Dispositions and Responsibility

This strand focuses on what data is and all of the ways students should think about and frame it as a concept and tool.

The nature of data is complex, diverse, and humanistic. When engaging with data you must consider the form it takes, where it can come from, and what it can and should be used for. Working with data is non-linear and often raises new questions while seeking answers to others. Additionally the data process is influenced at all stages by the humans working with it which can lead to biases and concerns about ethics and responsibility. However, data can also be powerful for supporting the advancement of discovery or enactment of change.

Substrand A1

Nature of Data

The nature of data is complex, variably, humanistic, and often incomplete. Data can take many forms and may come from many different sources. Additionally, data is integral to the field of AI.

Concept A.1.1

Data types and forms

Recognize that data can exist as quantitative, ordinal, categorical, and other values. Data also can be “nontraditional” forms such as graphical or other media.

A.1.1a

Analyze the way categorical and numeric data shapes its interpretation and analysis.

A.1.1b

Recognize that numerical variables may be either discrete or continuous.

Concept A.1.2

21st-century skills

AI literacy

Data are produced by people

Recognize that data represent decisions about measurement and inclusion involving people who are and are not immediately present.

A.1.2a

Ask questions regarding the origins of specific automated measures (e.g., webtracking, email meta-data, user accounts).

A.1.2b

Recognize the limits of the information the data can provide and the story it can tell.

A.1.2c

Recognize that conclusions may need to be revised in the future as more knowledge and data become available.

Concept A.1.3

Variability of data

Recognize that variability is a foundational component of data.

A.1.3a

Make sense of the variability of data through an iterative process of refinement by questioning.

Concept A.1.4

Data provides partial information

Recognize that data captures certain aspects of a model of a target phenomenon or set of objects in the world but does not represent it completely.

A.1.4a

Specify ways that data provide incomplete information relative to the object being studied.

A.1.4b

Approach data and evidence-based claims with reasonable skepticism and apply the process of evaluating the validity of claims while remaining open-minded.

Concept A.1.5

Data and AI

Recognize that data “fuels” AI, that AI can be compared to a function machine (math), algorithm (CS), or a prediction model (statistics) that relies on data to both operate and improve itself, and that AI tools can also be used to analyze complex data in research.

A.1.5a

Describe in plain language how AI uses and builds upon data in multiple ways. e.g., AI systems identify patterns in data by processing thousands of input-output pairs, and the system adjusts its internal mathematical model to minimize error, enabling it to predict outputs for new inputs such as a spam filter

A.1.5b

Identify how issues in data, such as bias, missing data, and errors, can affect the output of an AI tool and the training of an AI tool from the input-output pairs it learns from. e.g., If AI only sees pictures of cats in sunlight, it would fail to recognize cats in shadows

Substrand A2

Data Ethics and Responsibilities

The data process is influenced at all stages by the humans working with it which can lead to concerns about ethics and responsibility. It is important when working with data to consider the use risks as well as the benefits. Data can be powerful for supporting the advancement of discovery or enactment of change.

Concept A.2.1

Data use risks and benefits

21st-century skills

Durable skills

Recognize that data can pose risks but also benefits for individuals and groups, and understand its potential uses, limitations, and risks, including unintended consequences.

A.2.1a

Describe how social groups can be inadequately represented by existing data and data schemes. e.g., city planners using traffic data collected from weekday commuters overlook nighttime workers' needs, such as poor bus schedules for nurses on night shifts

A.2.1b

Acknowledge that options and choices are available for data collected about individuals, and recognize that what is gathered or excluded can have consequences. e.g., a survey claiming "most teens love math" is biased if only the math club members completed the survey

Concept A.2.2

Biases in data

21st-century skills

Durable skills

Recognize all data contains bias but data collection and analysis methods can increase or mitigate the effects of biases.

A.2.2a

Identify how biases in data affect inferences and questions.

Concept A.2.3

Power of data

21st-century skills

Media literacy and digital citizenship

Recognize data empowers discovery, decision-making, and advocacy across fields.

A.2.3a

Analyze how data is used to solve problems, persuade, and discover new ideas.

Substrand A3

Investigative Dispositions

Working with data is non-linear and often requires cycling between phases in various orders multiple times. The process of investigating with data often raises new questions while seeking answers to others. Additionally, data is influenced by the humans working with it and the contexts within which they work.

Concept A.3.1

The investigative process

21st-century skills

Durable skills

Recognize that making sense with data requires engaging with it in a particular way that includes combinations of the concepts and practices in the other four strands.

A.3.1a

Investigate real-world questions by cleaning, analyzing, and interpreting data to draw conclusions.

Concept A.3.2

Iteration

21st-century skills

Durable skills

Recognize that the investigative process is not linear but cyclic and iterative, with many of the phases repeating and looping back.

A.3.2a

Recognize that the investigative process is non-linear, often cycling between phases in various orders multiple times.

Concept A.3.3

Dynamic inferences

21st-century skills

Durable skills

Recognize that inferences from data are dynamic, evolving with new data and additional analysis.

A.3.3a

Revise initial conclusions when new data emerges and use evidence to support claims.

Concept A.3.4

Apply context

Recognize that the context surrounding the data and the investigation shapes interpretation. Many fields (biology vs. psychology; economics vs. sociology) have created very different frameworks to organize problems. Considering multiple approaches may reveal useful insights from the same data.

A.3.4a

Explain data interpretations from various disciplinary and community perspectives (e.g., social studies, families).

Concept A.3.5

Student data agency

21st-century skills

Durable skills

Cultivate the motivation to engage with data in all areas of life and understand how data impacts your own experiences.

A.3.5a

Embed data practices into everyday life and advocate for the benefits of doing so.

A.3.5b

Assess the accuracy, perspective, credibility, and relevance of various resources (e.g., information, media, data).

Creation and Curation

This strand focuses on where data comes from and how it should be collected, organized, and formatted in order to make it useful.

Data collected from real world scenarios is often complex and messy, and whether it is collected first hand, or retrieved second hand from an external source, it requires curation and cleaning before analysis. The context of data collection matters and affects the nature of errors in data collection. The methods and decisions made during data collection affect the usefulness of the data and its ability to answer different questions.

Substrand B1

Organization and Processing

In order for data to be useful for analysis and visualization, it often needs to be organized and formatted in particular ways. Organization can include both procedural cleaning up of errors or mistakes and processing or transforming the data through calculations and logic statements to create new or summative measures.

Concept B.1.1

Data cleaning

Identify and address data quality issues to ensure accuracy and reliability, progressing from simple error identification to using systematic approaches.

B.1.1a

Informally identify anomalies and outliers in a distribution of data and make an informed decision as to whether those observations should be removed or filtered out for analysis.

Concept B.1.2

Organizing and structure

Organize raw data into structured formats using categories, tables, and systematic recording methods.

B.1.2a

Use categorical variables or bins/groups of numerical variables in a dataset to restructure data into groups.

B.1.2b

Make sense of and use a dataset arranged in nested or hierarchical format.

Concept B.1.3

Processing and transformation

Transform and manipulate data through sorting, grouping, filtering, and combining datasets.

B.1.3a

Use existing numerical variables to create bins or groups based on benchmark values appropriate for the context, or bins based on numerical ranges (e.g., 0-4, 5-10, 11-15, etc...).

B.1.3b

Create a new variable from an existing variable that transforms (e.g., uses a formula to convert units of measure) or recodes data (e.g., blue-->B, red--> R).

Concept B.1.4

Summarizing groups

Calculate and analyze group-level statistics from detailed data to reveal patterns and relationships.

B.1.4a

Use summary measures of groups within a nested or hierarchical dataset.

Substrand B2

Designing for Data Collection

The design of a data investigation is as important as the data collection process. Framing a data-based investigation requires identifying a problem or question to be explored. Additionally, the methods must be carefully chosen and the values and tradeoffs considered.

Concept B.2.1

Designing data-based investigations

Identify problems and formulate questions that guide meaningful data collection and analysis.

B.2.1a

Construct data-based questions that explore relationships between variables and consider how data collection methods affect the quality of evidence.

Concept B.2.2

Data creation techniques and methods

21st-century skills

AI literacy

Explore various ways to generate data through simulations, sensors, and automated collection methods.

B.2.2a

Use data generated from simulations and models to investigate a question of interest.

B.2.2b

Deploy or trigger sensors or automated data collection methods and use the generated data to investigate a pre-defined problem or question.

Concept B.2.3

Creating data collection plans

21st-century skills

Durable skills

Develop systematic plans that specify what data to collect, how to collect it, and from what sources to answer investigation questions.

B.2.3a

Design data collection plans that include data types needed, collection methods, sample sizes, and timing considerations for investigations.

Concept B.2.4

Finding secondary data

21st-century skills

Media literacy and digital citizenship

Explore, locate, evaluate, and retrieve datasets collected by others to address research questions and data investigations.

B.2.4a

Search for and retrieve appropriate datasets from educational repositories and curated sources designed for middle school investigations.

B.2.4b

Evaluate potential datasets based on relevance, timeliness, and credibility of the source for answering specific questions.

B.2.4c

Use metadata and documentation to understand the context and limitations of secondary datasets.

Substrand B3

Measurement and Datafication

The methods and decisions made during data collection affect the usefulness of the data and its ability to answer different questions. It is important to consider the potential effects of methodological decisions when collecting data and to determine the methodological decisions made by others when using secondary data. It is also important to consider ethical practices of using other's data.

Concept B.3.1

Creating your own data

Collect, measure, and document data accurately using appropriate tools and methods.

B.3.1a

Create an ordinal scale of measurement.

B.3.1b

Understand that data is information collected and recorded with a purpose.

B.3.1c

Distinguish between human-derived data from images, sounds, and text vs. computer-derived data from images, sounds, and text.

Concept B.3.2

Working with data created by others

21st-century skills

Durable skills

Evaluate and interpret others' datasets by examining collection methods, context, and quality.

B.3.2a

Consider how the data were measured, with what tool and precision.

B.3.2b

Consider who collected these data and for what purpose.

B.3.2c

Consider when and where the data were collected.

Concept B.3.3

Ethics of data collection and usage

21st-century skills

Durable skills

AI literacy

Collect and use data ethically, considering privacy, fairness, and potential impacts.

B.3.3a

Design data collection methods that address privacy, consent, and fair representation of different groups.

B.3.3b

Examine historical examples of harmful data practices to inform ethical data use.

Substrand B4

Complexity of Data

Data collected from real world scenarios is often complex across many dimensions including messiness, size, and structure. In order to be able to work with authentic real-world datasets of high complexity, these dimensions must be scaffolded such that increasingly higher levels of complexity are encountered as one approaches mastery.

Concept B.4.1

Cleanliness

Work with datasets at increasing levels of cleanliness and identify how datasets need to be curated to address messiness issues.

B.4.1a

Identify and handle missing values marked by special codes (-99) or blank cells.

B.4.1b

Distinguish between true zero values and blank cells.

Concept B.4.2

Complexity of variables

Explore datasets containing various types of data and understand how each type serves different analytical purposes.

B.4.2a

Work with datasets that include rates and derived variables that combine multiple measurements.

B.4.2b

Work with datasets that have multiple variables that can suggest or answer different questions.

B.4.2c

Work with datasets that show natural variation and understand why values differ.

Concept B.4.3

Size

Work with datasets of increasing size in both number of observations and variables and arrange data in increasingly complex formats to facilitate meaningful analysis.

B.4.3a

Work with datasets with up to 20 variables and over 100 observations.

B.4.3b

Understand how categorical variables can be used to create meaningful subsets.

Concept B.4.4

Complexity of structure

Manipulate and combine data in increasingly complex ways to reveal new insights and patterns.

21st-century skills

AI literacy

B.4.4a

Work with datasets where the row isn't a single observation but something more complex (e.g., average, nested cases).

B.4.4b

Work with datasets that include derived or transformed variables, including creating categorical variables from numerical data.

B.4.4c

Understand how categorical variables can be used to create meaningful subsets.

Analysis and Modeling Techniques

This strand focuses on the process of analyzing data.

Analyzing data includes many different techniques such as examining single and multi-variable patterns, measures of centrality, variability, and uncertainty. Knowing which techniques to use on which types of data to answer which questions is as important as the skills to conduct analysis techniques. Additionally, understanding simulation and the relational nature of data is important to the analysis process, as is the use of technological tools for analysis and modeling.

Substrand C1

Summarizing Data

Raw data often is not useful for answering questions, making claims, or telling a story. In order to derive understanding it is usually useful to have a summary of the data which provides measures of the centrality, spread, and shape of the dataset.

Concept C.1.1

Measures of center

Analyze large datasets by measuring their central tendency while considering the context and distribution of the data.

C.1.1a

Identify measures of center as statistical values that represent the central tendency of data sets.

C.1.1b

Explain what measures of center are useful for and their limitations.

Concept C.1.2

Measures of spread

Examine dataset variability by applying measures of spread to identify and quantify outliers.

C.1.2a

Categorically identify the presence of potential outliers in a dataset.

Concept C.1.3

Shape

Identify the distribution of data points, including clusters, gaps, symmetry, skewness, and modes. Use these patterns to understand data spread and their impact on measures like the mean and median.

C.1.3a

Describe whether data is symmetric or asymmetric and the number of modes.

Concept C.1.4

Frequency tables

Organize data into frequency tables based on shared characteristics. Summarize data using counts, fractions, relative frequencies, or proportions to enable comparisons and generalizations. Understand the implications of choices made when creating and interpreting frequency tables.

C.1.4a

Generate a frequency table to summarize raw categorical data.

Concept C.1.5

Missingness

Identify and describe missing data numerically and categorically. Distinguish between missing values and true zeros. Understand how missing data impacts relationships, patterns, and models in data interpretation.

C.1.5a

Numerically measure missing data.

C.1.5b

Recognize the difference between the absence of data, and "zero."

Concept C.1.6

Metadata

21st-century skills

AI literacy

Recognize metadata as information about data, including its source, type, and structure. Use metadata to organize, summarize, and analyze data effectively, supporting interpretation and decision-making.

C.1.6a

Comprehend, in an informal sense, the value of information contained in metadata (e.g., data and time, text, continuous, geolocation).

Substrand C2

Identifying Patterns and Relationships in Data

A primary use of data is in understanding patterns and relationships across different variables and scenarios. As all data contains variability it is important to understand and analyze distributions both within and across variables.

Concept C.2.1

Comparing variables

Identify similarities and differences between variables and explore potential associations. Use distributions, numerical summaries, and simulations to compare groups based on numerical or categorical data.

C.2.1a

Use reasoning about distributions to compare two groups based on quantitative variables.

Concept C.2.2

Understanding distributions

Represent data visually and numerically to describe how outcomes occur and compare groups. Use variability to interpret distribution shape, support statistical reasoning, and assess population estimates.

C.2.2a

Represent the variability of numerical variables using appropriate displays (e.g., dotplots, boxplots).

Concept C.2.3

Defining relationships

Organize, visualize, and analyze data to identify patterns, trends, and associations. Use statistical measures and graphs to interpret relationships and make predictions.

C.2.3a

Employ complex graphs (e.g., bar graphs, line graphs) and basic statistical concepts (e.g., mean, median, mode) to describe patterns and identify trends, similarities, and differences within data.

C.2.3b

Create scatterplots and add line of best fit.

Concept C.2.4

Analyzing non-traditional data

21st-century skills

AI literacy

Examine data beyond numbers, including sounds, textures, and text. Categorize sensory inputs, track word frequencies, and analyze data from sensors and IoT devices to identify patterns and trends.

C.2.4a

Compare word frequencies across multiple texts to identify patterns and create simple visualizations from that text data.

C.2.4b

Explore patterns in audio data (e.g., analyzing sound waves for volume and frequency).

Concept C.2.5

Machine learning

21st-century skills

AI literacy

Use data to build decision trees, explore classification and clustering, and understand how machine learning optimizes predictions through algorithms like gradient descent.

C.2.5a

Learn to use simple diagrams (e.g., decision trees using small relatable examples) to make important decisions for everyday choices.

Substrand C3

Variability in Data

Variability is omnipresent within data and datasets. Working with data depends on understanding, explaining, and quantifying variability of all forms (variability within a group, between different groups, or between samples).

Concept C.3.1

Describing variability

Identify differences within data by sorting, grouping, and organizing characteristics. Use statistical and simulation methods to represent and analyze variability, connecting it to real-world uncertainty and probabilistic processes.

C.3.1a

Identify probabilistic processes that simulate various forms of categorical variability, including uniform and normal distributions. e.g., spinner, dice, random draw

C.3.1b

Illustrate variability in a dataset by determining how key descriptive features are represented.

C.3.1c

Evaluate how visualizations, models, or predictions account for variation at an appropriate level.

Concept C.3.2

Comparing variability

Examine differences between groups by analyzing measures of spread, such as range and standard deviation. Utilize visualizations like box plots and apply statistical methods, including mean, median, and standard deviation, to compare datasets, assess variability, and uncover patterns in data distributions and models.

C.3.2a

Use visualizations (e.g., box plots) to compare variability across datasets.

Concept C.3.3

Understanding sources of variability

Recognize measurement errors and natural variability in data. Assess data quality, identify outliers, and refine models using statistical and contextual analysis.

C.3.3a

Consider both context and the characteristics of a dataset to determine whether a given data point is reasonable. e.g., meaningful outliner, erroneous outlier

C.3.3b

Relate sources of variability to domain-specific phenomena as described in the relevant domain standards. e.g., Next Gen Science Standards, Mathematics Common Core State Standards, C3 Framework

Concept C.3.4

21st-century skills

AI literacy

Variability in our computational world

Explore how AI model outputs vary based on training data, labeling, and bias. Understand how generative AI and pre-trained models use large datasets to make inferences and how variability in data impacts outcomes.

C.3.4a

Conceptualize how the output of AI models such as LLMs vary along a variety of dimensions.

C.3.4b

Determine how labeling happens and how it affects the variability of the output of models. e.g., training set that labels dogs vs. cats, consider connections to bias

Substrand C4

Digital Tools of Data Analysis

While some datasets can be explored by hand, as they get bigger and more complex it becomes necessary to use digital tools for analysing data. It is important to understand which tools to use for which application or scenario, the affordances and tradeoffs, and the ethical considerations of using certain tools.

Concept C.4.1

Tool application

Use digital tools to summarize data and create visualizations. Apply these tools to identify patterns, clean and prepare data, perform analysis, and build models for simulations to explore relationships and trends.

C.4.1a

Summarize data across multiple categories using a digital tool.

C.4.1b

Create single variable visualizations using a digital tool.

C.4.1c

Identify relationships and patterns using a digital tool.

Concept C.4.2

21st-century skills

AI literacy

Tool ethics

Examine how digital tools influence access, privacy, and bias, shaping opportunities and challenges in technology use. Consider the broader ethical and societal impacts of AI, including its role in decision-making, accountability, and policy.

C.4.2a

Describe how digital tools can be used to provide equitable access to learning experiences.

Concept C.4.4

Tool selection

Choose the appropriate no-code, low-code, or high-code digital tool based on the task. Use multiple tools throughout the data investigation process and explore how digital tools are applied in the workforce.

C.4.4a

Select a no-code or low-code digital tool that is suited for the intended task.

Concept C.4.5

The role of code in data analysis

21st-century skills

AI literacy

Explore how block coding and computer code automate and enhance data analysis. Understand how coding enables reproducible processes and compare its advantages and limitations to no-code and low-code tools.

C.4.5a

Explore the basics of block coding in data investigation processes.

C.4.5b

Explore the basics of block coding in data analysis processes.

Concept C.4.6

Tool accessibility for diverse learners

Understand how digital tools can support a broad range of diverse learners. Evaluate their effectiveness and impact, and explore inclusive data representations.

C.4.6a

Explore why it's important to present data in multiple formats to ensure all learners can understand it.

C.4.6b

Explore how digital tools support diverse learners to analyze data. e.g., immersive readers, speech-to-text, translators, sonification

C.4.6c

Discuss how the lack of accessible digital tools can exclude people from participating in data analysis.

Substrand C5

Models of Data

Interpreting, creating, and using models is a central component of working with data. Models are both a way to analyze data and a source of data.

Concept C.5.1

Understanding modeling

21st-century skills

AI literacy

Analyze patterns and relationships in data using graphs, tables, and models. Explore tools like decision trees and neural networks, assess assumptions, and distinguish correlation from causation in real-world contexts.

C.5.1a

Explore how relationships in data connect characteristics, including patterns like increasing, decreasing, or no connection.

C.5.1b

Recognize that models simplify complex systems and have limitations.

C.5.1c

Recognize that relationships in data do not always imply causation. e.g., ice cream sales and shark attacks both increase in the summer, but one doesn't cause the other

Concept C.5.2

Creating models

21st-century skills

AI literacy

Develop an understanding of patterns and relationships. Use data and technology to build and refine models. Advance these skills by constructing complex models that incorporate multiple variables, assess assumptions, and improve predictions.

C.5.2a

Identify relationships between variables and represent them using tables, graphs, or diagrams (e.g., decision trees, flowcharts).

C.5.2b

Use simple mathematical or computational models (e.g., statistical summaries, spreadsheet formulas) to describe patterns and relationships in data.

C.5.2c

Test and refine models by comparing predictions to actual data values.

Strand D

Tools and Techniques

Interpreting Problems and Results

This strand focuses on justification and explanation of reasoning when making inferences, claims, or suggestions from data within the context and processes of the dataset collection and analysis.

An important component of interpreting results is understanding the relationship between questions, problems and datasets. Formulating a strong question or identifying a problem that can be addressed with data affects the opportunities for interpretation and results from the data. Additionally, the applicability of inferences and claims that are made are constrained by the sample, population, and context of the data.

Substrand D1

Making and Justifying Claims

As all data contains variability, it is important to use probabilistic thinking and language when making claims from data. This requires paying attention not only to patterns and comparisons within and across variables but also such things as expected and prior values, sample sizes, and significance.

Concept D.1.1

Probabilistic language

21st-century skills

Durable skills

When communicating with others, employ both plain-language and clear vocabulary to regularly describe degrees of uncertainty, both formally and informally as a thinking habit.

D.1.1a

Express a finding and quantify your confidence in it by stating the degree of certainty regarding the result. e.g., I am highly confident that a majority of students in my area ride the bus to school, based on separate sources' estimations of 60%, 62.3%, and 65% of students ride the bus to school

Concept D.1.2

Priors and updates

21st-century skills

Durable skills

Media literacy and digital citizenship

When encountering new data, integrate probabilistic thinking into everyday situations by explicating prior assumptions and the impact of new data / evidence on those assumptions.

D.1.2a

Recognize that an assumption should change somewhat, but may not need to change entirely, based on the "strength of" or degree of confidence in new evidence.

D.1.2b

Connect previous assumptions about a problem to the level of certainty in a finding by using the terms “prior assumption,” “new data/evidence," and “my updated assumption.”

Concept D.1.3

Expected value

21st-century skills

Financial literacy

When making a decision about uncertain outcomes in the future, integrate probabilistic thinking into everyday decisions by applying expected value (magnitude x probability) to appropriate situations.

D.1.3a

Numerically compare the impact of two events with different magnitudes and probabilities to determine which scenario is preferable. e.g., financial problem; 10% chance of receiving $100 vs. 50% chance of receiving $50

Concept D.1.4

Explaining significance

21st-century skills

Media literacy and digital citizenship

Clearly describe the basic logic of statistical significance to others, differentiating between significance, the size of an effect, and the statistical power of an analysis. Recognize what statistical significance can reveal and cannot reveal about a phenomenon.

D.1.4a

Recognize and describe random chance in a given situation, and explain whether a result is unusual by comparing it to what is expected from random chance. e.g., flipping a coin 10 times and getting 8 heads is less common than 5 heads and 5 tails, but still possible due to random chance

D.1.4b

Recognize that a unique result may be considered significant if it is substantially different from outcomes in similar situations.

D.1.4c

Recognize that a unique result may be considered significant if it falls far from the typical range of outcomes in a visualized distribution of results.

Concept D.1.5

Sampling and simulation

Comfortably identify the purpose of sampling and simulation for making arguments about data, and employ techniques using software to differentiate a real-data result from random chance or “happenstance.”

D.1.5a

Evaluate how different sampling methods impact the accurate representation of a population and their ability to generalize findings to other groups.

D.1.5b

Assess how sample size impacts the accuracy of estimates representing population characteristics.

D.1.5c

Identify the sources of potential bias in a sample or population, and describe how bias may impact the results of an investigation.

D.1.5d

Describe what it means for an event to be likely or unlikely using probability. e.g., probability of 0 is unlikely, 1 is very likely, 1:2 is neither likely or unlikely

Concept D.1.6

Correlation versus causation

21st-century skills

Media literacy and digital citizenship

Comfortably separate correlation from causation in a wide variety of situations, building a “first-reaction” thinking habit over time.

D.1.6a

Assess simple data to identify potential associations between two variables while considering that correlations do not imply causation and may arise from unobserved factors.

D.1.6b

Using graphical displays, identify and categorize the type of potential relationship between pairs of numerical variables with terms such as independent, dependent, and covariate.

Concept D.1.7

Randomization

When identifying a potential cause of a phenomenon, clearly describe the usefulness of randomization for constructing an argument with data.

D.1.7a

Explain why randomization is an effective way to reduce other potential influences, and as a result, successfully isolate the impact of an independent variable.

Concept D.1.8

21st-century skills

Durable skills

Multi-variable decision-making

Clearly describe how to leverage additional variables or additional outside data to make a logical argument, and identify potential risks of overdoing it.

D.1.8a

Distinguish direct vs. inverse relationships in multivariate data, such as associations between two categorical groups within the same visualization.

D.1.8b

Use color to differentiate categories in a scatterplot and identify patterns in their relationships.

D.1.8c

Calculate and compare the slopes and intercepts of multiple trend lines within the same graph to analyze differences between categories and their relationships.

Substrand D2

Problem Identification and Question Formation

Formulating a question or identifying a problem that can be addressed with data affects the opportunities for interpretation and results from the data. The ability to make and justify strong claims relies on identifying questions that are testable and can be answered with data. Additionally, identifying the uncertainty or limitations within the problem space is an important component of formulating conclusions

Concept D.2.1

Verifiable questions and statements

Identify and create the type of questions that can be answered by data, and are eventually verifiable using a combination of modeling and experimentation.

D.2.1a

Ask or identify a question that can be verified with data collected through observations.

D.2.1b

State a guess or potential answer to a question for later verification or testing via a hypothesis.

Concept D.2.2

21st-century skills

Durable skills

Iteration, validation, and multiple explanations

Regularly practice identifying alternative explanations for a result from data, both for interim steps and post-analysis conclusions.

D.2.2a

Predict whether the variability of one variable tends to increase or decrease in relation to another variable, providing evidence and reasoning to support the prediction.

D.2.2b

State a prediction or answer to an investigation question at the beginning, midway, and at the end of the analysis exercise while asking why this may be true each time.

Concept D.2.3

Uncertainty statements and limitations

Clearly explain the limitations and caveats of a conclusion from data, including the risks of extending the conclusion to another group or situation.

D.2.3a

Assess the data to determine which aspects of the original question can be answered and identify which areas still require further investigation for a confident conclusion.

Concept D.2.4

Relevant conclusions

21st-century skills

Durable skills

Ensure that increasingly complex analysis steps remain useful for the original question, and that the method does not distract from the problem.

D.2.4a

Generate an original statement that answers the original investigation question in a direct way and provides relevant statistical data to support one's statistical conclusion.

D.2.4b

Identify a statement that does NOT answer the original investigation question in a direct way and provides relevant and sufficient data to support one's statistical conclusion.

Substrand D3

Generalization

Though there is often an instinct to use data to make large generalized claims, the applicability of inferences and claims that are made are constrained by the sample, population, and context of the data.

Concept D.3.1

Application fitness

21st-century skills

Media literacy and digital citizenship

AI literacy

Regularly identify generalization issues, with frequent comparisons between significant real-world examples and a current analysis.

D.3.1a

Identify various factors that may cause data in a dataset to insufficiently represent or apply to other situations.

D.3.1b

Identify characteristics of data-based predictions that easily and do not easily generalize to many situations.

Concept D.3.2

Sample versus population

21st-century skills

AI literacy

Given a dataset, identify constraints and opportunities for what can be logically inferred about a broader population.

D.3.2a

Evaluate a population based on a sample by making informal arguments for the sample's sufficiency in answering the question.

D.3.2b

Identify potential weaknesses in a given sample that may limit its ability to represent a broader population or phenomenon.

Concept D.3.3

Sample size

When full information is hidden or inaccessible, recognize the logical relationship between a sufficient number of chances and a sufficiently large sample to reasonably represent something.

21st-century skills

Media literacy and digital citizenship

D.3.3a

Recognize that a sample must be sufficiently large to well-represent a broader population, based on the concept of the Law of Large Numbers. e.g., flipping a coin 10 times might give 7 heads, but 1000 flips will trend towards 50/50

D.3.3b

Identify examples of too-small sample sizes in the media or other real-world examples. e.g., medical drug trials, prior debunked research

Concept D.3.4

Simple bias

When information is completely hidden or unavailable, be aware of possible underlying issues in the sample and apply strategies to identify and address them.

21st-century skills

Media literacy and digital citizenship

D.3.4a

Acknowledge that a sample may be systematically skewed due to collection methods, data availability, survey design, or other factors, as demonstrated in a direct data collection activity.

Concept D.3.5

Extension statements

Following an initial analysis, list and implement opportunities for increasing the strength of an argument, a generalization claim, or ideas for a new analysis. Explore risks of the same approaches as well.

21st-century skills

Durable skills

D.3.5a

Identify additional possible scenarios for which a data-based conclusion may apply, beyond the original question or inquiry.

Concept D.3.6

Subset effects

Recognize that important information may be hidden or may even change a major conclusion when data is filtered into categories and/or groups.

21st-century skills

Media literacy and digital citizenship

D.3.6a

Summarize variables in a dataset with measures of central tendency with both the full data and with subsets (e.g., occupation, race, gender, income, zipcode, education).

Concept D.3.7

Meta-analysis and facts

Recognize the relationship between many trials, uncertainty, and whether a claim is a “fact.”

21st-century skills

Durable skills

Media literacy and digital citizenship

D.3.7a

Acknowledge that examining the same data with identical methods can yield different results due to varying factors, and that a "fact" is not always quickly or easily proven. e.g., data collection issues, analysis approaches, analysis errors, model assumptions

Strand E

Tools and Techniques

Visualization and Communication

This strand focuses on how to communicate about data through the creation and examination of visualizations.

Visualizations are a vital component of the sensemaking process when working with data. Being able to communicate with and about data using visualizations that are clear and tailored to a purpose and audience are an important step for creating action and impact through data. Also important are skills and habits for how to read, interpret, and critique other's data communication, paying attention to context, audience and purpose.

Substrand E1

Representations and Dynamic Visualizations

The creation and interpretation of graphic and interactive visualizations are vital components of the sensemaking process when working with data. Working with data visualizations requires an understanding of conventional components and best practices along with graphical literacy and representational fluency.

Concept E.1.1

Sense-making with visualizations

Practice creating visualizations to summarize many things at once, relationships between things in one place, or exceedingly complex ideas in one place. Recognize that visuals can be more efficient or compelling than other forms of communication.

E.1.1a

Create data visualizations that use multiple variables.

E.1.1b

Create a data visualization, collect feedback from the target audience, and revise the visualization based on feedback.

E.1.1c

Create map visualizations to display location data. e.g., events at certain spots on a map, data by state or region

Concept E.1.2

21st-century skills

Durable skills

Investigate with visualizations

Create data visualizations to directly support the analysis steps of data.

E.1.2a

Use visualizations of common data distributions to identify potential errors in the data. e.g., outliers, out-of-bounds values

E.1.2b

Visualize the distribution of data to illustrate the shape, spread, and measures of center informally.

E.1.2c

Create scatterplots for pairs of numerical variables in the data set and evaluate whether the relationships or non-relationships are as expected.

Concept E.1.3

21st-century skills

Durable skills

Clear design for user interpretation

Identify conventional components and best practices of data visualization from a user-centered or audience perspective.

E.1.3a

Clearly label a data visualization to demonstrate what the data is, what the unit of measure is, and where it came from.

E.1.3b

Choose or create a representation and color palette for one or two-variable data, and explain or defend their choice.

Concept E.1.4

Graphical literacy

Comfortably read graphs with accuracy and make sense of data visualizations by answering questions about how the data is represented with precision.

21st-century skills

Media literacy and digital citizenship

E.1.4a

Answer questions about continuous numerical scaled data, location data, and/or categorical data represented visually with multiple variables.

6-8 E.1.4b

Describe the relationships (or lack thereof) represented in scatterplots (e.g., direct vs. inverse, positive vs. negative).

E.1.4c

Review non-standard data representations that appear in popular media, identify the key visual elements and what they mean, and describe the intent and evaluate whether or not it is successful.

Concept E.1.5

Representational fluency

Identify how layout (ordering, scale, and axes) choices increase clarity or potentially mislead an audience.

21st-century skills

Media literacy and digital citizenship

E.1.5a

Compare and/or contrast various representations of data sets with multiple features and describe what is emphasized, de-emphasized, or obscured in each representation.

E.1.5b

Describe how different ways of representing data can improve clarity or mislead.

Concept E.1.6

Parallel visual-type construction

Align the type of data (numeric, categorical, string, other) to a visualization type designed for that use-case.

E.1.6a

Describe and discuss the typical visualization characteristics of numeric, categorical, and string data while identifying and outlining the differences between them.

Substrand E2

Data Storytelling

Being able to communicate with and about data using visualizations connected to a narrative is an important step for creating action and impact through data. Understanding the audience for the narrative is vital to clear communication.

Concept E.2.1

Connect narratives and data visualizations

Understand the relationship between a data visualization and its associated narrative.

E.2.1a

Evaluate the degree to which visualizations and their surrounding text and context match and support one another.

E.2.1b

Recognize that data visualizations need explanations to tell their story.

Concept E.2.2

Write data stories

21st-century skills

Durable skills

Structure effective stories about data when complex jargon and technical ideas are involved.

E.2.2a

Explain what the data reveals and whether it supports or contradicts any claims initially made.

E.2.2b

Create a visualization based on a 3-5 sentence narrative describing a particular environmental phenomenon involving multiple variables.

E.2.2c

Create a provocative question, support that question with relevant data, and reveal the story the data is telling, including connections with real-life scenarios and potential solutions.

Concept E.2.3

Adapt storytelling

21st-century skills

Durable skills

Tailor storytelling for different audiences.

E.2.3a

Present data in a way that is accessible and engaging, while considering the specific needs, interests, and knowledge level of the audience.

E.2.3b

Use visuals to enhance understanding and/or incorporate interactive discussion about the data and the narrative.

Substrand E3

Acting on Data to Benefit Society

One of the ultimate goals of working with data is applying interpretation and conclusions to real-world problems and scenarios in order to engage in civic practice and enact positive change on the world.

Concept E.3.1

Intent and authorship of analyses

21st-century skills

Media literacy and digital citizenship

Regularly interrogate the point of view of a data author, and transparently share your own.

E.3.1a

Identify the reason a data representation was created. e.g., to persuade, present factual information

E.3.1b

Identify potential biases of the source of data used to create a visualization.

E.3.1c

Communicate the limitations of data visualizations based on the source of data used in to create it.

Concept E.3.2

Advocacy with data arguments

Recognize how data can provide evidence for/persuade others toward positive change and how it can benefit society.

E.3.2a

Collect personal data and use it to benefit their family or classroom.

E.3.2b

Assess a current events news story featuring a data visualization and evaluate how effectively the graphic communicates the situation while allowing for a valid comparison.

Concept E.3.4

Impacts of technology use

Appreciate how AI and other data-driven technology may affect people and resources globally.

21st-century skills

AI literacy

E.3.4a

Recognize that data collection practices, tools, representations and resulting consequences are unevenly distributed across the globe.