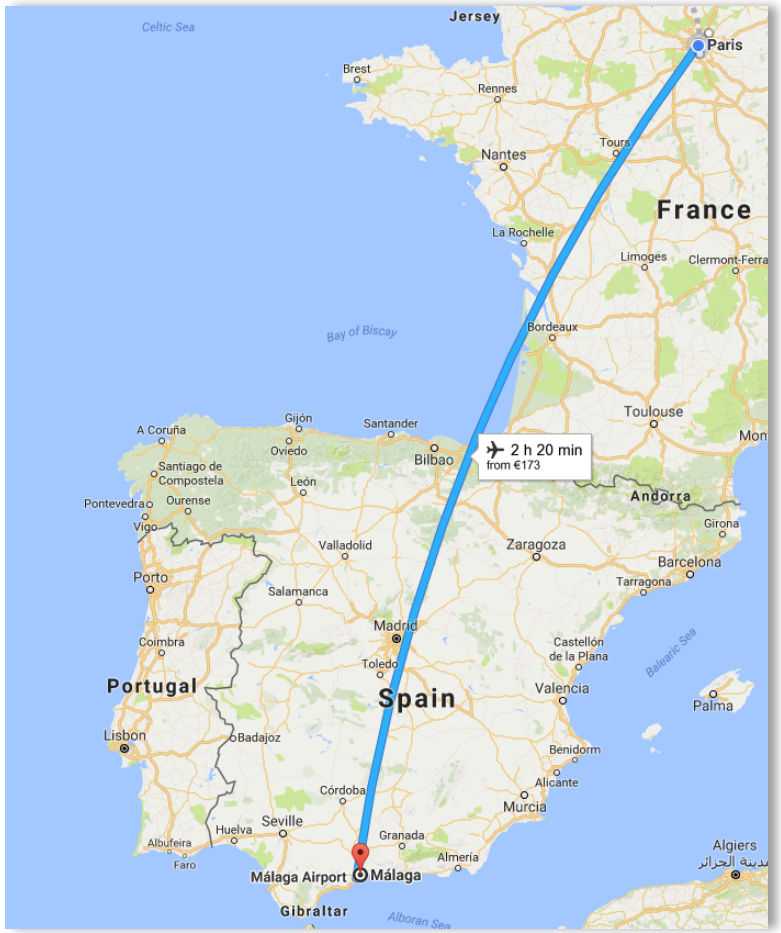
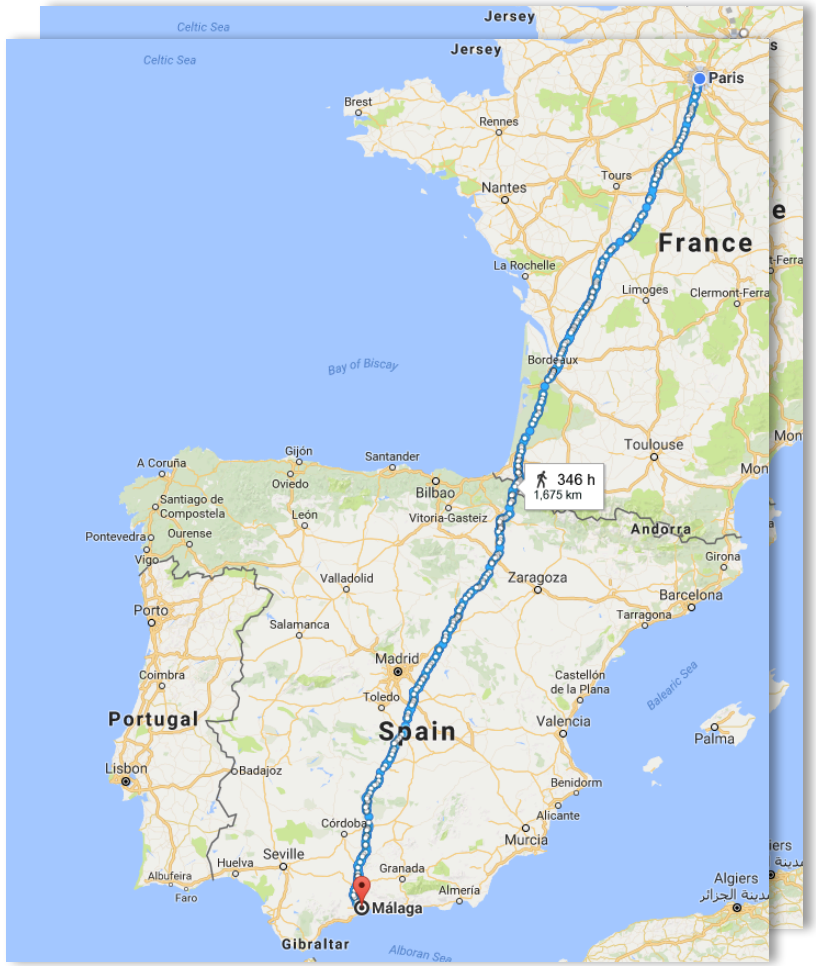
Reproducible research: your project as a journey

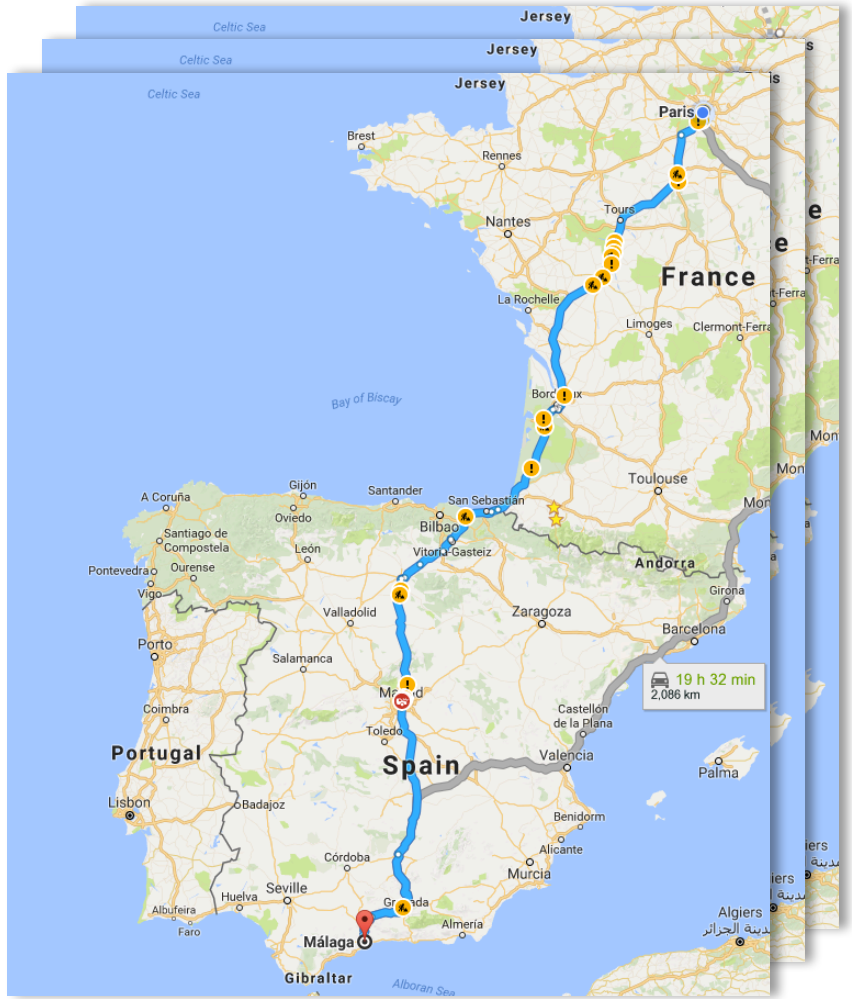
Your project in the grant



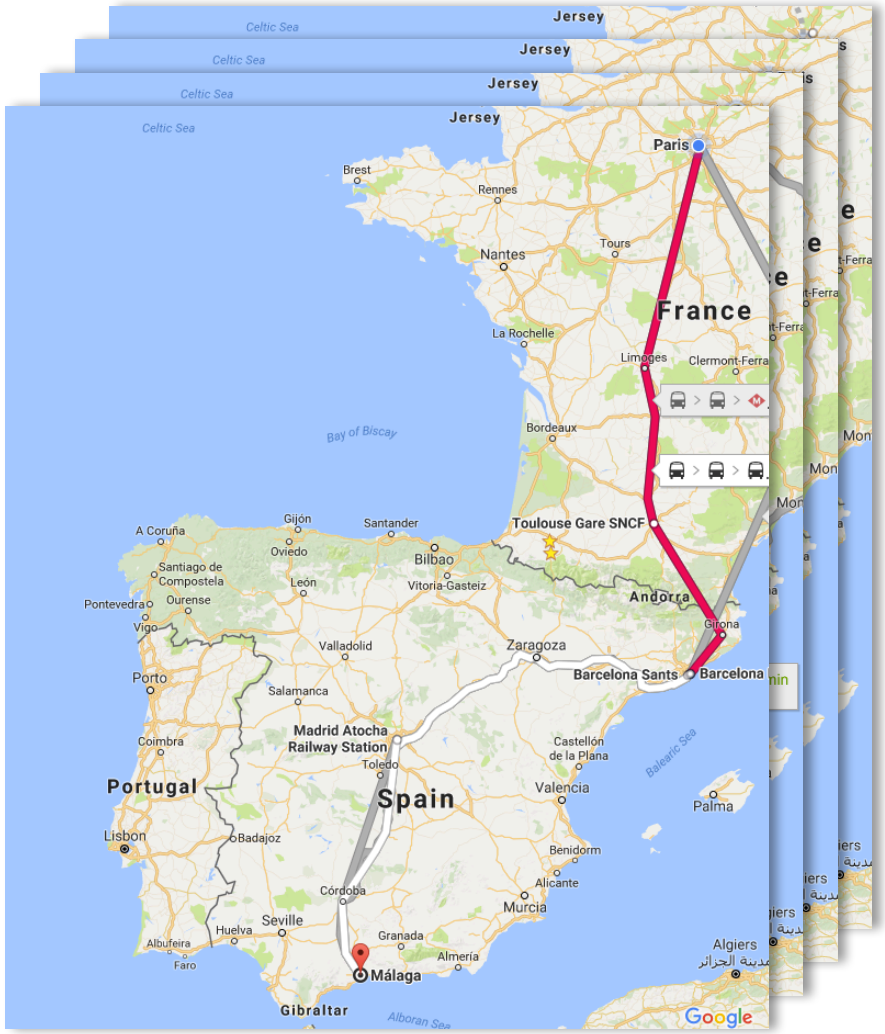
Walking is slooow…



Too many roads under construction at the moment…



Somehow this works!



What is reproducible research?

zenodo

1:15 PM–12:22 PM (Thursday) 23 h 7 min
Ouibus › Ouibus › R3 ›
AVE › AVE
1:15 PM from Paris Bercy (centre ville)
4 min
DETAILS

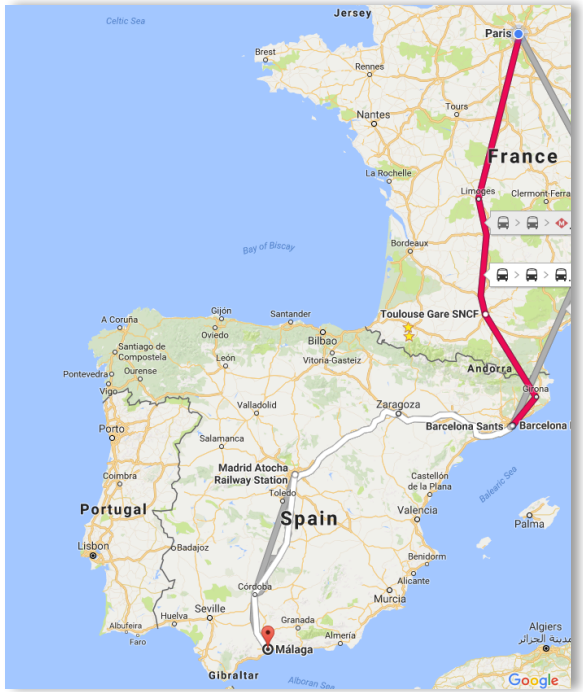
1:15 PM–12:22 PM (Thursday) 23 h 7 min
Ouibus › Ouibus › L1 ›
AVE › AVE

5:00 PM–2:24 PM (Thursday) 21 h 24 min
Ouibus › R4 › AVE

5:00 PM–2:17 PM (Thursday) 21 h 17 min
Ouibus › R4 › AVE ›
AVE

Publish everything

GitHub

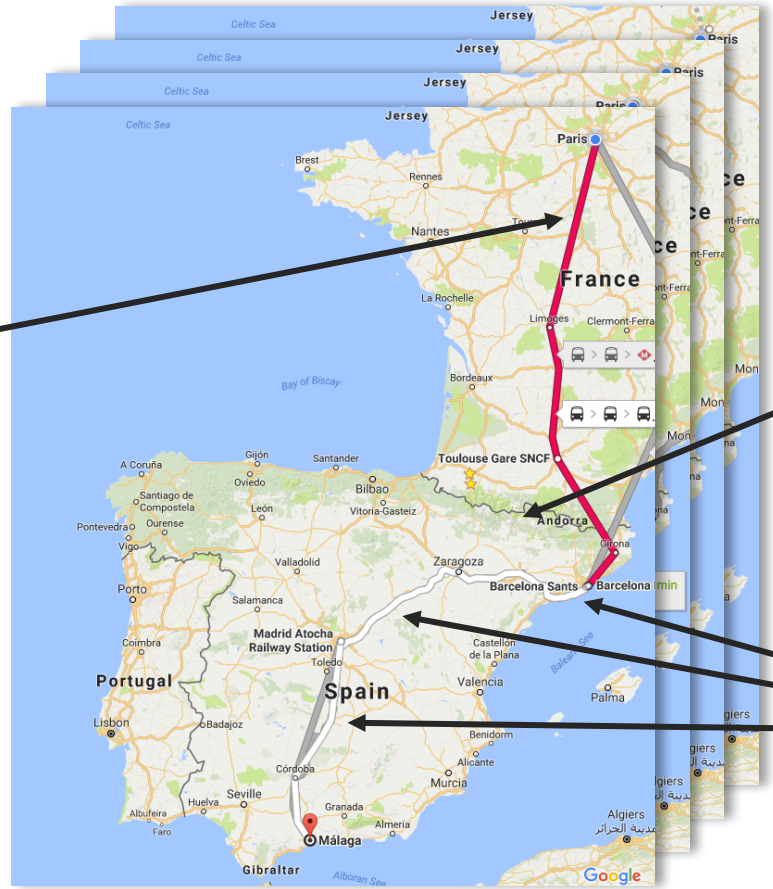


OPEN ACCESS

“It is good to have an end to journey toward, but it is the journey that matters in the end.”

Ursula K. Le Guin

Expensive: bring sandwiches

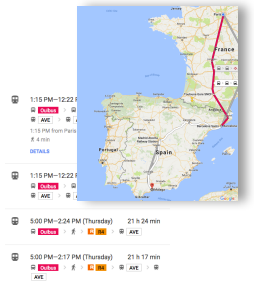


Mountains here

Weird arc

Reproducible research:
Reproducing the results AND the rationale.

The whole iceberg



OPEN ACCESS

GitHub

zenodo

git

jupyter

Publish everything

Control the environment

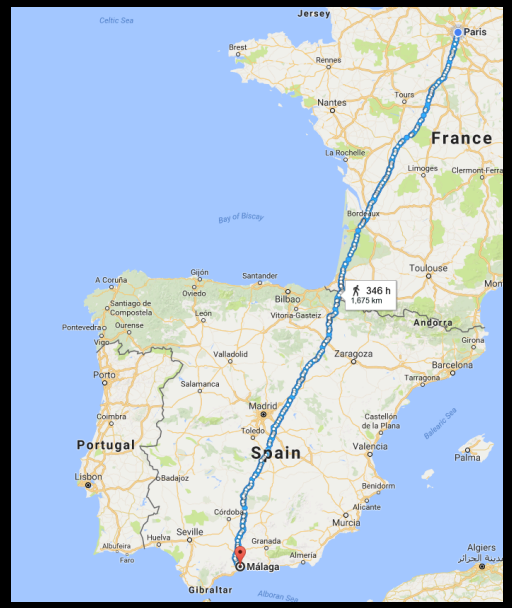
Track changes and thoughts

Quality code

docker

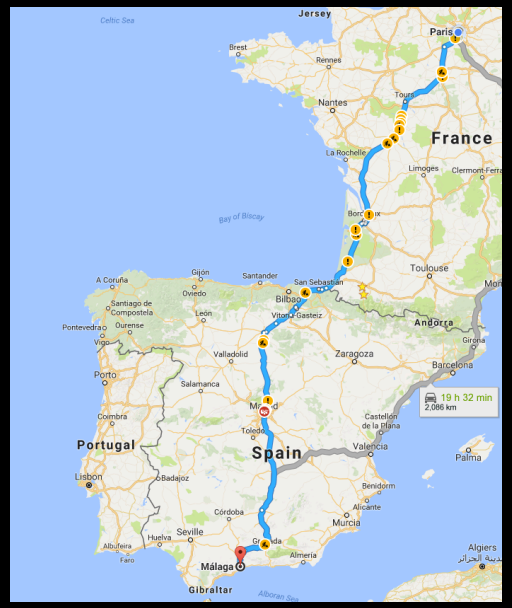
jupyter

The question of lab notebooks in bioinformatics



Feet hurt

Wheels next time?



Readable and executable notebooks

- Code + output + text, images, equations…
- Available kernels: python/R/Julia/Ruby.
- Run in the command line, access from your browser.

Goals

Dataset + code version

Analysis:

- code
- comments

Conclusions/next steps

https://github.com/hclimente/smartas

jupyter

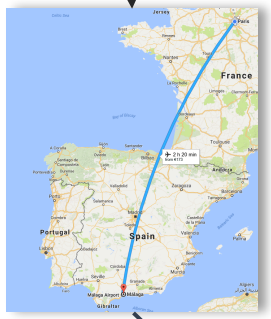
Hands-on time

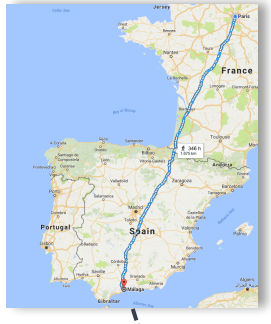
https://github.com/PeterJackNaylor/ReproducibleResearch

git

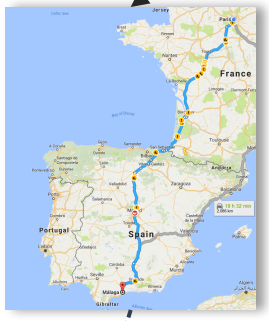
A timeline for your project

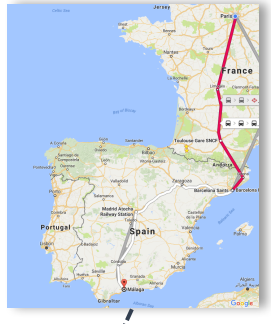
When I get my airplane…



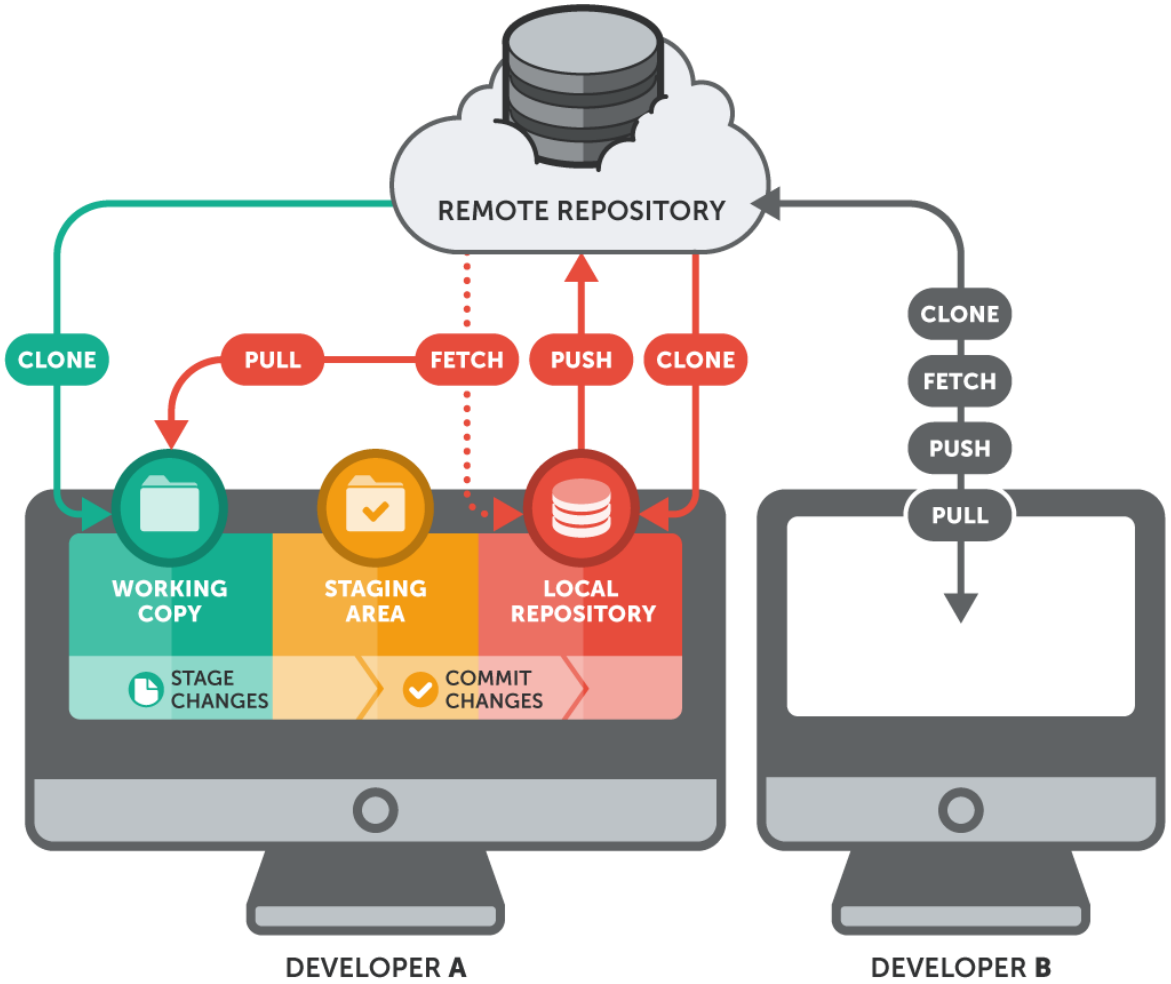


Maybe when roads get fixed?

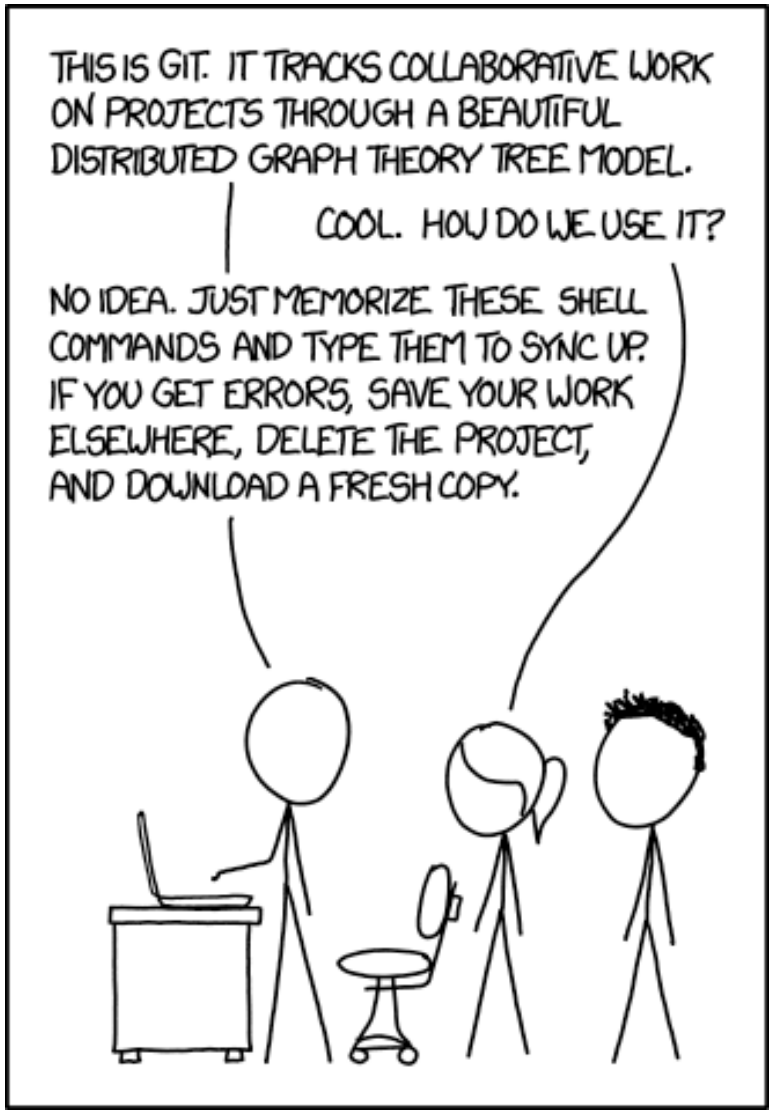




What git looks like



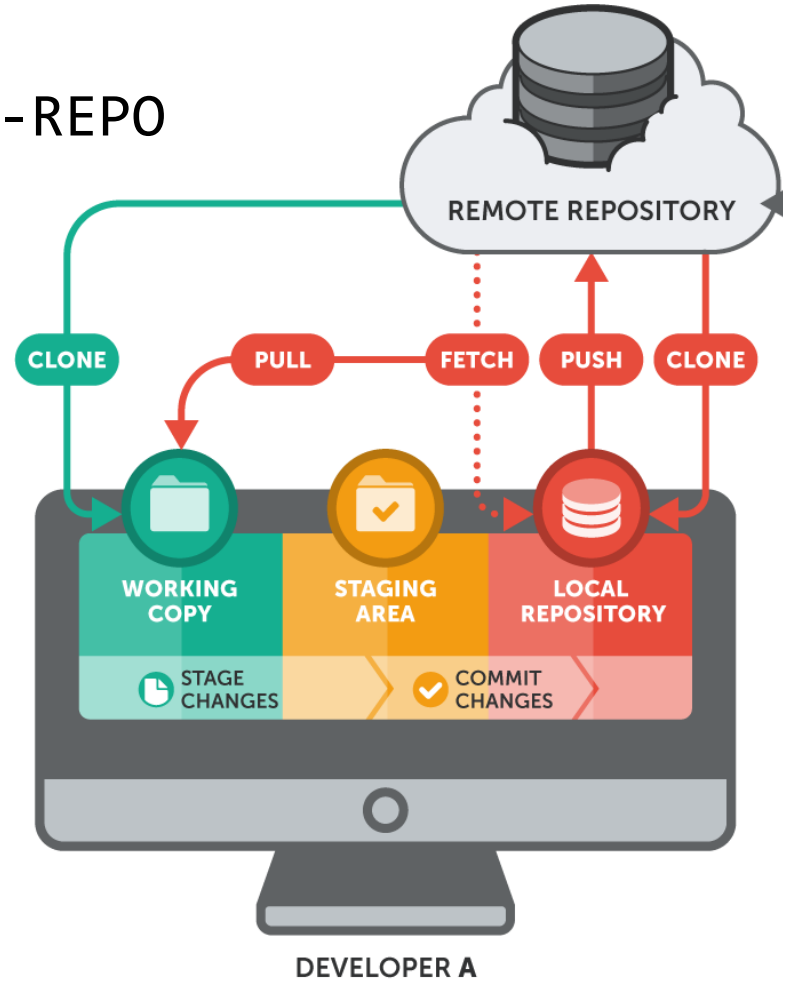
From https://www.git-tower.com/learn/git/ebook/en/command-line/remote-repositories/introduction



From https://xkcd.com/1597/

The bread and butter

```
$ git clone https://github.com/UR-USER/UR-REPO

$ git add newfileA.txt

$ git commit newfileA.txt oldfileB.txt

$ git push

$ git pull
```



From https://www.git-tower.com/learn/git/ebook/en/command-line/remote-repositories/introduction

(some) good practices

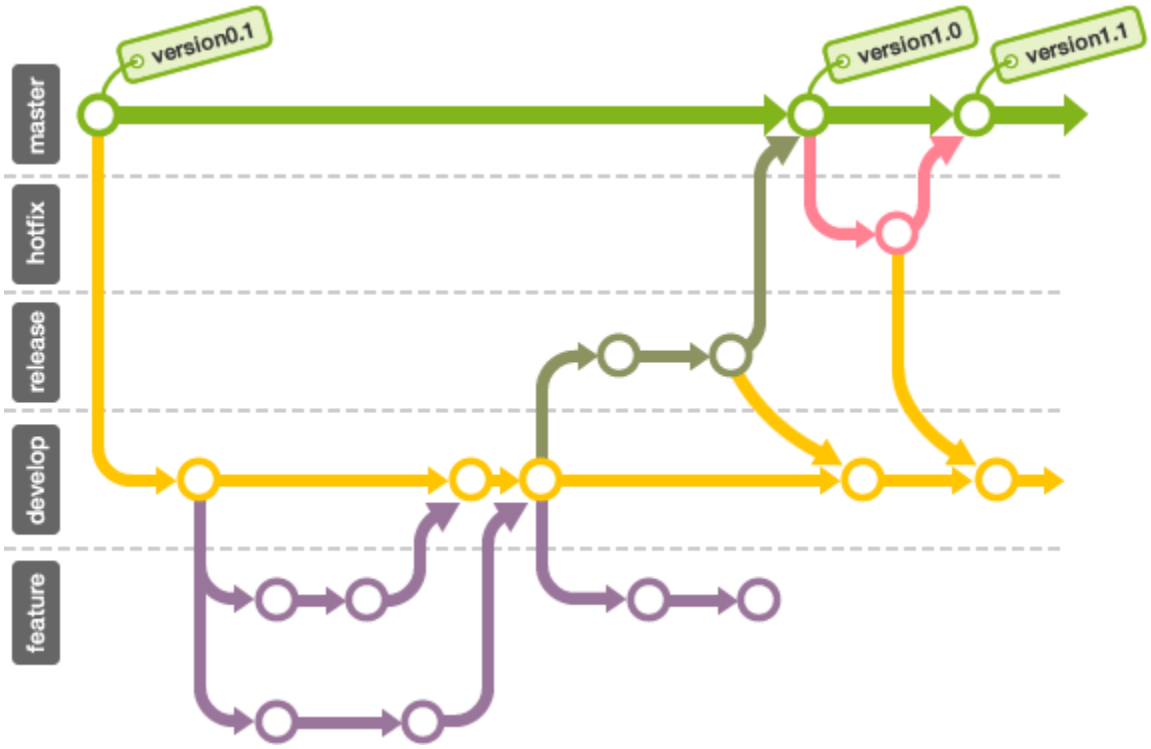
Keep up the quality of your commits

Keep commits informative

From https://xkcd.com/1296/

.gitignore up-to-date keeps useless files away

Master your branches



Join the GitHub community

- Integration with Jupyter notebooks and Markdown
- Fork repositories easily
- Public and private repositories
- Visibility

https://github.com/dekkerlab/giorgetti-nature-2016


docker

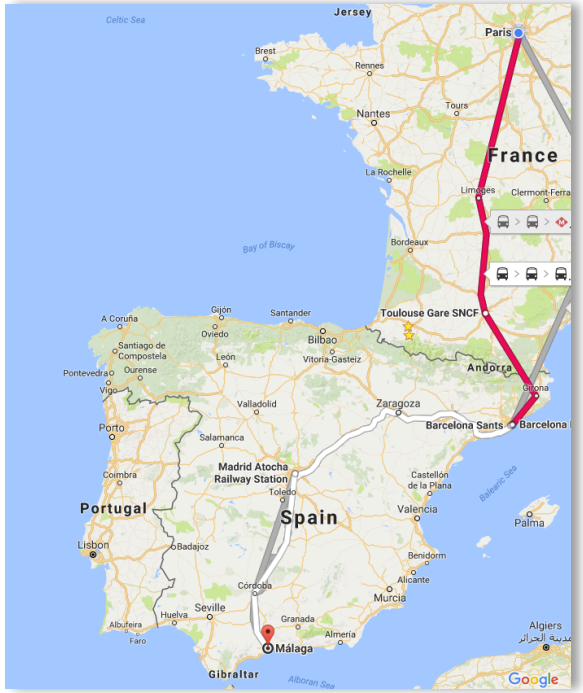
Soo a bus, huh?

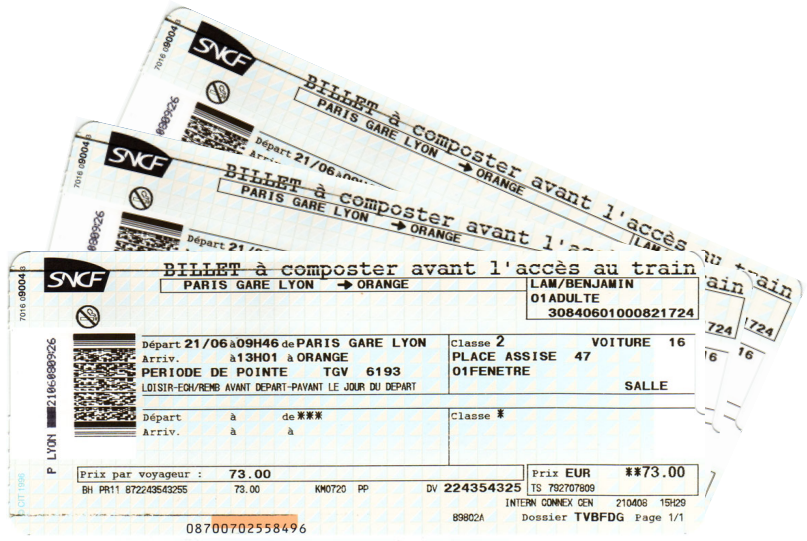
… with which bus company?

… so you took the 9:13 am bus?

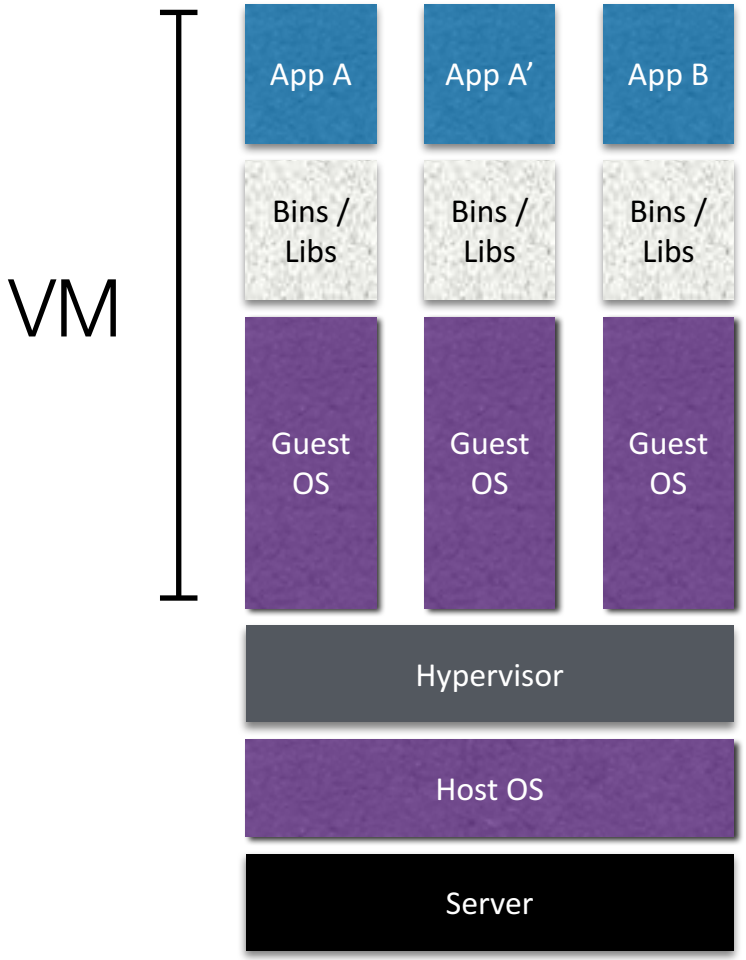
… and it was on a Monday in February?

Just follow the map with these tickets!

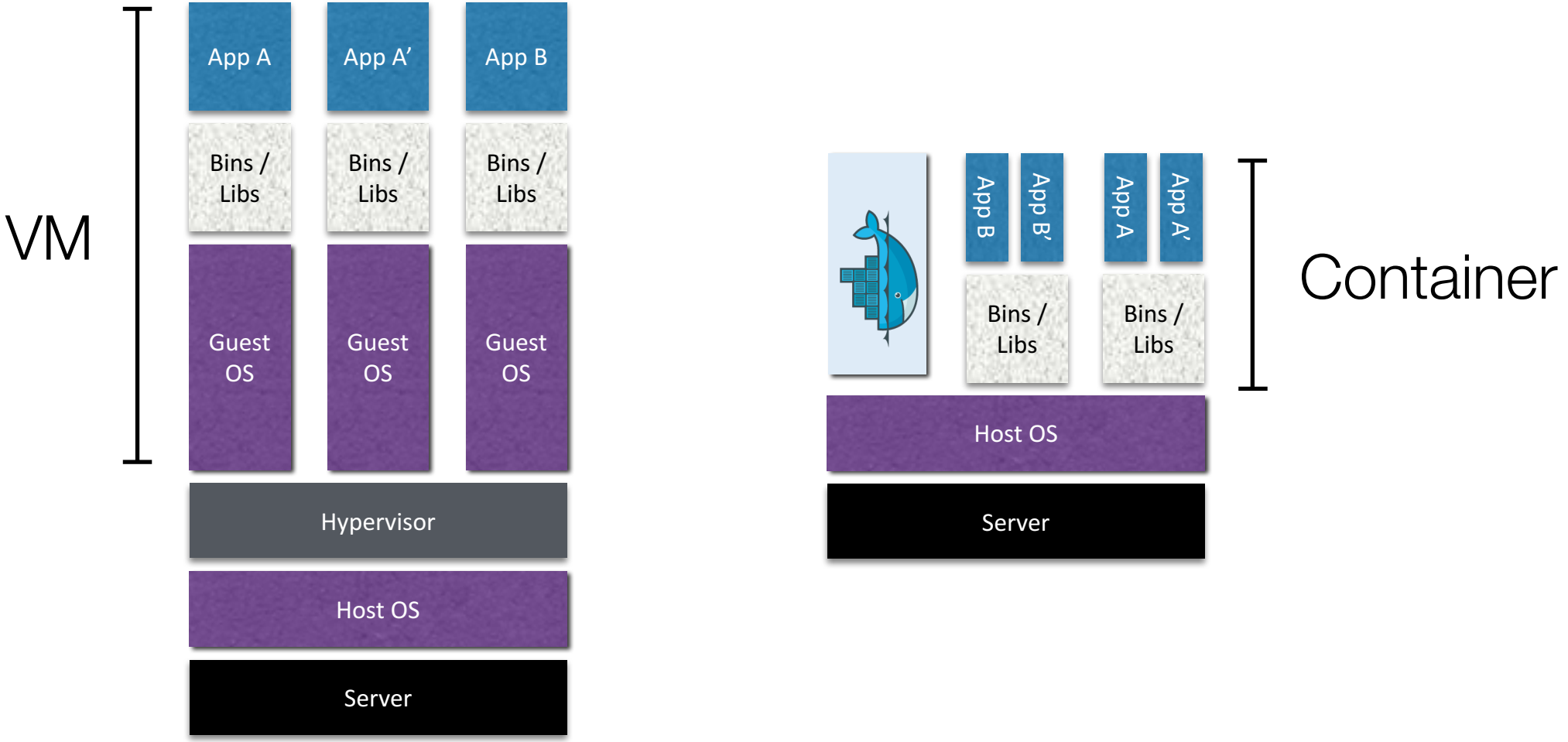




Reproducing an environment



docker: lighter, quicker and better integrated



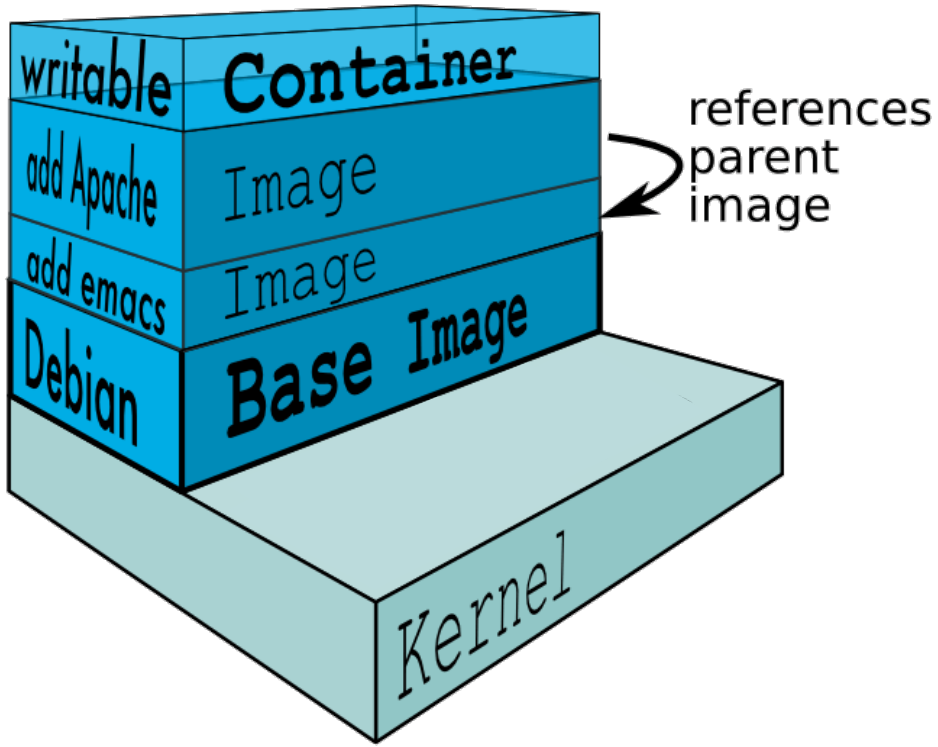
What is docker?

Open-source tool for deployment of software as containers run directly through the Linux kernel.*

* On mac and Windows docker invokes a lightweight Linux virtual machine to emulate the virtualization.

- Dependencies are just shipped along with the software!

Container ≈ *portable environment*

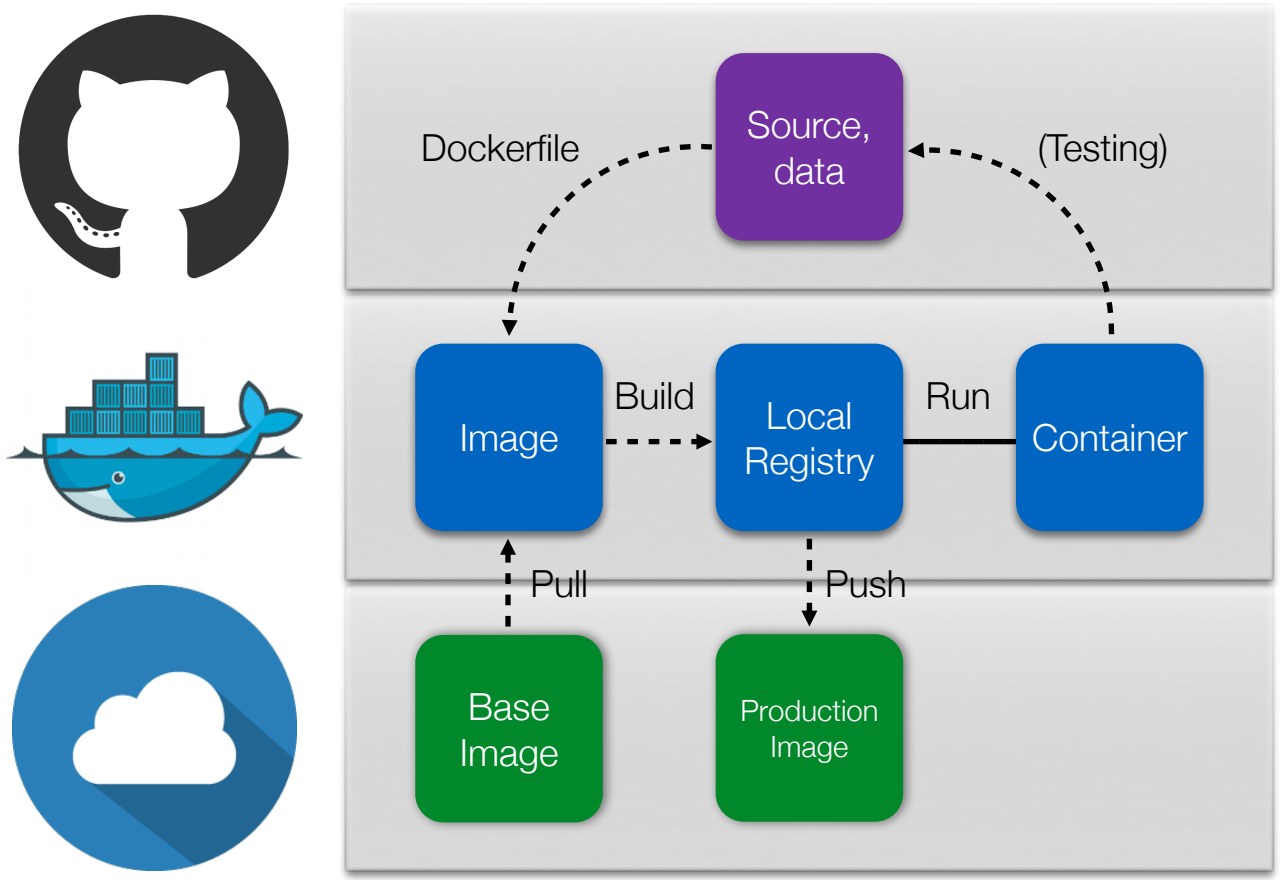


- Everything needed is in the container.
- Containers instantiate *images*.
- 100,000+ public images: Ubuntu, CUDA, Anaconda, Tensorflow, Rstudio, and more.

https://github.com/gerstung-lab/AML-multistage

Adapted from https://blog.risingstack.com/operating-system-containers-vs-application-containers/

(A) docker workflow



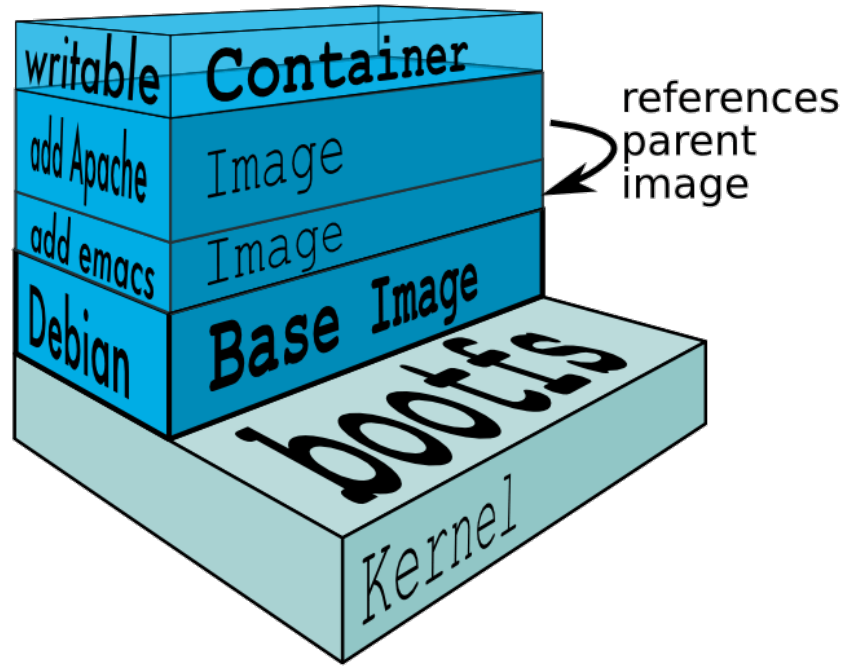
docker bread and butter

```
$ docker build DOCKERFILE

$ docker run IMAGE

$ docker pull USER/IMAGE

$ docker push USER/IMAGE
```

Different ways to create an image

A) Interactively:

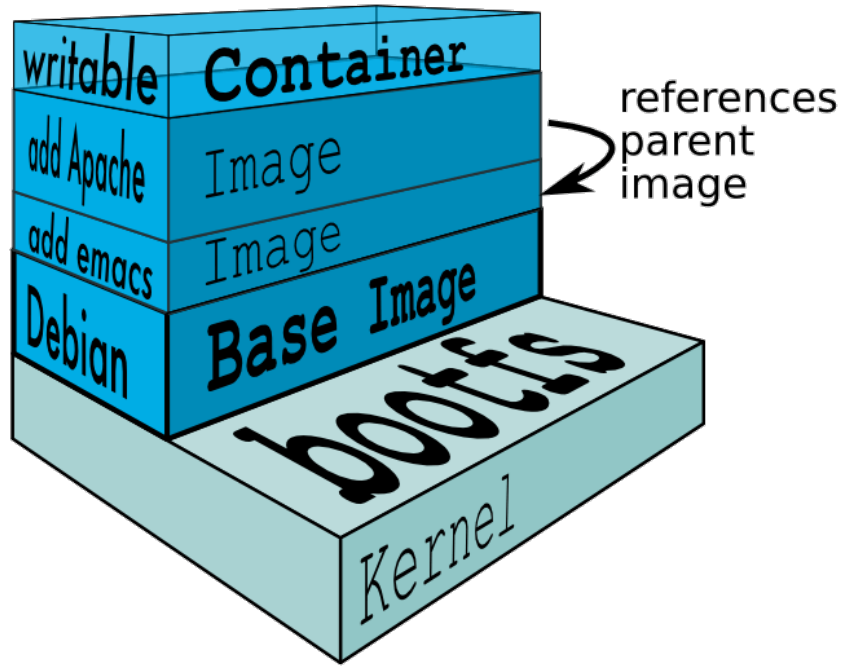
1. Run a base image
2. Install new source, data, packages (apt-get, pip, etc.)
3. Commit changes to a new image



Different ways to create an image

B) Write a build file (**Dockerfile**): bash commands run in sudo mode

```
1  FROM gcr.io/tensorflow/tensorflow:1.0.1 #-gpu
2  MAINTAINER Joseph Boyd <joseph.boyd@curie.fr>
3
4  COPY mnist_classification.ipynb /notebooks
5  WORKDIR /notebooks
6
7  RUN chmod 777 mnist_classification.ipynb
8
9  RUN mkdir img
10 ADD img img/
11
12 RUN mkdir src
13 ADD src src/
14
15 RUN mkdir data
16 COPY t10k-images-idx3-ubyte.gz data/
17 COPY t10k-labels-idx1-ubyte.gz data/
18 COPY train-images-idx3-ubyte.gz data/
19 COPY train-labels-idx1-ubyte.gz data/
```



Questions? Comments? Experiences?