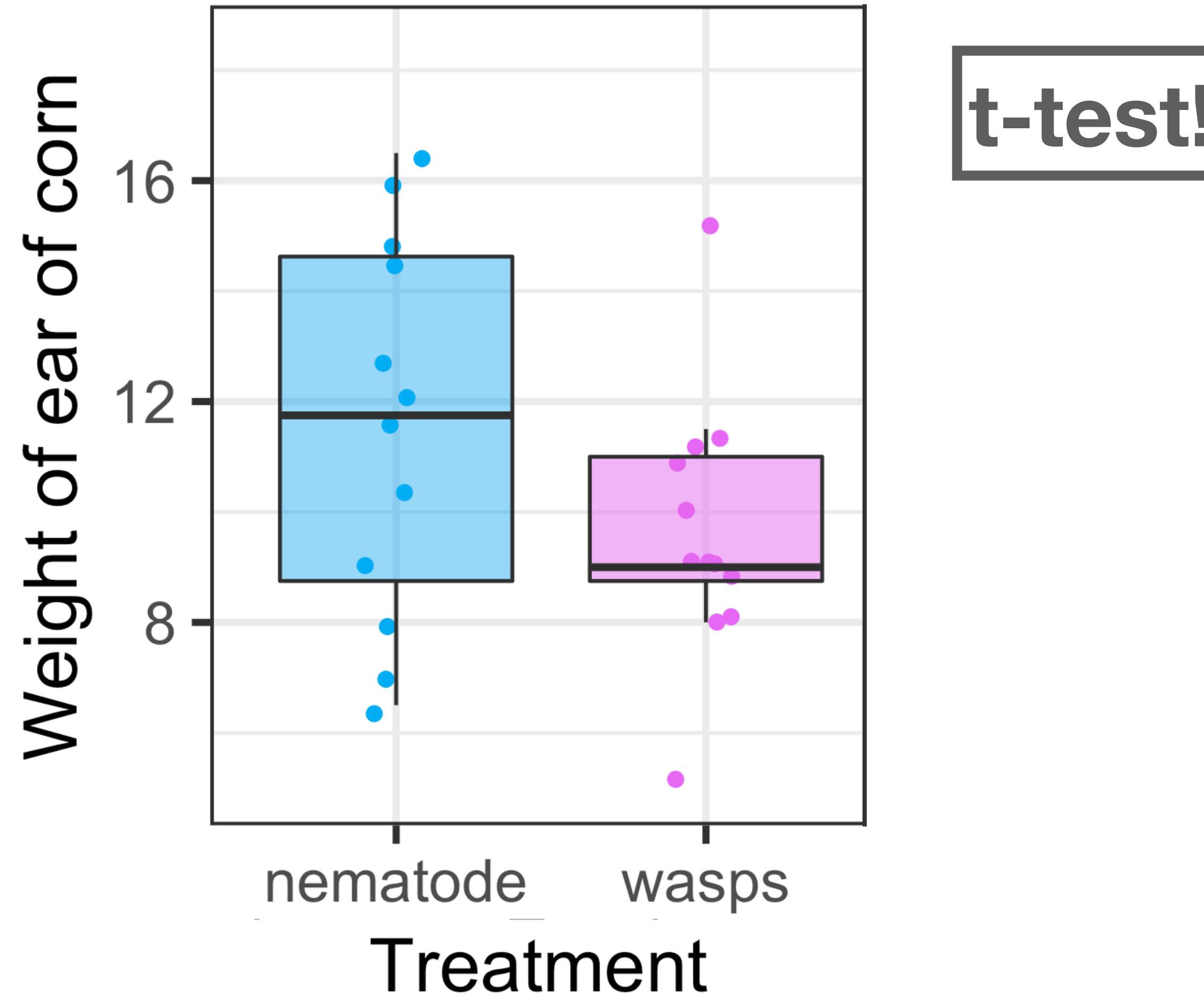


Lecture 15

11.23.21

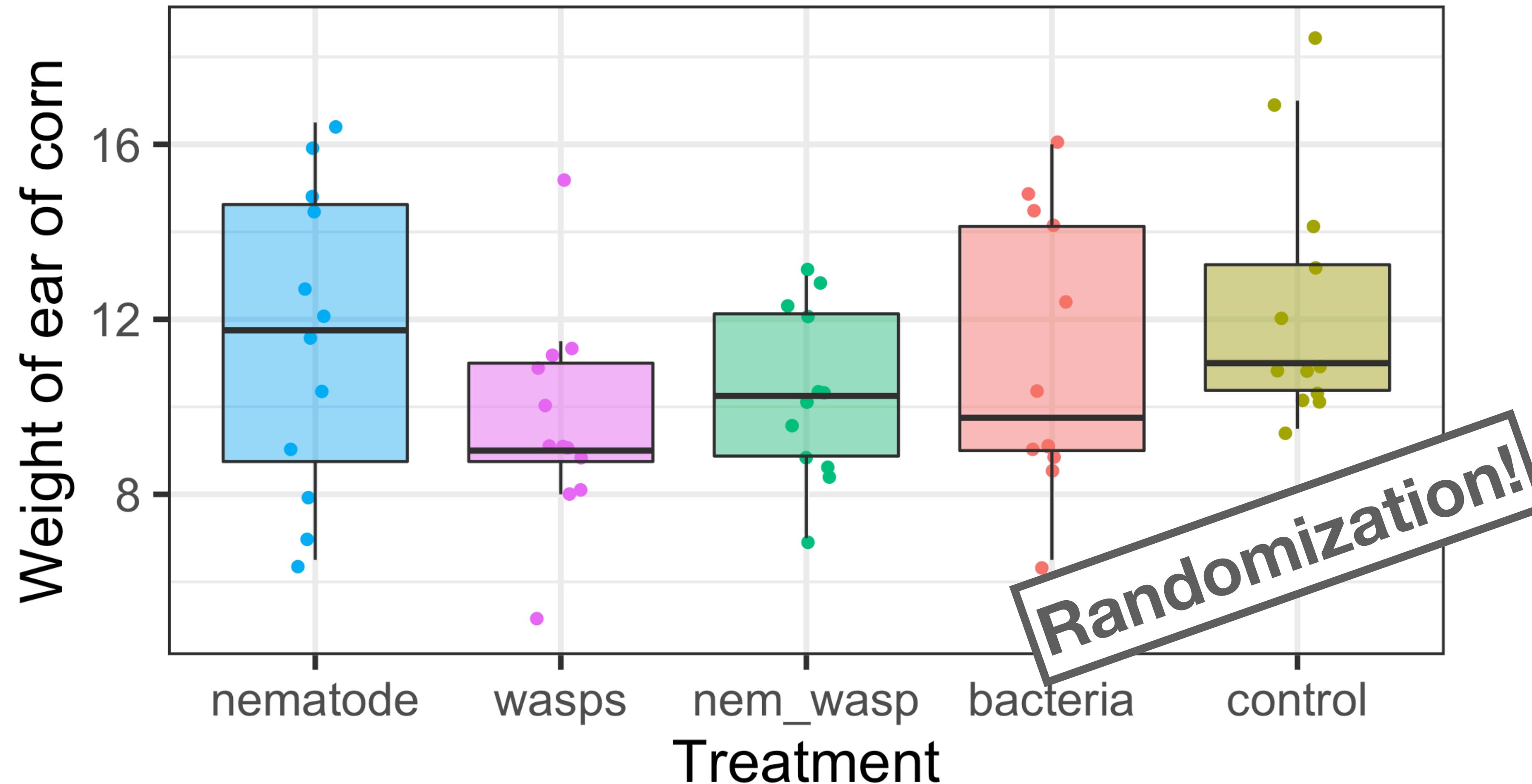
Comparing means of many samples

Q: Is there a difference between these groups?



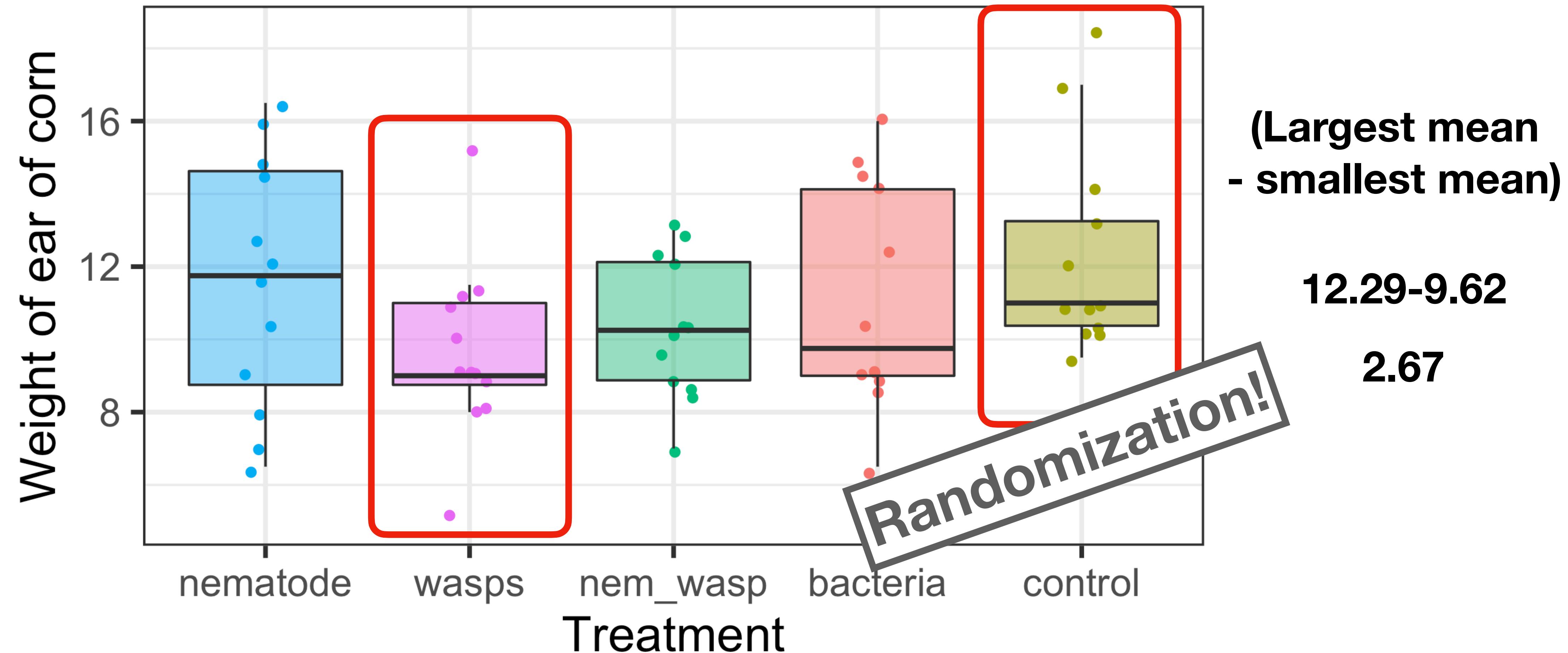
Comparing means of many samples

Q: Is there a difference between these groups?



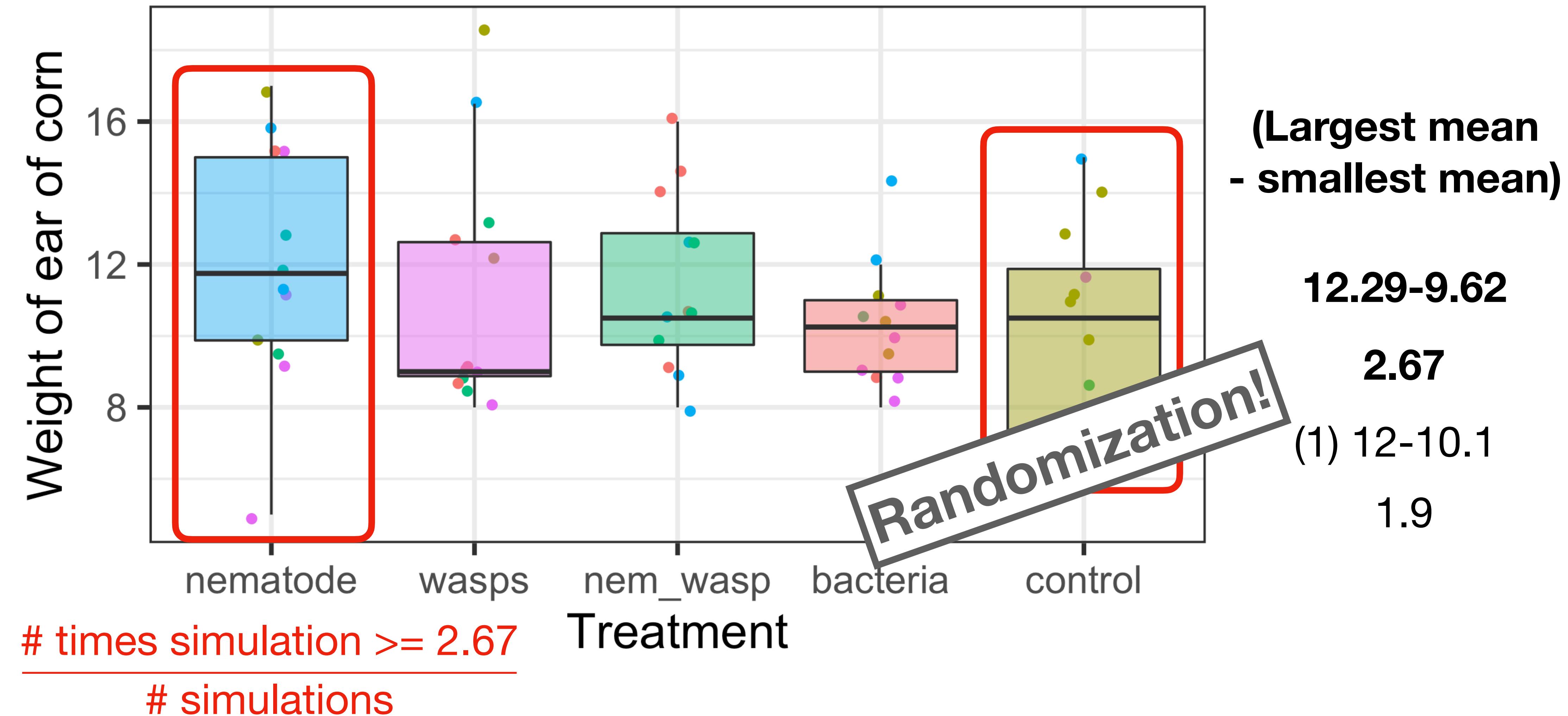
Comparing means of many samples

Q: Is there a difference between these groups?



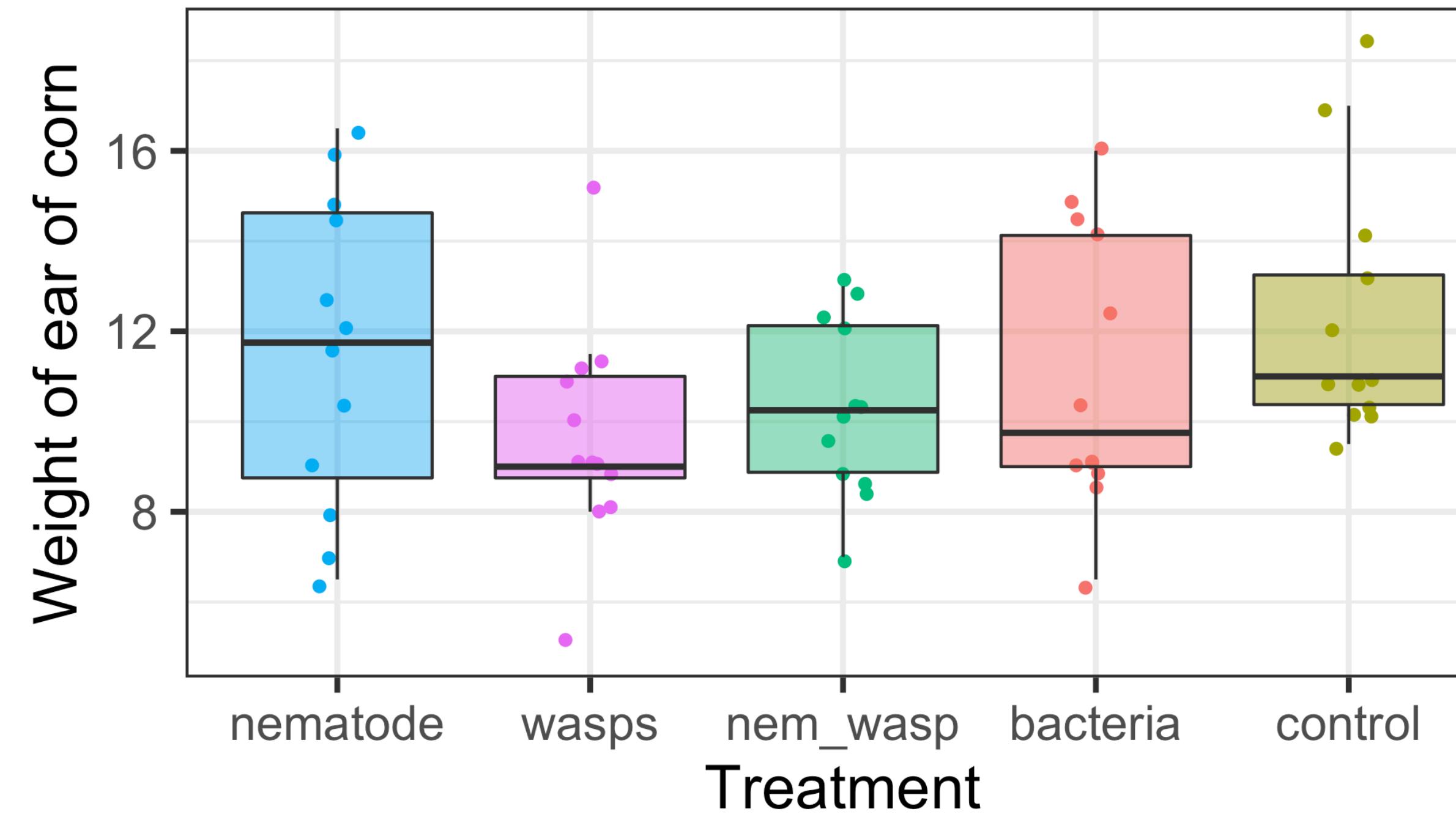
Comparing means of many samples

Q: Is there a difference between these groups?



Comparing means of many samples

ANOVA (analysis of variance)

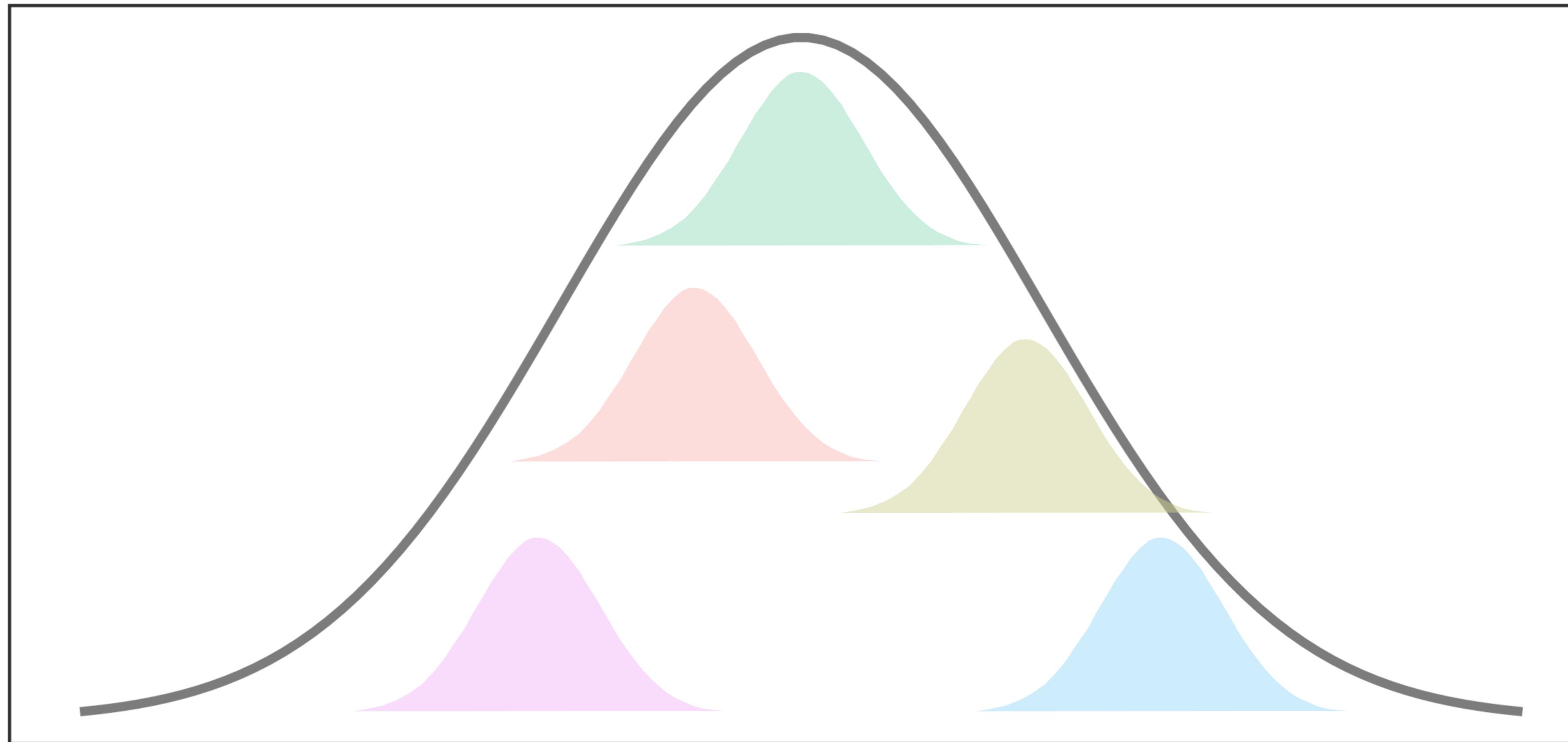


$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

H_A : Null is false (i.e. at least some of the means are different)

Comparing means of many samples

ANOVA (analysis of variance)

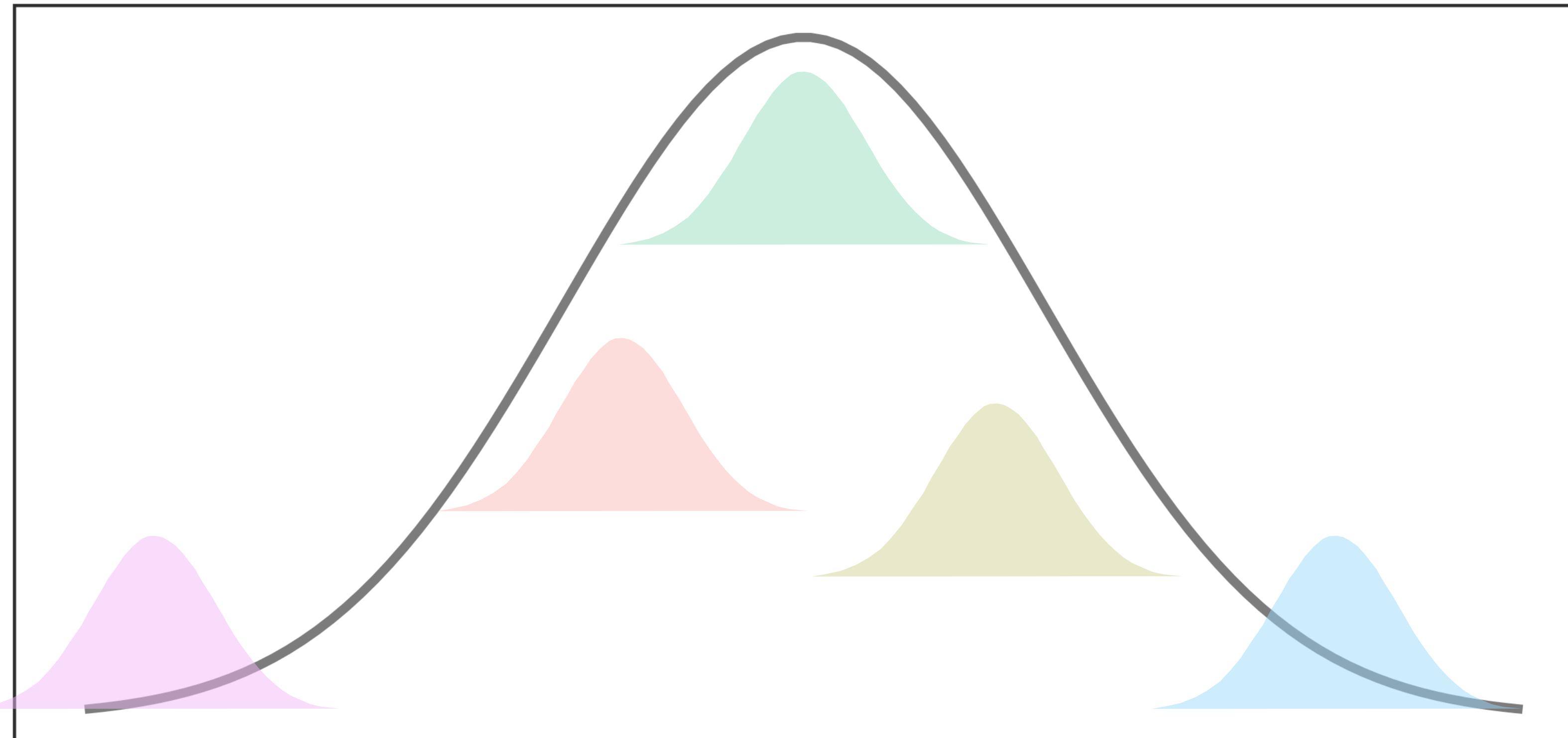


★ H_0 : All groups came from same population distribution

H_A : Null is false (i.e. at least some of the means are different)

Comparing means of many samples

ANOVA (analysis of variance)



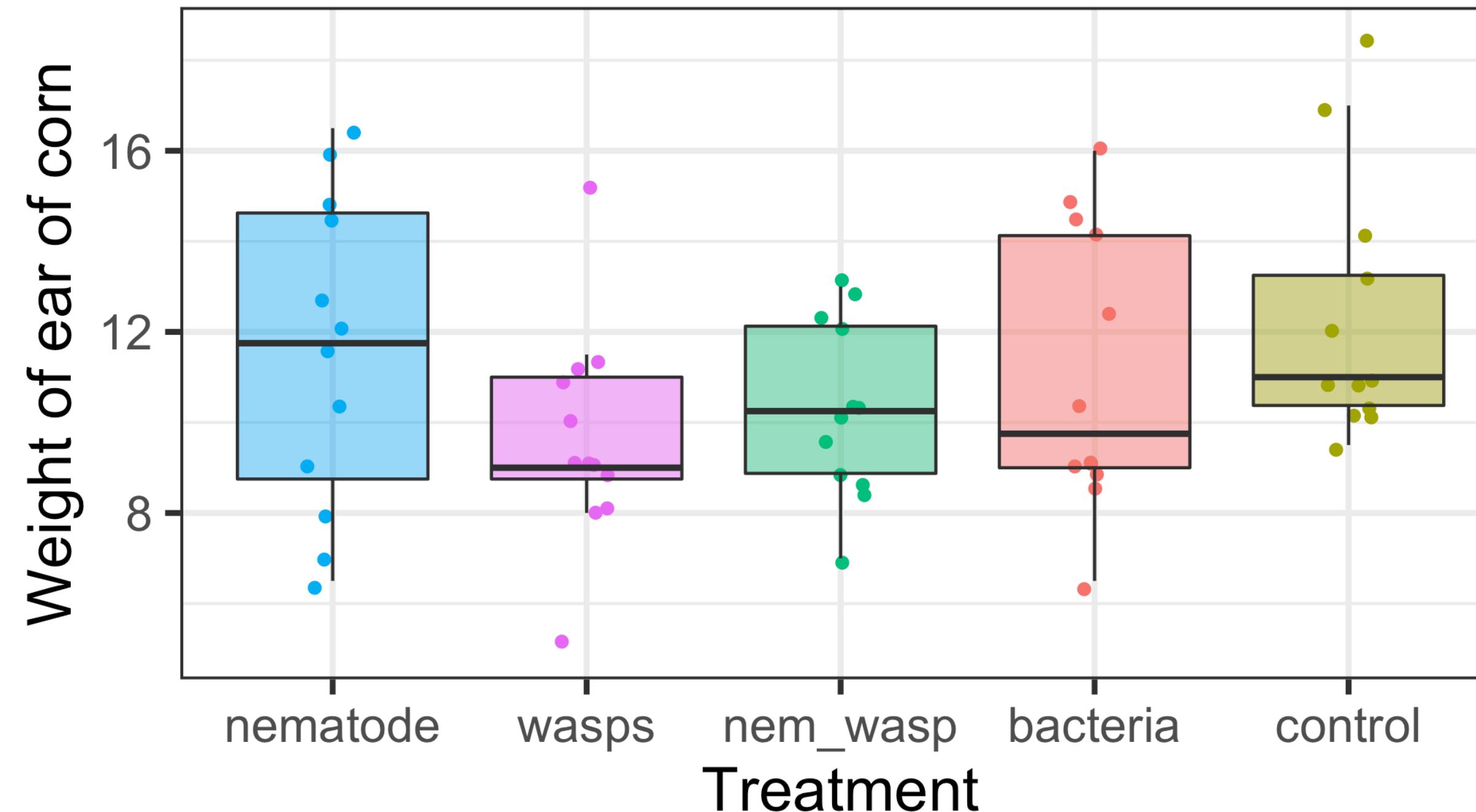
H_0 : All groups came from same population distribution

★ H_A : Null is false (i.e. at least some of the means are different)

Comparing means of many samples

X Pairwise t -tests?

Problem of multiple comparisons!



$$P(+) = 1 - (1 - 0.05)^n$$

# Groups	# Tests	P(+)
2	1	0.05
3	3	0.142
4	6	0.264
5	10	0.401

$$H_0 : \mu_1 = \mu_2$$

$$H_0 : \mu_1 = \mu_4$$

$$H_0 : \mu_2 = \mu_3$$

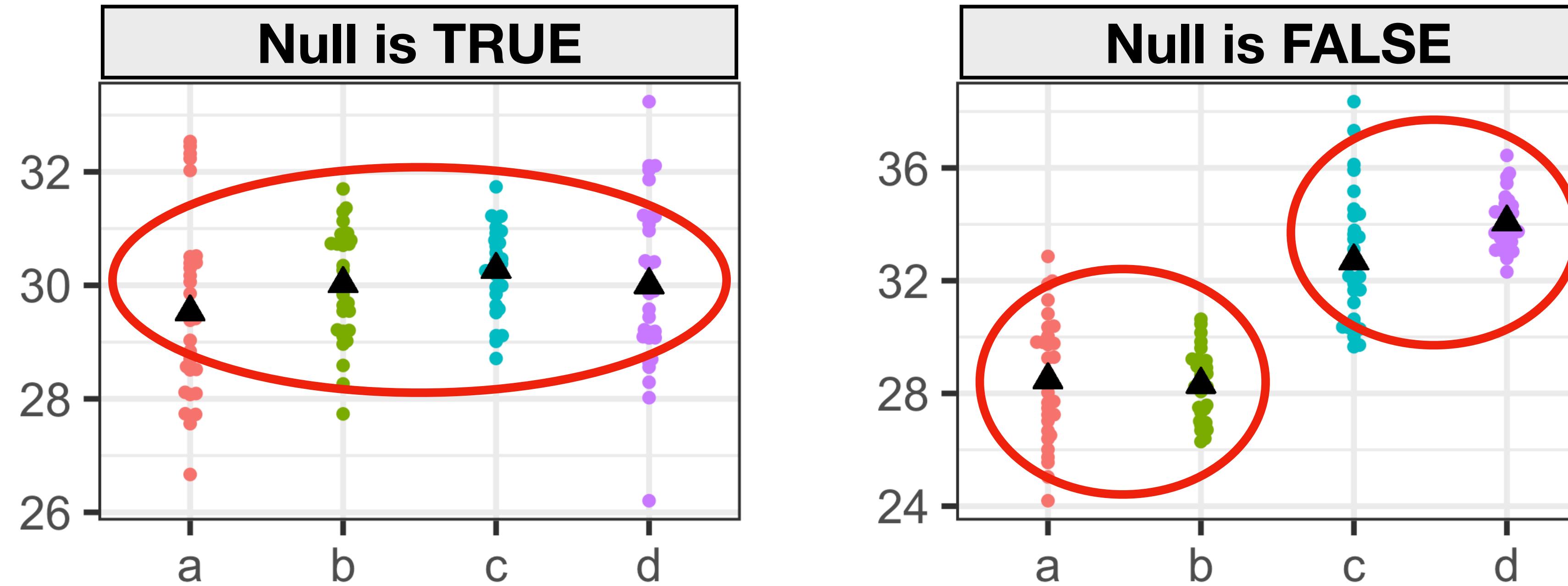
...

$$H_0 : \mu_1 = \mu_3$$

$$H_0 : \mu_1 = \mu_5$$

$$H_0 : \mu_2 = \mu_4$$

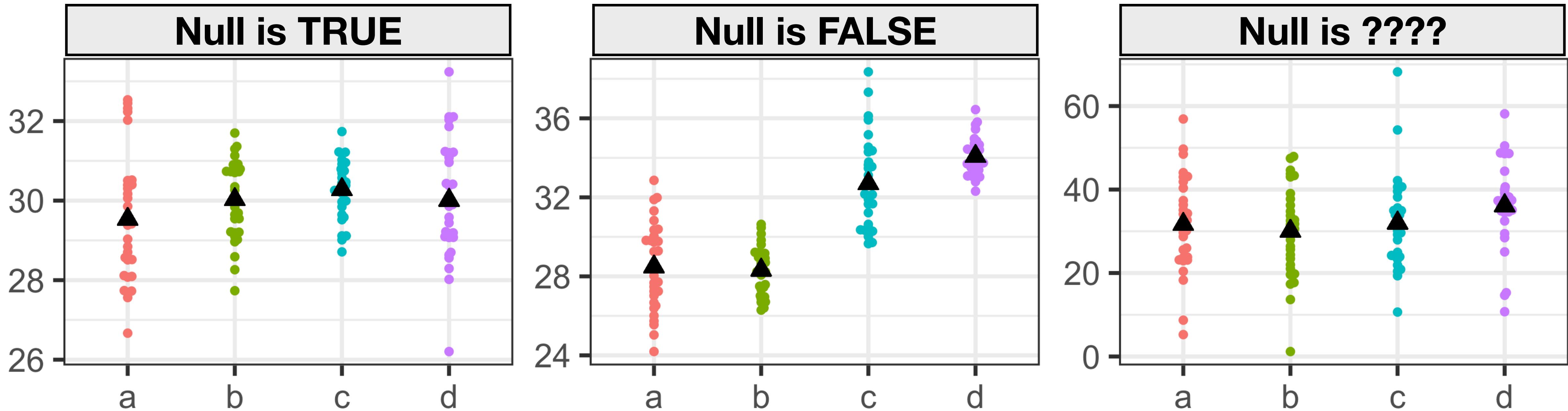
A graphical view of ANOVA



$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

H_A : Null is false (i.e. at least some of the means are different)

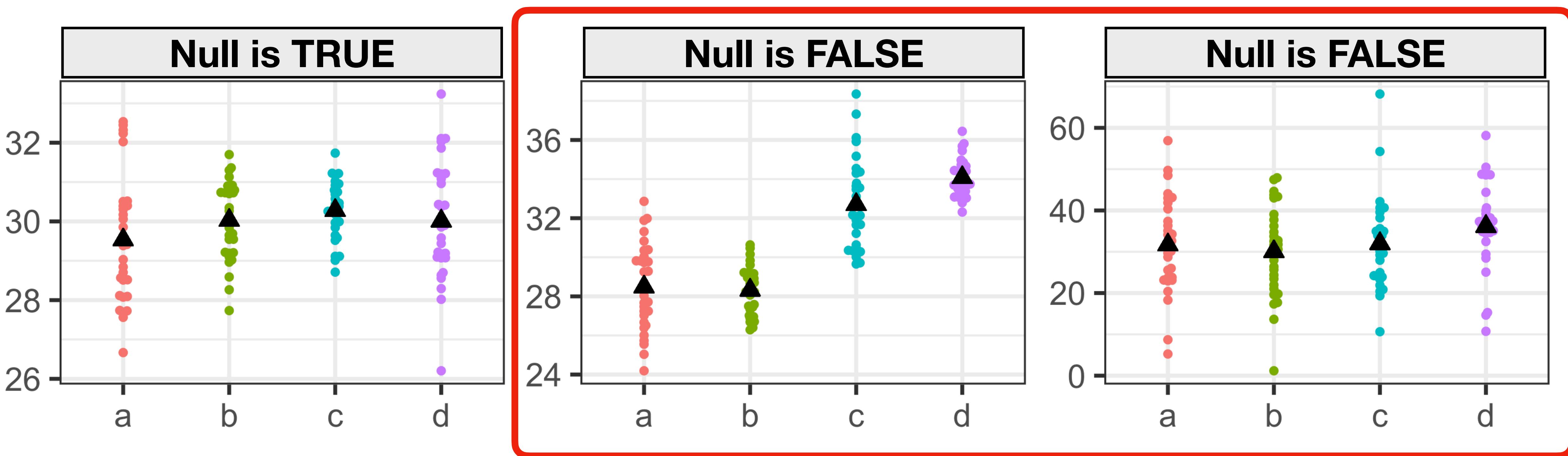
A graphical view of ANOVA



$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

H_A : Null is false (i.e. at least some of the means are different)

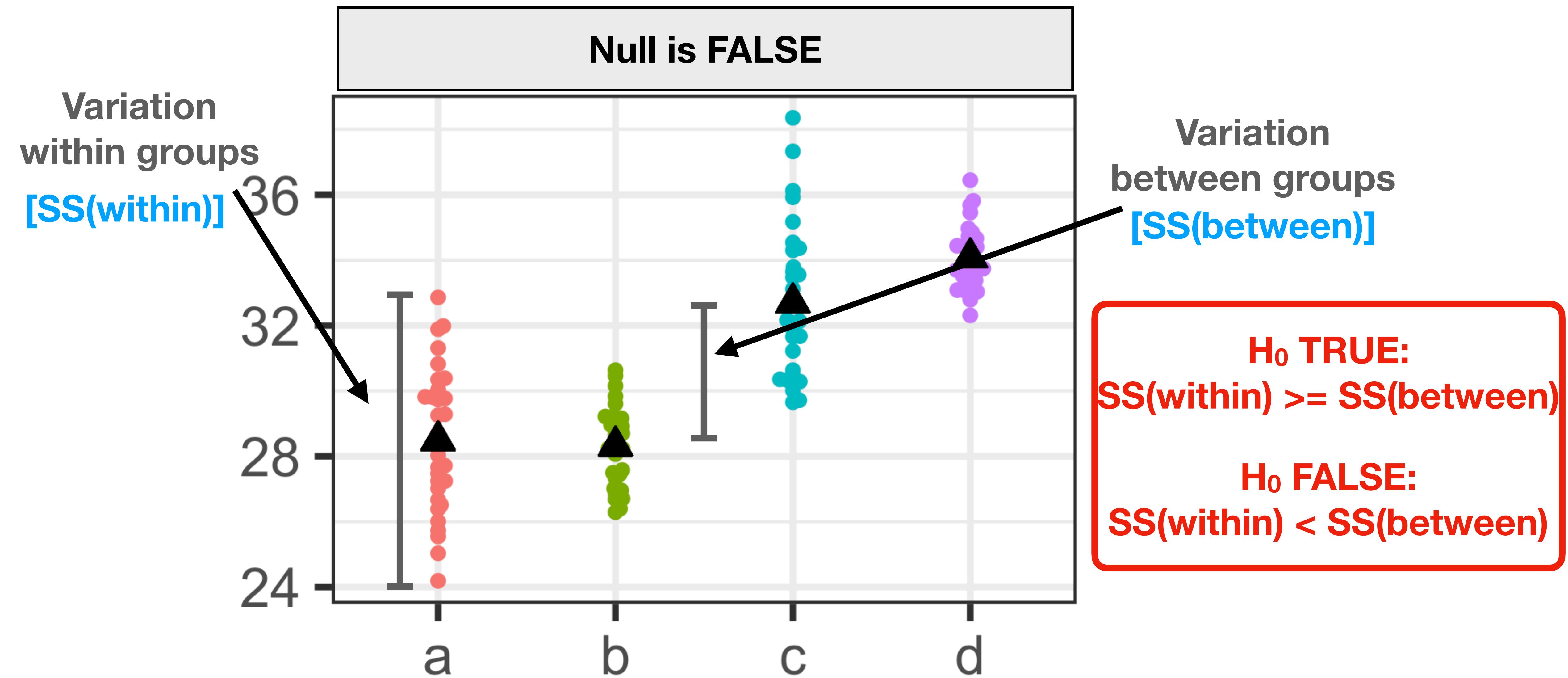
A graphical view of ANOVA



$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

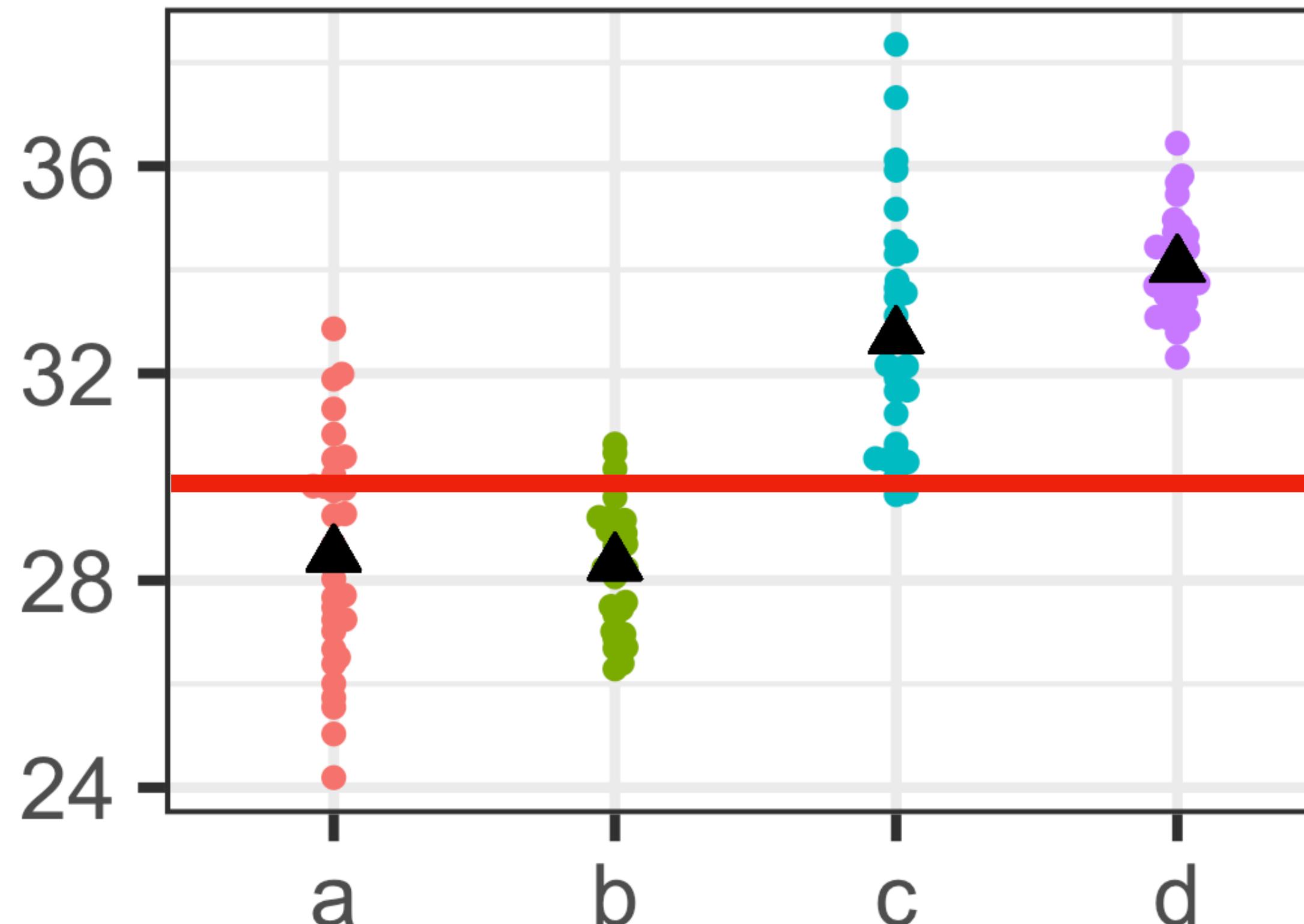
H_A : Null is false (i.e. at least some of the means are different)

“Analysis of variance” (ANOVA)



But first, some notation

Grand mean ($\bar{\bar{y}}$) [Mean of all the observations]



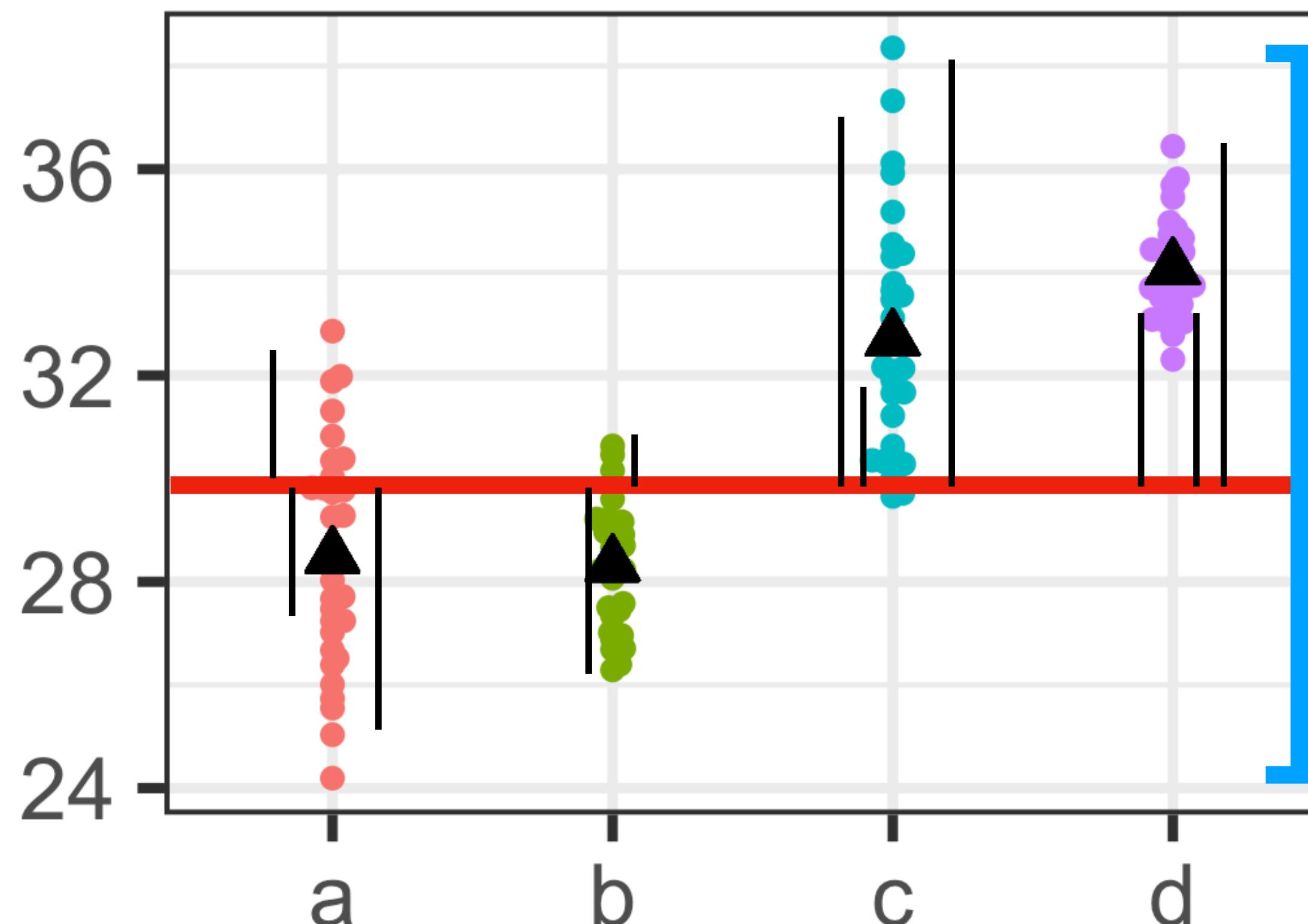
$$= \frac{\text{Sum of all values}}{\text{Total number of values}}$$

= *Mean of group means*
IF group n is same...

But first, some notation

Total sum of squares (SST)

(df = total observations - 1)



[Total variability among all groups & observations]

$$SST = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$$

= Sum of squared deviations between each data point and the grand mean

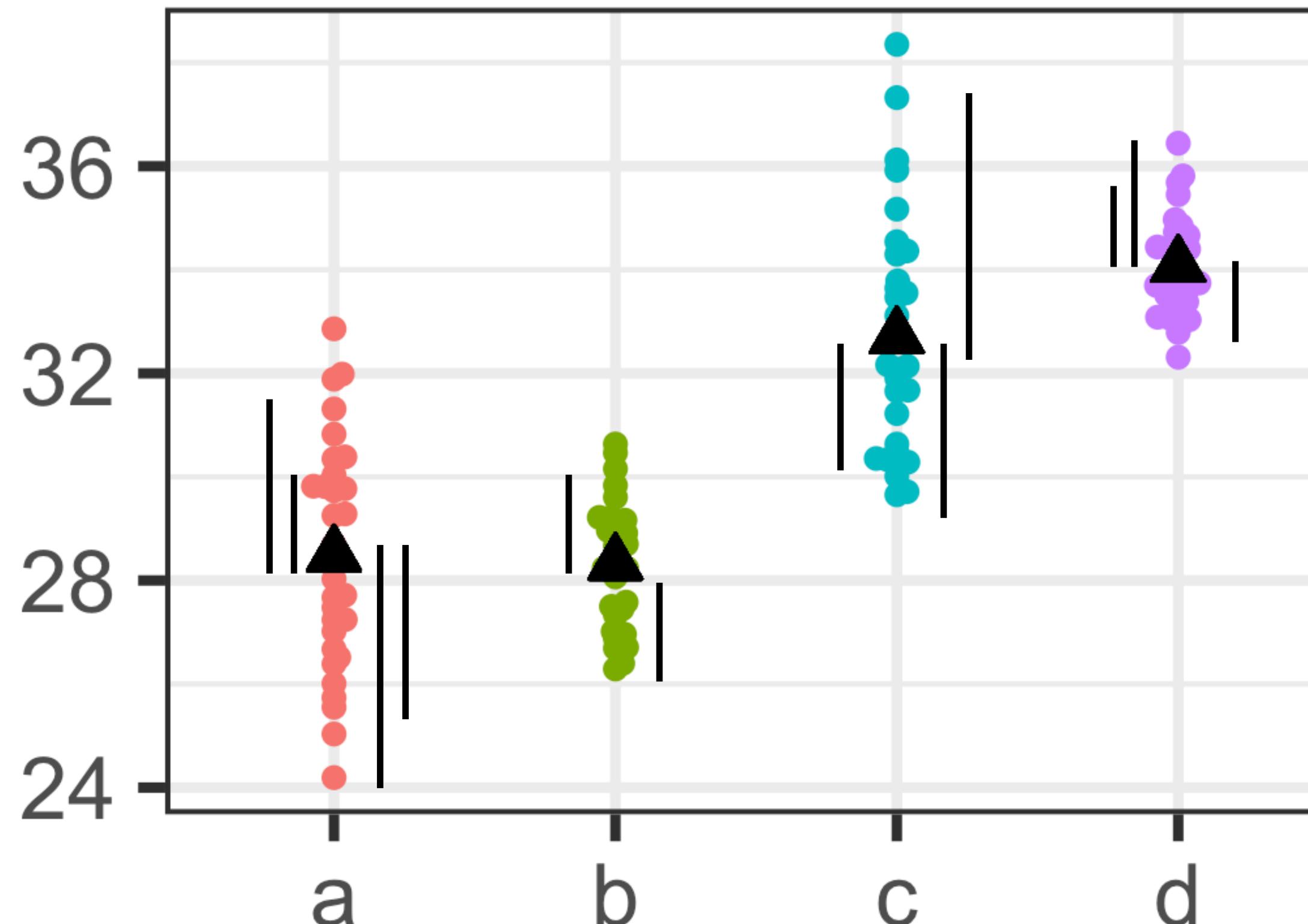
$$SST = SSW + SSB$$

= Sum of squared WITHIN + sum of square BETWEEN

But first, some notation

Sum of squares within (SSW) [Variability within groups]

(df = total observations - number of groups)



$$SSW = \sum_{i=1}^I \sum_{j=1}^{n_j} (y_{ij} - \bar{y}_i)^2$$

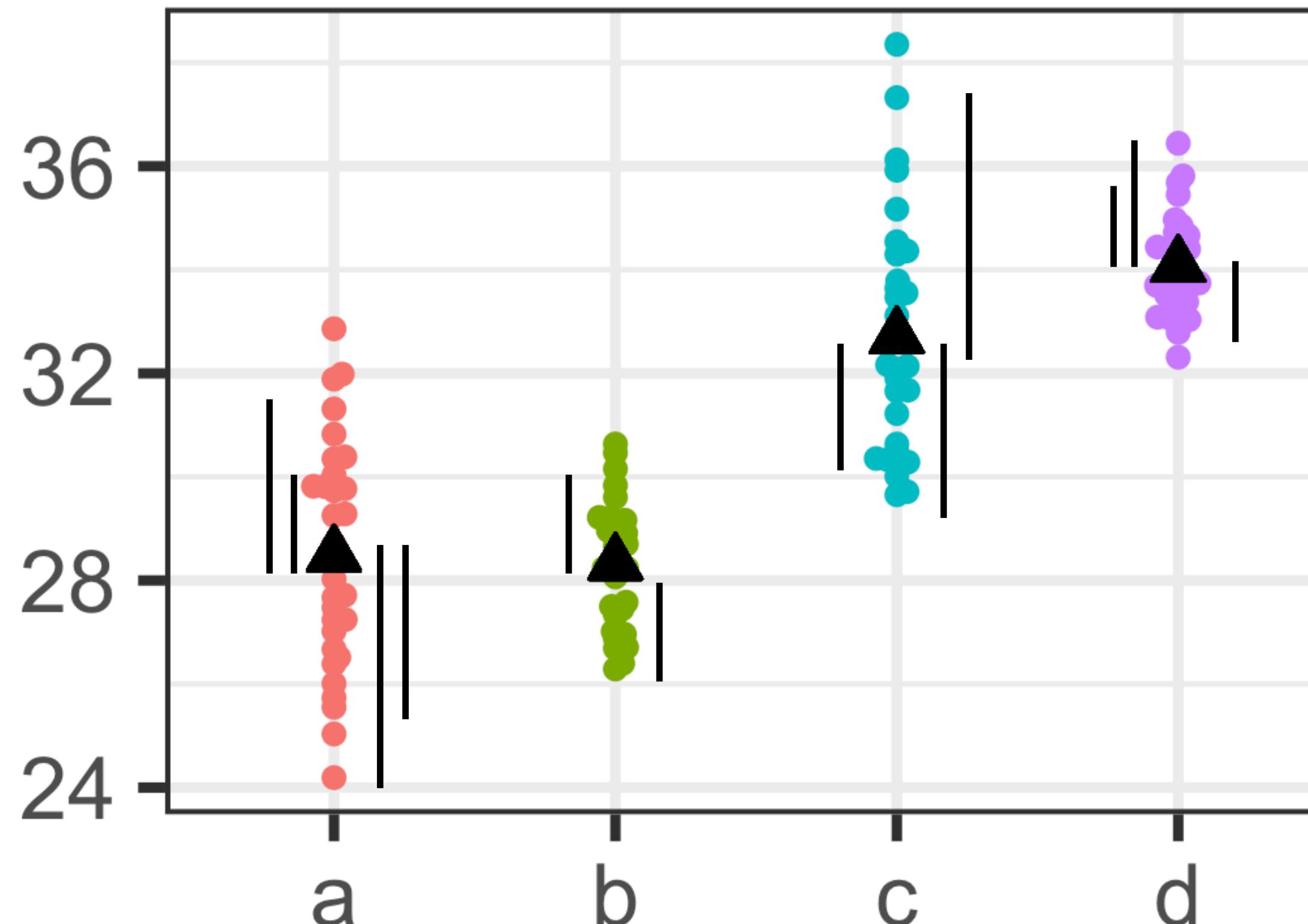
= Sum of squared deviations between
each data point and their group mean

$$SSW = \sum_{i=1}^I (n_i - 1) s_i^2 \frac{(y_i - \bar{y})^2}{n - 1}$$

But first, some notation

Sum of squares within (SSW) [Variability within groups]

(df = total observations - number of groups)



$$SSW = \sum_{i=1}^I \sum_{j=1}^{n_j} (y_{ij} - \bar{y}_i)^2$$

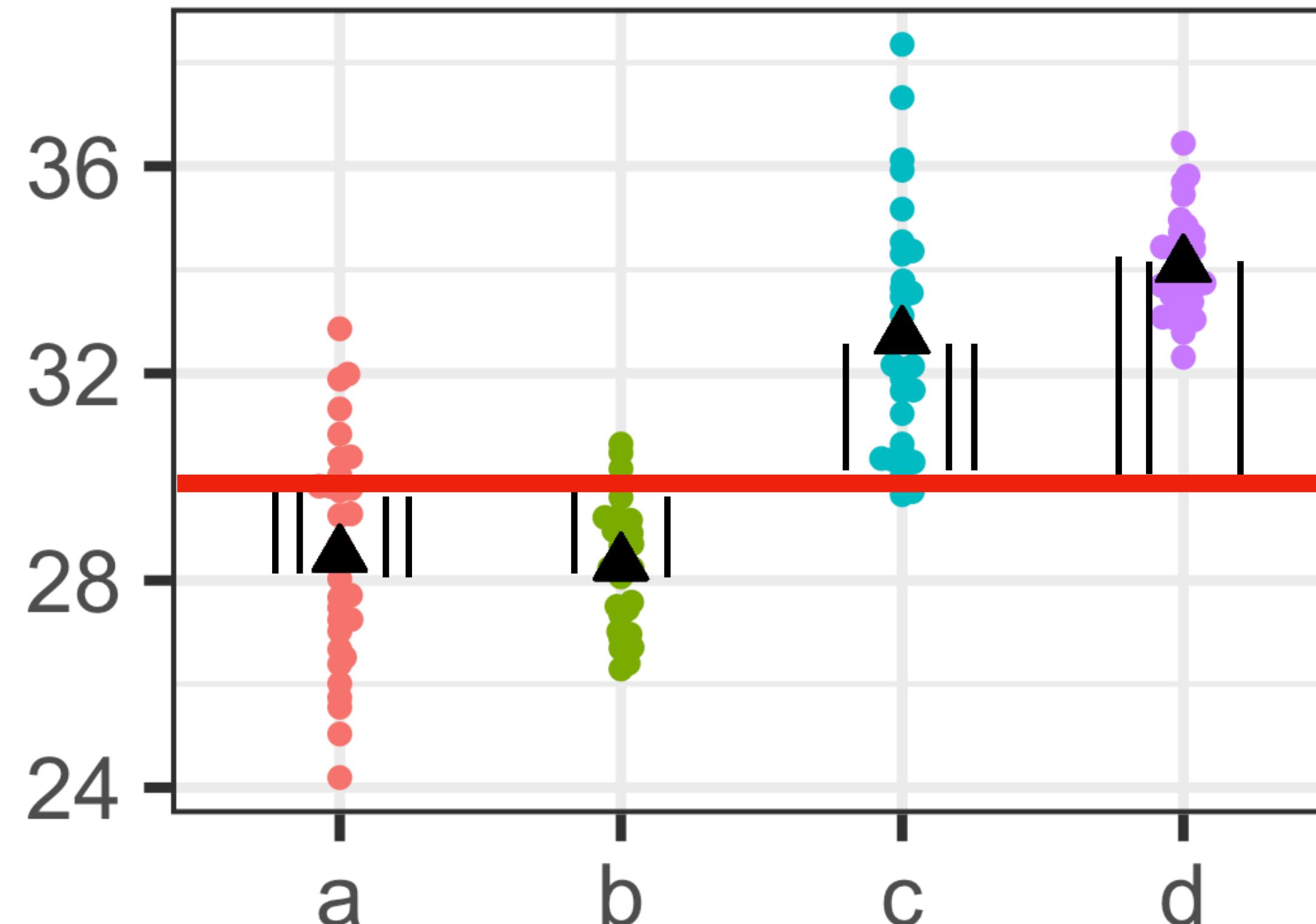
= Sum of squared deviations between
each data point and their group mean

$$SSW = \sum_{i=1}^I (n_i - 1) \frac{(y_i - \bar{y})^2}{n - 1}$$

But first, some notation

Sum of squares between (SSB) [Variability between groups]

(df = total groups - 1)



$$SSB = \sum_{i=1}^I \sum_{j=1}^{n_j} (\bar{y}_i - \bar{\bar{y}})^2$$

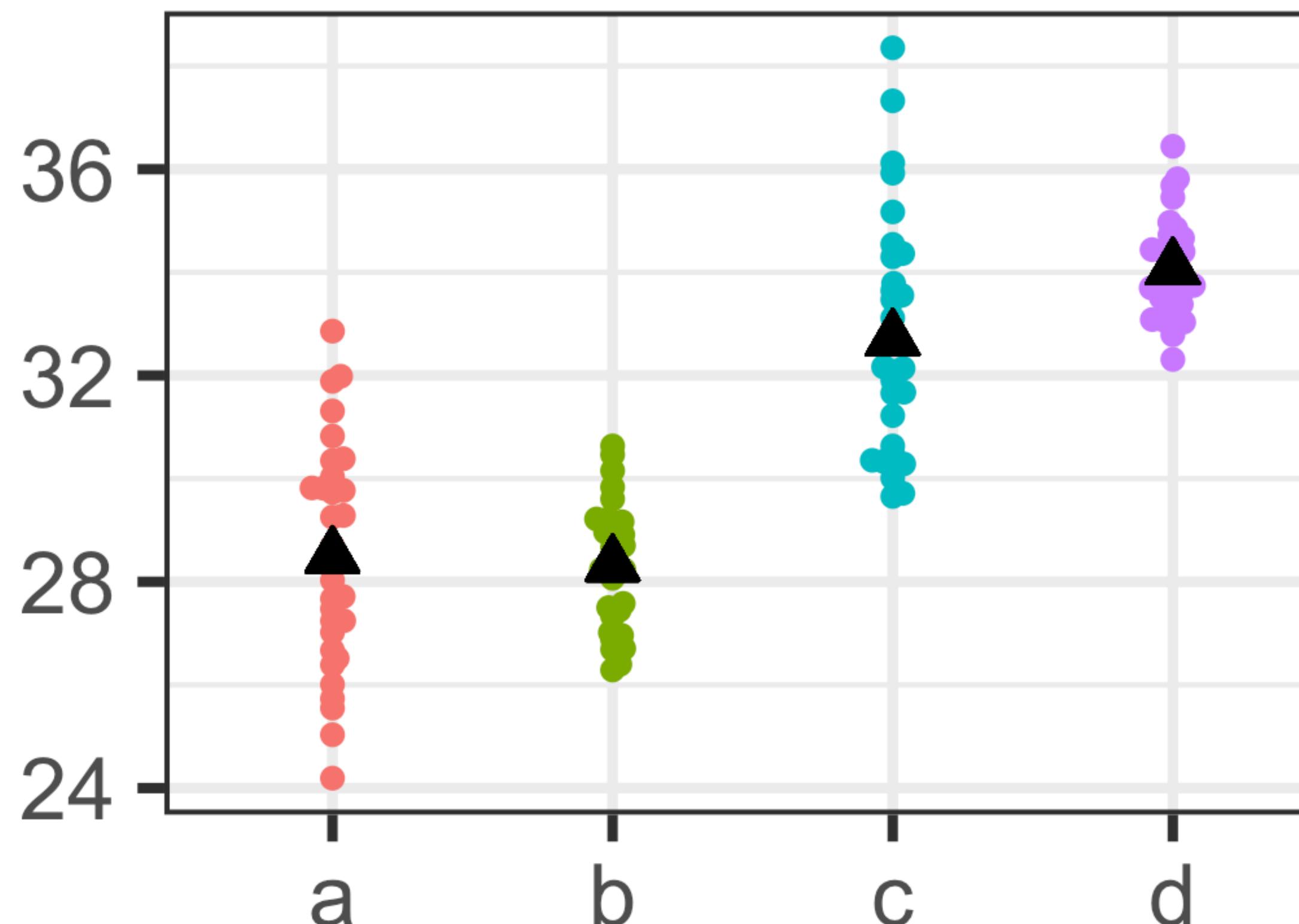
= Sum of squared deviations between
each group mean and the grand mean
(repeated for each data point)

$$SSB = \sum_{i=1}^I n_i (\bar{y}_i - \bar{\bar{y}})^2$$

But first, some notation

Mean square (MS)

[Ratio of SS / degrees of freedom]



$$MS = \frac{SS}{df}$$

= Variation between (MSB) or within (MSW) groups

But first, some notation

Understand the concept but don't need to memorize these equations

Source	df	ss (Sum of squares)	MS (Mean square)
Between groups	Groups - 1	$SSB = \sum_{i=1}^I n_i(\bar{y}_i - \bar{\bar{y}})^2$	SSB/df
Within groups	Observations - Groups	$SSW = \sum_{i=1}^I (n_i - 1)s_i^2$	SSW/df
Total	Observations - 1	$SST = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{\bar{y}})^2$	

Let's try out some real data...

Data: weight gain for three different diets

	Diet 1	Diet 2	Diet 3
	8	9	15
	16	16	10
	9	21	17
		11	6
		18	

Let's try out some real data...

Data: weight gain for three different diets

	Diet 1	Diet 2	Diet 3
	8	9	15
	16	16	10
	9	21	17
		11	6
		18	
n_i	3	5	4
Sum = $\sum y_{ij}$			
Mean = \bar{y}_i			
SD = s_i			

Let's try out some real data...

Data: weight gain for three different diets

	Diet 1	Diet 2	Diet 3
8	9	15	
16	16	10	
9	21	17	
	11	6	
	18		
n_i	3	5	4
Sum = $\sum y_{ij}$			
Mean = \bar{y}_i			
SD = s_i			

Sum:

$$8 + 16 + 9 = 33$$

Mean:

$$33 / 3 = 11$$

SD:

$$\sqrt{\frac{(8 - 11)^2 + (16 - 11)^2 + (9 - 11)^2}{3 - 1}}$$

$$= 4.359$$

Let's try out some real data...

Data: weight gain for three different diets

	Diet 1	Diet 2	Diet 3
8	9	15	
16	16	10	
9	21	17	
	11	6	
	18		
n_i	3	5	4
Sum = $\sum y_{ij}$	33		
Mean = \bar{y}_i	11		
SD = s_i	4.359		

Sum:

$$8 + 16 + 9 = 33$$

Mean:

$$33 / 3 = 11$$

SD:

$$\sqrt{\frac{(8 - 11)^2 + (16 - 11)^2 + (9 - 11)^2}{3 - 1}}$$

$$= 4.359$$

Let's try out some real data...

Data: weight gain for three different diets

	Diet 1	Diet 2	Diet 3
8	9	15	
16	16	10	
9	21	17	
	11	6	
	18		
n_i	3	5	4
$\text{Sum} = \sum y_{ij}$	33	75	48
$\text{Mean} = \bar{y}_i$	11	15	12
$\text{SD} = s_i$	4.359	4.950	4.967

Sum:

$$8 + 16 + 9 = 33$$

Mean:

$$33 / 3 = 11$$

SD:

$$\sqrt{\frac{(8 - 11)^2 + (16 - 11)^2 + (9 - 11)^2}{3 - 1}}$$

$$= 4.359$$

Let's try out some real data...

Source	df	ss (Sum of squares)	MS (Mean square)
Between groups	$3 - 1 = 2$	$SSB = \sum_{i=1}^I n_i(\bar{y}_i - \bar{\bar{y}})^2$	SSB/df
Within groups	$12 - 3 = 9$	$SSW = \sum_{i=1}^I (n_i - 1)s_i^2$	SSW/df
Total	$12 - 1 = 11$	$SST = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{\bar{y}})^2$	

Let's try out some real data...

	Diet 1	Diet 2	Diet 3
8	9	15	
16	16	10	
9	21	17	
	11	6	
	18		
n_i	3	5	4
Sum = $\sum y_{ij}$	33	75	48
Mean = \bar{y}_i	11	15	12
SD = s_i	4.359	4.950	4.967

$$SSW = \sum_{i=1}^I (n_i - 1)s_i^2$$

$$(3 - 1)(4.359)^2 + (5-1)(4.950)^2 + \\ (4-1)(4.967)^2$$

$$= 210.025$$

Let's try out some real data...

Source	df	ss (Sum of squares)	MS (Mean square)
Between groups	$3 - 1 = 2$	$SSB = \sum_{i=1}^I n_i(\bar{y}_i - \bar{\bar{y}})^2$	SSB/df
Within groups	$12 - 3 = 9$	210.025	$210.025/9 = 23.336$
Total	$12 - 1 = 11$	$SST = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{\bar{y}})^2$	

Let's try out some real data...

	Diet 1	Diet 2	Diet 3
8	9	15	
16	16	10	
9	21	17	
	11	6	
	18		
n_i	3	5	4
Sum = $\sum y_{ij}$	33	75	48
Mean = \bar{y}_i	11	15	12
SD = s_i	4.359	4.950	4.967

$$SSB = \sum_{i=1}^I n_i(\bar{y}_i - \bar{\bar{y}})^2$$

Grand mean:

$$(33 + 75 + 48) / (3 + 5 + 4) = 13$$

$$(3)(11-13)^2 + (5)(15-13)^2 + (4)(12-13)^2$$

$$= 36$$

Let's try out some real data...

Source	df	ss (Sum of squares)	MS (Mean square)
Between groups	$3 - 1 = 2$	36	$36/2 = 18$
Within groups	$12 - 3 = 9$	210.025	$210.025/9 = 23.336$
Total	$12 - 1 = 11$	$SST = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$	

Let's try out some real data...

	Diet 1	Diet 2	Diet 3
	8	9	15
	16	16	10
	9	21	17
		11	6
		18	
n_i	3	5	4
Sum = $\sum y_{ij}$	33	75	48
Mean = \bar{y}_i	11	15	12
SD = s_i	4.359	4.950	4.967

$$SST = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$$

$$(SST = SSW + SSB = 246.025)$$

$$(8-13)^2 + (16-13)^2 + (9-13)^2$$

$$(9-13)^2 + (16-13)^2 + (21-13)^2 +$$

$$(11-13)^2 + (18-13)^2 + (15-13)^2 +$$

$$(10-13)^2 + (17-13)^2 + (6-13)^2$$

$$= 246$$



Let's try out some real data...

Source	df	ss (Sum of squares)	MS (Mean square)
Between groups	$3 - 1 = 2$	36	$36/2 = 18$
Within groups	$12 - 3 = 9$	210.025	$210.025/9 = 23.336$
Total	$12 - 1 = 11$	246	

Let's try out some real data...

	Diet 1	Diet 2	Diet 3
8	9	15	
16	16	10	
9	21	17	
	11	6	
	18		
n_i	3	5	4
$\text{Sum} = \sum y_{ij}$	33	75	48
$\text{Mean} = \bar{y}_i$	11	15	12
$SD = S_i$	4.359	4.950	4.967

Pooled variance:

$$\text{MS(within)}$$

Pooled standard deviation:

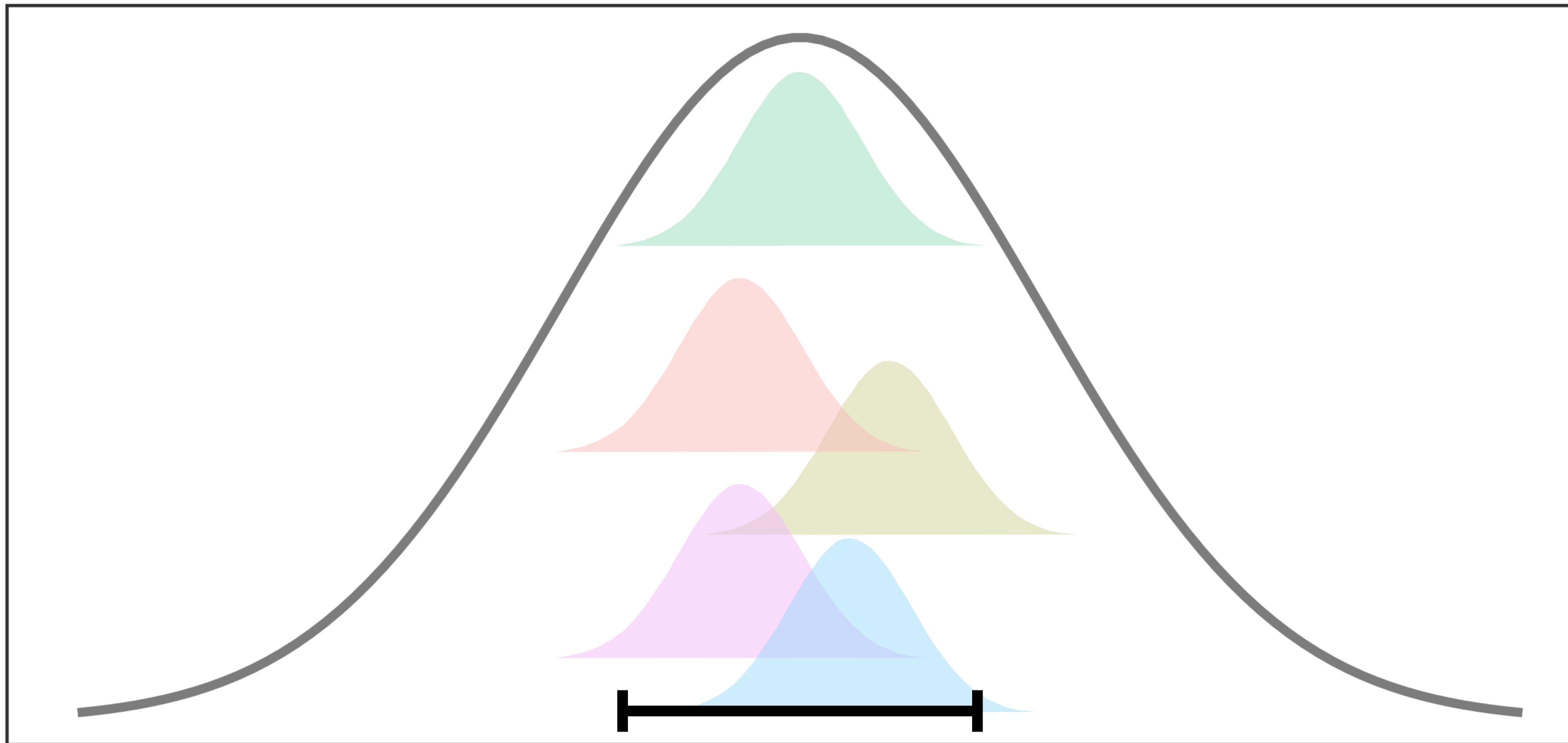
$$\sqrt{\text{MS}(within)}$$

$$\sqrt{23.336} = 4.831$$

If we assume the population SD is same for all 3 diets, we would estimate the common value to be 4.831.

Comparing means of many samples

ANOVA (analysis of variance)



★ H_0 : All groups came from same population distribution

H_A : Null is false (i.e. at least some of the means are different)

Example: filling out the ANOVA table

(1) Complete the table

Source	df	SS	MS
Between	3		45
Within	12	337	
Total			

Bonus: *How many groups were there in the study?
How many total observations?*

Example: filling out the ANOVA table

(1) Complete the table

Source	df	ss	MS
Between	3		45
Within	12	337	
Total	15		

$$\text{Total df} = \text{df(between)} + \text{df(within)}$$

$$\text{Total df} = 3 + 12 = 15$$

Bonus: How many groups were there in the study?
How many total observations?

Example: filling out the ANOVA table

(1) Complete the table

Source	df	SS	MS
Between	3	135	45
Within	12	337	28.08
Total	15		

$$MS = SS / df$$

$$\text{Between: } 45 = SS / 3$$

$$SS = 45 * 3 = 135$$

$$\text{Within: } MS = 337 / 12$$

$$MS = 28.08$$

Bonus: How many groups were there in the study?
How many total observations?

Example: filling out the ANOVA table

(1) Complete the table

Source	df	SS	MS
Between	3	135	45
Within	12	337	28.08
Total	15	472	

$$SST = SSB + SSW$$

$$SST = 135 + 337 = 472$$

Bonus: How many groups were there in the study?
How many total observations?

Example: filling out the ANOVA table

✓ 1) Complete the table

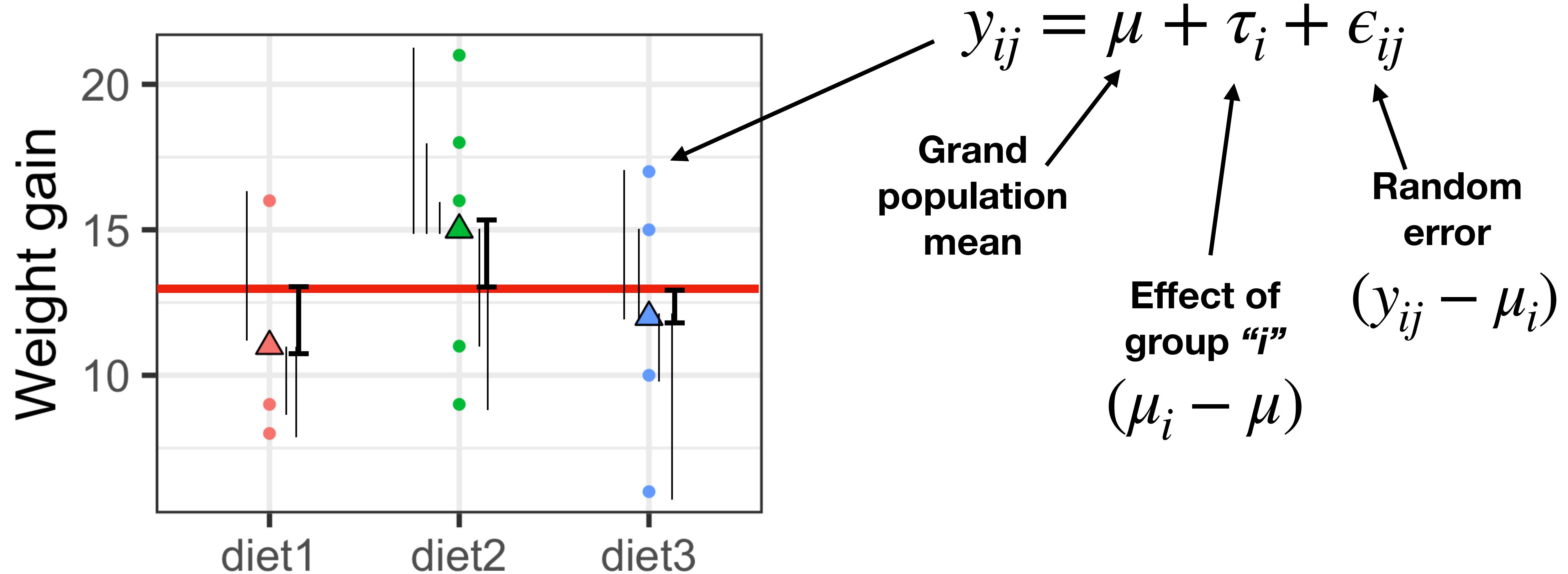
Source	df	SS	MS
Between	3	135	45
Within	12	337	28.08
Total	15	472	

✓ Bonus: How many **groups** were there in the study?
How many **total observations**?

df(between) = Groups - 1 → 4 groups

df(total) = Obs. - 1 → 16 obs.

The Analysis of Variance Model



But first, some notation

$$y_{ij} = \bar{\bar{y}} + (\bar{y}_i - \bar{\bar{y}}) + (y_{ij} - \bar{y}_i)$$
$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

Source	df	ss (Sum of squares)	MS (Mean square)
Between groups	Groups - 1	$SSB = \sum_{i=1}^I n_i(\bar{y}_i - \bar{\bar{y}})^2$	SSB/df
Within groups	Observations - Groups	$SSW = \sum_{i=1}^I (n_i - 1)s_i^2$	SSW/df
Total	Observations - 1	$SST = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{\bar{y}})^2$	

But first, some notation

$$y_{ij} = \bar{\bar{y}} + (\bar{y}_i - \bar{\bar{y}}) + (y_{ij} - \bar{y}_i)$$
$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

Source	df	ss (Sum of squares)	MS (Mean square)
Between groups	Groups - 1	$SSB = \sum_{i=1}^I n_i \tau_i^2$	SSB/df
Within groups	Observations - Groups	$SSW = \sum_{i=1}^I (n_i - 1) s_i^2$	SSW/df
Total	Observations - 1	$SST = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{\bar{y}})^2$	

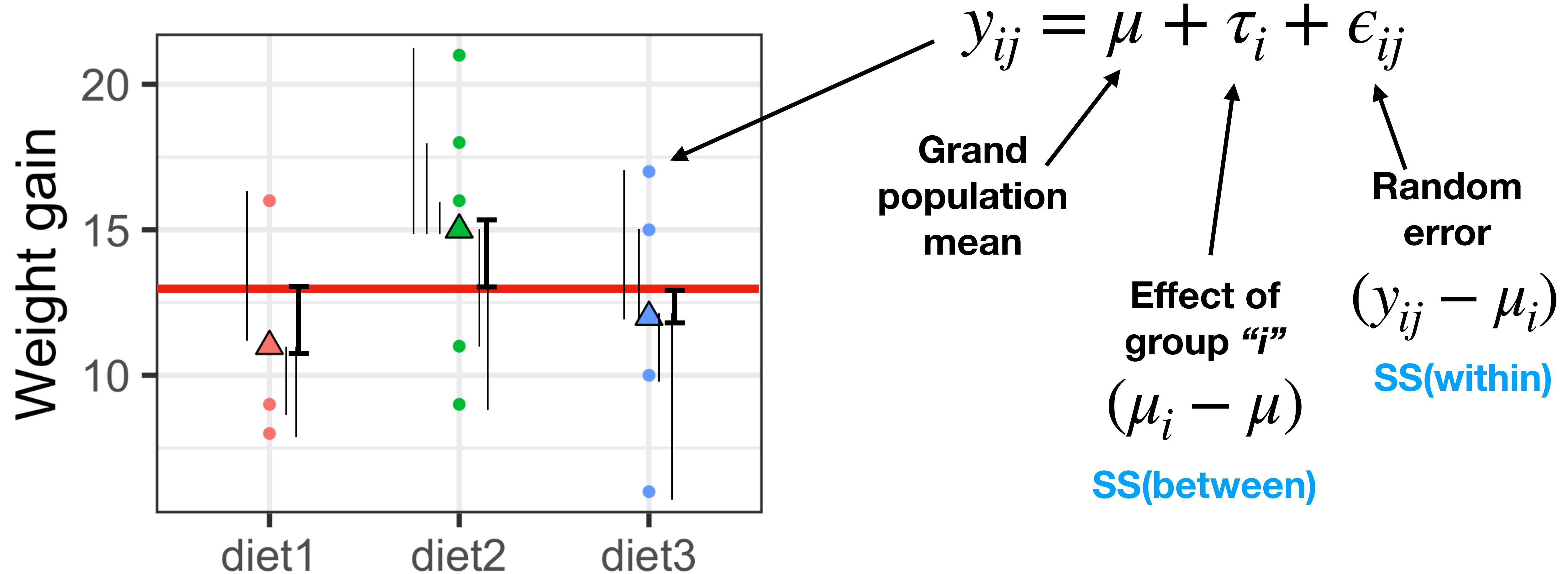
But first, some notation...

$$y_{ij} = \bar{\bar{y}} + (\bar{y}_i - \bar{\bar{y}}) + (y_{ij} - \bar{y}_i)$$

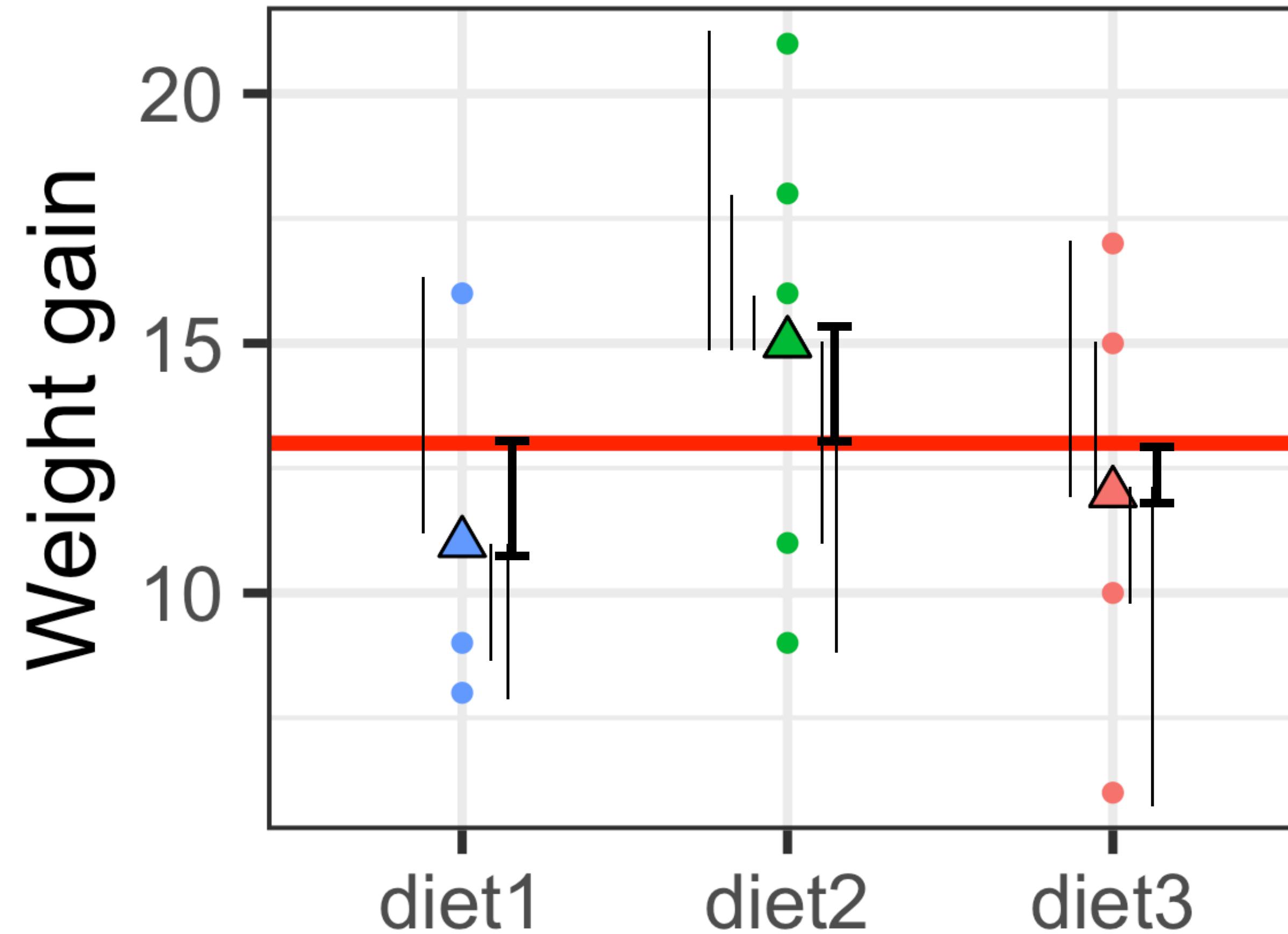
$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

Source	df	ss (Sum of squares)	MS (Mean square)
Between groups	Groups - 1	$SSB = \sum_{i=1}^I n_i \tau_i^2$	SSB/df
Within groups	Observations - Groups	$SSW = \sum_{i=1}^I \epsilon_{ij}^2$	SSW/df
Total	Observations - 1	$SST = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{\bar{y}})^2$	

The Analysis of Variance Model



The Analysis of Variance Model



$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$H_0 : \tau_1 = \tau_2 = \tau_3$$

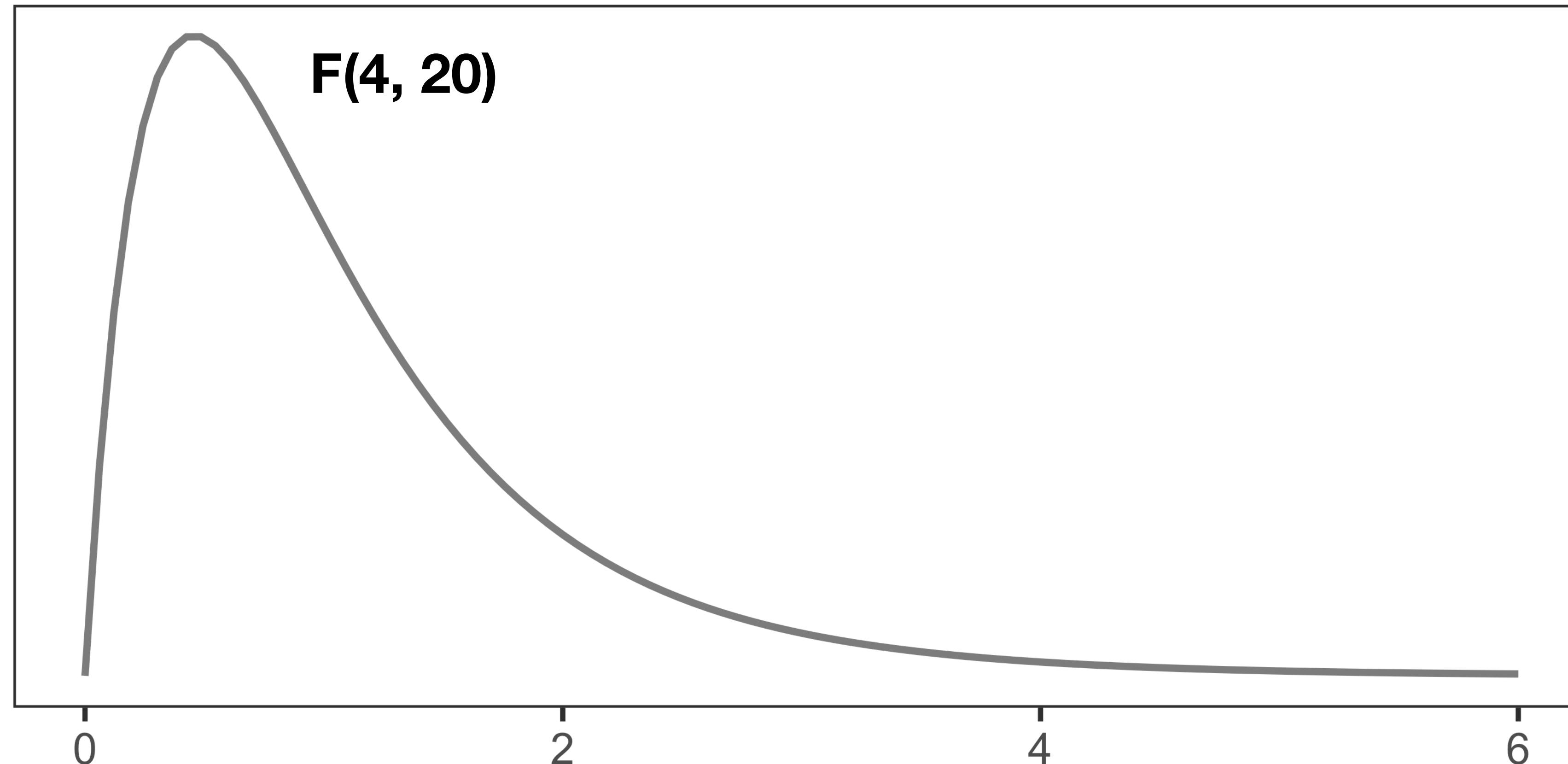
H_A : Null is false (i.e. at least some of the means are different)

“Compound null hypothesis”

Rejection of H_0 does not specify WHICH means are different...

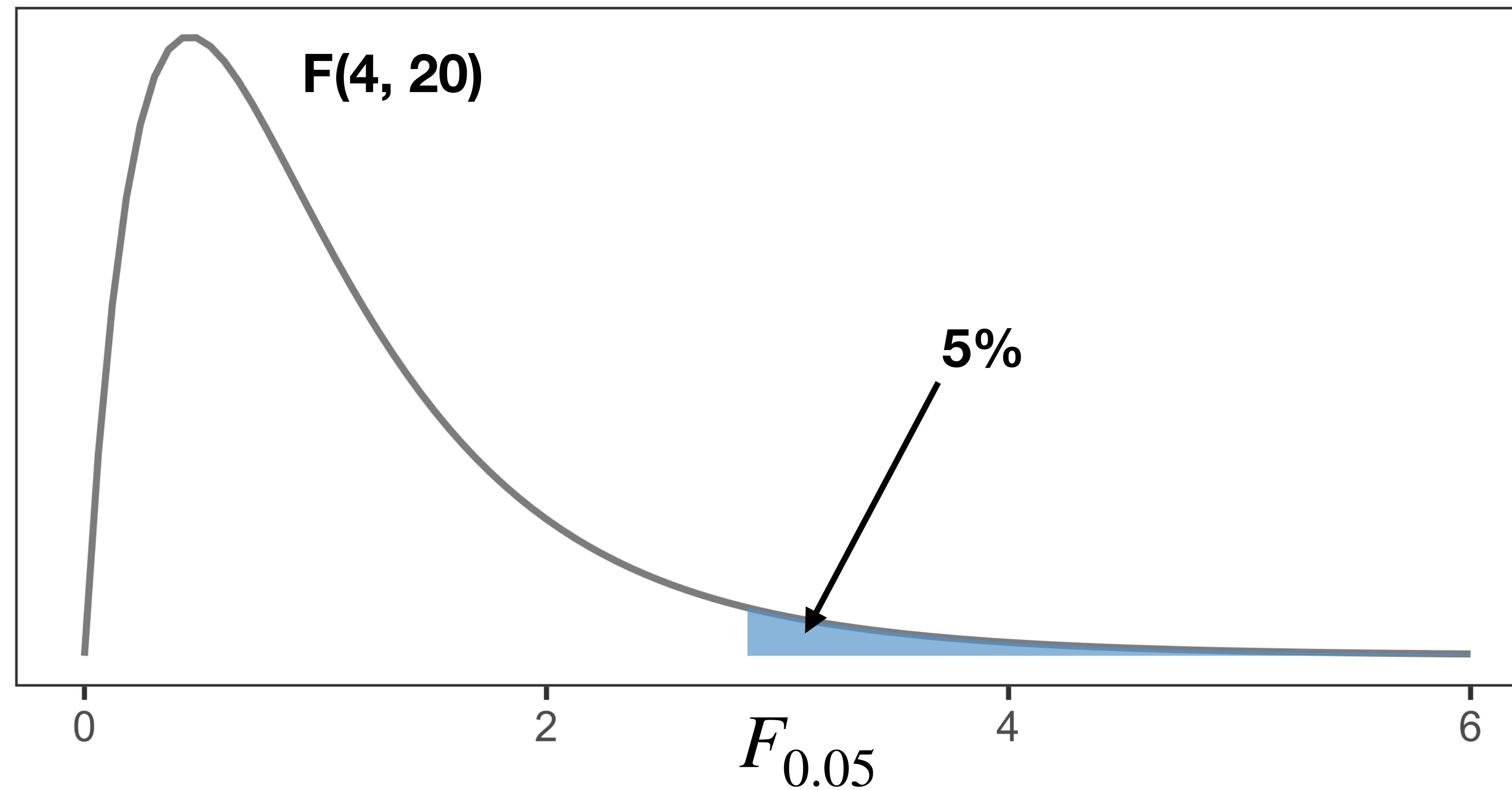
Introduction to the F distribution

Distribution used in several statistical analyses and depends on the numerator degrees of freedom and the denominator degrees of freedom



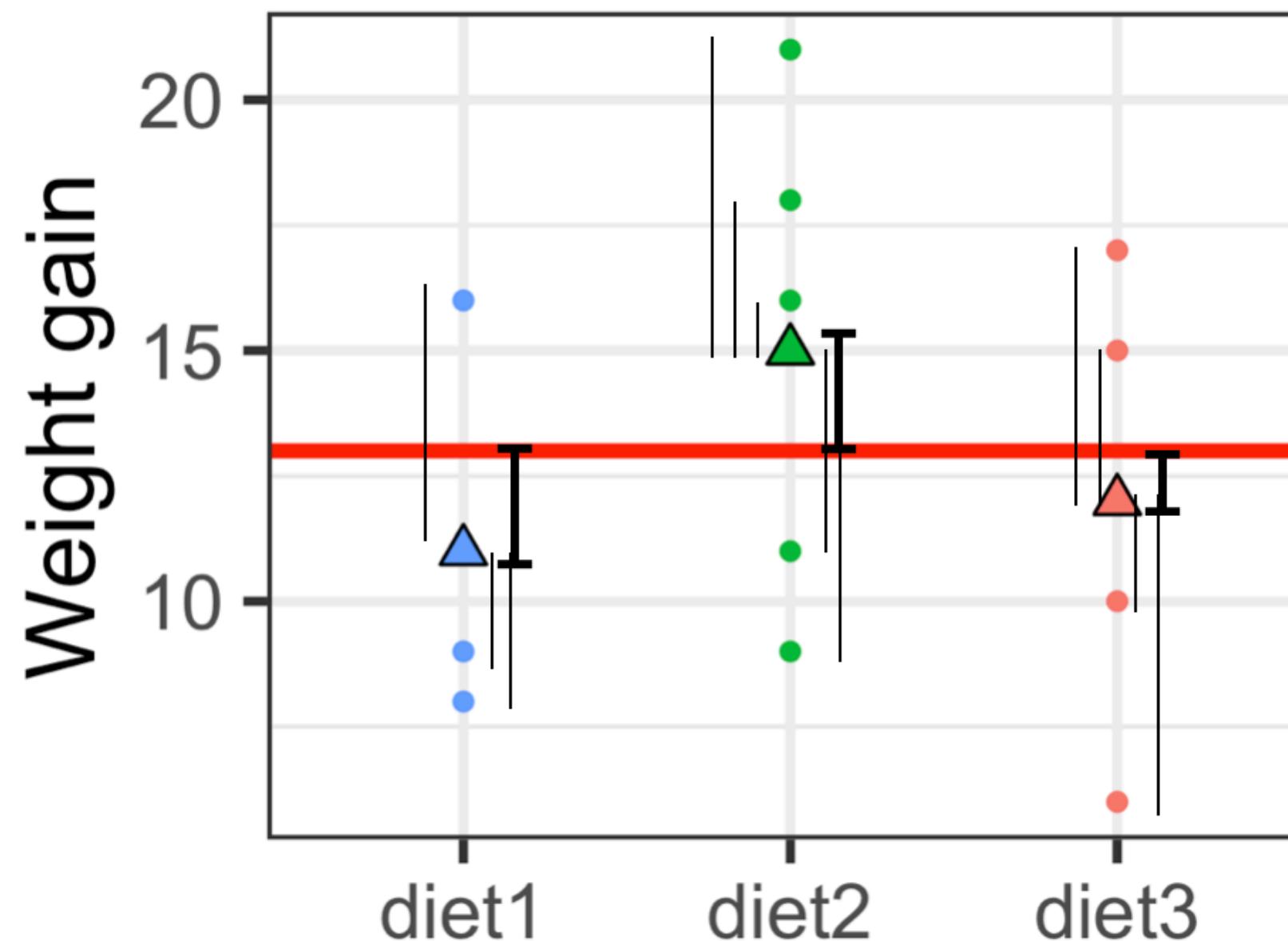
Introduction to the F distribution

Distribution used in several statistical analyses and depends on the numerator degrees of freedom and the denominator degrees of freedom



Introduction to the F distribution and test

Distribution used in several statistical analyses and depends on the numerator degrees of freedom and the denominator degrees of freedom



$$F_s = \frac{MS(\text{between})}{MS(\text{within})}$$

F_s is LARGE if discrepancies among the groups are large relative to variability within the groups

Large $F_s \rightarrow$ small p-value (evidence against null)

Numerator df = df(between)
Denominator df = df(within)

Introduction to the F distribution and test

Source	df	ss	MS
Between	2	36	18
Within	9	210.025	23.336
Total	11	246	

$$F_s = \frac{MS(between)}{MS(within)}$$

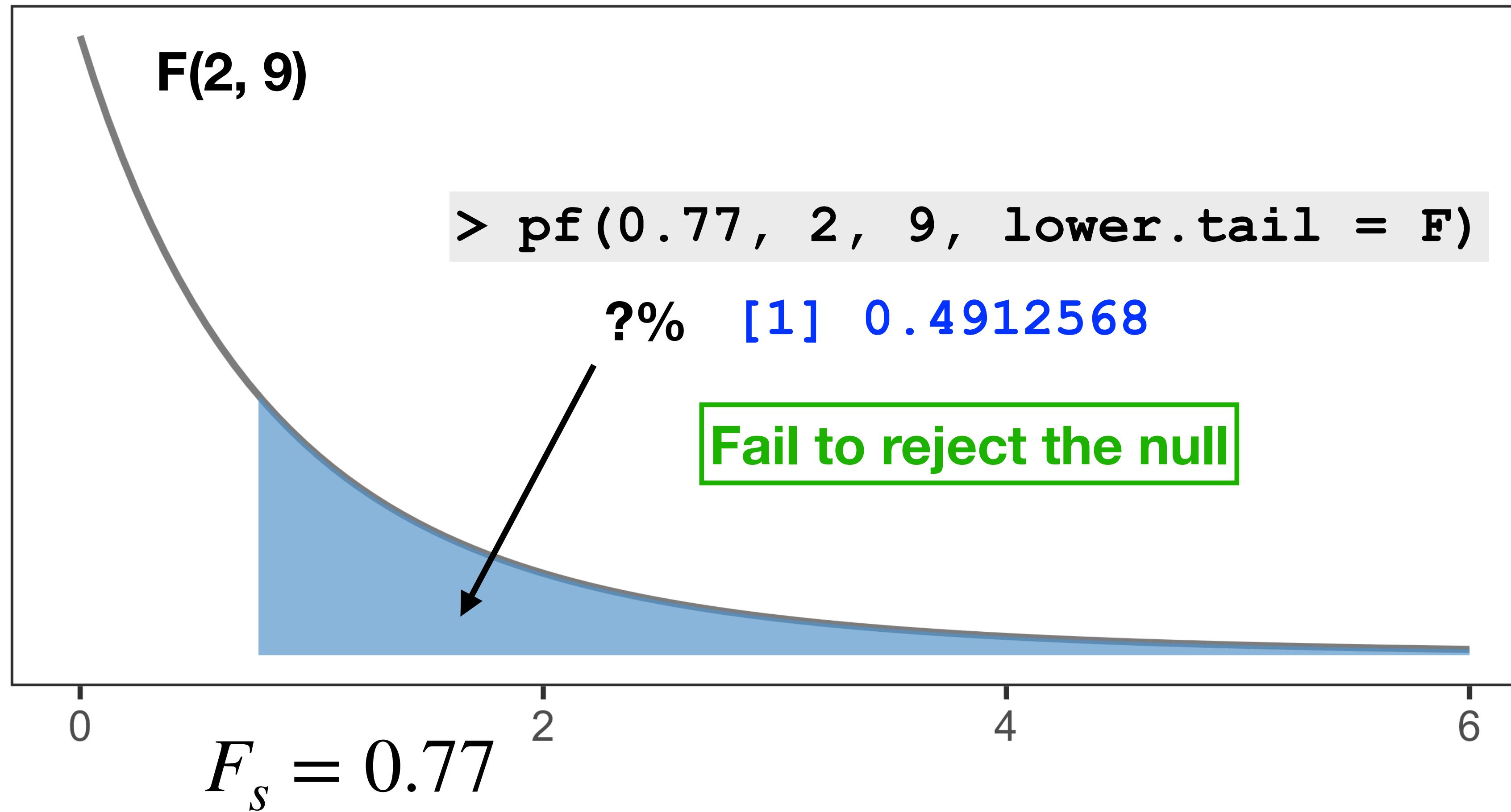
$$F_s = \frac{18}{23.336} = 0.77$$

H_0 : Mean weight gain is the same on all three diets

$$F(2,9)$$

H_A : Mean weight gain is not the same on all three diets

Introduction to the F distribution and test



Example: filling out the ANOVA table

✓) Complete the table then ✓ calculate the F statistic

Source	df	SS	MS
Between	3	135	45
Within	12	337	28.08
Total	15	472	

$$F_s = \frac{MS(between)}{MS(within)}$$

$$F_s = \frac{45}{28.08} = 1.602$$

```
> pf(1.602, 3, 12, lower.tail = F)
```

✓ Bonus: How many **groups** were there in the study?
How many **total observations**?

[1] 0.2406565

Fail to reject the null

df(between) = Groups - 1 → 4 groups

df(total) = Obs. - 1 → 16 obs.

Relationship between F test and regression

```
> summary(lm(happiness ~ income, data = income_data))
```

Call:

```
lm(formula = happiness ~ income, data = income_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.5033	-0.8873	0.0459	0.9218	3.1279

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.34291	0.16505	8.136	3.3e-15 ***
income	0.22765	0.01727	13.182	< 2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.234 on 496 degrees of freedom

Multiple R-squared: 0.2595, Adjusted R-squared: 0.258

F-statistic: 173.8 on 1 and 496 DF, p-value: < 2.2e-16

Relationship between F test and regression

```
> summary(lm(happiness ~ income, data = income_data))
```

Call:

```
lm(formula = happiness ~ income, data = income_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-----	----	--------	----	-----

$H_0 : y = \beta_0$ model fits data best (i.e. all independent variables = 0)

Coefficients:

$H_A : y = \beta_0 + \beta_1 X$ model fits data best

(Intercept)	1.34291	0.16505	8.136	3.3e-15	***
income	0.22765	0.01727	13.182	< 2e-16	***

Here, we reject the null and conclude that there is a relationship between happiness and income.

Residual standard error: 1.234 on 496 degrees of freedom

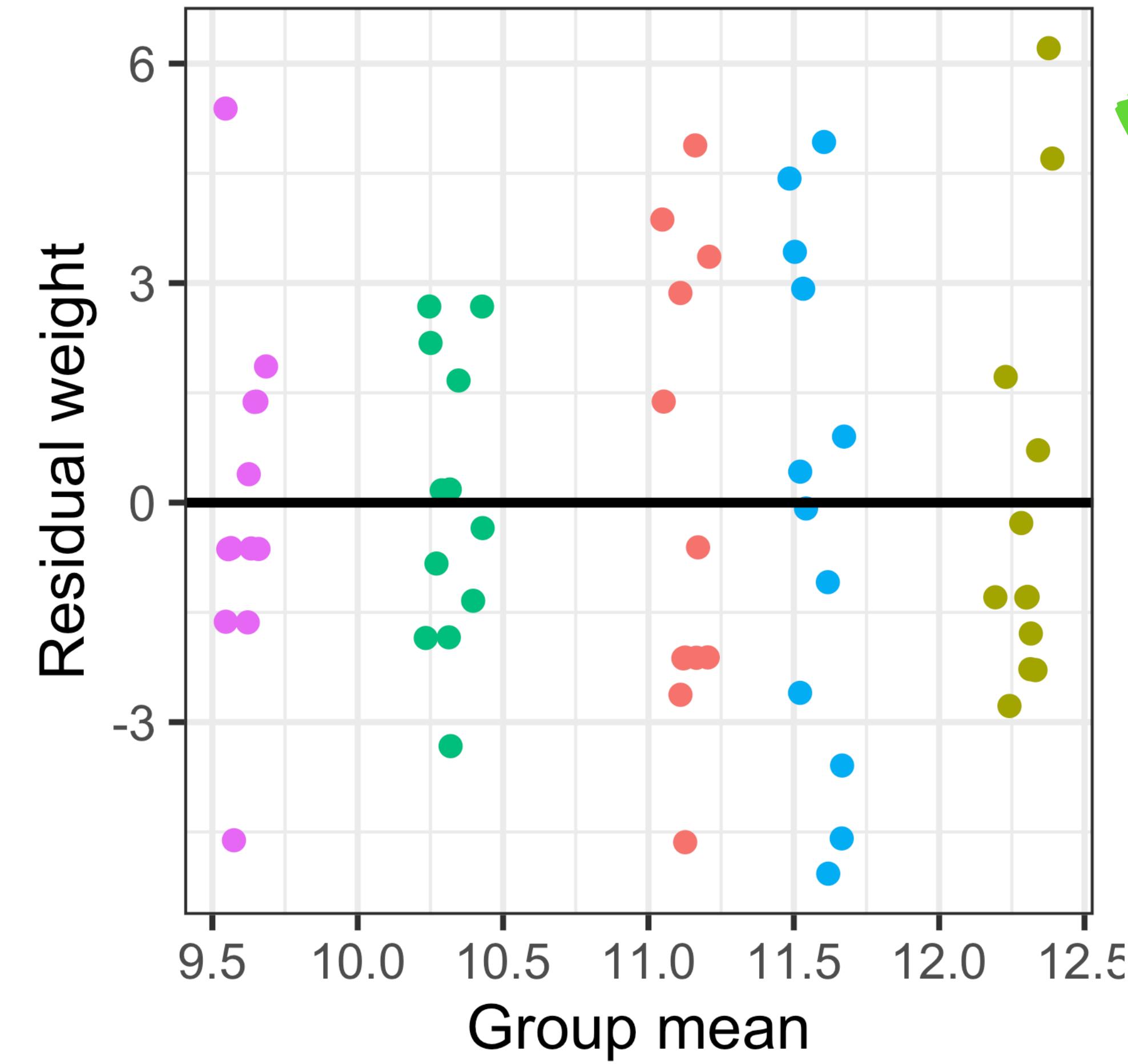
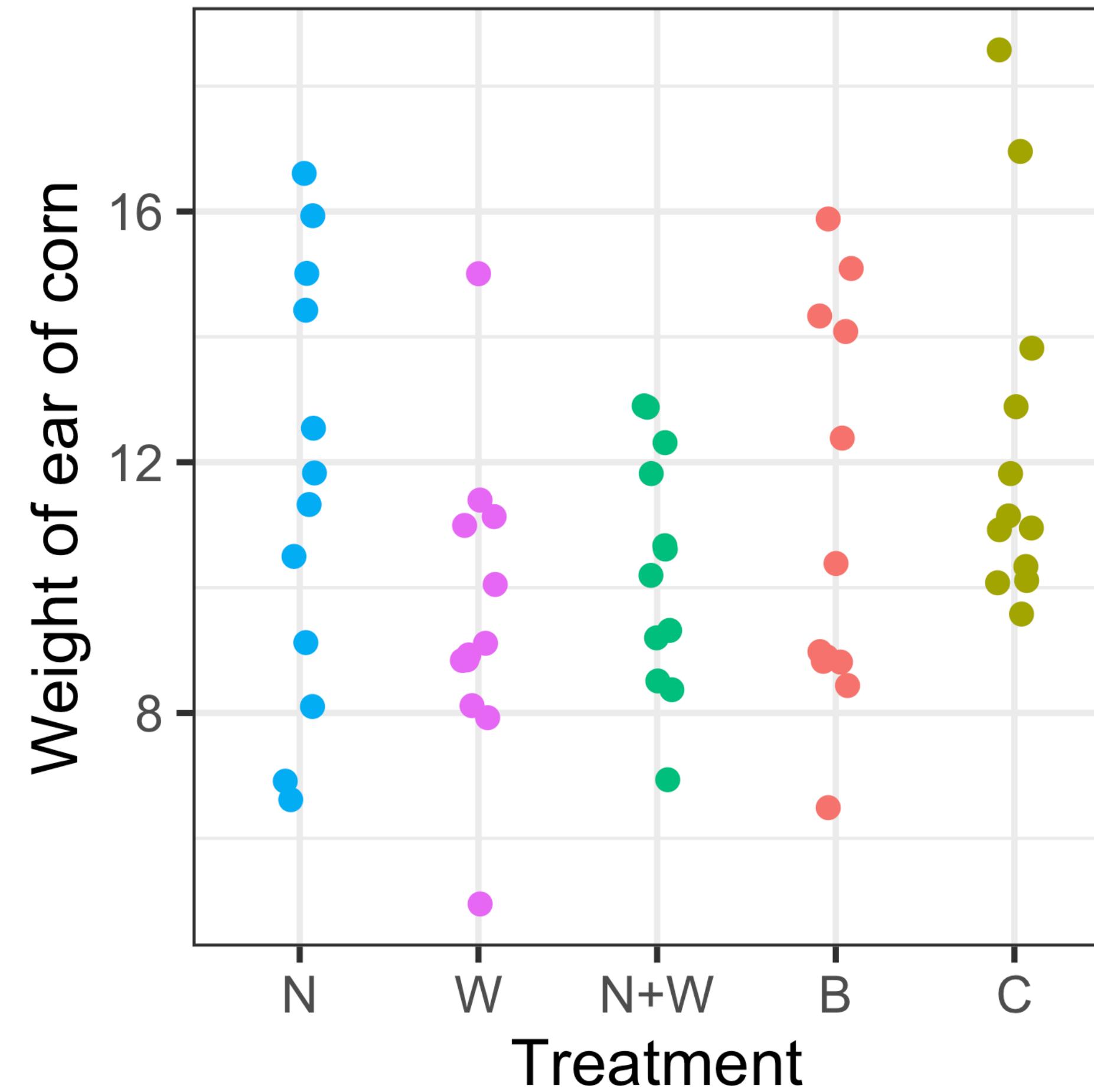
Multiple R-squared: 0.2595, Adjusted R-squared: 0.258

F-statistic: 173.8 on 1 and 496 DF, p-value: < 2.2e-16

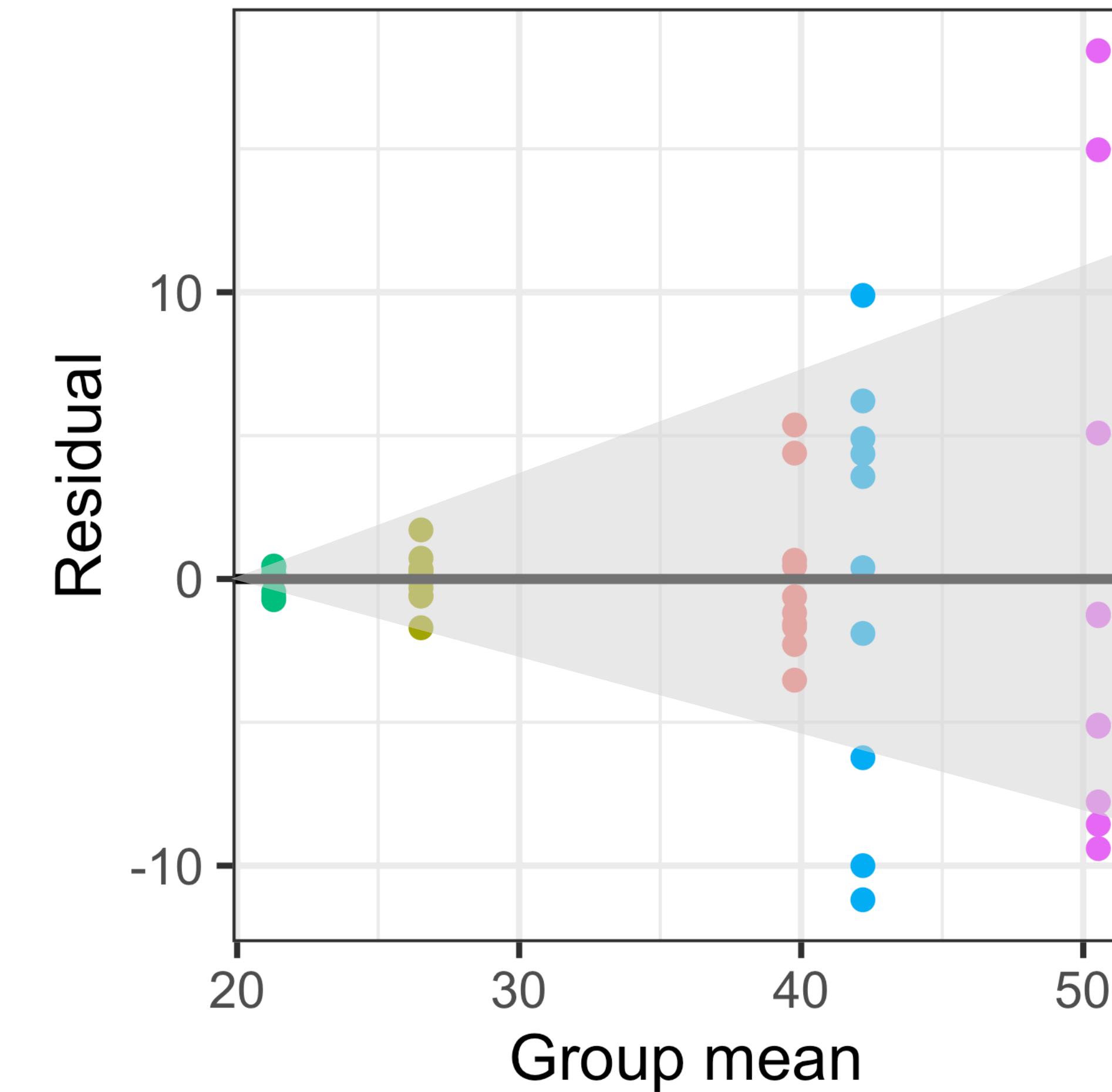
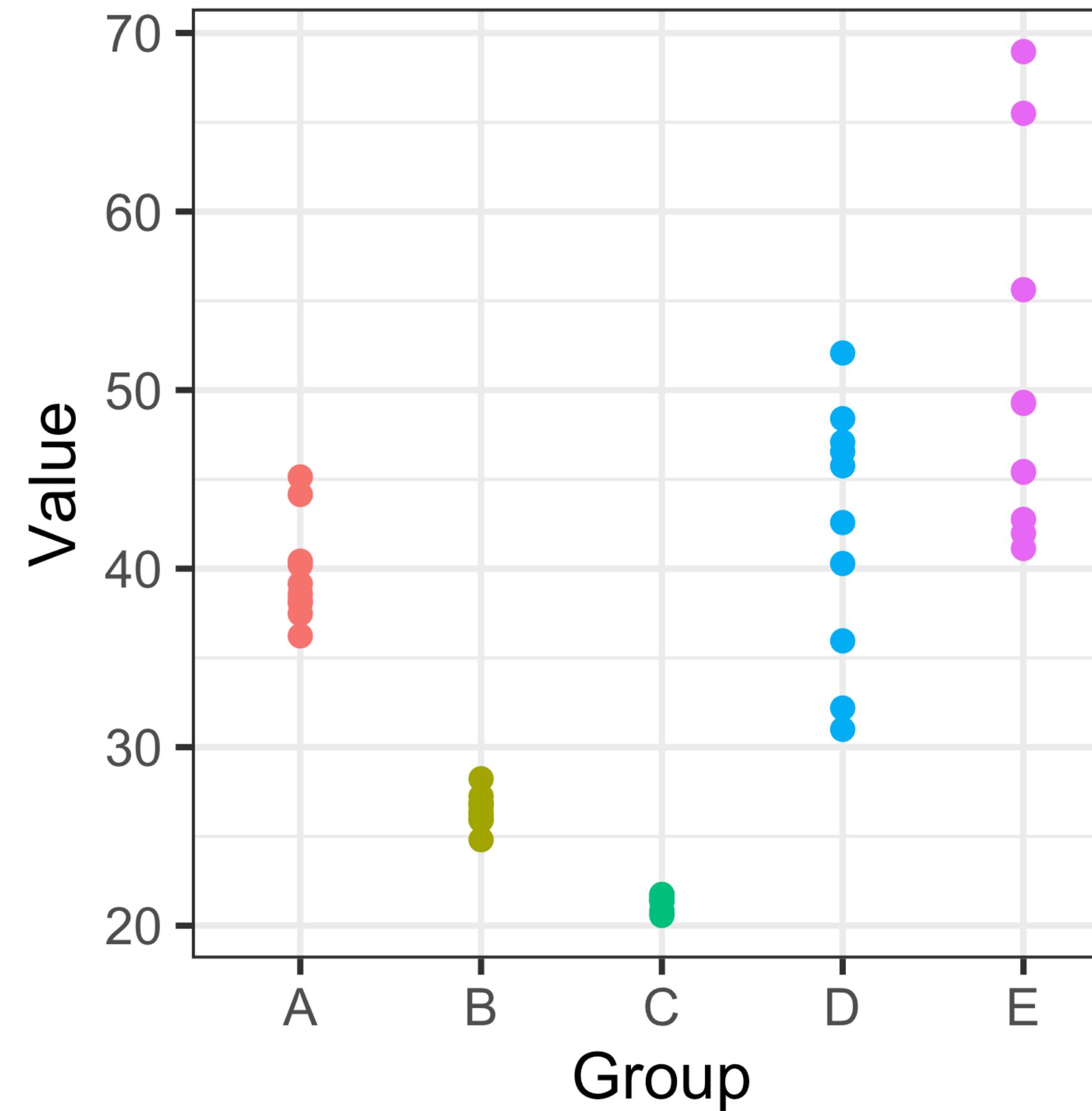
Conditions for using an ANOVA

- **Design conditions**
 - Groups of observations are random samples from their respective populations
 - The groups (I) must be independent of each other
- **Population conditions**
 - The group (I) population distributions must be approximately normal with equal standard deviations
 - *This is less crucial for larger sample size (if $n_1 = n_2 = n_3$)*

Conditions for using an ANOVA



Conditions for using an ANOVA



~Largest SD / smallest SD ≤ 2

Calculating ANOVA using R

diet	weight
diet1	8
diet2	9
diet3	15
diet1	16
diet2	16
diet3	10
diet1	9
diet2	21
diet3	17

Data: weight gain for three different diets

Source	df	SS	MS
Between	2	36	18
Within	9	210.025	23.336
Total	11	246	

```
> anova(lm(weight ~ diet, data = weight))
```

Analysis of Variance Table

Response: weight

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
diet	2	36	18.000	0.7714	0.4907
Residuals	9	210	23.333		

Calculating ANOVA using R

LONG

diet	weight
diet1	8
diet2	9
diet3	15
diet1	16
diet2	16
diet3	10
diet1	9
diet2	21
diet3	17

Data: weight gain for three different diets

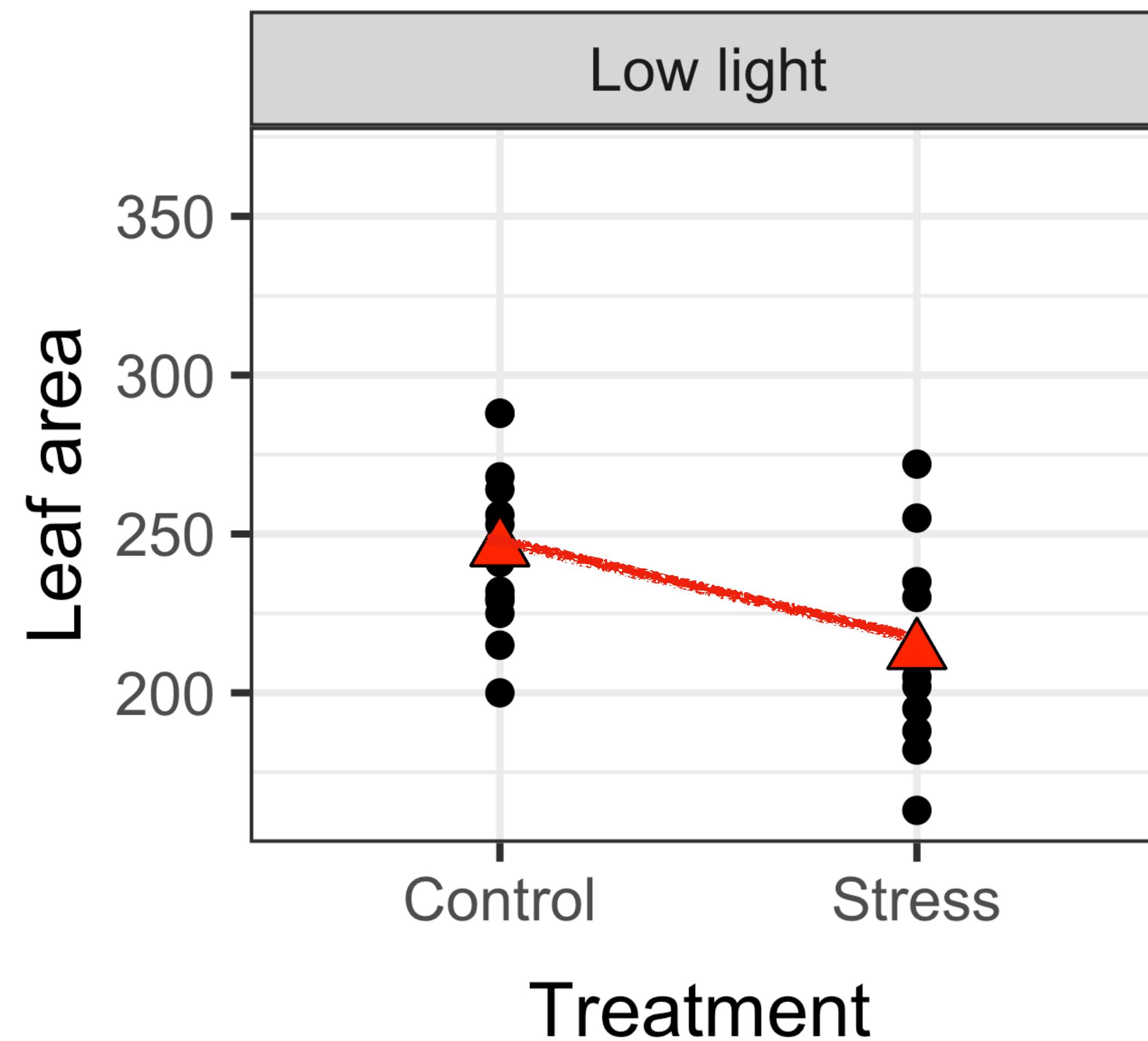
WIDE

diet1	diet2	diet3
8	9	15
16	16	10
9	11	17
NA	NA	NA

```
tidyverse::pivot_longer(diet1:diet3,  
                        names_to = "diet",  
                        values_to = "weight")
```

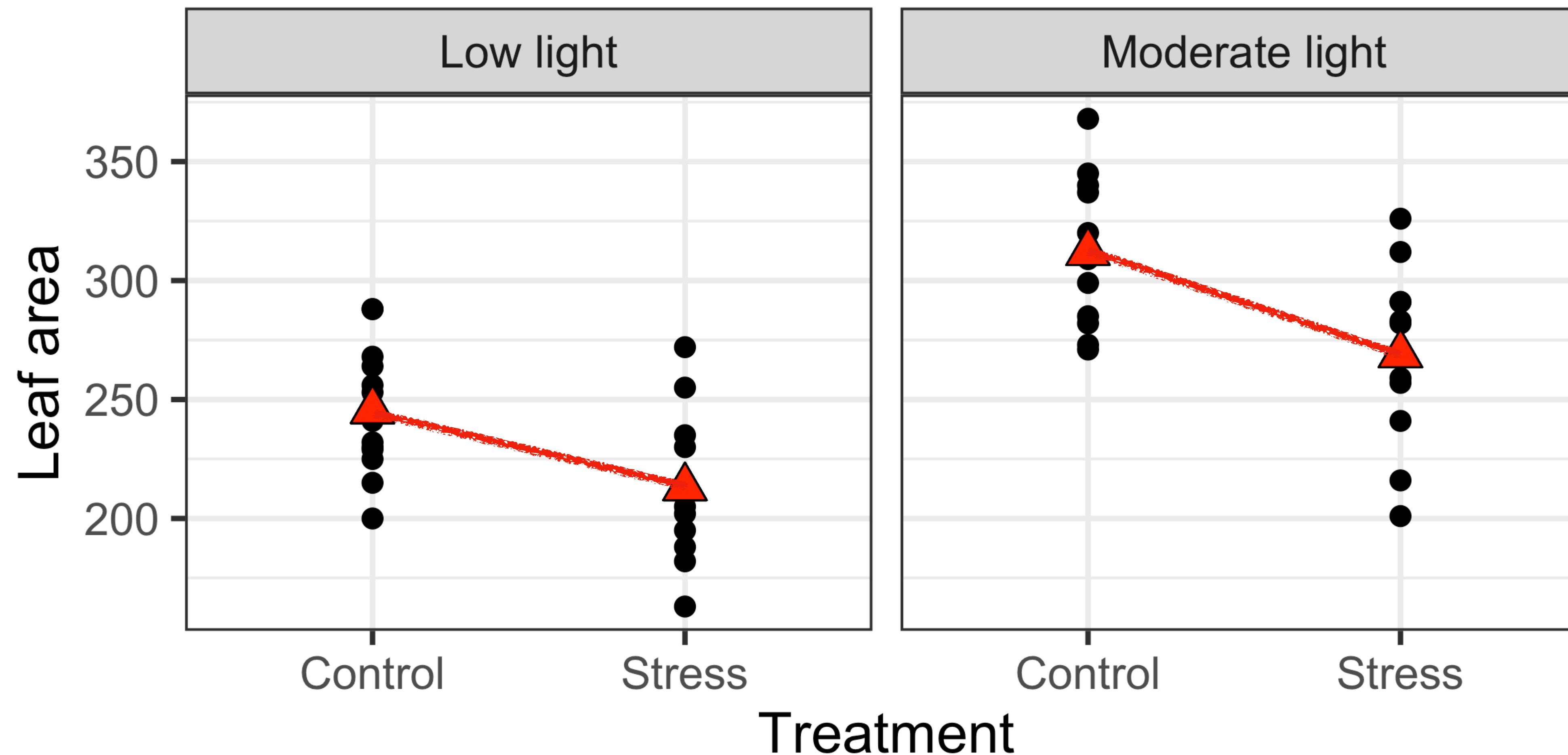
```
> anova(lm(weight ~ diet, data = weight))
```

Two-way (Factorial) ANOVA

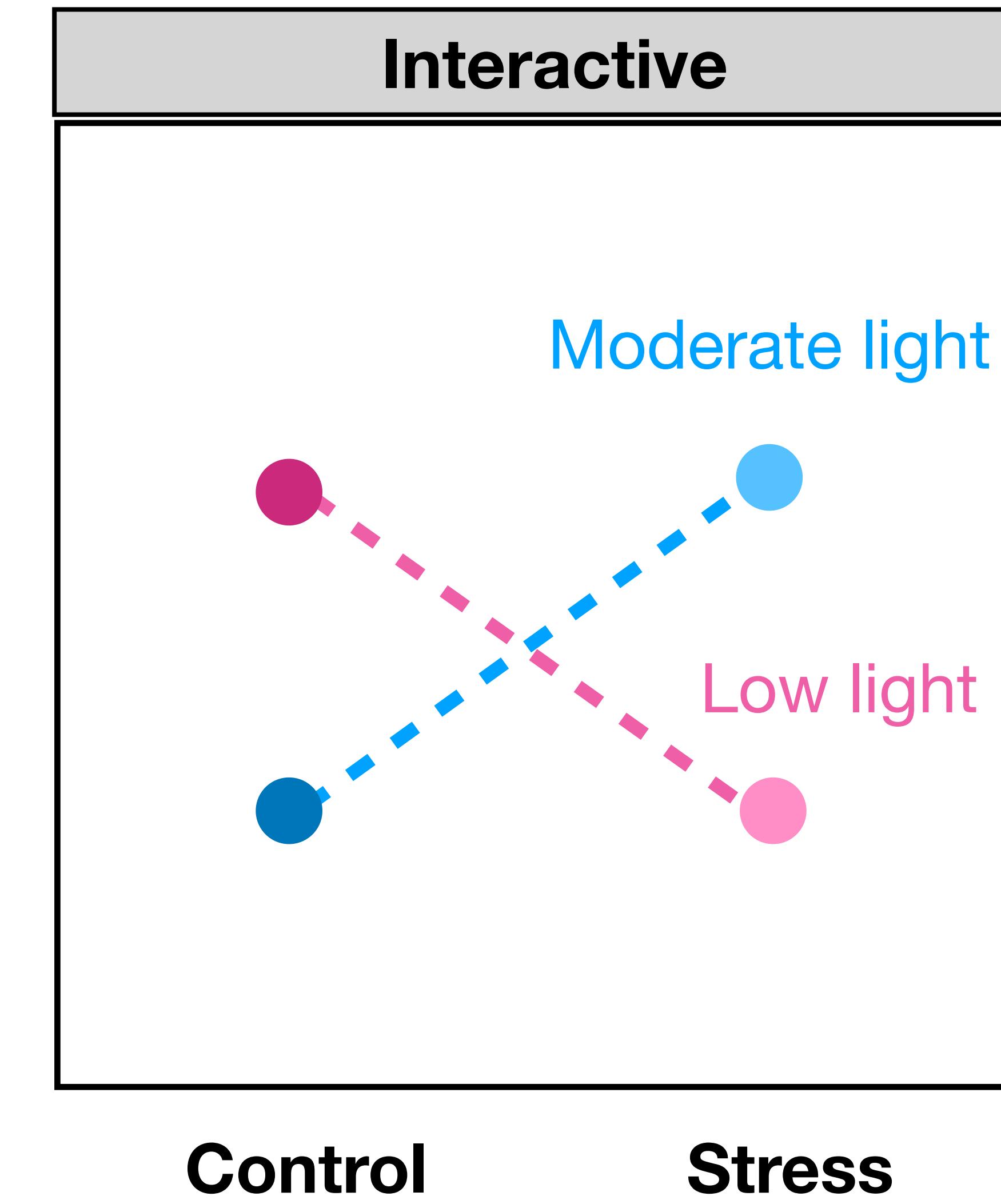
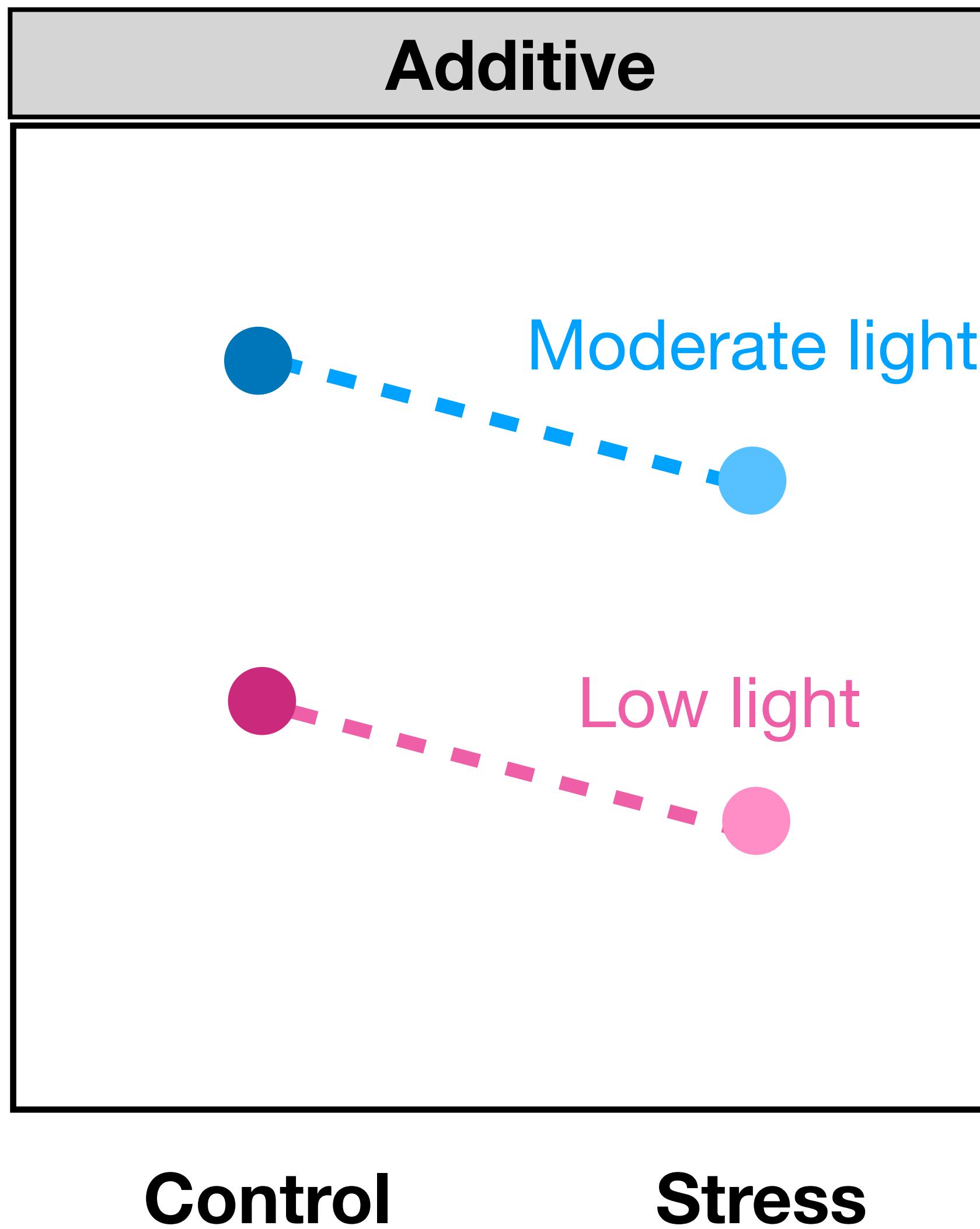


Two-way (Factorial) ANOVA

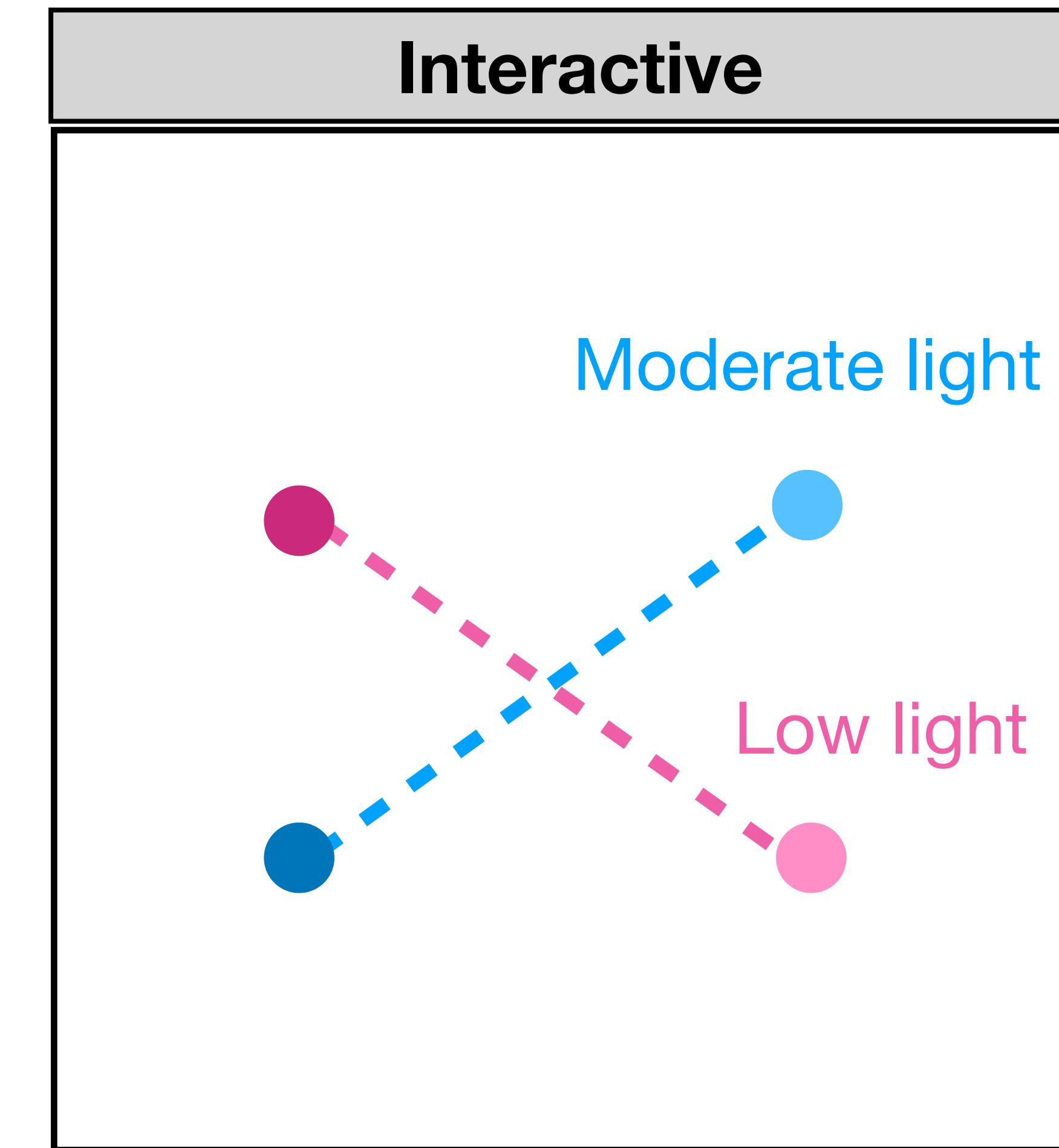
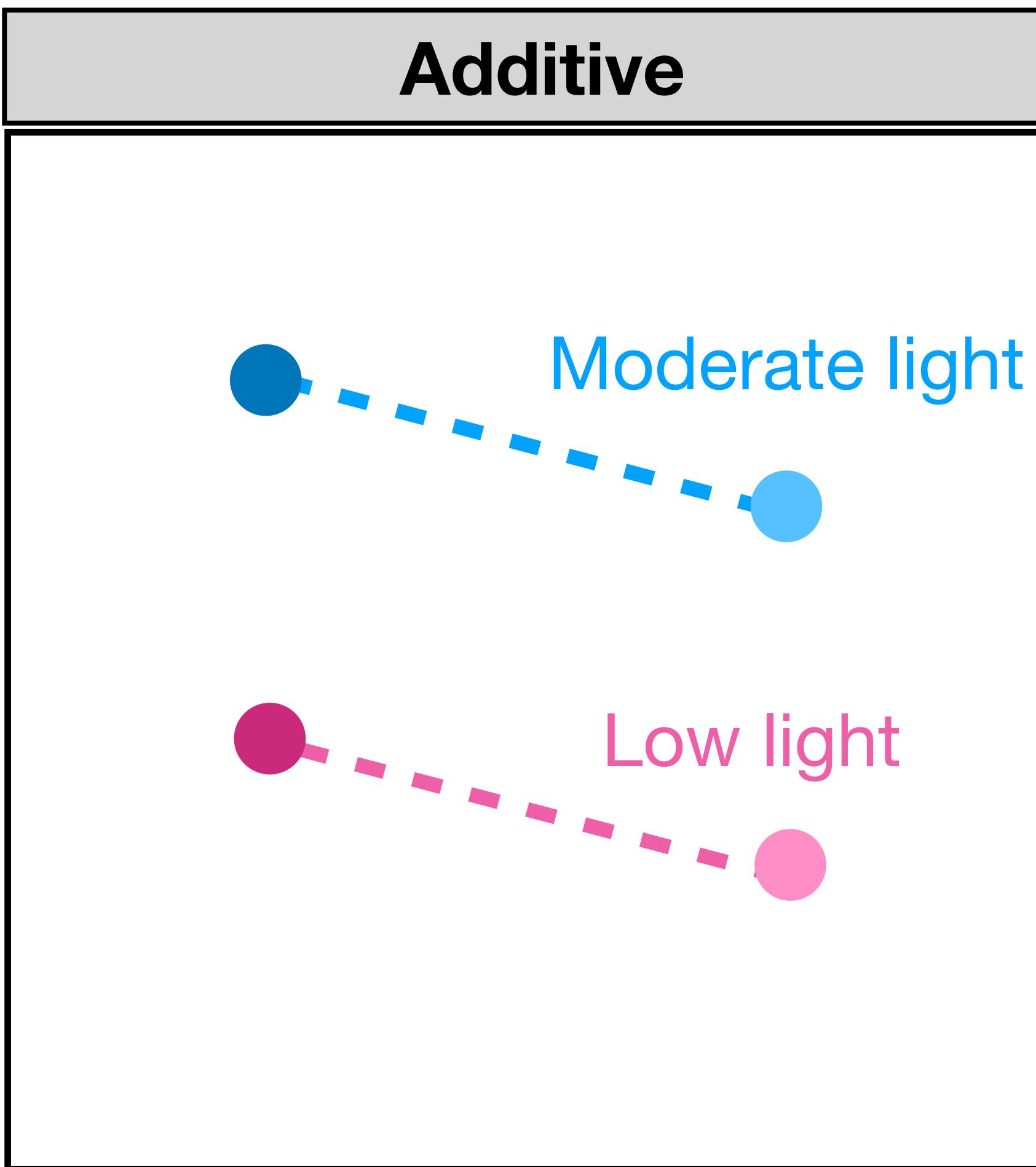
$$y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk}$$



Additive or interactive factors?



Additive or interactive factors?



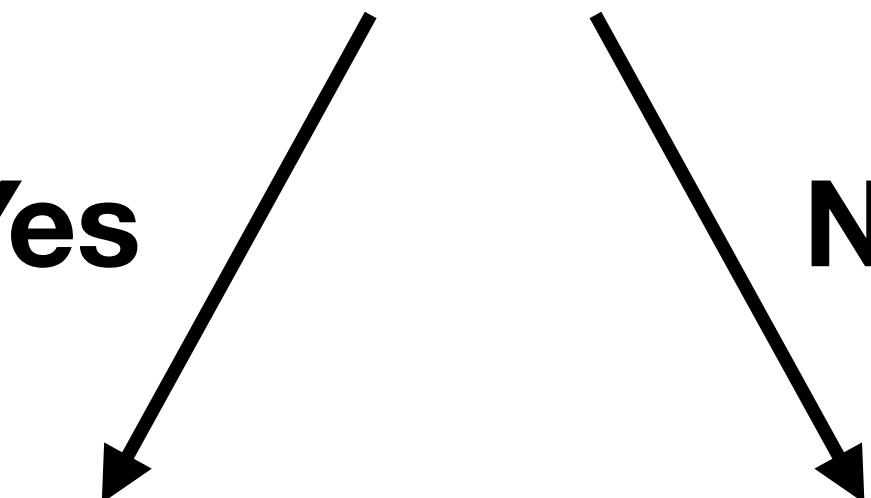
$$y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk}$$

$$y_{ijk} = \mu + \tau_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

Two-way (Factorial) ANOVA

Check for interactions

Yes



No

Done! Reject H_0 .

Test each factor for additive effects

```
> anova(lm(area ~ light * stress))
```

Response: leaf_area

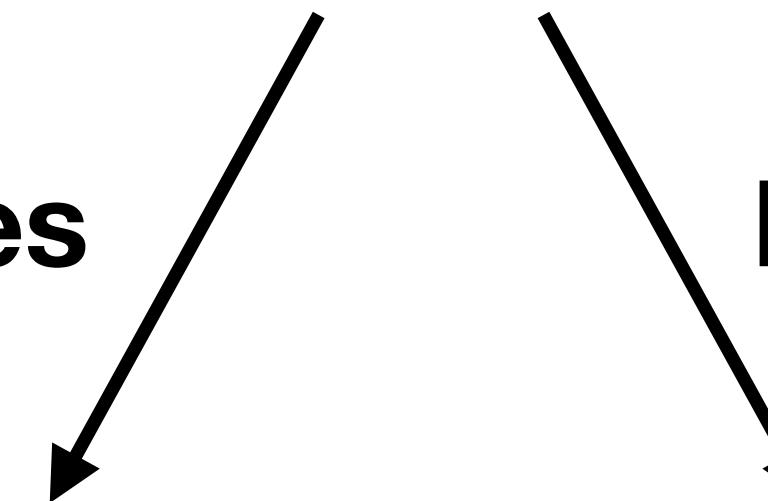
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
stress	1	18432	18432	19.6640	5.364e-05	***
light	1	48678	48678	51.9333	3.550e-09	***
stress:light	1	361	361	0.3851	0.5378	
Residuals	48	44992	937			

Interaction X

Two-way (Factorial) ANOVA

Check for interactions

Yes



No

Done! Reject H_0 .

Test each factor for additive effects

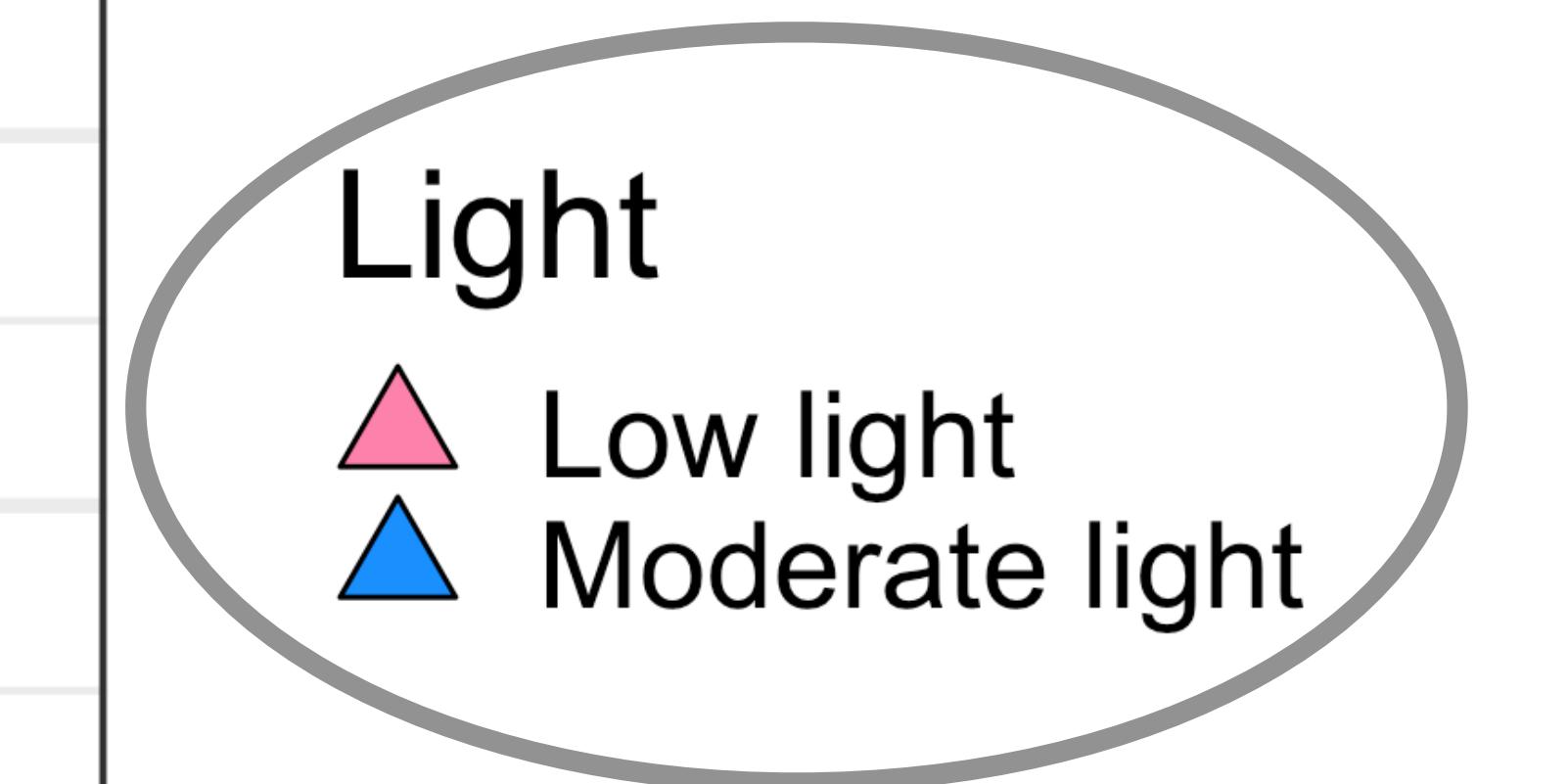
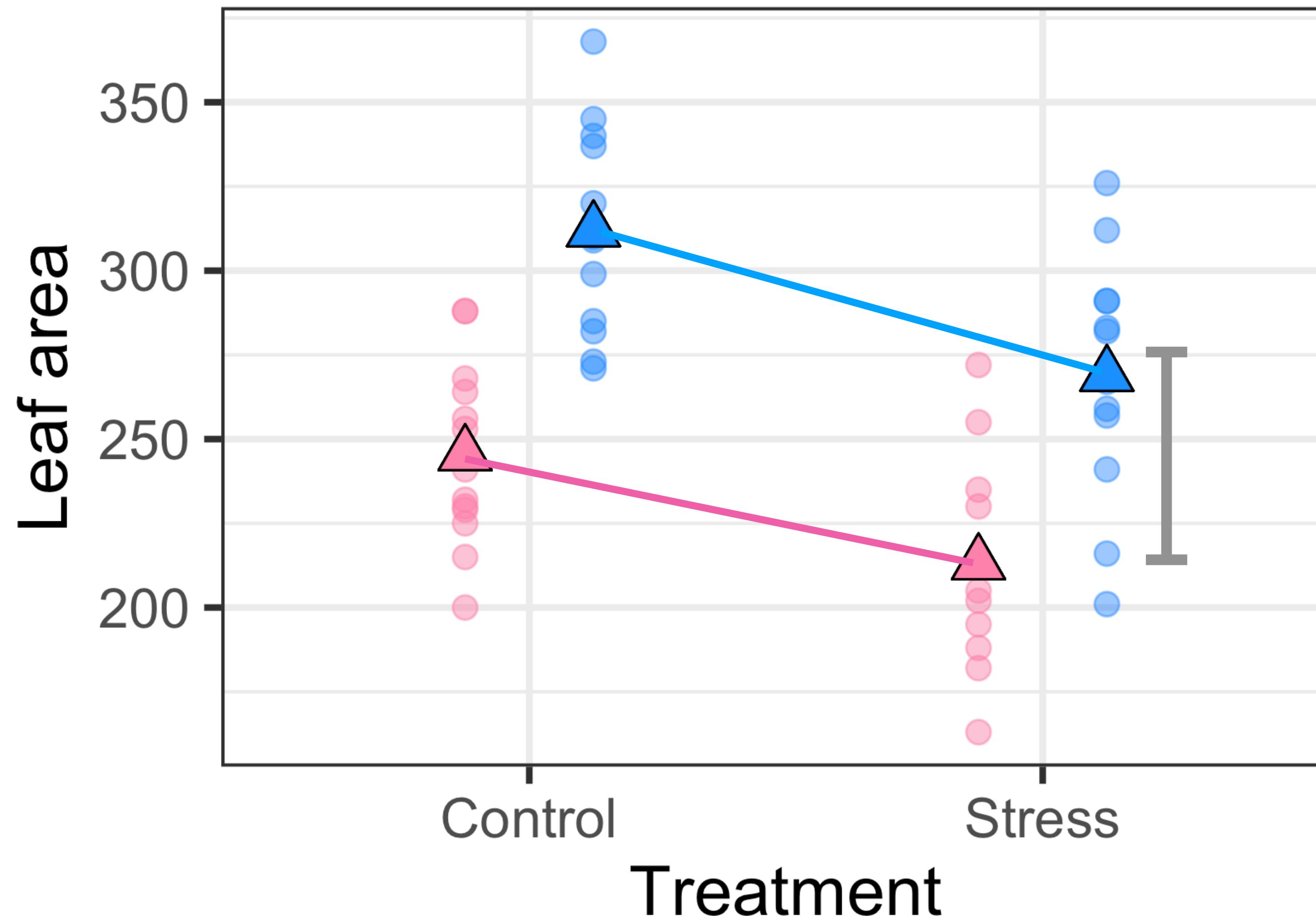
```
> anova(lm(area ~ light * stress))
```

Response: leaf_area

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
stress	1	18432	18432	19.6640	5.364e-05	***
light	1	48678	48678	51.9333	3.550e-09	***
stress:light	1	361	361	0.3851	0.5378	
Residuals	48	44992	937			



Two-way (Factorial) ANOVA

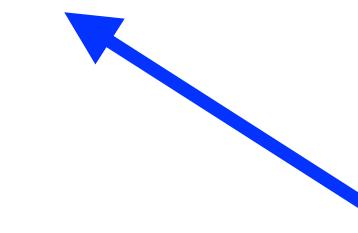


What to do next? You know that the group means are not equal...

Next steps for significant ANOVAs

- Pairwise t-tests to see which groups are significantly different
 - ***Need to make sure you adjust for multiple testing!***
- Many options: one of my favorites is **TukeyHSD(model)**

```
> TukeyHSD(aov(leaf_area ~ light + stress, data = leaf_area))
```



```
anova(lm(. . .))
```

Next steps for significant ANOVAs

```
> TukeyHSD(aov(leaf_area ~ light + stress, data = leaf_area))
```

Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = leaf_area ~ light + stress, data = leaf_area)

\$light

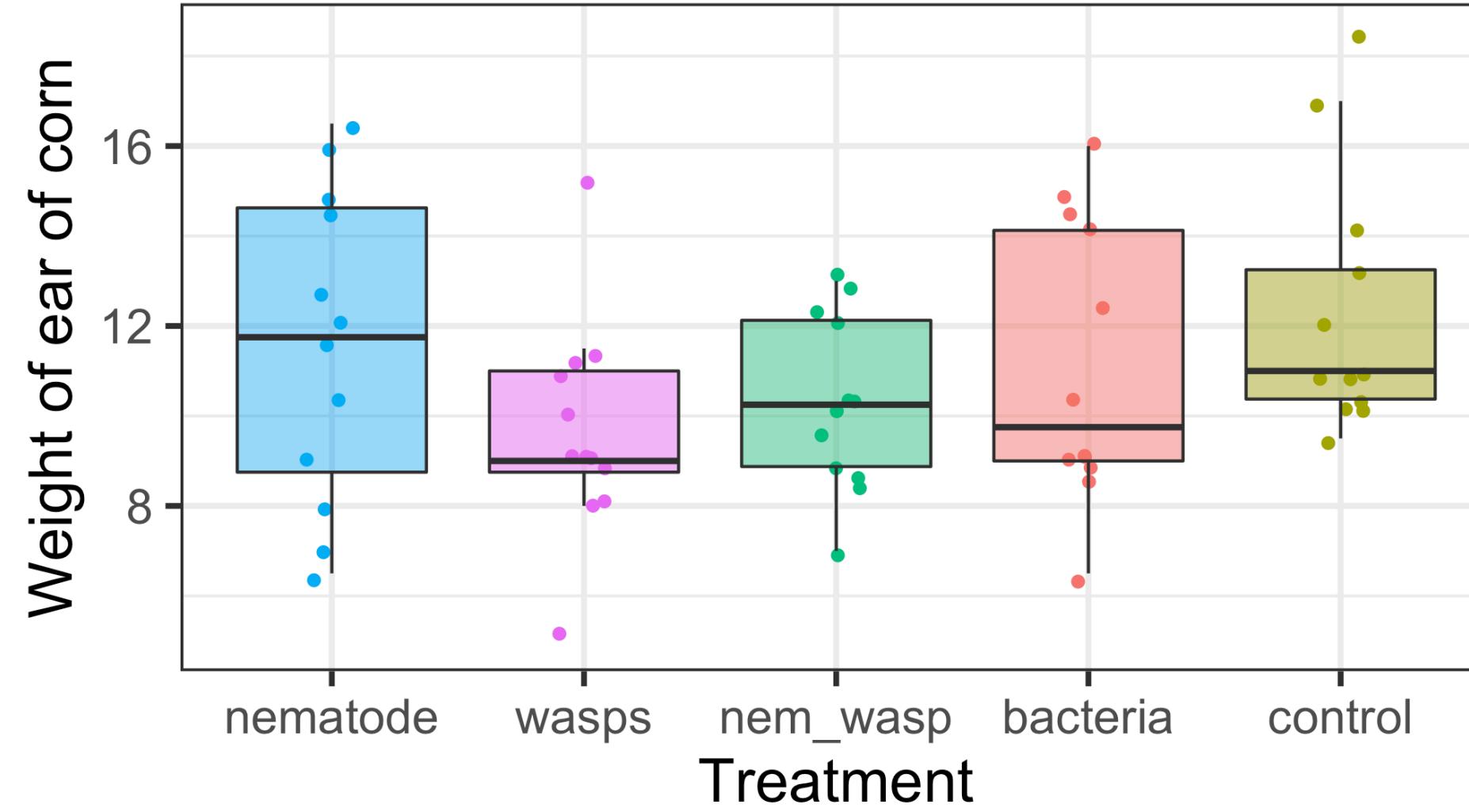
	diff	lwr	upr	p	adj
Moderate light-Low light	61.19231	44.23582	78.14879	0	

\$stress

	diff	lwr	upr	p	adj
Stress-Control	-37.65385	-54.61033	-20.69736	4.75e-05	

Next steps for significant ANOVAs

```
> TukeyHSD(aov(weight ~ treatment, data = corn))
```



Shows pairwise t-test results

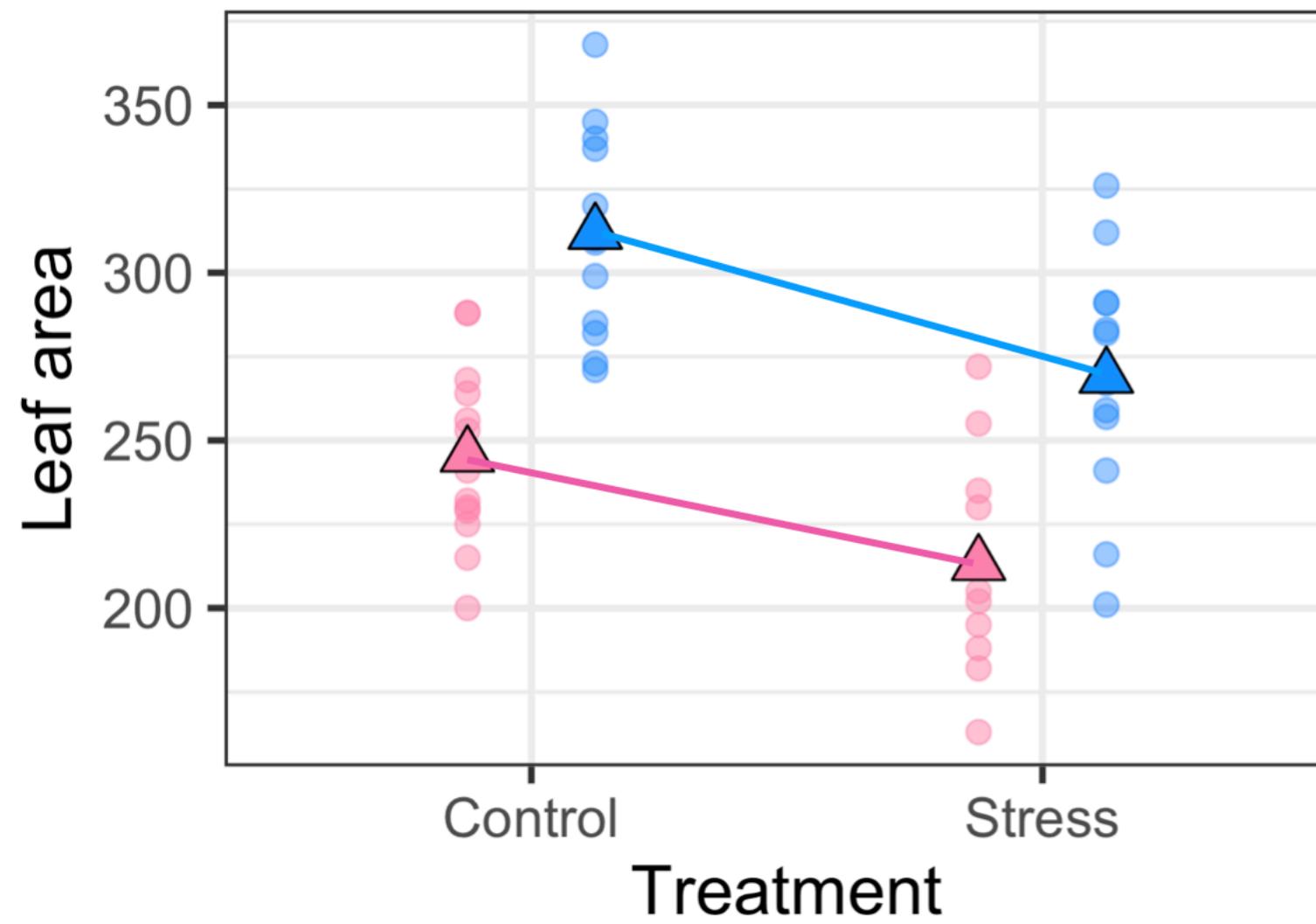
Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = weight ~ treatment, data = long_corn)

\$treatment

	diff	lwr	upr	p adj
control-bacteria	1.1666667	-2.078614	4.4119471	0.8478963
nem_wasp-bacteria	-0.7916667	-4.036947	2.4536138	0.9582644
nematode-bacteria	0.4583333	-2.786947	3.7036138	0.9945376
wasps-bacteria	-1.5000000	-4.745280	1.7452805	0.6901281
nem_wasp-control	-1.9583333	-5.203614	1.2869471	0.4412802
nematode-control	-0.7083333	-3.953614	2.5369471	0.9720121
wasps-control	-2.6666667	-5.911947	0.5786138	0.1548970
nematode-nem_wasp	1.2500000	-1.995280	4.4952805	0.8127831
wasps-nem_wasp	-0.7083333	-3.953614	2.5369471	0.9720121
wasps-nematode	-1.9583333	-5.203614	1.2869471	0.4412802

Aside: using ANOVA to compare models



```
> anova(model1, model3)
```

Model 1: leaf_area ~ light

Model 2: leaf_area ~ stress + light

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	50	63784				
2	49	45353	1	18432	19.914	4.748e-05 ***

Stress AND light > only light

Which variables (light, stress, or both) best predict leaf area?

```
# light as a predictor
```

```
> model1 <- lm(leaf_area ~ light)
```

```
# stress as a predictor
```

```
> model2 <- lm(leaf_area ~ stress)
```

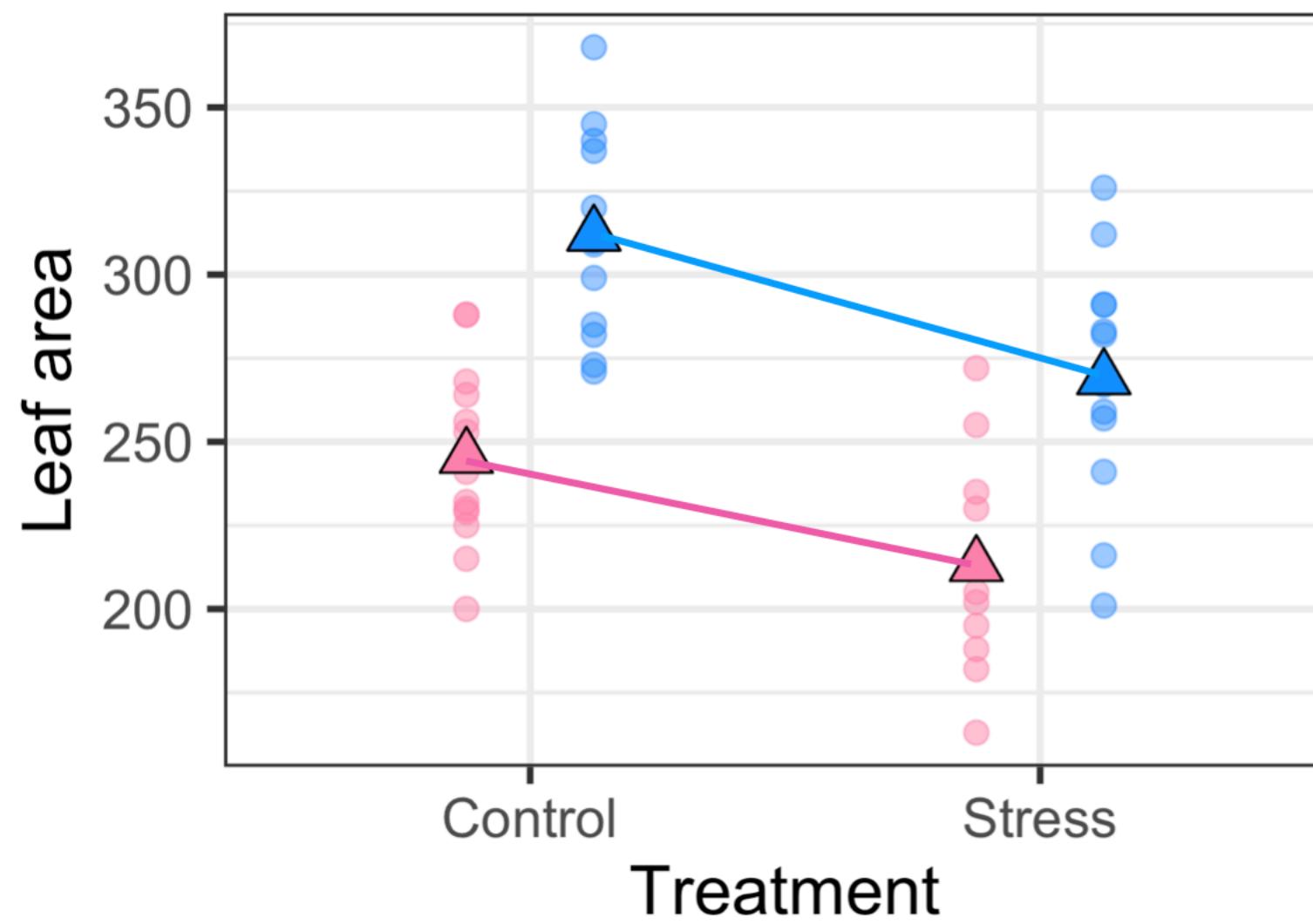
```
# light and stress as predictors (additive)
```

```
> model3 <- lm(leaf_area ~ light + stress)
```

```
# light and stress as predictors (interacting)
```

```
> model4 <- lm(leaf_area ~ light * stress)
```

Aside: using ANOVA to compare models



```
> anova(model3, model4)
```

Model 1: leaf_area ~ stress + light

Model 2: leaf_area ~ stress * light

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	49	45353				
2	48	44992	1	360.94	0.3851	0.5378

Which variables (light, stress, or both) best predict leaf area?

light as a predictor

```
> model1 <- lm(leaf_area ~ light)
```

stress as a predictor

```
> model2 <- lm(leaf_area ~ stress)
```

light and stress as predictors (additive)

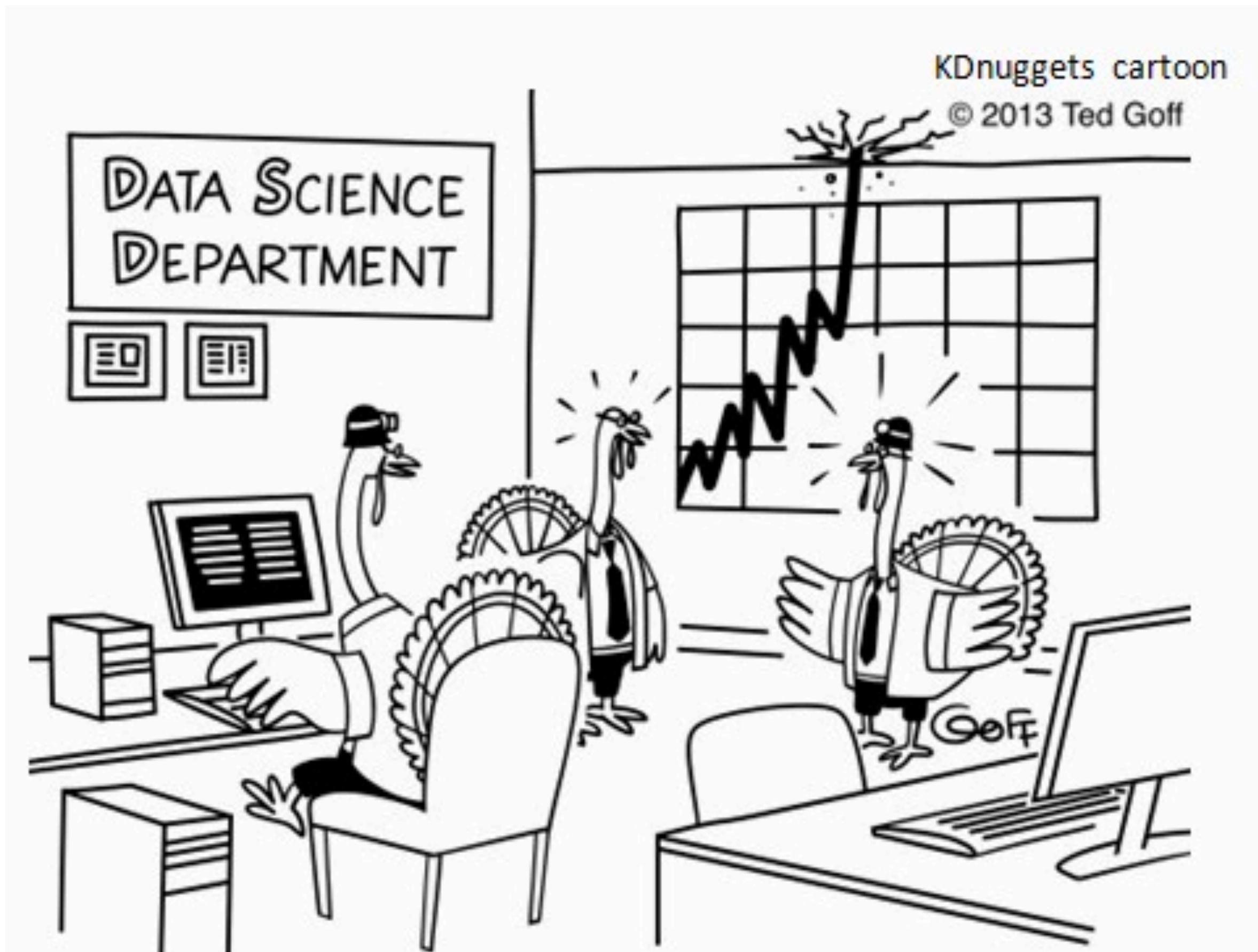
```
> model3 <- lm(leaf_area ~ light + stress)
```

light and stress as predictors (interacting)

```
> model4 <- lm(leaf_area ~ light * stress)
```

Interaction is not better than additive

Announcements



**"I don't like the look of this.
Searches for gravy and turkey stuffing
are going through the roof!"**

No class Thursday (Happy Thanksgiving!)

No homework this week

Next Tuesday: Practicum 3
(bring your laptops!)

Remember to look at the data for your project! (And if you want to change projects that's okay! No need for a new proposal)