

# 05 Data Science - Water in Your Life

Content Area: **Math**  
Course(s):  
Time Period: **Full Year**  
Length: **3 weeks**  
Status: **Published**

## **General Overview, Course Description or Course Philosophy**

This course combines the study of Statistics and Probability with Data Science. The goal is to have students think critically about data in today's data-driven world and understand its role in the 21st Century economy. Furthermore, students will become familiar with the concepts, topics, and techniques used by data scientists and statisticians in their day-to-day work.

Throughout this course, students will engage in project-based observational studies and experiments to develop their understanding of data analysis, sampling, correlation/causation, bias and uncertainty, probability, modeling with data, as well as making and evaluating data-based arguments. Students will also learn about the roles of data scientists, the power of data in society, machine learning, and how data scientists extract knowledge and insights from real-world data.

In this unit, students will learn about bivariate data through discussions and data explorations around the theme of water usage. Students will explore scatter plots as a visual way to represent the relationship between two variables, draw their own lines of best fit, and learn how data scientists determine and analyze lines of best fit. Throughout the unit, students will use the analytic tools of Google Sheets, CODAP and Tableau to make and refine claims about water usage based on both self-collected data and large, publicly available data sets.

## **OBJECTIVES, ESSENTIAL QUESTIONS, ENDURING UNDERSTANDINGS**

### Essential Questions

- How can we use trends in data to make predictions?
- How do regression lines play an important role in data science?

### Enduring Understandings

- Data scientists used information to create regression lines which are used to predict

behavior

- Business models often rely on findings from data to run more efficiently

## **CONTENT AREA STANDARDS**

---

### **S.ID**

**A. Summarize, represent, and interpret data on a single count or measurement variable**

**B. Summarize, represent, and interpret data on two categorical and quantitative variables**

**C. Interpret linear models**

### **S.IC**

**A. Understand and evaluate random processes underlying statistical experiments**

**B. Make inferences and justify conclusions from sample surveys, experiments, and observational studies**

### **S.CP**

**A. Understand independence and conditional probability and use them to interpret data**

**B. Use the rules of probability to compute probabilities of compound events in a uniform probability model**

### **S.MD**

**A. Calculate expected values and use them to solve problems**

**B. Use probability to evaluate outcomes of decisions**

MA.S-ID.B.6a	Fit a function to the data (including with the use of technology); use functions fitted to data to solve problems in the context of the data.
MA.S-ID.B.6c	Fit a linear function for a scatter plot that suggests a linear association.
MA.S-ID.C.9	Distinguish between correlation and causation.
MA.A-REI.D.10	Understand that the graph of an equation in two variables is the set of all its solutions plotted in the coordinate plane, often forming a curve (which could be a line).

## **RELATED STANDARDS (Technology, 21st Century Life & Careers, ELA Companion)**

---

## Standards are Required)

---

9.1.12.RM.4: Determine when and why it may be appropriate for the government to provide insurance coverage rather than private industry. • 9.1.12.RM.5: Explain what self-insuring is and determine when it is appropriate. • 9.1.12.RM.6: Differentiate the costs benefits and features (e.g., riders, deductibles, umbrella policies) of renter's and homeowner's insurance.

LA.RI.11-12.7	Integrate and evaluate multiple sources of information presented in different media or formats (e.g., visually, quantitatively) as well as in words in order to address a question or solve a problem.
LA.11-12.SL.11-12.2	Integrate multiple sources of information presented in diverse formats and media (e.g., visually, quantitatively, orally) in order to make informed decisions and solve problems, evaluating the credibility and accuracy of each source and noting any discrepancies among the data.
CRP.K-12.CRP2	Apply appropriate academic and technical skills.
CRP.K-12.CRP4	Communicate clearly and effectively and with reason.
CRP.K-12.CRP7	Employ valid and reliable research strategies.
CRP.K-12.CRP8	Utilize critical thinking to make sense of problems and persevere in solving them.
CRP.K-12.CRP11	Use technology to enhance productivity.
TECH.8.1.12.A.4	Construct a spreadsheet workbook with multiple worksheets, rename tabs to reflect the data on the worksheet, and use mathematical or logical functions, charts and data from all worksheets to convey the results.
TECH.8.1.12.A.5	Create a report from a relational database consisting of at least two tables and describe the process, and explain the report results.
TECH.8.1.12.C.CS2	Communicate information and ideas to multiple audiences using a variety of media and formats.
TECH.8.1.12.C.CS4	Contribute to project teams to produce original works or solve problems.
TECH.8.1.12.E.CS2	Locate, organize, analyze, evaluate, synthesize, and ethically use information from a variety of sources and media.
TECH.8.1.12.E.CS4	Process data and report results.
TECH.8.1.12.F.CS3	Collect and analyze data to identify solutions and/or make informed decisions.

## STUDENT LEARNING TARGETS

---

### Declarative Knowledge

---

Students will understand that:

- Linear regression and bivariate data
- Probability can be used to analyze the fit of a regression
- There are connections between the trend and the context to make predictions
- Correlation does not imply causation.

## **Procedural Knowledge**

---

Students will be able to:

- Obtain current data and analyze trends and make predictions.
- Decide how efficient the best line of fit is for a specific linear model
- Create a least squares regression line to predictions
- Use the correlation coefficient to determine how efficient the least squares regression line is.
- Explore confounding reasons why 2 variables are related.

## **EVIDENCE OF LEARNING**

---

### **Alternate Assessments**

---

- Portfolios
- Verbal Assessment (instead of written)
- Multiple choice
- Modified Rubrics
- Performance Based Assessments

### **Formative Assessments**

---

Observations

Task completion

Student journals and notebooks

Cooperative team work

### **Summative Assessments**

---

Project based Learning

Unit assessment

### **RESOURCES (Instructional, Supplemental, Intervention Materials)**

---

<https://hsdatascience.youcubed.org/curriculum/>

Unit 3

### **INTERDISCIPLINARY CONNECTIONS**

---

Educational tech applications

Current Events

Experimentation

### **ACCOMMODATIONS & MODIFICATIONS FOR SUBGROUPS**

---

See link to Accommodations & Modifications document in course folder.