




Data Science in R - Data Wrangling

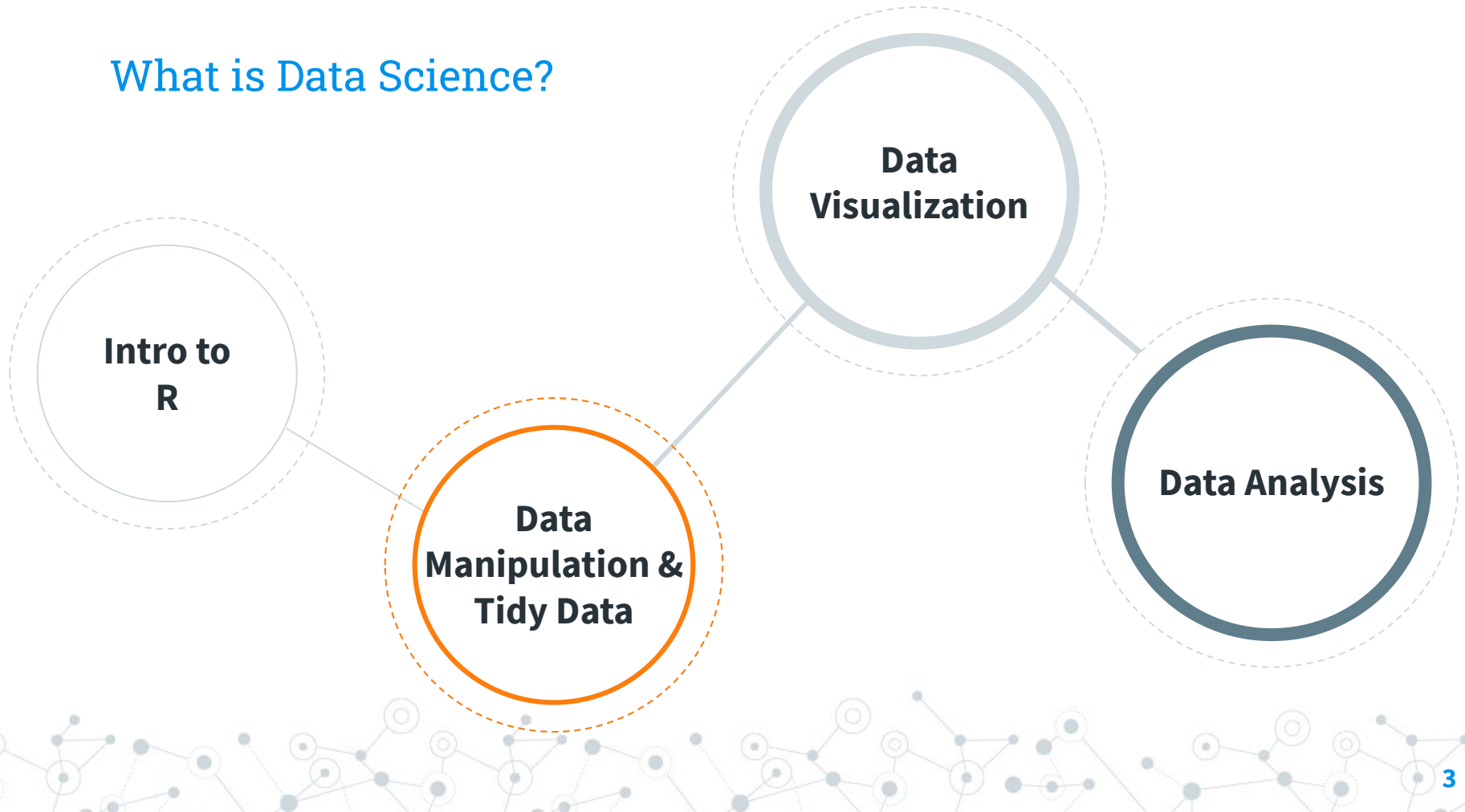
Fall 2021
Yama Chang
9/17/2021



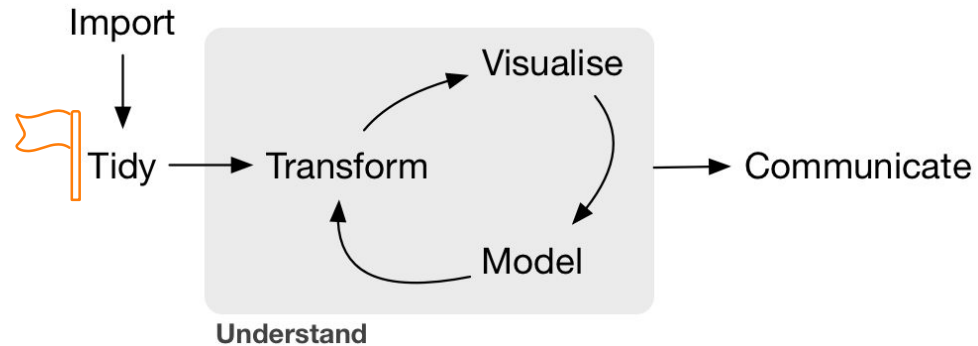
Overview: learning goal

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

What is Data Science?



A typical data project



R Studio interface

The screenshot displays the R Studio IDE interface. The top-left pane is the **Code editor/script**, showing an R script with comments and the `library(tidyverse)` command. The top-right pane is the **Environment/history** pane, which is currently empty. The bottom-left pane is the **Console**, showing the output of the `library(tidyverse)` command, including a list of attached packages and their versions. The bottom-right pane is the **Files/output/packages/help** pane, showing a file explorer view of the current project directory.

Code editor/script

```
1 #####
2 # Sept 10, 2021
3 # Yama Chang
4 #
5 # Topic: Introduction to R - session 1
6 #####
7
8 library(tidyverse)
9
10 # RStudio is an integrated development environment (IDE) for R.
11 # Left top: code editor/scripts - where commands are saved for reproducibility
12 # Left bottom: console - where commands are executed
13 # Right top: workspace and history
14 # Right bottom: files, plots, packages installation, and help
```

Environment/history

Environment is empty

Console

```
> library(tidyverse)
Attaching packages: tidyverse 1.3.1
  ggplot2 3.3.5  purrr  0.3.4
  tibble  3.1.4  dplyr  1.0.7
  tidyr   1.1.3  stringr 1.4.0
  readr   2.0.1  forcats 0.5.1
Conflicts:
  dplyr::filter() masks stats::filter()
  dplyr::lag()    masks stats::lag()
>
```

Files/output/packages/help

| Name | Size | Modified |
|---------------------------|-------|-----------------------|
| .. | | |
| .gitignore | 40 B | Sep 9, 2021, 10:50 PM |
| R_workshop_2021fall.Rproj | 205 B | Sep 9, 2021, 10:50 PM |
| README.md | 21 B | Sep 9, 2021, 10:50 PM |
| session1_0910 | | 11:05 PM |



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!

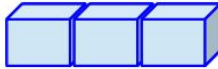
```
x <- 5 + 7
```

Object

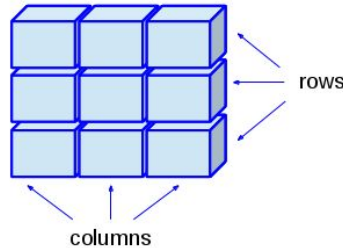
Value

- Computation: operation and objects
- Data Structure

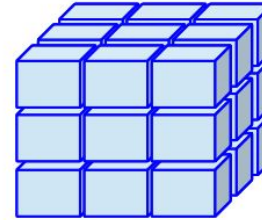
Vector



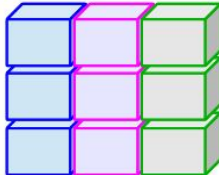
Matrix



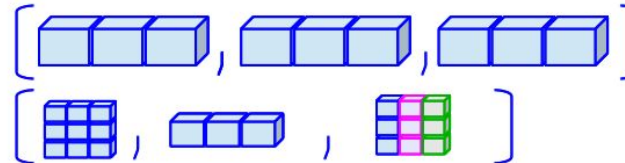
Array



Data Frame
(Table)



Lists



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 |
|--------------------|---|
| <code>~</code> | Home directory |
| <code>.</code> | Current working directory |
| <code>..</code> | One directory up from current working directory |
| <code>../..</code> | 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

