Introduction to using Docker for reproducibility in R

R-Ladies Brisbane

Malindrie Dharmaratne
PhD Candidate | Mar Group
Australian Institute for Bioengineering and Nanotechnology
University of Queensland

@MalDharm2

June 17, 2021

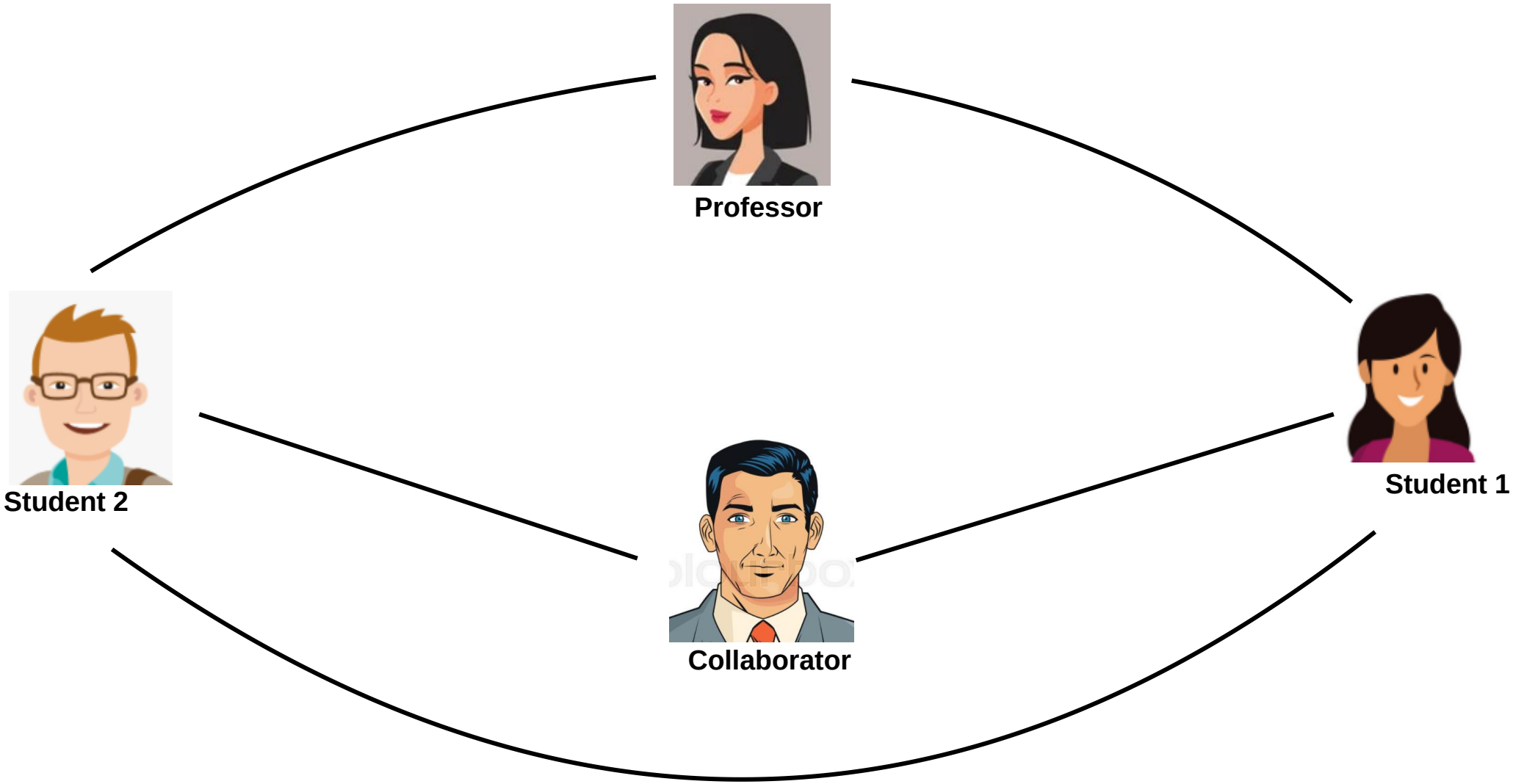
Agenda

- Introduction to Docker
- Docker with R
- Demo
- Q & A

Docker for R pre-requisites

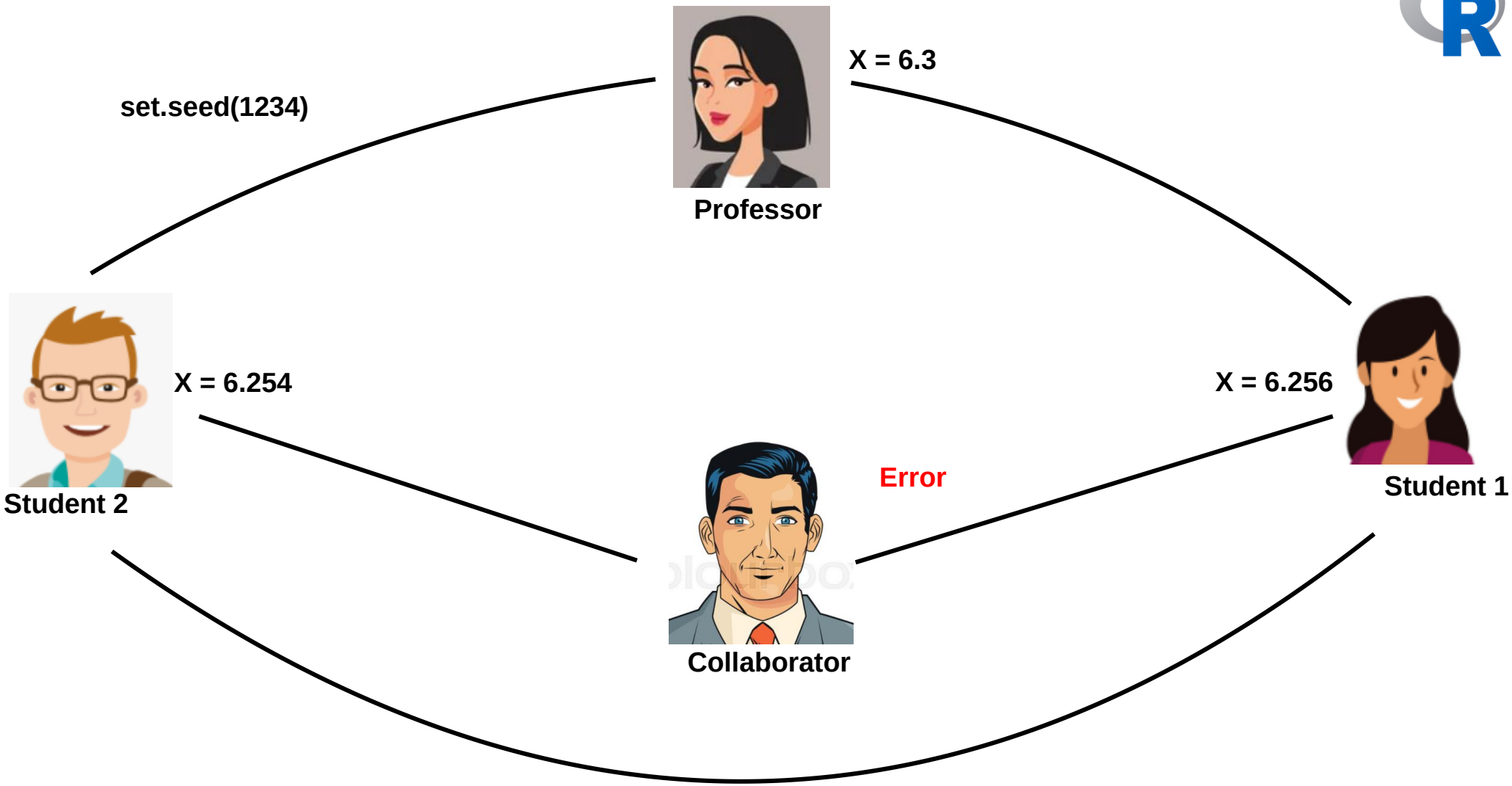
- R
- Docker
- Command line
- Bash/shell scripting
- Workshop materials - https://github.com/Malindrie/R-ladies_Brisbane

Reproducibility



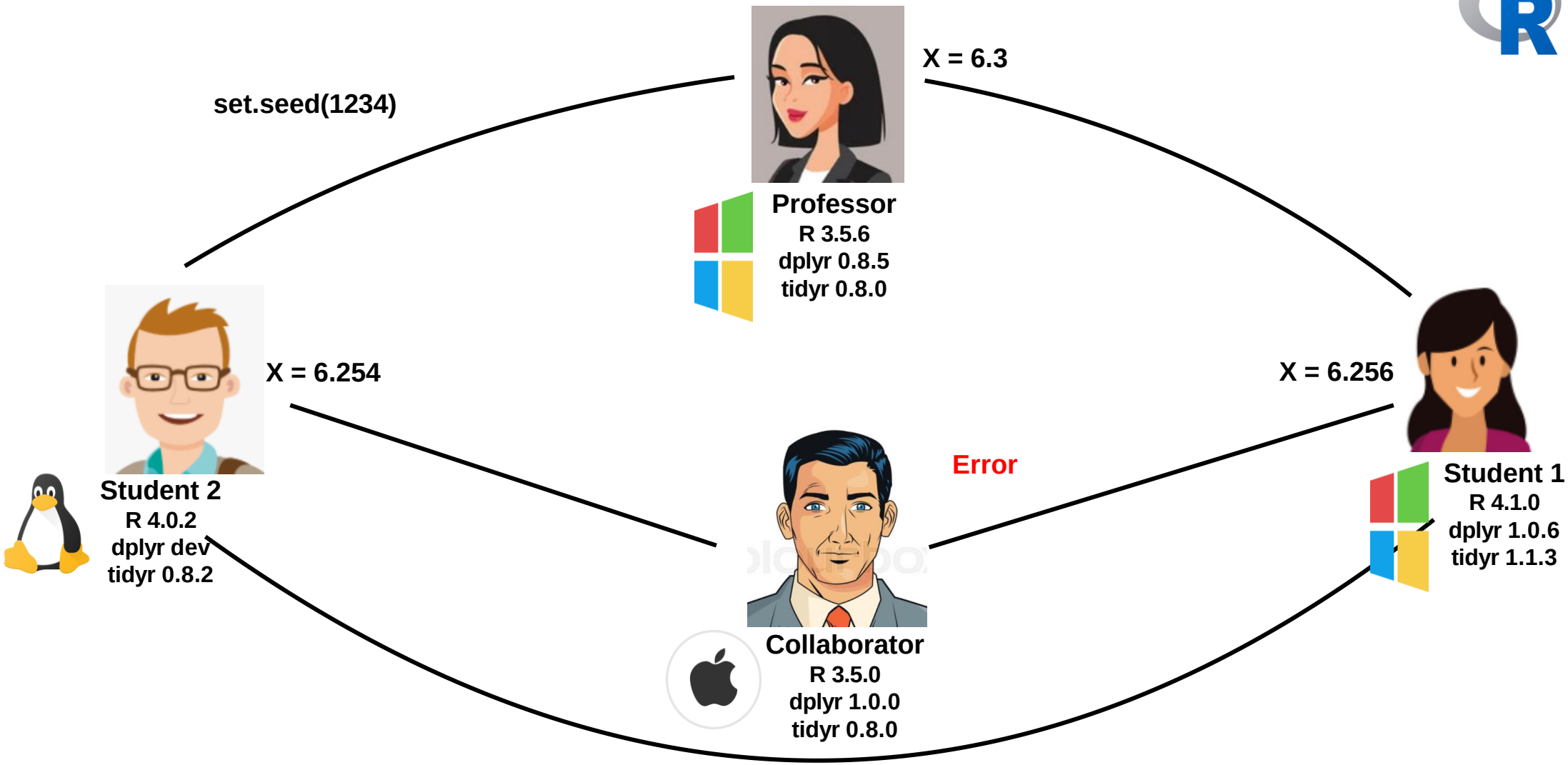
Source: Rami Krispin

Reproducibility



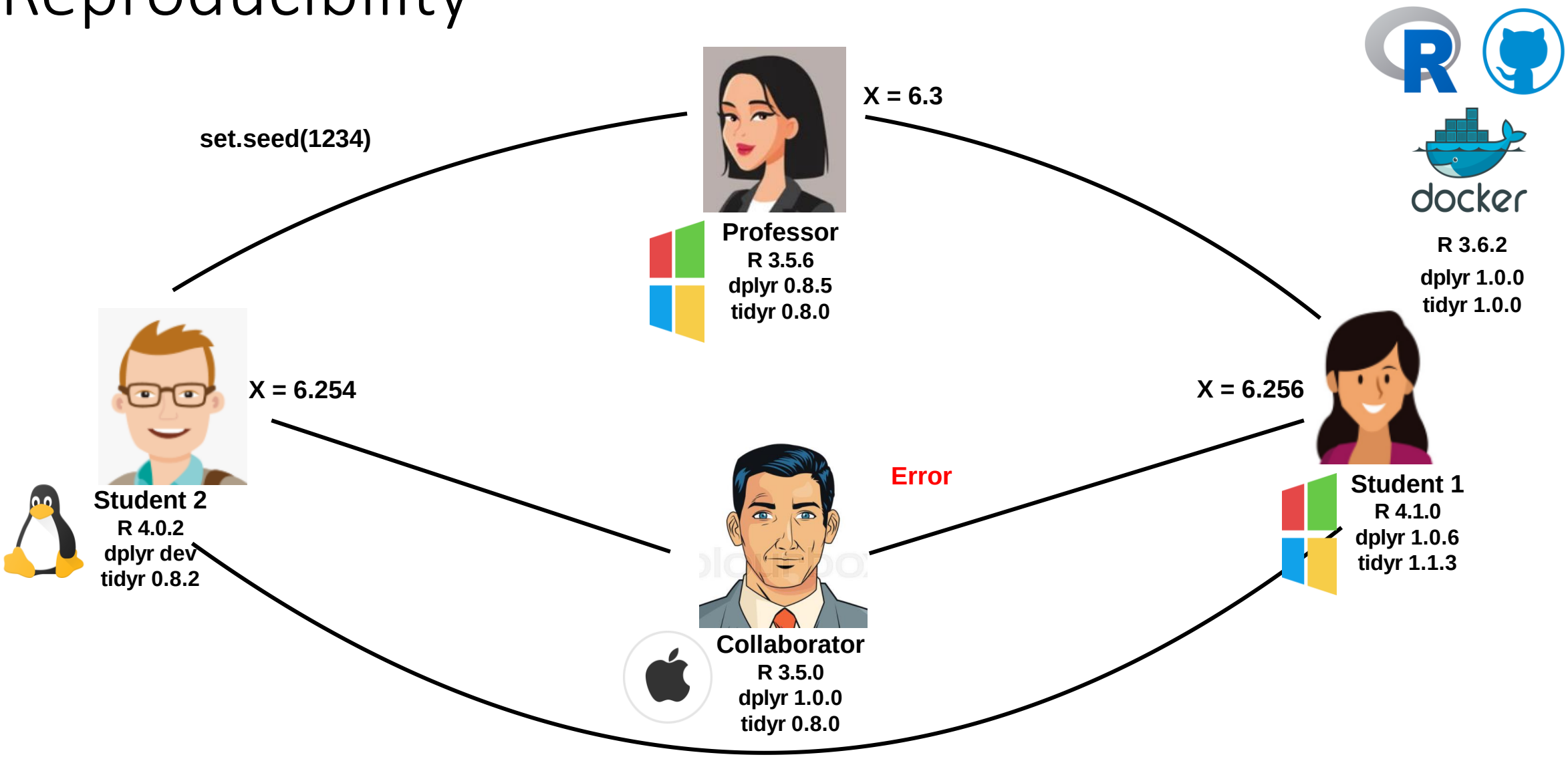
Source: Rami Krispin

Reproducibility



Source: Rami Krispin

Reproducibility



Source: Rami Krispin

Introduction to Docker

- A program that allows to manipulate multiple operating systems on your machine.
- Enable seamless shipment and deployment
- Open source and free (enterprise version also available)

What can I use Docker for?

- Continuous integration and continuous delivery (CI/CD) workflows
- Data automation
- Deploy Shiny apps
- R package deployment
- Reproducibility

Why use Docker?

- Fast, consistent delivery of your applications
- Responsive deployment and scaling
- Running more workloads on the same hardware

Installing Docker Engine

- Linux direct installation
- Mac & Windows – Docker Desktop

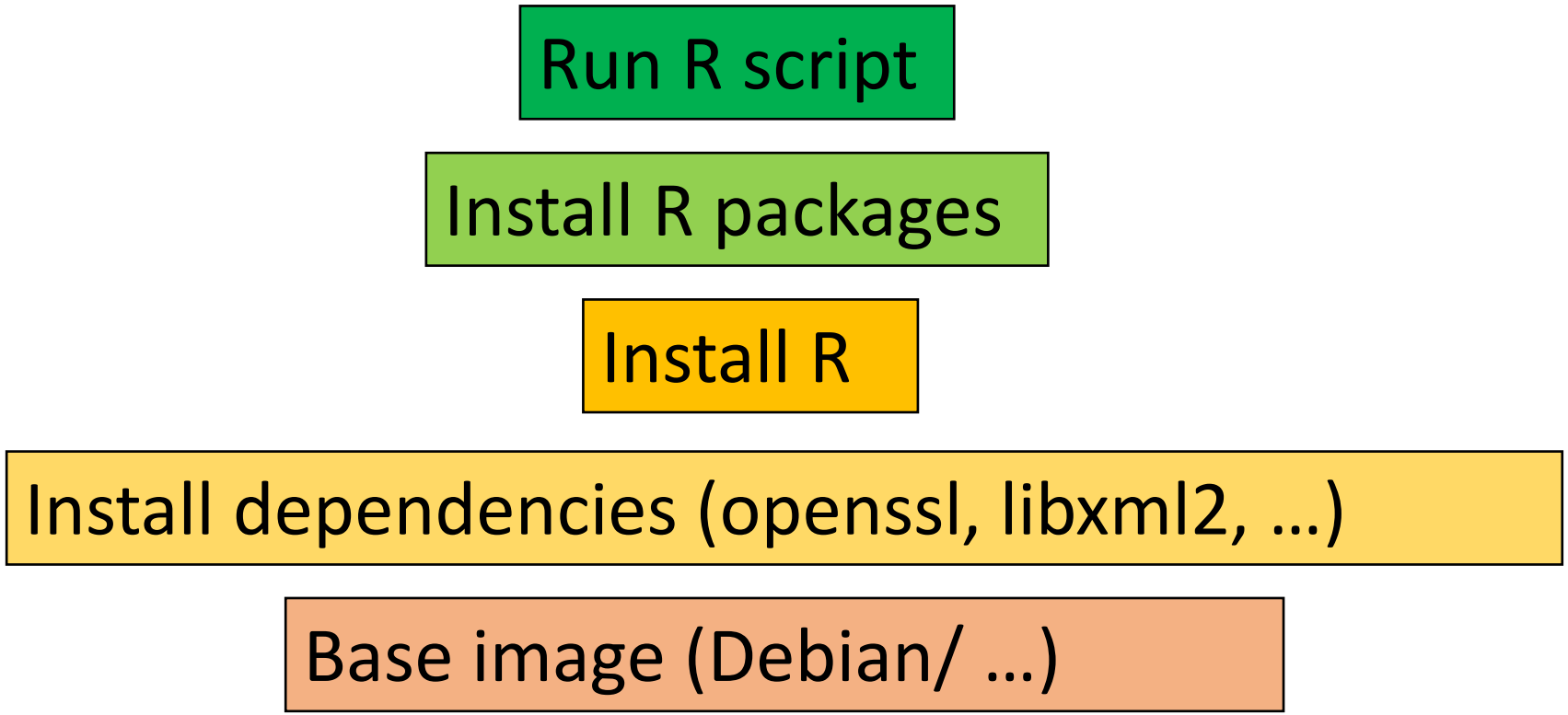
https://docs.docker.com/get-docker/

Docker objects

- Images - read-only template with instructions for creating a Docker container

- Containers - a runnable instance of an image

Docker R workflow

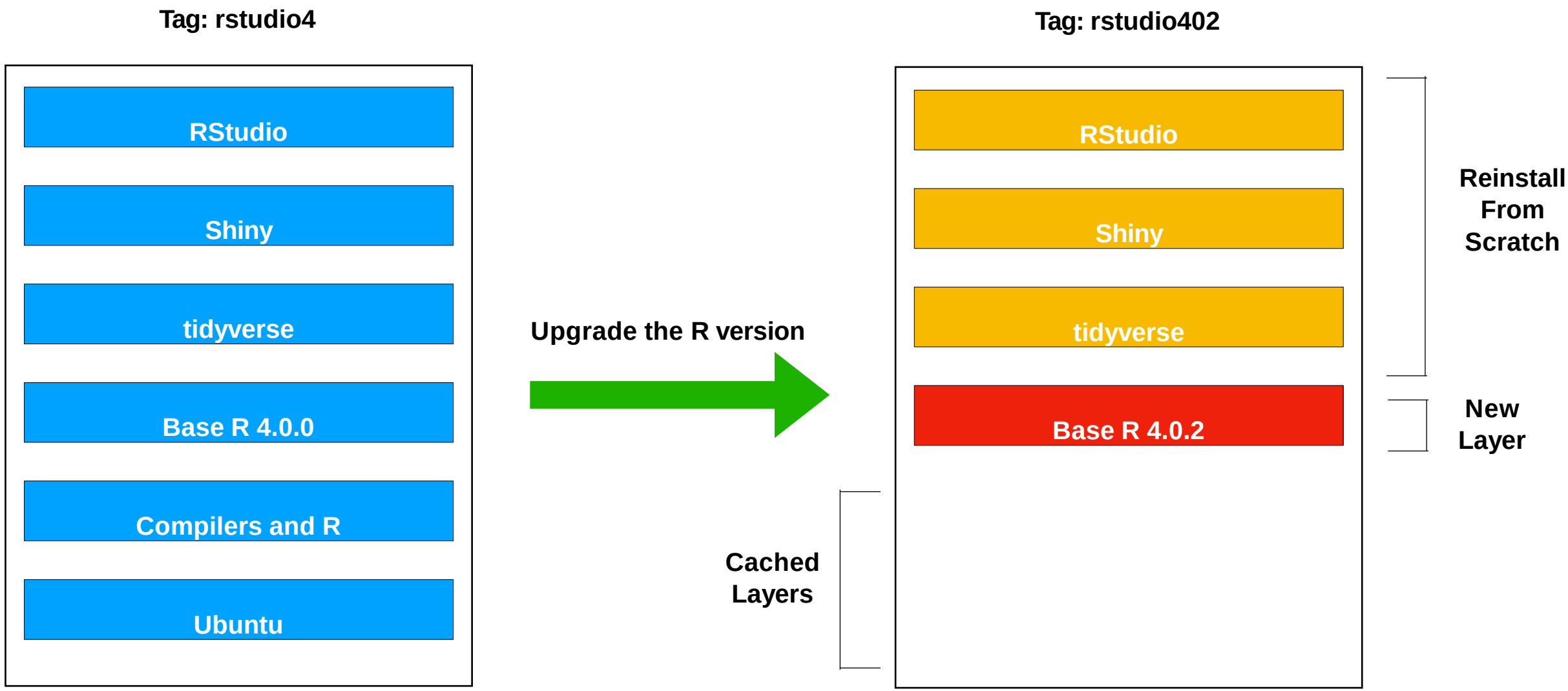


Docker R Workflow

Tag: rstudio4

Source: Rami Krispin

Docker R Workflow



Source: Rami Krispin

Docker R Approaches

Free Style – Build docker from scratch

Pros

- High customization
- Better optimization
- Learning experience

Cons

- Require advance knowledge
- More testing

Docker R Approaches

Built-in R Docker – Use out-of-box images

Pros

- Robust and tested
- Minimal effort
- Low barriers to entry

Cons

- Limited to specific versions
- Less optimization options

Dockerfile

- FROM
- LABEL
- RUN
- COPY
- ENV
- CMD

Command Line Tools

- sudo - substitute user do / super user do
- wget - web get
- curl - command-line tool for transferring data specified with URL syntax
- Rscript- calling R from the terminal
- bash- Unix shell and command language

Resources

- Workshop materials - https://github.com/Malindrie/R-ladies_Brisbane
- Docker documentation - https://docs.docker.com/get-started/overview/
- Docker Hub – https://hub.docker.com/
- Rocker - https://www.rocker-project.org/

THANK YOU