

# RESEARCH ETHICS BOARDS APPLICATION FORM

## Prospective Research

This form should only be used if new data will be collected. For research involving only secondary use of existing information (such as health records, student records, survey data or biological materials), use the *REB Application Form – Secondary Use of Information for Research*.

This form should be completed using the [Guidance for Submitting an Application for Research Ethics Review](#).

### SECTION 1. ADMINISTRATIVE INFORMATION

[File No:

office only]

Indicate the preferred Research Ethics Board to review this research:

Health Sciences OR  Social Sciences and Humanities

**Project Title:** Explaining User Selected Patterns from Global Indicators using Machine Learning and Visualization to Promote Data Analytics Democratization

1.1 Research team information				
Lead researcher (at Dalhousie)	Name	Leonardo Milhomem Franco Christino		
	Email (@dal)	<a href="mailto:christinoleo@dal.ca">christinoleo@dal.ca</a>	Phone	6046559491
	Banner #	B00839709	Academic Unit	Faculty of Computer Science
Co-investigator names, affiliations, and email addresses	Martha Dais, Post Doctoral Fellowship – Faculty of Computer Science, Dalhousie University, dais.martha@dal.ca			
Contact person for this submission (if not lead researcher)	Name	Fernando Vieira Paulovich		
	Email	<a href="mailto:paulovich@dal.ca">paulovich@dal.ca</a>	Phone	9024941986

Study start date	September 1, 2020	Study end date	October 31, 2020
------------------	-------------------	----------------	------------------

<b>1.2 For student submissions</b> (including medical residents and postdoctoral fellows)			
Degree program	PhD		
Supervisor name and department	Fernando Paulovich Faculty of Computer Science		
Supervisor Email (@dal)	<a href="mailto:paulovich@dal.ca">paulovich@dal.ca</a>	Phone	9024941986
Department/unit ethics review (if applicable). <b>Undergraduate minimal risk research only.</b>			
Attestation: <input type="checkbox"/> I am responsible for the unit-level research ethics review of this project and it has been approved.			
Authorizing name:			
Date:			

<b>1.3 Other reviews</b>			
Other ethics review (if any) for this research	Where?		
	Status?		
Scholarly/scientific peer review (if any)			
Is this a variation on, or extension of, a previously approved Dal REB submission?	<input type="checkbox"/> No <input type="checkbox"/> Yes Dal REB file # _____		
<b>If yes</b> , describe which components of the current submission are the same as the previously approved submission (list section numbers), and which components are different from the previously approved submission (list section numbers). You may also use highlighting to clearly indicate revised text.			

<b>1.4 Funding</b>		[x] Not Applicable
Funding (list on	Agency	
	Award Number	

consent form)	Institution where funds are/will be held	<input type="checkbox"/> Dalhousie University <input type="checkbox"/> Other: _____
	Was a Release of Funds issued for this award?	<input type="checkbox"/> No <input type="checkbox"/> Yes    Date of RoF Agreement: _____

<p><b>1.5 Attestation(s).</b> The appropriate boxes <i>must</i> be checked for the submission to be accepted by the REB</p>
<p><input checked="" type="checkbox"/> I am the <b>lead researcher</b> (at Dalhousie) named in section 1.1. I agree to conduct this research following the principles of the Tri-Council Policy Statement <i>Ethical Conduct for Research Involving Humans</i> (<a href="#">TCPS</a>) and consistent with the University <a href="#">Policy on the Ethical Conduct of Research Involving Humans</a>.</p> <p>I have completed the TCPS Course on Research Ethics (<a href="#">CORE</a>) online tutorial.</p> <p><input checked="" type="checkbox"/> Yes    <input type="checkbox"/> No</p> <p>For Supervisors (of student / learner research projects):</p> <p><input checked="" type="checkbox"/> I am the <b>supervisor</b> named in section 1.2. I have reviewed this submission, including the scholarly merit of the research, and believe it is sound and appropriate. I take responsibility for ensuring this research is conducted following the principles of the <a href="#">TCPS</a> and University <a href="#">Policy</a>.</p> <p>I have completed the TCPS Course on Research Ethics (<a href="#">CORE</a>) online tutorial.</p> <p><input checked="" type="checkbox"/> Yes    <input type="checkbox"/> No</p>

## SECTION 2. PROJECT DESCRIPTION

<p><b>2.1 Lay summary</b></p>
<p>2.1.1 In <b>plain language</b>, describe the rationale, purpose, study population and methods to be used. Include a summary of background information or literature to contextualize the study. What new knowledge, or public or scientific benefit is anticipated? [maximum 500 words]</p> <p>In the past decades, massive efforts involving companies, non-profit organizations, governments, and others have been put into supporting the concept of data democratization, promoting initiatives to educate people to confront information with data. Although this represents one of the most critical advances in our free world, access to data without concrete facts to check or the lack of an expert to help on understanding the existing patterns hampers its intrinsic value</p>

and lessens its democratization. So, the benefits of giving full access to data will only be impactful if we go a step further and support the Data Analytics Democratization, assisting users in transforming findings into insights without the need of domain experts to promote unconstrained access to data interpretation and verification. In this work, we will explore user responses when executing pre-defined actions on the visual analytics software called Explainable Patterns (ExPatt), which is a new framework to support lay users in exploring and creating data storytelling. Through its use, lay users analyse plausible explanations for observed or selected findings within world indicators visualizations using relationships of said findings extracted from Wikipedia texts and visualizations, reducing or even avoiding the need for domain experts, such as historians, economists and world indicator experts, to explain the aforementioned finding.

This is a pilot study.

This is a fully developed study.

2.1.2 Phased review. If a phased review is being requested, describe why this is appropriate for this study, and which phase(s) are included for approval in this application. Refer to the [guidance document](#) before requesting a phased review.

Not applicable

## 2.2 Research question

State the research question(s) or research objective(s).

We have four hypotheses and one research objective:

1. **Hypothesis I**: The proposed interface allows the user to identify patterns and successfully request and understand possible explanations of the selected patterns.
  2. **Hypothesis II**: The proposed interface is easy to understand and to use without the need of a human expert for guidance.
  3. **Hypothesis III**: The proposed interface was relevant to confirm a previously known historical fact by the user.
  4. **Hypothesis IV**: The proposed interface was useful to users by allowing them to learn about a previously unknown fact from world indicator data.
- **Objective**: We aim for confirmation of ease-of-use, usefulness, and relevancy of connecting graphical patterns to textual explanation through the webapp interface being tested.

## 2.3 Recruitment

2.3.1 Identify the study population. Describe and justify any inclusion / exclusion criteria. Also describe how many participants are needed and how this was determined.

We need 15 participants to have a credible statistical analysis but will attempt to reach 30 participants if possible. This number of participants is selected based on the similar user studies performed in this area, such as *Explainable Matrix – Visualization for Global and Local*

*Interpretability of Random Forest Classification Ensembles and RuleMatrix: Visualizing and Understanding Classifiers with Rules.*

2.3.2 Describe recruitment plans and append recruitment instruments. Describe who will be doing the recruitment and what actions they will take, including any screening procedures.

The main population for our study is the general population, with no strict limitations other than the ability to perform actions on a website. The study will reach out by email to Dalhousie University students, staff, and/or professors, and contacts in local communities in Halifax. There is no social, cultural or safety consideration, and no inclusion or exclusion of participants. Emails will be sent to Dalhousie students through existing Email lists, social media pages, and through Dalhousie events calendars (i.e. Notice Digest, notice.digest@dal.ca ), and emails to contacts in local communities in Halifax will be made through points of contacts from these communities known by the researchers.

2.3.3 If you require permission, cooperation, or participation from a community, organization, or company to recruit your participants, describe the agreement obtained from the relevant group(s). Attach correspondence indicating their cooperation and/or support (required). Describe any other community consent or support needed to conduct this research. (If the research involves Indigenous communities complete section 2.11).

Not applicable

## **2.4 Informed consent process**

2.4.1 Describe the informed consent process:

A) How, when and by whom will the study information be conveyed to prospective participants? How will the researcher ensure prospective participants are fully informed?

See appendices. The participants will be presented the informed consent form before they start the study. This form contains a brief introduction of the study, information about confidentiality and anonymity of the participant's data, the participant's right to withdraw and the compensation. The form will be provided and administered by PI at the very beginning of the study.

B) Describe how consent will be documented (e.g. written signature, audio-recorded, etc).

The consent will be digitally signed before the study and will be saved with the study results.

Append copies of all consent information that will be used (e.g. written consent document, oral consent script, assent document/script, etc).

*Note: If the research will involve third party consent (with or without participant assent), and/or ongoing consent, ensure these are described above.*

2.4.2 Discuss how participants will be given the opportunity to withdraw their participation (and/or their data) and any time (or content) limitations on this. If participants will not have opportunity to withdraw their participation and/or their data explain why.

The participants will be told that they can withdraw from the study at any time they want without penalty. This is mentioned in the consent form.

2.4.3 If an alteration/exception to the requirement to seek prior informed consent is sought, address the criteria in TCPS article [3.7A](#). If the alteration involves deception or nondisclosure, also complete section 2.4.4.

[x] Not applicable

2.4.4 Describe and justify any use of deception or nondisclosure and explain how participants will be debriefed.

[x] Not applicable

## 2.5 Methods, data collection and analysis

2.5.1

A) Where will the research be conducted?

The research will be conducted online through online forms using the Microsoft Office platform, and safely stored in the SharePoint secured vault of PI's university account Microsoft Office.

B) What will participants be asked to do?

**Five** problems whose solutions exist in the system are given to each participant. They are required to investigate the data through the system to find their answers. This task is done using two requirements from the participant: whether they found the expected answer and their opinion of their experience while searching for the task's answer. The measured metrics from both scenarios are compared to prove or reject our hypotheses. The user's interaction with the proposed system, such as mouse clicks and hover, are not captured.

- a. The users are asked to perform a specific series of actions on the system and search for the task's answer within the resulting visualization(s). When a relevant answer is found, the user should answer the questionnaire. This process continues until the user is satisfied with the result and submits the answer.  
Interaction and understandability feedback: The users are then asked to evaluate their experience while performing the task.

The user interface is implemented as a web page, and participants will have a specific amount of time to perform all parts of the study. Participants will be asked to perform the following tasks in 70 minutes:

- Participants will sign the consent form and fill the demographic questionnaire. (10 minutes)
- Participants will be given a short training on the system and tasks to be performed. (10 minutes)
- The post-condition (evaluation) questionnaire will be given to the users as well.
- The participants will receive five sets of pre-defined instructions, and they have up to 10 minutes to complete each problem.

- After finishing the user study, the participant will answer the post-condition (evaluation) questionnaire (10 minutes).
- Total time of the study is 70 minutes.

At any point, participants may take breaks of any amount of time as long as they submit the form by the deadline specified in the consent form.

C) What data will be collected using what research instruments? *(Note that privacy and confidentiality of data will be covered in section 2.6)*

The data collected will be the answers of the forms, which are found in the appendix. This questionnaire is of type multiple-choice, and will require the participants to perform several actions onto the webapp under analysis by this user-test and then to either rate an affirmation (from strongly agree to strongly disagree) or to specify what output they have found while executing the steps provided. The resulting data will be a table of answers for each participant.

D) How much of the participant's time will participation in the study require?

70 minutes

Append copies of all research instruments (questionnaires, focus group questions, standardized measures, etc)

This is a clinical trial (physical or mental health intervention) – ensure section 2.12 is completed

2.5.2 Briefly describe the data analysis plan. Indicate how the proposed data analyses address the study's primary objectives or research questions.

Data analysis will focus on verifying two main concepts: how well the user felt informed by the interface and how well the interface understood. Our questionnaire has two types of questions: direct and opinion-based. The direct questions will verify if the user is able to accurately perform a given task through the system, which on our end will indicate how well users were able to utilize and understand the system's functionalities. The latter type of questions (opinion-based) asks for the user's opinion on specifics of the system and opinion over user satisfaction, which will indicate to us how confident and engaged the users are with the system's capabilities. In addition to the question types, question 3 is split into part a and part b, where participants will answer in a random order. This means that some participants will answer 3.b first and then answer 3.a and other participants will answer 3.a first and then 3.b. This is done in order to measure our system's capability to enhance the user's analysis, since we will be able to compare participant's experiences and answers when first exposed to the system's automatic capabilities (3.a) to when first being required to perform the same procedure manually (3.b). By statistically checking the results by analysing the distribution of answers and correlation between answer distributions, we will extract information of user confidence and engagement with task accuracy in order to confirm if our goal of providing a data analysis tool for lay users is met or what are the main concepts/features which most hinders this goal to be achieved. Notice that our main objective with this research is to reduce the need of help from experts to explain findings to lay users, allowing lay users to explore and learn from world indicator data through the explanations provided by Wikipedia summaries and visualizations, therefore we also include a demographics questionnaire to include in the statistical analysis an expertise weight of how well versatile the participant is regarding visual analysis and self-guided investigation. With this we can verify whether users are able to use our

system for analysis just as well as experts or not, and what areas of the system can be improved towards this goal. The participants are considered as a sample of the population.

2.5.3 Describe any compensation that will be given to participants and how this will be handled for participants who do not complete the study. Discuss any expenses participants are likely to incur and whether/how these will be reimbursed.

Participants who accept participating in the study will enter a draw of four \$25 CAD amazon gift cards. This will be clearly outlined in the consent form. There would be no incurred expenses on the participants. The gift card will be sent by email. The participants will accept participating in the study when answering to the email described in the appendix and answering the second email containing the consent form. Participants will participate on the draw even if they withdraw their participation during the study.

## 2.6 Privacy and confidentiality

2.6.1

- A) Describe who will have knowledge of participants' identities.
- B) Describe the level of identifiability of the study data (anonymous, anonymized, de-identified/coded, identifying) (see [TCPS Chapter 5A – types of information](#) for definitions).
- C) Specify which members of the research team (or others) will have access to participants' data and for what purpose.
- D) Describe measures to ensure privacy and confidentiality of study documents and participant data during the data collection and analysis phase. *[Note that plans for long term storage will be covered in 2.6.2]*
  - Address: handling of documents/data during data collection; transportation or transfer of documents/data; storage of documents/data (during the study).
  - If a key-code will be maintained, describe how it will be kept secure.
  - For electronic data, describe electronic data security measures, including file encryption and/or password protection [as applicable](#).
  - For hard copy documents, describe physical security measures (specify location).

There will not be any personal or sensitive data collected in this study. Our software (online user interface) will not collect relevant interactivity data, only the questionnaire answers. The demographic questionnaire will also not collect sensitive data, and participants are to their discretion to answer them or not. Data will be stored in either a text-based format (such as comma separated lists) or in a spreadsheet format (such as xlsx). All data will be anonymous. Only the three researchers listed at the researcher team will have access to the participant's data. The records will be stored in the university's SharePoint and allocated only to the PI's account, allowing only him and his supervisor to have access to it. The data will be deleted from the university's secure vault five years after the experiment or final publication.

[ ] This research involves personal health records (ensure section 2.13 is completed)



2.6.2 Describe plans for data retention and long-term storage (i.e. how long data will be retained, in what form and where). Will the data eventually be destroyed or irreversibly anonymized? If so, what procedures will be used for this? Discuss any plans for future use of the data or materials beyond the study currently being reviewed.

The records will be deleted from PI's university Microsoft SharePoint vault five years after the experiment or publication. The results may be used for further revisions of the same publication within the above timeframe, if needed.

This research will be deposited in a data repository (ensure section 2.14 is completed)

### 2.6.3

Describe if/how participant confidentiality will be protected when research results are reported:

A) For quantitative results - In what form will study data be disseminated?

Only aggregate data will be presented

Individual de-identified, anonymized or anonymous data will be presented

Other. If "other", briefly describe dissemination plans with regard to identifiability of data.

Not applicable, only qualitative data will be presented

B) For qualitative results - Will identifiable data be used in research presentations/publications? If participants will be quoted, address consent for this and indicate whether quotes will be identifiable or attributed.

Not applicable, only quantitative data will be presented

Anonymous excerpts of the two open-ended questions will be used only as confirmation of conclusions found by analysing the close-ended quantitative questions. No direct quotes or any form of participant identification will be used.

2.6.4 Address any limits on confidentiality, such as a legal duty to report abuse or neglect of a [child](#) or [adult in need of protection](#), and how these will be handled. Ensure these are clear in the consent documents. (See the [guidance document](#) for more information on legal duties and professional codes of ethics).

Not applicable

2.6.5 Will any information that may reasonably be expected to identify an individual (alone or in combination with other available information) be accessible outside Canada? And/or, will you be using any electronic tool (e.g. survey company, software, data repository) to help you collect, manage, store, share, or analyze personally identifiable data that makes the data accessible from outside Canada?

No

Yes. If yes, refer to the University [Policy for the Protection of Personal Information from Access Outside Canada](#), and describe how you comply with the policy (such as securing participant consent and/or securing approval from the Vice President Research and Innovation).

## 2.7 Risk and benefit analysis

2.7.1 Discuss what risks or discomforts are anticipated for participants, how likely risks are and how risks will be mitigated. Address any particular ethical vulnerability of your study population. Risks to privacy from use of identifying information should be addressed. If applicable, address third party or community risk. (If the research involves Indigenous communities also complete section 2.11)

No risks or discomforts are anticipated for participants caused due to the research. No identifiable data will be collected or stored. Participants are to their discretion to answer demographic-related questions, and all results will be reported as an aggregate statistic. Even considering the 70 minutes of expected time to complete the research, participants are allowed as many breaks they wish and each break can be as long as they wish after considering the deadline of the from submission of 1 week.

2.7.2 Identify any direct benefits of participation to participants (other than compensation), and any indirect benefits of the study (e.g. contribution to new knowledge).

Participants will be contributing to confirm ease-of-use of a specific style of visual analytics tool, providing valuable insight of future steps of the research of data storytelling within our laboratory.

## 2.8 Provision of results to participants and dissemination plans.

2.8.1 The TCPS encourages researchers to share study results with participants in appropriate formats. Describe your plans to share study results with participants and discuss the process and format.

The publication resulting from the study will be made available to participants on request.

2.8.2 If applicable, describe how participants will be informed of any material incidental findings – a discovery about a participant made in the course of research (screening or data collection) that is outside the objectives of the study, that has implications for participant welfare (health, psychological or social). See [TCPS Article 3.4](#) for more information.

[x] Not applicable

2.8.3 Describe plans for dissemination of the research findings (e.g. conference presentations, journal articles, public lectures etc.).

The results will be used as a foundation to justify research in simple visual analytics tools for non-expert users towards conference presentations and/or journal articles publications.

## 2.9 Research Team

2.9.1 Describe the role and duties of all research team members (including students, RA's and supervisors) in relation to the overall study.

The lead researcher will reach out to possible participants, deliver the online research process webpage, and collect the results. He will be the main author of articles with the results of this study, therefore all statistical analysis, data retrieval and future communications with participants who requests results of the study will also be handled by him.

The rest of the research team will supervise and double-check the collected data and results. The co-investigator is focused on technical aspects of a portion of the system in question, therefore she will be responsible to investigate relevant related works depending on the results of the study. The supervisor will manage the operation to assert good progress on article publications and that the results of the study is used in a useful and relevant manner to make the system and the area of research better.

2.9.2 Briefly identify any previous experience or special qualifications represented on the team relevant to the proposed study (e.g. professional or clinical expertise, research methods, experience with the study population, statistics expertise, etc.).

The research supervisor has previous experiences with many other similar user-test researches, and this one will be equivalent as the others, only differentiating on what is the tool being tested. Several articles of systems, applications and techniques were published by the Supervisor (e.g. Explainable Matrix – Visualization for Global and Local Interpretability of Random Forest Classification Ensembles) , and he will follow on the same procedures, some of which also involved similar study population. The research lead and co-researcher are specialists in the technical aspect (research-wise and programming-wise) of the specific work required to build the system and will use the results of the study to improve certain aspects of the work in order to publish material and verify the usability and ease of use of the webapp system to guide future research.

## 2.10 Conflict of interest

Describe whether any dual role or conflict of interest exists for any member of the research team in relation to potential study participants (e.g. TA, fellow student, teaching or clinical relationship), and/or study sponsors, and how this will be handled.

Not applicable

Possible students with the same supervisor might participate in the study. All data will be anonymous.

## 2.11 Research involving Indigenous peoples

Consult TCPS [Articles 9.1 and 9.2](#) in determining whether this section is applicable to your research.

Not applicable – go to 2.12

2.11.1 If the proposed research is expected to involve people who are Indigenous, describe the plan for community engagement (per TCPS Articles [9.1 and 9.2](#)). Attach supporting letters, research agreements and other relevant documents, if available. If community engagement is not sought, explain why the research does not require it, referencing TCPS article 9.2.

2.11.2 State whether ethical approval has been or will be sought from [Mi'kmaw Ethics Watch](#) and if not, why the research does not fall under their purview. If the research falls under the purview of other Indigenous ethics groups, state whether ethical approval has been or will be sought.

2.11.3 Describe any plans for returning results to the community and any intellectual property rights agreements negotiated with the community with regard to data ownership (see also 2.11.4 if applicable). If there are specific risks to the community involved, ensure these have been addressed in section 2.8.1.

2.11.4 Does this research incorporate OCAP (Ownership, Control, Access, and Possession) principles as described in TCPS [Article 9.8](#)?

Yes. Explain how.

No. Explain why not.

## 2.12 Clinical trials

Not applicable – go to 2.13

2.12.1 Will the proposed clinical trial be registered?

No. Explain why not.

Yes. Indicate where it was/will be registered and provide the registration number.

2.12.2 If a novel intervention or treatment is being examined, describe standard treatment or intervention, to indicate a situation of clinical equipoise exists (TCPS [Chapter 11](#)). If placebo is used with a control group rather than standard treatment, please justify.

2.12.3 Clearly identify the known effects of any product or device under investigation, approved uses, safety information and possible contraindications. Indicate how the proposed study use differs from approved uses.

Not applicable

2.12.4 Discuss any plans for blinding/randomization.

2.12.5 What plans are in place for safety monitoring and reporting of new information to participants, the REB, other team members, sponsors, and the clinical trial registry (refer to TCPS [Articles 11.6, 11.7, 11.8](#))? These should address plans for removing participants for safety reasons, and early stopping/unblinding/amendment of the trial. What risks may arise for participants through early trial closure, and how will these be addressed? Are there any options for continued access to interventions shown to be beneficial?

## **2.13 Use of personal health information**

Not applicable

2.13.1 Research using health information may be subject to Nova Scotia's [Personal Health Information Act](#). Describe the personal health information ([definition explained in the guidance document](#)) required and the information sources, and explain why the research cannot reasonably be accomplished without the use of that information. Describe how the personal health information will be used, and in the most de-identified form possible.

2.13.2 Will there be any linking of separate health data sets as part of this research?

No

Yes

If yes:

A) Why is the linkage necessary?

B) Describe how the linkage will be conducted (it is helpful to append a flow diagram)

C) Does that linkage increase the identifiability of the participants?

2.13.3 Describe reasonably foreseeable risks to privacy due to the use of personal health information and how these will be mitigated.

## 2.14 Data Repositories

Not applicable

2.14.1 Identify and describe the data repository in which the research data will be deposited. What is its focus, who are its target users, who can access deposited data and under what circumstances? For how long will the data be kept in the repository?

2.14.2 Describe the data set to be released to the repository. If there is personal and/or sensitive information in the data, describe how you will prepare the data for submission to the repository and mitigate risks to privacy. Identify all fields that will be included in the final data set (include as an appendix).

2.14.3 Is agreeing to have one's data deposited a requirement for participation in the study? If yes, provide a justification. If no, indicate how participants can opt in or out.

## SECTION 3. APPENDICES

**Appendices Checklist.** Append all relevant material to this application in the order they will be used. This may include:

List of References

Permission letters (e.g. Indigenous Band Council, School Board, Director of a long-term care facility)

Support/cooperation correspondence

- Recruitment documents (posters, oral scripts, online postings, invitations to participate, etc.)
- Screening documents
- Consent/assent documents or scripts
- Research instruments (questionnaires, interview or focus group questions, etc.)
- Debriefing forms
- List of data fields included in data repository

### **Consent Form Templates**

Sample consent forms are provided on the [Research Ethics website](#) and may be used in conjunction with the information in the [Guidance](#) document to help you develop your consent form.

## Appendix A – Email Recruitment Notice

We are recruiting 30 participants to take part in a research study to collect and examine your experience when using a new web application called Explainable Patterns, which enables users to investigate explanations of historical events impacts on demographic indicators. We are looking for anyone interested in participating in this study. You receive an overview of the study details and a consent form. If you accept the consent form, you will then watch a short video on how to use Explainable Patterns for about 5 minutes and then you will perform pre-defined operations on the webpage while answering several questionnaires. The entire study is online. After the study is completed, you will also fill out and submit an evaluation questionnaire on your experience. You should be able to complete the study in 70 minutes and are allowed breaks of your own discretion, being only required to submit your form within 1 week of accepting the consent form. **Participants who complete the study will enter a draw of four 25\$ amazon gift cards.** Only the first 30 interested participants will be contacted for the study. If you are interested in participating, please contact Leonardo Christino (christinoleo@dal.ca).





# Appendix B – Consent Form

## CONSENT FORM

### User evaluation of Relevance Feedback for Explainable Patterns Webpage in Community Question Answering

You are invited to take part in a research study being conducted by, Leonardo Christino, a PhD graduate student in the Faculty of Computer Science at Dalhousie University. The purpose of this research is to analyze and verify the ease-of-use and usefulness of our Explainable Patterns Webapp for visual analytics. The study is funded by the Visual Analytics Laboratory of Dalhousie's FCS.

If you choose to participate in this research, you will be asked to perform pre-set operations and analysis through our Webapp and anonymously answer questions regarding its usability, which are listed below. The survey should take approximately 70 minutes and you are allowed breaks of your own discretion.

- You will complete a demographic questionnaire regarding only your level of education and knowledge and experience with data/visualization analysis and text interpretation from Wikipedia.
- You will be given a tutorial on how to use the software.
- You will be given a practice session to use the software.
- You will be given an evaluation questionnaire.
- You will perform four tasks of searching answers the system.
- You will submit the post-study questionnaire and comment.

Your participation in this research is entirely your choice. You do not have to answer questions that you do not want to answer (by selecting prefer not to answer), and you are welcome to stop the survey at any time if you no longer want to participate. All you need to do is close your browser or browser window. I will not include any incomplete surveys in my analyses. If you do complete your survey and you change your mind later, I will not be able to remove the information you provided as I will not know which response is yours. After accepting this consent, you will have 1 week to perform the study, after which time you will be considered as no longer wanting to participate in it.

Your responses to the survey will be anonymous. This means that there are no questions in the survey that ask for identifying details such as your name or email address. All responses will be saved on a secure Dalhousie computer. Only Leonardo Christino, Martha Dais and Prof. Fernando Paulovich will have access to the survey results.

I will describe and share general findings of this research in a journal and/or conference publication, including aggregate/statistical data from this study. All data with answers collected from participants will be deleted 5 years after reporting the results, where only aggregated data will be available through the published material.

The risks associated with this study are no greater than those you encounter in your everyday life.

To thank you for your time you will enter a draw to four 25\$ amazon gift cards. Drawn participants will be contacted after all participants have completed the survey. Your contact information will not be linked in any way to your survey responses.

You should discuss any questions you have about this study with Leonardo Christino or Prof. Fernando Paulovich. Please ask as many questions as you like before or after participating. My contact information is [christinoleo@dal.ca](mailto:christinoleo@dal.ca).

If you have any ethical concerns about your participation in this research, you may contact Research Ethics, Dalhousie University at (902) 494-3423, or email [ethics@dal.ca](mailto:ethics@dal.ca) (and reference REB file # 20XX-XXXX).”

If you agree to complete the survey, please answer this email with “I accept the consent agreement”, and the link to the survey will be sent to you.

## Appendix C–Demographic Questionnaire

- 1 How long have you used a computer?
  - Less than one week
  - 1 week to less than 1 month
  - 1 month to less than 1 year
  - 1 year to less than 2 years
  - 2 years to less than 4 years
  - 4 or more years
  - I prefer not to answer
- 2 On average, how much time do you spend per week on a computer?
  - Less than one hour
  - One to less than 4 hours
  - to less than 10 hours
  - 10 to less than 20 hours
  - 20 to less than 40 hours
  - Over 40 hours
  - I prefer not to answer
- 3 How comfortable are you at using an “interactive” user interface, such as websites?
  - Extremely comfortable
  - Very comfortable
  - Comfortable
  - Uncomfortable
  - Very uncomfortable
  - Extremely uncomfortable
  - I prefer not to answer
- 4 How often do you use interactive user interfaces where you drag objects from one place to another?
  - Extremely often
  - Very often
  - Often
  - Not often
  - Rarely
  - Never
  - I don't understand the question
  - I prefer not to answer
- 5 How familiar are you with Data Analytics Tools, such as Microsoft Excel or Tableau?
  - Very well
  - Well
  - Neutral
  - Not well
  - Not well at all
  - I prefer not to answer
- 6 How familiar are you with Information Retrieval from said data analytics tools?
  - Very well
  - Well
  - Neutral
  - Not well
  - Not well at all
  - I prefer not to answer
- 7 At what level do you think your understanding of written English is?

- Excellent
  - Very good
  - Good
  - Acceptable
  - Bad
  - Very bad
  - None
  - I prefer not to answer
- 8 What is the highest level of education you have completed?
- Little or no formal education
  - High school or equivalent
  - Currently pursuing an undergraduate/diploma
  - College or university
  - Master
  - Doctoral
  - Post-Doctoral
  - I prefer not to answer
- 9 How often do you use Wikipedia?
- Every day
  - Once two days
  - Once four days
  - Once a week
  - Once a month
  - Once a year
  - Never
  - I prefer not to answer
- 10 How often do you use tools that use some kind of visual/graphical representation, and not only text and/or numbers?
- Every day
  - Once two days
  - Once four days
  - Once a week
  - Once a month
  - Once a year
  - Never
  - I prefer not to answer
- 11 In case you have or are pursuing a degree, what is your primary area of study?
- Computer Science
  - Information technology
  - Internetworking
  - Social Science
  - Health Science
  - Other\_\_\_\_\_
  - I have no primary area of study
  - I prefer not to answer

## Appendix D –Questionnaires

### Questionnaire 1 - Video tutorial

Please view this [video tutorial](#) only once and respond to the following statements about the visualization-based interface, using the given scale:

Question	Answers				
The video was clear to understand	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I believe that now I can perform the same operations shown in the video, such as selecting datasets or a timeframe in the graph	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I understood what the data shown in the main line <i>graph</i> represents	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I understood what the data shown in the main <i>map</i> represents, including how to understand the map's colors	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
<b>From the video,</b> I understood what the <i>theme river</i> graph shows and how to use it	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I understood why when a timeframe is selected in the graph visualization, the <i>theme river</i> graph is updated	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I understood why the <i>title</i> and the <i>text</i> shown after clicking at the theme river graph is relevant to the timeframe selection in the <i>line</i> graph	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I understood how to use the <i>related</i> or <i>similar</i> dataset recommendations panel for further analysis of the data	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I understood very well how to apply this software to real-world investigations of world indicators, such as life expectancy or economy data	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I believe on my own I won't understand the information shown through this software, and I would need an expert in world history or world indicators to explain to me any findings, such as the ones shown in the video	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>

### Questionnaire 2 - Pre-defined execution Rating Questionnaire

Now open the website at [this link](#) and perform the task described below. Then respond to the following statements about your experience using the given scale:

1. Select the "LIFE\_EXPECTANCY" dataset to be shown at the main *line* graph.

2. Select “United States” and “Russia” from the map, so that two lines are shown in the line graph, one for each country.
3. Select “United States” as the **primary** country for further comparisons.
4. Select (*approximately*) the timeframe of 1860 though 1866 in the *line* graph.
5. Discover the largest *prominent valley* of this timeframe. This information is at the *related information tabs* to the left of the screen. is from the country of “Mauritius” by using screen.

Question	Answers				
I was able to follow the steps without any problems	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I was able to quickly understand what I needed to do to perform the required steps in the webapp	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I understood much more of the workflow now than I did before being asked to perform these pre-defined steps	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
While performing the steps, I also noticed other interesting patterns and/or findings in the <i>line</i> graph not related to the steps given by the questionnaire	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
While performing the steps, I totally ignored other information not relevant to these pre-defined steps given by the questionnaire	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
Looking at the software, the most prominent valleys in the selected timeframe is from this country:	<i>United States</i>	<i>Mauritius</i>	<i>Canada</i>	<i>Denmark</i>	<i>Iceland</i>

## Questionnaire 3.a – Follow up by utilizing the system’s automatic capabilities

Following the previous steps, use the theme river, which is the area chart in the center of the screen, to answer the following questions.

Question	Answers				
Identify the result present among the <i>theme river’s</i> titles	<i>Snake War</i>	<i>Oregon in the American Civil War</i>	<i>1864 in the United States</i>	<i>A, B and C were not present</i>	<i>A, B and C were present</i>
By clicking on “ <i>Union (American Civil War)</i> ” among the <i>theme river’s</i> titles, I see that:	<i>Russia is a country mentioned in the selected Document</i>	<i>“Lincoln” is the largest word in the word cloud</i>	<i>The document summary does not say anything</i>	<i>There is no “Union (American Civil War)” among the theme</i>	<i>“AID RECEIVED” is a related indicator</i>

			about the American civil war	river's titles	
I now understood much more how to apply this software to real-world problems	Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
I believe on my own I won't understand the information shown through this software, and I would need an expert in world history to explain to me any findings, such as the ones shown in the video	Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree

### Questionnaire 3.b – Follow up by not utilizing the system's automatic capabilities, and instead requiring the use of Wikipedia

You have selected one of the valleys of the line-chart. Go now to Wikipedia through this link (<https://en.wikipedia.org/>) and search for information related to the selection. You may choose any keywords you wish for this search. In case you need ideas, you may try keywords such as “United States”, “life expectancy” and some of the years selected. Do not worry about finding the answers. We are not keeping track of the time you use for answering, but we suggest avoiding using more than 5 minutes for this task.

Question	Answers				
I think was able to search Wikipedia easily for the information I was seeking	Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
Select the documents you found through your search.	Snake War	Oregon in the American Civil War	1864 in the United States	At least two among A, B and C were present	I wasn't able to use Wikipedia
In case you see any document related to the American Civil War among the results, click it and select the best answer:	I didn't read the Wikipedia document	There is no “American Civil War” among the results	I see references to other countries in the document	“Lincoln” is mentioned in the document	From the document, I am able to go back to the system of this survey and follow on using it
I now understood much more how to apply this software to real-world problems	Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree

I believe on my own I won't understand the information shown through this software, and I would need an expert in world history to explain to me any findings, such as the ones shown in the video	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
--	--------------------------	--------------------------	----------------	-----------------------	-----------------------

## Questionnaire 4 - Pre-defined execution Rating Questionnaire

Now open the website at [this link](#) and perform the task described below. Then respond to the following statements about your experience using the given scale:

1. Make so that the datasets shown in the *line* graph are only "LIFE\_EXPECTANCY", "CHILDREN\_DEATHS" and "CHILDREN\_FERTILITY".
2. Select "Separate indicator graphs" to visualize the three lines in different axis, allowing to better compare them.
3. Make so that "United States" is the only selected country
4. Answer questions 1-7.
5. Make so that the datasets shown in the *line* graph are only "CHILDREN\_FERTILITY" and select the timeframe of approximately 1935 and 1975.
6. Answer the rest of the questions.

Question	Answers				
I was able to quickly understand what I needed to do to perform the required steps in the webapp	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I understood much more of the workflow now than I did before being asked to perform these pre-defined steps	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I understand what the "Separate indicator graphs" button does	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
In my opinion, the year of 1918 has some event that affected the data of all three datasets	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
In my opinion, the period between 1935 and 1975 had a relevant impact in which indicator dataset:	<i>Life expectancy</i>	<i>Children Deaths</i>	<i>Children Fertility</i>	<i>All</i>	<i>None</i>
By analysing on my own on other panels of the software, Australia is <u>similar</u> to United States' children fertility within "1935 and 1975"	<i>Disagree</i>	<i>Neutral</i>	<i>Agree</i>	<i>I wasn't able to find this information</i>	
By analysing on my own on other panels of the software, Argentina is <u>similar</u> to United States' children fertility within "1935 and 1975"	<i>Disagree</i>	<i>Neutral</i>	<i>Agree</i>	<i>I wasn't able to find this information</i>	



By analysing on my own on other panels of the software, the country with highest peak of children fertility within “1935 and 1975” is:	<i>United States</i>	<i>Albania</i>	<i>Russia</i>	<i>Suriname</i>	<i>Guinea-Bissau</i>
I now understood much more how to apply this software to real-world problems	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I believe on my own I will not understand the information shown through this software, and I would need an expert in world history or world indicators to explain to me any findings, such as the ones shown in the video	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>

## Questionnaire 5 - Pre-defined execution Rating Questionnaire

Now open the website at [this link](#) and perform the task described below. Then respond to the following statements about your experience using the given scale:

1. Select the “Manual Query” button and search for starvation.
2. Select the “starvation” title in the *theme river*.
3. Answer the questions.

Question	Answers				
I was able to use the Manual Query with no problems	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
By hovering my mouse over the word “starvation” on the <i>theme river</i> , I am shown that:	<i>All 10 entries have the starvation keyword</i>	<i>Eight of the entries have the starvation keyword</i>	<i>Six of the entries have the starvation keyword</i>	<i>Three of the entries have the starvation keyword</i>	<i>No entries have the starvation keyword</i>
By looking at the map resulting from the starvation selection, it seems that Wikipedia has more data related to starvation about:	<i>United States</i>	<i>Canada</i>	<i>Sweden</i>	<i>Russia</i>	<i>Japan</i>
By looking at the summary text, I can understand what that Wikipedia page is talking about	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I understand what the “Related Indicators” selection column shows about the Wikipedia result and how to use it to analyse data related to starvation	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
From here, I believe I could analyse more on my own about the reason why Canada and United States are as I answered above by selecting timeframes of interest in the <i>line graph</i> and checking the Wikipedia results	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>

I now understood much more how to apply this software to real-world problems	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I believe on my own I will not understand the information shown through this software, and I would need an expert in world history or world indicators to explain to me any findings, such as the ones shown in the video	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>

## Questionnaire 6 - Interface Features Questionnaire

Please respond to the following statements about the visualization-based interface, using the given scale:

<b>Question</b>	<b>Answers</b>				
The main line-chart graph was easy to understand	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
The meaning of the map graph's colors was easy to understand	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
The map's color scale is useful for identifying similar and dissimilar countries	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
When I selected a timeframe in the main graph, I understood why the map's colors changed	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I am able to use well the <i>theme river's</i> graph of terms present in each Wikipedia result	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
The word cloud allowed me to better understand what a specific Wikipedia document is talking about	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
The summary text allowed me to better understand what a specific Wikipedia document is talking about	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
The related indicators and mentioned countries map of a Wikipedia document allowed me to follow on analysing the data	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
The similar and dissimilar countries panels were useful to me to better understand other countries and find interesting new facts	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
The peaks and valleys panels were useful to me to better understand other countries and find interesting new facts	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>

## Questionnaire 7 - Software Usability Questionnaire<sub>2</sub>

Question	Answers				
I think that I would like to use this system frequently	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I found the system unnecessarily complex	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I thought the system was easy to use	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I think that I would need the support of a technical person to be able to use this system	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I found the various visualizations in this system were well integrated	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I thought there was too much inconsistency in this system	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I would imagine that most people would learn to use this system very quickly	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I found the system very cumbersome to use	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
The video was very effective for me to quickly learn to use the system	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
The pre-defined step by step was essential for me to discover the usefulness of the system	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I feel I would not be able to use the system on my own and would need an expert of world history or world indicators with me to understand the findings of the system	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>
I would be able to investigate other datasets and/or countries on my own now	<i>Strongly Disagree</i>	<i>Somewhat Disagree</i>	<i>Neutral</i>	<i>Somewhat Agree</i>	<i>Strongly Agree</i>

1. Please give us more comments about the system:

---



---



---



---



---



---



---



---



---



---



---



---



---



---



---



---



---



---



---



---



---



---

