DATA TYPES

Statistics Tutorial
Day 2

Prabesh Dhakal
15 April 2019

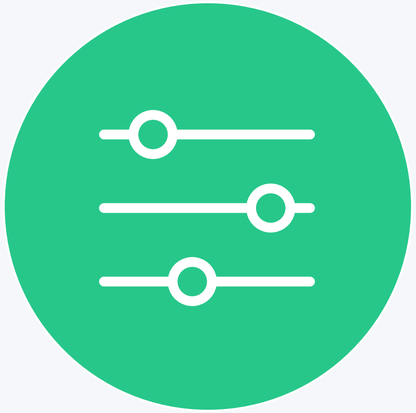
REVIEW FROM LAST SESSION

1. Statistics is the science of collecting, preparing, analyzing, and interpreting data.
2. Statistics allows us to: (i) make sense of the world around us, (ii) make decisions, and (iii) make predictions.
3. The tools we will use to work with the data and perform calculations is going to be R-Studio.

WHAT ARE WE DOING TODAY?

R RECAP

We briefly revisit the R tasks covered last week.

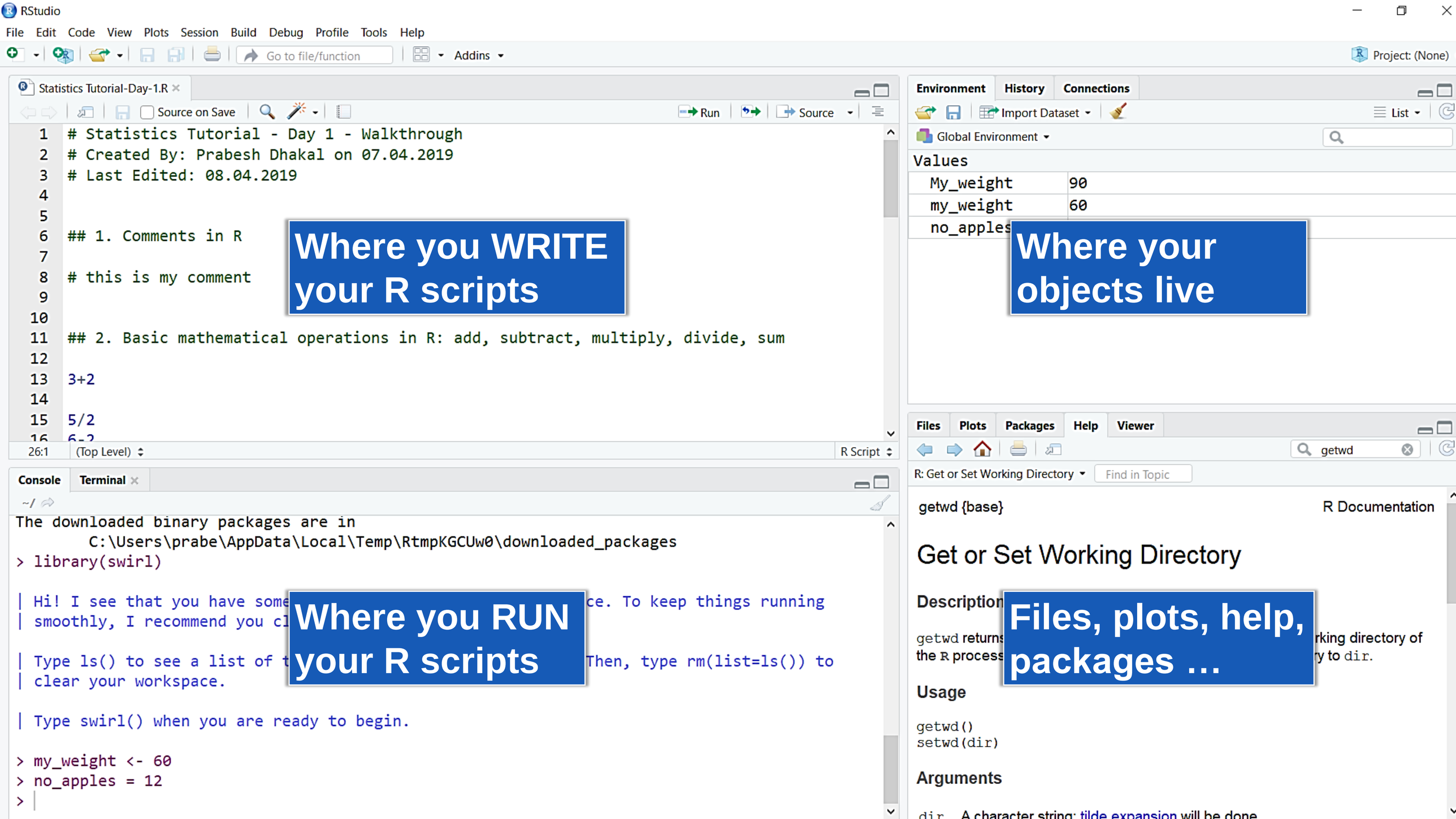


DATA TYPES

We talk about different data types and what we can do with them.

MORE R EXERCISES

We do more exciting things with R.

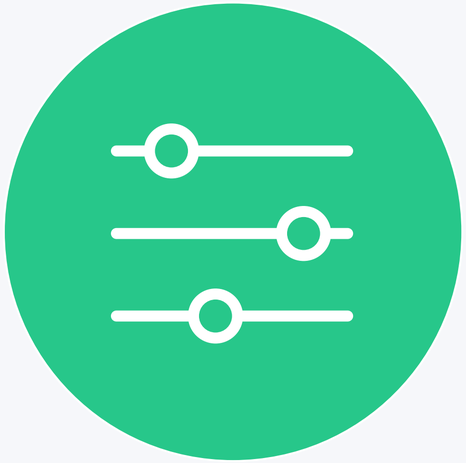

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Statistics Tutorial-Day-1.R
Source on Save
Run
Source
1 # Statistics Tutorial - Day 1 - Walkthrough
2 # Created By: Prabesh Dhakal on 07.04.2019
3 # Last Edited: 08.04.2019
4
5
6 ## 1. Comments in R
7
8 # this is my comment
9
10
11 ## 2. Basic mathematical operations in R: add, subtract, multiply, divide, sum
12
13 3+2
14
15 5/2
16 6-2
Where you WRITE your R scripts
26:1
(Top Level)
R Script
Environment
History
Connections
Import Dataset
List
Global Environment
Values
My_weight 90
my_weight 60
no_apples
Where your objects live
Console
Terminal
~/
The downloaded binary packages are in
C:\Users\prabe\AppData\Local\Temp\RtmpKGCUw0\downloaded_packages
> library(swirl)
| Hi! I see that you have some
ce. To keep things running
| smoothly, I recommend you cl
| Type ls() to see a list of t
Then, type rm(list=ls()) to
| clear your workspace.
| Type swirl() when you are ready to begin.
> my_weight <- 60
> no_apples = 12
>
Where you RUN your R scripts
Files
Plots
Packages
Help
Viewer
getwd
R: Get or Set Working Directory
Find in Topic
getwd {base}
R Documentation
Get or Set Working Directory
Description
getwd returns
rking directory of
the R process
y to dir.
Usage
getwd()
setwd(dir)
Arguments
dir A character string; tilde expansion will be done
Files, plots, help, packages …

CLASS EXERCISE - 1

Please do the following tasks:

1. Open R Studio
2. Open a new script
3. Write a line of code that stores the value `22/7` in an object named `custom_pi`
4. In another line, write this: `custom_pi == pi`
5. Save the file somewhere on your device
6. Close R Studio
7. Open R Studio
8. Run the lines you wrote

What observations could you make? What questions do you have?



Data Types

1. Differentiate between the types of data
2. Determine how we can approach them

CLASS EXERCISE - 2

1. Where was your starting point this morning in relation to this classroom?
   - Possible entries: N, S, W, E
2. What distance (in KM) did you travel to get here?
3. How are you feeling this morning?
   - Possible entries: very good, good, okay, bad, very bad
4. How did you get here?
   - Possible entries: on foot, bike, car, train, other

BASIC TYPES OF DATA IN STATISTICS

Qualitative Data

Nominal Data

- Variables without an order or a rank
- E.g. Gender, religion, ZIP Codes (PLZ)

Ordinal Data

- Variables with an order or a rank
- E.g. letter grades (A/B/C), star ratings

Binary Data

- Variables with only two possible states (also called dichotomous data)
- E.g. Yes/No, Pass/Fail, Head/Tail

Quantitative Data

Discrete Data

- Has clear spaces between values, and is generally counted in whole numbers.
- E.g.: no. of students in a class, shoe sizes, etc.

Continuous Data

- Falls on a continuous sequence, and is measurable.
- E.g. height, weight, speed, time to finish a project.

Interval Data

Interval Data

- Ordinal data with constant differences between observations
- No true zero point
- E.g. time, temperature

Ratio Data

- Same as interval data, but with a **true zero** (a clear definition of 0)
- E.g. age, height, income

WHAT DATA LOOKS LIKE

	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
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3

BASIC CALCULATIONS FOR DATA TYPES

	Nominal Data	Ordinal Data	Interval Data	Ratio Data
Frequency distribution	✓	✓	✓	✓
Median, quartiles, percentiles	✗	✓	✓	✓
Add or subtract	✗	✗	✓	✓
Mean, standard deviation	✗	✗	✓	✓
Ratio, coefficient of variation	✗	✗	✗	✓

BASIC STATISTICAL TESTS

		Outcome Variables			
		Binary	**Categorical**	**Ordinal**	**Continuous**
Predictor Variables	**Binary**	Chi-square, Fisher-Exact Tests	Chi-square Test	Wilcoxon	T-test/ Wilcoxon
	Categorical	Chi-square, Fisher-Exact Tests	Chi-square Test	Kruskal-Wallis	ANOVA
	Ordinal	Wilcoxon	Kruskal-Wallis	Spearman Correlation	ANOVA
	Continuous	T-test/Wilcoxon/ Logistic Regression	ANOVA/ Class Prediction	Ordinal Regression	Correlation/ Linear Regression

R Exercises

1. Basic data types in R
2. Basic data structures in R

BASIC DATA TYPES IN R

	Example	R Script
Character (or string)	`'apple'`	`a <- 'apple'`
Numeric	`1.618`	`b <- 1.618`
Integer	`3L, -5L`	`c <- 3L`
Booleans	`TRUE, FALSE`	`d <- TRUE`
Missing Values	`NA`	`e <- NA`

VECTOR OBJECTS IN R

	Example	R Script
Vectors	`c(1, 3, 4)`	`eg_vec <- c(1, 3, 4)`

Run the following in R Studio:

```
1. eg_vec <- c(1, 3, 4)
2. class(eg_vec)
3. typeof(eg_vec)
4. eg_vec * 3
5. eg_vec * eg_vec
6. View(eg_vec)
7. eg_vec_1 <- eg_vec * 3
8. View(eg_vec_1)
```

GENERAL R TIPS

Here are some functions that you will find useful:

- `help(x), builtins(), ?x`
- `typeof(x), class(x)`
- `str(x), summary(x)`
- `getwd(), setwd()`
- `list.files()`
- `length(x)`

CHECKING DATA TYPES & STRUCTURES

1. Know the basic data types and data structures
2. Useful functions: `typeof()` , `class()` : when in doubt, use them both
3. Check the structure of the data structure using: `str()`
4. Check the summary of the data structure using: `summary()`

PLAN FOR NEXT WEEK

That’s it for today! :-)

Next week, we are going to discuss:

1. Distribution of the data
2. Basic statistics

If you want to reach me, mail me at:

`prabesh.dhakal@stud.leuphana.de`