

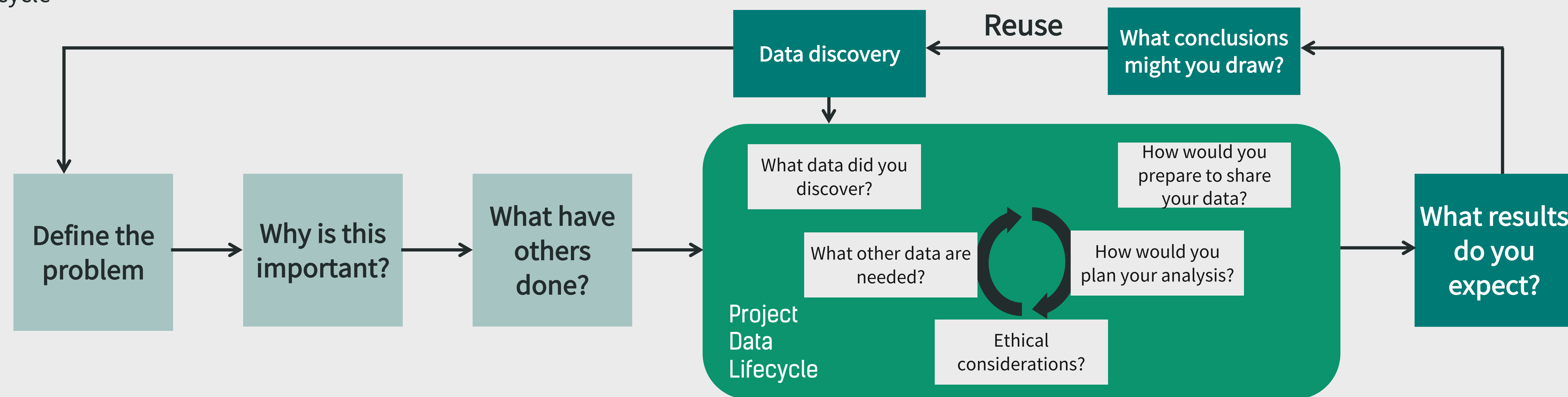
Getting Your Hands Dirty With Data

Ted Laderas¹, Melissa Haendel^{1,2}, Nicole Vasilevsky^{1,2}, Bjorn Pederson¹, Jackie Wirz³, William Hersh¹, David A. Dorr^{1,4}, Shannon McWeeney¹

¹Department of Medical Informatics and Clinical Epidemiology, ²Library, ³Graduate Student Affairs, ⁴Department of Medicine Oregon Health & Science University, Portland, OR

Approach

Developing **skills courses** that connect the dots to help researchers understand how to apply data science techniques in the context of the research life cycle



Topics



DEFINING THE PROBLEM

- ❖ What problems are amenable to analytics?
- ❖ Importance of question
- ❖ Team definitions
- ❖ Scope
- ❖ When we do this wrong: methods don't match



DATA IDENTIFICATION AND RESOURCES

- ❖ Finding the right data
- ❖ Search methods
- ❖ Use of metadata
- ❖ Data management



WRANGLING DATA

- ❖ Exploratory Data Analysis (EDA)
- ❖ Data Dictionaries
- ❖ Identifying pitfalls of analysis



METHODS, TOOLS AND ANALYSIS

- ❖ Interactive Visualization
- ❖ Machine Learning



SCIENTIFIC COMMUNICATION

- ❖ Lessons Learned
- ❖ Reporting Findings and Limitations
- ❖ Giving "Elevator Pitch" on approach to problem
- ❖ Critique of related problems

The Skills Courses



Taught by BD2K Faculty, Post-doc and Staff



In person format, with hands on learning



Targeted to a variety of skill levels

Course 1 Intro Course JULY 2015

WHO
Interns and undergraduates

WHAT
Taught basics of data science in the context of the research life cycle

Course 2 Data After Dark JAN 2016

WHO
OHSU students, staff and faculty

WHAT
Emerging data science activities/research impact

Course 3 Advanced Data After Dark MAY 2016

WHO
OHSU students, staff and faculty

WHAT
Hands on Data wrangling, interactive visualization, EDA, Machine Learning, and Interpretation

Course 4 Data and Donuts JUNE 2016

WHO
OHSU summer interns

WHAT
Basics of data science

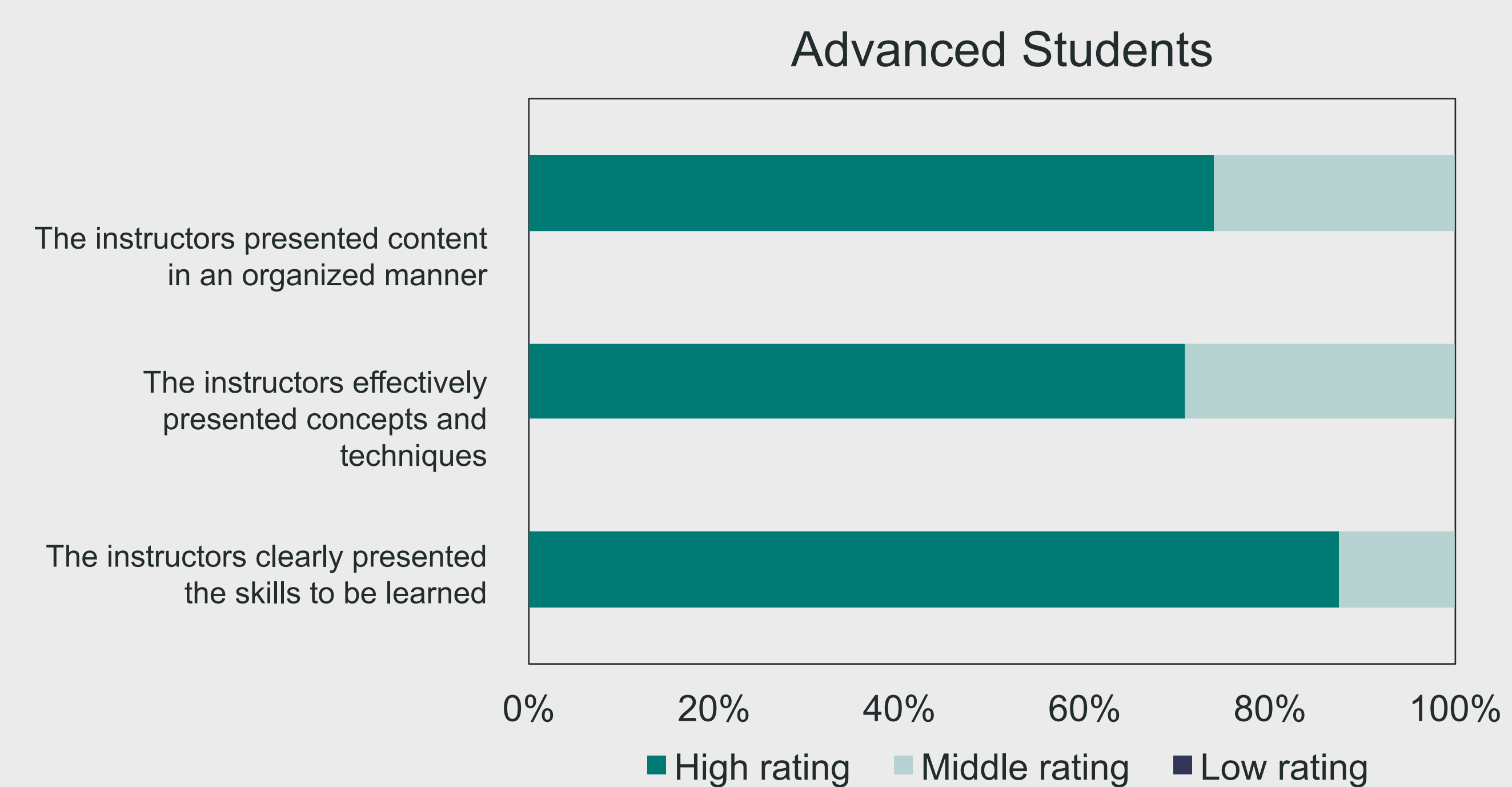
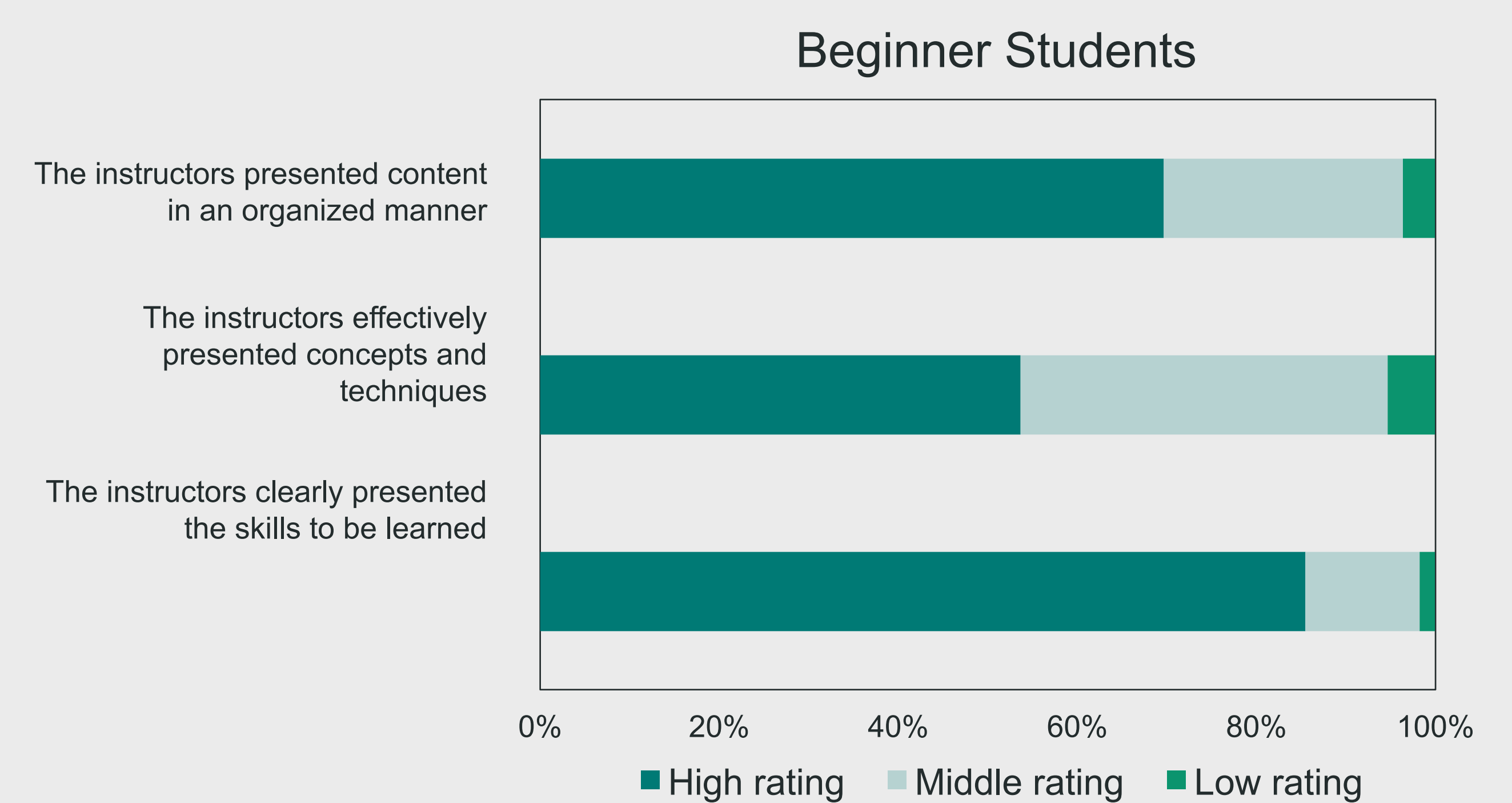
Course 5 Data and Donuts JULY 2016

WHO
OHSU summer interns (West campus)

WHAT
Basics of data science

Evaluations

- ❖ Students filled out evaluation forms with questions on a 1-7 scale, with 1 being the lowest and 7 the highest rating
- ❖ Low ratings were 1-2, middle ratings were 3-5, and high ratings were 6-7
- ❖ Feedback for both beginning and advanced students was strong
- ❖ Hands-on learning workshops were preferred by students



Next Steps

The next skills course will include more hands on components with a realistic synthetic dataset with clinical and genomics data for machine learning and interpretation.

www.ohsu.edu/bd2k

ACKNOWLEDGEMENTS:

This work is supported by NIH Grant 1R25EB020379-01

