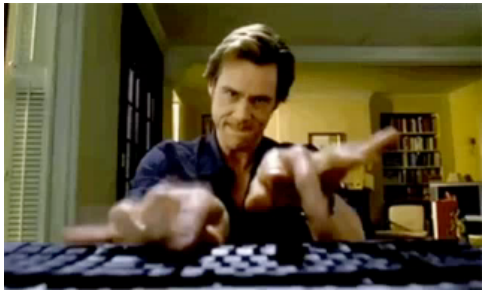
Practice with

Illustration by Allison Horst

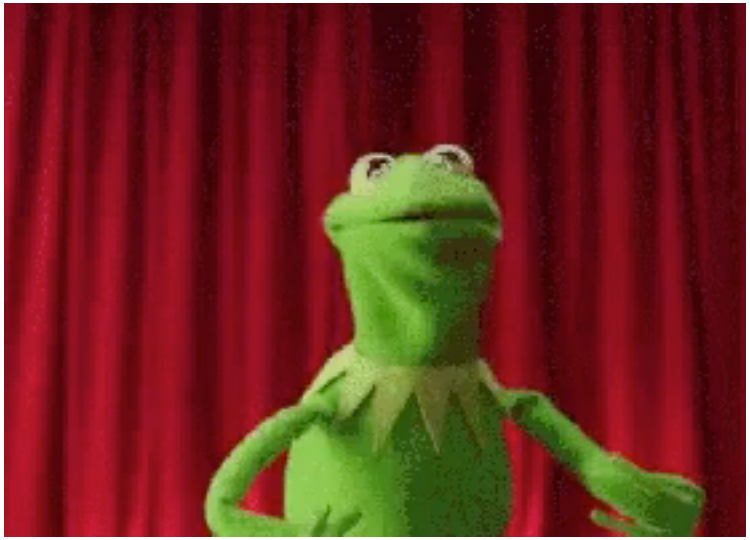
Viviana Ortiz · Paulo Izquierdo

13 Feb 2021

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**

- RStudio projects 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

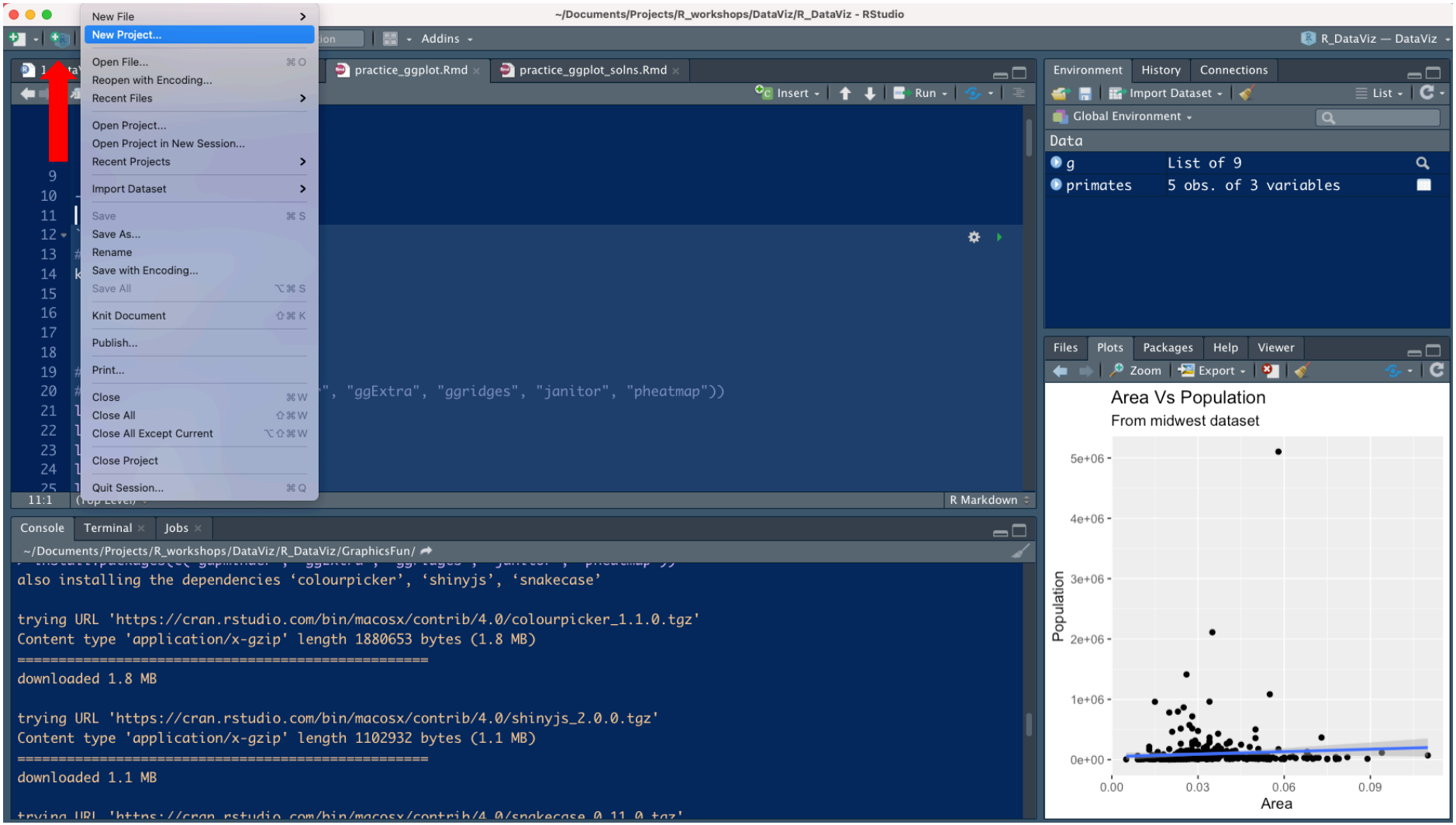
file paths relative to .Rproj file:
Data/Data1.xlsx.

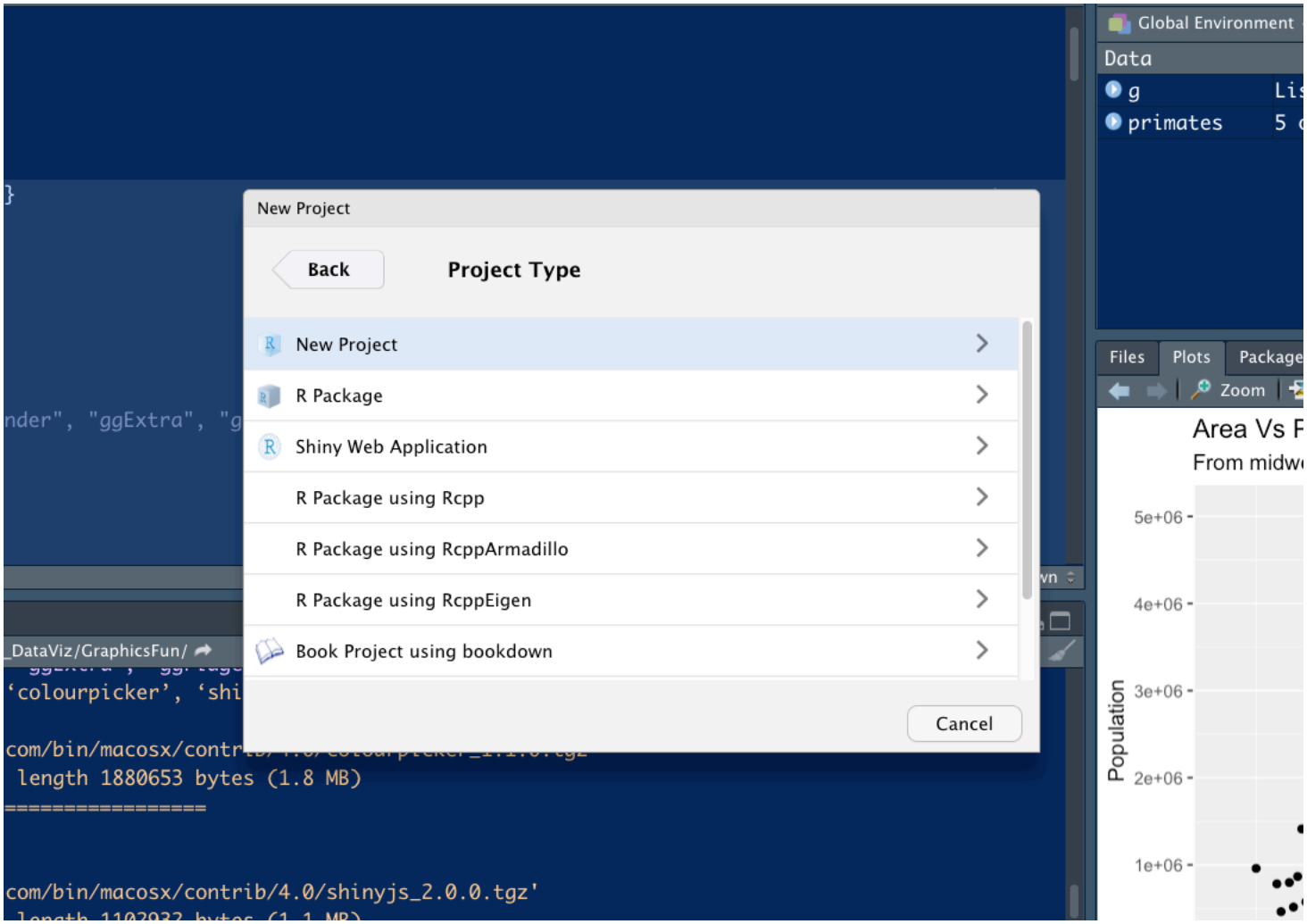
Organizing your project

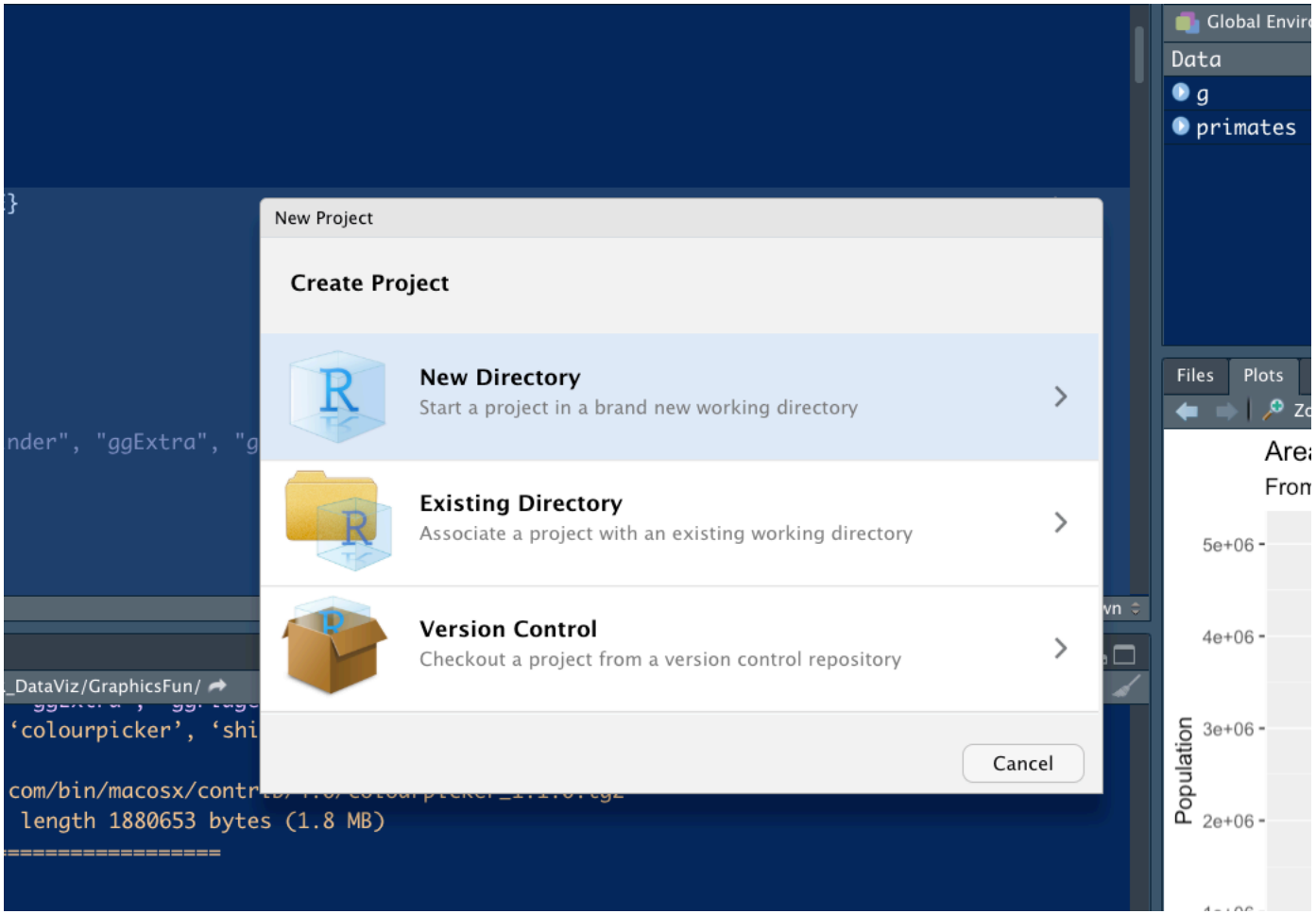
Helps anybody else you are collaborating with - or a future version of you trying to reproduce some analysis

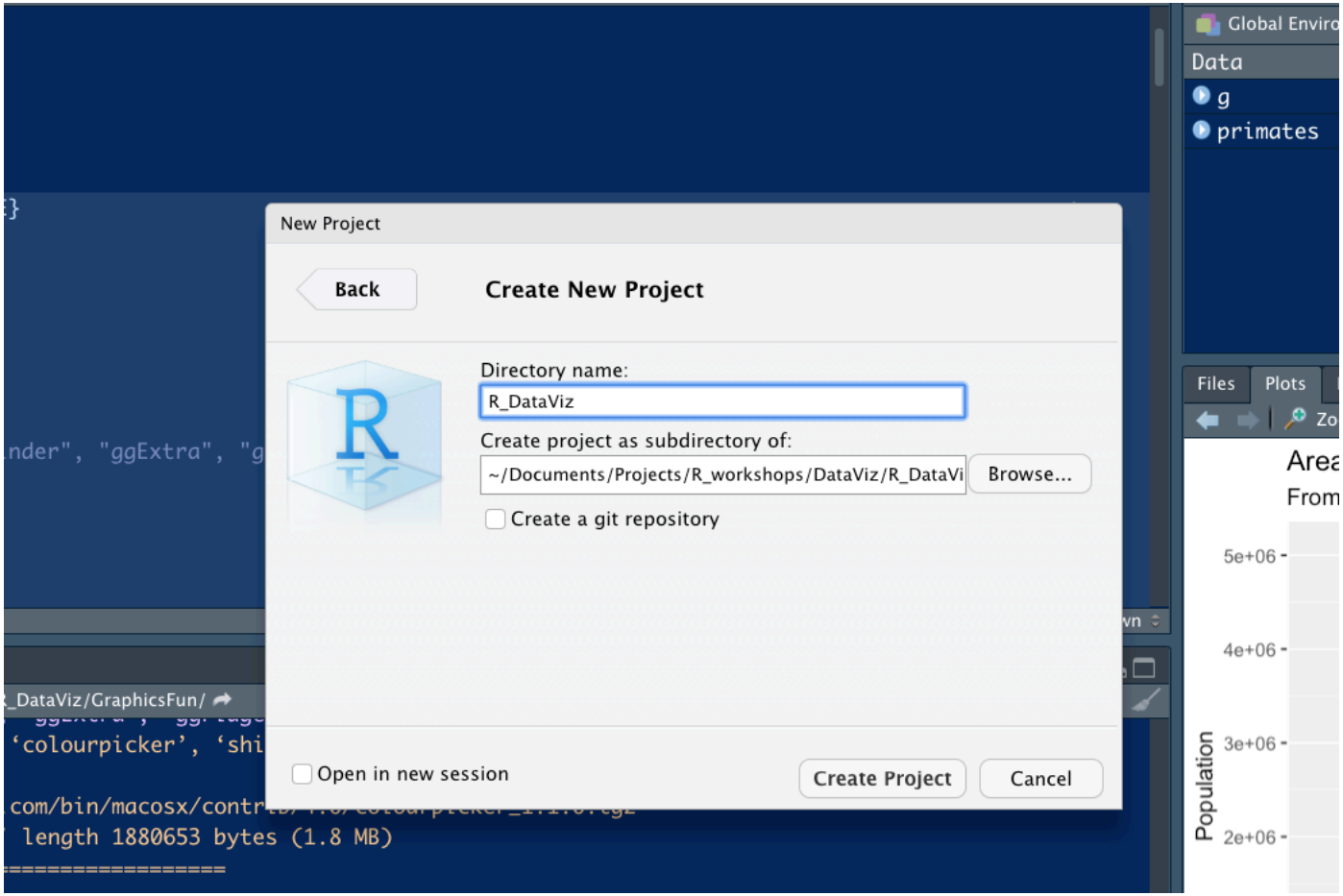
Start your R project

- 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

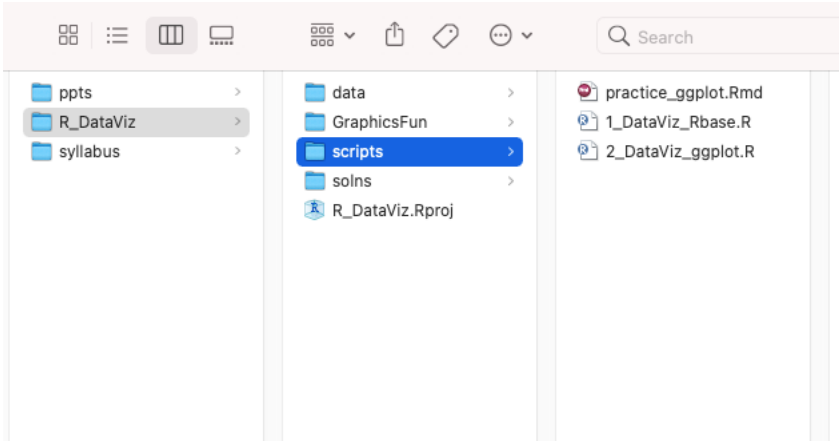

~/Documents/Projects/R_workshops/DataViz/R_DataViz - RStudio
New File
New Project...
Open File... ⌘O
Reopen with Encoding...
Recent Files
Open Project...
Open Project in New Session...
Recent Projects
Import Dataset
Save ⌘S
Save As...
Rename
Save with Encoding...
Save All ⌥⌘S
Knit Document ⇧⌘K
Publish...
Print...
Close ⌘W
Close All ⇧⌘W
Close All Except Current ⌥⇧⌘W
Close Project
Quit Session... ⌘Q
Addins
R_DataViz — DataViz
practice_ggplot.Rmd
practice_ggplot_solns.Rmd
Insert
Run
"ggExtra", "ggridges", "janitor", "pheatmap"))
11:1
R Markdown
Environment
History
Connections
Import Dataset
List
Global Environment
Data
g List of 9
primates 5 obs. of 3 variables
Files
Plots
Packages
Help
Viewer
Zoom
Export
Area Vs Population
From midwest dataset
Population
Area
5e+06
4e+06
3e+06
2e+06
1e+06
0e+00
0.00
0.03
0.06
0.09
Console
Terminal
Jobs
~/Documents/Projects/R_workshops/DataViz/R_DataViz/GraphicsFun/
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


New Project
Back
Project Type
New Project
R Package
Shiny Web Application
R Package using Rcpp
R Package using RcppArmadillo
R Package using RcppEigen
Book Project using bookdown
Cancel
Global Environment
Data
g
primates
Files
Plots
Zoom
Area Vs P
From midw
Population
5e+06
4e+06
3e+06
2e+06
1e+06
_DataViz/GraphicsFun/
'colourpicker', 'shi
length 1880653 bytes (1.8 MB)
com/bin/macosx/contrib/4.0/shinyjs_2.0.0.tgz'

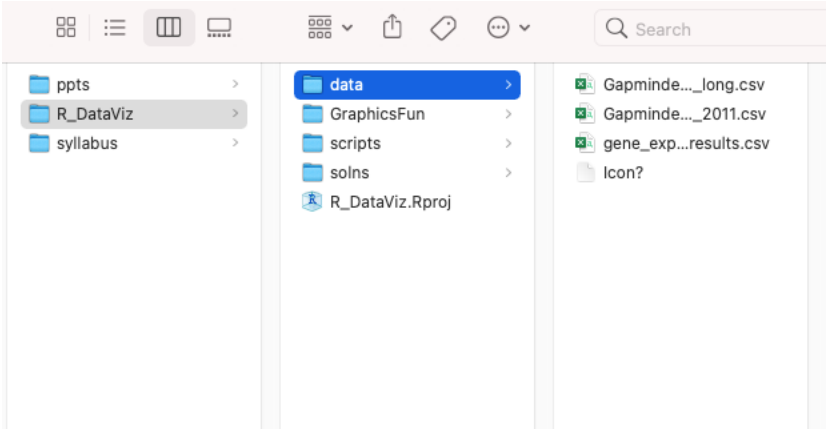

Global Enviro
Data
g
primates
New Project
Create Project
New Directory
Start a project in a brand new working directory
Existing Directory
Associate a project with an existing working directory
Version Control
Checkout a project from a version control repository
Cancel
Files
Plots
Area
From
5e+06
4e+06
3e+06
2e+06
Population
_DataViz/GraphicsFun/
'colourpicker', 'shi
length 1880653 bytes (1.8 MB)


New Project
Back
Create New Project
Directory name:
R_DataViz
Create project as subdirectory of:
~/Documents/Projects/R_workshops/DataViz/R_DataVi
Browse...
Create a git repository
Open in new session
Create Project
Cancel
Global Enviro
Data
g
primates
Files
Plots
Area
From
5e+06
4e+06
3e+06
2e+06
Population
length 1880653 bytes (1.8 MB)

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



Move “Gapminder_vars_2011.csv” into the “data” folder



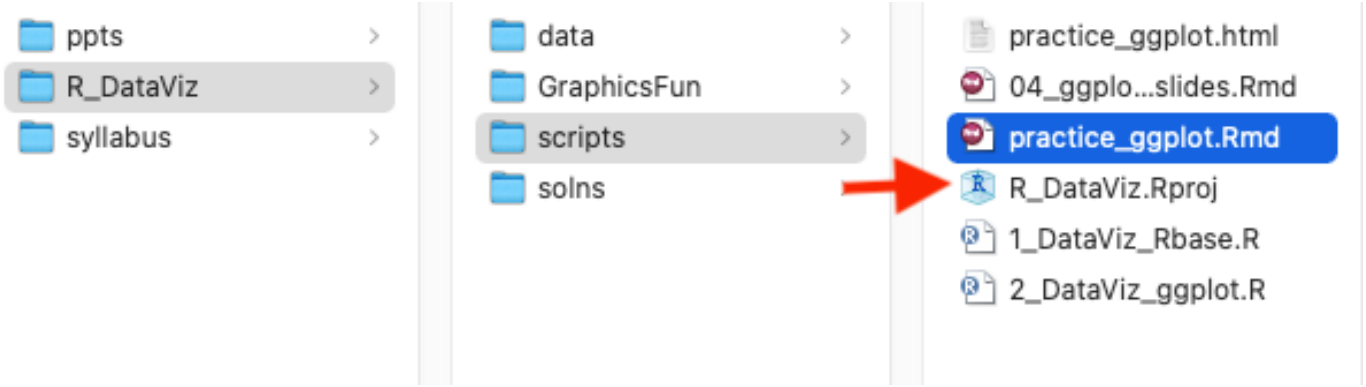
Working with .Rmd files

- By default, the working directory for R code chunks is the directory that contains the Rmd document.

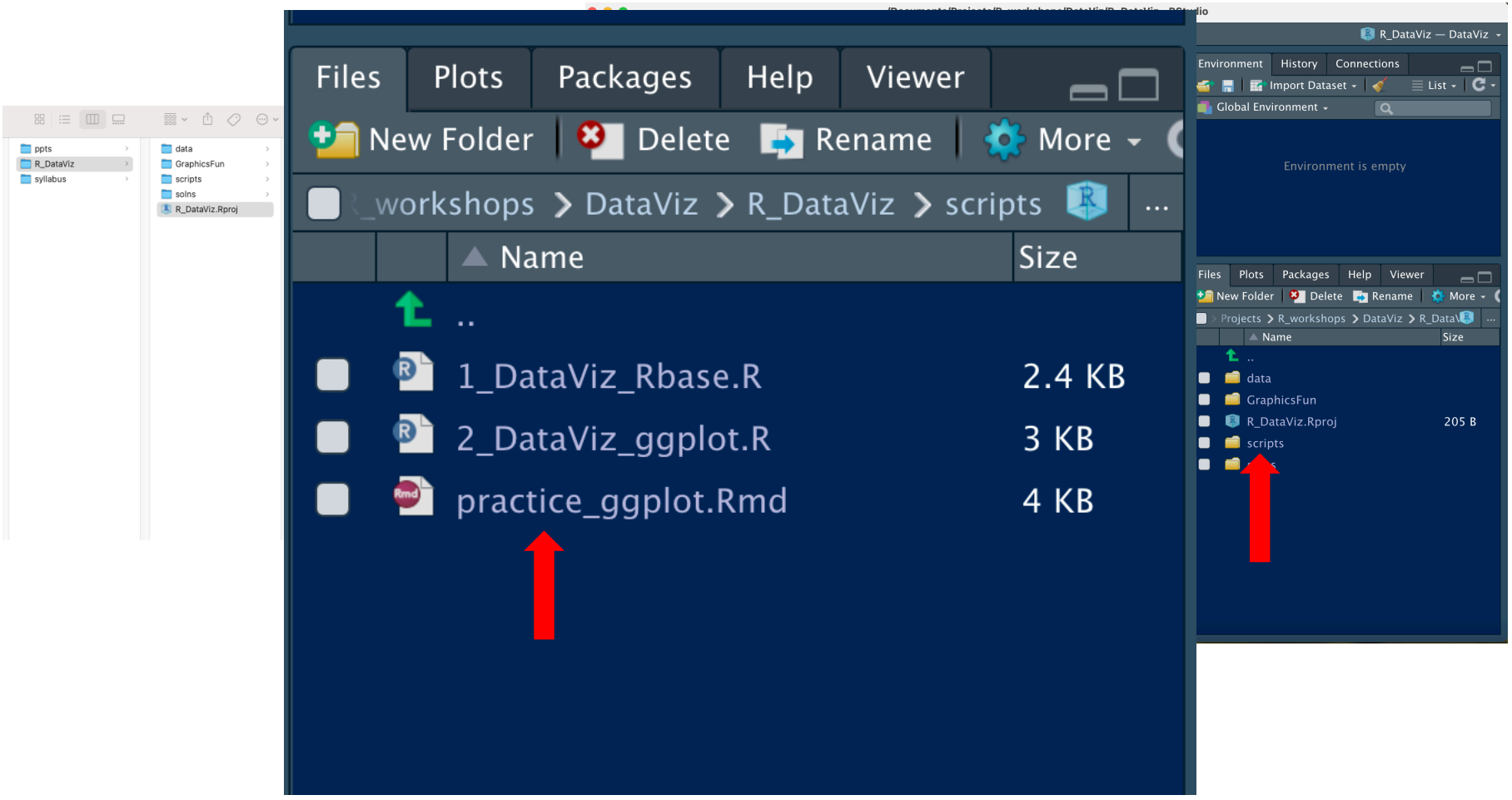
For example, if the path of an Rmd file is ~/scripts/practice_ggplot.Rmd, **the working directory** under which R code chunks are evaluated **is ~/scripts/** (not the directory that contains de .Rproj file).

- To avoid discrepancies with the working directories:

Move the .Rproj file to the scripts folder



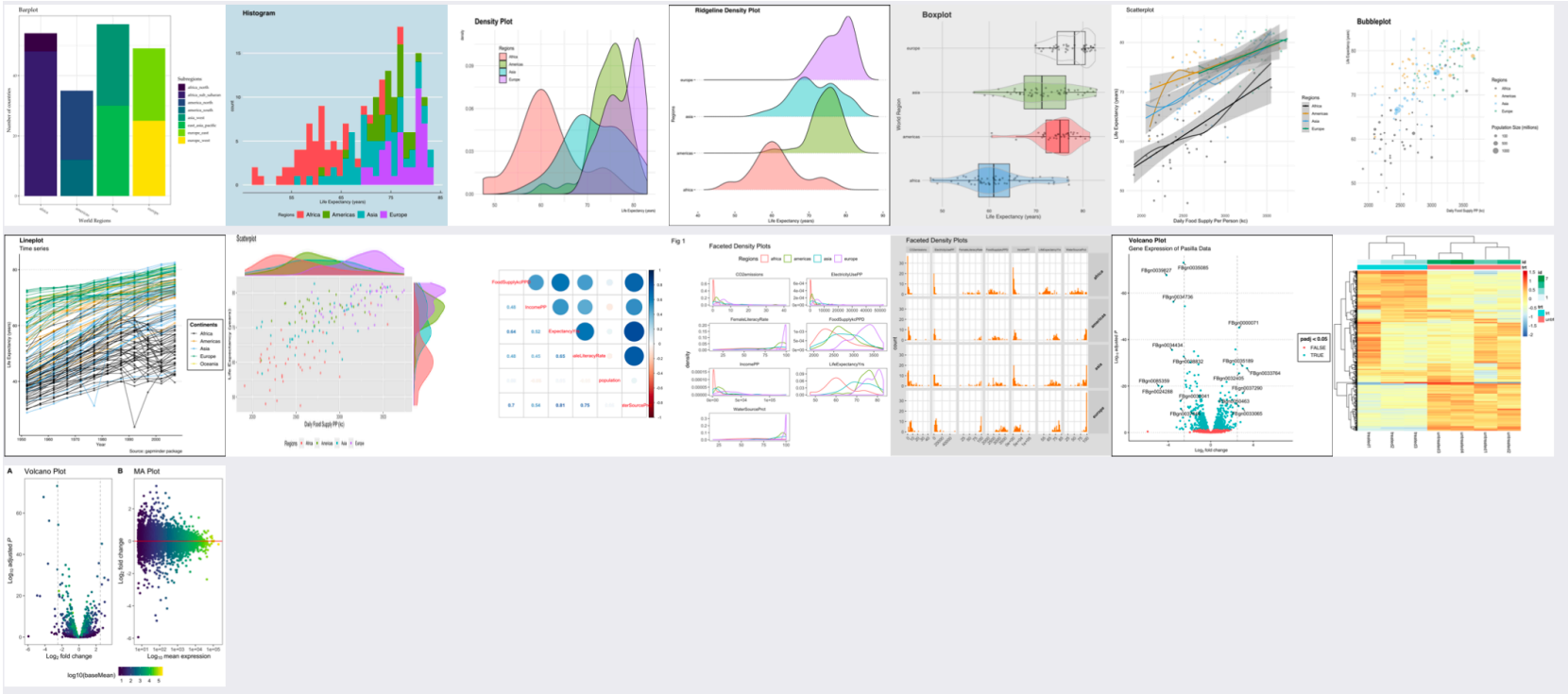
Open the practice_ggplot.Rmd file



Read in the dataset

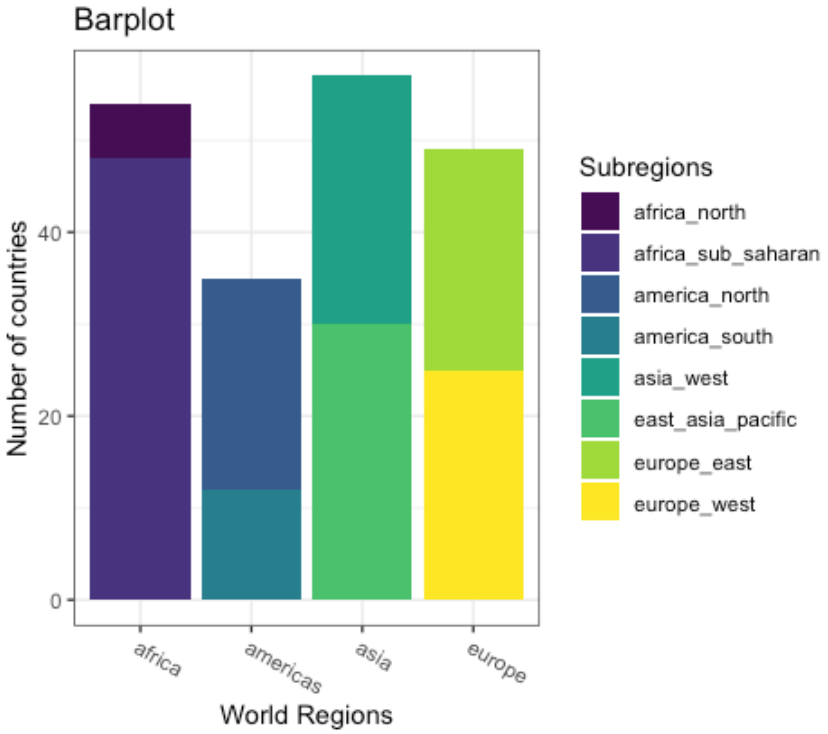
```
gapminder2011 <-
read_csv(”../data/Gapminder_vars_2011.csv"
)
```

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", ...

Exercise

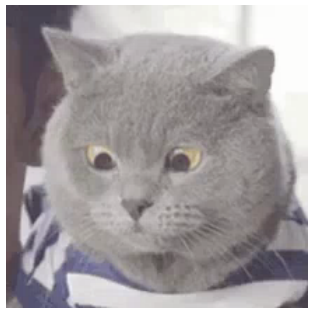
- use the variable `**members_oecd_g77**` for the bars

hint: `aes(y = members_oecd_g77,fill=…`

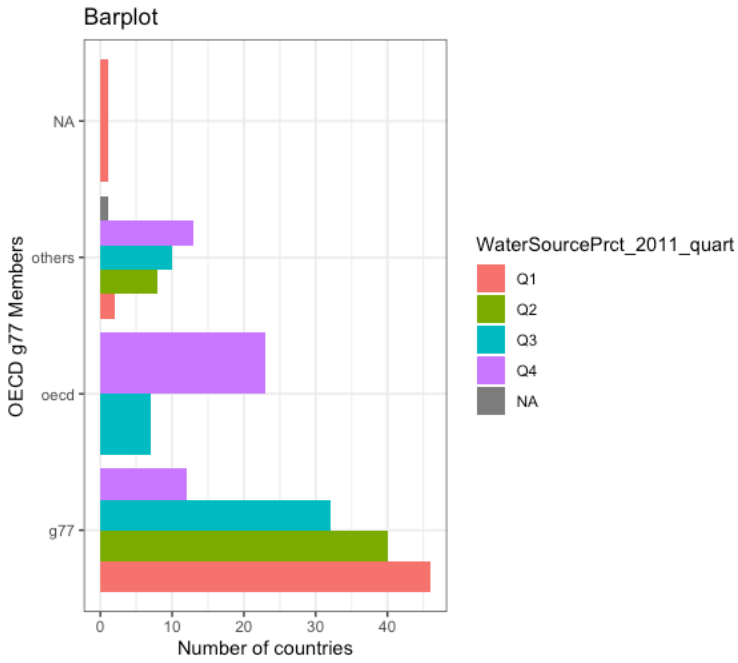
- change the fill to `**WaterSourcePrct_2011_quart**`

- use `position="dodge"` as an argument to `` `geom_bar()` ``

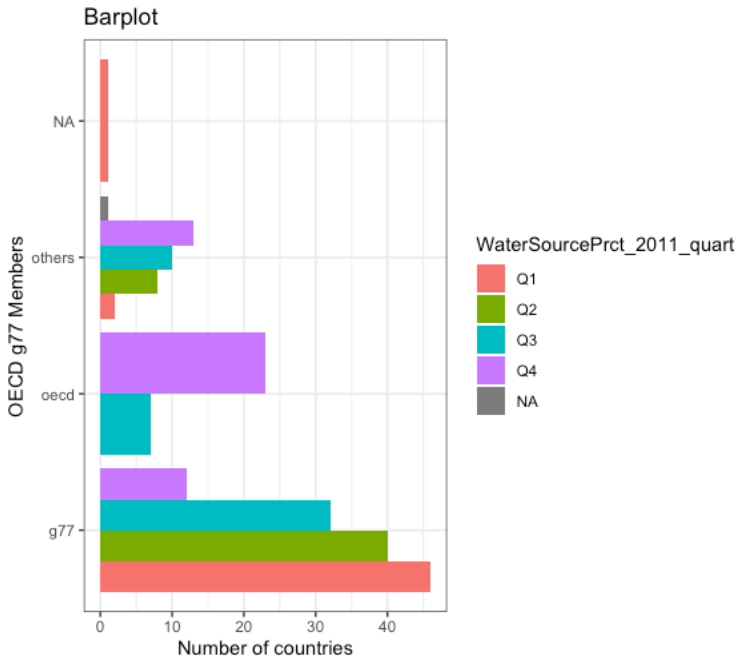
- change x axis label to "Number of countries" and the y axis label to "OECD g77 Members"



```
ggplot(data = gapminder2011,
       aes(y = members_oecd_g77,
           fill = WaterSourcePrct_2011_quart)) +
geom_bar(…
```



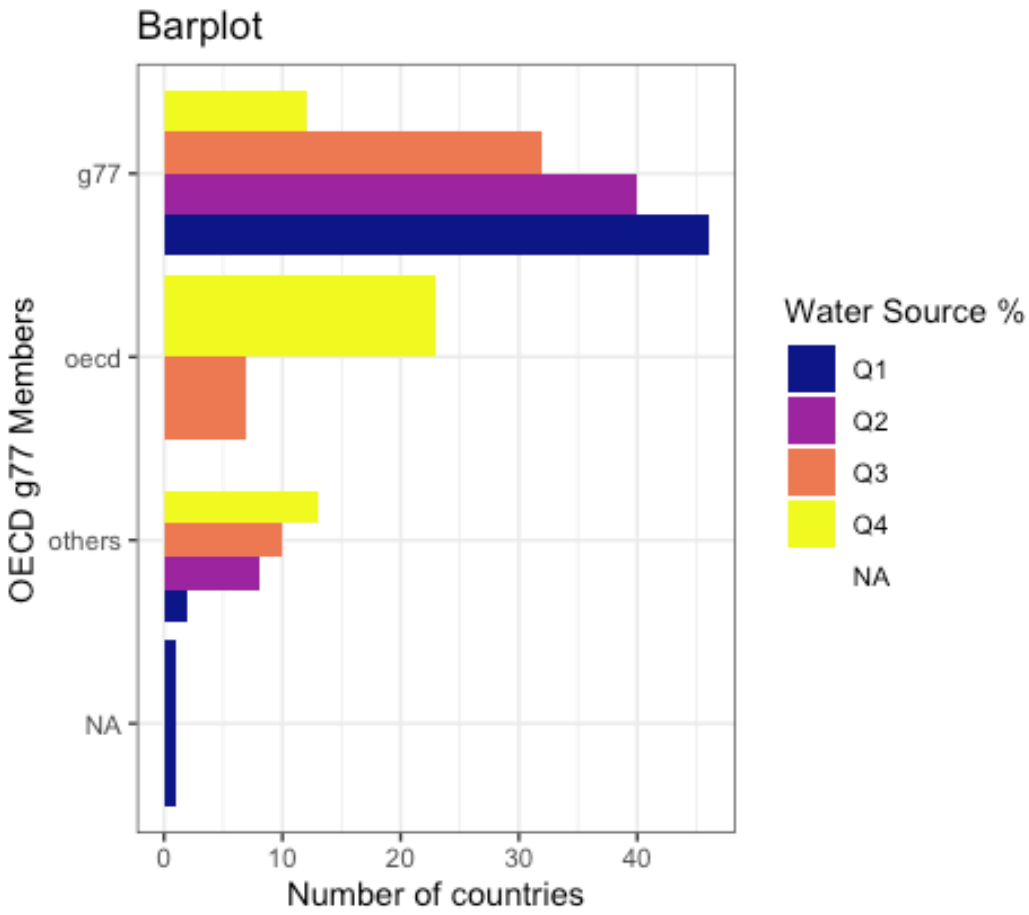
```
ggplot(data = gapminder2011,
       aes(y = members_oecd_g77,
           fill = WaterSourcePrct_2011_quart)) +
geom_bar(position="dodge") +
labs(x = "Number of countries",
       y = "OECD g77 Members",
       title = "Barplot") +
  theme_bw()
```



Homework

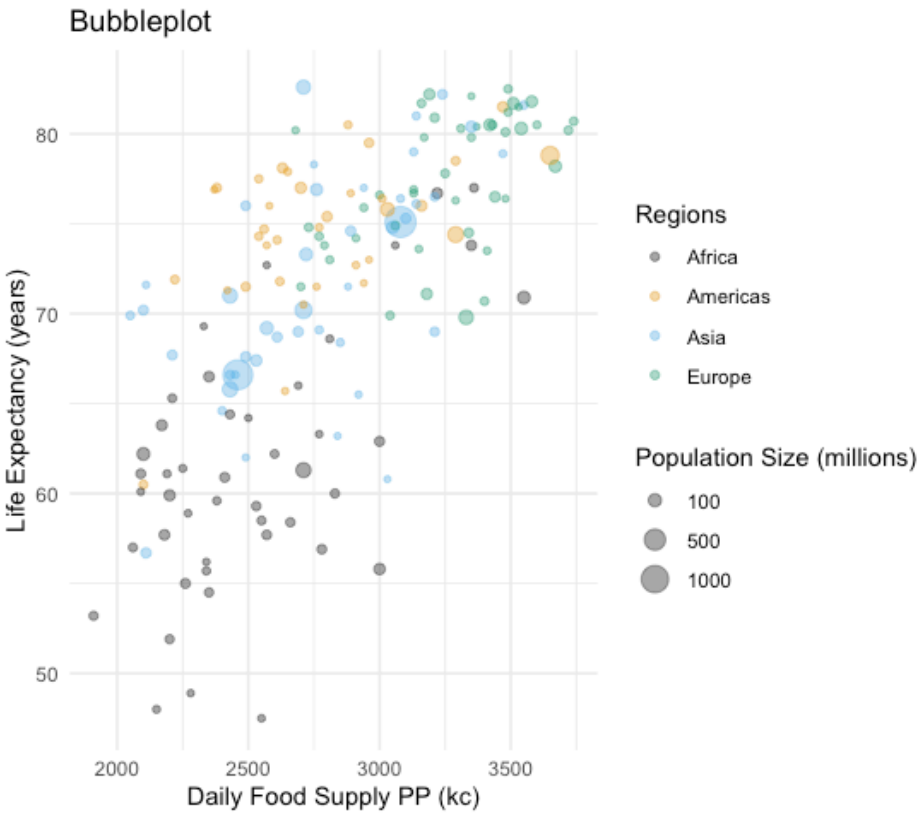
- change the order of the y axis with `scale_y_discrete(limits = c(NA,"others","oecd","g77"))`
- change the color of the bars to a different palette

hint: type `?scale_fill_viridis_d()` and look for the argument `option=""`

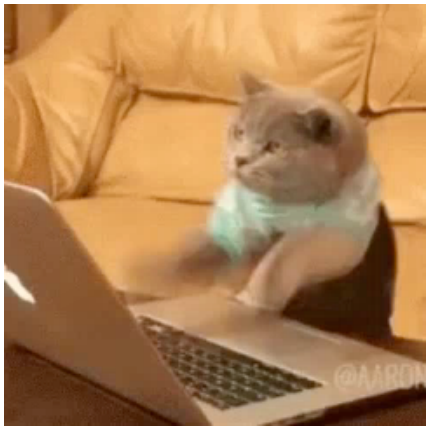


Bubbleplot

```
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"
    )
```



Exercise

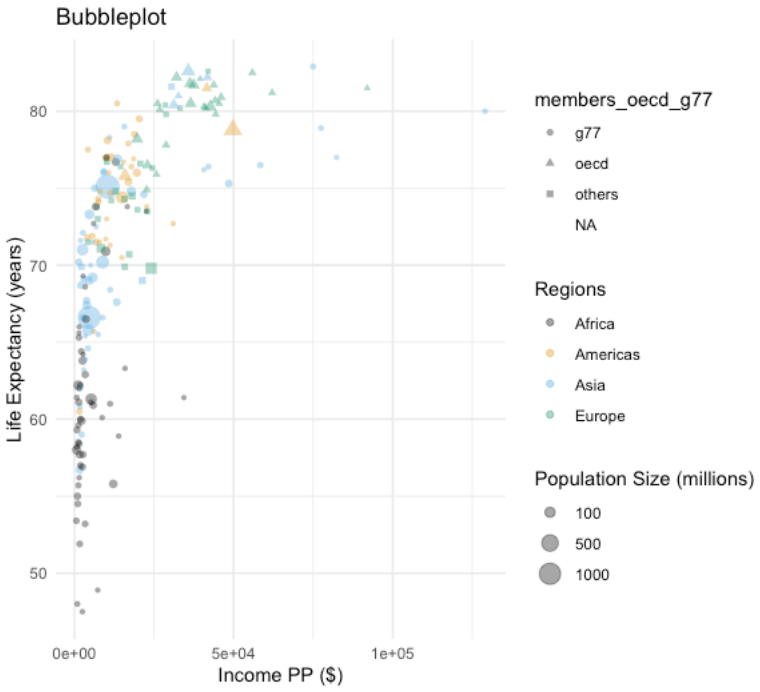


- Change x = `IncomePP`
- Change the x label to be accurate
- Map shape to `members_oecd_g77` aes

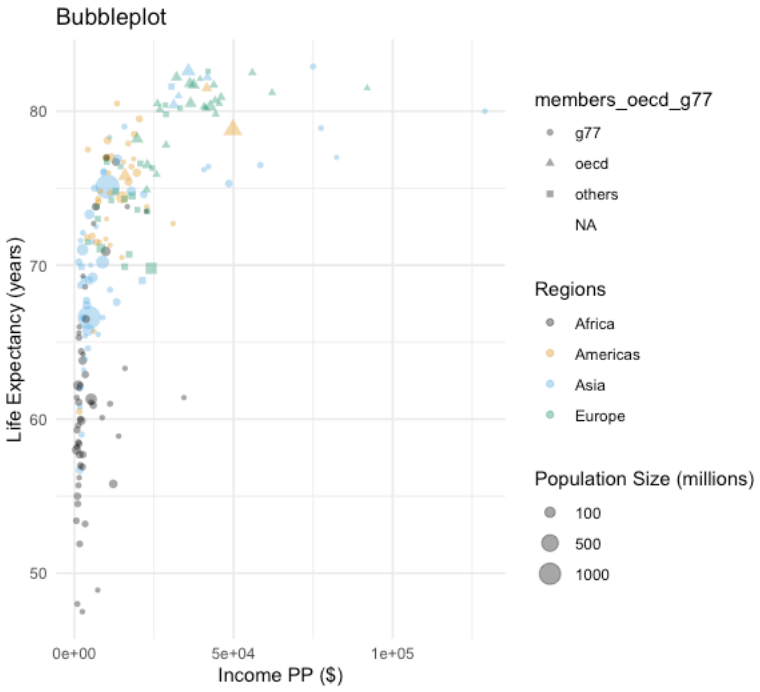
Hint: use shape in:

```
aes(…size = population,
      shape =…
```

```
ggplot(data = gapminder2011,
       aes(x = IncomePP,
           y = LifeExpectancyYrs,
           color = four_regions,
           size = population,
           shape = members_oecd_g77)) +
```

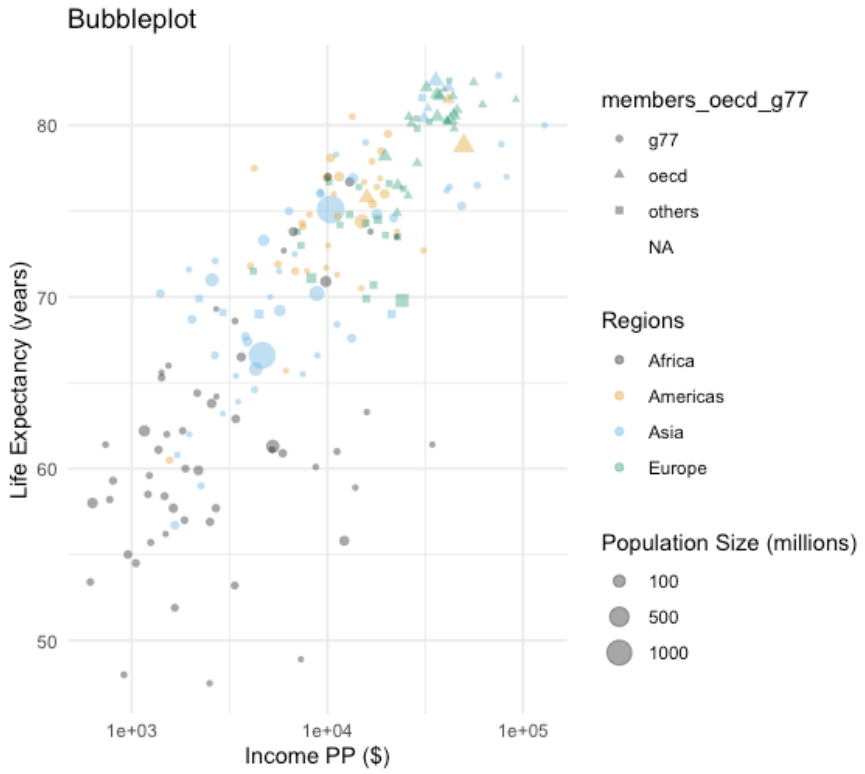


```r
ggplot(data = gapminder2011,
       aes(x = IncomePP,
           y = LifeExpectancyYrs,
           color = four_regions,
           size = population,
           shape = members_oecd_g77)) +
  geom_point(alpha = 0.4) +
  scale_color_colorblind(
    name = "Regions",
    labels = c("Africa", "Americas",
               "Asia", "Europe")
  ) +
  scale_size(
    name = "Population Size (millions)",
    breaks = c(1e08,5e08,1e09),
    labels = c(100,500,1000)
  ) +
  theme_minimal() +
  labs(
    x = "Income PP ($)",
    y = "Life Expectancy (years)",
    title = "Bubbleplot"
    )
```



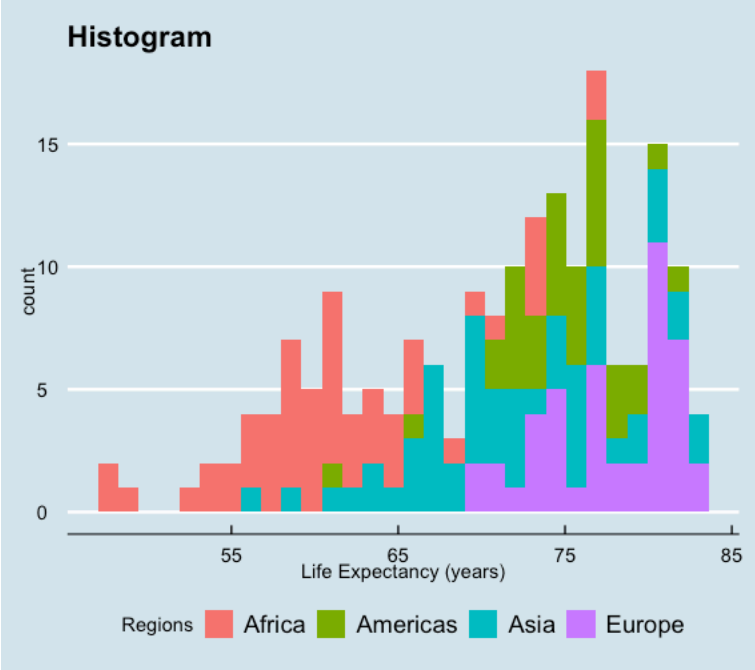
Homework

Change the scale of the x-axis to be on the log10 scale using `` `scale_x_log10()` ``



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"
  )
```



Exercise

- color by `four_regions`, fill by `eight_regions`

- change the width of the histogram bins to 1.5

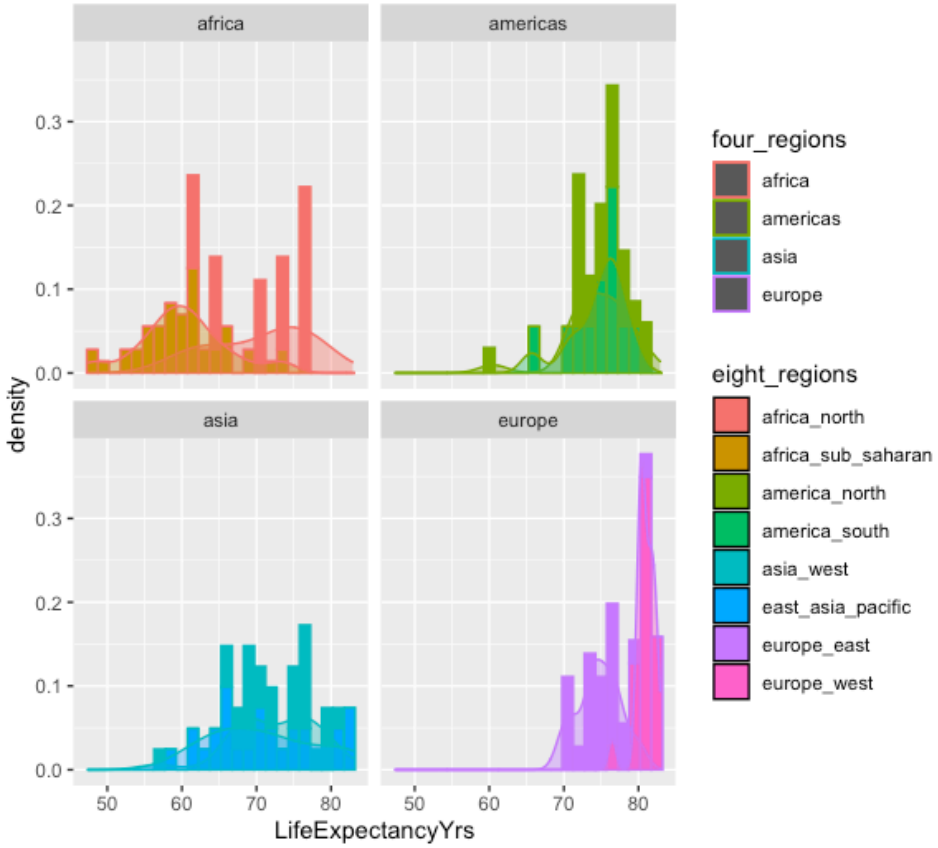
(hint: type `?geom_histogram` in the console to find the argument)

- add a layer of `geom_density()` with `alpha = 0.4`

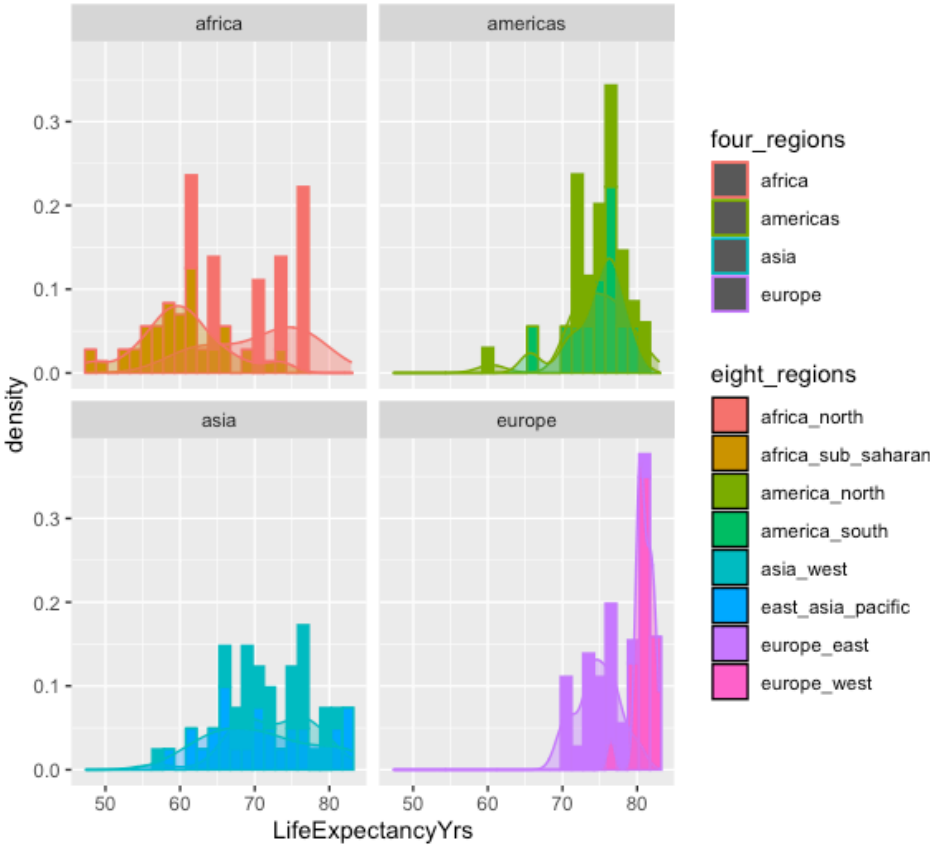
Why doesn't the density line show up? Add `aes(y=..density..)` to the histogram function arguments.

- facet by `four_regions` `facet_wrap(~)`

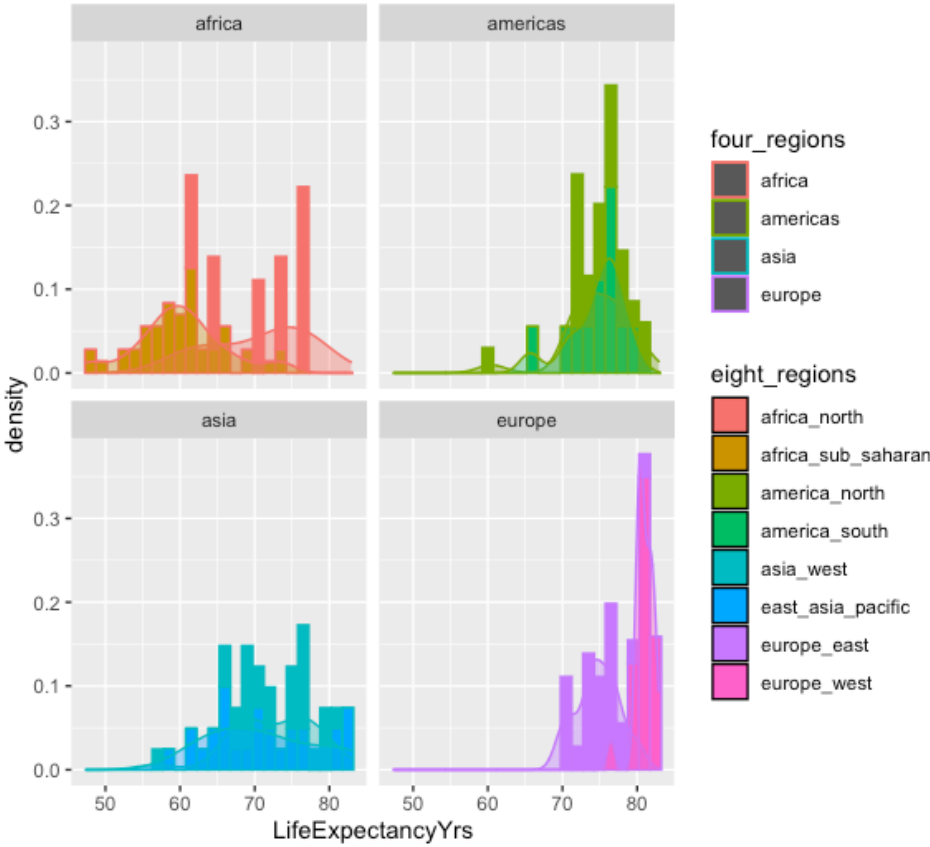
```r
ggplot(data = gapminder2011,
       aes(x = LifeExpectancyYrs,
           color = four_regions,
           fill = eight_regions
           )) +
  # geom_histogram(binwidth = 1.5) +
  geom_histogram(aes(y=..density..),
                 binwidth = 1.5) +
```



```
ggplot(data = gapminder2011,
       aes(x = LifeExpectancyYrs,
           color = four_regions,
           fill = eight_regions
           )) +
  #geom_histogram(binwidth = 1.5) +
  geom_histogram(aes(y=..density..),
                 binwidth = 1.5) +
  geom_density(alpha = 0.4) +
```

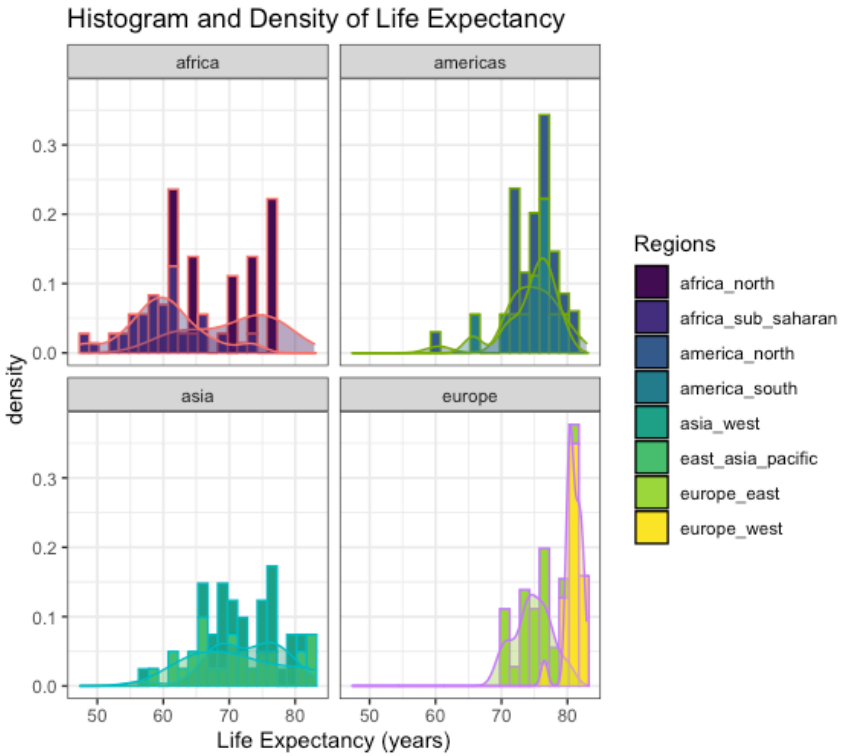


```
ggplot(data = gapminder2011,
       aes(x = LifeExpectancyYrs,
           color = four_regions,
           fill = eight_regions
           )) +
  # geom_histogram(binwidth = 1.5) +
  geom_histogram(aes(y=..density..),
                 binwidth = 1.5) +
  geom_density(alpha = 0.4) +
  facet_wrap(~four_regions) +
```



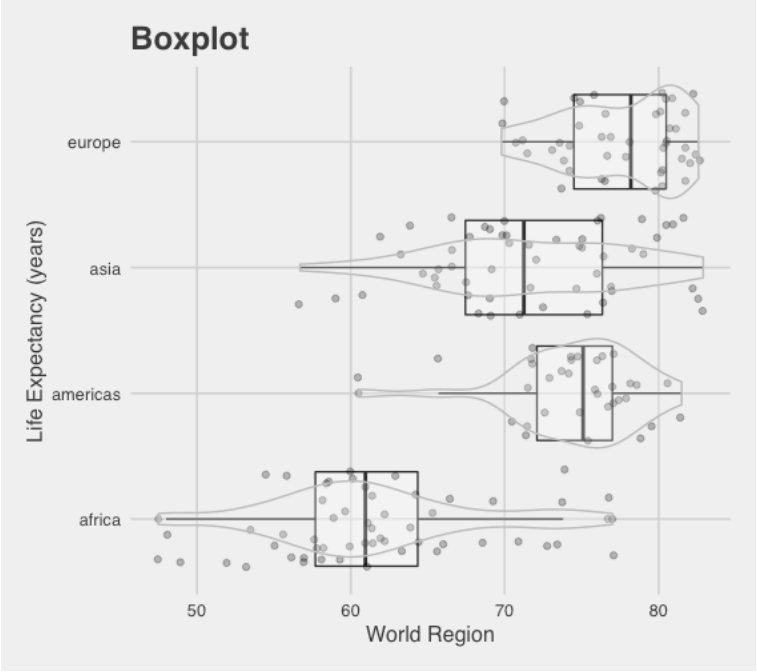
Homework

- change the theme to `theme_bw()`
- change the fill scale to viridis (discrete version is `scale_fill_viridis_d`)
- remove the color legend
- change the title to "Histogram and Density of Life Expectancy"



Boxplot

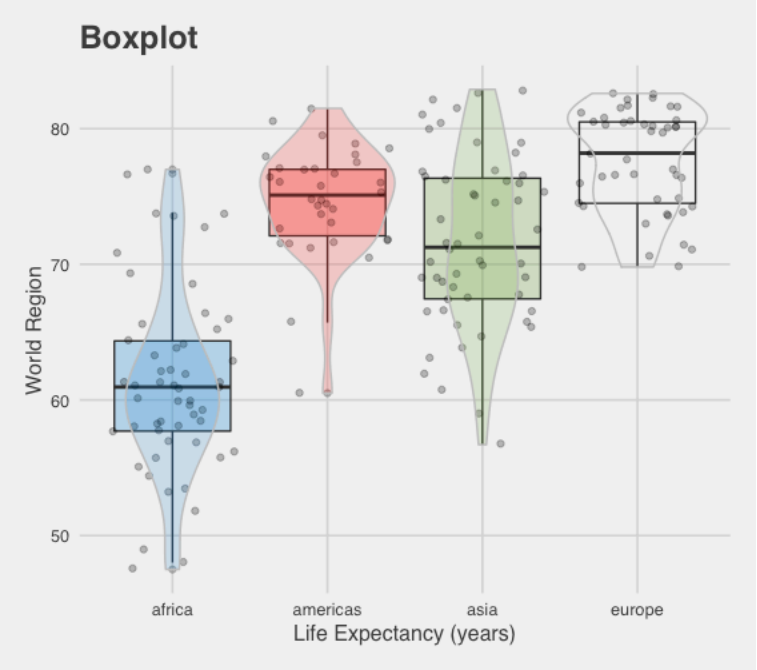
```
ggplot(data = gapminder2011,
       aes(x = LifeExpectancyYrs, # New!
           y = four_regions
           )
       ) +
  geom_boxplot(alpha = 0.3) +
  theme_fivethirtyeight() +
  theme(axis.title = element_text()) +
  scale_fill_fivethirtyeight() +
  theme(legend.position = "none") +
  geom_jitter(width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



Exercise

- fill by `four_regions`
- make the boxplot vertical instead of horizontal

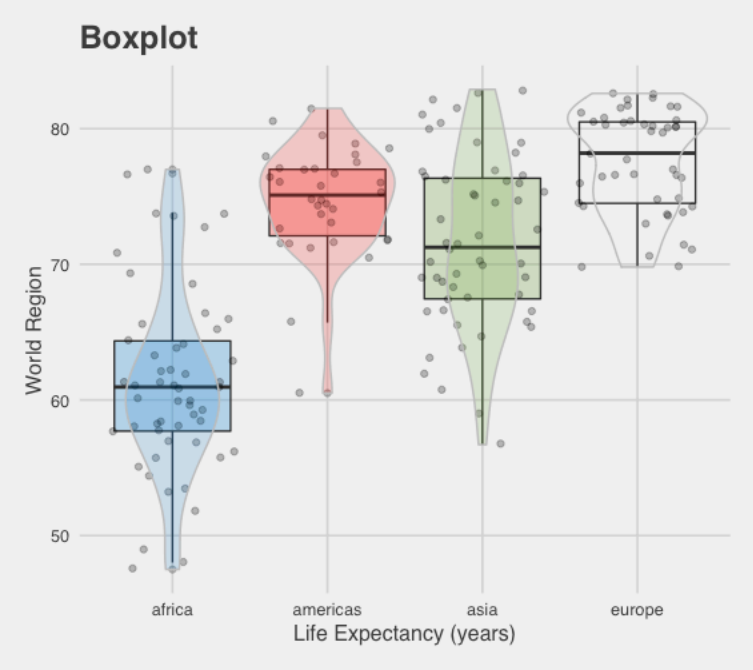
hint: use `coord_flip()`



Exercise

- fill by `four_regions`
- make the boxplot vertical instead of horizontal

```
ggplot(data = gapminder2011,
       aes(x = LifeExpectancyYrs, # New!
           y = four_regions,
           fill = four_regions)
       ) +
  geom_boxplot(alpha = 0.3) + # add outlier.shape = NA
  coord_flip() +
  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

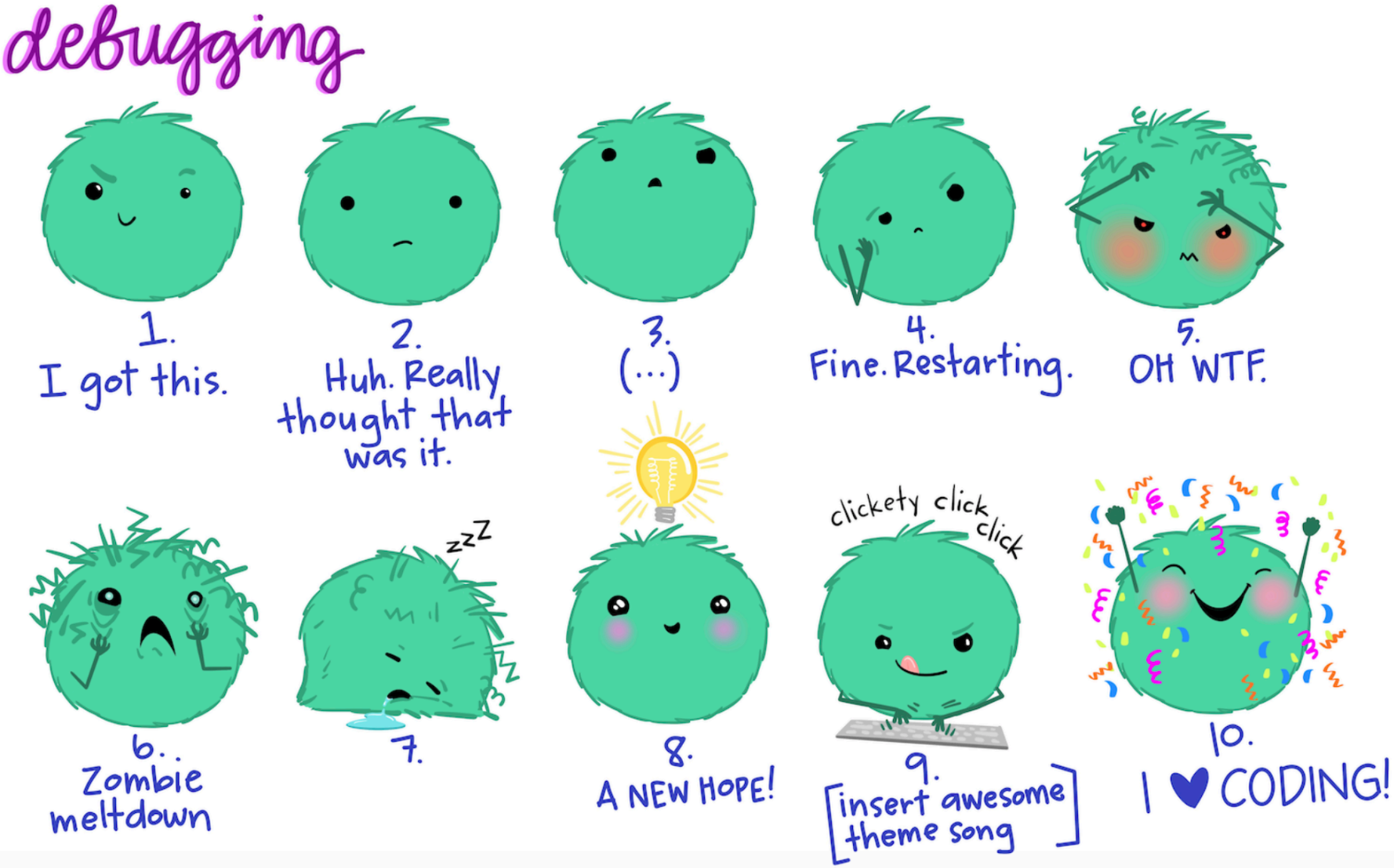


Illustration by Allison Horst