Intro to Docker Containers for Reproducible Research in Computational Science

June 29, 2020

Paul Stey,
Bradford Roarr,
Isabel Restrepo

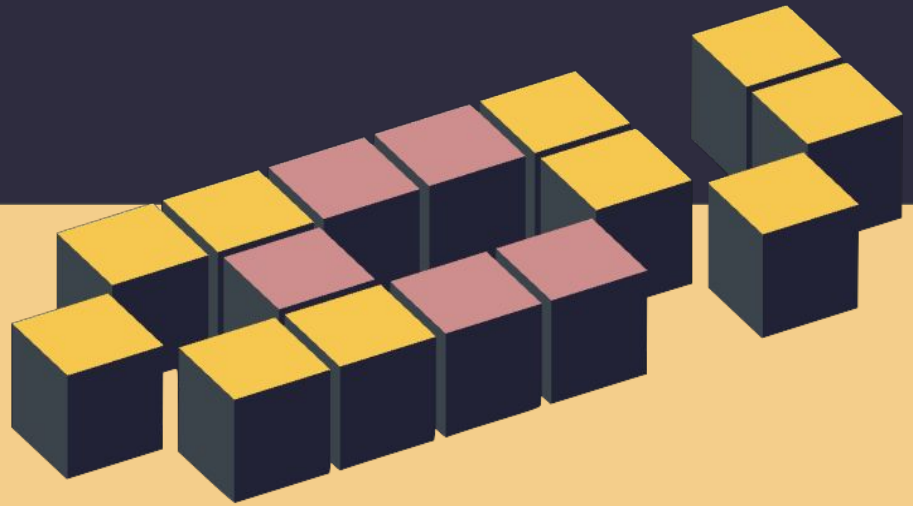
BROWN

BROWN
Center for
Computation &
Visualization

ccv.brown.edu
ccv@brown.edu

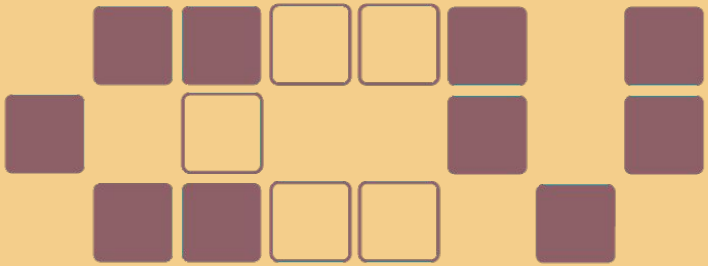
Outline

1. About CCV
2. Crisis of reproducibility
3. Improving reproducibility
4. Containers 101
    a. What is Docker?
5. Example
    a. “Hello, world!”
    b. Fibonacci
    c. Interactive
    d. Using host GPU



ccv.brown.edu
ccv@brown.edu

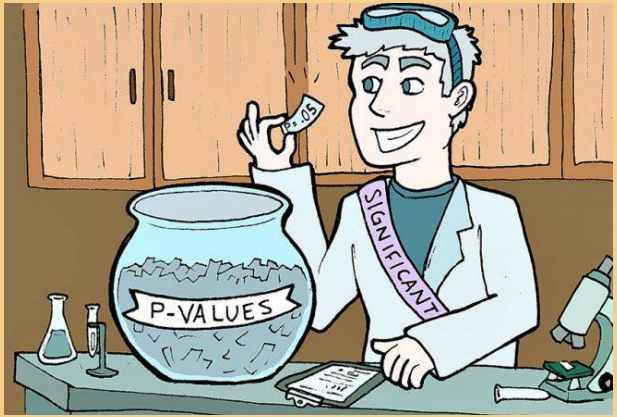
ccv.brown.edu



BROWN
Center for
Computation &
Visualization

Crisis of Reproducibility

1. Ioannadis (2005)
   a. "*Why Most Published Research Findings Are False*"
      i. "*p*-hacking"
      ii. Hypothesizing-after-the-fact
      iii. Small studies with low power
2. Misaligned Incentives in Academia
   a. "*Publish or perish!*"
   b. Priority of "Oh, wow!" findings
   c. Dis-incentives for replication studies



BROWN

BROWN
Center for
Computation &
Visualization

How do we Improve Reproducibility?

1. Use free, open-source software
2. Make code public
3. Share data (F.A.I.R.)
    a. Findable, Accessible, Interoperable, Reusable
4. Share entire compute environment
    a. Code, 3rd-party libraries, runtime env
        i. Use containers!!

julia

BROWN

BROWN
Center for
Computation &
Visualization

What is a Container?

1. Container is a virtualized environments
2. Similar to VMs (i.e., virtual machines)
    a. Containers have less performance overhead
3. Container Software
    a. Docker
    b. Singularity
    c. Linux-VServer

docker

BROWN

BROWN
Center for
Computation &
Visualization

Docker

1. Container engine for Windows, macOS, and Linux
2. Easy-to-use
3. Can use host's GPUs
4. Free (as in beer and speech)!!!

BROWN

BROWN
Center for
Computation &
Visualization

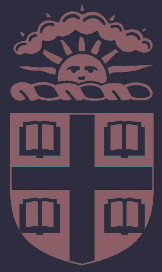
Live Demo

Code available here:

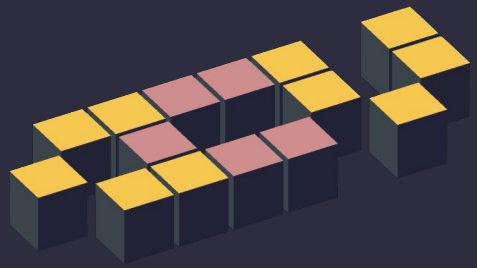
https://github.com/brown-ccv/docker-reprod-research

Any questions?

¯_(ツ)_/¯



BROWN



Thank you!!!

ccv.brown.edu