**Mathematics**

**2016 Standards of Learning**



Data Science

Curriculum Framework

**Board of Education**

**Commonwealth of Virginia**

Copyright © 2022

by the

Virginia Department of Education

P.O. Box 2120

Richmond, Virginia 23218-2120

[http://www.doe.virginia.gov](http://doe.virginia.gov/)

All rights reserved. Reproduction of these materials for instructional purposes in public school classrooms in Virginia is permitted.

**Superintendent of Public Instruction**

James F. Lane

**Chief Academic Officer/Assistant Superintendent** **for Instruction**

Michael Bolling

**Office of STEM and Innovation**

Brendon Albon, STEM and Innovation Director

Tina Mazzacane, Mathematics Coordinator

Kristin Williams-Faus, Mathematics and Special Education Specialist

Debra Delozier, Mathematics Specialist

**NOTICE**

The Virginia Department of Education does not unlawfully discriminate on the basis of race, color, sex, national origin, age, or disability in employment or in its educational programs or services.

**Virginia 2016 *Mathematics Standards of Learning* *Curriculum Framework***

**Introduction**

The 2016 *Mathematics Standards of Learning* *Curriculum Framework*, a companion document to the 2016 *Mathematics Standards of Learning*, amplifies the *Mathematics Standards of Learning* and further defines the content knowledge, skills, and understandings that are measured by the Standards of Learning assessments. The standards and *Curriculum Framework* are not intended to encompass the entire curriculum for a given grade level or course. School divisions are encouraged to incorporate the standards and *Curriculum Framework* into a broader, locally designed curriculum. The *Curriculum Framework* delineates in greater specificity the minimum content that all teachers should teach and all students should learn. Teachers are encouraged to go beyond the standards as well as to select instructional strategies and assessment methods appropriate for all students.

The *Curriculum Framework* also serves as a guide for Standards of Learning assessment development. Students are expected to continue to connect and apply knowledge and skills from Standards of Learning presented in previous grades as they deepen their mathematical understanding. Assessment items may not and should not be a verbatim reflection of the information presented in the *Curriculum Framework*.

Each topic in the 2016 *Mathematics Standards of Learning* *Curriculum Framework* is developed around the Standards of Learning. The format of the *Curriculum Framework* facilitates teacher planning by identifying the key concepts, knowledge, and skills that should be the focus of instruction for each standard. The *Curriculum Framework* is divided into two columns: Understanding the Standard and Essential Knowledge and Skills. The purpose of each column is explained below.

*Understanding the Standard*

This section includes mathematical content and key concepts that assist teachers in planning standards-focused instruction. The statements may provide definitions, explanations, examples, and information regarding connections within and between grade level(s)/course(s).

*Essential Knowledge and Skills*

This section provides a detailed expansion of the mathematics knowledge and skills that each student should know and be able to demonstrate. This is not meant to be an exhaustive list of student expectations.

**Mathematical Process Goals for Students**

The content of the mathematics standards is intended to support the following five process goals for students: becoming mathematical problem solvers, communicating mathematically, reasoning mathematically, making mathematical connections, and using mathematical representations to model and interpret practical situations. Practical situations include real-world problems and problems that model real-world situations.

**Mathematical Problem Solving**

Students will apply mathematical concepts and skills and the relationships among them to solve problem situations of varying complexities. Students also will recognize and create problems from real-world data and situations within and outside mathematics and then apply appropriate strategies to determine acceptable solutions. To accomplish this goal, students will need to develop a repertoire of skills and strategies for solving a variety of problems. A major goal of the mathematics program is to help students apply mathematics concepts and skills to become mathematical problem solvers.

**Mathematical Communication**

Students will communicate thinking and reasoning using the language of mathematics, including specialized vocabulary and symbolic notation, to express mathematical ideas with precision. Representing, discussing, justifying, conjecturing, reading, writing, presenting, and listening to mathematics will help students clarify their thinking and deepen their understanding of the mathematics being studied. Mathematical communication becomes visible where learning involves participation in mathematical discussions.

**Mathematical Reasoning**

Students will recognize reasoning and proof as fundamental aspects of mathematics. Students will learn and apply inductive and deductive reasoning skills to make, test, and evaluate mathematical statements and to justify steps in mathematical procedures. Students will use logical reasoning to analyze an argument and to determine whether conclusions are valid. In addition, students will use number sense to apply proportional and spatial reasoning and to reason from a variety of representations.

**Mathematical Connections**

Students will build upon prior knowledge to relate concepts and procedures from different topics within mathematics and see mathematics as an integrated field of study. Through the practical application of content and process skills, students will make connections among different areas of mathematics and between mathematics and other disciplines, and to real-world contexts. Science and mathematics teachers and curriculum writers are encouraged to develop mathematics and science curricula that support, apply, and reinforce each other.

**Mathematical Representations**

Students will represent and describe mathematical ideas, generalizations, and relationships using a variety of methods. Students will understand that representations of mathematical ideas are an essential part of learning, doing, and communicating mathematics. Students should make connections among different representations – physical, visual, symbolic, verbal, and contextual – and recognize that representation is both a process and a product.

**Instructional Technology**

The use of appropriate technology and the interpretation of the results from applying technology tools must be an integral part of teaching, learning, and assessment. However, facility in the use of technology shall not be regarded as a substitute for a student’s understanding of quantitative and algebraic concepts and relationships or for proficiency in basic computations. Students must learn to use a variety of methods and tools to compute, including paper and pencil, mental arithmetic, estimation, and calculators. In addition, graphing utilities, spreadsheets, calculators, dynamic applications, and other technological tools are now standard for mathematical problem solving and application in science, engineering, business and industry, government, and practical affairs.

Calculators and graphing utilities should be used by students for exploring and visualizing number patterns and mathematical relationships, facilitating reasoning and problem solving, and verifying solutions. However, according to the National Council of Teachers of Mathematics, “… the use of calculators does not supplant the need for students to develop proficiency with efficient, accurate methods of mental and pencil-and-paper calculation and in making reasonable estimations.” State and local assessments may restrict the use of calculators in measuring specific student objectives that focus on number sense and computation. On the grade three state assessment, all objectives are assessed without the use of a calculator. On the state assessments for grades four through seven, objectives that are assessed without the use of a calculator are indicated with an asterisk (\*).

**Computational Fluency**

Mathematics instruction must develop students’ conceptual understanding, computational fluency, and problem-solving skills. The development of related conceptual understanding and computational skills should be balanced and intertwined, each supporting the other and reinforcing learning.

Computational fluency refers to having flexible, efficient and accurate methods for computing.  Students exhibit computational fluency when they demonstrate strategic thinking and flexibility in the computational methods they choose, understand and can explain, and produce accurate answers efficiently.

The computational methods used by a student should be based on the mathematical ideas that the student understands, including the structure of the base-ten number system, number relationships, meaning of operations, and properties. Computational fluency with whole numbers is a goal of mathematics instruction in the elementary grades.  Students should be fluent with the basic number combinations for addition and subtraction to 20 by the end of grade two and those for multiplication and division by the end of grade four.   Students should be encouraged to use computational methods and tools that are appropriate for the context and purpose.

**Algebra Readiness**

The successful mastery of Algebra I is widely considered to be the gatekeeper to success in the study of upper-level mathematics. “Algebra readiness” describes the mastery of, and the ability to apply, the *Mathematics Standards of Learning*, including the Mathematical Process Goals for Students, for kindergarten through grade eight. The study of algebraic thinking begins in kindergarten and is progressively formalized prior to the study of the algebraic content found in the Algebra I Standards of Learning. Included in the progression of algebraic content is patterning, generalization of arithmetic concepts, proportional reasoning, and representing mathematical relationships using tables, symbols, and graphs. The K-8 *Mathematics Standards of Learning* form a progression of content knowledge and develop the reasoning necessary to be well-prepared for mathematics courses beyond Algebra I, including Geometry and Statistics.

**Equity**

**“**Addressing equity and access includes both ensuring that all students attain mathematics proficiency and increasing the numbers of students from all racial, ethnic, linguistic, gender, and socioeconomic groups who attain the highest levels of mathematics achievement.”
 – National Council of Teachers of Mathematics

Mathematics programs should have an expectation of equity by providing all students access to quality mathematics instruction and offerings that are responsive to and respectful of students’ prior experiences, talents, interests, and cultural perspectives. Successful mathematics programs challenge students to maximize their academic potential and provide consistent monitoring, support, and encouragement to ensure success for all. Individual students should be encouraged to choose mathematical programs of study that challenge, enhance, and extend their mathematical knowledge and future opportunities.

Student engagement is an essential component of equity in mathematics teaching and learning. Mathematics instructional strategies that require students to think critically, to reason, to develop problem-solving strategies, to communicate mathematically, and to use multiple representations engages students both mentally and physically. Student engagement increases with mathematical tasks that employ the use of relevant, applied contexts and provide an appropriate level of cognitive challenge. All students, including students with disabilities, gifted learners, and English language learners deserve high-quality mathematics instruction that addresses individual learning needs, maximizing the opportunity to learn.

| **Data and Society -** Understanding data science facilitates critical examination of questions in different parts of society and supportsinformed data-driven decision-making.**DS.1**† **The student will identify specific examples of societal problems that can be effectively addressed using data science.** |
| --- |
| **Understanding the Standard** | **Essential Knowledge and Skills** |
| * There are characteristics of problems in society that best lend themselves to be analyzed using the data cycle.
* Solutions addressed by Data Science include conjectures that can be supported or refuted by measurements or observations.
* The iterative stages of the data cycle include:
* Question/Problem Formulation - Identify the driving question for the problem being solved
* Data Acquisition & Collection - Collect and clean data to assist with multiple ways to solve a problem
* Data Processing - Manipulate data to make it usable through a predetermined process
* Data Visualization & Representation - Connect visual representations to brainstorm solutions
* Data Modeling & Analysis - Build a prototype of a model, test, and iterate
* Data Communication - Effectively communicate data driven solution based on context and audience

**Data Cycle*** The data science cycle is an iterative process.
 | **The student will use problem solving, mathematical communication, mathematical reasoning, connections, and representations to*** Identify and explain characteristics that best lend themselves to a data driven approach to problem solving.
* Formulate questions based on context.
* Understand the type of data relevant to the context of the question at hand.
* Define relationships between variables and constant relationships.
* Create a hypothesis of interest in terms of measurable data.
* Define the stages of the data cycle and how each stage is related to the other.
* Identify and explain constraints of the data-driven approach.
 |

† Standard should be included in a one-semester course in Data Science.

|  |
| --- |
| **Data and Society -** Understanding data science facilitates critical examination of questions in different parts of society and supportsinformed data-driven decision-making.**DS.2 The student will be able to formulate a top-down plan for data collection and analysis based on the context of a problem.** |
| **Understanding the Standard** | **Essential Knowledge and Skills** |
| * A data project plan ensures effective communication and agreement at all phases of the data science project.
* A data project plan allows effective execution on time and under budget.
* A data project plan allows us to understand the tools, resources and architecture needed to ensure a successful project.
* Project deliverables are the things you create to help you fulfill the objective while KPI stands for key performance indicator, a quantifiable measure of success of the project as a whole.
 | **The student will use problem solving, mathematical communication, mathematical reasoning, connections, and representations to*** Design a data project plan, which is aligned with the data science cycle, that includes the following components:
* Definition of the goal of the project as it pertains to a societal problem;
* Identification of the various stakeholders;
* A timeline for the project with deliverables;
* Key Performance Indicators (KPI) for the successful data project deliverables;
* Resource needs and tools for the project;
* Ethical considerations around the project; and
* Limitations of the project.
 |

|  |
| --- |
| **Data and Ethics -** Ethical implications may result from the types of methods used for data collection, processing, representation, analysis and use.**DS.3**† **The student will recognize the importance of data literacy in global citizenship and develop an awareness of how the analysis of data can be used to affect positive changes and mitigate negative consequences.** |
| **Understanding the Standard** | **Essential Knowledge and Skills** |
| * Data literacy is the ability to read data, work with data and communicate about data by putting it in proper context and asking relevant/clarifying questions to determine/identify bias.
* Data literacy helps to recognize, sort and filter through biases that leads to improved decision-making.
* Data privacy and consumer protection are important issues that affect individuals and organizations.
* Historical instances of government and private data breaches provide examples of the considerations of privacy in data.
 | **The student will use problem solving, mathematical communication, mathematical reasoning, connections, and representations to*** Formulate relevant/clarifying questions to identify potential biases in data presented in existing analyses/visualizations.
* Effectively read data summaries and visualizations and explain/translate into non-technical terms in proper context.
* Identify potential biases in terms of data presented and discuss the potential effects of such biases in terms of how they affect decision-making.
* Identify privacy and consumer protection issues that might be a result of how data is presented.
* Describe the types of data that business, industry, and government entities collect about people and possible ways the data is used.
 |

† Standard should be included in a one-semester course in Data Science.

|  |
| --- |
| **Data and Ethics -** Ethical implications may result from the types of methods used for data collection, processing, representation, analysis and use.**DS.4 The student will be able to identify biases in the data collection process, and understand the basic ethical implications and privacy issues surrounding data collection.** |
| **Understanding the Standard** | **Essential Knowledge and Skills** |
| * Ethical implications can result from the types of data collection methods used.
* Privacy and consumer protection are considerations when data is collected.
* Different types of biases can occur while collecting data, including implicit and explicit biases.
* There are producers, publishers, consumers and decision-makers of data.
* Producer of data: data is obtained through some source - open source, sensor equipment, third party organization/source, external source
* Publisher of data: entity that acquires, manages, stores, makes available the data
* Consumer of data: develops products/applications to support the decision-making
* Decision maker of data: uses the products/applications to make decisions
 | **The student will use problem solving, mathematical communication, mathematical reasoning, connections, and representations to*** Identify biases in the data collection process that include, but are not limited to, confirmation, selection, outliers, overfitting / underfitting, and confounding and describe mitigation strategies for these biases.
* Provide examples of biases in terms of data collection and the potential effects of such biases.
* Identify and describe biases apparent in the data given the biases of the producer as well as those of the consumer*/*decision maker of the data
* Identify potential sources of bias, given a specific data collection scenario.
* Describe how the data collection process should be focused, relevant, and limited to the scope of the data project plan.
* Describe basic ethical/privacy issues possible in the collection of data.
 |

|  |
| --- |
| **Data and Communication -** Datavisualizations are used to communicate insights about complex data sets to support an audience in making decisions. **DS.5**† **The student will use storytelling as a strategy to effectively communicate with data.** |
| **Understanding the Standard** | **Essential Knowledge and Skills** |
| * Storytelling with data involves combining context, visualizations and a narrative to communicate the idea behind a data science project effectively. Narrative, which is the crux of storytelling, is the way we simplify and make sense of complex data by supplying context, insight, interpretation to make the analysis more applicable and relevant.
* Communicating with data using storytelling involves concrete steps:
* Understanding context,
* Selecting a visual,
* Eliminating clutter,
* Focus attention, and
* Telling a story.
* Data storytelling done incorrectly can lead to incomplete or misleading information and conclusions.
 | **The student will use problem solving, mathematical communication, mathematical reasoning, connections, and representations to*** Define storytelling and explain the importance of storytelling as a strategy to communicate the idea behind and results of a data science project effectively.
* Explain the steps involved in data storytelling and how it relates to the data cycle.
* Effectively identify a story worth telling based on the data (looking for trends, correlations, outliers) and by asking a question or forming a hypothesis based on insight and audience.
* Effectively selecting visualizations that simplify the information, highlight the most important data, and communicate key points quickly.
* Effectively simplifying the information presented to make it more concise and focusing the audience's attention on the key parameters that support the student’s hypothesis.
* Effectively form a narrative based on data available to provide context, insight, interpretation to make the analysis more relevant to a given audience.
* Explain how data storytelling done incorrectly can lead to incomplete or misleading information and conclusions by eliminating/substituting, manipulating, cherry picking data, or creating inconsistent visuals or ineffective narratives.
 |

† Standard should be included in a one-semester course in Data Science.

| **Data and Communication -** Datavisualizations are used to communicate insights about complex data sets to support an audience in making decisions. **DS.6**† **The student will justify the design, use and effectiveness of different forms of data visualizations.** |
| --- |
| **Understanding the Standard** | **Essential Knowledge and Skills** |
| * The goal of data visualization is to distill large datasets into visual graphics to allow for easy understanding of complex relationships within data.
* Computer-based visualization systems provide visual representations of data sets designed to help end users to carry out tasks more effectively. Data visualization includes analysis, design, and construction.
* Task questions may include: What questions does the user want to answer? What problem is to be solved? Which decisions is the user trying to make? What outcomes are desired? What story does the user want to tell? What tasks should the user perform?
* Choosing a visualization based on data type and the message communicated reveals trends so the audience can easily understand the significance of the findings from the data set.
* Data set types in visualizations include but are not limited to: tabular; network; spatial; and textual. Tabular data may be represented in two-dimensional (row bycolumn) or multidimensional tables. Networks may include nodes and links and trees. Spatial data sets may be categorized as continuous fields as in grids of position and geometric such as in maps.
* Inputs for visualizations include data set types and tasks. Data attributes may be categorical, ordinal or quantitative with special cases for time and space.
* Data visualizations may include both conventional and emerging types based on function in the context of the data.
* Data insights from visualizations can be shared in different ways including: live or virtual presentations; dashboards; embedded into applications; and/or broadcast to audiences through data-driven alerts or communications.
* The choice of a suitable technological tool allows students to create and compare multiple visualizations of the same data set.
* Connections can be made among summary information from statistical analysis to visualizations of the same data set.
 | **The student will use problem solving, mathematical communication, mathematical reasoning, connections, and representations to*** Conduct exploratory data analysis using visualization.
* Formulate questions from exploration of a data set to consider how data will communicate a story.
* Determine the effectiveness of different data visualization choices based on the data context from conventional statistical charts to unconventional/emerging data visualizations to more complex visualizations.
* Create a visualization of a data set and summarize the representation using the context of the data.
* Compare two or more different representations to ensure the design communicates the features and behavior of data sets.
* Justify design choices (based on data set type, size, context and audience) of data visualizations to highlight important features, trends, and insights.

  |
| Chart Selection for Data Visualization by Function

|  | Comparisons | Proportions | Relationships | Hierarchy | Location | Distribution | Patterns | Range | Data Over Time | Analyzing Text | Movement/Flow | Financial | Uncertainty/Error |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Area Graph/PlotStacked Area Graph/Plot | X |  |  |  |  |  | X |  | X |  |  |  |  |
| Area Bands |  |  |  |  |  |  |  |  |  |  |  |  | X |
| Bar GraphStacked Bar Graph | X | X |  |  |  |  | X |  |  |  |  |  |  |
| Box and Whisker Plot | X |  |  |  |  | X | X | X |  |  |  |  |  |
| Bubble Chart/ Map | X | X | X |  | X | X | X |  | X |  |  |  |  |
| Candlestick Chart |  |  |  |  |  |  |  | X | X |  |  | X |  |
| Chord Diagram |  |  | X |  |  |  |  |  |  |  |  |  |  |
| Choropleth Map |  |  |  |  | X |  |  |  |  |  |  |  |  |
| Circle Packing |  | X |  | X |  |  |  |  |  |  |  |  |  |
| Confidence Strips |  |  |  |  |  |  |  |  |  |  |  |  | X |
| Connections Map |  |  | X |  |  | X |  |  |  |  | X |  |  |
| Data Over Geographical Region |  |  |  |  | X |  |  |  |  |  |  |  |  |
| Density Chart/Plot |  |  |  |  |  | X | X |  |  |  |  |  |  |
| Donut Chart |  | X |  |  |  |  |  |  |  |  |  |  |  |
| Dot Map |  |  |  |  | X | X | X |  |  |  |  |  |  |
| Dot Matrix |  | X |  |  |  | X |  |  |  |  |  |  |  |
| Error Bars |  |  |  |  |  |  |  |  |  |  |  |  | X |
| Flow Map |  |  |  |  | X | X |  |  |  |  | X |  |  |
| Gantt Chart |  |  |  |  |  |  | X | X |  |  |  |  |  |

 |

|  |
| --- |
| **Data Modeling –** Mathematical models are used to predict future, unobserved data values.**DS.7 The student will be able to assess reliability and validity of source data in preparation for mathematical modeling.** |
| **Understanding the Standard** | **Essential Knowledge and Skills** |
| * Understanding the characteristics of a reliable data source will allow for more effective analysis.
* Understanding the difference between a reliable and a valid data source compared to statistical validity and reliability in research analysisis important.
* Validity – is the data correctly formatted and stored in the right way?
* Completeness – does the dataset include values for all the fields required by your system?
* Uniqueness – is the data free from duplicates and dummy entries?
* Data validation or input validation is a method for checking the accuracy and quality of source data, typically performed prior to importing and processing so that data analysis results are accurate.
 | **The student will use problem solving, mathematical communication, mathematical reasoning, connections, and representations to*** Explain why determining the reliability and validity of big data sources is a key skill that data scientists use to build data trust across an organization.
* Assess the validity of different data sources.
* Assess processing source data for reliability based on validity, completeness and uniqueness.
 |

|  |
| --- |
| **Data Modeling –** Mathematical models are used to predict future, unobserved data values.**DS.8**† **The student will be able toacquire and prepare big data sets for modeling and analysis.** |
| **Understanding the Standard** | **Essential Knowledge and Skills** |
| * Data can be collected or acquired from reliable existing data sources.
* The purpose of sampling is to provide sufficient information so that population characteristics may be inferred.
* Data preparation helps catch errors before processing.
* Cleaning and reformatting data sets ensures that all data used in analysis will be high quality.
* Higher quality data can be processed and analyzed more quickly and efficiently.
* The process involved in preparing the data set for modeling and analysis involves one or more of the following sub-steps:
* Ingesting/wrangling the data,
* Cleaning the data,
* Formatting and Enriching the data, and
* Combining and storing the data.
 | **The student will use problem solving, mathematical communication, mathematical reasoning, connections, and representations to*** Explain the pros and cons of collecting data vs. acquiring it from existing sources.
* Explain various sampling techniques and their effectiveness - probability (simple random, systematic, stratified, and cluster sampling methods) vs. non-probability (convenience, quota, judgment, snowball).
* Read data from different sources for preparation and analysis.
* Identify important parameters about a big data set based on the context of data collected/acquired.

  |

† Standard should be included in a one-semester course in Data Science.

|  |
| --- |
| **Data Modeling –** Mathematical models are used to predict future, unobserved data values.**DS.9**† **The student will select and analyze data models to make predictions, while assessing accuracy and sources of uncertainty.** |
| **Understanding the Standard** | **Essential Knowledge and Skills** |
| * Data prediction involves extrapolating the data beyond the current data set and providing confidence values for those estimates.
* It is important to be able to distinguish between the “noise” in the data and relevant data. Every measurement is composed of true value, bias and random noise. This noise is the source of uncertainty.
* Mathematical models will be used to make data predictions based on the behavior of the data.
* Data prediction is limited by the fact that it is based on an assumption that historical patterns are a good predictor of future outcomes.
* Overfitting the data can lead to inaccurate results.
* Ethical considerations need to be taken into account during feature selection when trying to predict future outcomes.
* Students will understand the fundamentals of numerical methods, especially their application, limitations, and potentials.
 | **The student will use problem solving, mathematical communication, mathematical reasoning, connections, and representations to*** Identify factors that contribute to the overall behavior of a data set, including true values, bias and noise.
* Fit modelsbased on the behavior of the data, including models of univariate and bivariate data, in order to make predictions.
* Distinguish between linear and non-linear associations between variables using visualizations.
* Identify models that are overly complex and therefore fitting to random noise which decreases their predictive accuracy.
* Perform feature selection to choose features which are relevant to study while recognizing the potential ethical implications of removing features.
* Select the best model for a data set from among a large collection of models, using technological tools.
 |

† Standard should be included in a one-semester course in Data Science.

|  |
| --- |
| **Data Modeling –** Mathematical models are used to predict future, unobserved data values.**DS.10**† **The student will be able to summarize and interpret data represented in both conventional and emerging visualizations.** |
| **Understanding the Standard** | **Essential Knowledge and Skills** |
| * Characteristics of data sets can be summarized graphically by using visual representations of the distribution and numerically with measures of central tendency and measures of variation or dispersion.
* Descriptive statistics summarize the characteristics of a data set. Inferential statistics allow you to test a hypothesis or generalize findings from a sample group to a larger population.
* Statistical summaries lose information. Representing all the data through visualizations is important to confirm expected patterns, find unexpected patterns, and to assess the validity of the selected statistical model.
* Define emerging visualizations and describe summarization of characteristics and relationships among variables including:
* A heat map uses color to show changes and magnitude of a third variable to a two-dimensional plot.
* A bubble chart is a multivariable graph that is a cross between a scatterplot and a proportional area chart. Each plotted point then represents a third variable by the area of its circle.
* Visualizations are a key to validating underlying assumptions such as data being normally distributed and having no correlation between independent variables.
* Selected Charts for Data Visualization:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Univariate** | **Bivariate** | **Three Variables of Higher** |
| *Quantitative* | DotplotsStemplotsHistogramsBox and Whisker Plots | ScatterplotsLine Plots2-D Histograms | 3-D Scatterplot3-D LineplotHeat MapBubble Chart |
| *Categorical* | Bar ChartsPie Charts | Two-Way TablesSegmented Bar Graphs | Multivariate Bar Graphs |

 | **The student will use problem solving, mathematical communication, mathematical reasoning, connections, and representations to*** Apply descriptive statistics to explain measures of central tendency and measures of variability/dispersion to describe center and spread in visualizations of distributions.
* Define emerging visualizations and describe summarization of characteristics and relationships based on audience and purpose.
* Interpret various emerging visualizations by describing patterns, trends and relationships between and among the variables.
 |

† Standard should be included in a one-semester course in Data Science.

|  |
| --- |
| **Data Modeling –** Mathematical models are used to predict future, unobserved data values.**DS.11 The student will use hypothesis formulation and testing to extract actionable knowledge directly from data.** |
| **Understanding the Standard** | **Essential Knowledge and Skills** |
| * There are key differences between observed and theoretical probabilities.
* The different types of distribution of data vary according to the context and are important to predict future outcomes
* While causation and correlation can exist at the same time, correlation does not imply causation.
* Categorical variables can also be analyzed using specific tests.
* Unsupervised learning includes identifying meaningful clusters of data and associated sets of data points.
* It is important to have a toolbox of different statistical models for modeling a variety of phenomena (Binomial, Poisson, exponential, etc.)
* Histogram comparisons, Chi-squared tests, and other methods are used to test goodness of fit.
 | **The student will use problem solving, mathematical communication, mathematical reasoning, connections, and representations to*** Calculate the theoretical probability of random events and compare them to the observed frequencies.
* Describe the normal curve determined by the mean and standard deviation of a univariate data set.
* Fit non-linear models to data sets and use these models to predict unobserved data values.
* Select pairs of variables that identify meaningful clusters of data.
* Select an appropriate statistical distribution and test its goodness of fit including the normal distribution and other discrete and continuous distributions.
 |

|  |
| --- |
| **Data and Computing –** Technology is used to effectively prepare, analyze and communicate with data.**DS.12**† **The student will be able to select and utilize appropriate technological tools and functions within those tools to process and prepare data for analysis.** |
| **Understanding the Standard** | **Essential Knowledge and Skills** |
| * Data can be imported, processed, and exported (if necessary) using technology tools.
* Organizing data using technology tools aids in exploration.
* Technology tools can be used to handle missing entries, errors, or duplicates in the data.
* The process of decision-making that occurs during the importing or extracting, processing, cleaning and formatting of data uses a choice of tools: technological applications, coding and web.
* It is important that this data pre-processing technology process is clearly explained and documented for future replication and decision-making.
 | **The student will use problem solving, mathematical communication, mathematical reasoning, connections, and representations to*** Utilize technology tools to be able to access data effectively from multiple sources (e.g., tables, column separated values, spreadsheets, documents, databases).
* Utilize tools and functions (in tools) to effectively explore the data for issues and errors before beginning to process it.
* Define the (tools and technological) process to optimally ingest data and to export data after processing.
* Utilize tools to format and store the data appropriately to allow for effective analysis.
* Utilize tools and functions (in tools) to clean and validate data by:
* Removing data that is incomplete, incorrect or duplicated;
* Removing extraneous data or outliers; and
* Standardizing data to conform to contextual norms (e.g., privacy, sensitive data).
* Combine and store data by:
* Merging multiple data sets for efficiency purposes; and
* Optimizing the storage of data based on volume, velocity and variety.
* Define and document the process of ingesting, formatting and cleaning data for future decision making by:
* Making data more easily understood by a wider audience; and
* Connecting data with existing contextual data.
 |

 † Standard should be included in a one-semester course in Data Science.

|  |
| --- |
| **Data and Computing –** Technology is used to effectively prepare, analyze and communicate with data.**DS.13**† **The student will be able to select and utilize appropriate technological tools and functions within those tools to analyze and communicate data effectively.** |
| **Understanding the Standard** | **Essential Knowledge and Skills** |
| * Certain technological tools can be used to generate conventional and unconventional visualizations of data to explore patterns and/or analyze a large data set.
* Various technological tools have pre-built mathematical and statistical functions that allow for efficient exploration and analysis.
* Coding tools can allow for effective storage and extraction of data for more efficient analysis.
* Some technological tools have other functions that are useful to organize, summarize and gain insight from data.
* Visualization tools offer a variety of conventional and unconventional visualizations to help communicate our ideas to a wide audience.
 | **The student will use problem solving, mathematical communication, mathematical reasoning, connections, and representations to*** Select and utilize technology tools to effectively generate conventional and unconventional visualizations of data to explore patterns and/or analyze a large data set.
* Utilize specific functions in technology tools to perform descriptive and inferential statistical analysis.
* Utilize coding to store and extract data more effectively for data analysis.
* Select and apply features of technology tools effectively to organize, summarize and gain insight from data.
* Select the appropriate visualization based on context and audience and create it using technology tools to effectively communicate an idea.
 |

† Standard should be included in a one-semester course in Data Science.