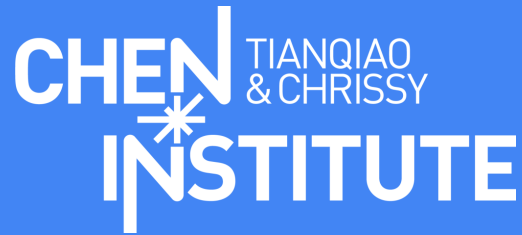


Chen Data Science & AI for Neuroscience Summer School



Caltech

Data Processing Principles

Sabera Talukder

What are your data processing principles?

- Visualize Your Data → Intuition Development

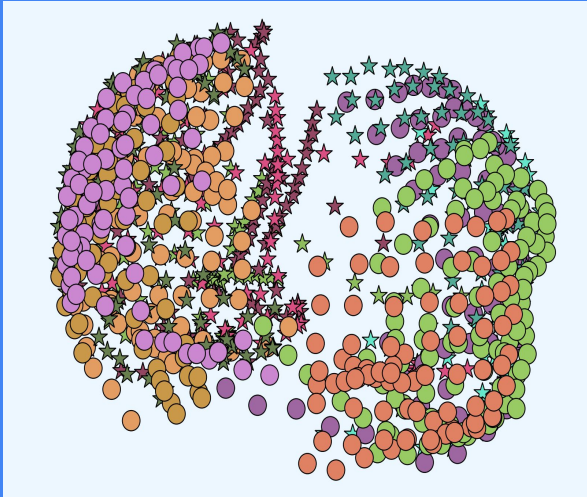
- Visualize Your Data → Intuition Development
- Signal Extraction (aka Denoising)

- Visualize Your Data → Intuition Development
- Signal Extraction (aka Denoising)
- Dataset Augmentation

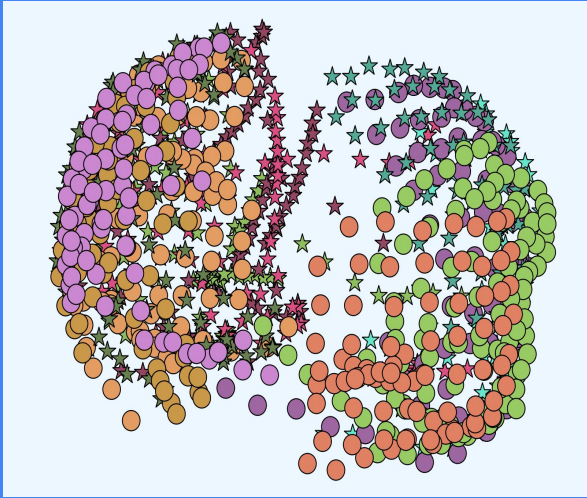
- Visualize Your Data → Intuition Development
- Signal Extraction (aka Denoising)
- Dataset Augmentation
- Normalization / Standardization

- Visualize Your Data → Intuition Development
- Signal Extraction (aka Denoising)
- Dataset Augmentation
- Normalization / Standardization
- Train / Validation / Test Splits

Visualize Your Data

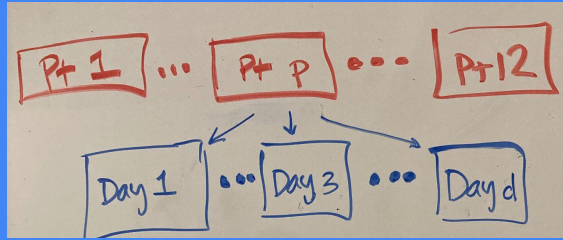
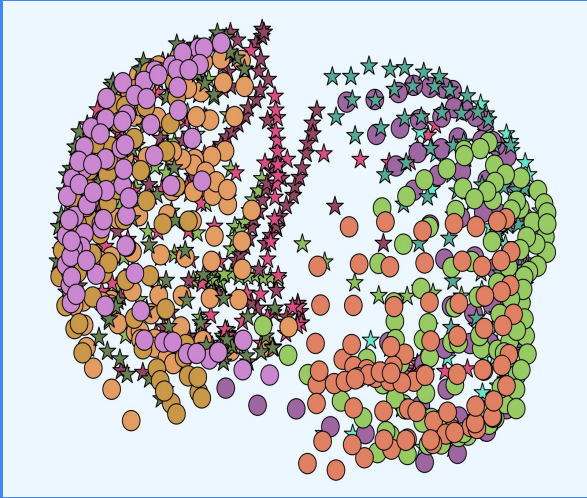


Visualize Your Data

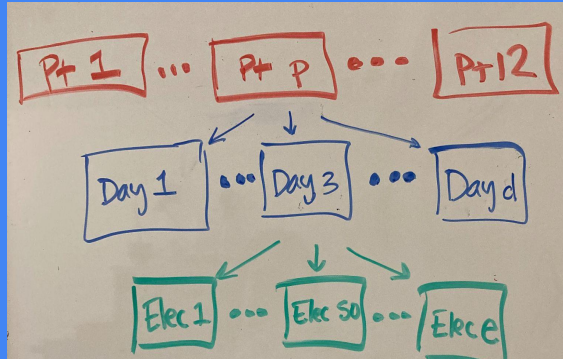
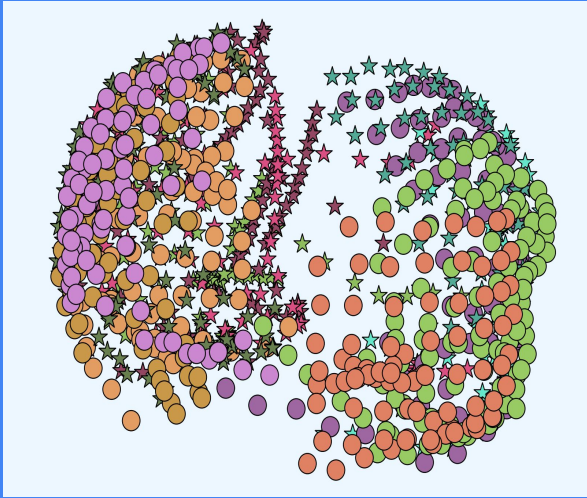


P_1 ... P_p ... P_{12}

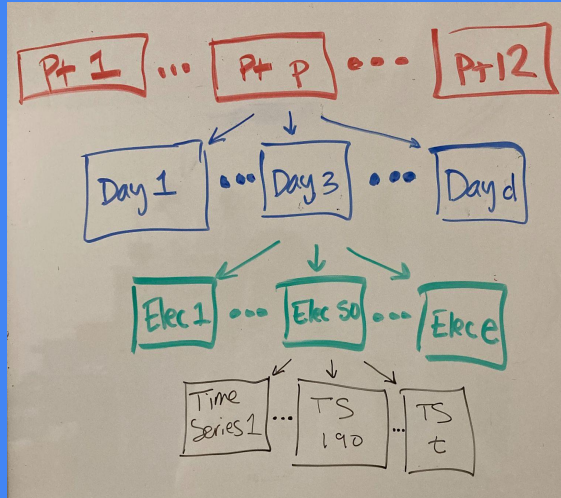
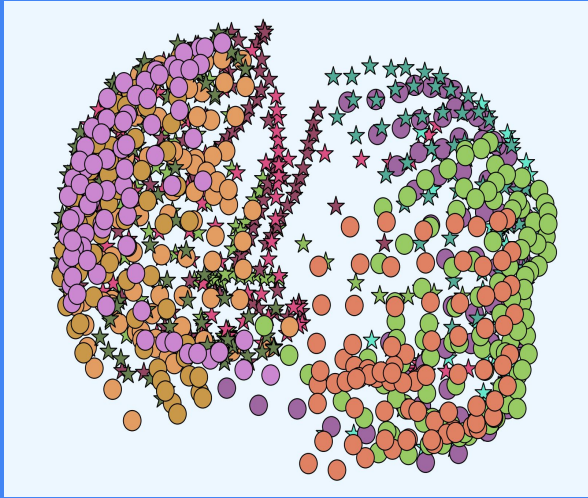
Visualize Your Data



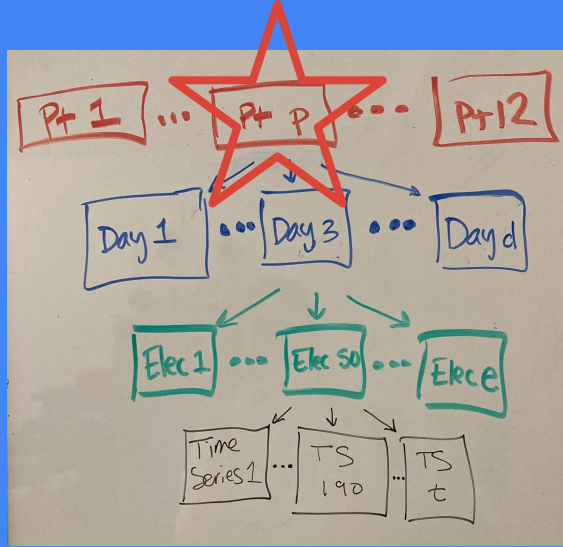
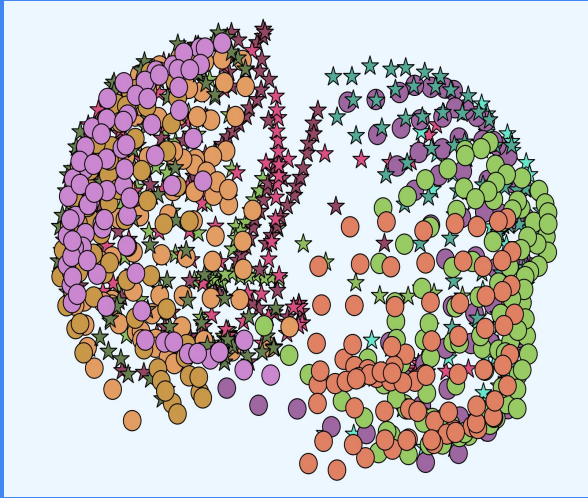
Visualize Your Data



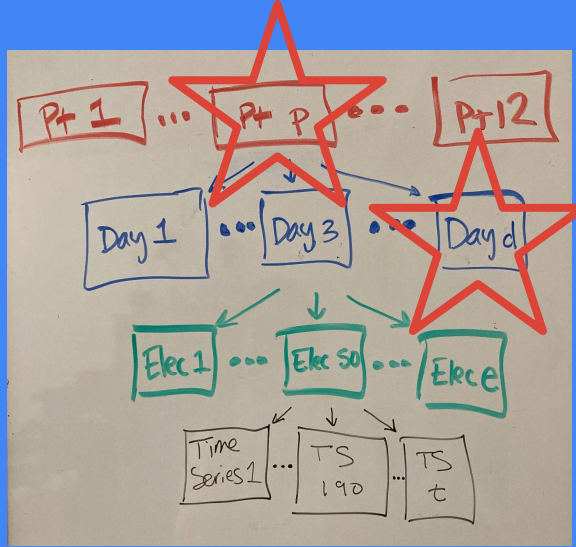
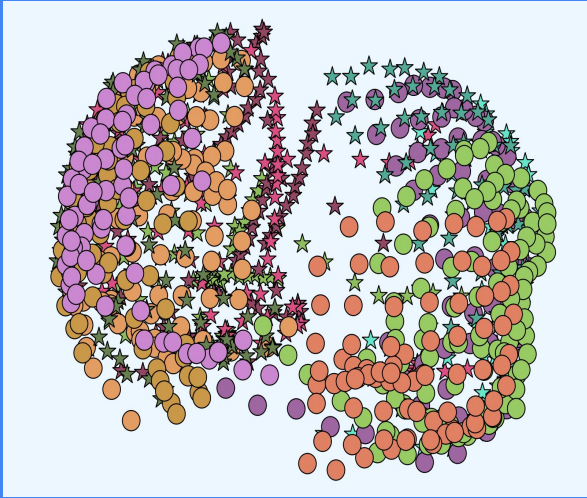
Visualize Your Data



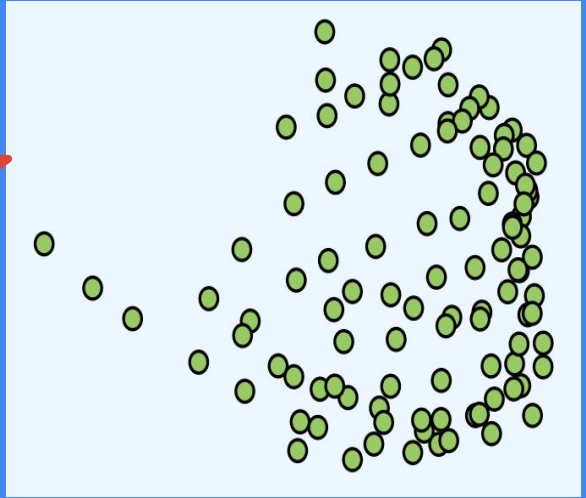
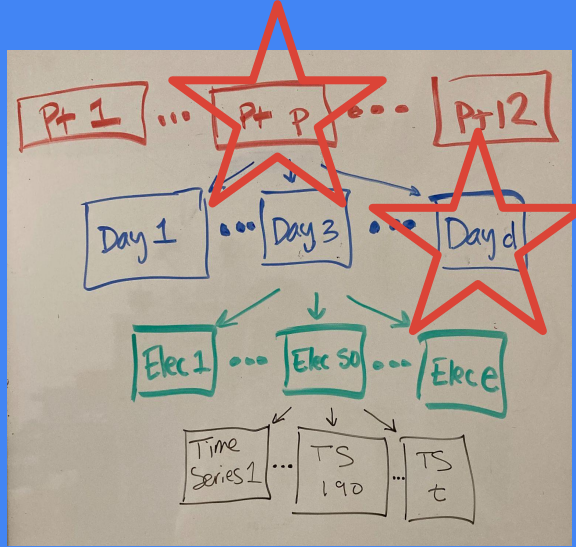
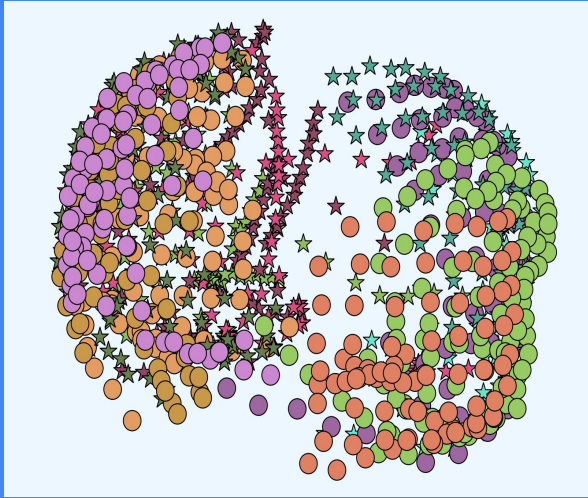
Visualize Your Data



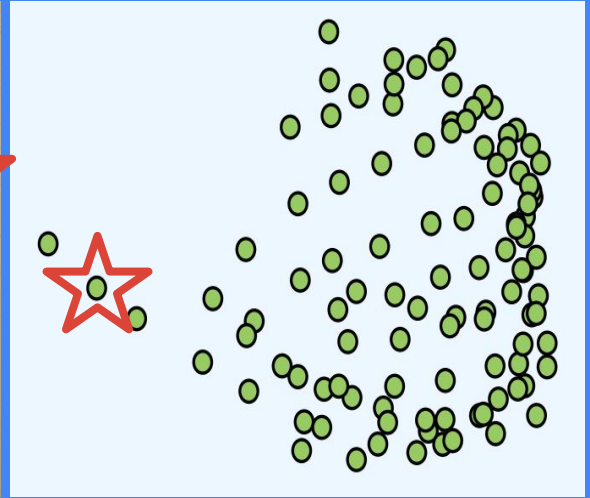
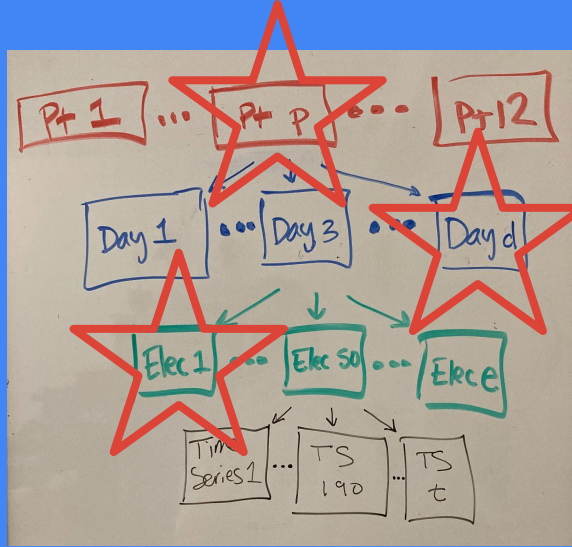
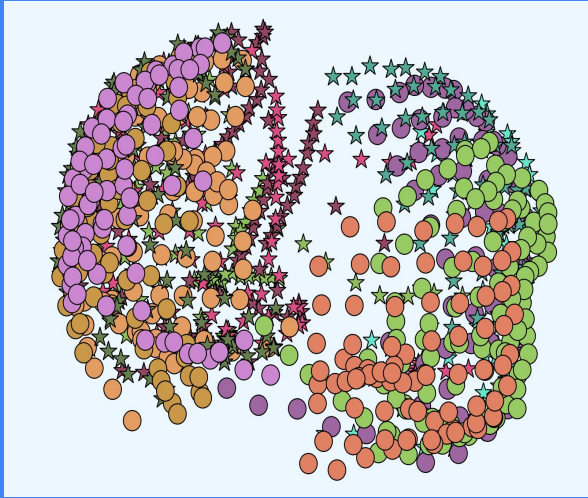
Visualize Your Data



Visualize Your Data



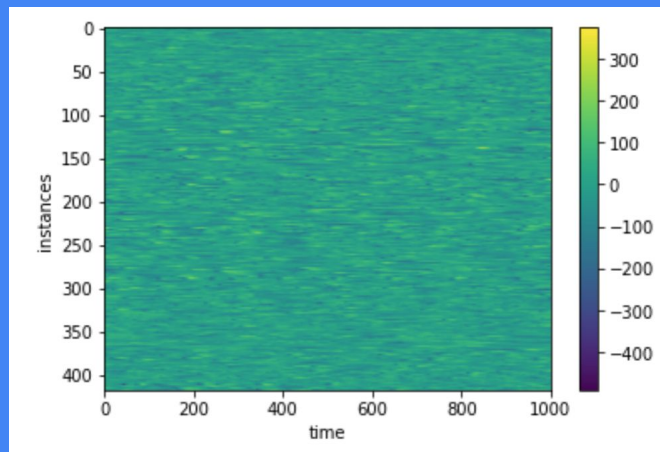
Visualize Your Data



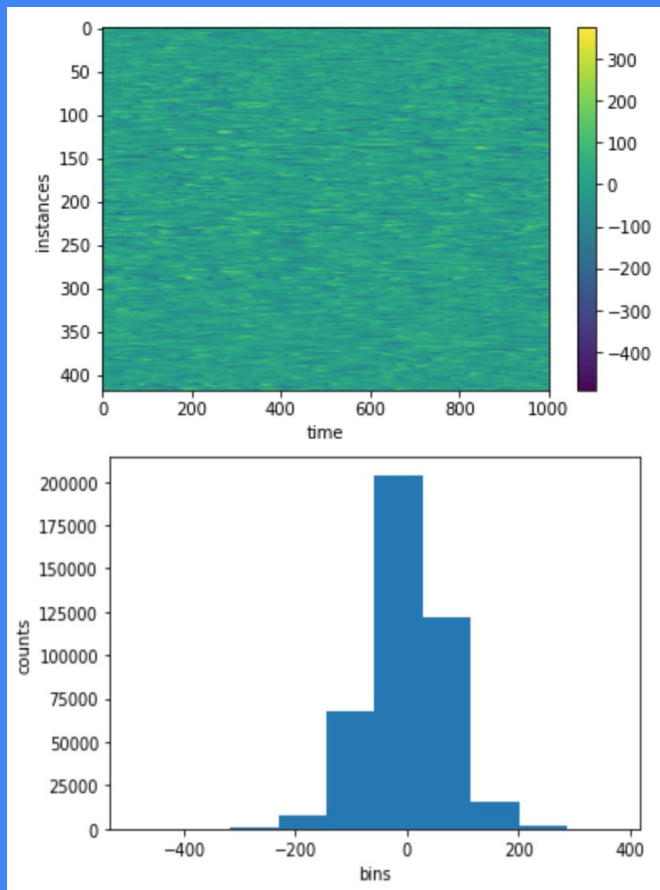
Visualize Your Data



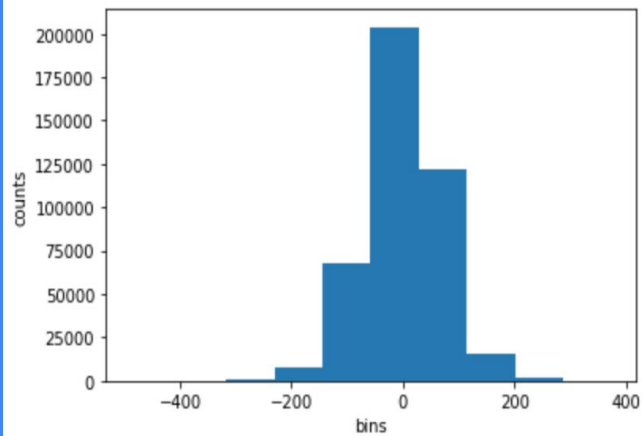
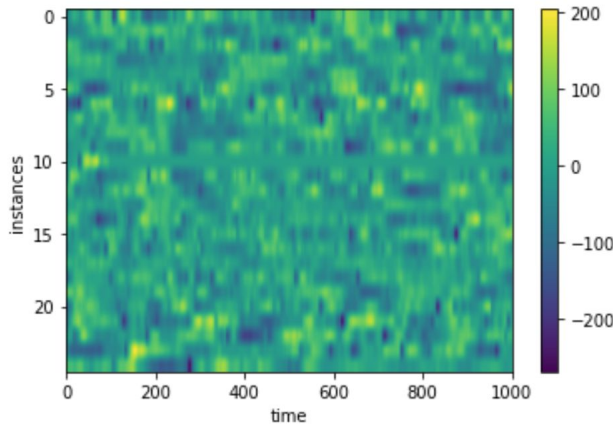
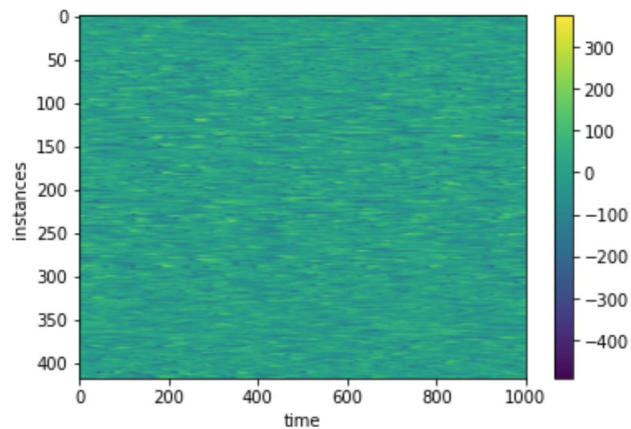
Visualize Your Data



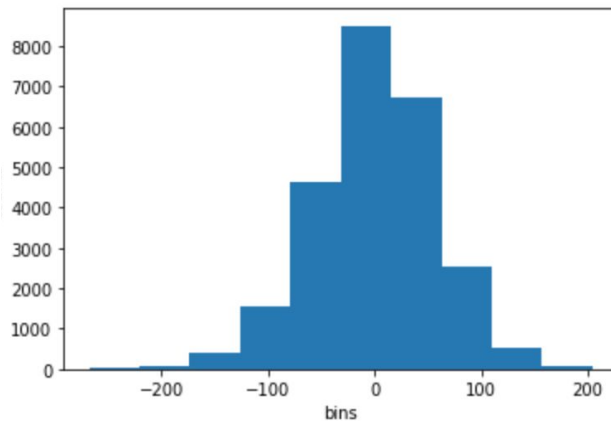
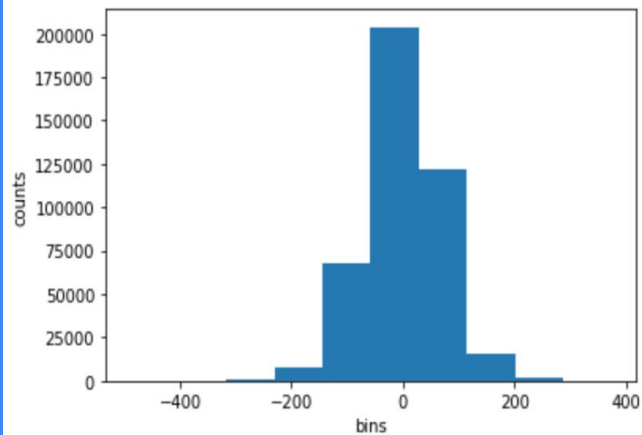
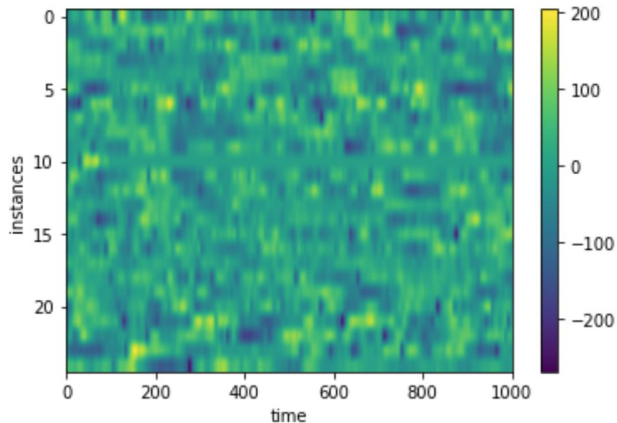
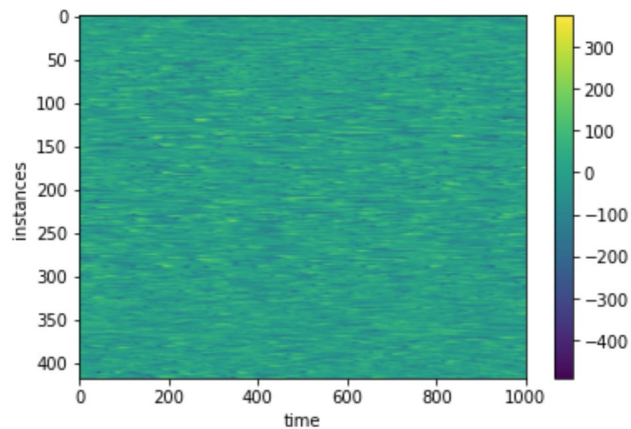
Visualize Your Data



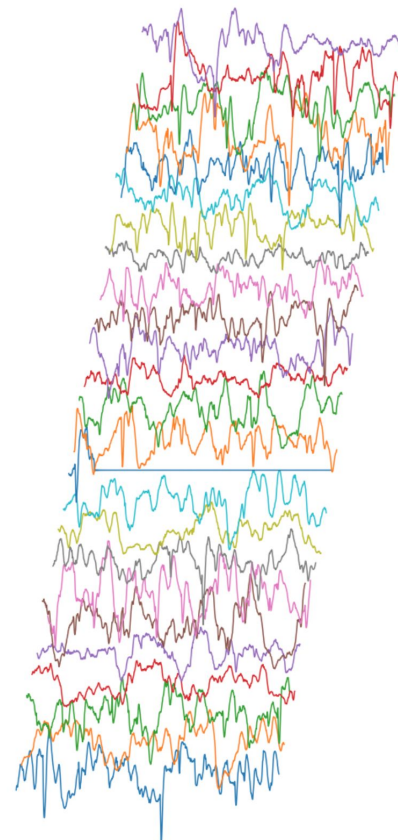
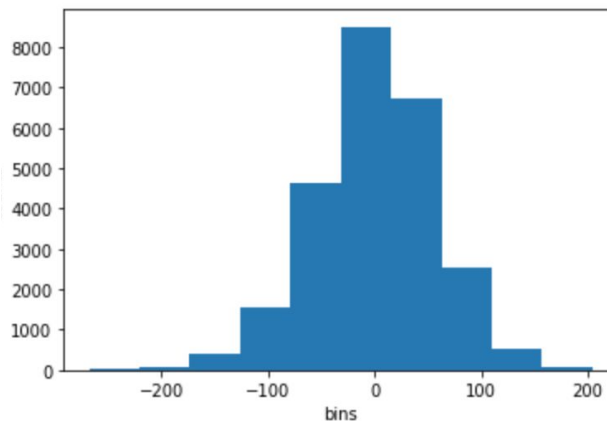
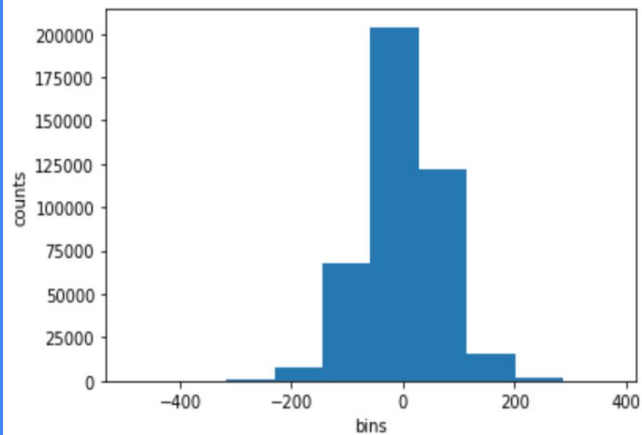
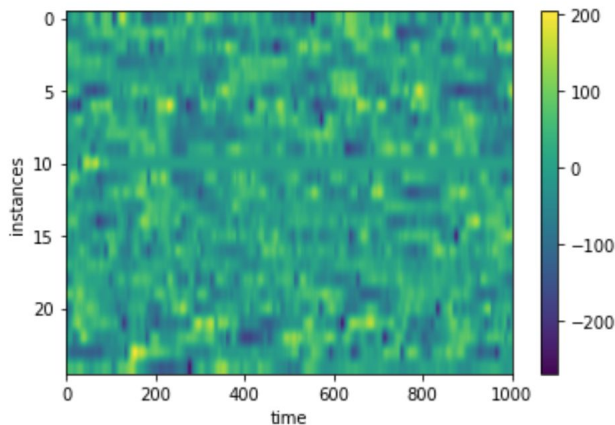
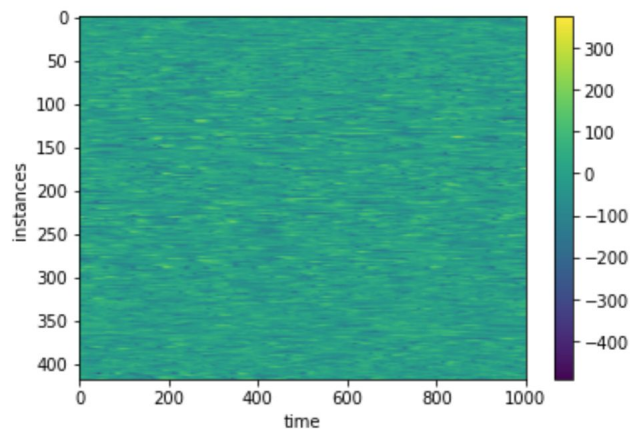
Visualize Your Data



Visualize Your Data



Visualize Your Data



Signal Extraction

Signal Extraction

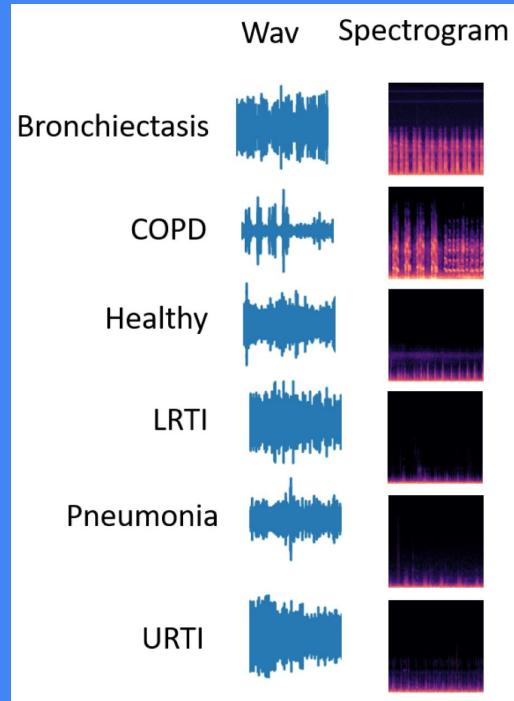
What types of signals might you have in your data?

Signal Extraction

- Lung Artifacts
- Heart Artifacts
- Stimulation Artifacts
- Movement Artifacts
- Electrical Noise
- ...

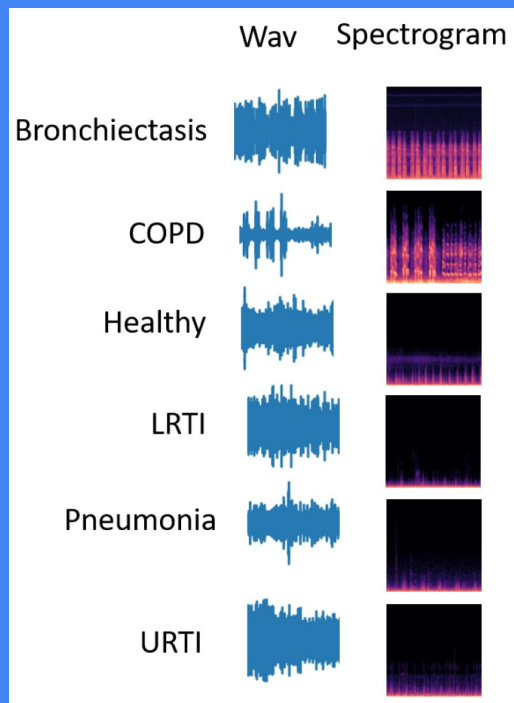
Signal Extraction

Lung Artifacts

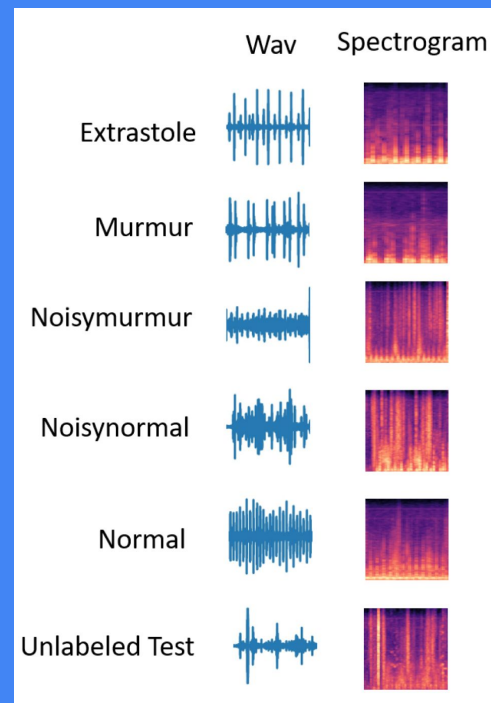


Signal Extraction

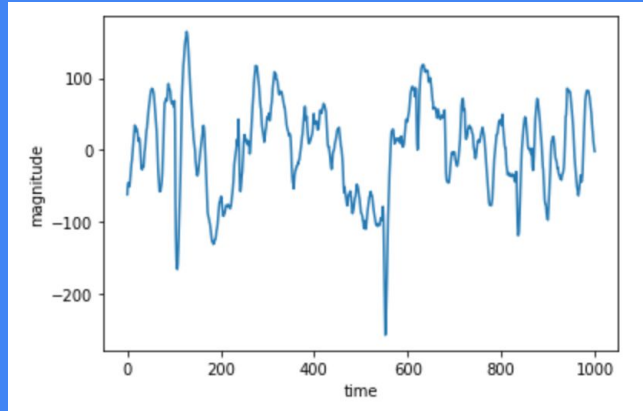
Lung Artifacts



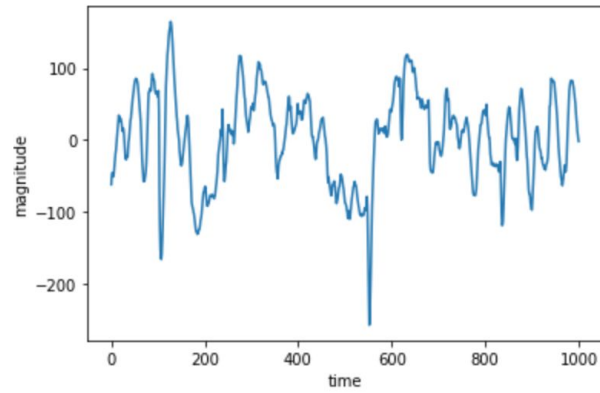
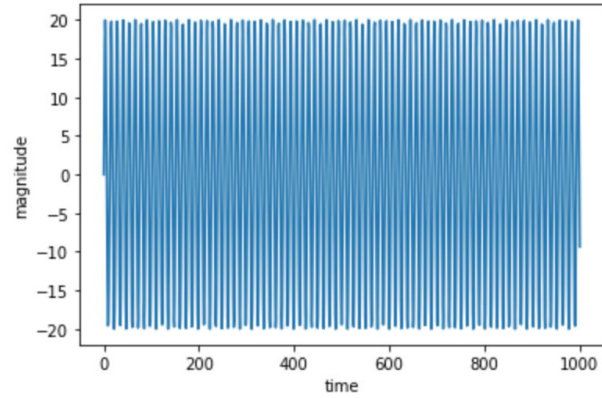
Heart Artifacts



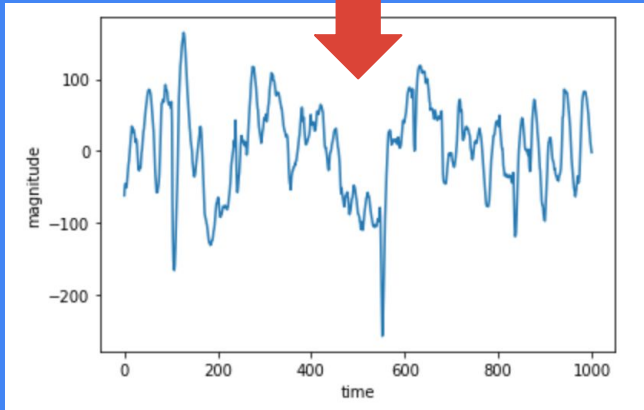
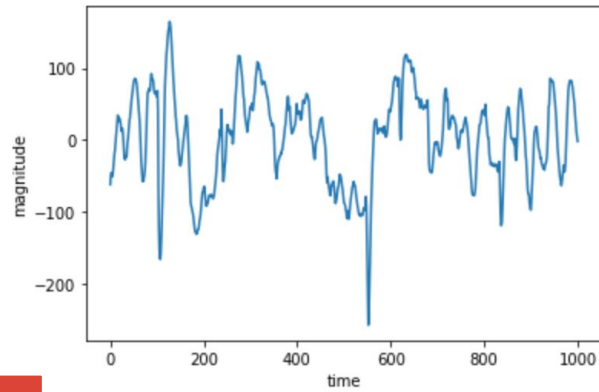
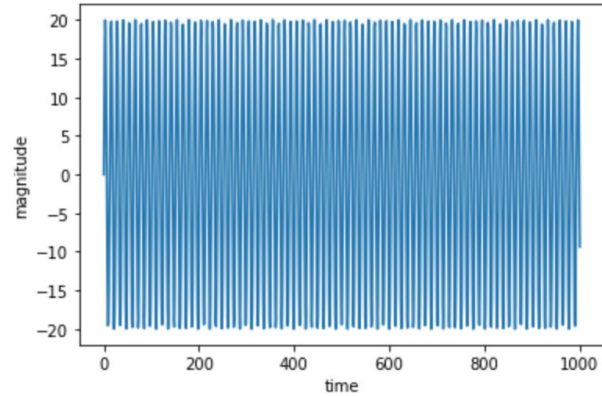
Signal Extraction



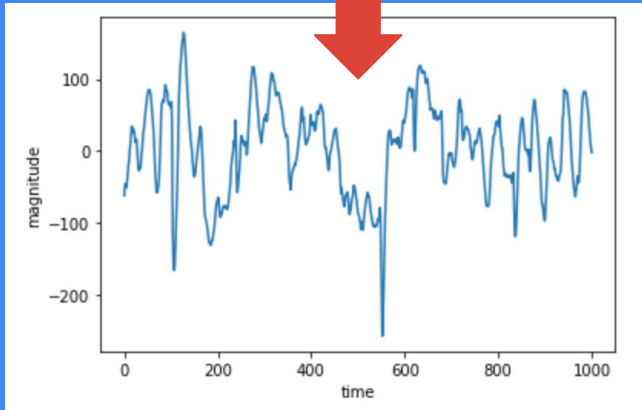
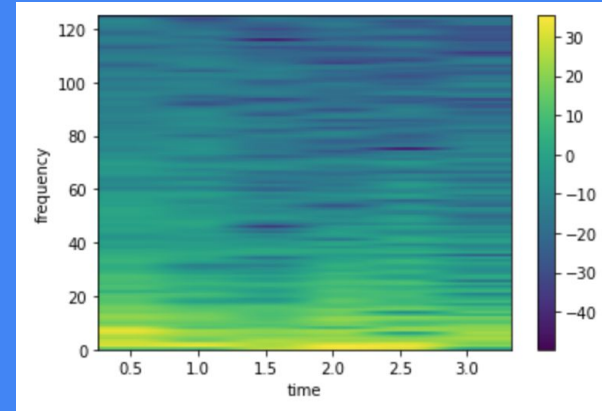
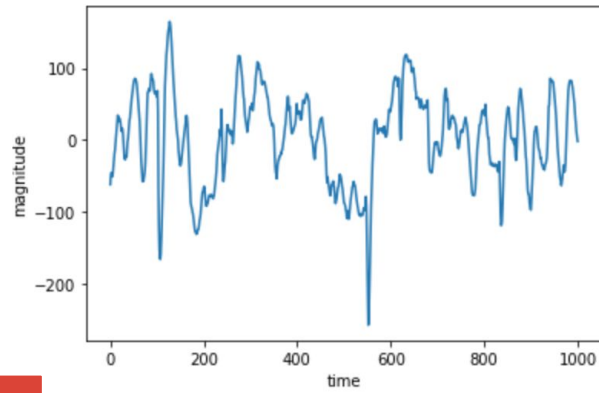
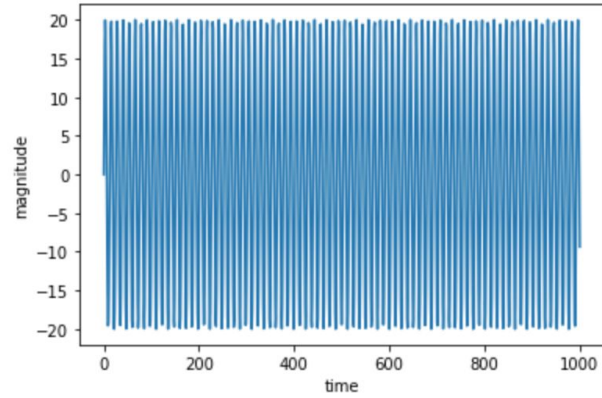
Signal Extraction



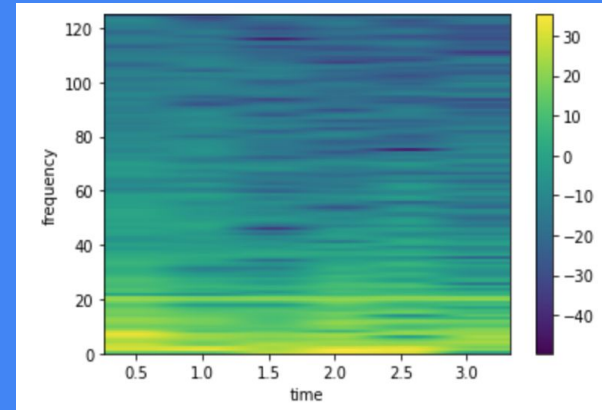
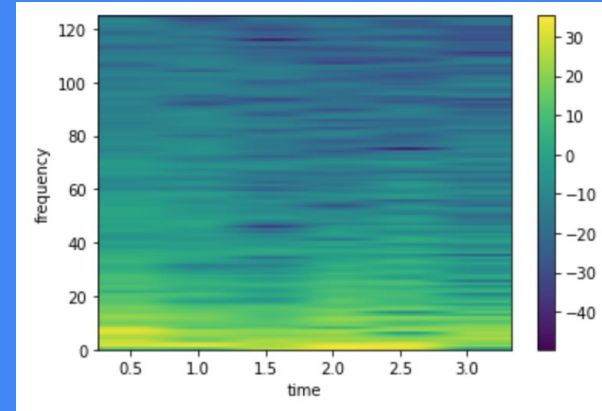
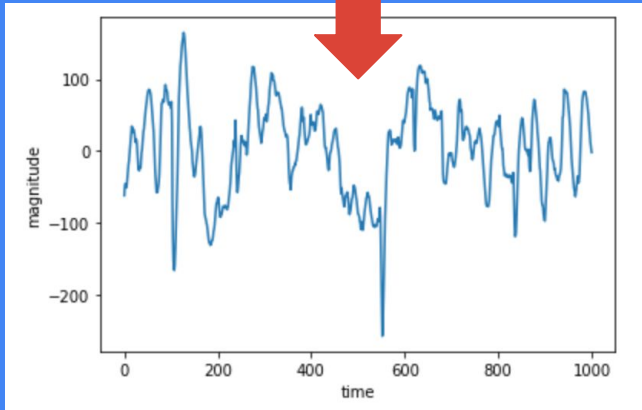
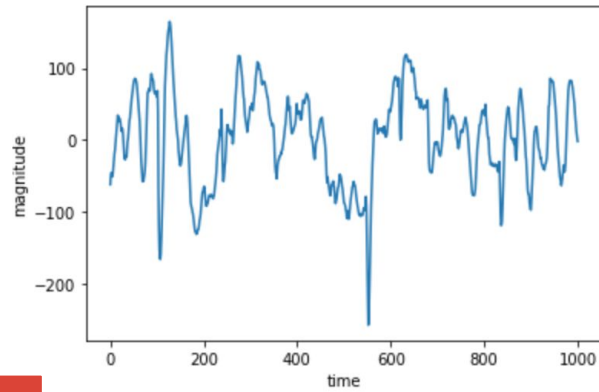
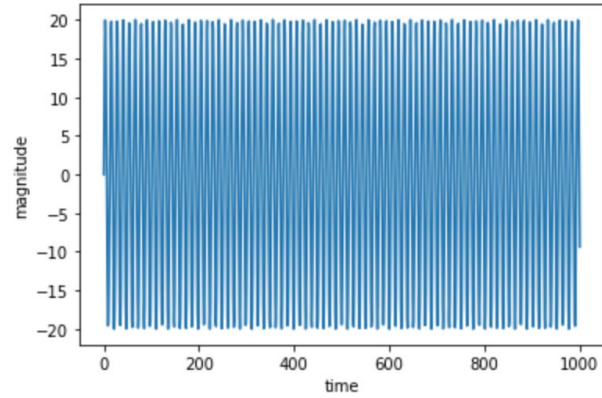
Signal Extraction



Signal Extraction



Signal Extraction



Dataset Augmentation

Dataset Augmentation

What is data augmentation?

Dataset Augmentation

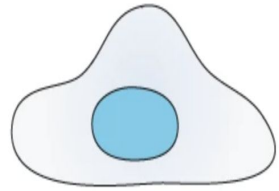
Data augmentation in data analysis are techniques used to **increase the amount of data** by adding slightly modified copies of already existing data or newly created synthetic data from existing data.

Dataset Augmentation

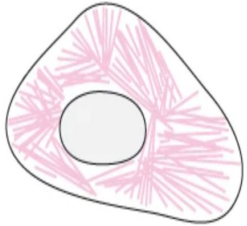
Data augmentation in data analysis are techniques used to **increase the amount of data** by adding slightly modified copies of already existing data or newly created synthetic data from existing data.

It acts as a regularizer and helps reduce overfitting when training a machine learning model.

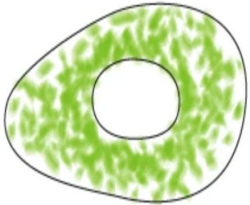
Dataset Augmentation



→ Nuclear

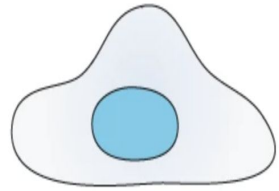


→ Cytoplasmic

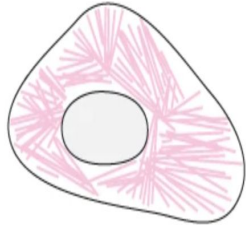


→ Mitochondrial

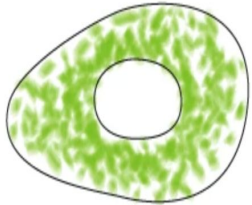
Dataset Augmentation



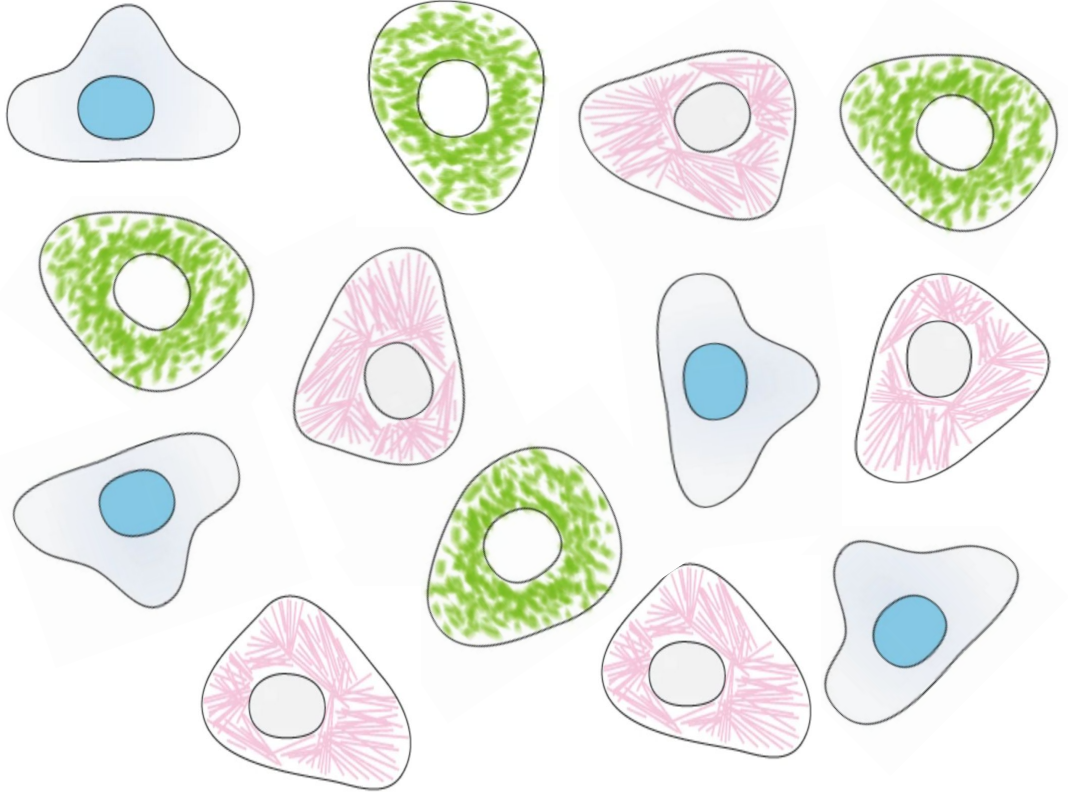
→ Nuclear



→ Cytoplasmic



→ Mitochondrial



Normalization / Standardization

Normalization / Standardization

What is dataset normalization?

Normalization / Standardization

Normalization: Is adjusting values measured on different scales to a notionally common scale... [or] more sophisticated adjustments where the intention is to bring the entire probability distributions of adjusted values into alignment.

Normalization / Standardization

Normalization: Is adjusting values measured on different scales to a notionally common scale... [or] more sophisticated adjustments where the intention is to bring the entire probability distributions of adjusted values into alignment.

Standardization: Is when we subtract the population mean from an individual raw score and then dividing the difference by the population standard deviation (aka Z-scoring).

Normalization / Standardization

Normalization: Is adjusting values measured on different scales to a notionally common scale... [or] more sophisticated adjustments where the intention is to bring the entire probability distributions of adjusted values into alignment.

Standardization: Is when we subtract the population mean from an individual raw score and then dividing the difference by the population standard deviation (aka Z-scoring).

$$z = \frac{x - \mu}{\sigma}$$

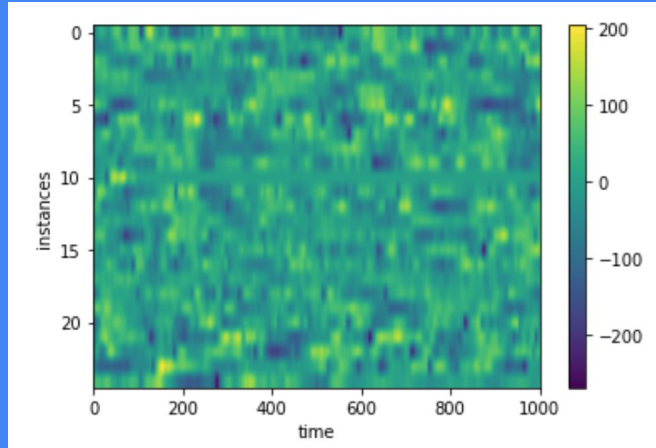
Normalization / Standardization

What does this help correct for in our data?

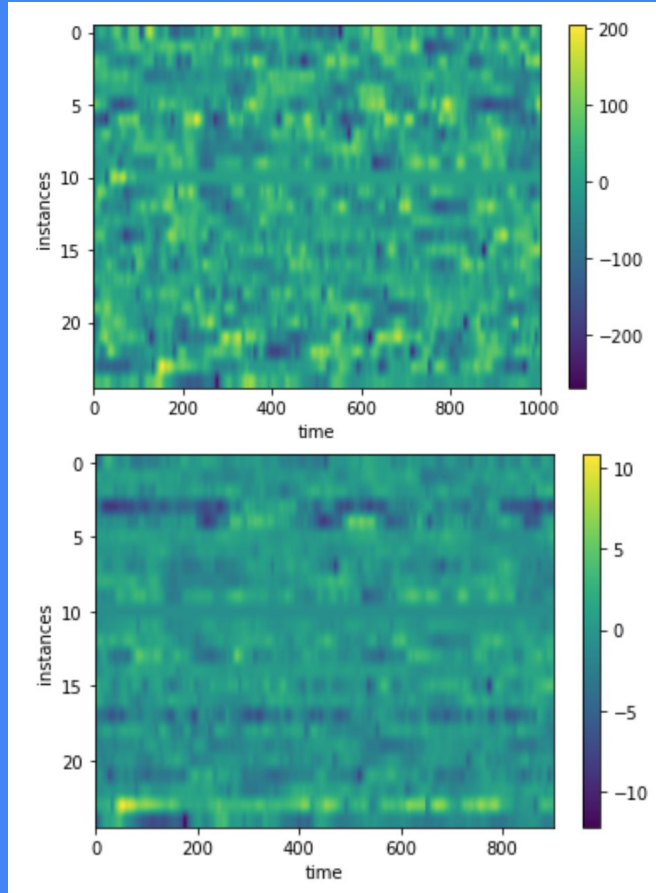
Normalization / Standardization

- Neural population drift across time.
- Electrode shift across days.
- Different dynamic ranges across patients.
- Habituation to stimuli.
- ...

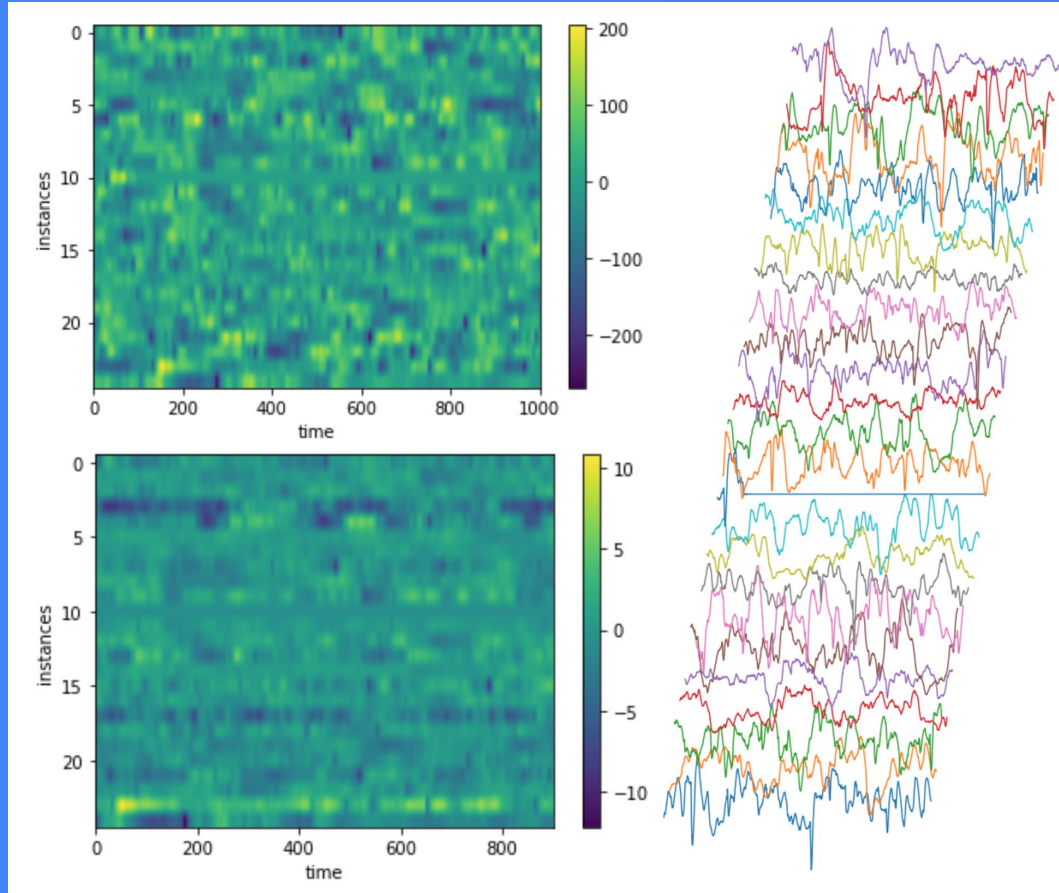
Normalization / Standardization



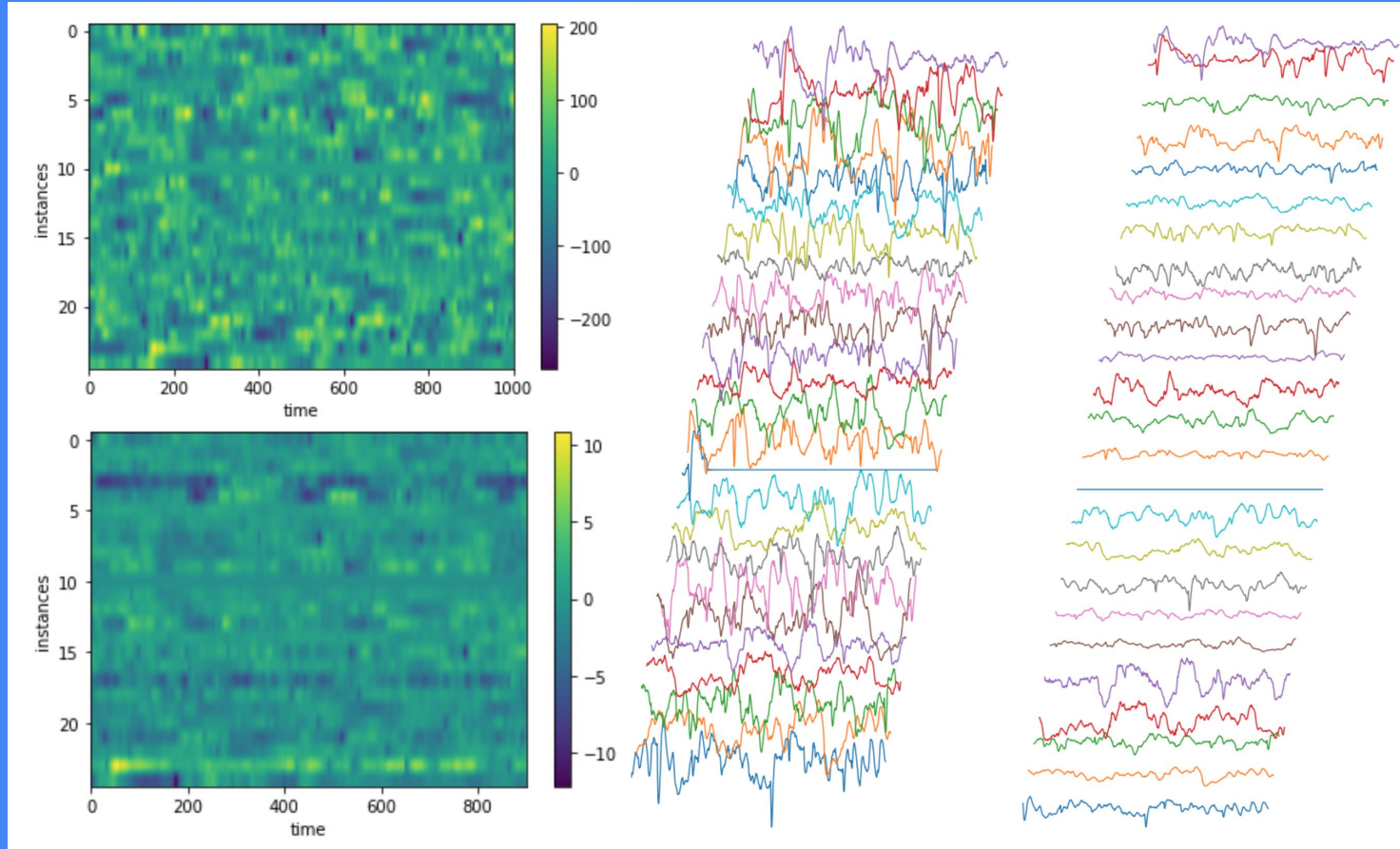
Normalization / Standardization



Normalization / Standardization

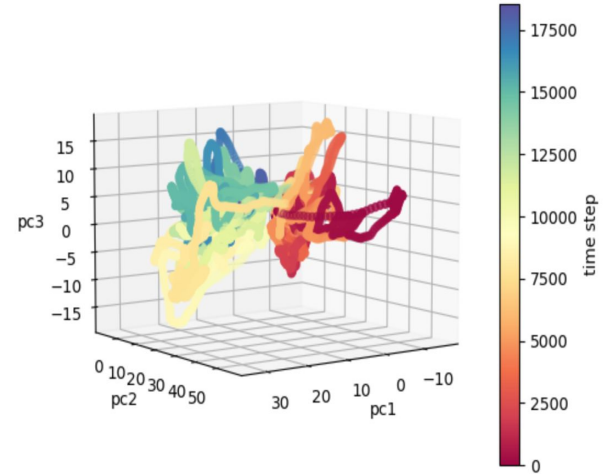
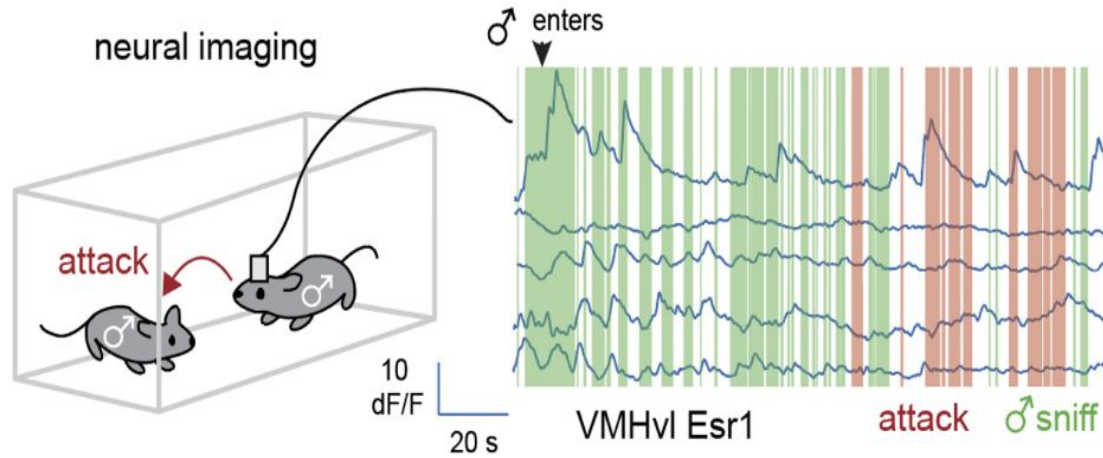


Normalization / Standardization

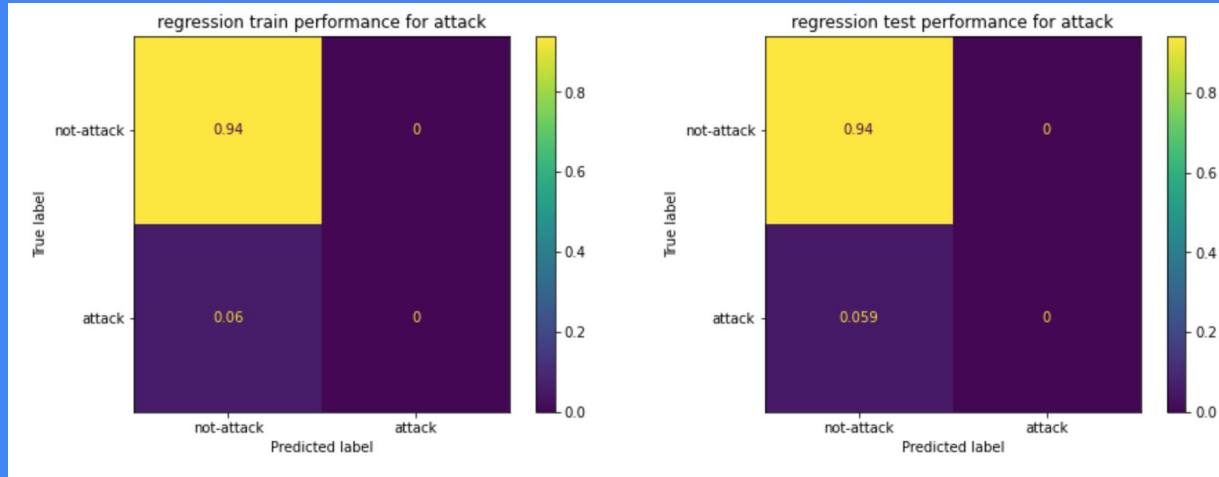


Train / Validation / Test Splits

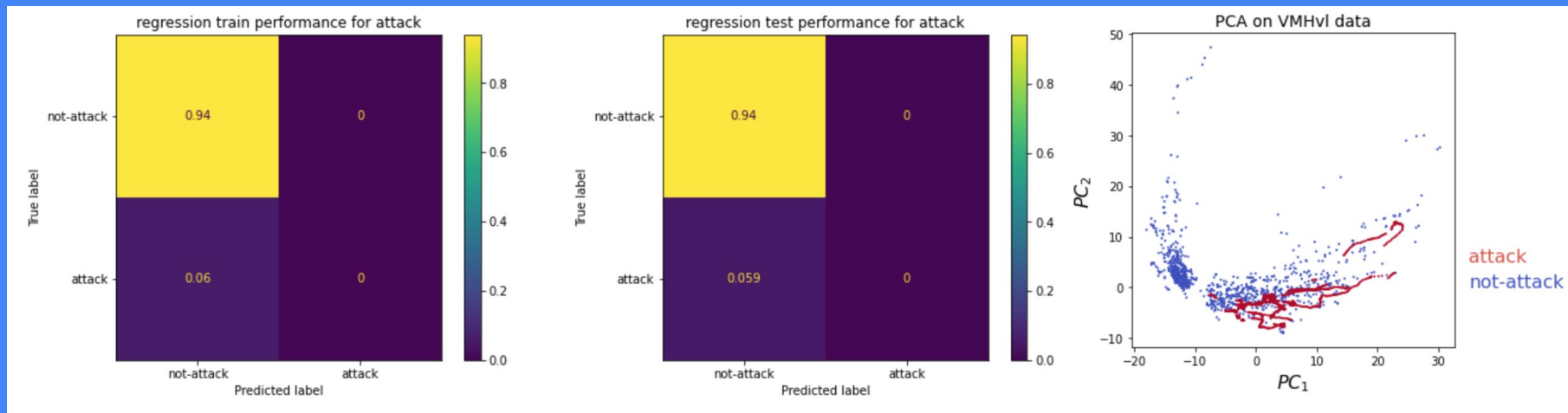
Train / Validation / Test Splits



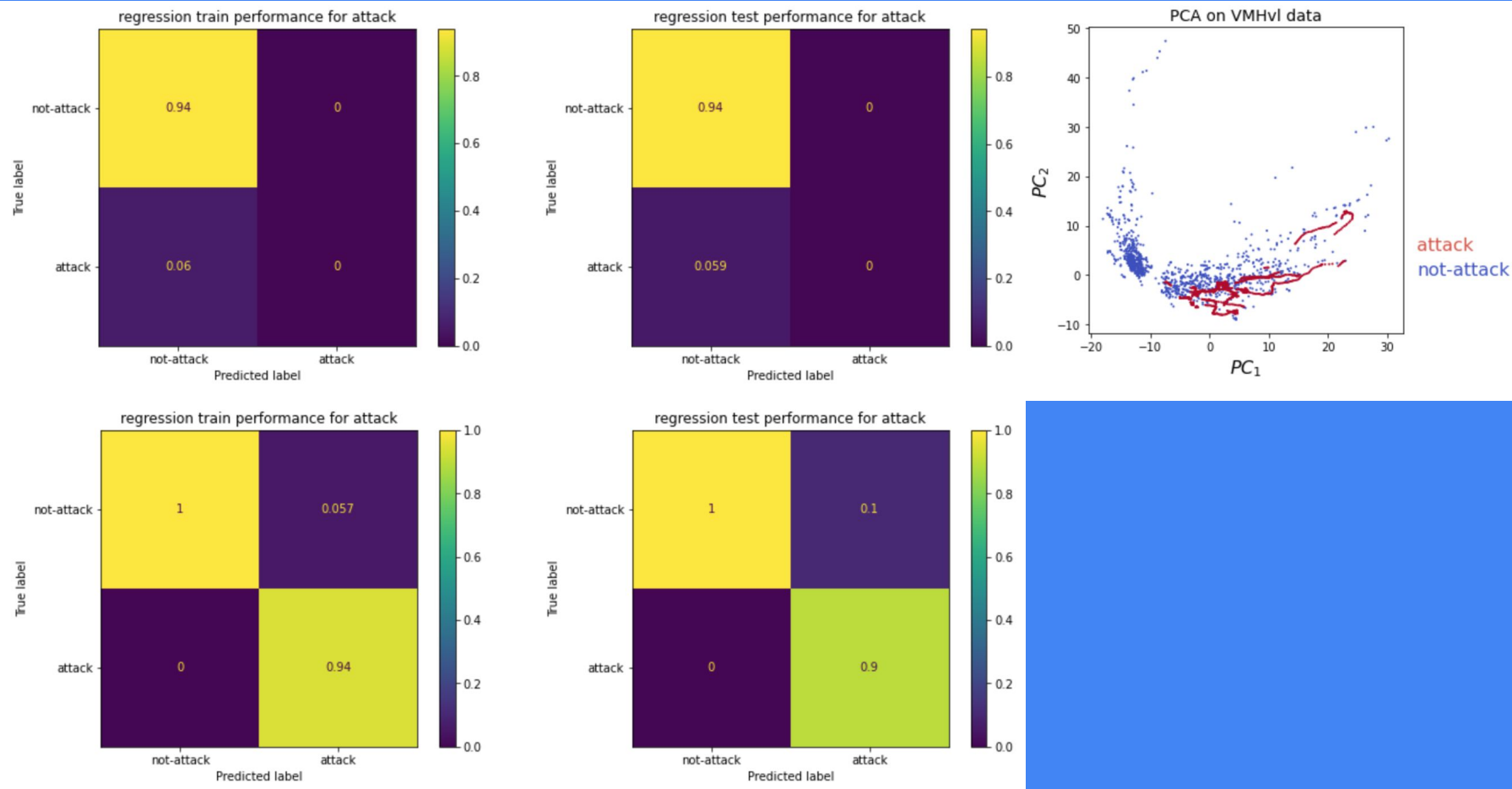
Train / Validation / Test Splits



Train / Validation / Test Splits



Train / Validation / Test Splits



Train / Validation / Test Splits

https://github.com/SaberaTalukder/Chen_Institute_DataSAI_for_Neuroscience/blob/main/07_05_22_day1_overview/code/diy_notebooks/dataset_engineering.ipynb