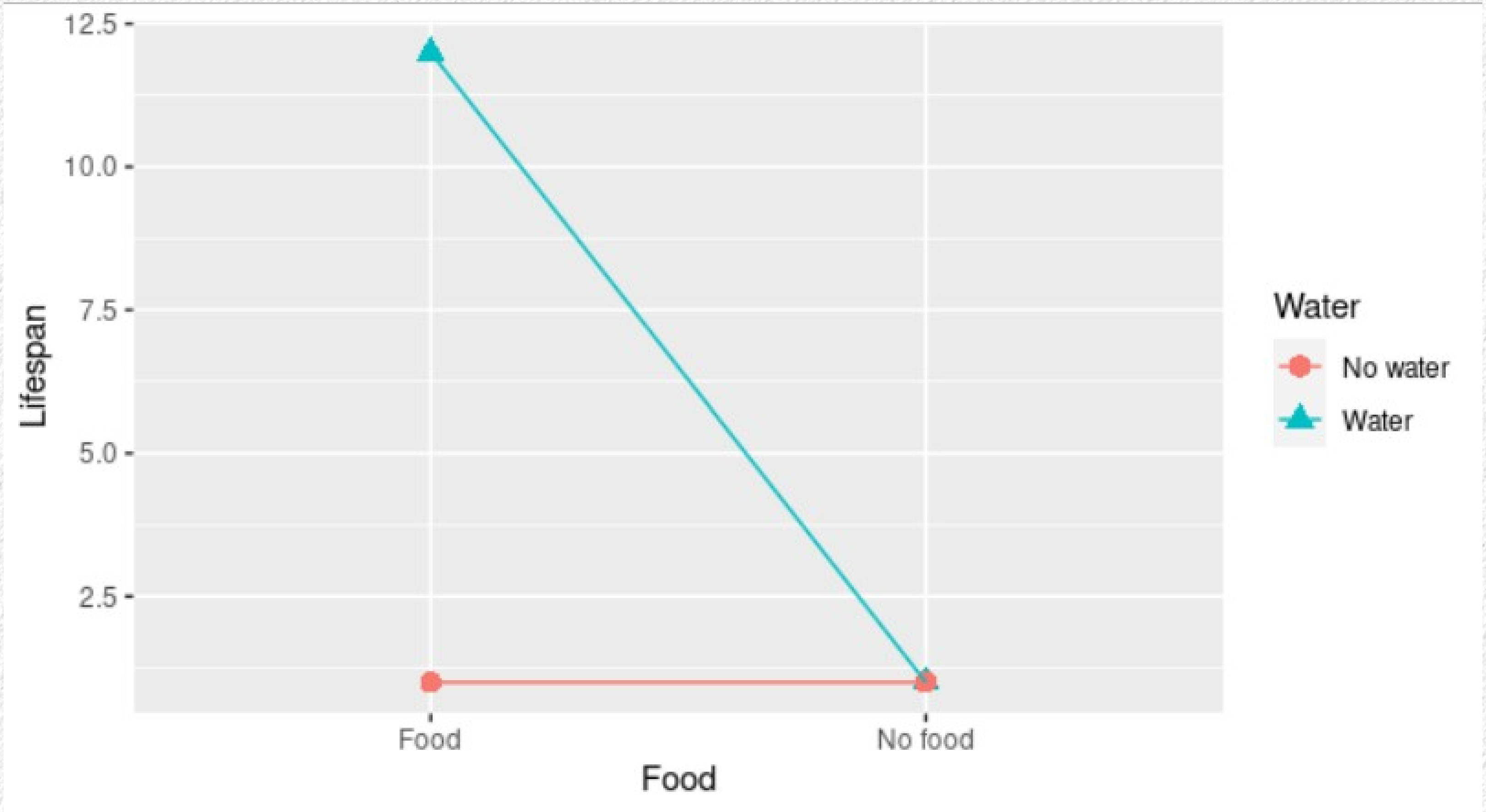
Factorial designs and interactions



Study of yield of wheat

	S1	S2	S3	S4
V1	3	3	3	3

- Expect that two different factors are important:
  - Sowing rate (density of seeds, 4 different levels)
  - Variety of wheat (2 varieties)
- The "ideal experiment" would isolate each of these factors, one at a time
- If you had 24 fields to do the experiment, you could use ½ to study the effect of sowing rate, and ½ to study differences between varieties

	S1
V1	6
V2	6

The “ideal experiment” has a less than ideal feature

- Ideal experiment hold everything constant, except for the one factor that is being tested
- Strictly applied, we would only study a single predictor at a time
- Problems:
  - Inefficient use of resources
  - Sometimes response to one predictor variable depends on another

The factorial principle

THE ARRANGEMENT OF FIELD EXPERIMENTS

R. A. FISHER, SC.D.,
Rothamsted Experimental Station.

Journal of the Ministry of Agriculture of Great Britain, 33: 503-513, (1926).

No aphorism is more frequently repeated in connection with field trials, than that we must ask Nature few questions, or, ideally, one question, at a time. The writer is convinced that this view is wholly mistaken. Nature, he suggests, will best respond to a logical and carefully thought out questionnaire; indeed, if we ask her a single question, she will often refuse to answer until some other topic has been discussed.

Two factors crossed: sowing rate and crop variety

Factorial approach

(4 x 2 factorial design)

	S1	S2	S3	S4
V1	3	3	3	3
V2	3	3	3	3

Study sowing rate and variety simultaneously

Doing this gives us “hidden replication”

3 replicates for each combination of S and V (24 total)

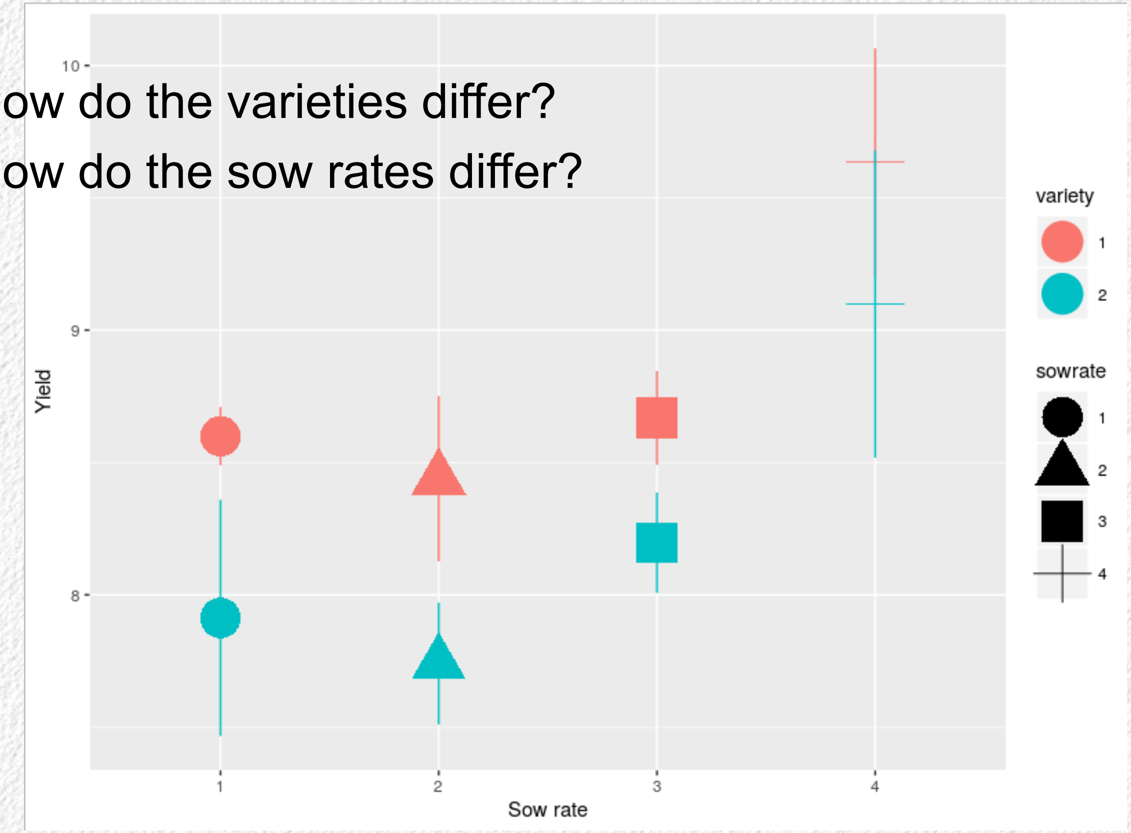
→ 6 replicates for each sowing rate (24 total),

→ 12 replicates for each variety (24 total)

This is like doubling our sample size!

Since the design is balanced and complete, we can measure effects of sow rate on yield independent of the effects of variety on yield (orthogonal design)

The data



How do the varieties differ?

How do the sow rates differ?

Do both varieties of wheat respond to sow rates the same way?

Analysis: one approach

Group	Yield	Group
S2V2	7.74	a
S1V2	7.91	a
S3V2	8.20	ab
S2V1	8.44	ab
S1V1	8.60	ab
S3V1	8.67	ab
S4V2	9.10	ab
S4V1	9.64	b

```
Response: YIELD
          Df Sum Sq Mean Sq F value  Pr(>F)
Group      7 8.0776  1.1539  3.2388 0.02443
Residuals 16 5.7006  0.3563
```

Each combination of S and V is a group, compare the means with one-way ANOVA

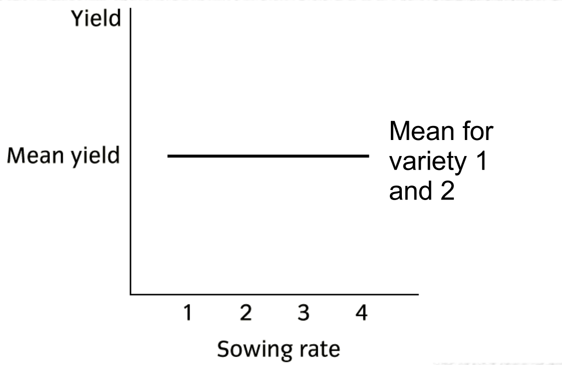
Problem: can we tell if S is more important than V, or if response to S depends on V?

Better analysis: factorial ANOVA

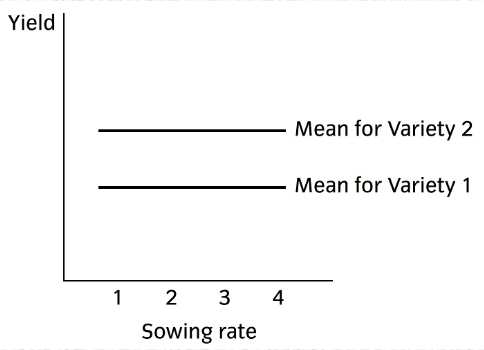
- Distinguish between **main effects** of the predictors and **interactions** between them
- Main effects – based on marginal means, one predictor at a time
  - Do sowing rates give different yields?
  - Do varieties have different yields?
- Interaction – does the effect of one predictor depend on the level of the other?
  - Does the difference between varieties change depending on the sow rate?

    ...or, equivalently...
  - Do sow rates have different effects depending on the variety?

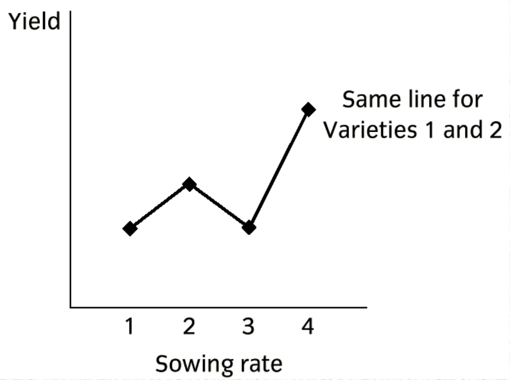
Sowing and variety experiment: possible results



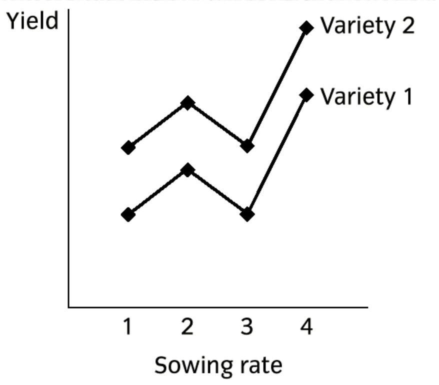
No effect of variety or sowing rate, no interaction



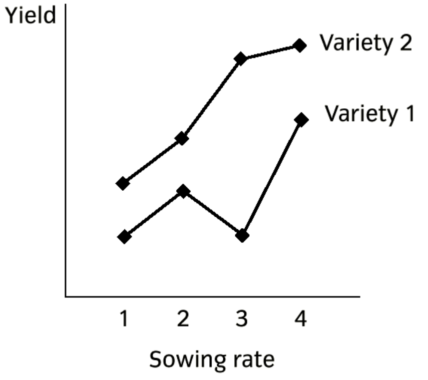
Main effect of variety, no effect of sowing rate, no interaction



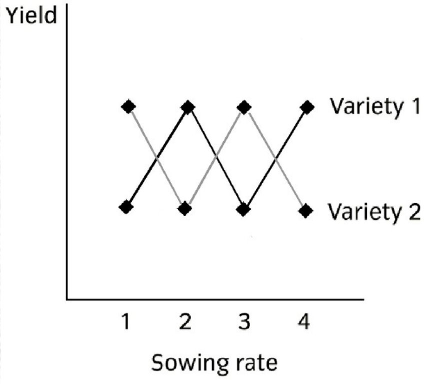
No effect of variety, main effect of sowing rate, no interaction



Both main effects, no interaction



Both main effects and interaction



Interaction, no main effects

All the means needed for the ANOVA

			Sow rate (c = 4)			
Variety (r = 2)	**1**	**2**	**3**	**4**	***Variety means***
1	8.82	7.83	8.34	8.82	
	8.49	8.64	8.95	9.82	8.84
	8.49	8.85	8.72	10.27	
Subgroup means	8.60	8.44	8.67	9.64	
2	8.53	7.76	8.53	8.99	
	8.17	7.33	8.19	10.15	8.24
	7.05	8.13	7.88	8.15	
Subgroup means	7.91	7.74	8.20	9.10	
Sow rate means	8.26	8.09	8.43	9.37	
Grand mean	8.54				

Individual data values, marginal means, and ***subgroup means***

Calculating each sums of squares

		Sow rate (c = 4)			
Variety (r = 2)	**1**	**2**	**3**	**4**	*Variety means*
1	8.82	7.83	8.34	8.82	
	8.49	8.64	8.95	9.82	8.84
	8.49	8.85	8.72	10.27	
Subgroup means	8.60	8.44	8.67	9.64	
2	8.53	7.76	8.53	8.99	
	8.17	7.33	8.19	10.15	8.24
	7.05	8.13	7.88	8.15	
Subgroup means	7.91	7.74	8.20	9.10	
Sow rate means	8.26	8.09	8.43	9.37	
Grand mean	8.54				

Subgroup SS	8.08
Within SS	5.70
Variety SS	2.15
Sow rate SS	5.87
Interaction SS	0.06

$$n\sum^{rc}\left(\bar{Y}-\bar{\bar{Y}}\right)^2$$

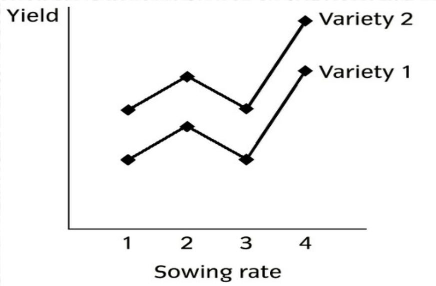
$$\sum^{rc}\sum^{n}\left(Y-\bar{Y}\right)^2$$

$$\sum^{r}cn\left(\bar{R}-\bar{\bar{Y}}\right)^2$$

$$\sum^{c}rn\left(\bar{C}-\bar{\bar{Y}}\right)^2$$

$$Subgroup\ SS-Variety\ SS-Sow\ rate\ SS$$

			Sow rate (c = 4)		
Variety (r = 2)	1	2	3	4	*Variety means*
1	8.82	7.83	8.34	8.82	
	8.49	8.64	8.95	9.82	8.84
	8.49	8.85	8.72	10.27	
Subgroup means	8.60	8.44	8.67	9.64	
2	8.53	7.76	8.53	8.99	
	8.17	7.33	8.19	10.15	8.24
	7.05	8.13	7.88	8.15	
Subgroup means	7.91	7.74	8.20	9.10	
Sow rate means	8.26	8.09	8.43	9.37	
Grand mean	8.54				

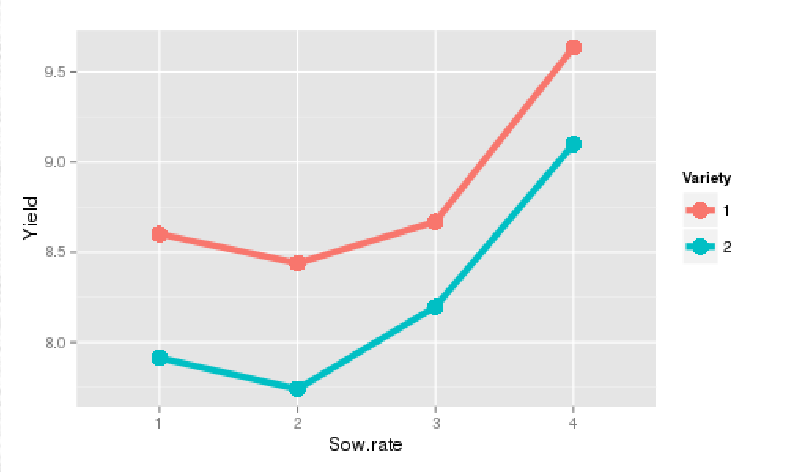


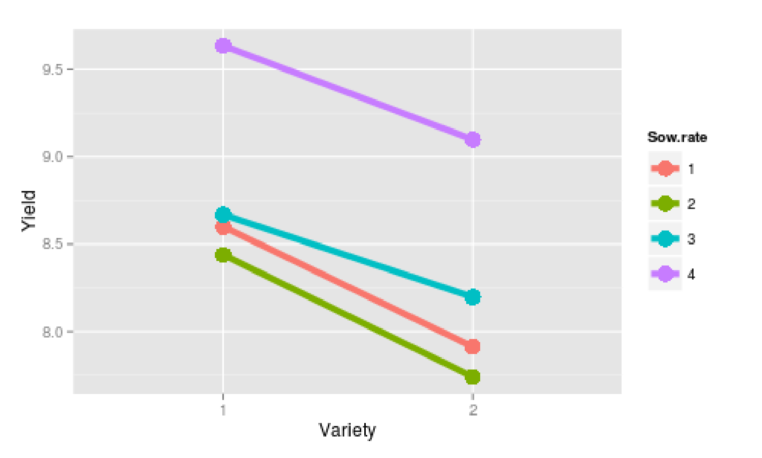
It was this one

Complete ANOVA table assembled

```
Analysis of Variance Table
Response: YIELD
                Df Sum Sq Mean Sq F value  Pr(>F)
VARIETY          1 2.1474 2.14742  6.0273 0.02591
SOWRATE          3 5.8736 1.95786  5.4952 0.00866
VARIETY:SOWRATE  3 0.0566 0.01887  0.0530 0.98334
Residuals       16 5.7006 0.35628
```

Only main effects are significant





Interaction plot

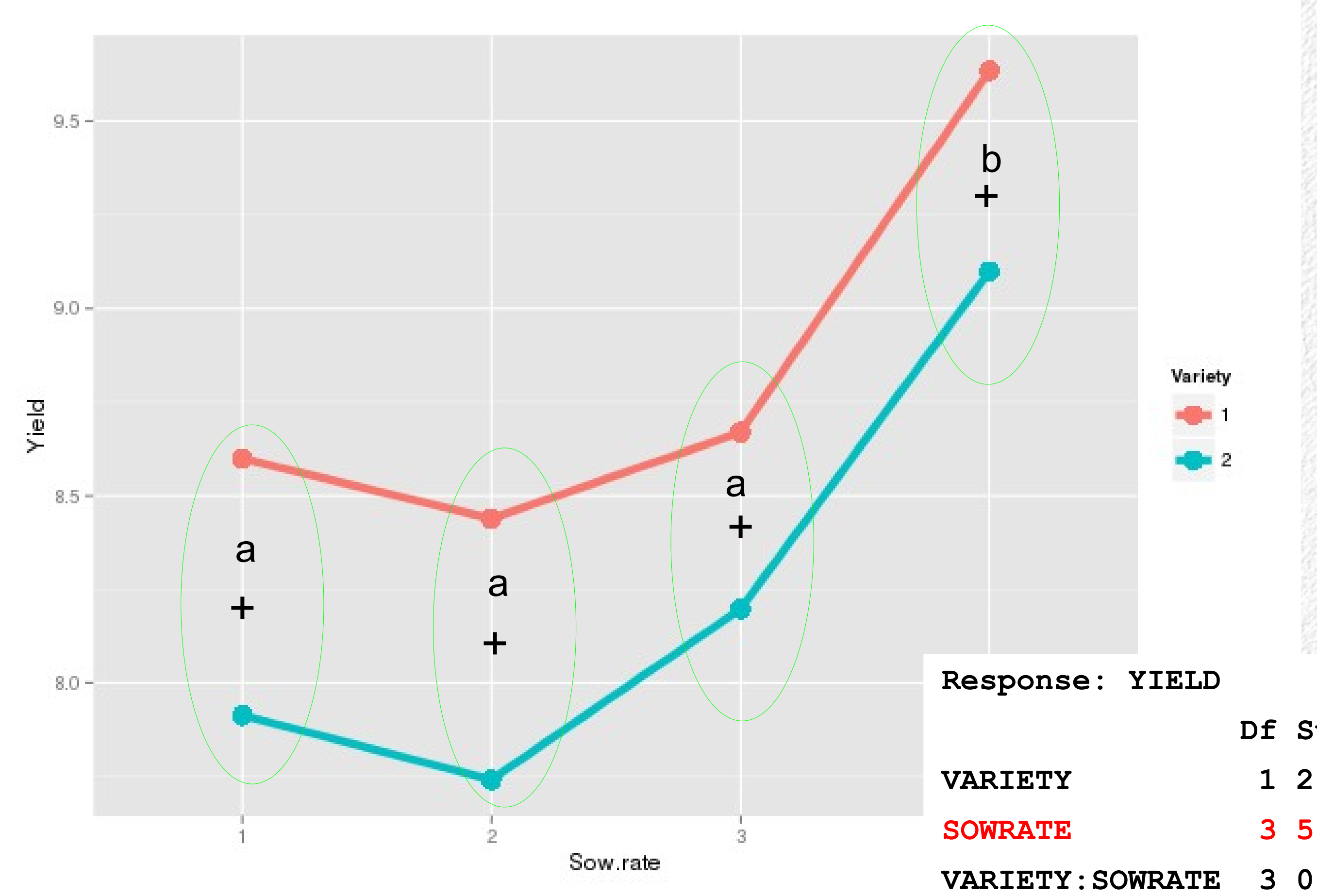
Standard approach to presenting a factorial design

- *Plot of the subgroup means (combinations of variety and sow rate)*
- *One of the predictors on the x-axis*
- *Other predictor indicated by color/plot symbol*

Parallel lines indicate no interaction

Lines that are not parallel indicate a possible interaction

Sow rate main effect

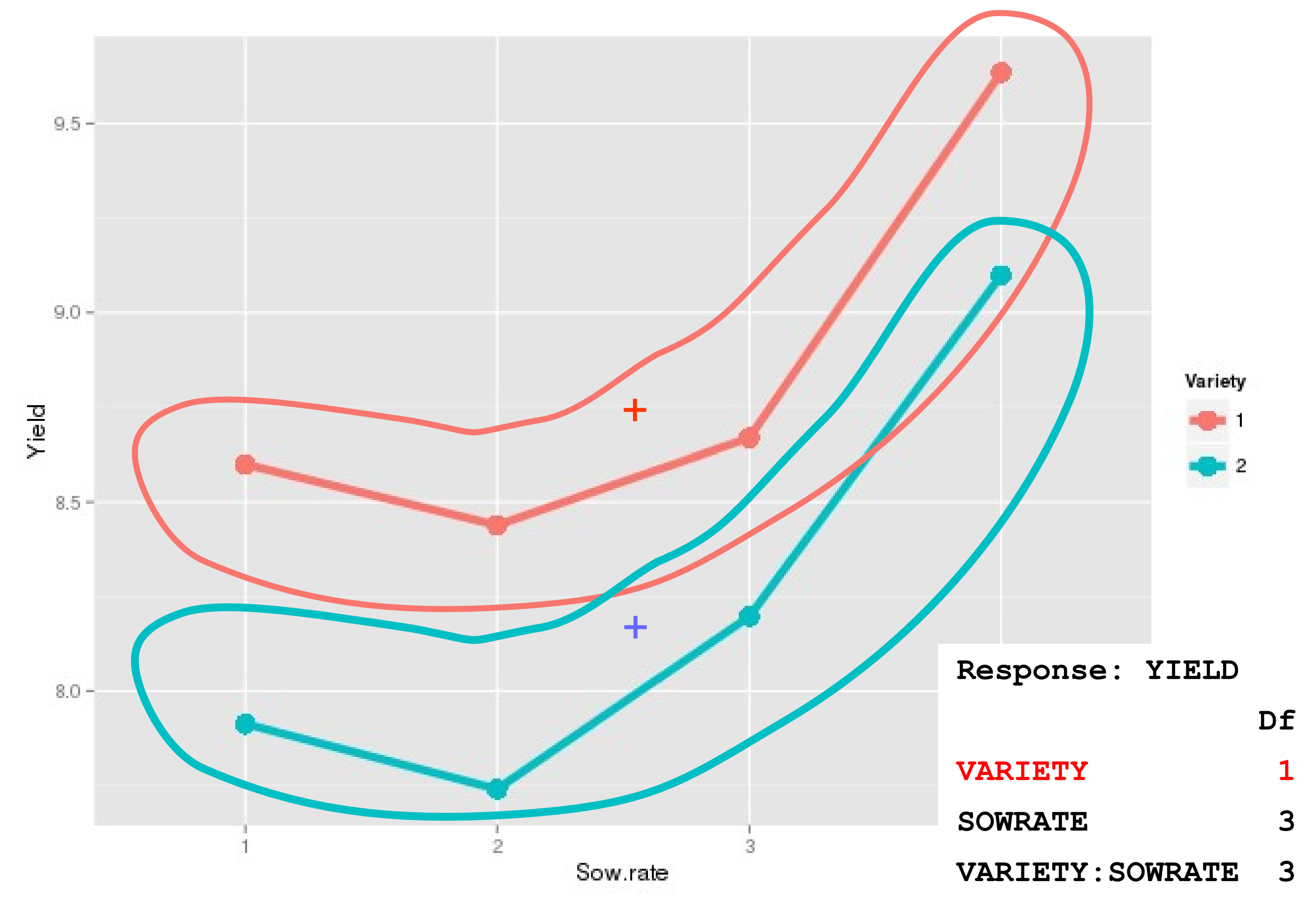


Comparison of the marginal means for each sow rate = the average for sow rates, across the two varieties

Response: YIELD

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
VARIETY	1	2.1474	2.1474	6.0273	0.02591
SOWRATE	3	5.8736	1.9579	5.4952	0.00866
VARIETY:SOWRATE	3	0.0566	0.0189	0.0530	0.98334
Residuals	16	5.7006	0.3563		

Variety main effect

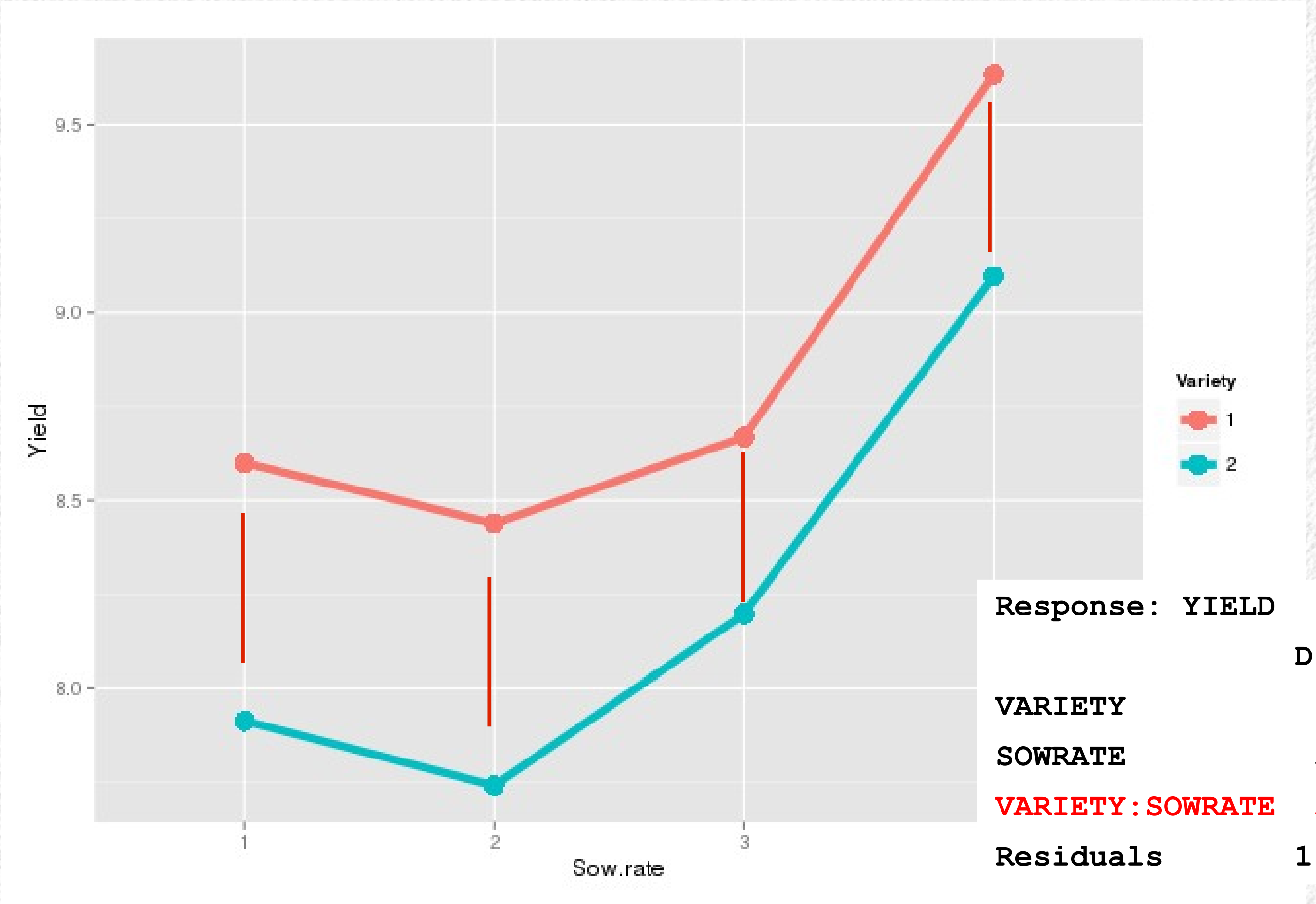


Comparison of the marginal means for each variety, across the four sow rates

Response: YIELD

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
VARIETY	1	2.1474	2.1474	6.0273	0.02591
SOWRATE	3	5.8736	1.9579	5.4952	0.00866
VARIETY:SOWRATE	3	0.0566	0.0189	0.0530	0.98334
Residuals	16	5.7006	0.3563		

Variety x Sow rate interaction



No significant interaction → amount of difference between varieties is the same for every sow rate

...or...

amount of difference between sow rates the same for each variety

Response: YIELD

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
VARIETY	1	2.1474	2.1474	6.0273	0.02591
SOWRATE	3	5.8736	1.9579	5.4952	0.00866
VARIETY:SOWRATE	3	0.0566	0.0189	0.0530	0.98334
Residuals	16	5.7006	0.3563		

The GLM: dummy-coded predictors

SOW-RATE	VARIETY
1	1
2	1
3	1
4	1
1	2
2	2
3	2
4	2

→

Sow rate			*Variety*	*Sow rate x variety interaction*		
S2	S3	S4	V2	S2V2	S3V2	S4V2
0	0	0	0	0	0	0
1	0	0	0	0	0	0
0	1	0	0	0	0	0
0	0	1	0	0	0	0
0	0	0	1	0	0	0
1	0	0	1	1	0	0
0	1	0	1	0	1	0
0	0	1	1	0	0	1

Why 3 df for interaction?

Predicting means from the model

$$YIELD = Intercept + \alpha\, SOWRATE + \beta\, VARIETY + \gamma\, SOWRATE * VARIETY$$

$$YIELD = 8.60 + \begin{matrix} -0.16\ SOWRATE\ 2 \\ 0.07\ SOWRATE\ 3 \\ 1.04\ SOWRATE\ 4 \end{matrix} - 0.69\ VARIETY\ 2 + \begin{matrix} -0.01\ VARIETY\ 2 * SOWRATE\ 2 \\ 0.21\ VARIETY\ 2 * SOWRATE\ 3 \\ 0.15\ VARIETY\ 2 * SOWRATE\ 4 \end{matrix}$$

SOWRATE 1 and VARIETY 1

$$YIELD = 8.60$$

Including an interaction allows each subgroup mean to be predicted exactly

SOWRATE 1 and VARIETY 2

$$YIELD = 8.60 - 0.69\ VARIETY\ 2 = 7.91$$

SOWRATE 2 and VARIETY 2

$$YIELD = 8.60 - 0.16\, SOWRATE\ 2 - 0.69\ VARIETY\ 2 - 0.01\, VARIETY\ 2 * SOWRATE\ 2 = 7.74$$

Interpreting results: main effects only

- If the interaction is not significant, interpretation is simple
  - Response to each variable is independent of the other
  - Each main effect can be interpreted separately, as in a block ANOVA

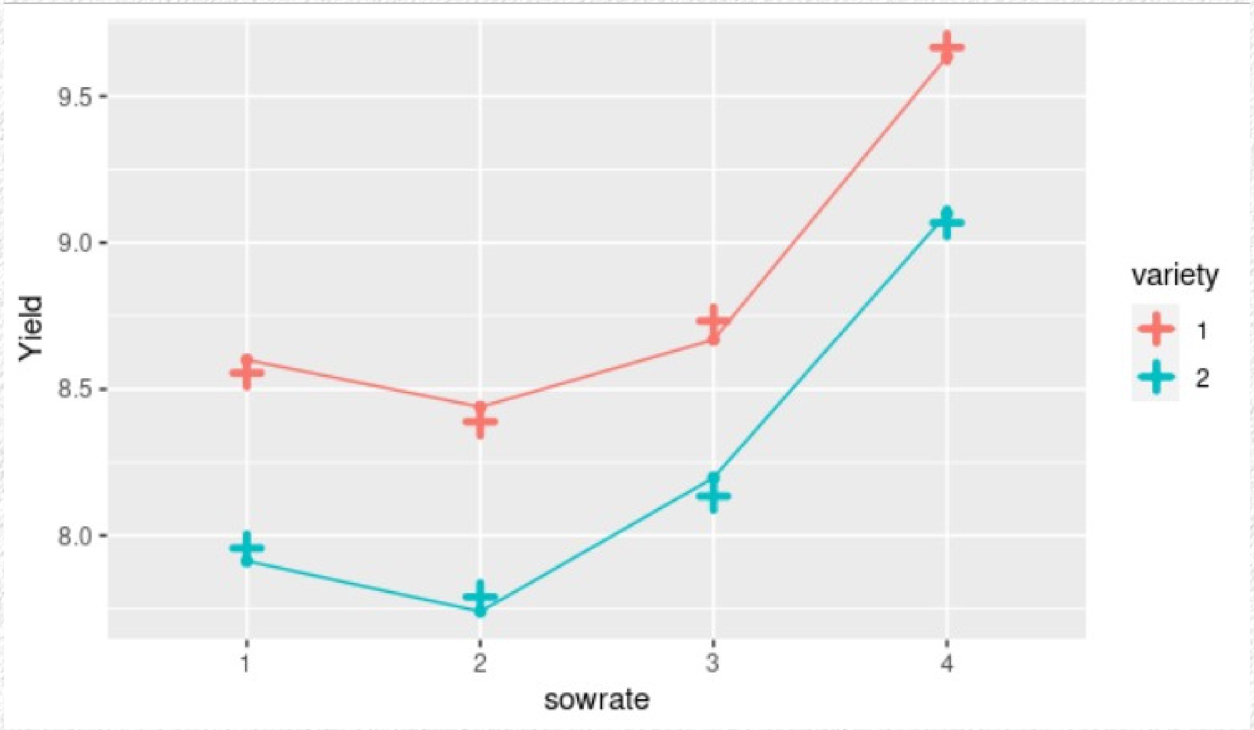
```
Sowing rate post-hocs:

             Estimate Std. Error t value Pr(>|t|)
2 - 1 == 0  -0.1665     0.3178  -0.524  0.95226
3 - 1 == 0   0.1770     0.3178   0.557  0.94351
4 - 1 == 0   1.1110     0.3178   3.496  0.01176 *
3 - 2 == 0   0.3435     0.3178   1.081  0.70508
4 - 2 == 0   1.2775     0.3178   4.020  0.00378 **
4 - 3 == 0   0.9340     0.3178   2.939  0.03855 *
```

- Tukey tests are conducted on marginal means for each predictor
- Variety not included because there are only two

What is the model without the interaction?

- With interaction, predicted values are subgroup means (dots)
- Without interaction, only main effects are used → parallel responses assumed (pluses)
- Main effects are nearly as good at predicting response as subgroup means, so interaction is not significant

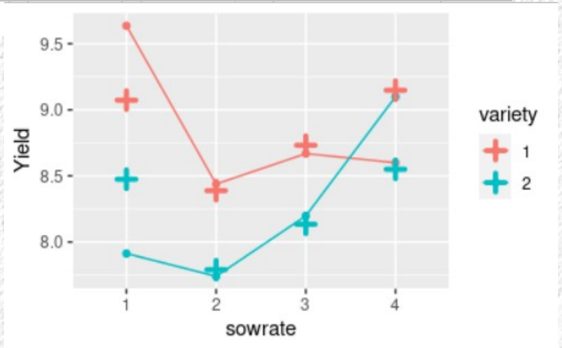


Adding a dependency

Subgroup SS	8.08
Within SS	5.70
Variety SS	2.15
Sow rate SS	2.19
Interaction SS	3.74

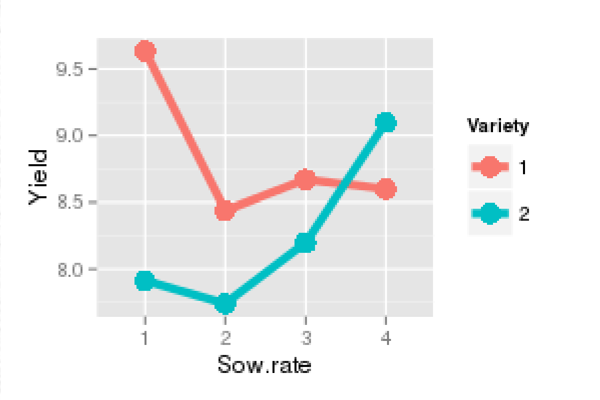
	Sow rate (c = 4)				
Variety (r = 2)	1	2	3	4	*Variety means*
1	8.82	7.83	8.34	8.82	
	9.82	8.64	8.95	8.49	8.84
	10.27	8.85	8.72	8.49	
Subgroup means	9.64	8.44	8.67	8.60	
2	8.53	7.76	8.53	8.99	
	8.17	7.33	8.19	10.15	8.24
	7.05	8.13	7.88	8.15	
Subgroup means	7.91	7.74	8.20	9.10	
Sow rate means	8.77	8.09	8.43	8.85	
Grand mean	8.54				

```
Analysis of Variance Table
Response: yield
                 Df Sum Sq Mean Sq F value  Pr(>F)
sowrate           3 2.1875 0.72917  2.0466 0.14786
variety           1 2.1474 2.14742  6.0273 0.02591
sowrate:variety   3 3.7427 1.24756  3.5016 0.04000
Residuals        16 5.7006 0.35628
```

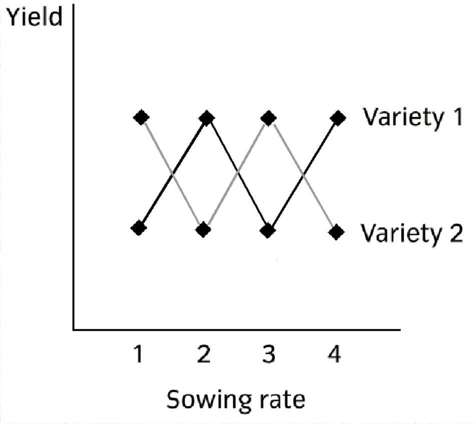


Significant interactions complicate interpretation

- Models with significant interactions may or may not have significant main effects
- The usual interpretation of the main effects may not be correct in the presence of an interaction



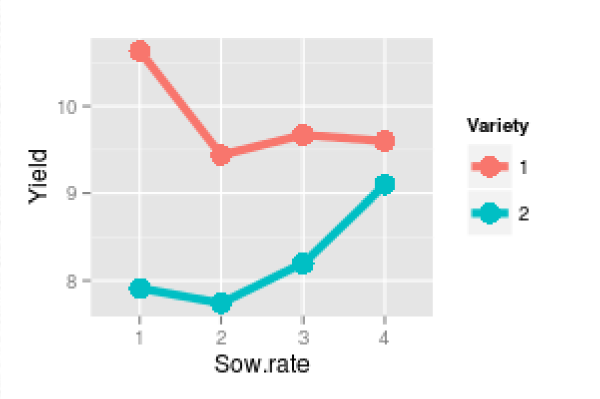
Sow rate main effect and SxV interaction significant

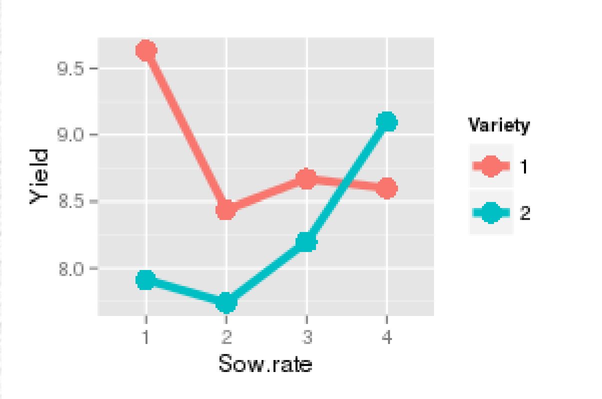


Interaction significant, main effects not

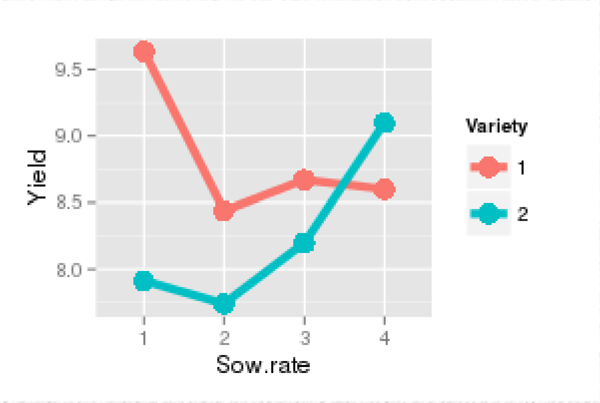
How to proceed when you have a significant interaction

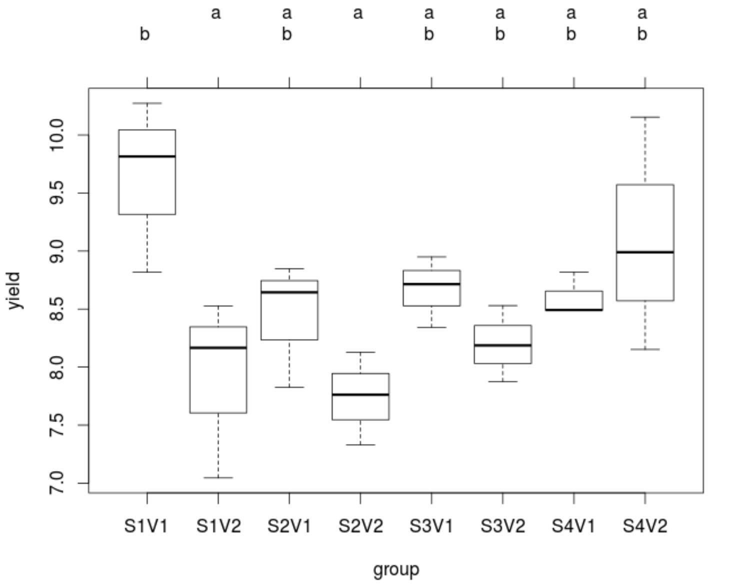
- Start with an interaction plot!
- Ask: do the lines touch or cross?
  - No – can still claim variety 1 > variety 2, but the amount of difference in yield depends on sow rate
  - Yes – now variety 1 > variety 2 only for sow rate 1, 2, and 3, so no general statement possible
- Options for post-hocs...





Group	Yield
S1V1	9.64
S1V2	7.91
S2V1	8.44
S2V2	7.74
S3V1	8.67
S3V2	8.20
S4V1	8.60
S4V2	9.10





Can do Tukey comparison of all 28 combinations of

S and V

Problem: low power (note p-values in list)

```
                   Estimate Std. Error t value Pr(>|t|)
S1V2 - S1V1 == 0 -1.72267    0.48736  -3.535   0.0436
S2V1 - S1V1 == 0 -1.19667    0.48736  -2.455   0.2806
S2V2 - S1V1 == 0 -1.89500    0.48736  -3.888   0.0222
S3V1 - S1V1 == 0 -0.96667    0.48736  -1.983   0.5205
S3V2 - S1V1 == 0 -1.43800    0.48736  -2.951   0.1258
S4V1 - S1V1 == 0 -1.03600    0.48736  -2.126   0.4401
S4V2 - S1V1 == 0 -0.53667    0.48736  -1.101   0.9474
S2V1 - S1V2 == 0  0.52600    0.48736   1.079   0.9524
S2V2 - S1V2 == 0 -0.17233    0.48736  -0.354   0.9999
S3V1 - S1V2 == 0  0.75600    0.48736   1.551   0.7707
S3V2 - S1V2 == 0  0.28467    0.48736   0.584   0.9986
S4V1 - S1V2 == 0  0.68667    0.48736   1.409   0.8406
S4V2 - S1V2 == 0  1.18600    0.48736   2.434   0.2895
S2V2 - S2V1 == 0 -0.69833    0.48736  -1.433   0.8297
S3V1 - S2V1 == 0  0.23000    0.48736   0.472   0.9997
S3V2 - S2V1 == 0 -0.24133    0.48736  -0.495   0.9995
S4V1 - S2V1 == 0  0.16067    0.48736   0.330   1.0000
S4V2 - S2V1 == 0  0.66000    0.48736   1.354   0.8643
S3V1 - S2V2 == 0  0.92833    0.48736   1.905   0.5659
S3V2 - S2V2 == 0  0.45700    0.48736   0.938   0.9772
S4V1 - S2V2 == 0  0.85900    0.48736   1.763   0.6509
S4V2 - S2V2 == 0  1.35833    0.48736   2.787   0.1667
S3V2 - S3V1 == 0 -0.47133    0.48736  -0.967   0.9730
S4V1 - S3V1 == 0 -0.06933    0.48736  -0.142   1.0000
S4V2 - S3V1 == 0  0.43000    0.48736   0.882   0.9837
S4V1 - S3V2 == 0  0.40200    0.48736   0.825   0.9889
S4V2 - S3V2 == 0  0.90133    0.48736   1.849   0.5995
S4V2 - S4V1 == 0  0.49933    0.48736   1.025   0.9635
```

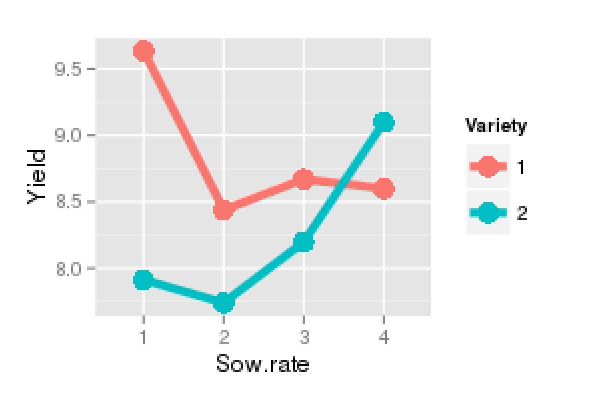
Can compare within levels of one variable

	Estimate	Std. Error	t value	Pr(>|t|)
1:2 - 1 == 0	-1.19667	0.48736	-2.455	0.192
1:3 - 1 == 0	-0.96667	0.48736	-1.983	0.395
1:4 - 1 == 0	-1.03600	0.48736	-2.126	0.323
1:3 - 2 == 0	0.23000	0.48736	0.472	0.998
1:4 - 2 == 0	0.16067	0.48736	0.330	1.000
1:4 - 3 == 0	-0.06933	0.48736	-0.142	1.000
2:2 - 1 == 0	-0.17233	0.48736	-0.354	1.000
2:3 - 1 == 0	0.28467	0.48736	0.584	0.995
2:4 - 1 == 0	1.18600	0.48736	2.434	0.199
2:3 - 2 == 0	0.45700	0.48736	0.938	0.946
2:4 - 2 == 0	1.35833	0.48736	2.787	0.107
2:4 - 3 == 0	0.90133	0.48736	1.849	0.471

Can compare sow rates within each variety

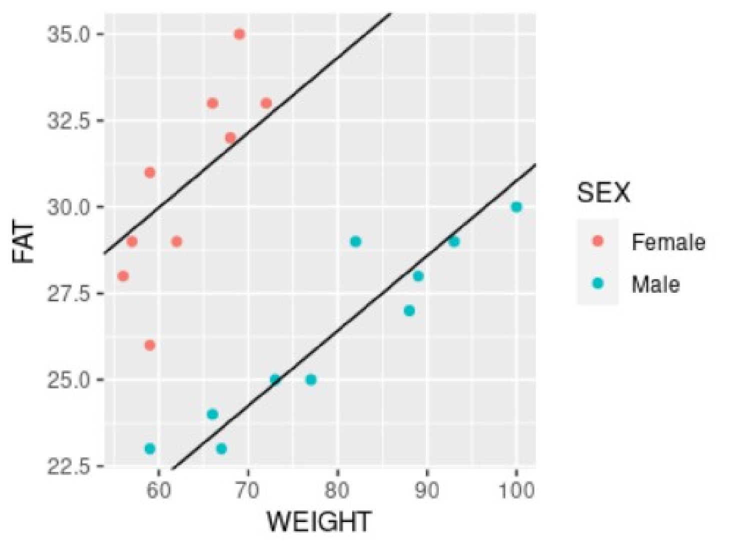
	Estimate	Std. Error	t value	Pr(>|t|)
1:2 - 1 == 0	-1.7227	0.4874	-3.535	0.0106
2:2 - 1 == 0	-0.6983	0.4874	-1.433	0.5052
3:2 - 1 == 0	-0.4713	0.4874	-0.967	0.8008
4:2 - 1 == 0	0.4993	0.4874	1.025	0.7674

Can compare varieties within each sow rate



Interactions with mixes of continuous and categorical predictors

- Without an interaction, ANCOVA gives parallel lines
  - A line for each group
  - Same slope, different intercepts
- An interaction between the numeric predictor and the grouping variable gives lines that aren't parallel
- Example of fat data – relationship between fat and weight for men and women



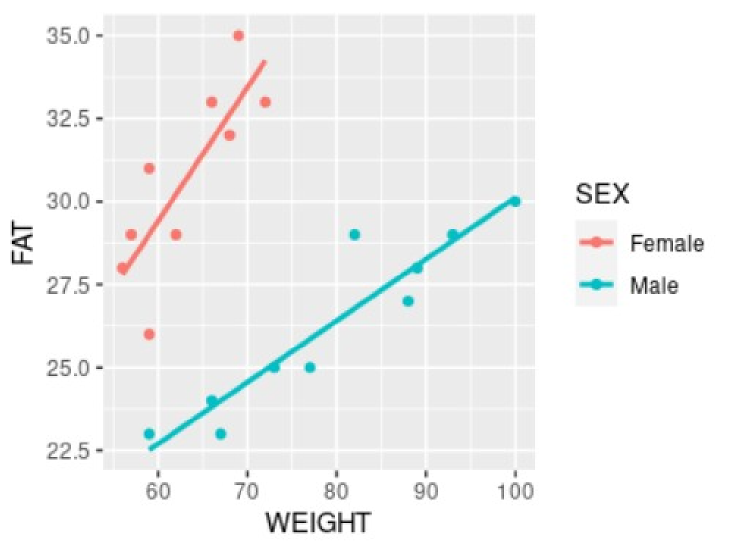
No interaction term: parallel lines

`FAT ~ SEX + WEIGHT`

Interaction between sex and weight

```
Anova Table (Type II tests)

Response: FAT
             Sum Sq Df F value    Pr(>F)
WEIGHT       87.105  1 43.3535 8.667e-06 ***
SEX         176.098  1 87.6467 1.181e-07 ***
WEIGHT:SEX   10.857  1  5.4039   0.03454 *
Residuals    30.138 15
```



With interaction term:
different slope and intercept for each

`FAT ~ SEX * WEIGHT`

The model equation...

$$FAT = Intercept + \alpha\,\mathrm{SEX} + \beta\,\mathrm{WEIGHT} + \gamma\,\mathrm{SEX*WEIGHT}$$

$$FAT = 5.23 + 6.33\,\mathrm{SEXMale} + 0.40\,\mathrm{WEIGHT} - 0.22\,\mathrm{SEXMale} * \mathrm{WEIGHT}$$

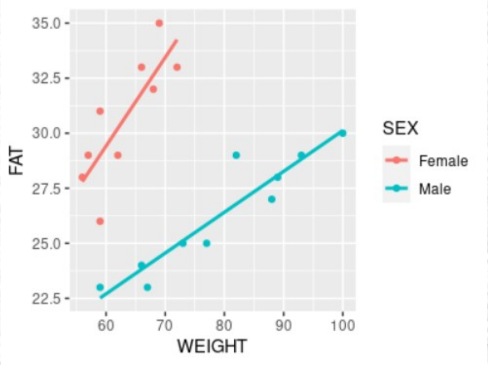
Mean for females with weight = 0

Main effects of being male

Main effect of weight

Effect of being male on

effect of weight



...produces two non-parallel lines

$$FAT = 5.23 + 6.33\,\text{SEXMale} + 0.40\,\text{WEIGHT} - 0.22\,\text{SEXMale} * \text{WEIGHT}$$

Females

$$FAT = 5.23 + 6.33(0) + 0.4\,\text{WEIGHT} - 0.22(0)\,\text{WEIGHT}$$

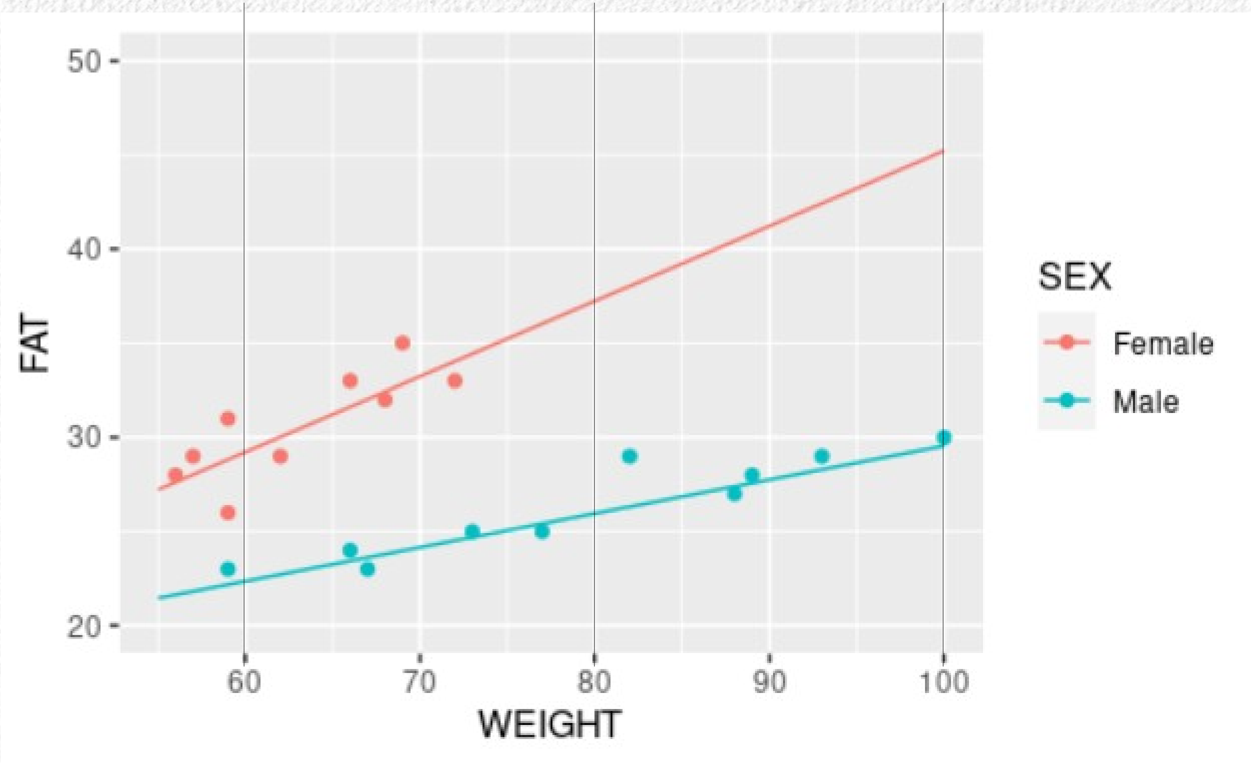
$$FAT = 5.23 + 0.4\,\text{WEIGHT}$$

Males

$$FAT = 5.23 + 6.33(1) + 0.4\,\text{WEIGHT} - 0.22(1)\,\text{WEIGHT}$$

$$FAT = 11.56 + 0.18\,\text{WEIGHT}$$

Can you interpret the main effect of sex?

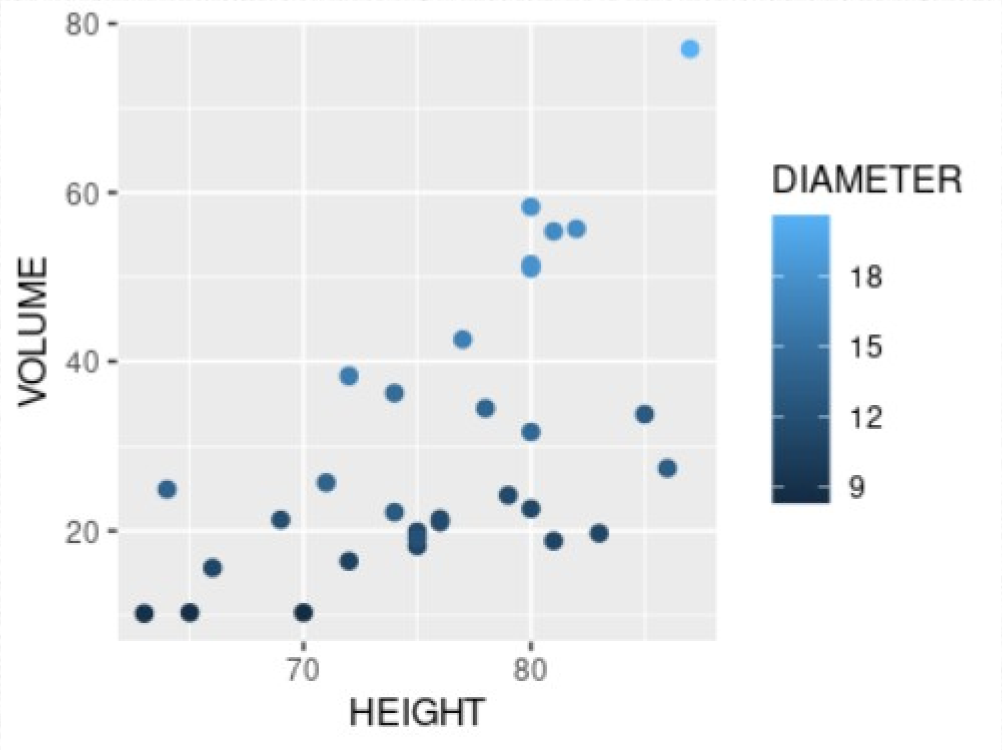


Would you get the same difference between sexes if you set weight to 60, 80, or 100?

Why is there a significant main effect of sex?

Interactions with 2 continuous variables

- Like the others, in that an interaction is expressed as a multiplicative effect
- The equation is a bit simpler
- The response is a bit more complex
- Example: tree volume as a function of diameter and height



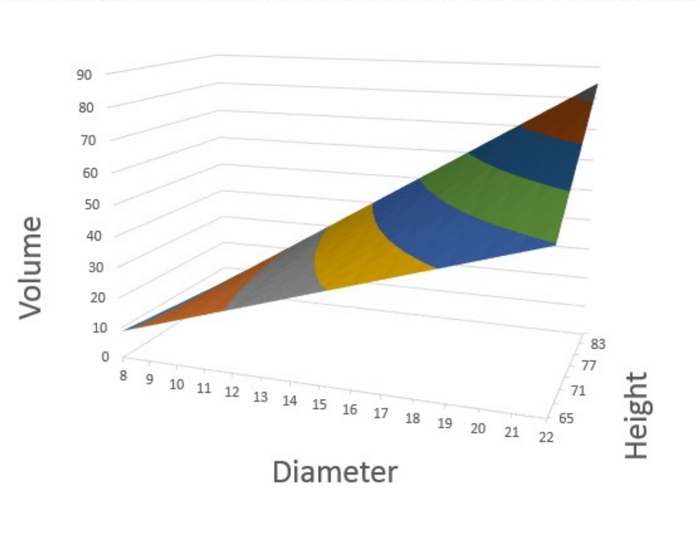
Coefficients:

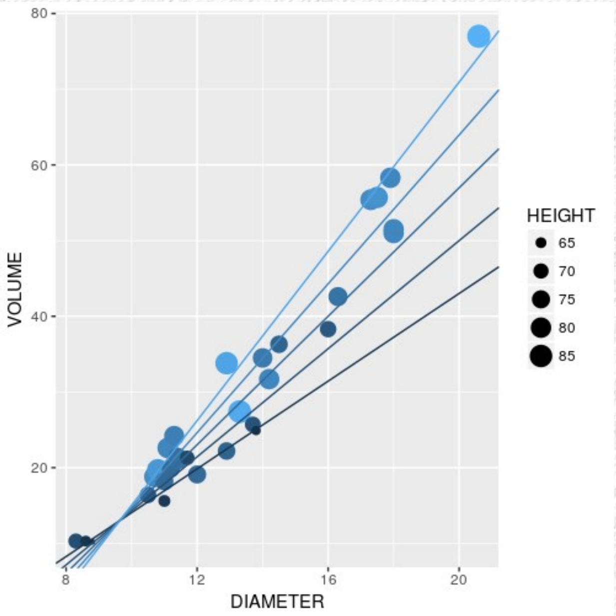
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	69.39631	23.83576	2.911	0.00713
DIAMETER	-5.85585	1.92134	-3.048	0.00511
HEIGHT	-1.29708	0.30984	-4.186	0.00027
DIAMETER:HEIGHT	0.13465	0.02438	5.524	7.48e-06

Response: VOLUME

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
DIAMETER	1	7581.8	7581.8	1033.469	< 2.2e-16
HEIGHT	1	102.4	102.4	13.956	0.0008867
DIAMETER:HEIGHT	1	223.8	223.8	30.512	7.484e-06
Residuals	27	198.1	7.3		

- Think of this as a set of lines with different slopes...
  - set height to a constant value
  - predict volume at that height across the range of diameters
- ...or, as a curved 3-D surface





Predicted values

Multiple regression equation with coefficients

$$VOLUME = 69.40 - 1.30\ HEIGHT - 5.86\ DIAMETER + 0.13\ HEIGHT * DIAMETER$$

Predicted values with height = 65

$$VOLUME = 69.40 - 1.30(65) - 5.86\ DIAMETER + 0.13(65) * DIAMETER$$

$$VOLUME = -15.1 + 2.59\ DIAMETER$$

Predicted values with height = 85

$$VOLUME = 69.40 - 1.30(85) - 5.86\ DIAMETER + 0.13(85) * DIAMETER$$

$$VOLUME = -41.1 + 5.58\ DIAMETER$$

Last words

- Factorial experiments are efficient ways to test more than one predictor at a time
- Major advantage: interactions can be tested
- Any combination of variable types can interact
  - Two categorical
  - Two numeric
  - Mix of numeric and categorical
- More complex cause/effect relationships can be studied
  - Synergistic, antagonistic responses
- In the presence of an interaction, be VERY careful about interpreting main effects! Make heavy use of graphs to avoid big mistakes

What's the model?

Predictor?

Response?

Main effect of location significant?

Main effect of age significant?

Interaction?