

Day 11

Topic Covered: Data View Manipulation in R using `tidyr` and `reshape2`

Summary:

Today's session focused on data view manipulation in R, specifically converting data between wider and longer formats using the `tidyr` and `reshape2` libraries. Reshaping data is important for creating tidy datasets suitable for analysis, visualization, and modeling. The session covered two main tasks: converting wider view to longer view, and converting longer view to wider view. These techniques helped reorganize student marks data efficiently.

New Concepts Learned:

1. Using `tidyr` Library

Example data frame:

```
st_test1 <- data.frame(  
  id = c(1,2,3),  
  name = c("alice","bob","john"),  
  test1 = c(92,91,85),  
  test2 = c(98,80,87),  
  test3 = c(92,93,90)  
)
```

Key functions learned:

- Wider to longer view:

```
pivot_longer(st_test1, cols = starts_with("test"), names_to = "test", values_to = "score")
```

This converts multiple test columns into a single column of test names and scores.

- Longer to wider view:

```
st_test2 %>% pivot_wider(names_from = "test", values_from = "score")  
st_test3 <- spread(st_test2, test, score)
```

These commands reshape long data back to wide format with separate test columns.

2. Using reshape2 Library

Example data frame is the same as above.

Key function learned:

- Wider to longer view using melt():

```
melt(st_test1, id.vars = c("id", "name"), variable.name = "test", value.name = "score")
```

This converts wide-format test columns into long-format data, similar to pivot_longer().

Activity:

- Created a student marks data frame
- Converted data from wider view to longer view using tidyverse and reshape2
- Converted data back from longer view to wider view using pivot_wider() and spread()
- Compared outputs to understand the differences between the libraries

Challenges Faced:

Remembering the correct syntax for pivot_longer, pivot_wider, and melt was slightly confusing. Ensuring that column names and values were correctly mapped during reshaping required careful checking. Managing intermediate data frames while switching between views also needed attention.

Key Takeaway:

Both tidyverse and reshape2 provide useful tools for reshaping datasets in R. Converting between longer and wider views is essential for tidy data analysis. Mastering these functions improves skills in data cleaning, preparation, and organization.