Consent to Take Part in a Research Study

Q1.1.

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

Research Study Consent

Title of research study: Data science workshops for biomedical and health professionals: Persona identification and workshop assessment

Principal Investigator: Anne M Brown, PhD

Other study contact(s): Daniel Chen

Department: Research and Informatics, University Libraries

Telephone Number: 540-231-9231

Email Addresses: ambrown7@vt.edu, chend@vt.eu

Key Information: The following is a short summary of this study to help you decide whether or not to be a part of this study. More detailed information is listed later on in this form. This confidential survey is intended to identify key characteristics of learners attending a data science workshop geared towards medical and biomedical practitioners. Your responses are anonymous and will become part of summary data included in a report that is compiled by Daniel Chen from the Genetics, Bioinformatics, and Computational Biology (GBCB) PhD program at Virginia Tech.

Detailed Information: The following is more detailed information about this study in addition to the information listed above.

Who can I talk to?

If you have questions, concerns, or complaints, or think the research has hurt you, talk to the research team at Anne Brown (ambrown7@vt.edu) or Daniel Chen (chend@vt.edu).

This research has been reviewed by the Virginia Tech Institutional Review Board (IRB). You may communicate with them at 540-231-3732 or irb@vt.edu if:
You have questions about your rights as a research subject Your questions, concerns, or complaints are not being answered by the research team You cannot reach the research team You want to talk to someone besides the research team to provide feedback about this research.

How many people will be studied?
We plan to include about 100-150 people in this research study.

What happens if I say yes, I want to be in this research?
Once you consent to participate in this study, you will follow a next arrow to the study survey and begin by answering some population-specific questions. One of these questions will include creating a unique identifier which will be used for future surveys in this study. This unique identifier will also be used to remove your data if you choose to not remain in the study. From there you will complete a pre-workshop questionnaire about your experiences with programming, data processing, project management, and what you plan to get out of the workshop. The survey should take approximately 10-15 minutes to complete. After the workshop you will be invited to participate in a post-workshop survey. You may opt to complete these surveys to provide feedback to the researchers about the materials without agreeing to participate in the research study.

Your responses in this survey will be used to assess the quality of a workshop and its learning materials. The researchers will then take feedback from the

workshop to improve materials for each type of learner. You have the option to provide feedback without having your data used in the final research analysis report.

This information will be used to improve the workshop materials and give instructors a sense of who is attending the workshop and what needs learners need before, during, and after the workshop ends. It will also help determine what kind of workshop is better suited for each kind of learner.

This consent form can be taken anywhere and survey can be taken anywhere. Workshops are planned for Fall 2020-Spring 20201. Your responses from this survey will be matched up with subsequent workshop assessments, if you do attend workshops, using your de-identified ID. You may choose to attend and register for the workshop even if you do not wish to partake in this study. You may also take the surveys to provide feedback without participating in the research study.

This study is designed that individuals will take a student self-assessment prior to workshop creation. As workshops are delivered, pre and post assessment will be performed on each workshop. A subsequent long-term survey may be issued. You are invited to participate in any of the surveys as well as the current student self-assessment.

Your de-identified (anonymous) survey responses will be shared on an open science platform such as the Open Science Framework (https://osf.io/), GitHub (https://github.com/), Zenodo (https://zenodo.org/), and/or VTechData (https://data.lib.vt.edu/).

Q1.2.

What happens if I say yes, but I change my mind later?

You can leave the research at any time, for any reason, and it will not be held against you.

You will still be allowed to participate in the workshop without completing this survey.
You may also take the survey to provide feedback and also opt-out of the research study.

If you decide to leave the research, you can discontinue filling out the survey by closing your web browser. Incomplete responses will not be used in the data analysis. If you already filled out the survey and wish to withdraw your participation, please let one of the study investigators know You can decide to pull all your information from analysis, or only withdraw from any follow-up surveys. If you decide to leave the research, no consequences will occur.

Is there any way being in this study could be bad for me? (Detailed Risks)
During the process of completing the survey, you will be asked questions about programming experience and your thoughts and attitudes surrounding the subject of statistics and data management. If there are any questions you would rather not answer or that you do not feel comfortable answering, you can move on to the next question.

There is minimal risk the by being a part of this study you could experience physical, psychological, privacy, legal, social, economic, or emotional distress given the subject of the survey.

This study is not meant to gather information about specific individuals, and the information you provide will be combined with that of other survey participants to gather information.

What happens to the information collected for the research?

We will make every effort to limit the use and disclosure of your personal information, including research study and medical records, only to people who have a need to review this information. We cannot promise complete confidentiality. Organizations that may inspect and copy your information include the IRB, Human Research Protection Program, and other authorized representatives of Virginia Tech.

Your de-identified (anonymous) survey responses will be shared on an open science platform such as the Open Science Framework (https://osf.io/), GitHub (https://github.com/), Zenodo (https://zenodo.org/), and/or VTechData (https://data.lib.vt.edu/).

Your data could be used for future research studies or distributed to another investigator for future research studies without your additional informed consent.

The results of this research study may be presented in summary form at conferences, in presentations, reports to the sponsor, academic papers, and as part of a thesis/dissertation.

Can I be removed from the research without my OK?
The person in charge of the research study or the sponsor can remove you from the research study without your approval. Possible reasons for removal include incomplete responses.

What else do I need to know?
Any expenses accrued for seeking or receiving medical or mental health treatment will be your responsibility and not that of the research project, research team, or Virginia Tech.

Are you at least 18 years of age?
Yes. I am 18 years of age or older.

No. I am not at least 18 years of age.

Do you agree to participate in the research study?

Yes. I have read the consent form and this response will serve as my consent to participate in the research study
No. I do not want to participate in the research study.

Would you like to provide pre-workshop feedback? Your responses would not be used for the research study.
Yes. I would like to provide feedback about the workshop and its learning materials
No. I do not want to provide feedback.

Q1.3. Are you at least 18 years of age?

- ○ Yes. I am 18 years of age or older.
- ○ No. I am not at least 18 years of age.

Q1.4.
Do you agree to participate in the research study?

- ○ Yes. I have read the consent form and this response will serve as my consent to participate in the research study.
- ○ No. I do not want to participate in the research study.

Q1.5. Would you like to provide pre-workshop feedback? Your responses would not be used for the research study.

- ◯ Yes. I would like to provide feedback about the workshop and its learning materials
- ◯ No. I do not want to provide feedback.

Demographics

Q2.1.
Hello:

Welcome to the "Data Science for Medical and Biomedical Practitioners" workshop.

This is a student pre-workshop and/or pre self-learning survey.

This survey helps determine learner personas and create tailored (but broad!) workshop materials for individuals to learn data science in the biomedical and health fields.

These survey results will allow us to create materials specific to needs of the group.

Q2.2.
Please create a unique identifier. This unique identifier will be used for long-term assessment but keep your personal information anonymous.

To create an identifier type in:

- Number of siblings (as numeric) +

- First two letters of the city you were born in (lowercase) +
- First three letters of your current street (lowercase).

E.g., (Sherlock Homes has **1** brother, was born in **Po**rsmouth, and lives on **Bac**ker Street - **1pobac**)

Q2.3. Please select the first date of your workshop

- ○ Tuesday, June 29, 2021
- ○ Monday, May 17, 2021
- ○ Tuesday, February 2, 2021
- ○ Wednesday, December 9, 2020
- ○ Tuesday, October 20, 2020
- ○ I went through the online materials on my own
- ○ Other

Q2.4. What is your current affiliation?

- ○ Virginia Tech
- ○ Virginia Tech Carilion School of Medicine
- ○ Fralin Biomedical Research Institute at VTC
- ○ VCOM-Virginia
- ○ Virginia-Maryland College of Veterinary Medicine
- ○ Other

Q2.5. What is your current occupation/career stage (select all that apply).

- [ ] DO/MD
- [ ] DVM
- [ ] RN/PA
- [ ] PhD
- [ ] Academic
- [ ] Analyst
- [ ] Student (Masters e.g., MPH)
- [ ] Student (MD/DO)
- [ ] Student (Nurse, PA)
- [ ] Student (Graduate)
- [ ] Student (Undergraduate)
- [ ] iTHRIV Scholar
- [ ] Other, please describe

Q2.6. What operating system will be on the computer you are using at the workshop or to participate in the online materials?

- ○ Windows
- ○ macOS
- ○ Linux
- ○ Not sure

Q2.7. What gender do you most identify with?

- ○ Female
- ○ Gender variant/non-conforming
- ○ Male
- ○ Prefer not to say
- ○ Prefer to self describe

Q2.8.

How would you describe your ethnic background? Choose one or more of the following groups.

- ☐ American Indian or Alaska Native (Having origins in any of the original peoples of North and South America (including Central America), and who maintains a tribal affiliation or community attachment.)
- ☐ Asian (Having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, Indonesia, the Philippine Islands, Thailand, and Vietnam.)
- ☐ Black or African American (Having origins in any of the Black racial groups of Africa – includes Caribbean Islanders and others of African origin.)
- ☐ Hispanic or Latino(a) (A person of Spanish-speaking origin or ancestry and/or Latin American origin or ancestry – includes Portuguese and Brazilians.)
- ☐ Native Hawaiian or Other Pacific Islander (Having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.)
- ☐ White (Having origins in any of the original peoples of Europe, the Middle East, or North Africa.)
- ☐ I prefer not to say.
- ☐ Prefer to self describe

Persona

Q3.1.

Which of the below personas do you most identify with? Be less concerned about the actual occupation, and more with what relates to your skill and workshop needs.

More detailed descriptions of Alex Academic, Clare Clinician, Patricia Programmer, and Samir Student can be found here: https://ds4biomed.tech/who-is-this-book-for.html

Alex Academic
Alex performs their research using a combination of Excel spreadsheets and specialized software,
but is switching to R or Python (which they taught themself during a sabbatical).
They have never taken a formal programming course, and suffers from impostor syndrome in discussions about programming.
Alex would like to learn more about how programming can help their research and keep up with the tools their students are learning in class.

Alex needs workshops (so they can allocate focused time) and how-to guides (for research).
They would like ready-to-use lesson material that could be remixed for their students and some orientation material to demystify jargon (what is "tidy data"?).
Alex also wants to be able to use the same tools in their research as in their teaching to amortize learning costs and stay in practice.

Clare Clinician
Clare keeps up with medical research, but has little to no experience in doing medical research

process behind data analysis to perform their own analysis and study one day.

Clare wants self-paced tutorials with practice exercises, plus forums where they can ask for help.
They also need short overviews to orient them and introductory tutorials that include videos or animated GIFs showing exactly how to drive the tools, and that use datasets they can relate to.
Clare wishes they had a community of other people in the medical field who are interested in learning how to do data work so they can learn and ask questions.

Patricia Programmer
Patricia regularly connects to a remote server to do their work.
They write SQL statements to pull data out of Epic and processes the data in both Python and R to generate reports and dashboards for their team and management.
Patricia writes data pipelines for all their work either by combining shell scripts or build scripts.

Patricia wants how-to guides and reference material for their day-to-day work and short, intensive online training for very specific topics.
Because they often jump around between various tools, Patricia wants a way to quickly review topics before starting a new project.

Samir Student
Samir is fairly proficient in Excel and does works with spreadsheets regularly and
knows how to load up Excel spreadsheets into R and do basic data

Samir wants a formal workshop and reference materials that can be used to build a good foundation of the programming skills they were never taught. They want a better understanding of the terminology and jargon used in data science so they have the vocabulary to
search for and understand solutions posted online. They are also looking for a community to help in their growth as a student in this domain.

- ○ Alex Academic
- ○ Clare Clinician
- ○ Patricia Programmer
- ○ Samir Student

Prior and background knowledge

Q4.1. How familiar are you with interactive programming languages like Python or R?

- I do not know what those are
- I have heard of them but have never used them before
- I have installed it, but have only done simple examples with them
- I have written a small program with them before
- I use it to automate certain repetitive tasks
- I have small side projects that I program in it
- I program in them for work

Q4.2. Do you know what "long" and "wide" data are?

- I have never heard of the term
- I have heard of it but don’t remember what it is.
- I have some idea of what it is, but am not too clear
- I know what it is and could explain what it pertains to

Q4.3. If you were given a dateset containing an individual's smoking status (binary variable) and whether or not they have hypertension (binary variable), would you know how to conduct a statistical analysis to see if smoking has an increased relative risk or odds of hypertension? Any type of model will suffice.

- I wouldn't know where to start
- I could struggle through, but not confident I could do it
- I could struggle through by trial and error with a lot of web searches
- I could do it quickly with little or no use of external help

Workshop Framing and Motivation

Q5.1. Why are you participating in this workshop? Please check all that apply.

- [ ] To learn new skills
- [ ] To refresh or review my skills
- [ ] To learn skills that I can apply to my current work
- [ ] To learn skills that I can apply to my work in the future
- [ ] To learn skills that will help me get a job or a promotion
- [ ] As a requirement for my program or current position

Q5.2. Please rate your level of agreement with the following statements

	Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
I believe having access to the original, raw data is important to be able to repeat an analysis.	○	○	○	○	○	○	○
I can write a small program, script, or macro to address a problem in my own work.	○	○	○	○	○	○	○
I know how to search for answers to my technical questions online.	○	○	○	○	○	○	○
While working on a programming project, if I got stuck, I can find ways of overcoming the problem.	○	○	○	○	○	○	○
I am confident in my ability to make use of	○	○	○	○	○	○	○

programming language (like R or Python) can make my analyses easier to reproduce.	○	○	○	○	○	○	○
Using a programming language (like R or Python) can make me more efficient at working with data.	○	○	○	○	○	○	○

Q5.3. Please rate your level of agreement with your ability to do the following tasks

	Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
Name the features of a tidy/clean dataset	○	○	○	○	○	○	○
Transform data for analysis	○	○	○	○	○	○	○
Identify when spreadsheets are useful	○	○	○	○	○	○	○
Assess when a task should not be done in a spreadsheet software	○	○	○	○	○	○	○
Break down data processing into smaller individual (and more manageable) steps	○	○	○	○	○	○	○
Construct a plot and table for exploratory data analysis	○	○	○	○	○	○	○
Build a data processing pipeline that can be used in multiple programs	○	○	○	○	○	○	○

statistical analysis of the data

Q5.4. Please share what you most hope to learn from participating in this workshop and/or workshop series.

Q5.5. What do you want to know or be able to do after this workshop (or series of sessions) that you don't know or can't do right now?