



# The VADA Program

Visual and Automated Disease Analytics  
Graduate Training Program

## Wednesday Program

Time	Activity/Topic	Speaker/Led by
7:45am – 8:30am	Breakfast - A206 Chown	
8:30am – 12:00pm	Statistical methods for Neuroimaging Data Analysis – A&B207 Chown	<b>Dr. Linglong Kong</b> – University of Alberta
12:00pm – 1:00pm	Lunch – John Buhler Research Centre Atrium	
1:00pm – 3:00pm	Techniques for Reproducible Visualization in R: Graph as a Sequence of Three Functions	<b>Dr. Andriy Koval</b> – University of Central Florida
3:30pm – 5:00pm	Networking Reception – Canadian Museum for Human Rights Sifton Terrace	

## Thursday Program

Time	Activity/Topic	Speaker/Led by
7:45am – 8:30am	Breakfast - A206 Chown	
8:30am – 10:30am	Using Github For Data Science: Version Control, Project Management, Promotion – A&B207 Chown	<b>Dr. Andriy Koval</b> – University of Central Florida
10:30am – 12:00pm	TBD	TBD
12:00pm – 1:00pm	Lunch – John Buhler Research Centre Atrium <b>Separate Round Table Discussions:</b> Using data analytics to monitor and transform health care – Dr. Michael Routeledge – 405 Brodie What kind of training do you need to become a successful health data scientist – Dr. Joon Lee – 404 Brodie	<b>Dr. Michael Routeledge</b> – Manitoba Health – 404 Brodie <b>Dr. Joon Lee</b> – University of Calgary – 405 Brodie
1:00pm – 2:30pm	Data Science, AI, and Health: Where are we Headed? - A&B207 Chown	<b>Dr. Joon Lee</b> – University of Calgary
2:30pm – 2:45pm	Refreshment Break – 206 Chown	
2:45pm – 4:30pm	Data visualization strategies and tools for microbial genomic epidemiology - A&B207 Chown	<b>Anamaria Crisan</b> – University of British Columbia



UNIVERSITY OF MANITOBA



NSERC  
CRSNG



University of Victoria



# The VADA Program

Visual and Automated Disease Analytics  
Graduate Training Program

## Friday Program

Time	Activity/Topic	Speaker/Led by
7:45am – 8:30am	Breakfast - A206 Chown	
8:30am – 10:30am	Jupyter Notebook: A Modern Tool for Open, Reproducible, and Distributable Data Science – A&B207 Chown	<b>Adrian Zetner</b> – Public Health Agency of Canada
10:30am – 10:45am	Break	
10:45am – 12:15pm	Career Panel - A&B207 Chown	<b>Dr. Andriy Koval</b> – University of Central Florida <b>Anamaria Crisan</b> – University of British Columbia
12:15pm – 1:15pm	Lunch - John Buhler Research Centre Atrium	
1:15pm – 2:15pm	Debrief and Evaluations	

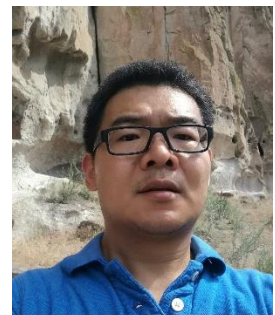
## Session Descriptions and Speaker Biographies

### Statistical Methods for Neuroimaging Data Analysis

Wednesday June 12<sup>th</sup> 8:30am – 12:00pm

**Session description:** With the development of modern imaging techniques, massive imaging data can be observed over both time and space, for example, magnetic resonance imaging (MRI), functional magnetic resonance imaging (fMRI), and diffusion tensor imaging (DTI), advanced statistical methods on imaging data have been proposed, studied and applied in various fields. This short course aims to provide a practical introduction to and an overview of recent advanced statistical challenges and development to analyze and model medical image data quantitatively. The course material is applicable to a wide variety of medical and biological imaging problems. The topics cover tract-based analysis, multi-scaled statistical methods, fMRI processing methods, diffusion imaging methods, brain image and genetics. While presenting the statistical fundamentals, we emphasize the concepts, methods and their real-world implementation. Furthermore, the course material is applicable to a wide variety of medical and biological imaging problems.

**Bio:** Linglong Kong is an associate professor at the department of Mathematical and Statistical Sciences at the University of Alberta. Currently, Currently, Linglong is serving as associate editor of Journal of the American Statistical Association, Applications & Case Studies, International Journal of Imaging Systems and Technology, Canadian Journal of Statistics, and as the ASA Statistical Imaging Session



UNIVERSITY  
OF MANITOBA



NSERC  
CRSNG



University  
of Victoria



# The VADA Program

Visual and Automated Disease Analytics  
Graduate Training Program

program chair. His research interests include high-dimensional data analysis, neuroimaging data analysis, robust statistics and quantile regression, and statistical machine learning.

## Techniques for Reproducible Visualization in R: Graph as a Sequence of Three Functions

Wednesday June 12<sup>th</sup> 1:00pm – 3:00pm

**Session Description:** The session demonstrates a technique for organizing workflows that generate reproducible data visualizations. Rarely do applied data science projects produce any given graph only once. The need to generate plots of the same form using different inputs, different options, and in multiple contexts requires the analyst to structure the operations involved in graph production as customizable functions. The technique demonstrated in this session divides the production of data visualizations into three sections, each governed by a dedicated function. The first function prepares the data for graphing, the second produces the graphic, and the third prints the image to disk. This applied session will walk the learner through the stages of developing such a chain of functions and demonstrate the advantages of such an approach in reproducible projects. Data and starter scripts will be provided. Software in focus: RStudio, R, specifically ggplot2 package.

**Bio:** Dr. Koval is an assistant professor with the Department of Health Management and Informatics at University of Central Florida. His background is in quantitative methods and he has a strong interest in data-driven models of human aging. He received his Ph.D. in Quantitative Methods in 2014 from Vanderbilt University, where he studied statistical modeling, design of graphical displays of information, and reproducible research. Interest in data-driven narratives of human aging lead Dr. Koval to explore the use of electronic health records (EHR) for research and improvement of patient care. Dr. Koval's program of work proposes to develop a system for population health surveillance that would focus on chronic diseases, with particular focus on mental health and substance use (MHSU) conditions, which tend to have high comorbidity rates, polysubstance use patterns, and slowly progressing pace of development.



## Using GitHub for Data Science: Version Control, Project Management, Promotion

Thursday June 13<sup>th</sup> 8:30am – 10:30am

**Session Information:** This session offers a hands-on walk through a minimalistic project designed to introduce the learner to basic operational utility of git and GitHub in data science projects. While git(hub) is best known for providing version control and streamlining collaboration in software development, its applications in data analytic projects can enhance the robustness of the produced solutions and help make results more visible and accessible to the community. The topics to be covered include: 1) project kickoff from a [standard stencil](#), 2) using GitHub client for version control 3) using GitHub Issues for structuring tasks in project development 3) team control and communication and 4) designing an effective README page that faces the public. Data and starter scripts will be provided. Software in focus: R, RStudio, GitHub client.

**Bio:** Dr. Koval is an assistant professor with the Department of Health Management and Informatics at University of Central Florida. His background



UNIVERSITY  
OF MANITOBA



NSERC  
CRSNG



University  
of Victoria



# The VADA Program

Visual and Automated Disease Analytics  
Graduate Training Program

is in quantitative methods and he has a strong interest in data-driven models of human aging. He received his Ph.D. in Quantitative Methods in 2014 from Vanderbilt University, where he studied statistical modeling, design of graphical displays of information, and reproducible research. Interest in data-driven narratives of human aging lead Dr. Koval to explore the use of electronic health records (EHR) for research and improvement of patient care. Dr. Koval's program of work proposes to develop a system for population health surveillance that would focus on chronic diseases, with particular focus on mental health and substance use (MHSU) conditions, which tend to have high comorbidity rates, polysubstance use patterns, and slowly progressing pace of development.

## Data Science, AI, and Health: Where are we Headed?

Thursday June 13th 1:00pm – 2:30pm

**Session Information:** Currently, there is a lot of, and rapidly growing, interest and enthusiasm for how data technologies, artificial intelligence in particular, are poised to transform the entire field of health. This talk will discuss how we got to this point, the current state of cutting-edge health data science research, and where we are headed next. The truly multidisciplinary nature of health data science will be dissected including ethical, equity, and training issues. An overview of research happening at the Data Intelligence for Health Lab in the Cumming School of Medicine, University of Calgary will also be provided.



**Bio:** Dr. Joon Lee is the Director of the [Data Intelligence for Health Lab](#) and an Associate Professor of Health Data Science in the Cumming School of Medicine, University of Calgary. Prior to joining the University of Calgary, he held a faculty appointment in the School of Public Health and Health Systems at the University of Waterloo for 6 years. He holds a PhD in Biomedical Engineering from the University of Toronto and completed postdoctoral training in Medical Data Science at the Harvard-MIT Division of Health Sciences and Technology. His research applies data science, machine learning, artificial intelligence, natural language processing, mobile technology, and biostatistics to a wide range of health domains including intensive care medicine, aging, and population health surveillance. He received the Early Researcher Award from the Ontario Ministry of Research, Innovation and Science in 2016.

## Data Visualization Strategies and Tools for Microbial Genomic Epidemiology

Thursday June 13th 2:45pm – 4:30pm

**Session Information:** New technologies are enabling public health agencies to collect more data of many different types, which can be used to inform public health policy and practice. Yet, this new "big data" is challenging to analyze and to communicate to stakeholders that need to make decisions. In this session I will give a brief introduction to data visualization research – what it is, how you can use it in your own work, and what tools exist to help you visualize your data. I will also discuss my own work for visualization public health microbial genomic epidemiology. I will provide an overview of the different visualization techniques that are used in public health and show how I've used this knowledge to build data visualization tools and a visualization recommender system. As public health continues to evolve, it is becoming critical to build robust and actionable data visualization tools that support growing challenges of public health data-driven decision making.



UNIVERSITY  
OF MANITOBA



NSERC  
CRSNG



University  
of Victoria



# The VADA Program

Visual and Automated Disease Analytics  
Graduate Training Program

**Bio:** Anamaria Crisan is a Vanier Canada Scholar and UBC Public Scholar in her final year of study in Computer Science at the University of British Columbia. Under the joint supervision of Dr. Tamara Munzner and Dr. Jennifer Gardy, she is researching how to visualize the heterogenous collections of genomic and public health data that support investigations of disease outbreaks. Prior to her PhD, Ana worked as a researcher with the British Columbia Centre for Disease Control and, separately, at a Vancouver-based start-up where she was responsible for the research and development of a commercially deployed prostate cancer genomic classifier. She holds a MSc in Bioinformatics from the University of British Columbia and a BSc in Computer Science from Queen's University. You can learn more about her research on her [website](#).



## Jupyter Notebook: A Modern Tool for Open, Reproducible, and Distributable Data Science

Friday June 14th 8:30am – 10:30am

**Session Information:** In the world of modern data science there exists a plethora of development and distribution tools available to the budding data scientist. Foremost amongst these options is the Jupyter Notebook: an open-source, web-based application that simplifies the process of sharing code, results, and associated documentation across multiple languages. This workshop will teach the basics of Jupyter from installation up to distribution using real world examples in R and Python. At the end of this 2-hour session you will have learned the tools and developed an understanding of using modern computational notebooks for reproducible data science.



**Bio:** Adrian Zetner is a computational biologist and member of the Bioinformatics section at the National Microbiology Laboratory. He collaborates with fellow scientists on genomics-related public health projects as well as heading the department's online and in-person training programs. Despite his initial reluctance, he has now embraced Jupyter notebooks as the ideal vehicle for computer-aided scientific analysis and communication. He is a graduate of the University of Manitoba Microbiology program, and a self-taught R champion.



UNIVERSITY  
OF MANITOBA



NSERC  
CRSNG



University  
of Victoria