



PHPM 672 Data Science for HSR  
PHPM 677 Data Science in Public Health

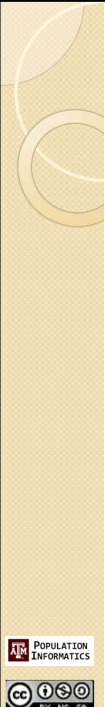
Hye-Chung Kum  
Population Informatics Research Group  
<http://pinformatics.org/>

**License:**  
Data Science in the Health Domain by Hye-Chung Kum is licensed under a  
[Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/)

**Course URL:**  
<http://pinformatics.tamhsc.edu/phpm672>



 POPULATION INFORMATICS  


1



## Who am I ?

- PhD in Computer Science
  - Datamining
  - KDD (Knowledge discovery and datamining)
- MSW: Policy & Management Track
  - Certificate of nonprofit management
- UNC-CH
  - School of Social Work, UNC-CH
  - Department of Computer Science, UNC-CH
- TAMU
  - Department of Health Policy and Management, School of Public Health
  - Department of Computer Science and Engineering
  - Department of Industrial and Systems Engineering
  - The Center for Remote Health Technologies and Systems (CRHTS)
- Teaching
  - I love teaching. I put a lot into it & I expect a lot from students
  - Slides are my personal notes so I wont forget (don't use as example of good presentations)
- Questions?

 POPULATION INFORMATICS  


2

## Who are you ?

- Program
  - PhD in HPM ? (MUST be 672)
  - MPH? (Should be 677)
  - Anyone else? (Not HSC?)
- Majors before
- Experience in programming & statistics
  - Class in statistics
  - Have used STATA, R, SQL
  - Have used SAS
  - Have any programming experience
- What would you like to get from this class



3

## What is this class about

- Website
  - <http://pinformatics.org/phpm672>
  - Very Important to check regularly!
  - Some links, not there yet. Will add as we go.
- E-Campus (limited)
  - Submit homework
  - Post grades
  - Midterm
- Syllabus (review together)
- Schedule
- Resource on pinformatics website
  - <http://pinformatics.org>



4

## What is this class REALLY about

- How to think critically, logically about data
- Think “tables”(=columns & rows)
- Communicate with your computer
  - Tell it what you want it to do (with the data)
  - Listen when the computer is lost
  - Fix your instructions so the computer wont be lost
    - The computer is ALWAYS right
    - ... just like a 4 year old ☺



5

## One Previous Student

- “PHPM 677 is probably one of the most valuable classes I have taken in my graduate coursework. I would recommend it to any public health student (in ANY discipline) who wants to learn how to think about and use data. I came into this class wanting to learn SAS and came out beginning to think like a data scientist.”



6

## Background: direct quote

- “Before taking this class,
  - I had **limited experience with SAS and data science.**
  - I knew **how to copy and paste basic SAS commands,**
  - **but I didn’t really understand it, much less how to think about data.**
- That changed throughout the 15 weeks of this class.
- Fair warning:
  - **you will probably get very frustrated and discouraged along the way.**
  - You will **spend hours working on weekly assignments and still feel lost and behind.**
  - There’s **definitely a steep learning curve.**
- But at some point in the semester, **it finally begins to “click.”**
  - You **begin to understand how to troubleshoot**
  - **and find new ways of thinking about problems.**
  - After a while, **you even start to feel confident in your ability to transform and analyze a dataset.**
- I still have a lot to learn and improve upon,
  - but I think now I have **a foundation to build upon throughout the rest of my career.”**

AIM POPULATION INFORMATICS  
CC BY-NC-SA

7

## WARNING:

### Be Prepared to Work Hard

- Previous Students: “What to Expect”
- This was the only part of the class that will go at this nice pace
- There is a LOT of materials for me to cover
- You will not have another dedicated time to learn this, but programming will hit you as you work on your dissertation, when you are looking for a job, and on the job.
- So I want to teach you as much as I can this semester.
- If you feel you are lost, please come talk to me. I will try to help. If majority of you come talk to me, I will slow down. If you don’t give me input, I will assume the pace is fine.
- **You will get AS MUCH AS you put into this class**

AIM POPULATION INFORMATICS  
CC BY-NC-SA

8

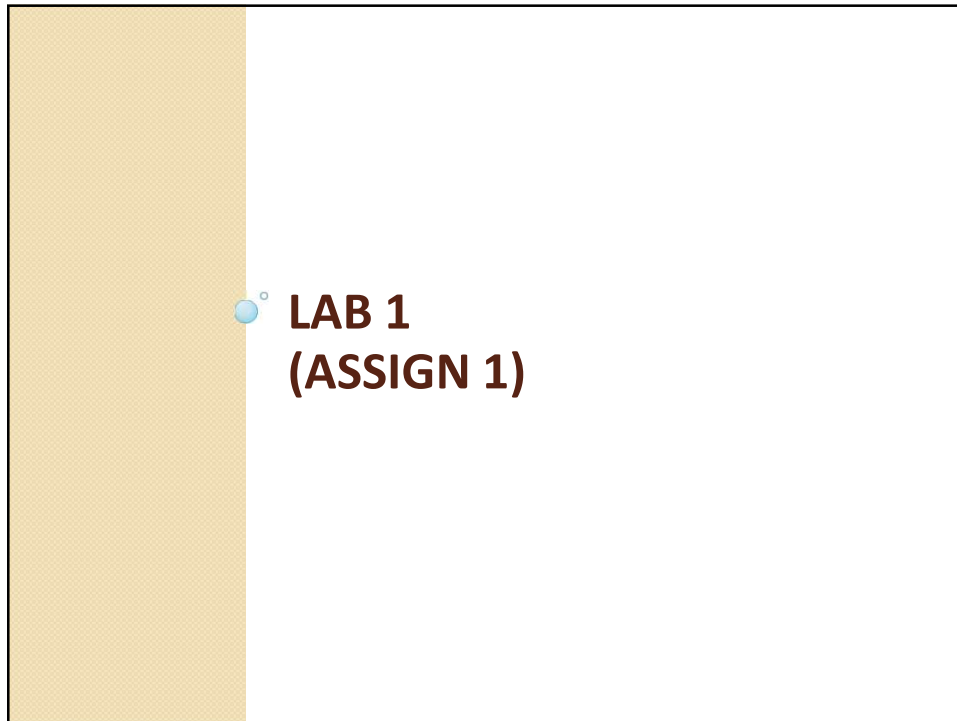


9

## Last thoughts

- Very wide background
  - Try to spend time moving up from where you are
  - If you work hard, you should be able to get the basics in 15 weeks
- Programming
  - Bottom line, you have to DO this.
  - READ (lecture), WATCH (lab), DO (assignment).
  - Not easy, but really worth it to take the time to learn. Like your multiplication tables.
- Data Science
  - Very new. I didn't read any textbooks, no one taught me
  - So mostly my opinion on an evolving topic based on many things I have experienced.
  - Share your thoughts. Younger generation born into the digital world have something I don't.
- Data & Programming CAN be FUN !!
  - It's my favorite hobby

10



11

A slide with a light beige background and a vertical beige bar on the left. The word "Agenda" is written in a large, dark brown font. Below it is a bulleted list of topics. At the bottom left, there are logos for "AIM POPULATION INFORMATICS" and a Creative Commons license "CC BY-NC-SA".

## Agenda



- What is Data Science/Population Informatics?
  - How does it relate to HPM? HSR?
  - How does it relate to Public Health?
- Examples of population informatics
  - NC-DHHS (Dept. Health & Human Svc) Management Assistance Project
  - County self evaluation
  - Research

AIM POPULATION INFORMATICS  
CC BY-NC-SA

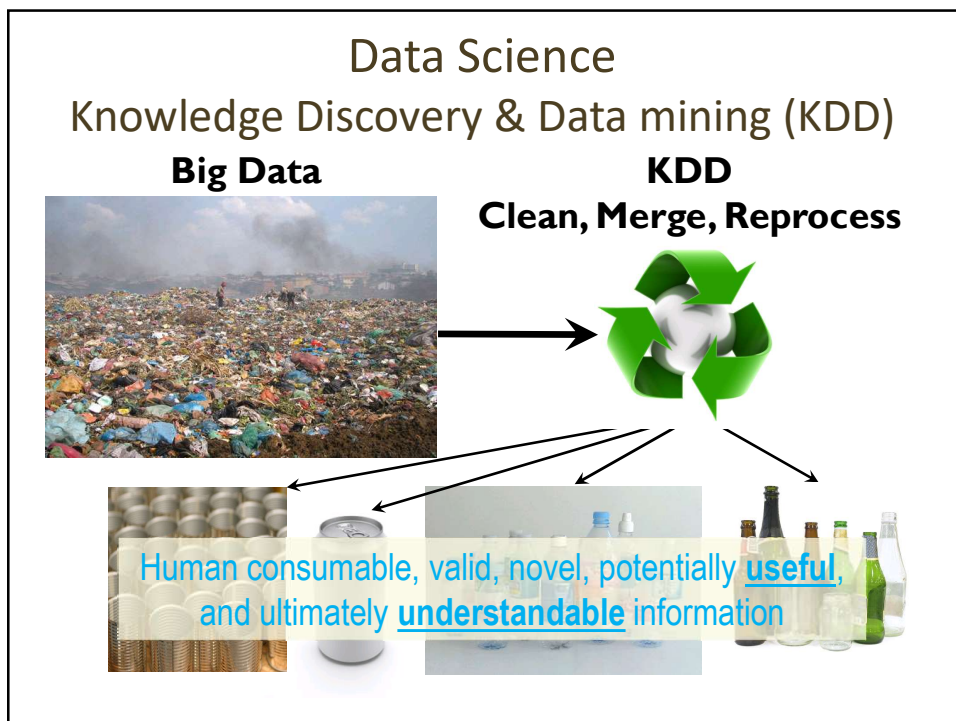
12

# Agenda

- What is Data Science/Population Informatics?
  - How does it relate to HPM? HSR?
  - How does it relate to Public Health?
- Examples of population informatics
  - NC-DHHS (Dept. Health & Human Svc) Management Assistance Project
  - County self evaluation
  - Research

13




14

## Properties of BIG DATA : 4V

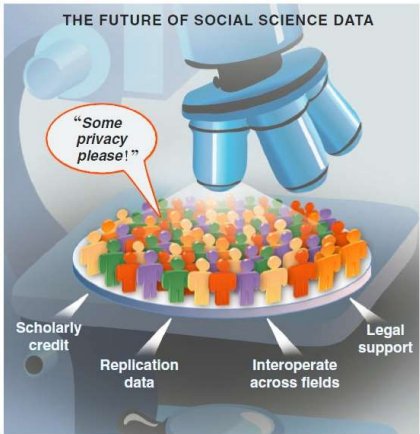
- Volume : constantly generating
- Velocity : constantly changing
- Variety : expressed in many ways
- Veracity : lots of errors

**What do you do to find information/knowledge on the Internet?**




15

## Population Informatics (Massive secondary data analysis)



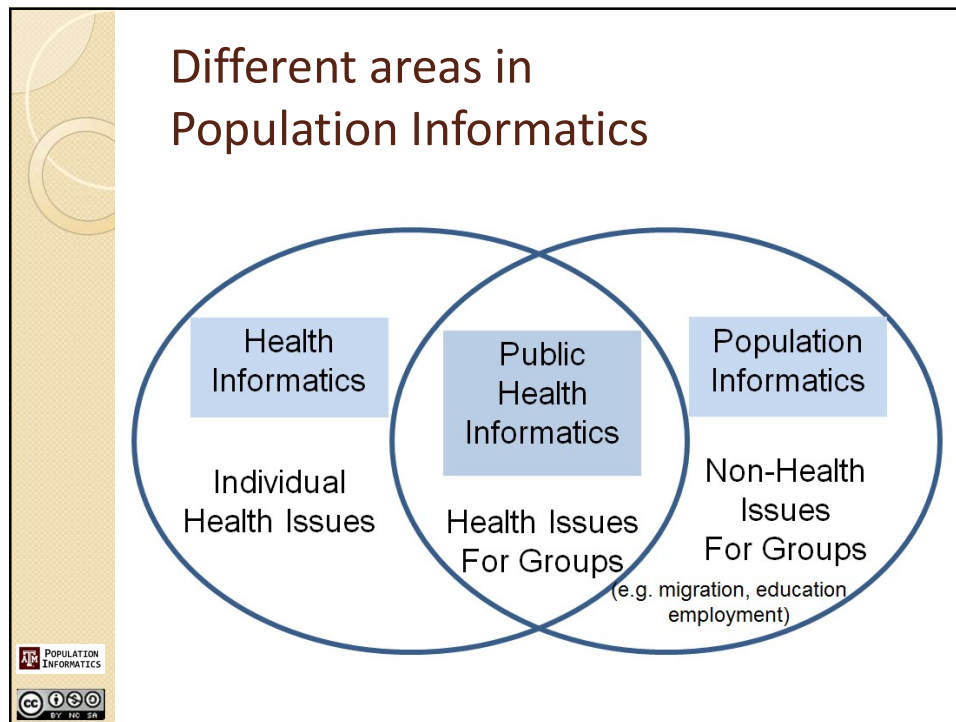
- CS+Statistics+Social Science
- Big data analysis about people
- **Health Population Informatics: Analyzing Big Data about People for Better Healthcare**
- **E-government: Analyzing Big Administrative Data about People to better manage government resources**
- Gary King. Ensuring the Data-Rich Future of the Social Sciences, *Science*, vol 331, 2011, pp 719-721.

**Fig. 1.** New types of research data about human behavior and society pose many opportunities if crucial infrastructural challenges are tackled.

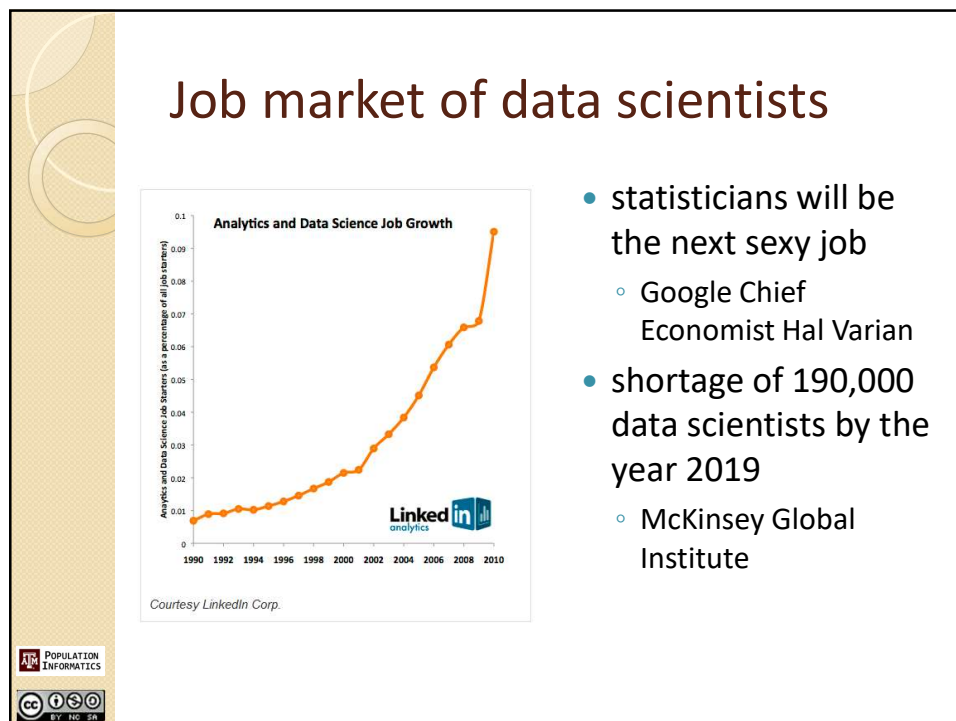


16

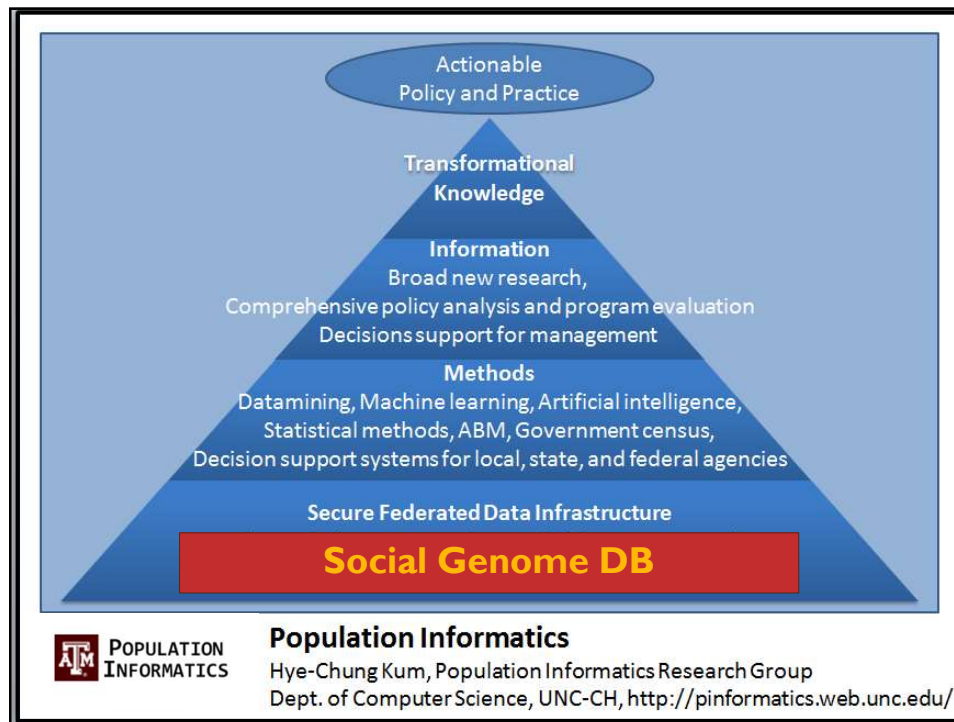




17



18



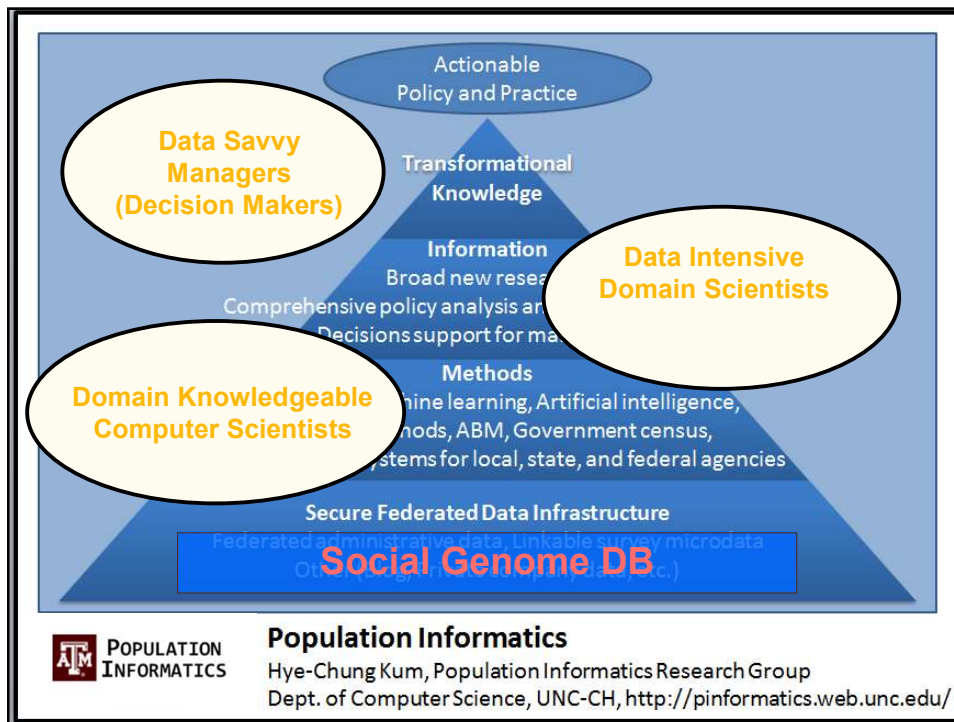
19

## Data experts in the next century

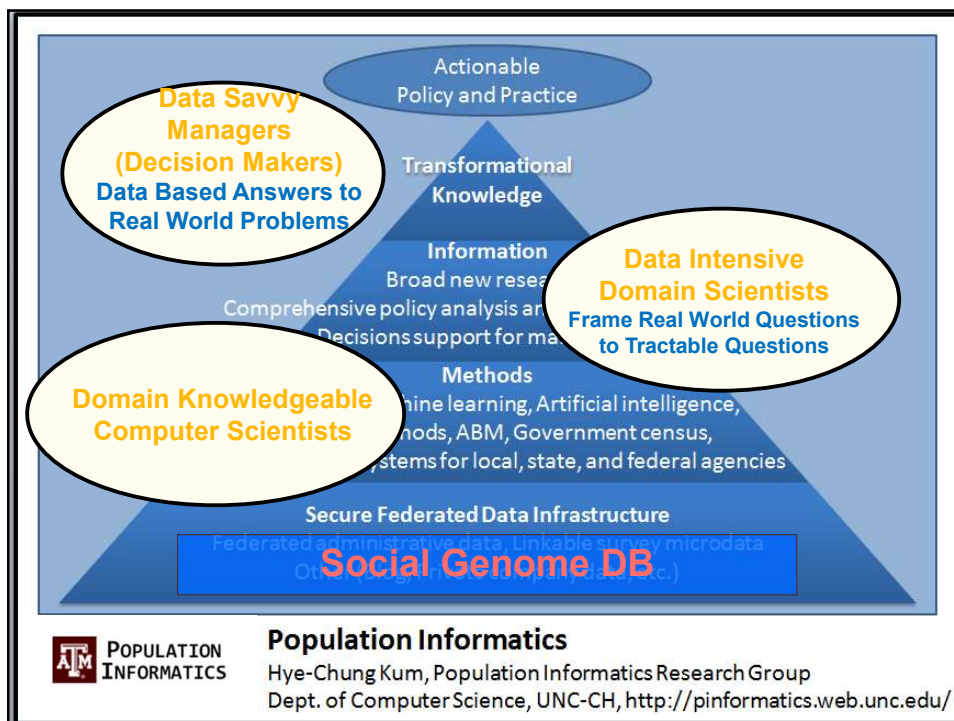
- **Data Savvy Mangers (decision makers)**
  - **MHA & MPH**
  - who can understand and use data for decisions and actions.
- **Data intensive domain scientists**
  - **PhD, MSPH, MPH**
  - experts in the domain science with intensive training on data science and analytics
- **Domain knowledgeable computer scientists**
  - Collaborators in CS (undergraduates, MS, PhD).
  - **HPM department role : teach them the domain science!**
  - Build tools, manage data, and run analytics

**POPULATION INFORMATICS**  
CC BY-NC-SA

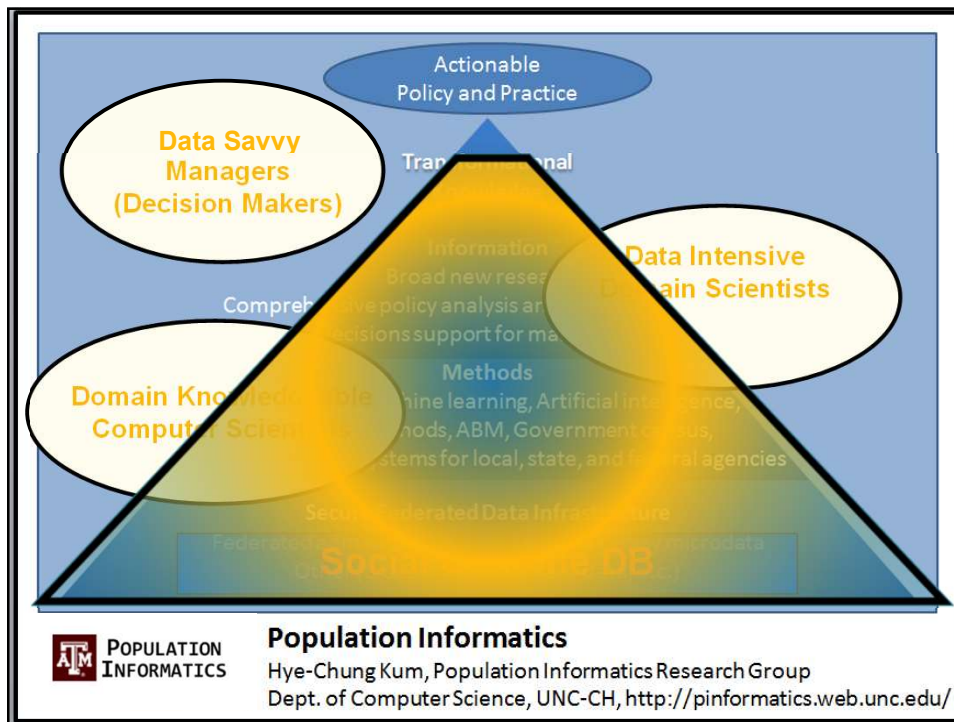
20



21



22



23

## Thomas Davenport

### *Competing on Analytics*

- Skill set for good data scientists
  - IT & Programming skills
  - Statistical skills
  - Business skills:
    - Understand pros/cons of decisions & actions
    - Communication skills
    - Excel / PowerPoint
  - Intense curiosity: the most important skill or trait. “a desire to go beyond the surface of a problem, find the question at its heart, and distill them into a very clear set of hypothesis that can be tested”

POPULATION INFORMATICS

CC BY-NC-SA

24

## New Era in Science : Big Data Science

- **Data** is the new raw material of business: an economic **input almost on par with capital and labor.**(Microsoft's Craig Mundie)
- **Those who can harness the power of data will lead the next century** and drive innovation in commerce, scientific discovery, healthcare, finance, energy, government, and countless other fields.
- Students who learn to be a data science will be in high demand.

AIM POPULATION INFORMATICS  
CC BY NC SA

25

## High International Interest Doing Good Research with Big Data

- Figuring how to do good research with Big Data
- EUDAT (Oct 2011): EU
  - €16-million initiative to develop an international data management infrastructure over 3 years
- White house (Mar 2012) : US
  - a national effort to fund Big Data research across the federal agencies including NSF, NIH, DOD (Dept of Defense), and DOE (Dept of Energy)





AIM POPULATION INFORMATICS  
CC BY NC SA

26

## International Population Health Informatics Research



- US : LEHD (Census Bureau) – 2010 Nobel Prize in economics
- Australia & New Zealand
  - National Centre for Epidemiology and Population Health (NCEPH), The Australian National University
  - Australian Institute of Health and Welfare
  - Centre for Health Record Linkage
  - Centre for the Study of Assessment and Prioritisation in Health, School of Medicine and Health Science (NZ)
- EU
  - Health Information Research Unit, School of Medicine, Swansea University, Wales, UK
  - Health Services Research Unit, University of Aberdeen, Scotland
- Canada
  - Canadian Institute for Health Information
  - Child and Youth Data Lab, Alberta Centre for Child, Family and Community Research




27

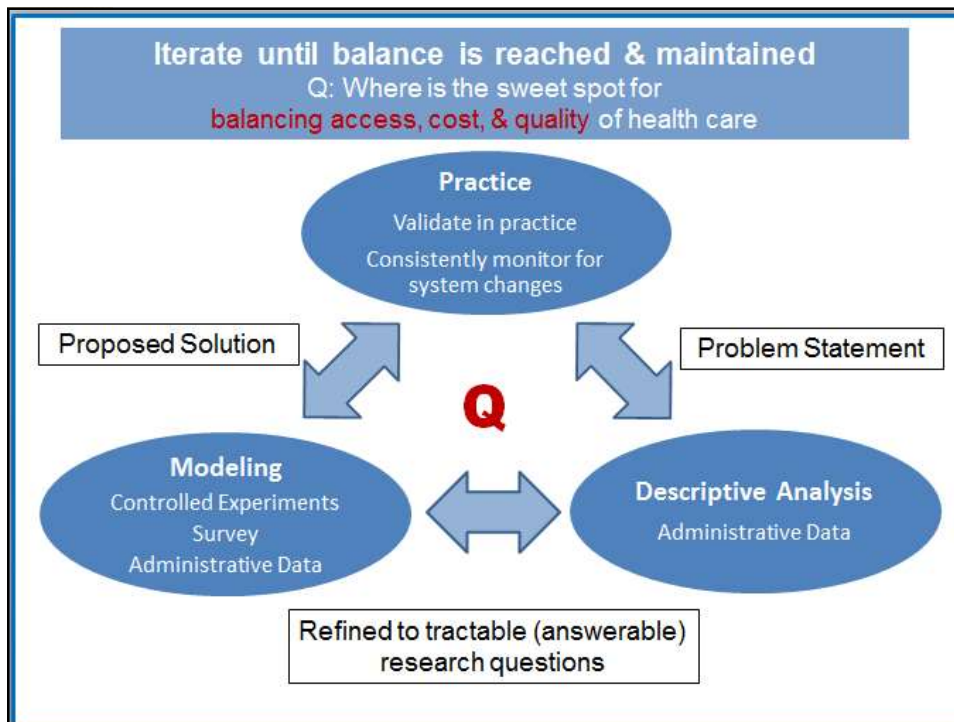
## LEHD : US Census Bureau

- Vertically integrated in one domain
  - Wage : UI (Unemployment Insurance) Data
- Decision support : LEHD website
- By building an integrated data that **“permits the real world of the US economy to be interrogated by the models of unemployment dynamics”** Peter Diamond, Dale Mortense, and Christopher Pissarides shared the Nobel Prize in economics last year (David Warsh, economicprinciple.com)

28





29



30