

NeuroNex



Data Joint

Workshop 2018 Day 1

Sponsored by



NeuroNex Innovation Award
#1707359

Presented by
Edgar Y. Walker
April 19, 2018



Session 0: Getting Connected

Getting setup for the workshop

- Visit <https://datajoint.io/workshop>
- Sign up with an email address and workshop code*
- Be sure to join Slack group for online help during the workshop!
- You will need a GitHub account to access the workshop JupyterHub
- You will receive an email with database username and password

*workshop codes are released only at workshops



Session 1:
Getting started with DataJoint

Session 1 Goals

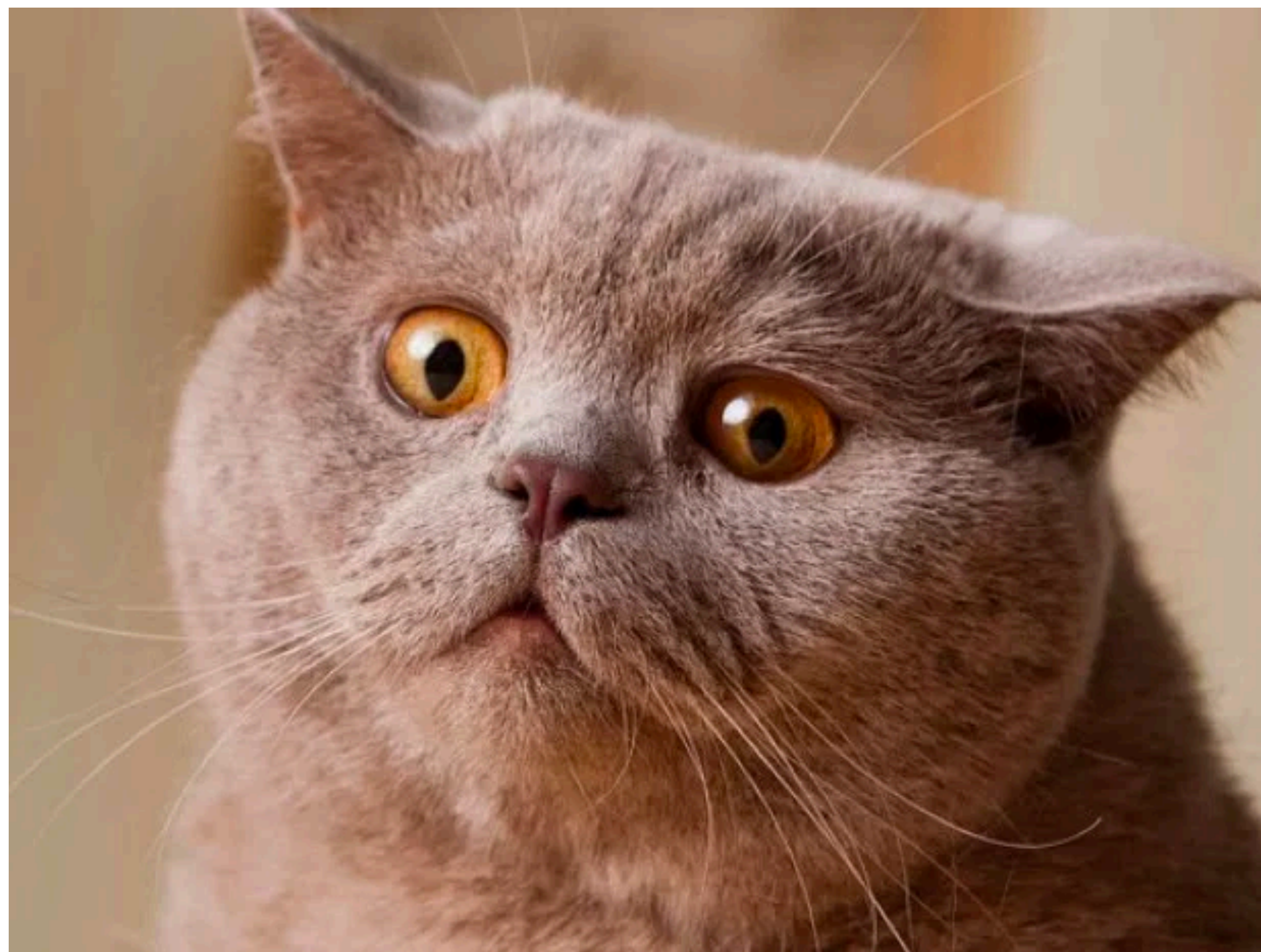
1. Learn what a pipeline is
2. Start designing our first data pipeline based on experiment requirements
3. Create the data pipeline in DataJoint
4. Insert some data into the pipeline
5. Perform basic queries to explore and fetch the data from the pipeline

What is a data pipeline?

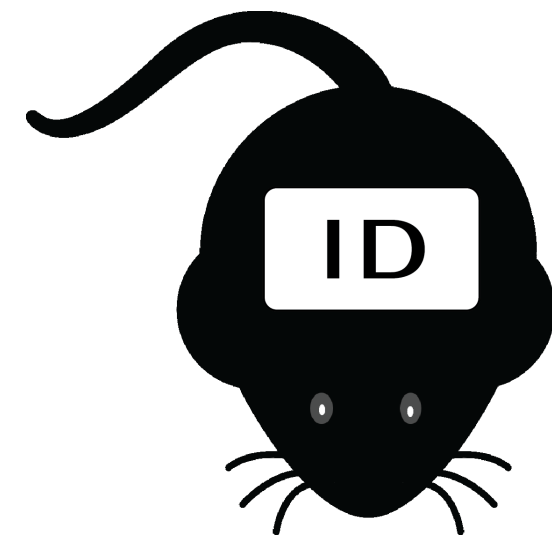
“A data pipeline is a sequence of steps (more generally a directed acyclic graph) with integrated storage at each step. These steps may be thought of as nodes in a graph”

~ from DataJoint documentation (<https://docs.datajoint.io>)

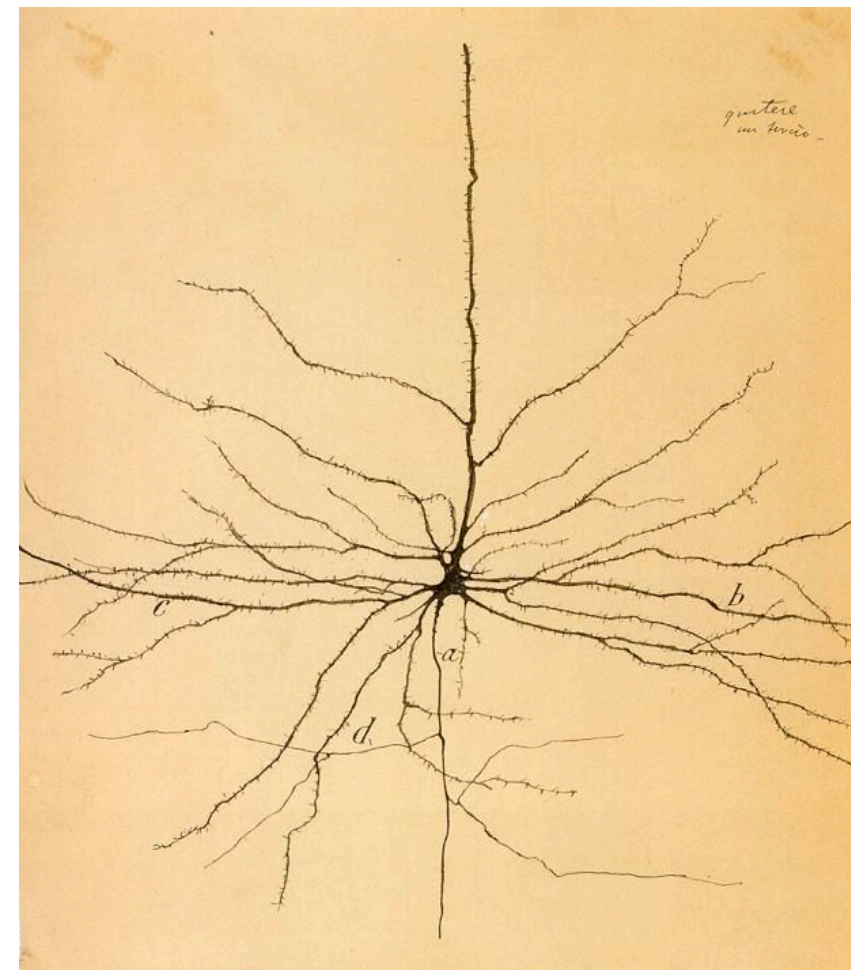
...?



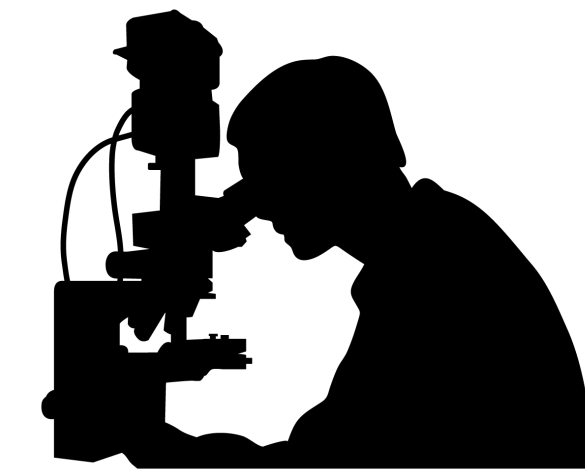
Data pipeline are about “things” in your experiment!



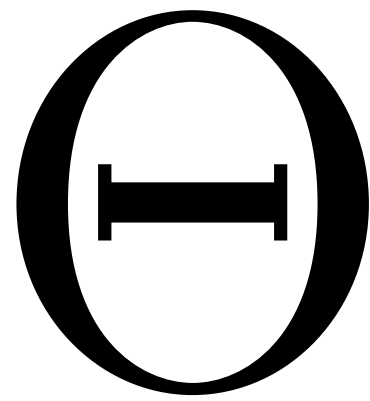
mouse



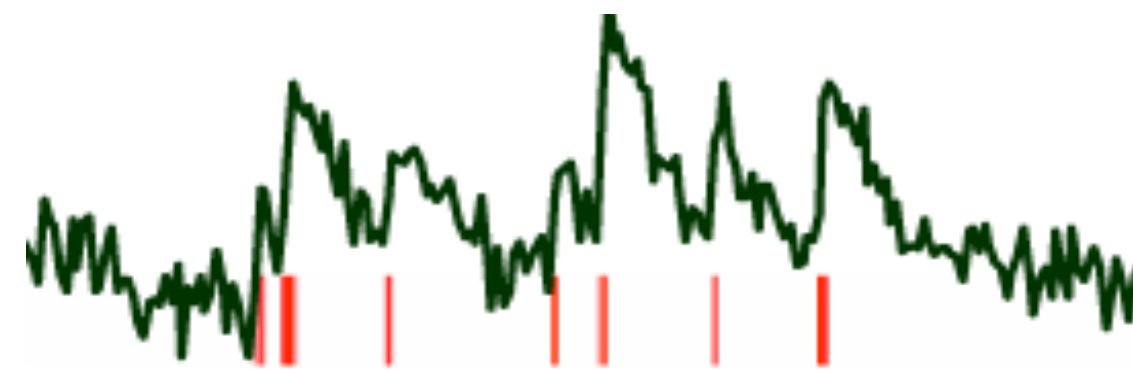
neuron



experimental
session



parameter

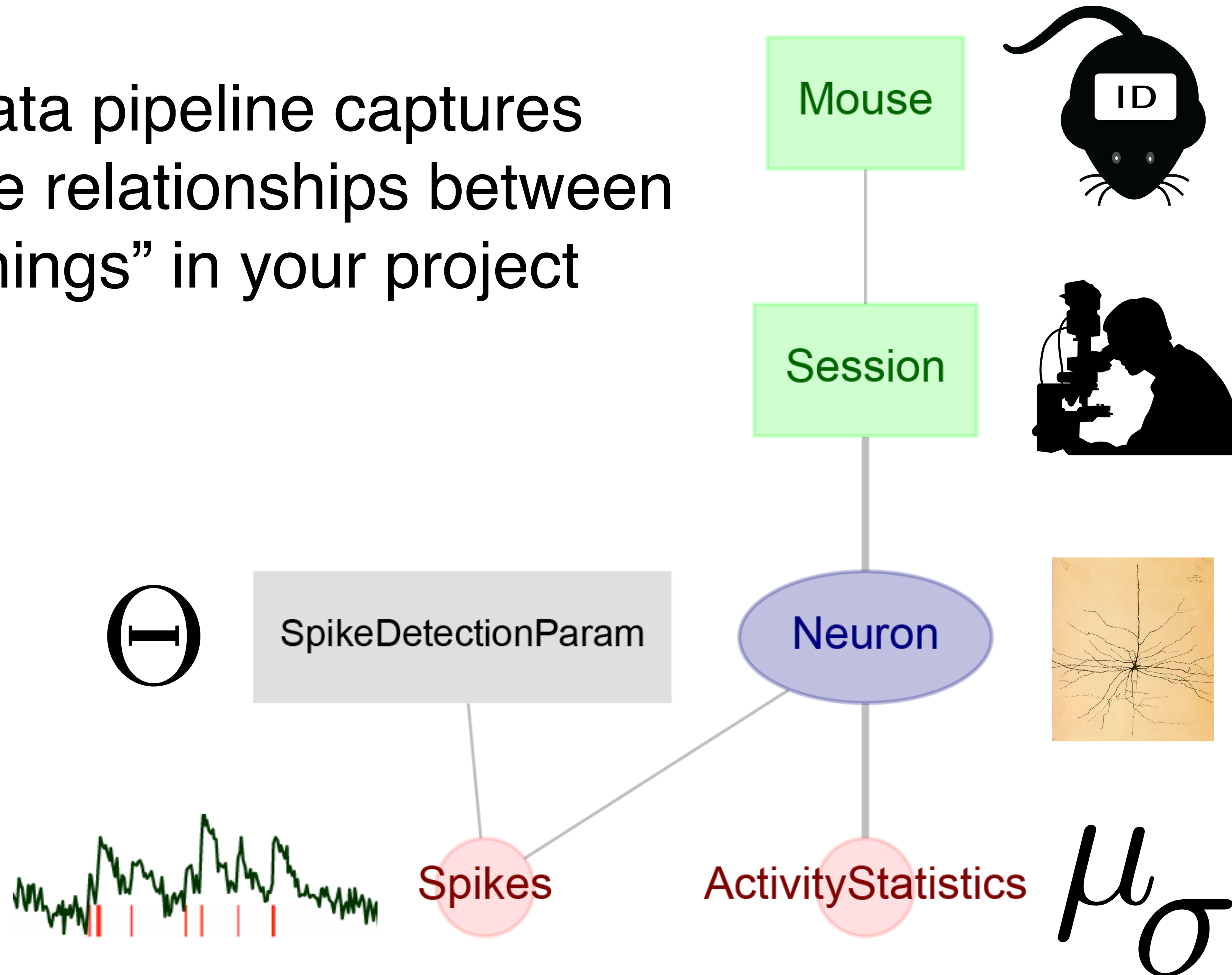


spikes

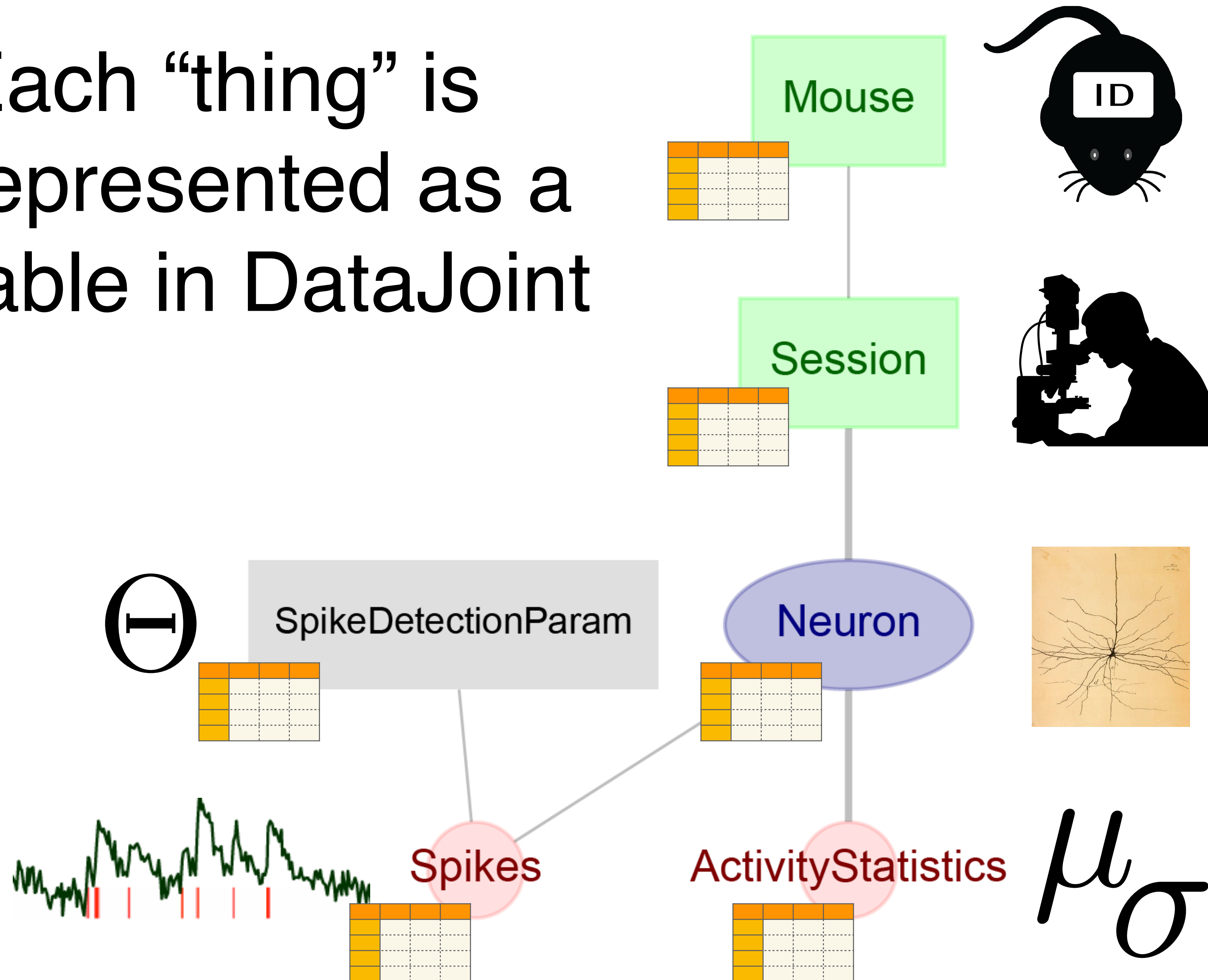


statistics

Data pipeline captures the relationships between “things” in your project



Each “thing” is represented as a table in DataJoint



Building your first pipeline

- Your lab houses many mice, and each mouse is identified by a unique ID. You also want to keep track of information about each mouse such as their date of birth, and gender.
- As a hard working neuroscientist, you perform experiments every day, sometimes working with more than one mouse in a day! However, on any given day, a mouse undergoes at most one recording session.
- For each experimental session, you would like to record what mouse you worked with and when you performed the experiment. You would also like to keep track of other helpful information such as the experimental setup you worked on.
- In each experimental session, you record electrical activity from a single neuron. You use recording equipment that produces separate data files for each neuron you recorded.
- Neuron's activities are recorded as raw traces. Neuron's spikes need to be detected for further analysis to be performed.

Building your first pipeline

- Your lab houses many **mice**, and each mouse is identified by a *unique ID*. You also want to keep track of information about each mouse such as their *date of birth*, and *gender*.
- As a hard working neuroscientist, you perform *experiments* every day, sometimes *working with more than one mouse in a day!* However, on any given day, a *mouse undergoes at most one recording session*.
- For each **experimental session**, you would like to record *what mouse you worked with* and *when you performed the experiment*. You would also like to keep track of other helpful information such as the *experimental setup* you worked on.
- In each **experimental session**, you record *electrical activity* from a single **neuron**. You use recording equipment that produces *separate data files* for each neuron you recorded.
- Neuron's activities are recorded as raw traces. **Neuron's spikes** needs to be detected for further analysis to be performed.

“Things” in our project

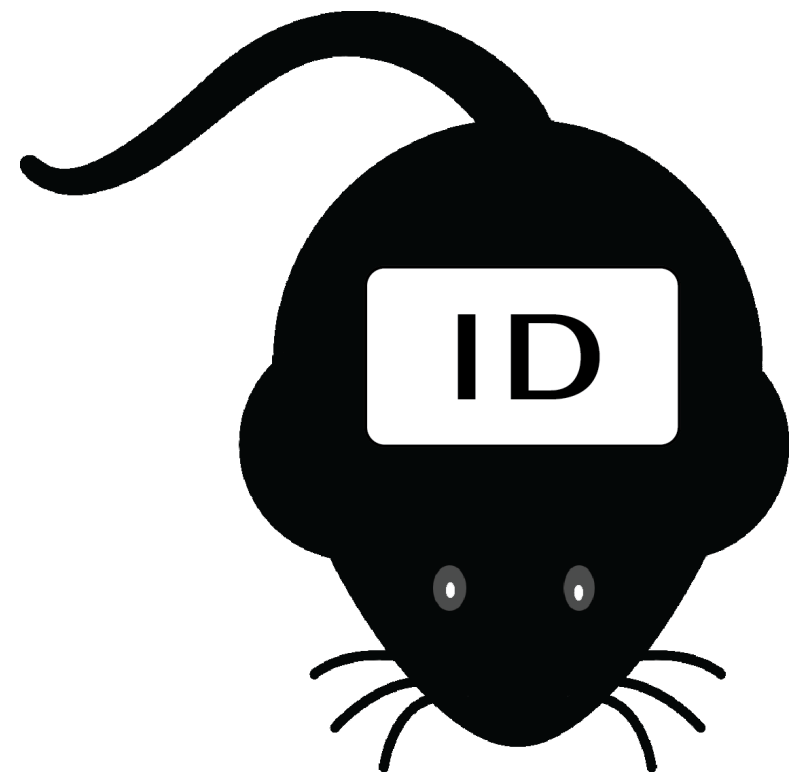
- Mouse
- Experimental session
- Neuron
- Spikes

“Things” in our project

- **Mouse**
- Experimental session
- Neuron
- Spikes

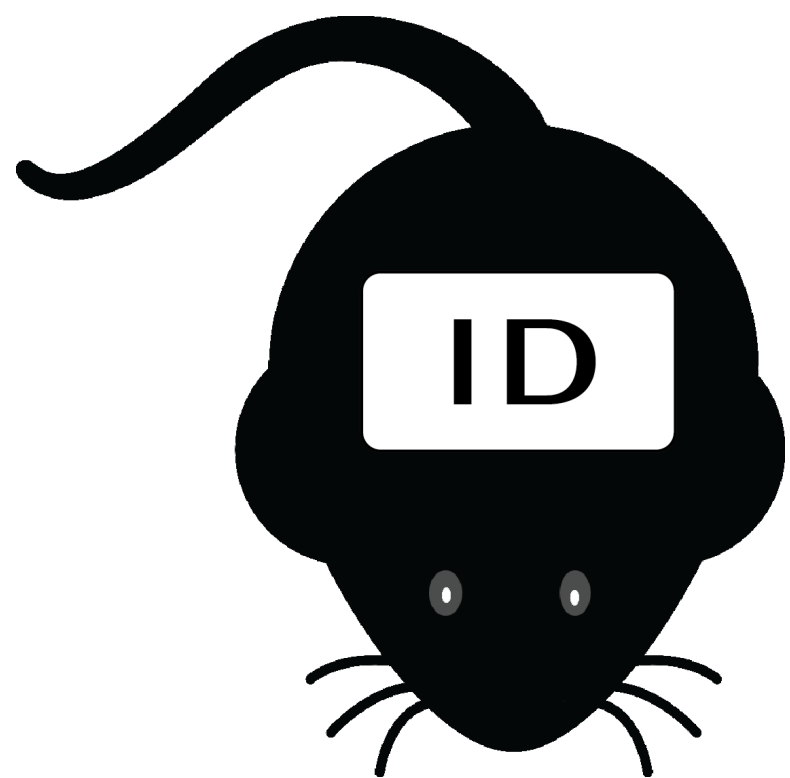
Representing a mouse

What would uniquely
identify a mouse?



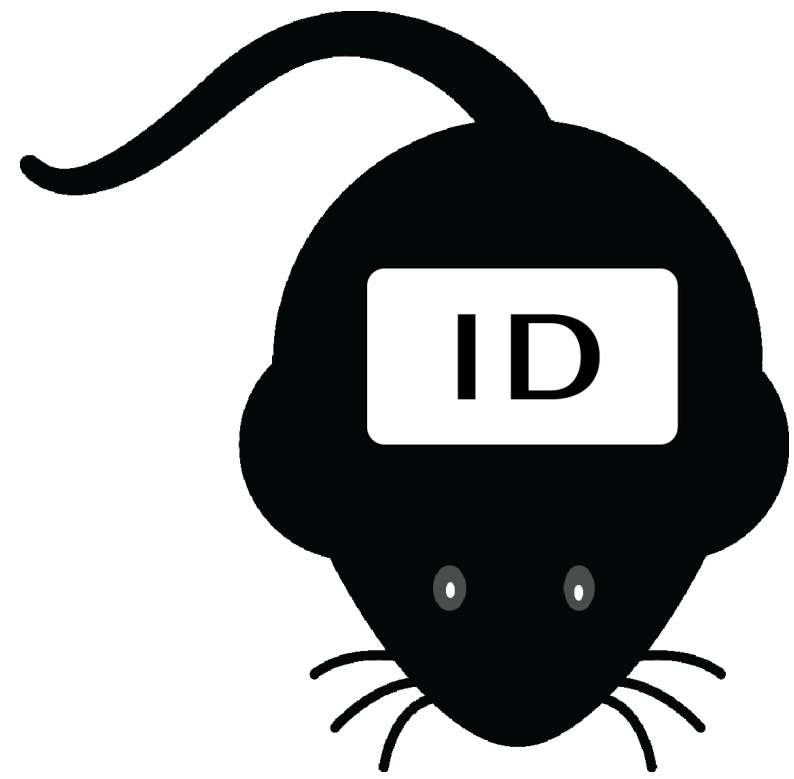
Representing a mouse

mouse_id!



mouse_id*		
12123		
15302		
1243		

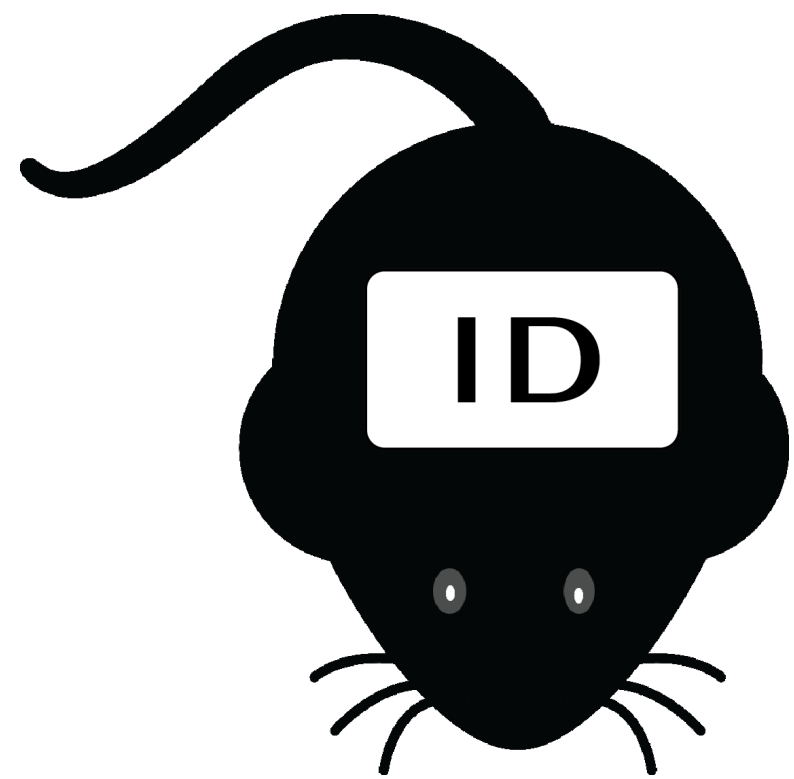
Representing a mouse



This is the **primary key**

mouse_id*		
12123		
15302		
1243		

Representing a mouse

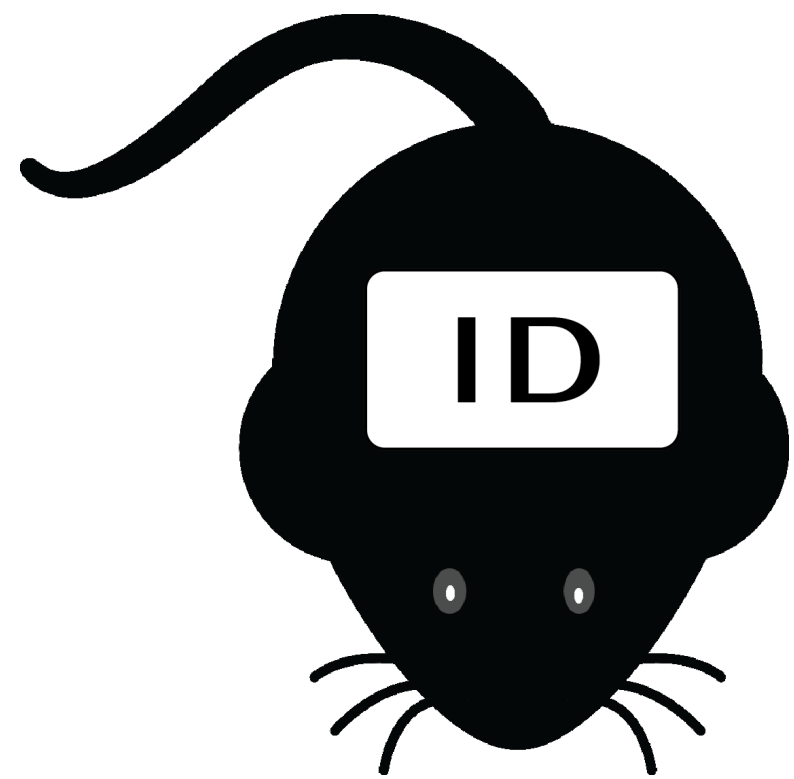


mouse_id*		
12123		
15302		
1243		

Each row is **a mouse**

Representing a mouse

Adding other attributes
(columns) **about each
mouse**



mouse_id*	dob	sex
12123	2017-01-12	M
15302	2018-01-01	F
1243	2016-03-05	Unknown

Let's now go build the
pipeline in DataJoint!



Session 2: Imported and Computed tables

Session 2 Goals

1. import neuron activity data from data files into an **Imported table**
2. compute various statistics for each neuron by defining a **Computed table**
3. define a **Lookup table** to store parameters for computation
4. define another **Computed table** to perform spike detection and store the detected spikes
5. automatically trigger computations for all missing entries with **populate**



Session 3:
Design patterns and
advanced queries

Session 3 Goals

1. highlight various design patterns found in our data pipeline
2. exercise writing more complex DataJoint queries

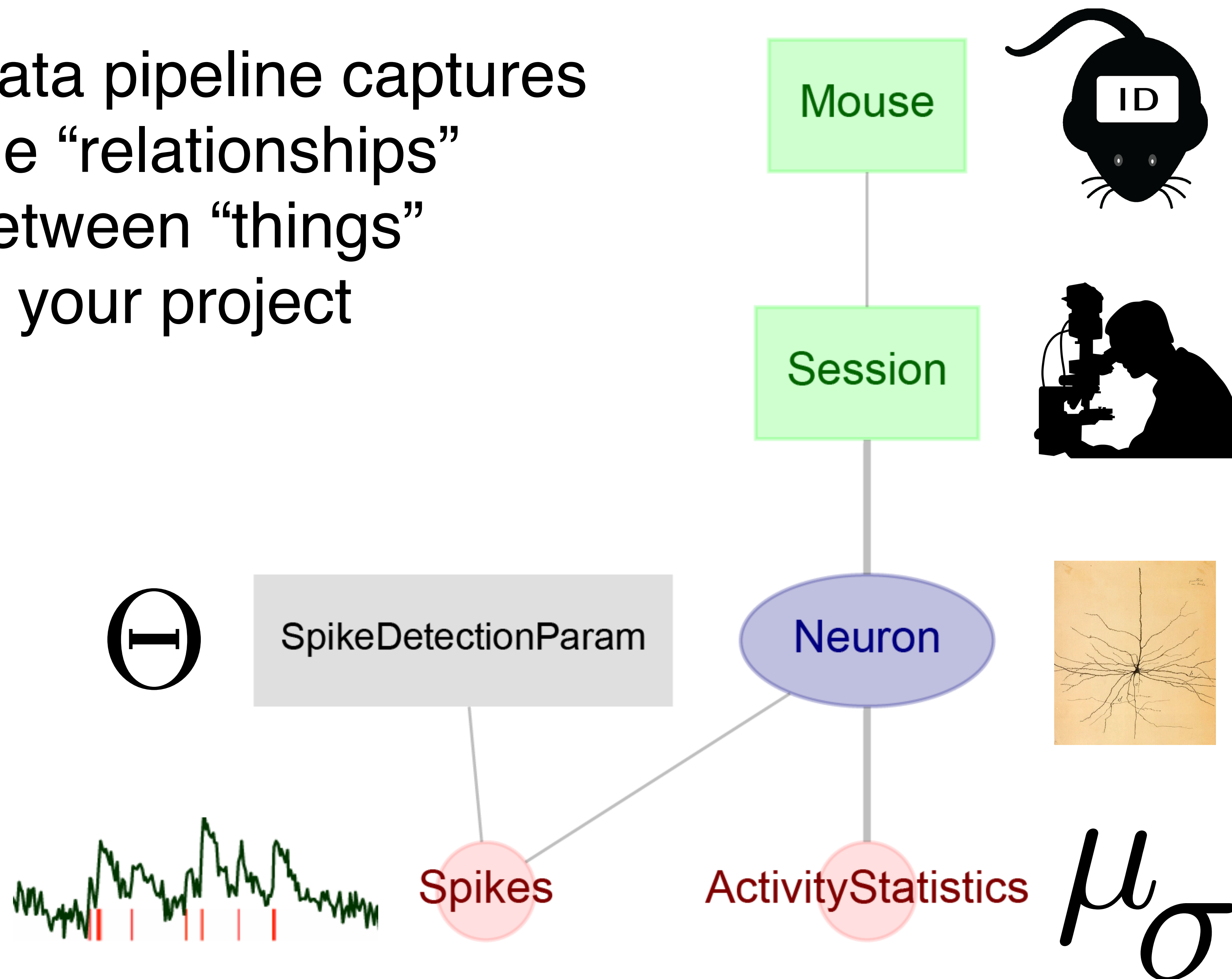


Recap of Day 1

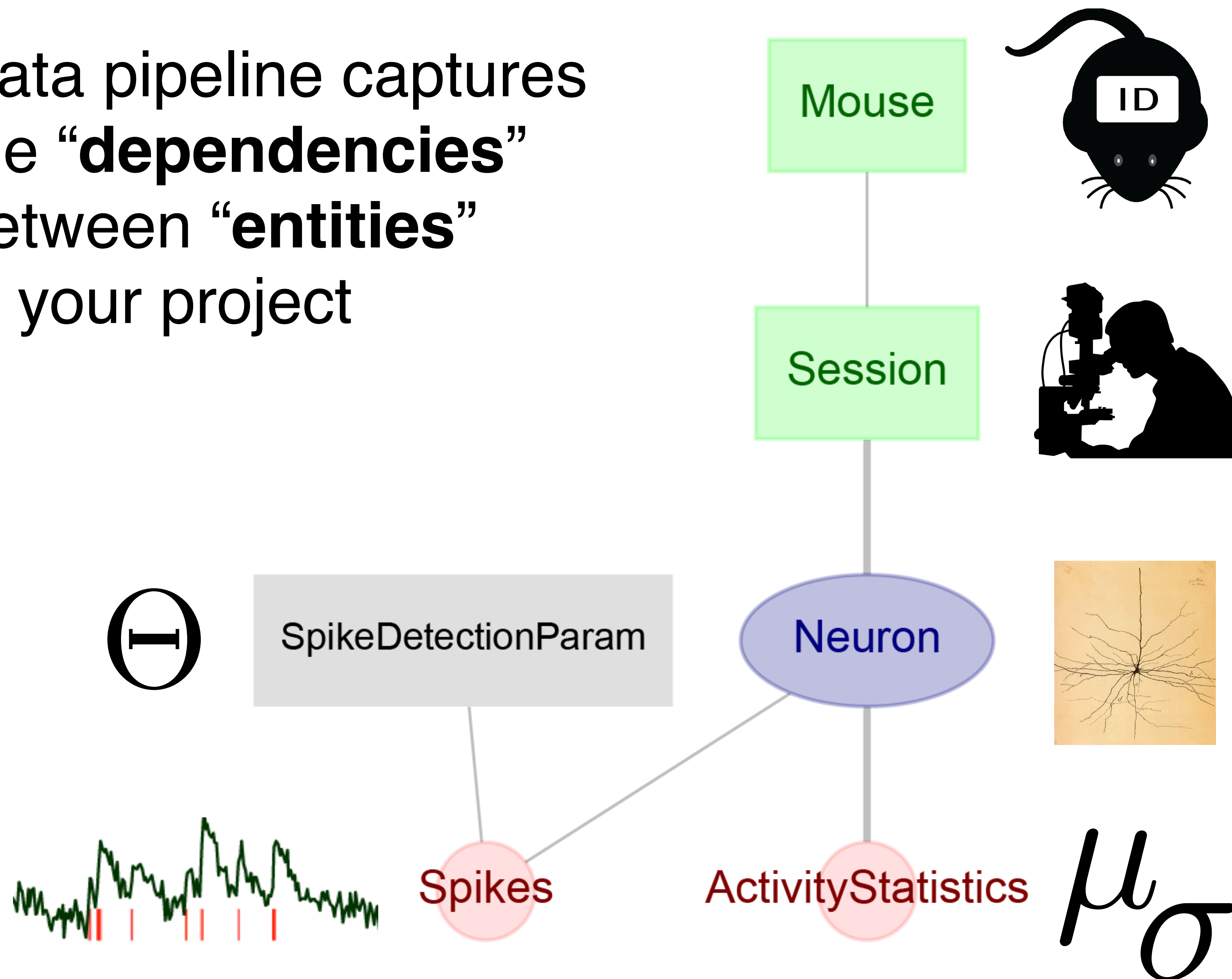
We covered a lot!

- Designed our first data pipeline
- Learned to insert, query and fetch data
- Learned to define computations as tables in data pipeline
 - Computing “statistics”
 - Detected “spikes”
- Learned to use `make` and `populate` logic to automatically “populate” tables
- Studies common design patterns in data pipeline

Data pipeline captures the “relationships” between “things” in your project



Data pipeline captures the “**dependencies**” between “**entities**” in your project



What are we covering tomorrow?

- You will work in a team to:
 - Session 4: Practice designing a pipeline from scratch based on project requirements
 - Session 5: Extend an existing pipeline with a table defining a new analysis
- Session 6: Learn about best practices in data organization and sharing, and survey various tools/technologies to help you achieve the goals

Additional learning resources



- Visit <https://datajoint.io> for more information about DataJoint - the free open-source libraries for Python 3 and MATLAB
- Documentation and tutorials are available at <https://docs.datajoint.io> and <https://tutorials.datajoint.io>
- DataJoint Slack group is an excellent place to interact with developers and other users.
- More learning resources are up and coming!

That's it for Day 1!

Thank you for attending!

You will receive an email with a survey - we would really like to hear your feedback on the workshop day 1! Tell us what you liked and what could be improved!