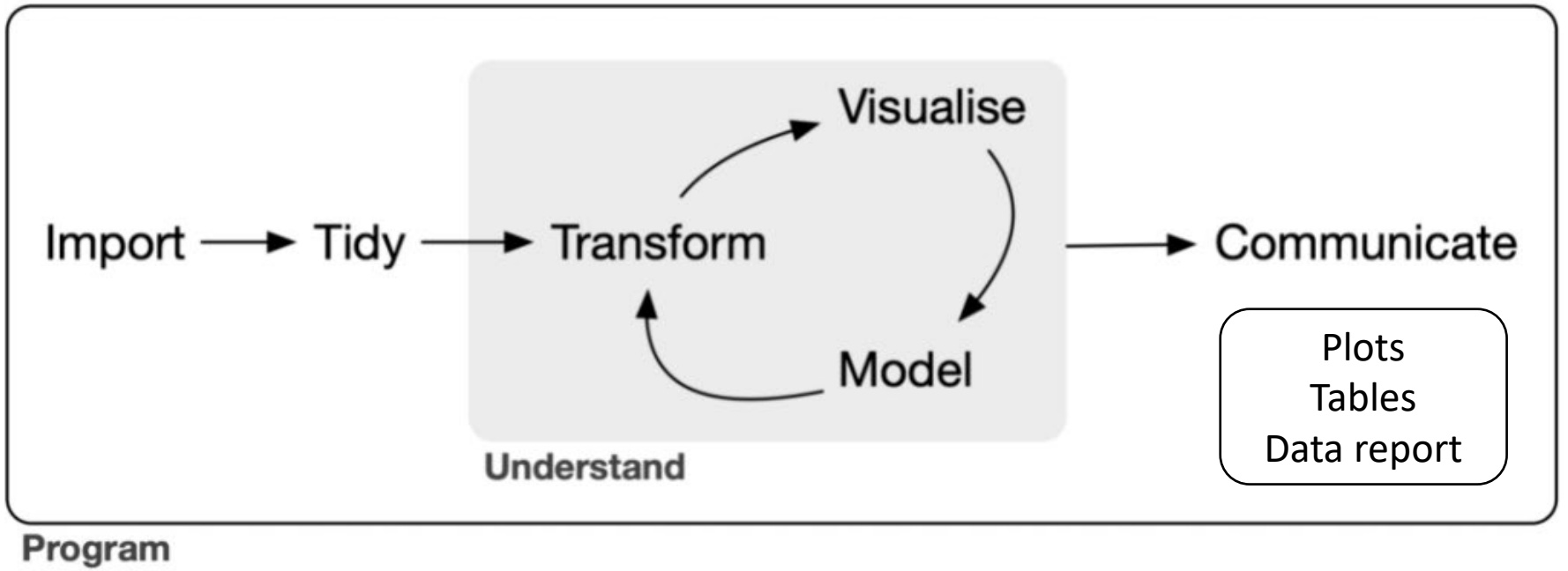
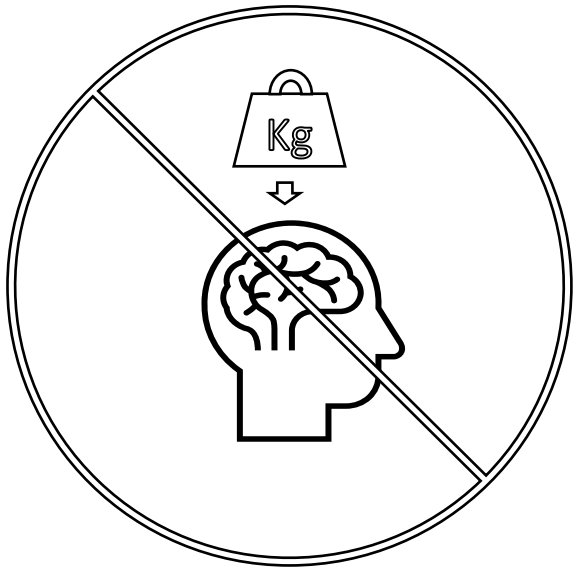
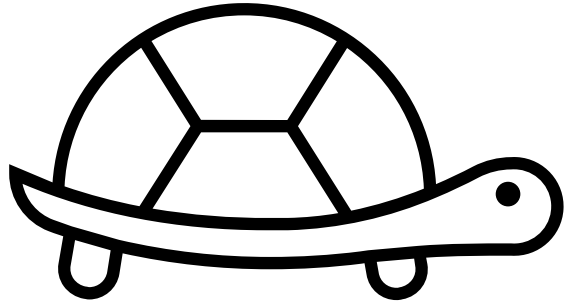
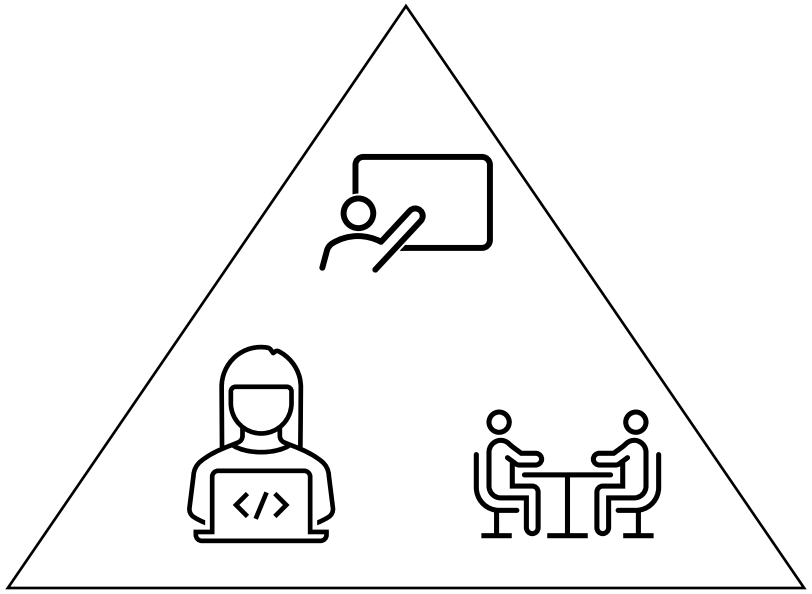
R, the Tidyverse, and basic data science principles



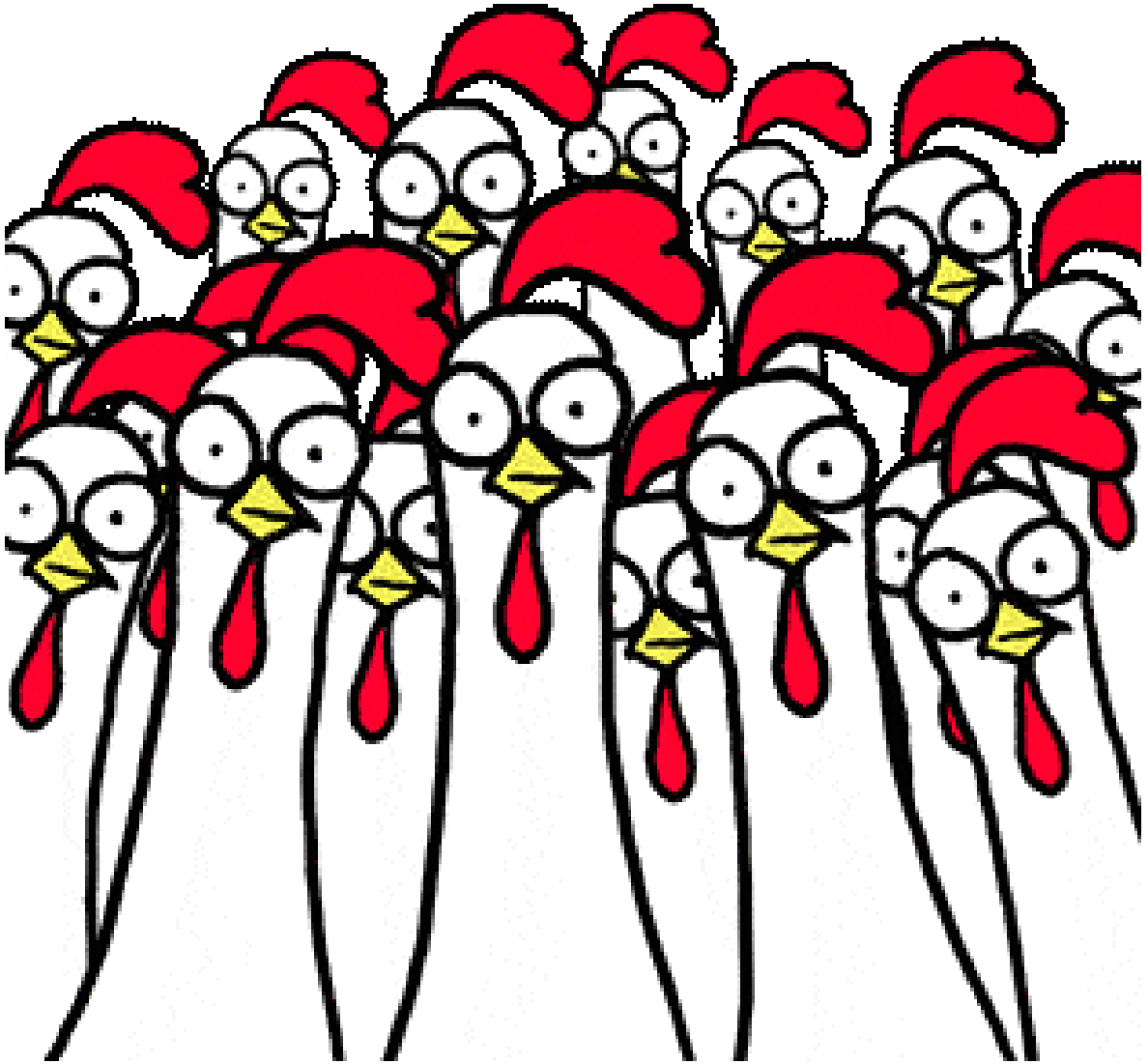
Wickham et al., 2019. Welcome to the Tidyverse. Journal of Open Source Software 4, 1686. https://doi.org/10.21105/joss.01686


Kg





Who are we?



If statistics programs/languages were cars...

Top 20 Programming Languages to Learn in 2025

The list of the best programming languages is sourced from the **Stack Overflow Developer Survey**, **GitHub Octoverse**, and the **TIOBE Index**. We've also analyzed job market demand through platforms like **Indeed**, **Glassdoor**, and **LinkedIn** to highlight the skills companies are currently hiring for. We have provided information about the **top programming languages to learn**, as well as those offering the highest salaries:

9. R

R is a **statistical computing** and **graphics language and environment**. It is very much extensible and has a large collection of abilities and techniques in its niche, thus being a favorite choice for doing **data analysis** and **academic research**.

Key Features

- **Designed for statistical computing and data visualization; excels in data manipulation and graphical output.**
- Performance lags in non-statistical tasks; challenging for non-statisticians.
- Preferred for data analysis, statistics, and academic research.
- Widely used in academia and data science communities.

Category	Details
Learning Curve	Moderate to steep
Average Salary	$105,000 per year
Platforms	Cross-platform
Level	Intermediate to Advanced
Key Skills	Data Analysis, Statistics, Visualization
Companies Using	Google, Facebook, Twitter, Airbnb
Community Size	Large and active
Ecosystem Maturity	Mature with extensive statistical libraries

https://www.geeksforgeeks.org/top-programming-languages/#20-best-programming-languages-in-2025

10 Best Language for Machine Learning

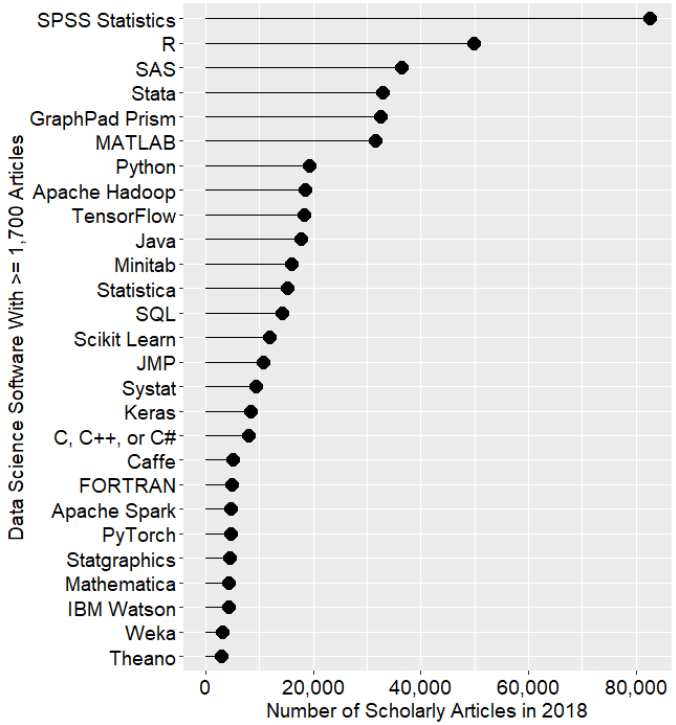
https://www.geeksforgeeks.org/best-language-for-machine-learning/

The Popularity of Data Science Software

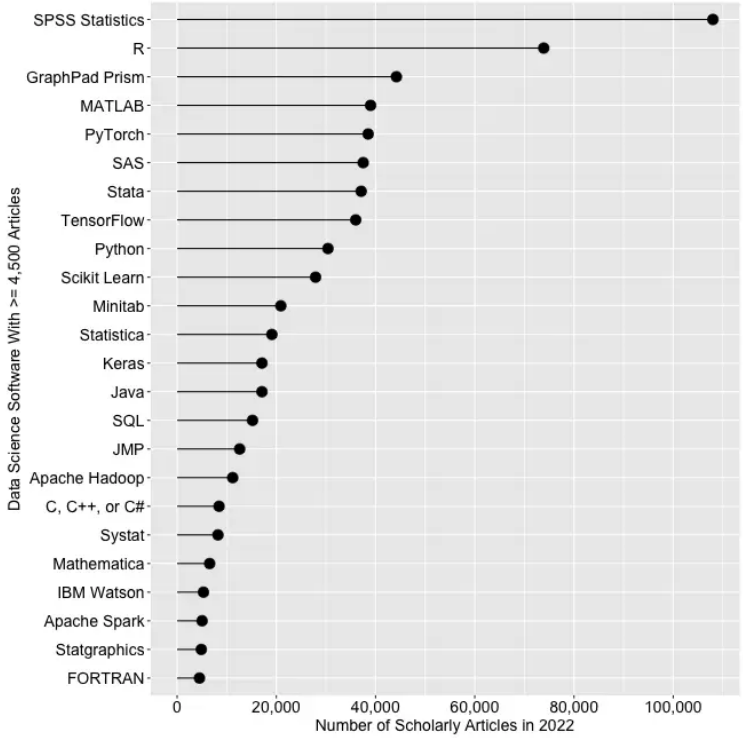
by Robert A. Muenchen

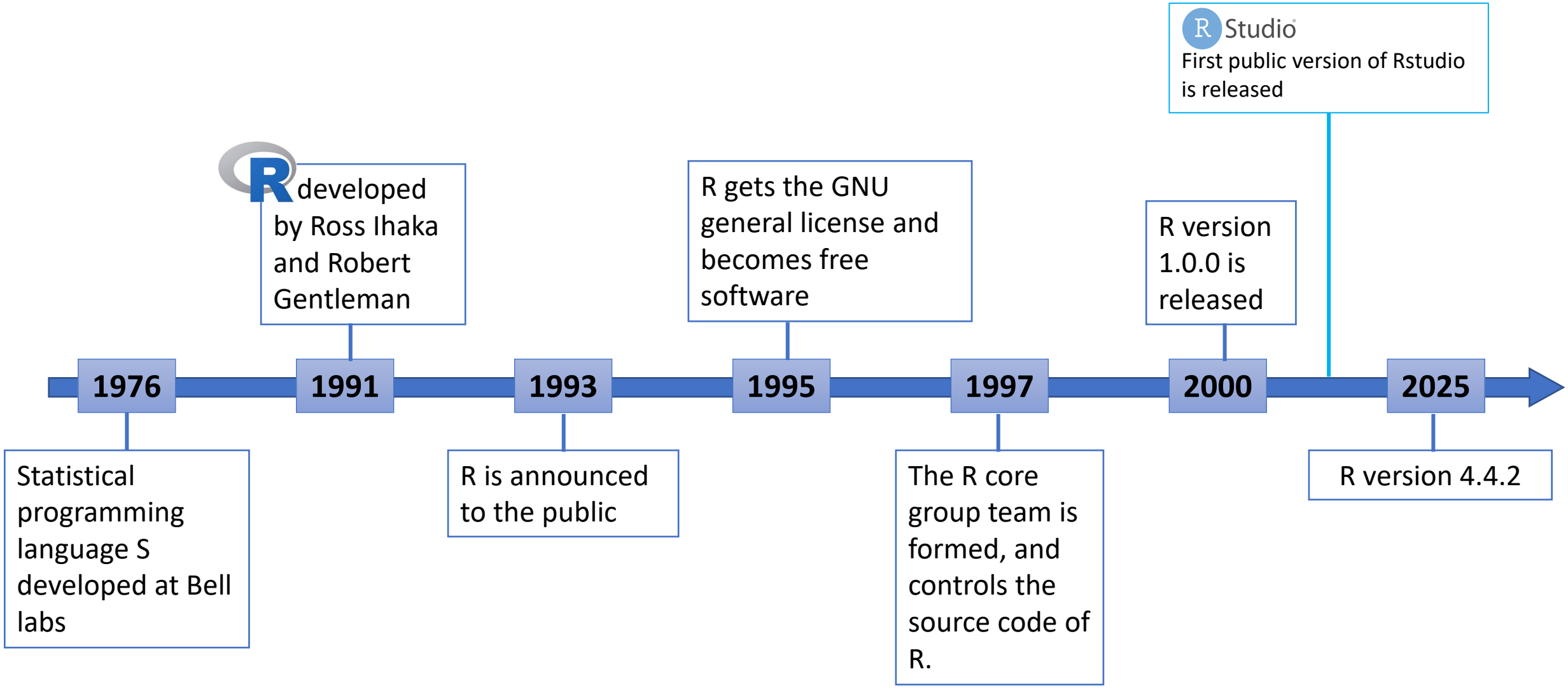
http://r4stats.com/articles/popularity/

2018



2022




RStudio
First public version of Rstudio is released
R developed by Ross Ihaka and Robert Gentleman
R gets the GNU general license and becomes free software
R version 1.0.0 is released
1976
1991
1993
1995
1997
2000
2025
Statistical programming language S developed at Bell labs
R is announced to the public
The R core group team is formed, and controls the source code of R.
R version 4.4.2

base R is conservative

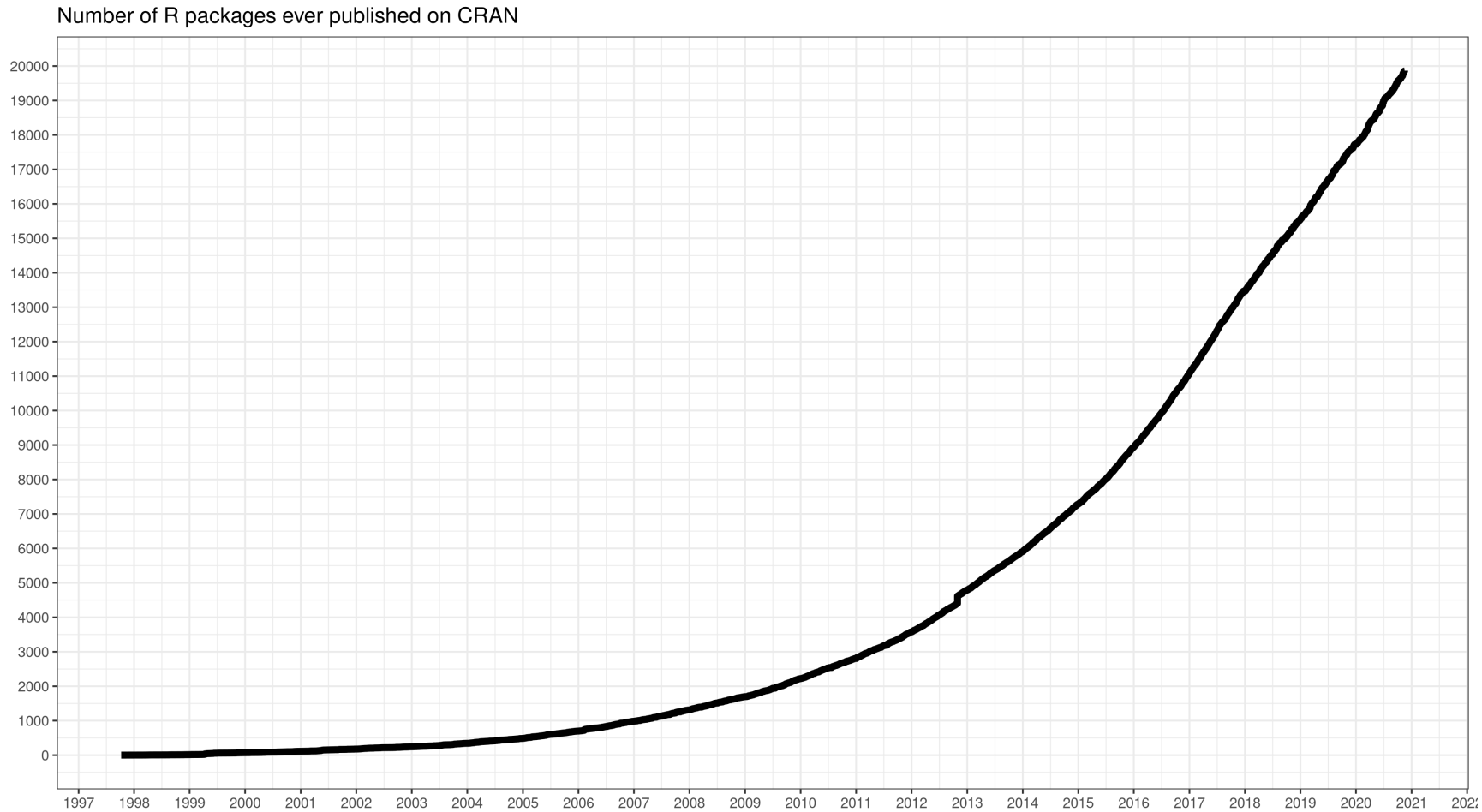
highly focussed on stability

CRAN

The Comprehensive R Archive Network

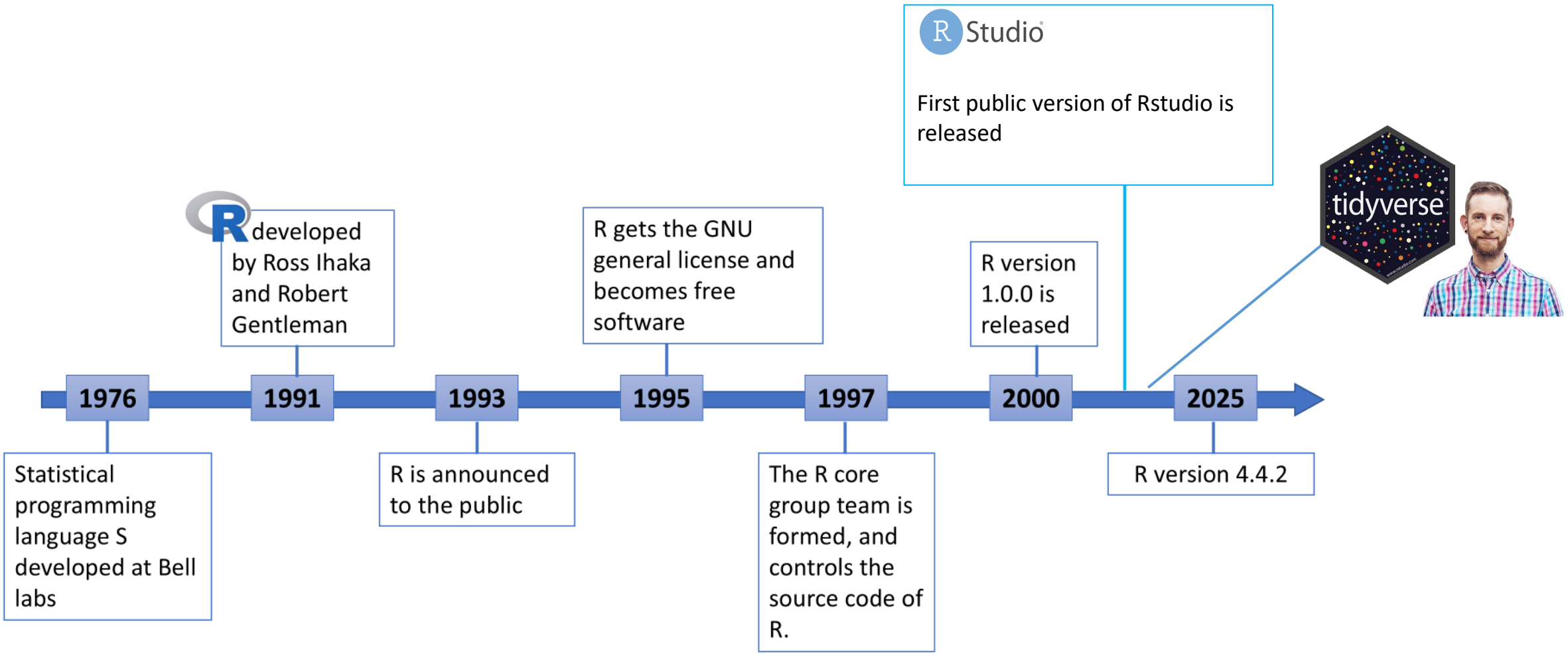
Currently, the CRAN package repository features 19875 available packages.

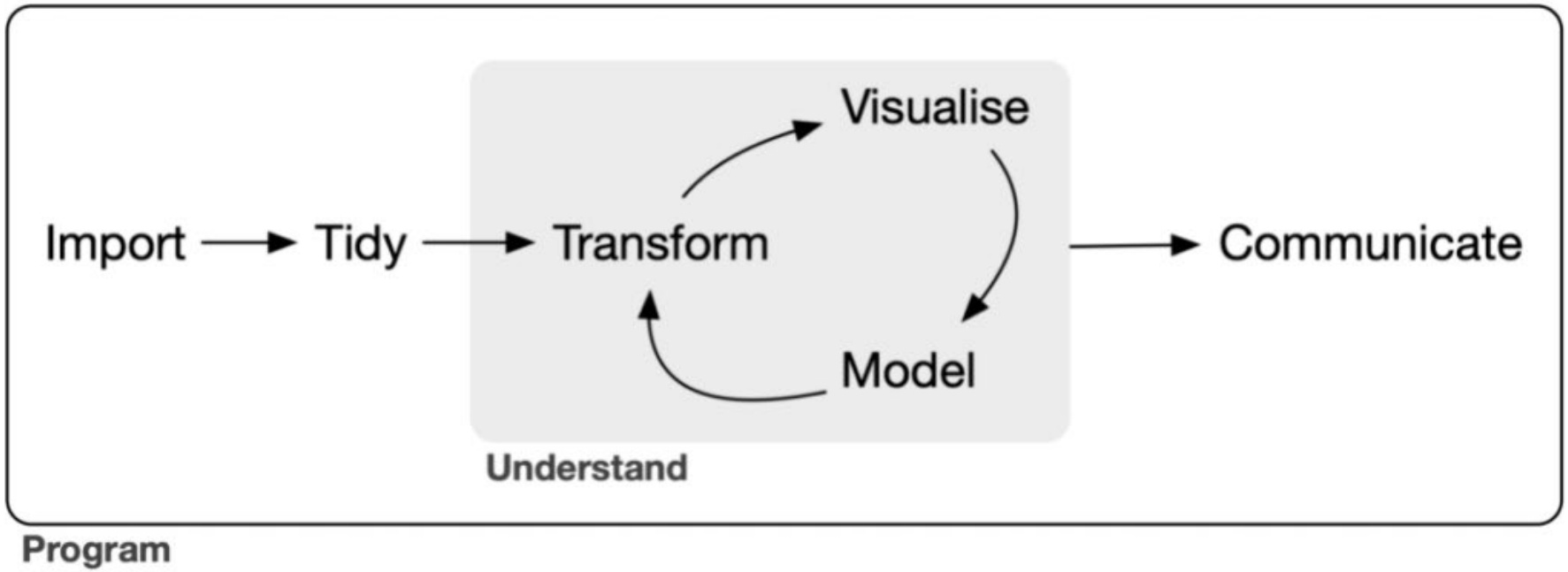
2023-08-08



Plot by Gergely Daróczi

https://gist.github.com/daroczig/3cf06d6db4be2bbe3368

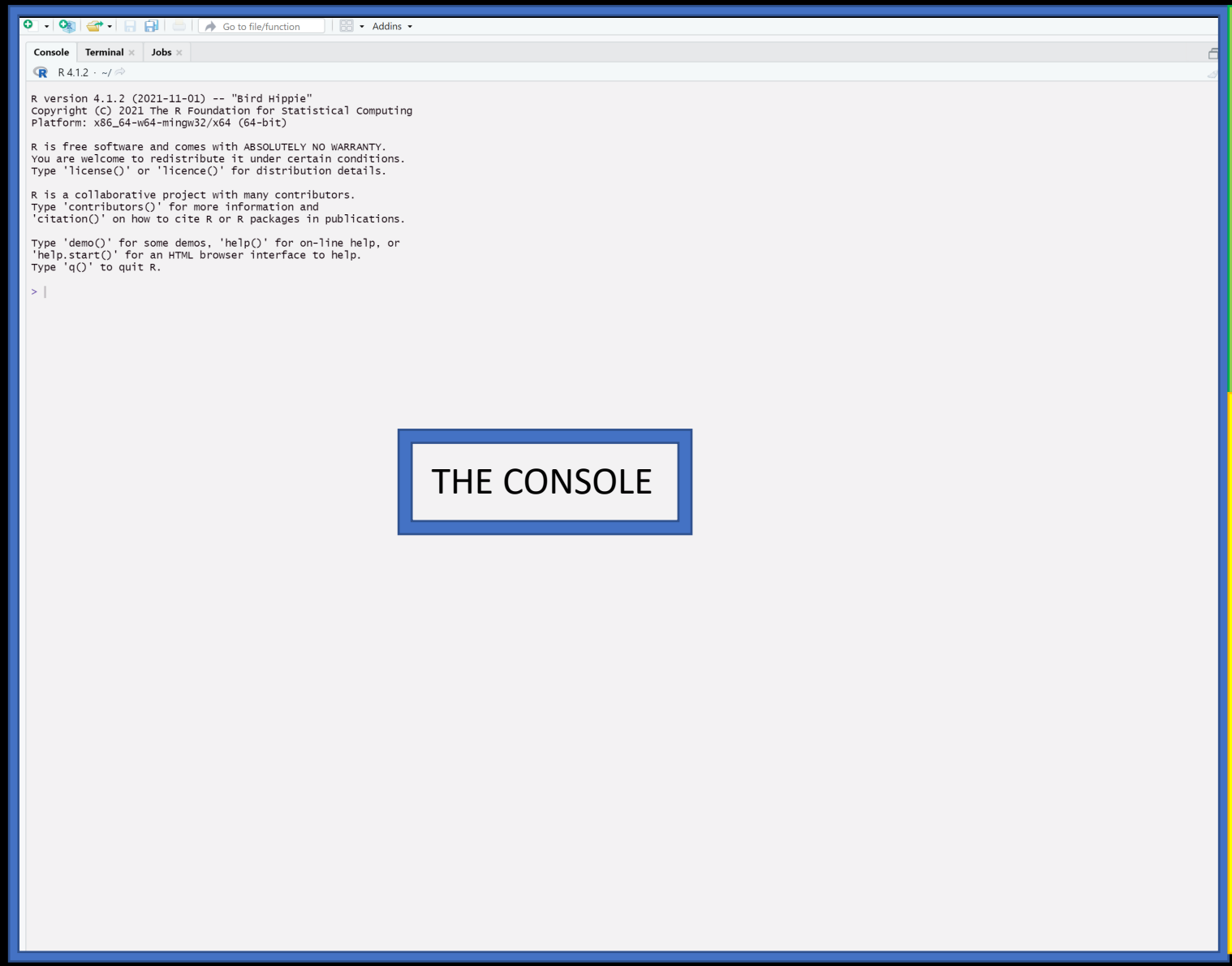

RStudio
First public version of Rstudio is released
tidyverse
developed by Ross Ihaka and Robert Gentleman
R gets the GNU general license and becomes free software
R version 1.0.0 is released
1976
1991
1993
1995
1997
2000
2025
Statistical programming language S developed at Bell labs
R is announced to the public
The R core group team is formed, and controls the source code of R.
R version 4.4.2

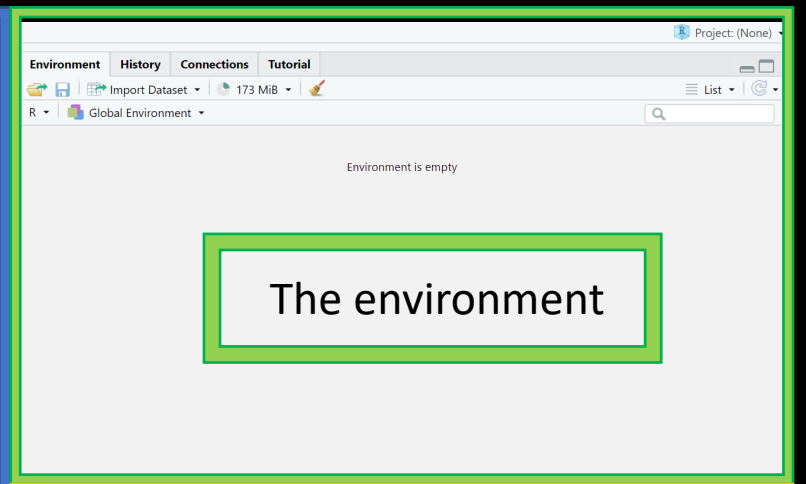


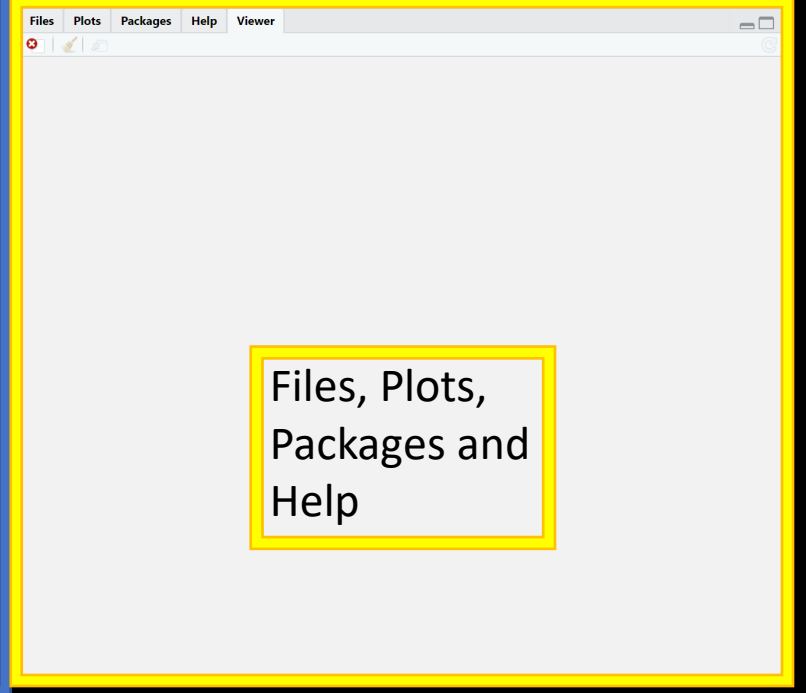
Wickham et al., 2019. Welcome to the Tidyverse. Journal of Open Source Software 4, 1686. https://doi.org/10.21105/joss.01686

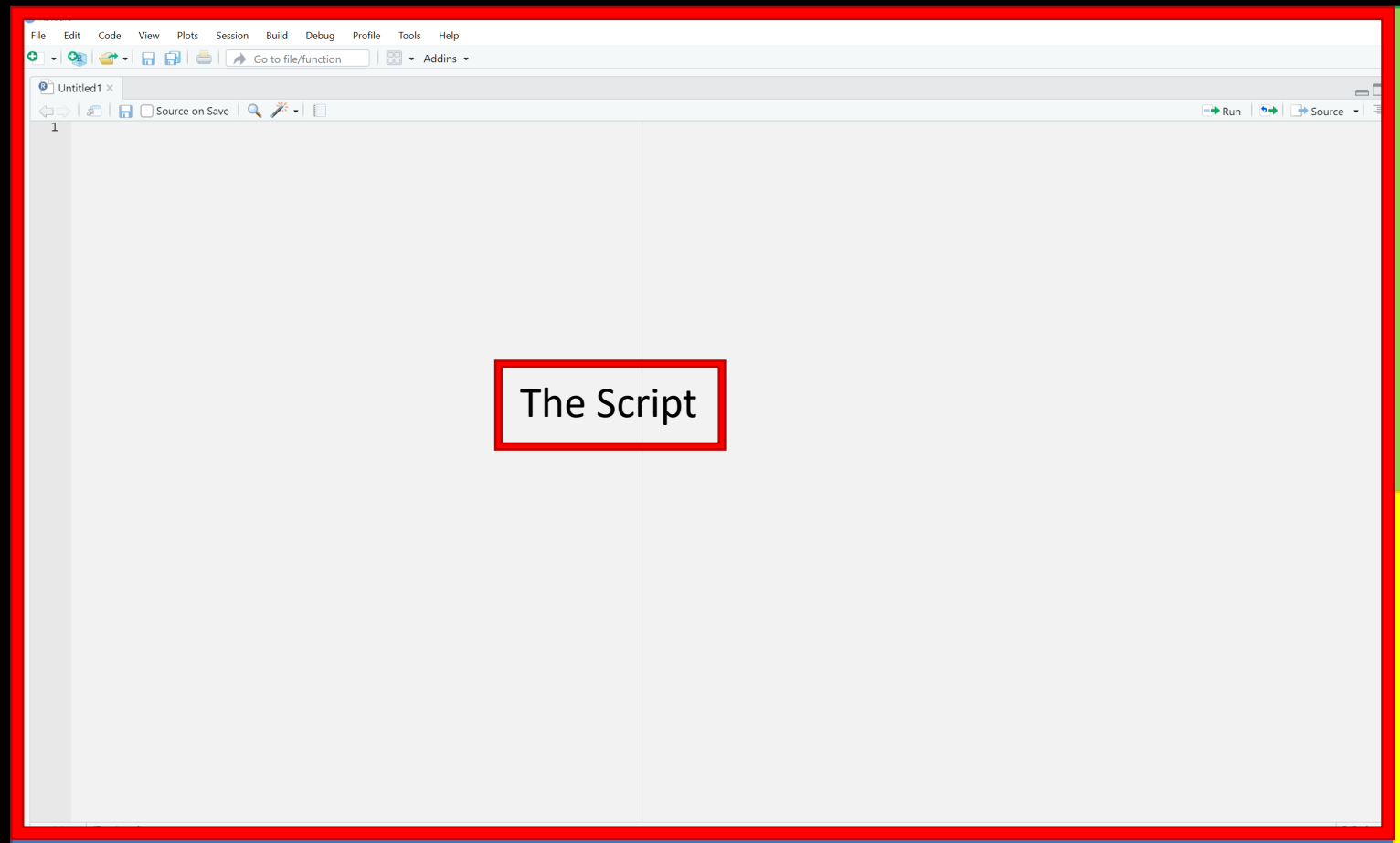

R
Studio®

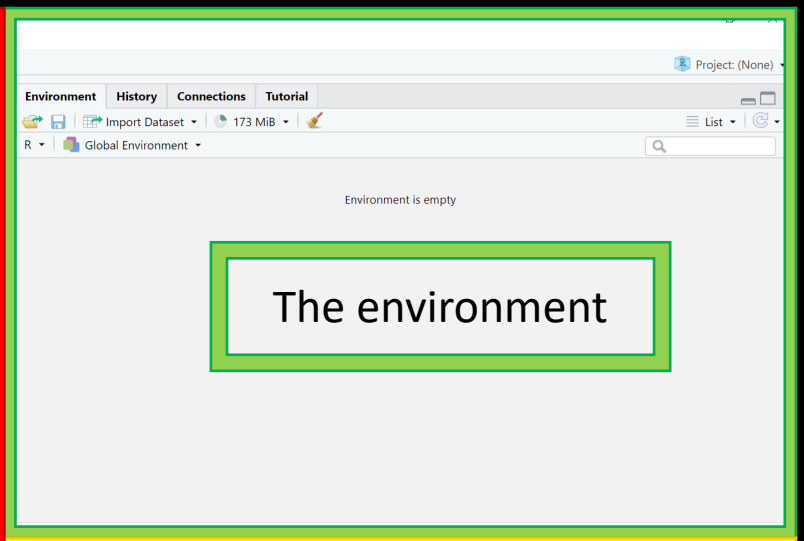
- Navigating Rstudio
  - Panes (Console, Environment, Plot, Script)
- Scripts and .Rmd files
- Rstudio Projects
- Autocomplete
- F1 for help

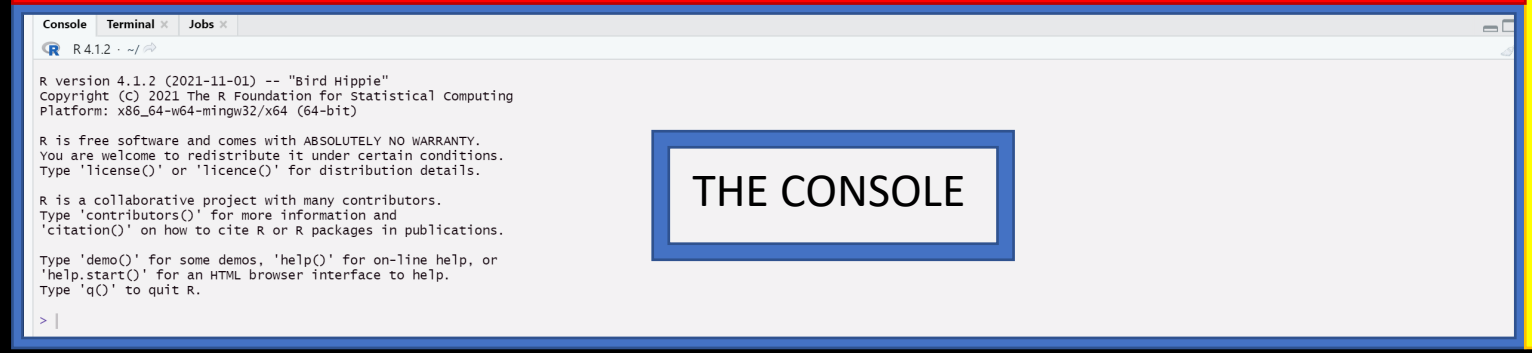

Go to file/function
Addins
Console
Terminal
Jobs
R 4.1.2 · ~/
R version 4.1.2 (2021-11-01) -- "Bird Hippie"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
>
THE CONSOLE

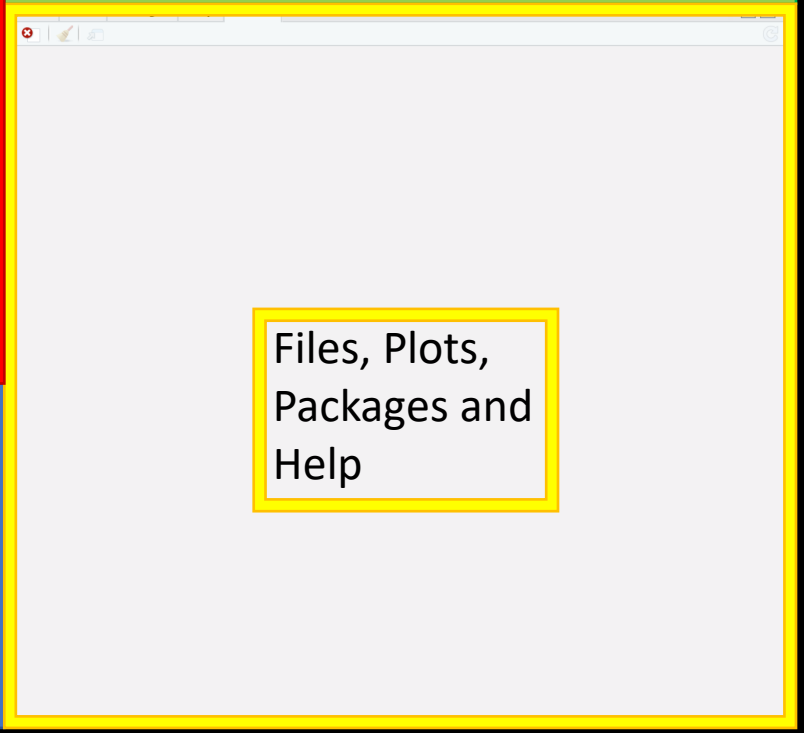

Project: (None)
Environment
History
Connections
Tutorial
Import Dataset
173 MiB
List
R
Global Environment
Environment is empty
The environment


Files
Plots
Packages
Help
Viewer
Files, Plots, Packages and Help

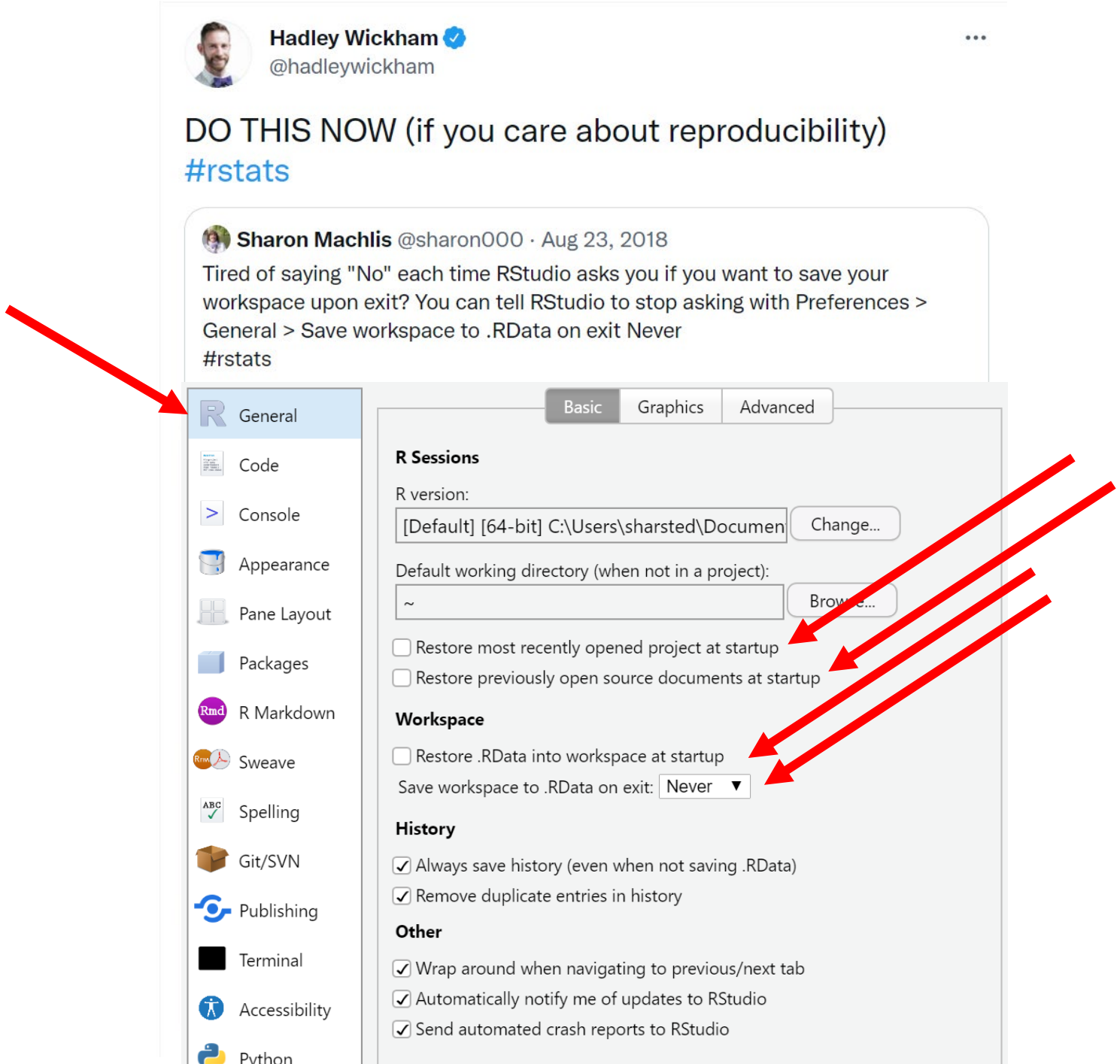

File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Untitled1
Source on Save
Run
Source
1
The Script


Project: (None)
Environment History Connections Tutorial
Import Dataset
173 MiB
List
R
Global Environment
Environment is empty
The environment

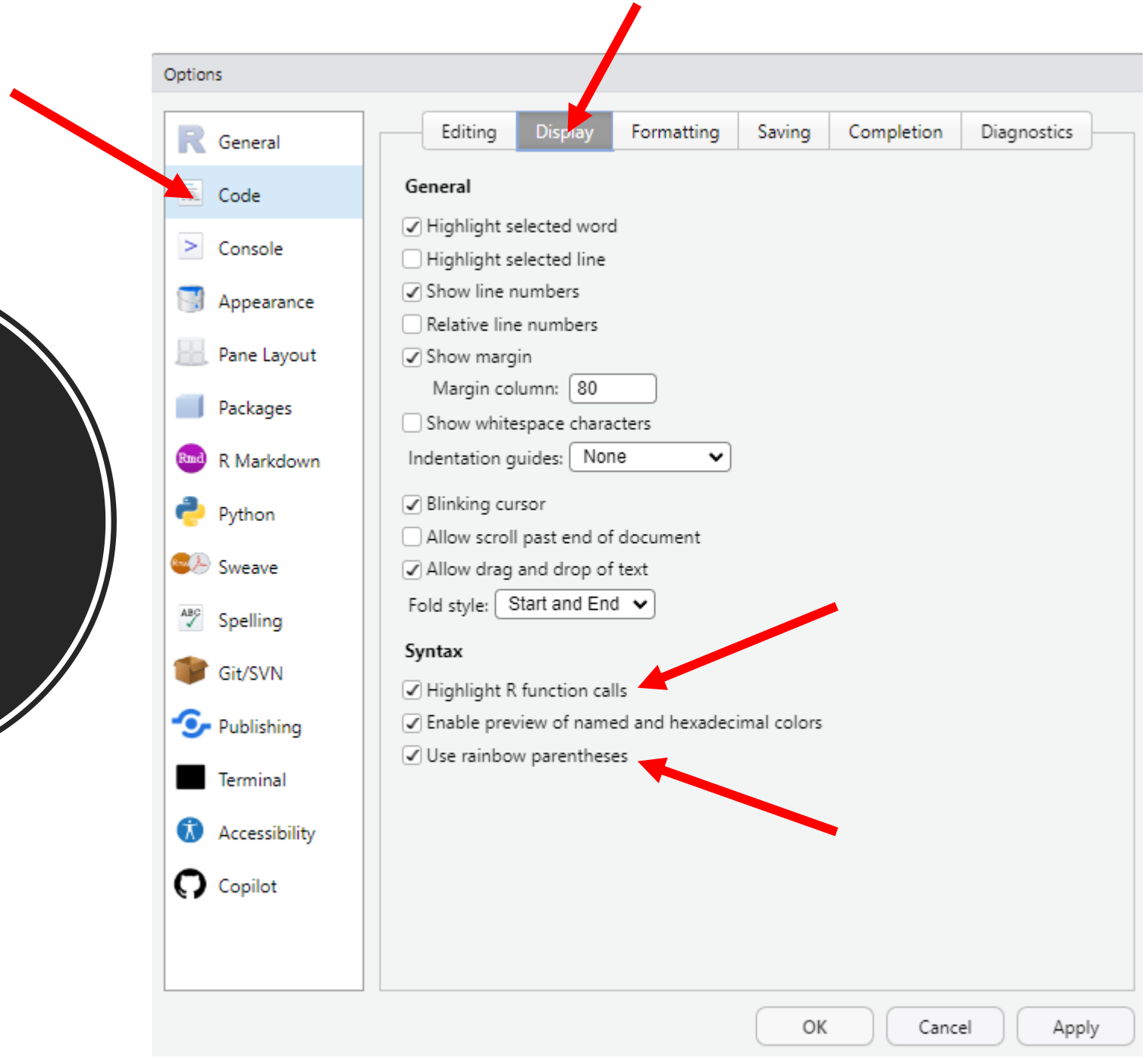

Console Terminal Jobs
R 4.1.2 · ~/
R version 4.1.2 (2021-11-01) -- "Bird Hippie"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
>
THE CONSOLE


Files, Plots, Packages and Help

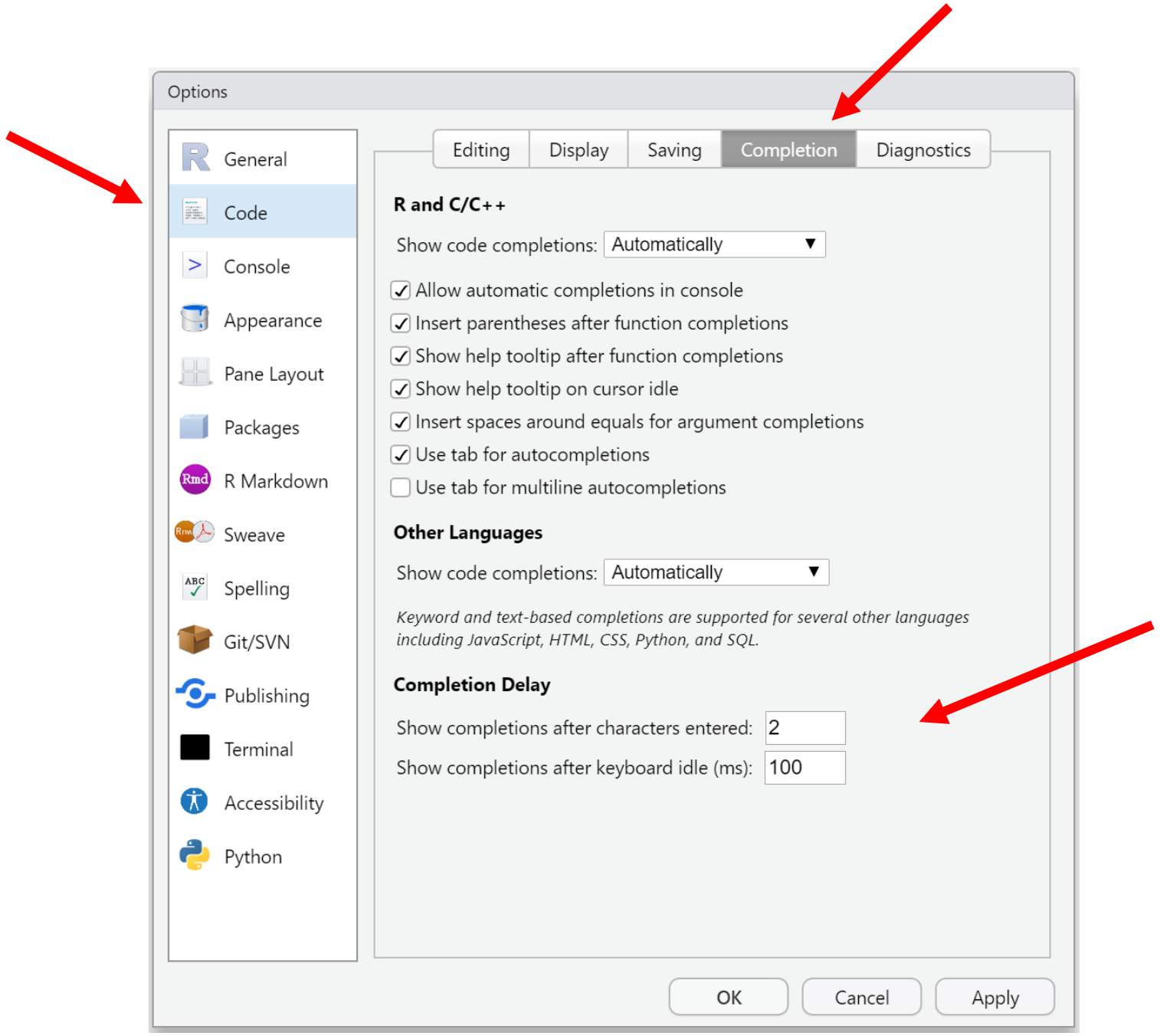
Rstudio setup



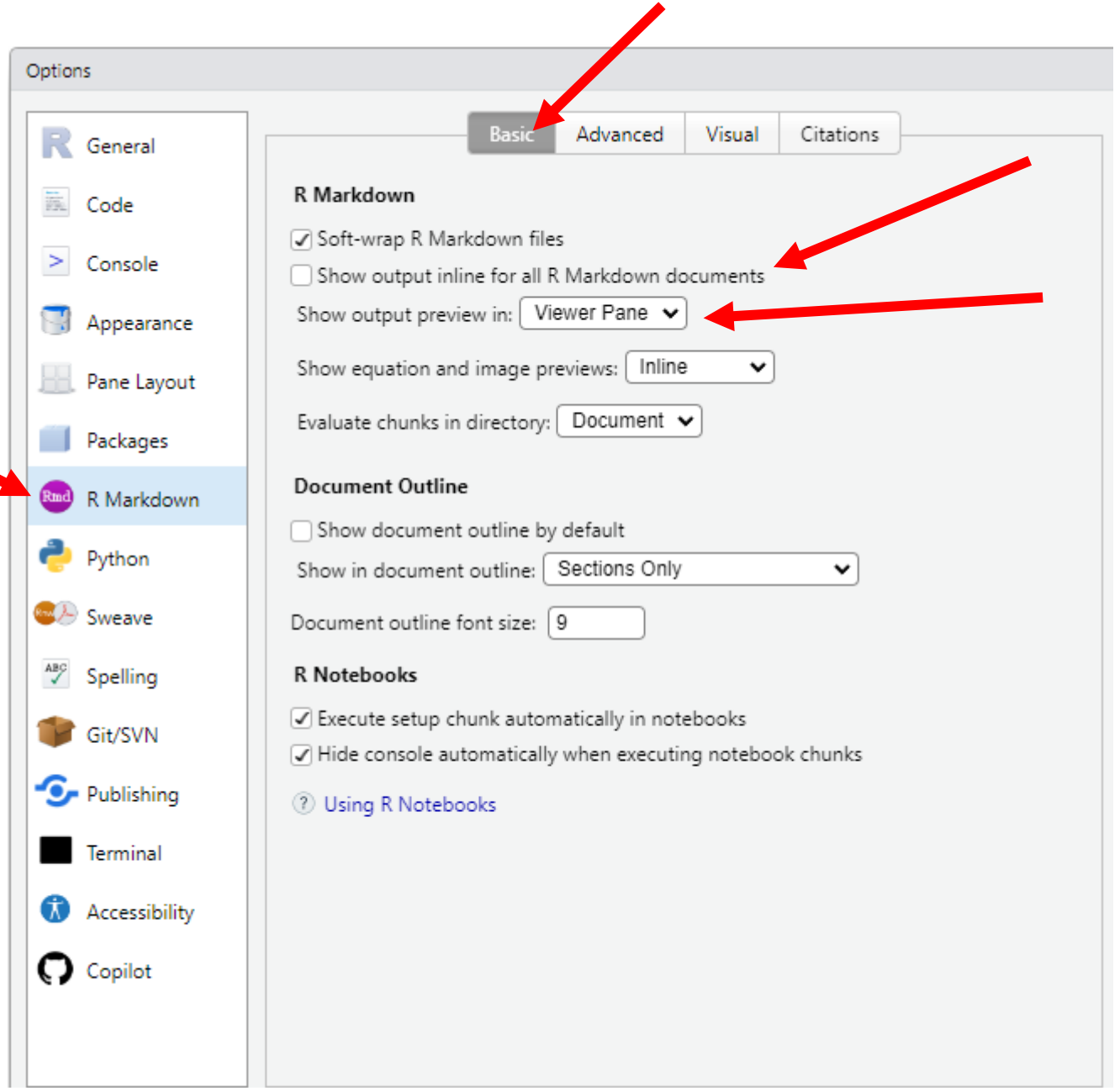
Rstudio setup Rainbow (())



Rstudio setup Autocomplete



RmarkDown setup



- <-
- #
- Functions
- |>

<- (The Assignment operator)

```
2+2
4
```

```
a <- 2+2
```

No Output?
Where is the a?

```
a
4
```

```
a <- a + 2
a
6
```

(comments)

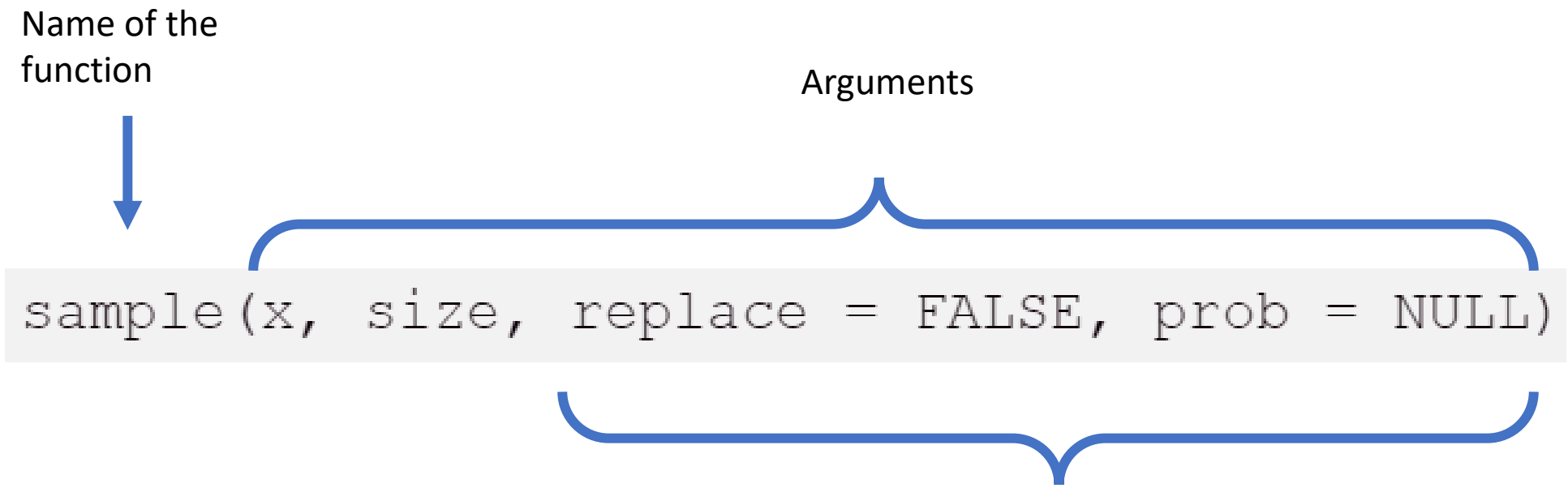
```
# THIS IS A COMMENT
# THIS LINE WILL NOT BE EVALUATED

2+3  # THIS PART WILL ALSO NOT BE EVALUATED
```

Comments are useful for many things.

First and foremost - ***explaining your code***

The person who will most likely benefit from your comments is ***future you***!

Functions


Name of the function
Arguments
sample(x, size, replace = FALSE, prob = NULL)
These two arguments have a default value.
How can you see this?

Functions

sample {base} R Documentation

Random Samples and Permutations

Description

`sample` takes a sample of the specified size from the elements of `x` using either with or without replacement.

Usage

```
sample(x, size, replace = FALSE, prob = NULL)

sample.int(n, size = n, replace = FALSE, prob = NULL,
           useHash = (!replace && is.null(prob) && size <= n/2 && n > 1e7))
```

Arguments

`x`	either a vector of one or more elements from which to choose, or a positive integer. See 'Details.'
`n`	a positive number, the number of items to choose from. See 'Details.'
`size`	a non-negative integer giving the number of items to choose.
`replace`	should sampling be with replacement?
`prob`	a vector of probability weights for obtaining the elements of the vector being sampled.
`useHash`	`logical` indicating if the hash-version of the algorithm should be used. Can only be used for `replace = FALSE`, `prob = NULL`, and `size <= n/2`, and really should be used for large `n`, as `useHash=FALSE` will use memory proportional to `n`.

```
sample(1:6, 6)
3 4 6 5 2 1
```

```
 sample(1:6, 6, replace = TRUE)
6 5 1 4 6 2
```

Nested Functions

- You can put a function inside another function

```
sample(1:6, 1000, replace = TRUE)
[5, 6, 3, 1, 5, 3….]
```

```
mean(sample(1:6, 1000, replace = TRUE) )
3.521
```

```
round(mean(sample(1:6, 1000, replace = TRUE)))
 4
```

```
round(mean(sample(1:6, 1000, replace = TRUE)), digits = 1)
 3.5
```

The Pipe |>

Nested functions tend to be difficult for humans to read.

Technical:
The pipe transfers the output of one function to the first argument of the following function.

Practical:
Read the pipe as “AND THEN”

Compare:

```
round(mean(sample(1:6, 1000, replace = TRUE)), digits = 1)
```

```
sample(1:6, 1000, replace = TRUE) |>
  mean() |>
  round(digits = 1)
```

```
my_data  <- mtcars
my_data
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3

```
my_data |>
 group_by(cyl) |>
 summarise(
  n = n(),
  mean_hp = mean(hp),
  sd_hp = sd(hp)
  )
```

A tibble: 3 × 4

cyl	n	mean_hp	sd_hp
<dbl>	<int>	<dbl>	<dbl>
1 4	11	82.6	20.9
2 6	7	122.	24.3
3 8	14	209.	51.0

Daily usage Stack Overflow