

Analysis by Group

Working with Rows

```

• adding, removing, and replacing
> (people <- rbind(people, Joe=list(25, "M", 64)))
  age gender height
Stan   24     M    61
Ruoja  20     F    55
Sam    18     F    52
Ahmed  20     M    57
Felicia 19    F    57
Joe    25     M    64
> people[row.names(people) != "Stan",] # or just people[-1,]
  age gender height
Ruoja  20     F    55
Sam    18     F    52
Ahmed  20     M    57
Felicia 19    F    57
Joe    25     M    64
> people[1,] <- list(18, "M", 58); people
  age gender height
Stan   18     M    58
Ruoja  20     F    55
Sam    18     F    52
Ahmed  20     M    57
Felicia 19    F    57
Joe    25     M    64
>

```

21

```

> mtcars
      mpg  cyl  disp  hp drat   wt  qsec vs am gear carb
Mazda RX4      21.0   6 160.0 110 3.90 2.620 16.46 0 1  4  4
Mazda RX4 Wag  21.0   6 160.0 110 3.90 2.875 17.02 0 1  4  4
Datsun 710     22.8   4 108.0  93 3.85 2.320 18.61 1 1  4  1
[ . . . ]
> tapply(mtcars$mpg, mtcars$am, mean)
 0      1
17.14737 24.39231
> tapply(mtcars$mpg, list(mtcars$am, mtcars$gear), mean)
 3      4      5
0 16.10667 21.050  NA
1      NA 26.275 21.38
> by(mtcars[,c("mpg", "hp", "qsec")],
     list(manual=mtcars$am, gears=mtcars$gear), mean)
manual: 0
gears: 3
      mpg      hp      qsec
16.10667 176.13333 17.69200
-----
manual: 0
gears: 4
      mpg      hp      qsec
21.050 100.750 20.025
-----
[ . . . ]
> aggregate(mtcars[,c("mpg", "hp", "qsec")],
            list(am=mtcars$am, gear=mtcars$gear), mean)
  am gear      mpg      hp      qsec
1  0   3 16.10667 176.1333 17.692
2  0   4 21.05000 100.7500 20.025
3  1   4 26.27500  83.8750 18.435
4  1   5 21.38000 195.6000 15.640
>

```

22

Reshaping

```

> michaelson
  expt run1 run2 run3 run4 run5      run20
1     1  850  740  900 1070  930      960
2     2  960  940  960  940  880      800
3     3  880  880  880  860  720      840
4     4  890  810  810  820  800      780
5     5  890  840  780  810  760      870
> (m.long <- reshape(michaelson, idvar="expt",
                    varying=list(paste("run", 1:20, sep="")),
                    timevar="run", v.names="speed",
                    direction="long"))
  expt run speed
1.1    1  1  850
2.1    2  1  960
3.1    3  1  880
4.1    4  1  890
5.1    5  1  890
1.2    1  2  740
2.2    2  2  940
[ . . . ]
> summary(aov(speed ~ factor(run) + factor(expt), data=m.long))
      Df Sum Sq Mean Sq F value Pr(>F)
factor(run) 19 113344    5965  1.1053 0.363209
factor(expt)  4  94514   23629  4.3781 0.003071 **
Residuals    76 410166    5397
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>

```

23

Cross-Tabulation

Questionnaire Data:

```

> names(quest)
[1] "prereq.course" "prereq.when" "stat305" "stat404"
[5] "stat441"      "stat442"      "stat450" "windows"
[9] "unix"         "s"            "r"        "sas"
[13] "matlab"       "emacs"        "c"        "java"
[17] "perl"
> attach(quest)
>

```

Frequency tables:

- One factor:

```

> prereq.course
[1] n    t    y    t    y    y    y    y    y
[ . . . ]
Levels: n < t < equiv < y
> table(prereq.course)
prereq.course
      n    t equiv  y
      4    3    2   38
>

```

- Two factor:

```

> table(prereq.course, r)
      r
prereq.course 0  1  2  3
n              2  2  0  0
t              0  2  1  0
equiv         2  0  0  0
y             2  4 20 12
>

```

24

Cross-Tabulation

Cross-Tabulation

- Multi-factor:

- Using table:

```
> table(prereq.course,stat305,stat404,stat441)
, , stat404 = -, stat441 = -
```

```
      stat305
prereq.course - n y
n             1 0 0
t             0 0 1
equiv        1 0 0
y             0 0 2
```

```
, , stat404 = n, stat441 = -
```

```
      stat305
prereq.course - n y
n             0 0 0
t             0 0 0
equiv        0 0 0
y             0 0 0
[ . . . about 60 more lines . . . ]
>
```

25

- Multi-factor:

- Summarized in a data frame:

```
> aggregate(row.names(quest),
            list(prereq.course,stat305,stat404,stat441),
            length)
  Group.1 Group.2 Group.3 Group.4 factor(x)
1      n      -      -      -           1
2  equiv      -      -      -           1
3      t      y      -      -           1
4      y      y      -      -           2
[ . . . ]
```

```
> aggregate(list(freq=row.names(quest)),
            list(prereq=prereq.course,s305=stat305,
                s404=stat404,s441=stat441),
            length)
  prereq s305 s404 s441 freq
1      n      -      -      -           1
2  equiv      -      -      -           1
3      t      y      -      -           1
4      y      y      -      -           2
5      y      y      y      -           1
6      n      n      n      n           2
7      t      n      n      n           1
8  equiv      n      n      n           1
9      y      n      n      n           2
10     n      y      n      n           1
11     t      y      n      n           1
12     y      y      n      n           8
13     y      y      y      n          10
14     y      y      -      y           1
15     y      y      y      y          14
>
```

26