Introduction to Tidyverse

Claire Liow

University of Tokyo

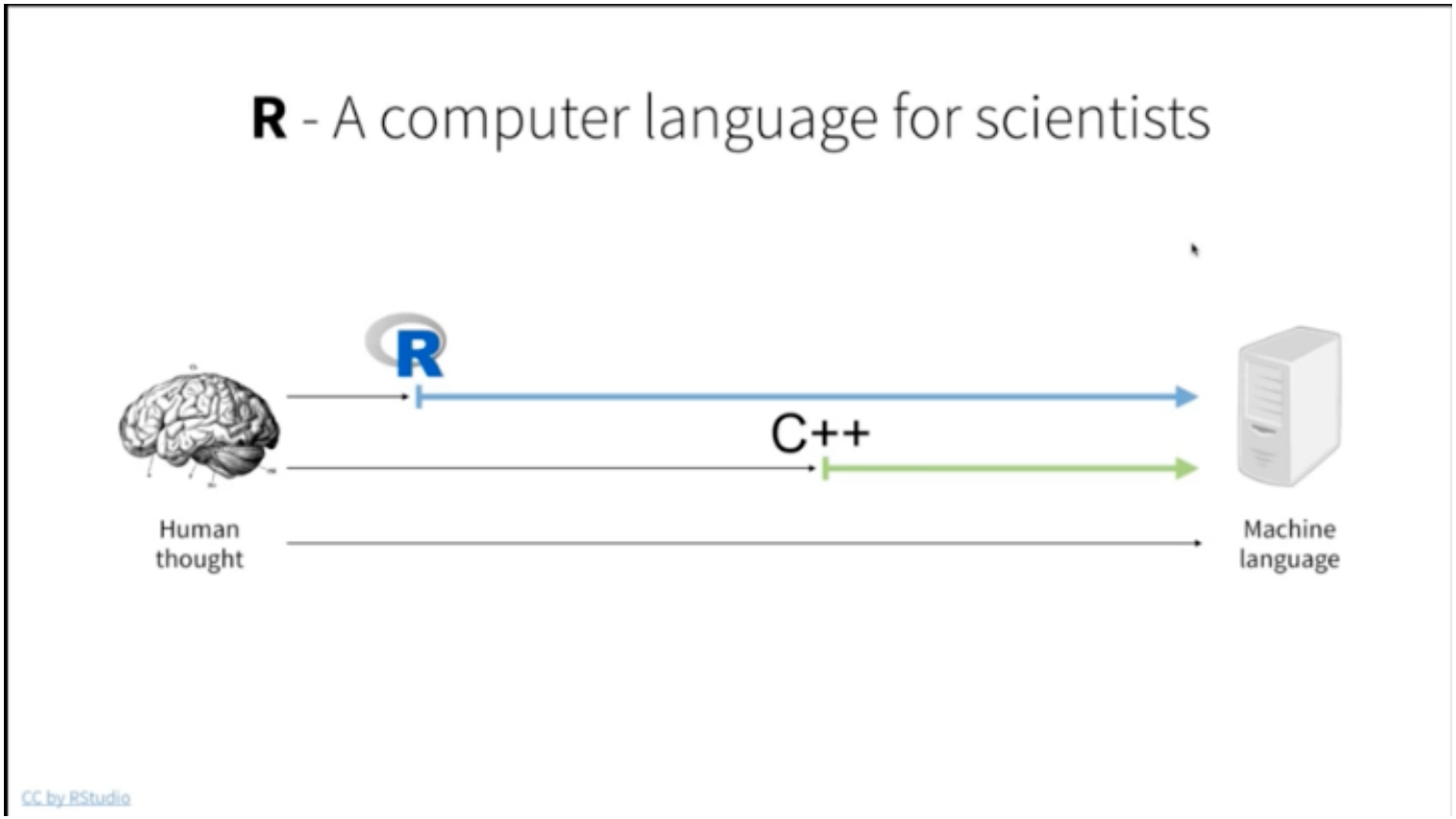
5/24/2022

Section 1

What is Tidyverse?

What is R?

A program language!



Source: RStudio

Then, what is Tidyverse?

- A collection of R packages
- Support natural workflow of data analysis
- Data import, tidying, manipulation, visualization, programming

Load tidyverse

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

Section 2

Base R versus tidyverse syntax

Example

```
UNpop <- read.csv("data/UNpop.csv")
UNpop
```

```
##   year world.pop
## 1 1950   2525779
## 2 1960   3026003
## 3 1970   3691173
## 4 1980   4449049
## 5 1990   5320817
## 6 2000   6127700
## 7 2010   6916183
```

Let's calculate the % of increase from 1950!

Base R syntax

```
UNpop_base <- UNpop
# calculate the ratio compared to 1950
UNpop_base$vs_1950 <-
  UNpop_base$world.pop / UNpop_base$world.pop[1]
# convert to percentage increase and round
UNpop_base$percent_increase <-
  round((UNpop_base$vs_1950 - 1) * 100, 1)
UNpop_base
```

```
##   year world.pop  vs_1950 percent_increase
## 1 1950   2525779 1.000000              0.0
## 2 1960   3026003 1.198047             19.8
## 3 1970   3691173 1.461400             46.1
## 4 1980   4449049 1.761456             76.1
## 5 1990   5320817 2.106604            110.7
## 6 2000   6127700 2.426063            142.6
## 7 2010   6916183 2.738238            173.8
```

tidyverse syntax

```
UNpop_tidy <- UNpop %>%
          # calculate the ratio compared to 1950
  mutate(vs_1950 = world.pop / first(world.pop),
          # convert to percentage increase and round
         percent_increase = round(100 * (vs_1950 - 1), 1)
         )
UNpop_tidy
```

```
##   year world.pop  vs_1950 percent_increase
## 1 1950   2525779 1.000000              0.0
## 2 1960   3026003 1.198047             19.8
## 3 1970   3691173 1.461400             46.1
## 4 1980   4449049 1.761456             76.1
## 5 1990   5320817 2.106604            110.7
## 6 2000   6127700 2.426063            142.6
## 7 2010   6916183 2.738238            173.8
```

Same task, with different syntax

Base-R

```
UNpop_base$vs_1950 <-
  UNpop_base$world.pop / UNpop_base$world.pop[1]
UNpop_base$percent_increase <-
  round((UNpop_base$vs_1950 - 1) * 100, 1)
```

Tidyverse

```
UNpop_tidy <- UNpop %>%
  mutate(vs_1950 = world.pop / first(world.pop),
         percent_increase = round(100 * (vs_1950 - 1), 1))
```

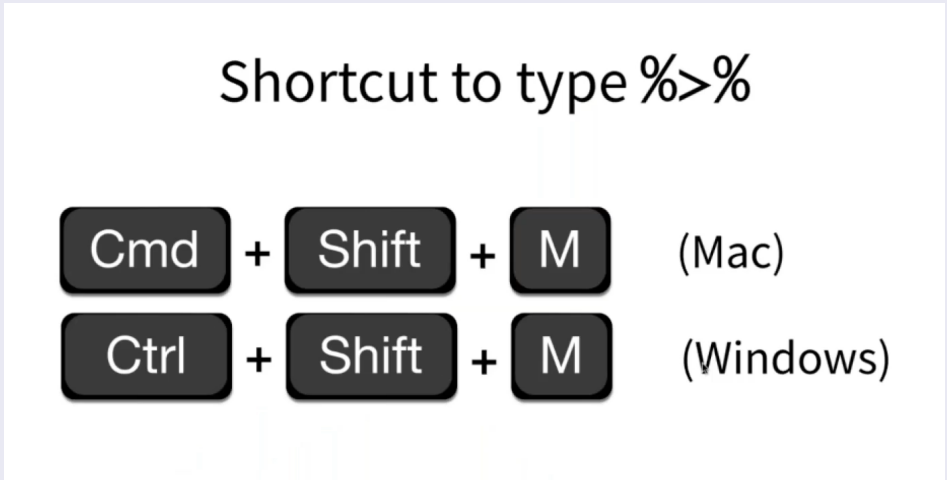
Section 3

Summary

Summary

Differences between Base R/tidyverse?

- Base R: Lots of **$** and **[[]]**
- Tidyverse: **%>%** (Forward pipe operator)



Source: RStudio

When to use Base R/tidyverse?

- Combination of both