The Indiana Training Program in Public & Population Health Informatics

Data Science Curriculum
Competencies and Topics/Skills

I. Data generation, acquisition and management

1. Demonstrate an understanding of data wrangling
   - 1.1. Identify as well as access structured and semi-structured electronic data sets
     - 1.1.1. Use an API to create or gain access to a data set
     - 1.1.2. Download a public data set using data.gov or similar site
     - 1.1.3. Create datasets by extracting data from websites (e.g. using cURL)
   - 1.2. Transformation of raw data to formats more suitable for downstream use cases
   - 1.3. Apply data merging/linking and reshaping methods
   - 1.4. Assess data quality
   - 1.5. Develop an understanding of processing different kinds of data (e.g. string processing)
   - 1.6. Familiarize with a variety of tools that facilitate activities in competencies (I),1.1–(I),1.5

2. Develop skills related to database management (including SQL)
   - 2.1. Acquire an understanding of Relational Database Management Systems (RDBMS) and NoSQL databases
   - 2.2. Demonstrate the ability to create and modify tables as well as to run basic queries using SQL
   - 2.3. Develop an understanding of practices related to data persistence
   - 2.4. Explain the concept of indexing and apply an index to a table

3. High-performance Computing (HPC) and cloud computing
   - 3.1. Develop an understanding of high performance computing systems
   - 3.2. Develop an understanding of various cloud computing platforms and the use cases when these platforms may be used

4. Data representation
   - 4.1. Identify an appropriate data standard for a given data type
   - 4.2. Establish and use metadata to describe a data set
   - 4.3. Explain the concept of data provenance
   - 4.4. Assign a digital object identifier (DOI) to a resource and explain its purpose
   - 4.5. Explain the purpose of data standards and when to use them

II. Data Analysis

1. Apply techniques related to exploratory data analysis, data visualization and descriptive statistics
   - 1.1. Demonstrate conceptual understanding of probabilities and distributions.

1.2. Demonstrate the ability to describe data (find means, standard deviations, outliers, evaluate correlations etc.)
1.3. Acquire skills related to visualize data to discover patterns (including interactive visualization techniques)
1.4. Interpret findings from implementation of competencies (II),1.1–(II),1.3 using statistical software or object-oriented programming languages

2. Apply inferential statistical methods
   2.1. Demonstrate a conceptual understanding of regression
   2.2. Understand the model selection approach
   2.3. Implement knowledge from competencies (II)2.1, (II),2.2 using statistical software or object-oriented programming languages

3. Apply predictive analytics methods
   3.1. Demonstrate a conceptual understanding of:
      3.1.1. machine learning approaches
      3.1.2. dimensionality reduction approaches
      3.1.3. artificial intelligence and deep learning approaches
   3.2. Understand the model selection approach including cross-validation methods
   3.3. Implement knowledge from competencies (II)3.1–(II),3.3 using statistical software or object-oriented programming languages

III. Evidence generation and reproducibility

1. Develop an understanding of the scientific method, study design and quality of evidence
   1.1. Understand the scientific method and develop critical thinking abilities so as to facilitate generation of feasible research questions
   1.2. Demonstrate a conceptual understanding of various observational, quasi-experimental and experimental study designs
   1.3. Determine the quality of evidence based on the rigor and robustness of design
2. Understand the importance of transparency, replicability, reproducibility, and ethics
   2.1. Demonstrate the ability to accurately document and archive data analysis process to facilitate replicability as well as reproducibility
   2.2. Utilize various software tools which facilitate reproducibility and replicability

IV. Dissemination and Implementation

1. Acquire effective dissemination skills
   1.1. Apply technical writing and oral skills for effective communication and interpretation of quantitative analysis to the scientists and researchers.
   1.2. Apply technical writing and oral skills for effective communication and interpretation of quantitative analysis to industry as well as lay audiences
2. Demonstrate the ability to utilize data science methods to answer research questions

2.1. Utilize implementation science approaches to integrate tools and applications into practice settings
2.2. Evaluate research questions and evidence generated from above tools/applications