An Introduction to R 2.1 Descriptive statistics

- Commands:
 - Calculate means using mean()
 - Calculate medians using median()
 - Find the mode using modeOf() [Isr package]

- Commands:
 - Calculate means using mean()
 - Calculate medians using median()
 - Find the mode using modeOf() [lsr package]

```
> mean( expt$age )
[1] 25.25
```

- Commands:
 - Calculate means using mean()
 - Calculate medians using median()
 - Find the mode using modeOf() [lsr package]

```
> mean( expt$age )
[1] 25.25
```

```
> median( expt$age )
[1] 25
```

- Commands:
 - Calculate means using mean()
 - Calculate medians using median()
 - Find the mode using modeOf() [lsr package]

```
> mean( expt$age )
[1] 25.25
```

```
> median( expt$age )
[1] 25
```

```
> library(lsr)
> modeOf( expt$age )
[1] 25
```

- Sometimes there are missing data
 - These are represented as NA values
 - Different functions handle NA values differently

- Sometimes there are missing data
 - These are represented as NA values
 - Different functions handle NA values differently
- What is the mean of 3, 4, 5 and NA?
 - Pragmatic answer: ignore the missing data, and calculate the average of 3,4 and 5... i.e., $\underline{\text{mean}} = 4$
 - Cautious answer: we don't know the missing value, so we don't know the mean either... i.e. mean = NA

```
> age <- c( 32, 19, NA, 64 )
> mean( age )
[1] NA
```

By default, mean() gives the conservative "don't know" answer

```
> age <- c( 32, 19, NA, 64 )
> mean( age )
[1] NA
```

```
> mean( age, na.rm=TRUE )
[1] 38.33333
```

But we can force it to be a pragmatist: tell R to remove the NA values by specifying na.rm=TRUE

(the na.rm argument shows up in quite a lot of functions)

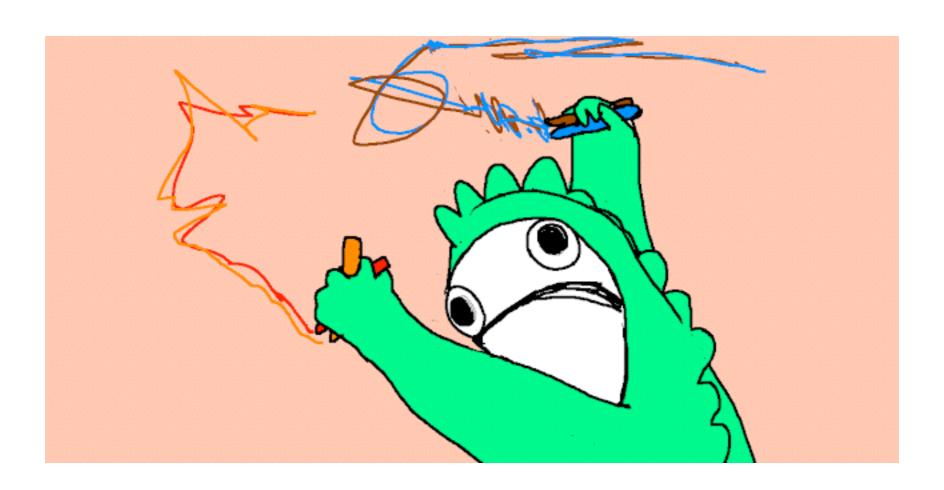
Calculating a trimmed mean

Sometimes the mean isn't a compelling measure of central tendency, but we'd prefer not to resort to the median because the sample size is so small

Calculating a trimmed mean

This gives the 10% trimmed mean, a more robust measure of central tendency than the mean

Try it yourself (Exercise 2.1.1)



- Standard deviation: sd()
- Range: range()
- Interquartile range: IQR()
- Specific quantiles: quantile()

- Standard deviation: sd()
- Range: range()
- Interquartile range: IQR()
- Specific quantiles: quantile()

```
> sd( expt$age )
[1] 3.250874
```

- Standard deviation: sd()
- Range: range()
- Interquartile range: IQR()
- Specific quantiles: quantile()

```
> sd( expt$age )
[1] 3.250874
```

```
> range( expt$age )
[1] 19 30
```

- Standard deviation: sd()
- Range: range()
- Interquartile range: IQR()
- Specific quantiles: quantile()

```
> sd( expt$age )
[1] 3.250874
```

```
> range( expt$age )
[1] 19 30
```

```
> IQR( expt$age )
[1] 4.25
```

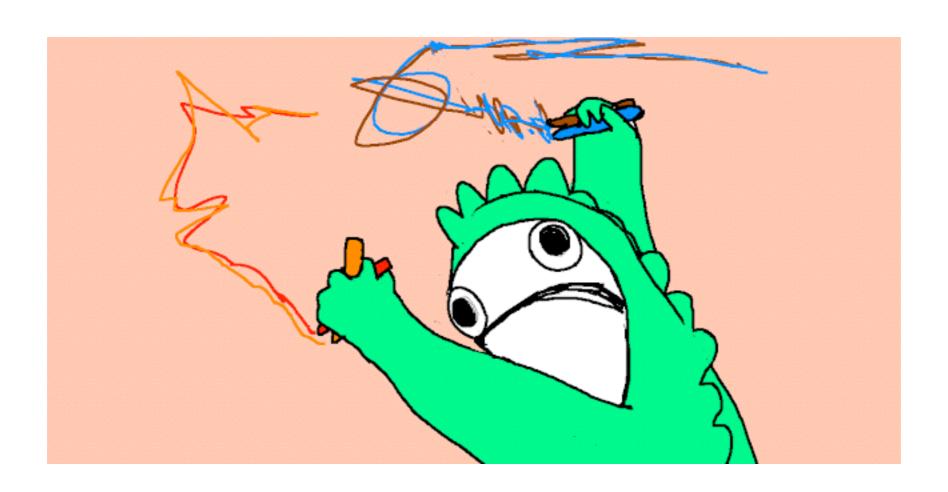
- Standard deviation: sd()
- Range: range()
- Interquartile range: IQR()
- Specific quantiles: quantile()

```
> range( expt$age )
[1] 19 30
```

```
> IQR( expt$age )
[1] 4.25
```

```
> quantile( expt$age, probs=c(.05,.25,.5,.75,.95))
   5%   25%   50%   75%   95%
20.10 23.75 25.00 28.00 29.45
```

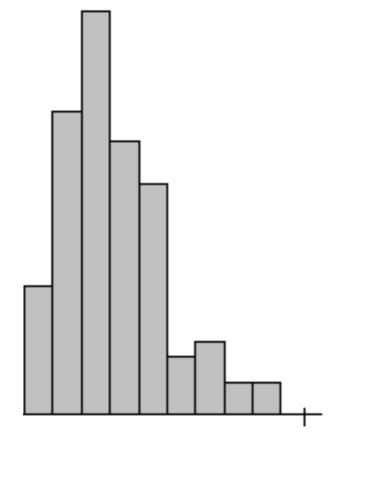
Try it yourself (Exercise 2.1.2)



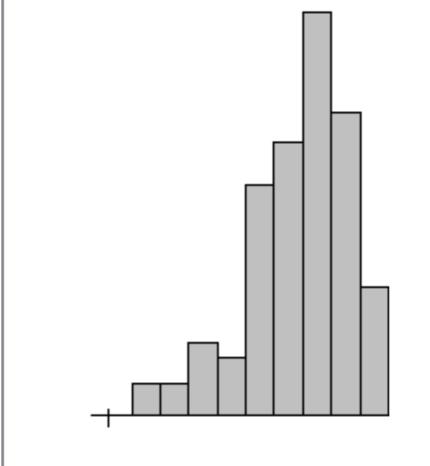
Higher order moments: Skew and kurtosis (briefly)

Skewness = asymmetry

positive skewness... the data "skews" out to the <u>right</u> (i.e. a long tail of large values)



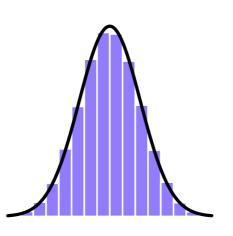
negative skewness... the data "skews" out to the <u>left</u> (i.e. a long tail of small values)



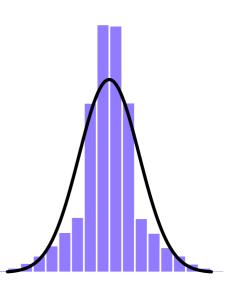
Kurtosis = pointiness

Platykurtic ("too flat")

Mesokurtic



Leptokurtic ("too pointy")



kurtosis < 0

kurtosis = 0

kurtosis > 0

Skew and kurtosis

- Skew: skew() [psych package]
- Kurtosis: kurtosi() [psych package]

Skew and kurtosis

- Skew: skew() [psych package]
- Kurtosis: kurtosi() [psych package]

```
> library(psych)
> skew( expt$age )
[1] -0.3210882
> kurtosi( expt$age )
[1] -0.9841378
```

Tabulating and cross-tabulating categorical variables

R always has lots of ways to do things

- Here are two ways to tabulate variables
 - The table() function
 - The xtabs() function
- Normally I wouldn't bother showing both, but there's a very good reason in this case...

Tabulating using table()

```
> table( expt$treatment )
control drug1 drug2
4 4 4
```

Frequency table for the treatments

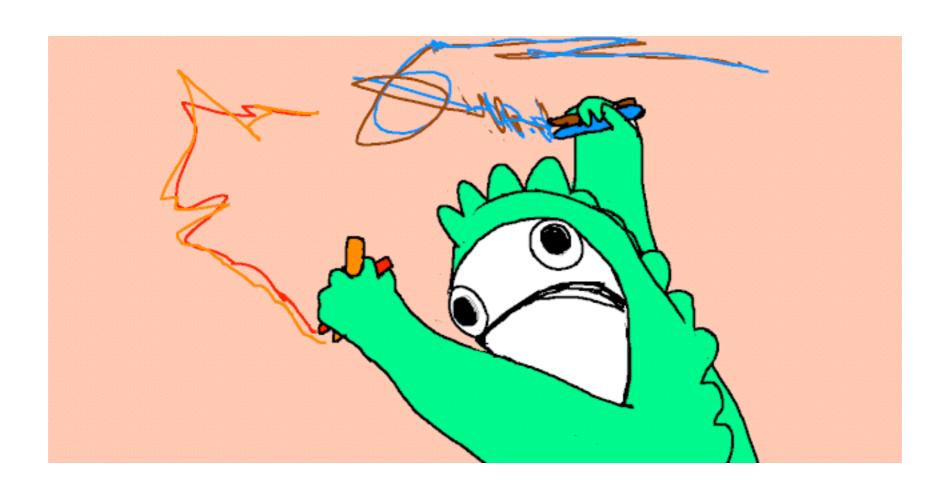
Tabulating using table()

```
> table( expt$treatment )
control drug1 drug2
4 4 4
```

We can get a cross tabulation simply by listing more variables in the input

```
> table( expt$age, expt$treatment, expt$gender )
     = male
     control drug1 drug2
                                       Adding a third variable
  19
                  0
                                         gives a three way
  21
                                          cross-tabulation
  23
  24
                        0
                        1
  25
                        0
  26
  28
                        0
  29
                  0
                        0
                  0
  30
     = female
     control drug1 drug2
  19
  21
  23
                        0
  24
                        0
  25
                  0
  26
  28
                  0
  29
                        0
                  0
                        1
  30
```

Try it yourself (Exercise 2.1.3)



table() versus xtabs()

When we do the cross tabulation using table(), we type in a list of variable names

table() versus xtabs()

xtabs() works a bit differently.

We specify the name of the data frame (ie. expt), and a formula that indicates which variables need to be crosstabulated

Digression: Formulas

Formulas

- A formula is an abstract way to write down variable relationships
 - The precise meaning depends on the context
 - Formulas get used a lot in R, so it's helpful to see some examples...

Examples

• In xtabs(), a "one-sided" formula is used to specify a set of variables to cross tabulate...

```
~ variable1 + variable2
```

~ variable1

Examples

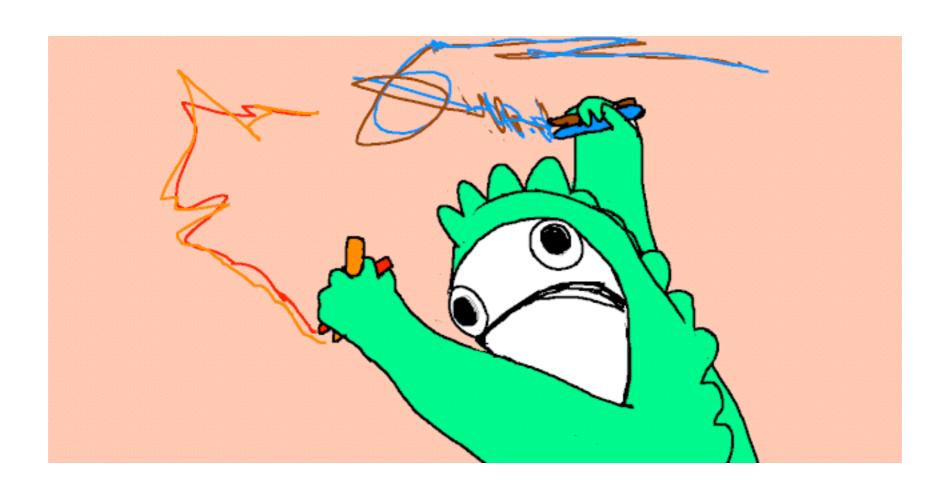
• In xtabs(), a "one-sided" formula is used to specify a set of variables to cross tabulate...

```
~ variable1 + variable2
~ variable1
```

 In Im(), a "two-sided" formula is used to specify a regression model...

```
outcome ~ predictor1 + predictor2
outcome ~ predictor1 * predictor2
```

Try it yourself (Exercise 2.1.4)



Getting lots of descriptive statistics quickly and easily...

Useful commands

- Getting lots of descriptive information for several variables at once:
 - describe() [psych package]
 - summary()
- Getting descriptive statistics separately for several groups:
 - describeBy() [psych package]
 - by() and summary()
 - aggregate()

```
> library( psych )
> describe( expt )
```

```
sd median trimmed
                                                                          skew kurtosis
                    mean
                                                 mad
                                                        min
                                                              max range
           var
                                                                                           se
             1 12
                    6.50
id
                          3.61
                                 6.50
                                          6.50
                                                4.45
                                                       1.00 12.00 11.00
                                                                          0.00
                                                                                   -1.50 1.04
                                25.00
             2 12 25.25
                                                3.71 19.00 30.00 11.00 -0.32
age
                          3.25
                                         25.40
                                                                                  -0.980.94
             3 12
                    1.50
                          0.52
                                  1.50
                                          1.50
                                                0.74
                                                       1.00
                                                             2.00
                                                                    1.00
                                                                          0.00
                                                                                   -2.16 0.15
gender*
                                                                                   -1.740.25
             4 12
                    2.00
                          0.85
                                 2.00
                                          2.00
                                                1.48
                                                       1.00
                                                             3.00
                                                                    2.00
                                                                          0.00
treatment*
                                42.15
                                         41.80 23.28
                                                                                   -0.397.18
             5 12 43.59 24.86
                                                       6.70 98.40 91.70
                                                                          0.51
hormone
happy
             6 12
                    3.71
                          1.02
                                 3.42
                                          3.69
                                                0.84
                                                       2.00
                                                             5.69
                                                                    3.69
                                                                          0.28
                                                                                   -0.880.29
                    3.42
                          1.60
                                          3.46
                                                       0.34
                                                             6.12
sad
             7 12
                                 3.23
                                                1.91
                                                                    5.78 - 0.16
                                                                                   -0.960.46
```

```
> library( psych )
> describe( expt )
```

```
sd median trimmed
            var
                                                   mad
                                                          min
                                                                             skew kurtosis
                                                                 max range
                     mean
id
                     6.50
                           3.61
                                   6.50
                                            6.50
                                                              12.00
                                                                             0.00
                                                  4.45
                                                         1.00
                   25.25
                                  25.00
                                           25.40
                                                        19.00
                                                              30.00
                                                                            -0.32
                                                                                      -0.980.94
age
              3
                     1.50
                           0.52
                                   1.50
                                            1.50
gender*
                                                         1.00
                                                                2.00
                                                                             0.00
treatment*
                    2.00
                           0.85
                                   2.00
                                            2.00
                                                  1.48
                                                         1.00
                                                                3.00
                                                                             0.00
                   43.59
                          24.86
                                  42.15
                                           41.80 23.28
                                                         6.70
                                                              98.40
                                                                             0.51
hormone
                                                                                      -0.397.18
                12
happy
                           1.02
                                   3.42
                                            3.69
                                                  0.84
                                                         2.00
                                                                5.69
                                                                             0.28
                                                                                      -0.880.29
sad
                12
                    3.42
                           1.60
                                   3.23
                                            3.46
                                                  1.91
                                                         0.34
                                                                6.12
                                                                      5.78 - 0.16
                                                                                      -0.960.46
```

Each row contains descriptive statistics for one of the variables in the data frame.

Variables with asterisks next to the names are factors: the asterisk here is a reminder that most of these measures are inappropriate for nominal scale variables

```
> library( psych )
> describe( expt )
                             sd median trimmed
                                                                            skew kurtosis
                                                  mad
                                                         min
                                                                max range
           var
                    mean
id
                    6.50
                           3.61
                                  6.50
                                           6.50
                                                 4.45
                                                        1.00 12.00
                                                                            0.00
                                                                                    -1.501.04
                                                                    11.00
                12
                                                             30.00
                   25.25
                           3.25
                                 25.00
age
                                          25.40
                                                       19.00
                                                                          -0.32
                                                                                    -0.98 0.94
                12
                    1.50
                                           1.50
                           0.52
                                  1.50
                                                  0.74
                                                              2.00
                                                                            0.00
                                                                                    -2.160.15
gender*
                                                        1.00
                                                                     1.00
              4 12
                    2.00
                                           2.00
                                                              3.00
                           0.85
                                  2.00
                                                  1.48
                                                        1.00
                                                                     2.00
                                                                            0.00
                                                                                    -1.740.25
treatment*
                12
                                          41.80 23.28
                   43.59
                          24.86
                                 42.15
                                                             98.40
                                                                            0.51
hormone
                                                        6.70
                                                                    91.70
                                                                                    -0.397.18
                12
                    3.71
happy
                           1.02
                                  3.42
                                           3.69
                                                 0.84
                                                        2.00
                                                              5.69
                                                                     3.69
                                                                            0.28
                                                                                    -0.880.29
                                                                     5.78 -0.16
                12
                    3.42
                                           3.46
                                                        0.34
                                                              6.12
sad
                           1.60
                                  3.23
                                                 1.91
                                                                                    -0.960.46
```

Number of observations

```
> library( psych )
> describe( expt
                             sd median trimmed
                                                                           skew kurtosis
                    mean
                                                  mad
                                                         min
                                                               max range
           var
                    6.50
                                  6.50
id
                          3.61
                                           6.50
                                                 4.45
                                                                           0.00
                                                                                    -1.501.04
                                                        1.00
                                                             12.00
                                 25.00
                   25.25
                          3.25
                                          25.40
                                                       19.00
                                                             30.00
                                                                          -0.32
                                                                                    -0.980.94
age
             3 12
                    1.50
                          0.52
                                  1.50
                                           1.50
                                                 0.74
gender*
                                                        1.00
                                                              2.00
                                                                     1.00
                                                                           0.00
                    2.00
             4 12
                                  2.00
                          0.85
                                           2.00
                                                 1.48
                                                        1.00
                                                              3.00
                                                                     2.00
                                                                           0.00
treatment*
                  43.59
             5 12
                         24.86
                                 42.15
                                          41.80 23.28
                                                                           0.51
                                                        6.70
                                                             98.40
                                                                    91.70
                                                                                    -0.397.18
hormone
                                  3.42
             6 12
                    3.71
happy
                          1.02
                                           3.69
                                                 0.84
                                                        2.00
                                                              5.69
                                                                     3.69
                                                                           0.28
                                                                                    -0.880.29
                    3.42
                                           3.46
                                                        0.34
sad
             7 12
                           1.60
                                  3.23
                                                 1.91
                                                              6.12
                                                                     5.78 -0.16
                                                                                    -0.960.46
```

Mean, standard deviation and median

```
> library( psych )
> describe( expt )
                             sd median trimmed
                                                        min
                                                                           skew kurtosis
                                                  mad
                                                               max range
                    mean
id
                          3.61
                                  6.50
                                           6.50
                    6.50
                                                 4.45
                                                        1.00 12.00
                                                                           0.00
                                                                                   -1.501.04
                                                                   11.00
                                 25.00
age
                          3.25
                                         25.40
                                                 3.71
                                                      19.00 30.00
                                                                          -0.32
                                                                                   -0.980.94
                          0.52
                                  1.50
                                           1.50
gender*
                    1.50
                                                 0.74
                                                              2.00
                                                                                   -2.160.15
                                                        1.00
                                                                    1.00
                                                                           0.00
                                           2.00
                                  2.00
                    2.00
                          0.85
                                                 1.48
                                                        1.00
                                                              3.00
                                                                    2.00
                                                                           0.00
                                                                                    -1.740.25
treatment*
                                                23.28
                                 42.15
                                         41.80
                   43.59
                         24.86
                                                             98.40
                                                                           0.51
hormone
                                                        6.70
                                                                   91.70
                                                                                   -0.397.18
happy
                                  3.42
                                           3.69
             6 12
                          1.02
                                                 0.84
                                                       2.00
                                                              5.69
                                                                    3.69
                                                                           0.28
                                                                                   -0.880.29
                                                                    5.78 -0.16
                          1.60
                                                       0.34
                                                              6.12
sad
             7 12
                    3.42
                                  3.23
                                           3.46
                                                 1.91
                                                                                   -0.960.46
```

10% trimmed mean

```
> library( psych )
> describe( expt )
                             sd median trimmed
                                                                           skew kurtosis
                                                         min
                                                  mad
                                                               max range
                    mean
id
                    6.50
                          3.61
                                  6.50
                                           6.50
                                                 4.45
                                                        1.00 12.00
                                                                           0.00
                                                                                    -1.501.04
                                                                    11.00
                                          25.40
                   25.25
                          3.25
                                 25.00
                                                 3.71
                                                       19.00 30.00
                                                                          -0.32
                                                                                    -0.980.94
age
                                                 0.74
                    1.50
                          0.52
                                  1.50
                                           1.50
gender*
                                                        1.00
                                                              2.00
                                                                     1.00
                                                                           0.00
                                                                                    -2.160.15
                    2.00
                           0.85
                                  2.00
                                           2.00
                                                 1.48
                                                        1.00
                                                              3.00
                                                                     2.00
treatment*
                                                                           0.00
                                                                                    -1.740.25
                                                23.28
                   43.59
                         24.86
                                 42.15
                                          41.80
                                                        6.70
                                                             98.40 91.70
                                                                           0.51
hormone
                                                                                    -0.397.18
              6 12
                    3.71
                          1.02
                                  3.42
                                           3.69
                                                 0.84
                                                        2.00
                                                              5.69
                                                                     3.69
                                                                           0.28
                                                                                    -0.88 0.29
happy
sad
             7 12
                    3.42
                          1.60
                                  3.23
                                           3.46
                                                 1.91
                                                        0.34
                                                              6.12
                                                                     5.78 - 0.16
                                                                                    -0.960.46
```

A robust estimator of the standard deviation that is computed by a transformation of the median absolute deviation (mad) from the sample median

```
> library( psych )
> describe( expt )
                                                                           skew kurtosis
                             sd median trimmed
                                                         min
                                                               max range
                                                  mad
                    mean
                                                                                            se
id
                                                        1.00 12.00 11.00
                    6.50
                          3.61
                                  6.50
                                           6.50
                                                 4.45
                                                                           0.00
                                                                                    -1.501.04
                   25.25
                                         25.40
                                                 3.71 19.00 30.00 11.00
                                 25.00
                          3.25
                                                                           -0.32
                                                                                    -0.980.94
age
                                                        1.00
                                                              2.00
                    1.50
                          0.52
                                  1.50
                                           1.50
                                                 0.74
                                                                    1.00
                                                                           0.00
                                                                                    -2.160.15
gender*
                                           2.00
                    2.00
                          0.85
                                  2.00
                                                 1.48
                                                        1.00
                                                              3.00
                                                                    2.00
                                                                           0.00
                                                                                    -1.740.25
treatment*
                                                        6.70 98.40 91.70
                                                                           0.51
                   43.59
                                         41.80 23.28
                         24.86
                                 42.15
                                                                                    -0.397.18
hormone
happy
             6 12
                    3.71
                          1.02
                                  3.42
                                           3.69
                                                 0.84
                                                        2.00
                                                              5.69
                                                                    3.69
                                                                           0.28
                                                                                   -0.880.29
                                                                           -0.16
                                           3.46
                                                        0.34
                                                              6.12
sad
             7 12
                    3.42
                          1.60
                                  3.23
                                                 1.91
                                                                    5.78
                                                                                   -0.960.46
```

Information about the range

```
> library( psych )
> describe( expt )
                                                                            skew kurtosis
                             sd median trimmed
                                                   mad
                                                          min
                                                                max range
                                                                                              se
                    mean
id
                                                                             0.00
                                                                                     -1.50
                    6.50
                           3.61
                                   6.50
                                            6.50
                                                  4.45
                                                         1.00 12.00
                                                                     11.00
                                                                                           1.04
                                                                                     -0.98 0.94
                                  25.00
                                                                            -0.32
age
                           3.25
                                          25.40
                                                        19.00
                                                              30.00
                                                                     11.00
                                            1.50
                                                                             0.00
                                                                                     -2.16
                    1.50
                           0.52
                                   1.50
                                                  0.74
                                                         1.00
                                                               2.00
                                                                      1.00
                                                                                            0.15
gender*
                                                                                     -1.74
                                                                             0.00
                    2.00
                           0.85
                                   2.00
                                            2.00
                                                  1.48
                                                         1.00
                                                               3.00
                                                                      2.00
                                                                                           0.25
treatment*
                                                                             0.51
                                                                                     -0.39
                                                                                            7.18
                   43.59
                          24.86
                                  42.15
                                          41.80 23.28
                                                              98.40
                                                                     91.70
hormone
                                                         6.70
                                                                                     -0.88 0.29
happy
              6 12
                           1.02
                                   3.42
                                           3.69
                                                  0.84
                                                         2.00
                                                               5.69
                                                                      3.69
                                                                             0.28
                                                                                     -0.96 | 0.46
                                                                      5.78
                           1.60
                                            3.46
                                                         0.34
                                                               6.12
sad
              7 12
                    3.42
                                   3.23
                                                  1.91
                                                                            +0.16
```

Skew and kurtosis

```
> library( psych )
> describe( expt )
                             sd median trimmed
                                                                             skew kurtosis
                                                   mad
                                                          min
                                                                max range
                                                                                              se
                    mean
id
                    6.50
                           3.61
                                   6.50
                                           6.50
                                                  4.45
                                                         1.00 12.00
                                                                             0.00
                                                                                     -1.50 | 1.04
                                                                     11.00
                                                                                     -0.98 0.94
                           3.25
                                  25.00
                                          25.40
                                                        19.00
                                                              30.00
                                                                           -0.32
age
                                           1.50
                                                                                     -2.16
                                                                                           0.15
                    1.50
                           0.52
                                   1.50
                                                  0.74
gender*
                                                         1.00
                                                               2.00
                                                                      1.00
                                                                             0.00
                                                                                     -1.74 0.25
                    2.00
                           0.85
                                   2.00
                                           2.00
                                                  1.48
                                                         1.00
                                                               3.00
                                                                      2.00
                                                                             0.00
treatment*
                                                                                     -0.39 7.18
                   43.59
                                          41.80 23.28
                                                              98.40
                                                                             0.51
                          24.86
                                  42.15
                                                         6.70
                                                                     91.70
hormone
                                                                                     -0.880.29
happy
              6 12
                    3.71
                           1.02
                                   3.42
                                           3.69
                                                  0.84
                                                         2.00
                                                               5.69
                                                                      3.69
                                                                             0.28
                                                                      5.78 -0.16
                                                         0.34
                                                               6.12
sad
              7 12
                    3.42
                           1.60
                                   3.23
                                           3.46
                                                  1.91
                                                                                     -0.96 0.46
```

Standard error of the mean (computed by the usual normal theory estimate)

Summary

> summary(expt)

```
id
                                 gender
                                        treatment
                   age
               Min.
Min.
                              male :6
                                        control:4
      : 1.00
                     :19.00
                              female:6
1st Qu.: 3.75
               1st Qu.:23.75
                                        drug1:4
Median : 6.50
               Median :25.00
                                        drug2:4
Mean : 6.50
                     :25.25
               Mean
3rd Qu.: 9.25 3rd Qu.:28.00
Max. :12.00
                     :30.00
               Max.
  hormone
                 happy
                                   sad
                     :2.000
Min. : 6.70
               Min.
                              Min.
                                     :0.340
1st Qu.:23.82
               1st Qu.:3.107
                              1st Qu.:2.540
Median :42.15
               Median :3.425
                              Median :3.235
      :43.59
                     :3.712
               Mean
                              Mean
                                     :3.423
Mean
3rd Qu.:55.88
               3rd Qu.:4.522
                              3rd Qu.:4.633
Max. :98.40
                     :5.690
                                     :6.120
               Max.
                              Max.
```

Summary

gender

male :6

female:6

treatment

: 4

control:4

drug1 :4

drug2

> summary(expt)

hormone

Min. : 6.70

1st Qu.:23.82

Median :42.15

Mean :43.59

3rd Qu.:55.88

Max. :98.40

```
id age
Min.: 1.00 Min.: 19.00
1st Qu.: 3.75 1st Qu.:23.75
Median: 6.50 Median: 25.00
Mean: 6.50 Mean: 25.25
3rd Qu.: 9.25 3rd Qu.:28.00
Max.: 12.00 Max.: 30.00
```

happy sad
Min. :2.000 Min. :0.340
1st Qu.:3.107 1st Qu.:2.540
Median :3.425 Median :3.235
Mean :3.712 Mean :3.423
3rd Qu.:4.522 3rd Qu.:4.633
Max. :5.690 Max. :6.120

Summary produces a frequency table for the factor variables

Summary

> summary(expt)

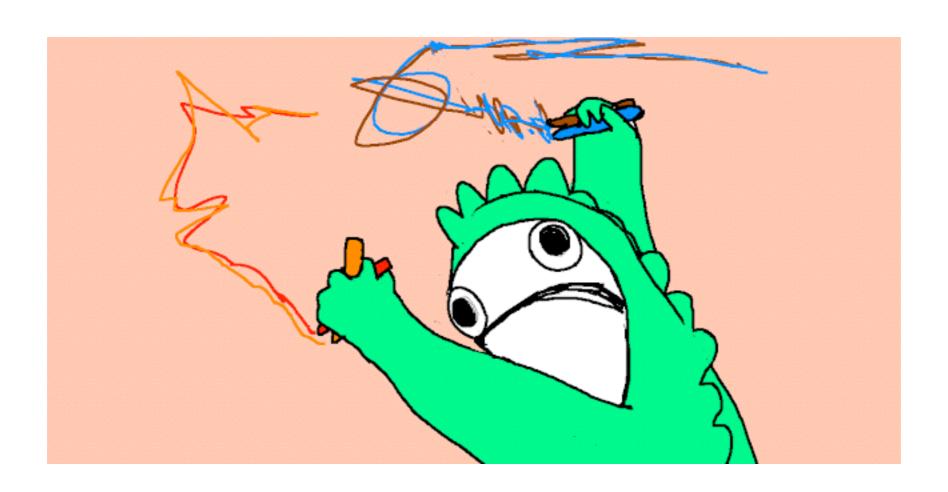
```
id
                      age
                 Min.
                        :19.00
Min.
       : 1.00
1st Qu.: 3.75
                 1st Qu.:23.75
                 Median :25.00
Median : 6.50
       : 6.50
                        :25.25
                Mean
Mean
                 3rd Qu.:28.00
3rd Qu.: 9.25
       :12.00
                        :30.00
                 Max.
Max.
```

```
gender treatment
male :6 control:4
female:6 drug1 :4
drug2 :4
```

```
hormone
                    happy
                                      sad
       : 6.70
Min.
                Min.
                        :2.000
                                 Min.
                                        :0.340
1st Qu.:23.82
                1st Qu.:3.107
                                 1st Qu.:2.540
Median :42.15
                Median :3.425
                                 Median :3.235
       :43.59
                        :3.712
                Mean
                                 Mean
Mean
                                        :3.423
3rd Qu.:55.88
                3rd Qu.:4.522
                                 3rd Qu.:4.633
                                        :6.120
       :98.40
                        :5.690
Max.
                Max.
                                 Max.
```

For numeric variables it gives the mean, plus the 0th, 25th, 50th, 75th and 100th percentiles

Try it yourself (Exercise 2.1.5)



Useful commands

- Getting lots of descriptive information for several variables at once:
 - describe() [psych package]
 - summary()
- Getting descriptive statistics separately for several groups:
 - describeBy() [psych package]
 - by() and summary()
 - aggregate()

describeBy()

describeBy(expt, group=expt\$gender)

data frame containing all the
 variables to be described the groups

describeBy()

> describeBy(expt, group=expt\$gender)

6 6

7 6

happy

sad

3.77

2.87

0.76

1.03

3.67

2.67

```
group: male
                            sd median trimmed
                                                                           skew kurtosis
                                                        min
                                                 mad
                                                               max range
                   mean
                                                                                             se
            var n
                                                 2.22
              1 6
                   3.50
                                 3.50
                                          3.50
                                                       1.00
                                                             6.00
                                                                    5.00
id
                          1.87
                                                                           0.00
                                                                                   -1.80
                                                                                           0.76
                  25.50
                          2.07
                                25.00
                                         25.50
                                                2.22 23.00 28.00
                                                                    5.00
                                                                           0.22
                                                                                           0.85
                                                                                   -1.90
age
                                 1.00
                                          1.00
                                                 0.00
gender*
                   1.00
                          0.00
                                                       1.00
                                                              1.00
                                                                    0.00
                                                                            NaN
                                                                                     NaN
                                                                                           0.00
                   2.00
                                          2.00
              4 6
                          0.89
                                 2.00
                                                 1.48
                                                       1.00
                                                              3.00
                                                                    2.00
                                                                           0.00
                                                                                   -1.96
                                                                                           0.37
treatment*
hormone
                  38.55 32.02
                                31.75
                                         38.55 16.38
                                                       6.70 98.40
                                                                   91.70
                                                                           0.88
                                                                                   -0.7813.07
              6 6
happy
                   3.65
                                          3.65
                          1.30
                                 3.38
                                                1.24
                                                       2.00
                                                              5.69
                                                                    3.69
                                                                           0.32
                                                                                           0.53
                                                                                   -1.49
                   3.98
                          1.97
                                 4.53
                                          3.98
                                                0.96
                                                       0.34
                                                             6.12
                                                                    5.78 -0.82
                                                                                   -0.81
                                                                                           0.80
sad
group: female
                            sd median trimmed
                                                                         skew kurtosis
                                                       min
            var n
                   mean
                                                mad
                                                             max range
                                                                                           se
                                          9.50 2.22
                                                      7.00
              1 6
                   9.50
                                 9.50
                                                           12.00
                                                                   5.00
id
                          1.87
                                                                         0.00
                                                                                  -1.800.76
              2 6 25.00
                          4.34
                                25.50
                                         25.00 5.93 19.00
                                                           30.00 11.00 -0.18
                                                                                  -1.85 1.77
age
              3 6
                                                      2.00
gender*
                   2.00
                          0.00
                                 2.00
                                          2.00 0.00
                                                            2.00
                                                                   0.00
                                                                          NaN
                                                                                    NaN 0.00
              4 6
                   2.00
                          0.89
                                          2.00 1.48
                                                      1.00
                                                            3.00
                                                                   2.00
treatment*
                                 2.00
                                                                         0.00
                                                                                  -1.960.37
hormone
              5 6 48.63 16.54
                                54.90
                                         48.63 8.75
                                                    18.50
                                                           65.20 46.70 -0.82
                                                                                  -0.986.75
```

3.77 0.99

2.87 0.52

2.83

1.87

4.78

4.82

1.95

2.95

0.12

0.94

-1.900.31

-0.670.42

```
> aggregate(
    formula = age ~ gender + treatment,
    data = expt,
    FUN = mean
)
```

```
> aggregate(
    formula = age ~ gender + treatment,
    data = expt,
    FUN = mean
)
```

Tells R you want to summarise "age", broken down separately by "gender" and "treatment"

```
> aggregate(
    formula = age ~ gender + treatment,
    data = expt,
    FUN = mean
)
```

Tells R that the variables are all stored in the data frame called "expt"

```
> aggregate(
    formula = age ~ gender + treatment,
    data = expt,
    FUN = mean
)
```

The name of the function that produces the descriptive statistic that you want... e.g., mean, sd, IQR, etc

```
> aggregate(
    formula = age ~ gender + treatment,
    data = expt,
    FUN = mean
)
```

```
gender treatment age

1 male control 26.5

2 female control 25.5

3 male drug1 23.5

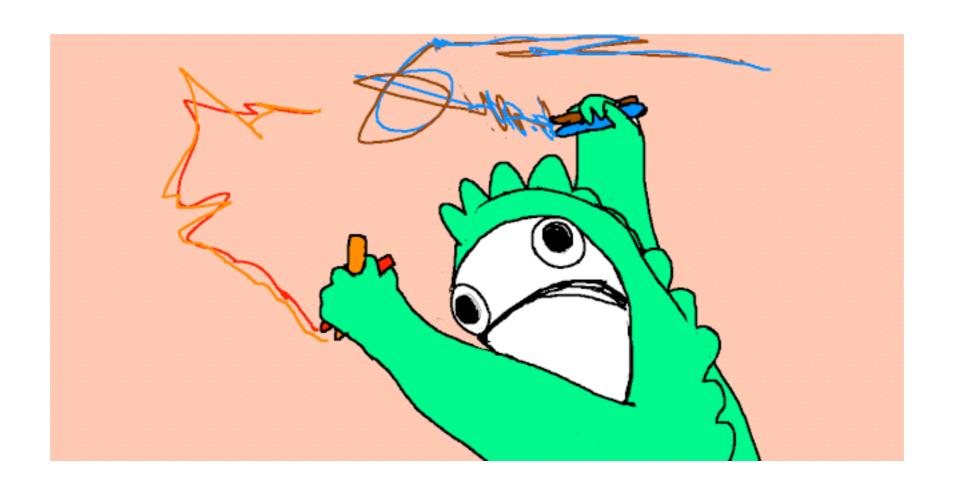
4 female drug1 24.0

5 male drug2 26.5

6 female drug2 25.5
```

The output contains the mean age for every group

Try it yourself (Exercise 2.1.6)



Briefly... another trick is to use by()

```
> by( expt, INDICES=expt$gender, summary )
expt$gender: male
       id
                                    gender
                                               treatment
                                                             hormone
                                                                              happy
                      age
                                                                                                sad
                        :23.00
 Min.
        :1.00
                Min.
                                 male :6
                                             control:2
                                                         Min.
                                                                 : 6.70
                                                                          Min.
                                                                                  :2.000
                                                                                           Min.
                                                                                                  :0.340
 1st Qu.:2.25
                1st 0u.:24.25
                                 female:0
                                             drug1 :2
                                                         1st Qu.:21.48
                                                                          1st Qu.:3.007
                                                                                           1st 0u.:3.768
 Median :3.50
                Median :25.00
                                                    :2
                                                                          Median :3.380
                                             drug2
                                                         Median :31.75
                                                                                           Median :4.525
 Mean
        :3.50
                Mean
                        :25.50
                                                         Mean
                                                                 :38.55
                                                                          Mean
                                                                                  :3.650
                                                                                           Mean
                                                                                                  :3.977
 3rd Qu.:4.75
                3rd Ou.:27.25
                                                         3rd Ou.:41.42
                                                                          3rd Qu.:4.270
                                                                                           3rd 0u.:4.758
        :6.00
                Max.
                        :28.00
                                                                 :98.40
                                                                          Max.
                                                                                  :5.690
                                                                                                   :6.120
 Max.
                                                         Max.
                                                                                           Max.
expt$gender: female
       id
                                     gender
                                                treatment
                                                              hormone
                                                                               happy
                                                                                                 sad
                       age
                                  male :0
                                                          Min.
                                                                  :18.50
                                                                                                    :1.870
        : 7.00
                 Min.
                         :19.00
                                              control:2
                                                                           Min.
                                                                                   :2.830
                                                                                            Min.
 Min.
 1st Qu.: 8.25
                 1st Qu.:22.00
                                  female:6
                                              drug1 :2
                                                          1st Qu.:44.95
                                                                           1st Ou.:3.248
                                                                                            1st Qu.:2.340
 Median : 9.50
                 Median :25.50
                                              drug2 :2
                                                          Median :54.90
                                                                           Median :3.675
                                                                                            Median :2.675
        : 9.50
                         :25.00
                                                                  :48.63
                                                                                   :3.775
                                                                                                    :2.870
                 Mean
                                                          Mean
                                                                           Mean
                                                                                            Mean
 Mean
 3rd Qu.:10.75
                 3rd Qu.:28.25
                                                          3rd Qu.:56.23
                                                                           3rd Qu.:4.357
                                                                                            3rd Qu.:2.882
```

:65.20

Max.

Max.

:4.780

Max.

:4.820

:12.00

Max.

Max.

:30.00

Descriptives 2: Correlating two variables

Correlations

```
> cor( expt$happy, expt$sad )
[1] -0.7331656
```

Pearson correlation

Correlations

```
> cor( expt$happy, expt$sad )
[1] -0.7331656
```

```
> cor( expt$happy, expt$sad, method="spearman" )
[1] -0.7075317
```

Spearman correlation

Correlations

```
> cor( expt$happy, expt$sad )
[1] -0.7331656
```

```
> cor( expt$happy, expt$sad, method="spearman" )
[1] -0.7075317
```

```
> cor( expt$happy, expt$sad, method="kendall" )
[1] -0.5343667
```

Kendall's tau

All pairwise correlations

```
> library( lsr )
> correlate( expt )

Just a reminder that you need to have the lsr
package loaded for this command to work!
```

The correlate command itself

All pairwise correlations

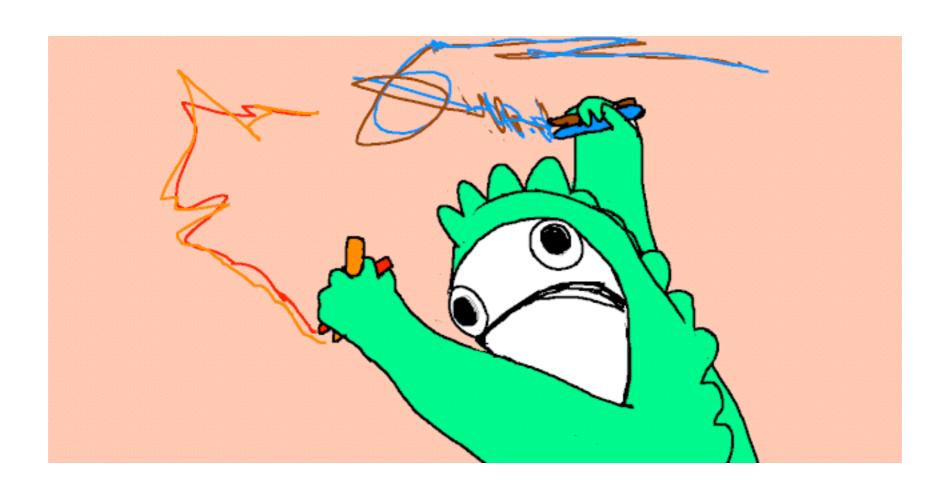
```
> library( lsr )
> correlate( expt )
```

CORRELATIONS

- correlation type: pearson
- correlations shown only when both variables are numeric

```
age gender treatment hormone happy sad
             id
               -0.012
                                        0.316 0.120 - 0.486
id
                                        0.238 0.203 -0.265
         -0.012
age
gender
treatment
                                             0.880 -0.952
          0.316 0.238
hormone
       0.120 0.203
                                    . 0.880
                                                  -0.733
happy
         -0.486 - 0.265
sad
                                       -0.952 - 0.733
```

Try it yourself (Exercise 2.1.7)



End of this section