Interactive Exploration of Large-scale UI Datasets with Design Maps

Luis A. Leiva[1,*], Asutosh Hota[2], Antti Oulasvirta[3]

[1] *University of Luxembourg, Luxembourg*
[2] *Aalto University, Finland*
[*]*Corresponding author:* `name.surname@uni.lu`

Designers are increasingly using online resources for inspiration. How to best support design exploration without compromising creativity? We introduce and study Design Maps, a class of point-cloud visualizations that makes large UI datasets explorable. Design Maps are computed using dimensionality reduction and clustering techniques, which we analyze thoroughly in this paper. We present concepts for integrating Design Maps into design tools, including interactive visualization, local neighborhood exploration, and functionality to integrate existing solutions to the design at hand. These concepts were implemented in a wireframing tool for mobile apps, which was evaluated with actual designers performing realistic tasks. Overall, designers find Design Maps supporting their creativity (avg. CSI score of 74/100) and indicate that the maps producing consistent whitespacing within cloud points are the most informative ones.

Research Highlights

- **Design Maps are interactive point-cloud visualizations of large-scale UI datasets.**
- **Designers prefer consistent whitespace within cloud points.**
- **Designers find Design Maps supporting their creativity and perspective-taking.**

1. INTRODUCTION

The collection and curation of *design examples*, or sources serving as a starting point to influence a final design (Eckert and Stacey, 2000), is an important aspect of creative design practice and ideation (Herring et al., 2009; Stolterman, 2008). Designers traditionally seek inspiring design examples from external sources, such as design magazines, books, websites, as well as from colleagues and other peers' work (Gonçalves et al., 2014; Vasconcelos and Crilly, 2016; Wallace et al., 2020). They take breaks and change environments to incubate new ideas and obtain distance from solutions at hand. Recently, designers have been reported to turn to online resources for inspiration (Koch et al., 2018). They use search engines and online social design platforms, such as Behance, Pinterest, and Dribble. These offer example galleries, curated metadata and discussion on examples, as well as search functionality. Paradoxically, even if there is an inexhaustible number of examples available in these sources, it is nonetheless challenging to find examples that are relevant to the task at hand (Koch et al., 2018). In particular, finding relevant user interface (UI) examples, especially in large-scale datasets, is a highly challenging task because UIs have aesthetic and functional properties that are only indirectly reflected by their corresponding pixels and associated metadata.

We believe that the available methods support design exploration poorly, because they have been designed for other purposes. For example, the prevailing search

method is the familiar keyword-based search. However, if a designer does not have a specific goal in mind, or cannot express the defining aspect verbally, this method fails (Sharmin et al., 2009). Image-based search is similarly limited in this regard. Overall, designers using query-based approaches may not get exposed to a broad set of alternatives (Herring et al., 2009). On the other hand, browsing galleries curated by people or generated algorithmically, may have low relevance to the task at hand. Furthermore, novice designers may struggle to find appropriate examples if they lack appropriate annotations. Even when successful, they may struggle to understand how the particularities of an example apply to their project (Kang et al., 2018). Generally, it has been shown that failing to find suitable materials may compromise the creativity and novelty of produced ideas (Perttula and Sipilä, 2007). This article contributes to a line of research to support inspiration-seeking and exploration in large UI design datasets. For practical reasons, as explained later, we will focus on mobile UI designs.

1.1. Background

We introduce and study *Design Maps*, a novel class of interactive visualizations of large-scale design datasets. A Design Map visualizes a design dataset as a colorful point cloud on a two-dimensional display. Design Maps are computed using machine learning methods that yield a low-dimensional representation of a higher dimensional design space. They use mathematically defined visuospatial encodings to convey structure, similar to what geographic maps do. The benefit is that they can efficiently summarize very large design datasets, in our case in the order of thousands or tens of thousands of design objects, being for example web pages, app screenshots, scene renderings, or layout plans. We should mention that not every design dataset can be easily summarized and in fact each one may require a particular treatment; see 'Discussion' section. Therefore, to narrow down the scope of this paper, we will focus on UI design datasets and, in particular, mobile UI designs. In 'Future work' section we discuss possible extensions to create Design Maps for other types of design datasets.

Notice that our approach is orthogonal to query-based methods. The main difference between Design Maps over previous concepts, such as interactive design galleries (Lee et al., 2010) or design querying (Kumar et al., 2013), is that Design Maps present *all* design objects (even in a very large dataset) for interactive exploration. A Design Map can be panned and zoomed, and individual design objects inspected and interacted on. In practical terms, Design Maps allow inspecting several designs in a matter of seconds, with glance and mouse/finger cursor. Figure 2

and Figure 3 provide examples of Design Maps, here shown for the purpose of exploring a large dataset of mobile apps which we use to exemplify the Design Maps concept and further discuss in Section 4.1.

Our work is motivated by a recent review of research on fixation in design (Crilly and Cardoso, 2017) that concluded that there is a need for knowledge on how and what kinds of stimuli to show and what types of representations to use, and how to best trade off nearby vs. distantly related stimuli. Broadly, the goal of Design Maps is to support inspiration-seeking processes in design, and in particular the exploration–exploitation problem therein. By 'inspiration' we refer to example designs that can influence the final design by serving as a starting point for the design, offer reusable partial ideas, or patterns (Eckert and Stacey, 2000). Explorative visualizations in general can help ensure that designers do not fixate on one design idea but explore the space before intensifying effort around a design candidate (Tohidi et al., 2006). However, it is not known how they could help design exploration.

Empirical research suggests multiple possible benefits of having design examples available. They can lower the barrier to entry for novices (Lee et al., 2010; Woodbury et al., 2017) and, when provided at the right time, can improve the quantity and quality of generated ideas (Siangliulue et al., 2015). Design Maps could inspire designers by serving as a starting point for the design and offer reusable partial ideas and patterns. Producing numerous design ideas is believed to increase the probability to producing creative solutions (Perttula and Sipilä, 2007). As previously mentioned, key aspects of a suitable Design Map visualization include: identify intrinsic dimensions, color codings, and other visual cues that allow the designer to understand the structure of the design space and efficiently navigate it.

Design Maps also expose similar examples and diverse examples. This can help when reaching an impasse in ideation (Bernard A. Nijstad and Baas, 2010). However, while design examples can help a designer 'on a roll' to explore the design space more, or get ideas 'when stuck', researchers have raised concerns with digital example galleries (Lee et al., 2010; Marks et al., 1997). Existing examples found in the web often lack additional information such as author, purpose, method, or approach used, which hampers understanding and evaluating of the quality and credibility of the example (Koch et al., 2018). Moreover, examples are not often related to the design at hand and their hit rate may be low overall. While exposing designers to ideas that are semantically far from their own can trigger novel combinations of ideas, it is suggested that far inspirational ideas can harm creativity if received during productive ideation (Chan et al., 2017). Furthermore, they can evolve as new data comes in or new

goals and constraints are learned. Moreover, irrelevant examples can distract designer and decrease the quality of produced ideas (Chan et al., 2017). Examples may also fixate the designer to a predefined idea of a 'good design', resulting in homogeneous design solutions. In addition, being exposed to too similar design examples can led to designer's block. These challenges call for research on approaches like Design Maps, which can offer an overview and structure to a dataset, means for quick diversification and intensification, as well as integration in design tools. Prior to this work, there has been no study of how different machine learning methods can be used to visualize design datasets in a way that support creativity. Further, it is not known which methods work and what their effect on designer's creativity may be.

1.2. Summary of Contributions

We present and compare machine learning methods to create Design Maps, and study them empirically in a UI wireframing application in the context of mobile apps design. The computational problem is particularly challenging. There is a high number of aspects differentiating any two designs, and prior to this work it has not been known which methods are suitable for inspiration-seeking in UI wireframing. The technical solutions proposed here build on known clustering and dimensionality reduction methods from machine learning, including classic methods like PCA as well as more recent ones like UMAP and tSNE, to be described later. These methods have been previously used in complex visualization problems for example in molecular science and biology. Off-the-shelf, these methods perform poorly with pixel input only, because color differences can be trivial and unrelated to design structure or contents. We address this problem by using semantic labels describing the *type* of each part of a design. In our case, each UI element is labeled according to their view hierarchy. The result is a visualization that breaks down the design dataset into clusters that can be navigated and positioned against the design at hand. Eventually, the goal in designing a suitable visualization is to identify intrinsic dimensions and color and other codings that allow the designer to understand the structure of the design space and efficiently navigate it. To this end, we report differences in each technique's capability to distinguish meaningful and recognizable aspects of a given dataset.

Informed by an interview study we conducted with UI and graphic designers, aimed at understanding design practice to inform the concept of Design Maps, we learned that a passive visualization of a design dataset is insufficient: the interactive visualizations enabled by the Design Maps concept should be integrated into design tools in such a way that designers can directly exploit the ideas in their project. This allows the designer to go back and forth between design ideas and their visual form. Consequently, we developed a wireframing tool for mobile app design to study how this integration could be made in practice. In this scenario, designers imagine and assess possible designs to learn about the problem at hand (Moggridge, 2006) and finally come up with a particular visual form (Tohidi et al., 2006), in this case a wireframe of the envisioned UI layout. Specifically, we present concepts for integrating them into design tools, including interactive support for (1) exploring large-scale visualization, (2) exploring local neighborhoods of a particular design, and (3) exploiting designs by incorporating them to an existing wireframe. Finally, we evaluated our wireframing tool with twelve professional designers carrying out realistic tasks. To sum up the results, we observed that PCA and tSNE based maps are the most informative visualizations, as they produced consistent whitespacing within cloud points. Overall, designers found Design Maps to be a valuable means for interactive exploration.

2. COMPUTATIONAL SUPPORT FOR INSPIRATION-SEEKING

The concept of Design Maps builds on methods from machine learning and data visualization, but also relies on previous research on creativity support tools and design mining.

2.1. Visualization of large datasets

Data visualizations provide perceptually intuitive summarizations that allow users to interactively explore and analyze complex datasets, identifying interesting patterns, inferring correlations and causalities, and supporting sense-making activities. (For reviews, we recommend Bikakis (2019) and Card and Mackinlay's seminal work about the information visualization landscape (Card and Mackinlay, 1997)).

A key problem data visualization methods have to address is the problem of *excess information* (Moacdieh and N., 2015). To be perceptually understandable, and fit on a display, data visualization methods should provide efficient and effective abstraction and summarization mechanisms. To this end, many papers report the use of approximation (i.e. data reduction) techniques, in which abstractions of data are computed (Bikakis, 2019). Considering the existing approaches, most of them are based on (1) sampling and filtering and/or (2) aggregation (e.g., binning, clustering). By sampling the data, a system can produce approximate answers fast enough for

exploratory visualization, at the cost of accuracy and trust (Moritz et al., 2017). By aggregating the data, interesting patterns can be identified. Design Maps follow this latter approach.

2.2. Knowledge discovery in multi-dimensional data

Previous work has applied machine learning to visualization for engineering design purposes, though it is mainly focused on design decision making rather than on enhancing creativity. For example, Richardson and Winer (2011) used self-organizing maps to break the final visualization up into three maps containing separate contextual information, so that researchers can quickly obtain information about the design space. Cloud Visualization (Eddy and Lewis, 2002) is a non-interactive representation of sets of points as clouds aimed at conveying engineering information in a multidimensional context. Cityplot (Knerr and Selva, 2016) presents a superposition of a dimensionally reduced representation of design decisions and bar plots representing the multiple criteria of the objective space. Overall, in trade space exploration of large design data sets, designers need to select a subset of data of interest and examine data from different data dimensions and within data clusters at different granularities (Zhang et al., 2012).

2.3. Collaborative ideation tools

The use of data visualizations for supporting divergent and convergent thinking has been studied earlier in the research literature. For example, Dynagrams (Eppler and Kernbach, 2016) are collaborative visualization tools in the form of schematic drawings to foster interactive thinking and deliberation. They help to produce a joint solution space taking all participants' contributions into account. The use of Dynagrams has shown that they work well as consolidation tools that help to get an overview to see the big picture, to focus discussions, and to make better informed decision that are also comprehensible for others. However, Dynagrams are limited in terms of the number of concepts that can be represented and the number of elements they can carry (Eppler and Kernbach, 2016). In contrast, Design Maps are highly scalable and highly performant, as confirmed by our integration into a wireframing tool (to be described later) that has been tested with a dataset comprising near hundred of thousands data points.

2.4. Interactive computational design

Design Maps build on creativity support tools and related areas that computationally *generate* design ideas related to the designer's draft of an initial idea, and show them e.g. in an example gallery next to the tool. For example, DesignScape (O'Donovan et al., 2015) aids the design process by making interactive UI layout suggestions using optimization methods. It uses two distinct but complementary types of suggestions: refinement suggestions, which improve the current UI layout, and brainstorming suggestions, which change the style. Rewire (Swearngin et al., 2018a) is an interactive system that helps UI designers leverage example screenshots. Rewire automatically infers a vector graphics representation of screenshots, so that each UI component becomes a separate object with editable shape and style properties. Scout (Swearngin et al., 2018b) is a mixed-initiative explorative tool for UI design variations under high-level constraints based on usability and visual design principles. Swire (Huang et al., 2019) enables UI designers to retrieve similar designs by means of sketching. Finally, GUIComp (Lee et al., 2020) is a design assistant with real-time multi-faceted feedback that allows a novice designer to browse UI design datasets and use the relevant examples they find to guide their designs. Design Maps borrow some of these ideas, to be discussed in a later section. In contrast to computational design where novel designs are generated by the computer, Design Maps only visualize already existing designs.

2.5. Design mining and exploratory tools

Design Maps also build on design mining; i.e., knowledge discovery and retrieval techniques to understand design demographics, automate design curation, and support data-driven design tools. Design Maps subscribe to underlying goals of design mining such as scalability (e.g. how to handle thousands of design examples) and extensibility (e.g. how to incorporate different visualization schemes). In this regard, we should mention Dream Lens (Matejka et al., 2018), which allows searching and filtering through a large collection of 3D model design variations, and Webzeitgeist (Kumar et al., 2013), a platform for large-scale design mining of websites that inspired a series of query-by-ui-example applications, including e.g. GUIfetch (Behrang et al., 2018) and Rico (Deka et al., 2017; Liu et al., 2018). Design Maps differs from design mining in its attempt to directly visualize, without querying, a design dataset.

Finally, a large body of research has been conducted on computational methods that support browsing high-dimensional data such as movies (Jorge et al., 2017; Martins et al., 2011) and music (Lübbers and Jarke, 2009; Pampalk et al., 2002; Vad et al., 2015). However, we are not aware of any alternative for exploring large datasets comprising visual material, and concretely UI designs.

Design Maps thus fill this existing gap in the current research literature.

3. PRE-STUDY: DESIGNER INTERVIEWS

Prior to implementing Design Maps, we carried out an interview pre-study to sensitize us to designers' viewpoint. We interviewed six professional UI designers (aged 24–44), to understand the inspiration-seeking practices they follow. To direct discussions to a concrete domain, we focused on UIs for mobile apps. Five of the interviewees had prior experience of at least 2 years in designing mobile apps, and had worked with wireframes in their initial design phases. One designer had no experience in mobile app design but was proficient (4+ years of experience) in graphical design applications and tools such as those from Adobe (Photoshop, Illustrator, Indesign) and Sketch.

We used the following general questions to conduct the interviews:

(i) What are the steps you follow when you are given a design task by a client?

(ii) Where do you find ideas/inspiration while creating the initial draft of your design?

(iii) What are the applications you normally use for tasks such as wireframing or prototyping?

(iv) How do you explore mobile UI designs (and how many) on app marketplaces like Google Play or App Store?

By using the questions above as a guiding thread, we inquired about inspiration-seeking of UI design examples, and how designers reflect on their everyday practices and design choices.

3.1. Findings

All the interviewees mentioned that they currently use some wireframing tool to bring their app designs into reality after getting the client requirements. *"Normally, a client briefs me with the requirements for his application and the UI of his interest and then my task is to start drafting the rough ideas into an appropriate structure using Sketch. I keep looking for competitor's app on the Play store and try to beat them in terms of design and usability."* They browse app marketplaces and search for existing applications to know more design details about their competitors as well as to gain inspiration from other designs. However, each designer has different methods to get inspiration, for example: borrowing concept ideas from websites like Behance, Dribble, Pinterest, and using other methods like mood boards (Garner and McDonagh-Philp,

2001). *"Every designer has their own process... Some sketch on article, post-its, and some like me use Sketch or even a whiteboard. It's important to put everything visually together so that I get an idea where I can start from."* Most interviewees said that they look at 10+ apps based on popularity (number of downloads, number of ratings, and stars) in the Google Play Store well before they start sketching their first idea.

Designers mentioned that they had not used any application that would enable them to explore existing apps and create wireframes side by side. Interestingly, the category of a mobile app does not matter very much to them in terms of inspiration. Instead, designers care a lot more about visual design characteristics, such as element structure, aesthetics, typography, and color. *"If I had to design a mobile app, it might be very likely that even after getting a look at several apps, I may still choose a design aesthetic seen on a different kind of app or a web application. It really depends what I can get in the most efficient ways."* They pointed out that they would like to interact with a design space visualization based on some UI similarity. They also pointed out that it would very valuable if mobile apps could be grouped together based on some similarity-based features, rather than category-based. They mentioned that UI design properties are dependent on the content of the app they are designing, and that such content is different for every app. They think it would be highly valuable if apps could be displayed based on their popularity and provide designers with some filtering capabilities.

Designers also mentioned that if they could use a visualization tool for exploring app designs, they would not explore *all* existing designs on their own, but instead would welcome a set of design suggestions based on some criteria. For example, design recommendations could be given based on an existing design, akin to the query-by-image feature used by major image search engines. Designers also mentioned that it would be extremely interesting if they could click on a design they like and then borrow specific parts of it.

At the end of the interview, participants were shown paper-based mockups of different layout options for a hypothetical UI wireframing tool. The tool would feature a visualization panel with Design Maps, a wireframing panel, and a recommendation panel. Participants discussed their visual preferences regarding panel arrangements and how they could interact with them. In sum, participants provided valuable feedback about their everyday design practices, what features they would expect from a wireframing tool, and seemed to be keen of the Design Maps concept.

4. DESIGN MAPS

Computing an understandable visualization of a design dataset is challenging, primarily because a design is a high-dimensional object and cannot be directly visualized in two dimensions. There are multiple dimensionality reduction techniques to achieve this. To the best of our knowledge, prior to this work they have not been investigated for the goal of representing UI design datasets. Moreover, as in any visualization scheme, there are many visual encodings that can be used to show the results, ranging from color and marker size to connectors, etc. Ultimately, a design map should offer intuitive 'geography' for navigating that space. Here we discuss problems and approaches in the computation of Design Maps, addressing these aforementioned challenges.

4.1. Dataset

To exemplify Design Maps, we use the Rico dataset (Deka et al., 2017), a repository of ~10,000 Android mobile app designs spanning 27 categories. It features ~72,000 individual screenshots with associated metadata such as annotations of element types, visual and structural data, and interactive design properties.[1] Rico provides semantic annotations for ~3 million UI elements, classified into 25 categories, such as icon, button, text, navigation, etc. A semantic annotation is an abstract representation of a given UI, identifying both structural and functional roles that UI elements play in the screen design (Liu et al., 2018). Rico also provides semantic screenshots, where each UI element is rendered as a bounding box with an associated color code; see Figure 1.

Figure 1: We use the Rico dataset for computing Design Maps. Examples shown here together with their associated semantic screenshots.

4.2. Algorithms

In Appendix A.1 we discuss all the algorithmic choices we took into consideration when creating the Design Maps for the Rico dataset. To summarize, semantic annotations are converted to RGB feature vectors and then transformed

---

[1]Rico is available at `http://interactionmining.org/rico`

to a 2D point cloud (Figure A1) using the following dimensionality reduction techniques (Section A.1.2):

Principal Component Analysis (PCA) Produces the best *linear* approximation to the original dataset, often with even spacing between data points.

Locally Linear Embedding (LLE) Each object is projected using a linear combination of their nearest neighbors. As a result, the resulting point cloud tends to be dense.

Uniform Manifold Approximation and Projection (UMAP) Produces a non-linear representation of the dataset with minimimun cross-entropy, ensuring that similar objects lie together and dissimilar objects are spread out.

t-distributed Stochastic Neighbor Embedding (tSNE) Converts distances between two objects into similarity probabilities, so that points close to one another in the original dataset will tend to be close in the resulting point cloud.

Finally, the point cloud is clustered using the K-means algorithm (Section A.1.3). As can be observed in Figure A2, we decided to choose $K = 5$ clusters since it seemed to be the best compromise solution in light of the validity measures we analyzed. These results suggest that designers exploring the Rico dataset should primarily focus on 5 large groups of UIs. Therefore, designers can quickly inspect one cluster by hovering over some UI examples and if no design is satisficing enough they can switch to exploring another cluster.

Figures 2 and 3 illustrate the resulting Design Maps we have created for the Rico dataset, each map comprising about 72k mobile UI designs. Each cluster found is highlighted in a different color, chosen at random. All figures provide three UI examples that are close to their respective cluster centroids, to illustrate the types of UI designs that can be expected in each cluster. As can be observed in the figures, each dimensionality reduction technique leads to a different map topology, informed by the mathematical principles we discus in the appendix, which is further reinforced by the color mapping assigned after clustering. We should emphasize that the main purpose of clustering the dimensionality-reduced data is to delimit the map areas that are likely to group similar UI designs together. Designers, however, are free (and encouraged) to explore each map in their way.

We should mention that there is only one map visualization (PCA, LLE, UMAP, or tSNE) for a given dataset, since it is created *deterministically*. On the one hand, we use the same seed for the random number generator used by the dimensionality reduction and clustering algorithms. On the other hand, the order in which the data is provided to the algorithms does not

matter, since both dimensionality reduction and clustering are order-agnostic algorithms. In sum, each visualization remains the same for the same dataset.

4.3. Integration in a design tool

Informed by the interview study (Section 3) and the comparison of Design Map visualizations (Section 4), we sought to develop a prototype to illustrate how Design Maps could be integrated into a design tool in a readily actionable manner, in order to empirically learn about them and how they can support creative processes in UI design. To this end, we created a design application that implements the Design Maps idea, see Figure 4. We named the application 'MobiSketch', in reference to the mobile dataset we used and to make it clear that this application is just one integration example of Design Maps. Indeed, because Design Maps are a general-purpose technique, they can generalize to other design datasets and even can be implemented in different ways as an end-user application.

MobiSketch is a web application that exposes 3 graphical components to the user (Figure 4):

(i) A **visualization** panel, that implements the Design Maps, enabling thus interactive exploration of the design space, and provides on-demand app information.

(ii) A **recommendation** panel, featuring a set of related UI screenshots that can be used as the basis for new designs.

(iii) A **wireframing** panel, that allows for creating UI wireframes and even borrowing parts from the recommended designs.

4.4. Visualization panel

The visualization panel (Figure 4.3) implements all the Design Maps we have analyzed in Section 4, based on the examined clustering and dimensionality reduction techniques. This panel also implements the visual information-seeking mantra (VISM) (Shneiderman, 1996):

Overview: The user is provided with a map of the design space, where apps are clustered based on UI semantics and are represented by a cloud point of size proportional to app popularity (number of downloads).

Zoom: The user can zoom in on items of interest and pan the maps, as well as explore nearby recommendations.

Filter: The user can filter out uninteresting items, based on app stars, number of ratings, downloads, and app category.

Details on demand: The user can get app details at any moment by hovering over each cloud point.

Relate: The user can notice how UI designs relate to each other within each semantic cluster.

History: The user can keep track of the designs they have browsed, and the actions they have performed over the wireframing panel.

Extract: The user can select a particular design and incorporate it to the wireframing panel.

4.5. Recommendation panel

Designers habitually seek inspiration from pre-existing related designs. However, the diversity of the set of examples provided is key for inspiring the user (Chan et al., 2015; Siangliulue et al., 2015). Morevover, in other contexts, it has been highlighted that users want to be shown more but "not just more of the same" (Iacobelli et al., 2010).

Therefore we devised the concept of "creative neighbors" to recommend within-cluster UIs that are similar to a given design but not too similar, to foster creativity. In a nutshell, the creative neighbors algorithm proceeds as follow: (i) set a threshold distance (we use $r/2$, where r is the radius of the candidates hypersphere), (ii) use the threshold distance to find the closest candidates to the user-specified design, (iii) sort the candidates found from larger to smaller distance and choose the top k candidates. Figure 5 illustrates the result of this procedure.

The design recommendations are presented as a list of actual UI screenshots from the Rico dataset (Figure 4.2), so that designers can get inspiration from real-world mobile apps. Previously selected designs and recommendations are available as a "design history", and they can be easily accessed through scrolling down this panel. As previously mentioned, it is possible to select a particular design and incorporate it to the wireframing panel. We implemented this feature by loading the semantic annotation of the selected design, which is available in the Rico dataset as a JSON file, and placing the corresponding wireframe icons on the drawing canvas.

4.6. Wireframing panel

The Wireframing panel (Figure 4.1) implements a standard HTML5 canvas element with basic drawing capabilities and a toolbar that allows for common actions such as undo/redo, group/ungroup, arrangement, or alignment. The set of wireframe icons was derived from the Rico dataset. The wireframe icons are incorporated to the drawing canvas by means of click or drag and drop. Once they are placed in the canvas, they can be scaled and
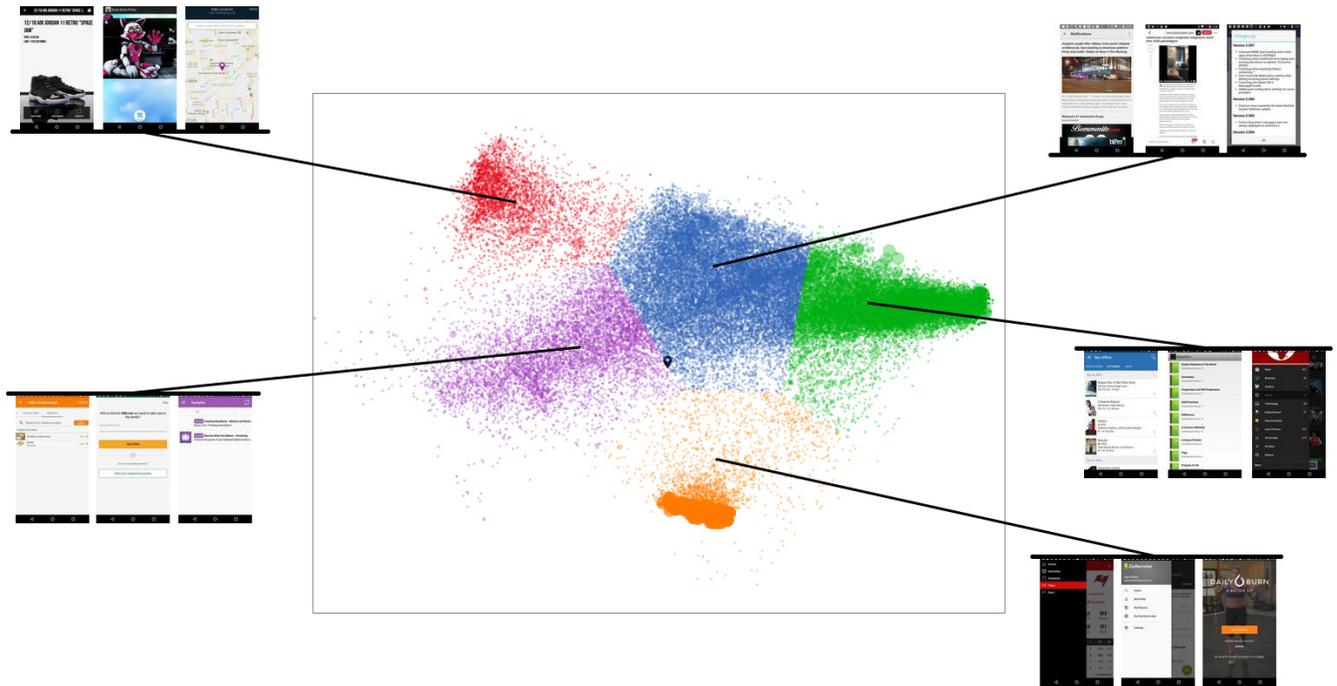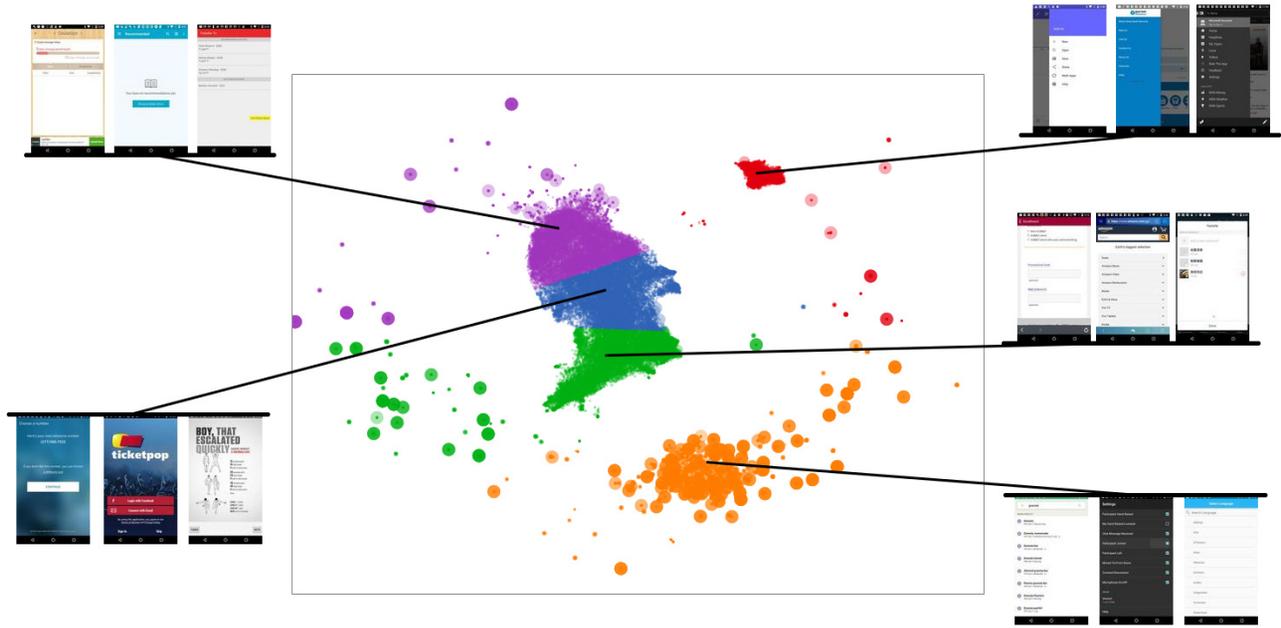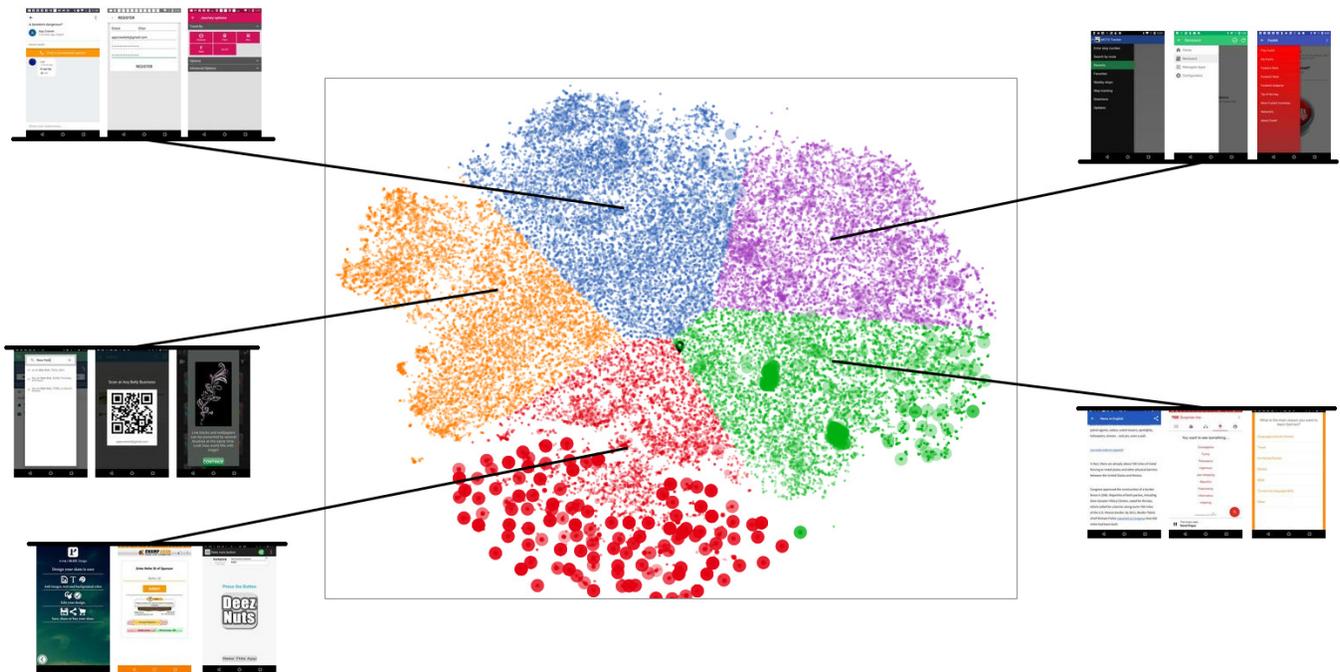
(a) PCA
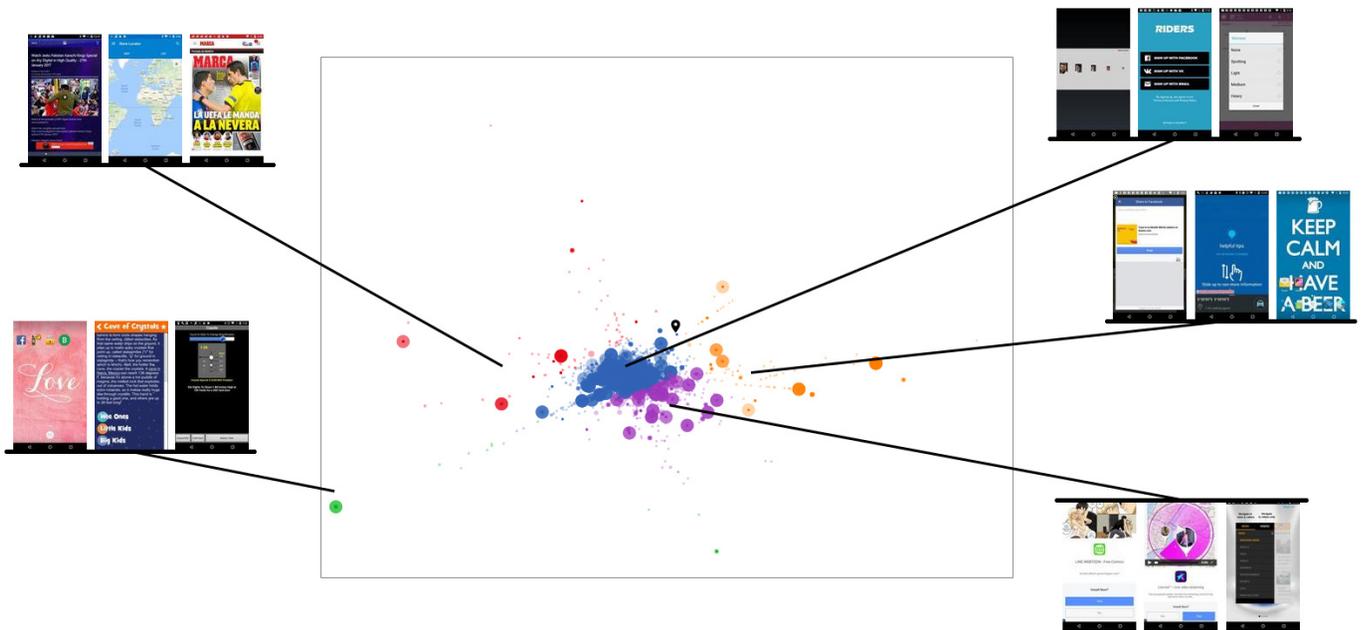
(b) LLE

Figure 2: Design Maps are based on clustering and different dimensionality reduction techniques. Each cluster found is highlighted in a different color. We provide three examples that are close to their respective cluster centroids, to illustrate the types of UI designs that can be expected in each cluster.

(a) UMAP

(b) tSNE

Figure 3: Continuation of Figure 2. Design Maps are based on clustering and different dimensionality reduction techniques. Each cluster found is highlighted in a different color. We provide three examples that are close to their respective cluster centroids, to illustrate the types of UI designs that can be expected in each cluster.
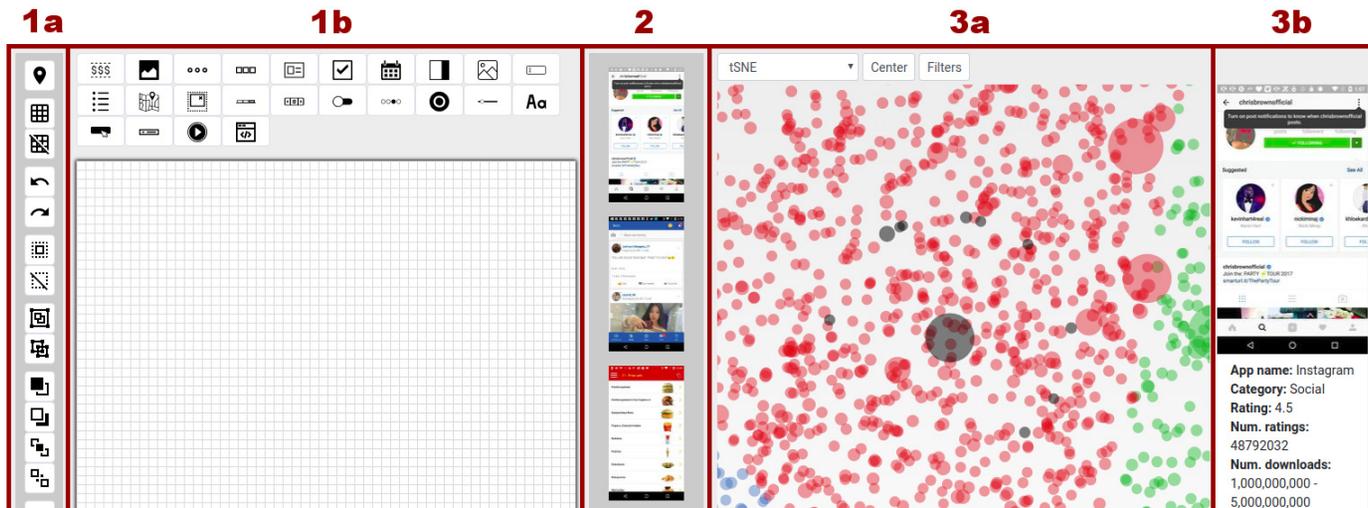
Figure 4: We developed a design tool for Design Maps featuring three components: a *wireframing panel*, comprising a toolbar (1a) and a canvas component (1b), a *recommendations panel*, comprising designs related to the user's selected design (2), and a *visualization panel*, comprising different Design Maps of existing mobile apps (3a) and an information component (3b) about the app UI that is selected on the design map.
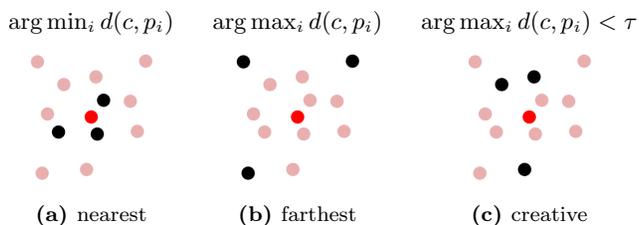

$$\arg\min_i d(c, p_i) \qquad \arg\max_i d(c, p_i) \qquad \arg\max_i d(c, p_i) < \tau$$

(a) nearest **(b)** farthest **(c)** creative

Figure 5: Looking for 3 recommendations (denoted in black color) starting from a point candidate c denoted in red color: nearest neighbors (a), farthest neighbors (b), and creative neighbors (c).

moved along the available canvas space. Each icon has an indicative shape and color, since each icon represents a particular UI element with a particular semantics.

Once the user has laid out some elements on the wireframing panel, it is possible to know the closest semantic cluster where the user's design lies within, in the toolbar (Figure 4.1a). Upon clicking on the map marker icon, MobiSketch automatically places a marker on the design space map, inspired by the you-are-here maps (Levine, 1982). This visual hint is meant to narrow down the search of inspirational UIs, by showing the most interesting cluster according to their current design. We implemented this feature with k-nearest-neighbor search over semantic UI embeddings, for which we followed the same autoencoder architecture proposed by Liu et al. (Liu et al., 2018) and used the learned latent space

of each semantic UI as feature vectors. Notice that nearest-neighbor search cannot be performed directly over the dimensionality reduced space, since it would require recomputing e.g. the covariance matrix in the case of PCA and the conditional point probabilities in the case of tSNE.

5. EVALUATION

The goal of Design Maps is to support creative processes in UI design, and in particular the exploration-exploitation problem therein. Design Maps offer a fast and structured way to navigate and explore a larger set of designs. However, it is not known how this affects the process and thinking in design. Therefore, to investigate this question, we conducted a controlled design study with professional designers. We used representative, challenging wireframing tasks. Design Maps were offered as part of a design tool, especially MobiSketch. This section details the experimental procedure and reports both quantitative and qualitative observations.

5.1. Participants

We recruited 12 designers (8 male, 4 female) with ages ranging from 25–44 years old. None of these participants took part in the pre-study. We aimed for a representative sample of end-users, who in our case are UI designers. Nine participants were professional designers with at least 2 years of experience in mobile app design (Android and iOS). The remaining three participants were studying

design courses at the master's level, and all of them did internships in a design company. All participants volunteered under informed consent and agreed to anonymized publication of results. We anticipated about one hour per participant, therefore they were compensated with a cinema voucher upon finishing the study.

5.2. Tasks

We devised three realistic tasks that would be accomplished by the participants.

(i) **Warm-up task:** The first task is a warm-up condition, in which participants had to create the splash screen of a photography app. For this task, participants had to use pen and paper to create the wireframes. They also could browse any website and take inspiration from any resource they would like.

(ii) **Wireframing from scratch:** The second task was aimed at gathering an overall impression of a dataset by using Design Maps. The participants had to create a login screen for an app that sells furniture online. This time participants had access to the full MobiSketch application, and were told that the visualization panel is available for getting inspiration.

(iii) **Completing a partial wireframe:** The third task was aimed at testing Design Maps under a design-constrained scenario, for which participants had to improve a product page for an app from a fashion company. This time participants were given a partial wireframe to begin with and had to complete it. This task is more complex than the previous task, since the designer has to build upon a set of design constraints and is expected to lay out more elements.

5.3. Procedure

At the beginning of each task, participants were given a design brief about the task at hand. Upon completing the first task, the experimenter demonstrated the Design Maps as a standalone application. Participants were told to explore the four different Design Maps (Figure 2 and Figure 3) and explain verbally what they think about each one. Then, the experimenter gave a walkthrough of the MobiSketch application and let the participants try themselves the different panels. After this, the experimenter proceeded with the second and third tasks.

At the end of the experiment, the experimenter conducted a semi-structured interview focusing on user experience, perceived issues, and the value proposition of Design Maps. Participants could also verbalize their thoughts and how we could improve Design Maps in the future. We also asked about general satisfaction with the outcomes produced with the wireframing tool.

In addition, also at the end of the experiment, each participant was administered the Creativity Support Index (CSI) questionnaire (Cherry and Latulipe, 2014). CSI is a standardized psychometric scale for assessing the perceived creativity support of a system/tool, looking at 6 factors: 1) collaboration, 2) enjoyment, 3) exploration, 4) expressiveness, 5) immersion, and 6) worthiness of effort. Since our design tool does not provide any collaboration capabilities, we skipped this factor and set the two corresponding scores to "neutral" (value of 5 out of 10 in the CSI scale) in order not to alter the calculation of the final CSI score (Cherry and Latulipe, 2014). The CSI score is a reflection of how well that tool supports creativity for the particular task or activity the user was engaged in and that is likely dependent on both individual preferences and the individual's level of expertise with the tool.

5.4. Quantitative results

Participants were told to rank the four Design Maps (Figure 2) in order of preference, from 1 (highest) to 4 (lowest). As can be observed in Table 1, participants chose often PCA as the most preferred choice to represent Design Maps. tSNE was on par with UMAP as the second most preferred option, and LLE was consistently ranked in the last position of the list (11/12 participants). When asked about their choices, most of the participants complained about LLE being too dense. They felt that having some whitespace or "semantic islands" was helpful. In the next section we shed more light about these impressions.

PCA	tSNE	UMAP	LLE
1	2	2.5	4

Table 1. Median user rankings of each visualization.

The Kruskal-Wallis rank sum test revealed a statistically significant difference between the four Design Maps presented to the participants ($\chi^2(3) = 27.28$, $p < .001$). We therefore ran pairwise comparisons using Wilcoxon rank sum test as post-hoc test, using the Bonferroni correction to guard against over-testing the data. The Wilcoxon test revealed no statistically significant differences between PCA, tSNE, and UMAP; but all of them significantly scored higher than LLE ($p < .001$).

Finally, as can be observed in Table 2, users rated MobiSketch high as a creativity support tool, with an average score of 73.97 (SD=13.26), which according to previous work is more than satisfactory (Cherry and Latulipe, 2014). The authors of the CSI questionnaire

reported that a score below 50 indicates that the tool does not support creative work very well. In the next section we comment on each of the CSI factors separately.

Factor	Mean	SD	95% Conf. Int.
Collaboration	–	–	–
Enjoyment	16.50	2.84	[14.89, 18.11]
Exploration	15.17	3.01	[13.46, 16.87]
Expressiveness	11.42	2.97	[9.74, 13.10]
Immersion	10.67	5.99	[7.28, 14.06]
Results Worth Effort	16.25	3.55	[14.24, 18.26]
Total	73.97	13.26	[66.47, 81.48]

Table 2. Creativity Support Index scores, the higher the better. The maximum score for each factor is 20.

5.5. Qualitative observations and discussion

We begin by distilling comments regarding the different factors of the CSI questionnaire, then discuss on different aspects of Design Maps as part of the MobiSketch wireframing tool.

5.5.1. Overall satisfaction
Participants highly appreciated Design Maps and the functionality of the recommendation panel. This was reflected by the average score for the 'Enjoyment' factor of the CSI questionnaire (16.5 out of 20). *"I normally use many applications and websites when I get started to drawing a layout or a wireframe, the best part which I felt that I now have access to lot of data in a small space and I can drag and drop designs directly."* All participants stated that they would like to use Design Maps for their professional work. *"This would save a lot of time and effort in crafting the initial designs and if you have a lot of work on your desk, it can be a quick way to get going and having something on board!"*

5.5.2. Decreasing effort and facilitating exploration
Although few participants (4/12) struggled to find the designs they wanted in LLE and UMAP-based Design Maps, they could actually create complete wireframes in less than ten minutes. The CSI 'Exploration' factor had an average score of 15.17. They felt Design Maps were not invasive of their creative space and let them take relevant decisions on the designs they create. *"According to me, designers are artists with a purpose. And I am surprised that I could use the suggestions to draw a design which I generally would have designed after browsing the web for a while on different platforms. Here my purpose is solved and that is great to be honest!"*

5.5.3. Reducing distractions
Participants felt that Design Maps were not deviating their attention from the actual task. *"I am good that the app is subtle and doesn't distract me too much."* The 'Immersion' factor of the CSI questionnaire has the lowest score on average (10.67/20), suggesting that participants were focused more on the activity rather than the tool itself. *"It's not one of those apps which keep bugging you even when you don't want them to. It's good the way it works right now."*

5.5.4. Perceived effect on productivity
Most of the participants (8/12) felt that they produced high-quality results given the amount of time and effort spent on the design tasks. *"If I had the initial wireframes ready very quickly, then I could focus on more important features like the functionality and design of my app workflow."* The CSI 'Results Worth Effort' factor had an average score of 16.25 and most participants were satisfied with the designs they created using the recommended UIs. *"I am quite satisfied because I did not have to do the boring stuff of creating the initial wireframes from the designs I look on web. I just drag and drop it and then build upon it. It was easy."*

5.5.5. Thoughts on design practice
Most participants (10/12) were happy with the concept of Design Maps as a visualization and exploration helper. They acknowledged the fact that Design Maps have the potential to be widely used e.g. as a plugin to any existing wireframing tools in the market. *"It might be a helper tool which can be used anytime I want to or I am stuck at some design or just too lazy to search at different places for better ideas."* *"Overall I think this can be a great addition to my toolkit that I regularly use."* Three participants mentioned an interesting use case of Design Maps beyond the context of wireframing. *"This could have a use case when you are actually designing the visual components of an app, may be the color, may be the different shapes as I sometimes take ideas from completely random designs. For instance, you can ask me to draw a cooking app and I can take ideas from a photo editing app, who knows!"*

Participants used the Design Maps extensively to explore the UI designs they wanted while performing the assigned tasks and used the recommendation panel to create their initial wireframes very quickly. While participants did not fully understand the structural differences of the four visualizations produced by each dimensionality reduction algorithm, they stated that Design Maps do encourage exploration and allow designers to get inspiration from many different designs at a single place. Because having too much information on the screen can become overwhelming, one participant suggested to incorporate semantic zooming, so that the selection and

structure of the data can change whenever they delve deeper into some clusters.

On another line, color is an important visual demarcator while exploring large datasets and most participants (9/12) asked the significance of the color scheme used in Design Maps. When they realized that each color represented a semantic cluster, they felt we could have made a better use of those, for example to highlight app similarity. However these concerns can be considered as improvements for the next iteration of Design Maps, and should thus not be seen as detrimental to their utility, as indicated by the participants both quantitatively and qualitatively.

Finally, participants appreciated the feature of showing the current design creation in the visualization panel, and mentioned that they really would like to use it to know their position in the Design Maps. *"There is no way to quantify this information now in any existing tool I've ever used, so I would use it for sure."* One participant also suggested that this feature could be used to even evaluate designers' work based on the designs they produce. In the supplementary materials accompanying this submission, we show some of the wireframes created by the participants.

5.5.6. Design recommendations

Sometimes the recommended designs were not similar to the designs our participants were searching for. For example, because Design Maps perform clustering based on UI semantics, a background image with a text on top of it can be either a login screen, a pop up screen, or a modal window. This can be frustrating if designers get very dissimilar suggestions constantly. However, this was taken as an opportunity by three participants, who acknowledged that those examples would help them to assess more diverse designs, that otherwise they would have not thought of.

5.5.7. Suggestions for improvement

A few participants (4/12) mentioned they would like to see the evolution of their designs, i.e, an art board where there are all the previous designs they created. This can be also used by the users to save all the liked designs from the recommendation panel. Other participants (2/12) commented on the fact that they do not only look for visual inspirations on mobile app marketplaces, but also they like test the usability and workflow of the different apps. Based on this observation, they suggested that MobiSketch could make this process more seamless by e.g. adding a link to each mobile app from Google Play so that designers can simply click on the link within MobiSketch.

6. DISCUSSION

While there is a lot of previous work on information visualization and dimensionality reduction techniques, as well as applications in broad areas of sciences, little work has been conducted to understand what types of visualizations can benefit exploration of design datasets without compromising designer's creativity. In this paper we have looked at the problem that there is a huge amount of design-related contents available on the Web, possibly to the extent that it is overwhelming, and in fact designers report struggling finding appropriate design examples (Koch et al., 2018).

Our results suggest that Design Maps have potential for supporting and encouraging design exploration. Design Maps allow designers to quickly get inspiration from several UI designs at the same place and at the same time. Our results also suggest that Design Maps can help designers explore diverse design examples that are meaningful. Since designers use prototypes for testing and evaluation (Deininger et al., 2017), Design Maps, when integrated in interactive design tools, can transport the wireframe to code that can be used in testing. Beyond finding an individual example, designers can also use Design Maps to store, organize, and maintain collections of inspirational materials (Keller et al., 2009).

Design Maps support exploration by projecting a high-dimensional design dataset into a few navigable dimensions on a display. We found that Design Maps that produce consistent whitespace *within* cloud points like tSNE and PCA resulted in better performance in terms of finding the UI designs that designers were looking for. The key characteristic of tSNE is that it solves the so-called *crowding* problem, but PCA is a linear projection and unexpectedly it was highly appreciated by the participants. On the contrary, Design Maps that produce consistent whitespace *between* cloud points (such as UMAP) or dense areas (such as LLE) were perceived as less helpful by designers.

We should mention that the position of UI designs is determined by the underlying dimensionality reduction algorithm but no axis labels are shown in the resulting design map. This was a deliberate choice in order not to mislead end-users. Take for example a PCA-generated Design Map: each axis represents "directions of maximum variance" (Bishop, 2006), which is difficult to communicate to and interpret by designers; e.g. What does variance mean for UI design? Is it an intrinsic property of a design dataset? May it guide visual inspection meaningfully? Previous work on music visualization projected high-dimensional data in two orthogonal axes with well-defined semantics, such as tempo vs timbre (Zhu and Lu, 2005) and arousal vs valence (Kim et al., 2011). For UI datasets, however,

we find it challenging to find a straightforward semantic mapping like those. Therefore, we eventually decided not to show axis labels in Design Maps.

We firmly believe that it is vital to study computational methods that can support designers cognitive and creative processes as opposed to automation of tasks. The type of machine learning techniques we have used in this work is not aiming at automating design, nor at being a proactive collaborative partner, but instead it is leveraged as a cognitive aid that helps structuring and navigating a complex design dataset. For example, by moving further away within a cluster, we have observed that the designs change in some determinate way; a property that can benefit divergent and convergent thinking. A virtue of a Design Map is that its use is entirely controlled by the designer, who can turn to it when needed, but is not required to act on it. Another virtue over previous approaches, like query-based search, is that design ideas can be skimmed very rapidly. The designers from our final evaluation reported this to be particularly helpful. We believe that this work has barely scratched the surface of the many kinds of work enabled by Design Maps.

Our participants reported highly appreciating Design Maps and the functionality enabled by the wireframing application, as reflected by the CSI 'Enjoyment' and 'Exploration' factors. They used the Design Maps extensively to explore the UI designs they wanted and used the recommendation panel to quickly create their initial wireframes. They rated the Design Maps high on performance as they believed they produced good results given the efforts they have put into the wireframing tool.

Research on divergent thinking and creative ideation suggests that being exposed to numerous examples will give researchers a concise but thorough picture of the many options available (Hocevar, 1979; Hocevar and Bachelor, 1989). Although irrelevant examples can negatively affect the quality of design ideas, and prior to this study it was not known if machine learning based visualizations can be supportive. Participants reported that Design Maps helped them avoid deviating their attention from the actual task. However, although 'Exploration' factor was high, 'Expressiveness' and 'Immersion' factors were relatively low.

We felt that designers were unable to internalize a high-fidelity mental model of the design space, which is understandable given the otherwise huge number of design examples available at a glance (more than 72,000 cloud points shown at once). However, they stated that Design Maps do encourage exploration and allow designers to get inspiration from many different designs at their own convenience. This finding deserves new ways of creating novel demarcators, which we discuss in the next section.

Finally, participants reported having used Design Maps extensively to explore the UI designs they wanted and find design examples that are meaningful and relevant to the design at hand. Most of the participants felt that they produced high-quality results given the amount of time and effort spent on the design tasks. This was further supported by the CSI 'Results Worth Effort' factor. We conclude that Design Maps, when integrated as part of a design tool, help designers reflect on their design choices.

6.1. Future work

While it is possible to create Design Maps from any design dataset comprising images only, we have observed that the resulting maps are more informative if we consider semantic information instead. Therefore, any domain that can provide labeled images together with their structural information is a good candidate for Design Maps. One such domain is e.g. websites, since existing URLs can be crawled and then processed to generate a screenshot together with the associated HTML structure. Another domain of interest are designs created with the Sketch application, which is very popular among professional UI designers. Sketch files are highly structured, so they could be used with little engineering effort to create Design Maps. We should also mention other popular professional applications among graphic designers, such as PhotoShop or InDesign, that produce images from structured data (e.g. layers in PhotoShop or XML nodes in InDesign) and therefore these applications could benefit from Design Maps if integrated e.g. as a plugin. Obviously, not every design dataset can be easily summarized as a Design Map, at least not without any modification of our existing solution. These include e.g. 3D objects, physical designs, motion captures, etc. As previously hinted, we need two basic components to create Design Maps in the same way as we did in this paper: (1) a collection of design images, for previsualization purposes, and (2) associated structural information, which can be featurized for later dimensionality reduction and clustering. Therefore, different design datasets may require a particular treatment, such as the ones we have discussed at the beginning of this section, namely: websites, Sketch projects, and related graphic design applications.

We plan to investigate hierarchical techniques for our design maps to allow for progressive exploration (Bikakis, 2019). Hierarchical visualization techniques have been extensively used in large graphs visualization, where the graph is recursively decomposed into smaller subgraphs that form a hierarchy of abstraction layers. Progressiveness can significantly improve efficiency in exploration scenarios, where it is common that users attempt to find something interesting, without knowing what exactly they are searching for beforehand. Also, we took note and will make a better use of color in

the future, for example using a warm-color palette to indicate how cluttered is a cluster, to express similar design objects, or encode additional metadata about the dataset at hand. Further visual encoding techniques could be explored to better structure the outputs, such as marker sizes, symbols, or contours. We could even make an explicit use of geographic metaphors, for example, creating correspondents of streets, districts, elevations, lakes, etc. It would be interesting to consider what categories or features might be added to the source dataset to accomplish this goal, though we acknowledge that some datasets may be limited, feature-wise. Therefore we leave these ideas as opportunities for future work.

Beyond the benefits on exploration, which our results suggest, several intriguing effects remain open. First, Design Maps could provide an overview of the design space via analogical transfer. It is known that the perception of analogues (structural similarities) among design instances can drive transfer of solutions in design (Gentner and Colhoun, 2010; Gick and Holyoak, 1983). Browsing a design map could help designers internalize a model of the design space that diversifies their thinking (Gomes et al., 2006). Further, example-driven analogical transfer is argued to be a cross-cutting aspect of creative design, shown for example in drawing (Chinmay Kulkarni and Klemmer, 2010), ideation (Siangliulue et al., 2015), and writing (Venables and Summit, 2003).

Design Maps allow designers to get a higher-level representation of a dataset that could be used like a mental map to find suitable designs. What if all datasets were shown the same way? That is, every dataset could be projected in a two-dimensional space with the same geographical shape. While this could help to transfer the designer's knowledge from one dataset to another, we believe it would eventually cause confusion. We argue that each dataset should have their own Design Map representation, to clearly inform the designers that they are exploring a particular dataset, but we are open to exploring this possibility in future work.

Further, designers usually take into account multi-faceted aspects of design features while seeking for inspiration, including e.g. the purpose and functionality of the design, user interactions and context, requirements of implementation, etc. It might be beneficial if we could make a good use of these features, provided that they are available in the source dataset. Rico, for example, provides user traces (sequences of interactions), animations capturing transition effects in response to user interaction, and rich app metadata (such as marketplace category, ratings, and number of downloads) that could be used to create more cohesive clusters.

It is interesting to note that curation activities, like one can engage in with Design Maps, can develop cognitive schemata (Lawson, 2004) that enable designers to recognize patterns in design problems. Leiva et al. (2020) have recently revised 10k Rico UIs and have concluded to a curated set of 1460 high-quality designs that have been categorized according to 20 topics; e.g. login screen, media player, gallery, etc. As an extension to this work, it would be interesting to recompute our Design Maps for these UI categories or even predict automatically the most likely topics for a set of new, unseen designs.

Finally, getting an overview of a design space can provide new input for reflection. Recognizing and weighing alternatives is known to be valuable throughout the design process (Tohidi et al., 2006). Examples that complement expert-created rubrics can facilitate a deeper understanding of design principles among feedback providers (Kang et al., 2018; Paivio, 1990). Projecting even further, having such design alternatives can facilitate comparative reasoning, ground team discussion, and enable situated exploration (Hartmann et al., 2008).

7. CONCLUSION

We have presented *Design Maps*, a novel class of colorful point cloud visualizations that can represent a large design dataset for interactive exploration. Design Maps are computed using dimensionality reduction and clustering techniques, the result being a two-dimensional representation of a higher dimensional design space that is broke down into clusters that can be navigated and positioned against the design at hand.

We exemplified Design Maps to enable large-scale exploration of mobile apps UIs, for which we devised wireframing (low-fidelity prototyping) as our use-case scenario. Professional designers reported that having Design Maps available in a wireframing tool supports their creativity. Designers also reported that Design Maps with equidistant spacing between markers were the most informative. Besides interactive exploration, Design Maps enable local neighborhood search and rapid exploitation of design examples. Both features were highly appreciated by our participants. Taken together, our results suggest that Design Maps support exploration and have potential for inspiration-seeking at large. Looking forward, we believe this work opens new possibilities to leverage novel visualization techniques to the benefit of design thinking.

A.1. ALGORITHMS FOR DESIGN MAPS

This section describes the algorithmic choices we took into consideration for Design Maps, namely dimensionality reduction and clustering algorithms.

A.1.1. Feature vectors

First of all, we need to define the *features* on which Design Maps computations are carried out. In early explorations, we quickly learned that clustering *raw* UI screenshots (i.e., using only pixel colors of the UI) is inadequate and, moreover, the resulting clusters are of low quality; i.e., as expected, UIs were basically clustered by color rather than structure. In further explorations we learned that using semantic screenshots produces more meaningful, semantically related clusters; e.g. UIs comprising clocks and calendars were grouped together. This got us closer to our goal of providing designers with a suitable (abstract) representation of the design space.

Further, to make clustering computationally effective without sacrificing image quality, such an abstract representation of the UI designs is converted to RGB feature vectors using a downsampled version of the semantic UI image. Then, the RGB vectors are further reduced to a small number of dimensions, since it is highly recommended to do so when the number of features is very high. This will suppress some noise and speed up later computations. For this, we used Principal Component Analysis (PCA) since it is highly performant. Then, we tried several techniques (to be discussed later) to compress the PCA-reduced feature vectors to 2 dimensions. Figure A1 illustrates this process.
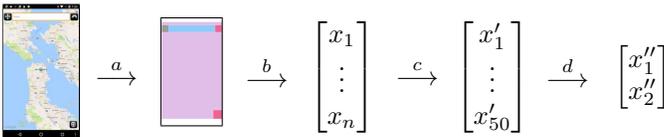

Figure A1: Computation of Design Maps: Each UI screenshot is replaced by their semantic screenshot downsampled to 300px height (a), converted to RGB feature vector (b), PCA-reduced to 50 dimensions to remove noise (c), then further reduced to 2 dimensions (d) using different algorithms. The resulting 2D data points are finally clustered using K-means.

A.1.2. Dimensionality reduction

We have tried several techniques to reduce our feature vectors to 2 dimensions (step e in Figure A1). All of the following techniques have proved in earlier research to be well-suited for reducing high-dimensional data for visualization in a low-dimensional space of e.g. two or three dimensions. However, their relative merits in Design Maps have not been studied before.

At a high level, given a D-dimensional dataset $\mathcal{X} \in \mathbb{R}^{N \times D}$ with N observations or data points, dimensionality reduction creates a projection $\mathcal{Z} \in \mathbb{R}^{N \times d}$ of $\mathcal{X}$ where $d \ll D$, in a way that the significant structure of the high-dimensional data $\mathcal{X}$ is preserved as much as possible in the low-dimensional space $\mathcal{Z}$.

A.1.2.1. Principal Component Analysis

PCA is the classic dimensionality reduction technique in machine learning. The goal of PCA is to reduce the number of variables (or features) of a dataset, while preserving as much information as possible. PCA uses an orthogonal linear transformation $z = x^T u$ to convert a set of (possibly correlated) observations into a set of linearly uncorrelated observations, called principal components (Abdi and Williams, 2010). The new data projection ensures that the greatest variance lies on the first axis (called the first principal component), the second greatest variance on the second axis, and so on. Geometrically speaking, principal components represent the directions of the data that explain a maximal amount of variance. Formally, PCA tries to minimizes the reconstruction error of the original data:

$$\text{maximize } u_k^T \Sigma u_k \text{ s.t. } ||u_k|| = 1 \qquad (A.1)$$

where k is the k-th dimension, μ is the sample mean, and Σ is the sample covariance. PCA thus finds the top-k eigenvalues u_k of the sample covariance matrix. Intuitively, PCA produces a point cloud that represents the best linear approximation to the original dataset.

A.1.2.2. Locally Linear Embedding

LLE seeks a lower-dimensional non-linear projection (embedding) of the data which preserves distances within local neighborhoods. It can be thought of as a series of local PCAs which are globally compared to find the best projection. LLE is advantageous because, unlike PCA, it involves no convergence (minimization) criteria. LLE tries to characterize the local geometry of the data by linear coefficients that reconstruct each data point $x \in \mathcal{X}$ from its neighbors (Roweis and Saul, 2000). Reconstruction errors are then measured by the cost function:

$$\mathcal{E}(w) = \sum_i ||x_i - \sum_j w_{ij} x_j||^2 \qquad (A.2)$$

which adds up the squared distances between all the data points and their reconstructions. The weights w_{ij} summarize the contribution of the jth data point to the ith reconstruction. Because LLE replaces each feature vector by a linear combination of their nearest vectors, the resulting point cloud tends to be dense.

A.1.2.3. Uniform Manifold Approximation and Projection

UMAP is a non-linear dimensionality reduction technique that uses local manifold approximations (McInnes et al.,

2018) to construct a topological representation of the high-dimensional data $\mathcal{P}$. A manifold is a data representation that maps a complex space to a simpler, smoother subset. UMAP iterates by creating several low-dimensional topological representations $\mathcal{Q}$ and selecting the one that minimizes the cross-entropy between both representations:

$$H(\mathcal{P}, \mathcal{Q}) = -\sum_{x \in \mathcal{X}} \mathcal{P}(x) \log \mathcal{Q}(x) \qquad (A.3)$$

One of the hyperparameters of UMAP is the desired separation between close points in the embedding space. This leads to spreading dissimilar objects out but also creates potentially densely packed regions for similar objects in the resulting point cloud.

A.1.2.4. t-distributed Stochastic Neighbor Embedding
tSNE is a non-linear dimensionality reduction technique that converts the Euclidean distances between two high-dimensional data points x_i and x_j into conditional probabilities that represent similarities (van der Maaten and Hinton, 2008):

$$p(j|i) = \frac{\exp\left(-||x_i - x_j||^2 / 2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-||x_i - x_k||^2 / 2\sigma_i^2\right)} \qquad (A.4)$$

where σ_i is the variance of a Gaussian centered at x_i, and for the low-dimensional counterparts z_i and z_j of the high-dimensional data points:

$$q(j|i) = \frac{\exp\left(-||z_i - z_j||^2\right)}{\sum_{k \neq i} \exp\left(-||z_i - z_k||^2\right)} \qquad (A.5)$$

And finally minimizes the sum of Kullback-Leibler divergences over all data points:

$$C(\mathcal{X}, \mathcal{Z}) = \sum_i \sum_j p(j|i) \log \frac{p(j|i)}{q(j|i)} \qquad (A.6)$$

The tSNE algorithm adapts its notion of distance to regional density variations in the data. As a result, it naturally expands dense areas, and contracts sparse ones, evening out the point cloud. The main advantage of tSNE is the ability to preserve the local structure of the data. This means that points which are close to one another in the high-dimensional dataset will tend to be close to one another in the resulting 2D point cloud.

A.1.3. Clustering

Clustering algorithms allow for discovering "natural" groups in a dataset, where objects in a group are similar to each other and different from objects in other groups. The idea is to divide a dataset $\mathcal{X} = \{x_1, \ldots, x_N\}$ of N feature vectors into a set $\prod = \{\mathcal{C}_1, \ldots, \mathcal{C}_K\}$ of K disjoint homogeneous classes with $1 < K \ll N$. One way to tackle this problem is to define a criterion function that measures the quality of the clustering partition and then find a partition $\prod^*$ that optimizes such a criterion function.

There exist several algorithms for data clustering, among which we used the K-means algorithm, which is known for its simplicity, relative robustness, and fast convergence to local minima. The most common version of this algorithm, generally attributed to Lloyd (Lloyd, 1982), uses a heuristic minimum-distance criterion, where in each iteration all the data points are assigned to their closest cluster mean (centroid) and convergence is achieved when the assignments no longer change. There exists, however, a more interesting version, often attributed to Duda and Hart (Duda and Hart, 1973), which uses a sample-by-sample iterative optimization refinement scheme. At each step, the sum of squared errors is evaluated and the considered sample is reallocated to a different cluster if and only if that reassignment decreases the overall cluster variance. Clearly, such a greedy optimization guarantees that the resulting partition corresponds to a local minimum. In this article we follow this approach.

The cost function that K-means tries to minimize is the sum of squared errors *within* clusters, which emphasizes the local structure of the data (Veenman et al., 2002):

$$SSE_w = \sum_{j=1}^{K} \sum_{x \in \mathcal{C}_j} ||x - \mu_j||^2 \qquad (A.7)$$

where μ_j is the centroid of cluster $\mathcal{C}_j$.

A fundamental choice in clustering techniques is that of the similarity metric. However, K-means implicitly assumes Euclidean distances, since the algorithm minimizes within-cluster variance and it is equal to the sum of squared Euclidean distances of each cluster member to its centroid. In fact one cannot use arbitrary distance functions, otherwise the algorithm may not converge. Therefore, we use the Euclidean distance over the 2-dimensional feature vector space.

A.1.4. Creating Design Maps

Armed with the theoretical and mathematical knowledge previously introduced, we are now ready to create different Design Maps, discuss the challenges they face, and illustrate how the look like in practice.

Clustering a design dataset may help further understanding the underlying structure in it. In Design Maps, the main purpose of clustering the above-mentioned dimensionality-reduced data is to delimit the map areas that are likely to group similar UI designs together. This can provide designers with a better visual understanding
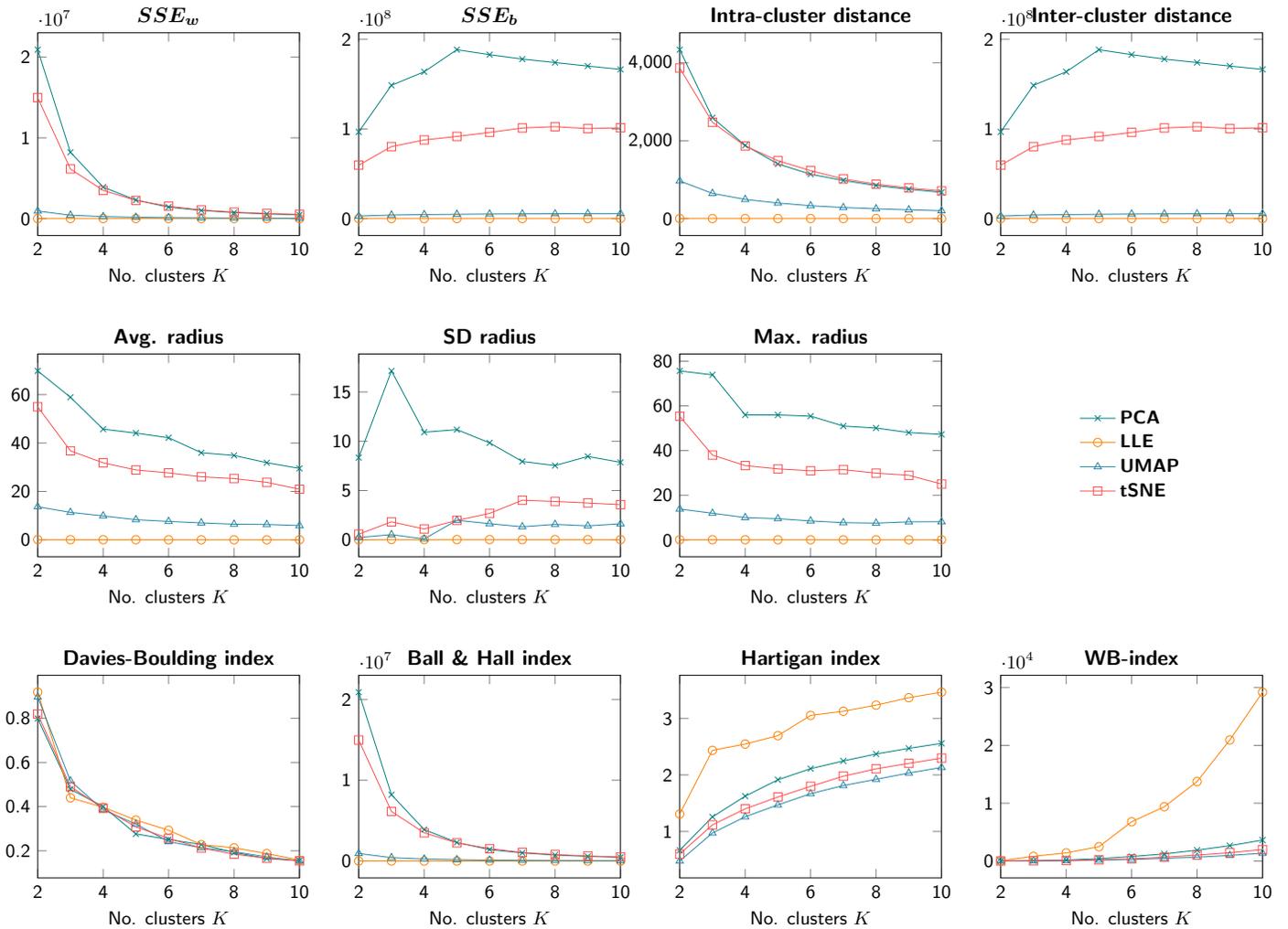
Figure A2: Evaluation of UI clustering results as a function of the number of clusters K, for different dimensionality reduction techniques.

of a large dataset, since the clusters found can be assigned a distinct color and therefore they can be easily noticed at a glance.

Because K-means is an unsupervised learning algorithm, it can make inferences about the data using only the feature vectors, without having to label every data instance. As a downside, however, the algorithm requires as input the value of K because the "best" value of K depends a lot on the dataset. Therefore, in order to find a suitable K we used the *elbow* criterion (Leiva and Vidal, 2010) or 'knee point detection': one should choose K so that adding another cluster does not give much better modeling of the data. To achieve this goal, the algorithm is increasingly run from $K = 1$ to $K = N$ and stops as soon as some key measure does not improve w.r.t. the previous value of K. Figure A2 illustrates this process.

Since the K-means algorithm minimizes the SSE_w (Equation A.7), we will choose as best configuration the one that does not improve significantly the percentage of variance explained from one step to another.

The fitness of a clustering configuration can be measured in multiple ways. We explore these to understand the different dimensionality reduction techniques. In addition to the sum of squared errors within clusters, we report macro-measures of the clustering geometry such as mean and standard deviation of cluster radius. These measures provide an overview of each clustering configuration and help to quantify the impact of adding or removing a new cluster.

We also report several well-known evaluation measures for clustering configurations: Davies-Boulding index (Davies and Bouldin, 1979), Ball & Hall index (Ball and Hall, 1965), Hartigan index (Hartigan, 1975), and

WB-index (Zhao and Fränti, 2014). These evaluation measures favor clustering configurations with high similarity within a cluster (compactness) and low similarity between clusters (separation). However, according to theoretical analysis (Liu et al., 2010), some of these measures may be affected by the configuration and structure of the data. Therefore, a thorough analysis becomes necessary.

As can be observed in Figure A2, the most suitable number of clusters gravitates around $K = 5$ or $K = 6$ in most cases; see e.g. the plots in the top row of the figure. This observation can also be inferred from the geometry of the clustering configurations: a plateau in the cluster radii can be easily noticed for both values of K; see middle row of Figure A2. This is further supported by the different cluster validity measures, e.g. the Davies-Boulding and Ball & Hall indices show no noticeable improvement for $K \geq 6$. A similar trend can be observed in the plots pertaining Hartigan and WB indices. All in all, we decided to choose $K = 5$ since it seemed to be the best compromise solution in light of the above-mentioned measures. These results suggest that designers exploring the Rico dataset should primarily focus on 5 large groups of UIs. Therefore, designers can quickly inspect one cluster by hovering over some UI examples and if no design is satisficing enough they can switch to exploring another cluster.

A.2. MOBISKETCH REMARKS

We would like to emphasize on a couple of key capabilities of our MobiSketch application that, we believe, will improve designers' productivity while navigating Design Maps. The first key capability is that of *filtering*, by which it is possible to reduce the number of design objects shown on the map by indicating e.g. a minimum or maximum number of downloads, app ratings, or even indicating multiple app categories by means of check boxes. The second key capability is that of *copying*, by which it is possible to drag and drop an existing UI preview (thumbnail) to the design canvas and then borrow the desired design parts. Figure A1 and Figure A2 illustrate these filtering and copying capabilities, respectively.

A.3. SUPPLEMENTARY MATERIALS

We provide a video of Design Maps and sample screenshots of the wireframes created by our participants. We will release our software to create Design Maps upon this article's publication.

ACKNOWLEDGEMENTS

REFERENCES

Abdi, H., Williams, L. J., 2010. Principal component analysis. WIREs Comput. Stat. 2 (4), 433–459.

Ball, G., Hall, D., 1965. ISODATA, a novel method of data analysis and pattern classification. Tech. rep., Stanford Research Institute.

Behrang, F., Reiss, S. P., Orso, A., 2018. GUIfetch: Supporting app design and development through GUI search. In: Proc. MOBILESoft. pp. 236–246.
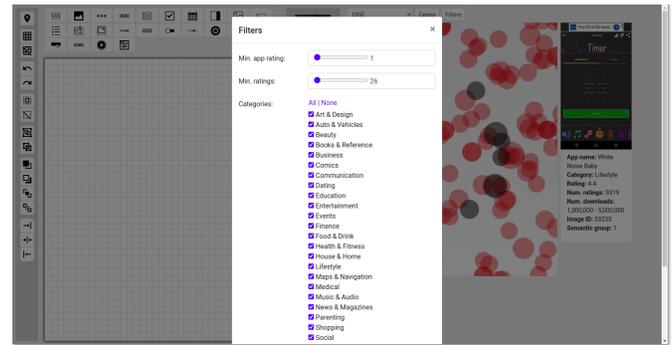

Figure A1: Filtering capabilities. In MobiSketch it is possible to reduce the number of design objects show on the map by selecting the desired filters, e.g. number of received app ratings or app categories.
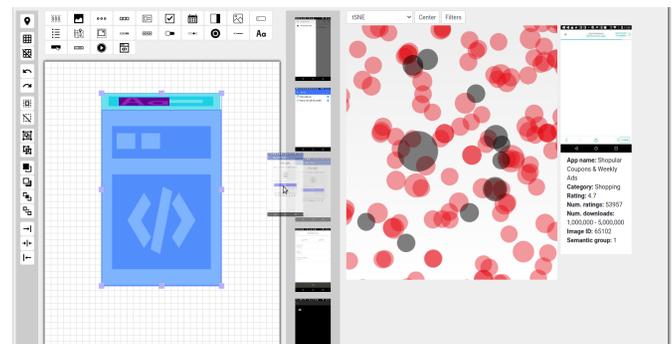

Figure A2: Copying capabilities. In MobiSketch it is possible to borrow existing design parts from the pool of recommended designs (central panel) by means of a drag and drop operation, shown here with a semi-transparent picture of the thumbnail being dragged.

Bernard A. Nijstad, Carsten K. W. De Dreu, E. F. R., Baas, M., 2010. The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence. Eur. Rev. Soc. Psychol. 21, 34–77.

Bikakis, N., 2019. Big Data Visualization Tools. Springer.

Bishop, C., 2006. Pattern Recognition and Machine Learning. Springer-Verlag.

Card, S. K., Mackinlay, J., 1997. The structure of the information visualization design space. In: Proc. InfoVis. pp. 92–100.

Chan, J., Dow, S. P., Schunn, C. D., 2015. Do the best design ideas (really) come from conceptually distant sources of inspiration? Design Stud. 36, 31–58.

Chan, J., Siangliulue, P., Qori McDonald, D., Liu, R., Moradinezhad, R., Aman, S., Solovey, E. T., Gajos, K. Z., Dow, S. P., 2017. Semantically far inspirations considered harmful?: Accounting for cognitive states in collaborative ideation. In: Proc. C&C. pp. 93–105.

Cherry, E., Latulipe, C., 2014. Quantifying the creativity support of digital tools through the creativity support index. ACM Trans. Comput.-Hum. Interact. 21 (4), 21:1–21:25.

Chinmay Kulkarni, S. P. D., Klemmer, S. R., 2010. Early and Repeated Exposure to Examples Improves Creative Work. Springer.

Crilly, N., Cardoso, C., 2017. Where next for research on fixation, inspiration and creativity in design? Design Stud. 50, 1–38.

Davies, D. L., Bouldin, D. W., 1979. A cluster separation measure. IEEE Trans. Pattern Anal. Mach. Intell. 1 (2), 224–227.

Deininger, M., Daly, S. R., Sienko, K. H., Lee, J. C., 2017. Novice designers' use of prototypes in engineering design. Design Stud. 51, 25–65.

Deka, B., Huang, Z., Franzen, C., Hibschman, J., Afergan, D., Li, Y., Nichols, J., Kumar, R., 2017. Rico: A mobile app dataset for building data-driven design applications. In: Proc. UIST. pp. 845–854.

Duda, R. O., Hart, P. E., 1973. Pattern Classification and Scene Analysis. John Wiley & Sons.

Eckert, C., Stacey, M., 2000. Sources of inspiration: a language of design. Design Stud. 21 (5), 523–538.

Eddy, J., Lewis, K. E., 2002. Visualization of multidimensional design and optimization data using cloud visualization. In: Proc. DAC. pp. 899–908.

Eppler, M. J., Kernbach, S., 2016. Dynagrams: Enhancing design thinking through dynamic diagrams. Design Stud. 47, 91–117.

Garner, S., McDonagh-Philp, D., 2001. Problem interpretation and resolution via visual stimuli: the use of 'mood boards' in design education. Int. J. Art Des. Educ. 20 (1), 57–64.

Gentner, D., Colhoun, J., 2010. Analogical Processes in Human Thinking and Learning. Springer.

Gick, M. L., Holyoak, K. J., 1983. Schema induction and analogical transfer. Cogn. Psychol. 15 (1), 1–38.

Gomes, P., Seco, N., Pereira, F. C., Paiva, P., Carreiro, P., Ferreira, J. L., Bento, C., 2006. The importance of retrieval in creative design analogies. Knowl.-Based Syst. 19 (7), 480–488.

Gonçalves, M., Cardoso, C., Badke-Schaub, P., 2014. What inspires designers? preferences on inspirational approaches during idea generation. Design Stud. 35 (1), 29–53.

Hartigan, J. A., 1975. Clustering algorithms. John Wiley & Sons.

Hartmann, B., Yu, L., Allison, A., Yang, Y., Klemmer, S. R., 2008. Design as exploration: Creating interface alternatives through parallel authoring and runtime tuning. In: Proc. UIST. pp. 91–100.

Herring, S. R., Chang, C.-C., Krantzler, J., Bailey, B. P., 2009. Getting inspired!: Understanding how and why examples are used in creative design practice. In: Proc. CHI. pp. 87–96.

Hocevar, D., 1979. Ideational fluency as a confounding factor in the measurement of originality. J. Educ. Psychol. 71 (2), 191–196.

Hocevar, D., Bachelor, P., 1989. A taxonomy and critique of measurements used in the study of creativity. Springer, pp. 53–75.

Huang, F., Canny, J. F., Nichols, J., 2019. Swire: Sketch-based user interface retrieval. In: Proc. CHI. pp. 1–10.

Iacobelli, F., Birnbaum, L., Hammond, K. J., 2010. Tell me more, not just "more of the same". In: Proc. IUI. pp. 81–90.

Jorge, A., Correia, N., Chambel, T., 2017. Designing interactive spatiotemporal visualizations to enhance movie browsing. In: Proc. INTERACT. pp. 352–355.

Kang, H. B., Amoako, G., Sengupta, N., Dow, S. P., 2018. Paragon: An online gallery for enhancing design feedback with visual examples. In: Proc. CHI. pp. 606:1–606:13.

Keller, I., Visser, F. S., van der Lugt, R., Stappers, P. J., 2009. Collecting with cabinet: Or how designers organise visual material, researched through an experiential prototype. Design Stud. 30 (1), 69–86.

Kim, J., Lee, S., Kim, S., Yoo, W. Y., 2011. Music mood classification model based on arousal-valence values. In: Proc. ICACT. pp. 292–295.

Knerr, N., Selva, D., 2016. Cityplot: Visualization of high-dimensional design spaces with multiple criteria. J. Mech. Des. 138 (9), 1–9.

Koch, J., László, M., Lucero, A., Oulasvirta, A., 2018. Surfing for inspiration: Digital inspirational material in design practice. In: Design Research Society 2018 Catalyst (DRS2018). pp. 1247–1260.

Kumar, R., Satyanarayan, A., Torres, C., Lim, M., Ahmad, S., Klemmer, S. R., Talton, J. O., 2013. Webzeitgeist: Design mining the web. In: Proc. CHI. pp. 3083–3092.

Lawson, B., 2004. Schemata, gambits and precedent: some factors in design expertise. Design Stud. 25 (5), 443–457.

Lee, B., Srivastava, S., Kumar, R., Brafman, R., Klemmer, S. R., 2010. Designing with interactive example galleries. In: Proc. CHI. pp. 2257–2266.

Lee, C., Kim, S., Han, D., Yang, H., Park, Y.-W., Kwon, B. C., Ko, S., 2020. GUIComp: A GUI design assistant with real-time, multi-faceted feedback. In: Proc. CHI. pp. 1–13.

Leiva, L. A., Hota, A., Oulasvirta, A., 2020. Enrico: A high-quality dataset for topic modeling of mobile UI designs. In: Proc. MobileHCI Extended Abstracts.

Leiva, L. A., Vidal, E., 2010. Assessing users' interactions for clustering web documents: A pragmatic approach. In: Proc. HT. pp. 277–278.

Levine, M., 1982. You-are-here maps: Psychological considerations. Environ. Behav. 14 (2), 221–237.

Liu, T. F., Craft, M., Situ, J., Yumer, E., Mech, R., Kumar, R., 2018. Learning design semantics for mobile apps. In: Proc. UIST. pp. 569–579.

Liu, Y., Li, Z., Xiong, H., Gao, X., Wu, J., 2010. Understanding of internal clustering validation measures. In: Proc. ICDM. pp. 911–916.

Lloyd, S., 1982. Least squares quantization in PCM. IEEE Trans. Inf. Theory 28 (2), 129–137.

Lübbers, D., Jarke, M., 2009. Adaptive multimodal exploration of music collections. In: Proc. ISMIR. pp. 195–200.

Marks, J., Andalman, B., Beardsley, P. A., Freeman, W., Gibson, S., Hodgins, J., Kang, T., Mirtich, B., Pfister, H., Ruml, W., Ryall, K., Seims, J., Shieber, S., 1997. Design galleries: A general approach to setting parameters for computer graphics and animation. In: Proc. SIGGRAPH. pp. 389–400.

Martins, P., Langlois, T., Chambel, T., 2011. MovieClouds: Content-based overviews and exploratory browsing of movies. In: Proc. MindTrek. pp. 133–140.

Matejka, J., Glueck, M., Bradner, E., Hashemi, A., Grossman, T., Fitzmaurice, G., 2018. Dream Lens: Exploration and visualization of large-scale generative design datasets. In: Proc. CHI. pp. 369:1–369:12.

McInnes, L., Healy, J., Melville, J., 2018. UMAP: Uniform manifold approximation and projection for dimension reduction. ArXiv e-prints.

Moacdieh, N., N., S., 2015. Display clutter: a review of definitions and measurement techniques. Hum. Factors 57 (1), 61–100.

Moggridge, B., 2006. Designing Interactions. MIT Press.

Moritz, D., Fisher, D., Ding, B., Wang, C., 2017. Trust, but verify: Optimistic visualizations of approximate queries for exploring big data. In: Proc. CHI. pp. 2904–2915.

O'Donovan, P., Agarwala, A., Hertzmann, A., 2015. Design-Scape: Design with interactive layout suggestions. In: Proc. CHI. pp. 1221–1224.

Paivio, A., 1990. Mental representations: A dual coding approach. Oxford University Press.

Pampalk, E., Rauber, A., Merkl, D., 2002. Content-based organization and visualization of music archives. In: Proc. MM. pp. 570–579.

Perttula, M., Sipilä, P., 2007. The idea exposure paradigm in design idea generation. J. Eng. Design 18 (1), 93–102.

Richardson, T., Winer, E., 2011. Visually exploring a design space through the use of multiple contextual self-organizing maps. In: Proc. DAC. pp. 857–866.

Roweis, S., Saul, L., 2000. Nonlinear dimensionality reduction by locally linear embedding. Science 290 (5500), 2323–2326.

Sharmin, M., Bailey, B. P., Coats, C., Hamilton, K., 2009. Understanding knowledge management practices for early design activity and its implications for reuse. In: Proc. CHI. pp. 2367–2376.

Shneiderman, B., 1996. The eyes have it: a task by data type taxonomy for information visualizations. In: Proc. IEEE Symposium on Visual Languages. pp. 336–343.

Siangliulue, P., Chan, J., Gajos, K. Z., Dow, S. P., 2015. Providing timely examples improves the quantity and quality of generated ideas. In: Proc. C&C. pp. 83–92.

Stolterman, E., 2008. The nature of design practice and implications for interaction design research. Int. J. Des. 2 (1), 55–65.

Swearngin, A., Dontcheva, M., Li, W., Brandt, J., Dixon, M., Ko, A. J., 2018a. Rewire: Interface design assistance from examples. In: Proc. CHI. pp. 504:1–504:12.

Swearngin, A., Ko, A. J., Fogarty, J., 2018b. Scout: Mixed-initiative exploration of design variations through high-level design constraints. In: Adj. Proc. UIST. pp. 134–136.

Tohidi, M., Buxton, W., Baecker, R., Sellen, A., 2006. Getting the right design and the design right. In: Proc. CHI. pp. 1243–1252.

Vad, B., Boland, D., Williamson, J., Murray-Smith, R., Steffensen, P. B., 2015. Design and evaluation of a probabilistic music projection interface. In: Proc. ISMIR. pp. 134–140.

van der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res., 2579–2605.

Vasconcelos, L. A., Crilly, N., 2016. Inspiration and fixation: Questions, methods, findings, and challenges. Design Stud. 42, 1–32.

Veenman, C. J., Reinders, M. J. T., Baker, E. L., 2002. A maximum variance cluster algorithm. IEEE Trans. Pattern Anal. Mach. Intell. 24 (9), 1273–1280.

Venables, A., Summit, R., 2003. Enhancing scientific essay writing using peer assessment. Innov. Educ. Teach. Int. 40 (3), 281–290.

Wallace, S., Le, B., Leiva, L. A., Haq, A., Kintisch, A., Bufrem, G., Chang, L., Huang, J., 2020. Sketchy: Drawing inspiration from the crowd. Proc. ACM Hum.-Comput. Interact. 4 (CSCW2).

Woodbury, R., Mohiuddin, A., Cichy, M., Mueller, V., 2017. Interactive design galleries: A general approach to

interacting with design alternatives. Design Stud. 52, 40–72.

Zhang, X. L., Simpson, T., Frecker, M., Lesieutre, G., 2012. Supporting knowledge exploration and discovery in multi-dimensional data with interactive multiscale visualisation. J. Eng. Des. 23 (1), 23–47.

Zhao, Q., Fränti, P., 2014. WB-index: A sum-of-squares based index for cluster validity. Data Knowl. Eng. 92, 77–89.

Zhu, J., Lu, L., 2005. Perceptual visualization of a music collection. In: Proc. ICME. pp. 1058–1061.