



Learning and teaching health data science using real-world data science tools

Sanja Lujic and Oscar Perez Concha

Centre for Big Data Research in Health (CBDRH) (on behalf of the postgraduate programs in Health Data Science teaching team)

Abstract:

Health data science (HDS) is an interdisciplinary field lying at the nexus of health, epidemiology, statistics and computer science. Postgraduate HDS programs combine multifaceted areas of healthcare systems, research design, theoretical statistical models, computer programming skills and machine-learning methods in order to manipulate and analyse high-volume, unstructured and complex datasets. To meet the challenge of enabling HDS learners with the necessary content and analytical expertise, the interdisciplinary CBDRH teaching team have developed an integrated learning and teaching environment utilising real-world data science tools as a platform for both blended and fully online streams. The environment combines the leading open-source data analytics platforms (R and Python) and state-of-the art interactive teaching resources (learnr and Jupyter Notebooks) supported by industry-standard version control software (Git, GitHub Classroom and Google Colaboratory). At the heart of the environment are interactive tutorials which weave together narrative text, instructional videos, quizzes and hands-on computer programming exercises in a single easy-to-use interface, run on students' own laptops and deployable in online environments. These are all the same tools used every day by research and industry health data scientists. Demonstration of the integrated environment will be showcased, and its benefits to both students and teaching faculty discussed.