

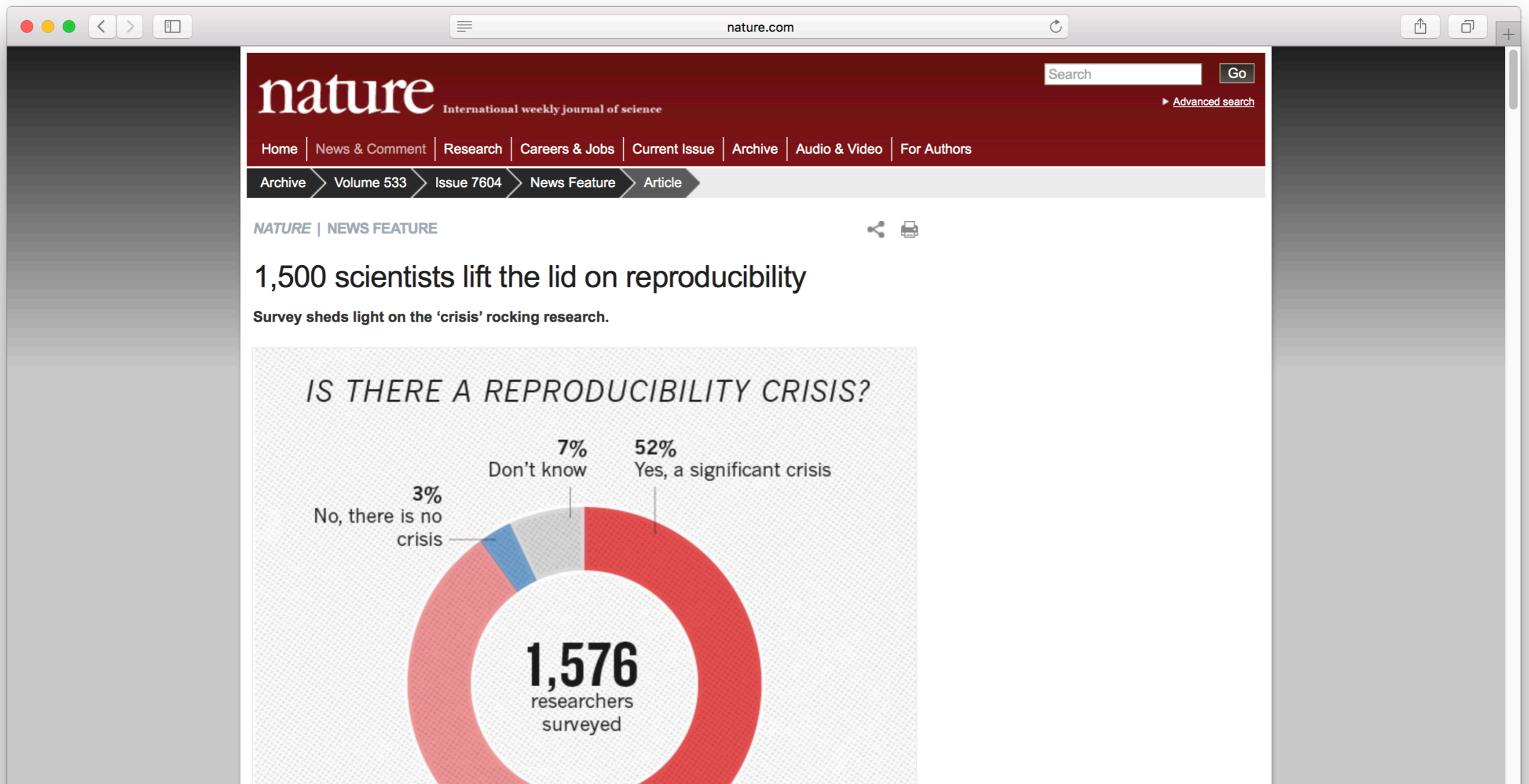
liftr: an R Package for Persistent Reproducible Research

JSM 2018

Nan Xiao
[@road2stat](#)

The Reproducibility Crisis

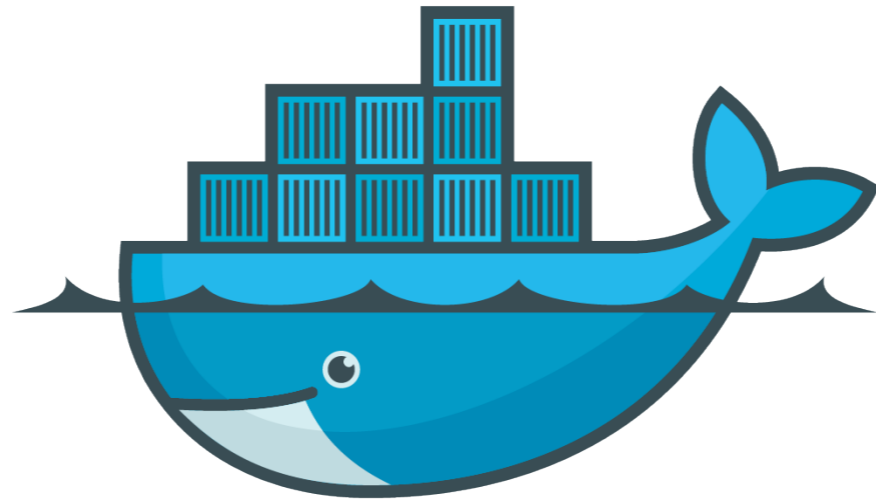
- Always a concern in both academia & industry.
- R Markdown + knitr pretty much saved the day.



The New Challenge

Even higher reproducibility for statistical computing:
regardless of *time* or *environment*.

Docker to the Rescue



docker

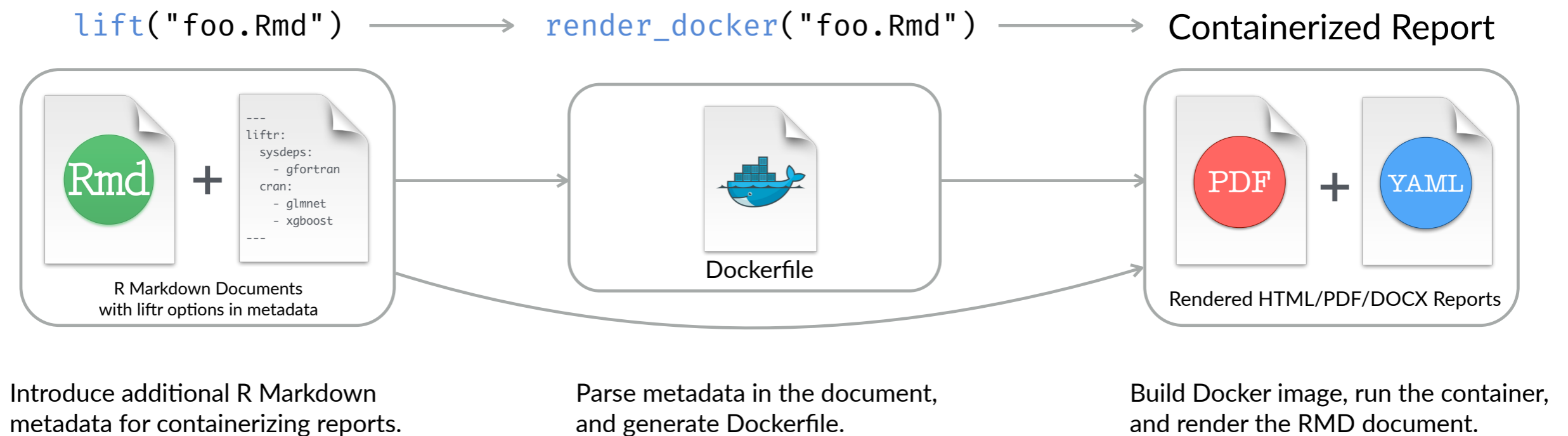
- Docker allows applications and their dependencies to be packaged into discrete runtime environments, called [containers](#).
- Apps packaged in this way can run from diverse infrastructures.

Our Solution: liftr

Persistent, OS-level reproducibility for R Markdown documents.



Containerize R Markdown Documents as Easy as 1-2-3



~/liftr - master - RStudio

liftr-tidyverse.Rmd

```
1 ---
2 title: "Explore tidyverse with liftr"
3 author: "Nan Xiao <me@nanx.me>"
4 date: "`r Sys.Date()`"
5 output:
6   rmarkdown::pdf_document:
7     toc: true
8     number_sections: true
9 liftr:
10   from: "rocker/tidyverse:latest"
11   maintainer: "Nan Xiao"
12   email: "me@nanx.me"
13   pandoc: false
14   texlive: true
15   cran:
16     - nycflights13
17 ---
18
19 \clearpage
20
21 # ggplot2
22
23 The example is from: https://github.com/tidyverse/ggplot2.
24
25 ```{r}
26 library("ggplot2")
27
```

9:7 # Explore tidyverse with liftr R Markdown

Console

CLIPR

- Output to clipboard
- Value to clipboard

LIFTR

- Containerize
- Render
- Prune Dangling
- Remove Image

PKGDOWN

- Build pkgdown

RHANDSONTABLE

- Edit a Data Frame

SEVENBRIDGES

Tool UI

Environment

Environment is empty

Files Plots Packages Help Viewer

- New Folder
- Delete
- Rename
- More

Home > liftr > inst > examples

	Name	Size	Modifie
↑	..		
📄	bioc-rnaseq.bib	49 KB	Apr 10
📄	bioc-rnaseq.Rmd	74.5 KB	Apr 14
📄	liftr-minimal.Rmd	1018 B	Apr 10
📄	liftr-tidyverse.Rmd	1.5 KB	Dec 12

IDE integration: RStudio addins (emojified): 📦 🎉 ✂️ 🗑️

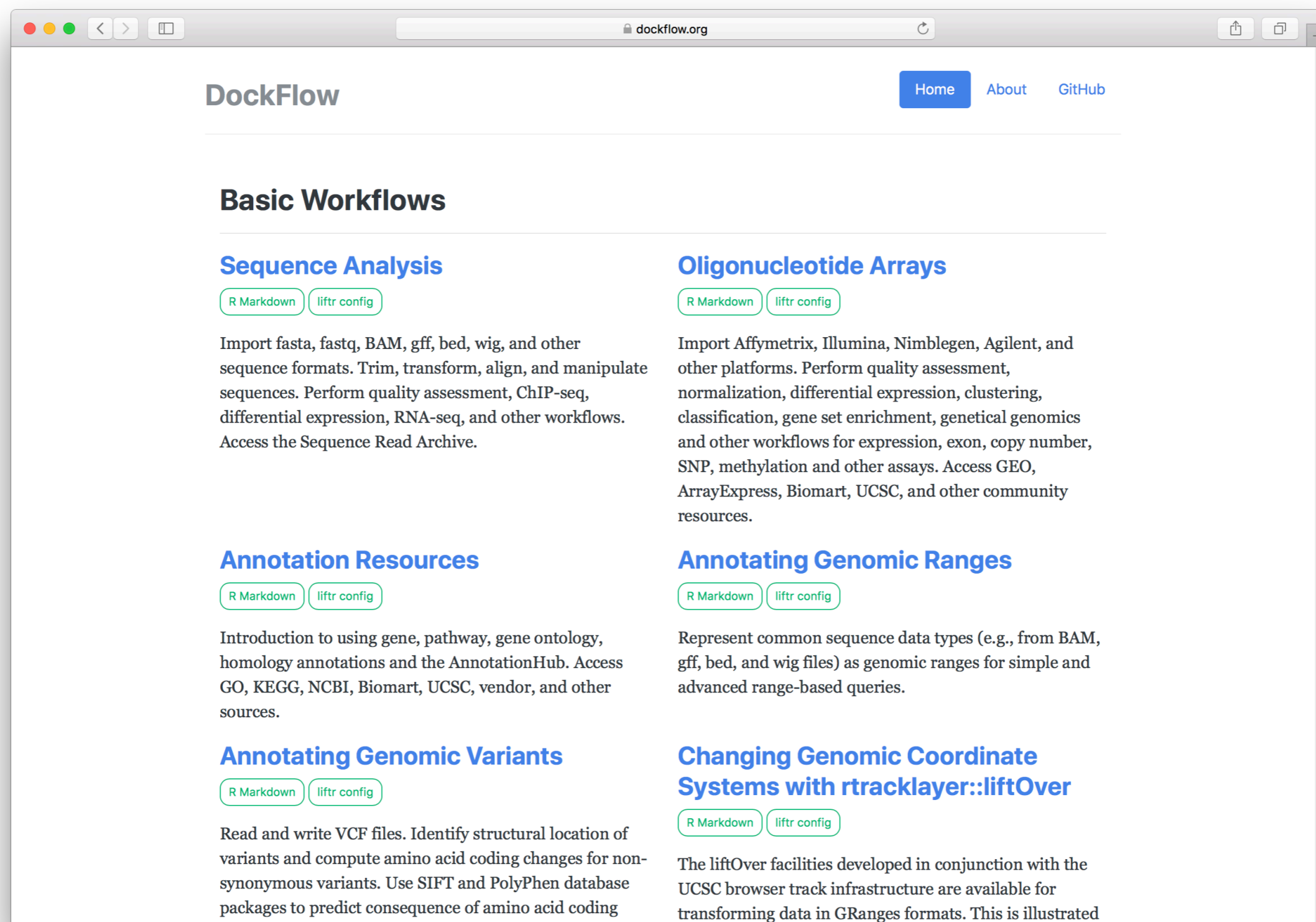
Philosophies

- **Continuous reproducibility.** Reproducible research should be a continuous process, instead of simply archiving code/data.
- **Document first.** R Markdown documents should be the center. Everything should be driven by documents, not packages.
- **Minimal footprint.** Connect R Markdown and Docker wisely, achieve more flexibility by doing less.

Applications

- **Individuals**: off-the-shelf solution for achieving persistent, environment-irrelevant reproducibility for data analysis.
- **Institutions**: key backend component for automated, large-scale report compilation/orchestration services.

dockflow.org



Easily containerized ~20 complex R Markdown workflows from Bioconductor.

Feature Roadmap

- Automatic inference of document **dependencies** (packrat)
- New **renderers** for bookdown, xaringan, and blogdown
- Improve CLI message **interface** (cli + crayon)
- Better Docker **integration** (reticulate + Docker API)

Acknowledgements

A special thanks to:

- Prof. Qing-Song Xu (Central South University)
- Prof. Matthew Stephens (University of Chicago)
- Dr. Tengfei Yin (Seven Bridges Genomics)
- Dr. Miao Zhu Li (Duke University)

for offering me all the freedom, advice, and encouragement to develop better software for the statistical and R community.

Thank you! Questions?

