

W07 Hands-On Full Life Cycle Data Science Workshop Preparation



You should have received an email prior to this workshop with download instructions for Anaconda (Ananconda.com). If you did not receive or install you can still do the workshop by going to <https://colab.research.google.com>

Go to AMIA program website <https://informaticssummit2019.zerista.com/>, log in or register, go to the WS-07 section and you will see a link to Workshop Notebook Download Instructions.

All of the materials for this workshop are available on github at:
https://github.com/joh06288/AMIA2019_W07





W07 Hands-On Full Life Cycle Data Science Workshop

Steve Johnson PhD, University of Minnesota
Tamara Winden PhD, University of Kansas
Lisiane Pruinelli PhD, University of Minnesota



Disclosure



Lisiane Pruinelli: I have no relevant relationships with commercial interests to disclose.

Steve Johnson: I and my spouse/partner have no relevant relationships with commercial interests to disclose.

Tamara Winden: I and my spouse/partner have no relevant relationships with commercial interests to disclose.

Learning Objectives

After participating in this session the learner should be better able to:

- Apply data science techniques to health care data. Participants will have discussed the challenges of working with EHR data including issues of data quality, data complexity and non-standardization.
- Conduct hands-on data science including exploratory data analysis, data preparation and model development.
- Interact with and take away a fully functioning environment with synthetic data and step-by-step recipes for a data science project.

Introduction



Introductions

- Speakers
 - Steve Johnson, Director, Informatics Innovation, Assistant Professor, University of Minnesota
 - Tamara Winden, Chief Research Informatics Officer, Assistant Professor, University of Kansas Med Ctr
 - Lisiane Pruinelli, Assistant Professor, School of Nursing, University of Minnesota

Housekeeping

- Break if needed after Exploratory Data Analysis section
- This is an interactive session and a large group so we encourage asking for help and helping each other where possible
- We're here to learn together

Agenda

1. Background
2. Research question
3. Jupyter Notebook
4. Data Preparation
5. Exploratory Data Analysis
6. Modeling and Analytics
7. Model Evaluation
8. Deployment

Background

- EHR = large electronic data sets that will help answer operational and clinical questions.
- Combining data sets from multiple organizations results in truly big data of tens of millions of patients to address population health and inform clinical research.
- Challenge of health care data
 - Data quality
 - Lack of discrete data: benefit vs burden of documentation
 - Lack of standards
 - Extremely complex
 - Data privacy and security

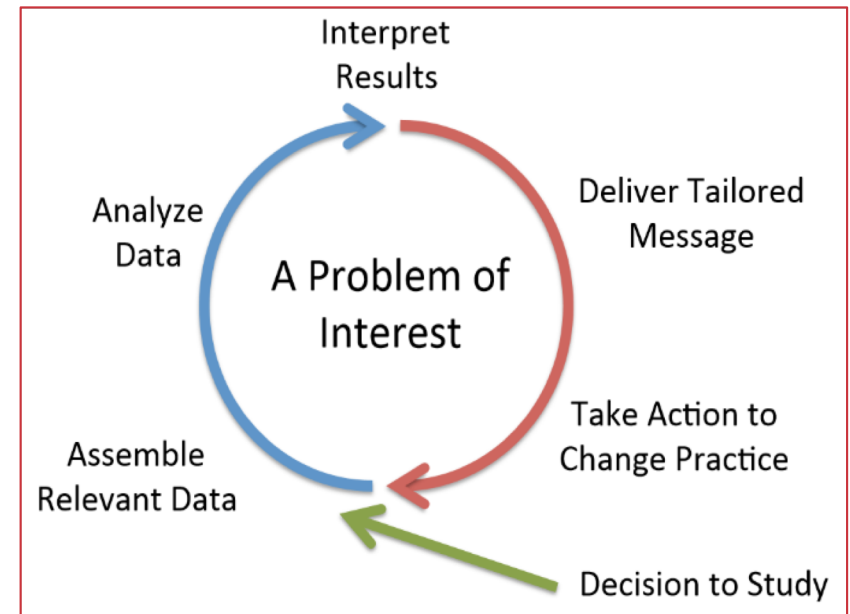
Background

- **Today's Goal:** Convince you that future data science could be as easy as spreadsheets
- Solving real-world problems using data science
 - A data scientist is a data analyst who lives in California!
 - “Data science is a field with a broad scope, encompassing approaches for generation, characterization, management, storage, analysis, visualization, integration and use of large, heterogeneous data sets that have relevance to population health.¹”
- This workshop will use a hands-on approach to demonstrate big data science to illustrate these issues and discuss approaches to dealing and analyzing them for better health care initiatives.
- Hands on workshop experiment...don't panic!

1. NOT-LM-17-006: Request for Information (RFI): Next-Generation Data Science Challenges in Health and Biomedicine.

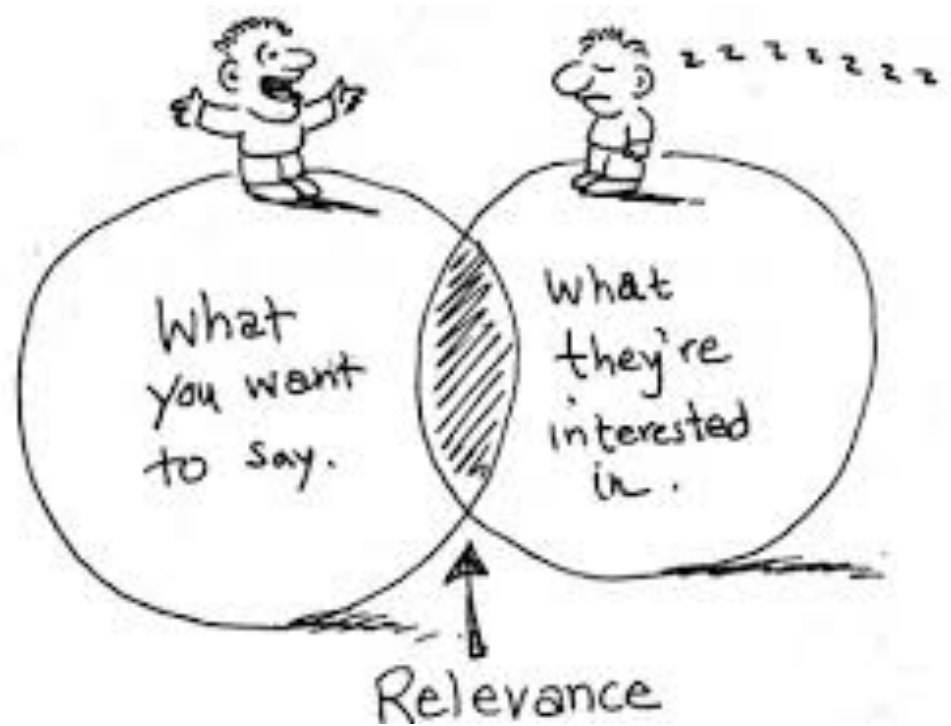
The Learning Health System

- “Science, informatics, incentives, and culture are aligned for continuous improvement and innovation, with best practices seamlessly embedded in the delivery process and new knowledge captured as an integral by-product of the delivery experience.” IOM
<http://www.learninghealthcareproject.org/index.php>
- *Enabling multi-disciplinary research to accelerate clinical outcomes*
- **Toward Complete & Sustainable Learning Systems** Charles P. Friedman, PhD Josiah Macy, Jr. Professor Chair, Department of Learning Health Sciences Professor of Information and Public Health University of Michigan. December 8, 2014 https://medicine.umich.edu/sites/default/files/2014_12_08-Friedman-IOM%20LHS.pdf
- **Mind the Gap: Putting Evidence into Practice in the Era of Learning Health Systems.** [Quise JM](#), [Savitz LA](#), [Friedman CP](#).



Relevance of the Research Question

- Clinical Relevance
- Funding Alignment
- Data Availability
- Domain Expertise on the Research Team



The Research Question

FINER Criteria for Good Research Questions

| | | |
|---|-------------|---|
| F | Feasible | <ul style="list-style-type: none">•Adequate number of subjects•Adequate technical expertise•Affordable in time and money•Manageable in scope |
| I | Interesting | <ul style="list-style-type: none">•Getting the answer intrigues investigator, peers and community |
| N | Novel | <ul style="list-style-type: none">•Confirms, refutes or extends previous findings |
| E | Ethical | <ul style="list-style-type: none">•Amenable to a study that institutional review board will approve |
| R | Relevant | <ul style="list-style-type: none">•To scientific knowledge•To clinical and health policy•To future research |

Farrugia P, Petrisor BA, Farrokhyar F, Bhandari M. Research questions, hypotheses and objectives. *Canadian Journal of Surgery*. 2010;53(4):278-281.

Our Research Question Today



Can we predict which patients might
overdose on opioids?



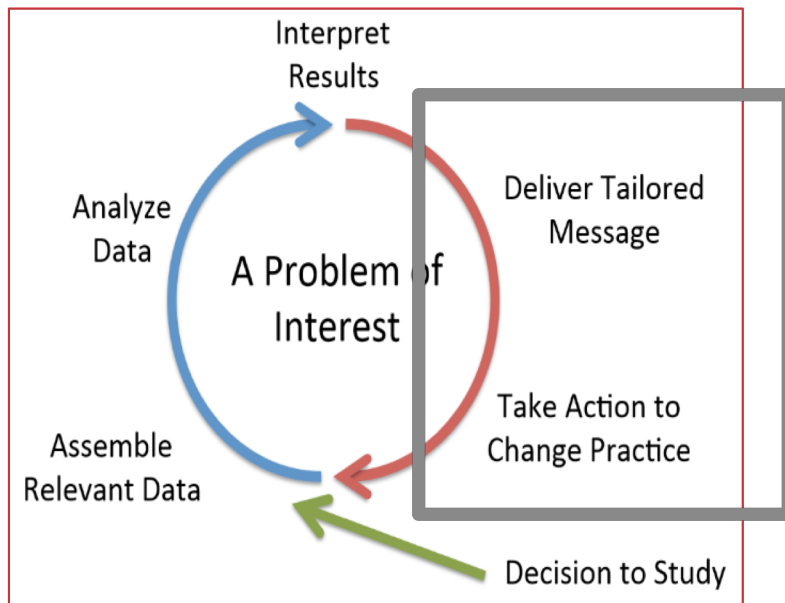
Agenda

1. Background
2. Research question
3. Jupyter Notebook
4. Data Preparation
5. Exploratory Data Analysis
6. Modeling and Analytics
7. Model Evaluation
8. Deployment

Agenda

1. Background
2. Research question
3. Jupyter Notebook
4. Data Preparation
5. Exploratory Data Analysis
6. Modeling and Analytics
7. Model Evaluation
- 8. Deployment**

Clinical Decision Support



- “A process for enhancing health-related decisions and actions with pertinent, organized clinical knowledge and patient information to improve health and healthcare delivery¹.”
- Deliver the right information to the right person in the right format through the right channel at the right time in the workflow.

Toward Complete & Sustainable Learning Systems Charles P. Friedman, PhD Josiah Macy, Jr. Professor Chair, Department of Learning Health Sciences Professor of Information and Public Health University of Michigan. December 8, 2014
https://medicine.umich.edu/sites/default/files/2014_12_08-Friedman-IOM%20LHS.pdf

1. Osheroff, J.A., Teich, J. M., & Levick, D. (2012). Improving outcomes with clinical decision support: An implementer's guide, second ed. Healthcare Information and Management Systems Society (HIMSS), Chicago, IL.

Deployment: To the bedside...

- Clinical Decision Support (CDS) Tools

Clinical decision support (CDS) provides clinicians, staff, patients or other individuals with knowledge and person-specific information, intelligently filtered or presented at appropriate times, to enhance health and health care. CDS encompasses a variety of tools to enhance decision-making in the clinical workflow.

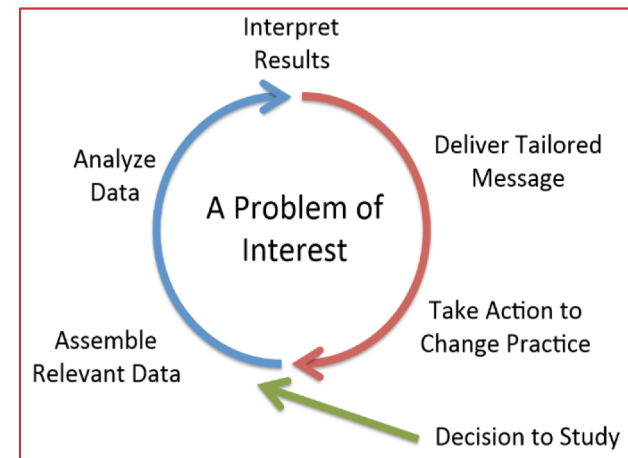
- Pop-up message
- Passive messages
- Snapshot views
- Reporting tools (reports, dashboards)
- Orders/order sets

- Data visualization techniques

- Edward Tufte, cognitive science theory

- Post Go-live

- Outcomes analysis, metrics, and continuous improvement



Toward Complete & Sustainable Learning Systems Charles P. Friedman, PhD Josiah Macy, Jr. Professor Chair, Department of Learning Health Sciences Professor of Information and Public Health University of Michigan. December 8, 2014
https://medicine.umich.edu/sites/default/files/2014_12_08-Friedman-IOM%20HS.pdf

Deployment: Process, Process, Process



- **Organizational and Clinician Support**
 - Must have a clinician champion
 - Support from impacted user group, align with strategy
 - Who is going to do data entry, view, act on Clinical Decision Support tool?
 - Check with your EHR vendor/community to see if a tool already exists
- **IT Change Management**
 - Involve IT early and often (clinical informatics or optimization analysts)
 - Governance and approvals
 - Testing
 - Workflow documentation, training, and materials
 - Implementation

CDS Exemplars



- CDS Hooks
 - A protocol that connects EHRs to clinical decision support services
- SMART on FHIR
 - An ecosystem of apps that can be plugged into EHR systems
 - A set of open specifications to integrate apps with EHRs, portals, Health Information Exchanges, and other Health IT systems
 - Helps make structured healthcare data available to 3rd-party applications
- CDS Connect
 - AHRQ-supported online web presence that functions as a repository of CDS artifacts and create prototype infrastructure for sharing CDS across health care settings & technologies
 - A project to demonstrate how evidence-based care can be more rapidly incorporated into clinical practice through interoperable decision support
 - MITRE corporation; <https://cds.ahrq.gov/cdsconnect>

Thank you!

Email us at:

- Steve Johnson joh06288@umn.edu
- Tamara Winden twinden@kumc.edu
- Lisiane Pruinelli pruin001@umn.edu



This workshop is partially funded by the University of Minnesota, School of Nursing Foundation Grant