Learning to Automate Chart Layout Configurations
用机器学习自动化图表布局

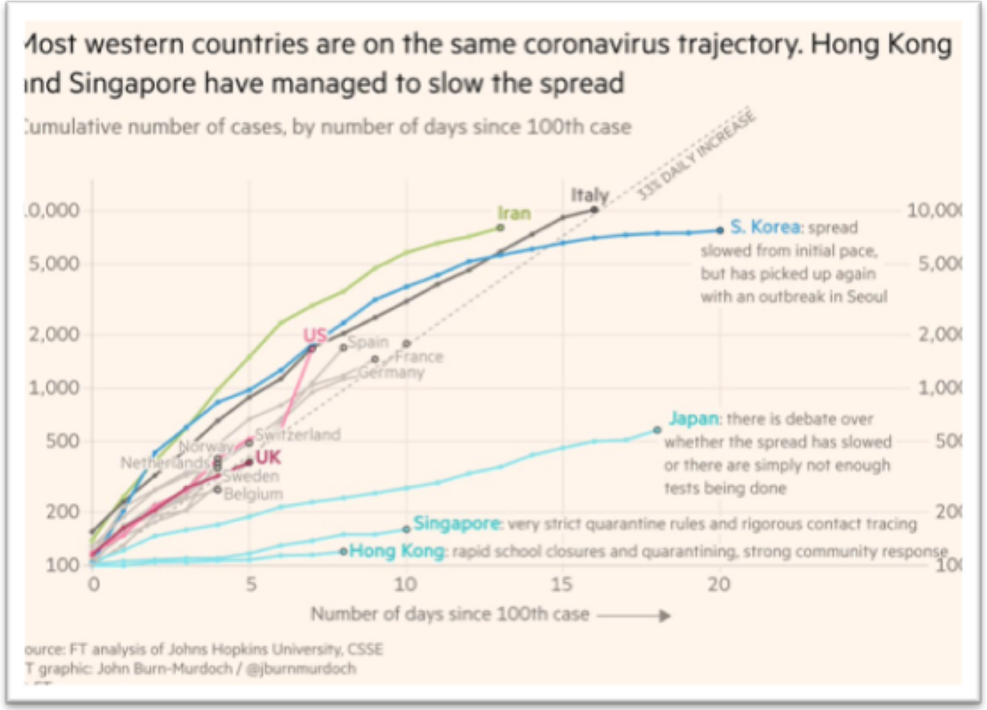
WU Aoyu

Homepage: http://awuac.student.ust.hk

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

0. Background: Chart

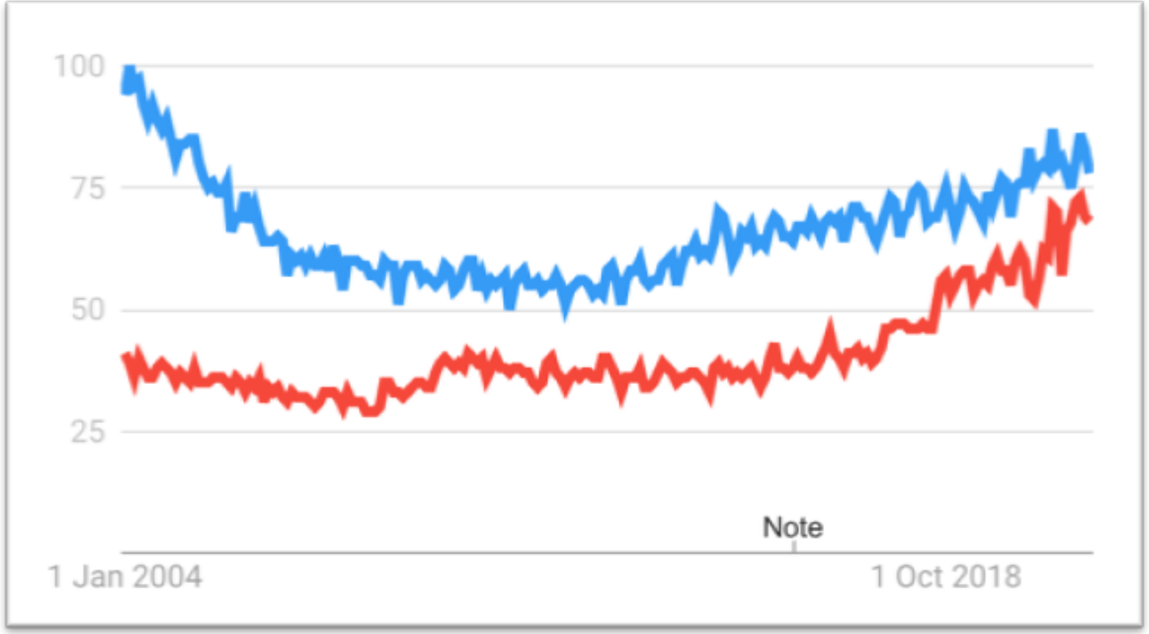
Charts are easy to read, and arguably one of the most easiest way for the masses to assess data.



By far the most visited page in NY Times

0. Background: Chart

Charts have been created and shared at an unprecedented speed.



Google Search Trend Globally: Data v.s. Chart

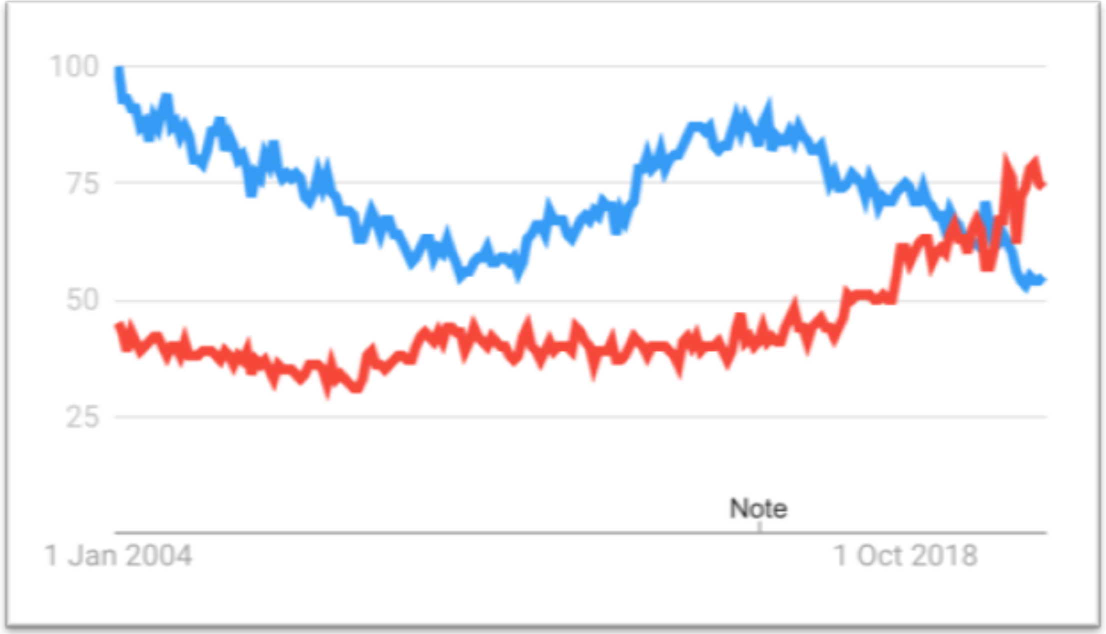
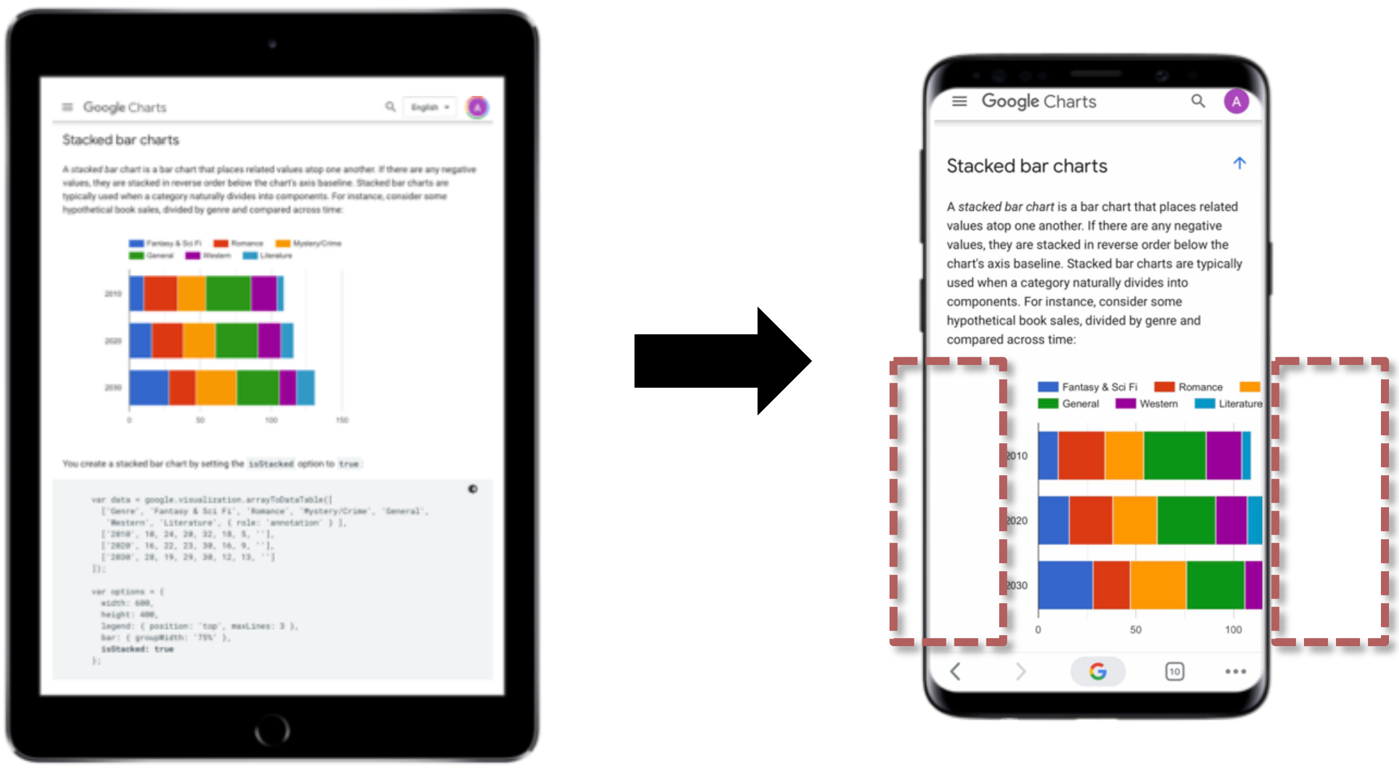


Chart has surpassed Image globally in Google Search Trend since Jan 2020

0. Background: Chart Layout

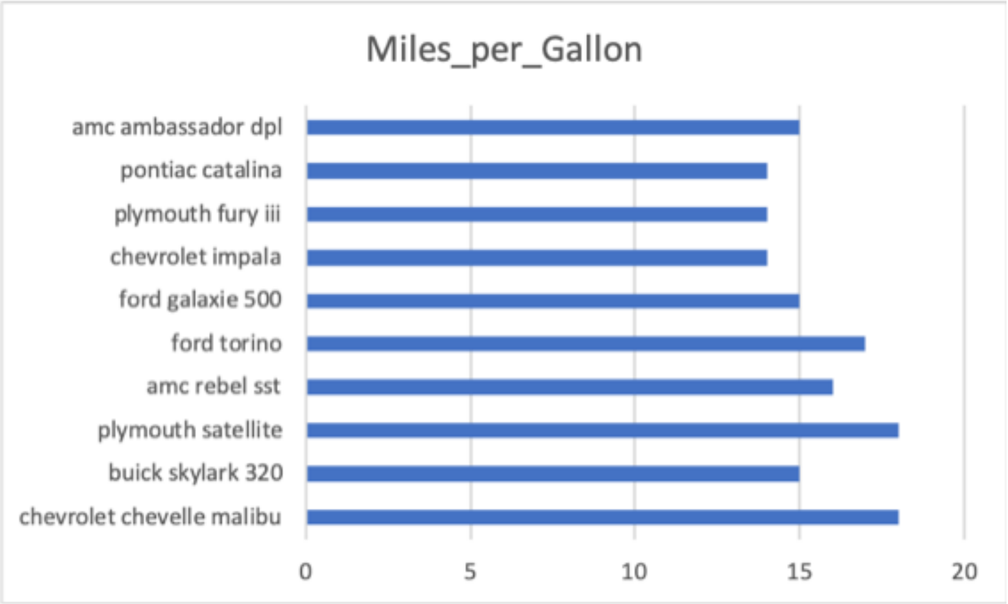
Chart layouts directly influence the readability and aesthetics.

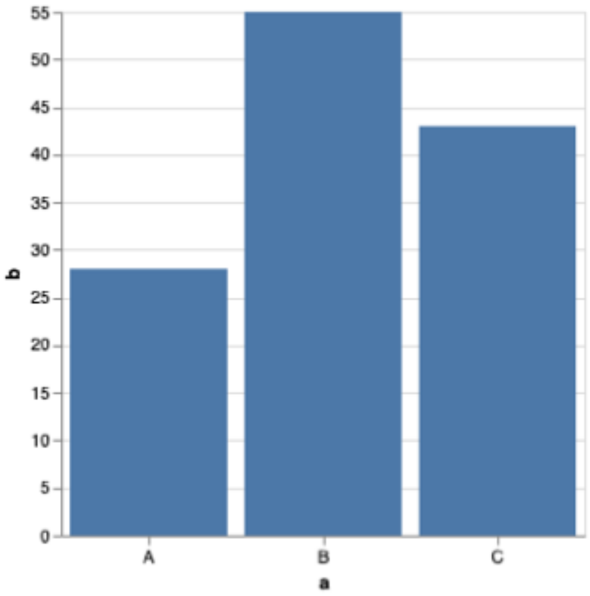


Mobile friendliness

0. Background: Chart Layout

Chart layouts directly influence the readability and <u>aesthetics</u>.





Default styles in Excel (charting software) and Vega-Lite (charting library)

0. Background: Chart Layout

Manually adjusting chart layouts faces problems.

Responsive Settings

Multiple Parameters

Poor User Experience (UX)

0. Background: Chart Layout

Manually adjusting chart layouts faces problems.

Responsive Settings

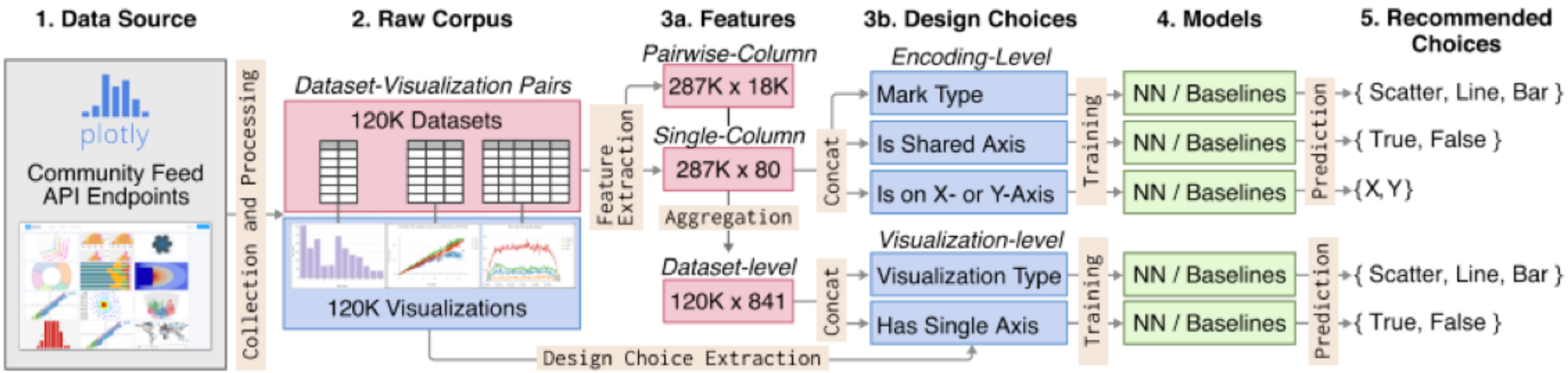
Multiple Parameters

Poor User Experience (UX)

How to automatically optimize multiple parameters for chart layouts given constraints (such as screen widths)?

1. Related Work: Visualization Recommendation

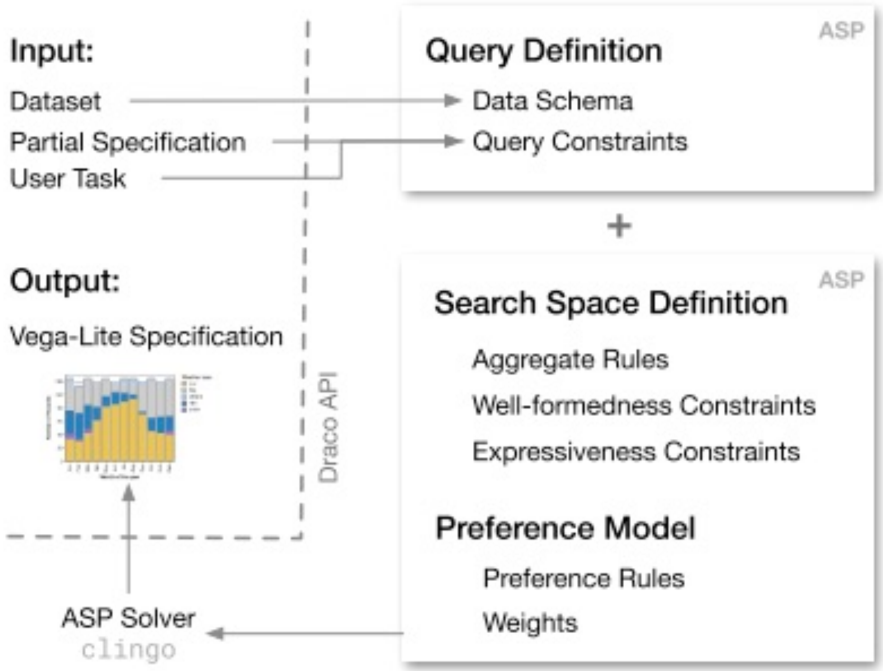
Visualization recommendation is the problem of automatically recommending visual encodings.



VizML [CHI 2019] – Machine Learning

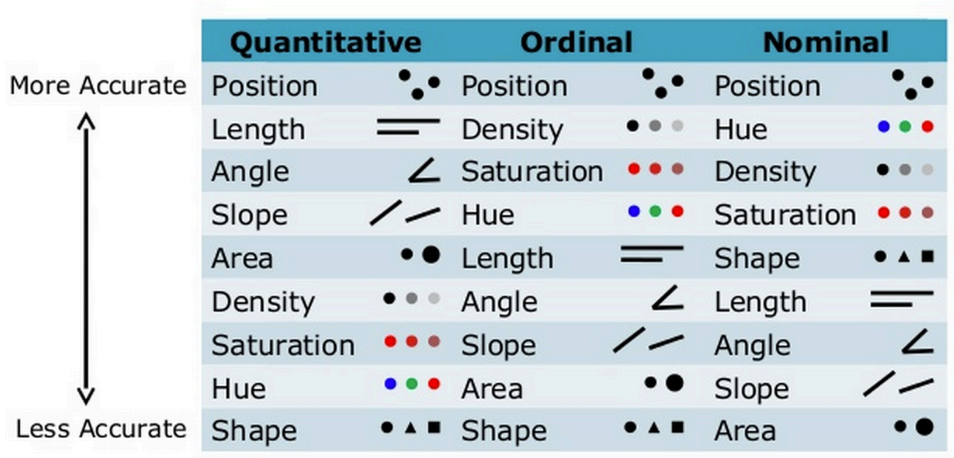
1. Related Work: Visualization Recommendation

Most work focused on recommending <u>data-encodings</u>.



Draco [VIS 2018]

Incorporating decision rules from design knowledge.

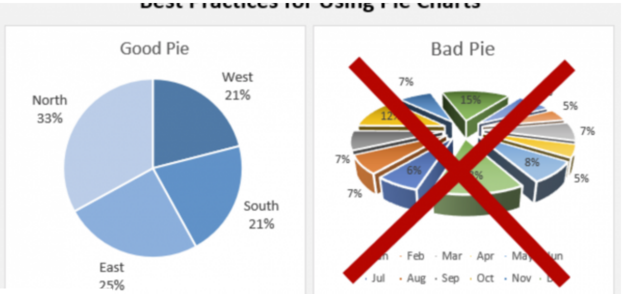
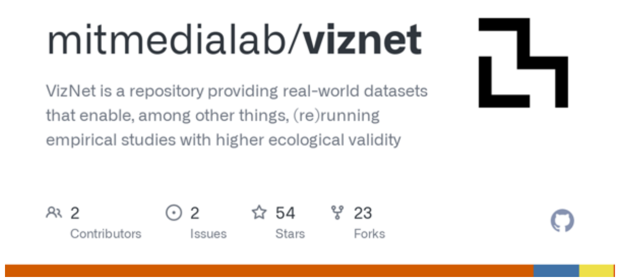
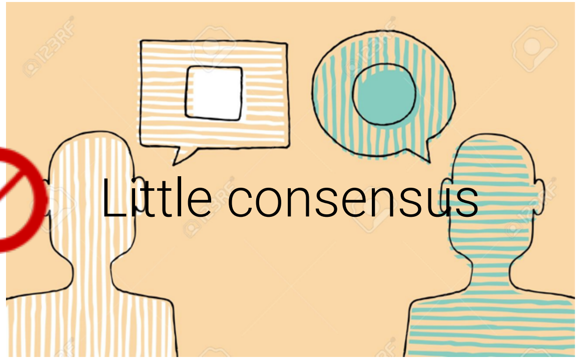


Visual encoding by data type

	Quantitative	Ordinal	Nominal
More Accurate	Position	Position	Position
	Length	Density	Hue
	Angle	Saturation	Density
	Slope	Hue	Saturation
	Area	Length	Shape
	Density	Angle	Length
	Saturation	Slope	Angle
	Hue	Area	Slope
Less Accurate	Shape	Shape	Area

1. Related Work: Visualization Recommendation

Less research addresses non-data-encodings (such as layouts).

	Rule-Based Approach	ML-based Approach
Data Encodings	 Much research about what's good/bad data encodings.	 Much dataset
Non-data Encodings (Layout)	 Little consensus	Little dataset

2. Rule-based Approach

Problem: Automated responsive visualization

How to automatically optimize multiple parameters for chart layouts given a screen size?

What are good/bad chart layouts?

- Mobile-friendliness issue

2. Rule-based Approach

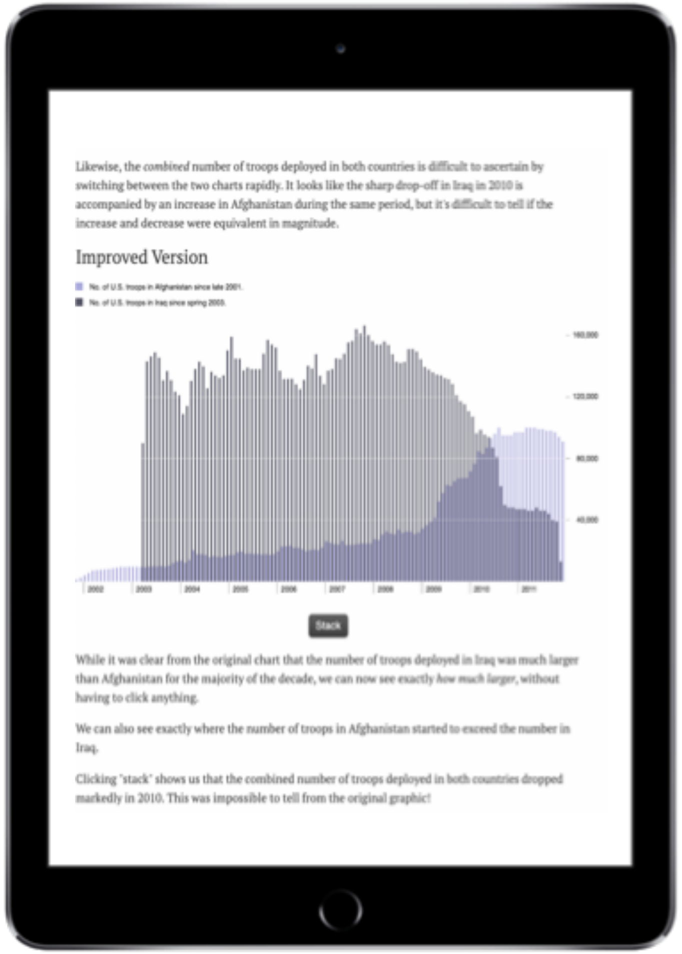
Mobile-friendliness issue

- Out-of-view-box

2. Rule-based Approach

Mobile-friendliness issue

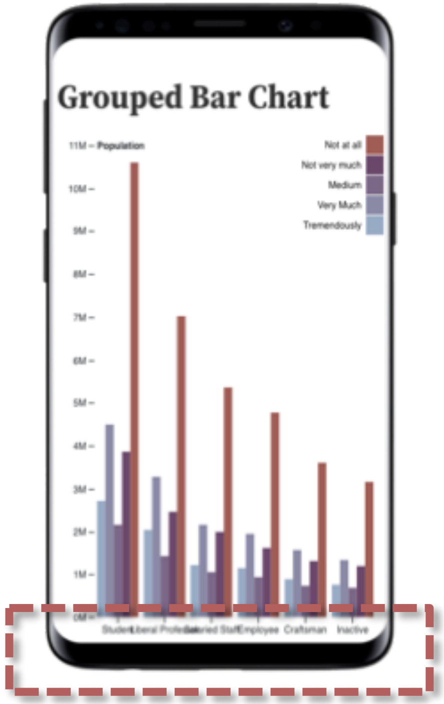
- Out-of-view-box
- Unreadable font-size



2. Rule-based Approach

Mobile-friendliness issue

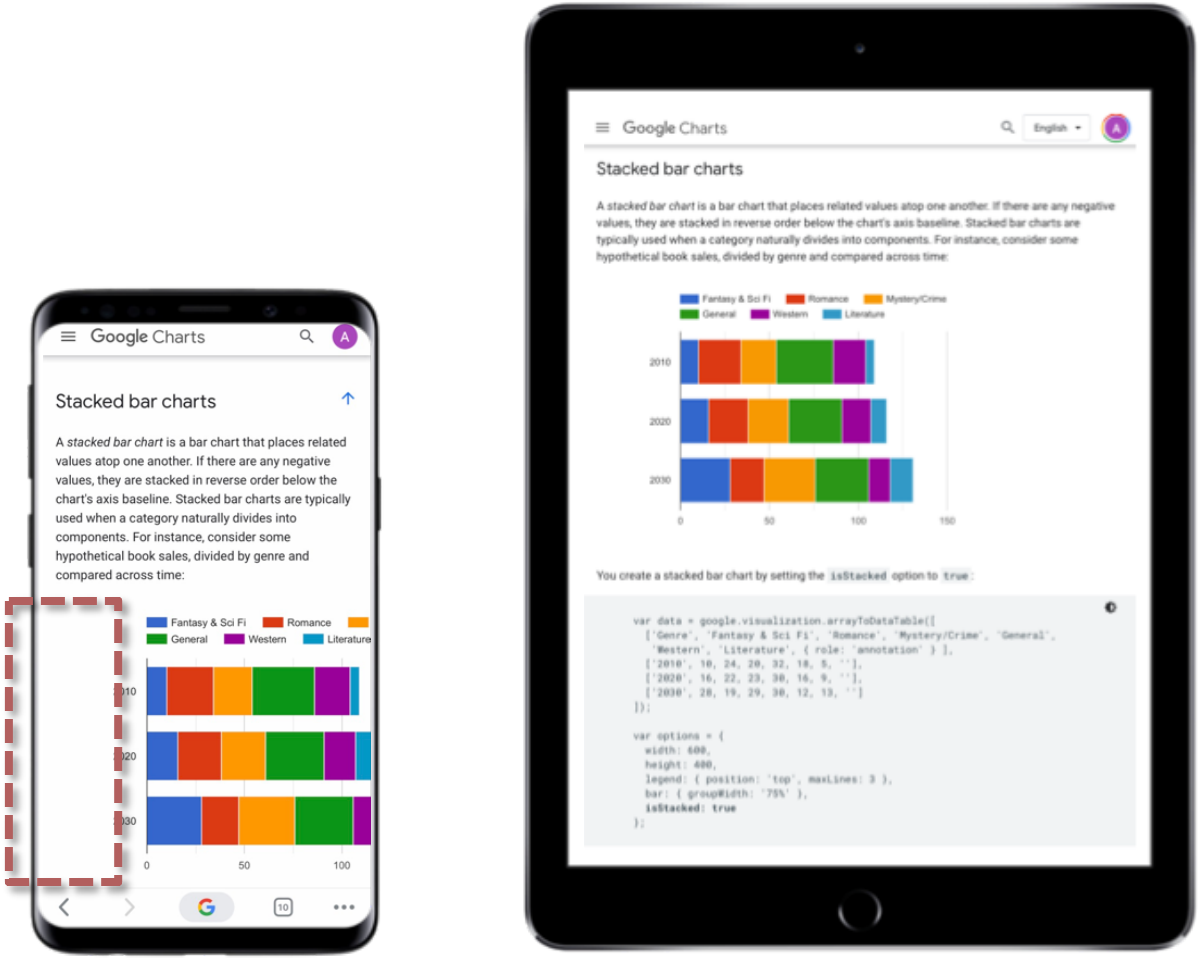
- Out-of-view-box
- Unreadable font-size
- Overlapping text



2. Rule-based Approach

Mobile-friendliness issue

- Out-of-view-box
- Unreadable font-size
- Overlapping text
- Unwanted white space
- …



2. Rule-based Approach

Approach: Reinforcement Learning

Environment - *Chart*

State(S) – *Mobile-friendly Issue*
Reward(R) – *Change of loss (L)*

Example:
S: *SmallFontSize*
L: *8px*

Action(A) – *Modify the chart*
Policy P(A|S)

Example:
A1: *IncreaseFontSize*
A2: *DecreaseFontSize*

2. Rule-based Approach

Approach: Reinforcement Learning

Environment - *Chart*

State(S) – *Mobile-friendly Issue*
Reward(R) – *Change of loss (L)*

Example:
S: *SmallFontSize*
L: *8px*

Action(A) – *Modify the chart*
Policy P(A|S)

Example:
A1: *IncreaseFontSize*
A2: *DecreaseFontSize*

2. Rule-based Approach

Approach: Reinforcement Learning

Environment - *Chart*

State(S) – *Mobile-friendly Issue*
Reward(R) – *Change of loss (L)*

Example:
S: *SmallFontSize*
L: *8px + 1px = 9px*
R: 1px

Increase P(A1|SmallFontSize)

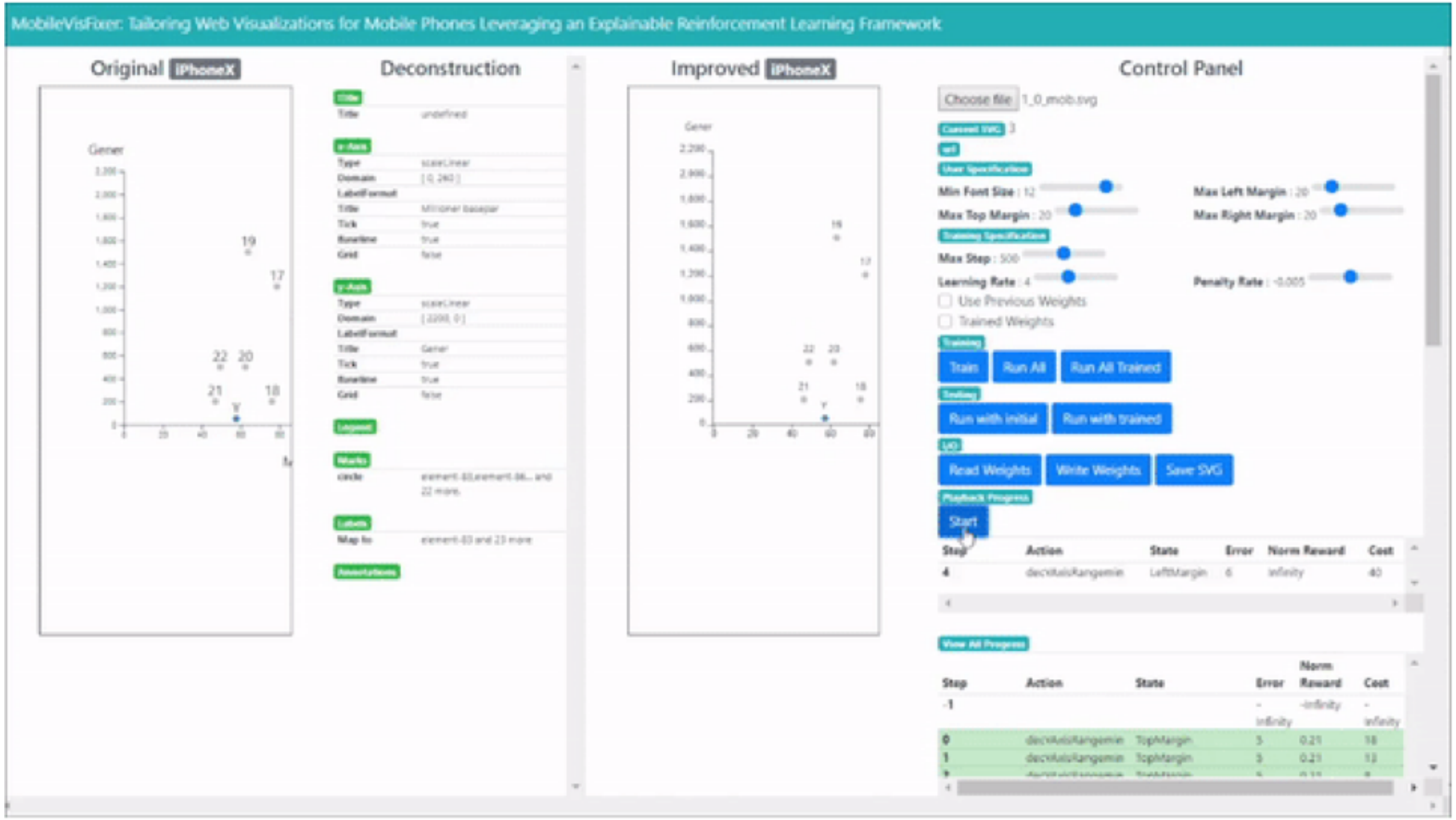
Action(A) – *Modify the chart*
Policy P(A|S)

Example:
A1: *IncreaseFontSize*
A2: *DecreaseFontSize*

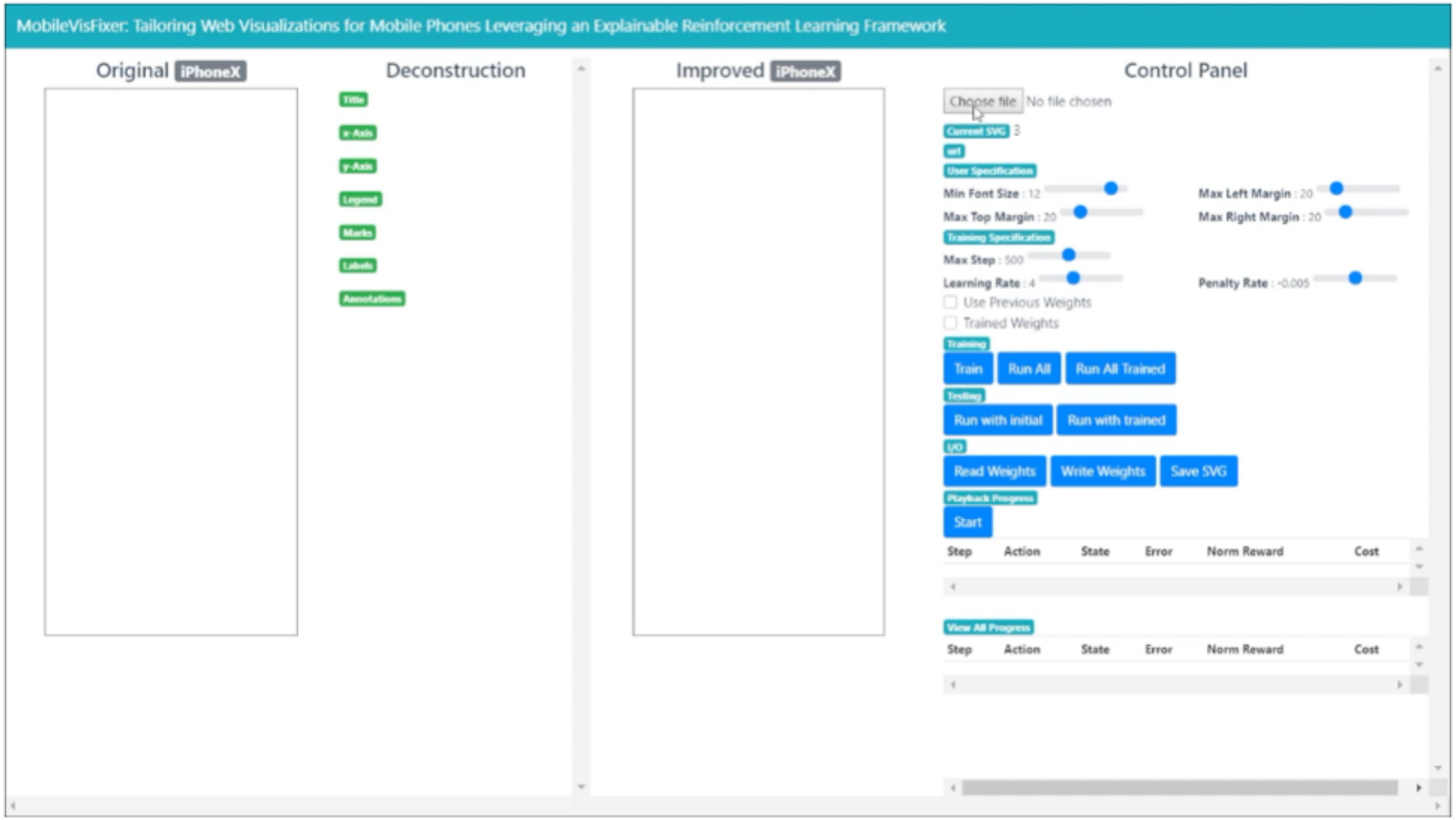
2. Rule-based Approach

29 states

23 actions

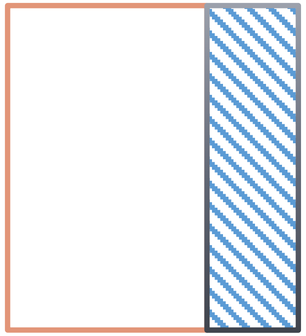


2. Rule-based Approach: Limitation



When the axis exceeds the right of the view box, the agent will increase the YRangeMin.

2. Rule-based Approach: Limitation



The cost was the area exceeding the view box.

Compress height reduces this cost as well.

2. Rule-based Approach: Limitation

The performances highly rely on manually-crafted rules and/or cost functions.

Even seemingly reasonable cost functions might be problematic.

Can we directly learn layout qualities from human perception data?

3. Machine-Learning Based Approach

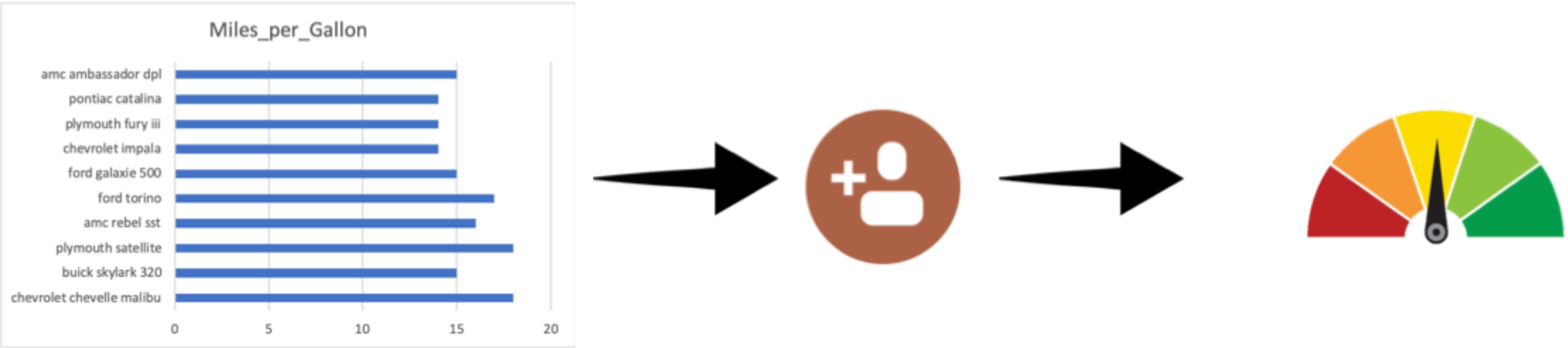


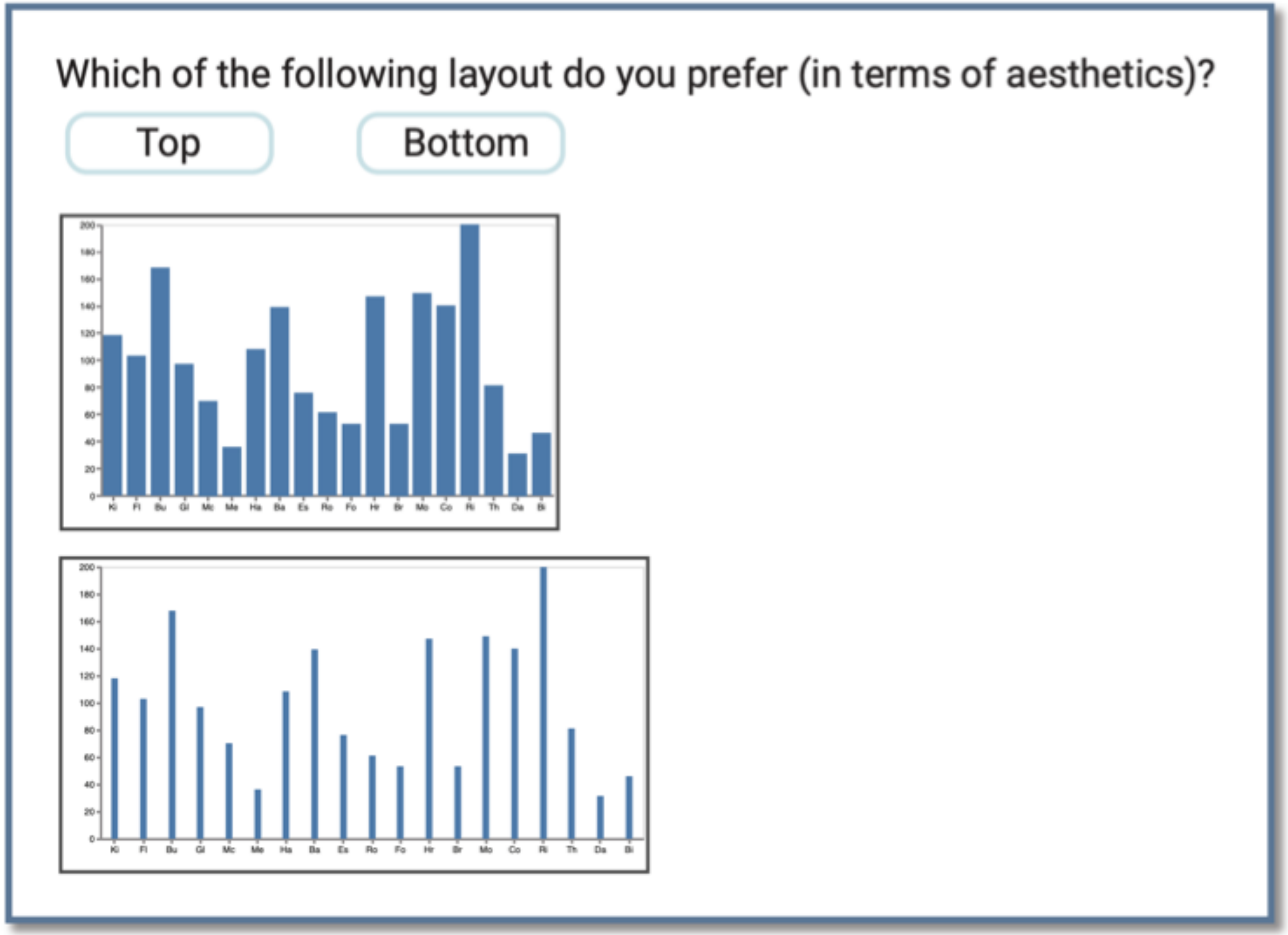
Chart Human Score

Difficulty in giving an accurate score

Inconsistent scoring scales among participants

3. Machine-Learning Based Approach

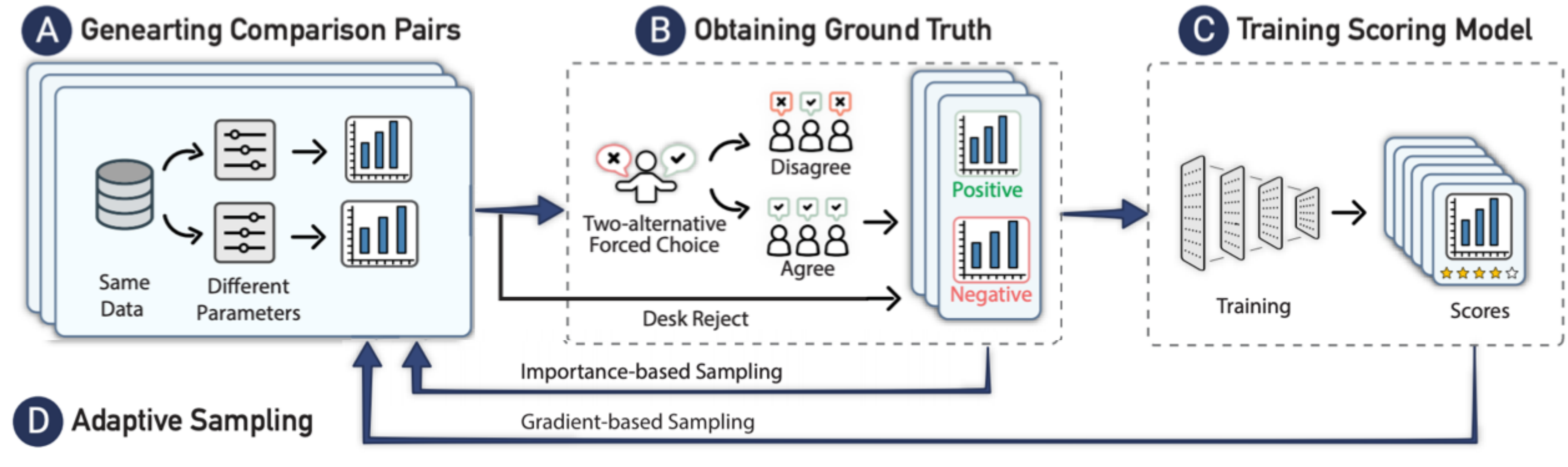
<u>Comparison</u> is an easier task than scoring a chart.



Two-forced Alternative Choice

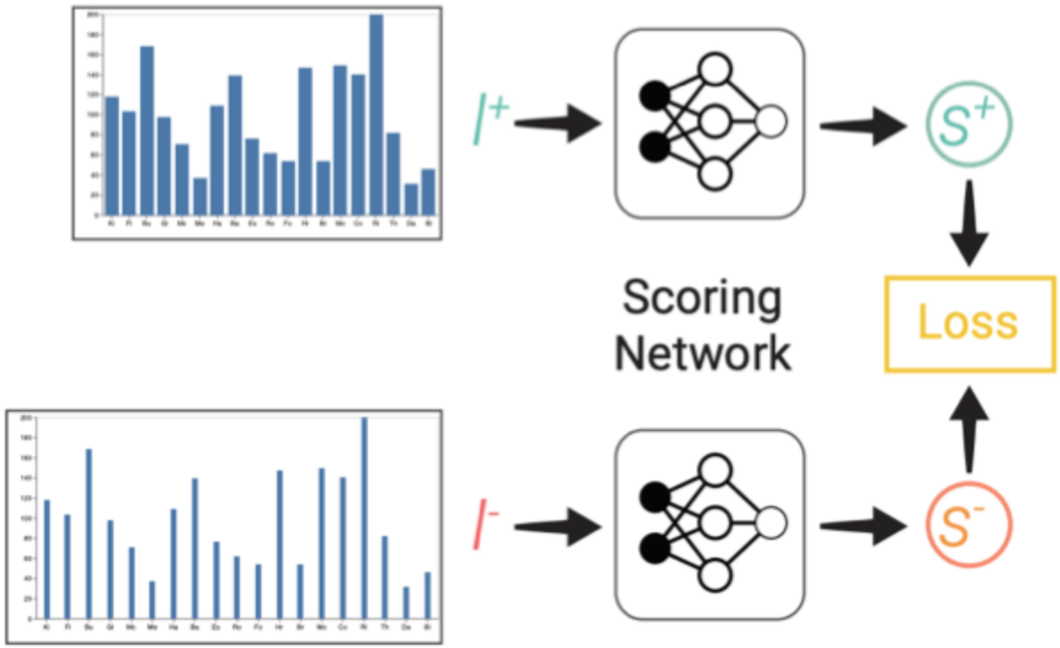
3. Machine-Learning Based Approach

Learning to automate chart configurations from crowdsourced paired comparison



3. Machine-Learning Based Approach

Deep learning on paired comparison data



To predict a higher score
for the positive chart in a pair

3. Machine-Learning Based Approach: Evaluation

Two experiments:

- Exp. 1: 3 parameters in bar charts
- Exp. 2: 6 parameters in bar charts

Baseline:

- RankSVM
- Hand-crafted cost functions for layout qualities

	Ours	RankSVM	White Space	Scale	Unity	Balance	All
Exp. 1 (N = 1,177)	76.60	70.83	57.28	56.26	52.00	56.08	60.81
Exp. 2 (N = 1,333)	78.27	64.48	58.24	61.72	56.21	63.18	68.73

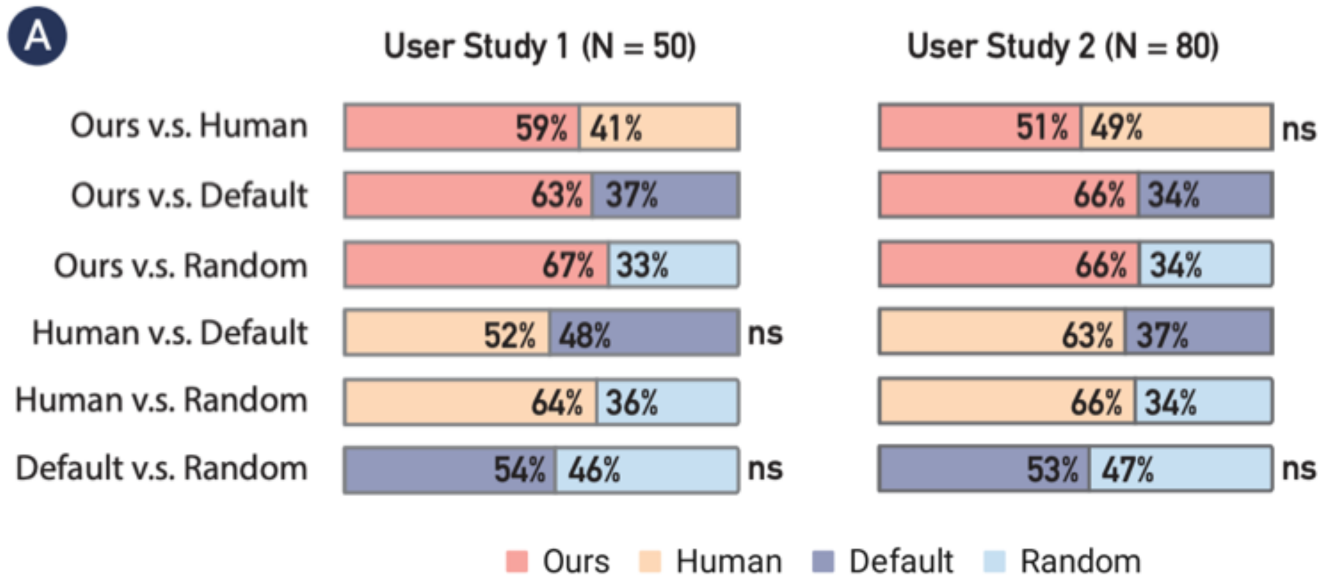
Model Performance in Accuracy

3. Machine-Learning Based Approach: Evaluation

Our method recommended better chart configurations

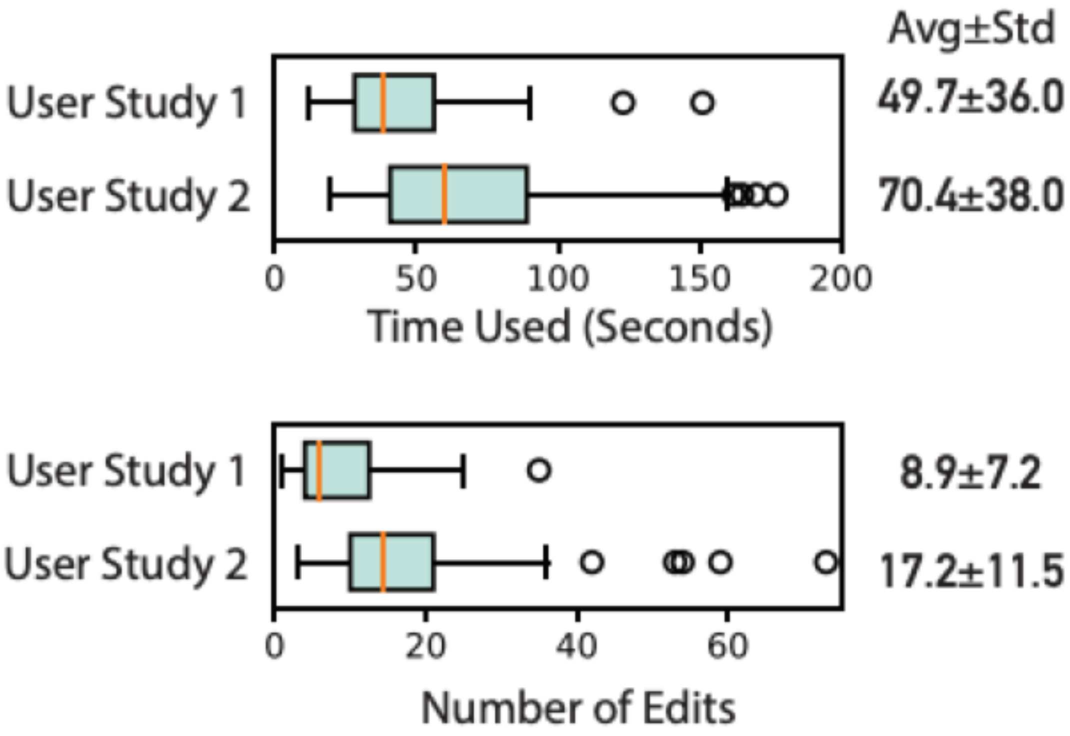
Baseline:

- Human
- Default (Excel and Vega-lite)
- Random



3. Machine-Learning Based Approach: Evaluation

Our method <u>saved user time.</u>



Baseline: Human

3. Machine-Learning Based Approach: Limitation

Paired comparison has exponential complexity.

6 parameters

10 possible values per parameter

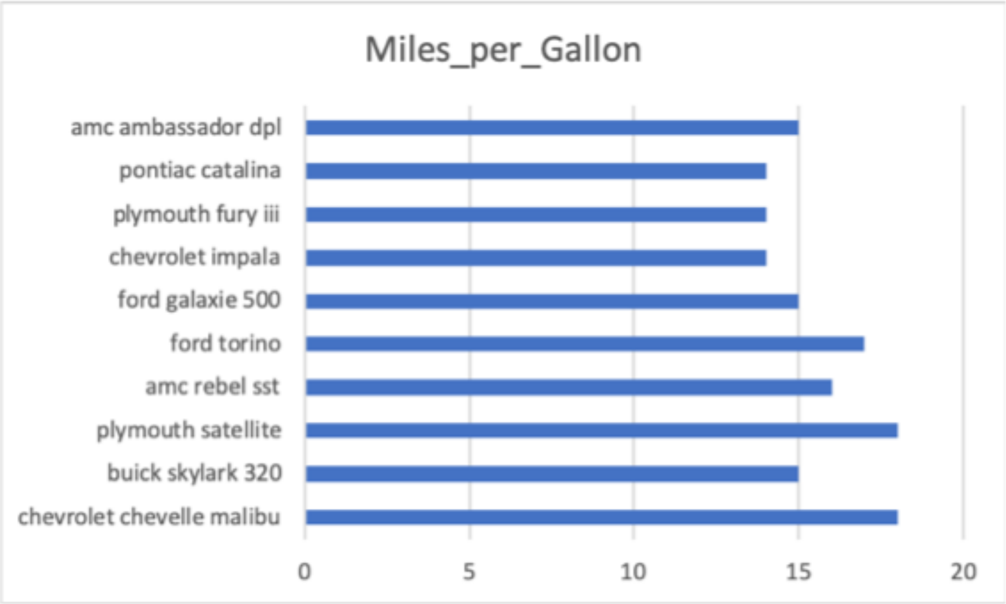
= 10^6 possible combinations of parameter values

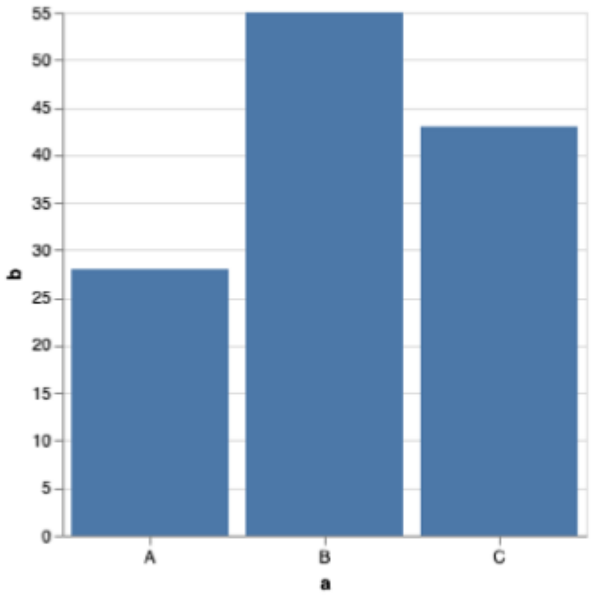
Full paired comparison requires 10^6 C 2.

❓ How to (adaptively) sample important pairs for comparison?

4. Discussions and Implications

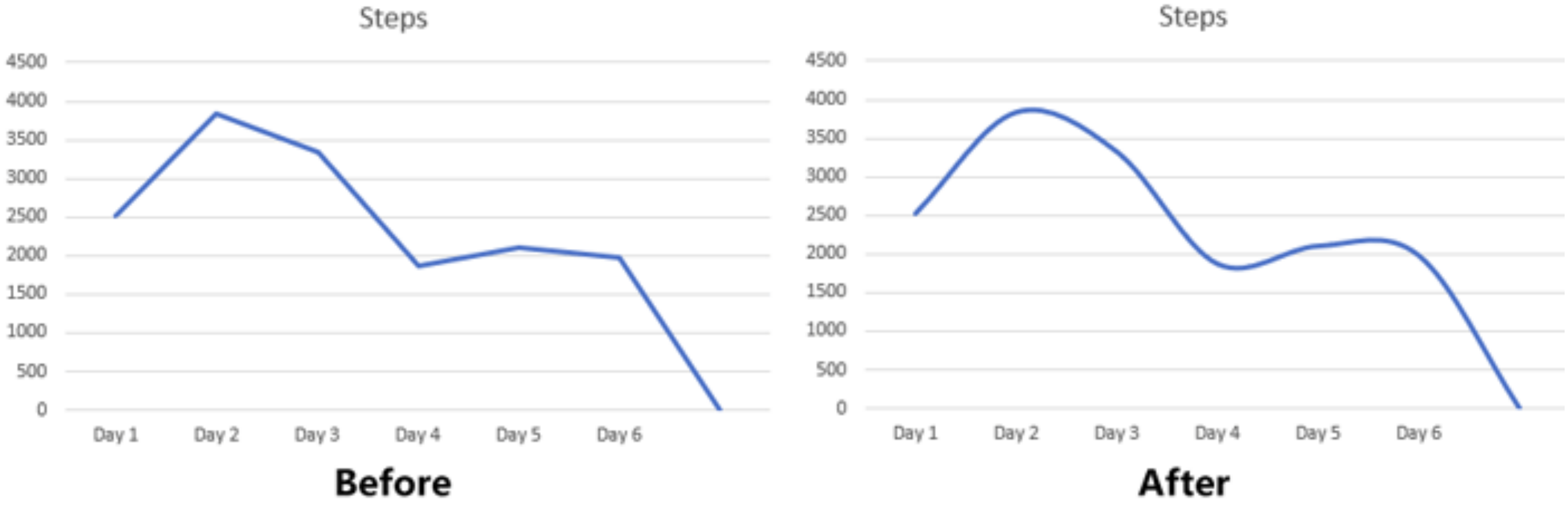
- Do not trust the defaults.





4. Discussions and Implications

- "What is beautiful is good"?



Google's Smooth Line Chart

4. Discussions and Implications

From Data Visualization to <u>Visualization Data</u>

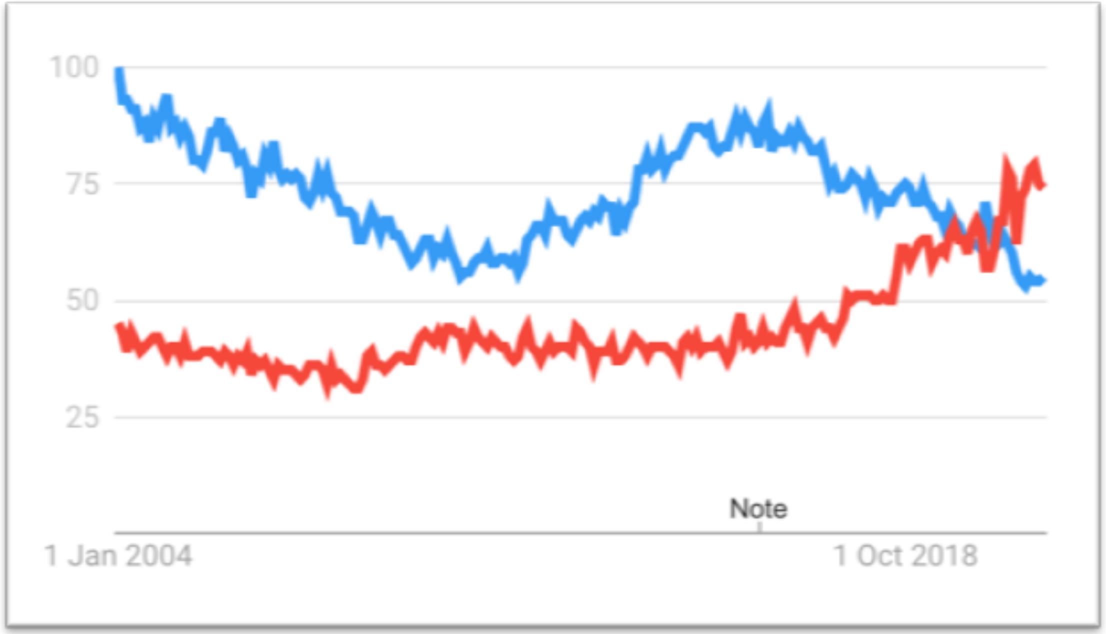


Chart has surpassed Image globally in Google Search Trend since <u>Jan 2020</u>

How AI could process and analyze visualizations?

- Recommendation
- Querying
- Summarizing
- ...

4. Discussions and Implications

What makes visualization data different?

- <u>Visualization data is characterized as multi-modal data.</u>

Image	Parameter	Text	Data

? How to represent and fuse different modals?

? How to translate from one modal to another?

? How to align one modal with another?

4. Discussions and Implications

What makes visualization data different?

- Visualization data is characterized as multi-modal data.
- <u>Visualization data is unnatural artefacts purposefully constructed with domain knowledge, which is difficult to learn.</u>

? How to develop ML models that are more tailored to visualization data?

? How to combine ML approaches with domain knowledge?

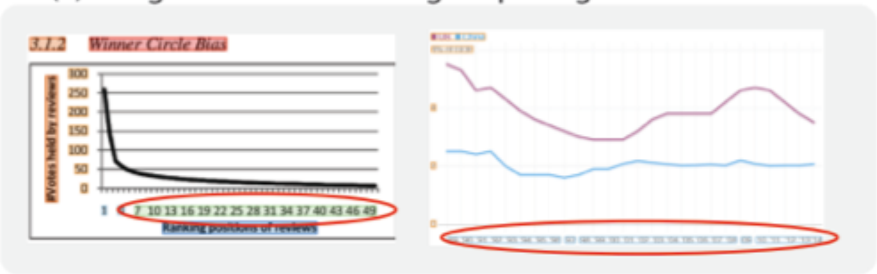
4. Discussions and Implications

What makes visualization data different?

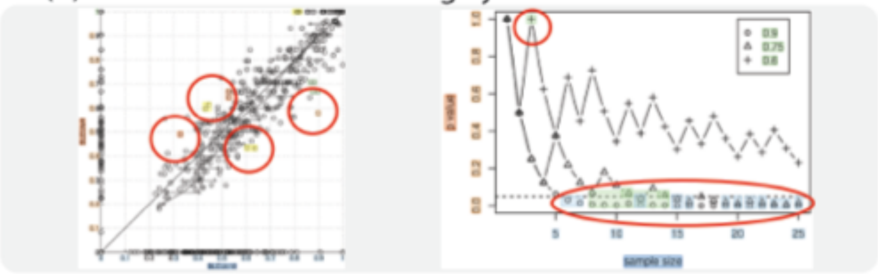
- Visualization data is characterized as multi-modal data.
- Visualization data is unnatural artefacts purposefully constructed with domain knowledge, which is difficult to learn.
- <u>Visualization data is more susceptible to detailed local information.</u>

How to develop ML models with high accuracy and granularity?

(a) Merged Labels due to Tight Spacing



(b) Text Characters as Plotting Symbols



4. Discussions and Implications

What makes visualization data different?

- Visualization data is characterized as multi-modal data.
- Visualization data is unnatural artefacts purposefully constructed with domain knowledge, which is difficult to learn.
- Visualization data is more susceptible to detailed local information.
- <u>The encoded data brings up several data-specific challenges such as mathematical reasoning.</u>

? How to learn mathematical reasoning?

5. Take-home Messages

- Chart layouts influence user experiences (UX) such as readability and aesthetics.
- It is helpful to automate parameter configurations via ML, saving user time and improving results.
- Visualization is becoming a new type of data that warrants deep research.