Stat 412/512

Multifactor Studies without Replication

Feb 12 2015

Charlotte Wickham

stat512.cwick.co.nz

Replication

Multiple measurements at a specific combination of explanatory variable values, are called **replicates**.

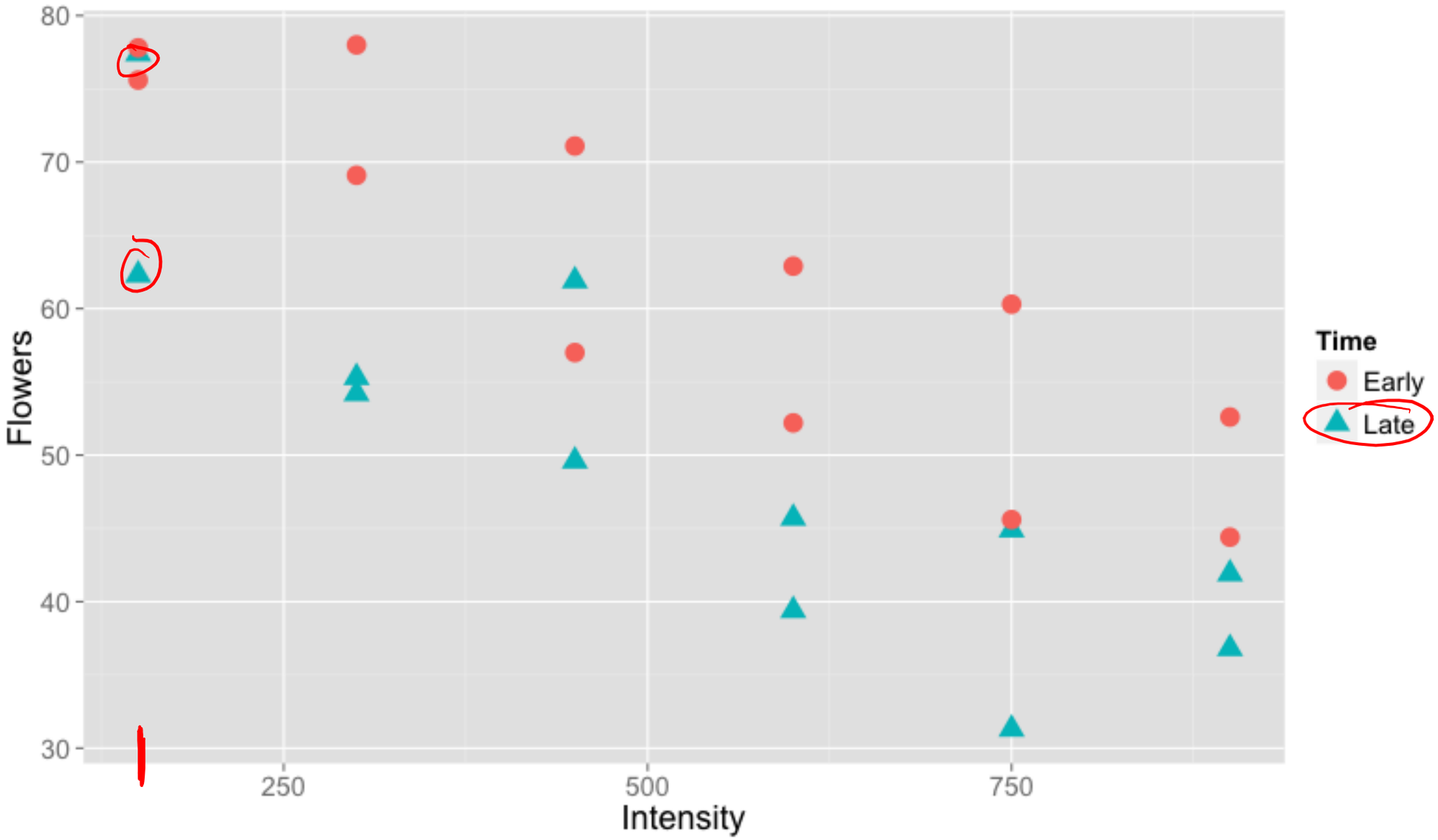
Meadowfoam, seaweed grazers had replicates.

This chapter (14) focusses on examples without replicates.

Meadowfoam

Two factors: Intensity & Timing

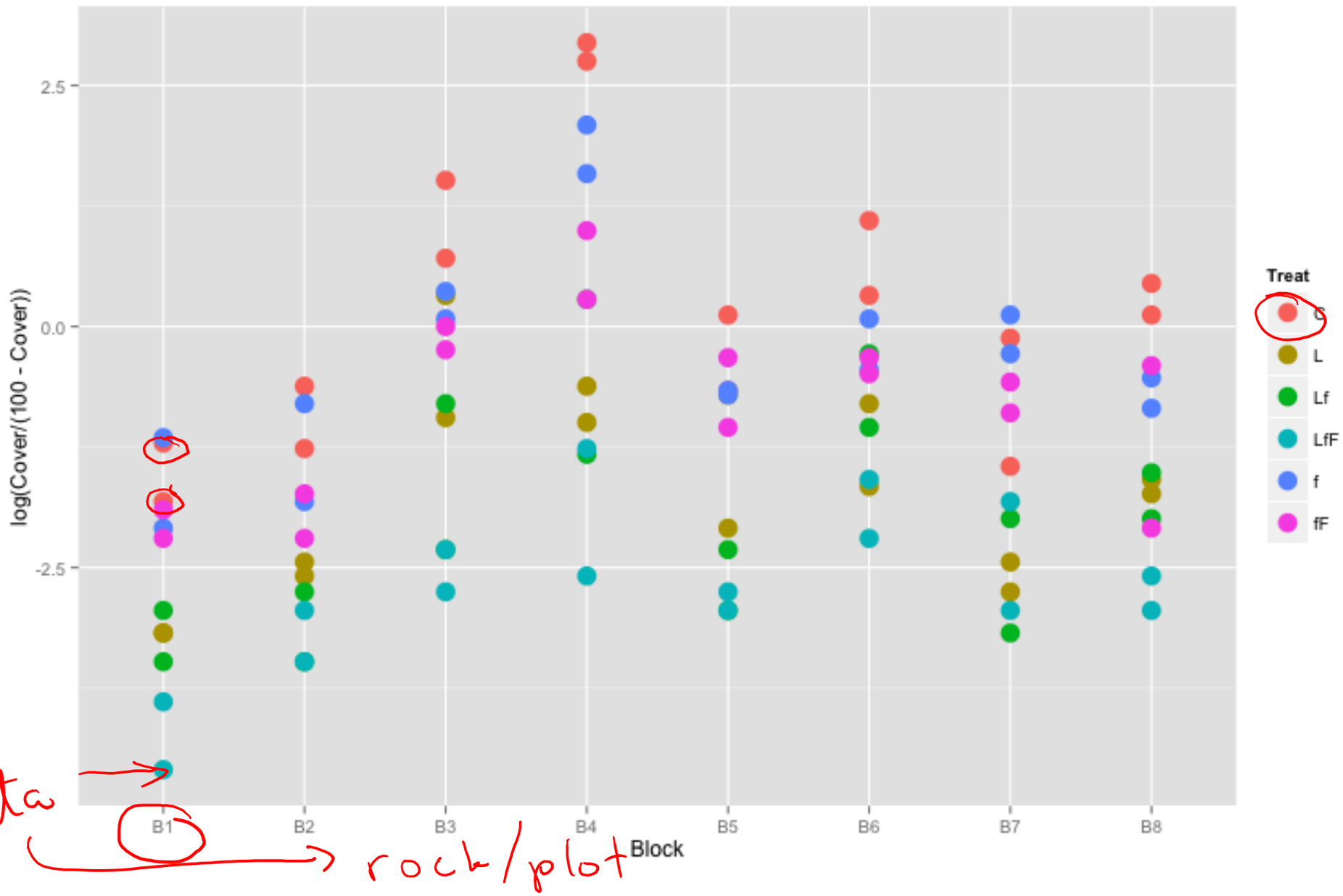
For each combination we have two **replicates**



Seaweed grazers

Two factors: Microhabitat & Treatment

For each combination we have two **replicates**



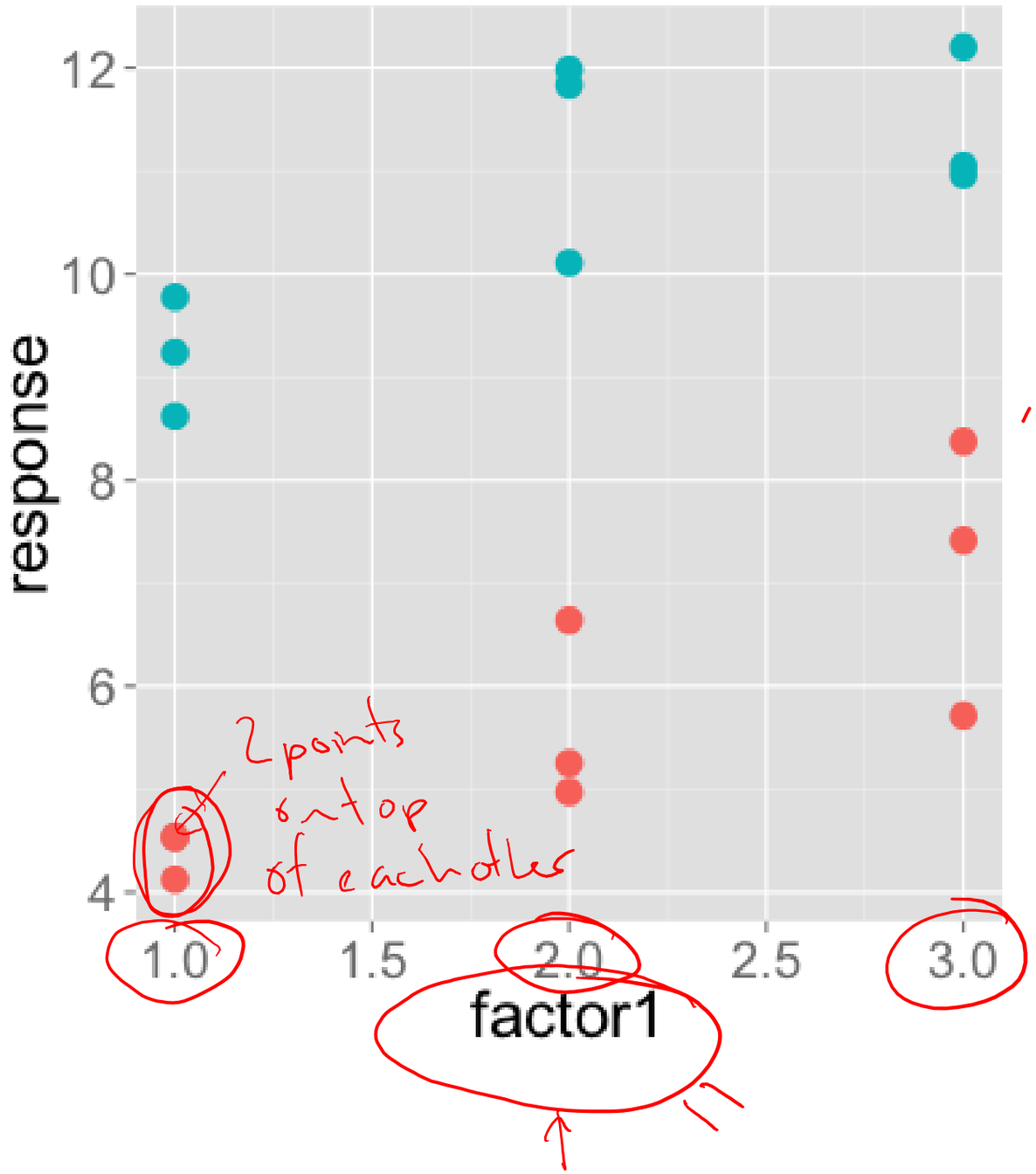
Today

Why is replication good to have?

What do you do when you don’t have replication?

An example, a two factor study with no replication.

Two-factor example with replicates

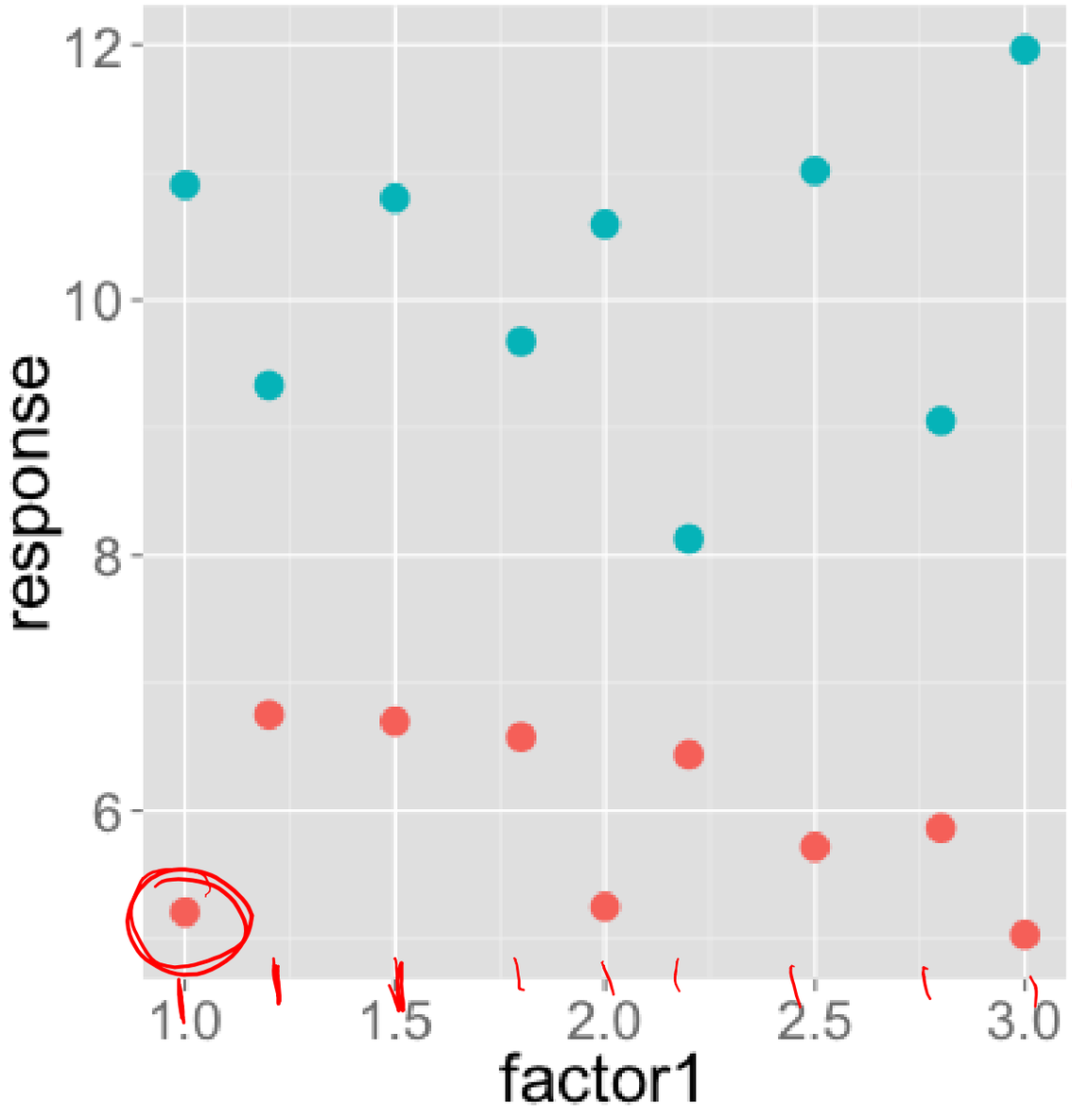


factor 1: categorical with two levels, A and B.

factor 2: continuous, but only set to three levels, 1, 2 and 3.

3 replicates at each combination of factor 1 and factor 2.

Two-factor example without replicates



factor 1: categorical with two levels, A and B.

factor 2: continuous, set to nine levels between 1 and 3.

NO replicates.

Why are replicates good?

Replicates allow a "model-free" estimate of variation, σ^2.

Lack of fit F-tests are available for any model.

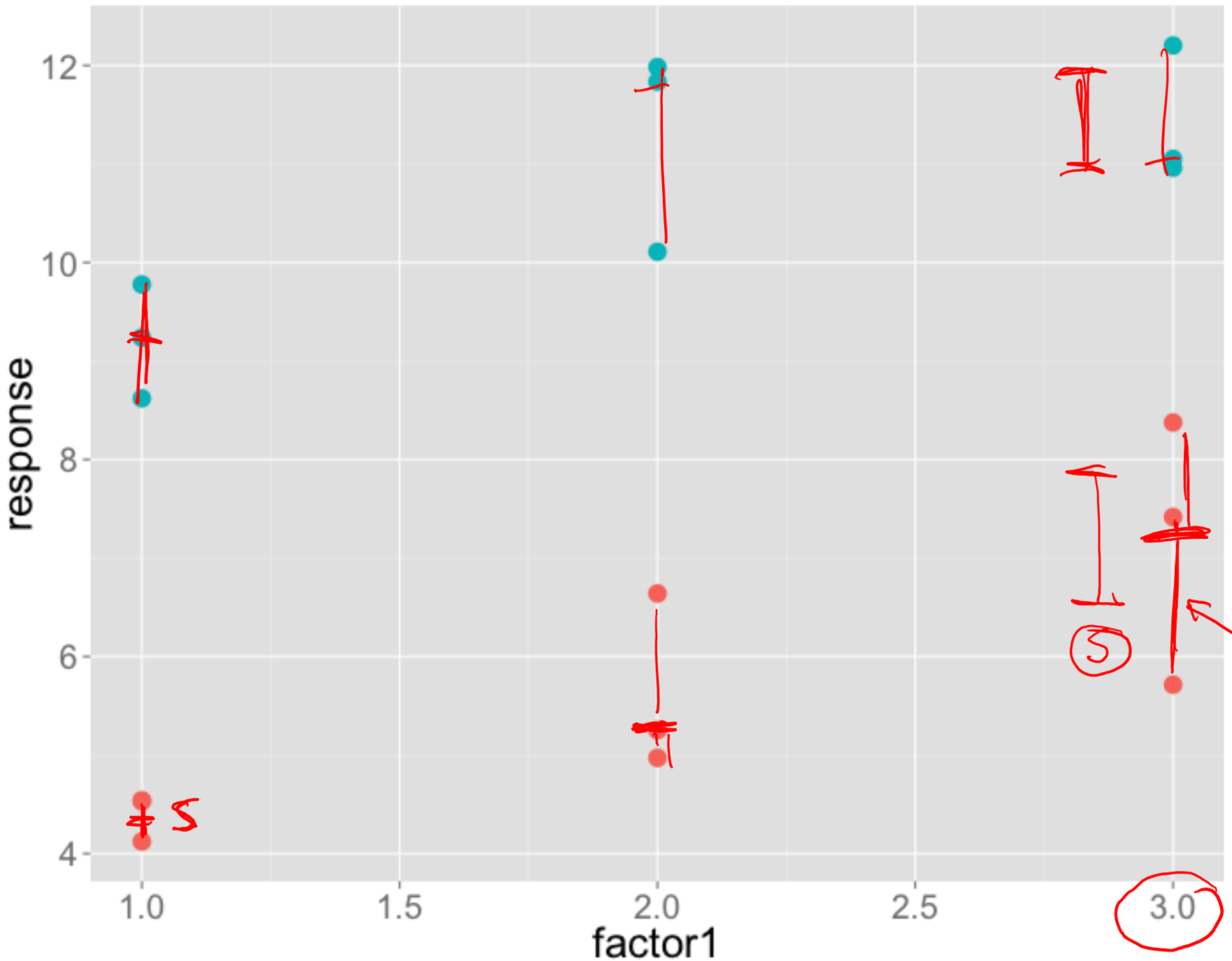
Without replicates we rely on our model being adequate, and using the residuals to estimate, σ^2.

If the saturated model is fitted, there are no degrees of freedom left for estimating σ^2.

perfect fit

with replicates

σ^2 = variation we would expect if we could make multiple observations at a set of explanatory values



deviations from the mean tells about σ^2

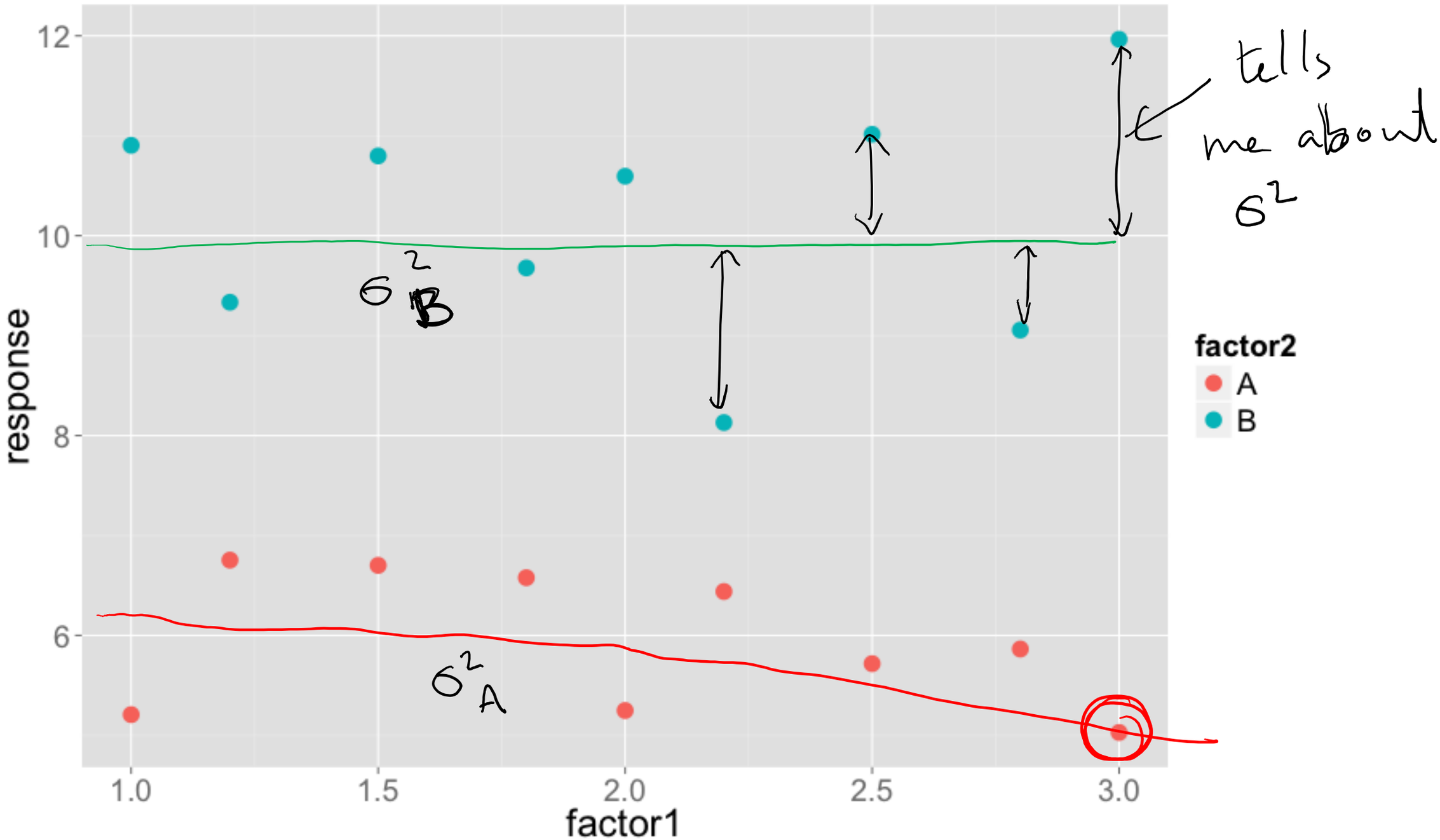
saturated model

without replicates

Separate lines

$\mu\{\text{response} \mid \text{factor1 \& F2}\} = \beta_0 + \beta_1 \text{Factor} + \ldots$

↑ means this isn't "model free"



Strategy

We are still working in the multiple regression world.

Fit tentative model, check for transformations outliers, refine and check model. Interpret.

Ways to deal with non-replication:

→ Assume some interactions don't exist.

→ Treat numerical factors as continuous not categorical.

case1401: Chimp data

Teach 10 American Sign Language Signs to four chimpanzees.

Response: "time in minutes it took to learn the sign"

Are some signs easier to learn? Do some chimps take longer to learn words?

Minutes to acquisition of American Sign Language signs by four chimps

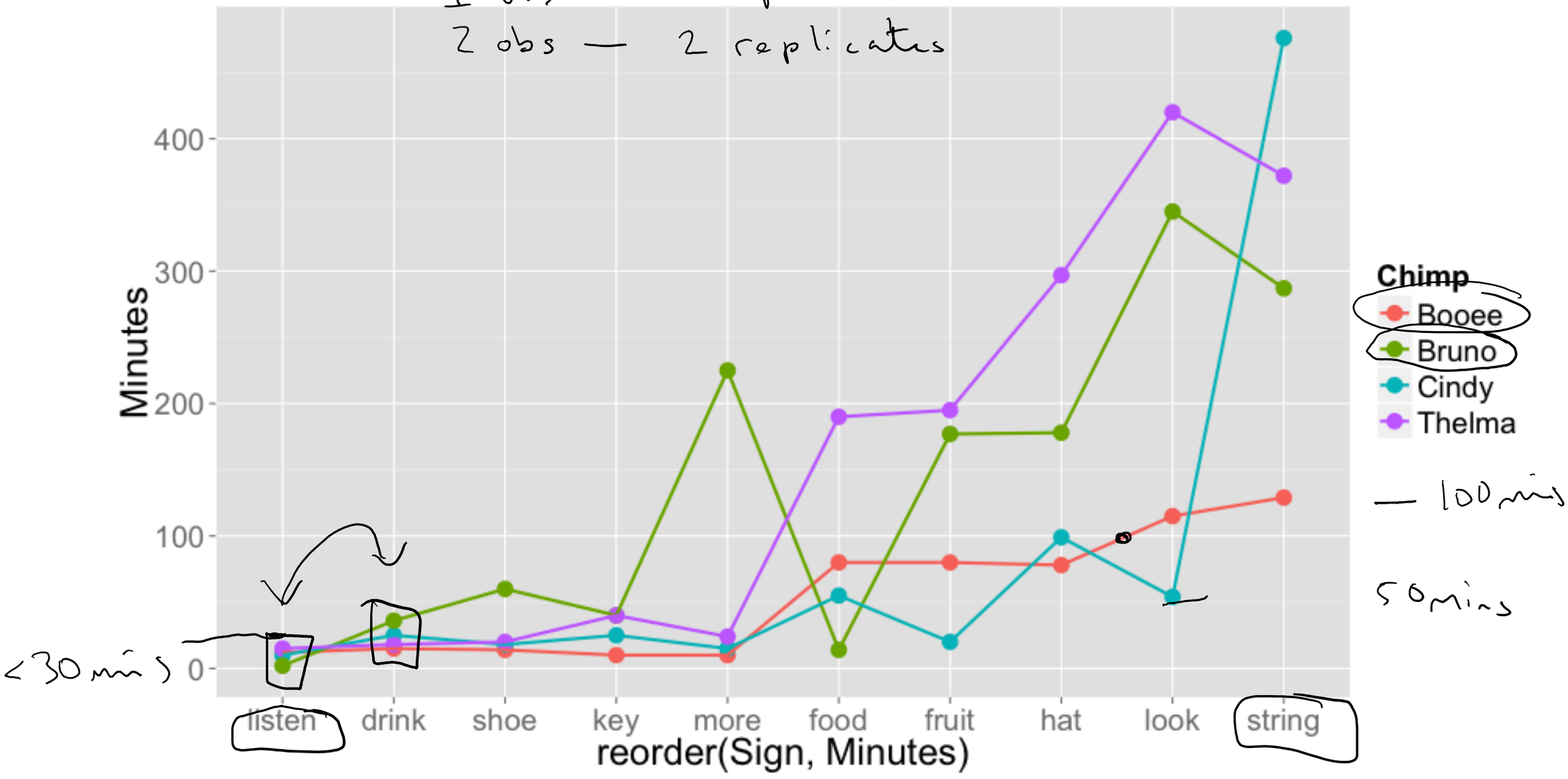
Signs

	listen	*drink*	*shoe*	*key*	*more*	*food*	*fruit*	*hat*	*look*	*string*
Booee	12	15	14	10	10	80	80	78	115	129
Cindy	10	25	18	25	15	55	20	99	54	476
Bruno	2	36	60	40	225	14	177	178	345	287
Thelma	15	18	20	40	24	190	195	297	420	372

Chimp

1 obs — no replicates

2 obs — 2 replicates



< 30 mins

— 100 mins

50 mins

Tentative model

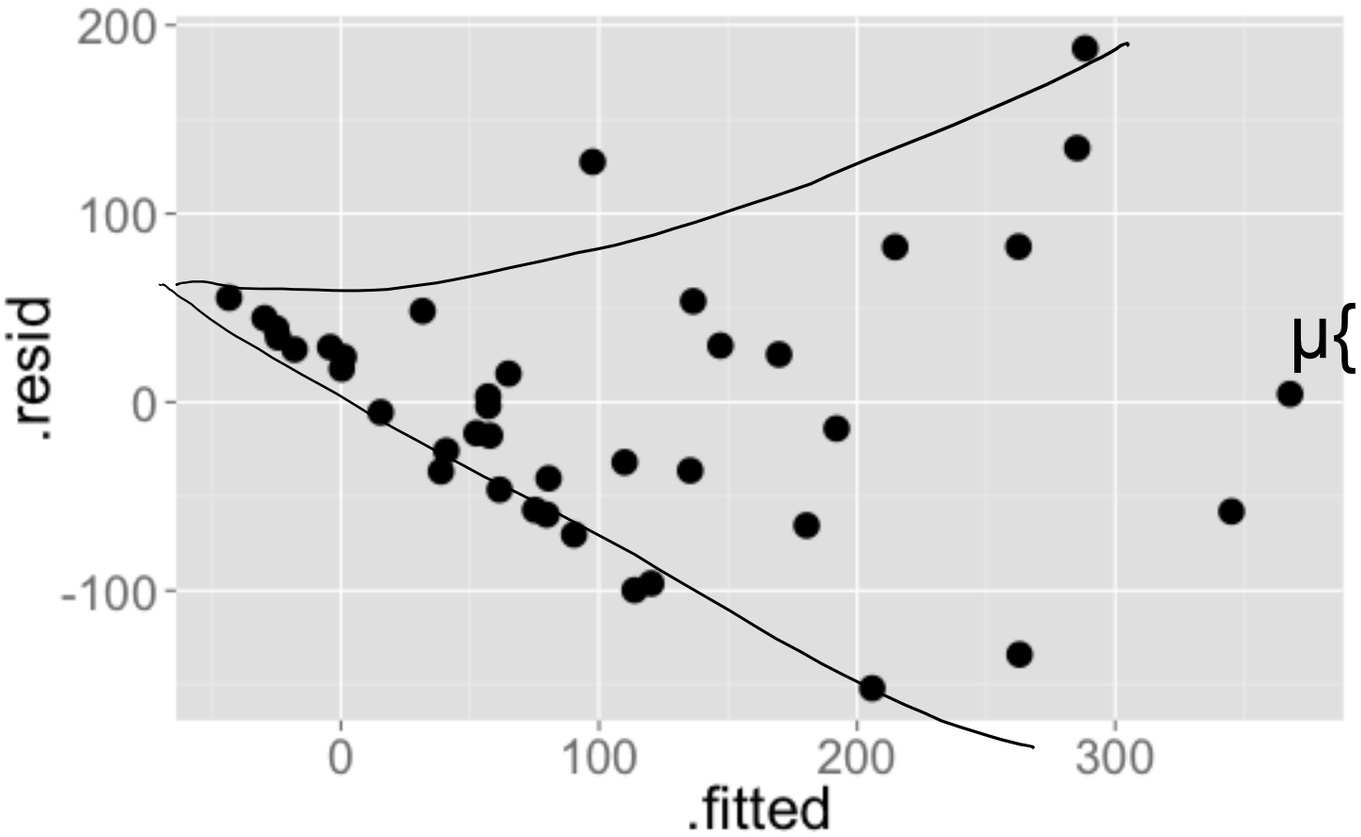
The saturated model:

μ{ Minutes | Chimp, Sign } = CHIMP + SIGN + ~~CHIMP×SIGN~~

leaves us with no d.f. to estimate σ.

We are going to assume there are **no interactions,** and fit the additive model:

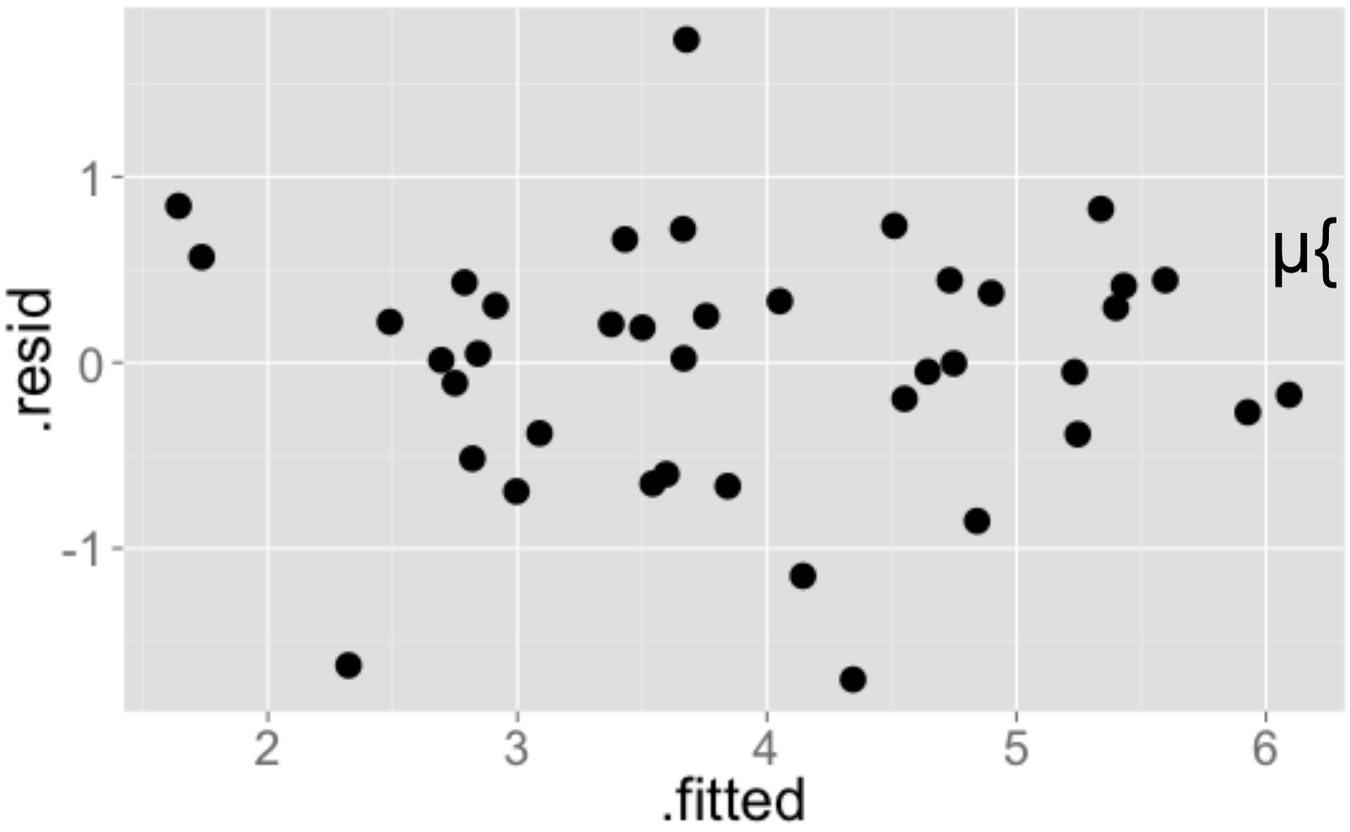
μ{ Minutes | Chimp, Sign } = CHIMP + SIGN



Residual plot from:

μ{ Minutes | Chimp, Sign } =

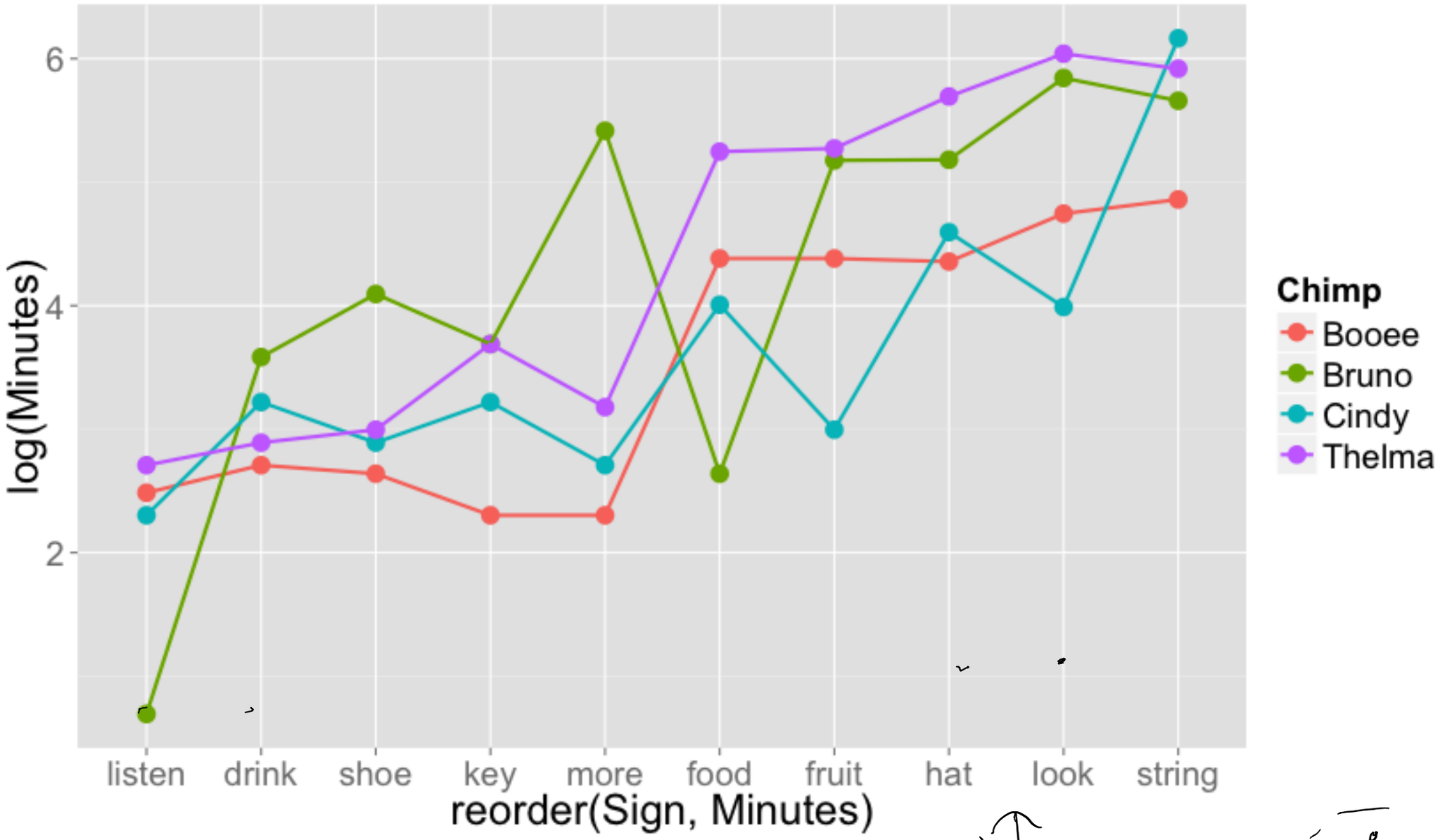
CHIMP + SIGN



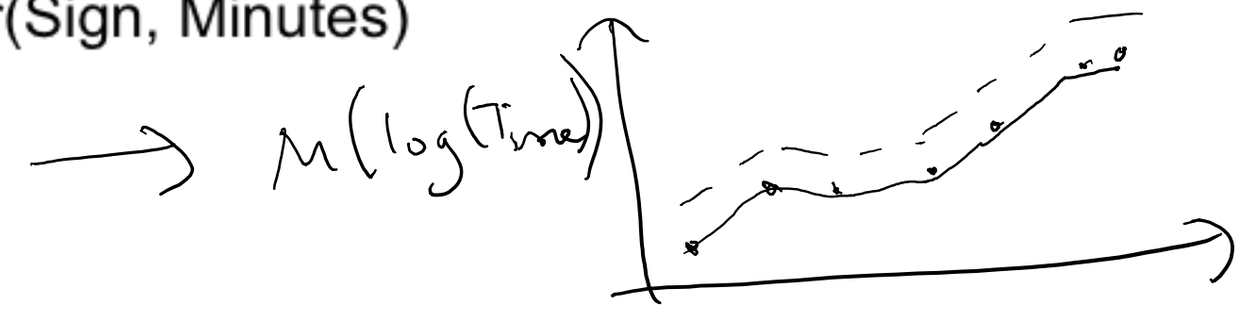
Residual plot from:

μ{ **log**(Minutes) | Chimp, Sign } =

CHIMP + SIGN



Additive model?



We can’t evaluate this with an F-test, we have to make an argument based on this plot (or outside knowledge).

Analysis of variance for the additive model fit to log(acquisition times)

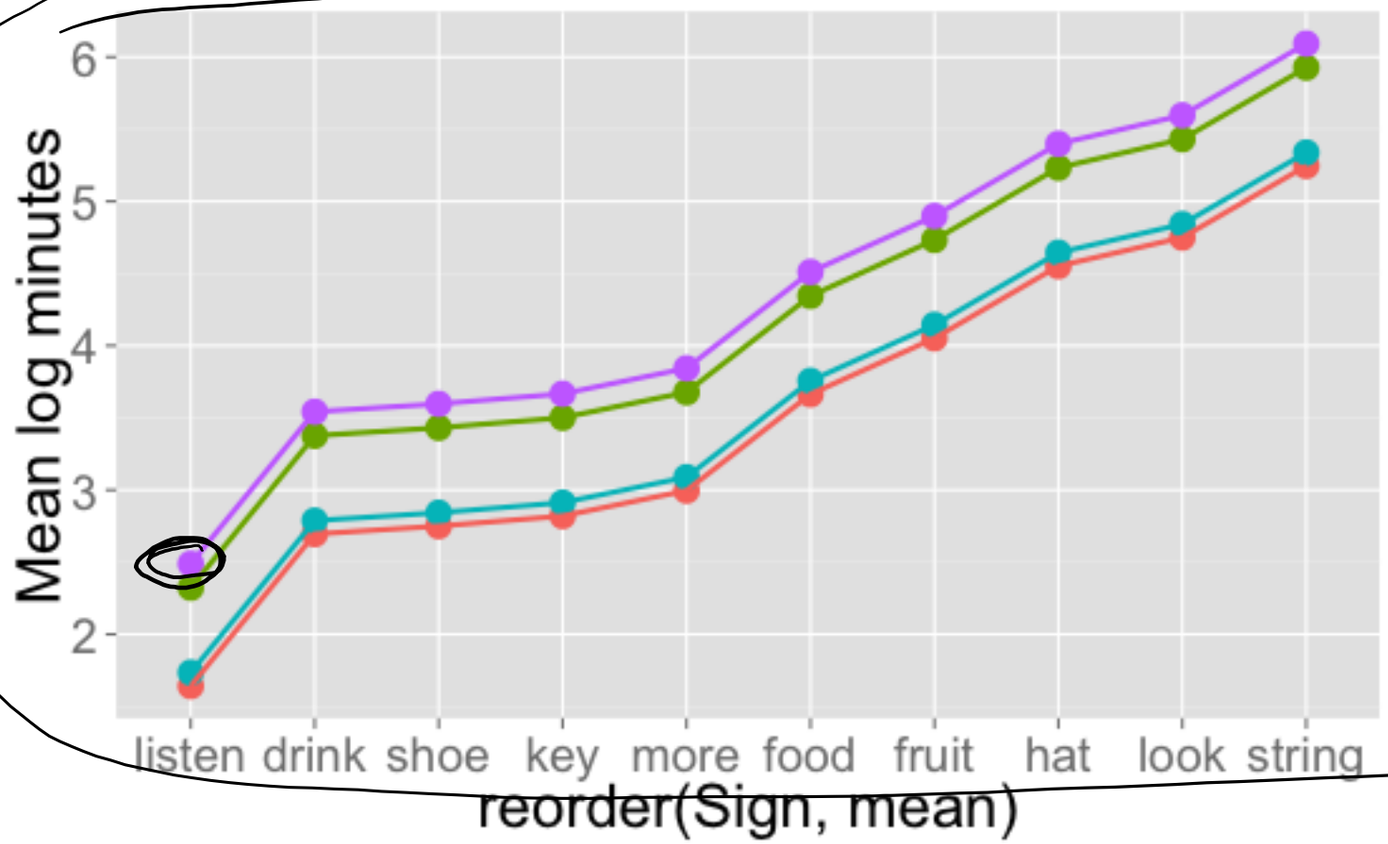
Source of Variation	Sum of Squares	df	Mean Square	F-Statistic	p-value
Signs	45.6900	9	5.0767	7.7649	0.00001
Chimpanzees	5.3329	3	1.7776	2.7190	0.0642
Residual	17.6526	27	0.6538		
Total	68.6755	39			

R-squared = 74.3%

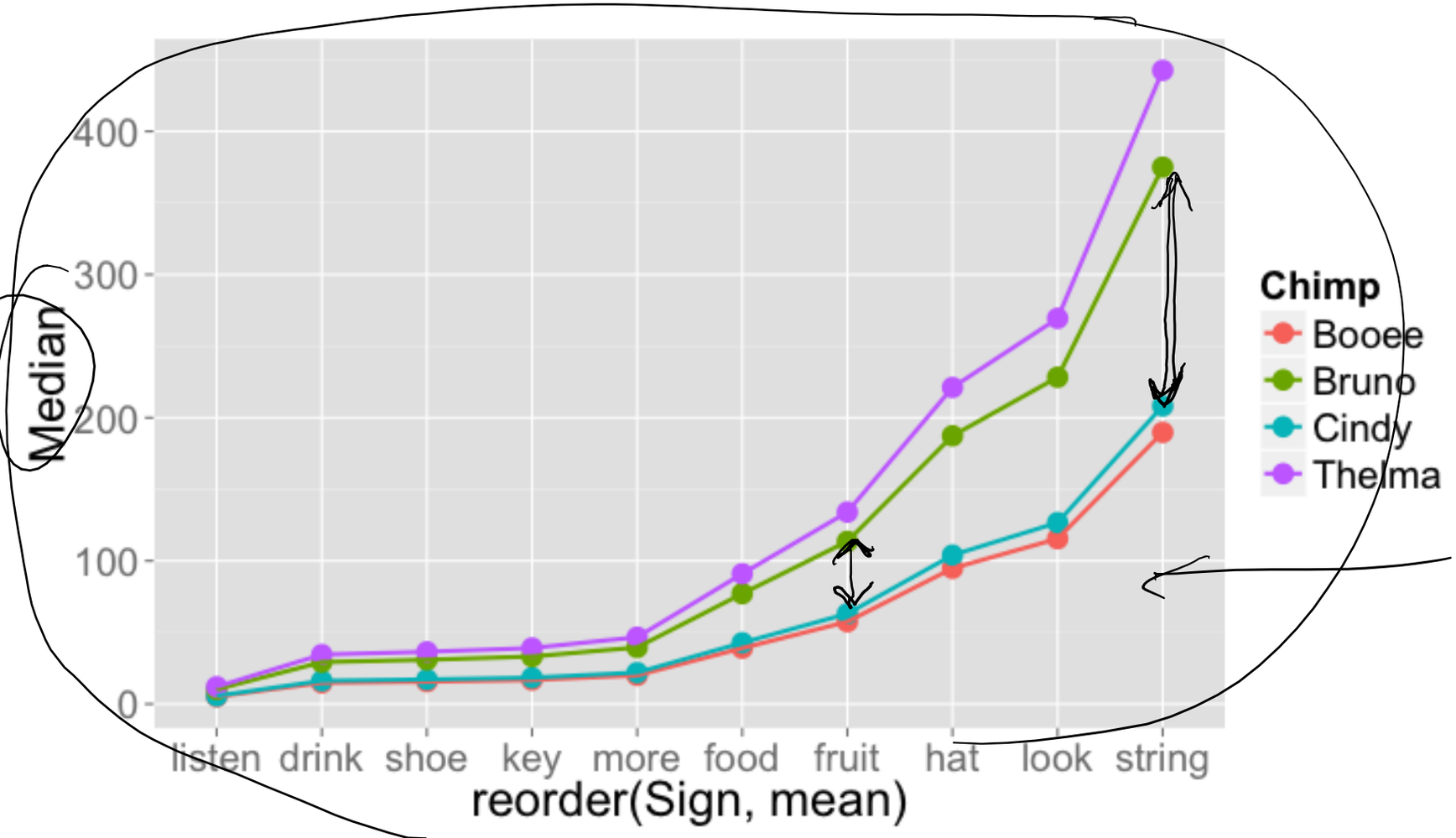
Estimated SD = 0.8086

There is strong evidence that sign is associated with a change in mean log time to learn word (extra sum of squares F-test on 9 and 27 degrees of freedom, pvalue < 0.0001), after accounting for chimp.

There is weak evidence that chimp is associated with a change in median time to learn word (extra sum of squares F-test on 3 and 27 degrees of freedom, pvalue < 0.06), after accounting for sign.



on log scale



backtransformed

multiplicative effects

Multiple comparisons

If we want to make all pairwise comparisons between Signs we should adjust for multiple comparisons.

Tukey-Kramer is appropriate here.

Multiple comparisons of sign means on the log scale