In Praise of Small Data

Statistical Science and Data Science

Nancy Reid University of Toronto

6 February 2019



Rochester Institute of Technology 2019

Outline

- big data
- data science
- statistical science
- 4 examples
- how do statistical science and data science intersect?
- building collaborations



www.fields.utoronto.ca/programs/scientific/14-15/bigdata

Big Data



Big data: are we making a big mistake?

"Big data" has arrived, but big insights have not

Economist, journalist and broadcaster **Tim Harford** delivered the 2014 *Significance* lecture at the Royal Statistical Society International Conference. In this article, republished from the *Financial Times*, Harford warns us not to forget the statistical Rochester Institute of Technology 2019

Big Data

Figure 1. Hype Cycle for Emerging Technologies, 2018



Plateau will be reached:

Big Data

STATISTICS &

Statistics and Probability Letters 136 (2018) 42-45

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

The role of Statistics in the era of big data

Edited by Laura Sangalli Volume 136, Pages 1-170 (May 2018)

Rochester Institute of Technology 2019



- the data are here to stay
- their size is not the most interesting aspect
- big is in the eye of the beholder
- Literary Digest poll, 1936

2.4 million (Harford 2014)



Big Data



Statistical Inference, Learning and Models in Data Science



9/59

Data Science

- environmental science
- visualization
- social policy
- health policy
- networks
- inference
- optimization
- machine learning
- industry day

Rochester Institute of Technology 2019





September 24 - 27, 2018 at **THE FIELDS INSTITUTE** September 28, 2018 at **MARS**

This is a retrospective workshop for the 2015 thematic program *Statistical Models, Learning and Inference* for Big Data. We will reflect on recent progress and the shift in emphasis to data science in the intervening three years.

INVITED SPEAKERS

Edoardo Airoldi, Havaad University Jimmy Ba, University of Toronto Jelena Bradic, Univenity of Toronto Michael Correll, Takibau Debbie Dupuis, HC Montraal Ruth Etzioni, Tred Huchinon Cancer Research Center Mark Fox, University of Toronto Marzych Chassemi, Mt Laura Haftleid, Havand Meedeal School Simon Lacoste-Julien, University of Montreal Rahul Mazumder, Iutr Staan School Isabel Meirelles, OCAU University Raymond Ng, University of British Columbia Sofia Olhede, University Calego London George Paliouras, III Autors Greg Ridgeway, University of Pennyhennia Veronika Rockova, University of Entago Mark Schmidt, University of Entago Mark Schmidt, University of Entago

10/59

... what is data science?

- · short-hand for "lots of data", "complicated data", "data of uncertain provenance"
- an undergraduate or post-graduate program of training
- a job description
- a new multi-disciplinary field of study



- combining mathematics, statistics, computer science, domain science
- increased emphasis on privacy, fairness, communication, visualization, impact on policy, workflow and reproducible research
 Blake & Olhede, 2016
- data curation and provenance, digital humanities, privacy vs utility, statistical vs computational efficiency, data flows and markets, ... HDSR, Meng, 2019

Rochester Institute of Technology 2019

data acquisition data preservation

Making data trustable and usable Management of data Modelling and Analysis Dissemination and Visualization

Security and privacy

Ethics, policy and social impact

Rochester Institute of Technology 2019

Making data trustable and usable Management of data

Modelling and Analysis Dissemination and Visualization provenance, sampling, cleaning, digitizing size, speed, accessibility

interpretable vs predictive methods accessibility and impact

mathematics statistics computer science domain expertise

Security and privacy

disclosure limitation, anonymization, encryption

Ethics, policy and social impact

fairness and transparency

- an art and a science
- a set of tools and methods, ... [and] computation
- a way of thinking about data
- what statisticians do and how they think about what they do
- interacts vigorously with astronomy, biology, engineering, geology, medicine and public health, and many social sciences, including political science, law, sociology, psychology, anthropology, archeology, history, ... "
- 'A distinguishing feature of the statistics profession, and the methodology it develops, is the focus on a set of cautious principles for drawing scientific conclusions from data.' Lindsay 2004



Statistics Workflow

- start with a scientific question
- assess how data could shed light on this
- plan data collection
- consider of sources of variation and how careful planning can minimize their impact
- develop strategies for data analysis: modelling, computation, methods of analysis
- assess the properties of the methods and their impact on the question at hand
- communicate the results: accurately

but not pessimistically

• visualization strategies, conveyance of uncertainties

NEWS · 28 JUNE 2018

There's no limit to longevity, says study that revives human lifespan debate

Death rates in later life flatten out and suggest there may be no fixed lin human longevity, countering some previous work. LONGEVITY UNLIMITED model based on data from 3,8

Elie Dolgin

"the study included fewer than 100 people who lived to 110 or beyond" "even small inaccuracies in the Italian longevity records could lead to a spurious conclusion" Dolgin. Nature, 28 June 2018

Rochester Institute of Technology 2019

Nature News June 28 2018

A person's chances of dying tend to increase throughout adulthood, but a model based on data from 3,836 people aged 105 or older predicts that this trend flattens out in the very elderly.



HUMAN DEMOGRAPHY

The plateau of human mortality: Demography of longevity pioneers

Elisabetta Barbi¹*, Francesco Lagona², Marco Marsili³, James W. Vaupel^{4,5,6,7}, Kenneth W. Wachter⁸

Theories about biological limits to life span and evolutionary shaping of human longevity depend on facts about mortality at extreme ages, but these facts have remained a matter of debate. Do hazard curves typically level out into high plateaus eventually, as seen in other species, or do exponential increases persist? In this study, we estimated hazard rates from data on all inhabitants of Italy aged 105 and older between 2009 and 2015 (born 1896–1910), a total of 3836 documented cases. We observed level hazard curves, which were essentially constant beyond age 105. Our estimates are free from artifacts of aggregation that limited earlier studies and provide the best evidence to date for the existence of extreme-age mortality plateaus in humans. Science June 29 2018



"We observed level hazard curves, which were essentially constant beyond age 105" "... provide the best evidence to date for the existence of extreme-age mortality plateaus" Rochester Institute of Technology 2019 "This study is unlikely to be the last word on the age-limit dispute, says Haim Cohen, a molecular biologist at Bar-Ilan University in Ramut-Gan Israel 'I'm sure that the debate is going to continue'." Dolgin, Nature, June 2018



FORMAL COMMENT

Plane inclinations: A critique of hypothesis and model choice in Barbi et al

"The capacity for data entry and age inflation errors provides a sufficient model to explain late-life mortality Abstract patterns observed by Barbi and colleagues "

Rochester Institute of Technology 2019

Saul Justin Newman 6*

Research School of Biology, The Australian National University, Acton, ACT, Australia

* saul newman@anu.edu.au

This study highlights how the mortality plateau in Barbi and colleagues can be generated by low-frequency, randomly distributed age-misreporting errors, Furthermore, sensitivity of the late-life mortality plateau in Barbi and colleagues to the particular age range selected for regression is illustrated. Collectively, the simulation of age-misreporting errors in late-life human mortality data and a less-specific model choice than that of Barbi and colleagues



FORMAL COMMENT

Hypothetical errors and plateaus: A response to Newman

"... claims of Barbi and colleagues Kenneth W. Wachter@* rest on nearly 4,000 carefully validated cases from an established registration system. A critique like Newman's, ... can hardly carry force."

Rochester Institute of Technology 2019

Department of Demography, University of California, Berkeley, California, United States of America

* wachter@demog.berkelev.edu

Abstract

Newman questions recent claims about a plateau in mortality rates for Italians beyond age 105 on the basis of a hypothetical model. His model implies implausibly high error rates for / 59 extreme ages. For individuals over 110, for whom birth certificates have been collected, the

- claims that age-misreporting can generate spurious late life plateaus
- Barbi et al (2018) fit a parametric model and used likelihood ratio test to compare to a constant hazard for age > 105
- Newman argued that a modelling choice they made influenced their results
- "of the 861 ... combinations tested, the model selected by Barbi et al generated the single largest late-life mortality plateau"
- statistics: Gompertz model, LRT, power analysis

$$h(x) = ae^{bx}e^{\beta_1 C + \beta_2 M}$$

• data science: 861 such fits, plus simulated errors

with probabilities ranging from 10^{-3} to 10^{-6}

- domain: all inhabitants of Italy aged \geq 105 years 2009–2015 (3836 cases) + Human Mortality Database

... is there a limit to human longevity?

Barbi et al 2018



Fig. 1. Yearly hazards on a logarithmic scale for the cohort of Italian women born in 1904. Confidence intervals were derived from Human Mod Database (HMD) data for ages up to 105 and from ISTAT data beyond age 105. (A) Closeup with 95% confidence intervals based solely on single-cohort (B) Broad view with estimated plateau beyond age 105 (black dashed line) and 95% confidence bands (orange) predicted from the model parameters base full ISTAT database, along with a straight-line prediction (black) from fitting a Gompertz model to ages 65 to 80.
 23/59

- looking for extremes big data becomes small data
- complex dependencies space, time, networks, images, ...
- combining data at multiple scales, or of multiple types

NEWS · 28 JUNE 2018

There's no limit to longevity, says study that revives human lifespan debate

Death rates in later life flatten out and suggest there may be no fixed limit on Rochesthuman longevity;countering some previous work.

In praise of small data: Example 2



Figure 1 | A glacier at Mount Robson Provincial Park, British Columbia, Canada. An analysis by Schildgen and colleagues' confirms that the rate of mountain erosion by glaciers has increased during the past few million years in certain places (such as in British Columbia) in response to climate cooling. but casts doubt on the idea that this was global effect.

"Mountain erosion is thought to have sped up globally ... A reassessment of the data suggests that this acceleration was limited..."

Kirby, Nature July 2018

EARTH SCIENCE

Global erosion by glaciers revisited

Mountain erosion is thought to have sped up globally over the past few million years as the climate cooled and glaciers grew. A reassessment of the data suggests that this acceleration was limited to just a few regions. SEE LETTER P.89

Rochester Institute of Technology 2019

LETTER

https://doi.org/10.1038/s41586-018-0260-6

Spatial correlation bias in late-Cenozoic erosion histories derived from thermochronology

Taylor F. Schildgen^{1,2,6}*, Pieter A. van der Beek^{3,6}, Hugh D. Sinclair⁴ & Rasmus C. Thiede^{2,5}





... Global erosion by glaciers revisited



Rochester Institute of Technology 2019

Fig. 2 | Constructing erosion histories from multiple samples or multiple thermochronometers. The lower figure illustrates thermochronological data from a tilted normal-fault-bounded block and

... Global erosion by glaciers revisited



Rochester Institute of Technology 2019

³ig. 3 | Thermochronology data and modelled erosion-rate changes or the western European Alps. a. Thermochronology data are derived rom 52 different sources compiled by refs^{6,25}. Equivalent colours for hermochronometers correspond to equivalent one-dimensional steady-

Statistics and data science



Figure 1 | A glacier at Mount Robson Provincial Park, British Columbia, Canada. An analysis by Schildgen and colleagues' confirms that the rate of mountain erosion by glaciers has increased during the past few million years in certain places (such as in British Columbia) in response to climate cooling. but casts doubt on the idea that this was a global effect.

"... as we use increasingly sophisticated analyses of **'big data'** to gain insight into global trends in geology, we must not lose sight of the physical processes that operate locally"

Kirby, Nature July 2018

EARTH SCIENCE

Global erosion by glaciers revisited

Mountain erosion is thought to have sped up globally over the past few million years as the climate cooled and glaciers grew. A reassessment of the data suggests that this acceleration was limited to just a few regions. SEE LETTER P.89

Statistical and data science challenges

- looking for extremes big data becomes small data
- complex dependencies space, time, networks, images, ...
- combining data at multiple scales, or of multiple types

LETTER

https://doi.org/10.1038/s41586-018-0260-6

Spatial correlation bias in late-Cenozoic erosion histories derived from thermochronology

Taylor F. Schildgen^{1,2,6}*, Pieter A. van der Beek^{3,6}, Hugh D. Sinclair⁴ & Rasmus C. Thiede^{2,5}

In praise of small data: Example 3



Science, Apr 6 2018

Features: au M. Broadfoot

Antibiotic treatment of premature infants

- "the vast majority of the nearly half-million infants born prematurely in the United States are given antibiotics, even without evidence of infection."
- " 'Were beginning to recognize that the risk of giving that antibiotic may actually outweigh the benefit,' says Josef Neu, a neonatologist at the University of Florida in Gainesville
- "Neu hopes to provide hard evidence with a small clinical trial: A random selection of premature infants who would have been given antibiotics automatically will instead be placed in a nontreatment control group."
- "At first, he says, the university's institutional review board balked at his proposal for a randomized control trial"
- "Neu and his team plan to enroll 150 premature infants. However ... Doctors taking part in the study often put an infant on antibiotics at the first sign of illness, even if the baby had been randomized to the control group."

	v Help 😽 🚱	I O W 💿 🖬	. 🖁 🕙 🛜 🜒 8	0% 🗩 Tue 5:45 PM 🔍 🔕 😑
● ● ● ■ Be pi × औ Why × È Obse × 🗅 Scien × 🗰 Rethi	× E Stati: × CT Antib	× <i>CT</i> Archi × €	U.S. IX S∰S RITC X	👹 Ten S × │ New Tab × │ +
← → C https://clinicaltrials.gov/ct2/show/NCT02784821			¥	😧 🗢 🏩 🔤 🖉 🐑 🗄
👯 Apps 🔟 TD Canada Trust 📄 Apple 🗎 News 📊 Create LaTeX tabl 🦿	15 Questions All R 🖺 Ha	ppy Git and Git 📋 E	JMS: Login 🕒 JSTOR: #	Advanced » 🛅 Other Bookmarks
NIH U.S. National Library of Medicine				
ClinicalTrials.gov	Find Studies ▼	About Studies -	Submit Studies 🕶	Resources ▼ About Site ▼

Antibiotic "Dysbiosis" in Preterm Infants

The safety and scientific validity of this study is the responsibility of the study sponsor and investigators. Listing a study does not mean it has been evaluated A by the U.S. Federal Government, Know the risks and potential benefits of clinical studies and talk to your health care provider before participating. Read our disclaimer for details.

ClinicalTrials.gov Identifier: NCT02784821

Recruitment Status () : Recruiting First Posted (): May 27, 2016 Last Update Posted () : September 12, 2018

See Contacts and Locations

Sponsor:

University of Florida

Rochester Reliaberaters f Technology 2019 Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) Society for Pediatric Dermatology

Study Design

Study Type 🚯 :	Interventional (Clinical Trial)
Estimated Enrollment ():	420 participants
Allocation:	Randomized
Intervention Model:	Crossover Assignment
Masking:	None (Open Label)
Primary Purpose:	Treatment
Official Title:	Antibiotic Effects on the Developing Microbiome, Metabolome and Morbidities in Preterm Neonates
Study Start Date 1 :	January 2017
Estimated Primary Completion Date ():	September 2019
Estimated Study Completion Date ():	December 2019

- clinical trial (phase 2)
- 420 participants
- randomized
- crossover assignment
- no masking

Rochester Institute of Technology 2019

- Four Arms: antibiotics control, no antibiotics control, randomized to pre-emptive antibiotics, randomized to no pre-emptive antibiotics
- primary outcome measure NEC
- secondary outcome measures ...
- protocol for sampling fluids

Go to 🔻

15 in total

Study Record Versions

Version	A	в	Date	Changes
1			May 24, 2016	Nothing (earliest Version on record)
2	0	0	November 29, 2016	Study Status
3			December 23, 2016	Study Status
4	0	0	February 2, 2017	Recruitment Status, Oversight, Study Status and Contacts/Locations
5			August 28, 2017	Recruitment Status, Arms and Interventions, Study Status, Outcome Measures, Study Design, Conditions and Study Description
6	0	0	September 22, 2017	Study Status, Study Design, Sponsor/Collaborators and Study Identification
7			September 26, 2017	Contacts/Locations, Oversight and Study Status
8	0		January 17, 2018	Study Status
9			April 4, 2018	Study Status and Study Identification
10	0	0	April 19, 2018	Arms and Interventions, Contacts/Locations, Sponsor/Collaborators and Study Status
11			August 21, 2018	Study Status
12	0	0	September 11, 2018	Contacts/Locations and Study Status

- · A study version is represented by a row in the table.
- · Select two study versions to compare. One each from columns A and B.
- Choose either the "Merged" or "Side-by-Side" comparison format to specify how the two study versions are to be displayed. The Side-by-Side format only applies to the Protocol section of the study.
- · Click "Compare" to do the comparison and show the differences.
- · Select a version's date link to see a rendering of the study for that version.
- · Edits or deletions will be displayed in red.
- · Additions will be displayed in green.
- The yellow choices in the table indicate the study versions currently compared below. A yellow row indicates the study version being viewed.
- Hover over the "Recruitment Status" to see how the study's recruitment status changed.

Rochester Institute of Technology 2019

Changes (Side-by-Side) for Study: NCT02784821 May 24, 2016 (v1) -- September 11, 2018 (v12)

Changes in: Study Status, Arms and Interventions, Contacts/Locations, Oversight, Study Identification, Outcome Measures, Study Design, Conditions, Study Description and Sponsor/Collaborators

Show only changed modules

	May 24, 2016	September 11, 2018
Study Identification		
Unique Protocol ID:	IBB201501045	IBB201501045- N
chique i fotocorito.		
Brief Title:	Antibiotic "Dysbiosis" in Preterm Infants	Antibiotic "Dysbiosis" in Preterm Infants
Official Title:	Antibiotic Effects on the Developing Microbiome, Metabolome and Morbidities in Preterm Neonates	Antibiotic Effects on the Developing Microbiome, Metabolome and Morbidities in Preterm Neonates
Secondary IDs:		R21HD088005 [U.S. NIH Grant/Contract]
Study Status		
Record Verification:	May 2016	September 2018
Overall Status:	Unknown status [Previously: Not yet recruiting]	Recruiting
Study Start:	September 2016	January 2017
Primary Completion:	September 2017 [Anticipated]	September 2019 [Anticipated]
Study Completion:	September 2018 [Anticipated]	December 2019 [Anticipated]
First Submitted:	May 3, 2016	May 3, 2016

In praise of small data: Example 3



Science, Apr 6 2018

Features: au M. Broadfoot

In praise of small data: Example 4

CANADA

& LOG IN SURSCRIBE REGISTER 0 D AdChoices

B.C. wildfires stoked by climate change, likely to become worse: study

JEFF LEWIS > ENVIRONMENT REPORTER PUBLISHED JANUARY 8 2019 UPDATED 18 HOURS AGO

Jeff Lewis Jan 8 2019 Globe & Mail



Rochester Institute of Technology 2019 a wildfire southwest of the town of Cache Creek, B.C., on July 18, 2017. BEN NELMS/REUTERS

TRENDING					
1	OPINION				
	As parents of complex special-needs kids, we know inclusive education doesn't work				
	PHIL RICHMOND AND HAYLEY AVRUSKIN				
2	New Canadian telescope detecting more brief, powerful radio blasts from far beyond our galaxy				
3	Jagmeet Singh gets his chance as Trudeau calls three by-elections, including in Burnaby South				
4	BMO slices 1,000 points from its Toronto stock market forecast				
5	Toronto's Vena secures \$115-million in financing from U.S. private-equity				

firms -

38/59

- "anthropogenic climate change increased the area burned by a factor of 7 11"
- "We use a large ensemble of CanRCM4 ... consisting of 50 realizations on a 50-km grid. Each realization is driven by a member of the CanESM2 ... large ensemble ... We utilize data from 1961 to 2020."
- "A data set of gridded maximum (and minimum) temperature and precipitation anomalies was created by interpolating monthly values calculated from surface station observations relative to a 30-year climatology. Observational data was acquired from numerous sources and interpolated using a thin plate spline methodology."
- "... values for each year and large ensemble realization were pooled together for two time periods: 1961-1970 and 2011-2020, resulting in 500 values for each decade (10 years x 50 realizations). "

... event attribution



Figure 5. Time series (a, log scale) of regression-predicted annual burned area in the BC Southern Cordillera for bias-corrected CanRCM4 realizations (gray) and ensemble mean (bold), reanalysis (turquoise/purple), and observation (green). The dashed line marks the observed 2017 value. Probability distributions (b) for area burned amounts (log sc from decades outlined in corresponding colors in (a). The gray bar indicates the area burned amount in the distributi Rochestewith reduced anthropogenic influence (blue) of a corresponding percentile to the 2017 amount (dashed line) 4n/the distribution of the current decade, which includes anthropogenic influence (red). The distributions in (b) are used to

- complex computer simulation of global climate mathematics, numerical analysis
- creation of regional climate scenarios
- combined with available observational data
- · modelled with regression and kernel density estimation



statistics, data science

statistics, mathematics

Français



Rochester Institute of Technology 2

41/59

- · looking for extremes big data becomes small data
- · complex dependencies space, time, networks, images, ...
- combining data at multiple scales, or of multiple types



- collecting data
 - design of experiments
 - design of observational studies
- analysing data
 - data \rightarrow conclusions
 - data \rightarrow uncertainty about conclusions
- communication
 - explaining conclusions
 - explaining limitations

	BROWSE	PUBLISH	ABOUT	SEARCH	C
					advanced sear
OPTH ACCESS EDITORIAL				948 Save	21 Citation
Ten Simple Rules for Effective Statistica Robert E. Kase, Brian S. Caffo, Marie Davidan, Xiao-Li Meng, Bin Yu, Nancy Reid Published: June 9, 2016 - Intervited con/10.1371/Journal och: 1004061	al Practio	ce		226,840 View	3,148 Share

Theory of Statistical Inference

- how to get from data to conclusions
- with generalizable strategies
- what principles do we use to develop these strategies
- · how are these strategies to be evaluated

efficiency, precision

• a long history of the subject; using probability to both develop statistical methods and to evaluate their performance

Bayes, Laplace, Gauss; Student, Fisher, Neyman, Pearson, Jeffreys, ...

• leading to confidence intervals, *p*-values, estimates and standard errors, etc.

Rochester Institute of Technology 2019

- probability to describe physical haphazard variability
 - probabilities represent features of the "real" world in somewhat idealized form
 - subject to empirical test and improvement
- probability to describe the uncertainty of knowledge
 - measures rational or "impersonal" degree of belief, or
 - measures a particular person's degree of belief
 - linked to personal decision making

frequentist

Bavesian

Jeffreys, 1939,1961

F.P. Ramsey, 1926

... role of probability

- confidence intervals or *p*-values refer to empirical probabilities
- inference is assessed by behaviour of the procedure under hypothetical repetition
- the Bayesian approach to inference describes uncertainty of knowledge
- this can be interpreted empirically by appeal to a notion of calibration





"[7 – 11]"

Collaborations



- Canadian Statistical Sciences Institute
 CANSSI
- launched in 2012
- funded 2014–2019 by Natural Sciences and Engineering Research Council
- national scope, virtual institute
- Collaborative Research Teams multidisciplinary, multi-institution, statistical leadership, scientific engagement

State Space Models for Fisheries Science 2014 – 2017



Computer and Physical Models in Earth, Atmospheric and Ocean Sciences 2014 - 2017



Collaborative Research Teams

Copula Dependence Modeling: theory and applications 2014 - 2017



Marked Point Processes and Wildfire Modeling 2015 – 2018



Statistical Inference for Complex Surveys 2015 - 2018



Modern Spectrum Methods in Time Series Analysis 2015 – 2018



Collaborative Research Teams

Joint Analysis of Neuro-imaging Data 2016 – 2019



Rare DNA Variants and Human Complex Traits 2016 - 2019



Statistical Analysis of Large Administrative Health Databases

2017 - 2020



Spatial Modeling of Infectious Diseases: Environment and Health

2018 - 2021



Rochester Institute of Technology 2019

Collaborative Research Teams

Sustainable Fisheries and Stock Assessments 2018 – 2021



2019 - 2022

Contingent Capital and Calibration of Capital Structure Models

Statistical Methods for Public Health Microbiology 2018 - 2021



Statistical methods for the analysis of genetic data with survival outcomes

Rochester Institute of Technology 2019

Other Programs



postdoctoral fellows

in 2019 joint with SAMSI at University of Victoria

• Health Science Collaborating Centers

11 centers across the country – emphasis on experiential training

- partial support for workshops and conferences
- support for ASA Datafest and other data-thons
- Distinguished Visitors Program
- KickStart Collaborations Rochester Institute of Technology 2019



Some statistical challenges in data science

- inference in very high dimensions ($p \to \infty)$

requires dimension reduction or sparsity assumptions

- standard statistical methods may not scale well ($n
ightarrow \infty$)

distributed computing

data provenance and quality

"big data has arrived, but big insights have not" Harford, 2014

- learning causal relations from observational data
- · data is everywhere

requires rapid increase in training



Embrace data chaos.

You're buried in raw data. Traditional tools require you to structure it before it can be useful. With Splunk, you can start digging for actionable insights immediately, no matter what state that data is in.

splunk.com/chaos



 $R\!\cdot\!I\!\cdot\!T\quad \text{Rochester Institute of Technology}$

Directories Search RIT



IN THIS SECTION

About the College \sim

Mission and Vision

Overarching Goals

Strategic Plan

Undergraduate Programs

Graduate Programs



Rochester Institute of Technology 2019

Thank You!