

Data Science Skills & SOCR Interactive Tools

Ivo D. Dinov

Statistics Online Computational Resource

Health Behavior & Biological Sciences
Computational Medicine & Bioinformatics
Neuroscience Graduate Program
Michigan Institute for Data Science

University of Michigan



<http://SOCR.umich.edu>

Slides Online:
"SOCR News"



SCHOOL OF NURSING

STATISTICS ONLINE COMPUTATIONAL RESOURCE (SOCR)
UNIVERSITY OF MICHIGAN

SOCR Cloud Services

- ☐ Datasets & Case-studies
- ☐ Visualization/EDA, e.g., Big Neuroscience Data
- ☐ Student Learning/Refresher Modules (Activities)
- ☐ R&D Projects & Collabs (DataSifter, CBDA, Spacekime)
- ☐ Instructional Resources (EBooks)
- ☐ Tools, Apps, Demos, Workflows
 - ☐ Java Applets (500+), require Java-enabled browsers
 - ☐ HTML5 Apps
 - ☐ End-to-end computational protocols (workflows)

It's Online, Therefore, it Exists!

http://wiki.socr.umich.edu/index.php/SOCR_News_BrockU_Symposium_2019



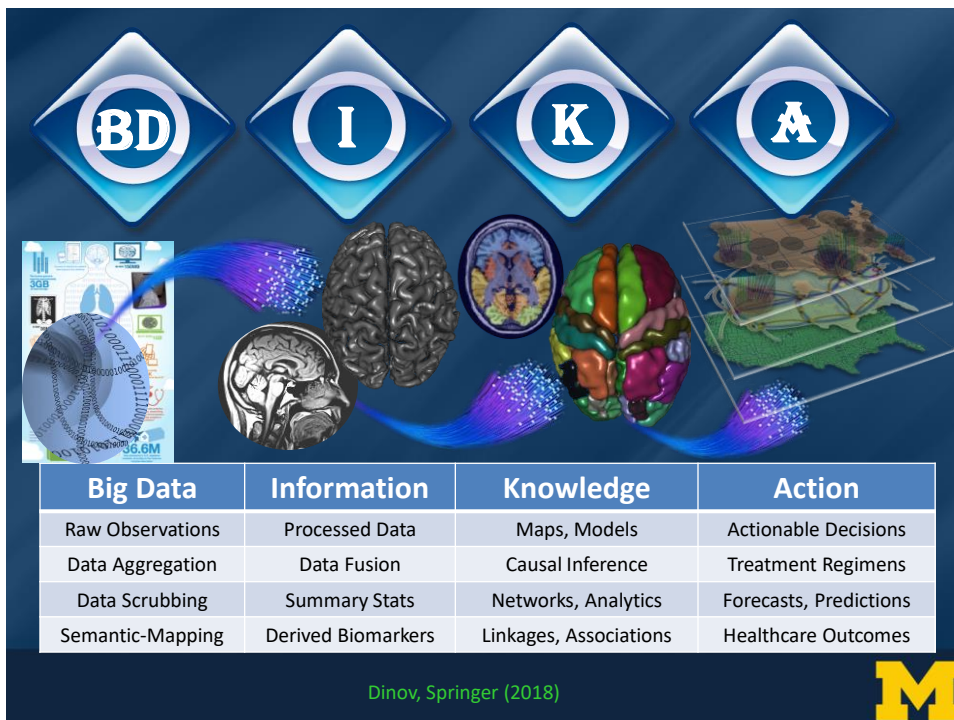
- OOC Topics Covered: Data Science and Predictive Analytics
- DSPA Motivation
- Examples of Big Biomedical Challenges (AD, PD, ALS, RWD)
- Brain Visualization
- Neurodegeneration
- Genomics computing
- Neuroimaging pipelines
- Common Characteristics of Big (Biomedical and Health) D
- High-throughput Big Data Analytics
- Statistical Software Pros/Cons Comparison
- Getting started
- Python-based R
- Reproducible Research (RStudio)
- R Notebook Layout
- Applications of Matrix Algebra: Linear modeling
- Simplification of complex nonlinear models using calculus
- Least Square Estimation
- T/S: R Im Function
- Eigenvalues and Eigenvectors
- Other important functions
- Variance Inflation
- Linear Regression Manipulation
- Kernels and Principal Component Analysis (internal statistical validation)
- Latent Dirichlet Model
- Random Forest and PCA
- Support Vector Classification and SVM performance
- Parameter sensitivity for automated parameter tuning
- Confusion matrix and ROC curve (AUC)
- Local Binary Patterns for Image Embedding (t-SNE)
- Empirical performance with meta-learning
- Distance metrics for finding nearest neighbors
- Distance Metric
- Euclidean distance
- Manhattan Distance
- Preprocessing for Support k
- Preprocessing for Support KNN
- Preprocessing for Support Performance
- Mining Biometrics
- Software tooling for data structures/databases
- File formats and file management Files
- Formalizing the concept of context of web pages
- Formalizing the concept of context
- Using Statistical Methods to Analyze mean and median
- Using Statistical Methods to Analyze variance and standard deviation

- Analytics (S, RWD)
 - Stats/Prob Tables: <http://socr.umich.edu/Applets>
 - SMHS: <http://wiki.socr.umich.edu/index.php/SMHS>
 - Probability Distributome: <http://Distributome.org>
- R&D Projects & Collaborations
 - DataSifter: <http://DataSifter.org>
 - CBDA: <http://socr.umich.edu/HTML5/CBDA>
 - Spacekime: <http://Spacekime.org>
- Instructional Resources (EBooks)
 - Prob & Stats: <http://wiki.socr.umich.edu/index.php/EBook>
 - SMHS: <http://wiki.socr.umich.edu/index.php/SMHS>
 - DSPA: <http://DSPA.predictive.space>



Tools, Apps, Demos, Workflows

- ❑ SOCR Navigators:
<http://socr.umich.edu/html/Navigators.html>
- ❑ Java Applets (500+), require Java-enabled browsers
 - ❑ <http://www.socr.umich.edu>
- ❑ HTML5 Apps
 - ❑ t-SNE: http://socr.umich.edu/HTML5/SOCR_TensorBoard_UKBB/UsersData.html
 - ❑ Motion Charts: <http://socr.umich.edu/HTML5/MotionChart>
- ❑ End-to-end computational protocols (workflows)
 - ❑ Pipeline workflows: <http://pipeline.loni.usc.edu>
 - ❑ SOCRAT: <http://socr.umich.edu/HTML5/SOCRAT>



Acknowledgments

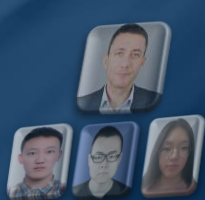
Slides Online:
"SOCR News"

Funding

NIH: P20 NR015331, U54 EB020406, P50 NS091856, P30 DK089503, UL1TR002240, R01CA233487
NSF: 1916425, 1734853, 1636840, 1416953, 0716055, 1023115

Collaborators

- ❑ **SOCR:** Milen Velez, Yongkai Qiu, Zhe Yin, Yufei Yang, Alexandr Kalinin, Selvam Palanimalai, Syed Husain, Matt Leventhal, Ashwini Khare, Rami Elkest, Abhishek Chowdhury, Patrick Tan, Pratyush Pati, Brian Zhang, Juana Sanchez, Dennis Pearl, Kyle Siegrist, Rob Gould, Nicolas Christou, Hanbo Sun, Tuo Wang, Yi Wang, Lu Wei, Lu Wang, Simeone Marino
- ❑ **LONI/INI:** Arthur Toga, Roger Woods, Jack Van Horn, Zhuowen Tu, Yonggang Shi, David Shattuck, Elizabeth Sowell, Katherine Narr, Anand Joshi, Shantanu Joshi, Paul Thompson, Luminita Vese, Stan Osher, Stefano Soatto, Seok Moon, Junning Li, Young Sung, Carl Kesselman, Fabio Macciardi, Federica Torri
- ❑ **UMich MIDAS/MNORC/AD/PD Centers:** Cathie Spino, Chuck Burant, Ben Hampstead, Kayvan Najarian, Stephen Goutman, Stephen Strobbe, Hiroko Dodge, Hank Paulson, Bill Dauer, Roger Albin, Chris Monk, Issam El Naqa, HV Jagadish, Brian Athey



<http://SOCR.umich.edu>