algorithms

Linear Algebra

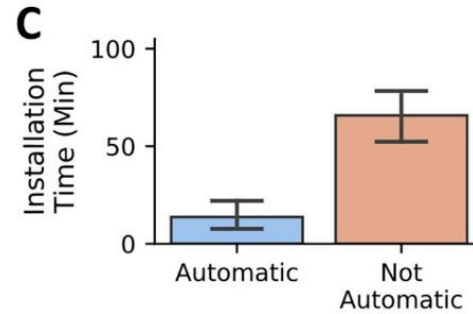
Introduction to Data Science

CSCI Senior Capstone

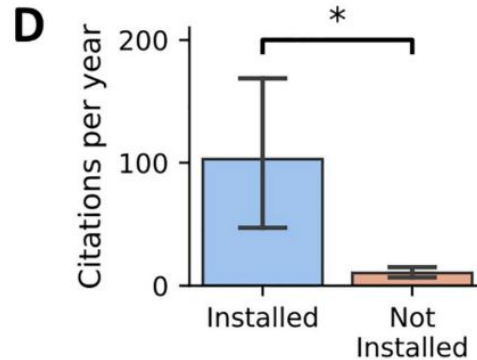
Foundations of Software Engineering



Installation is hard



and slow



and it is affects impact

# 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>

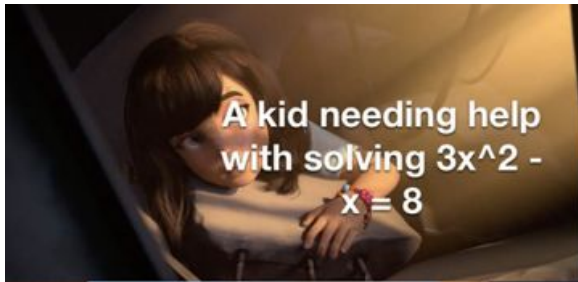
```
&showLog("Calculating het probability for $sample...");  
`/home/yerui/miniconda2/bin/perl $home/bin/Het-prob.pl -f $ref -m $snps -r $precnv -i $bam -o $outdir/$precnvbase.het.prob`;
```



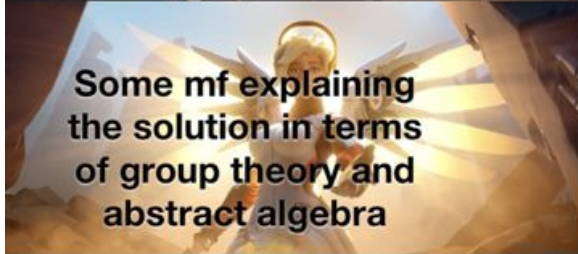
Coding best practices  
Version control  
Git and GitHub  
Testing  
Continuous integration  
Code reviews  
Benchmarking  
Workflows  
Algorithms  
Data structures  
Some data science  
Project



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**

# Where to go for:

- Course communication and announcements: Slack
- Course notes, slides, etc.: course GitHub page
- Assignments and grades: Canvas
- Assignment submission: GitHub Classroom

<https://github.com/swe4s/lectures>

<https://canvas.colorado.edu/courses/127183>

**Erik's office hours (while I'm teaching):  
T, Th 9-9:50am, JSCBB 315D**

# Development environment setup

- Since we'll mostly be using/covering the command line, shell scripting, Python, and git, every student will need access to a Unix-like command line and a text editor for writing code.
- Text editor: here are many different text editors and people have their preferences. In my experience, most Python (and other language) software developers use PyCharm or Visual Studio Code (VS Code). I suggest using [VS Code](#).
- Notes on [development environment](#) and [Conda](#)

## Checklist for next time:

1. Make sure you have access to a Unix-like command line
  - a. For Mac and Linux users, I suggest using your personal computer rather than CU's reference environment. Macs and Linuxs come with a Unix-like command line. Just make sure you know how to access it (e.g., via the "Terminal " app on Macs)
  - b. For Windows users, I suggest spending some time trying to set up your personal computer and, if that's proving too painful, then use CU's reference environment. For a Unix-like command line, Windows users have a few options one of which is described in the first/top-voted answer in [this thread](#) (assumes you're using Visual Studio Code). Another option is [here](#)
2. Pick and install a text editor (again, I highly suggest using Visual Studio Code)
3. Make sure you have git installed and can access it via the command line
4. Install mamba and make sure you can create and activate a new environment

## For next time...

- Get your development environment set up
  - Read the [development environment](#) and [Conda](#) notes
  - Make sure you've completed the checklist on the previous, "Development environment setup" slide. This might take some time!
  - We'll start coding on Tuesday so do this before class.
  - Come to office hours on Tuesday if you want help or to talk about anything
- Join the class Slack channel (via link in syllabus)
- Let me know (via Slack or office hours) if you have any questions