How do communicative goals guide which data visualizations people think are effective?

*Holly Huey[a], *Lauren A. Oey[a], Hannah S. Lloyd[a], & Judith E. Fan[a,b]

[a]University of California San Diego, [b]Stanford

Overview

Data visualizations are powerful communication tools.

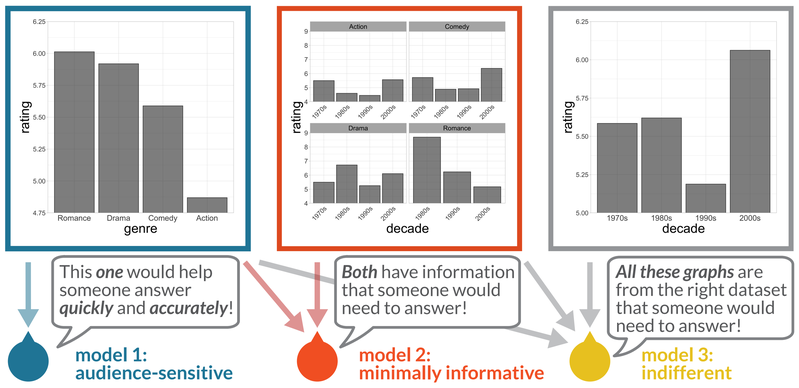
Psychophysical studies have largely focused on comprehension of data visualizations rather than how people *generate* informative ones.

We explored how well people can select graphs that make it easy for other people to understand key patterns in data.

Hypotheses

We considered three strategies people might use:

"On average, what is the rating of Action movies?"



This *one* would help someone answer *quickly* and *accurately*!

model 1: audience-sensitive

Both have information that someone would need to answer!

model 2: minimally informative

All these graphs are from the right dataset that someone would need to answer!

model 3: indifferent

Stimuli

Graphs & questions generated from 8 datasets varying in topic (e.g., storms, test scores, flowers)

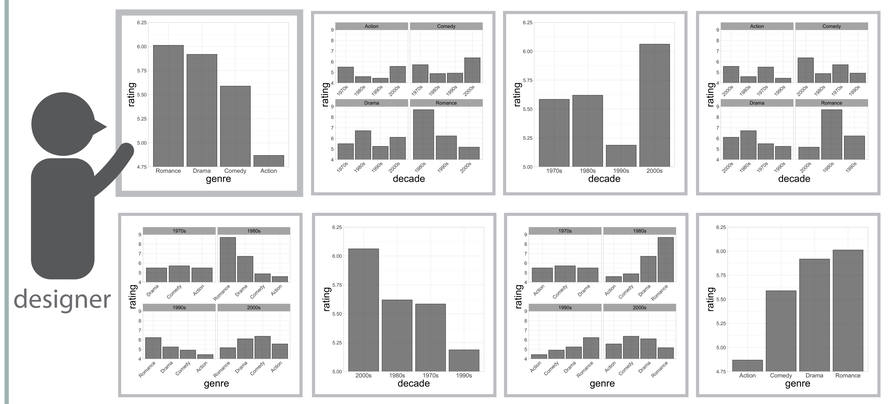
question goal	example dataset: movie ratings
retrieve values needs 1 panel to answer	On average, what is the rating of Action movies?
retrieve values needs 2+ panels to answer	On average, what is the rating of 1990s movies within the Action genre?
make comparisons needs 1 panel to answer	On average, how much higher are ratings of Drama movies compared to Comedy movies?
determine range needs 1 panel to answer	How much higher are ratings of movies from the decade with the highest ratings compared to the decade with the lowest rating?

Methods

Graph selection task

n=398

Do people believe different graphs are better for answering different questions?

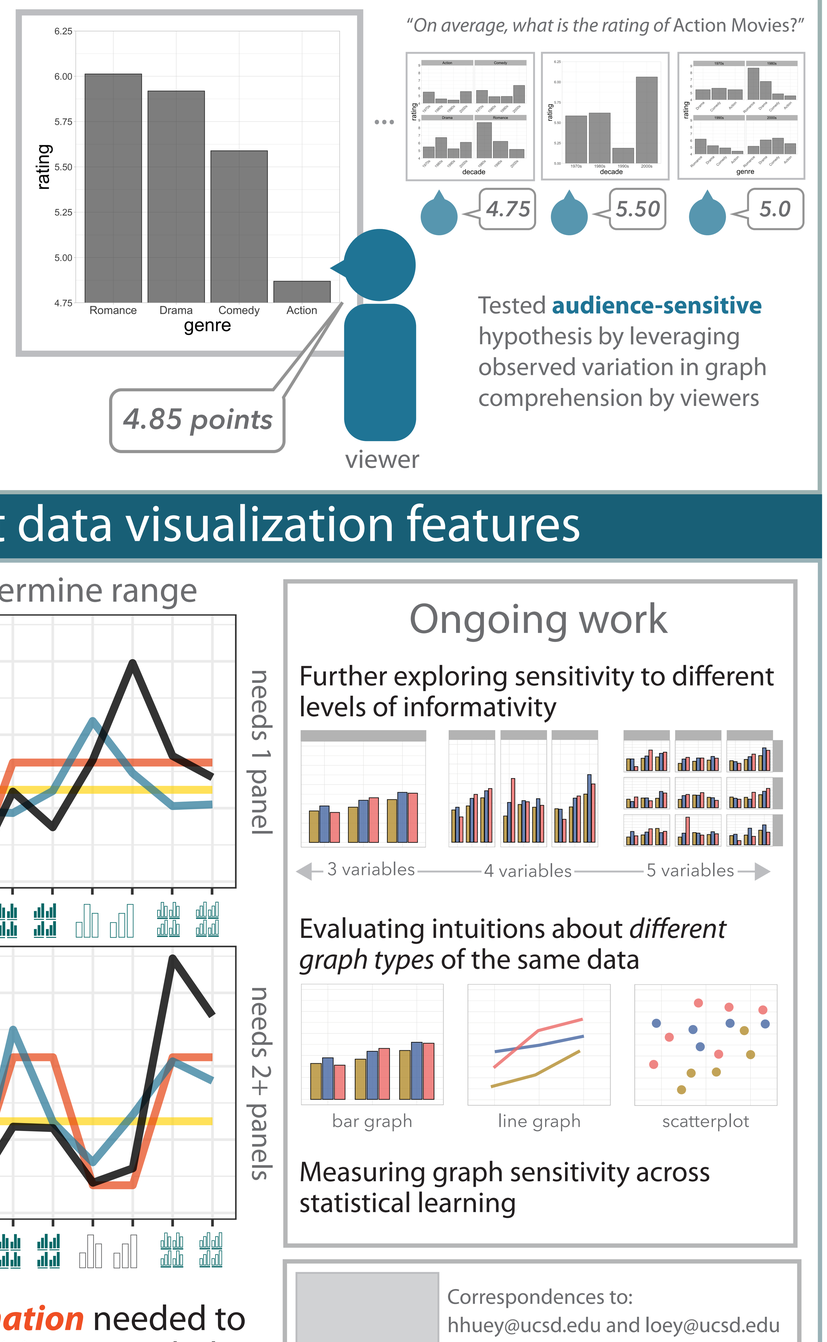


Select the best graph to answer:
"On average, what is the rating of Action Movies?"

Graph comprehension task

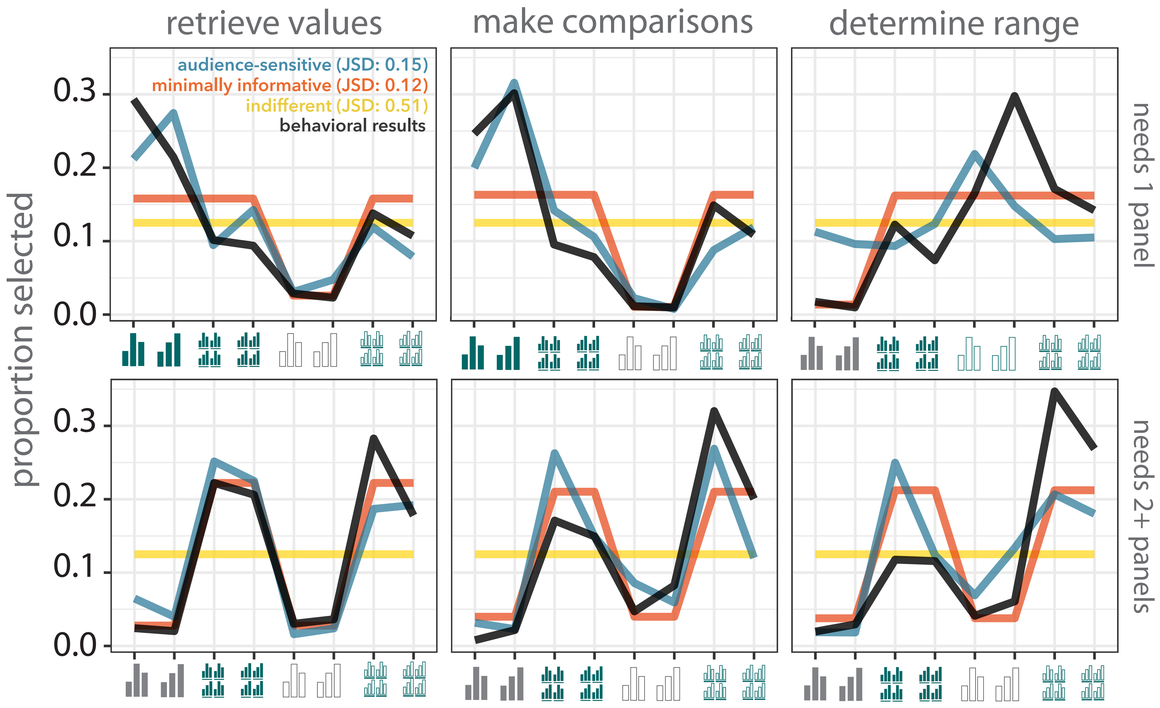
n=542

Which graphs are better at helping people answer those questions?



Tested **audience-sensitive** hypothesis by leveraging observed variation in graph comprehension by viewers

Results: Evaluating sensitivity to different data visualization features



People selected graphs with the ***minimal information*** needed to answer prompts, but were not necessarily sensitive to graded variation in audience comprehension

Ongoing work

Further exploring sensitivity to different levels of informativity

Evaluating intuitions about *different graph types* of the same data

Measuring graph sensitivity across statistical learning

Correspondences to:
hhuey@ucsd.edu and loey@ucsd.edu

data: https://github.com/cogtoolslab/davinci_public2023