

Exploratory Big Data Analytics

Ivo D Dinov

Center for Complexity and Self-management of Chronic Disease (CSCD)
Statistics Online Computational Resource (SOCR)
Michigan Institute for Data Science (MIDAS)
School of Nursing
University of Michigan



www.socr.umich.edu/CSCD
www.MIDAS.umich.edu





The Center for Complexity and Self-management of Chronic Disease (CSCD) is funded by the National Institute of Health (NIH) and the National Institute for Nursing Research (NINR), [grant P20 NR015331](#). CSCD investigators espouse that optimal living with one or more chronic conditions and/or adopting and maintaining behaviors to promote health, avoid disease and maximize function involves the integration of self-management behaviors into one's life. This integration is influenced by many factors at various points in time. The chronic illness trajectory is not static, nor are contextual or process factors and, therefore, research in SM needs to be able to competently address numerous changing relationships. Chaos theory can help modeling processing and data analytics as it takes a complex systems approach and makes the assumption that behavior can be chaotic with variance that is not neatly or linearly explained. Complexity theoretical approaches study random events that can cause change and where small changes in behavior can have huge effects at the system level. CSCD investigators employ complexity theory and multivariate data analytics methods to address challenges in managing chronic disease.

[About CSCD >](#)



CSCD Team

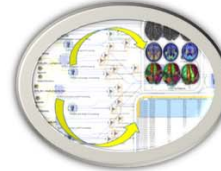
The CSCD Center advances self-management research and strives to partner with a variety of stakeholders including researchers, patients, policy makers and others interested in effectively and efficiently managing chronic diseases. Junior researchers and investigators may seek CSCD pilot funds and mentorship with establishing interdisciplinary research teams and exposure to innovative methods and analytics. Mid-career and senior level investigators may engage in CSCD

Computational Infrastructure



CSCD Research projects are supported by databases (MySQL/NoSQL), computational services and high-throughput processing workflow. In addition to the [UMich FLUX services](#), CSCD investigators have access to a powerful **1.5TB RAM, 20TB storage, 40-core SOCR Server** facilitating storage, retrieval, processing and visualization of heterogeneous data.

Methodological Consulting



CSCD faculty provide consultations on study-designs, methodological challenges and service recipients. For instance, collectively the CSCD investigators have **broad and deep expertise in Scientific Visualization**, which is essential in interrogating Big healthcare data. **Pilot project** investigators and other researchers are encouraged to [contact the Center](#) and discuss their specific needs.

Data Analytics



Data acquisition, management, aggregation, processing and analysis are important components of all quantitative healthcare studies. CSCD investigators provide support for data analytics, high-throughput data analysis, protocol design, validation and open-dissemination. You can [see here](#) one example of a high-throughput data analytics workflow ([trans-proteomics pipeline](#)).

Training Opportunities



CSCD provides learning materials, instructional resources, refreshers and self-guided activities supporting studies of complexity, self-management of chronic disease, and [data analytics](#). The Center also periodically organizes [training events and educational activities](#).



Home	→
Data Science Initiative	→
About MIDAS	+
Affiliated Faculty	→
Education & Training	+
Challenges Initiatives	→
Industry Engagement	→
Events	→
Grant Opportunities	→
Contact Us	→

MICDE
computational science

MIDAS
data science

Please join us for the MIDAS inaugural symposium, **The Future of Data Science: A Convergence of Academia, Industry, and Government** at the Rackham Building on Oct. 6.

« All Events

The Future of Data Science: A Convergence of Academia, Industry, and Government

October 6

You are cordially invited to join top industry, academic and government data scientists in Ann Arbor for an inaugural symposium to mark the launch of the University of Michigan Data Science Initiative.

On October 6, experts from around the country will convene at Rackham Amphitheater to focus on the future of data science and to highlight current research, trends and emerging issues in the field that transcend all aspects of our global community.

University of Michigan (U-M) leaders will share details regarding major new investments in data science research, instruction, and infrastructure. Internationally renowned scientists will present on cross-cutting data

University leaders

- *Martha Pollack, Provost and EVPA*
- *Jack Hu, Interim Vice President for Research*
- *Eric Michielssen, Associate Vice President, ARC*
- *Brian Athey, Co-Director, MIDAS*
- *Alfred Hero, Co-Director, MIDAS*

Internationally renowned scientists

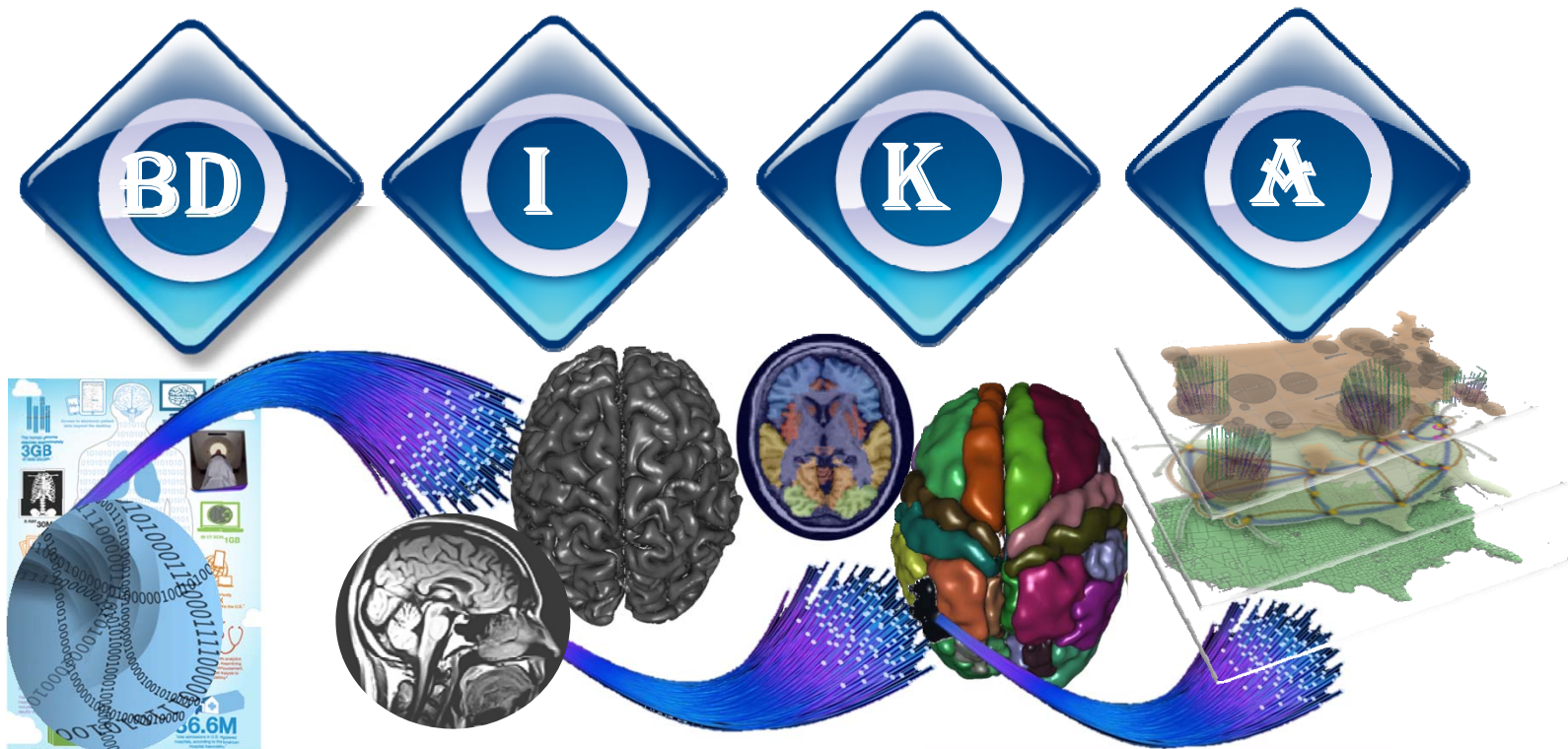
- *Robert Nowak, Engineering, Wisconsin-Madison*
- *Susan Murphy, Statistics, Michigan*
- *Kathleen McKeown, Data Sciences and Engineering, Columbia*

An introduction to the MIDAS Education and Training Program

- *Ivo Dinov, Associate Director, MIDAS*
- *Erin Shellman, Research Scientist, AWS*
- *Patrick Harrington, Director of Engineering, WalmartLabs*
- *Nandit Soparkar, Chief Executive Officer, Ubiquiti*

Keynote presentations

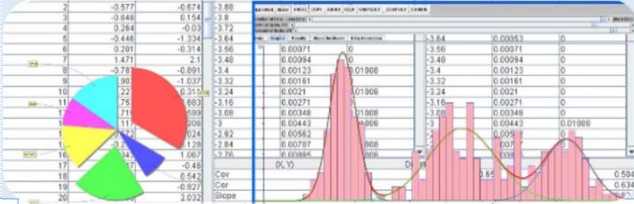
- *Daniel Goroff, VP, Sloan Foundation*
- *George Poste, Health Innovation, Arizona State*
- *Bror Saxberg, Chief Learning Officer, Kaplan*
- *Kathleen Carley, Computation, Carnegie Mellon*
- *Jonathan Owen, Operations Research, General Motors*
- *Ed Seidel, Physics and Astronomy, Illinois*
- *Kathleen McKeown, Data Sciences and Engineering, Columbia*
- *Ratna "Babu" Chinnam, Engineering, Wayne State*
- *Yike Guo, Computing Science, Imperial College London*
- *Keith Elliston, CEO, tranSMART Foundation*



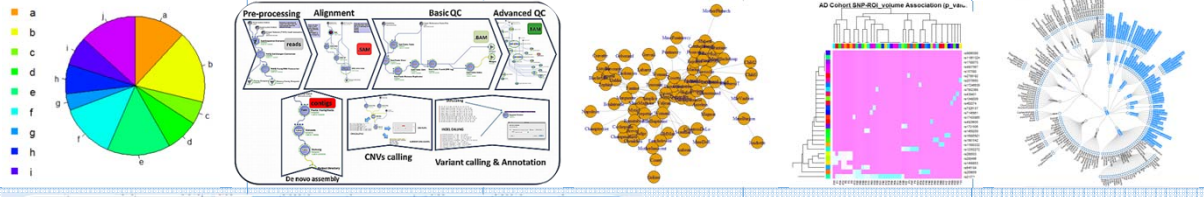
Big Data	Information	Knowledge	Action
Raw Observations	Processed Data	Maps, Models	Actionable Decisions
Data Aggregation	Data Fusion	Causal Inference	Treatment Regimens
Data Scrubbing	Summary Stats	Networks, Analytics	Forecasts, Predictions
Semantic-Mapping	Derived Biomarkers	Linkages, Associations	Healthcare Outcomes

Neuroimaging-Geno-Pheno Development/Aging

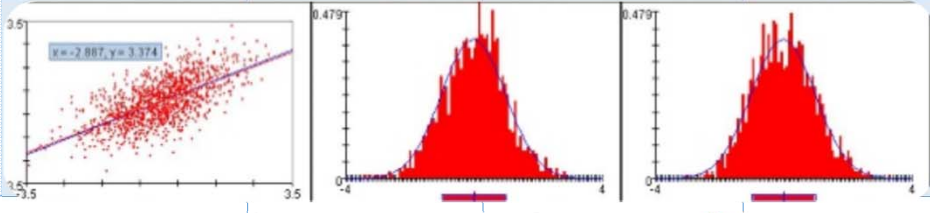
**1
Factor**



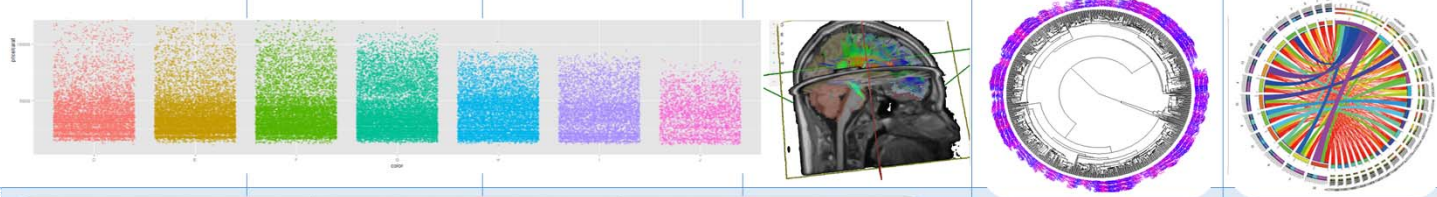
**Many
Factors**



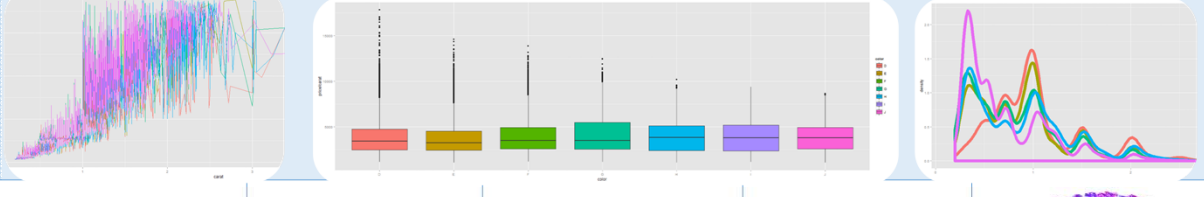
**Uni-
Variate**



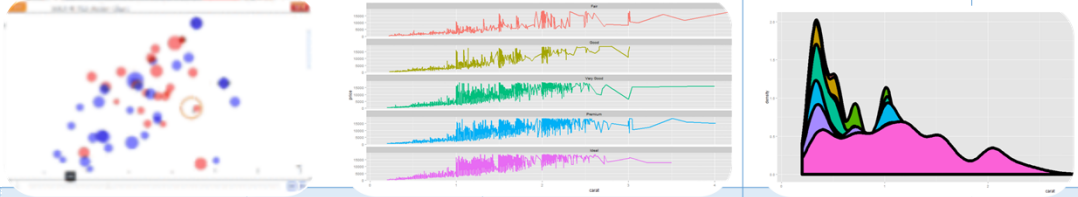
**Multi-
Variate**



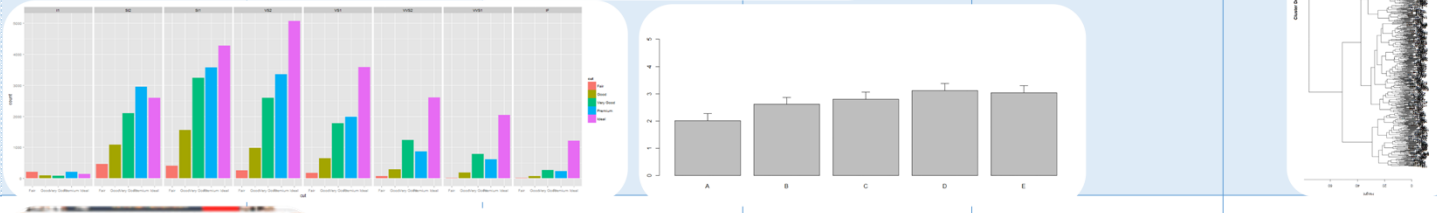
**2
Variables**



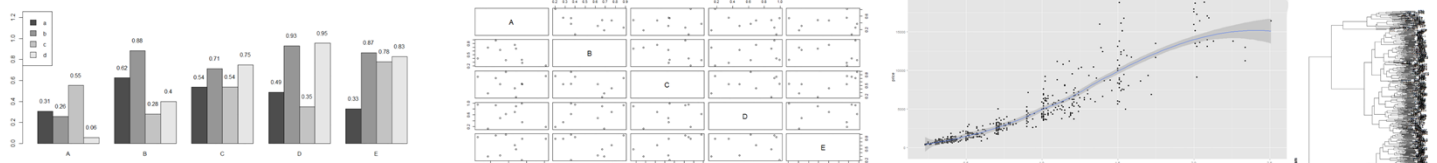
**Multi-
Variate**



**Across
Categories**



**Across
Time**



Demos

- **SOCR Charts** (<http://www.socr.umich.edu/html/cha/>)
 - EDA: Pie Chart (Ring Chart)
 - 3D Chart (http://socr.umich.edu/html/cha/SOCR_3DCharts.html)
 - GIS/Cartography (http://socr.umich.edu/html/cha/SOCR_Cartography.html)
 - Motion Charts (<http://socr.umich.edu/HTML5/MotionChart>),
HPI (default) and Ozone data
(http://wiki.socr.umich.edu/index.php/SOCR_Data_121608_OzoneData)
- **Data Dashboard** (<http://socr.umich.edu/HTML5/Dashboard>), Husain, *et al.*, 2015
- **Brain Viewer** (<http://socr.umich.edu/HTML5/BrainViewer>)
- **Hierarchical Data Visualization**
 - SOCR Navigators (<http://socr.umich.edu/html/Navigators.html>)
 - EBook Navigator (http://socr.umich.edu/html/navigators/D3/SOCR_D3_Viewer.html)
 - Circos (<http://mkweb.bcgsc.ca/tableviewer>), AD Data, Moon, Dinov, *et al.*, 2015
- **Cloud Data**
 - Alzheimer's Disease Neuroimaging Initiative (ADNI), <http://adni.loni.usc.edu>
 - dbGaP (<http://www.ncbi.nlm.nih.gov/gap>)
 - Parkinson's Progression Markers Initiative (PPMI), <http://www.ppmi-info.org/data>
- **Pipeline Workflows** (<http://pipeline.loni.usc.edu>), PWS, Navigator

