"The human understanding, on account of its own nature, readily supposes a greater order and uniformity in things than it finds. And ... it devises parallels and correspondences and relations which are not there."

—Francis Bacon, 1620

"The human understanding, on account of its own nature, readily supposes a greater order and uniformity in things than it finds. And ... it devises parallels and correspondences and relations which are not there.

—Francis Bacon, 1620

Is what we see really there?

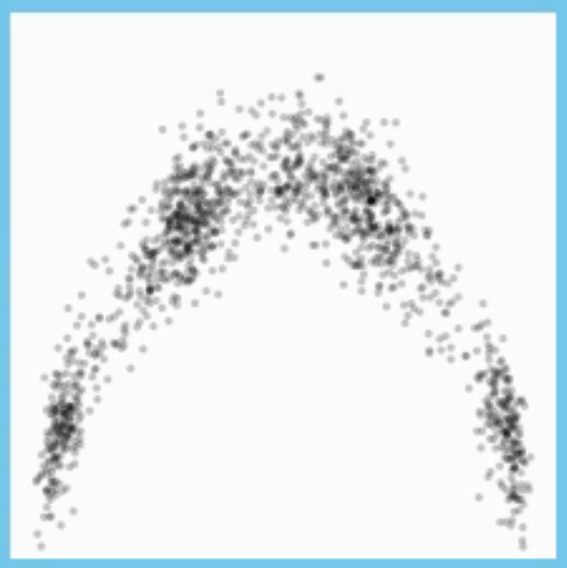
Graphical inference

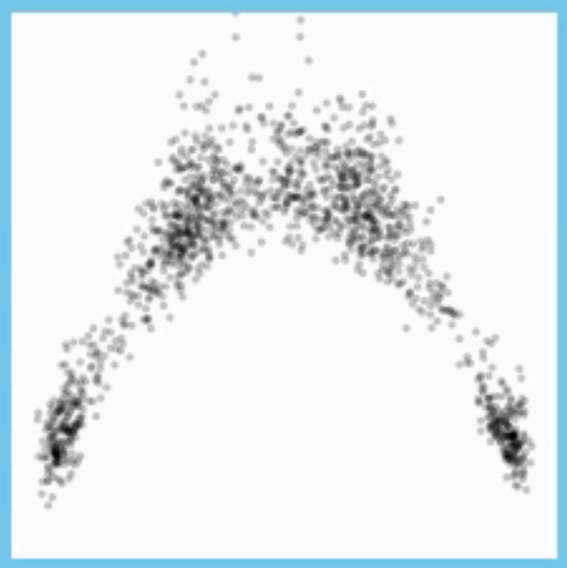
Hadley Wickham, Dianne Cook,
Heike Hofmann, Andreas Buja,
Mahbubul Majumder

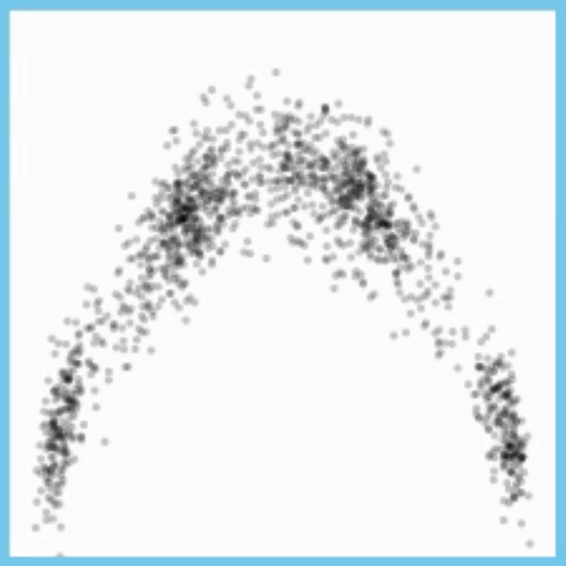
May 2011

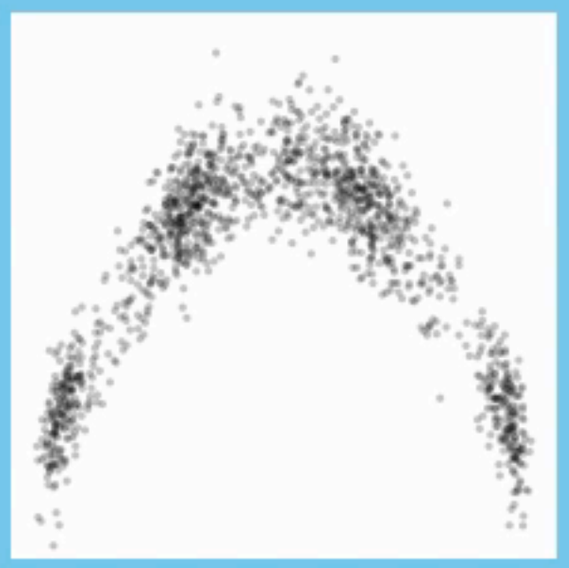
1. Line up protocol
2. Rorschach protocol
3. Case study
4. Future work

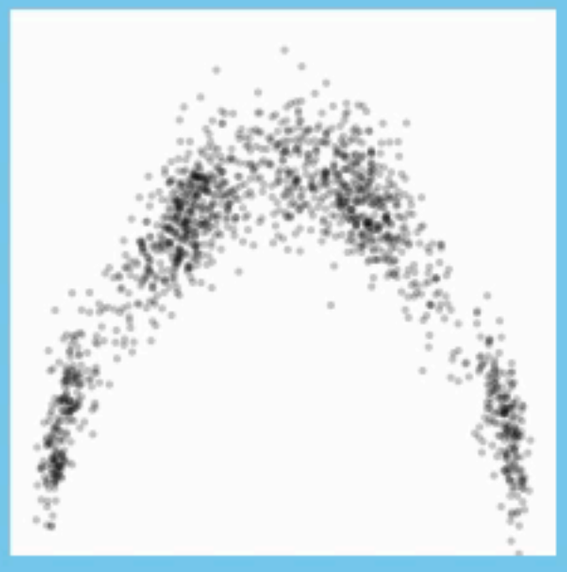
Line up

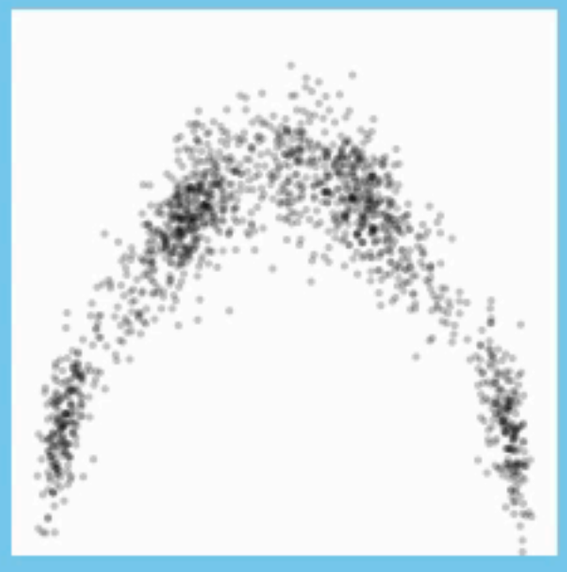


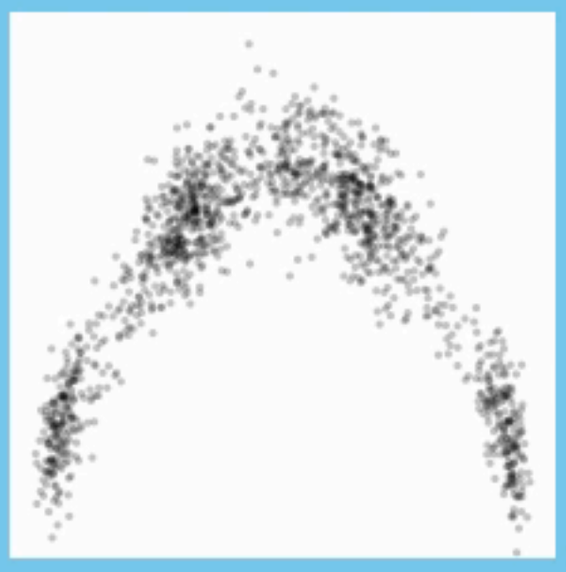


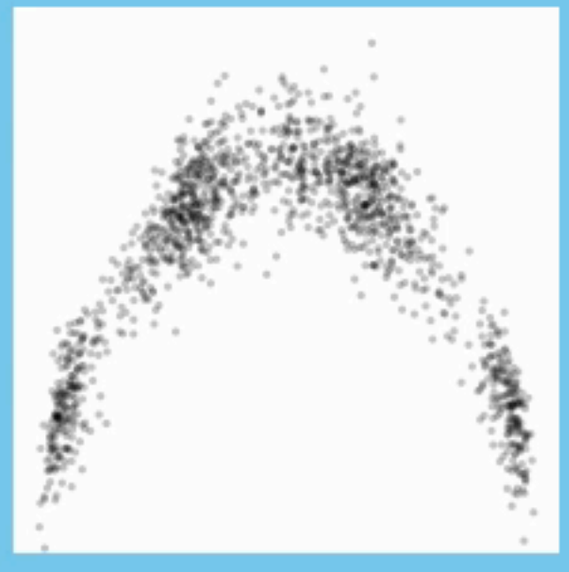






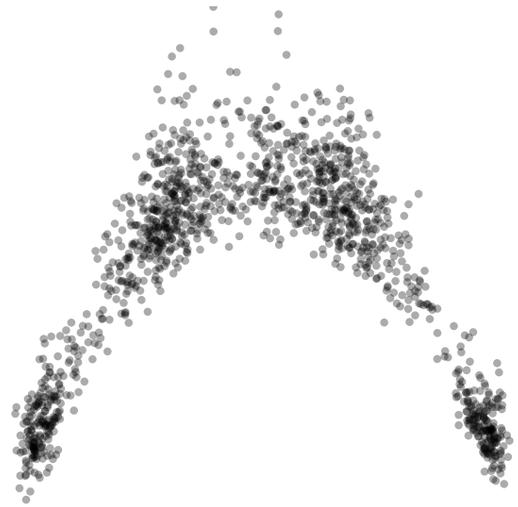






7 of those plots were **null plots**, plots of data drawn from the null hypothesis: a quadratic relationship between x and y. 1 plot was the real data.

Under the null hypothesis, there is a 1/20 chance of picking the correct plot. If we do pick it as being different, we have a p-value of 0.05



We have just performed a statistically valid test!

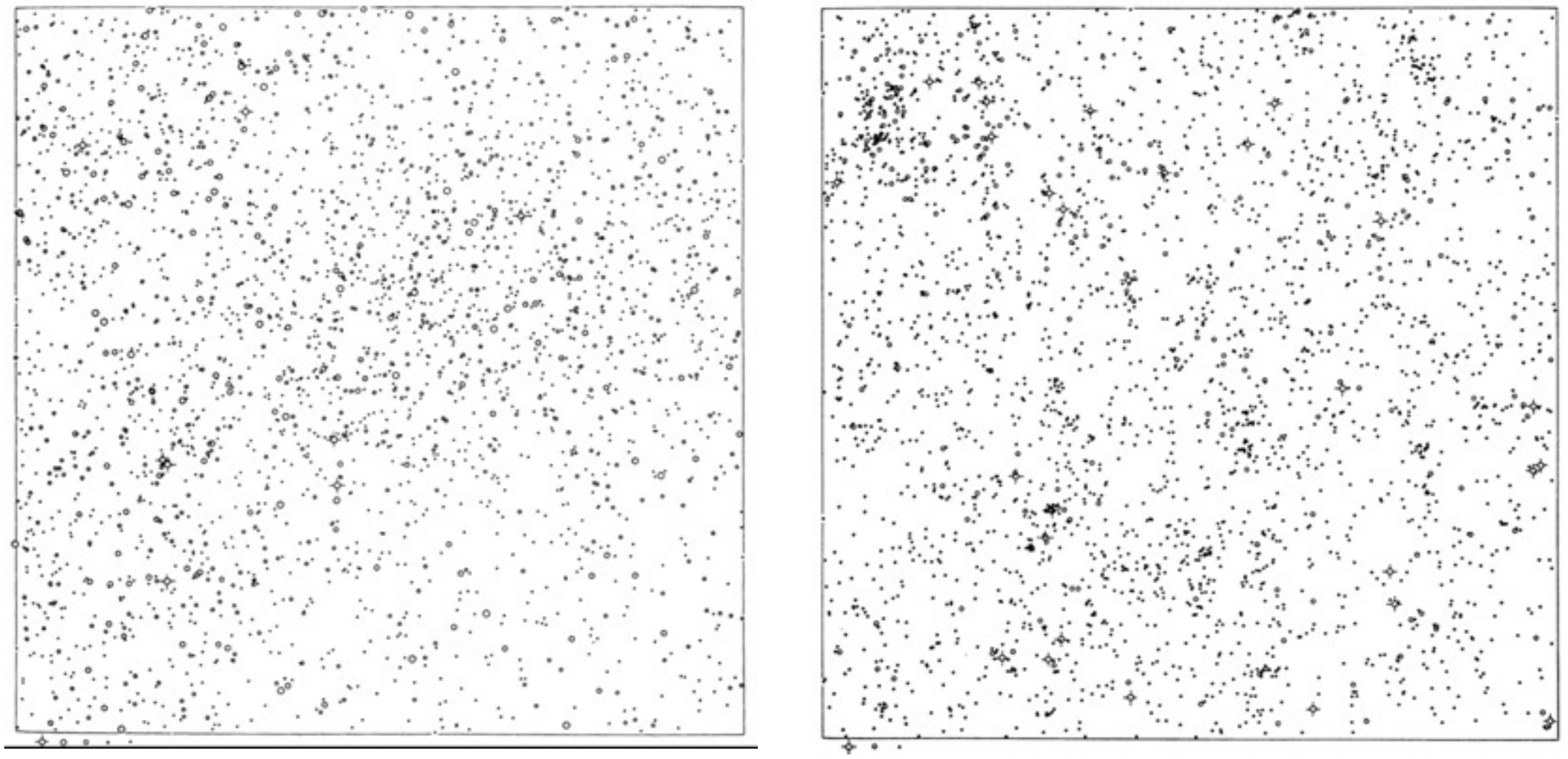
Protocol

Generate n-1 decoys
(null datasets)

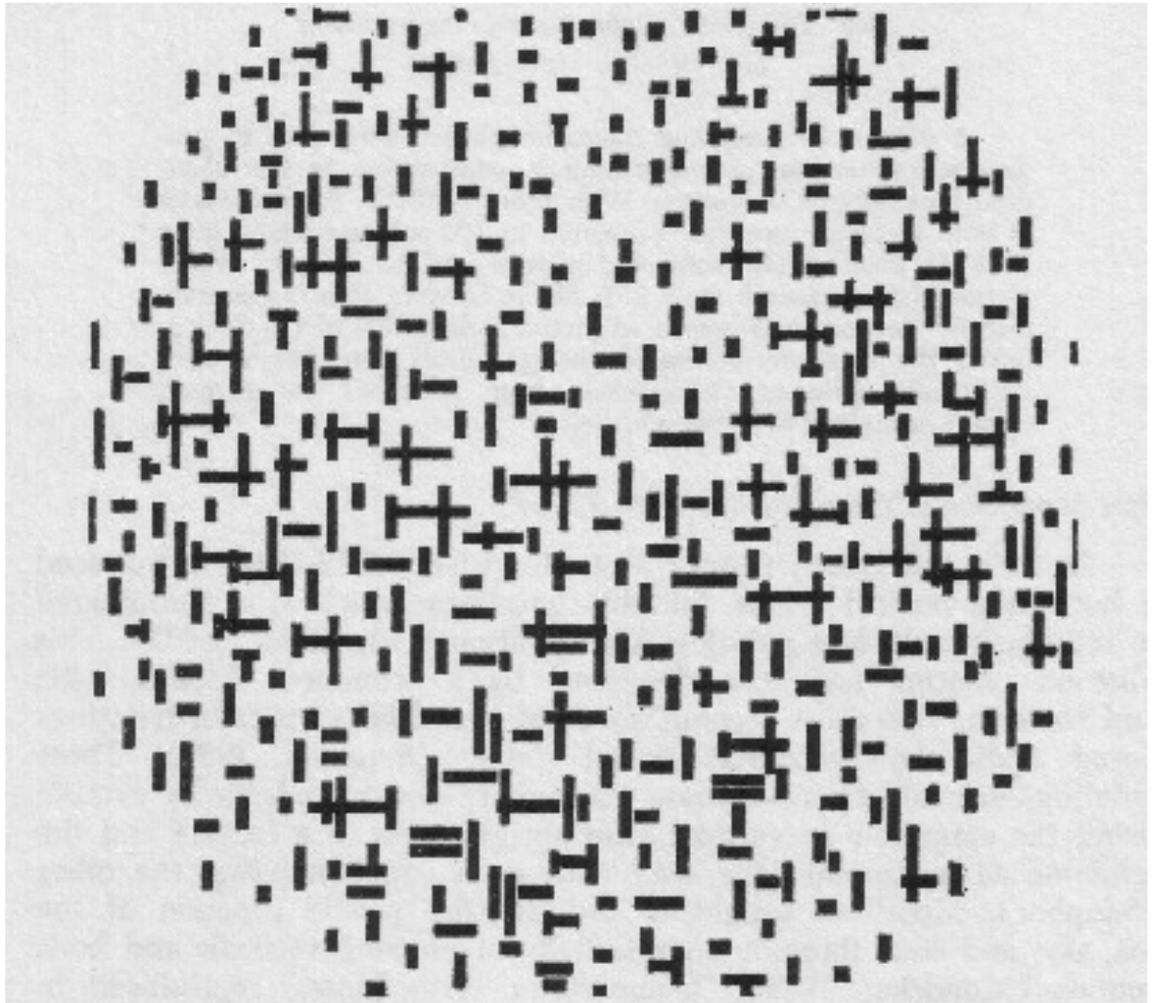
Plot the decoys + the real data
(randomly positioned)

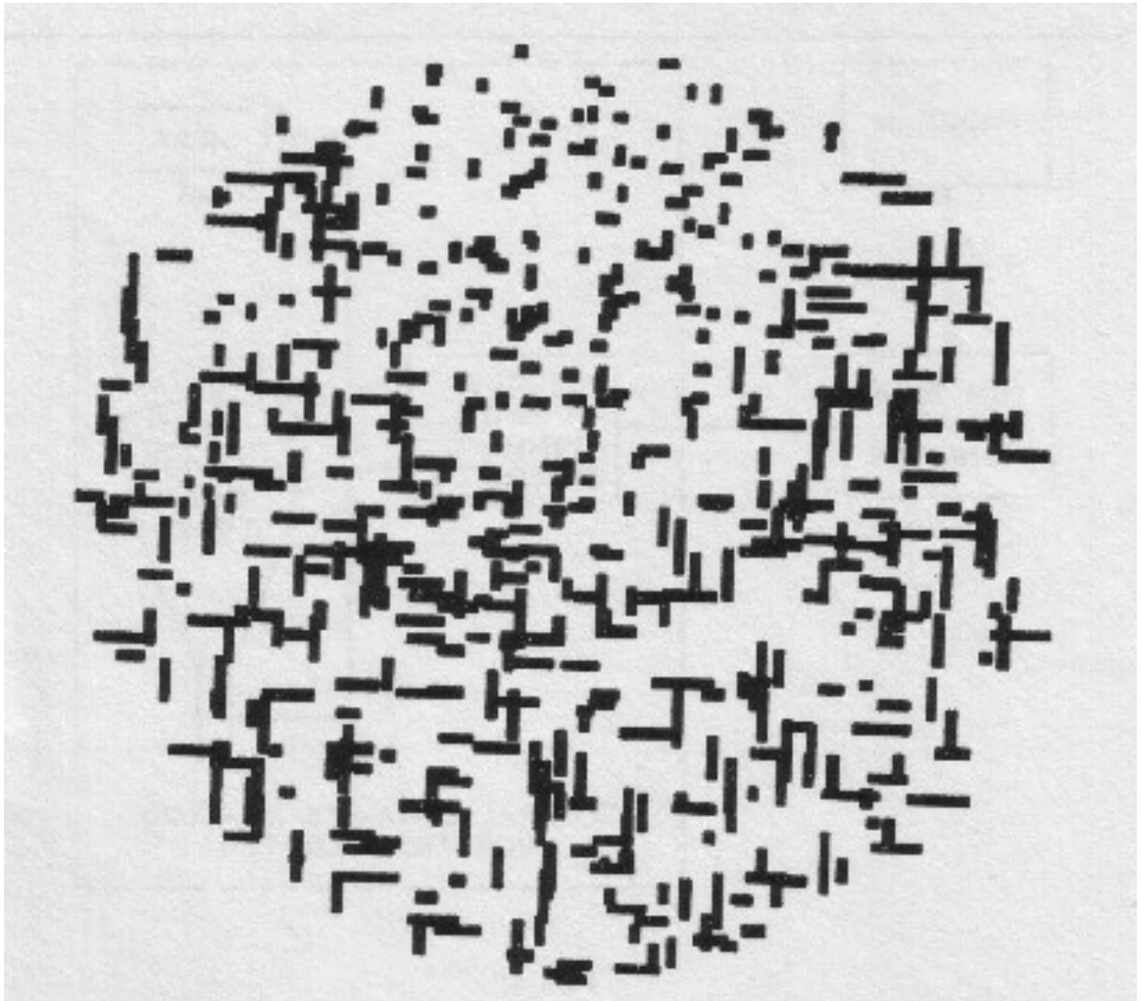
Show to an **impartial** observer.
Can they spot the real data?

If so, you have evidence for true difference
(p-value = 1/n)



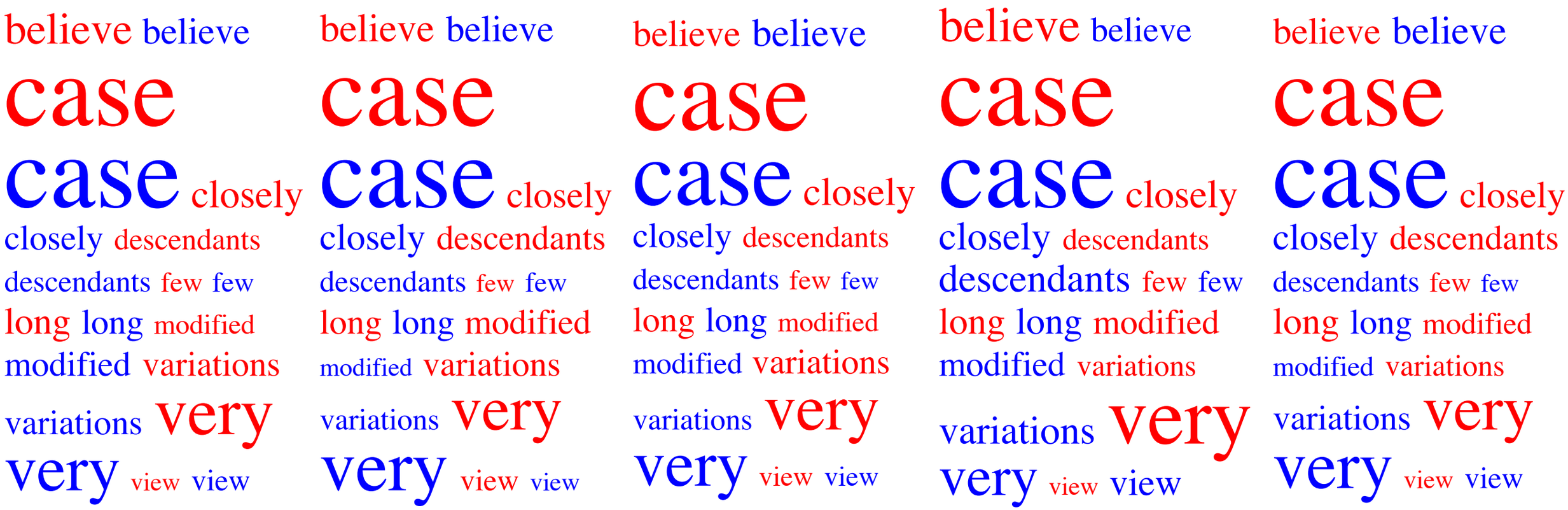
E. L. Scott, C. D. Shane, and M. D. Swanson. Comparison of the synthetic and actual distribution of galaxies on a photographic plate. *Astrophysical Journa*l, 119:91–112, Jan. 1954.



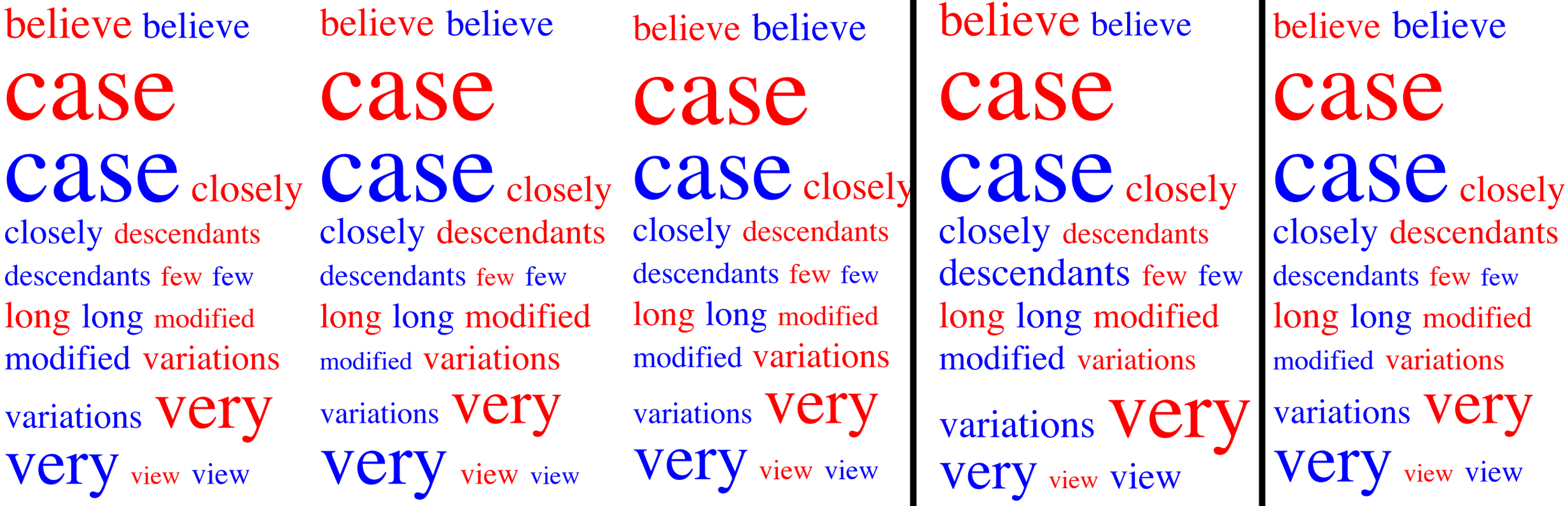


A. M. Noll. Human or machine: A subjective comparison of Piet Mondrian's "composition with lines" (1917) and a computer-generated picture. *The Psychological Record*, 16:1–10, 1966.

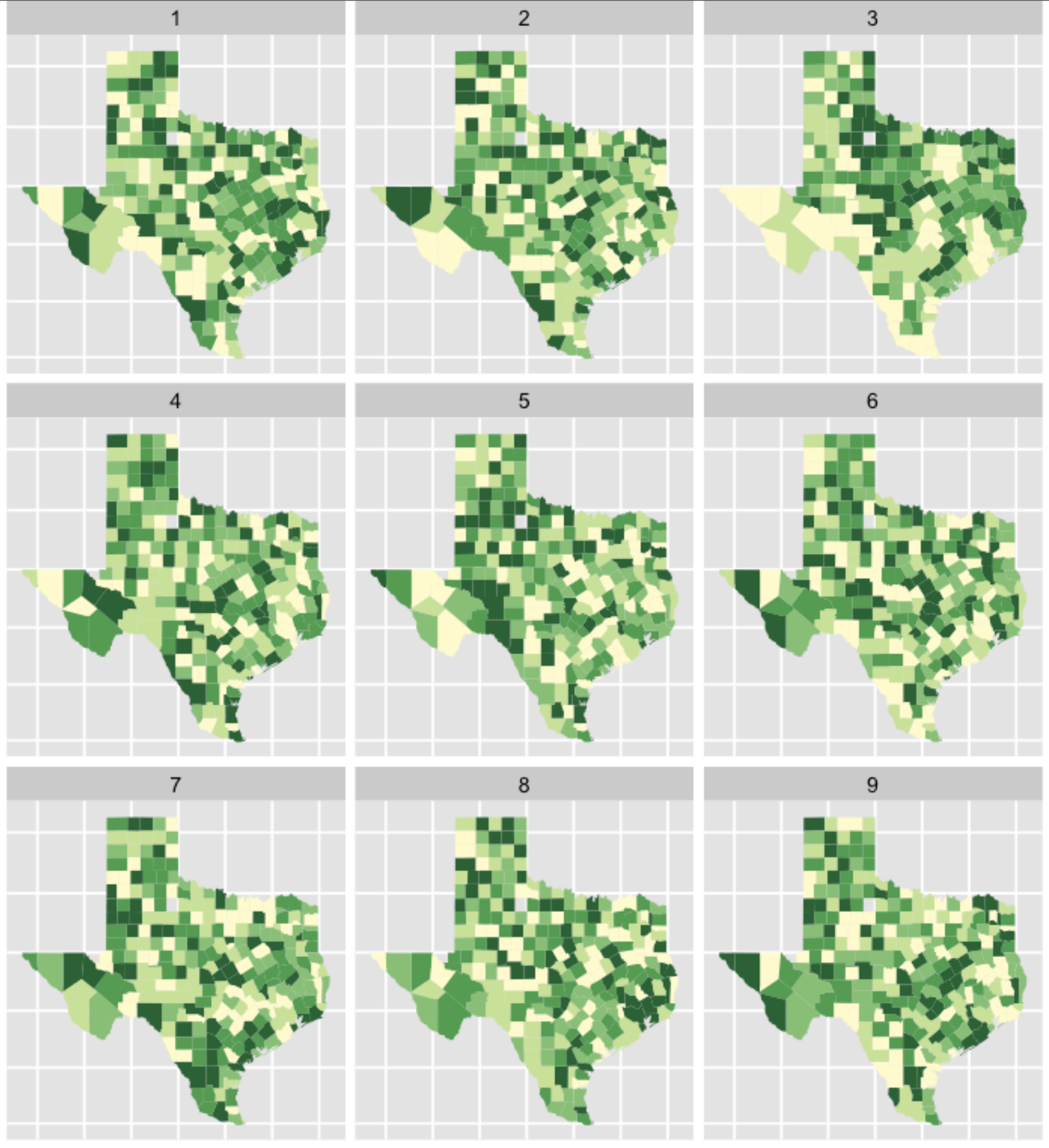
Plot	Task
Scatterplot	Are the two variables independent?
Tag cloud	Do the words come from the same distribution?
Time series	Is there a trend in mean or variability?
Choropleth map	Is there a spatial trend?

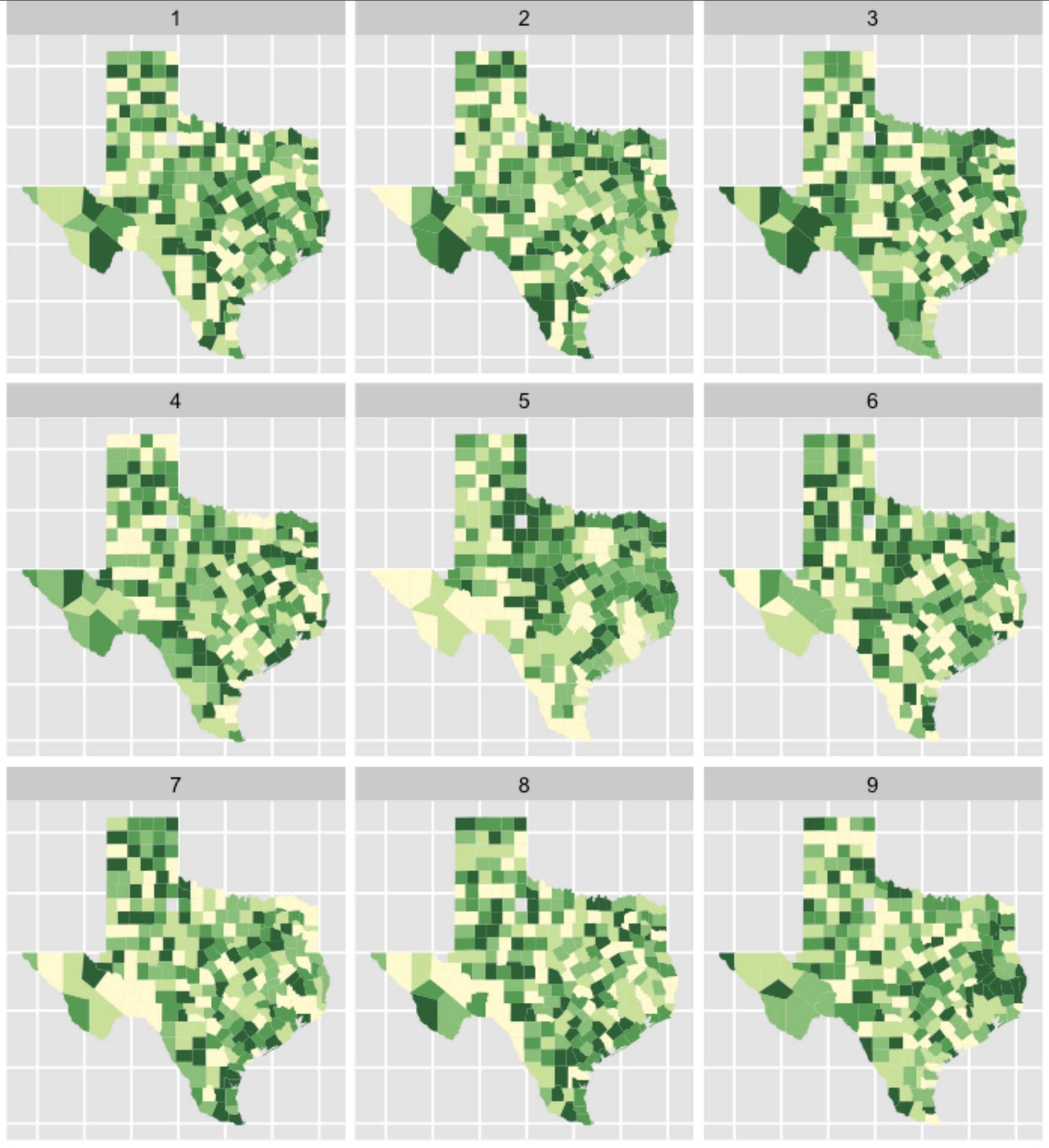


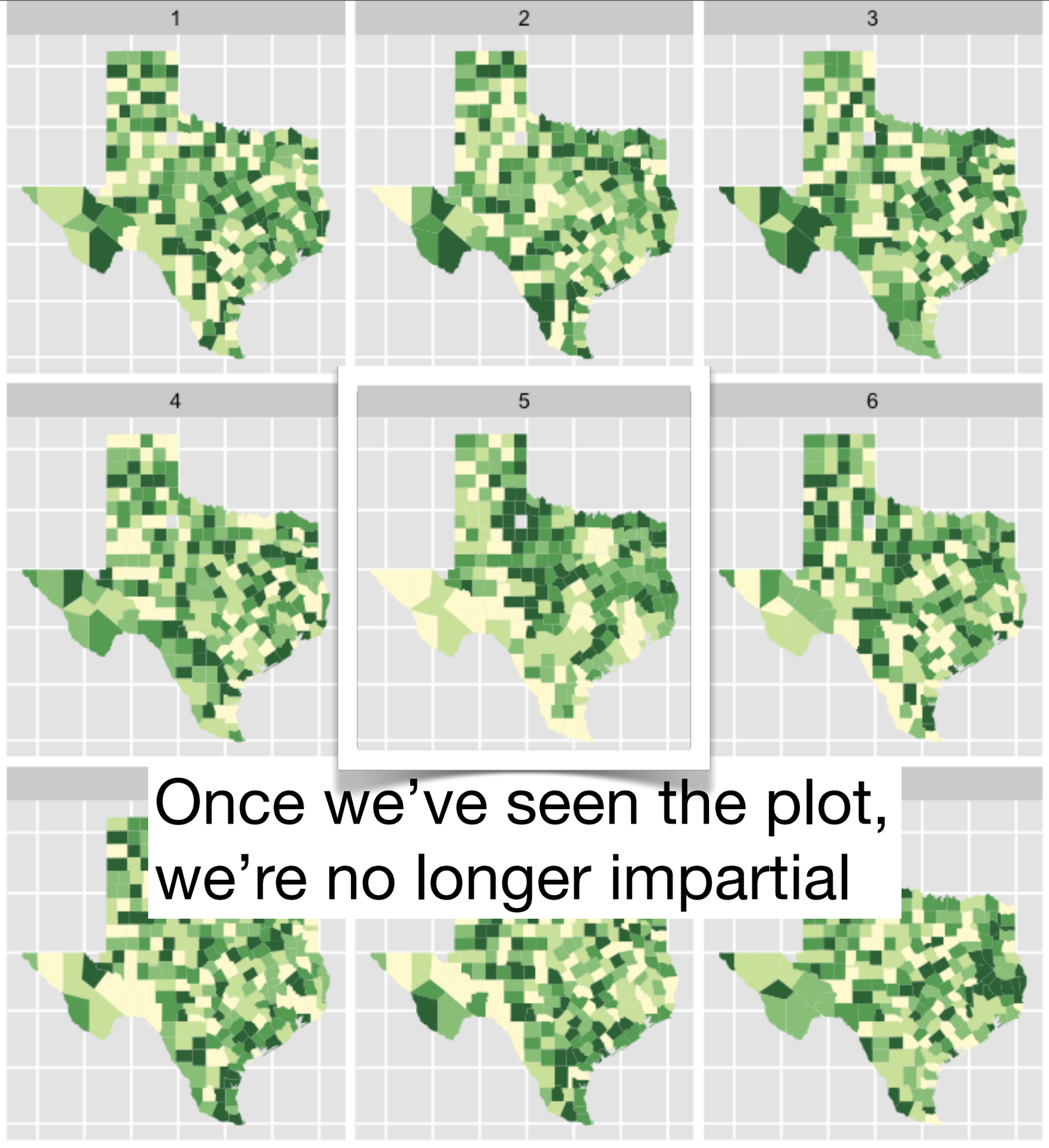
Five tag clouds of selected words from the 1st (red) and 6th (blue) editions of Darwin's "Origin of Species". Four of the tag clouds were generated under the null hypothesis of no difference between editions, and one is the true data. Can you spot it?



Five tag clouds of selected words from the 1st (red) and 6th (blue) editions of Darwin's "Origin of Species". Four of the tag clouds were generated under the null hypothesis of no difference between editions, and one is the true data. Can you spot it?


1
2
3
4
5
6
7
8
9


1
2
3
4
5
6
7
8
9



Once we've seen the plot, we're no longer impartial

Solutions

Show to colleagues/collaborators

Automated visual testing service using amazon mechanical turk

Multiple Quantitative Testing:

Null Hypothesis

$\Downarrow$

Collection of Test Statistics:
$T^{(i)}(\mathbf{y}) \quad (i \in I)$

$\Downarrow$

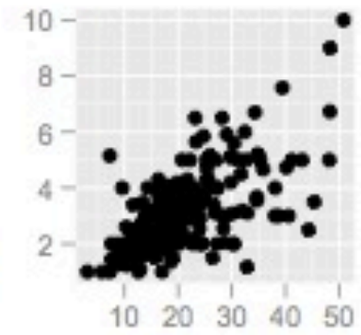
Tests: Any Rejections?
For which $i \in I$ is $T^{(i)}(\mathbf{y}) > c^{(i)}$?

Visual Discovery:

Null Hypothesis

$\Downarrow$

Plot of $\mathbf{y}$: Visible Features



$\Downarrow$

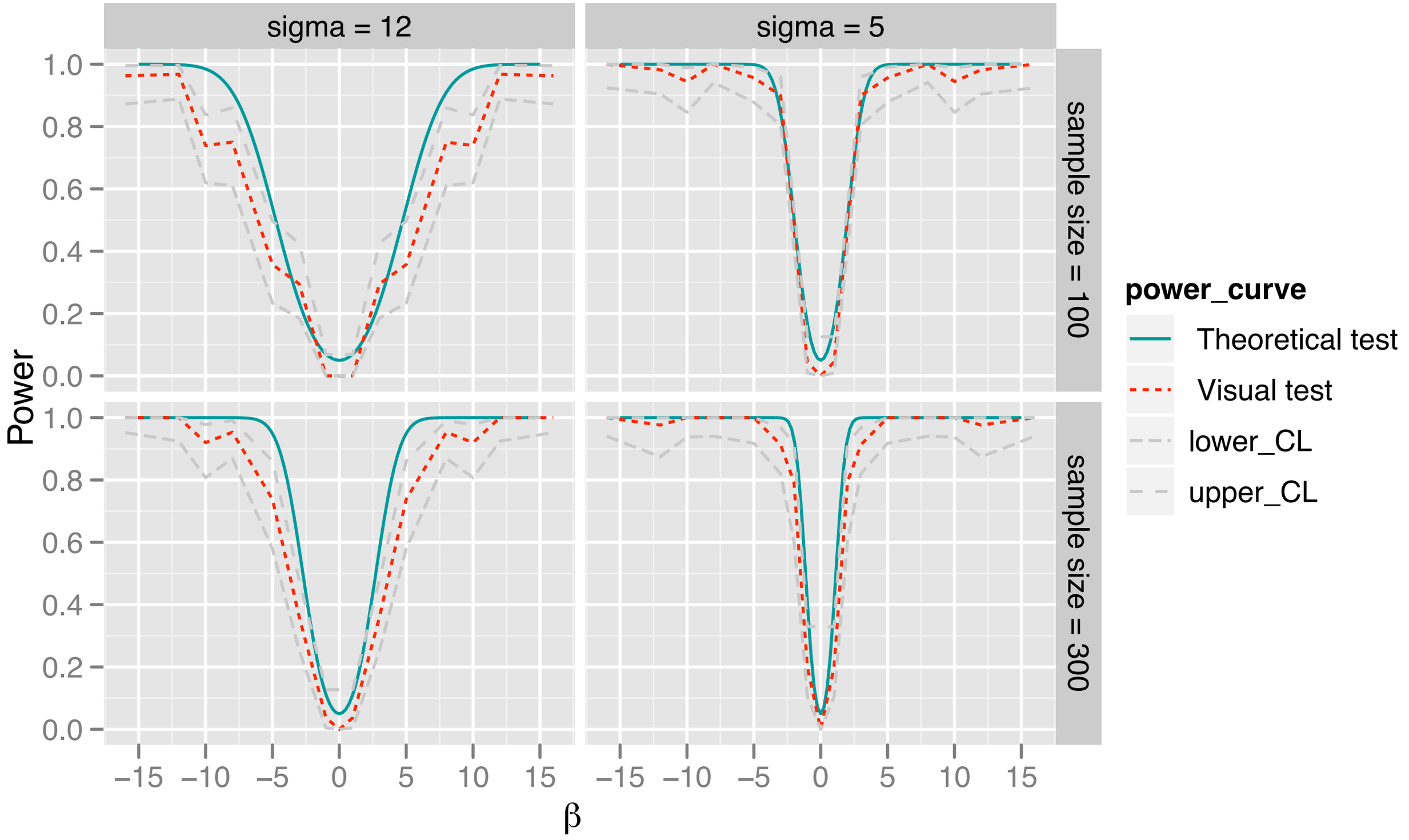
Human Viewer: Any
Discoveries? What kind?

vs. classical tests

Of course, if we know what we're looking for, we can always develop an algorithm or numerical test.

The **advantage** of visual inference is that works for very general tasks, including when you don't know exactly what you're looking for.

Recent work suggest that power only a little worse than classical test



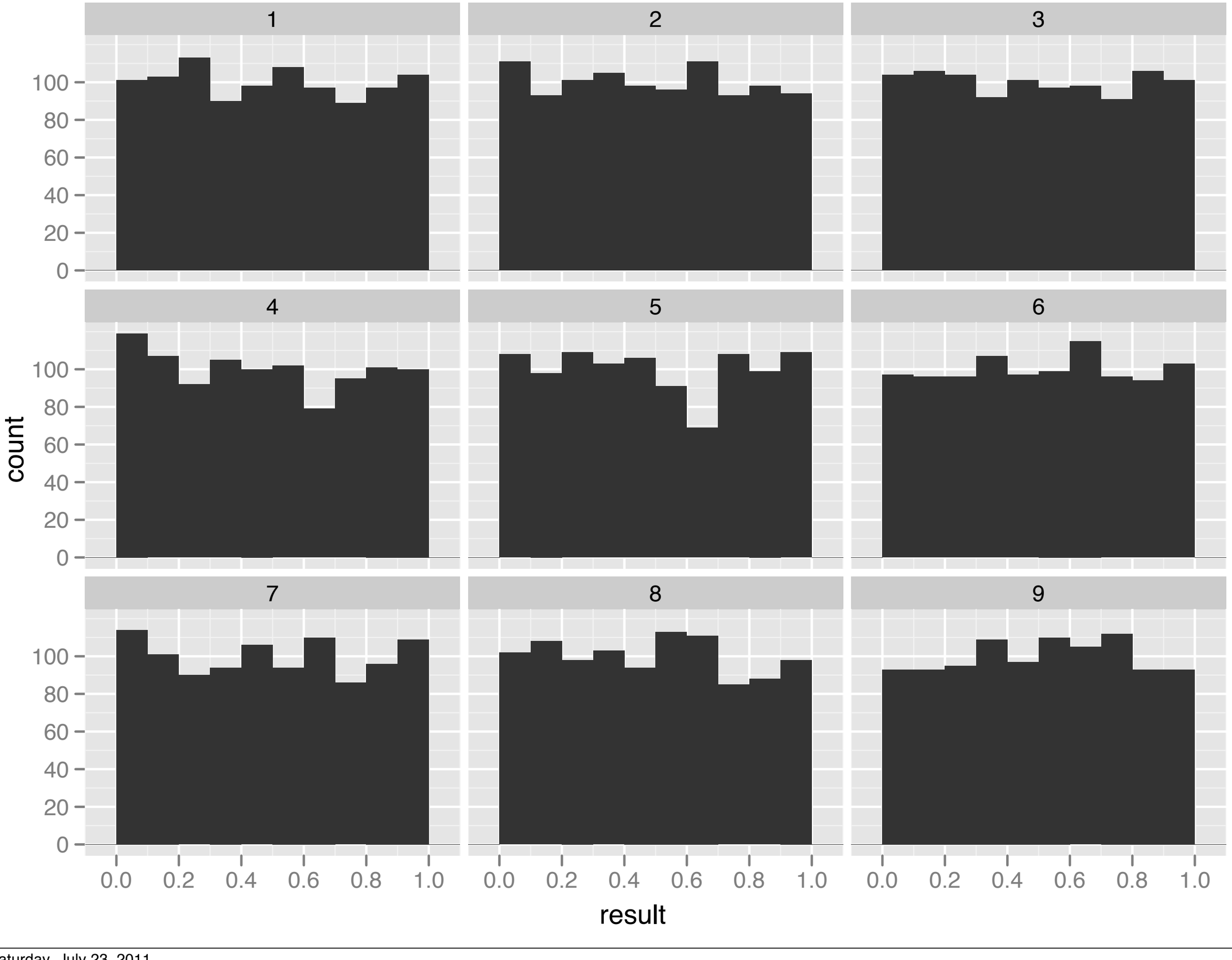
Rorschach

Rorschach

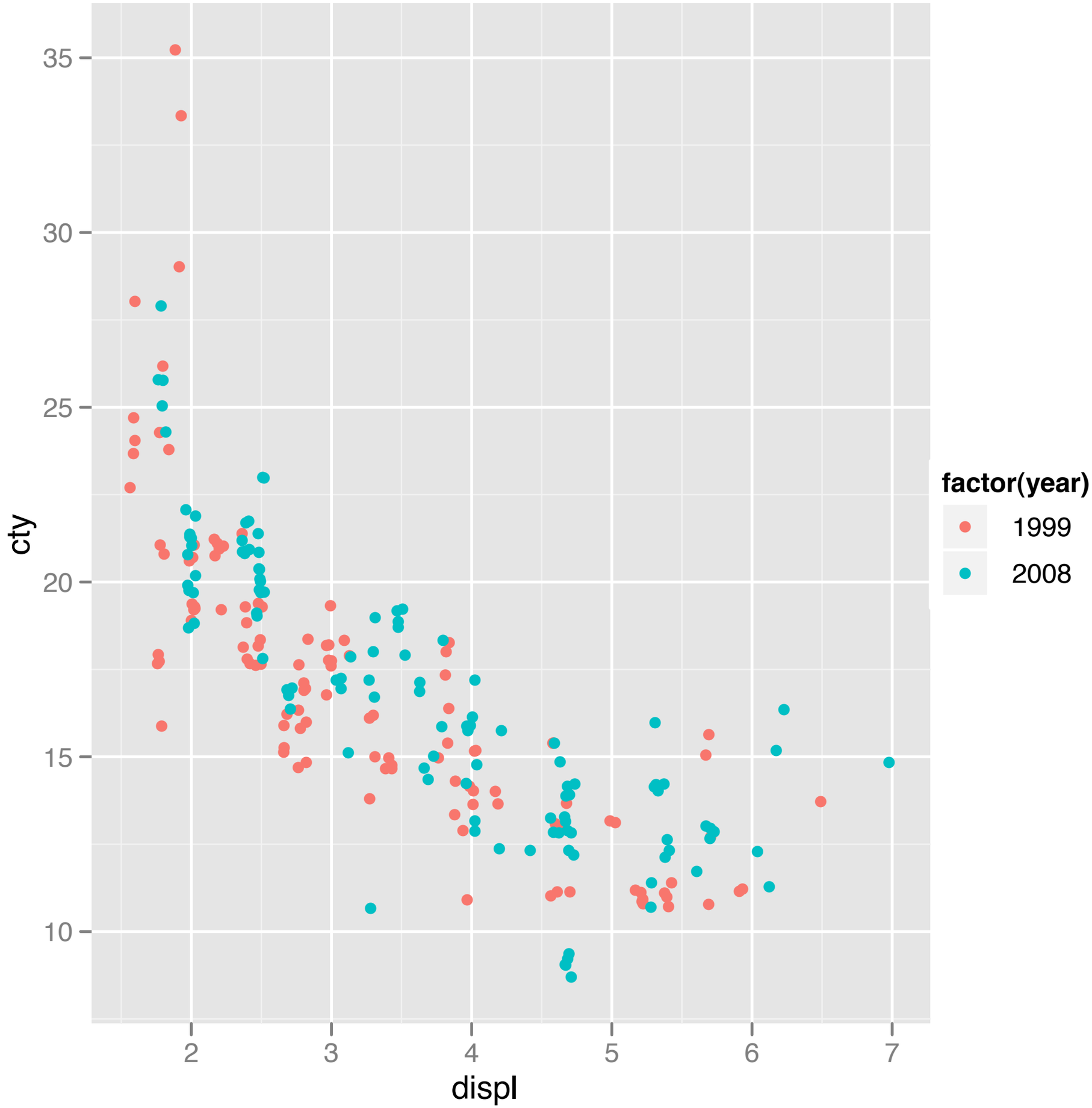
We're surprisingly bad at appreciating the amount of variation in random data.

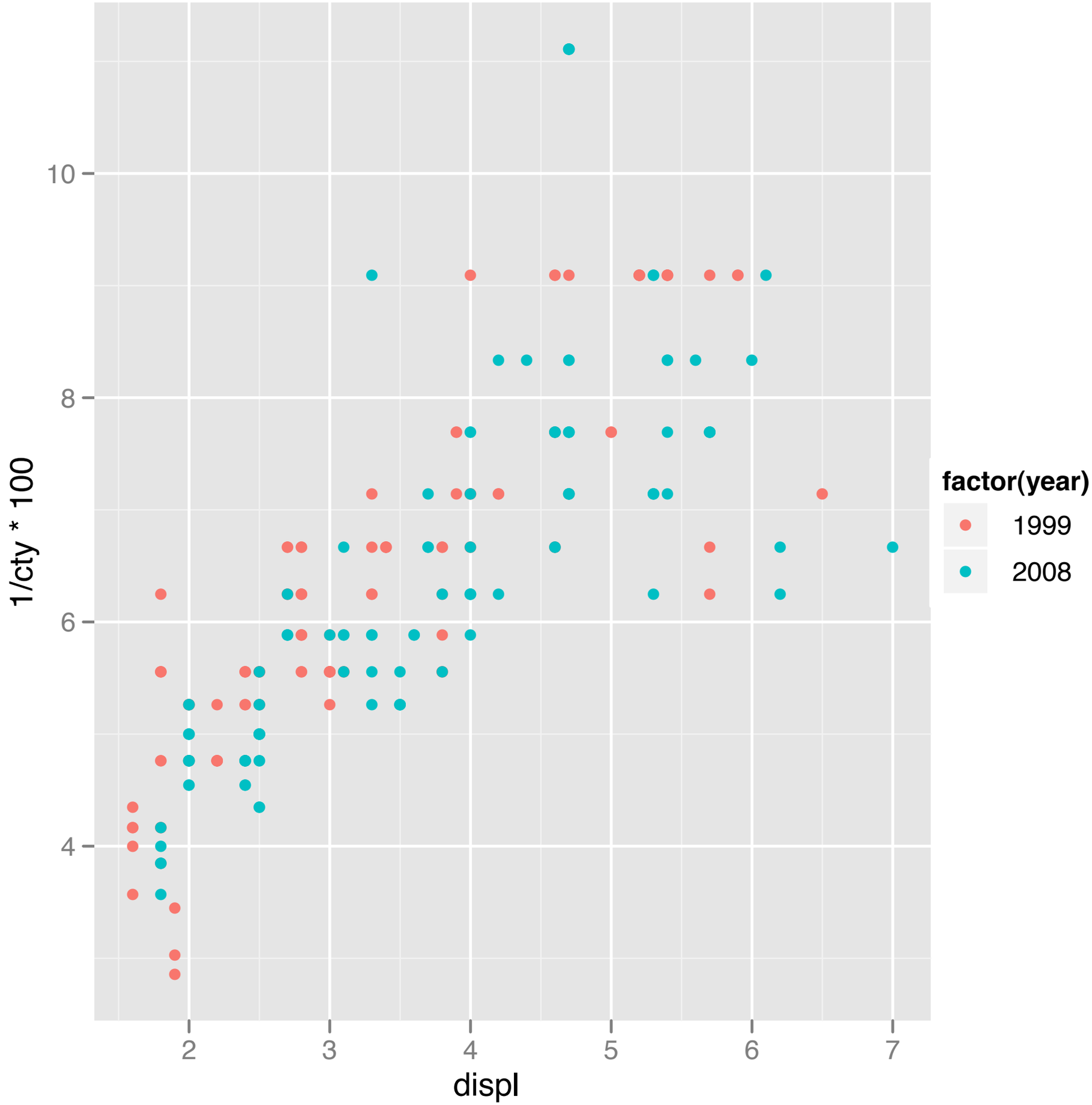
Showing only null plots is a good way to calibrate our intuition.

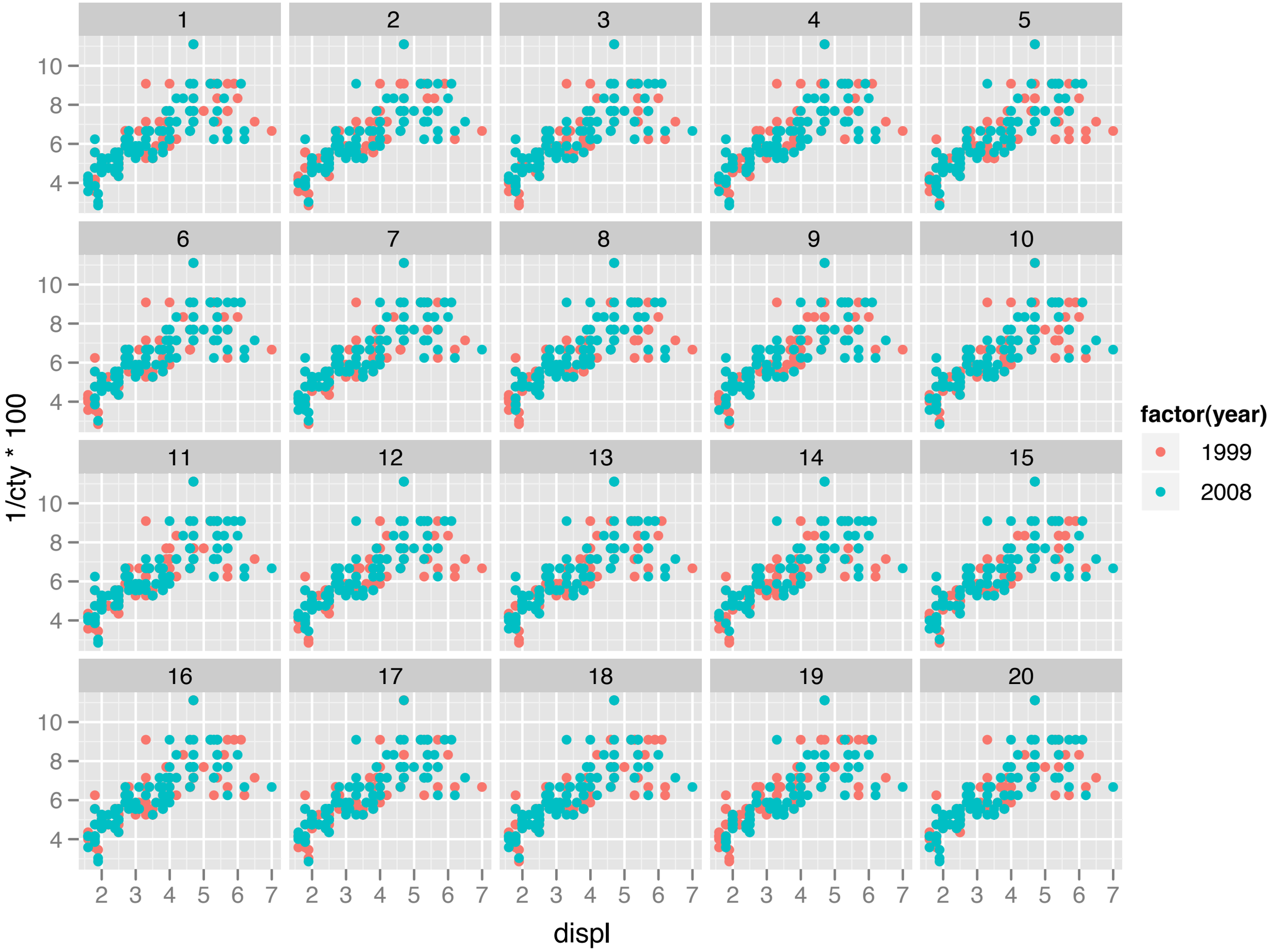
We also plan on using these plots as an empirical tool to understand what features people pick up on. Anecdotally, undergrads focus too much on outliers


1
2
3
4
5
6
7
8
9
count
result
0
20
40
60
80
100
0.0
0.2
0.4
0.6
0.8
1.0

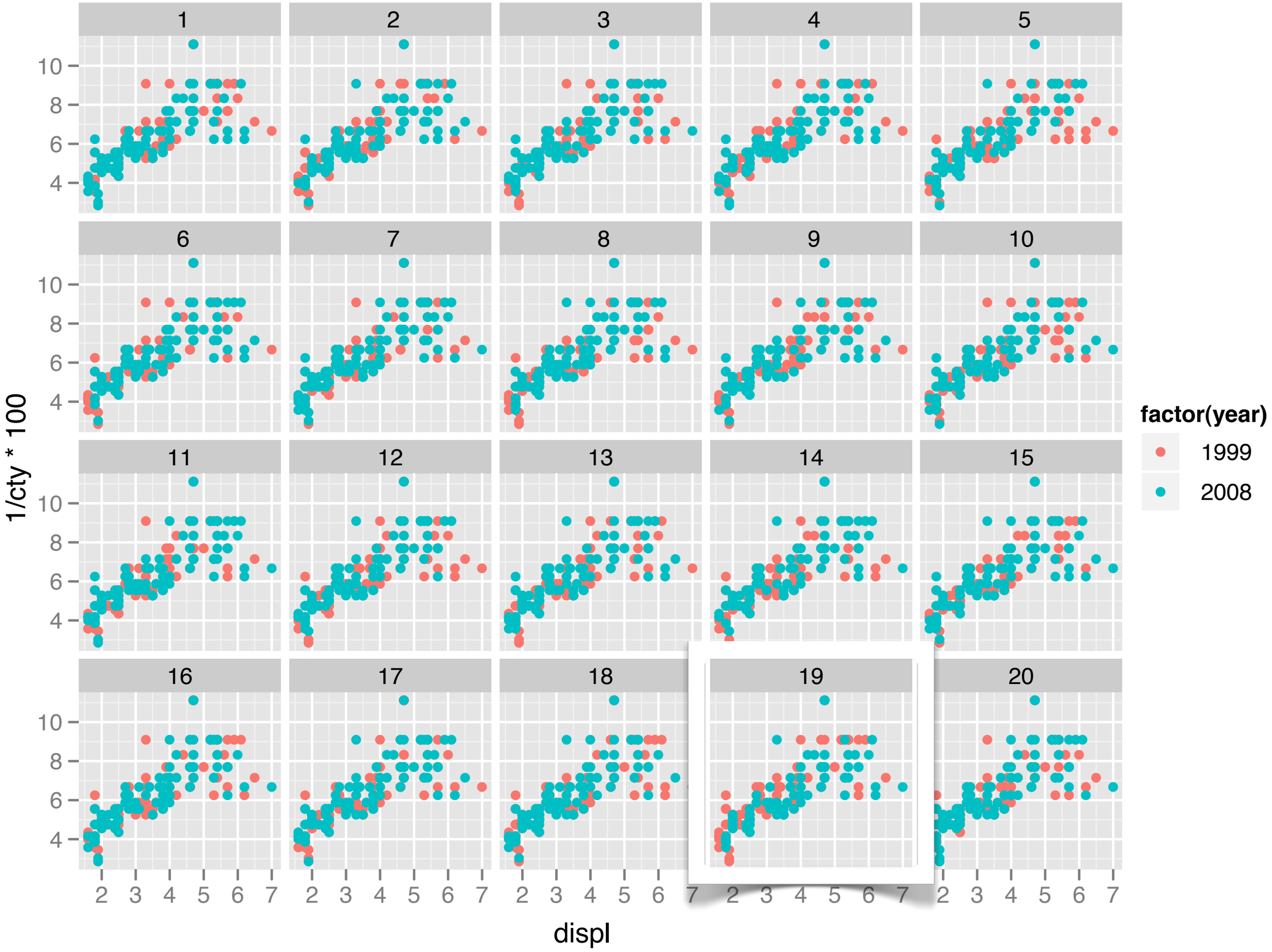
Case study


cty
35
30
25
20
15
10
2
3
4
5
6
7
displ
factor(year)
1999
2008


10
8
6
4
1/cty * 100
2
3
4
5
6
7
displ
factor(year)
1999
2008

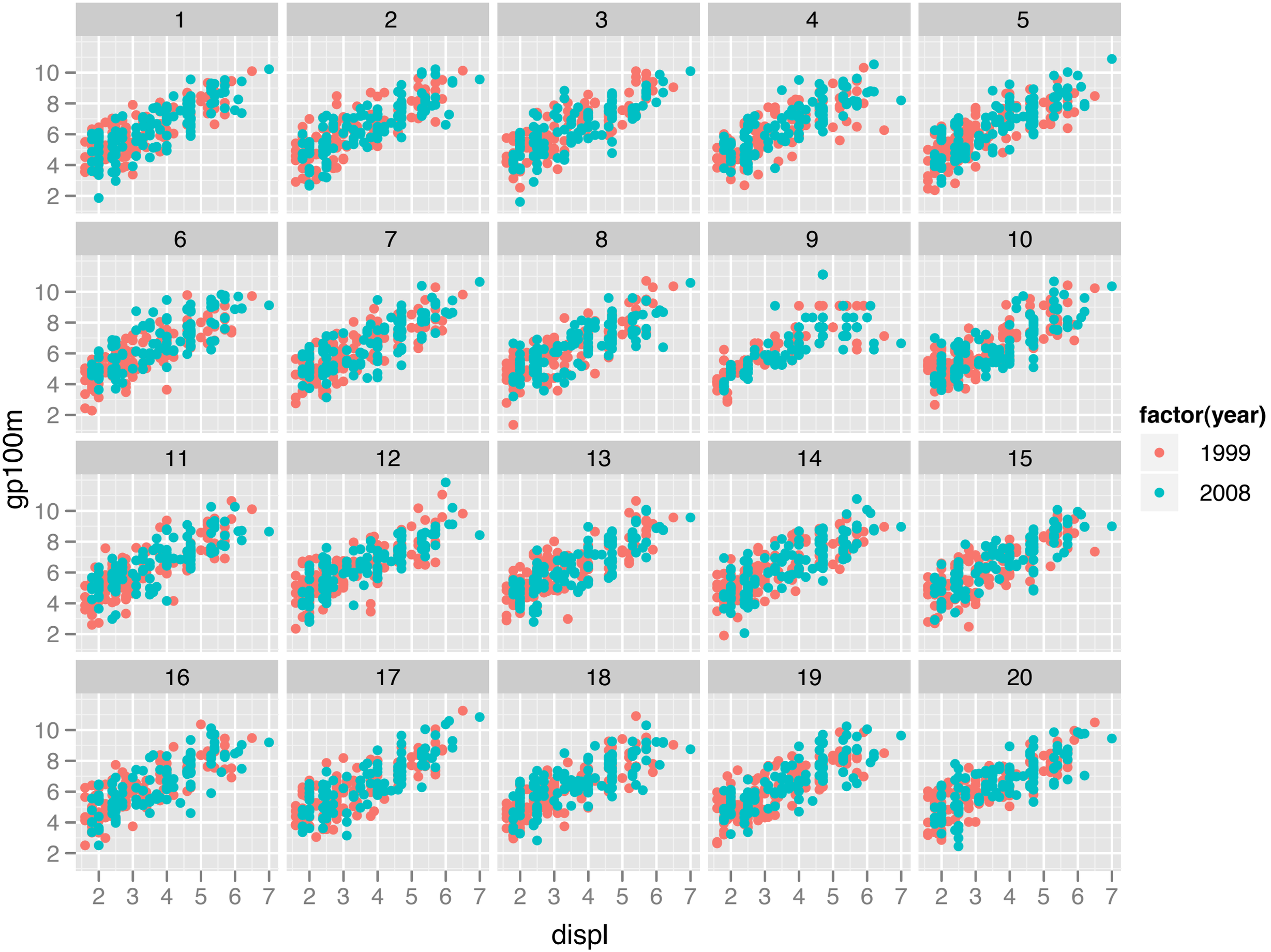


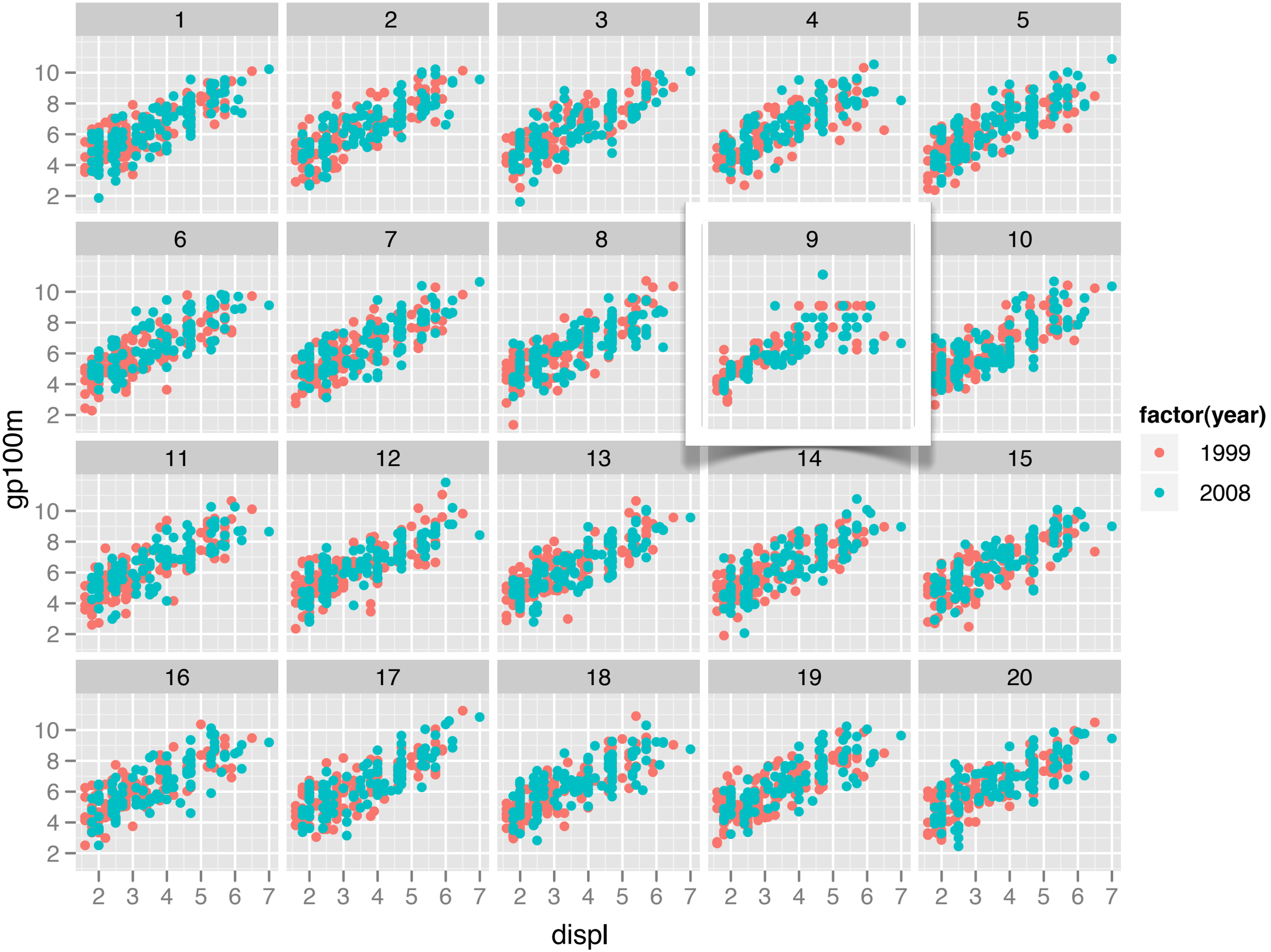
Saturday, July 23, 2011



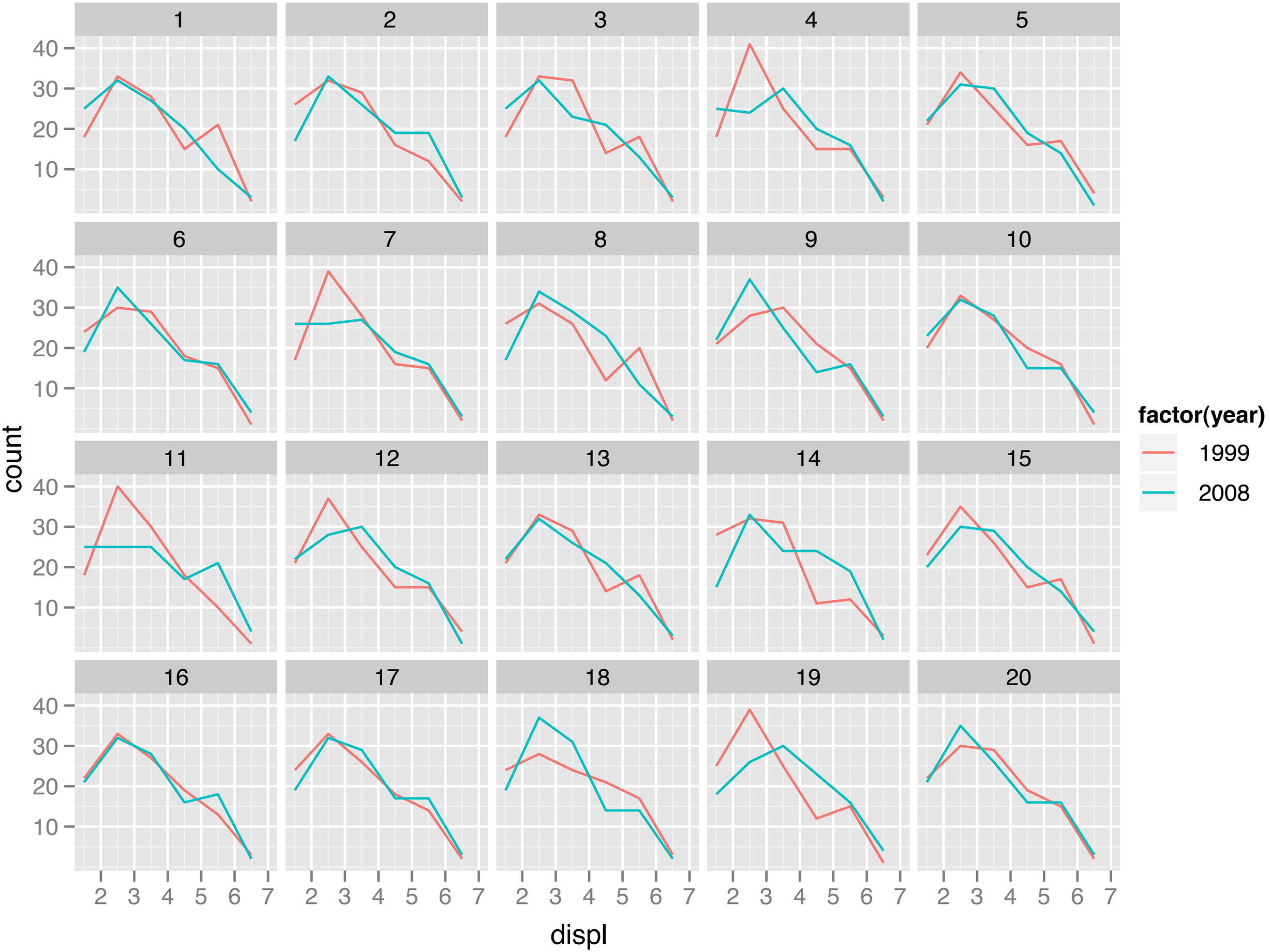
Saturday, July 23, 2011

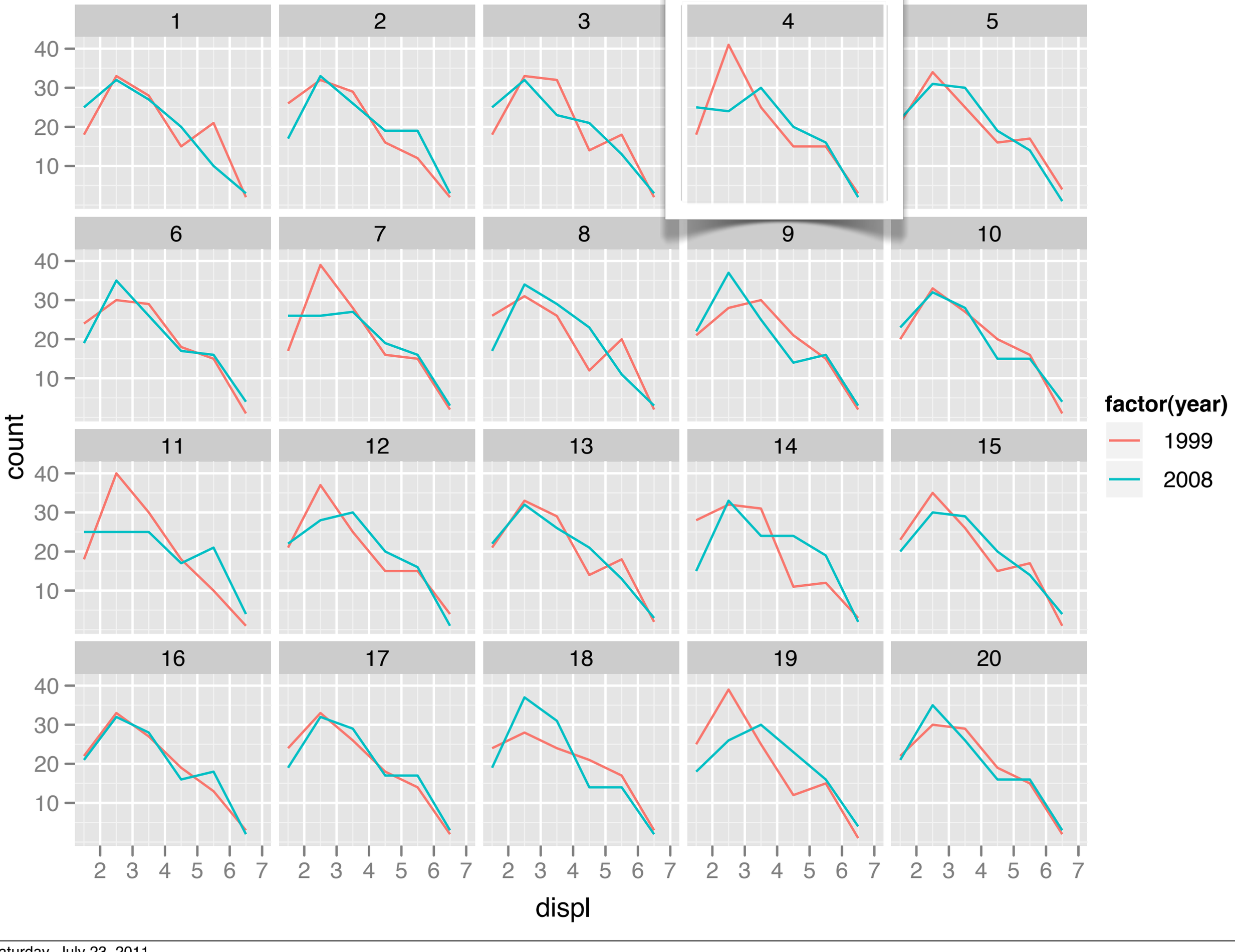
Is a linear model with displacement as single predictor adequate?


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
gp100m
displ
factor(year)
1999
2008


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
gp100m
displ
factor(year)
1999
2008

Maybe there are fewer bigger cars?


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
40
30
20
10
2
3
4
5
6
7
count
displ
factor(year)
1999
2008


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
40
30
20
10
2
3
4
5
6
7
count
displ
factor(year)
1999
2008

Future work

Future work

How can visual inference be integrated into visualisation software at a fundamental level? Is it possible to guess plausible null hypotheses from the plot specification?

How does training affect results? How do novices and experts differ?

What patterns do people pick up on? What are the alternatives that people respond to?

Questions?