Dana Center Data Science Design Principles

May 13, 2022

Contents

Introduction to the workshop and the design principles

There has been dramatic growth in the teaching of introductory data science at two- and four- year colleges, with more anticipated in the future. Beyond content and curriculum, what course designs are effective in developing a "coherent and engaging" (Dana Center, 2021) introduction to data exploration? In this faculty development workshop, we will provide an overview of course design principles and practices that can help a diverse set of students succeed. We will utilize the Data Science Course Design framework from the Dana Center to illustrate how:

1. active learning,

2. growth mindset,

3. problem solving,
4. authenticity,
5. context and interdisciplinary connections,
6. communication,
7. technology, and
8. assessment can be organized within an introductory data science course.

Participants will have the opportunity to explore some activities and approaches, to discuss best practices, and to share ideas and approaches. Faculty from two-year colleges are particularly welcomed given the expected growth of courses and programs at the associate's level.

Our goal for the workshop is that participants will be able to:

- describe the data science course framework developed by the Dana Center
- identify for each of the principles a practice that can support that framework
- list specific ideas and resources for further study or exploration

The Course Design framework can be found at https://www.utdanacenter.org/sites/default/files/2021-05/data_science_course_framework_2021_final.pdf.

Design Principle: Active Learning

The course provides regular opportunities for students to actively engage in data explorations using a variety of different instructional strategies (e.g., hands-on and technology-based activities, projects, small group collaborative work, facilitated student discourse, interactive lectures).

STUDENTS WILL:

- Be active and engaged participants in discussion, in working on data explorations with classmates, and in making decisions about the direction of instruction based on their work.
- Discuss results of their data explorations with the instructor and/or classmates in class.
- Develop and evaluate data- based arguments.
- Think critically about data and be open to changing their mind after considering data-based arguments presented in class.
- Consider the implications of their conclusions within the context and as part of a broader picture, including consideration of data ethics.

TEACHERS WILL:

- Provide low-floor, high-ceiling activities and explorations that all students can access and that extend to high levels. Such activities should provide meaningful opportunities for exploration and co-creation of mathematical understanding and data literacy.
- Engage students through relevant contexts by providing local data sets and inviting students to ask questions about the data. Encourage different students to pose and investigate different questions that can be addressed by exploring data, and to come together to discuss findings.
- Ensure that all students are provided equitable opportunities to engage successfully throughout the course.
- Facilitate students' active learning of data science through a variety of instructional strategies, including inquiry, problem solving, critical thinking, and reflection.
- Create a safe, student-driven classroom environment in which all students feel a sense of belonging to the class and the discipline, are encouraged to take risks and embrace mistakes, and are able to make decisions about the direction for instruction through the results of their exploration of data science. Students' ideas are at the center of the conversation.

Design Principle: Growth Mindset

The course supports students in developing the tenacity, persistence, and perseverance necessary for learning data science, for using mathematics and statistics to tackle authentic problems, and for being successful in post-high school endeavors.

STUDENTS WILL:

- Make sense of data explorations by drawing on and making connections with their prior understanding and ideas.
- Persevere in solving problems and realize that it is acceptable to say, "I don't know what to do next," but that it is not acceptable to give up.
- Understand that productive struggle is valuable for brain growth and that times of struggle should be valued.
- Identify productive struggle and have coping mechanisms for destructive struggle.
- Reflect on mistakes and misconceptions to improve their mathematical understanding and data literacy.
- Seek help from different sources to move forward in their investigations, or be willing to start from a different perspective.
- Compassionately help one another by sharing strategies and solution paths rather than simply giving answers.
- Develop/strengthen a growth mindset to continue to apply in mathematics, data science, and other areas of their post-high school life.

TEACHERS WILL:

- Provide information about and model the importance of having a growth mindset.
- Facilitate discussions on the value of mistakes, misconceptions, and struggles.
- Demonstrate a growth mindset and value mistakes in their own experience with students.
- Provide students with low-stakes opportunities where they can make mistakes and learn from those mistakes.
- Give students time to struggle with tasks and ask questions that scaffold students' thinking without stepping in to do the work for them.
- Provide regular opportunities for students to self-monitor, evaluate, and reflect on their learning, both individually and with their peers.
- Encourage students to work beyond their comfort zone.

Design Principle: Problem Solving

The course provides opportunities for students to engage in the entire statistical problem- solving process.

STUDENTS WILL:

- Apply intuition, life experience, and previous learning to develop a strategy for solving unfamiliar problems.
- Explore and use multiple solution methods.
- Share and discuss different solution pathways and methods.
- Use tools and representations, as needed, to support their thinking and problem solving.
- Develop and justify their own strategies to approach new problems.
- Be willing to make and learn from mistakes in the problem-solving process.

TEACHERS WILL:

- Present tasks that require students to find or develop an approach that is appropriate for exploring data to reach a data-based conclusion.
- Provide data sets that allow for multiple exploration and visualization methods, including transfer of previously developed skills and strategies to new contexts.
- Provide opportunities to share and discuss different data analysis and visualization methods.
- Model the problem-solving process using various strategies.

- Encourage and support students to explore and use a variety of approaches and strategies to make sense of data and reach data-based conclusions.

Design Principle: Authenticity

The course presents data explorations that allow students to address relevant questions that arise in their communities.

STUDENTS WILL:

- Recognize specific ways in which mathematics and data are used in everyday decision making.
- Recognize questions that arise in the real world that can be addressed by exploring appropriate data.
- Contribute meaningful questions that can be addressed by exploring appropriate data.
- Identify bias and sources of bias in data, and describe the impact of bias in data on people and society.
- Experience in the process of collecting, cleaning, analyzing, and visualizing data to answer a data-based question of interest.

TEACHERS WILL:

- Provide opportunities to ask questions of data sets that are relevant to students, both in class and on assessments.
- Provide opportunities for students to ask questions about their school, community, or world that can be addressed by exploring appropriate data.
- Provide opportunities to investigate bias and the source(s) of bias in data and to discuss how bias impacts people and society.
- Provide students with real data, including data that require data processing and cleaning.

Design Principle: Context and Interdisciplinary Connections

The course presents data science in context and connect data science to various disciplines and everyday experiences.

STUDENTS WILL:

- Contribute personal experiences, where appropriate, that connect to classroom experiences.
- Actively seek connections between classroom experiences and the world outside of class.
- Describe connections between personal experiences or personal aspirations and the world outside the classroom through data analysis.
- Examine the ways in which data are collected in their day-to-day lives, and consider the ethics and consequences of collecting and using data to make decisions.

TEACHERS WILL:

- Provide opportunities for students to share their personal backgrounds and interests, including cultural and societal values, and help make the connection between what is important in students' lives and future aspirations, and what they are learning in data science.
- Provide real and interesting data sets, including those that are local to students.
- Invite students into data explorations that illustrate authentic applications.
- Provide data explorations that include applications from a variety of academic disciplines, programs of study, and careers, and which are culturally sustaining.

Design Principle: Communication

The course develops students' ability to communicate insights from their data explorations and findings in varied ways, including with words, data visualizations and numbers.

STUDENTS WILL:

- Present and explain ideas, reasoning, and representations to one another in pair, small- group, and whole-class discourse using discipline- specific terminology, language constructs, and symbols.
- Seek to understand the approaches used by peers by asking clarifying questions, trying out others' strategies, and describing the approaches used by others.
- Listen carefully to and critique the reasoning of peers using data to support arguments or counterexamples to refute arguments.
- Develop the skills to communicate data-based arguments with clarity and precision.
- Practice constructing data-based arguments with specific audiences in mind.
- Consider matters of accessibility in designing and executing their communications.
- Consider the pros and cons of various types of data visualizations for communicating with data in different situations.

TEACHERS WILL:

- Introduce concepts in a way that connects students' experiences to course content and that bridges from informal contextual descriptions to formal definitions.
- Clarify the use of data science terminology and symbols, especially those also used in different contexts or different disciplines.
- Engage students in purposeful sharing of data explorations and approaches using varied representations.
- Support students in developing active listening skills and in asking clarifying questions to their peers in a respectful manner that deepen understanding.
- Facilitate discourse by positioning students as authors of ideas who explain and defend their approaches.
- Provide regular opportunities for students to communicate with data using a variety of data visualizations.
- Scaffold instruction to support students in developing the required reading and writing skills.

Design Principle: Technology

The course introduces students to current technologies appropriate for data exploration and visualization, and prepares them to learn and use new ones.

STUDENTS WILL:

- Use technology to visualize data and support data-based conclusions.
- Understand the necessity of digital tools in cleaning and analyzing large data sets, and select appropriate tools for different situations.
- Develop experience in learning new tools, which will allow them to use emerging technology tools for analyzing data in the future
- Explore how technology can enhance data analysis as well as creativity in data visualization.
- Understand that the use of tools or technology does not replace the need for evaluating the reasonableness of conclusions or how the conclusions apply to a given context.

TEACHERS WILL:

- Introduce students to various data analysis and visualization technology tools that students can use beyond the classroom and support them in understanding the best uses for each tool.
- Facilitate student learning of technology platforms through exploration, as this will aid in transferring the knowledge to future platforms.
- Empower students to be creative and to use technology in support of their own goals
- Not be experts in the use of every platform, but are willing to experiment in response to students' questions and will model good practices for seeking answers to such questions.

Design Principle: Assessment

The course uses project- based assessments both as formative assessments and to evaluate student progress.

STUDENTS WILL:

- Assemble a collection of their work, which includes both data explorations that demonstrate understanding of the statistical problem- solving process and reflections on their learning process and their evolving understanding of the field of data science.
- At the end of the course, have a portfolio of data science work that showcases their knowledge of data science and their technology skills. This portfolio might be shared with a potential employer or educational institution.

TEACHERS WILL:

- Provide students with projects through which they are exposed to new content and can demonstrate their ability to use this new content to answer questions through exploration of appropriate data. These projects will include products that demonstrate student learning and will be part of students' portfolios.
- Evaluate student progress throughout the course by considering students' evolving portfolios as well as their reflections on their learning.
- In the final project of the course, allow students freedom to decide the topic and methods used in their data exploration, so that they can bring together the various skills they will have developed over the course, allowing the teacher to assess student progress.

Social Emotional, and Academic Development

Use collaboration and communication.

Students should recognize situations for which collaboration is an effective strategy, identify the features of collaborative work groups, and develop strategies for overcoming group work challenges. They should work collaboratively with students from various cultural and ethnic backgrounds while examining alternate points of view, accepting constructive criticism and revising personal views when evidence warrants.

Utilize resources to overcome obstacles

Students should engage in productive academic behaviors, including recognizing when help is needed with a task, and developing and applying a variety of strategies and sources for seeking help; monitoring and adjusting attitudes, emotions, and thoughts when facing challenging tasks or academic setbacks; and seeking and using feedback to improve performance.

Recognize and improve individual behaviors

Students should maintain motivation and persistence through a variety of strategies, including identifying and adjusting habits and beliefs that have interfered with success; applying metacognitive awareness to plan, monitor, evaluate, and reflect on their learning; and setting and monitoring goals.

Content Outcomes

Understand the role of data in the world.

Students need to demonstrate an understanding of data and the many different types of data that exist, including nontraditional data types such as photos, text, and sounds. They should understand the ways in which data are generated and collected, recognizing that primary data are collected directly by researchers from main sources while secondary data have already been collected and are readily available for use. It is crucial that students recognize the extent of their digital footprint and that it is based on the data created by their digital lives, while also considering issues surrounding data privacy and the ethical use of data. Examples of proficiency include the ability to distinguish between different types of data and between primary and secondary data; represent, summarize, and interpret data; recognize trends in data; and reflect on implications of data privacy and data use policies.

Ask data-based questions

In order to understand the vast information that can come from data, students should recognize the types of questions that can be answered through data exploration and be able to formulate their own questions. Students should determine what data might be collected to answer data-based questions and consider what questions can be answered from readily available data. Performing exploratory data analysis, drawing preliminary conclusions, and using what they have learned are key to formulating new questions for further exploration. Examples of proficiency include the ability to ask new questions based on what is learned from a data exploration and to identify the relevant data to address a data-based question of interest.

Collect and manage data

Sources of data are vast, so students should consider the different ways in which data might be generated, including through sampling (observational studies), experimentation (statistical experiments), and simulation, explaining the role of random selection in sample selection and the role of random assignment in statistical experiments. In order to appreciate the challenges of data collection, students should collect and organize data about their own lives and communities. They should identify bias and sources of bias in data, and describe how bias in data impacts people and society. While considering data collection, students should discuss the ethics and consequences of collecting and using data, including consideration of the bias that may be present in data collection or selection processes. Students should also acquire data in different formats and work with large, real-world, publicly available data sets. Combining two or more data sources might be necessary to investigate a question of interest. Students should become familiar with different types of data structures such as arrays, stacks, and queues, understanding that data (both traditional and nontraditional) are not always collected/shared/received in a form that is ready for analysis and often require the use of different digital tools to clean and prepare data for analysis (e.g., merge data sets, deal with incomplete data, normalize data, create new variables). Students should explore the basics of programming as needed, and be comfortable editing and documenting code, or finding the appropriate tools to transform the data to be useful in their own data analysis. Examples of proficiency include the ability to distinguish between observational studies and statistical experiments; explain why random selection is important in observational studies and why random assignment is important in statistical experiments; use data scraping to obtain data from an online source; use transformations to create new variables for analysis; and clean data sets to address extraneous, incorrect, or missing data values in preparation for analysis.

Explore data to make sense of and represent the story that the data are telling

Interpreting and critiquing data visualizations are fundamental skills for building data acumen. Students should use technology, and programming where appropriate, to create a variety of data visualizations to explore data and to share insights based on what the data reveal. Students should be able to look for patterns, describe data distributions, and compare distributions while examining graphical displays. They should also be able to create graphical displays, data visualizations, and tables to explore relationships. Technology should be used to explore correlation between two numerical variables visually and numerically, while tables and conditional relative frequencies should be used to explore associations between categorical variables. Interpreting graphical displays, data visualizations, and tables using more than two variables allows students the opportunity to demonstrate multivariable thinking. Examples of proficiency include the ability to draw insights from a data visualization and to communicate those insights to others; use and interpret graphical displays and tables to describe relationships between two variables; and create and interpret data visualizations that demonstrate multivariable thinking.

Analyze data to create data-based arguments and to reach data-based conclusions

Students need to understand the importance of communicating with data and making data-based arguments. In order to communicate with data effectively, students should use appropriate summary measures to describe data distributions and to compare data distributions. They should understand that variability is present in data and take sampling variability into account when formulating data-based arguments or making data-based

decisions. Students should combine their knowledge of probability, technology, and programming where appropriate, to construct simulations to estimate probabilities and to assess statistical significance. Since conclusions based on sample data are subject to misinterpretation, students should acknowledge potential errors and their possible consequences in the data collection process. Examples of proficiency include the ability to describe data distributions, including shape, center, variability and any unusual features for numerical data distributions; compare two or more data distributions using graphical and numerical summaries; design and implement a simulation to assess statistical significance and interpret the results of the simulation; contrast statistical significance and practical significance in a given context; describe potential errors and possible consequences of a data-based argument; and interpret a margin of error in context.

Understand limitations of data sources and data-based conclusions

When working with publicly available data sets, students should be able to determine if conclusions are appropriate based on the study design and the way in which the data were collected. They should evaluate and critique data-based claims and arguments, understanding that data and data-based arguments may have inherent sources of bias and they should seek to identify them. Examples of proficiency include the ability to evaluate whether a conclusion from a data analysis or exploration is appropriate given the data source and data collection method; evaluate a data-based argument and identify potential sources of bias; and critique a data-based argument.

Use data to make predictions

A key component in using data is the ability to make predictions accurately. Students use data to build models (including linear models, nonlinear models, and models with multiple predictor variables) to describe relationships between variables. They should evaluate the appropriateness and usefulness of prediction models, while also using models to make their own predictions. Examples of proficiency include the ability to use graphical displays to make informal predictions; fit linear and nonlinear models and evaluate the usefulness of models; and use fitted models to make predictions.

Use data to inform decision making

The use of probability, including conditional probability, to make decisions and to quantify uncertainty about real-world situations is necessary. Students need to be able to understand and interpret results from classification and decision tree algorithms. Examples of proficiency include the ability to calculate and interpret probabilities, including conditional probabilities; estimate probabilities empirically and by using simulation; use relevant probabilities to inform a decision; and use a given classification or decision tree to reach a decision, describing the steps in the process of reaching that decision.