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 (i.e., learner personas) 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 and approved 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 student self-assessment 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.

Your responses in this survey will be used to create learner personas. These personas will then be used for the creation and design of workshop materials. The researchers will then take feedback from the workshop to improve materials for each type of learner persona.

This information will be used to create learner personas which gives 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 persona.

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 later surveys, if you do attend workshops, using your de-identified ID.

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/).

Q1.12. **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.

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.

Q1.2. 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.

Q1.3.

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.

Demographics

Q2.1.

Hello:

Welcome to the "data science for medical and biomedical practitioners" workshop.

This is a student self-assessment 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. What is your current occupation/career stage

- ○ DO/MD
- ○ RN/PA
- ○ Academic
- ○ Analyst
- ○ Student (Masters e.g., MPH)
- ○ Student (MD/DO)

- ○ Student (Nurse, PA)
- ○ Student (Graduate)
- ○ Student (Undergraduate)
- ○ Other, please describe

Q2.4. What operating system will be on the computer you are using at the workshop?

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

Q2.5. What gender do you most identify with?

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

Q2.6.

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

Programming Experience

Q3.1. In general, which of these best describes your experience with programming?

- ○ I have none
- ○ I have written a few lines now and again
- ○ I have written programs for my own use that are a couple of pages long
- ○ I have written and maintained larger pieces of software

Q3.2. What programming languages have you used in the past? Select all that apply.

- ☐ VBA (Visual Basic for Applications)
- ☐ Python
- ☐ R
- ☐ Perl
- ☐ Matlab
- ☐ Javascript
- ☐ C
- ☐ C++
- ☐ Fortran
- ☐ Other, please list

Q3.3. 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

Q3.4. How often do you currently use programming languages (R, Python, etc.)?

- ○ Never
- ○ Less than once per year
- ○ Several times per year
- ○ Monthly
- ○ Weekly
- ○ Daily

Q3.5. Which of these best describes how easily you could write a program (in any language) to find the largest number in a list?

- ○ I wouldn't know where to start
- ○ 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

Q3.6. How often do you currently use a specialized software with a point-and-click graphical user interface (e.g., for statistical analysis: SPSS, SAS, …; for Geospatial analysis: ArcGIS, QGIS, … ; for Genomics analysis: Geneious, …)?

- ○ Never
- ○ Less than once per year
- ○ Several times per year
- ○ Monthly
- ○ Weekly
- ○ Daily

Q3.7. How often do you currently use Databases (SQL, Access, etc.)

- ○ Never
- ○ Less than once per year
- ○ Several times per year
- ○ Monthly
- ○ Weekly
- ○ Daily

Data Cleaning and Processing Experience

Q4.1. How familiar are you with Microsoft Excel?

- ○ I have never used it, or I have tried it but can't really do anything with it.
- ○ I have used it as an electronic todo list and planner putting schedules and task deadlines in a single place
- ○ I've used it to store datasets and able to calculate basic aggregate values, such as mean and sums
- ○ I've used data aggregation, pivot tables, formulas, and plotting feature to understand how my data breaks down.
- ○ I've coded up VBA macros and made VOOKUP calls integrating multiple sheets for a simulation task

Q4.2. If you were given a dataset (e.g., Excel file, CSV file) and asked to do some preliminary analysis on it, which of these best describe how easily you can accomplish the task?

- ○ I wouldn't know where to start
- ○ 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

Q4.3. Are you familiar with the term "tidy data"?

- ○ 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

Q4.4. 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

Project and Data Management

Q5.1. Please rate your level of satisfaction with your current data management and analysis workflow (e.g. how you collect, organize, store and analyze your data).

- ○ Very unsatisfied
- ○ Unsatisfied
- ○ Neutral
- ○ Satisfied
- ○ Very satisfied

- ○ Not sure
- ○ Not applicable
- ○ Never thought about this

Q5.2. How do you manage your data and analysis?

- ○ I don't do data and/or analysis work
- ○ My data and analysis are all in excel files, possibly with multiple sheets.
- ○ I work on carefully time-stamped excel files for my version control and analysis
- ○ I use some programming language to load in my data sets for analysis, but sometimes modify my original data files when cleaning the data
- ○ I hold my original data sacred, and only work on it from another program and save out intermediate and final data projects as separate files
- ○ I have a very specific project structure where data and analysis are kept in separate areas and have a version control system (e.g., Git, SVN)
- ○ I have version controlled project templates along with build scripts (e.g., Makefile) to reproduce various aspects of the analysis

Statistics

Q6.1. If you were given a dateset containing 2 cholesterol treatment options (drug and placebo) and a patients cholesterol value 4 weeks after treatment has started, would you

know how to conduct a statistical analysis to see if there is a difference between the 2 groups? Any type of model will suffice.

- ○ I wouldn't know where to start
- ○ 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

Q6.2. 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 by trial and error with a lot of web searches
- ○ I could do it quickly with little or no use of external help

Q6.3. If you were given a dateset comparing different treatment methods for cancer patients. Would you know how to conduct an analysis to see which treatment had a higher survival rate of patients?

- ○ I wouldn't know where to start
- ○ 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

Q6.4. Are you familiar with the term "dummy variable"? It is sometimes also called "one-hot encoding".

- ○ 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

Workshop Framing and Motivation

Q7.1. 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

Q7.2. Please rate your level of agreement with the following statements

	Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
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 software to work with data.	○	○	○	○	○	○	○

	Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
Using a 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.	○	○	○	○	○	○	○

Q7.3. Please share what you most hope to learn from participating in this workshop and/or workshop series.

Q7.4. What do you want to know or be able to do after this workshop that you don't know or can't do right now?