Introduction to using R and RStudio

Matthew Galbraith
Linda Crnic Institute for Down Syndrome

Data Science for Developing Scholars in Down Syndrome Research (DS3) 2025

INCLUDE
INvestigation of Co-occurring conditions across the Lifespan to Understand Down syndromE

Linda Crnic Institute for Down Syndrome
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

GLOBAL
DOWN SYNDROME FOUNDATION®

Links for this session

https://github.com/DS3-2025/installing_updating_R-RStudio
https://docs.posit.co/ide/user/ide/get-started/
https://rstudio.github.io/cheatsheets/base-r.pdf

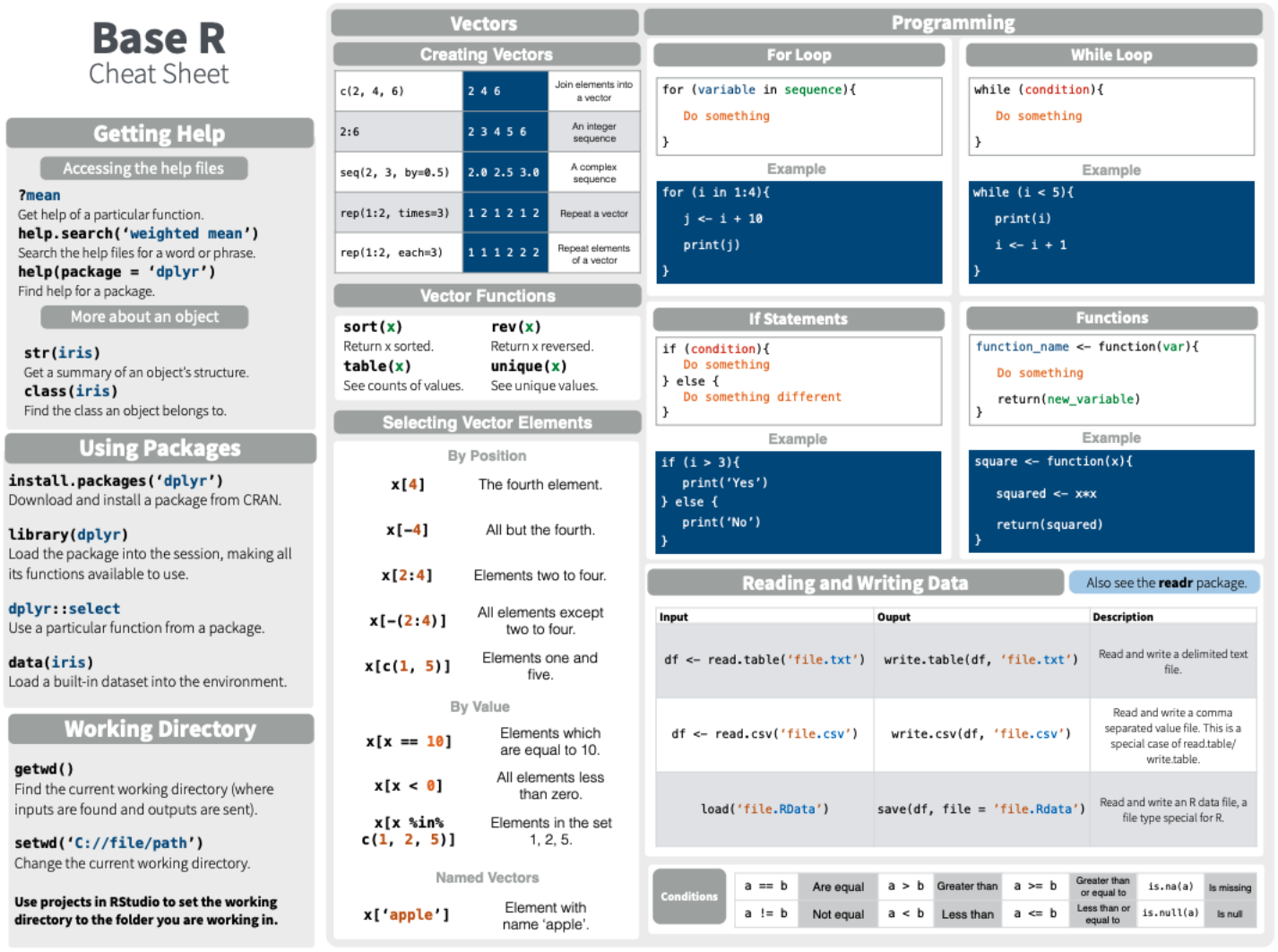
https://posit.co/download/rstudio-desktop/
https://rstudio.github.io/cheatsheets/html/rstudio-ide.html

https://support.rstudio.com/hc/en-us/articles/200526207-Using-RStudio-Projects
https://r4ds.had.co.nz/workflow-projects.html
https://github.com/DS3-2025/Rproject_template

https://raw.githubusercontent.com/rstudio/cheatsheets/main/data-visualization.pdf
https://www.data-to-viz.com/caveats.html
https://r-graph-gallery.com/index.html

https://tidyverse.tidyverse.org/

R



Base R
Cheat Sheet

Getting Help

Accessing the help files

?mean
Get help of a particular function.
help.search('weighted mean')
Search the help files for a word or phrase.
help(package = 'dplyr')
Find help for a package.

More about an object

str(iris)
Get a summary of an object's structure.
class(iris)
Find the class an object belongs to.

Using Packages

install.packages('dplyr')
Download and install a package from CRAN.

library(dplyr)
Load the package into the session, making all its functions available to use.

dplyr::select
Use a particular function from a package.

data(iris)
Load a built-in dataset into the environment.

Working Directory

getwd()
Find the current working directory (where inputs are found and outputs are sent).

setwd('C://file/path')
Change the current working directory.

Use projects in RStudio to set the working directory to the folder you are working in.

Vectors

Creating Vectors

c(2, 4, 6)	2 4 6	Join elements into a vector
2:6	2 3 4 5 6	An integer sequence
seq(2, 3, by=0.5)	2.0 2.5 3.0	A complex sequence
rep(1:2, times=3)	1 2 1 2 1 2	Repeat a vector
rep(1:2, each=3)	1 1 1 2 2 2	Repeat elements of a vector

Vector Functions

sort(x)
Return x sorted.
rev(x)
Return x reversed.
table(x)
See counts of values.
unique(x)
See unique values.

Selecting Vector Elements

By Position

x[4]	The fourth element.
x[-4]	All but the fourth.
x[2:4]	Elements two to four.
x[-(2:4)]	All elements except two to four.
x[c(1, 5)]	Elements one and five.

By Value

x[x == 10]	Elements which are equal to 10.
x[x < 0]	All elements less than zero.
x[x %in% c(1, 2, 5)]	Elements in the set 1, 2, 5.

Named Vectors

x['apple']	Element with name 'apple'.

Programming

For Loop

```
for (variable in sequence){
   Do something
}
```

Example

```
for (i in 1:4){
   j <- i + 10
   print(j)
}
```

While Loop

```
while (condition){
   Do something
}
```

Example

```
while (i < 5){
   print(i)
   i <- i + 1
}
```

If Statements

```
if (condition){
   Do something
} else {
   Do something different
}
```

Example

```
if (i > 3){
   print('Yes')
} else {
   print('No')
}
```

Functions

```
function_name <- function(var){
   Do something
   return(new_variable)
}
```

Example

```
square <- function(x){
   squared <- x*x
   return(squared)
}
```

Reading and Writing Data

Also see the **readr** package.

Input	Ouput	Description
df <- read.table('file.txt')	write.table(df, 'file.txt')	Read and write a delimited text file.
df <- read.csv('file.csv')	write.csv(df, 'file.csv')	Read and write a comma separated value file. This is a special case of read.table/ write.table.
load('file.RData')	save(df, file = 'file.Rdata')	Read and write an R data file, a file type special for R.

Conditions

a == b	Are equal	a > b	Greater than	a >= b	Greater than or equal to	is.na(a)	Is missing
a != b	Not equal	a < b	Less than	a <= b	Less than or equal to	is.null(a)	Is null

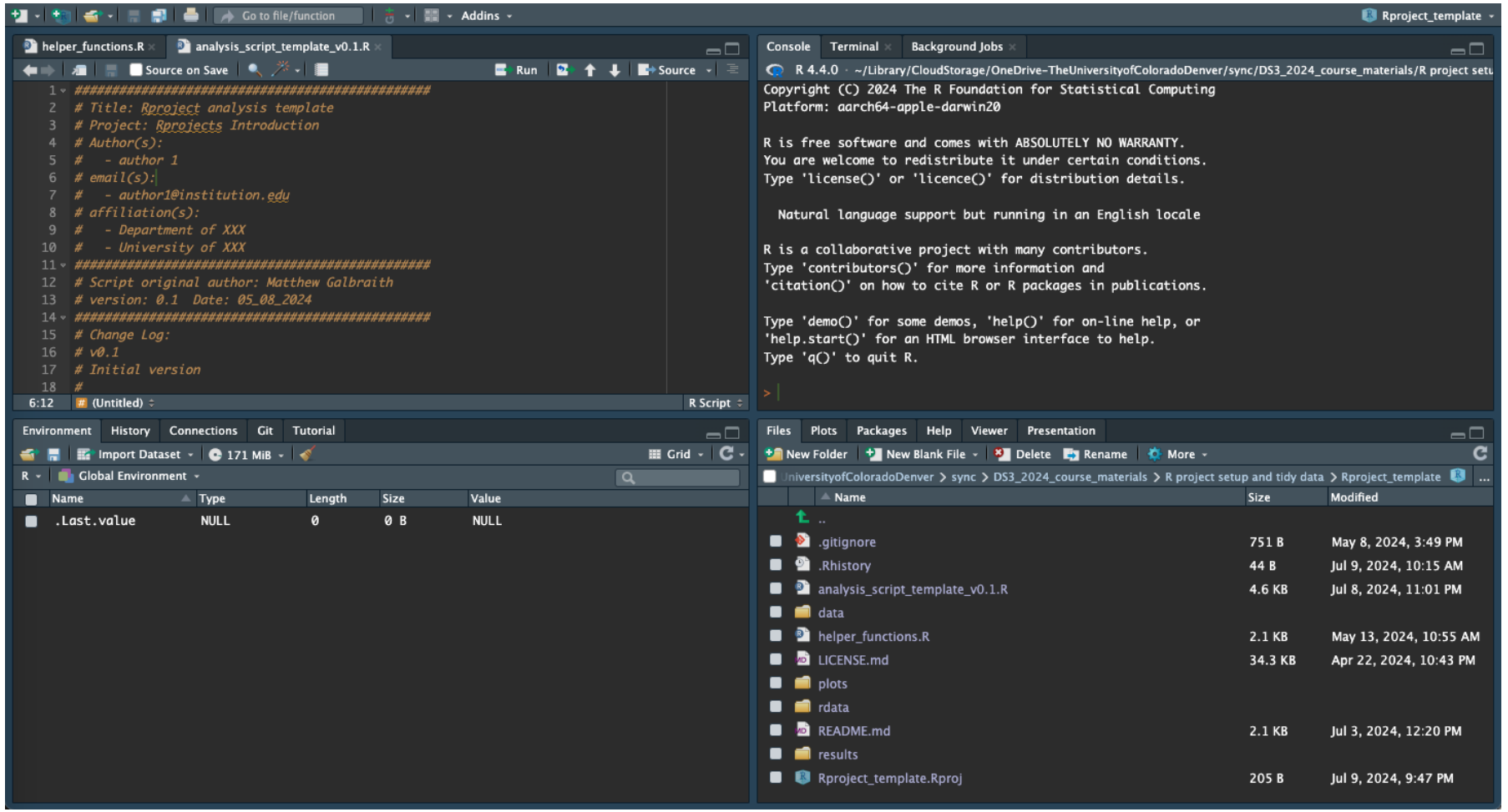
Learn more at web page or vignette • package version • Updated: 3/15

https://rstudio.github.io/cheatsheets/base-r.pdf

RStudio

Integrated Development Environment (IDE) for R

- Edit + execute R code
- Syntax highlighting, code completion, debugging
- View output, plots, help, environment



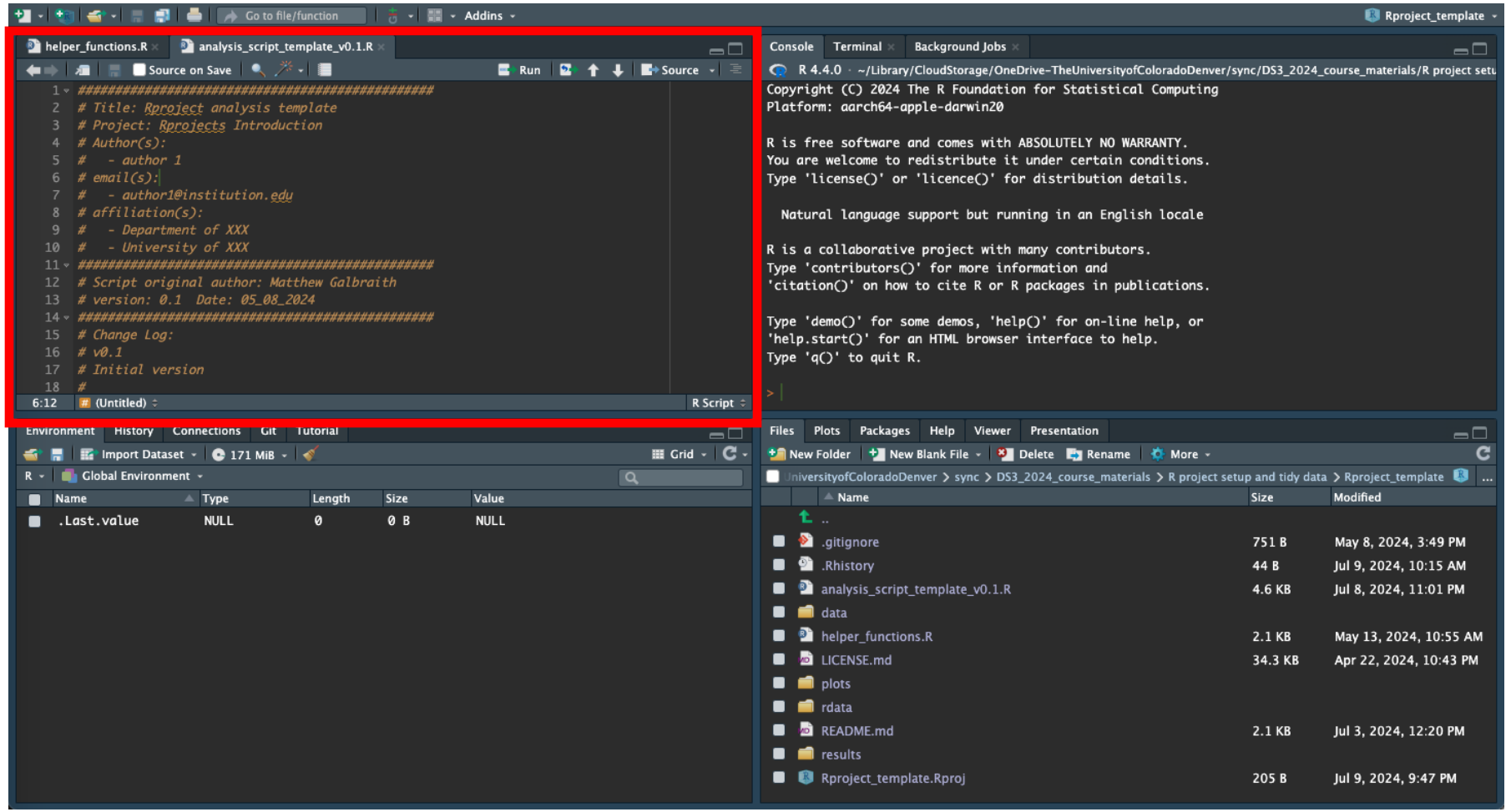
https://posit.co/download/rstudio-desktop/

RStudio

Integrated Development Environment (IDE) for R

- Edit + execute R code
- Syntax highlighting, code completion, debugging
- View output, plots, help, environment

Source pane

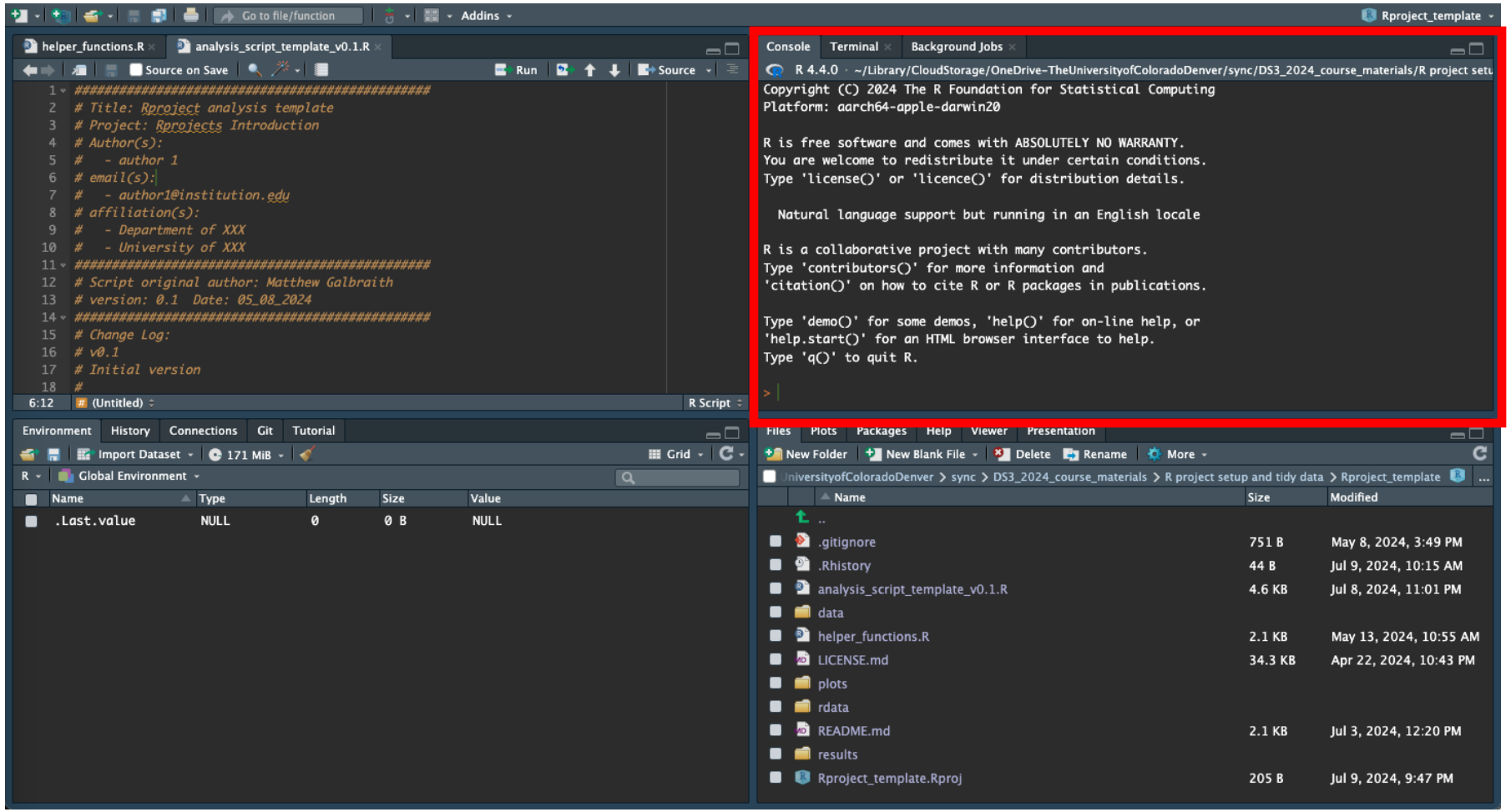


Panes can be rearranged and customized

RStudio

Integrated Development Environment (IDE) for R

- Edit + execute R code
- Syntax highlighting, code completion, debugging
- View output, plots, help, environment



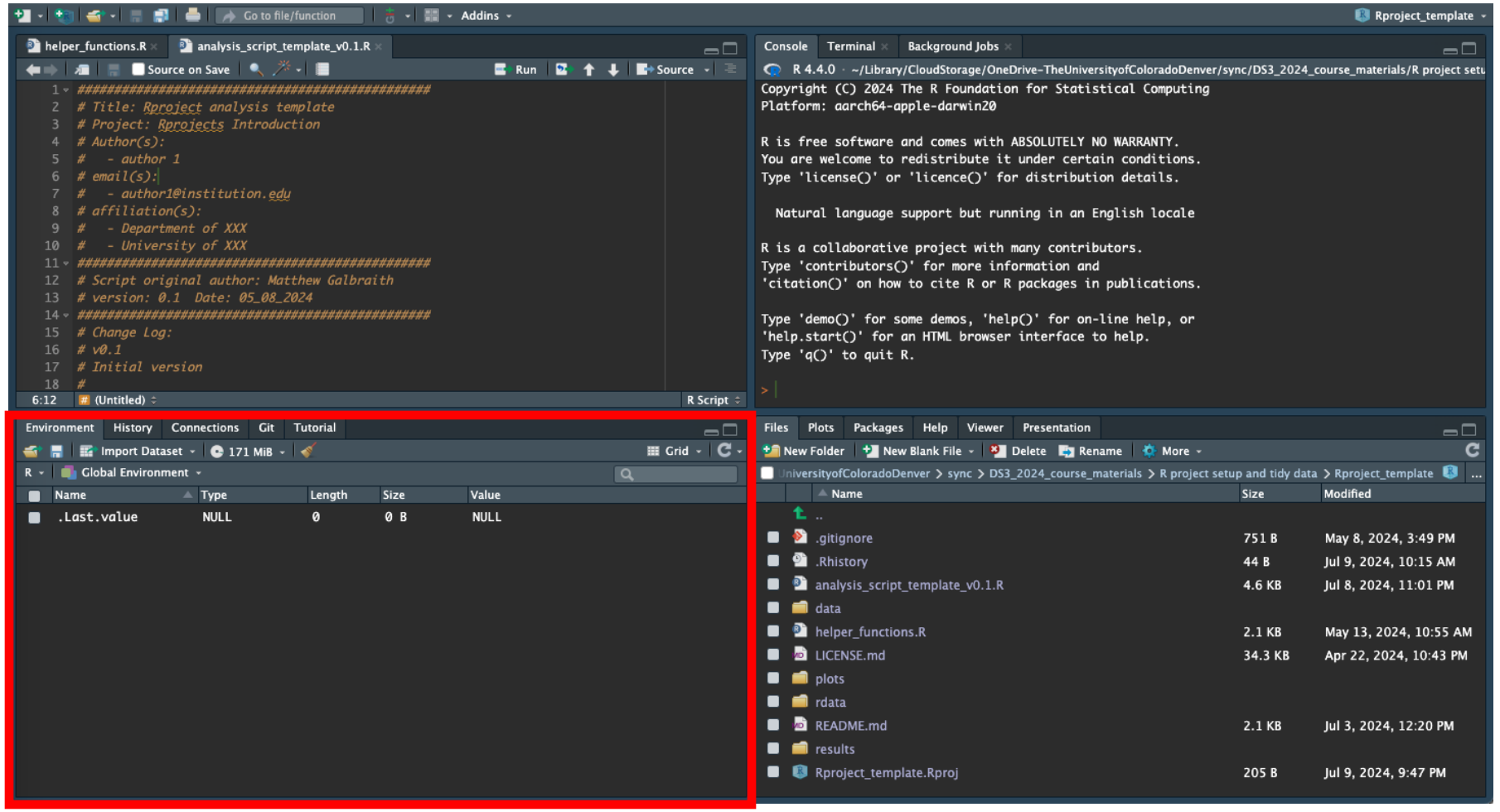
Console pane

RStudio

Integrated Development Environment (IDE) for R

- Edit + execute R code
- Syntax highlighting, code completion, debugging
- View output, plots, help, environment

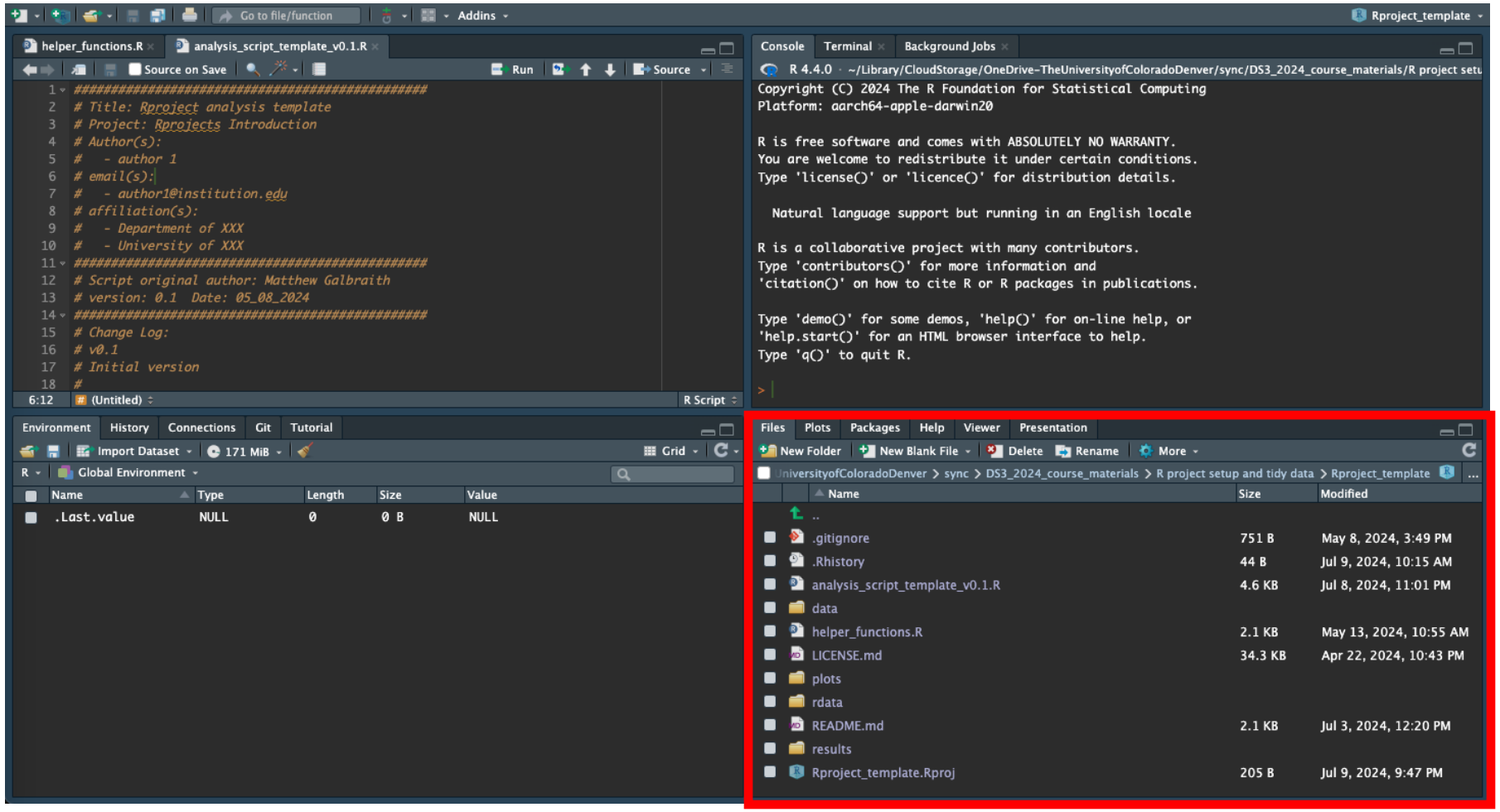
Environment etc pane



RStudio

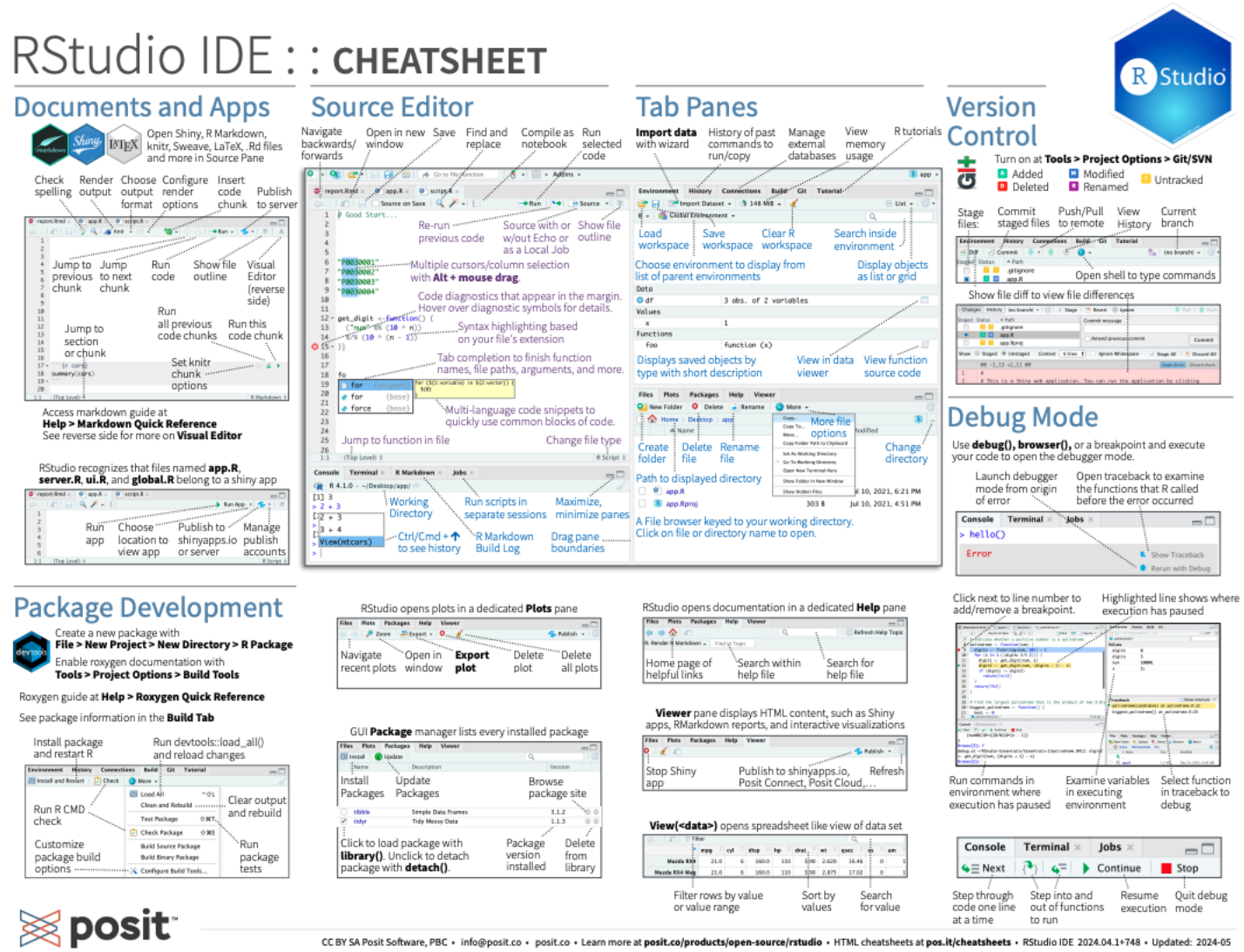
Integrated Development Environment (IDE) for R

- Edit + execute R code
- Syntax highlighting, code completion, debugging
- View output, plots, help, environment



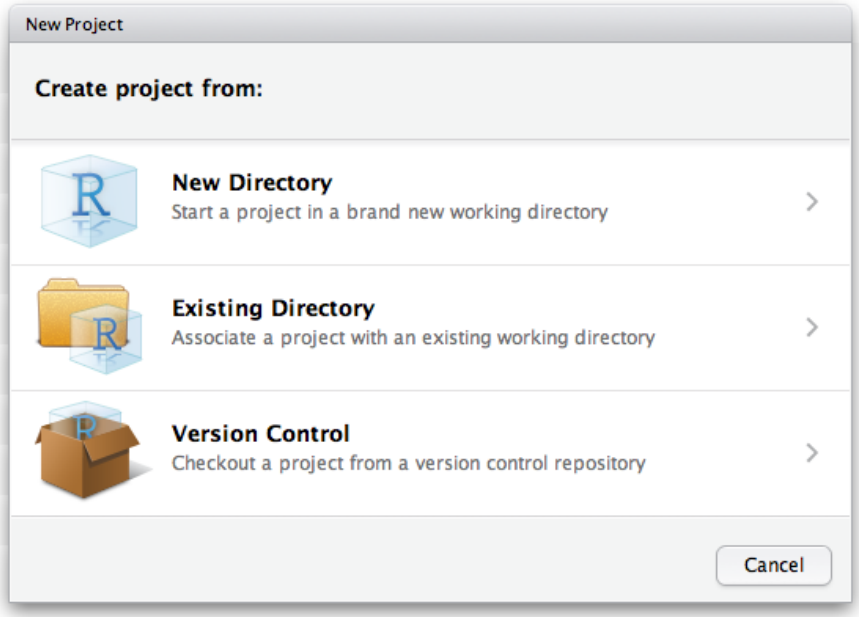
Files, Plots, Help etc pane

RStudio cheatsheet



https://rstudio.github.io/cheatsheets/html/rstudio-ide.html

RStudio Projects



- Open existing projects via .Rproj file
- Automatically sets your working directory
- Self-contained set of directories, scripts, and data files (very important for multiple projects)

Organizing your Rstudio Projects

- Only **/data** and R scripts are required - everything else can be recreated (incl. earlier versions)
- Treat **/data** directory as read-only
- Analysis outputs go to **/results** or **/plots** (with version info)
- R workspace and large intermediate files stored in **/rdata**
- Additional directories added as needed, eg /Archive
- Compatible with manual or other version control

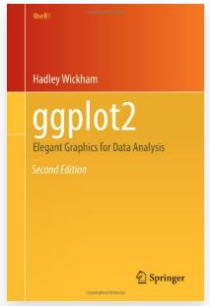
Project_directory
- **/data**
- /results
- /plots
- /rdata
- **analysis_script.R**
- **helper_functions.R**
- project.Rproj

https://support.rstudio.com/hc/en-us/articles/200526207-Using-RStudio-Projects
https://r4ds.had.co.nz/workflow-projects.html
https://github.com/DS3-2025/Rproject_template

Into the Tidyverse: Visualization using ggplot2

Publication-quality data visualization

- Implements a “grammar of graphics”
- Start by defining the data to be plotted (“aesthetics”):
  `ggplot(aes(x, y, color, fill, shape, alpha, linetype))`
- Then add layers (“geoms”) to specify how data is plotted, eg:
  `+ geom_point()`
- Add additional geom layers, eg:
  `+ geom_boxplot()`
- Can split into separate plots, eg male vs. female, by “faceting”:
  `+ facet_wrap(~ Sex)`
- Finally add title and modify theme:
  ```
  + labs(title = “Plot title”, subtitle = “plot details”
  + theme(aspect.ratio = 1)
  ```

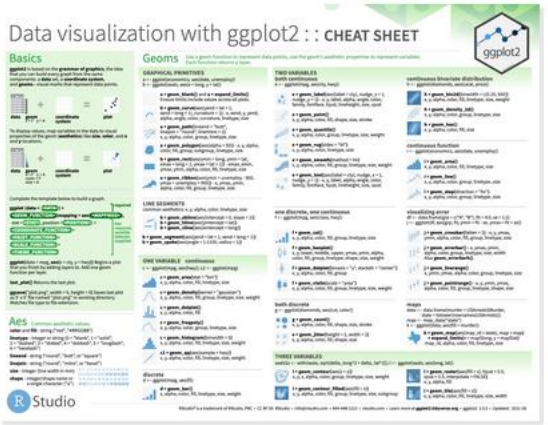


https://ggplot2-book.org/

https://raw.githubusercontent.com/rstudio/cheatsheets/main/data-visualization.pdf

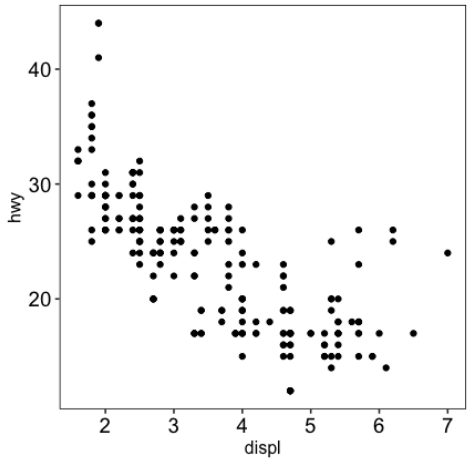
https://www.data-to-viz.com/caveats.html

https://r-graph-gallery.com/index.html



Visualization using ggplot2


```
mpg %>%
    ggplot(aes(displ, hwy)) +
    geom_point()
```




```
mpg %>%
    ggplot(aes(displ, hwy, color = class)) +
    geom_point() +
    theme(aspect.ratio = 1) +
    labs(title = "Highway mpg (hwy) vs. displacement")
```