2021 THE VADA PROGRAM

Optional Workshops Schedule

July 5, 6, 8, 22 & 23

9:00 am - 12:00 pm (PDT)

The VADA Program

Visual and Automated Disease Analytics
Graduate Training Program

Monday July 5, 2021

9:00 am – 12:00 pm (PDT)
11:00am – 2:00 pm (CDT)

Workshop Level:
Intermediate

Data Wrangling in the Tidyverse

Speakers: Dr. Robert Balshaw (PhD, Senior Biostatistician, University of Manitoba) and Olawale Ayilara (PhD Candidate, Community Health Sciences, University of Manitoba)

Bio: Rob Balshaw is a Senior Biostatistician with the Data Science Platform at the George and Fay Yee Centre for Healthcare Innovation. Before this, Rob was a Senior Scientist for 5 years at the BC Centre for Disease Control, and for 14 years was head of biostatistics at Syreon Corporation (Vancouver), a contract research organization conducting clinical trials for the pharmaceutical industry. He has been writing programs using statistical languages like SAS and R for more than 25 years, and he is always trying to learn new tools and techniques to make his programs more elegant and effective.

Bio: Olawale's research involves developing and applying methods to automate the analyses of patient-reported outcomes for longitudinal data. He focuses on differential item functioning and latent variable mixture models.

Abstract: If 80% of our analysis time is spent wrangling data (i.e, cleaning it and getting it into the right shape for analysis), improving our data wrangling skills could pay large dividends.

We will review the principles of tidy data and learn how tools from the tidyverse make data wrangling easier and more elegant. We'll illustrate by wrangling a messy data set into shape and see how much easier it is to analyse.

If you are already an R user but you haven't yet discovered tidy data and the tidyverse, you'll never look at R programming the same way after this course. And, whether you use R or some other programming language(*) you'll find that your own data analysis projects will be cleaner, more readable, and easier to build. (* It's okay. We understand. Some of our best friends use SAS, SPSS — even Python.)

University of Manitoba NSERC CRSNG University of Victoria

The VADA Program

Visual and Automated Disease Analytics Graduate Training Program

Requirements: We assume at least basic familiarity with R, RStudio, and writing code using the tidyverse. If this does not apply to you, but you are proficient in another language and/or programming environment, you should do fine. The principles of tidy data apply pretty much universally in data analysis.

Read the paper: Tidy Data paper: Wickham, Hadley. 2014. "Tidy Data." *Journal of Statistical Software* 59 (10). https://www.jstatsoft.org/article/view/v059i10

If you want to run the examples: please install R version 4.0 (or greater), RStudio 1.4 (or greater) and tidyverse package version 1.3 (or greater). However, we probably won't have time for "hands on" exercises.

Excellent overall reference: Grolemund, G., & Wickham, H. (2017). R for Data Science. O'Reilly Media. https://r4ds.had.co.nz/

The VADA Program

Visual and Automated Disease Analytics
Graduate Training Program

Tuesday July 6, 2021

9:00 am – 12:00 pm (PDT)
11:00am – 2:00 pm (CDT)

Workshop Level:
Introductory

How to Data Science with Python

Speaker: Simon Minshall (PhD Student, Assistant Teaching Professor, School of Health Information Science, University of Victoria)

Bio: Simon Minshall, B.Sc. (Westminster, UK) M.Sc. (UVic), Ph.D. Student (UVic). Simon is an Assistant Teaching Professor and a PhD Student at the School of Health Information Science, University of Victoria, BC, Canada.

With over 20 years of professional experience with imaging technologies and software development in the healthcare and entertainment sectors, Simon brings a wide range of skills and knowledge to the School. His research interests include data visualisation, analytics, usability, education, and medical imaging.

Abstract: Computer programming exists to be used as a tool for all scientists. Computer code's concise syntax and complex structures may cause many to leave programming to expert programmers; but, modern, easy to learn languages do not require expertise. Python, with its beautifully expressive code, remains the second most popular programming language. It offers a "batteries-included" approach to writing software. Writing Python scripts that can easily filter and sift through millions of data points remains an essential and accessible data-handling and analytics skill.

About the Workshop: How to Data Science with Python: No Experience Necessary
This short workshop will introduce three main topics: (a) the fundamentals—the parts, often skipped— that you need to understand more difficult concepts, (b) the iteration tools you need to process data, and (c) the helper libraries we leverage to perform complex tasks such as graphing and visualization. All code is delivered using notebook software allowing you to run the examples on your own. The brief lessons are structured with opportunities for you to follow along with the examples shown during the session. Have you never written a line of code during your career? Fear not, this workshop can get you started.

Workshop Setup

University of Manitoba NSERC CRSNG University of Victoria

The VADA Program

Visual and Automated Disease Analytics Graduate Training Program

Unfortunately, there is no time in the workshop to cover installing software; so, please complete the following two tasks before we begin:

1. **Install Python 3**. Follow the instructions on https://installpython3.com/
2. **Install JupyterLab**.

• On the commandline (see apple or windows for how-tos), run the following command:
– pip3 install jupyterlab
– Need greater detail? Refer to: https://jupyter.org/install,
note: follow the instructions under Installation with pip

Running the Notebook

Once you've installed Python 3 and the JupyterLab software, you are ready to begin. JupyterLab is a notebook that allows you to run Python code and intermix it with notes written in markdown format (https://daringfireball.net/projects/markdown/syntax). Do not install markdown, JupyterLab has everything you need.

To start the JupiterLab server, run the following on the commandline:

jupyter-lab

When the JupyterLab server starts on your computer, it opens automatically opens a web browser and shows the notebook. The notebook will be used for writing and reviewing our code. When the notebook appears, you have successfully completed the requirements for running the code notebooks. To shut it down, choose Shutdown from the File menu.

The Workshop Plan:

1. Why programming, why Python?
2. Handling Small Data
3. Don't Repeat Yourself, Iterate
4. Handling large datasets
5. *Break for lunch*
6. Using libraries
7. Visualizing Prevalence
8. Wrapup

The VADA Program

Visual and Automated Disease Analytics Graduate Training Program

Thursday July 8, 2021

9:00 am – 12:00 pm (PDT)
11:00am – 2:00 pm (CDT)

Workshop Level: Intermediate

Information Visualizations for Health Care: A Human Factors Perspective

Speakers: Dr. Helen Monkman (PhD, Assistant Professor, the School of Health Information Science at the University of Victoria) and Susan Martin (MSc, Teaching Assistant, School of Health Information Science at the University of Victoria)

Bio: Dr. Helen Monkman is an Assistant Professor in the School of Health Information Science at the University of Victoria. Her mission is to improve consumer health information systems by making them easier for people to use and the information therein easier to understand. Her work seeks to empower people and help them make better health decisions as well as have better conversations with their health care providers. Her research interests include human factors, user experience, usability, eHealth literacy, digital health literacy, information visualization, and how these factors impact the use and understandability of consumer health information systems.

Bio: Susan Martin is a Teaching Assistant in the School of Health Information Science at the University of Victoria. In the fall of 2020 she was the online Lab Instructor for the inaugural lab component of an existing Health Information Science Course – HINF350: Human Aspects of Health Care Information Systems. She introduced students to the basics of prototyping using Figma and will do so again in the fall of 2021 – hopefully in person! Susan graduated from Health Information Science in 2019 and also has a BA in Anthropology from UVic (2011) and a Masters in International Health from Uppsala University (2014).

Abstract: Information visualizations can be extremely useful for synthesizing data, garnering insight, and communicating complex information quickly and with minimum effort. However, to achieve these goals, visualizations must be designed to accommodate for human capabilities and mitigate human limitations. In this workshop we will explain some of the key human factors principles that should be considered when designing visualizations, let you see them in action, and demonstrate the potential consequences of ignoring them. We will also give students hands on practice applying these principles.

University of Manitoba NSERC CRSNG University of Victoria

The VADA Program

Visual and Automated Disease Analytics Graduate Training Program

Requirements: ensure that Tableau installed on your computer. Free student licences here: https://www.tableau.com/academic/digital-data-skills

The data set/ tableau file will be shared with participants during the workshop.

Thursday July 22, 2021

9:00 am – 12:00 pm (PDT)
11:00am – 2:00 pm (CDT)

Workshop Level:
Intermediate

Creating Decision-Support Dashboard from eICU Relational Database Using Tableau and PowerBI
Speaker: Dr. Dillon Chrimes (PhD, Assistant Teaching Professor, School of Health Information Science, University of Victoria)

Bio: Dr. Dillon Chrimes works as assistant teaching professor at the School of Health Information Science, University of Victoria. Previously, Dillon studied in Canada, Sweden, Austria, Japan and US. He is currently teaching and researching on technical middleware influences on health informatics that embraces big data analytics of hospital system; predictions of usability of electronic health records (EHR); and natural language processing (NLP) aspects of machine learning for text.

Abstract: The workshop lesson investigates electronic records in intensive care units (ICUs) across multiple hospitals while comparing analytical software of Tableau and Power BI. The data set is a relational database demo called eICU (https://physionet.org/content/eicu-crd-demo/2.0.1/). Collaborative Research Database, made publically available this past May 2021. The data were collected through the Philips Healthcare eICU program in collaboration with Massachusetts Institute of Technology (MIT) for ~200,000 admissions (2014-2015) across the United States. The eICU data are unique, de-identified and includes vital signs, care plans, severity of illness measures, diagnosis, treatment, length-of-stay and more. Creating eICU dashboards establishes usability comparisons to effectively visualize clinical decision making. Furthermore, establishing eICU dashboards in this workshop could be further utilized when using the actual, and much larger, eICU (https://physionet.org/content/eicu-crd/2.0/) big data database. Moreover, the freely available nature of the data can support a number of applications including the development of machine learning algorithms, decision support, and clinical research.

University of Manitoba NSERC CRSNG University of Victoria

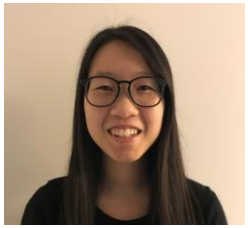
The VADA Program

Visual and Automated Disease Analytics
Graduate Training Program

Friday July 23, 2021

9:00 am – 12:00 pm (PDT)
11:00am – 2:00 pm (CDT)

Workshop Level:
Introductory



Healthcare Big Data Mining Concepts, Techniques and Practice
Speakers: Dr. Alex Kuo (PhD, Professor at the School of Health Information Science, University of Victoria) and Cherry Cheng (MSc)

Bio: Dr. Alex Kuo holds a PhD from the Department of Computer Science, University of Nottingham, UK. He is a full time professor at the School of Health Information Science, University of Victoria (UVic). He was a visiting scholar at the Center for Expanded Data Annotation and Retrieval (CEDAR), School of Medicine, Stanford University, USA (2016), and the Electronic Commerce Resource Centre (ECRC), Georgia Tech (2000).

He is the chair of the IEEE Big Data Education Track at the Big Data Initiative (BDI), and the chair of Special Interest Group on Big Data for Healthcare, Medicine and Biology at IEEE Technical Committee on Big Data.

With over 20 years of programming and data analysis practical as well as research experience, he has over 170 peer-reviewed publications. His research interests include Cloud Computing & Big Data application to healthcare, health data interoperability, health database & data warehousing, data mining application in healthcare, e-health and clinical decision support system.

Bio: Cherry Cheng is an Emergency Registered Nurse currently working as a Clinical Information Specialist in BC. She received an Honours Bachelor of Science in Nursing from The University of British Columbia and an Emergency Nursing Specialty with Distinction from British Columbia Institute of Technology. She has worked at the emergency department (ED) and pioneered colour-coded status events in the ED information system to enhance patient flow and communication, which was adopted regionally across 13 hospitals in the Lower Mainland. Currently, Cherry is part of the provider documentation and electronic medical record design team of a new healthcare information system. In her Master of Science in Health Informatics at UVIC, Cherry has developed her research area focusing on machine learning and the ED overcrowding problem. Her preliminary work was presented at the ICIMTH 2020 conference which received the Best Student Paper Award.

University of Manitoba NSERC CRSNG University of Victoria

The VADA Program

Visual and Automated Disease Analytics Graduate Training Program

Abstract: Big Data in healthcare/life science is different from other disciplines such as social network or transactional business data in that it includes standardized structured, coded data (e.g. ICD, SNOMED CT), semi-structured data (e.g. HL7 messages), unstructured clinical notes, medical images (e.g. MRI, X-rays), genetic lab data, and other types of data (e.g. public health and mental or behavioral health). Huge volumes of very heterogeneous raw data are generated daily by a variety of hospital systems such as Electronic Health Records, Computerized Physician Order Entry, Picture Archiving and Communication Systems, Clinical Decision Support Systems, and Laboratory Information Systems.

However, healthcare Big Data is not usable until they can be aggregated and integrated into a manner that computer can process to generate knowledge. Extracting useful knowledge from such data can be considered as a processing pipeline that involves multiple distinct configuration stages. Each stage faces several specific challenges include: Data Aggregation, Data Maintenance, Data Integration, Data Analysis and Pattern Interpretation.

The objective of this presentation is to teach students Healthcare Big Data Analytics (mining) concepts (challenges and opportunities), techniques (platform and algorithms) and practice examples.

Requirements: no software required.

University of Manitoba NSERC CRSNG University of Victoria