Introduction to Tidyverse

Claire Liow

University of Tokyo

5/24/2022

1. Introduction
2. What is tidyverse?
3. Base R versus tidyverse syntax
4. Exercise

Section 1

Introduction

Section 2

What is tidyverse?

What is tidyverse?

```
#![Alt text] (~/Desktop/tidyverse_packages.png)
```

- 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 3

Base R versus tidyverse syntax

Base R versus tidyverse syntax

Base R syntax

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

UNpop$world.pop # access individual variable
UNpop[, "world.pop"] # extract the column "world.pop"
UNpop[1:3, "year"] # subset first three rows of year
```

tidyverse syntax

```
UNpop.tidy <- read_csv("data/UNpop.csv")
class(UNpop.tidy)
glimpse(UNpop.tidy)

select(UNpop.tidy, world.pop) # extracting the world.pop vari
UNpop.tidy %>% # subset the first three rows of years
```

Section 4

Exercise

Subsetting Data and Pipe operator

```r
setwd("~/Dropbox/GitHub/qss-inst-tidyverse/Introduction")
UNpop <- read.csv("data/UNpop.csv")

glimpse(UNpop)
dim(UNpop)s

## Subset the first three rows of UNpop with tidyverse
slice(UNpop, n = 1:3)

## Extract the world.pop variable
select(UNpop, world.pop)

## Pipe operator %>% to link commands together
UNpop %>%
  select(world.pop) %>%
  slice(1:3)
```

Adding new variables: mutate()

```r
## Create an additional column based on existing ones
UNpop.mill <- UNpop %>%
                mutate(world.pop.mill = world.pop / 1000)

## Conditional statement: if_else()
UNpop.mill <- UNpop.mill %>%
  mutate(after.1980 = if_else(year >= 1980, "yes", "no"))

## Conditional symbol: %in%
target.years <- c(1950, 1980, 2000)

UNpop.mill <- UNpop.mill %>%
  mutate(year.of.interest = if_else(year %in% target.years, "y
```

Summarizing data frames: summarize()

```r
## Base R
summary(UNpop.mill)
mean(UNpop.mill$world.pop)

## Tidyverse
UNpop.mill %>%
  summarize(mean.pop = mean(world.pop.mill),
            median.pop = median(world.pop.mill))

UNpop.mill %>%
  group_by(after.1980) %>%
  summarize(mean.pop = mean(world.pop.mill))
```