Project: Dataset Exploration

This lesson plan has students use CODAP. (Using another tool? Please select it now: Pyret.)

Students choose a real world dataset that is interesting to them and practice making and interpreting a range of visualizations using that dataset. *This project spans up to nine of our Data Science lessons, each of which includes an optional section with project-specific directions. We have built a Library of Datasets to support this project.*

Lesson Goals	Students will be able to… • Compute measures of center including mean, median and mode(s) of a dataset in CODAP. • Compute measures of spread of a dataset including range and IQR in CODAP. • create a variety of visualizations from a dataset of their choosing
Student-facing Lesson Goals	• Let's use a real world dataset that we find intriguing to get more practice making visualizations, computing measures of center & spread, and interpreting data. • Let's record the interesting questions we generate while exploring our dataset.
Prerequisites	• Choosing Your Dataset
Materials	• PDF of all Handouts and Pages ▼ • Data Exploration Project Slide Template • Printable Lesson Plan (a PDF of this web page)
Supplemental Materials	• Global Food Supply & Production Starter File

Preparation	• This project spans multiple lessons in the data science curriculum. Decide whether you intend to have your students engage with all or some of those lessons and modify the Data Exploration Project Slide Template accordingly.
	• Decide how much choice you're ready to offer your students about what dataset they focus on before you begin. Research shows that choice increases student engagement! But focusing the whole class on a single dataset is also an option.
	  ◦ Would focusing all of your students on a single dataset make this doable for you? Because you teach younger students who might need more scaffolding? Or because you are new to teaching data science and managing fewer moving parts would increase your confidence? We recommend the Global Food Supply and Production dataset.
	  ◦ Are you ready to jump straight into supporting your students in working on a wide range of topics of their choosing? We have a full Library of Datasets!
	  ◦ Want to give students choice from a shorter curated list…to shorten the decision-making process, focus on topics related to curricular goals, or just to have fewer options to manage during class? We've assembled descriptions of individual datasets here. For those looking for a pre-curated shorter list, we've starred a few of the datasets in our library for you.
	  ◦ **If you have time** , you may want to complete all of the lessons with everyone getting extra practice analyzing the Global Food Supply & Production dataset and then have your students choose a dataset to analyze for their culminating research papers!
	• Teachers are welcome and encouraged to edit and adapt page 1 and page 2 of the student-facing rubrics for their unique classroom context and distribute them to help students understand the scope of the project - and your expectations - at the outset.

Dataset Exploration Project

Overview

This is a multi-lesson project focused on exploring and interpreting a real world dataset.

> *Depending on how much time you have and what your learning goals are, this project may be used in preparation for a more focused analysis of the dataset and a more curated telling of the data story through our Project: Research Capstone.*

Eight of our Data Science lessons have a section at the end that support students in applying what they learn by creating and interpreting a range of categorical and quantitative data visualizations, as well as computing and analyzing various measures of spread and center. *(The scope of the project can be easily modified if you do not intend to teach all 8 of those lessons.)*

There are several options for which dataset students focus this project on:

- *Standard Option:* Students choose from our Library of Datasets, which is introduced in the Choosing Your Dataset lesson.
- *Simpler Option:* Focus your class on the Global Food Supply & Production dataset.

> *See the preparation section for more guidance on identifying which option makes the most sense for your class!*

Launch

The official launch of this project occurs during the Choosing Your Dataset lesson.

- Students copy, save, and explore the Data Exploration Project Slide Template.
  Invite them to peruse the slide deck and familiarize themselves with it; they'll be revisiting and adding to it often.
- Students complete the "About My Dataset" section of the slide deck, reflecting on the dataset's source, structure, and relevance.
- This is also a great time to distribute your adapted version of page 1 and page 2 of the student-facing rubrics so students understand the scope of the project - and your expectations - at the outset.

Investigate

Once students have a real world dataset to focus on, they will apply what they learn about each new type of display to that dataset. Visualizations will be added during the conclusion of the following Bootstrap:Data Science lessons:

- Choosing Your Dataset (since bar charts and dot plots are introduced in Bar Charts and Dot Plots, which comes before they'll have a dataset to focus on)
- Dot Plots
- Histograms: Interpreting "Shape"
- Measures of Center
- Introduction to Box Plots
- Standard Deviation
- Scatter Plots

- Correlations
- Linear Regression

If you do not plan to teach all of those lessons, be sure to modify the Data Exploration Project Slide Template and rubrics accordingly, before distributing them to students!

This project is designed as a way to record your exploration process, much like you might "show your work" when solving a word problem.

- Some visualizations will reveal interesting information and lead to new thoughtful questions.
- Some visualizations will be useless.

All visualizations you make should be added to the "Making Data Visualizations" section of your Data Exploration Project, with discussion about how they informed your thinking or a note that they didn't reveal much new information.

Recognizing whether or not a display is useful or interesting is an important step in becoming data literate!

*Encourage students to create **multiple** visualizations of each kind and to write about why they can't make more of a display if their options are limited.*

Real world datasets are not one size fits all! If your students aren't all working on the same dataset, there will be a different number of each display that students can make using the available data. (For example, some students might be able to make 10 scatter plots with their data while other students may not be able to make any!)

★ *If your students are up to the challenge, you might ask them to make ALL possible visualizations of each kind from their dataset.*

Synthesize

- Invite students to customize their slides, add graphics, and beautify their slide decks.
- Encourage students to self-assess and revise their work. Peer review is a powerful tool if time allows! The rubric (both page 1 and page 2) is a useful resource for facilitating self and peer review.
- Finally, celebrate students' work! In many instances, students will want to share their project, given how much time they have invested. Class or public presentations can instill a sense of pride.

Optional: Once students have completed this preliminary exploration of their dataset, the Project: Research Capstone is an opportunity to investigate the questions they've developed and use their data to tell a story.