A Case for Design Localization: Diversity of Website Designs in 44 Countries

Manuel Nordhoff, Tal August, Nigini A. Oliveira, and Katharina Reinecke
Paul G. Allen School of Computer Science & Engineering, DUB Group
University of Washington
{nordhoff,taugust,nigini,reinecke}@cs.washington.edu

ABSTRACT
Adapting the visual designs of websites to a local target audience can be beneficial, because such design localization increases users' appeal, trust, and work efficiency. Yet designers often find it difficult to decide when to adapt and how to adapt the designs, mainly because there are currently no guidelines that describe common website designs in various countries. We contribute the first large-scale analysis of 80,901 website designs across 44 countries, made available via an interactive web-based design catalog. Using computational image metrics to compare the ~2,000 most visited websites per country, we found significant differences between several design aspects, such as a website's colorfulness, visual complexity, the number of text areas and the average saturation of colors. Our results contribute a snapshot of web designs that users in 44 countries frequently see, showing that the design of websites with a global reach are more homogenized compared to local websites between countries.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords
Website design; Localization; Quantified Aesthetics

INTRODUCTION
Design localization involves adapting visual aspects of a user interface to other countries, such as the layout and overall aesthetics, beyond simply changing the language [24]. The process is time-consuming and expensive [48], but research has shown that it is worthwhile [52, 17]. In particular, research has demonstrated that people in different countries have varying visual preferences for website designs [50] and that adapting the visual designs of websites to specific countries and cultures can improve user satisfaction and work efficiency [49].

However, companies and designers face two challenges when deciding whether and how to pursue design localization. First, they need to decide whether localizing their website is indeed needed: Do the website designs in a specific country differ from those in their own? To inform their decision, they may turn to the literature, which offers contradictory findings. While some studies found that aesthetic preferences and website designs differ between countries [50, 41, 10, 13], research has also found that the design of university websites does not significantly differ between countries, and that "rather than being a forum for different cultures, the Web has promoted the emergence of a cosmopolitan online culture" [55]. One problem is that prior results are based on subjective analyses of small samples of websites and usually focus on specific website categories. For example, Stenger et al. [55], who found that web designs across countries do not significantly differ, analyzed 40 university websites for each of four cultural dimensions from 10 countries. Cyr and Trevor-Smith [13], who found many differences between website designs, compared 90 municipal sites from three countries. Prior comparisons rarely factored in what Internet users mostly see (i.e., what the most popular websites are), and are often too old to be a reliable source of information. To help companies and designers decide whether to pursue design localization, a large-scale study of current website designs is needed.

Second, if designers decide to adapt their websites, they still need to find out what design elements to change. They can turn to marketing reports that provide an overview of competitors (e.g., [40]), but these reports do not include concrete design guidelines. The same holds for literature that has reported on design differences between countries (e.g., [14]), which, for the most part, only provides fuzzy design guidelines derived from comparisons between few countries.

To fill this gap, we analyzed a data sample of 80,901 websites from 44 countries, including the ~2,000 most visited URLs per country. To enable objective comparisons between website designs, we quantified each website's aesthetics by computing a set of 32 image metrics, such as the colorfulness, visual complexity, the number of image and text areas, and the saturation of colors.

Our findings show significant differences between the visual design of the ~2,000 most visited websites per country and that the design of websites with a global reach are much more homogenized compared to local websites between countries. More specifically, we make the following contributions:

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

DOI: https://doi.org/10.1145/3173574.3173911

Website design diversity across 44 countries: Our work shows that, broadly, website designs are different across countries; however, websites with a global reach (i.e., websites that have users from multiple countries) usually do not significantly localize their designs, while being among the most popular in every country. The results suggest that (1) adapting web designs to the visual norms of a specific country is needed in some cases, but that (2) a website's purpose, reputation, and network effect can trump the need for design localization.

Tangible design guidelines for specific countries: Using computational image metrics, we conducted quantitative comparisons between the designs of the most popular websites from 44 countries to detect visual differences. The results can help designers determine how to localize the designs of their websites.

A website design catalog: We developed an interactive tool, available at http://www.juxtapose.labinthewild.org, that provides access to our library of website screenshots and their corresponding image metric scores (e.g., the website's average saturation). Users can browse and sort the data to compare how specific image metrics manifest themselves in the website designs in different countries.

A snapshot of current web design: Our analysis and public dataset (available at http://www.juxtapose.labinthewild.org) provide the foundation for future research, such as conducting a longitudinal analysis to determine changes in web designs on a global level or between countries.

RELATED WORK

Aesthetic appeal is a fundamental visual design goal and an important factor in the decision making of website visitors [56]. It influences their ability to build trust [35], their willingness to further engage [44], their mental workload [5], and their purchase intentions [16, 22]. A positive aesthetic judgment was also found to increase the perceived credibility of websites [52]. In combination with good usability, aesthetics facilitate the evaluation of content, resulting in faster reaction times [30]. Clear and orderly designs, as described with the notion of classical aesthetics [33] have been found to improve perceived usability [57]. Similarly, expressive aesthetics describe creative and novel designs, which can invoke curiosity due to a perceived novelty of the site [33].

Website Design and Design Preferences Between Countries

Prior work has mostly compared websites at a small scale. For example, Callahan [6] compared 20 university websites from 8 countries and found similarities in the layout, structure, and differences in the use of graphical elements like colors, images and animations. Marcus and Gould [41] examined websites from 10 different countries and produced guidelines for the design localization of websites based on Hofstede's cultural dimensions. These guidelines were only partly confirmed by Goyal et al. [19] when comparing governmental websites from 5 countries. While Goyal et al. found similar differences in the design between websites in the US and China, they did not find the design guidelines confirmed for Brazil, Russia and India. In particular, the website samples from Brazil, Russia, and India did not have as many images as Marcus and Gould's guidelines suggested. They also did not have a high number of internal links, which Marcus and Gould found common in Russia and India. Such differences in findings are likely due to changing designs over time (Marcus and Gould's study was published in 2000, while Goyal's study is from 2012), a focus on specific website categories (e.g., Goyal et al. analyzed only governmental websites), and due to insufficiently large and representative website samples overall. Our work addresses these concerns by analyzing a large number of different website categories, comparing the designs between several countries, and contributing an updated snapshot of webdesign trends.

In contrast to the work mentioned above that compared existing websites, researchers have also analyzed how visual preferences might differ between countries. For example, Canadians were shown to prefer less colorful interface designs than do Nigerians [47]. Simialrly, Cyr [12] found significant differences between the visual preferences of users from eight countries. Reinecke and Gajos [50] confirmed and extended these findings by studying the visual preferences of around 40,000 participants from more than 200 countries. Their work resulted in quantified models of visual appeal that can predict a person's visual preference for a given website depending on the person's country, age, gender, and education level. Researchers have also found that preferences can be influenced by a person's prior experience with and exposure to design [20, 8], suggesting that visual preferences are dynamically changing over time. This is also suggested by the results of a recent analysis on the evolution of website design over the last decades [7]. The findings confirm that people tend to like designs that they have been previously exposed to – an exposure effect that means that familiar designs might be more likely to be perceived as appealing compared to unknown designs [61].

To cater to such diverse visual preferences, research suggests to not only localize a website's language, but also its design [24, 48]. One way to achieve this is by referring to the aforementioned design guidelines (e.g. [41, 10]), which compared websites from different countries and cultures. However, these studies were usually based on small samples and are often outdated given that website designs change over time [7]. Designers can also find information about country-specific design preferences in marketing reports [46], online magazines [53], handbooks [34, P. 71-76] and blogs [4]. However, these resources usually provide vague guidelines, such as "If your core market is any of the Arabic countries, bear in mind that they read right to left."

Quantification of Website Design

Due to the intangible nature of the design guidelines discussed above, researchers have increasingly attempted to quantify aesthetics in order to predict individual visual preferences [54, 62, 43, 2]. For example, Reinecke and Gajos [50] collected 2.4 million visual appeal ratings from people with diverse demographic and geographic backgrounds and correlated these ratings with a number of image metrics pertaining to the visual complexity and colorfulness of websites. Their results include quantified website design guidelines according to what people find most appealing. Our work aims to extend this work by showing what website designs Internet users are currently most frequently exposed and analyzing how this might differ from people's visual preferences.

A DATASET AND TOOL FOR WEBSITE COMPARISONS

To enable statistical comparisons of website designs between countries, we first collected a dataset of website URLs from 44 countries, took screenshots of these websites, and quantified their visual design with computational image metrics. To facilitate a visual analysis of the screenshots, including comparisons of specific image metrics between countries, we additionally built a web-based tool. We will describe these different steps and artifacts in the following sections.

Dataset of Website URLs

To compare the designs of the most frequently visited websites between a variety of countries, we purchased a dataset of the top 2000 website rankings for each of 44 countries listed in Figure 2 using the API provided by the Alexa Top Sites (ATS) service [26]. We chose the same 44 countries studied in [50] to allow comparisons between their results (which focused on differences between users' visual preferences) and our own study (which focuses on what designs users are exposed to). We additionally decided to cap the number of websites at 2000 per country to obtain a large enough sample to include both more and less visited sites (given that the average user visits less than 100 websites in a month [9]).

The dataset includes several metrics that enable us to define popularity scores on both a country and a global level: First, the *popularity rank* and *access frequency* on a global and a country level enable us to distinguish between websites which are popular only within a country and those which reach a global audience (e.g., a website from one provider which is popular in numerous countries). The ATS service provides statistically meaningful numbers for rankings above rank 100.000 since the data quality and therefore accuracy of the rank formula increases or decreases synchronously to the traffic of a website. The data is derived from website owners who use the ATS service and millions of Internet users who installed browser extensions which collect Internet usage data, like the Alexa Toolbar [25]. The *popularity rank* on a global and country level is aggregated from a website's unique visitor count and the URL requests made by individual users. Higher combinations of both numbers lead to higher ranks. To enable an efficient comparison of a country's global reach, we calculated a website's *global score* by averaging its country ranks across all countries for which it appears in the top 2000 websites. For any country where the website was not included in the top 2000, we added a country rank of 2000 (the lowest possible rank).

We supplemented the ATS data with information about each website's category, which indicates its purpose, such as whether it is an e-commerce site, a news provider, or social network. Our aim was to control for differences in website designs depending on their purpose [28]. The categories were obtained using the service WebShrinker [38], which uses machine learning and human feedback to categorize websites and has a distinct set of top level categories, such as *shopping* and *news and media*. Note that while the categories are included in our dataset, they are not fully reliable as numerous URLs are assigned to wrong categories. We therefore did not include them in our main analysis and instead relied on manually comparing whether the distribution of categories differed between countries (they only did so for a minority of countries).

There are several other options for obtaining the data needed for our study, such as the free and crowdsourced DMOZ database [15], the WhoIs records [37], NetCraft [39], or SimilarWeb [40]. However, these services either did not fulfill our requirements for obtaining the website diversity, popularity, and access times for each URL (Netcraft and WhoIs), or they lacked a rigorous methodology for ranking websites based on their popularity (SimilarWeb and DMOZ).

Image Metrics to Quantify Website Aesthetics

To enable comparisons between websites, we first took screenshots of all 88,000 website URLs in our dataset using the WebKit tool PhantomJS [23]. Around 10% of our URLs contained dynamic content, such as videos, which were not properly captured by PhantomJS. For these URLs, we retook the screenshots with SlimerJS [29], which renders the content in a real browser; the tool is therefore slower, but able to capture dynamic content. All screenshots were taken with a size of 1024x768. While this format does not capture the entire website as a user would see when scrolling down (or with a higher resolution), it enables us to compare the website designs using statistical image metrics, which would otherwise be flawed given the varying website lengths. A positive side effect of this choice is that it measures what most users see at first sight, which has been subject to numerous previous research studies [36, 35, 51, 50, 58].

To quantify the visual design of all websites in our dataset, we computed a set of 15 image metrics (listed in Table 1) for each screenshot using the algorithms provided by the open source project VizWeb [42]. For each screenshot, we calculated its perceived colorfulness and perceived visual complexity based on computational models first introduced in [51]. The outcome of this calculation is a score between 1 (lowest complexity/colorfulness) and 10 (highest complexity/colorfulness). Colorfulness and visual complexity are highly predictive of visual appeal [50]. They are also two design aspects that people notice at first sight [51].

The computational models of perceived colorfulness and perceived visual complexity were developed based on subjective ratings of website screenshots and can almost accurately predict a person's perception of these design aspects [51]; the results of the models are therefore valuable to compare an overall design impression. To explain specific design differences between our websites in more detail, we supplemented these two metrics with additional image metrics from Reinecke et al. [51] used as independent variables in the models of perceived colorfulness and perceived visual complexity (listed in Table 1). This allowed us to derive design comparisons, such as "Websites in country X use more saturated colors than those in country Y", which is more detailed than comparing their overall colorfulness.

From the 88,000 website URLs in our dataset, we removed 6,408 entries for which we could not obtain screenshots. We

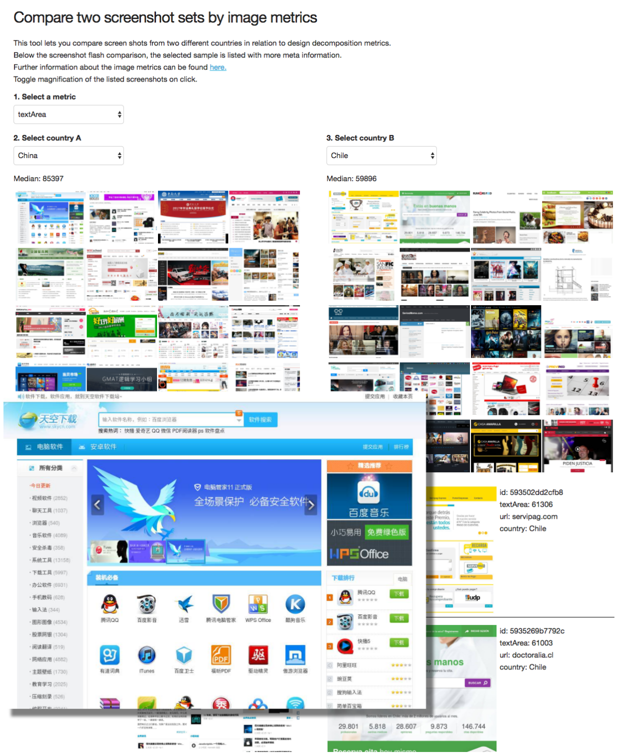

Figure 1. The interactive website design catalog, showing two selected countries compared by the average number of text areas. One screenshot from the details list is magnified.

then removed 689 entries for which we were unable to compute image metrics, resulting in a total number of 80,901 website screenshots along with their country-specific ranking and the computed image metrics. Of the 80,901 websites that we analyzed, 57% were represented in more than one country. Hence, the distinct count of URLs in our dataset is 37,856 (an average of 1,839 unique entries per country, sd = 100).

Visual Comparison of Websites Between Countries

While the image metrics described above enable us to empirically compare website designs, we additionally wanted to evaluate whether the results of the statistical analysis correspond to visually discernible differences between the website screenshots. To enable these additional analyses, we developed Juxtapose (`http://juxtapose.labinthewild.org`), an interactive design tool shown in Figure 1. Juxtapose's main functionality is the comparison of website designs between two countries. Users can select specific image metrics and observe similarities and differences between these countries both with the help of summary statistics that the tool reports, and with the help of 20 website samples from each country. Researchers and designers can use Juxtapose to (1) further understand our statistical results by visually comparing websites between countries and according to different image metrics, (2) extend our results by analyzing differences between the websites of various countries not captured by our image metrics, and (3) gain inspiration for designing websites for specific countries.

STUDY

Using the dataset and our Juxtapose tool introduced in the previous section, we conducted a comparative analysis of website designs in 44 countries. Four underlying research questions guided our analysis:

1. How does the visual design of the 2,000 most popular websites differ across countries?
2. Do globally popular websites differ across countries?
3. How do local websites across countries differ from the average global website design?
4. How do local websites differ across countries?

Analysis

To answer our first question, how website designs differ across countries, we ran analyses over the entire dataset. While our dataset corresponds to the 2000 most popular websites per country, the average Internet user is likely only exposed to a small subset of these sites, since Internet users rarely visit more than 100 websites per month [9]. However, depending on a person's interests, these 100 websites do not necessarily consist of the 100 most popular websites. Instead, a user could spend their time on generally less popular websites, or distribute his/her time between websites at the top and at the bottom of our dataset. We therefore employed a resampling technique to better represent a user's experience on the web in a given country. We sampled the data 1,000 times per country. Each sample included 100 websites randomly drawn from the 2,000 websites per country, for which we then calculated the mean value for a given metric (e.g., for colorfulness). We computed a one-way ANOVA over the resampled dataset. To identify top differing country pairs, we conducted post-hoc analyses using Tukey's tests for pairwise comparisons [59]. All p-values were adjusted for multiple hypotheses testing using Benjamini Hochberg correction [3]; effect sizes of our findings use Cohen's d [32]. Note that the resampling reduced variance in our dataset, increasing the effect sizes of differences between countries and likelihood of significance. We utilized our tool Juxtapose to observe visually salient differences of our analyses.

To answer our second question, we began by defining what constitutes a global website. The global score by itself does not tell us whether the website is truly global; in order to define this, we needed to find a threshold for the global score that defines the number of countries in which a website appears. To find this threshold, three authors from three different countries coded all website URLs with global vs. local based on their own knowledge of the sites and their brands and later resolved any discrepancies by discussing specific websites. The majority of global sites were identified to be among those that have a global score of 400 or less; hence, we define a global website as a website with a maximum score of 400. We conducted one-way ANOVAs to compare global players in each visual design metric between countries, and again applied a Benjamini Hochberg correction for multiple hypotheses testing.

We addressed question three, how local websites differ from an average global website design, by excluding all global websites from each country and collecting them into a separate

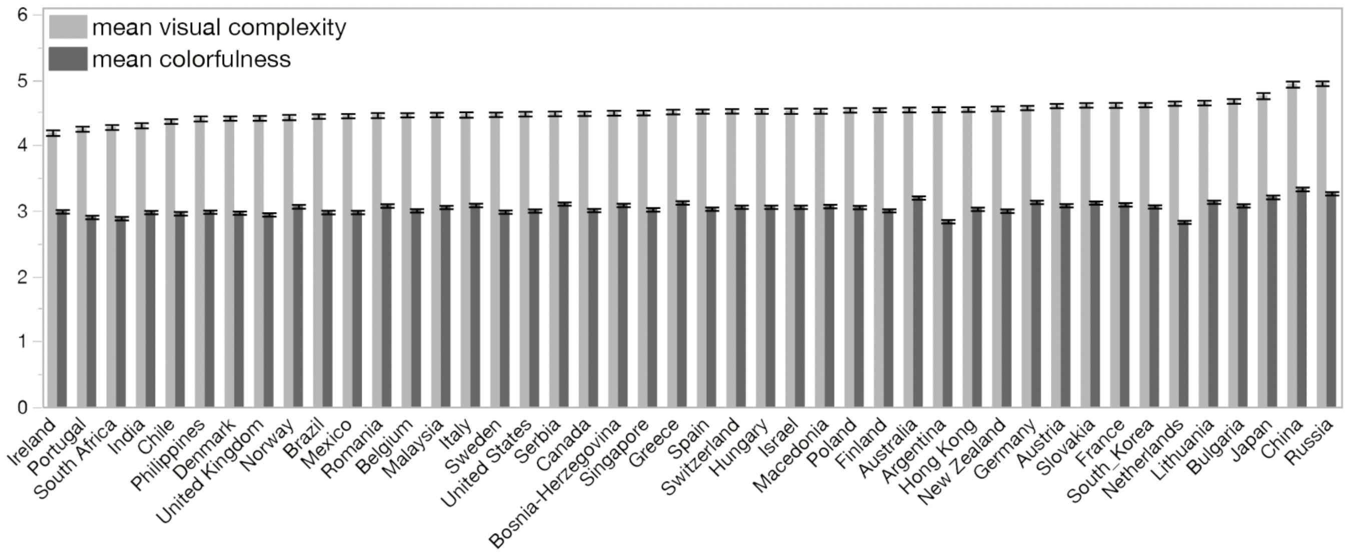

Figure 2. The mean scores of perceived visual complexity and perceived colorfulness (computed using the perceptual models from [51]) for the 2000 most popular websites in each of 44 countries, sorted by visual complexity. The scale for the models ranges from 0 (low) to 10 (high).

subset. We compared this subset (and its previously computed image metrics) against the local websites of each country using one-way ANOVAs and Tukey tests.

For question four, how local websites differ across countries, we resampled our dataset again with only local websites for each country and calculated one-way ANOVAs and Tukey tests on this resampled local dataset.

Results

Differences in Website Designs

Our first question asked how website designs differ across countries. The results from ANOVAs across the resampled dataset showed a significant variation of website designs between countries in the model scores for overall colorfulness ($F_{43,43956} = 834.55$, $p < .001$) and visual complexity ($F_{43,43956} = 931.26$, $p < .001$). Website designs also varied significantly across all image sub metrics, as shown in Table 1. The results suggest that when looking at all top 2000 websites per country, website designs show statistically measurable differences in their colorfulness, visual complexity, and other sub-metrics.

To find out which countries have significantly different website designs, we conducted follow-up analyses using Tukey's tests. The results revealed that 86% of all country pairs differed in average visual complexity and 87% for colorfulness. The fact that most countries, but not all, differed in their website complexity and colorfulness can also be seen in Figure 2.

Figure 2 also shows that the variation between website colorfulness and complexity between countries was less than 1 point on a 10-point scale. Irish websites had the lowest average visual complexity score (m = 4.12, sd = 0.17) and Russian websites had the highest (m = 4.87, sd = 0.14, $t_{1998} = -107.25$, $p < .0001$, $d = 4.80$). The large effect sizes and high number of significantly differing pairs was due to the reduced variance that resampling the dataset provided. For example, the previous comparison between the visual complexity of Irish and Russian websites on the original (not resampled) dataset had an effect size of $d = 0.47$. The difference was small enough that it was barely visually noticeable when comparing the websites around the mean scores from both countries using Juxtapose.

For colorfulness, the lowest average score was for websites in the Netherlands (m = 2.78, sd = 0.10) and the highest was for China (m = 3.28, sd = 0.11, $t_{1998} = 105.49$, $p < .0001$, $d = 4.62$). Again, the difference was small and barely visually noticeable when comparing the screenshots.

Part of the reason why these differences were not visually noticeable is that the colorfulness and complexity models factor in a variety of sub-metrics. Any differences between these sub-metrics can be balanced out by human perception. For example, when looking at a group of websites at once, the models may provide similar scores for websites that are all complex or colorful for different reasons, muddying the visual comparison due to differing scores on sub metrics within each screenshot. To find out which metrics provide the most visually noticeable differences between the website designs of different countries, we therefore turned to our sub-metrics. The sub-metric most visually noticeable in the context of visual complexity was number of text areas (see Figure 3), which measures the overall occurrence of text across an image [51]. The number of text areas positively impacts complexity scores, meaning the higher the average text area score, the more complex the website layout.

The number of text areas differed most between websites in Chile (m = 64306.12, sd = 4561.85) and China (m = 91869.59, sd = 6572.84, $t_{1998} = -108.94$, $p < .0001$, $d = 4.87$). The

Table 1. Overview of the results of ANOVAs. Column "all websites" reports on differences between the top 2,000 websites between countries; column "global" reports on differences between 3,551 websites that have global reach. The sub-metrics contribute to perceived visual complexity and colorfulness as found in the perceptual models in [51].

Image Metric	Explanation	all websites ($p < .0001$)	global websites ($p > .05$, n.s.)
Visual Complexity			
Perceived complexity score	predicted perceived visual complexity based on [51]	$F_{43,43956} = 895.72$	$F_{43,3462} = 0.19$
Number of leaves	the number of leaves of a space-based decomposition	$F_{43,43956} = 464.61$	$F_{43,3507} = 0.24$
Number of text areas	the number of leaves classified as text	$F_{43,43956} = 1202.48$	$F_{43,3507} = 0.31$
Number of non-text areas	the number of leaves classified as non-text	$F_{43,43956} = 420.86$	$F_{43,3507} = 0.10$
Number of text groups	the number of horizontal groups of text characters	$F_{43,43956} = 898.41$	$F_{43,3507} = 0.25$
Number of image areas	the number of image areas (adjacent images count as one)	$F_{43,43956} = 1705.94$	$F_{43,3507} = 0.35$
Colorfulness [45]	average saturation of pixels as chroma divided by lightness	$F_{43,43956} = 277.15$	$F_{43,3507} = 0.63$
Hue	the average pixel value for hue in the HSV color space	$F_{43,43956} = 516.653$	$F_{43,3507} = 0.17$
Colorfulness			
Perceived colorfulness score	predicted perceived colorfulness based on [51]	$F_{43,43956} = 763.75$	$F_{43,3462} = 0.40$
Saturation	the average pixel value for saturation	$F_{43,43956} = 509.47$	$F_{43,3507} = 0.30$
Colorfulness [21]	measures the average color difference of pixels against grey	$F_{43,43956} = 174.45$	$F_{43,3507} = 0.59$
Number of image areas	the number of separate images	$F_{43,43956} = 1705.94$	$F_{43,3507} = 0.35$
Number of quadtree leaves	the number of leaves of a quadtree decomposition using minimum color entropy as a criterion for division	$F_{43,43956} = 1569.79$	$F_{43,3507} = 0.72$
Number of text areas	the number of leaves classified as text	$F_{43,43956} = 1202.48$	$F_{43,3507} = 0.31$
Number of non-text areas	the number of leaves classified as non-text based	$F_{43,43956} = 420.86$	$F_{43,3507} = 0.10$

statistical difference corresponds to a noticeable variation between the amount of text in Chilean and Chinese websites, as shown in Figure 3.

The most visually noticeable difference between sub metrics that contribute to the perceptual colorfulness model was saturation, which measures the average pixel value of saturation in an image [51], and is positively correlated with the colorfulness score. The countries whose websites differed most for saturation were South Korea (m = 41.74, sd = 3.82) and Brazil (m = 57.12, sd = 4.34, $t_{1998} = 84.09$, $p < .0001$, $d = 3.76$). Figure 4 compares South Korea and Brazil websites, showing that the average saturation of a site noticeably differs between the Brazilian websites (right) compared to the South Korean websites (left). Similar to China, South Korean websites also featured many more text areas (though the average number was slightly lower than that of Chinese websites) and more white space. Brazilian websites, in contrast, showed larger image areas and more saturated colors, albeit less so than the Chilean websites shown in Figure 3.

Our tool Juxtapose lists the average scores for all image metrics and for all countries in our dataset, enabling designers to look up specific differences between countries that they may choose to design for.

Differences Between Global Website Designs

Our second research question asked whether globally popular websites differ across countries. Using the criteria for global players described in the analysis section, we selected a subset of 3,551 websites to analyze.

On average, these global websites made up 4.45% (sd = 0.26%) of the top websites per country, with Japan having the highest percentage (5.26%) and Russia having the lowest (3.52%). The overall low fraction of global websites among popular websites in each country could suggest that global players do not dominate the web in any of our 44 countries; however, many of the global companies whose websites made the top 2000 websites per country were near the top of these rankings. Global websites made up an average of 40.51% (sd = 9.02%) of the top 100 websites in each country. The Netherlands had the highest percentage, with 55.56% of its top 100 websites classified as global players, while Russia had the lowest with 15%. Given that the top 100 websites per country constitute the most popular websites with regards to their traffic, these higher percentages show that a decent portion of website traffic is occupied by global players in most of the countries analyzed.

The website designs of these global players did not significantly differ in their visual complexity ($F_{43,3462} = 0.19$, $p = 1.00$) or colorfulness ($F_{43,3462} = 0.40$, $p = 1.00$), nor were there any significant differences between the websites for any of the sub-metrics (see Table 1). These results suggest that although there are examples of international companies that localize their websites to specific countries, such as Adobe or SAS [27], the majority of global players do not adapt their website designs to other countries. For example, we did not observe any design changes between countries for the websites of Google and YouTube. Combined with our finding that these global websites predominantly ranked among the top 100 websites in each country, our result suggests that most users are exposed to the relatively homogeneous website designs of global companies on a regular basis.

Countries that differ from Global Design

Our third question asked how countries' local websites differ from the average global website design. If we found countries with websites that differed from the overall global websites, this would suggest that some countries buck a global design trend. We selected all global websites and combined them into a single subset. The country samples were reduced to only include locally popular websites and compared to the subset which represents a global website design average. This analysis was made on the original, not resampled, dataset in order to make accurate comparisons to percentage statistics.

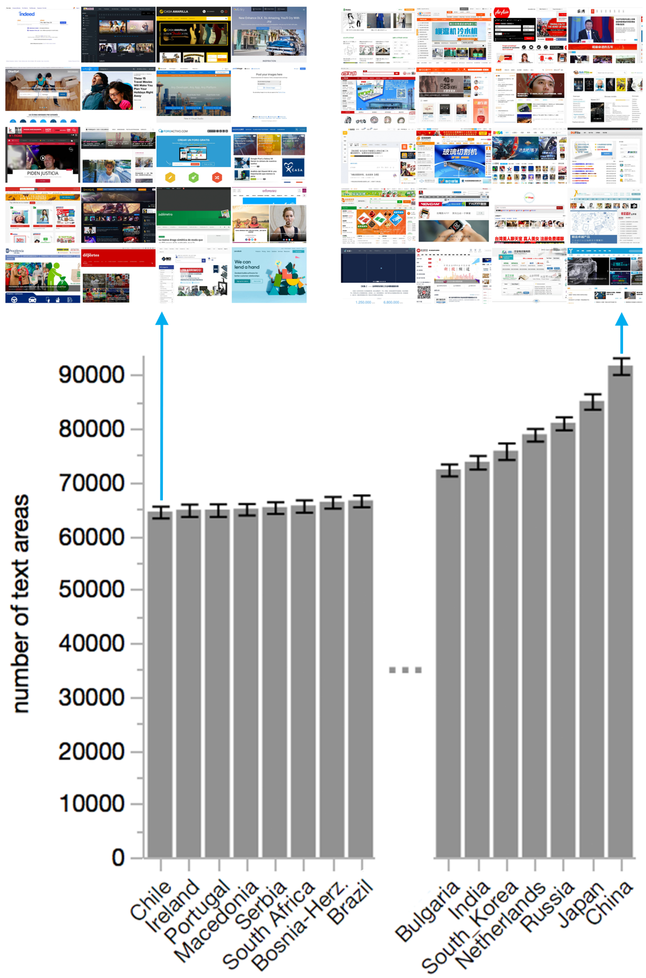

Figure 3. Comparison of a random sample of websites from Chile and China. Websites shown have roughly the same number of text areas as the average website from Chile (left, m = 64306.12, se = 1104.03) and China (right, m = 91869.59, se = 1571.99). Error bars show the standard errors.

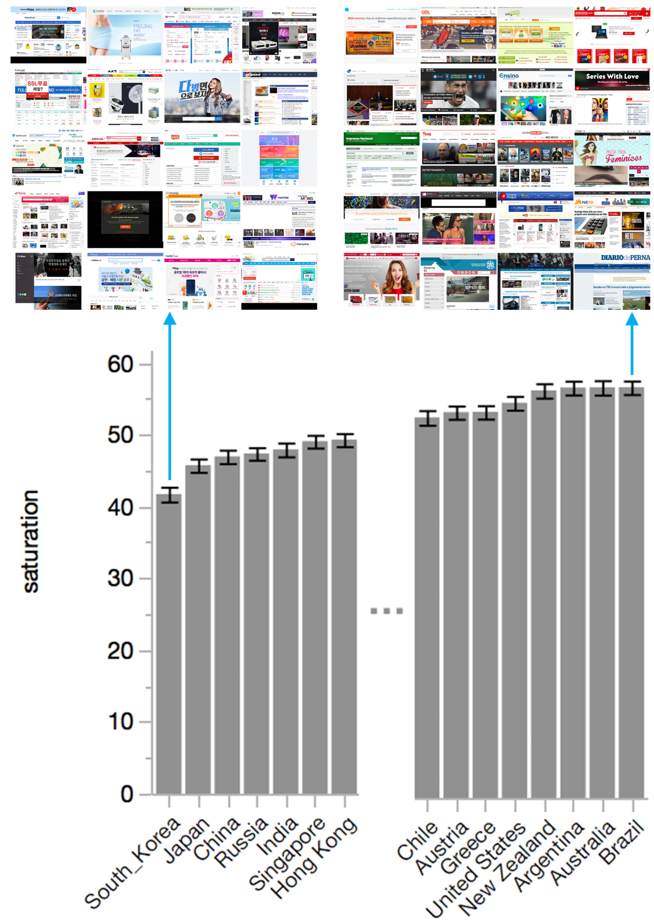

Figure 4. Comparison of a random sample of websites from South Korea and Brazil. Websites shown have roughly the same saturation value as the average website from South Korea (left, m = 41.74, se = 1.04) and Brazil (right, m = 57.12, se = 1.05). Error bars show the standard errors.

Russia (m = 4.90, sd = 1.50) was the top country whose websites differed from the global subset (m = 4.21, sd = 1.45) in visual complexity ($t_{1945} = -4.20$, $p < .01$, $d = 0.46$). The second top differing country in website visual complexity was China (m = 4.89, sd = 1.75, $t_{1540} = 4.89$, $p < .01$, $d = 0.39$). In average number of text areas, Chinese websites (m = 92924.89, sd = 64576.13) also differed significantly from the global subset (m = 66184.76, sd = 45642.03, $t_{1731} = 5.09$, $p < 0.001$, $d = 0.42$), although no other countries did. Russia and China also scored among the countries with the lowest number of global players in their top 100 most popular websites. Russia was first with 15% of global players in the top 100 websites, and China second with 20%.

Differences Between Local Website Designs

Our final question asked how local websites differ across countries. To look at only local websites, we excluded all global player websites when resampling our database.

Local websites differed significantly in visual complexity ($F_{43,43956} = 957.02$, $p < .0001$) and colorfulness ($F_{43,43956} = 833.63$, $p < .0001$). The top differing country pair for visual complexity was Ireland (m = 4.12, sd = 0.17) and Russia (m = 4.90, sd = 0.15, $t_{1998} = -110.04$, $p < .0001$, $d = 4.92$). For colorfulness, the top differing pair was the Netherlands (m = 2.79, sd = 0.11) and China (m = 3.30, sd = 0.11, $t_{1998} = 108.59$, $p < .0001$, $d = 4.86$). The differences were again relatively small (less than 1 point on a 10-point scale for both colorfulness and visual complexity). However, while the overall perceived colorfulness and visual complexity were similar, local websites in different countries strongly vary between their number of text areas, images areas, and saturation.

Local websites also differed significantly in saturation ($F_{43,43956} = 582.72$, $p < .0001$) and average number of text areas ($F_{43,43956} = 1273.10$, $p < .0001$). The countries whose websites differed most for saturation were South Korea (m = 40.68, sd = 3.62) and Argentina (m = 57.41, sd = 4.36, $t_{1998} = 93.39$, $p < .0001$, $d = 4.18$). For saturation, the countries whose websites differed most were Chile (m = 64488.22, sd = 4422.47) and China (m = 92805.68, sd = 6662.27, $t_{1998} = -111.98$, $p < .0001$, $d = 5.01$).

The similar results to the overall website comparison were expected due to the relatively small global subset excluded (only an average of 4.45% of each country's websites). The results strengthen our previous findings that differences exist between countries' website design and show that they are mainly driven by differences between the designs of local websites. In contrast, global websites are a homogenized, but popular, subset within each country's diverse set of website designs.

SUMMARY AND DISCUSSION

The aim of this paper was to provide a snapshot of current website designs comparing 44 countries. In particular, we investigated whether designers should localize their websites, and if yes, which design elements they should adapt to local preferences. In the following, we discuss our main findings:

Design localization is needed between some countries: Our results indicate that design localization is needed between *some* countries, but that there are a number of countries with relatively similar website designs. This also suggests that the decision to perform design localization needs to be made on a case-by-case basis; if a Chilean designer aims to localize for the Irish market, they might not have to change the number of text areas given that both of these countries were found to have low number of text areas. However, if the same designer localizes for the South Korean, Japanese, or Chinese markets, for example, the differences are much more profound and addressing them will enable the localized designs to blend in with the local design trends. According to Reinecke et al. [51], such adaptations can result in a significant increase in visual appeal, making the localized website more likely to be perceived as trustworthy and usable by local users.

Interestingly, we found that the websites from culturally similar countries (e.g., anglo-saxxon, European, or East Asian countries) are often similar in their visual complexity, colorfulness, number of text areas, and saturation. For example, Japan, China, and South Korea usually ranked close together on various image metrics, such as the number of text areas (high), the average saturation of colors (low), or the visual complexity (high). Websites from the United States, New Zealand, and Australia were among those with the highest saturation (along with Brazil and Argentina) and scored among the middle of all countries for visual complexity and number of text areas. This is similar to the findings in Reinecke and Gajos [50] who showed that neighboring countries often share similar visual preferences for certain website designs. In addition, our results show that visual preferences often go hand in hand with local website designs. For example, according to [50], the average Chilean prefers a visual complexity of 4.56, which closely matches the average visual complexity of the top 2000 websites in Chile (4.55). However, in other cases, our results show a mismatch between people's website design preferences and the website designs that they are exposed to. For example, Russians were found to prefer websites with a very low visual complexity [50]. In contrast, our findings show that Russian websites are those with the highest visual complexity in our dataset. It is possible that Russians actually prefer simple websites but that this is not reflected in local design trends. If this is the case, users might not be provided with websites that they find most appealing, most trustworthy, and usable [35], and website owners might forgo the chance to optimize the success of their site. More work is needed to further shed light on these differences between a country's average visual appeal and its average website design.

The number of text areas and the average saturation of colors reveal clear visual differences between countries: While there are statistically significant differences between the website designs of all 44 countries that we analyzed in numerous design metrics, we especially found visually discernible differences between the number of text areas and the average saturation of colors. These variations were most noticeable between China and Chile (with China's websites exhibiting a significantly higher number of text areas than Chile) and between Brazil and South Korea (with Brazilian websites having a significantly higher average saturation of colors than South Korean websites).

Variations between websites' overall colorfulness and visual complexity (as computed by perceptual models first presented in [51]) across countries were relatively small, with the largest difference being less than one point on a 10-point scale. While previous work showed that even a 0.5-point difference between a website's perceived colorfulness and visual complexity can have a large positive or negative impact on a person's first impression of a website's visual appeal [51, 50], we were nevertheless surprised to find relatively small differences between website designs across many countries. This result suggests that there are only few countries that buck an international design trend. The finding is in contrast to those of prior work (e.g. [41, 19]), which has repeatedly found larger differences between website designs in different countries. There are two possible explanations for the divergences in results: First, while prior work has manually annotated websites to compare their designs, we used computational image metrics, which might be more robust to human bias, but less sensitive to picking up on design differences. Second, website designs might be homogenizing. This could be due to the increasing presence of templates and frameworks, such as Twitter's bootstrap, or because designers might increasingly draw inspiration from globally popular designs. For example, in line with theories on the priming effects of frequent exposure to website designs [20, 8, 7, 61], the presence of common website designs might decrease design diversity over time. In general, prior work has found that website design undergoes trends [7, 18]. We therefore assume that website designs follow these trends independent of country borders, with the exception of a few countries in which the websites remain fairly isolated in their own design.

The designs of global websites are relatively similar across countries: We also found that most users are exposed to the relatively homogeneous website designs of global companies, which often rank among the top 100 websites in each country. These global players dominate the popularity ranking in most countries, making up an average of 40.51% of the top 100 websites in each country. Their designs do not significantly differ when looking at any of our image metrics, such as visual complexity, colorfulness, number of text areas, or saturation. This indicates a homogenized, globally popular, website design. Comparing example screenshots, we found that many of these globally popular websites, such as Google, YouTube, or Reddit, do not localize their design. Their global reach and popularity, while evidently ignoring country-specific aesthetic preferences, suggests that potential benefits of design localization might be outperformed by brand recognition and network effects.

The websites in countries with the lowest percentage of global players differ most from global design trends: When comparing the design metrics of all countries with the subset of global players, we found that only websites in China and Russia were significantly different from global designs in their visual complexity—the two countries that also had the lowest percentage of global players (around 3.52% of the top 2000 websites for Russia and 4.39% for China). Both of these countries are known for internet censorship [60, 11]; additionally, China has increasingly invested in efforts to retain local customs and arts, such as by supporting local artists and companies. Our findings indicate that decreasing the number of global website designs—such as through censorship, reducing the number of foreign business, or supporting local businesses—increases the diversity of local designs. In these environments, any priming effects from frequent exposure to website designs or designers following specific design trends might strengthen local designs.

Limitations and Future Work

Our study has several limitations: First, we were unable to obtain robust website category data. While we ensured that all countries had similar percentages of specific website categories (by manually checking the dataset), there is a chance that some of the detected differences are due to diverging fractions of categories (e.g., one country having a higher number of news paper sites than entertainment sites) rather than due to actual design differences. In addition, we cannot make predictions on the average website design in a given category and how they might differ across countries. Second, we defined local websites as those that do not have a global reach, thereby excluding global websites even if they might be considered as local designs in a specific country (e.g., Google in the US). In some countries, this can slightly change the average website design scores that we report on. Third, we did not include additional characteristics, such as the origin of a website, the default language, or the market the website aims at, which could further help to investigate local design preferences.

Our results also laid the foundation for much future work. Because we focused on comparing website designs in various countries, our study did not enable us to additionally evaluate whether these website designs do in fact correspond to people's visual preferences. We hope to use our website dataset to investigate whether these websites align with people's visual preferences, and to find out about potential reasons for why they might not.

In addition, future work could extend existing attempts to understand how website design changes over time [7, 31, 18] by conducting longitudinal analyses of design diversity for the most popular websites around the world. Monitoring how diversity is affected by the (presumably increasing numbers of) designs of global players could indicate whether design localization is a promising investment, or if people align more and more with a global design taste. Our snapshot of current website designs lays the foundation for such analyses.

Based on our dataset, it is also possible to investigate country-specific design elements and characteristics beyond the image metrics that we used for our analysis, such as the content of images, or the location of specific design elements. This could result in more fine-grained design guidelines and add to our results as well as to prior research (e.g., [1, 41]) by providing designers with guidelines and templates for the creation of localized website designs.

CONCLUSION

We contributed a large-scale analysis of current website designs in 44 countries. Our results demonstrate that website designs significantly differ between countries, and that these differences are predominantly driven by variations in the designs of local websites. In contrast, global websites that are popular in a number of different countries often do not localize their designs.

Our work allows designers to determine whether to adapt their website to another country and how to do that. In order to support these decisions, we provided findings that suggest when it is worthwhile adapting (e.g., to appear more "local" in a given country), a dataset that includes tangible image metrics that can serve as design guidelines, and a tool that allows visually comparing websites from our 44 countries. These guidelines, orchestrated with other important website localization aspects, like translation, content, image composition, color interpretation, and symbols, will point designers into the direction of successful solutions.

Our contributions will help designers engage with target audiences across the world, and adapt in an ever changing online ecosystem of local and global aesthetics.

DATASET AND TOOL

We make available our dataset and tool at `http://juxtapose.labinthewild.org`.

ACKNOWLEDGEMENTS

This project was funded by NSF award 1651487.

REFERENCES

1. Tanveer Ahmed, Haralambos Mouratidis, and David Preston. 2008. Website design and localisation: A comparison of Malaysia and Britain. *International Journal of Cyber Society and Education* 1, 1 (2008), 3–16.

2. Ahamed Altaboli and Yingzi Lin. 2011. *Objective and Subjective Measures of Visual Aesthetics of Website Interface Design: The Two Sides of the Coin*. Springer

Berlin Heidelberg, 35–44. https://doi.org/10.1007/978-3-642-21602-2_4

3. Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)* (1995), 289–300.
4. Oliver Bock. 2017. The Essential Guide To Website Localization. (2017). https://phraseapp.com/blog/posts/the-essential-guide-to-website-localization/.
5. Jens Bölte, Thomas M. Hösker, Gerrit Hirschfeld, and Meinald T. Thielsch. 2017. Electrophysiological correlates of aesthetic processing of webpages: a comparison of experts and laypersons. *PeerJ* 5 (June 2017). https://doi.org/10.7717/peerj.3440
6. Ewa Callahan. 2005. Cultural similarities and differences in the design of university web sites. *Journal of Computer-Mediated Communication* 11, 1 (2005), 239–273.
7. Wen Chen, David J. Crandall, and Norman Makoto Su. 2017. Understanding the Aesthetic Evolution of Websites: Towards a Notion of Design Periods. In *Proc. CHI'17*. ACM, New York, NY, USA, 5976–5987. http://doi.acm.org/10.1145/3025453.3025607
8. Aline Chevalier, Anne-Claire Maury, and Nicolas Fouquereau. 2014. The influence of the search complexity and the familiarity with the website on the subjective appraisal of aesthetics, mental effort and usability. *Behaviour & Information Technology* 33, 2 (2014), 117–132.
9. Nielsen Company. 2010. Nielsen Provides Topline U.S. Web Data for March 2010. (2010). https://tinyurl.com/ya5rgm7o.
10. Jack Cook and Mike Finlayson. 2005. The impact of cultural diversity on web site design. *SAM Advanced Management Journal* (2005), 15–23.
11. Charles Crabtree, Christopher Fariss, and Holger Kern. 2015. Truth Replaced by Silence: Private Censorship in Russia. In *NYU CESS 8th Annual Experimental Political Science Conference*.
12. Dianne Cyr. 2013. Website design, trust and culture: An eight country investigation. *Electronic Commerce Research and Applications* 12, 6 (2013), 373–385.
13. Dianne Cyr and Haizley Trevor-Smith. 2004. Localization of Web design: An empirical comparison of German, Japanese, and United States Web site characteristics. *Journal of the Association for Information Science and Technology* 55, 13 (2004), 1199–1208.
14. L.C. de Castro Salgado, C.F. Leitão, and C.S. de Souza. 2012. *A Journey Through Cultures: Metaphors for Guiding the Design of Cross-Cultural Interactive Systems*. Springer London.
15. DMOZ. 2017. Directory Mozilla. (2017). http://dmoztools.net/.
16. Andrea Everard and Dennis F. Galletta. 2005. How Presentation Flaws Affect Perceived Site Quality, Trust, and Intention to Purchase from an Online Store. *Journal of Management Information Systems* 22, 3 (2005), 56–95.
17. Boudhayan Ganguly, Satya Bhusan Dash, Dianne Cyr, and Milena Head. 2010. The effects of website design on purchase intention in online shopping: the mediating role of trust and the moderating role of culture. *International Journal of Electronic Business* 8, 4-5 (2010), 302–330.
18. Gili Korman Golander, Noam Tractinsky, and Ilanit Kabessa-Cohen. 2012. "Trends in Website Design,". *AIS Transactions on Human-Computer Interaction* (2012), 169âĂŞ189. http://aisel.aisnet.org/thci/vol4/iss3/1
19. Nitesh Goyal, William Miner, and Nikhil Nawathe. Cultural Differences Across Governmental Website Design. In *Proc. ICIC'12*.
20. Jan Hartmann, Alistair Sutcliffe, and Antonella De Angeli. 2008. Towards a Theory of User Judgment of Aesthetics and User Interface Quality. *ACM Trans. Comput.-Hum. Interact.* 15, 4, Article 15 (Dec. 2008), 30 pages. DOI:http://dx.doi.org/10.1145/1460355.1460357
21. David Hasler and Sabine E. Suesstrunk. 2003. Measuring Colourfulness in Natural Images. 5007 (06 2003), 87–95.
22. Marc Hassenzahl. 2004. The interplay of beauty, goodness, and usability in interactive products. *Human-computer interaction* 19, 4 (2004), 319–349.
23. Ariya Hidayat. 2017. PhantomJS headless WebKit browser simulation. (2017). http://phantomjs.org.
24. Zhang Hua, Maria Taslim, and Matt Keating. 2014. Same but Not the Same: How Much Do We Look into It for Localization Design?. In *Proceedings of HCI Korea (HCIK '15)*. Hanbit Media, Inc., South Korea, 83–87. http://dl.acm.org/citation.cfm?id=2729485.2729499
25. Alexa Internet Inc. 2017a. The Alexa browser extension. (2017). https://www.alexa.com/toolbar.
26. Alexa Internet Inc. 2017b. The Alexa Top Sites Service. (2017). https://www.alexa.com/topsites.
27. Konstantin Dranch James Austin. 2016. Only 11 of 80 companies localize web content. (2016). https://tinyurl.com/y7u6r5hj.
28. Jeff Johnson. 2014. *Designing with the Mind in Mind, Second Edition: Simple Guide to Understanding User Interface Design Guidelines* (2nd ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
29. Laurent Jouanneau. 2017. SlimerJS scriptable browser. (2017). https://slimerjs.org.
30. Nahye Kim, Bonseung Koo, Jaehwa Yoon, and Kwangsu Cho. Understanding the Formation of User's First Impression on an Interface Design from a Neurophysiological Perspective - EEG Pilot Study. In *Proceedings of HCI Korea*.

31. Ranjitha Kumar, Arvind Satyanarayan, Cesar Torres, Maxine Lim, Salman Ahmad, Scott R. Klemmer, and Jerry O. Talton. 2013. Webzeitgeist: Design Mining the Web. In *Proc. CHI'13*. ACM, New York, NY, USA, 3083–3092. http://doi.acm.org/10.1145/2470654.2466420

32. Daniel Lakens. 2013. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology* 4 (2013), 863. DOI: http://dx.doi.org/10.3389/fpsyg.2013.00863

33. Talia Lavie and Noam Tractinsky. 2004. Assessing dimensions of perceived visual aesthetics of web sites. *International journal of human-computer studies* 60, 3 (2004), 269–298.

34. Dave Lawrence and Soheyla Tavakol. 2006. *Balanced website design: optimising aesthetics, usability and purpose*. Springer Science & Business Media.

35. Gitte Lindgaard, Cathy Dudek, Devjani Sen, Livia Sumegi, and Patrick Noonan. 2011. An Exploration of Relations Between Visual Appeal, Trustworthiness and Perceived Usability of Homepages. *ACM Trans. Comput.-Hum. Interact.* 18, 1, Article 1 (May 2011), 30 pages. DOI:http://dx.doi.org/10.1145/1959022.1959023

36. Gitte Lindgaard, Gary Fernandes, Cathy Dudek, and J. Brown. 2006. Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour & Information Technology* 25, 2 (2006), 115–126. DOI: http://dx.doi.org/10.1080/01449290500330448

37. Whois API LLC. 2017. WhoIsXmlAPI. (2017). https://www.whoisxmlapi.com/.

38. Web Shrinker LLC. 2017. WebShrinker content categorization. (2017). https://www.webshrinker.com/content-categorization-api-network-providers/.

39. Netcraft Ltd. 2017. Netcraft Internet Security and Data Mining. (2017). https://www.netcraft.com/.

40. SimilarWeb LTD. 2017. SimilarWeb. (2017). https://www.similarweb.com/.

41. Aaron Marcus and Emilie West Gould. 2000. Crosscurrents: Cultural Dimensions and Global Web User-interface Design. *interactions* 7, 4 (July 2000), 32–46. DOI:http://dx.doi.org/10.1145/345190.345238

42. R. Mardiko. 2017. Collection of image metric computers. (2017). https://bitbucket.org/rmardiko/vizweb.git.

43. Aliaksei Miniukovich and Antonella De Angeli. 2014. Visual impressions of mobile app interfaces. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*. ACM, 31–40.

44. Morten Moshagen and Meinald T. Thielsch. 2010. Facets of visual aesthetics. *International Journal of Human-Computer Studies* 68, 10 (2010), 689 – 709. DOI: http://dx.doi.org/10.1016/j.ijhcs.2010.05.006

45. S N Yendrikhovskij, F J J Blommaert, and Huib Ridder. 1998. Optimizing color reproduction of natural images. (01 1998).

46. Unbabel News. 2014. Why you should localize your website. (2014). https://tinyurl.com/ycyv8pp9.

47. Kiemute Oyibo, Yusuf Sahabi Ali, and Julita Vassileva. 2016. An Empirical Analysis of the Perception of Mobile Website Interfaces and the Influence of Culture.. In *PPT@ PERSUASIVE*. 44–56.

48. Pei-Luen Patrick Rau and Sheau-Farn Max Liang. 2003. Internationalization and localization: evaluating and testing a website for Asian users. *Ergonomics* 46, 1-3 (2003), 255–270.

49. Katharina Reinecke and Abraham Bernstein. 2011. Improving Performance, Perceived Usability, and Aesthetics with Culturally Adaptive User Interfaces. *ACM Trans. Comput.-Hum. Interact.* 18, 2, Article 8 (July 2011), 29 pages. DOI: http://dx.doi.org/10.1145/1970378.1970382

50. Katharina Reinecke and Krzysztof Z. Gajos. 2014. Quantifying Visual Preferences Around the World. In *Proc. CHI'14*. 11–20. http://doi.acm.org/10.1145/2556288.2557052

51. 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 *Proc. CHI '13*. 2049–2058.

52. David Robins and Jason Holmes. 2008. Aesthetics and Credibility in Web Site Design. *Inf. Process. Manage.* 44, 1 (Jan. 2008), 386–399. DOI: http://dx.doi.org/10.1016/j.ipm.2007.02.003

53. Julia Rozwens. 2014. How to conduct website localization. (2014). https://www.smashingmagazine.com/2014/12/how-to-conduct-website-localization/.

54. Mirjam Seckler and Alexandre N. Tuch. 2012. Linking Objective Web-design Factors to Facets of Subjective Aesthetic Perception. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design (NordiCHI '12)*. ACM, New York, NY, USA, 809–810. http://doi.acm.org/10.1145/2399016.2399159

55. Hélène Stengers, Olga De Troyer, Martine Baetens, Frank Boers, and A Mushtaha. 2004. Localization of Web Sites: Is there still a need for it. In *international workshop on web engineering (held in conjunction with the ACM HyperText 2004 conference), Santa Cruz, USA*.

56. Meinald T. Thielsch, Iris Blotenberg, and Rafael Jaron. 2014. User Evaluation of Websites: From First Impression to Recommendation. *Interacting with Computers* 26, 1 (2014), 89–102. +http://dx.doi.org/10.1093/iwc/iwt033

57. Noam Tractinsky, Adi S Katz, and Dror Ikar. 2000. What is beautiful is usable. *Interacting with computers* 13, 2 (2000), 127–145.

58. Alexandre N. Tuch, Sandra P. Roth, Kasper Hornbäk, Klaus Opwis, and Javier A. Bargas-Avila. 2012. Is beautiful really usable? Toward understanding the relation between usability, aesthetics, and affect in HCI. *Computers in Human Behavior* 28, 5 (2012), 1596 – 1607. http://www.sciencedirect.com/science/article/pii/S0747563212000908

59. John W Tukey. 1949. Comparing individual means in the analysis of variance. *Biometrics* (1949), 99–114.

60. Xueyang Xu, Z. Morley Mao, and J. Alex Halderman. 2011. *Internet Censorship in China: Where Does the Filtering Occur?* Springer Berlin Heidelberg, Berlin, Heidelberg, 133–142. DOI: http://dx.doi.org/10.1007/978-3-642-19260-9_14

61. Robert B. Zajonc. 1968. Attitudinal effects of mere exposure. 1. *Journal of Personality and Social Psychology Monograph Supplement* (1968), 1–27.

62. Xianjun Sam Zheng, Ishani Chakraborty, James Jeng-Weei Lin, and Robert Rauschenberger. 2009. Correlating Low-level Image Statistics with Users - Rapid Aesthetic and Affective Judgments of Web Pages. In *Proc. CHI'09 (CHI '09)*. ACM, New York, NY, USA, 1–10. http://doi.acm.org/10.1145/1518701.1518703