RESEARCH USING CODING
TEACHER PAGES

DESCRIPTION
This activity provides techniques for guiding students doing an independent research project. Students identify a research question, find a large data set from which to select data for analysis, and use advanced data processing techniques to answer the research question. These projects usually involve coding techniques explored in earlier activities.

STANDARDS ADDRESSED
Next Generation Science Standards

Science and Engineering Practices

1. Asking Questions and Defining Problems
2. Developing and Using Models
3. Planning and Carrying Out Investigations
4. Analyzing and Interpreting Data
5. Using Mathematics and Computational Thinking
6. Constructing Explanations and Designing Solutions
7. Engaging in Argument from Evidence
8. Obtaining, Evaluating, and Communicating Information

Crosscutting Concepts

1. Patterns.
2. Cause and effect.
3. Scale, proportion, and quantity.
4. Systems and system models.
5. Energy and matter.
6. Structure and function.
7. Stability and Change

Common Core Literacy Standards

Reading

9-12.4 Determine the meaning of symbols, key terms . . .

9-12.7 Translate quantitative or technical information . . .

Common Core Mathematics Standards

MP1. Make sense of problems and persevere in solving them.

MP2. Reason abstractly and quantitatively.

MP4. Model with mathematics.

ENDURING UNDERSTANDING
- Scientists form and refine research questions, experiments, and models using observed patterns in large data sets.

LEARNING OBJECTIVES
Students will know and be able to:
- Locate and import a large data set into a coding notebook.
- Determine a research question, or claim, that can be addressed using the large data set.
- Design a data analysis technique which will provide evidence to support the claim.

PRIOR KNOWLEDGE

Students must be able to:

- Design or adapt a coding notebook for developing the coding techniques.
- Import a large data set into a coding notebook.
- Select a data analysis technique appropriate to the research question.

BACKGROUND MATERIAL

- **Reading from a file** Open in Colab or download.
- **Math with tabular data**: Open a data file, add a new column, and fill a new column with calculated values. Open in Colab or download.
- **Descriptive statistics and a histogram**: Open a data file, view descriptive statistics (e.g., mean, median, counts), and visualize the distribution with a histogram. Open in Colab or download.
- **Error bars**: Make plots with error bars. Open in Colab or download.
- **Model fitting**: Add a trendline or curve to a set of data. Open in Colab or download.
- **Monte Carlo**: Embrace randomness to find the area under a curve without explicitly calculating an integral. Open in Colab or download.

RESOURCES.

The link below provides a sampling of large data sets that are appropriate for research. These data sets reside in the repository at University of California at Irvine:

https://archive.ics.uci.edu/ml/index.php

The link below provides help with the basics and additional links to data sets:

http://codingink12.org/

IMPLEMENTATION

Implementation starts with a discussion about available large data sets and/or the interests of individual students. A student brainstorms possible questions suitable to available data sets. Having selected a data set and research question, the student develops the coding notebook. Check in with students to help them overcome hurdles that may arise.

ASSESSMENT

Assessment is based on accuracy of results as well as claims, evidence, and reasoning. The final product can be a report, but where possible, the students should present their findings.