

Shifting Expectations for Encoding Rules Mitigates Misinterpretation of Connected Scatterplots

Wen Xu and Lace Padilla

Accepte for IEEE Transactions on Visualization and
Computer Graphics

Why this paper?

- Misinterpretation of graphs
- Provides systematic evaluation of the efficacy of suggested solutions by graph designers
-
- Includes theory on graph interpretation

Pro & con list

Well done!

- Preprint available
(<https://www.lacepadilla.com/pubs.html>)
- Preregistration
- No salami slicing - 3 experiments in 1 paper
- Gender inclusive, but still politically correct in USA in Trump time?
*The sample size after exclusion was $n = 404$, **of whom 203 were not men.***
- Paper ends with list of 4 recommendations

Room for improvement:

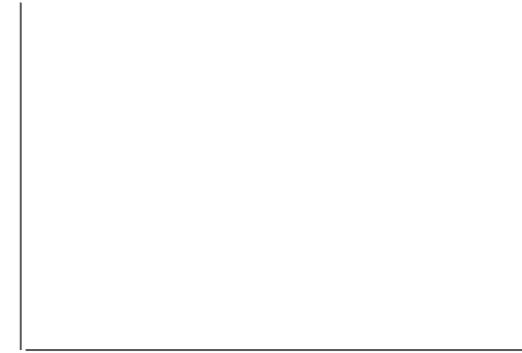
- "Look at our $p = 0.052$ result"!
- Graph literacy is measured and included in the statistical models, but results are not reported in the paper.

Graph comprehension - *Expectations*

Graph conventions

Conventions = Rules used and pervasively understood within the community that shares them

- E.g. from linguistic or cultural conventions
- Taught in schools



Writing direction

Graph schemas

Pinker's theory:

graph schemas = mental structures that describe how a data visualization is interpreted

Result: Schemas "prime" specific encoding rules



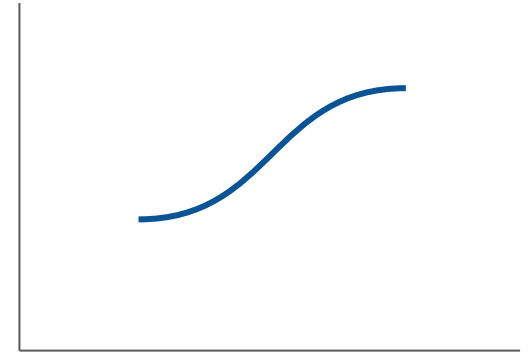
Flat line = no change

Chart-type schemas

These author's theory:

chart-type schemas = abstract mental representations of visualization types, containing the prototypical appearance of specific categories of data visualizations and the rules that govern them

Speculation: various aspects of the appearance of a visualization can activate a chart-type schema



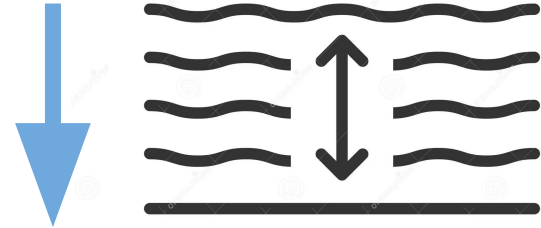
Conceptual Metaphors

Conceptual Metaphor Theory: metaphors we use in everyday language reflect the way we structure thoughts; many metaphors are embodied, or grounded in our physical and bodily experiences

Up is more



down is deeper



WATER DEPTH

Expectations

any intuitive, initial assumptions about encoding rules that may have stemmed from conventions, metaphors, or both.

Can be contradictory!



Mitigating Misinterpretation of Connected Scatterplots

Connected Scatterplots

To visualize time-series data:
connect points on a
scatterplot on temporal
sequence

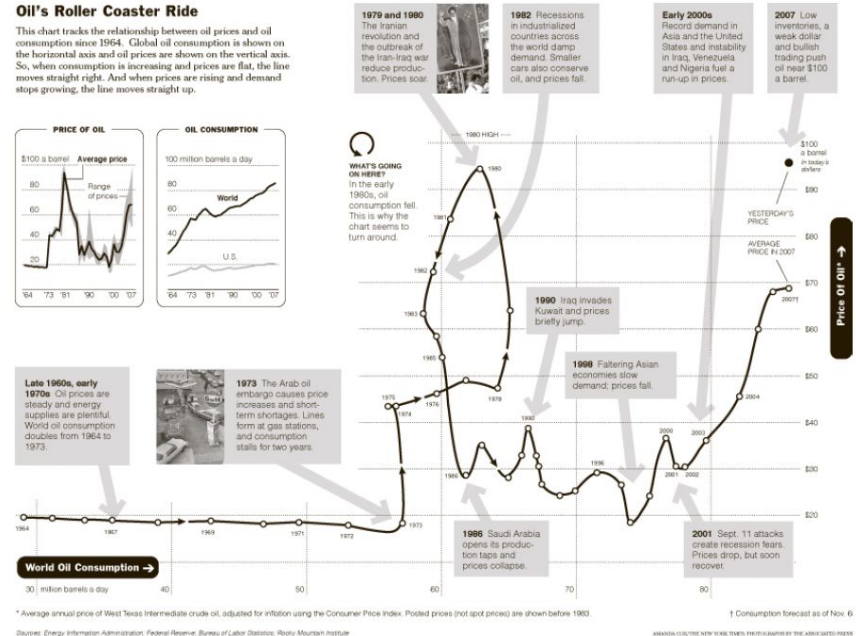


Fig. 1: *New York Times*' Oil's Roller Coaster Ride visualization depicts the relationship between world oil consumption and oil price over time [9].

Expectation conflict!

Connected scatterplot:

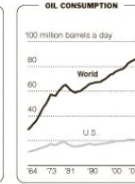
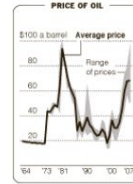
- Time is a line

Line graph schema:

- Right is later

Oil's Roller Coaster Ride

This chart tracks the relationship between oil prices and oil consumption since 1964. Global oil consumption is shown on the horizontal axis and oil prices are shown on the vertical axis. So, when consumption is increasing and prices are flat, the line moves straight right. And when prices are rising and demand stops growing, the line moves straight up.



Late 1960s, early 1970s Oil prices are steady and energy supplies are plentiful. World oil consumption doubles from 1964 to 1973.

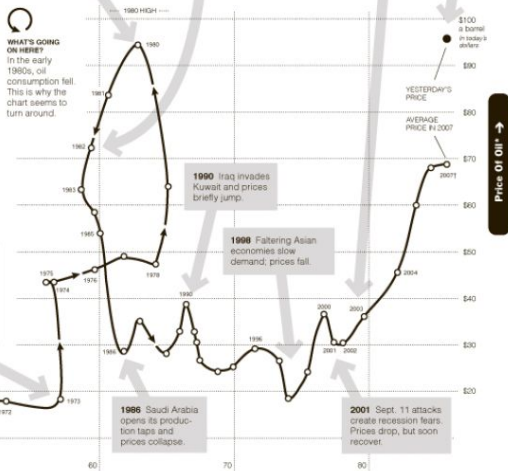
1973 The Arab oil embargo causes price increases and short-term shortages. Lines form at gas stations, and consumption stalls for two years.

1979 and 1980 The Iranian revolution and the outbreak of the Iran-Iraq war reduce production. Prices soar.

1982 Recessions in industrialized countries across the world damp demand. Smaller cars also conserve oil, and prices fall.

Early 2000s Record demand in Asia and the United States and instability in Iraq, Venezuela and Nigeria fuel a run-up in prices.

2007 Low inventories, a weak dollar and bullish trading push oil near \$100 a barrel.



* Average annual price of West Texas Intermediate crude oil, adjusted for inflation using the Consumer Price Index. Pooled prices (not spot prices) are shown before 1983.

Sources: Energy Information Administration, Federal Reserve, Bureau of Labor Statistics, Rocky Mountain Institute

† Consumption forecast as of Nov. 6

ARABIAN OIL THE NEW YORK TIMES PHOTOGRAPH BY THE ASSOCIATED PRESS

Fig. 1: *New York Times*' Oil's Roller Coaster Ride visualization depicts the relationship between world oil consumption and oil price over time [9].

Goal of this study

Designers have proposed various design strategies to mitigate misinterpretation of connected scatterplots

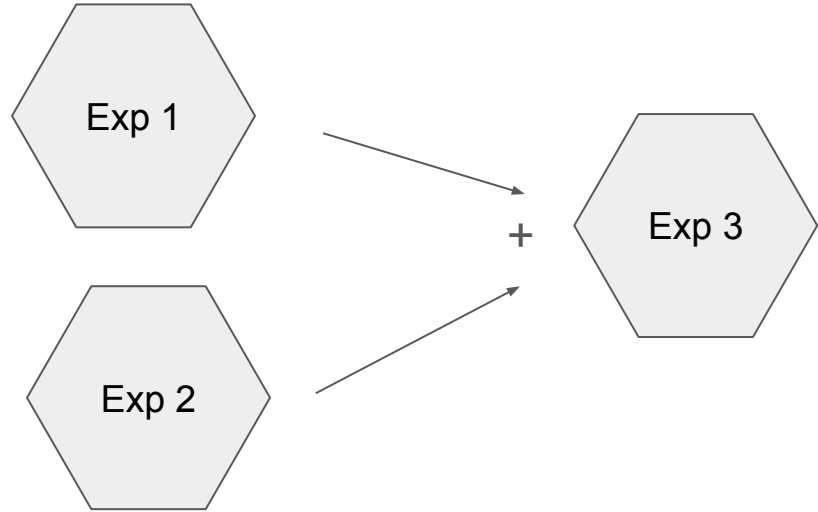
But were never tested.. Until now!



Two approaches:

1) Suppress incorrect chart-type schema

2) Emphasize correct expectation



Experiment 1

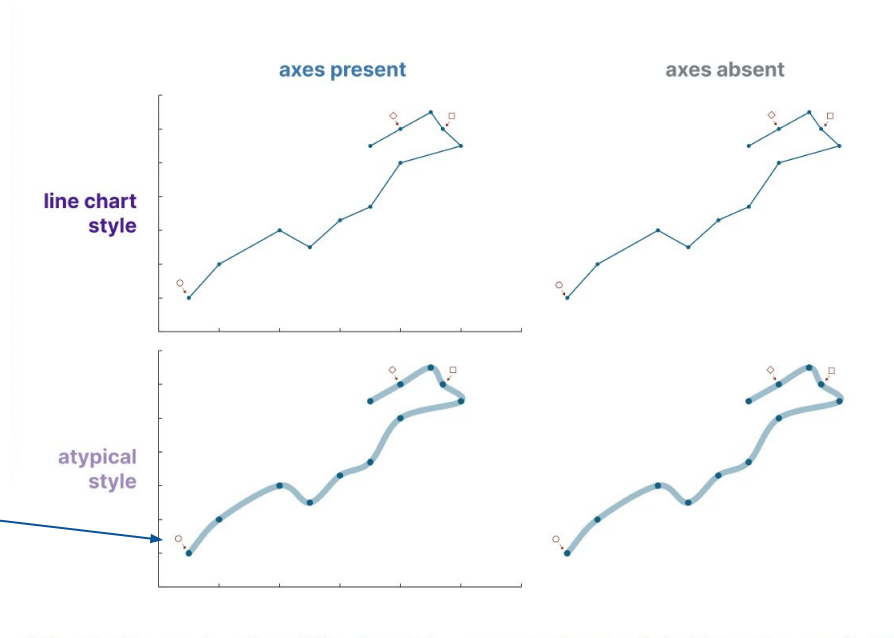
Approach 1: Suppress incorrect chart-type schema

Reduce visual similarity of connected scatterplots to line charts:

- Remove axes
- Atypical lines
 - Curved
 - Thick
 - Translucent

In task instruction:

This is the first year in the dataset.



Tasks:

1) Chart-type judgment:

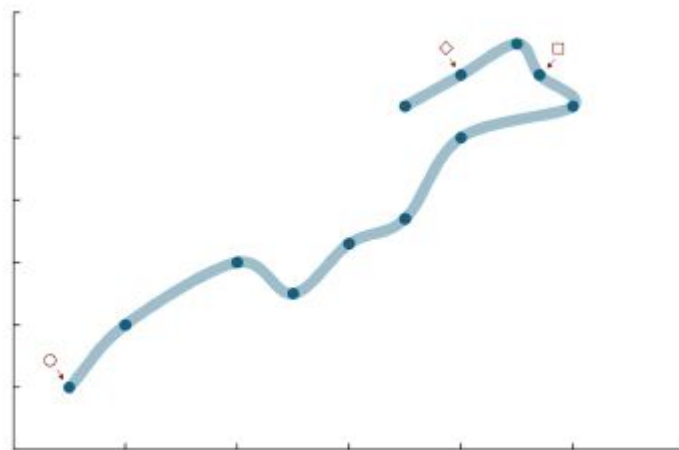
What type of chart is this?

[textbox]

2) Rule expectation:

Indicate which one of the two highlighted points show data for a later year than the other, based on your intuition?

- *Right is later*
- *Time is a line*



Tasks:

- 1) Chart-type judgment:

What type of chart is this?

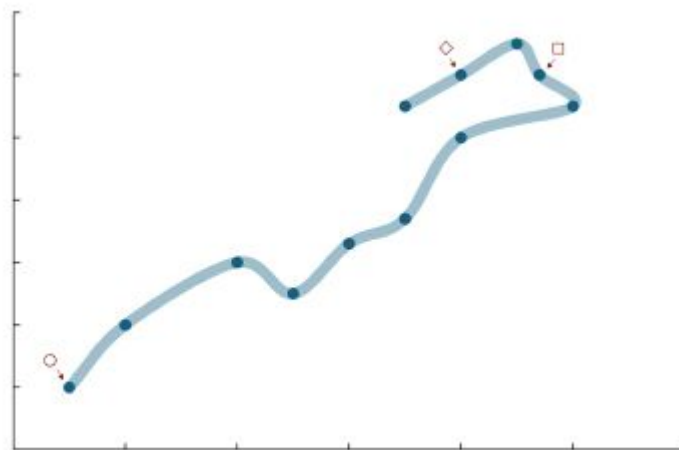
[textbox]

- 2) Rule expectation:

Indicate which one of the two highlighted points show data for a later year than the other, based on your intuition?

☐ *Right is later*

☐ *Time is a line*



Participants

U.S. population on Prolific

- 18 years and older
- Fluent in English
- Approval rate of $\geq 80\%$
- Used desktop

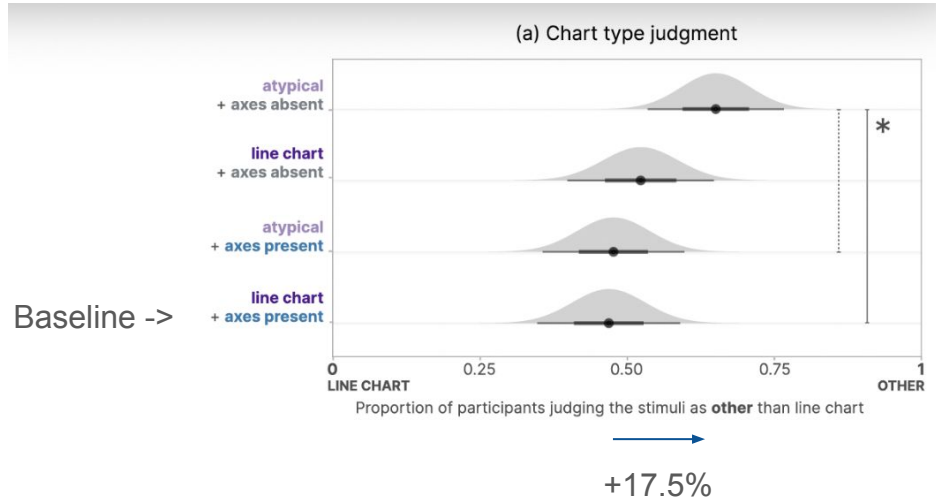
Exclusion based on:

- Attention check
- Task comprehension check



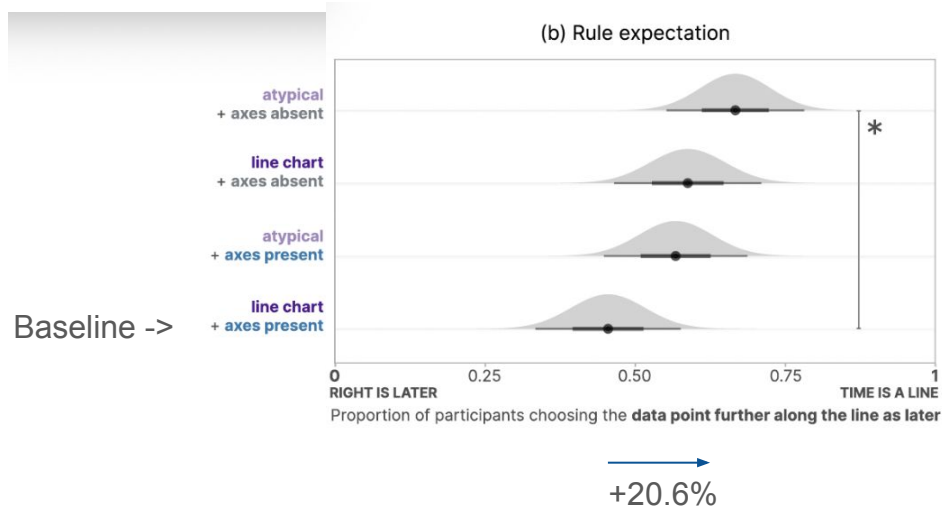
Results - chart type judgment

What type of chart is this?



Results - Rule expectation

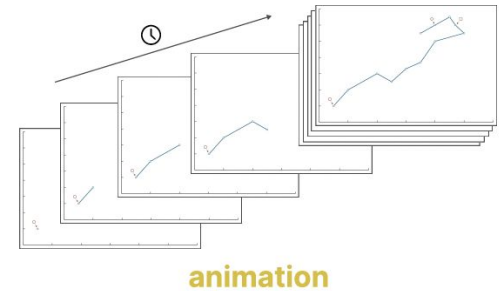
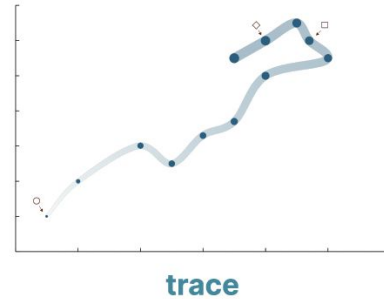
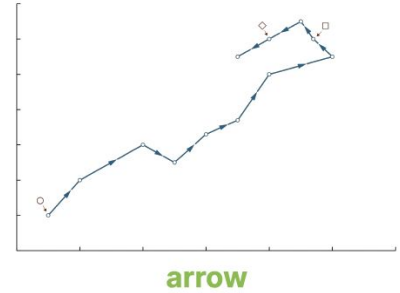
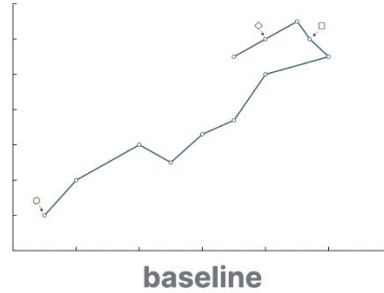
What point is the later year?



Experiment 2

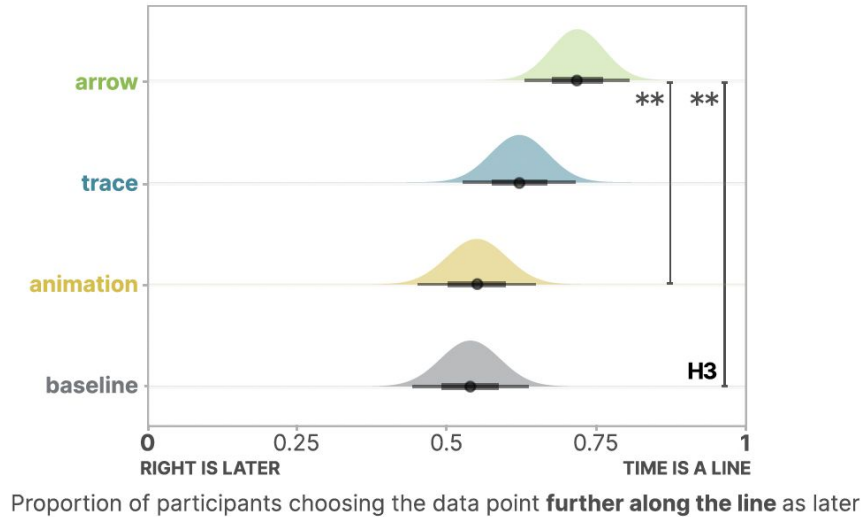
Approach 2: Emphasize correct expectation with directional cues

- Arrows
- Trace-line effect
- Animation (time = sequence)



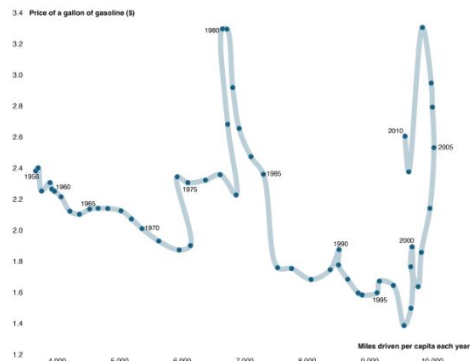
Results - Rule expectation

What point is the later year?

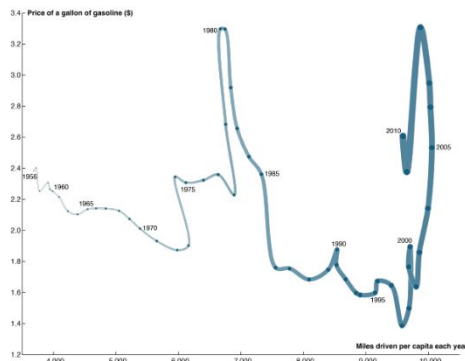


Experiment 3

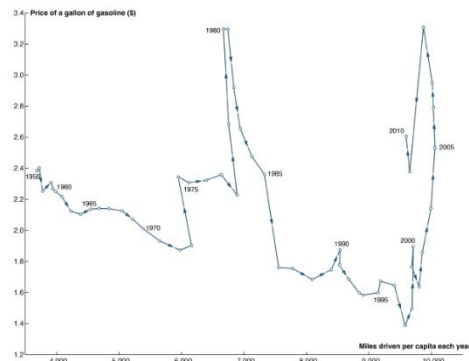
Goal 1: Generalization to more realistic graphs



schema

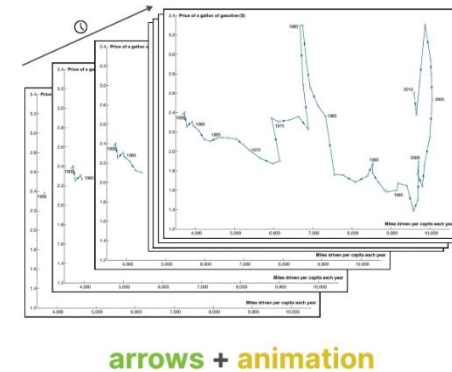
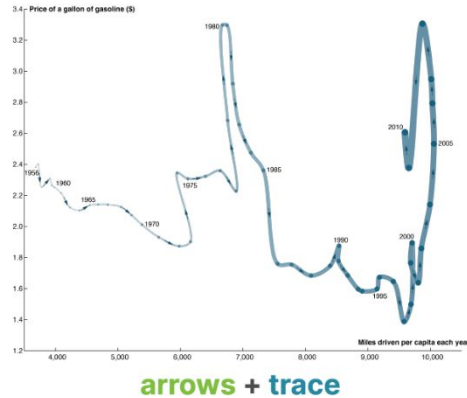
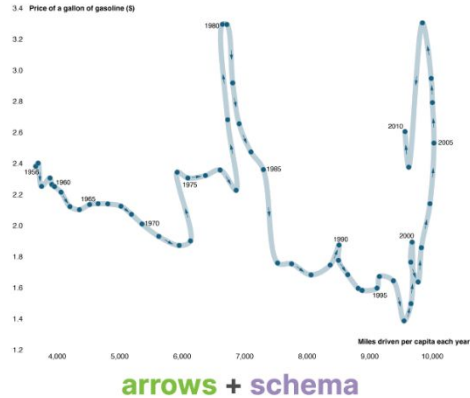
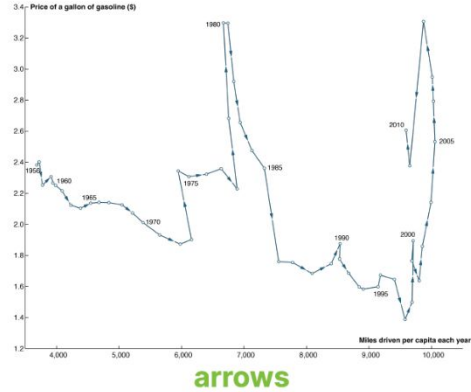
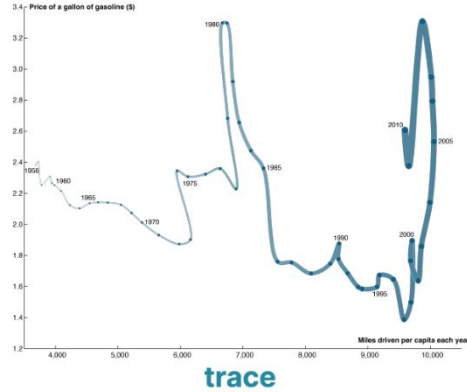
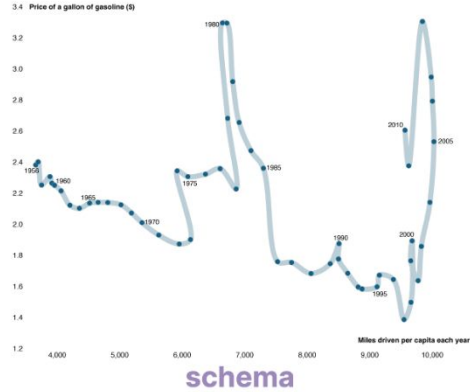


trace

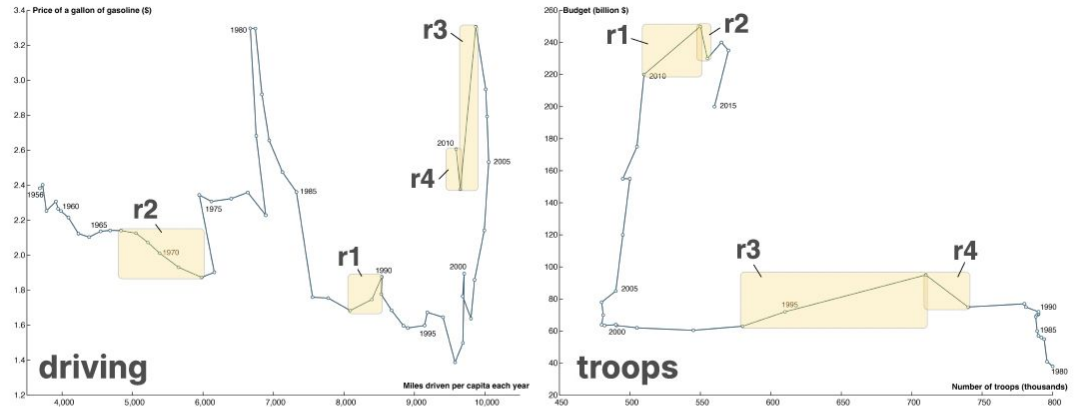


arrows

Goal 2: Combine approaches?







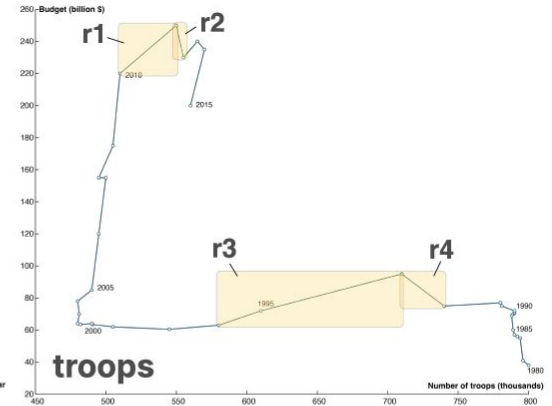
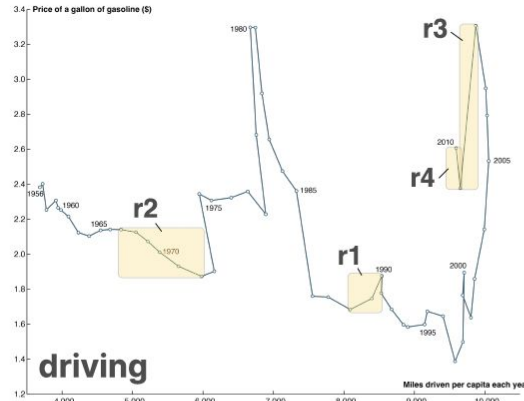
Conditions:



Questions:

- 1) “How did [the x variable] change in the highlighted region?”
- 2) “How did [the y variable] change in the highlighted region?”

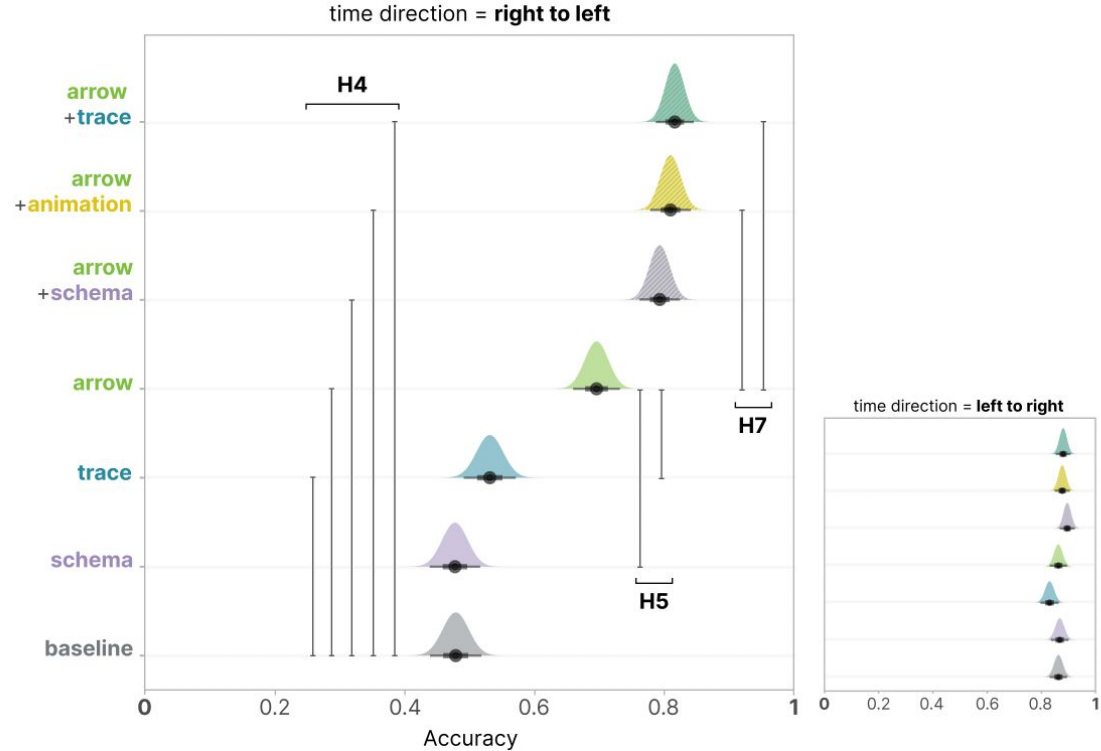
region	line shape	time direction	variable direction	
			x	y
r1		→	+ over time	+ over time
r2		→	+ over time	- over time
r3		←	- over time	- over time
r4		←	- over time	+ over time



Results:

Conclusions:

- Highlighting the correct expectation is more efficient than suppressing the incorrect schema
- Redundant coding improves performance



Take-home messages

Take-home message:

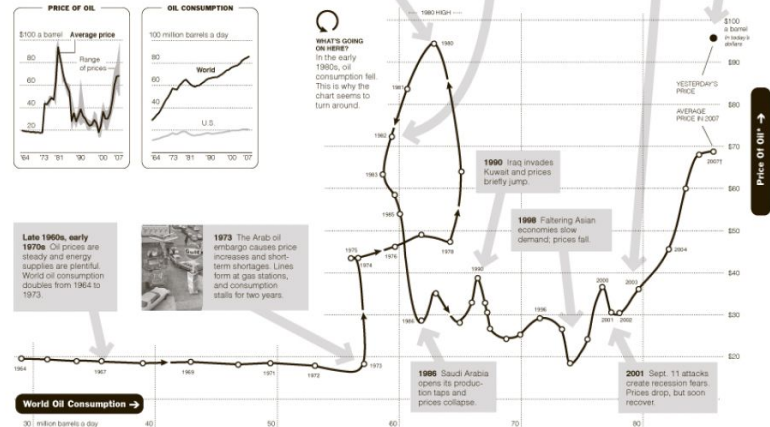
- Understanding of expectations can be leveraged to inform design interventions for a commonly misinterpreted visualization format

Design implications for connected scatterplots:

- Directional cues are necessary
- Arrows are the most effective cue to clarify the direction of time (that we tested)
- Redundant encoding helps

Oil's Roller Coaster Ride

This chart tracks the relationship between oil prices and oil consumption since 1964. Global oil consumption is shown on the horizontal axis. So, when consumption is increasing and prices are flat, the line moves straight right. And when prices are rising and demand stops growing, the line moves straight up.



* Average annual price of West Texas Intermediate crude oil, adjusted for inflation using the Consumer Price Index. Plotted prices (not spot prices) are shown before 1983.

Source: Energy Information Administration, Federal Reserve, Bureau of Labor Statistics, Rocky Mountain Institute

† Consumption forecast as of Nov. 6

ANALYSIS COURTESY OF THE NEW YORK TIMES. FIGURES PROVIDED BY THE ANALYSTS PROVIDED.