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.

K–2	3–5	6–8	9–10	11–12
K-2.A.1.1a Utilize both categorical and numeric data.	3-5.A.1.1a Distinguish when data is categorical versus numeric and define the difference.	6-8.A.1.1a Analyze the way categorical and numeric data shapes its interpretation and analysis.	9-10.A.1.1a Define "qualitative" and "quantitative" and understand how they relate to categorical and numeric data.	11-12.A.1.1a Recognize that multiple types of data can provide valuable insights into the same inquiry.
K-2.A.1.1b Recognize that data can be derived from many different forms of sources (e.g., photographs, written text, audio recordings, videos, people, and other non-traditional places).	3-5.A.1.1b Recognize that non-traditional forms (e.g., photographs, written text, audio recordings) of data are informative and supportive of inquiry.	6-8.A.1.1b Recognize that numerical variables may be either discrete or continuous.	9-10.A.1.1b Understand that forms of media (e.g., photographs, written text, audio recordings) can be represented in quantitative and qualitative terms.	
K-2.A.1.1c Understand that data can be used to ask and answer questions.	3-5.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.

K–2	3–5	6–8	9–10	11–12
K-2.A.1.2a Recognize the importance of asking questions about how data were collected.	3-5.A.1.2a Ask questions about how data are collected or considered.	6-8.A.1.2a Ask questions regarding the origins of specific automated measures (e.g., webtracking, email meta-data, user accounts).	9-10.A.1.2a Explain how data-based decisions are revisited as new evidence or societal needs emerge (e.g., blood pressure cut-off numbers, dietary guidance, medical benchmarks).	11-12.A.1.2a Explore the origins of some standardized unit measurements (e.g., horsepower, mole, scores on AP exams).
	3-5.A.1.2b Understand that data is generated by people who make decisions about what and how to measure.	6-8.A.1.2b Recognize the limits of the information the data can provide and the story it can tell.	9-10.A.1.2b Evaluate why data models require updates to maintain accuracy and relevance.	11-12.A.1.2b Identify the risks and tradeoffs of using traditional measurements (e.g., IQ, BMI).
		6-8.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.

K–2	3–5	6–8	9–10	11–12
K-2.A.1.3a Observe that data can have many different answers or results.	3-5.A.1.3a Multiple conclusions can be drawn from the same set of data. 3-5.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	6-8.A.1.3a Make sense of the variability of data through an iterative process of refinement by questioning.	9-10.A.1.3a Recognize the different types of variability (e.g., natural, measurement, sampling).	11-12.A.1.3a Explore different types of variability for making inferences (e.g., confidence intervals, various tests, classification models).

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.

K–2	3–5	6–8	9–10	11–12
K-2.A.1.4a Understand that data can show some things but not others. e.g., categorizing the colors worn in a kindergarten class does not indicate the clothing item or size	3-5.A.1.4a Select variables of interest for data investigations while recognizing those selections will retain inherent limits.	6-8.A.1.4a Specify ways that data provide incomplete information relative to the object being studied. 6-8.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.	9-10.A.1.4a Evaluate claims derived from data by questioning how phenomena are measured, categorized, or represented.	11-12.A.1.4a Design and compare alternative data representations, justifying choices to address inherent uncertainty.

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.

K–2	3–5	6–8	9–10	11–12
K-2.A.1.5a Recognize that computing tools (e.g., computers, smartphones, IoT buttons, sensors) and AI need data from human inputs (e.g., a function machine: if x input, then y action) to perform actions. e.g., smart thermostat turns on heat when the temperature sensor detects the room temperature is colder than the temperature a human programmed, such as 68°F	3-5.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	6-8.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	9-10.A.1.5a Describe the basic mathematical features of an AI model in terms of independent variables (e.g., inputs), dependent variables (e.g., outputs), and predictors or weights (e.g., slopes of many variables). e.g., AI models use math to weigh inputs, such as a music recommendation model might calculate: (play_count × $weight_1$) + (listen_duration × $weight_2$) + (skip_count × $weight_3$) = recommendation_score, and weights are adjusted automatically to minimize mismatches between predicted and actual user preferences.	11-12.A.1.5a Identify and label a simple prediction algorithm or equation for a very basic AI prediction model. e.g., a basic AI model's equation looks like: Prediction = ($weight_1 \times input_1$) + ($weight_2 \times input_2$), such as a college admission model might weight GPA ($input_1$) and test scores ($input_2$) to output an acceptance likelihood, and training involves automatically adjusting weights to match historical data.
K-2.A.1.5b Understand that AI tools use data from people to do tasks. e.g., chatbots learn from typed questions	3-5.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).	6-8.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	9-10.A.1.5b Describe and explore how it is possible for data in a variety of formats (e.g., images) to be translated into organized, numerical information for an AI model to process.	11-12.A.1.5b Understand that algorithms use cost functions to measure errors and adjust predictions.
			9-10.A.1.5c Identify how biases in training data can lead to biases in AI models by directly affecting predictors or weights. e.g., If AI only sees pictures of cats in sunlight, it would fail to recognize cats in shadows	11-12.A.1.5c Recognize that some AI tools can be used to explore complex data with many variables.
				11-12.A.1.5d Recognize the types of problems that are ideal for using an AI tool to analyze complex data.

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

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

21st-century skills

Durable skills

K–2	3–5	6–8	9–10	11–12
K-2.A.2.1a	3-5.A.2.1a	6-8.A.2.1a	9-10.A.2.1a	11-12.A.2.1a
Recognize how data can be useful in understanding the world around us. e.g., counting rainy days to plan outdoor activities	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.	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	Analyze how data use can perpetuate biases or systemic inequities (e.g., predictive policing, hiring algorithms).	Recognize that data risk can change based on time, circumstance, and purpose.
K-2.A.2.1b	3-5.A.2.1b	6-8.A.2.1b	9-10.A.2.1b	11-12.A.2.1b
Understand that some data about people should not be collected or shared with technology.	Evaluate how datasets can benefit society (e.g., solving problems, improving designs) while considering potential risks to individuals.	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	Evaluate context-specific risks and benefits of data interpretations (e.g., health tracking for improving care vs. privacy concerns).	Identify data benefits that can appear well into the future and in unexpected ways.

Concept A.2.2

Biases in data

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

21st-century skills

Durable skills

K–2	3–5	6–8	9–10	11–12
K-2.A.2.2a	3-5.A.2.2a	6-8.A.2.2a	9-10.A.2.2a	11-12.A.2.2a
Recognize how data are affected by decisions made around data design, collection, and interpretation.	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	Identify how biases in data affect inferences and questions.	Recognize how biases can obscure inferences drawn from data.	Propose multiple perspectives on data to mitigate inherent biases.
	3-5.A.2.2b		9-10.A.2.2b	11-12.A.2.2b
	Understand the importance of considering the context, scope, and purpose of data in order to mitigate bias.		Consider how the consolidation or combination of different data can create additional biases.	Understand the difference between implicit and explicit bias.

Concept A.2.3

Power of data

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

21st-century skills

Media literacy and digital citizenship

K–2	3–5	6–8	9–10	11–12
K-2.A.2.3a Use data to answer questions and see how it improves guesses.	3-5.A.2.3a Compare arguments with and without data.	6-8.A.2.3a Analyze how data is used to solve problems, persuade, and discover new ideas.	9-10.A.2.3a Evaluate how data drives innovation in fields and informs community choices.	11-12.A.2.3a Use data to support arguments, design solutions, or challenge inequities. 11-12.A.2.3b Investigate case studies where data advanced scientific, economic, or social progress. 11-12.A.2.3c Identify when data alone is insufficient and complementary methods are needed. e.g., Data may quantify the number of people affected by a policy, but personal testimonies are needed to illustrate its human impact

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

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.

21st-century skills

Durable skills

K–2	3–5	6–8	9–10	11–12
K-2.A.3.1a Recognize there is an investigative process for exploring questions about the world.	3-5.A.3.1a Plan and conduct investigations to answer questions using basic data organization and visualization.	6-8.A.3.1a Investigate real-world questions by cleaning, analyzing, and interpreting data to draw conclusions.	9-10.A.3.1a Design and refine investigations to address contextual problems (e.g., social, educational, business, medical, governmental issues), evaluating limitations and biases.	11-12.A.3.1a Conduct independent investigations to inform decisions, leveraging advanced tools and addressing uncertainty. 11-12.A.3.1b Compare investigative approaches across fields to critique strengths and limitations.

Concept A.3.2

Iteration

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

21st-century skills

Durable skills

K–2	3–5	6–8	9–10	11–12
K-2.A.3.2a Utilize different views such as pictures, tallies, or charts to help answer questions and notice patterns.	3-5.A.3.2a Revise questions and methods at each stage of investigation based on new findings.	6-8.A.3.2a Recognize that the investigative process is non-linear, often cycling between phases in various orders multiple times.	9-10.A.3.2a Employ iteration in an investigation to strengthen interpretations or inspire new investigations.	11-12.A.3.2a Propose new approaches for leveraging the investigative process to strengthen inferences and arguments.

Concept A.3.3

Dynamic inferences

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

21st-century skills

Durable skills

K–2	3–5	6–8	9–10	11–12
K-2.A.3.3a	3-5.A.3.3a	6-8.A.3.3a	9-10.A.3.3a	11-12.A.3.3a
Utilize a single set of data to generate multiple inferences for various inquiries.	Explain how inferences shift as new data emerges during an investigation.	Revise initial conclusions when new data emerges and use evidence to support claims.	Use digital tools to test and refine inferences from large or complex datasets.	Critically evaluate and update inferences as data scales or methods advance.

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.

K–2	3–5	6–8	9–10	11–12
K-2.A.3.4a	3-5.A.3.4a	6-8.A.3.4a	9-10.A.3.4a	11-12.A.3.4a
Share data and interpretations with others.	Recognize that data interpretation varies across social and cultural contexts.	Explain data interpretations from various disciplinary and community perspectives (e.g., social studies, families).	Reinterpret data from multiple perspectives, disciplines, and historical frames of reference.	Interpret data drawn from different fields and topics based on accepted norms within those fields.
			9-10.A.3.4b	11-12.A.3.4b
			Compare and contrast problem solving approaches and the resulting findings.	Compare multiple problem-solving approaches, and identify how those differences may compound over time and when repeated.

Concept A.3.5

Student data agency

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

21st-century skills

Durable skills

K–2	3–5	6–8	9–10	11–12
K-2.A.3.5a	3-5.A.3.5a	6-8.A.3.5a	9-10.A.3.5a	11-12.A.3.5a
Develop curiosity about data and how it can be used in the world.	Describe the ways in which data can affect your personal life and habits.	Embed data practices into everyday life and advocate for the benefits of doing so.	Utilize data science tools and methods to engage in personal and collective inquiry relevant to one's own life and interests.	Establish accountability by basing claims and decisions on relevant data.
K-2.A.3.5b		6-8.A.3.5b		11-12.A.3.5b
Exhibit the capacity to work with open-ended problems.		Assess the accuracy, perspective, credibility, and relevance of various resources (e.g., information, media, data).		Explore career fields and their intersection with data collection, curation, storytelling, and societal impact.

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.

K–2	3–5	6–8	9–10	11–12
K-2.B.1.1a	3-5.B.1.1a	6-8.B.1.1a	9-10.B.1.1a	11-12.B.1.1a
Recognize and explain any missing data (e.g., a student was absent when data was collected) or data recording errors (e.g., "10" recorded as a "1").	Look through data to identify missing data, and add additional cases or values for variables if needed.	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.	Use data dictionaries to identify codes for missing or incomplete data (e.g., NA, 99999, 0, " "), and either recode or filter data to remove those observations.	Develop comprehensive data validation procedures, including automated checks.
K-2.B.1.1b	3-5.B.1.1b		9-10.B.1.1b	11-12.B.1.1b
Record responses so that you can tell if everyone has been asked.	Look through data to identify unreasonable values or recording errors in data values, and correct these if the correct values are known.		Apply basic cross-validation techniques to verify data quality across multiple sources, including source comparison, split sampling, internal consistency checks, and domain range validation.	Implement verification protocols for complex datasets with multiple dependencies.

Concept B.1.2

Organizing and structure

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

K–2	3–5	6–8	9–10	11–12
K-2.B.1.2a	3-5.B.1.2a	6-8.B.1.2a	9-10.B.1.2a	11-12.B.1.2a
Collect and record data on case cards, wherein each card represents a single observation.	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).	Use categorical variables or bins/groups of numerical variables in a dataset to restructure data into groups.	Create and manage complex data structures with multiple related tables, understanding primary and foreign key relationships between datasets.	Develop and implement data organization systems that accommodate both structured and unstructured data types.
K-2.B.1.2b		6-8.B.1.2b	9-10.B.1.2b	11-12.B.1.2b
Create categories from individual categorical responses (e.g., scary things).		Make sense of and use a dataset arranged in nested or hierarchical format.	Transform and restructure hierarchical or nested data into normalized tabular formats suitable for analysis.	Create scalable data organization strategies that maintain data integrity while handling missing values, irregular structures, and evolving data requirements.
K-2.B.1.2c			9-10.B.1.2c	11-12.B.1.2c
Define the categories used to measure the qualities of an object (e.g., color, shape).			Design efficient organizational schemas for large datasets with multiple variables and complex relationships.	Design and implement metadata documentation systems to track data lineage, transformations, and organizational structures.

Concept B.1.3

Processing and transformation

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

K–2	3–5	6–8	9–10	11–12
K-2.B.1.3a Sort case cards so that observations with similar values for a variable are grouped together.	3-5.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).	6-8.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...).	9-10.B.1.3a Use calculations and logic statements to create new categorical variables based on existing categorical (e.g., if (employment=”employed”, Yes, No)) or quantitative variables (e.g., if (weight<30, light, if(weight>60,heavy ,medium)).	11-12.B.1.3a Use an identifying variable (e.g., index, case ID) to merge two separate datasets that have the same observation, but contain different variables to merge datasets together.
K-2.B.1.3b Order case cards so that a numerical variable is ordered from smallest to largest or largest to smallest.		6-8.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).	9-10.B.1.3b Filter data based on groups or subsets of data relevant to the problem and context.	11-12.B.1.3b Use appropriate procedures to join two datasets together that have different observations with the same variables measured.

Concept B.1.4

Summarizing groups

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

K–2	3–5	6–8	9–10	11–12
K-2.B.1.4a Count the number of items in different groups when data is organized into simple categories (e.g., counting how many students chose each favorite color).	3-5.B.1.4a Compare characteristics across groups using basic numerical summaries (e.g., comparing the typical recess activity across different grade levels).	6-8.B.1.4a Use summary measures of groups within a nested or hierarchical dataset.	9-10.B.1.4a Create summary measures for groups that can then be used as a measure at the group level (e.g., for salary data, compute average salary for different occupational groups).	11-12.B.1.4a Use datasets with derived variables, based on other variables in the dataset.
	3-5.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.

K–2	3–5	6–8	9–10	11–12
K-2.B.2.1a Formulate simple questions that guide data collection and analysis about familiar contexts, using appropriate support.	3-5.B.2.1a Design an investigation requiring collection of data involving the collection or gathering of multiple variables.	6-8.B.2.1a Construct data-based questions that explore relationships between variables and consider how data collection methods affect the quality of evidence.	9-10.B.2.1a Construct data-based questions about the design of a study to determine causality and make predictions.	11-12.B.2.1a Construct data-based questions that address complex systems with multiple interacting variables, including consideration of confounding factors and effect modifiers.
	3-5.B.2.1b Design an investigation that require collecting numerical data, including looking at a variable over a period of time.		9-10.B.2.1b Identify comparison and association data-based questions appropriate for addressing problems of interest.	11-12.B.2.1b Design research questions that incorporate multiple levels of analysis and account for both direct and indirect relationships between variables.
				11-12.B.2.1c Formulate questions that address the validity and reliability of data collection methods, including considerations of systematic bias and measurement error.

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

K–2	3–5	6–8	9–10	11–12
K-2.B.2.2a Recognize that simulations and models can act as sources of data.	3-5.B.2.2a Record outcomes of simple random simulations or processes.	6-8.B.2.2a Use data generated from simulations and models to investigate a question of interest.	9-10.B.2.2a Describe benefits and drawbacks of using proxy variables.	11-12.B.2.2a Design simulations (e.g., using an RNG or computer software) and underlying models to generate data specific to a problem of interest.
	3-5.B.2.2b Use data generated by sensors or automated techniques. e.g., weather stations record temperature every hour	6-8.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.	9-10.B.2.2b Use and/or change parameters of simulations to generate data to address a problem of interest.	11-12.B.2.2b Identify optimal sensors or automated data collection methods for answering a data-based question or designing an experiment.
	3-5.B.2.2c Describe the procedures and tools to be used to measure a quantity of an object or an event.		9-10.B.2.2c Design sensor-based experiments or automated data collection scenarios to explore a problem or question and identify the scenarios' limitations and trade-offs.	11-12.B.2.2c Distinguish between surveys, observational studies, and experiments.
			9-10.B.2.2d Describe the features, benefits, limitations, and ethical thinking that went into a data collection process.	
			9-10.B.2.2e Design and implement traditional data collection methods (e.g., surveys, observations, field studies) to investigate research questions and evaluate their strengths and limitations compared to automated approaches.	

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

K–2	3–5	6–8	9–10	11–12
K-2.B.2.3a Identify what information is needed and where it might be found to answer simple questions.	3-5.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.	6-8.B.2.3a Design data collection plans that include data types needed, collection methods, sample sizes, and timing considerations for investigations.	9-10.B.2.3a Develop comprehensive data collection plans that address potential limitations, specify quality control measures, and include contingency strategies.	11-12.B.2.3a Create sophisticated data collection plans that incorporate ethical considerations, cost-benefit analysis, and protocols for ensuring data integrity and reproducibility.

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

K–2	3–5	6–8	9–10	11–12
K-2.B.2.4a Recognize that data can be found in various sources such as books, websites, and classroom resources to help answer questions.	3-5.B.2.4a Locate and retrieve simple datasets from educational resources and child-friendly data repositories to investigate specific questions.	6-8.B.2.4a Search for and retrieve appropriate datasets from educational repositories and curated sources designed for middle school investigations.	9-10.B.2.4a Locate and retrieve relevant datasets from publicly available scientific, civic, or government databases using search tools and filters.	11-12.B.2.4a Identify (and know you can request access to) non-publicly available datasets by contacting researchers, reading scientific literature, or communicating with public officials.
K-2.B.2.4b Explore simple, age-appropriate data sources provided by teachers or educational websites that show information about familiar topics.	3-5.B.2.4b Identify basic criteria for determining whether a dataset is relevant to a given question (e.g., topic match, timeframe, geographic relevance).	6-8.B.2.4b Evaluate potential datasets based on relevance, timeliness, and credibility of the source for answering specific questions.	9-10.B.2.4b Evaluate datasets from multiple sources to determine which best addresses a research question, considering factors such as data quality, sample size, and collection methods.	11-12.B.2.4b Develop strategies for finding and accessing datasets that require special permissions, logins, or formal data requests.
		6-8.B.2.4c Use metadata and documentation to understand the context and limitations of secondary datasets.	9-10.B.2.4c Use data catalogs, repositories, and open data portals to find datasets that meet specific criteria for investigations.	11-12.B.2.4c Evaluate and navigate licensing and citation requirements when using secondary data sources for research.
				11-12.B.2.4d Combine multiple secondary datasets to create more comprehensive or useful data for specific investigations.

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.

K–2	3–5	6–8	9–10	11–12
K-2.B.3.1a Anticipate variability in measurement.	3-5.B.3.1a Understand that a variable measures the same characteristic on several individuals or objects.	6-8.B.3.1a Create an ordinal scale of measurement.	9-10.B.3.1a Create a data dictionary to document the data collection process.	11-12.B.3.1a Evaluate and critique measurement validity, reliability, and bias in data collection methods, and design comprehensive datafication strategies that address ethical considerations and potential sources of measurement error.
K-2.B.3.1b Use either standard (e.g., inches, feet, miles) or nonstandard (e.g., paperclips, shoes) units to determine a physical quantity (e.g., width of a table) and understand the importance of standard units for consistency.	3-5.B.3.1b Recognize and apply measurement precision, including why repeated measurements may vary and how to choose appropriate precision levels.	6-8.B.3.1b Understand that data is information collected and recorded with a purpose.		
K-2.B.3.1c Begin to coordinate multiple variables of the same observation (e.g., measure more than one variable of an object or event).	3-5.B.3.1c Identify the characteristics of an event or object that can be measured.	6-8.B.3.1c Distinguish between human-derived data from images, sounds, and text vs. computer-derived data from images, sounds, and text.		
	3-5.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

21st-century skills

Durable skills

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

K–2	3–5	6–8	9–10	11–12
K-2.B.3.2a Understand how other people measured their data.	3-5.B.3.2a Consider the reasonable values for each of the variables and note those that are suspect.	6-8.B.3.2a Consider how the data were measured, with what tool and precision.	9-10.B.3.2a Make use of metadata and data dictionary to understand a data set.	11-12.B.3.2a Work with data collected over time and consider how to aggregate appropriately.
		6-8.B.3.2b Consider who collected these data and for what purpose.	9-10.B.3.2b Consider who or what was included in the data collection and who or what was not.	11-12.B.3.2b Work with data collected over space and consider how to aggregate appropriately.
		6-8.B.3.2c Consider when and where the data were collected.		11-12.B.3.2c Create strategies for dealing with data that is constantly updated.

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

K–2	3–5	6–8	9–10	11–12
K-2.B.3.3a	3-5.B.3.3a	6-8.B.3.3a	9-10.B.3.3a	11-12.B.3.3a
Understand that collecting data about people requires their permission. e.g., asking before writing down a classmates favorite color	Recognize that personal information needs to be used respectfully and that this hasn't always been done in the past.	Design data collection methods that address privacy, consent, and fair representation of different groups.	Evaluate and address ethical implications of data collection choices, including privacy, bias, and representation.	Develop data collection protocols that prevent bias, protect privacy, and ensure ethical representation across diverse populations.
K-2.B.3.3b	3-5.B.3.3b	6-8.B.3.3b	9-10.B.3.3b	11-12.B.3.3b
Ask permission before sharing others' information.	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	Examine historical examples of harmful data practices to inform ethical data use.	Analyze existing datasets for potential bias, discrimination, or unfair representation.	Apply validation techniques to prevent bias and ensure ethical use of secondary data, including AI tools.

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.

K–2	3–5	6–8	9–10	11–12
K-2.B.4.1a	3-5.B.4.1a	6-8.B.4.1a	9-10.B.4.1a	11-12.B.4.1a
Work with datasets that are relatively clean (e.g., don't have missing data or errors).	Work with datasets that require some cleaning (e.g., resolution of missing data or blank cells).	Identify and handle missing values marked by special codes (-99) or blank cells.	Work with datasets requiring multiple types of cleaning such as missing values, errors, and anomalies.	Apply advanced data cleaning techniques to handle complex data quality issues such as outliers, inconsistencies, and systematic errors.
	3-5.B.4.1b	6-8.B.4.1b	9-10.B.4.1b	11-12.B.4.1b
	Verify data by comparing recorded values to original sources when possible.	Distinguish between true zero values and blank cells.	Clean and prepare datasets before merging multiple sources.	Develop and document reproducible data cleaning workflows that maintain data integrity.
				11-12.B.4.1c
				Evaluate and validate cleaned datasets using statistical methods and domain knowledge.

Concept B.4.2

Complexity of variables

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

K–2	3–5	6–8	9–10	11–12
K-2.B.4.2a Use datasets that include only numerical or only categorical variables.	3-5.B.4.2a Use datasets that include only variables necessary to answer the stated question.	6-8.B.4.2a Work with datasets that include rates and derived variables that combine multiple measurements.	9-10.B.4.2a Work with datasets that include time-series data at different intervals to detect various patterns.	11-12.B.4.2a Create and use expected value models to support data-based decision making.
	3-5.B.4.2b Use datasets with both numerical and categorical variables.	6-8.B.4.2b Work with datasets that have multiple variables that can suggest or answer different questions.	9-10.B.4.2b Understand and work with different observation structures beyond individual units.	11-12.B.4.2b Work with multiple datasets that combine multiple types of data and combine and transform the different types.
		6-8.B.4.2c Work with datasets that show natural variation and understand why values differ.	9-10.B.4.2c Work with merged datasets that align different time scales and observation structures.	11-12.B.4.2c Work with complex derived variables and understand their calculation methods.

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.

K–2	3–5	6–8	9–10	11–12
K-2.B.4.3a Work with datasets with 1 - 2 variables and 10 - 30 observations (e.g., size of a class).	3-5.B.4.3a Work with datasets with up to 4 variables and up to 50 observations.	6-8.B.4.3a Work with datasets with up to 20 variables and over 100 observations.	9-10.B.4.3a Work with datasets with over 20 variables and over 1000 observations.	11-12.B.4.3a Work with very large datasets multiple thousands of observations.
	3-5.B.4.3b Recognize the difference between numerical and categorical data and choose the appropriate type for a particular measurement.	6-8.B.4.3b Understand how categorical variables can be used to create meaningful subsets.	9-10.B.4.3b Transform data between wide and long formats based on analysis needs.	11-12.B.4.3b Use selection, sampling, and transformation tools to navigate very large datasets.

Concept B.4.4

Complexity of structure

21st-century skills

AI literacy

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

K–2	3–5	6–8	9–10	11–12
K-2.B.4.4a Work with datasets already formatted into structure necessary for analysis.	3-5.B.4.4a Combine information from two simple datasets about the same objects or events.	6-8.B.4.4a Work with datasets where the row isn't a single observation but something more complex (e.g., average, nested cases).	9-10.B.4.4a Merge multiple datasets while maintaining appropriate observation structure.	11-12.B.4.4a Design and implement data structures that can accommodate longitudinal data and multiple levels of aggregation.
K-2.B.4.4b Create new categories from existing data through basic grouping rules.	3-5.B.4.4b Create new variables through simple calculations or combinations of existing data.	6-8.B.4.4b Work with datasets that include derived or transformed variables, including creating categorical variables from numerical data.	9-10.B.4.4b Transform complex variables into more interpretable forms using student-relatable benchmarks.	11-12.B.4.4b Handle data aggregation across different observation structures and time scales.
	3-5.B.4.4c Convert data between different basic formats (e.g., from tally marks to numbers).	6-8.B.4.4c Understand how categorical variables can be used to create meaningful subsets.		11-12.B.4.4c Create flexible organizational systems that can handle both structured and unstructured data sources.
				11-12.B.4.4d Develop documentation systems for complex data structures that track relationships and dependencies between variables.

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.

K–2	3–5	6–8	9–10	11–12
K-2.C.1.1a	3-5.C.1.1a	6-8.C.1.1a	9-10.C.1.1a	11-12.C.1.1a
Recognize that categorical data does not have a measure of center.	Calculate summaries for categorical and numeric data, focusing on total and typical values.	Identify measures of center as statistical values that represent the central tendency of data sets.	Identify appropriate ways to summarize numerical or categorical data using frequency tables, graphical displays, and numerical summary statistics.	Explore the sensitivity of the mean to outliers compared to the median.
K-2.C.1.1b		6-8.C.1.1b		11-12.C.1.1b
Describe the center of numeric data categorically using phrases like “most popular”.		Explain what measures of center are useful for and their limitations.		Discuss instances when to use the mean or median based on the context and data distribution (e.g., skewed vs. symmetric distribution).

Concept C.1.2

Measures of spread

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

K–2	3–5	6–8	9–10	11–12
K-2.C.1.2a	3-5.C.1.2a	6-8.C.1.2a	9-10.C.1.2a	11-12.C.1.2a
Describe the upper and lower bounds of a set of objects. e.g., tallest and shortest, biggest and smallest	Calculate the range for numerical data.	Categorically identify the presence of potential outliers in a dataset.	Calculate standard deviation from mean or interquartile range.	Numerically operationalize the meaning of an "outlier" using standard deviation as a measure of variability and a modified boxplot.
			9-10.C.1.2b	
			Use standard deviation as a measure of variability and a modified boxplot for identifying outliers.	

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.

K–2	3–5	6–8	9–10	11–12
K-2.C.1.3a	3-5.C.1.3a	6-8.C.1.3a	9-10.C.1.3a	11-12.C.1.3a
Describe the shape of the data categorically. e.g., "all grouped together", "spread out", "lots of small groups"	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	Describe whether data is symmetric or asymmetric and the number of modes.	Acknowledge that in a tie for the mode the distribution is bi-modal.	Explain how the shape of a distribution influences the relationship between measures of center. e.g., in symmetric distributions - the mean and median are close, in a right-skewed distribution - the mean is greater than the median, in a left-skewed distribution - the mean is less than the median
			9-10.C.1.3b	
			Understand how the data is distributed across the range of data. e.g., if the data is skewed to one side of the range	

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.

K–2	3–5	6–8	9–10	11–12
K-2.C.1.4a	3-5.C.1.4a	6-8.C.1.4a	9-10.C.1.4a	11-12.C.1.4a
Sort objects into a frequency table based on shared characteristics.	Summarize data with fractions, relative frequencies, proportions, or percentages to make comparisons.	Generate a frequency table to summarize raw categorical data.	Generate a relative frequency table to make comparisons and to generalize results.	Discuss implications of choices made when generating a frequency table.

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.

K–2	3–5	6–8	9–10	11–12
K-2.C.1.5a	3-5.C.1.5a	6-8.C.1.5a	9-10.C.1.5a	11-12.C.1.5a
Identify the absence of data.	Categorically describe the absence of data.	Numerically measure missing data.	Adjust analyses in light of missing values.	Describe how missing data affects analysis and resulting relationships, patterns, or models.
		6-8.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.

K–2	3–5	6–8	9–10	11–12
K-2.C.1.6a	3-5.C.1.6a	6-8.C.1.6a	9-10.C.1.6a	11-12.C.1.6a
Discuss the context of data. e.g., where or when it was collected	Understand the definition and use of metadata (e.g., data and time, text, continuous, geolocation).	Comprehend, in an informal sense, the value of information contained in metadata (e.g., data and time, text, continuous, geolocation).	Apply understanding of metadata (e.g., data and time, text, continuous, geolocation) to summarize and analyze data numerically, in tables and through visualizations.	Reasonably ideate on some potential modeling approaches when given the metadata (e.g., data and time, text, continuous, geolocation) for a dataset.

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.

K–2	3–5	6–8	9–10	11–12
K-2.C.2.1a	3-5.C.2.1a	6-8.C.2.1a	9-10.C.2.1a	11-12.C.2.1a
Describe similarities or differences across two variables.	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	Use reasoning about distributions to compare two groups based on quantitative variables.	Use numerical measures such as average, standard deviation and quartiles to compare two groups.	Use simulations to investigate associations between two categorical variables and to compare groups.

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.

K–2	3–5	6–8	9–10	11–12
K-2.C.2.2a	3-5.C.2.2a	6-8.C.2.2a	9-10.C.2.2a	11-12.C.2.2a
Work with visual aids (e.g., colorful charts) and hands-on activities to sort objects (e.g., color).	Understand that the distribution of a categorical or numerical variable represents how often a specific outcome occurs.	Represent the variability of numerical variables using appropriate displays (e.g., dotplots, boxplots).	Quantify variability in distributions using numerical measures.	Use variability in distributions to engage in statistical reasoning.
K-2.C.2.2b	3-5.C.2.2b		9-10.C.2.2b	11-12.C.2.2b
Use primary data (e.g., favorite fruit) and represent data with tally marks or pictographs.	Recognize that distributions can be used to compare two groups.		Recognize the relationship between variability and the shape of a distribution.	Understand and interpret variability in sampling distributions and how it impacts population estimates.

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.

K–2	3–5	6–8	9–10	11–12
K-2.C.2.3a Organize objects by size, color, shape, etc.	3-5.C.2.3a Create time-series graphs to determine change in variable over time.	6-8.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.	9-10.C.2.3a Describe associations between two categorical variables using measures such as difference in proportions and relative risk.	11-12.C.2.3a Conduct linear regression analysis to find the best-fit.
K-2.C.2.3b Use language like "goes with" "belongs to", or "matches" to group items together.	3-5.C.2.3b Use data collected through surveys or experiments (e.g., heights of fellow classmates) and use spreadsheets to visualize trends and relationships 3-5.C.2.3c Use no-code or low-code data science tools. e.g., CODAP, Desmos, Google sheets	6-8.C.2.3b Create scatterplots and add line of best fit.	9-10.C.2.3b Analyze data to uncover correlations, trends, and groupings such as clustering that inform decision-making processes across diverse fields.	11-12.C.2.3b Construct prediction intervals and confidence intervals to determine plausible values of a predicted observation or a population characteristic.

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.

K–2	3–5	6–8	9–10	11–12
K-2.C.2.4a Analyze sensory data by counting occurrences of sounds (e.g., claps or animal noises).	3-5.C.2.4a Identify word frequencies from a simple text (e.g., paragraph or story).	6-8.C.2.4a Compare word frequencies across multiple texts to identify patterns and create simple visualizations from that text data.	9-10.C.2.4a Analyze data from sensors and IoT devices to track trends and monitor changes over time. e.g., smart thermostats and lighting systems for energy monitoring, wearable fitness trackers for health and activity data	11-12.C.2.4a Generate a word cloud of a given text after standardizing (e.g., all lower case), stemming, and removing stop words.
K-2.C.2.4b Categorize sensory data by type (e.g., loud vs. soft).	3-5.C.2.4b Collect and analyze simple sensor data (e.g., temperature readings over a day).	6-8.C.2.4b Explore patterns in audio data (e.g., analyzing sound waves for volume and frequency).	9-10.C.2.4b Understand that geographic data can be visualized using maps, and it can be represented as points (e.g., latitude and longitude) and areas (e.g., GeoJSON).	
K-2.C.2.4c Sort and compare objects based on textures (e.g., smooth, rough, or bumpy).				

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.

K–2	3–5	6–8	9–10	11–12
	3-5.C.2.5a Use data from surveys (e.g., favorite snacks) and then have students use this data to build a decision tree.	6-8.C.2.5a Learn to use simple diagrams (e.g., decision trees using small relatable examples) to make important decisions for everyday choices.	9-10.C.2.5a Explore machine learning basics (e.g., classification and clustering) to make predictions with data.	11-12.C.2.5a Explore how gradient descent optimizes loss functions and powers machine learning applications like neural networks.

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.

K–2	3–5	6–8	9–10	11–12
K-2.C.3.1a Describe how similar objects can differ based on characteristics such as color, shape, and size.	3-5.C.3.1a Sort, order, group, or otherwise organize objects or their representations to answer questions. 3-5.C.3.1b Categorically describe the center, spread, and shape of a simple distribution and understand what each of these descriptions refer to.	6-8.C.3.1a Identify probabilistic processes that simulate various forms of categorical variability, including uniform and normal distributions. e.g., spinner, dice, random draw 6-8.C.3.1b Illustrate variability in a dataset by determining how key descriptive features are represented. 6-8.C.3.1c Evaluate how visualizations, models, or predictions account for variation at an appropriate level.	9-10.C.3.1a Describe methods (e.g., statistical, simulation) to analyze variability in data and connect it to known or hypothesized processes in a specific domain.	11-12.C.3.1a Apply statistical or simulation methods to model variability to explore uncertainty in real-world situations.

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.

K–2	3–5	6–8	9–10	11–12
K-2.C.3.2a Describe how two things or groups are different from one another (e.g., more or less, bigger or smaller).	3-5.C.3.2a Understand how data varies by exploring spread (e.g., range) and comparing qualities (e.g., brightness or temperature).	6-8.C.3.2a Use visualizations (e.g., box plots) to compare variability across datasets.	9-10.C.3.2a Use simple statistics including mean, median, range, standard deviation, etc. to compare data distributions.	11-12.C.3.2a Explore variability through statistical methods, such as analyzing residuals or variance in linear models.

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.

K–2	3–5	6–8	9–10	11–12
	3-5.C.3.3a Identify and explain simple measurement error. e.g., different students' get varying results when measuring the same object	6-8.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	9-10.C.3.3a Consider variability as a key component of informal inference by questioning whether observed differences are meaningful or not. e.g., phone battery lasts 6 hours one day and 4 the next—is this a real difference in battery life, or just normal variation from daily use	11-12.C.3.3a Estimate and describe errors between predictions and actual outcomes. e.g., residuals, misclassification rates
	3-5.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	6-8.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	9-10.C.3.3b Identify categorical options for measuring "best" fit from data points to provided estimates. e.g., line or curve for a scatterplot, mean for a distribution	11-12.C.3.3b Analyze error patterns to assess model performance. e.g., residual plot, confusion matrix
			9-10.C.3.3c Consider both context and the characteristics/source of a dataset to determine how "messy" a dataset may be due to measurement error. e.g., faulty sensors, inaccurate or inappropriate measurements	11-12.C.3.3c Use insights from error analysis to improve the model. e.g., in linear regression, add a variable or use a curve; in classification, balance the groups or adjust the cutoff
			9-10.C.3.3d Use errors to improve the AI and/or machine learning model.	

Concept C.3.4

Variability in our computational world

21st-century skills: AI literacy

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.

K–2	3–5	6–8	9–10	11–12
		6-8.C.3.4a Conceptualize how the output of AI models such as LLMs vary along a variety of dimensions.	9-10.C.3.4a Acknowledge how variability in the training data for generative AI influences bias in its output. e.g., facial recognition, ownership of DNA data	11-12.C.3.4a Appreciate that many AI tools are pre-trained with large quantities of data so that inferences can be drawn on smaller sample sizes.
		6-8.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.

K–2	3–5	6–8	9–10	11–12
	3-5.C.4.1a Summarize data that is represented in a digital tool.	6-8.C.4.1a Summarize data across multiple categories using a digital tool. 6-8.C.4.1b Create single variable visualizations using a digital tool. 6-8.C.4.1c Identify relationships and patterns using a digital tool.	9-10.C.4.1a Identify relationships and patterns using a digital tool. 9-10.C.4.1b Clean and wrangle data using a digital tool. 9-10.C.4.1c Create multi-variable visualizations using digital tools.	11-12.C.4.1a Create models to perform simulations using a digital tool. 11-12.C.4.1b Perform data analysis using a digital tool.

Concept C.4.2

Tool ethics

21st-century skills

AI literacy

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.

K–2	3–5	6–8	9–10	11–12
		6-8.C.4.2a Describe how digital tools can be used to provide equitable access to learning experiences.	9-10.C.4.2a Describe the ethical limitations (e.g., environmental, privacy, copyright, hallucination) of using AI tools.	11-12.C.4.2a Critique the societal effect of AI by exploring issues surrounding bias, accountability, and transparency in decision-making using AI tools, as well as the effects on privacy, jobs, and policy.

Concept C.4.3

Tool evaluation

Assess the technical limitations of digital tools and compare no-code, low-code, and high-code solutions based on their capabilities and use cases.

K–2	3–5	6–8	9–10	11–12
			9-10.C.4.3a Identify the technical limitations of a digital tool.	11-12.C.4.3a Identify differences between a no-code, low-code, or high-code 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.

K–2	3–5	6–8	9–10	11–12
	3-5.C.4.4a Select a no-code digital tool that is suited for the intended task.	6-8.C.4.4a Select a no-code or low-code digital tool that is suited for the intended task.	9-10.C.4.4a Select a no-code, low-code or high-code digital tool that is suited for the intended task.	11-12.C.4.4a Select multiple digital tools suited for different tasks throughout the data investigation process. 11-12.C.4.4b Describe how digital tools are used in the workforce.

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.

K–2	3–5	6–8	9–10	11–12
		6-8.C.4.5a Explore the basics of block coding in data investigation processes. 6-8.C.4.5b Explore the basics of block coding in data analysis processes.	9-10.C.4.5a Recognize how computer code can automate data investigation processes. 9-10.C.4.5b Recognize how computer code can automate data analysis processes.	11-12.C.4.5a Recognize how computer code can be used to produce reproducible data analysis processes. 11-12.C.4.5b Recognize the advantages and limitations of using computer code compared to no-code or low-code tools.

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.

K–2	3–5	6–8	9–10	11–12
K-2.C.4.6a Recognize that some digital tools help people who have difficulty seeing, hearing, or using technology.	3-5.C.4.6a Identify tools that make data more accessible, such as screen-readers, captions, or tactile graphs.	6-8.C.4.6a Explore why it's important to present data in multiple formats to ensure all learners can understand it. 6-8.C.4.6b Explore how digital tools support diverse learners to analyze data. e.g., immersive readers, speech-to-text, translators, sonification 6-8.C.4.6c Discuss how the lack of accessible digital tools can exclude people from participating in data analysis.	9-10.C.4.6a Explore how to communicate with data while prioritizing accessibility. 9-10.C.4.6b Critique the levels of accessibility of digital tools and representations of data.	11-12.C.4.6a Design data visualizations that include accessible features such alt-text and text descriptions. 11-12.C.4.6b Examine how policies, limitations, and technological advancements impact the development of accessible digital tools.

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.

K–2	3–5	6–8	9–10	11–12
K-2.C.5.1a Understand that objects can be grouped based on similar characteristics. e.g., "all blue items go here," "this group has ground shapes"	3-5.C.5.1a Understand that grouping objects by shared characteristics creates rules that can be used to classify and categorize new objects.	6-8.C.5.1a Explore how relationships in data connect characteristics, including patterns like increasing, decreasing, or no connection.	9-10.C.5.1a Recognize that bivariate relationships between numerical features can be examined using both linear and non-linear associations.	11-12.C.5.1a Discern that different models, such as decision trees and neural networks, analyze patterns and relationships in data to make predictions.
K-2.C.5.1b Recognize that criteria for sorting objects helps to organize them and identify patterns. e.g., grouping buttons by shape reveals which shapes are most common	3-5.C.5.1b Recognize that patterns and relationships in data provide different kinds of information.	6-8.C.5.1b Recognize that models simplify complex systems and have limitations.	9-10.C.5.1b Investigate real-world examples where correlation does not imply causation.	11-12.C.5.1b Assess relationships in the context of uncertainty, bias, and reliability of the data.
K-2.C.5.1c Predict whether an object belongs to a group or category based on its characteristics.	3-5.C.5.1c Discuss how data relationships help describe real-world phenomena. e.g., taller plants tend to be older	6-8.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		11-12.C.5.1c Investigate how assumptions and bias influence a model's results.

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.

K–2	3–5	6–8	9–10	11–12
K-2.C.5.2a Articulate simple rules for sorting. e.g., "all blue items go here," "this group has ground shapes"	3-5.C.5.2a Predict whether an object belongs to a group or category based on its characteristics.	6-8.C.5.2a Identify relationships between variables and represent them using tables, graphs, or diagrams (e.g., decision trees, flowcharts).	9-10.C.5.2a Construct and analyze models to represent linear and non-linear relationships in data.	11-12.C.5.2a Develop models that incorporate multiple variables and explicitly consider interactions between them.
K-2.C.5.2b Classify objects based on their observed similarities and characteristics.	3-5.C.5.2b Distinguish patterns from relationships in data.	6-8.C.5.2b Use simple mathematical or computational models (e.g., statistical summaries, spreadsheet formulas) to describe patterns and relationships in data.	9-10.C.5.2b Use technology to create, test, and refine models.	11-12.C.5.2b Use computational methods, coding, or machine learning techniques to build and refine models.
K-2.C.5.2c Extend simple patterns based on observable characteristics. e.g., arranging objects by size		6-8.C.5.2c Test and refine models by comparing predictions to actual data values.	9-10.C.5.2c Evaluate and improve models by comparing predictions to observed data.	11-12.C.5.2c Assess assumptions, limitations, and biases in models to evaluate their impact on predictions in real-world scenarios.

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

K–2	3–5	6–8	9–10	11–12	Advanced
K-2.D.1.1a	3-5.D.1.1a	6-8.D.1.1a	9-10.D.1.1a	11-12.D.1.1a	
Recognize that some situations are not binary and find appropriate vocabulary to describe them. e.g., a classmate riding the bus to school could be "always," "sometimes," or "never"	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	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	Clearly state a result or finding and indicate the level of certainty regarding a formal statistical concept alongside an informal evaluation of the likelihood of the event.	Clearly state the result or finding and indicate the level of certainty regarding the statistical analysis and the quality of the evidence (e.g., dataset or source characteristics, similar findings in alternative data) as justification.	Clearly state a result or finding, along with the degree of certainty, using two or more advanced statistical methods (e.g., probability distributions, t-tests, z-tests, or bootstrapping/simulation), while justifying the conclusions with evidence (e.g., dataset or source characteristics, similar findings in alternative data) quality indicators like dataset characteristics, source reliability, and corroborating findings from alternative data.

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

K–2	3–5	6–8	9–10	11–12	Advanced
K-2.D.1.2a	3-5.D.1.2a	6-8.D.1.2a	9-10.D.1.2a	11-12.D.1.2a	
Discuss a guess or hypothesis before an investigation, compare the initial guess to findings, and informally express how new data changes a prior guess.	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.	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.	Analyze how confirmation bias and availability bias influence the way individuals evaluate new information, especially regarding their existing beliefs and assumptions.	Summarize previous assumptions and potential updates in written conclusions from a data analysis, and identify any known contradictory findings to mitigate confirmation bias.	Explain Bayes Theorem in formal conditional probability statements: P(A|B) = (P(A) * P (B|A)) / P(B), where A is the event in question and B is the event of new evidence related to A.
		6-8.D.1.2b	9-10.D.1.2b	11-12.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."	Relate assumptions about a problem to the certainty of findings based on new evidence.	Describe Bayes Theorem by explaining how it relates to conditional probability, which includes the probability of an event occurring, the probability of that event given the evidence is true, and the probability that the evidence itself is true.	Apply Bayes Theorem to an example result in an academic research finding or discussion.
				11-12.D.1.2c	
				Apply the logic of Bayes Theorem to determine whether a data-based claim in the media was accurately explained.	

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

K–2	3–5	6–8	9–10	11–12	Advanced
	3-5.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	6-8.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	9-10.D.1.3a Identify and accurately employ the Expected Value equation (EV = P (Xi) * Xi) across multiple contexts to compare scenarios involving multiple trials. e.g., insurance policies, lotteries	11-12.D.1.3a Justify the Expected Value equation (EV = P (Xi) * Xi) with formal probability statements and by explaining the Law of Large numbers.	
			9-10.D.1.3b Solve a real-world comparison problem using a digital spreadsheet, such as selecting insurance policies or entering different lotteries. e.g., choosing insurance policies, entering different lotteries	11-12.D.1.3b Apply the Expected Value equation to assess its fitness for the problem by determining the accuracy of the estimate based on the number of trials conducted. e.g., flipping a coin 100 times and determining if getting heads 30 times is reasonable when the expected value is getting heads 50 times	

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

K–2	3–5	6–8	9–10	11–12	Advanced
	3-5.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	6-8.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	9-10.D.1.4a Identify situations when distinguishing from random chance is especially important. e.g., medical drug trial, public policy implementation	11-12.D.1.4a Explain the concept of statistical significance (e.g., including its role in distinguishing meaningful results from random chance) in plain language and the limitations of significance testing (e.g., inability to address study design flaws, confounding variables, or real-world validity beyond a narrow model comparison).	Identify examples of p-value misuse in the media or academic research.
		6-8.D.1.4b Recognize that a unique result may be considered significant if it is substantially different from outcomes in similar situations.	9-10.D.1.4b Describe probability distributions and give real-world examples of how they can represent different types of random events.	11-12.D.1.4b Describe how statistical significance tests are constructed, calculated, and interpreted in the context of chosen probability models and/or assumptions.	Describe a p-value to without using the language of the "null hypothesis" or "alternative hypothesis."
		6-8.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.	9-10.D.1.4c Identify and describe a normal distribution as a possible model for random chance that can be used to determine whether a result is statistically significant.	11-12.D.1.4c Identify real-world instances where assessing statistical significance is crucial (e.g., scientific studies to distinguish actual effects from random variation) while also evaluating the significance claims made by others and recognizing situations where statistical significance is necessary but not sufficient for proving a point.	
				11-12.D.1.4d Differentiate statistical significance, effect size, and statistical power in simple terms with real-world examples, explaining how each addresses distinct questions in research. e.g., whether outcomes could be connected to random chance, the meaningfulness of impacts in context, the suitability of the analysis approach to the specific data and problems	

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.”

K–2	3–5	6–8	9–10	11–12	Advanced
K-2.D.1.5a Describe characteristics of a population and recognize that variability exists within any population. e.g., jelly beans in a jar	3-5.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?	6-8.D.1.5a Evaluate how different sampling methods impact the accurate representation of a population and their ability to generalize findings to other groups.	9-10.D.1.5a Use simulations in a digital software to help determine whether the results of an experiment are likely due to something other than random chance.	11-12.D.1.5a Use simulation-based inferential methods at large N to draw conclusions from a dataset using digital software.	Execute and correctly interpret the margin of error, confidence interval, and standard deviation in a data analysis software for a given summary statistic.
K-2.D.1.5b Describe events as either likely, meaning they happen a lot, or unlikely, meaning they happen a little. e.g., pulling a purple jelly bean out of a jar with mostly red jelly beans	3-5.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	6-8.D.1.5b Assess how sample size impacts the accuracy of estimates representing population characteristics.	9-10.D.1.5b Analyze how dataset bias impacts sample results over time by introducing intentional bias sources in digital simulations and observing their effects.	11-12.D.1.5b Identify why simulation can be used to infer conclusions about a population referencing the Law of Large Numbers.	Describe the relationship between the margin of error, confidence intervals, and standard deviation, in both words and in their formal mathematical definitions.
		6-8.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.	9-10.D.1.5c Answer probabilistic questions resulting from a simulation.	11-12.D.1.5c Interpret margin of error and confidence intervals for a given sample.	
		6-8.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.

K–2	3–5	6–8	9–10	11–12	Advanced
	3-5.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.	6-8.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.	9-10.D.1.6a Recognize that a randomized experiment is the best way to establish evidence for causation and justify a claim through isolating the effect of only one independent variable on another variable at a time.	11-12.D.1.6a Independently identify examples of two dependent variables that are both influenced by a third variable in real-world data. e.g., coffee consumption and lower risk of disease are both affected by an active lifestyle	Justify a causal relationship in a multivariable dataset with real-world data, including additional datasets gathered from outside sources and connect the analysis to existing research literature.
		6-8.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.	9-10.D.1.6b Identify spurious correlations in the media and explore other potential causes that may explain these associations when applicable.	11-12.D.1.6b Identify spurious correlations in the media and analyze how they relate to media claims and AI recommendations. e.g., ice cream sales and shark attacks both increase in the summer; they're both linked to hot weather, not each other	Explain why a chosen analysis method effectively isolates an effect.

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.

K–2	3–5	6–8	9–10	11–12	Advanced
	3-5.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	6-8.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.	9-10.D.1.7a Explain why randomization mitigates many potential sample biases (e.g., observation bias, collection errors, selection bias) concurrently in a variety of examples.	11-12.D.1.7a Recognize that randomization can happen in various settings, regardless of the intervention or events involved. e.g., artificial interventions, accidental or chance events, unrelated to the question of interest	Implement randomization using a random seed in a simulation technique using a computer-based analysis tool to compare sampling techniques (e.g., sampling with replacement or without replacement).
				11-12.D.1.7b Differentiate between lab experiments and natural experiments in scenario-based questions.	

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

K–2	3–5	6–8	9–10	11–12	Advanced
	3-5.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	6-8.D.1.8a Distinguish direct vs. inverse relationships in multivariate data, such as associations between two categorical groups within the same visualization.	9-10.D.1.8a Use computer software to explore how adding additional numerical variables to a linear model changes the interpretation of the results.	11-12.D.1.8a Use computer software to analyze the relationship between an independent and dependent variable in a linear model by changing the number and combination of dependent variables.	Use computer software to incorporate categorical variables into a linear regression model.
		6-8.D.1.8b Use color to differentiate categories in a scatterplot and identify patterns in their relationships.	9-10.D.1.8b Use computer software to analyze the relationship between two or more numerical variables by interpreting the strength and direction (e.g., positive, negative, none) of the association using computed values.	11-12.D.1.8b Evaluate how changes to the number and combination of dependent variables affect the model by interpreting R-squared and regression coefficients.	Analyze and interpret the regression coefficients to understand the effect of the categories on the model.
		6-8.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.		11-12.D.1.8c Explore how polynomials of different degrees fit scatterplots.	Create an “ideal” multi-variable model for real-world data in a computer-based software that explains as much variance as possible, without overfitting a model. Justify how you have found the “ideal” model by comparing R^2, covariance, and the number of variables chosen in relation to their real-world context.
				11-12.D.1.8d Analyze how increasing or decreasing the degree of a polynomial can lead to potential overfitting or underfitting the data.	

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.

K–2	3–5	6–8	9–10	11–12	Advanced
K-2.D.2.1a Ask a question that you answer by counting or measuring something.	3-5.D.2.1a Ask or identify a question that you answer by counting or measuring results from different groups.	6-8.D.2.1a Ask or identify a question that can be verified with data collected through observations.	9-10.D.2.1a Differentiate query-based, hypothesis-based, and causal questions by their focus on trends, uniqueness of outcomes, and causal relationships, respectively.	11-12.D.2.1a Develop a causal diagram to map relationships among multiple variables and create an iterative analysis plan to test each relationship with data.	
	3-5.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).	6-8.D.2.1b State a guess or potential answer to a question for later verification or testing via a hypothesis.	9-10.D.2.1b Assess query-based questions by establishing a threshold of satisfaction for certainty in interval estimates (e.g., if it applies 95% of the time, I find it acceptable).	11-12.D.2.1b For query-based questions, estimate a confidence interval and margin of error in a real-world data analysis project with software.	
			9-10.D.2.1c Assess hypothesis-based questions by debating the condition of uniqueness (e.g., if it occurs 5% of the time or less).	*11-12.D.2.1c* For hypothesis-based questions, estimate a p-value based on a proposed statistical model for real-world data with software.	

Concept D.2.2

Iteration, validation, and multiple explanations

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

21st-century skills

Durable skills

K–2	3–5	6–8	9–10	11–12	Advanced
K-2.D.2.2a	3-5.D.2.2a	6-8.D.2.2a	9-10.D.2.2a	11-12.D.2.2a	
Categorize, count, or measure variables and verify the categorization with peers for agreement.	Estimate the total count of a characteristic within a group, providing several reasons to support the accuracy of your estimate.	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.	Identify various possible explanations for an observed association by investigating and comparing relationships between variables within a dataset.	Highlight unusual associations or outcomes in an analysis document by validating analysis steps and investigating other parts of the dataset.	Document analysis steps and errors while implementing validation checks in the software for data wrangling.
	3-5.D.2.2b	6-8.D.2.2b	9-10.D.2.2b	11-12.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.	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.	Regularly log questions during data analysis and identify additional factors that may clarify associations. e.g., knowing X would be helpful because it would explain or rule out Y	Identify potential counter-arguments or alternative explanations that may refute one's conclusions drawn from data, and suggest mitigation strategies that could be tried in the future with additional data or new research.	Execute an alternative analysis plan to validate a significantly different result from the initial method.

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.

K–2	3–5	6–8	9–10	11–12	Advanced
	3-5.D.2.3a	6-8.D.2.3a	9-10.D.2.3a	11-12.D.2.3a	
	Identify reasons to support and refute conclusions when drawing insights from data.	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.	Identify potential issues in data investigations and state what cannot be reasonably concluded from the available data and approach, noting areas that may require further investigation.	Evaluate the potential limitations of statistical findings by considering the data collection methods, sample selection, and simplifications that may not capture the complexity of real-world scenarios.	

Concept D.2.4

Relevant conclusions

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

21st-century skills

Durable skills

K–2	3–5	6–8	9–10	11–12	Advanced
K-2.D.2.4a	3-5.D.2.4a	6-8.D.2.4a	9-10.D.2.4a	11-12.D.2.4a	
Identify from among a set of given examples what types of data are needed to answer a given investigation question.	Propose types of data and/or data comparisons that are relevant for answering a given investigation question.	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.	Formulate a statement that directly addresses the original investigation question, incorporates relevant statistical data to substantiate the conclusion, and interprets the statistical results to explain their broader implications in practice. e.g., statistical claims are not solely about numbers, they also interpret what the results signify and why they are important for solving a problem or answering a question	Determine if a causal claim can be established based on the investigation's design (e.g., natural experiments, real-world observations) and describe the differences between expectations and the design.	
	3-5.D.2.4b	6-8.D.2.4b	9-10.D.2.4b		
	Identify types of data and/or data comparisons that are NOT relevant for answering a given investigation question.	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.	Identify statements that do NOT include descriptions of the data and context implications that address the original 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

K–2	3–5	6–8	9–10	11–12	Advanced
K-2.D.3.1a Identify whether a pattern applies to another situation or not.	3-5.D.3.1a Recognize that a result or pattern from data does not always extend to other situations.	6-8.D.3.1a Identify various factors that may cause data in a dataset to insufficiently represent or apply to other situations.	9-10.D.3.1a Examine and identify common generalization issues from data-based conclusions in the media.	11-12.D.3.1a Analyze a data generalization issue in media or real-world situations and discuss its significant impacts and the importance of addressing generalization errors.	
		6-8.D.3.1b Identify characteristics of data-based predictions that easily and do not easily generalize to many situations.	9-10.D.3.1b Identify and list analysis strategies for a given data-driven conclusion to better generalize to other populations or situations.	11-12.D.3.1b Implement multiple strategies to generalize data-based conclusions to new populations or situations. e.g., add additional context or control variables, repeat the analysis with new collection or sample, test a model with a different dataset	
				11-12.D.3.1c Evaluate the advantages and disadvantages of automated tools that rely on large datasets for universal predictions. e.g., prediction algorithm for airline ticket prices or home mortgage application assessment, AI model for facial recognition, autonomous vehicle model trained on city roads	

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

K–2	3–5	6–8	9–10	11–12	Advanced
	3-5.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	6-8.D.3.2a Evaluate a population based on a sample by making informal arguments for the sample's sufficiency in answering the question.	9-10.D.3.2a Analyze a population through a sample by clearly articulating how the chosen sampling method relates to the research question.	11-12.D.3.2a Evaluate the suitability of different sampling methods (e.g., random sample with or without replacement) for the specific question and available data.	Identify machine learning methods such as supervised, unsupervised, and reinforcement learning, and discuss the pros and cons of each when data on the entire population or a very detailed sample with many variables is available.
		6-8.D.3.2b Identify potential weaknesses in a given sample that may limit its ability to represent a broader population or phenomenon.		11-12.D.3.2b Identify situations in which data on the full population is easily available or even critical to answer a question of interest, and traditional sampling-methods are not required.	

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

K–2	3–5	6–8	9–10	11–12	Advanced
	3-5.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	6-8.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 6-8.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	9-10.D.3.3a Recognize there are formal methods to determine the minimum sample size needed to make a well-supported claim about a population. 9-10.D.3.3b Explain "statistical power" of a statistical test as the general probability that an outcome "lands" more "extremely," beyond an arbitrary pivotal value set for statistical significance that a researcher chooses. 9-10.D.3.3c Explain "statistical power" as the probability that a statistical test properly detects a real effect when one exists.	11-12.D.3.3a Make an informal power analysis for an analysis or experimental setup using real-world data and a hypothesis, including claims about the 1) Effect Size 2) Sample Size 3) Statistical Significance and 4) Statistical Power. 11-12.D.3.3b Use the simple equation Power = 1 - β to visually show the difference between a normal distribution of outcomes and an abnormal distribution of outcomes.	Make a formal Power Analysis by identifying a sufficient sample size for a real-world data exploration. Students should mathematically isolate "n" in a t-test or z-test, and estimate Power with a software tool.

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

K–2	3–5	6–8	9–10	11–12	Advanced
		6-8.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.	9-10.D.3.4a Acknowledge that a sample may be systematically skewed due to collection methods, data availability, survey design, or other reasons, particularly in a secondary data context. 9-10.D.3.4b Identify examples of sample bias in the media or other real-world examples. e.g., medical drug trials, prior debunked research	11-12.D.3.4a Propose and implement at least two methods to mitigate sample bias in a real-world dataset. e.g., adding additional data, making a new variable with a correction, explicitly stated assumption	Estimate bias by interpreting and applying the formula for a biased estimator.

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

K–2	3–5	6–8	9–10	11–12	Advanced
		6-8.D.3.5a Identify additional possible scenarios for which a data-based conclusion may apply, beyond the original question or inquiry.	9-10.D.3.5a Identify additional scenarios for which a data-based conclusion may apply and list the similarities and differences of the new scenario. 9-10.D.3.5b Identify the risks of extending the original analysis to a new scenario. e.g., data that might not be captured, incorrect assumptions	11-12.D.3.5a Identify and implement at least two strategies in a project-based activity that utilize original data to address questions in a new scenario. 11-12.D.3.5b Describe potential ethical and statistical issues with the extension strategies, including explicit caveats on any conclusions reached with real-world data.	

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

K–2	3–5	6–8	9–10	11–12	Advanced
		6-8.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).	9-10.D.3.6a Create and compare subsets of a dataset with software.	11-12.D.3.6a Identify and explain Simpson's Paradox: an average trend may disappear or even reverse when individual subsets and/or groupings are examined.	
			9-10.D.3.6b Discuss examples of aggregate measures of data that missed important subsets in the media or other real-world contexts.	11-12.D.3.6b Review examples of Simpson's Paradox in the media and in well-known research studies.	

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

K–2	3–5	6–8	9–10	11–12	Advanced
	3-5.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	6-8.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	9-10.D.3.7a Recognize that one study or data analysis may be insufficient to prove something is "true" for certain.	11-12.D.3.7a Recognize the importance of many trials, study validation, and meta-analyses in academic research.	
			9-10.D.3.7b Document data analysis steps in a shareable and reproducible format that can be repeated.	11-12.D.3.7b Document data analysis steps in a shareable and reproducible format for collaboration platforms (e.g., GitHub, Bitbucket).	

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.

K–2	3–5	6–8	9–10	11–12	Advanced
K-2.E.1.1a Create data visualizations to represent an aspect of the student's daily life. e.g., draw a map of the playground and stick figures to represent each student's favorite apparatus	3-5.E.1.1a Create data visualizations to summarize categorical data.	6-8.E.1.1a Create data visualizations that use multiple variables.	9-10.E.1.1a Use computer-based analysis tools to make basic descriptive summaries of a dataset. e.g., bar charts, histograms, line graphs, scatterplots	11-12.E.1.1a Create data visualizations that illustrate complex bivariate relationships. e.g., exponential, quadratic	Demonstrate presentation skills to fully communicate depth and breadth of a visualization to an audience.
K-2.E.1.1b Create bar graphs and picture graphs to represent a small data set. e.g., whole number scales, concrete situations	3-5.E.1.1b Display groups or categories in visualizations using complementary or contrasting colors to highlight differences.	6-8.E.1.1b Create a data visualization, collect feedback from the target audience, and revise the visualization based on feedback.	9-10.E.1.1b Quickly or informally estimate relationships visually by adding lines of best fit with a computer-based tool.	11-12.E.1.1b Edit data visualizations to optimize it for your intended audience and the audience's different needs. e.g., "chartjunk" can be distracting for some audience but necessary for others	Present both 1) basic visual summaries of the data 2) additional visualizations that "go deeper" into the story the data is telling, and relationships discovered within the data visualization e.g., new relationships within subsets, significant outliers, complex or overlapping control variables
	3-5.E.1.1c Display continuously scaled data in visualization using shading.	6-8.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

Investigate with visualizations

21st-century skills

Durable skills

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

K–2	3–5	6–8	9–10	11–12	Advanced
K-2.E.1.2a Ensure students can use tally marks or manipulatives (e.g., stickers, blocks) to represent data.	3-5.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	6-8.E.1.2a Use visualizations of common data distributions to identify potential errors in the data. e.g., outliers, out-of-bounds values	9-10.E.1.2a Visualize the distribution of raw data to identify outliers and out-of-bounds values in context.	11-12.E.1.2a Create data visualizations of raw data and increasingly aggregated forms of the same data to help understand the nuances of the data.	
	3-5.E.1.2b Organize and present collected data visually to highlight relationships and to support a claim.	6-8.E.1.2b Visualize the distribution of data to illustrate the shape, spread, and measures of center informally.	9-10.E.1.2b Communicate key features of distribution (e.g., measures of center, spread, shape) formally and with precision.	11-12.E.1.2b Strategically use data visualization to identify potential outliers, errors, and unexpected findings, while clearly stating and justifying any reasons for excluding certain potentially erroneous observations.	
		6-8.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

Clear design for user interpretation

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

21st-century skills

Durable skills

K–2	3–5	6–8	9–10	11–12	Advanced
K-2.E.1.3a Identify and describe the parts (e.g., titles, labels, legends, colors) of bar graphs and picture graphs and what they communicate.	3-5.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.	6-8.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.	9-10.E.1.3a Properly cite data sources near visuals to ensure transparency and credibility.	11-12.E.1.3a Provide context for the data to help viewers understand the background and implications.	Apply design principles such as balance, emphasis, and simplicity to make visualizations clear and engaging.
	3-5.E.1.3b Reliably use the parts (e.g., titles, labels, legends, colors) of bar graphs, picture graphs, and line graphs.	6-8.E.1.3b Choose or create a representation and color palette for one or two-variable data, and explain or defend their choice.	9-10.E.1.3b Recognize how complementary or contrasting features (e.g., color, texture, shape) can be used to represent dichotomous ideas in data visualizations.	11-12.E.1.3b Recognize how color theory (e.g., tint, saturation, shading) can be used to represent continuously scaled data (e.g., darker color =higher concentration of occurrence).	Understanding the basics of interactive visualizations (e.g., tooltips, zooming) and their advantages in data exploration.
			9-10.E.1.3c Describe how human color/contrast perception varies and apply this to select accessible data visualization palettes.	11-12.E.1.3c Recognize that we have culturally-influenced or domain-specific ways of using and interpreting chart elements. Consider the conventions that are known to or expected by your audience when developing data visualizations.	

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

K–2	3–5	6–8	9–10	11–12	Advanced
K-2.E.1.4a Answer questions about whole number numerical data or categorical data represented visually (e.g., bar graphs, picture graphs).	3-5.E.1.4a Answer questions about fractional valued numerical data or categorical data represented visually with one or two variables.	6-8.E.1.4a Answer questions about continuous numerical scaled data, location data, and/or categorical data represented visually with multiple variables.	9-10.E.1.4a Answer questions about and explain the data in a variety of data visualizations, including non-standard visualizations. Extract key insights, trends, and patterns from the data.	11-12.E.1.4a Understand how uncertainty around point and effect estimates are communicated on data visualizations with error bars.	Describe the potential relationships (or lack thereof) represented in scatterplots (including linear, exponential, logarithmic, polynomial, and piecewise) and debate which function is the best representation for the shape and context.
K-2.E.1.4b Recognize how data is organized into categories or groups. e.g., each bar or picture represents a certain quantity in a graph	3-5.E.1.4b Recognize unusual data points and consider reasons why they might appear.	6-8 E.1.4b Describe the relationships (or lack thereof) represented in scatterplots (e.g., direct vs. inverse, positive vs. negative).	9-10.E.1.4b Describe the potential relationships (or lack thereof) represented in scatterplots (including linear, exponential, and logarithmic) and debate which function is the best representation for the shape and context.	11-12.E.1.4b Evaluate the effectiveness of data visualizations, including the risk of misleading the reader.	Visualize confidence intervals or margins of error using error bars with computer-based software.
	3-5.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.	6-8.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.			Visualize margins of error of a continuous variable using error bands with a computer-based software.

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

K–2	3–5	6–8	9–10	11–12	Advanced
K-2.E.1.5a	3-5.E.1.5a	6-8.E.1.5a	9-10.E.1.5a	11-12.E.1.5a	
Compare and/or contrast different representations of the same data, including physical models (e.g., block towers), bar graphs, and picture graphs, and describe how they differ.	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	Compare and/or contrast various representations of data sets with multiple features and describe what is emphasized, de-emphasized, or obscured in each representation.	Compare and/or contrast visualizations of the same numerical data at different scales and understand how the scale affects people's interpretation. e.g., accurately representing the relative magnitudes vs. exaggerating them	Compare and/or contrast various representations of relative frequencies and proportions, identify elements of each representation that facilitate or hinder the identification of relative proportions, and explain the reasoning behind conventions. e.g., ordered or unordered stacked bar graph	Compare and/or contrast 2D and 3D bar graphs and pie charges and identify how unnecessary use of three dimensions obfuscates the relative frequencies and/or proportions of the data.
K-2.E.1.5b		6-8.E.1.5b	9-10.E.1.5b	11-12.E.1.5b	
Use visual cues to interpret data. e.g., bar graph, taller bars mean higher frequency		Describe how different ways of representing data can improve clarity or mislead.	Critique misleading visualizations, such as those with truncated axes, cherry-picked data points, confusing colors, or manipulated scales. e.g., graph starting at 50 (not 0) can make a 5% drop look like a crash	Compare and/or contrast various ways to represent distributions and their measures of center (e.g., histograms, density plots, box plots) by plotting two distributions on the same graph and explaining how different representations facilitate or hinder the visibility of differences and associations.	Compare and/or contrast varying bin sizes to demonstrate how different degrees of granularity in a histogram or other visualization type can lead to different interpretations.

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.

K–2	3–5	6–8	9–10	11–12	Advanced
K-2.E.1.6a	3-5.E.1.6a	6-8.E.1.6a	9-10.E.1.6a	11-12.E.1.6a	
Explore hands-on activities with students, helping them understand the difference between data types. e.g., bar graphs for numerical data such as votes of favorite fruit, picture graph for categorical data such as favorite colors	Visualize multiple types of data (e.g., numeric, categorical, string data) during in-class data collection exercises.	Describe and discuss the typical visualization characteristics of numeric, categorical, and string data while identifying and outlining the differences between them.	Demonstrate the wrong type of data (e.g., numeric, categorical, string) entered into a misaligned visualization package (e.g., scatterplot of categorical data) and explain why the visualization fails to work or clearly represent the data.	Produce a data visualization parallel to the type of data (e.g., numeric, categorical, string, image, unstructured).	
				11-12.E.1.6b	
				Defend your visualization choice to others and explain the data type and visualization type including suitability for continuous or discrete variables.	

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.

K–2	3–5	6–8	9–10	11–12	Advanced
K-2.E.2.1a	3-5.E.2.1a	6-8.E.2.1a	9-10.E.2.1a	11-12.E.2.1a	
Understand how visualizations (e.g., graphs, charts, maps, diagrams) can be created and used to tell a story about ourselves. e.g., picture graph showing how many students walk vs. ride the bus	Evaluate the effectiveness of text, visualization, and text plus a visualization to communicate a particular story.	Evaluate the degree to which visualizations and their surrounding text and context match and support one another.	Evaluate the degree to which visualizations and their surrounding text match and support real-world context.	Evaluate the degree to which visualizations and their surrounding text match and support a real-world argument or broader explanation of social, economic, scientific, or political factors.	
K-2.E.2.1b	3-5.E.2.1b	6-8.E.2.1b			
Make a prediction based on a visualization and the trend of the data relationship it conveys using designations "it will happen, it won't happen, it might happen."	Make a prediction based on a visualization using the terms: "likely, unlikely, certain, and impossible."	Recognize that data visualizations need explanations to tell their story.			

Concept E.2.2

Write data stories

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

21st-century skills

Durable skills

K–2	3–5	6–8	9–10	11–12	Advanced
K-2.E.2.2a Describe data represented in very basic ways and begin to link observations.	3-5.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."	6-8.E.2.2a Explain what the data reveals and whether it supports or contradicts any claims initially made.	9-10.E.2.2a Explain how the data directly supports or contradicts any claims made about it while also being open about limitations such as sample size or external factors that may influence results, and anticipate potential counterarguments.	11-12.E.2.2a Make and defend arguments using key features from a data visualization.	Use complex visualizations like multivariable graphs, scatter plots, heat maps, or interactive dashboards to present data clearly. Then, develop a research paper or presentation to explain the background, methodology, and context of the data, using visualizations to provide evidence of their findings and conclusions.
K-2.E.2.2b Demonstrate the ability to utilize simple questions to spark interest in a story conveyed by data, and effectively illustrate the data's narrative through visual representations such as pictographs or charts based on student responses. e.g., utilizing smiley face stickers to answer "do you like rain or sun more?"	3-5.E.2.2b When describing the data, decide whether any claim made about the data makes sense.	6-8.E.2.2b Create a visualization based on a 3-5 sentence narrative describing a particular environmental phenomenon involving multiple variables.	9-10.E.2.2b Support claims by citing expert opinions or research studies that corroborate the data.	11-12.E.2.2b Clearly define the claim by making it specific, measurable, and actionable.	
	3-5.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.	6-8.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.	9-10.E.2.2c Use data to explain trends and predict future outcomes based on those trends.	11-12.E.2.2c Ensure the data directly addresses the claim being defended.	
				11-12.E.2.2d Address potential confounding variables and factors in claim-making, and if possible, demonstrate how the data controls for those confounding variables and factors.	
				11-12.E.2.2e Discuss a claim's broader implications in writing, including societal effects. e.g., a graph showing declining crime might ignore rising cybercrime	

Concept E.2.3

Adapt storytelling

Tailor storytelling for different audiences.

21st-century skills

Durable skills

K–2	3–5	6–8	9–10	11–12	Advanced
K-2.E.2.3a Share simple data findings with peers and adults using age-appropriate language. e.g., "more kids like apples than bananas," "this tower has 3 more blocks than that one"	3-5.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	6-8.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.	9-10.E.2.3a Identify an audience of interest, and tailor data stories to that audience, presenting the data in a way that ensures it resonates with them.	11-12.E.2.3a Write data analyses and stories using plain-language vocabulary along with relevant problem-specific terms, ensuring adaptability to various audiences, both technical and non-technical, with clear explanations of why the content is important for each audience.	
	3-5.E.2.3b Provide the appropriate level of context for various audiences.	6-8.E.2.3b Use visuals to enhance understanding and/or incorporate interactive discussion about the data and the narrative.	9-10.E.2.3b Explain the implications and takeaways by detailing how the information can be utilized in their daily lives or work experiences, while offering actionable advice that aligns with their interests and needs.	11-12.E.2.3b Provide multiple representations of data relevant to individual arguments. e.g., visualizations, summary statistics, and descriptions of processes or methodologies	

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

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

21st-century skills

Media literacy and digital citizenship

K–2	3–5	6–8	9–10	11–12	Advanced
K-2.E.3.1a Investigate the origins of the data underlying a visualization and ask students about the sources of the data.	3-5.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.	6-8.E.3.1a Identify the reason a data representation was created. e.g., to persuade, present factual information	9-10.E.3.1a Evaluate the source, methodology, sample size, and any potential biases in data collection that may impact the reliability of the data narrative.	11-12.E.3.1a Communicate and present the source of the data used for the data visualization to ensure transparency.	
	3-5.E.3.1b Identify situations when data can be used to make decisions at school or at home.	6-8.E.3.1b Identify potential biases of the source of data used to create a visualization.	9-10.E.3.1b Evaluate the potential agenda(s) or motivation(s) of the author of a data visualization.	11-12.E.3.1b Examine the significance of the data being visualized by understanding what it measures and its relevance to real-world issues or scenarios.	
		6-8.E.3.1c Communicate the limitations of data visualizations based on the source of data used in to create it.	9-10.E.3.1c Understand standard journalistic practices, including fact checking and source verification, that support accurate reporting and help combat misinformation.	11-12.E.3.1c Examine how institutions (e.g., government, businesses, nonprofit organizations) utilize big data to achieve policy goals while considering the benefits and harms to the public and their implications for civic behavior.	
			9-10.E.3.1d Analyze situations when institutions have made big decisions based on untrustworthy data and describe the consequences.		

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.

K–2	3–5	6–8	9–10	11–12	Advanced
K-2.E.3.2a Understand how data can reflect real-world situations and help answer simple questions.	3-5.E.3.2a Develop creative data visualizations to depict an aspect of the student's community or social connections.	6-8.E.3.2a Collect personal data and use it to benefit their family or classroom.	9-10.E.3.2a Construct a data story to enact change in your community.	11-12.E.3.2a Explain how data science connects to other disciplines to solve major problems around the globe.	
	3-5.E.3.2b Draw simple conclusions about the data from a narrative.	6-8.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.	9-10.E.3.2b Analyze data narratives related to social and/or political issues and explore how different presentations of the data could alter its impact on communities and daily life.	11-12.E.3.2b Discuss strategies to mitigate harmful predictions derived from a data story, such as the varying injury rates from crash test dummies among different groups of drivers.	

Concept E.3.3

Civic data practices

Engage in civic practice and dispositions through recognition of the role data plays in civic society.

K–2	3–5	6–8	9–10	11–12	Advanced
			9-10.E.3.3a Access open government data from local, state, and/or Federal websites.	11-12.E.3.3a Develop democratic dispositions through evaluation of local data. e.g., review local election data, housing data in local city or county	Pick a local issue of student interest and draft a Letter to the Editor (LTE) to a local news outlet or to a local politician based on conclusions from public-access datasets.
			9-10.E.3.3b Leverage open government data to supplement or contextualize a data analysis project. e.g., U.S. Census	11-12.E.3.3b Pick a local issue of student interest and based on a data analysis project, submit a Public Comment.	

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

K–2	3–5	6–8	9–10	11–12	Advanced
K-2.E.3.4a Recognize that technology needs energy (e.g., electricity, batteries) to work and that we should use technology wisely.	3-5.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	6-8.E.3.4a Recognize that data collection practices, tools, representations and resulting consequences are unevenly distributed across the globe.	9-10.E.3.4a Recognize the environmental cost of running large data centers and AI/ML models while considering the costs versus benefits of nuclear power and evaluating solar and wind options for clean energy. 9-10.E.3.4b Evaluate an impactful data story and its societal implications. e.g., historical heart disease research impacts for men and women	11-12.E.3.4a Consider the environmental and human costs of harvesting natural resources for the creation of modern technologies. e.g., mining of lithium, geopolitical issues with high precision silicon	