

short.concord.org/a5

Using CODAP to introduce data science at the pre-college level

A workshop conducted at USCOTS, Thursday May 16, 1:00 pm – 4:30 pm

Led by [William \(Bill\) Finzer](#), Senior Scientist, [The Concord Consortium](#)

Ask questions and share resources in [Workshop Participant Resources](#)

Introductions

The Idea and the Plan

In the long term we expect that students will be learning with data in all subject areas. Over time they will become data literate, develop data habits of mind, learn data science skills, develop statistical intuition, and make appropriate use of informal and formal inference. How do we help bring that about?

This afternoon we'll be messing around with data using CODAP. We're going to get very dirty with data. This may prove quite *challenging*, but hopefully also quite *interesting*! It will be most fun if you work with a partner. In between data activities we'll discussing our ideas about how we think data science education should come about, especially in grades K–12.

Goals

1. You become a “confident” CODAP user.
2. Lots of ideas about data science education, especially pre-college, are percolating in your brain.

About the Common Online Data Analysis Platform (CODAP)

- Free, browser-based, open source
- NSF-supported
- Designed for use in grades 5–14
 - Design strongly influenced by experience with *Fathom* and *TinkerPlots*
- Classroom friendly: No login, no installation. Give students a link to CODAP documents they can work on independently.
- Used by over 20 curriculum development projects and counting
- Under continuous development

Projects (21 cases)			
Project	Organization	Discipline	Status
InquirySpace I	CC	STEM	Complete
OceanTracks	EDC	Marine Biology	Complete
IPUMS Terra	UMinn	Geography	Complete
Students Discover	NCSU	Biology, Citizen Sc...	Complete
UKDE	Physics Front	Learning Analytics	Complete
NetLogo	Northweste...	Modeling	Complete
i2Sea	Stanford	Marine Ecology	Occasional
Data Science Games	CC	CT, DSE, STEM	Ongoing
Building Models	CC	Modeling, STEM	Complete
Zoom-In Data (ZiSci)	EDC	STEM	Ongoing
ESTEEM	NCSU	Statistics	Ongoing
InquirySpace 2	CC	STEM	Ongoing
Inspect	CC	CT, DSE, STEM	Ongoing
Linking Complex Sy...	CC	STEM	Occasional
SDLC	EDC	Statistics, Social S...	Ongoing
Data Clubs	TERC	DSE	Ongoing
Data Sketch	UCB	DSE	Ongoing
Data Fundamentals	U of Melbou...	Statistics	Ongoing
Vital Signs	GMRI	Citizen Science	Ongoing
Writing Data Stories	UCB/CC	DSE	Ongoing
Data Jam	Learning Pa...	STEM	Ongoing

This material is based on work supported by the National Science Foundation under Grant Nos. IIS-1530578 and DRL-1435470. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

1—Getting Started

Purpose: To think together about what it takes to get students (and teachers) comfortable using data exploration software in general and CODAP in particular.

1. Go to codap.concord.org.
2. Press the **Try CODAP** button.
3. (Optional) Bookmark CODAP.
4. Choose OPEN DOCUMENT OR BROWSE EXAMPLES.
5. Open **Getting started with CODAP**.
6. Complete the five tasks.



Spend about 10 minutes poking around.

 **Four Seals**

 **Getting started with CODAP**

 Learn how to use CODAP using the CODAP plugin onboarding



Discussion

Keep notes in the **Workshop Participant Resources** document.

What does it take to get students (and teachers) comfortable using data exploration software in general and CODAP in particular?

2—Census Microdata and American Community Survey Data

Purpose: To explore a large dataset by taking samples from it and to think about how to make use of this kind of dataset in various classroom contexts.

[Here](#) is a shared CODAP document that contains a “plugin” that gives access to microdata back to 1860 and over 40 attributes. Spend some time looking at the plugin and some data so we can brainstorm some “quickie” questions we might investigate?

Keep track of brainstormed questions in the **Workshop Participant Resources** document.

- We pick a single question and spend ~15 minutes (possibly more) in investigation mode.
- Make a *shared link* to a view of a document that shows the results of your investigation.

Discussion

What are some ways we can make use of this kind of dataset in various classroom contexts?

How might using these data in a classroom “smell more like data science” than most intro stats activities?

3—OceanTracks and Marrying Maps with Graphs

Purpose: To get some experience with simple GIS data in CODAP. To consider how data and rich STEM contexts can be woven together.

[This document](#) has a plugin that serves as an interface to the GPS tracks of a large number of animals belonging to four different species that travel in the Pacific Ocean.

- Get familiar with the track data.
- Take a look at the [OceanTracks library](#). We will find useful information about the four species there.
- A somewhat more evolved prototype of these data is [here](#). You can read about its development in this [blog post](#).

Notes

1. You can make new attributes (with or without formulas) *in the case table*. If you do so in the *case card*, you will break the plugin!
2. A formula you may find useful is this one for distance between two points on earth:

```
greatCircleDistance(latitude, longitude, prev(latitude),  
prev(longitude))
```

Our Task

- Let's work together to characterize the tracks of different animals. A group effort! Decide what that means. Divide the work. Discuss the results.

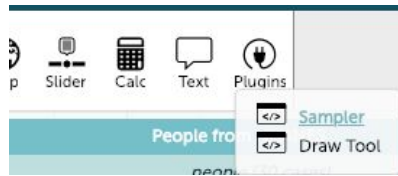
Discussion

1. These data have an incredibly rich context. Science teachers (e.g. marine biology) is often preoccupied with getting kids to understand concepts, and feel that having students do interesting data-rich investigations will get in the way. Let's talk about ways to deal with this first in the short term, and then in the long term.
2. What data science experiences did you have while working with these data that are different from working with the census data?

4—A Little Bit of Bootstrapping

Unlike *Fathom* and *TinkerPlots*, CODAP is not (yet) a very powerful simulation platform. But, thanks to Hollylynne Lee's ESTEEM project at NCSU, CODAP has a built-in plugin Sampler that can be used for simple simulations and a bit of bootstrapping.

[This CODAP document](#) contains a sample of 30 people between the ages of 20 and 30 taken from the [NHANES - National Health and Nutrition Examination Survey - CDC](#). Take a look at the limited set of attributes we're given, and think about questions about the population at large that might be answerable using this small sample. Let's brainstorm a short list of these and choose one or two as our focus.



You can add a Sampler by choosing it from **Plugins**.

A sampler has three modes: *Mixer*, *Spinner*, and *Collector*. By choosing the *Collector* mode you're on your way to setting up for bootstrapping.

Discussion

How might bootstrapping be useful in this situation. How and why does it work? How can we set it up with the *Collector*?

Action

Do it!

Share results!

Discussion

Most students (not to mention the rest of us) regard bootstrapping with suspicion. Are there ways to use the *Collector* to help demystify bootstrapping?

5—Importing and Munging Data

You've worked with documents that have plugins for accessing data and with documents that already have data. What if you're starting from scratch with a dataset that you find on the Web? Our goal here is to experience together the problem solving and thinking that goes along with preparing data and its surround for use by students. And we'll see what we think about the circumstances in and the extent to which students need to become familiar (competent?) with this process.

Here are a few sites with data:

- [Awesome Public Datasets](#)
- [Baseball](#)
- [Cities](#)
- [Bureau of Transportation Statistics](#)

You may also have brought some data or a link to some data. You may also have a context for which you think there *must* be interesting, usable data. Poke around. But not for too long. Settle in on some data you'd like to work with, at least for a short time. See how far you can get in bringing these data into CODAP and massaging it into a usable form.

[Here](#) is a CODAP document with a plugin that helps import data from HTML tables on a web page. It may prove useful to you!

Keep track of strategies you find helpful as you attempt to import and munge data. Add to the **Data Munging Strategies** list in the **Workshop Participant Resources** document. We'll use this list as an anchor for sharing what we learned.

Discussion

By some estimates 80% of what data analysts and data scientists end up doing in the world is variations on this data munging process we just been engaged in. Doing it well requires some pretty sophisticated problem solving. Doesn't that mean we should give students plenty of exposure and practice in data munging?

6—The Search for Earth-like Exoplanets

The search for planets outside our solar system that might harbor life is very much in the news. We're going to start with a dataset of about 9,500 *Objects of Interest* that emerged as of April 26, 2018. [Here](#) is a document that contains that data.

We want to end up with a list of exoplanets that are “earth-like.” Here are the constraints on such exoplanets:

1. They are confirmed exoplanets; i.e. not merely “objects of interest.”
2. The earth's sun has a temperature of 5778°K. We're looking for planets whose star has a temperature near that of the earth's sun.
3. Each star has a habitable zone. An earth-like planet should lie within its star's habitable zone.

Discuss: What percentage of the objects of interest do you think will turn out to be earth-like?

Action

Proceed to identify earth-like exoplanets. We'll share what we found and how we went about it.

Source

This task is modified from the work of EDC's [Zoom In! Teaching Science with Data](#) project. For an overview of the project see their [STEM Video Showcase entry](#).