The Quality 4.0 Revolution:

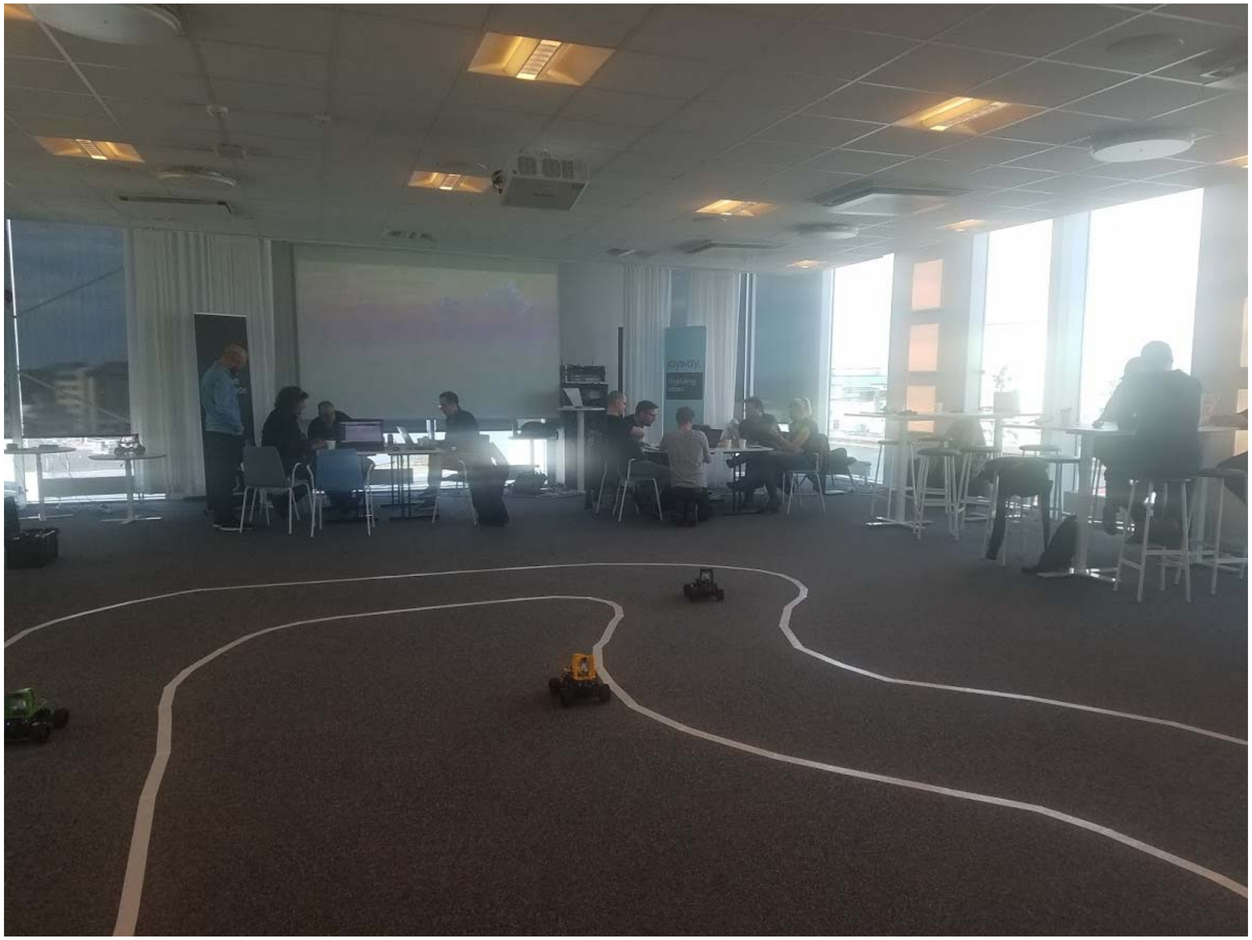
Reveal Hidden Insights Now With Data Science and Machine Learning

Nicole M. Radziwill, PhD, MBA

ASQ Fellow & Editor, *Software Quality Professional*

Quality Practice Leader

INTELEX



https://aws.amazon.com/blogs/machine-learning/build-an-autonomous-vehicle-on-aws-and-race-it-at-the-reinvent-robocar-rally/

Cameras on front of remote-controlled cars sampled at 20 images/second

Based on patterns in the photos, give car steering instructions

Goal: Keep them on the road up to max speed!

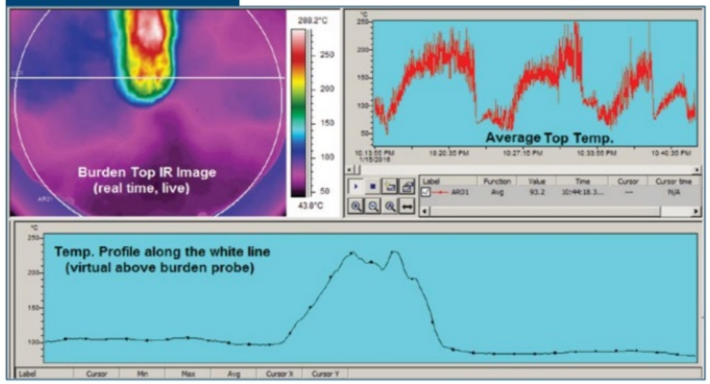
(AWS with multilayer neural net in Python/TensorFlow in Jupyter notebooks)

INTELEX

"… a leading car manufacturer… was experiencing **recurrent faults**, and therefore **high warranty costs**, on the rear tail-lights. The engineers had been looking at the rear of the vehicle for the answer (and not succeeding) however **the software quickly found the root cause of the fault to be located in the roof of the vehicle, a part of the car that had not even been investigated as a possible source**. Sometimes prior experience, or being too close to a problem, can inhibit a solution if an old hypothesis is applied to a new problem."

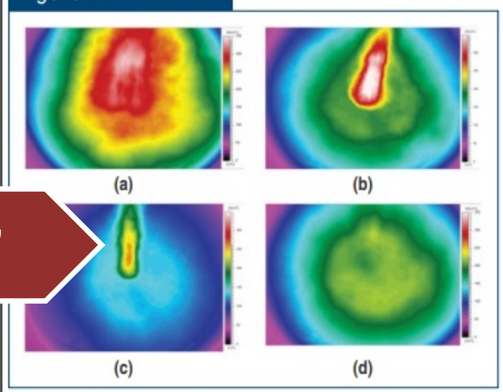
From http://www.qmtmag.com/display_eds.cfm?edno=9665074

Figure 5



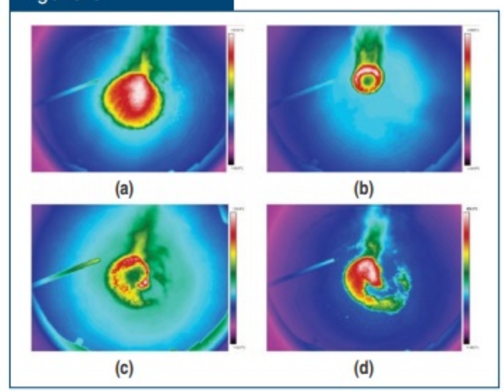
“chimney”

Figure 7



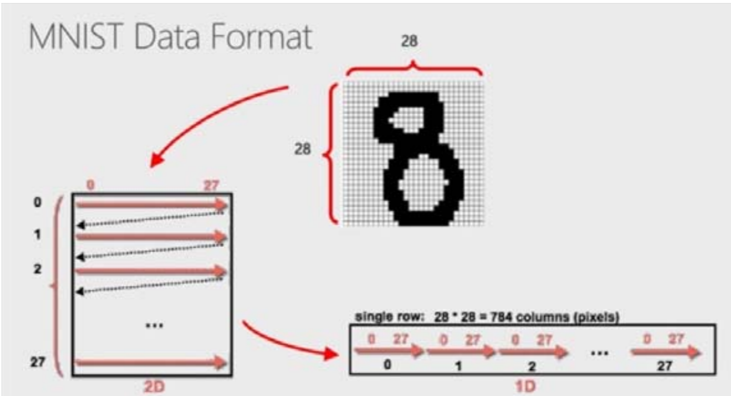
Size of the central gas flow: large chimney (a), medium chimney (b), small chimney (c), no clear chimney (d).

Figure 8

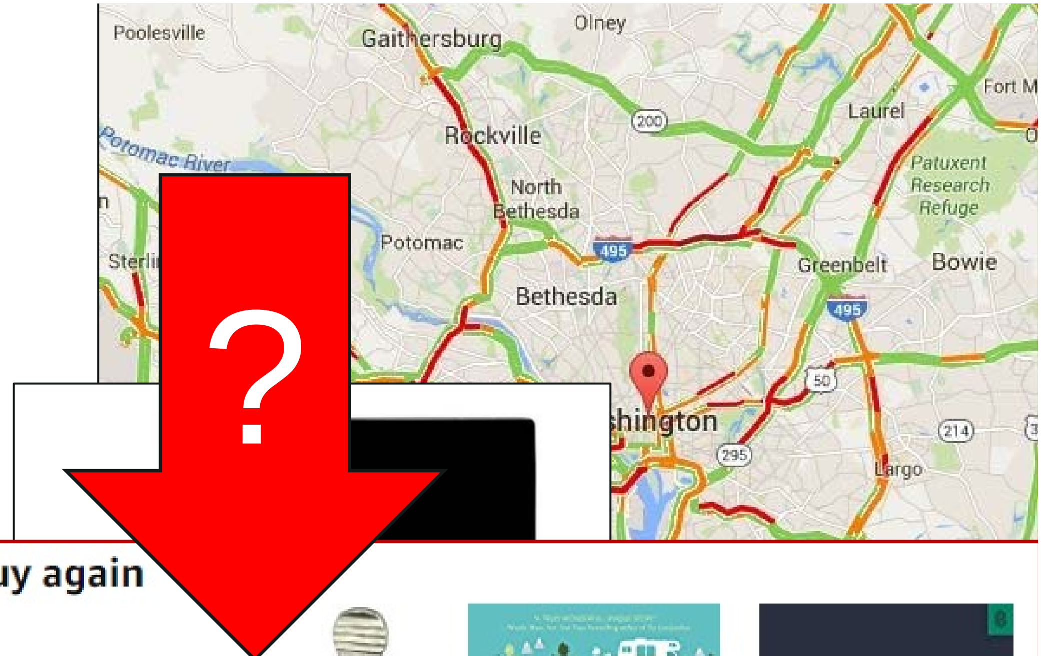


Shape of the central gas flow: solid shape (a), hollow or donut shape (b), “6” shape (c), “G” shape (d).

Chen, X., Liu, F., Hou, Q., & Lu, Y. (2009, August). Industrial high-temperature radar and imaging technology in blast furnace burden distribution monitoring process. In Electronic Measurement & Instruments, 2009. ICEMI'09. 9th Intl Conference on (pp. 1-599). IEEE.



From https://www.slideshare.net/WindowsDev/build-2017-p4005-microsoft-r-server-development-to-deployment-workflow



New Ways to Solve Old Problems

You will:

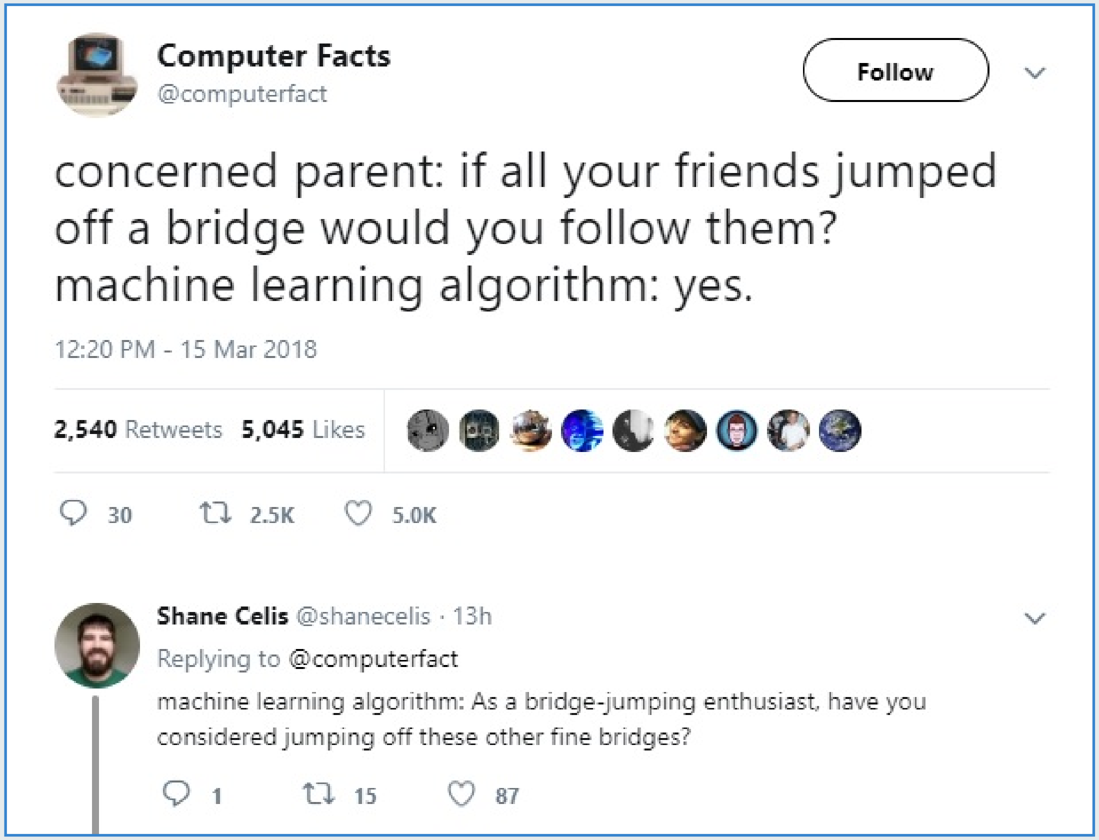
1. Understand **what Quality 4.0 is** and the value propositions it presents.

2. Identify the **types of problems** to which you can apply these new methods in your organization.

3. Explain how these technologies support **engagement and collaboration** to promote a connected and empowered workforce.

4. Explore **ways to incorporate data science and machine learning into your work**, whether you are drowning in data or just starting your digital transformation.

INTELEX


Computer Facts
@computerfact
Follow
concerned parent: if all your friends jumped off a bridge would you follow them?
machine learning algorithm: yes.
12:20 PM - 15 Mar 2018
2,540 Retweets 5,045 Likes
30 2.5K 5.0K
Shane Celis @shanecelis · 13h
Replying to @computerfact
machine learning algorithm: As a bridge-jumping enthusiast, have you considered jumping off these other fine bridges?
1 15 87

INTELEX

Nicole Radziwill

Quality Practice Lead, **Intelex**

Fellow, American Society for Quality (ASQ)
CSSBB #11962 & CMQ/OE #9583
Ph.D. Quality Systems, Indiana State
Editor, *Software Quality Professional*

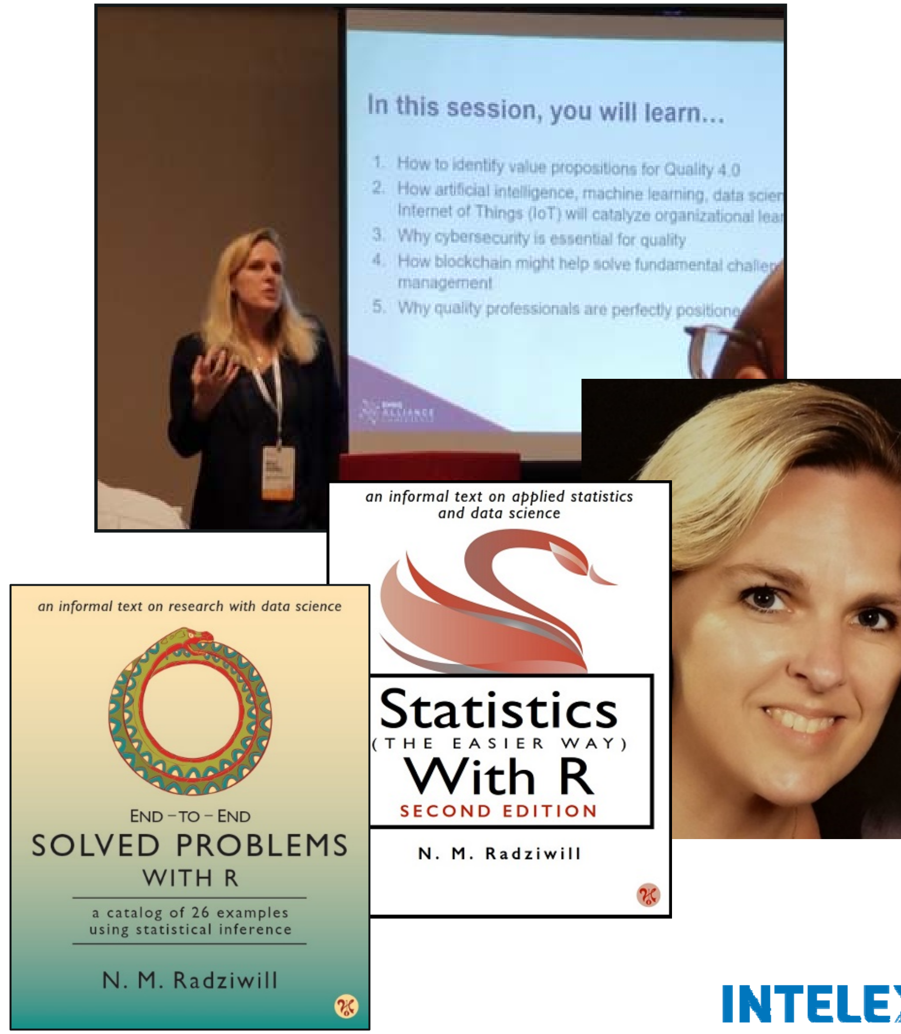
Previously*:*

•Project Manager, Solution Architect, Engagement Manager // Meteorological Research, Telecom, Manufacturing, Software, High Tech 1995-2002

•Division Head, Software, Green Bank Observatory, 2002-2006

•Assistant Director of End to End Operations, National Radio Astronomy Observatory, 2006-2009

•Associate Professor of Data Science & Production Systems, James Madison University, 2009-2018; taught data science & applied machine learning from 2009-2018



INTELEX

1: Quality 4.0

and value propositions enabled by data science

INTELEX

Quality as:

1.0 - Inspection

2.0 - Design

3.0 - Engagement

Quality Planning	Quality Assurance/ Quality Control	Quality Improvement

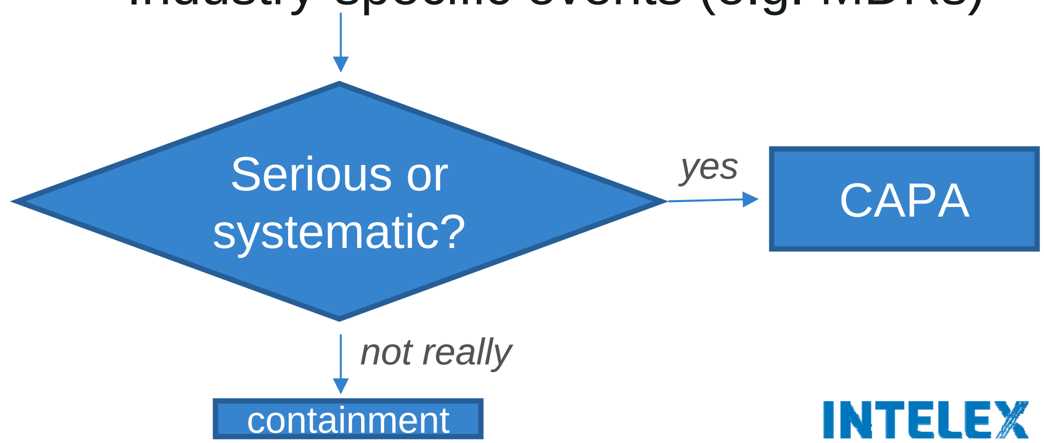
• Establish quality goals
• Understand organizational context (mission, vision, values, strategy, position, capabilities)
• Understand business environment
• Identify customer & stakeholder needs – Voice of Customer (VoC)
• Design products/services to meet needs, implement design controls
• Design processes for consistent production & related controls
	• Evaluate process performance
	• Compare performance with quality goals
	    • Manage “quality events” and update controls
	        • Audits
	        • Management reviews
	        • Nonconformances
	        • Incidents/near misses
	        • Complaints
	• Out-of-control production
	• Act to close the gaps
	
	Integrate lessons learned into Quality Planning
		• Build capabilities (people and technical)
		• Build infrastructure
		• Identify and justify needs and gaps
		• Conduct improvement activities and establish controls to maintain them
		    • Quick wins (PDCA)
		    • Process improvement (DMAIC, Lean)
		    • Process design or redesign (DMADV)
		    • Business process mgmt (BPM) or Robotic Process Automation (RPA)
		• Confirm/Validate changes

INTELEX

Quality Events

indicate that quality goals are not being met and action is needed

- Nonconforming product
- Incidents/near misses
- Customer complaints
- Audit findings
- Recalls/warranty calls
- Deviations (from SOP)
- Out-of-control Action Plans
- Industry-specific events (e.g. MDRs)



INTELEX

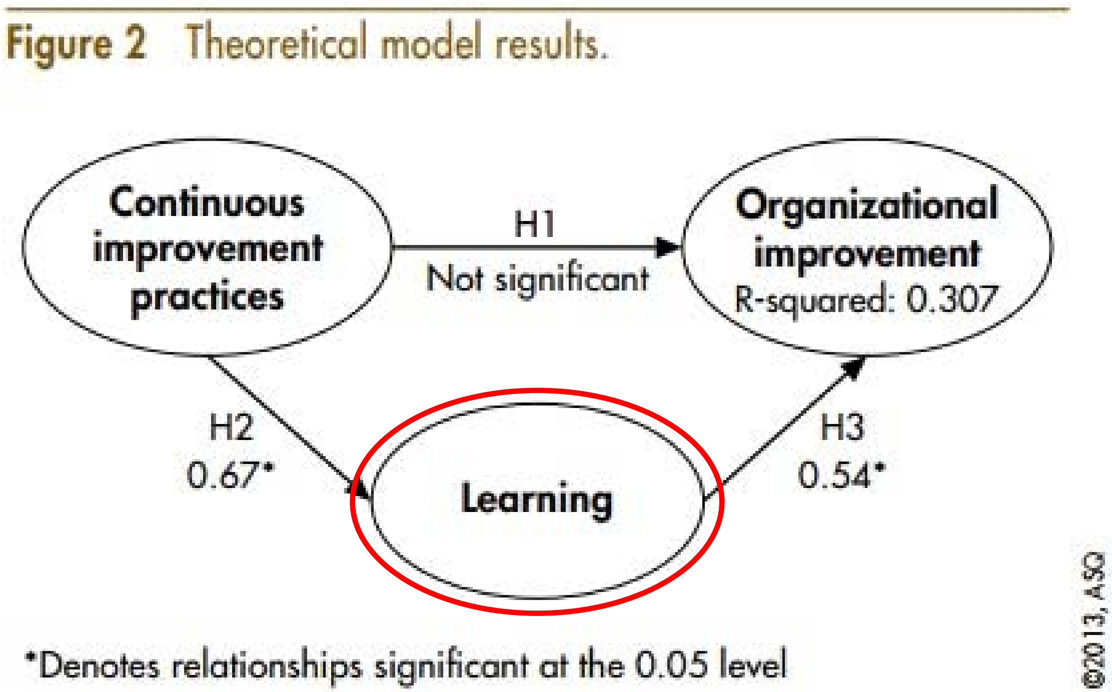
Quality Controls

to prevent or correct unwanted or unexpected change → stability and consistency

- Calibrations
- Maintenance
- Inspections
- Sampling incoming parts
- Process validation
- Mistake-proofing
- In-situ process monitoring
- Environment monitoring
- Professional testing/ competency assessment
- Training programs and reminders
- Corrective actions taken
- Information security/ network security

INTELEX

Why Quality Systems Work



From Kovach, J. V., & Fredendall, L. D. (2013). The influence of continuous improvement practices on learning: An empirical study. *The Quality Management Journal*, *20*(4), 6.

AI/ML Can Enhance & Accelerate Organizational Learning

People	Intelligent Systems
As people learn about the work environment, they update: • policies, • procedures, • practices, and • heuristics to improve the performance of people, projects, processes & the bottom line.	When AI/ML systems learn, they update how they: • predict, • classify, • find patterns, and • identify important predictors to improve the performance of people, projects, processes & the bottom line.

* Jim Duarte (ASQ TV) also noticed that the algorithms do what we've been used to doing as quality professionals… just iteratively, and much faster.

INTELEX

Quality as:

1.0 - Inspection

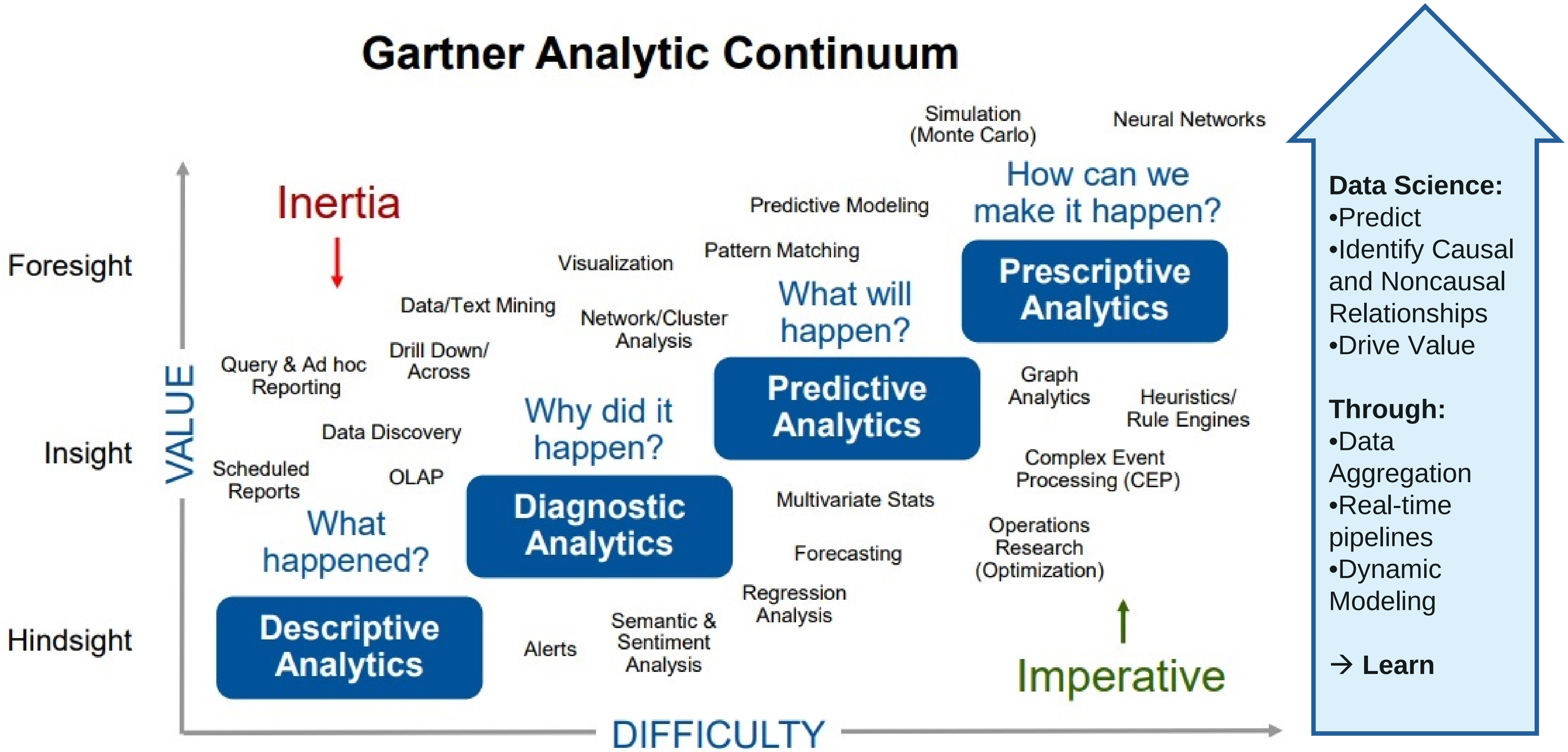
2.0 - Design

3.0 - Engagement

4.0 - Discovery

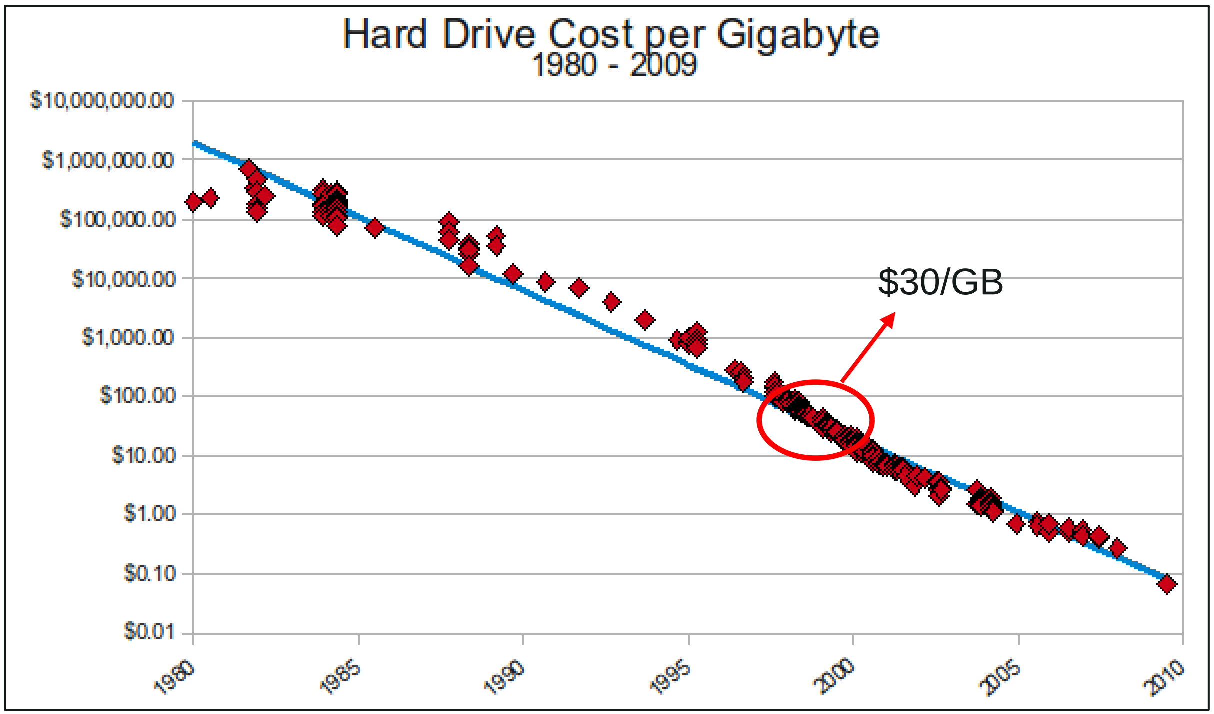
AI/ML evolves data-driven decision making to be *self-aware* & adapt to changing environments, circumstances, and customer/stakeholder needs

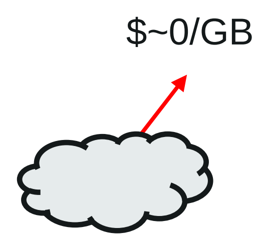
ASQ



@doug_laney

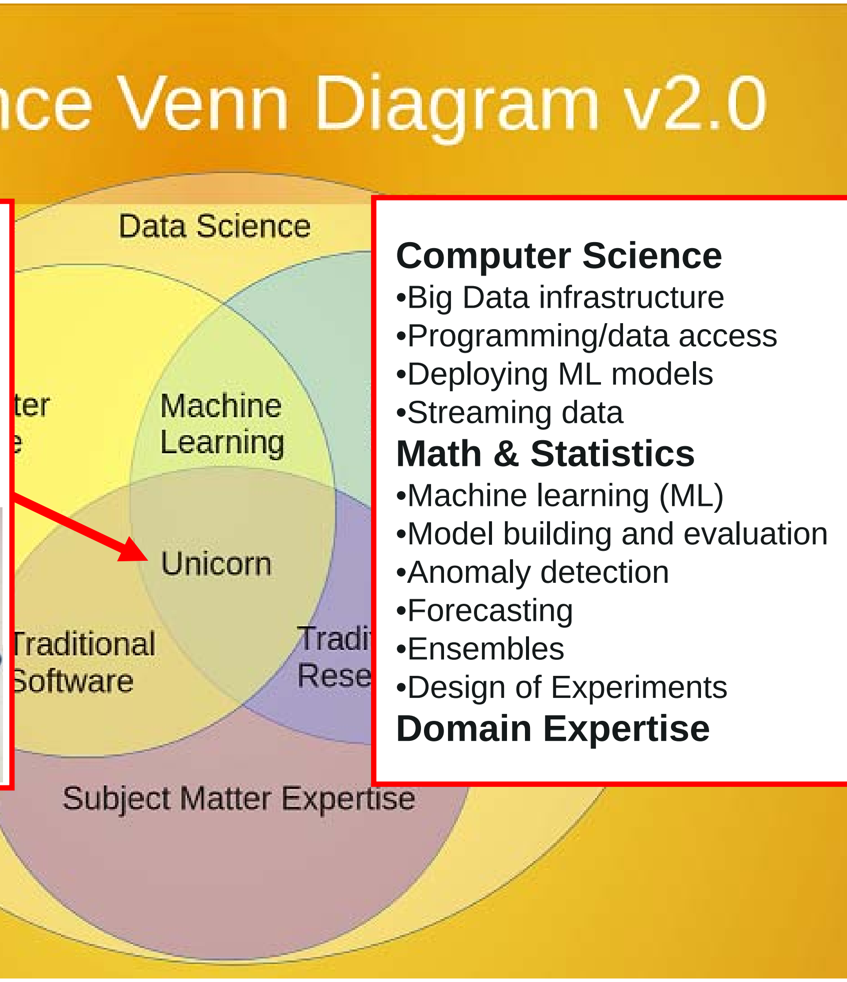
INTELEX





From http://www.mkomo.com/cost-per-gigabyte

INTELEX



c/o 92 !

Computer Science

- Big Data infrastructure
- Programming/data access
- Deploying ML models
- Streaming data

Math & Statistics

- Machine learning (ML)
- Model building and evaluation
- Anomaly detection
- Forecasting
- Ensembles
- Design of Experiments

Domain Expertise

Scientific Method vs Data Science

1. Develop a hypothesis based on theory
2. Take a random, representative sample from the population
3. Compare characteristics to your prediction
4. Draw conclusions

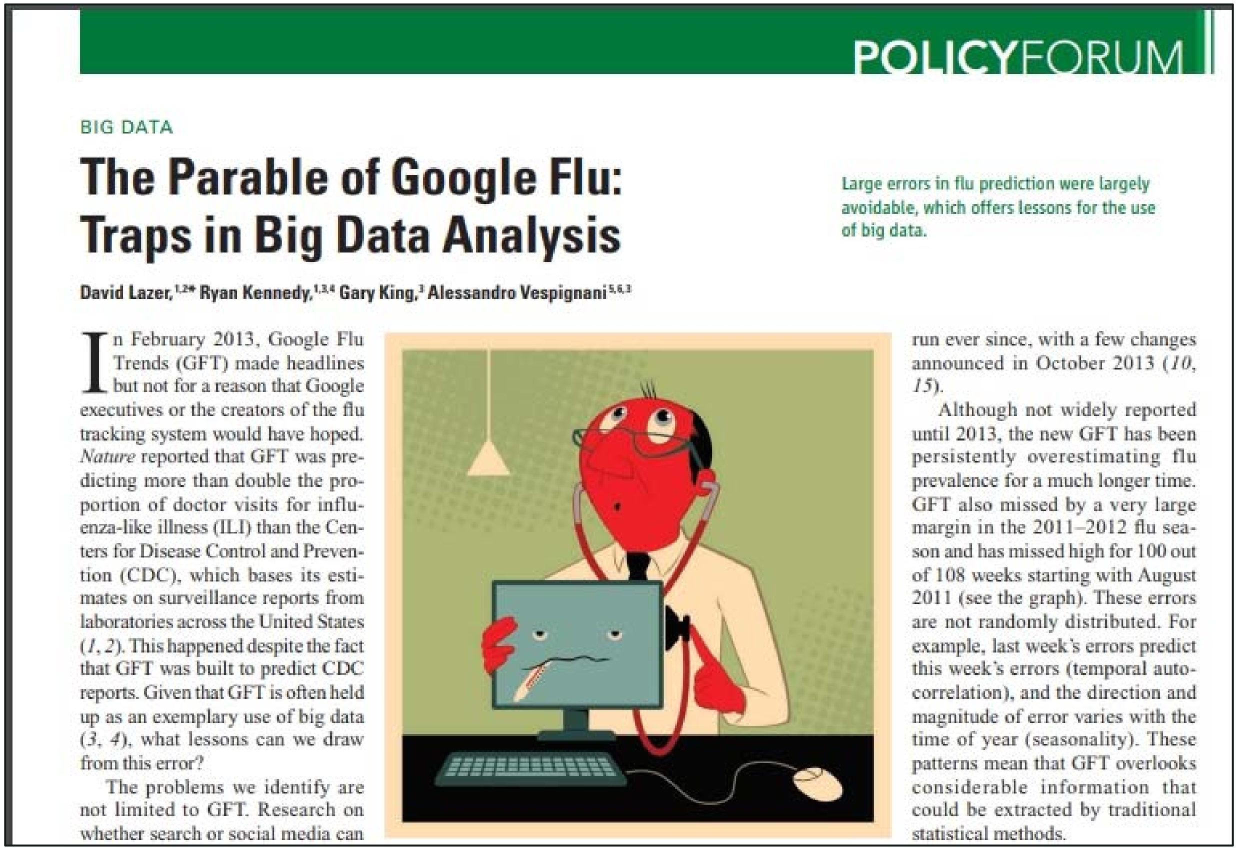
Traditional

1. Just look at the entire population of data and see what you find… we have the tools to do it and the data is cheap! Who needs theory or a hypothesis?

Data Science

Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, *45*(2), 167-256.

INTELEX



POLICYFORUM

BIG DATA

The Parable of Google Flu: Traps in Big Data Analysis

Large errors in flu prediction were largely avoidable, which offers lessons for the use of big data.

David Lazer,[1,2]* Ryan Kennedy,[1,3,4] Gary King,[3] Alessandro Vespignani[5,6,3]

In February 2013, Google Flu Trends (GFT) made headlines but not for a reason that Google executives or the creators of the flu tracking system would have hoped. *Nature* reported that GFT was predicting more than double the proportion of doctor visits for influenza-like illness (ILI) than the Centers for Disease Control and Prevention (CDC), which bases its estimates on surveillance reports from laboratories across the United States (*1*, *2*). This happened despite the fact that GFT was built to predict CDC reports. Given that GFT is often held up as an exemplary use of big data (*3*, *4*), what lessons can we draw from this error?

The problems we identify are not limited to GFT. Research on whether search or social media can

run ever since, with a few changes announced in October 2013 (*10*, *15*).

Although not widely reported until 2013, the new GFT has been persistently overestimating flu prevalence for a much longer time. GFT also missed by a very large margin in the 2011–2012 flu season and has missed high for 100 out of 108 weeks starting with August 2011 (see the graph). These errors are not randomly distributed. For example, last week's errors predict this week's errors (temporal autocorrelation), and the direction and magnitude of error varies with the time of year (seasonality). These patterns mean that GFT overlooks considerable information that could be extracted by traditional statistical methods.

' …GFT overlooks information that could be extracted by traditional statistical methods.'

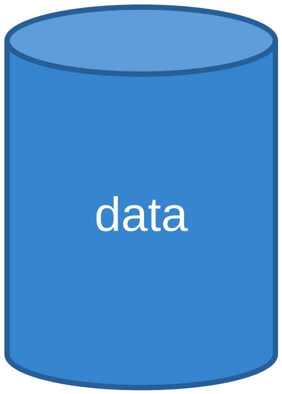
' "**Big data hubris**" is the often implicit assumption that big data are a substitute for, rather than a supplement to, traditional data collection and analysis.'

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. Science, 343(6176), 1203-1205.

INTELEX

What is Big Data?

- Lots of it (**volume**)
- It's coming at you fast (**velocity**)
- Different formats and sampling frequencies (**variety**)
- Huge variations in data quality (**veracity**)
- Different people/organizations produce it or own it (**governance**)
- It could easily change or disappear (**control**)
- There may be restrictions on how you use it (**policy**)



"… Big Data is anything bigger or more complex than what your organization is currently prepared to handle."

-- one of the world's experts on cyberinfrastructure for Big Data at a National Science Foundation panel (2013)

INTELEX

Type of data	Inventory management	Transportation management	Customer and supplier relationship management
Sales	How can sales data be used with detailed customer data to improve inventory management either in terms of forecasting or treating some inventory as "committed" based on specific shoppers requirements?	How can more current sales data be used to re-direct shipments in transit? How can sales data, integrated with detailed customer data, be used for more efficient and effective merge-in-transit operations?	How can more granular sales " data from the wide variety of sources that exist be used to improve visibility on the one hand and trust on the other, between trading partners?
Consumer	How can face profiling data for shopper identification, emotion detection, and eye-tracking data be used to determine which items to carry and stock at particular shelf locations?	How can delivery preferences captured in online purchases be used to manage transportation mode and carrier selection decisions?	How can customer sentiment about products purchased based on "Likes," "Tweets," and product reviews be used to collaborate on forecasts?
Location and time	How can sensor data used to detect location in store, be used to improve inventory management, including departmental merchandising decisions?	How can sensor data in the distribution center be used to anticipate transportation requirements?	How can location and time-stamp data of shoppers be used for collaborative assortment and merchandising decisions?

From Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics*, 34(2), 77-84.

INTELEX

FEATURE
QUALITY 4.0

Let's Get Digital

The many ways the fourth industrial revolution is reshaping the way we think about quality | by Nicole Radziwill

Just the Facts

To adapt to the disruptive changes brought about by the fourth industrial revolution, a new perspective on quality is required.

Organizations must leverage Quality 4.0 tools—such as artificial intelligence, machine intelligence, pervasive computing and connectivity—to improve their quality and performance.

Quality professionals possess the skills required to help their organizations succeed in the fourth industrial revolution.

The technology landscape is richer and more promising than ever before. In many ways, cloud computing, big data, virtual reality (VR), augmented reality (AR), blockchain, additive manufacturing, artificial intelligence (AI), machine learning (ML), Internet Protocol Version 6 (IPv6), cyber-physical systems and the Internet of Things (IoT) all represent new frontiers. These technologies can help improve product and service quality, and organizational performance.

In many regions, the internet is now as ubiquitous as electricity. Components are relatively cheap. And, a robust ecosystem of open-source software libraries means that engineers can solve problems 100 times faster than just two decades ago.

This digital transformation is leading us toward connected intelligent automation: smart, hyperconnected agents deployed in environments where humans and machines cooperate—and leverage data—to achieve shared goals. This isn't the world's first industrial revolution. In fact, it is its fourth—and the disruptive changes it will bring suggest we'll need a fresh perspective on quality to adapt to it.

24 QP October 2018 | qualityprogress.com

Radziwill, N. M. (2018, October). Let's Get Digital: The many ways the fourth industrial revolution is reshaping the way we think about quality. *Quality Progress*, p. 24-29. http://qualityprogress.com

Quality 4.0 = C I A

Connectedness

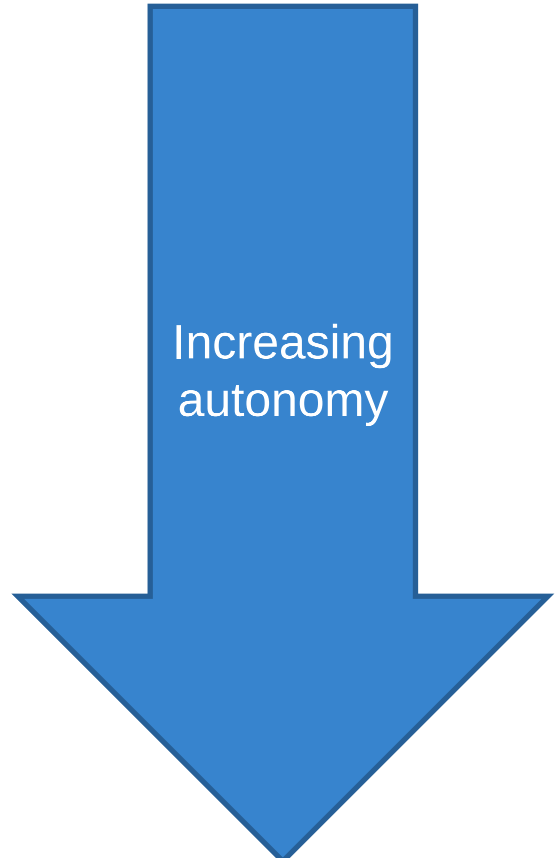
Intelligence

Automation

INTELEX

Degrees of Automation

1. Human specifies process and computer only implements
2. Computer assists by determining options
3. Computer assists by determining options and suggests choice
4. Computer assists by determining options and selects choice - human has option to do it, or not
5. Computer selects and implements options if human approves
6. Computer selects option and automatically implements, but gives human chance to stop the process flow
7. Computer selects and implements options, then tells human about results
8. Computer selects and implements options, but only tells human about results if asked for report
9. Computer selects and implements options, but only tells human about results if asked, only reports some information
10. Computer performs whole job; may or may not tell human



From Sheridan, T. B., & Verplank, W. L. (1978). Human and computer control of undersea teleoperators. MASSACHUSETTS INST OF TECH CAMBRIDGE MAN-MACHINE SYSTEMS LAB.

INTELEX

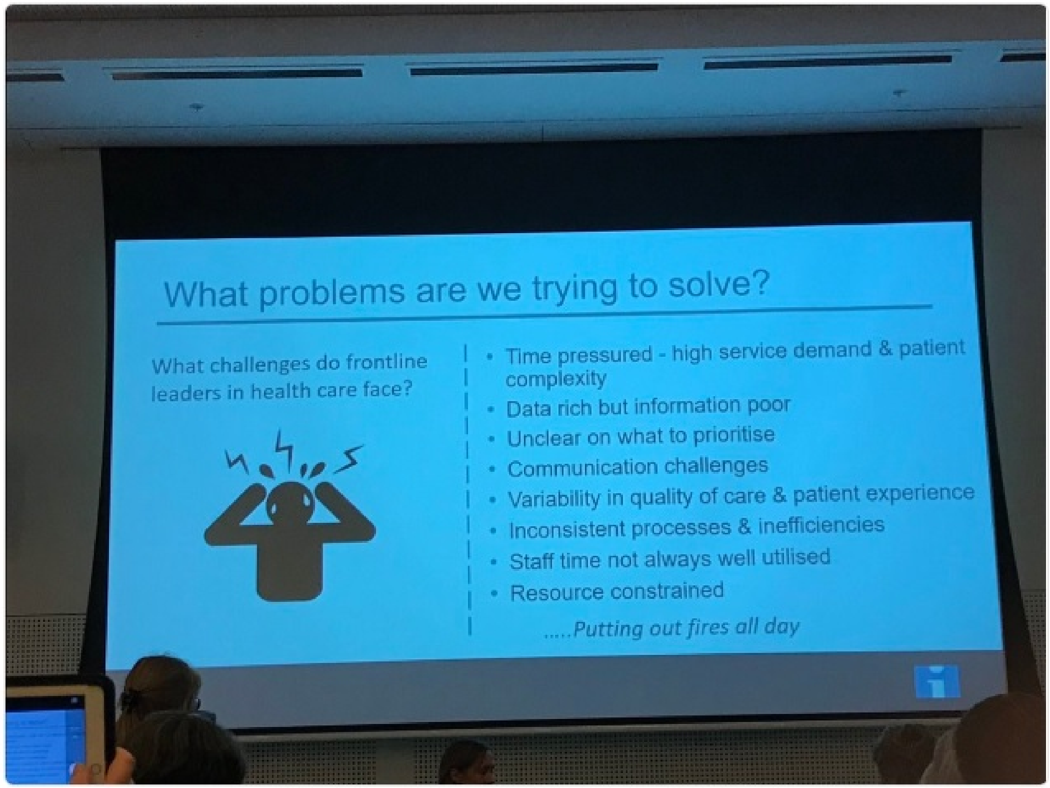
2: Types of Problems

& how to identify challenges that are well supported by DS/ML

INTELEX

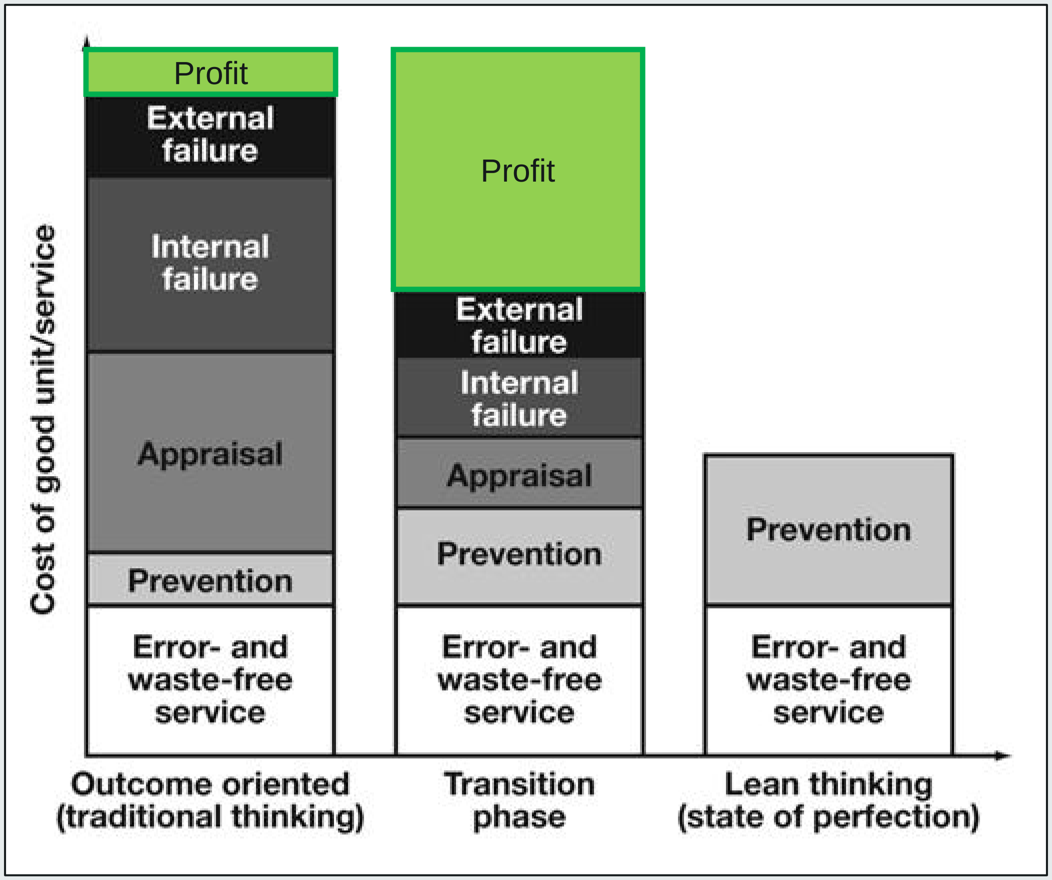
Adam Gruer @AdamGruer · 24m

Sit down if you think your org has solved any of these problems for front line staff.   Narrator: only 2 people out of 200 sat down #quality2018



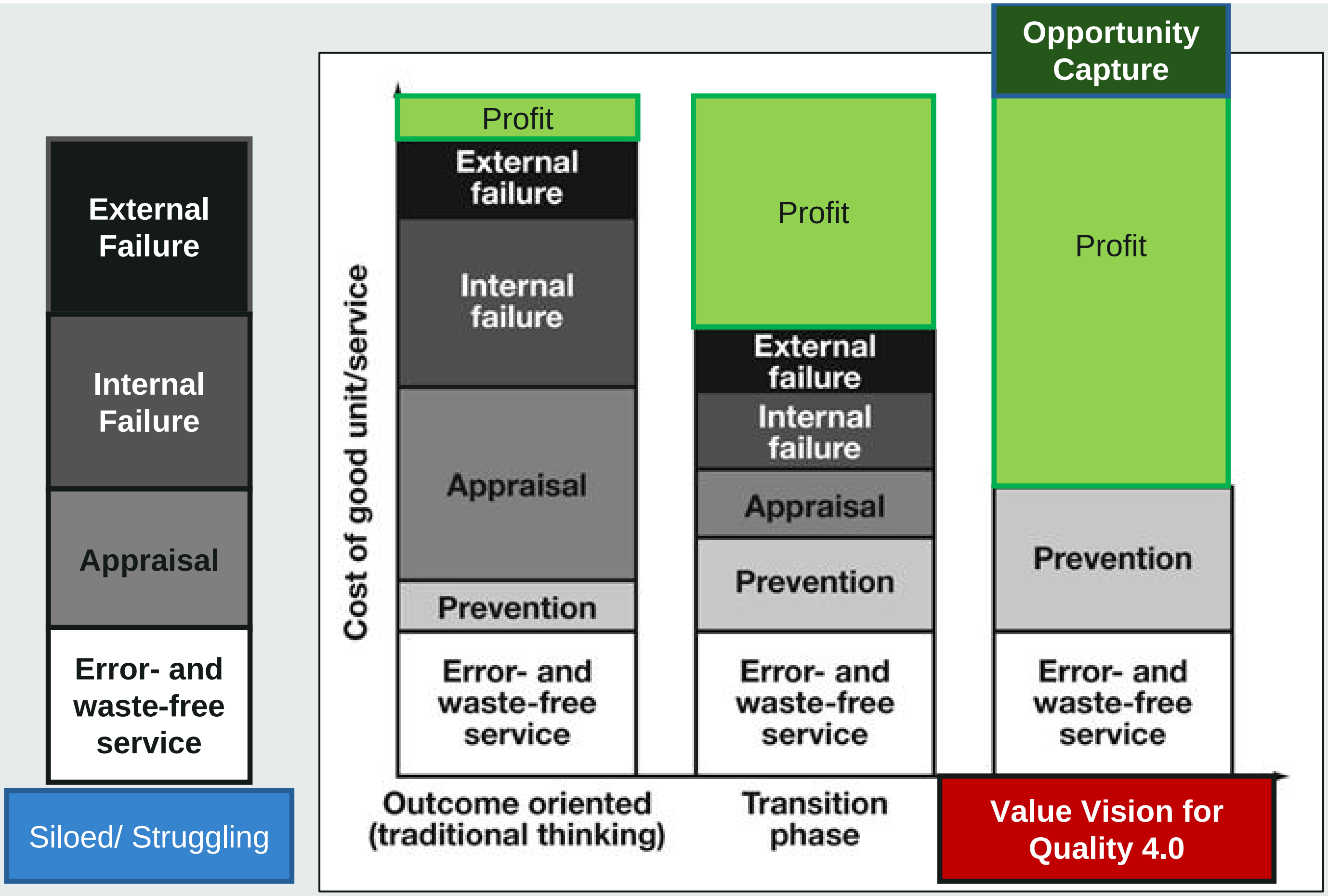
2

INTELEX



From Carlson, R. O., Amirahmadi, F., & Hernandez, J. S. (2012). A primer on the cost of quality for improvement of laboratory and pathology specimen processes. *American journal of clinical pathology*, *138*(3), 347-354.

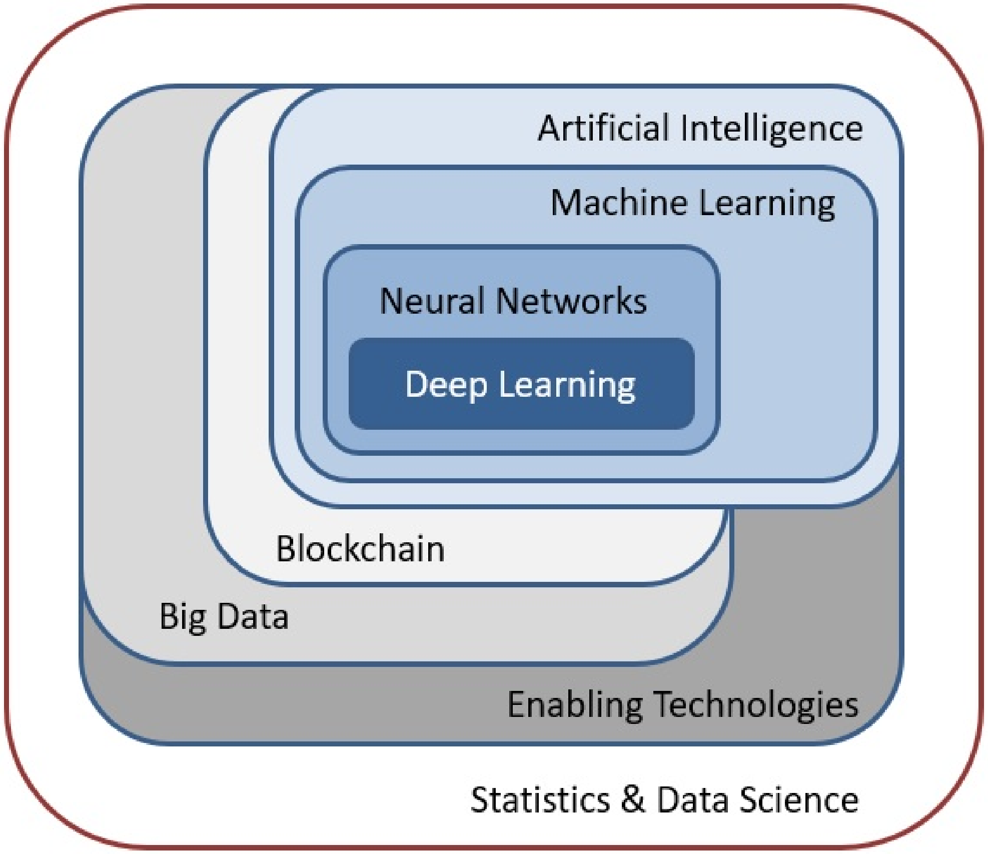
INTELEX



From Carlson, R. O., Amirahmadi, F., & Hernandez, J. S. (2012). A primer on the cost of quality for improvement of laboratory and pathology specimen processes. *American journal of clinical pathology*, *138*(3), 347-354.

INTELEX

Quality 4.0 Ecosystem



Radziwill, N. M. (2018, October). Let's Get Digital: The many ways the fourth industrial revolution is reshaping the way we think about quality. *Quality Progress*, p. 24-29. http://qualityprogress.com

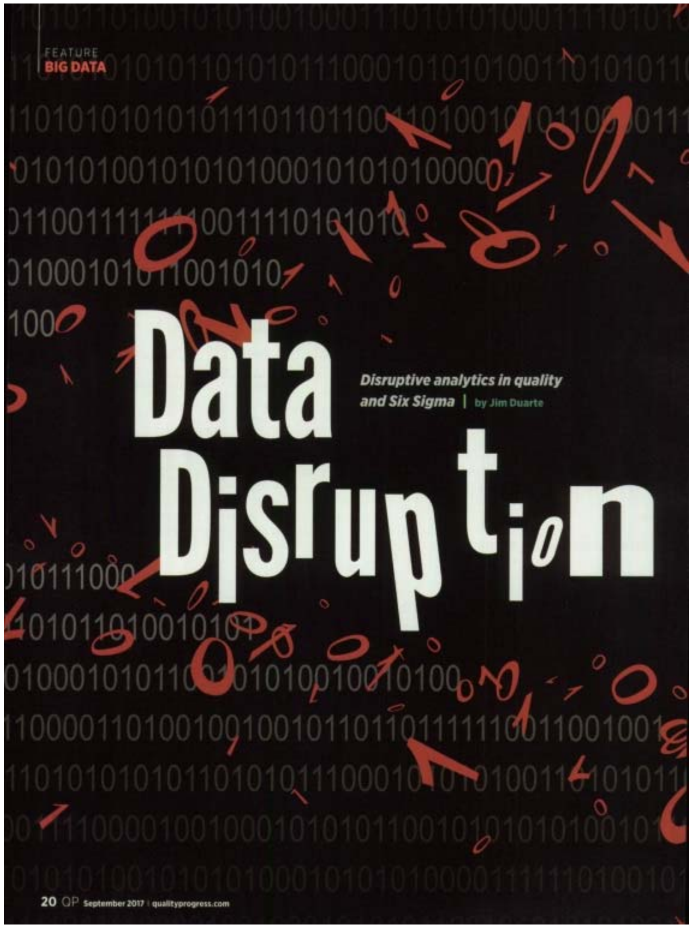
INTELEX

Machine Learning Tasks

1. Prediction
2. Classification
3. Pattern Identification
4. Data Reduction
5. Anomaly Detection
6. Pathfinding

Data Science Tasks

1. Frame the problem(s)
2. Explore the data
3. Build models
4. Evaluate model performance
5. Deploy models and real-time processing **pipelines**
6. Uncover high-value insights
7. Tell stories
8. Do it again… and again



From Duarte, J. (2017). Data Disruption. Quality Progress, 50(9), 20-24.

“Data can take several forms, such as data at rest and data in motion. Data that are sitting in warehouses waiting to be analyzed are referred to as **data at rest**…

Data in motion (data analyzed in real time as the event occurs, such as click streams and sensors) flow through a connected device to a database on the receiving end for immediate analysis… used to fine tune the [predictive] models… and analytics at the edge… deploys information to control processes in real time.”

-- Jim Duarte

INTELEX

Quality Problem	Machine Learning	Feasibility
Audio anomalies	Classification (CNN)	Feasible
Checking for leakage	Image classification	Potential
Setting up engines	Reinforcement learning	Unnecessarily complicated
Supervising testing	Image classification	Potential
Feature detection	Image classification	Unnecessarily complicated
Cause diagnosis	Multidimensional Regression, association rule learning	Potential
Machine adjustments	Multidimensional Regression	Potential
Surface defects	Image classification	Feasible
Assembly gaps	Image Regression	Unnecessarily complicated
Part identification	Image classification	Unnecessarily complicated
Machine error cause	Image classification	Unnecessarily complicated
Field quality detection	Association rule learning	Unnecessarily complicated
Field quality cause analysis	Association rule learning	Feasible
Predictive maintenance	Classification	Feasible
Control chart pattern detection	Classification	Feasible
Optimizing production parameters	Classification	Potential

Granstedt Möller, E. (2017). The Use of Machine Learning in Industrial Quality Control. Thesis, KTH Vetenskap Och Konst (Stockholm, Sweden).

INTELEX

Value Propositions for Quality 4.0

1. **Augment** (or improve upon) human intelligence
2. **Increase** speed and enhance the quality of decision-making
3. **Anticipate** changes, reveal biases, and adapt to new circumstances and data sources
4. **Learn** how to learn: cultivate self-awareness and other-awareness
5. **Reveal** opportunities for continuous improvement
6. **Improve** transparency, traceability, and auditability

MACHINE LEARNING:
Any place where **HEURISTICS** are used for decision making

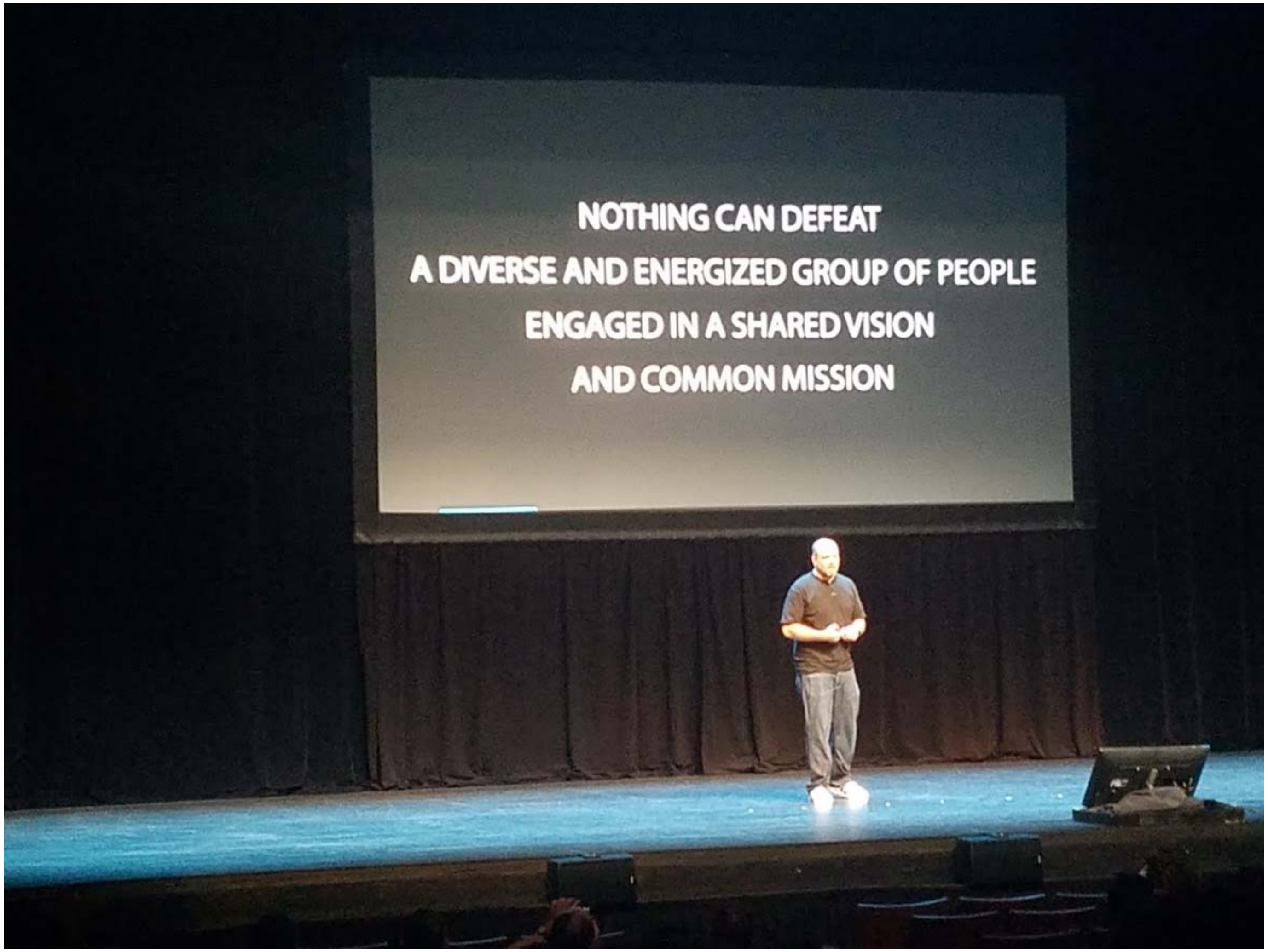
BLOCKCHAIN:
Any place there are **TRANSACTIONS** (pre, during, post)

Radziwill, N. M. (2019). Value Propositions for Quality 4.0. *Manuscript in review.*

3: Engagement & Collaboration

with Quality 4.0

INTELEX


NOTHING CAN DEFEAT
A DIVERSE AND ENERGIZED GROUP OF PEOPLE
ENGAGED IN A SHARED VISION
AND COMMON MISSION

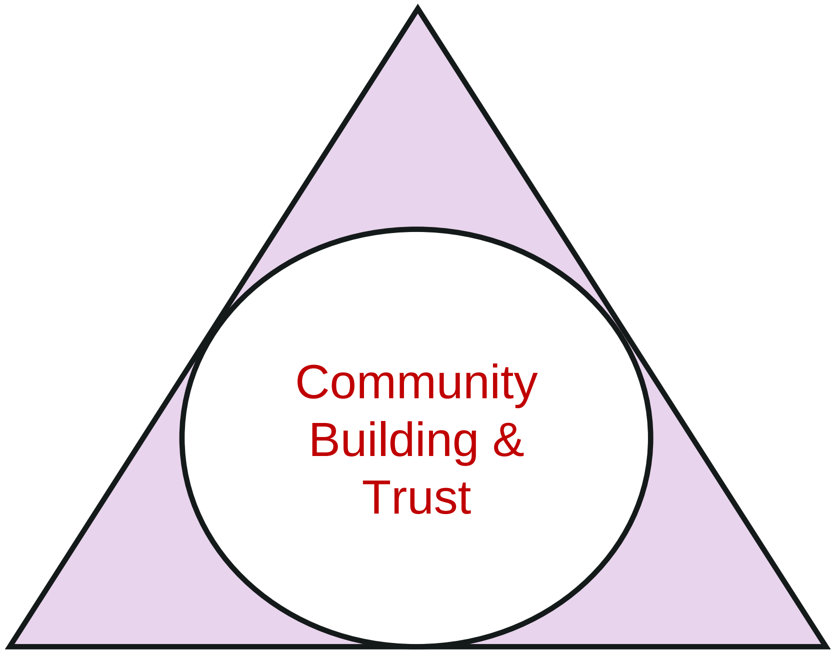
INTELEX

What is Engagement?

A Review of ~120 Research Papers

- Drive out fear
- Build safety and quality culture
- Find meaning and purpose

Excitement



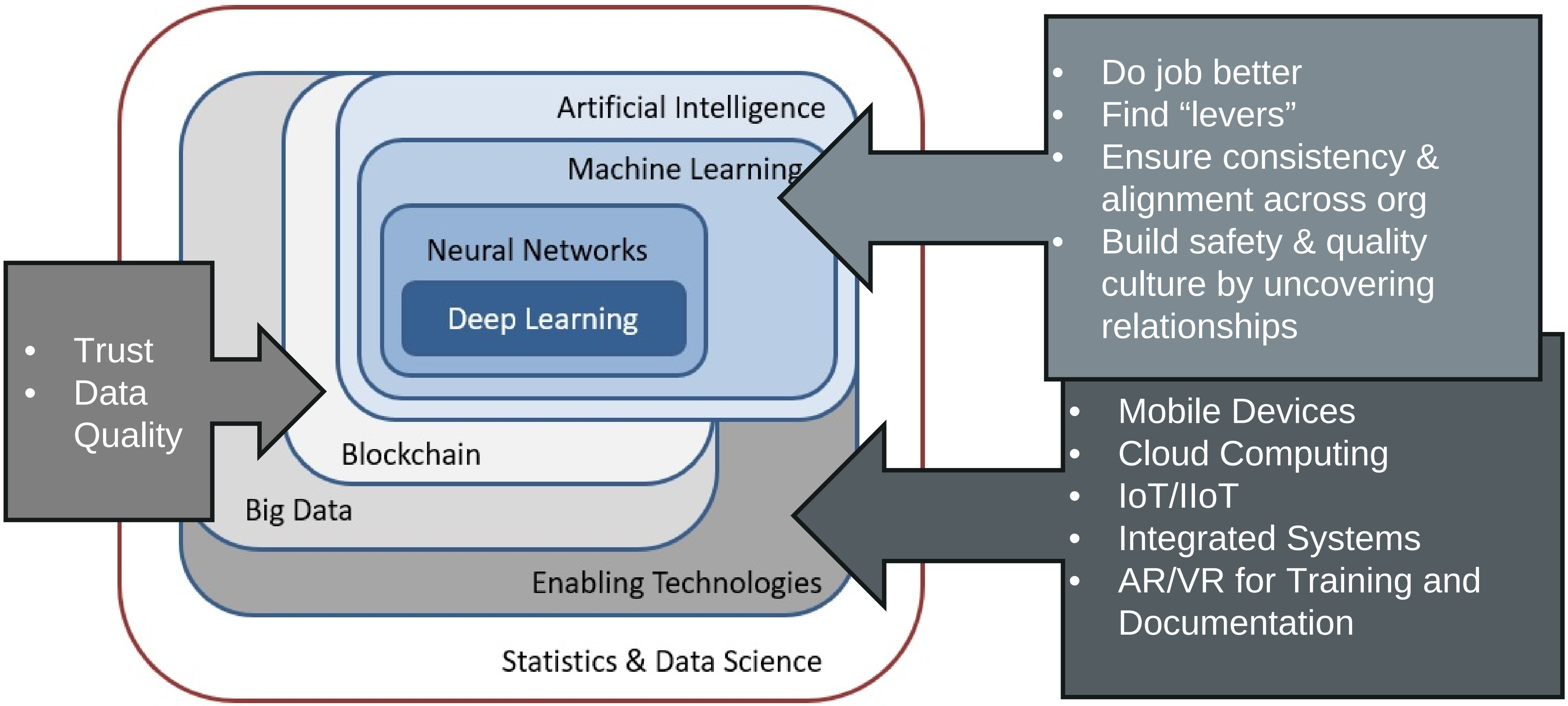
- Helps you do your job
- Understand expectations
- Understand process
- Get into flow

Execution

Extension

- Ensures alignment
- Supports collective efforts
- Achieves consistency
- Deploys strategy broadly & deeply

Radziwill, N. M. (2019). The Relationship Between Engagement and Quality Outcomes. *Manuscript in review.*



Radziwill, N. M. (2018, October). Let's Get Digital: The many ways the fourth industrial revolution is reshaping the way we think about quality. *Quality Progress*, p. 24-29. http://qualityprogress.com

INTELEX

Not Just for Manufacturing, Not Just for Germany

While the focus of Plattform Industrie 4.0 has, without doubt, been very much on German manufacturing, it is by no means an exclusive club for any particular industry or region. Gartner's recent definition of the term Industrie 4.0 underlines this point, calling it "a business-outcome-driven digital transformation approach to generate value from the collaboration of multiple partners in ecosystems across value chains and industries."

In other words, it is not simply about technologies such as the Internet of Things (IoT) and artificial intelligence (AI) increasing production efficiency at any one site or even one company. Industrie 4.0 enables integration across entire enterprises, powering the creation of brand new ecosystems that can span multiple industries.

I have spoken many times about the importance of openness – in terms of technologies, development mindsets, partnerships, and collaborations. Industrie 4.0 is for me another great example of the crucial role openness plays in successful digitalization. On the one hand, there are the individual interorganizational networks that Industrie 4.0 helps to create – ecosystems that connect companies and industries in completely new value chains. But on a broader scale, cooperation and openness between many different types of organizations across Germany, Europe, and indeed the globe are required to fully exploit the advantages of digitalization. And among the various goals the Plattform Industrie 4.0 pursues, the promotion of networking between global consortia and stakeholders, and the coordination of this exchange is a clear priority.

Bernd Leukert, Executive Board SAP, at https://news.sap.com/2018/01/industrie-4-0-why-openness-and-collaboration-make-all-the-difference/

INTELEX

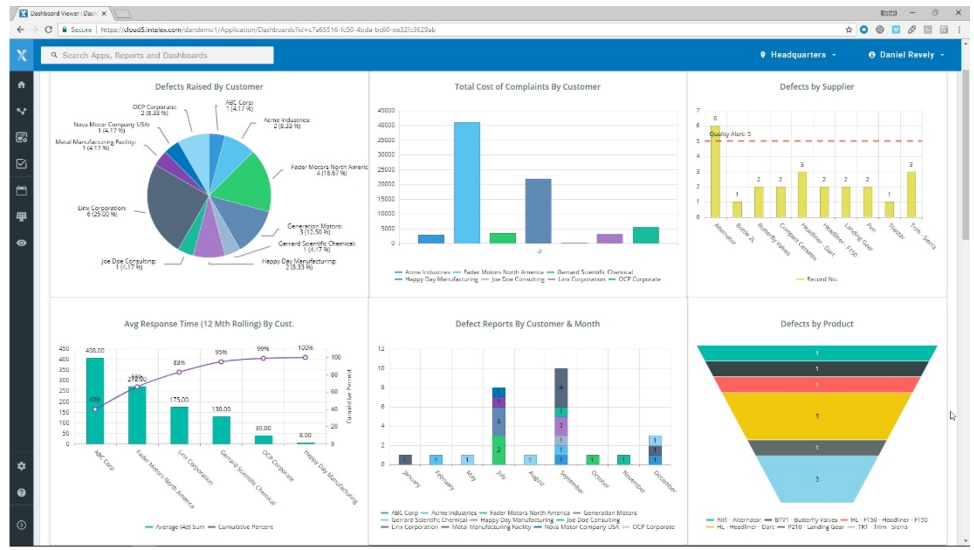
4: Reveal Hidden Insights Now

incorporate DS/ML & ensure readiness for the future

INTELEX

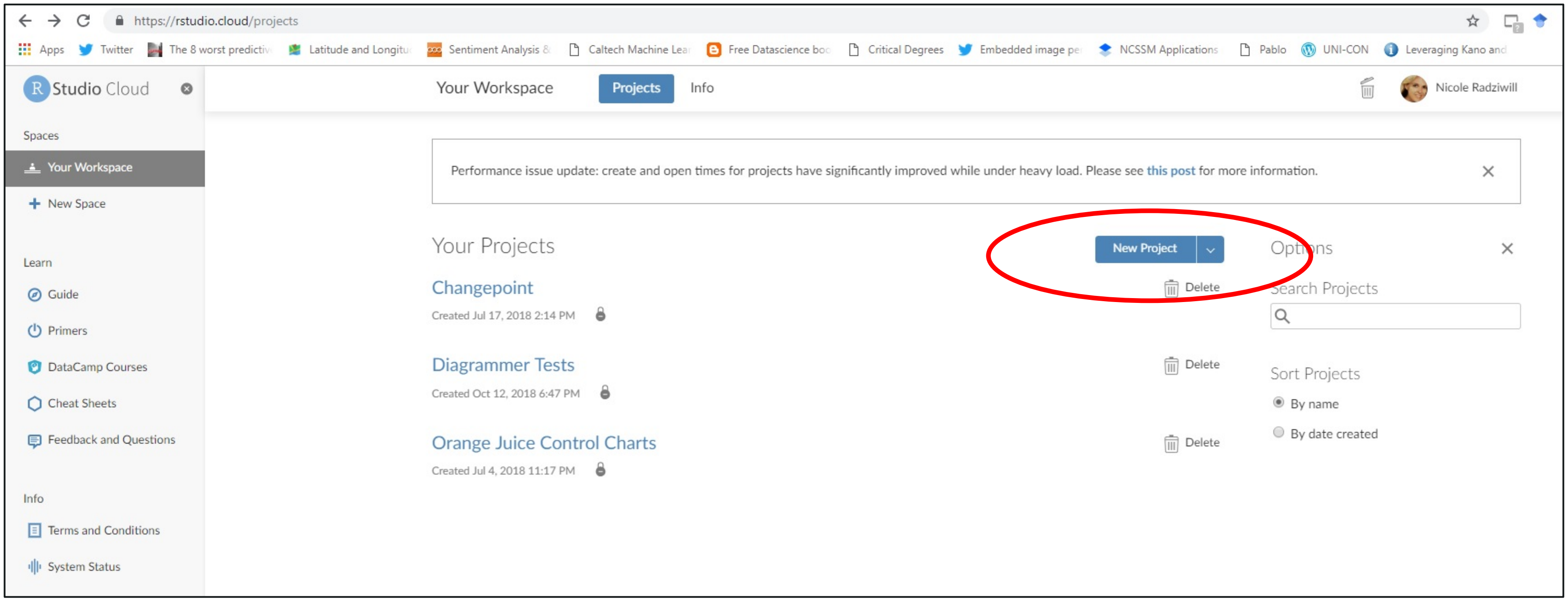
I want my QMS/IMS to…

- Help me **make better decisions** about my processes
- **Audit itself and alert me** when I need to do something
- Tell me where to **focus** my resources for maximum added value
- Based on risk assessments and theory of constraints, tell me how should I **prioritize** work
- Alert me when an important **change** occurs at the systems level

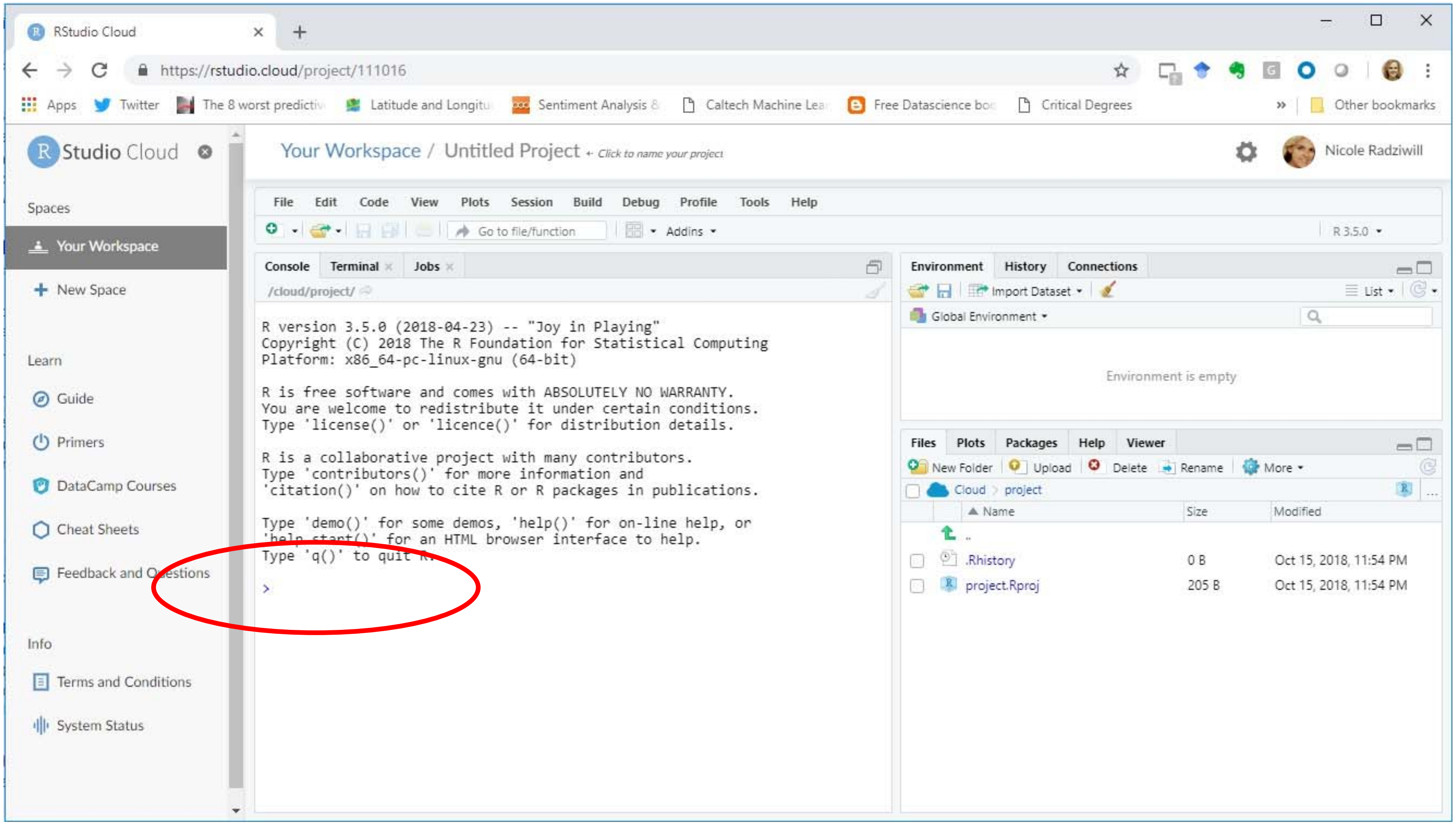


INTELEX

DS in the Cloud: http://rstudio.cloud

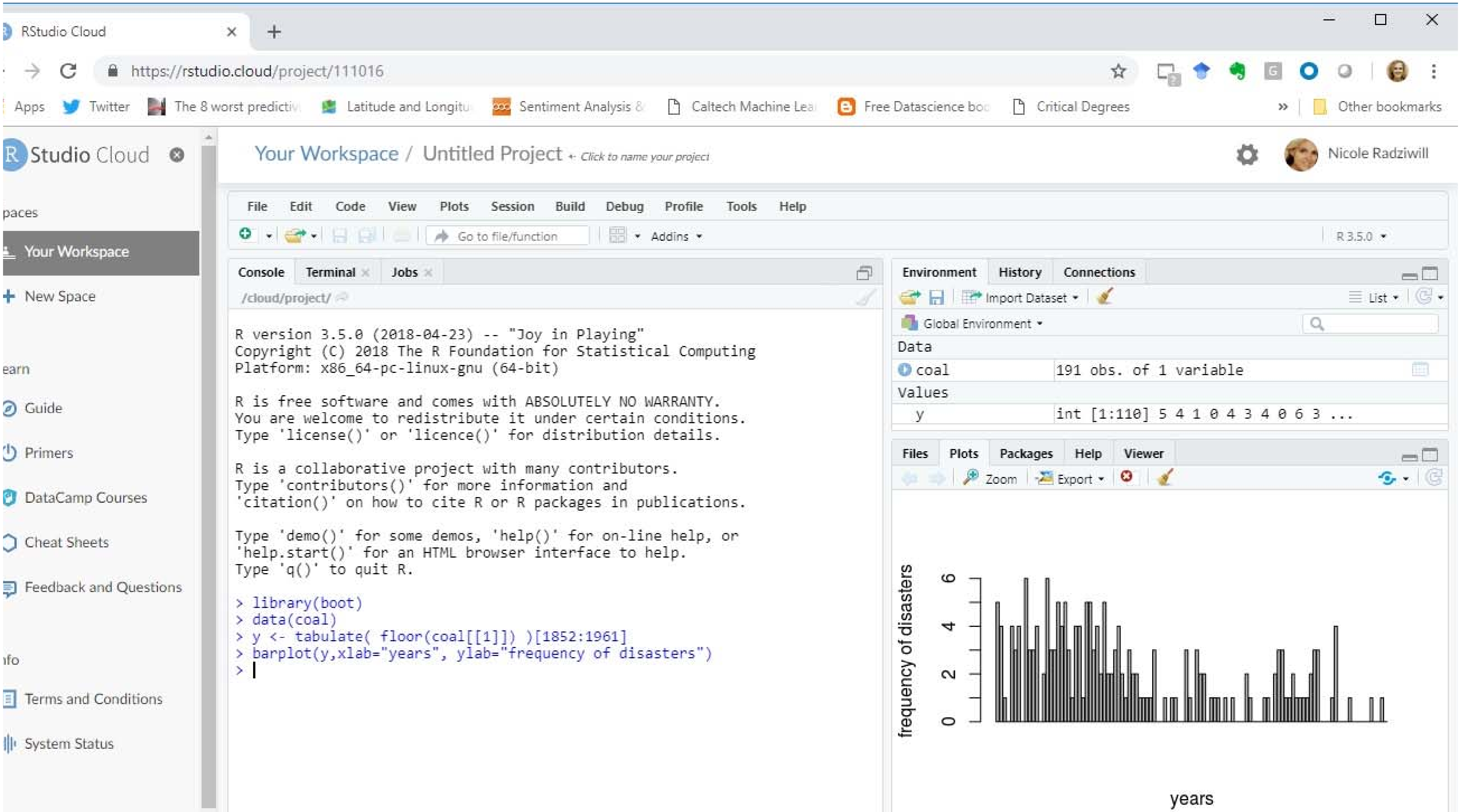


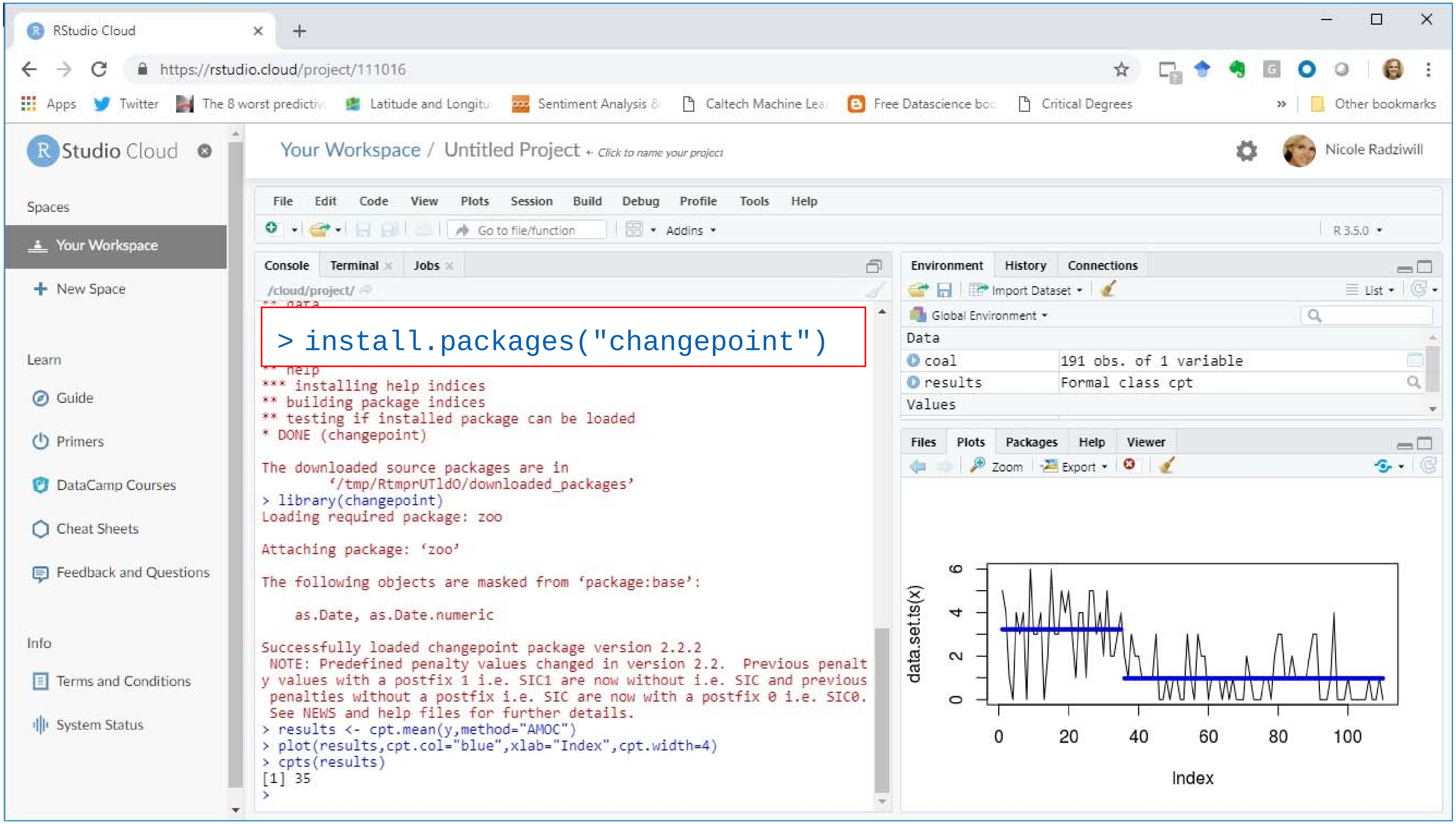
Example to follow: https://qualityandinnovation.com/2015/07/14/a-simple-intro-to-bayesian-change-point-analysis/



RStudio Cloud
https://rstudio.cloud/project/111016
Studio Cloud
Your Workspace / Untitled Project ← Click to name your project
Nicole Radziwill
Spaces
Your Workspace
New Space
Learn
Guide
Primers
DataCamp Courses
Cheat Sheets
Feedback and Questions
Info
Terms and Conditions
System Status
File Edit Code View Plots Session Build Debug Profile Tools Help
Console Terminal Jobs
/cloud/project/
R version 3.5.0 (2018-04-23) -- "Joy in Playing"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
Environment History Connections
Global Environment
Environment is empty
Files Plots Packages Help Viewer
Cloud > project
Name Size Modified
.Rhistory 0 B Oct 15, 2018, 11:54 PM
project.Rproj 205 B Oct 15, 2018, 11:54 PM

INTELEX


RStudio Cloud
https://rstudio.cloud/project/111016
Your Workspace / Untitled Project ← Click to name your project
Nicole Radziwill
Your Workspace
New Space
Guide
Primers
DataCamp Courses
Cheat Sheets
Feedback and Questions
Terms and Conditions
System Status
File Edit Code View Plots Session Build Debug Profile Tools Help
Console Terminal Jobs
R version 3.5.0 (2018-04-23) -- "Joy in Playing"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
> library(boot)
> data(coal)
> y <- tabulate(floor(coal[[1]]))[1852:1961]
> barplot(y,xlab="years", ylab="frequency of disasters")
Environment History Connections
Global Environment
Data
coal 191 obs. of 1 variable
Values
y int [1:110] 5 4 1 0 4 3 4 0 6 3 ...
Files Plots Packages Help Viewer
frequency of disasters
years


RStudio Cloud
https://rstudio.cloud/project/111016
Apps
Twitter
The 8 worst predictiv
Latitude and Longitu
Sentiment Analysis &
Caltech Machine Lea
Free Datascience bo
Critical Degrees
Other bookmarks
RStudio Cloud
Spaces
Your Workspace
New Space
Learn
Guide
Primers
DataCamp Courses
Cheat Sheets
Feedback and Questions
Info
Terms and Conditions
System Status
Your Workspace / Untitled Project
Click to name your project
Nicole Radziwill
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R 3.5.0
Console Terminal Jobs
/cloud/project/
> install.packages("changepoint")
*** installing help indices
** building package indices
** testing if installed package can be loaded
* DONE (changepoint)
The downloaded source packages are in
'/tmp/RtmprUTldO/downloaded_packages'
> library(changepoint)
Loading required package: zoo
Attaching package: 'zoo'
The following objects are masked from 'package:base':
as.Date, as.Date.numeric
Successfully loaded changepoint package version 2.2.2
NOTE: Predefined penalty values changed in version 2.2. Previous penalt
y values with a postfix 1 i.e. SIC1 are now without i.e. SIC and previous
penalties without a postfix i.e. SIC are now with a postfix 0 i.e. SIC0.
See NEWS and help files for further details.
> results <- cpt.mean(y,method="AMOC")
> plot(results,cpt.col="blue",xlab="Index",cpt.width=4)
> cpts(results)
[1] 35
Environment History Connections
Import Dataset
List
Global Environment
Data
coal 191 obs. of 1 variable
results Formal class cpt
Values
Files Plots Packages Help Viewer
Zoom
Export
data.set.ts(x)
0
2
4
6
0
20
40
60
80
100
Index

Example 2: Wine Quality

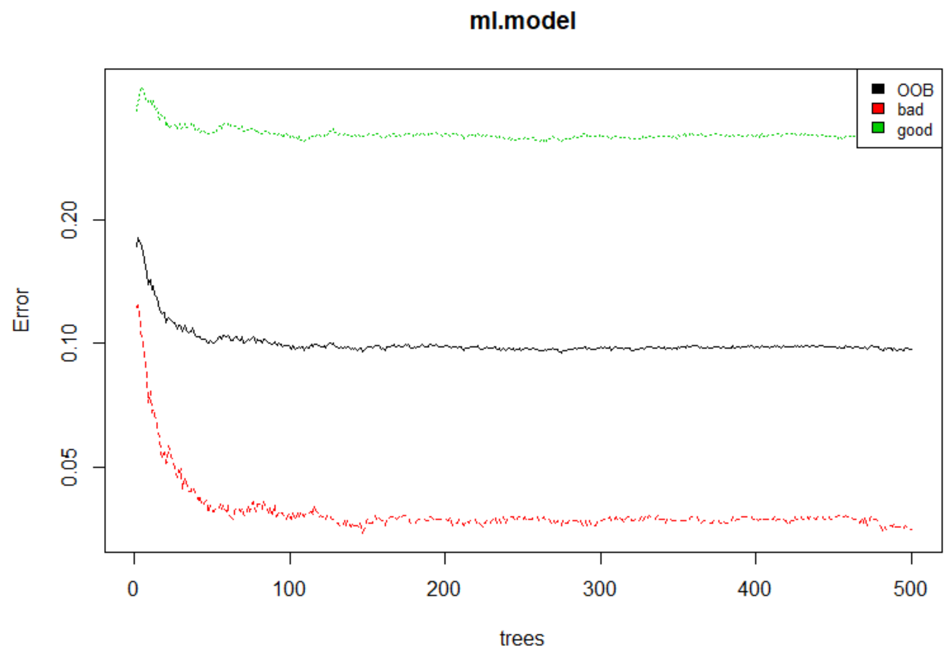
```
library(tidyverse)
library(randomForest)

url <- "https://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality-white.csv"
wine <- as.tibble(read.delim(url,header=TRUE,sep=";"))

> wine
# A tibble: 4,898 x 12
   fixed.acidity volatile.acidity citric.acid residual.sugar chlorides free.sulfur.diox~
<dbl>            <dbl>        <dbl>          <dbl>     <dbl>       <dbl>
17.00            0.270       0.360        20.7       0.0450       45
2 2              6.30        0.300        0.340      1.60         41

# ... with 4,888 more rows, and 5 more variables: density <dbl>, pH <dbl>, sulphates <dbl>,
alcohol <dbl>,quality <int>

> wine$taste <- as.factor( ifelse(wine$quality <= 6,'bad','good') )
```

Build the ML Model

```
> train <- wine[1:3000, ]
> test  <- wine[3001:4898, ]
> ml.model <- randomForest(taste ~ . - quality, data=train)
> print(ml.model)

Call:
 randomForest(formula = taste ~ . - quality, data = train)
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 3

        OOB estimate of  error rate: 9.7%
Confusion matrix:
      bad good class.error
bad  2276   83   0.0351844
good  208  433   0.3244930
```

Check Performance of ML Model

```
> pred <- predict(ml.model, newdata=test)
> table(pred, test$taste)

pred    bad good
  bad  1233  191
  good  246  228
> ml.model$importance
                     MeanDecreaseGini
fixed.acidity                64.47684
volatile.acidity             84.14794
citric.acid                  61.37304
residual.sugar               91.95227
chlorides                    86.69503
free.sulfur.dioxide          82.76855
total.sulfur.dioxide         80.33434
density                     125.57005
pH                           81.95264
sulphates                    78.01374
alcohol                     169.45978
```




```
plot(ml.model, log="y")
legend("topright",
      colnames(ml.model$err.rate),
      col=1:4,cex=0.8,fill=1:4)
```

Drowning in Data

Set up data **governance** processes. Bust silos, encourage collaboration, continue systems integration.

Invest in **cybersecurity**: physical, behavioral, and technical

Be selective: archive only most critical information, or “compute at the edges” & store aggregated values

Put **controls** in place to avoid the Data Swamp/Graveyard problem

Haven’t Started Yet

Digitize! Make data capture part of what’s expected in each work process

Practice good data hygiene: **one observation per row, one variable per column**

Keep learning key concepts, get your feet wet and play

Wait it out! Gartner says ~40% of data science tasks will be automated by 2020.


In a perfect world, my QMS would tell me…
From Kendall, K. (2017). The Increasing Importance of Risk Management in an Uncertain World.
The Journal for Quality and Participation, 40(1), 4.

INTELEX

Integrated Management Systems for **Environment, Health & Safety, and Quality Management (EHSQ)**

Established in 1992.

Worldwide client base.

More than 500 employees.

Over 1,200 global customers.

Over 1.5 million users.

Peer reviewed as best managed company.

Quality 4.0 is…

Connectedness

Intelligence

Automation

… for discovering insights into performance.

QUALITY 4.0 SUMMIT ON DISRUPTION, INNOVATION, AND CHANGE

Organizational Excellence in the Digital Age

November 12 – 13, 2018
Dallas, TX

Nicole Radziwill

Quality Practice Lead

nicole.radziwill@intelex.com

@nicoleradziwill

https://www.linkedin.com/in/nicoleradziwill/

INTELEX