



**Karolinska  
Institutet**

# R workshop

## Data manipulation & Rmarkdown

Bolin Wu

NEAR, Aging Research Center

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



# Data manipulation

assign, tidyverse

# Introduction

## Install the workshopr package

```
# run the install command if did not do it before  
install.packages("remotes")  
remotes::install_github("Bolin-Wu/workshopr", subdir
```

## Load the package

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

This session is 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.

The content includes:

- transform data shape
- get the label from DTA and SPSS in R
- assign function
- apply, lapply, sapply
- filter variables based on name pattern
- check missing values
- bin variables by percentiles

# Transform data shape

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

In the beginning, 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 But as you can see from their documentation, their 'lifecycle' is 'superseded'.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 assume 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



