Lab 4 How to merge two dataframes

Xiaoqin Yang
yangxiaoqin@suda.edu.cn

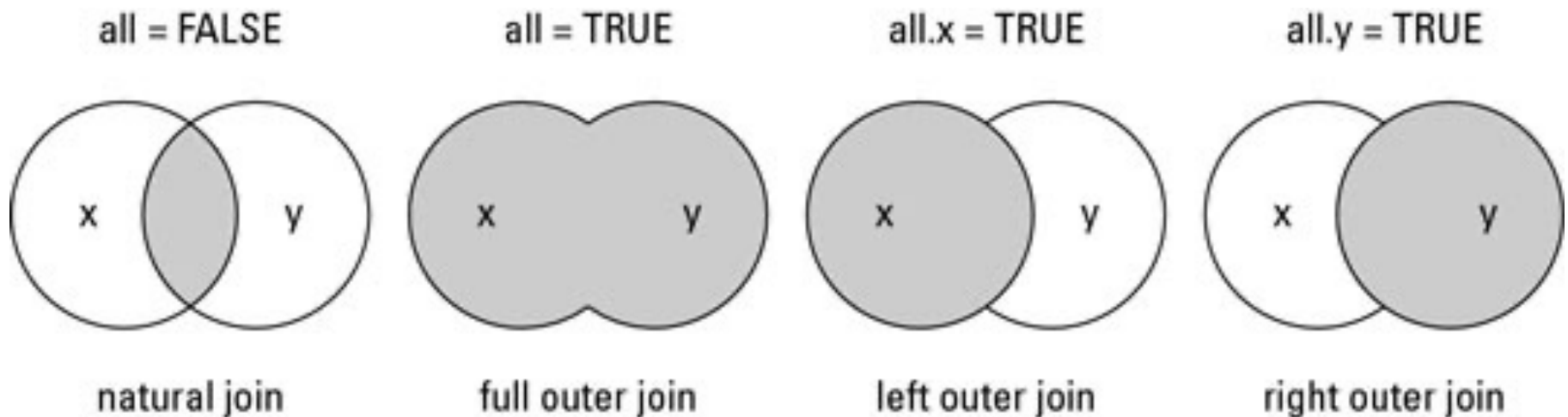# Merging Data Frames by Common Column

- **Problem**
  - You have two data frames that share a common column. You want to merge their rows into one data frame by matching on the common column.

- **Solution**
  - Use the **merge** function to join the data frames into one new data frame based on the common column:

```
> born <- data.frame(name=c("Moe","Larry","Curly", "Harry"),
+         year.born=c(1887,1902,1903, 1964))
> died <- data.frame(name=c("Curly","Moe","Larry"),
+         year.died=c(1952,1975,1975))
> merge(born, died, by="name")
  name year.born year.died
1 Curly     1903     1952
2 Larry     1902     1975
3   Moe     1887     1975
```
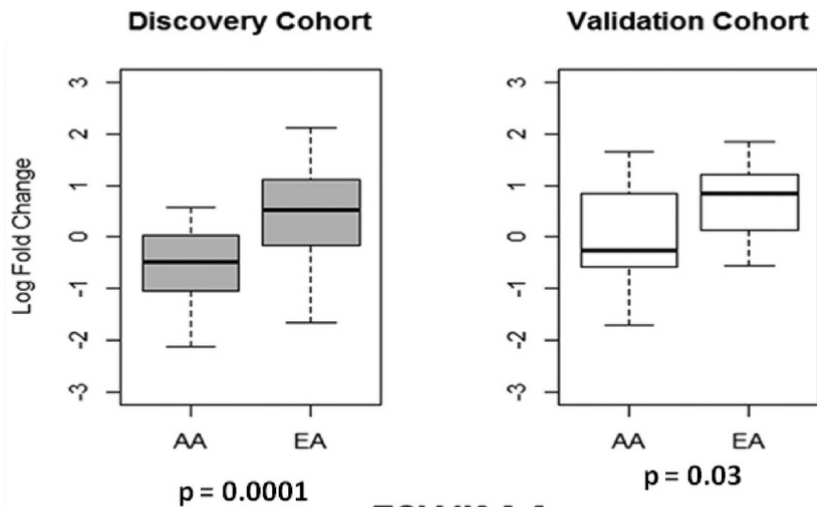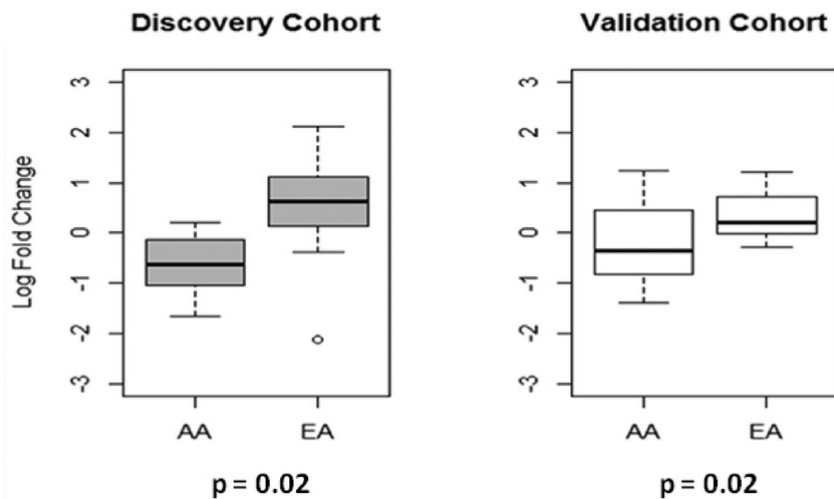
Option "all" for merge function

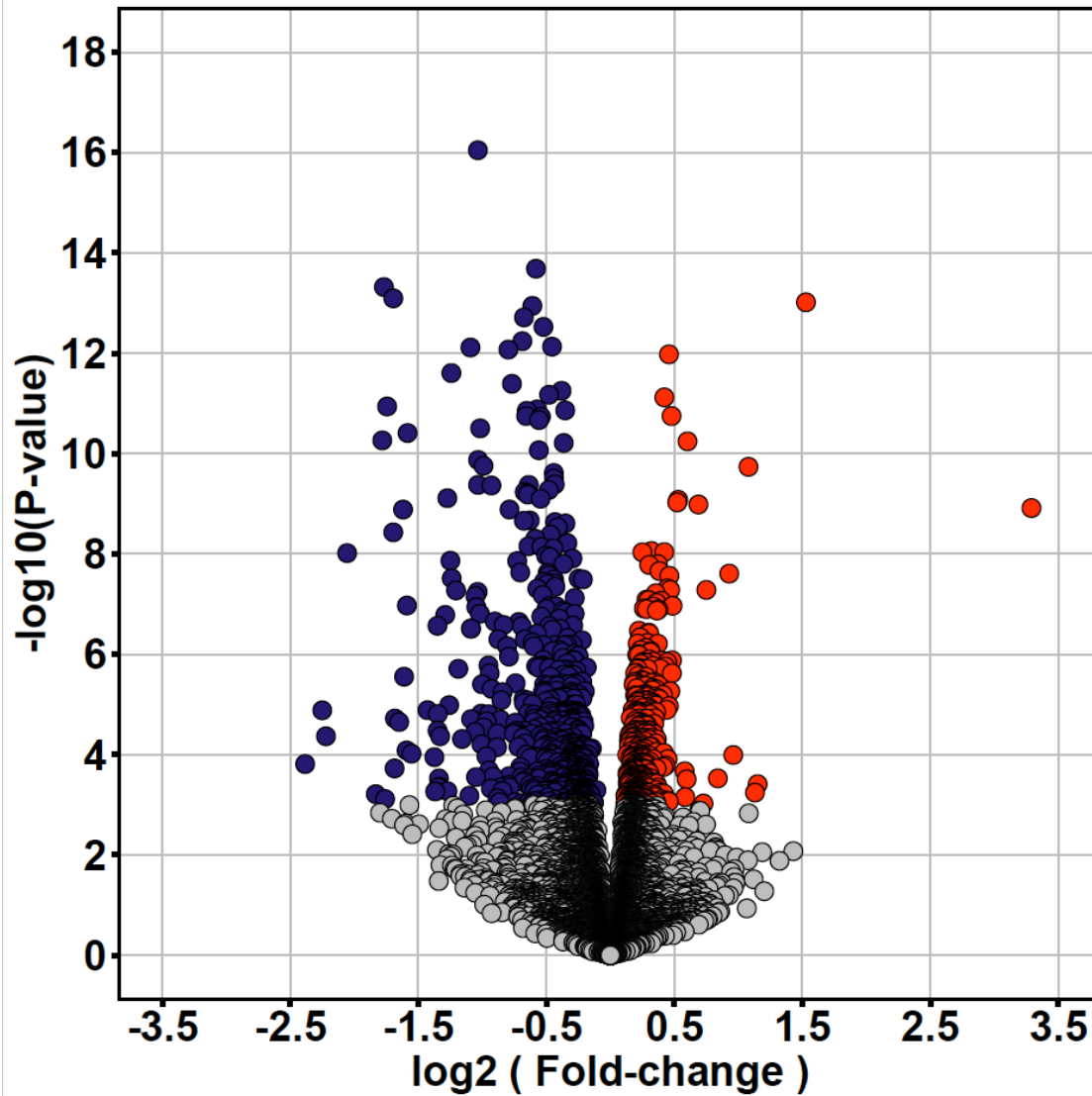| all = FALSE	all = TRUE	all.x = TRUE	all.y = TRUE
natural join	full outer join	left outer join	right outer join

Individual plot for the difference

An example for Volcano plot

We have a problem ….

- A volcano plot is a type of scatterplot that shows statistical significance (P-value) versus magnitude of change (fold change).

- The dataset we have contains the P-values and fold changes, but with some missing values.

- The order in the two data.frames are also different.

- And we should set a significance cutoff (P < 0.05).

The task

- Load in the data, match according to the row names.

- Transform the values to fit the volcano plot.

- Annotate the difference by different colors.