Using Visualizations to Encourage Blind-Spot Exploration

Jayachithra Kumar
Delft University of Technology
j.kumar-1@student.tudelft.nl

Nava Tintarev
Delft University of Technology
n.tintarev@tudelft.nl

ABSTRACT

In this paper, we help users to better understand their consumption profiles by exposing them to their unexplored regions, thereby indirectly nudging them to diverse exploration. We refer to these regions as a user's *blind-spots*, and we visualize these by enabling comparisons between a user's consumption pattern with that of other users of the system. We compare the effectiveness of two visualizations – a bar-line chart and a scatterplot — for increasing a user's intention to explore new content. The results suggest that users can understand both visualizations. Furthermore, our results confirmed that users with higher understanding of their profile tend to explore their blind-spot categories more. This experiment is a first step towards increasing user's awareness of their choices as well as providing the kind of user control that encourages users to explore new types of items.

CCS CONCEPTS

• **Information systems** → **Decision support systems**; • **Human centered computing** → **Human computer interaction (HCI)**;

KEYWORDS

Visualization, Recommender Systems, Blind Spots, Filter Bubble, Scatterplot, Bar-line chart

1 INTRODUCTION

While personalized recommendations can help people to cope with the information overload problem, over time, using recommender systems can decrease the diversity of content that we consume [15], thereby, limiting our exposure to some novel content, views and opinions contrary to our own. Our current preferences often reflect our past preferences, and our behaviors may also interact with online filtering and ranking algorithms to further narrow our views. This phenomena of algorithmic narrowing, or over-tailoring, is called 'filter bubbles' [3, 6, 16]. However, there may be design choices for recommender systems that could decrease over-tailoring. Flaxman et al. found evidence that recent technological changes both increase and decrease various aspects of over-tailoring [7].

This work addresses this possibility by helping users understand the limitations of their consumption patterns using visualizations. Specifically, we propose a novel approach for recognizing 'blind-spots' in user profiles - regions of the preference space that are under-represented - and describe techniques for revealing these blind-spots to users. By helping users to recognize these blind-spots, we also study if this has a demonstrable effect on their consumption; whether this recognition encourages them to further explore the recommendation space. In the following sections, we describe the results of a user experiment to evaluate the efficacy of two

different techniques for blind-spot visualization and its effect on users' exploration of the recommendation space.

In the next section, we describe related work. This is followed by a description of the method used to generate the visualizations used in this study (Section 3). Next in Section 4, we describe a lab study with 23 participants which investigates the relationship between understanding the visualizations and exploring blind-spot regions in a user's profile. Section 5 outlines the main results. This is followed by discussion of qualitative findings, and post-hoc analysis for surprising results in Section 6. We conclude with suggestions for future work in Section 7

2 RELATED WORK

This work sits at the intersection between two important recommender systems themes: 1) the use of visualization to aid transparency and explanation, and 2) techniques for dealing with filter bubbles. One important objective of this work is to increase user's awareness of their filter-bubble, to improve decision making by better informing users about their consumption pattern. To help users understand their own consumption patterns we propose an approach for visualizing user profiles. This builds on work for visualizing consumption blind-spots in movie recommender systems [18], and visualizing consumption profiles in music [11].

When it comes to mitigating filter-bubbles, there are two common responses in the literature. *The first approach* is to develop recommendation algorithms that are more responsive to the risks inherent in the filter bubble. This can be achieved by focusing on tuning algorithms to increase beyond accuracy aspects (such as diversity, serendipity, coverage and novelty), in addition to relevance of recommendations (c.f., [1, 2, 4, 17, 20]), and re-ranking recommendation lists to include diversity in an optimization function (c.f., [12, 19]).

While improving recommendation diversity can go some way to coping with the filter bubble, it is far from a complete solution. For example, it does not increase user awareness of the filter bubble itself. *A second approach* helps users to better understand the available options – the recommendation space – so as to inform them about the compromises that are inherent in any set of recommendations, relative to a wider set of items. In this regard, the work of [14, 18] is pertinent, showing how visualization was found to increase user awareness of the filter bubble, understandability of the filtering mechanism, and a user's sense of control.

In this paper, we address the blind-spot issue by showing the consumption behaviours of users, and highlighting blind-spots that may exist in their consumption relative to a larger user population. We further study whether by making users aware of their blind-spots, we may be able to influence them to explore items in the under-explored parts of their catalogue.

3 METHOD

In this section, we provide a brief overview of the stages involved in the extraction and visualization of consumption pattern. With our visualization we aim to give users a holistic view of their filter-bubble by enabling them to compare *their* consumption pattern (user profile) with the (aggregate) consumption pattern of other users of the system ('global' consumption pattern or 'global' profile). However, in doing so, we do not aim to explain individual items to users, but rather highlight the important aspects of their profile as a whole (i.e., by grouping tracks based on genres). That way visualization could scale better and still provide an accurate representation of global and user's preferences.

In comparing global and user's preferences, we not only enable comparisons between different categories, but also within the same category between user and global profiles (i.e., within the same genre, we highlight the differences between user's preferences and global preferences). To further emphasize significant categories, in addition to representing a range of categories, we also represent interaction between these categories (i.e., when a track belongs to more than one genre). This enables us to highlight a user's most familiar categories thereby increasing their trust in the visualization. In the following sections we describe the design decisions that went into the extraction of consumption data and creation of visualizations.

Figure 1 provides a brief overview of the stages involved in the extraction and visualization of consumption pattern. Steps 1 & 2 involve feature extraction and data collection respectively. Step 3 involves extracting global and local preferences using frequent item-set mining algorithm. Once the global and local preferences are extracted, visualizations are constructed to represent this data (step 4). The following sections describe in detail, the design decisions that went into each of these stages.
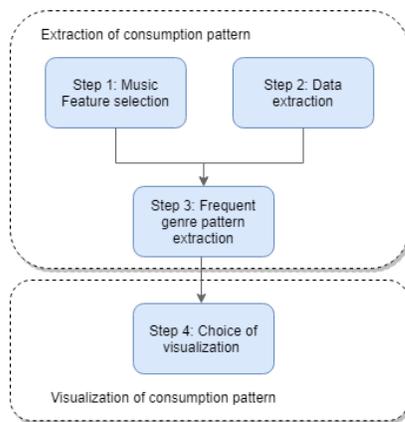

Figure 1: Steps involved in the extraction and visualization of consumption pattern

3.1 Feature Selection

For visualization, we categorize tracks based on their genre tags. Genres provide a good collective representation of a user's preferences compared to other acoustic features (such as tempo, pitch

etc). Besides, users can easily relate to a genre-based categorization, since it is used in existing recommender systems like Spotify.

In addition to providing genre-level categorization between user and global profiles, it is also important for the system to be able to distinguish between items in the same genre, between user and global profiles. In order to achieve this, a second dimension is added to the visualization. To select the most representative feature we looked into the Million Song Dataset (MSD) which provides a total of 55 features for each track, and we chose the feature 'Artist hotness'. 'Artist hotness' is a value (0 to 1) assigned by MSD for each artist, which corresponds to how much buzz the artist is getting right now. This value is computed algorithmically based on information derived from several undisclosed sources, including mentions in the web, mentions in music blogs, music reviews, play counts etc. In comparison to other features, artist hotness is proven to provide a stable representation of user's preferences [10].

3.2 Data extraction

We used the Million Song Dataset [13], which is the largest available music feature dataset containing audio features, song and artist meta-data for a million contemporary music tracks. It is also the only dataset that provides artist hotness value for tracks.

To obtain global consumption pattern, we used one of the complementary datasets of MSD, the 'Taste Profile Subset' (TPS) and merged this with the MSD dataset. TPS dataset provides a list of tracks listened by a number of users of last.fm, along with the play count of these tracks. We retained users who listened to at least 20 tracks. The artist hotness values of these tracks were obtained by merging TPS and MSD. Since MSD does not provide genre information for tracks, we obtained this information from a third dataset provided by tagtraum [8].

To build a user's individual profile, we obtained a specific user's real time music listening pattern from Spotify. Similar to global profile, this entails all the track preferences of the user, the genre/genre-combinations, and artist hotness values of these tracks. We used Spotify since it is the only API that provides all these three required information for research. Besides, Spotify is one of the largest music service providers, and hence it is relatively easy to find real users for evaluation.

3.3 Frequent genre-set extraction

We applied frequent itemset mining algorithm (RElim, [5]) in order to obtain the most frequently listened genres/genre-combinations. Frequent itemset mining algorithms work by identifying all common sets of items in a given list, and it is used for discovering regularities between frequently co-occuring items in large datasets.

We used 'Recursive Elimination Algorithm' (RElim) provided by the 'pymining' package of Python. For both global and user's profile, this algorithm gives a set of most frequently consumed genres/genre-combinations and their frequency values (i.e., how many times the item appears in the profile). By visualizing this information we believe that we can enable users to compare their consumption pattern with the global consumption pattern and subsequently to identify the blind-spots in their profiles.

RElim was parameterized at a minimum frequency value (minimum *support*) of 2. This means that all itemsets that occur less than

two times will be eliminated from the global profile and user profile. This support value was chosen to ensure a faster computation time while still preserving significant genres.

Table 1 shows the top 20 most frequent genre/genre-combinations along with their (normalized) frequencies, for the global data set. Certain genres ('Rock', 'Pop') are highly preferred globally compared to others. We also notice that certain genre-combinations are preferred more than other individual genres. For example, 'Alternative, Rock' has higher frequency compared to Rap or Metal. For each of the top-20 genre/genre-combinations, we compute the average artist hotness values of all the tracks listened in that genre (Table 1). For all the genres, the average artist hotness value lies closer to the center (0.5) which accounts for the diverse music consumption of users.

Table 1: Top 20 frequent item-sets for global dataset with a minimum support value of 2. 'Alt' represents 'Alternative'

Genres (1-10)	Frequency	Genres (11-20)	Frequency
Rock	0.308	Metal	0.029
Pop	0.108	Rock, Punk	0.029
Alt	0.075	Rock, Metal	0.028
Alt, Rock	0.071	Country	0.027
Hip-Hop	0.038	Dance	0.023
Electronic	0.036	Rock,Pop	0.023
Rap	0.032	Alt,Punk	0.021
Rap, Hip-Hop	0.032	Alt,Rock,Punk	0.021
R&B	0.032	Latin, Indie	0.020
Punk	0.029	Indie	0.020

3.4 Choice of Visualization

For our visualizations we represent the top-20 most frequent genre-sets for user and global profiles. The choice of visualizations was made based on their ability to represent all the required dimensions (i.e., genre/genre-combinations, frequency of genres and average of artist hotness values for each genre, for top 20 genres), to span across global and user profiles, and to be able to represent all required data points. We used scatterplot as our main visualization and we compare the performance of scatterplot with the baseline - bar-line chart. In this section, we describe both these charts.

3.4.1 Visualization 1: Scatterplot. Scatterplot is the type of chart in which data is represented as a collection of points, with each point having the value of its first variable determining its position along the horizontal (x-) axis, and the second variable determining its position along the vertical (y-) axis. Traditional scatterplots are capable of representing only two dimensions, however, with the inclusion of visual attributes such as color, size and shape, it is possible to represent up to five dimensions.

An example scatterplot used in our study is shown in Figure 2. Here the size of the bubbles represent the frequency of the item sets. So larger the bubble, higher the frequency of the genre corresponding to that bubble. To distinguish between genres we use color hues. The horizontal orientation of a bubble represents its average artist hotness value and its vertical orientation represents whether

it belongs to the global profile or the user's profile ('yours' label). We also implemented a hover feature wherein on hovering over a bubble, the genre corresponding to the bubble gets highlighted in both global and user profile. This enables easy comparison between both profiles. Furthermore, on hovering over a bubble, the genre name, frequency and average artist hotness value corresponding to the bubble gets displayed. From the given visualization, we can infer the following:

(1) For the given user, *Pop* is the most frequently consumed genre, since it corresponds to the largest bubble under 'yours' category of vertical axis.
(2) Pop is also highlighted under the global category, which means that it is also globally one of the most (but not *the* most) frequent genre(s).
(3) The user prefers more popular artists compared to the average user of the system since the user's bubbles are generally aligned more towards the right.

3.4.2 Visualization 2: Bar-line chart. We compare the performance of scatterplot with the base-line visualization bar-line chart. Bar-line chart is a combination of bar chart and line chart and it can represent up to three variables. A bar chart based visualization was chosen as the base-line for the following reasons:

(1) It is proven to be the most compelling and persuasive means to convey explanations in recommender systems [9].
(2) It is used in existing recommender systems such as Movie-Lens[1] to represent user's ratings across genres, and frequency of ratings (Figure 3).

Figure 4 shows an example of a bar-line chart. Here the horizontal axis represents the itemset name; left-vertical axis corresponding to the line-chart represents the genre frequency, and right-vertical axis corresponding to the bar-chart represents the average artist hotness of the genre. Unlike scatterplot, a single bar-line chart is not capable of showing both user's and global data, and hence we use two separate charts. This separation might make comparison between both profiles cumbersome, but it accounts for the relative simplicity of the chart.

4 EXPERIMENT

We performed an online evaluation of our system to compare the effectiveness of visualizations, and to study changes in user's preferences. For ease of explanation, we divide our evaluation process into two conceptual stages: Stage 1 - where we evaluate user's understanding of visualizations, and Stage 2 - where we observe a user's music exploration pattern after they are exposed to their blind-spots. It is important to note here that this classification is introduced solely for the purpose of better representation of concepts, and from participant's perspective the whole evaluation process is staged as a single experimental session. In the following sections, we explain the experimental design and research hypotheses for Stage 1 and Stage 2 in Sections 4.1 and 4.2 respectively. We then brief about the materials (Section 4.3) and detailed procedures (Section 4.4) involved in the study.

---

[1]https://movielens.org/, retrieved June 2018
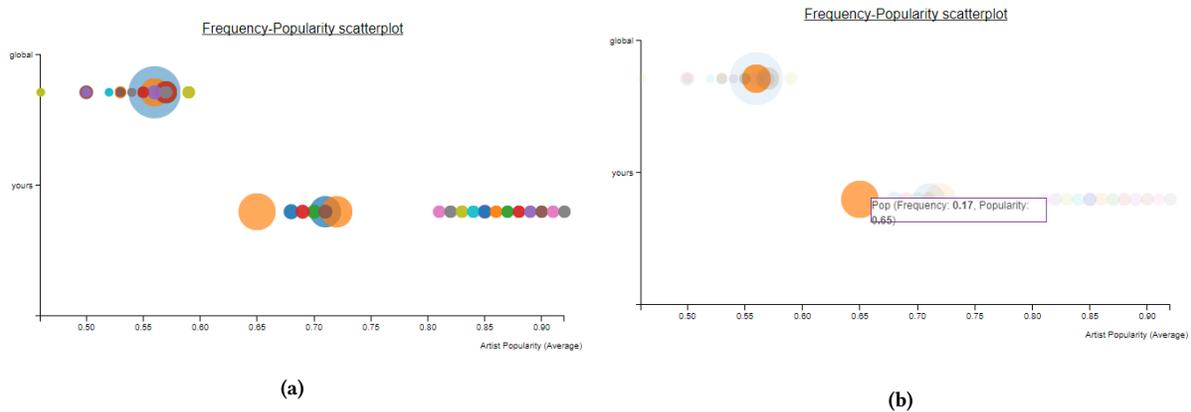
(a)

(b)

Figure 2: (a) Example scatterplot visualization used in the study. (b) On hovering over a bubble its corresponding genre name gets highlighted in both global and user's profile.
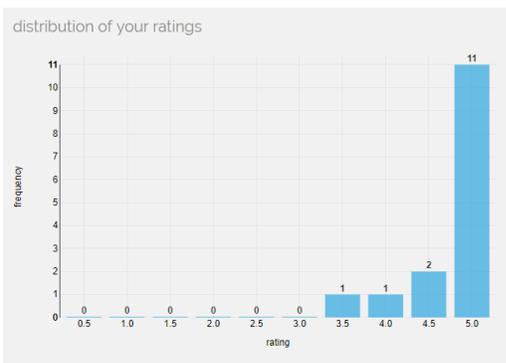

Figure 3: Example bar chart visualization used in MovieLens system to represent the distribution of ratings (retrieved June 2018).

4.1 Stage 1: To study the understanding of visualization

4.1.1 Design. For stage 1 of our evaluation, we used a *within-subjects* repeated measures design, where each participant was presented with both scatterplot and bar-line chart. In order to minimize order effects we performed counterbalancing by changing the order of visualization for each participant.

4.1.2 Independent variable. For each user we show both types of visualizations (bar-line chart and scatterplot), and study the effectiveness of each of these visualizations in increasing the understanding of a user's consumption pattern and blind-spots. Hence *type of visualization* is our independent variable.

4.1.3 Dependent variables.

(1) **Correctness of understanding**: Understandability of a visualization is measured by asking users to answer questions about information represented in the visualization. These questions test a user's understanding of their consumption pattern and their blind-spots.

(2) **Confidence**: In addition to measuring user's actual understanding, we also measure the perceived ease of understanding for both the visualizations. These are self-suggested confidence scores provided by the user for each question about their consumption pattern and blind-spots. It says how confident the users are in the answers they provide.

4.1.4 Hypotheses.

- **H1**: Users are able to answer questions about their consumption pattern more accurately with scatterplot than with bar-line chart.
- **H2**: Users have more confidence in their answers about their consumption pattern for scatterplot more than bar-line chart.
- **H3**: Users are able to answer questions about their blind-spots more accurately with scatterplot than with bar-line chart
- **H4**: Users have more confidence in their answers about their blind-spots for scatterplot more than bar-line chart.

4.2 Stage 2: To study user's music exploration

4.2.1 Design. For stage 2, we perform a simple **correlation analysis** to study the relation between a user's understanding of their profile and their music exploration pattern.

4.2.2 Independent variable. For all users, we measure if their understanding in their profile has an impact on their exploration of blind-spot genres. Hence a user's **correctness of understanding** is the independent variable. This value is directly computed for each user as a dependent variable in Stage 1 (Section 4.1.3).

4.2.3 Dependent variable. **Exploration factor**: Exploration factor is measured for each user by computing the proportion of tracks the user has explored in their blind-spot category, and it quantifies a user's exploration in that category.

4.2.4 Hypothesis.

- **H5**: Users who score higher for their questions about their blind-spots, explore their blind-spot genres more.

(a) All users: 'Global'
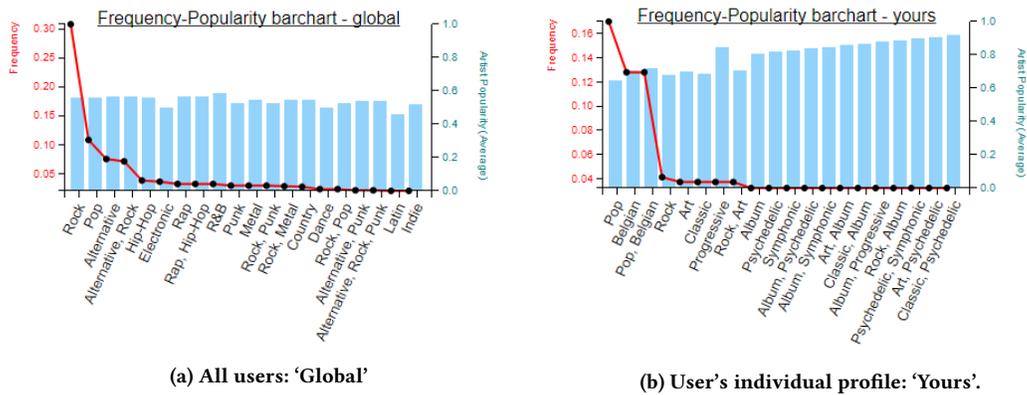

(b) User's individual profile: 'Yours'.

Figure 4: Example bar-line chart visualization used in the study.

4.3 Materials

Visualizations were designed using D3.js Javascript visualization library [2]. The online interfaces for web-based survey were developed using Python Flask framework[3].

4.4 Procedure and Tasks

Each participant goes through six steps to complete the experiment. Participants start by providing their basic demographic information (Step 1) after which, they log in with their Spotify account (Step 2). From the user's account, we collect his/her top 50 tracks using spotify's API. We then use frequent pattern mining algorithm (Section 3.3) on the genres of these tracks to compute the user's top 20 frequent genres/genre-combinations, and their average artist hotness values.

In the next two steps, users are presented with each of the two visualizations (bar-line chart and scatterplot), accompanied by a set of instructions on how to read the visualization. After a minimum buffer time of 1 minute to read and understand the visualization, questionnaires are shown below the visualization for the users to answer. The questionnaire is designed in such a way that, for each visualization, they evaluate user's understanding of the system in all four aspects - global consumption pattern, user's consumption pattern, user's blind-spots and artist hotness values. More particularly, we ask users to identify - the top *first* and *second* highest consumed genres (globally and in their profile, i.e., 2x2 = 4 questions), their top *first* and *second* highest blind-spots (i.e., genres with high frequency in global profile but not found in their profile, 2 questions) and the artist hotness values of the all genres chosen for these questions (6 questions). In order to reduce the learning effect, we split the above 12 questions, and performed counterbalancing to assign half of the questions for each chart. For each question, the user is also asked to provide their confidence in their answers in a 5-point Likert scale.

Once users examine both visualizations, in the next step, we study user's music exploration pattern. We provide an interface where users can listen to music from different genres and genre

combinations from their blind-spot and frequent genre categories. More specifically, users are asked to select one or more genres to listen to from these categories. Based on their chosen genres, songs are recommended using Spotify's recommendation API. Users are asked to listen to tracks that they find interesting, and if they like any track they are asked to "add" it to their list. Our interface (Figure 6) was inspired by Spotify's old exploration interface (Figure 5). We use color coding to differentiate user's frequent and blind-spot genres (green = frequent, red = blind-spots).

Figure 5: Original interface from Spotify that lets users select multiple genres

Once users have listened and rated songs for at least five genre/genre-combinations, in the final-step users fill-out a post-stage assessment survey. Here users are provided with a set of questions to test their overall impression of the visualizations used in the study, with respect to their perceived - ease of understanding, ease of interaction, usefulness and interest. The answers are collected in a five-point likert scale.

5 RESULTS

In this section, we summarize the results of our online experiment with respect to our proposed hypotheses.

5.1 Participants

There were a total of 23 participants. 83% of the participants (n = 19) were male and 17% female (n = 4). Participants were between age-groups 19-35. 20 participants had computer science background

[2]https://d3js.org, retrieved March 2018
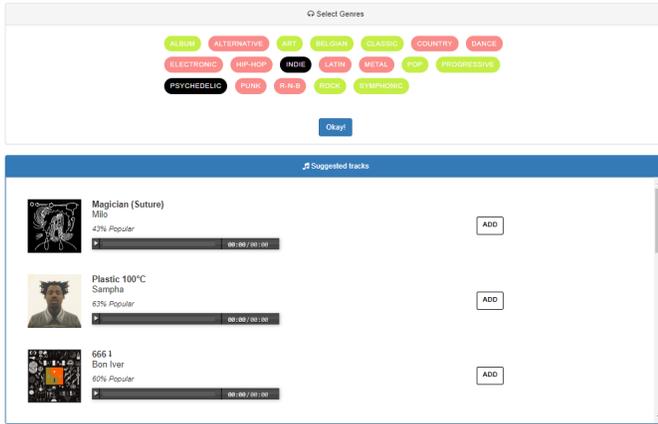[3]http://flask.pocoo.org, retrieved March 2018

Figure 6: Exploration interface used in the study

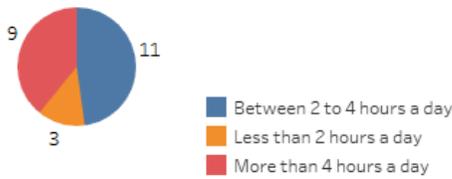(PhD and MSc). They all had diverse music backgrounds and music consumption behavior (Figure 7).

Figure 7: Participant's demographics by their music listening frequency

5.2 Understandability 1: Genres

Participants were asked to identify their first and second most consumed genres. Understandability was measured by how accurately participants could identify these genres. For each answer, a score was provided based on its correctness. For example, when answering about their first most consumed genre, a score of 1 is assigned if the answer is right, a score of 0.5 is assigned if the participant provided the name of their second most consumed genre, and a score of 0 is assigned for all other wrong answers.

The average scores for all participants for identifying their first and second most consumed genre, and the *artist hotness* values of these genres, is given in table 2. The difference in the mean scores are not statistically significant (Mann-Whitney U-test at p<0.05). Thus the results provide **no support** for hypothesis 1, which stated that participants would be able to answer questions about their consumption pattern more accurately with scatterplot than with bar-line chart.

5.3 Confidence 1: Genres

Participants were asked to provide their confidence values in their answers for identifying their first and second most consumed genres. The average scores are summarized in Table 3. The trends show that the participants had higher confidence with bar-line chart for

Table 2: Mean (std) scores of all participants for identifying their first and second most consumed genres and their artist hotness values. Scores lie between 0 to 1 with 1 being the highest score representing maximum understanding. * indicates significance (p<0.05).

	Bar-line	Scatter
1st place (frequency)	1.00 (0)	0.95(0.14)
2nd place (frequency)	1.00 (0)	0.92(0.29)
1st place (artist hotness)	0.87 (0.31)	1.00 (0.00)
2nd place (artist hotness)	1.00 (0)	0.92 (0.29)

identifying their first most consumed genre. For their second most consumed genre, they had higher confidence with scatterplot. However, the results of statistical tests show that the obtained scores are significant for identification of artist hotness values of **first most consumed genre** (Mann-Whitney U-test, U-value = 29, p<0.05). Hypothesis 2 predicted that participants will have higher confidence for answers about their genres with scatterplot more than bar-line chart. The trend is significant in the reverse direction and the hypothesis is consequently discarded.

Table 3: Mean (std) confidence scores of all participants for identifying their first and second most consumed genres and their artist hotness values. Scores lie between 1 to 5 with 5 representing maximum confidence. * indicates significance (p<0.05).

	Bar-line	Scatter
1st place (frequency)	4.60 (0.49)	4.00 (0.89)
2nd place (frequency)	4.18 (0.75)	4.45 (0.96)
1st place (artist hotness)*	4.90 (0.28)	4.27 (0.64)
2nd place (artist hotness)	4.30 (0.67)	4.70 (0.49)

5.4 Understandability 2: Blind-spots

Participants were asked to identify their first and second highest blind-spots. For each answer, a score of 1 is assigned if the answer is right, a score of 0.5 is assigned if the participant provided the second best answer, and a score of 0 is assigned for all other answers.

The average scores for all participants for identifying their first and second highest blind-spot and their artist hotness values are shown in Table 4. The average scores are slightly higher for scatterplot than for bar-line chart, but the results are not statistically significant. Hence hypothesis 3, which stated that participants would be able to answer questions about their consumption pattern more accurately with scatterplot than with bar-line chart, **is not confirmed**.

Table 4: Mean (std) scores of all participants for identifying their top 2 blind-spot genres and their artist hotness values. Scores lie between 0 to 1 with 1 representing maximum understanding. * indicates significance (p<0.05).

	Bar-line	Scatter
1st place (frequency)	0.66 (0.49)	0.68 (0.46)
2nd place (frequency)	0.82 (0.34)	0.83 (0.39)
1st place (artist hotness)	0.66 (0.49)	0.70 (0.46)
2nd place (artist hotness)	0.81 (0.34)	0.83 (0.39)

5.5 Confidence 2: Blind-spots

Participants were asked to provide their confidence values for their answers about their first and second highest blind-spots. The average scores are summarized in Table 5. The trends show that the participants had higher confidence with scatterplot for identifying their first most consumed genre. For their second most consumed genre, they had higher confidence with bar-line chart. The observed trends are significant for identification of artist hotness values (Mann-Whitney U-test, U-value = 33 at p<0.05 for artist hotness of first highest blind-spot, and U-value = 16.5 at p<0.05 for artist hotness of second highest blind-spot). Hypothesis 4 predicted that participants will have higher confidence for answers about their blind-spots with scatterplot more than bar-line chart. The trend is significant in the both directions, and the hypothesis is **not confirmed**.

Table 5: Mean (std) confidence scores of all participants for identifying their top 2 blind-spots, and their artist hotness values. Scores lie between 1 to 5 with 5 representing maximum confidence. * indicates significance (p<0.05).

	Bar-line	Scatter
1st place (frequency)	3.80 (0.93)	4.20 (0.90)
2nd place (frequency)	4.30 (0.82)	3.90 (0.67)
1st place (artist hotness)*	4.25 (0.62)	4.80 (0.40)
2nd place (artist hotness)*	4.90 (0.30)	4.00 (0.50)

5.6 Exploration

In the exploration stage, participants were asked to explore music from their frequent and blind-spot genres/genre-combinations. Hypothesis 5 states that users who have higher understanding of their profile explore their blind-spot genres more. For each user, an exploration factor (EF_{bs}) was computed to quantify their exploration in their blind-spot genres:

$$EF_{bs} = N_{bs} * w_{bs},$$

where N_{bs} is the number of genres listened in blind-spot category, and w_{bs} is the number of tracks listened in each of these genres. We compared this exploration factor with the user's understanding of their consumption pattern and blind-spots (obtained from their total

scores for all their answers, from Stage 1 evaluation - Section 4.1). A positive Spearman's correlation of 0.44 was obtained between user's understanding of their profile and their exploration in blind-spot genres (Significant at p<0.05). Thus hypothesis H5 is **confirmed**.

5.7 Post-hoc analysis

We did a post-hoc analysis to confirm that the positive correlation obtained between a user's exploration in blind-spot category and their understanding of their profile (Section 5.6) is exclusive, and not observed in frequent and bridge (i.e., frequent + blind-spot combination) categories. The results of correlation analysis for frequent and bridge categories are shown in Table 6. The results state that user's understanding of their profile has a negative correlation (p<0.05) with their exploration in frequent category and an insignificant weak positive correlation with bridge category. This observation implies that the positive correlation between user's exploration and their understanding is exclusive to blind-spot category.

Table 6: Spearman's correlation between exploration factor in frequent (EF_f) and bridge (EF_b) genre categories, and user's understanding of their profile. * indicates significance (p<0.05).

	EF_f	EF_b
Actual understanding	-0.45*	0.01

6 DISCUSSION

In the first stage of our evaluation (Section 4.1) we aimed to understand the effectiveness of visualizations at conveying information about (a) user's consumption pattern, and (b) user's blind-spots.

The correctness scores show that conventional bar-line chart is better at conveying information that is *explicit* about user's profiles (i.e., information about consumption pattern). For conveying information about blind-spots, or *implicit* information, scatterplot obtained higher scores. But the obtained results were not significant, and therefore, we did a post-hoc analysis on user's comments for each of the visualization. We found that a large number of the users agreed that bar-line charts were easier to get detailed information (8 users agreed and no one disagreed), while scatterplot was easier for comparison of their profile with global profile (9 users agreed and no one disagreed). This reasoning supports the scores obtained for both the charts, especially for scatterplot for the identification of blind-spots, since, the ability of a chart to compare global and individual profile is significant for blind-spot recognition.

In Stage 2 of evaluation (Section 4.2), we study the impact of user's understanding of their profile, on their intention to explore blind-spot genres. A positive correlation concluded that users who are more aware of their profile tend to explore their blind-spot genres more. Furthermore, the results of post-hoc analysis (Section 5.7) showed that the observed positive correlation (between user's exploration in blind-spot category and their understanding) is exclusive to blind-spot category, and not observed in frequent or bridge (frequent + blind-spot combination) categories, thereby

further reinforcing the fact that users with higher understanding of their profile explore their blind-spot category more.

Additionally, during exploration, we found that users show interest in mixing genres from their frequent and blind-spot categories (i.e., bridge genres), to discover new songs. The total number of genres that users explored in bridge category is almost as high as the number of genres explored in purely frequent or blind-spot genres/genre-combinations (Table 7). This suggests that, irrespective of their understanding in their profile, users are equally inclined to combined genres from different categories. During exploration phase, we used different color codes to distinguish between frequent and blind-spot genres. This might have stimulated an urge among users to combine genres from these two categories.

Table 7: Total number of genre/genre-combinations explored by the user in different categories.

Frequent	Blind-spot	Bridge
50	52	47

6.1 Limitations

In this section, we delineate the limitations and delimitations of our system which restrict the scope of our results.

Firstly, when it comes to the data used in our experiment, the global consumption data obtained from Million Song Dataset's taste profile subset (TPS), is available only until the year 2011. There is no known way to extract data beyond this time period, and hence it is quite possible that recent changes in trends are not reflected in our global profile. Secondly, for comparison of visualizations, we only compare between the scores of bar-line chart and scatterplot. However, there could be other visualizations that obtain higher scores than these two visualizations. Future studies could focus on exploring better means of representation.

Finally, when studying the correlation between user's exploration and their understanding, our study is restricted to user's exploration at that specific point during the experiment. Neither do we confirm if users continue to explore diverse music, nor do we consider the impacts of contextual factors such as user's mood, time of the day etc. In future work, these factors should be taken into account.

7 CONCLUSIONS AND FUTURE WORK

Recommender systems continue to inform our beliefs and opinions as they influence the information we consume in the world around us, ranging from the music we listen and movies we watch, to the news we read and food we consume. This raises the bar in terms of the ethics of responsible recommendation, and if recommender systems are to earn our trust then they must help us understand why certain suggestions are being made and why others are not. We have presented a user-centered study to assess the effectiveness of a visualizations to improve human decision making. The results suggest that users can understand the two visualizations, and that these visualizations are effective for helping users to identify their consumption blind-spots.

Furthermore, on studying users' exploration pattern we found that users who have more understanding of their profile, also actively explore their blind-spots more. Together, our findings suggest that it is possible to break a user's filter-bubble by increasing a user's awareness of their choices, and providing user control to explore new item-sets.

In our future work, we will learn to detect a user's exploration preferences and incorporate this information to refine our recommendations. Our first step will be to differentiate between content that a user is not consuming because they are not aware of it, from content that the user does not engage with because they are not interested. We also plan to continue this work in other domains than music, such as news recommendations.

REFERENCES

[1] Zeinab Abbassi, Vahab S. Mirrokni, and Mayur Thakur. 2012. *Diversity Maximization Under Matroid Constraints*. Technical Report. Department of Computer Science, Columbia University.

[2] Gediminas Adomavicius and YoungOk Kwon. 2011. Improving Aggregate Recommendation Diversity Using Ranking-Based Techniques. *IEEE Transactions on Knowledge and Data Engineering* 24 (2011), 896–911.

[3] Eytan Bakshy, Solomon Messing, and Lada A. Adamic. 2015. Exposure to Ideologically Diverse News and Opinion on Facebook. *Science* 348, 6239 (2015), 1130–1132.

[4] Derek Bridge and John Paul Kelly. 2006. Ways of Computing Diverse Collaborative Recommendations. In *Adaptive Hypermedia and Adaptive Web-based Systems*. 41–50.

[5] Borgelt C. 2005. Keeping things simple: finding frequent item sets by recursive elimination. (2005).

[6] Michael D Conover, Bruno Gonçalves, Alessandro Flammini, and Filippo Menczer. 2012. Partisan asymmetries in online political activity. *EPJ Data Science* 1, 1 (2012), 6.

[7] Seth Flaxman, Sharad Goel, and Justin M Rao. 2016. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly* 80, S1 (2016), 298–320.

[8] Schreiber H. 2015. Improving Genre Annotations for the Million Song Dataset. (2015).

[9] Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. 2000. Explaining collaborative filtering recommendations. In *ACM conference on Computer supported cooperative work*. 241–250.

[10] Li D Hu Y. 2013. Evaluation on Feature Importance for Favorite Song Detection. (2013).

[11] Yucheng Jin, Nava Tintarev, and Katrien Verbert. 2018. Effects of Individual Traits on Diversity-aware Music Recommender User Interfaces. In *UMAP*.

[12] Michael Jugovac, Dietmar Jannach, and Lukas Lerche. 2017. Efficient optimization of multiple recommendation quality factors according to individual user tendencies. *Expert Systems with Applications* 81 (2017), 321–331.

[13] Ellis DP Lanckriet GR McFee B, Bertin-Mahieux T. 2012. The million song dataset challenge. (2012).

[14] Sayooran Nagulendra and Julita Vassileva. 2014. Understanding and controlling the filter bubble through interactive visualization: a user study. In *Proceedings of the 25th ACM conference on Hypertext and social media*. ACM, 107–115.

[15] Tien T Nguyen, Pik-Mai Hui, F Maxwell Harper, Loren Terveen, and Joseph A Konstan. 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proceedings of the 23rd international conference on World wide web*. ACM, 677–686.

[16] Eli Pariser. 2011. *The filter bubble: What the Internet is hiding from you*. Penguin Books.

[17] Barry Smyth and Paul McClave. 2001. Similarity vs. Diversity. In *4th International Conference on Case-Based Reasoning*.

[18] Nava Tintarev, Shahin Rostami, and Barry Smyth. 2018. Knowing the unknown: visualising consumption blind-spots in recommender systems. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing, SAC 2018, Pau, France, April 09-13, 2018*. 1396–1399.

[19] Saúl Vargas and Pablo Castells. 2011. Rank and relevance in novelty and diversity metrics for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*. ACM, 109–116.

[20] Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, and Georg Lausen. 2005. Improving Recommendation Lists Through Topic Diversification. In *WWW'05*. 22–32.