Data science from a bioinformatic viewpoint

Hacker Dojo Python Club
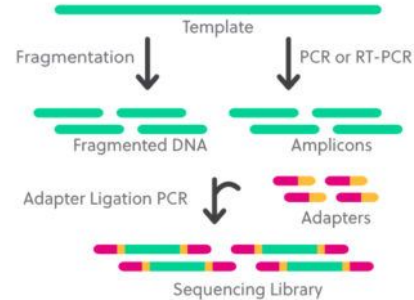Presenter - Jessica Huynh-Westfall, Ph.D.

Agenda

- Introduce Data science and machine learning from a Biologist viewpoint
- Overview concept map
- Python example of neural network
- Gather feedback on if there's interest in future talks

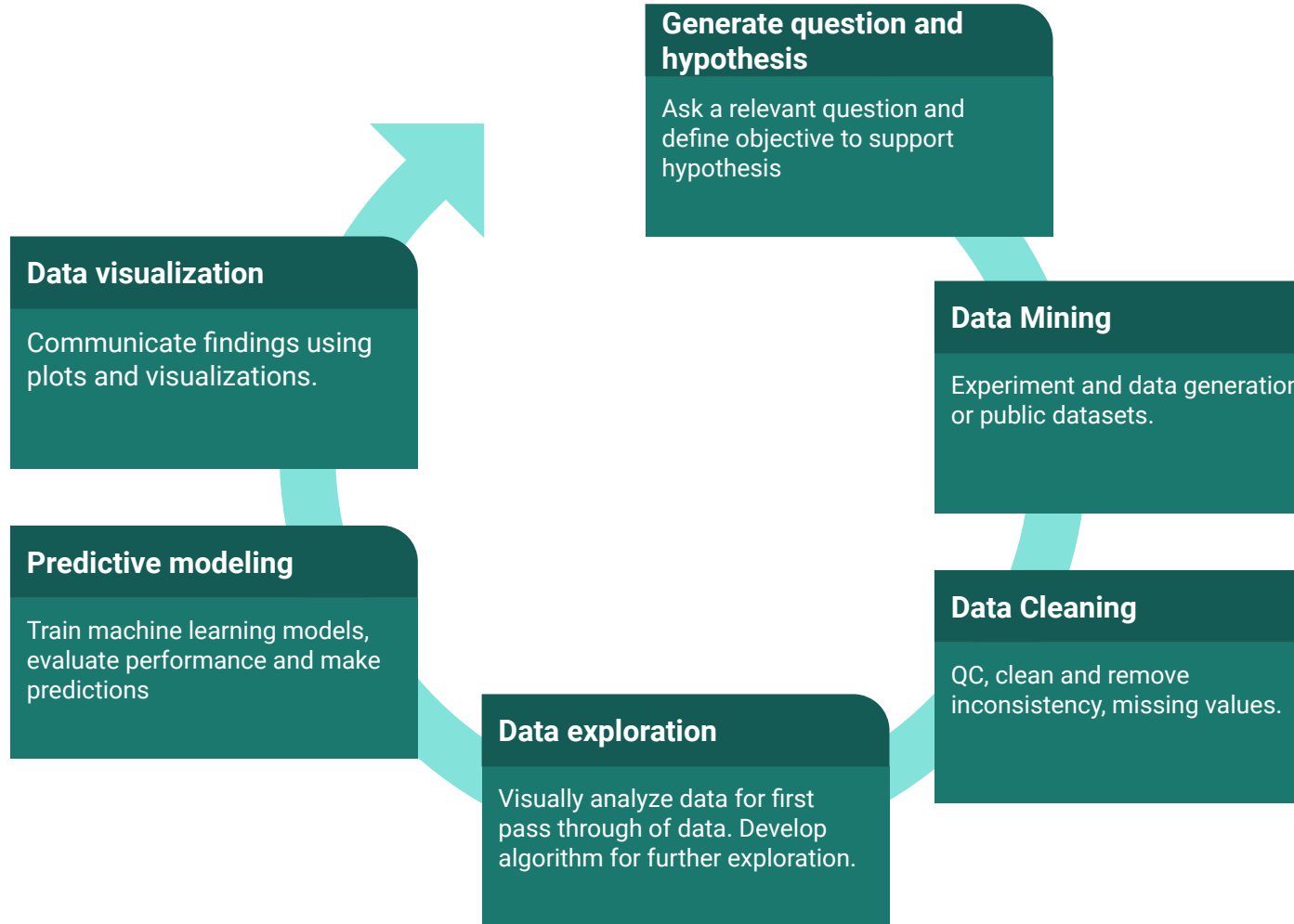# What type of data do I typically work with?

Data science life cycle

Generate question and hypothesis

Ask a relevant question and define objective to support hypothesis

Data Mining

Experiment and data generation or public datasets.

Data Cleaning

QC, clean and remove inconsistency, missing values.

Data exploration

Visually analyze data for first pass through of data. Develop algorithm for further exploration.

Predictive modeling

Train machine learning models, evaluate performance and make predictions

Data visualization

Communicate findings using plots and visualizations.

Machine Learning

- Supervised learning — classification/regression — Linear models / Logistic regression
- Unsupervised learning — clustering/prediction — K-mean
  - t-SNE, PCA, UMAP
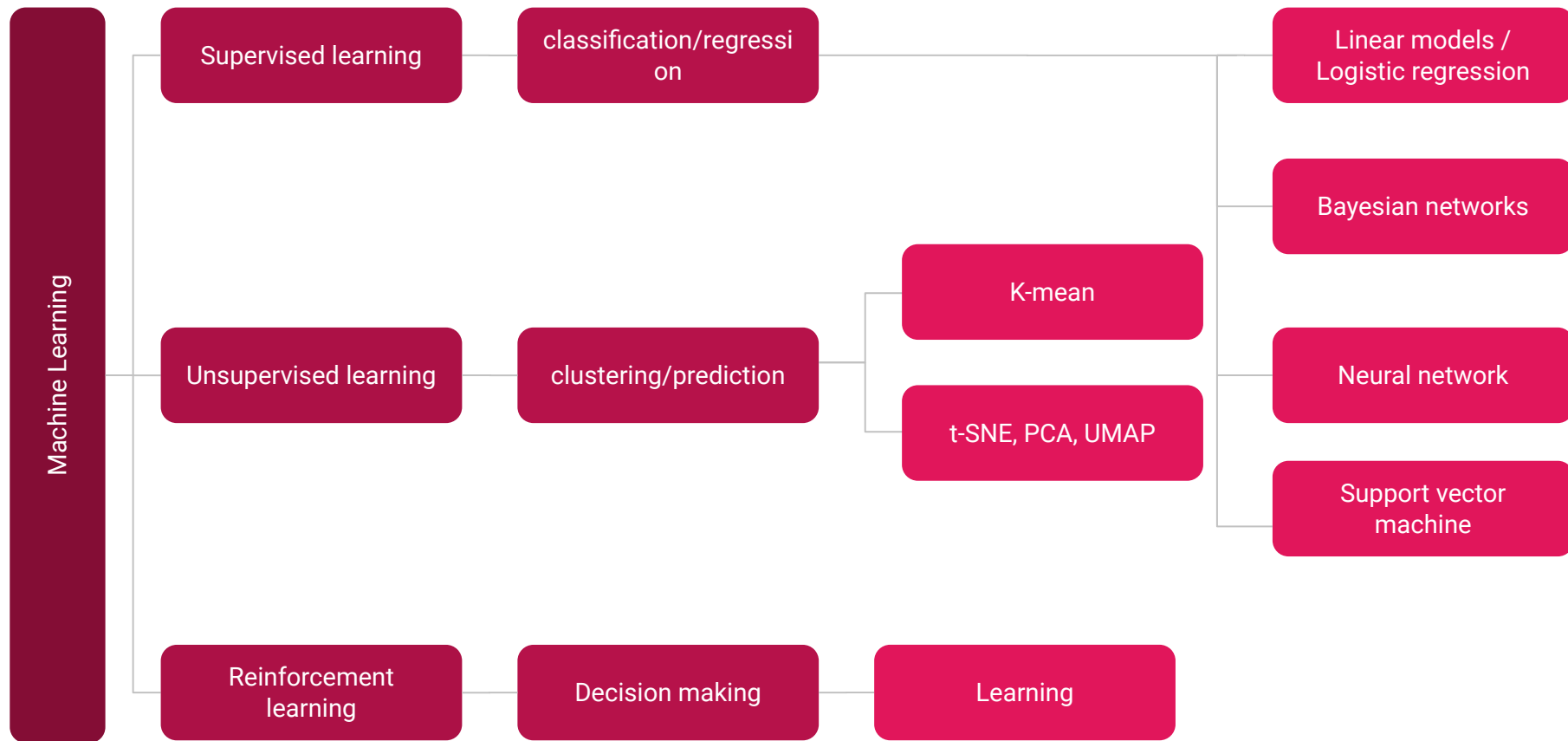- Reinforcement learning — Decision making — Learning

Bayesian networks

Neural network

Support vector machine

Demonstrate PCA, linear, and logistic regression

All code can be found on Github
(https://github.com/jessicatwes/HackerDojo_PythonClub/blob/main/20240319_introDS/)