Strand A

Essential Core Ideas

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

Distinguish when data is categorical versus numeric and define the difference.

A.1.1b

Recognize that non-traditional forms (e.g., photographs, written text, audio recordings) of data are informative and supportive of inquiry.

A.1.1c

Understand case structure as a way to identify the defining "case" of the data where a case is a data point which may have many variables associated with it, each with a possible value.

Concept A.1.2

Data are produced by people

21st-century skills

AI literacy

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

A.1.2a

Ask questions about how data are collected or considered.

A.1.2b

Understand that data is generated by people who make decisions about what and how to measure.

Concept A.1.3

Variability of data

Recognize that variability is a foundational component of data.

A.1.3a

Multiple conclusions can be drawn from the same set of data.

A.1.3b

Recognize that variability of data contributes to uncertainty. e.g., measuring plant growth daily shows natural variation which makes predicting exact height for the next day difficult

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

Select variables of interest for data investigations while recognizing those selections will retain inherent limits.

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

Recognize AI as a computing tool that adapts its functions by acquiring knowledge from organized data inputs and outputs. e.g., AI tools improve their tasks by comparing outputs to correct answers such as a photo-sorting app checks if its ‘cat’ labels match human-provided tags, then updates its sorting rules to reduce mistakes

A.1.5b

Recognize that many inputs and outputs can be organized into a structure that is easily readable by a machine (e.g., data-table).

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

Identify how data collection can create risks (e.g., medical information, location, privacy, exclusion) for individuals or groups, and describe ways to protect personal information.

A.2.1b

Evaluate how datasets can benefit society (e.g., solving problems, improving designs) while considering potential risks to individuals.

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

Recognize that some biases in data are neutral, while others can be harmful when making decisions and inferences, and some may not cause harm at all. e.g., neutral, preference of apples over oranges in a fruit study; harmful, surveying the coding club to generalize about all students

A.2.2b

Understand the importance of considering the context, scope, and purpose of data in order to mitigate bias.

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

Compare arguments with and without data.

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

Plan and conduct investigations to answer questions using basic data organization and visualization.

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

Revise questions and methods at each stage of investigation based on new findings.

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

Explain how inferences shift as new data emerges during an investigation.

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

Recognize that data interpretation varies across social and cultural contexts.

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

Describe the ways in which data can affect your personal life and habits.

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

Look through data to identify missing data, and add additional cases or values for variables if needed.

B.1.1b

Look through data to identify unreasonable values or recording errors in data values, and correct these if the correct values are known.

Concept B.1.2

Organizing and structure

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

B.1.2a

Collect and organize data about objects or events with multiple variables, progressing from simple case cards to structured tables with labeled rows (e.g., observations) and columns (e.g., variables).

Concept B.1.3

Processing and transformation

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

B.1.3a

Manipulate tabular data by grouping cases based on categorical variables (e.g., grouping roller coaster cases so that all wood coasters are together and all steel coasters are together) and ordering cases based on numerical variables (e.g., ordering roller coaster cases "top speed" from slowest to fastest).

Concept B.1.4

Summarizing groups

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

B.1.4a

Compare characteristics across groups using basic numerical summaries (e.g., comparing the typical recess activity across different grade levels).

B.1.4b

Create basic summaries that describe what is the same or different about groups in a dataset (e.g., summarizing how children of different ages differ in their favorite sports).

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

Design an investigation requiring collection of data involving the collection or gathering of multiple variables.

B.2.1b

Design an investigation that require collecting numerical data, including looking at a variable over a period of time.

Concept B.2.2

Data creation techniques and methods

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

21st-century skills

AI literacy

B.2.2a

Record outcomes of simple random simulations or processes.

B.2.2b

Use data generated by sensors or automated techniques. e.g., weather stations record temperature every hour

B.2.2c

Describe the procedures and tools to be used to measure a quantity of an object or an event.

Concept B.2.3

Creating data collection plans

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

21st-century skills

Durable skills

B.2.3a

Create a basic data collection plan that specifies what data to gather, what tools to use, and what sources to access for an investigation.

Concept B.2.4

Finding secondary data

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

21st-century skills

Media literacy and digital citizenship

B.2.4a

Locate and retrieve simple datasets from educational resources and child-friendly data repositories to investigate specific questions.

B.2.4b

Identify basic criteria for determining whether a dataset is relevant to a given question (e.g., topic match, timeframe, geographic relevance).

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

Understand that a variable measures the same characteristic on several individuals or objects.

B.3.1b

Recognize and apply measurement precision, including why repeated measurements may vary and how to choose appropriate precision levels.

B.3.1c

Identify the characteristics of an event or object that can be measured.

B.3.1d

Plan and conduct measurements by identifying measurable characteristics and collecting both categorical and numerical variables of objects/events.

Concept B.3.2

Working with data created by others

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

21st-century skills

Durable skills

B.3.2a

Consider the reasonable values for each of the variables and note those that are suspect.

Concept B.3.3

Ethics of data collection and usage

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

21st-century skills

Durable skills

AI literacy

B.3.3a

Recognize that personal information needs to be used respectfully and that this hasn't always been done in the past.

B.3.3b

Consider how data categories might affect different people in different ways. e.g., asking students about the language they speak at home and not including that language as an option may make students feel excluded

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

Work with datasets that require some cleaning (e.g., resolution of missing data or blank cells).

B.4.1b

Verify data by comparing recorded values to original sources when possible.

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

Use datasets that include only variables necessary to answer the stated question.

B.4.2b

Use datasets with both numerical and categorical variables.

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 4 variables and up to 50 observations.

B.4.3b

Recognize the difference between numerical and categorical data and choose the appropriate type for a particular measurement.

Concept B.4.4

Complexity of structure

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

B.4.4a

Combine information from two simple datasets about the same objects or events.

B.4.4b

Create new variables through simple calculations or combinations of existing data.

B.4.4c

Convert data between different basic formats (e.g., from tally marks to numbers).

Strand C

Tools and Techniques

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

Calculate summaries for categorical and numeric data, focusing on total and typical values.

Concept C.1.2

Measures of spread

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

C.1.2a

Calculate the range for numerical data.

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 the number of clusters, symmetric or not, and gaps. e.g., dot plot of test scores might show a cluster at 80-90% meaning most students did well and a gap at 50-60% meaning few students struggled

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

Summarize data with fractions, relative frequencies, proportions, or percentages to make comparisons.

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

Categorically describe the absence of data.

Concept C.1.6

Metadata

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.

21st-century skills

AI literacy

C.1.6a

Understand the definition and use of 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

Observe whether or not there appears to be an association between two variables. e.g., student height compared to shoe size vs. student height compared to favorite color

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

Understand that the distribution of a categorical or numerical variable represents how often a specific outcome occurs.

C.2.2b

Recognize that distributions can be used to compare two groups.

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

Create time-series graphs to determine change in variable over time.

C.2.3b

Use data collected through surveys or experiments (e.g., heights of fellow classmates) and use spreadsheets to visualize trends and relationships

C.2.3c

Use no-code or low-code data science tools. e.g., CODAP, Desmos, Google sheets

Concept C.2.4

Analyzing non-traditional data

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.

21st-century skills

AI literacy

C.2.4a

Identify word frequencies from a simple text (e.g., paragraph or story).

C.2.4b

Collect and analyze simple sensor data (e.g., temperature readings over a day).

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

Use data from surveys (e.g., favorite snacks) and then have students use this data to build a decision tree.

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

Sort, order, group, or otherwise organize objects or their representations to answer questions.

C.3.1b

Categorically describe the center, spread, and shape of a simple distribution and understand what each of these descriptions refer to.

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

Understand how data varies by exploring spread (e.g., range) and comparing qualities (e.g., brightness or temperature).

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

Identify and explain simple measurement error. e.g., different students' get varying results when measuring the same object

C.3.3b

Identify potential sources of natural variability in a given measure based on knowledge of the data context. e.g., plants can be different heights, plants grow taller over time, plants grow differently in different areas in the garden

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 that is represented in a digital tool.

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 digital tool that is suited for the intended task.

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

Identify tools that make data more accessible, such as screen-readers, captions, or tactile graphs.

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

21st-century skills

AI literacy

Understanding modeling

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

Understand that grouping objects by shared characteristics creates rules that can be used to classify and categorize new objects.

C.5.1b

Recognize that patterns and relationships in data provide different kinds of information.

C.5.1c

Discuss how data relationships help describe real-world phenomena. e.g., taller plants tend to be older

Concept C.5.2

Creating models

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.

21st-century skills

AI literacy

C.5.2a

Predict whether an object belongs to a group or category based on its characteristics.

C.5.2b

Distinguish patterns from relationships in data.

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

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.

21st-century skills

Durable skills

D.1.1a

Formulate a guess or hypothesis and identify informal vocabulary to convey your level of confidence. e.g., I strongly believe most of my classmates ride the bus to school because XX or YY

Concept D.1.2

Priors and updates

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

21st-century skills

Durable skills

Media literacy and digital citizenship

D.1.2a

Record a guess about the world, compare the initial assumption to new findings from data, and assess the extent to which the original assumption should change in light of new evidence.

Concept D.1.3

Expected value

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.

21st-century skills

Financial literacy

D.1.3a

Discuss the relationship between magnitude and probability. e.g., is a small chance of a large-sized event equivalent to a medium chance of a medium-sized event

Concept D.1.4

Explaining significance

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.

21st-century skills

Media literacy and digital citizenship

D.1.4a

Describe how "unusual" a result may be compared to an otherwise expected outcome in a given situation. e.g., flipping a coin 10 times and getting 10 heads is highly unlikely

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

Recognize that a sample of a group may or may not reflect the entire group. e.g., if the class's favorite drink for lunch is chocolate milk, does that mean the school's favorite drink for lunch is chocolate milk?

D.1.5b

Relate the effect of repeated samples to the representativeness of an entire group. e.g., pulling 10 jellybeans from a jar 5 times gives a better estimate of the color distribution than just one handful

Concept D.1.6

Correlation versus causation

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

21st-century skills

Media literacy and digital citizenship

D.1.6a

Using graphical displays, informally assess whether or not there is an association between two phenomenon in a data visualization and discuss whether one observation or trend may affect the other.

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

Recognize that randomization ensures fairness in selection processes and consider the potential consequences of non-blind selection methods. e.g., picking a raffle champion, prizes from a jar, candy of different sizes from a treat bag without looking

Concept D.1.8

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.

21st-century skills

Durable skills

D.1.8a

Describe patterns in two-variable data, such as data that show trends that increase or decrease, or relationships shown in different types of graphs. e.g., side-by-side bar charts and line graphs

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 you answer by counting or measuring results from different groups.

D.2.1b

Identify from among a set of given examples what types of questions can be answered with real-world data (e.g., values, opinions, non-observables).

Concept D.2.2

Iteration, validation, and multiple explanations

21st-century skills

Durable skills

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

D.2.2a

Estimate the total count of a characteristic within a group, providing several reasons to support the accuracy of your estimate.

D.2.2b

Evaluate whether the count of a characteristic in one group differs from that in another group, considering various reasons for this difference.

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

Identify reasons to support and refute conclusions when drawing insights from data.

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

Propose types of data and/or data comparisons that are relevant for answering a given investigation question.

D.2.4b

Identify types of data and/or data comparisons that are NOT relevant for answering a given investigation question.

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

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

21st-century skills

- Media literacy and digital citizenship
- AI literacy

D.3.1a

Recognize that a result or pattern from data does not always extend to other situations.

Concept D.3.2

Sample versus population

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

21st-century skills

- AI literacy

D.3.2a

Recognize that in some situations, a small amount of data can represent or estimate a larger unknown, saving time and effort. e.g., dice rolling, jars of jelly beans

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 in a scenario of random chance (e.g., dice rolls, jar of jelly beans), too few trials can skew conclusions. e.g., flipping a coin twice and getting heads both times doesn't mean it's always heads and more flips will provide a clearer picture

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 errors can arise in analysis due to both human and technological factors, especially when the analysis is duplicated. e.g., different sensors, multiple data collections, multiple people

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 to summarize categorical data.

E.1.1b

Display groups or categories in visualizations using complementary or contrasting colors to highlight differences.

E.1.1c

Display continuously scaled data in visualization using shading.

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

Recognize how frequency distributions can help identify outliers and errors in the data. e.g., data contains values that shouldn't be possible

E.1.2b

Organize and present collected data visually to highlight relationships and to support a claim.

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

Identify and support how different colors and/or patterns can be used in visualizations to represent different groups/categories/scales in the data.

E.1.3b

Reliably use the parts (e.g., titles, labels, legends, colors) of bar graphs, picture graphs, and line graphs.

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 fractional valued numerical data or categorical data represented visually with one or two variables.

E.1.4b

Recognize unusual data points and consider reasons why they might appear.

E.1.4c

Work with a variety of data types, including numerical data, charts, graphs, and visual representations to draw conclusions and understand the story the data is telling.

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 visualizations of the same data by altering different features (e.g., reordering bars, changing colors), and explain how these changes affect what is highlighted or obscured in each representation. e.g., bar graph sorted by size highlights the most popular option, while sorting alphabetically can make comparison challenging

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

Visualize multiple types of data (e.g., numeric, categorical, string data) during in-class data collection exercises.

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 effectiveness of text, visualization, and text plus a visualization to communicate a particular story.

E.2.1b

Make a prediction based on a visualization using the terms: “likely, unlikely, certain, and impossible.”

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

Describe the data clearly by identifying any trends or patterns found using descriptive language and terms such as “most,” “least,” “greater than,” “less than,” and “equal to."

E.2.2b

When describing the data, decide whether any claim made about the data makes sense.

E.2.2c

Find ways to generate interest in the story by crafting a hook that captivates the audience, then supporting it with data examples that reveal the narrative the data conveys.

Concept E.2.3

Adapt storytelling

21st-century skills

Durable skills

Tailor storytelling for different audiences.

E.2.3a

Understand various audiences and adapt storytelling to suit their needs and comprehension levels. e.g., using straightforward language with peers vs. more analytical explanations with teachers

E.2.3b

Provide the appropriate level of context for various audiences.

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

Assess the purpose and effectiveness of a data story by identifying why it is being told, its goal, and whether it achieves that goal.

E.3.1b

Identify situations when data can be used to make decisions at school or at home.

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

Develop creative data visualizations to depict an aspect of the student's community or social connections.

E.3.2b

Draw simple conclusions about the data from a narrative.

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

Use familiar examples of energy consumption (e.g., tablets, laptops, cell phones) to draw conclusions about the energy use of large data centers and systems like AI. e.g., one laptop charging uses 50 watts and an AI data center uses as much energy as 50 million laptops running together