



## R from Containers

---

# Collaboration beyond company boundaries

Share data and analysis code

→ Challenge: **reproducibility**

- ◆ R version
- ◆ R package versions
- ◆ System libraries, for instance, math libraries (BLAS, Armadillo)

→ Inconsistencies are normally easy to detect, not so easy to fix, and maybe numeric differences are hiding there

→ Traditional recommendation:

Describe computing environment with

- ◆ `Sys.info()` and
- ◆ `sessionInfo()`

---

# Reproducibility via container

<https://www.rocker-project.org>



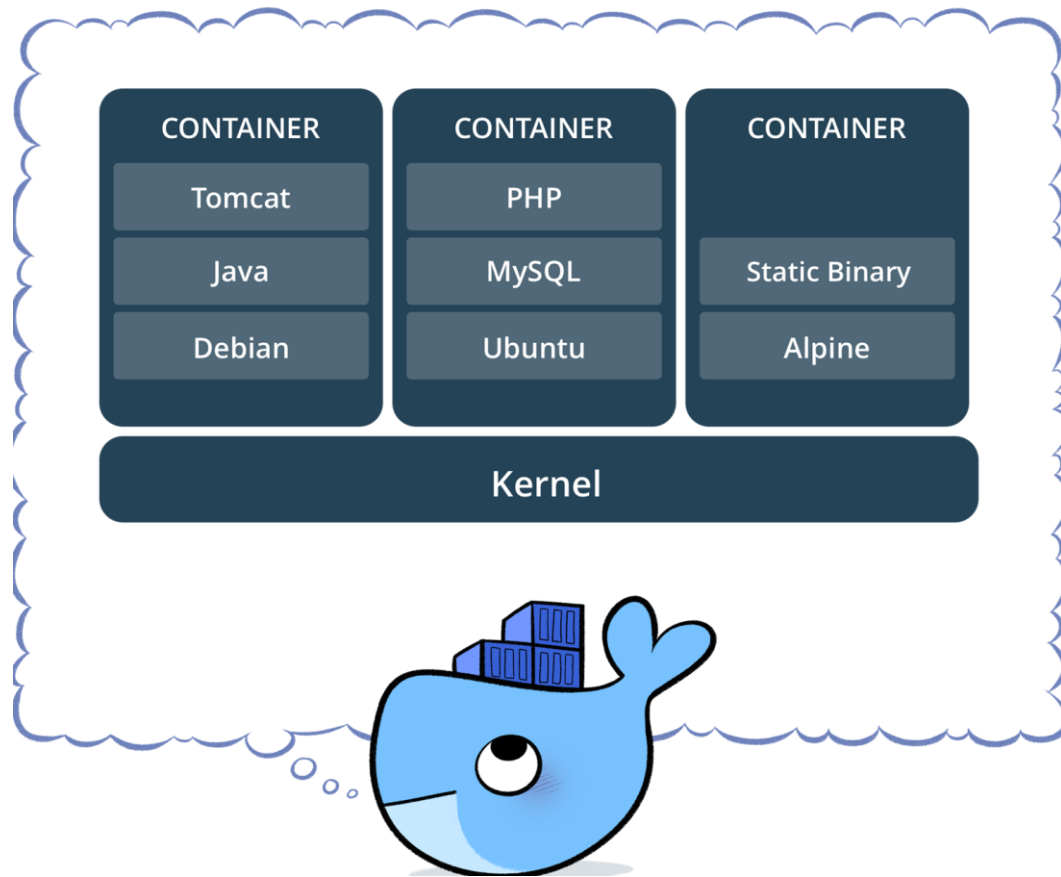
R in **docker**

# Docker container

## Packages

1. Scripts
2. R, R packages, and collateral software
3. System libraries
4. Operating system

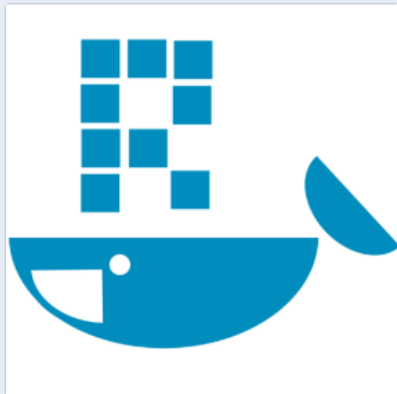
into **one** portable unit



---

# Dockerfile

```
FROM rocker/rstudio:latest
RUN apt-get update -qq && apt-get -y --no-install-
recommends install \
  libxml2-dev \
  ...
  unixodbc-dev \
  && R -e "source('https://bioconductor.org/biocLite.R')"
\
&& install2.r --error \
  --deps TRUE \
  tidyverse \
  dplyr \
  ggplot2 \
  devtools \
  formatR \
```






# rocker

## Rocker Project

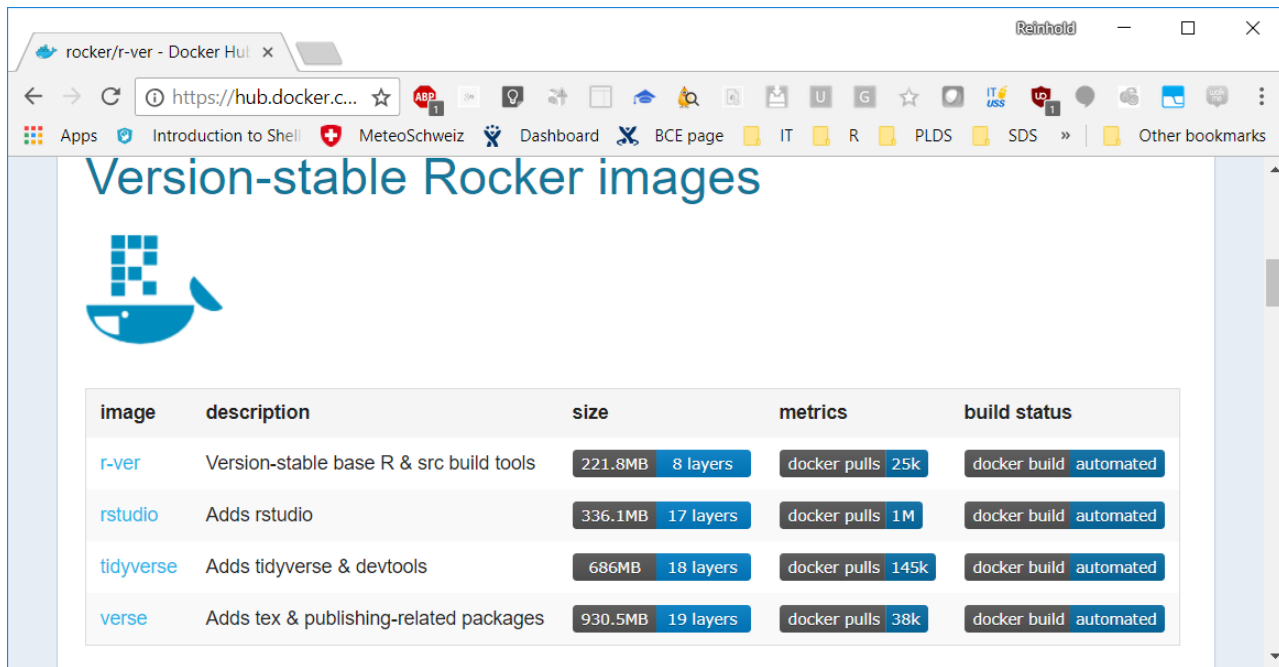
<https://github.com/rocker-org/rocker>

Joined October 2014

 <a href="#">rocker/rstudio</a> public   automated build	187 STARS	1M+ PULLS	<a href="#">DETAILS</a>
 <a href="#">rocker/shiny</a> public   automated build	83 STARS	100K+ PULLS	<a href="#">DETAILS</a>
 <a href="#">rocker/tidyverse</a> public   automated build	28 STARS	100K+ PULLS	<a href="#">DETAILS</a>
 <a href="#">rocker/hadleyverse</a> public   automated build	55 STARS	50K+ PULLS	<a href="#">DETAILS</a>

# Docker Image Store

hub.docker.com



The screenshot shows a web browser window displaying the Docker Hub page for 'rocker/r-ver'. The browser's address bar shows 'https://hub.docker.c...'. The page title is 'Version-stable Rocker images' and features the Docker logo. Below the title is a table listing four Docker images: 'r-ver', 'rstudio', 'tidyverse', and 'verse'. Each row in the table provides details on the image name, description, size, number of layers, Docker pulls, and build status.

image	description	size	metrics	build status
<a href="#">r-ver</a>	Version-stable base R & src build tools	221.8MB <a href="#">8 layers</a>	<a href="#">docker pulls 25k</a>	<a href="#">docker build automated</a>
<a href="#">rstudio</a>	Adds rstudio	336.1MB <a href="#">17 layers</a>	<a href="#">docker pulls 1M</a>	<a href="#">docker build automated</a>
<a href="#">tidyverse</a>	Adds tidyverse & devtools	686MB <a href="#">18 layers</a>	<a href="#">docker pulls 145k</a>	<a href="#">docker build automated</a>
<a href="#">verse</a>	Adds tex & publishing-related packages	930.5MB <a href="#">19 layers</a>	<a href="#">docker pulls 38k</a>	<a href="#">docker build automated</a>

---

# Get a Docker Environment

Requires

- a modern 64-bit operating system,
- administrator privileges.

Download from <https://www.docker.com/community-edition>,

install docker.



# RStudio via Docker

From command prompt:

```
docker run --rm -p 8787:8787 rocker/rstudio:3.4.4
```

Then point your browser to <http://localhost:8787>,  
login as "rstudio" with password "rstudio".

Easy way to explore different versions of R **without**  
disturbing your local R installation.

**And:**

You are using a global standard **everybody** can  
access - not a difficult to replicate local installation!!

---

# Write Reproducible Research in R

Package **rrtools** from

<https://github.com/benmarwick/rrtools>

```
rrtools::use_compendium("iris")
```

```
usethis::use_mit_license(name = 'Reinhold  
Koch')
```

```
devtools::use_github(".", auth_token =  
Sys.getenv('GITHUB_PAT'), private =  
FALSE, protocol = 'https')
```

```
rrtools::use_analysis()
```

```
rrtools::use_dockerfile()
```

---

# Containerize Analysis

