distinct() and arrange() functions

Pema Lama

April 5, 2017

In this session we will learn distinct() and arrange() functions.

How do I set the current working directory using setwd() function?

```
setwd("C:/Users/Owner/Desktop/dplyr")
```

How do I load R-packages using 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
```

How do I read data set using read.csv() function?

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

How do I see the structure of a data set using 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 ...
```

How do I see the unique values of a variable with distinct() function?

```
private <- distinct(college, Private)
private
```

```
##   Private
## 1     Yes
## 2      No
```

How do I select unique values of multiple variables using distinct() function?

```
phd <- distinct(college, Private, PhD)
phd
```

```
##     Private PhD
## 1       Yes  70
## 2       Yes  29
## 3       Yes  53
## 4       Yes  92
## 5       Yes  76
## 6       Yes  67
## 7       Yes  90
## 8       Yes  89
## 9       Yes  79
## 10      Yes  40
## 11      Yes  82
## 12      Yes  73
## 13      Yes  60
## 14      Yes  36
## 15      Yes  78
## 16      Yes  93
## 17      Yes  48
## 18      Yes  62
## 19       No  60
## 20      Yes  69
## 21       No  83
## 22      Yes  55
## 23       No  88
## 24       No  57
## 25       No  85
## 26      Yes  65
## 27      Yes  66
## 28      Yes  81
## 29      Yes  59
## 30      Yes  58
## 31      Yes  68
## 32      Yes  57
## 33      Yes  98
## 34      Yes  83
## 35      Yes  71
## 36      Yes  74
## 37      Yes  61
## 38      Yes  87
## 39      Yes  64
## 40      Yes  35
## 41      Yes  80
## 42      Yes  63
```

```
## 43       No  66
## 44       No  81
## 45      Yes  75
## 46      Yes  39
## 47      Yes  99
## 48      Yes 100
## 49      Yes  95
## 50      Yes  77
## 51      Yes  72
## 52       No  72
## 53       No  90
## 54      Yes  10
## 55      Yes  86
## 56       No  89
## 57      Yes  22
## 58      Yes  50
## 59      Yes  41
## 60      Yes   8
## 61       No  69
## 62       No  67
## 63       No  80
## 64      Yes  94
## 65       No  82
## 66      Yes  56
## 67       No  73
## 68      Yes  46
## 69      Yes  54
## 70      Yes  84
## 71      Yes  85
## 72       No  87
## 73      Yes  97
## 74      Yes  51
## 75      Yes  42
## 76       No  49
## 77      Yes  88
## 78       No  50
## 79      Yes  52
## 80      Yes  43
## 81       No  74
## 82       No  71
## 83       No  62
## 84      Yes  37
## 85       No  75
## 86      Yes  45
## 87       No  76
## 88      Yes  47
## 89       No  93
## 90      Yes  91
## 91       No  92
## 92      Yes  25
## 93       No  64
## 94      Yes  31
## 95       No  77
## 96      Yes  49
```

```
## 97       No  63
## 98      Yes  96
## 99      Yes  34
## 100      No  48
## 101      No  47
## 102      No  68
## 103     Yes  33
## 104      No  52
## 105     Yes  44
## 106     Yes  32
## 107      No  99
## 108      No  65
## 109      No  79
## 110      No  78
## 111      No  84
## 112      No  55
## 113      No  86
## 114      No  70
## 115     Yes  14
## 116      No  58
## 117      No  91
## 118      No 103
## 119      No  95
## 120      No  96
## 121     Yes  26
## 122      No  46
## 123      No  97
## 124      No  56
## 125      No  94
## 126     Yes  16
## 127      No  39
## 128      No  61
## 129      No  33
## 130      No  53
```

How do I sort data set by a variable with arrange() function?

```
sorted <- arrange(college, Accept)
head(sorted)
```

```
##                                X Private Apps Accept Enroll Top10perc
## 1              Christendom College     Yes   81     72     51        33
## 2                  Capitol College     Yes  100     90     35        10
## 3            College of St. Joseph     Yes  141    118     55        12
## 4       Wisconsin Lutheran College     Yes  152    128     75        17
## 5 Saint Mary-of-the-Woods College     Yes  150    130     88        23
## 6                   Wilson College     Yes  167    130     46        16
##   Top25perc F.Undergrad P.Undergrad Outstate Room.Board Books Personal PhD
## 1        71         139           3     8730       3600   400      800  92
## 2        52         282         331     8400       2812   300     2134  10
## 3        21         201         173     8300       4850   450     1300  53
## 4        41         282          22     9100       3700   500     1400  48
## 5        50         341         768    10300       4130   500     1700  44
## 6        50         199         676    11428       5084   450      475  67
##   Terminal S.F.Ratio perc.alumni Expend Grad.Rate
```

```
## 1       92       9.3          17  10922        58
## 2       50      12.1          24   7976        52
## 3       53       9.5          19   6936        76
## 4       48       8.5          26   8960        50
## 5       58      10.2          37   9678        75
## 6       76       8.3          43  10291        67
```

How do I sort data set by multiple variables with arrange() function?

```
sorted <- arrange(college, Apps, Accept)
head(sorted)
```

```
##                                X Private Apps Accept Enroll Top10perc
## 1             Christendom College     Yes   81     72     51        33
## 2                 Capitol College     Yes  100     90     35        10
## 3           College of St. Joseph     Yes  141    118     55        12
## 4 Saint Mary-of-the-Woods College     Yes  150    130     88        23
## 5      Wisconsin Lutheran College     Yes  152    128     75        17
## 6                  Wilson College     Yes  167    130     46        16
##   Top25perc F.Undergrad P.Undergrad Outstate Room.Board Books Personal PhD
## 1        71         139           3     8730       3600   400      800  92
## 2        52         282         331     8400       2812   300     2134  10
## 3        21         201         173     8300       4850   450     1300  53
## 4        50         341         768    10300       4130   500     1700  44
## 5        41         282          22     9100       3700   500     1400  48
## 6        50         199         676    11428       5084   450      475  67
##   Terminal S.F.Ratio perc.alumni Expend Grad.Rate
## 1       92       9.3          17  10922        58
## 2       50      12.1          24   7976        52
## 3       53       9.5          19   6936        76
## 4       58      10.2          37   9678        75
## 5       48       8.5          26   8960        50
## 6       76       8.3          43  10291        67
```

How do I sort data set in descending order with desc() function?

```
sorted <- arrange(college, desc(Accept))
head(sorted)
```

```
##                                   X Private  Apps Accept Enroll
## 1            Rutgers at New Brunswick      No 48094  26330   4520
## 2 Purdue University at West Lafayette      No 21804  18744   5874
## 3           Michigan State University      No 18114  15096   6180
## 4   Indiana University at Bloomington      No 16587  13243   5873
## 5                   Boston University     Yes 20192  13007   3810
## 6 University of Michigan at Ann Arbor      No 19152  12940   4893
##   Top10perc Top25perc F.Undergrad P.Undergrad Outstate Room.Board Books
## 1        36        79       21401        3712     7410       4748   690
## 2        29        60       26213        4065     9556       3990   570
## 3        23        57       26640        4120    10658       3734   504
## 4        25        72       24763        2717     9766       3990   600
## 5        45        80       14971        3113    18420       6810   475
## 6        66        92       22045        1339    15732       4659   476
##   Personal PhD Terminal S.F.Ratio perc.alumni Expend Grad.Rate
## 1     2009  90       95      19.5          19  10474        77
## 2     1060  86       86      18.2          15   8604        67
## 3      600  93       95      14.0           9  10520        71
```

```
## 4     2000  77       88      21.3          24   8686        68
## 5     1025  80       81      11.9          16  16836        72
## 6     1600  90       98      11.5          26  14847        87
```