Data Science Foundations Scope and Sequence 2024-2025

The course addresses the standards for Data Science Foundations that can be found on the DEW's website: [DSF Standards]

Unit 0 – Communication and Analysis (11 Days)				
Standards: SMP 1 - 8				
Days	**Lessons**	**Campaign**	**Topics**	**Essential Concepts**
1	Lesson 1: Collaboration		Group Activity: Closest to the Pin	Students will work in groups to improve their collaboration skills by completing the Closest to the Pin Activity.
2	Lesson 2: Collaboration		Group Activity: Carter's Number Generate collaboration Norms	Students will work in groups to improve their collaboration skills by completing the Carters Number Challenge.
3	Lesson 3: Launching a Lesson		Notice and Wonder	Sets the routine of launching a lesson using the "Notice and Wonder" protocol.
4	Lesson 4: Launching a Lesson		3 - Act Task	Sets the routine of launching a lesson using the "3 Act Task" protocol.
5	Lesson 5: Mathematical Mindset- Productive Struggle and Perseverance		Would You Rather?	Students will learn about what a growth mindset is and the need for perseverance in problem solving.

Department of Education & Workforce

Unit 0 – Communication and Analysis (11 Days)				
Standards: SMP 1 - 8				
Days	**Lessons**	**Campaign**	**Topics**	**Essential Concepts**
6	Lesson 6: Mathematical Mindset - Making Mistakes		Learning the value that can come from Mistakes	Students will learn that making mistakes is expected and is necessary to learn.
7	Lesson 7: Overview of the Mathematical Practices		Number/Data Talks	Students will be introduced to the practice of Number/Data talks and how they can be used to make sense of a problem.
8	Lesson 8: Communication, Reasoning, Precision		Communication Origami	Students will improve their communication skills.
9	Lesson 9: Looking For Structure and Regularity in Repeated Reasoning		Stairs Problem	Students will learn how to look for structure and regularity in problems.
10	Lesson 10: Use Tools Strategically and with Precision		Back-to-Back Sketches	Students will learn how to use various tools strategically.

Unit 0 – Communication and Analysis (11 Days)				
Standards: SMP 1 - 8				
Days	**Lessons**	**Campaign**	**Topics**	**Essential Concepts**
11	Lesson 11: Convincing, Defending, Proof		Convince Me	Students will learn how to make proper arguments to support their thinking.

Unit 1: Data and Visualization **Big Idea:** Introduces students to fundamental notions of data analysis such as distribution and multivariate associations and emphasizes creating and interpreting visualizations of real-world processes as captured by data.				
Lessons	**Campaign**	**Topics**	**Essential Concept**	**Days**
Section 1: Data are all around (7 days) *Standards: S.ID.1, 2, 5, SMP.3, 5*				
Lesson 1: Data Trails		Defining data, consumer privacy	Data are a collection of recorded observations. Data are gathered by people and by sensors. Patterns in data can reveal previously unknown patterns in our world. Data play a large, and sometimes invisible, role in our lives.	1
Lesson 2: Stick Figures		Organizing & collecting data	Data consists of records of particular characteristics of people or objects. Data can be organized in many different ways, and some ways make it easier than others for achieving particular purposes.	1
Lesson 3: Data Structures		Organizing data, rows & columns, variables	Variables record values that vary. By organizing data into rectangular format, we can easily see the characteristics of observations by reading across a row, or we can see the variability in a variable by reading down the column. Computers can easily process data when it is in rectangular format.	1
Lesson 4: The Data Cycle		Data cycle, statistical questions	A statistical investigation consists of cycling through the four stages of the Data Cycle; statistical questions are questions that address variability and are productive in that they motivate data collection, analysis, and interpretation. The Data Collection phase might	1

			consist of collecting data through Participatory Sensing or some other means, or it might consist of examining previously collected data to determine the quality of the data for answering the statistical questions. Data Analysis is almost always done on the computer and consists of creating relevant graphics and numerical summaries of the data. Data Interpretation is involved with using the analysis to answer the statistical questions.	
Lesson 5: So Many Questions		Statistical questions, variability	Statistical questions address variability.	1
Lesson 6: What Do I Eat?	Food Habits – data	Collecting data, statistical questions	After raising statistical questions, we examine and record data to see if the questions are appropriate.	1 (begin data collection)
Lesson 7: Setting the Stage	Food Habits – data	Participatory sensing	In Participatory Sensing, we humans behave as if we are robot sensors, collecting data whenever a "trigger" event occurs. Our ability to learn about the patterns in our life through these data depends on our being reliable data collectors.	1
Section 2: Visualizing Data (14 days) *S.ID.1, 3, 6, SMP.3, 5*				
Lesson 8: Tangible Plots	Food Habits – data	Dot plots, minimum/maximum, frequency	Distributions organize data for us by telling us (a) which values of a variable were observed, and (b) how many times the values were observed (their frequency).	1
Lesson 9: What is Typical?	Food Habits – data	Typical value, center	The “center” of a distribution is a deliberately vague term, but it is one way to answer the subjective question "what is a typical value?" The center could be the perceived balancing	1

			point or the value that approximately cuts the area of the distribution in half.	
Lesson 10: Making Histograms	Food Habits – data	Histograms, bin widths	Histograms can be created through the use of an algorithm. The distributions displayed in a histogram can be classified using the technical terms for the shapes of distributions. Learning to describe routine tasks through an algorithm is an important component of computational thinking.	1
Lesson 11: What Shape Are You In?	Food Habits – data	Shape, center, spread	Identifying the shape of a histogram is part of the interpret step of the Data Cycle.	1
Lesson 12: Exploring Food Habits	Food Habits – data	Single & multi-variable plots	Once Participatory Sensing data has been collected, the Dashboard and PlotApp perform the analysis step of the Data Cycle, though humans need to tell the computer which plots to examine.	1
Lesson 13: Google Sheets Basics	Food Habits – data	Intro to Google Sheets	The computer has a syntax, and it can only understand if you speak its language.	1
Lab 1A: Data, Code & Google Sheets	Food Habits – data	Google Sheets		1
Lab 1B: Get the Picture?	Food Habits – data	Variable types, bar graphs, histograms		1 (end data collection)
Lab 1C: Export, Upload, Import	Food Habits	Importing data		1
Lesson 14: Variables, Variables, Variables		Multi-variable plots	To examine whether two (or more) variables are related, we can plot their distributions on the same graph.	1
Lab 1D: Zooming Through Data		Sub-setting		1
Lab 1E: What's the Relationship?		Multi-variable plots		1

Practicum: The Data Cycle & My Food Habits and Presentations	Food Habits	Data cycle, variability		2
Section 3: Would You Look at the Time (9 days) *S.ID.5, S.IC.6, SMP.3, 5*				
Lesson 15: Americans' Time on Task	Time Use – data	Evaluating claims	Learning to examine other analyses is an important part of statistical thinking.	1 (begin data collection)
Lab 1F: A Diamond in the Rough	Time Use – data	Cleaning names, categories, and strings		1
Lesson 16: Categorical Associations	Time Use – data	Joint relative frequencies in 2-way tables	A two-way table is a summary of the association/relationship between two categorical variables. Joint relative frequencies answer questions of the form "what proportion of the people/objects had *this* value on the first variable and *this* value on the second?"	1
Lesson 17: Interpreting Two-Way Tables	Time Use – data	Marginal & conditional relative frequencies	Marginal (relative) frequencies tell us about the distribution of a single variable. Conditional relative frequencies tell us about the distribution of one variable when "subsetting" the other.	1
Lab 1G: What's the FREQ?	Time Use – data	2-way tables, tally		1 (end data collection)
Practicum: Teen Depression and Presentations	Time Use	Statistical questions, interpreting plots		2
Lab 1H: Our Time		Data cycle, synthesis		2
Unit 1 Project (5 days)				
				5

Unit 2: Distributions, Probability and Simulations				
Big Idea: Students use numerical summaries to describe distributions and introduces probability through the lens of computer simulations for informal inference.				
Lessons	**Campaign**	**Topics**	**Essential Concept**	**Days**
Section 1: What is Your True Color? (10 days) *S.ID.2, 3, S.IC.6, SMP.4,5*				
		Subsets, relative frequency	Students will understand that the 'typical' value is a value that can represent the entire group, even though we know that not all members of the group share the same value.	1
Lesson 2: What Does *Mean* Mean?	Personality Color	Measures of center – mean	The center of a distribution is the 'typical' value. One way of measuring the center is with the mean, which finds the balancing point of the distribution. The mean gives us the typical value but does not tell the whole story. We need a way to measure the variability to understand how observations might differ from the typical value.	1
Lesson 3: Median In the Middle	Personality Color	Measures of center – median	Another measure of center is the median, which can also be used to represent the typical value of a distribution. The median is preferred for skewed distributions or when there are outliers, because it better matches what we think of as 'typical.'	1
Lesson 4: How Far Is It from Typical?	Personality Color	Measures of spread – MAD	MAD measures the variability in a sample of data - the larger the value, the greater the variability. More precisely, the MAD is the typical distance of observations from the mean. There are other measures of	1

			spread as well, notably the standard deviation and the interquartile range (IQR).	
Lab 2A: All About Distributions	Personality Color	Measures of center & spread – mean, median, MAD		1
Lesson 5: Human Boxplots		Boxplots, IQR	A common statistical question is “How does this group compare to that group?” This is a hard question to answer when the groups have lots of variability. One approach is to compare the centers, spreads, and shapes of the distributions. Boxplots are a useful way of comparing distributions from different groups when all of the distributions are unimodal (one hump).	1
Lesson 6: Face Off		Comparing distributions	Writing (and saying) precise comparisons between groups in which variability is present based on the (a) center, (b) spread, (c) shape, and (d) unusual outcomes help to make statements in context of the data. Actual comparison statements should use terms such as "less than," "about the same as," etc.	1
Lesson 7: Plot Match		Comparing distributions	Boxplots are an alternative visualization of histograms or dot plots. They capture most, but not all, of the features we can see in a dot plot or histogram.	1
Lab 2B: Oh, the Summaries…	Personality Color	Boxplots, IQR, numerical summaries, custom functions		1

Practicum: The Summaries	Food Habits or Time Use	Statistical questions, comparing distributions		1
Section 2: How Likely Is It? (7 days) *S.CP.2, 9, S.IC.6, SMP.4,5*				
Lesson 8: How Likely Is It?		Probability, simulations	Probability is an area about which we humans have poor intuition. Probability measures a long-run proportion: 50% chance means the event happens 50% of the time if you repeated it forever. When we don't repeat forever, we see variability	1
Lesson 9: Bias Detective		Simulations to detect bias	In the short-term, actual outcomes of chance experiments vary from what is 'ideal.' An ideal die has equally likely outcomes. But that does not mean we will see exactly the same number of one dots, two dots, etc.	1
Lesson 10: Marbles, Marbles		Probability, with replacement	There are two ways of sampling data that model real-life sampling situations: with and without replacement. Larger samples tend to be closer to the "true" probability.	1
Lab 2C: Which Song Plays Next?		Probability of simple events, do loops, set. seed		1
Lesson 11: This AND/OR That		Compound probabilities	What does "A or B" mean versus "A and B" mean? These are compound events and two-way tables can be used to calculate probabilities for them.	1
Lab 2D: Queue It Up!		Probability with & without		1

		replacement, sample()		
Practicum: Win, Win, Win		Probability estimation through repeated simulations		1
Section 3: Are You Stressing or Chilling? (8 Days) *S.IC.2,.6, SMP.4,5*				
Lesson 12: Don't Take My Stress Away	Stress/Chill – data		Generating statistical questions is the first step in a Participatory Sensing campaign. Research and observations help create applicable campaign questions.	1 (begin data collection)
Lesson 13: The Horror Movie Shuffle	Stress/Chill – data		We can "shuffle" data based on categorical variables. The statistic we use is the difference in proportions. The distribution we form by shuffling represents what happens if chance were the only factor at play. If the actual observed difference in proportions is near the center of this shuffling distribution, then we would conclude that chance is a good explanation for the difference. But if it is extreme (in the tails or off the charts), then we should conclude that chance is NOT to blame. Sometimes, the apparent difference between groups is caused by chance.	1
Lab 2E: The Horror Movie Shuffle	Stress/Chill – data			1
Lesson 14: The Titanic Shuffle	Stress/Chill – data		We can also "shuffle" data based on numerical variables. The statistic we use is the difference in means. The distribution	1

			we form by this form of shuffling still represents what happens if chance were the only factor at play. When differences are small, we suspect that they might be due to chance. When differences are big, we suspect they might be 'real.'	
Lab 2F: The Titanic Shuffle	Stress/Chill – data			1
Lesson 13b: RStudio Basics			Students will be introduced to the statistical package R and RStudio	1
Lab 2E: Horror Movie Shuffle - R version	Stress/Chill & Personality Color		Students will repeat lab 2E using RStudio	1
Lab 2F: Titanic Shuffle - R version	Stress/Chill & Personality Color		Students will repeat lab 2F using RStudio	1
Practicum: What Stresses us out?	Stress/Chill & Personality Color		Answering statistical questions of merged data.	
What's Normal? (6 Days) *S.ID.4, S.IC.6, SMP.4,5*				
Lesson 16: What Is Normal?		Introduction to normal curve	The Normal curve, also called the Gaussian distribution and the "bell curve," is a model that describes many real-life distributions and is usually called the Normal Model.	1
Lesson 17: Normal Measure of Spread		Measures of spread - SD	The standard deviation is another measure of spread. This is commonly used by statisticians because of its role in common models and distributions, such as the Normal Model.	1
Lesson 18: What's Your Z-Score?		z-scores, shuffling	Z-scores allow us a way to measure how extreme a value is, regardless of the units of measurement. Usually, z-scores will range between -3 and +3, and so values	2

			that are at or more extreme than -3 or +3 standard deviations are considered large.	
Lab 2H: Eyeballing Normal		Normal curves overlaid on distributions & simulated data		1
Lab 2I: Normal Distribution Alphabet		Normal probability and skew		1
Unit 2 Project (4 days)				
End of Unit Project and Oral Presentations: Asking and Answering Statistical Questions of Our Own Data	Stress/Chill, Personality Color, Food Habits, or Time Use		Synthesis of above	5

Unit 3: Data Collection Methods: Traditional and Modern **Big Idea:** Prepares students to learn about the various ways of collecting data including participatory sensing, and the effect that data collection has on their interpretation of the patterns they discover.				
Lessons	**Campaign**	**Topics**	**Essential Concept**	**Days**
Section 1: Testing, Testing...1, 2, 3... (7 days) *S.IC.1, 3, 6, S.MP.1, 4, 8*				
Lesson 1: Anecdotes vs. Data		Reading articles critically, data	Data beat anecdotes. In science, we need to closely examine the quality of evidence in order to make sound conclusions. Anecdotes can contain personal bias, might be carefully selected to represent a particular point of view, and, in general, may be completely different from the general trend.	1
Lesson 2: What is an Experiment?		Experiments, causation	Science is often concerned with the question "What causes things to happen?" To answer this, controlled experiments are required. Controlled experiments have several key features: (1) there is a treatment variable and a response variable, and we wish to see if the treatment causes a change that we can measure with the response variable; (2) There is a comparison/control group; (3) Subjects are assigned randomly to treatment or control (randomized assignment); (4) Subjects are not aware of which group they are in (a 'blind'). This may require the use of a placebo for those in the control group; and (5) those who measure the response variable do not know which group the subjects were in (if both 4 and 5 are satisfied, this is a 'double blind' experiment).	1

Unit 3: Data Collection Methods: Traditional and Modern **Big Idea:** Prepares students to learn about the various ways of collecting data including participatory sensing, and the effect that data collection has on their interpretation of the patterns they discover.				
Lessons	**Campaign**	**Topics**	**Essential Concept**	**Days**
Lesson 3: Let's Try an Experiment!		Random assignments, confounding factors	Randomized assignment is required to determine cause-and-effect.	1
Lesson 4: Predictions, Predictions		Visualizations, predictions	Designing an experiment requires making many decisions, including what to measure and how to measure it.	1
Lesson 5: Time Perception Experiment		Elements of an experiment	Designing and carrying out an experiment helps us answer specific statistical questions of interest.	1
Lab 3A: The results are in!		Analyzing experiment data		1
Practicum: Music to my Ears		Design an experiment		1
Section 2: Would You Look at That? (4 days) *S.IC.1, 3, 6, S.MP.1, 4, 8*				
Lesson 6: Observational Studies		Observational study	Observational studies are those for which there is no intervention applied by researchers.	1
Lesson 7: Observational Studies vs. Experiments		Observational study, experiment	Experiments are not always possible because of various factors such as ethics, cost limitations, and feasibility.	1
Lesson 8: Monsters that Hide in Observational Studies		Observational study, confounding factors	Confounding factors/variables make it difficult to determine a cause-and-effect relation between two variables.	1
Lab 3B: Confound it all!		Confounding factors		1

Section 3: Are You Asking Me? (8 days) *S.IC.1, 3, 6, S.MP.1, 4, 8*				
Lesson 9: Survey Says…		Survey	Surveys ask simple, straightforward questions in order to collect data that can be used to answer statistical questions. Writing such questions can be hard (but fun)!	1
Lesson 10: We're So Random		Data collection, random samples	Another popular data collection method involves collecting data from a random sample of people or objects. Percentages based on random samples tend to 'center' on the population parameter value.	1
Lesson 11: The Gettysburg Address		Sampling bias	Statistics vary from sample to sample. If the typical value across many samples is equal to the population parameter, the statistic is 'unbiased.' Bias means that we tend to "miss the mark." If we don't do random sampling, we can get biased estimates.	1
Lab 3C: Random Sampling		Random sampling		1
Lesson 12: Bias in Survey Sampling		Bias, sampling methods	Another popular data collection method involves collecting data from a random sample of people or objects. Percentages based on random samples tend to 'center' on the population parameter value.	1
Lesson 13: The Confidence Game		Confidence intervals	We can estimate population parameters. This means that we can give an estimate "plus or minus" some amount that we are confident contains the true value (the population parameter).	1

Lesson 14: How Confident Are You?		Confidence intervals, margin of error	We can estimate population parameters. This means that we can give an estimate "plus or minus" some amount that we are confident contains the true value (the population parameter).	1
Practicum: Let's Build a Survey!		Non-biased survey design		1
Section 4: What's the Trigger? (5 days) *S.IC.3, 6, S.MP.1, 4, 8*				
Lesson 15 Ready, Sense, Go!		Sensors, data collection	Sensors are another data collection method. Unlike what we have seen so far, sensors do not involve humans (much). They collect data according to an algorithm.	1
Lesson 16: Does it have a Trigger?		Survey questions, sensor questions	A key feature that distinguishes the way sensors collect data from more traditional approaches is that sensors collect data when a 'trigger' event occurs. In Participatory Sensing, this event is something we humans agree upon beforehand. Every time that trigger happens, we collect data.	1
Lesson 17: Creating Our Own Participatory Sensing Campaign		Participatory sensing campaign creation	Creating a Participatory Sensing Campaign requires that survey questions must be completed whenever they are "triggered". Research questions provide an overall direction in Participatory Sensing Campaign.	1
Lesson 18: Evaluating Our Own Participatory Sensing Campaign		Statistical questions, evaluate campaign	Statistical questions guide a Participatory Sensing Campaign so that we can learn about a community or ourselves. These Campaigns should be evaluated before implementing to make sure they are reasonable and ethically sound.	1

Lesson 19: Implementing Our Own Participatory Sensing Campaign	Class Campaign—data	Mock-implement campaign, campaign creation, data collection	Practicing data collection prior to implementation allows optimization of a Participatory Sensing Campaign.	1 (begin collecting data)
Section 5: Webpages (6 days) *S.IC.3, 6, S.MP.1, 4, 8*				
Lesson 20: Online Data-ing	Class Campaign—data	Data on the internet	We stretch students' conception of data, to help them see that many web pages present information that can be turned into data.	1
Lab 3E: Scraping web data	Class Campaign—data	Scraping data from the internet		1
Lab 3F: Maps	Class Campaign—data	Making maps with data from the internet		1
Lesson 21: Learning to Love XML	Class Campaign—data	Data storage, XML	XML is a programming language that we use with our campaigns. We create basic XML "tags" in the code, which help us store data in a format we understand.	1
Lesson 22: Changing Orientation	Class Campaign—data	Converting XML files	Converting XML to spreadsheet format helps us better understand and view our data.	1 (end data collection)
Practicum: What Does Our Campaign Data Say?	Class Campaign	Statistical questions, visualizations, numerical summaries		1
Unit 3 Project (5 days)				
End of Unit Project (5 days)	Class Campaign	Simulation using experiment data		5

Unit 4: Predictions and Models **Big Idea:** Students learn to make and how to use mathematical and statistical models to predict future observations and how data scientists measure success of these predictions.				
Lessons	**Campaign**	**Topics**	**Essential Concept**	**Days**
Section 1: Predictions and Models (15 days) *S.ID.6, 7, 8, S.IC.6, SMP.2, 4, 7*				
Lesson 1: Water Usage		Data cycle, official data sets	Data can be used to make predictions. Official data sets rely on censuses or random samples and can be used to make generalizations. On the other hand, data from Participatory Sensing campaigns are not random and rely on the sensors, in our case, humans, to be gathered and limits the ability to generalize.	1
Lesson 2: Exploring Water Usage		Exploratory data analysis, campaign creation	Exploring different data sets can give us insight about the same processes. Information from an official data set compared with a Participatory Sensing data set can yield more information than one data set alone. Research questions provide an overall direction to make comparisons between data sets.	1
Lesson 3: Evaluating and Implementing a Water Campaign	Water Campaign—data	Statistical questions, evaluate & mock implement campaign	Statistical questions guide a Participatory Sensing campaign so that we can learn about a community or ourselves. These campaigns should be evaluated before implementing to make sure they are reasonable and ethically sound.	1
Lesson 4: Refining the Water Campaign	Water Campaign—data	Revise and edit campaign, data collection	Statistical questions guide a Participatory Sensing campaign so that we can learn about a community or ourselves. These campaigns should be tried before implementing to make	1 (begin data collection)

Department of Education & Workforce

Unit 4: Predictions and Models **Big Idea:** Students learn to make and how to use mathematical and statistical models to predict future observations and how data scientists measure success of these predictions.				
Lessons	**Campaign**	**Topics**	**Essential Concept**	**Days**
			sure they are collecting the data they are meant to collect and refined accordingly.	
Lesson 5: Statistical Predictions Using One Variable	Water Campaign—data	One-variable predictions using a rule	Anyone can make a prediction. But statisticians measure the success of their predictions. This lesson encourages the classroom to consider different measures of success.	1
Lesson 6: Statistical Predictions by Applying the Rule	Water Campaign—data	Predictions applying mean square deviation, mean absolute error	If we use the squared residuals rule, then the mean of our current data is the best prediction of future values. If we use the mean absolute error rule, then the median of the current data is the best prediction of future values.	1
Lesson 7: Statistical Predictions Using Two Variables	Water Campaign—data	Two-variable statistical predictions, scatterplots	When predicting values of a variable *y*, and *if y* is associated with *x*, then we can get improved predictions by using our knowledge about *x*. Basically, we “subset” the data for a given value of *x*, and use the mean *y* for those subset values. If the resulting means follow a trend, we can model this trend to generalize to as-yet unseen values of *x*.	1
LAB 4A: If the Line Fits…	Water Campaign—data	Estimate line of best fit		1
LAB 4B: What’s the Score?	Water Campaign—data	Comparing predictions to real data		1
Lesson 8: What’s the Trend?	Water Campaign—data	Trend, associations, linear model	Associations are important because they help us make better predictions; the stronger the trend, the better the prediction we can make. “Better”	1

Department of Education & Workforce

Unit 4: Predictions and Models **Big Idea:** Students learn to make and how to use mathematical and statistical models to predict future observations and how data scientists measure success of these predictions.				
Lessons	**Campaign**	**Topics**	**Essential Concept**	**Days**
			in this case means that our mean squared residuals can be made smaller.	
Lesson 9: Spaghetti Line	Water Campaign—data	Estimate line of best fit, single linear regression	We can often use a straight line to summarize a trend. "Eyeballing" a straight line to a scatterplot is one way to do this.	1
LAB 4C: Cross-Validation	Water Campaign—data	Use training and testing data for predictions		1
Lesson 10: Predicting Values	Water Campaign—data	Predictions based on linear models	The regression line can be used to make good predictions about values of *y* for any given value of *x*. This works for exactly the same reason the mean works well for one variable: the predictions will make your score on the mean squared residuals as small as possible.	1
Lesson 11: How Strong Is It?	Water Campaign—data	Correlation coefficient, strength of trend	A high absolute value for correlation means a strong linear trend. A value close to 0 means a weak linear trend.	1
LAB 4D: Interpreting Correlations	Water Campaign—data	Use correlation coefficient to determine best model		1
Section 2: Piecing it Together (6 days) S.ID.6, *SMP.2, 4, 7*				
Lesson 12: More Variables to Make Better Predictions	Water Campaign—data	Multiple linear regression	We can use scatterplots to assess which variables might lead to strong predictive models. Sometimes using several predictors in one model can produce stronger models.	1

Unit 4: Predictions and Models **Big Idea:** Students learn to make and how to use mathematical and statistical models to predict future observations and how data scientists measure success of these predictions.				
Lessons	**Campaign**	**Topics**	**Essential Concept**	**Days**
Lesson 13: Combination of Variables	Water Campaign—data	Multiple linear regression	If multiple predictors are associated with the response variable, a better predictive model will be produced, as measured by the mean absolute error.	1
LAB 4E: This Model Is Big Enough for All of Us	Water Campaign—data	Multiple linear regression		1
Practicum: Predictions	Water Campaign—data	Linear regression		1
Lesson 14: Improving Your Model	Water Campaign—data	Non-linear regression	If a linear model is fit to a non-linear trend, it will not do a good job of predicting. For this reason, we need to identify non-linear trends by looking at a scatterplot or the model needs to match the trend.	1
LAB 4F: Some Models Have Curves	Water Campaign—data	Non-linear regression		1

Section 3: Decisions, Decisions! (3 days) *S.IC.2*				
Lesson 18: Grow Your Own Decision Tree	Water Campaign—data	Multiple predictors, classifying into groups, decision trees	Many data sets have multiple predictors and are very non-linear. We can still use this data, but need to model it differently, such as in a decision tree. Decision trees are a useful tool for classifying observations into groups.	1
Lesson 19: Data Scientists or Doctors?	Water Campaign—data	Decision trees based on training and testing data	We can determine the usefulness of decision trees by comparing the number of misclassifications in each.	1
LAB 4G: Growing Trees	Water Campaign—data	Decision trees to classify observations		1
Unit 4 Project (10 days)				
Introduction to Data Dashboards		Data Dashboard	Students will be introduced to what data dashboards are and how to create them	2
Data Dashboard Creation		Data Dashboard	Students will construct a data dashboard of a data set of their choice	6
Data Dashboard Presentation		Data Dashboard	Students will present their dashboards	2