Two-way ANOVA

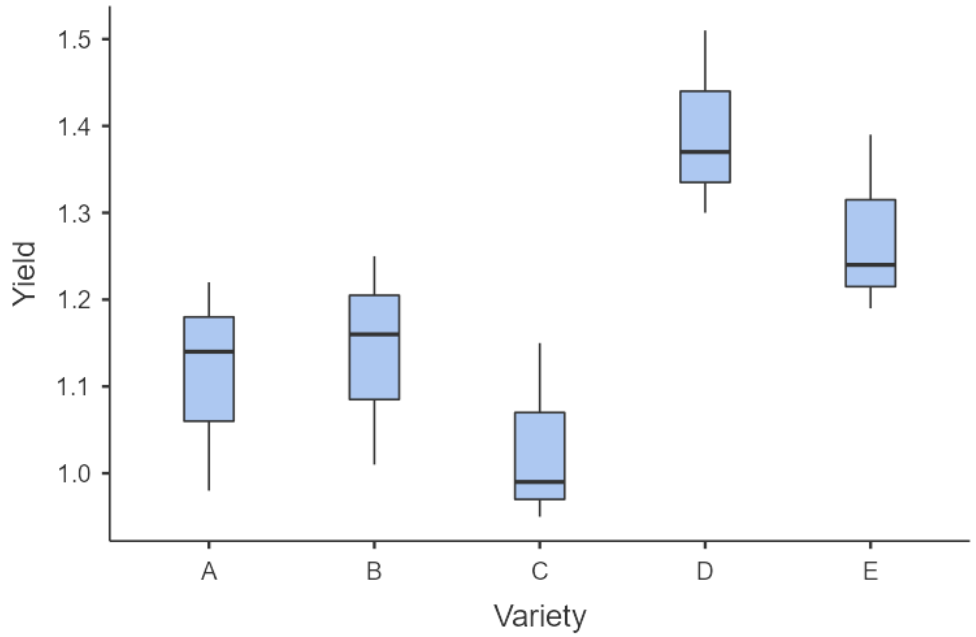
Questions or announcements?

One-way ANOVA - difference in 3+ groups

- One-way refers to a single FACTOR with many levels
- `Yield ~ Seed`$_{\texttt{variety}}$

Review

- Null and alternative hypotheses?
- H_O: All μ's are the same
- H_A: NOT all μ's are the same

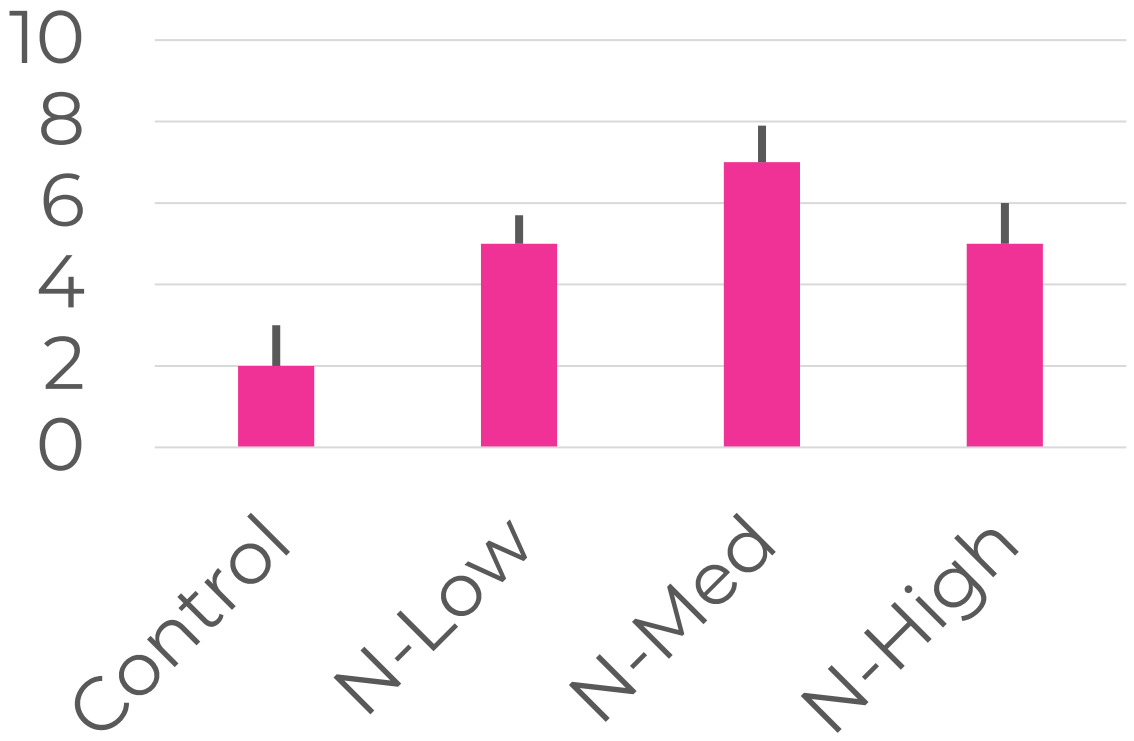


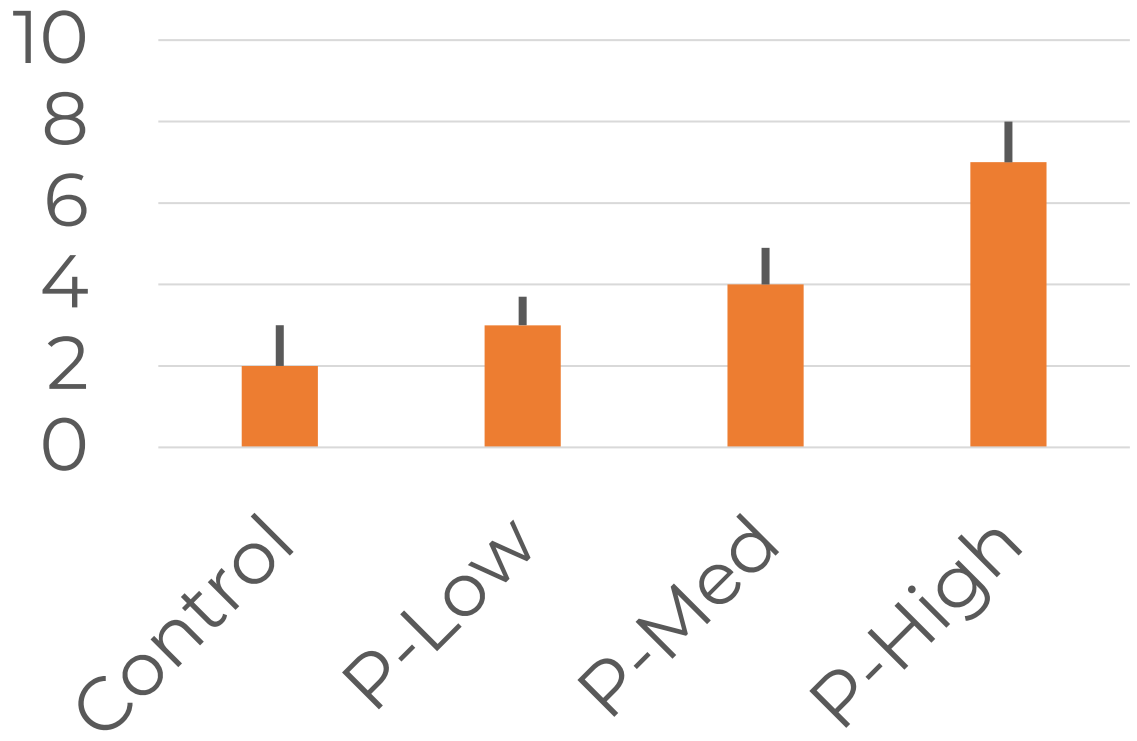
ANOVA terminology review

- Dependent Variable
  - response variable, **continuous**
- Independent Variable(s)
  - predictor variable(s), **categorical**, (`factor( , levels = c( )`) in R
- Factors
  - may be > 1 Independent variable
  - Factor = Nitrogen fertilizer
- Treatment Levels within Factors
  - Nitrogen levels: control (no N), low N, high H
- Replicates
  - Usually the number of observations, *n*

How to test more than one factor?

- Could do many one-way ANOVA studies
- How does N addition affect corn production?
- How does P addition affect corn production?
- Test for N + P?





Two-way ANOVA: Main + Interaction effects

- Account for multiple independent variables (Factors)
  - Factorial analysis
- Example: Effects on N and P on Corn production

	Control	+N
Control	10	15
+P	16	25

Interaction Effect of N & P

2 x 2 **Fully Factorial Design**

2 levels of N & 2 levels of P

Hypotheses in 2 x 2 factorial design

1. Main effect of factor A
   - H_O: There is no effect of factor A
   - H_A: There is an effect of factor A
2. Main effect of factor B
   - H_O : There is no effect of factor B
   - H_A : There is an effect of factor B
3. Interaction
   - H_O : The effect of one factor does **not depend** on another factor
   - H_A : The effect of one factor **does depend** on another

Main effects

- Impact of effect of one factor **averaged *across*** the other factor

	Control	+N	Avg.
Control	10	15	12.5
+P	16	25	20.5
Avg.	13	20	Global Avg. 16.5

Main effects - Nitrogen

- Impact of effect of one factor averaged *across* the other factor
- Effect of N, *ignoring* P

	Control	+N	Avg.
Control	10	15	12.5
+P	16	25	20.5
Avg.	13	20	Global Avg. 16.5

Main effects - Phosphorous

- Impact of effect of one factor averaged *across* the other factor
- Effect of P, *ignoring* N

	Control	+N	Avg.
Control	10	15	12.5
+P	16	25	20.5
Avg.	13	20	Global Avg. 16.5

Group and global averages → Sum of Squares

- Marginal Means
- Tedious hand calculations (see video in topic page)
- R does it automatically (Important to set up data correctly!)

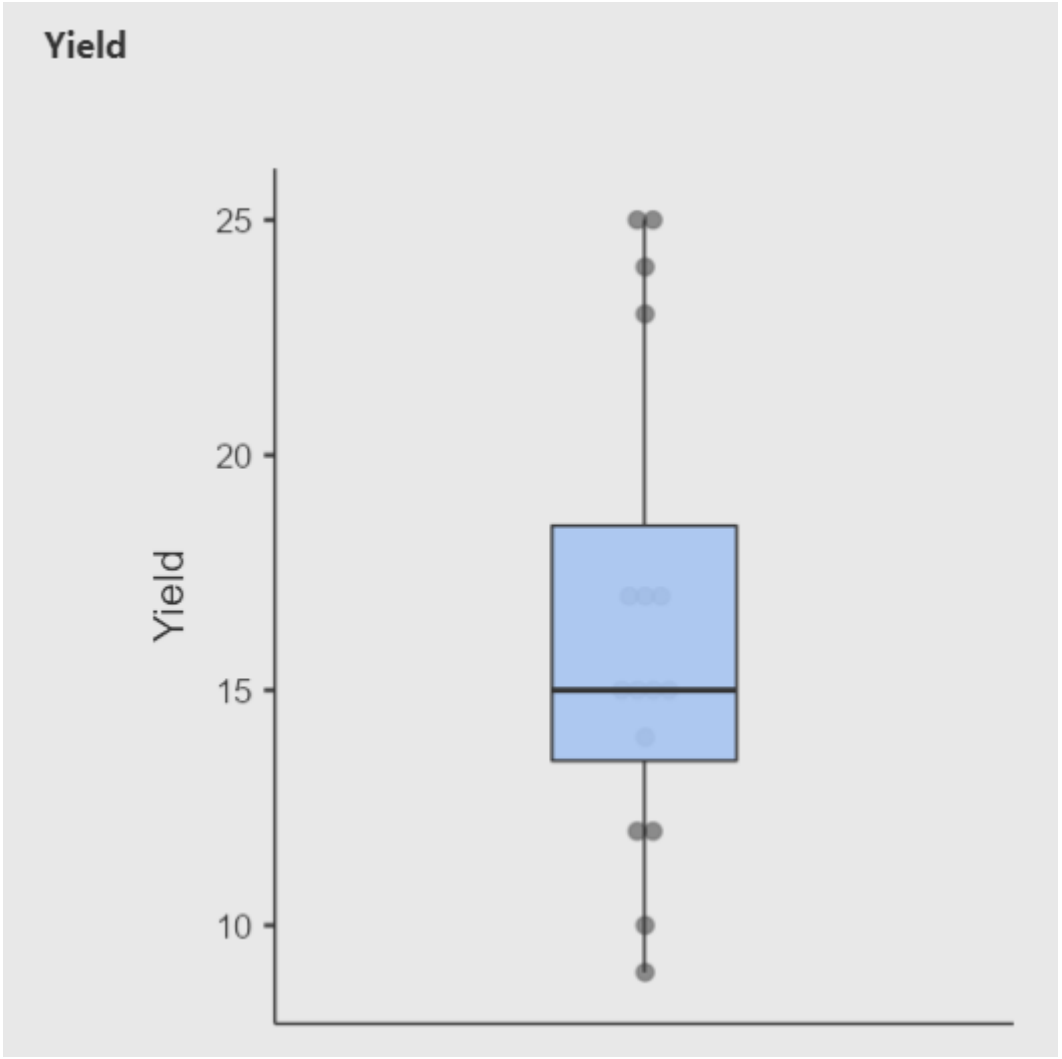
	Control	+N	Avg.
Control	10	15	12.5
+P	16	25	20.5
Avg.	13	20	Global Avg. 16.5

Global Average

- Response = Yield
- No groups
- Total variation observed

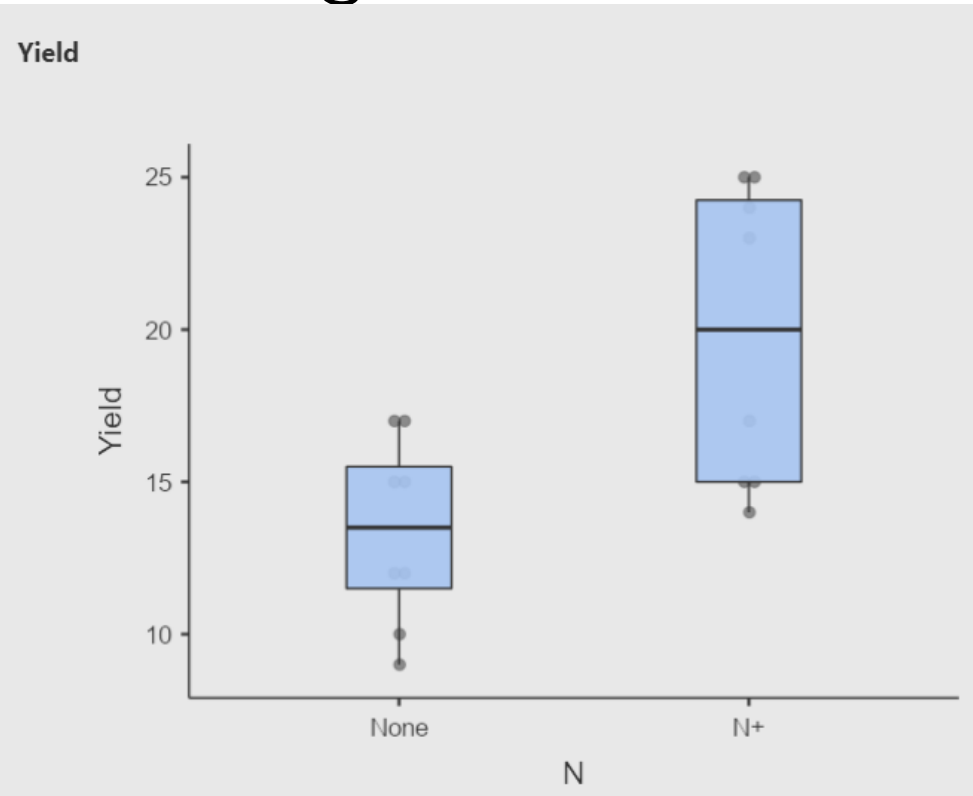
Descriptives

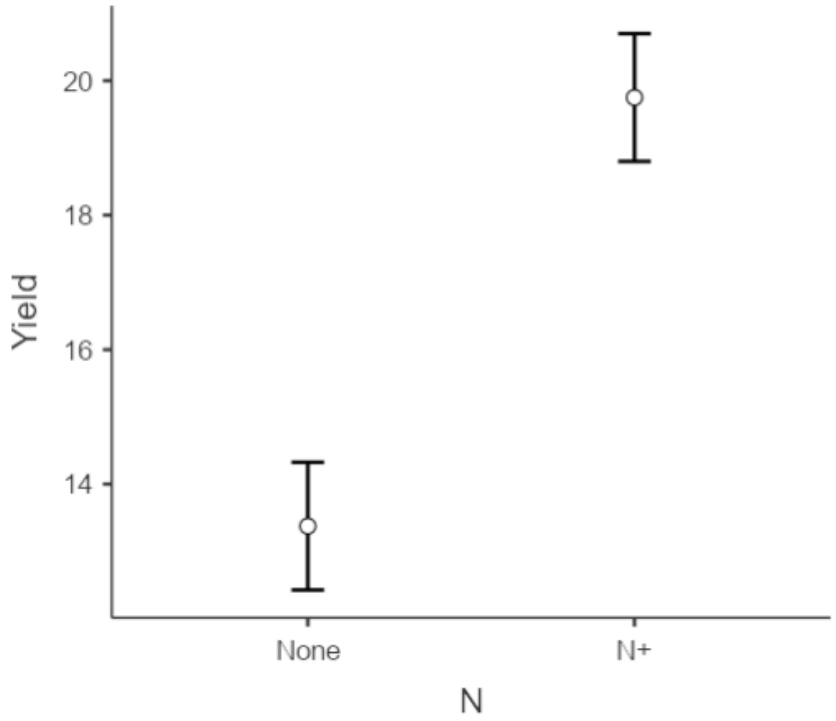
	Yield
Mean	16.6
Standard deviation	5.15



Main effect of N

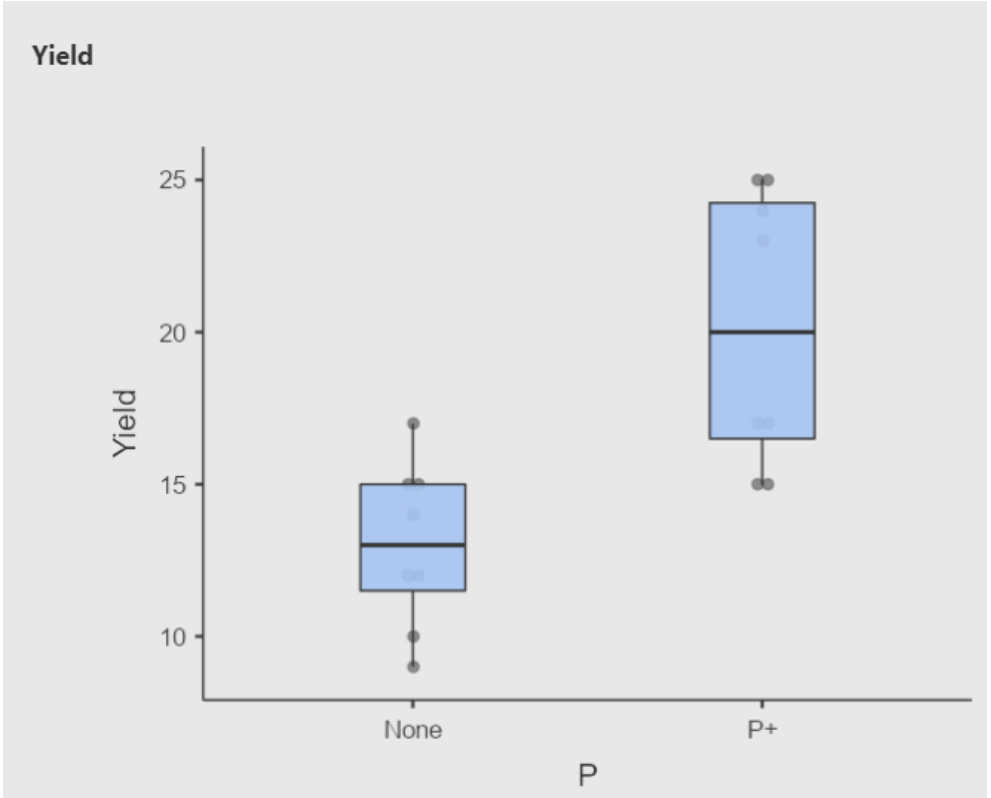
- Averaged *across* P
- Group-level variation
- Averages different?

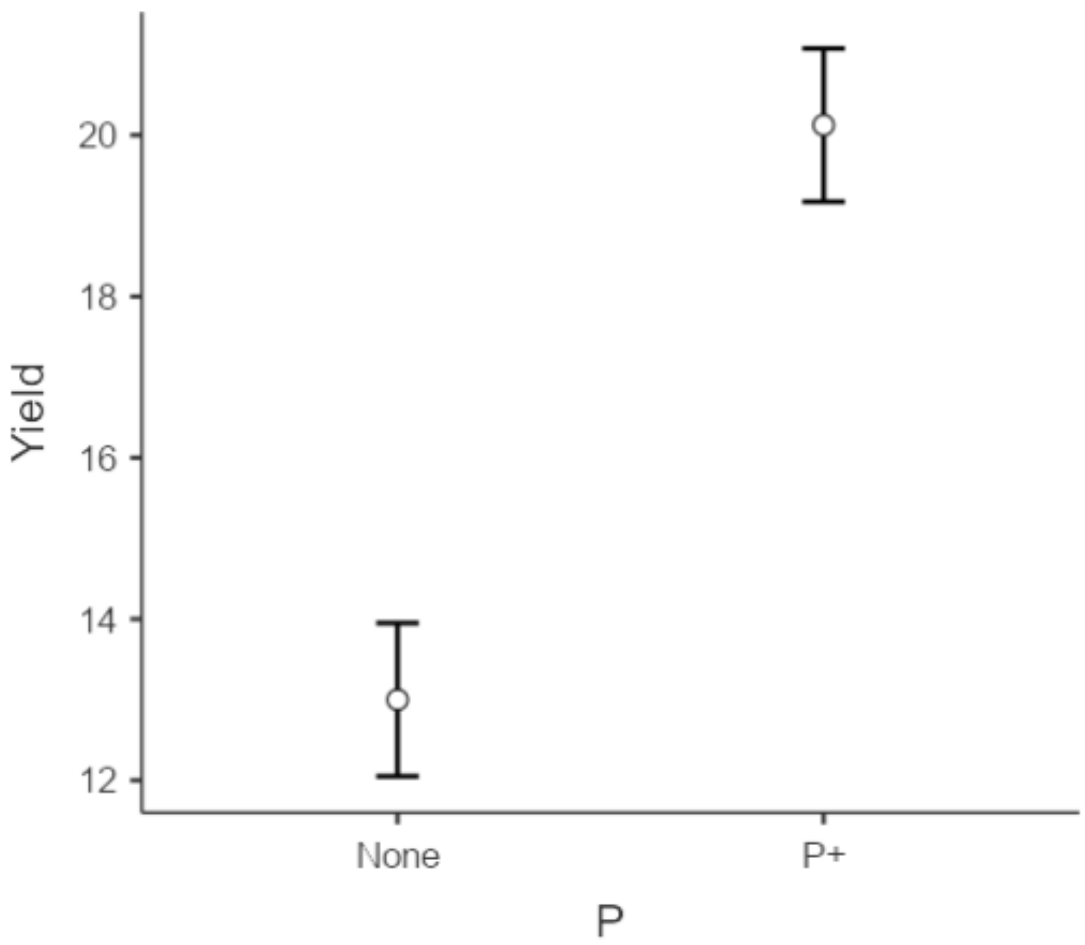




Main effect of P

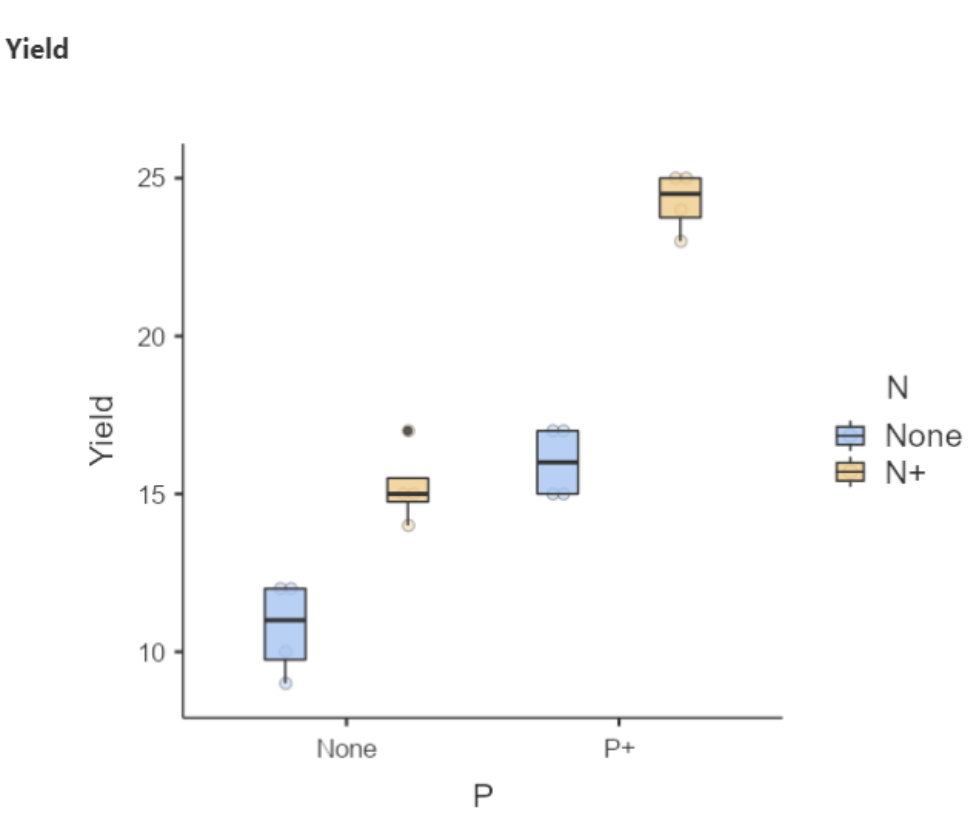
- Averaged *across* N
- Group-level Variation
- Averages different?

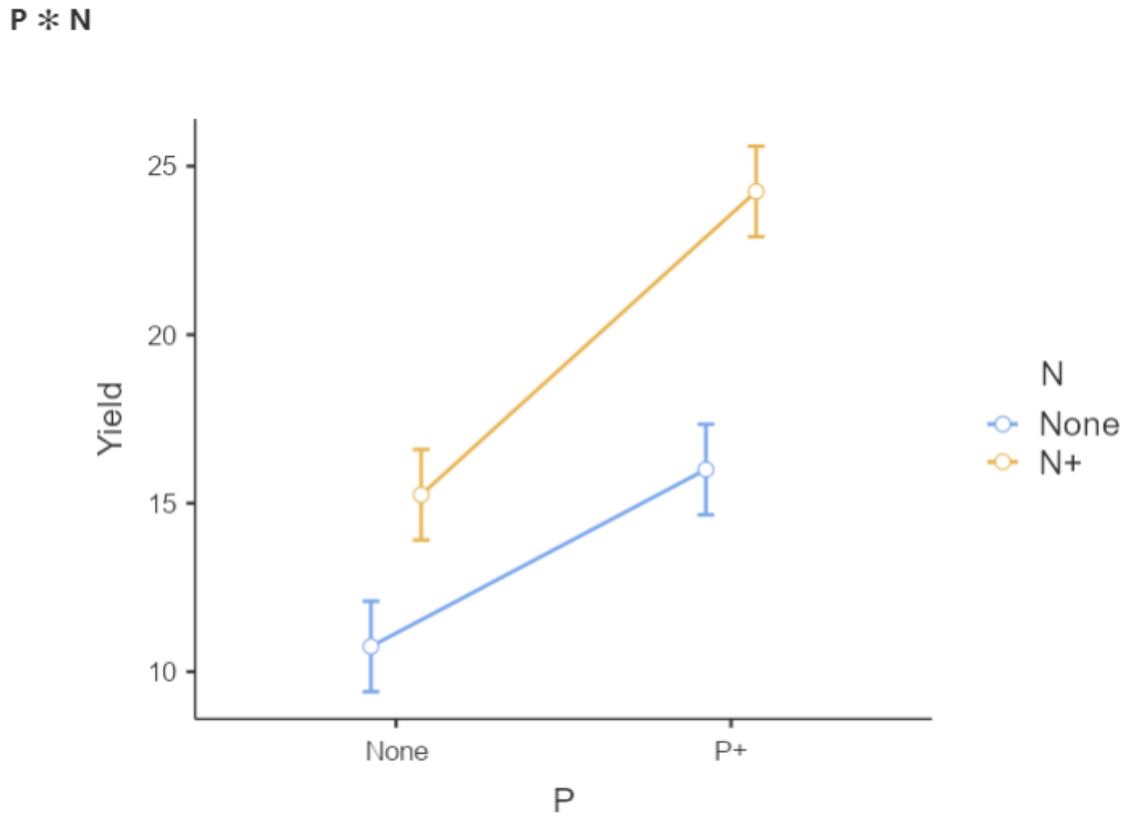




Interactive Effect

- Grouped by N & P
- ***SLOPES*** different?





Two-way ANOVA – 3 Hypotheses

1. Main Effect of Factor A
    1. H_{O1}: Mean *not* affected by factor A
    2. H_{A1}: Mean *is affected* by factor A
2. Main Effect of Factor B
    1. H_{O2}: Mean *not* affected by factor B
    2. H_{A2}: Mean *is affected* by factor B
3. Interaction effect
    1. H_{O3}: The effect of one factor *does not* depend on the other factor
    2. H_{A3}: The effect of one factor *does* depend on the other factor

No direction in hypotheses; either = or ≠

2-way ANOVA with interaction in R

- `lm(yield ~ N + P + N:P, data = corn)`
- `lm(yield ~ N * P, data = corn)`
- The `N * P` syntax will automatically include main effects and interaction
- Later, this will be important because we may not want *ALL* interactions:
- `lm(yield ~ N + P + Block + Mechanical+ N:P, data = corn)`

N+P on Corn yield: 2-way ANOVA

- H_{A1}: Phosphorous affects corn production (Factor A)
- H_{A2}: Nitrogen affects corn production (Factor B)
- H_{A3}: The effect of N on corn production will change when P is also added (or vice versa)

ANOVA - Yield

	Sum of Squares	df	Mean Square	F	p
P	203.1	1	203.06	133.52	< .001
N	162.6	1	162.56	106.89	< .001
P ✻ N	14.1	1	14.06	9.25	0.010
Residuals	18.3	12	1.52		

Results sentence(s) from ANOVA table

Based on a 2-way ANOVA, there were significant main effects of P ($F_{1, 12}$ = 133.52, p < 0.001) and N ($F_{1, 12}$ = 106.89, p < 0.001) on corn yield. In addition, the effects of one factor depended on the presence of the other factor, as demonstrated by a significant interaction term ($F_{1, 12}$ = 9.25, p = 0.01)

ANOVA - Yield

	Sum of Squares	df	Mean Square	F	p
P	203.1	1	203.06	133.52	< .001
N	162.6	1	162.56	106.89	< .001
P ✻ N	14.1	1	14.06	9.25	0.010
Residuals	18.3	12	1.52		

Results sentence(s) from ANOVA table

Based on a 2-way ANOVA, there were significant main effects of P ($F_{1, 12}$ = 133.52, p < 0.001) and N ($F_{1, 12}$ = 106.89, p < 0.001) on corn yield. In addition, the effects of one fact
presence of the other factor, as demons
interaction term ($F_{1, 12}$ = 9.25, p = 0.01)

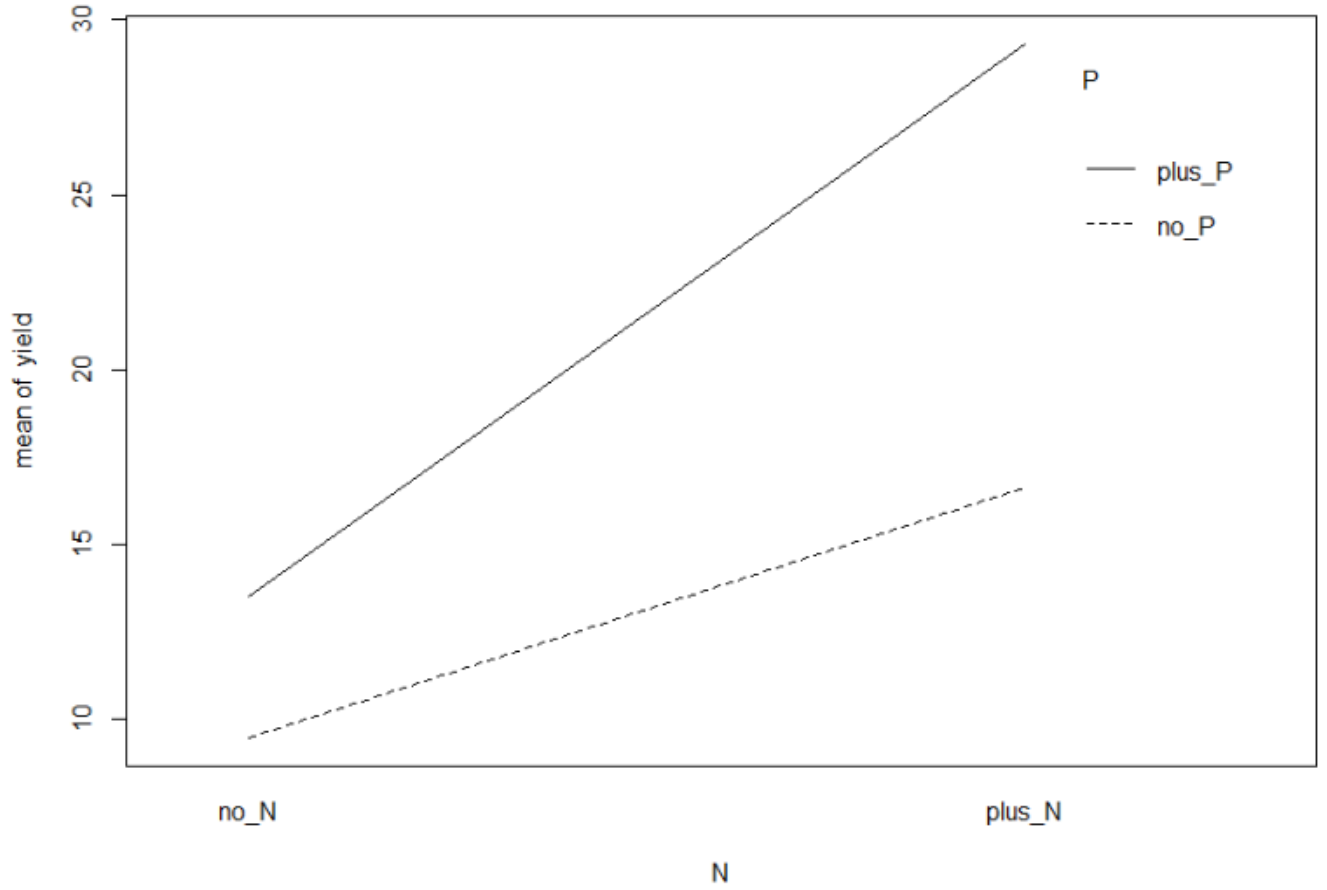
Note: do NOT try and interpret the interaction. Just say that there *is* an interaction effect.

ANOVA - Yield

	Sum of Squares	df	Mean Square	F	p
P	203.1	1	203.06	133.52	< .001
N	162.6	1	162.56	106.89	< .001
P ✻ N	14.1	1	14.06	9.25	0.010
Residuals	18.3	12	1.52		

Interaction plots, Base R

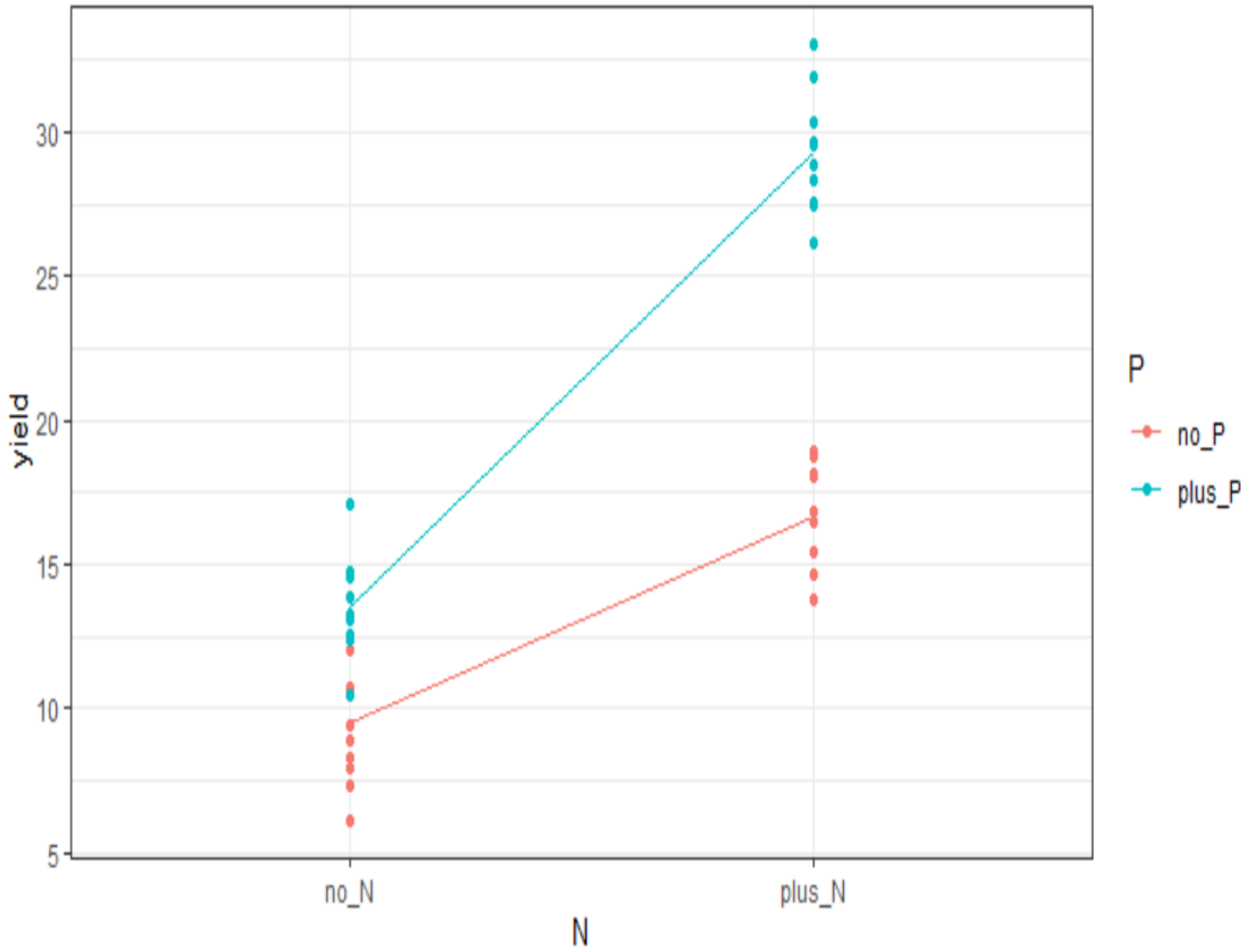
```
with(data = corn, interaction.plot(N, P, yield))
```



Interaction plots, ggplot2

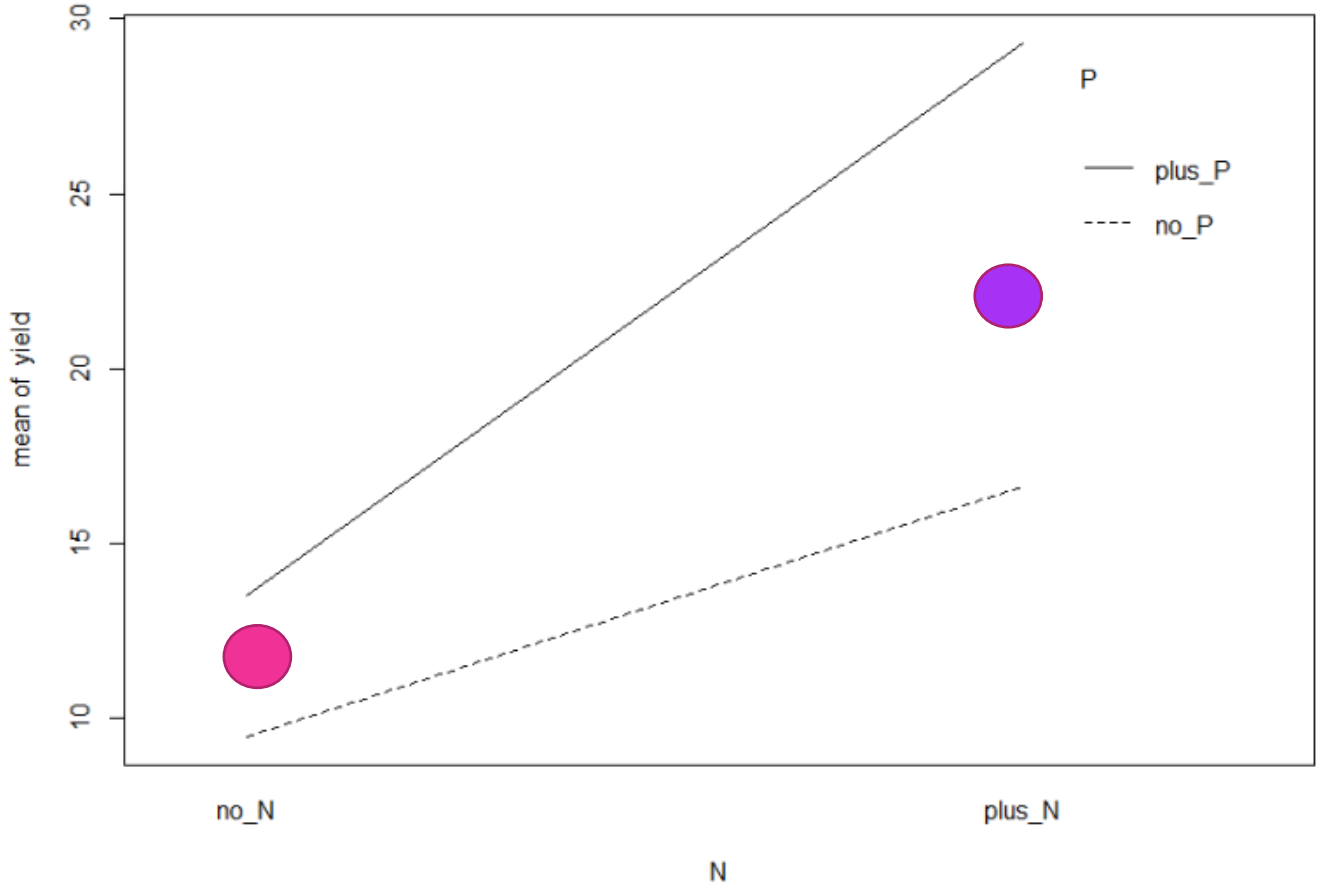
```
ggplot(dat_out,
       aes(x = N,
           y = yield,
           group = P,
           color = P)) +
  geom_point() +
  stat_summary(fun = mean,
               geom = "line") +
  theme_bw()
```

New commands in **bold**



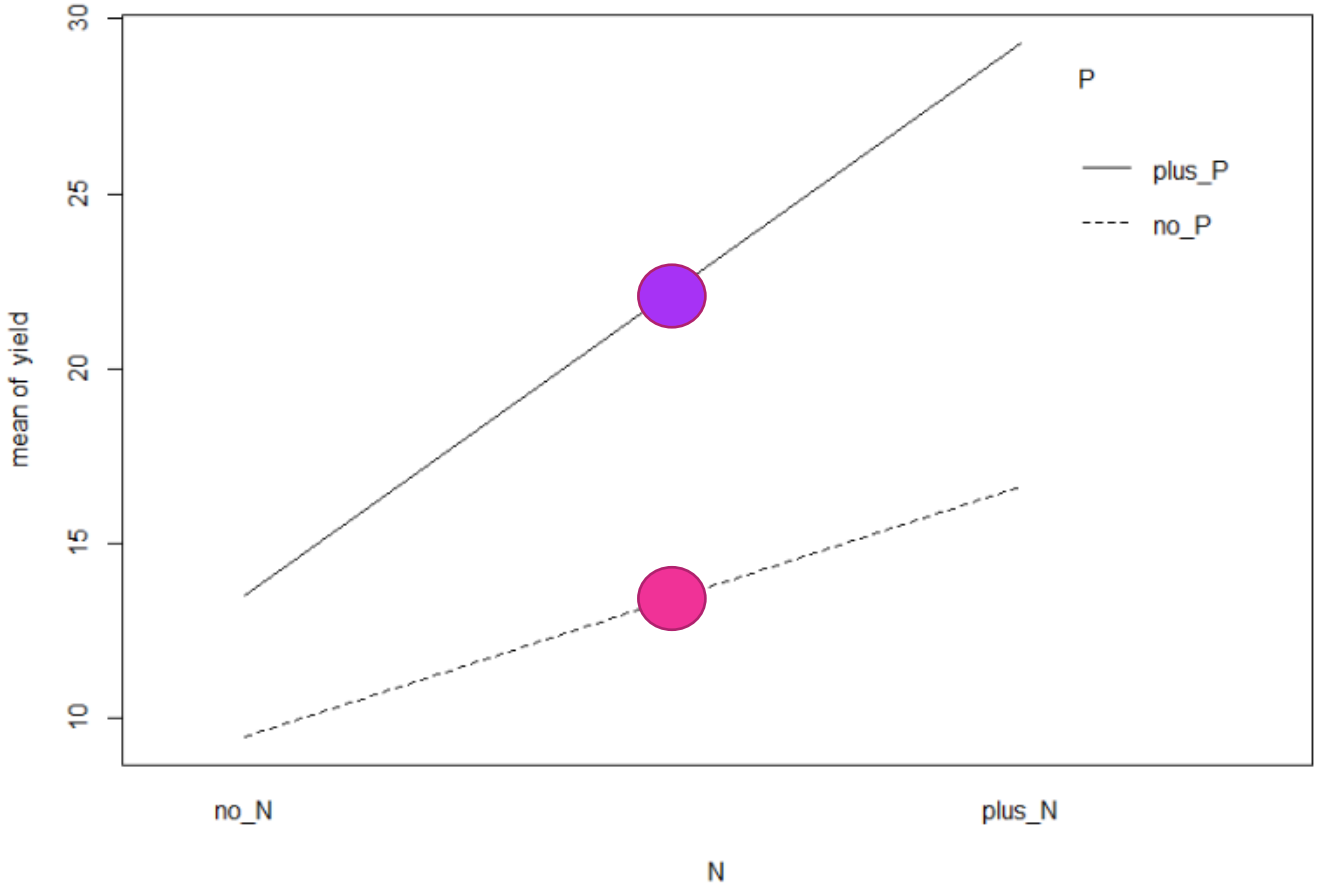
Interpreting interaction plots

- **Main effect of N**: Average L and R points between lines
- L and R different heights
- Main effect of N is likely significant



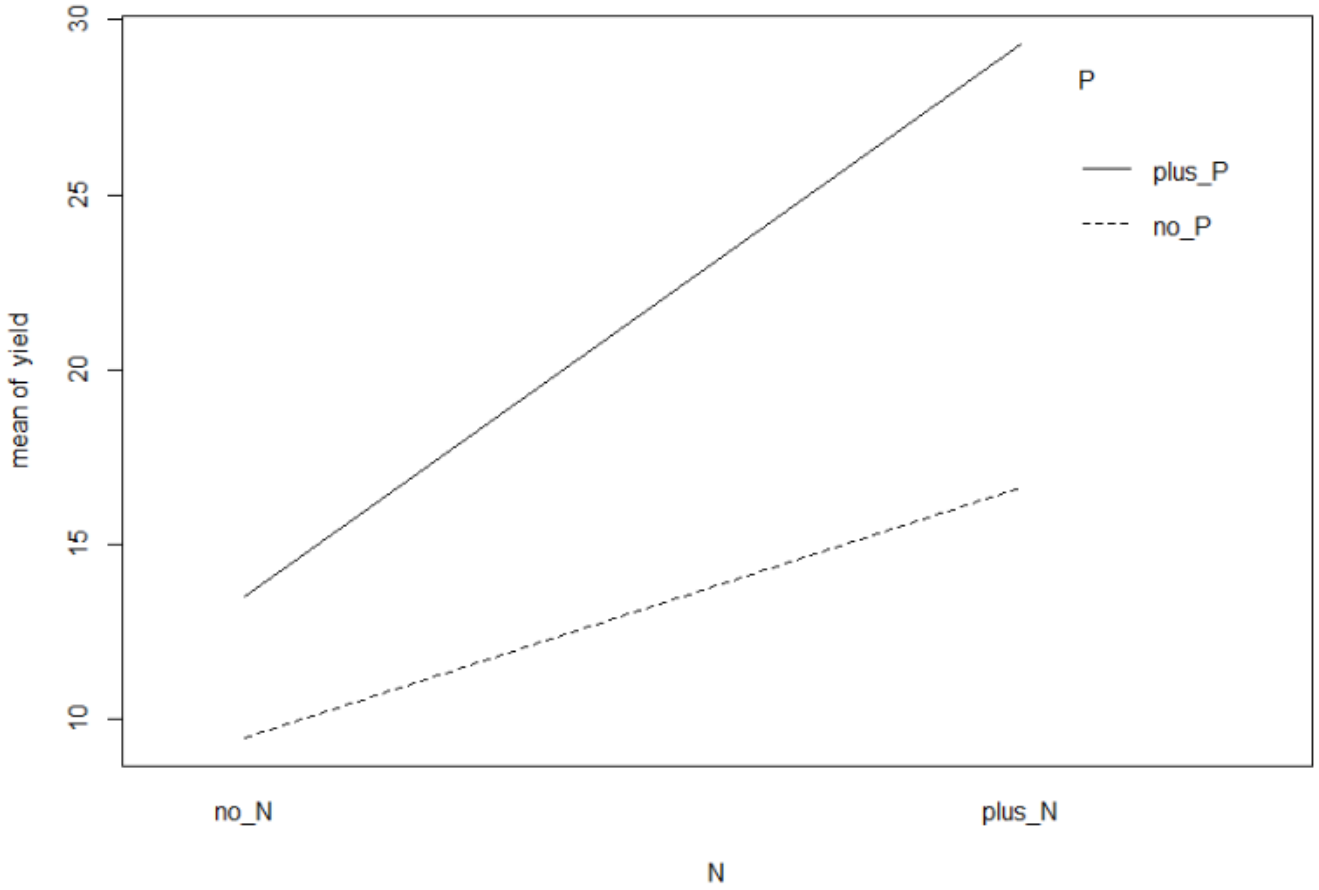
Interpreting interaction plots

- **Main effect of P**: Average lines between points
- Top and bottom are at different heights
- Main effect of P is likely significant



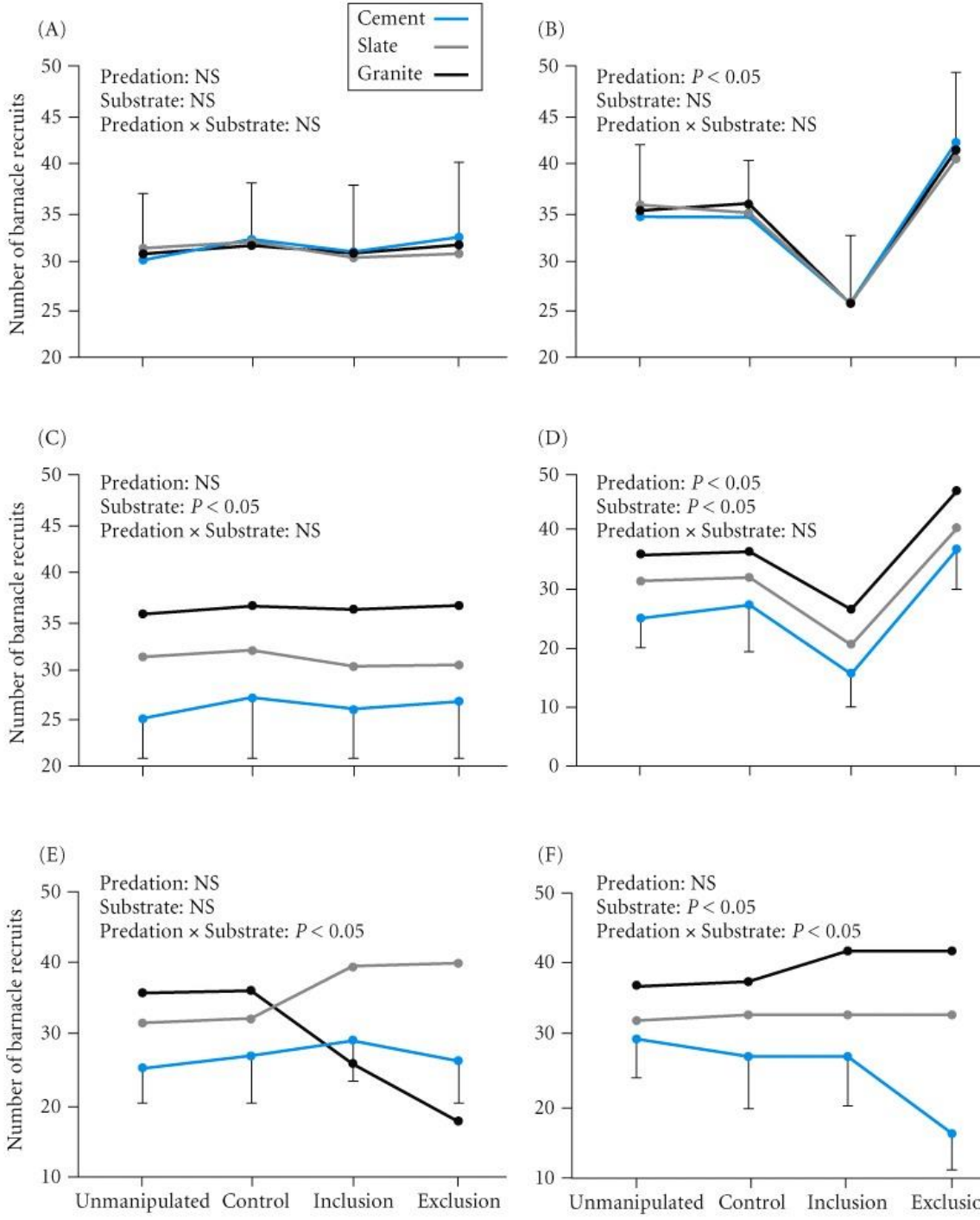
Interpreting interaction plots

- **Interactive effect**: Angle of slopes
- Lines have different slopes
- Interactive effect is likely significant
  - The effect of N *depends* on the level of P (and vice versa)



2-way with many levels

- Response: number of barnacle recruits
- 2 Factors
- Substrate: 3 levels
  - Cement, Slate, Granite
- Predation: 4 levels
  - Unmanipulated, Cage-control, Inclusion, Exclusion
- All possible responses in graph to Right
- Can be difficult to interpret (need stats!) but used as quick “Guess-stimate”



Multi-way ANOVA

- Can have more than 2 factors
  - Fire, aspect, restoration, predator removal, bird feeders, nest boxes, …
- Each factor can have more than 2 levels
  - Elevation: 6K, 8K, 9K, 10K, …
  - Treatment: Clear Cut, Thinning, Chaining, …

Hypotheses

- One hypothesis *per main effect*
- One hypothesis *for each interaction*

More complex factorial designs

Aspect

West slope

Elevation	Control	Clearcut
8,000	26	19
10,000	17	16

East slope

Elevation	Control	Clearcut
8,000	34	22
10,000	26	11

3-way ANOVA → 2 x 2 x 2 Factorial

How many interactions possible?
How many Hypotheses?

More complex factorial designs

Aspect

West

Elevation	Control	Clearcut
8,000	26	19
10,000	17	16

East

Elevation	Control	Clearcut
8,000	34	22
10,000	26	11

3-way ANOVA → 2 x 2 x 2 Factorial

How many interactions possible?
How many Hypotheses?

3 Main Effects:
Aspect
Elevation
Treatment

4 Interactions:
Aspect * Elevation
Aspect * Treatment
Elevation * Treatment
Aspect * Treat * Elevation

7 Hypotheses:
3 main effects
+ 4 interactions

M

- 3-way interactions can be difficult to interpret.
- More than 3 can be (are?) impossible to interpret.
- Best to limit design to 2-way interactions only

West

Aspect

East

10,000	26	11

3 Main Effects:
Aspect
Elevation
Treatment

4 Interactions:
Aspect * Elevation
Aspect * Treatment
Elevation * Treatment
Aspect * Treat * Elevation

3-way ANOVA → 2 x 2 x 2 Factorial

How many interactions possible?
How many Hypotheses?

7 Hypotheses:
3 main effects
+ 4 interactions

Summary

- Factorial designs let you test the effects of > 1 treatment variable
- It's important to always assess the interaction first, before assessing main effects
- If the interaction is significant, then the effect of one factor depends on the other, so they can't be interpreted in isolation
- Factorial designs are extremely powerful but hard to pull off when there are more than 2 factors (with many levels per factor)