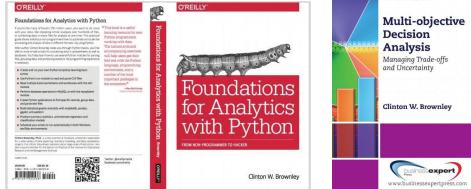
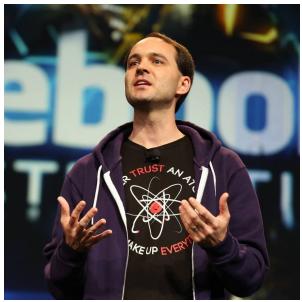


https://github.com/cbrownley/2024ODSC_AIPLUS

Visualization in Bayesian workflow

Clinton Brownley

About Me



Overview

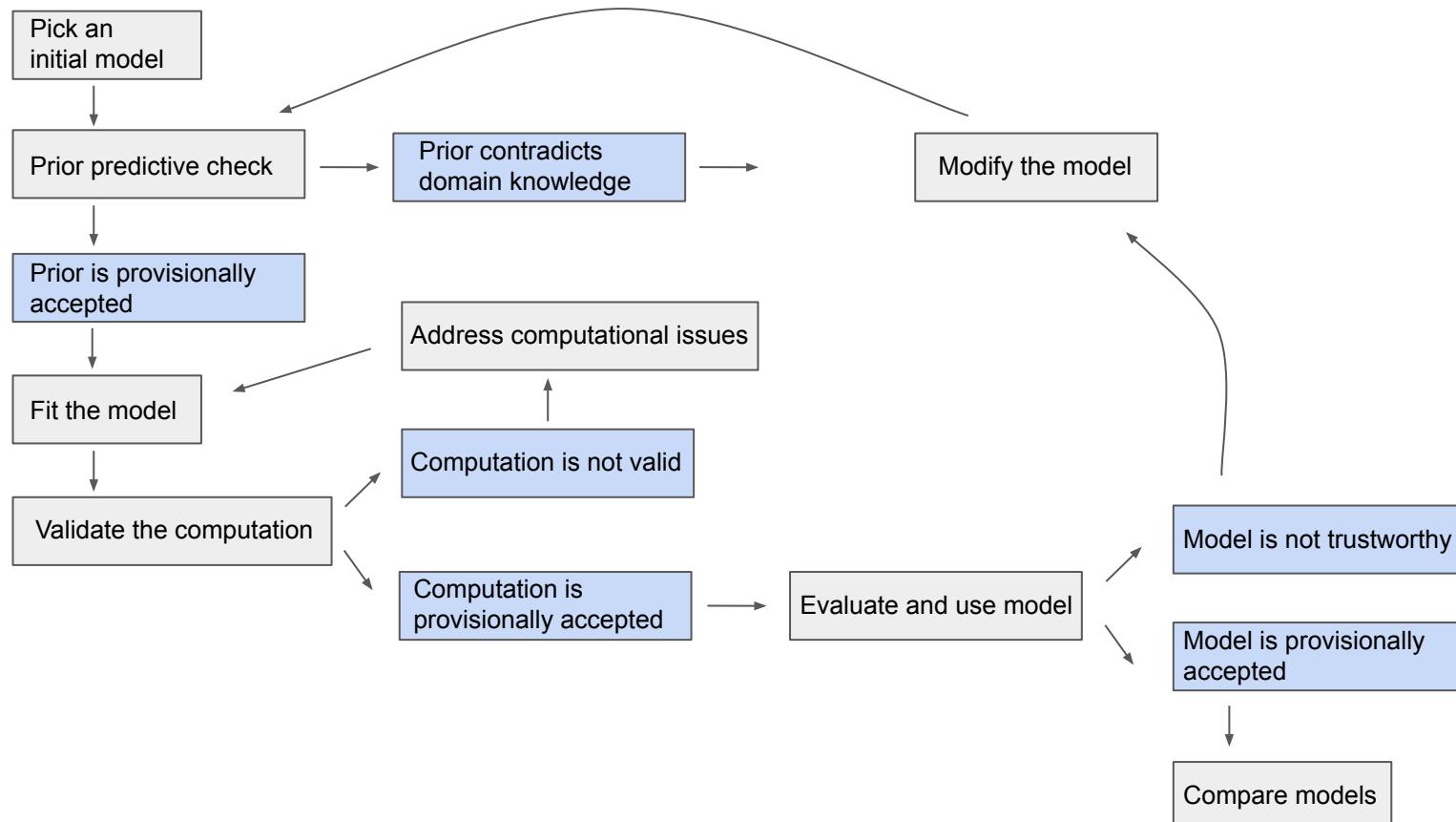
- Research questions and Bayesian workflow
- Audience
- Foundational visualization concepts
- Visualization guidelines and best practices
- Layered grammar of graphics
- Storytelling
- Visualization in Bayesian workflow

Research questions and Bayesian workflow

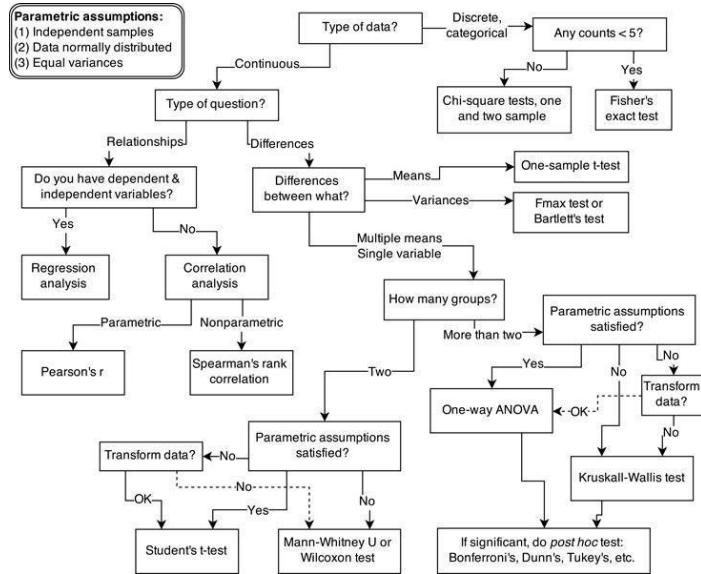
Research Questions

- **What** is your thesis -- your research question?
- **Why** is it important?
- **How** can you communicate your message effectively?
- **Who** is your audience?

What is Bayesian workflow?

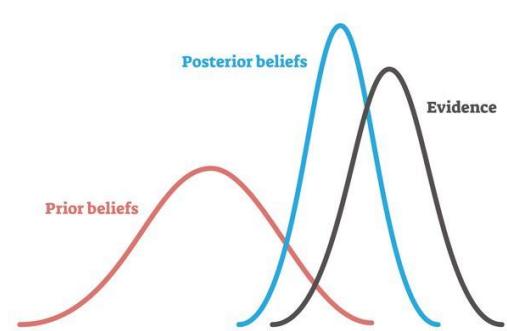


Bayesian workflow propels us from inflexible, fragile pre-made tools to a unified theory of statistical analysis



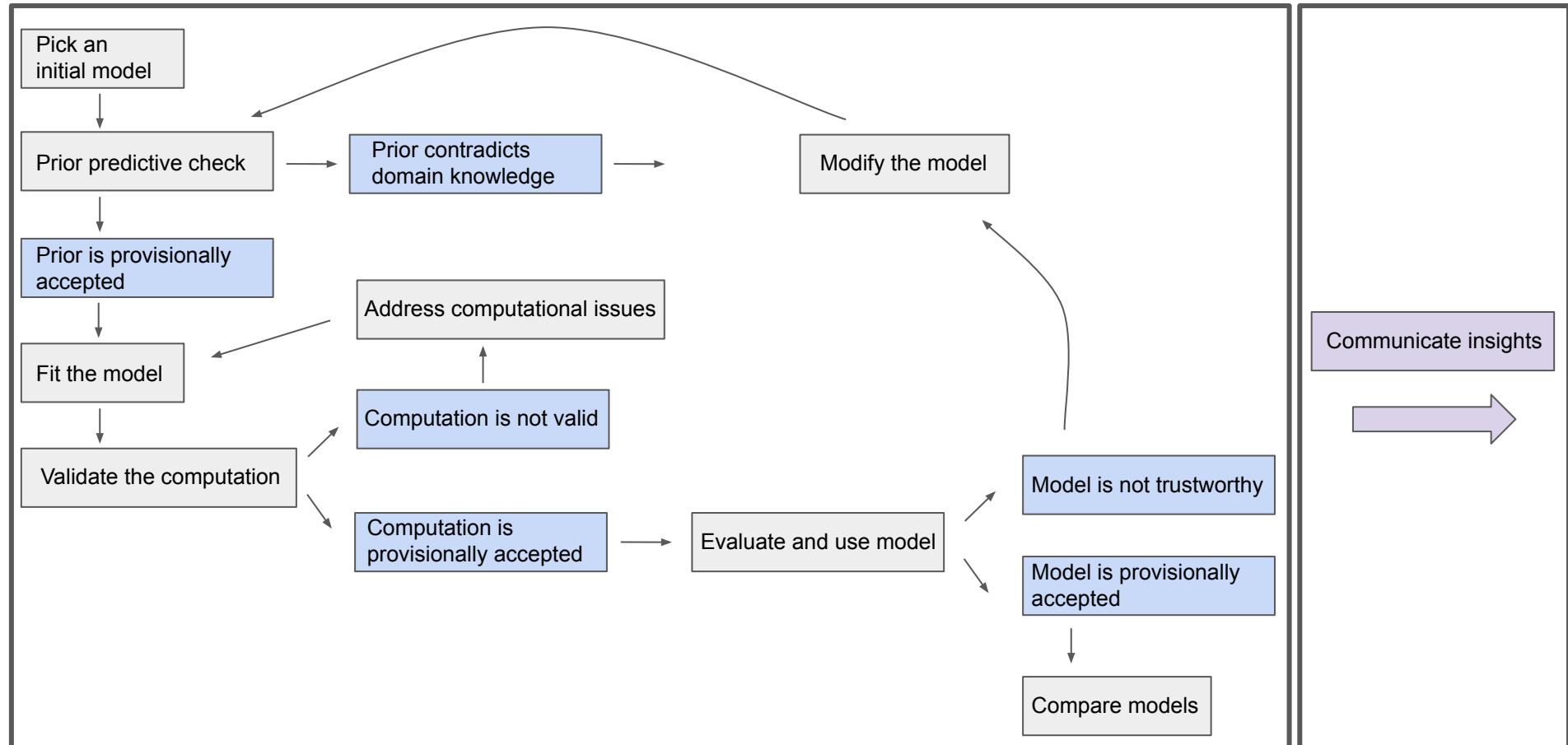
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

BAYESIAN ANALYSIS



Exploration

Explanation

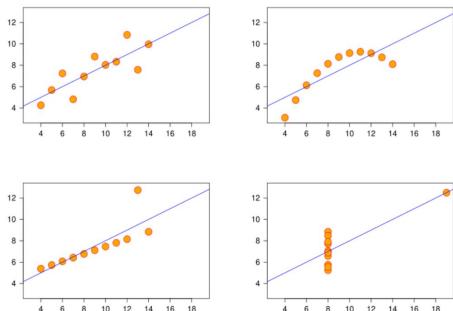


Visualization is important for both **exploration** and **explanation**

Exploration

Visualization helps you *discover* insights that might otherwise remain hidden

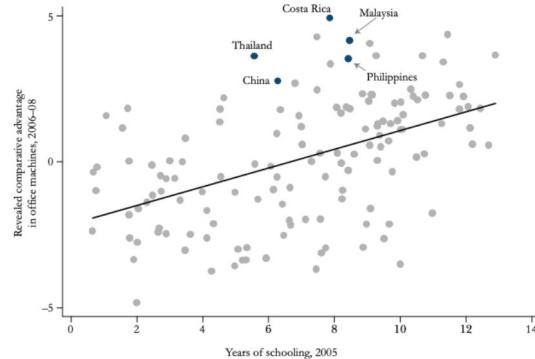
I		II		III		IV		
x	y	x	y	x	y	x	y	
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58	
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76	
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71	
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84	
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47	
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04	
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25	
4.0	4.28	4.0	3.10	4.0	5.39	19.0	12.50	
12.0	10.84	12.0	9.31	12.0	8.15	8.0	5.56	
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91	
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89	
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Var.	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Corr.	0.816	0.816	0.816	0.816	0.816	0.816	0.816	0.816



Explanation

Visualization helps you *present* insights clearly and effectively

Education and Exports of Office Machines



What are some **key qualities** of an effective visualization?

- **Truthful**
 - is based on thorough and honest research
- **Functional**
 - constitutes an accurate depiction of the data, and enables people to accomplish intended tasks
- **Beautiful**
 - is attractive, intriguing, and even aesthetically pleasing
- **Insightful**
 - reveals evidence that would be hard to see otherwise
- **Enlightening**
 - changes our minds for the better, if we grasp and accept the evidence it depicts

What knowledge helps us create effective visualizations?

- Audience
 - Who is your audience?
 - What do you want them to know or do?
 - How can you communicate your message effectively?
 - What does a success outcome look like?

What knowledge helps us create effective visualizations?

- Audience
- Foundational visualization concepts: marks, channels, and channel rankings
 - Marks: basic graphical elements
 - Channels: ways to control the appearance of marks
 - Channel rankings: how effectively channels express the information in the data

What knowledge helps us create effective visualizations?

- Audience
- Foundational visualization concepts: marks, channels, and channel rankings
- Visualization guidelines and best practices
 - Choose effective graphic forms
 - Create visual hierarchy
 - Direct attention
 - Use color strategically
 - Remove clutter
 - Facilitate comparisons
 - Annotate with text

What knowledge helps us create effective visualizations?

- Audience
- Foundational visualization concepts: marks, channels, and channel rankings
- Visualization guidelines and best practices
- Layered grammar of graphics
 - Data and aesthetic mappings
 - One or more layers, each with geometric mark, statistical transformation, and position adjustment
 - Scale for each aesthetic mapping
 - Coordinate system
 - Facet specification

What knowledge helps us create effective visualizations?

- Audience
- Foundational visualization concepts: marks, channels, and channel rankings
- Visualization guidelines and best practices
- Layered grammar of graphics
- Storytelling
 - Big Idea and SPAR
 - Bing, Bang, Bongo
 - Horizontal and vertical logic
 - Storyboarding

Audience

Audience

- **Who** is your audience?
- **What** do you want them to know or do?
- **How** can you use data and visualizations to make your point effectively?
- **What** does a success outcome look like?

Foundational visualization concepts

Foundational visualization concepts: *marks, channels, and channel rankings*

Marks

A **mark** is a basic geometric element depicting an item or link, classified by the number of spatial dimensions it requires

→ Points



→ Lines



→ Areas

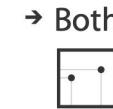
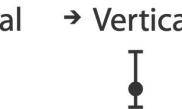
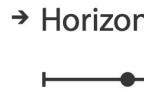


Foundational visualization concepts: *marks, channels, and channel rankings*

Channels

A visual **channel** is a way to control the appearance of marks, independent of the dimensionality of the marks

⇒ Position



⇒ Color



⇒ Shape



⇒ Tilt



⇒ Size

→ Length



→ Area



→ Volume



Foundational visualization concepts: *marks, channels, and channel rankings*

Channel Rankings

The **effectiveness** of a channel for encoding data depends on its type:

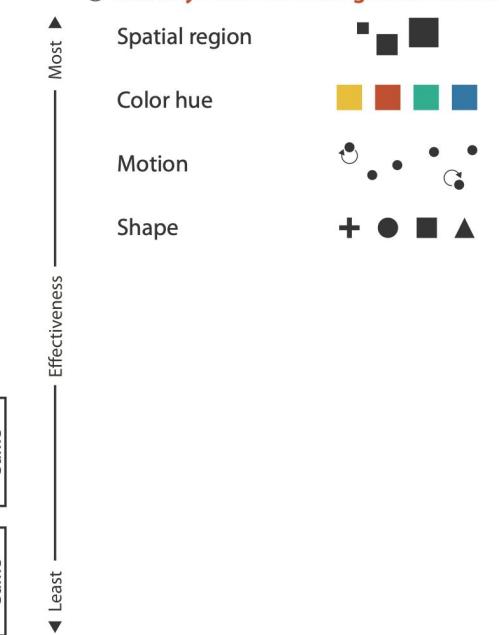
- channels that perceptually convey *magnitude* are good for *ordered* data
- channels that convey *identity* are good for *categorical* data

Channels: Expressiveness Types and Effectiveness Ranks

⊕ Magnitude Channels: Ordered Attributes

Position on common scale	
Position on unaligned scale	
Length (1D size)	
Tilt/angle	
Area (2D size)	
Depth (3D position)	

⊕ Identity Channels: Categorical Attributes



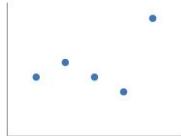
Visualization guidelines and best practices

Visualization guidelines and best practices:

Choose effective graphic forms

91%

Simple text



Scatterplot

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

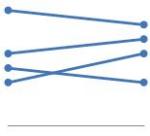
Table



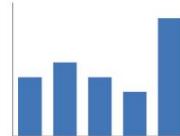
Line

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Heatmap



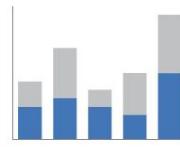
Slopegraph



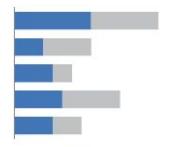
Vertical bar



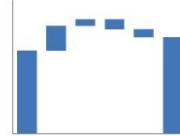
Horizontal bar



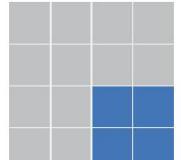
Stacked vertical bar



Stacked horizontal bar

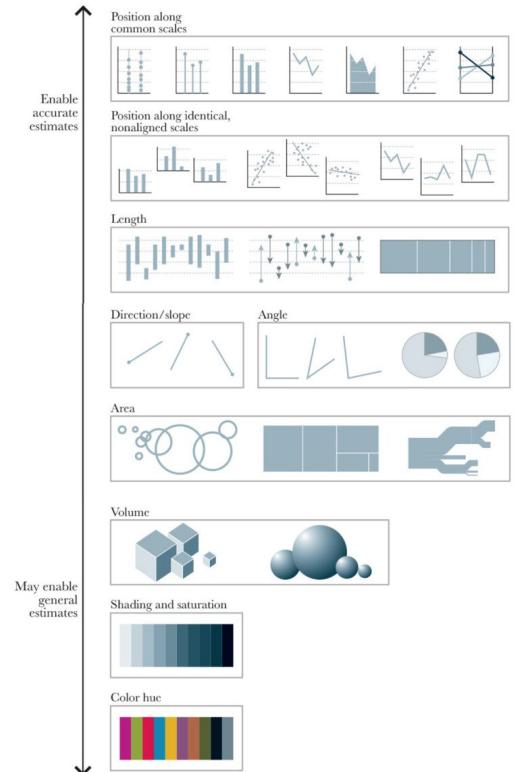


Waterfall



Square area

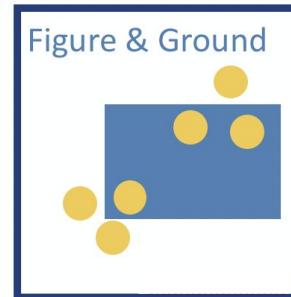
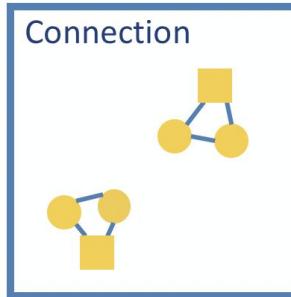
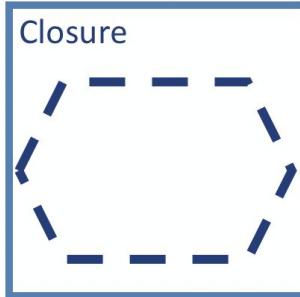
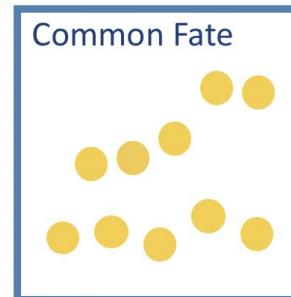
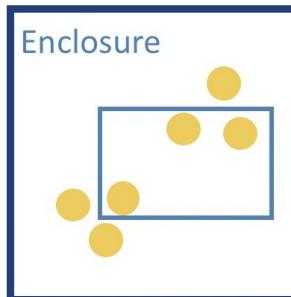
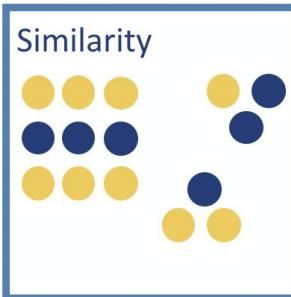
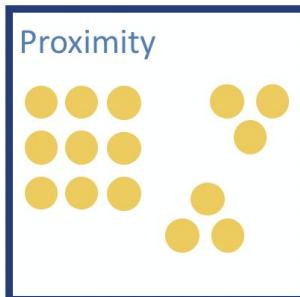
Alberto Cairo's version of hierarchy of elementary perceptual tasks



Visualization guidelines and best practices:

Create visual hierarchy

Gestalt Principles of Visual Perception



Visualization guidelines and best practices:

Direct attention with pre-attentive attributes



Orientation Length Width Size



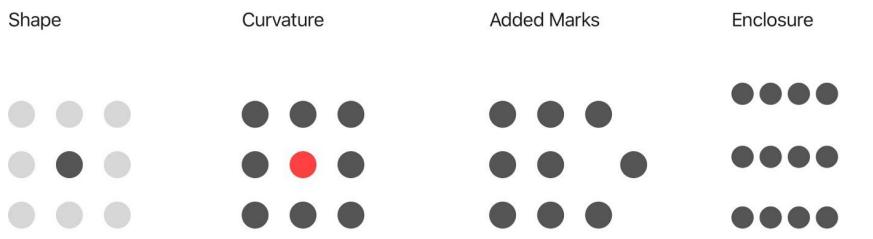
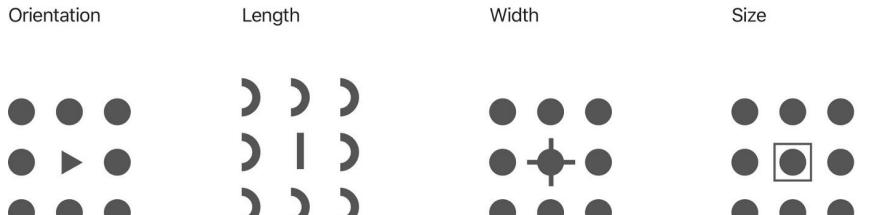
Shape Curvature Added Marks Enclosure



Contrast Colour Position Spatial Grouping

Visualization guidelines and best practices:

Direct attention with pre-attentive attributes



No preattentive attributes

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.
You have a great company – keep up the good work!

Color

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.
You have a great company – keep up the good work!

Size

What are we doing well? Great Products. These products are the best in their class. Replacement parts are shipped when needed. You sent gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.
You have a great company – keep up the good work!

Outline (enclosure)

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.
You have a great company – keep up the good work!

Bold

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.
You have a great company – keep up the good work!

Italics

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.
You have a great company – keep up the good work!

Separate spatially

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask.

Problems are resolved promptly

Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours. You have a great company – keep up the good work!

Underline (added marks)

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.
You have a great company – keep up the good work!

Visualization guidelines and best practices:

Use color thoughtfully to convey groupings or quantities

Hue

Color **hue** is an *identity* channel that visually conveys categories and groupings

Color hue



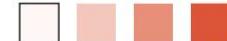
Luminance and Saturation

Color **luminance** and **saturation** are *magnitude* channels that we automatically perceive as ordered

Color luminance

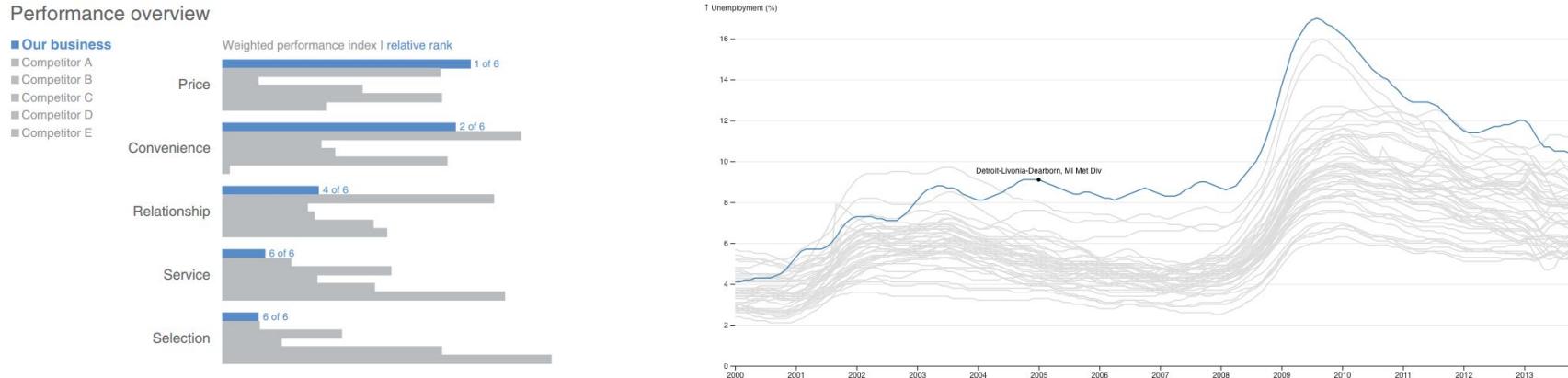


Color saturation



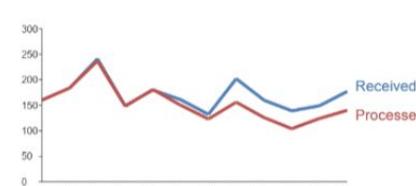
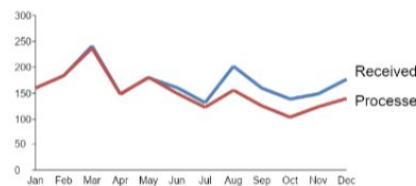
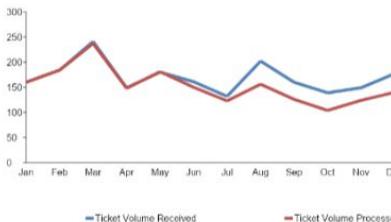
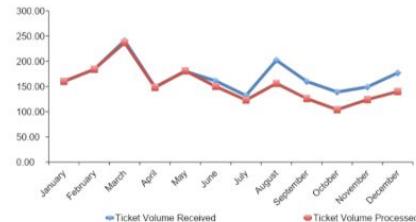
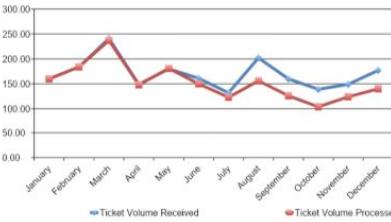
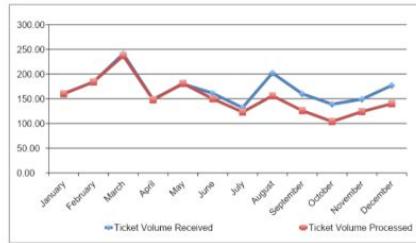
Visualization guidelines and best practices:

Use color thoughtfully to direct attention

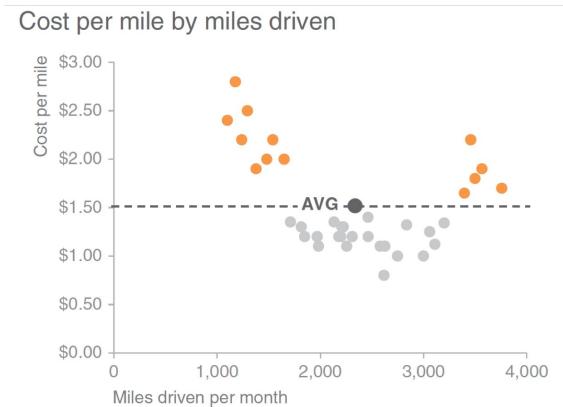
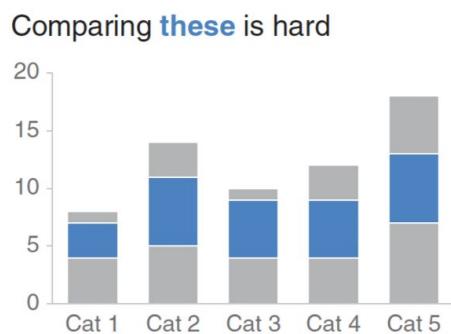
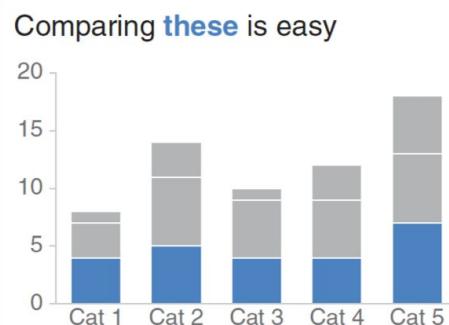


Visualization guidelines and best practices:

Remove clutter (enhance data ink, remove non-data ink)



Visualization guidelines and best practices: *Facilitate comparisons*

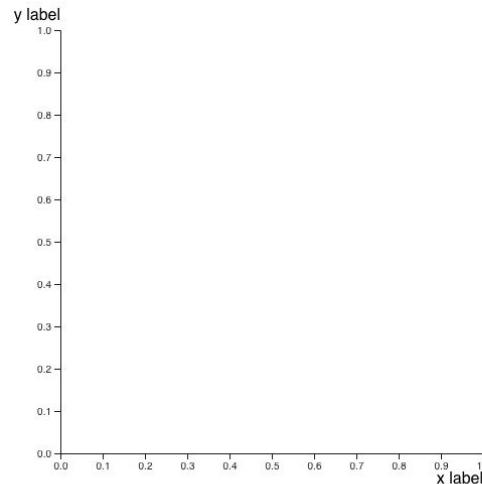


Visualization guidelines and best practices:

Annotate with text to explain and highlight

Header title that describes findings

Lead-in text is your chance to provide more details on what the data is about, where it's from, and what the audience should see or take away



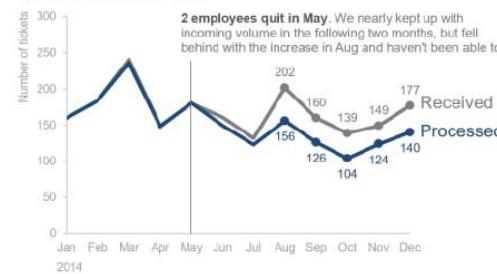
By Clinton Brownley

Source: [This is where the data is from](#)

Please approve the hire of 2 FTE

to backfill those who quit in the past year

Ticket volume over time



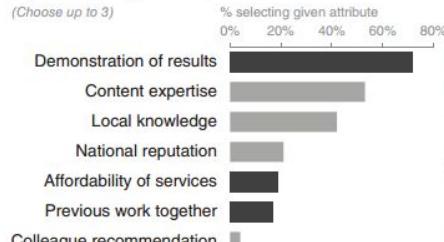
Data source: XY2 Dashboard, as of 12/31/2014 | A detailed analysis on tickets processed per person and time to resolve issues was undertaken to inform this request and can be provided if needed.

Visualization guidelines and best practices:

Annotate with text to explain and highlight

Demonstrating effectiveness is most important consideration when selecting a provider

In general, **what attributes are the most important** to you in selecting a service provider?



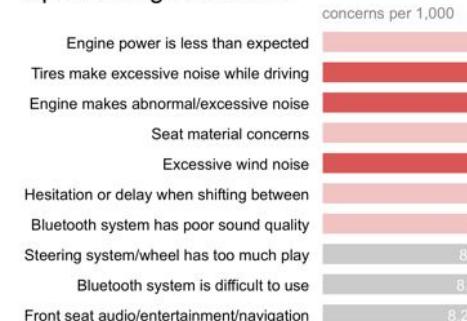
Data source: xyz; includes N number of survey respondents.
Note that respondents were able to choose up to 3 options.

Survey shows that **demonstration of results** is the single most important dimension when choosing a service provider.

Affordability and experience working together previously, which were hypothesized to be very important in the decision making process, were both cited less frequently as important attributes.

Of the top design concerns, three are noise related.

Top 10 Design Concerns



Comments indicate that **noisy tire issues** are most apparent in the rain.

Complaints about **engine noise** commonly cited the issue after the car **had not been driven for a while**.

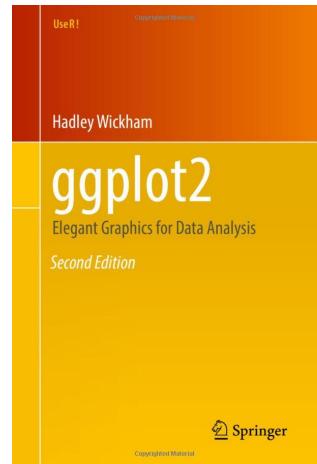
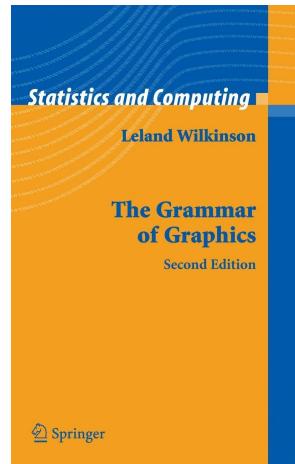
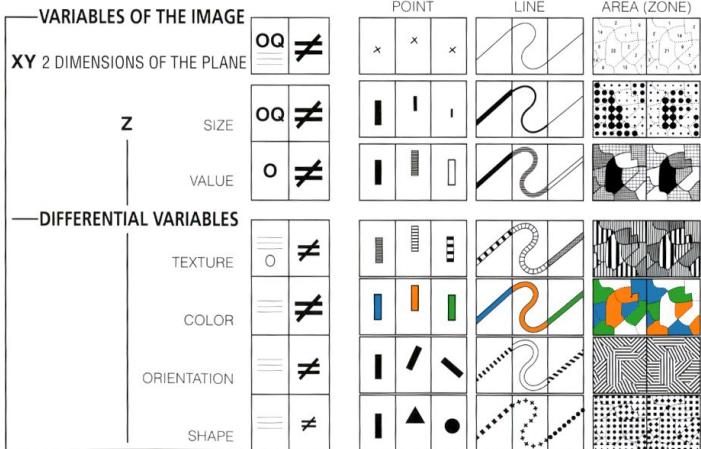
The **excessive wind noise** complaint is surprising given efforts already made to address this issue; we recommend **additional research be done here to better understand and remedy the issue**.

Layered grammar of graphics

Layered grammar of graphics

Development of the grammar

Semiology of graphics
by Jacques Bertin



Layered grammar of graphics

Components of the grammar

Components

The layered grammar defines a plot as the combination of:

- A default **dataset** and set of **mappings** from variables to aesthetics.
- One or more layers, each composed of a **geometric object**, a **statistical transformation**, a **position adjustment**, and optionally, a dataset and aesthetic mappings.
- One **scale** for each aesthetic mapping.
- A **coordinate system**.
- The **faceting specification**.

```
ggplot(data = <DATA>) +  
  <GEOM_FUNCTION>(  
    mapping = aes(<MAPPINGS>),  
    stat = <STAT>,  
    position = <POSITION>  
  ) +  
  <COORDINATE_FUNCTION> +  
  <FACET_FUNCTION>
```

Layered grammar of graphics

Benefits of a layered grammar

Benefits

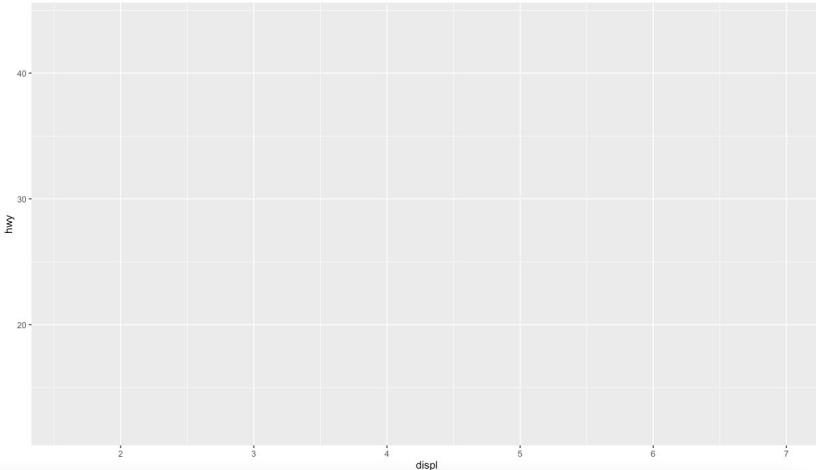
A layered grammar of graphics affords the following benefits:

- Enables us to concisely describe the components of a graphic.
- Allows us to gain insight into the composition of complicated graphics, and reveal unexpected connections between seemingly different graphics.
- Provides a strong foundation for understanding a diverse range of graphics.
- Helps guide us on what a well-formed graphic looks like – like the English language, good grammar is just the first step in creating a good sentence.

Layered grammar of graphics

Build a plot layer by layer

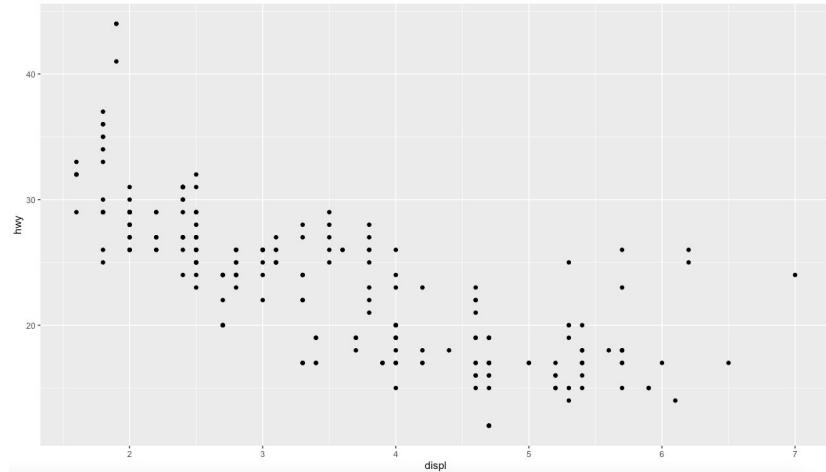
```
p <- ggplot(data = mpg, aes(x = displ, y = hwy))  
p
```



Layered grammar of graphics

Build a plot layer by layer

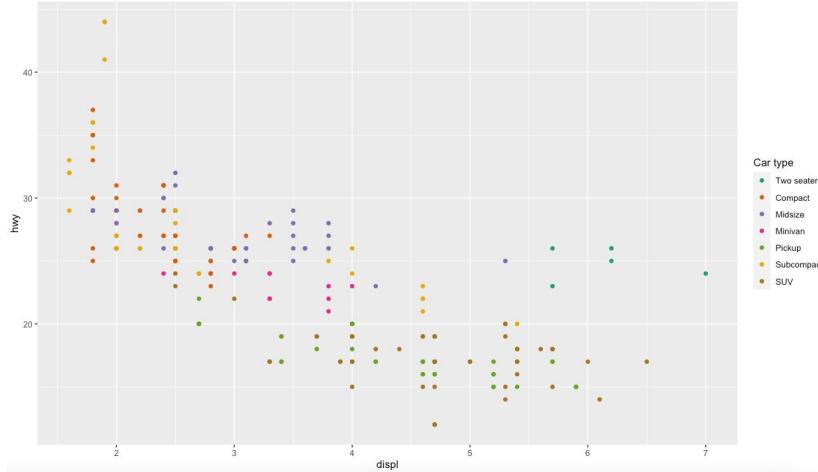
`p + geom_point()`



Layered grammar of graphics

Build a plot layer by layer

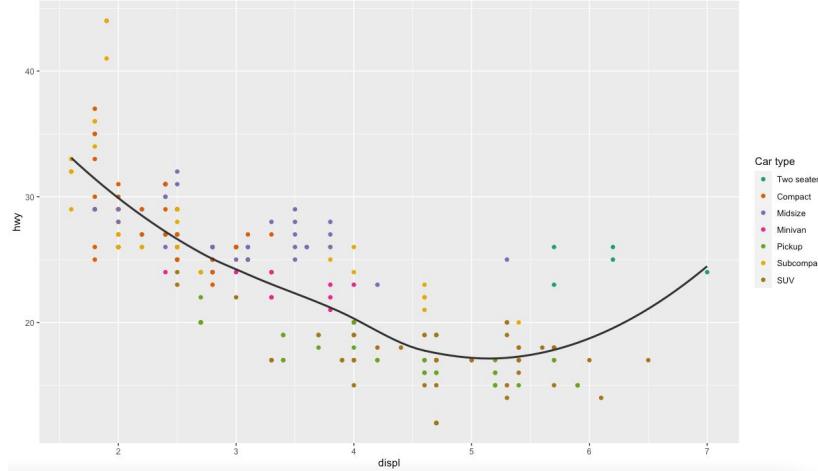
```
p +  
  geom_point(aes(color = class)) +  
  scale_color_brewer(labels = c("Two seater",  
    "Compact",  
    "Midsize",  
    "Minivan",  
    "Pickup",  
    "Subcompact",  
    "SUV"),  
    palette = "Dark2") +  
  labs(color = "Car type")
```



Layered grammar of graphics

Build a plot layer by layer

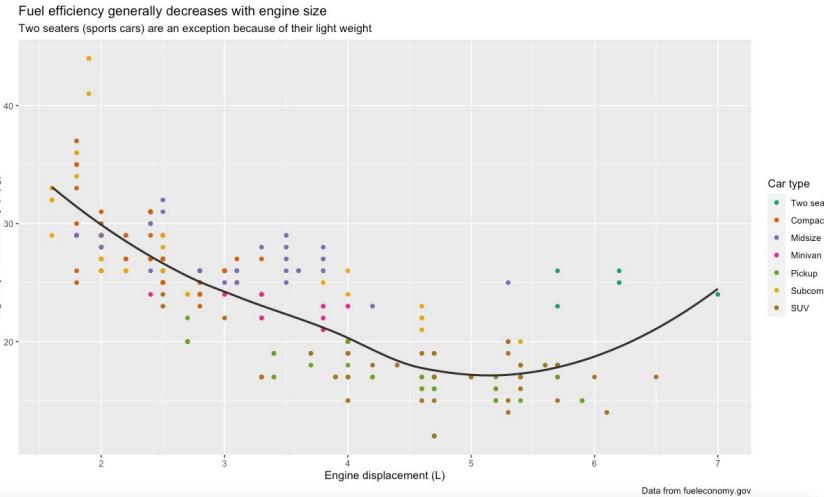
```
p +  
  geom_point(aes(color = class)) +  
  scale_color_brewer(labels = c("Two seater",  
    "Compact",  
    "Midsize",  
    "Minivan",  
    "Pickup",  
    "Subcompact",  
    "SUV"),  
    palette = "Dark2") +  
  geom_smooth(method = "loess", se = FALSE, color = "#333333") +  
  labs(color = "Car type")
```



Layered grammar of graphics

Build a plot layer by layer

```
p +  
  geom_point(aes(color = class)) +  
  scale_color_brewer(labels = c("Two seater",  
    "Compact",  
    "Midsize",  
    "Minivan",  
    "Pickup",  
    "Subcompact",  
    "SUV"),  
    palette = "Dark2") +  
  geom_smooth(method = "loess", se = FALSE, color = "#333333") +  
  labs(  
    title = "Fuel efficiency generally decreases with engine size",  
    subtitle = "Two seaters (sports cars) are an exception  
      because of their light weight",  
    caption = "Data from fueleconomy.gov",  
    x = "Engine displacement (L)",  
    y = "Highway fuel economy (mpg)",  
    color = "Car type")  
)
```



Storytelling

Big Idea

That one key message you want to communicate

Big Idea

Three components of a big idea:

1. Articulate your unique point of view
2. Convey what's at stake
3. Be a complete sentence

These Are Not Big Ideas

Lunar Mission

Client Sales Call

Third-Quarter Update

These Are Big Ideas

The United States should lead in space achievement because it holds the key to our future on Earth

Our software gives your customers access to their records, which saves your employees time and increases your margins by 2 percent

Third-quarter numbers are down; and to stay in the game, every department needs to support the sales initiative

SPAR

Four characteristics of a good so-what

SPAR

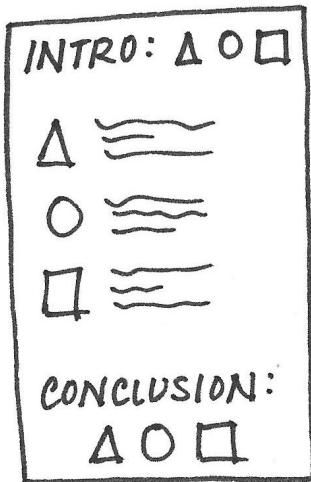
A good so-what should be:

- **S - Succinct**
 - Is your so-what short, clear, and to the point. Would someone be able to say it back to you easily?
- **P - Provocative**
 - Is your so-what something that someone could realistically disagree with? Will it start a conversation?
- **A - Actionable**
 - Is it clear what someone should do if they're convinced by your argument? You don't need to provide narrow solutions, but it should offer clear direction. "If we agree on this provocative thing I've said, we should move concretely in this direction."
- **R - Relevant**
 - Finally, is your so-what related to an ongoing discussion in your field or team, or does it speak to assumptions, questions, or debates that are currently in progress in your field or team?



Bing, Bang, Bongo

Beginning, Middle, End



POINT YOU WANT TO MAKE		STORY ABOUT ORGANIZATIONAL CHANGE ¹⁰	STORY ABOUT CUSTOMER INTEREST
BEGINNING	When, Who, Where	Every cross-divisional function could benefit from a steering committee.	Mid-sized companies would save money if they bought this software.
MIDDLE	Context	A few years ago, the sales team tackled a problem that demonstrates the cross-divisional issues I'm talking about.	Last year I met with Susan, the CEO from a company very similar to yours.
	Conflict	At the time, all sales groups were independent.	She was strategically wicked-smart, and just like you, she was curious whether our software could help her business.
	Proposed Resolution	This means we were confusing the customers with many different rules, processes, and formats.	She knew that her organization wouldn't scale if she didn't have software that worked in a global environment.
	Complication	So we decided to create a sales-steering committee.	We installed a trial version for the employees in the Dallas office only.
END	Actual Resolution	You can imagine how hard it was to reach agreement on anything.	She was concerned that the employees would have a dip in productivity while learning a new program.
		But we agreed to meet every two weeks to discuss common ground. Over the next year, we standardized all our processes and learned a lot from each other. The customers were much happier with our service.	Instead, employee productivity increased, and Susan received numerous e-mails about how the software will help them gain market advantage.
			It took her less than a week to agree to an organization-wide installation.
MOST IMPORTANT POINT		I think every cross-divisional function could benefit from a steering committee.	Your company has the same challenges and would benefit, too.

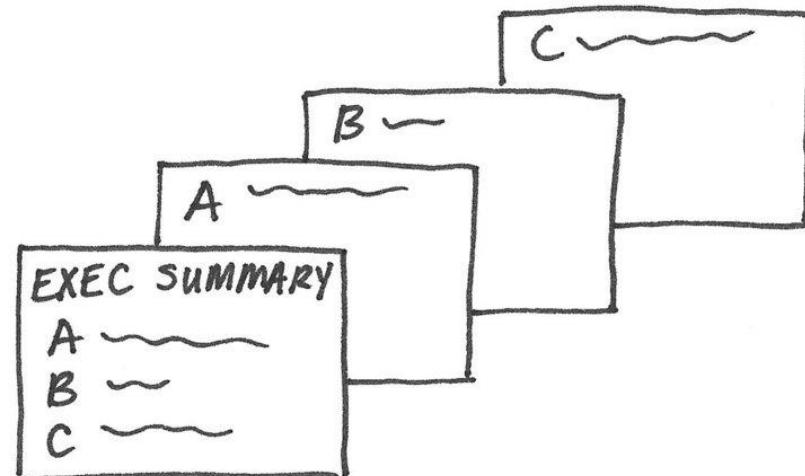
Horizontal logic

Slide titles tell the overarching story you want to communicate

Horizontal logic

“You can read *just the slide title* of each slide and, together, these snippets tell the overarching story

It is important to have *action titles* (not descriptive titles) for this to work well”
– Cole



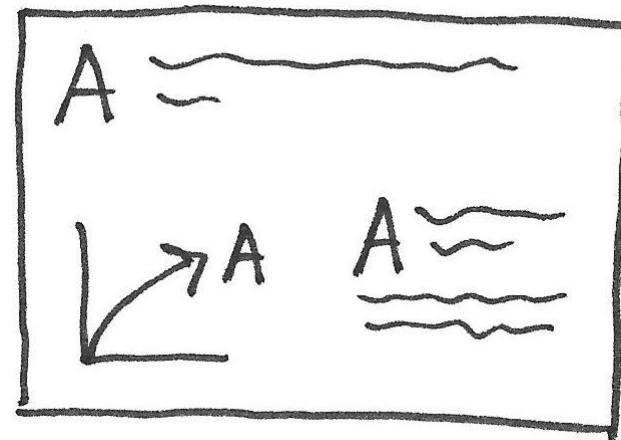
Vertical logic

Information on a given slide is self-reinforcing

Vertical logic

“The content reinforces the title and vice versa. The words reinforce the visual and vice versa.

There isn’t any extraneous or unrelated information.” – Cole



Storyboarding

Establish a structure for your communication

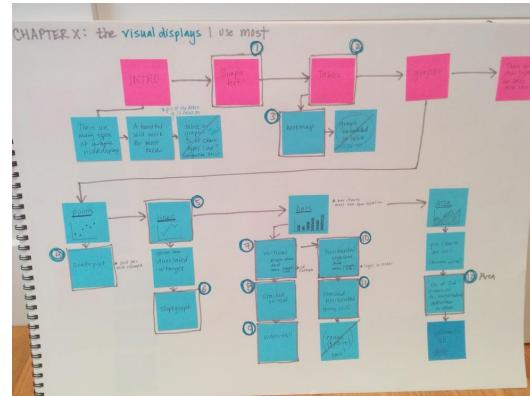
Storyboarding

“When it comes to storyboarding, don’t start with presentation software.

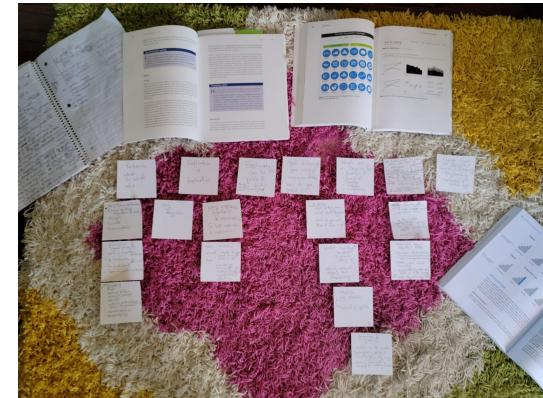
Instead, use a whiteboard, Post-it notes, or plain paper.

It’s much easier to put a line through an idea on a piece of paper or recycle a Post-it note without feeling the same sense of loss as when you cut something you created on your computer.” – Cole

Cole storyboarding a chapter

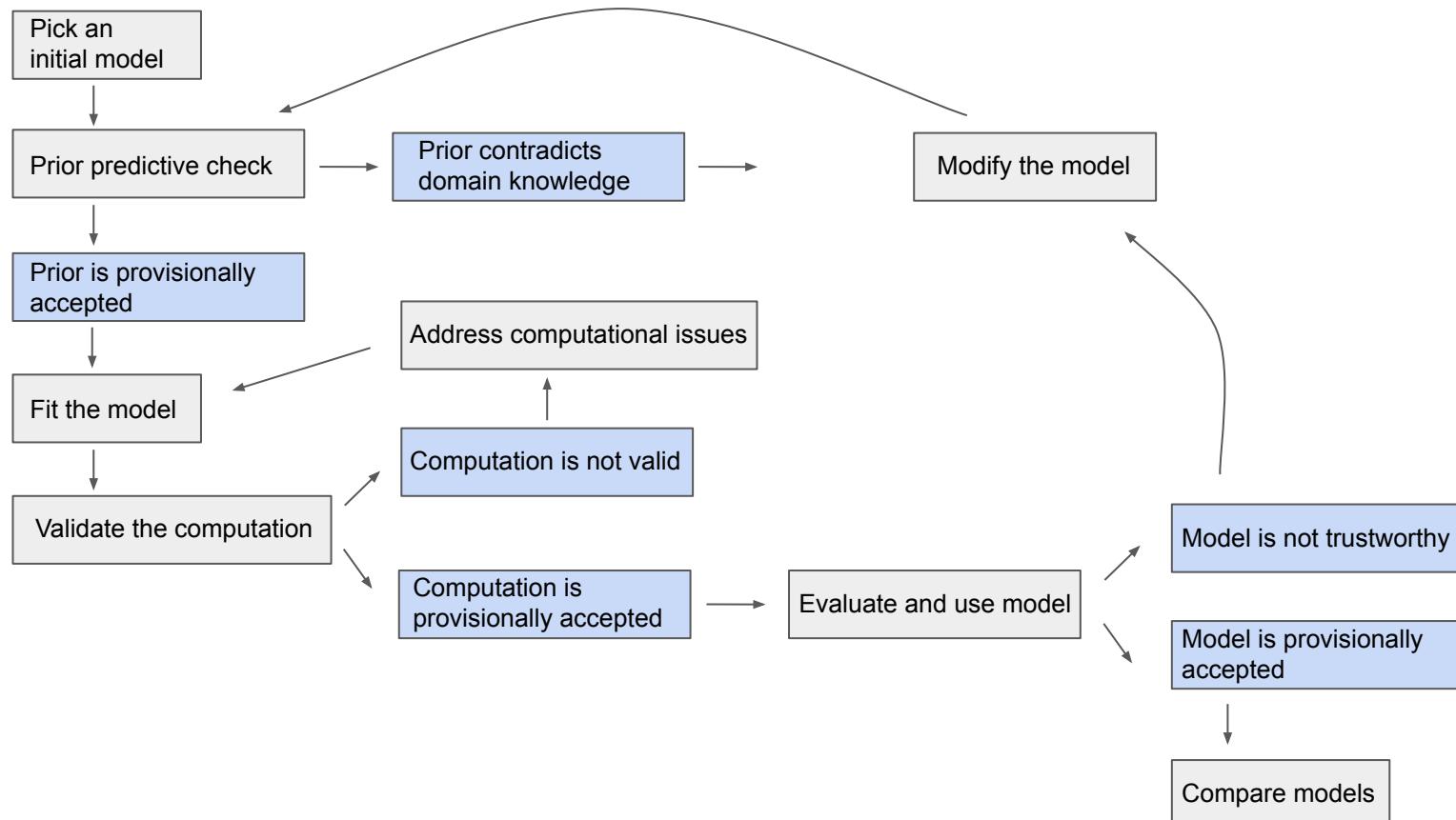


Me storyboarding this presentation

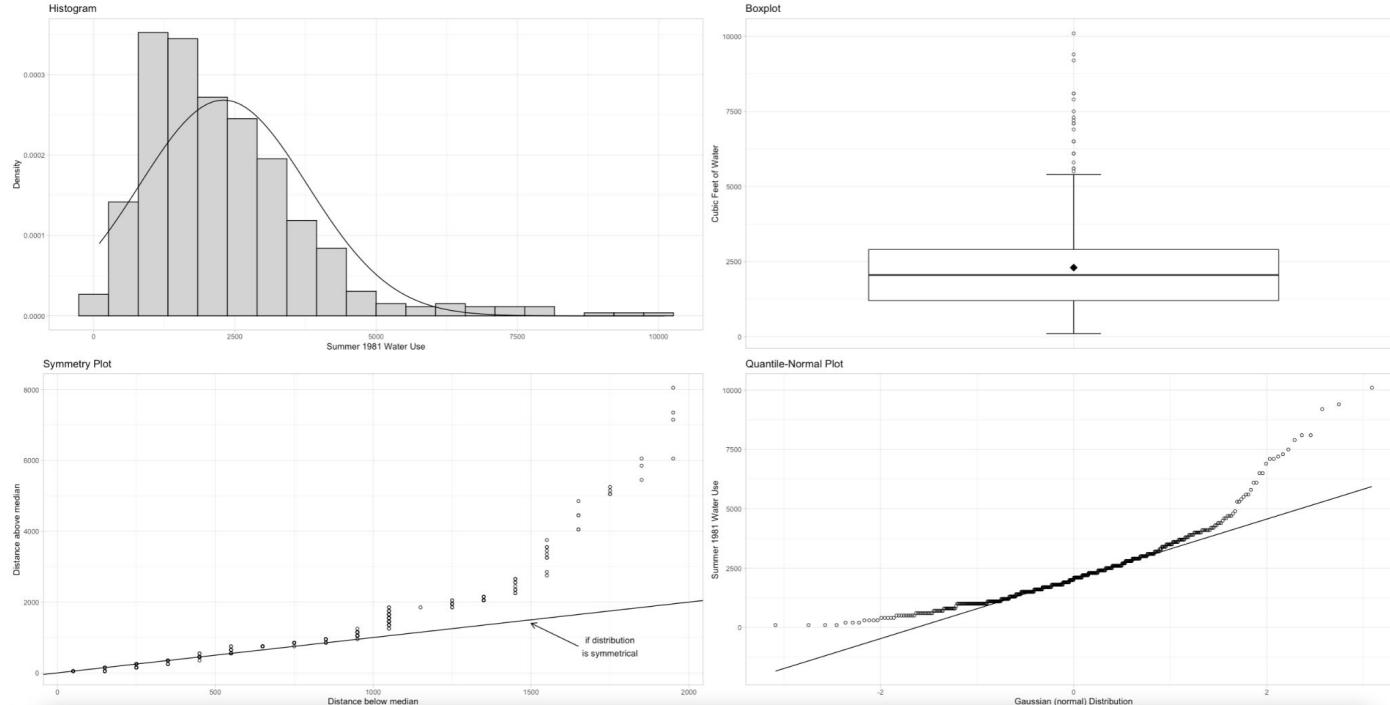


Visualization in Bayesian workflow

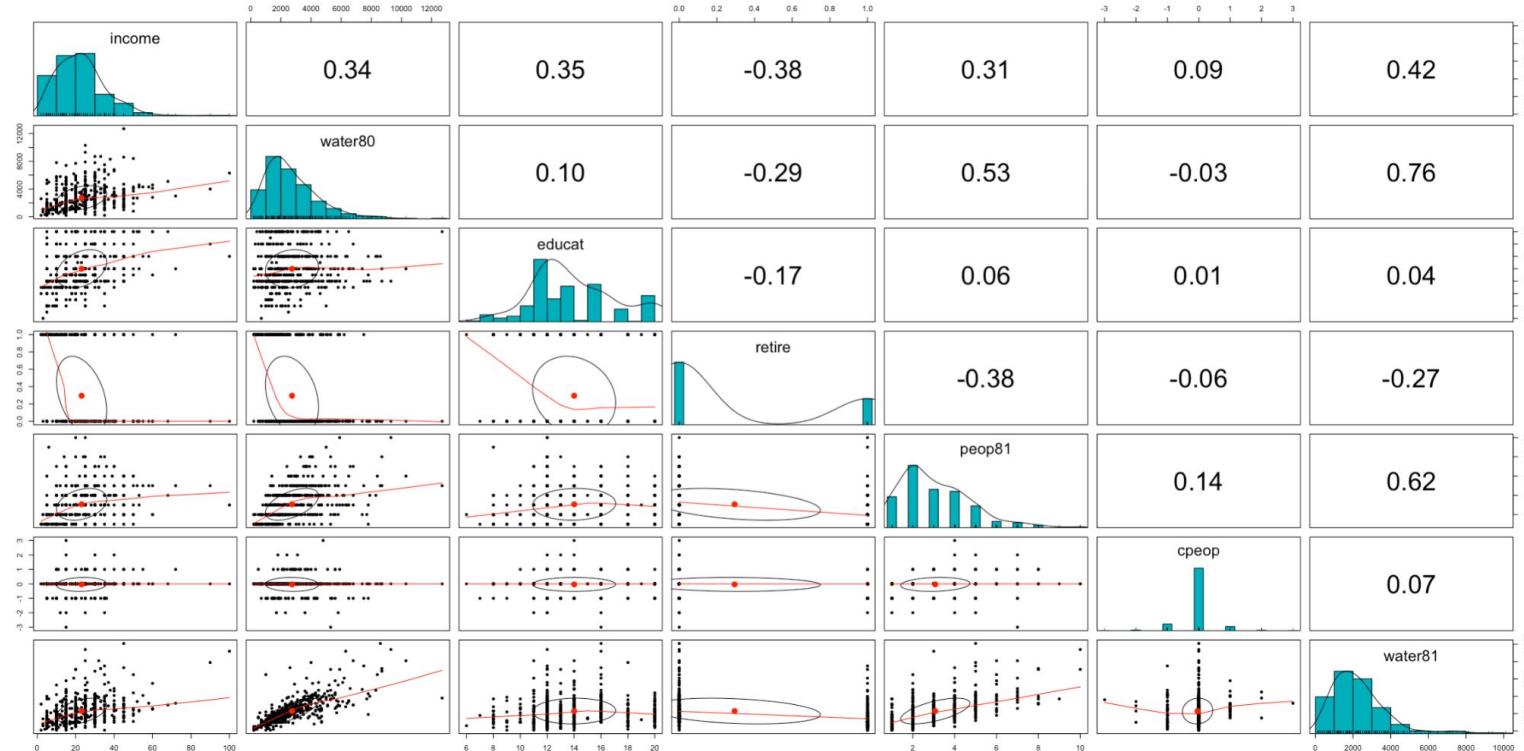
Bayesian workflow



Exploratory data analysis helps you *check distributions of variables before analysis*

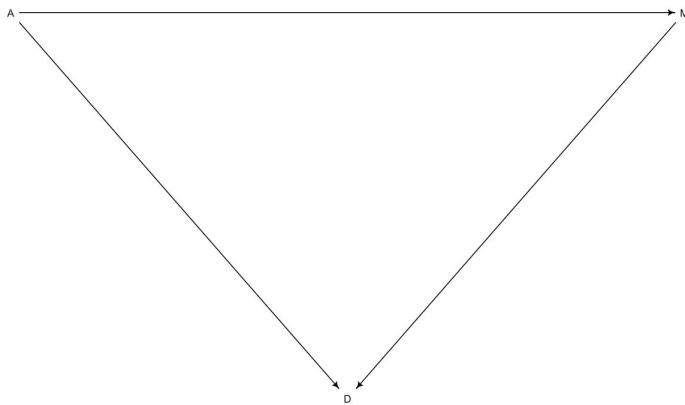


Exploratory data analysis helps you *check relationships among variables* before analysis

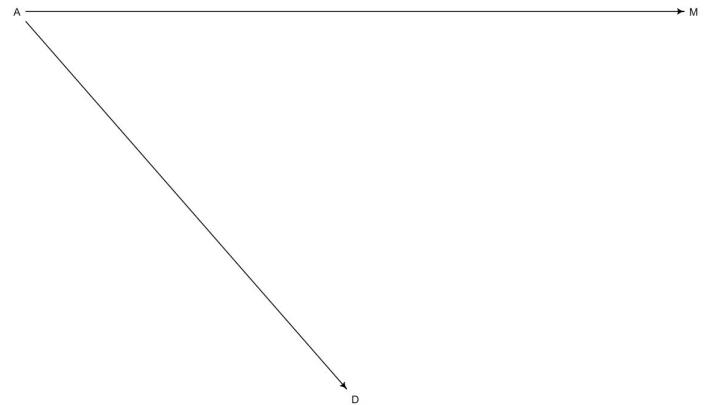


Conceptual (causal) diagrams illustrate the *assumptions* and *testable implications* of your models

Model 1



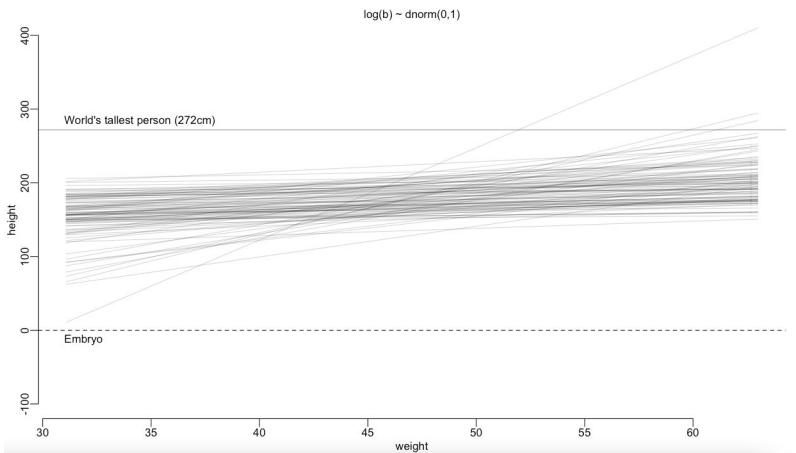
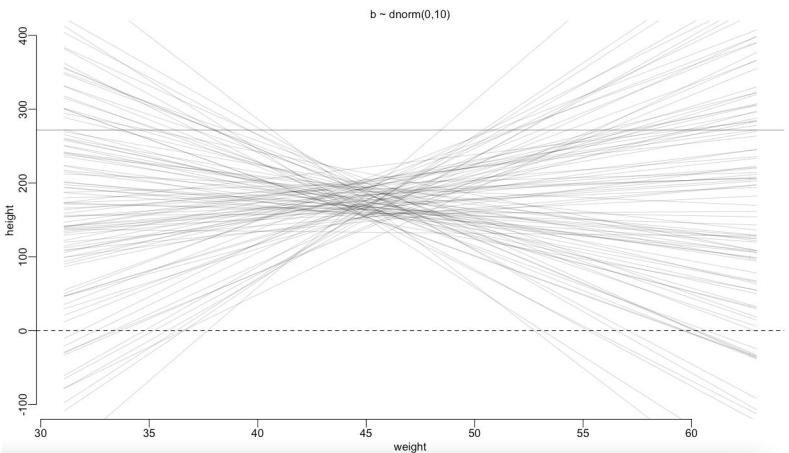
Model 2



```
> DMA_dag1 <- dagitty('dag{ D <- A -> M -> D }')
> impliedConditionalIndependencies( DMA_dag1 )
>
```

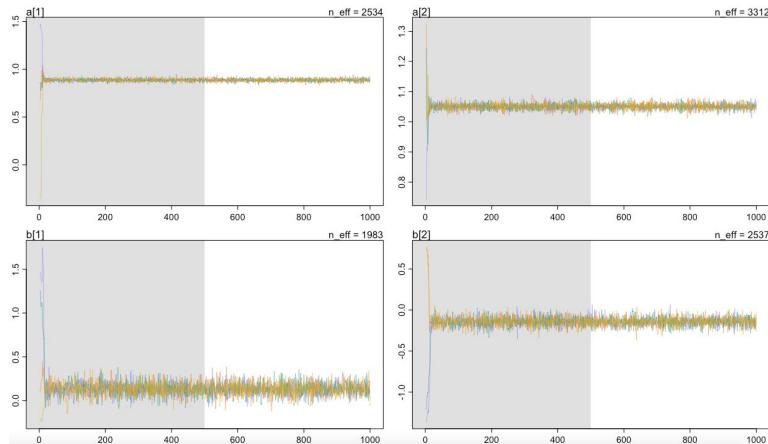
```
> DMA_dag2 <- dagitty('dag{ D <- A -> M }')
> impliedConditionalIndependencies( DMA_dag2 )
D _||_ M | A
>
```

Prior predictive simulations help you *check the implications of prior distributions before analysis*

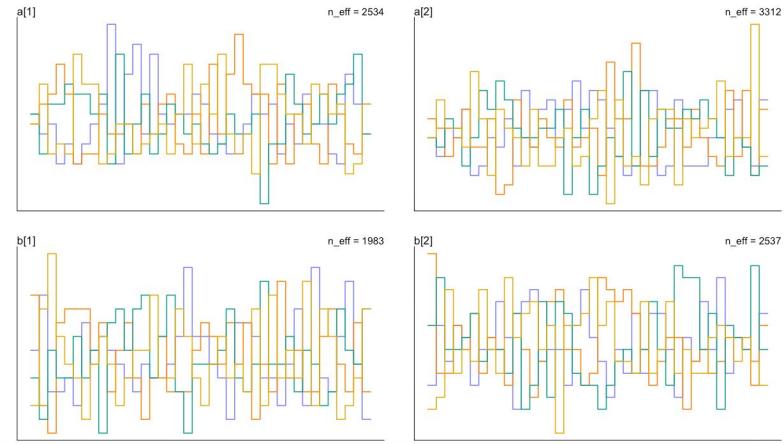


Markov chain Monte Carlo (MCMC) diagnostics help you catch computational problems before interpreting the results

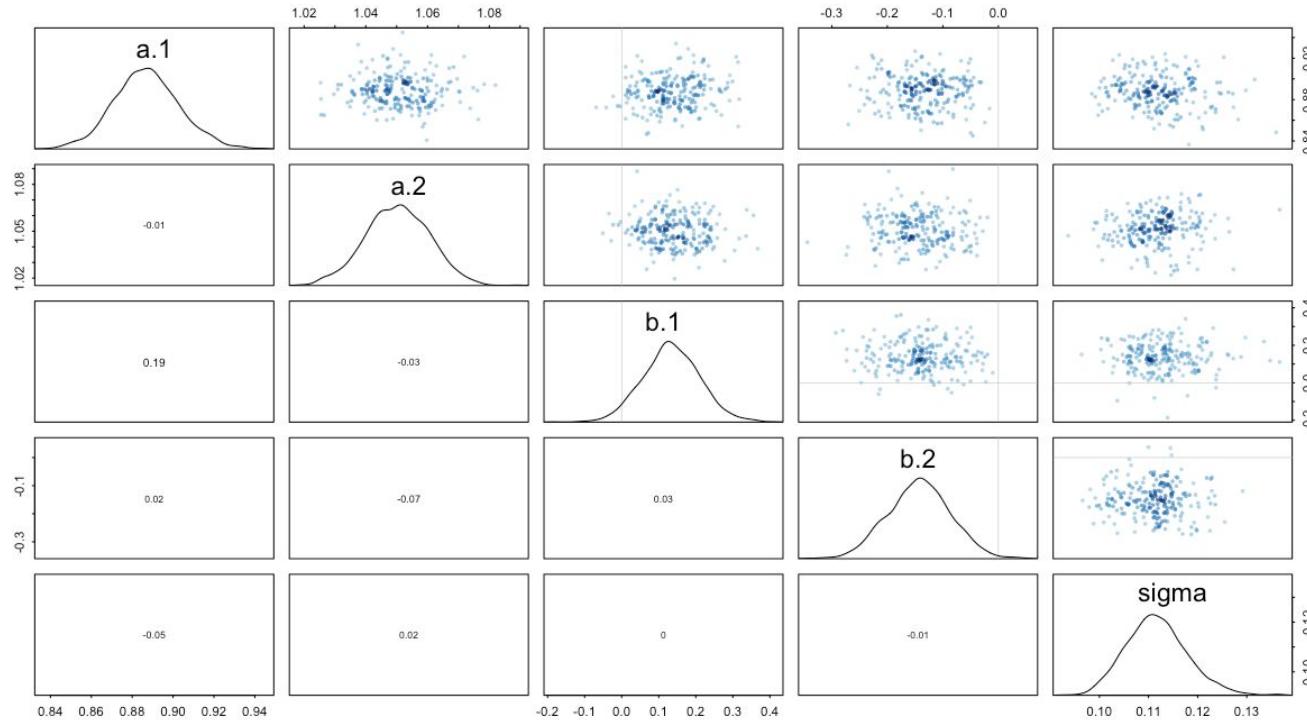
Trace plots: plot samples in sequential order, joined by a line



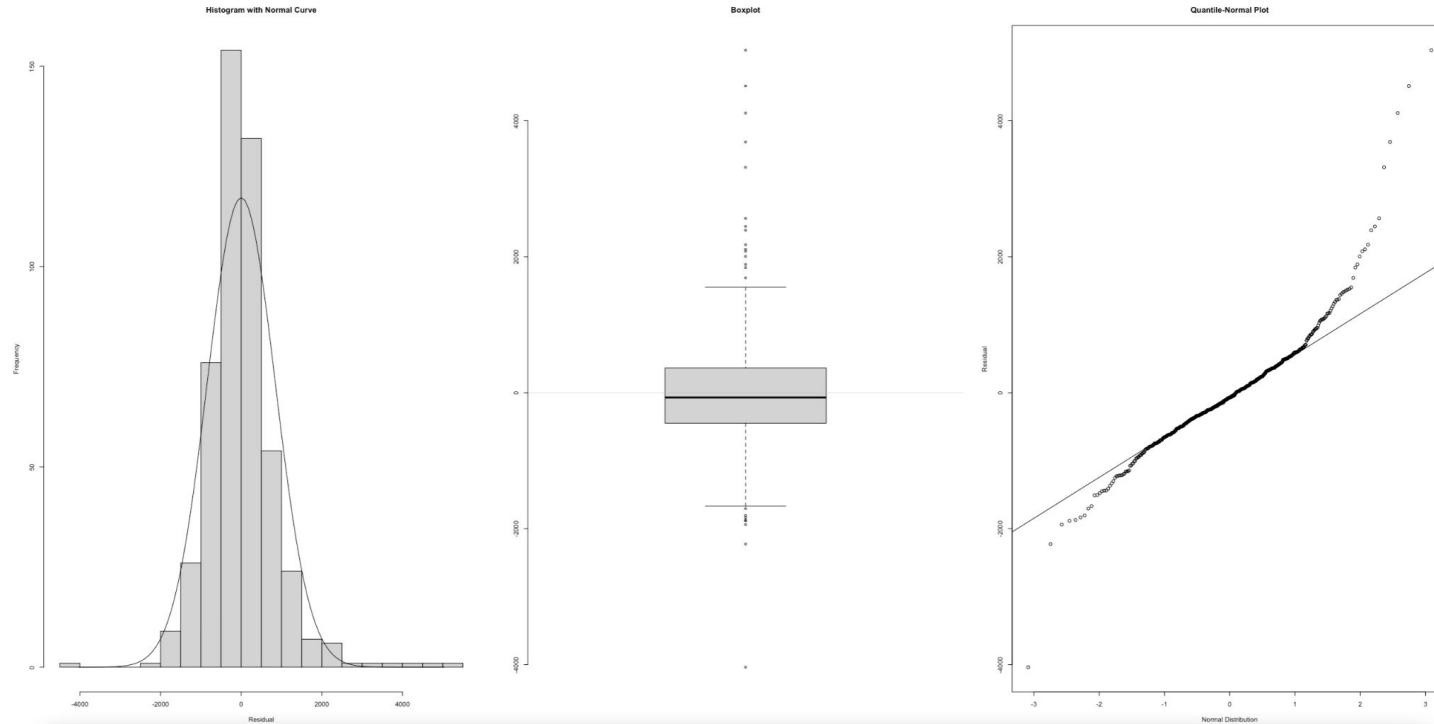
Rank plots: plot the distribution of ranked samples, as overlapping histograms



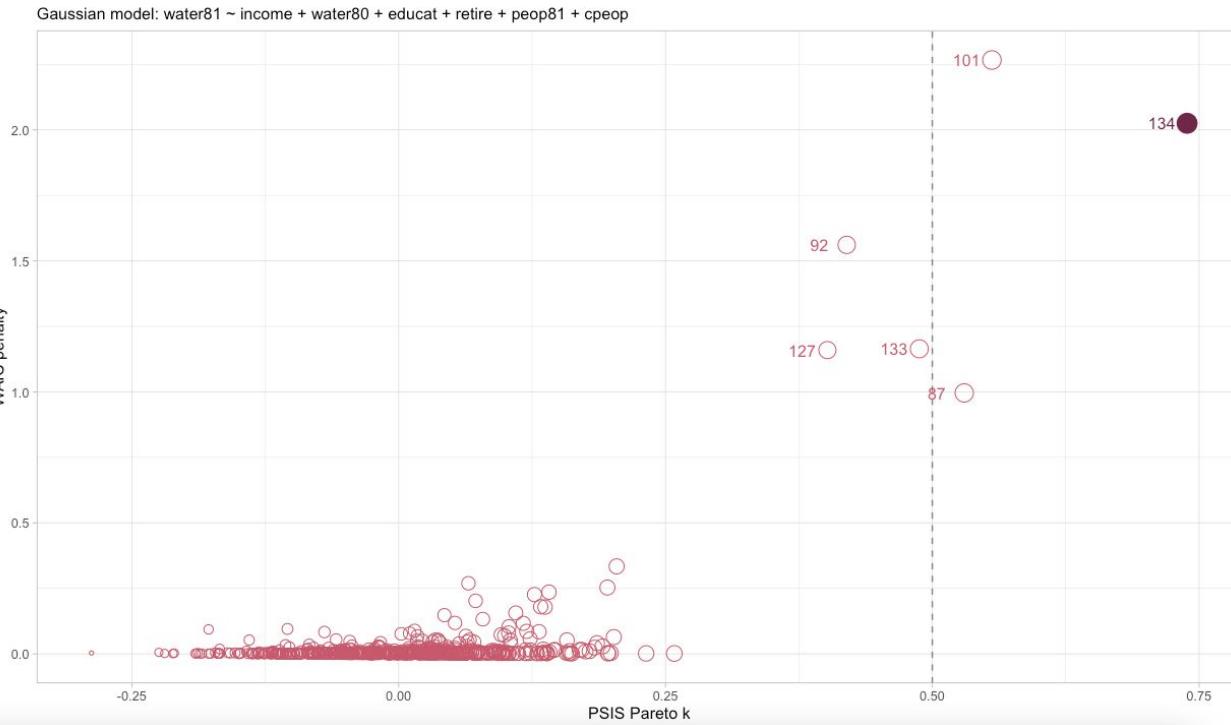
Posterior parameter distributions illustrate compatible parameter values and relationships between parameters



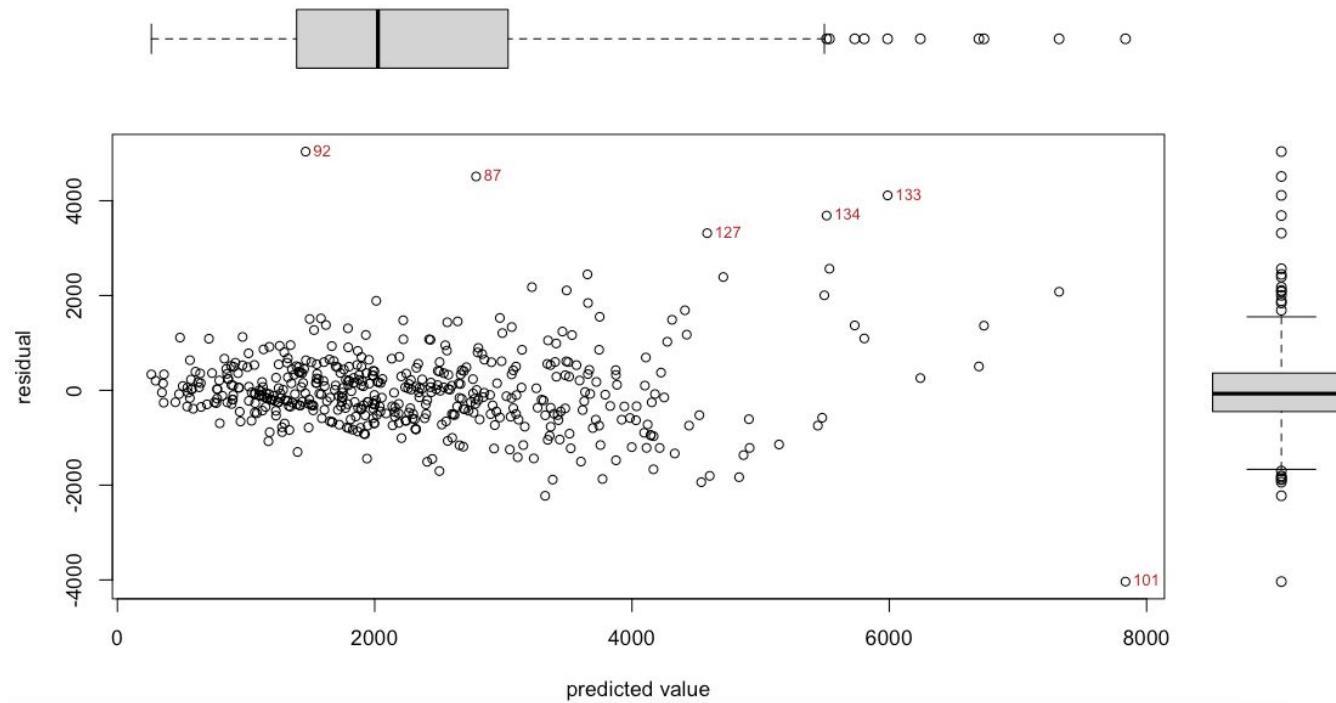
Residual plots can uncover problems such as *nonnormality*



PSIS Pareto k plots can uncover problems such as *influential outliers*

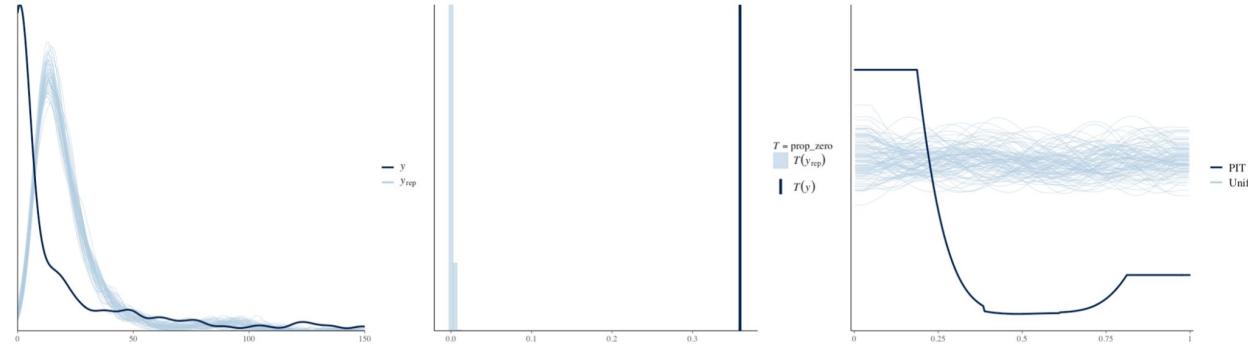


Residual vs predicted plots can uncover problems such as *curvilinearity, heteroscedasticity, nonnormality, or outliers*

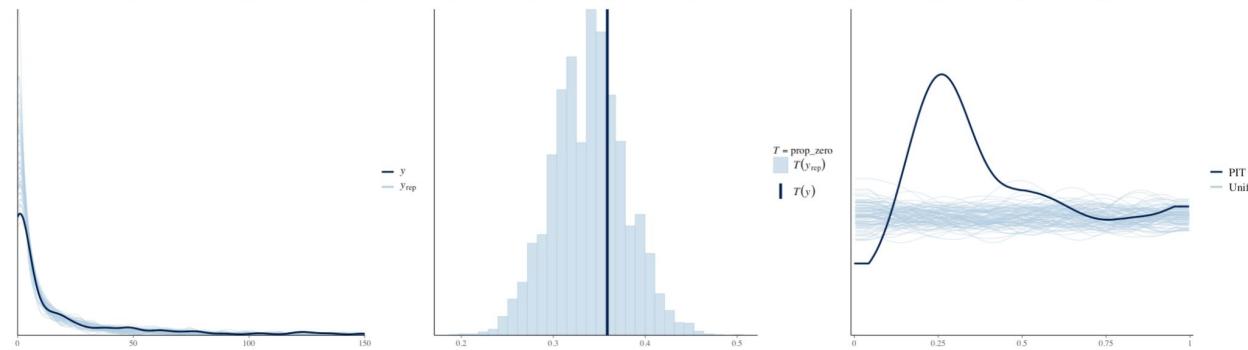


Posterior predictive checks illustrate whether a *fitted model can generate data resembling the observed data*

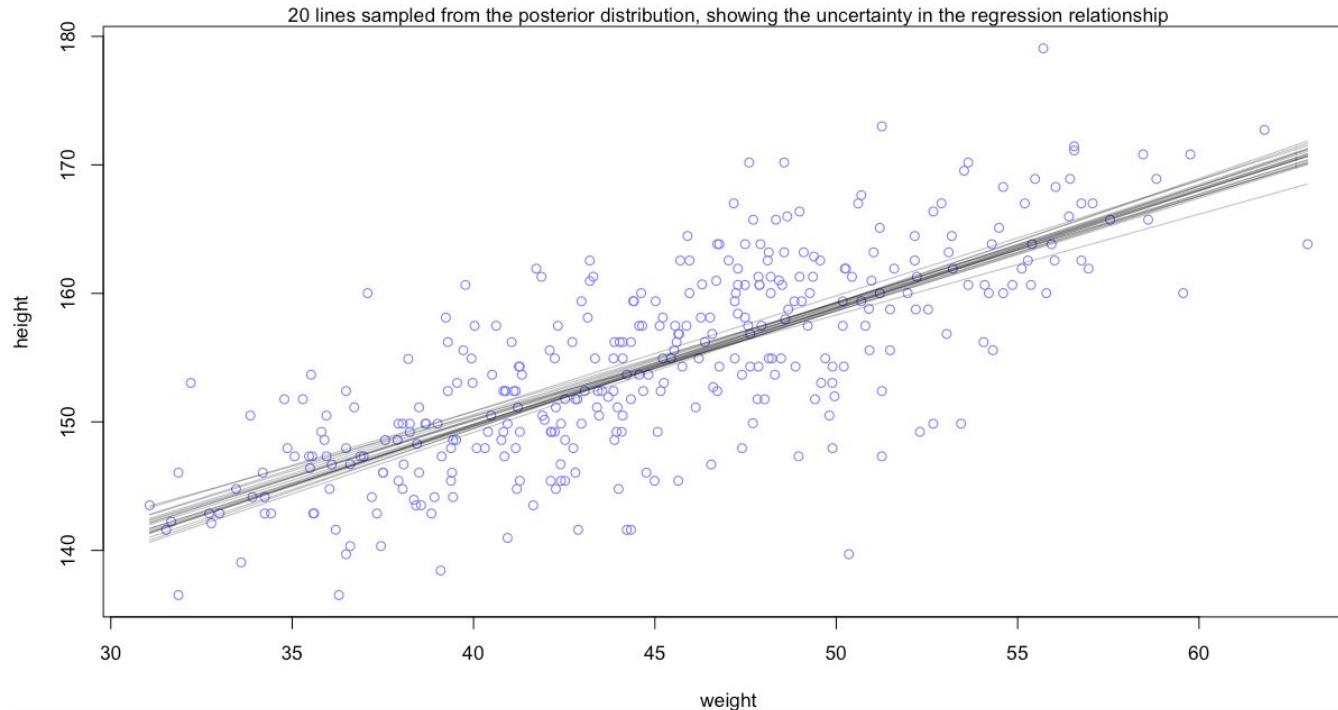
Model 1



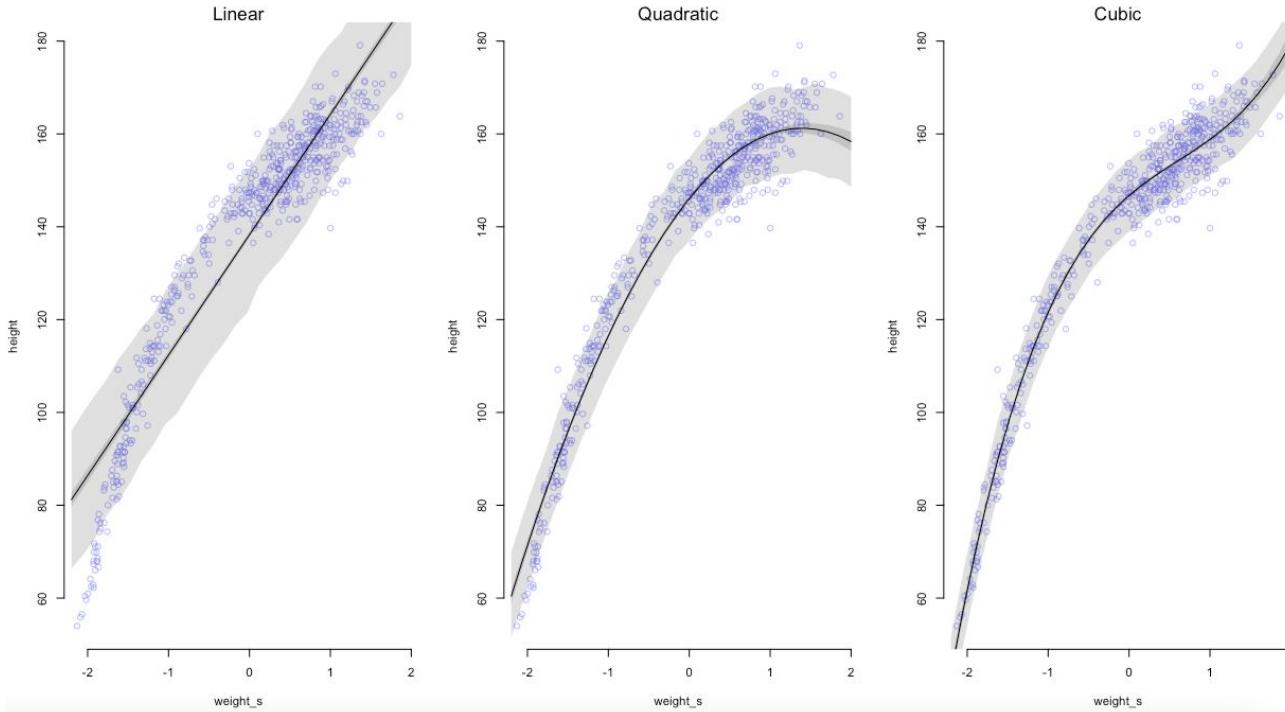
Model 2



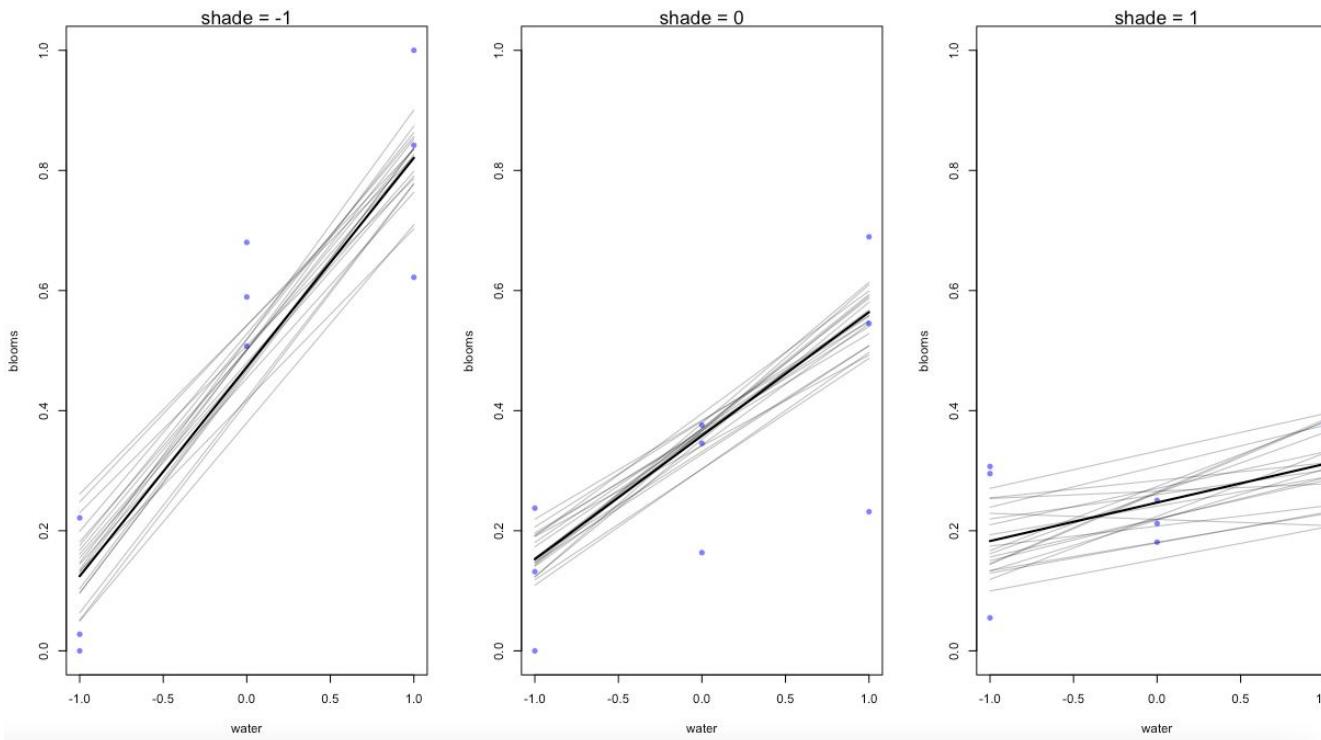
Plotting posterior inference against the data helps you *interpret* the model and provides an *informal check* on model assumptions



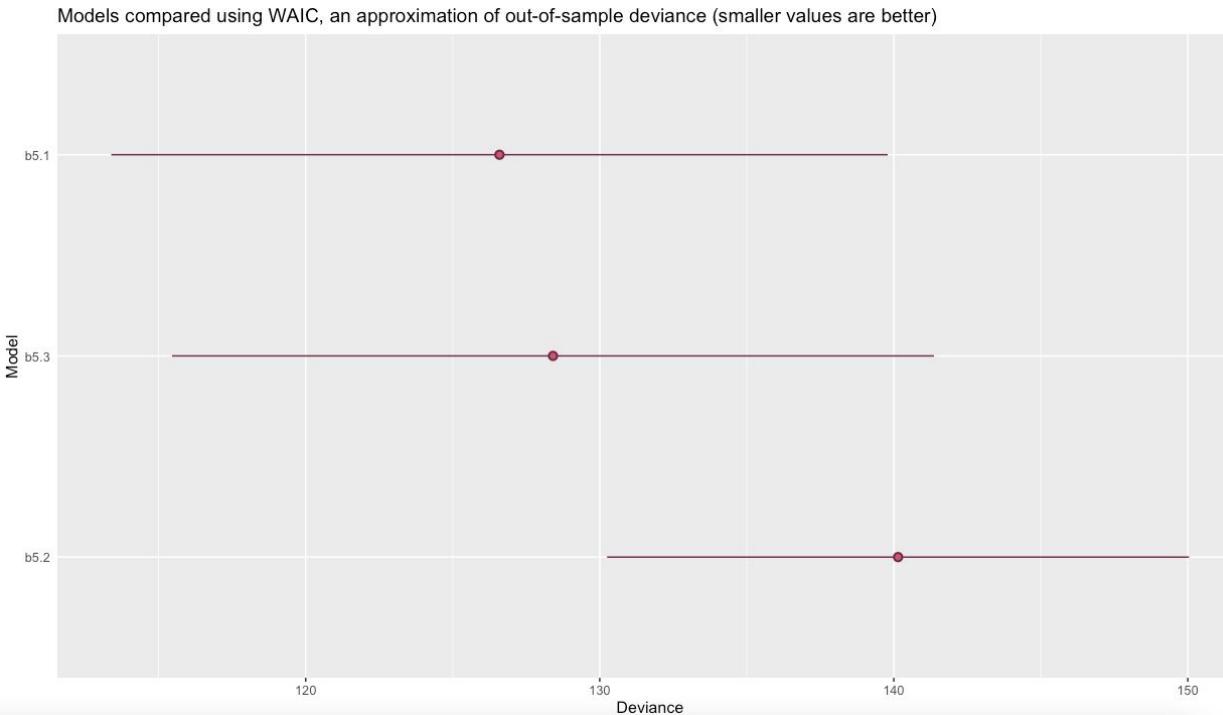
Plotting posterior inference against the data helps you *interpret* the model and provides an *informal check* on model assumptions



Plotting posterior inference against the data helps you *interpret* the model and provides an *informal check* on model assumptions



Model comparison plots help you compare the predictive performance of competing models



Appendix

References

- Cairo, Alberto. *The Truthful Art: Data, Charts, and Maps for Communication*. New Riders, 2016.
- Cleveland, William. *Visualizing Data*. New Jersey: Hobart Press, 1993.
- Duarte, Nancy. *Resonate: Present Visual Stories that Transform Audiences*. Hoboken: Wiley & Sons, 2010.
- Dykes, Brent. *Effective Data Storytelling: How to drive change with data, narrative, and visuals*. Hoboken: Wiley & Sons, 2020.
- Few, Stephen. *Show Me the Numbers: Designing Tables and Graphs to Enlighten, 2nd Edition*. El Dorado Hills: Analytics Press, 2012.
- Gabry, Jonah, et al. "Visualization in Bayesian workflow." *J. R. Stat. Soc. A* (2019) 182: 389-402. <https://doi.org/10.1111/rssc.12378>.
- Gelman, Andrew, et al. "Bayesian workflow." arXiv (2020) [stat.ME] 1-77. <https://doi.org/10.48550/arxiv.2011.01808>.
- Gelman, Andrew, et al. *Regression and Other Stories*. New York: Cambridge University Press, 2021.
- Hamilton, Lawrence. *Regression with Graphics: A Second Course in Applied Statistics*. Belmont: Duxbury Press, 1992.
- Kirk, Andy. *Data Visualisation, 2nd Edition*. Thousand Oaks: SAGE Publications, 2019.
- Knaflic, Cole. *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Hoboken: Wiley & Sons, 2015.
- McElreath, Richard. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan, 2nd Edition*. Boca Raton: CRC Press, 2020.
- Munzner, Tamara. *Visualization Analysis & Design*. Boca Raton: A K Peters Visualization Series, CRC Press, 2014.
- Satyanarayan, Arvind, et al. "Vega-Lite: A Grammar of Interactive Graphics." IEEE TVCG (Proc. InfoVis), 2017.
- Schwabish, Jonathan. "An Economist's Guide to Visualizing Data." JEP (2014) 28: 209-234.
- Schwabish, Jonathan. *Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks*. New York: Columbia University Press, 2021.
- Tufte, Edward. *The Visual Display of Quantitative Information, 2nd Edition*. Cheshire: Graphics Press, 2001.
- Wickham, Hadley. "A Layered Grammar of Graphics." JCGS (2010) 19: 3-28.
- Yau, Nathan. *Data Points: Visualization That Means Something*. Indianapolis: Wiley & Sons, 2013.

