Color Me *Curious*

Dina Bseiso and Natalie Goldberg
Information Visualization, Spring 2016
http://groups.ischool.berkeley.edu/ColorMeCurious/

Project Goals

We would like to compare differences across fashion lines as designed by particular fashion designers. A few dimensions of initial interest are dominant colors and whether they vary across lines depending on the global location of the fashion show, the nation of origin of the designer, the season the line is intended for, and whether each designer has a particular color scheme they stick to (as in, are designer color choices more similar across their lines than when compared against lines from their peers?).

The visualization is meant to be interactive, and we hope to allow the user to explore the data by selecting certain dimensions of interest to order designer lines by. The user should be able to refine their investigation in the following ways: designer, designer's country of citizenship, Fashion Week city location, Spring/Summer year, and Fall/Winter year. After much consideration, we believed allowing a user to refine their investigation in these ways would not constrain their intellectual and creative curiosities, and would also be clear in presentation. We did not want the user to feel wanting, that they would have liked to compare the data in a certain way that our model simply did not allow. The designer lines are ordered in chronological order within a designer, and in alphabetical order by last name of the designer. So, as a user looks down a row of data, they should be able to observe the designer house's choice in color schemes over the five year period, in both spring/summer and fall/winter seasons (or as they so choose to filter).

The visualization is also meant to be informative, while avoiding superfluous additions. We wanted it to be clear that the merged color images were compiled from actual outfit data (in the event that a user did not read our methods section), and so superimposed these designer thumbprints on a very simple, modest female figure. The female figure remained consistent across all visible lines, as to allow the user to habituate to her being there and accentuate the details in the colors she wears instead. Initially, we provided additional, more granular designer-line information upon hover; however, after some usability studies we found that users wanted that information upfront. After additional user-testing on a version that allowed that information to be readily available and persistent on the page, it was determined that the added information was not distracting the user from exploring the designer thumbprints. More usability-related findings will be expanded upon below in our Results section.

This visualization is intended for the lay-user, as opposed to professionals within the fashion industry. A visualization intended for a professional in such an industry may include color palettes, so that one designer's shade of red may be compared to another designer's shade of red. These nuances are fascinating to the fashion professional niche, but less-so captivating to a general audience that clusters colors in ROYGBIV-ways (ROYGBIV stands for red-orange-yellow-green-blue-indigo-violet and is a pneumonic children learn at a young age to remember the colors of the rainbow; a note, though, is that people also strongly associate color clusters with black, white, and grey).

Related Work

1. NY Times Fashion

— http://www.nytimes.com/newsgraphics/2013/09/13/fashion-week-editors-picks/
This example shown in class gave us motivation for this project. It mainly displays work but doesn't go into much detail about designer color choices in trends over time, which is something we'd like to provide.

2. Computer Vision Academic Literature and Object Recognition
Much work has been done with regard to computer vision and fashion, in trying to best identify articles of clothing. While that methodology is outside the scope of our objective, the existence of such work allowed us to believe it was feasible to embark on familial algorithmic endeavors with fashion.

3. Menu Journeys
This InfoViz final project was done last year, and was showcased as an example of an appropriate final project idea to our class. What was inspirational about this visualization is that it took atypical data — data that originated from media, and not raw statistics — and transformed it in a way that could be digested by a user. Menu Journeys scraped menu items from historical and archived menus, originally in a paper medium, and parsed the information such that trends on popularity of ingredients and dish prices could be observed. Our project relates to this former InfoViz project in the form of the input being used — not that the input are alike in content, but in nature. Our project utilizes raw image data, values from which (in the form of pixel color data in our case) are extracted, digested, and then transformed in a way that communicates overarching trends to users.

4. Retail Catalogues
While not explicitly information visualizations, per say, online retail catalogues served as foundational examples to model our information architecture. Online retail catalogues have existed long enough to have undergone ample usability testing, and so implement intuitive interactive elements. Online retail catalogues also implement extensive filtering flexibility, which we thought was important for a user's exploration of our dataset. Since our visualization would be rendered useless if found unusable, we took

great care to research various filtering displays and considered the intentions behind using certain filters. It was after this consideration that we resolved to use a combination of filter types: checkbox lists and dropdown lists.

5. FashionWeek itself
We aimed to embody the culture of FashionWeek in our visualization in addition to using it as the source of all our data. We investigated important properties of FashionWeek — that it happens twice a year: one in spring, and one in fall; that it mostly occurs in the following cities: Milan, New York, and Paris. We did not believe that straying from these important attributes of FashionWeek would make for a polished study, and would have us running the risk of looking at the wrong dimensions for investigation (such as exceptionally rare FashionWeek city locations and one-hit designers). Our preliminary research on the details of FashionWeek allowed us to select our dimensions of analysis, as well as the designers through which to examine those dimensions.

An image to portray just how colorful Fashion Week can be.

6. Fashion Blogs
Fashion blogs inspire and give readers new ideas on how to dress, and provide updates on designer lines. We hoped to give our viewers a similar benefit in providing insight into which designers they would most feel a connection with when making fashion decisions. Also, blog writers instill their own choices and opinions in their blog posts, which is what we hoped to show in our visualization by giving a holistic picture of the designer beyond a single line or piece of clothing.

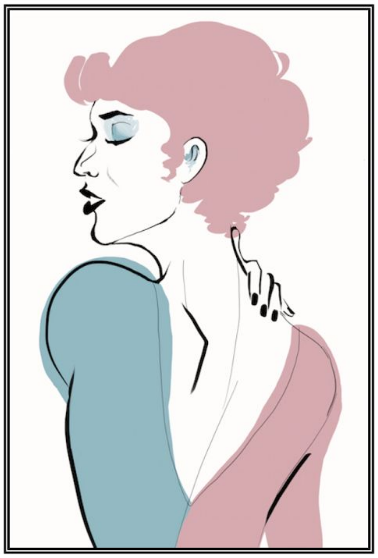
7. … and the greater fashion scene — in print and online.
But because not everyone can make it out to these FashionWeek cities to observe firsthand the debut of fashion trends, we looked toward more accessible artifacts of fashion archival (be it magazines, fashion blogs, fashion design sketchbooks) to create an aesthetic that would be immediately familiar to the fashion scene. Our intended user group is not the fashiona professional, but an everyday soul who may have curiosities of fashion every now and again, or who bear strong assumptions about color in fashion that have never been challenged before due to a lack of means to do so.

8. Evocative Imagery and Holmes
Ever since the beginning of the term, it has been a goal of Dina's to attempt to incorporate illustration into a data visualization in a meaningful way that enhances the meaning and clarity of the visualization. Holmes not only produces beautiful visualizations that invite the user to gaze at them for longer, but they also effectively convey relevant information. As a result, literature has shown that such evocative imagery has assisted in recall of the information.

Visualization Description

The visualization begins with our title (Figure 1) and an introduction, which includes a description of our motivations and the methods used (Figure 2, Figure 3).

Figure 1. Our project title. Font family chosen was Bodoni, resembling the typographical aesthetic of fashion publications (mostly Vogue).



Fashion and color are both universal phenomena — one societal, and one (arguably) scientific.

While fashion trends are known to be erratic and seemingly unpredictable, the domain of color is finite in comparison. Despite a variation in culture, do different designer lines necessarily reach for disimilar color palettes depending on where they are from, where their work will be presented, or what season it is?

Join us in our *curiosity*.

Step through the seasons of Fashion Week, where the most prolific designers show their *colors*.

Figure 2. An introduction written in a voice common to fashion publications in magazines and in fashion blogs. Intended to set the tone and inform the user of the intent of the webpage, and ultimately the visualization below. It's accompanied by an image intended to emulate sketches found in fashion designer sketch books, with a woman bearing Pantone's colors of the year for timely relevance. Drawn by Dina.

Every Artist has a Method

From NowFashion.com, we scraped content (don't worry — we checked their Terms of Service to make sure we wouldn't get in trouble) from Ready To Wear designer lines from Spring/Summer 2012 through Fall/Winter 2016, for a total of 10 seasons.

We then filtered out the designers missing a season here or there, and found the most prevalent Fashion Week locations: Milan, New York, and Paris. If designers didn't present their work at these popular sites, we filtered them out, too.

Only the most productive designers made the cut, giving us over 500 images across 14 designers from 11 different countries. We made sure to clean the datasets of irrelevant or dirty images, and cropped each set separately from one another to focus on the bulk of the outfit at hand.

We implemented color quantization to each outfit for its top three colors via k-means clustering. Background color became a non-issue, as it took on a hue represented on the female figure. Skin color was also treated as an important trait, as models are relevant design decisions as well.

Once the images for a particular season's line were merged, they were then compressed and fashioned upon a lovely lady below. Feel free to filter our data in any way you like, and observe commonalities and differences across designers, seasons, designer citizenship, and fashion show location.

We hope you enjoy the fruits of our labor!

Your Data-Obsessed (Fashionably Dressed) Friends,
Dina & Natalie

Figure 3. A brief write-up of our methods, as to demystify how we created the final data represented in the designer thumbprints in the visualization below.

Users then move onto the visualization section (Figure 4). Each model's dress represents all of the colors from a single designer's ready to wear outfits that were shown at a fashion show for a particular year and season. The dress is the result of the quantified, cropped, and merged images that we scraped. Here, users have the opportunity to filter the visualization by designer, country of citizenship, fashion week city, season, or year. They can also toggle the filter menu for unobstructed viewing of the data.

The models are shown side by side to help users notice patterns and challenge assumptions they may have about designer color choice. Additionally, users have the option to scroll left and right for each designer to easily view all of the various years at once. When filtered, only the selected options will be shown, allowing the user to compare certain properties they are interested in.

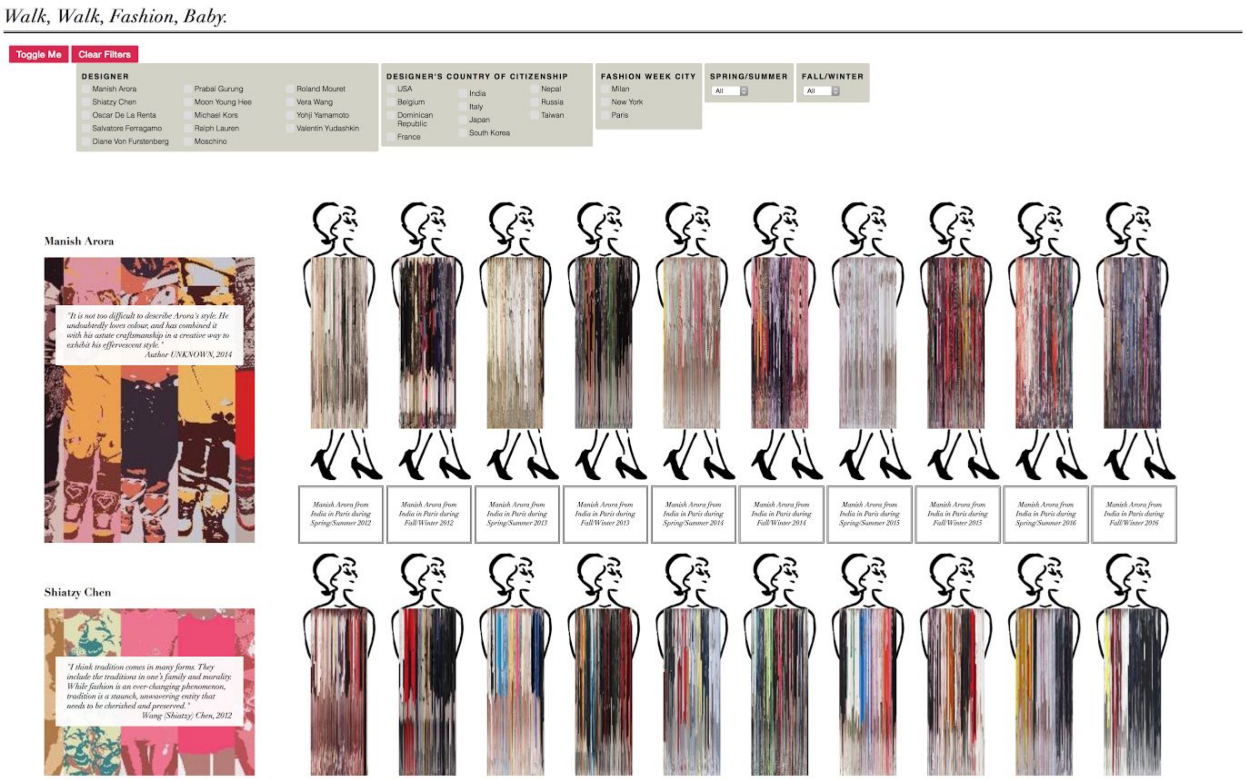


Figure 4. The interactive visualization, with filter menu visible. Screen captured on a 27" monitor viewed at fullscreen, so all seasons over the course of the 5 year period are visible.

Lastly, we have provided a quote corresponding to each designer. The purpose of this is to provide insight into the designer's motivations and ideally give the user a better understanding of the designer's color choices. The background of the quote is a selection of that designer's outfits that underwent color quantization. Remaining images can be seen in the Appendix.

Methods

A. Data and Preparation

Data was collected from nowfashion.com, a fashion-centric website that uploads images to its site in a live-fashion. We scraped images from designer houses with the most image data available. The scraping process involved several steps:

1. Scraped the link subject lines and URLs for all of the ready to wear lines.
2. Imported data into tableau, splitting up the information into the following columns: designer, season, year, URL.
3. Used EDA with tableau to discover the designers with a complete set of values - designers that had a URL value for each year and season (Spring/Summer 2012 - Fall/Winter 2016).

4. Exported a new dataset including only these complete designers and cleaned the data. This involved removing any extraneous items, such as links to videos or ads.
5. Ran the cleaned csv file of URLs through ImageScraper on terminal to download all of the relevant images.

Figure 5. An example of the raw data scraped from NowFashion.com. This specific frame was collected from Yohji Yamamoto's fall/winter line of 2012 in Paris.

After exploring the downloaded images, we decided to collect a diverse set of designers from different countries. Inevitably, we used a higher number of American and Italian designers because there were a higher proportion of both of these available. We also collected some that we personally connected with or felt that users would appreciate seeing in the visualization. The final list of designers used included: Diane Von Furstenberg, Manish Arora, Michael Kors, Moon Young Hee, Moschino, Oscar De La Renta, Prabal Gurung, Ralph Lauren, Roland Mouret, Salvatore Ferragamo, Shiatzy Chen, Valentin Yudashkin, Vera Wang, and Yohji Yamamoto. Selected designers originated from 11 different countries and participated in fashion shows in Milan, New York City, and Paris. These cities are where the majority of Fashion Week shows take place.

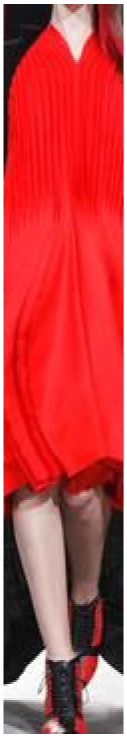


Figure 6. An example of the cropping of the aforementioned raw photo. All crops had a width of 60px, centered on the model from the neck down to the toes to minimize background noise/capture.

After we selected our raw dataset, we cropped each individual photo to center on the designer's torso such that the bounds of the image were no higher than their neck and down to her toes. Each crop took on a width of 60px. Height pixel length was variable +-25px depending on the standardized focus of the photographer's lens (how far they were from the models when they took a photo, for example). This cropping was done using the Python Image Library (PIL).

Figure 7. An example of a cropped image that underwent color quantization. Note the skin color dominates the color of the floor, and the shadow takes on the color of the dark fabric of the dress.

Once each set of photos was cropped, each individual photo then underwent color quantization via a k-means clustering method. In this algorithm, we selected for the three most dominant colors within the frame. We toggled for more clusters — first six, then four, finally settling on three as we noted that after three dominant colors, shadow was picked up on. Considering shadow to be an artifact of choice, and not an explicit design decision, we chose to mitigate that variable as much as possible with three clusters. A concern here was whether the background color, if it snuck into our crop, would impact the results of the clustering. Through much scrutiny, we found it to be a non-issue, as the background color would never be so dominant enough that it would dictate the color scheme of the image. In fact, we found that any background color would take on the hue represented in either the flesh tones or the fabric. This note brings up an additional point about skin color. We chose to be stingy in our crop by not cropping pixels below the knee or above the collarbone. Model skin color within the fashion world is a design choice that a designer chooses to diversify in their selection of models (or not!). Therefore, we found it relevant to the color schemes we hope to investigate in the completely prepped dataset.

At this point in our cleaned dataset, 6373 images were used for the following steps in our data-prep protocol. In other words, an average of 45 photos went into the compressed thumbprint of each designer, each season, over the course of five years, for a final total of 140 designer thumbprints.

Figure 8. An example of a portion of merged images before compression. Notice that the color of the floor — something that is consistent on the actual runway — takes on the dominate light tone reflected on the model, be it flesh or cloth.

Once the images were quantized, we merged and compressed them into one photo. This merging was done using Python Imaging Library. The Python script ran through all of the images for a particular designer, season and year, resized the image to be 3 x 300 pixels, and pasted them onto a new image that was around 100 x 300 pixels. The image was then resized to be exactly 100 x 300 pixels so all images were at precisely the same size.

Figure 9. An example of a final, compressed designer thumbprint, dressed upon a vector-image of a lady. This is representative of Diane Von Furstenberg's line of Fall/Winter 2012.

This process involved a lot of trial and error to decide what size was most appropriate. In addition, during this process, we noticed that some of the images were so diverse in color that it was difficult to notice any prominent colors. Thus, images were manually reorganized to group similar colors together to represent these dominant colors. Additional merging at a wider length was used to identify any prominent errors that could be identified when all the pictures were side by side. This method was also used to merge a few quantized images together to create the background image for the quotes.

B. Tools

- ImageScraper (https://pypi.python.org/pypi/ImageScraper)
- Tableau
- OpenCV, numpy, and matplotlib Python Libraries for k-means clustering for color quantization
- Python Imaging LIbrary (PIL)

- HTML, CSS, Javascript
- MixItUp (jQuery library)

Results

From the Data Itself

We came into this project with a few assumptions (some we found were globally had amongst users as well). That spring and summer seasons are dominated by brights and whites, and that fall and winter seasons are instead more bleak and dark in color regardless of country of citizenship. Due to the small sample size of designers from shared countries outside of the US, we could not adequately test hypotheses around cultural perspectives and social stereotypes. An additional hypothesis we had was that color schemes would be more similar if they were showcased at the same city location, as to appeal to the immediately surrounding audience (despite it being a broadcasted event). An additional curiosity we had was whether flesh tones within a designer's line would be consistent or varied. Our assumption rested with the former.

Our first assumption about the trends over seasons was indeed challenged. We found that, for the most part, designers did divulge in vibrant colors more often in spring and summer and darker colors in fall and winter; however, some designers we sampled had more stubborn color schemes that spanned throughout the years. A few examples are Salvatore Ferragamo (until 2015), Moon Young Hee (with blacks, navys, and whites being a persistent color scheme), Vera Wang (2014 and onward), and Yohji Yamamoto (with seldom splashes of color here and there).

Our second assumption surrounded Fashion Week city location color scheme consistency could really be adequately compared in Paris and New York only, as Milan was comprised of two designers. Despite Paris containing two of the four stubbornly-schemed designers above, we found Paris to exhibit a broader array of strong, vibrant colors across the seasons. While New York also has quite a bit of color showcased there, there seems to be a tendency to gravitate toward warmer, duller colors, or have but a few particularly vibrant outfits amongst the set. What this says about showcasing work within particular city locations at this time is unclear, and a larger sampling size is probably necessary.

Per the flesh tone question, we needed to keep in mind that accessories like knee-high boots (Salvatore Ferragamo really likes a distinct shoe style per season), socks (Moon Young Hee is of the very few designers represented in our selection to incorporate black knee high socks), and stockings (less frequently observed here) could morph our impression of what we believe may be flesh in the compressed images. So, we took special care to follow-up our observations with a look at the unmerged photos individually. Overall, we found a lack of diversity amongst lines. It was much easier to find the cases of an introduction to flesh-tone diversity because it was more rare. Moschino, Salvatore Ferragamo, Manish Arora and Roland Mouret were the examples we found.

While we did not really encourage investigating cultural or stereotypical curiosities due to our sample sizes outside of the United States, users themselves would utter their surprise. A broadly shared opinion people have of Russia is that it is a bleak place where everyone is supposedly depressed, and creativity or being outside of a norm is highly discouraged; however, upon viewing Valentin Yudashkin's work, people are amazed at how daringly vibrant the lines are, even across seasons. An additional utterance heard a handful of times was that it "made sense" that Roland Mouret would use many blues, whites, and reds (however irregular or alongside other hues), because he was French. Additionally, it "made sense" that Manish Arora, an Indian designer, expressed lines with a great deal of vibrancy. It was interesting to hear these thoughts out loud, and a future investigation of stereotypical schemas of various cultures could speak to their strength in our conceptions of fashion, something tied so closely to societal culture.

Usability Testing

Many users' initial impression involved confusion over which dress belonged to which designer or season. Because the labels were initially encased in a hover, it wasn't readily apparent which designer a particular dress represented. The hover initially completely obscured the designer thumbprint. This hover was burdensome for some in this way, as their cursor functions as the center of their gaze. As they look to the dress, they expect to still see the dress and not for it to disappear. Based on this feedback, the hovers were moved to appear below the dress so that the user could see both the label and the dress simultaneously. However, after further feedback, we found that it was still difficult for users to identify which dress belonged to which designer. Thus, we made this portion of our visualization persistent, in that you never needed a hover to see the information relevant to the designer's line that season. In addition, we added the designer's name at the header of each of the designer's respective sections. These "sections" were introduced after some user feedback, elaborated on in the next paragraph below (see figures 10-11).

We also received feedback about the inability to compare a variety of designer thumbprints because each image was so large. After much deliberation, we were not willing to compromise on the size of the designer thumbprints, as we felt thumbnails any smaller would erase the nuance of color represented in certain seasons (see Yohji Yamamoto's line, for example, which is predominantly bleak with some flourishes of color). Through some preliminary user-testing on resizing these thumbnails, we found our hunch to be verified. Based on other feedback, we decided to present each designer within one horizontally scrollable box, as opposed to on many rows on the page. This presented each designer in a separate segment, making it easier for viewers to compare designers explicitly and understand what information they were viewing. Initially, we thought this might present with some usability issues, because the entirety of a designer's work would not be present at any given time unless the pixel resolution of the monitor was high enough. Through some user testing, we discovered that this interface was preferred, as users enjoyed the physicality of scrolling "across time" (if needed) and the cognitive clarity of one designer's work per row. There is some superficial incompatibility on iOS devices, however, as the way we have it coded is through CSS WebKit tagging. iOS devices have presets to how scrollbars appear, and so while our code will function perfectly fine, the designer rows seem glaringly misaligned about four designer's down. This issue should not be too concerning (however much concerning it is), as the objective of this interface is to reasonably compare a few designers at once, as opposed to the entire set of designers (see figure 11).

Another functional and experiential piece of feedback we received centered around the filters themselves. Due to the flexibility of filtering we wanted to affords our users, we accumulated a breadth of different filters that take up space on the page. On screens with lower resolution, this is neither visually appealing nor practical for viewing the visualization. Therefore, we implemented through JavaScript a means to toggle the filter menu. This minimizes the amount of distraction people have as they view the content of the visualization, but keeps the menu within a reasonable and easy reach (see figure 4). Other bits of feedback we received about the filters was that people would prefer options represented in adjacent sections to be unhighlighted as they become unavailable. For example, if USA is selected as a filter, all non-USA citizen designers would recede into the background as to indicate that they would not produce compatible options. After initializing themselves with the filters, though, people quickly picked up on the way in which the filters were organized and intended for use. For this reason, we did not focus our efforts on correcting this problem that would require additional JavaScript to resolve.

Additionally, some users were confused about the general purpose of our visualization, thus we added a methodology section in the beginning to help users understanding the technology and purpose. We found that both introduction and methodology sections of our page helped to orient people who chose to glance through them, and piqued curiosity and interest in those who had glanced through them. We found users who traversed through these sections spent more time exploring through the visualization via the filters than those who had not glimpsed the sections. Additionally, people who visited these sections spent more time analyzing the designer color palettes in more varied ways than those who did not.

A change to our interface that helped clarify the methodology of color quantization for those who had not read our methods section was to enlarge the image behind the designer quotes. Previously, people saw them, but dismissed them as pretty colors and nothing deliberate or relevant to the analysis itself; however, once pointed out, people understood with an “Oh!” and stated they wished it was more apparent that the background image of the quotes represented images compiled within the thumbprints. Once we made the background images of the quotes large, both former and new users could tell with ease that the background images contained images processed from the runways (see figures 11 and 4 to compare).

We originally had a less obvious color scheme - a light gray that appeared white on certain monitors, despite playing with monitor color options. This occurred on the day of the final project showcase. Aesthetically, it made it seem like we had a white-space issue. But more importantly, it also made the filter menu options more difficult to see. We decided to darken these areas to make the contrast more obvious and the filters more easily identifiable (see figures 1 and 16, 4 and 17 to compare).

In asking our users if they were left wanting for something more that our visualization did not make available, the sole answer was the following: the option to filter by color. This request is a reasonable one, but also detracts from the deliberate role of delightful discovery on the user’s part. With this filter, all the work would be done for the user. We do think that there is some value in this suggestion, though, and so in a future iteration we believe it would be worthwhile to add this sort of functionality. With our current structure, it would require us to classify upfront each designer thumbprint with predominant colors.

Should we find an alternative way to automate this, it would be preferable, as we have almost 150 thumbprints currently.

We ascertained user feedback from ~20 individuals, ranging from students/peers, to people within the tech, business, and design industries. One person we surveyed was specifically from the fashion industry as a color analyst. Her feedback was particularly insightful should we choose to refine our target user group. As of now, our visualization is targeting the general population; however, should we choose to aim toward a niche within the fashion industry, it would be particularly useful to provide distinct color swatches such that fashion professionals may compare one designer's red to another's, for example. Such discrimination in the data was seen as too granular for this project, but the idea is a thought-provoking one that warrants additional consideration. We did attempt to look into this tactic briefly, but due to time constraints it fell low on our priority list given our target audience.

Additionally, to those who had seen the New York Time's Runway Fashion information visualization, we received feedback on incorporating an expansive view of the designer's whole line for that season upon hovering over a thumbprint. We noted that this feedback was never mentioned by those who had not witnessed the NYT information visualization. While they liked our visualization as it was, they stated that "it would be cool" to have the additional functionality of adding the raw images for a more detailed view of each outfit. Because the focus of this project was to focus on the dominant colors of each outfit -- and not the details of the craftsmanship of each outfit -- to define a designer's line, it would have been outside of the scope of our objective. However, no one can deny that it would be a bolstering property to add to our visualization.

Step	**Dina**	**Natalie**
Data Scraping	0%	100%
Image Cropping	100%	0%
Color Quantization	100%	0%
Image Merging	0%	100%
Alternative Color Analyses	0%	100%
Front End Development	60%	40%
Graphic Design	100%	0%
Scientific Literature Review	100%	0%
Pop-Culture Literature Review	0%	100%
Report	50%	50%

Final visualization can be seen here: http://groups.ischool.berkeley.edu/ColorMeCurious/

Appendix

Figure 10. The full-hover problem and the multi-designer per row problem.

Figure 11. The edit to represent the horizontal scrollbar, a fix to the multiple-designer-per-row problem discussed earlier.

Figure 12. Old filters section

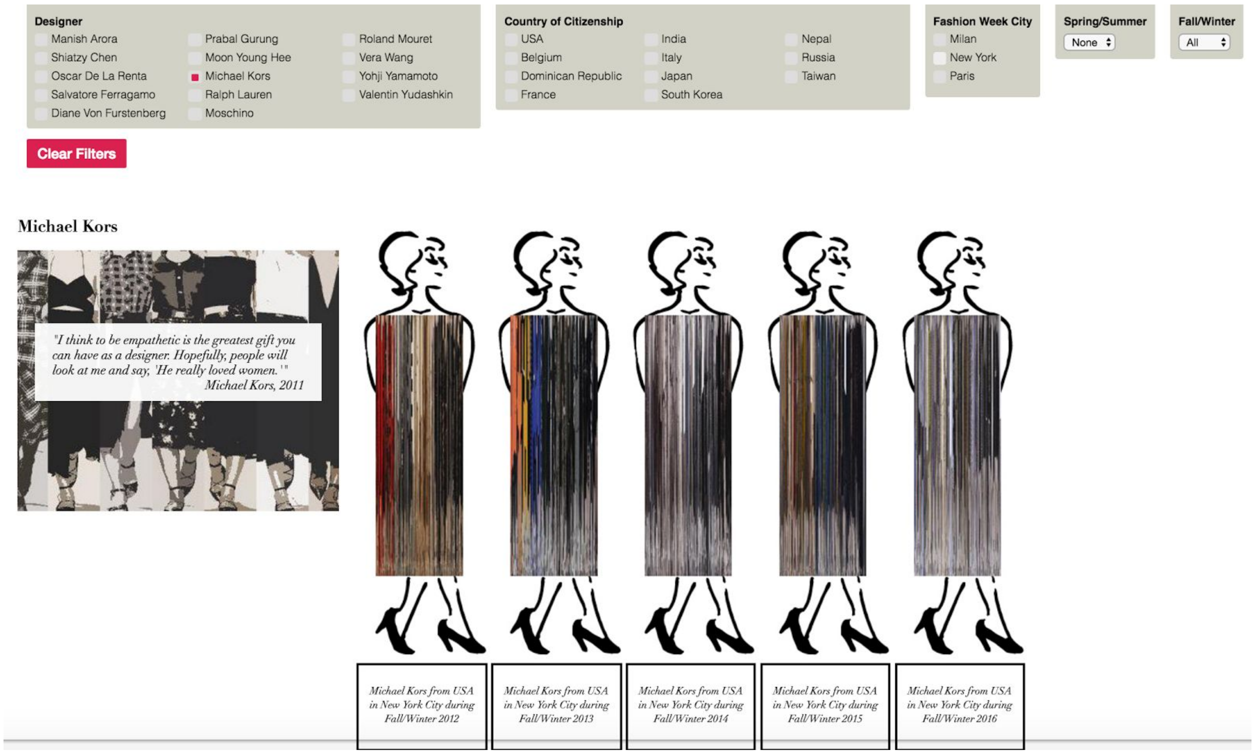


Figure 13. Visualization with checkboxes selected.

Figure 14. Scrollbar at the bottom allows users to scroll left and right to see all of one designer's dresses.

Figure 15. Sample of Valentin's images for Spring/Summer seasons only.

Figure 16. Old color scheme header.

Figure 17. Old filter menu with old color scheme..

Figure 18. Iterations of the walking lady.