Sociology Quant Camp

Introduction to R

Module 1: Intro and basics

Monica Alexander, Statistical Sciences and Sociology

Welcome

- Hello!
- Overview of today
  - Module 1: Intro, navigating RStudio, R scripts and R code basics
  - Module 2: Piping, the tidyverse
  - Module 3: Intro to plotting with ggplot2
- Will put materials here: https://www.monicaalexander.com/teaching/

Hello!

- Statistical Sciences and Sociology departments
- Demographer
- On sabbatical (lol) usually teach either 6302 or 6707
- Love R
- monica.alexander@utoronto.ca

Introduction to R and RStudio

Downloads

- To install R:
  - Go to https://cran.r-project.org/
  - Select the download link that is relevant to you: if you have a Mac, select “Download R for (Mac) OS X”, if you have a Windows machine, select “Download R for Windows”.
  - If you have a Mac, click on the “R-4.4.1pkg” link. If you have a Windows machine, click on the “base” link, then click on the “Download R 4.4.1 for Windows” link. Open the downloaded file and follow the install instructions on your machine.

Downloads

- To install RStudio
  - Go to https://posit.co/download/rstudio-desktop/
  - Scroll down and click the “Download RStudio” button (it will either say for Mac or Windows)
  - Open the downloaded file and follow the install instructions on your machine.

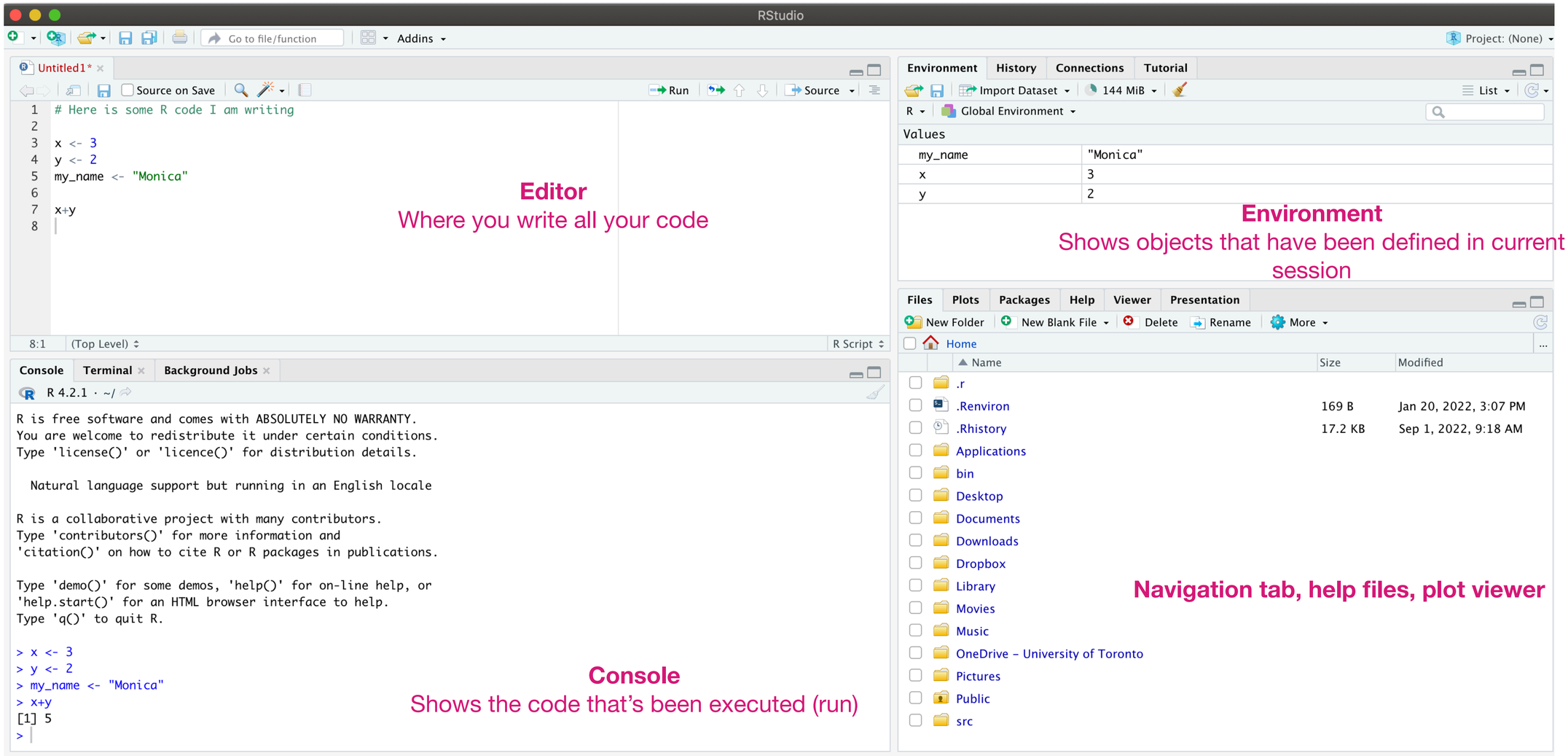
Install check?

What is R?

- R is a programming language for statistical computing and graphics
- Using R is like speaking another language (but you type it)
- You may have used other programs to do statistical calculations before (Excel, SPSS)
- With R you have to give the computer typed commands in order for it to do statistics (rather than clicking buttons)
- Much more powerful methods available

What is RStudio?

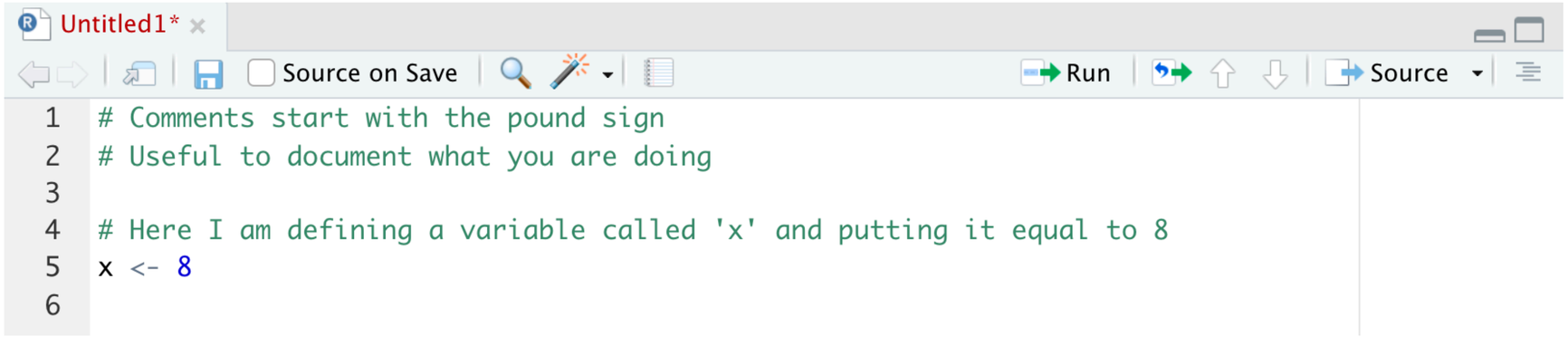
- RStudio is an integrated development environment for R
- It makes it easier to write R code and visualizes inputs and outputs
- Car analogy:
  - R is the engine
  - RStudio is the car dashboard


RStudio
Go to file/function
Addins
Project: (None)
Untitled1*
Source on Save
Run
Source
1 # Here is some R code I am writing
2
3 x <- 3
4 y <- 2
5 my_name <- "Monica"
6
7 x+y
8
8:1
(Top Level)
R Script
Editor
Where you write all your code
Console
Terminal
Background Jobs
R 4.2.1 · ~/
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.
Natural language support but running in an English locale
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.
> x <- 3
> y <- 2
> my_name <- "Monica"
> x+y
[1] 5
>
Console
Shows the code that's been executed (run)
Environment
History
Connections
Tutorial
Import Dataset
144 MiB
List
R
Global Environment
Values
my_name "Monica"
x 3
y 2
Environment
Shows objects that have been defined in current session
Files
Plots
Packages
Help
Viewer
Presentation
New Folder
New Blank File
Delete
Rename
More
Home
Name
Size
Modified
.r
.Renviron 169 B Jan 20, 2022, 3:07 PM
.Rhistory 17.2 KB Sep 1, 2022, 9:18 AM
Applications
bin
Desktop
Documents
Downloads
Dropbox
Library
Movies
Music
OneDrive - University of Toronto
Pictures
Public
src
Navigation tab, help files, plot viewer

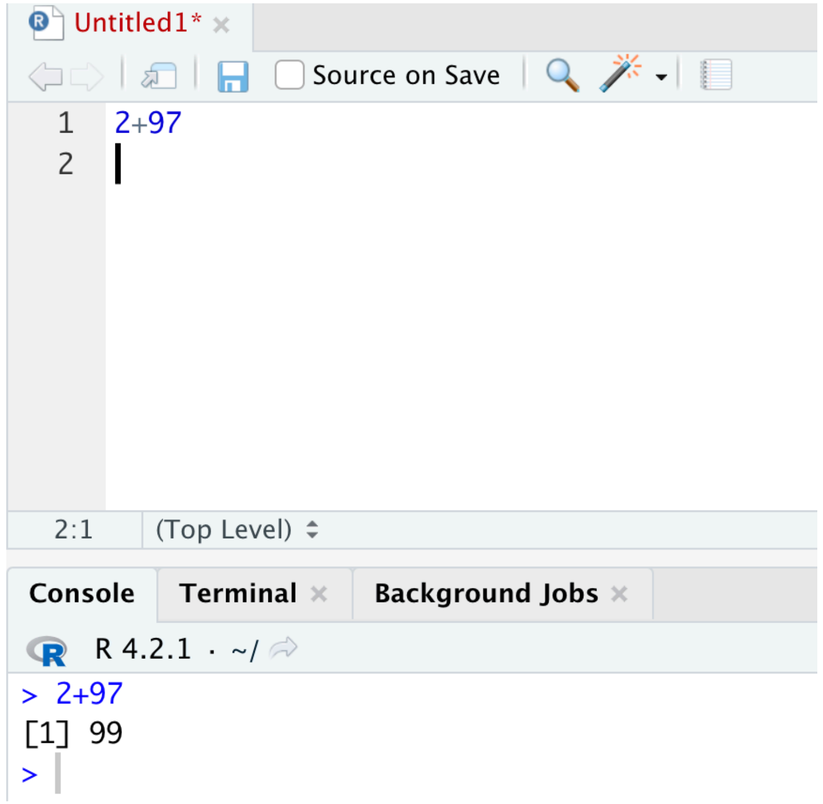
Demo: opening RStudio, opening and saving a R script

R code basics

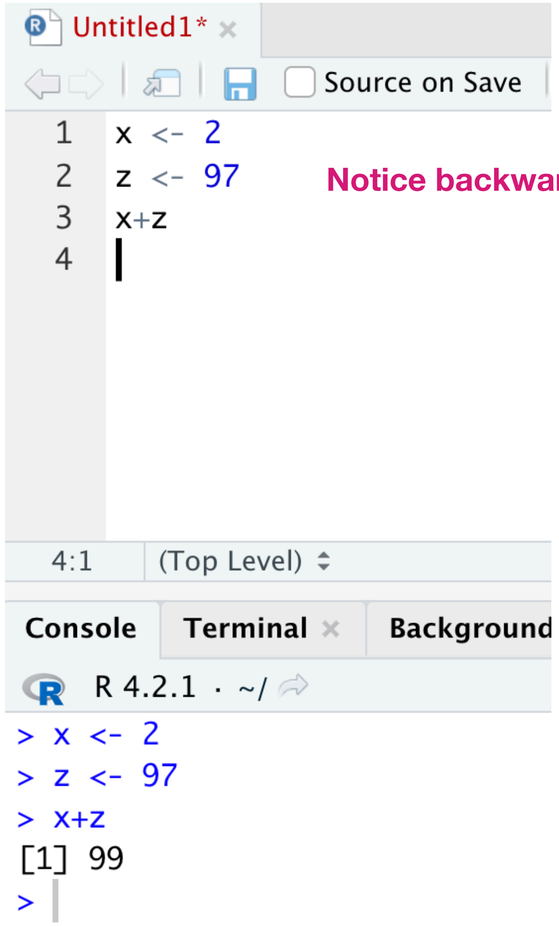
Code versus comments



R as a calculator



… versus defining objects



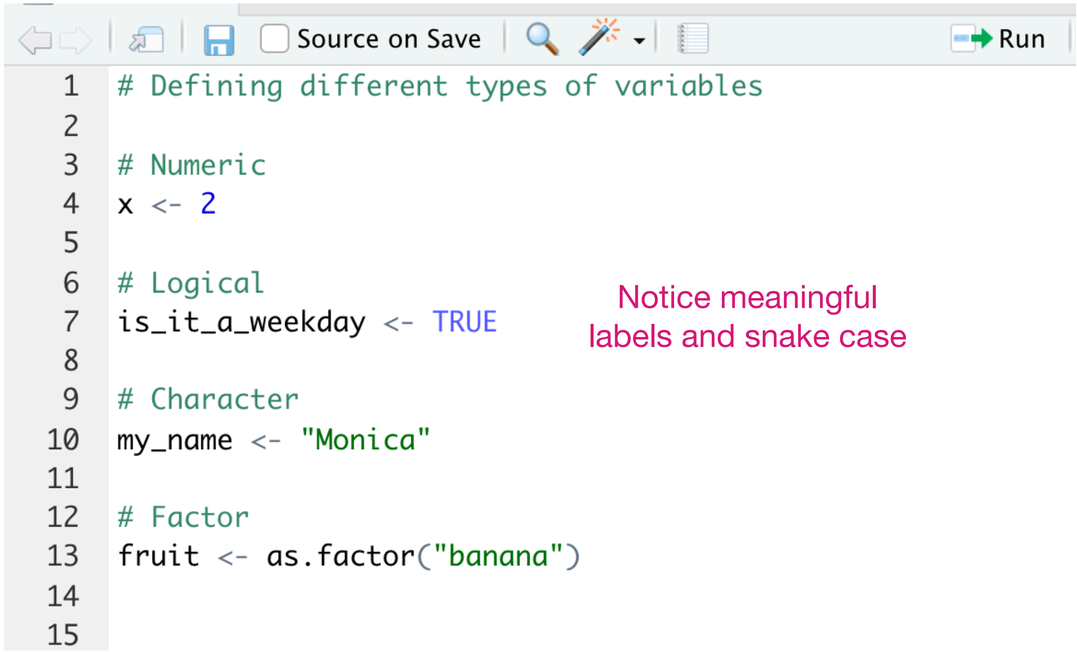
Notice backwards arrows

All standard mathematical operations are available

Demo: mathematical operations, defining objects

Types of variables in R

- Numeric
- Logical (TRUE/FALSE)
- Character
- Factor (categories with levels defined)



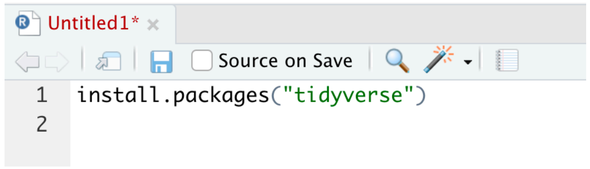
Demo: different variable types, checking variable types

Packages

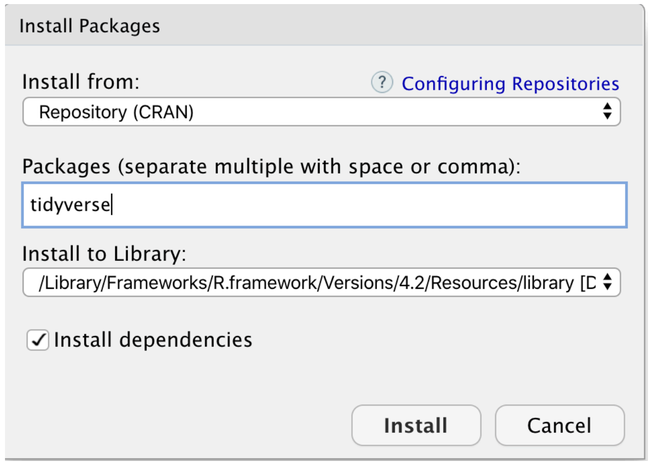
- People have written R Packages, which are add ons to base R that increase functionality
- Phone analogy
  - R/RStudio is a phone
  - Packages are apps
- One package that is very useful is the `tidyverse`
- Has graphing capabilities, tiydverse grammar (more later)

Installing packages

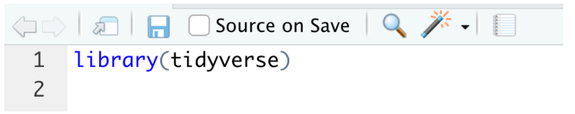
- Via code
- Or using menu (Tools -> Install packages…)
- Once a package is installed, don't need to do it again!



OR



- To use the functionality in a package, need to load it in at the start of your code using the `library` function



Different types of objects in R

- Single values
- Vectors:
  - contain two or more values
  - Defined with the c() function (“concatenate”)
  - Values must be of the same type
- Data frames (tibbles)
  - Closest thing to a dataset that we deal with
  - Each column is a different variable, each row is an observation
  - Columns (variables) can be different types

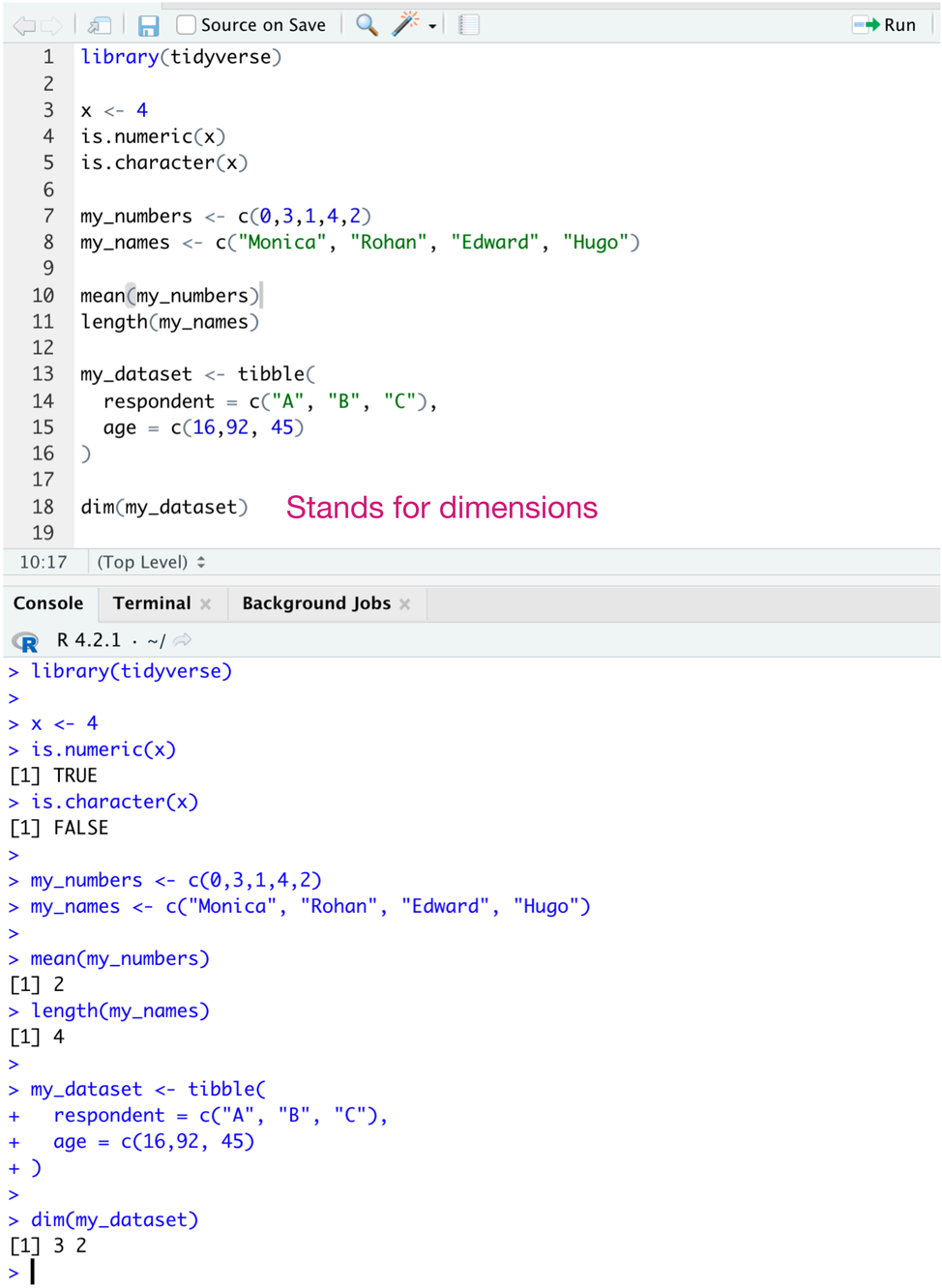
```r
library(tidyverse)


# single value
x <- 2
color <- "red"

# vector
my_numbers <- c(0,3,1,4,2)
my_names <- c("Monica", "Rohan", "Edward", "Hugo")

# tibble
my_dataset <- tibble(
  respondent = c("A", "B", "C"),
  age = c(16,92, 45)
)
```

We can define columns of a tibble using vectors

Demo: defining different types of objects

Functions

- Do stuff to your variables!
- Have already seen some: as.factor(), c(), tibble()
- Examples:
  - mean(), median()
  - min(), max()
  - length(), dim()
  - paste()
  - is.numeric() etc

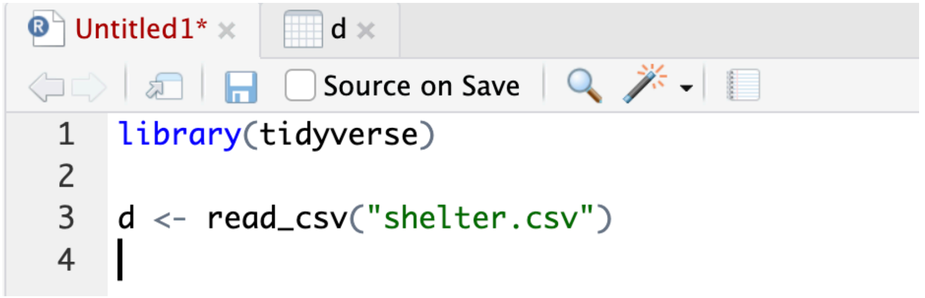


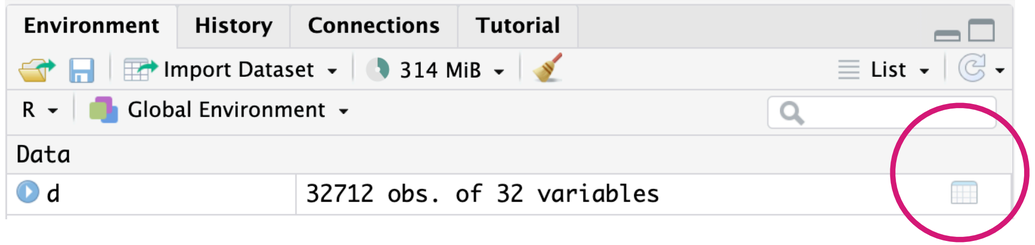
Demo: functions

Opening files

Need to be careful with file paths!

- One of the most powerful ways we can use R is to analyze and visualize data
- Need to be able to read in files of different formats (csv, excel, Stata…)
- Can use a variety of `read_*()` functions e.g. `read_csv()`
- Note these are from the `tidyverse` package so need to make sure this has been loaded in
- Data are from https://open.toronto.ca/dataset/daily-shelter-overnight-service-occupancy-capacity/





Untitled1* d

Filter

	_id	OCCUPANCY_DATE	ORGANIZATION_ID	ORGANIZATION_NAME
1	1	2024-01-01	24	COSTI Immigrant Services
2	2	2024-01-01	24	COSTI Immigrant Services
3	3	2024-01-01	24	COSTI Immigrant Services
4	4	2024-01-01	24	COSTI Immigrant Services
5	5	2024-01-01	24	COSTI Immigrant Services
6	6	2024-01-01	14	Christie Ossington Neighbourhood Centre
7	7	2024-01-01	14	Christie Ossington Neighbourhood Centre
8	8	2024-01-01	14	Christie Ossington Neighbourhood Centre
9	9	2024-01-01	14	Christie Ossington Neighbourhood Centre
10	10	2024-01-01	14	Christie Ossington Neighbourhood Centre

Demo: opening files, setting working directory, viewing datasets

Where to get help

- Lots of good, free online sources
  - R for Data Science: https://www.tidyverse.org/learn/
  - Telling stories with data: https://tellingstorieswithdata.com/
  - Tidyverse skills for data science: https://jhudatascience.org/tidyversecourse/intro.html
- Google/Stack Overflow
- Email
- Practice, practice, practice; don’t be afraid of mistakes