

# It Takes a Village to Raise a Prospect: Data Mining in a Large Institution



Steven Sorensen, Senior Prospect  
Research Analyst, University of Alberta

SESSION DATE: October 14, 2016

SESSION TIME: 9:00 AM – 10:15 AM

## ***Overview of Data Mining***

**Data Mining** – the process of collecting, searching through and analyzing a large amount of data from a database in order to discover patterns or relationships

General Process of Data Mining:

*Step 1: Picking Variables*

- **Variables** – attributes or data elements that can undergo change
- A variable may be quantitative (e.g., giving), categorical (e.g., email address) or a combination of both (e.g., the number of children)



# *Overview of Data Mining*

## General Process of Data Mining:

### *Step 1: Picking Variables – Continued*

- In order to examine patterns or relationships within your database, select a variable (“predictor variable”) that you wish to test its correlation with the outcome of another variable (“outcome variable”) – for example, testing the relationship of an alum’s primary degree graduation year (“predictor variable”) with his/her giving total (“outcome variable”)
  
- Sometimes the presence or absence of data can be a predictor of an outcome



# Overview of Data Mining

## General Process of Data Mining:

### *Step 2: Drawing a Sample*

- **Sampling** – the process of selecting a representative part of a population for the purpose of determining parameters or characteristics of the whole population



"Here's a list of 100,000 warehouses full of data. I'd like you to condense them down to one meaningful warehouse."



# Overview of Data Mining

## General Process of Data Mining:

### *Step 2: Drawing a Sample – Continued*

- Basic types of sampling:
  - a) Random Sampling** – selecting a specific subset of records drawn from a larger population in such a way that no one record in the population has any more chance of being drawn than any other record
  - b) Systematic Sampling** – selecting a subset of records according to a random starting point and a fixed periodic interval. The number of records in the population is divided by the number of records to be drawn for the sample. The quotient is called  $k$ , and every  $k^{th}$  record is drawn from the sample



# ***Overview of Data Mining***

## General Process of Data Mining:

### *Step 2: Drawing a Sample - Continued*

- The larger your sample size, the more accurate your results will be
- Split your sample in half in order to cross-check the results
- Divide your variables into segments – for example, you can report your giving amounts like \$0-\$100, \$101-\$1,000, and so on, and graduation years like 1930s, 1940s, and so on



# ***Overview of Data Mining***

## General Process of Data Mining:

### *Step 3: Looking for Promising Predictors*

- **Mean** or **Average** – the sum of the values of a variable in a sample divided by the number of records in the sample
- **Median** – mid-point of the value of a variable within the sample (50% of the sample records fall below and 50% of sample records fall above)



# *Overview of Data Mining*

## General Process of Data Mining:

### *Step 3: Looking for Promising Predictors - Continued*

- The mean of a sample is sensitive to extreme values (for instance, a significantly large donation). In order to get a sense if a variable is a reliable predictor of another variable's outcome, it is recommended to compare the sample's mean with its median. If the mean and median amounts are close, the predictor variable is likely good

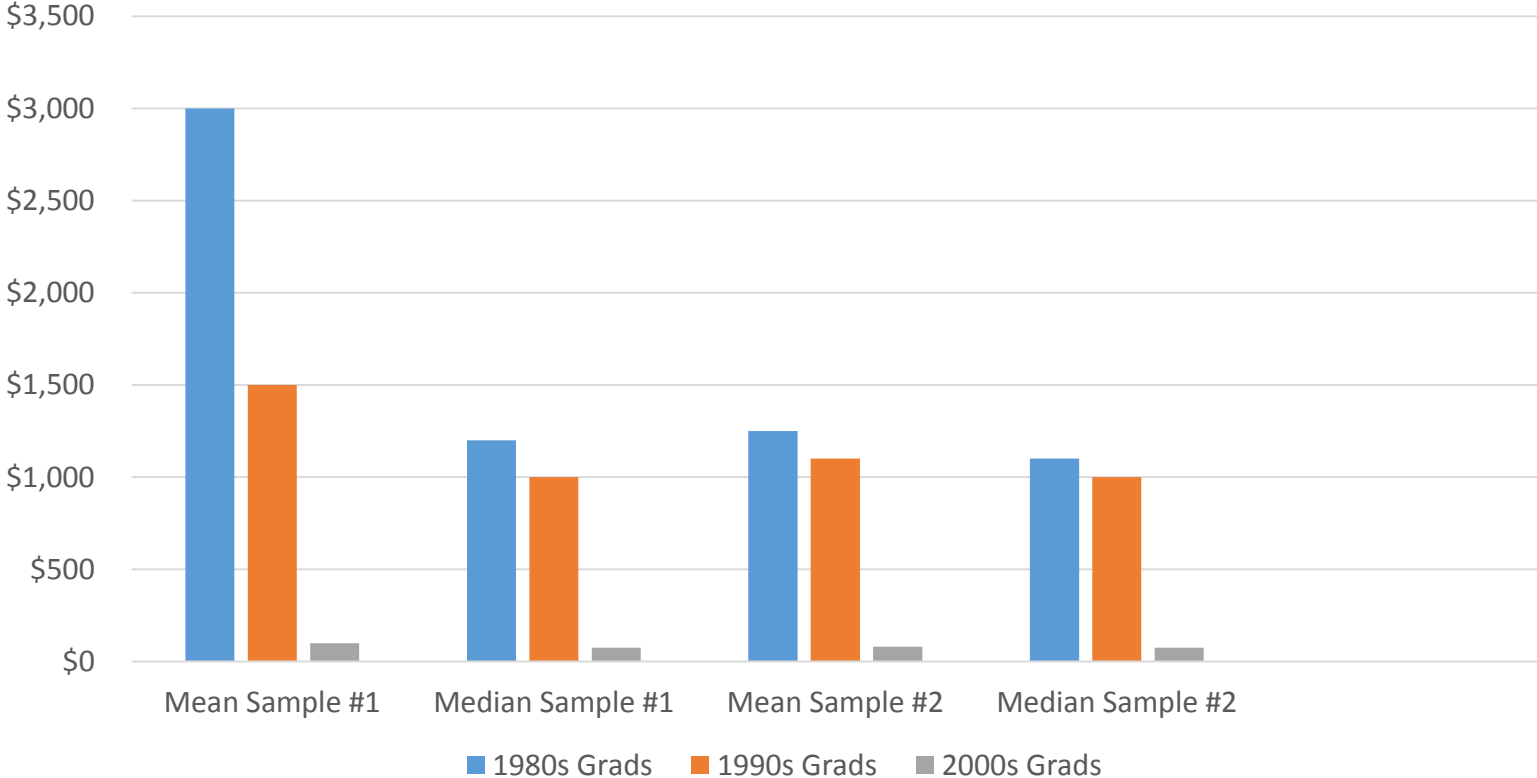




# Overview of Data Mining

## General Process of Data Mining:

Mean and Median Total Giving by Primary Degree Graduation Year



# *Overview of Data Mining*

## General Process of Data Mining:

### *Step 4: Building a Scoring System*

- **Binary Variable (aka 1/0 Variable; Dummy Variable)** – provides a way of using a categorical predictor variable in various kinds of estimation models – basically, the variable takes the value of 0 or 1 to indicate the absence or presence of some categorical effect [e.g., past giving = 1 (higher likelihood to give), no giving = 0 (lower likelihood to give)]
- Scoring should incorporate weighting to account for the fact that some predictors are better than others



# ***Overview of Data Mining***

## General Process of Data Mining:

### *Step 4: Building a Scoring System - Continued*

- The score should be simple, concise and understandable
  - An example of a gift rating would be a number/letter combination representing an entity's capacity to give (represented by the number) and inclination to give (represented by the letter)

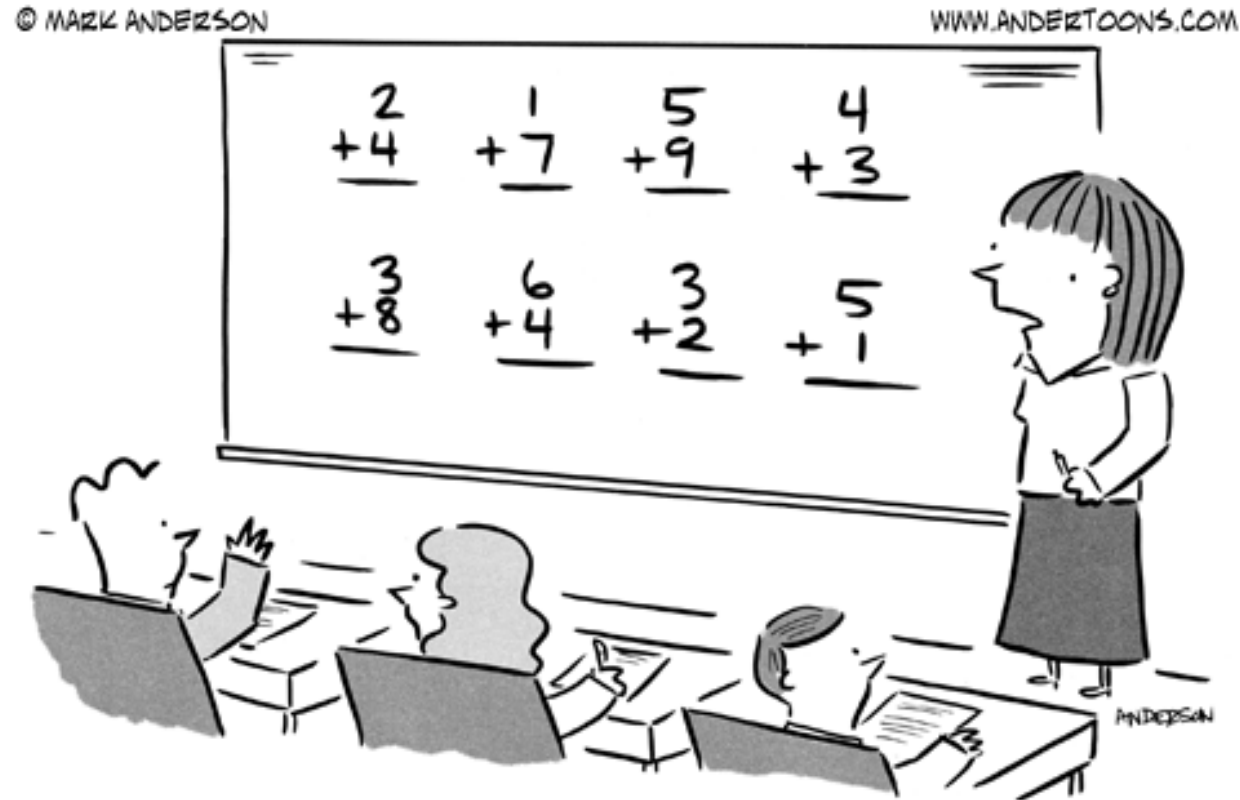
**Recommended Reading:** *Data Mining for Fund Raisers: How to Use Simple Statistics to Find Gold in Your Donor Database (Even if You Hate Statistics)* by Peter B. Wylie (2004, The Council for Advancement and Support of Education)



# Essential Elements for Effective Data Mining

In order for data mining to be effective, data elements in your database need to be:

- *Clear*
- *Accurate*
- *Concise*
- *Comprehensive*
- *Locatable*



"Let's solve these first. We can worry about data mining later."

APRA  
CANADA 2016

October 12-14  
Toronto

## ***“Welcome to my Village” – An overview of a large institution***

As of August 2016, the University of Alberta’s central alumni and donor database consisted of:

- 292 active database users with varying access rights from 40 different faculties/departments/units
- 353,455 records of individuals, companies, foundations, associations/organizations, institutions and governing bodies (348,818 active records + 4,637 inactive records)



## *“Welcome to my Village” – An overview of a large institution*

- 12,553 different data fields
- Millions of different data elements

An average of 12,500 new records are added to the database annually



## ***Challenges with data mining within a large institution***

- “Too Many Cooks in the Kitchen” – Numerous database users are likely to enter data into a database in different and/or incorrect ways
- Vastness of database can make comprehending, locating and entering data challenging

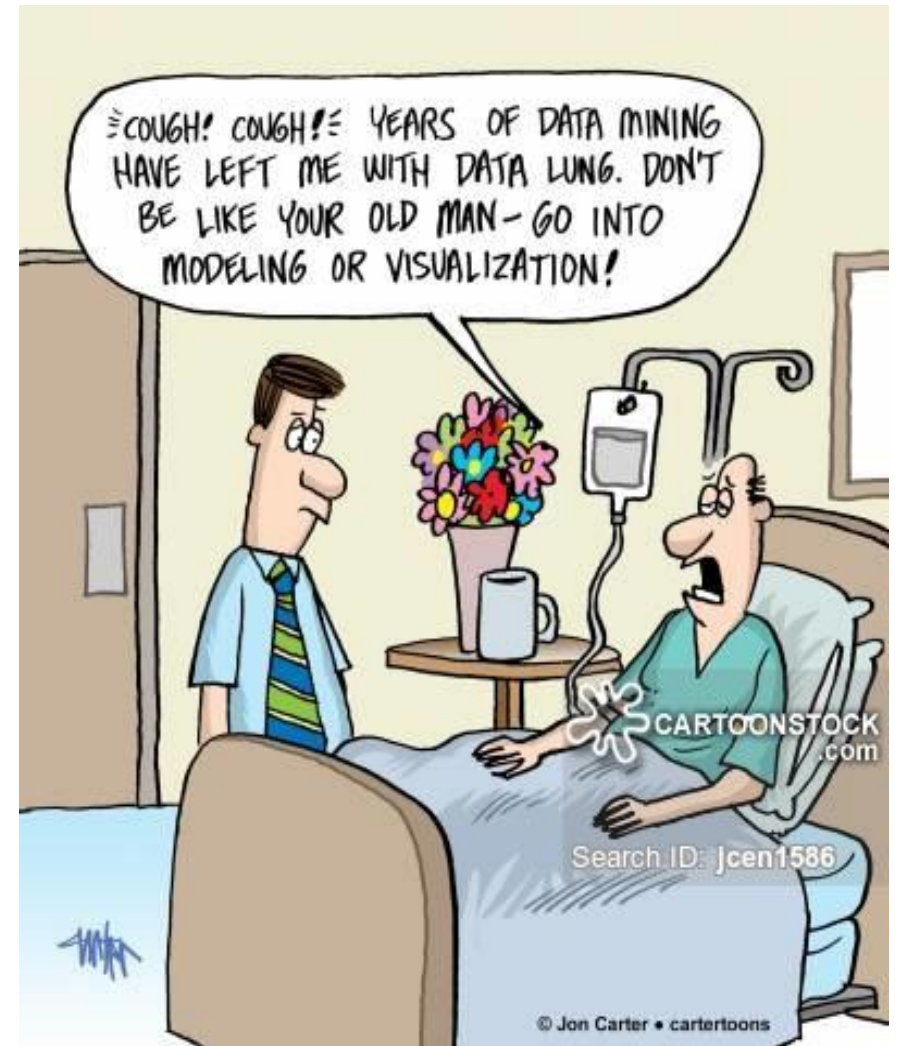


## ***Challenges with data mining within a large institution***

- Rapid and frequent updates and changes to the database can be hard to track
- Large institutions likely have a long history of data retention, which leads to conversion of data from older databases that might not necessarily fit nicely within the current database







## ***Solutions to challenges facing data mining within large institutions***

- Allocate management of specific database fields to a department or select departments
  - Specific department(s) that utilize certain database fields on a regular basis are better equipped to handle rapid changes, recognize issues and avoid mistakes within aforementioned fields
  - Data Audits



## ***Solutions to challenges facing data mining within large institutions***

- Central / Interdepartmental database groups
  - Regular and coordinated examinations of database issues from a broader viewpoint
- Communication
  - Maintain consistent language and terms in database
  - Provide major database updates and data mining strategies at meetings
  - Interdepartmental Intranet postings



## ***Solutions to challenges facing data mining within large institutions***

- Departmental Interaction
  - Attend other departmental meetings to present data mining issues
  - Have other departments make presentations at your department's staff meetings
  - Spend time within other departments to better understand their needs and identify potential database variables



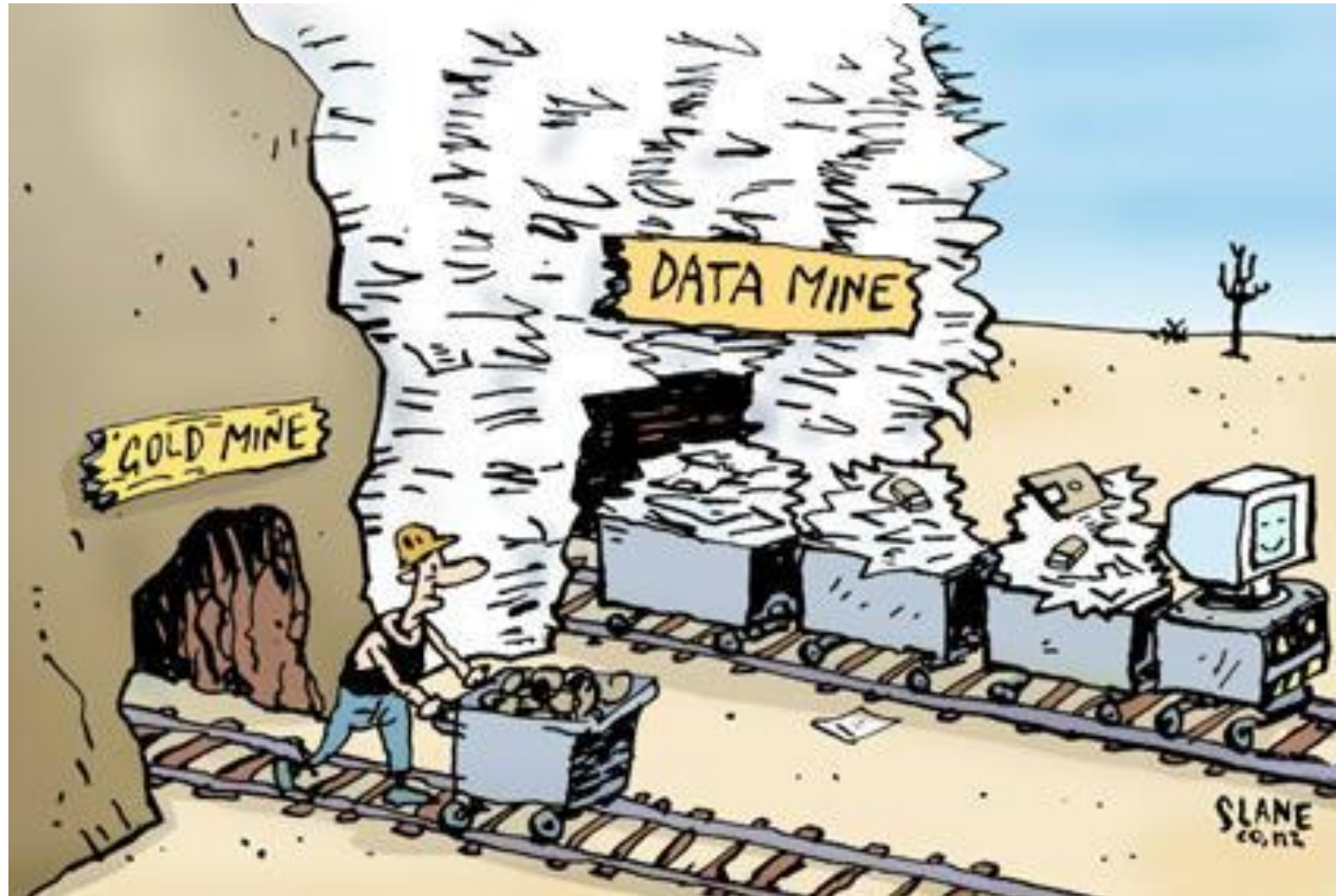
# ***Solutions to challenges facing data mining within large institutions***

- Learning
  - Encourage / attend regular database training sessions
  - Hold departmental introduction and overview sessions
- “Playing”
  - Spend time exploring the database
- Confer with other “villages”
  - Check with other institutions about their data mining methods and issues
  - Refer to books and articles about data mining





Questions / Comments  
E-mail: [steven.sorensen@ualberta.ca](mailto:steven.sorensen@ualberta.ca)



**APRA**  
CANADA 2016

October 12-14  
Toronto