

How to merge datasets

A. Blejec

July 30, 2013

1 Countries dataset

Read the country data

```
> sData <- read.table("../data/sData.txt",header=TRUE)
> dim(sData)

[1] 52 11

> names(sData)

[1] "country"      "continent"    "infantMort"   "lifeExpect"
[5] "population"   "birthRate"    "nPhysicians"  "physicians"
[9] "gdp"          "internetUsers" "tvSets"
```

For our purpose we will split the data into two datasets. We will keep the variable country in both. We will also have different countries in each dataset, with few countries in both sets.

```
> first <- sData[1:10,1:3]
> head(first)

  country continent infantMort
1  Albania   Europe      20.0
2  Angola    Africa     184.4
3  Australia Pacific      4.6
4  Austria   Europe      4.5
5  Bangladesh Asia       59.1
6   Brazil S America    27.6

> second <- sData[5:20,c(1,3,7:8)]
> head(second)

  country infantMort nPhysicians physicians
5  Bangladesh     59.1      38485         0.26
6   Brazil       27.6     198153         1.15
7   Canada       4.6      66583         2.14
8    Chile       8.4      17250         1.09
9    China      22.1    1364000         1.06
10 Costa Rica     9.5       5204         1.32
```

Merge the country data into the new dataset

```
> new <- merge(x=first,y=second,by="country",  
+             all=TRUE)  
> dim(new)
```

```
[1] 20 6
```

Show first six lines

```
> head(new)
```

	country	continent	infantMort.x	infantMort.y	nPhysicians	physicians
1	Albania	Europe	20.0	NA	NA	NA
2	Angola	Africa	184.4	NA	NA	NA
3	Australia	Pacific	4.6	NA	NA	NA
4	Austria	Europe	4.5	NA	NA	NA
5	Bangladesh	Asia	59.1	59.1	38485	0.26
6	Brazil	S America	27.6	27.6	198153	1.15

Show all

```
> new
```

	country	continent	infantMort.x	infantMort.y	nPhysicians	physicians
1	Albania	Europe	20.0	NA	NA	NA
2	Angola	Africa	184.4	NA	NA	NA
3	Australia	Pacific	4.6	NA	NA	NA
4	Austria	Europe	4.5	NA	NA	NA
5	Bangladesh	Asia	59.1	59.1	38485	0.26
6	Brazil	S America	27.6	27.6	198153	1.15
7	Canada	N America	4.6	4.6	66583	2.14
8	Chile	S America	8.4	8.4	17250	1.09
9	China	Asia	22.1	22.1	1364000	1.06
10	Costa Rica	S America	9.5	9.5	5204	1.32
11	Cyprus	<NA>	NA	6.9	1864	2.34
12	Czech Republic	<NA>	NA	3.9	35960	3.51
13	Denmark	<NA>	NA	4.5	15653	2.93
14	Ecuador	<NA>	NA	22.1	18335	1.48
15	Egypt	<NA>	NA	30.1	38485	0.54
16	Finland	<NA>	NA	3.5	16446	3.16
17	France	<NA>	NA	4.2	203487	3.37
18	Germany	<NA>	NA	4.1	277885	3.37
19	Greece	<NA>	NA	5.3	47944	4.38
20	Guatemala	<NA>	NA	29.8	9965	0.90