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:

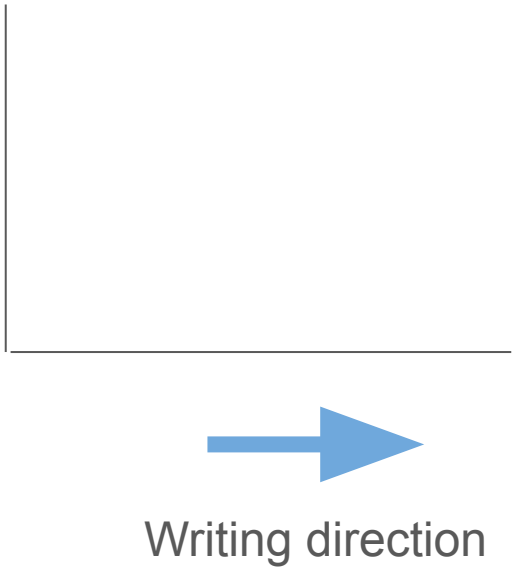
- "Look at our p = 0.052 result"!
- Graph literacy is measured and included in the statistical models, but results are not reported in the paper.

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



Graph schemas

Pinker’s theory:
graph schemas = mental structures that describe how a data visualization is interpreted

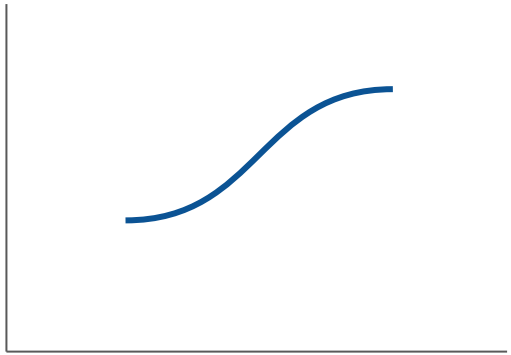
Result: Schemas "prime" specific encoding rules

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

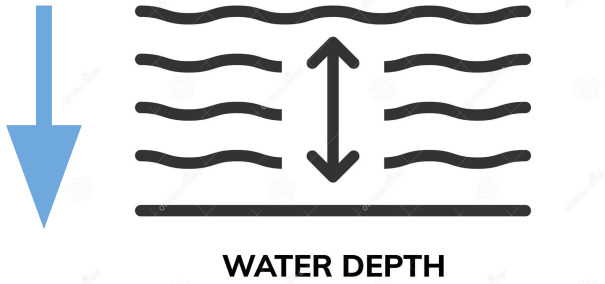


Conceptual Metaphors

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

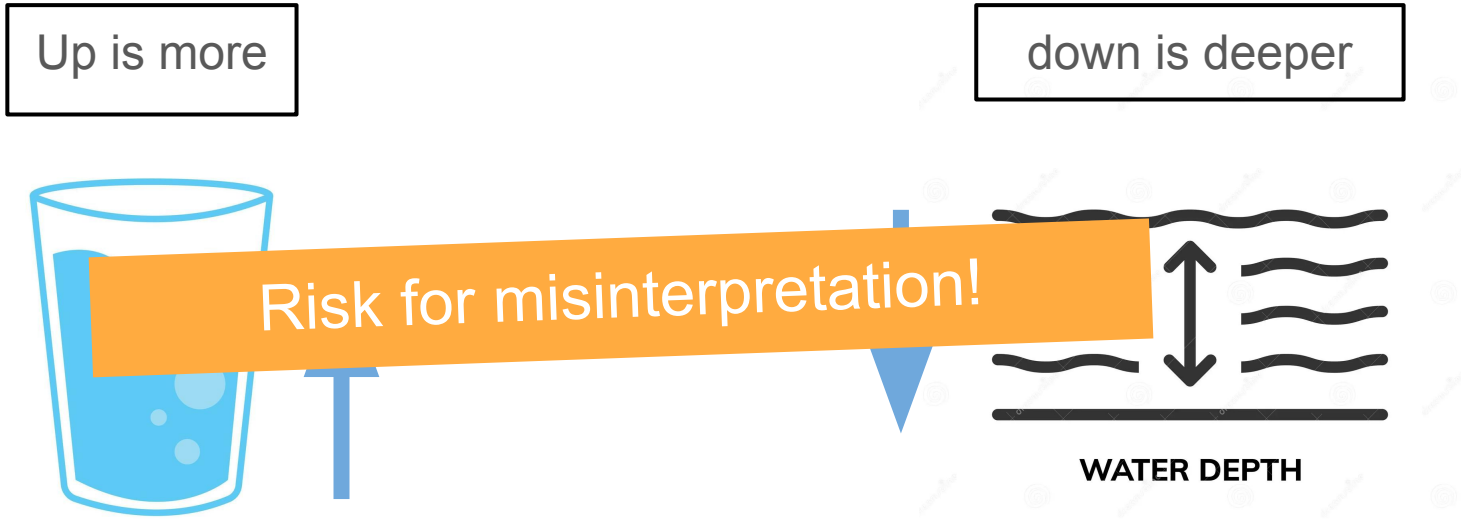
down is deeper



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

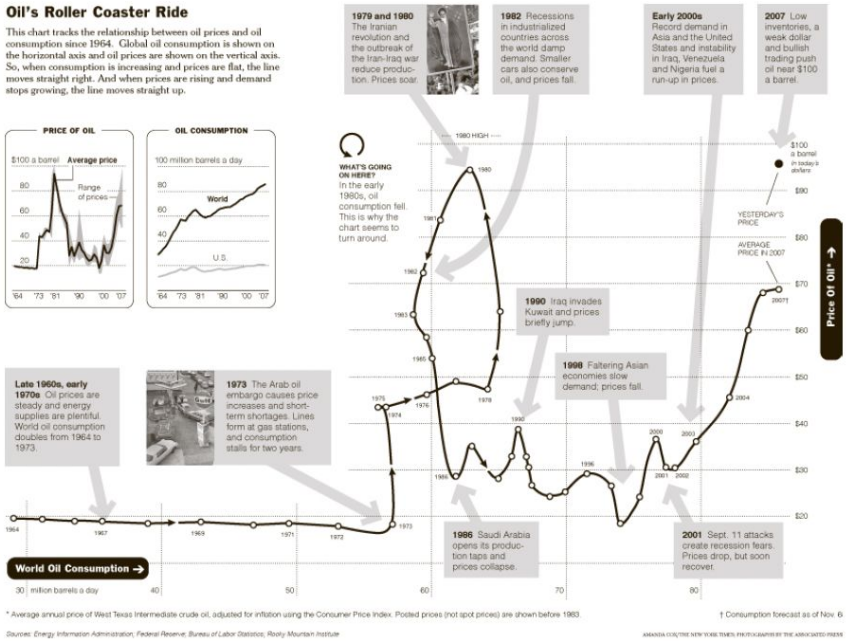


Fig. 1: *New York Times*' Oil's Roller Coaster Ride visualization depicts the relationship between world oil consumption and oil price over time [9].

Expectation conflict!

Connected scatterplot:

- Time is a line

Line graph schema:

- Right is later

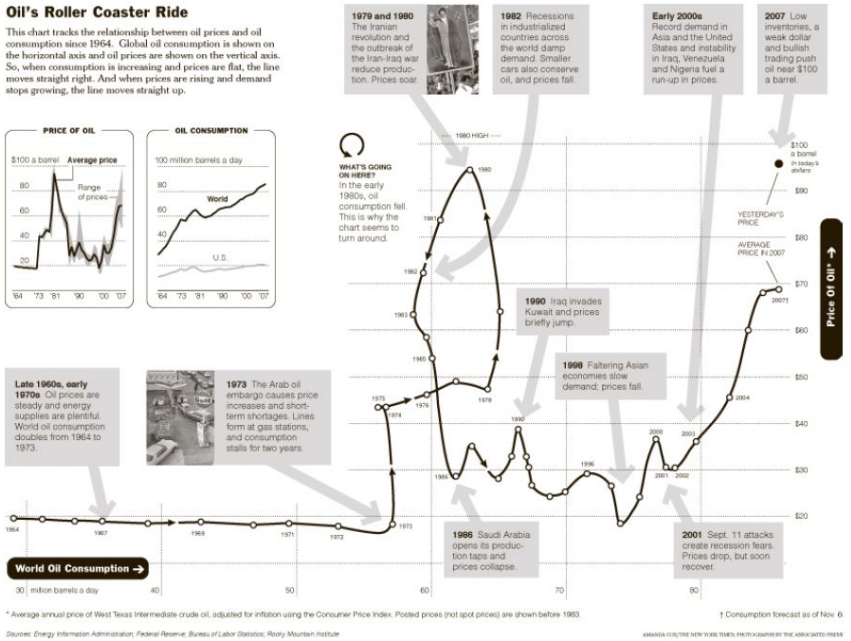


Fig. 1: *New York Times*' Oil's Roller Coaster Ride visualization depicts the relationship between world oil consumption and oil price over time [9].

Goal of this study

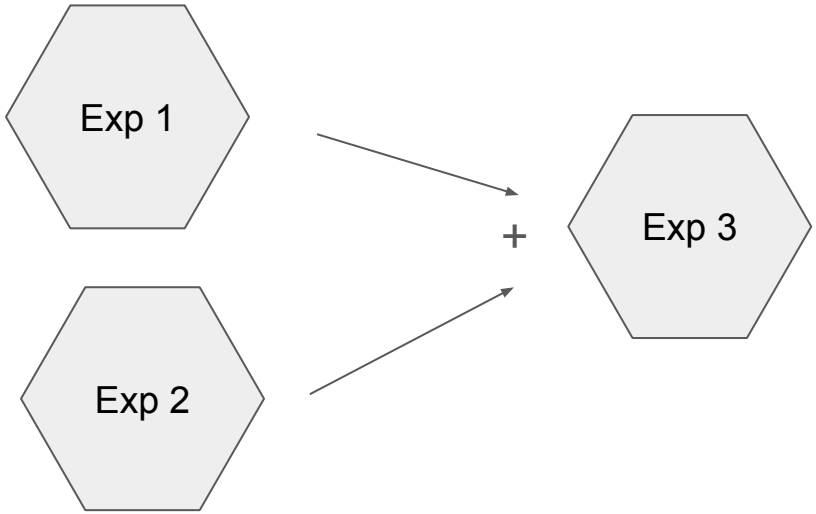
Designershave proposed various design strategies to mitigate misinterpretation of connected scatterplots

But were never tested.. Until now!

Two approaches:

1) Suppress incorrect chart-type schema

2) Emphasize correct expectation



Experiment 1

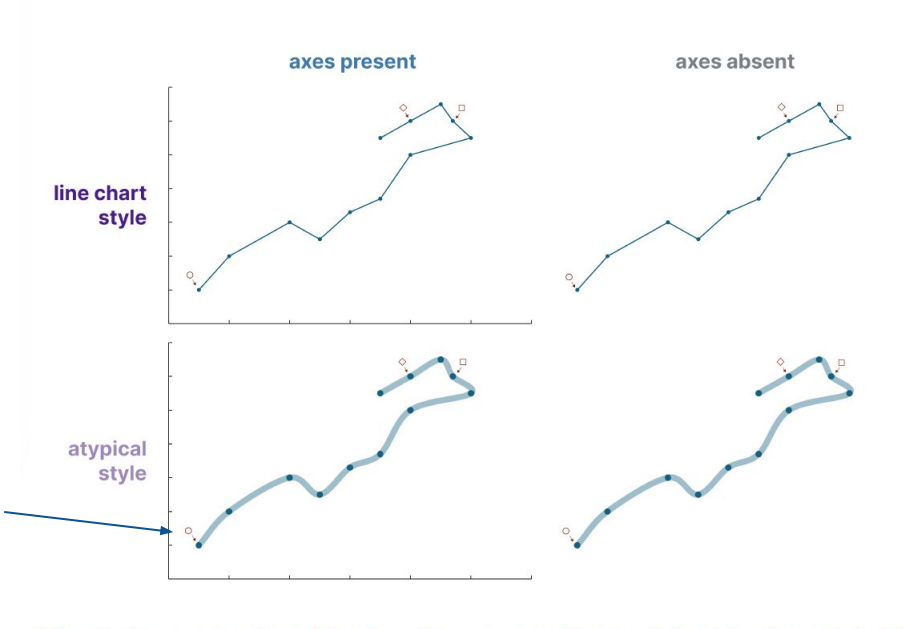
Approach 1: Suppress incorrect chart-type schema

Reduce visual similarity of connected scatterplots to line charts:

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

In task instruction:

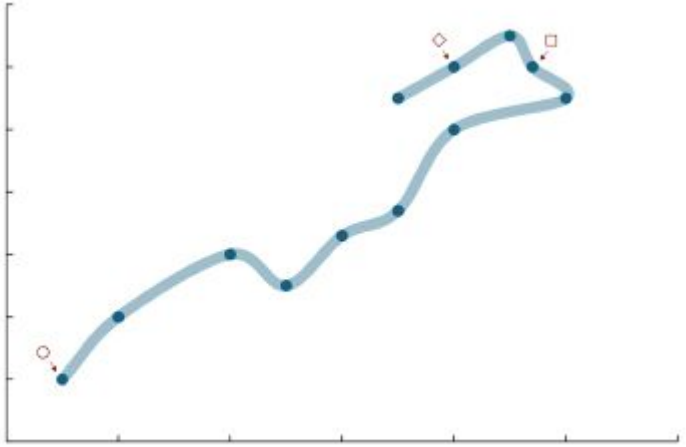
This is the first year in the dataset.



Tasks:

1) Chart-type judgment:
   What type of chart is this?
   [textbox]

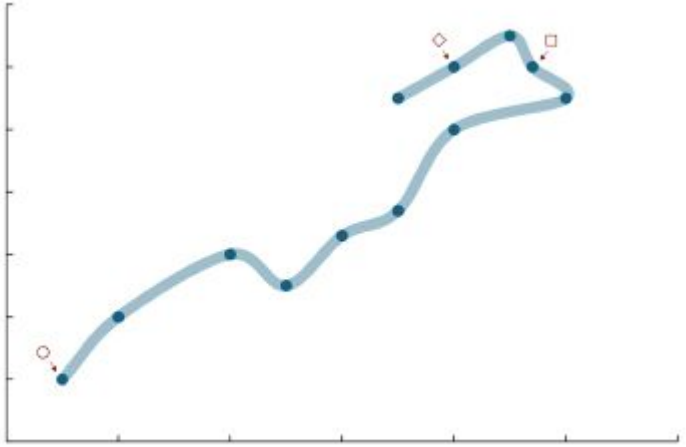
2) Rule expectation:
   Indicate which one of the two highlighted points show data for a later year than the other, based on your intuition?
   - *Right is later*
   - *Time is a line*



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

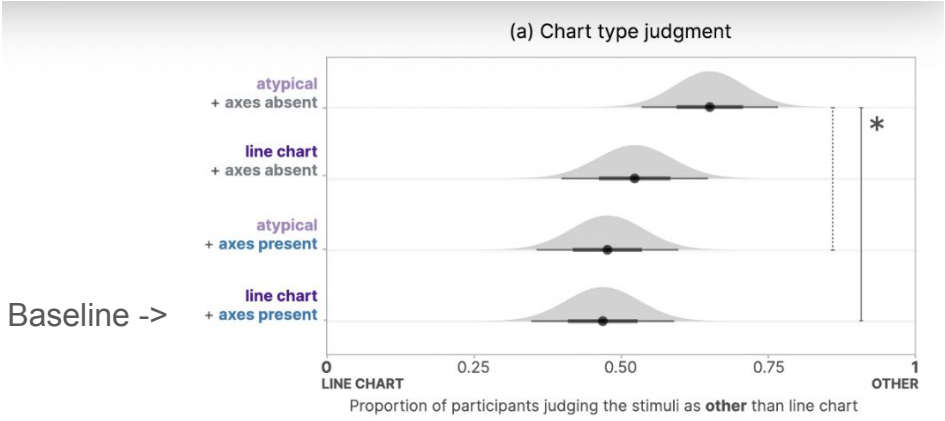
- 18 years and older
- Fluent in English
- Approval rate of >= 80%
- Used desktop

Exclusion based on:

- Attention check
- Task comprehension check

Results - chart type judgment

What type of chart is this?

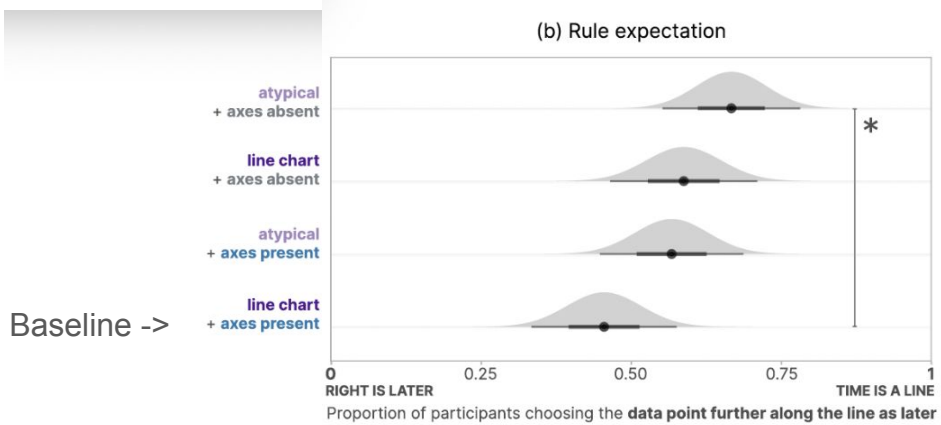


Baseline ->

+17.5%

Results - Rule expectation

What point is the later year?



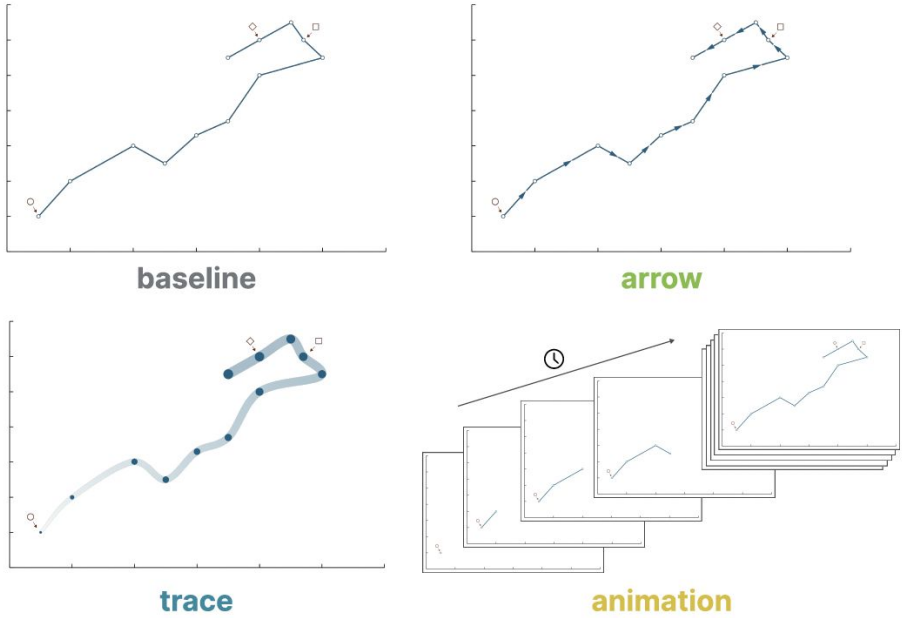
Baseline ->

+20.6%

Experiment 2

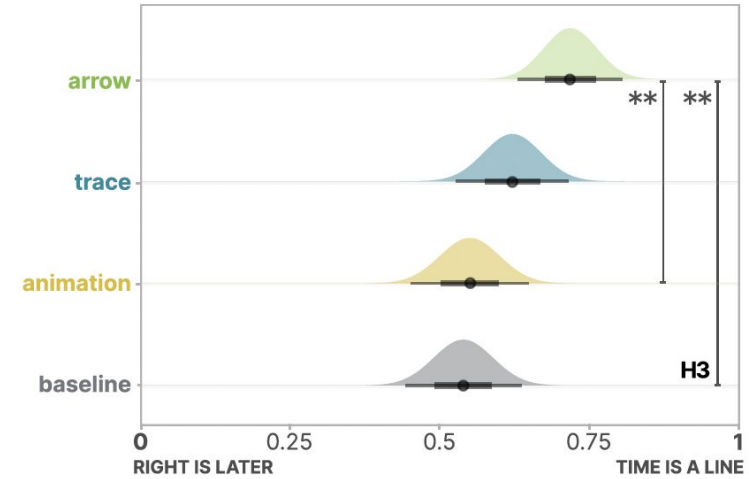
Approach 2: Emphasize correct expectation with directional cues

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



Results - Rule expectation

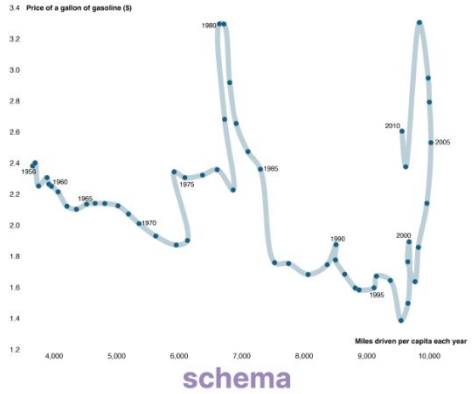
What point is the later year?



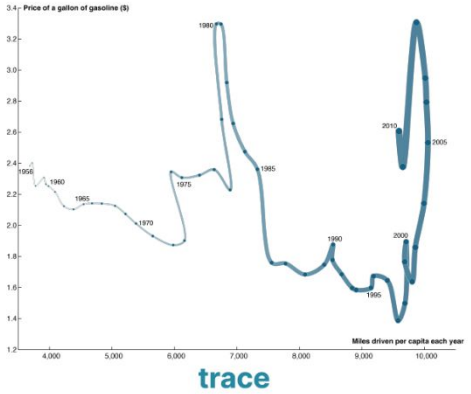
Proportion of participants choosing the data point **further along the line** as later

Experiment 3

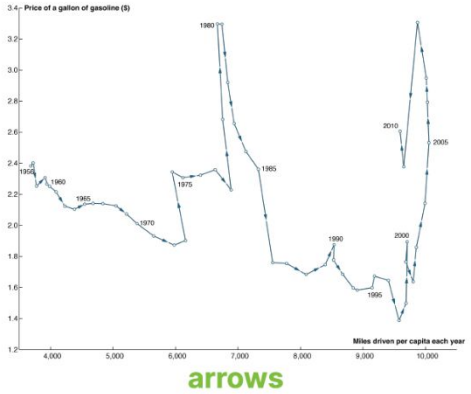
Goal 1: Generalization to more realistic graphs



schema

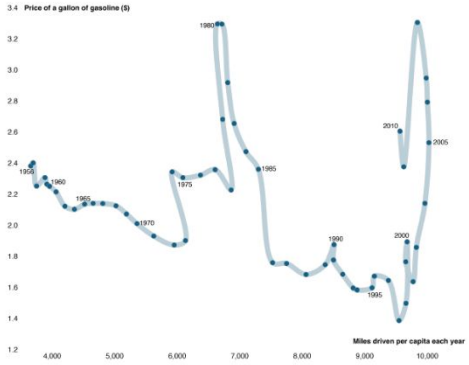


trace

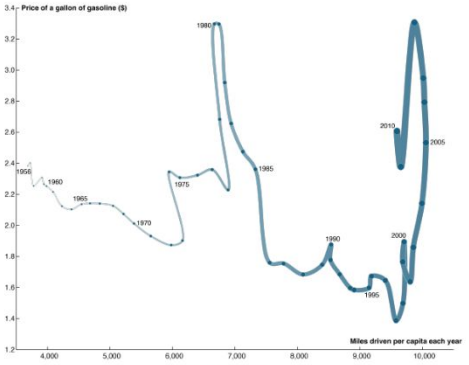


arrows

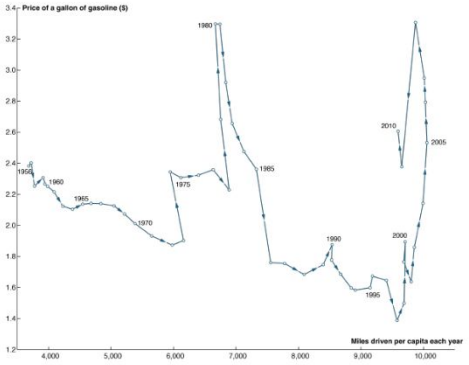
Goal 2: Combine approaches?



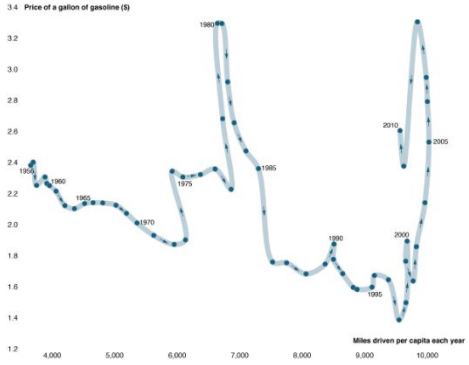
schema



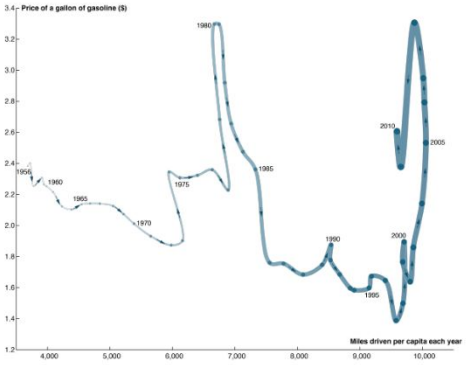
trace



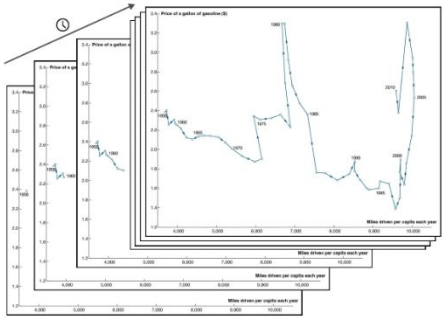
arrows



arrows + schema

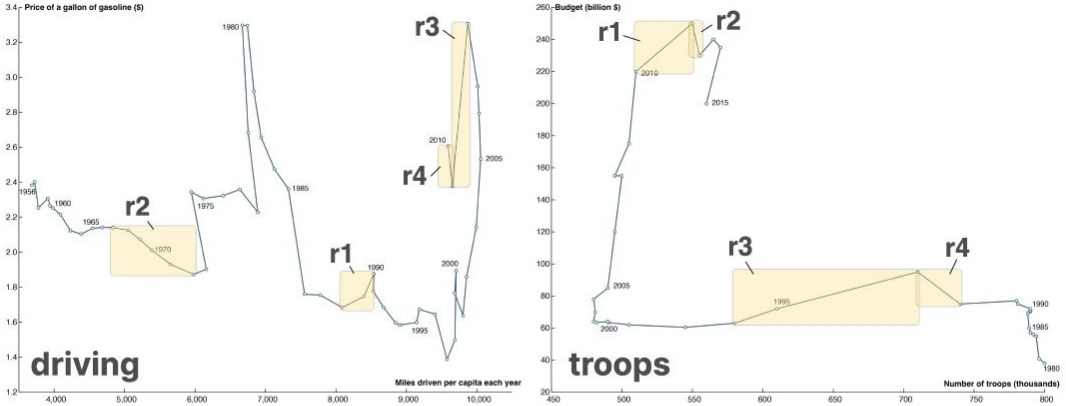


arrows + trace



arrows + animation

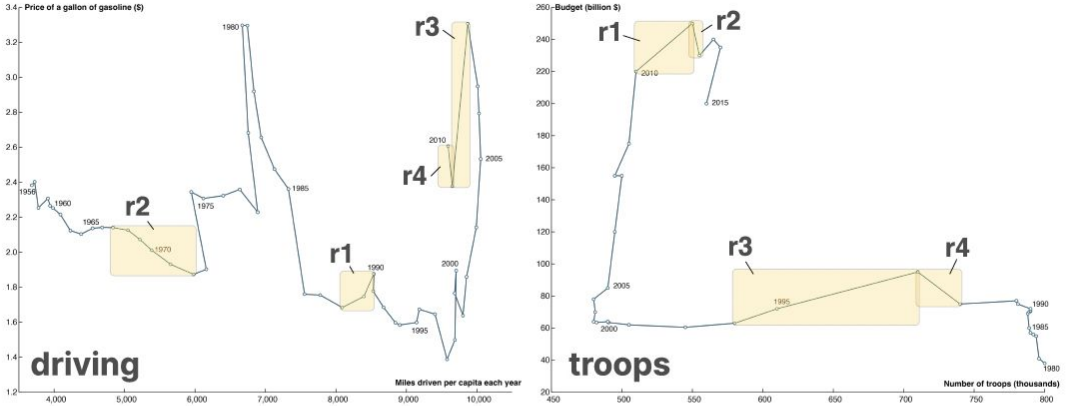
Conditions:


Price of a gallon of gasoline ($)
Miles driven per capita each year
driving
r1
r2
r3
r4
Budget (billion $)
Number of troops (thousands)
troops
r1
r2
r3
r4

Questions:

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

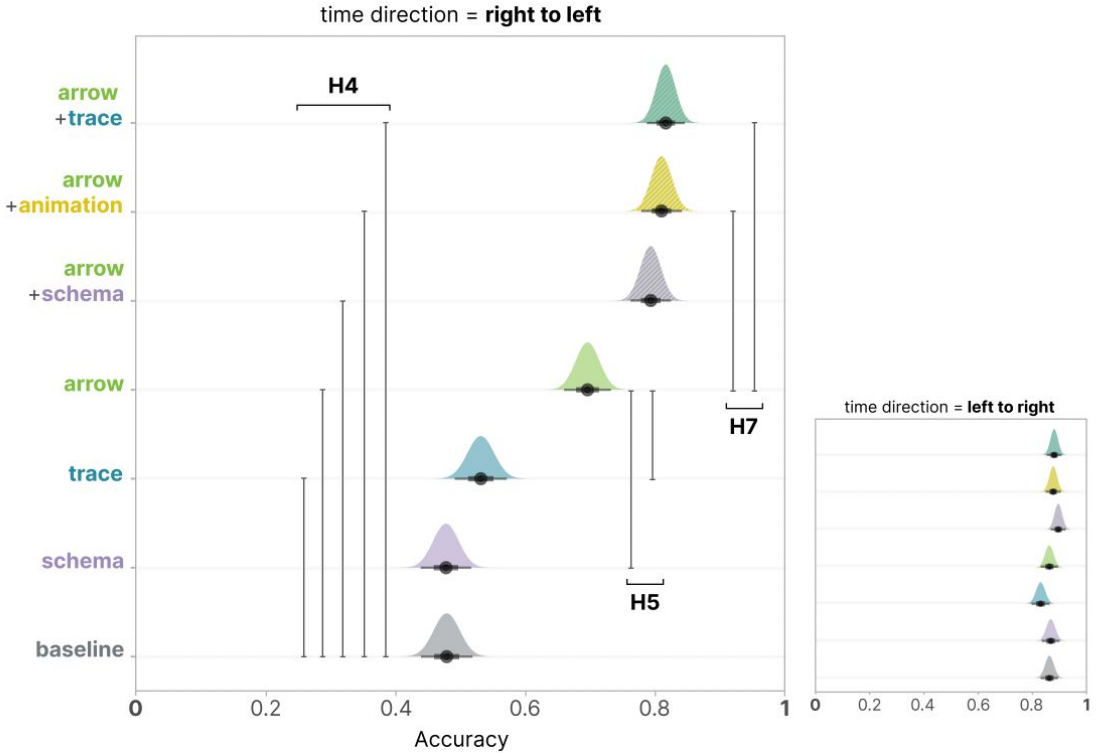
region	line shape	time direction	variable direction	
			x	y
r1	/	→	+ over time	+ over time
r2	\	→	+ over time	- over time
r3	/	←	- over time	- over time
r4	\	←	- over time	+ over time



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