R workshop

Data manupulation & Rmarkdown

Bolin Wu
NEAR, Aging Research Center

2023/01/26 (updated: 2023-01-26)

workshopr package

Install the package

```
install.packages("remotes")
remotes::install_github("Bolin-Wu/workshopr",
  subdir = "rpackage",
  force = TRUE
)
```

Load the package

```
library(workshopr)
library(tidyverse)
```

Data manupulation

tidyverse, assign

Introduction

This session is to share useful data manipulation skills at daily epidemiology work. My main goal is to follow the "don't repeat yourself" (DRY) principle.

It can make our code more readable and reduce our chance of making mistakes.

Content

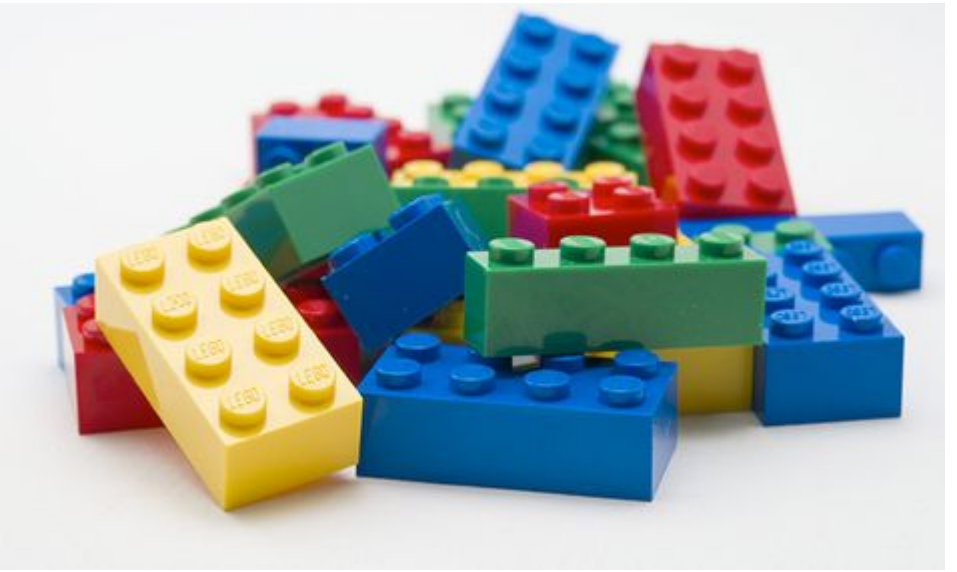
- %>% syntax
- transform data shape
- get the label from DTA and SPSS in R
- filter variables based on name pattern
- check missing values
- bin variables by percentiles
- assign function

Tidyverse

Tidyverse is a huge umbrella which includes several useful visualization and data manipulation package.

We have to

Beautiful syntax with pipeline, just like playing LEGO.



Transform data shape

Transform data shape for many people, including me, sounds troublesome. In R, its relevant functions are evolving overtime as well.

In the beginning (2019), I used spread and gather. Every time I use `spread` and `gather`, it takes me a while to figure out how to fill in 'key' and 'value'. But as you can see from their documentation, their 'lifecycle' is 'superseded'.

Now I only use `pivot_longer` and `pivot_wider` for transforming data. You can find their comprehensive documentation here. They come with better documentation, more powerful application, and better integration with tidyverse syntax.

Example

Let's assumme we received a wide format data:

```
head(fake_snack_df)
```

```
## # A tibble: 6 x 15
##   Lopnr    sex Date_wave1 education Date_wave2 Date_wave3
##   <int>  <dbl> <date>         <dbl> <date>     <date>
## 1  1284 -1.62  1848-01-16   -1.00   1703-06-26 1807-08-10
## 2  2013  0.384 1788-07-16    0.883  1785-09-24 1811-08-26
## 3  1618 -0.363 1833-09-17    0.0753 1792-10-19 1781-01-06
## 4  3190  0.682 1803-04-08    0.0943 1829-04-27 1744-02-04
```

R markdown

Basics and daily work uses