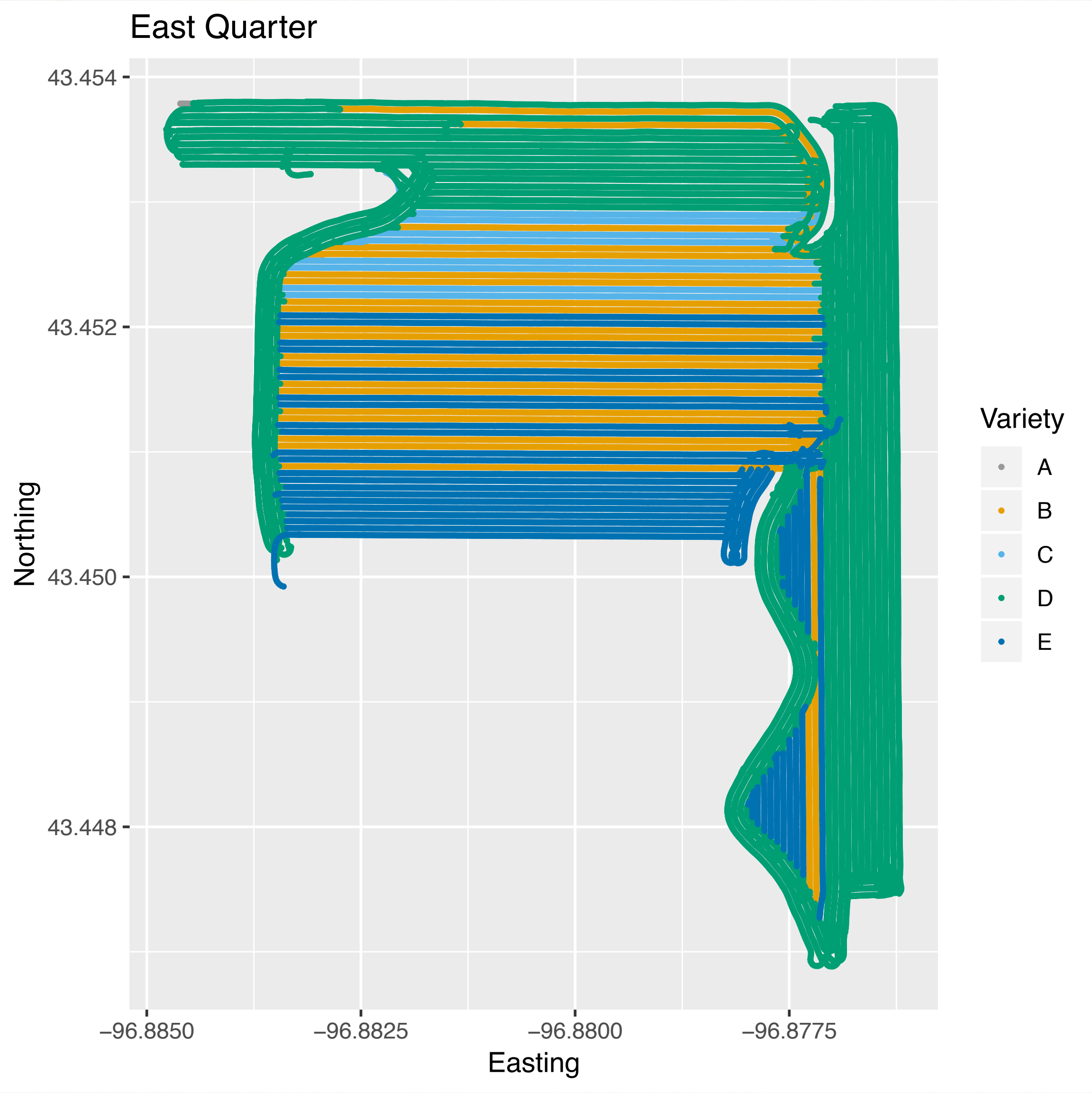
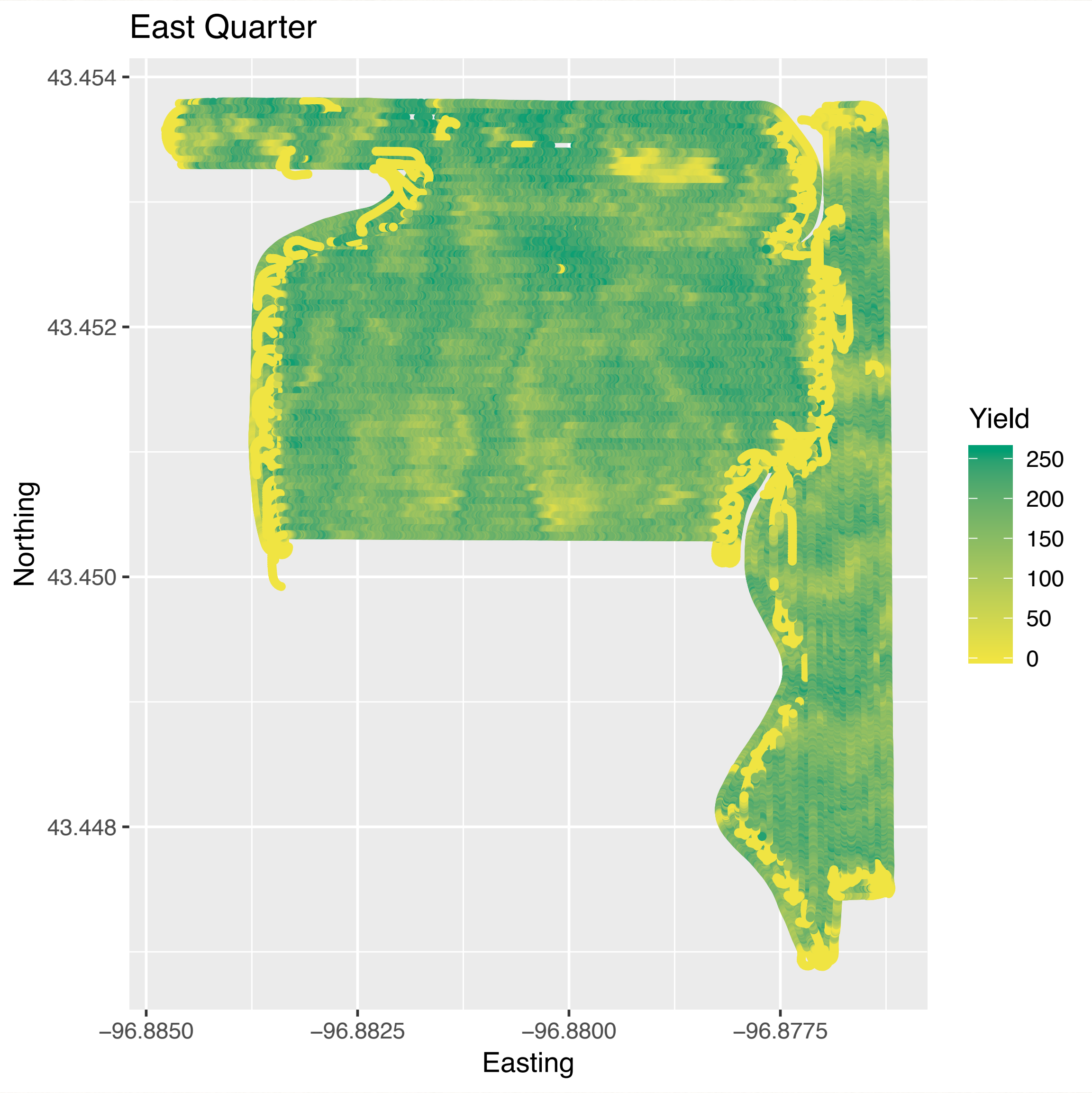
Thoughts about Power and Sample Size for On-Farm Strip Trials



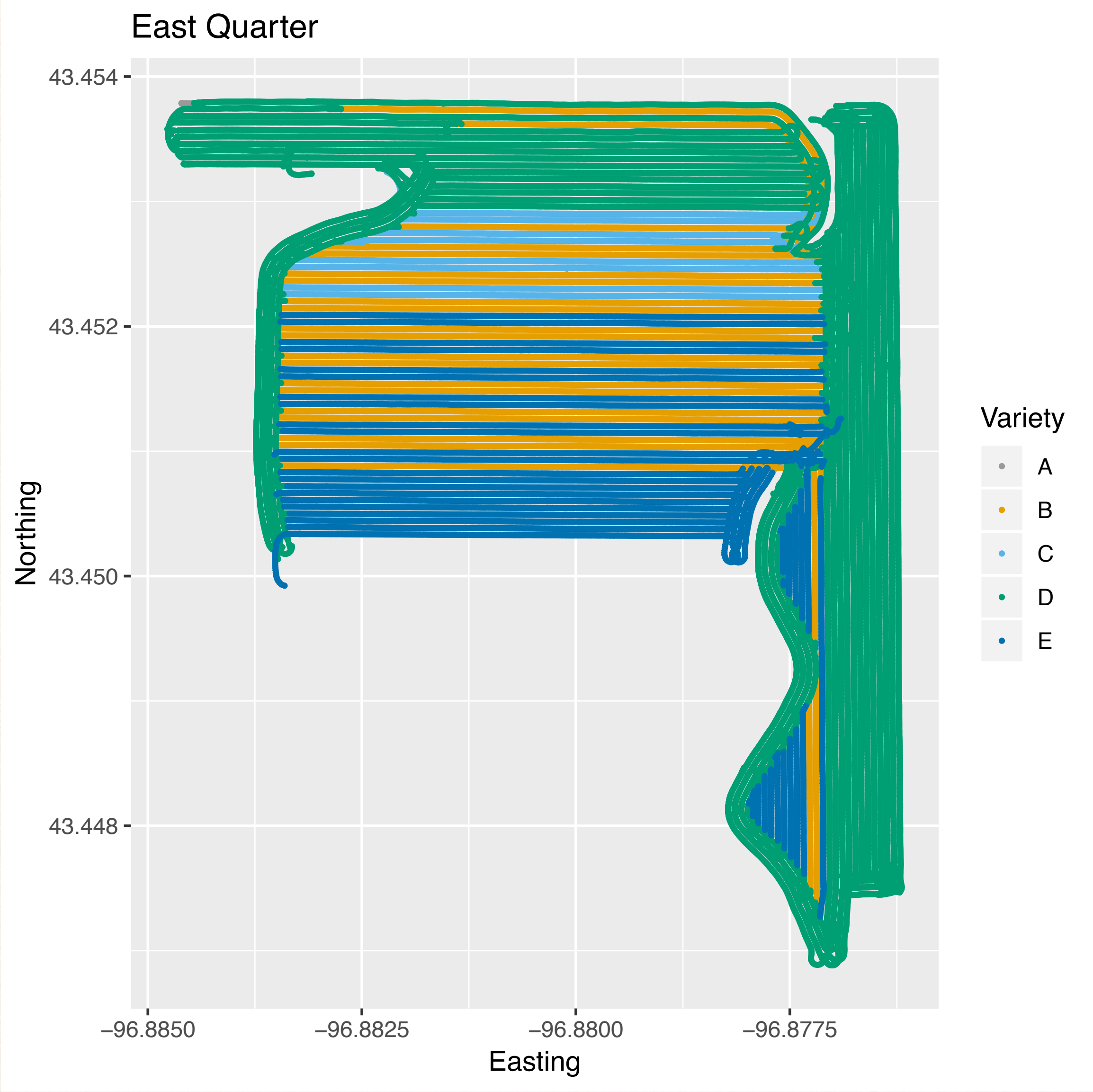


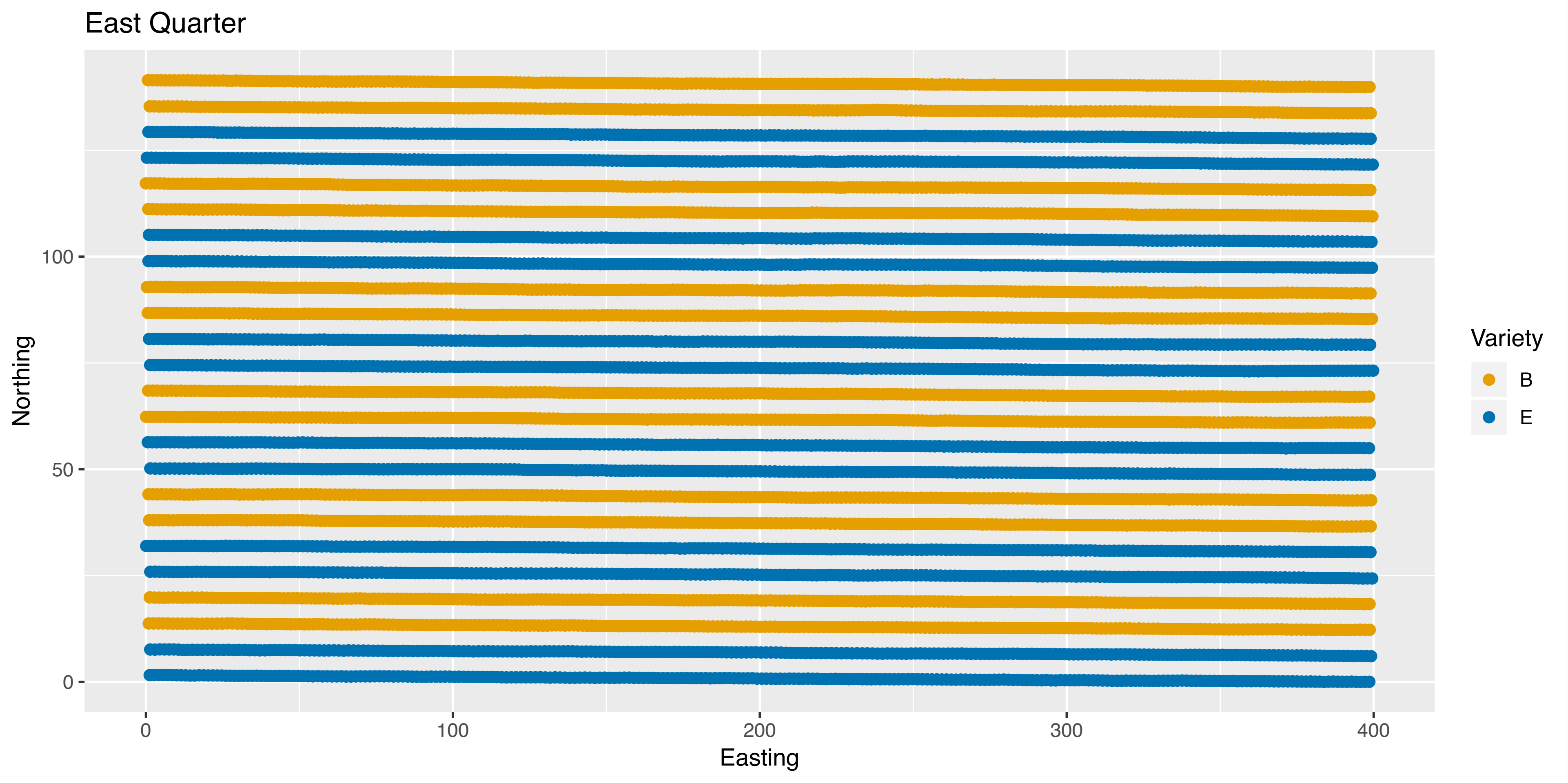
Example Strip Trial

Planting and Yield Maps

Design

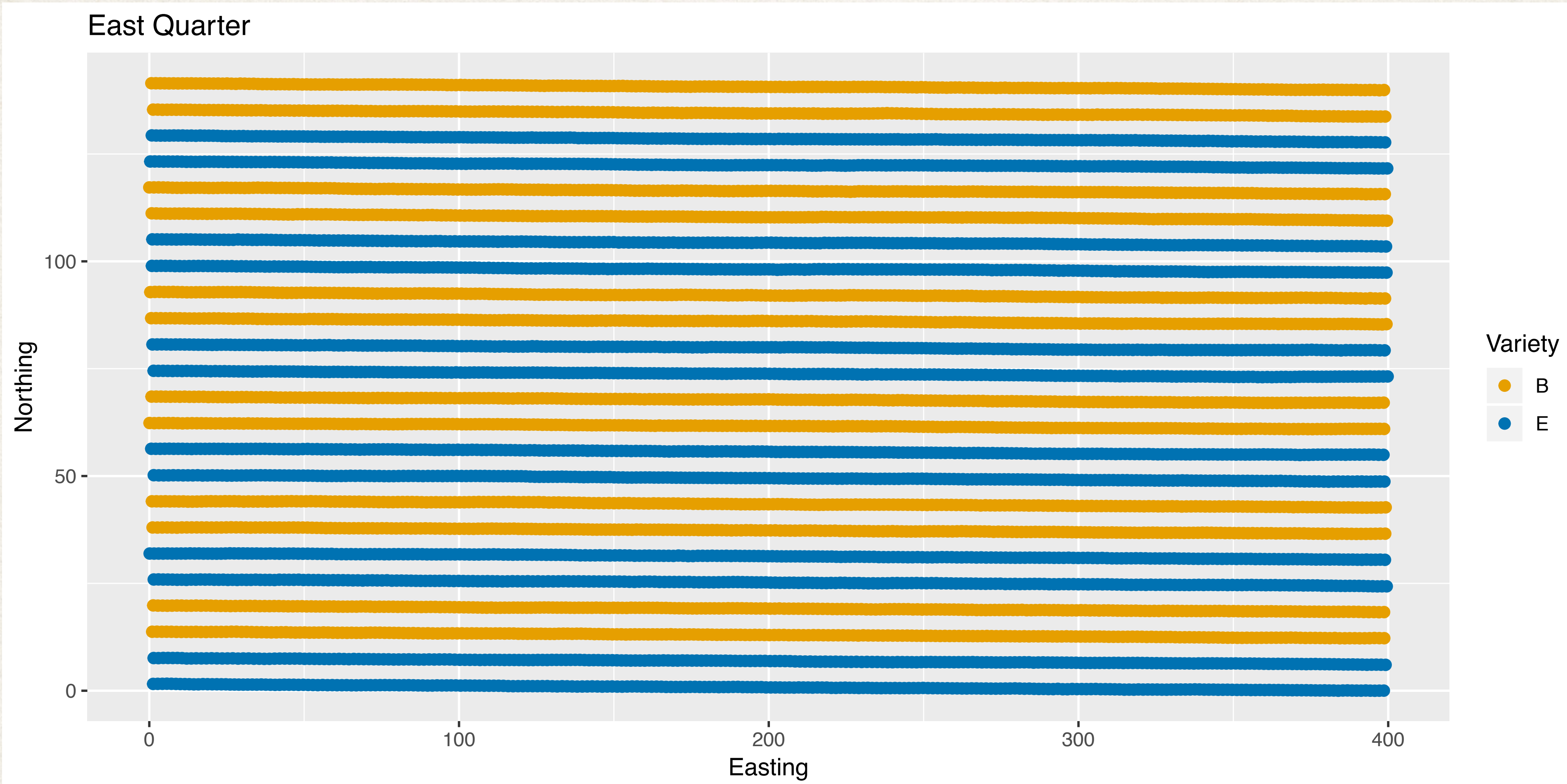
- corn variety trial
- split-planter comparison
- treatment unit is a strip one-half the width of the planter
- for simplicity, experimental unit is two treated strips of the same variety

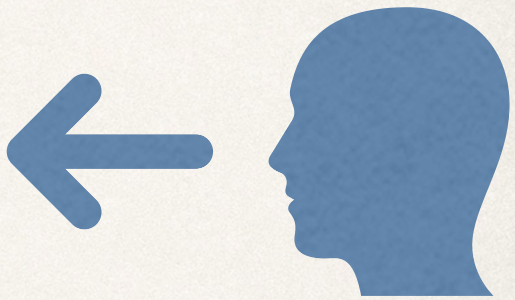




Analyzed Strips

Use only two varieties from the middle of the field to illustrate method





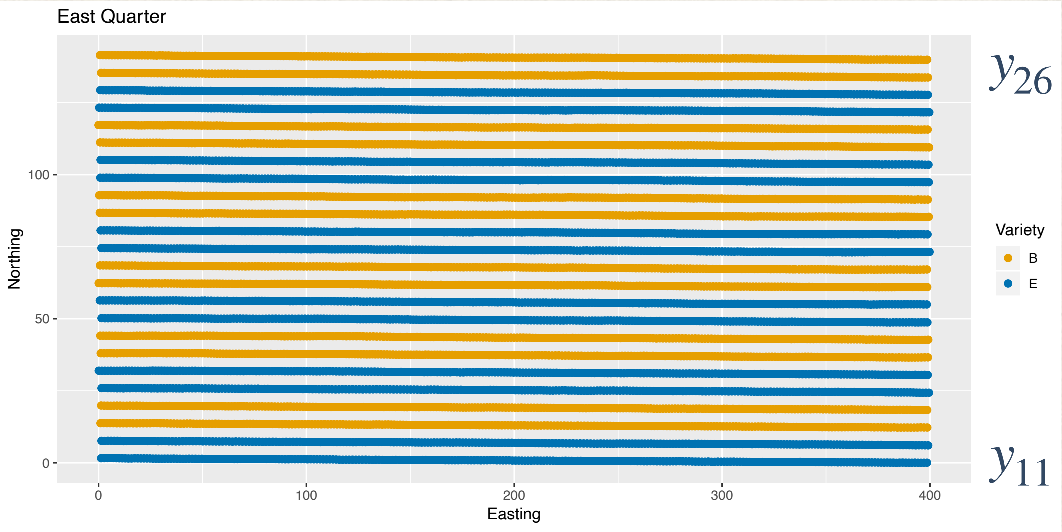
Strips as experimental units

Average yield over the length of the strip.

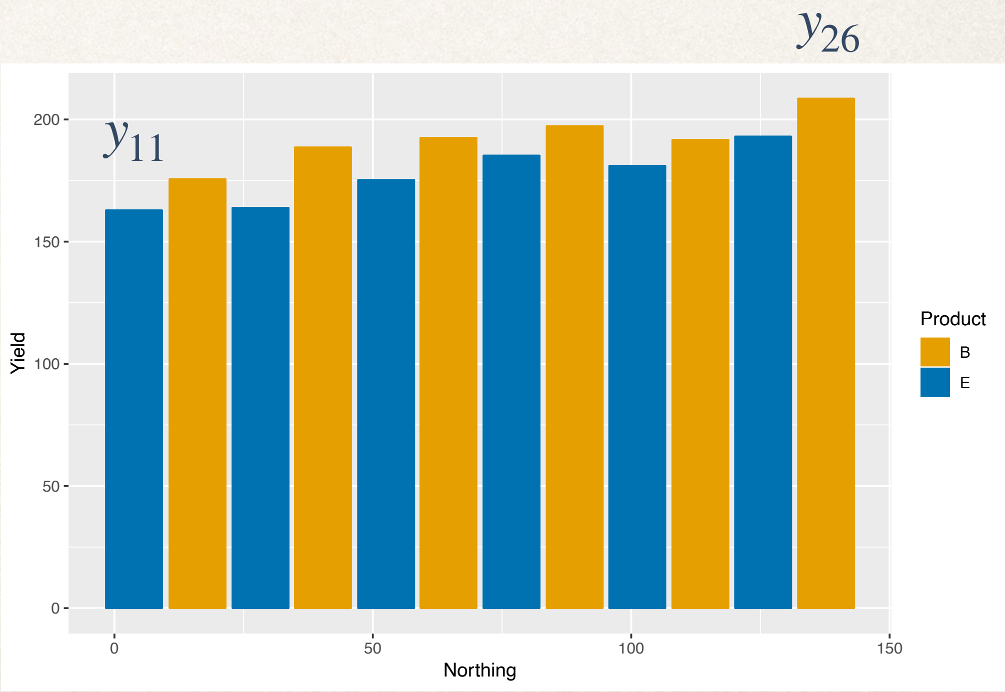
Strips as single experimental units

Calculate plot estimate as if yield data points are sub-samples

- Avoid pseudo-replication



$$y_{ij} = \bar{y}_{ij.} = \frac{\sum_{k=1}^{K} y_{ijk}}{K}$$



Statistical Model

- This implies the simple means model:

$$y_{ij} = \mu_i + e_{ij}$$

- with estimates given by

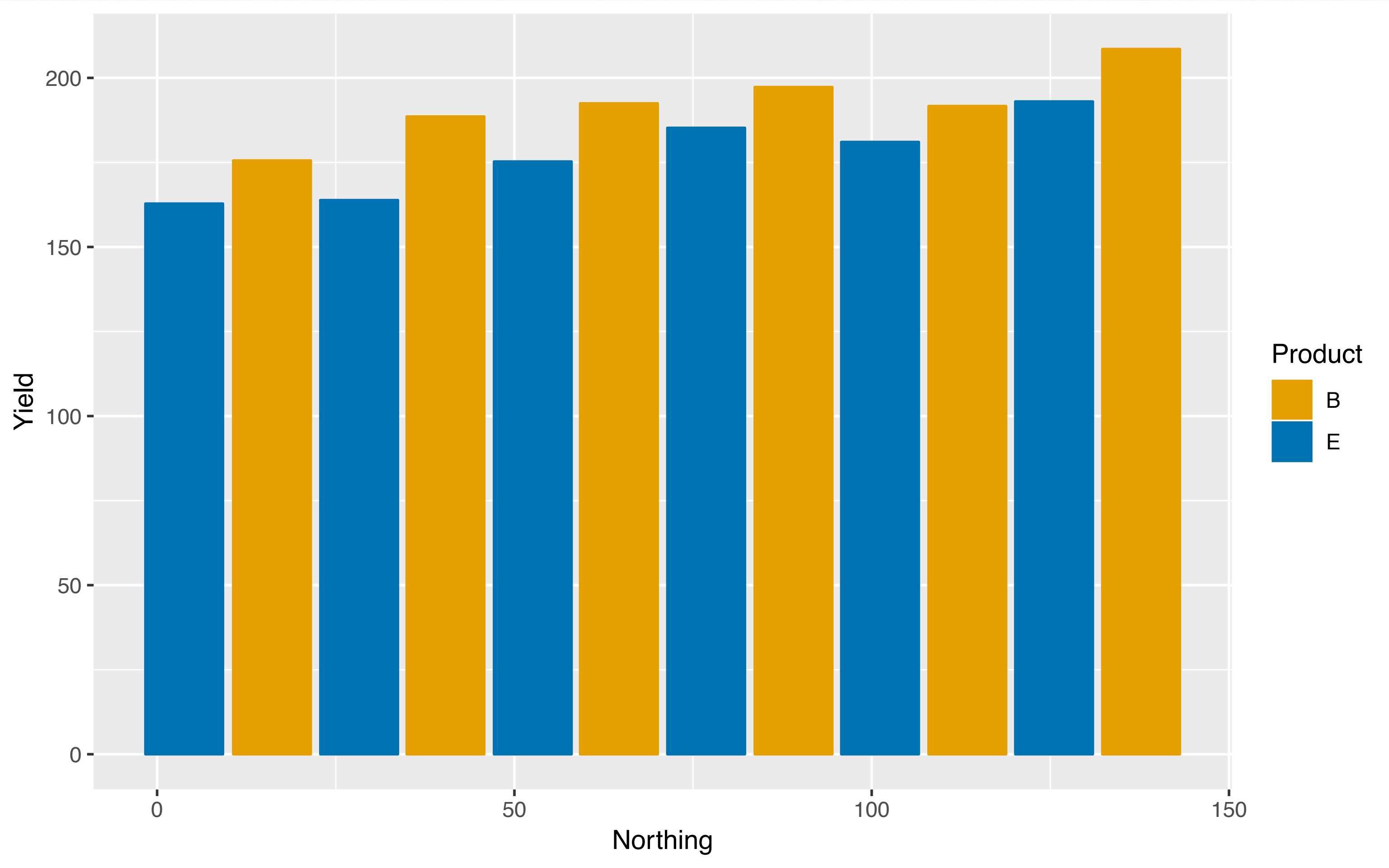
$$\widehat{\mu}_i = \frac{\sum_{j=1}^{N_i} y_{ij}}{N_i}, \widehat{\sigma}^2 = \frac{\sum_{j=1}^{N} (y_{ij} - \widehat{\mu}_i)^2}{N - k}$$

t-test

- We test the difference between means with

$$t = \frac{\delta}{\sqrt{2\sigma^2/n}} \qquad \delta = \hat{\mu}_1 - \hat{\mu}_2$$

- so…



$$\widehat{\mu}_1 = 192.35$$

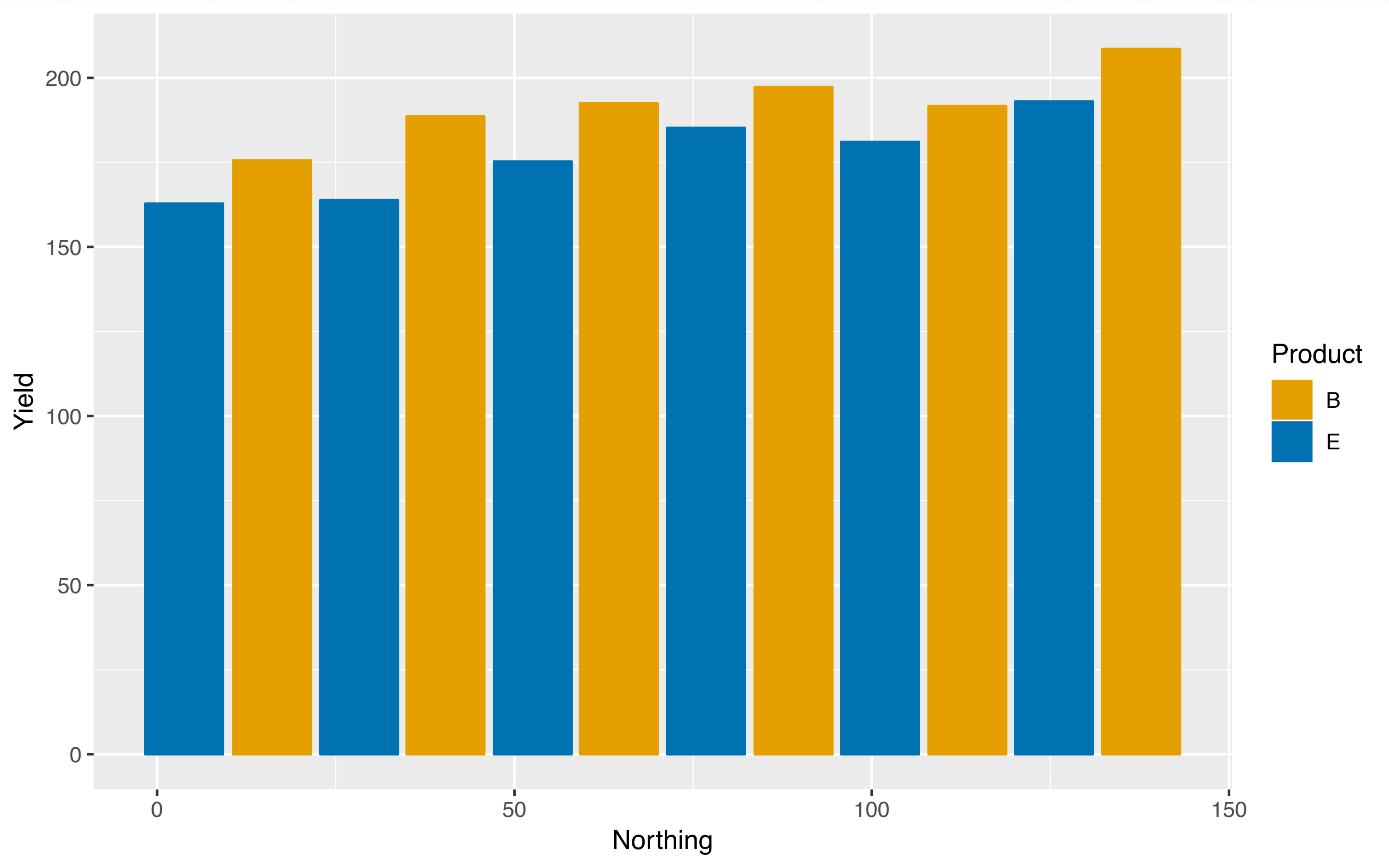
$$\widehat{\mu}_2 = 176.88$$

$$\delta = 15.48$$

$$\widehat{\sigma}^2 = 11.38$$

$$t = 2.35$$

$$p(>t) = 0.04$$



$\hat{\mu}_1 = 192.35$

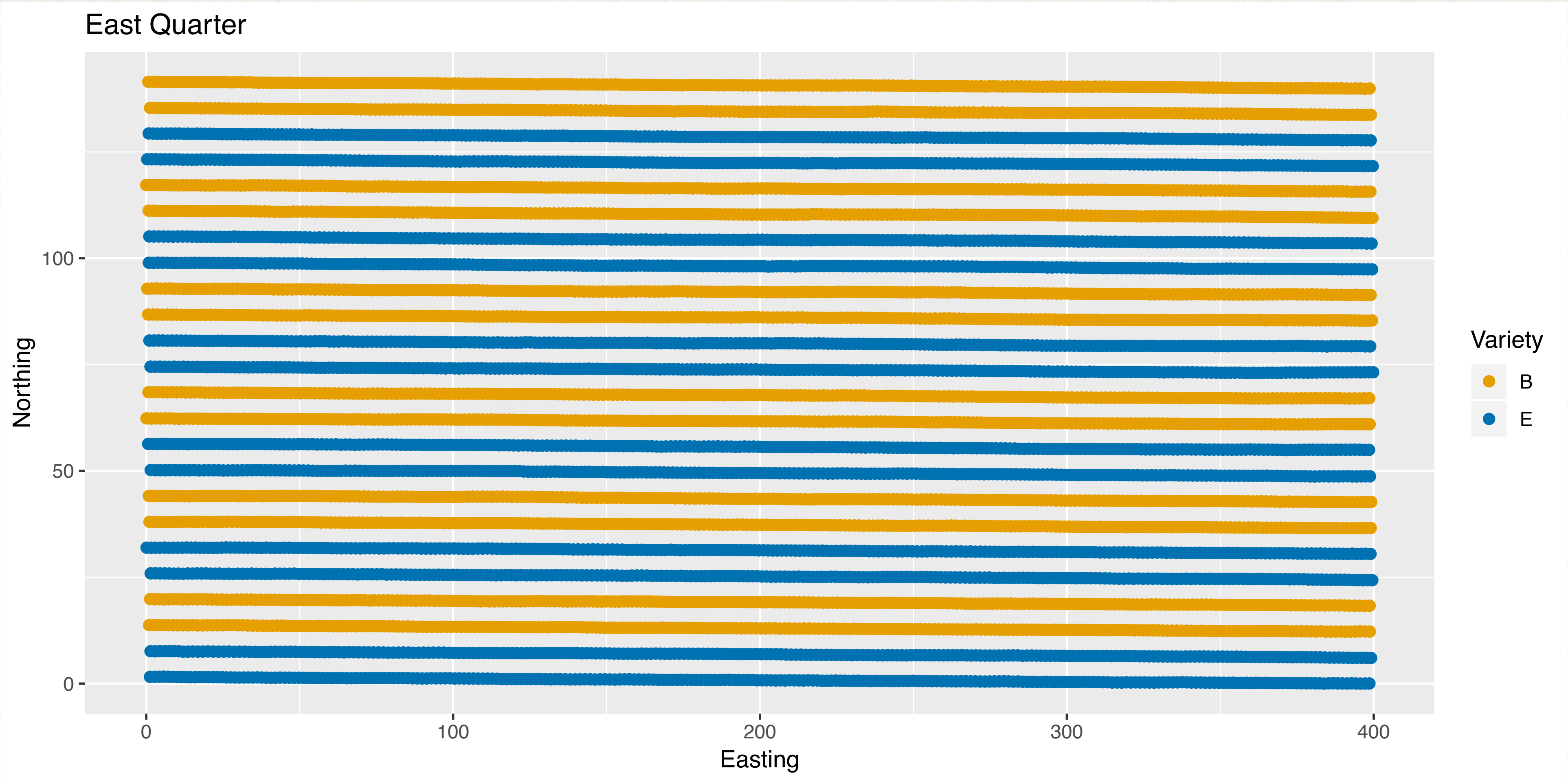
$\hat{\mu}_2 = 176.88$

$\rightarrow$ $\delta = 15.48$

$\hat{\sigma}^2 = 11.38$

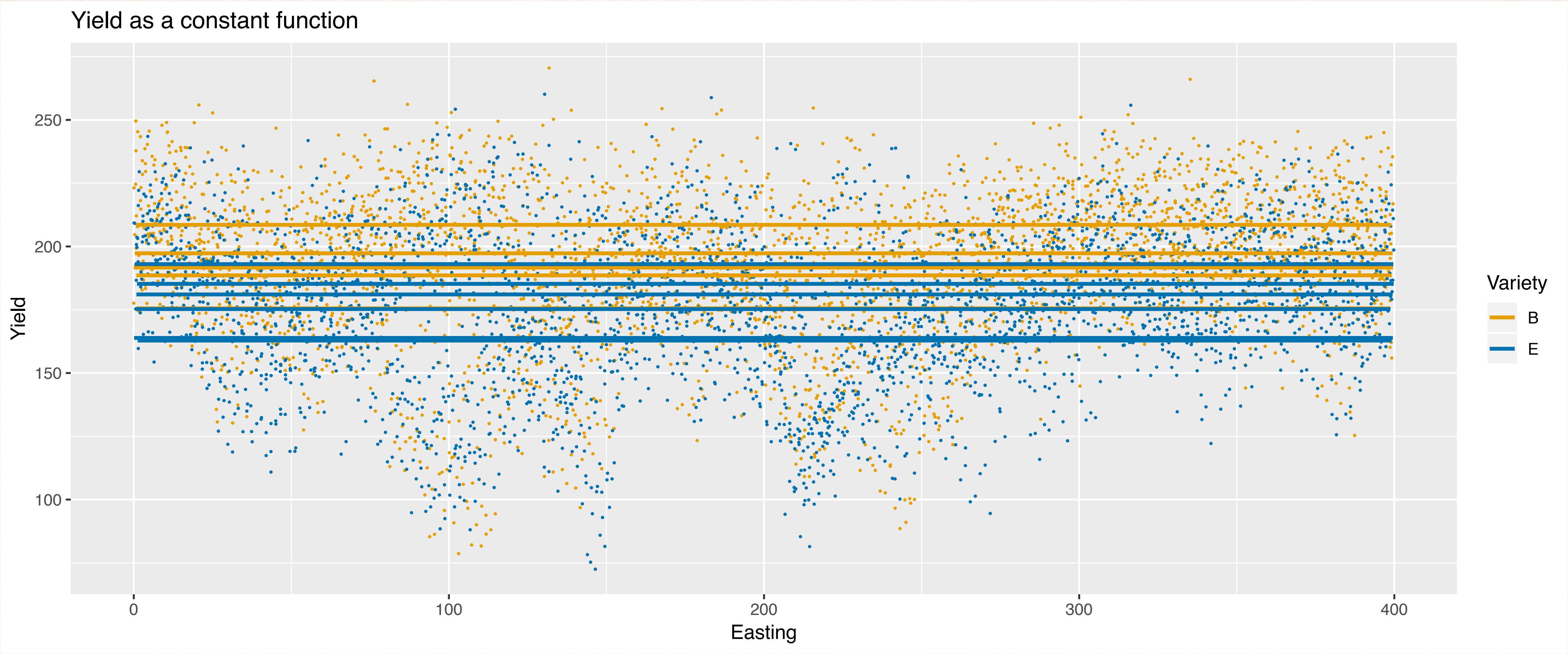
$\rightarrow$ $t = 2.35$ $\rightarrow$ $p(>t) = 0.04$

For this, I planted quarter mile strips?



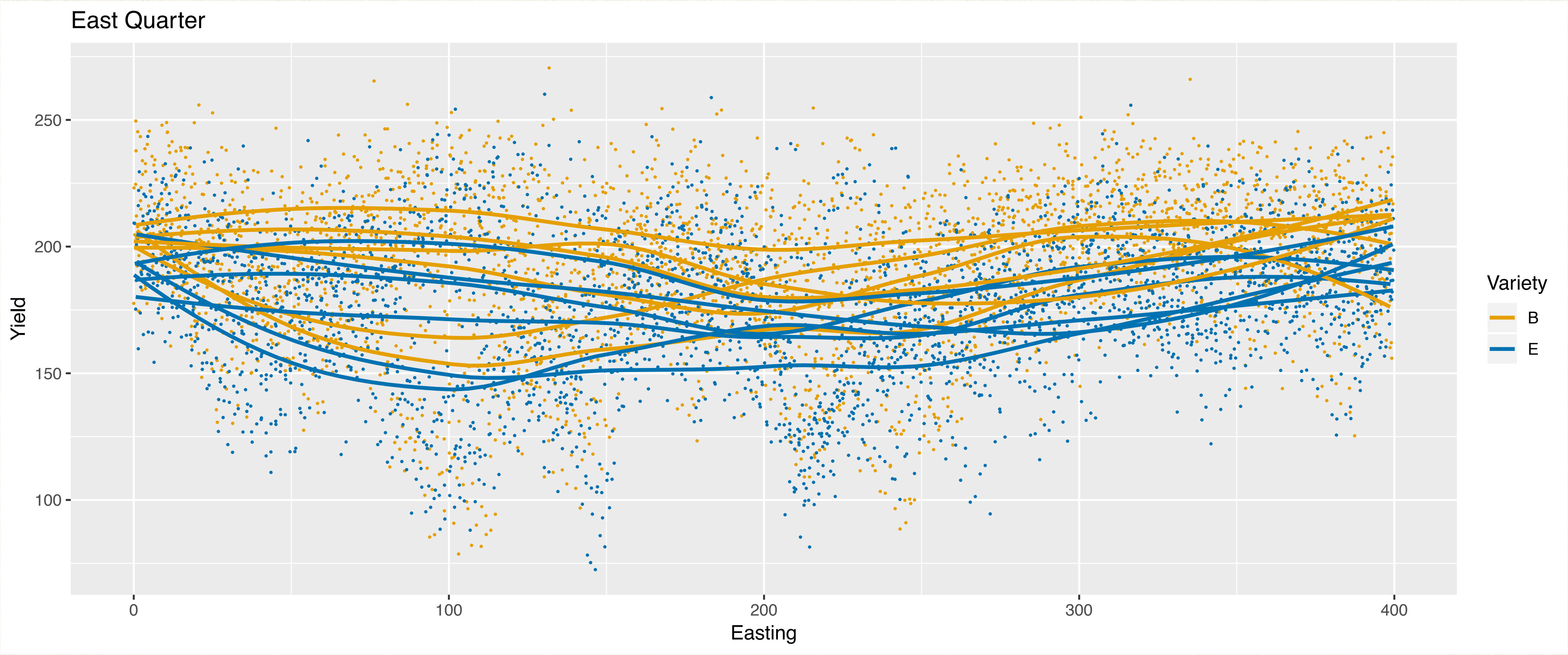
Strips as experimental units

Model strip as a function over position



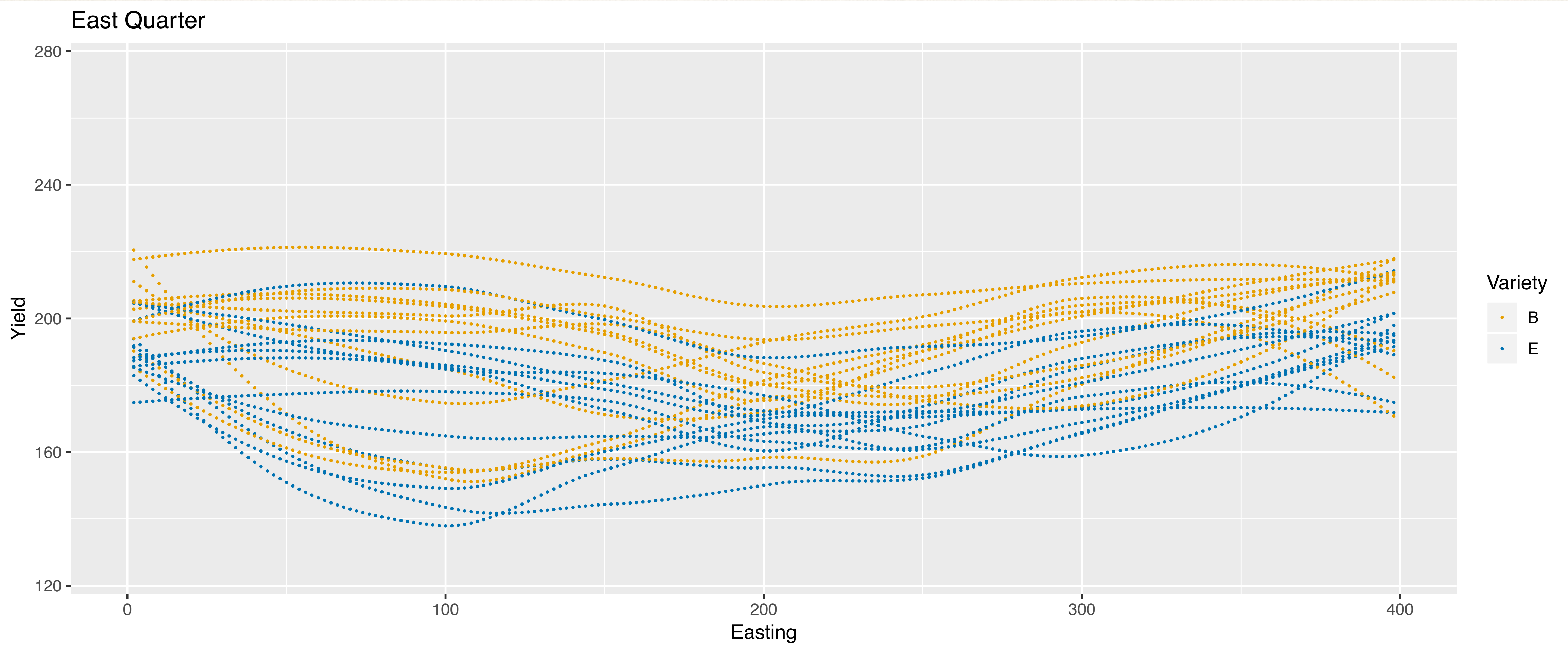
Yield Estimate

A single point estimate for mean implies that yield is constant along the length of the strip



Yield Function

A more realistic model allows yield to vary along the length of the strip



Functional analysis

Model each strip as a function of position, west to east.

Functional t-test

- Given a function for yield at distance *d,*

$$y_{ij}(d) = \text{Yield at Easting } d$$

- we can write the remaining statistics as functions:

$$y(d)_{ij} = \mu(d)_i + e(d)_{ij}$$

$$\hat{\mu}_i(d) = \frac{\sum_{j=1}^{N_i} y_{ij}(d)}{N_i}, \hat{\sigma}^2(d) = \frac{\sum_{j=1}^{N_i} (y_{ij}(d) - \hat{\mu}_i(d))^2}{N - k}$$

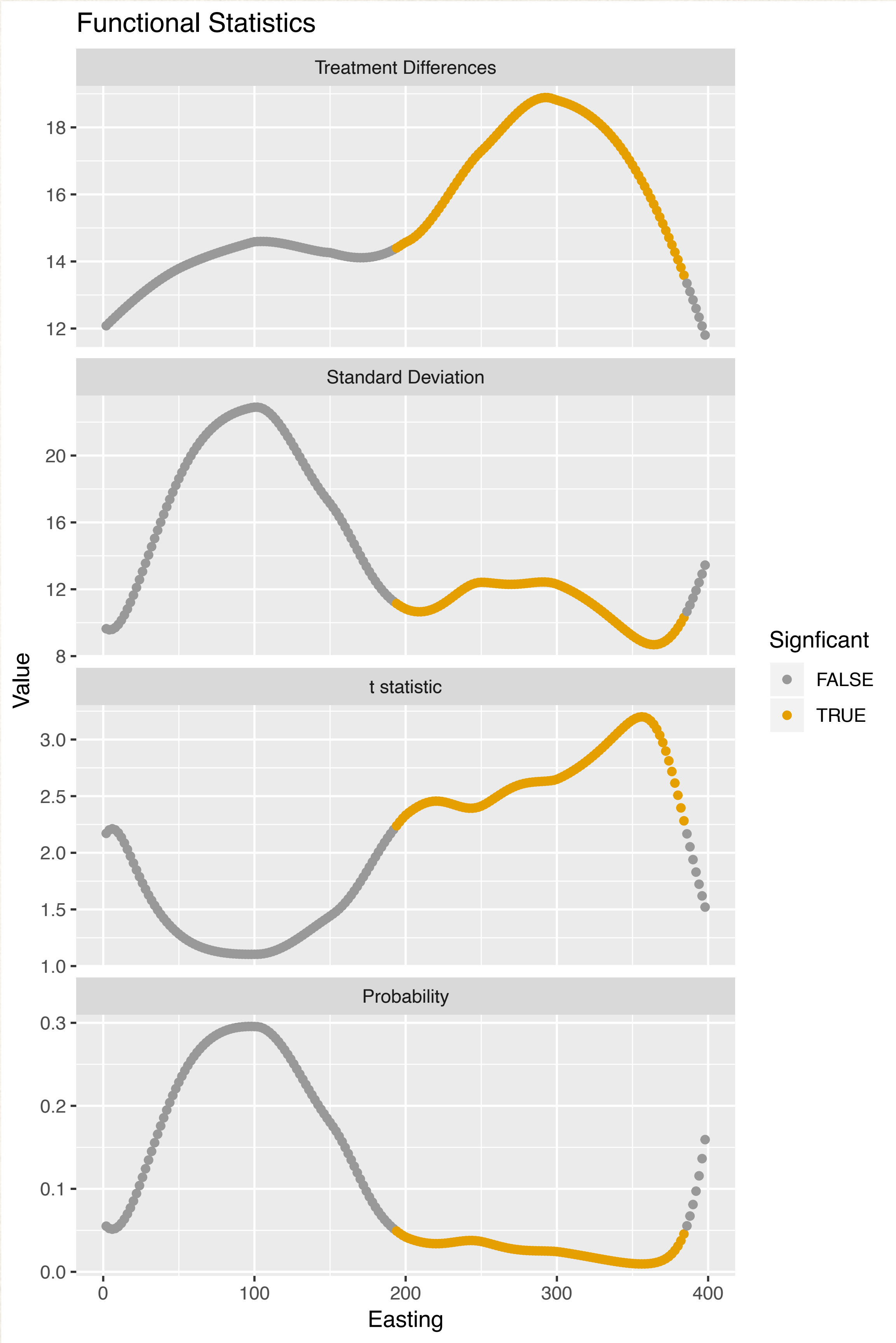
$$\delta(d) = \hat{\mu}_1(d) - \hat{\mu}_2(d) \qquad t(d) = \frac{\delta(d)}{\sqrt{2\sigma^2(d)/n}}$$

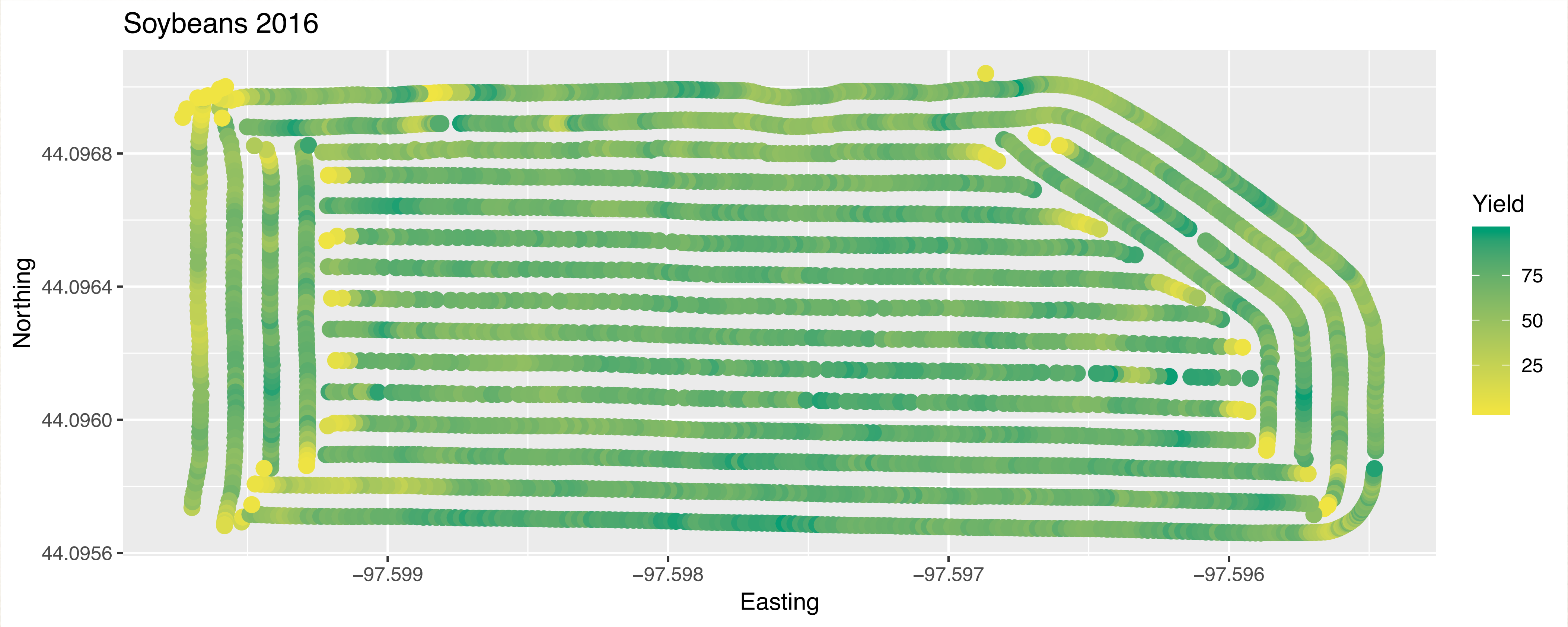
Functional t-test

Moving west to east,

- difference increases
- standard deviation decreases
- t ratio increases

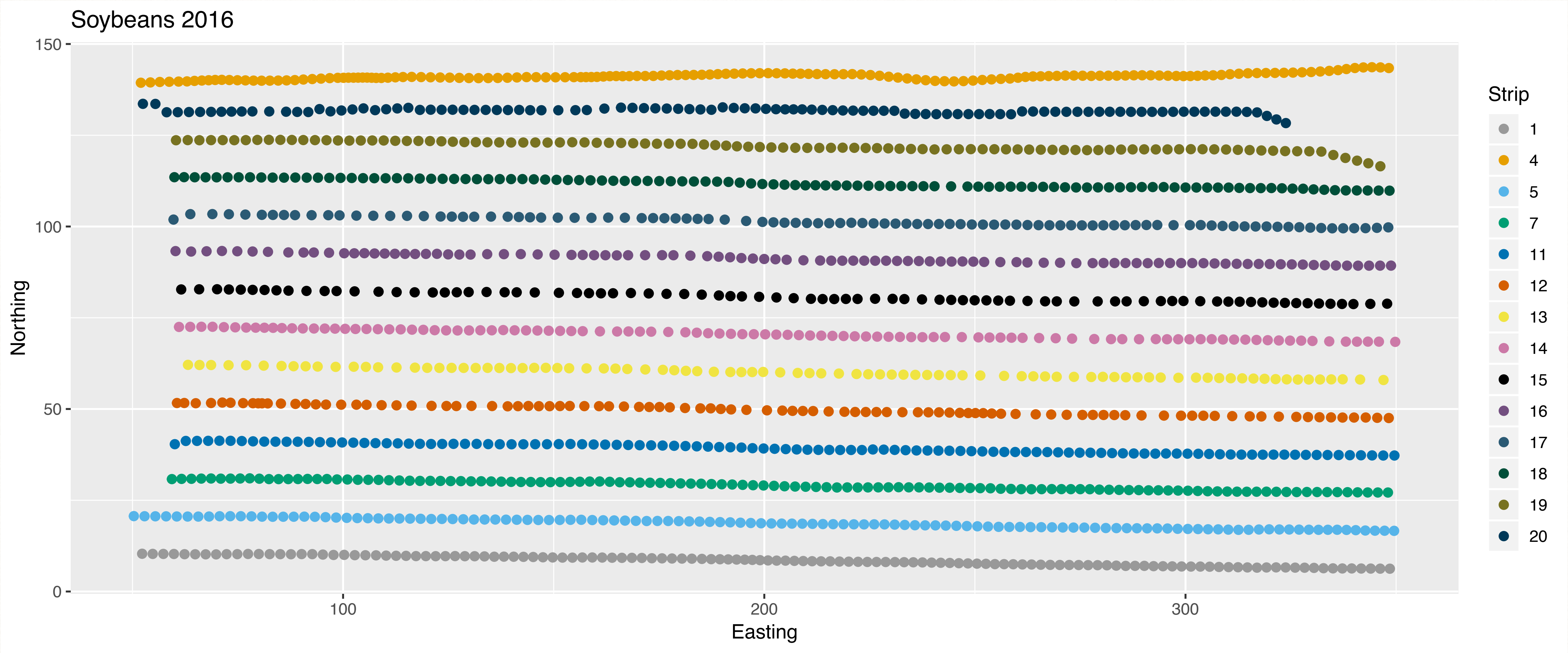
We can be more confident in variety differences in the eastern part of the field.





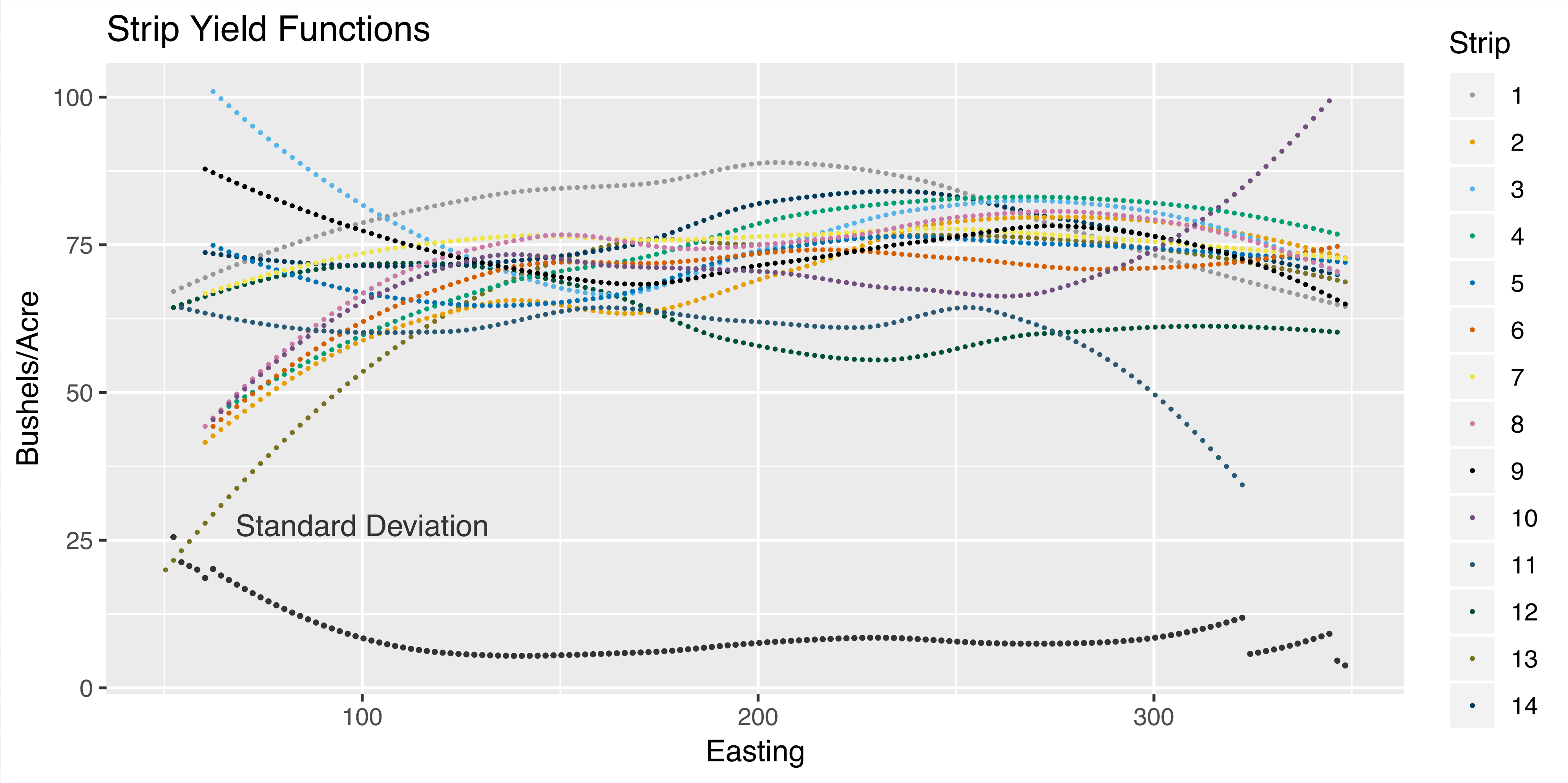
Power analysis

Suppose we have yield monitor data from a field that can be entered into strip trials.



Potential experimental units

How many strips would be available in this field? How many should we treat?



Convert strips to functions

Calculate a standard deviation for the strips across the length of the field.

Power calculations

- Given a t-test of the form

$$t(d) = \frac{\delta(d)}{\sqrt{2\sigma^2(d)/n(d)}}$$

- Required replicates

$$n(d) \geq \left(\frac{2\sigma^2(d)}{\delta(d)}\right)^2 (t_{\alpha/2} - t_\beta)^2$$

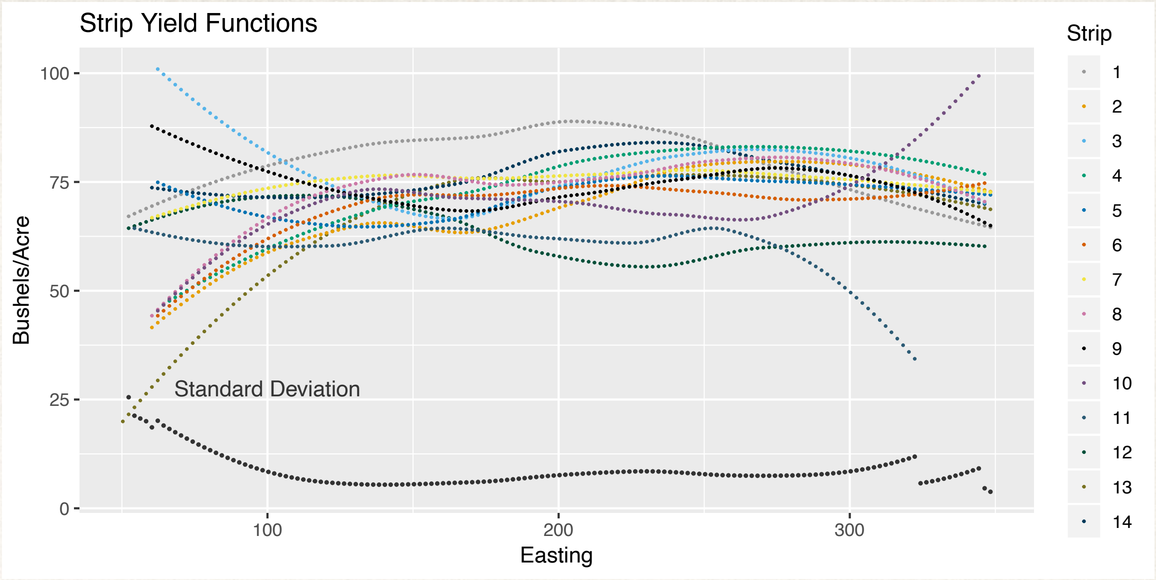
- Detectable difference

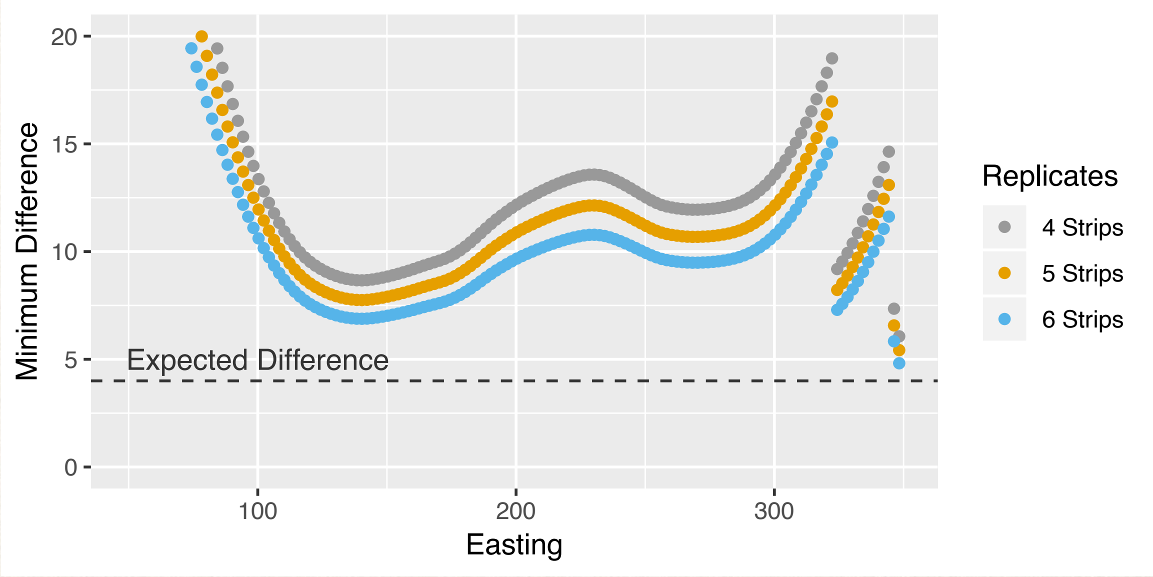
$$\delta(d) = \frac{2\sigma^2(d)(t_{\alpha/2} - t_\beta)}{\sqrt{n(d)}}$$

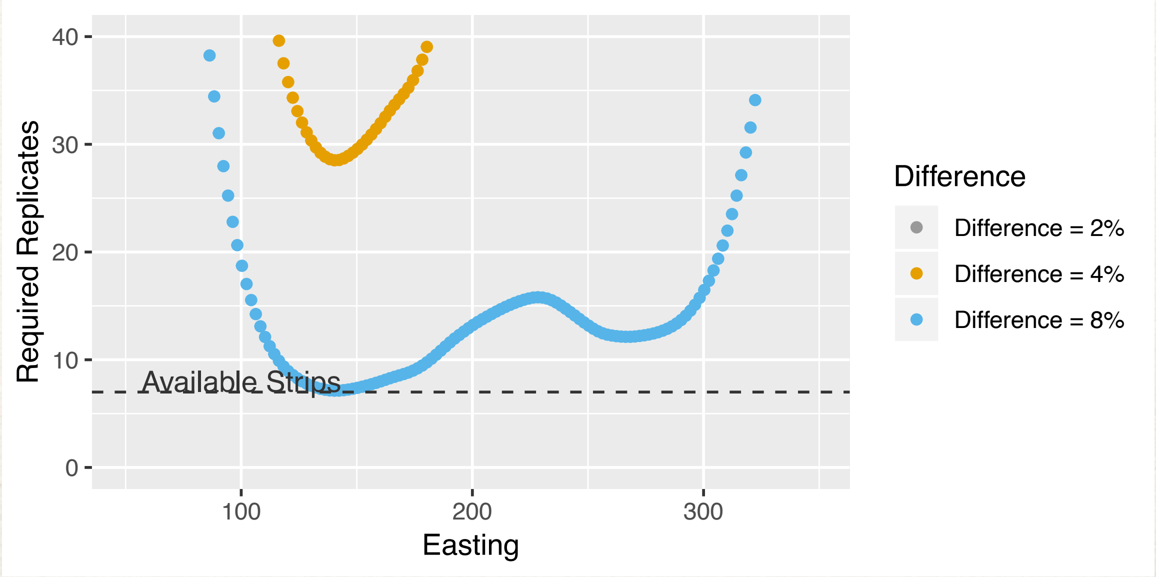
Power Curves

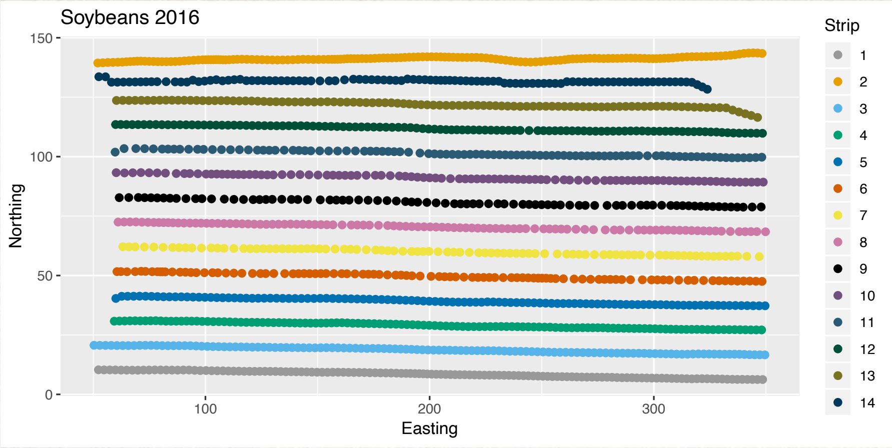
Given a set of potential strips, what differences can we expect to detect?

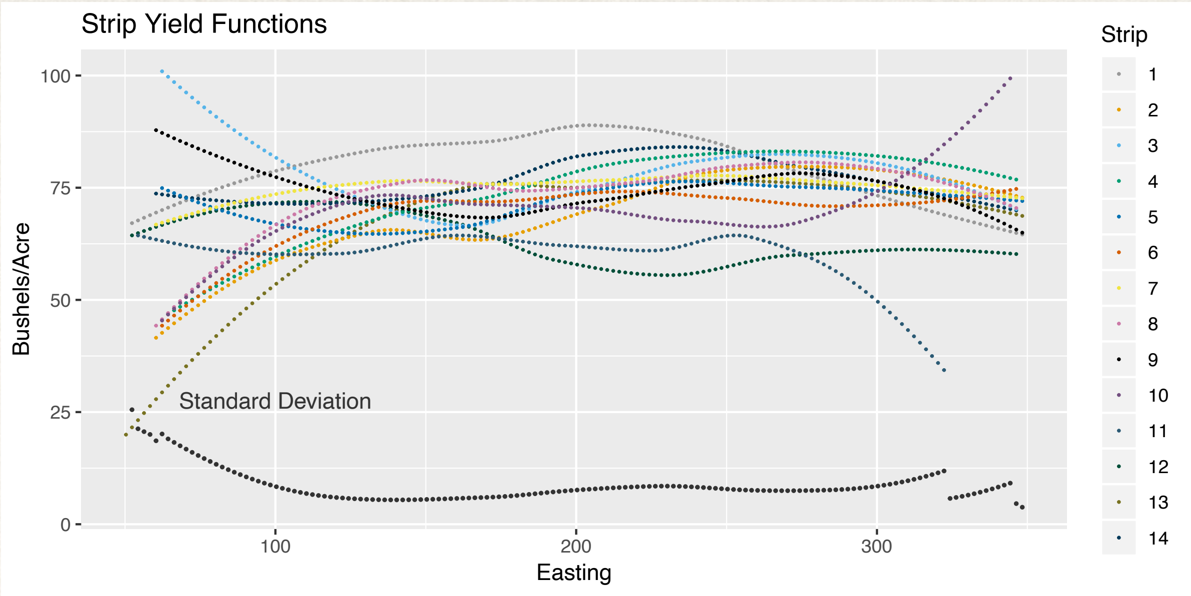
- Assume a seed treatment that may improve yield by 4 bu/acre
- Assume we can entered up to 12 strips into a trial

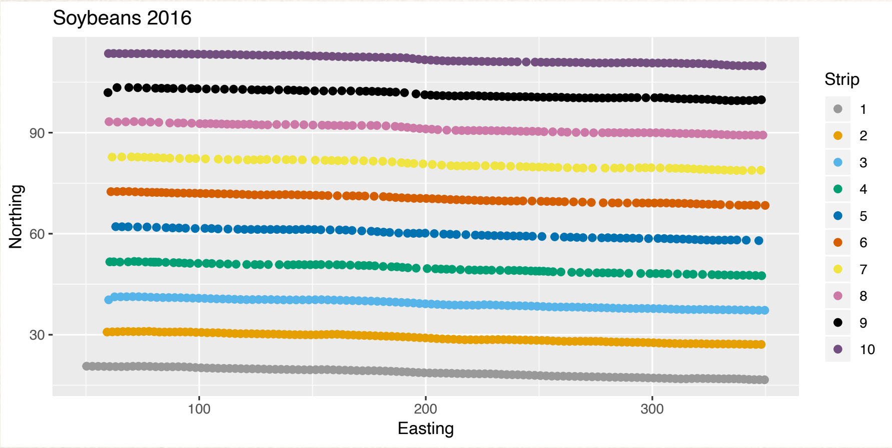


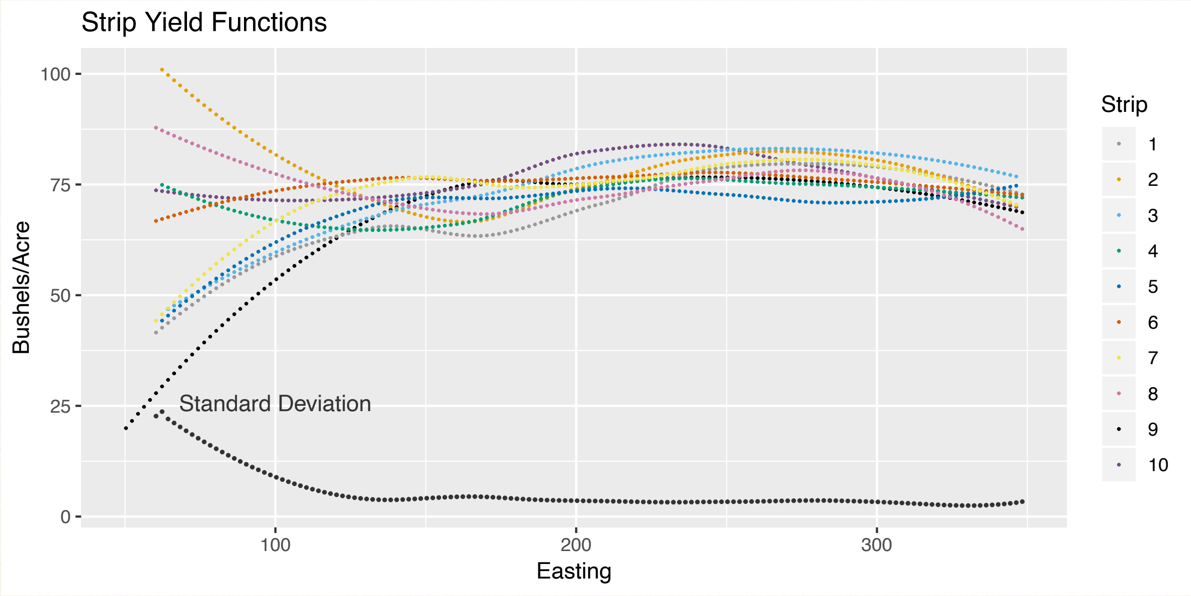












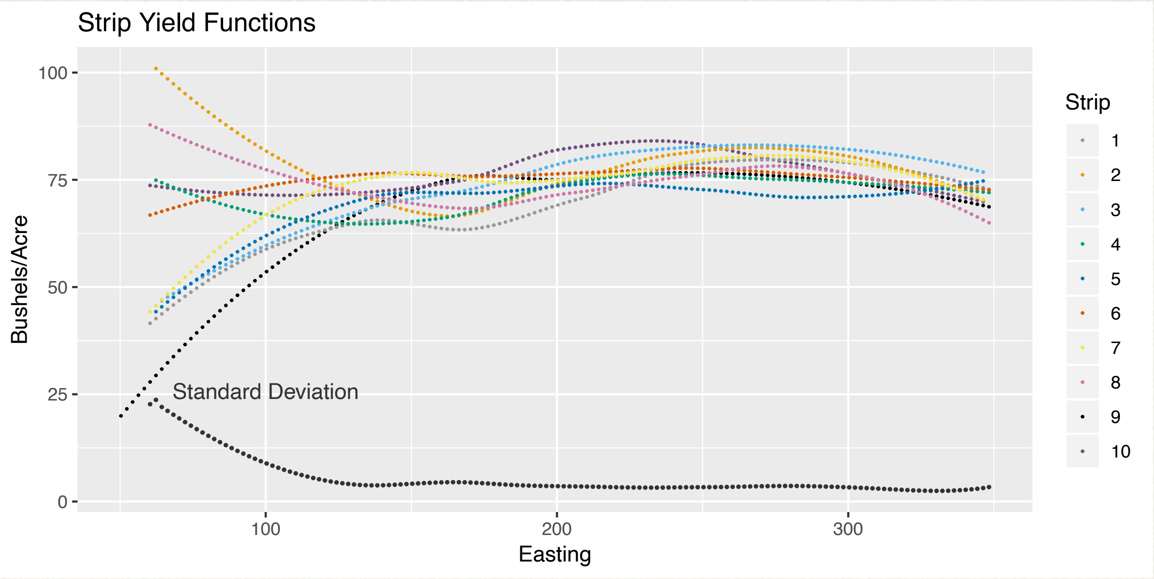
Remove Outliers

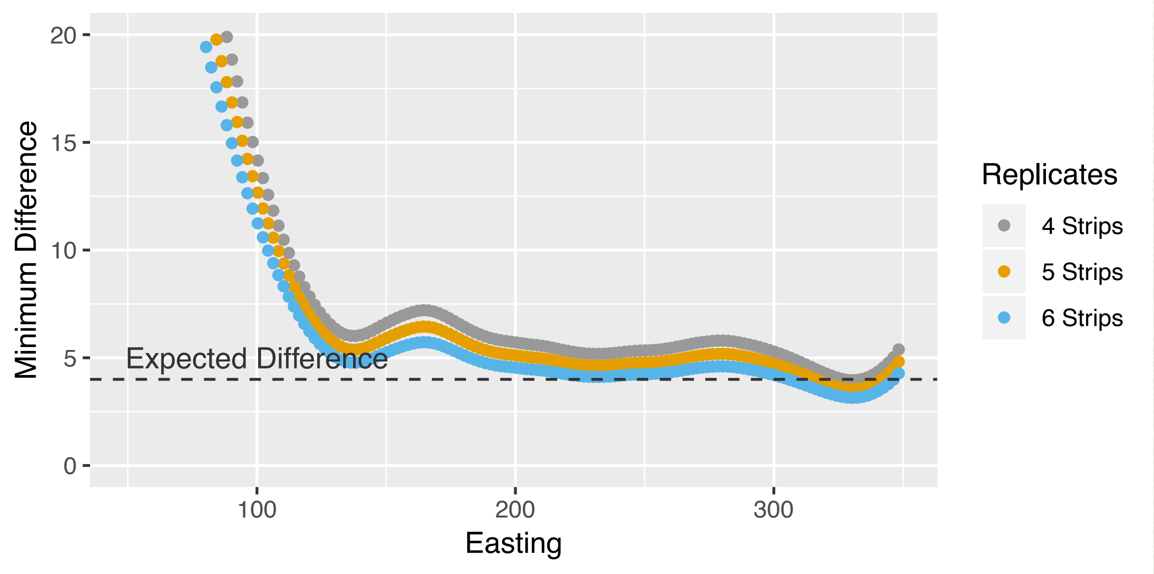
Strips in the center are more uniform than border strips

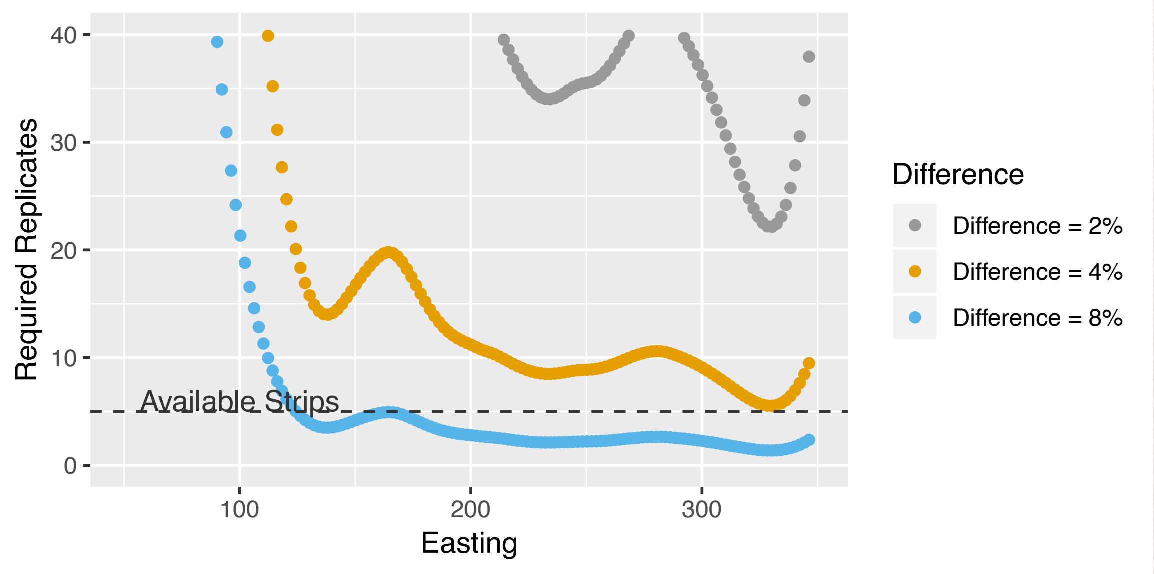
Power Functions

Removing border strips reduces standard deviation and increases power

For this field, fewer strips might provide more power

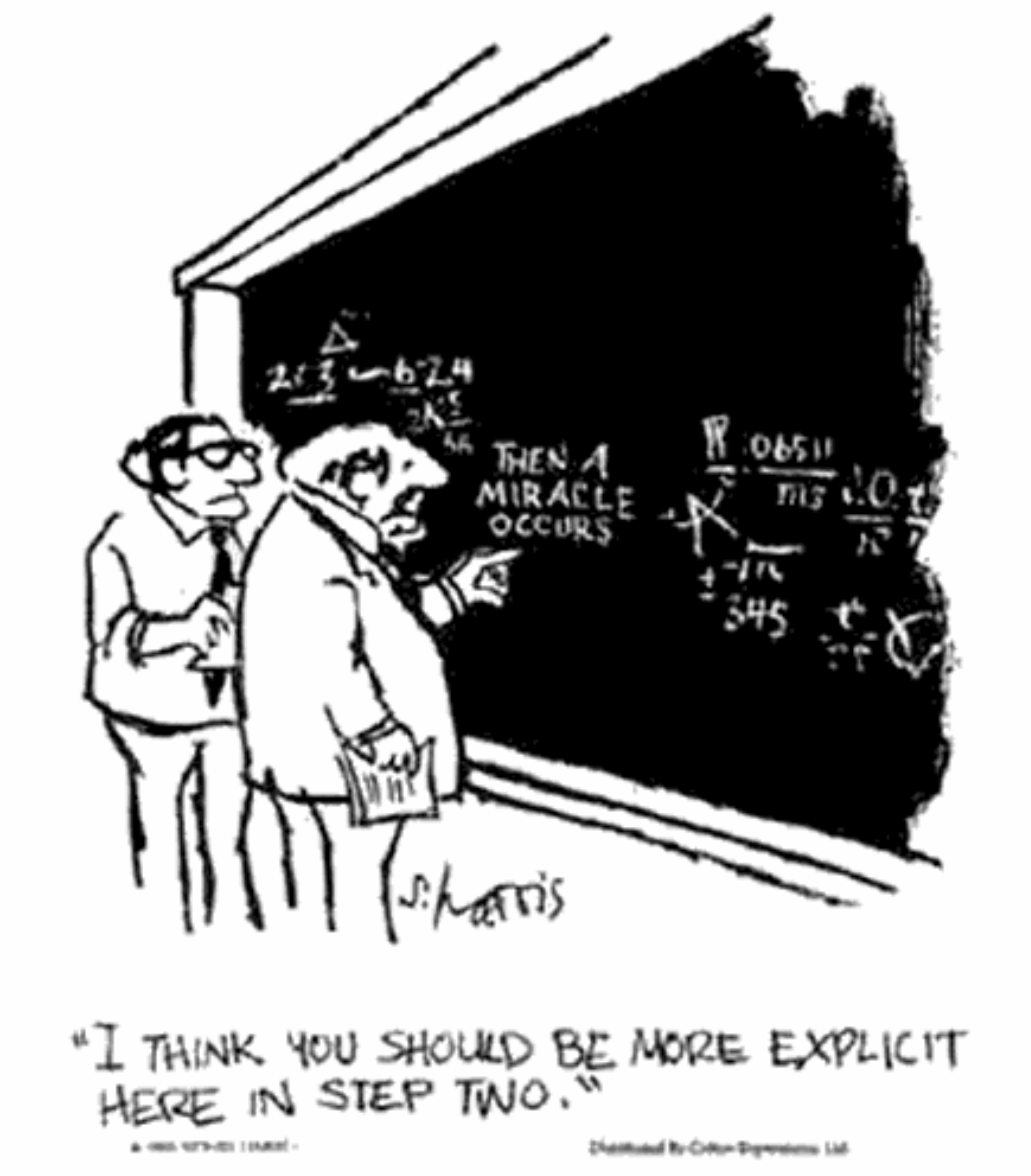




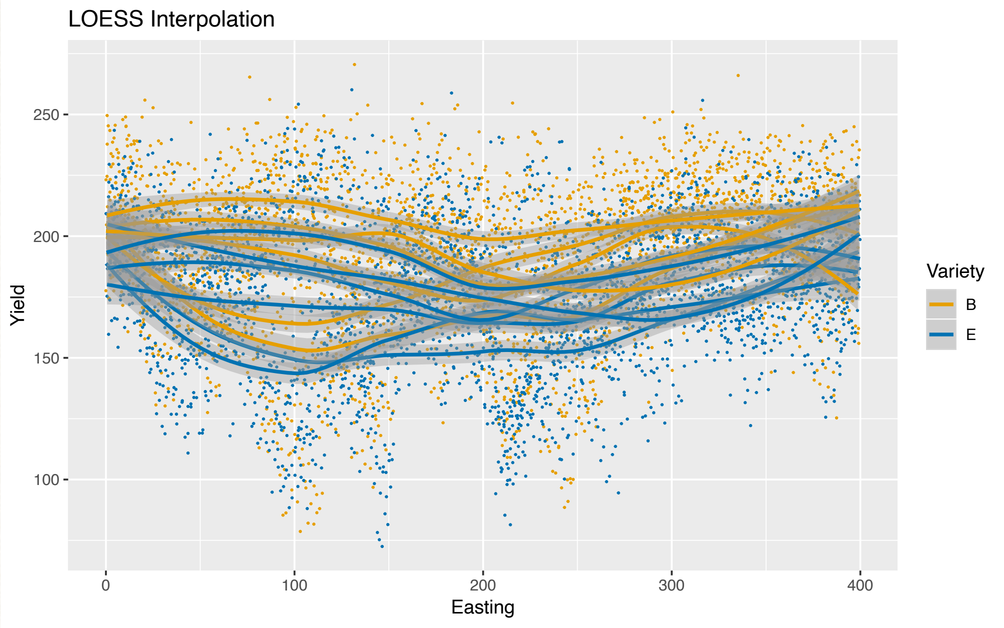


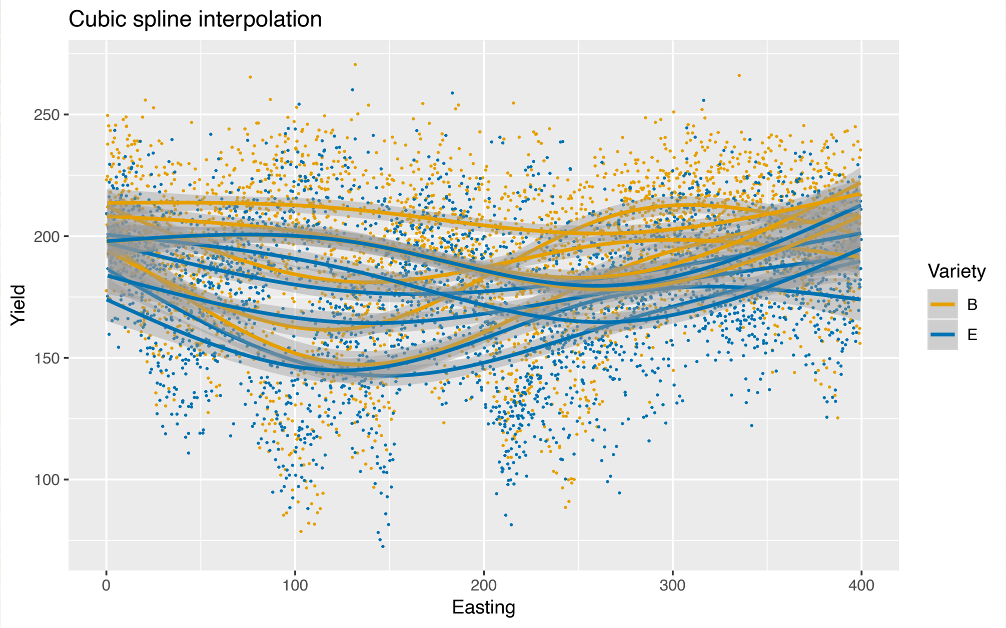
Yield Function

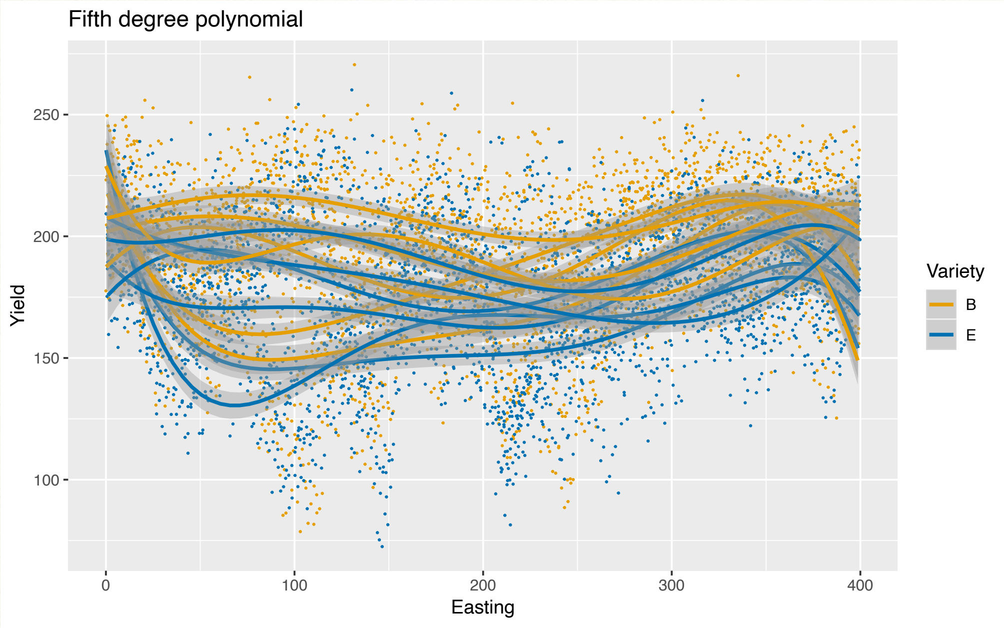
$y_{ij}(d)$ = Yield at Easting d

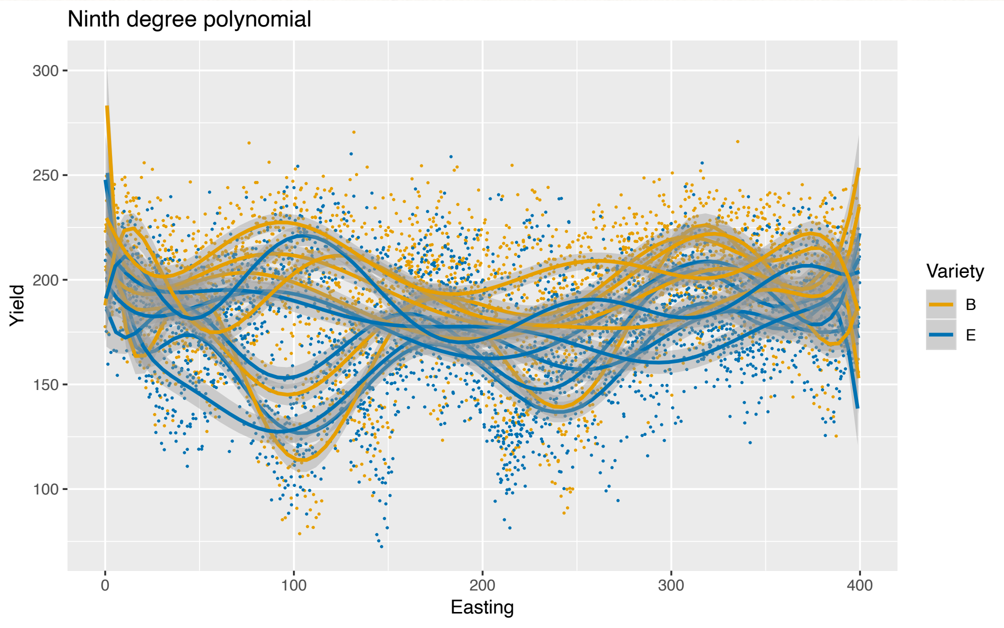


ScienceCartoonPlus.com







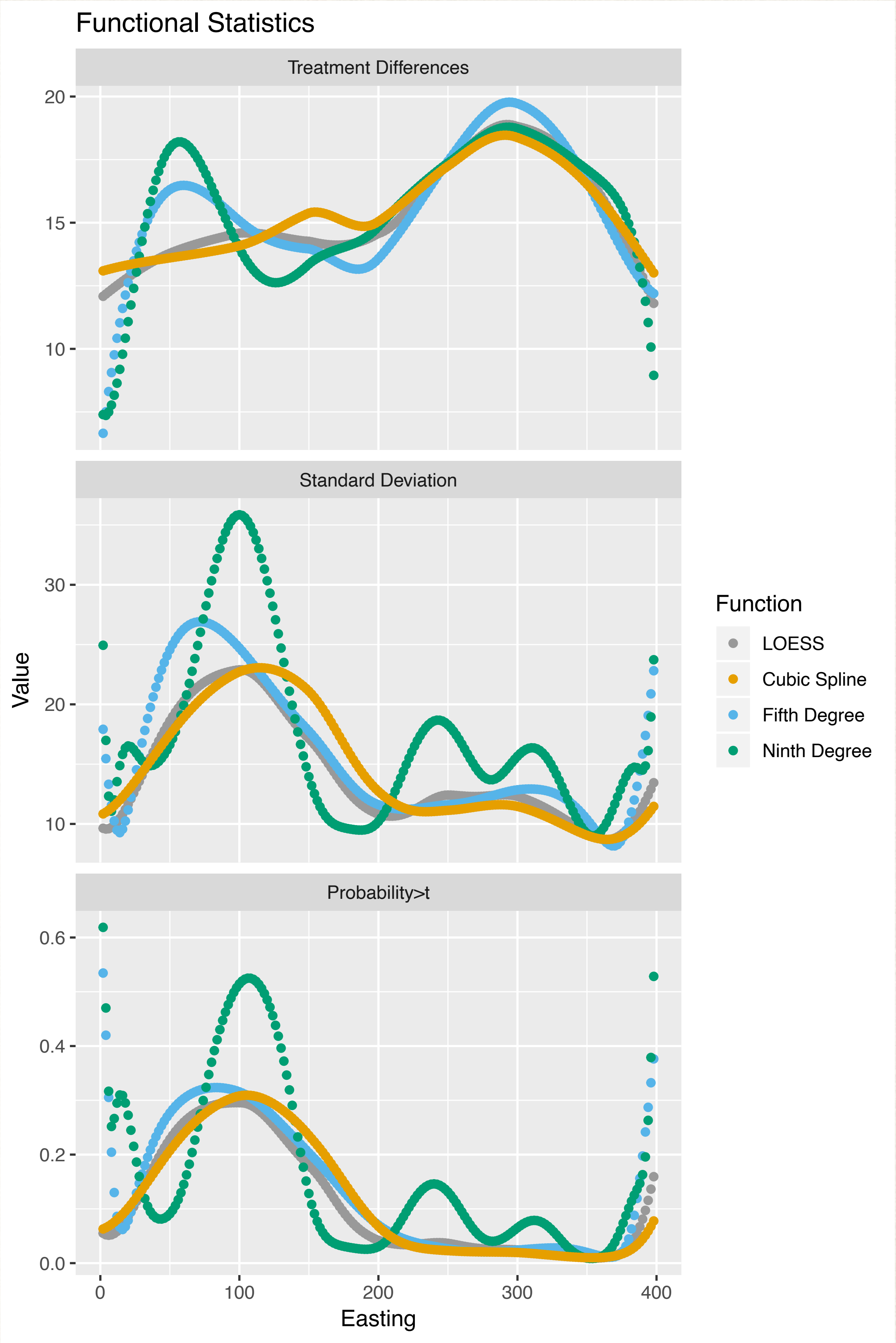


Alternative Yield Functions

Smoothing functions can be global functions, or local smoothing functions with different ranges

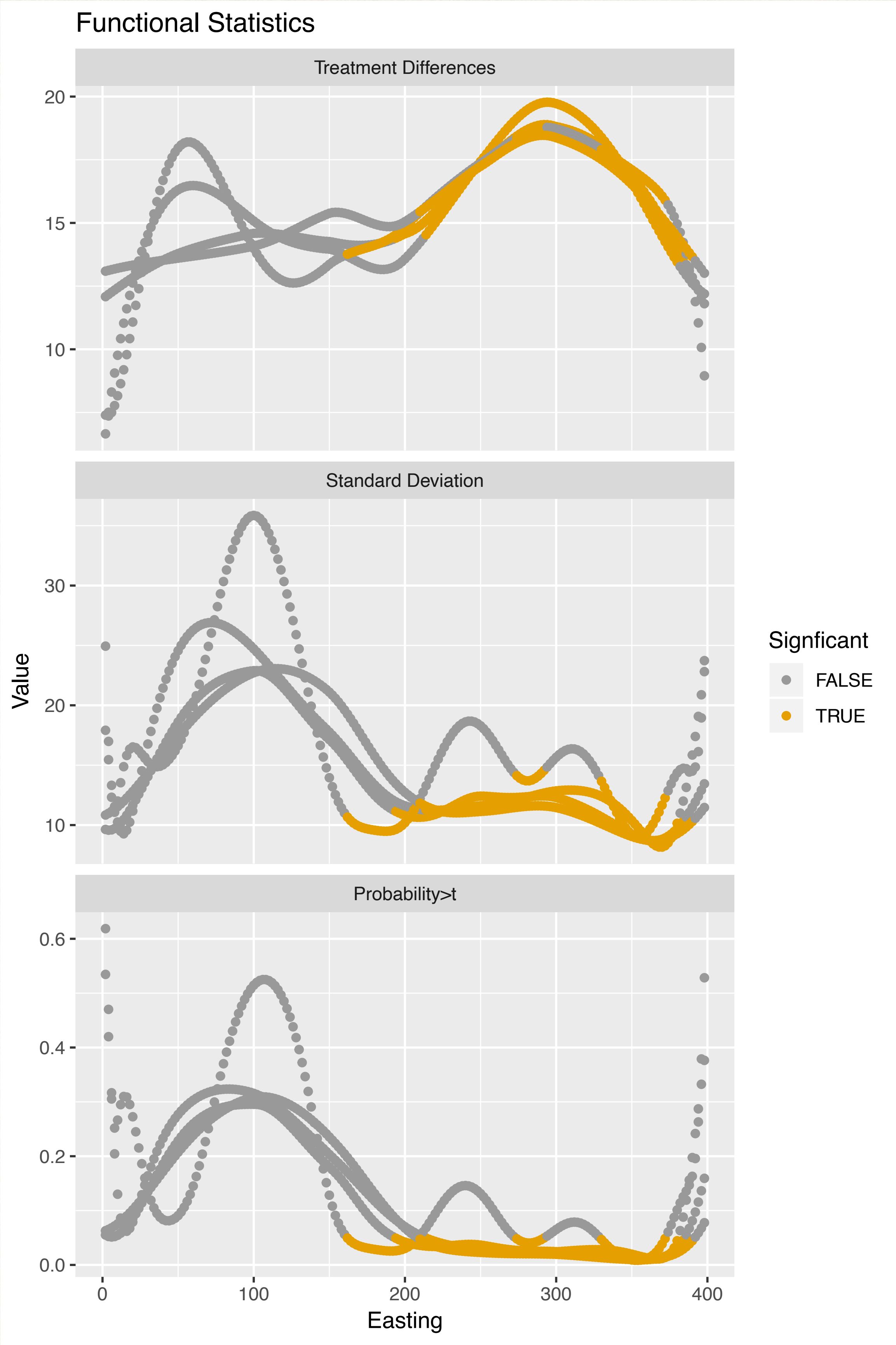
Alternative Functions

The choice of function can influence estimates of treatment differences and standard deviations in different parts of the field.



Alternative Functions

In this case, though, there is some general consensus with regards to which part of the field treatment differences are significant.



Conclusion

- Functional analysis provides an option for the analysis of strip trials that remains faithful to classically designed experiments, while taking advantage of the extra information available in strips that are larger than traditional small plots.

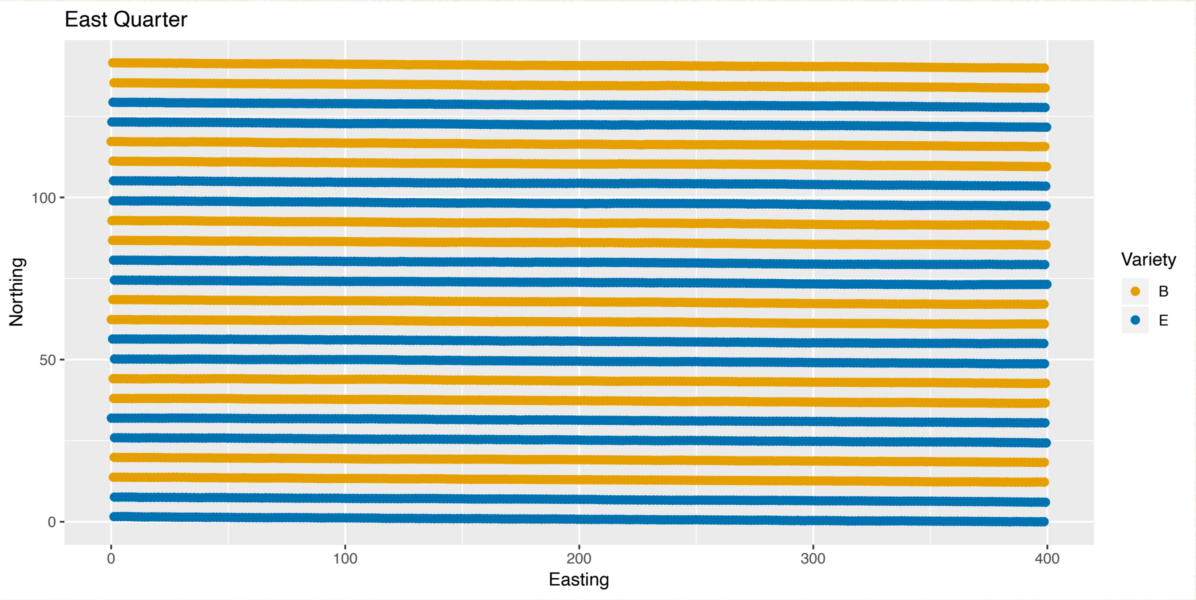
- Care must be taken in the selection of functions used to model strips.

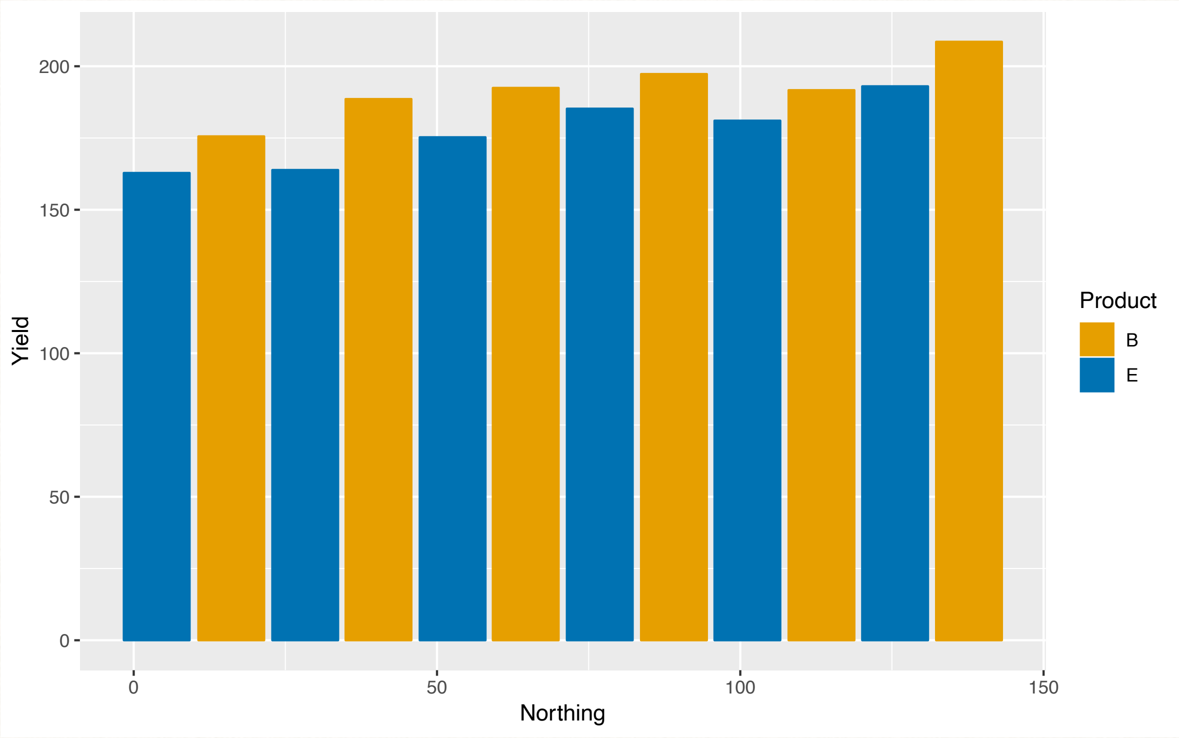
Additional Thoughts

Blocking and Randomization

It is not always practical to expect treatments to be applied to strips completely at random. This can lead to bias.

How should we model bias?

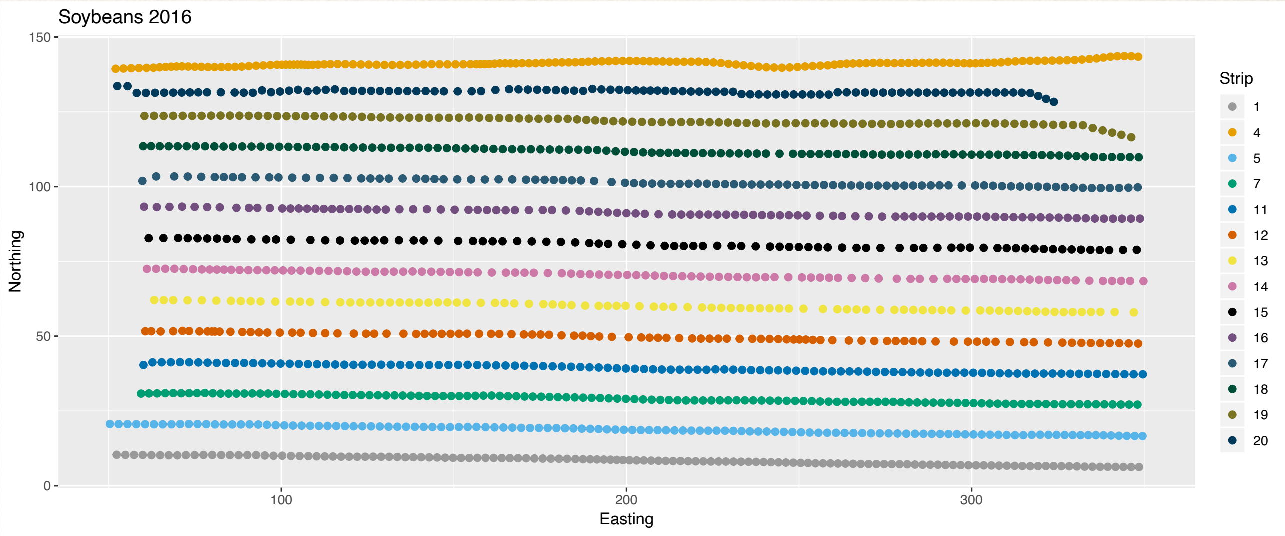


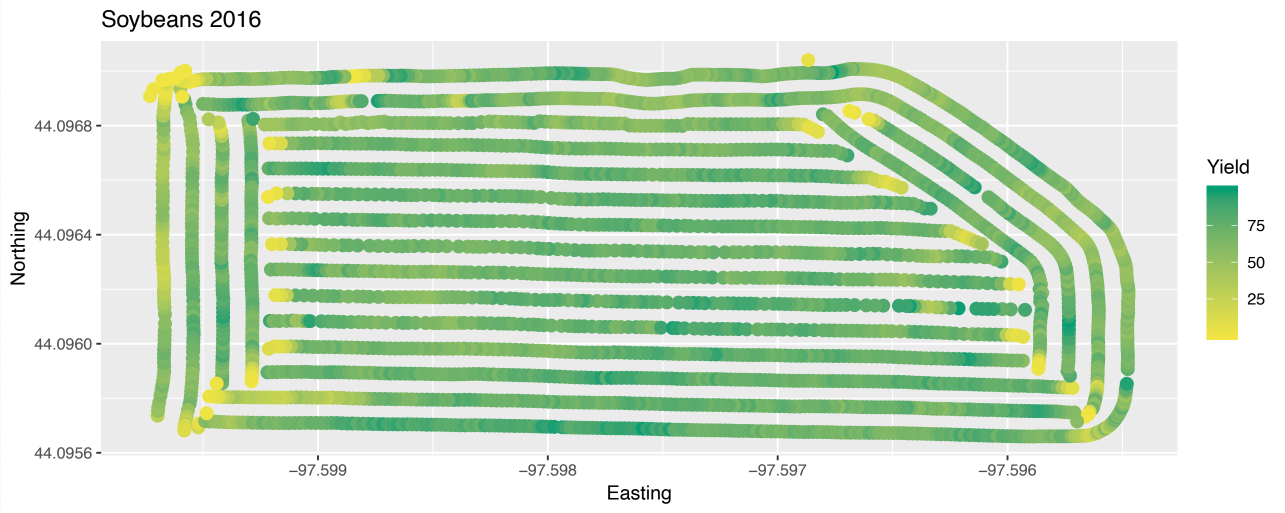


Correlated Errors

It is not always practical to include buffer strips between treated strips. There will be some degree of correlation between adjacent strips.

How should we model this correlation? Does it matter if strips are harvested in parallel or anti-parallel directions?

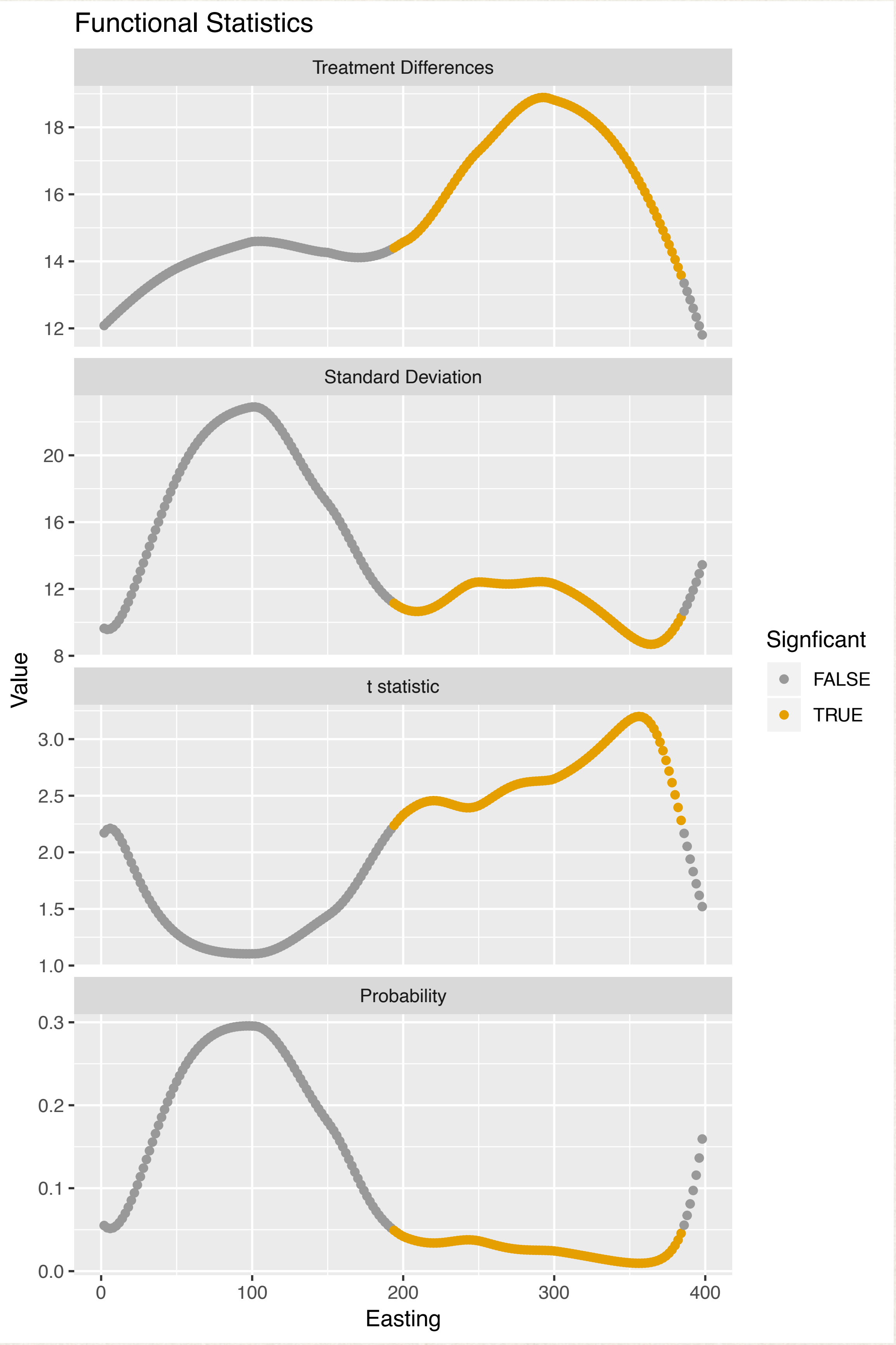




Multiple testing

What does the null-hypothesis test imply in a functional context?

Is there a family-wise error rate?



How do we extend simple strips to more advanced experimental designs?

- Blocking

$$y(d)_{ij} = \mu(d) + \tau_i(d) + b_j(d) + e(d)_{ij}$$

- Covariates

$$y(d)_{ij} = \mu(d) + \tau_i(d) + \beta x_{ij}(d) + e(d)_{ij}$$

- Factors

$$y(d)_{ij} = \mu(d) + \alpha_i(d) + \beta_j(d) + \alpha\beta_{ij}(d) + e(d)_{ij}$$