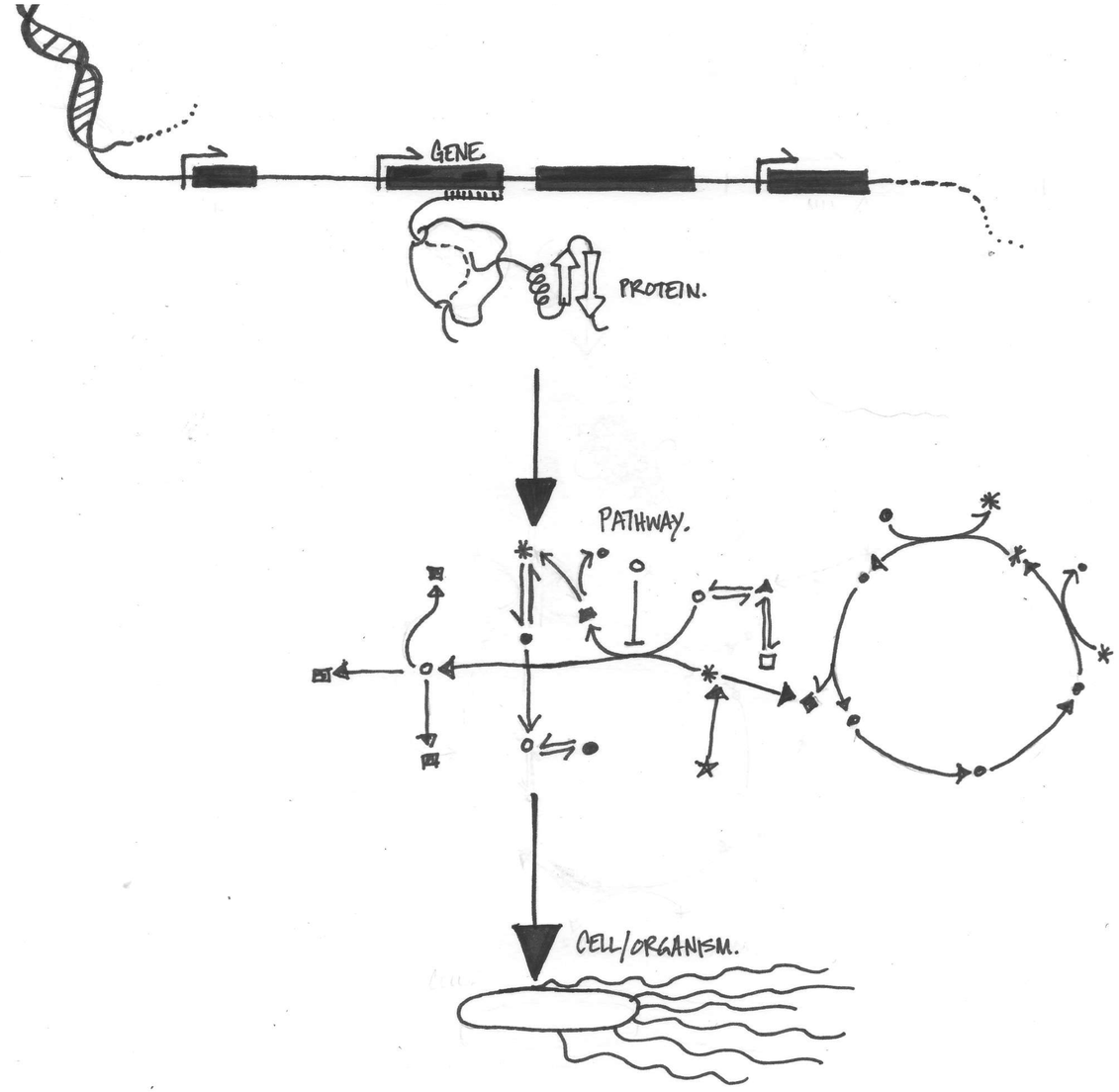
Welcome!

Python Programming for Enabling Your Science

STARS 2024
UT Southwestern Medical Center

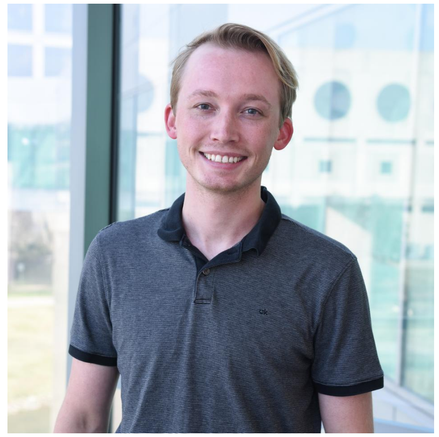
My own lab uses a combination of computation and experiment in our research.
We want to understand:



How individual mutations (variation) at the genetic level…

ultimately shape appearance, behavior, and growth
at the organismal level.

These are the folks that have made this class possible:



Ryan Otto

Dominique Lagasca

Suzette Palmer

Ashley Vu

Tasia Bos

Phil Brown

Jerry Dinan

Eryn Sale

Let’s get started with a download: https://www.anaconda.com/

If you are using a Mac for this course you will want the 64 bit graphical installer. You will need to pay intention to whether or not you have an Apple silicon or intel processor.

If you wish to double check, you can go to the apple menu, then select:

“**About This Mac**”

Under **Processor**, it will specify the type

If you are using a PC, and not sure if you need 64 or 32 bit… you probably need 64.

If you wish to double check, you can go to the START button, then select:

Settings -> **System** -> **About**

Under **Device specifications**, see **System type**

Let's get started with a download: https://www.anaconda.com/

If you are using a Mac for this course you will want the 64 bit graphical installer.

If you are using a PC, and not sure if you need 64 or 32 bit… you probably need 64.

If you wish to double check, you can go to the START button, then select:

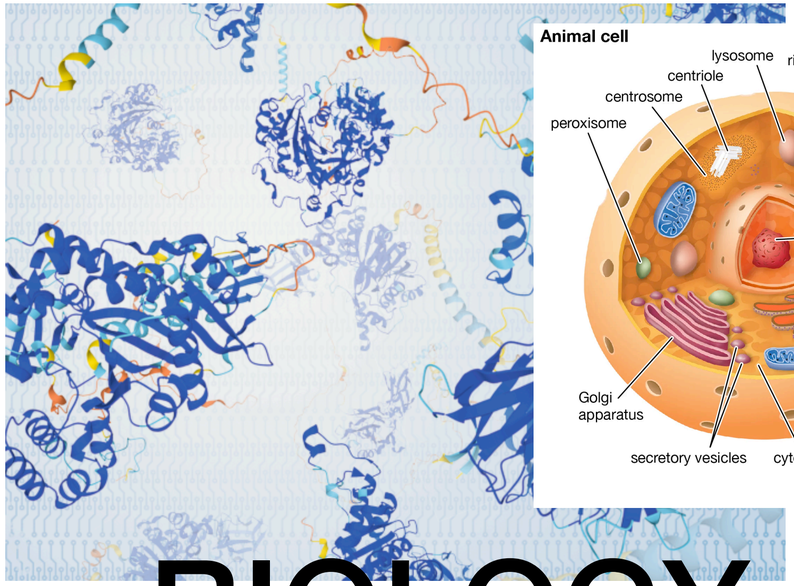
Settings -> **System** -> **About**

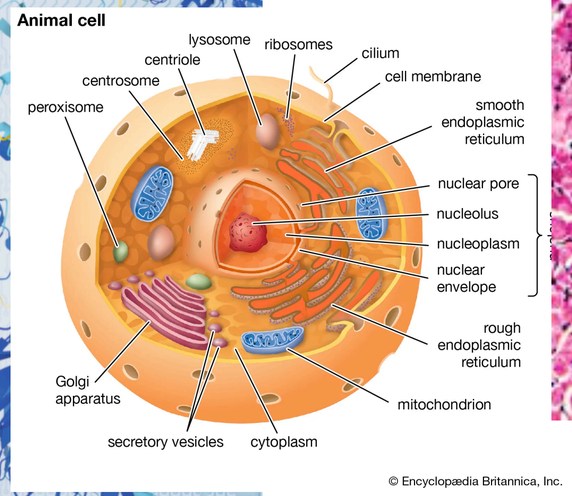
Under **Device specifications**, see **System type**

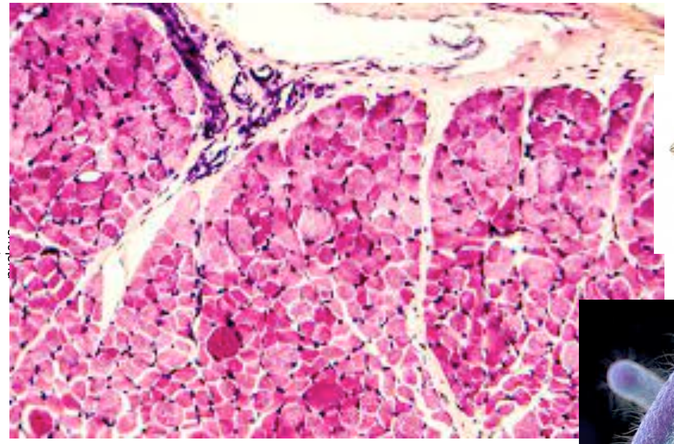
Anaconda is free.

It is one (of several) possible Python distribution platforms.

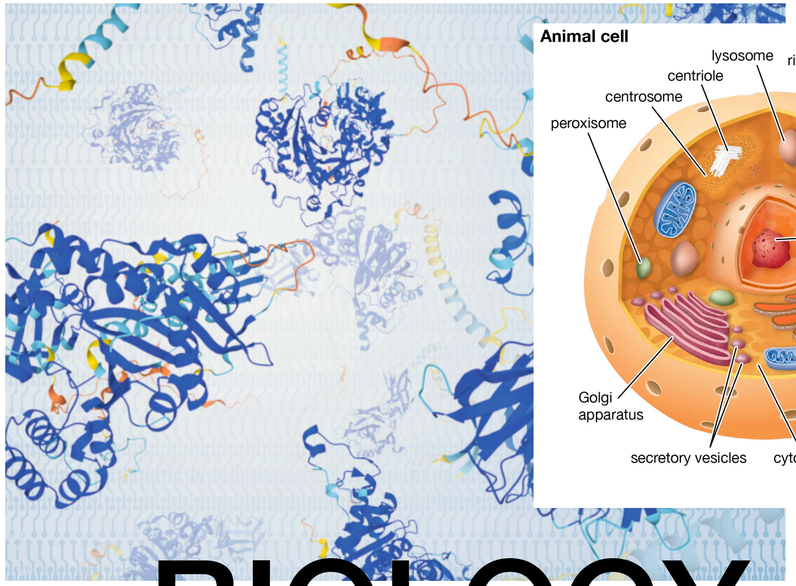
It includes a number of other tools that are useful to scientific computing and which we will make use of in this class.

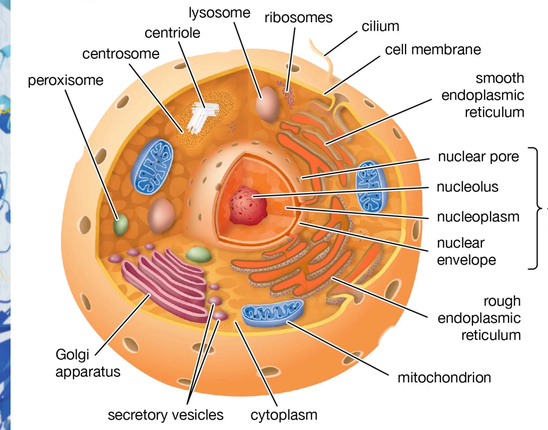


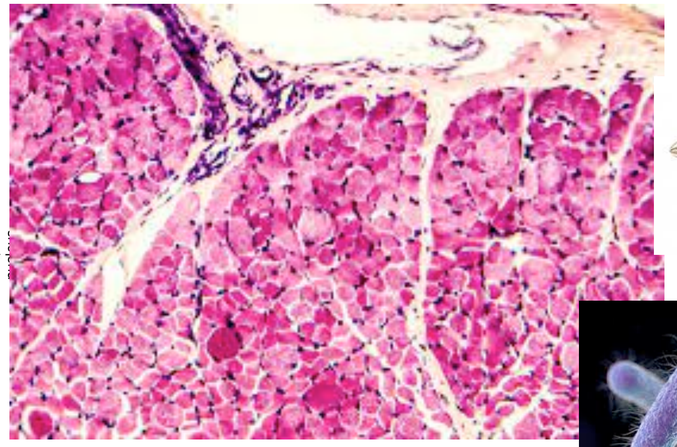

Animal cell
lysosome
ribosomes
centriole
cilium
centrosome
cell membrane
peroxisome
smooth endoplasmic reticulum
nuclear pore
nucleolus
nucleoplasm
nuclear envelope
rough endoplasmic reticulum
Golgi apparatus
mitochondrion
secretory vesicles
cytoplasm
© Encyclopædia Britannica, Inc.



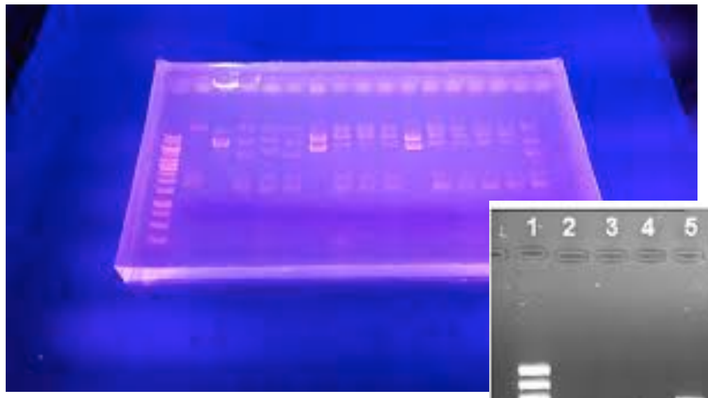
BIOLOGY…

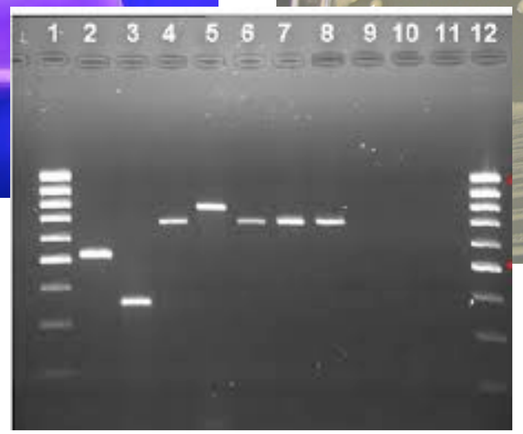


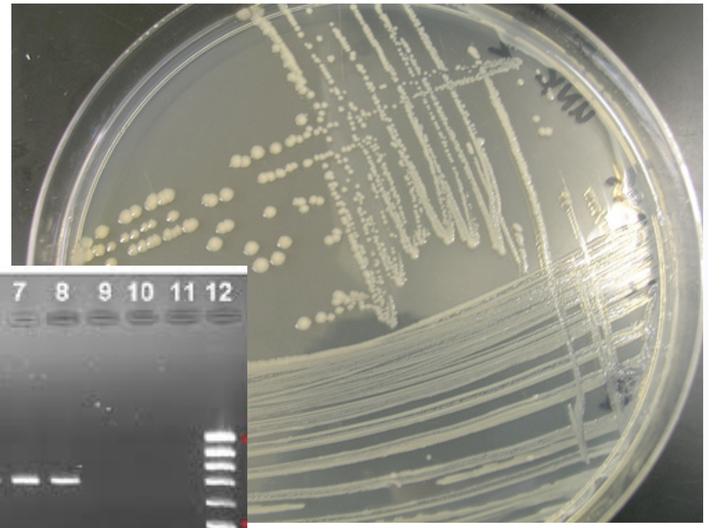


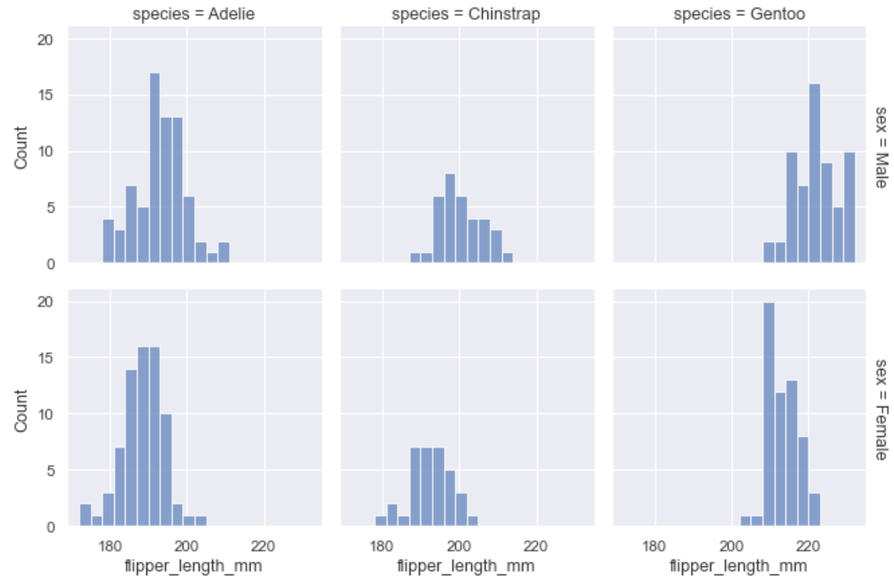


BIOLOGY… WHY CODE?

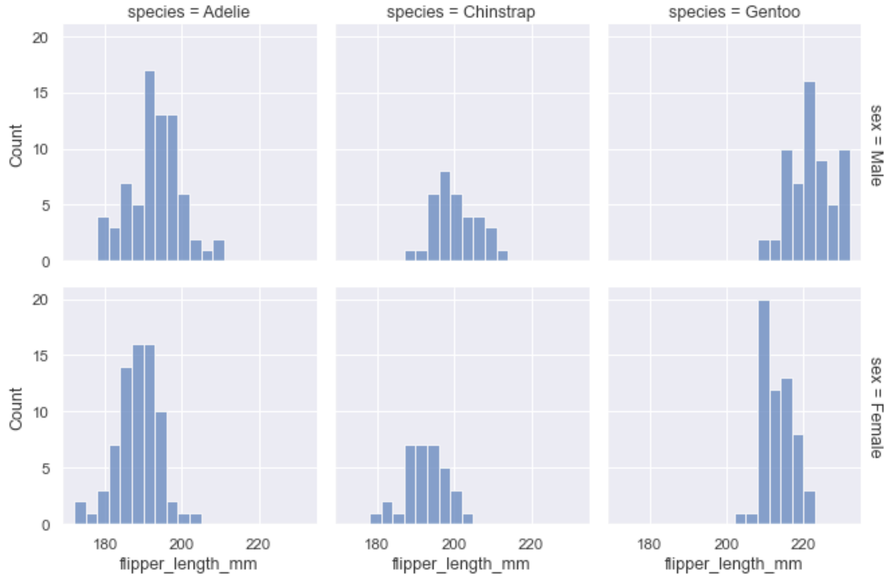


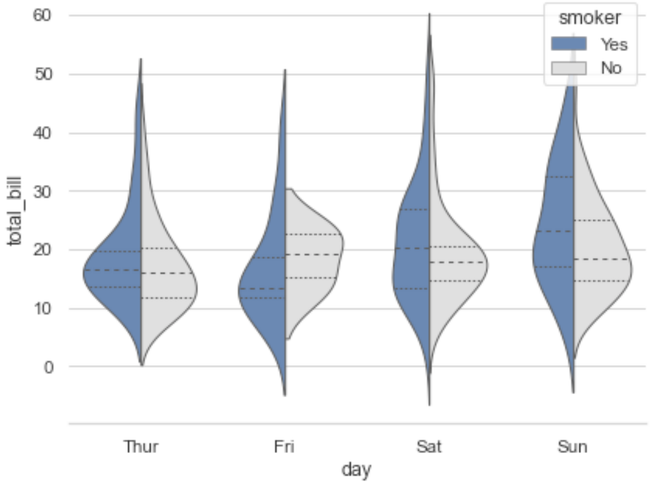


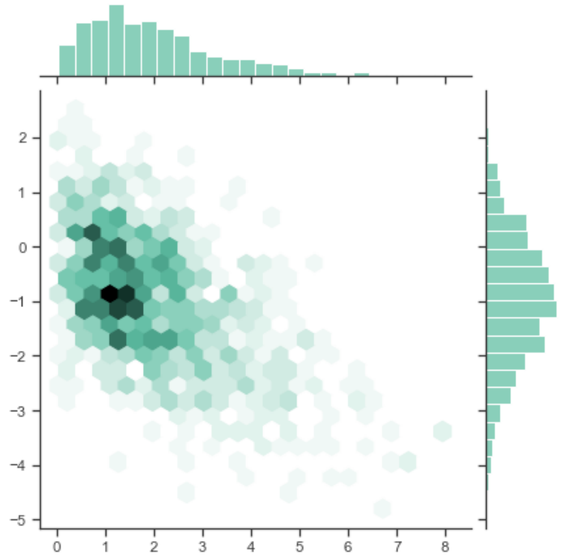




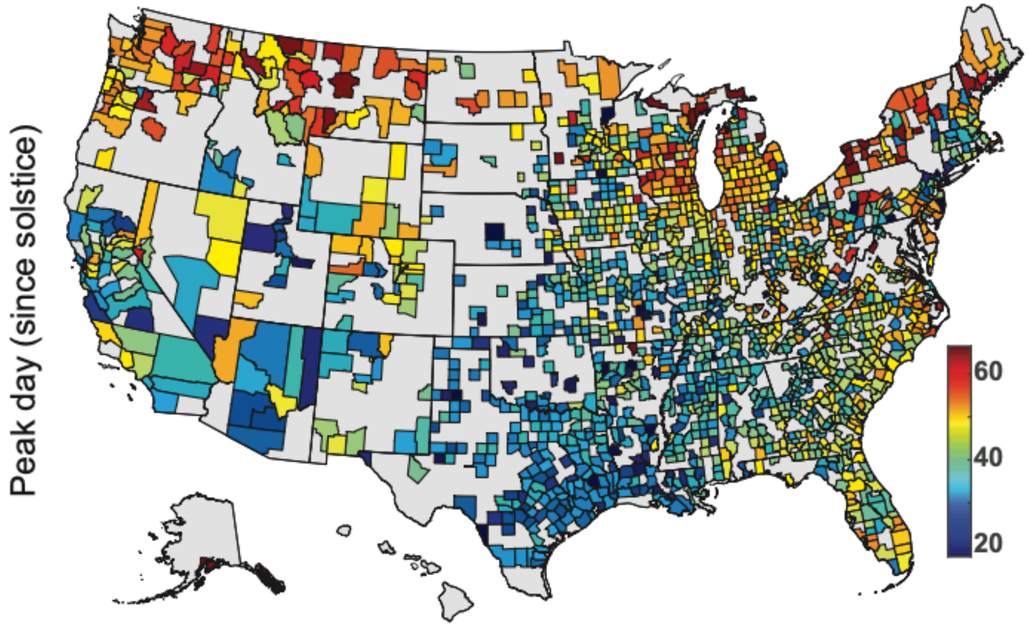
Data Visualization.

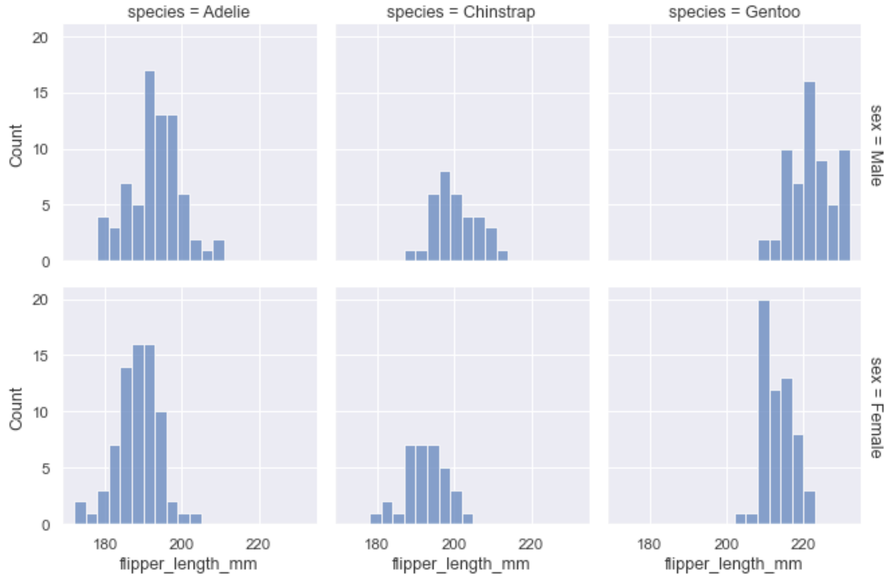


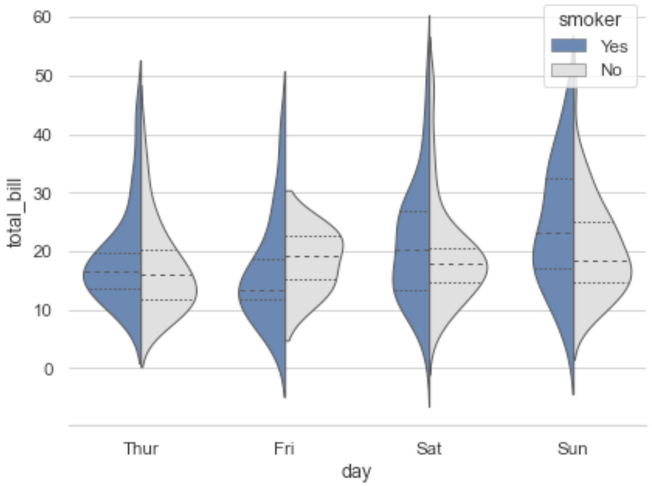


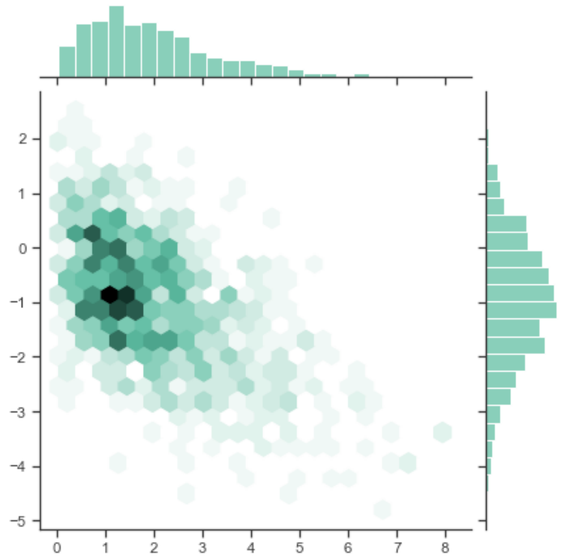


Data Visualization.

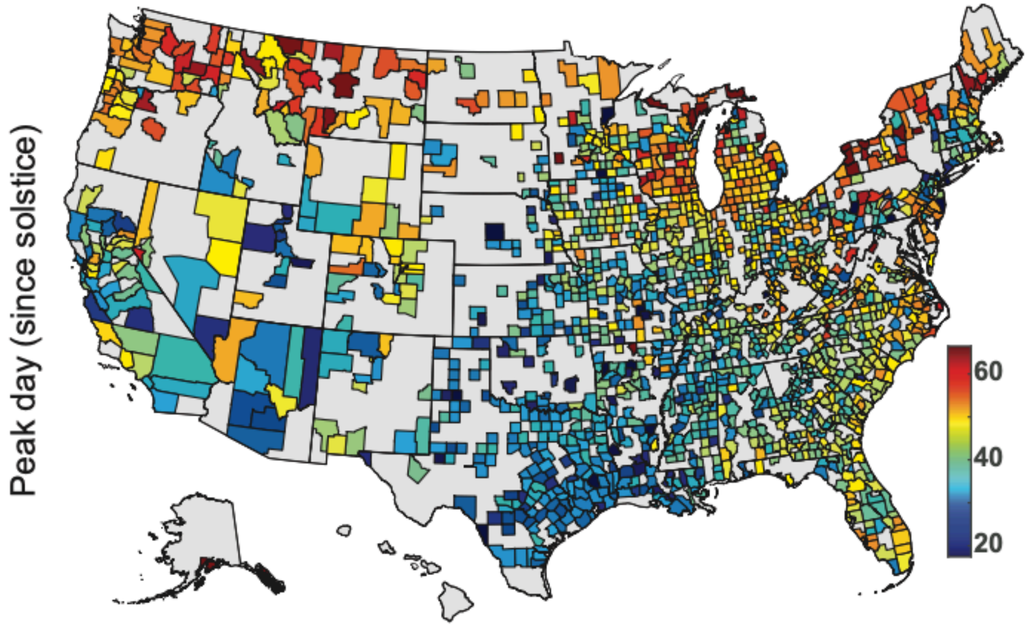








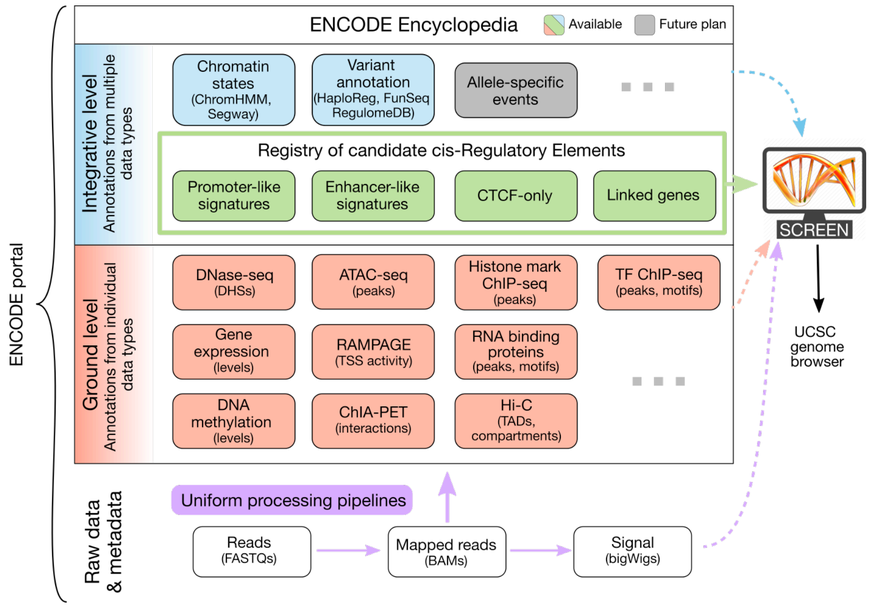
Data Visualization.



matplotlib

seaborn

plotly



Big Data!

NATIONAL CANCER INSTITUTE
THE CANCER GENOME ATLAS

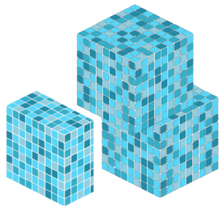
TCGA BY THE NUMBERS

TCGA produced over

2.5

PETABYTES

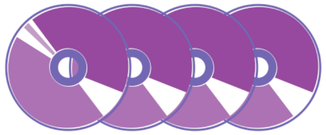
of data



To put this into perspective, **1 petabyte** of data is equal to

212,000

DVDs



TCGA data describes

33

DIFFERENT TUMOR TYPES

...including

10

RARE CANCERS

...based on paired tumor and normal tissue sets collected from

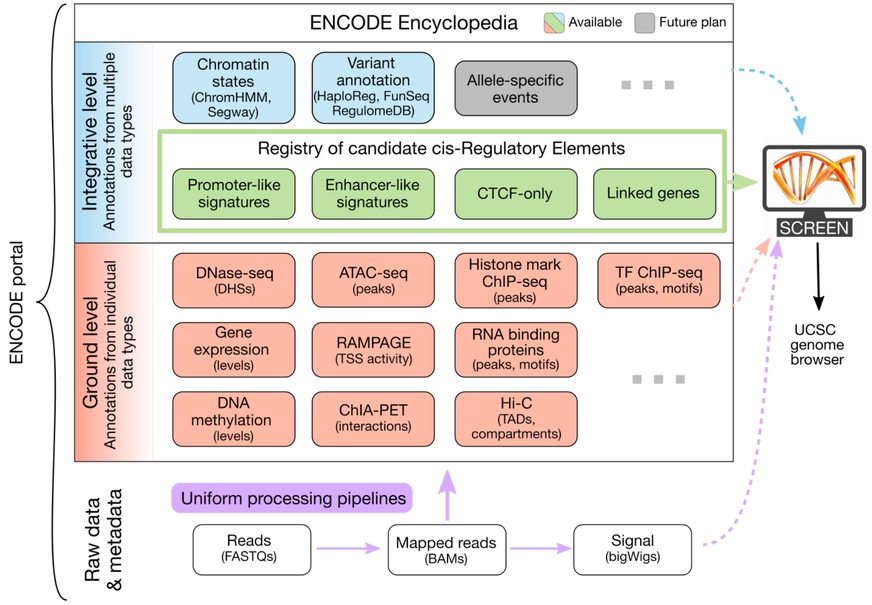
11,000

PATIENTS

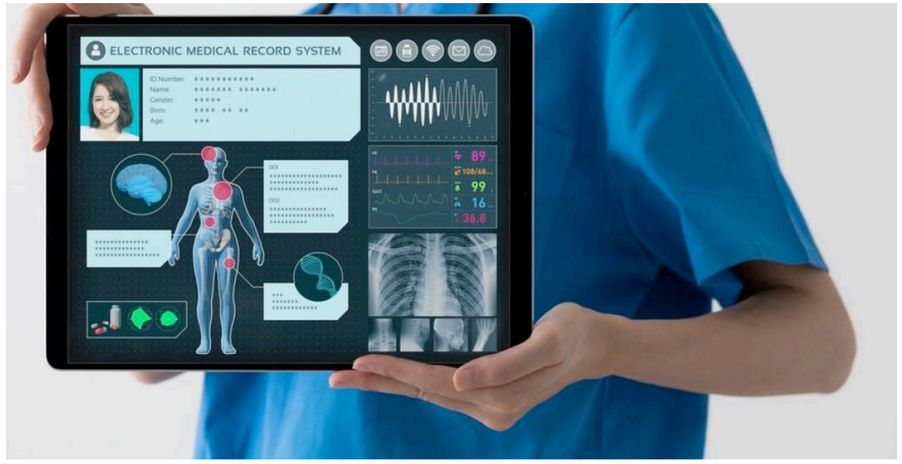
...using

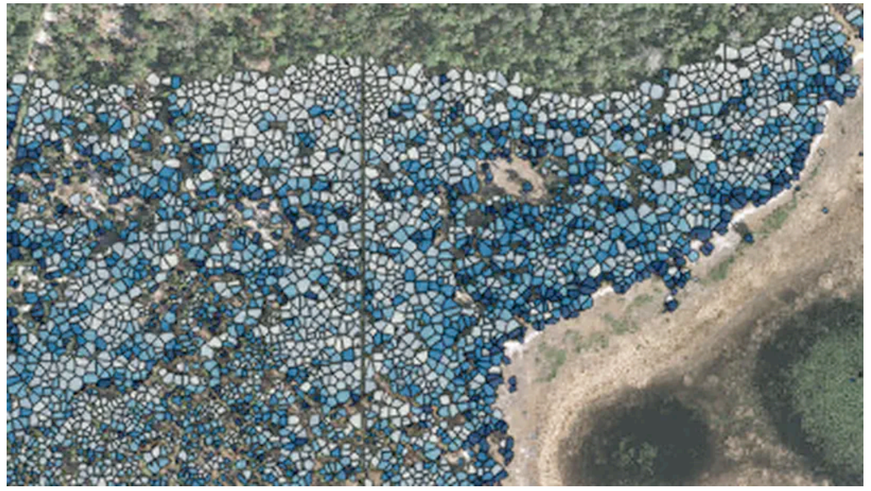
7

DIFFERENT DATA TYPES



Big Data!



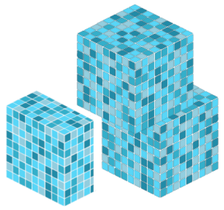


NATIONAL CANCER INSTITUTE
THE CANCER GENOME ATLAS

TCGA BY THE NUMBERS

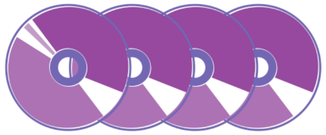
TCGA produced over

2.5 PETABYTES of data



To put this into perspective, **1 petabyte** of data is equal to

212,000 DVDs



TCGA data describes

33 DIFFERENT TUMOR TYPES

...including

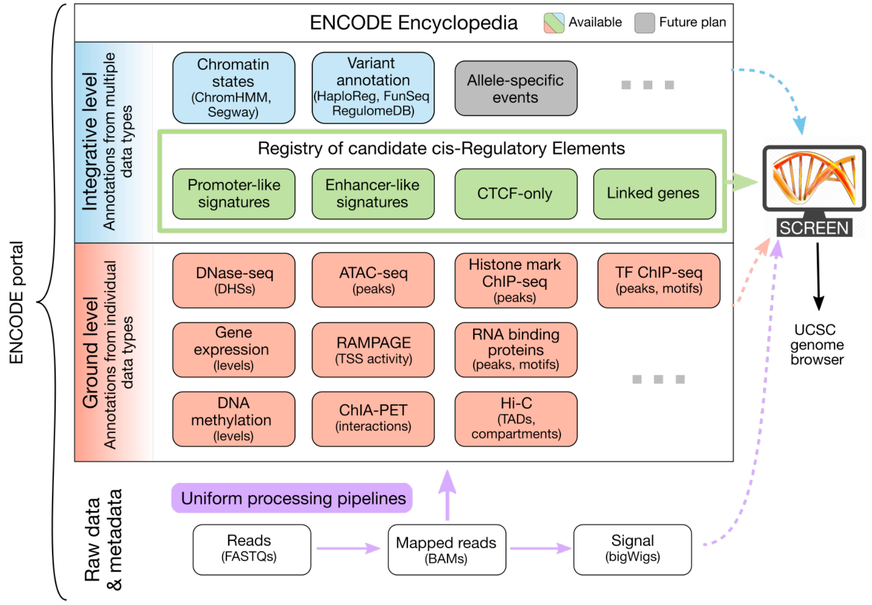
10 RARE CANCERS

...based on paired tumor and normal tissue sets collected from

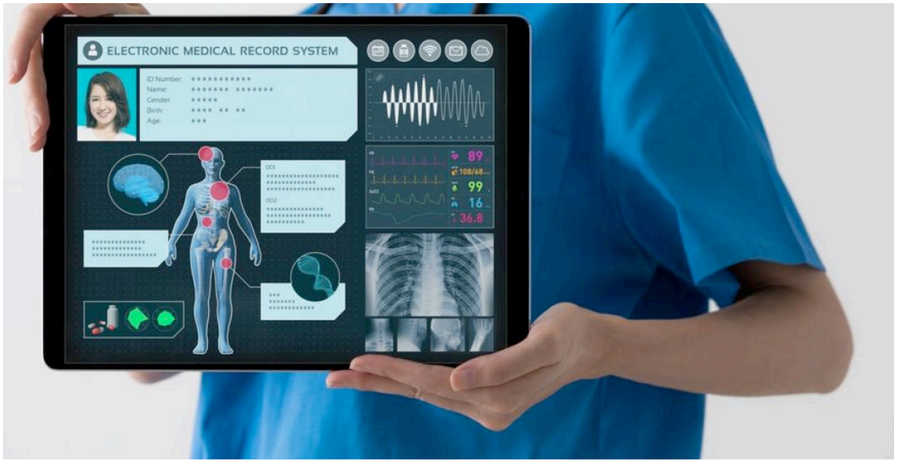
11,000 PATIENTS

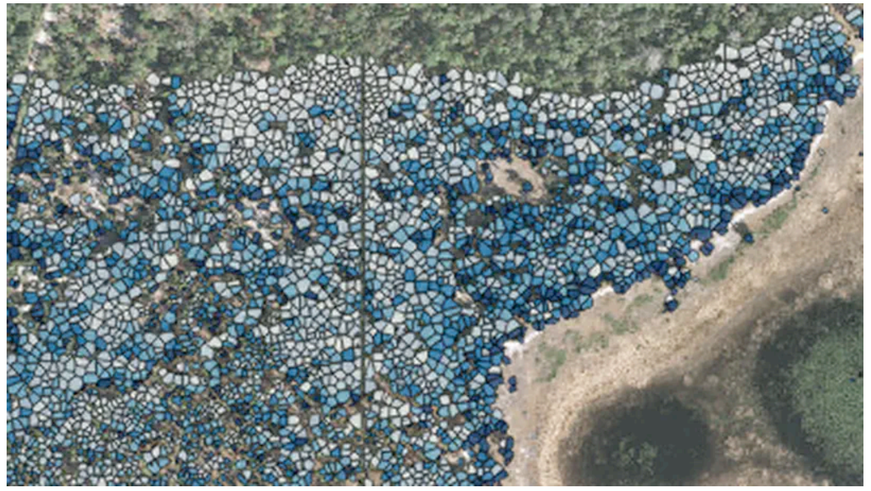
...using

7 DIFFERENT DATA TYPES



Big Data!



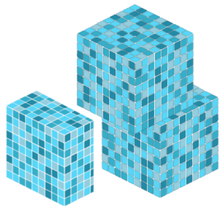


NATIONAL CANCER INSTITUTE
THE CANCER GENOME ATLAS

TCGA BY THE NUMBERS

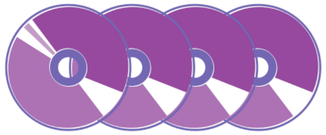
TCGA produced over

2.5 PETABYTES of data



To put this into perspective, **1 petabyte** of data is equal to

212,000 DVDs



TCGA data describes **33** DIFFERENT TUMOR TYPES

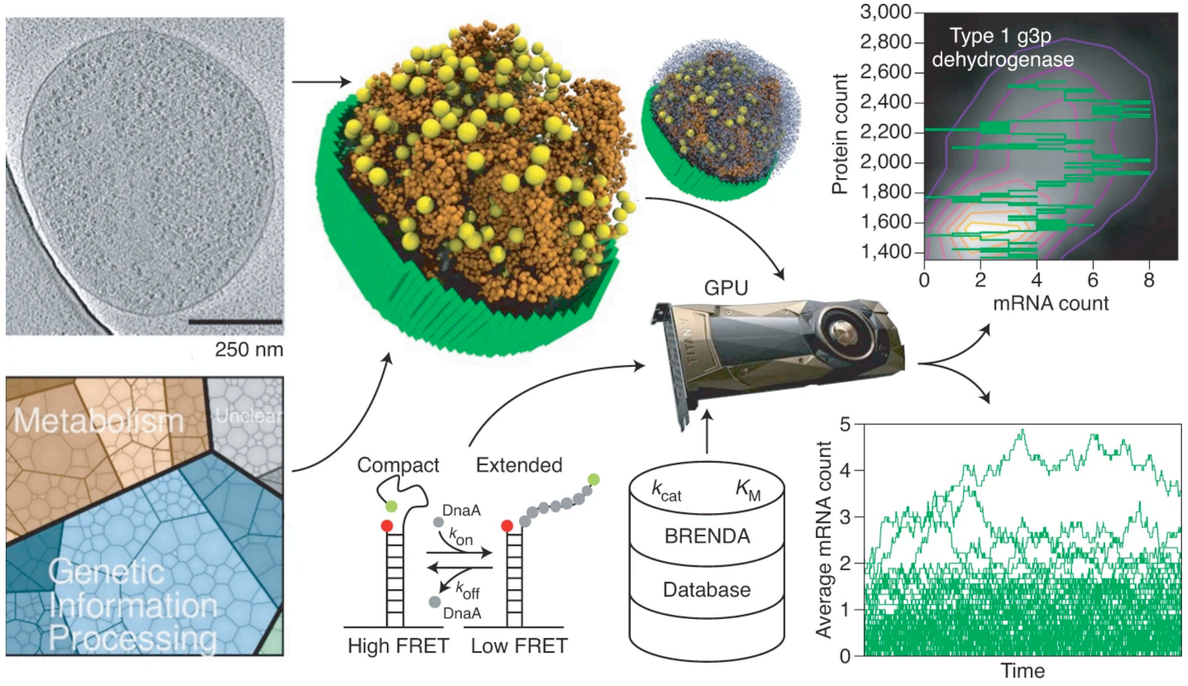
...including **10** RARE CANCERS

...based on paired tumor and normal tissue sets collected from

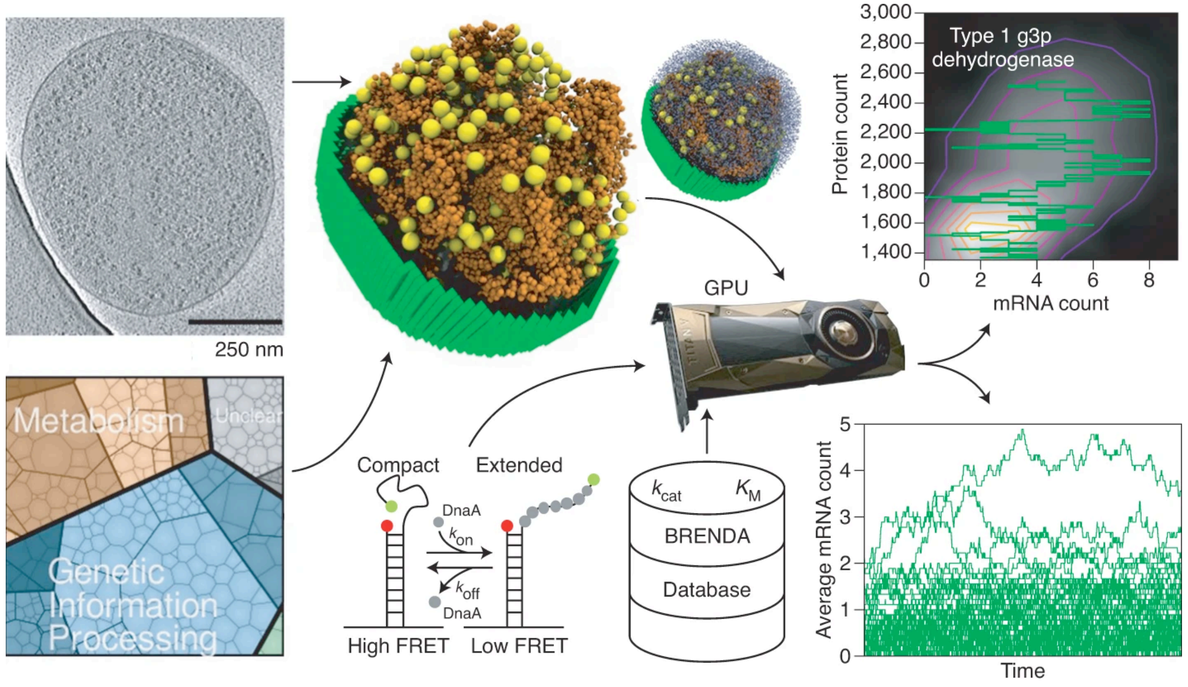
11,000 PATIENTS

...using **7** DIFFERENT DATA TYPES

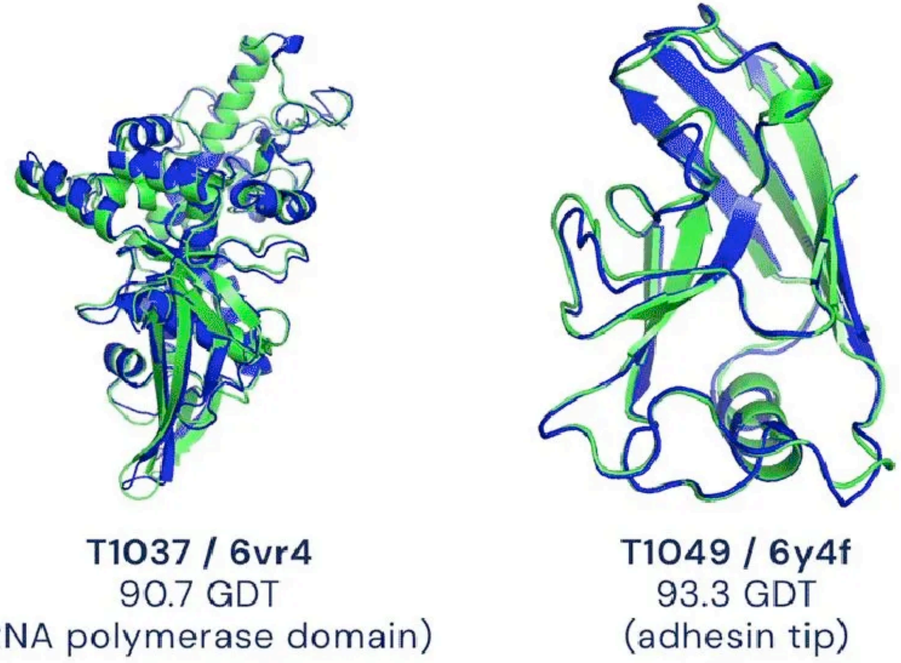
pandas



Modeling, simulation, and AI



Modeling, simulation, and AI



T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)

T1049 / 6y4f
93.3 GDT
(adhesin tip)

Experimental result
Computational prediction

A widely used programming language that is valued for its read-ability and ease of use.

Is a common (if not the most common) language in data science, and interfaces well with a number of code libraries for machine learning and AI.

Our learning objectives for this summer:

- Build familiarity with Python data types (integers, float, boolean, list, dictionary) and flow control (if/then/else, while , for loops)
- Learn to read in and write out data
- Use Python to perform basic statistics (mean, variance)
- Use Python to create well-annotated plots

We will orient our lessons around a single (toy) data set.

Toy (or idealized, or fake) data play a useful role in research.
You can often use them to build your analyses or think about how to approach a problem before considering more challenging and less well-behaved real data.

But it is essential to be clear that they are not real, and are never presented as real

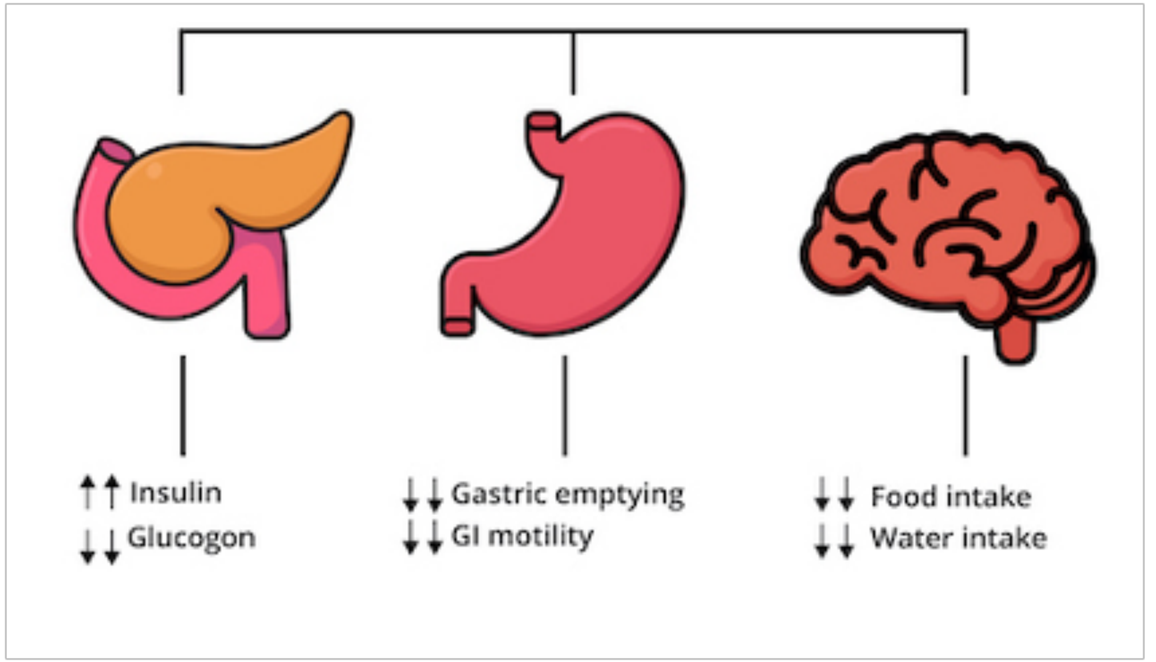
Mock Data Set: Mouse Obesity

Glucagon-Like Peptide -1

GLP-1

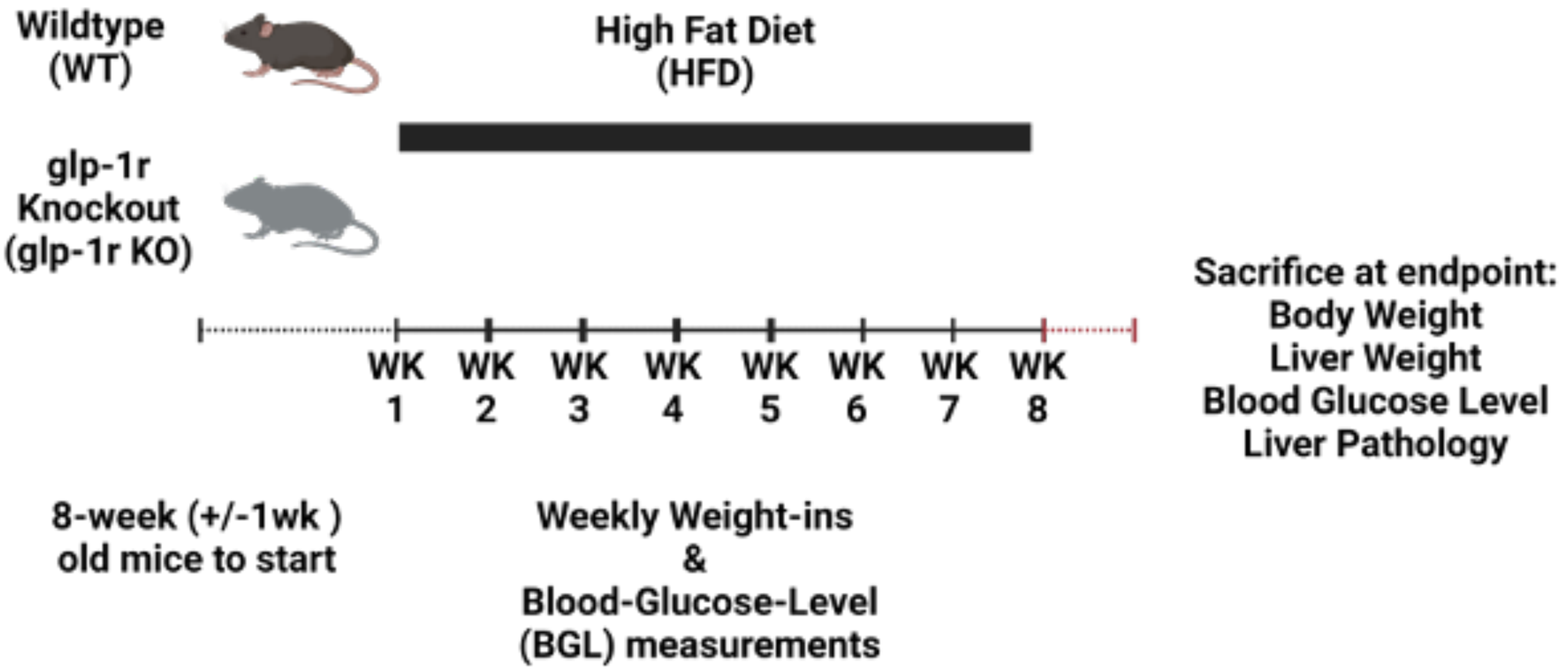
GLP-1 Agonist
(e.g. Liraglutide, Ozempic, etc.)

GLP-1 Receptor

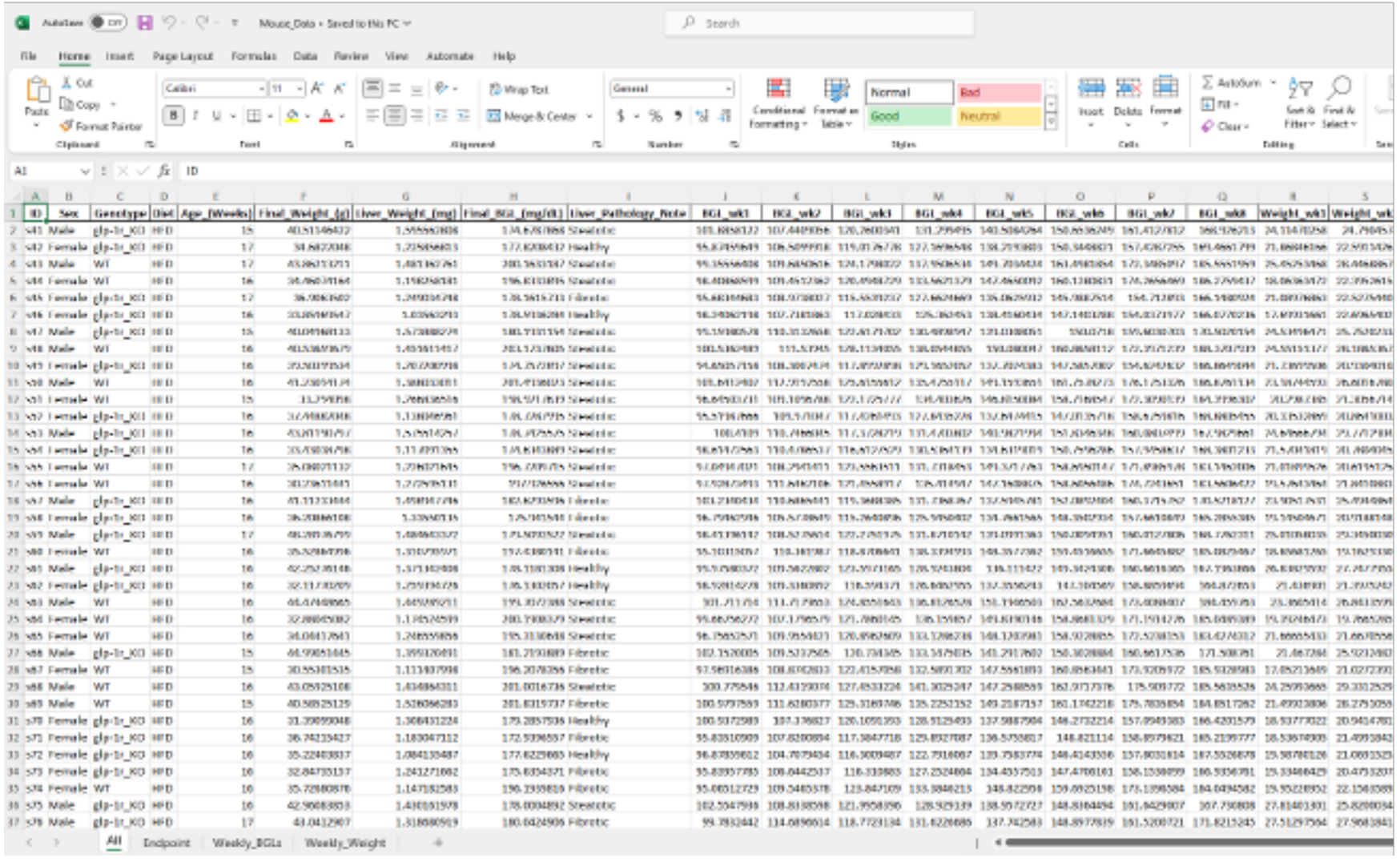


https://utswmed.org/medblog/obesity-medication-weight-program/

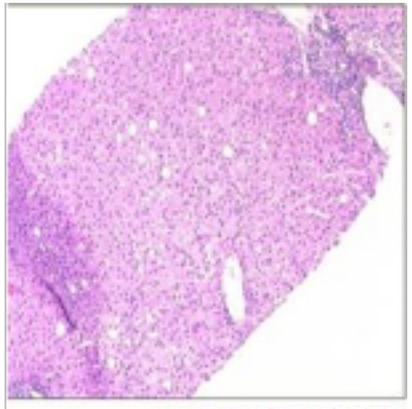
Mock Data Set: “Experiment” Layout



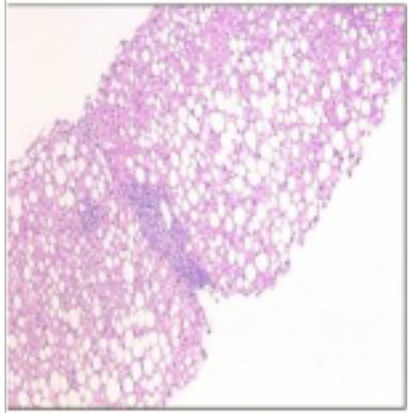
Mock Data Set:



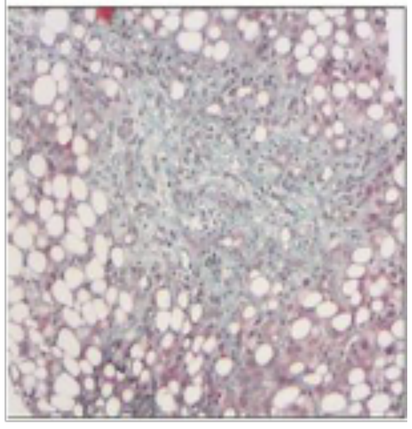
Normal Liver (Healthy)



Fatty Liver (Steatotic)



Liver Fibrosis (Fibrotic)



Ok, so with this in mind, let's get started with installing anaconda.

Step-by-step:

- Navigate to the directory where you downloaded anaconda
- Double click to open the installer
- Click "next" to start the installation
- Agree to the license agreement
- Select to install the software for "just me" (not All Users/system wide)
- Choose Install Location. The default is usually a good choice (on Mac this will be your /Applications folder). Take note of where it is being installed.
- Do not add Anaconda to your PATH, but do register Anaconda as your default Python
- Wait for the install…
- You don't need to check the "learn more" buttons, you can click finish after the install is completed.

https://www.geeksforgeeks.org/how-to-install-anaconda-on-windows/

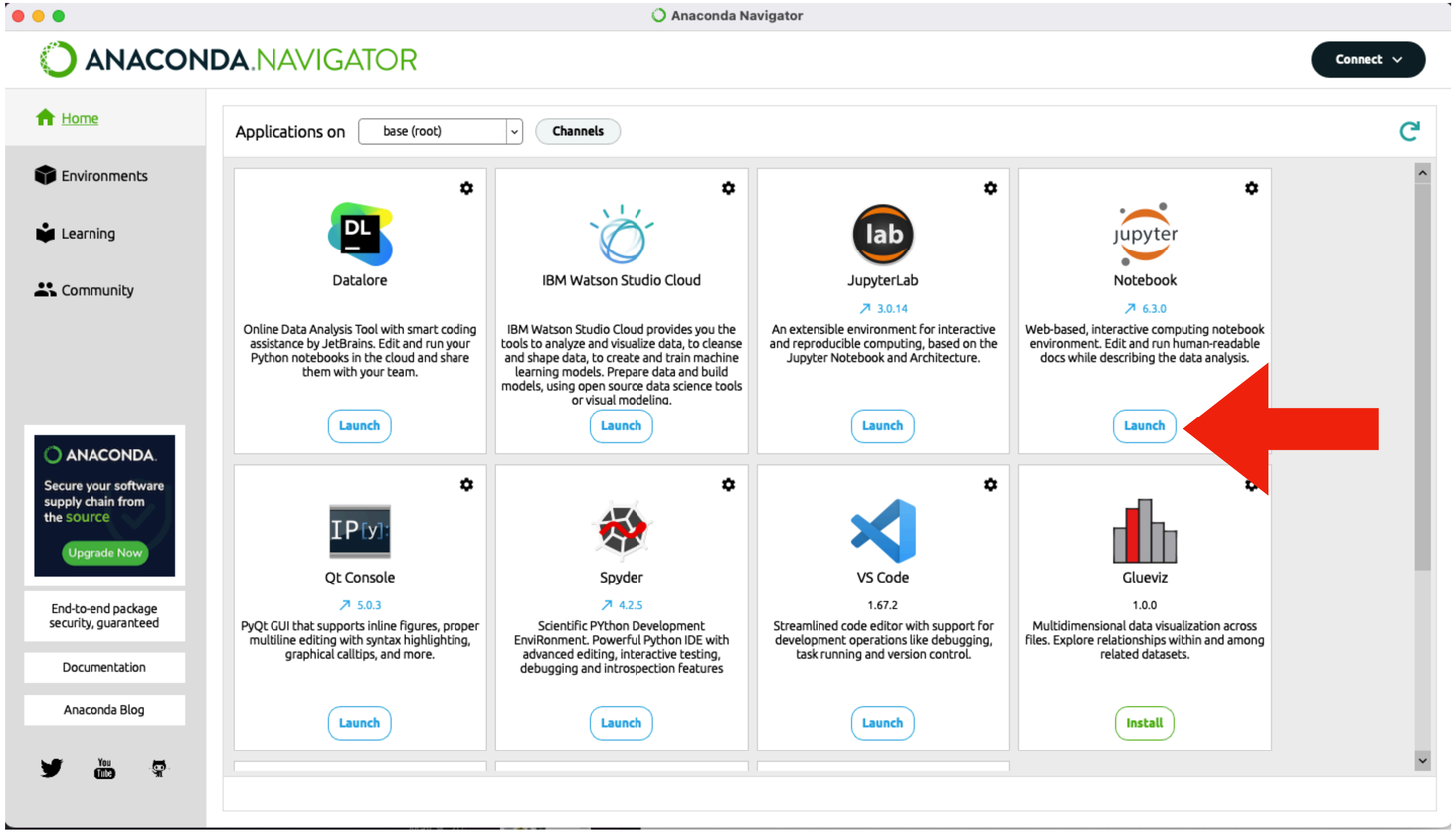
Now that you have anaconda installed, let's start by loading a Jupyter notebook!

On Mac:

Use the Finder to navigate to the applications folder
Double click to open the Anaconda-Navigator

On Windows:

Use the start menu to find the anaconda navigator
Double click to open the Anaconda-Navigator



So what is a Jupyter notebook?

It is a notebook for code, comments and plots that runs within a browser window.

Even though it is in a browser, it is not on the internet. It is running locally on your computer, and only you see your notebook

You can use the file navigator to find a spot where you want to create your notebook

Then click New -> Python 3

Now for our final exercise… let's get ready for the next class.

* Close your Jupyter notebook and the Anaconda navigator
* Download next week's notebook from here: https://reynolds-lab.net/python-workshop-2024/
* Re-open the Anaconda navigator and open next week's notebook