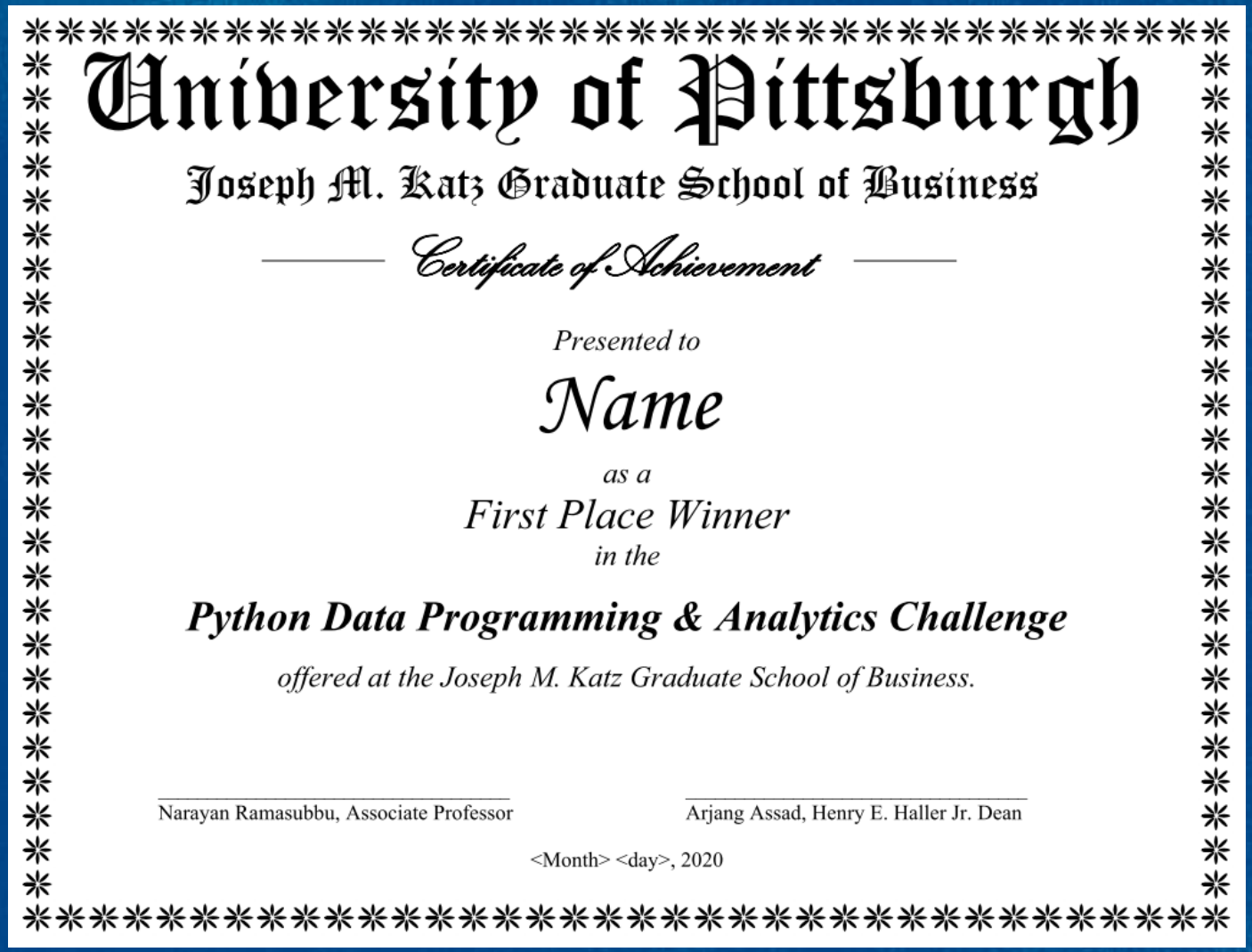
BMIS-2542: Data Programming Essentials *with Python*

AY 2019-20 Spring; Narayan Ramasubbu

5MT Analytics Challenge

University of Pittsburgh | Joseph M. Katz Graduate School of Business and College of Business Administration

Do well in the 5MT project and earn Katz certificate



University of Pittsburgh

Joseph M. Katz Graduate School of Business

Certificate of Achievement

Presented to

Name

as a

First Place Winner

in the

Python Data Programming & Analytics Challenge

offered at the Joseph M. Katz Graduate School of Business.

Narayan Ramasubbu, Associate Professor

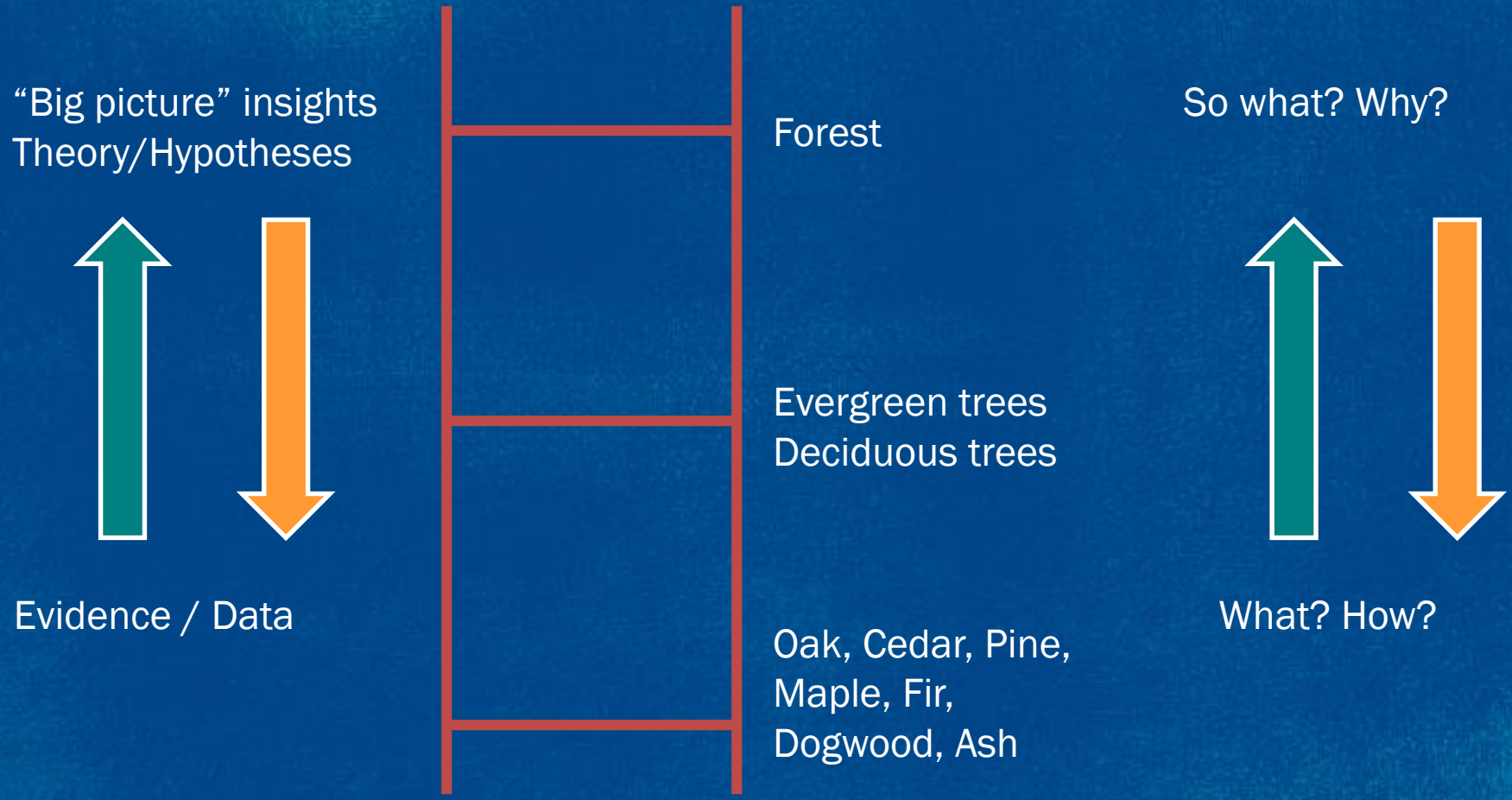
Arjang Assad, Henry E. Haller Jr. Dean

<Month> <day>, 2020

Key items to remember

- Explanation vs. prediction
  - Don't have to be "either…or" for 5MT
  - Rely on "theory" or prior knowledge for explanation
  - Emphasize the practical applications of your prediction
  - Remember it is all about data-driven story telling
- Correlation is not causation
- Bias-variance tradeoff
  - What can we do about it?
  - Always perform cross-validation before finalizing evidence

Story-telling tip: move up and down the abstraction ladder



Expected analysis workflow

1. Clean and pre-process data
2. Explore patterns in data
3. Identify options for investigations / story lines / narratives
4. "Train" models (classification and/or regression)
5. Evaluate models through cross-validation
   - Classification: confusion matrix
   - Regression: RMSE

- Advanced tool: model tuning with GridSearchCV()

Wrangling and modeling approaches

- The "Pandas way"
- Linear models
  - Transformations, interactions
  - Explanation-orientation (hypothesis testing)
  - Prediction-orientation
- Logistic regression
- KNN
- Tree-based approaches
  - Bagging, RandomForest, Boosting

Final 5MT Project Deliverables

Submission packet:

1. Jupyter Notebook
   - Well-organized with comments to identify the steps of your analysis
   - Use Markdown cell for comments
2. Any additional data (if used)
   - Store it in an accessible cloud folder (e.g., Pitt Box) and provide link.
3. Link to your video file
   - Panopto hosted or video file (.mp4) stored in accessible cloud folder
   - Provide a link to your video
   - Make sure all links are set with correct “share” settings

5MT: Got Questions?

- Post all questions in Piazza forum
  - Keep in mind the individual evaluation context
  - Pitt Academic Integrity policies apply
  - Don't discuss with anyone and don't share/post your work
- I will respond to all challenging / technical issues to help all students