Reproducible R Development with Docker and renv

phb 243b

2025-03-10

Table of contents

1 Introduction

- Ensuring reproducibility in R Markdown can be challenging.
- Package version mismatches and OS differences cause issues.
- Solution: **renv** (package management) + **Docker** (containerization).

2 Workflow Overview

- Developer 1: Sets up **renv** & Docker, pushes to GitHub & DockerHub.
- Developer 2: Clones repo, pulls Docker image, runs & modifies analysis.
- Uses **bind mounting** to ensure local files persist.

3 Developer 1 - Setup GitHub Repository

```
git clone https://github.com/username/penguins-ana
cd penguins-analysis
```

4 Developer 1 - Initialize renv

```
install.packages("renv")
renv::init()
```

- Creates **renv.lock** to track package versions.

5 Developer 1 - Install Required Packages

```
install.packages("ggplot2")
install.packages("palmerpenguins")
renv::snapshot()
```

- **renv::snapshot()** saves exact package versions.

6 Developer 1 - Create Dockerfile

```
FROM rocker/r-ver:4.1.0
WORKDIR /workspace
RUN R -e "install.packages('renv', repos='https://
COPY renv.lock renv/activate.R /workspace/
RUN R -e "renv::restore()"
CMD ["/bin/bash"]
```

- **Does NOT include peng1.Rmd** (handled locally by Developer 2).

7 Developer 1 - Build & Push Docker Image

```
docker build -t username/penguins-analysis:v1 .
docker login
docker push username/penguins-analysis:v1
```

- Image now available on **DockerHub**.

8 Developer 1 - Push to GitHub

```
git add .
git commit -m "Initial renv setup and Docker envir
git push origin main
```

9 Developer 2 - Clone & Pull Docker Image

```
git clone https://github.com/username/penguins-ana
cd penguins-analysis
docker pull username/penguins-analysis:v1
```

10 Developer 2 - Run Docker Interactively

```
docker run --rm -it -v "$(pwd):/workspace" -w /wor
```

- Mounts **local repository** to `/workspace` inside the container.

11 Developer 2 - Write peng1.Rmd

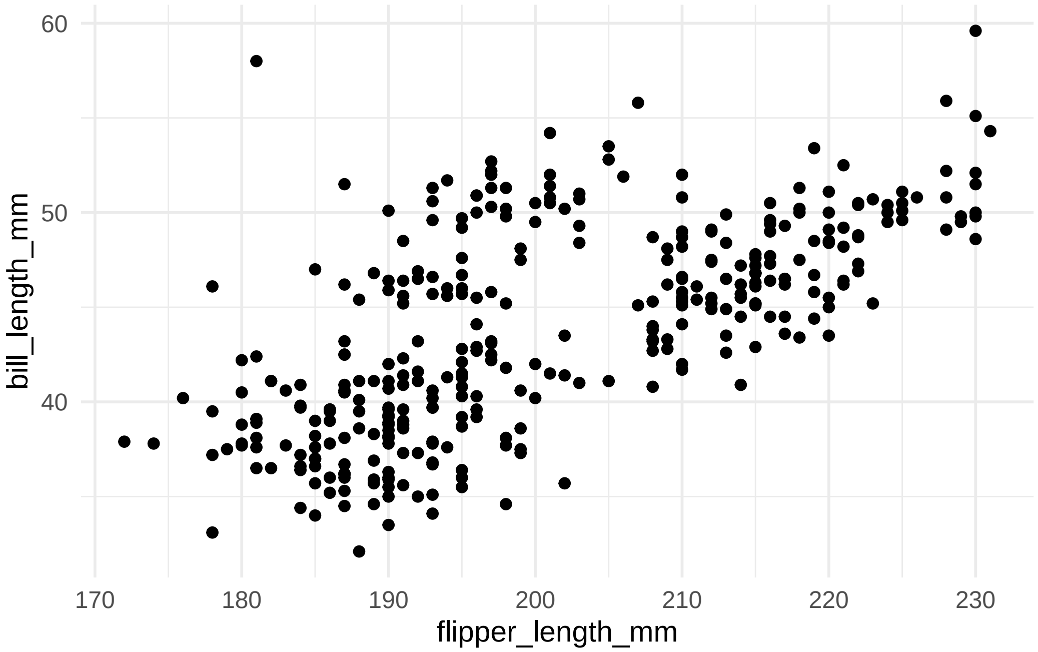
~~~
---
title: "Palmer Penguins Analysis"
author: "Developer 2"
date: "2025-03-10"
output: html_document
---

::: {.cell}

```{.r .cell-code}
library(ggplot2)
library(palmerpenguins)

ggplot(penguins, aes(x = flipper_length_mm, y = bi
  geom_point() + theme_minimal()
~~~

```
Warning: Removed 2 rows containing missing values
```



:::

~~~~
## Developer 2 - Render R Markdown
```r
rmarkdown::render("peng1.Rmd")
~~~~

- Output **peng1.html** is saved **locally** (due to bind mount).

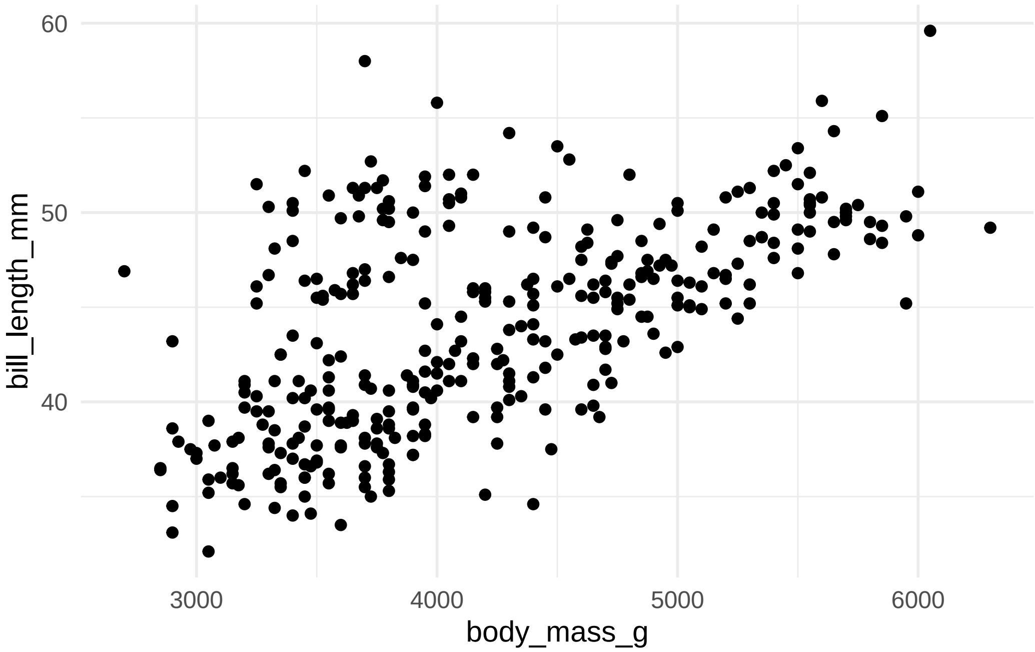
12 Developer 2 - Extend Analysis & Push to GitHub

~~~
## Body Mass vs. Bill Length

::: {.cell}

```{.r .cell-code}
ggplot(penguins, aes(x = body_mass_g, y = bill_len
  geom_point() + theme_minimal()
~~~

```
Warning: Removed 2 rows containing missing values
```



:::

~~~
```sh
git add peng1.Rmd
git commit -m "Added second plot: Body Mass vs. Bi
git push origin main
~~~

13 Conclusion

renv ensures package reproducibility.
Docker ensures OS & R version consistency.

GitHub & **DockerHub** enable collaboration.

Bind mounting ensures local modifications persist.

14 Thank You!

- Questions?
- Try setting up your own Dockerized R environment!