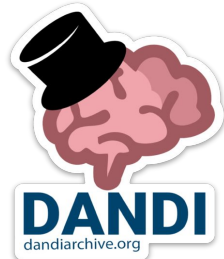# NWB User Days 2023

Visualization Discussion

Data Science Pyramid

THE DATA SCIENCE
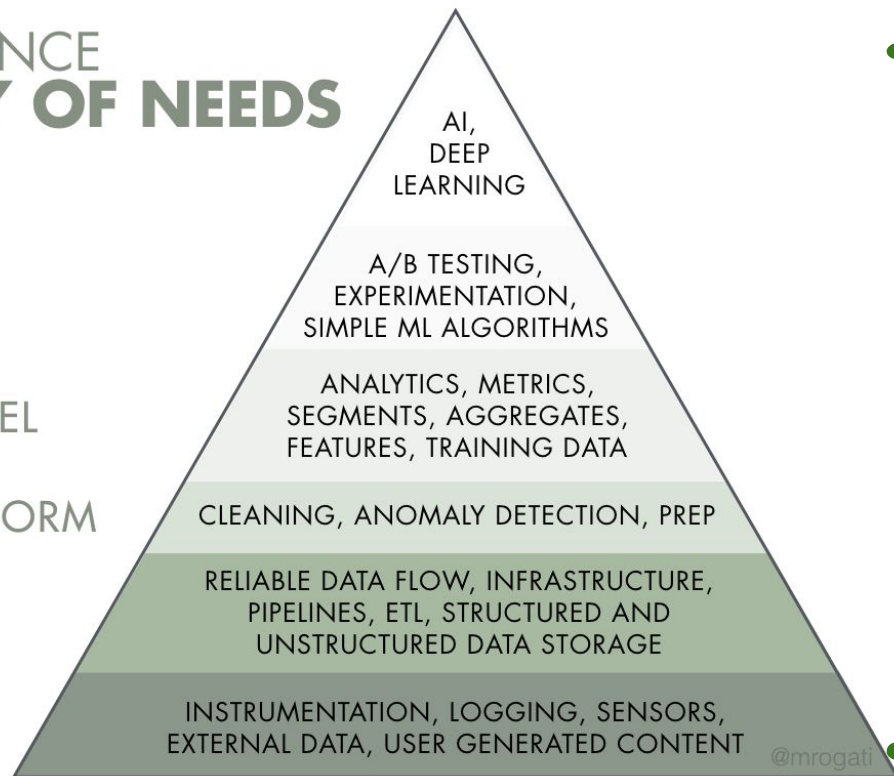**HIERARCHY OF NEEDS**

LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT

AI, DEEP LEARNING

A/B TESTING, EXPERIMENTATION, SIMPLE ML ALGORITHMS

ANALYTICS, METRICS, SEGMENTS, AGGREGATES, FEATURES, TRAINING DATA

CLEANING, ANOMALY DETECTION, PREP

RELIABLE DATA FLOW, INFRASTRUCTURE, PIPELINES, ETL, STRUCTURED AND UNSTRUCTURED DATA STORAGE

INSTRUMENTATION, LOGGING, SENSORS, EXTERNAL DATA, USER GENERATED CONTENT

@mrogati

Visualizations of data are used for communication at each stage

https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007

CATALYST
NEURO

Data Science Pyramid

THE DATA SCIENCE **HIERARCHY OF NEEDS**

LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT

AI, DEEP LEARNING

A/B TESTING, EXPERIMENTATION, SIMPLE ML ALGORITHMS

ANALYTICS, METRICS, SEGMENTS, AGGREGATES, FEATURES, TRAINING DATA

CLEANING, ANOMALY DETECTION, PREP

RELIABLE DATA FLOW, INFRASTRUCTURE, PIPELINES, ETL, STRUCTURED AND UNSTRUCTURED DATA STORAGE

INSTRUMENTATION, LOGGING, SENSORS, EXTERNAL DATA, USER GENERATED CONTENT

@mrogati
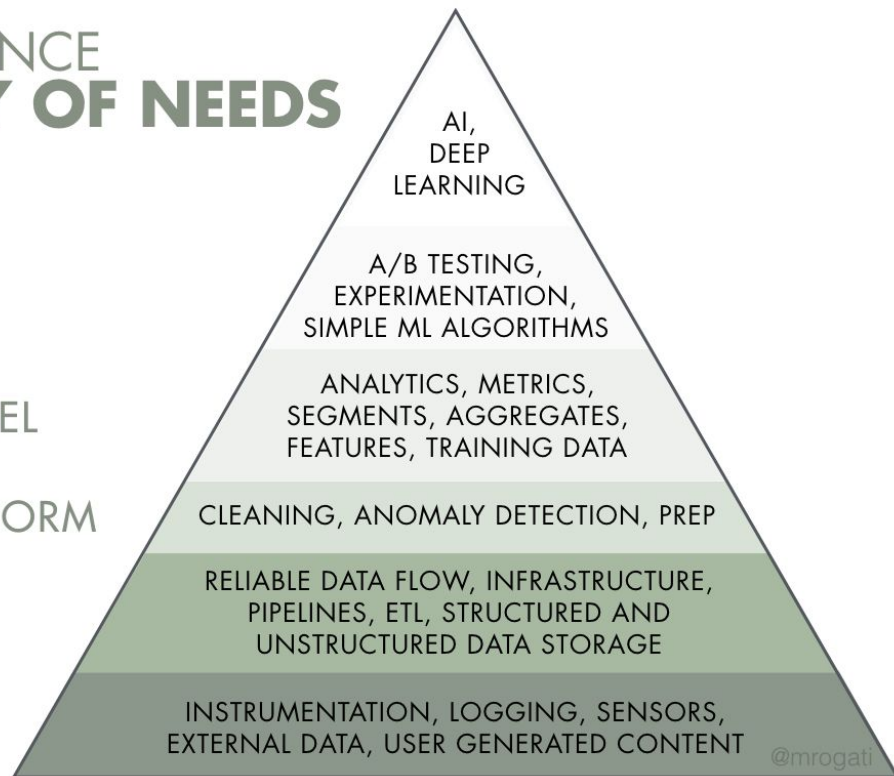
Some recording systems give real-time traces

Most acquisition stores metadata files, logs etc.

Experimenters write down notes

CATALYST NEURO

Data Science Pyramid
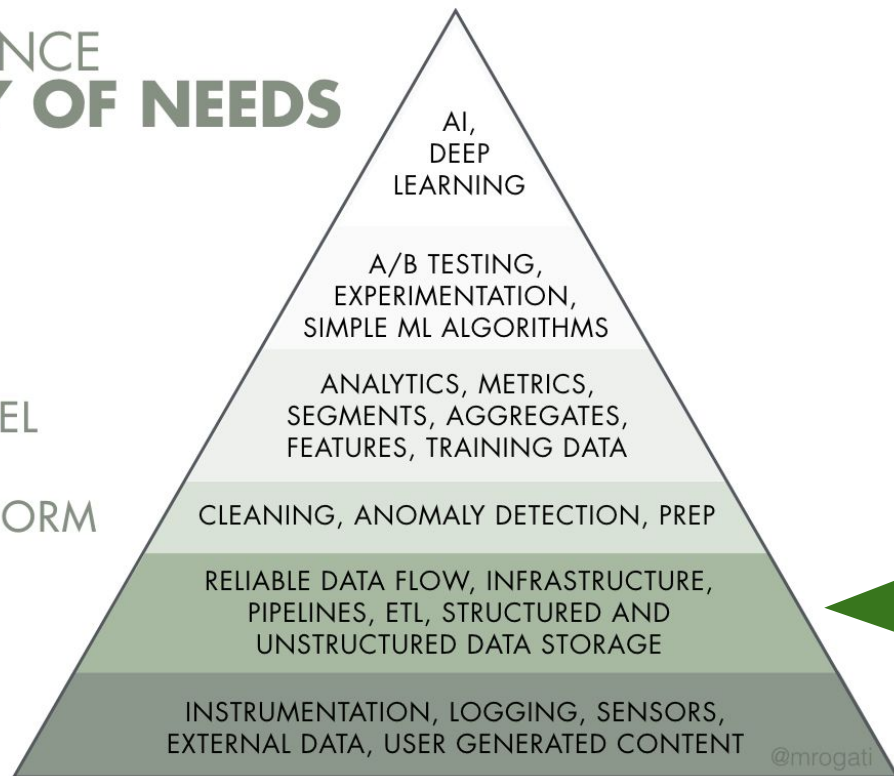
Processing pipelines increasingly visualize workflows using 'graphical programming' (nodes are operations, arrows indicate input/output)

Progress bars

THE DATA SCIENCE
HIERARCHY OF NEEDS

LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT

AI, DEEP LEARNING

A/B TESTING, EXPERIMENTATION, SIMPLE ML ALGORITHMS

ANALYTICS, METRICS, SEGMENTS, AGGREGATES, FEATURES, TRAINING DATA

CLEANING, ANOMALY DETECTION, PREP

RELIABLE DATA FLOW, INFRASTRUCTURE, PIPELINES, ETL, STRUCTURED AND UNSTRUCTURED DATA STORAGE

INSTRUMENTATION, LOGGING, SENSORS, EXTERNAL DATA, USER GENERATED CONTENT

@mrogati

CATALYST
NEURO

Data Science Pyramid

THE DATA SCIENCE
HIERARCHY OF NEEDS

LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT

AI, DEEP LEARNING

A/B TESTING, EXPERIMENTATION, SIMPLE ML ALGORITHMS

ANALYTICS, METRICS, SEGMENTS, AGGREGATES, FEATURES, TRAINING DATA

CLEANING, ANOMALY DETECTION, PREP

RELIABLE DATA FLOW, INFRASTRUCTURE, PIPELINES, ETL, STRUCTURED AND UNSTRUCTURED DATA STORAGE

INSTRUMENTATION, LOGGING, SENSORS, EXTERNAL DATA, USER GENERATED CONTENT

@mrogati

Manual curation

ROI tracing or rejection

etc.

CATALYST
NEURO

Data Science Pyramid

THE DATA SCIENCE
HIERARCHY OF NEEDS

AI,
DEEP
LEARNING

A/B TESTING,
EXPERIMENTATION,
SIMPLE ML ALGORITHMS

ANALYTICS, METRICS,
SEGMENTS, AGGREGATES,
FEATURES, TRAINING DATA

CLEANING, ANOMALY DETECTION, PREP

RELIABLE DATA FLOW, INFRASTRUCTURE,
PIPELINES, ETL, STRUCTURED AND
UNSTRUCTURED DATA STORAGE

INSTRUMENTATION, LOGGING, SENSORS,
EXTERNAL DATA, USER GENERATED CONTENT

LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT

@mrogati
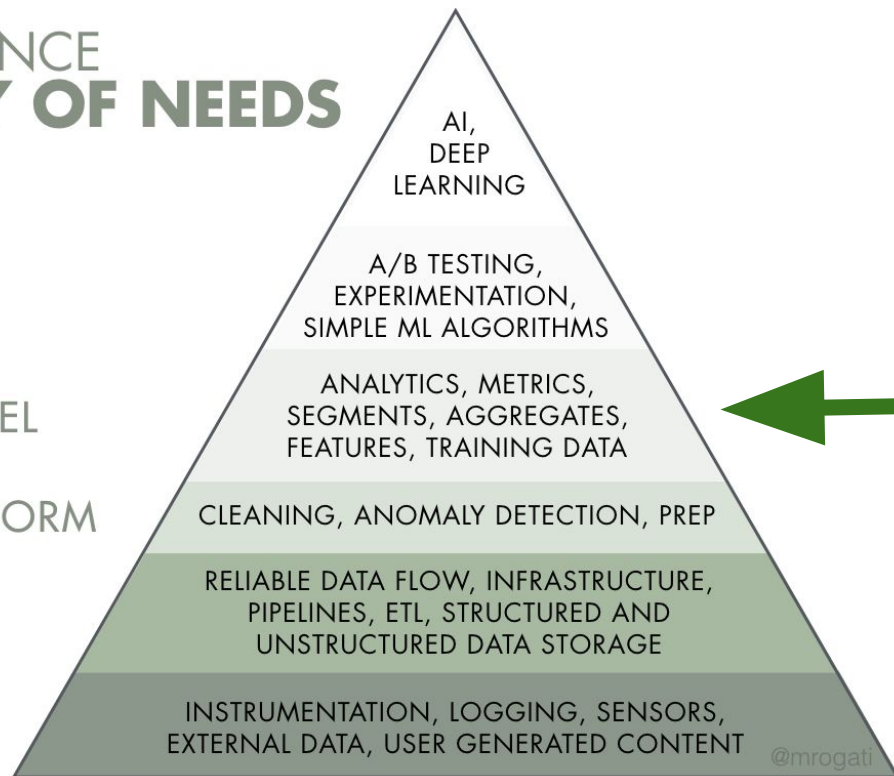
PSTH

Receptive fields

Labeling frames for
pose estimation

… many more

Data Science Pyramid
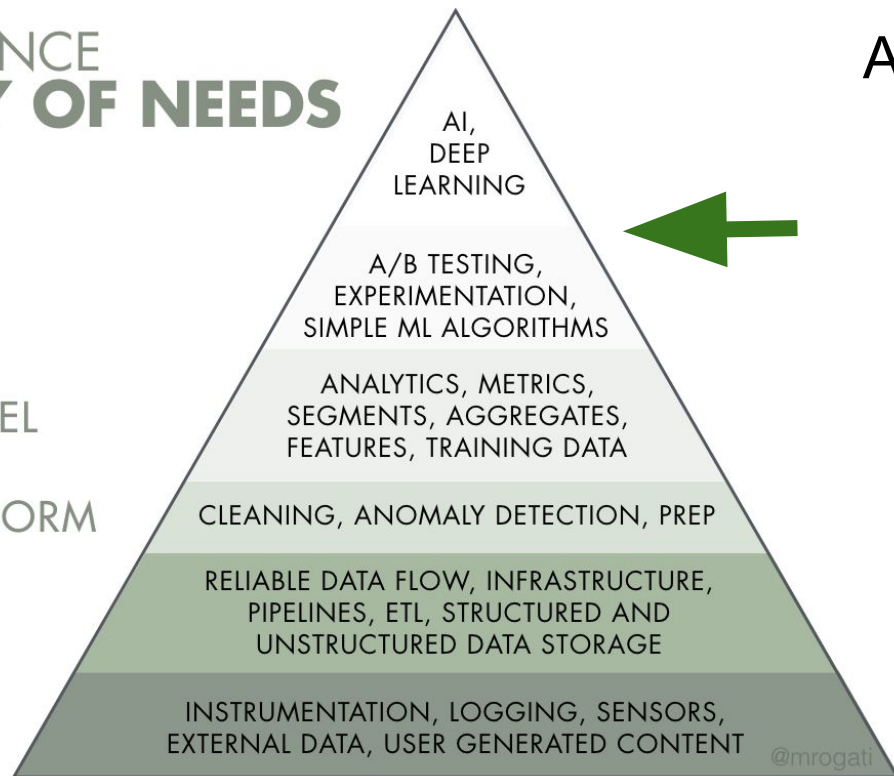
Dataset level summaries

Average or variability of metrics over subjects and sessions

Reference atlases

Latent space models
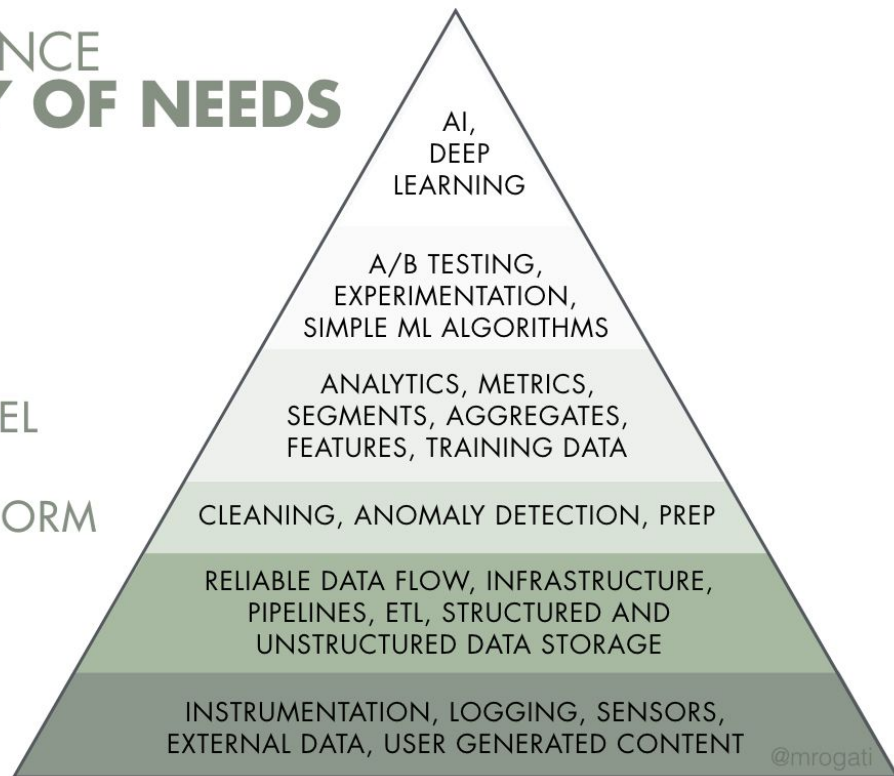
…many more

THE DATA SCIENCE
HIERARCHY OF NEEDS

LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT

AI, DEEP LEARNING

A/B TESTING, EXPERIMENTATION, SIMPLE ML ALGORITHMS

ANALYTICS, METRICS, SEGMENTS, AGGREGATES, FEATURES, TRAINING DATA

CLEANING, ANOMALY DETECTION, PREP

RELIABLE DATA FLOW, INFRASTRUCTURE, PIPELINES, ETL, STRUCTURED AND UNSTRUCTURED DATA STORAGE

INSTRUMENTATION, LOGGING, SENSORS, EXTERNAL DATA, USER GENERATED CONTENT

@mrogati

CATALYST
NEURO

Data Science Pyramid

THE DATA SCIENCE
HIERARCHY OF NEEDS

LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT

AI,
DEEP
LEARNING

A/B TESTING,
EXPERIMENTATION,
SIMPLE ML ALGORITHMS

ANALYTICS, METRICS,
SEGMENTS, AGGREGATES,
FEATURES, TRAINING DATA

CLEANING, ANOMALY DETECTION, PREP

RELIABLE DATA FLOW, INFRASTRUCTURE,
PIPELINES, ETL, STRUCTURED AND
UNSTRUCTURED DATA STORAGE

INSTRUMENTATION, LOGGING, SENSORS,
EXTERNAL DATA, USER GENERATED CONTENT

@mrogati

Diversity
and
complexity
of
visualizations
increase

Challenges from experience

- Figures are harder to reproduce (exactly) than they are to initially create
  - Sharing generative code that acts on data can help alleviate
  - Examples: example-notebooks on DANDI

- The exact version of package dependencies used to create those can make a big difference and is often not tracked
  - Possible solutions: Code Ocean, GitHub + codespaces, rigorous pipelines that log everything in the background

- Building a visualization app that completely supports all platforms and device architectures is difficult and time consuming
  - Visualization packages typically have deep dependencies, may rely on configuration of GPU
  - Web-based tends to be more stable
  - Worth the time for software developers; probably not for average scientist

CATALYST
NEURO

Questions around-the-room

When performing experiments, analyzing data, or developing software/algorithms…

- Broadly: what general information do you hope to communicate with visualizations you create?

- Specifically: how exactly do you accomplish that (what specific viewers do you support)

- Challenges: what difficulties do you encounter?

- Ideally: what would you do if there were no tech limitations (framework difficulties or data limits)

- NWB: In what ways can the NWB format or the DANDI archive help you get closer to the ideal?

CATALYST
NEURO