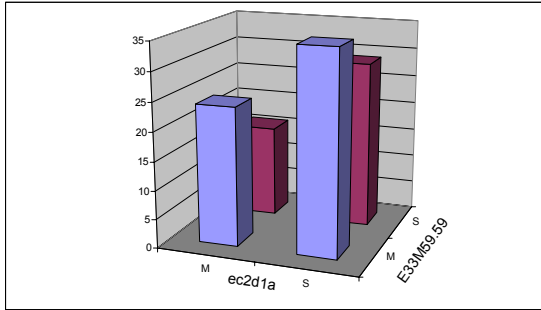


counts: recombination

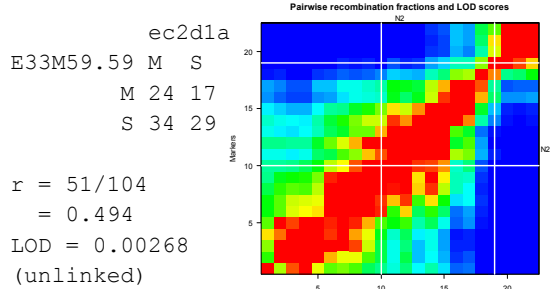


sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

7

counts: recombination

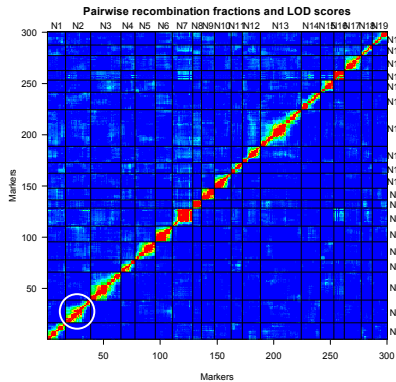


sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

8

counts: recomb & LOD scores

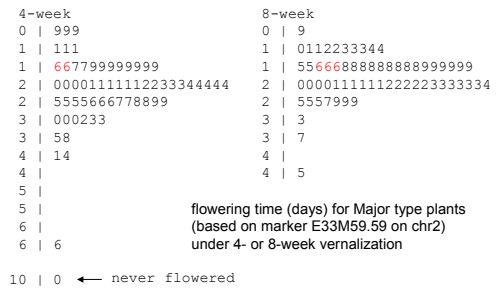


sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

9

flowering time: stem-leaf plots (Ferreira et al. 1995)

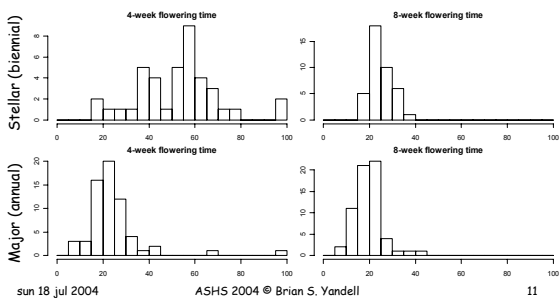


sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

10

histogram summaries

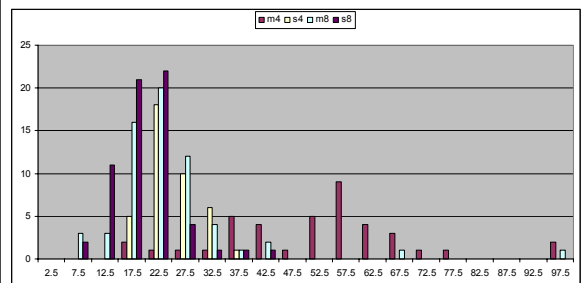


sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

11

side-by-side histograms (colors, background, grid lines, group order)

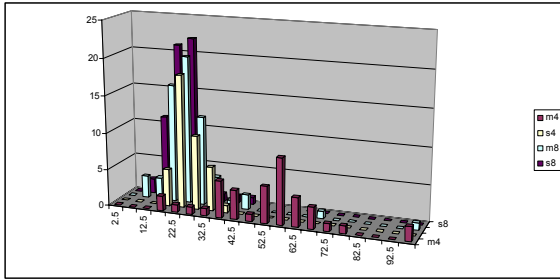


sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

12

3-D histograms (group order, grid lines, view angle)

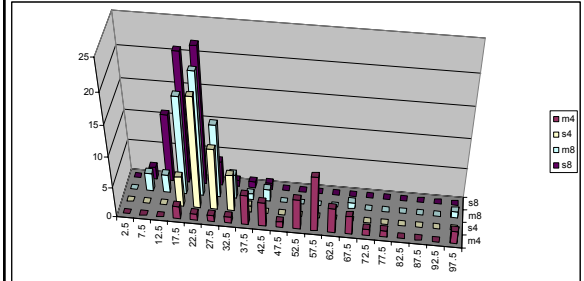


sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

13

3-D histograms (another view angle)

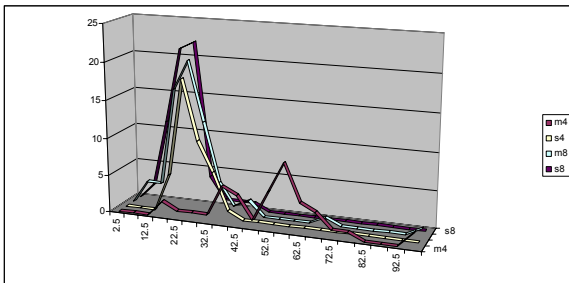


sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

14

3-D ribbon plot (group order, view angle)

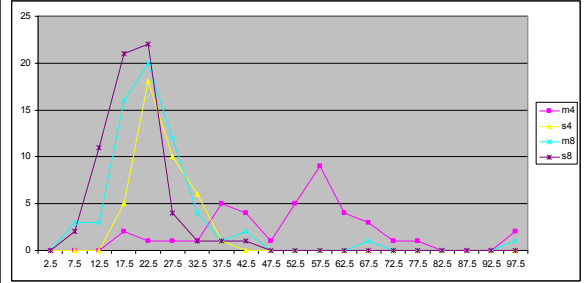


sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

15

polygon histograms (colors, background, grid lines)

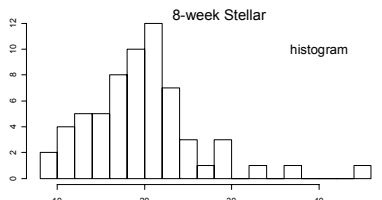


sun 18 jul 2004

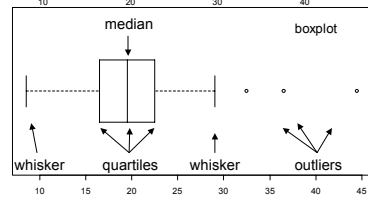
ASHS 2004 © Brian S. Yandell

16

histogram shows most of the data and broad shape, but takes up a lot of space



boxplot shows key features (summaries) of shape, hiding most of the data

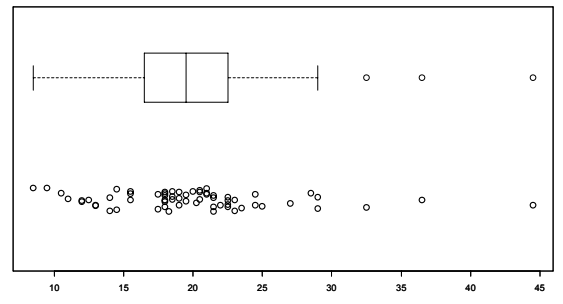


sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

17

boxplot vs. all data

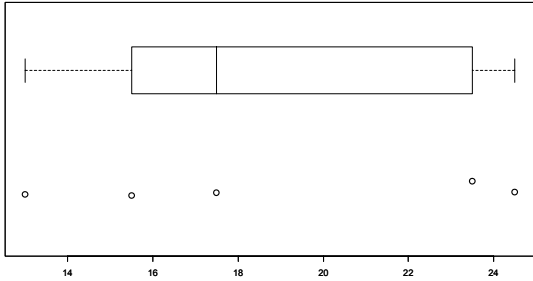


sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

18

boxplot for n = 5!

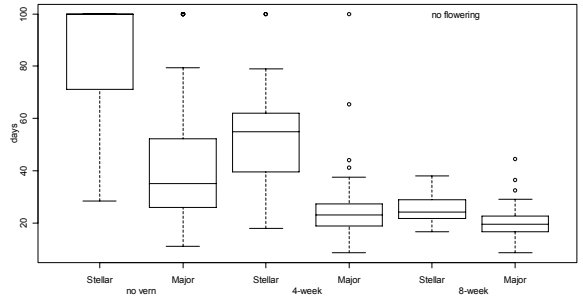


sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

19

box plots show skew in days but hide details (gap to no flowering)

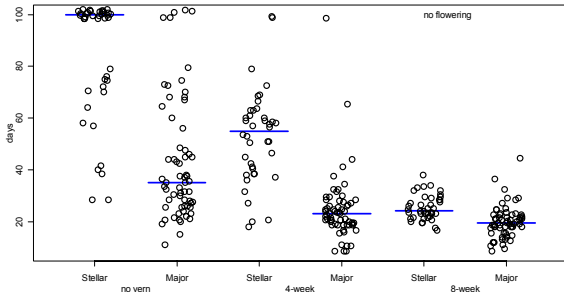


sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

20

jittered plot of data by group (add small noise to x axis; lines for medians)

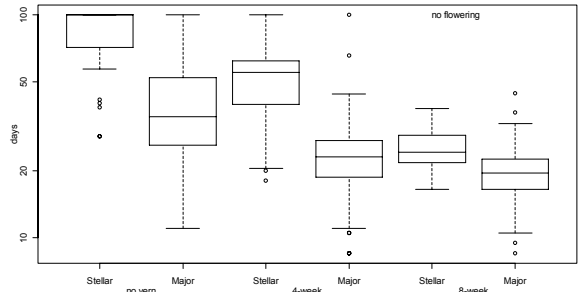


sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

21

box plots on log scale show symmetric, but still hide details

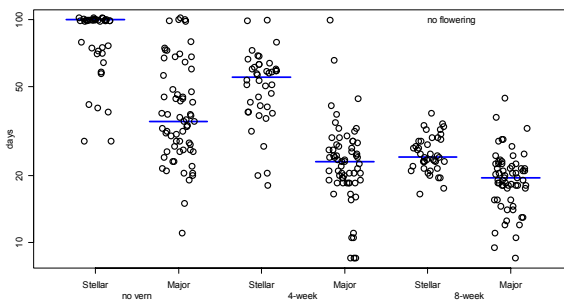


sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

22

jittered plots of log days

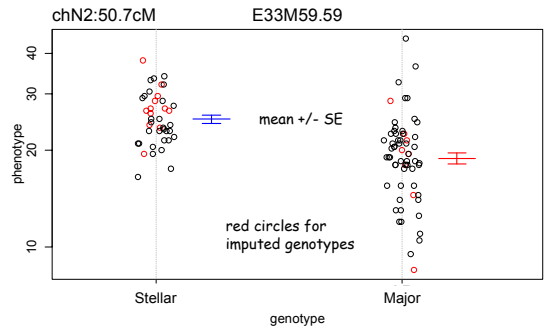


sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

23

phenotype by locus genotype

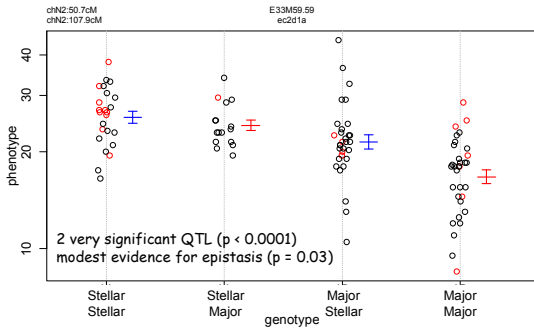


sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

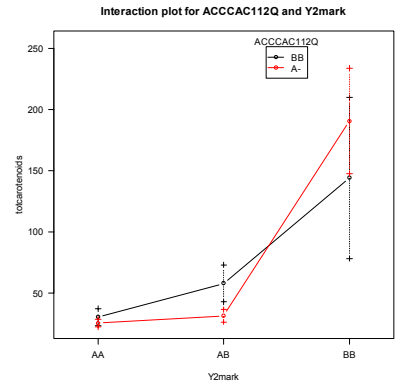
24

phenotype by two loci genotypes



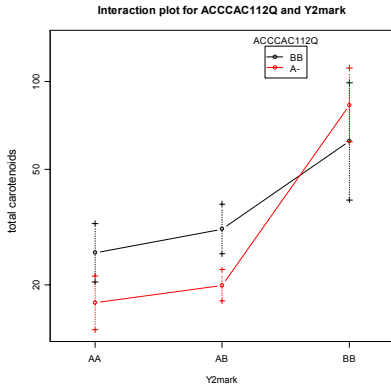
interaction plot

shows phenotype vs. two factors (genetic markers)
includes SE spread



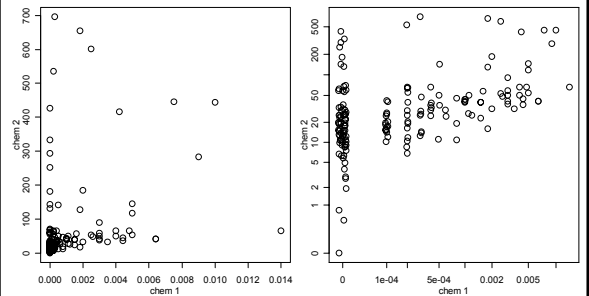
interaction plot

log-transformed phenotype
more equal SEs
similar interaction pattern

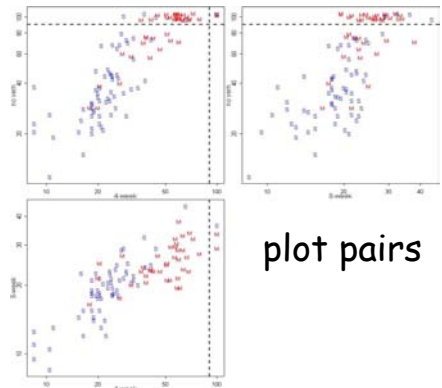
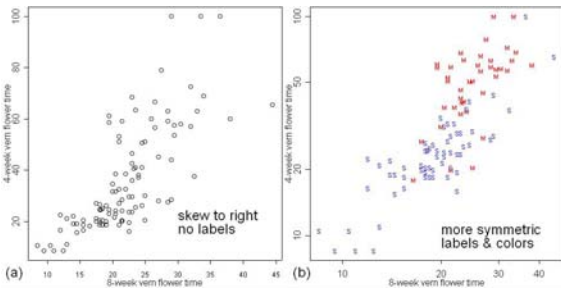


highly skewed data with zeroes

(plot on log scale; add half the min; axis in units)

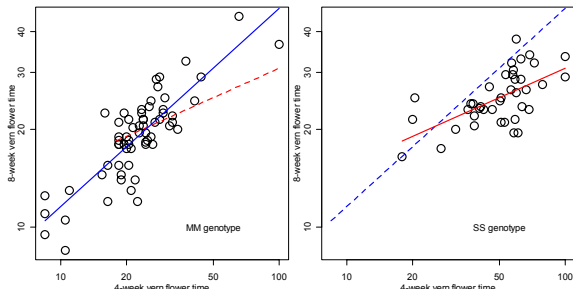


annotated scatter plots can show relationships



plot pairs

regression lines



sun 18 jul 2004

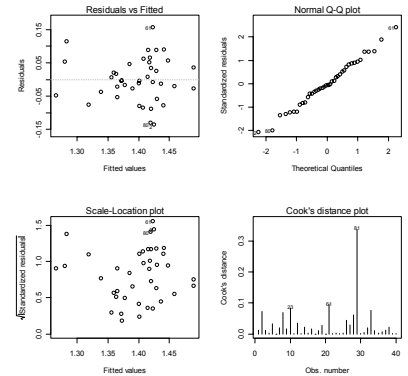
ASHS 2004 © Brian S. Yandell

31

regression diagnostics for MMs

large resid
ind 2,61,80

hi influence
ind 23,61,81

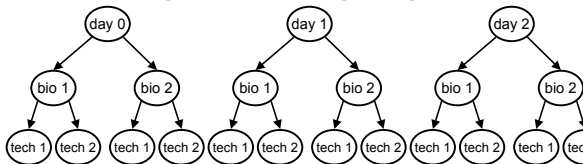


sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

32

design: what's going on?



- 2 biological reps (whole plant tissue)
 - 2 bioreps per day? (6 bioreps) or 2 bioreps total?
- 3 days (0,1,2)
 - independent samples or repeated measurements
- 2 technical reps on each day
 - also known as sub-samples or pseudo-replicates

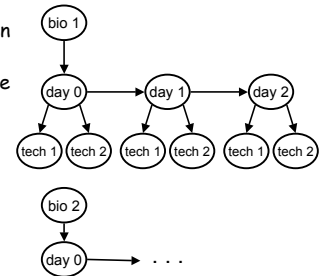
sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

33

design: repeated measures

- repeated measures on each biological rep
- measurements may be correlated over days



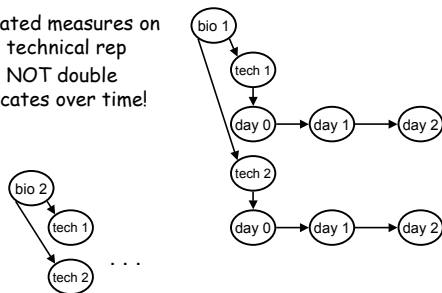
sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

34

design: pseudo-replication

- repeated measures on each technical rep
- does NOT double replicates over time!



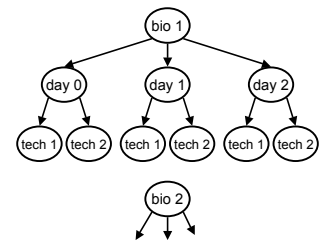
sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

35

design: independent samples

- 2 biological reps (whole plant tissue)
- independent samples from each biorep over 3 days
- 2 technical reps (subsamples or pseudoreps) on each day



sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

36

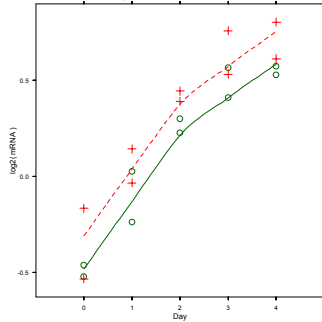
one trait (gene expression)

color/symbol = biorep

technical reps
aligned vertically

smooth parallel lines
show average day trend

p-value for days is small
 $p < 0.00001$

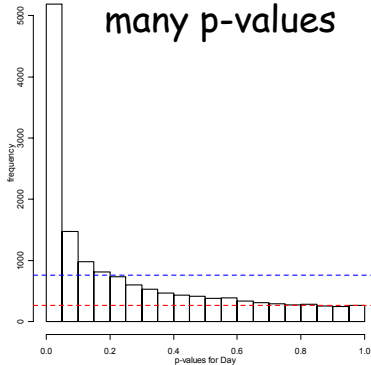


microarray studies

- experimental design & analysis per entry
 - spot reference vs. general design concepts
 - field plots on a smaller scale
 - multiple measurements for same experiment
 - 1000s rather than 10s
 - combining information across entries (variance estimation)
 - multiple testing takes on new meaning
- finding key results
 - shades of gray (grey?)
 - p-value and q-value ideas
 - what is your goal? follow-up studies?
 - confirmation: true positives
 - biochemical pathway studies (KOs, genetics, etc.)

1500 genes
p-values for
day effect
small p = large
day effect

flat histogram
if null true
false discovery
rate ...



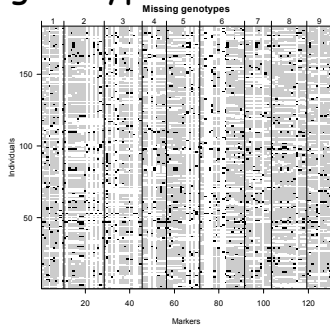
QTL studies

- map information
 - linkage map (shown earlier)
 - missing data and genotypes
- phenotype information
 - histograms (shown earlier)
 - phenotype x genotype plots (shown earlier)
 - LOD plots: 1D and 2D

missing genotype data

black = missing
gray = dominant
white = codominant

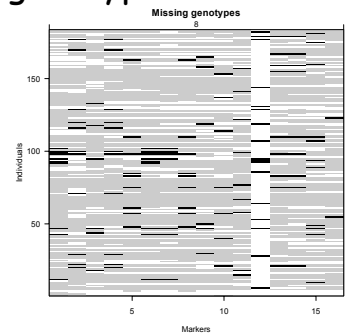
note many dominant
markers but no pattern
to missing data



missing genotype data

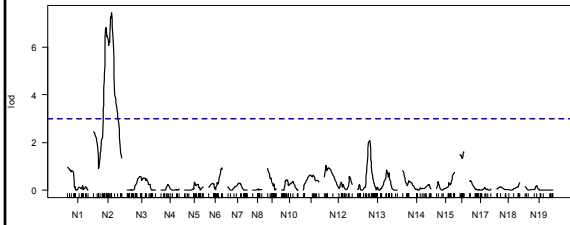
black = missing
gray = dominant
white = codominant

detailed view of chr 8,
individuals 90:110



1-QTL scan

(tall is good; small may still be interesting)



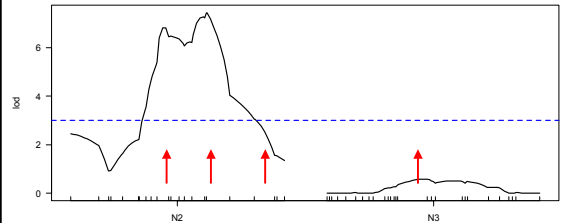
sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

43

1-QTL scan

(model is too simple: misses multiple QTL)



sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

44

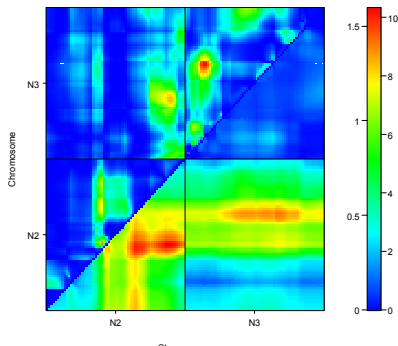
2-QTLs and flowering time

model allows 2 QTL at a time

upper triangle for epistasis (no evidence)

lower triangle for joint effects (strong evidence)

LOD score using color scale



sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

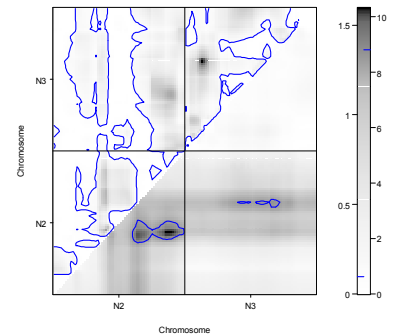
45

2-QTLs and flowering time

upper triangle for epistasis (no evidence)

lower triangle for joint effects (strong evidence)

LOD score using gray scale



sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

46

References

- D Huff
 - *How to Lie with Statistics* (1993 WW Norton & Co)
- H Wainer
 - *How to display data badly* (1984 *Amer Statist* 38: 137)
- ER Tufte (Graphics Press)
 - *The Visual Display of Quantitative Information* (1983)
 - *Envisioning Information* (1990)
 - *Visual Explanation* (1997)
- WS Cleveland
 - *Visualizing Data* (1993 Hobart Press)
 - *The Elements of Graphing Data* (1994 CRC Press)

sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

47

Thanks

- Karl Broman, Johns Hopkins U Biostat
 - notes & ideas
- my hort & agron colleagues
 - Bob Stupar/Jiming Jiang
 - Brian Just/Phil Simon
 - Tom Osborn/Marcio Ferreira/C Kole
- Funding
 - USDA/CSREES Hatch grants
 - NIH/NIDDK grants
 - ASHS

sun 18 jul 2004

ASHS 2004 © Brian S. Yandell

48