

Practice with

ggplot2:

Build a data
MASTERPIECE



Viviana Ortiz · Paulo Izquierdo

Did you get the datasets and the
“practice_ggplot.Rmd” file?



Starting a project!!!



What is a RStudio project, and why?

- The RStudio project file is a file that sits in the root directory, with the extension .Rproj.
- The **working directory** points to the root folder where that .Rproj file is saved
- Solve the problems associated with `setwd()`:
 - Links break very easily
 - Reproducibility



Illustration by Allison Horst

What is a RStudio project, and why?

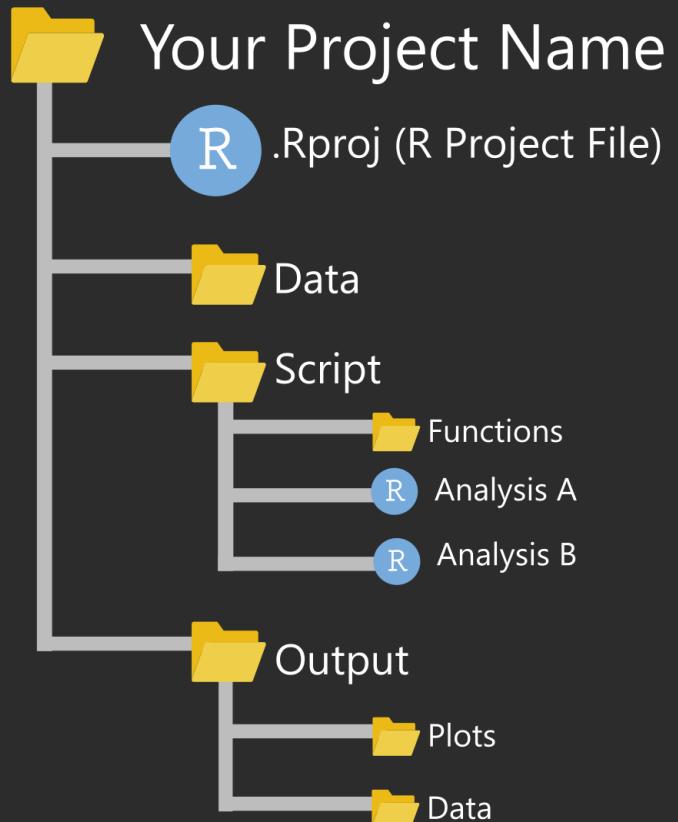
Easy file path referencing with RStudio projects

file paths *relative*:
Data/Data1.xlsx.

Organizing your project

Helps anybody else you are collaborating with - or a future version of you trying to reproduce some analysis

A basic R project set up



<https://martinctc.github.io>

Start your R project

- Create a new R Project
- Move the R script: “practice_ggplot.Rmd” inside a “scripts” folder
- Move dataset: “Gapminder_vars_2011.csv” into the “data” folder
- Open “practice_ggplot.Rmd” file
- Read in the dataset

A screenshot of the RStudio IDE interface. A red arrow points to the 'New Project...' option in the 'File' menu.

The 'File' menu is open, showing the following options:

- New File
- New Project...**
- Open File...
- Reopen with Encoding...
- Recent Files
- Open Project...
- Open Project in New Session...
- Recent Projects
- Import Dataset
- Save
- Save As...
- Rename
- Save with Encoding...
- Save All
- Knit Document
- Publish...
- Print...
- Close
- Close All
- Close All Except Current
- Close Project
- Quit Session...

The main workspace shows two R Markdown files: 'practice_ggplot.Rmd' and 'practice_ggplot_sols.Rmd'. The code in 'practice_ggplot.Rmd' includes a section for installing packages:

```
#> " ", "ggExtra", "ggridges", "janitor", "pheatmap")
```

The 'Console' tab displays the output of the package installation command:

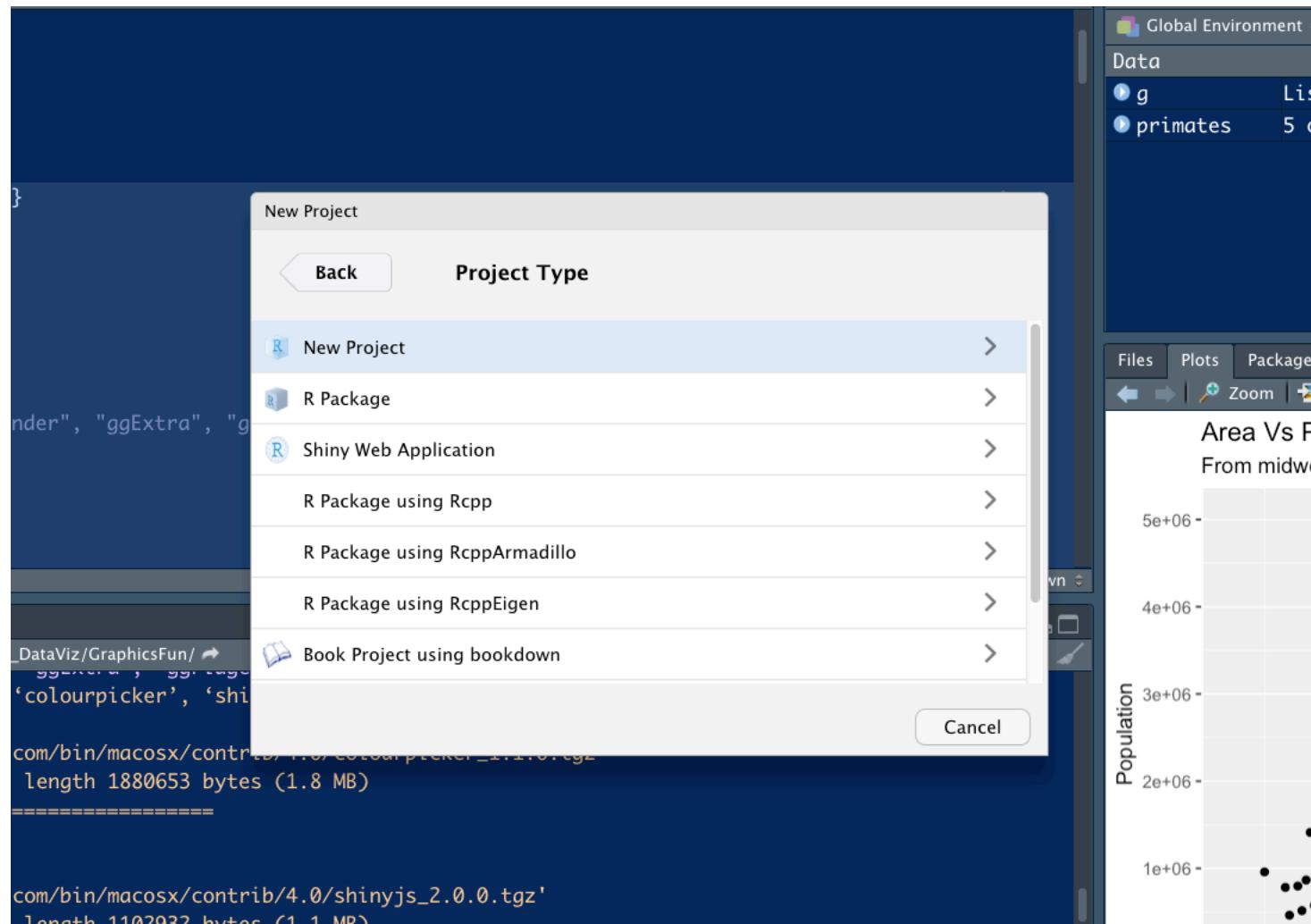
```
also installing the dependencies 'colourpicker', 'shinyjs', 'snakecase'

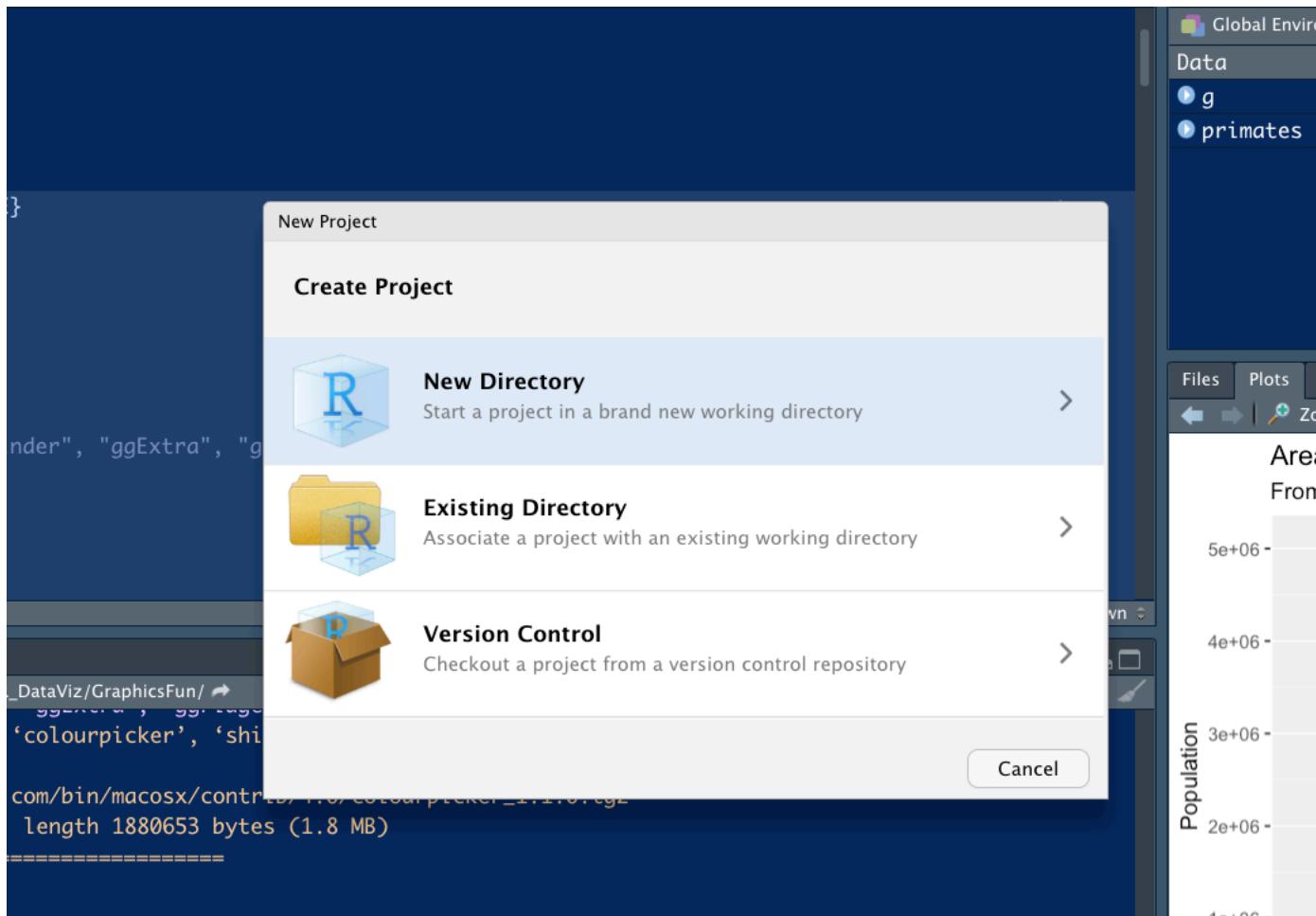
trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.0/colourpicker_1.1.0.tgz'
Content type 'application/x-gzip' length 1880653 bytes (1.8 MB)
=====
downloaded 1.8 MB

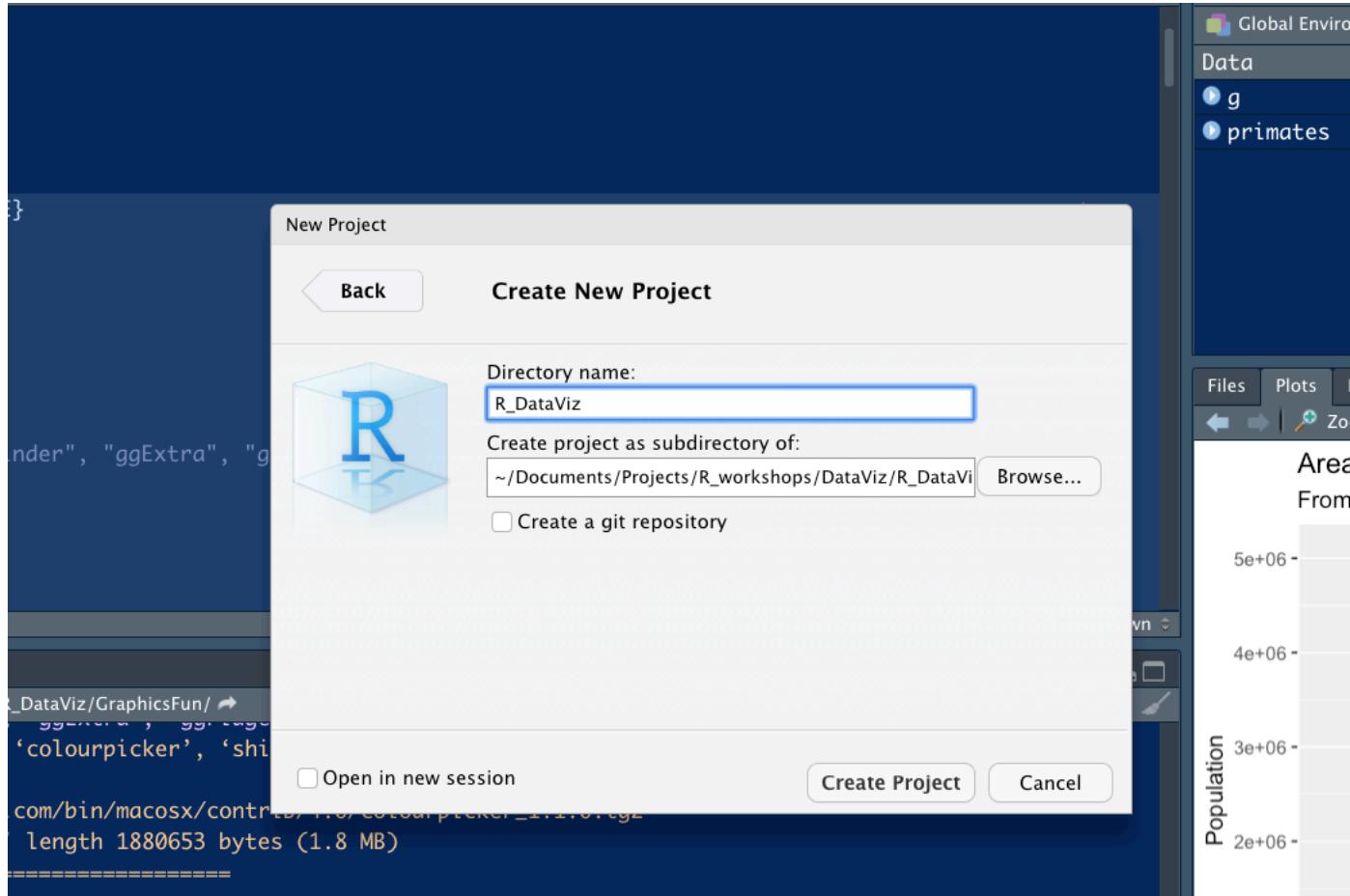
trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.0/shinyjs_2.0.0.tgz'
Content type 'application/x-gzip' length 1102932 bytes (1.1 MB)
=====
downloaded 1.1 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.0/snakecase_0.11.0.tgz'
```

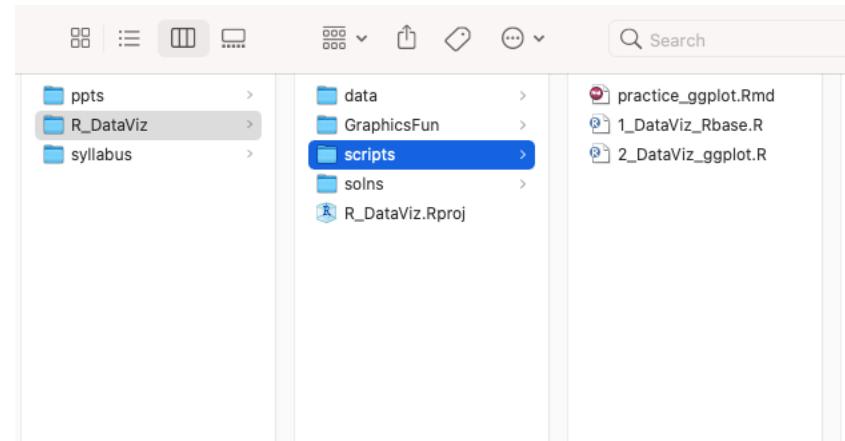
The 'Plots' panel shows a scatter plot titled 'Area Vs Population' from the 'midwest' dataset. The x-axis is labeled 'Area' and ranges from 0.00 to 0.09. The y-axis is labeled 'Population' and ranges from 0e+00 to 5e+06. The plot shows a positive correlation, with most data points clustered between Area 0.01 and 0.03 and Population 0e+00 and 1e+06, and a few outliers at higher values.



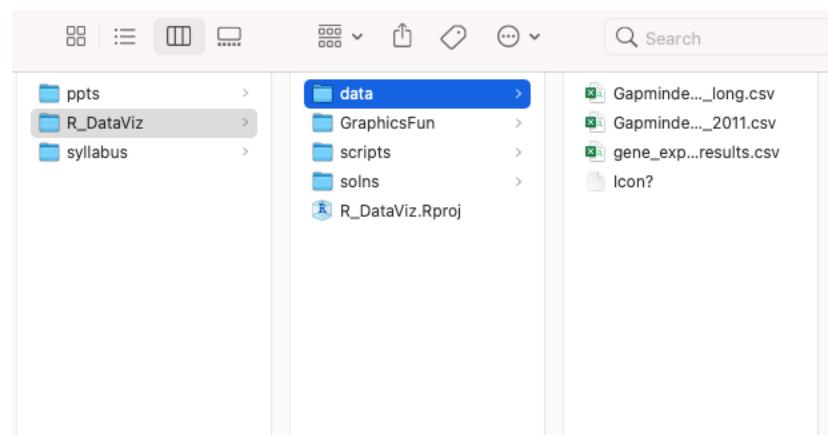




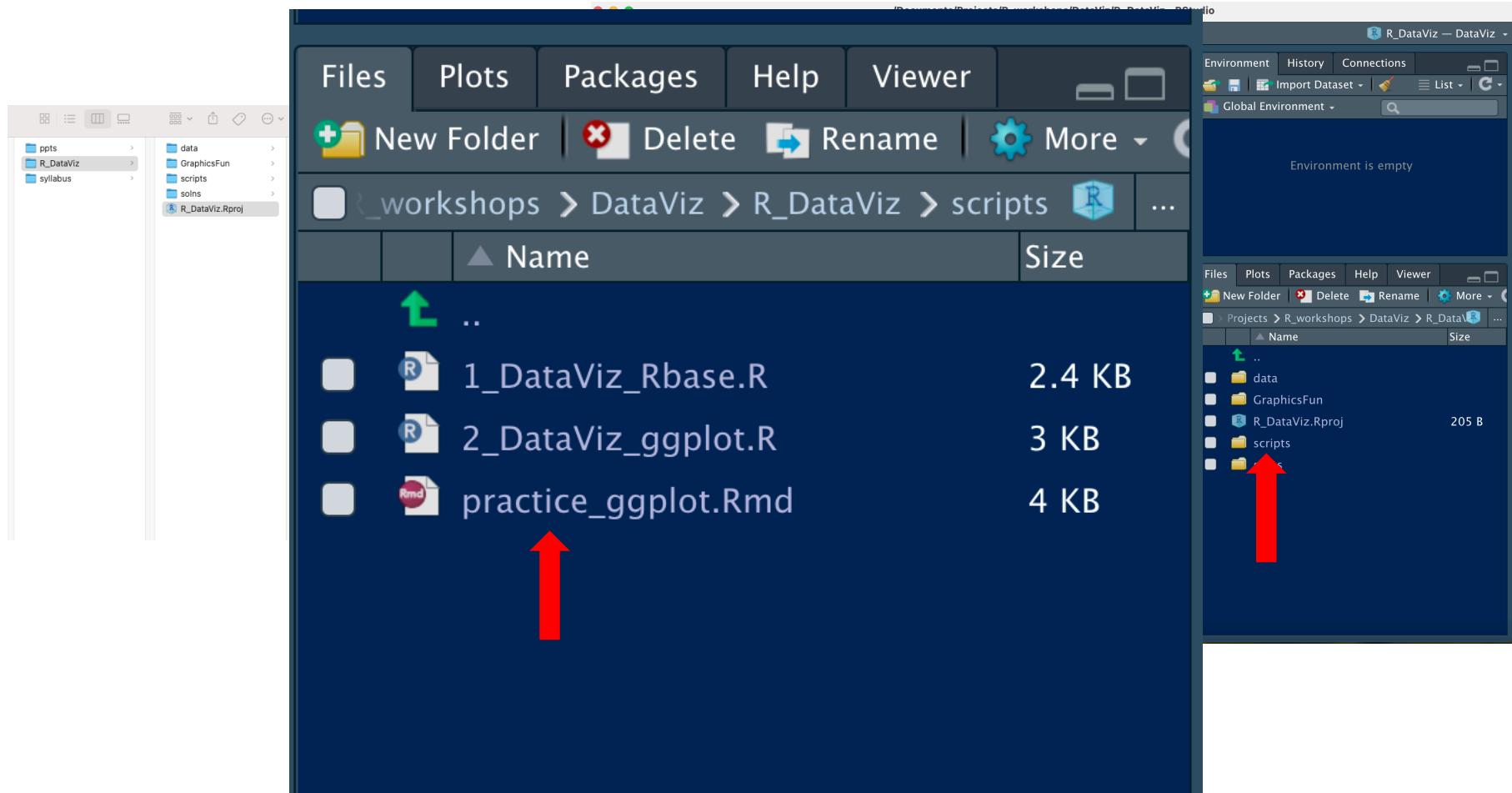
Move the R script: “practice_ggplot.Rmd”
inside a “scripts” folder



Move data into the “data” folder



Open the practice_ggplot.Rmd file



Read in the dataset

```
gapminder2011 <-  
read_csv("data/Gapminder_vars_2011.csv")
```

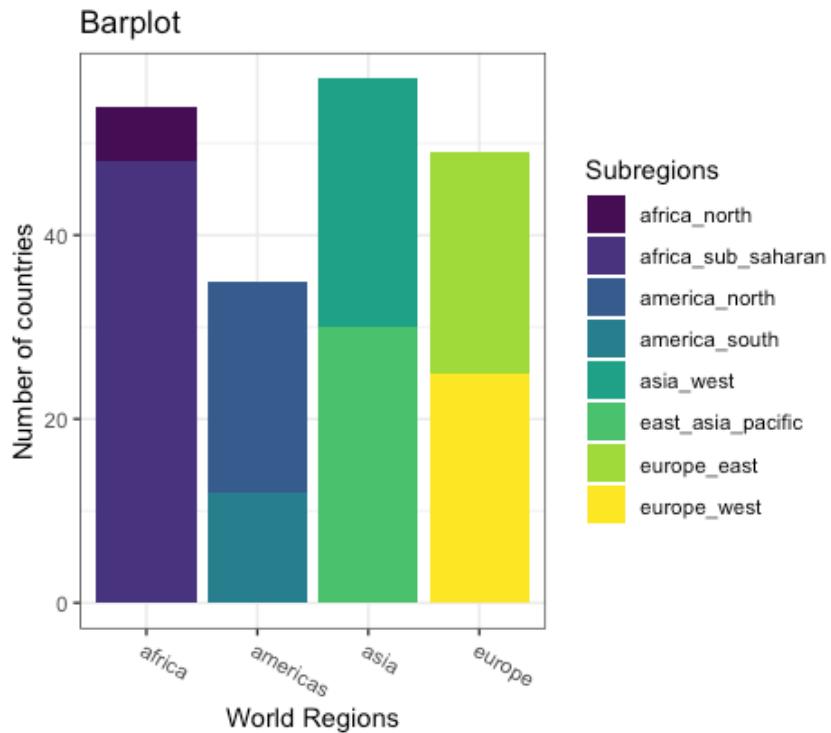
Exercises!



practice_ggplot.Rmd file authors: Jessica Minnier and Meike Niederhausen
Oregon Health & Science University
Downloaded at: github.com/jminnier/berd_ggplot_project

Barplot

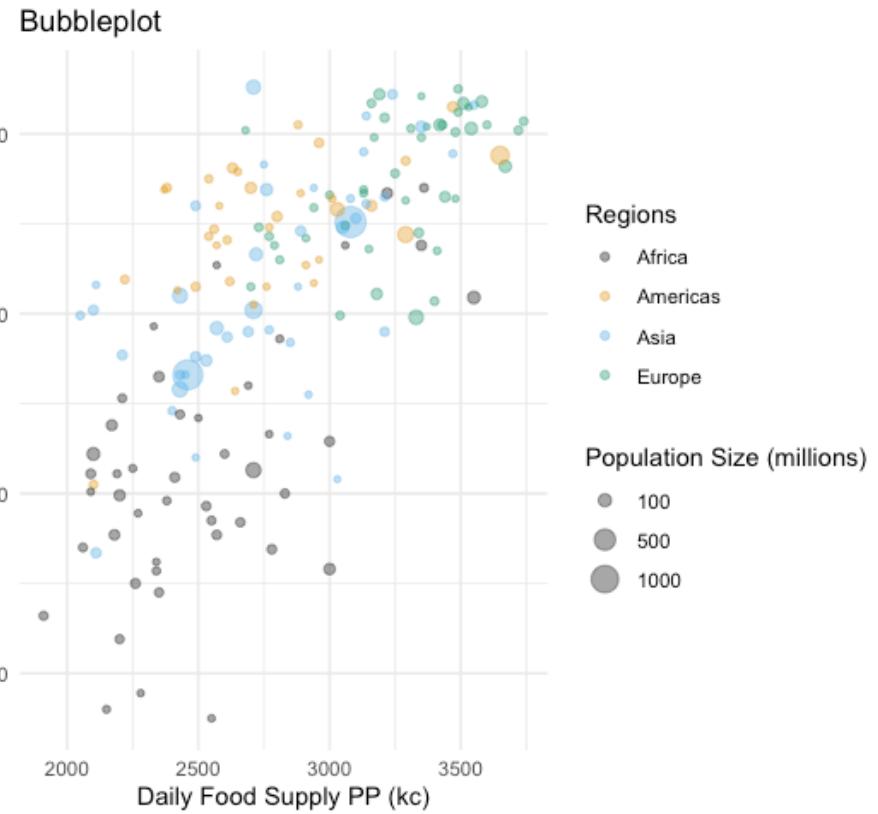
```
ggplot(data = gapminder2011,  
       aes(x = four_regions,  
            fill = eight_regions)) +  
  geom_bar() +  
  labs(x = "World Regions",  
       y = "Number of countries",  
       title = "Barplot") +  
  theme_bw() +  
  theme(  
    axis.text.x = element_text(angle = -30,  
                                hjust = 0)) +  
  scale_fill_viridis_d(name = "Subregions")
```



?geom_bar() geom_bar(mapping = NULL, data =
NULL, stat = "count", position = "stack", ...

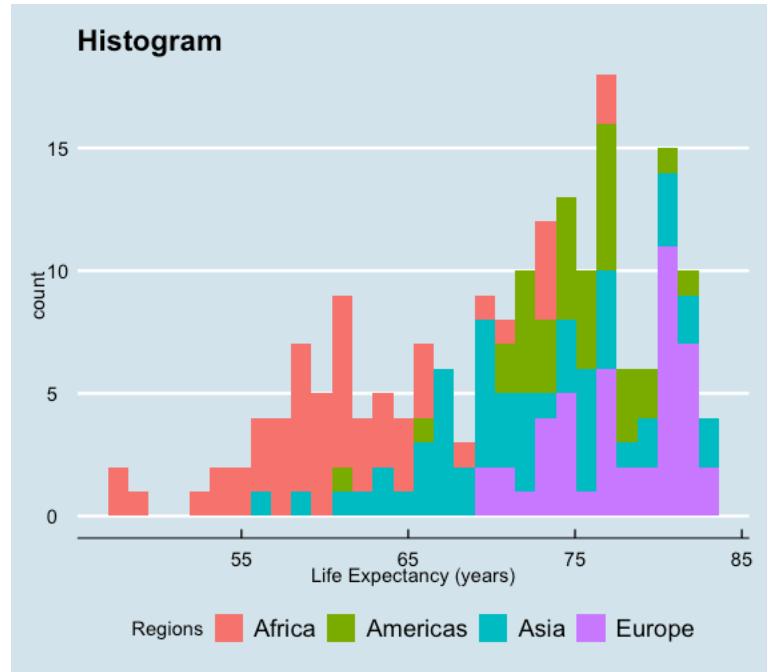
Bubbleplot

```
ggplot(data = gapminder2011,
       aes(x = FoodSupplykcPPD,
            y = LifeExpectancyYrs,
            color = four_regions,
            size = population)) +
  geom_point(alpha = 0.4) +
  scale_color_colorblind(
    name = "Regions",
    labels = c("Africa", "Americas",
              "Asia", "Europe"))
) +
  scale_size(
    name = "Population Size (millions)",
    breaks = c(1e08,5e08,1e09),
    labels = c(100,500,1000)
) +
  theme_minimal() +
  labs(
    x = "Daily Food Supply PP (kc)",
    y = "Life Expectancy (years)",
    title = "Bubbleplot"
)
```



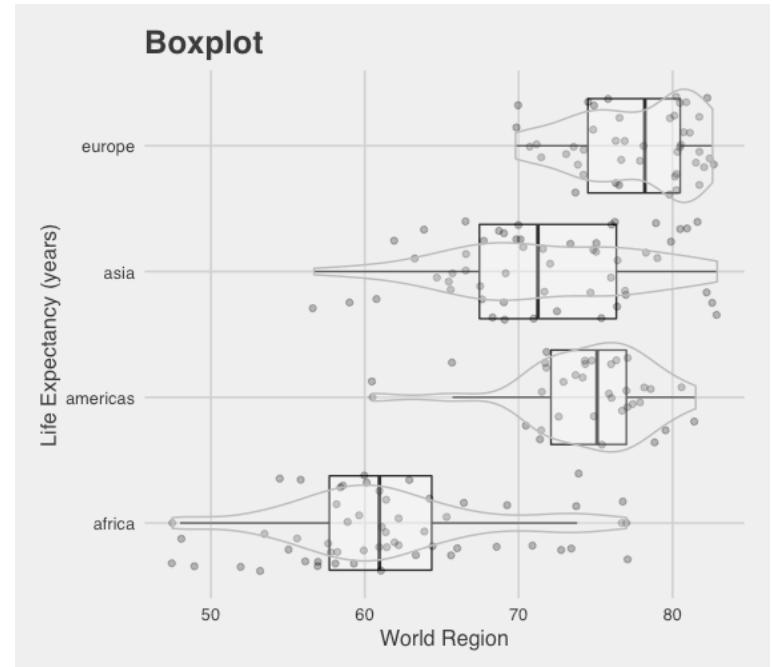
Histogram

```
ggplot(data = gapminder2011,  
       aes(x = LifeExpectancyYrs,  
            fill = four_regions)) +  
  geom_histogram() +  
  scale_fill_discrete(  
    name = "Regions",  
    labels = c("Africa", "Americas",  
              "Asia", "Europe"))  
  ) +  
  labs(  
    x = "Life Expectancy (years)",  
    title = "Histogram"  
  ) +  
  ggthemes::theme_economist() +  
  theme(  
    legend.position="bottom"  
  )
```



Boxplot

```
ggplot(data = gapminder2011,
       aes(x = LifeExpectancyYrs, # New!
            y = four_regions
            )
      ) +
  geom_boxplot(alpha = 0.3) +
  theme_fivethirtyeight() +
  theme(axis.title = element_text()) +
  scale_fill_fivethirtyeight() +
  theme(legend.position = "none") +
  geom_jitter(width = .1, alpha = 0.3) +
  geom_violin(colour = "grey", alpha = .2) +
  labs(
    x = "World Region",
    y = "Life Expectancy (years)",
    title = "Boxplot"
  )
```



Learning R can be scary.



Illustration by Allison Horst

And it's an investment.



Illustration by Allison Horst

debugging

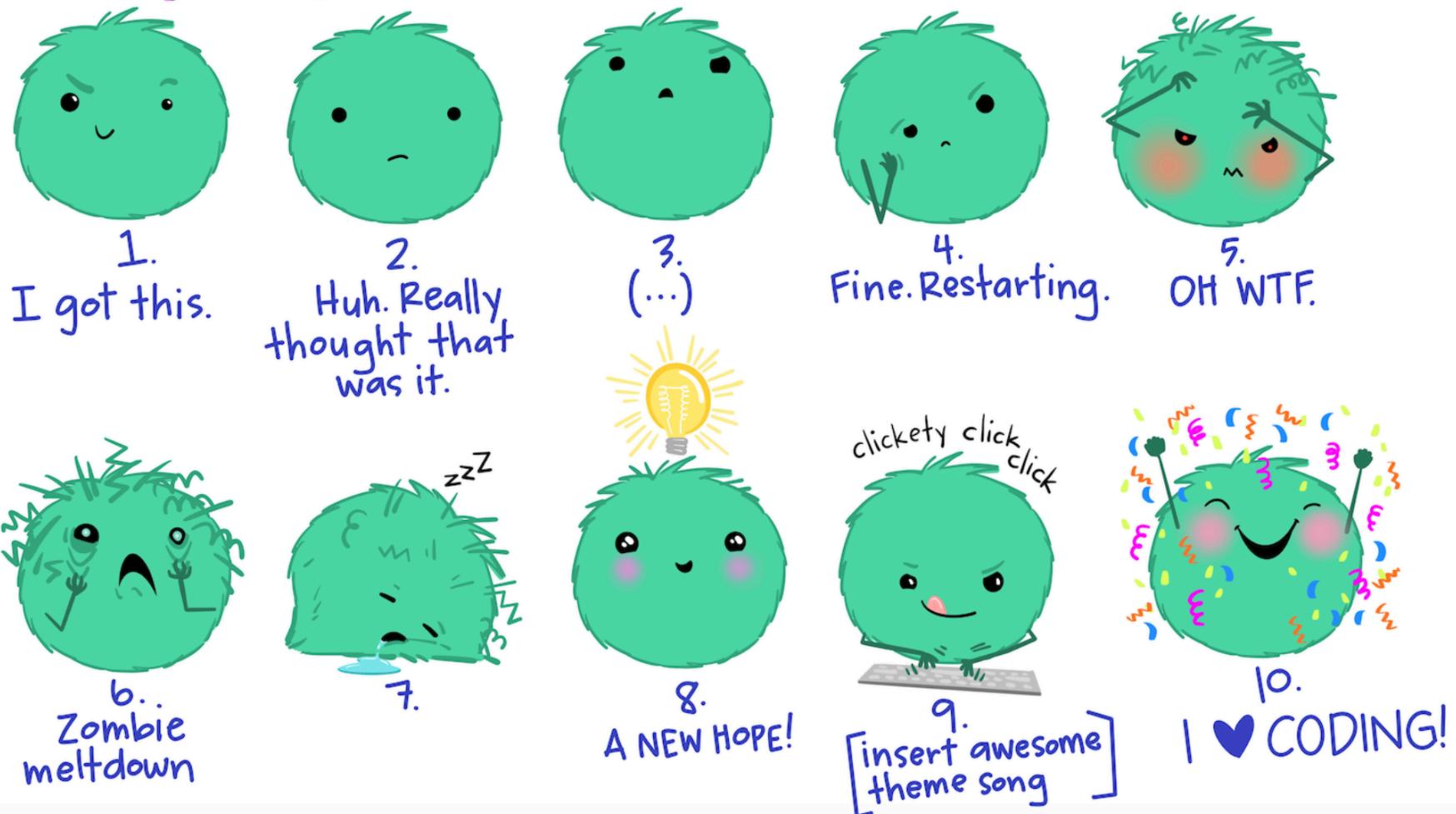


Illustration by Allison Horst