



NOAA VIIRS
Satellite
Imagery

Commerce Data Usability Project

With tens of thousands of datasets ranging from satellite imagery to standards research to demographic surveys, the U.S. Department of Commerce has long been in the business of Open Data. But we can do more to enhance the American economy's ability to innovate using data.

Data not only needs to be disseminated in machine readable formats, but also requires context for how and what it can be used. The Commerce Data Service is launching the Usability Project, designed to illustrate the robust uses of commerce data assets, make basic code available in order to help data scientists get started on new innovative projects, and catalog areas of success and in need of improvement.

For this to work, we need your help.

We're inviting researchers and practitioners from academia, government, and industry to collaborate, demonstrate usability, and improve our data assets.

Questions? datausability@doc.gov.



NTIA National
Broadband Map

Severe
Weather
Data
Inventory

NOAA
NEXRAD

National
Economic
Accounts

American Housing Survey
+ American Community Survey

Census Foreign Trade
+ ITA Market Reports

NOAA Fishing
Landings

Commodity
Flow Survey

Commerce Data Usability Project

Get Involved

Overview

Launched from the Office of the Commerce Secretary, the Commerce Data Usability Project (CDUP) focuses on usability of data products at the Department of Commerce to promote and disseminate methods to a wider technical audience. To illustrate the utility of the data, the Commerce Data Service is working with internal partners (e.g. our technologists and scientists) and external collaborators (e.g. companies and academia) to create tutorials and user stories, focused on high-value datasets from NOAA, Census, USPTO, and BEA. Through cross-sectoral collaboration, this initiative demonstrates the promise of cross-sector data use.

What makes a good tutorial?

Each tutorial and blog entry in CDUP derive its value from the combination of the following:

- A novel analysis or question posed to the data (e.g. satellite imagery for economic inference, cross dataset mashups);
- Visually arresting graphics and engagement with the public;
- Open code and data for the public to use and replicate; and
- Materials are available through at least a free tier service, but open preferred.

What's a home run

The home run for the CDUP effort is a case in which the audience incorporates code and data from a tutorial into their workflow. If an analysis is adopted as a textbook play due to the openness of the code, methods and data, then it's a grand slam.

What are some good examples?

To show the potential of these tutorials, the Commerce Data Service has developed a few examples of tutorials using high-value datasets:

- *[Data science for nighttime lights](#)*. Correlating NOAA satellite imagery for tracking population ([link](#)).
- *[Is there enough fire in your firewall?](#)* An exploratory cybersecurity-focused analysis using NIST's National Vulnerability Database ([link](#)).
- *[Goodness gracious, great balls of hail](#)*. Basic processing of radar-derived hail events to understand long term climate risks ([link](#)).

With thousands of datasets produced across the 12 agencies within the U.S. Department of Commerce, there are many opportunities to uncover and elevate compelling use cases.

Commerce Data Usability Project

How to Contribute

Guiding questions for choosing a tutorial

When developing a project, think of the following:

- How can one craft a narrative that turn heads? This can be achieved through a combination of identifying unconventional use of data,
- Which data is accessible? Commerce has a variety of [Commerce APIs](#) for more developer-centric work as well as a list of priority datasets (link).
- What tools make sense? The Commerce Data Service is 'open by default', meaning that open source tools should be considered first, such as Python, R, JavaScript, Julia as well as free tier software.

Tutorial Structure

CDUP tutorials are commonly structured as follows:

- Show the thing. A catchy set of visualizations (at least two visualizations with one interactive) with narrative to make an interesting point. This is largely to engage non-technicals as well as give technical audience members a taste of what's possible.
- Under the hood. Expose annotated code and data endpoints so that motivated audience members can replicate the 'thing' and create new products.. Often times the methodology is documented in an iPython Notebook, RMarkdown, or other language with a clean script checked into a Github repo.

Hosting

Posts should get as many miles as possible. The proposed structure is as follows:

- The post is published to a CDUP partner's website linking back to commerce.gov/datausability index page as well as links to a Github repo containing code and materials;
- [Commerce.gov/datausability](https://commerce.gov/datausability) points to the post on the partner's website; and
- Where possible, code is checked into a Commerce Data Service Github repo for longevity purposes.

About the Commerce Data Service

The Commerce Data Service is a public startup within the Office of the Secretary, tasked with enhancing the contribution of data to the economy and American well-being. Working with the 12 Bureaus of the U.S. Department of Commerce, the Data Service rapidly creates and develops projects to advance the Department's mission. Organized as a new shared services model, the Data Service tackles core infrastructure issues around the Commerce Department and helps enrich human capital.

To contribute, contact us at datausability@doc.gov