

# Technology lowering barriers: get started with R at the snap of a finger

slides & demo materials at

<https://github.com/mine-cetinkaya-rundel/2016-05-17-tech-lower-barriers-ecots>

mine çetinkaya-rundel  
duke university

nick horton  
amherst college

[mine@stat.duke.edu](mailto:mine@stat.duke.edu) / [nhorton@amherst.edu](mailto:nhorton@amherst.edu) 

@minebocek / @askdrstats 

mine-cetinkaya-rundel / nicholasjhorton 

# demo - later in the session

1

Go to <http://tinyurl.com/ecots-rstudio> and claim a container:

- Click on the URL
- Use guest as username, and password provided
- Mark the container you use with X to denote that it has been claimed

2

Download the sample template by typing the following in the console:

```
download.file("http://tinyurl.com/ecots-sample", destfile = "ecots_sample.Rmd")
```

# roadmap

assuming you  
want to teach with  
R & want to know  
how to get started

more on  
why R?

more on  
how R?

## **Less Volume, More Creativity – Getting Started Teaching with R**

with Randall Pruim and Nicholas Horton  
3pm

## **Notebooks with R Markdown**

with JJ Allaire  
3:45pm

## **Lowering the barriers to inclusive, collaborative, reproducible analyses**

with Chester Ismay and Andrew Bray  
4:30pm

# context

A blurred background image of a classroom full of students sitting at desks, looking towards the front. Some students are holding pens or pencils.

first course  
in stats,  
for majors or  
non-majors

only stats course  
taken by students  
or first of many

deliberate  
emphasis on  
computation  
in class or lab  
(or both)

# why R?

## why R?

free & open  
source

powerful &  
flexible

relevant  
beyond intro  
stat

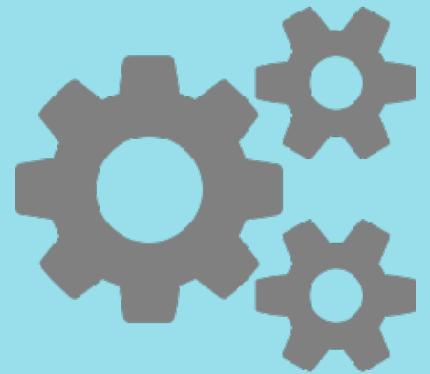
## why not R?

challenge of  
teaching  
programming in  
addition to stats  
concepts

command line  
more  
intimidating  
than GUI

**getting started  
involves  
cumbersome /  
OS dependent  
installation**

# how R?

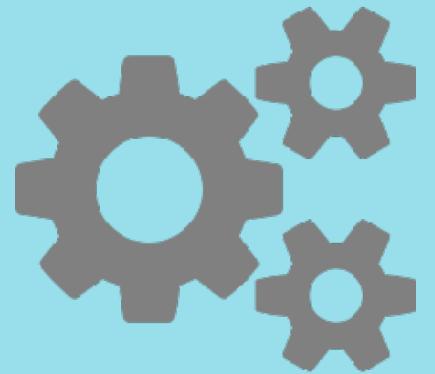


technical



pedagogical

# how R?



technical

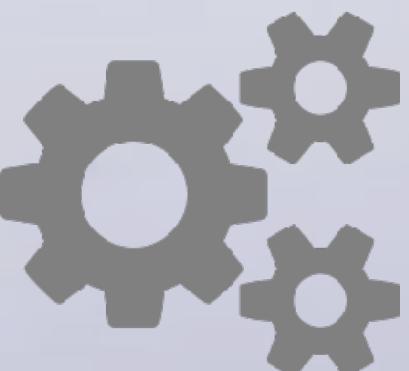


**getting started:**  
“like a knife  
through butter”

avoid local  
installation

student only  
need to bring a  
browser

preinstalled &  
preloaded  
packages



## implementation:

**option 1:**  
monolithic  
RStudio  
server  
instance

simpler  
setup

load  
prediction

account  
management +  
security  
consideration  
(large # of  
non-dept students)

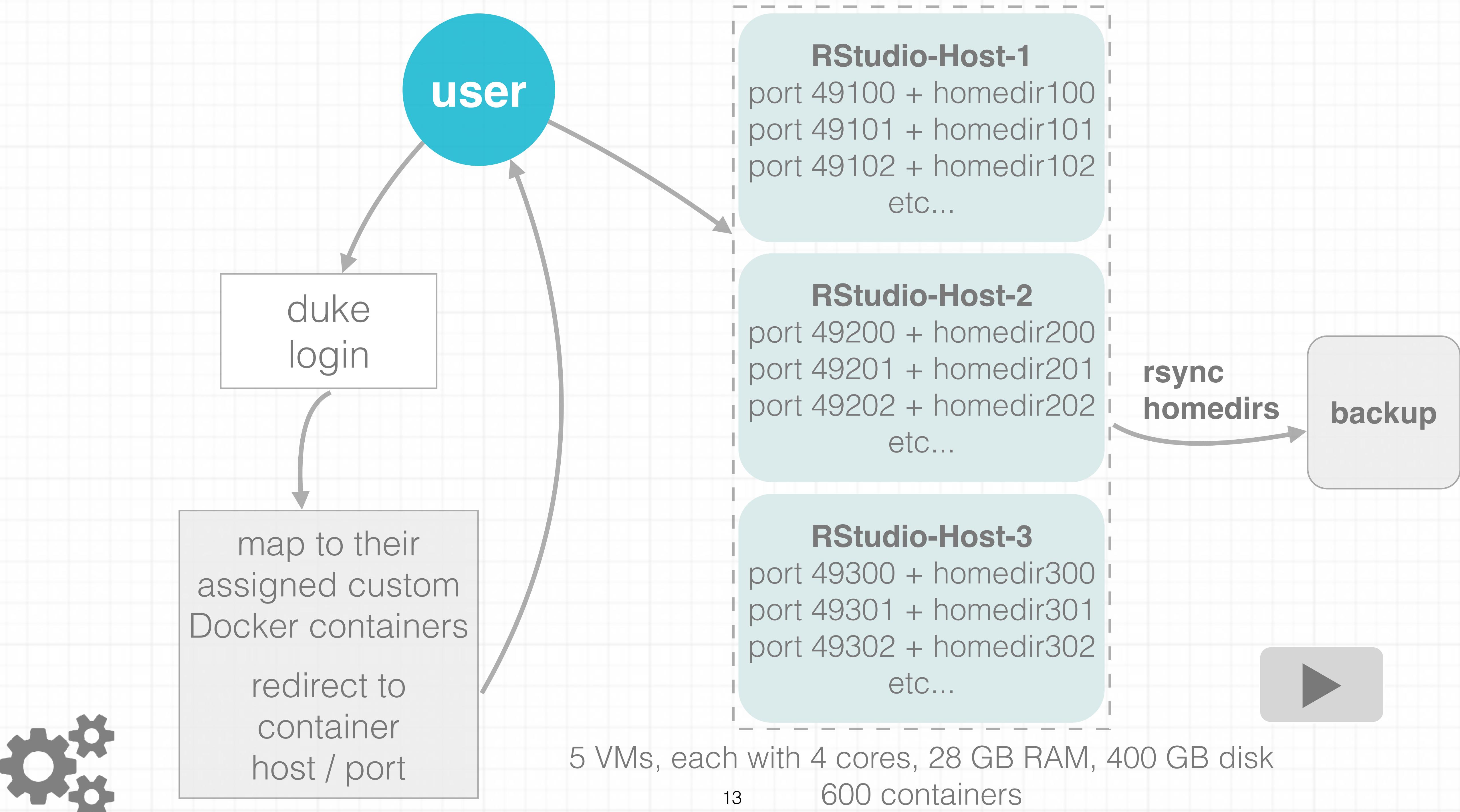
**option 2:**  
docker  


lightweight  
(with many  
virtues of  
individual VMs)

sandbox  
individual  
students

spin up new  
servers on the  
fly as needed







# mark mccahill, duke OIT

## why R Studio in ?

1

Ability to move individual users between servers to balance the load across a couple servers since I wasn't sure how to size the servers — the users don't notice this because we have them go through a different web site ([vm-manage.oit.duke.edu](http://vm-manage.oit.duke.edu)) where they log in with their Net ID, and we then redirect them to their personal container.

2

Ease of managing changes in R packages. Since there is an explicit script for building the container, it is easy to track changes as we update packages between semesters. We have evolved to a fairly complex set of packages, but it is possible to build off previous semesters' work it has not been a big deal to keep up with the instructors' requests.

## how R Studio in ?

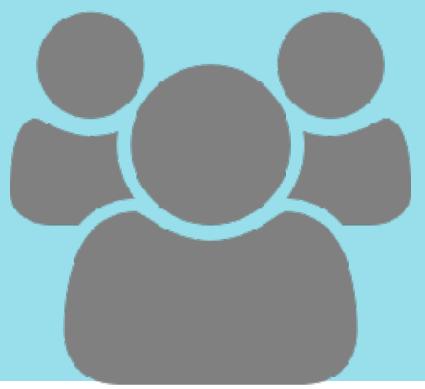
<https://github.com/mccahill/docker-rstudio>

**README** explains how we run a large-scale RStudio docker container farm.

- Typically peak at around 600 student RStudio containers during the semester, and run the containers on a set of 5 virtual machines, each with 4 cores, 28 GB RAM and 400 GB disk.
- Could probably get away with smaller machines, but this set up makes sure that if instructors assign large datasets we will not get into trouble.

**Build scripts** for the containers also available at the repo.

# how R?



pedagogical

**reproducible:**  
literate  
programming

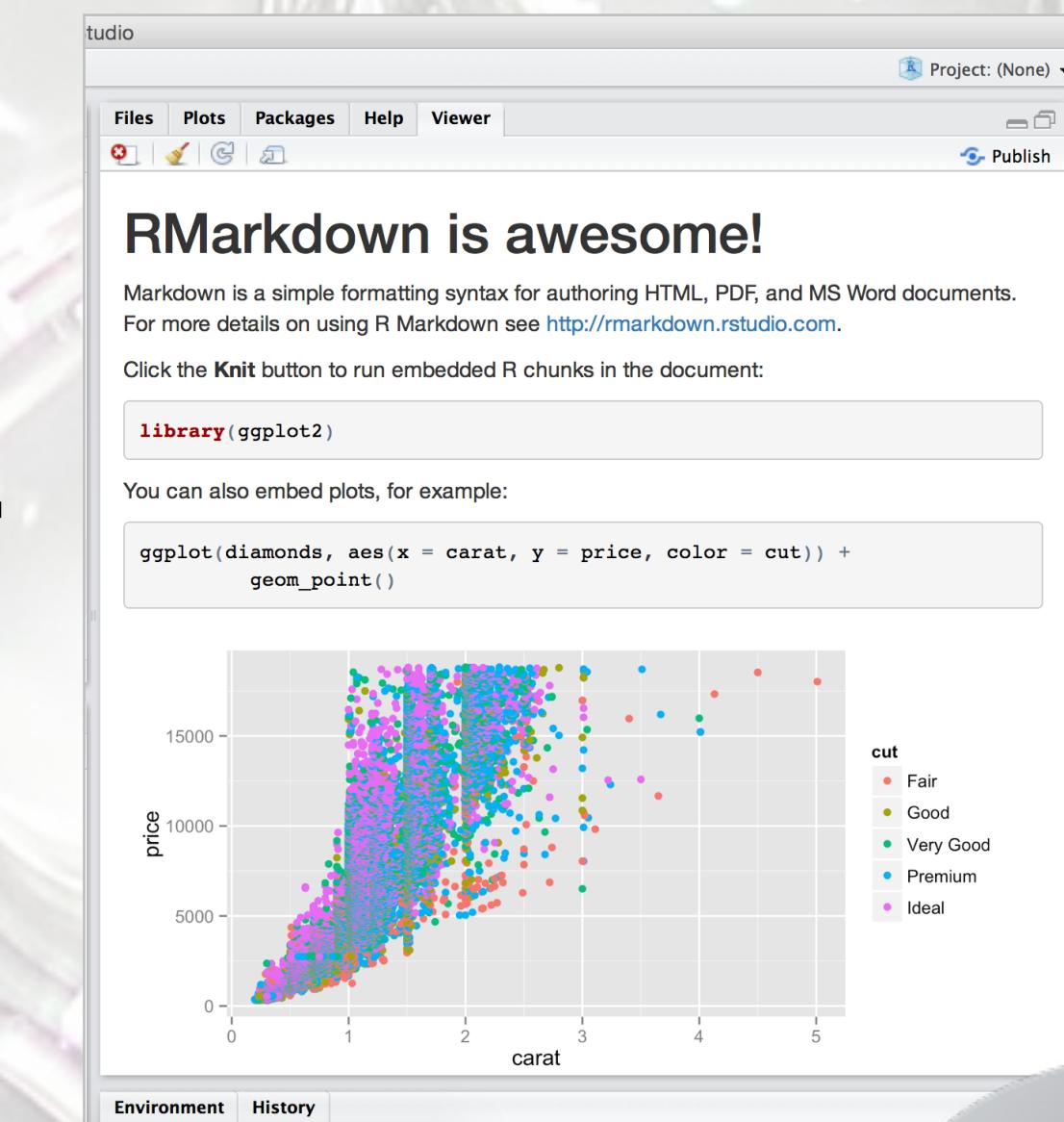
toolkit

train new  
researchers  
whose only  
workflow is a  
reproducible  
one

don't touch  
the raw data

keep track of  
all analysis  
steps

avoid copy-  
paste



= Literate programming in





**support:**  
lots to less

start with  
templates  
including code  
and answers

slowly remove  
handholding



# R Markdown learning outcomes (beyond reproducibility)

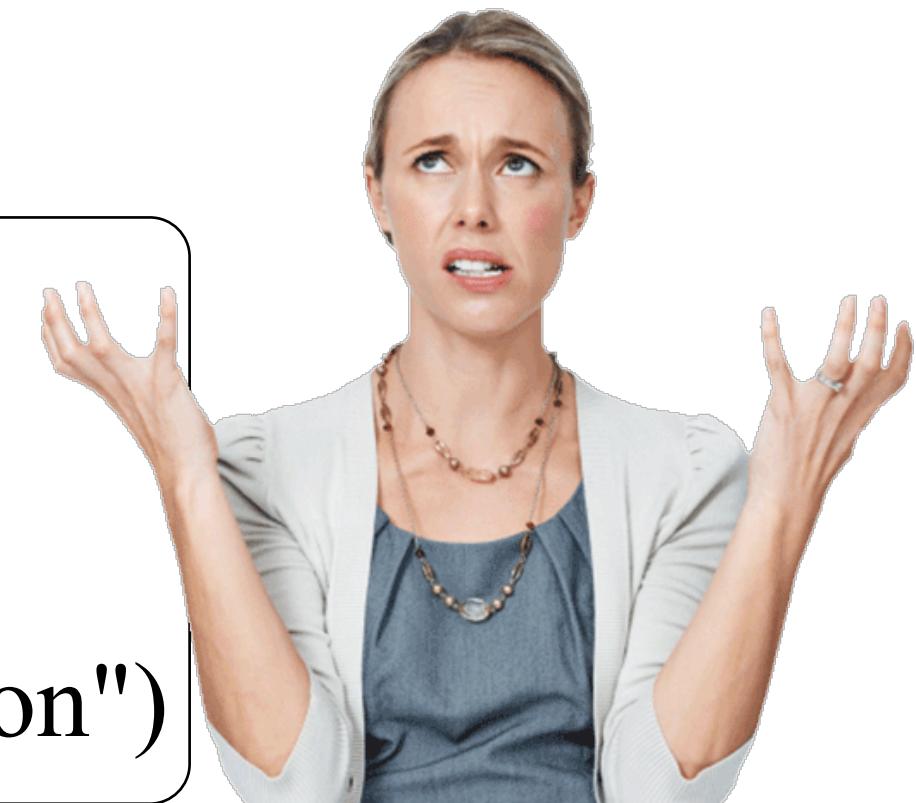


learn R

avoid the  
messy /  
frustrating  
console

built-in and  
consistent  
syntax  
highlighting

```
n <- 1000
p <- seq(0, 1, 0.01)
me <- 2 * sqrt(p * (1 - p)/n)
plot(me ~ p, ylab = "Margin of Error", xlab = "Population Proportion")
```



```
n <- 1000
p <- seq(0, 1, 0.01)
me <- 2 * sqrt(p * (1 - p)/n)
plot(me ~ p, ylab = "Margin of Error", xlab = "Population Proportion")
```

# R Markdown learning outcomes (beyond reproducibility)



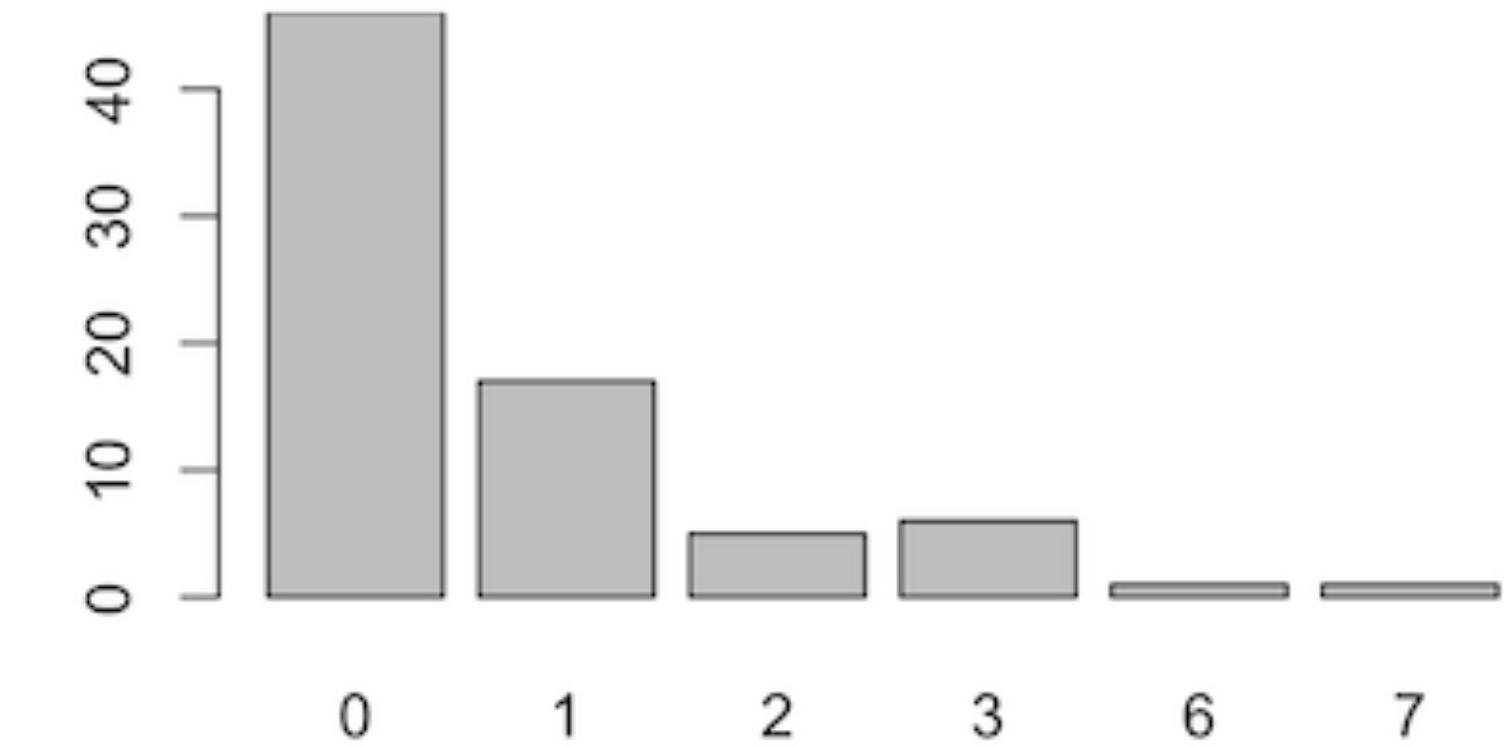
learn R

avoid the  
messy /  
frustrating  
console

built-in and  
consistent  
syntax  
highlighting

code and  
output always  
together

```
sim_streak <- calc_streak(sim_basket)  
barplot(table(sim_streak))
```



```
median(sim_streak)
```

```
## [1] 0
```

```
IQR(sim_streak)
```

```
## [1] 1
```

# R Markdown learning outcomes (beyond reproducibility)



learn R

avoid the  
messy /  
frustrating  
console

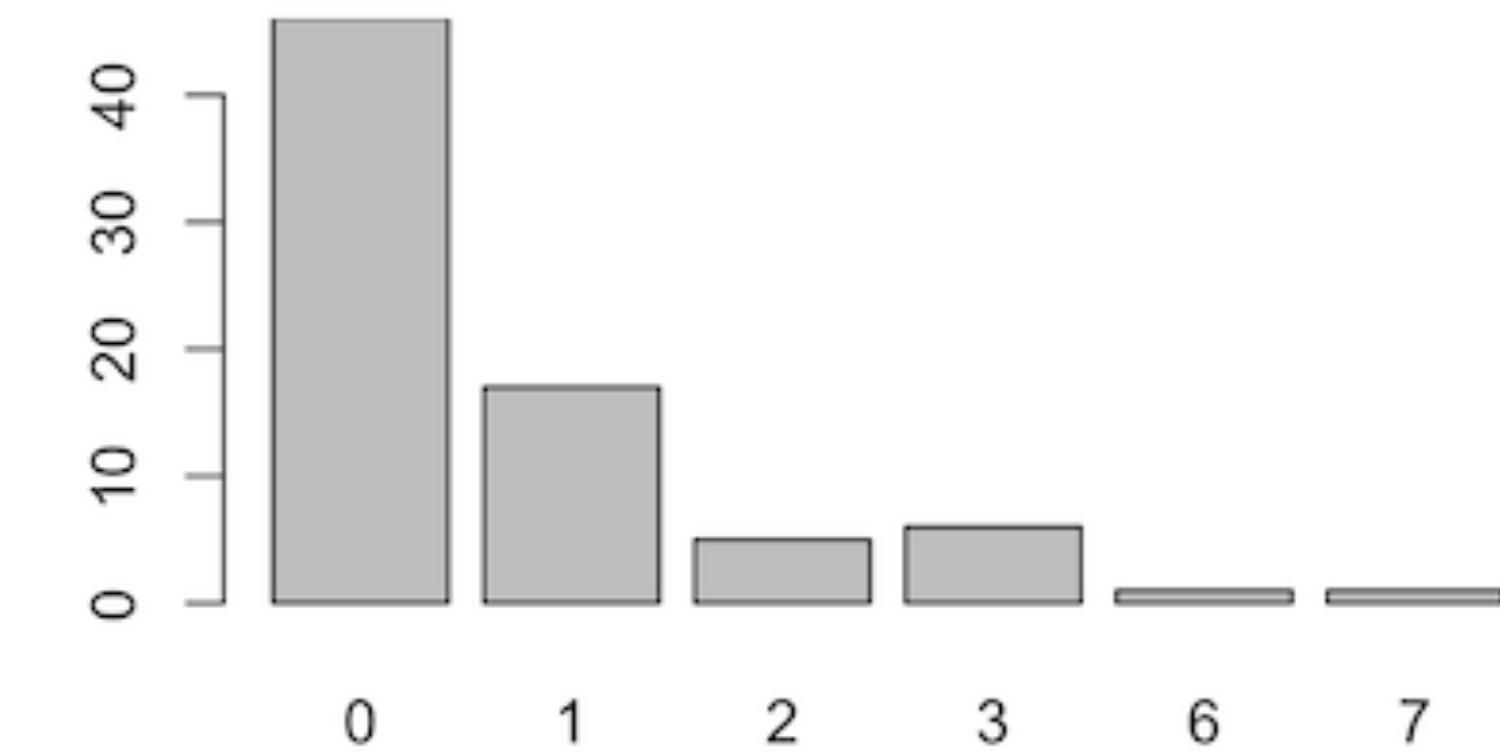
built-in and  
consistent  
syntax  
highlighting

code and  
output always  
together

feedback +  
grading

ambiguity  
removed

```
sim_streak <- calc_streak(sim_basket)  
barplot(table(sim_streak))
```



```
median(sim_streak)
```

```
## [1] 0
```

```
IQR(sim_streak)
```

```
## [1] 1
```

# R Markdown learning outcomes (beyond reproducibility)



learn R

feedback +  
grading

collaboration

avoid the  
messy /  
frustrating  
console

ambiguity  
removed

just share  
the Rmd

built-in and  
consistent  
syntax  
highlighting

code and  
output always  
together

# demo

# if you would like to follow along

1

Go to <http://tinyurl.com/ecots-rstudio> and claim a container:

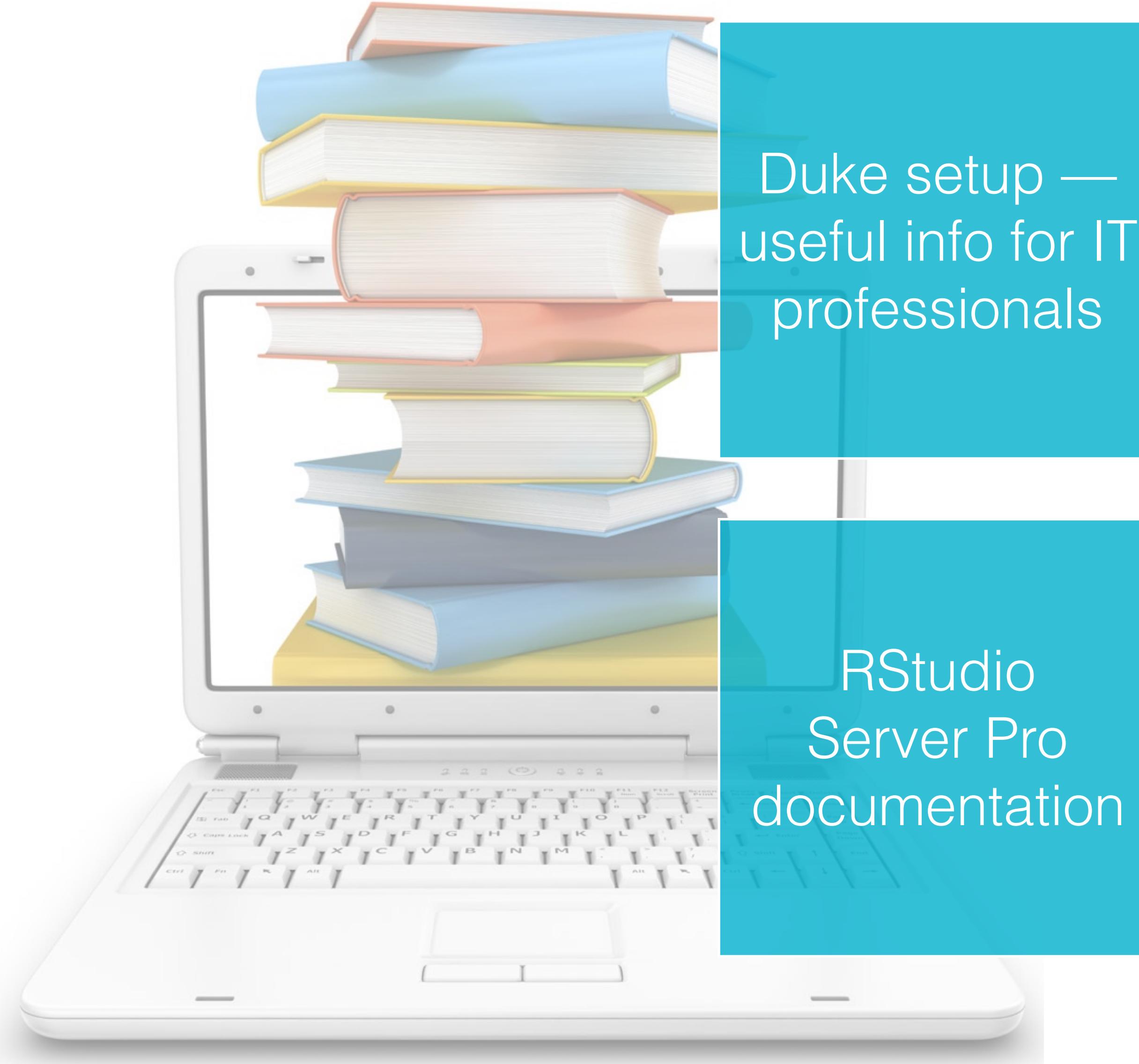
- Click on the URL
- Use guest as username, and password provided
- Mark the container you use with X to denote that it has been claimed

2

Download the sample template by typing the following in the console:

```
download.file("http://tinyurl.com/ecots-sample", destfile = "ecots_sample.Rmd")
```

# resources



<https://github.com/mccahill/docker-rstudio>

<http://docs.rstudio.com/ide/server-pro/>

# takeaways



reduce  
learning curve  
for students

reduce  
technical  
hurdles for  
instructors

minimize  
time to  
first plot/output

# thank you!

## comments / questions?



[mine@stat.duke.edu](mailto:mine@stat.duke.edu) / [nhorton@amherst.edu](mailto:nhorton@amherst.edu)



@minebocek / @askdrstats



[mine-cetinkaya-rundel](https://github.com/mine-cetinkaya-rundel) / [nicholasjhorton](https://github.com/nicholasjhorton)