

Applied Statistics for Data Scientists with R

Class 06: Data Wrangling with Base R and dplyr



Learning Objective

- 1. Base R data manipulation vs dplyr
- 2. dplyr: select, filter, arrange, mutate, group by, summarize

Key Data Wrangling Tasks



Data wrangling means transforming raw data into a structured and usable format for analysis.

- Filter rows
- Select columns
- Create / remove columns
- Edit columns / rows
- Arranging
- Data reshaping
- Join data frames

Reshape Data Frames



"Long" format

country	year	metric
Х	1960	10
Х	1970	13
Х	2010	15
у	1960	20
у	1970	23
у	2010	25
Z	1960	30
Z	1970	33
Z	2010	35

"Wide" format

country	yr1960	yr1970	yr2010
X	10	13	15
у	20	23	25
Z	30	33	35

Joining Datasets



ID	T1	T2
1	а	d
2	b	е

inner_join(df1, df2)

INNER join



left_join(df1, df2)

ID	T1	T2
1	а	d
2	b	е
3	С	NA

LEFT join



df1

df2

ID	T1	ID	T2
1	а	1	d
2	b	2	е
3	С	99	f

. . 1



RIGHT join	

ID	T1	T2
1	а	d
2	b	е
99	NA	f

right_join(df1, df2)



FULL join

full_join(df1, df2)

ID	T1	T2
1	а	d
2	b	е
3	С	NA
99	NA	f

Some Add-ins to Make Life Easier



- questionr
- Styler
- reprex