

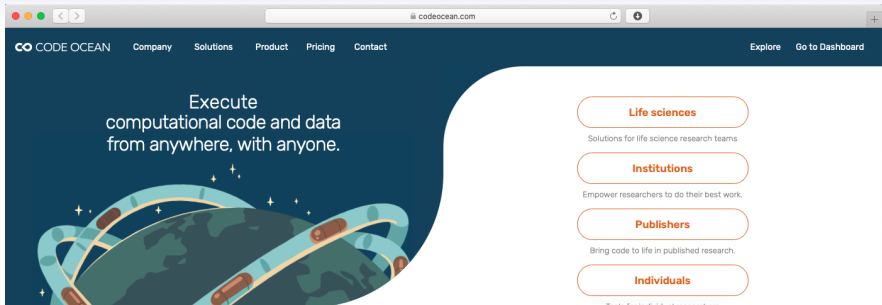
Using Cloud Computing Resources for Big Data and Code Reproducibility

—
Code Ocean

Winter Institute in Data Science and Big Data

Simon Heuberger

13 January 2020



Code Ocean transforms computational research.

We've created an integrated computational research platform for increased productivity, guaranteed computational reproducibility, and seamless collaboration. The platform captures the exact version of code, data, and development environment that generated every result. It also automates labor-intensive compute infrastructure tasks to let research teams focus on science and optimize computing costs.

[Go to Dashboard](#)

[Learn More](#)



What is a capsule?

- Code + (optional) Data + Computational environment = The minimum required for computational reproducibility
- At the core of the capsule: Docker image
- Potential to revolutionize data reproduction and replication

Safari File Edit View History Bookmarks Window Help

dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/H9X2AB

HARVARD
Dataverse

Add Data Search About User Guide Support Sign Up Log In

Files Metadata Terms Versions







Change View Table Tree

Search this dataset... Find

Filter by
File Type: All Access: All

Sort

1 to 10 of 32 Files Download

<input type="checkbox"/>	 00_run_scripts.sh Shell Script - 237 B - Nov 12, 2019 - 5 Downloads MD5: 51ae1311d1b49a03c6078931b8b4be26	Download
<input type="checkbox"/>	 01_stan_model.R scripts/ R Syntax - 6.4 KB - Nov 12, 2019 - 7 Downloads MD5: cf0b717e2712cf8c3378c4d64a960be1	Explore Download
<input type="checkbox"/>	 02_score_in_database.R scripts/ R Syntax - 4.7 KB - Nov 12, 2019 - 5 Downloads MD5: 1a2f8ac3e21842ebf5ca669b0ecodda	Explore Download
<input type="checkbox"/>	 03_county_intercept_correction.R scripts/ R Syntax - 3.0 KB - Nov 12, 2019 - 4 Downloads MD5: 274a8697195528696419331375127563	Explore Download
<input type="checkbox"/>	 04_figures_1_to_4.R scripts/ R Syntax - 8.3 KB - Nov 12, 2019 - 4 Downloads MD5: 8bf4730cf5ad85c6c0fd131a1dda43f9	Explore Download
<input type="checkbox"/>	 05_figures_5_to_7.R scripts/	Feedback

What is a capsule?

- Code?
- Data?
- Computational environment?

Computational environment

- Base environment
 - ▶ Operating system
 - ▶ Programming language
- Packages/Dependencies

codeocean.com/capsule/5033784/tree

PrivateUntitled Capsule Dec 19, 2020 14:14

Collaborate

Files

Core Files

environment178 B

code0 B

dataManage Datasets0 B

results

Other Files

Upload

or

Start with Sample Files

Environment

Environment

R (4.0.3)

R is a language and environment for statistical computing and graphics
Ubuntu 18.04 R

Change

Additional Packages

Customize the selected environment with any other packages you need. You can also use these package managers to install other package managers, such as for different languages. Packages will be installed on the next capsule run. [Learn more.](#)

Package Managers	Packages
apt-get	+ Add
Bioconductor	+ Add
R (CRAN)	+ Add
R (GitHub)	+ Add

Post-Install Script

If a package isn't available via the above package managers, use this script to download, extract and install it. Please note: this script should not be used to download data and cannot access any capsule folders. [Learn more.](#)
[Edit Post-Install Script](#)

Reproducible Run

or launch a cloud workstation

lab, DataLab, JupyterLab, RStudio, Docker

Timeline

You have 2 uncommitted changes

Describe what changed:

Added Dockerfile, metadata.yml

Commit Changes

Dec 19, 2020
Created capsule

Example: Actual capsule

Exercises

On our GitHub repository: 08-03-cloud-authors_how_to.pdf

- 1 Create a Code Ocean account at <https://codeocean.com>
- 2 Create a new capsule
- 3 Add R as the base environment
- 4 Add packages along with their specific versions
- 5 Upload data
- 6 Upload one .R file
 - ▶ The file should read in the data
 - ▶ The file should create and save one plot with ggplot2
- 7 Create a run script for the uploaded .R file
- 8 Upload a very rudimentary Readme file
- 9 Commit the changes
- 10 Execute one successful reproducible run

Extra exercises

- Add Python 3 to your R base environment via `apt-get`
- Include a full log of the files in `run`
- Save your plot to the subfolder `/figures`
- Create another new capsule by importing your Git repository