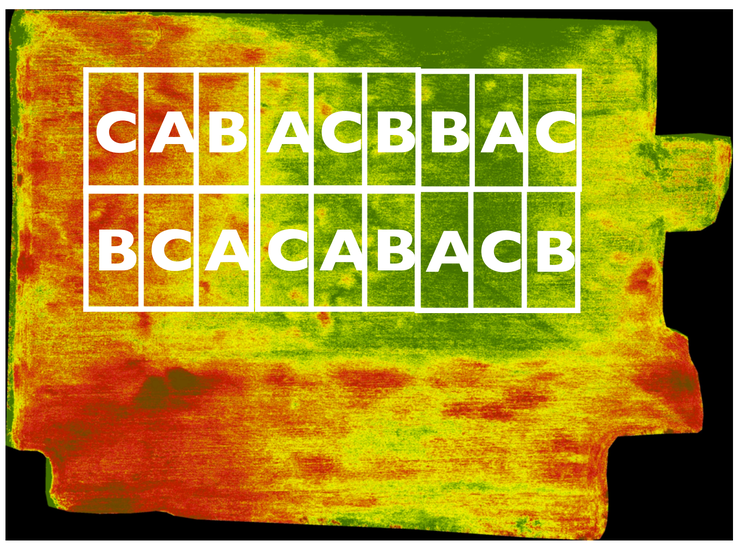
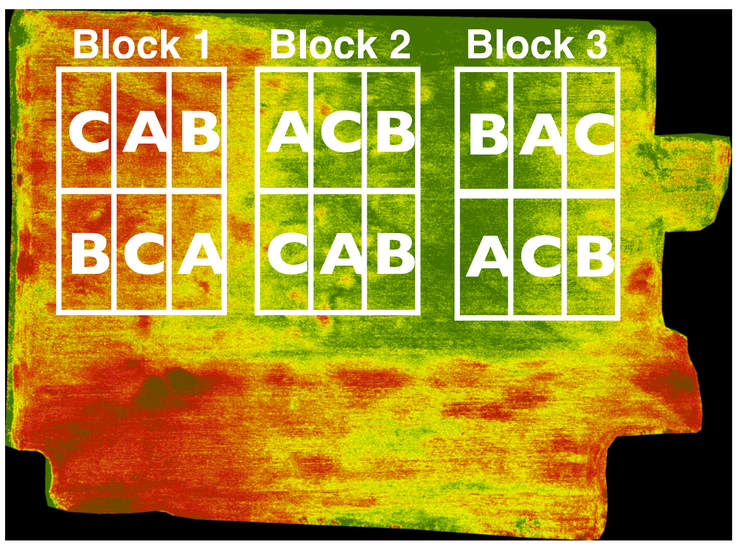
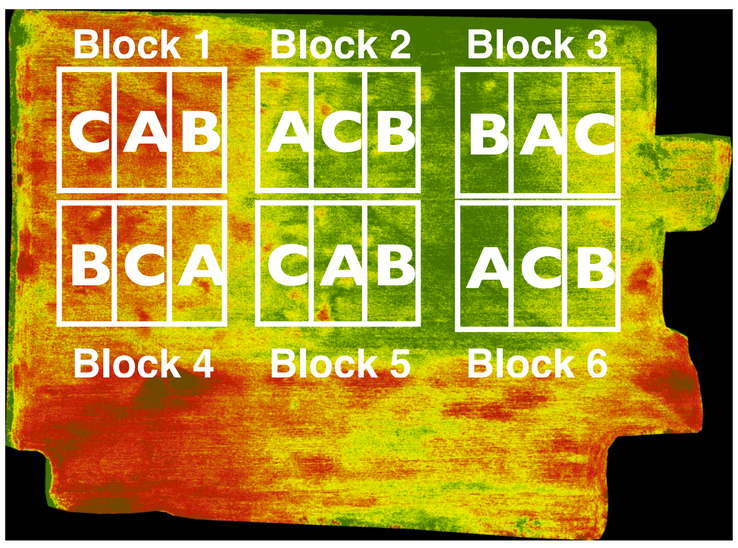
Optimal number of blocks



No Blocks



3 Blocks



6 Blocks

Compare similar EU s^2_{pooled} or s^2_{error}

Worst → Best

Degrees of Freedom - average effect (main effect)

$k*(n_i - 1) = 15$	$(k-1)*(b-1) = 4$	$(k-1)*(b - 1) = 10$

Degrees of Freedom - specific effects

0	$b*k*(n_{ij}-1) = 9$	0

Variability of treatment effects across the field

Violates assumptions Detect through diagnostics	Test Treatment:Block interactions	Increases s^2, SED

More replicates per block means:

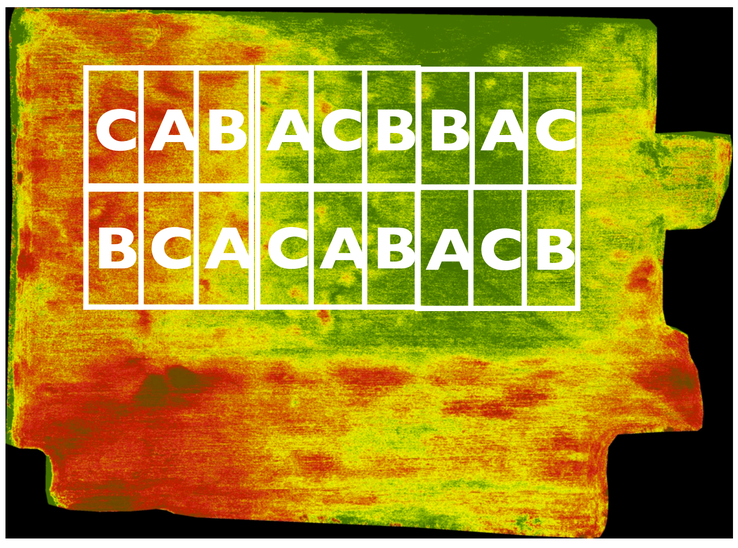
More confidence in Treatment:Block interactions

Less power for main effects

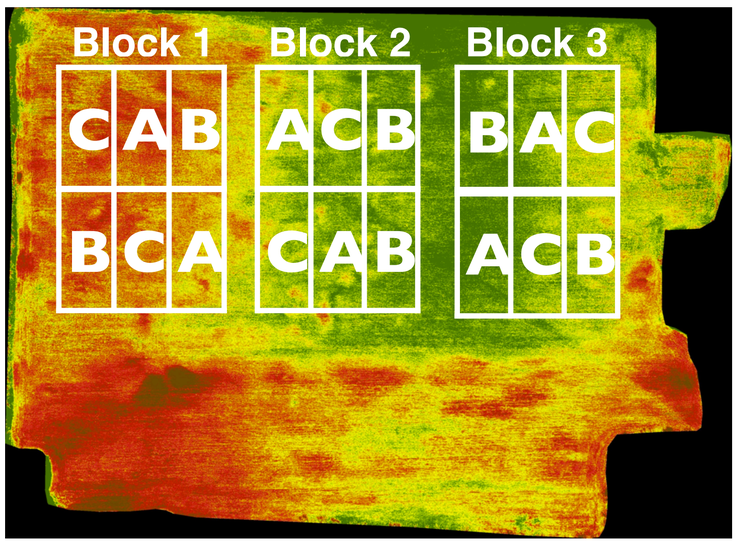
Higher DF

Less similar EU (higher s^2)

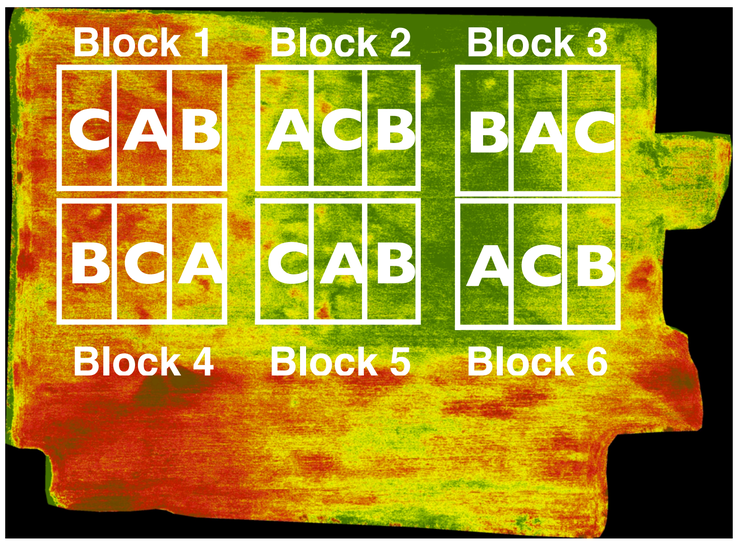
Optimal number of blocks



No Blocks



3 Blocks



6 Blocks

Recommendations:

Don’t block unless you can identify **clusters** of EU

More (small) blocks are best for studying **main effects**

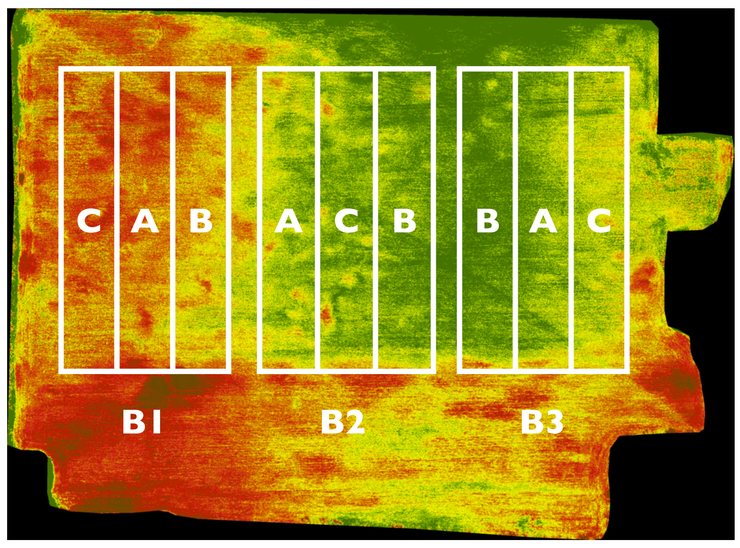
But, blocks should be representative of your target population

Only do replication of treatments within blocks if:

1) You want to know if treatment effects vary in your population

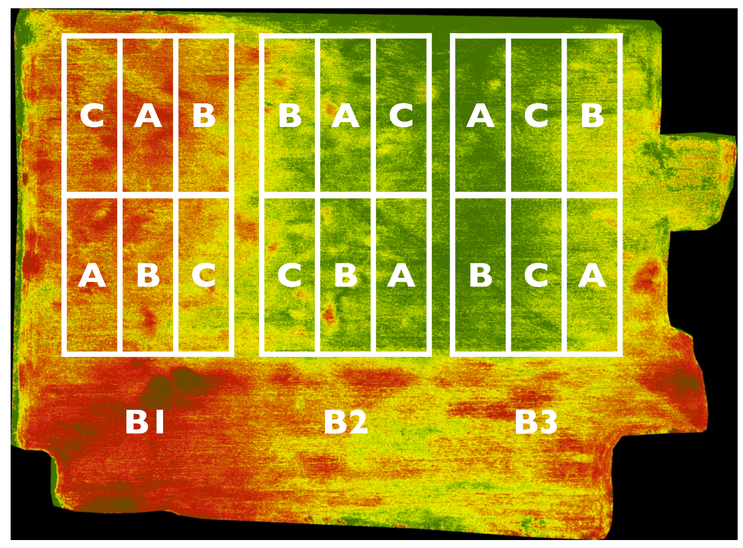
2) It’s not feasible to do smaller blocks that “make sense”

Recommending Designs



RCBD

CRD



RCBD with Reps

Which design would you use to make an Insecticide recommendation to **this farmer** (in this field)?

Reps are overkill, replicate plots within blocks are sub-samples; not-interspersed

Say the farmer could target regions within a field? Which would you use?

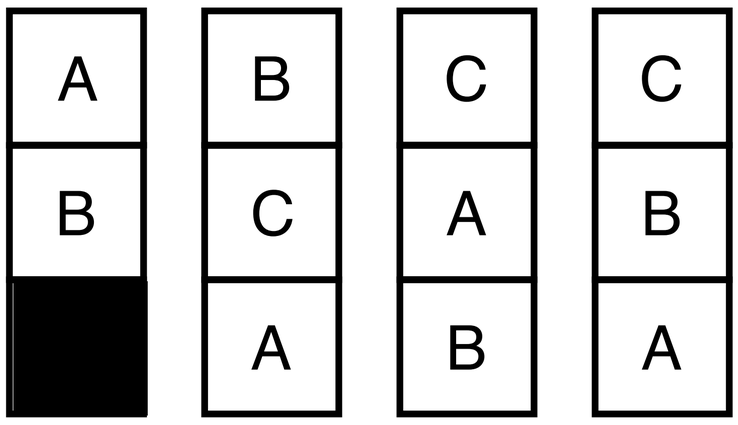
RCBD with reps is necessary (more like a factorial)

What design would you use to make a recommendation in a new field?

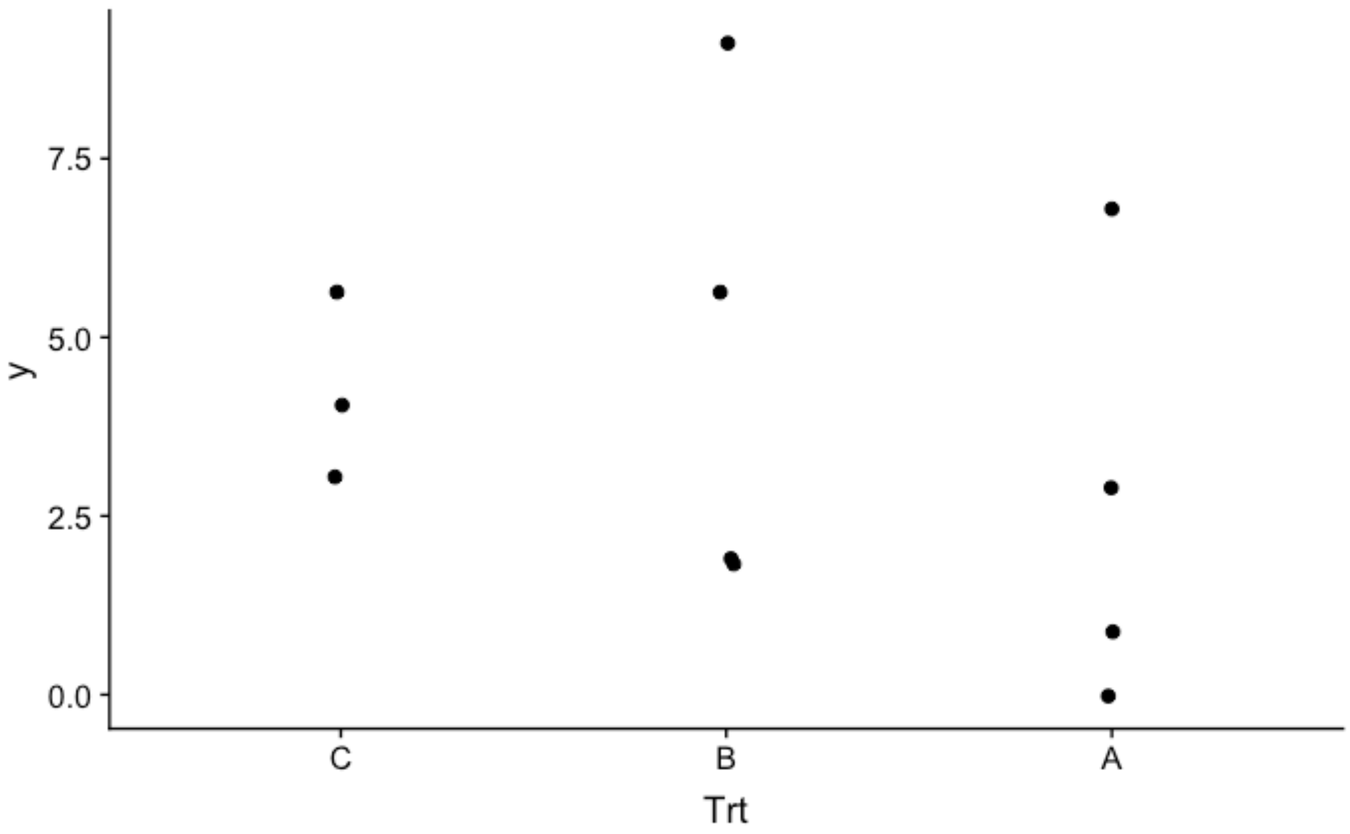
RCBD with Fields as blocks to estimate **main effect** of insecticide across all fields

Factorial with **Field types** as moderator to estimate **specific effects** in **certain types of fields**

This would be an example of a **split-plot experiment**



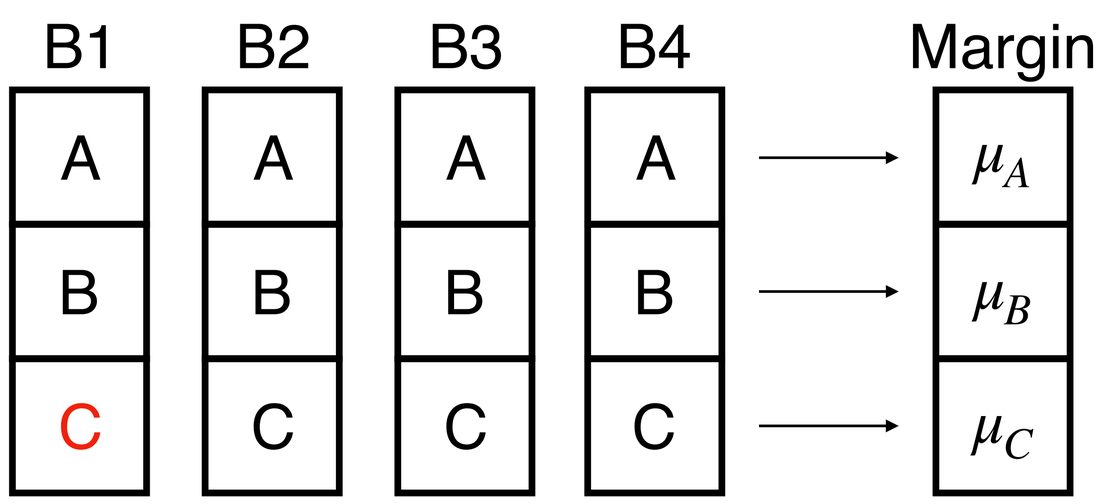
	μ_A	μ_B	μ_C
#reps	4	4	3
estimate	$\frac{\sum y_{Ai}}{4}$	$\frac{\sum y_{Bi}}{4}$	$\frac{\sum y_{Ci}}{3}$
$\sigma^2_{\hat{\mu}}$	$\frac{\sigma^2_{EU}}{4}$	$\frac{\sigma^2_{EU}}{4}$	$\frac{\sigma^2_{EU}}{3}$




```r
aggregate(y~Trt,data_m,FUN=mean)
```

Trt <fctr>	y <dbl>
C	4.333333
B	4.750000
A	2.750000

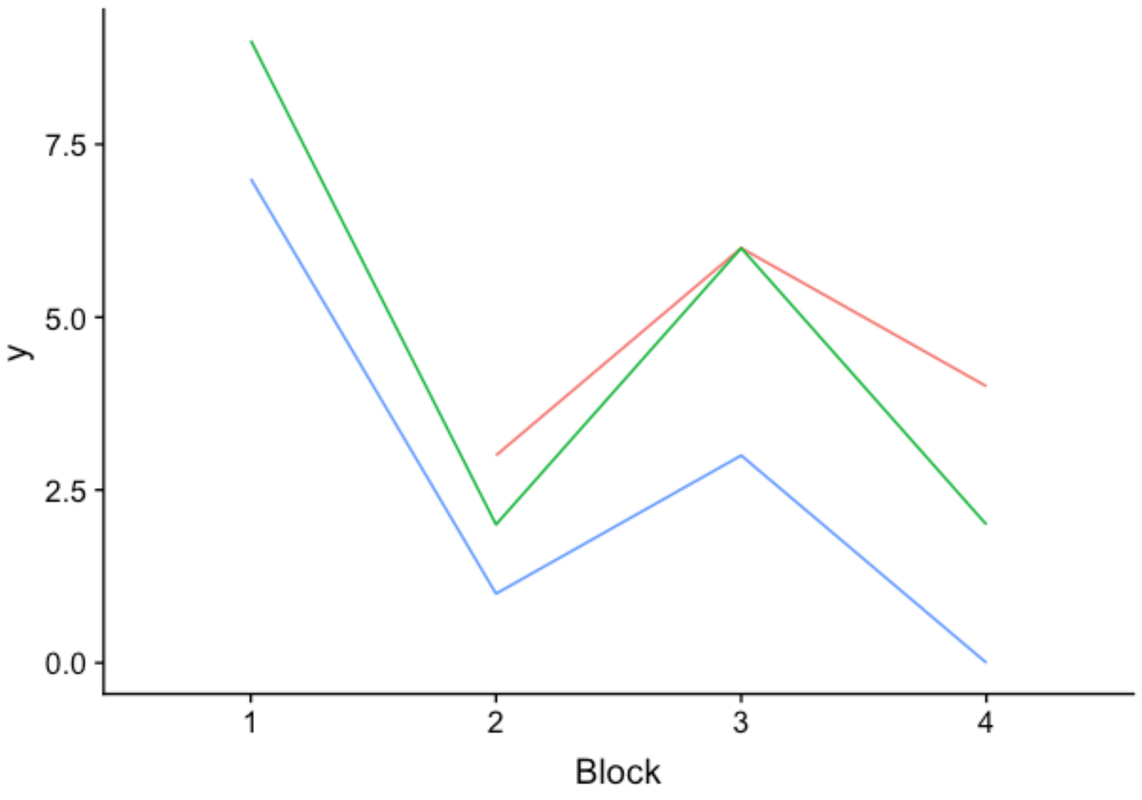
3 rows



"emmeans" = Expected Marginal Means

predict the mean had the experiment been complete

SEs and df adjusted for actual data




```{r}
emmeans(lm(y~Block+Trt,data_m),~Trt)
```

```
 Trt emmean        SE df lower.CL upper.CL
 C     5.75 0.3872983  5 4.754418 6.745582
 B     4.75 0.3162278  5 3.937111 5.562889
 A     2.75 0.3162278  5 1.937111 3.562889

Results are averaged over the levels of: Block
Confidence level used: 0.95
```

$\hat{\delta}_{B-A}$ = ave (B-A) in 4 blocks

$\hat{\delta}_{C-A}$ = ave (C-A) in 3 blocks — low SE (within blocks)

= (C-B) - (B-A) — high SE (across blocks)

B2-B4 B1

A1 is **connected** to C(2-4) through B

Analysis with missing cells

emmeans: same as normal

ANOVA: order of terms in formula matters

```{r}
lm1 = lm(y~Block+Trt,data_m)
anova(lm1)
```

```
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
Block      3 58.909  19.636  49.091 0.0003933 ***
Trt        2 16.000   8.000  20.000 0.0041152 **
Residuals  5  2.000   0.400
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```{r}
lm_switch = lm(y~Trt+Block,data_m)
anova(lm_switch)
```

```
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
Trt        2  8.742  4.3712  10.928 0.0149561 *
Block      3 66.167 22.0556  55.139 0.0002972 ***
Residuals  5  2.000  0.4000
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

*put blocks first

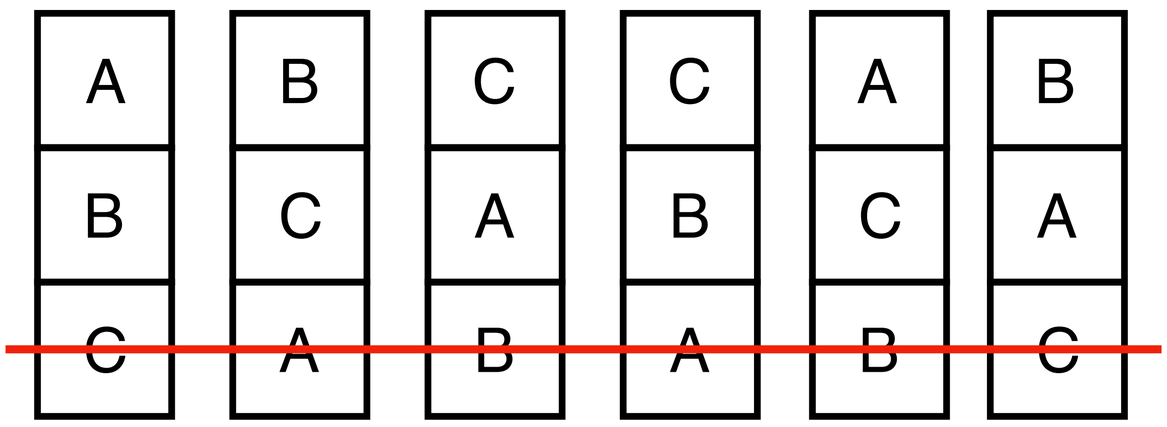
Can declare Blocks as random, may reduce SE of differences if there are many blocks

y ~ (1|Block) + Trt

Say we want to evaluate three types of eyeglass materials

Can we use blocks in this experiment?

 can be a block, but cannot be a **complete block**



	1	2	3	4	5	6
A	1		1		1	1
B	1	1		1		1
C		1	1	1	1	

Balanced Incomplete Block Design (BIBD)

Incomplete Block: not every treatment in every block

t: # treatments k: # treatments/block k < t

Balanced: every pair occurs together same # of times

λ: ave # times pairs occur together λ = 2

smaller blocks than RCBD

more homogeneous, easier to run

all pairwise comparisons equally powerful

optimized for “pairwise”

hard with many treatments

needs to be very large to achieve balance

t=8, k=4 => need at least 14 blocks

Description:

b: # blocks

t: # treatments

k: # trt/block

r: # reps/trt

λ: ave # times pairs occur together

λ = r(k-1)/(t-1)

Augmented design

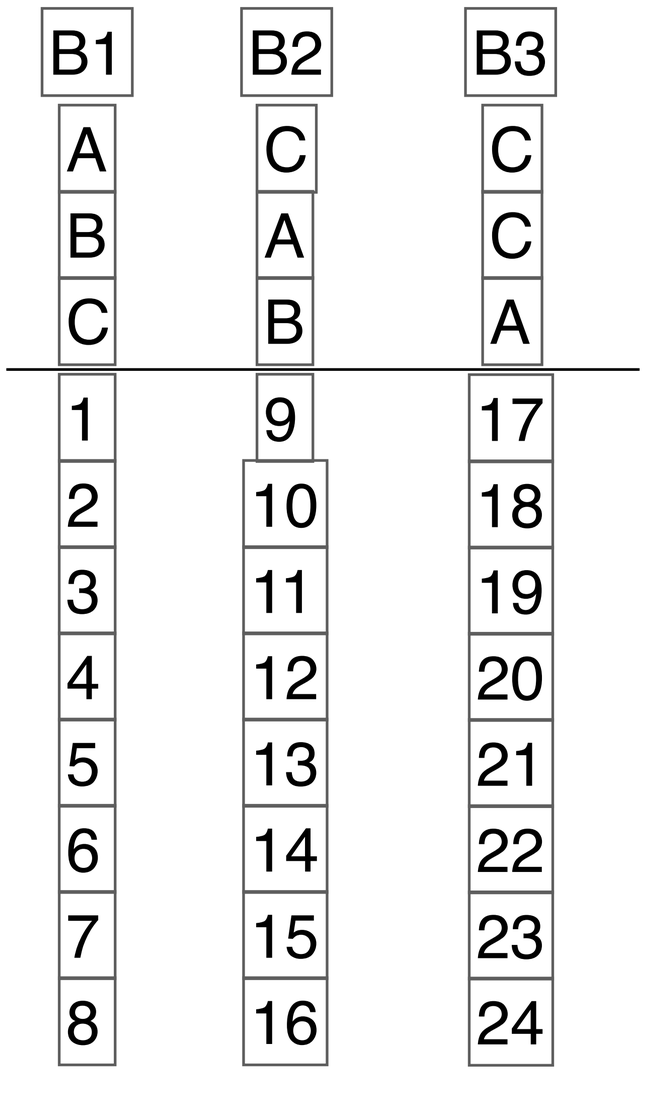
Incomplete block design

Control treatments

New treatments

optimized for $\hat{\delta}_{\text{ctrl-new}}$

Dunnett-type



RCBD for the controls

Un-replicated for new treatments

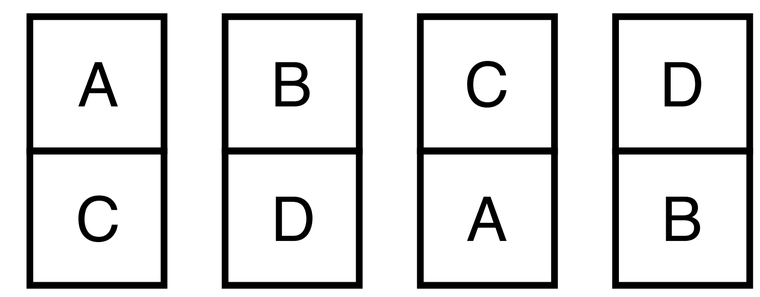
But all connected

(11-1) = (11-A) - (A-1)

$\hat{\delta}_{B-A}$

$\hat{\delta}_{2-A}$ Different SEs

$\hat{\delta}_{12-2}$



Unconnected design

```{r}
table(data$Block,data$Trt)
```

```
  1 2 3 4
1 1 0 1 0
2 0 1 0 1
3 1 0 1 0
4 0 1 0 1
```

```{r}
lm1 = lm(y~Block+Trt,data)
emmeans(lm1,pairwise~Trt)$contrasts
```

```
 contrast   estimate        SE df t.ratio p.value
 1 - 2         nonEst        NA NA      NA      NA
 1 - 3    -0.4306599 0.4025535  2  -1.070  0.7381
 1 - 4         nonEst        NA NA      NA      NA
 2 - 3         nonEst        NA NA      NA      NA
 2 - 4     0.2626295 0.4025535  2   0.652  0.9068
 3 - 4         nonEst        NA NA      NA      NA

Results are averaged over the levels of: Block
P value adjustment: tukey method for comparing a family
of 4 estimates
```

Incomplete block designs

Goal: smaller blocks can be more homogeneous

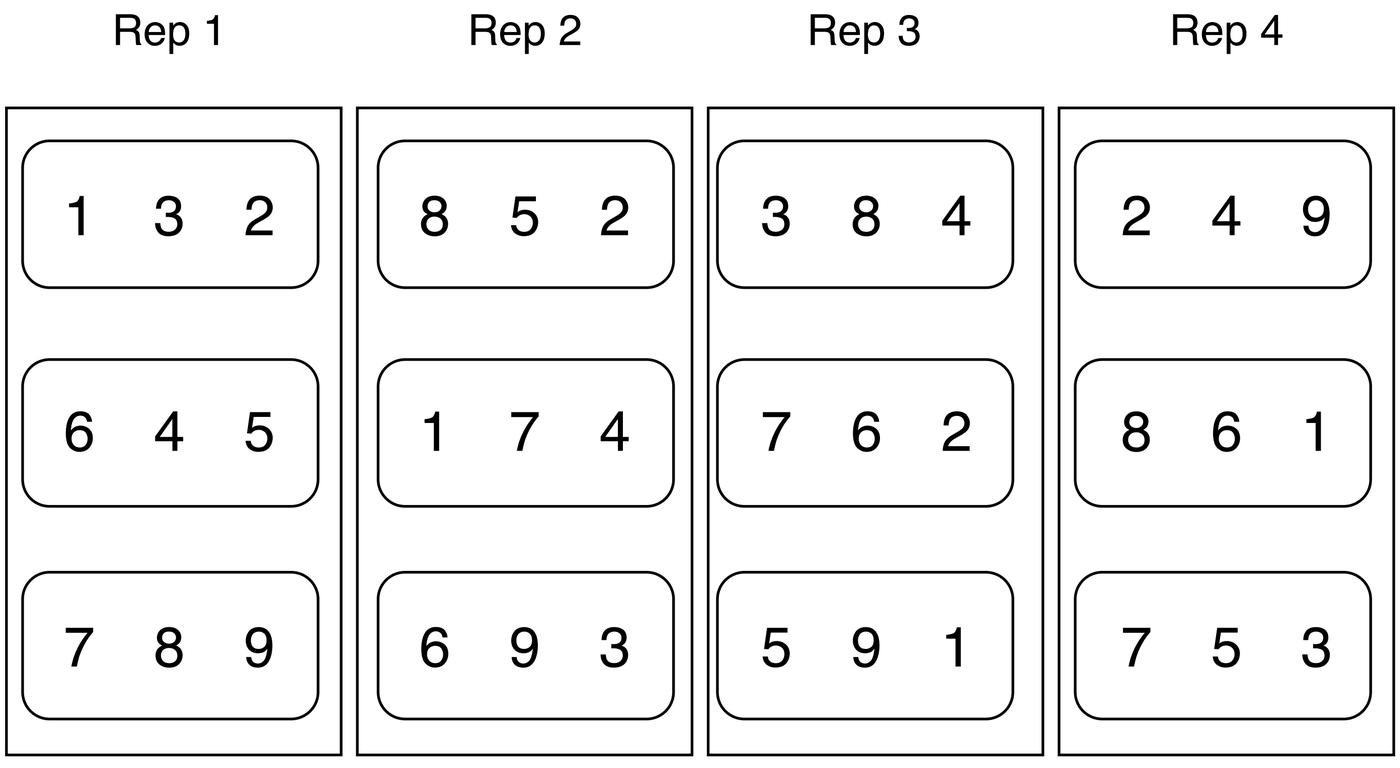
Strategy: Think about **balance** and **connectedness**

Analysis:

Put Blocks first in formula. Consider declaring Block to be random

No change in emmeans

Use emmeans before experiment to check connectedness



Blocks?

Incomplete?

Balanced?

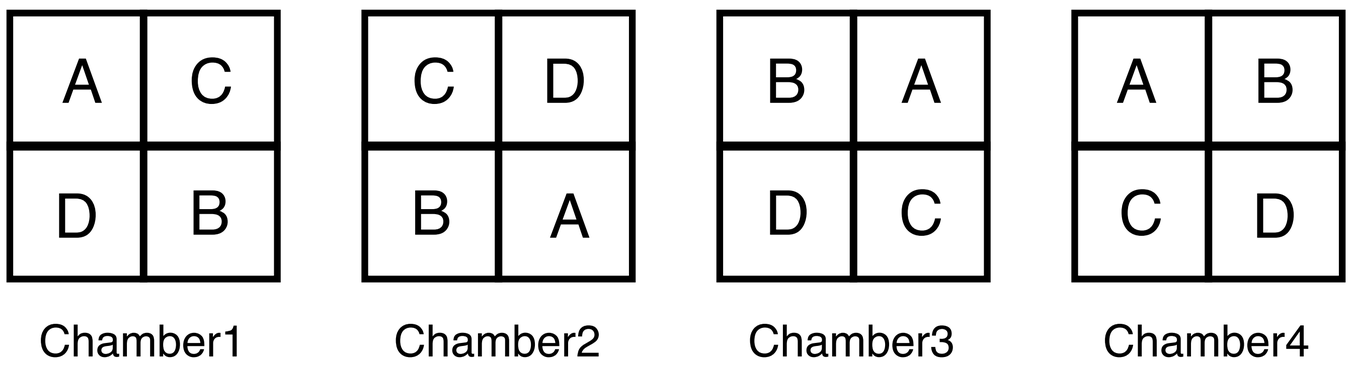
Resolvable Incomplete Block Design

Resolvable: each Rep is a complete block

Can analyze each Rep separately, combine. Or drop an entire rep.

An experiment was run to compare 4 tomato genotypes for their flowering times

The experiment was done in growth chambers with 1 plant of each in each chamber



What type of experiment is this?

What is the EU?

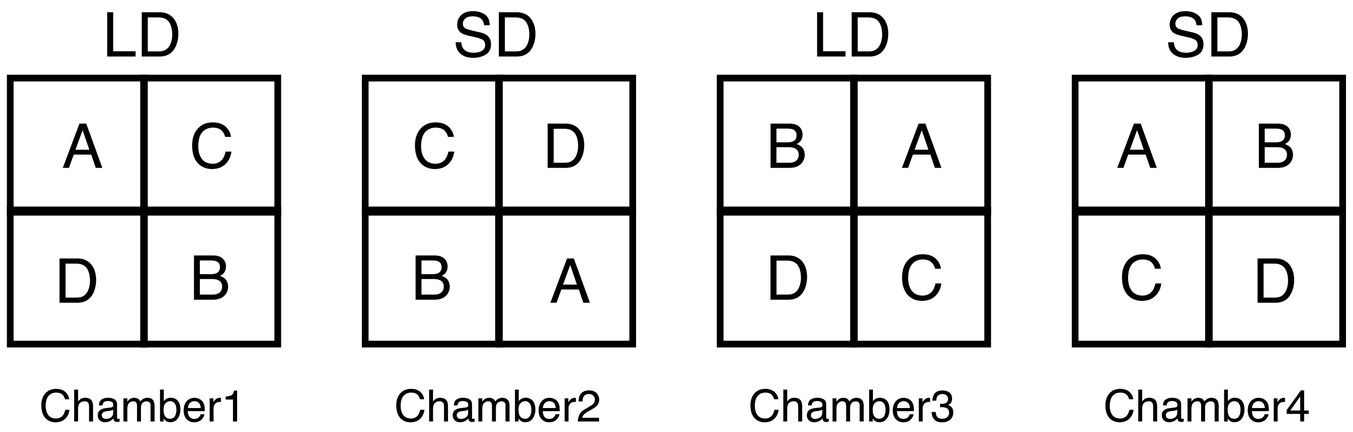
What estimates does the researcher want to report?

An experiment was run to compare 4 tomato genotypes for their flowering times

The experiment was done in growth chambers with 1 plant of each in each chamber

The experimenter was interested if daylength altered the flowering effects among genotypes

Two chambers were set to 16h days (Long days = LD) and two to 8h (Short days = SD)



What type of experiment is this? Factorial? RCBD? Split-plot factorial

What is the EU? Depends on the treatment

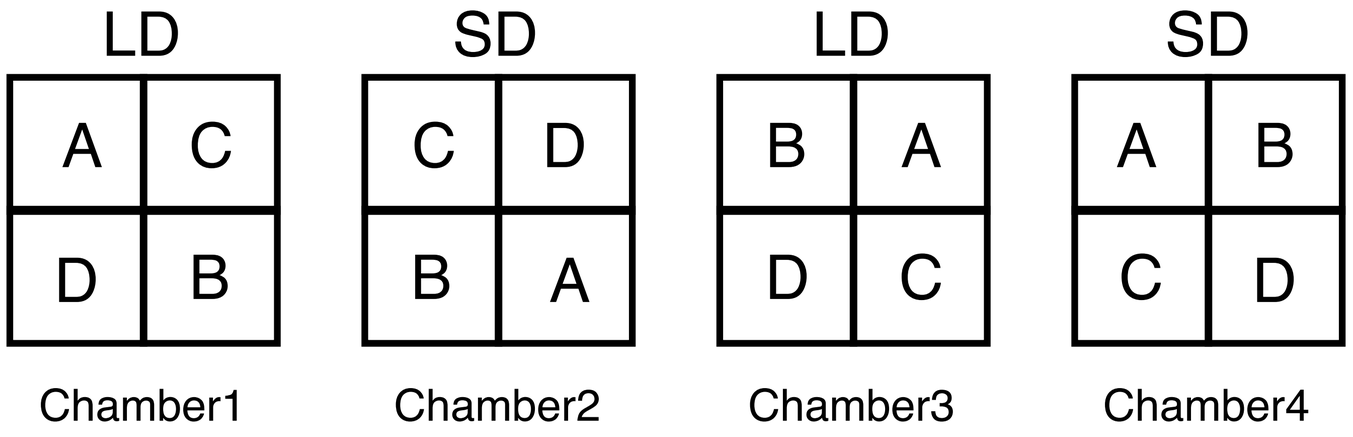
What estimates does the researcher want to report? focal = Genotype moderator = Daylength

Specific effects (Genotype effects at each Daylength)

Interaction effects (Daylength effects on Genotype differences)

What are the consequences of the Split-plot design for the analysis?

Multiple EU, different SEs for different types of effects



Split-plot design

Design used for factorial experiments

Two-stage randomization

1) Daylength treatments randomized to chambers

2) Genotypes randomized to pots **within** each chamber

Naming:

Chambers = **Main Plots**

Plant = **Sub Plots / Split-plots**

EUs:

Different for each treatment variable Be sure to declare them all!

Blocks:

Different for each treatment variable Be sure to declare them all!

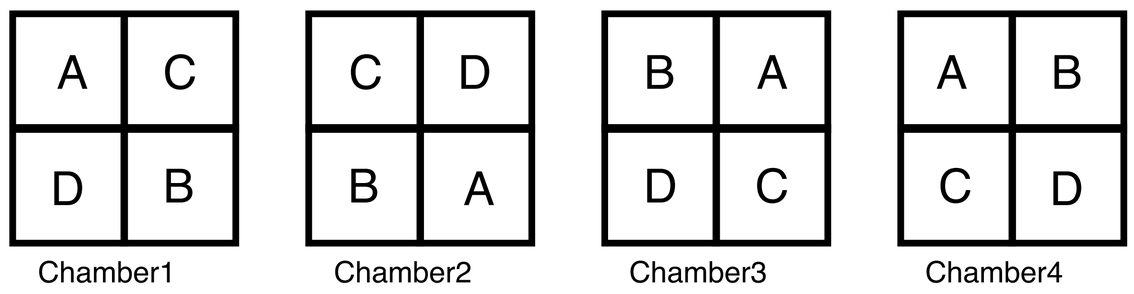
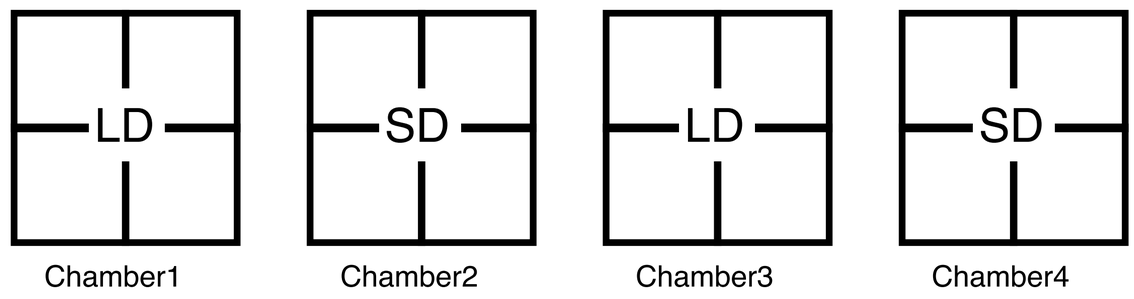
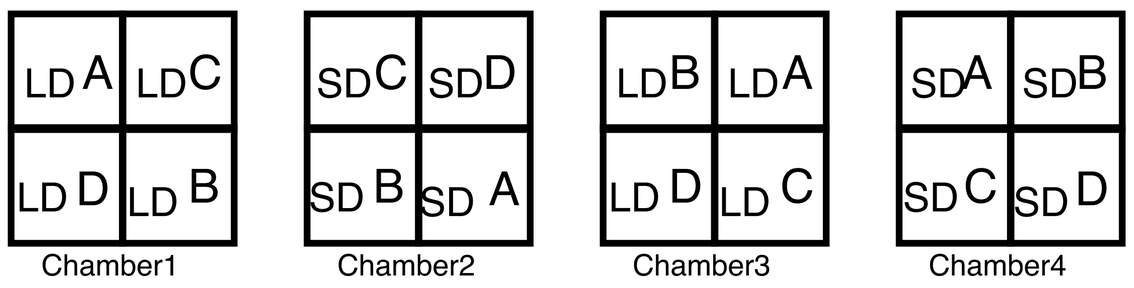
Idea

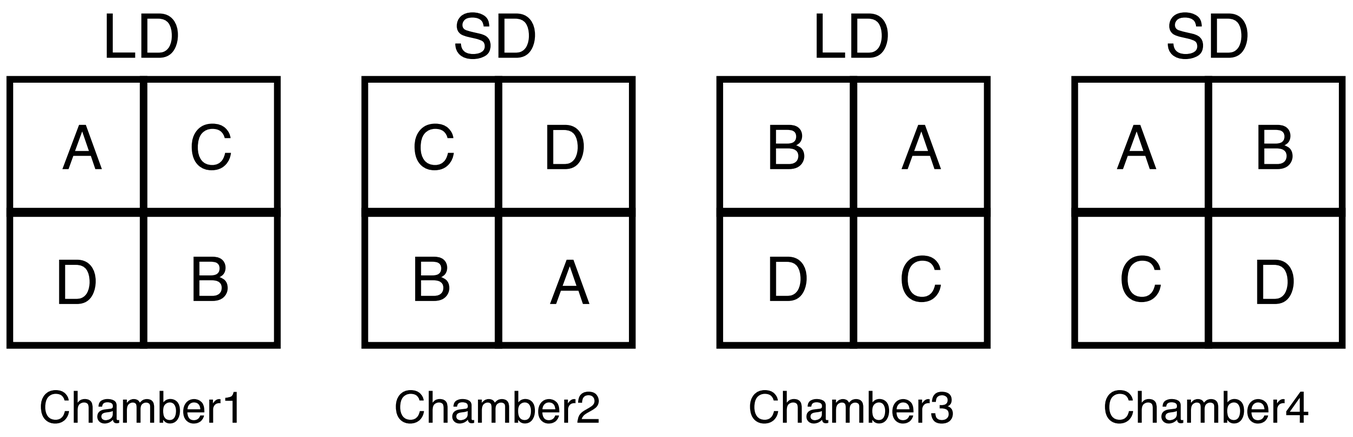
Sacrifice replication of MainPlot treatment

to increase replication of SubPlot treatment

and treatment combinations

Three experiments

		EU?	**Block?**
1) Genotype		Plant	Chamber
2) Daylength		Chamber	None
3) Genotype:Daylength		Plant	Chamber Incomplete Block



Focal: Geno Moderator: Daylength

Structure	Variable	#levels	Block	Experimental Unit
Focal	Geno	4	Dayl, Chamber	Plant
Moderator	Dayl	2	None	Chamber
Combo	Geno:Dayl	8	Chamber	Chamber
Design	Chamber	4		
	Plant	16		
	Geno:Chamber	16		
	Geno:Dayl:Chamber	16		
Response	Days	16		

lmer(Days~Geno+Dayl+Geno:Dayl +(1|Chamber))

	level relationships	Keep term?	Make combinations?
aliased	levels are 1:1	only one of set	No
nested	1-many	all	No
crossed	many-many	all	Yes

Repeat with each new Variable that you create from a combination

Keep all terms needed for clarity (usually Plot/SampleID)

Always keep EU as a Variable