# 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

School of Nursing
Center for Complexity and
Self-management of
Chronic Disease (CSCD)
University of Michigan

[www.SOCR.umich.edu](http://www.SOCR.umich.edu)
[www.MIDAS.umich.edu](http://www.MIDAS.umich.edu)

Michigan Institute
for Data Science
University of Michigan

SCHOOL OF NURSING
STATISTICS ONLINE COMPUTATIONAL RESOURCE (SOCR)
UNIVERSITY OF MICHIGAN

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*

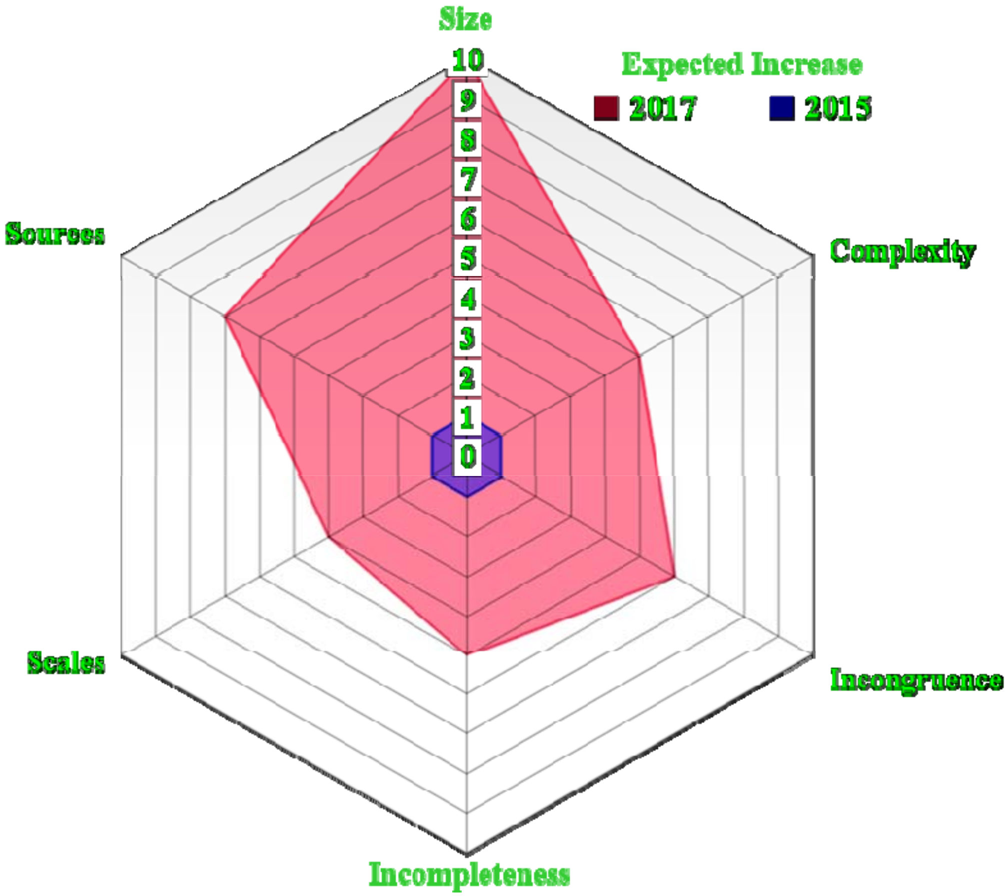
An introduction to the MIDAS Education and Training Program

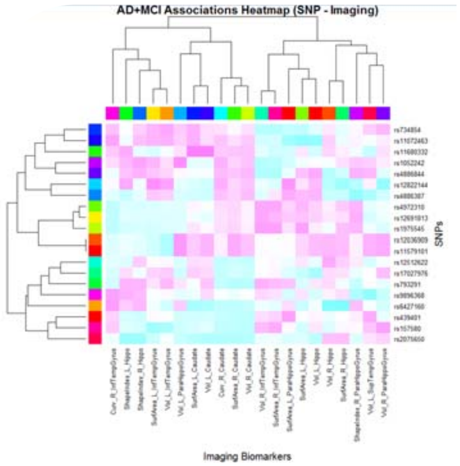
- *Ivo Dinov, Associate Director, MIDAS*
- *Erin Shellman, Research Scientist, AWS*
- *Patrick Harrington, Director of Engineering, WalmartLabs*
- *Nandit Soparkar, Chief Executive Officer, Ubiquiti*

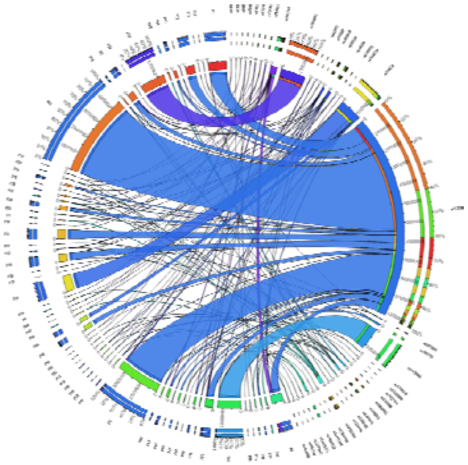
Keynote presentations

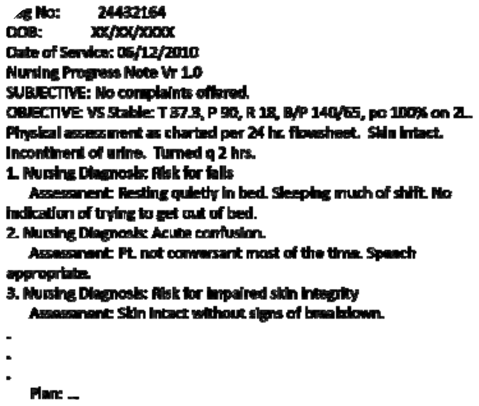
- *Daniel Goroff, VP, Sloan Foundation*
- *George Poste, Health Innovation, Arizona State*
- *Bror Saxberg, Chief Learning Officer, Kaplan*
- *Kathleen Carley, Computation, Carnegie Mellon*
- *Jonathan Owen, Operations Research, General Motors*
- *Ed Seidel, Physics and Astronomy, Illinois*
- *Kathleen 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 Characteristics







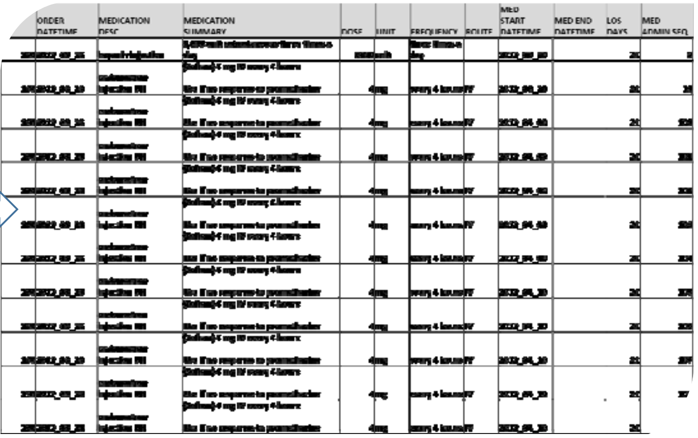


g No: 24432164
DOB: XX/XX/XXXX
Date of Service: 06/12/2010
Nursing Progress Note Vr 1.0
SUBJECTIVE: No complaints offered.
OBJECTIVE: VS Stable: T 37.3, P 90, R 18, B/P 140/65, po 100% on 2L.
Physical assessment as charted per 24 hr. flowsheet. Skin intact.
Incontinent of urine. Turned q 2 hrs.

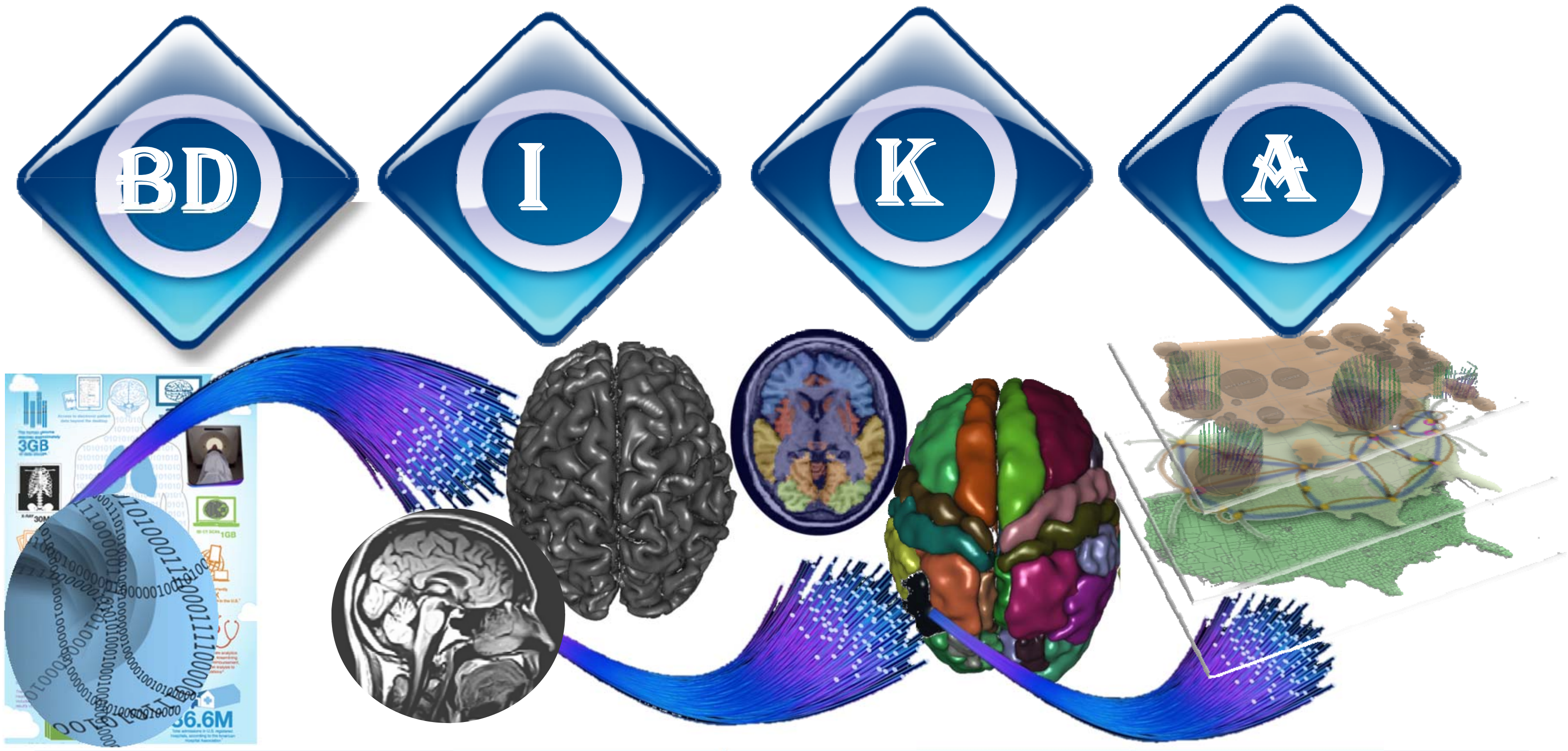
1. Nursing Diagnosis: Risk for falls
   Assessment: Resting quietly in bed. Sleeping much of shift. No indication of trying to get out of bed.
2. Nursing Diagnosis: Acute confusion.
   Assessment: Pt. not conversant most of the time. Speech appropriate.
3. Nursing Diagnosis: Risk for impaired skin integrity
   Assessment: Skin intact without signs of breakdown.

-
-
-

Plan: ...

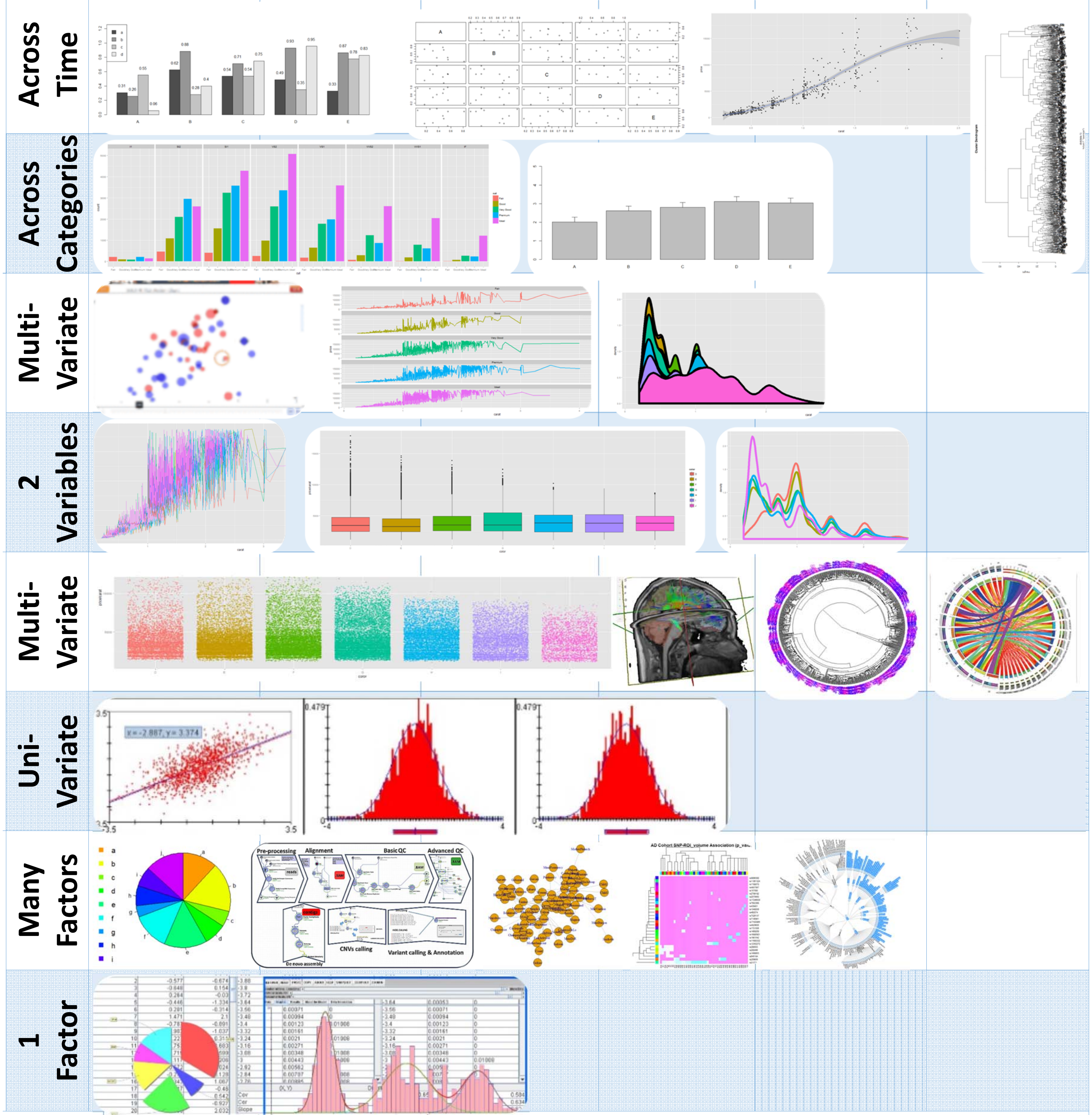


| ORDER DATETIME | MEDICATION DESC | MEDICATION SUMMARY | DOSE | UNIT | FREQUENCY | ROUTE | MED START DATETIME | MED END DATETIME | LOS DAYS | MED ADMIN SEQ |
|---|---|---|---|---|---|---|---|---|---|---|



| 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


Across Time
Across Categories
Multi-Variate
2 Variables
Multi-Variate
Uni-Variate
Many Factors
1 Factor
Pre-processing
Alignment
Basic QC
Advanced QC
De novo assembly
CNVs calling
Variant calling & Annotation
AD Cohort SNP-ROI_volume Association (p_valu
0.479
-4
4
-3.5
3.5

# 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