Chap02. Bind_Merge_Condition

Peter

2018-11-28

cbind test

```
a = c('A', 'B', 'C', 'D', 'E')
b = c(0.280, 0.257, 0.312, 0.266, 0.295)
c = cbind(a, b)
c
```

```
##      a   b
## [1,] "A" "0.28"
## [2,] "B" "0.257"
## [3,] "C" "0.312"
## [4,] "D" "0.266"
## [5,] "E" "0.295"
```

```
colnames(c) = c('player', 'avg')
c
```

```
##      player avg
## [1,] "A"    "0.28"
## [2,] "B"    "0.257"
## [3,] "C"    "0.312"
## [4,] "D"    "0.266"
## [5,] "E"    "0.295"
```

```
age = c(26, 23, 31, 27, 24)
d = cbind(c, age)
d
```

```
##      player avg     age
## [1,] "A"    "0.28"  "26"
## [2,] "B"    "0.257" "23"
## [3,] "C"    "0.312" "31"
## [4,] "D"    "0.266" "27"
## [5,] "E"    "0.295" "24"
```

rbind test

```
f = rbind(a, b)
f
```

```
##   [,1]   [,2]    [,3]    [,4]    [,5]
## a "A"    "B"     "C"     "D"     "E"
## b "0.28" "0.257" "0.312" "0.266" "0.295"
```

merge test

```
d = matrix(c('C', 'D', 'E', 'B', 'A', 26, 23, 31, 27, 24), ncol=2)
d
```

```
##      [,1] [,2]
## [1,] "C"  "26"
## [2,] "D"  "23"
## [3,] "E"  "31"
## [4,] "B"  "27"
## [5,] "A"  "24"
```

Change name of variables

```
colnames(d) = c('player', 'age')
d
```

```
##      player age
## [1,] "C"    "26"
## [2,] "D"    "23"
## [3,] "E"    "31"
## [4,] "B"    "27"
## [5,] "A"    "24"
```

Do Merge

```
e = merge(c, d, by='player')
e
```

```
##   player   avg age
## 1      A  0.28  24
## 2      B 0.257  27
## 3      C 0.312  26
## 4      D 0.266  23
## 5      E 0.295  31
```

Convert discrete variables into Categorical variables

```
str(e) # check table structures
```

```
## 'data.frame':    5 obs. of  3 variables:
##  $ player: Factor w/ 5 levels "A","B","C","D",..: 1 2 3 4 5
##  $ avg   : Factor w/ 5 levels "0.257","0.266",..: 3 1 5 2 4
##  $ age   : Factor w/ 5 levels "23","24","26",..: 2 4 3 1 5
```

```
e$age = as.numeric(e$age)
e
```

```
##   player   avg age
## 1      A  0.28   2
## 2      B 0.257   4
## 3      C 0.312   3
## 4      D 0.266   1
## 5      E 0.295   5
```

What's Wrong?

```
e$age = c(26, 23, 31, 27, 24) # compensate previous wrong
e$age = as.character(e$age)
g = ifelse(e$age>25, 1, 0)
g
```

```
## [1] 1 0 1 1 0
```

```
#e$age = c(26, 23, 31, 27, 24)
h = cbind(e, g)
h
```

```
##   player   avg age g
## 1      A  0.28  26 1
## 2      B 0.257  23 0
## 3      C 0.312  31 1
## 4      D 0.266  27 1
## 5      E 0.295  24 0
```

Conditional Expressions

```
library(Lahman)
aa = subset(Batting, playerID=='altuvjo01' | playerID=='zobribe01')
aa
```

```
##         playerID yearID stint teamID lgID   G  AB   R   H X2B X3B HR RBI
## 88691   zobribe01   2006     1    TBA   AL  52 183  10  41   6   2  2  18
## 90076   zobribe01   2007     1    TBA   AL  31  97   8  15   2   0  1   9
## 91461   zobribe01   2008     1    TBA   AL  62 198  32  50  10   2 12  30
## 92849   zobribe01   2009     1    TBA   AL 152 501  91 149  28   7 27  91
## 94205   zobribe01   2010     1    TBA   AL 151 541  77 129  28   2 10  75
## 94226   altuvjo01   2011     1    HOU   NL  57 221  26  61  10   1  2  12
## 95595   zobribe01   2011     1    TBA   AL 156 588  99 158  46   6 20  91
## 95618   altuvjo01   2012     1    HOU   NL 147 576  80 167  34   4  7  37
## 97003   zobribe01   2012     1    TBA   AL 157 560  88 151  39   7 20  74
## 97022   altuvjo01   2013     1    HOU   AL 152 626  64 177  31   2  5  52
## 98411   zobribe01   2013     1    TBA   AL 157 612  77 168  36   3 12  71
## 98440   altuvjo01   2014     1    HOU   AL 158 660  85 225  47   3  7  59
## 99846   zobribe01   2014     1    TBA   AL 146 570  83 155  34   3 10  52
## 99874   altuvjo01   2015     1    HOU   AL 154 638  86 200  40   4 15  66
## 101330 zobribe01   2015     1    OAK   AL  67 235  39  63  20   2  6  33
## 101331 zobribe01   2015     2    KCA   AL  59 232  37  66  16   1  7  23
## 101360 altuvjo01   2016     1    HOU   AL 161 640 108 216  42   5 24  96
## 102814 zobribe01   2016     1    CHN   NL 147 523  94 142  31   3 18  76
##        SB CS BB  SO IBB HBP SH SF GIDP
## 88691   2  3 10  26   1   0  2  3    2
## 90076   2  0  3  21   0   1  2  2    1
## 91461   3  0 25  37   1   2  0  2    4
## 92849  17  6 91 104   4   2  1  4    7
## 94205  24  3 92 107   1   3  7 12   10
## 94226   7  3  5  29   0   2  5  1    5
## 95595  19  6 77 128   1   2  2  5    9
## 95618  33 11 40  74   0   6  4  4    8
## 97003  14  9 97 103   7   3  2  6   13
## 97022  35 13 32  85   5   2  4  8   24
## 98411  11  3 72  91   4   7  1  6   18
## 98440  56  9 36  53   7   5  1  5   20
## 99846  10  5 75  84   4   1  2  6    8
## 99874  38 13 33  67   8   9  3  6   17
## 101330  1  1 33  26   2   0  0  3    5
## 101331  2  3 29  30   1   1  0  2    3
```

```
## 101360 30 10 60  70  11   7  3  7   15
## 102814  6  4 96  82   6   4  4  4   17
```

```
bb = subset(aa, yearID > 2011 & yearID < 2017)
bb
```

```
##         playerID yearID stint teamID lgID   G  AB   R   H X2B X3B HR RBI
## 95618  altuvjo01   2012     1    HOU   NL 147 576  80 167  34   4  7  37
## 97003  zobribe01   2012     1    TBA   AL 157 560  88 151  39   7 20  74
## 97022  altuvjo01   2013     1    HOU   AL 152 626  64 177  31   2  5  52
## 98411  zobribe01   2013     1    TBA   AL 157 612  77 168  36   3 12  71
## 98440  altuvjo01   2014     1    HOU   AL 158 660  85 225  47   3  7  59
## 99846  zobribe01   2014     1    TBA   AL 146 570  83 155  34   3 10  52
## 99874  altuvjo01   2015     1    HOU   AL 154 638  86 200  40   4 15  66
## 101330 zobribe01   2015     1    OAK   AL  67 235  39  63  20   2  6  33
## 101331 zobribe01   2015     2    KCA   AL  59 232  37  66  16   1  7  23
## 101360 altuvjo01   2016     1    HOU   AL 161 640 108 216  42   5 24  96
## 102814 zobribe01   2016     1    CHN   NL 147 523  94 142  31   3 18  76
##        SB CS BB  SO IBB HBP SH SF GIDP
## 95618  33 11 40  74   0   6  4  4    8
## 97003  14  9 97 103   7   3  2  6   13
## 97022  35 13 32  85   5   2  4  8   24
## 98411  11  3 72  91   4   7  1  6   18
## 98440  56  9 36  53   7   5  1  5   20
## 99846  10  5 75  84   4   1  2  6    8
## 99874  38 13 33  67   8   9  3  6   17
## 101330  1  1 33  26   2   0  0  3    5
## 101331  2  3 29  30   1   1  0  2    3
## 101360 30 10 60  70  11   7  3  7   15
## 102814  6  4 96  82   6   4  4  4   17
```

```
cc = subset(bb, !(yearID == 2014 | yearID == 2015))
cc
```

```
##         playerID yearID stint teamID lgID   G  AB   R   H X2B X3B HR RBI
## 95618  altuvjo01   2012     1    HOU   NL 147 576  80 167  34   4  7  37
## 97003  zobribe01   2012     1    TBA   AL 157 560  88 151  39   7 20  74
## 97022  altuvjo01   2013     1    HOU   AL 152 626  64 177  31   2  5  52
## 98411  zobribe01   2013     1    TBA   AL 157 612  77 168  36   3 12  71
## 101360 altuvjo01   2016     1    HOU   AL 161 640 108 216  42   5 24  96
## 102814 zobribe01   2016     1    CHN   NL 147 523  94 142  31   3 18  76
##        SB CS BB  SO IBB HBP SH SF GIDP
## 95618  33 11 40  74   0   6  4  4    8
## 97003  14  9 97 103   7   3  2  6   13
## 97022  35 13 32  85   5   2  4  8   24
## 98411  11  3 72  91   4   7  1  6   18
## 101360 30 10 60  70  11   7  3  7   15
## 102814  6  4 96  82   6   4  4  4   17
```

```
dd = subset(cc, select=c('playerID','HR','X3B'))
dd
```

```
##         playerID HR X3B
## 95618  altuvjo01  7   4
## 97003  zobribe01 20   7
## 97022  altuvjo01  5   2
## 98411  zobribe01 12   3
```

```
## 101360 altuvjo01 24   5
## 102814 zobribe01 18   3
```