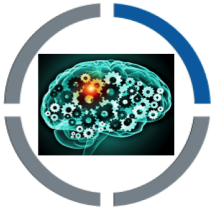
SpeedPerception: Phase-1 Update Results Overview

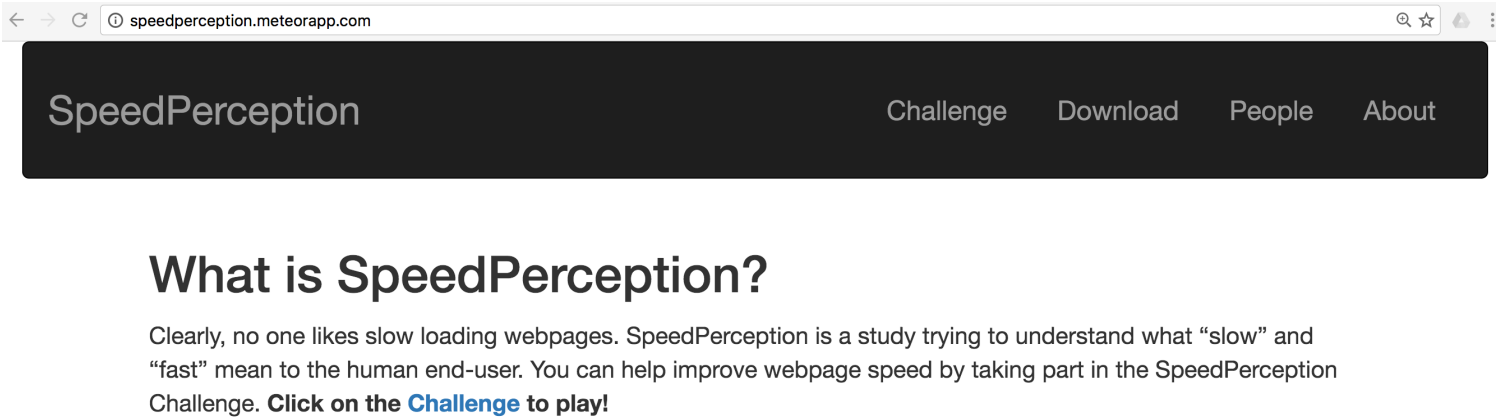
Parvez Ahammad
Head of Data Science and Machine Learning Group
Instart Logic Inc.

October 2016

What is SpeedPerception?



"SpeedPerception is a large-scale web performance crowdsourcing study focused on the perceived loading performance of above-the-fold content."



instartlogic

SpeedPerception: Team

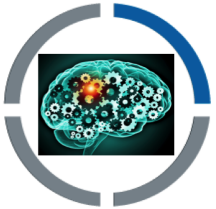
- Clark Gao
- Parvez Ahammad (@perceptPA)
- Prasenjit Dey



Collaborators:

- Estelle Weyl (@estellevw)
- Pat Meenan (@patmeenan)

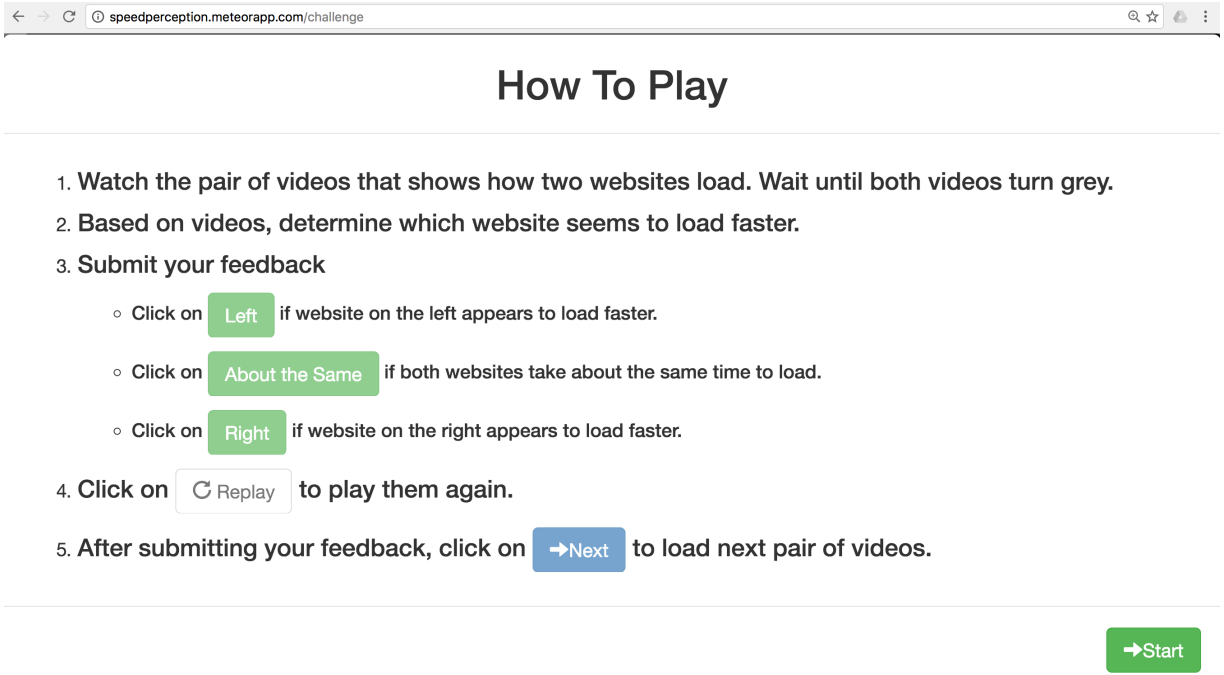

Why did we do SpeedPerception?



- Fill a critical gap for A/B comparisons of human end-user web UX
- Facilitate shift from "*anecData*™" to real human Ground Truth on webperf UX
- Provide a large-scale quantitative benchmark
    - Following MNIST / ImageNet / ActivityNet tradition
- Facilitate modeling / machine learning efforts in future
- Consistent dataset for algorithmic comparisons
- Open source framework / benchmark → reproducible results

SpeedPerception Challenge: Design (1/2)

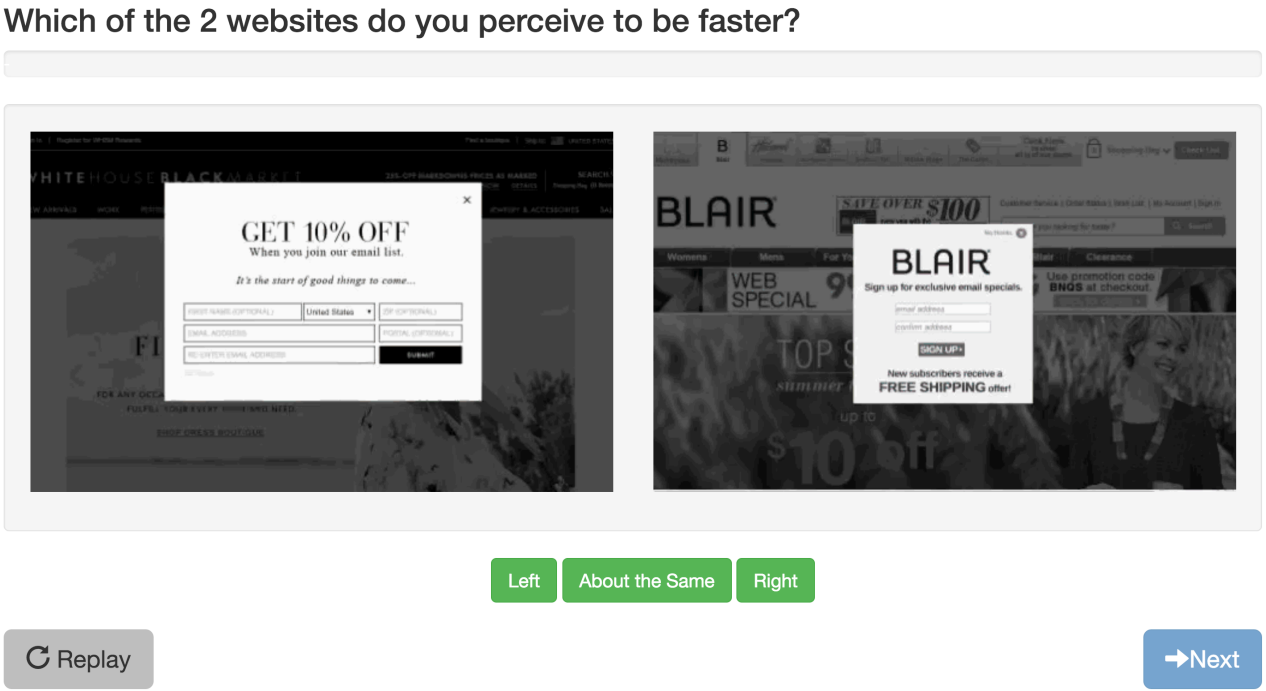
- Premise: perception of above-the-fold performance (like perceived webpage speed) is *relative*. http://speedperception.meteorapp.com/challenge



instartlogic

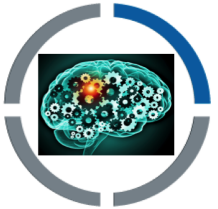
SpeedPerception Challenge: Design (2/2)

- Premise: perception of above-the-fold performance (like perceived webpage speed) is *relative*. http://speedperception.meteorapp.com/challenge

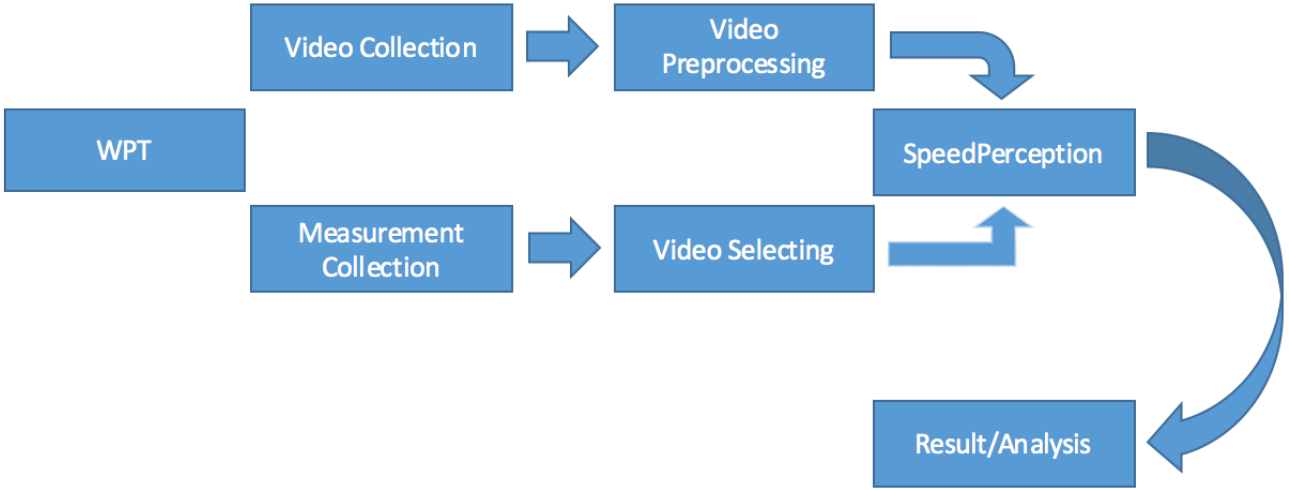


instartlogic

SpeedPerception Challenge: Phase-1 Data (1/2)

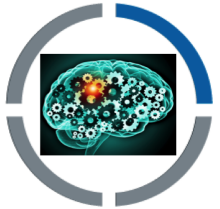


- Source: Internet Retailer top-500 URLs
- WebPagetest (private instance) to collect HAR / metrics / videos (June 2016)
- Chrome browser / Cable connection speed / Desktop rendering mode
- All steps publicly available: https://github.com/pdey/SpeedPerceptionApp



instartlogic

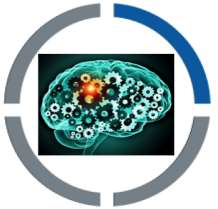
SpeedPerception Challenge: Phase-1 Data (2/2)



- After applying publicized rules for pair selection:
    - 115 URLs
    - 160 A/B pairs
- Each session: 16 test pairs + 5 honeypot* pairs = 21 pairs
- Record voluntary time to click for each pair (not publicized)

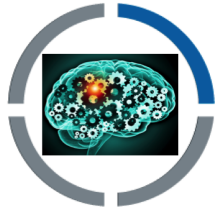
Honeypot* video pair = A video pair with known (very obvious) choice – used as a way to evaluate user input quality

SpeedPerception Challenge: Working hypotheses



- **H1:** No single metric can explain human choices with 90%+ accuracy
- **H2:** Visual metrics will perform better than non-visual/network metrics
- **H3:** User will not wait until “Visual Complete” to make their choice (despite the explicit instruction to wait until video turns grey)

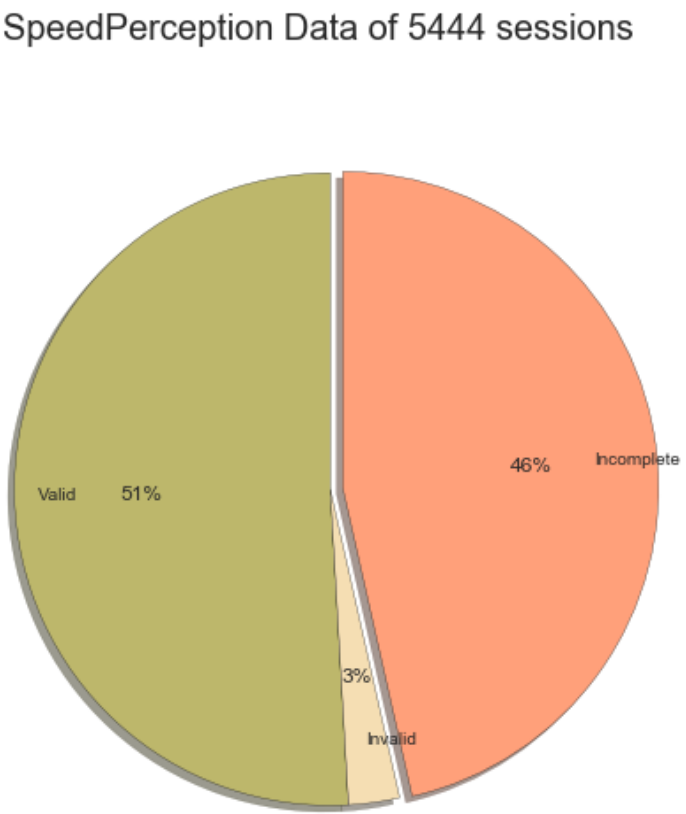
SpeedPerception Challenge



- Went public on 28th July 2016

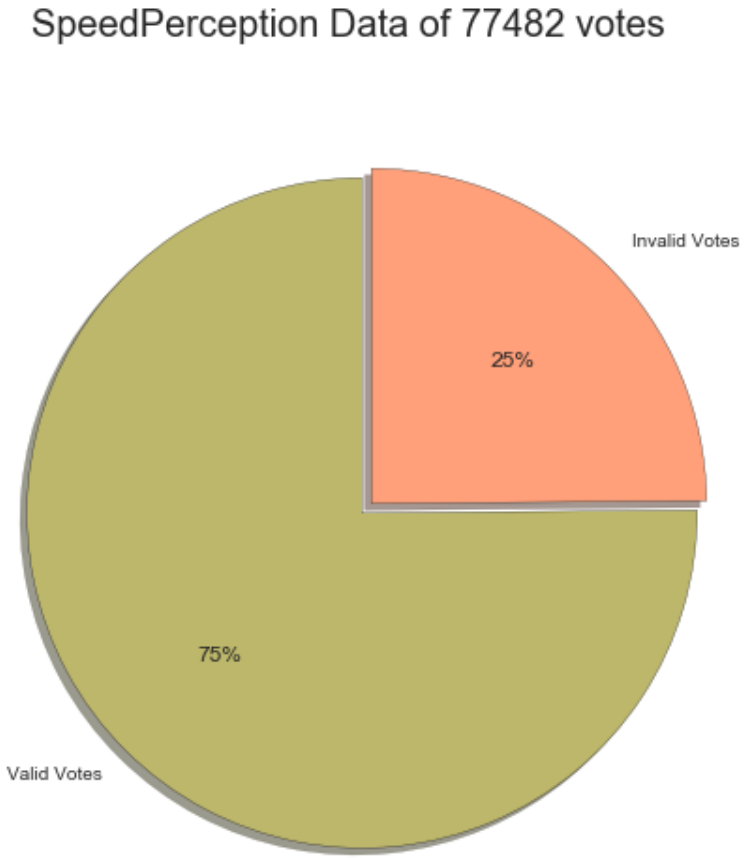
- Phase-1 ended on 30th September 2016
- Benchmark and findings available at:
  - http://SpeedPerception.org

SpeedPerception Benchmark / DeepDive



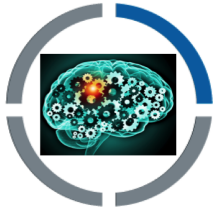
instartlogic

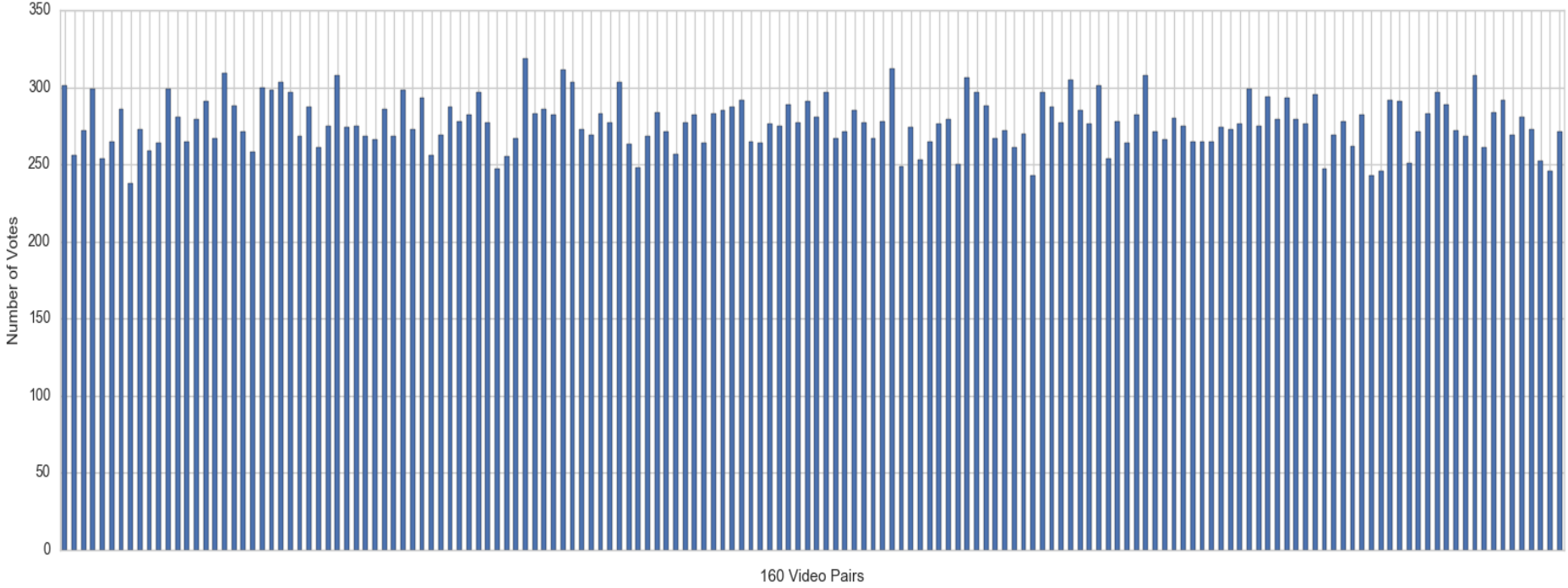
SpeedPerception Benchmark / DeepDive



instartlogic

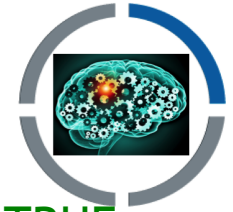
SpeedPerception Benchmark / DeepDive



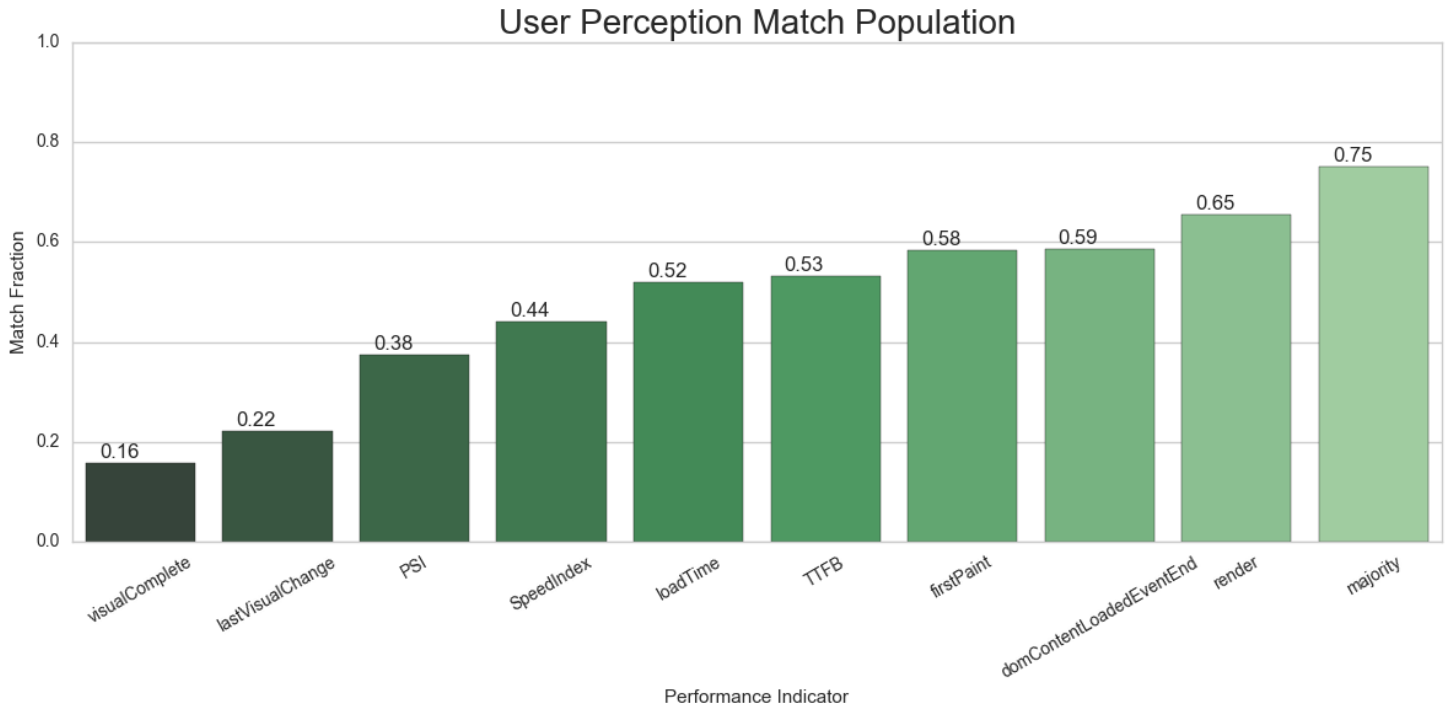


instartlogic

SpeedPerception Benchmark / Hypothesis # 1

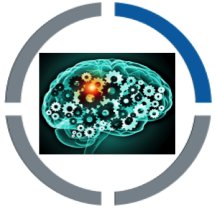


- **H1:** No single metric can explain human choices with 90%+ accuracy - TRUE

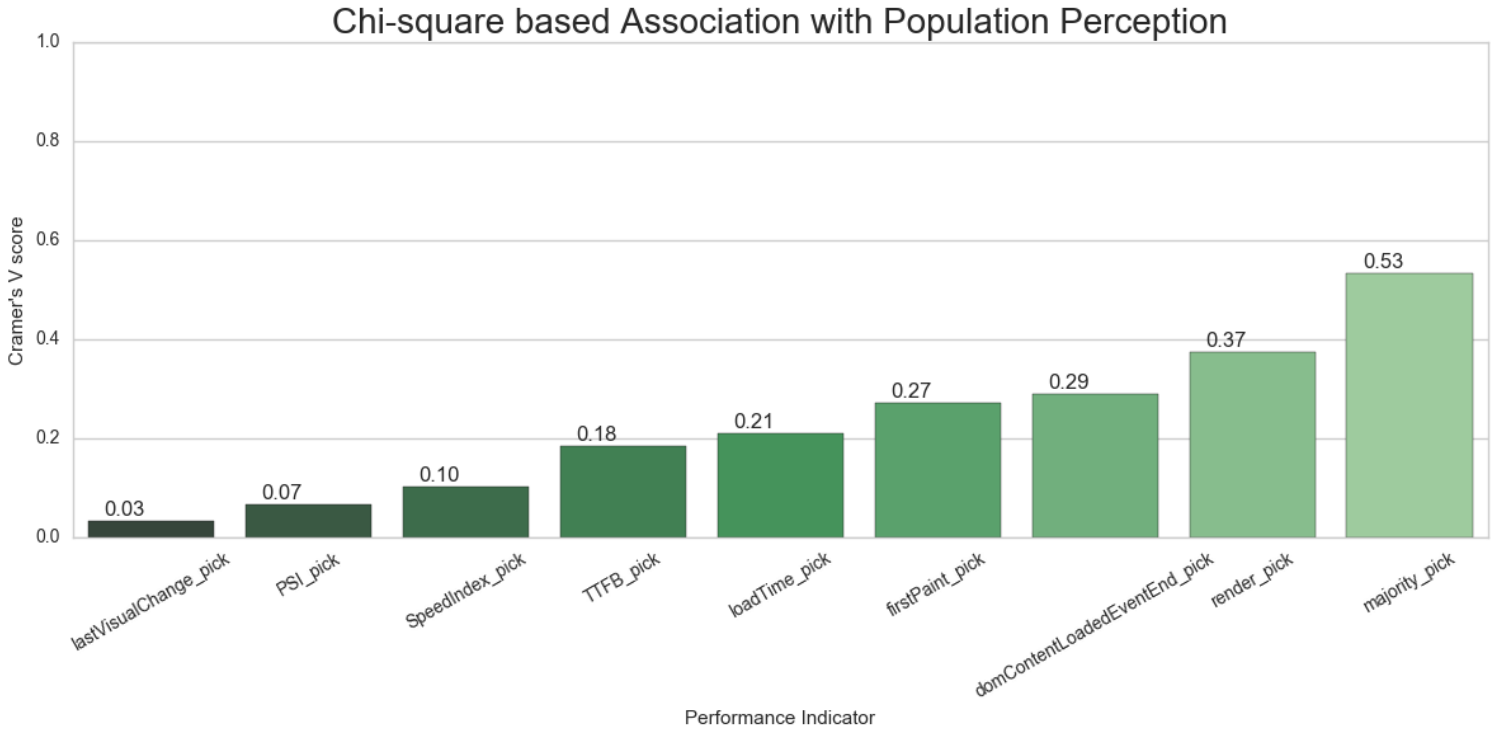


instartlogic

SpeedPerception Benchmark / Hypothesis # 1

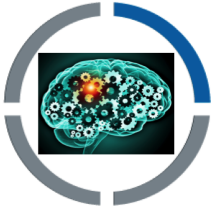


- **H1:** No single metric can explain human choices with 90%+ accuracy - TRUE

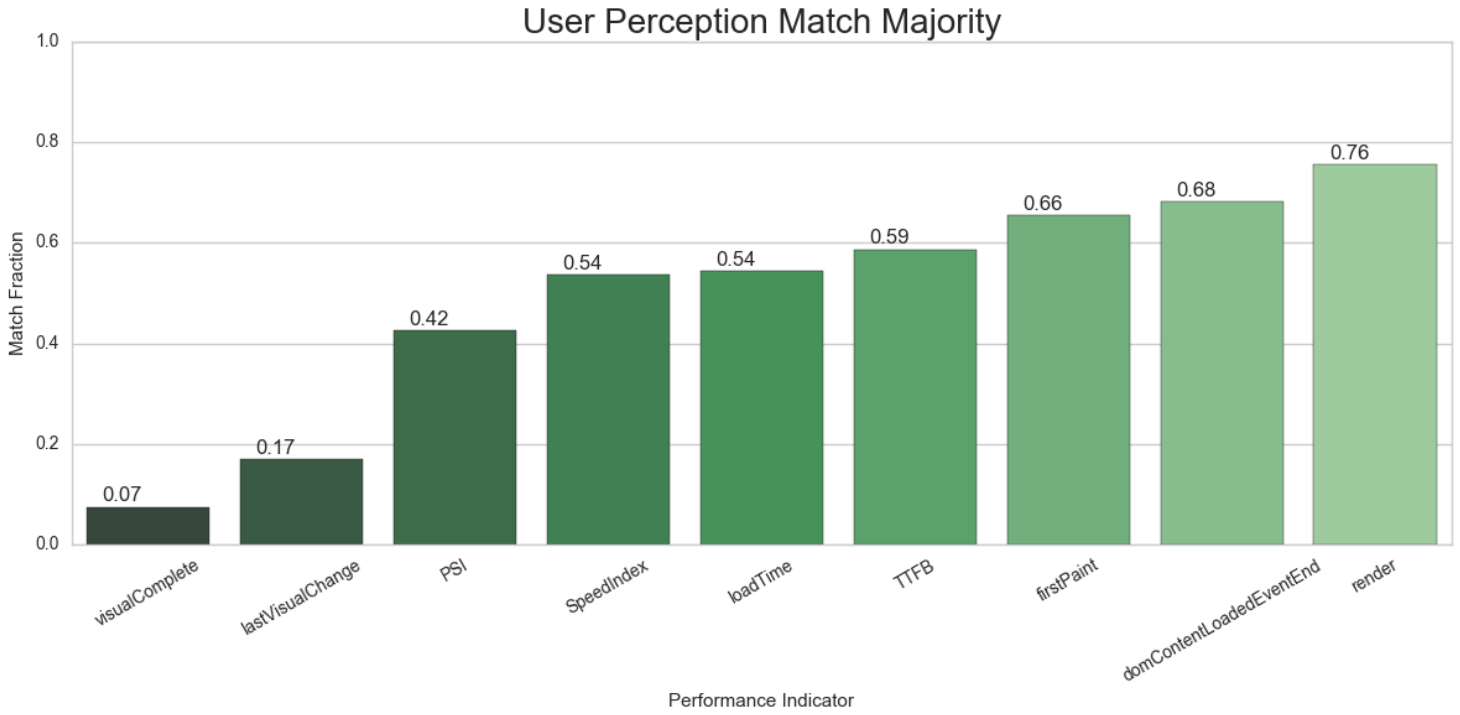


instartlogic

SpeedPerception Benchmark / Hypothesis # 1

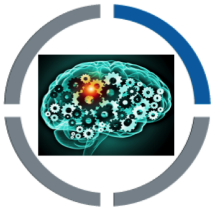


- **H1:** No single metric can explain human choices with 90%+ accuracy - TRUE

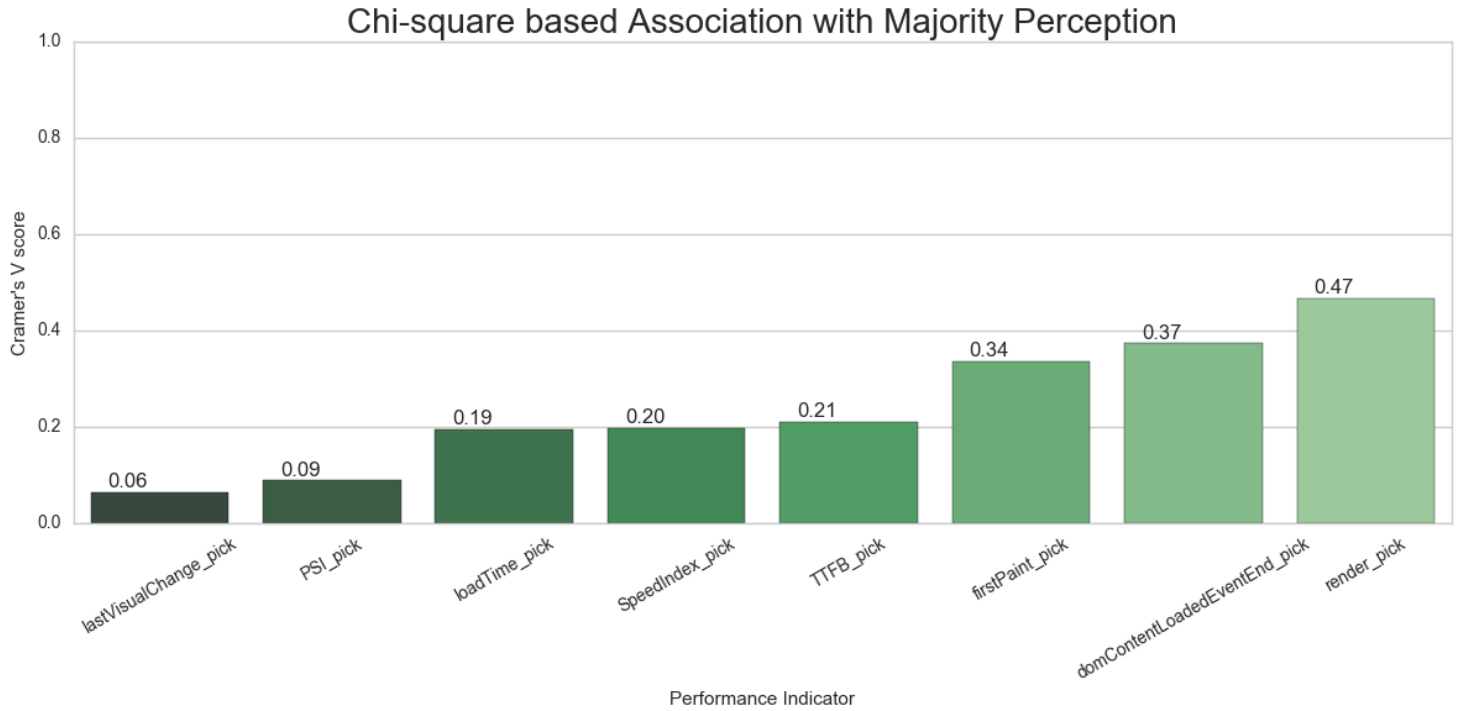


instartlogic

SpeedPerception Benchmark / Hypothesis # 1



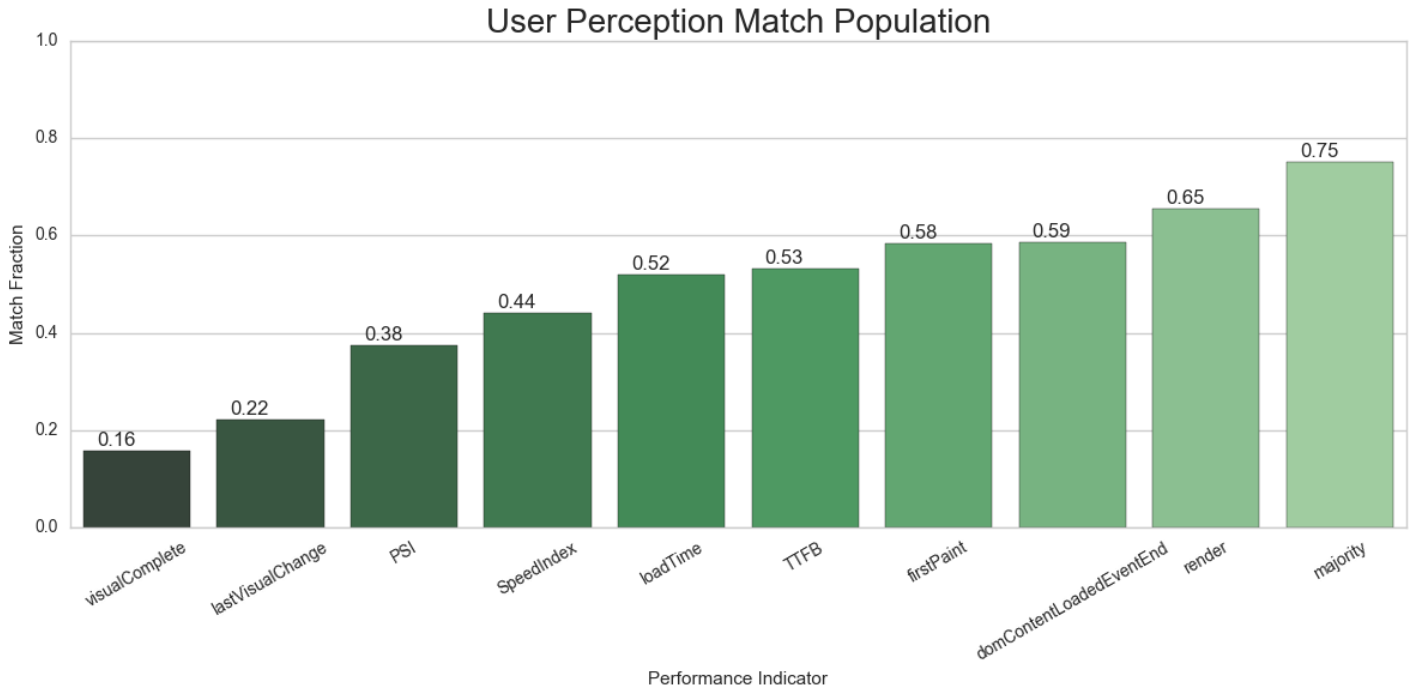
- **H1:** No single metric can explain human choices with 90%+ accuracy - TRUE



instartlogic

SpeedPerception Benchmark / Hypothesis # 2

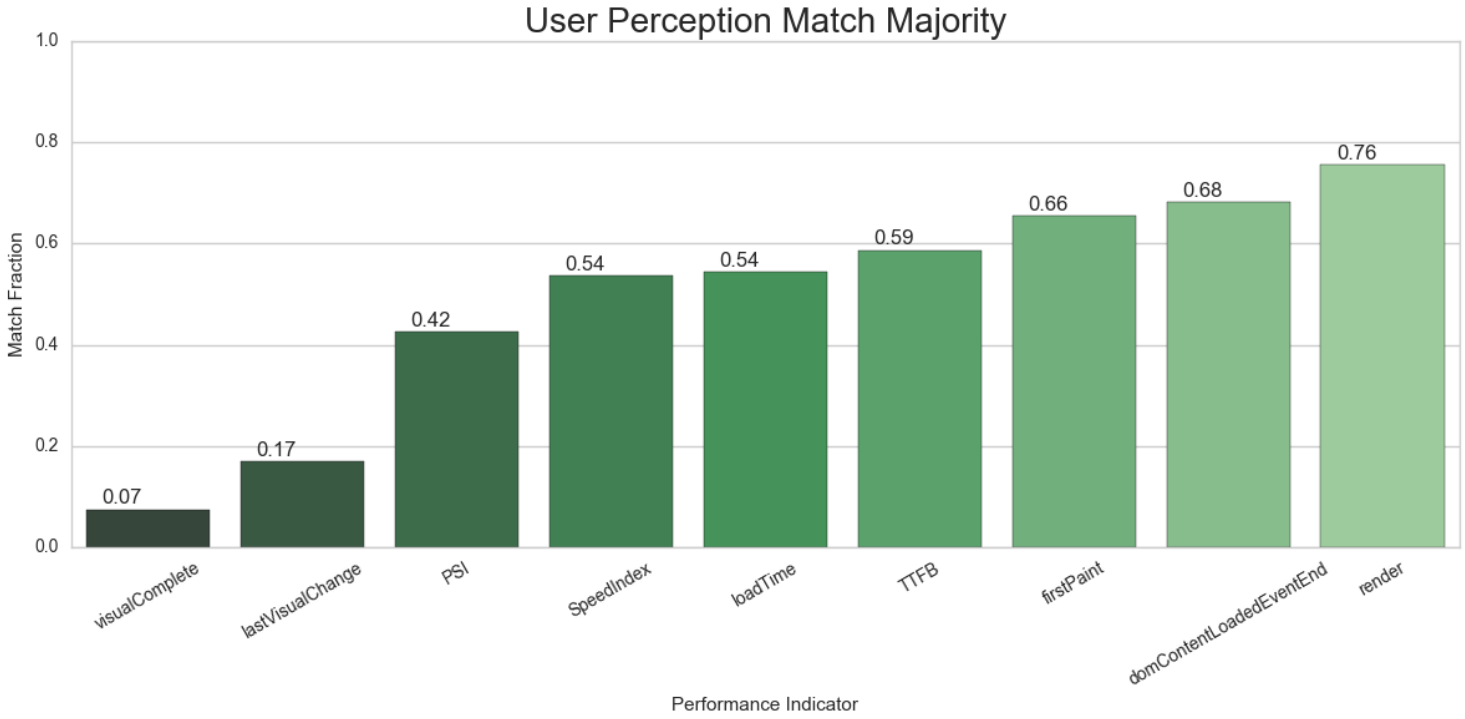
- **H2:** Visual metrics will perform better than non-visual metrics – NOT TRUE



instartlogic

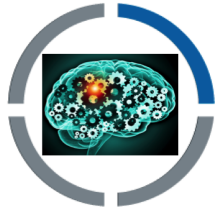
SpeedPerception Benchmark / Hypothesis # 2

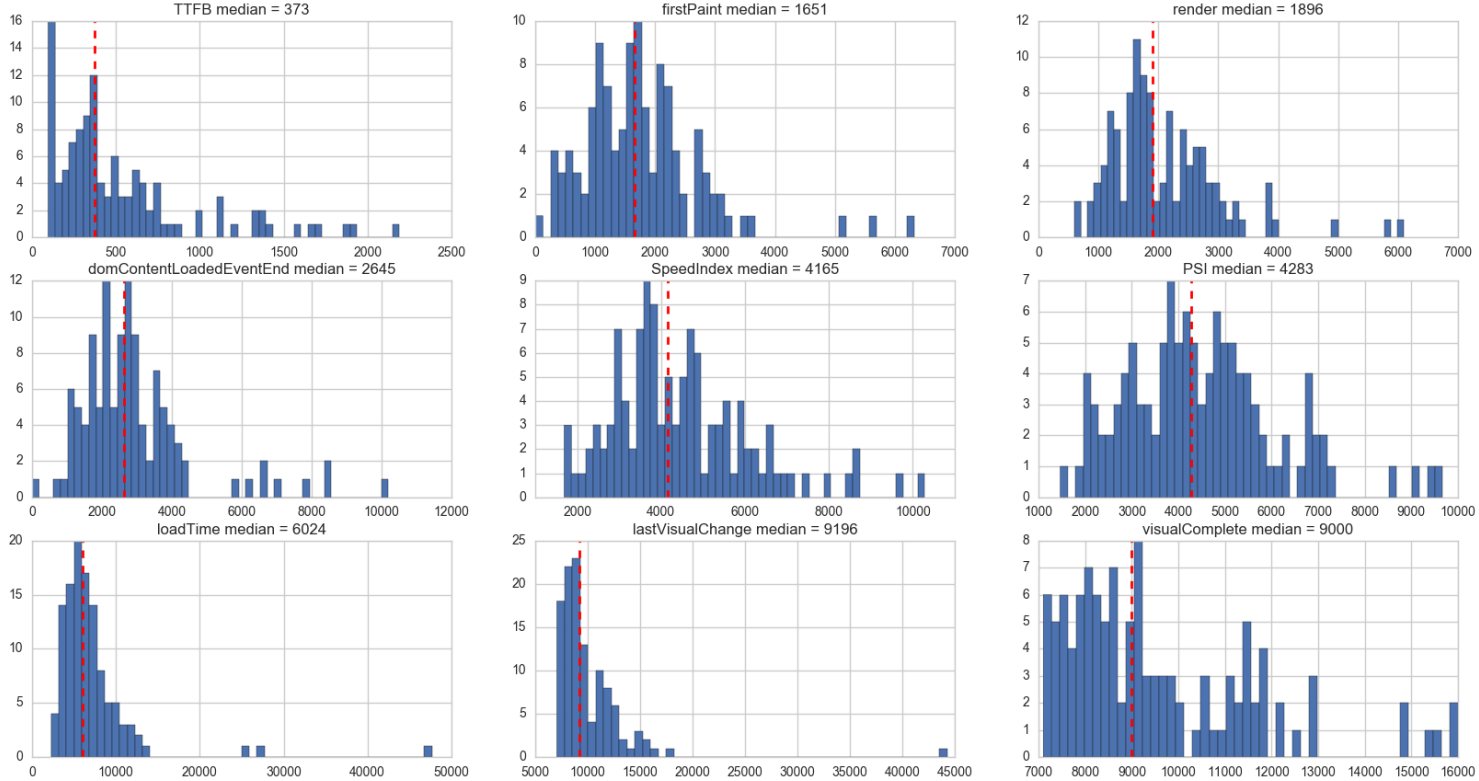
- **H2:** Visual metrics will perform better than non-visual metrics - NOT TRUE



instartlogic

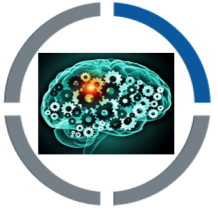
SpeedPerception Benchmark / Order of events



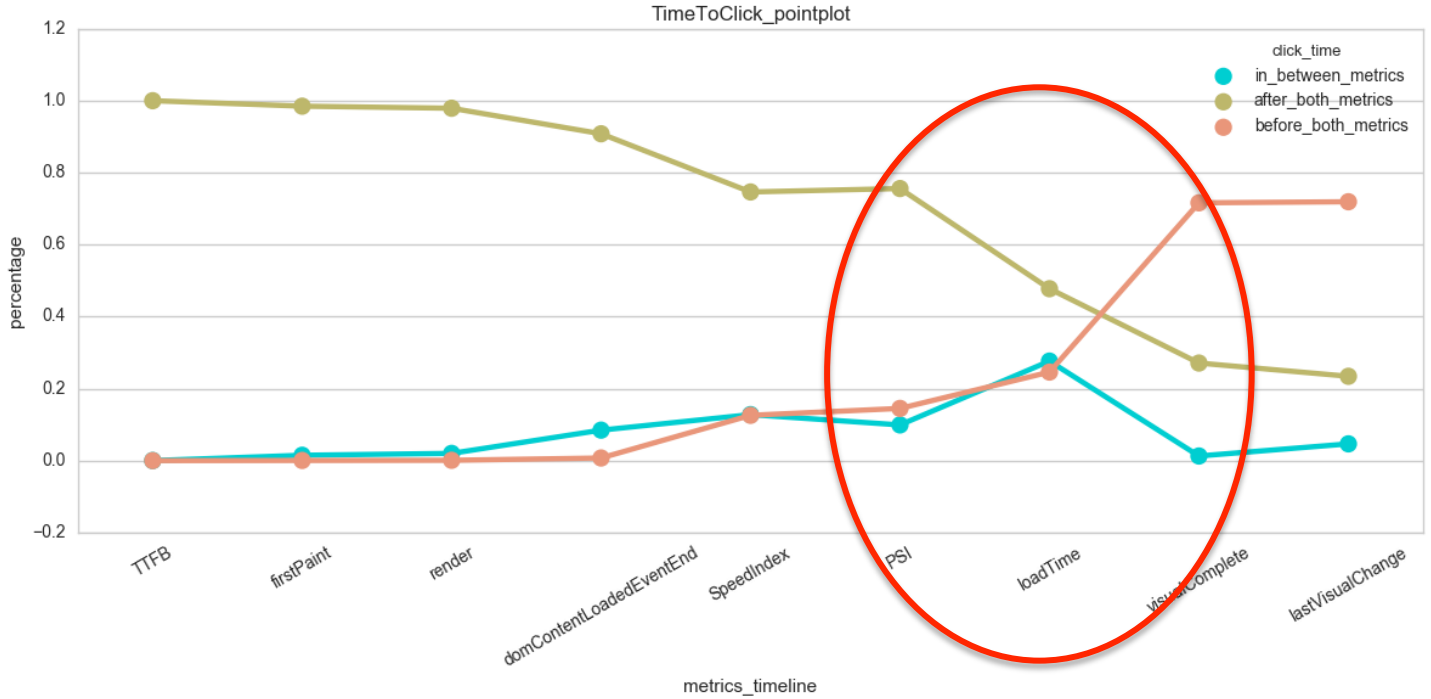


instartlogic

SpeedPerception Benchmark / Hypothesis # 3



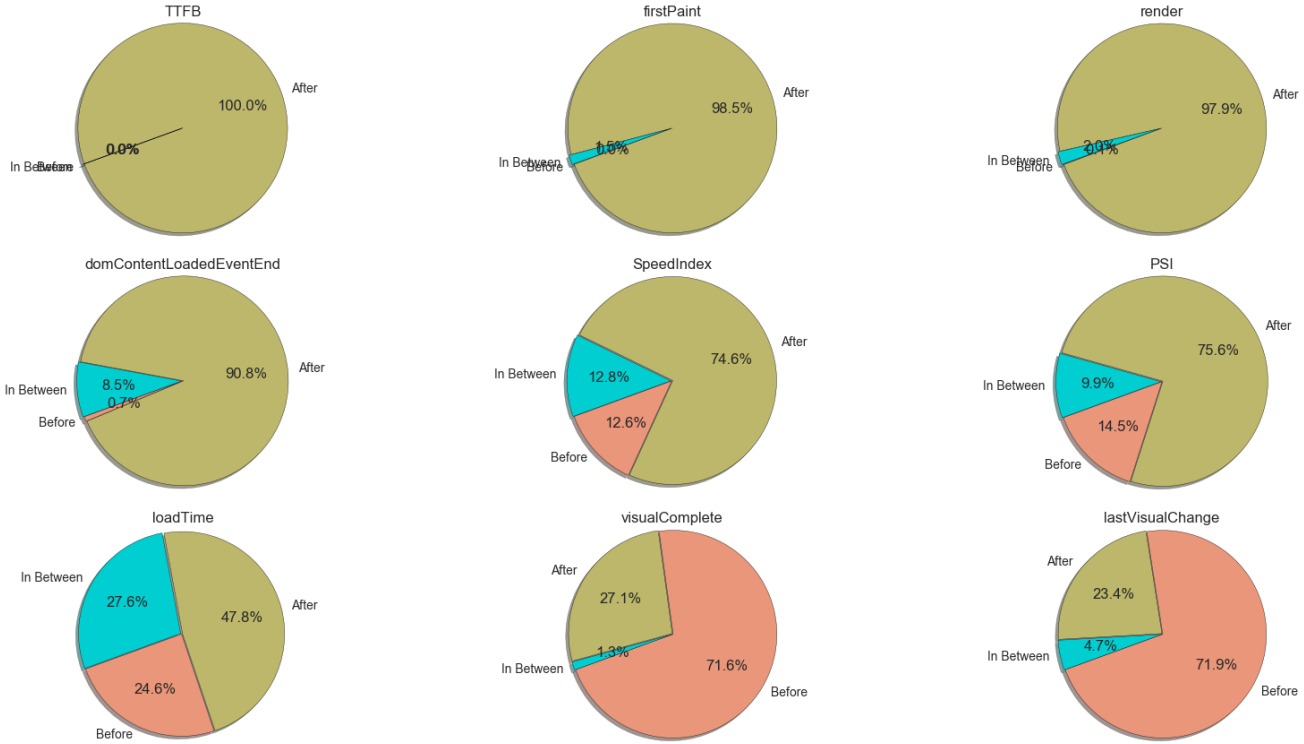
- **H3:** User will not wait until “Visual Complete” to make their choice - TRUE



instartlogic

SpeedPerception Benchmark / Hypothesis # 3

- **H3:** User will not wait until “Visual Complete” to make their choice - TRUE



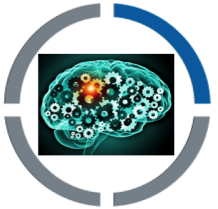
instartlogic

SpeedPerception Benchmark / User Feedback

- Perception of speed and UX strongly impacted by popups / overlays

instartlogic

Thoughts / Looking ahead..



- A rich source to explore & model webperf ideas with human A/B ground truth

- Need to look into Mobile rendering and its impact on results (Phase-2)
- URL sources ought to go beyond IR500 list / E-commerce vertical
- Other browsers beyond Chrome might be worth sampling
- Would it be worth capturing & including Dev. Timeline information?
- Better sampling of URLs to understand impact of interstitials / visual jitter

- Suggestions? Send to @perceptPA or parvez@ieee.org

SpeedPerception Challenge: Phase-2 planning underway

- Benchmark, analysis code and findings from Phase-1 available at:
  - http://SpeedPerception.org

- Send feedback to @perceptPA or parvez@ieee.org

instartlogic