

#### Container-driven Reproducible Research Made Simple

#### Ronald Lencevičius

Academic Data Science Alliance Annual Meeting October 31st, 2024





2024 Academic Data Science Alliance Annual Meeting





• 3rd year PhD in Statistics & Applied Probability







- 3rd year PhD in Statistics & Applied Probability
- Background: Computer Science, Applied Mathematics







- 3rd year PhD in Statistics & Applied Probability
- Background: Computer Science, Applied Mathematics
- Interests: software development, reproducibility/explainability in ML





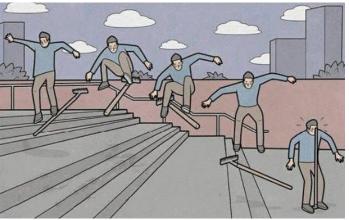






#### 🐱 A little about me...

- 3rd year PhD in Statistics & Applied Probability
- Background: Computer Science, Applied Mathematics
- Interests: software development, reproducibility/explainability in ML
- As with most things, tripped, stumbled, and fell into this project by good fortune during my 2nd year



2024 Academic Data Science Alliance Annual Meeting

#### 🐱 A little about me...

2024 Academic Data Science Alliance Annual Meeting

🦇 👻 🐻 🐟 🤠 🕅

- 3rd year PhD in Statistics & Applied Probability
- Background: Computer Science, Applied Mathematics
- Interests: software development, reproducibility/explainability in ML
- As with most things, tripped, stumbled, and fell into this project by good fortune during my 2nd year
- Since then I have built tooling, created documentation, and ran workshops/trainings for the department to spread the good words of containerized research





"An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures."

Buckheit & Donoho [1]







• An honorable task... one that requires effort and care in practice

- Simple truth: reproducibility is *hard*
- As a result, it is usually a burden or an afterthought

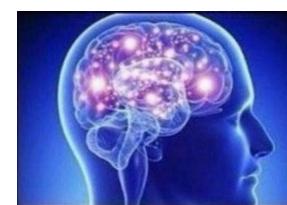




## Reproducibility Options – Good

• Naïve solution: README with dependency instructions

- e.g. sf package in  $R \rightarrow$  relies on multiple manually configured dependencies (geos, gdal, proj)
- Ok (at best): In depth test of reading comprehension
- Con: Assumes in depth knowledge on the part of maintainer and user





## Reproducibility Options – Better

• Naïve solution: README with dependency instructions

- e.g. sf package in  $R \rightarrow$  relies on multiple manually configured dependencies (geos, gdal, proj)
- Ok (at best): In depth test of reading comprehension
- Con: Assumes in depth knowledge on the part of maintainer and user
- Language Package Managers
  - e.g. requirements.txt, environment.yml, package-lock.json, etc.
  - Pro: Simple to use, usually built in with language or environment (Anaconda)
  - Con: Depends on underlying OS, not language agnostic, "works on mine, not on thine"





## Reproducibility Options – Complete

- Naïve solution: README with dependency instruction
  - e.g. sf package in R  $\rightarrow$  relies on multiple manually configured of
  - Ok (at best): In depth test of reading comprehension
  - Con: Assumes in depth knowledge on the part of maintainer a
- Language Package Managers



- e.g. requirements.txt, environment.yml, package-lock.json, etc.
- Pro: Simple to use, usually built in with language or environment (Anaconda)
- Con: Depend on underlying OS, not language agnostic, "works on mine, not on thine"
- Virtual Machines
  - Emulation/virtualization of a computer from hardware to software (computer in computer)
  - e.g. VirtualBox, VMWare
  - Pro: Portable via images, fully reproducible from ground up
  - Con: "If you wish to make an apple pie from scratch, you must first invent the universe."
    - ~ Carl Sagan

2024 Academic Data Science Alliance Annual Meeting

## Reproducibility Options – Just right…

- Containerization
  - Benefits of a VM, usage similar to a package manager
  - Access to full system resources minimum overhead
    - No hypervisor needed to manage system resources
  - Contains *just* the required packages and configuration (OS/language(s)/packages/tools)
    - Only abstracts away software applications
  - Each container is isolated from the system and from each other (in some cases, other users)
    - e.g. 3 containers for separate projects with different language versions and dependencies
  - Easily shareable for other people to spin up and use
  - Widely used in software engineering for the specific purpose of isolating applications and development environments





## Why use containers for research?

- Scenario 1:
  - Want two different versions of R
  - Non-trivial task will likely require you to build R from source or use renv
  - Solution: Can use RStudio professional or Anaconda
- Scenario 2:
  - Attempting to install dependencies for a research project via a package manager...
  - ...only to realize it will *only* run on a Unix system
  - Solution: Shell out for a Macbook or install Linux
- Scenario 3:
  - A new student joins your lab and needs to get ramped up on an existing project
  - Will need to replicate language, package, and tool configurations...
  - Solution: Either leave them to their own devices or sit down and take time ironing out their configuration given their unique system
- Impossible to address all these scenarios at once



## Why use containers for research?

- Reproducibility should be something we simply do everyday
- Be able to configure a wide range of languages, packages, and development tools
- At the same time research should be available wherever you take or share it
- More than that it should be simple to use, run, and interface with
- ...as a result, we should dive straight into containers!









## Containers for Reproducibility [2]

- Lightweight virtualization → Only requires a container engine (Docker/Podman/Apptainer) for the underlying hardware
- Relies on a "Dockerfile" to completely define a working environment  $\rightarrow$  easily portable between people and systems
- Can create arbitrary operating system, software, and language configurations





#### Containers for Reproducibility [2] (Example)

Dock	kerfile ×				
devcon	tainer > 👉 Dockerfile				
1	FROM ubuntu:latest				
2					
3	RUN apt-get update &	& apt-get install -y \			
4	g++ \	ronaldas .devcontainer docker build -t my_awesome_container . &&			
5	git	<pre>&gt; docker run -it my_awesome_container</pre>			
6	15.00	STEP 1/4: FROM ubuntu:latest			
7	WORKDIR /workspace	STEP 2/4: RUN apt-get update && apt-get install -y g++ git			
8		> Using cache cdc10c83e5ae34f5220b5f9b7b1d5e22ed341f67a760de53b56bc6381019946			
9	CMD ["bash"]	> cdc10c83e5ae			
10		STEP 3/4: WORKDIR /workspace > e3f5f9ea162d			
		STEP 4/4: CMD ["bash"]			
		COMMIT my_awesome_container			
		> 4732f67f3271			
		Successfully tagged localhost/my_awesome_container:latest			
		4732f67f3271ab2dedf824632b5d71eeda18090c9b215abd00c273456507ed7b			
	uis Data Caisana Alliana Annual I	root@7a3595cbcbd2:/workspace#			

2024 Academic Data Science Alliance Annual Meeting

## 🔥 Challenges – Docker CLI

- As you can see, we are interacting with terminal which is not ideal
- A more complex configuration using a Jupyter container:

docker run -it --rm -p 10000:8888 -v "\${PWD}":/home/jovyan/work
quay.io/jupyter/r-notebook:2024-10-07

- No easy way to manage containers other than learning yet another CLI tool
- Cumbersome! (and not even the most complicated configuration)
- Writing the actual configs can get complicated as well (more on that later)



# Development Containers with Visual Studio Code

- "Development containers" (devcontainers) [3] simplify the process of interacting with and managing your containers
- Adds one additional spec file: devcontainer.json in it...
  - ....we can specify either a pullable image or a Dockerfile
  - ...configure build and run arguments
  - ...add development environment customizations through extensions
- This interfaces with Visual Studio Code (VS Code), a powerful integrated development environment that helps manage our containers, remote servers, and GitHub repositories through convenient extensions
- Our project software *and* development tools are now made easily reproducible, shareable, and (most importantly) interactive



👉 Docl	kerfile ×		
.devcor 1 2 3	ntainer > 👉 Dockerfile FROM ubuntu:latest RUN apt-get update &	apt-get install −y \	
4 5 7 8 9	<pre>g++ \ git WORKDIR /workspace CMD ["bash"]</pre>	<pre>ronaldas .devcontainer docker build -t my_awe</pre>	come_container g++ git
		STEP 4/4: CMD ["bash"] COMMIT my_awesome_container > 4732f67f3271 Successfully tagged localhost/my_awesome_container 4732f67f3271ab2dedf824632b5d71eeda18090c9b215abd00 root@7a3595cbcbd2:/workspace#	

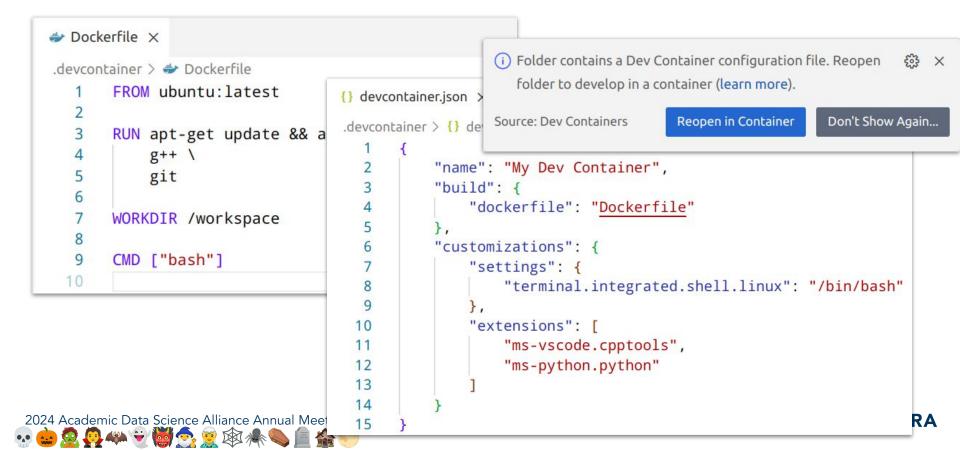
2024 Academic Data Science Alliance Annual Meeting

#### Devcontainers with VS Code (Example)

```
→ Dockerfile ×

    .devcontainer > 🗇 Dockerfile
            FROM ubuntu: latest
                                         {} devcontainer.json ×
       2
                                         .devcontainer > {} devcontainer.json > ...
       3
            RUN apt-get update && a
       4
                 g++ \
                                                     "name": "My Dev Container",
                                            2
       5
                git
                                            3
                                                     "build": {
       6
                                            4
                                                          "dockerfile": "Dockerfile"
       7
            WORKDIR /workspace
                                            5
       8
                                            6
                                                     "customizations": {
       9
            CMD ["bash"]
                                                          "settings": {
                                            7
      10
                                                              "terminal.integrated.shell.linux": "/bin/bash"
                                            89
                                           10
                                                          "extensions": [
                                           11
                                                              "ms-vscode.cpptools",
                                           12
                                                              "ms-python.python"
                                           13
                                           14
 2024 Academic Data Science Alliance Annual Meet
                                                                                                                  RA
                                           15
💀 🎃 🧟 🗛 🦇 👻 👹 🗙 🕱 🦚
```

## Devcontainers with VS Code (Example)

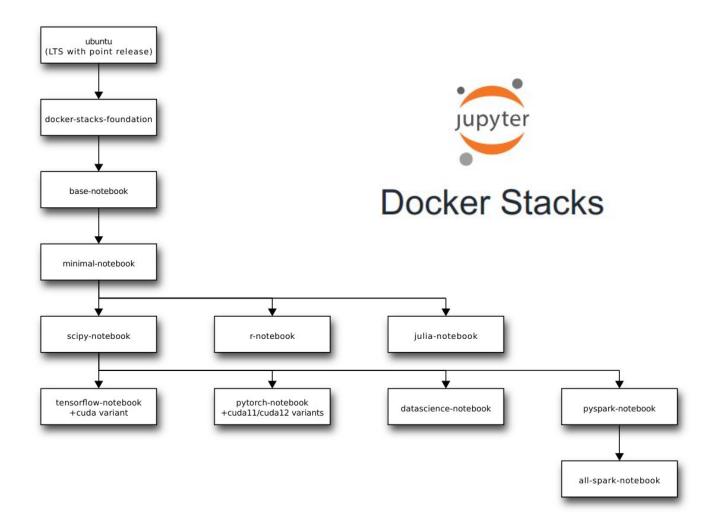


## Devcontainers with VS Code (Jupyter Stacks)

- Previous container isn't too exciting so let's check out one with a built in RStudio server
- To do this we will use Jupyter Docker Stacks [4], a set of Docker images containing Jupyter apps and tools
- If you or your institution uses JupyterHub instances for students then you may have been using this the entire time without realizing it!







```
FROM quay.io/jupyter/r-notebook:r-4.3.1
 2
3
     ENV GITHUB CLI VERSION 2.30.0
 4
     ENV R_STUDIO_VERSION 2023.12.1-402
 6
    USER root
 8
     RUN apt-get update && \
 9
         apt-get install -y --no-install-recommends \
10
                                                                                                                                           10
             1modern \
                                                                                                                                           11
11
             file \
                                                                                                                                           12
12
             curl \
                                                                                                                                           13
13
             g++ \
14
             tmux \
                                                                                                                                           14
                                                                                                                                           15
15
             psmisc \
16
                                                                                                                                           16
             lsb-release \
17
                                                                                                                                           17
             libssl-dev \
                                                                                                                                           18
18
             libclang-dev '
19
             libpg5 \
                                                                                                                                           19
20
             libtiff-dev \
                                                                                                                                           20
21
             88 1
                                                                                                                                           21
                                                                                                                                           22
22
         apt-get clean -v && \
23
         rm -rf /var/lib/apt/lists/* /tmp/library-scripts
                                                                                                                                           23
24
                                                                                                                                           24
     RUN R -e "dotR <- file.path(Sys.getenv('HOME'), '.R'); if(!file.exists(dotR)){ dir.create(dotR) }; Makevars <- file.path(dotR, 'M
                                                                                                                                          25
25
26
     RUN R -e "dotRprofile <- file.path(Sys.getenv('HOME'), '.Rprofile'); if(!file.exists(dotRprofile)){ file.create(dotRprofile) }; d
                                                                                                                                          26
27
                                                                                                                                           27
28
     RUN wget -q https://download2.rstudio.org/server/jammy/amd64/rstudio-server-${R_STUDIO_VERSION}-amd64.deb && \
                                                                                                                                           28
                                                                                                                                           29
29
         apt-get install -yg --no-install-recommends ./rstudio*.deb && \
                                                                                                                                           30
30
         rm -f ./rstudio*.deb && \
31
         apt-get clean && \
                                                                                                                                           31
32
         chmod 777 /var/run/rstudio-server && \
                                                                                                                                           32
33
                                                                                                                                           33
         chmod +t /var/run/rstudio-server
34
                                                                                                                                           34
35
                                                                                                                                           35
    USER ${NB USER}
                                                                                                                                           36
36
37
     RUN mamba install -y -c conda-forge --freeze-installed \
                                                                                                                                           37
                                                                                                                                           38
38
             iupvter-server-proxv=4.1.0 \
                                                                                                                                           39
39
             jupyter-rsession-proxy=2.2.0 \
40
             22 \
                                                                                                                                           40
         mamba clean --all
                                                                                                                                           41
41
                                                                                                                                           42
42
43
     RUN pip install \
                                                                                                                                           43
44
             nbgitpuller \radian==0.6.11 \
45
             22 1
46
         jupyter labextension enable nbgitpuller
47
48
     RUN R -q -e 'remotes::install version("markdown", version="1.12", repos="cloud.r-project.org")' && \
49
         R -q -e 'remotes::install version("languageserver", version="0.3.16", repos="cloud.r-project.org")' && \
         R -q -e 'remotes::install_version("httpgd", version="2.0.1", repos="cloud.r-project.org")' && \
50
51
         R -q -e 'remotes::install github("ManuelHentschel/vscDebugger")' && \
52
         R -q -e 'remotes::install_version("syuzhet", version="1.0.7", repos="cloud.r-project.org")' && \
53
         echo
54
55
     RUN wget https://github.com/cli/cli/releases/download/v${GITHUB_CLI_VERSION}/gh_${GITHUB_CLI_VERSION}_linux_amd64.tar.gz -0 - | \
56
         tar xvzf - -C /opt/conda/bin gh ${GITHUB CLI VERSION} linux amd64/bin/gh --strip-components=2
57
58
     RUN echo "echo \"Jupyter server token: \$(jupyter server list 2>&1 | grep -oP '(?<=token=)[[:alnum:]]*')\"" > ${HOME}/.get-jupyter
59
         echo "sh \${HOME}/.get-jupyter-url.sh" >> ${HOME}/.bashrc
```

```
"name": "my-awesome-project",
"build": {
    "dockerfile": "Dockerfile".
    "options": ["--format=docker"]
},
"updateRemoteUserUID": false,
"overrideCommand": false,
"shutdownAction": "none",
"workspaceMount": "source=${localWorkspaceFolder},target=/home/jovyan/work,type=bind,z",
"workspaceFolder": "/home/jovyan/work",
"runArgs": [
    "--name=my-awesome-project", // needs to be unique
    "--hostname=my-awesome-project container" // needs to be unique
],
"forwardPorts": [8888].
"portsAttributes": {
    "8888": {
        "label": "Jupyterlab".
        "onAutoForward": "ignore"
},
"customizations": {
    "vscode":
        "settings": {
            "r.rterm.linux": "/opt/conda/bin/radian",
            "r.bracketedPaste": true,
            "r.plot.useHttpgd": true
        },
        "extensions": [
                                            // for R
            "reditorsupport.r".
            "RDebugger.r-debugger",
                                            // for R
            "ms-vscode.live-server".
            "analytic-signal.preview-pdf"
```

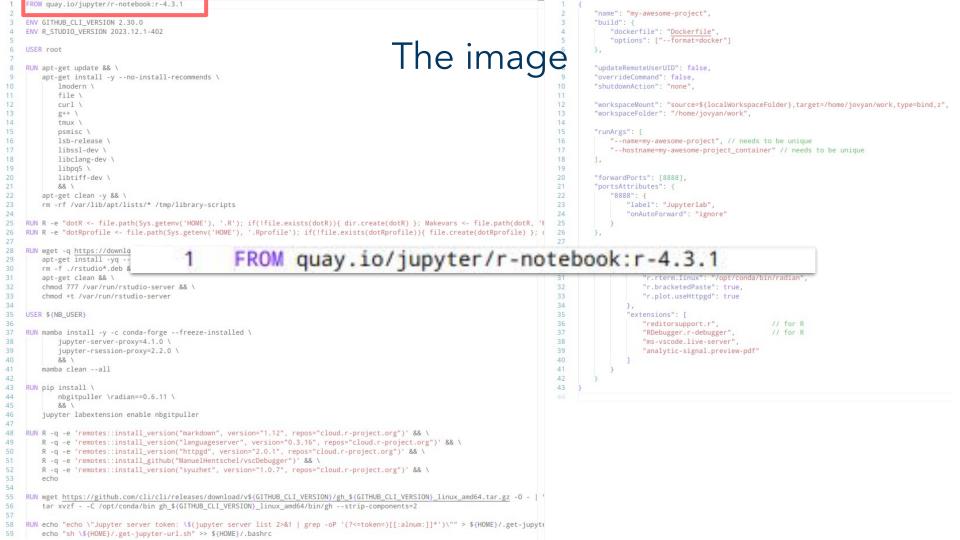
3

6

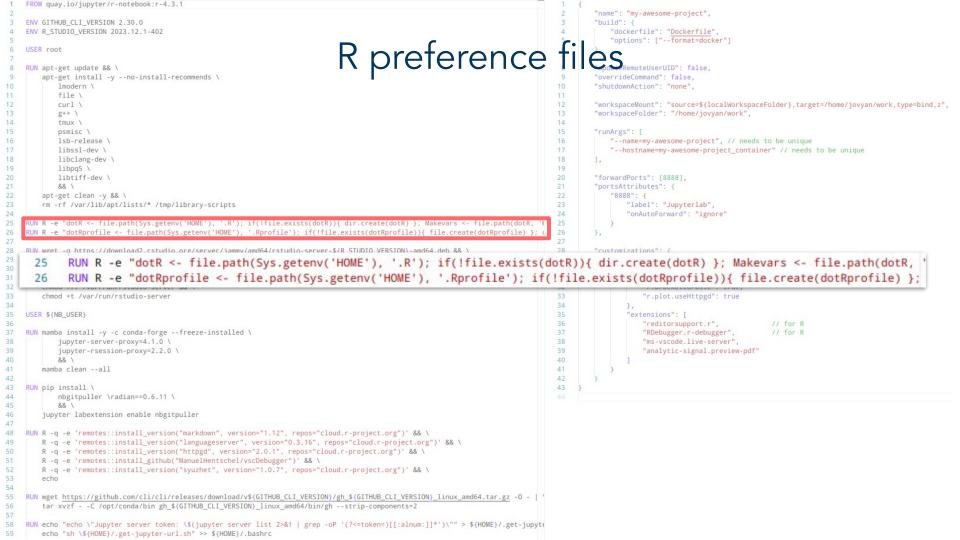
7

8

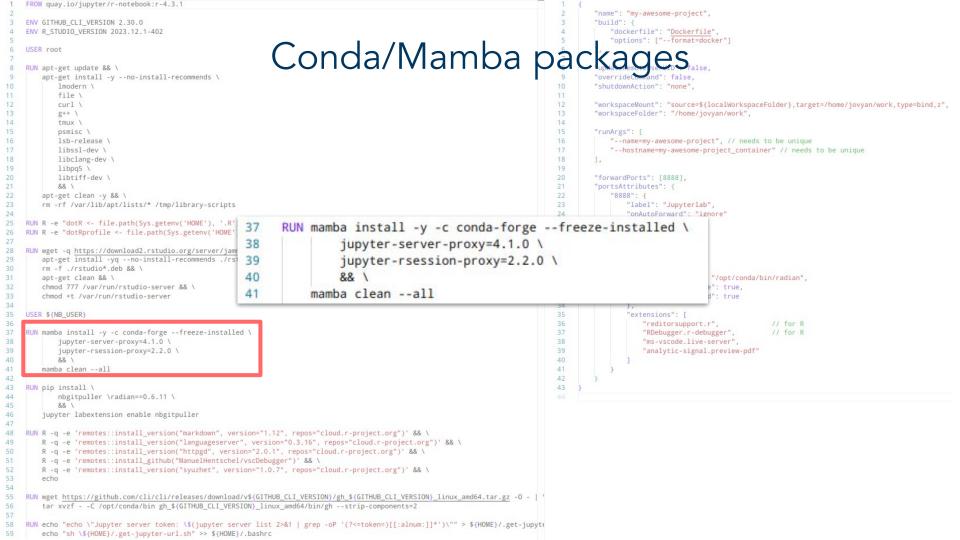
9

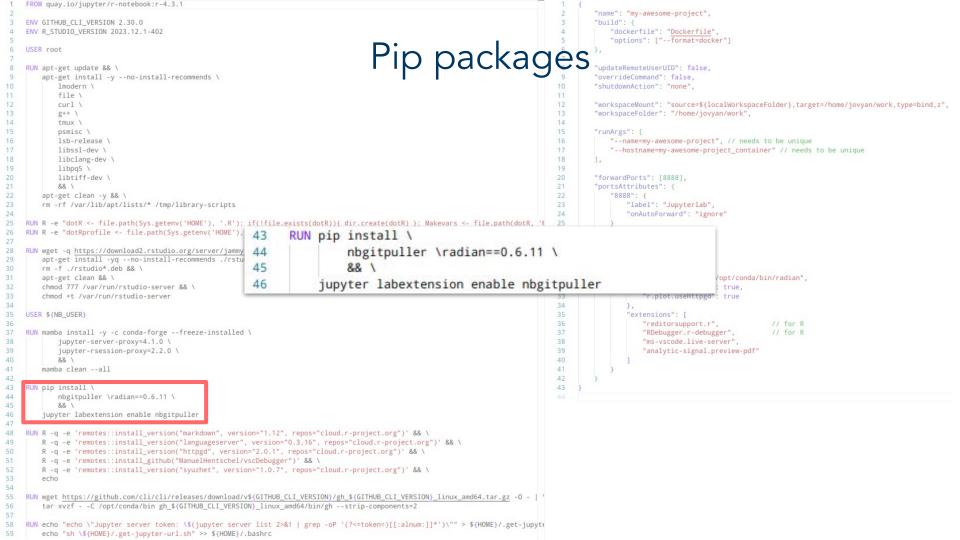


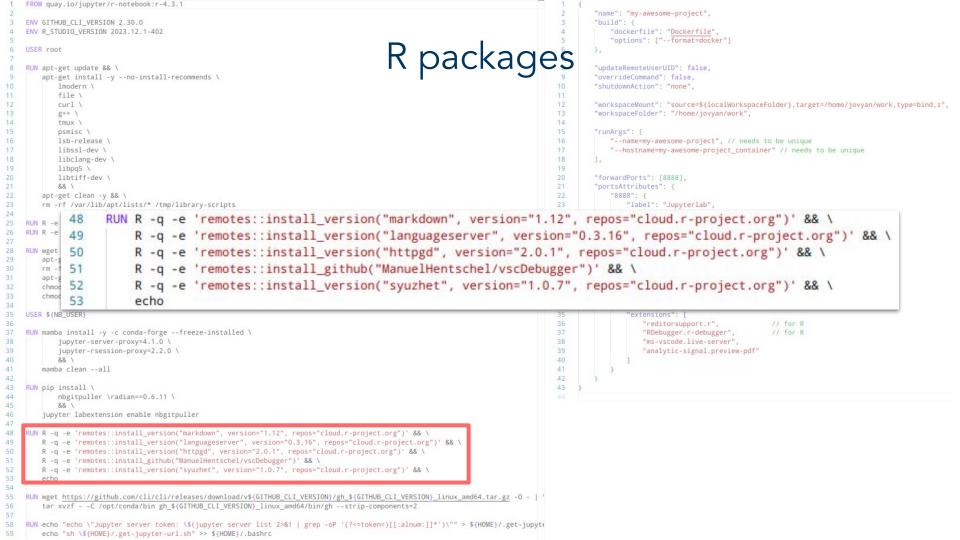
```
FROM quay.io/jupyter/r-notebook:r-4.3.1
                                                                                                                                     "name": "my-awesome-project",
    ENV GITHUB CLI VERSION 2.30.0
                                                                                                                                     "build": {
    ENV R_STUDIO_VERSION 2023.12.1-402
                                                                                                                                        "dockerfile": "Dockerfile".
                                                                                                                                        "options": ["--format=docker"]
                                                                          System Packages.
6
   USER root
8
     RUN apt-get update && \
        apt-get install -y --no-install-recommends \
                                                                                                                                     "overrideCommand": false,
                                                                                                                                     "shutdownAction": "none",
            1modern \
           file \
           curl \
                                                                                                                                     "workspaceMount": "source=${localWorkspaceFolder},target=/home/jovyan/work,type=bind,z",
                                                                                                                                     "workspaceFolder": "/home/jovyan/work",
           g++ \
14
                                                                                                                            14
           tmux \
           psmisc \
                                                                                                                                    "runArgs": [
16
           lsb-release \
                                                                                                                                                                     needs to be unique
                                                 8
                                                        RUN apt-get update && \
           libssl-dev \
                                                                                                                                                                    container" // needs to be unique
18
           libclang-dev \
                                                               apt-get install -y --no-install-recommends \
                                                 9
19
           libtiff-dev \
                                                                     1modern \
                                                10
           1 28
        apt-get clean -v && \
                                                11
                                                                     file \
        rm -rf /var/lib/apt/lists/* /tmp/librarv-
24
                                                12
                                                                     curl \
    RUN R -e "dotR <- file.path(Sys.getenv('HOME'
                                               13
                                                                     g++ \
26
    RUN R -e "dotRprofile <- file.path(Sys.getenv(
                                                14
                                                                     tmux \
28
   RUN wget -q https://download2.rstudio.org/serv
29
        apt-get install -yq --no-install-recommend
                                                15
                                                                     psmisc \
30
        rm -f ./rstudio*.deb && \
31
        apt-get clean && \
                                                                     lsb-release \
                                                                                                                                                                    conda/bin/radian".
                                                16
32
        chmod 777 /var/run/rstudio-server && \
                                                                     libssl-dev \
                                                17
        chmod +t /var/run/rstudio-server
34
                                                18
                                                                     libclang-dev \
35
   USER ${NB USER}
36
                                                                                                                                                                            // for R
                                                                     libpg5 \
                                               19
    RUN mamba install -y -c conda-forge --freeze-i
                                                                                                                                                                            // for R
38
           jupyter-server-proxy=4.1.0 \
                                                                     libtiff-dev \
                                                20
           jupyter-rsession-proxy=2.2.0 \
                                                                                                                                                                    w-pdf"
40
           22 \
                                                21
                                                                     1 88
41
        mamba clean --all
                                                22
                                                               apt-get clean -y && \
42
43
    RUN pip install \
                                                               rm -rf /var/lib/apt/lists/* /tmp/library-scripts
                                                23
44
           nbgitpuller \radian==0.6.11 \
45
           22 1
46
        jupyter labextension enable nbgitpuller
47
    RUN R -q -e 'remotes::install_version("markdown", version="1.12", repos="cloud.r-project.org")' && \
48
49
        R -q -e 'remotes::install version("languageserver", version="0.3.16", repos="cloud.r-project.org")' && \
50
        R -q -e 'remotes::install version("httpgd", version="2.0.1", repos="cloud.r-project.org")' && \
        R -q -e 'remotes::install_github("ManuelHentschel/vscDebugger")' && \
        R -q -e 'remotes::install_version("syuzhet", version="1.0.7", repos="cloud.r-project.org")' && \
        echo
54
    RUN wget https://github.com/cli/cli/releases/download/v${GITHUB_CLI_VERSION}/gh_${GITHUB_CLI_VERSION}_linux_amd64.tar.gz -0 - | \
        tar xvzf - -C /opt/conda/bin gh ${GITHUB CLI VERSION} linux amd64/bin/gh --strip-components=2
   RUN echo "echo \"Jupyter server token: \$(jupyter server list 2>&1 | grep -oP '(?<=token=)[[:alnum:]]*')\"" > ${HOME}/.get-jupyte
58
59
        echo "sh \${HOME}/.get-jupyter-url.sh" >> ${HOME}/.bashrc
```



```
FROM quay.io/jupyter/r-notebook:r-4.3.1
                                                                                                                                      "name": "my-awesome-project",
    ENV GITHUB CLI VERSION 2.30.0
                                                                                                                                      "build": {
                                                                                                                                          "dockerfile": "Dockerfile".
    ENV R_STUDIO_VERSION 2023.12.1-402
                                                                                                                                          "options": ["--format=docker"]
                                                                                             RStudio
6
   USER root
    RUN apt-get update && \
                                                                                                                                      "updateRemoteUserUID": false,
        apt-get install -y --no-install-recommends \
                                                                                                                                      "overrideCommand": false,
                                                                                                                                      "shutdownAction": "none",
            1modern \
            file \
            curl \
                                                                                                                                      "workspaceMount": "source=${localWorkspaceFolder},target=/home/jovyan/work,type=bind,z",
            g++ \
                                                                                                                                      "workspaceFolder": "/home/jovyan/work",
14
            tmux \
                                                                                                                              14
            psmisc \
                                                                                                                                      "runArgs": [
                                                                                                                              16
                                                                                                                                         "--name=my-awesome-project", // needs to be unique
            lsb-release \
            libssl-dev \
                                                                                                                                          "--hostname=my-awesome-project container" // needs to be unique
18
            libclang-dev \
                                                                                                                              18
19
            libpg5 \
                                                                                                                              19
            libtiff-dev \
                                                                                                                                      "forwardPorts": [8888].
            1 28
                                                                                                                                      "portsAttributes": {
        apt-get clean -v && \
                                                                                                                                          "8888": {
        rm -rf /var/lih/ant/lists/* /tmn/lihrarv-scrints
                                                                                                                                             "lahel" ... "lunvterlah"
24
                    RUN wget -q https://download2.rstudio.org/server/jammy/amd64/rstudio-server-${R_STUDIO_VERSION}-amd64.deb && \
            28
25
    RUN R
26
    RUN R
            29
                           apt-get install -yq --no-install-recommends ./rstudio*.deb && \
28
    RUN wg
            30
                           rm -f ./rstudio*.deb && \
29
        apt
30
            31
                           apt-get clean && \
        rm
        apt
                           chmod 777 /var/run/rstudio-server && \
            32
        chr
        chr
                           chmod +t /var/run/rstudio-server
            33
34
36
                                                                                                                                                 "reditorsupport.r".
                                                                                                                                                                               // for R
    RUN mamba install -y -c conda-forge --freeze-installed \
                                                                                                                                                 "RDebugger.r-debugger",
                                                                                                                                                                              // for R
                                                                                                                              38
                                                                                                                                                 "ms-vscode.live-server".
38
            iupvter-server-proxv=4.1.0 \
            jupyter-rsession-proxy=2.2.0 \
                                                                                                                                                 "analytic-signal.preview-pdf"
40
            1 28
                                                                                                                              40
        mamba clean --all
                                                                                                                              41
41
                                                                                                                              42
42
43
    RUN pip install \
                                                                                                                              43
44
            nbgitpuller \radian==0.6.11 \
45
            22 1
46
        jupyter labextension enable nbgitpuller
47
    RUN R -q -e 'remotes::install_version("markdown", version="1.12", repos="cloud.r-project.org")' && \
48
49
        R -q -e 'remotes::install version("languageserver", version="0.3.16", repos="cloud.r-project.org")' && \
        R -q -e 'remotes::install version("httpgd", version="2.0.1", repos="cloud.r-project.org")' && \
        R -q -e 'remotes::install_github("ManuelHentschel/vscDebugger")' && \
        R -q -e 'remotes::install_version("syuzhet", version="1.0.7", repos="cloud.r-project.org")' && \
        echo
54
    RUN wget https://github.com/cli/cli/releases/download/v${GITHUB_CLI_VERSION}/gh_${GITHUB_CLI_VERSION}_linux_amd64.tar.gz -0 - | \
        tar xvzf - -C /opt/conda/bin gh_${GITHUB_CLI_VERSION}_linux_amd64/bin/gh --strip-components=2
58
    RUN echo "echo \"Jupyter server token: \$(jupyter server list 2>&1 | grep -oP '(?<=token=)[[:alnum:]]*')\"" > ${HOME}/.get-jupyte
59
        echo "sh \${HOME}/.get-jupyter-url.sh" >> ${HOME}/.bashrc
```







1	FROM quay.io/jupyter/r-notebook:r-4.3.1	1	
2 3 4 5	ENV GITHUB_CLI_VERSION 2.30.0 ENV R_STUDIO_VERSION 2023.12.1-402	2 3 4 5	<pre>"name": "my-awesome-project", "build": {     "dockerfile": "Dockerfile",     "options": ["format=docker"]</pre>
6	USER root GitHub C	6	
7		7	
8	RUN apt-get update && \	-8	"updateRemoteUserUID": false,
9	apt-get install -yno-install-recommends \	9	"overrideCommand": false,
10	lmodern \	10	"shutdownAction": "none",
11	file \	11	
12	curl \	12	<pre>"workspaceMount": "source=\${localWorkspaceFolder},target=/home/jovyan/work,type=bind,z",</pre>
13	g++ \	13	"workspaceFolder": "/home/jovyan/work",
14	tmux \	14	
15	psmisc \	15	"runArgs": [
16	lsb-release \	16	"name=my-awesome-project", // needs to be unique
17	libssl-dev \	17	"hostname=my-awesome-project_container" // needs to be unique
18	libclang-dev \	18	1,
19	libpq5 \	19	
20	libtiff-dev \	20	"forwardPorts": [8888],
21	8.8. \	21	"portsAttributes": {
22	apt-get clean -y && \	22	"8888": {
23	rm -rf /var/lib/apt/lists/* /tmp/library-scripts	23	"label": "Jupyterlab",
24		24 25	"onAutoForward": "ignore"
25	RUN R -e "dotR <- file.path(Sys.getenv('HOME'), '.R'); if(!file.exists(dotR)){ dir.create(dotR) }; Makevars <- file.path(dotR, 'M	}	
26	RUN R -e "dotRprofile <- file.path(Sys.getenv('HOME'), '.Rprofile'); if(!file.exists(dotRprofile)){ file.create(dotRprofile) }; c	26	3,
55			
32	chmod 777 /var/run/rstudio-server && \	32	"r.bracketedPaste": true,
33	chmod +t /var/run/rstudio-server	33	"r.plot.useHttpgd": true
34		34	},
35	USER \$ (NB_USER)	35	"extensions": [

36 37 RUN mamba install -y -c conda-forge --freeze-installed \ 38 jupyter-server-proxy=4.1.0 \ 39 jupyter-rsession-proxy=2.2.0 \ 40 1 28 41 mamba clean --all 42 43 RUN pip install \ 44 nbgitpuller \radian==0.6.11 \ 45 1 28 46 jupyter labextension enable nbgitpuller 47 48 RUN R -q -e 'remotes::install\_version("markdown", version="1.12", repos="cloud.r-project.org")' && \ 49 R -q -e 'remotes::install\_version("languageserver", version="0.3.16", repos="cloud.r-project.org")' && \ 50 R -q -e 'remotes::install\_version("httpgd", version="2.0.1", repos="cloud.r-project.org")' && \ 51 R -q -e 'remotes::install\_github("ManuelHentschel/vscDebugger")' && \ 52 R -q -e 'remotes::install\_version("syuzhet", version="1.0.7", repos="cloud.r-project.org")' && \ echo 54 55 RUN wget https://github.com/cli/cli/releases/download/v\${GITHUB\_CLI\_VERSION}/gh\_\${GITHUB\_CLI\_VERSION}\_linux\_amd64.tar.gz -0 - | 56 tar xvzf - -C /opt/conda/bin gh\_\${GITHUB\_CLI\_VERSION}\_linux\_amd64/bin/gh --strip-components=2

 32
 "r.bracketedPaste": true,

 33
 "r.plot.useHttpgd": true

 34
 },

 35
 "extensions": [

 36
 "reditorsupport.r", // for R

 37
 "RDebugger.r-debugger", // for R

 38
 "ms-vscode.live-server",

 39
 "analytic-signal.preview-pdf"

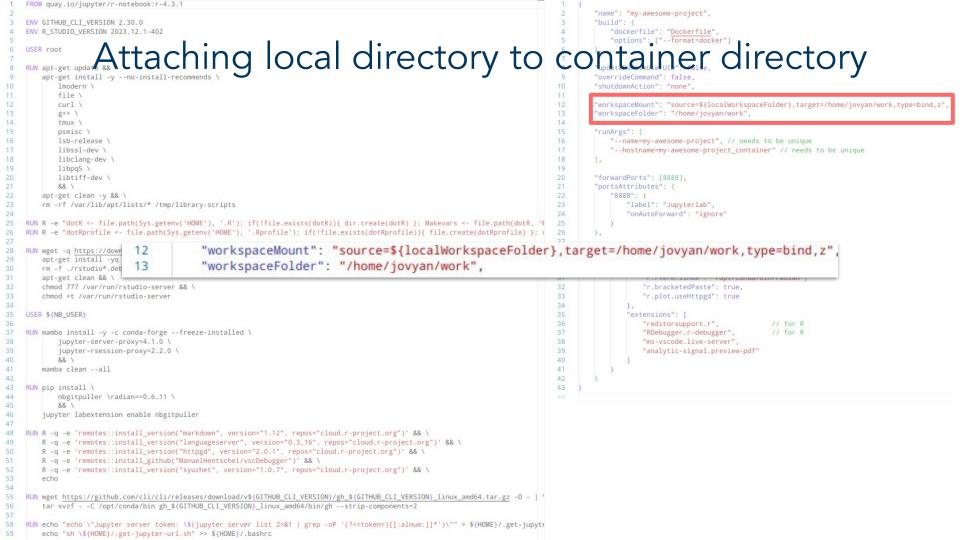
 40
 ]

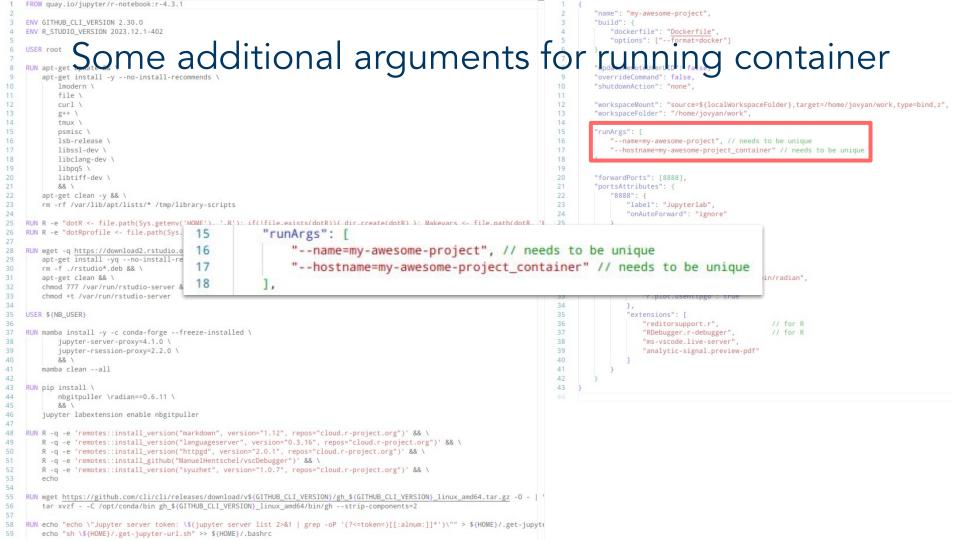
 41
 }

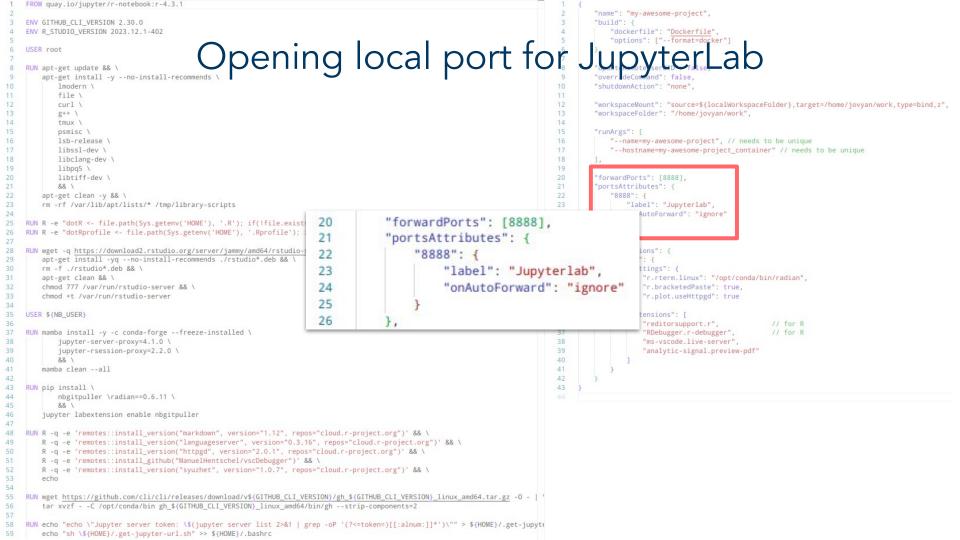
 42
 }

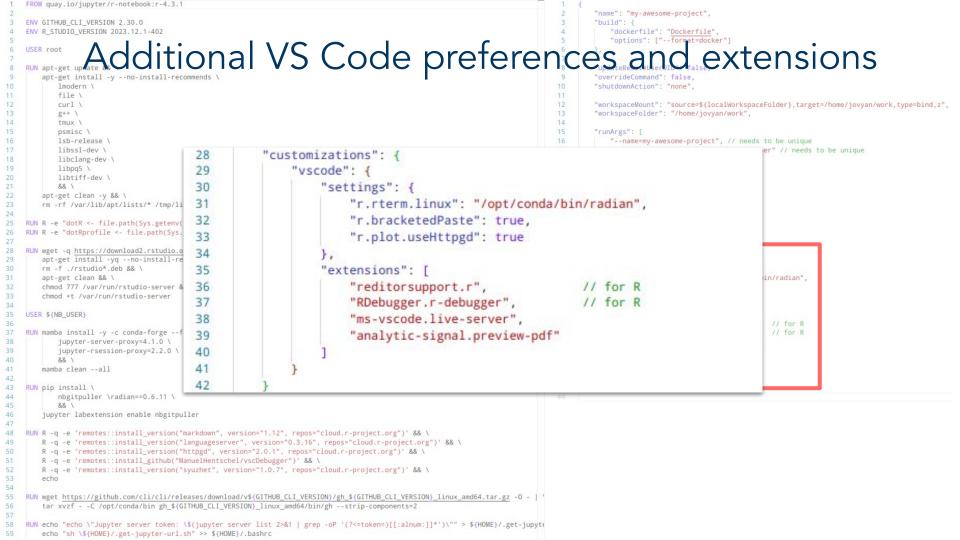
58 RUN echo "echo \"Jupyter server token: \\$(jupyter server list 2>&1 | grep -oP '(?<=token=)[[:alnum:]]\*')\"" > \${HOME}/.get-jupyte
59 echo "sh \\${HOME}/.get-jupyter-url.sh" >> \${HOME}/.bashrc













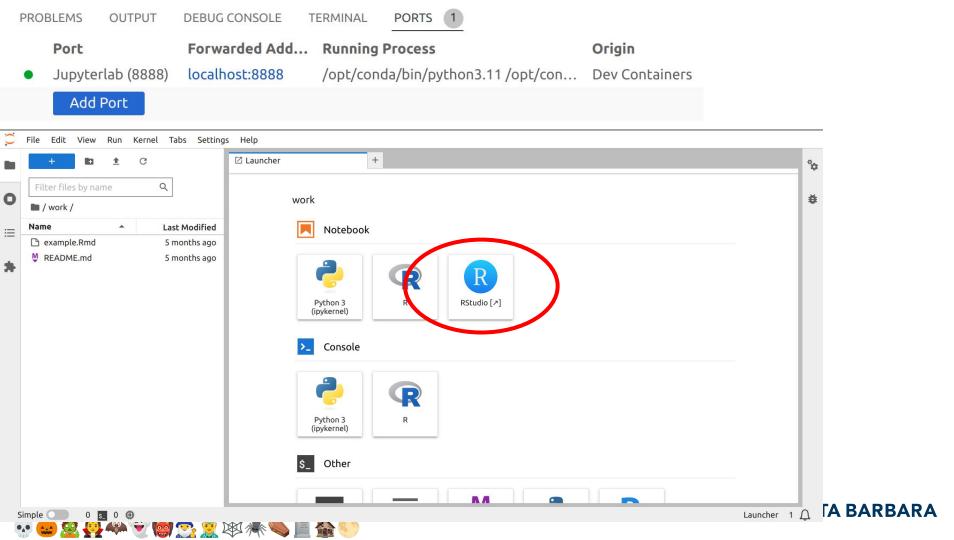
2024 Academic Data Science Alliance Annual Meeting

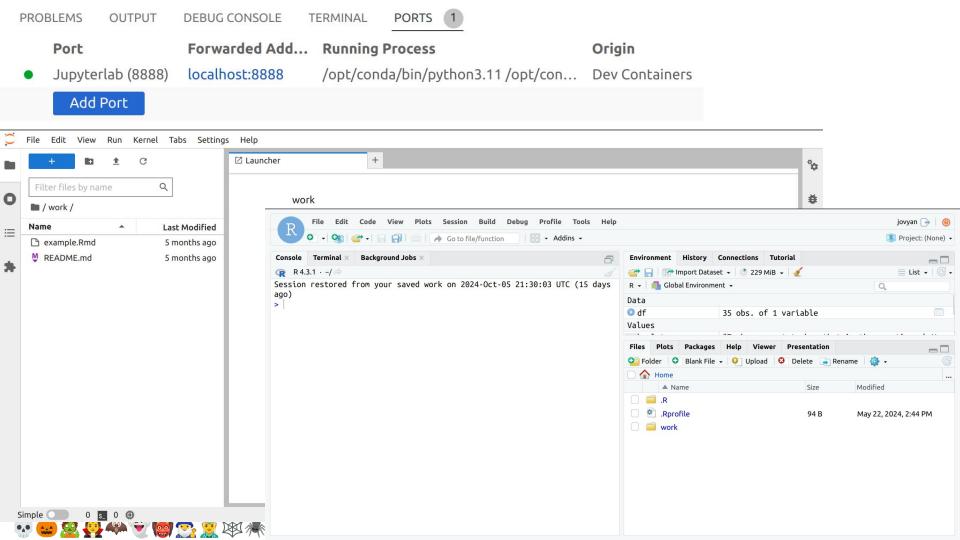


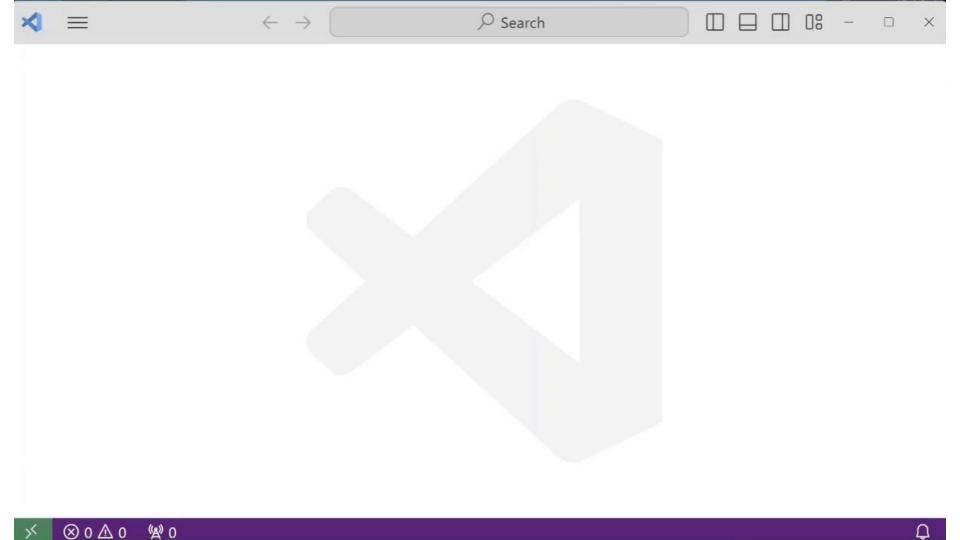
🗙 File Edit Selection View Go ·	\cdots bibliography.bib - work [Dev Container: paper-containers] - Visual Studio Code 🛛 🔲 🛞 🎇 \cdots 💷 🔍
EXPLORER ····	<pre>\$ containers.qmd × {} devcontainer.json 1, M</pre>
✓ WORK [DEV CONTAINER: PAPER-CONTAINERS]	src >
✓ .devcontainer	300 <b># Portability</b> 6235 @article{Meurer2017,
<pre>{} devcontainer.json 1, M</pre>	<pre>whether it is their local desktop, a remote department server, or an HPC/ supercomputing facility. In regards to the latter, HPCs tend to use a different container platform called Apptainer (Singularity) to build out the batch job instances. Fortunately, Apptainer is compatible with the Docker platform and allows for Dockerfiles to be directly converted into Apptainer compatible files as seen in @lst-apptainer-conversion.</pre>
<ul> <li>➡ containers.qmd</li> <li>Ξ sn-apacite.bst</li> </ul>	307       6252       year = {2024},         308       :::{#lst-apptainer-conversion       6253       urldate = {2024-06-13},         lst-cap="Building a local Dockerfile       6254       burnthlicked       (bttract (container a local Dockerfile)
≡ sn-aps.bst	PROBLEMS 1 OUTPUT DEBUG CONSOLE TERMINAL PORTS 1 X
<ul> <li>≡ sn-basic.bst</li> <li>≡ sn-chicago.bst</li> <li>&gt; sn-jnl.cls</li> </ul>	PortForwarded AddressRunning ProcessOriginJupyterlab (8888)localhost:8888/opt/conda/bin/python3.11/opt/conda/binDev Containers
<ul> <li>Sn-mathphys.bst</li> <li>Sn-nature.bst</li> </ul>	Add Port
> TIMELINE	

PRO	BLEMS	OUTPUT	DEBUG CONSOLE	TERMINAL	PORTS 1	
	Port		Forwarded Add	Running	Process	Origin
•	Jupyterl	ab (8888).	localhost:8888	/opt/con	da/bin/python3.11 /opt/con	Dev Containers
	Add P	Port				









## **6** Further Extending Usability

- With additional extensions we can build and manage our containers (and images) interactively
- This simplifies the overall management process of your project (and more)

DOCKER          ✓ CONTAINERS       ✓         ✓       ☐         Individual Containers       >         ✓       ☐         Individual Containers       >         ✓       ☐         Ubuntu sharp_hodgkin - Ex       >	View Logs Attach Shell Attach Visual Studio Code	REMOTE EXPLORER       Remotes (Tu ∨       ··         REMOTES (TUNNELS/SSH)       ✓       SSH         >       □       alta         >       □       roble       →       ±	<ul> <li>✓ SOURCE CONTROL</li> <li>Message (Ctrl+Enter to commit on</li> <li>✓ Commit</li> </ul>
IMAGES Image: A state of the state	Inspect Open in Browser Stop Restart Remove	☐ denali ☐ sandbox ☐ guthix	<ul> <li>SOURCE CONTROL GR &amp; Auto </li> <li>Changes and edits origin/main</li> <li>Various updates and s origin/main</li> <li>updated build image 392781</li> <li>changed apt to apt-get (apt has an u</li> <li>devcontainer example updates Sang</li> </ul> UC SANTA BARBAR

## 🟫 Real World Usage - Our Department

- Want to revamp department server...
  - Original server was centrally managed
  - Languages, packages, software were all shared by users
  - Tools like RStudio were only accessible via a specific server
- 🕴 Needless to say this limited users...
  - Locked users to specific software versions
  - New packages needed to be requested
  - Dependency conflicts everywhere
  - Access limited to terminal interaction
- 🔽 Solution?: Devcontainers

2024 Academic Data Science Alliance Annual Meeting

🖇 👻 🐻 🐟 🤨 🕅





- Ability to reproduce and share work
- Work on a multiuser system in isolation
  - Something that is possible to configure for containers but somewhat nontrivial
- Ease of use

2024 Academic Data Science Alliance Annual Meeting

**.0. «». 👻 🐻 🐟 💇**  :

- As seen previously, creating more complex Dockerfiles can be, again, a nontrivial task
- Researchers should be able to spin up a custom image with ease
- More than that, they should be able to have a starting point for configuration
- Should also be able to easily connect to and manage remote servers!





• Because writing container files can be a difficult process...

...we created an extendable template!

○[ronaldas@alta ~]\$





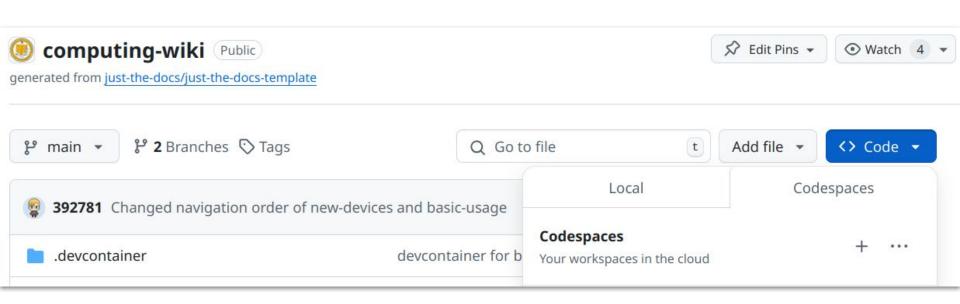
- We include our template tooling with an image on Jetstream2
- Usage is exactly the same between different systems



- Additionally, devcontainer files can be run without any additional tooling once uploaded on GitHub via a feature called CodeSpaces
- This is a feature that is available with GitHub Pro (which is free for all academics)











		$\leftrightarrow  ightarrow$ Computing-wiki [Codespaces: wretched superstition]	
≡	EXPLORER ····	同 [Preview] README.md ×	с <b>л</b> Ш …
<b>D</b>	<pre>     COMPUTING-WIKI [CODESPACES: WRETCHED SUP     _includes     _sass </pre>	Documentation with Codelabs	
r fo	<ul> <li>&gt; .devcontainer</li> <li>&gt; .github</li> <li>&gt; assets</li> </ul>	Build Steps	
	> docs ! _config.yml	Build website	
₿	<ul> <li>◆ .gitignore</li> <li>■ DEVELOPMENT</li> <li>■ docs.csv</li> </ul>	<b>Container</b> Simply build the Dockerfile via Docker/Podman (or using VS Code). Once built, run the following inside the container:	
	<ul> <li>Gemfile</li> <li>Gemfile.lock</li> <li>index.md</li> <li>LICENSE</li> </ul>	bundle exec jekyll servelivereload	
	<ul> <li>README.md</li> <li>wiki.py</li> <li>wiki.sh</li> </ul>	PROBLEMS OUTPUT DEBUG CONSOLE <u>TERMINAL</u> PORTS ② COMMENTS ○ root@codespaces-f36fd8:/workspaces/computing-wiki#	∽ [] @ … ^ ×
8			
€33 × Cor	> OUTLINE     > TIMELINE despaces: wretched superstition     & main ↔ ⊗ 0 ▲	0 <sup>(k)</sup> 2	Layout: us
	024 Academic Data Science Alliance	Annual Meeting UC SANTA E	



[1] Buckheit, Jonathan B, and David L Donoho. 1995. "Wavelab and Reproducible Research." *Springer*.

[2] Moreau, David, Kristina Wiebels, and Carl Boettiger. 2023. "Containers for Computational Reproducibility." *Nature Reviews Methods Primers* 3 (1): 50

[3] Microsoft. 2024. "Development Containers." <u>https://containers.dev/</u>.

[4] Jupyter. 2024. "Jupyter Docker Stacks — Docker Stacks Documentation."<u>https://jupyter-docker-stacks.readthedocs.io</u>.





UCSB Department of Statistics & Applied Probability Computing

Dr. Sang-Yun Oh (<u>syoh.org</u>) Ronald Lencevičius (<u>ronaldas.dev</u>)

