Guidelines for proposing courses for the data science minor

Proposing a course to be used as an elective

Any department in the university may offer courses that can serve as electives for the data science minor. Elective courses should focus on data science methods, applications of data science, or innovative uses of data. To be included as eligible electives, courses must meet the following criteria:

- The course must be second or higher level courses in the discipline and be at least 3 credits.
- It should complement the core requirements by either expanding on the methods or algorithms covered in the core requirements, and/or applying them to specific domains of inquiry.

Examples of courses that satisfy the elective requirement are:

- Courses designed to teach advanced methods and algorithms beyond the topics covered in the core requirements
- Courses designed to deepen students' programming skills in areas that are relevant to data science
- Courses with theoretical components aimed at helping students better interpret the results and implications of data science
- Courses with a meaningful practical component that exposes students to the types of data and analysis tools typically used in the field
- Methodological courses that cover specialized techniques and algorithms that are particular to a field of study
- Courses with theoretical components aimed at strengthening the connection between the quantitative results obtained through data science techniques and the theoretical understanding of the field

Proposing a course to satisfy a core requirement

All courses proposed to satisfy a core requirement should be at least 3 credits, and no higher than 4xx level. Additionally, they must fall into one of the three core requirements in the data science minor at Carolina:

- Data & Computational Thinking
- Data & Statistical Thinking
- Data, Culture & Society

There are no restrictions on which departments can offer courses satisfying any of the core requirements. However, in order for a course to satisfy a core requirement, it must cover a specific set of topics, as outlined by this guideline.

Topics for Data & Computational Thinking

This core requirement will provide students with an introduction to the computing tools and coding methods needed to gather, manipulate, visualize, and analyze data in an efficient manner. The list below does not imply a specific order in which the topics are to be covered. The course should be taught in Python and/or R.

Real-world data, computer data representations and data wrangling

- **Data in the real world***:* different data types in the world (names, temperature, area, gene snip, BMI etc.).
- **How computers see data:** scalar data types, vector data types, matrix and higher order data types.
- **Data structures**: Lists, arrays, data frames, libraries such as numpy, pandas and dplyr.
- **Reading and writing data:** data file formats, high-level functions to read specific data files such as CSV.
- **More on managing and manipulating data:** filtering data, data transformation, merging data sets, dealing with missing data.
- **Advanced data types:** language specific topics.

Programming fundamentals

- **Logical statements and Boolean operators:** True/False, combining booleans (truth tables).
- **Loops:** for and while loops.
- **Conditionals and control sequences***:* if-elseif and derivatives.
- **Problem description and decomposition:** mapping a question to a computational process, decomposing the computation process into steps.
- **Writing your own functions:** abstraction, detail hiding, mapping computation decomposition to functions.

Putting data to use

- **How to find and acquire publicly available data:** downloadable data, data access using APIs.
- **Data visualization:** various types of plots, graphs and image visualization; data visualization tools like GGPlot2 (R) and MATPLOTLIB (Python).
- **Examples from selected domains**.

Topics for Data & Statistical Thinking

This core requirement will provide students with an introduction to data driven statistical analysis, focusing on a hands-on approach to making inferences and predictions to learn from data. The list below does not imply a specific order in which the topics are to be covered. The course should be taught in Python and/or R.

Data Manipulation and Visualization

- **Data Types:** numbers, strings, comparisons.
- **Sequences:** arrays and tables.
- **Tables:** sort, select, group, join.
- **Visualizations:** bar plot, histogram, line plot, scatter plot, overlaid graphs.
- **Functions:** Defining functions, applying to tables.

Inference

- **Randomness:** iteration, simulation, probability rules.
- **Sampling:** types of samples, probability distribution, empirical distribution, empirical sampling distribution.
- **Testing hypotheses:** assessing models through simulation, making decisions with uncertainty, error probabilities, A/B testing.
- **Estimation:** percentiles, bootstrapping, confidence intervals.
- **Central limit theorem:** normal distribution, variability of the sample statistic.

Prediction

- **Regression***:* correlation, least squares, visual and numerical diagnostics, nonlinear and multiple regression models.
- **Regression inference:** bootstrapping for inference for the slope and prediction intervals.
- **Classification:** nearest neighbors, training and testing, implementing and determining the accuracy of a classifier, updating predictions, Bayes' Theorem.

Topics for Data, Culture and Society

The courses in this core requirement are expected to be centered in the theories and methods specific to the discipline in which they are taught; however, they should address the major questions below.

What is data and information?

- What are the foundational concepts of data and information in the field of study in which the course is taught?
- How have conceptions of data and information changed over time?
- How do these conceptions of data inform how knowledge is produced and valued in the field?
- How does the field communicate data to different audiences?

What counts as data?

- What (or who) count as data in the field of study in which the course is taught?
- How is data analyzed in the field of study in which the course is taught?
- How is data organized and categorized?
- What do these organizational systems and modes of analysis tell us about how value judgements are made?
- What are ethical standards of data use and analysis in the field?

What is the relationship between data and structures of power?

- How does data influence and/or shape structures of power?
- What are the political, social, economic, legal and/or cultural implications of how data is used?
- What is the relationship between individual privacy and data use?
- What is the relationship of notions of race, gender, class, sexuality and other markers of identity and conceptions of data?

How does data shape our lived realities?

- Using concrete case studies or examples, how does data impact the lived realities of people?
- Some possible areas of exploration include, but are not limited to social media, public policy, public opinion, archival structures, mapping systems, copyright, security, legal standards, economic infrastructures, business models, search engines, cultural production, new media, etc.