Mutate & Transmute Functions

Pema Lama

April 5, 2017

We will use mutate() and transmute() functions of dplyr package.

We can the set current working directory with setwd() function.

We load dplyr package with library() function.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

We load titanic data set with read.csv() function.

```
college <- read.csv('College.csv', stringsAsFactors = TRUE, header = TRUE)
```

To see the structure of the data set, we use str() function.

```
str(college)
```

```
## 'data.frame':    777 obs. of  19 variables:
##  $ X          : Factor w/ 777 levels "Abilene Christian University",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ Private    : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Apps       : int  1660 2186 1428 417 193 587 353 1899 1038 582 ...
##  $ Accept     : int  1232 1924 1097 349 146 479 340 1720 839 498 ...
##  $ Enroll     : int  721 512 336 137 55 158 103 489 227 172 ...
##  $ Top10perc  : int  23 16 22 60 16 38 17 37 30 21 ...
##  $ Top25perc  : int  52 29 50 89 44 62 45 68 63 44 ...
##  $ F.Undergrad: int  2885 2683 1036 510 249 678 416 1594 973 799 ...
##  $ P.Undergrad: int  537 1227 99 63 869 41 230 32 306 78 ...
##  $ Outstate   : int  7440 12280 11250 12960 7560 13500 13290 13868 15595 10468 ...
##  $ Room.Board : int  3300 6450 3750 5450 4120 3335 5720 4826 4400 3380 ...
##  $ Books      : int  450 750 400 450 800 500 500 450 300 660 ...
##  $ Personal   : int  2200 1500 1165 875 1500 675 1500 850 500 1800 ...
##  $ PhD        : int  70 29 53 92 76 67 90 89 79 40 ...
##  $ Terminal   : int  78 30 66 97 72 73 93 100 84 41 ...
##  $ S.F.Ratio  : num  18.1 12.2 12.9 7.7 11.9 9.4 11.5 13.7 11.3 11.5 ...
##  $ perc.alumni: int  12 16 30 37 2 11 26 37 23 15 ...
##  $ Expend     : int  7041 10527 8735 19016 10922 9727 8861 11487 11644 8991 ...
##  $ Grad.Rate  : int  60 56 54 59 15 55 63 73 80 52 ...
```

We create new variables from the existing ones with mutate() function.

```
college <- mutate(college, newEnroll = Enroll/10, notAccepted = Apps - Accept)
```

To see the column names we use names() function.

```
names(college)
```

```
##  [1] "X"           "Private"     "Apps"        "Accept"      "Enroll"
##  [6] "Top10perc"   "Top25perc"   "F.Undergrad" "P.Undergrad" "Outstate"
## [11] "Room.Board"  "Books"       "Personal"    "PhD"         "Terminal"
## [16] "S.F.Ratio"   "perc.alumni" "Expend"      "Grad.Rate"   "newEnroll"
## [21] "notAccepted"
```

To see only the new variables we use transmute() function.

```
new_college <- transmute(college, newEnroll = Enroll/10, notAccepted = Apps - Accept)
```

To see the names of the new variables we use names() function.

```
names(new_college)
```

```
## [1] "newEnroll"   "notAccepted"
```

To see the summary statistics of PhD students we use summary() function.

```
summary(college$PhD)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    8.00   62.00   75.00   72.66   85.00  103.00
```

With following R-command, we create new variable using if_else() function with mutate() function.

```
college <- mutate(college, newEnroll = if_else(PhD <= 75, 'belowAvg', 'aboveAvg'),
                  notAccepted = Apps - Accept)
```

To view the first 20 rows of a data set.

```
head(college, 20)
```

```
##                                        X Private Apps Accept Enroll
## 1             Abilene Christian University     Yes 1660   1232    721
## 2                       Adelphi University     Yes 2186   1924    512
## 3                           Adrian College     Yes 1428   1097    336
## 4                     Agnes Scott College     Yes  417    349    137
## 5                  Alaska Pacific University     Yes  193    146     55
## 6                         Albertson College     Yes  587    479    158
## 7                   Albertus Magnus College     Yes  353    340    103
## 8                           Albion College     Yes 1899   1720    489
## 9                         Albright College     Yes 1038    839    227
## 10                Alderson-Broaddus College     Yes  582    498    172
## 11                        Alfred University     Yes 1732   1425    472
## 12                       Allegheny College     Yes 2652   1900    484
## 13 Allentown Coll. of St. Francis de Sales     Yes 1179    780    290
## 14                             Alma College     Yes 1267   1080    385
## 15                          Alverno College     Yes  494    313    157
## 16          American International College     Yes 1420   1093    220
## 17                          Amherst College     Yes 4302    992    418
## 18                       Anderson University     Yes 1216    908    423
## 19                        Andrews University     Yes 1130    704    322
## 20                   Angelo State University      No 3540   2001   1016
##    Top10perc Top25perc F.Undergrad P.Undergrad Outstate Room.Board Books
## 1         23        52        2885         537     7440       3300   450
## 2         16        29        2683        1227    12280       6450   750
## 3         22        50        1036          99    11250       3750   400
```

```
## 4         60        89         510          63    12960       5450   450
## 5         16        44         249         869     7560       4120   800
## 6         38        62         678          41    13500       3335   500
## 7         17        45         416         230    13290       5720   500
## 8         37        68        1594          32    13868       4826   450
## 9         30        63         973         306    15595       4400   300
## 10        21        44         799          78    10468       3380   660
## 11        37        75        1830         110    16548       5406   500
## 12        44        77        1707          44    17080       4440   400
## 13        38        64        1130         638     9690       4785   600
## 14        44        73        1306          28    12572       4552   400
## 15        23        46        1317        1235     8352       3640   650
## 16         9        22        1018         287     8700       4780   450
## 17        83        96        1593           5    19760       5300   660
## 18        19        40        1819         281    10100       3520   550
## 19        14        23        1586         326     9996       3090   900
## 20        24        54        4190        1512     5130       3592   500
##    Personal PhD Terminal S.F.Ratio perc.alumni Expend Grad.Rate newEnroll
## 1      2200  70       78      18.1          12   7041        60  belowAvg
## 2      1500  29       30      12.2          16  10527        56  belowAvg
## 3      1165  53       66      12.9          30   8735        54  belowAvg
## 4       875  92       97       7.7          37  19016        59  aboveAvg
## 5      1500  76       72      11.9           2  10922        15  aboveAvg
## 6       675  67       73       9.4          11   9727        55  belowAvg
## 7      1500  90       93      11.5          26   8861        63  aboveAvg
## 8       850  89      100      13.7          37  11487        73  aboveAvg
## 9       500  79       84      11.3          23  11644        80  aboveAvg
## 10     1800  40       41      11.5          15   8991        52  belowAvg
## 11      600  82       88      11.3          31  10932        73  aboveAvg
## 12      600  73       91       9.9          41  11711        76  belowAvg
## 13     1000  60       84      13.3          21   7940        74  belowAvg
## 14      400  79       87      15.3          32   9305        68  aboveAvg
## 15     2449  36       69      11.1          26   8127        55  belowAvg
## 16     1400  78       84      14.7          19   7355        69  aboveAvg
## 17     1598  93       98       8.4          63  21424       100  aboveAvg
## 18     1100  48       61      12.1          14   7994        59  belowAvg
## 19     1320  62       66      11.5          18  10908        46  belowAvg
## 20     2000  60       62      23.1           5   4010        34  belowAvg
##    notAccepted
## 1          428
## 2          262
## 3          331
## 4           68
## 5           47
## 6          108
## 7           13
## 8          179
## 9          199
## 10          84
## 11         307
## 12         752
## 13         399
## 14         187
## 15         181
```

```
## 16          327
## 17         3310
## 18          308
## 19          426
## 20         1539
```