Bringing Visual Inference to the Classroom

SDSS 2020

Adam Loy

Carleton College

@aloy

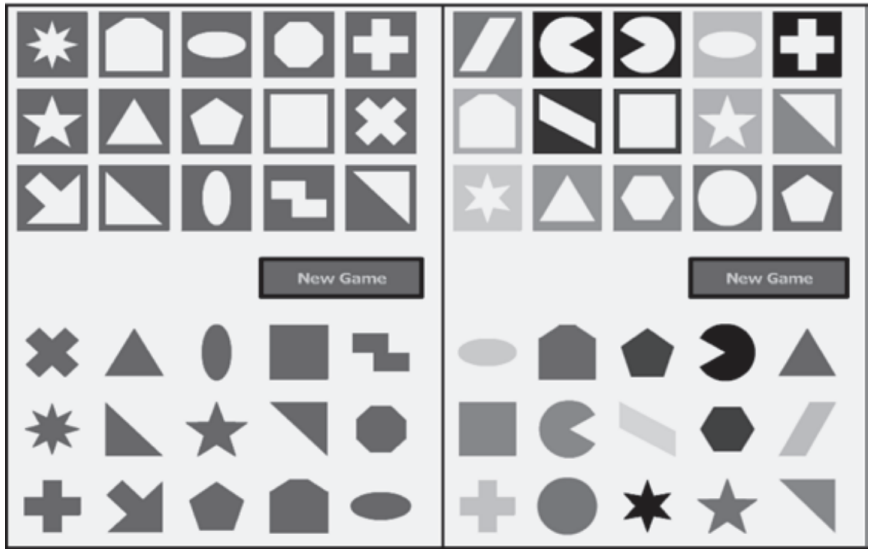
aloy.rbind.io

The move to a simulation-based curriculum

- Since 2007, we've seen a shift to simulation-based inference in the intro course
- Validation studies (Tintle et al. 2014; Maurer & Lock 2014; Hildreth et al. 2018)
- Implementation in other courses
  - Statistical inference (Cobb 2011; Chihara & Hesterberg 2011)
  - Throughout curricula (Tintle et al. 2015)
- All have similar approach to visualization of the inferential process

Do Distracting Colors Influence the Time to Complete a Game?

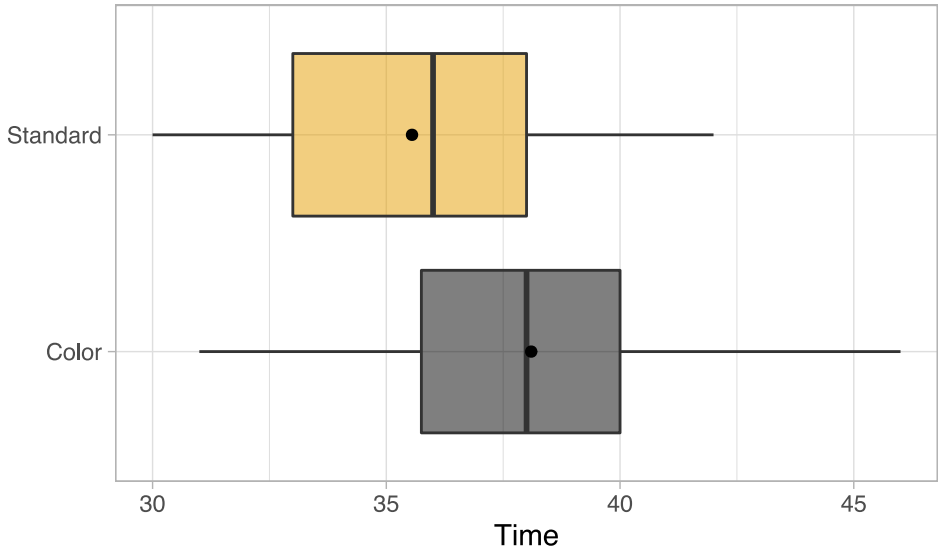
- 20 students randomly assigned to the standard game (left),
  20 students a game with a color distracter (right)
- Subjects played the game in the same area with similar background noise
- Collected the the time, in seconds, required to complete the game



Example taken from Kuiper and Sklar (2013).

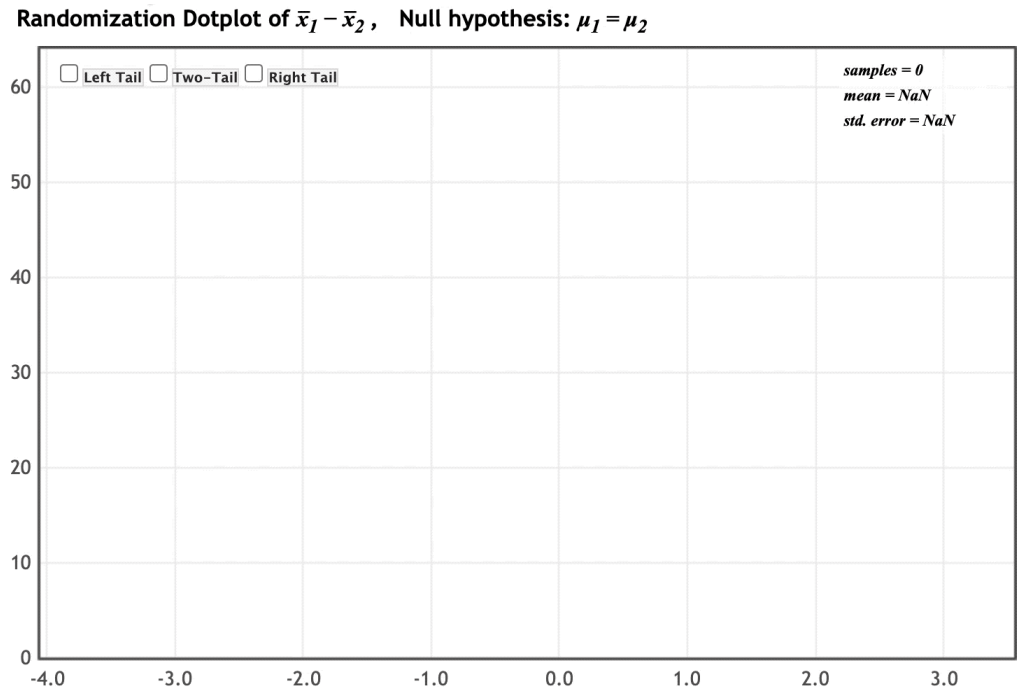
Initial activity

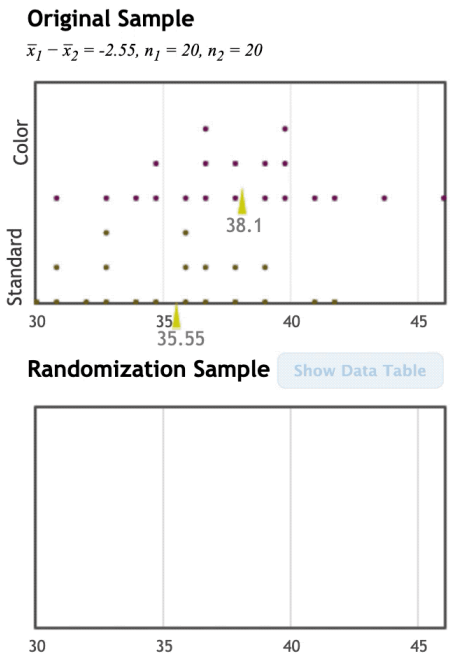
- What competing claims are being investigated in this study?
- What do the sample data have to say?



- What evidence does the observed plot provide?

The gap between apps and understanding





- Look at a few resamples
- Build up a distribution that describes behavior of the statistic

Using a lineup

Choose which plot is most different from the others and justify your choice

Using a lineup

Choose which plot is most different from the others and justify your choice

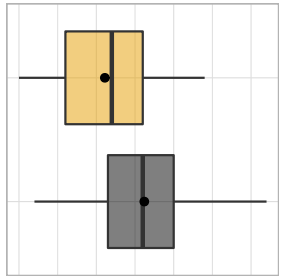
What did we just do?

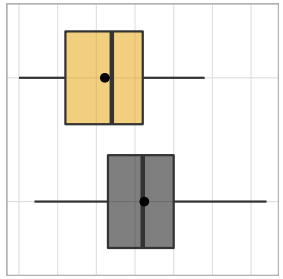
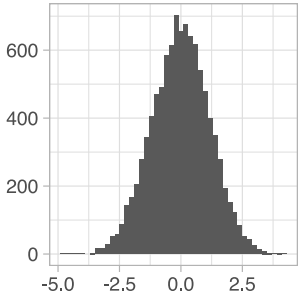
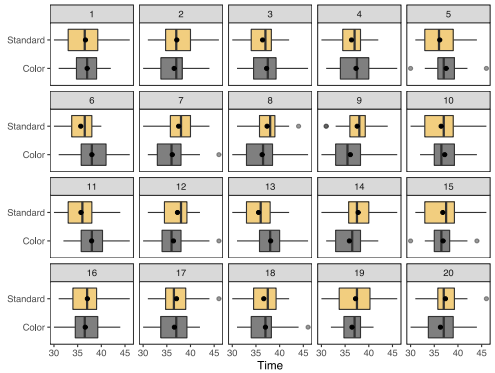
We compared the **data plot** with **null plots** of samples where, by construction, there is no association

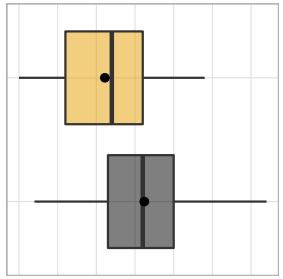
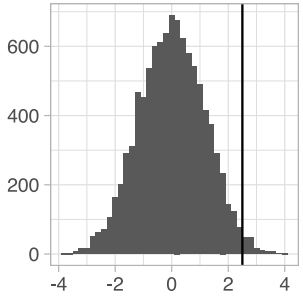
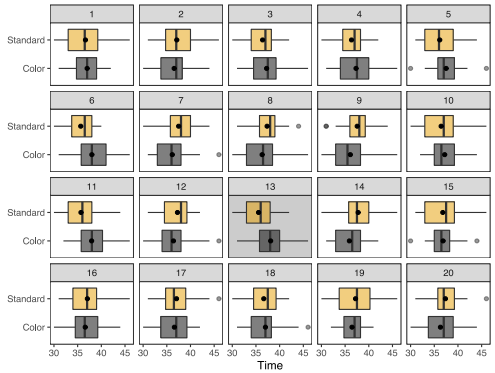
This forces us to make decisions by comparing what we observe to what we would expect under the null

All of this is done using "Sesame Street logic"

	Simulation-based Inference	**Visual Inference**
Hypotheses	H_0: equal means	H_a: larger mean for color distractor
Test statistic		
Reference distribution		
Evidence against H_0 if...		

	Simulation-based Inference	**Visual Inference**
Hypotheses	H_0: equal means	H_a: larger mean for color distractor
Test statistic	$T(x) = \overline{x}_1 - \overline{x}_2$	
Reference distribution		
Evidence against H_0 if...		

	Simulation-based Inference	Visual Inference
Hypotheses	H_0: equal means	H_a: larger mean for color distractor
Test statistic	$T(x) = \overline{x}_1 - \overline{x}_2$	
Reference distribution		
Evidence against H_0 if...		

	Simulation-based Inference	Visual Inference
Hypotheses	H_0: equal means	H_a: larger mean for color distractor
Test statistic	$T(x) = \overline{x}_1 - \overline{x}_2$	
Reference distribution		
Evidence against H_0 if...	the test statistic is "extreme"	the data plot is identifiable

Where else is the lineup protocol useful?

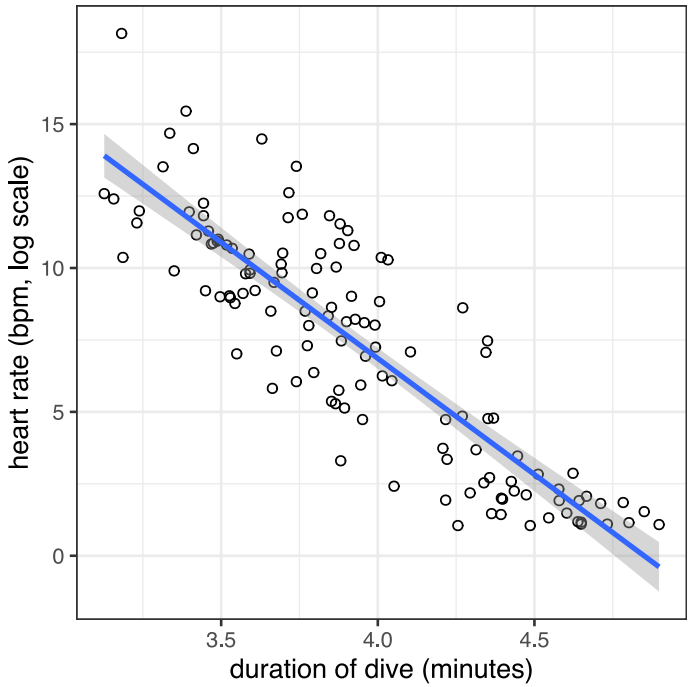
apophenia

the tendency to perceive a connection or meaningful pattern between unrelated or random things (such as objects or ideas)

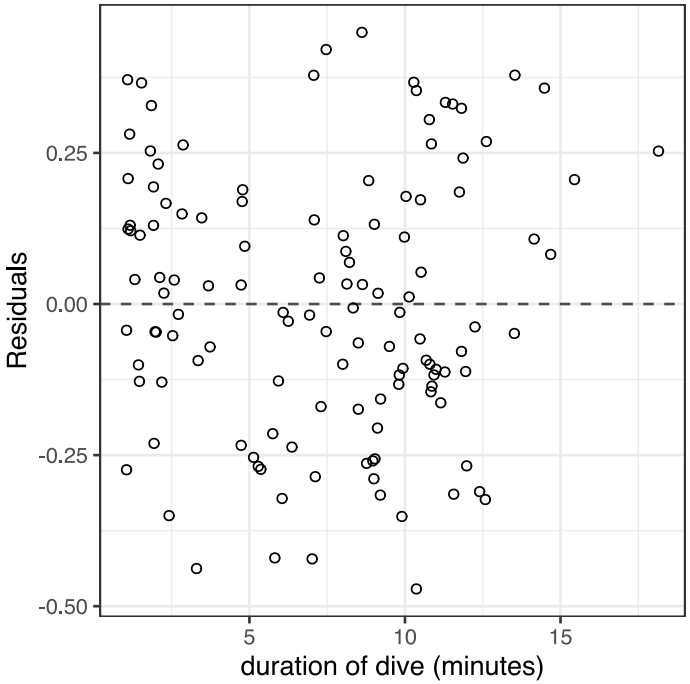
"apophenia" Meriam-Webster Dictionary Online, September 2019, merriam-webster.com

Interpreting residual plots

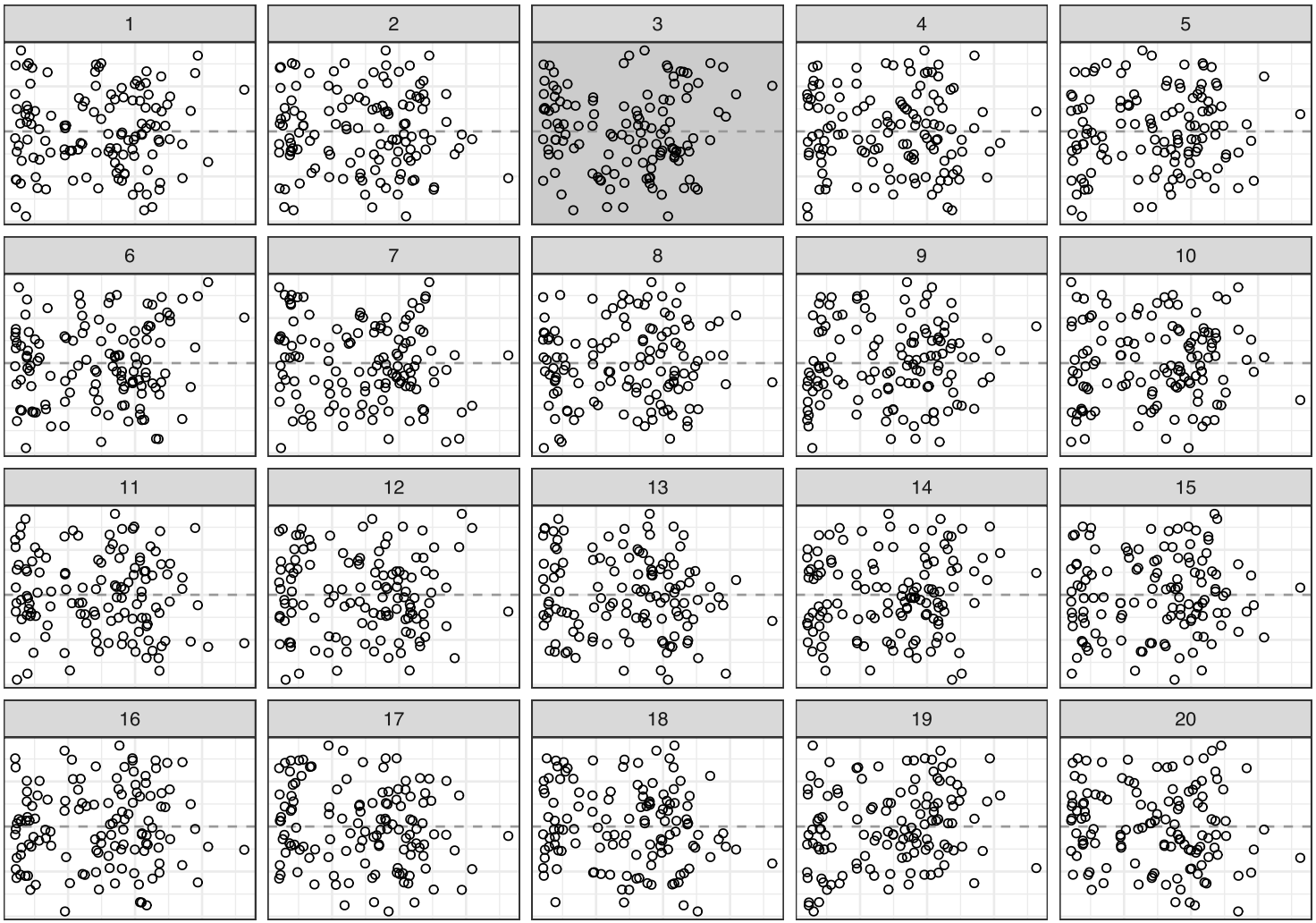
$$\widehat{\texttt{heart.rate}} = b_0 + b_1 \cdot \texttt{duration}$$



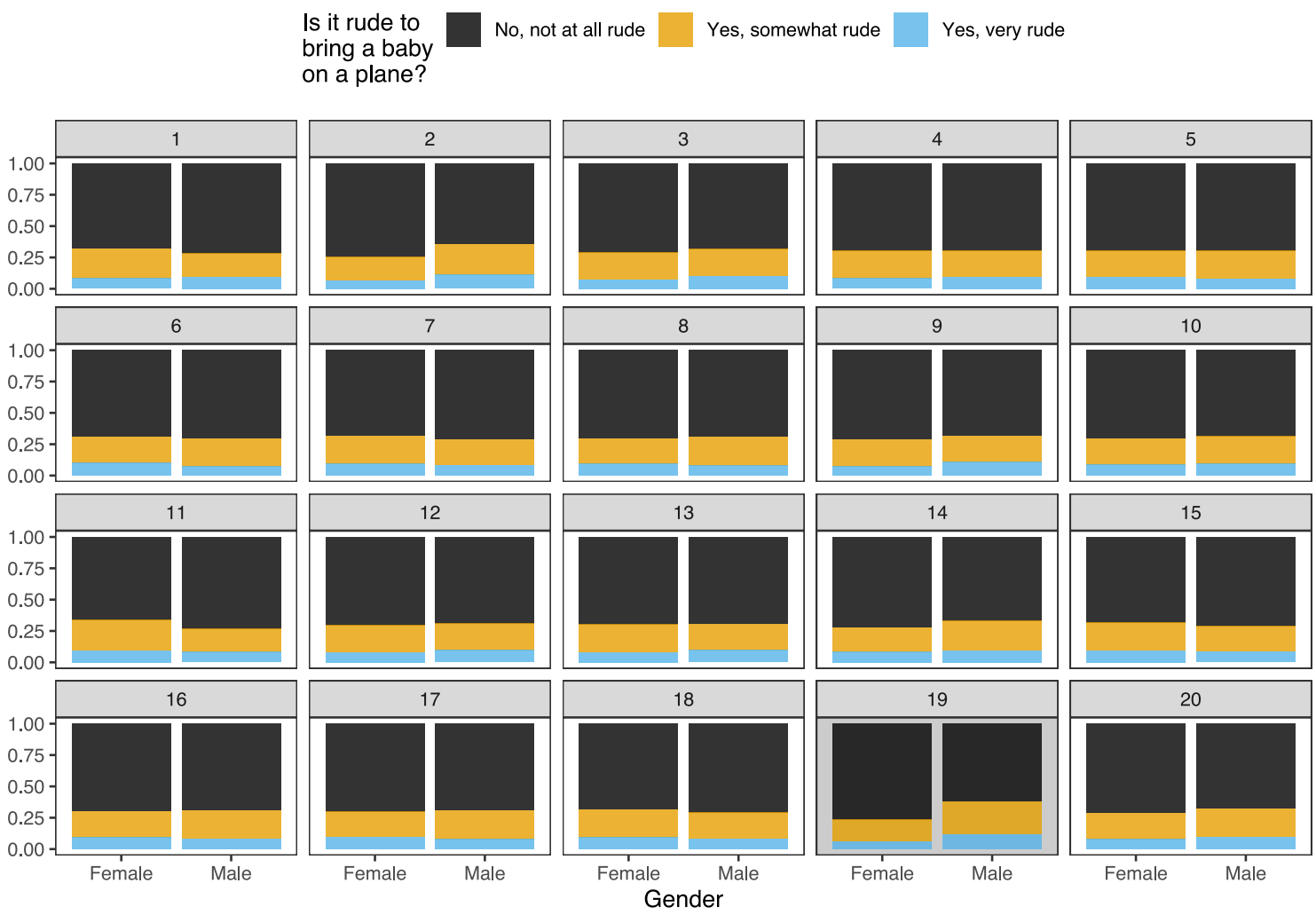
Is there any evidence of structure?



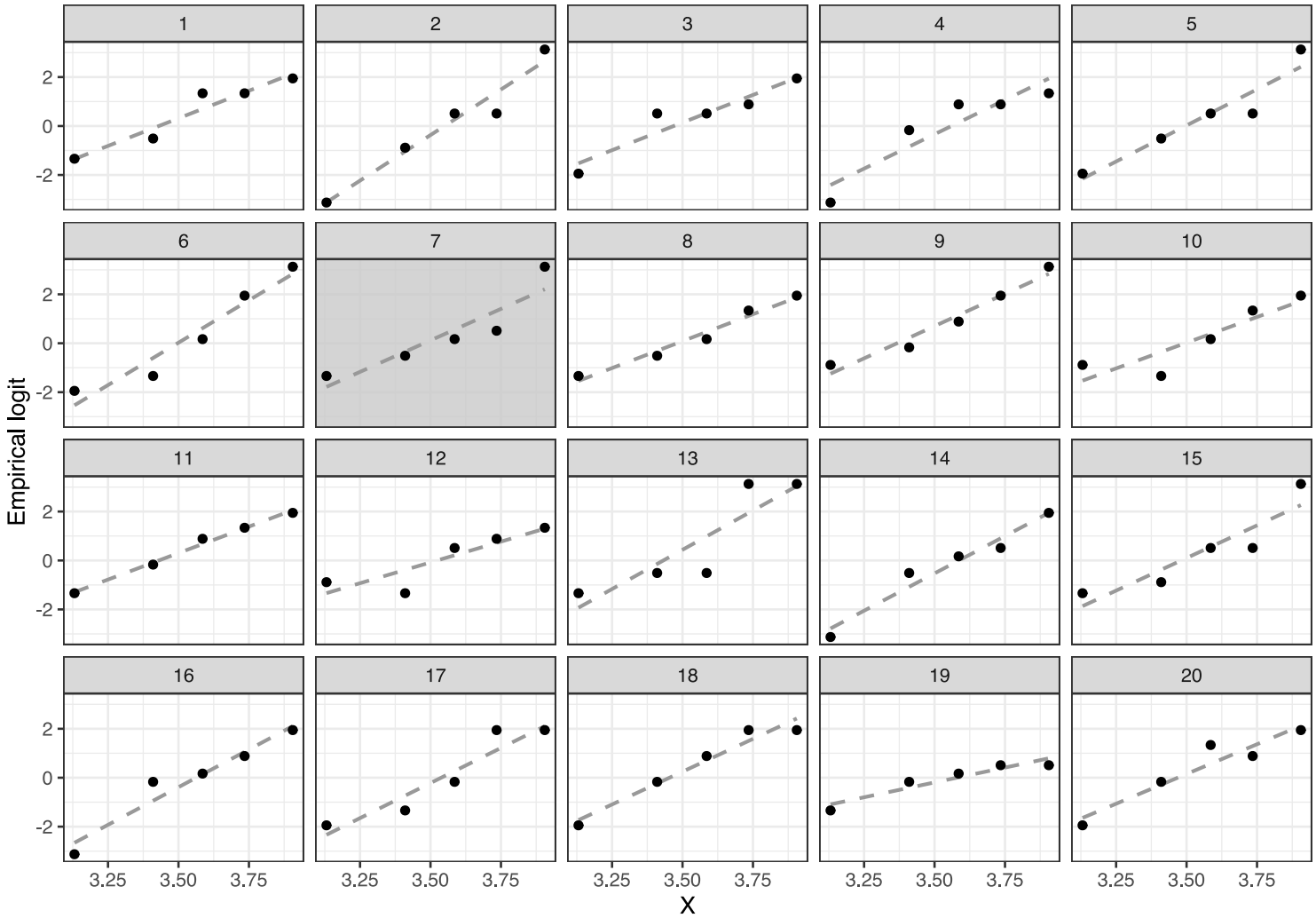
Does the observed residual plot stand out?



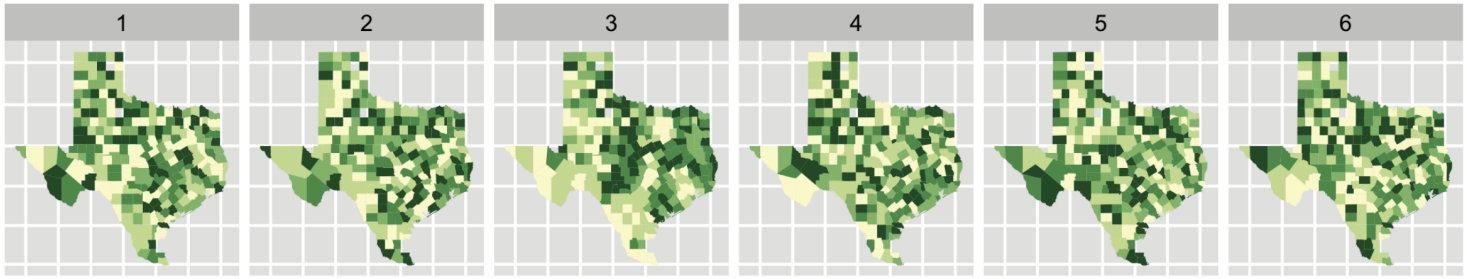
Is it rude to bring a baby on a plane?



Are the empirical odds linear?



Is there spatial association in this chloropleth map?



Wickham et al (2010)

How can I create lineups?

- Suite of Shiny apps: https://github.com/aloy/shiny-vizinf
- In R: `nullabor` + `tidyverse` tools

  Tutorial: https://aloy.github.io/classroom-vizinf/

```
library(nullabor)
library(tidyverse)
stroop %>%
  lineup(method = null_permute("Type"), true = ., n = 20) %>%
  ggplot(aes(x = Type, y = Time, fill = Type)) +
  geom_boxplot(alpha = 0.5) +
  coord_flip() +
  facet_wrap(~ .sample, ncol = 5) +
  scale_fill_colorblind()
```

```
decrypt("vkyo RpNp l2 6UtlNlU2 uA")
```

StatTLC blogpost: https://stattlc.com/2019/10/30/intro-visual-inference/

Conclusions

- Lineup introduces students to logic behind testing without need for technical discussions
- Lineup provides a framework to help students interpret new statistical graphics
- Lineup is a rigorous tool for statistical investigation later in the curriculum

References: Simulation-based curricula

- Cobb (2007) The introductory statistics course: a ptolemaic curriculum? *TISE*
- Cobb (2011) Teaching statistics: Some important tensions, *Chilean J. Stat.*
- Chihara & Hesterberg (2011) *Mathematical Statistics with Resampling and R*, Wiley
- Tintle et al. (2014) Quantitative evidence for the use of simulation and randomization in the introductory statistics course, *ICOTS9*
- Maurer & Lock (2014) Comparison of learning outcomes for randomization-based and traditional inference curricula in a designed educational experiment, *TISE*
- Tintle et al. (2015) Combating Anti-Statistical thinking using simulation-based methods throughout the undergraduate curriculum, *TAS*
- Hildreth at al. (2018) Comparing student success and understanding in introductory statistics under consensus and Simulation-Based curricula, *SERJ*

References: Visual inference

- Buja et al (2009) Statistical Inference for Exploratory Data Analysis and Model Diagnostics, *Roy. Soc. Ph. Tr., A*
- Majumder et al (2013) Validation of Visual Statistical Inference, Applied to Linear Models, *JASA*
- Wickham et al (2010) Graphical Inference for Infovis, *InfoVis*
- Hofmann et al (2012) Graphical Tests for Power Comparison of Competing Design, *InfoVis*
- Loy et al (2017) Model Choice and Diagnostics for Linear Mixed-Effects Models Using Statistics on Street Corners, *JCGS*