Investigating the Homogenization of Web Design: A Mixed-Methods Approach

Sam Goree
Indiana University Bloomington
Luddy School of Informatics, Computing, and Engineering
Bloomington, Indiana, USA
sgoree@indiana.edu

Bardia Doosti
Indiana University Bloomington
Luddy School of Informatics, Computing, and Engineering
Bloomington, Indiana, USA
bdoosti@indiana.edu

David J. Crandall
Indiana University Bloomington
Luddy School of Informatics, Computing, and Engineering
Bloomington, Indiana, USA
djcran@indiana.edu

Norman Makoto Su
Indiana University Bloomington
Luddy School of Informatics, Computing, and Engineering
Bloomington, Indiana, USA
normsu@indiana.edu

ABSTRACT

Visual design provides the backdrop to most of our interactions over the Internet, but has not received as much analytical attention as textual content. Combining computational with qualitative approaches, we investigate the growing concern that visual design of the World Wide Web has homogenized over the past decade. By applying computer vision techniques to a large dataset of representative websites images from 2003–2019, we show that designs have become significantly more similar since 2007, especially for page layouts where the average distance between sites decreased by over 30%. Synthesizing interviews from 11 experienced web design professionals with our computational analyses, we discuss causes of this homogenization including overlap in source code and libraries, color scheme standardization, and support for mobile devices. Our results seek to motivate future discussion of the factors that influence designers and their implications on the future trajectory of web design.

CCS CONCEPTS

• **Human-centered computing** → **User interface design**.

KEYWORDS

web design, design homogenization, historical analysis, computational social science, interviews

1 INTRODUCTION

As our interactions with the world become increasingly digital, designers of digital interfaces gain increasing influence over our lives. In the same way that architecture and interior design shape our offline behavior patterns by, for example, suggesting levels of formality or directing our attention [9], web design shapes our online behaviors. This influence is not simply about functionality: the *aesthetics* of an interface and the subtextual messages that that interface conveys about who or what a site is for can be just as influential [25]. For example, a website with large text and bright colors may signal that it was designed for children. If designers have this kind of influence over users' online lives, then the forces that shape those design choices may have a surprising amount of indirect impact on users as well.

As technical constraints on the web (e.g., bandwidth) have loosened and CSS and JavaScript have matured into expressive design tools, we would expect the variety of visual designs on the web to explode. Certainly the accessibility and power of such tools have led to an influx of creative remixes in at least the video medium [48]. Instead, at least anecdotally, the opposite has happened: a number of design blogs have posed questions like, "Why do all websites look the same now?" [2, 3, 8, 22, 27, 49, 50, 53, 55, 62, 66]. These posts typically point out common web design patterns and surmise that templates, common web frameworks, and the complexity required of today's sites (e.g., ensuring that they are accessible and responsive to multiple screen sizes) have led to the current state of designs. For some, homogenization is positive because it makes the Internet more usable and inclusive [66], while others argue that homogenization stifles creativity and negatively affects user experience by divorcing form from content [55].

For similar reasons, HCI research has begun to examine trends in website design. Chen et al. [20] speculate that web design has evolved in concert with technological advancements such as changing screen resolutions and sizes, new media formats, and the popularity of templates. These are described as distinct web "design periods." Moran [54] and Brage [13] address the rise of Web Brutalism, a related counter-trend against homogeneous designs. Additionally, there is a growing area of interest in the study of design history [29] and history is a growing area of interest in the study of

web design [41]. Purely aesthetic design qualities have been shown to influence perceived usability [12, 52], making design history important for studying the past, present, and future of interaction. Some of this work is motivated by a belief that understanding the history of web design can expand designers' repertoire, leading to innovative designs drawing from the past, for example.

Our central research question is: how has the visual homogeneity of the Internet changed between 2003 and 2019? We investigate these dates in particular because of the design periods observed by Chen et al. [20]: we believe that the homogenization trends that designers and scholars have observed may have occurred within and between the "Chaos," "Formative," and "Condensation" periods that they identify.

We use a mixed-methods approach to identify and explain homogenization in website design. We introduce and apply computational approaches to characterize visual properties of website designs, first focusing on a small dataset of 100 representative websites, and then scaling up to three larger datasets which contain over 200,000 snapshots of over 10,000 websites. To contextualize our computational findings, we conducted semi-structured interviews with 11 expert website design professionals (each having at least about 15 years of experience) involving their own historical portfolios. Our design history approach [24] is a first step to investigating, on a wide scale, the design evolution of the web. (However, we acknowledge that due to challenges of data availability and biases in our own experiences of the web (cf. Section 3.1.1), our sampling strategies tend towards sites from the American English-speaking web.) Additionally, our work contributes a modeling exercise to an existing conversation in the HCI community on how to quantify websites' aesthetic qualities [12, 40, 45, 60, 63]. While no computational approach may perfectly capture our intuitive notions of design similarity, by analyzing when and why our metrics are right and wrong, we can come to a better understanding of those intuitions themselves.

We make the following contributions:

(1) We introduce two complementary approaches for quantifying the perceptual visual similarity between pairs of website images, using (a) deep convolutional neural networks, which are data-driven and do not start from prior knowledge or assumptions about which characteristics of sites are important; and (b) hand-engineered features based on color and layout, which are easy to interpret and can be engineered to match our intuition.

(2) We find that the average difference between website layouts in our datasets *declined significantly—43%*—between 2010 and 2019, providing the first—to our knowledge—quantitative evidence that website designs are becoming more similar to each other.

(3) We investigate several potential explanations for this increase in homogeneity on the web including increasing overlap in front-end source code library usage, color scheme standardization, and support for mobile devices. Our findings suggest that use of a few specific libraries is a strong predictor of layout similarity.

(4) We also find that changes in the design process (which themselves are driven by an amalgam of cultural, professional, and technological trends) are at least partially responsible for the homogenization of the web.

The combination of two computational approaches applied to multiple large-scale datasets as well as qualitative analyses gives us high confidence in our results above.

Homogenization is not necessarily a cause for alarm. Consistent use of familiar design patterns can improve usability, since common patterns may have had rigorous usability testing and may already be familiar to new users [67, 74]. Websites which conform to style trends may also be more likely to satisfy accessibility standards [17] (although the converse is not necessarily true: accessible sites do not have to look visually homogeneous). Yet we believe that design in technology should be scrutinized in terms of its potential to be marginalized and commodified. Just as recent scholarship in HCI has, for instance, highlighted the often insidious ways in which algorithms support entrenched systems of patriarchy [42], we might wish to see how the designs we encounter on the web every day are shaped by dominant players. For example, the Mozilla Foundation argues that decentralization is essential to the long-term health of the Internet [31] to prevent companies from undermining privacy, openness, and competition. Arguably, the homogenization of design may signal that a few corporations have gained influence over what constitutes proper design on web. While the scope of our paper does not allow settling the debate on the benefits or detriments of homogenization in web design, we anticipate that it will provide the necessary groundwork to motivate such a debate in future work.

2 RELATED WORK

Our work lies at the intersection of two disciplines that crisscross HCI: design history and cultural analytics.

2.1 Design History

Design history argues that the seemingly mundane and anonymous work of countless engineers and designers is worthy of serious academic study [30]. While much work has focused on either design of objects and material culture or graphic design in printed material, the World Wide Web has also been analyzed by design historians. Engholm [28] argues that web design should be part of the corpus considered in design history and describes how concepts of genre and style from design history can be applied to the web. Chen et al.'s interaction design criticism sessions identified a series of design periods in the history of the web [20]. The latter argues that larger-scale work is needed to more holistically examine the design patterns in the history of the web. Our work takes a first step towards this vision, using quantitative methods to analyze large datasets in archives of the web to more broadly and systematically measure trends in web design.

Our research also aligns with efforts by design historians in constructing online galleries of web design, such as the Web Design Museum [43] and webmuseum.dk [1]. The process of acquiring, registering, and preserving websites as cultural artifacts is challenging to do at scale [29]. Research on information retrieval and extraction techniques such as ours may help future web design curators search the vast unorganized material available in collections like the Internet Archive [5].

2.2 Cultural Analytics

Online trends have been the subject of much study in cultural analytics research. Consumption patterns of social media videos [58] and photos [77] are significantly related to our cultural preferences. Various social media have been shown to predict trends in music as well [71]. Specific to web design, some work has analyzed raw HTML [40, 45, 60], but Javascript, CSS, and other code makes it difficult to study the visual design of a page without actually fully rendering it.

Visual analysis methods have been used in cultural analytics as well. Ushizima et al. [73] use visual features related to color, spatial organization, edges, and texture to quantify the visual differences between groups on Flickr. Metrics from computer vision have been used to identify stylistic and cultural trends in art history data. Saleh and Elgammal use metric learning on low-level image features to find metrics useful for classifying style, genre, and artist [65]. Influenced by their metric learning approach, we focus on finding useful interpretable representations and metrics specific to color and layout.

Cultural analytics researchers have also studied web design. Brady and Phillips investigate the relationships between color, balance, and usability [12], finding that websites modified to have spatially balanced pages and triadic color schemes are rated more usable than unmodified sites. Ben-David et al. [11] use a "visual distant reading" of color to study the websites of Yugoslavia between 1997 and 2000 and observe both a decrease in color diversity and a shift away from the colors of the national flag at the start of the Kosovo war. Cocciolo [21] automatically measures the quantity of text on web pages and finds that text density has been declining since a peak around 2005. Wobbrock et al. [76] find that the perceived credibility of news websites is heavily related to the visual design, and that the "overall gestalt of a page is responsible for participants' credibility judgments more than any single factor." Overall this work lends further credence to how web design is intimately entangled with its users and their societal context [20].

Recent work has used Convolutional Neural Networks (CNNs)—the state-of-the-art technique in computer vision—to analyze visual design. CNNs have the advantage of being able to learn "holistic" and domain-specific feature representatives instead of relying on hand-engineered features. Jahanian et al. [41] use both color and CNN features to study web history, and Doosti et al. [26] use CNN features to classify sites based on genre and year. Our work builds directly on these last two papers—which address methods for quantitatively characterizing designs—by focusing on a specific research question about the homogeneity of the web and directly comparing learned and interpretable features. In addition, we conducted semi-structured interviews with web design professionals to better understand, contextualize, and validate the results suggested by our quantitative analysis.

3 METHODOLOGY

We use a mixed-methods approach in our research, developing and employing computational methods (computer vision) to uncover large scale patterns of website design—namely, its increasing homogenization—and conducting semi-structured interviews with experienced web designers to identify sources of these patterns.

3.1 Computational Methods: Uncovering Large-scale Patterns in Website Design

One of our contributions is to try to quantitatively measure whether website design is becoming more homogeneous. We do this by collecting and automatically analyzing large-scale historical corpora of rendered images of web pages. While computational analysis of the semantic structure and textual content of the web has yielded important insights (e.g., ties between organizations) [15], recent work suggests that the visual design of websites encodes information about changes in design standards, technological innovations, and aesthetics [20, 26]. We thus use rendered website screenshots as our main source of data (see Figure 1 for an example).

To try to measure visual design differences at scale, we developed computer vision-based distance metrics that try to evaluate perceptual similarity of website images, and then applied these metrics to our corpora. Due to the subjectivity of the problem, there is no general computational metric that can accurately predict human perception of visual similarity in all contexts. Given this lack of a gold standard metric, we applied two very different approaches, with orthogonal strengths and weaknesses, to help avoid basing any findings on biases or artifacts of the metrics themselves.

First, using the *de facto* standard state-of-the-art techniques in computer vision, deep convolutional neural networks, we took a data-driven approach to learn classifiers that estimate the visual difference between web page images [26]. These deep learning models have become ubiquitous across nearly all problems in computer vision because they can learn complex features that may be difficult to describe algorithmically, including ones that predict human perceptual similarity [79]. However, this advantage is also their weakness: while these models typically perform very well, they are "black-boxes" [61] for which it is notoriously difficult to interpret the features they learn or how they make their decisions.

Our second approach is designed to be complementary by using hand-engineered representations of color and layout that we designed. Unlike the deep features, these hand-designed features are constrained by our prior assumptions about the design features that may be relevant, but (also unlike deep features) they allow us to better interpret the similarity results.

3.1.1 Data Collection. We collected a large-scale dataset of website images over time by using the historical crawls of the Internet Archive [5]. Just as there is no gold standard metric for comparing the visual similarity of websites, there is no obvious choice of URL corpus to use for our analysis. It may seem that the entire web (or a randomly-chosen subset) would be the ideal dataset, but such a corpus would not actually reflect the average web users' perception of the web: it would over-represent sites having many distinct pages, pages of the dark and deep web (which vastly outnumber those of the surface web [36]), and spam sites that arguably do not reflect the mainstream web. Moreover, random sampling of the web is difficult in practice because of its decentralized nature [14, 35], and the Internet Archive's collection, while impressively expansive, does not include a dense history of the entire web.

As a baseline dataset, we collect the homepages of large public corporations of the Russell 1000 stock index, as it existed in 2018 when we began the study. Indices such as Russell are commonly used to gauge the health of the overall economy, and we reasoned

that this sample might represent overall web trends, albeit biased towards corporate designs.

As a check against that bias, we verify our findings with two additional datasets based on different selection criteria: Alexa rankings [4] and Webby Award nominated URLs [57]. Alexa's rankings of websites have been used extensively, but its methodology is proprietary and has changed over time: it originally counted page loads through a custom user-installed browser toolbar, but now uses a variety of data. Historical Alexa rankings are also incomplete and the Internet Archive does not have regular snapshots of the full rankings. Webby Award-nominated URLs belong to websites nominated for an award in any "Websites" category. Since the structure of the Webby awards has changed over time, this is also an inconsistent sample, and only includes websites owned by individuals or organizations who pay an entry fee. In total, our three datasets constitute 227,802 images of 10,482 websites.

Verifying our findings on separate corpora collected with different selection criteria helps reduce biases related to any one dataset, and allows us to compare design trajectories of these different sets as well. We use Russell as our baseline dataset, since its selection criteria are publicly available and consistent (unlike Alexa and the Webby awards described above, which use proprietary or subjective criteria), but we replicate our main results on all three. We believe that variety in collection methodology and the datasets themselves lends robustness to our findings.

For each of the above three corpora of URLs, we use the Internet Archive [5] to fetch historical snapshots of the source code every 15 days, as available, from 2003–2017. For the Alexa and Webby sets, we do not look at the entire history of each site, but rather use snapshots within one year of each year that it appeared on the URL list. This makes this dataset reflect the changing population of popular and award-winning sites. Once we have downloaded the front-end source code for each site at each point in time, we render it as a 1200x1200 pixel image using PhantomJS [6].

Unfortunately, the Internet Archive does not have a dense historical collection for each website, and some sites have been indexed much less frequently than others. When working with data which is noisy and irregular, it is possible to find "trends" that are actually artifacts of methodological changes in the ranking and archiving processes. We thus conduct some experiments on what we call the *Dense Russell* dataset, which is a more complete dataset consisting of 100 websites of the Russell-1000 dataset which are available at least once every 15 days over the entire period 2003-2017. (This is not the same as the Russell 100 stock market index). Most of the 100 companies are *Consumer Services* (N=31) or *Technology* (N=29) according to the Russell 1000 categorization, while 15 are *Finance*, 13 are *Capital Goods*, 5 are *Energy*, 3 are *Health Care*, 2 are *Transportation*, and 2 are *Basic Industries*.

3.1.2 CNN Representation. We use deep learning with Convolutional Neural Networks (CNNs) to create a representation which automatically tries to quantify visual differences between websites. This representation is not based on our prior ideas about what elements of a design are important, but is instead learned automatically from training data. To do this, we train a CNN model on a classification task: given a website image (taken at some point in the period 2003–2019), identify which of the 100 companies in

(a)   (b)

(c)
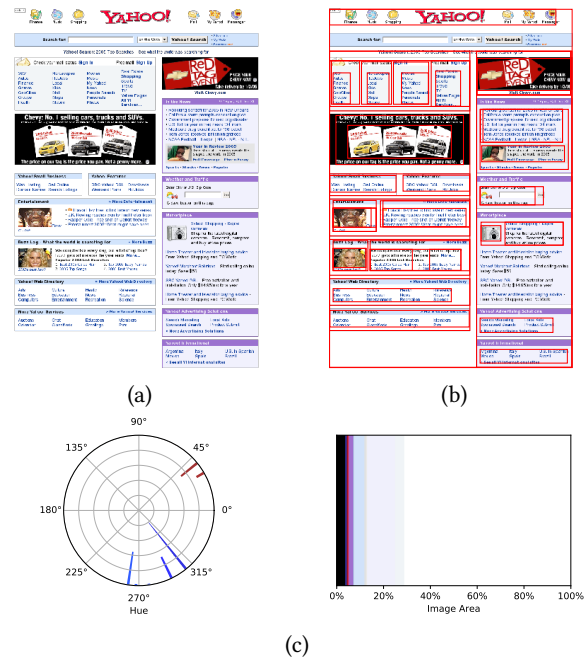
Figure 1: An example of our hand-engineered visual representations: (a) a sample website from the Alexa dataset, (b) nodes of the XY-tree decomposition, visualized as red boxes overlaid on the raw image, and (c) circular hue histogram and area plot (inspired by visualizations [11] showing the website's color distribution as fractions of the X-axis).

the Dense Russell that each page belongs to. In addition to the reasons above, we use the Dense Russell instead of the full Russell dataset because CNN models are known to be sensitive to class imbalance [16], so it is important that each of the classes has a roughly equal number of snapshots. Our goal is *not* to produce a classifier to solve this 100-way classification problem—we could have used a different task, such as autoencoding the input [75]. Instead, we are interested in the *features that the network learns* while learning to solve this problem, because these features are likely distinctive for characterizing and comparing images [70].

In more detail, we train a CNN model with 273 randomly-sampled images from each of the 100 company websites, each resized to 227×227 pixels, which is standard for CNN approaches. We use a canonical model architecture [44], the same model used in [26] except that the final fully-connected layer has 100 outputs, corresponding to the 100 classes of our classification problem; please see [44] for network details. Each output is a number between 0 and 1 that indicates the estimated similarity to that class (website). We use these outputs as a feature vector, and quantify the difference between two website images as the Euclidean distance between their vectors. Since we are using our deep learning model not as an accurate classifier but as a "data mining model" that lets us characterize any website using a feature vector, we use an established and well-studied model architecture rather than newer, more complex techniques.
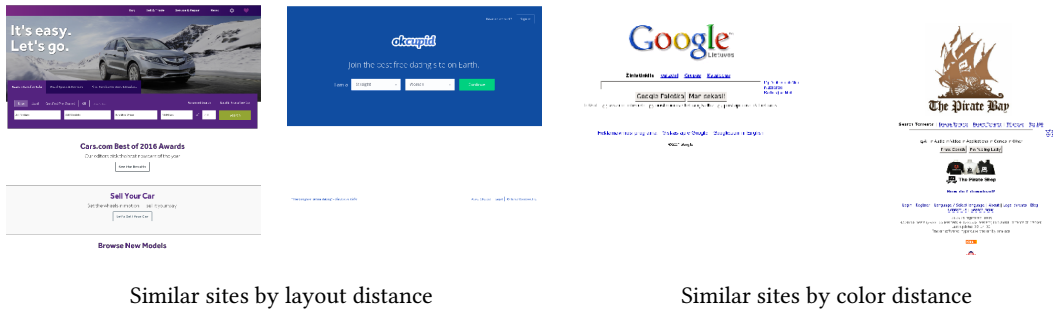
Similar sites by layout distance   Similar sites by color distance

Figure 2: Sample pairs of websites from Alexa dataset, having *left:* low layout distance, because both have a single large element across the top half and minimal content below, and *right:* low color distance, since they both are mostly white with blue and black text.

3.1.3 Color and Layout Representations. For our hand-engineered distance metrics, we make the modeling assumption that the visual design of a site is captured by two major characteristics: color scheme and spatial layout. While other research [63] has focused on collecting specific numerical features which capture these characteristics, we define distance metrics on our data representations directly. By using ideas from early query-by-image systems that search for images with similar color [64] and layout [51], we maintain the interpretability of our metrics without the added complexity of enumerating and weighting all the ways that two website images could look similar to one another.

We represent color scheme using color histograms (Figure 1), and use the Earth Mover's Distance (EMD) for measuring distances between them. Intuitively, the EMD captures the amount of "work" it would take to transform the color scheme of one website image into that of another, where one unit of work increments or decrements the color value of one pixel. Unlike simpler metrics like cosine or Euclidean distance which simply compare histograms on a bin-by-bin basis, the EMD incorporates the fact that nearby bins—e.g., similar colors—should be considered more similar to each other than distant bins. We use the CIEL*c*h* color space and measure the distance between two *individual* colors using the CIE 1976 delta E metric [23], which is a validated measure of perceptual color similarity. Our CIEL*c*h* histograms have $100 \times 182 \times 360 = 100 \times 256 \times 256$ bins. Figure 2 (top) shows an example of two non-identical sites with low color distance. We use the variant of EMD described in Rubner et al. [64] and the implementation by Pele and Werman [59].

We represent layout using XY-trees, which are created with a structure-based tree decomposition algorithm similar to that proposed in Ha et al. [34], although our implementation uses the algorithm from Reinecke et al. [63] (see Figure 1). The basic idea is to break up the page into a hierarchical structure of page elements by decomposing along gutters and solid colors of the page. Measuring the similarity of two trees is non-trivial; after much experimentation, we settled on using the tree edit distance, which measures the amount of work needed to transform the layout of one image into another by inserting, deleting, or relabeling elements. The cost of insertion or deletion is the size of the region being inserted or deleted, and the cost of relabeling is the symmetric difference between the

Participant ID	Web Design Experience	Job Title
P1	25 years	Coordinator of Instructional Design
P2	20 years	Retired Freelance Web Designer
P3	20 years	Systems Engineer
P4	14 years	Digital Editor & Web Designer
P5	20 years	Lecturer & Freelance Web Designer
P6	25 years	Vice President of Digital Marketing
P7	27 years	Graduate Student
P8	25 years	UX Strategist, Designer & Trainer
P9	22 years	Java & Web Development Trainer
P10	27 years	Web Accessibility Officer
P11	26 years	Web Designer & Joomla Certified Admin.

Table 1: Interview participant information.

old and new regions. We compute edit distances using an open-source implementation [38] of the Zhang-Shasha algorithm [78]. Measuring website layout difference using XY-tree edit distance is a novel contribution, to the best of our knowledge, though edit distances on XY-trees have been used before for document image retrieval [51]. Figure 2 (left) shows an example of two non-identical sites with low layout distance.

3.2 Qualitative Methods: Identifying Changes in Web Design Practices

Additionally, since images only hold a piece of the story of a website's design, and automated computer vision on large-scale image collections is a relatively blunt instrument, we also conduct a series of semi-structured interviews with experienced web designers. These interviews provide important historical context and link the computational trends to, for examples, specific changes in tools, techniques, and practices used in web design, as well as societal trends.

We sought out professionals and semi-professionals with at least 15 years of web design experience via snowball sampling over email and social media. We recruited 11 participants, of which 7 (64%) were men and 4 (36%) were women. Six (55%) were in North America, while 2 (18%) were in Europe, and 1 (9%) was in each of

South America, Asia, and Oceania, respectively. Semi-structured interviews were conducted remotely over the Zoom videoconferencing platform between May and July 2020, and each interview lasted between one and two hours. Before each interview, the participant was asked to prepare a portfolio of 4–7 representative websites in which they had been involved in the design process. To avoid confirmation bias, we did not pose direct questions regarding any hypothesized homogenization of web design; rather, the semi-structured interview protocol focused on identifying factors which shaped the participants' design decisions regarding the visual appearance of their chosen websites. The portfolios proved useful in grounding participant stories in concrete details of their web design practices [68, p.88]. All interviews were transcribed and anonymized.

3.2.1 Analytic Approach. We analyze our interview transcripts using a constructivist grounded theory approach [19]. Transcripts were coded by the authors to identify emergent themes with a focus on tools, processes, technologies, and historical events which shaped the designs of specific sites. Constructivist grounded theory was particularly apt because its inductive approach draws from both a literature review (e.g., work by Chen et al. [20]) and multiple focal points in data, such as our computational analyses. We initially coded interviews with open codes, then with axial codes focused on concepts which may relate to broad trends in visual design practices. Some codes described the tools that played key roles in the design process (e.g., "JQuery," "Adobe Photoshop," "Wordpress") while other codes described professional, cultural, and technological shifts in the design process (e.g., "negotiation," "designing for web vs. print," and "mobile-first design"). Throughout the interviewing and analysis, an iterative process of memoing led to a final set of themes, which was the genesis of our reported qualitative findings.

3.3 Limitations

Our study has several limitations. While we use two different modeling approaches for measuring the visual design similarity of pairs of websites, and compare results across three datasets, we cannot claim our measures of similarity are definitive. The CNN distance is taken from a model used to classify website identities and is not guaranteed to measure any particular visual qualities, although we find that it tends to correlate with our interpretable color and layout distances (see Section 4.1). Color and layout distances were hand-engineered based on our ideas about what makes sites look similar. All metrics for measuring subjective, aesthetic qualities are necessarily imperfect, and we welcome critical dialogue which may contribute towards better perceptual design similarity modeling in the future.

Also, as we discussed above, it is difficult to collect a representative sample of the web—or even to define what a representative sample should be. We use three different datasets with different selection criteria to try to avoid artifacts of any single one, but our selections can only reveal trends in the appearance of the English-language, and primarily American, web. Future work is needed to study how homogenization trends spread internationally and across language barriers.

While our interview participants were involved with the designs of several of the specific websites in our datasets, we cannot say for certain that their experiences explain the trends we observed through our computational analyses. Our selection criteria, which prioritized designers who had experienced the whole temporal range of our sample, does not capture the changing nature of web design as a profession.

4 RESULTS

We now turn to presenting the results of our mixed-methods analysis. We start by introducing our central finding: according to our measures, visual similarity on the web has been increasing, particularly since 2007. We then describe our investigations of potential underlying causes for this homogenization in layout and color as well as the significance of the time period from 2007. We organize these findings into three sections which investigate the relationships, respectively, between layouts and software libraries, color and image content, and the 2007 catalyst for mobile support and responsive design.

## 4.1 Homogenization in the Dense Russell Subset Over Time

We begin our analysis by examining the Dense Russell subset (cf. Section 3.1.1), which includes 100 websites for which snapshots are consistently available every 15 days in the Internet Archive. For each 15-day period, we computed the color feature for each of the 100 websites, and computed the average Earth Mover's Distance (EMD) between all pairs of these sites. We also computed the average Euclidean distance between CNN features as well as the average tree distance between layout features at each time period. Figure 3 presents the results, showing three significant trends:

- Sites became less homogeneous (higher values) between 2003 and 2007 and more homogeneous (lower values) afterwards.
- Color distance declines 32% between 2008 and 2019.
- Layout distance declines 44% from 2010 to 2019.

The CNN distance roughly follows the color distance metric with a decline of 30% between 2007 and 2019. Overall, this suggests that *websites have homogenized since 2007, and that layout in particular has seen a significant decrease in diversity.*

We emphasize the large scale of this analysis: while it only contains 100 pages, each page has a snapshot every 15 days for 17 years (408 snapshots per site), for a total of about 40,800 snapshots. At each temporal snapshot, each of the pages is compared to every other ($\frac{100 \times 99}{2} = 4950$), for a total of about 2 million comparisons (4950×408).

As mentioned in Section 3.1.1, we curated additional datasets, collecting according to different criteria, in order to verify our computational results. For both the Alexa rankings and Webby Awards dataset, we randomly chose a subset of 100 sites that had data available for each month, computed the pairwise distances between them for each month, and plot the results over time with confidence intervals and trend lines in Figure 4. Notably, all of the layout trend lines have negative slope, indicating homogenization over time, and the layout trend for the Webby data is particularly steep. On the other hand, the Alexa and Webby datasets show different color trends, with Alexa decreasing in homogeneity and the Webby dataset remaining relatively consistent across time.
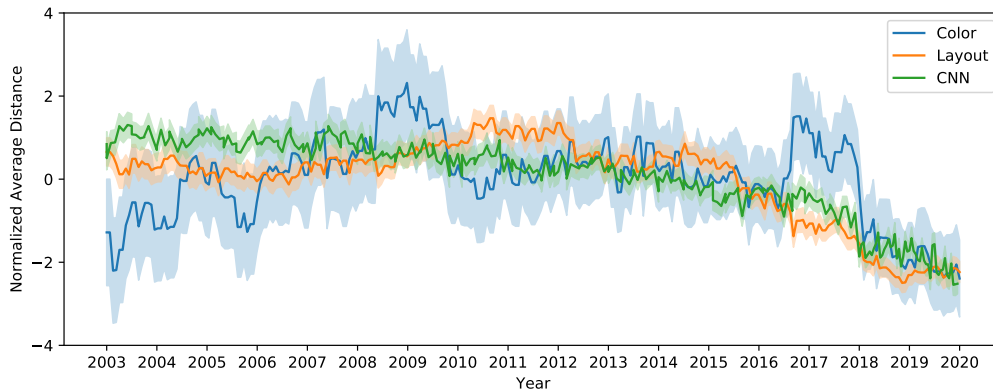
Figure 3: Average pairwise distance between the Dense Russell websites, plotted as a function of time, for color, layout, and CNN features. Shaded regions show 95% confidence intervals. To allow comparison across the three metrics (which have different dimensions), each plot was normalized to 0 mean and unit standard deviation. Lower values indicate more homogeneous.
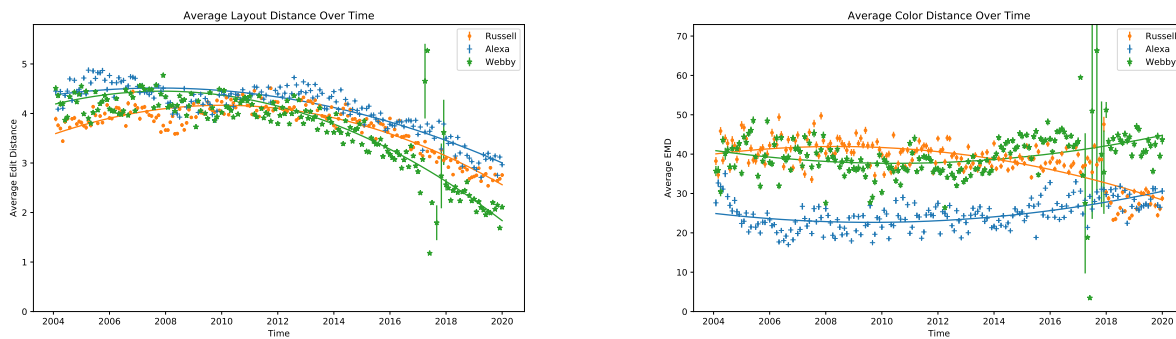
Figure 4: Average layout (left) and color (right) distances in all three datasets, with 95% confidence intervals and quadratic regression lines. All layout distance trends and the Russell color distance trend are negative with $p < 0.001$. The Alexa and Webby color distance trends are positive with $p < 0.001$. Large confidence intervals in the Webby data in 2017 are due to inconsistent data availability.

These results confirm the downward layout trend we observed with the Dense Russell subset, though each dataset has slightly different micro trends. The color distance plot highlights the noise in the Alexa and Webby data: since the set of highly-ranked Alexa websites changes over time, the mean similarity jumps up and down month to month. Meanwhile, the small number of Webby categories causes large confidence intervals in early years. The significantly lower color diversity of the Alexa data makes sense given the average color distributions in Figure 8: the Alexa websites have much more whitespace and fewer colored backgrounds than websites from the other two datasets.

One might argue that our trends align because they are cueing on the same image characteristics. To address this concern, we explored the relationship between our three distance metrics using multiple regression. The color and layout distances have no correlation with each other, as expected. We found that while the CNN distance is correlated with both color and layout distance, the coefficients are

relatively small (0.2 for color and 0.18 for layout, respectively (both $p < 0.001$)), indicating that our CNN distance is incorporating additional information beyond color and layout. The CNN has access to far more information about the image, e.g., it can use texture and shape features which do not appear in either the color or layout representations.

4.2 Layout Similarity and Libraries

Our analysis of our large-scale website corpora in the last section showed evidence of homogenization, especially in the website layouts. But what is driving this effect? In this section, we find evidence that increasing dependence on libraries, frameworks, and content management systems (CMS) in web design spurred layout homogenization. Moreover, over different periods of time, particular libraries have exerted an oversized influence on layout across the web. We also find that increasing expectations for responsive, accessible, and usable sites has made it increasingly difficult to
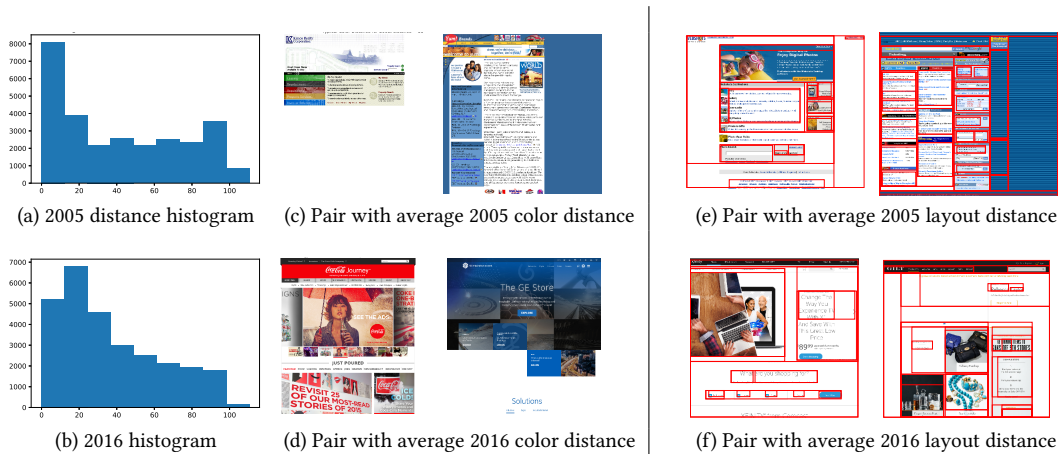
Figure 5: Visualization of average color and layout distances in 2005 versus 2016. (a) and (b): Histograms of color distances in 2005 and 2016, respectively. Note that this is a histogram of distances between color distributions, with units of image area × CIE color distance. The observed change between 2005 and 2016 is due to a shift from colored backgrounds to off-white and image backgrounds. (c) A sample pair of images with the average color distance between pairs of sites in 2005. (d) Same, for average color distance in 2016. (e) and (f) Same, with average layout distance in 2005 and 2016, respectively.

create unique, complex layouts that web designers of the past once could.

To give some intuition for our measure of layout similarity and how it has changed over time, Figure 5(e) shows a sample pair of websites from 2005 whose layout distance is the average distance in 2005, and (f) shows a sample pair in 2016 whose layout distance is the average of 2016. The red rectangles show the decomposition found by our layout analysis. As seen in these "prototypical" examples, early sites often use box-based layouts with fixed size and shape, while later layouts often have flat designs with fewer boxes and more images, presumably allowing them to scale to different sizes more easily.

4.2.1 Dependence on Packaged Code. One hypothesis for the increasing similarity of website layouts is that the rise of libraries, frameworks, and content management system (CMS) templates makes it easy to create new sites with a given predefined "look." Indeed, many of the participants in our interviews mentioned the impact that these have had on the design of the web. For example, participants P6, P8, and P11 pointed to CMS templates allowing users without web design expertise to produce high quality sites. P1, P4, and P6 described instances where they made use of code snippets copied from other sites on the Internet, because they were not experienced enough with Javascript or it was simply easier to appropriate existing snippets. Participant P5 identified an influx of software engineers into web design around 2012-2014 who over-use framework defaults due to a lack of design experience.

To test the hypothesis that reuse of website source code and/or the use of libraries, frameworks, and CMS templates may be driving the increasing website layout similarity, we performed analysis directly on the source code files of our Dense Russell dataset. We first compared the similarity of source code files themselves. For every pair of pages at a given time point, we used a string matching algorithm to compute the length of the substrings in common

between the two source code files over the sum of their lengths. Figure 6(a) presents the results. Surprisingly, we found that the content of websites (including their native CSS and JavaScript code) have become *less* similar over time, despite that visual appearance has become more similar. One possible explanation is that as the demands for online information content increase, the diversity and complexity of that content increases as well. Another explanation is that over time the amount of native code has decreased, and instead web designers have increasingly used libraries instead of writing (or copying) native code.

To investigate library usage, we used Wappalyzer [7] to detect libraries and back-end platforms from front-end web page source code for each page at each time point of our Dense Russell dataset. The tool detects 101 JavaScript and 47 CSS libraries. For every possible pair of websites at each time point, we computed the Jaccard similarity of the sets of libraries they use (i.e., the ratio of the size of intersection of the two sets over the size of the union). This metric ranges from 0 if two pages do not share any libraries in common or do not use any libraries at all, to 1.0 if they use exactly the same libraries. As Figure 6(b) shows, average similarity according to library usage has increased over time, especially after 2007, mirroring the results we found for visual similarity; the correlation between our CNN's measure of visual similarity and our measure of library similarity is 0.77 ($p < 0.001$).

Taken together, these results suggest that the increasing homogeneity of the visual web is not caused by increasing similarity of website content, but by increasing homogeneity in the choice of libraries that web designers use.

4.2.2 Library Monopolies. If increasing use of common libraries is causing websites to look more similar, then it should be that two sites using the same library should predict a decrease in visual layout distance. We perform multiple regression to predict layout similarity using binary features indicating whether two websites

(a) Similarity of code


(b) Similarity of library usage
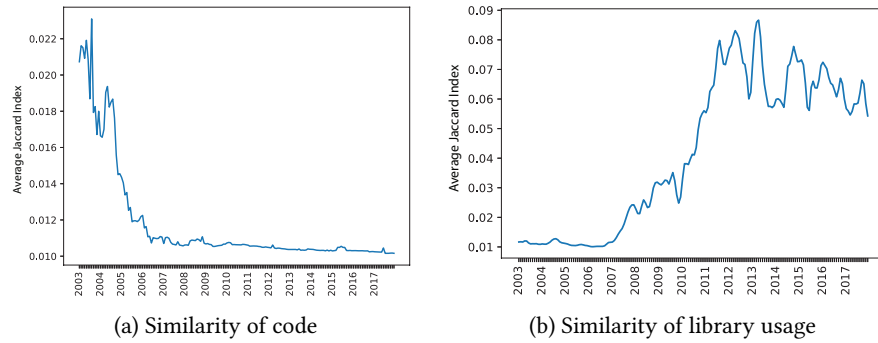
Figure 6: Average pairwise similarity between the Dense Russell websites, plotted as a function of time, according to (a) similarity of front-end source code, and (b) similarity of library use. We see that the source code itself has become less similar over time, while websites have used more and more libraries in common.

share each of the 16 most common libraries. The regression coefficients in Table 2 suggest that, yes, particular libraries are associated with increased similarity of web page layouts. Bootstrap correlates strongly with decreased layout distance, relative to other libraries. Other libraries, like SWFObject, which is used to embed Adobe Flash content, and jQuery tools, a Javascript user interface library, correlate with more varied layouts.

Several of our interview participants commented on the relationship between software tools and source code similarity. P5 and P10 commented on the low quality and complexity of code that has been auto-generated by a development environment (like Dreamweaver), CMS, or framework, indicating that the rise of libraries and frameworks may be contributing to the drop in code overlap. P4, P6, and P10 emphasize that these tools are essential time-savers to handle the complexity of the web, and P5 identifies a split between highly technical and non-technical tooling for the web: *"We have tools where we're splitting the middle. We have tools that are getting easier and easier…to use…where people can build websites with no knowledge of code whatsoever and communicate their ideas and live with whatever limitations those technologies put in place. And on the other end, if you want to get into the professional side of this, it is unbelievable the barriers to entry that have been put up."*

4.2.3 Shifting Layout Practices. Participants also discussed changing best practices in web design related to layout. For example, they recalled some of their considerations during the earlier days of the web: P1, P8, and P11 discussed using fixed-width wallpapers and background images for layout, P3 and P9 referenced a desire to keep content "above the fold"—within the first screenful of content (Figure 7)—so that users did not have to scroll, P5 described using search engines like Yahoo and AltaVista, which have highly complex layouts, as design examples, and P4, P5 and P9 described a process of designing a fixed layout in Photoshop and then "slicing" it into small images to put into an HTML table. These practices all but disappeared with the advent of CSS, which allowed sophisticated and reusable layout and styling of text content, and which was better for load times and search engine optimization.

In addition to saving time, libraries and frameworks often promote better usability and accessibility — a common topic in six of
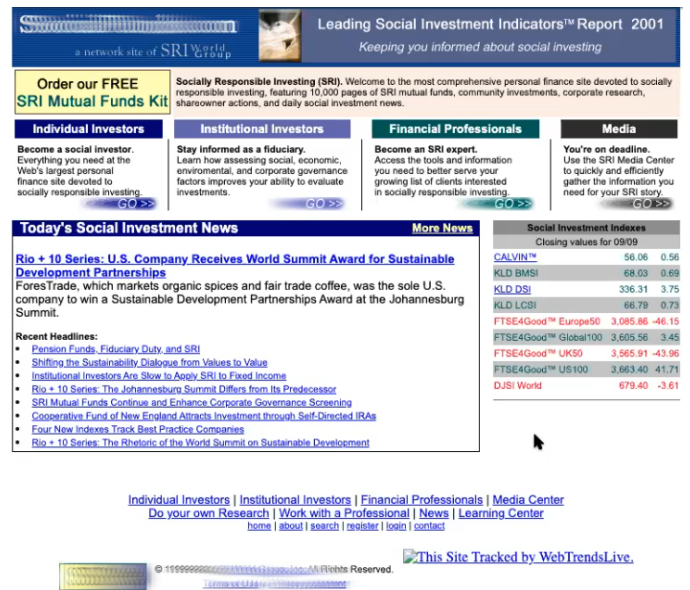
Figure 7: A screenshot from one of our interviews regarding layout practices. P5 explains, *"The boss was very, very concerned with above the fold…And so not a lot of white space and a lot of words and a lot of text that's a lot the same size."*

our interviews (P1, P2, P6, P7, P8, P10). P10 emphasized the deep relationship between accessibility, usability, and resilience: *"Make your designs as resilient as possible…They should work if somebody wants to blow the screen up to 300%, the interface should still work as intended."* Unlike the earlier mindset inherited from print design, web designers now see themselves designing flexible interfaces, not documents. P6 discussed that a web framework *"comes with a lot of accessibility code built in. So it's got ARIA [an accessibility standard for web applications] and stuff like that."*

| Library	Normalized Coefficient	N pairs
Bootstrap	-0.410	287
Font Awesome	-0.190	340
jQuery UI	-0.187	598
Underscore	-0.184	54
Moment.js	-0.173	25
Modernizr	-0.129	1489
ZURB Foundation	-0.007	52
Scriptaculous	0.055	31
yepnope	0.067	470
Prototype	0.089	62
jQuery	0.118	12523
React	0.199	36
Lightbox	0.204	23
SWFObject	0.373	1719
jQuery Tools	0.562	76

Table 2: Normalized regression coefficients for library overlap and layout distance. Negative numbers indicate that the presence of this library makes it more likely that two websites look similar, positive numbers indicate the opposite. N Pairs indicates the number of site pairs that had that library in common out of 65703 total pairs.

4.3 Color Similarity and Technological/Cultural Constraints

While it stands to reason that web frameworks would influence the layout of websites, it is less clear if this would explain shifts in color, given that frameworks allow color to be customized. So how can we explain the homogenization of color schemes observed in Figure 3? In this section, we observe that the homogenization in color schemes is related to the shift from monochromatic, usually white backgrounds to more off-white backgrounds and image backgrounds, and we find that this shift comes from a combination of technological and cultural constraints.

By plotting the full histogram of color distances, rather than just the average, we observe that, year by year, the color distances start as a bimodal distribution with modes around 0 and 90, then converge to a single mode around 10 (see Figure 5). As backgrounds have a large number of pixels, they cause the distance to swing heavily towards low values when both sites have similar backgrounds. Photos, on the other hand, have many different colors of pixels, so their distances are less polarized. We can also look at the average color distributions across time (Figure 8) and observe that off-white and dark colors have increased while black and white have decreased, which supports the notion that off-white and photo backgrounds are more common.

Interview data connects color distribution changes to changing technological and cultural constraints. In terms of technological constraints, Participants P5 and P9 identify the "web-safe palette," an early design constraint to ensure that web colors would appear correctly on different monitors. P9 mentions, "*I tend to compose colors in multiples of 51. So I would say, oh, that's a 51, 51, 102. Yeah, I still tend to do that despite the fact that we don't need to use web safe colors anymore.*" P10 points to early differences between monitor gamuts (the range of colors that can be displayed) as a reason to
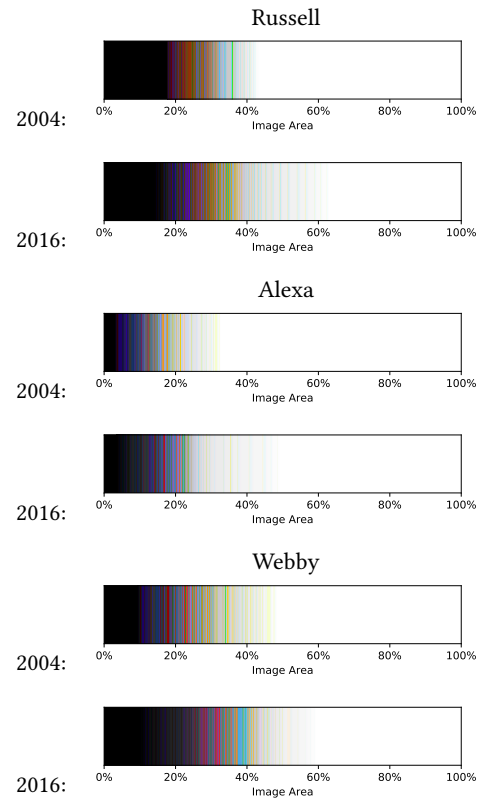
Figure 8: Average color distributions for three datasets in 2004 vs. 2016, in which bar width indicates the average fraction of each image made up of pixels of that color. In each dataset, the amount of dark and off-white increases and the amount of white and black decreases.

avoid color on the early web. P4 and P5 mentioned that bandwidth limitations prevented them from using much image content in the early days of the web. Since images were large files, they would load slowly over poor connections and negatively impact user experience. P7 identifies CSS and SVG as prerequisites to expressive use of color in web design, since they allowed a large degree of control over visual presentation without significantly increasing page load times.

Cultural constraints come from the changing nature of website design. P1 and P7 describe how, in the early days of the web, their stakeholders like clients or organizational leadership generally weren't interested in the details of the website. As the web has grown, these stakeholders have become increasingly invested in the look and feel of their organization's web presence, and website design and redesign has become a long, negotiated processes. Five participants (P1, P2, P3, P4, and P10) referenced negotiations spread over weeks or months focused on visual details like typeface or accent color during website redesigns. P1 describes the process for a website redesign in 2007: "*They had focus groups. They come in and they meet with large numbers of people and talk about their needs. And then they do a sitemap which kinda lays out the content, how*

it's going to be organized. And then…they gave us like five options of color schemes…And then our administration and faculty would basically vote on it." Because redesigns are more expensive and attract attention from stakeholders, designers are under more pressure to comply with branding guidelines and avoid unusual color combinations. P10 explains how a background color choice was made: "*This sky background kind of came out of the branding efforts to do with wings…And there was, this kind of illustrated the transition from IT running the show to marketing and communications running the show.*"

Wireframes — i.e., visual prototypes — are essential for facilitating these long design cycles. P2 describes wireframing with pencil and paper in 2004 while P8 references wireframing with software tools like InVision and Axure regularly from 2011 to the present. These wireframes function as drafts: they serve as a fixed point and ground discussions through the design process for specialists who may not have the time or expertise to edit the source code directly.

4.4 Catalyst for Mobile Support and Responsive Design

We observed that the abrupt change in web design visual similarity after 2007 (Figure 3) coincides with the release of Apple's first-generation iPhone. We hypothesized that this is not coincidental: the rapid need for mobile support may have caused the web to grow more homogeneous as companies began using similar front-end libraries to support mobile platforms. In this section, we show that our findings do suggest that visual similarity of website design has moved in tandem with increasing mobile support. Mobile design and therefore responsive design became de rigueur in the industry, pushed in part by search engine optimization (SEO) strategies.

To track mobile support in company websites over time, we analyze the CSS code in each web page's source code. If any of the CSS code has the "@mobile" keyword, we flag that website as supporting mobile screens in their design. The correlation between the average website CNN similarity in each month and the corresponding rate of mobile support is 0.84 ($p < 0.001$), which suggests that websites which support mobile tend to look more similar than websites which do not.

Our interview data strongly supports this relationship between mobile support and visual homogenization. Participants P5, P6, and P10 pointed to the rise of mobile web browsers as a major factor in their decision to adopt the philosophy of responsive design (i.e., where content "responds" to the size and shape of the browser, rearranging to best use the screen space, usually with a CSS framework). While any web page design can be made responsive by hand, doing so requires tedious work specifying content rearrangement and resizing rules using CSS media queries, and it is difficult from a design perspective to create layouts that are visually appealing in several arrangements. P5 and P9 point towards Apple's design decisions not to support Flash and emphasize scrolling over other forms of navigation on the iPhone. An increasing share of web traffic coming from mobile devices has led to a rise in "mobile-first" design practice that several participants (P2, P6, P8, P9, P10) would recommend to new designers. Interestingly, all five of these participants were quick to clarify that they do not personally use mobile-first practices in their design.

Four participants (P2, P4, P6, and P9) reference search engine optimization (SEO), and Google's policies in particular, as factors in their work. While not claiming to be "SEO expert[s]," they identified strategies for improving SEO. P4 said that he stopped using Flash for interaction because Google would not index text inside Flash components. P4 and P6 point to Google search's 2015 "Mobilegeddon" update, which suddenly prioritized search results which would display well on mobile, as the reason the web adopted responsive design. P2 and P9 say they use alt tags on images to improve SEO and P6 referenced using Google's Accelerated Mobile Pages framework as a new SEO strategy, describing Google as "*the 800-pound gorilla at this point. They're driving a lot of what happens on the web in terms of design and stuff.*"

5 DISCUSSION

We set out to investigate a very straightforward question — has visual design on the web become more homogeneous? Answering this question turned out to be a much larger-scale endeavor than we had imagined. We first collected a large-scale set of over 200,000 historical snapshots (over 15 years) of over 10,000 websites. To avoid bias associated with any single selection criteria, the dataset consisted of websites selected from three different sources: corporate sites of the Russell 1000 stock index, winners and nominees for the Webby awards, and top-visited sites according to Alexa. We then developed novel computational methods for measuring and characterizing the similarity of website images. To avoid bias associated with any single measure, we developed three, one based on deep learning and two hand-engineered features that characterize color and spatial layout features, respectively.

Across datasets and metrics, this large-scale, quantitative analysis showed strong evidence that website designs—especially with respect to page layouts—have become more similar, starting between 2007 and 2010 and continuing until at least 2016. To understand why, we analyzed, again at large scale, the similarity of source code, the use of libraries, and support for mobile devices. We found that the use of a relatively small number of frameworks and libraries has expanded significantly, and that the use of similar libraries strongly correlates with visual similarity—suggesting that the rise of these tools may be contributing significantly to the homogenization that we observed.

But this large-scale quantitative analysis could not reveal what was causing the uptake of libraries, or other more subtle factors that might be driving homogenization. We thus recruited and interviewed 11 web design professionals, each having at least 15 years of experience, to reflect on the changes in their design process over time. These qualitative data contextualized our quantitative results and confirmed or introduced several explanations, including the rise of software libraries and frameworks, increased use of large images, and support for mobile devices.

Layout homogenization appears to have begun after the release of the iPhone. Our interviews suggest that as mobile browsers grew in market share, responsive design and the libraries and frameworks that support them became an essential part of professional web design. This transformation was driven by libraries promoted by major tech companies, such as Twitter's Bootstrap and Font Awesome which is incorporated into the BootstrapCDN (content

delivery network), and jQuery and Modernizr which are both included in templates like Microsoft ASP.Net MVC. Responsive Web Design has changed how we experience the web [39] and largely replaced older best practices like placing content "above the fold." As developing for the responsive web requires more technical expertise, many designs shifted closer to library and framework defaults to avoid time-consuming, difficult development work.

Color homogenization offers a less clear story. Histogram data (Figures 1 and 8) suggests a shift from colored backgrounds to off-white backgrounds featuring images. Rather than being driven by the display limitations of new mobile devices, these changes respond to lifting bandwidth restrictions and increased support for CSS and SVG. While homogenization was the trend in the Russell 1000 sites, the opposite trend was true for the Alexa and Webby datasets, indicating that the shift to off-white backgrounds and large images does not consistently lead to color homogenization on less corporate websites.

We choose to reserve judgment on whether the trend of visual homogenization is good or bad. We suggest, however, that if the diversity of visual designs is indeed shrinking, this may limit the perceived repertoire of possible and legitimate designs that future website designers draw from, constraining the creativity and innovation of future websites.

Two decades ago, studies by Newman & Landay [56] noted that web designers used many informal, flexible representations of websites (e.g., sketches) for wireframing; these unconstrained, sometimes low-fidelity representations afford usual and unusual designs equally, limited only by the designer's imagination. Designers turned to media like paper for practical reasons: creating polished, high fidelity prototypes at the time was laborious.

Now, designers have a myriad of tools, including wireframing tools like InVision and Axure, CMS like Wordpress and Joomla, and frontend libraries and frameworks to rapidly prototype high fidelity websites. These tools allow designers to search for or choose from exemplar widgets and styles that follow the current landscape of *legitimate* designs and studied interaction patterns [47, 74], at the cost of making unusual designs more difficult to create [37]. The bias towards legitimate designs is helpful for both amateur and professional designers: using such tools and strategies leads to faster development cycles, reduced complexity, better accessibility, and use of learned affordances to achieve greater usability.

However, as HCI takes up other concerns such as value-sensitive, authentic, and reflective design [32, 69, 72], we may rethink how visual design (and closely related ideas about interaction) can go beyond concerns of usability, efficiency, or even marketing. We do not suggest turning back progress on, for example, creating inclusive websites or open-source software, but we should investigate how to enable enjoyable, diverse, and/or provocative forms within the space of inclusive design. For example, while animated GIFs may violate accessibility guidelines, the aesthetic qualities that make them popular indicate that we should not prematurely limit the creation of novel designs for accessibility reasons (see work by Gleason et al. [33] on making GIFs accessible).

Future work investigating this concern over homogenization should examine prototyping and development tools, both in the context of their literal interfaces, as well as how they serve as an interaction point for designers, developers, usability experts, and stakeholders. One could argue that such tools are inclusive, empowering a diverse group of people to become creators in the online world. However, the "deskilling" [18] of Internet users and increasing reliance on automated tools may be creating unhealthy conditions that restrict innovation to a small subset of web designers with deep technical expertise. To address these concerns, we will likely need to turn to methods and theories from the humanities [10]. Future work will need to revisit questions such as,

- How do prototyping tools both reflect and shape the space of legitimate designs?
- How do amateur designers, who did not meet the participation criteria for our interviews, create websites, and to what extent are their designs guided by the designs of these tools?
- How could prototyping tools be reimagined to better support unusual designs?

We anticipate that our method and strategies can empower designers, historians, activists, and policymakers based on their values and priorities. Brügger and Schoreder [15] argue that unpacking the history of websites allows us to analyze "societal developments at large." Organizations such as the Mozilla Foundation [46] posit that consolidation is detrimental to the "health" of the Internet. For example, our tools may be a first step for researchers to investigate whether or how homogenization of website designs may signal, overall, monopolization of design by a few companies.

6 CONCLUSION

We see our research as starting a larger conversation over whether the very conveniences developed to make the web more accessible to creators and users may also narrow the openness of design. Our mixed-methods analyses combining computational and qualitative approaches suggest that the range of designs we see online are reducing and that this may be driven, at least in part, by reliance on libraries, templates, and services that mitigate the complexity of creating modern web designs. Interwoven with these causes are broader technological and societal shifts in the web design profession and what users now expect in the web medium. The challenge then remains to see whether we can or should create an online design landscape that is diverse—without giving up the advantages that maturity and standardization provide users and creators.

ACKNOWLEDGMENTS

REFERENCES

[1] 2009. webmuseum.dk. http://webmuseum.dk/
[2] 2016. Longing For Innovation: Why Do All Websites Look The Same? Article on webydo.com. Retrieved from https://cmd-t.webydo.com/longing-for-innovation-why do-all-websites-look-the-same-9c80f5c41c61?gi=25360bd9ad1b.
[3] 2016. Why All New Websites Look the Same. Blog post on fontbundles.net. Retrieved from https://fontbundles.net/blog/why-all-new-websites-look-the-same.
[4] 2019. Alexa: About Us. https://www.alexa.com/about.
[5] 2019. The Internet Archive. http://www.archive.org.
[6] 2019. PhantomJS. http://www.phantomjs.org.
[7] 2019. Wappalyzer. https://www.wappalyzer.com/.

[8] 2019. Why Do All Websites Look the Same? Blog post on bigtuna.com. Retrieved from https://bigtuna.com/why-do-all-websites-look-the-same/.

[9] Thomas Allen, Gunter Henn, and Gunter Henn. 2019. *The Organization and Architecture of Innovation.* Routledge. https://doi.org/10.4324/9780080545370

[10] Jeffrey Bardzell and Shaowen Bardzell. 2015. Humanistic HCI. 8, 4 (2015), 1–185. https://doi.org/10.2200/S00664ED1V01Y201508HCI031

[11] Anat Ben-David, Adam Amram, and Ron Bekkerman. 2016. The Colors of the National Web: Visual Data Analysis of the Historical Yugoslav Web Domain. 19, 1 (2016), 95–106. https://doi.org/10.1007/s00799-016-0202-6

[12] Laurie Brady and Christine Phillips. 2003. Aesthetics and Usability: A Look at Color and Balance. *Usability News* 5, 1 (February 2003).

[13] Ellen Brage. 2019. The rise of brutalism and antidesign. http://www.diva-portal.org/smash/get/diva2:1304924/FULLTEXT01.pdf

[14] Niels Brügger. 2018. *The Archived Web: Doing History in the Digital Age.* The MIT Press.

[15] Niels Brügger and Ralph Schroeder. 2017. *The Web as History: Using Web Archives to Understand the Past and the Present.* UCL Press, London.

[16] Mateusz Buda, Atsuto Maki, and Maciej A Mazurowski. 2018. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks* 106 (2018), 249–259.

[17] Ben Caldwell, Michael Cooper, Loretta Guarino Reid, and Gregg Vanderheiden. 2008. Web content accessibility guidelines (WCAG) 2.0. *WWW Consortium (W3C)* (2008).

[18] Nicholas Carr. 2014. *The Glass Cage: Automation and Us* (1 edition ed.). W. W. Norton & Company.

[19] Kathy Charmaz. 2006. *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis.* Pine Forge Press.

[20] Wen Chen, David J. Crandall, and Norman Makoto Su. 2017. Understanding the Aesthetic Evolution of Websites: Towards a Notion of Design Periods. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. ACM, New York, NY, USA, 5976–5987. https://doi.org/10.1145/3025453.3025607

[21] Anthony Cocciolo. 2015. The Rise and Fall of Text on the Web: A Quantitative Study of Web Archives. 20, 3 (2015). https://eric.ed.gov/?id=EJ1077827

[22] Mary Collins. 2016. Web Design Trends: Why Do All Websites Look The Same? Blog post on friday.ie. Retrieved from https://www.friday.ie/blog/why-do-all-websites-look-the-same/.

[23] Comission Internationale de l'Eclairage. 1978. Recommendations on uniform color spaces, color-difference equations, psychometric color terms. *Paris: CIE* (1978).

[24] Carl DiSalvo. 2014. The Need for Design History in HCI. *Interactions* 21, 6 (Oct. 2014), 20–21. https://doi.org/10.1145/2669617

[25] Judith Donath. 2014. *The Social Machine: Designs for Living Online.* MIT Press.

[26] Bardia Doosti, David J. Crandall, and Norman Makoto Su. 2017. A Deep Study into the History of Web Design. In *Proceedings of the 2017 ACM on Web Science Conference* (Troy, New York, USA) *(WebSci '17)*. ACM, New York, NY, USA, 329–338. https://doi.org/10.1145/3091478.3091503

[27] Dave Ellis. 2015. All Websites Look the Same. Blog post on novolume.co.uk. Retrieved from http://www.novolume.co.uk/blog/all-websites-look-the-same/.

[28] Ida Engholm. 2007. Design History of the WWW: Website Development from the Perspective of Genre and Style Theory. *Artifact* 1, 4 (Jan. 2007), 217–231. http://dx.doi.org/10.1080/17493460802127757

[29] Ida Engholm. 2010. Research-Based Online Presentation of Web Design History: The Case of webmuseum.dk. In *Web History*, Niels BrÃijgger (Ed.). Peter Lang Inc.

[30] Kjetil Fallan. 2010. *Design History: Understanding Theory and Method.* Berg.

[31] Mozilla Foundation. 2019. Internet Health Report v. 1.0 2018. https://internethealthreport.org/2019

[32] Batya Friedman, Peter H. Kahn, and Alan Borning. 2006. Value Sensitive Design and Information Systems. In *Human-Computer Interaction and Management Information Systems: Foundations Advances in Management Information Systems*. Advances in Management Information Systems, Vol. 5. M.E. Sharpe, 348–372. https://doi.org/10.1007/978-94-007-7844-3_4

[33] Cole Gleason, Amy Pavel, Himalini Gururaj, Kris M. Kitani, and Jeffrey P. Bigham. 2020. Making GIFs Accessible. In *The SIGACCESS Conference on Computers and Accessibility (ASSETS '20)*. ACM, New York, NY, USA.

[34] Jaekyu Ha, Robert M. Haralick, and Ihsin T. Phillips. 1995. Recursive X-Y cut using bounding boxes of connected components. *Proceedings of 3rd International Conference on Document Analysis and Recognition* 2 (1995), 952–955 vol.2.

[35] Scott A. Hale and Victoria D. Alexander. 2017. Live versus Archive: Comparing a Web Archive and to a Population of Webpages. In *The Web as History: Using Web Archives to Understand the Past and the Present*. 45–61. OCLC: 1057656067.

[36] Bin He, Mitesh Patel, Zhen Zhang, and Kevin Chen-Chuan Chang. 2007. Accessing the deep web: A survey. *Commun. ACM* 50, 5 (2007), 94–101.

[37] Scarlett R. Herring, Chia-Chen Chang, Jesse Krantzler, and Brian P. Bailey. 2009. Getting Inspired!: Understanding How and Why Examples Are Used in Creative Design Practice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 87–96.

[38] Tim Hinderson. 2014. Zhang-Shasha: Tree edit distance in Python. https://github.com/timtadh/zhang-shasha

[39] Azham Hussain and Emmanuel OC Mkpojiogu. 2015. The effect of responsive web design on the user experience with laptop and Smartphone devices. *Jurnal Teknologi* 77, 4 (2015).

[40] Melody Yvette Ivory. 2001. *An Empirical Foundation for Automated Web Interface Evaluation*. Ph.D. Dissertation. UC Berkeley.

[41] Ali Jahanian, Phillip Isola, and Donglai Wei. 2017. Mining Visual Evolution in 21 Years of Web Design. In *CHI EA*.

[42] Matthew Kay, Cynthia Matuszek, and Sean A Munson. 2015. Unequal representation and gender stereotypes in image search results for occupations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 3819–3828.

[43] Petr Kovář and Ondřej Letocha. 2019. Web Design Museum. https://www.webdesignmuseum.org/

[44] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *NIPS*.

[45] Ranjitha Kumar, Arvind Satyanarayan, Cesar Torres, Maxine Lim, Salman Ahmad, Scott R. Klemmer, and Jerry O. Talton. 2013. Webzeitgeist: Design Mining the Web.

[46] Solana Larsen, Kasia Odrozek, and Jairus Khan. 2018. Internet Health Report. https://internethealthreport.org/2018/

[47] Brian Lee, Savil Srivastava, Ranjitha Kumar, Ronen Brafman, and Scott R. Klemmer. 2010. Designing with Interactive Example Galleries. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2257–2266.

[48] Lawrence Lessig. 2008. *Remix: Making Art and Commerce Thrive in the Hybrid Economy.* Penguin.

[49] Benek Lisefski. 2019. Data-Driven Design Is Killing Our Instincts. Article on medium.com. Retrieved from https://modus.medium.com/data-driven-design-is-killing-our-instincts-d448d141653d.

[50] Lexie Lu. 2015. Why Do All Modern Websites Look the Same? Blog post on designroast.org. Retrieved from https://designroast.org/why-do-all-modern-websites-look-the-same/.

[51] S. Marinai, E. Marino, and G. Soda. 2005. Layout based document image retrieval by means of XY tree reduction. In *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*. 432–436 Vol. 1. https://doi.org/10.1109/ICDAR.2005.150

[52] Eleni Michailidou, Simon Harper, and Sean Bechhofer. 2008. Visual complexity and aesthetic perception of web pages. *Proceedings of the 26th Annual ACM International Conference on Design of Communication*, 215–224. https://doi.org/10.1145/1456536.1456581

[53] Dan Miller. 2019. Why do all websites look the same? Blog post on awareweb.com. Retrieved from https://www.awareweb.com/blog/why-do-all-websites-look-the-same.

[54] Kate Moran. 2017. Brutalism and Antidesign. Article on nngroup.com. Retrieved from https://www.nngroup.com/articles/brutalism-antidesign/.

[55] Boris Müller. 2018. Why Do All Websites Look the Same. Article on Medium.com. Retrieved from https://medium.com/s/story/on-the-visual-weariness-of-the-web-8af1c969ce73.

[56] Mark W. Newman and James A. Landay. 2000. Sitemaps, Storyboards, and Specifications: A Sketch of Web Site Design Practice. In *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (New York, NY, USA, 2000) *(DIS '00)*. ACM, 263–274. https://doi.org/10.1145/347642.347758

[57] International Academy of Digital Arts and Sciences. 2019. Webby Awards. http://www.webbyawards.com

[58] Minsu Park, Jaram Park, Young Min Baek, and Michael Macy. 2017. Cultural values and cross-cultural video consumption on YouTube. *PloS one* 12, 5 (2017).

[59] O. Pele and M. Werman. 2009. Fast and robust Earth Mover's Distances. In *2009 IEEE 12th International Conference on Computer Vision*. 460–467. https://doi.org/10.1109/ICCV.2009.5459199

[60] Xiaoguang Qi and Brian Davison. 2009. Web Page Classification: Features and Algorithms. *ACM Comp. Surv.* 41, 2, Article 12 (Feb. 2009), 31 pages. https://doi.org/10.1145/1459352.1459357

[61] Zhuwei Qin, Fuxun Yu, Chenchen Liu, and Xiang Chen. 2018. How convolutional neural networks see the world — A survey of convolutional neural network visualization methods. *Mathematical Foundations of Computing* 1 (2018), 149. https://doi.org/10.3934/mfc.2018008

[62] Paul Regensburg. 2014. 5 Reasons Why Most Websites Look the Same Today. Blog post on blog.raincastle.com. Retrieved from https://blog.raincastle.com/bid/105839/5-Reasons-Why-Most-Websites-Look-the-Same-Today.

[63] Katharina Reinecke, Tom Yeh, Luke Miratrix, Rahmatri Mardiko, Yuechen Zhao, Jenny Liu, and Krzysztof Z. Gajos. 2013. Predicting Users' First Impressions of Website Aesthetics with a Quantification of Perceived Visual Complexity and Colorfulness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France). ACM, 2049–2058.

[64] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. 2000. The earth mover's distance as a metric for image retrieval. *International journal of computer vision* 40, 2 (2000), 99–121.

[65] Babak Saleh and Ahmed Elgammal. 2016. Large-scale Classification of Fine-Art Paintings: Learning the Right Metric on the Right Feature. *International Journal for Digital Art History* 2 (October 2016). https://journals.ub.uni-heidelberg.de/index.php/dah/article/view/23376

[66] Yaron Schoen. 2016. In Defense of Homogeneous Design. Article on medium.com. Retrieved from https://medium.com/@yarcom/in-defense-of-homogeneous-design-b27f79f4bb87.

[67] Ahmed Seffah. 2010. The evolution of design patterns in HCI: from pattern languages to pattern-oriented design. In *Proceedings of the 1st International Workshop on Pattern-Driven Engineering of Interactive Computing Systems.* ACM, 4–9.

[68] Irving Seidman. 2006. *Interviewing as Qualitative Research: A Guide for Researchers in Education and the Social Sciences* (3rd ed ed.). Teachers College Press.

[69] Phoebe Sengers, Kirsten Boehner, Shay David, and Joseph 'Jofish' Kaye. 2005. Reflective Design. In *Proceedings of the 4th Decennial Conference on Critical Computing: Between Sense and Sensibility* (New York, NY, USA, 2005) *(CC '05).* ACM, 49–58. https://doi.org/10.1145/1094562.1094569

[70] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. 2013. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229* (2013).

[71] Benjamin Shulman, Amit Sharma, and Dan Cosley. 2016. Predictability of Popularity: Gaps between Prediction and Understanding. In *ICWSM.* https://aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13129/12754

[72] Norman Makoto Su and Erik Stolterman. 2016. A Design Approach for Authenticity and Technology. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (New York, NY, USA, 2016) *(DIS '16).* ACM, 643–655.

https://doi.org/10.1145/2901790.2901869

[73] Daniela Ushizima, Lev Manovich, Todd Margolis, and Jeremy Douglas. 2012. Cultural analytics of large datasets from flickr. In *Sixth International AAAI Conference on Weblogs and Social Media.*

[74] Douglas K. Van Duyne, James A. Landay, and Jason I. Hong. 2007. *The Design of Sites: Patterns for Creating Winning Web Sites* (2nd ed ed.). Prentice Hall. OCLC: ocm70911156.

[75] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. 2008. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning.* 1096–1103.

[76] Jacob O. Wobbrock, Anya K. Hsu, Marijn A. Burger, and Michael J. Magee. 2019. Isolating the Effects of Web Page Visual Appearance on the Perceived Credibility of Online News Among College Students. In *Proceedings of the 30th ACM Conference on Hypertext and Social Media* (Hof, Germany) *(HT '19).* ACM, New York, NY, USA, 191–200. https://doi.org/10.1145/3342220.3343663

[77] Quanzeng You, Dario Garcia-Garcia, Mahohar Paluri, Jiebo Luo, and Jungseock Joo. 2017. Cultural Diffusion and Trends in Facebook Photographs. In *ICWSM.* https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15704/14800

[78] Kaizhong Zhang and Dennis Shasha. 1989. Simple Fast Algorithms for the Editing Distance Between Trees and Related Problems. *SIAM J. Comput.* 18 (12 1989), 1245–1262. https://doi.org/10.1137/0218082

[79] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 586–595.