ACTUARIAL SCIENCE VS DATA SCIENCE:

FILLING IN THE BLANKS

CAS ANTITRUST STATEMENT

The Casualty Actuarial Society is committed to adhering strictly to the letter and spirit of the antitrust laws. Seminars conducted under the auspices of the CAS are designed solely to provide a forum for the expression of various points of view on topics described in the programs or agendas for such meetings.

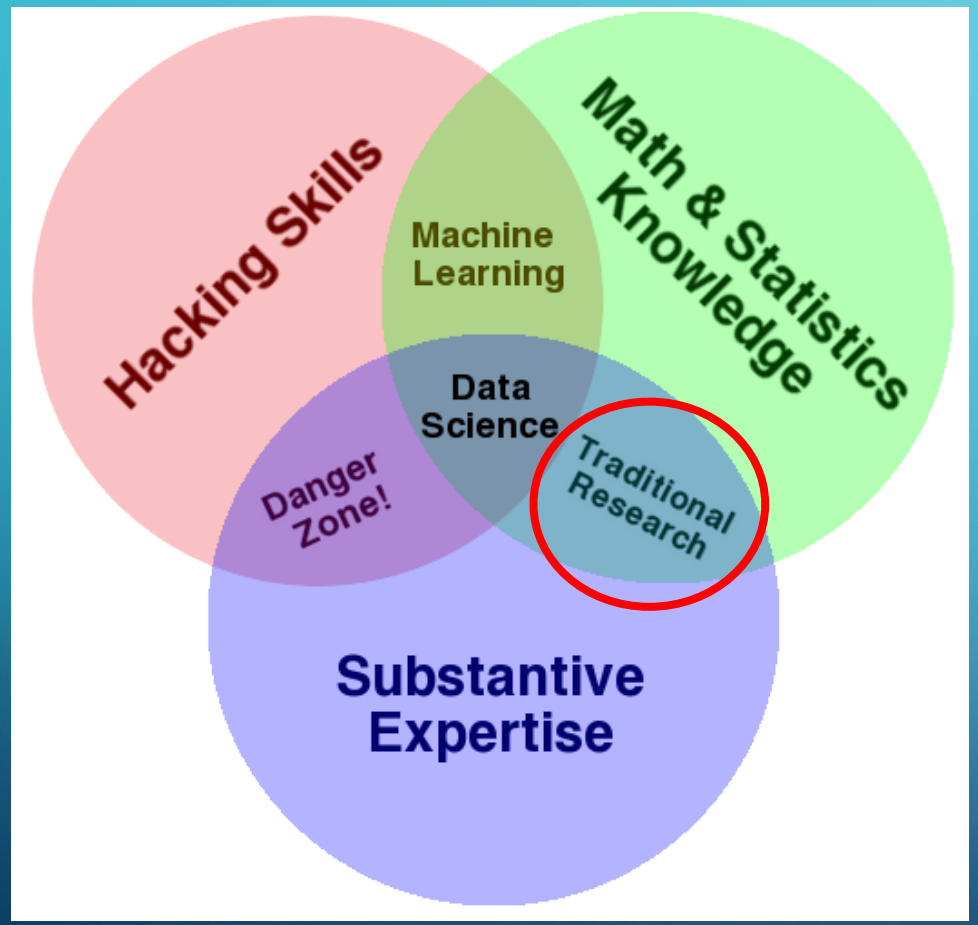
Under no circumstances shall CAS seminars be used as a means for competing companies or firms to reach any understanding – expressed or implied – that restricts competition or in any way impairs the ability of members to exercise independent business judgment regarding matters affecting competition.

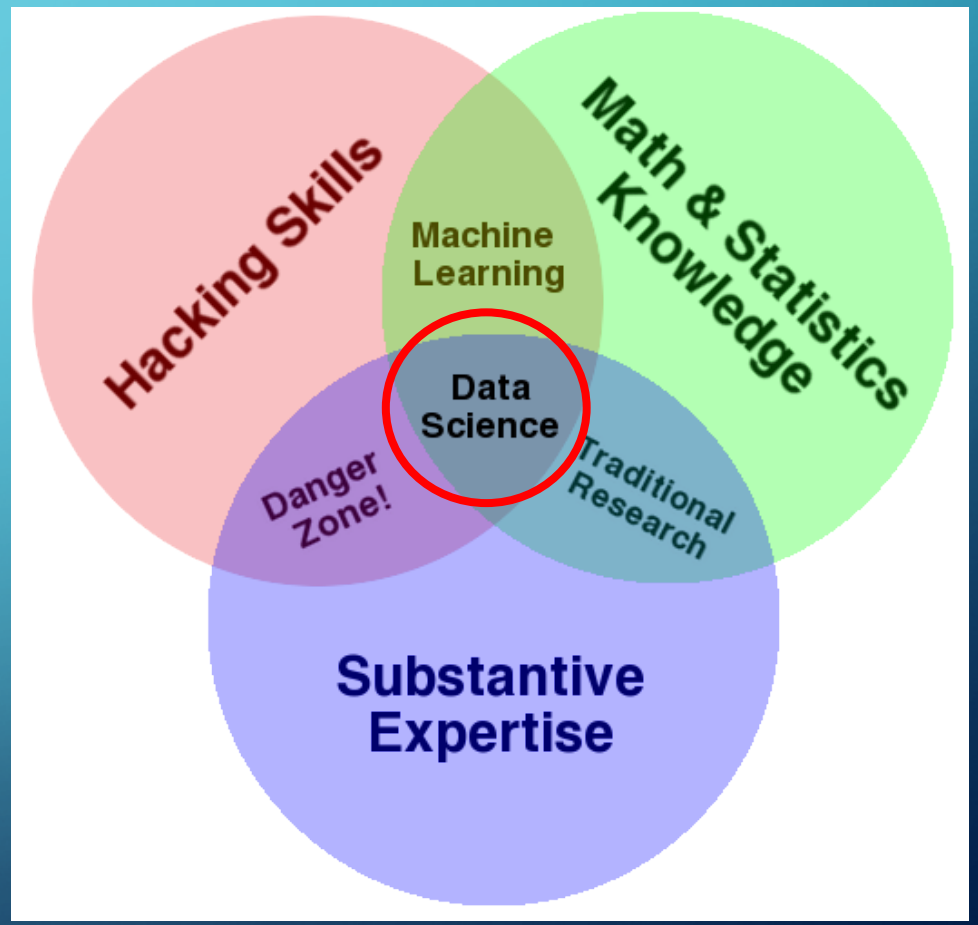
It is the responsibility of all seminar participants to be aware of antitrust regulations, to prevent any written or verbal discussions that appear to violate these laws, and to adhere in every respect to the CAS antitrust compliance policy.

AGENDA

- If I'm an actuary and not a data scientist, what am I missing?
- How much should I care?
- Is it too late?
- What can I do to fill in the blanks?

ACTUARIAL VS DATA SCIENCE?





http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

TWO LEVELS OF KNOWLEDGE TO CONSIDER

LEVEL 1 – PARTNER

- Knowledgeable
- Conversant
- Reading books & papers
- Data & Tech Working Party papers

LEVEL 2 – PRACTITIONER

- Skilled & Experienced
- Capable
- Courses, degrees & certifications
- iCAS

DATA & TECHNOLOGY WORKING PARTY

Mission:

Working party members will research and define the knowledge and skills required for actuaries to *successfully partner* with IT to participate in the Data and Analytics revolution – including:

- Data Quality
- Databases
- Business Intelligence
- Data Science

PARABLE: A TALE OF TWO INSURANCE COMPANIES

Company A

- Actuaries took the lead introducing data warehousing and data mgt
- Actuaries took the lead on hiring predictive modelers
- Actuaries partnered with IT on projects like Big Data and 3PD

Company B

- Actuaries struggled and left data warehousing and data mgt to IT
- Predictive modelers were hired into separate units in other departments
- IT led the introduction of Big Data and the CDO is a technology job

IS IT TOO LATE?

- If my organization is already on the path to Company B?
- If I have no interest in taking more exams?
- If IT has no interest in working with me?

THE D&T WORKING PARTY PAPERS

- Data Quality
- Databases
- Business Intelligence
- Data Science

DATA QUALITY PRINCIPLES

Data quality – data science requires an understanding the data's value for analysis and paleontologist's passionate curiosity

- Validity
- Accuracy
- Completeness
- Timeliness
- Reasonability
- Data Transparency

POLL: DATA GOVERNANCE

1. Does your Organization have a Data Governance Board?
   a. Yes
   b. No
   c. I don't know
2. Do you participate as
   a. A data owner
   b. A data steward
   c. A Board member
   d. I don't participate
3. Does your Organization have a Metadata Repository?
   a. Yes
   b. No
   c. I don't know
4. I have
   a. Used our data glossary
   b. Used our data profiling tools
   c. Contributed metadata content
   d. None of the above

WHAT YOU SHOULD KNOW ABOUT DATA GOVERNANCE

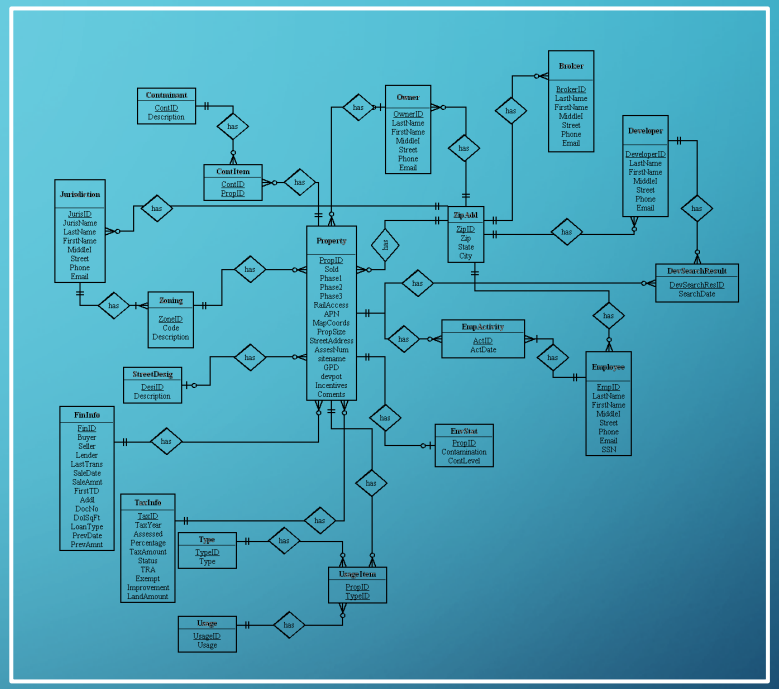
- Data Governance – defining and enforcing policy regarding data quality
- Data Stewardship – Clear ownership and accountability for data quality
- Data Quality – suitability for an intended use including validity, accuracy, completeness, timeliness, reasonability, and traceability
- Metadata – documented definitions, valid values, profiles, and lineage
- Master Data Management (MDM) – the process of reconciling critical data that is shared across the organization (e.g. customer)
-

DATABASES

Want data?

Go get it…

Does an actuary know *how*?



Data scientists understand how the data is stored, updated and transmitted – with this knowledge they know how and where to hack the data they need

POLL: DATABASES

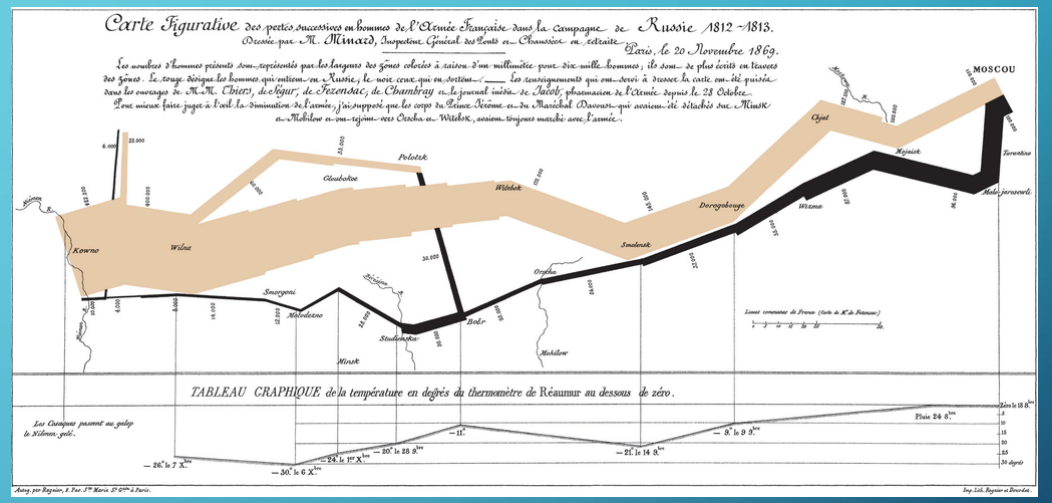
1. Does your Organization have a Data Warehouse?
   a. Yes
   b. No
   c. I don't know
2. Where is the data you use?
   a. In a data exploration environment
   b. In a Business Intelligence environment
   c. In a spreadsheet
   d. I don't use raw data any more
3. How is your data organized for you?
   a. Unstructured using schema on read
   b. In a relational star schema
   c. In a flat and wide extract
   d. In a pivot table
4. How do you get your data?
   a. I write my own SQL queries
   b. I use a self-service BI tool
   c. I have someone pull it for me
   d. I don't use raw data any more

WHAT YOU SHOULD KNOW ABOUT DATABASES

- Application Databases vs Analytic Databases – transactional vs batch
- Database Structures – Flat and Wide vs Relational vs Columnar vs Graph vs NoSQL vs Unstructured
- Structured Query Language (SQL) – The universally accepted standard
- Extract, Transform & Load (ETL) – The process of preparing data for analysis
- Fact Tables – the tables that store the metrics of interest to an organization (e.g. premium)
- Dimension Tables – the tables that describe the context of the facts (e.g. line of business)
- Keys – a field used to join data across tables
- Grain – the level of detail in a database
- Conformed – consistent fact and dimension definitions across tables

BUSINESS INTELLIGENCE

- Conclusions should be easy to understand and compelling
- BI aims to enable fast and easy data access and analysis for everyone
- The data scientist uses BI to tell a story with data
- Knowledge of business intelligence concepts improves insight and communication



The devastation of Napoleon's march on Moscow presented visually

http://www.edwardtufte.com/tufte/minard

POLL: BUSINESS INTELLIGENCE

1. Does your Organization have a Business Intelligence team(s)?
   a. Yes
   b. No
   c. I don't know
2. Do you participate as
   a. A solution developer
   b. A BI designer or architect
   c. A provider of requirements
   d. I don't participate
3. Does your Organization have a mature self-service BI capality?
   a. Yes
   b. No
   c. I don't know
4. The most sophisticated visualization tool I have used is:
   a. Tableau
   b. Excel charts
   c. Something else
   d. None of the above

WHAT YOU SHOULD KNOW ABOUT BUSINESS INTELLIGENCE

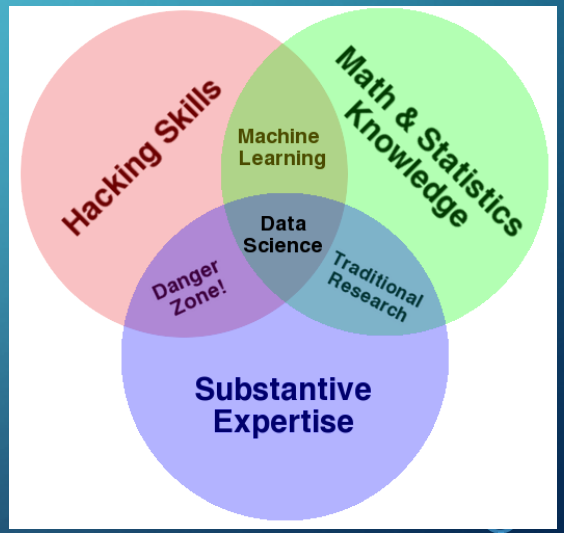
- Business Intelligence – integrated decision support applications and databases
- Applications – Tools used to organize and present the data (e.g. Cognos)
- Databases – Data Warehouses, Data Marts, OLAP Cubes, etc.
- BI Projects – Planning, Analysis, Design, Develop, Test, Implement
- The Actuary's Role – Typically a Subject Matter Expert (SME)
- The "Hybrid Team" – Participate throughout – not just requirements and UAT

DATA SCIENCE

Data Science is…

- methodologies to extract knowledge and insights..
- from data in various forms, either structured or unstructured
- and a continuation of data analysis fields such as data mining, operations research, and predictive analytics..

The depth of domain knowledge and analytic rigor defines the difference between dangerous, misleading hacking and true data



POLL: DATA SCIENCE

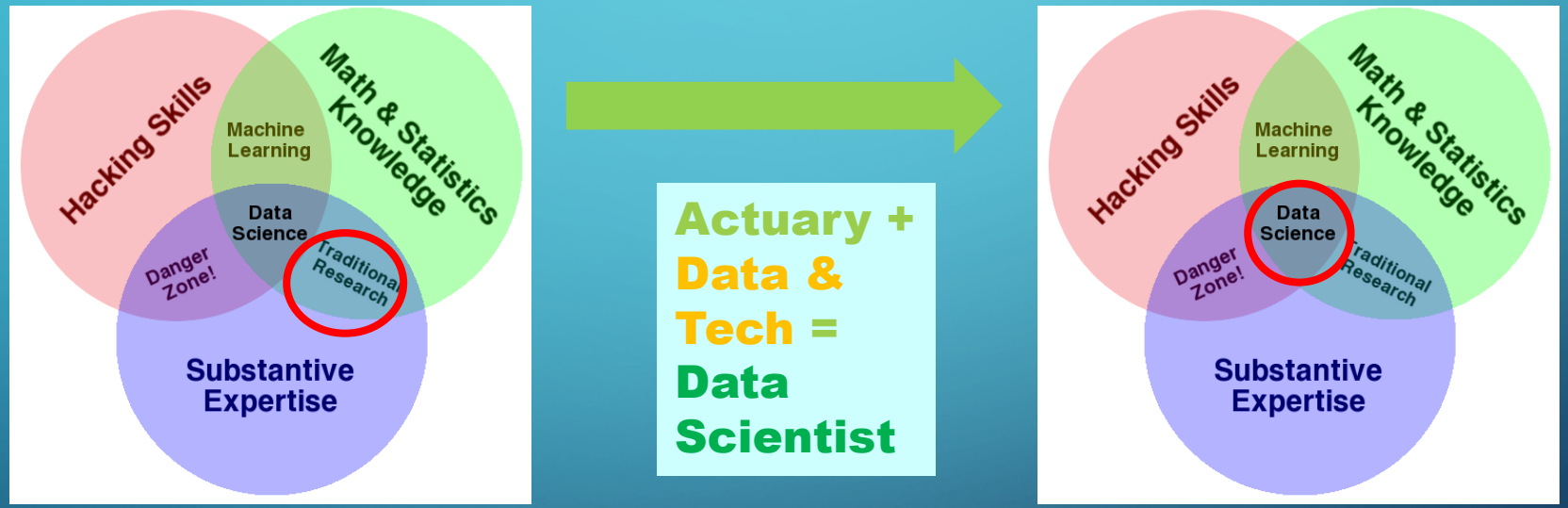
1. Does your Organization have a Data Science Team?
   a. Yes
   b. No
   c. I don't know
2. Do you participate as
   a. A data scientist
   b. An actuarial partner
   c. A customer
   d. I don't participate
3. If you have a Data Science team, does it report to?
   a. Actuarial (e.g. The Chief Actuary)
   b. The business (e.g. the CFO)
   c. IT (e.g. The CDO or CIO)
   d. We don't have a DS team
4. I have
   a. Built a predictive model
   b. Been part of a predictive modeling team
   c. Helped implement a predictive model
   d. None of the above

WHAT YOU SHOULD KNOW ABOUT DATA SCIENCE

- Analytic Complexity – Descriptive, Diagnostic, Predictive & Prescriptive
- Data Science – Joins stats, OR, data mining, machine learning & visualization
- Big Data – Velocity, Variety, Volume, Veracity precludes traditional analysis
- Test and Learn – A/B testing, designed experimentation accelerate insight
- Unstructured Data – Text mining, Telematics, IoT, audio recordings, video
- Best Practices
  - Define the Target Variable & identify the potential Independent variables
  - Prepare the data
  - Train, Test & Validate
  - Implementation & Adoption

ACTUARY AS DATA SCIENTIST

In insurance, the actuary has a tremendous head start in domain knowledge and analytical rigor



The Working Party's papers are aimed at beginning to fill any remaining gaps in data and technology knowledge

WORKING PARTY MEMBERS

- Pete Bothwell, Co-Chair
- Mary Jo Kannon, Co-Chair
- Benjamin Avanzi
- Joe Izzo
- Stephen Knobloch
- Ray Nichols
- James Norris
- Andrea Pan
- Dimitri Semenovich
- Linda Waite
- Dom Yarnell
- Cheri Widowski
- Tracy Spadola
- Michele Wetzel