An Introduction to R
2.1 Descriptive statistics

Dan Navarro (daniel.navarro@adelaide.edu.au)
School of Psychology, University of Adelaide
ua.edu.au/ccs/people/dan
DSTO R Workshop, 27-Apr-2015

Central tendency

Central tendency

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

Central tendency

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

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

Central tendency

- 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
```

Central tendency

- 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
```

What if there are missing data?

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

What if there are missing data?

- 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., mean = 4
  - Cautious answer: we don't know the missing value, so we don't know the mean either… i.e. mean = NA

What if there are missing data?

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

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

What if there are missing data?

```
> 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

```
> score <- c( 3, 2, 1, 5, 7,
              12, 3, 1, 4, 10000 )

> mean( score )
[1] 1003.8
```

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

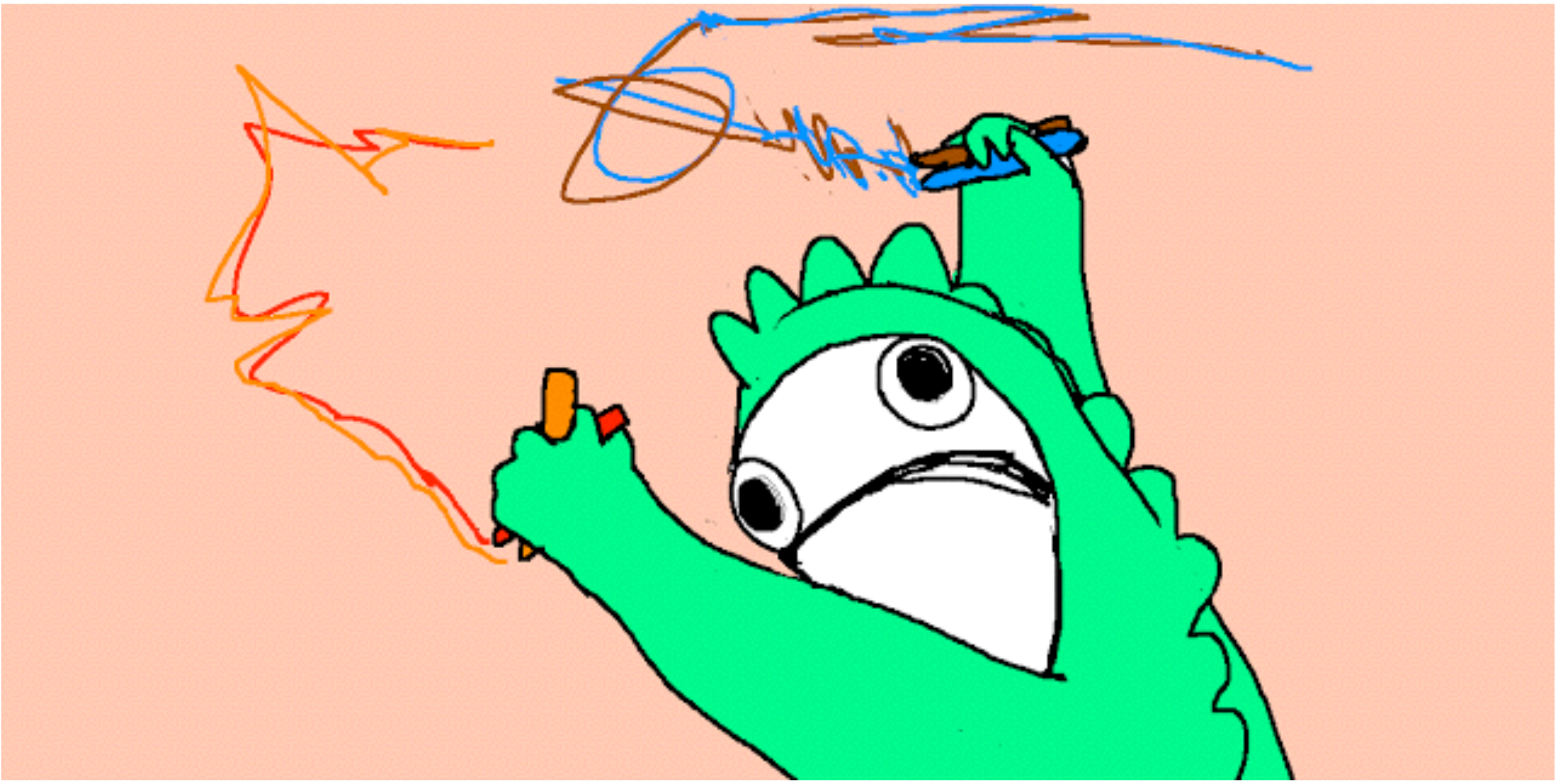
Calculating a trimmed mean

```
> score <- c( 3, 2, 1, 5, 7,
             12, 3, 1, 4, 10000 )

> mean( score )
[1] 1003.8
```

```
> mean( score, trim=.1 )
[1] 4.625
```

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

Try it yourself (Exercise 2.1.1)



Spread

Spread

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

Spread

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

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

Spread

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

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

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

Spread

- 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
```

Spread

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

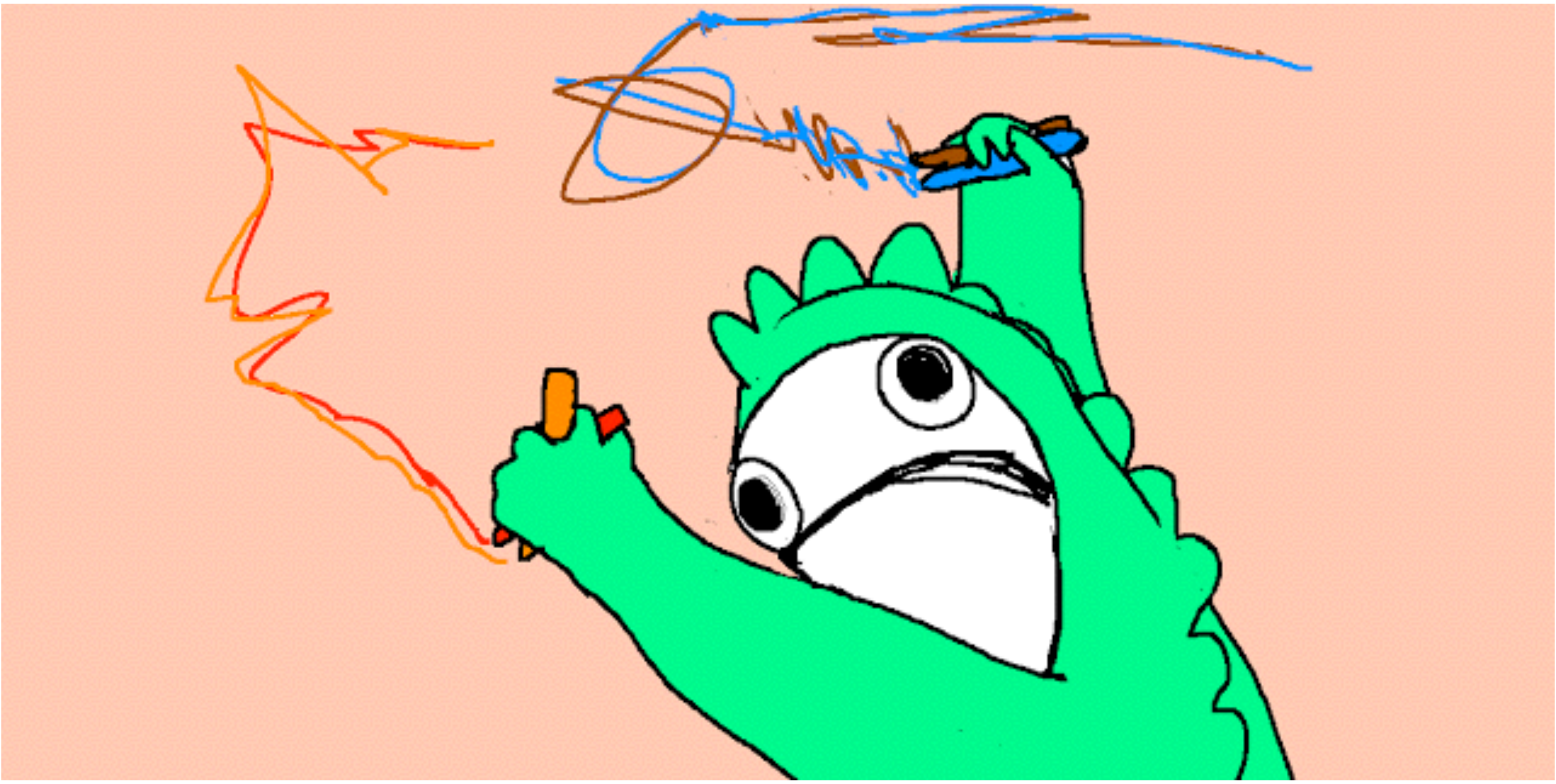
```
> sd( expt$age )
[1] 3.250874
```

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

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

```
> 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 
```

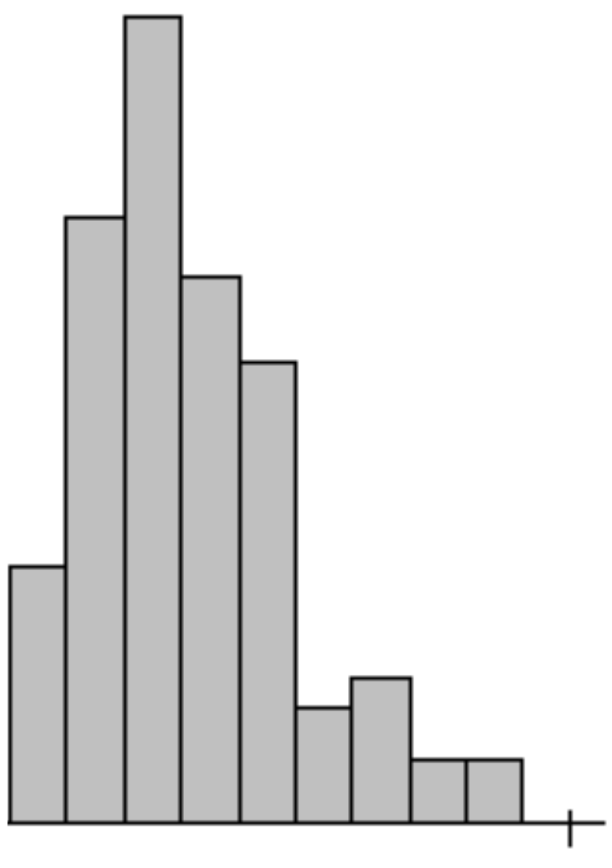
Try it yourself (Exercise 2.1.2)



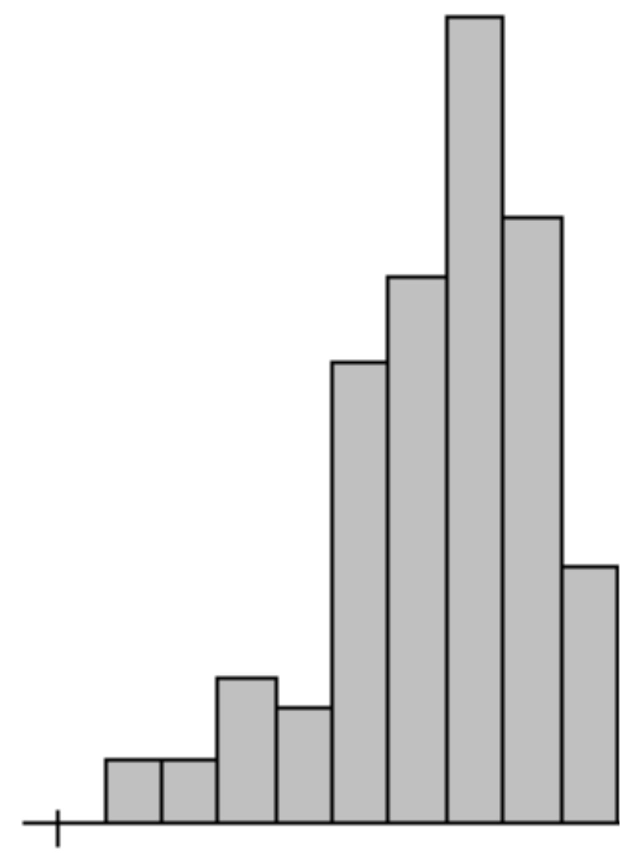
Higher order moments: Skew and kurtosis (briefly)

Skewness = asymmetry

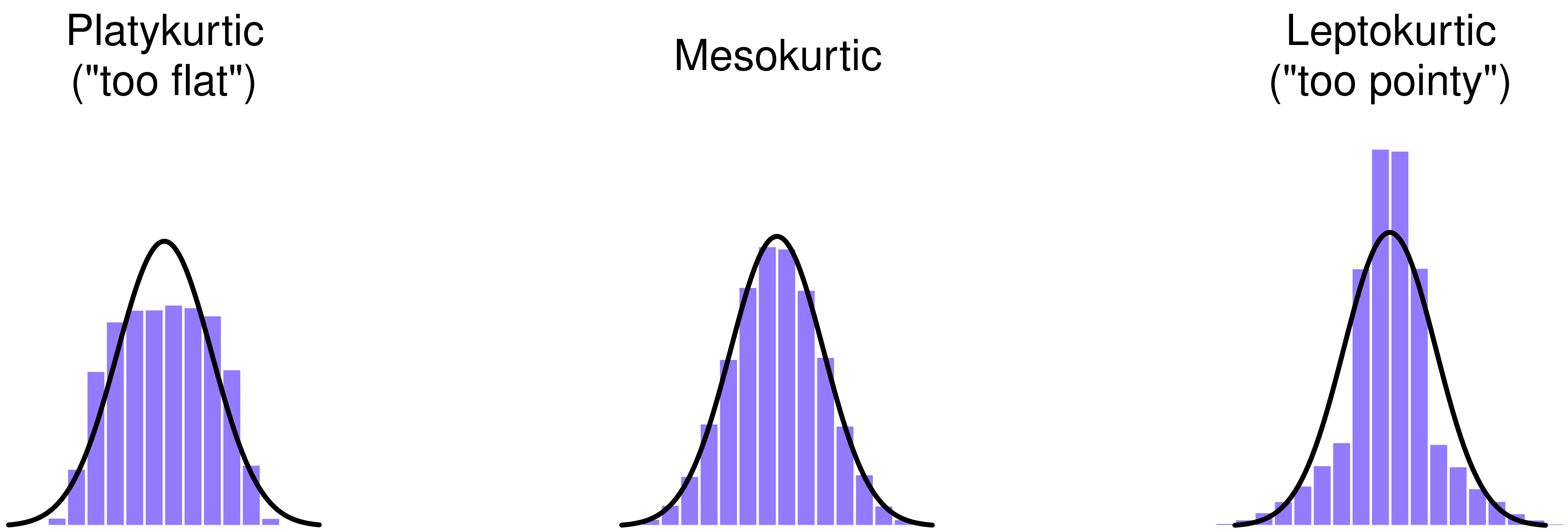
positive skewness… the data "skews" out to the right (i.e. a long tail of large values)



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



Kurtosis = pointiness



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
```

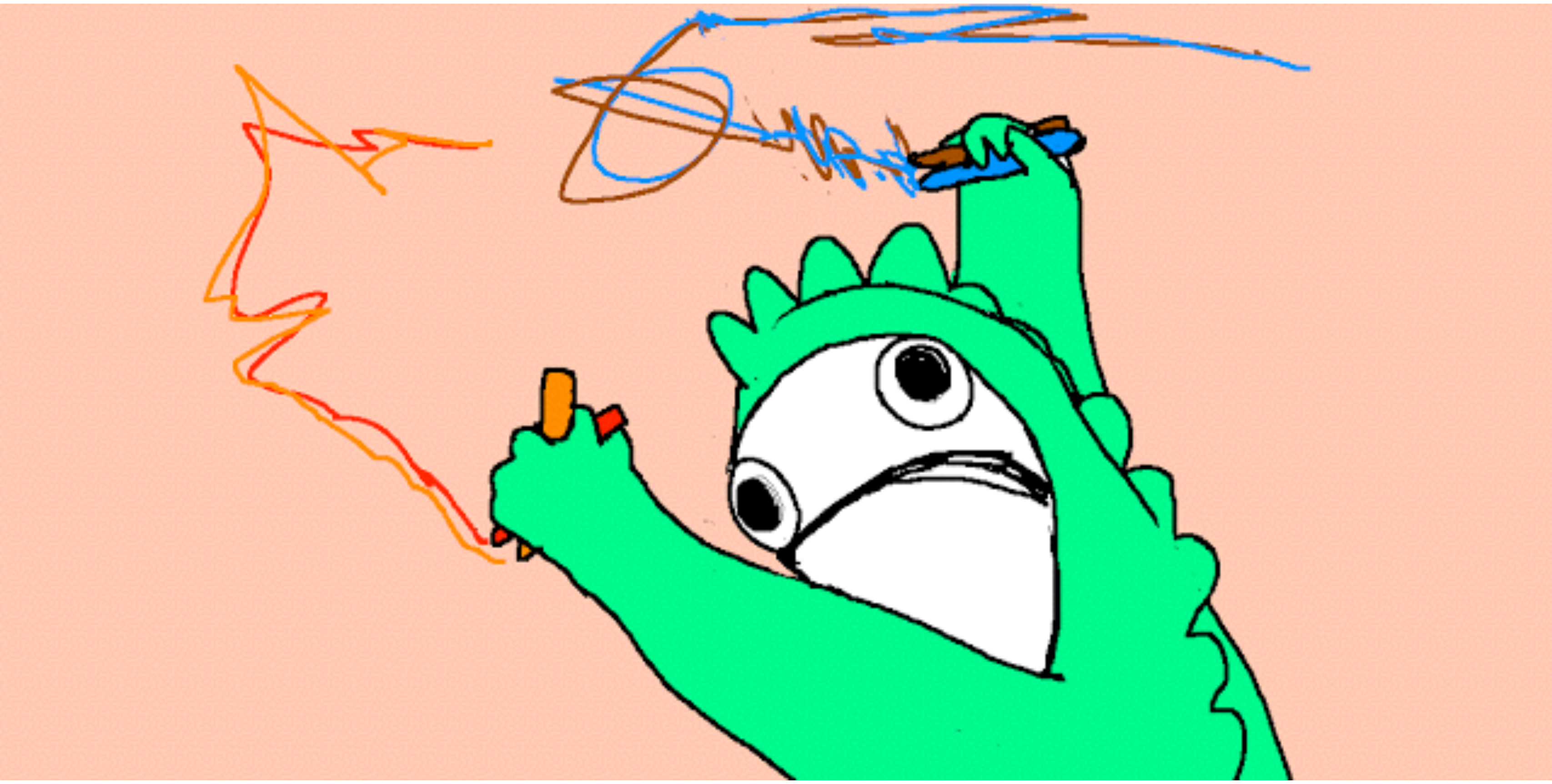
```
> table( expt$treatment, expt$gender )

          male female
  control    2      2
  drug1      2      2
  drug2      2      2
```

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
  19       0     0     0
  21       0     0     0
  23       0     1     0
  24       0     1     0
  25       1     0     1
  26       0     0     0
  28       1     0     1
  29       0     0     0
  30       0     0     0

, ,  = female


    control drug1 drug2
  19       0     1     0
  21       0     0     1
  23       0     0     0
  24       0     0     0
  25       1     0     0
  26       1     0     0
  28       0     0     0
  29       0     1     0
  30       0     0     1
```

Adding a third variable gives a three way cross-tabulation

Try it yourself (Exercise 2.1.3)



table() versus xtabs()

```
> table( expt$treatment, expt$gender )

          male female
  control    2      2
  drug1      2      2
  drug2      2      2
```

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

table() versus xtabs()

```
> table( expt$treatment, expt$gender )

          male female
  control    2      2
  drug1      2      2
  drug2      2      2
```

```
> xtabs(
     formula = ~ treatment + gender,
     data = expt
  )

          gender
treatment male female
  control    2      2
  drug1      2      2
  drug2      2      2
```

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 cross-tabulated

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

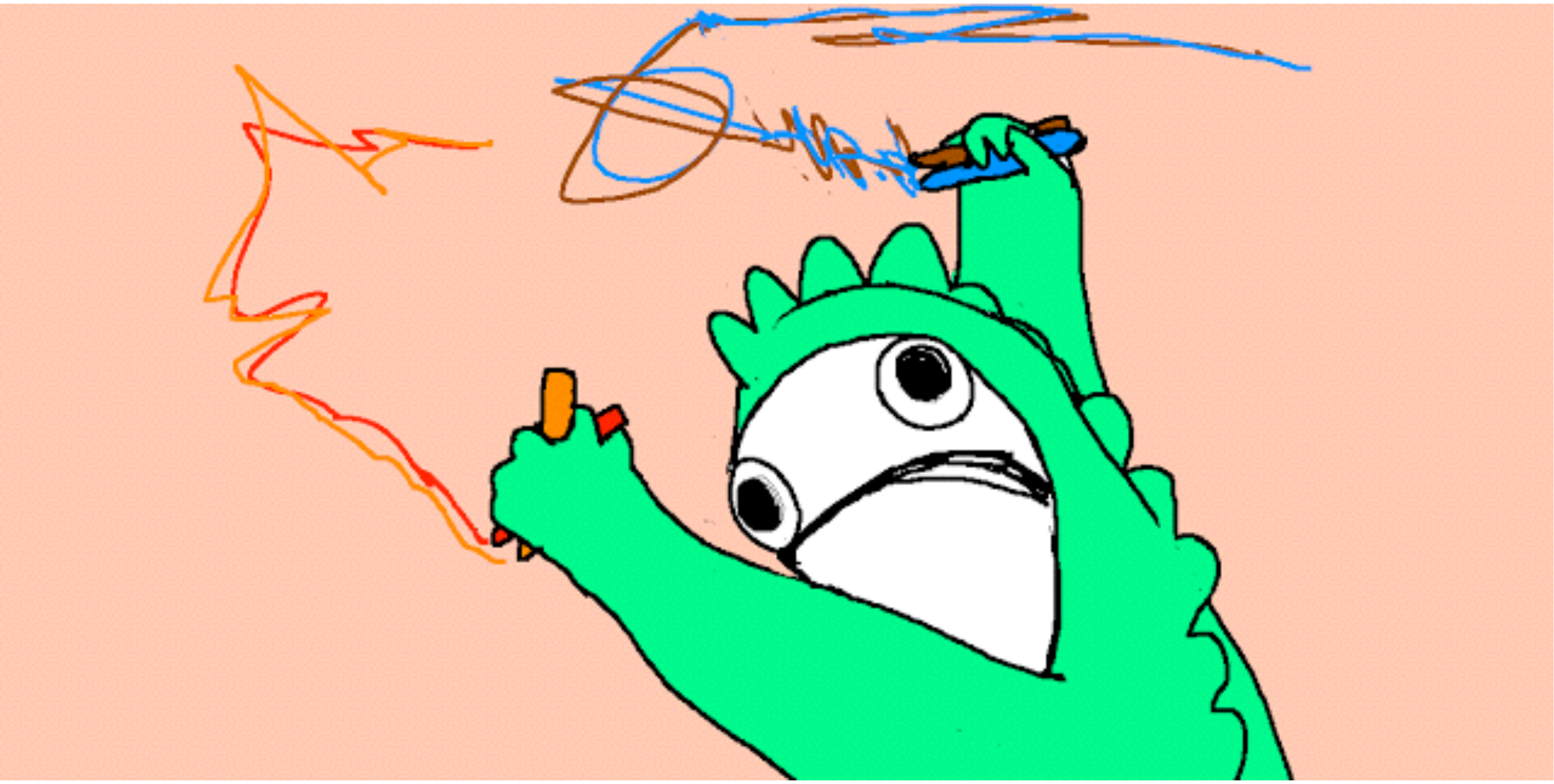
- In xtabs(), a "one-sided" formula is used to specify a set of variables to cross tabulate...

```
~ variable1 + variable2

~ variable1
```

- In lm(), 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()

Describe

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

           var  n  mean    sd median trimmed   mad   min   max range  skew kurtosis   se
id           1 12  6.50  3.61   6.50    6.50  4.45  1.00 12.00 11.00  0.00    -1.50 1.04
age          2 12 25.25  3.25  25.00   25.40  3.71 19.00 30.00 11.00 -0.32    -0.98 0.94
gender*      3 12  1.50  0.52   1.50    1.50  0.74  1.00  2.00  1.00  0.00    -2.16 0.15
treatment*   4 12  2.00  0.85   2.00    2.00  1.48  1.00  3.00  2.00  0.00    -1.74 0.25
hormone      5 12 43.59 24.86  42.15   41.80 23.28  6.70 98.40 91.70  0.51    -0.39 7.18
happy        6 12  3.71  1.02   3.42    3.69  0.84  2.00  5.69  3.69  0.28    -0.88 0.29
sad          7 12  3.42  1.60   3.23    3.46  1.91  0.34  6.12  5.78 -0.16    -0.96 0.46
```

Describe

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

	var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
id	1	12	6.50	3.61	6.50	6.50	4.45	1.00	12.00	11.00	0.00	-1.50	1.04
age	2	12	25.25	3.25	25.00	25.40	3.71	19.00	30.00	11.00	-0.32	-0.98	0.94
gender*	3	12	1.50	0.52	1.50	1.50	0.74	1.00	2.00	1.00	0.00	-2.16	0.15
treatment*	4	12	2.00	0.85	2.00	2.00	1.48	1.00	3.00	2.00	0.00	-1.74	0.25
hormone	5	12	43.59	24.86	42.15	41.80	23.28	6.70	98.40	91.70	0.51	-0.39	7.18
happy	6	12	3.71	1.02	3.42	3.69	0.84	2.00	5.69	3.69	0.28	-0.88	0.29
sad	7	12	3.42	1.60	3.23	3.46	1.91	0.34	6.12	5.78	-0.16	-0.96	0.46

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

Describe

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

	var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
id	1	12	6.50	3.61	6.50	6.50	4.45	1.00	12.00	11.00	0.00	-1.50	1.04
age	2	12	25.25	3.25	25.00	25.40	3.71	19.00	30.00	11.00	-0.32	-0.98	0.94
gender*	3	12	1.50	0.52	1.50	1.50	0.74	1.00	2.00	1.00	0.00	-2.16	0.15
treatment*	4	12	2.00	0.85	2.00	2.00	1.48	1.00	3.00	2.00	0.00	-1.74	0.25
hormone	5	12	43.59	24.86	42.15	41.80	23.28	6.70	98.40	91.70	0.51	-0.39	7.18
happy	6	12	3.71	1.02	3.42	3.69	0.84	2.00	5.69	3.69	0.28	-0.88	0.29
sad	7	12	3.42	1.60	3.23	3.46	1.91	0.34	6.12	5.78	-0.16	-0.96	0.46

Number of observations

Describe

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

	var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
id	1	12	6.50	3.61	6.50	6.50	4.45	1.00	12.00	11.00	0.00	-1.50	1.04
age	2	12	25.25	3.25	25.00	25.40	3.71	19.00	30.00	11.00	-0.32	-0.98	0.94
gender*	3	12	1.50	0.52	1.50	1.50	0.74	1.00	2.00	1.00	0.00	-2.16	0.15
treatment*	4	12	2.00	0.85	2.00	2.00	1.48	1.00	3.00	2.00	0.00	-1.74	0.25
hormone	5	12	43.59	24.86	42.15	41.80	23.28	6.70	98.40	91.70	0.51	-0.39	7.18
happy	6	12	3.71	1.02	3.42	3.69	0.84	2.00	5.69	3.69	0.28	-0.88	0.29
sad	7	12	3.42	1.60	3.23	3.46	1.91	0.34	6.12	5.78	-0.16	-0.96	0.46

Mean, standard deviation and median

Describe

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

	var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
id	1	12	6.50	3.61	6.50	6.50	4.45	1.00	12.00	11.00	0.00	-1.50	1.04
age	2	12	25.25	3.25	25.00	25.40	3.71	19.00	30.00	11.00	-0.32	-0.98	0.94
gender*	3	12	1.50	0.52	1.50	1.50	0.74	1.00	2.00	1.00	0.00	-2.16	0.15
treatment*	4	12	2.00	0.85	2.00	2.00	1.48	1.00	3.00	2.00	0.00	-1.74	0.25
hormone	5	12	43.59	24.86	42.15	41.80	23.28	6.70	98.40	91.70	0.51	-0.39	7.18
happy	6	12	3.71	1.02	3.42	3.69	0.84	2.00	5.69	3.69	0.28	-0.88	0.29
sad	7	12	3.42	1.60	3.23	3.46	1.91	0.34	6.12	5.78	-0.16	-0.96	0.46

10% trimmed mean

Describe

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

	var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
id	1	12	6.50	3.61	6.50	6.50	4.45	1.00	12.00	11.00	0.00	-1.50	1.04
age	2	12	25.25	3.25	25.00	25.40	3.71	19.00	30.00	11.00	-0.32	-0.98	0.94
gender*	3	12	1.50	0.52	1.50	1.50	0.74	1.00	2.00	1.00	0.00	-2.16	0.15
treatment*	4	12	2.00	0.85	2.00	2.00	1.48	1.00	3.00	2.00	0.00	-1.74	0.25
hormone	5	12	43.59	24.86	42.15	41.80	23.28	6.70	98.40	91.70	0.51	-0.39	7.18
happy	6	12	3.71	1.02	3.42	3.69	0.84	2.00	5.69	3.69	0.28	-0.88	0.29
sad	7	12	3.42	1.60	3.23	3.46	1.91	0.34	6.12	5.78	-0.16	-0.96	0.46

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

Describe

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

           var  n  mean    sd median trimmed   mad   min   max range skew kurtosis   se
id           1 12  6.50  3.61   6.50    6.50  4.45  1.00 12.00 11.00 0.00    -1.50 1.04
age          2 12 25.25  3.25  25.00   25.40  3.71 19.00 30.00 11.00 -0.32   -0.98 0.94
gender*      3 12  1.50  0.52   1.50    1.50  0.74  1.00  2.00  1.00 0.00    -2.16 0.15
treatment*   4 12  2.00  0.85   2.00    2.00  1.48  1.00  3.00  2.00 0.00    -1.74 0.25
hormone      5 12 43.59 24.86  42.15   41.80 23.28  6.70 98.40 91.70 0.51    -0.39 7.18
happy        6 12  3.71  1.02   3.42    3.69  0.84  2.00  5.69  3.69 0.28    -0.88 0.29
sad          7 12  3.42  1.60   3.23    3.46  1.91  0.34  6.12  5.78 -0.16   -0.96 0.46
```

Information about the range

Describe

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

           var  n  mean    sd median trimmed   mad   min   max range  skew kurtosis   se
id           1 12  6.50  3.61   6.50    6.50  4.45  1.00 12.00 11.00  0.00    -1.50 1.04
age          2 12 25.25  3.25  25.00   25.40  3.71 19.00 30.00 11.00 -0.32    -0.98 0.94
gender*      3 12  1.50  0.52   1.50    1.50  0.74  1.00  2.00  1.00  0.00    -2.16 0.15
treatment*   4 12  2.00  0.85   2.00    2.00  1.48  1.00  3.00  2.00  0.00    -1.74 0.25
hormone      5 12 43.59 24.86  42.15   41.80 23.28  6.70 98.40 91.70  0.51    -0.39 7.18
happy        6 12  3.71  1.02   3.42    3.69  0.84  2.00  5.69  3.69  0.28    -0.88 0.29
sad          7 12  3.42  1.60   3.23    3.46  1.91  0.34  6.12  5.78 -0.16    -0.96 0.46
```

Skew and kurtosis

Describe

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

	var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
id	1	12	6.50	3.61	6.50	6.50	4.45	1.00	12.00	11.00	0.00	-1.50	1.04
age	2	12	25.25	3.25	25.00	25.40	3.71	19.00	30.00	11.00	-0.32	-0.98	0.94
gender*	3	12	1.50	0.52	1.50	1.50	0.74	1.00	2.00	1.00	0.00	-2.16	0.15
treatment*	4	12	2.00	0.85	2.00	2.00	1.48	1.00	3.00	2.00	0.00	-1.74	0.25
hormone	5	12	43.59	24.86	42.15	41.80	23.28	6.70	98.40	91.70	0.51	-0.39	7.18
happy	6	12	3.71	1.02	3.42	3.69	0.84	2.00	5.69	3.69	0.28	-0.88	0.29
sad	7	12	3.42	1.60	3.23	3.46	1.91	0.34	6.12	5.78	-0.16	-0.96	0.46

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

Summary

```
> summary( expt )

       id             age            gender    treatment
 Min.   : 1.00   Min.   :19.00   male  :6   control:4
 1st Qu.: 3.75   1st Qu.:23.75   female:6   drug1  :4
 Median : 6.50   Median :25.00              drug2  :4
 Mean   : 6.50   Mean   :25.25
 3rd Qu.: 9.25   3rd Qu.:28.00
 Max.   :12.00   Max.   :30.00

    hormone          happy            sad
 Min.   : 6.70   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
 Mean   :43.59   Mean   :3.712   Mean   :3.423
 3rd Qu.:55.88   3rd Qu.:4.522   3rd Qu.:4.633
 Max.   :98.40   Max.   :5.690   Max.   :6.120
```

Summary

```
> summary( expt )

       id             age             gender     treatment
 Min.   : 1.00   Min.   :19.00   male  :6   control:4
 1st Qu.: 3.75   1st Qu.:23.75   female:6   drug1  :4
 Median : 6.50   Median :25.00              drug2  :4
 Mean   : 6.50   Mean   :25.25
 3rd Qu.: 9.25   3rd Qu.:28.00
 Max.   :12.00   Max.   :30.00

    hormone          happy            sad
 Min.   : 6.70   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
 Mean   :43.59   Mean   :3.712   Mean   :3.423
 3rd Qu.:55.88   3rd Qu.:4.522   3rd Qu.:4.633
 Max.   :98.40   Max.   :5.690   Max.   :6.120
```

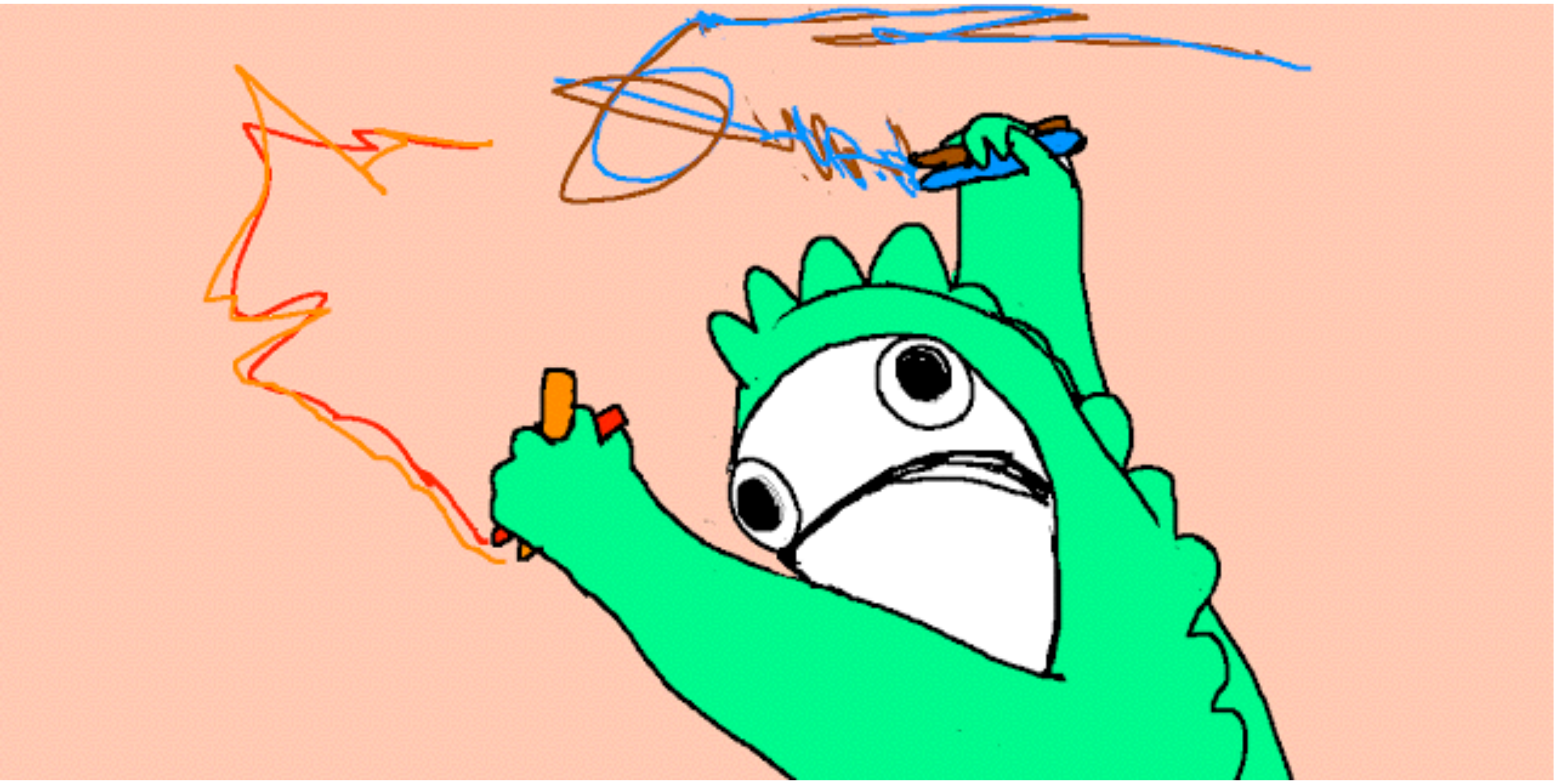
Summary produces a frequency table for the factor variables

Summary

```
> summary( expt )
       id             age          gender     treatment
 Min.   : 1.00   Min.   :19.00   male  :6   control:4
 1st Qu.: 3.75   1st Qu.:23.75   female:6   drug1  :4
 Median : 6.50   Median :25.00              drug2  :4
 Mean   : 6.50   Mean   :25.25
 3rd Qu.: 9.25   3rd Qu.:28.00
 Max.   :12.00   Max.   :30.00

    hormone          happy            sad
 Min.   : 6.70   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
 Mean   :43.59   Mean   :3.712   Mean   :3.423
 3rd Qu.:55.88   3rd Qu.:4.522   3rd Qu.:4.633
 Max.   :98.40   Max.   :5.690   Max.   :6.120
```

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 variable used to define the groups

describeBy()

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

group: male
           var n  mean    sd median trimmed   mad   min   max range  skew kurtosis    se
id           1 6  3.50  1.87   3.50    3.50  2.22  1.00  6.00  5.00  0.00    -1.80  0.76
age          2 6 25.50  2.07  25.00   25.50  2.22 23.00 28.00  5.00  0.22    -1.90  0.85
gender*      3 6  1.00  0.00   1.00    1.00  0.00  1.00  1.00  0.00   NaN      NaN  0.00
treatment*   4 6  2.00  0.89   2.00    2.00  1.48  1.00  3.00  2.00  0.00    -1.96  0.37
hormone      5 6 38.55 32.02  31.75   38.55 16.38  6.70 98.40 91.70  0.88    -0.78 13.07
happy        6 6  3.65  1.30   3.38    3.65  1.24  2.00  5.69  3.69  0.32    -1.49  0.53
sad          7 6  3.98  1.97   4.53    3.98  0.96  0.34  6.12  5.78 -0.82    -0.81  0.80
---------------------------------------------------------------------------------------------
group: female
           var n  mean    sd median trimmed  mad   min   max range  skew kurtosis   se
id           1 6  9.50  1.87   9.50    9.50 2.22  7.00 12.00  5.00  0.00    -1.80 0.76
age          2 6 25.00  4.34  25.50   25.00 5.93 19.00 30.00 11.00 -0.18    -1.85 1.77
gender*      3 6  2.00  0.00   2.00    2.00 0.00  2.00  2.00  0.00   NaN      NaN 0.00
treatment*   4 6  2.00  0.89   2.00    2.00 1.48  1.00  3.00  2.00  0.00    -1.96 0.37
hormone      5 6 48.63 16.54  54.90   48.63 8.75 18.50 65.20 46.70 -0.82    -0.98 6.75
happy        6 6  3.77  0.76   3.67    3.77 0.99  2.83  4.78  1.95  0.12    -1.90 0.31
sad          7 6  2.87  1.03   2.67    2.87 0.52  1.87  4.82  2.95  0.94    -0.67 0.42
```

Using aggregate()

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

Using aggregate()

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

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

Using aggregate()

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

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

Using aggregate()

```
> 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

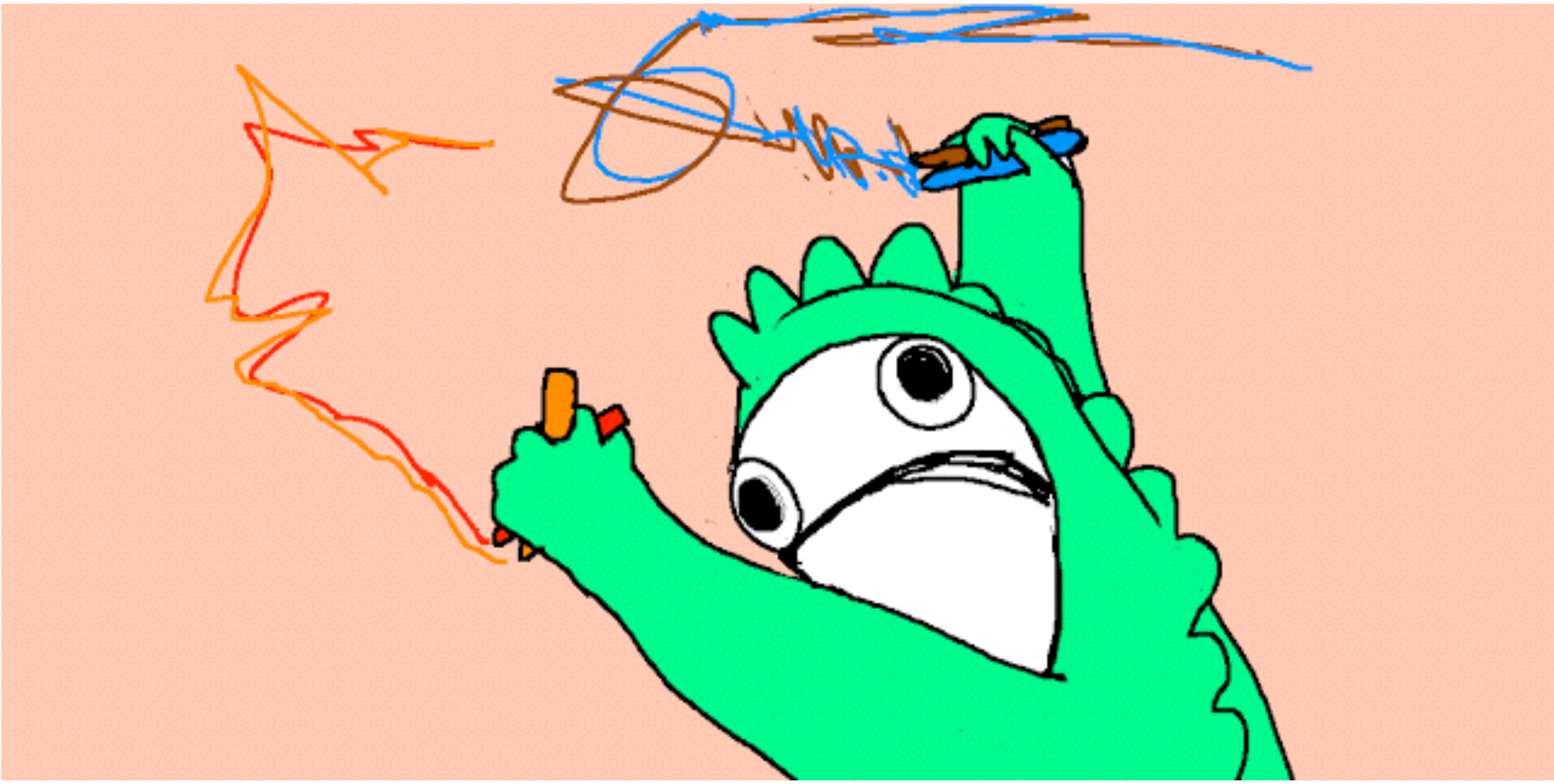
Using aggregate()

```
> 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$gender: male
       id            age            gender    treatment     hormone          happy            sad       
 Min.   :1.00   Min.   :23.00   male  :6   control:2   Min.   : 6.70   Min.   :2.000   Min.   :0.340  
 1st Qu.:2.25   1st Qu.:24.25   female:0   drug1  :2   1st Qu.:21.48   1st Qu.:3.007   1st Qu.:3.768  
 Median :3.50   Median :25.00              drug2  :2   Median :31.75   Median :3.380   Median :4.525  
 Mean   :3.50   Mean   :25.50                          Mean   :38.55   Mean   :3.650   Mean   :3.977  
 3rd Qu.:4.75   3rd Qu.:27.25                          3rd Qu.:41.42   3rd Qu.:4.270   3rd Qu.:4.758  
 Max.   :6.00   Max.   :28.00                          Max.   :98.40   Max.   :5.690   Max.   :6.120  
------------------------------------------------------------------------------------------- 
expt$gender: female
       id             age            gender    treatment     hormone          happy            sad       
 Min.   : 7.00   Min.   :19.00   male  :0   control:2   Min.   :18.50   Min.   :2.830   Min.   :1.870  
 1st Qu.: 8.25   1st Qu.:22.00   female:6   drug1  :2   1st Qu.:44.95   1st Qu.:3.248   1st Qu.:2.340  
 Median : 9.50   Median :25.50              drug2  :2   Median :54.90   Median :3.675   Median :2.675  
 Mean   : 9.50   Mean   :25.00                          Mean   :48.63   Mean   :3.775   Mean   :2.870  
 3rd Qu.:10.75   3rd Qu.:28.25                          3rd Qu.:56.23   3rd Qu.:4.357   3rd Qu.:2.882  
 Max.   :12.00   Max.   :30.00                          Max.   :65.20   Max.   :4.780   Max.   :4.820  
```

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 )
```

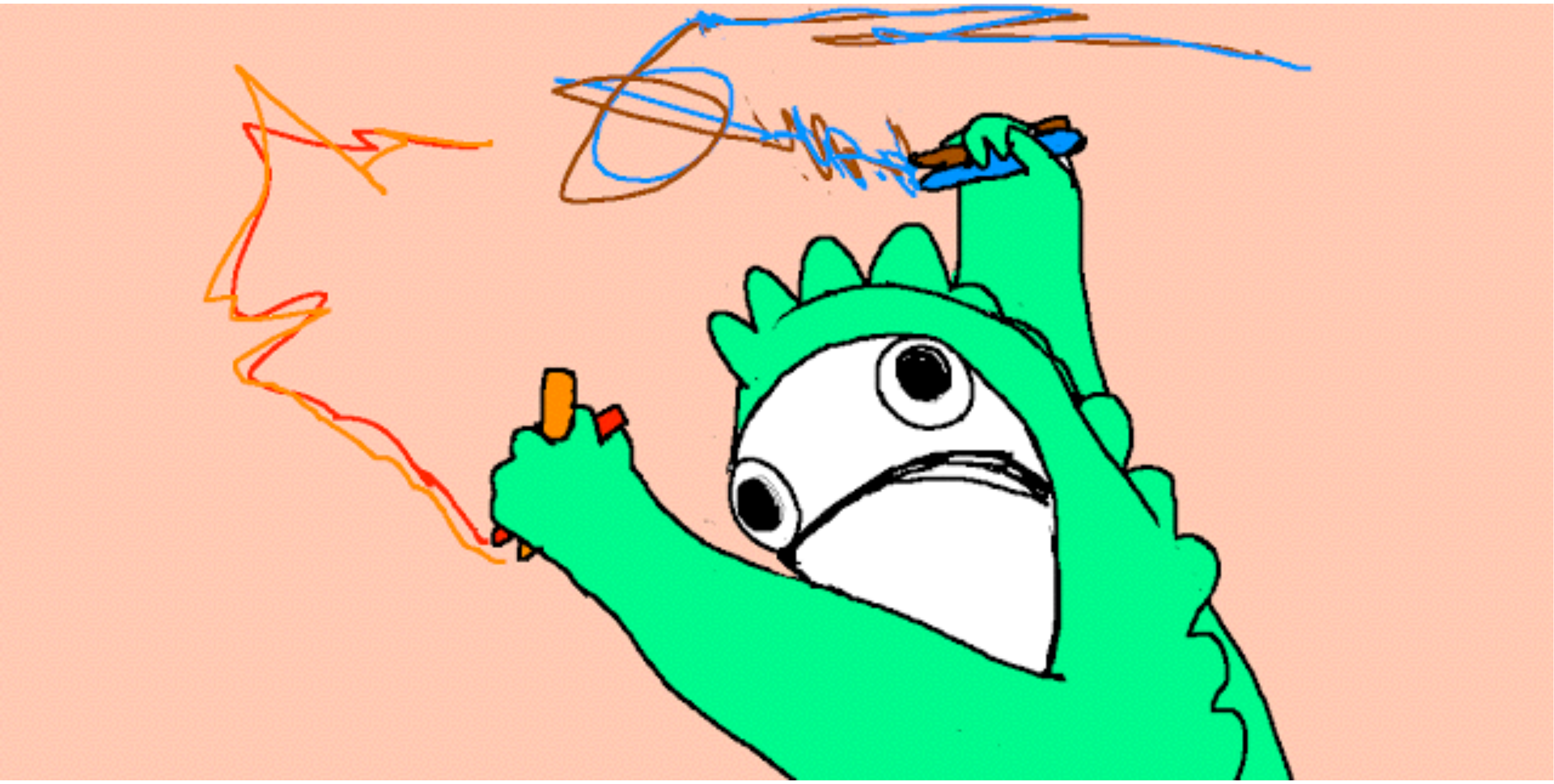
Just a reminder that you need to have the lsr package loaded for this command to work!

The correlate command itself

All pairwise correlations

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

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

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

Try it yourself (Exercise 2.1.7)



End of this section