Evaluation protocol for a data visualization web page displaying a Choropleth map

January 18, 2021

Introduction

The following document presents a Data Visualization application that relies on the deep learning WASABI dataset and proposes an evaluation protocol to assess its usability.

The application is a web page displaying a Choropleth map, coded with the D3.js Javascript library.

As a reminder, a Choropleth map displays divided geographical areas or regions that are coloured, shaded or patterned in relation to a data variable. As such, our application offers the following:

> The user can display the total population of, and the total number of bands within, each country in the world (along with a per capita ratio and 'market share' in %) over a range of decades (1960s to the 2000s) and among seven genre families (rock, metal, punk, country&folk, hip-hop&rap, jazz, and electro).

It is available on Github[1] with how-to guidelines to replicate the visualization on any computer.
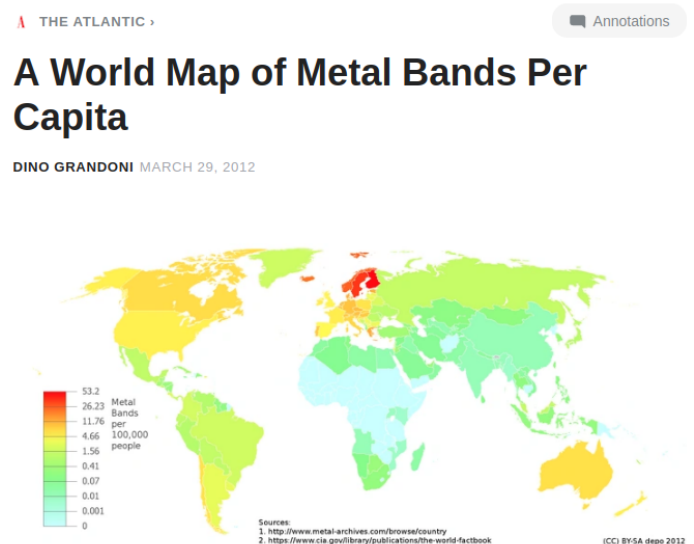

Figure 1: Example of Choropleth use in the Atlantic

1 Creating the Data Visualization

1.1 Why a Choropleth?

The idea of a choropleth to display the presence of music bands in a country, and pinning countries against each other to show which has the most band/concentration of bands for certain genres is a staple of music news media (see Fig. 1).

The goal is to reproduce and expand such a visualization to show, as highlighted in the Introduction:
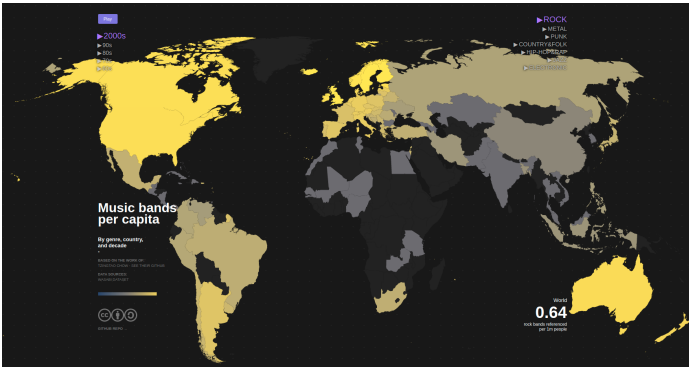
---

[1]Link to Github

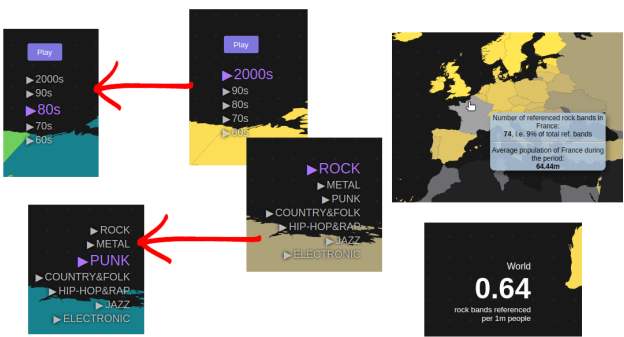Figure 2: Choropleth Map WASABI implementation

Figure 3: Examples of available interactions

→ the distribution of bands per capita, per genres, per decades, across the world since the 1960s.

The final Choropleth that will presented to the evaluation participants can be seen in Fig. 2.

1.2 Available Interactions

Multiple interactions are available to the users (see Fig. 3):

- Switching the displayed decade: 1960s, 1970s, 1980s, 1990s, 2000s

- Switching the displayed genre: rock, metal, punk, country & folk, hip-hop & rap, jazz, electronic

- Hovering on countries to display tooltips (with country demographics data, and band number/'market' share)

- A static display showing the concentration/density of bands per 1m people in the area currently hovered on

- A play button that triggers a scrolling animation over the whole decade range (from 1960s to 2000s) for the currently selected genre

1.3 How Does It Work?

Folder Structure

The folder structure of the project is fairly straightforward. Discarding all non-superfluous folders, the application's folder structure can be seen in Fig. 4.

```
.
├── css
│   └── base.css
├── data
│   ├── genre-summary
│   │       JSON with tree-like struct: {genre:{decade:{band number}}}
│   ├── music-data
│   │       JSON with tree-like struct: { country:{ genre:{ decade:{ band number, population}}}}
│   └── music-world-data
│           JSON with tree-like struct: {country:{ decade:{band number}}}
├── js
│   └── main.js
├── index.html
└── README.md
│
```

Superfluous folders and files omitted

Figure 4: Folder structure of the project

D3.js Render Function

The rendering is performed by a single Javascript function that loads three JSON files that contains the processed WASABI dataset at different levels of granularity (see Fig. 5).

The reason why the WASABI data is stored in the three JSON files GENRE-SUMMARY, MUSIC-DATA, and MUSIC-WORLD-DATA is to reduce the amount of calculations performed at each render (i.e. every time the user interacts with the decades or genres by clicking on a button, the map is rendered again). By adding two separate JSON files, which weight 20Kb in total, we were able to cut the render time of the map to a seamless degree.
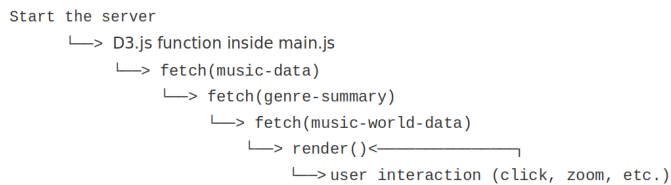
```
Start the server
    └──> D3.js function inside main.js
        └──> fetch(music-data)
            └──> fetch(genre-summary)
                └──> fetch(music-world-data)
                    └──> render()<───────────────┐
                        └──>user interaction (click, zoom, etc.)
```

Figure 5: D3.js render function process

Data Processing Involved

We saw that the data visualization relies on 3 JSON files: GENRE-SUMMARY, MUSIC-DATA, and MUSIC-WORLD-DATA. To obtain those files, two datasets are used:

- WASABI dataset[2]. Tables extracted: ALBUMS, ARTISTS

- World Bank Population dataset[3]. Table extracted: POPULATION

For each table, we keep the following data columns:

- ALBUMS: _ID, location.country

- ARTISTS: ID_artist, genre, publicationDate

- POPULATION: year, country.name

Our data processing follows the following pipeline:

1. All string data is lower-cased

2. The 'genre' column in the ARTISTS table is standardized: genre names are filtered for keywords and sorted into the seven genre families we are interested in. The other artists are dropped. Ex. {'punk *rock*', ... , 'math *rock*'} → 'rock'

3. The 'publicationDate' column in the ARTISTS table, containing years, is aggregated into decades. Ex. {1980, ..., 1989} → 1980s.

4. The process is repeated for the column 'year' in the table POPULATION

5. The table ALBUMS and ARTISTS are joined into a single table WASABI by their respective [primary] keys _ID and ID_artist

---

[2]Link to WASABI
[3]Link to WBP

6. the columns 'location.country' in the joined table WASABI and 'country.name' in the table POPULATION are standardized and matched to create shared values, making them both primary keys (*This involved a fair amount of manual work as location can be defunct countries, countries that changed names, cities, regions, etc.* **No row was dropped at this step**)

7. Rows in the joined table WASABI that had a NA value in the column 'location.country' had the value replaced with *world* so they did not have to be dropped (34052 artists concerned)

8. The joined table WASABI and the table POPULATION are joined via their primary key

9. The final joined table is *grouped by* countries, decades, and genres

10. The final joined table is dumped into 3 distinct JSON files:
    - genre-summary:
      {GENRE:{DECADE:{# BANDS}}}
    - music-data:
      {COUNTRY:{GENRE:{DECADE:{POP.,# BANDS}}}}
    - music-world-data:
      {COUNTRY:{DECADE:{# BANDS}}}

1.4 Summary

To summarize, a user can use the data visualization to perform the following:

- Explore band density per country, genre, and decade

- make a tooltip appear by hovering on a country. The tooltip provides demographic and 'market share' data for the currently selected genre and decade

- Play a scrolling animation over the decade range (from the 1960s tothe 2000s) for each genre

This can represent a complex number of interactions. As such, we can be interested in evaluating how users would interact with them.

1.5 Possible changes and further developments

The final question is whether the current data visualization application can be further improved. To this, four items have been identified:

- Zooming is tied to the rendering function of the map, i.e., every time the user switches the decade or genre, the zoom is reset to its starting value – which can be frustrating if a user is unaware of it beforehand. A solution could be to create a global variable that holds a value independent of the rendering.

- The current data visualization only shows decades. Replacing decades with years could add a further layer of details, albeit it would imply finding a solution to replace the current button implementation to avoid clutter.

- We saw that 34052 artists had an NA value for a localization. Finding the actual localization for those would provide an increase accuracy to the data visualization

- The play button cannot be paused once it is launched. Allowing pausing could be an improvement to the user interactivity.

2 Evaluation Format of the Data Visualization

This document must not be shared with the participants as before-hand knowledge of the application is not permitted. The goal is to avoid preparedness to the task.

The evaluation process is structured in three steps:

- Welcoming, briefing, and interview of the participants (with a questionnaire)

- The participants discover the application and must go through a series of monitored tasks

- Debriefing and interview of the participants (questionnaire and end survey)

Participant-Oriented Evaluation Method

We choose a participant-oriented method with the following user-oriented hypothesis as our driver:

> Can a participant complete a set of increasingly demanding tasks on a data visualization format they have never encountered before?

We focus on efficiency (i.e. how quickly a participant can achieve a task) and participant satisfaction.

Hypothesis Regarding The Data Visualization Format

The experimental design is not duplicated between various forms of UI or between different conditions. As such, there is no hypothesis associated with the current UI implementation of a data visualization.

Metrics

To measure the participants' handling of their tasks, the following metrics are considered:

- For effectiveness:
  - Number of successful task completions
  - Number of clicks performed to achieve a task compared to the minimum achievable
  - Tracking of eye movements
  - Tracking of mouse pointer movements

- For satisfaction: end-questionnaire to assess usability (e.g. SUS)

2.1 Test Session Checklist

A session corresponds to running a single participant through the whole process from start to finish, unless a stopping event is reached. Any session *must* follow the following steps:

- Welcoming the participant and introducing them to each involved party: the tester/observer, etc.

- *Each of the following steps must be preceded with asking the participant if they have any questions, are comfortable, and wish to proceed*

- Introducing the participant to:
    - the **evaluation process**: reason(s) for evaluating participants on the application, goals for compiling their data (see Annex 1)
    - the **context**: they will be tested on a Data Visualization application (see Annex 2)
    - the **place**: the testing area
    - the **hardware setup**, e.g., computer monitor (output to the participant), mouse (input from the participant), the means of observation of the participant and (mouse tracker, camera, etc.)

- Have the participant sign the disclosure agreement/letter of understanding (see Annex 3)

- Have the participant answer a profile questionnaire before starting the tasks (see Annex 4)

- The tester/observer removes themselves from the test area, and mentions to the participant that they can start

- The participant reveals their instructions and proceeds (see Annex 5)

- Once each task is performed, the tester/observer comes back to the test area

- Have the participant answer a system usability survey (see Fig. 1, and Annex 6) and a post-task questionnaire (see Annex 7)

- Thank the participant as they leave

2.2 Pre-Task Questionnaire

During the first part of the test, and after a disclosure agreement/letter of understanding has been signed (see Annex 3), the participant is invited to complete a questionnaire split in two parts (see Annex 4):

Figure 6: Original System Usability Scale questionnaire

- **Participant info**: Name, contact information, etc. which make the participant reachable

- **Participant prior experience** (meant to help understand the participant's profile)

2.3 Tasks

Once the pre-task questionnaire is filled and the participant is ready to continue, they are given a series of tasks to perform (see Annex 5 for the exhaustive list of tasks for the project). Each task is formatted as follow:

- Task description

- Evaluation question from the participant's perspective (e.g. how hard they found the task)

- Open-ended question for the participant on how they would improve the question

2.4 System Usability Scale survey

Once the tasks have been completed or attempted, the participant is given a post-task survey. The one used in this project follows the **System Usability Scale** standard (see Fig. 6) from J. Brooke[1]. It is a standard questionnaire used for assessing the overall usability of a system/tool/etc. Based on the contained 10 standard questions (see Annex 6), it allows to calculate a score from 0 to 100 (from worst to best) to inform testers/observers of the overall usability and efficient learning of a system/tool/etc. by a novice user.

2.5 End Questionnaire

Once the survey is completed, a final questionnaire is provided to the participant (see Annex 7).

2.6 Data Visualization Results

Based on a preliminary run through a single person, the application yielded a SUS score of 85 out of 100. Overall, the test participant found the application easy to navigate, finding difficulties not in the tool but the way the evaluation was organized – notably that the questions were asked by voice rather than printed on paper.

The main criticism the application received was related to the color scale of a genre ("Jazz") which was not sharp enough. The second criticism is related to the zoom function which is reset every time the user changes a parameter.

Annexes

Annex 1 - Introduction to the Choropleth evaluation

"Welcome, my name is [tester/observer]. We are going to walk you through today's session. It should take between twenty and thirty minutes of your time. Our purpose is to evaluate the usability and engagement of participants towards a newly developed Data Visualization tool. This tool will allow you to explore data related to music genres across the world and decades.

Before we proceed, we want to thank you for clearing your time schedule for us. Your input is valuable for our research and will help us determine the current state of our development and where to improve the tool."

Annex 2 - Statement of purpose for the Choropleth evaluation

"This session is organized in a few steps, the first of which is asking you to review a document called a 'Letter of Understanding' that will describe the goal of your participation. If you agree to it and sign it – it works as consent –, we will then proceed with a questionnaire that will ask you about your personal background. Once filled, we will ask you to take seat in front of a computer and try accomplishing a series of tasks on an application that will be displayed on the monitor. The application itself is displayed inside a web browser such as Internet Explorer or Google Chrome. You will be given a mouse and keyboard to perform those tasks.

While performing those tasks, your reactions and movements will be recorded to assess your primary reaction to the application. Please note that there are no

wrong ways to arrive an an answer. As you are performing the task, we will ask you to speak out loud your thought process and remarks as often as possible. The goal is to capture your thought process, what you are looking at, what you are trying to do. This is a tremendous help for our research.

You can ask us any question at any time as you continue through the process. We may not be able to provide an answer right away as we are interested in your reaction in a seemingly unsupervised manner.

After you are done with the tasks, or after you stop the process, we will go proceed to a final interview where you will be asked a survey about the application you just used, and finally a questionnaire about your general comprehension of the tool and process involved."

Annex 3 - Letter of Understanding/Disclosure Agreement for the Choropleth evaluation

"*Please read this page carefully*

By Participating in this evaluation, you will be assessed to help our program improve a data visualization tool related to the evolution of music data across the world and past decades. You will be observed and the way you interact with the application will be collected. You will also be asked to fill out 2 questionnaires, 1 survey, and answer interview questions if need be.

By signing the present form, you give your permission for your data (verbal statements, computer inputs such as mouse movements, visual data such as eye movements, etc.) to be collected for the purpose of evaluating, improving, and displaying the application. Your personal

identifiable data such as your name will not be used.

You can withdraw from this test at any time and unilaterally. You may ask any question at any time.

If you agree with these terms, please indicate your agreement by providing your name, signature, and current date."

Annex 4 - Profile questionnaire for the Choropleth evaluation

- Personal Information:
  - Name, Age, & Email
  - Gender (Male, Female, Other)
  - Education Level (None, High School, Undergrad, Graduate, PhD)
  - Current Occupation (Student, Employee, etc.)

- Participant's Profile Information:
  - Do you play a music instrument? [Yes/No]
  - Are you subscribed to an audio streaming platform (e.g. Spotify)? [Yes/No]
  - How many hours per week do you dedicate to listening to music? [<1h, 1-3h, 3h+]
  - How many music genres do you mainly listen to? [1, 2, 3+]
  - Do you read or follow (on social media) music bands or music-dedicated news media? [Yes/No]
  - Do you know what a Choropleth map is? [Yes/No]
  - On a scale of 1 (least) to 5 (very), how easy was it to answer the questions above? [1-5]

Annex 5 - Choropleth evaluation tasks

Training Explanation

"You have been given a Choropleth map: a visual representation of the density of an occurrence across a region.

Here you are able to explore the number of bands available in the WASABI dataset per country, per capita, per decade, and per genre across the world (You can select any decade since the 1960s and any of seven main genres, and select/hover on any country, to display the related data).

The color gradient represents the concentration of bands in a country.

Please find below a series of tasks to complete. You may stop at any time. Please remember to think out loud whenever possible to walk us through your mental process. Good luck!"

Tasks

Task	Description	Metrics
1	You have entered a new website displaying a world map. Find the average population for the United States of America in in the 2000s.	# of clicks Time spent Eye moves Mouse moves
Q1.	How easy was it to find the value?	1 to 5 scale
Q2	What would you change to make the process easier?	open question

Task	Description	Metrics
2	Find the number of rock bands in the United States in the 1980s.	# of clicks Time spent Eye moves Mouse moves
Q1.	How easy was it to find the value?	1 to 5 scale
Q2	What would you change to make the process easier?	open question

Task	Description	Metrics
3	Find the ratio of rock bands per one million people in the United States in the 1960s.	# of clicks Time spent Eye moves Mouse moves
Q1.	How easy was it to find the value?	1 to 5 scale
Q2	What would you change to make the process easier?	open question

Task	Description	Metrics
4	Find the ratio of electro bands in Germany in the 1990s.	# of clicks Time spent Eye moves Mouse moves
Q1.	How easy was it to find the value?	1 to 5 scale
Q2	What would you change to make the process easier?	open question

Task	Description	Metrics
5	Find the percentage share of hip-hop bands over the total number of bands in France in the 1990s.	# of clicks Time spent Eye moves Mouse moves
Q1.	How easy was it to find the value?	1 to 5 scale
Q2	What would you change to make the process easier?	open question

Task	Description	Metrics
6	Find whether Andorra (a small country between France and Spain) had metal bands in the 2000s.	# of clicks Time spent Eye moves Mouse moves
Q1.	How easy was it to find the value?	1 to 5 scale
Q2	What would you change to make the process easier?	open question

Task	Description	Metrics
7	Find which country had the highest ratio of jazz band per capita in the world in the 2000s.	# of clicks Time spent Eye moves Mouse moves
Q1.	How easy was it to find the value?	1 to 5 scale
Q2	What would you change to make the process easier?	open question

Task	Description	Metrics
8	Automatically play the sequence of map per decade for the genre Punk.	# of clicks Time spent Eye moves Mouse moves
Q1.	How easy was it to find the value?	1 to 5 scale
Q2	What would you change to make the process easier?	open question

Annex 6 - SUS survey of the Choropleth evaluation

Each item of the survey is rated from 1 (strongly disagree) to 5 (strongly agree).

- I think that I would like to use this type of data visualization more frequently

- I found this type of data visualization unnecessarily complex

- I thought the Choropleth map was easy to use

- I think that I would need the support of a technical person to be able to use such a map

- I found the various way to display information on the map were well integrated

- I thought there was too much inconsistency on the map

- I would imagine that most people would learn to use this type of visualization very quickly

- I found the map very cumbersome to use

- I felt very confident in navigating the map

- I needed to learn a lot of things before I could get going with this system

Annex 7 - Final questionnaire of the Choropleth evaluation

- What did you like and/or dislike about the evaluation process?

- Have you had any difficulties, if so where?

- For each task, were you confident beforehand that your interaction with the application would output the right results?

- What would you change to the data visualization application?

- What would you change to the evaluation process?

- Please describe in one sentence the data visualization application you used?

Annex 8 - Example participant feedback for the Choropleth evaluation

Pre-Task Questionnaire

- Personal Information:
    - **Name**: C****** C*********
    - **Email**: c**********@*****.fr
    - **Gender**: Male
    - **Age**: 23
    - **Education Level**: Undergrad
    - **Current Occupation**: Student

- Participant's Profile Information
    - Do you play a music instrument? **No**
    - Are you subscribed to an audio streaming platform (e.g. Spotify)? **Yes**
    - How many hours per week do you dedicate to listening to music? **1-3h**
    - How many music genres do you mainly listen to? **3+**

    - Do you read or follow (on social media) music bands or music-dedicated news media? **No**
    - Do you know what a Choropleth map is? **Yes**
    - On a scale of 1 (least) to 5 (very), how easy was it to answer the questions above? **5**

Tasks

- **Task 1**:
    - Was the task achieved? **Yes**
    - Easiness Rating by Participant (1-hardest to 5-easiest): **5**
    - Did the participant offer to change something? **No**

- **Task 2**:
    - Was the task achieved? **Yes**
    - Easiness Rating by Participant (1-hardest to 5-easiest): **5**
    - Did the participant offer to change something? **No**

- **Task 3**:
    - Was the task achieved? **Yes**
    - Easiness Rating by Participant (1-hardest to 5-easiest): **4**
    - Did the participant offer to change something? **Yes**, maybe include the ratio in the tooltip.

- **Task 4**:
    - Was the task achieved? **Yes**
    - Easiness Rating by Participant (1-hardest to 5-easiest): **5**
    - Did the participant offer to change something? **No**

- **Task 5**:
    - Was the task achieved? **Yes**
    - Easiness Rating by Participant (1-hardest to 5-easiest): **4**

– Did the participant offer to change something? **Yes**, maybe provide more information on what the ratio mean.

- **Task 6**:
  - Was the task achieved? **Yes**
  - Easiness Rating by Participant (1-hardest to 5-easiest): **3**
  - Did the participant offer to change something? **Yes**, remove automatic unzooming

- **Task 7**:
  - Was the task achieved? **Yes**
  - Easiness Rating by Participant (1-hardest to 5-easiest): **3**
  - Did the participant offer to change something? **Yes**, spread the range of color for Jazz

- **Task 8**:
  - Was the task achieved? **Yes**
  - Easiness Rating by Participant (1-hardest to 5-easiest): **5**
  - Did the participant offer to change something? **No**

SUS Survey

- I think that I would like to use this type of data visualization more frequently | 2

- I found this type of data visualization unnecessarily complex | 2

- I thought the Choropleth map was easy to use | 5

- I think that I would need the support of a technical person to be able to use such a map | 1

- I found the various way to display information on the map were well integrated |

- I thought there was too much inconsistency on the map | 1

- I would imagine that most people would learn to use this type of visualization very quickly | 5

- I found the map very cumbersome to use | 2

- I felt very confident in navigating the map | 4

- I needed to learn a lot of things before I could get going with this system | 1

To calculate the SUS score, first sum the score contributions from each item. Each item's score contribution will range from 0 to 4. For items 1,3,5,7,and 9 the score contribution is the scale position minus 1. For items 2,4,6,8 and 10, the contribution is 5 minus the scale position. Multiply the sum of the scores by 2.5 to obtain the overall value of SU.

The total score is: **85** out of 100.

End Questionnaire

- What did you like and/or dislike about the evaluation process? **Questions were asked via voice. Providing a paper version would be better for people who are not English native speakers**.

- Have you had any difficulties, if so where? **No**.

- For each task, were you confident beforehand that your interaction with the application would output the right results? **Most of the time**.

- What would you change to the data visualization application? **Some color scales could be made more pronounced**.

- What would you change to the evaluation process? **Nothing**.

- Please describe in one sentence the data visualization application you used? **Evolution of music genres by decade**.

References

[1] Brooke, John. (1995). *SUS: A quick and dirty usability scale*. Usability Eval. Ind.. 189.