Welcome!
MS in Data Science

Ryan T. Moore

American University

2025-08-22

- Political methodologist
  (Dept of Government, SPA)

- Political methodologist
  (Dept of Government, SPA)
- Assoc Director, Center for Data Science
  (AU, SPA)

- Political methodologist
  (Dept of Government, SPA)
- Assoc Director, Center for Data Science
  (AU, SPA)
- Senior Social Scientist
  (The Lab @ DC)

- Political methodologist
  (Dept of Government, SPA)
- Assoc Director, Center for Data Science
  (AU, SPA)
- Senior Social Scientist
  (The Lab @ DC)
- Methods Fellow
  (Office of Evaluation Sciences, US GSA)

Connections

- Winter Institute in Data Science

Connections

- Winter Institute in Data Science
- RAs with The Lab @ DC

Data Science

Particular intersection of

- Statistical practice
- Computational tools
- Substantive knowledge

- Stats: prediction (vs. explanation), algorithms (vs. models)

- Stats: prediction (vs. explanation), algorithms (vs. models)
- Computing: addressing problems with data *per se* (size, tidy-ness, un/structure, replicability)

- Stats: prediction (vs. explanation), algorithms (vs. models)
- Computing: addressing problems with data *per se* (size, tidy-ness, un/structure, replicability)
- Substance: social science

Three Tasks of Data Science

- Description

Three Tasks of Data Science

- Description
- Prediction

Three Tasks of Data Science

- Description
- Prediction
- Causal Inference

Three Tasks of Data Science

- Description
- Prediction
- Causal Inference

Three Tasks of Data Science

- Description
- Prediction
- Causal Inference

Models/algorithms central to all three.

Three Tasks of Data Science

- Description
- Prediction
- Causal Inference

Models/algorithms central to all three.

Hernán, Hsu, and Healy (2019)

Three Tasks of Data Science

Description

- Identifying patterns, etc.

Three Tasks of Data Science

Description

- Identifying patterns, etc.
- E.g., k-means clustering to discover groups

Three Tasks of Data Science

Prediction

- Components

Three Tasks of Data Science

Prediction

- Components
  - Inputs/outputs
    (predictors/outcomes, features/responses, ...)

Three Tasks of Data Science

Prediction

- Components
  - Inputs/outputs
    (predictors/outcomes, features/responses, ...)
  - Mapping from inputs to outputs
    (linear model, decision tree, ...)

Three Tasks of Data Science

Prediction

- Components
  - Inputs/outputs
    (predictors/outcomes, features/responses, ...)
  - Mapping from inputs to outputs
    (linear model, decision tree, ...)
  - Metric for evaluating mapping

Three Tasks of Data Science

Prediction

- Components
  - Inputs/outputs
    (predictors/outcomes, features/responses, ...)
  - Mapping from inputs to outputs
    (linear model, decision tree, ...)
  - Metric for evaluating mapping
- With these, model/machine learning algorithm does the work

Three Tasks of Data Science

Prediction

- Components
  - Inputs/outputs
    (predictors/outcomes, features/responses, ...)
  - Mapping from inputs to outputs
    (linear model, decision tree, ...)
  - Metric for evaluating mapping
- With these, model/machine learning algorithm does the work
- E.g., regression, random forests, neural networks, ...

Three Tasks of Data Science

Causal Inference

- Potential outcomes/counterfactual/interventionist perspective

Three Tasks of Data Science

Causal Inference

- Potential outcomes/counterfactual/interventionist perspective
- Requires *expertise* different to description/prediction

Three Tasks of Data Science

Causal Inference

- Potential outcomes/counterfactual/interventionist perspective
- Requires *expertise* different to description/prediction
- Requires more than summary statistics, metrics, etc.

Three Tasks of Data Science

Causal Inference

- Potential outcomes/counterfactual/interventionist perspective
- Requires *expertise* different to description/prediction
- Requires more than summary statistics, metrics, etc.
- Requires some knowledge of causal structure

Three Tasks of Data Science

Causal Inference

- Potential outcomes/counterfactual/interventionist perspective
- Requires *expertise* different to description/prediction
- Requires more than summary statistics, metrics, etc.
- Requires some knowledge of causal structure
  - Not all inputs treated same

Three Tasks of Data Science

Causal Inference

- Potential outcomes/counterfactual/interventionist perspective
- Requires *expertise* different to description/prediction
- Requires more than summary statistics, metrics, etc.
- Requires some knowledge of causal structure
  - Not all inputs treated same
  - T v. $\mathbf{X}$ v. M – very different!

Three Tasks of Data Science

Causal Inference

- Potential outcomes/counterfactual/interventionist perspective
- Requires *expertise* different to description/prediction
- Requires more than summary statistics, metrics, etc.
- Requires some knowledge of causal structure
  - Not all inputs treated same
  - T v. $\mathbf{X}$ v. M – very different!
  - (the more knowledge, the better!)

Three Tasks of Data Science

Causal Inference

- Potential outcomes/counterfactual/interventionist perspective
- Requires *expertise* different to description/prediction
- Requires more than summary statistics, metrics, etc.
- Requires some knowledge of causal structure
  - Not all inputs treated same
  - T v. $\mathbf{X}$ v. M – very different!
  - (the more knowledge, the better!)
- E.g., experiments, observational causal designs, ...

Goals

- Computational tools (R, Python, etc.)
- Varied data experience, original work
- Apply knowledge in academia, government, industry
- Communicate understanding

Requirements

- 10 courses / 30 credits

Requirements

- 10 courses / 30 credits
- Choose a track: **Applied Public Affairs**, Business Analytics, Computer Sci, Cybersecurity, Environmental Sci, Finance, Int'l Economic Relations, Investigative Journalism, Microeconomic Analysis, Specialized Methods

Requirements

- 10 courses / 30 credits
- Choose a track: **Applied Public Affairs**, Business Analytics, Computer Sci, Cybersecurity, Environmental Sci, Finance, Int'l Economic Relations, Investigative Journalism, Microeconomic Analysis, Specialized Methods
- Catalog ('25-'26): https://t.ly/Ib43y

  https://catalog.american.edu/preview_program.php?catoid=21&poid=10388

Courses

- DATA-612: R
- DATA-613: Data Science (modeling, Shiny, etc.)
- STAT-615: Regression
- STAT-627: Stat Machine Learning
- One of
  - GOVT/STAT-618: Bayes
  - GOVT-650: Political Analysis
- One of
  - GOVT-653: Intro Quant Methods in Pol Sci
  - STAT-614: Stat Methods
- DATA-793: Capstone
  - (The Lab @ DC DHS connections ...)
- Three more courses
  - (Electives, depend on track)
  - Substance or methods (computer vision, AI, ML, deep learning, NLP, survey sampling, ...)

Directors

- Jeff Gill: `jeffgill.org`
- Betty Malloy: `malloy@american.edu`

Questions?

Thanks!

rtm@american.edu
www.ryantmoore.org

References

Hernán, Miguel A., John Hsu, and Brian Healy. 2019. "A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks." *CHANCE* 32 (1): 42–49. https://doi.org/10.1080/09332480.2019.1579578.