Data Manipulation in R

Data Manipulation in R

Daniel J. Hocking

22 November 2013

R Working Group

R is a great resource and has become the lingua franca for statistics in ecology. R is not the best of languages but it has to big advantages: a large ecology user base and a large, centralized repository of contributed packages (CRAN). R is an incredibly flexible language, possibly to a fault. For example to extract a column from a data frame (we'll call toy) you can do any of the following:

```
a <- c("one", "two", "three", "four")
b <- c(1, 2, 3, 4)
c <- c(1.1, 2.2, 3.3, 4.4)
toy <- data.frame(a, b, c)

str(toy)
```

```
## 'data.frame':    4 obs. of  3 variables:
##  $ a: Factor w/ 4 levels "four","one","three",..: 2 4 3 1
##  $ b: num  1 2 3 4
##  $ c: num  1.1 2.2 3.3 4.4
```

```
summary(toy)
```

```
##      a           b              c       
##  four :1   Min.   :1.00   Min.   :1.10  
##  one  :1   1st Qu.:1.75   1st Qu.:1.93  
##  three:1   Median :2.50   Median :2.75  
##  two  :1   Mean   :2.50   Mean   :2.75  
##            3rd Qu.:3.25   3rd Qu.:3.58  
##            Max.   :4.00   Max.   :4.40  
```

```
toy[, 1]
```

```
## [1] one   two   three four
## Levels: four one three two
```

```
toy$a
```

```
## [1] one   two   three four
## Levels: four one three two
```

```
toy[, c("a")]
```

```
## [1] one   two   three four
## Levels: four one three two
```

```
toy[["a"]]
```

```
## [1] one   two   three four
## Levels: four one three two
```

As you can see, this all produces the same results. In other languages, you can usually only access a portion of the toy in 1 way. While the flexiblity of R can be useful at times, it can cause confusion and creates an extremely steep learning curve. It's difficult to read the code from other people without a much larger vocabularly.

Add a column to a toyframe

```
d <- seq(from = 100, to = 200, length.out = 4)
d
```

```
## [1] 100.0 133.3 166.7 200.0
```

```
(e <- seq(100, 200, length.out = 4))
```

```
## [1] 100.0 133.3 166.7 200.0
```

```
e[3]
```

```
## [1] 166.7
```

```
(e3 <- e[3] + 2e-06)
```

```
## [1] 166.7
```

```
print(e[3], dig = 10)
```

```
## [1] 166.6666667
```

```
print(e3, dig = 10)
```

```
## [1] 166.6666687
```

```
toy$d <- d
toy$new <- e  # name in the toyframe independent of original object name
toy$f <- rep(NA, times = 4)

toy$d
```

```
## [1] 100.0 133.3 166.7 200.0
```

```
toy$d <- c("o", "v", "e", "r")  # overwrites existing column with the same name
toy$d
```

```
## [1] "o" "v" "e" "r"
```

Now let's work with some real data. It will be bigger and messier than our toy dataset:

```
setwd("/Users/Dan/Documents/Teaching/R_intro/")
demo <- read.table("Salamander_Demographics.csv", header = TRUE, sep = ",")

head(demo)
```

```
##   line page    dates month day year time plot  mass svl tl  sex gravid
## 1 1861   60  4/21/09     4  21 2009    N    5 0.427  33 63 <NA>      N
## 2 1115   36   9/9/08     9   9 2008    N <NA> 0.633  37 68 <NA>      N
## 3  360   12  5/31/08     5  31 2008    N    3 0.639  42 63 <NA>      N
## 4 2897   92   5/7/11     5   7 2011    N    7 0.921  43 79 <NA>      N
## 5 1432   46 10/16/08    10  16 2008    N    9 0.943  45 74 <NA>      N
## 6  372   12  5/31/08     5  31 2008    N    3    NA  46 NA <NA>      N
##   group clutch color recap mark   id damage
## 1  <NA>     NA     R  <NA> <NA> 1371      N
## 2  <NA>     NA     R  <NA> <NA>   NA      N
## 3  <NA>     NA     R  <NA> <NA>  187      Y
## 4  <NA>     NA     R     N <NA> 2154      N
## 5  <NA>     NA     L  <NA> <NA> 1042      Y
## 6  <NA>     NA     R  <NA> <NA>  198      N
```

```
tail(demo)
```

```
##      line page    dates month day year time plot  mass svl tl sex gravid
## 3377 1435   46 10/16/08    10  16 2008    N    4 1.174  48 86   Y      N
## 3378 2765   88   5/4/11     5   4 2011    N    7 0.974  49 89   Y      N
## 3379 3248  103   6/9/11     6   9 2011    N    9 1.204  49 87   Y      N
## 3380 1503   49  11/6/08    11   6 2008    N    4 1.365  49 89   Y      N
## 3381 1475   48  11/1/08    11   1 2008    D   T1 1.295  50 93   Y      N
## 3382  494   16   6/4/08     6   4 2008    N    9 0.814  51 69   Y      N
##      group clutch color recap mark   id damage
## 3377     Y     NA     R  <NA> <NA> 1045      N
## 3378     Y     NA     R     N <NA> 2022      N
## 3379     Y     NA     R     N <NA> 2464      Y
## 3380     Y     NA     R  <NA> <NA> 1079      N
## 3381     Y     NA     R  <NA> <NA> 1101      N
## 3382     Y     NA     R  <NA> <NA>  292      N
```

```
summary(demo)
```

```
##       line           page          dates          month
##  Min.   :   1   Min.   :  1.0   4/21/09: 166   Min.   : 4.00
##  1st Qu.: 846   1st Qu.: 27.0   5/31/08: 158   1st Qu.: 5.00
##  Median :1692   Median : 55.0   6/9/11 : 147   Median : 6.00
##  Mean   :1692   Mean   : 54.3   5/29/09: 107   Mean   : 6.31
##  3rd Qu.:2537   3rd Qu.: 82.0   6/4/08 : 106   3rd Qu.: 6.00
##  Max.   :3382   Max.   :107.0   9/9/08 : 104   Max.   :11.00
##                                 (Other):2594
##       day            year      time         plot          mass
##  Min.   : 1.0   Min.   :2008   D: 206   5      :709   Min.   :0.061
##  1st Qu.: 8.0   1st Qu.:2008   N:3176   4      :671   1st Qu.:0.511
##  Median :15.0   Median :2008            3      :616   Median :0.718
##  Mean   :15.4   Mean   :2009            9      :615   Mean   :0.708
##  3rd Qu.:22.0   3rd Qu.:2009            7      :586   3rd Qu.:0.887
##  Max.   :31.0   Max.   :2011            (Other):181   Max.   :1.929
##                                         NA's   :  4   NA's   :2
##       svl             tl           sex        gravid      group
##  Min.   :15.0   Min.   : 20.0   U   : 812   D   : 128   GF  : 241
##  1st Qu.:34.0   1st Qu.: 59.0   UA  :   8   N   :2952   NG  : 775
##  Median :39.0   Median : 69.0   UI  : 226   Y   : 241   U   : 812
##  Mean   :38.1   Mean   : 66.9   X   :1077   NA's:  61   UA  :   8
##  3rd Qu.:43.0   3rd Qu.: 77.0   Y   :1249               UI  : 226
##  Max.   :55.0   Max.   :105.0   NA's:  10               Y   :1249
##  NA's   :3      NA's   :2                               NA's:  71
##      clutch          color       recap          mark            id
##  Min.   : 2.0   BLOTCHY:   3   N   : 600   XXXY   :   2   Min.   :   1
##  1st Qu.: 6.0   L      :  74   Y   :  48   OGGX   :   1   1st Qu.: 594
##  Median : 7.0   R      :3283   NA's:2734   OOOX   :   1   Median :1397
##  Mean   : 7.5   TAN    :  17               OORG   :   1   Mean   :1329
##  3rd Qu.: 9.0   NA's   :   5               ORGO   :   1   3rd Qu.:2012
##  Max.   :13.0                              (Other):  33   Max.   :2598
##  NA's   :3117                              NA's   :3343   NA's   :1003
##  damage
##  N:2106
##  Y:1276
##
##
##
##
##
```

```
str(demo)
```

```
## 'data.frame':    3382 obs. of  20 variables:
##  $ line  : int  1861 1115 360 2897 1432 372 231 2739 2236 543 ...
##  $ page  : int  60 36 12 92 46 12 8 87 72 17 ...
##  $ dates : Factor w/ 81 levels "10/1/08","10/16/08",..: 12 81 32 36 2 32 28 3 15 59 ...
##  $ month : int  4 9 5 5 10 5 5 10 5 6 ...
##  $ day   : int  21 9 31 7 16 31 27 24 14 5 ...
##  $ year  : int  2009 2008 2008 2011 2008 2008 2008 2009 2009 2008 ...
##  $ time  : Factor w/ 2 levels "D","N": 2 2 2 2 2 2 2 2 2 2 ...
##  $ plot  : Factor w/ 12 levels "1","3","4","5",..: 4 NA 2 5 7 2 7 9 4 5 ...
##  $ mass  : num  0.427 0.633 0.639 0.921 0.943 ...
##  $ svl   : int  33 37 42 43 45 46 47 48 NA NA ...
##  $ tl    : int  63 68 63 79 74 NA 75 89 87 NA ...
##  $ sex   : Factor w/ 5 levels "U","UA","UI",..: NA NA NA NA NA NA NA NA NA NA ...
##  $ gravid: Factor w/ 3 levels "D","N","Y": 2 2 2 2 2 2 2 2 2 2 ...
##  $ group : Factor w/ 6 levels "GF","NG","U",..: NA NA NA NA NA NA NA NA NA NA ...
##  $ clutch: int  NA NA NA NA NA NA NA NA NA NA ...
##  $ color : Factor w/ 4 levels "BLOTCHY","L",..: 3 3 3 3 2 3 3 3 3 3 ...
##  $ recap : Factor w/ 2 levels "N","Y": NA NA NA 1 NA NA NA NA NA NA ...
##  $ mark  : Factor w/ 38 levels "OGGX","OOOX",..: NA NA NA NA NA NA NA NA NA NA ...
##  $ id    : int  1371 NA 187 2154 1042 198 74 2036 1564 351 ...
##  $ damage: Factor w/ 2 levels "N","Y": 1 1 2 1 2 1 1 1 2 1 ...
```

Let's create a dataframe with just the size data

```
size.vars <- demo[c("svl", "tl", "mass")]
head(size.vars)
```

```
##   svl tl  mass
## 1  33 63 0.427
## 2  37 68 0.633
## 3  42 63 0.639
## 4  43 79 0.921
## 5  45 74 0.943
## 6  46 NA    NA
```

or maybe we just want the first 5 rows

```
demo5 <- demo[, c(1:5)]
demo5b <- demo[c(1:5)]

head(demo5)
```

```
##   line page    dates month day
## 1 1861   60  4/21/09     4  21
## 2 1115   36   9/9/08     9   9
## 3  360   12  5/31/08     5  31
## 4 2897   92   5/7/11     5   7
## 5 1432   46 10/16/08    10  16
## 6  372   12  5/31/08     5  31
```

```
head(demo5b)
```

```
##   line page    dates month day
## 1 1861   60  4/21/09     4  21
## 2 1115   36   9/9/08     9   9
## 3  360   12  5/31/08     5  31
## 4 2897   92   5/7/11     5   7
## 5 1432   46 10/16/08    10  16
## 6  372   12  5/31/08     5  31
```

Delete some variables

```
rm.vars <- names(demo) %in% c("id", "mark", "recap")
newdemo <- demo[!rm.vars]
head(newdemo)
```

```
##   line page    dates month day year time plot  mass svl tl  sex gravid
## 1 1861   60  4/21/09     4  21 2009    N    5 0.427  33 63 <NA>      N
## 2 1115   36   9/9/08     9   9 2008    N <NA> 0.633  37 68 <NA>      N
## 3  360   12  5/31/08     5  31 2008    N    3 0.639  42 63 <NA>      N
## 4 2897   92   5/7/11     5   7 2011    N    7 0.921  43 79 <NA>      N
## 5 1432   46 10/16/08    10  16 2008    N    9 0.943  45 74 <NA>      N
## 6  372   12  5/31/08     5  31 2008    N    3    NA  46 NA <NA>      N
##   group clutch color damage
## 1  <NA>     NA     R      N
## 2  <NA>     NA     R      N
## 3  <NA>     NA     R      Y
## 4  <NA>     NA     R      N
## 5  <NA>     NA     L      Y
## 6  <NA>     NA     R      N
```

```
newdemo2 <- demo[c(-1, -3)]
head(newdemo2)
```

```
##   page month day year time plot  mass svl tl  sex gravid group clutch
## 1   60     4  21 2009    N    5 0.427  33 63 <NA>      N  <NA>     NA
## 2   36     9   9 2008    N <NA> 0.633  37 68 <NA>      N  <NA>     NA
## 3   12     5  31 2008    N    3 0.639  42 63 <NA>      N  <NA>     NA
## 4   92     5   7 2011    N    7 0.921  43 79 <NA>      N  <NA>     NA
## 5   46    10  16 2008    N    9 0.943  45 74 <NA>      N  <NA>     NA
## 6   12     5  31 2008    N    3    NA  46 NA <NA>      N  <NA>     NA
##   color recap mark   id damage
## 1     R  <NA> <NA> 1371      N
## 2     R  <NA> <NA>   NA      N
## 3     R  <NA> <NA>  187      Y
## 4     R     N <NA> 2154      N
## 5     L  <NA> <NA> 1042      Y
## 6     R  <NA> <NA>  198      N
```

```
newdemo2$id <- newdemo2$mark <- NULL
head(newdemo2)
```

```
##   page month day year time plot  mass svl tl  sex gravid group clutch
## 1   60     4  21 2009    N    5 0.427  33 63 <NA>      N  <NA>     NA
## 2   36     9   9 2008    N <NA> 0.633  37 68 <NA>      N  <NA>     NA
## 3   12     5  31 2008    N    3 0.639  42 63 <NA>      N  <NA>     NA
## 4   92     5   7 2011    N    7 0.921  43 79 <NA>      N  <NA>     NA
## 5   46    10  16 2008    N    9 0.943  45 74 <NA>      N  <NA>     NA
## 6   12     5  31 2008    N    3    NA  46 NA <NA>      N  <NA>     NA
##   color recap damage
## 1     R  <NA>      N
## 2     R  <NA>      N
## 3     R  <NA>      Y
## 4     R     N      N
## 5     L  <NA>      Y
## 6     R  <NA>      N
```

Select Observations

```
# based on variable values
newdemo <- demo[which(demo$sex == "Y" & demo$mass > 1), ]
str(demo)
```

```
## 'data.frame':    3382 obs. of  20 variables:
##  $ line  : int  1861 1115 360 2897 1432 372 231 2739 2236 543 ...
##  $ page  : int  60 36 12 92 46 12 8 87 72 17 ...
##  $ dates : Factor w/ 81 levels "10/1/08","10/16/08",..: 12 81 32 36 2 32 28 3 15 59 ...
##  $ month : int  4 9 5 5 10 5 5 10 5 6 ...
##  $ day   : int  21 9 31 7 16 31 27 24 14 5 ...
##  $ year  : int  2009 2008 2008 2011 2008 2008 2008 2009 2009 2008 ...
##  $ time  : Factor w/ 2 levels "D","N": 2 2 2 2 2 2 2 2 2 2 ...
##  $ plot  : Factor w/ 12 levels "1","3","4","5",..: 4 NA 2 5 7 2 7 9 4 5 ...
##  $ mass  : num  0.427 0.633 0.639 0.921 0.943 ...
##  $ svl   : int  33 37 42 43 45 46 47 48 NA NA ...
##  $ tl    : int  63 68 63 79 74 NA 75 89 87 NA ...
##  $ sex   : Factor w/ 5 levels "U","UA","UI",..: NA NA NA NA NA NA NA NA NA NA ...
##  $ gravid: Factor w/ 3 levels "D","N","Y": 2 2 2 2 2 2 2 2 2 2 ...
##  $ group : Factor w/ 6 levels "GF","NG","U",..: NA NA NA NA NA NA NA NA NA NA ...
##  $ clutch: int  NA NA NA NA NA NA NA NA NA NA ...
##  $ color : Factor w/ 4 levels "BLOTCHY","L",..: 3 3 3 3 2 3 3 3 3 3 ...
##  $ recap : Factor w/ 2 levels "N","Y": NA NA NA 1 NA NA NA NA NA NA ...
##  $ mark  : Factor w/ 38 levels "OGGX","OOOX",..: NA NA NA NA NA NA NA NA NA NA ...
##  $ id    : int  1371 NA 187 2154 1042 198 74 2036 1564 351 ...
##  $ damage: Factor w/ 2 levels "N","Y": 1 1 2 1 2 1 1 1 2 1 ...
```

```
str(newdemo)
```

```
## 'data.frame':    119 obs. of  20 variables:
##  $ line  : int  872 3038 628 328 468 232 903 1268 895 2982 ...
##  $ page  : int  27 96 20 11 15 8 28 41 28 95 ...
##  $ dates : Factor w/ 81 levels "10/1/08","10/16/08",..: 51 22 60 32 58 28 52 78 52 19 ...
##  $ month : int  6 5 6 5 6 5 6 9 6 5 ...
##  $ day   : int  20 20 6 31 4 27 22 27 22 17 ...
##  $ year  : int  2008 2011 2008 2008 2008 2008 2008 2008 2008 2011 ...
##  $ time  : Factor w/ 2 levels "D","N": 2 2 1 2 2 2 2 2 2 2 ...
##  $ plot  : Factor w/ 12 levels "1","3","4","5",..: 5 5 3 5 5 5 3 4 7 7 ...
##  $ mass  : num  1.01 1.02 1.05 1.11 1 ...
##  $ svl   : int  41 41 41 41 42 42 42 42 42 42 ...
##  $ tl    : int  58 64 71 82 83 75 80 80 78 72 ...
##  $ sex   : Factor w/ 5 levels "U","UA","UI",..: 5 5 5 5 5 5 5 5 5 5 ...
##  $ gravid: Factor w/ 3 levels "D","N","Y": 2 2 2 2 2 2 2 2 2 2 ...
##  $ group : Factor w/ 6 levels "GF","NG","U",..: 6 6 6 6 6 6 6 6 6 6 ...
##  $ clutch: int  NA NA NA NA NA NA NA NA NA NA ...
##  $ color : Factor w/ 4 levels "BLOTCHY","L",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ recap : Factor w/ 2 levels "N","Y": NA 2 NA NA NA NA NA NA NA 1 ...
##  $ mark  : Factor w/ 38 levels "OGGX","OOOX",..: NA 6 NA NA NA NA NA NA NA NA ...
##  $ id    : int  662 2314 423 164 267 75 693 NA 685 2239 ...
##  $ damage: Factor w/ 2 levels "N","Y": 2 2 1 2 1 1 1 1 1 2 ...
```

Subset Function

```
newdemo <- subset(demo, sex == "X" & gravid == "Y", select = mass:tl)
str(newdemo)
```

```
## 'data.frame':    241 obs. of  3 variables:
##  $ mass: num  0.716 0.771 0.806 0.843 0.962 0.608 0.708 0.76 0.793 0.808 ...
##  $ svl : int  36 36 37 37 37 38 38 38 38 38 ...
##  $ tl  : int  62 67 67 73 80 53 60 62 66 72 ...
```

```
head(newdemo)
```

```
##       mass svl tl
## 1893 0.716  36 62
## 1894 0.771  36 67
## 1895 0.806  37 67
## 1896 0.843  37 73
## 1897 0.962  37 80
## 1898 0.608  38 53
```

Make table and export as csv file

```
write.table(x = newdemo, file = "Gravid_Female_Demographics.csv", sep = ",")
```

Sorting Data There is no undo key in R. If you write over or delete an object or column, it’s gone. Similarly, if you sort you can’t unsort. I like to have a primary key (line number) so can always return to orginal order.

```
demo$key <- seq(1, length(demo$svl))
head(demo, n = 10)
```

```
##    line page    dates month day year time plot  mass svl tl  sex gravid
## 1  1861   60  4/21/09     4  21 2009    N    5 0.427  33 63 <NA>      N
## 2  1115   36   9/9/08     9   9 2008    N <NA> 0.633  37 68 <NA>      N
## 3   360   12  5/31/08     5  31 2008    N    3 0.639  42 63 <NA>      N
## 4  2897   92   5/7/11     5   7 2011    N    7 0.921  43 79 <NA>      N
## 5  1432   46 10/16/08    10  16 2008    N    9 0.943  45 74 <NA>      N
## 6   372   12  5/31/08     5  31 2008    N    3    NA  46 NA <NA>      N
## 7   231    8  5/27/08     5  27 2008    N    9 1.073  47 75 <NA>      N
## 8  2739   87 10/24/09    10  24 2009    N    T 1.107  48 89 <NA>      N
## 9  2236   72  5/14/09     5  14 2009    N    5 0.626  NA 87 <NA>      N
## 10  543   17   6/5/08     6   5 2008    N    7 1.058  NA NA <NA>      N
##    group clutch color recap mark   id damage key
## 1   <NA>     NA     R  <NA> <NA> 1371      N   1
## 2   <NA>     NA     R  <NA> <NA>   NA      N   2
## 3   <NA>     NA     R  <NA> <NA>  187      Y   3
## 4   <NA>     NA     R     N <NA> 2154      N   4
## 5   <NA>     NA     L  <NA> <NA> 1042      Y   5
## 6   <NA>     NA     R  <NA> <NA>  198      N   6
## 7   <NA>     NA     R  <NA> <NA>   74      N   7
## 8   <NA>     NA     R  <NA> <NA> 2036      N   8
## 9   <NA>     NA     R  <NA> <NA> 1564      Y   9
## 10  <NA>     NA     R  <NA> <NA>  351      N  10
```

```
demo <- demo[order(demo$mass), ]  # if don't want to create extra dataframes
demo.sort <- demo[order(demo$mass), ]  # alt create new sorted dataframe
head(demo, n = 10)
```

```
##     line page    dates month day year time plot  mass svl tl sex gravid
## 834  186    6  5/22/08     5  22 2008    D    3 0.061  16 21  UI      N
## 835 1866   61  4/22/09     4  22 2009    N    5 0.065  16 26  UI      N
## 851  202    7  5/22/08     5  22 2008    D    9 0.074  18 29  UI      N
## 841 1498   49  11/1/08    11   1 2008    D    3 0.076  17 23  UI      N
## 836  221    8  5/27/08     5  27 2008    D    4 0.078  16 26  UI      N
## 837  262    9  5/31/08     5  31 2008    N    5 0.079  16 20  UI      N
## 842  216    7  5/27/08     5  27 2008    D    3 0.079  17 27  UI      N
## 843 1439   46 10/16/08    10  16 2008    N    4 0.081  17 28  UI      N
## 852 2825   90   5/4/11     5   4 2011    N    3 0.082  18 20  UI      N
## 853  175    6  5/22/08     5  22 2008    D    4 0.082  18 28  UI      N
##     group clutch color recap mark   id damage key
## 834    UI     NA     R  <NA> <NA>   31      Y 834
## 835    UI     NA     R  <NA> <NA> 1401      N 835
## 851    UI     NA     R  <NA> <NA>   47      N 851
## 841    UI     NA     L  <NA> <NA> 1128      Y 841
## 836    UI     NA     R  <NA> <NA>   65      N 836
## 837    UI     NA     R  <NA> <NA>  102      N 837
## 842    UI     NA     R  <NA> <NA>   60      N 842
## 843    UI     NA     R  <NA> <NA> 1049      N 843
## 852    UI     NA     R     N <NA> 2082      Y 852
## 853    UI     NA     R  <NA> <NA>   21      N 853
```

```
demo.sort <- demo[order(demo$sex, demo$svl), ]
head(demo.sort, n = 10)
```

```
##    line page   dates month day year time plot  mass svl tl sex gravid
## 27  777   25 6/16/08     6  16 2008    N    7 0.206  29 49   U      N
## 28 2007   65  5/5/09     5   5 2009    N    4 0.215  29 30   U      N
## 29 2118   68  5/7/09     5   7 2009    N    3 0.226  29 37   U      N
## 30 1752   57 4/21/09     4  21 2009    N    3 0.270  29 33   U      N
## 31 1014   32 8/16/08     8  16 2008    N    3 0.277  29 32   U      N
## 32 1840   60 4/21/09     4  21 2009    N    5 0.292  29 46   U      N
## 33 2144   69  5/9/09     5   9 2009    N    7 0.294  29 38   U      N
## 34 2708   87 8/29/09     8  29 2009    N    3 0.295  29 53   U      N
## 35 2693   87 8/22/09     8  22 2009    N    5 0.300  29 51   U      N
## 36 1123   36  9/9/08     9   9 2008    N    4 0.300  29 52   U      N
##    group clutch color recap mark   id damage key
## 27     U     NA     R  <NA> <NA>  568      N  27
## 28     U     NA     R  <NA> <NA> 1453      N  28
## 29     U     NA     R  <NA> <NA>   NA      Y  29
## 30     U     NA     R  <NA> <NA> 1280      Y  30
## 31     U     NA     R  <NA> <NA>   NA      Y  31
## 32     U     NA     R  <NA> <NA> 1384      N  32
## 33     U     NA     R  <NA> <NA>   NA      Y  33
## 34     U     NA     R  <NA> <NA> 2005      N  34
## 35     U     NA     R  <NA> <NA> 1990      N  35
## 36     U     NA     R  <NA> <NA>   NA      N  36
```

On your own:

Add column of random numbers from your favorite distribution (hint ?rnorm) and sort by tl and then that column from largest to smallest.

Further exploration:

built-in (base) functions: by aggregate

Packages to try: plyr reshape2