

CONTAINERIZE YOUR

CODE

DALE BARR

M&MS NOV 13, 2019

I can't reproduce your results on my machine

I can't even reproduce my own results on my own machine a few months later

- Code runs, but results differ
- Code won't run at all

Software infrastructure is unreliable across users, across machines, across time

ILLUSTRATION

```
library("lme4")  
  
set.seed(62)  
  
rnorm(5)  
## load data, fit model, etc...  
  
Loading required package: Matrix  
[1] 0.80127013 0.28445671 -1.51279698 0.07463783 2.26074486
```

Me (2019)

- Ubuntu Linux 18.04
- R 3.6.0
- lme4 1.1-21

Someone else (2019)

- macOS 10.14
- R 3.5.1
- lme4 1.1-17

DEPENDENCY HELL

```
library("lme4")
sessionInfo()

R version 3.6.1 (2019-07-05)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 18.04.3 LTS

Matrix products: default
BLAS: /usr/local/lib/R/lib/libRblas.so
LAPACK: /usr/local/lib/R/lib/libRlapack.so

locale:
 [1] LC_CTYPE=en_GB.UTF-8      LC_NUMERIC=C
 [3] LC_TIME=en_GB.UTF-8      LC_COLLATE=en_GB.UTF-8
 [5] LC_MONETARY=en_GB.UTF-8  LC_MESSAGES=en_GB.UTF-8
 [7] LC_PAPER=en_GB.UTF-8    LC_NAME=C
 [9] LC_ADDRESS=C            LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_GB.UTF-8 LC_IDENTIFICATION=C

attached base packages:
[1] stats    graphics  grDevices  utils      datasets  methods   base

other attached packages:
[1] lme4_1.1-21  Matrix_1.2-17

Loaded via a namespace (and not attached):
 [1] minqa_1.2.4    MASS_7.3-51.4  compiler_3.6.1  Rcpp_1.0.2
 [5] splines_3.6.1  nlme_3.1-140   grid_3.6.1      nloptr_1.2.1
 [9] boot_1.3-22   lattice_0.20-38
```

CONTAINERS

A container is a standard unit of software that packages up code and all its dependencies so the application runs quickly and reliably from one computing environment to another.

docker.com

CONTAINER IMAGE

A Docker container image is a lightweight, standalone, executable package of software that includes everything needed to run an application: code, runtime, system tools, system libraries and settings.

docker.com

- a container is an instance of an image

POTENTIAL DOWNSIDES

- Requires command-line expertise (BUT see cloud-based solutions e.g. Binder)
- Docker installation cumbersome on some platforms (Windows)
- Tricky to manage user permissions for accessing files/folders/resources from inside the container
- Will the Docker daemon/Dockerhub still be around in 2020? 2025? 2050?

BASIC COMMANDS

```
docker --help
docker image --help
docker container --help

docker image ls
docker container ls

docker run -it --rm imagename # create instance from image

docker run -it --rm \
-v /Users/Me:/home/user/work imagename

docker build -t imagename . # build image using Dockerfile

docker push imagename[:tag] # push image to repository
```


A FEW EXAMPLES

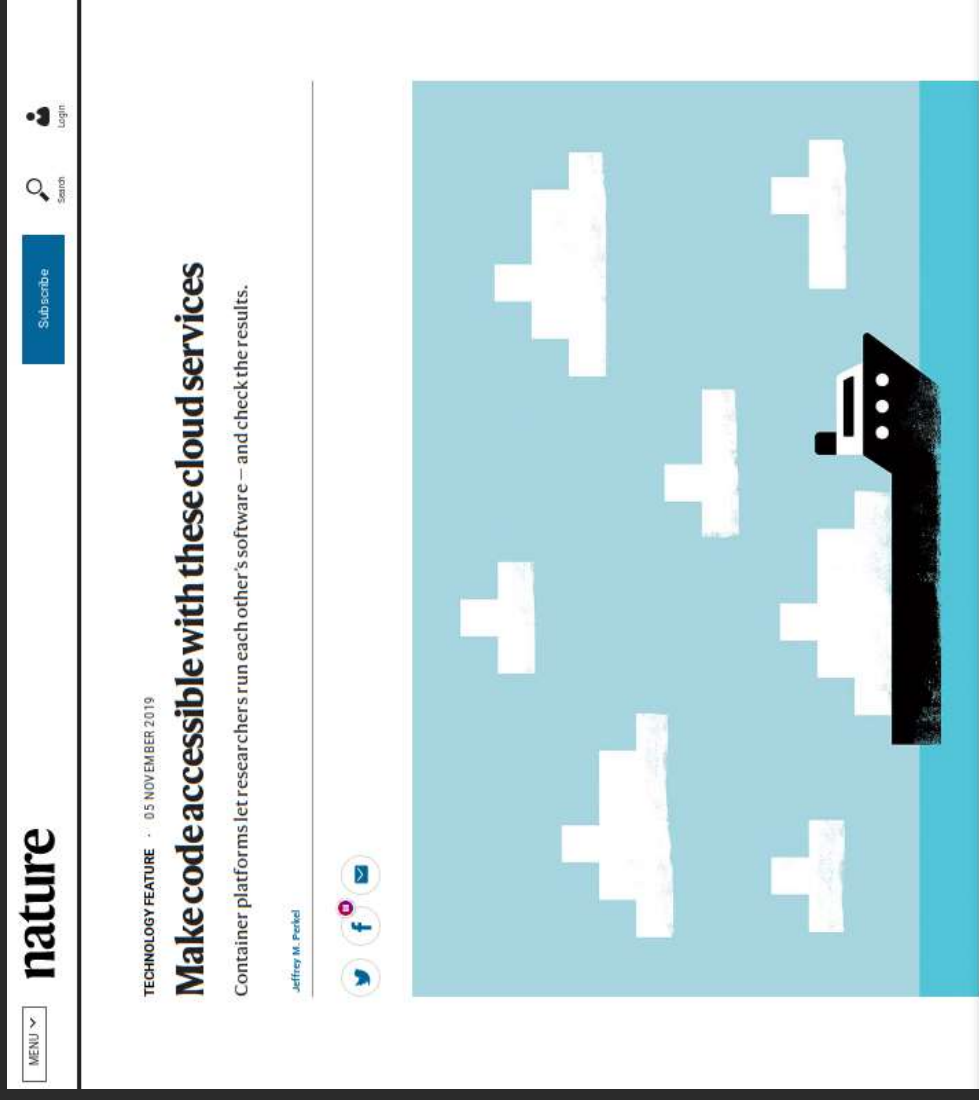
1. docker “hello-world”
2. estimating a linear-mixed effects model in Julia via R
3. fully reproducible data analysis and manuscript for 3 experiments

SOME TIPS

- Develop your analysis inside a container *from the beginning*
- Don't rely solely on a Docker image
 - keep your data outside the image
 - provide multiple 'entry-points' for reproduction
- Want to learn more about automating analyses in R?
Check out drake:
 - <https://books.ropensci.org/drake>
- Have a look at cloud-based services (esp. Binder)

Cloud-based services

- Binder
- Code Ocean
- Colaboratory
- Gigantum
- NextJournal



The image is a screenshot of a web article from the journal Nature. At the top left, there is a 'MENU' button. The 'nature' logo is prominently displayed. To the right of the logo are buttons for 'Subscribe' and 'Login'. A search bar is also visible. The article title is 'Make code accessible with these cloud services', categorized as a 'TECHNOLOGY FEATURE' dated '05 NOVEMBER 2019'. The author is 'Jeffrey M. Perkel'. The main text reads: 'Container platforms let researchers run each other's software — and check the results.' Below the text are social media sharing icons for Twitter, Facebook, and Email. The main visual is a stylized illustration of a black ship on a blue sea, with several white, pixelated clouds in the sky.

<https://www.nature.com/articles/d41586-019-03366-x>

RESOURCES

- [ROpenSciLabs, A Docker tutorial for reproducible research.](#)
- [Karthik Ram's holapunch package for R](#)
- [Marwick, Boettiger, & Mullen \(2018\). Packaging data analytical work reproducibly using R \(and friends\). PeerJ preprint.](#)