Data Science in R - Data Wrangling

Fall 2021

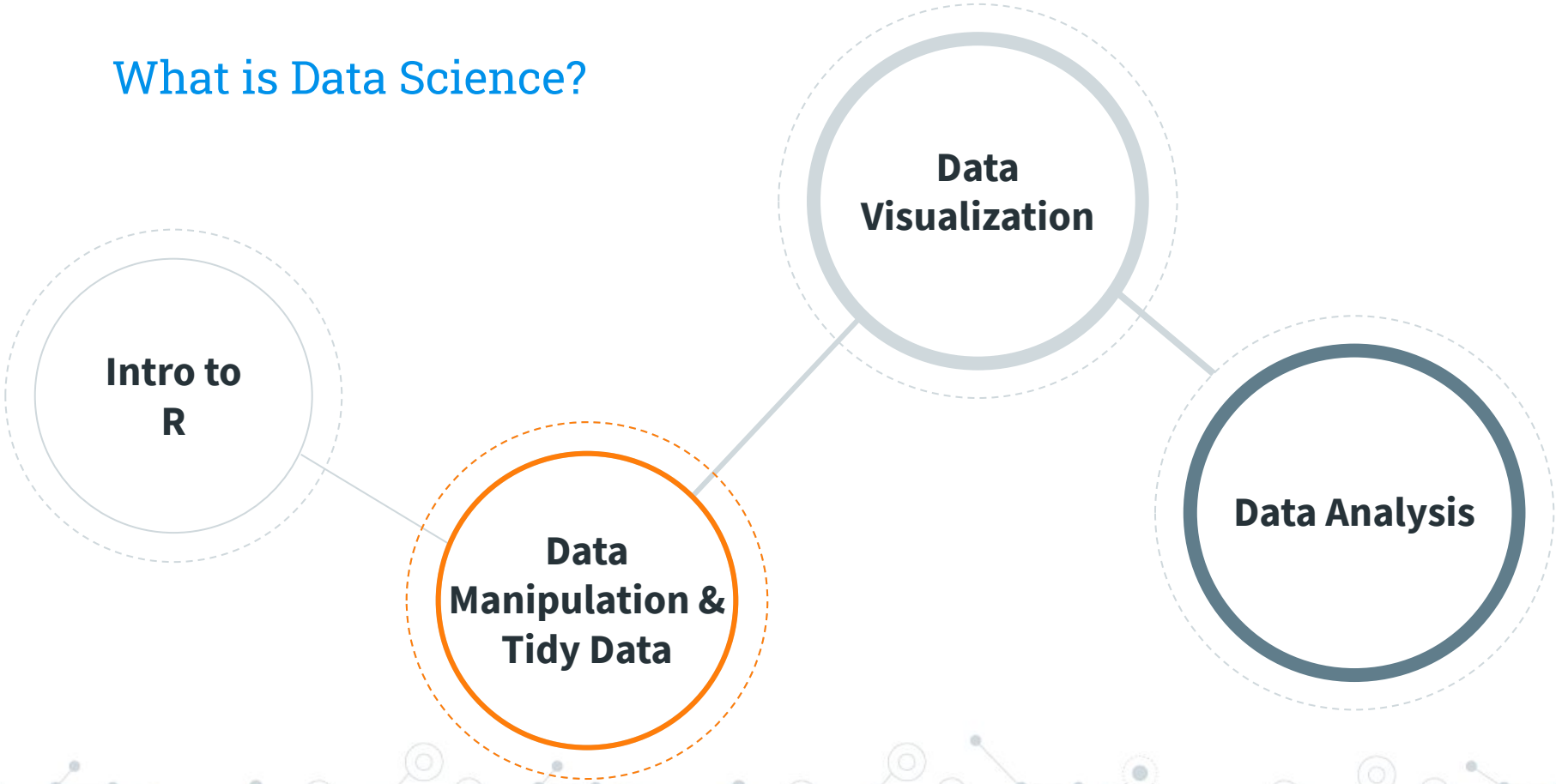
Yama Chang

9/17/2021

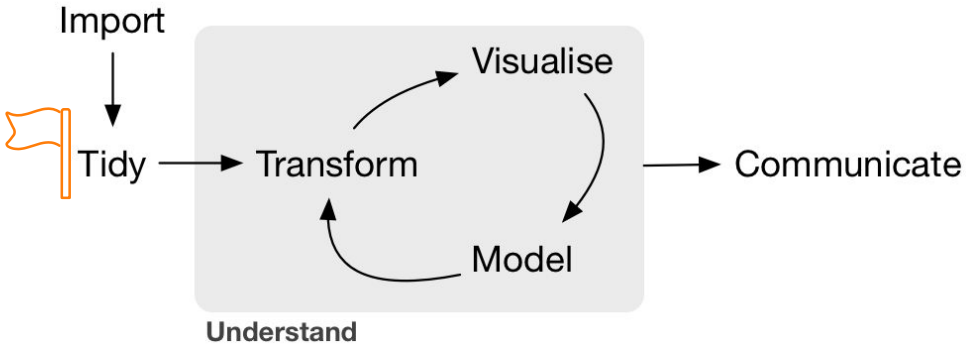
Overview: learning goal

- Import and export data
- Data manipulation
  - Piping
  - Select
  - Filter
  - Mutate
  - Arrange

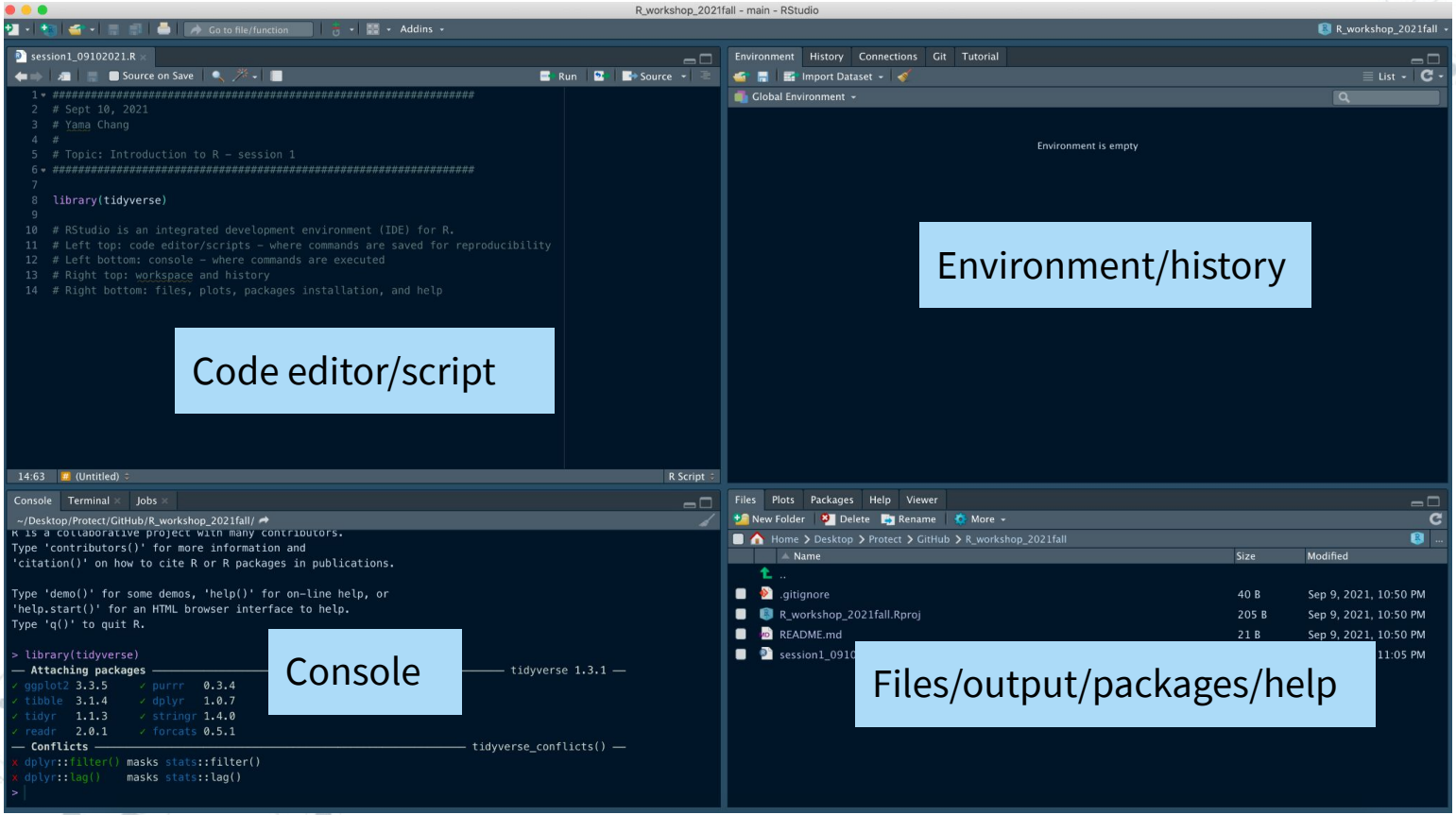
What is Data Science?



A typical data project



R Studio interface



Quick recap...

- Installation and load package

```
install.packages("tidyverse")
library(tidyverse)
```

Quick recap…

- Get your working directory and set working directory

```
getwd()
setwd("/Users/yama/Box/Yama/R workshop")
```

Quick recap...

- Assign value to an object so you can use the object later

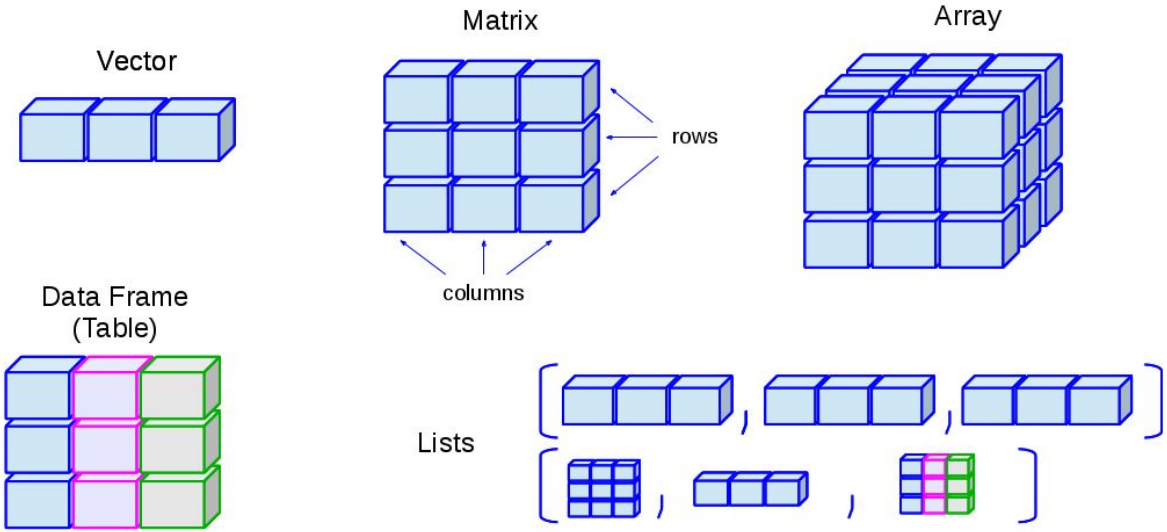
```
object <- value
a <- 5 # We can assign numeric values
b <- “protect” # We can assign character values
c <- c(1, 4, 7) # We can assign combination of values
```

Let's do some coding!

```
x <- 5 + 7
```

Object Value

- Computation: operation and objects
- Data Structure



http://venus.ifca.unican.es/Rintro/dataStruct.html

Import and Export files

- File path
  - Absolute path

    `/Users/yama/Desktop/Protect/GitHub/Recruitment/data/df_rworkshop_p3.csv`
  - Relative path (portable, recommended)

    `./data/df_rworkshop_p3.csv`

Shorthand	Meaning
`~`	Home directory
`.`	Current working directory
`..`	One directory up from current working directory
`../..`	Two directories up from current working directory

Import and Export files

1. Import
   a. Write the function in code editor for importing data frame stored in your folder

```
read_csv("./datafolder/dataframe.csv")
```

Has this shown in your environment? What do you need to do to show this in your environment?

Import and Export files

1. Import
   a. Write the function in code editor for importing data frame stored in your folder
   b. Assign this code to become an object.

```
df <- read_csv(file = "./datafolder/dataframe.csv")
```

Yes! You need to assign (=save) it as an object (we're calling it df here but you can name whatever you want)

Import and Export files

1. Import

```
df <- read_csv("./datafolder/dataframe.csv")
```

2. Export

```
write_csv(df_you_want_to_export, file = "./datafolder/name_your_df.csv")
```

Do you need to assign to an object?

Data Manipulation

- Using Tidyverse package
- Clean and organize data using **dplyr** verbs and piping
  - Select
  - Filter
  - Mutate
  - Arrange
- Practice!

Thank you!

You can find my slides and codes at my GitHub

Also find me at: changy11@upmc.edu