Eye Fixation Metrics for Large Scale Analysis of Information Visualizations

Zoya Bylinskii, *Massachusetts Institute of Technology*
Michelle A. Borkin, *Northeastern University*

Hi, I'm Zoya. Today I will tell you about my workshop paper in collaboration with Michelle Borkin at Northeastern. We analyze eye movements on visualizations and then visualize the eye movements.

We are interested in eye movement analyses

and corresponding visual analytics tools

that can facilitate the collective analysis of

information visualization designs.

First let me give you our motivation both for this workshop paper and our general line of research on eye movements and visualizations. [text] Now let me explain the text in blue a little more precisely.

We are interested in eye fixation metrics

and corresponding visualizations

that can facilitate the collective analysis of

information visualization designs.

across a population of users

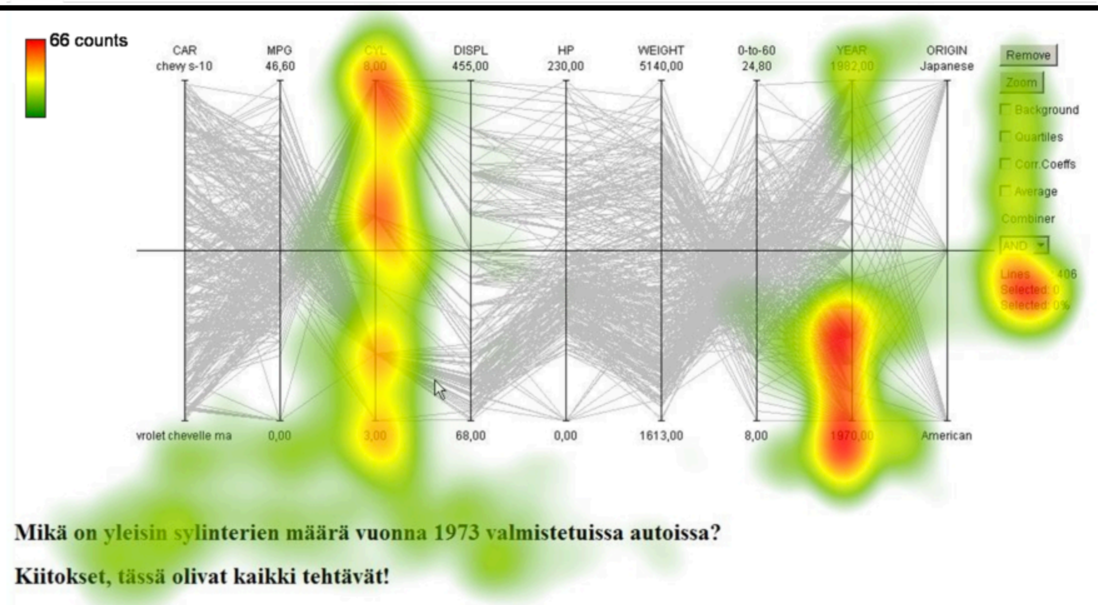
to evaluate and compare visualization types and sources

Although we are interested in eye movement analyses more generally, in this paper and presentation we focus specifically on fixation metrics. Also, although multiple visual analytics tools are available for examining eye movement data, here we restrict our scope to static visualizations. By 'collective analysis' I mean the ability to quantify and run statistics on a whole population of users. Our goal is to perform this eye movement analysis in order to evaluate and compare different visualization types and sources.

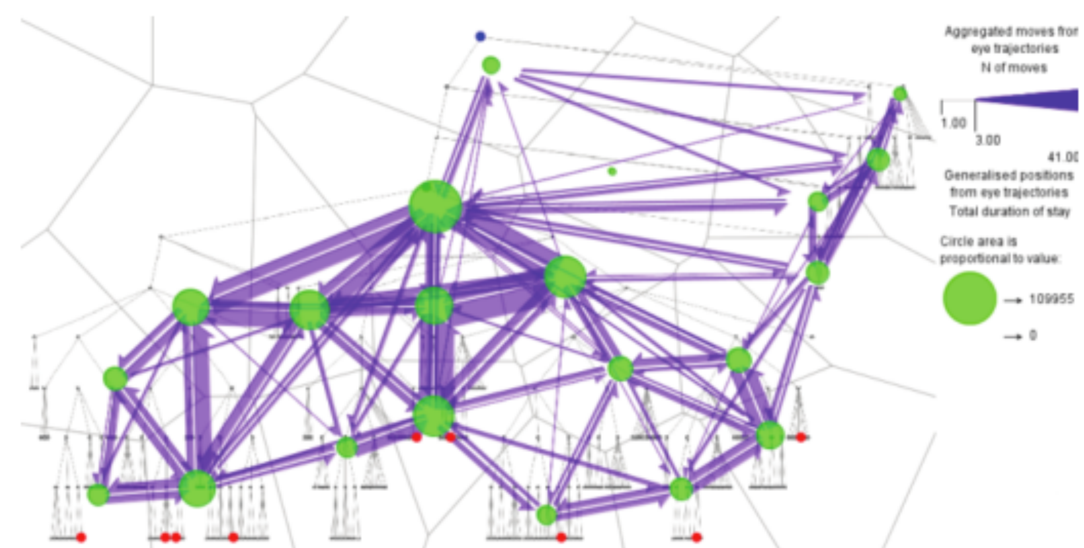
Design evaluation using eye movements

(a) SimulSort (b) Typical Sorting

Kim et al. (2012)



Siirtola et al. (2009)

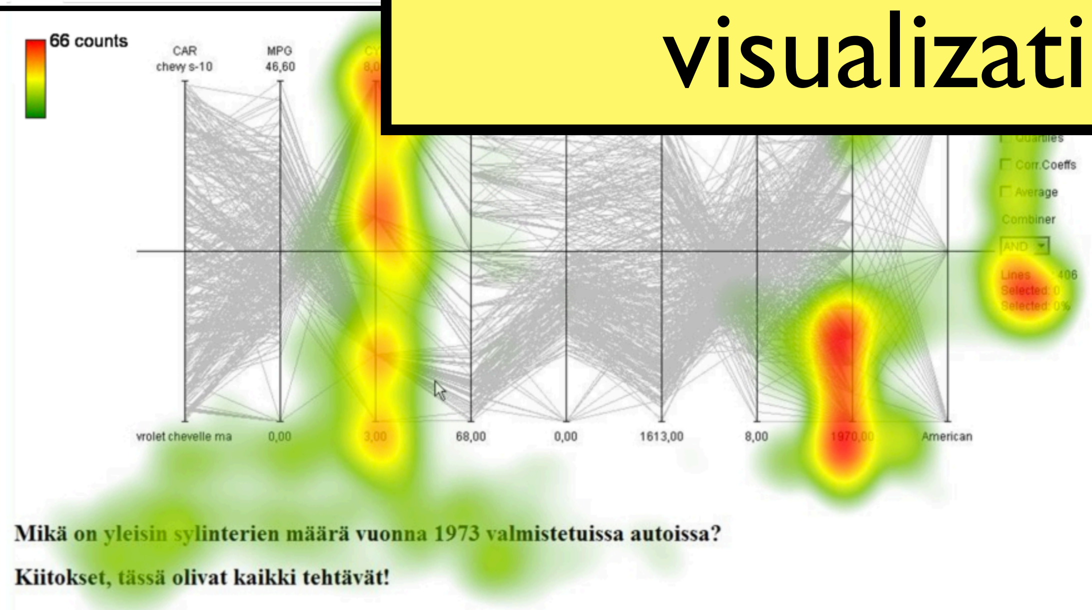


Andrienko et al. (2012)

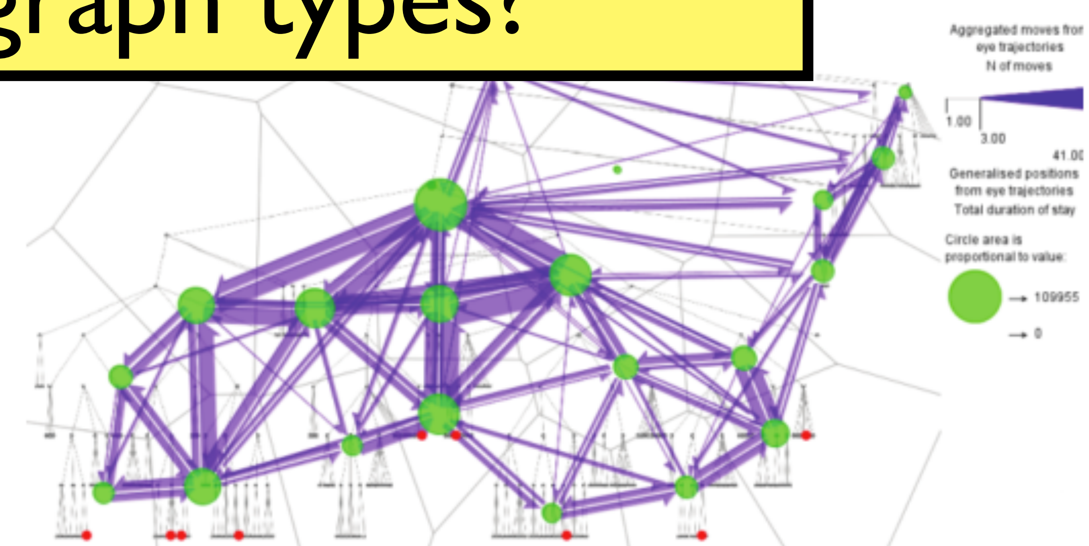
A number of prior works have visualized and analyzed eye movements to investigate the attentional patterns of a population of participants on a particular information visualization type.

Design evaluation using eye movements

What about comparisons across visualization/graph types?



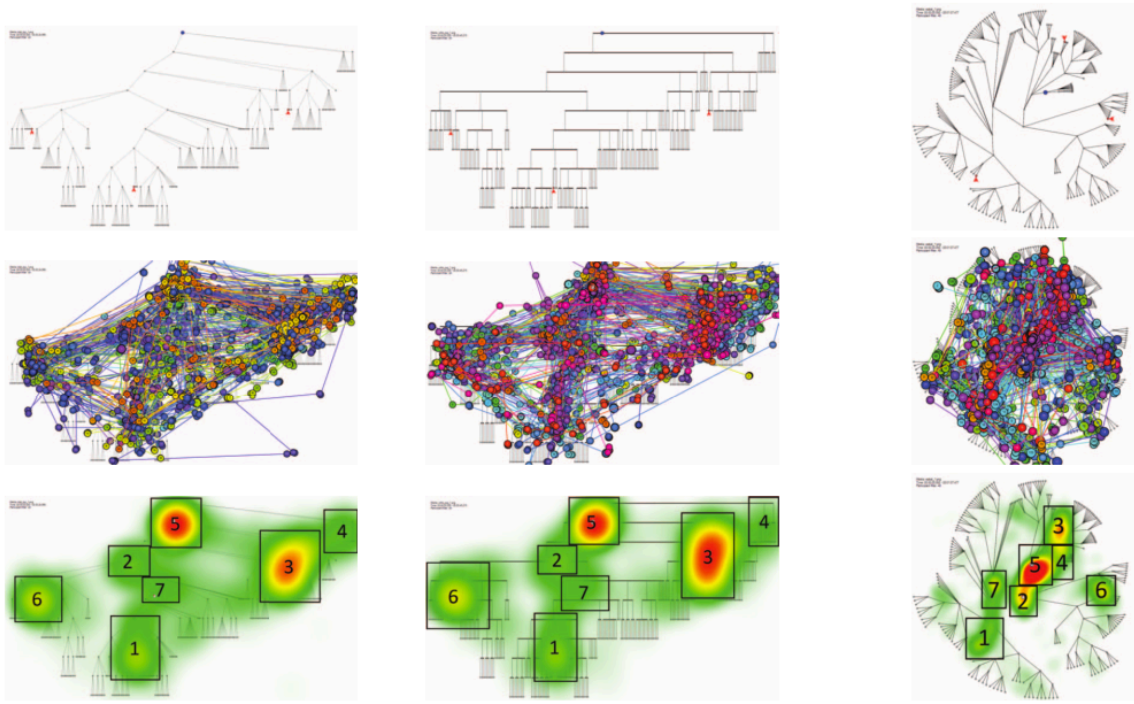
Siirtola et al. (2009)



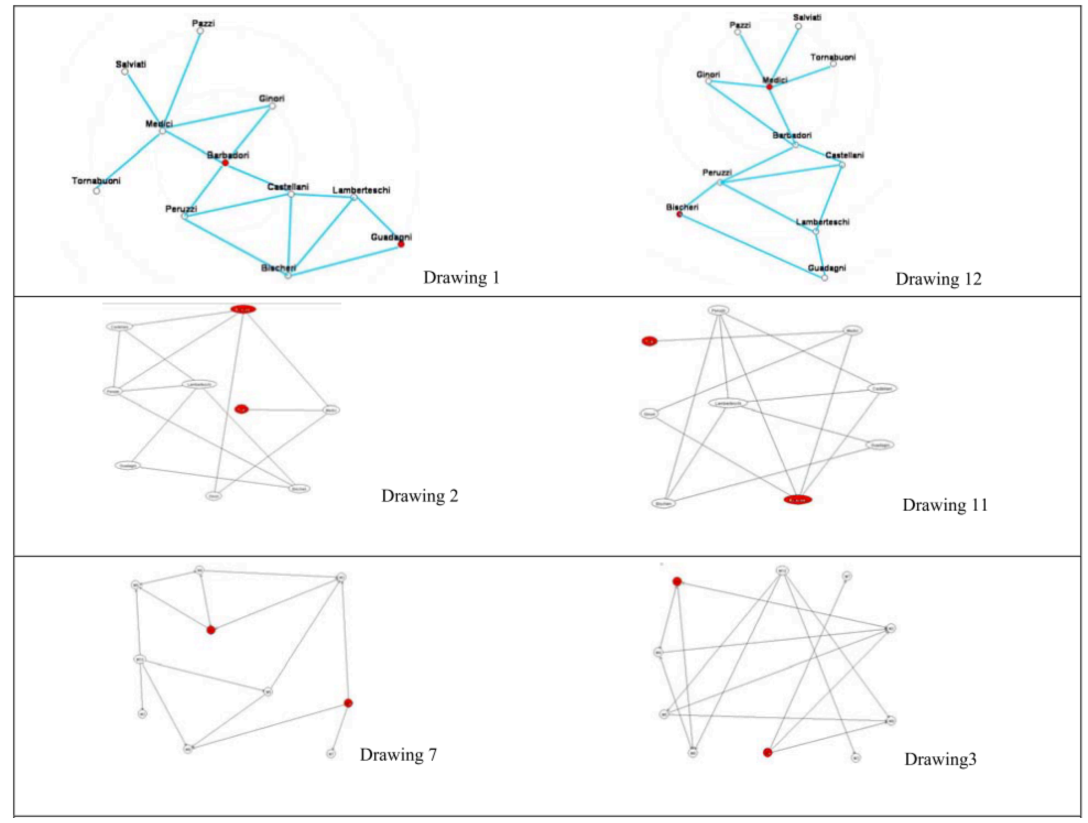
Andrienko et al. (2012)

Rather than validating and evaluating a particular visualization or design, here our focus is on the evaluation and comparison <u>across</u> a diverse set of visualization types.

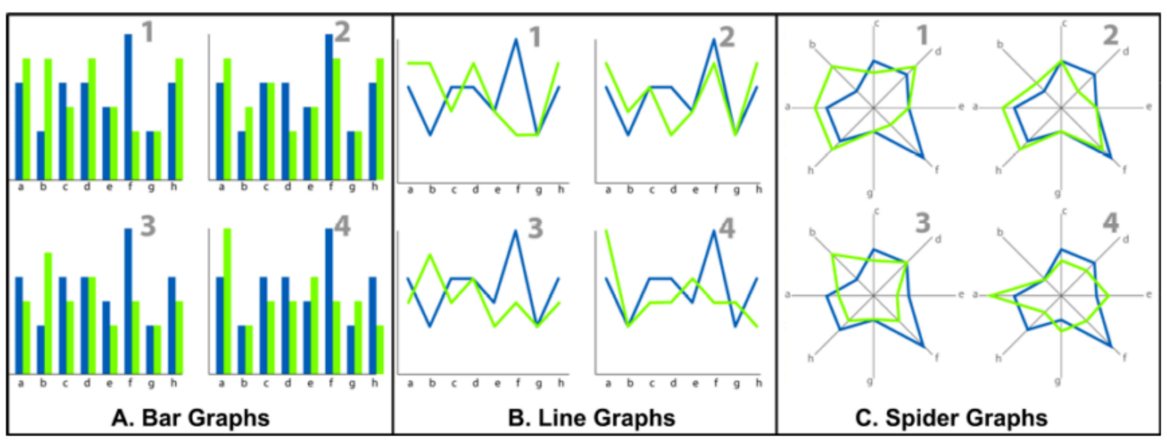
Design comparison using eye movements



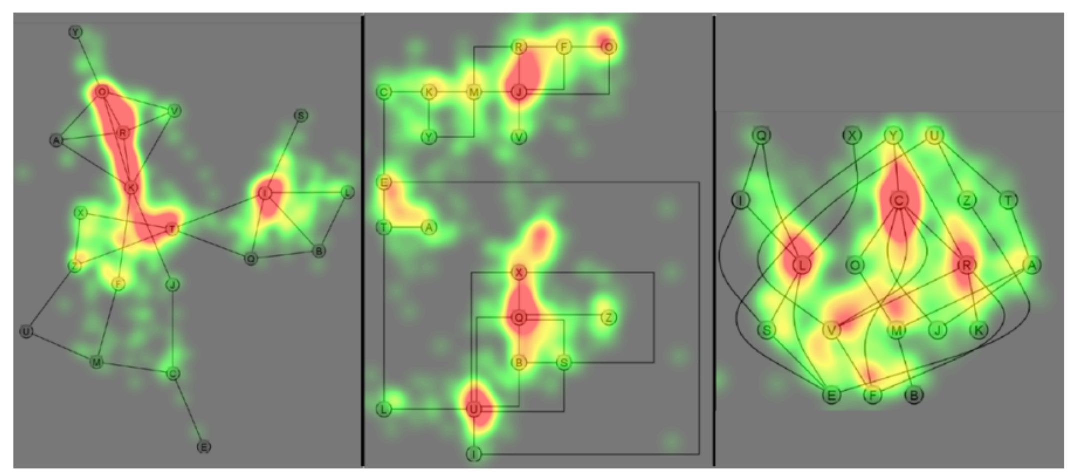
Burch et al. (2011)



Huang et al. (2005,2007,2009)



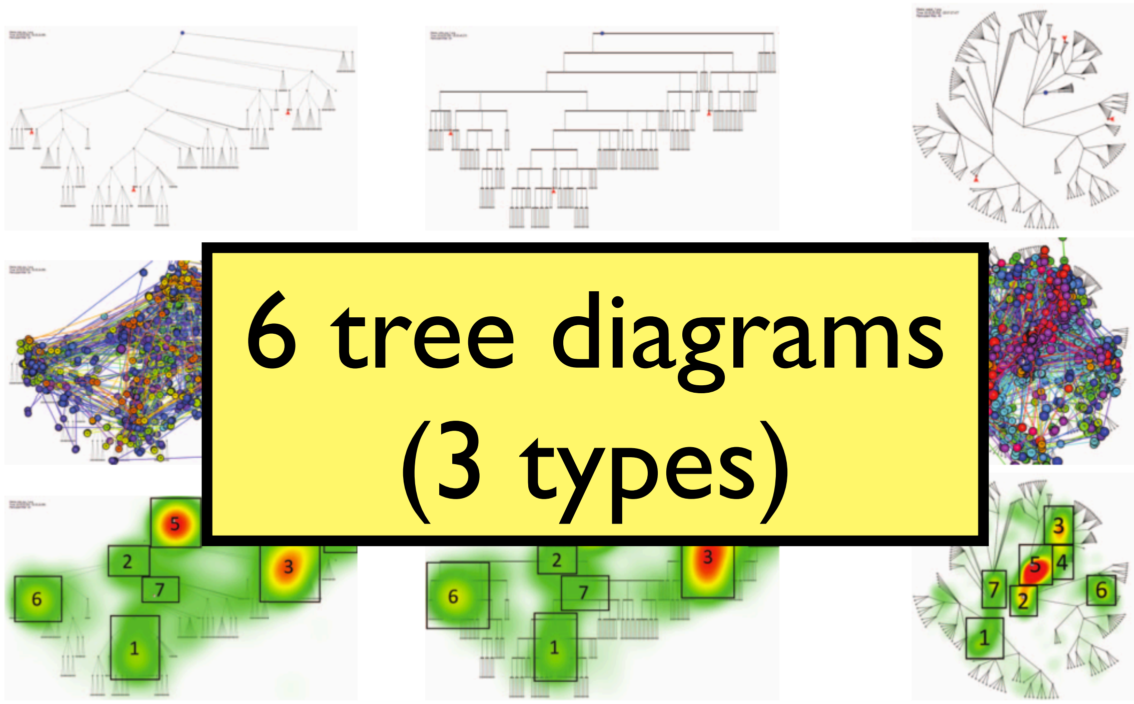
Goldberg & Helfman (2010)



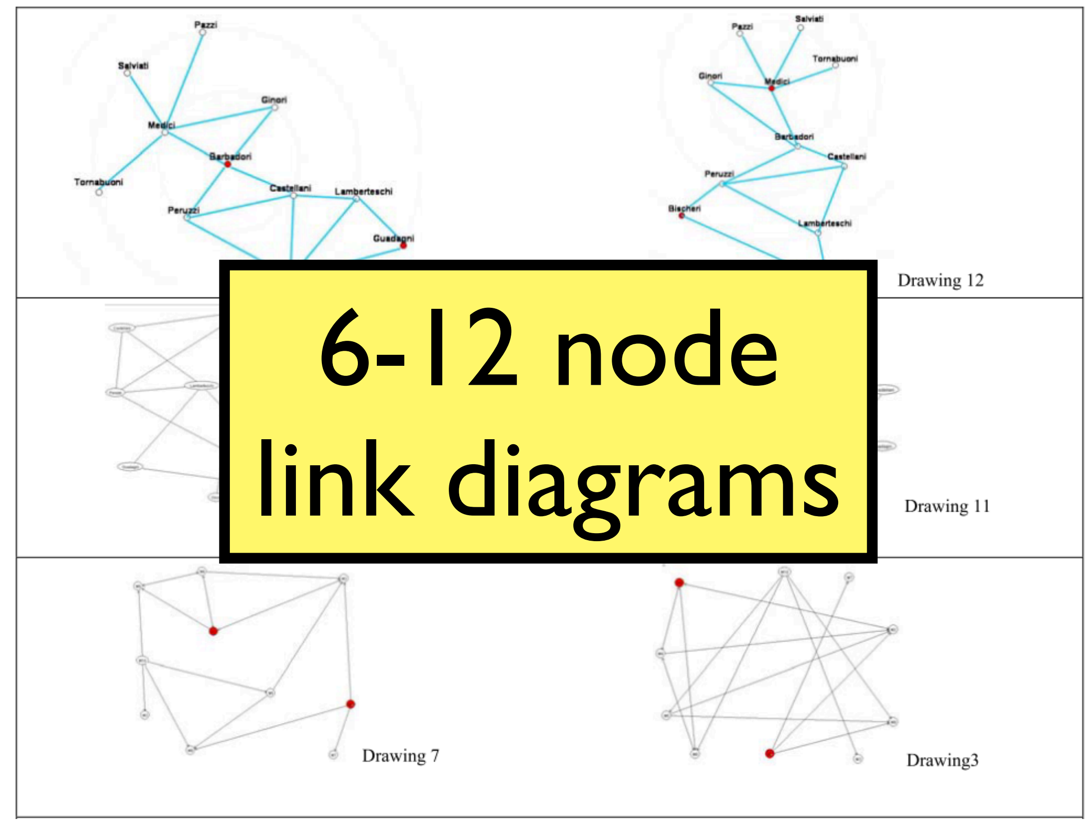
Pohl et al. (2009)

There have also been a number of works looking at directly comparing different designs. (there's many more works that won't fit here).

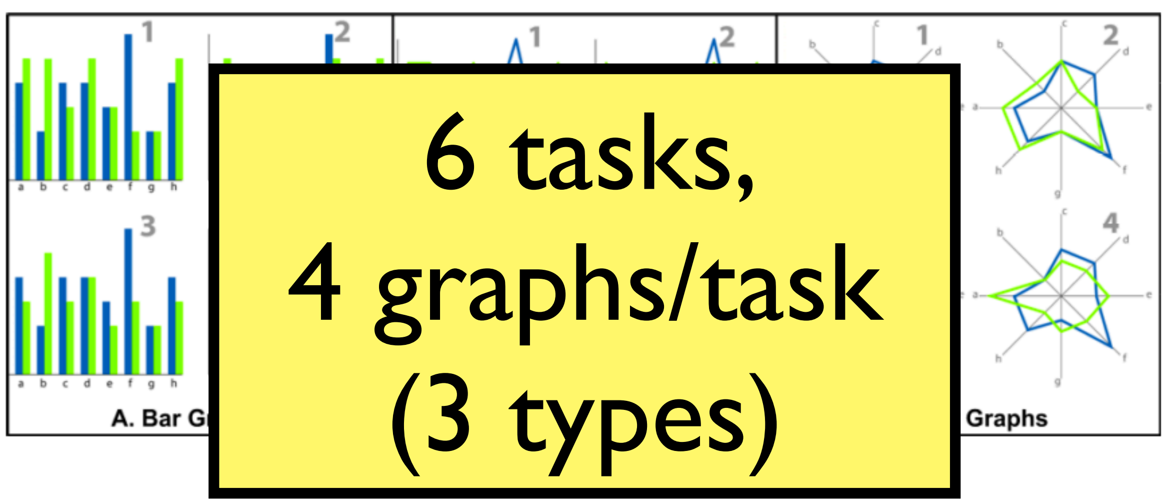
Design comparison using eye movements



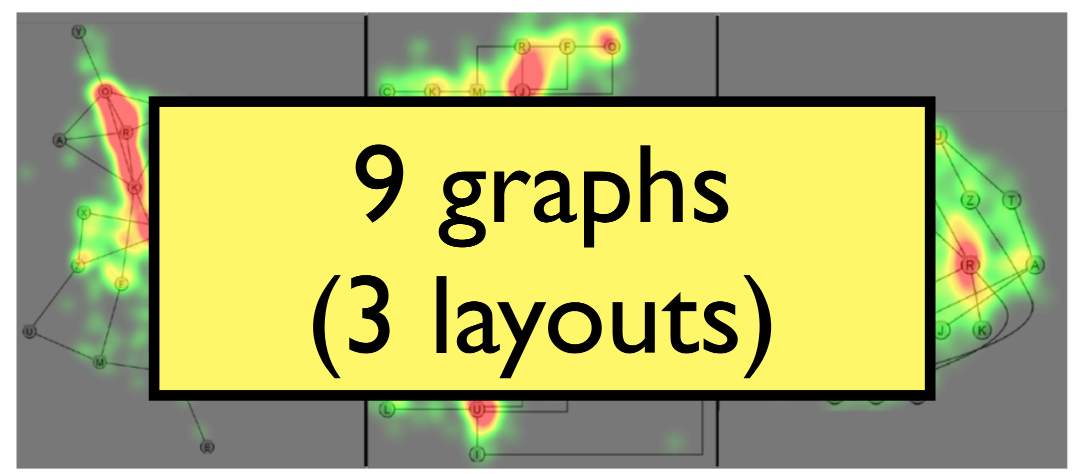
Burch et al. (2011)



Huang et al. (2005,2007,2009)



Goldberg & Helfman (2010)



Pohl et al. (2009)

However, these analyses have been relatively small scale – examining a small number of different graphs or visualizations each, with the designs often being simple and generated specifically for the study. This made it feasible to manually examine individual fixation maps and manually compare visualizations.

Scaling up

- large collection of "in the wild" visualizations
- different visualization types and sources
- many diverse participants

Instead, we are interested in scaling up analyses to evaluate and compare a large collection of diverse visualizations of different types and sources, for which we have the eye movements of a large collection of participants. At the end of the presentation, I'll mention how we continue to expand the amount of data we have, specifically observer attentional patterns. Here, we are interested in analyses and tools that can scale with large populations of visualizations and observers.

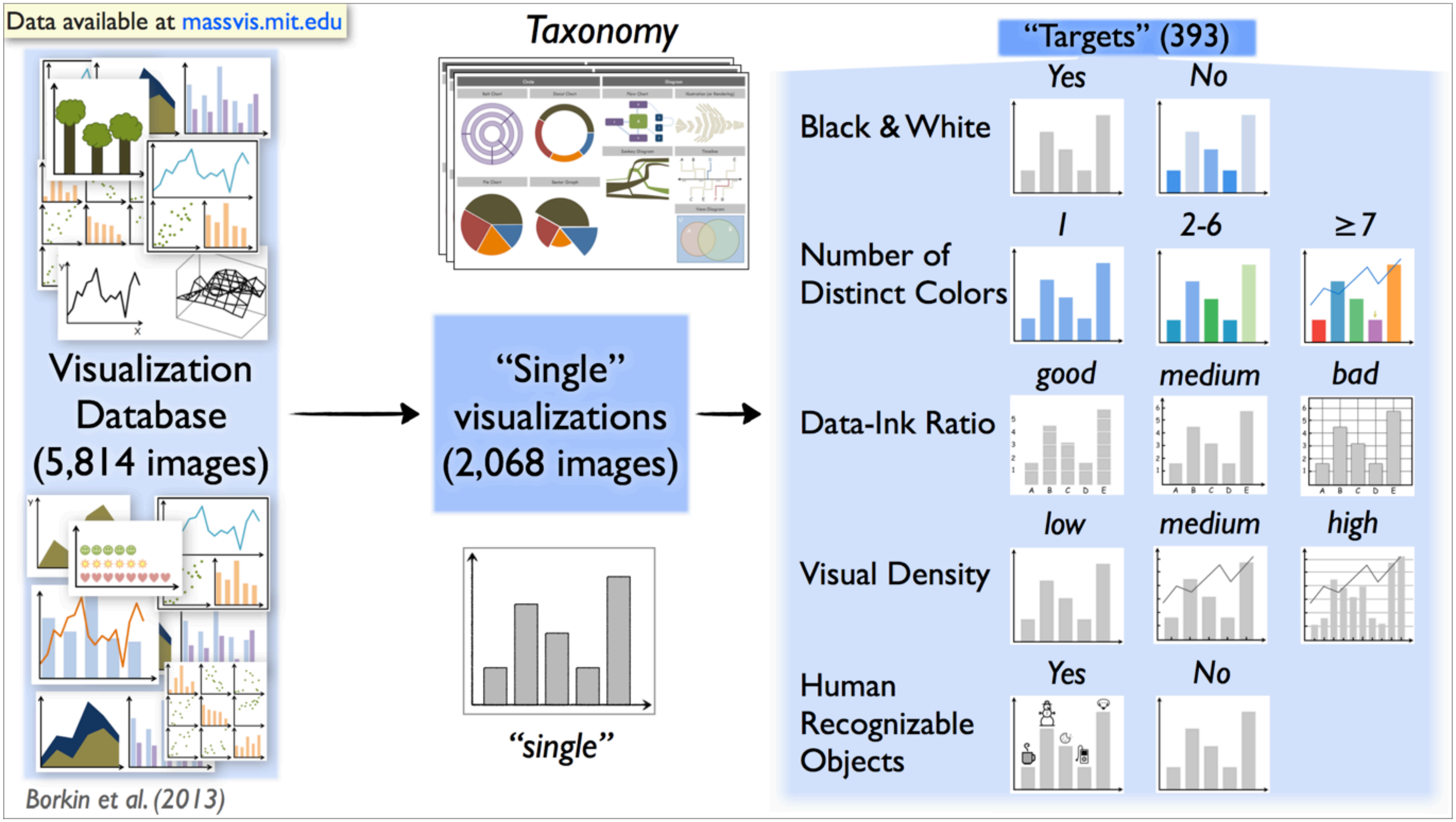
Scaling up

- large collection of "in the wild" visualizations
- different visualization types and sources
- many diverse participants

Goal: automatic, quantitative analysis

With the goal being to run automatic, quantitative analyses on all this data and be able to quantify the evaluations and comparisons.

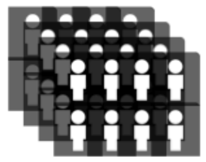
MASSVIS Dataset



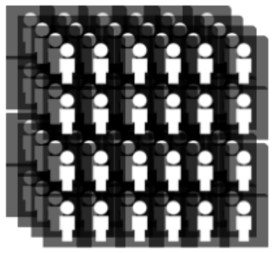
massvis.mit.edu

Naturally for this goal we need a large dataset. So I'm going to use this opportunity to mention that the dataset we (MIT and Harvard teams) have been working on for a few years is now being made available online at massvis.mit.edu. We have thousands of "in-the-wild" visualizations and for hundreds of them we have many manual annotations.

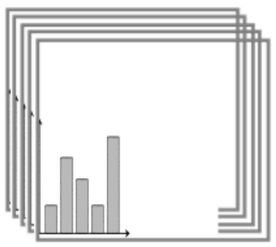
MASSVIS Dataset



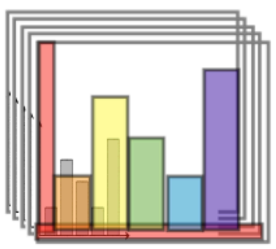
dozens of eye-tracking lab participants



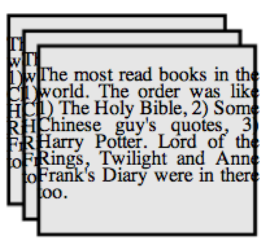
100s of online participants (MTurk)



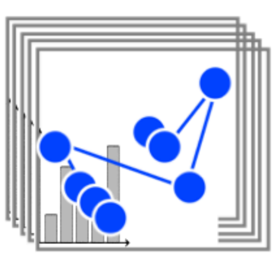
100s of diverse visualizations



1000s of labels and annotations



1000s of user text descriptions



10,000s of eye fixations

massvis.mit.edu

And among the data collected on this dataset, we have quite a bit of user data, including eye movements, which will be my focus for this presentation.

Beyond Memorability: Visualization Recognition and Recall

Michelle A. Borkin*, *Member, IEEE*, Zoya Bylinskii*, Nam Wook Kim, Constance May Bainbridge, Chelsea S. Yeh, Daniel Borkin, Hanspeter Pfister, *Senior Member, IEEE*, and Aude Oliva

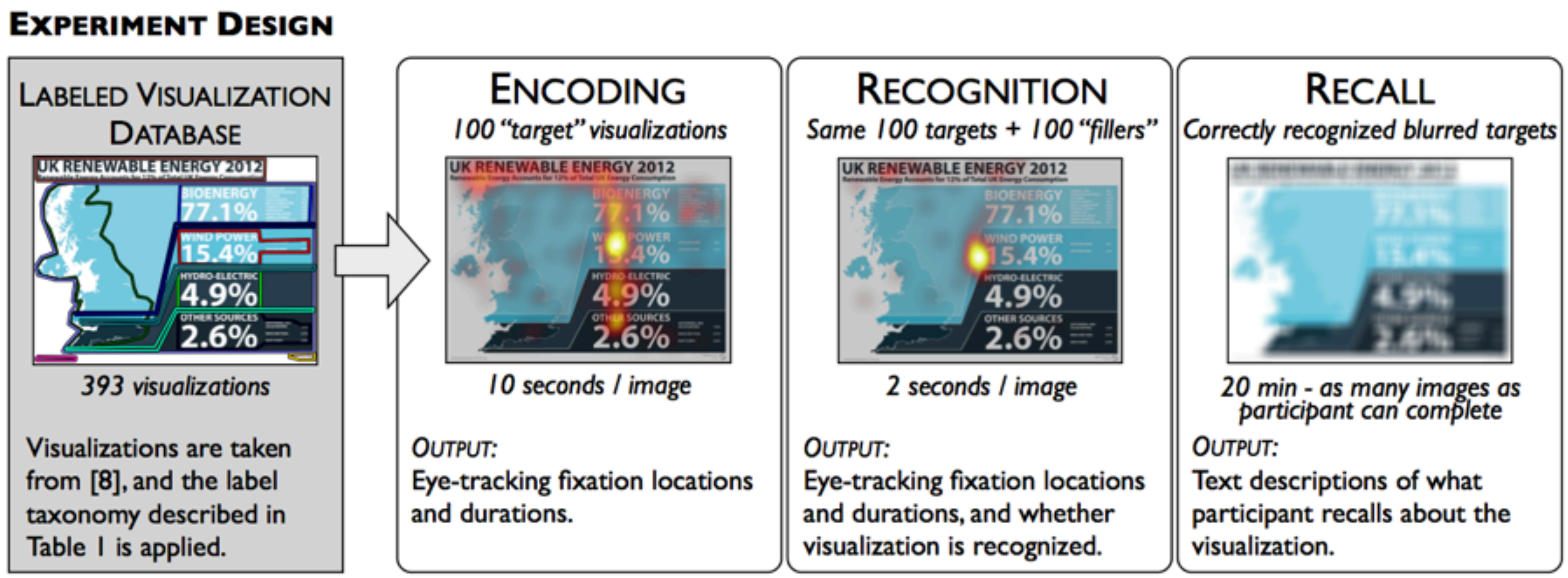


Fig. 1. Illustrative diagram of the experiment design. From left to right: the elements of the visualizations are labeled and categorized, eye-tracking fixations are gathered for 10 seconds of "encoding", eye-tracking fixations are gathered while visualization recognizability is measured, and finally participants provide text descriptions of the visualizations based on blurred representations to gauge recall.

Abstract— In this paper we move beyond memorability and investigate how visualizations are recognized and recalled. For this study we labeled a dataset of 393 visualizations and analyzed the eye movements of 33 participants as well as thousands of participant-generated text descriptions of the visualizations. This allowed us to determine what components of a visualization attract people's attention, and what information is encoded into memory. Our findings quantitatively support many conventional qualitative design guidelines, including that (1) titles and supporting text should convey
tograms do not interfere with understanding and can improve recogn
message. Importantly, we show that visualizations memorable "at-a-gla
the visualization. Thus, a memorable visualization is often also an effe

Index Terms—Information visualization, memorability, recognition, reca

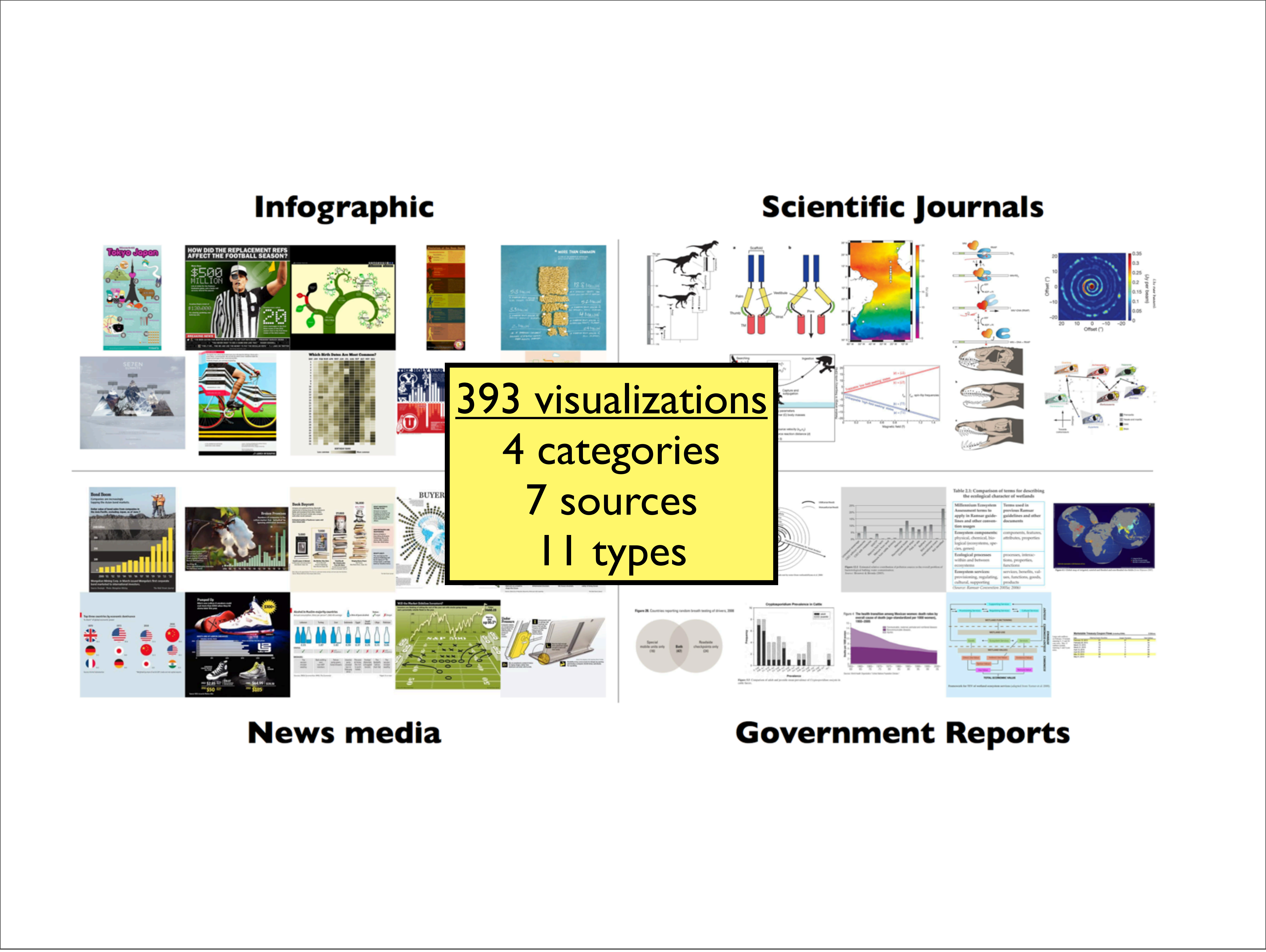
InfoVis: Human Reasoning
Thursday, Oct. 29
8:30-10:10 am, Grand

To find out more about the dataset, our experiments, and the conclusions we are able to make, I encourage you to attend our InfoVis talk this Thursday morning.

This paper

(a) ORIGINAL (b) SCAN PATH (c) DURATION PLOT (d) FIXATION HEATMAP (e) COVERAGE PLOT

The focus of this workshop paper and presentation is specifically on the eye movement analyses, going beyond what is presented in the InfoVis paper. Here we are interested in the general kind of inferences about design we can make from eye movements, how we can visualize them, and quantify them on this kind of data. We also provide a review of relevant visualization techniques and visualizations that scale to this task of large data analysis.

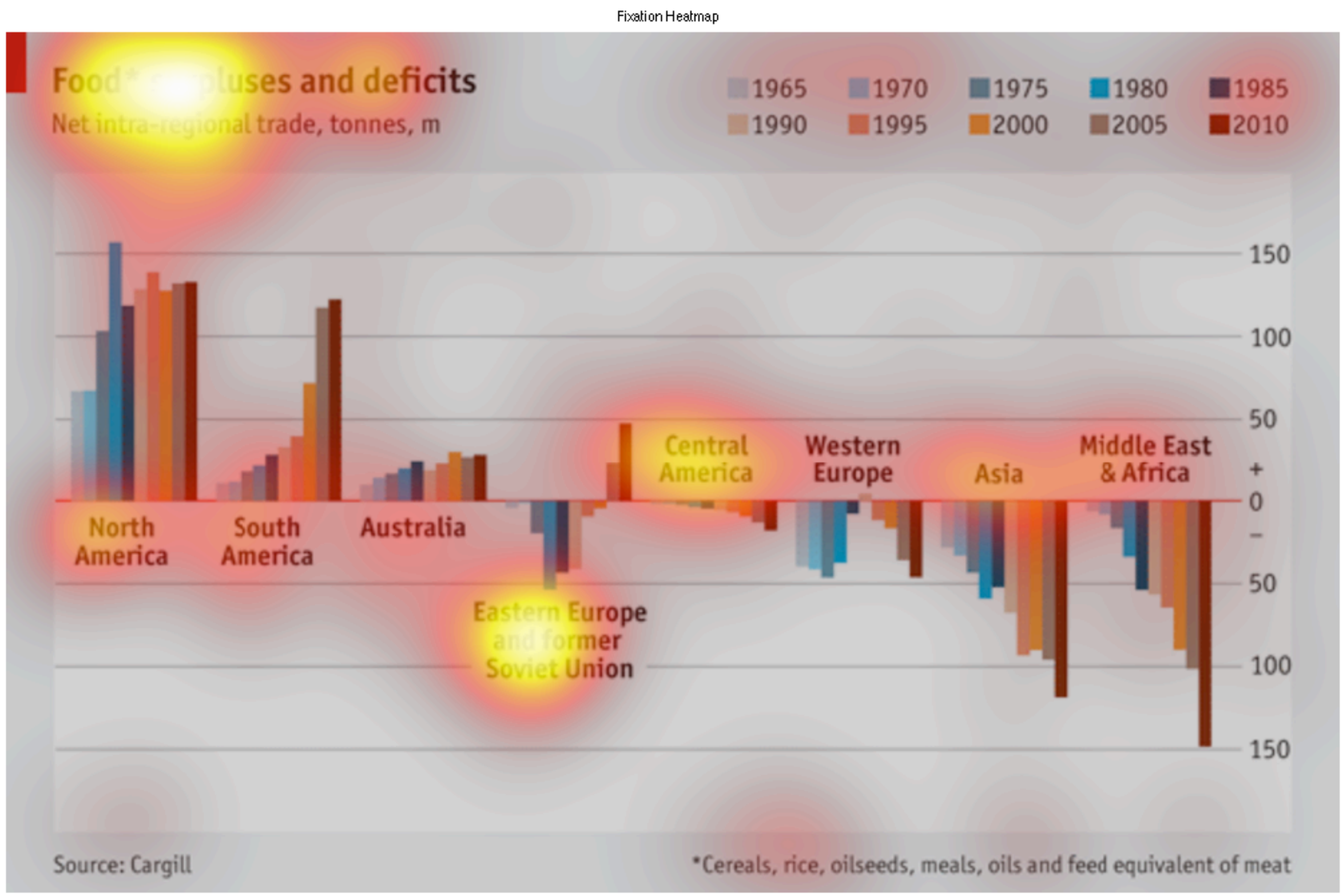


The kind of diversity we are dealing with in our analysis is hundreds of visualizations from different publication sources, divided into 4 different categories, and of 11 different types.

We consider
visualizations of eye fixation data of a population
to discover patterns across
different information visualizations,
and then we
quantify the patterns using eye fixation metrics.

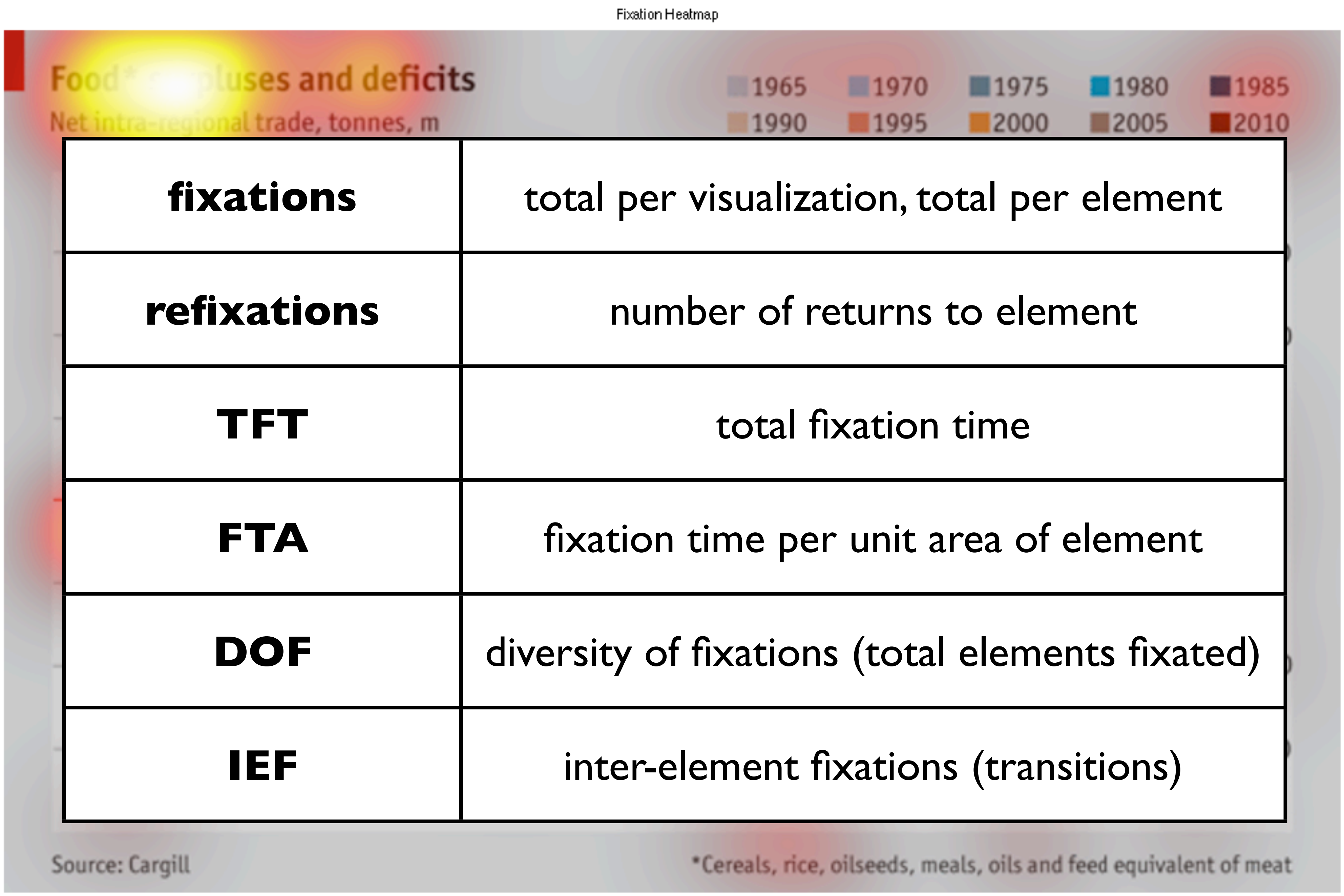
Given all of this data, we want to consider the kinds of visualizations that can help us discover patters in all of this eye movement data, and then metrics that can quantify these patterns. In the next few slides I'll mention some of these visualizations and metrics.

Fixation heatmaps



Many of you are probably familiar with fixation heatmaps, whereby fixations of a population of viewers are smoothed and overlayed to highlight the portions of a visualization that captured viewer attention.

Fixation heatmaps

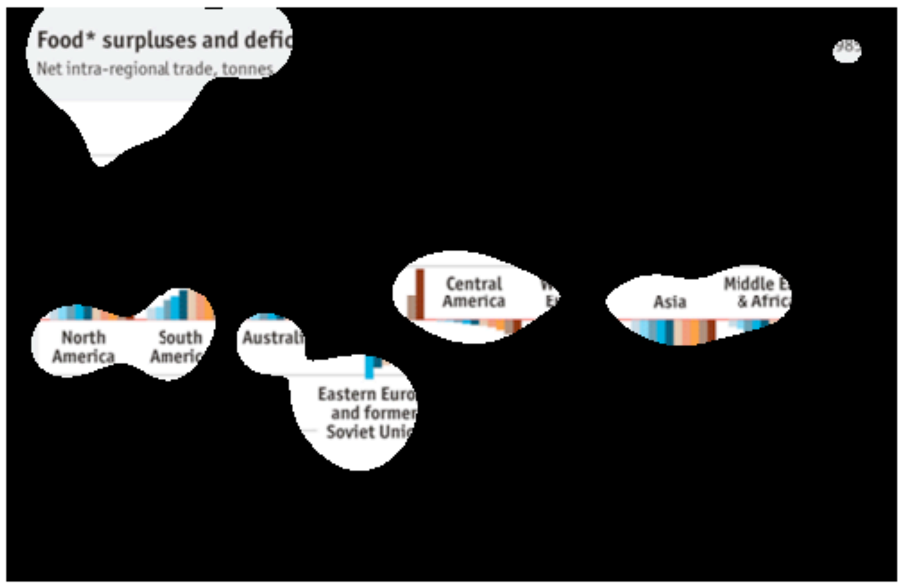


fixations	total per visualization, total per element
refixations	number of returns to element
TFT	total fixation time
FTA	fixation time per unit area of element
DOF	diversity of fixations (total elements fixated)
IEF	inter-element fixations (transitions)

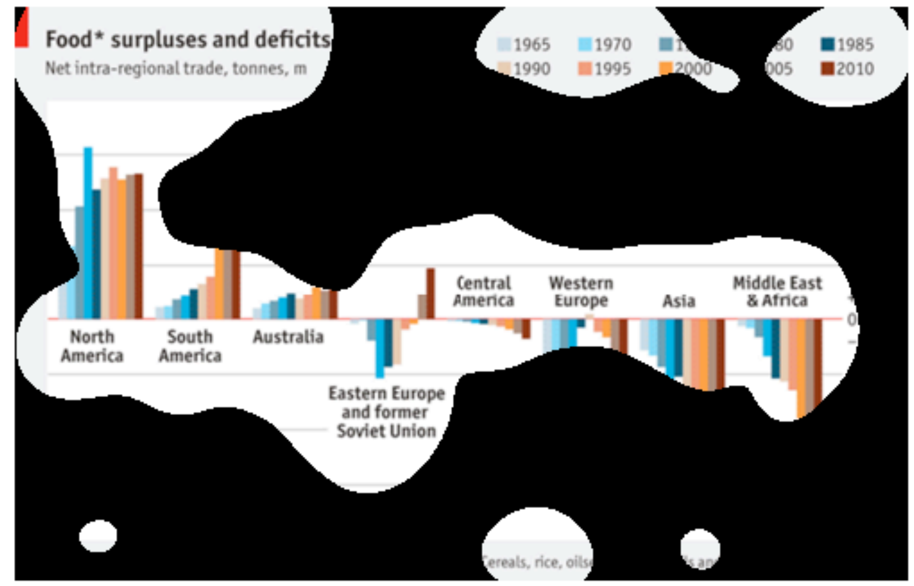
Given the location and duration of all of these fixations, there's a number of ways of quantifying this data, and running statistics over a large set of users and a large set of visualizations. For sake of time, I will not describe all of these here, but please refer to the paper for details.

Coverage

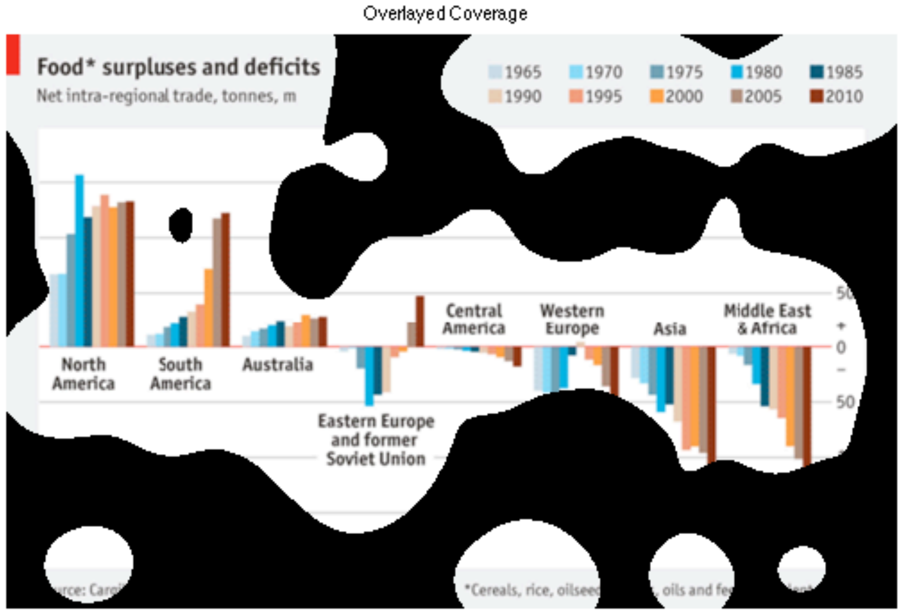
Overlayed Coverage



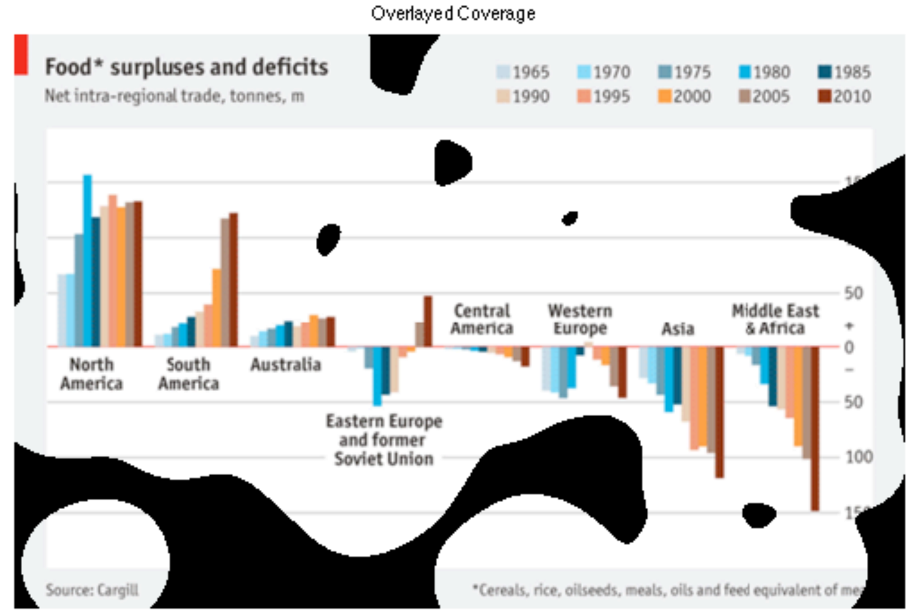
Overlayed Coverage



Overlayed Coverage

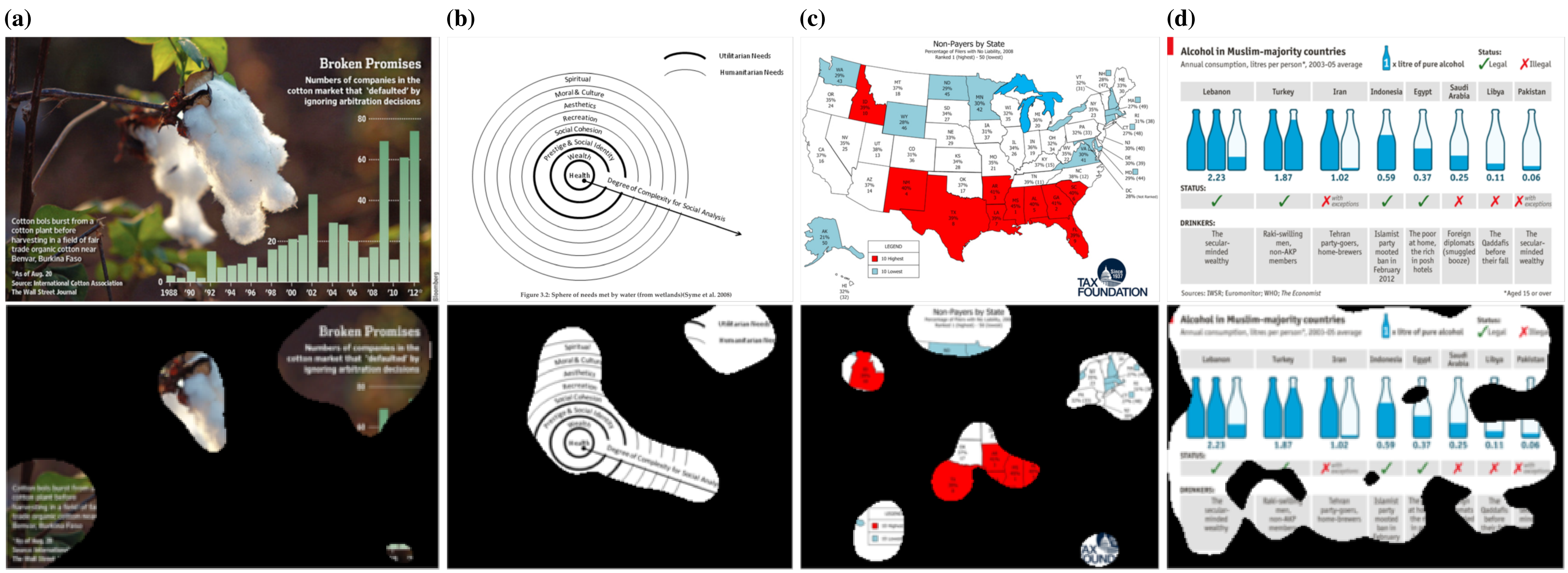


Overlayed Coverage



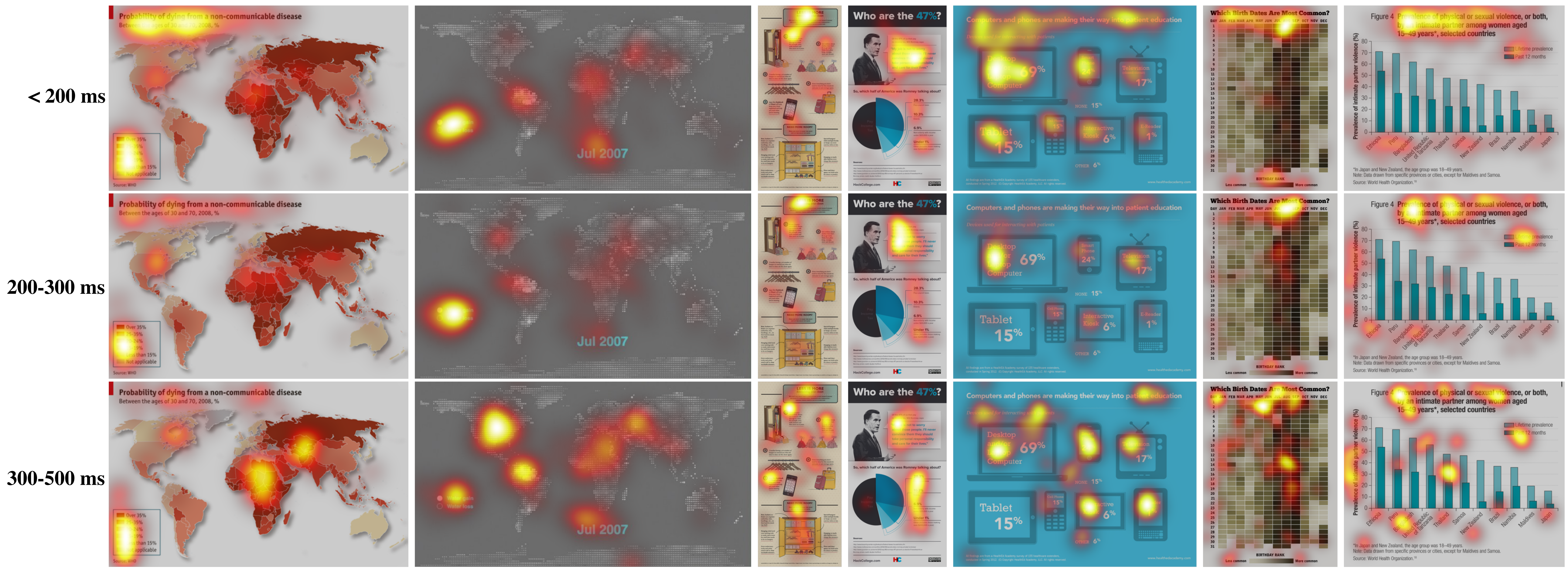
Coverage is another way to view the eye fixation data of a population, by thresholding the fixation heatmap at various thresholds. This gives us a sense of the relative importance of different areas of the display, and the portion of the visualization that most viewers visit.

Coverage



We can use this visualization to debug design and see common areas of the visualization missed by observers. Here, we used the same threshold for all 4 visualizations but have very different coverages. Note that in the first 2 cases, observers miss the bar graph and title, respectively. In the last visualization, observers visit most of the areas of the table, possibly integrating this information. Coverage is also easily quantifiable, since we can measure this area. Again, more details are available in the paper.

Duration plots



An alternative to considering all fixations at once, we can separately build fixation maps for fixations of different durations. This can be used to analyze which aspects of a visualization are examined for different amounts of time, and link this to cognitive levels of processing. For instance, fixations shorter than about 200 ms are considered automatic, and not necessarily conscious. We can see that with longer durations, participants begin to study more of the data aspects.

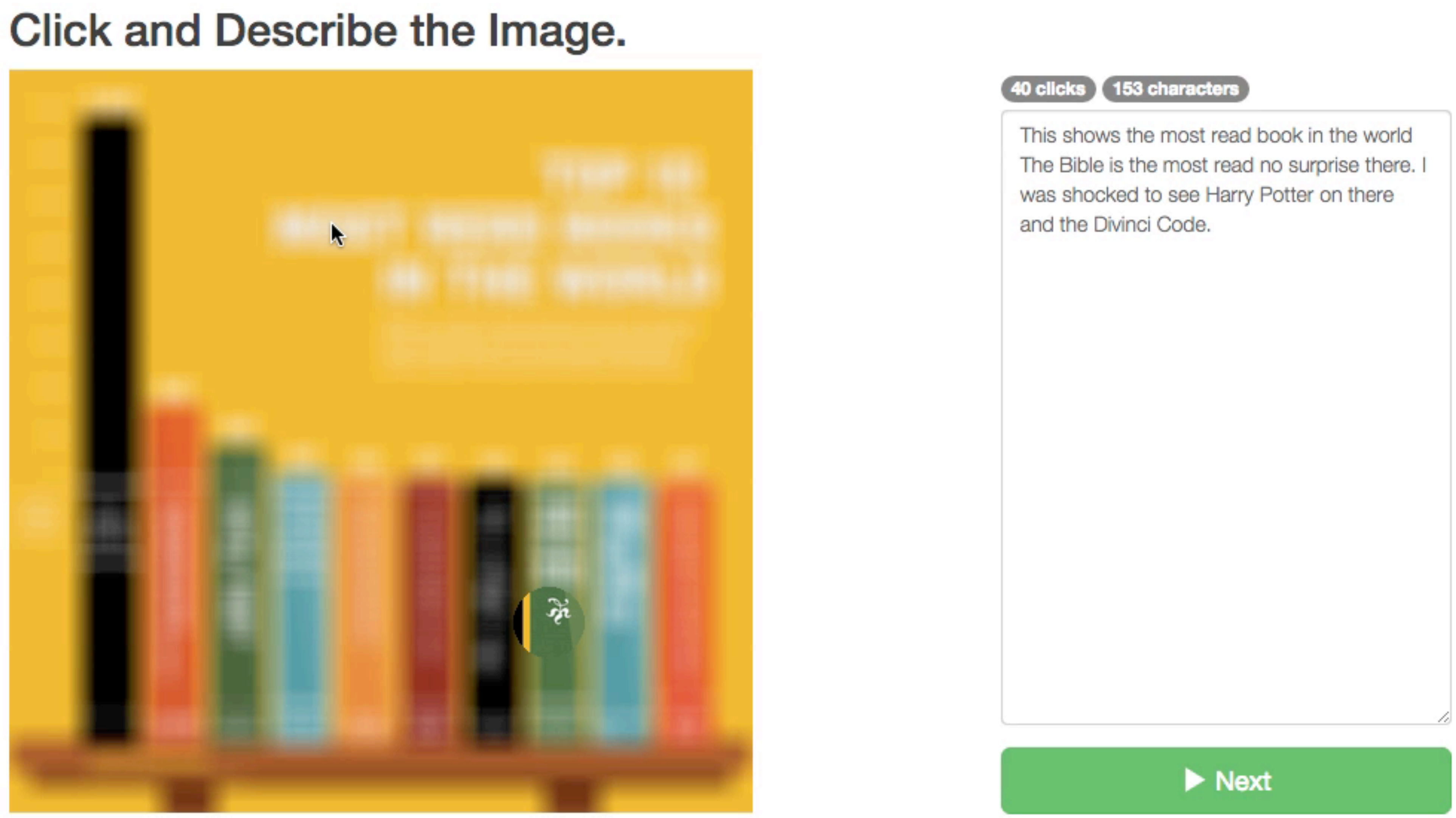
Inter-observer consistency

LOW IOC

HIGH IOC

By running similarity comparisons across the fixation heatmaps of different observers or groups of observers, we can quantify (IOC score) how consistent the eye movement patterns are across a population. This is informative of how well a design guides user attention. The relevant question then becomes: will everyone get the same information out of the information visualization?

Ongoing work



"A Crowdsourced Alternative to Eye-tracking for Visualization Understanding"

Kim, N.W., Bylinskii, Z., Borkin, M., Oliva, A., Gajos, K.Z., Pfister, H. (CHI EA'15)

These and other metrics are available in the paper. To give you a hint as to how we continue to expand our dataset to even larger groups of observers, please see our CHI workshop paper. Our preliminary studies demonstrate how click data on blurred visualizations (collected on MTurk) can approximate eye movement data collected in the lab, allowing us to move from tens of viewers to hundreds.

Thank you!

<u>Thanks also to:</u>

Nam Wook Kim (Harvard)

Aude Oliva (MIT)

Hanspeter Pfister (Harvard)