

## C12-5 Combining data frames

When thinking about combining two data frames one has to distinguish between merging them by the values given in a specific column or consecutively putting them together just rows after rows.

For the following examples, we will use these two data frames.

```
x <- c(1, 2, 3, 4, 5)
y <- c(1.4, 2.5, 3.6, 4.0, 5.5)
z <- c("A", "B", "C", "D", "E")
df1 <- data.frame(X = x, Y = y, Z = z)
df1
```

```
##   X   Y Z
## 1 1 1.4 A
## 2 2 2.5 B
## 3 3 3.6 C
## 4 4 4.0 D
## 5 5 5.5 E
```

```
y <- c(14, 25, 36, 40, 55)
x <- c(100, 200, 300, 400, 500)
z <- c("A", "C", "E", "D", "B")
df2 <- data.frame(X = x, Y = y, Z = z)
df2
```

```
##   X   Y Z
## 1 100 14 A
## 2 200 25 C
## 3 300 36 E
## 4 400 40 D
## 5 500 55 B
```

To combine two data frames, use the `rbind` function provided that both data frames have the same column names and the same number of columns.

```
dfr <- rbind(df1, df2)
dfr
```

```
##   X   Y Z
## 1  1 1.4 A
## 2  2 2.5 B
## 3  3 3.6 C
## 4  4 4.0 D
## 5  5 5.5 E
## 6 100 14.0 A
```

```
## 7 200 25.0 C
## 8 300 36.0 E
## 9 400 40.0 D
## 10 500 55.0 B
```

Note, that it is not necessary that the columns are in the same order since the concatenation is done by the column names, not the column sequence:

```
df3 <- df2[,c(1,3,2)]
df3
```

```
##      X Z  Y
## 1 100 A 14
## 2 200 C 25
## 3 300 E 36
## 4 400 D 40
## 5 500 B 55
```

```
dfr <- rbind(df1, df3)
dfr
```

```
##      X      Y Z
## 1     1   1.4 A
## 2     2   2.5 B
## 3     3   3.6 C
## 4     4   4.0 D
## 5     5   5.5 E
## 6  100 14.0 A
## 7  200 25.0 C
## 8  300 36.0 E
## 9  400 40.0 D
## 10 500 55.0 B
```

To merge two data frames, you have to specify the column containing the values which should be used in the matching. In this example, we use the values of the column Z to merge the two data frames.

```
dfc <- merge(df1, df2, by = "Z")
dfc
```

```
##      Z X.x Y.x X.y Y.y
## 1 A     1 1.4 100  14
## 2 B     2 2.5 500  55
## 3 C     3 3.6 200  25
## 4 D     4 4.0 400  40
## 5 E     5 5.5 300  36
```

Since both data frames have identical column names, X and Y is added to the column names in the resulting data frame to indicate if the column is from the first (i.e. X) or second (i.e. Y) data frame. You

can easily rename the columns using the `colnames` function.

If the columns which should be used for merging have different names, no problem: just supply the column names separately for the first (i.e. X) and second (i.e. Y) data frame:

```
colnames(df2) <- c("H", "I", "J")
dfc <- merge(df1, df2, by.x = "Z", by.y = "J")
dfc
```

```
##   Z X   Y   H I
## 1 A 1 1.4 100 14
## 2 B 2 2.5 500 55
## 3 C 3 3.6 200 25
## 4 D 4 4.0 400 40
## 5 E 5 5.5 300 36
```

Since not only the names of the columns used for merging but all column names are different, no X or Y is added in the output column names.