Data Science: An Introduction to the Computational Tools and Techniques for Large-scale Biomedical Research

John Darrell Van Horn, Ph.D.

USC Mark and Mary Stevens Neuroimaging and Informatics Institute, Laboratory of Neuro Imaging (LONI), University of Southern California

It has been noted that data science and its large-scale analysis represent a wholly new research paradigm (Szalay and Gray 2006). With the onslaught of rich biomedical data types related to the brain as well as to drugs and their abuse (Moningka *et al.* 2018), these methods form the basis for several lines of modern neuroscientific examination (Van Horn and Toga 2014, Toga *et al.* 2015). In this presentation, I will review the motivations behind data science, several computational tools, and feature examples of where and how it has utility for the 'big data' examination of the human brain. Important in this is the education and training of the next generation of biomedical data scientists (Van Horn 2016, Van Horn *et al.* 2018). This presentation seeks to set the tone for subsequent presentations in this novel and forward-thinking scientific session.

References
Moningka, H., Lichenstein, S., Worhunsky, P. D., DeVito, E. E., Scheinost, D. and Yip, S. W. (2018). "Can neuroimaging help combat the opioid epidemic? A systematic review of clinical and pharmacological challenge fMRI studies with recommendations for future research." Neuropsychopharmacology.
Szalay, A. and Gray, J. (2006). "2020 computing: science in an exponential world." Nature **440**(7083): 413-414.
Toga, A. W., Foster, I., Kesselman, C., Madduri, R., Chard, K., Deutsch, E. W., Price, N. D., Glusman, G., Heavner, B. D., Dinov, I. D., Ames, J., Van Horn, J., Kramer, R. and Hood, L. (2015). "Big biomedical data as the key resource for discovery science." J Am Med Inform Assoc **22**(6): 1126-1131, PMC: PMC5009918.
Van Horn, J. D. (2016). "Opinion: Big data biomedicine offers big higher education opportunities." Proceedings of the National Academy of Sciences **113**(23): 6322-6324.
Van Horn, J. D., Fierro, L., Kamdar, J., Gordon, J., Stewart, C., Bhattrai, A., Abe, S., Lei, X., O'Driscoll, C., Sinha, A., Jain, P. C., Burns, G., Lerman, K. and Ambite, J. L. (2018). "Democratizing data science through data science training." Proceeding of the Pacific Symposium on Biocomputing **23**: 240-246.
Van Horn, J. D. and Toga, A. W. (2014). "Human neuroimaging as a "Big Data" science." Brain Imaging Behav **8**(2): 323-331, PMC: PMC3983169.