reshape2:

- melt - takes a wide format data melts into a long format
- cast - takes a long format data cast into a wide format

The need for melt and cast:

- aggregation
- pivot tables
- plotting

```
data(state)

head(state.x77)
```

```
##            Population Income Illiteracy Life Exp Murder
HS Grad Frost
## Alabama          3615   3624        2.1    69.05   15.1
41.3    20
## Alaska            365   6315        1.5    69.31   11.3
66.7   152
## Arizona          2212   4530        1.8    70.55    7.8
58.1    15
## Arkansas         2110   3378        1.9    70.66   10.1
39.9    65
## California      21198   5114        1.1    71.71   10.3
62.6    20
## Colorado         2541   4884        0.7    72.06    6.8
63.9   166
##              Area
## Alabama     50708
## Alaska     566432
## Arizona    113417
## Arkansas    51945
## California 156361
## Colorado   103766
```

```
states <- data.frame(state.x77, state =
row.names(state.x77), region = state.region,
    row.names = 1:nrow(state.x77))
str(states)
```

```
## 'data.frame':    50 obs. of  10 variables:
##  $ Population: num  3615 365 2212 2110 21198 ...
##  $ Income    : num  3624 6315 4530 3378 5114 ...
##  $ Illiteracy: num  2.1 1.5 1.8 1.9 1.1 0.7 1.1 0.9 1.3
2 ...
##  $ Life.Exp  : num  69 69.3 70.5 70.7 71.7 ...
##  $ Murder    : num  15.1 11.3 7.8 10.1 10.3 6.8 3.1 6.2
10.7 13.9 ...
##  $ HS.Grad   : num  41.3 66.7 58.1 39.9 62.6 63.9 56
54.6 52.6 40.6 ...
##  $ Frost     : num  20 152 15 65 20 166 139 103 11 60
...
##  $ Area      : num  50708 566432 113417 51945 156361 ...
##  $ state     : Factor w/ 50 levels
"Alabama","Alaska",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ region    : Factor w/ 4 levels
"Northeast","South",..: 2 4 4 2 4 4 1 2 2 2 ...
```

```
head(states[states$state == "Illinois", ])
```

```
##    Population Income Illiteracy Life.Exp Murder HS.Grad
Frost  Area
## 13      11197   5107        0.9    70.14   10.3    52.6
127 55748
##       state        region
## 13 Illinois North Central
```

```
library(reshape2)
mstates <- melt(states)
```

```
## Using state, region as id variables
```

```
is(mstates)
```

```
## [1] "data.frame" "list"       "oldClass"   "vector"
```

```
mstates[mstates$state == "Illinois", ]
```

```
##        state        region   variable    value
## 13  Illinois North Central Population 11197.00
## 63  Illinois North Central     Income  5107.00
## 113 Illinois North Central Illiteracy     0.90
## 163 Illinois North Central   Life.Exp    70.14
## 213 Illinois North Central     Murder    10.30
## 263 Illinois North Central    HS.Grad    52.60
## 313 Illinois North Central      Frost   127.00
## 363 Illinois North Central       Area 55748.00
```

melt automatically assigns the state and region as the id, because it's a factor (same goes for character)

All measured variables must be of the same type, e.g., numeric, factor, date, as it's stored in a data frame.

```
mstatesByRegion <- melt(states, id.vars = c("region"))
mstatesByRegion[mstatesByRegion$region == "North Central" &
mstatesByRegion$variable ==
    "state", ]
```

```
##            region variable        value
## 413 North Central    state     Illinois
## 414 North Central    state      Indiana
## 415 North Central    state         Iowa
## 416 North Central    state       Kansas
## 422 North Central    state     Michigan
## 423 North Central    state    Minnesota
## 425 North Central    state     Missouri
## 427 North Central    state     Nebraska
## 434 North Central    state North Dakota
## 435 North Central    state         Ohio
## 441 North Central    state South Dakota
## 449 North Central    state    Wisconsin
```

By default, it converts the measured variables into two columns named:

- variable (which identifies which variable is being measured)
- value (which contains the actual values).

```
popDensity <- read.csv("pop_density.csv", skip = 3)[, 1:12]
colnames(popDensity) <- c("state", seq(1910, 2010, 10))
head(popDensity)
```

```
##           state     1910      1920      1930      1940
1950      1960
## 1 United States 92228531 106021568 123202660 132165129
151325798 179323175
## 2       Alabama  2138093   2348174   2646248   2832961
3061743   3266740
## 3        Alaska    64356     55036     59278     72524
128643    226167
## 4       Arizona   204354    334162    435573    499261
749587   1302161
## 5      Arkansas  1574449   1752204   1854482   1949387
1909511   1786272
## 6    California  2377549   3426861   5677251   6907387
10586223  15717204
##        1970      1980      1990      2000      2010
## 1 203211926 226545805 248709873 281421906 308745538
## 2   3444165   3893888   4040587   4447100   4779736
## 3    300382    401851    550043    626932    710231
## 4   1770900   2718215   3665228   5130632   6392017
## 5   1923295   2286435   2350725   2673400   2915918
## 6  19953134  23667902  29760021  33871648  37253956
```

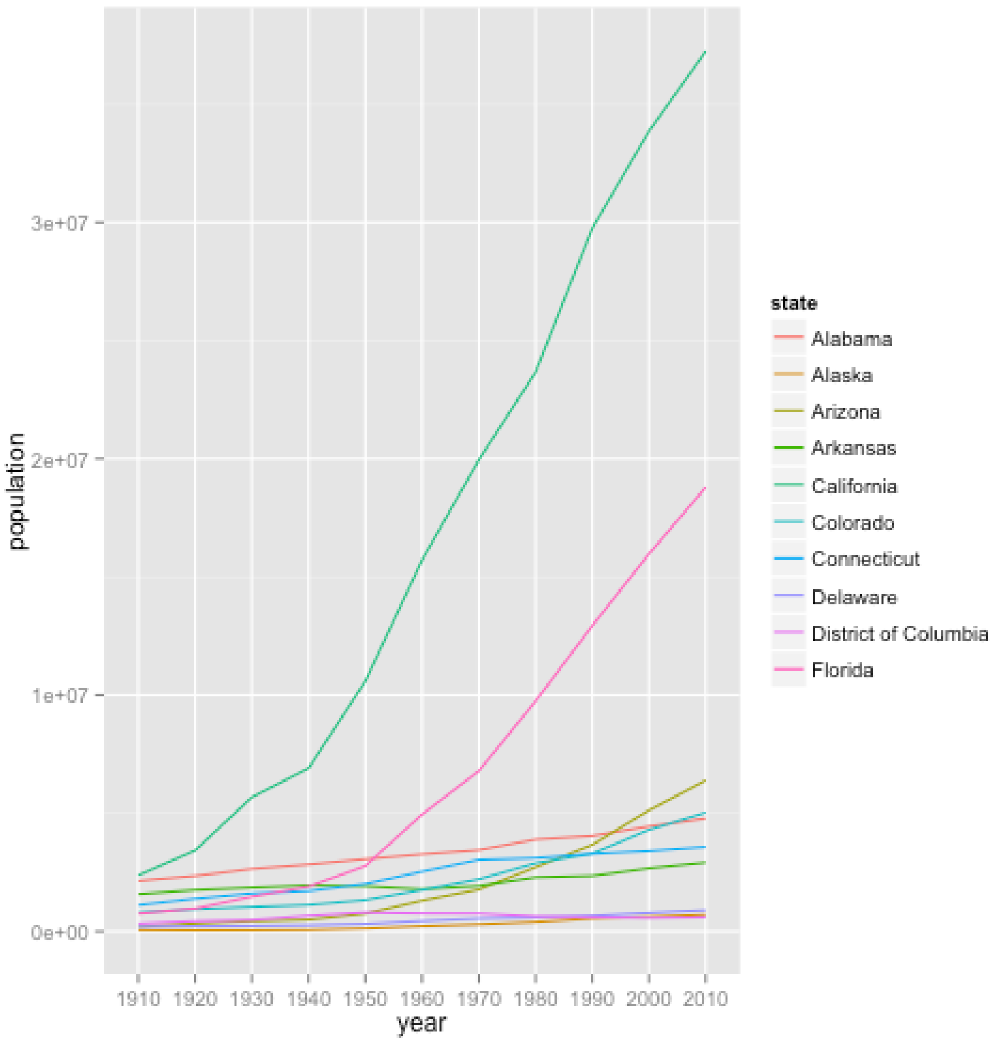
```
subPopDensity <- head(subset(popDensity, state != "United
States"), 10)
head(subPopDensity)
```

```
##        state    1910    1920    1930    1940     1950
1960     1970
## 2    Alabama 2138093 2348174 2646248 2832961  3061743
3266740  3444165
## 3     Alaska   64356   55036   59278   72524   128643
226167   300382
## 4    Arizona  204354  334162  435573  499261   749587
1302161  1770900
## 5   Arkansas 1574449 1752204 1854482 1949387  1909511
1786272  1923295
## 6 California 2377549 3426861 5677251 6907387 10586223
15717204 19953134
## 7   Colorado  799024  939629 1035791 1123296  1325089
1753947  2207259
##       1980     1990     2000     2010
## 2  3893888  4040587  4447100  4779736
## 3   401851   550043   626932   710231
## 4  2718215  3665228  5130632  6392017
## 5  2286435  2350725  2673400  2915918
## 6 23667902 29760021 33871648 37253956
## 7  2889964  3294394  4301261  5029196
```

```
msubPopDensity <- melt(subPopDensity, id.vars = "state",
variable.name = "year",
    value.name = "population")
head(msubPopDensity)
```

```
##         state year population
## 1    Alabama 1910    2138093
## 2     Alaska 1910      64356
## 3    Arizona 1910     204354
## 4   Arkansas 1910    1574449
## 5 California 1910    2377549
## 6   Colorado 1910     799024
```

```
library(ggplot2)
ggplot(msubPopDensity, aes(group = state)) +
geom_line(aes(x = year, y = population,
    color = state))
```



Now on to the casting

There are multiple cast() overrides:

- acast: vector, matrix, array
- dcast: data.frame

```
dcast(mstates, region ~ variable, mean)
```

```
##          region Population Income Illiteracy Life.Exp
Murder HS.Grad
## 1     Northeast       5495   4570      1.000    71.26
4.722   53.97
## 2         South       4208   4012      1.738    69.71
10.581   44.34
## 3 North Central       4803   4611      0.700    71.77
5.275   54.52
## 4          West       2915   4703      1.023    71.23
7.215   62.00
##    Frost   Area
## 1 132.78  18141
## 2  64.62  54605
## 3 138.83  62652
## 4 102.15 134463
```

```

# by state and region represents all other variables not
used in the formula
dcast(mstates, ... ~ variable, mean)
```

```
##             state        region Population Income
Illiteracy Life.Exp
## 1         Alabama         South       3615   3624
2.1    69.05
## 2          Alaska          West        365   6315
1.5    69.31
## 3         Arizona          West       2212   4530
1.8    70.55
## 4        Arkansas         South       2110   3378
1.9    70.66
## 5      California          West      21198   5114
1.1    71.71
## 6        Colorado          West       2541   4884
0.7    72.06
## 7     Connecticut     Northeast       3100   5348
1.1    72.48
## 8        Delaware         South        579   4809
0.9    70.06
## 9         Florida         South       8277   4815
1.3    70.66
```

```
## 10        Georgia         South       4931   4091
2.0    68.54
## 11         Hawaii          West        868   4963
1.9    73.60
## 12          Idaho          West        813   4119
0.6    71.87
## 13       Illinois North Central      11197   5107
0.9    70.14
## 14        Indiana North Central       5313   4458
0.7    70.88
## 15           Iowa North Central       2861   4628
0.5    72.56
## 16         Kansas North Central       2280   4669
0.6    72.58
## 17       Kentucky         South       3387   3712
1.6    70.10
## 18      Louisiana         South       3806   3545
2.8    68.76
## 19          Maine     Northeast       1058   3694
0.7    70.39
## 20       Maryland         South       4122   5299
0.9    70.22
## 21  Massachusetts     Northeast       5814   4755
1.1    71.83
## 22       Michigan North Central       9111   4751
0.9    70.63
## 23      Minnesota North Central       3921   4675
0.6    72.96
## 24    Mississippi         South       2341   3098
2.4    68.09
## 25       Missouri North Central       4767   4254
0.8    70.69
## 26        Montana          West        746   4347
0.6    70.56
## 27       Nebraska North Central       1544   4508
0.6    72.60
## 28         Nevada          West        590   5149
0.5    69.03
## 29  New Hampshire     Northeast        812   4281
0.7    71.23
## 30     New Jersey     Northeast       7333   5237
1.1    70.93
## 31     New Mexico          West       1144   3601
2.2    70.32
## 32       New York     Northeast      18076   4903
1.4    70.55
## 33 North Carolina         South       5441   3875
1.8    69.21
## 34   North Dakota North Central        637   5087
0.8    72.78
## 35           Ohio North Central      10735   4561
0.8    70.82
## 36       Oklahoma         South       2715   3983
1.1    71.42
## 37         Oregon          West       2284   4660
```

```
0.6    72.13
## 38   Pennsylvania     Northeast      11860   4449
1.0    70.43
## 39   Rhode Island     Northeast        931   4558
1.3    71.90
## 40 South Carolina         South       2816   3635
2.3    67.96
## 41   South Dakota North Central        681   4167
0.5    72.08
## 42      Tennessee         South       4173   3821
1.7    70.11
## 43          Texas         South      12237   4188
2.2    70.90
## 44           Utah          West       1203   4022
0.6    72.90
## 45        Vermont     Northeast        472   3907
0.6    71.64
## 46       Virginia         South       4981   4701
1.4    70.08
## 47     Washington          West       3559   4864
0.6    71.72
## 48  West Virginia         South       1799   3617
1.4    69.48
## 49      Wisconsin North Central       4589   4468
0.7    72.48
## 50        Wyoming          West        376   4566
0.6    70.29
##    Murder HS.Grad Frost   Area
## 1    15.1    41.3    20  50708
## 2    11.3    66.7   152 566432
## 3     7.8    58.1    15 113417
## 4    10.1    39.9    65  51945
## 5    10.3    62.6    20 156361
## 6     6.8    63.9   166 103766
## 7     3.1    56.0   139   4862
## 8     6.2    54.6   103   1982
## 9    10.7    52.6    11  54090
## 10   13.9    40.6    60  58073
## 11    6.2    61.9     0   6425
## 12    5.3    59.5   126  82677
## 13   10.3    52.6   127  55748
## 14    7.1    52.9   122  36097
## 15    2.3    59.0   140  55941
## 16    4.5    59.9   114  81787
## 17   10.6    38.5    95  39650
## 18   13.2    42.2    12  44930
## 19    2.7    54.7   161  30920
## 20    8.5    52.3   101   9891
## 21    3.3    58.5   103   7826
## 22   11.1    52.8   125  56817
## 23    2.3    57.6   160  79289
## 24   12.5    41.0    50  47296
## 25    9.3    48.8   108  68995
## 26    5.0    59.2   155 145587
## 27    2.9    59.3   139  76483
## 28   11.5    65.2   188 109889
## 29    3.3    57.6   174   9027
## 30    5.2    52.5   115   7521
## 31    9.7    55.2   120 121412
## 32   10.9    52.7    82  47831
## 33   11.1    38.5    80  48798
## 34    1.4    50.3   186  69273
## 35    7.4    53.2   124  40975
## 36    6.4    51.6    82  68782
## 37    4.2    60.0    44  96184
## 38    6.1    50.2   126  44966
## 39    2.4    46.4   127   1049
## 40   11.6    37.8    65  30225
## 41    1.7    53.3   172  75955
## 42   11.0    41.8    70  41328
## 43   12.2    47.4    35 262134
## 44    4.5    67.3   137  82096
## 45    5.5    57.1   168   9267
## 46    9.5    47.8    85  39780
## 47    4.3    63.5    32  66570
## 48    6.7    41.6   100  24070
## 49    3.0    54.5   149  54464
## 50    6.9    62.9   173  97203
```

```
#
dcast(mstates, region ~ ., mean)
```

```
##          region    NA
## 1     Northeast  3559
## 2         South  7877
## 3 North Central  9042
## 4          West 17791
```

The variable(s) on the left hand side of ~ will appear in the column(s) of the result, whereas the variable(s) on the right hand side of ~ will appear in the rows. The order of the variable matters, the first varies slowest, and the last fastest

To limit the variables that are used, we can use the subset= argument of cast. Since this argument uses the melted data, we need to refer to the variable named variable:

```
library(plyr)
dcast(mstates, region ~ variable, mean, subset = .(variable
%in% c("Population",
    "Life.Exp")))
```

```
##          region Population Life.Exp
## 1     Northeast       5495    71.26
## 2         South       4208    69.71
## 3 North Central       4803    71.77
## 4          West       2915    71.23
```

```
# introduce margins
dcast(mstates, region ~ variable, mean, subset = .(variable
%in% c("Population",
    "Life.Exp")), margins = "region")
```

```
##          region Population Life.Exp
## 1     Northeast       5495    71.26
## 2         South       4208    69.71
## 3 North Central       4803    71.77
## 4          West       2915    71.23
## 5         (all)       4246    70.88
```

```
# inline function
dcast(mstates, region ~ variable, function(x) mean(x),
subset = .(variable %in%
    c("Population", "Life.Exp")))
```

```
##          region Population Life.Exp
## 1     Northeast       5495    71.26
## 2         South       4208    69.71
## 3 North Central       4803    71.77
## 4          West       2915    71.23
```

```
# pass arguments
dcast(mstates, region ~ variable, sum, subset = .(variable
%in% c("Population",
    "Life.Exp")), trim = 0.1)
```

```
##          region Population Life.Exp
## 1     Northeast      49456    641.5
## 2         South      67330   1115.4
## 3 North Central      57636    861.3
## 4          West      37899    926.1
```

```
aggregate(state.x77, list(Region = state.region), mean)
```

```
##          Region Population Income Illiteracy Life Exp
Murder HS Grad
## 1     Northeast       5495   4570      1.000    71.26
4.722   53.97
## 2         South       4208   4012      1.738    69.71
10.581   44.34
## 3 North Central       4803   4611      0.700    71.77
5.275   54.52
## 4          West       2915   4703      1.023    71.23
7.215   62.00
##    Frost   Area
## 1 132.78  18141
## 2  64.62  54605
## 3 138.83  62652
## 4 102.15 134463
```

reshape vs reshape2

```
library(reshape)
```

```
##
## Attaching package: 'reshape'
##
## The following objects are masked from 'package:plyr':
##
##     rename, round_any
##
## The following objects are masked from
'package:reshape2':
##
##     colsplit, melt, recast
```

```
dstats <- function(x) (c(n = length(x), mean = mean(x), sd
= sd(x)))
dfm <- melt(mtcars, measure.vars = c("mpg", "hp", "wt"),
id.vars = c("am", "cyl"))
cast(dfm, am + cyl + variable ~ ., dstats)
```

```
##    am cyl variable  n    mean      sd
## 1   0   4      mpg  3  22.900  1.4526
## 2   0   4       hp  3  84.667 19.6554
## 3   0   4       wt  3   2.935  0.4075
## 4   0   6      mpg  4  19.125  1.6317
## 5   0   6       hp  4 115.250  9.1788
## 6   0   6       wt  4   3.389  0.1162
## 7   0   8      mpg 12  15.050  2.7744
## 8   0   8       hp 12 194.167 33.3598
## 9   0   8       wt 12   4.104  0.7683
## 10  1   4      mpg  8  28.075  4.4839
## 11  1   4       hp  8  81.875 22.6554
## 12  1   4       wt  8   2.042  0.4093
## 13  1   6      mpg  3  20.567  0.7506
## 14  1   6       hp  3 131.667 37.5278
## 15  1   6       wt  3   2.755  0.1282
## 16  1   8      mpg  2  15.400  0.5657
## 17  1   8       hp  2 299.500 50.2046
## 18  1   8       wt  2   3.370  0.2828
```

```
ddply(dfm, .(am, cyl, variable), summarise, n =
length(value), mean = mean(value),
    sd = sd(value))
```

```
##    am cyl variable  n    mean      sd
## 1   0   4      mpg  3  22.900  1.4526
## 2   0   4       hp  3  84.667 19.6554
## 3   0   4       wt  3   2.935  0.4075
## 4   0   6      mpg  4  19.125  1.6317
## 5   0   6       hp  4 115.250  9.1788
## 6   0   6       wt  4   3.389  0.1162
## 7   0   8      mpg 12  15.050  2.7744
## 8   0   8       hp 12 194.167 33.3598
## 9   0   8       wt 12   4.104  0.7683
## 10  1   4      mpg  8  28.075  4.4839
## 11  1   4       hp  8  81.875 22.6554
## 12  1   4       wt  8   2.042  0.4093
## 13  1   6      mpg  3  20.567  0.7506
## 14  1   6       hp  3 131.667 37.5278
## 15  1   6       wt  3   2.755  0.1282
## 16  1   8      mpg  2  15.400  0.5657
## 17  1   8       hp  2 299.500 50.2046
## 18  1   8       wt  2   3.370  0.2828
```

References:

- http://cran.r-project.org/web/packages/reshape2/reshape2.pdf
- http://had.co.nz/reshape/introduction.pdf