



Using Docker for analysis: Why? How?

What's the problem?

1. We do complicated stuff on computers.
 - 1.1 Systems differ, dependencies matter, stuff breaks.
2. We do science.
 - 2.1 Science should be repeatable.
3. When things change across platforms, or over time, the results may not be repeatable, and the reason why might not be clear.

How can we fix this?

There are multiple solutions including virtualization,
cloud computing, and *containerized computing*.

Containers

- Tiny set of instructions that can be run on ANY computer.
- Creates a "mini-computer" that is built to your requirements to mimic settings of your analyses.
- Allows others to run your code and presumably recreate your analyses.

Docker

- Docker is one containerization tool.
(www.docker.com)
 - It's commonly used for scientific computing.
 - It has a cute whale icon.



How to Docker from R

First: install Docker

Then, use `liftr` Package in R.

<https://liftr.me/articles/liftr-intro.html>

Uses metadata from your `.Rmd` file to set the stage.

```
---
```

```
title: "Mixomics TGIP sRNA and mRNA"
author: "Seth Baribeau"
date: "2021-07-16"
output: rmarkdown::html_document
liftr:
  maintainer: "Seth Baribeau"
  email: "seth.baribeau@gmail.com"
  from: "rocker/verse3.6.1" # <- this is where it finds the R version
  pandoc: true # <- stuff for making report docs
  texlive: false
  sysdeps:
    - gfortran
cran: # <- packages you need for your analysis
  - edgeR # from CRAN
  - pacman
  - igraph
bioc:
  - mixOmics/6.16.3 # from Bioconductor (note version no.)
---
```

Within R, you can use the `liftr` package to create a Docker.

Dockerfile:

Contains all the info to make your docker image.

```
FROM rocker/verse3.6.1
```

```
MAINTAINER Seth Barribeau <seth.barribeau@gmail.com>
```

```
# System dependencies for required R packages
RUN rm -f /var/lib/dpkg/available \
&& rm -rf /var/cache/apt/* \
&& apt-get update -qq \
&& apt-get install -y --no-install-recommends \
ca-certificates \
libssl-dev \
libcurl4-openssl-dev \
libxml2-dev \
git
```

```
RUN apt-get update -qq && apt-get install -y --no-install-recommends gfortran
```

I needed to tweak this on my computer. Might be easier on non-windows platforms.

See the Dockerfile at

https://github.com/sethbarr/tgip_mixomic_2/ for where I adjusted things.

To actually make the docker image, you need to run
the following commands:

```
docker build -t mixomic .
docker run --rm -it -p 8787:8787/tcp -v '/$(pwd):/home/rstudio mixomic:latest'
```

This should then make a website that you can access at:

<http://localhost:8787/>

Sign in to RStudio

Username:

rstudio

Password:

.....

Stay signed in when browser closes

You will automatically be signed out after 60 minutes of inactivity.

Sign In

localhost:8787

R studio 2 Project: (None)

File Edit Code View Plots Session Build Debug Profile Tools Help

tgipMixomics_spfiedCorr_allComp...* tgipmixomics.Rmd

ABC Knit

1 ---
2 title: "Mixomics TGIP sRNA and mRNA"
3 author: "Seth Baribeau"
4 date: "2021-07-16"
5 output: rmarkdown::html_document
6 liftr:
7 maintainer: "Seth Baribeau"
8 email: "seth.baribeau@gmail.com"
9 from: "rocker/verse3.6.1"
10 pandoc: true
11 texlive: false
12 sysdeps:
13 - gfortran
14 cran:
15 - edgeR
16 - pacman
17 - igraph
18 bioc:
19 # Mixomics TGIP sRNA and mRNA

Console Terminal Jobs

R 4.1.0 . ~/

, context: Error restoring session data (loading package eugeR)]; OCCURRED AT rstudio::core::Error rstudio::r::exec::{anonymous}::evaluateExpressionsUnsafe(SEXP, SEXP, SEXP**), rstudio::r::sexp::Protect*, rstudio::r::exec::{anonymous}::EvalType) src/cpp/r/RExec.cpp:186; LOGGED FROM: void rstudio::r::session::search_path::{anonymous}::loadPackage(const string&, const string&) src/cpp/r/session/RSearchPath.cpp:202
Error cleaning up session: Permission denied
26 Aug 2021 01:20:30 [rsession-rstudio] ERROR system error 13 (Permission denied) [path: /home/rstudio/.local/share/rstudio/sessions/active/session-5238fe2f/suspended-session-data, context: Error cleaning up session]; OCCURRED AT rstudio::core::Error rstudio::core::FilePath::remove() const src/cpp/shared_core/FilePath.cpp:1517; LOGGED FROM: bool rstudio::r::session::state::destroy(const rstudio::core::FilePath&) src/cpp/r/session/RSessionState.cpp:691
Error in .Internal(quit(save, status, runLast)) :
 Unable to quit (session cleanup failure)

Environment History Connections Tutorial

Import Dataset 436 MiB

R Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

Name	Size	Modified
_workflowr.yml	287 B	Aug 11, 2021, 10:55 AM
.gitattributes	148 B	Aug 11, 2021, 10:55 AM
.gitignore	174 B	Aug 17, 2021, 11:50 AM
.RData	38.6 MB	Aug 17, 2021, 10:54 PM
.Rhistory	303 B	Aug 26, 2021, 1:20 PM
.Rprofile	384 B	Aug 11, 2021, 10:55 AM
archive		
data		
Dockerfile_conflict	1.8 KB	Aug 26, 2021, 10:44 AM
Dockerfile.txt	3.1 KB	Aug 26, 2021, 1:14 PM
dockerTalk.md	61 B	Aug 26, 2021, 9:57 AM
getGeneNames.Rmd	3.3 KB	Aug 18, 2021, 10:02 PM
GettingRunning.md	1.1 KB	Aug 16, 2021, 10:03 PM
login.png	58.9 KB	Aug 26, 2021, 1:21 PM

SLAY ALL DAY



