

Introduction to R

(part 2)

Will Valdar
Oxford University

	A	B	C	D	E	F	G	H	I
1	sex	weight	country	age	snp.A	snp.B	family	phenotype	status
2	F	91.29	Europe	42	0	1	Family3	-2.6134576	FALSE
3	M	154.22	US	42	1	2	Family6	-7.7237116	FALSE
4	F	95.74	US	40	2	0	Family1	3.2361416	TRUE
5	F	110.25	Australia	30	0	1	Family10	-0.026201	TRUE
6	M	170.47	Australia	36	0	2	Family1	-11.132778	FALSE
7	M	127.57	US	45	1	1	Family6	-3.6016196	FALSE
8	M	125.56	Europe	44	1	1	Family9	-7.0376314	FALSE

data.frame

	A	B	C	D	E	F	G	H	I
1	sex	weight	country	age	snp.A	snp.B	family	phenotype	status
2	F	91.29	Europe	42	0	1	Family3	-2.6134576	FALSE
3	M	154.22	US	42	1	2	Family6	-7.7237116	FALSE
4	F	95.74	US	40	2	0	Family1	3.2361416	TRUE
5	F	110.25	Australia	30	0	1	Family10	-0.026201	TRUE
6	M	170.47	Australia	36	0	2	Family1	-11.132778	FALSE
7	M	127.57	US	45	1	1	Family6	-3.6016196	FALSE
8	M	125.56	Europe	44	1	1	Family9	-7.0376314	FALSE

data.frame

	A	B	C	D	E	F	G	H	I
1	sex	weight	country	age	snp.A	snp.B	family	phenotype	status
2	F	91.29	Europe	42	0	1	Family3	-2.6134576	FALSE
3	M	154.22	US	42	1	2	Family6	-7.7237116	FALSE
4	F	95.74	US	40	2	0	Family1	3.2361416	TRUE
5	F	110.25	Australia	30	0	1	Family10	-0.026201	TRUE
6	M	170.47	Australia	36	0	2	Family1	-11.132778	FALSE
7	M	127.57	US	45	1	1	Family6	-3.6016196	FALSE
8	M	125.56	Europe	44	1	1	Family9	-7.0376314	FALSE

```
setwd("F:valdar/ThursdayMorningR/")
```

```
data <- read.delim("disease.txt")
```

```
class(data)
```

```
head(data)
```

data.frame

	A	B	C	D	E	F	G	H	I
1	sex	weight	country	age	snp.A	snp.B	family	phenotype	status
2	F	91.29	Europe	42	0	1	Family3	-2.6134576	FALSE
3	M	154.22	US	42	1	2	Family6	-7.7237116	FALSE
4	F	95.74	US	40	2	0	Family1	3.2361416	TRUE
5	F	110.25	Australia	30	0	1	Family10	-0.026201	TRUE
6	M	170.47	Australia	36	0	2	Family1	-11.132778	FALSE
7	M	127.57	US	45	1	1	Family6	-3.6016196	FALSE
8	M	125.56	Europe	44	1	1	Family9	-7.0376314	FALSE

	sex	weight	country	age	snp.A	snp.B	family	phenotype	status
1	F	91.29	Europe	42	0	1	Family3	-2.61345764	FALSE
2	M	154.22	US	42	1	2	Family6	-7.72371165	FALSE
3	F	95.74	US	40	2	0	Family1	3.23614160	TRUE
4	F	110.25	Australia	30	0	1	Family10	-0.02620101	TRUE
5	M	170.47	Australia	36	0	2	Family1	-11.13277756	FALSE
6	M	127.57	US	45	1	1	Family6	-3.60161956	FALSE

a variable with other variables hanging off it

```
str(data)
```

```
'data.frame':  400 obs. of  9 variables:
 $ sex      : Factor w/ 2 levels "F","M": 1 2 1 1 2 ...
 $ weight   : num  91.3 154.2  95.7 110.2 170.5 ...
 $ country  : Factor w/ 3 levels "Australia","Europe",...: 2 3 ...
 $ age      : int  42 42 40 30 36 45 44 38 40 34 ...
 $ snp.A    : int  0 1 2 0 0 1 1 0 2 1 ...
 $ snp.B    : int  1 2 0 1 2 1 1 0 1 1 ...
 $ family   : Factor w/ 10 levels "Family1",...: 4 7 ...
 $ phenotype: num  -2.6135 -7.7237  3.2361 ...
 $ status   : logi  FALSE FALSE  TRUE  TRUE FALSE FALSE ...
```

a variable with other variables hanging off it

```
str(data)
```

```
'data.frame':  400 obs. of  9 variables:
 $ sex      : Factor w/ 2 levels "F","M": 1 2 1 1 2 ...
 $ weight   : num  91.3 154.2  95.7 110.2 170.5 ...
 $ country  : Factor w/ 3 levels "Australia","Europe",...: 2 3 ...
 $ age      : int  42 42 40 30 36 45 44 38 40 34 ...
 $ snp.A    : int  0 1 2 0 0 1 1 0 2 1 ...
 $ snp.B    : int  1 2 0 1 2 1 1 0 1 1 ...
 $ family   : Factor w/ 10 levels "Family1",...: 4 7 ...
 $ phenotype: num  -2.6135 -7.7237  3.2361 ...
 $ status   : logi  FALSE FALSE  TRUE  TRUE FALSE FALSE ...
```

```
data$weight
```

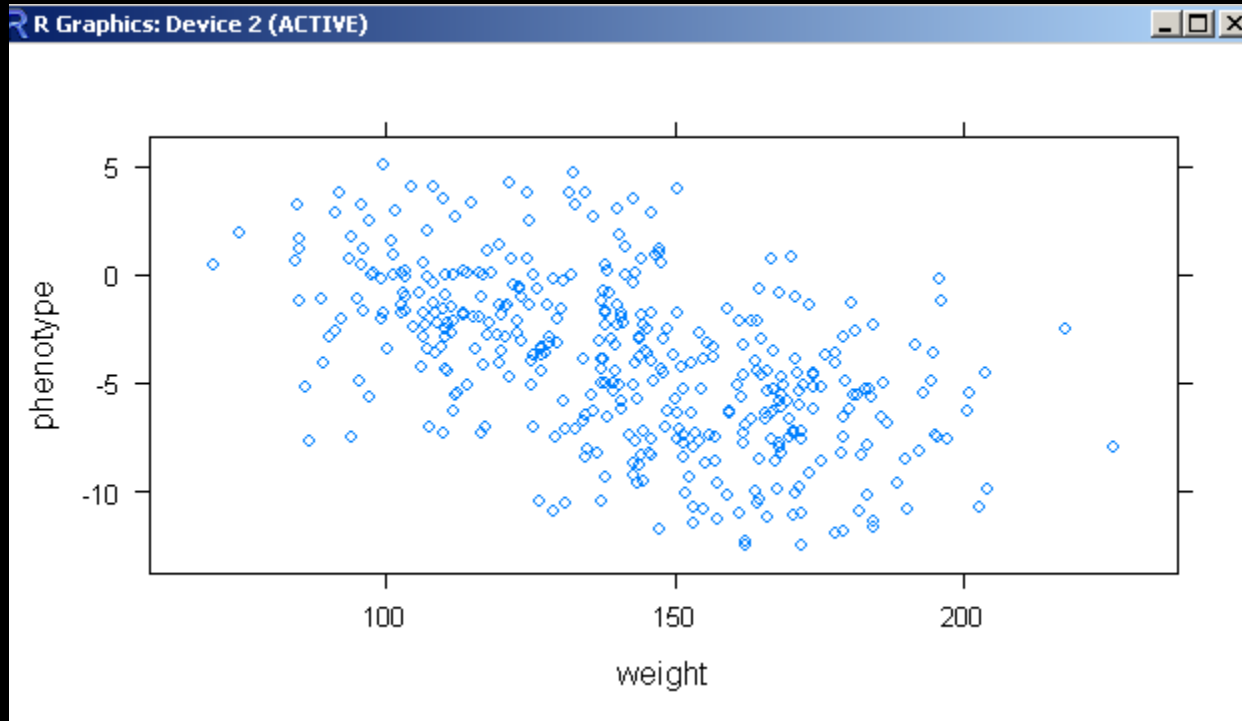
```
data$weight[3:10]
```

```
mean(data$weight)
```

```
library(lattice)
```

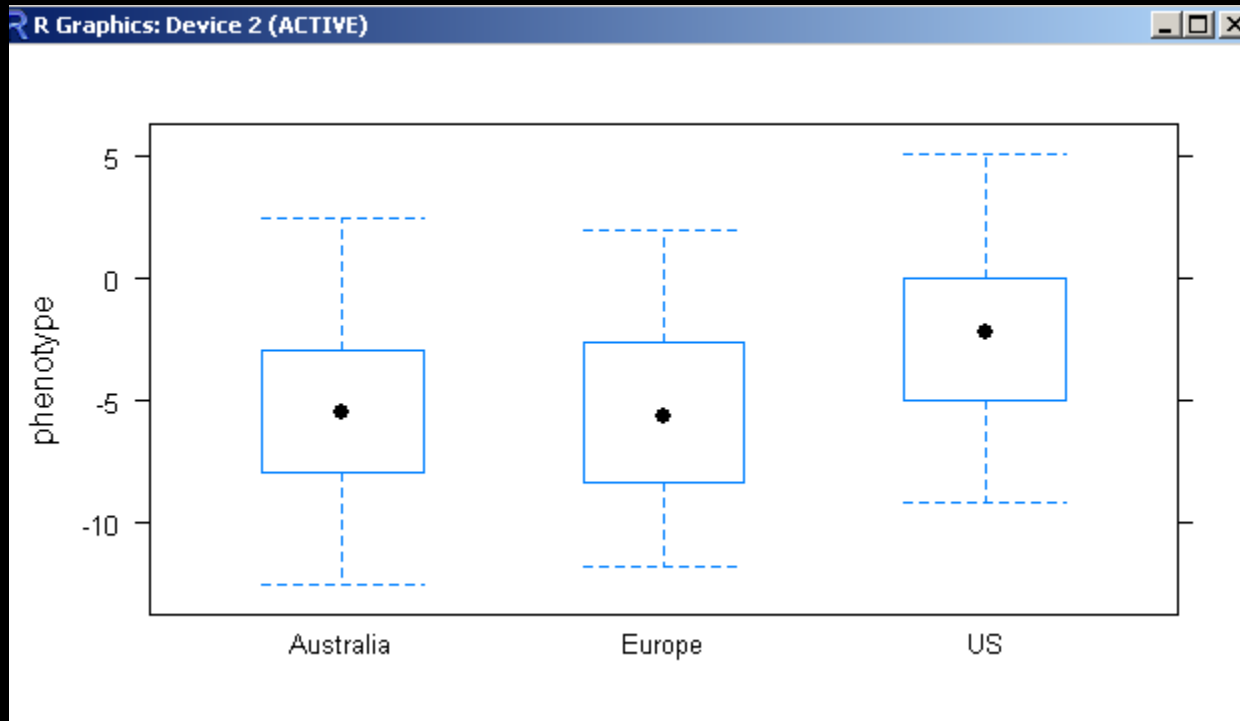
Scatter plots

```
xyp1ot( phenotype ~ weight, data=data)
```



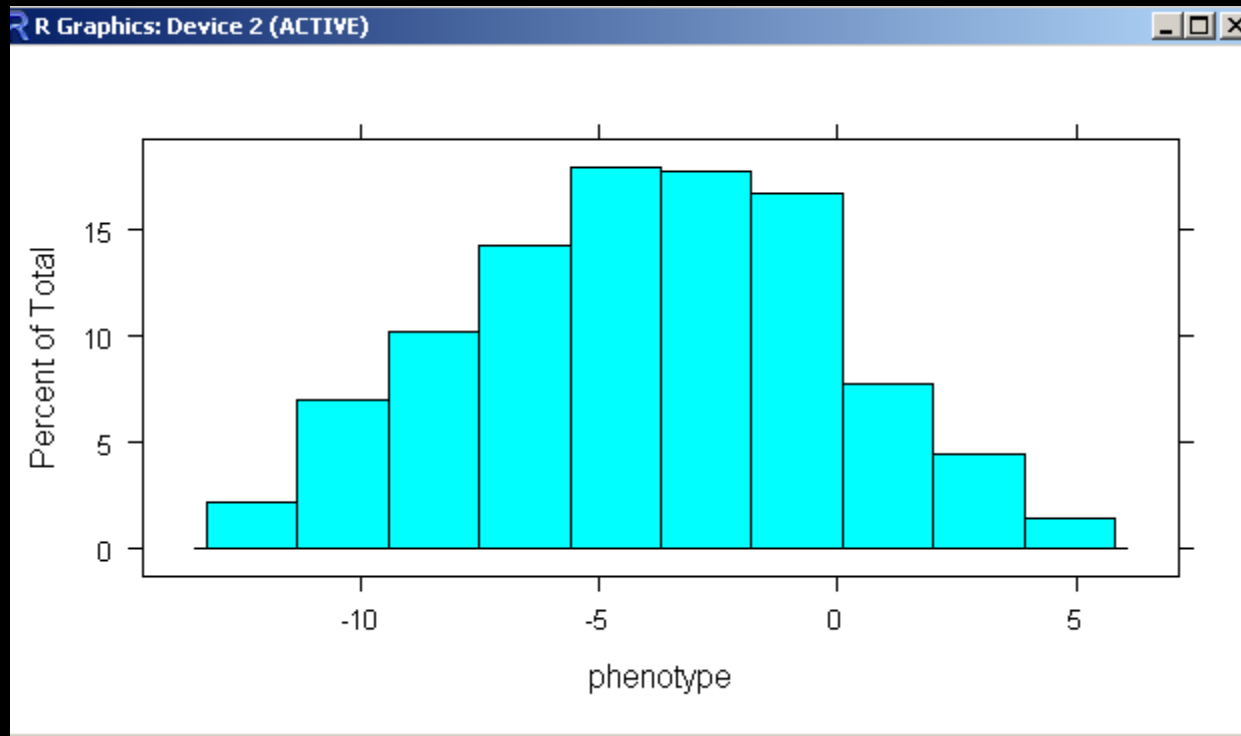
Box and whisker plot

```
bwplot( phenotype ~ country, data=data)
```



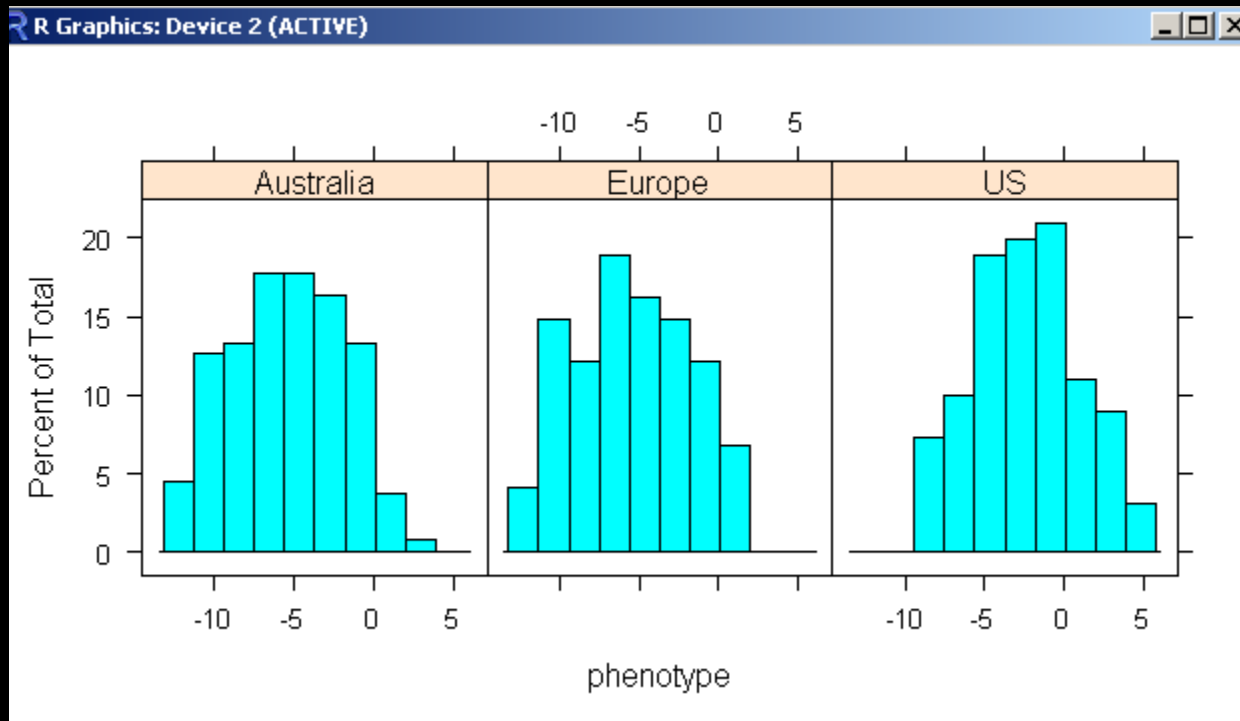
Histogram

```
histogram( ~ phenotype, data=data)
```



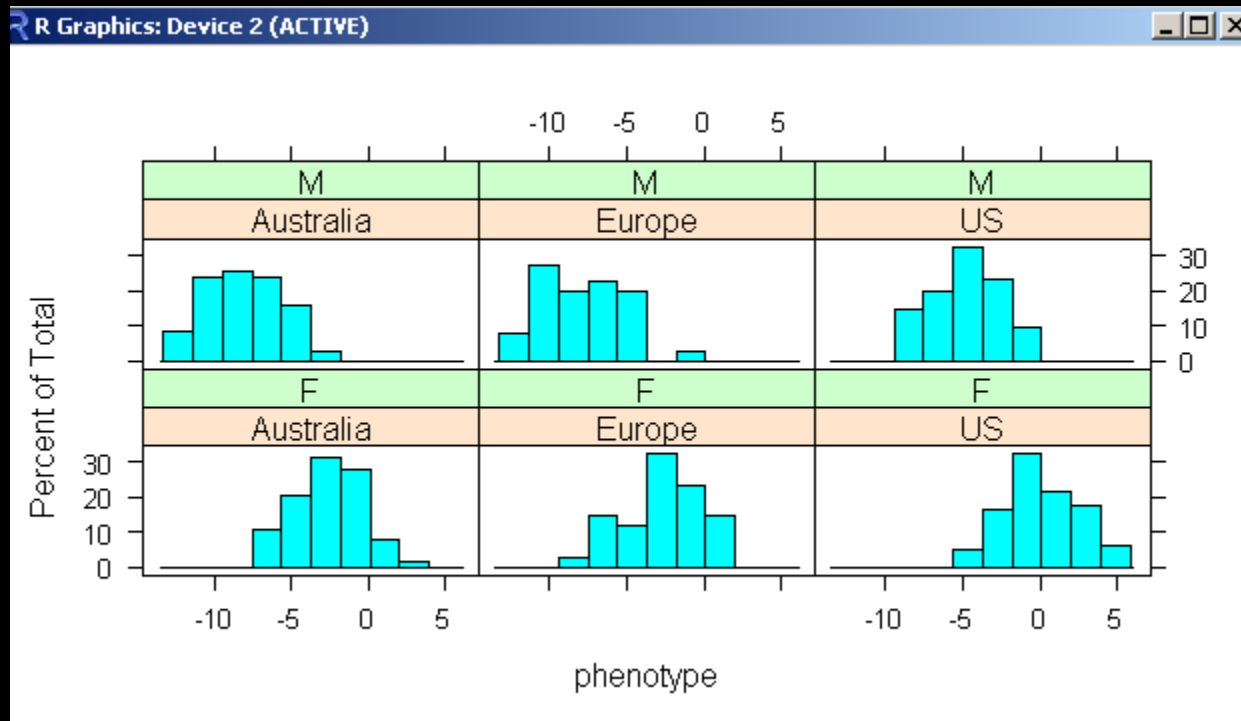
Conditioning

```
histogram( ~ phenotype | country, data=data)
```



Conditioning

```
histogram( ~ phenotype | sex * country, data=data)
```



Conditioning

Examples:

```
xyplot( phenotype ~ weight | country, data=data)
```

```
bwplot( phenotype ~ country | snp.A, data=data)
```

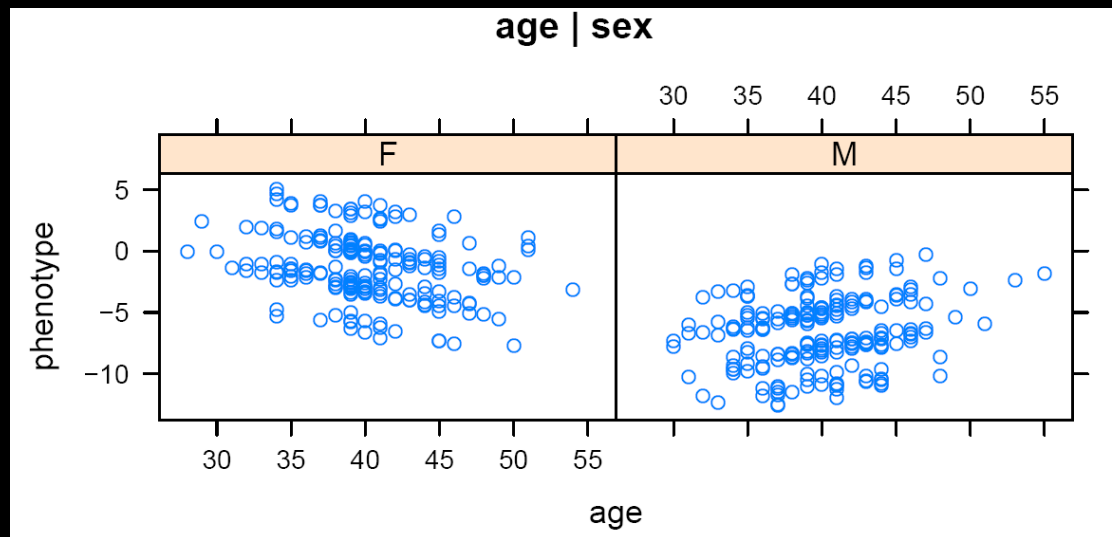
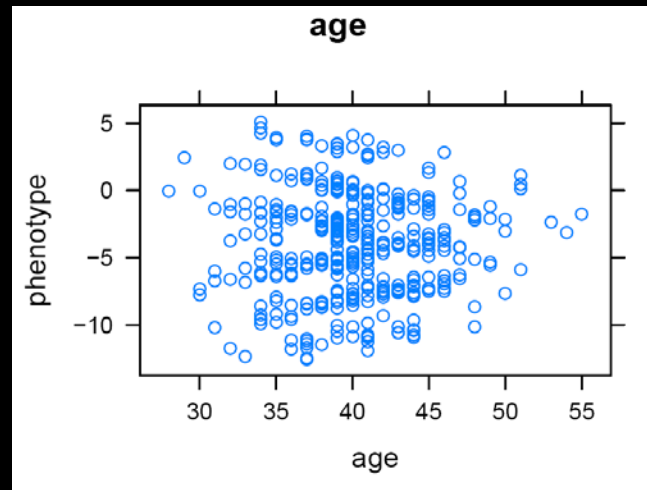
Exercises 2

- 1) What do you think the status variable means and how is it defined? (*conditional histogram*)
- 2) How does snp.B affect the phenotype? (*cond. hist*)
- 3) Is there an effect of age on the phenotype? Are you sure? (*conditional xyplot*)
- 4) Does weight affect the phenotype? Does sex? Does sex affect weight? Does weight affect sex? What's going on here? (*conditional xyplot*)

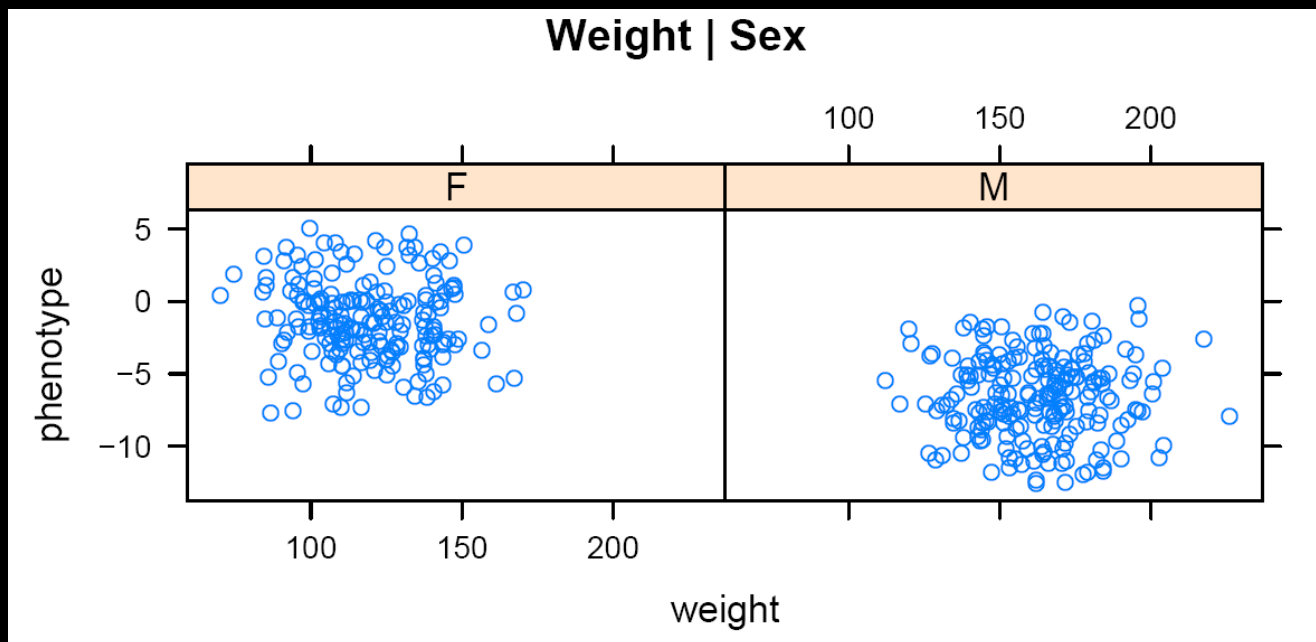
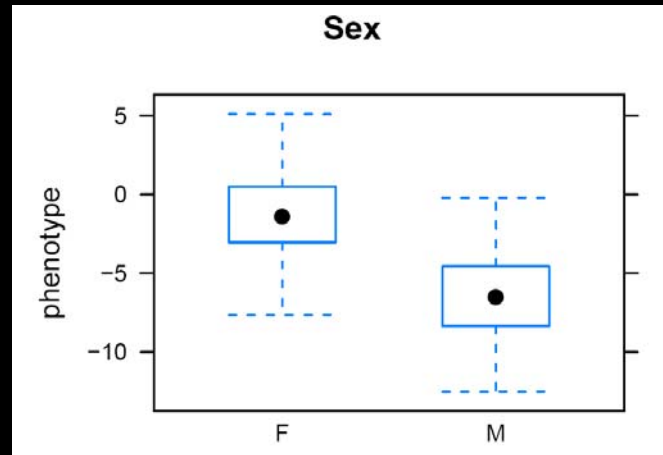
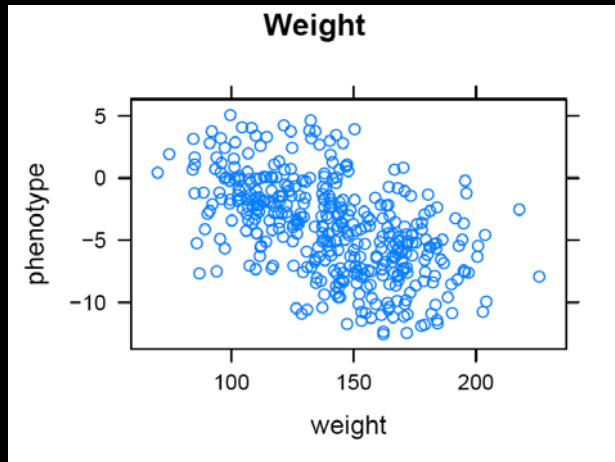
Answers 2

- 1) `histogram( ~ phenotype | status, data=data)`
- 2) `xyplot(phenotype ~ snp.B, data=data)`
`histogram( ~ phenotype | snp.B, data=data)`
- 3) `xyplot(phenotype ~ age, data=data)`
`xyplot(phenotype ~ age | sex, data=data)`
- 4) `bwplot(phenotype ~ sex, data=data)`
`xyplot(phenotype ~ weight, data=data)`
`xyplot(phenotype ~ weight | sex, data=data)`

Answers 2: question 3



Answers 2: question 4



Further reading

Books:

Introductory Statistics with R

by Peter Dalgaard

Modern Applied Statistics with S

by Venables and Ripley

Web sites:

<http://www.r-project.org/>

<http://wiki.r-project.org/rwiki/doku.php>

<http://news.gmane.org/gmane.comp.lang.r.general>

Graph gallery:

<http://addictedtor.free.fr/graphiques/>