Rotman

INTRO TO R PROGRAMMING

R Tutorial (RSM456) – Session 1

January 8, 2024 Prepared by Jay Cao / TDMDAL

Website: https://tdmdal.github.io/r-intro-2024-winter/

Rotman School of Management
UNIVERSITY OF TORONTO

Plan for Session 1

- What is R and what can R do?
- Setup R and RStudio, an R coding environment
- Get started
  - Navigate RStudio
  - Install and load R packages
  - Load/import a tabular dataset (in csv and Excel format)
- R programming basics
  - Expression and assignment
  - Basic data structures
  - Basic programming structures & functions

What's R?

- R = a language + an eco-system
  - A free and open-source programming language
  - An eco-system of many high-quality user-contributed libraries/packages

- In the past R is mostly known for its statistical analysis toolkits
- Nowadays R is capable of (and very good at) many other tasks
  - Tools that facilitates the whole data analysis workflow
  - Tools for web technology (e.g., web scraping, web app/dashboard development, etc.)
  - Many more…

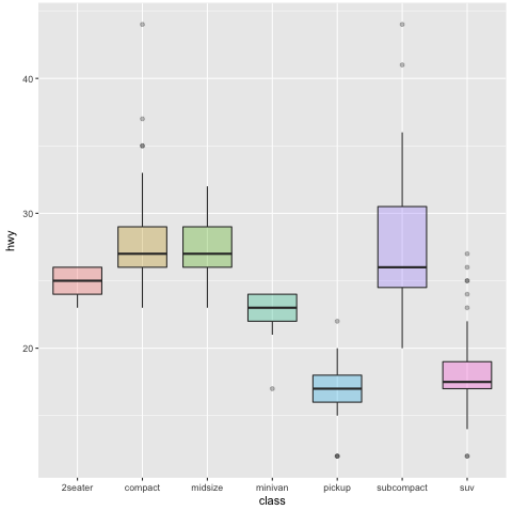
What can R do – Statistics & related

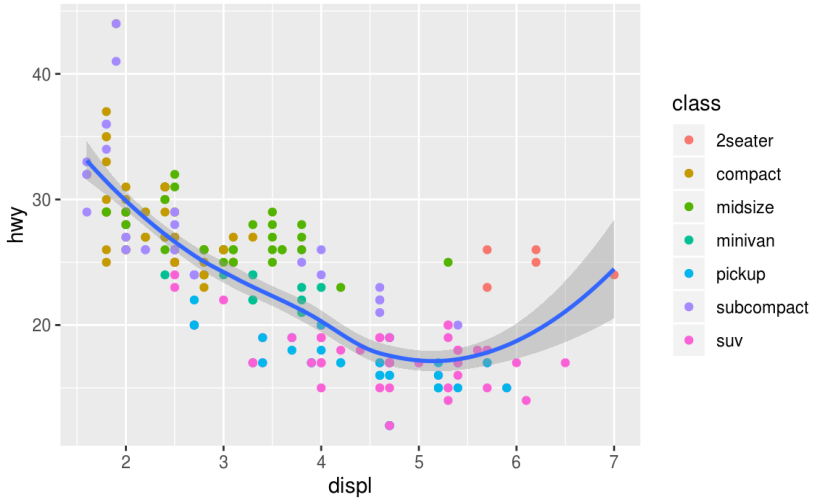
- Statistics & Econometrics
  - Regressions
  - Time series analysis
  - Bayesian inference
  - Survival analysis
  - …
- Numerical Mathematics
  - Optimization
  - Solver
  - Differential equations
  - …

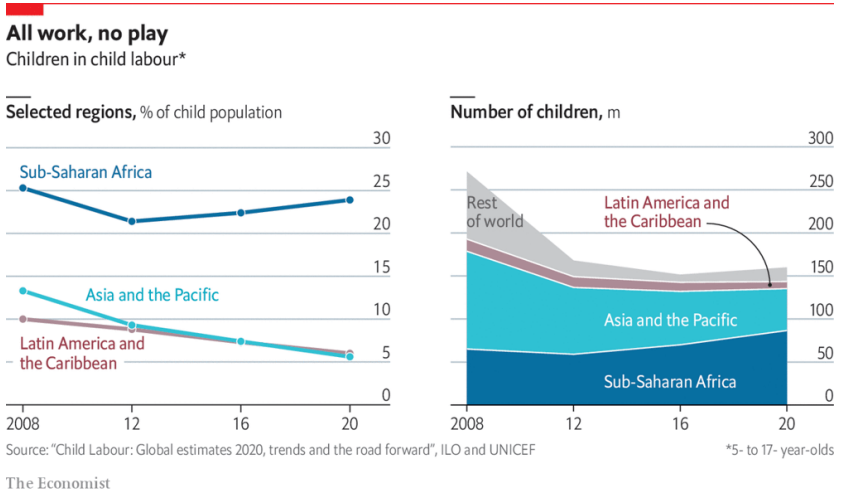
- Finance
  - Portfolio management
  - Risk management
  - Option pricing
  - …
- Machine learning
  - …

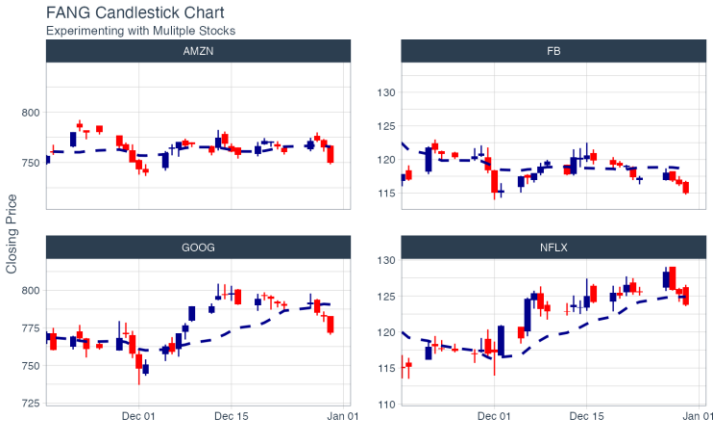
- see **R Task View** for more

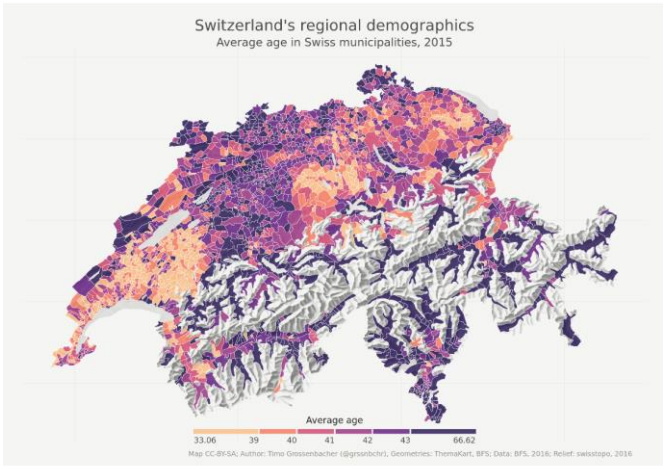
What can R do – Graphics

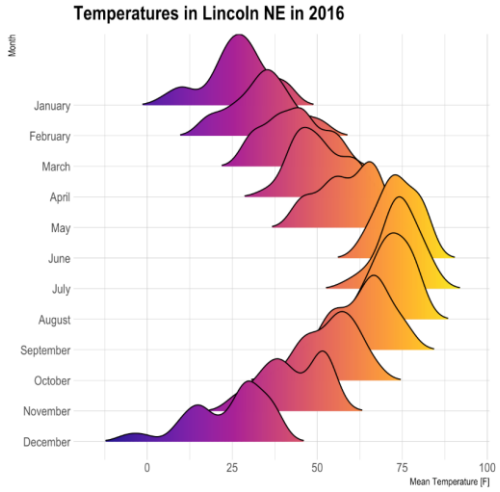












...

Ref: 1) https://www.r-graph-gallery.com/

2) https://timogrossenbacher.ch/2016/12/beautiful-thematic-maps-with-ggplot2-only/;

Setup R (Install R & its Coding Environment)

- **R** & **RStudio** on your **local computer** ← **Our Choice**
  - Install R (https://www.r-project.org/)
  - Install RStudio (https://rstudio.com/products/rstudio/download/)

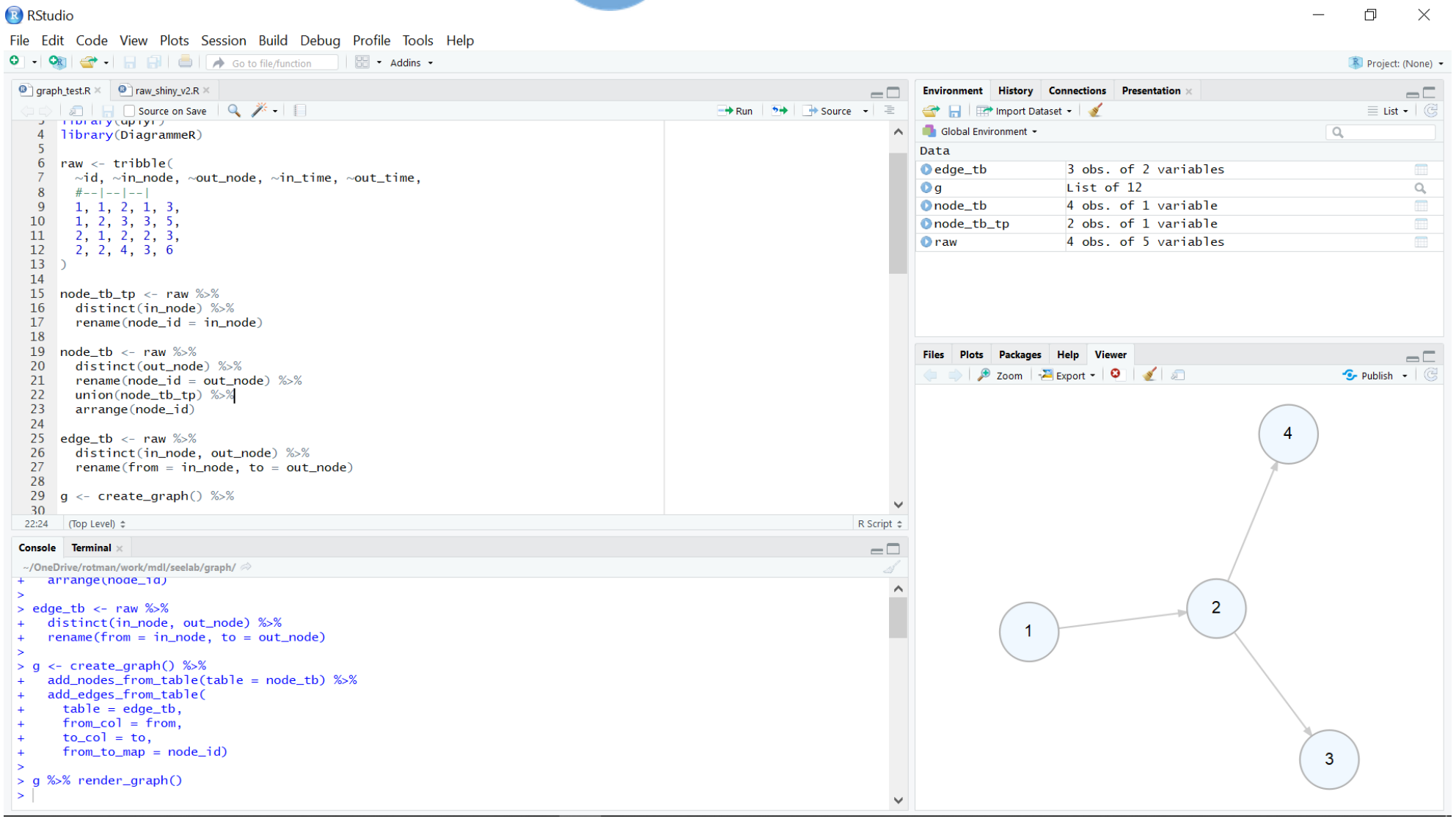
- R & RStudio **in the Cloud** (run R without installation) ← **Backup Options**
  - Option 1: RStudio Cloud (https://rstudio.cloud/)
  - Option 2: UofT JupyterHub RStudio (https://jupyter.utoronto.ca/hub/login)

Note. In this workshop, we will also occasionally use R in Google Colab (https://colab.to/r), a notebook coding environment in the cloud.

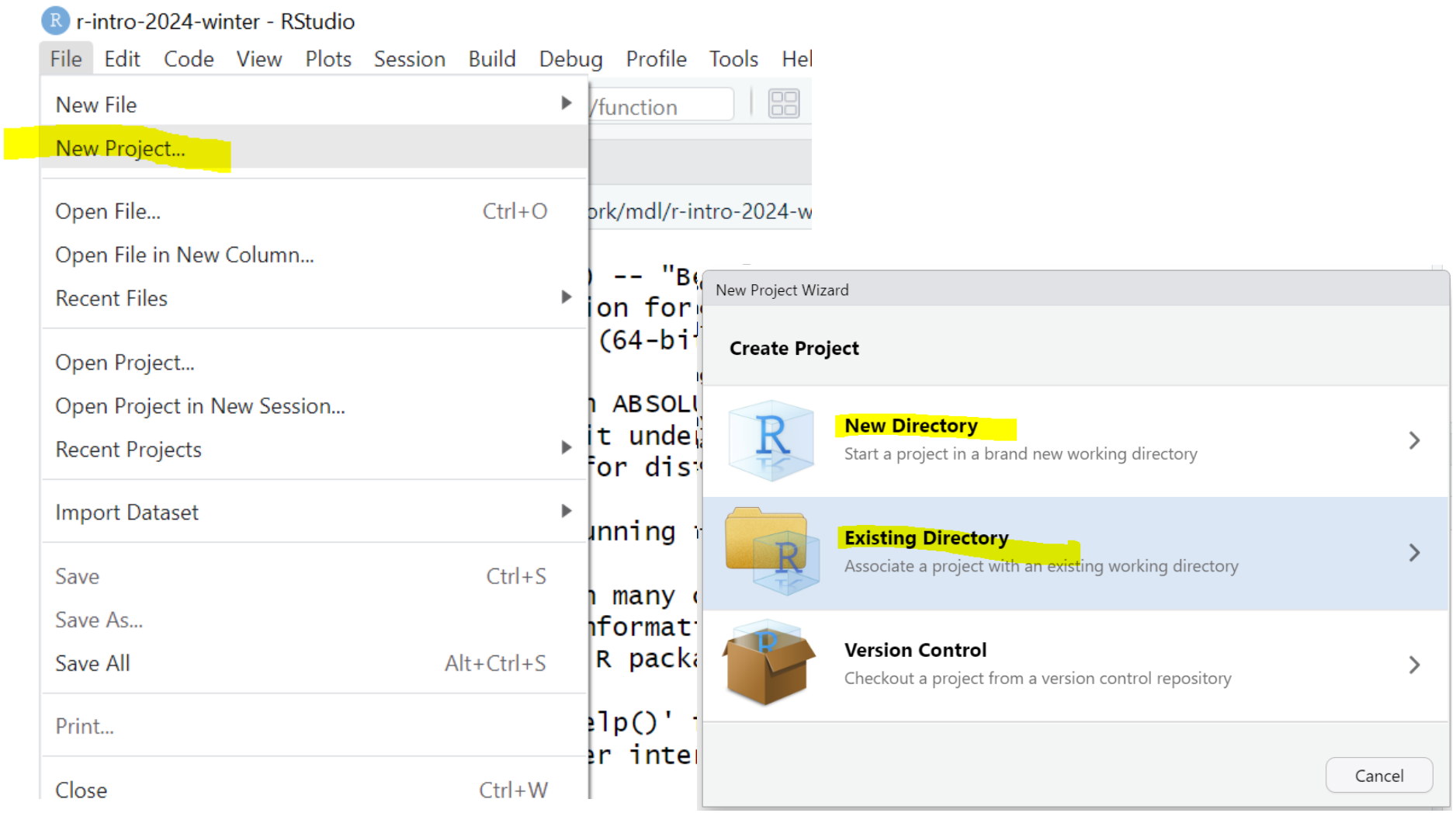
Plan for Session 1

- What is R and what can R do?
- Setup R and RStudio, an R coding environment
- Get started
  - Navigate RStudio
  - Install and load R packages
  - Load/import a tabular dataset
- R programming basics
  - Expression and assignment
  - Basic data structures
  - Basic programming structures & functions

Navigate RStudio



Create New Project – A Good Practice



Install and Load R packages/libraries

- Install an R library (only need to install a library once)

```
install.packages("library_name")
```

- Load an R library (before you use a library)

```
library(library_name)
```

- CRAN (The Comprehensive R Archive Network)
  - CRAN Task Views

Load a CSV file

- What's a CSV file
- `read_csv()` from the `readr`

```
read_csv(file)
```

```
e.g. hprice <- read_csv("hprice.csv")
```

- More about `read_csv()`
  - header row or not, missing values, etc.
- More about `readr`

Load an Excel file

- `read_excel()` from the `readxl`

```
read_excel(path, sheet, skip)
```

```
e.g. country_risk <- read_excel(path =
"country_risk.xlsx", sheet = "raw_kmeans", skip = 1)
```

- More about `read_excel()`
- More about `readxl`

Plan for Session 1

- What is R and what can R do?
- Setup R and RStudio, an R coding environment
- Get started
  - Navigate RStudio
  - Install and load R packages
  - Load/import tabular datasets
- R programming basics
  - Expression and assignment
  - Basic data structures
  - Basic programming structures & functions

Expression and Assignment

```
# expression
2 + sqrt(4) + log(exp(2)) + 2^2


# assignment
x <- 3
y <- (pi == 3.14)
```

R Data Structure - Overview

	Homogeneous	Heterogeneous
1-d	**Atomic vector**	**List**
2-d	Matrix	**Data frame**
n-d	Array	

http://adv-r.had.co.nz/Data-structures.html

R Data Structure - Overview

	Homogeneous	Heterogeneous
1-d	**Atomic vector**	**List**
2-d	Matrix	**Data frame**
n-d	Array	

http://adv-r.had.co.nz/Data-structures.html

Atomic Vectors

```
# create R vectors
vec_character <- c("Hello,", "World!")

vec_integer <- c(1L, 2L, 3L)

vec_double <- c(1.1, 2.2, 3.3)

vec_logical <- c(TRUE, TRUE, FALSE)
```

Hello,	**World!**

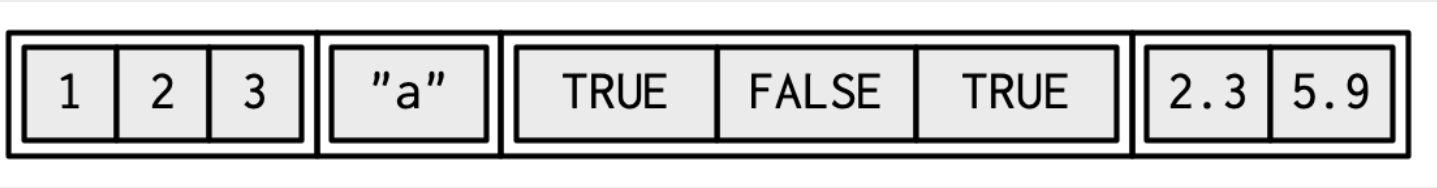
1	**2**	**3**

1.1	**2.2**	**3.3**

TRUE	**TRUE**	**FALSE**

List

```
# create an R list
l1 <- list(
  1:3,
  "a",
  c(TRUE, FALSE, TRUE),
  c(2.3, 5.9)
)
```



ref. https://adv-r.hadley.nz/vectors-chap.html#list-creating

Data Frame

```
# create a data frame
df1 <- data.frame(
  x = 1:3,
  y = letters[1:3],
  z = c(1.1, 2.2, 3.3)
)
```

x	y	z
1	"a"	1.1
2	"b"	2.2
3	"c"	3.3

Data Frame

```
# create a data frame
df1 <- data.frame(
  x = 1:3,
  y = letters[1:3],
  z = c(1.1, 2.2, 3.3)
)
```

x	y	z
1	"a"	1.1
2	"b"	2.2
3	"c"	3.3

Data Frame

```
# create a data frame
df1 <- data.frame(
  x = 1:3,
  y = letters[1:3],
  z = c(1.1, 2.2, 3.3)
)
```

x	y	z
1	"a"	1.1
2	"b"	2.2
3	"c"	3.3

A Cousin to Data Frame - Tibble

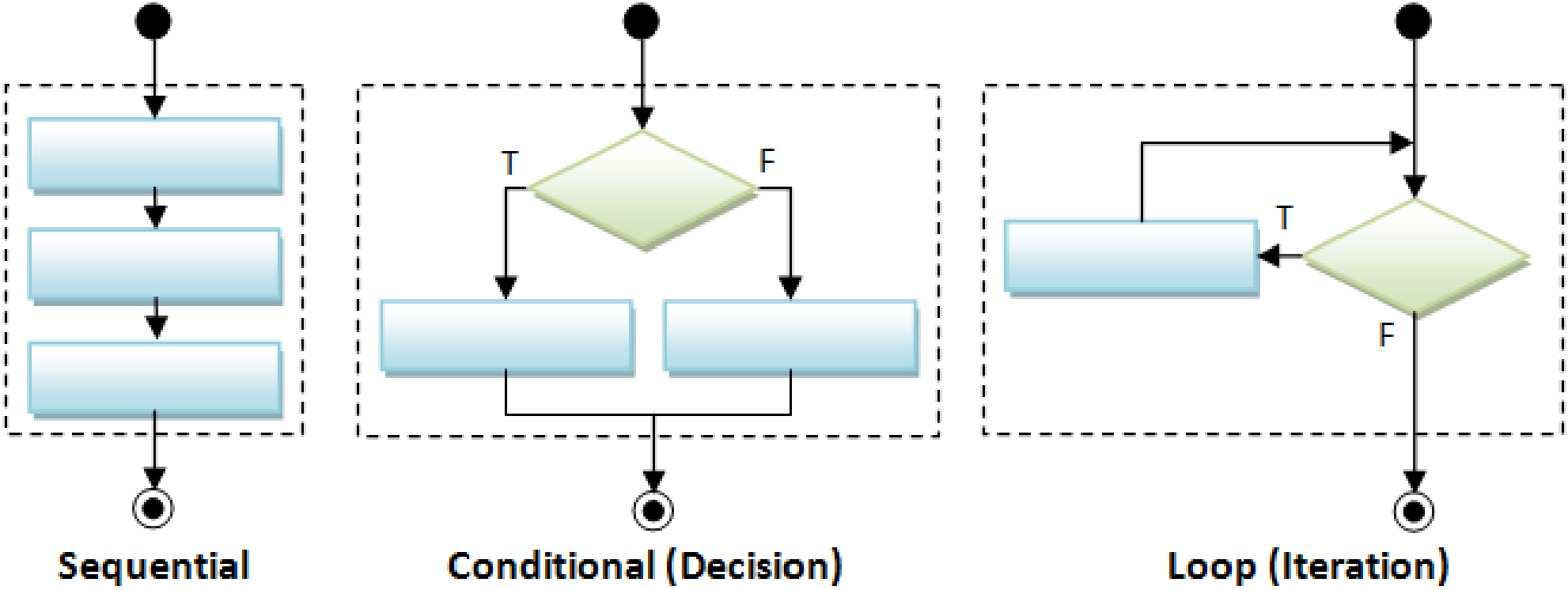
```
# load tibble library (part of tidyverse lib)
library(tibble)

# create a tibble
tb1 <- tibble(
  x = 1:3,
  y = letters[1:3],
  z = c(1.1, 2.2, 3.3)
)
```

x	y	z
1	"a"	1.1
2	"b"	2.2
3	"c"	3.3

https://r4ds.had.co.nz/tibbles.html#tibbles-vs.data.frame

Programming Structure: Control Flows



Sequential

- Example: Sum of Squares

$$\sum_{t=1}^{3} t^2$$

```
# sum of squares
t <- 1:3
y <- sum(t^2)
print(y)
```

Sequential

- Example: Sum of Squares

$$\sum_{t=1}^{3} t^2$$

```
# sum of squares
t <- 1:3
y <- sum(t^2)
print(y)
```

t	1	2	3

Sequential

- Example: Sum of Squares

$$\sum_{t=1}^{3} t^2$$

```
# sum of squares
t <- 1:3
y <- sum(t^2)
print(y)
```

t	1	2	3
t^2	1	4	9
sum(t^2)	14		

Conditional (if…else…)

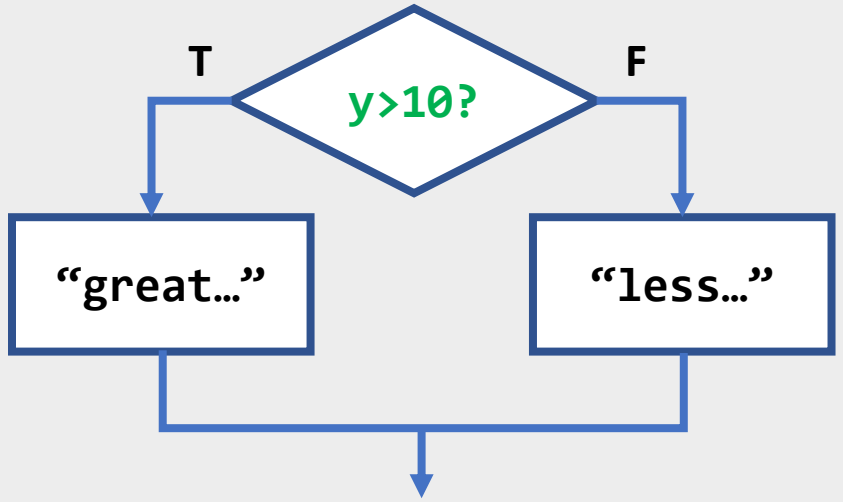
```
if (cond) {
  # run here if cond is TRUE
} else {
  # run here if cond is FALSE
}
```

```
# y greater than 10?
if (y > 10) {
  print("greater than 10")
} else {
  print("less or equal to 10")
}
```

Conditional (if…else…)

```
if (cond) {
  # run here if cond is TRUE
} else {
  # run here if cond is FALSE
}
```

```
# y greater than 10?
if (y > 10) {
  print("greater than 10")
} else {
  print("less or equal to 10")
}
```



Conditional (if…else if…else…)

```
if (cond1) {
  # run here if cond1 is TRUE
} else if (cond2) {
  # run here if cond1 is FALSE but cond2 is TRUE
} else {
  # run here if neither cond1 nor cond2 is TRUE
}
```

Iteration

```
for (var in seq) {
  do something
}

while (cond) {
  do something if cond is TRUE
}
```

```
# sum of squares
t <- 1:3
y <- 0

for (x in t) {
  y <- y + x^2
}

print(y)
```

Programming Structure: Functions

- What's a function
  - a logical block of code
  - input -> output
- Why write functions
  - Reusability
  - Abstraction
  - Maintainability
- Example: $\sum_{t=1}^{n} t^2$

```
# sum of squares from 1 to n
ss <- function(n) {
  t <- 1:n
  sum(t^2)
}

# calling the ss() function
print(ss(2))
print(ss(3))
```

Programming Structure: Functions

- What's a function
  - a logical block of code
  - input -> output
- Why write functions
  - Reusability
  - Abstraction
  - Maintainability
- Example: $\sum_{t=1}^{n} t^2$

```
# sum of squares from 1 to n
ss <- function(n) {
  t <- 1:n
  sum(t^2)
}

# calling the ss() function
print(ss(2))
print(ss(3))
```

Programming Structure: Functions

- What's a function
  - a logical block of code
  - input -> output
- Why write functions
  - Reusability
  - Abstraction
  - Maintainability
- Example: $\sum_{t=1}^{n} t^2$

```r
# sum of squares from 1 to n
ss <- function(n) {
  t <- 1:n
  sum(t^2)  # return(sum(t^2))
}

# calling the ss() function
print(ss(2))
print(ss(3))
```

Turn Ideas into Code

- Solve problems using code: three main ingredients
  - 1) Data Structure (vector, list, **data frame**, etc.)
  - 2) Programming Structure (**sequential**, conditional, iterative)
  - 3) Algorithm (sorting, searching, optimization, **modeling**, etc.)
  - Design to bind the above 3 together (functions, classes, design patterns, software architecture,…)

- Examples
  - Generate and solve Sudoku puzzles
  - Implement and backtest a trading rule/algorithm
  - **Import, manipulate, and model data**

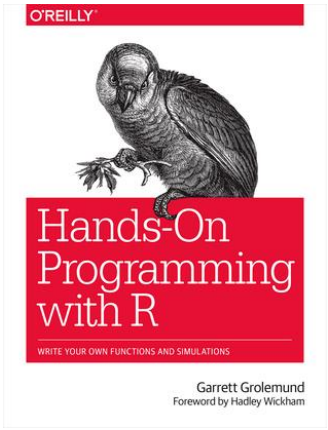
- For us (data analysis in RSM456), in most case,
  - Data frame manipulation + sequential programming flow + modeling (using algorithm already implemented by others)

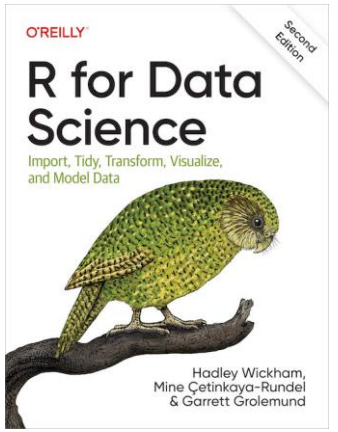
R Learning Road Map (From Zero to Hero)

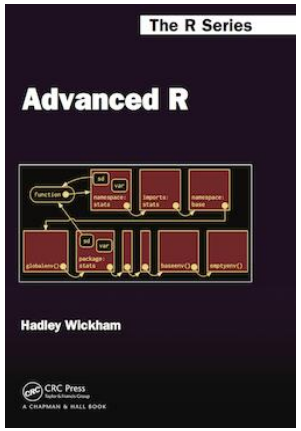
- Step 1. Basic R programming skills (Beginner)
  - Data and programming structure; how to turn an idea into code;
  - Book: Hands-On Programming with R

- Step 2. R Data Science skills (Intermediate)
  - Data wrangling, basic modeling, and visualization/reporting; Best practice;
  - Book: R for Data Science

- Step 3. Take your R Skill to the next level
  - Book: Advanced R







Ref. For other free R books, check bookdown.org often