Bringing Visual Inference to the Classroom

Adam Loy, Carleton College

Overview

Recent developments in graphics methodology provide us with a rigorous framework for graphical discovery: Visual inference (Buja et al., 2009) allows for the quantification of the significance of graphical findings. This framework can be used in the classroom to:

1. visualize permutation tests, providing visuals to augment class discussion, and
2. build student intuition when learning how to interpret new statistical graphics.

The lineup protocol

The lineup protocol was developed to evaluate and quantify the significance of graphical findings.

Key steps:

1. Formulate competing hypotheses
2. Create a **plot of the observed data** relevant to these hypotheses
3. Create a distribution of **null (typical) plots** by randomly generating data under the null hypothesis
4. **Randomly situate the observed plot** in a sample of null plots (usually 19) drawn from the reference distribution. **If the observed plot is identifiable, that is evidence it is "extreme."**

Parallel: Simulation-based inference

The lineup protocol parallels the ideas laid out by Cobb (2007) that form the core of the randomization-based introductory statistics curriculum:

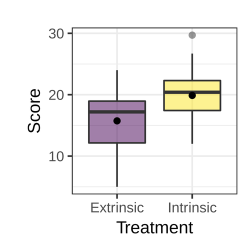
1. Formulate competing hypotheses
2. Calculate a test statistic that summarizes the information relevant to these hypotheses.
3. Create a distribution of "typical test statistics" by simulating data sets under the null hypothesis and calculating the test statistic.
4. If the observed test statistic is in the tail of this distribution, reject the null hypothesis.

Where are lineups useful in the curriculum?

Lineups visualize simulated data, so they are great tools to visualize first permutation tests

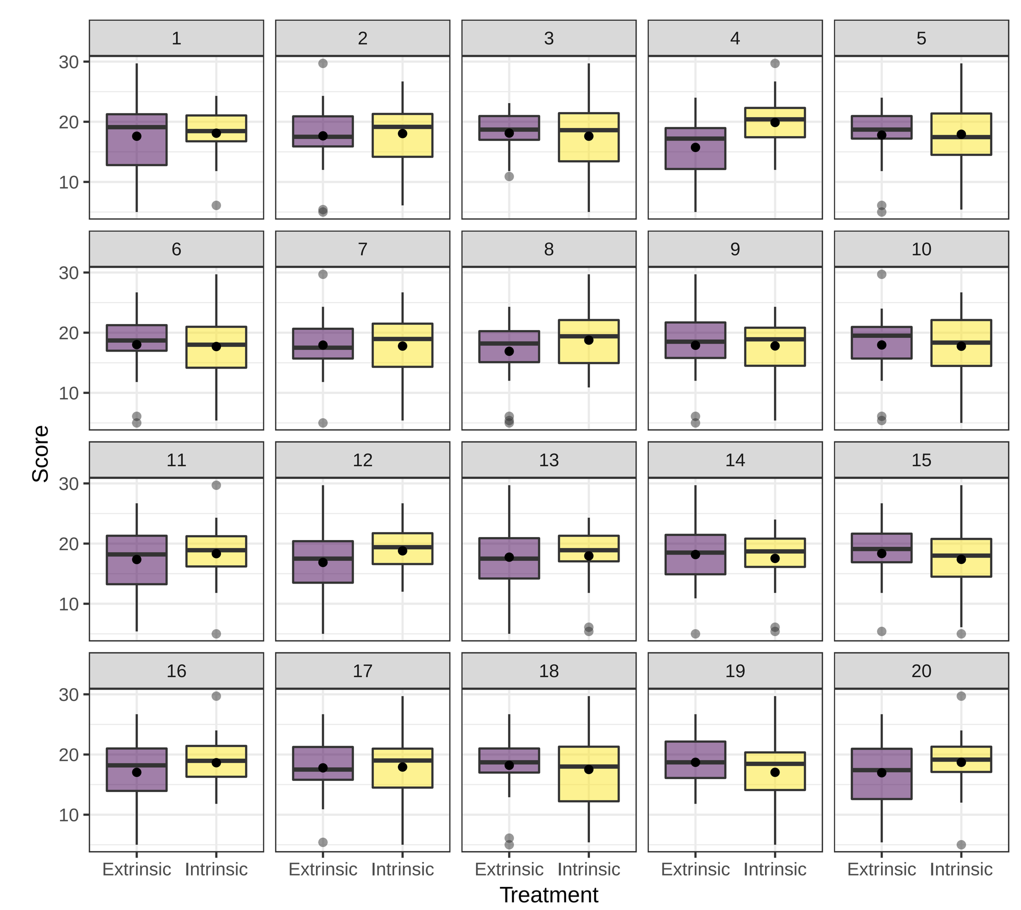
Lineups help students calibrate their intuition, so they are great tools to introduce visual model diagnostics

Introducing hypothesis tests

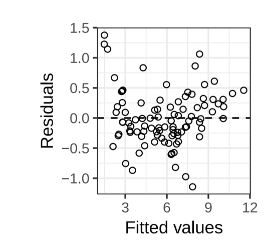


Question: Do these boxplots provide evidence of a significant treatment effect?

Answer: Let's compare the plot to what we would expect if there was in fact no treatment effect!

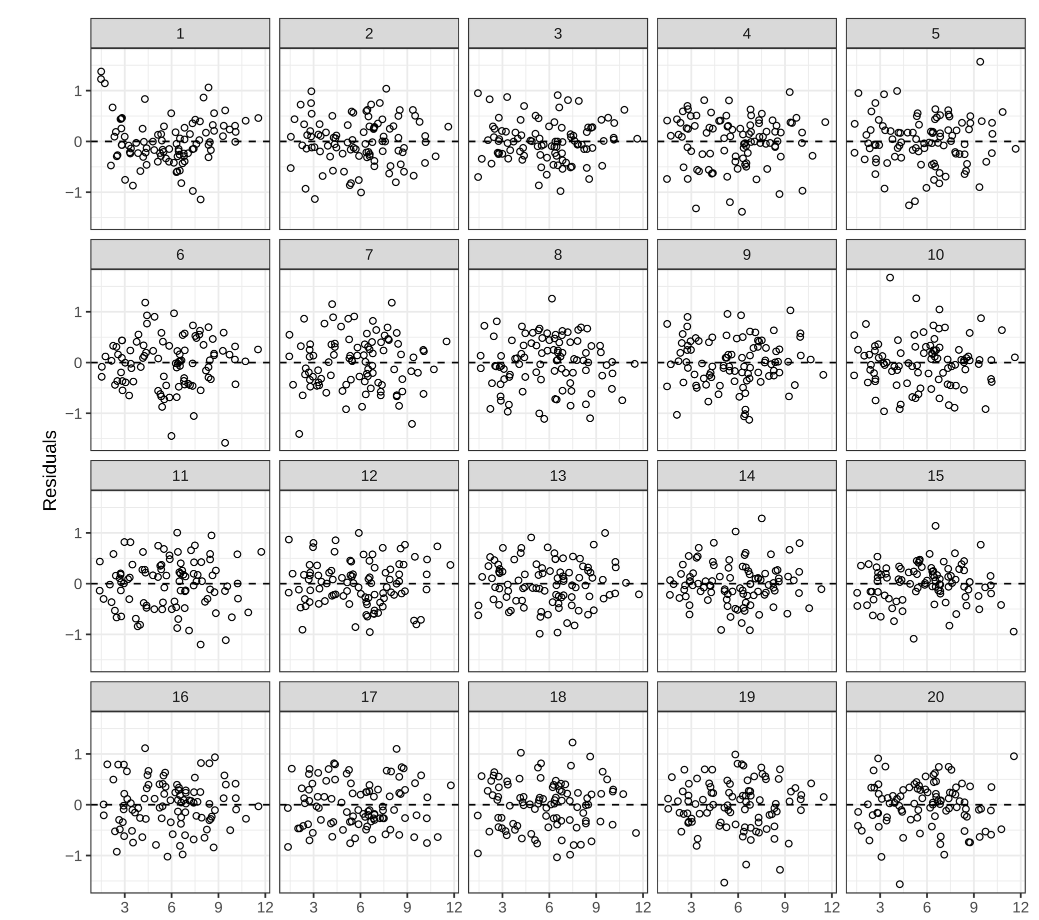


Diagnosing linear regression

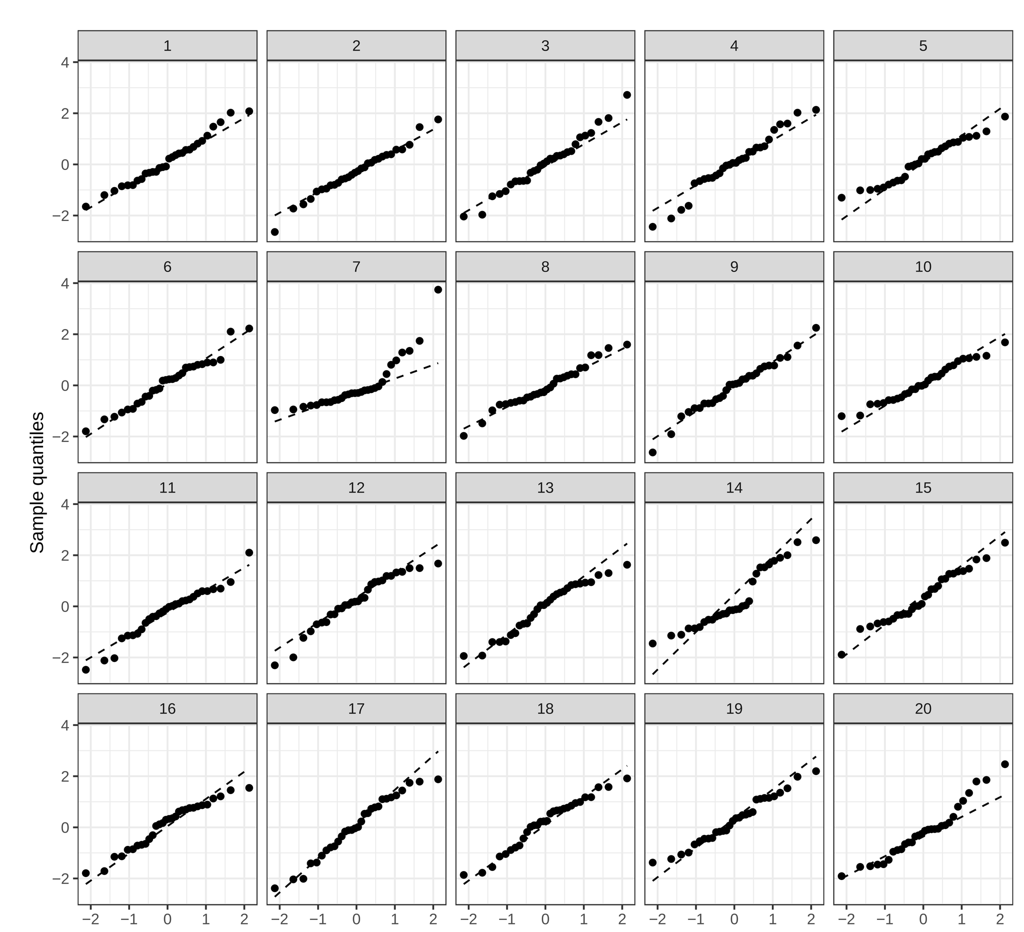


Question: Does this residual plot provide evidence of a model deficiency?

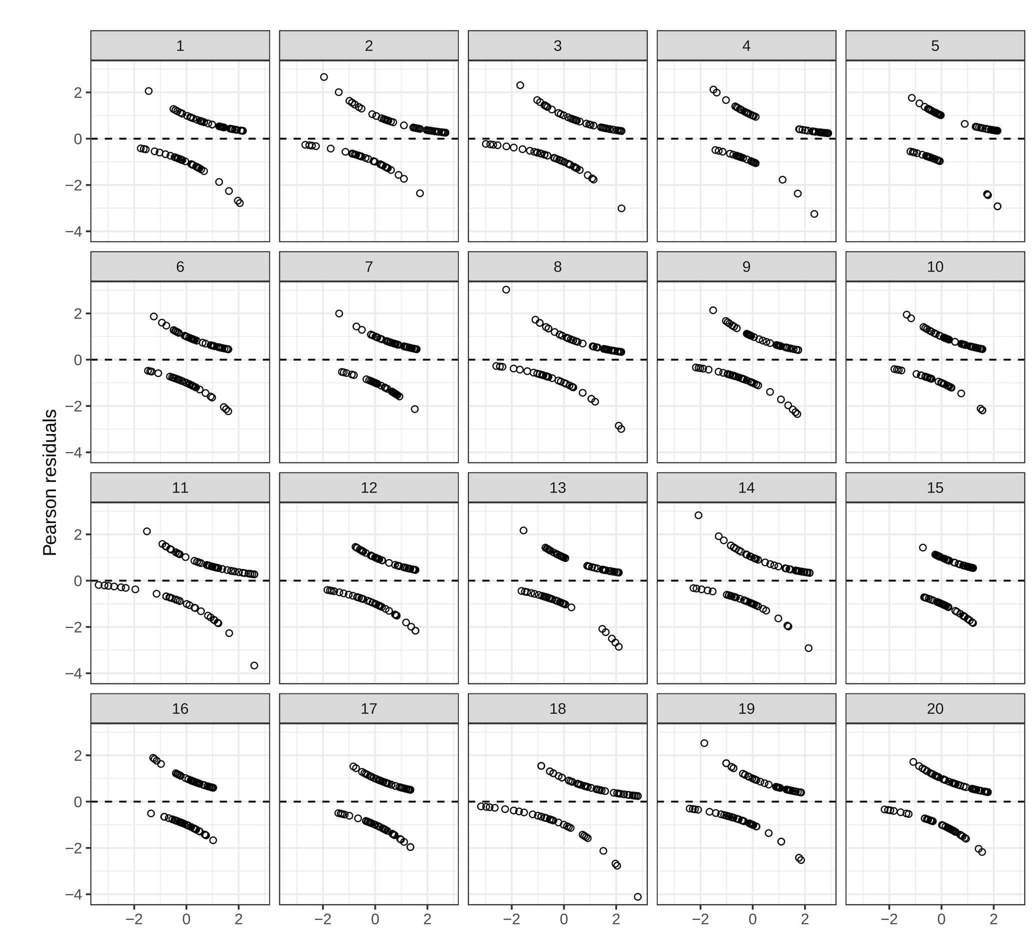
Answer: Let's compare the residual plot to what we would expect if the model was valid!



Reading Q-Q plots



Diagnosing logistic regression



Creating lineups

1. Create a plot of the observed data
2. Generate 19 sets of null plots
   - by permuting treatment labels
   - by simulating data from your regression model
   - by simulating data from a specified distribution
3. Randomly assign ID numbers to all 20 plots (19 nulls + 1 data), keeping track of the ID for the data plot
4. Facet the plots into a sensible grid

The lineups presented here were all rendered in R using `ggplot2`. The code used to generate these lineups is available on GitHub.

Contact Information

Adam Loy:

aloy.rbind.io
aloy@carleton.edu
@aloy

Get the code!

github.com/aloy/uscots2019

References

Buja, A., Cook, D., Hofmann, H., Lawrence, M., Lee, E. K., Swayne, D. F., and Wickham, H. (2009). Statistical inference for exploratory data analysis and model diagnostics. *Philosophical Transactions of the Royal Society A*, 367(1906):4361–4383.

Cobb, G. W. (2007). The introductory statistics course: a ptolemaic curriculum? *Technology Innovations in Statistics Education*, 1.

Where are the data plots?

1. Boxplots plots: #4
2. Residuals from linear regression: #1
3. Q-Q plots: #7
4. Residuals from logistics regression: #10