Instructions for Using the BEDE Network Data Science Skills Curriculum Map

Part of the following QUBES Educational Resource:

> *O'Donnell, K. L., Aiello-Lammens, M., Bledsoe, E., Bowlick, F. J., Broughton, L., Calderon, O., Crispo, E., Emery, N., Farrell, K., Ngiramahoro, M., Patel, N., Paudel, S., Richardson, L., Soares, B. E., Supp, S., Weigel, E. (2024). BEDE Network Data Science Skills Curriculum Map. Biological and Environmental Data Education (BEDE) Network, QUBES Educational Resources.*

Resource Contents:

- BEDE Network Data Science Skills Curriculum Map (available as an Excel Spreadsheet downloadable here or as a Google Sheet linked here).
- Instructions for Using the BEDE Network Data Science Curriculum Map (this document)

Who we are:

The Biological and Environmental Data Education (BEDE) Network is a community of scientists and educators whose mission is to support instructors as they integrate more data science skills into undergraduate biology and environmental science curricula. One of the ways the network will achieve this objective is through the creation of a data science skills framework, a "Curriculum Map," presented on QUBES as an open educational resource. Over the course of two years, members of the BEDE Network built the Curriculum Map as a flexible tool to help instructors highlight data science skills that are already being taught and to discover new areas to add data science skills that complement the content of their courses or program. The Curriculum Map was built by a committee of scientists and educators representing different schools, fields, positions, and career stages. There are representatives from four-year colleges and community colleges, research institutions and primarily undergraduate institutions. Committee members have backgrounds in multiple fields, including ecology, microbiology, geography, biology education, and data analytics.

Overview:

Welcome to the BEDE Network Data Science (DS) Skills Curriculum Map! We built this tool to help guide you through adding more DS skills to your biological science and environmental science courses and/or help guide you through the process of adding more DS skills across a program, department, or major. The intended audience for this resource is instructors who teach undergraduates in biological and environmental science courses and it was specifically built to focus on the DS skills needed to be successful in those types of courses. This tool is not intended to provide a complete list of data science skills that one would find in a data analytics program, for example.

The Curriculum Map is organized around seven major DS skill categories: Data Management, Analysis, Visualization, Modeling, Coding, Reproducibility, and Ethics. These categories are built off of those used in Hampton *et al.* 2017 and *Vision and Change* (Brewer and Smith 2011) with further modifications to the categories and subcategories based on the results of the BEDE Network incubator survey (Emery *et al.* 2021), Angra and Gardner's Graph Rubric (2018), and discussions with the BEDE Network members. One important change in particular was to break out Ethics into its own major category to highlight the importance of data provenance, responsible conduct, and integrity. We want our students to be thinking about the impacts of where data come from and how they are used and the Ethics-related skills pair very well with the skills in the other categories. We expect a justice and equity lens in our teaching regardless of which of the DS skills we are focused on.

In the Curriculum Map, each of the seven major categories is broken out into three to four sub-categories which contain specific skill areas (Figure 1). The skill areas are associated with specific student learning outcomes (SLOs) at three levels: Foundational, Next Level, Third Level (described below). For example, the SLO "create a graph with appropriate axes and axes titles with units under guidance" is in the Visualization category, Graph Mechanics sub-category, and Graph Elements skill area. You can see all of the SLOs in the Excel file or Google Sheet containing the Curriculum Map. We have numbered the items in the map to make them easier to find, so that particular SLO mentioned earlier would be *Vi 2.2.N1* (Figure 2). In this numbering system, the major categories have two letter abbreviations. Subcategory numbering does not indicate priority or foundational skills and, for the most part, neither does the numbering of the skill areas. In the current version of the Curriculum Map, the numbers merely indicate the position in the framework on each sheet. Planned future versions of this resource will revisit those positions to make it clearer when a Next Level or Third Level skill directly builds on the Foundational one in the same row.

BEDE Network Data Science Skills Curriculum Map Framework

Data Science Skills Category

Subcategory
specific skill areas
Subcategory
specific skill areas

EXAMPLE

Management

Data Handling
data type recognition, entry, organization, ownership
Project Organization
file organization, effective documentation, personnel responsibilities
Security, Sharing, and Accountability
data storage and security

Analysis

Data Structure
variability, sample size
Data Wrangling
cleaning data, data transformations, creating new variables
Making Inferences
formulating hypotheses, hypothesis testing, relating to experimental design
Summarizing and Communicating
basic descriptive statistics, communicating results

Visualization

Visualization Choice
visualization match, visualization type
Graph Mechanics
representing error, graph elements, table elements, map elements, figure or table creation
Communication
visual clarity, accessibility, interpretation

Modeling

Models and Their Applications
modeling, model need, model selection
Model Conceptualization
model limits, assumptions, data, selection
Running Models
software and tool selection, model testing
Interpreting and Communicating Results
communicating model results, interpreting model uncertainty

Ethics

Data Provenance
acknowledging ownership, data sovereignty and open science
Responsible Conduct
informed consent, privacy, certifications and approvals
Integrity
limitation and uncertainty reporting, recognizing misinformation, social consequences, ethical use

Coding

Computational Thinking
problem breakdown, computational tools, flow charts, pseudocode
Reading Code
coded variables, finding help, code errors, code logic
Writing Code
language choice, data import, flow control, iterations, coding notebooks, debugging, results export, expanding base language
Refining Code
existing code, commenting, documenting, optimizing

Reproducibility

Acknowledging
provenance
Sharing
sharing, storage, detailing
Versioning
testing code, compilation, version control, documenting versioning

Figure 1: An overview of the seven major data science skills categories in the Curriculum Map and their subcategories and skill areas.

	A	B	C	D	E
1	**Sub-Cat**	**Skill**	**Foundational**	**Next Level**	**Third Level**
7	**(2) Graph mechanics**	(1) Representing error	(F1) demonstrate an understanding of error	(N1) be able to draw error bars on a graph	(T1) choose among std dev, SE, confidence intervals to display on your graph and justify choice
8			(F2) demonstrate an understanding of what error bars are	(N2) be able to include measurement of error in a table	(T2) choose among std dev, SE, confidence intervals to display on your table and justify choice
9		(2) Graph Elements	(F1) demonstrate an understanding of how to interpret axes	(N1) create a graph with appropriate axes and axes titles with units under guidance	(T1) create a graph with appropriate axes and axes titles with units without guidance
10			(F2) demonstrate an understanding of how to read and use legends	(N2) include appropriate legends, symbols for variables when creating graphs under guidance	(T2) include appropriate legends, symbols for variables when creating graphs without guidance
11		(3) Table Elements	(F1) demonstrate an understanding of how to interpret column and row titles/headings	(N1) include proper column and row titles/headings when creating tables under guidance	(T1) include proper column and row titles/headings when creating tables without guidance
12		(4) Map Elements	(F1) demonstrate an understanding of how interpret the symbols and legends of a map	(N1) include appropriate symbols, color-coding, and legends with creating a map under guidance	(T1) include appropriate symbols, color-coding, and legends with creating a map without guidance

Figure 2: An excerpt from the Visualization (Vi) sheet of the Curriculum Map to demonstrate the numbering system. The red box is the SLO of interest and the arrows point to the numbers. The numbers are read from the left to right, so this SLO would be *Vi 2.2.N1.*

In addition to categorizing all of these skills, we have also indicated which SLOs are Foundational within each skill area in the Curriculum Map. In general, instructors will want to either include the Foundational SLOs for a particular skill or assess the students' proficiency with those Foundational SLOs before proceeding to add SLOs from the "Next Level" and "Third Level." We purposefully did not use course levels (100, 200, 300, etc.) here to make this tool more broadly applicable in different institutional and programmatic contexts. We will have a few examples of course pathways with these associated skills in the next version of this tool.

In each category sheet of the Curriculum Map, there are also some columns that make connections between the categories. The "Foundations Elsewhere" column indicates when there are Foundational SLOs for that skill area in a different major category. Figure 3 is a concept map of our categories and subcategories that illustrates these foundational connections. The "Applications Elsewhere" column indicates where to find SLOs that build on a specific skill area. The "Pairs Well With…" column has recommendations for where to add related SLO from another category. The final column on each sheet, "Resources," features links to activities that address one or more of the SLOs in that skill area. The SLOs in the Curriculum Map are written to be tool agnostic; instructors may choose the appropriate tools (pen and paper, Excel, R, etc.) for their course contexts and student populations. The Resources column is a work in progress and additional resources with examples using different kinds of tools will be added in the next version of the Curriculum Map.

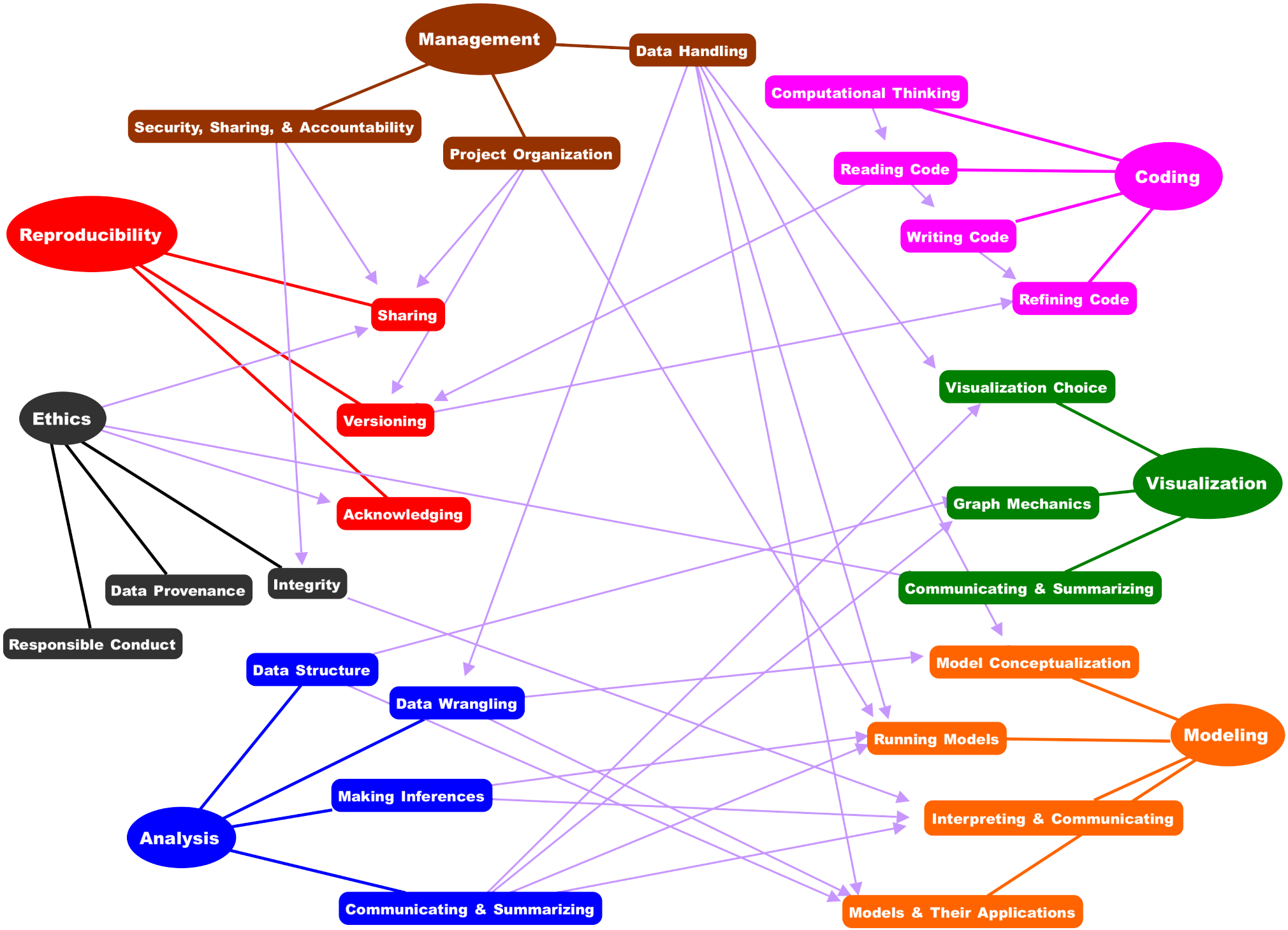

Figure 3: A concept map of the seven major data science skill categories and their subcategories. The arrows indicate where foundational skill SLOs may be found in other categories or subcategories. For example, the Data Handling subcategory of Management has SLOs that are foundational in five other subcategories across three major categories.

Data Science Skill Category Descriptions:

Management skills are those associated with the recording/collecting/generating, organization, storage, and processing of data. The Data Handling sub-category is foundational to other major categories including Visualization and Analysis, where many Data Handling skills are applied.

Analysis skills are those associated with how we summarize data and draw conclusions using statistics. Skill areas in this category are foundational to other major categories, including Visualization.

Visualization skills are associated with clear graphical communication of data analyses, including understanding different types of visualizations and when to use them. Subcategories in this category are inspired by Angra and Gardner's Graph Rubric (2018). This category has many good pairings in the other categories and some of its foundations can be found in Analysis and Management.

Modeling skills are associated with the selecting, fitting, and interpreting of models, including data use and assumptions, limitations and uncertainty, and the testing of models. Some foundations can be found in Management and Analysis.

Ethics skills concern the responsibility one has when holding sensitive information. People collect data or guide the direction of data collection via algorithms, so checks need to be in place to minimize harm and reduce bias. This category has close ties to Management.

Coding skills include those associated with being able to use a particular coding language to perform data analyses and understanding the logic of coding (not necessarily a specific language). In this category, the Computational Thinking subcategory is foundational to the other subcategories.

Reproducibility skills are associated with knowing what tools to use to ensure that aspects of data processing, organization, and analysis can be replicated by other people/groups, including those skills associated with version control. A few of the subcategories have foundations in the Management, Coding, and Ethics categories.

How to Use This Tool:

There are a few different ways instructors can approach this tool. The Curriculum Map can be used as a guide to help add more DS skills to a specific activity/lesson, a whole syllabus, or across a program. The first step is to figure out what your scale is and if you'll be starting with something you are already doing or working on creating something brand new. You may want to set aside some time and go through the entire Curriculum Map to check off what you are already doing, or if you're developing something new, what your target SLOs are.

Below are a few examples of different applications of the Curriculum Map based on what an instructor's goal is.

- *Auditing* - Use the Curriculum Map to evaluate an existing course (or courses across a program) to get a current picture of what DS skills are being taught.
- *Making Changes* - The Curriculum Map can be used to identify points in your existing course or program where SLOs can be added and assessed.
- *Develop Something New* - Use the Curriculum Map as a starting point when building out and scaffolding DS skills across a new course or series of courses that will make DS skills a focus.

Here are some guiding questions you can answer for yourself as you begin to work with this Curriculum Map:

1) Where am I looking to integrate more DS skills? In an activity/lesson, a course, a program, etc.?
2) Am I already doing activities/lessons that address some of the SLOs in the Curriculum Map?
    a) Consider reframing activities where you may feature these skills but don't make the DS skills explicit for your students.
3) Reflect on the spread of those SLOs. Are they from one or two categories? Do they cover many categories?
4) What are the connecting skills (Foundations/Applications/Pairings)?
5) Which skills do I want to focus on adding? For Next or Third Level SLOs, are the Foundations already set?
6) What are possible contexts for those new skills? What tools will students be using to demonstrate specific SLOs in that skill area?
7) What are some ways I may assess the new skill SLOs?
8) Do I anticipate any barriers to making these changes?
    a) Those barriers may be personal, student-based, and/or institutional. What kind of support will I need to overcome them?
9) Who are good collaborators for further discussion at my institution?

Preliminary Glossary of Selected Terms in the Curriculum Map:

The terms here are defined in the context of their use in our Curriculum Map. We will be adding to this list in future versions of this tool.

abstraction	focusing on the important information only, ignoring irrelevant detail
algorithm	a step-by-step solution to the problem, or the rules to follow to solve the problem
computational thinking	conceptualizing problems in a way that allows one to create a manageable process to solve them, often using tools such as software and coding languages
decomposition	breaking down a complex problem or system into smaller, more manageable parts
pattern recognition	looking for similarities among and within problems
visualization	for the purposes of this tool, a visualization is a visual representation of a data set or statistical analysis including graphs/charts, tables, maps, etc.

Works Cited:

Angra A, Gardner SM. 2018. The Graph Rubric: Development of a Teaching, Learning, and Research Tool. *CBE - Life Sciences Education.* 17:ar65,1-18. https://doi.org/10.1187/cbe.18-01-0007

Brewer CA, Smith D. 2011. *Vision and Change in Undergraduate Biology Education: A Call to Action*. American Association for the Advancement of Science.

Emery NC, Crispo E, Supp SR, Farrell KJ, Kerkhoff AJ, Bledsoe EK, O'Donnell KL, McCall AC, Aiello-Lammens ME. 2021. Data Science in Undergraduate Life Science Education: A Need for Instructor Skills Training. *BioScience*. 71(12):1274–1287. https://doi.org/10.1093/biosci/biab107

Hampton SE, Jones MB, Wasser LA, Schildhauer MP, Supp SR, Brun J, Hernandez RR, Boettiger C, Collins SL, Gross LJ, Fernández DS, Budden A, White EP, Teal TK, Labou SG, Aukema JE. 2017. Skills and knowledge for data-intensive environmental research. *BioScience* 67:546–557. https://doi.org/10.1093/biosci/bix025