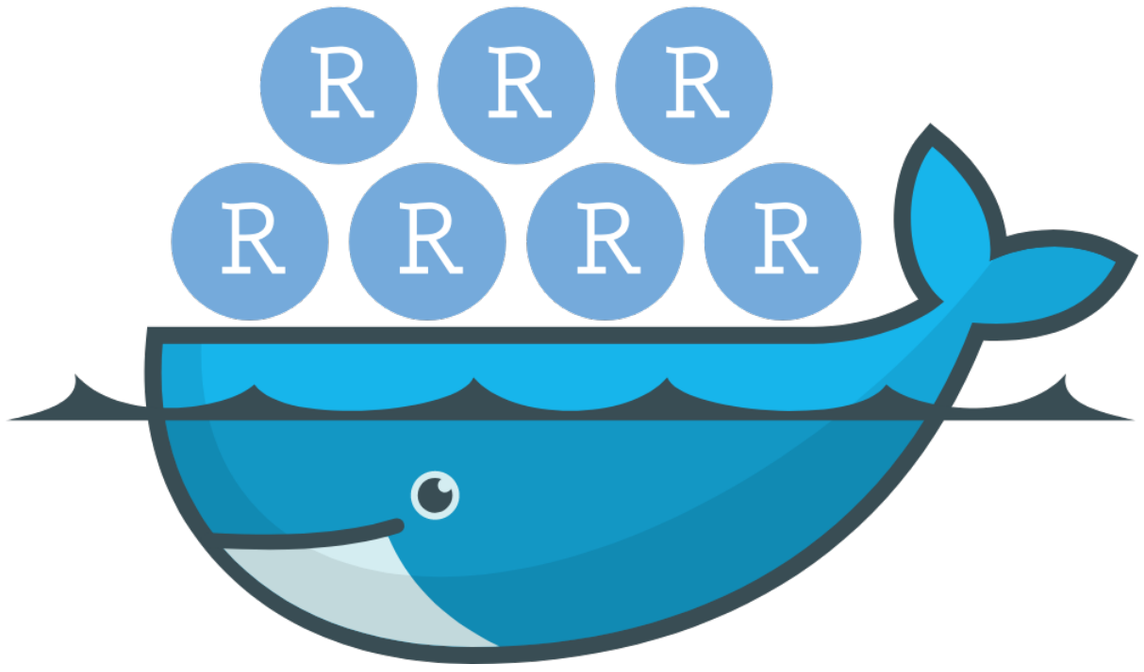
Using Rocker containers and CI for teaching R-based courses

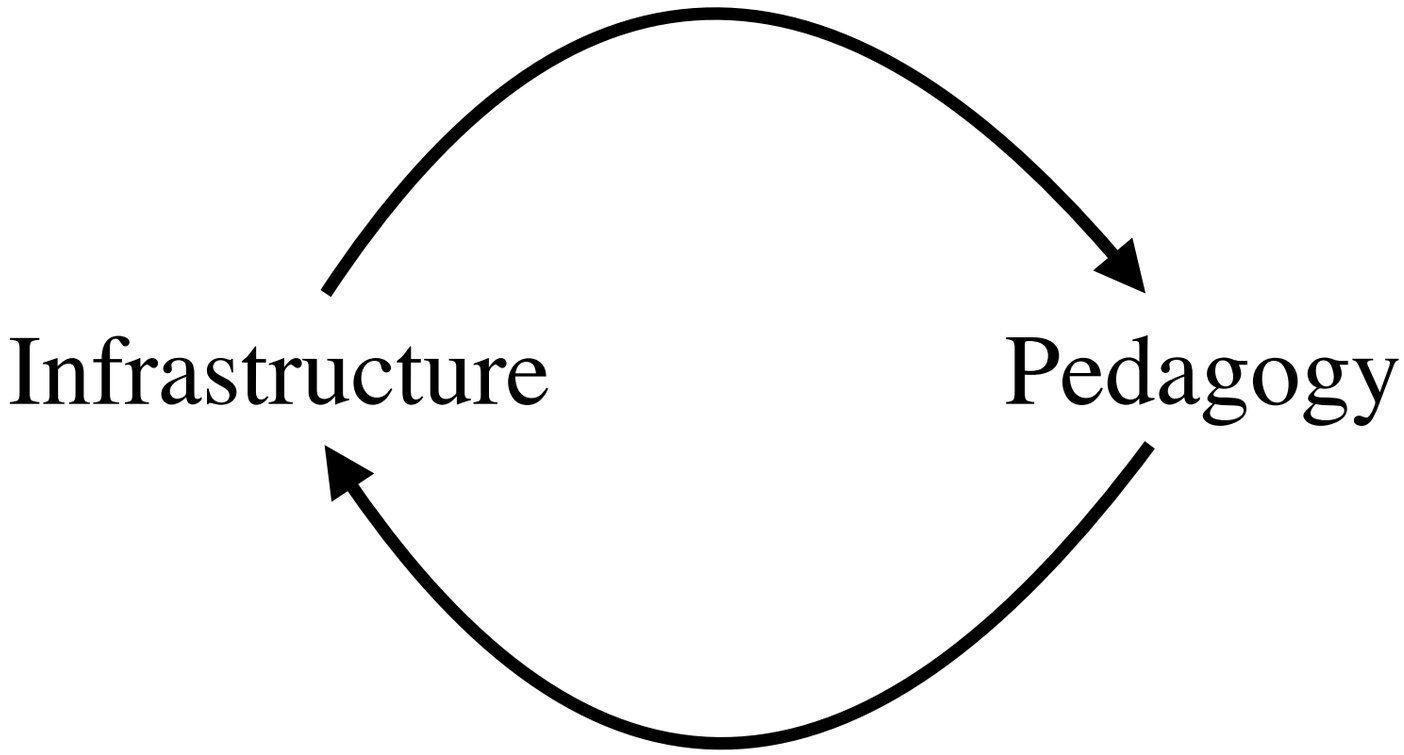


Colin Rundel
University of Edinburgh / Duke University

SDSS 2019

bit.ly/sdss_rocker

(Computational) Infrastructure Matters



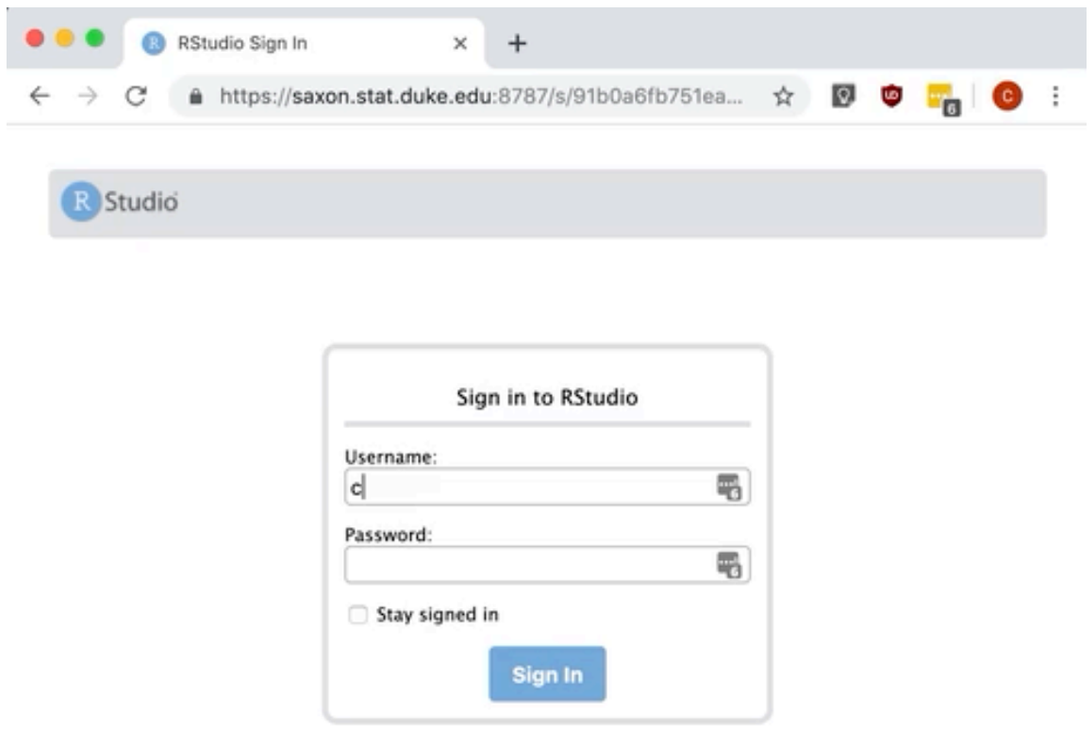
GUI vs. Programming?

Computer Lab vs. Student Laptops?

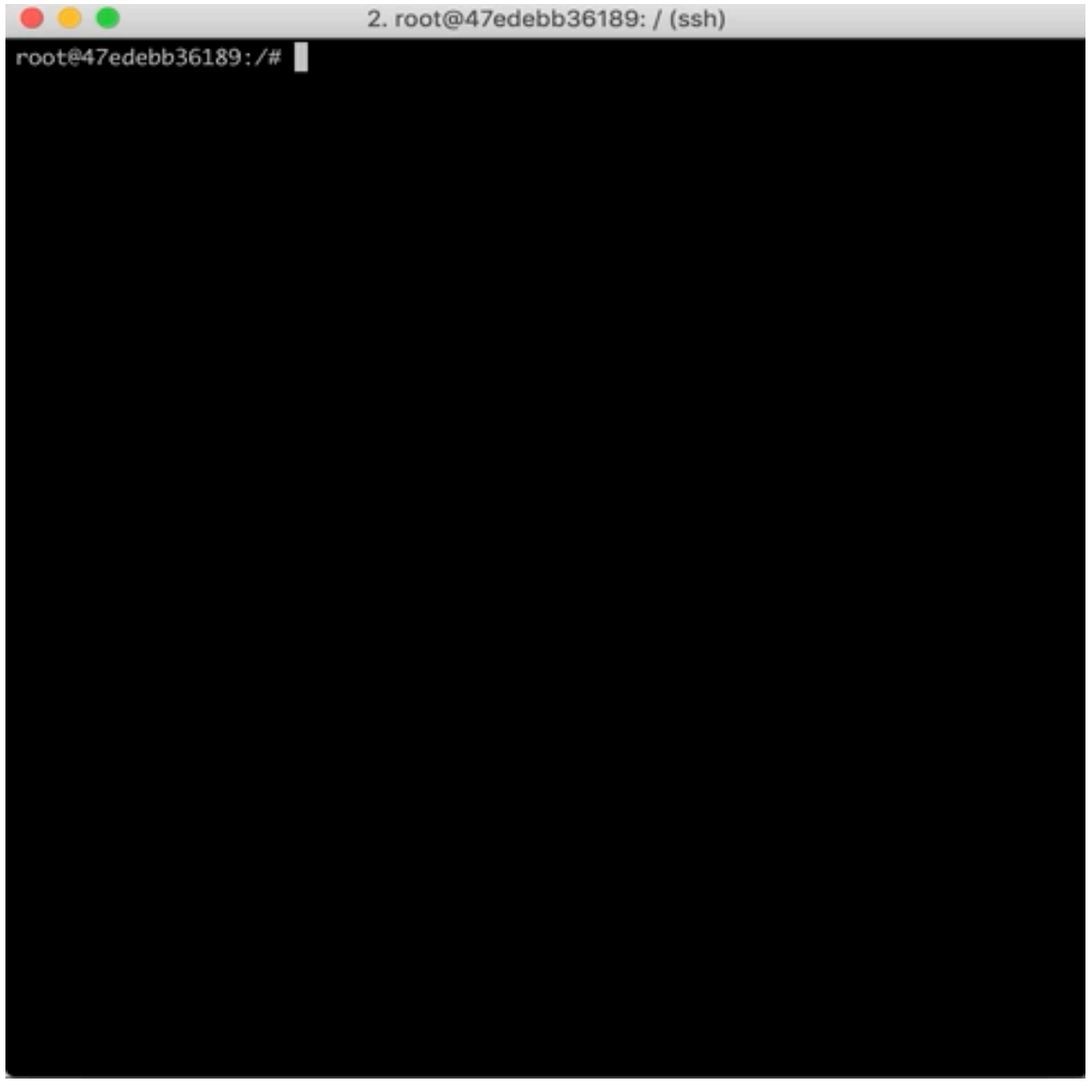
Local vs. Cloud Servers?

Stat 101 vs Thesis Student?

Day 1, hour 0



VS

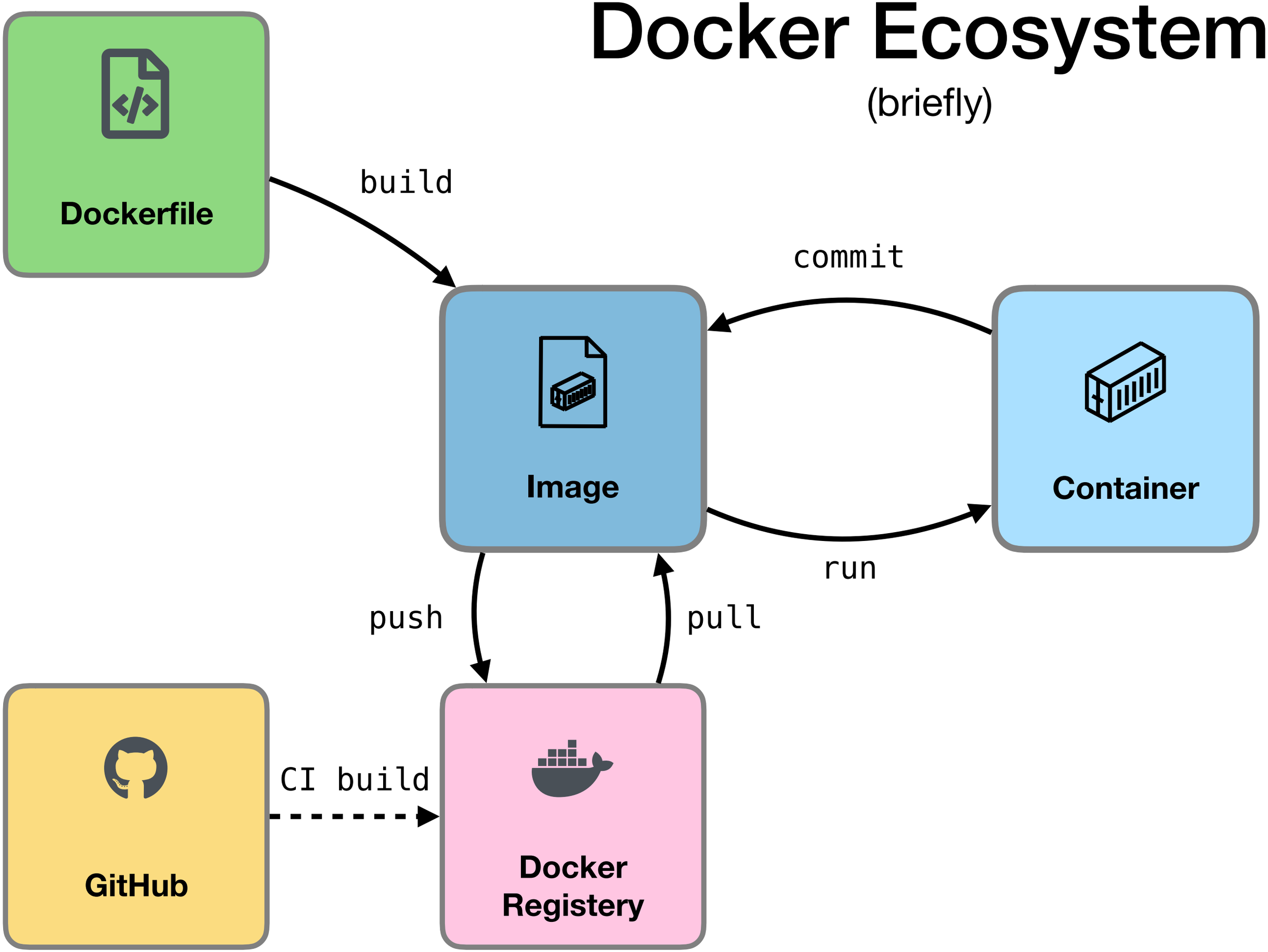


Containers?

GROSS
TARE
KG
LB
KG
LB
PAYLOAD
CU.CAP.
KG
LB
CU.M
CU.FT
CSV
+ metadata

Docker Ecosystem

(briefly)



Rocker Project

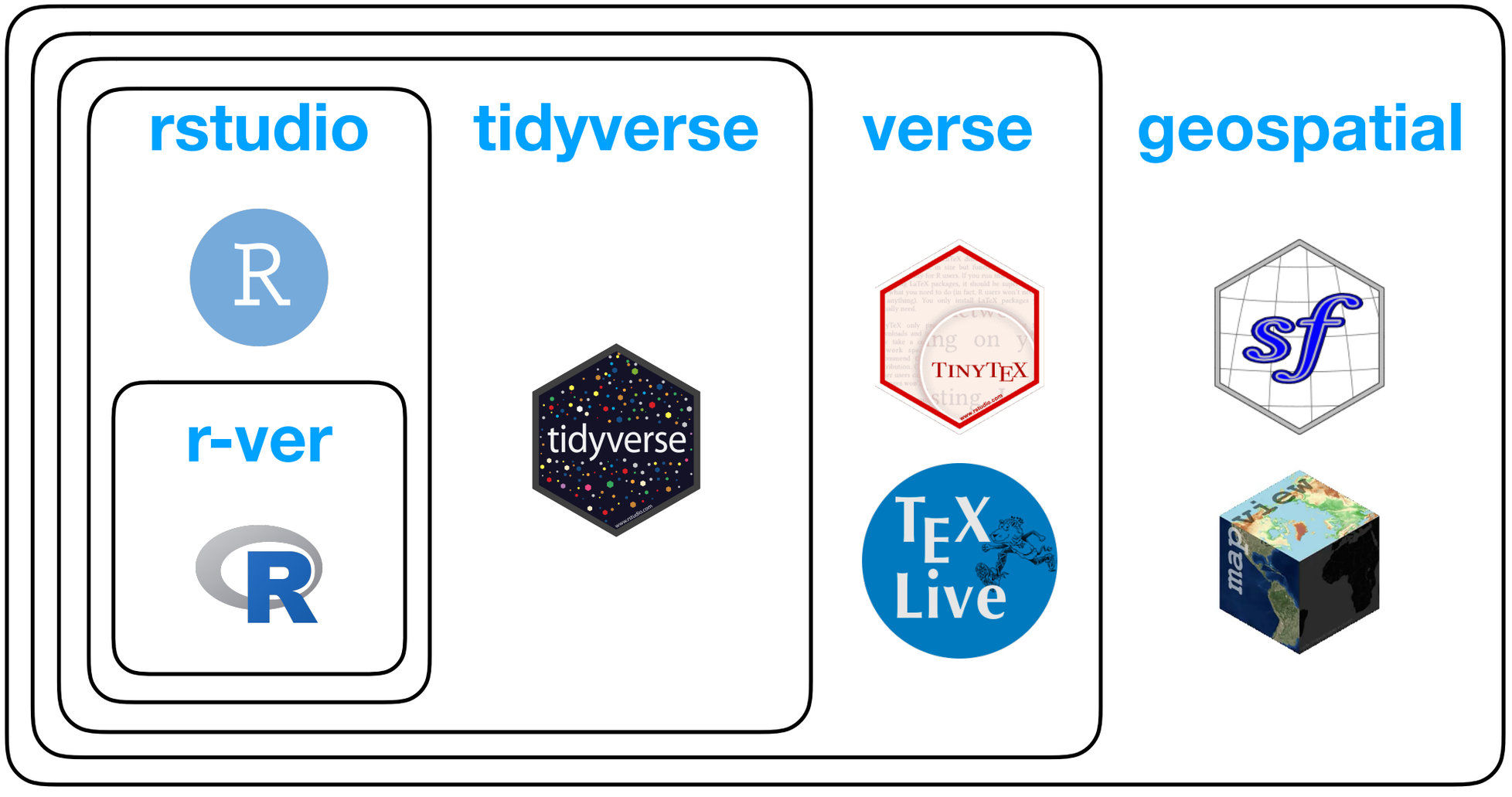
Docker Containers for the R Environment

Carl Boettiger
UC Berkeley

Dirk Eddelbuettel
Consolidated Trading LLC / UIUC

GitHub.com/rocker-org

rocker-org/**rocker-versioned**



rocker/<image name>:<version tag>

Supported Versions: 3.1.0 through 3.6.0 (latest)

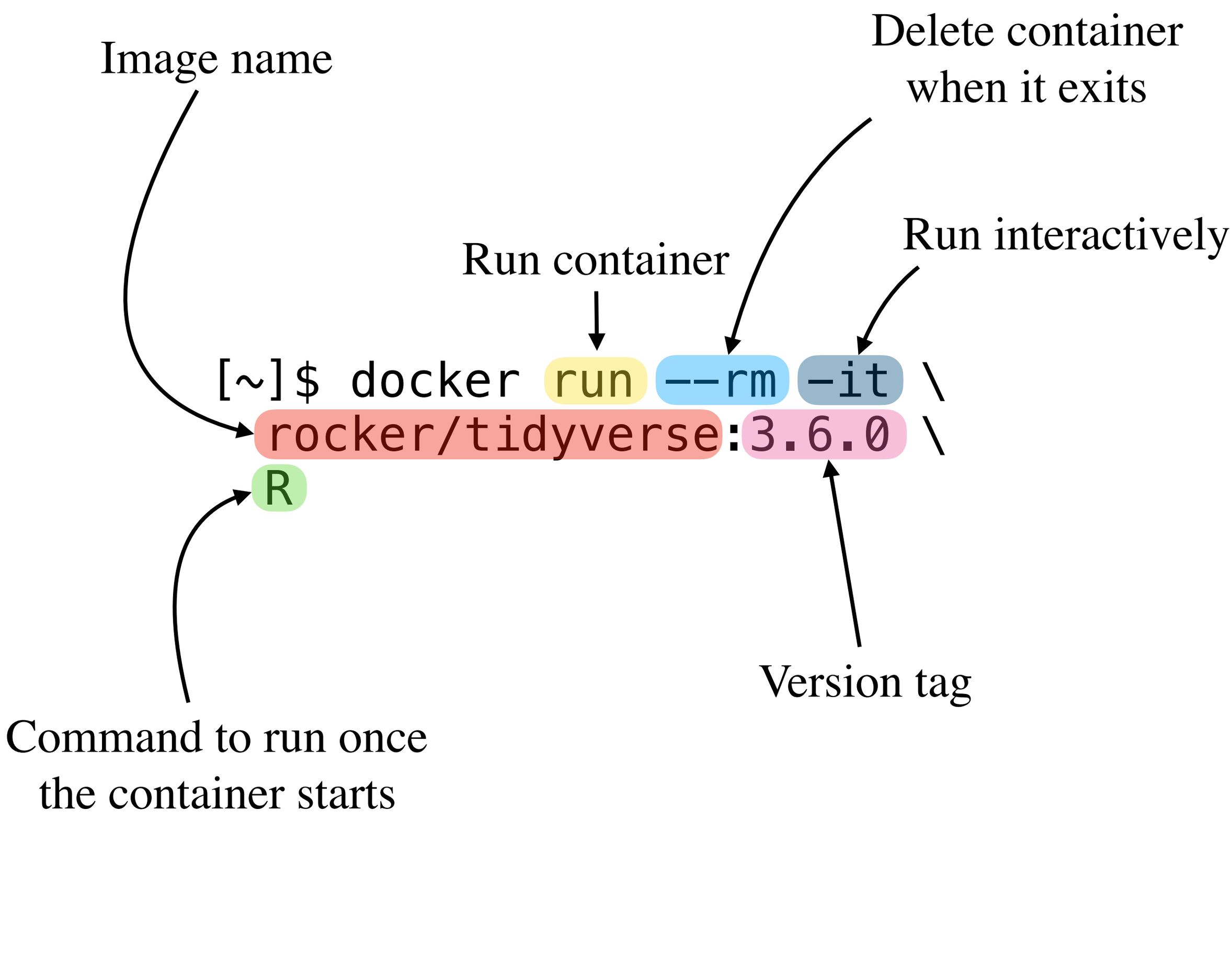


Image name
Delete container
when it exits
Run container
Run interactively
[~]$ docker run --rm -it \
rocker/tidyverse:3.6.0 \
R
Command to run once
the container starts
Version tag

```
[~]$ docker run --rm -it   \
    rocker/tidyverse:3.6.0 \
    R

R version 3.6.0 (2019-04-26) -- "Planting of a Tree"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)
…

> library(dplyr)
…
```

```
> installed.packages() %>%
    as_tibble() %>%
    select(Package, Version) %>%
    filter(Package %in% c("Rcpp","dplyr"))

# A tibble: 2 x 2
  Package Version
  <chr>   <chr>
1 dplyr   0.8.1
2 Rcpp    1.0.1
```

```
[~]$ docker run --rm -it   \
    rocker/tidyverse:3.4.3 \
    R

R version 3.4.3 (2017-11-30) -- "Kite-Eating Tree"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)
…

> library(dplyr)
…
```

```
> installed.packages() %>%
    as_tibble() %>%
    select(Package, Version) %>%
    filter(Package %in% c("Rcpp","dplyr"))

# A tibble: 2 x 2
  Package Version
  <chr>   <chr>
1 dplyr   0.7.4
2 Rcpp    0.12.15
```

Why Docker / Rocker?

1. Portability
2. Reproducibility
3. Scalability

Portability

Set a default password for RStudio login via environment variable

Delete container when it exits

Docker Command

```
[~]$ docker run --rm    \
  -e PASSWORD=testtest  \
  -p 8787:8787          \
  rocker/rstudio
```

Publish container port 8787 to host port 8787

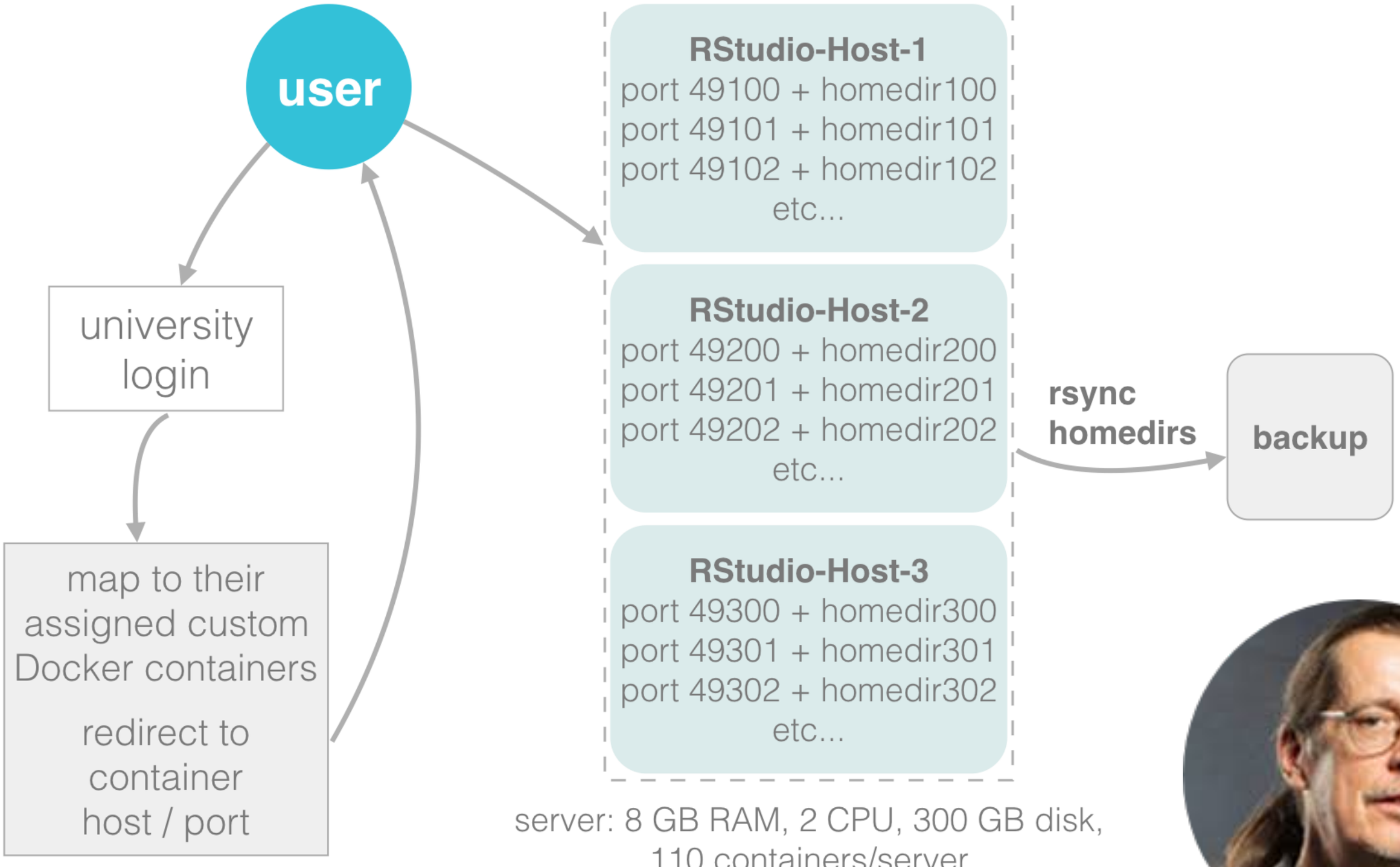
Image Name

```
[~]$ docker run --rm   \
  -e PASSWORD=testtest \
  -p 8787:8787         \
  rocker/rstudio

Unable to find image 'rocker/rstudio:latest' locally
latest: Pulling from rocker/rstudio
c5e155d5a1d1: Already exists
c60a10fed131: Already exists
fb10ae70f367: Already exists
392677a5d52c: Already exists
cbdd36353578: Already exists
f07af4782b4d: Already exists
6398787194a2: Already exists
Digest: sha256:d6caabec752b0e594dedbdea6433b13ff4a4e1208141c53a1b32a
Status: Downloaded newer image for rocker/rstudio:latest


[s6-init] making user provided files available at /var/run/s6/
etc...exited 0.
 …
[services.d] done.

[~]$ open http://localhost:8787
```



github.com/mccahill/docker-rstudio/

Mark McCahill
Duke University
IT System Architect

VM Manage

https://vm-manage.oit.duke.edu

Duke UNIVERSITY | VM Manage

Home Docker Logout

About VM-Manage

VM-Manage is a reservation service for compute environments such as RStudio, Eclipse, Jupyter, Spyder+Anaconda Python and other course-oriented Docker containers.

For virtual machine reservations visit the new **Virtual Computing Manager** service at: https://vcm.duke.edu

Docker containers

Students and instructors can reserve personal computer environments running applications such as **RStudio**, the **Eclipse IDE**, **Jupyter Notebooks**, **Matlab**, **Octave**, **Mathematica**, and others for a semester.

Click here to access your personal Docker containers.

Duke UNIVERSITY © 2012-2016 Duke University, Durham, NC 27708 | (919) 684-8111

Accessibility | Style Guide | About Duke.edu | Contact Us

ASA DATA FEST 2019
@ DUKE

436 participants

94 teams

8 schools

1 surprise dataset

Reproducibility

Server/Container

RStudio

Rmd

html

RStudio Server

push

pull

GitHub

class/hw-team

1 organization / class

1 repo / team / assignment

rundel/ghclass

hook

badge

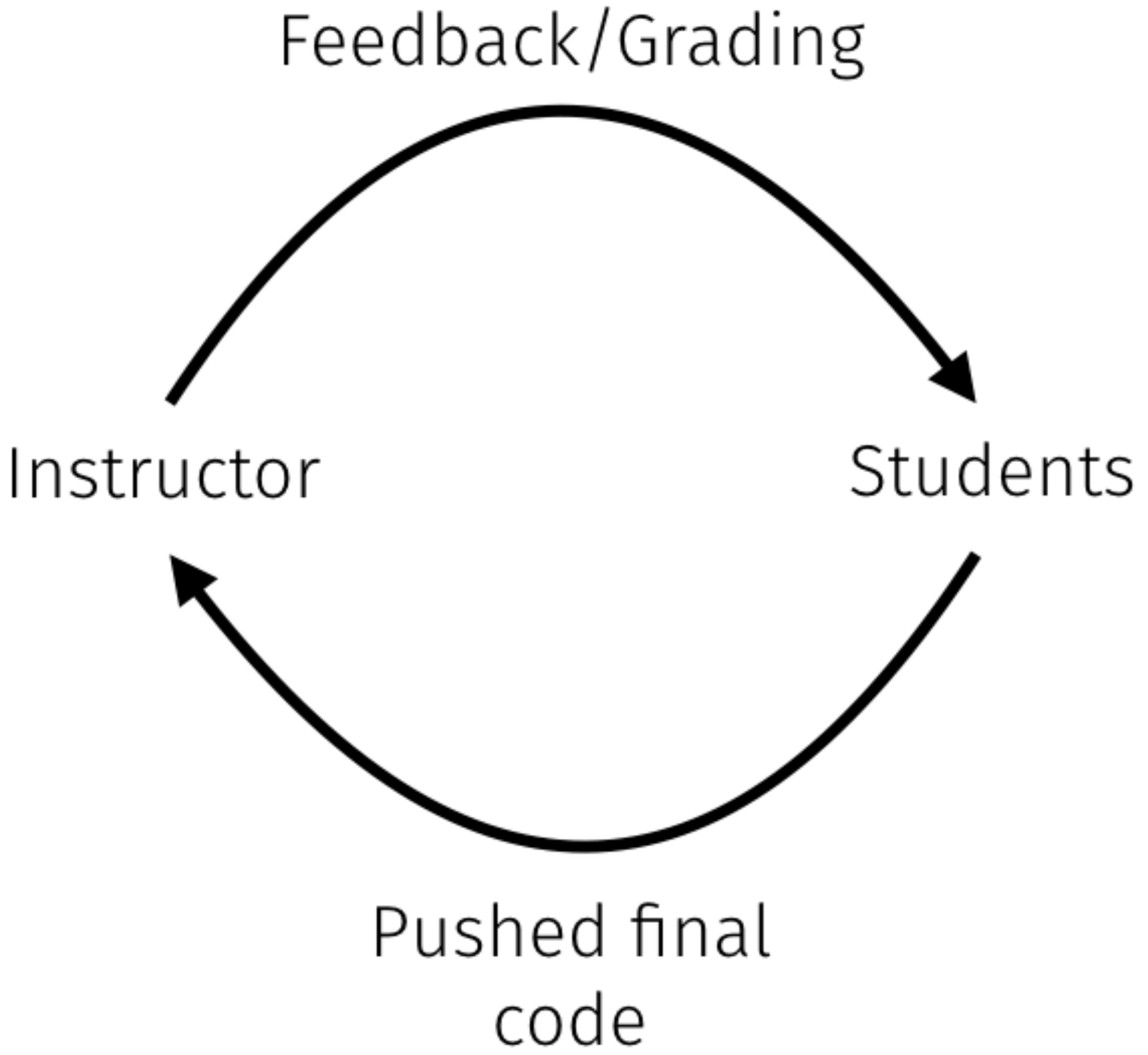
Wercker

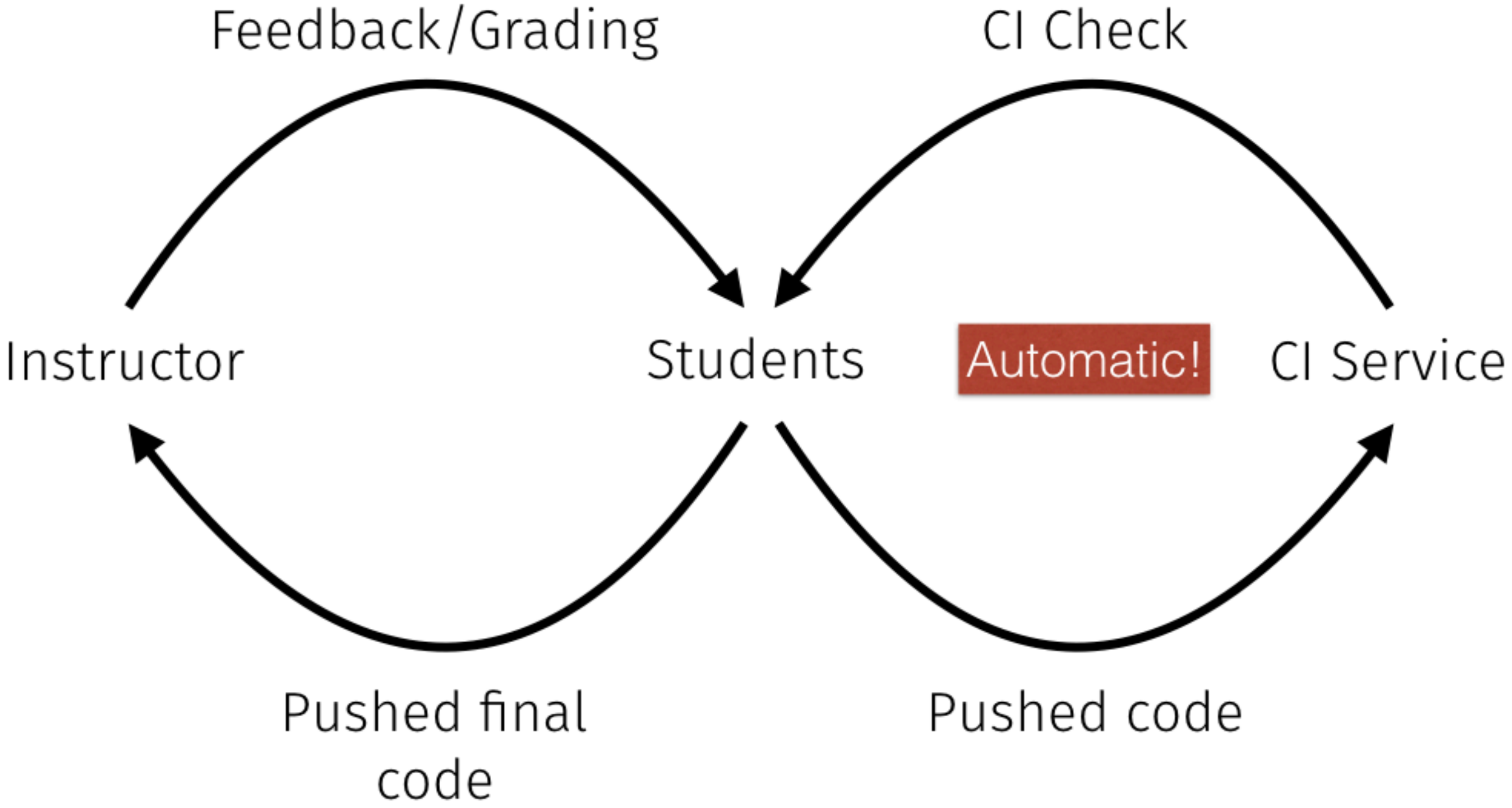
class/hw-team

1 organization / class

1 application / repo

rundel/wercker


Feedback/Grading
Instructor
Students
Pushed final code


Feedback/Grading
CI Check
Instructor
Students
Automatic!
CI Service
Pushed final code
Pushed code

```
# wercker.yml

box: rocker/verse

build:
  steps:
    - script:
        name: Check for allowed files
        code: |
          Rscript -e "source('hw4_whitelist.R')"
- script:
    name: Render R markdown
    code: |
      Rscript -e "library(rmarkdown);render('hw4.Rmd')"
```

Sta323-Sp19/hw4-team00

GitHub, Inc. [US] | https://github.com/Sta323-Sp19/h...

7 commits | 1 branch | 0 releases

Branch: master | New pull request | Create new file | Upload f

rundel Update hw4.Rmd

data	first commit
.gitignore	fix
README.md	Added wercker badge
hw4.Rmd	Update hw4.Rmd
hw4.Rproj	first commit
hw4_whitelist.R	first commit
wercker.yml	fix

README.md

build failed

Sta 323 - Homework 4 - Tidy Tuesday & TV's Golden Age

Due on Friday, February 22nd by 11:59 pm.

- Rundel, Colin - cr173@duke.edu

Sta323-Sp19/hw4-team00 | Sta323-Sp19/hw4-team00 - pi

https://app.wercker.com/Sta323-Sp19/hw4-team00/ru...

✗ Update hw4.Rmd

✗ build

Details

rundel | master | #ebf8471

Steps

✓ get code — 0 seconds

✓ setup environment — 1 minute, 23 seconds

✓ wercker-init — 0 seconds

✓ Check for allowed files — 1 second

✗ Render R markdown — 1 second

Command cancelled due to error

Sta323-Sp19/hw4-team00 - pi

https://app.wercker.com/Sta323-Sp19/hw4-team00/runs/build/5cf01eb687d3180008e21fb1

ORACLE + wercker Pipelines

Sta323-Sp19 / **hw4-team00**

Runs | Workflows | Access | Environment | Options

✗ Update hw4.Rmd master build Retry

✗ build *Render R markdown failed*

Steps

✓	get code	0 seconds
✓	setup environment	1 minute, 23 seconds
✓	wercker-init	0 seconds
✓	Check for allowed files	1 second
✗	Render R markdown	1 second

Command cancelled due to error

Sta323-Sp19/hw4-team00

GitHub, Inc. [US] | https://github.com/Sta323-Sp19/h...

6 commits | 1 branch | 0 releases

Branch: master | New pull request | Create new file | Upload f

rundel Added wercker badge

data	first commit
.gitignore	fix
README.md	Added wercker badge
hw4.Rmd	fix
hw4.Rproj	first commit
hw4_whitelist.R	first commit
wercker.yml	fix

README.md

build passing

Sta 323 - Homework 4 - Tidy Tuesday & TV's Golden Age

Due on Friday, February 22nd by 11:59 pm.

- Rundel, Colin - cr173@duke.edu

Sta323-Sp19/hw4-team00 - pi

https://app.wercker.com/Sta323-Sp19/hw4-team00/ru...

Sta323-Sp19 / **hw4-team00**

Runs | Workflows | Access | Env | Options

✓ Added wercker badge

✓ build

Details

rundel master #5ef3b76

Steps

✓ get code	0 seconds
✓ setup environment	1 minute, 25 seconds
✓ wercker-init	0 seconds
✓ Check for allowed files	1 second

Scalability

```
[~]$ gcloud compute firewall-rules create allow-rstudio  \
      --allow tcp:8787 --target-tags rstudio-server

Creating firewall...⠹Created [https://www.googleapis.com/compute/v1/projects/sta79
Creating firewall...done.
NAME           NETWORK  DIRECTION  PRIORITY  ALLOW     DENY  DISABLED
allow-rstudio  default  INGRESS    1000      tcp:8787        False

[~]$ gcloud compute instances create-with-container test \
      --machine-type n1-highcpu-16                       \
      --container-image rocker/tidyverse                 \
      --container-env PASSWORD=testtest                  \
      --tags rstudio-server

Created [https://www.googleapis.com/compute/v1/projects/sta790-02-spring-2019/zone
NAME  ZONE        MACHINE_TYPE   PREEMPTIBLE  INTERNAL_IP  EXTERNAL_IP    STATUS
test  us-east4-b  n1-highcpu-16               10.150.0.13  35.188.243.26  RUNNING

[~]$ gcloud compute instances delete test -q
```

Singularity Project

https://www.sylabs.io/singularity/

```
[~]$ singularity build R.simg docker://rocker/r-ver:3.5.3

Docker image path: index.docker.io/rocker/r-ver:3.5.3
Cache folder set to /home/rundel/.singularity/docker
[2/2] |===================================| 100.0%
Importing: base Singularity environment
Exploding layer: sha256:c5e155d5a1d130a7f8a3e24cee0d9e1349bff13f
Exploding layer: sha256:2524371be50d9c4849d98602793193ce9979872d
Exploding layer: sha256:5657cc96edf66b119f01f5df61f189f383e1a96f
Building Singularity image...
Singularity container built: R.simg
Cleaning up...

[~]$ ./R.simg

R version 3.5.3 (2019-03-11) -- "Great Truth"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)
…

>
```

```
[~]$ scp R.simg cr173@dcc-slogin.oit.duke.edu:~/
R.simg                                        100%  229MB  12.7MB/s   00:18
[~]$ ssh cr173@dcc-slogin.oit.duke.edu
cr173@dcc-slogin-02  ~ $ ./R.simg

R version 3.5.3 (2019-03-11) -- "Great Truth"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)
…

>
```

Questions / Comments?

rundel@gmail.com

@rundel

rundel

bit.ly/sdss_rocker

Çetinkaya-Rundel, Rundel (2018) Infrastructure and Tools for Teaching Computing Throughout the Statistical Curriculum, *The American Statistician*, 72:1, 58-65

bit.ly/tas_infrastructure

```
# Dockerfile: rocker/tidyverse

FROM rocker/rstudio:3.6.0

RUN apt-get update -qq \
 && apt-get -y --no-install-recommends install \
        libxml2-dev \
        libcairo2-dev \
        libsqlite3-dev \
        libmariadbd-dev \
        libmariadb-client-lgpl-dev \
        libpq-dev \
        libssh2-1-dev \
        unixodbc-dev \
        libsasl2-dev
&& install2.r --error --deps TRUE \
        tidyverse \
        dplyr \
        devtools \
        formatR \
        remotes \
        selectr \
        caTools \
        BiocManager
```