Visualizing Large Volumes of Feedback for Designers

by

Caleb D. Heidel

May 2020

Presented to the

Division of Science, Information Arts and Technologies

University of Baltimore

In Partial Fulfillment

of the Requirements for the Degree of

Master of Science

Approved by: _____

[name, Thesis Advisor]

_____

[name, Committee Member]

Abstract

Reviewing feedback is an important part of the creative process, allowing designers to determine the effectiveness of their work and make critical decisions. This practice becomes challenging when feedback is collected in high volumes. Without a way to organize, visualize, and explore data, such feedback is unusable. This research presents the design for an interface which automatically categorizes large volumes of feedback through machine learning, generates an interactive visualization which progressively discloses data to users, and provides tools for navigating and manipulating feedback, allowing designers to uncover and record insights. Eight professional designers were interviewed to determine the functionality of this interface and later took part in user tests to judge its effectiveness. User tests confirmed that the automatic categorization, interactive visualizations, navigation features, and data manipulation tools were largely helpful for designers. Results also revealed areas for further development, some of which include empowering users to train the machine learning, evolving the visualizations to show how data overlaps and changes over time, and increasing transparency so that users know what information was organized by a human versus machine learning.

Acknowledgements

I would like express my appreciation to Professor Greg Walsh, who has served as my thesis advisor for the IDIA program. His teaching introduced me to the world of interaction design and showed me that this practice is not just about designing interfaces, it is ultimately about solving problems for people. His approach to design demonstrates that an effective creative process requires working alongside users to invent solutions. Greg's guidance over the past year is what enabled me to complete this research.

Additionally, I would like to thank the Johns Hopkins University Applied Physics Laboratory for making it possible for me to study at the University of Baltimore.

Contents

List of Tables

List of Figures

Introduction

Successfully designed products result from the collaboration of two key players: the creator and the user. Professional designers across industries require feedback from users to maintain quality and accuracy in their work. Rather than settle on an initial idea, the dialogue between these two sides enables designers to pursue a number of options in order to find the best solution (Xu, Huang, & Bailey, 2014). Working with user feedback educates creators on what they have made, how it is perceived by others, and where gaps exist in the intended design. Furthermore, insights from reviewers can loosen up any creative blocks a designer is experiencing, freeing them to make important or risky decisions (Luther, et al., 2015; Xu, Huang, & Bailey, 2014).

In order for designers to collaborate with users on large design efforts, efficient methods of gathering feedback are required. Typically, data can be collected through formal surveys or user tests. These trusted methods can provide highly valuable information, but they come at the cost of creating, facilitating, and maintaining the study materials. To get around these barriers, many businesses in and outside of the design industry, are turning to crowdsourcing to bring in large volumes of feedback relatively quickly (Gamon, Aue, Corston-Oliver, & Ringger, 2005, Luther, et al., 2015; Xu, Huang, & Bailey, 2014).

However, in order to make use of vast amounts of crowdsourced and data mined feedback, a serious design problem has to be solved. How can the complex information gathered be presented clearly such that designers can comprehend the breadth of responses and uncover insights? One way to address this challenge is by building an interactive digital tool which aggregates and visualizes data for designers. With such a technology, professionals would be able to efficiently review large datasets in order spot issues in their creations and brainstorm new solutions.

This literature review will engage the problem of visualizing large volumes of feedback to professional designers by exploring four topics. First, it will investigate issues related to organizing and presenting large amounts of unstructured data. Building on this, it will review the design requirements for making large datasets accessible and navigable. Furthermore, it will discuss previous interfaces created to present large quantities of feedback. Finally, this document will conclude by mapping out recommendations for a design feedback reviewing tool.

Literature Review

What's the Deal? Understanding the Barriers to Presenting Design Feedback

When collecting feedback on a design, not every type of information is equally useful. Quantitative data, from a Likert or number scale, is easy to understand and synthesize. However, because of its simplicity, it is limited in its richness and cannot fully describe why a design does or does not meet a particular standard (Soriano & Palaoag, 2018). When customers are asked to rate an element quantitatively, their simple responses may lack complexity or nuance. For instance, a user could give an overall high score to a product after viewing multiple positive elements and one negative feature. The resulting favorable grade masks important details about the single flaw that the person observed. Designers can learn the most when they are able to gather qualitative text-based entries which describe each reviewer's detailed opinion. For this reason, companies are increasingly exploring the use of data mining to compile opinions from large volumes of text-based reviews (Oelke, et al., 2009).

When collecting text-based feedback on a product, information can be gathered in a structured form, making it easier to manage. This is done by guiding the user through a series of constructed questions in the form of a detailed survey or interview (Gamon, et al., 2005). Analyzing and running statistics on structured data is a much easier (Ziegler, Skubacz, & Viermetz, 2008). In some design feedback systems, special techniques are utilized to structure the data so that designers can uncover key insights at a faster rate (Luther, et al., 2015; Xu, Huang, & Bailey, 2014).

However, there are drawbacks to this type of data which can be avoided. Plenty of labor may be required to construct and administer a structured survey or interview. Furthermore, the methods used to collect structured feedback risk limiting the open communication of customers, potentially degrading the data's expressiveness. This leads to the additional possibility of not identifying important trends because customers hold back from communicating certain opinions. Finally, because some users will not have the patience to fill out a detailed survey, there is further risk of missing entire portions of a target audience. To avoid these difficulties, designers can utilize unstructured data, which avoids the need for highly choreographed studies. This type of information can be collected as text from all sorts of sources, such as online customer reviews, user tests, blog posts, and more (Gamon, et al., 2005). Data of this sort is far richer, more

expressive, and more detailed. Unlike quantitative and structured approaches, customers can freely name and describe what they like and dislike (Ziegler, et al., 2008).

The opinions contained in a collection of unstructured data is highly valuable to designers. However, without proper processing tools, this information is useless. Manually combing through unstructured data is highly labor intensive (Ziegler, et al., 2008). This is due to the fact that this data, as its name would suggest, is complex, unordered, and abstract (Gretarsson, et al., 2012). In addition, unstructured responses risk being inconsistent in a number of ways. The lengths of written reviews are often uncontrolled and the writing quality may vary from skilled to poorly constructed. The tones of responses might range from emotional to objective and some entries could be extremely detailed while others are rather unspecific (Apler, Yang, Haber & Kandogan, 2011). Furthermore, it is not uncommon for reviews to be contradictory. For example, in one review, a customer could state that they find a mobile application very helpful overall *and* that one of its features is hard to use. It requires an extra level of sophistication for a digital tool to effectively record and separate out those opposing opinions (Apler, et al., 2011).

The valuable insights contained in unstructured data is useless unless if it is made accessible to designers. So, users need some way to get around the time-consuming task of sorting feedback. This is because combing through large sets of data requires information to be explored in a parallel fashion, tracking opinions in context of other expressed sentiments. Furthermore, while sorting through piles of data, time may be wasted reviewing responses which are not usable or relevant (Apler, et al., 2011; Ziegler, et al., 2008).

To make use of rich unstructured data, machine learning must be utilized to give structure to the information, making it easier to visualize. Equipped with this technology, designers can efficiently cluster data and extract keywords from each grouping in order to spot themes. Utilizing machine learning to synthesize these datasets allows the discovery of general areas of concern and reoccurring opinions (Gamon, et al., 2005).

However, machine learning must be used as a tool, not as a replacement for human discernment. The issues related to unstructured data cannot be solved entirely by machine learning. The result of automatically mining out topics from chaotic collections of data will not be perfect (Apler, et al., 2011; Hu, et al., 2013). Machine learning cannot know everything about

a reviewer's opinion or judge an author's certainty. Therefore, it is possible that portions of information will be misinterpreted or organized incorrectly (Wu, et al., 2010).

Successfully organizing rich unstructured data still leaves a core issue unresolved, how to present the data visually and enable interactions for exploration and analysis. This is problematic because large unstructured data is overwhelming to users in its raw form. Interfaces which present feedback need to make the data comprehensible and accessible to users. However, in many cases, past efforts to visualize this sort of information have erred by including too much data. These designs are overly detailed and complex, confounding users. In other scenarios, they are too simple, limiting the amount of information available and restricting free exploration and analysis (Apler, et al., 2011).

How to Start Solving the Problem? Design Requirements for Visualizing Unstructured Design Feedback

As stated earlier, machine learning is required to organize large amounts of unstructured data. However, categorizing large datasets must be done carefully due to the multifaceted nature of the information. When gathering feedback, each response can be seen as the culmination of a mixture of topics, making it possible to group entire datasets in many ways (Soriano & Palaoag, 2018). Some methods of categorization will be more useful to designers than others. Therefore, measures must be taken in order to ensure that information is optimally grouped.

One way to optimize categorization is by limiting the number of topics that are generated (Käki, 2005). In a study that tested techniques for sorting over 800,000 text documents, it was found that only 20% of the initial set of categories were necessary for users to find information (Wetzker, et al., 2008). Similarly, research that sought to optimize the categorization of search results discovered that only 10-20 categories were required for use, rather than an original set of 40 topics. These studies show that less is more when it comes to generating categories from a set of information, assuming there is adequate coverage of the content. This is for two reasons: fewer topics will improve computational efficiency and provide reliable buckets of information for users without hurting the accuracy of the categories (Käki, 2005; Wetzker, et al., 2008).

Unstructured data can be made more accessible to users by designing easy navigation. In general, moving through complex data can be risky for users. An abrupt shift from one area of information to the next could result in a user being disoriented or not realizing they have moved at all. Furthermore, without some sort of reference point, users could lose track of where they are

in the data structure while navigating. Two systems which attempt to deal with these potential risks, *PivotPaths* and *Hierarchie*, provide examples of navigation features which make the exploration of data seamless (Dork, Riche, Ramos, & Dumais, 2012; Smith, Hawes, & Myers, 2014). Together, these tactics give users context for the information they are viewing and the confidence to explore freely.

The first of these examples, *PivotPaths*, enables casual exploration through large collections of documents by implementing smooth transitions from one view to the next (see Figure 1). Less abrupt movements help users to not lose track of where they are in the data structure. When operating *PivotPaths*, users narrow down results on screen by clicking on a piece of information. When this action is executed, three events happen in parallel: information currently displayed that is relevant to the user's selection remains on screen, new information fades into view, and unrelated data dissolves away. While these smooths transitions take place, the resulting information from the user selection smoothly rearranges into a new layout. These navigation features further the philosophy behind *PivotPaths*, that navigation should be cheap for users, not risky (Dork, et al., 2012).
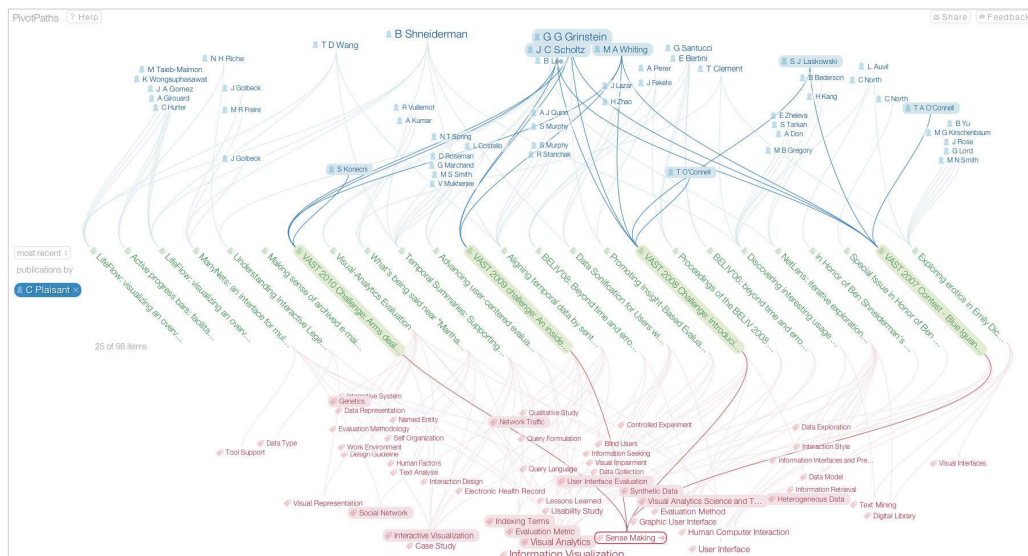

Figure 1. The PivotPaths interface features nodes connected by a network of curving lines. Interaction with one node highlights related nodes. Clicking on a node causes the entire page to reshuffle with relevant information (Dork, et al., 2012).

Hierarchie, which visualizes a hierarchy of topics on a layered radial diagram, enables easy navigation by implementing two forms of breadcrumbs (see Figure 2). First, a horizontal, inline graphic displays the layers of topics that the user has travelled through. Second, a

simplified thumbnail image of the radial diagram highlights the area of the data structure that the user is currently exploring (Smith, et al., 2014).
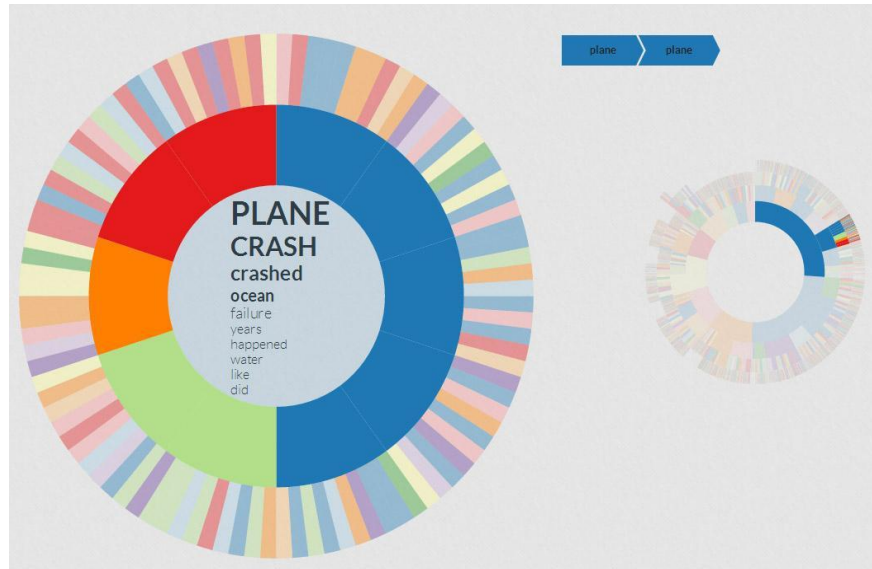
Figure 2. The Hierarchie interface features one central interactive graphic with accompanying breadcrumbs and thumbnail sized map (Dork, et al., 2012)

Large volumes of unstructured feedback are also made accessible by enabling features which allow users to address the imperfections of automation. This technology, although necessary to organize data, cannot perfectly categorize information or intuit the intent of the author. Therefore, a number of capabilities must be present to overcome these barriers. First, the user should have control of the parameters which dictate the number of categories created. From here, users should be able to manipulate the information shown on screen by editing category names, creating, merging, and removing topics, altering the metadata associated with a response, and moving data into new categories (Hu, et al., 2013; Ziegler, et al., 2008). This sort of control allows users to create their own meaning and correct for any mistakes found in the automated structure.

At every level of detail, a visual system designed to present large amounts of feedback should serve to increase functionality rather than create a barrier between the data and the user (Dork, et al., 2012). This can be done by condensing the raw data into buckets and then progressively disclosing information to users as they interact with those containers. When users first view the feedback, some type of labelled graphic identifier is required to denote each category of responses. This could take the form of a keyword or short text snippet attached to a

container or node (Gretarsson, et al., 2012). When content is contained in this way, the user is presented with a visual summary of the information, which is far less overwhelming than viewing the raw data (Apler, et al., 2011). A designer can use this perspective to locate portions of responses that are particularly interesting or useful. By interacting with a category of interest, either by clicking or hovering, users can navigate to a new layout which displays all of the content relating to that selection (Ziegler, et al., 2008). Visualization in this stepped approach allows designers to explore content on both global and granular levels.

Treemaps, when equipped with interactive navigation features, serve as effective visual frameworks for viewing information at both large and small scales. Constructed with dynamically sized rectangles, these infographics are able to efficiently cluster information into categories, display relative volume of data through scale, and communicate general sentiment with color coding. Simple interactions can be built into treemaps to enable access to all levels of information, which facilitates analysis of hierarchies of data (Ziegler, et al., 2008). With a set of information plugged into a treemap structure, users are able to spot patterns in the data with ease (Carvalho & Chaves, 2012).

Zooming in a bit, the microscopic details making up the visual system also serve as mechanisms to reinforce meaning and functionality. These finer points concern the styling of the interface's building blocks, including text, buttons, links, containers, and other elements. The systematic use of a particular size, color, brightness, texture, and shape for each item reinforces the meaning of the visualized data to users (Carvalho & Chaves, 2012).

Mentioned above, *PivotPaths* is one platform which uses deliberate and sophisticated styling of its visual elements to communicate the meaning of data. At a basic level, color is used to classify data nodes as a person, publication, or concept. *PivotPaths* then goes a step further by using color and curving paths to indicate how each node is related to the other. When a user hovers over a node, all related items and their pathways increase in darkness and saturation. This systematic and dynamic styling gives users a sense for what information is on screen and how it is all related (Dork, et al., 2012).

Where Are We Now? Exploring Past Efforts to Visualize Design Feedback

Voyant and *CrowdCrit* are two products built specifically to help designers collect and visualize feedback (see Figure 3; Figure 4). Both systems rely on crowdsourcing workers to generate large volumes of responses relatively quickly. To ensure that recruited workers provide

high quality feedback, each platform has reviewers complete a set of microtasks, where they respond to prompts about the design. In *Voyant*, users are shown an image of the design and then, depending on the nature of the prompt, respond by providing text, selecting a radio button, choosing a word from a list of options, or drawing a box around a detail in an image (see Figure 5). *CrowdCrit* goes a step further by implementing scaffolding in the form of language suggestions. As users draft responses, the system offers pre-authored statements to help participants formulate their opinions. When the microtasks are completed, the responses are aggregated behind the scenes into neat categories without the aid of machine learning (Luther, et al., 2015; Xu, Huang, & Bailey, 2014).
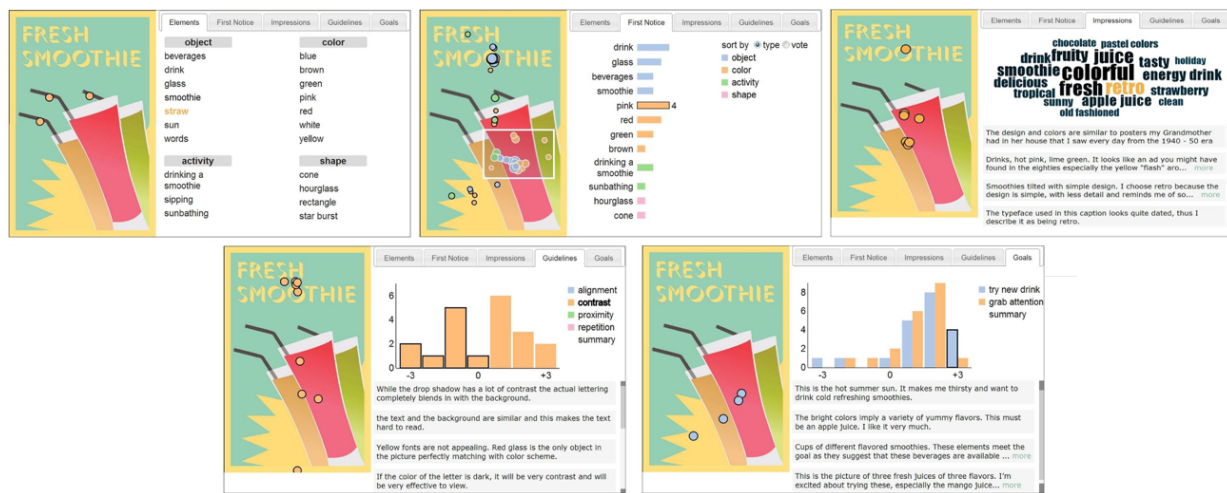

Figure 3. The Voyant interface visualizes feedback across five categories, separated by tabs (Xu, Huang, & Bailey, 2014).
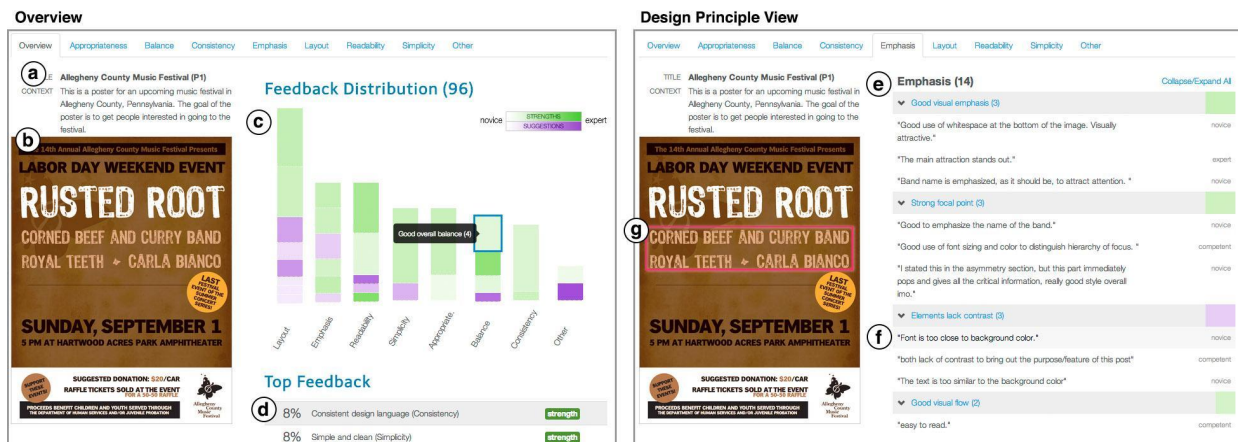

Figure 4. The CrowdCrit interface visualizes feedback across nine tabs (Luther, et al., 2015).

Figure 5. A snapshot of the design reviewing workflow for Voyant, where crowdsourced workers complete microtasks by responding to prompts. Afterwards, all answers are aggregated into a full report for the designer (Xu, Huang, & Bailey, 2014).

The information collected through microtasks is used to automatically populate an interface with all data separated into predefined categories. Users access each topic by selecting different tabs in the global navigation. *Voyant* categorizes data into five groups: what elements are seen, which details are noticed first, overall impressions, how well the design meets predetermined design guidelines, and how effectively it fulfills the designer's specified goals. *CrowdCrit* divides the feedback across nine tabs: *overview*, *appropriateness*, *balance*, *consistency*, *emphasis*, *layout*, *readability*, *simplicity*, and *other*. Each topic includes an interactive infographic depicting an overview of associated responses. Typically, these take the form of bar charts, but other methods are also used such as term lists and word clouds. Users must click through each tab and comb through individual responses in order to get a sense for the breadth of feedback (Luther, et al., 2015; Xu, Huang, & Bailey, 2014).

Voyant and *CrowdCrit* have a number of strengths when it comes to visualizing large volumes of data. Both are able to ensure relatively high-quality responses by utilizing microtasks in the data collection process. Configuration tools allow professionals to state their design goals and request responses from particular audiences. Once data is collected, both systems are able to present feedback in an organized manner. With interactions built-in to enable data exploration, designers have the ability to discover key insights by sifting through responses (Luther, et al., 2015; Xu, Huang, & Bailey, 2014).

However, the extent to which *Voyant* and *CrowdCrit* rely on structured data does not make them ideal tools for professionals. The use of microtasks, rather than machine learning, to organize responses limits the volume and variety of data that can be collected. Reviewers are guided in the way they respond, especially in *CrowdCrit*, where workers are given suggested snippets of text to use in their written responses. Furthermore, information is always sorted into the same static predefined groups rather than unique topics that reflect the nature of the responses. Finally, once information is presented to the designer, it cannot be edited in any way. Therefore, these platforms are largely inflexible and limit a designer's ability to organize, manipulate, and explore the information as they see fit (Luther, et al., 2015; Xu, Huang, & Bailey, 2014).

Another concern relates to the overall quality of the feedback collected. After testing *CrowdCrit*, concerns were raised that the aggregated responses did not result in designers making significant revisions. Rather, it was noted that most changes are minor in nature. It is possible that when using *CrowdCrit*, designers could be led to think that they are making large scale edits to their products. However, in reality, they are only making small adjustments which do not address potential large-scale issues (Luther, et al., 2015).

Although *Voyant* and *CrowdCrit* can consistently provide useful feedback to designers, they are largely inflexible. Data is organized and presented in a one-size-fits-all method for every product. As a result, designers are given little freedom to manipulate and organize information as they see fit in order to discover insights.

Ironically, features which would be most useful when reviewing feedback at scale are found in products not built for designers. Rather, they exist in technologies meant to help businesses analyze customer feedback on various products in order to respond to issues in a timely manner. The *Pulse* and *OpinionBlocks* systems, as well as a product developed by researchers at Heinrich-Heine-Universität, are such examples. Compared to *Voyant* and *CrowdCrit*, these systems do a better job at enabling free analysis, manipulation, and exploration of information (Gamon, et al., 2005; Hu, et al., 2013; Ziegler, et al., 2008).

The first strength in these systems is found in the type of information they use. Each platform collects feedback by mining large volumes of unstructured text-based data from online product reviews. To make sense of this information, machine learning is implemented to categorize the responses into meaningful topics (Gamon, et al., 2005; Hu, et al., 2013; Ziegler, et

al., 2008). As a result, the categories are unique, reflecting the nature of the product being reviewed and the collected responses. This removes any need for microtasks to organize the data.

To present feedback, each of these systems uses a single interactive visualization. Rather than separate all data with tabs, the information is contained within one visualization of the top-level topics, providing a summary view of all data. To accomplish this, all three systems implement some form of block layout to visualize the categories. *OpinionBlocks* populates a container for each topic (see Figure 6). These are each labelled by a keyword which summarizes the content. Within the container, a word cloud is displayed to describe the content in more detail. Below the word cloud, rows of tiny squares depict each response. Color coding of the words and squares communicates whether a piece of data is positive, neutral, or negative. The other two systems implement flexible treemaps which arrange topics into more dynamic block layouts, each labelled by a keyword (see Figure 7; Figure 8). In these visualizations, the scale and color of each block depicts both volume and average sentiment of responses (Gamon, et al., 2005; Hu, et al., 2013; Ziegler, et al., 2008).

Figure 6. The OpinionBlocks interface. Top Left: Hovering over a keyword reveals which colored comment blocks contain the selected phrase. Top Right: Clicking on a block reveals a snippet of text and selecting the plus icon reveals the entire comment. Bottom Left: Comments can be dragged into other colored rows to change their sentiment rating. Bottom Right: Titles of categories can be changed by clicking them to edit (Hu, et al., 2013).
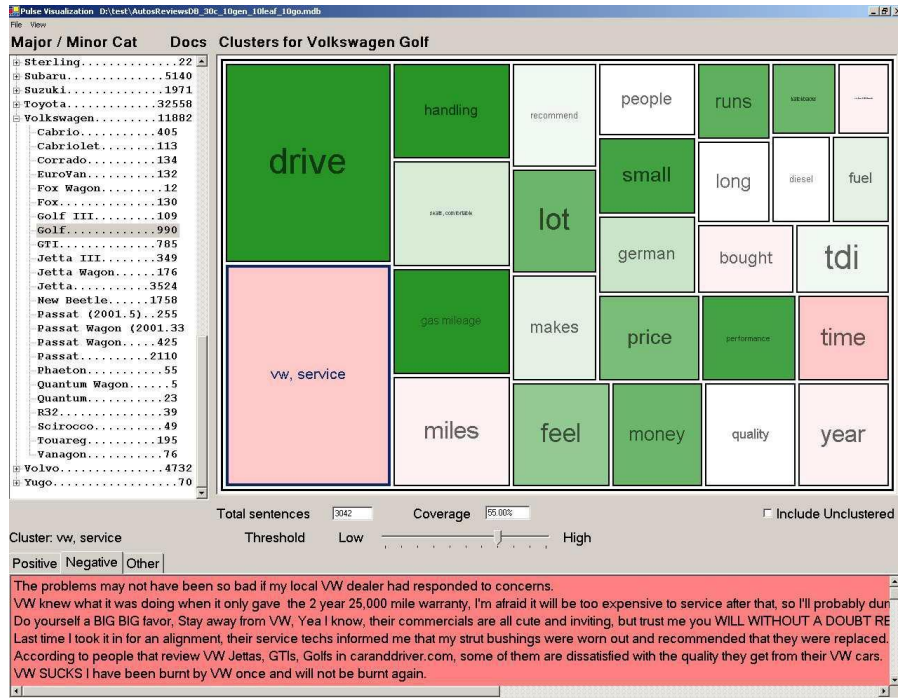
Figure 7. The Pulse interface uses a treemap to represent categories of data. Selecting a category reveals its details in the bottom container (Gamon, et al., 2005).
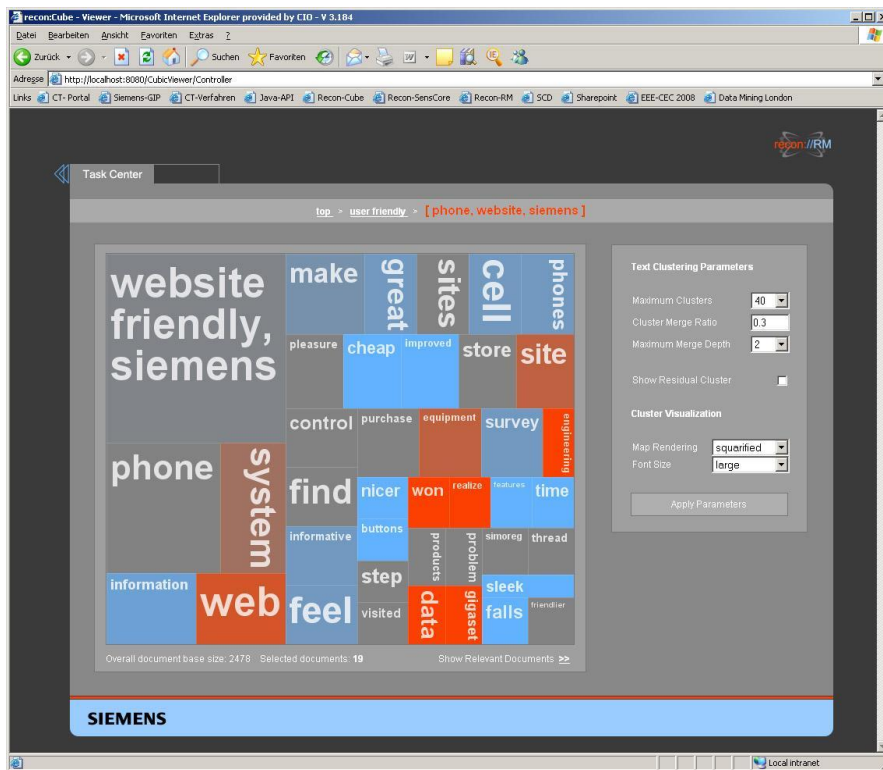

Figure 8. An interface using a treemap visualization to represent categories. Settings are displayed in the right column (Ziegler, et al., 2008).

All three systems are equipped with interactive features which progressively disclose information to users, such that they can explore feedback without being overwhelmed. Initially, only keywords or phrases are displayed in each category. To dive deeper, users must click on or hover over a topic to expose more information. In *OpinionBlocks*, hovering over any item in a word cloud will highlight the colored squares containing the selected text. When users hover over a square, the original text from that response appears on the screen. The other two systems, which utilize treemaps, function differently from each other. Clicking on a topic block in *Pulse* reveals all associated comments on the bottom of the screen and filters them by sentiment. The other treemap solution progressively refines the infographic as you click on topic blocks. For instance, clicking on one category will manipulate the treemap to only show the sub topics belonging to the user's selection. Clicking on a sub topic will reveal yet another level of grouped information, provided there is more data contained within. This feature allows the user to drill down until only the comments relating to their area of interest are displayed (Gamon, et al., 2005; Hu, et al., 2013; Ziegler, et al., 2008).

These systems also include interactive features that enable users to address the imperfections of data mining. Between these three systems, there exists basic editing features which allow users to make adjustments by changing topic names, moving content into different groups, merging categories, or changing the sentiment rating of a comment. If there is too much unnecessary information displayed, users can hide or remove data which is not needed, or adjust settings to reduce the number of categories created (Gamon, et al., 2005; Hu, et al., 2013; Ziegler, et al., 2008). In essence, these systems first build out a basic structure for the user to absorb the data. Then, they enable the user to explore and manipulate data as needed in order to investigate areas of interest and uncover key insights.

Conclusion: Putting It All Together

Numerous products exist which provide novel features for organizing and interacting with large volumes of text information. However, none provides an environment where designers can quickly view heaps of unstructured design feedback and perform their own analysis by freely exploring and manipulating data. *CrowdCrit* and *Voyant*, two systems built specifically for designers, limit users by organizing responses into static categories. Oddly enough, many of the features which would greatly help designers explore and analyze feedback are found in platforms designed for other purposes, such as *Pulse*, *OpinionBlocks*, and *PivotPaths*. Analysis of these

systems, along with other research, reveals the features required for a design feedback tool that enables open exploration and analysis of unstructured data.

A design feedback tool should be equipped to address the inefficiencies of machine learning. This technology needs to be optimized to generate enough categories to cover the information provided while not overwhelming users with too many options. Users should have the ability to adjust the number of categories after being served an initial set of topics. This way, if the automated categories are too broad or narrow, designers can quickly adjust the visualization for their purposes.

Feedback must be presented with a visualization which is meaningful to users and functions to enable exploration and analysis of data. The information should be grouped into categories and visualized through a dynamic interactive infographic, such as a treemap. In this framework, volume and sentiment can be easily communicated through the scale and color of each box.

To grapple with the complexity of large datasets, feedback must be progressively disclosed to designers through interaction. When users select a category, information within that topic should be revealed. Depending on the amount of information in a group, this interaction will unveil the original written responses or a new set of sub topics relating to their selection. While navigating through various levels of feedback, transitions to new information should be smooth and breadcrumbs are required to give users context for where they are in the dataset. Rather than being overwhelmed by information, these features will allow users to understand the full breadth of responses and explore its finer points with ease.

A tool for designers should do more than display information, it should enable users to perform analysis by manipulating data as they see fit. While reviewing information, users should be able to change category names, merge topics, rearrange information, or remove data. If a piece of information is grouped incorrectly, users should be empowered to move it into another area, create an entirely new group with it, or simply delete it. Similarly, users should be able to make any necessary changes to the metadata associated with a response, such as its sentiment value. For instance, if a complimentary review is labelled *neutral*, a designer should be able mark it as *positive*.

By implementing these features, a flexible platform can be created to help designers uncover insights from large volumes of feedback. Aided by machine learning and empowered by

a range of editing features, designers can conduct the analysis they feel is necessary for the scope of their project. As a result, large sets of data rich unstructured feedback can be harnessed and used efficiently in the creative process.

Methods

Participants

To conduct my research, eight professionals from the design industry volunteered to take part in two studies. The participants included graphic designers, user experience designers, a design strategist, and design thinker. The commonality among these individuals was their reliance on user feedback in their design processes.

Each person responded to an email requesting their participation in a study. The subjects consisted of eight young adult professionals in their mid 20s and 30s. Five participants were designers at The Johns Hopkins Applied Physics Laboratory (JHUAPL) in Laurel, Maryland and the other three worked as designers at Capital One in McLean, Virginia (see Table 1).

Table 1

The gender, age, job title, and employer of all participants.

Participant	Gender	Age	Job title
1	Female	31	User Interface Designer & Illustrator
2	Female	27	User Experience Design & User Research
3	Male	27	User Experience Design & Front-End Development
4	Female	27	Design Thinking & Graphic Design
5	Female	27	User Experience Design & Front-End Development
6	Female	30	User Experience Design
7	Male	36	User Experience Design & User Research
8	Female	25	Service Design & Design Strategy

The nature of these studies related very closely to the participants' professional work. From the recruitment email, subjects understood that the goal of this research was to develop a tool which would help designers utilize large volumes of feedback. Therefore, these participants had an interest in the subject matter and in the result of this research.

Additionally, each participant was compensated for their time. Those who participated in the first study received a $10 Starbucks gift card. Participants in the second study received a $15 Amazon gift certificate.

Materials

In both studies, a variety of materials were used to conduct sessions with participants. Audio was recorded with a smartphone recorder. A laptop was used to show wireframes and

interactive prototypes to participants. When interviews or user tests took place remotely, *Zoom* conferencing software was used to connect both parties over a call.

Other materials were used depending on the type of study being conducted. For interviews, *InDesign* was used to display wireframes. During user tests, *Axure* was utilized to provide a functioning prototype to the subjects. When user tests took place remotely, the prototype was shared with the participants through *Axure Cloud*. *Zoom's* screen share capability enabled remote users to share their screen with the researcher as they performed tasks. User tests of the prototype were recorded with the *QuickTime* screen capture functionality.

Four software packages were utilized to perform analysis on interviews and user tests. To review user tests, *Adobe Premiere* was used to view screen captures and recorded audio in parallel. For all studies, the audio was transcribed in *Rev.ai* and then analyzed in the *Text Analysis Markup System (TAMS)* and *Microsoft Excel*.

Design

This effort began with an initial phase of formative research to gather information about the value of feedback in the creative process and how it is used by designers. To do this, an interview was designed to include nine general questions to inspire discussion, an activity where subjects explain how they would review a hypothetical set of data, and a review session where the participants provide verbal feedback on wireframes of a potential feedback reviewing tool (see Table 2).

Table 2

Formative research interview questions. Questions are also available in Appendix A.

#	Question
1	What area(s) of design do you work in?
2	In what way(s) is user feedback important to your design process?
3	In what quantity and how often is feedback used in your design process?
4	Do you have an end goal when reviewing feedback on a design? Are there certain key insights you are looking to uncover?
5	When reviewing user feedback, what strategies/processes do you use to organize and sort responses?
6	What problems do you face when reviewing feedback?

7	How do you address the problem of analyzing feedback which is ambiguous, unconstructive, or contradictory?
8	Have you ever had to review large amounts of feedback, in the 100s of 1000s? In that scenario, how did you make that amount of feedback manageable?
9	How could a digital tool help you in making feedback review easier and more productive? Are there any tools you use now?
Activity Question	Assume you are reviewing the design of the homepage of a website. You have collected 1,500 individual text comments on this page. Could you walk me through how you might go about organizing that data and discovering key insights about the design?

The design of these wireframes featured two visualizations on the main page (see Figure 9). First, inspired by the feedback reviewing platforms discussed before, an interactive treemap was provided which used boxes of varying sizes to represent the volume of each feedback category. Although not represented in these wireframes, it was intended that the boxes would use coloring to communicate the average sentiment of responses in each category. To provide flexibility to users, a barchart visualization was also provided, displaying each category as a segmented line. The length of each line represented the volume of responses. The segments in each line indicated the breakdown of positive, neutral, and negative responses. Users would be able to adjust the main page visualizations by using filter options on the left side of the screen.

The subpages included a sticky note visualization, inspired by *OpinionBlocks*. Here, the responses were represented as boxes arranged into subgroups. The varying shades of gray were meant to represent color-coding that would indicate the sentiment of each piece of feedback. Hovering over a comment would preview its contents on the top right of the screen. Clicking on a sticky note would reveal its contents in a popup window. The information on screen could be manipulated by clicking and dragging sticky notes into other subgroups or top-level topics in the side navigation. To navigate out of a subpage, users could click the home button or select another category from the side navigation.

Figure 9. Wireframes created in InDesign that were shown during formative research interviews.

Top: Main page treemap visualization, displaying all categories of feedback. Middle: Main page

barchart visualization. Bottom: subpage sticky note visualization, showing one category of feedback. All wireframes from this design are featured in Appendix B.

The audio from each interview was recorded and transcribed for analysis. In *TAMS*, the dialogue was tagged by topic in order to track trends among participants. Topics were ranked by frequency of mention. As a result, a tiered list of topics, ranked by priority, was created to inform which wireframe features should be kept, developed, or discarded.

A second stage of research was designed to test a functioning prototype, the design of which was informed by findings from the interviews (see Figure 10). The design of this interface is discussed in the *Results* section. The prototype featured real user feedback from wireframes of *Coupon Buddy*, a mobile application created to help users store and redeem coupons (See Appendix D). 27 tasks were formulated for the user test with the intention of mimicking actions a designer would take to review feedback (see Table 3). By completing a user test session, each participant would be required to visualize the data in a number of ways, navigate through the prototype, create action items, flag comments, hide and filter information, and edit and rearrange feedback. Accompanying the user test was an exit interview with eleven questions written to assess key features of the prototype (see Table 4).
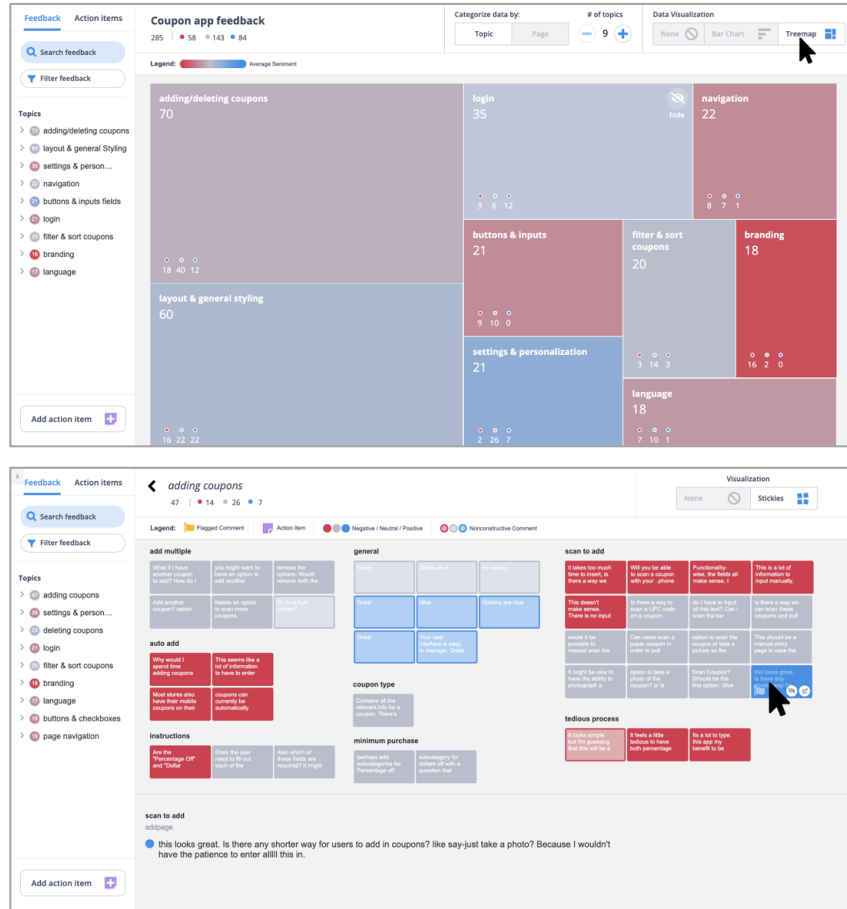
Figure 10. Interactive prototype built in Axure. All screens from this design are featured in Appendix C.

Table 3

User test tasks. Tasks can be found in Appendix E.

Task	Description
1	This is the landing page of the tool, providing an overview of the data. Without clicking, take a moment to view the page and describe what you see.
2	One of your other team members has already used this prototype to review this feedback. Find the notes he created about the feedback.
3	Without clicking, review this page and describe what you see. Then return to the previous page.
4	Find a way to view this data differently.
5	Looking at this data, you find that it's a bit too simplistic. You'd like to see this feedback arranged into more groupings. Find a way to do this.
6	Your team doesn't have the bandwidth currently to address issues concerning the layout, menu, or inputs. Find a way to remove feedback relating to these things from your view.

7   Now, locate the topic containing feedback about "scanning to upload" coupons.

8   Without clicking, take a moment to view the page and describe what you see.

9   This is too much feedback to review efficiently. We need to distill this feedback down a bit. Find a way to quickly remove all feedback which you perceive to be overly positive or otherwise unhelpful.

10   Create an action item addressing an important piece of feedback you see. And mark any comments which you would like to review later.

11   Now, find the topic containing feedback relating to the Coupon Buddy logo.

12   Check to see if the subgroup titles appropriately describe the comments. If not, correct the titles.

13   Review the comments. Check for any misspellings. And, confirm that the sentiment rating matches the opinions expressed in the comments. If you find any errors, correct them.

14   Create an action item addressing an important piece of feedback you see and mark any comments which you would like to review later.

15   Now, locate the topic containing feedback about relating to checkboxes.

16   Find a way to view this data differently.

17   Machine learning has done a less than perfect job organizing these comments. Find a way to have the comments in the "general" subgroup be in more appropriate subgroups.

18   Create an action item addressing an important piece of feedback you see. And mark any comments which you would like to review later.

19   Now, locate the topic containing feedback about relating to the back button.

20   Visualize the data and look for responses which don't belong in this topic. Find a way to store them in more appropriate topics.

21   Review the remaining feedback and confirm that they are in appropriate subgroups. If not, rearrange them as you see fit.

22   Create an action item addressing an important piece of feedback you see. And mark any comments which you would like to review later. After - return to the initial landing page.

23   You'd like to find feedback relating to "color coding" but you're not sure what topic it belongs in. Query the data to find this information.

24   Create an action item addressing an important piece of feedback you see. And mark any comments which you would like to review later. After - return to the initial landing page.

25   Find a quick way to determine which page has the most negative feedback. When you find this information, create an action item noting this page.

26   Your next meeting with the design team is about to start. Send your action items to Geoff, as he'll be running the meeting!

27   Before you go, confirm that the comments you flagged are stored for you to review later.

Table 4

User test exit interview questions. Questions can also be found in Appendix F.

#	Question
1	Did the auto generated categories improve your ability to review feedback? Why?
2	Did the data visualizations improve your ability to review feedback?
3	Was it easy to navigate through the feedback? Why?
4	Did the search functionality improve your ability to review the feedback? Why?
5	Did the editing functionality improve your ability to review the feedback? Why?
6	Did the ability move and rearrange data improve your ability to review the feedback? Why?
7	Are there any important features you think are missing?
8	Are there other steps that you would have taken in reviewing this feedback?
9	What did you not like about this tool?
10	What did you like about this tool?
11	Could this tool work for a design team when reviewing feedback? Why?

As in the first study, audio was transcribed and analyzed in order to track the opinions expressed by users. Additionally, the performances of users were assessed by reviewing screen recordings and audio in *Adobe Premiere*. This resulted in a report of the success and failure rates of each task. By comparing the actions taken by each user and the transcripts of each session, conclusions could be drawn about where the prototype failed and where it could be developed further.

Procedure

In the first study, subjects met in-person or through video conferencing. Once participants completed a consent form, the audio recorder began and an introduction was read aloud by the interviewer. Participants then worked through the questions, each of which was open ended, acting as a guideline for conversation. Next, participants completed an activity question by explaining how they would go about reviewing a potential set of 1,500 text-based user responses.

Finally, subjects were shown wireframes of a feedback reviewing tool. After the facilitator explained the designs, subjects were asked to respond freely to the designs with any thoughts or suggestions.

The second study required subjects to test a functioning prototype and complete an exit interview. Similar to before, these sessions took place either in-person or remotely. After completing a consent form, the facilitator began recording audio and the users' screen. An introductory statement was read aloud to users, explaining that they would be reviewing feedback for *Coupon Buddy*. They were then shown the design of this app along with the associated user feedback stored in a *Microsoft Excel* document. Users then opened the prototype and proceeded to complete 27 tasks. After this, subjects participated in a short exit interview consisting of 11 questions about their experience with the prototype.

Results

Formative Research Interviews

The data collected from interviews was used to expand upon findings described in the literature review and guide the development of a functioning prototype. These discussions revealed a variety of experiences and opinions relating to user feedback. The features brought up by interviewees, each of which are discussed below, have been ranked by frequency of mention.

View Any Feedback in Its Raw Form

Although visualizations are important for digesting large volumes of feedback, many designers described a desire to view the data in its raw form, such as a list of data entries. In some cases, it is preferable to read through the comments manually rather than have them obscured by a graphic. This slower approach allows designers to get a general sense of user sentiment and confirm that the software has categorized feedback properly. Participant 6 described, "I sometimes just like to read it… to make sure that… what the computer thinks is a negative comment isn't actually a neutral comment based on context".

View Trends of Data Over Time

In discussions with the participants, it became clear that feedback is rarely collected just once. Rather, it is collected multiple times through the design process. This is especially true for managers or strategists, who are constantly reevaluating their products. As participant 3 described, "we do a lot of small feedback collection". Describing their process, participant 6 explained "I would take each of the comments and I would categorize them based [on] what specific issues they were seeing... and would count how many people said that over a period of time. So usually like a month". With this in mind, it is important that a feedback reviewing platform allows users to add new feedback to an existing pool of data and track responses over time.

Build Report and Create Action Items

In addition to exploring and visualizing data, many users also wanted to create something with the feedback. These designers were looking to build a report featuring data to justify revisions. Participant 3 explained, "what's next?… am I going to use them to support… actions I want to take to make a change to the design? Do they map to like overall insights into my design?". This report could take the form of a document or interactive dashboard and would be used by a design team or shared with stakeholders. Designers would also like some way to create

shareable action items which are linked to data. Some interviewees were also interested in having the machine learning software analyze the feedback and suggest action items to the user. As participant 2 described, "it would be amazing if it looked at this and it said, 'hey, we see this common problem, here's the recommendation for how to fix that'... that is what this is missing for me".

View Feedback Metrics and Metadata

Designers also indicated the need to have metadata and metrics attached to the feedback. Participant 1 explained she would like to see "some metrics on the average amount of time that users spend [on] the page... also other metrics I would be interested in is the time that it takes them to complete an action on that page". Participants were also interested in seeing the number of comments stored in each category. As participant 7 asked, "is there any data that shows, like you have X amount of comments?... I know that you have the word like the size of the box, but actual numbers".

Searching and Filtering Data

An important part of reviewing feedback includes querying the data based on keywords and phrases. When explaining how they would organize a large amount of text feedback, many designers mentioned the need for search functionality. Participant 7 described, "So I think my first approach is trying to find specific keywords that we're trying to search for and see if … that comes up". Participant 8 explained "Honestly, my first idea is [to] throw it all on the Google doc... then do 'command+F' and find a word that I know popped up a lot".

A filtering tool was also mentioned by designers. Participant 1 suggested "I could narrow down [with] some sort of dynamic filtering... to focus on the people who spent 10 minutes on this page. What is their path that they're taking on this?" Participant 2 also described this, saying "you can do like filtering on data. So, I'd be like... anyone who responded that they're in [a specific department] and is a section supervisor, let's look at their responses". With search and filtering combined, users can perform complex actions, such as search by a keyword, then filter by a demographic, and then sort by some other criteria in order to locate data.

Ability to Rearrange Data in a Number of Ways

Two designers who use sticky notes in their work to record feedback, described the importance of being able to rearrange data in order to refresh their perspective and gain new insights. Participant 2 explained, "we're constantly rearranging... tons of different ways. Looking

at it, seeing what we see, writing that down". Participant 8 described, "I try to [force] myself to just take a new perspective... So, like mess up all the PostIts and format them in a totally different way".

<div align="center">

Design of Interactive Prototype

</div>

When interviews with designers concluded, existing features in the wireframes were developed further and new capabilities were added. Some features mentioned above were not added due to their complexity. These include the report builder, visualizations which track data over long periods of time, and ability to save sticky note arrangements. The changes detailed below capture the overall functionality of the interactive prototype created for user testing.

Existing Features That Were Developed Further

The features discussed in this section were developed from capabilities represented in the wireframes and others which were described to designers during interviews.

Color coding was integrated into the treemap and barchart visualizations on the main page (see Figure 11). Red indicated negative, gray indicated neutral, and blue indicated positive. In the treemap, colors were blended depending on the amount of each sentiment contained in a given category.
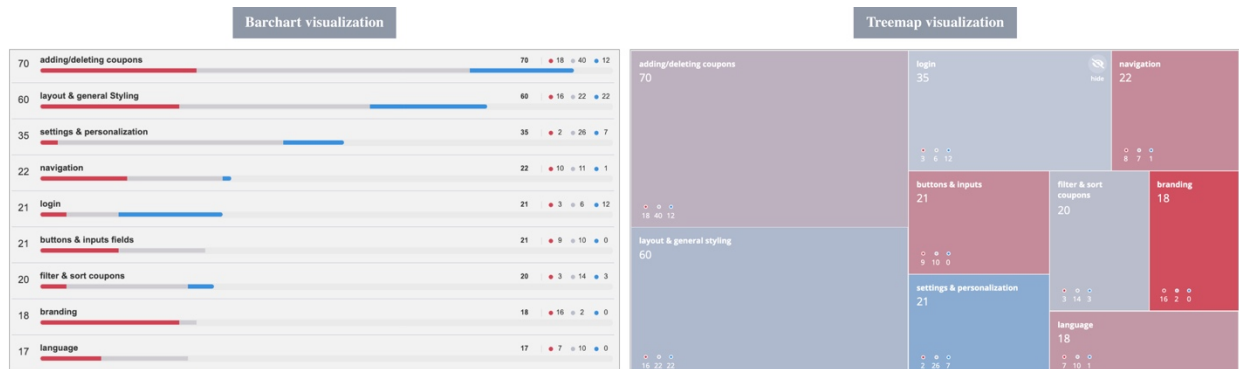

*Figure 11*. Color-coded barchart and treemap top-level visualizations.

The sticky note visualization on category pages was also developed further. Hovering over a note revealed text on the bottom of the screen. Clicking the *expand* icon on a note revealed the entire comment in a popup window. The size of each note was expanded to reveal a snippet of its text, allowing users to browse data without hovering to preview. Finally, sticky notes could be dragged and dropped in order to rearrange data. Notes could be dragged into existing or new subgroups within a given category. They could also be dragged into other top-level categories listed in the side navigation. Furthermore, new top-level categories could be

created by dropping notes into a labelled blue rectangle in the side navigation, which only

appeared when a note was being dragged (see *Figure 12*).
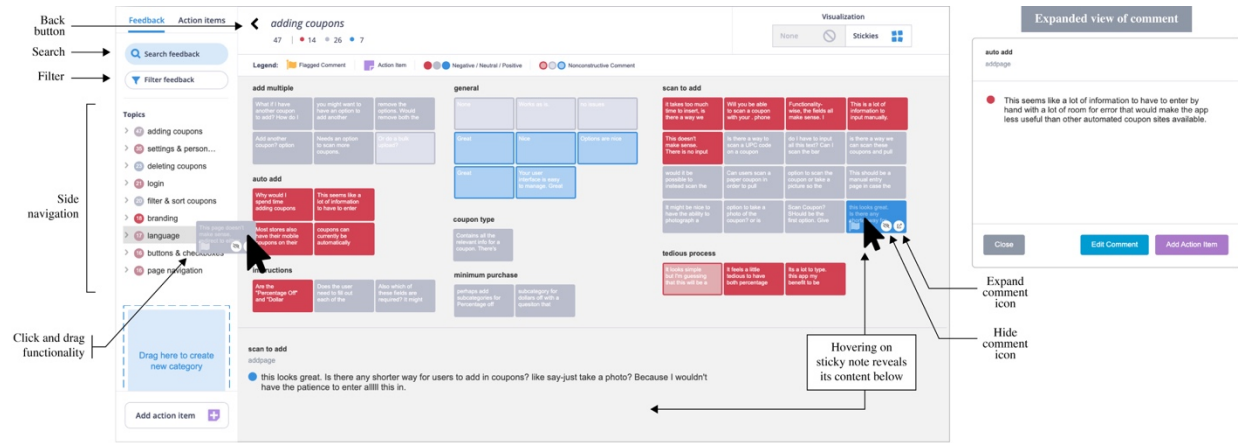
Figure 12. Existing features which were developed further, including sticky note interactions and

navigation tools.

Categorization controls were developed further on the main page to enable global

rearranging of data (see *Figure 13*). A toggle button was added which controlled the

categorization scheme. With this feature, users could arrange the data by *topic* or *page*. Selecting

topic categorized the data into organic themes discovered in the feedback. In this view, controls

were included to allow users to change the number of topics generated by machine learning.

Choosing *page* categorized the feedback based on the screens of *Coupon Buddy* that reviewers

were on when authoring a comment. For example, comments submitted when a user was

reviewing the *home* page would be stored in a *home* category. This aimed to show how data was

distributed across the screens of the application being reviewed.
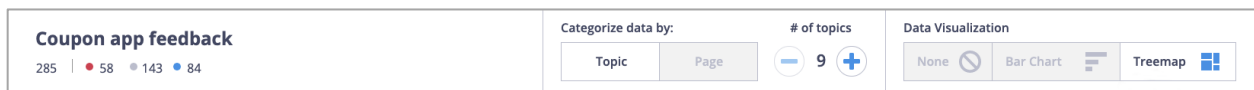
Figure 13. Categorization controls displayed on the top of the screen when viewing top-level

visualizations.

Editing capabilities were also developed (see *Figure 14*). When a comment was

expanded, an *edit* button was added to the resulting popup window. This allowed users to change

a comment's metadata, such as its sentiment and quality. Furthermore, the names of subgroups

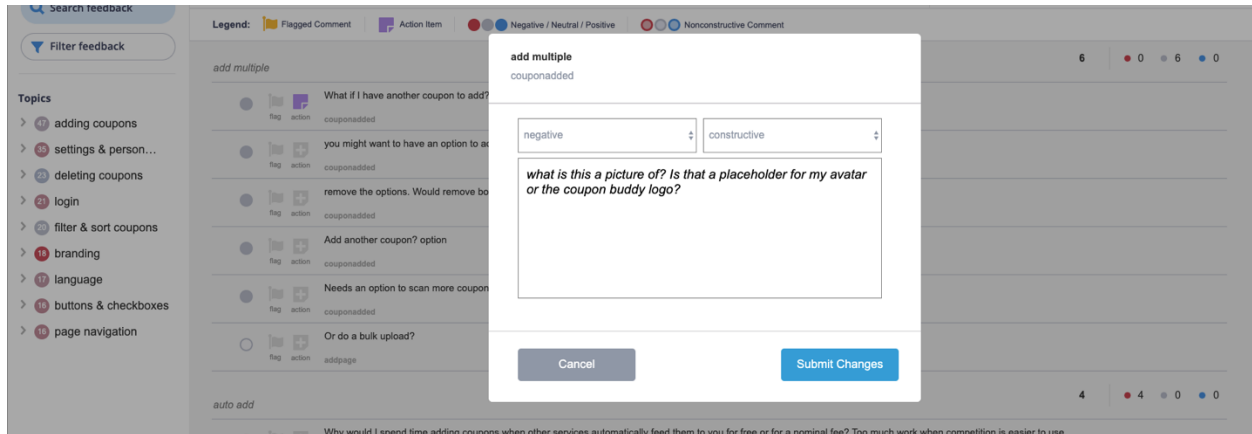within a category could be edited by clicking directly on the title.

Figure 14. Comment editing options.

Tools were implemented throughout the interface to enable users to filter and hide data. A *hide* icon on every visualized category and individual comment was added so users could remove any piece of information from their view. Furthermore, a *filter* button allowed users to choose what data appeared based on sentiment, quality, or the page of *Coupon Buddy* being referenced.

Other features were developed to improve navigation through the data. The side navigation was expanded to appear on all pages within the *feedback* tab. This listed all categories and subgroups in the data, allowing users to navigate to any portion of feedback without having to click the back button and interact with the main page visualizations.

New Features Added Based on Findings From Interviews With Designers

Another visualization option was added which displayed data as a list of comments organized by category. This list view would be the default setting when viewing data on the main page and any category page. Each list item included a color-coded circle icon to indicate sentiment and quality, using the same scheme as the treemap and barchart visualizations. Unconstructive comments were indicated by a circle with a faded center. Filled circles indicated constructive comments. Additionally, this styling was extended to the sticky notes to maintain a consistent visual experience and a legend bar was added above the visualized data to help users

understand the semantic styling (see *Figure 15*).
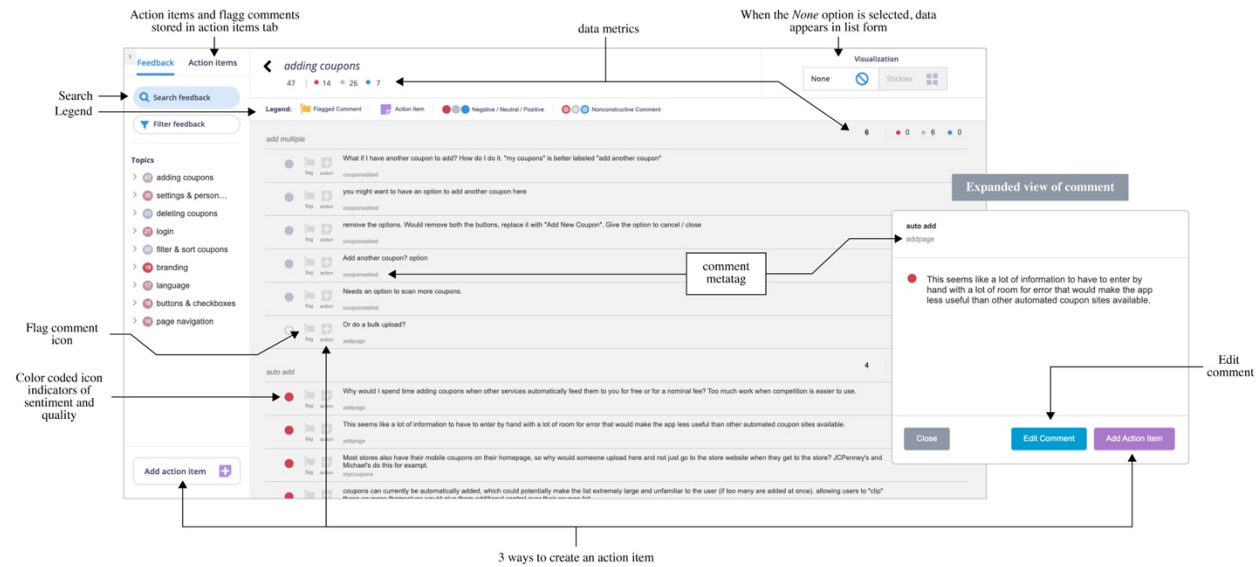
Figure 15. New features added include a list visualization, search functionality, color-coded iconography, data metrics, action items, comment flagging, and legend.

Features were added to enable action item creation and flagging of data. Action items could be created in three ways: by clicking a large action item button below the side navigation, selecting an action item icon attached to any list comment, or expanding a comment and clicking the action item button featured in the popup window. The first method allowed users to manually link a comment to their action item. In the other two methods, data is automatically linked to the action item. Data could be flagged by selecting the *flag* icon on any listed comment or sticky note. Action items and flagged comments were stored in the *action items* tab, accessed from the global navigation on the top left of the interface. From here, users can read all action items and flagged comments as well as share action items via email.

As suggested by designers, a search functionality was added to streamline navigation through the data. Now, users could select a search dialogue above the side navigation and find data based on a phrase or keyword.

To provide more information about the data being visualized, metrics and metadata were added to the interface. Each visualization now included metrics showing how much feedback was collected in total, as well as the number of comments in each category or subgroup. The metrics also showed a breakdown of the number of negative, neutral, and positive comments. Furthermore, a metatag was added to individual comments, indicating what page of *Coupon Buddy* the piece of feedback referenced.

User Tests

Each participant completed a set of 27 tasks. For each task, users received a score of 1, 2, or 3. 1 was given to users who completed a task on their first attempt. 2 was given to users if they required multiple tries to complete a task. Users scored a 3 if they failed to complete a task. Success rates for each task are based on how many users scored a 1 or 2 when performing a task.

Tasks were ranked by priority of high, medium, or low. High priority was placed on tasks where users were exposed to key functions of the prototype for the first time. This included features such as visualizing data, navigating through the interface to new information, creating action items, flagging comments, hiding and filtering information, and editing and rearranging feedback. Medium priority was given to tasks that were repeats of prior actions, such as the second time a user was asked to create an action item. Low priority was given to tasks where users were asked to provide verbal feedback on the design of a page without clicking on the interface (See Appendix G).

Visualizing Data

Tasks 4 and 16 required users to find a way to visualize the data, once while on the main page and again from a category page. In both instances, all 8 users succeeded in locating visualization options on the top right of the page.

Navigation and Search

Tasks 7, 11, 15, and 19 asked users to find feedback relating to a specific keyword or phrase. These tasks had success rates of 87.5%, 100%, 100%, 100%, respectively. In task 7, one user selected the list visualization, rather than the treemap or barchart, and were unable to find the information they were looking for. All other participants completed the tasks either by interacting with the treemap or barchart, search dialogue, or side navigation. As users became more familiar with the prototype, they relied more heavily on the side navigation to get around and less so on the main page treemap or bar chart. Although the search dialogue was also used, it was not the most popular option.

Task 26 tested the search functionality, asking users to query the data to locate feedback related to *color coding*. All users succeeded in finding the search dialogue box and performing a query to find the information. However, 2 users selected the filter button before interacting with the search dialogue.

Task 25 asked users to determine which page of *Coupon Buddy* had the most negative feedback. This would require users to arrange the data by page and then select the treemap or bar chart visualization. 6 out of 8 users succeeded in finding the information in this way, although half of them clicked on a few other options before arranging the data by page. One of the unsuccessful users did navigate to the *page* data, but proceeded to manually scan the list visualization to determine which page had the least favorable feedback.

Tasks 2, 26, and 27 had users locate information relating to action items or flagged comments, requiring them to navigate to the *action items* tab. Task 2 tested to see if users could locate action items. All but one user was able to do this by clicking the *action items* tab. Task 26 tested to see if users could share their action items with a coworker. All users were able to complete this by clicking the *share action items* button. Task 27 asked users to find a way to review their flagged comments. All but one user was able to do this by clicking the *flagged comments* category on the action items page.

Action Items and Flagging

Tasks 10, 14, 18, 22, and 24 checked to see if users could successfully create action items and flag comments in list view or sticky note view. All tasks had success rates of 100%. However, some difficulty was had in task 18, when users created an action item from the sticky note view for the first time. In this scenario, one user first clicked on the *action item* icon in the legend bar, rather than click the *expand* icon on a sticky note in order to create an action item from the popup window.

As described above, users could create an action item in one of three ways. While in list view, most users clicked the *action item* icon on a list comment to do this. While in sticky note view, for tasks 18 and 22, most users expanded the sticky note first and then hit the action item button in the popup window. However, in these two instances, there was a slight rise in use of the large *action item* button below the side navigation.

Hiding and Filtering Data

Task 6 and 9 tested to see if users could successfully use the hide and filter features. In task 6, 7 of out 8 users correctly made use of the *hide* icon to remove unwanted categories of data from the treemap or bar chart. However, two of these successful users clicked the filter dialogue box first before finding this feature. Similarly, the unsuccessful user misunderstood the purpose of the *hide* icon and tried to perform this task with the filter options instead.

In task 9, 5 of 8 users successfully removed unwanted comments from the *adding coupons* category by using the filter options. However, 3 of the successful users clicked on the icons in legend bar before selecting the *filter* button. The 3 unsuccessful users tried to perform the task by using the *hide* icons to manually remove individual comments.

Editing and Rearranging Data

Task 5 analyzed the user's ability to increase the number of categories displayed while the data was arranged by topic. 5 out of 8 users performed this task successfully. The 3 unsuccessful users did not complete this task and instead clicked on topics shown on the treemap in order to find more information.

Tasks 12 and 13 assessed the user's ability to edit information shown on screen. In task 12, 7 of 8 users successfully changed the titles of comment subgroups by clicking directly on the title. The unsuccessful user clicked to expand a list comment and then closed it before attempting to drag it into a different subgroup. In task 13, all 8 users succeeded in editing the details of a comment from the list view.

Tasks 17, 20, and 21 investigated the user's ability to rearrange sticky notes. In tasks 17 and 21, users had to drag comments into existing and new subgroups within a given category page. All 8 users completed this successfully in task 17, whereas 7 out of 8 completed task 21. In task 20, users had difficulty relocating sticky notes into existing and new top-level categories. Half of the users were unsuccessful, as they attempted to do this with the filter functionality and comment editing feature.

Observations

Success rates for tasks 1, 3, and 8 were not tracked. Findings from user observations will be included in the *discussions* section.

Summary

Users had the most difficulty with tasks 5, 9, and 20. These tasks had users attempt to increase the number of topics, filter data, and rearrange feedback into new and existing top-level topics. Each of these had a success rate of 62.5% or less, showing that these features were the most difficult for participants to understand.

Participants experienced moderate difficulty performing tasks 21 and 25. These tasks assessed the user's ability to reorganize sticky notes into a series of subgroups and find

information by arranging the data by page. Both of these tasks received a success rate of 75% percent.

A majority of tasks had success rates of 100% or 87.5%. A portion of these saw every user succeeded on their first attempt. These particular tasks tested the user's ability to select a new data visualization, rearrange sticky notes, share action items, and edit comments.

Exit Interviews

In response to question 1, many expressed that auto categorization enabled by machine learning can provide a great starting point and save time in the reviewing process. However, concern was raised about the visual hierarchy and terminology used to communicate the categorization scheme, as participant 2 expressed "I'm still confused… you use the word *categorize* then *topic* and then *feedback categories*". Further concern was raised about the rigidity of the categorization structure, as it did not indicate overlapping themes. Participant 8 described "categories overlap in my mind, which makes it a bit difficult [to sort] specific comments… *page navigation* is like an overarching thing being and then like *log in* and *adding coupons* underneath it for example".

Answers to question 2 were largely positive. Users expressed that the visualizations, metrics, and iconography allowed them to take in plenty of detail at a glance. Participant 3 described how the visualization options provided different perspectives on the data, saying "I used the bar chart because it gave a quick visual indicator of general spread area. Whereas in like the tree map, I could see what page had the most feedback associated with it." Many users loved the sticky note visualization, but some concern was raised over their small size. Participant 6 described "sticky notes are the way that I tend to... look at those individual data points. The only thing is I notice myself kind of tapping into them so that I could see the full comment". It was suggested that the sticky notes be larger so that more text could be read without having to use the hover to preview.

In question 3, many users explained that browsing through data was made easier by the side navigation, back button, search functionality, and hierarchy of labels. However, some concern was raised over the lack of an obvious *home* button. Participant 3 described this.

> "I do wish there was more of like a home button... I feel more comfortable when I'm starting a new task or something, just clicking *home* to get there... there was one point I got kind of far into an action item... and I had to press back a couple of times."

Responding to question 4, many users explained that search was helpful, especially for locating hard-to-find information. However, concern was raised about the location of the *search* dialogue. Participant 2 explained she would have found this feature more useful "it was placed in a different location". This relates to a reaction recorded during the user test, where participant 7 said "I wanted to search just [the data shown] on the screen, but the way that it's placed, it makes me feel like it might be searching ... everything".

Answering question 5, users were largely fond of the editing functionality but raised some concerns. Editing was seen as an opportunity to train the machine learning software and reorganize data. However, participant 5 explained that it could be dangerous to edit a comment, saying "if they have a lot of spelling mistakes and stuff, that can help build personas and you don't want to necessarily forget that information". Furthermore, participant 3 expressed that the editing process took too long, saying he would like to "jump straight [into] editing" without having to click through two screens.

In response to question 6, many designers explained that the ability to rearranging data brought value to the prototype. Participant 7 explained the rearranging data helps to fix errors, as "AI can't be 100% perfect". A number of users responded positively to the drag and drop functionality of the sticky notes. Participant 4 explained this, saying "moving the individual sticky notes, I could see that being really helpful... you can constantly regroup it and it might reveal different insights".

Responding to question 7, users named a number of items which they thought were missing from this prototype. They would like to be able to select multiple items at once when linking data to an action item. Participant 2 explained "one would think that you would be associating multiple comments with an action item". Echoing this, participant 3 described "if I see four pieces of similar feedback, I might want to create a one sentence summary for all of them". There was also a desire to add hashtags, tag teammates, and specify priority to any piece of feedback or action item. Users also described the need for transparency, as they were not sure what information on screen was authored by a reviewer, teammate, or machine learning. As participant 8 described, "I want some kind of way to know that this is verbatim what the customer [said]... what the machine learning technology thinks is correct. And… how other people on your team [have] made sense of this... I want a more specific breakdown."

In reference to question 8, a few users suggested other steps they would have taken in reviewing feedback that this prototype did not simulate. For participant 2, there was a strong desire to validate the work done by machine learning by checking *yes* or *no* on comment metadata and category names. Participant 5 wanted generate a report based on the action items they created, saying "I may have wanted to write a... report and export it into like an excel document or word document or something". Finally, participant 8 expressed the desire to view how the feedback maps to the screens of *Coupon Buddy*. Describing this, she said "one thing I'm thinking about is... are there visuals of the wireframes that can be added to this and how could you connect some of the comments to what actually was happening?".

Responding to question 9, users described a number of elements they did not like about the prototype. To participant 8, it was initially unclear how she ought to interpret the side navigation. Explaining, she says "it took me a second to figure out how the side nav related to the main content". Similarly, the global navigation tabs on the top left caused some confusion due to their placement in the interface. It was expected that these tabs would only affect the content in the left column, rather than affect the entire screen. Participant 1 explained she thought it "was only changing that side panel view when in fact it changed the entire context of the entire view". Finally, some users were confused about the legend bar and thought they could click on it to perform various actions. Participant 8 expressed "the legend up here feels clickable... it seems more like a filter thing".

Users described a number of details which they enjoyed about the prototype in response to question 10. Echoing earlier sentiments, they enjoyed the utility of visualizing data, the organization of categories, the ease of navigation, rearranging and editing data, and the ability to create action items and flag comments.

Finally, responding to question 11, users affirmed that this tool would be useful for a design team reviewing feedback. For many users, this prototype had potential to become a collaborative tool. Participant 6 described, "I think people working together would be able to see all the comments that have come in, be able [to] look through it pretty easily and edit it. So it's… very collaborative".

Discussion

Categorization

Although data categories helped participants find information, some were confused by the terminology used to described the organizational structure and concerned that the overall design was too rigid.

The terminology used to describe categories and subgroups was confusing to some designers. This is confirmed by user test results, as tasks 5, 20, and 25, where users moved data into new topics and manipulated the global arrangement of data, had low success rates. This confusion could have been caused by the moderator's script during user tests. However, the interface can still improve by using clear and consistent language when referring to the data structure. For instance, the words *topic* and *category* are both featured in the prototype but have similar meanings. Removing *topic* from the vocabulary and using *theme* instead may help to build more distinction. Furthermore, improving the typography of category and subgroup titles may help to reinforce their hierarchical relationship.

Furthermore, some users worried about the categorization structure being too rigid, not accounting for overlapping themes. In future designs, another data visualization should be considered which illustrates the relationship between categories, showing where they intersect. Furthermore, hashtags could be added to individual comments, allowing users to discover feedback without being bound to a rigid structure.

Machine Learning Capabilities

Most users were only interested in this capability as long they knew what data machine learning affected, understood how it made decisions, could use it to help them organize data in real time, and had the ability to train it.

A number of features need to be developed to address these concerns. To increase transparency, the author of any piece of data, category, subgroup, or action item should be indicated by a label. Furthermore, while the user is reviewing data, they need to be empowered to confirm what the machine learning has done correctly and incorrectly. Finally, more control needs to be given to the user, so they can adjust the criteria that determines the placement and quality of feedback.

Users also indicated that they wanted assistance from machine learning in the moment. For instance, when dragging notes together to create a subgroup, machine learning could

automatically suggest a name for that group of data. Participant 4 described this, saying "I think I'd like to see what the machine learning names [them]. But then have the ability to edit it". In future iterations, real-time assistance from machine learning should be tested. It may be helpful for machine learning and designers to work in parallel when reviewing data.

Navigation and Search

When navigating through the prototype to locate feedback, some concern was raised over the global navigation, side navigation, and search functionality.

Overall, navigation features were easy to use and made the data less daunting. However, some users took issue with the static nature of the side navigation, hoping it would change as they navigated. Another did not fully understand how the global navigation would affect the interface, thinking it would only change the left column. Users wanted more flexibility when searching. They desired the ability to query any data, including authors and action items, and perform searches at any time by pressing *ctrl + f* on their keyboard. Furthermore, users wanted the ability to search a portion of data rather than the entire set of feedback.

Overall, navigation features were easy to use and made the data less daunting. However, small changes to these features could improve the navigation experience. Expanding the global navigation bar to be the full width will better indicate that it affects the entire screen. The side navigation ought to visually highlight where a user is in the data and all times. Furthermore, drop down items in the side navigation should be clickable, as many users attempted to do this to no avail. Finally, search must evolve into a more powerful tool. This can be done by allowing users to determine which portion of data they are searching and enabling them to look up information outside of the user feedback.

Prototype Structure and Layout

A portion of users were dissatisfied with the general layout and styling of the prototype, expressing concern over the placement of specific features and the general look and feel of the interface.

Some participants felt that certain features would work better if moved to other parts of the prototype. Suggestions were made that the search and filter buttons would make more sense placed above the visualized data, rather than above the side navigation. Furthermore, some users said that the placement of action items and flagged comments on a separate tab made them feel disconnected from the feedback. Users wanted access to this information without having to

navigate away from the data. Finally, as mentioned before, some users raised concern over the lack of an obvious *home* page.

There were also a number of comments made about the graphic look and feel of the prototype. Concern was raised over the styling of text, that it was hard to read in some cases and did not always reinforce the hierarchy of data. Also, the styling of button states was confusing for some users. When toggling a button, one user could not determine which state meant *selected* or *unselected*. Finally, some users expressed that green would have been a better choice to indicate positive sentiment. This feedback about the look and feel of the design was expected, as the focus of this prototype was more on functionality than visual appearance.

Beyond a general graphic refresh to improve styling and typography, a number of other changes can be implemented to improve the general layout for users. Changing the *feedback* tab to *home* will more clearly indicate how to return to the default view. Alternatively, a landing page for the prototype could be created, where users enter before viewing the data. On this page, settings for the machine learning could be displayed, giving users more control over this technology. Furthermore, the *search* and *filter* features should be placed above the visualized data, so their functionality is more obvious to designers. Finally, two features should be implemented to reduce the feeling that action items and flagged comments are disconnected from the data. First, a button should be added on the data page which triggers a slide-in panel displaying all current action items. Secondly, the filter panel should include an option to filter all data that is flagged.

Data Visualizations

Overall, participants found the data visualizations to be valuable. However, the legend and metrics, which were added to expand upon the visualizations, were met with concern by many users.

All users reacted positively to the data visualization options, affirming that they helped them save time when analyzing data. Designers enjoyed having the list visualization available to display feedback in its raw form, along with colored iconography to indicate sentiment and quality. Between the treemap and barchart, the treemap was favored slightly more, as 5 of 8 users opted for this visualization as their primary view. However, it was indicated by some that both visualizations were useful, depending on what information was sought. Users also enjoyed the sticky note visualization, more so for its interactivity than its appearance. Many designers

hoped the notes were larger so more text could be displayed. Furthermore, they wanted a way to lock the hover preview, so they could drag their mouse over the content and highlight text.

The visualizations were accompanied by a legend bar and metrics. These were added to help users understand the semantic styling and comprehend the quantity of feedback received, broken down by positive, neutral, and negative data. Some designers made use of these features, but others found that they were not easy to understand. One hesitated when reading the quality indicator, asking "what would be considered a constructive versus a nonconstructive comment". While using the legend to make sense of the data display below, participant 5 expressed "I'm trying to match up what each thing goes to, but I'm not sure if these little circles… go to the topics or if they go to the negative non-constructive". In many instances, users incorrectly clicked on icons displayed in the legend when attempting to filter data or create action items.

Overall, the data visualizations proved to be valuable to the user experience. However, they should be more flexible, allowing for resizing of sticky notes and locking of the hover preview. Furthermore, the design of legend and metrics should be refined so that users have less difficulty interpreting their meaning and functionality. This can be accomplished by fine tuning their graphic appearance.

Feedback Context

Although each comment included a metatag that described the page of *Coupon Buddy* being referenced, users wanted more contextual information for the feedback being reviewed.

Many designers hoped there was more context provided for the feedback displayed. As participant 8 expressed, "it's hard for me to figure out what this comment [is] about". Similarly, participant 2 described, "[typically,] it's nice to know where in the application they were when they made this comment". As mentioned before, participant 8 asked if there was any way for an image of the design in question to be included in the feedback reviewing interface.

It is clear that each comment did not provide enough context for users. This can be solved first by improving the typography, so that the metadata for each comment is more clearly seen. Details such as the author and origination date can be included here. Furthermore, when a user clicks to expand a comment, the modal which appears could also include an image of the *Coupon Buddy* page being referenced. Alternatively, research could be done to determine how feedback could be mapped to each screen of the product being reviewed in order to create a new data browsing experience.

Creating Action Items and Flagging Comments

Although the action item and flagging features were easy for participants to operate, data was rarely flagged and some participants wanted to be able to create action items more quickly and specify their priority.

User testing showed that the action item functionality could use some improvements. The process for creating an action item was tedious, involving too many steps. Expressing some discontent, participant 5 said "Oh, okay. I've now clicked through like three screens". Users also wanted to be able to link multiple comments to an action item. Finally, there was a general desire to pack more information into an action item by assigning a teammate, adding a hashtag to indicate priority, or labelling it as a *question* rather than an *action*. As participant 8 described, "in [design] strategy, I just end up coming up with more and more questions. And so, it's not really like an action item… [it's] the big hairy questions that are bubbling up for us".

A number of changes can be implemented to make this feature more powerful for users. Improving this feature will require a simplification of the creation process. Users should only have to click once on a comment in order to begin creating an action item. Linking numerous comments to an action item could be made easy by selecting multiple comments before clicking the *action item* button. Users should also be able to tag teammates to an item, give it priority, and indicate whether it is a to-do item or something else, like a question or area for growth.

The flagging tool was used less frequently than the action item feature, as it was not clear what value this feature brought to the experience. Some designers mentioned that flagged data should be available within the data panel, rather than being stored elsewhere. As mentioned previously, this feature may be more useful as a filter option, so that only flagged comments are displayed.

Manipulating Data

Users were especially happy with data editing features as well as the drag and drop functionality for rearranging feedback. However, some participants were confused when they had to create new top-level categories and use tools to rearrange data on a global scale.

Rearranging data by dragging and dropping sticky notes proved to be very intuitive for users. Participant 1 affirmed, "I liked the dragging… it's very similar to things I've used before". Users were able to move notes into existing and new subgroups with this feature. In addition,

some designers were interested in declaring names for newly created subgroups, either by clicking a button to add text or by having machine learning automatically name it.

The drag and drop functionality should be expanded further in future iterations of the prototype. Participant 8 recommended "if you're going to do a drag and drop interaction, then make it prominent throughout the entire app." This means that comments should be rearrangeable when in list view as well.

Dragging and dropping was not the best method for everything. Many users did not realize they could create a new top-level category by dropping a comment into the blue rectangle appearing below the side navigation. To fix this, future designs should feature a button within the side navigation to create a new category. Furthermore, users should also be able to change a comment's categorization through the editing feature, as many users attempted to recategorize data in this way.

The editing feature, allowing users to change comments and subgroup titles, was very easy for designers to use. In general, participants responded positively to the idea of editing category and subgroup titles and comment sentiment and quality ratings. However, as mentioned before, some users were worried about the negative result of editing a comment's text. Therefore, in future iterations, a setting should be added which allows users to lock the ability to edit the text of a comment.

The features which controlled the arrangement of data from a global level, such as the number of topics and the arrangement of data by *page* or *topic*, were hard for users to understand. When increasing the number of topics displayed, one user expressed "I guess I'm surprised to see that [feature] *here*". Furthermore, when looking for the page with the largest amount of negative feedback, most users did not initially arrange the data by *page* to find the answer. These tools may be better understood if placed in a *settings* page or on a preliminary landing page where users can adjust the parameters of categorization.

Filtering and Hiding data

Features which enabled hiding and filtering data were met with some confusion. Hiding individual comments and categories by clicking the *hide* icon was relatively easy for users to understand. However, participant 8 did raise some concern over this feature, saying "I don't see any indicator that I'm hit anything on here". When large portions of comments had to be obscured at one time, many users still relied on the icon rather than leveraging the filter button.

Oddly enough, this button was often used to perform tasks that it was not built to do, such as locate action items or navigate to a category page. This indicates that users' expectations for the filter button were different from what it was designed to do.

Changes can be made to clear up this confusion. When a piece of information is hidden, a visual indicator needs to appear which confirms that data has been removed. Furthermore, placing the filter button directly above data on the right side will more clearly associate it with the data being displayed on the screen. Finally, to make the filter meet more of the users' expectations, it can be expanded to filter data by *page* and *topic*.

Conclusion

What a Designer Needs

This research aims to address the challenge of leveraging large volumes of unstructured feedback in the design process. To make full use of this data, designers need feedback categorized with tools to navigate and manipulate information. Current tools built to present design feedback do not effectively meet this need. Instead, many of them provide recommendations based on structured feedback, packaging data into predefined categories with limited tools for analysis.

Early research revealed how designers could be better served by a feedback reviewing tool. Rather than pour data into predefined categories, machine learning can be implemented to mine out the unique themes embedded in the data. The resulting data structure can be visualized into one or more interactive graphics which progressively discloses data to users. This allows designers to see the broad narrative of the data while also enabling them to easily explore it at a more granular level. As the user explores, visual indicators should provide them an awareness of where they are in the data at all times. Additionally, tools to edit, hide, filter, and rearrange the data will provide designers the power they need to use feedback in their creative process.

Discussions with professional designers brought to light other features which would further increase their ability to utilize large volumes of feedback. The first is an action item creator, where they can record any insight, question, or task and link it to data. Designers were interested in using such a feature as a way to keep track of their progress as well as collaborate with teammates.

Interviews also revealed other ideas for viewing and navigating data. Although graphics can help to tell the story behind a set of data, designers were also interested in seeing the data in its raw form. Furthermore, some interviewees described the desire to have the relationship between data visualized, displaying conflicts and intersections. Finally, although interactive visualizations were meant to serve as the main method of navigation, users were also interested in using a search function to locate very specific or hard-to-find data.

Designers were also interested in viewing metrics on the data. Many indicated that displaying the number of comments in each category would be helpful in combination with a graphic visualization. Other designers, who often had to collect data repeatedly for the same product, were interested in seeing how these metrics change overtime.

Analysis of The Prototype

Valuable Features

In testing the interactive prototype, a number of features proved to be valuable to designers, such as visualized data. The graphics provided a broad narrative of the data, using color, scale, and iconography to indicate sentiment, quantity, quality, and interactivity. This styling was applied to all visualizations throughout the tool, providing a consistent viewing experience. Having multiple visualization options gave users flexibility in how they explored the data.

Automatically generated categories also proved to be valuable. This capability saves time, removing the tedious task of manually sorting through data. However, for this feature to be valuable, the machine learning which enables it must be fully transparent and trainable.

Navigation features also benefited the user experience, making it easy to move through complex data. These included the interactive visualizations, side navigation, back button, and search dialogue. Of these options, the side navigation was most heavily used, as it gave a persistent snapshot of the entire dataset. Users expressed interest in having it provide even more visual indication of where they were in the data at any given moment.

The action item creator also added value. This feature was implemented to give users something to do with the data. Designers were able to keep track of their work and share insights through this tool. Moving forward, this feature could be made even more powerful by enabling users to prioritize action items, tag teammates, and create detailed reports for stakeholders.

Finally, allowing users to rearrange and edit data proved to be useful. By dragging and dropping information into other parts of the data structure, designers were enabled to reorganize feedback as they saw fit. Although this interaction was less useful when creating new categories, users responded well to its other uses and suggested it be applied to listed comments as well. Users were also able to change titles of subgroups as well as edit each comment's text, sentiment, and quality. Although there was some concern raised about changing the text of a comment, users indicated that the other aspects of editing helped them organize data and correct mistakes.

Areas for Improvement

A number of other features proved to be of little value, requiring revisions and further testing. Some users felt that the global navigation, which separated the tool into two tabs, made

the data feel disconnected from the action items. These designers wanted the action items to be better integrated into the feedback data. Furthermore, because this element was located in the top left corner of the screen, some users were not sure how it affected the screen. Finally, some users thought the global navigation design missed the opportunity to include a clear home page where other settings could be stored. In fact, two controls which were confusing to users, the button to increase topics and the feature to toggle data arrangements between *page* and *topic*, may be better placed on such a page.

The data shown on screen did not provide enough context or transparency for users, as many wondered where each category and piece of data came from. In user tests, participants were aware that the information on screen resulted from the efforts of the reviewers, teammates, and machine learning. This division of labor may have been what caused concern, as users did not fully understand how each party affected the data.

Flagging comments was a feature built to help designers prioritize data; however, it was seldom used. After flagging data, users could find all of their prioritized information stored within the *action items* tab. However, this seemed disconnected to participants. Users felt that flagged data should be accessed in the *feedback* tab, serving as a way to filter important feedback.

A filter feature was implemented in the prototype, enabling users to narrow down the data shown on screen. However, many users misinterpreted this tool, opting for more manual methods when hiding unnecessary information. Users needed a faster way to hide information, but the filter button, in its current design, did not present itself as a good method. In fact, some users attempted to filter and sort data by clicking on the legend bar. Reassessing the design, capabilities, and placement of this filter tool could improve its usability.

Limitations

Although machine learning was discussed heavily, it was never technically implemented into the design. Instead, it served more as a backdrop to the research. To create the interactive prototype, real feedback about *Coupon Buddy* had to be manually sorted and categorized in *Microsoft Excel*. In testing, users were aware that the interface was merely simulating machine learning capabilities.

The focus of this research was limited partly by the capabilities of *Axure*. Although this is a powerful tool, it was not possible to populate the entire interface with real, dynamic data

without crashing the program. So, in many cases, the prototype's functionality had to be simulated. Furthermore, specific features could only be tested on an individual page. For this reason, users often noticed that certain tools worked on some pages but not others. Whenever necessary, the moderator would step in to explain these gaps in functionality. Finally, limitations in *Axure* meant that a number of features mentioned in discussions with designers could not be tested due to their complexity. Efforts had to be focused on important features which could be developed and tested in a reasonable amount of time.

The complexity of this prototype and the limitations of *Axure* meant that the research had to focus more on functionality than on visual appearance. As a result, many users raised concern over the colors used, the way elements were styled, the placement of buttons, and poor typographic hierarchy. These responses were expected but were still important reactions to record and discuss.

Future Work

Aside from the smaller fixes which are discussed above, this research revealed a number of larger focus areas for future research. The first is collaboration. Many designers work with a team to analyze and synthesize data, so this tool should be built to be used within that context, allowing teammates to see each other's work, communicate, and track each other's progress on action items.

Users are also interested in context. More consideration should go into how the visual design being reviewed is brought into the feedback analysis process. Perhaps there is a way to explore the feedback by browsing the pages of the design.

Designers are interested in machine learning as long as they have control of it. Future research should investigate how users can most effectively train machine learning. Building this into the data reviewing process could help remedy any trust issues users have with automation.

It is important to show how data intersects and changes over time. Although data visualizations proved to be useful, they did not indicate where the feedback overlapped. Furthermore, the visualizations did not display trends through time, which is valuable information for designers who repeatedly collect data for the same product. Therefore, further research should investigate how data relationships and trends are displayed and navigated.

Appendix A

Formative Research Interview Questions

#	Question
1	What area(s) of design do you work in?
2	In what way(s) is user feedback important to your design process?
3	In what quantity and how often is feedback used in your design process?
4	Do you have an end goal when reviewing feedback on a design? Are there certain key insights you are looking to uncover?
5	When reviewing user feedback, what strategies/processes do you use to organize and sort responses?
6	What problems do you face when reviewing feedback?
7	How do you address the problem of analyzing feedback which is ambiguous, unconstructive, or contradictory?
8	Have you ever had to review large amounts of feedback, in the 100s of 1000s? In that scenario, how did you make that amount of feedback manageable?
9	How could a digital tool help you in making feedback review easier and more productive? Are there any tools you use now?
Activity Question	Assume you are reviewing the design of the homepage of a website. You have collected 1,500 individual text comments on this page. Could you walk me through how you might go about organizing that data and discovering key insights about the design?

Appendix B

Formative Research Wireframes
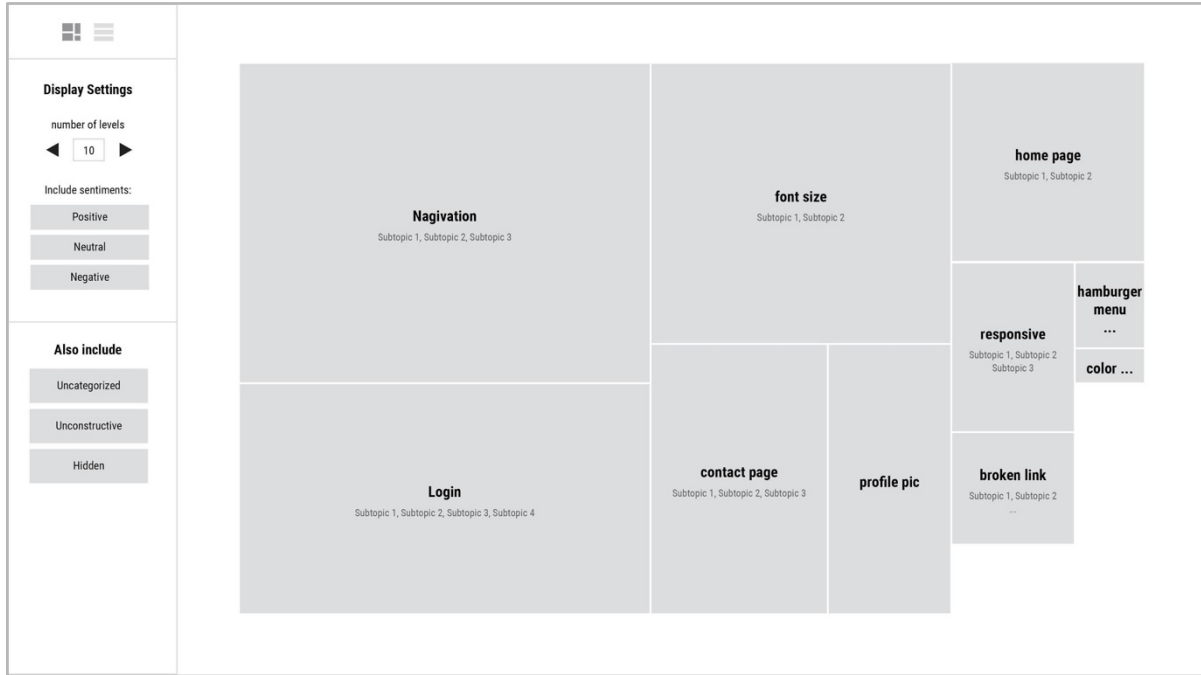
Main page: treemap visualization.
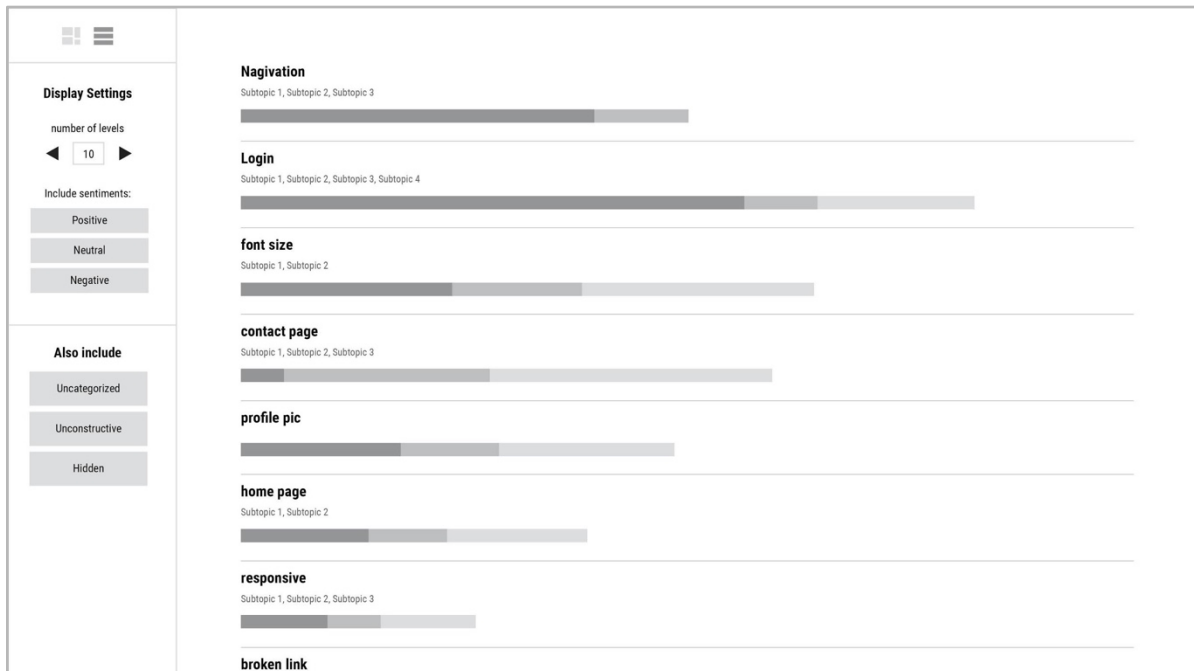
Main page: barhart visualization.
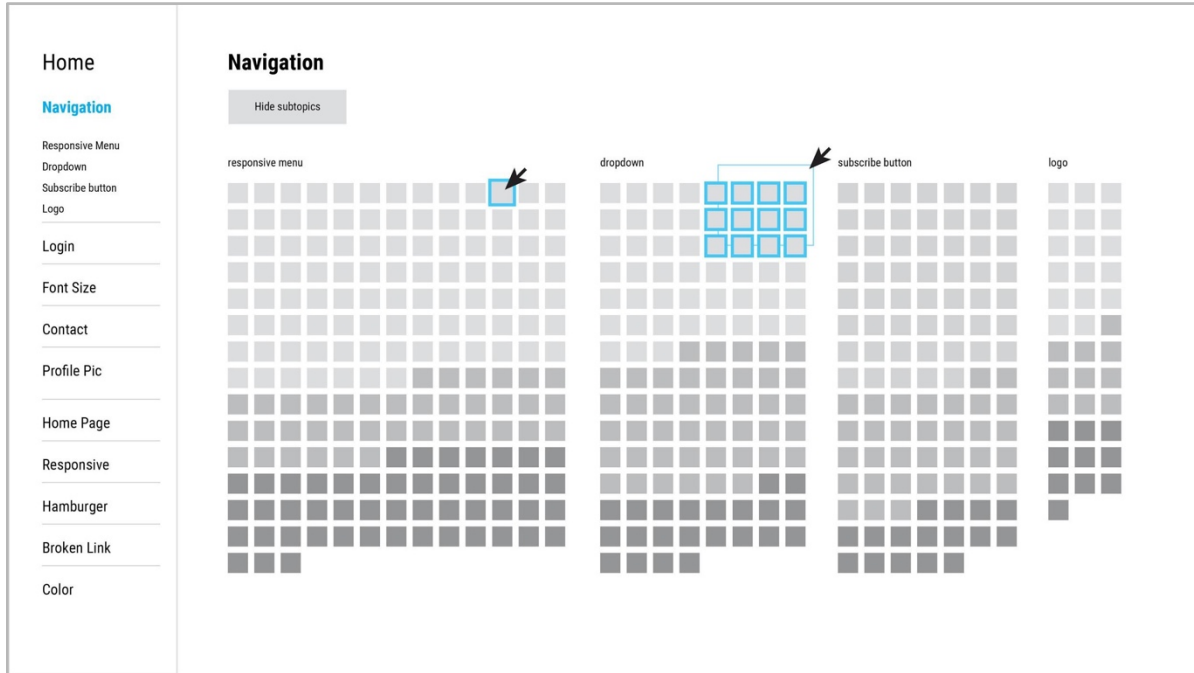
Category page: sticky note visualization.
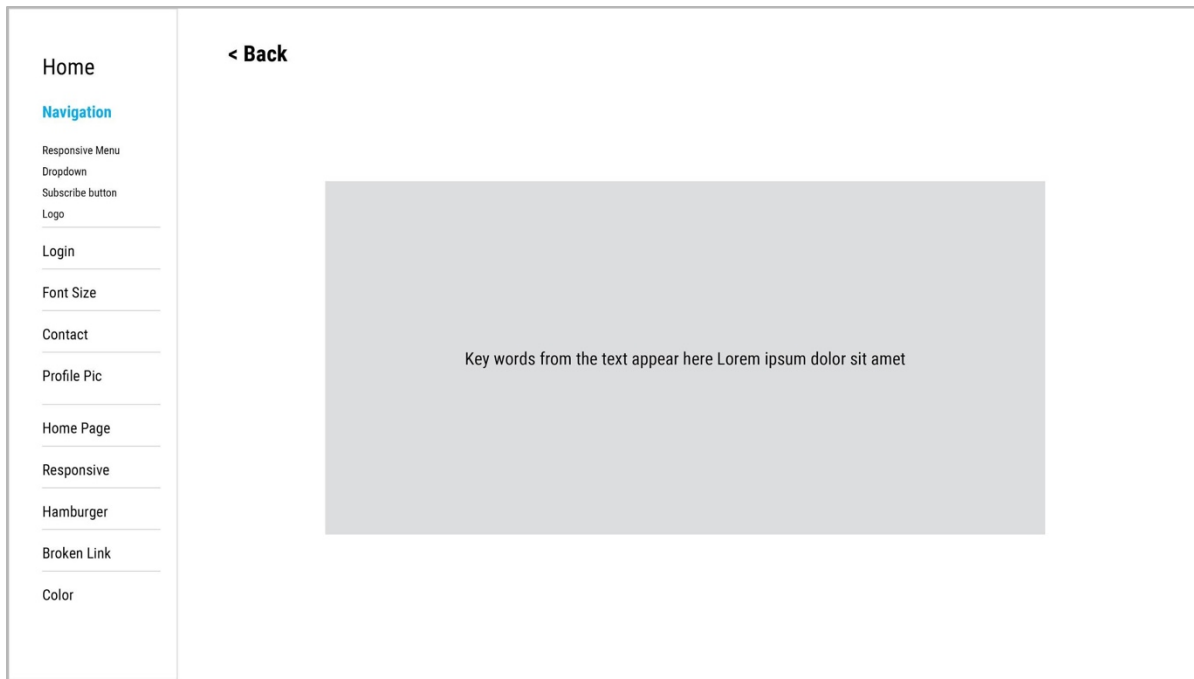
Category page: expanded comment view.

Appendix C

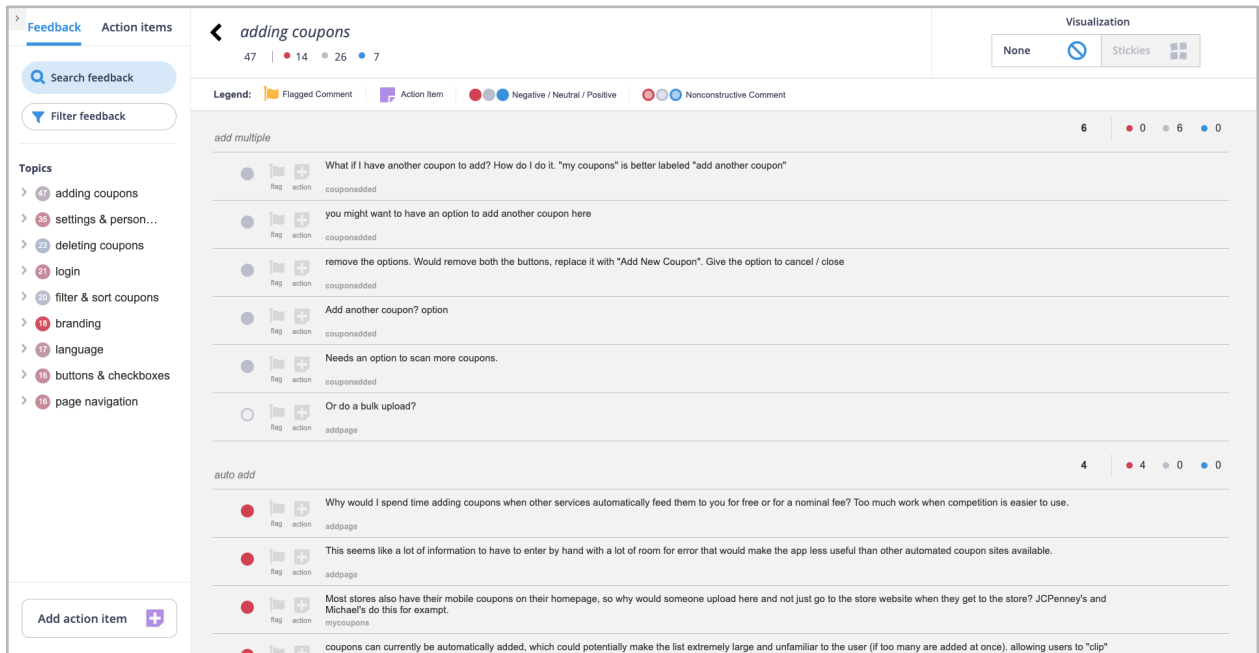**Interactive Prototype**
https://6y5h40.axshare.com
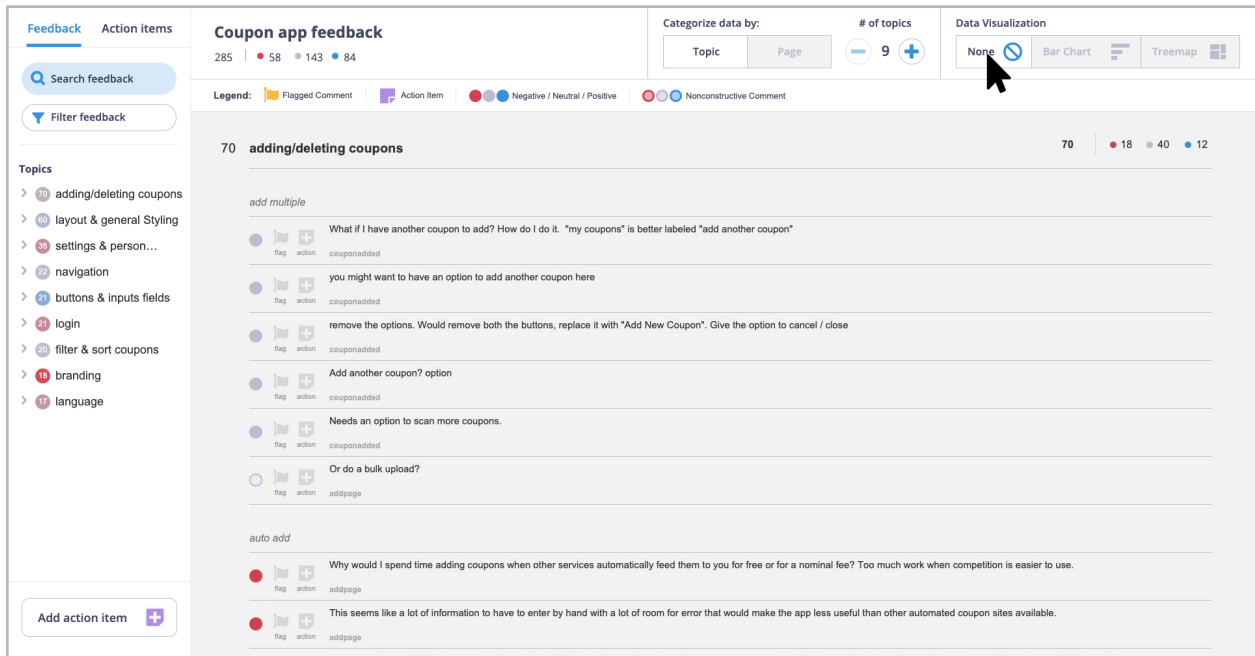
Feedback tab on the main page, list visualization.
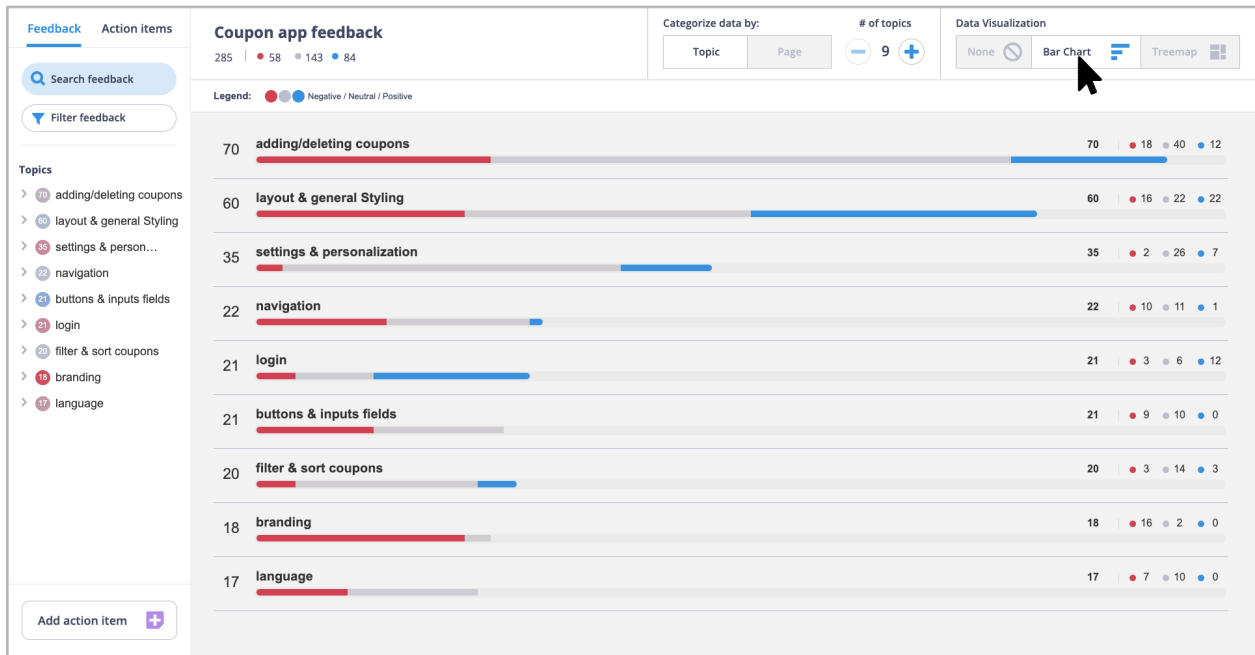
Feedback tab on the main page, barchart visualization.
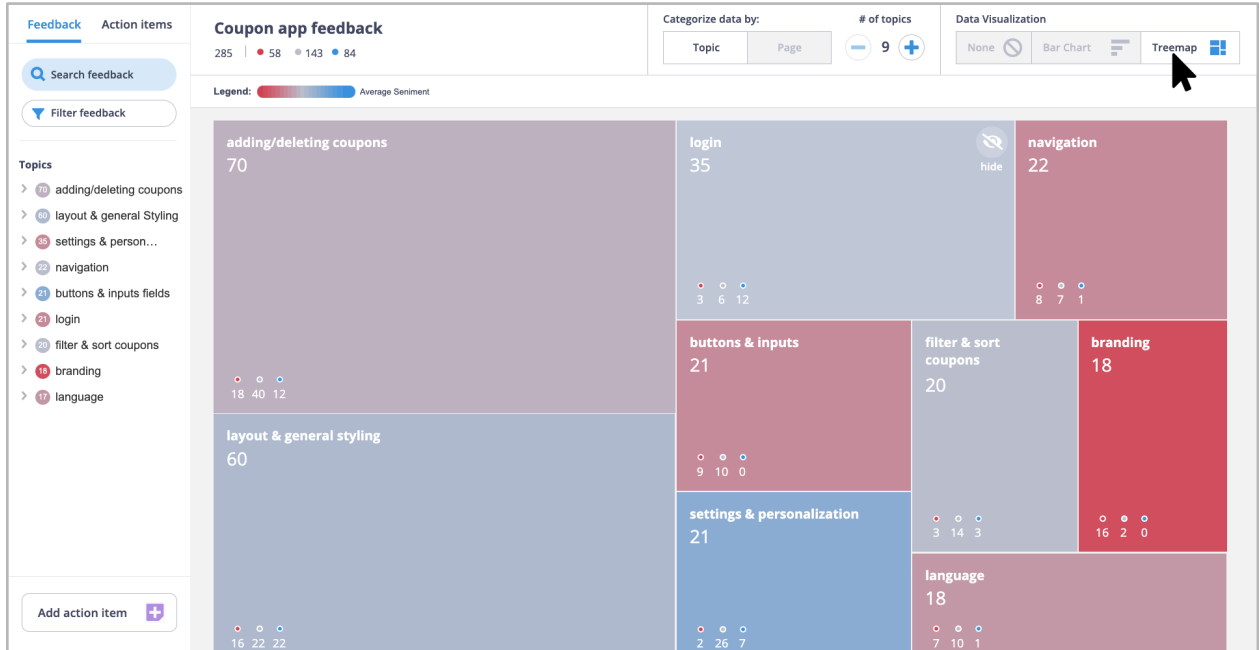
Feedback tab on the main page, treemap visualization.

Category page, list visualization.

Category page, sticky note visualization with hover preview.

Click and drag to rearranging sticky notes.

Side navigation dropdown items reveals subgroups.

Search dialogue.
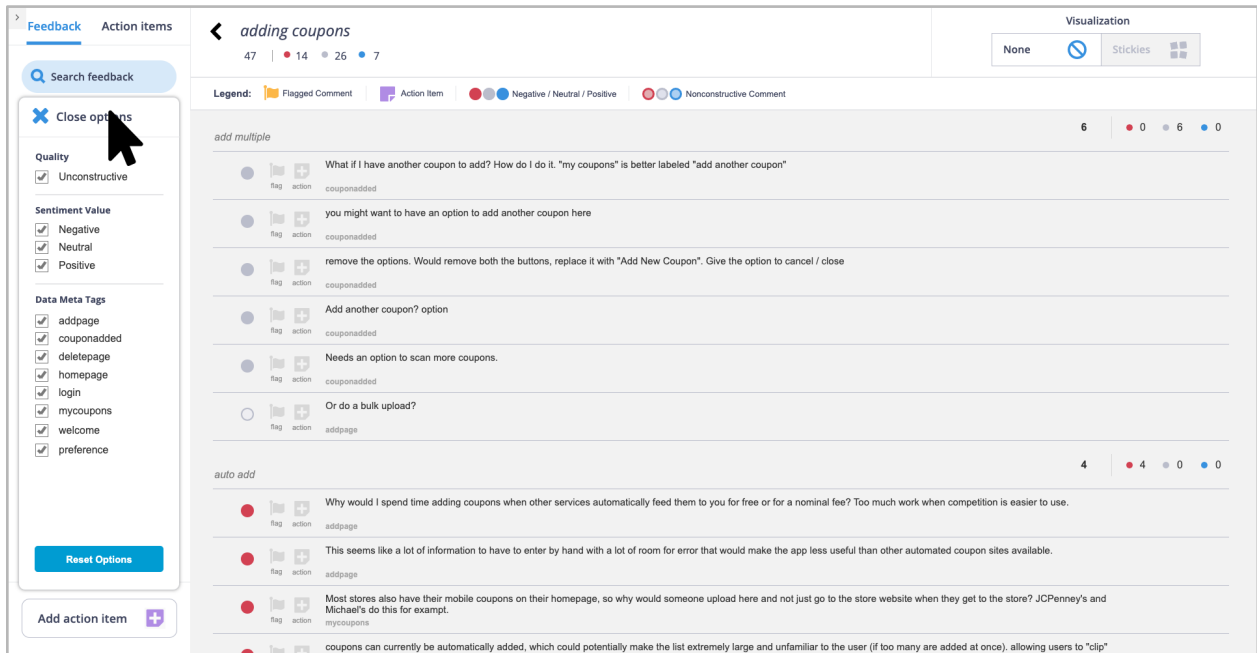
Filter options.
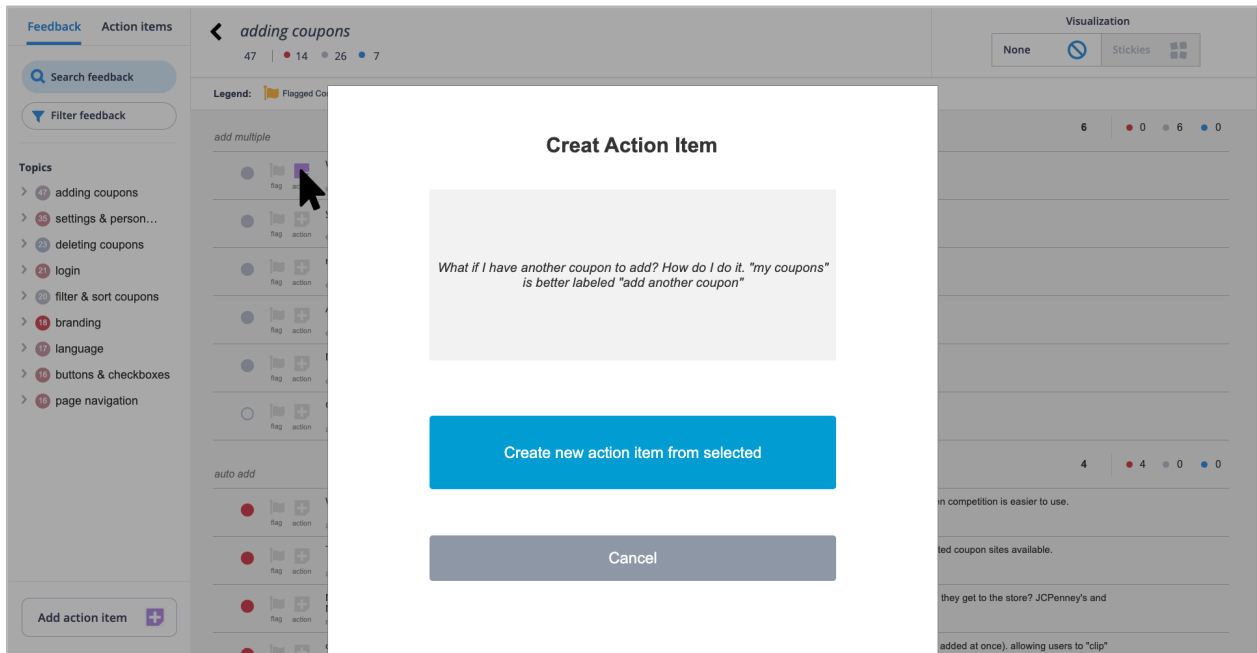
Action item popup screen when action item icon on comment is clicked (screen 1/2).
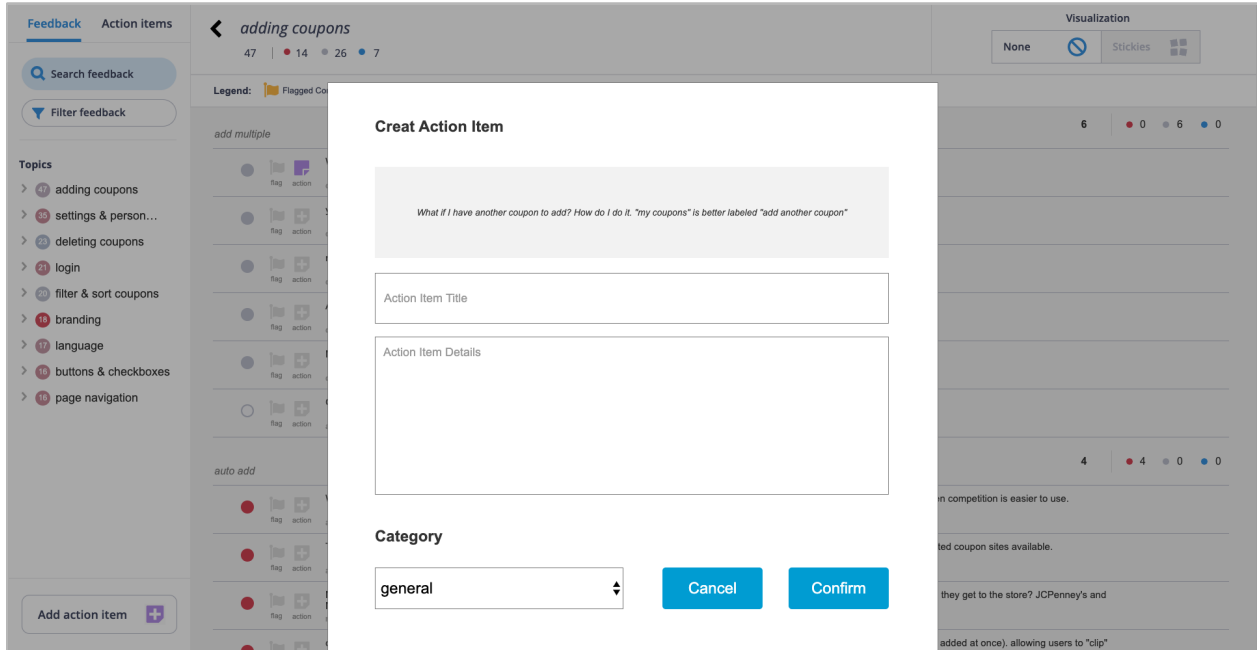
Action item popup screen when action item icon on comment is clicked (screen 2/2).
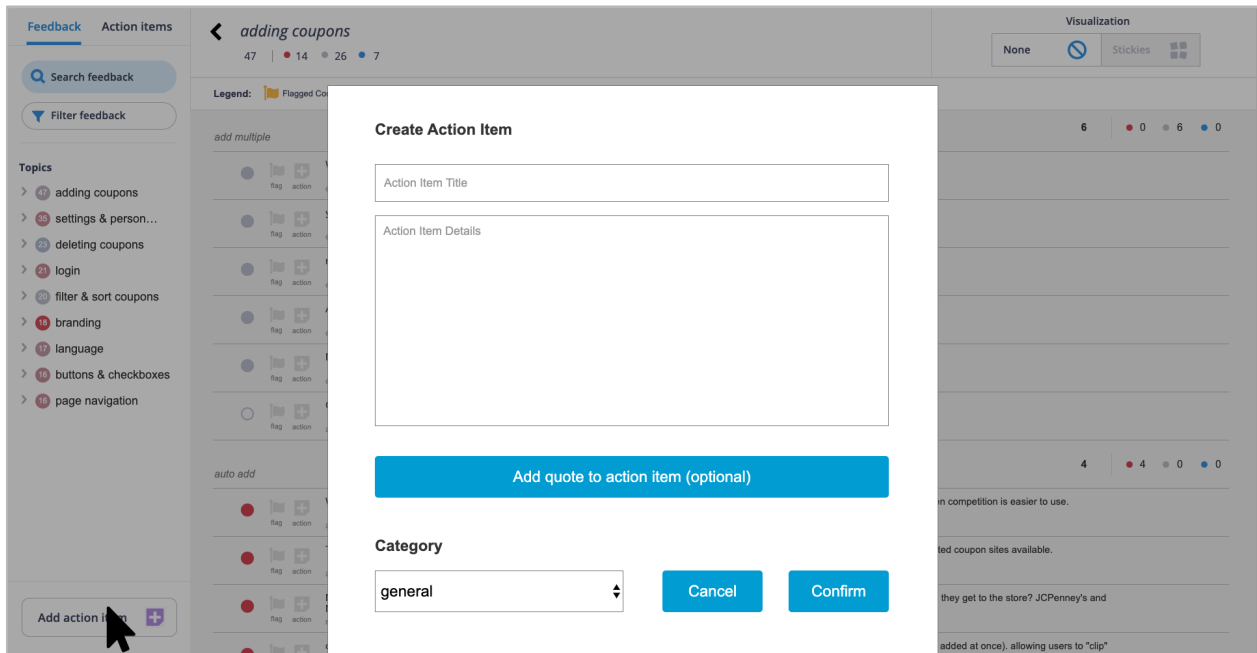
Action item popup screen when large action item button is clicked (screen 1/2).
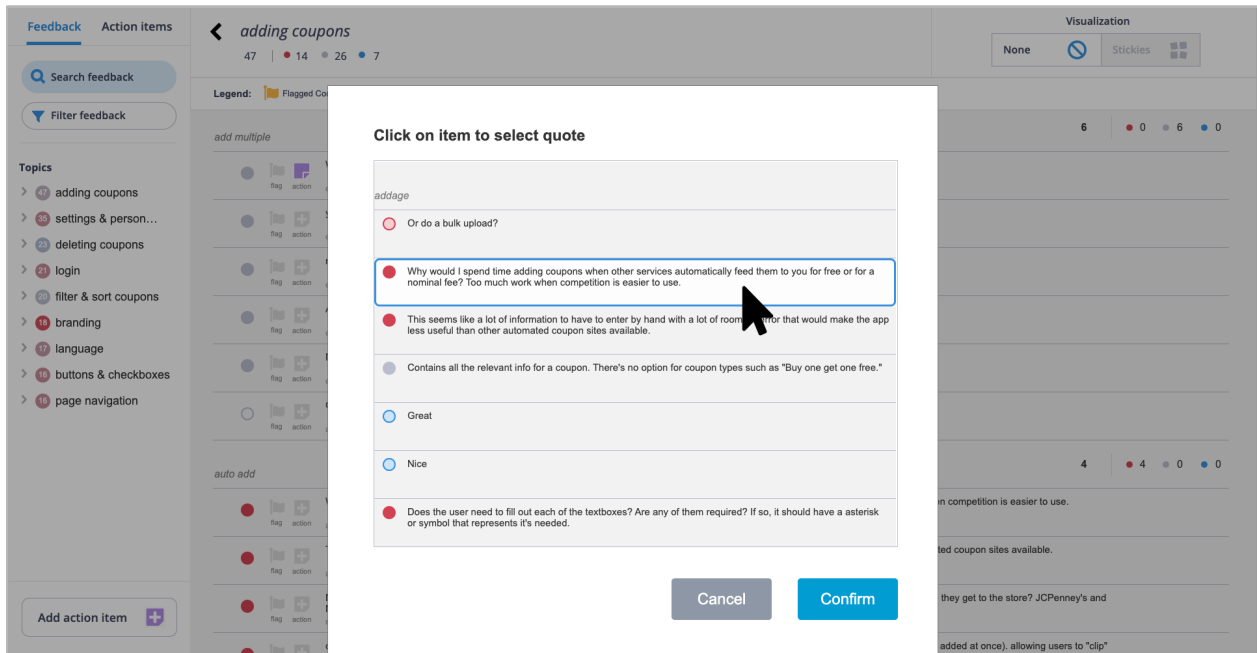
Action item popup screen when large action item button is clicked (screen 2/2).
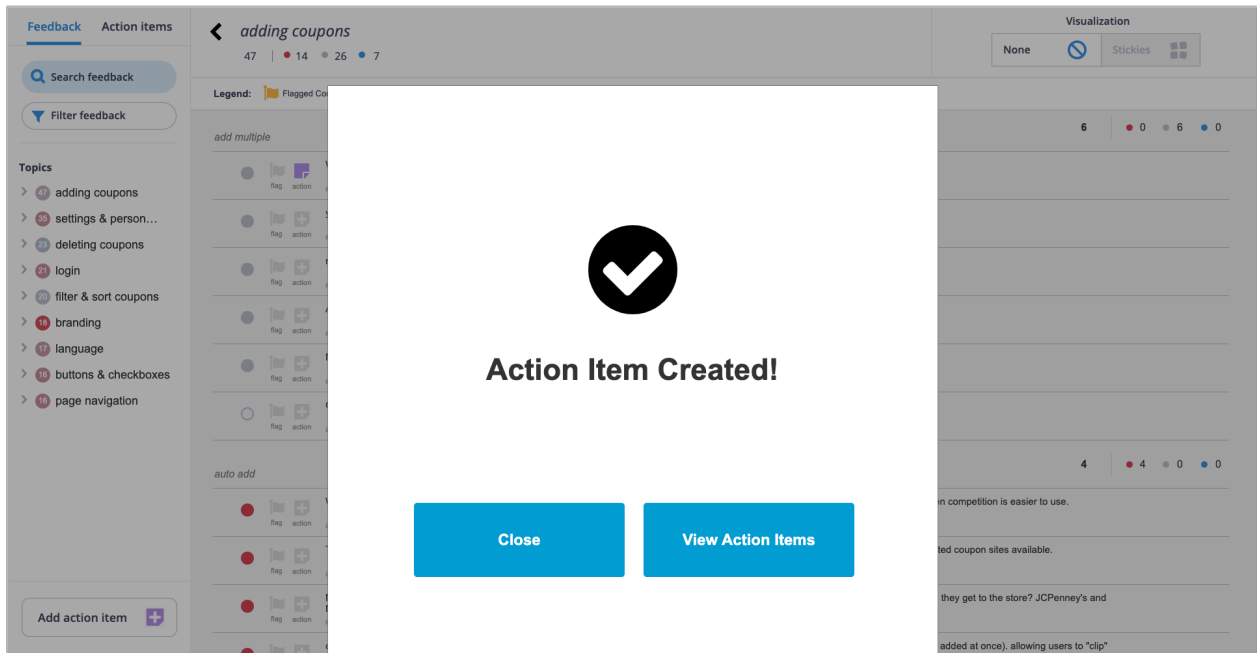
Action item creation confirmation screen.
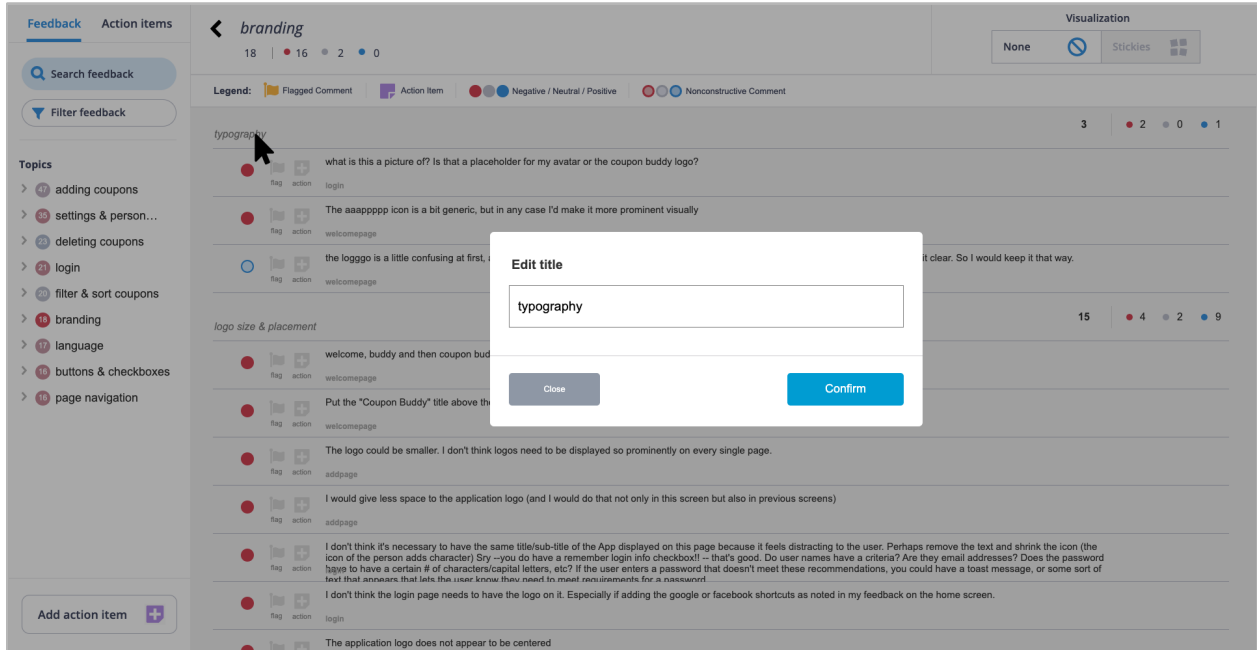
Subgroup title edit popup screen, triggered by clicking on subgroup title.
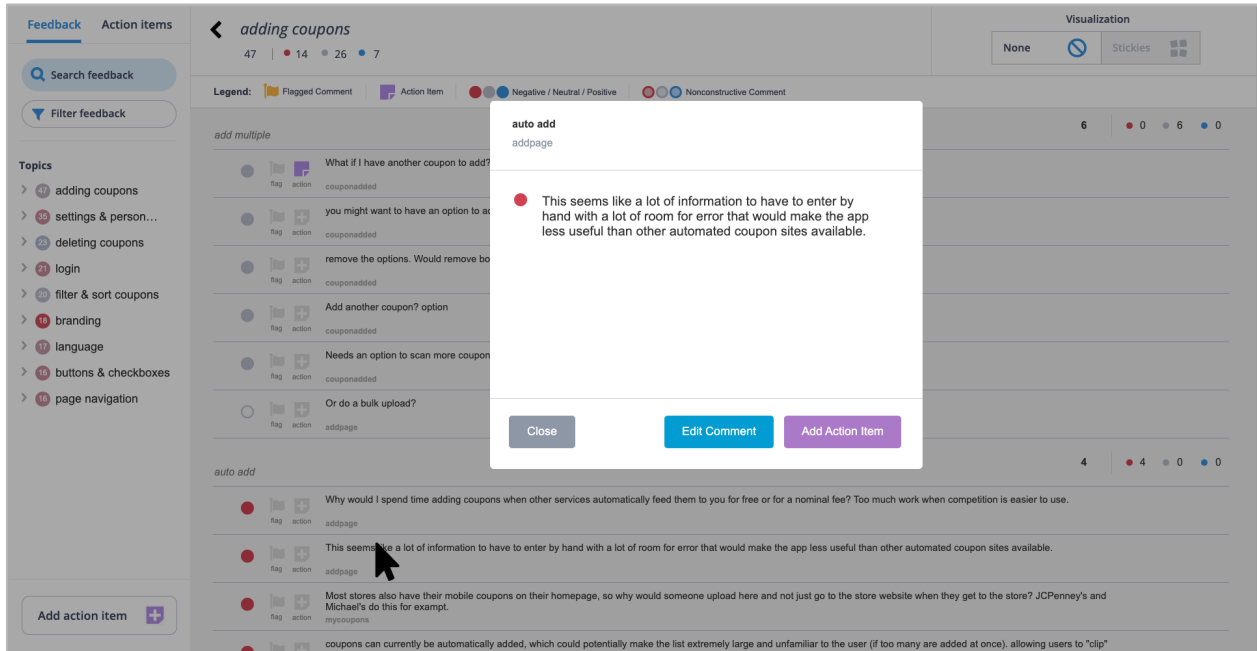
Expanded comment.
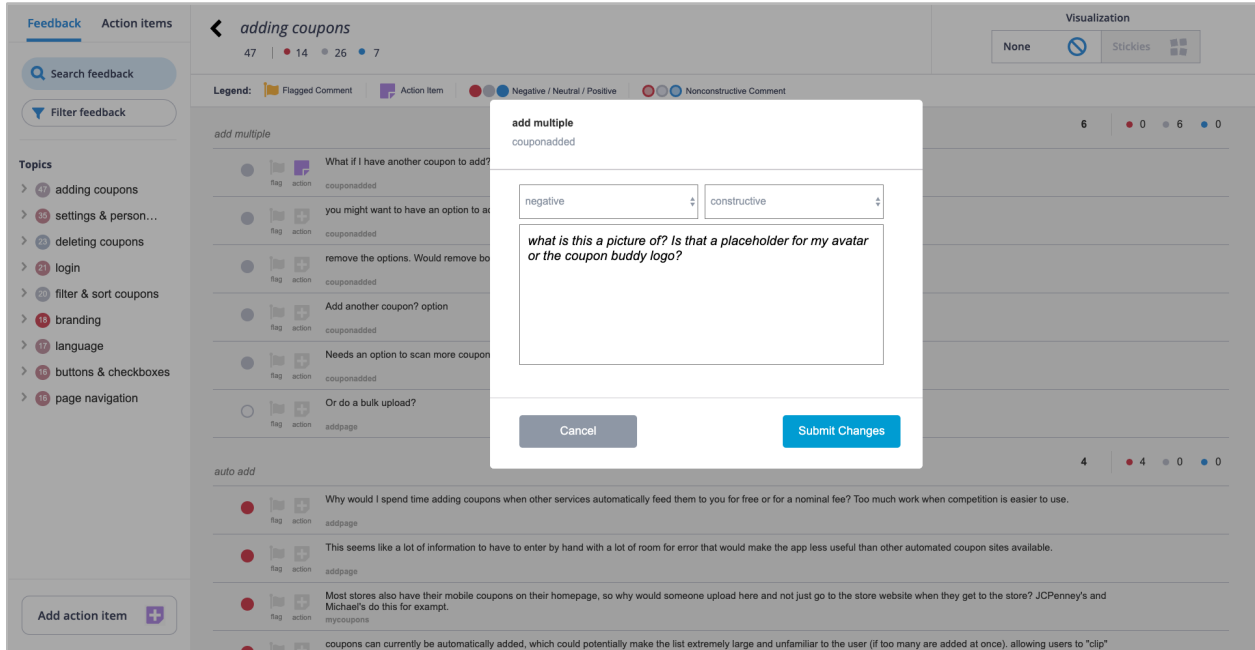
Edit comment.

Action items tab.

Feedback **Action items**

Share Action Items

All action items
General
Feature Requests
Bug Fixes
Visual Design
UX Improvements

Flagged comments

Add action item

Flagged comments

8comments flagged

what is this a picture of? Is that a placeholder for my avatar or the coupon buddy logo?
login

welcome, buddy and then coupon buddy listed below the picture are repetitive. Perhaps locate the "coupon buddy" below all the info.
welcomepage

Put the "Coupon Buddy" title above the logo or next to it. It gets lost as is currently
welcomepage

I don't think it's necessary to have the same title/sub-title of the App displayed on this page because it feels distracting to the user. Perhaps remove the text and shrink the icon (the icon of the person adds character) Sry --you do have a remember login info checkbox!! -- that's good. Do user names have a criteria? Are they email addresses? Does the password have to have a certain # of characters/capital letters, etc? If the user enters a password that doesn't meet these recommendations, you could have a toast message, or some sort of text that appears that lets the user know they need to meet requirements for a password.

I would move the buttons down a bit to give more emphasis to the logo.
welcomepage

Again, I would change the spacing so the fields and button are lower on the screen to give the logo more pop.
login

Making the logo smaller to allow the complete user input screen appears above the on-screen keyboard would be helpful.
login

Here again I don't think the logo serves any purpose other than to take up space. The user is in the app already know what the app is; no need to push the brand at the expense of screen real estate.
addpage

Flagged comments.

Appendix D

Coupon Buddy Wireframes
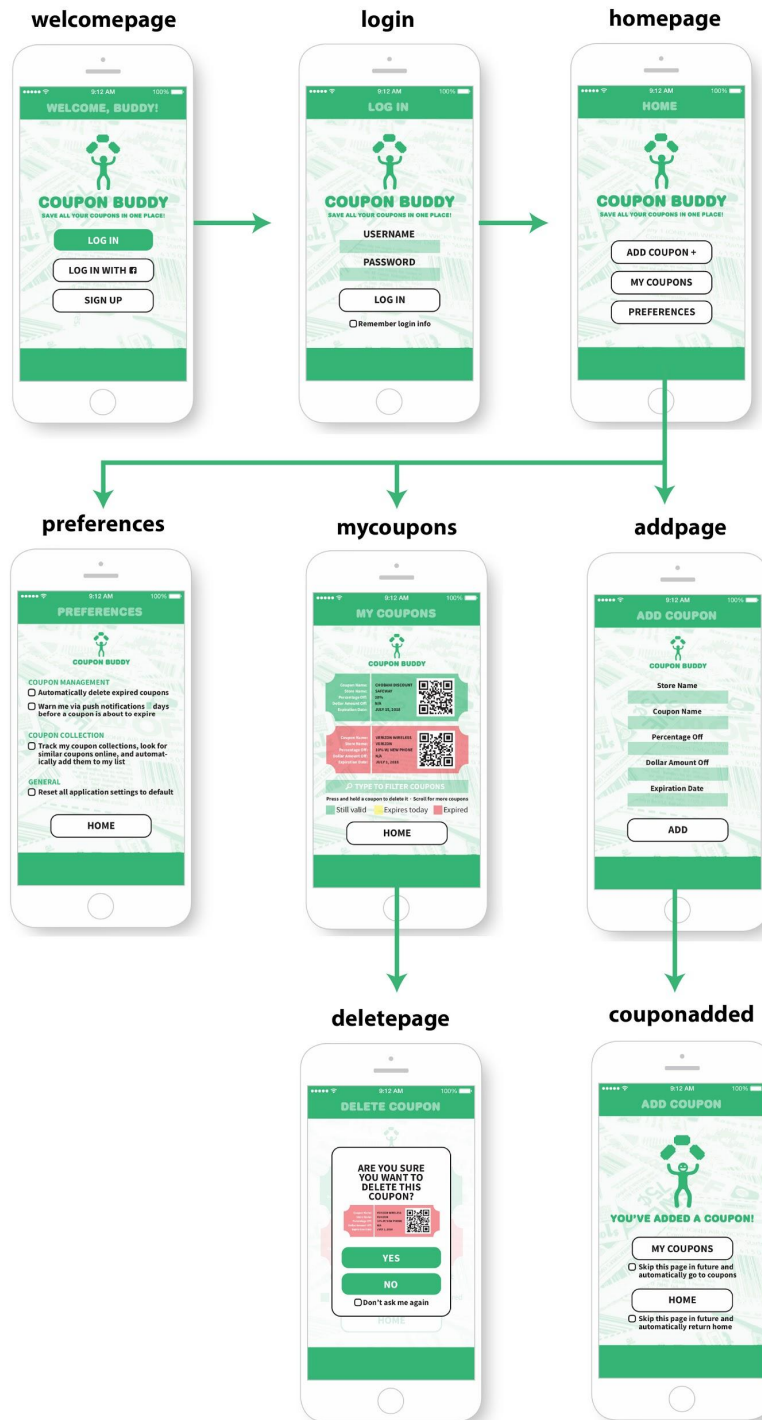
Appendix E

User Test Tasks

Task	Description
1	This is the landing page of the tool, providing an overview of the data. Without clicking, take a moment to view the page and describe what you see.
2	One of your other team members has already used this prototype to review this feedback. Find the notes he created about the feedback.
3	Without clicking, review this page and describe what you see. Then return to the previous page.
4	Find a way to view this data differently.
5	Looking at this data, you find that it's a bit too simplistic. You'd like to see this feedback arranged into more groupings. Find a way to do this.
6	Your team doesn't have the bandwidth currently to address issues concerning the layout, menu, or inputs. Find a way to remove feedback relating to these things from your view.
7	Now, locate the topic containing feedback about "scanning to upload" coupons.
8	Without clicking, take a moment to view the page and describe what you see.
9	This is too much feedback to review efficiently. We need to distill this feedback down a bit. Find a way to quickly remove all feedback which you perceive to be overly positive or otherwise unhelpful.
10	Create an action item addressing an important piece of feedback you see. And mark any comments which you would like to review later.
11	Now, find the topic containing feedback relating to the Coupon Buddy logo.
12	Check to see if the subgroup titles appropriately describe the comments. If not, correct the titles.
13	Review the comments. Check for any misspellings. And, confirm that the sentiment rating matches the opinions expressed in the comments. If you find any errors, correct them.
14	Create an action item addressing an important piece of feedback you see and mark any comments which you would like to review later.
15	Now, locate the topic containing feedback about relating to checkboxes.
16	Find a way to view this data differently.
17	Machine learning has done a less than perfect job organizing these comments. Find a way to have the comments in the "general" subgroup be in more appropriate subgroups.
18	Create an action item addressing an important piece of feedback you see. And mark any comments which you would like to review later.
19	Now, locate the topic containing feedback about relating to the back button.
20	Visualize the data and look for responses which don't belong in this topic. Find a way to store them in more appropriate topics.

21	Review the remaining feedback and confirm that they are in appropriate subgroups. If not, rearrange them as you see fit.
22	Create an action item addressing an important piece of feedback you see. And mark any comments which you would like to review later. After - return to the initial landing page.
23	You'd like to find feedback relating to "color coding" but you're not sure what topic it belongs in. Query the data to find this information.
24	Create an action item addressing an important piece of feedback you see. And mark any comments which you would like to review later. After - return to the initial landing page.
25	Find a quick way to determine which page has the most negative feedback. When you find this information, create an action item noting this page.
26	Your next meeting with the design team is about to start. Send your action items to Geoff, as he'll be running the meeting!
27	Before you go, confirm that the comments you flagged are stored for you to review later.

Appendix F

Exit Interview Questions

#	Question
1	Did the auto generated categories improve your ability to review feedback? Why?
2	Did the data visualizations improve your ability to review feedback?
3	Was it easy to navigate through the feedback? Why?
4	Did the search functionality improve your ability to review the feedback? Why?
5	Did the editing functionality improve your ability to review the feedback? Why?
6	Did the ability move and rearrange data improve your ability to review the feedback? Why?
7	Are there any important features you think are missing?
8	Are there other steps that you would have taken in reviewing this feedback?
9	What did you not like about this tool?
10	What did you like about this tool?
11	Could this tool work for a design team when reviewing feedback? Why?

Appendix G

User Test Results

Task	Priority Level	Success Rate	User Scores								Actions Taken by Users	Total
			P1	P2	P3	P4	P5	P6	P7	P8		
1	L	-	-	-	-	-	-	-	-	-	N/A	N/A
2	H	88%	1	1	1	1	1	1	1	3	Clicks action item tab	8
											Locates action items	7
											Clicks filter button	1
3	L	-	-	-	-	-	-	-	-	-	N/A	N/A
4	H	100%	1	1	1	1	1	1	1	1	Toggles visualization option	8
5	H	50%	1	1	1	3	3	3	1	2	Increases number of topics	5
											Toggles data arrangement buttons	3
											Selects visualization option	3
											Navigates to category page	1
6	H	63%	1	1	1	1	2	1	3	2	Clicks hide icon	7
											Clicks filter button	3
											Clicks search button button	1
											Clicks side navigation dropdown item	1
											Navigates to category page	1
											Toggles visualization option	1
7	H	88%	1	1	1	1	3	1	1	1	Clicks on topic in visualization	4
											Clicks side navigation dropdown item	2
											Toggles to list visualization	1
											Toggles data arrangement buttons	1
											Clicks search button button	2
8	L	-	-	-	-	-	-	-	-	-	N/A	N/A
9	H	25%	3	3	1	1	2	2	2	3	Clicks hide icon	3
											Clicks legend/metrics	3
											Clicks filter button button	5
											Toggles visualization option	1
10	H	100%	1	1	1	1	1	1	1	1	Clicks action item icon	7
											Clicks action item button	2
											Links quote, after cicking action item button	1
											Clicks legend	1
											Click flag icon	7
											Clicks to view action items	3
											Categorizes comment	7

Task	Priority Level	Success Rate	User Scores								Actions Taken by Users	Total
			P1	P2	P3	P4	P5	P6	P7	P8		
11	H	100%	1	1	1	1	1	1	1	1	Clicks category in side navigation	5
											Clicks side navigation dropdown item	2
											Clicks on topic in visualization	3
											Clicks on subgroup title, edit window	1
12	H	88%	1	1	1	1	1	1	1	3	Clicks to edit subgroup title	7
											Clicks to expand comment	1
											Attempts to drag list comment	1
13	H	88%	1	1	1	1	2	1	1	1	Clicks to expand comment	8
											Clicks button	8
											Edits content	8
											Edits sentiment	8
											Edits quality	5
14	M	100%	1	1	1	1	1	1	1	1	Clicks action item icon	7
											Clicks action item icon	1
											Click flag icon	5
15	M	88%	1	1	1	1	1	1	2	1	Clicks on topic in visualization	2
											Clicks category in side navigation	5
											Clicks search button button	1
											Clicks filter button	1
16	H	100%	1	1	1	1	1	1	1	1	Selects visualization option	8
17	H	100%	1	1	1	1	1	1	1	1	Drags notes into new subgroup	5
											Drag notes into other existing subgroups	8
18	H	88%	1	1	1	1	1	1	1	2	Clicks expand icon, then action item button in popup window	7
											Click flag icon	3
											Clicks action item button	2
											Clicks on legend action item icon	1
											Links quote, after cicking action item button	1
19	M	63%	1	1	2	1	1	2	1	2	Clicks on topic in visualization	1
											Clicks category in side navigation	3
											Clicks side navigation dropdown item	1
											Clicks search button button	4

Task	Priority Level	Success Rate	User Scores								Actions Taken by Users	Total
			P1	P2	P3	P4	P5	P6	P7	P8		
20	H	25%	1	3	2	2	1	3	3	3	Drags comment into side navigation category	4
											Drags comment to create new topic	4
											Expand comment to change categorization	4
											Clicks filter button	1
											Clicks side navigation dropdown item	1
21	H	75%	1	1	2	1	1	1	3	1	clicks and drags to create subgroups	6
											drags to create top level topic	1
											drags into top level topics	1
22	M	100%	1	1	1	1	1	1	1	1	Clicks action item icon	4
											Clicks action item icon	4
											Links quote (if button was pressed)	1
											Click flag icon	5
23	H	75%	2	2	1	1	1	1	1	1	Clicks filter button	2
											Navigates to another category	1
											Clicks search button button	8
24	M	100%	1	1	1	1	1	1	1	1	Clicks action item icon	6
											Clicks action item icon	2
											Click flag icon	4
25	H	38%	3	1	1	3	1	2	2	2	Clicks to arrange data by 'page'	7
											In 'page' arrangement, clicks to view the list visualization	1
											In 'page' arrangement, clicks to view the treemap or barchart	4
											Clicks action item icon	4
											Links quote (if button was pressed)	1
											Clicks action item icon from "page" category page	1
											Clicks filter button	3
											Clicks on legend	2
26	H	100%	1	1	1	1	1	1	1	1	Click share button	8
27	H	75%	1	3	2	1	1	1	1	1	Clicks flagged comments category	7

References

Alper, B., Yang, H., Haber E., & Kandogan, E. (2011). Opinionblocks: Visualizing consumer reviews. IEEE VisWeek Workshop on Interactive Visual Text Analytics for Decision Making, 2011

Carvalho, E. S., & Chaves, M. S. (2012). Exploring User-generated Data Visualization in the Accommodation Sector. *2012 16th International Conference on Information Visualisation*. doi: 10.1109/iv.2012.42

Dork, M., Riche, N. H., Ramos, G., & Dumais, S. (2012). PivotPaths: Strolling through Faceted Information Spaces. *IEEE Transactions on Visualization and Computer Graphics,18*(12), 2709-2718. doi:10.1109/tvcg.2012.252

Gamon, M., Aue, A., Corston-Oliver, S., & Ringger, E. (2005). Pulse: Mining Customer Opinions from Free Text. *Lecture Notes in Computer Science Advances in Intelligent Data Analysis VI*, 121–132. doi: 10.1007/11552253_12

Gretarsson, B., O'Donovan, J., Bostandjiev, S., Höllerer, T., Asuncion, A., Newman, D., & Smyth, P. (2012). TopicNets. *ACM Transactions on Intelligent Systems and Technology,3*(2), 1-26. doi:10.1145/2089094.2089099

Hu, M., Yang, H., Zhou, M. X., Gou, L., Li, Y., & Haber, E. (2013). OpinionBlocks: A Crowd-Powered, Self-improving Interactive Visual Analytic System for Understanding Opinion Text. *Human-Computer Interaction – INTERACT 2013 Lecture Notes in Computer Science*, 116–134. doi: 10.1007/978-3-642-40480-1_8

Käki, M. (2005). Optimizing the number of search result categories. *CHI 05 Extended Abstracts on Human Factors in Computing Systems - CHI 05*. doi:10.1145/1056808.1056955

Luther, K., Tolentino, J., Wu, W., Pavel, A., Bailey, B., Agrawala, M., et al. (2015). Structuring, Aggregating, and Evaluating Crowdsourced Design Critique. *UC Berkeley: Center for Information Technology Research in the Interest of Society (CITRIS)*. Retrieved from https://escholarship.org/uc/item/7ck2f48g

Oelke, D., Hao, M., Rohrdantz, C., Keim, D. A., Dayal, U., Haug, L.-E., & Janetzko, H. (2009). Visual opinion analysis of customer feedback data. *2009 IEEE Symposium on Visual Analytics Science and Technology*. doi: 10.1109/vast.2009.5333919

Smith, A., Hawes, T., & Myers, M. (2014). Hiearchie: Visualization for Hierarchical Topic Models. *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*. doi:10.3115/v1/w14-3111

Soriano, L. T., & Palaoag, T. D. (2018). A Machine Learning-Based Topic Extraction and Categorization of State Universities and Colleges (SUC) Customer Feedbacks. *Proceedings of the 8th International Conference on Information Communication and Management - ICICM 18*. doi:10.1145/3268891.3268897

Wetzker, R., Umbrath, W., Hennig, L., Bauckhage, C., Alpcan, T., & Metze, F. (2008). Tailoring Taxonomies for Efficient Text Categorization and Expert Finding. *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. doi:10.1109/wiiat.2008.179

Wu, Y., Wei, F., Liu, S., Au, N., Cui, W., Zhou, H., & Qu, H. (2010). OpinionSeer: Interactive Visualization of Hotel Customer Feedback. *IEEE Transactions on Visualization and Computer Graphics,16*(6), 1109-1118. doi:10.1109/tvcg.2010.183

Xu, A., & Bailey, B. P. (2011). A crowdsourcing model for receiving design critique. *Proceedings of the 2011 Annual Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA 11*. doi: 10.1145/1979742.1979745

Xu, A., Huang, S., & Bailey, B. (2014). Voyant. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW 14*. doi:10.1145/2531602.2531604

Ziegler, C.-N., Skubacz, M., & Viermetz, M. (2008). Mining and Exploring Unstructured Customer Feedback Data Using Language Models and Treemap Visualizations. *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. doi: 10.1109/wiiat.2008.69