The Framework for Data-Driven Operations

A Clear, Implementable Understanding

Richard G. Lamb, PE, CPA
Tel: 832-710-0755; Email: rchrd.lamb@gmail.com
Website (educational): https://analytics4strategy.com/

The first step toward becoming a data-driven operation is that its role holders must reach a clear, implementable understanding of data-drivenness and how to get there.

The purpose of the session is to be that first step.

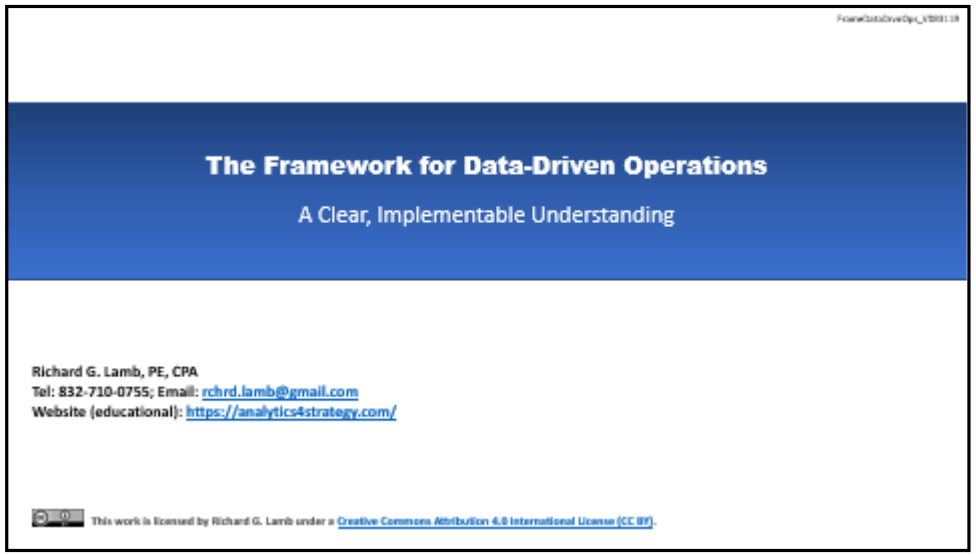
“Payload” approach to the session:

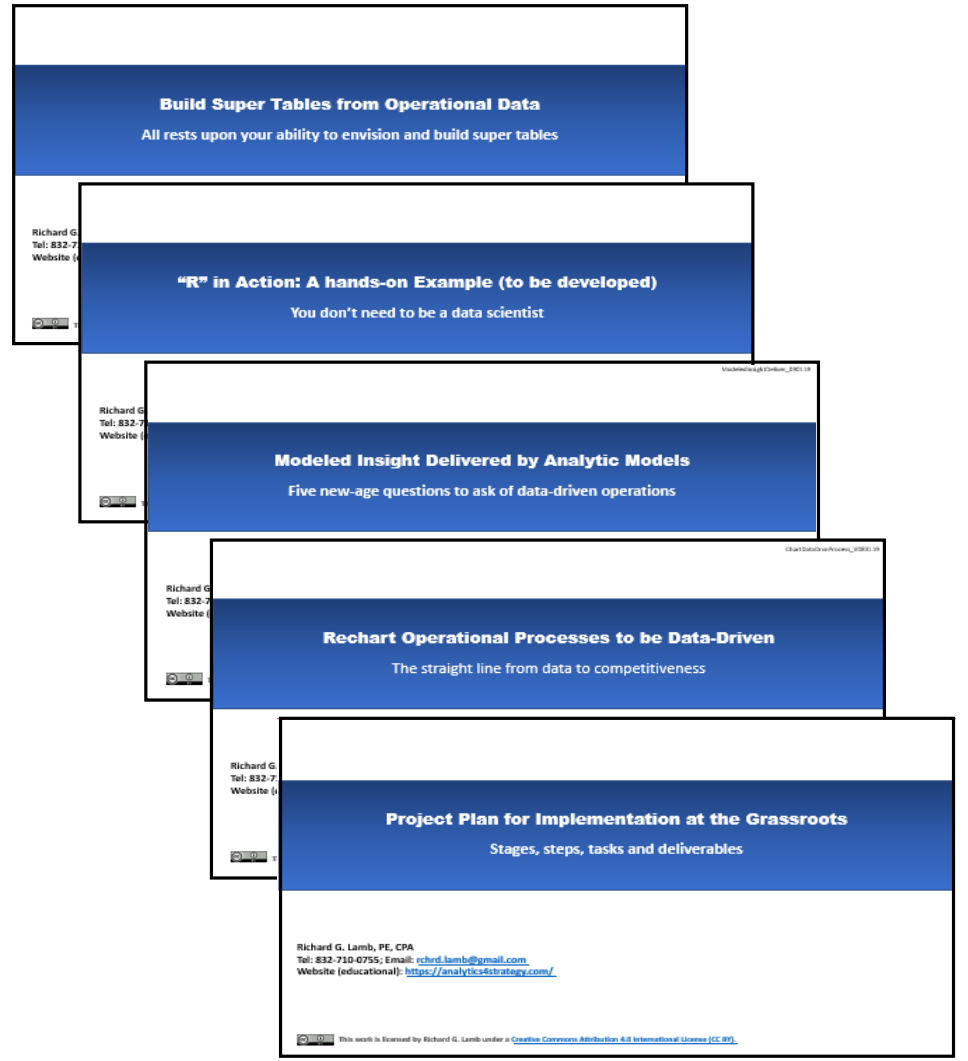
The presentation is formed as what I call a **“payload-style,”** defined as one formed to deliver, within limited allowed time, the full scope of information such that when you leave you will **“know what to do.”**

The criteria characteristics of a payload-type presentation are as follows:

- Introduces “all”—rather than “pieces”—that you **“must know that you must know”** for reaching data-drivenness.
- For every “must know,” map to “hands-on” instructions for self-directed learners.
- Non-expert attendees can, in turn, disperse the “know what to do” across their own organizations by emulating the presentation.

The collection of sessions are structured as a framework, expanded by specifics





All sets are available to download under the "Training Sessions" tab of the website https://analytics4strategy.com/

Agenda:

- The big picture of data-driven operations.
  - **Definition and depiction of a data-driven operation.**
  - Cost-free “Critical-Mass” strategy for reaching data-drivenness.
  - Essential definitions.
- Structure of methodologies.
  - “R”—as the analytic core of data-driven capability.
  - Gather and join subtables to form and cleanse super tables.
  - Layered charting in contrast to conventional charting.
  - Types of Insight Deliverables.
- Generalized implementation plan.
- Library of reference papers, presentations and texts.

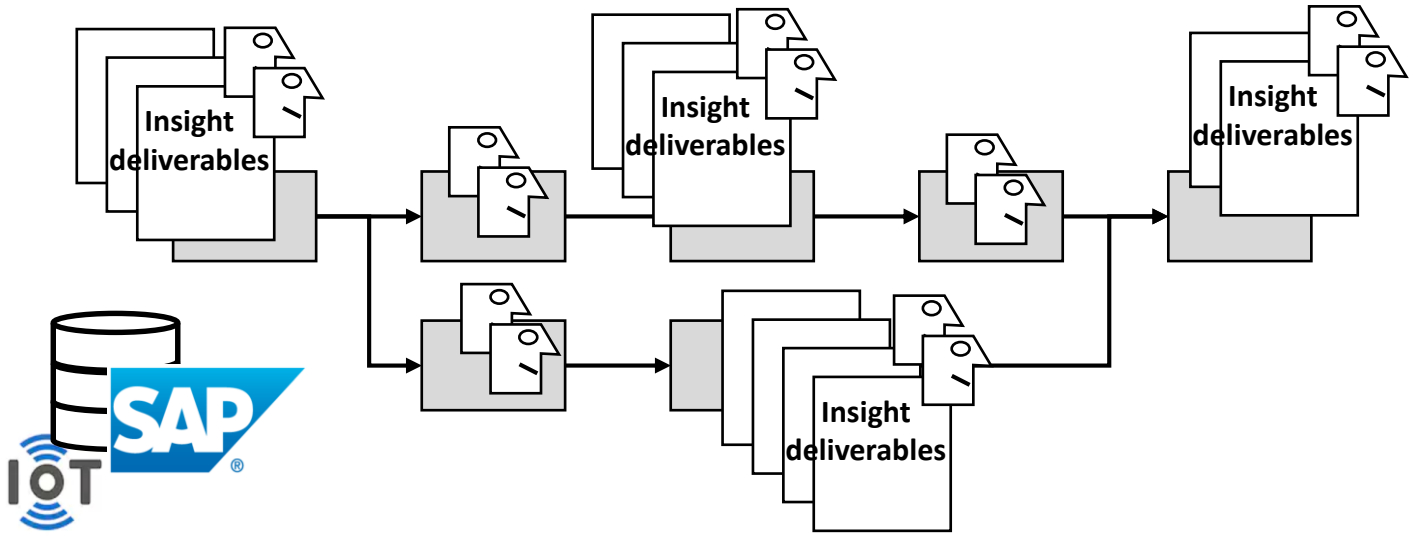
Data-driven defined:

A “Data-driven” operation is defined as one that harnesses its operational data to augment the experience and judgement of its operatives, managers, analysts and engineers as they plan, organize, conduct and control their processes.

What that looks like: An organization simply **improves** its processes to include all augmenting **"insight deliverables"** that will make a difference

At some places along any process, the "best outcomes" can only be realized when experience and judgement are augmented with "insight deliverables."

At each such place, a suite of system reports, tables, charts and models is recognized, built and worked to realize the best of outcomes.



The previous slide depicted a data-driven operational process. However, what is depicted is the end-point of a derivation top-down from how the firm competes and wins.

The top-down progression is demonstrated in the context of maintenance operation by the training slides titled, "Rechart Operational Processes to be Data-Driven," at https://analytics4strategy.com/trn-datadrvnmntcop

Agenda:

- The big picture of data-driven operations.
  - Definition and depiction of a data-driven operation.
  - **Cost-free "Critical-Mass" strategy for reaching data-drivenness.**
  - Essential definitions.
- Structure of methodologies.
  - "R"—as the analytic core of data-driven capability.
  - Gather and join subtables to form and cleanse super tables.
  - Layered charting in contrast to conventional charting.
  - Types of Insight Deliverables.
- Generalized implementation plan.
- Library of reference papers, presentations and texts.

Improvements can be **“grassroots”** because **“critical-mass”** for almost all of data-drivenness is not high-tech or new-tech, but modern-day knowledge, skills and software

- **“Critical-mass”** is defined as. . .

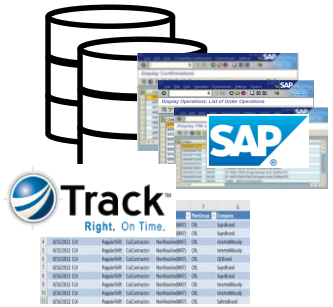
  The threshold set of knowledge, skills and software that must be in place to be fully, effectively and efficiently data-driven.

- Characteristics of critical-mass:

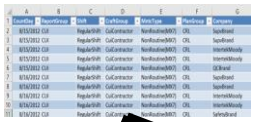
  - Knowledge and skills travel to up-teched and up-scaled strategies.
  - Up-teching from critical-mass will not practicably increase the power of the insight that is extracted from the operation’s data.

Action at the "Grassroots" is possible because there is a triad of software that we already have "use of" or "rights to"

Data in multiple operating systems and possibly multiple data bases.



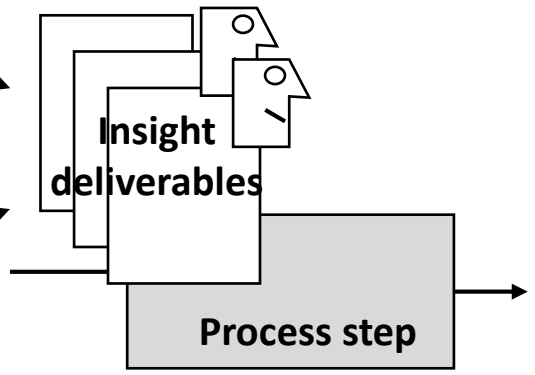
Data from process tasks conducted with Excel.



Build super tables from sourced data.

- Pivot as dashboards.
- Tables and calculations.
- Conventional charts.

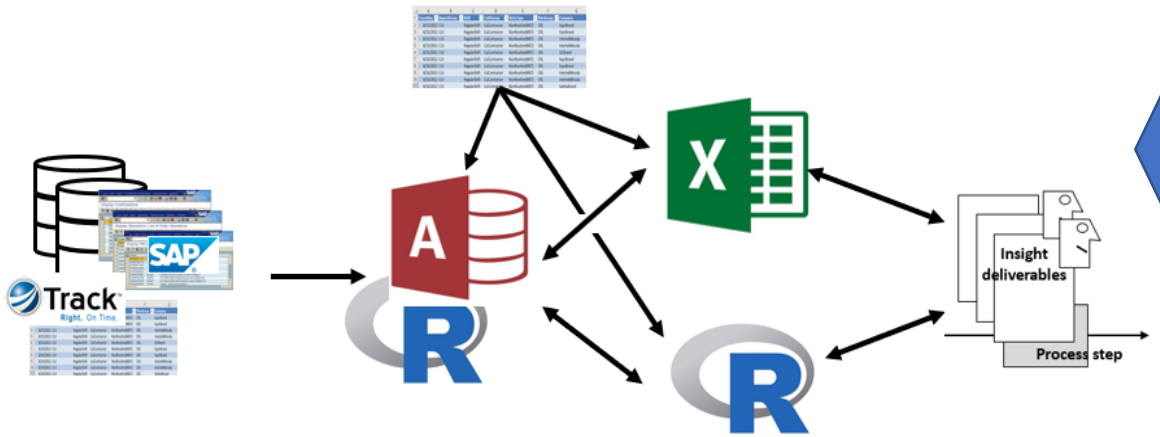
Experience and judgement augmented with a full range of relevant insight deliverables.



- Descriptive statistics.
- Layered charts with ggplot2.
- Machine learning and artificial intelligence based analytics.

Note:
For each of the triad, there are many commercial alternatives.

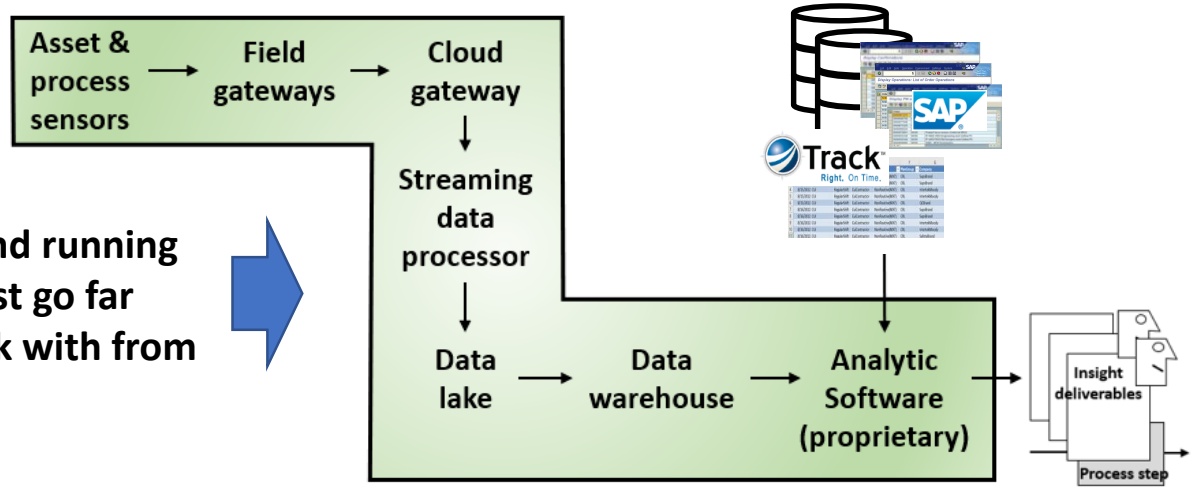
We can see why there can be action at the grassroots by comparing the **triad** to **IIoT-based** solutions for reliability and maintenance as alternative trails to deliver insight deliverables



The grassroot triad entails no infrastructure beyond what supports normal operations in any organization.

Over 90 percent of insight deliverables we can think of are possible through the triad.

IIoT-based solutions require deep capital and running cost to build and support a system that must go far beyond making use of what there is to work with from normal operations.



Condition-based maintenance (CBM) and IIoT—based CBM must not be mistaken as the whole of data-driven operations

- **Condition-based maintenance (CBM)** is a taking data through mobile and fixed sensors, and analyticis revealing changes to an asset's performance upstream (potential failure) to the functional failure we are watching for.
- **IIoT-based CBM** automates or makes continual the monitoring, data processing and insight—requiring a great deal of infrastructure.
- The design issue is whether IIoT-based solution is "worth it" for a particular case of CBM that, without consideration of IIoT, has already been found to be technically feasible and worth it?
- The reality is that for non-CBM insight, almost all modeled operational insights are technically feasible, but does not require an IIoT-based solution to get at it.

Note:
There is a distinction between on-condition and condition-based as one of four ways to seek potential failures to a functional failures.

Agenda:

- The big picture of data-driven operations.
  - Definition and depiction of a data-driven operation.
  - Cost-free “Critical-Mass” strategy for reaching data-drivenness.
  - **Essential definitions.**
- Structure of methodologies.
  - “R”—as the analytic core of data-driven capability.
  - Gather and join subtables to form and cleanse super tables.
  - Layered charting in contrast to conventional charting.
  - Types of Insight Deliverables.
- Generalized implementation plan.
- Library of reference papers, presentations and texts.

Data and **big-data** are distinctively different with respect to necessity, technology and organizational abilities

- We tend to think of “big data” in a colloquial sense—thousands or millions of rows seem “BIG” compared to our personal experience with Excel worksheets.

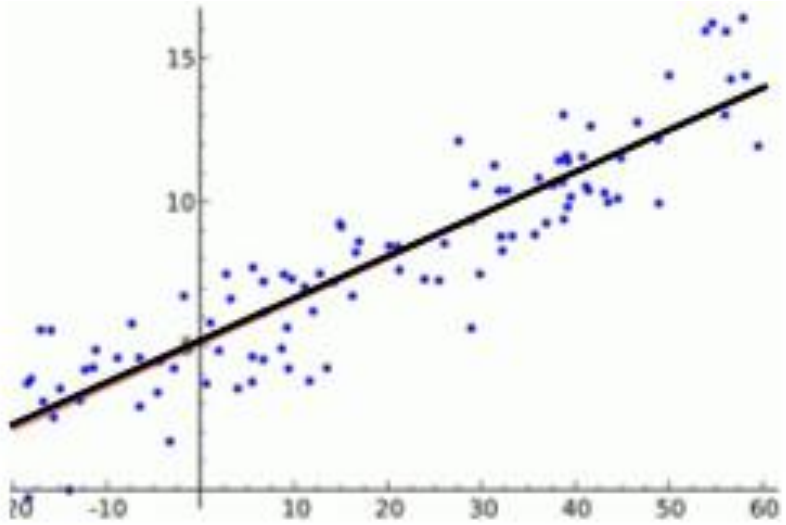
- “Big data” is the case in which data are massive or unstructured data (streaming sensor data, e-mail, document, video, photo, audio, webpage).
  - Litmus test—the data or analytics of the data cannot be processed on our notebook computers.
  - Big data entails high-tech systems, specialized skills and substantial organizational costs.

- However, the data analytics conducted in either arena are the same.

- The data of operational processes rarely entail big data, whereas, LLoT-based solutions typically do.

We need to set some clarity on **machine learning**, **artificial intelligence** and **algorithms**—using regression analysis as a frame of reference

- **Machine learning** takes place when we feed the data of variables to the set-up regression and its gut **algorithm** conducts a trial-and-error calculation until "learning" the best fit.
- Most often our interest is with the returned coefficients for each variable and other inferences telling us how much, if any, the variable plays in predicting the outcome.
- In contrast, **Artificial intelligence** feeds new data to the fitted model of coefficients to predict or forecast outcomes upon the "learned" coefficients.
- AI does not distinguish the model—all model types entail machine learning and most can be deployed as AI.



Until recently we just called everything a model—and still should.

When the model is to extend to be used as AI, the learning process will entail an additional stage

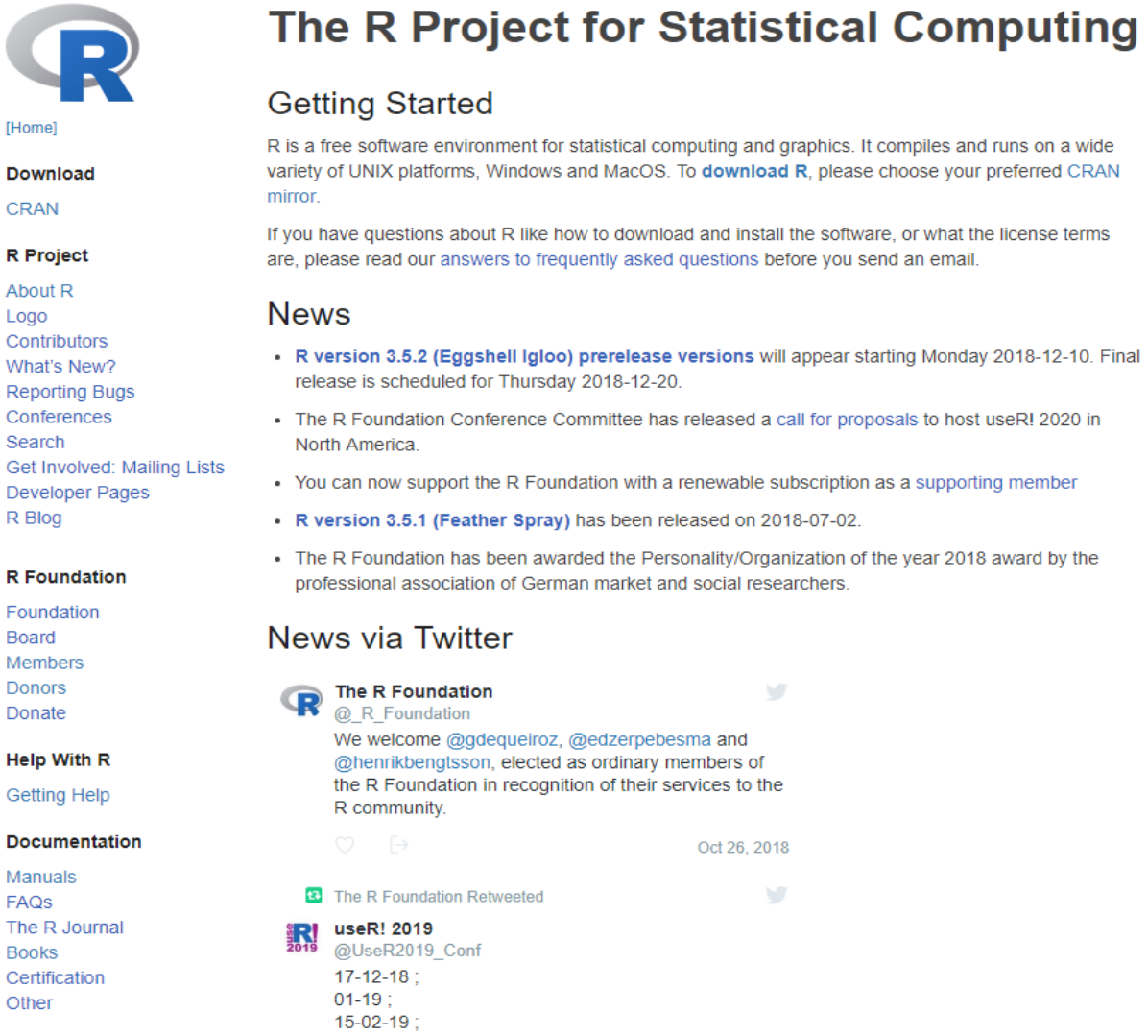
- A portion of the original data set is held out from building the model—e.g., one third—to be a test set.
- The remaining portion is fed to the model for learning.
- The test set is fed to the learned model to evaluate for how accurately the "trained" model estimates or calls the actual outcome of each case in the test set.
- If accuracy is acceptable to the intended use, model is deployed to make the learned judgement—augment human judgement.
  - Greater than 85 percent is considered as acceptable.
  - Which is why we must always think in terms of "augment" not "supplant" experience and judgement.

Agenda:

- The big picture of data-driven operations.
  - Definition and depiction of a data-driven operation.
  - Cost-free “Critical-Mass” strategy for reaching data-drivenness.
  - Essential definitions.
- Structure of methodologies.
  - **“R”—as the analytic core of data-driven capability.**
  - Gather and join subtables to form and cleanse super tables.
  - Layered charting in contrast to conventional charting.
  - Types of Insight Deliverables.
- Generalized implementation plan.
- Library of reference papers, presentations and texts.

“R” is available to download at https://www.r-project.org/ along with a manual in coding, instructions for download and more

YouTube video demonstrating how to download and install R on your computer. https://www.youtube.com/watch?v=ym8szN2Zim4



Aspects of “R” that make it critical-mass to achieving data-drivenness

- Only through analytics can there be descriptive statistics to our data, layered charting, maximal data cleansing and analytic model.
- For almost every imaginable analytic, there is a package of functions with arguments.
- All packages are accompanied with a full explanation and examples with data.
- Online support is highly evolved and vast—allowing google, copy and paste.
- Texts on data and analytics are plentiful in which “R” is used—relieving you to understand the analytic.

We do not program (we could), but select functions and set arguments to reflect the nature of the insight deliverable—just as we do in Excel

Example: In the "stat" package, lm() is a function—linear regression—with its arguments we would variously include and set to fit the analytic case.

```
lm(formula, data, subset, weights, na.action, method = "qr",
   model = TRUE, x = FALSE, y = FALSE, qr = TRUE,
   singular.ok = TRUE, contrasts = NULL, offset, ...)
```

"Never memorize something that you can look up."

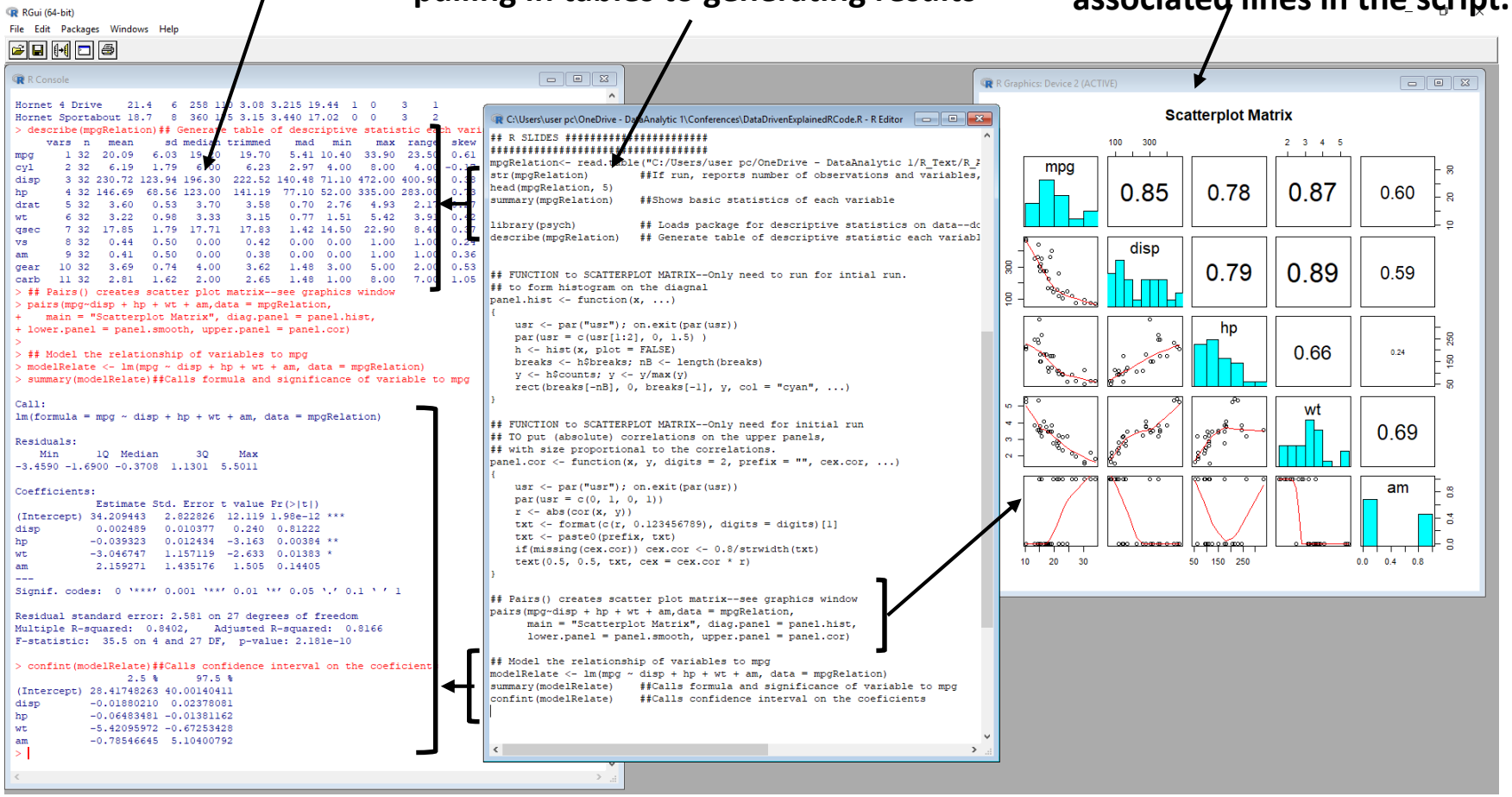
Albert Einstein

What it looks like, the frontend view of the “R” software

Textual results from running entire script or specific lines

Script that runs the process—from pulling in tables to generating results

Graphic results from running associated lines in the script.



How-to guidance for “R”

Knowledge and skills	Texts, papers and session slides
Download software	▪ **Website: Download at** https://r-project.org ▪ **YouTube: How to install** https://www.youtube.com/watch?v=ym8szN2Zim4
Coding—general	▪ **R for Data Science, Grolemund, Wickham, 2017. Free E-Book** https://r4ds.had.co.nz/index.html ▪ **R for Dummies, de Vries, Meys, 2015.** ▪ **Art of Programing R, Matloff, 2011.**
Coding in context of conducting analytics (1)	▪ **Discovering Statistics Using R, Field and Miles, 2012** ▪ **Multilevel Modeling Using R, Holmes, 2014** ▪ **Introductory Time Series with R, Cowpertwait and Metcalfe, 2009** ▪ **Event History Analytics with R, Bostrom, 2012** ▪ **Machine Learning with R, Lantz, 2015**

(1)
“Coding in context” is defined as, while explaining and building specific analytics, “R” coding to set up the analytic is explained in parallel.

The purpose of the section was to introduce “R” within the framework of data-driven operations.

A hands-on explanation of “R” as the training slides titled, “R in Action; Hands-On,” is available to download at https://analytics4strategy.com/trn-rhandson

Agenda:

- The big picture of data-driven operations.
  - Definition and depiction of a data-driven operation.
  - Cost-free “Critical-Mass” strategy for reaching data-drivenness.
  - Essential definitions.
- Structure of methodologies.
  - “R”—as the analytic core of data-driven capability.
  - **Gather and join subtables to form and cleanse super tables.**
  - Layered charting in contrast to conventional charting.
  - Types of Insight Deliverables.
- Generalized implementation plan.
- Library of reference papers, presentations and texts.

This section is an overview of what super tables are and how they are built within the framework of data-driven operations.

A deep-dive explanation of building super tables is available as the training slides titled, “Build Super Tables from Operational Data,” at https://analytics4strategy.com/train-builddatatables

There are three enabling realities to building super tables

- **Almost all operating systems allow their data to be extracted in table format as a standard report.**

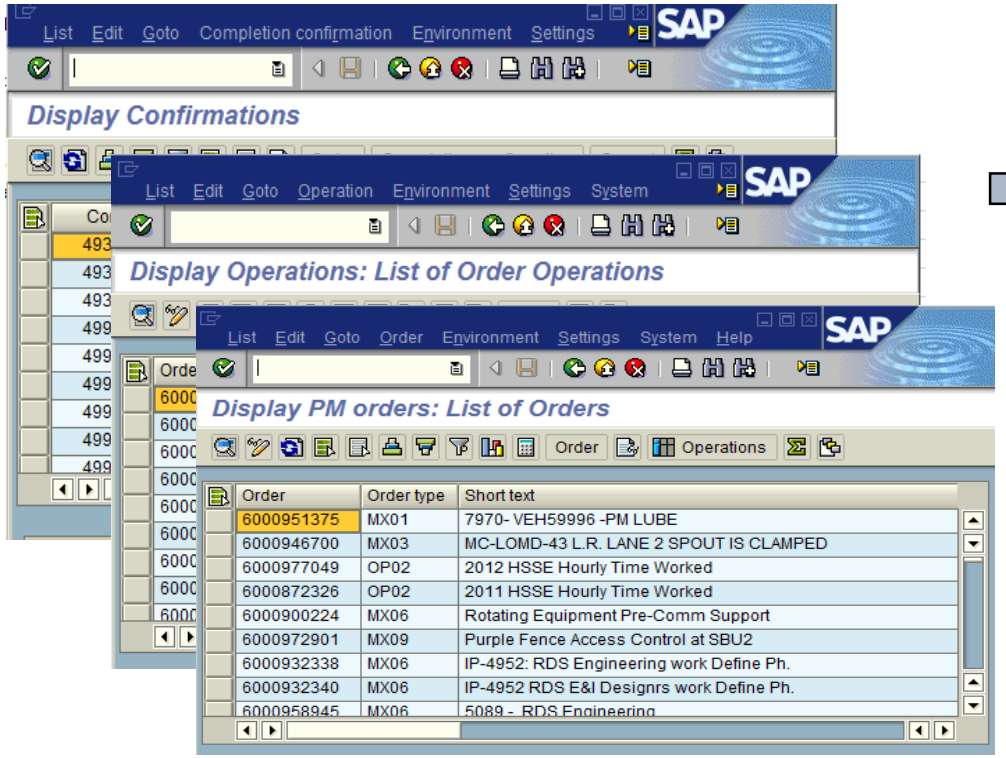
  When not, the IT data specialists can give us an on-demand tool to do so.

- **Individual data tables from any one or more systems or sources can be joined into one table by any identifying variable they have in common.**

  Only the data type (e.g., numeric, character) must match or be transformed to match.

- **Bad data is rarely a deal killer: Methods to "Cleanse" the data usually neutralize the flaws.**

In a nutshell: We extract topic-specific data from sources and fabricate a **super table** as required to build one or more intended insight deliverables

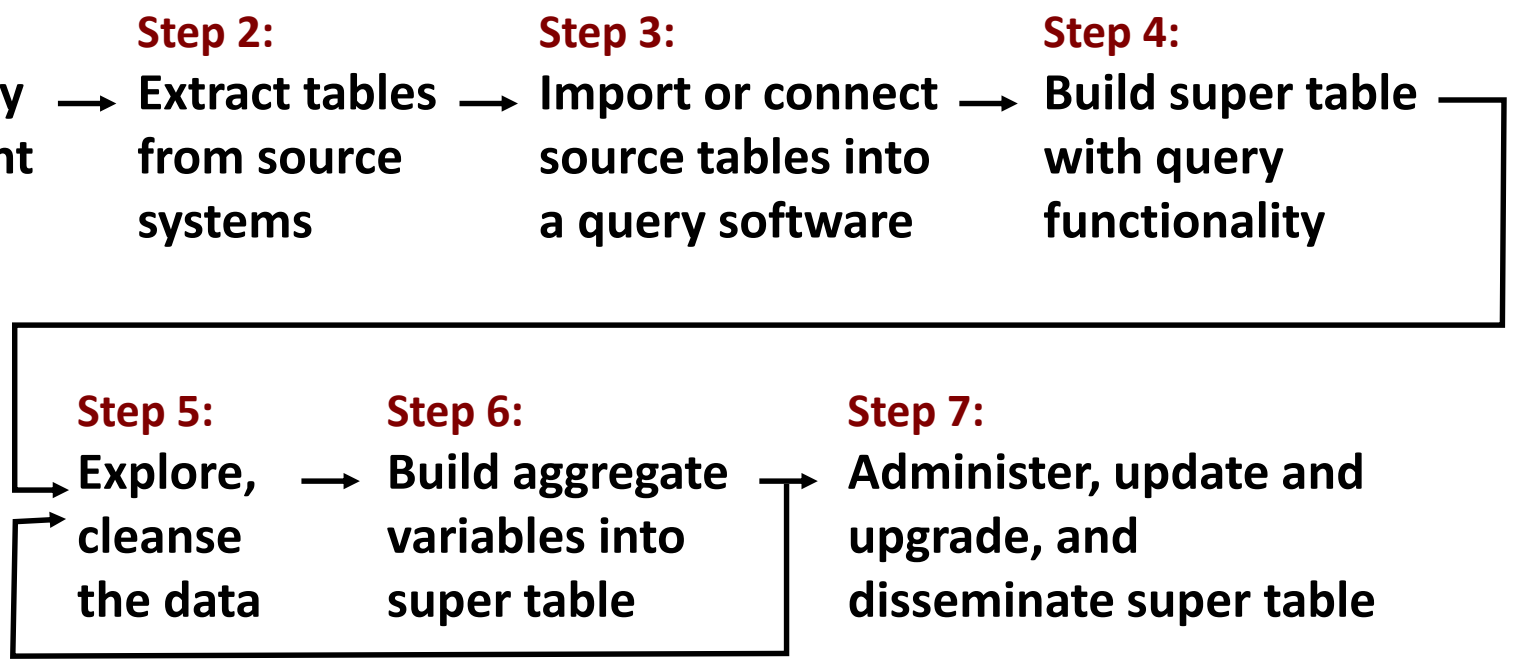


No one table has all needed variables to the envisioned insight deliverables.

Cost center	OrderNoText	StepNoText	MntcType	CraftTy
70160	6000707049: MA-DCU-PU8818 Install max impeller & 15h	180; DCU PU8818-JSA & INSTALL PUMP	Proactive	Machini
70160	6000707049: MA-DCU-PU8818 Install max impeller & 15h	30; DCU PU8818-JSA & LO/TO MOTOR	Proactive	Electrici
70160	6000707049: MA-DCU-PU8818 Install max impeller & 15h	60; DCU PU8818-OPERATION TO ENERGIZE MOTOR	Proactive	Machini
70160	6000707049: MA-DCU-PU8818 Install max impeller & 15h	80; DCU PU8818-LO/TO MOTOR	Proactive	Electrici
70160	6000707049: MA-DCU-PU8818 Install max impeller & 15h	80; DCU PU8818-LO/TO MOTOR	Proactive	Electrici
70160	6000812732: MC-DCU-Pull/Repair Dump Reg. on Jet Pump	40; DCU-Repair Dump Reg-INSTALL	Reactive	MultCra
70160	6000812732: MC-DCU-Pull/Repair Dump Reg. on Jet Pump	50; DCU-Repair Dump Reg-RECONNECT	Reactive	Instrum
70160	6000860441: MC-buff Tk1830 to add nozzles	27; DCU-TK1830-CENTER PUNCH AND BUFF AREAS O	Proactive	MultCra
70160	6000860441: MC-buff Tk1830 to add nozzles	27; DCU-TK1830-CENTER PUNCH AND BUFF AREAS O	Proactive	MultCra
70160	6000915285: MC-DCU-Bridge Crane AC unit installation	70; Crane to assist Electricians	Reactive	Electrici
70160	6000915285: MC-DCU-Bridge Crane AC unit installation	70; Crane to assist Electricians	Reactive	Electrici
70160	6000915285: MC-DCU-Bridge Crane AC unit installation	90; Motiva Inspector	Reactive	Electrici
70160	6000926113: EL-DCU-MOV open/close switch replacement	70; EL-DCU-MOV open/close switch replacement	Reactive	Electrici
70160	6000926113: EL-DCU-MOV open/close switch replacement	70; EL-DCU-MOV open/close switch replacement	Reactive	Electrici
70160	6000929188: IM-DCU-35304 tensionometer no indication	20; M-DCU-35304 tensionometer no indication	Reactive	Instrum
70160	6000929188: IM-DCU-35304 tensionometer no indication	20; M-DCU-35304 tensionometer no indication	Reactive	Instrum
70160	6000937432: MA-DCU-Pu8871seal leaking	130; DCU PU8871- INSTALL PUMP	Reactive	Machini
70160	6000937432: MA-DCU-Pu8871seal leaking	130; DCU PU8871- INSTALL PUMP	Reactive	Machini

- The "**super table**" does not, cannot and never will exist in any one operating system.
- Building the super table in Excel is too laborious to be practical.

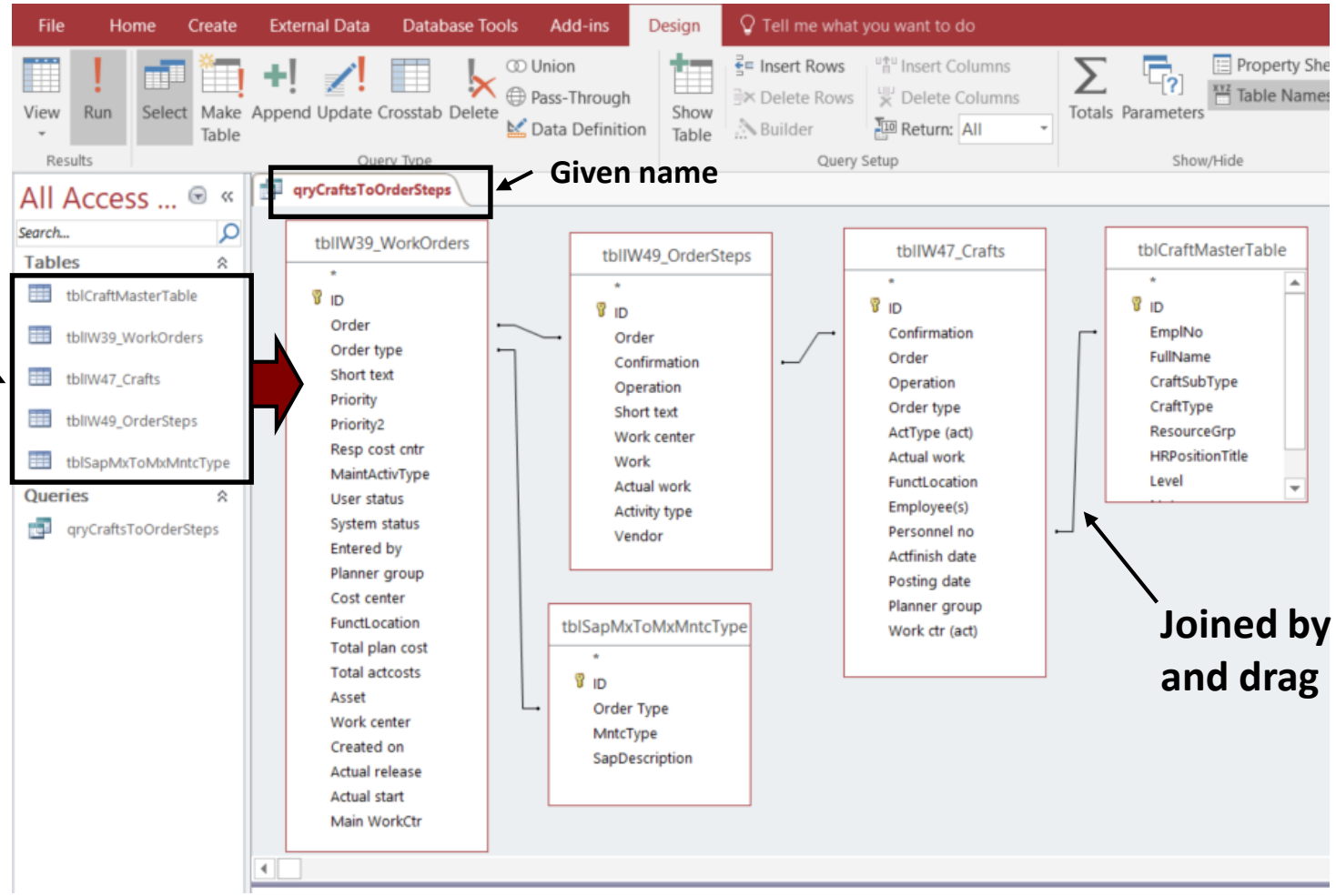
Building a super table from sub-tables follows a standard path



MS Access, of the triad, will be used to overview the charted process

With a query, the subtables are joined by the variables they have in common; creating a huge table with all the shown imported variables

Tables pulled into access, but could have been connected to source.

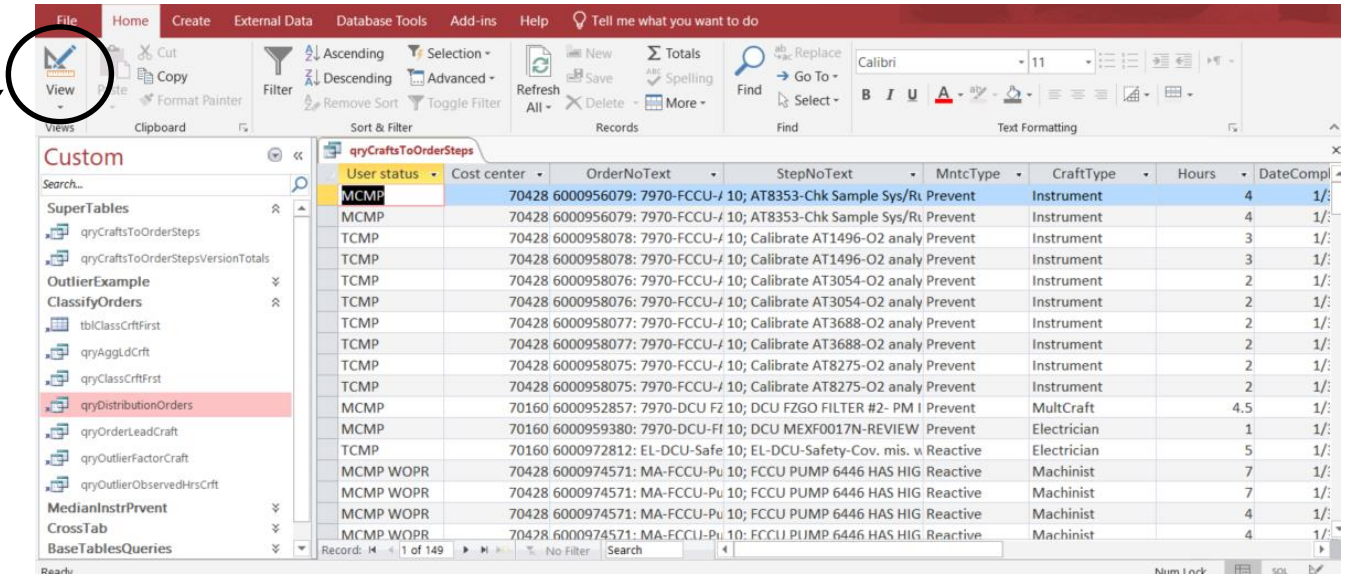


The super table is built in the design grid and along the way flipping back and forth between the “Design” and “Table” view to confirm we are getting what is intended and for discovery

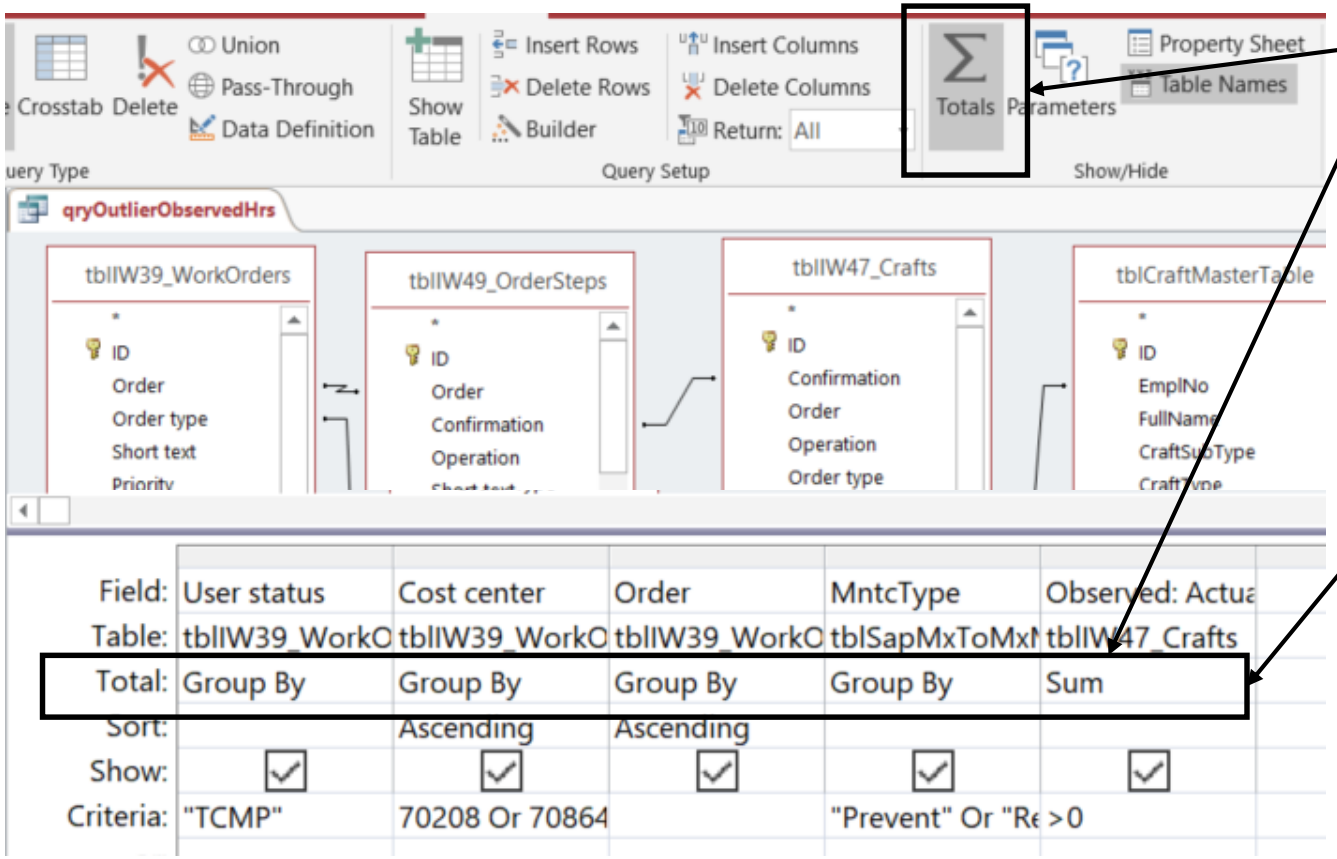
Created on
Actual release
Actual start
SapDescription

Field	User status	Cost center	OrderNoText: [tl	StepNoText: [tb	MntcType	CraftType	Hours: Actual w	DateComplete: /	DaysAftrCreatd:
Table	tblIW39_WorkO	tblIW39_WorkO			tblSapMxToMx	tblCraftMasterT	tblIW47_Crafts	tblIW47_Crafts	
Sort				Ascending					
Show	☑	☑	☑	☑	☑	☑	☑	☑	☑
Criteria		70208 Or 70864			"Prevent" Or "Pr			Between #1/1/2	
or:									

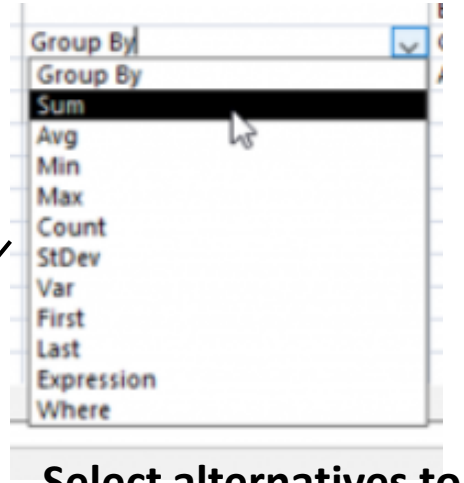
View icon



A powerful functionality is to build aggregations into the supper table to create calculated variables upon grouped summaries such as counts, sums, averages, spread, extremes and first-last

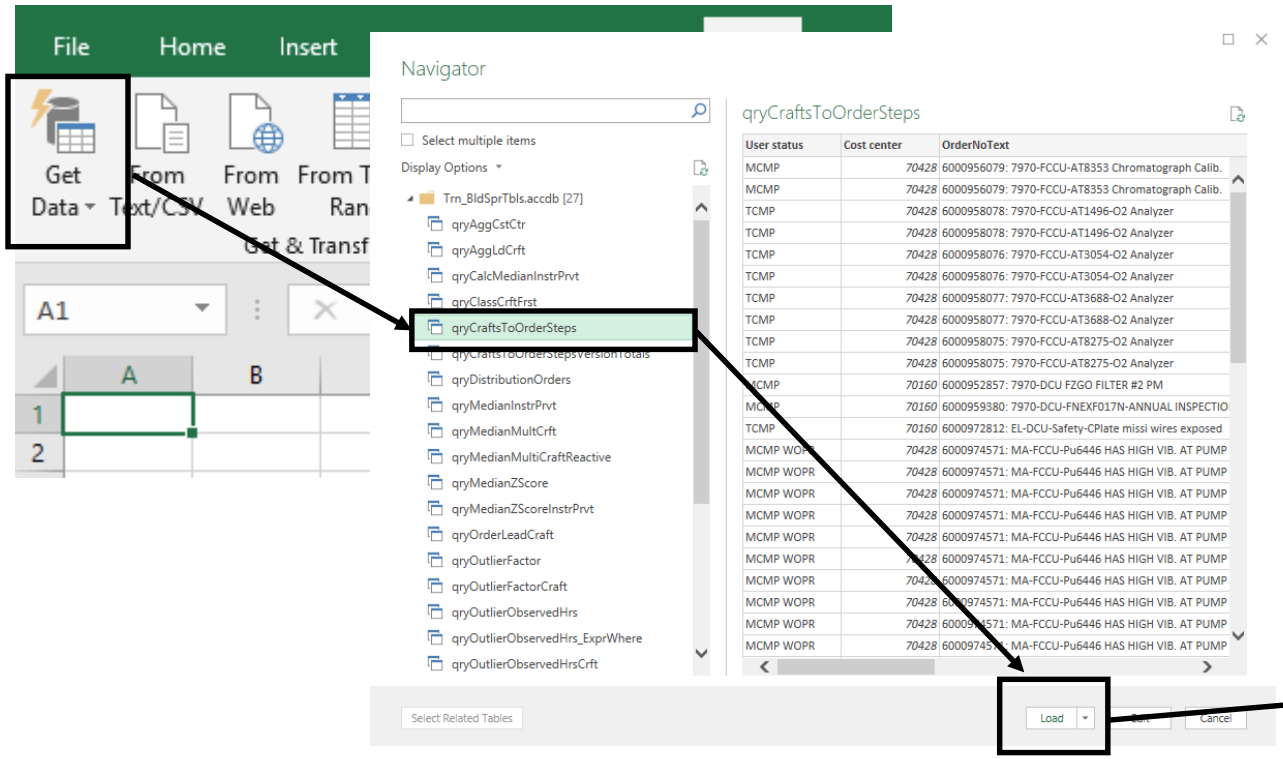


Clicking the icon adds the "Total" row of the query grid.

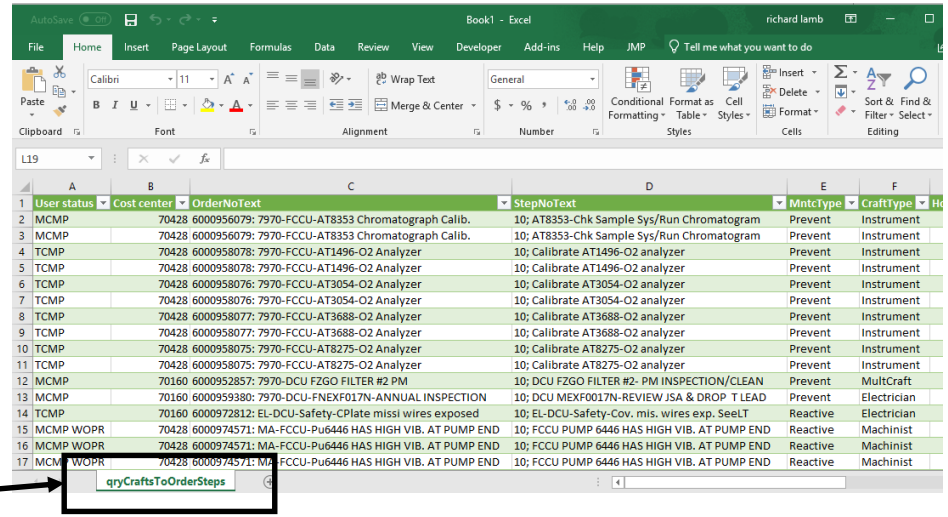


Select alternatives to the default "Group By"

Bring the query into Excel Table or Pivot by clicking the Excel “Get Data” button and following the path to select the query from Access—regardless of destination software there is a path



The super table will appear in Excel for immediate use.

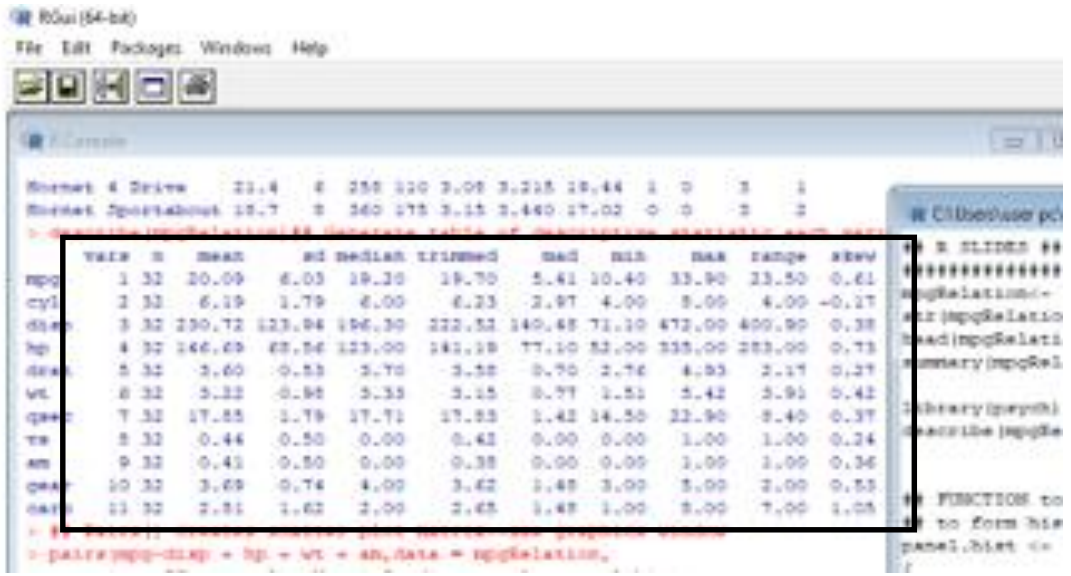


Note: **Could have connected to the super table rather than extracted, allowing refresh functionality.**

Get data >> From Data Base >> From Microsoft Access Database >> select File >> Select table from list >> click Load

The partnership of Access and “R”

- At times we must take data beyond what SQL, thus, Access can do.
- When more is needed the Access-built table or raw data can be pulled into “R” and powered up.
- Specialized tables that generate in the process of running an analytic in “R” can be exported to Access and joined with super tables.

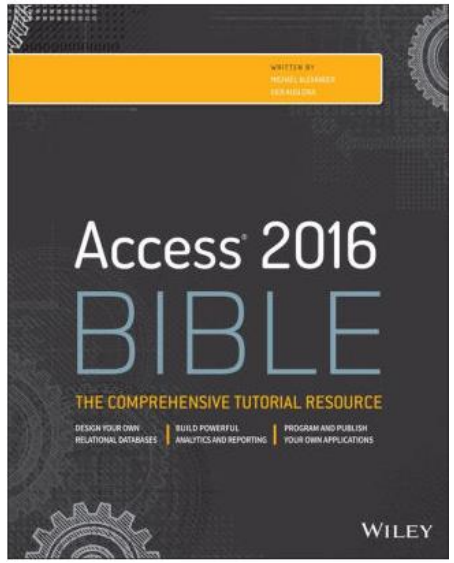


There are five types of bad data—the good news is that there are methods to deal with each

Type	Strategy
Duplicate cases	Seek cases with duplicate query—see YouTube explanation (https://www.youtube.com/watch?v=DPvJOWv6Ntc).
Empty cells Misclassifications	▪ Use analytic models applied to good cases to predict or classify what should-be versus what-is. ▪ Likely models are regressions (linear, logistic, Poisson), trees and K-Means. (1)
Misformatted	▪ Build translation tables for each bad-data case to a variable. ▪ Attach to super tables and use translated, rather than source dirty variable.
Outliers (numerical)	▪ Use aggregate functionality to build an outlier test variable into the super table. ▪ Utilize regressions to identify cases that would not have been predicted or have excessive influence on the model.

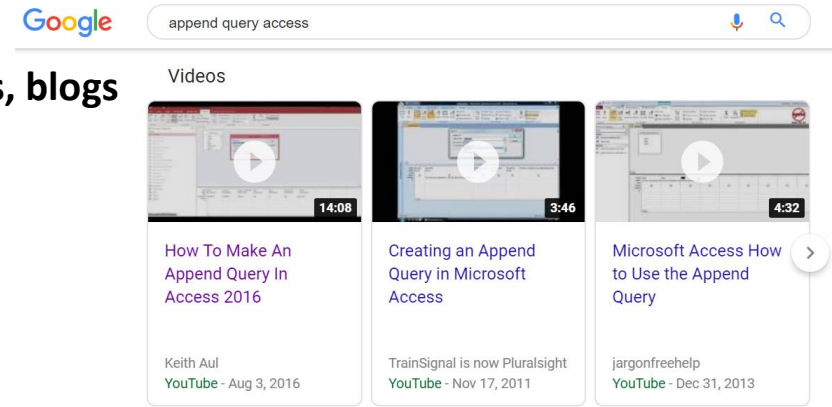
(1) See papers, "Find What Matters With Relationship Questions Of Operations" (https://analytics4strategy.com/relatqstoci) and "Dive Below The Surface With Apparency Questions" (https://analytics4strategy.com/apprqsblwfnctng) .

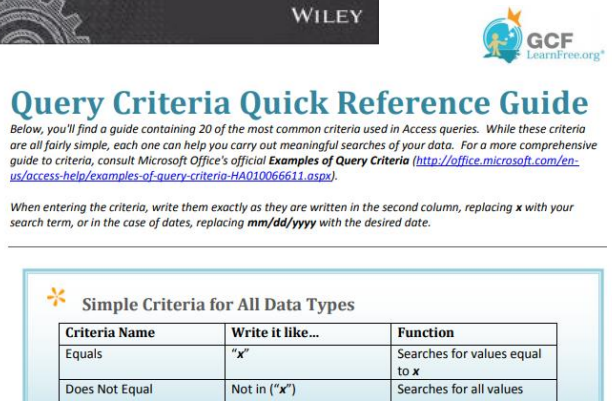
Another advantage of Access is that there is an immense support community



Chapters 8 through 13 explain everything there is to know about building super tables.

On line, there are YouTube videos, blogs and articles to everything Access.




GCF LearnFree.org

Query Criteria Quick Reference Guide

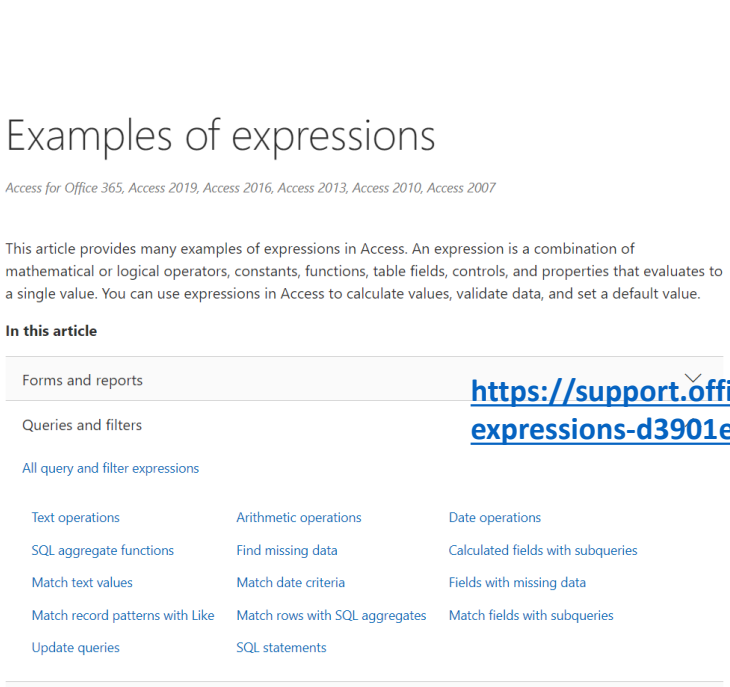
Below, you'll find a guide containing 20 of the most common criteria used in Access queries. While these criteria are all fairly simple, each one can help you carry out meaningful searches of your data. For a more comprehensive guide to criteria, consult Microsoft Office's official ***Examples of Query Criteria*** *(http://office.microsoft.com/en-us/access-help/examples-of-query-criteria-HA010066611.aspx).*

When entering the criteria, write them exactly as they are written in the second column, replacing ***x*** *with your search term, or in the case of dates, replacing* ***mm/dd/yyyy*** *with the desired date.*

Simple Criteria for All Data Types

Criteria Name	Write it like...	Function
Equals	"x"	Searches for values equal to x
Does Not Equal	Not in ("x")	Searches for all values

https://media.gcflearnfree.org/ctassets/topics/177/GCFAccessCriteriaGuide.pdf



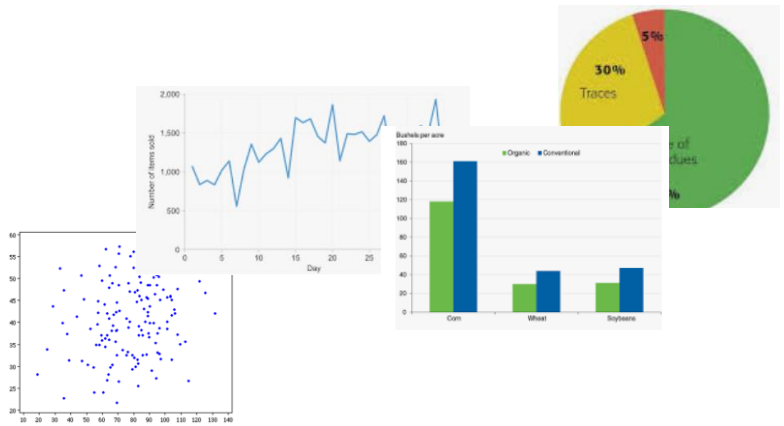
https://support.office.com/en-us/article/examples-of-expressions-d3901e11-c04e-4649-b40b-8b6ec5aed41f

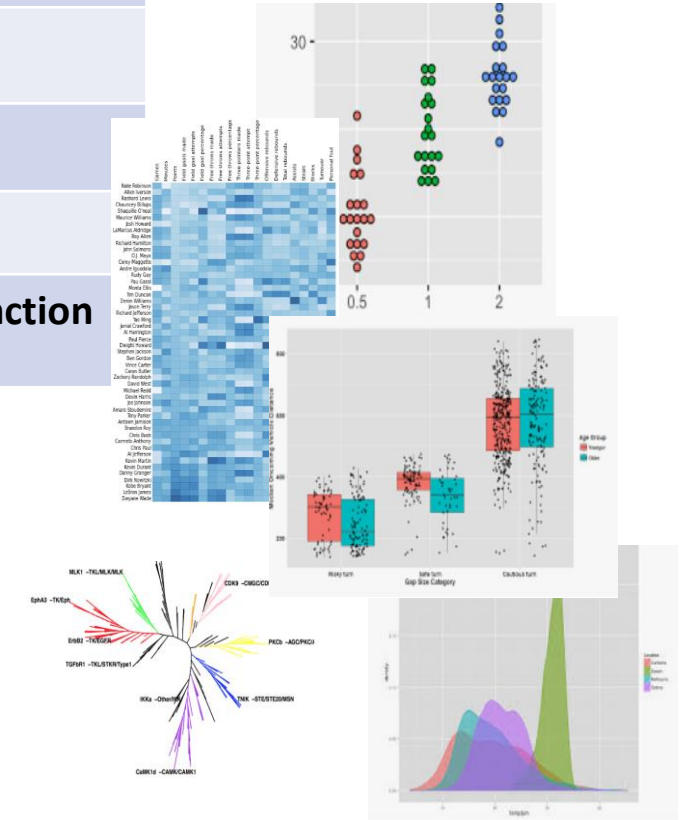
Agenda:

- The big picture of data-driven operations.
  - Definition and depiction of a data-driven operation.
  - Cost-free “Critical-Mass” strategy for reaching data-drivenness.
  - Essential definitions.
- Structure of methodologies.
  - “R”—as the analytic core of data-driven capability.
  - Gather and join subtables to form and cleanse super tables.
  - **Layered charting in contrast to conventional charting.**
  - Types of Insight deliverables.
- Generalized implementation plan.
- Library of reference papers, presentations and texts.

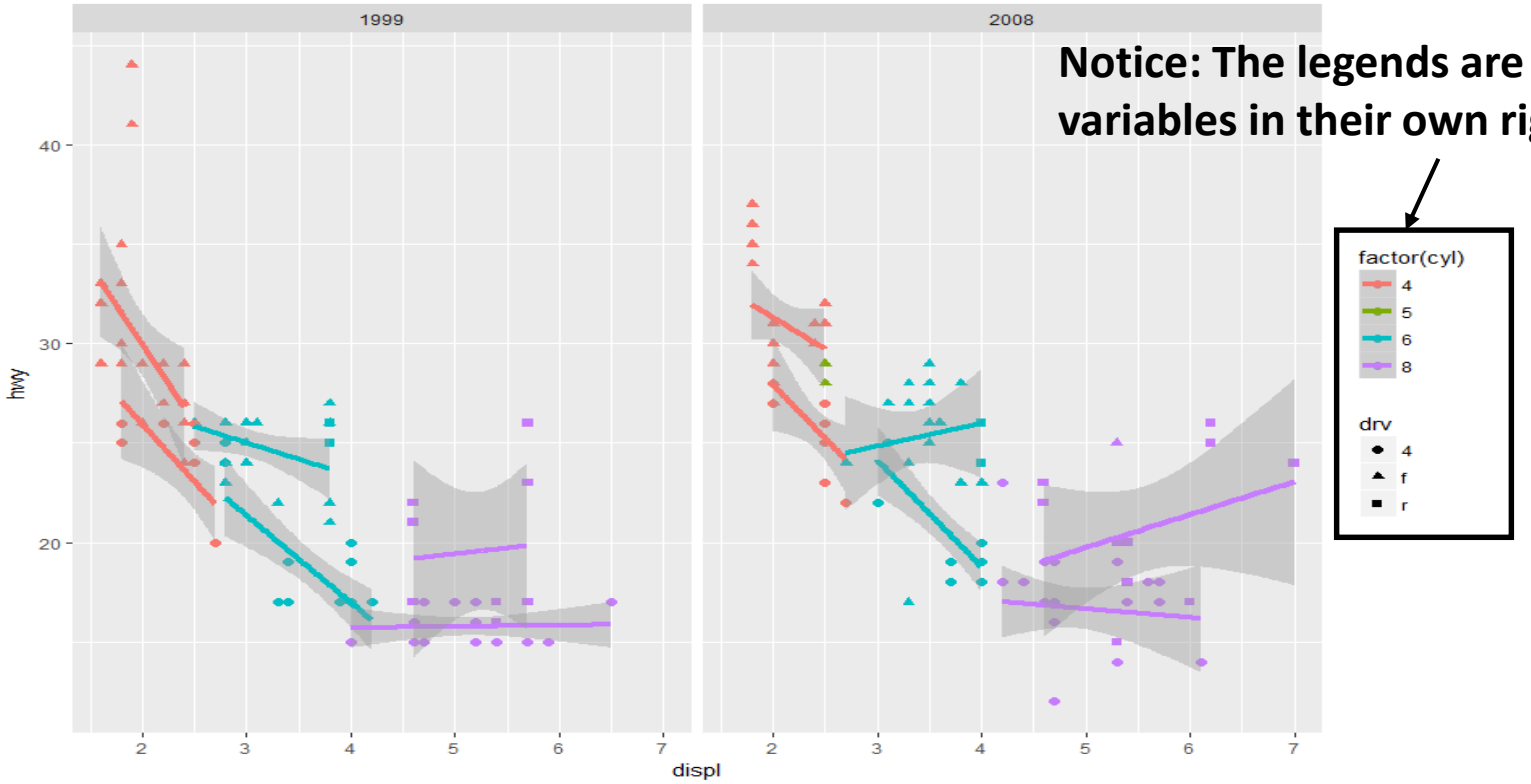
Conventional charts have been the staple since they were invented as early as the 1600s—now layered charting with ggplot2 in R allows perspectives we could not have before

Characteristic	Conventional	Layered
Types	limited	Unlimited
Number of variables	Two	Unlimited
Number of perspectives	One	Unlimited
Legends	Upon the variable	Can be variables
Data points	Lose distinction if many	Methods to create distinction





With this example you can begin to imagine how you would explore the interplay of KPIs—to name one idea

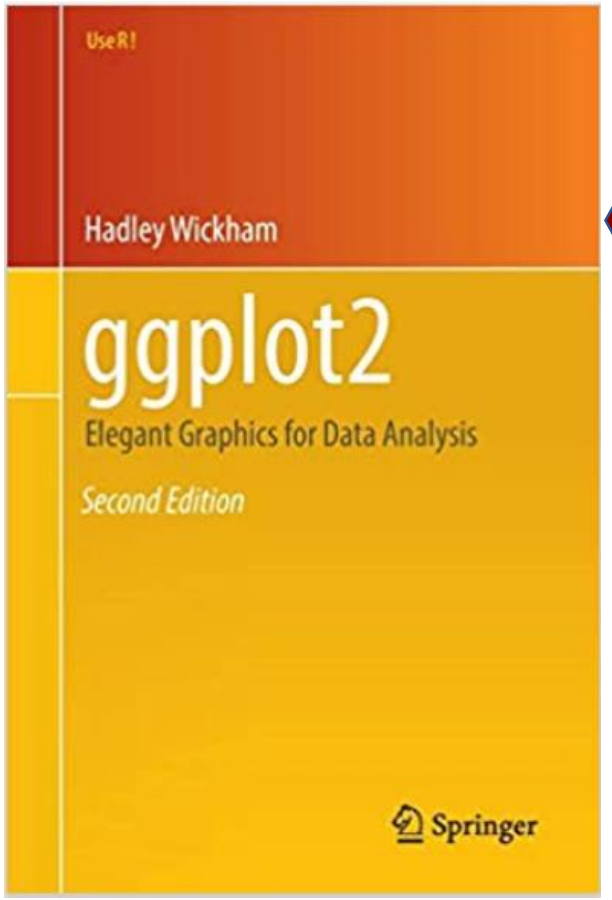


Notice: The legends are variables in their own right.

In this case:

- **Interplay of five variables: highway mileage, displacement, cylinders, drive and years.**
- **Three perspectives are layered on the five variables: cross plot, linear fit and confidence intervals.**

You likely have some ideas for where layered charting could play big in the work tasks you are responsible for—this will get you there



Work through the first eight chapters while imagining what you could do with what is being explained and shown along the way—thence, use the text as a cookbook.

Agenda:

- The big picture of data-driven operations.
  - Definition and depiction of a data-driven operation.
  - Cost-free “Critical-Mass” strategy for reaching data-drivenness.
  - Essential definitions.
- Structure of methodologies.
  - “R”—as the analytic core of data-driven capability.
  - Gather and join subtables to form and cleanse super tables.
  - Layered charting in contrast to conventional charting.
  - **Types of Insight Deliverables.**
- Generalized implementation plan.
- Library of reference papers, presentations and texts.

One or more of four types of insight deliverables may improve the outcome at locations along an operational process

Category	Description
System reports	Taken from operating systems as standard reports.
Know-thy-data	Data is explored in descriptive, graphic and statistical perspectives.
Recountive	Insight direct from data—without processing through analytics—to ask and answer questions of who, what, when, where, how much and metrics.
Modeled	Insight gained upon data flowing through ML/AI models that ask and answer questions of relationship, difference, time series, duration and apparency.

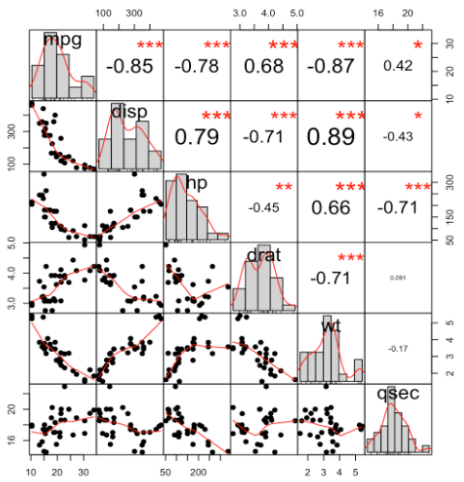
The place of the analytic core, “R,” in the scope of insight deliverables

Category	Role
System reports	None.
Know-thy-data	Insight is generated through various R functions specifically developed for the purpose of exploring data.
Recountive	R will come into play via the functions for layered charting.
Modeled	All are generated with “R” functions.

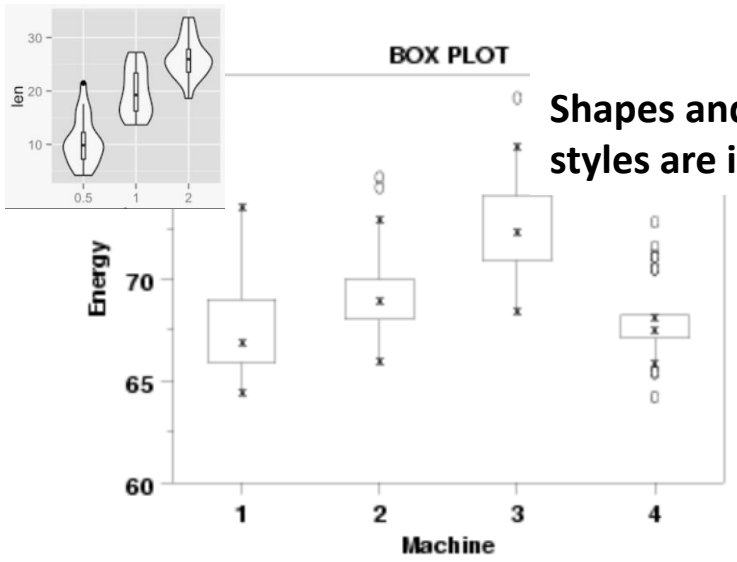
Next

Know-thy-data, recountive and modeled insight deliverables will be further explained. System reports require no further explanation.

R functions to "**Know-thy-data**" are combined as insight deliverables to probe and explore data



Unified view of eight insights into the nature and cross relationship of variables.(1)



Shapes and plotting styles are insightful.(2)

```
> summary(tdata)
  PassengerId       Survived          Pclass
Min.   :  1.0   Min.   :0.0000   Min.   :1.000
1st Qu.:223.5   1st Qu.:0.0000   1st Qu.:2.000
Median :446.0   Median :0.0000   Median :3.000
Mean   :446.0   Mean   :0.3838   Mean   :2.309
3rd Qu.:668.5   3rd Qu.:1.0000   3rd Qu.:3.000
Max.   :891.0   Max.   :1.0000   Max.   :3.000

                                    Name         Sex         Age
Abbing, Mr. Anthony                    :  1   female:314   Min.   : 0.42
Abbott, Mr. Rossmore Edward            :  1   male  :577   1st Qu.:20.12
Abbott, Mrs. Stanton (Rosa Hunt)       :  1                Median :28.00
Abelson, Mr. Samuel                    :  1                Mean   :29.70
Abelson, Mrs. Samuel (Hannah Wizosky):  1                3rd Qu.:38.00
Adahl, Mr. Mauritz Nils Martin         :  1                Max.   :80.00
(Other)                                :885                NA's   :177
     SibSp           Parch             Ticket         Fare
Min.   :0.000   Min.   :0.0000   1601     :  7   Min.   :  0.00
1st Qu.:0.000   1st Qu.:0.0000   347082   :  7   1st Qu.:  7.91
Median :0.000   Median :0.0000   CA. 2343:  7   Median : 14.45
Mean   :0.523   Mean   :0.3816   3101295  :  6   Mean   : 32.20
3rd Qu.:1.000   3rd Qu.:0.0000   347088   :  6   3rd Qu.: 31.00
Max.   :8.000   Max.   :6.0000   CA 2144  :  6   Max.   :512.33
                                 (Other)  :852
         Cabin     Embarked
             :687      :  2
B96 B98      :  4   C:168
C23 C25 C27:  4   Q: 77
G6           :  4   S:644
C22 C26      :  3
D            :  3
(Other)      :186
```

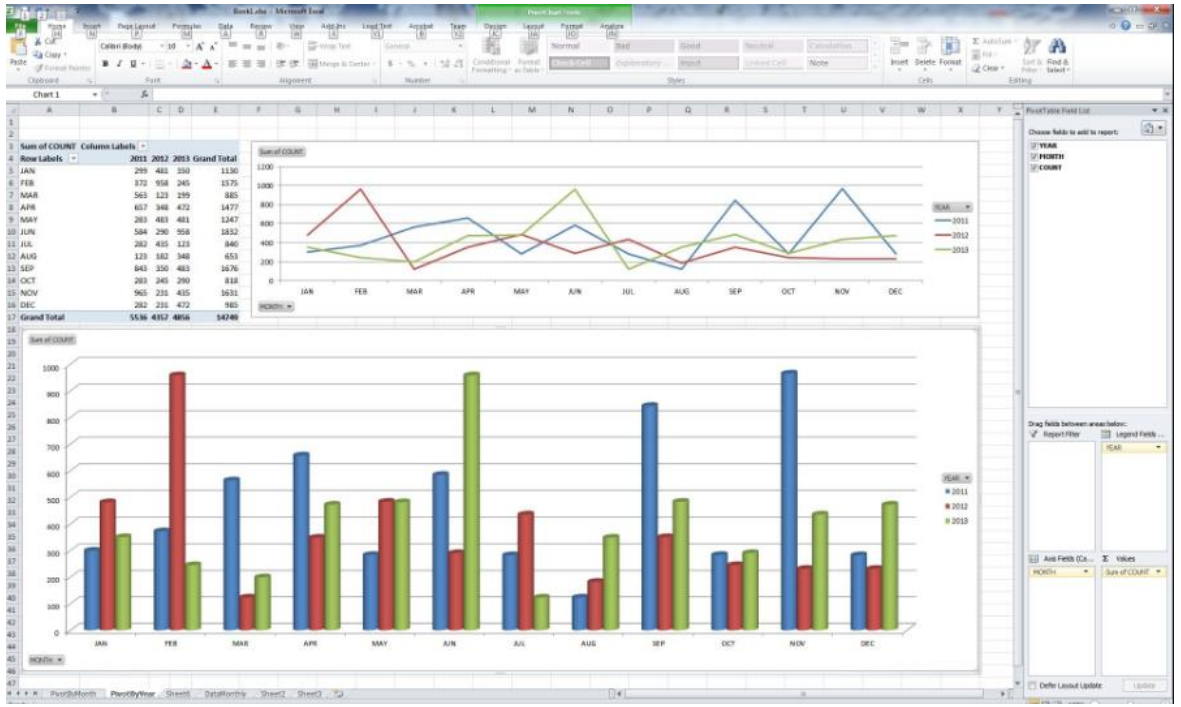
Min/max, median, mean, quartiles, categories and counts, and missing

See

1) **Correlation matrix, STHDA.** http://www.sthda.com/english/wiki/correlation-matrix-a-quick-start-guide-to-analyze-format-and-visualize-a-correlation-matrix-using-r-software
2) **R for Data Science, Grolemund, Wickham, 2017. eChap 7.** https://r4ds.had.co.nz/exploratory-data-analysis.html

“Recountive” insight deliverables are most widely observed current state to most organizations

Recountive insight can be packaged and disseminated through Pivot software directly from the super table.



See

- **Pivot Tables In-Depth for MS Excel 2016, Oesko, 2017.**
- **Excel Dashboards and Reports, Alexander and Walkenbach, 2013.**

For "modeled" insight deliverables, each of the five types of questions map to analytic models

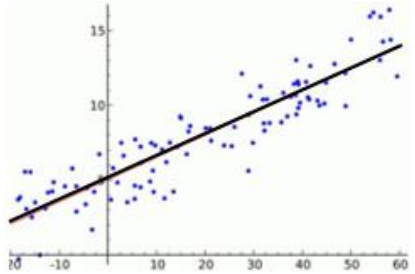
Question Type	Generic Question	Answering Model (1)
Relationship	Which asset and process variables are most strongly related to a performance of interest?	Linear, logistic and Poisson regression
Difference	How do slice-dice combinations of asset and process variables comparatively effect a performance of interest?	One-way and multi-way ANOVA, ANCOVA, repeated-measures and mixed ANOVA, and MANOVA.
Time series	What are the components that underlie the summary-level-only history that operating systems are limited to providing?	Holt-Winter, series regression, ARMA and ARIMA.
Duration	What is the probability an asset or process condition will hold for some time and then what is the probability the condition will end?	Cox regression, Cox proportional hazard, Cox mixed-effects, cumulative incidence, proportional hazard regression, Weibull and Crow-AMSAA.
Apparency	Are there hidden predictor variables to the performance of assets and processes?	Decision tree, regression tree, model tree, naïve Bayes, K-mean and principle component analysis.

(1) A set of explanatory papers to the models can be downloaded at https://analytics4strategy.com/new-age-five-questions

The next five slides are an overview of the five types of questions that can only be asked and answered by modeled insight.

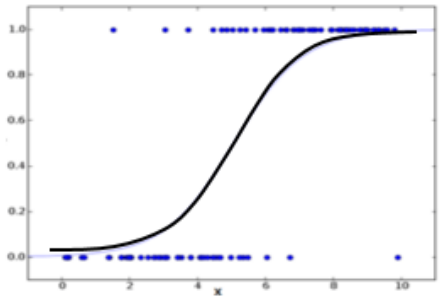
An extensive explanation is available with the training slides titled, "Modeled Insight Delivered by Analytic Models," at https://analytics4strategy.com/trn-modelinsigh

Relationships questions: Which asset and process variables are most strongly related to a performance of interest and by what type of relationship?



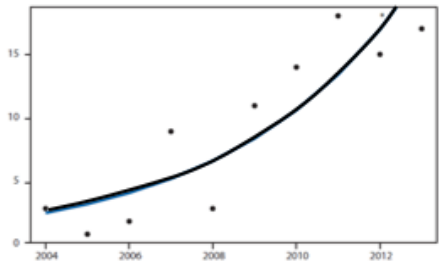
Linear regression:

- Numeric score—e.g., costs, hours, productivity and KPIs.
- Fit is linear.
- Example question: Which asset and process variables are most related to work order cost, hours, productivity and KPIs?



Logistic regression:

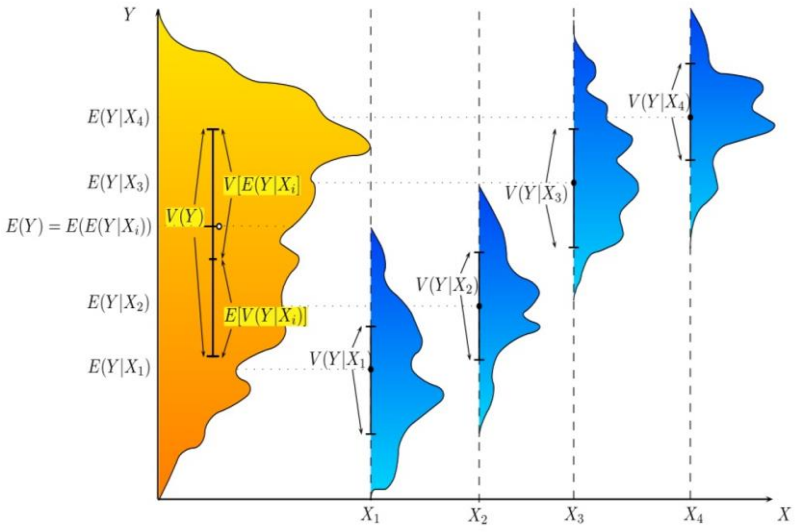
- Classifies between possible categorical outcomes.
- Binomial: Yes/No and multinomial: yes, no, maybe.
- Classification based on the probability of an outcome.
- Example question: Which asset and process variables most foretell the probability of finding crews engaged in non-value work?



Poisson regression:

- Outcomes of interest are occurrences—e.g., failures and emergency orders.
- Scored as counts (failures) or rates (failures per month).
- Example question: Which variables are most related to the count and rate of specific process non-compliances?

Difference questions: How do slice-dice combinations of asset and process variables comparatively effect a performance of interest?



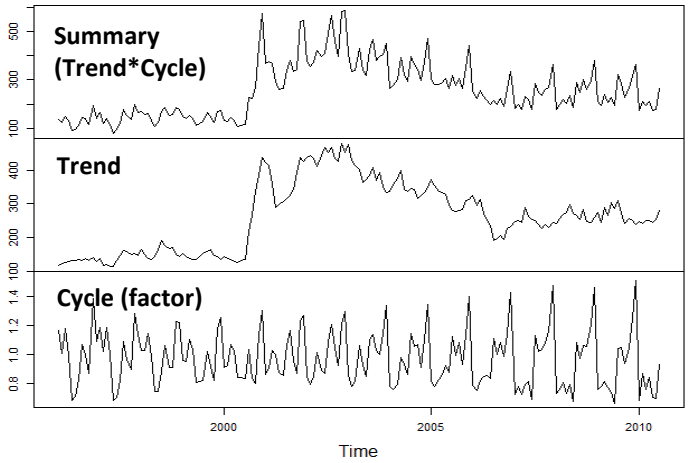
➢ Explanation:

- A data set made up of all variables in a model has a mean (expected value)—left-most.
- Grouping one or more variables results in a mean for each group.
- Difference models are necessary because simple pair-wise statistical comparisons are misinformation due to the math of error.
- The methods of comparison—post hoc and contrasts—are the power of difference models.

➢ Example Question:

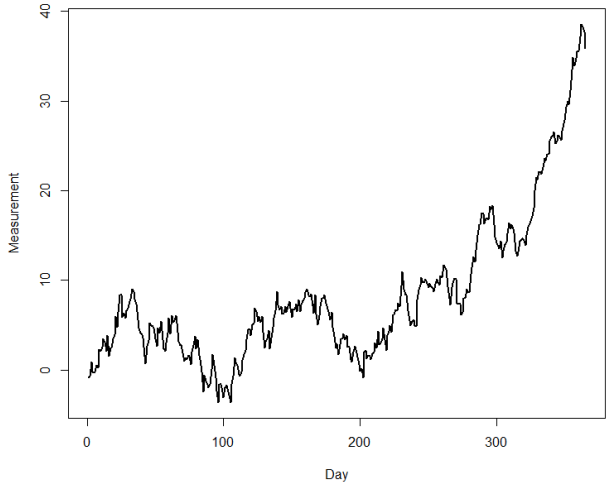
- In what situations are there the largest gap (non-value) between planned and actual craft hours for work orders?

Time series questions: What are the components that underlie the summary-level-only history that operating systems are limited to providing?



Operating systems only collect summary-level information.

The true story lies behind the summary-level data as the history of constituent trend and cycle.

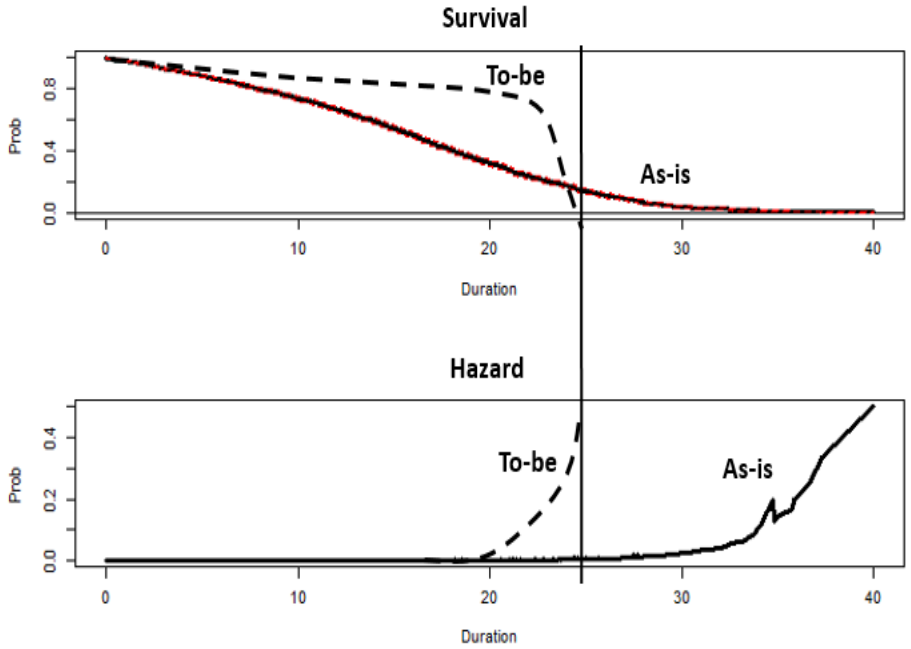


Just as important, what is the trend—deterministic, random or random walk?

Example question:

Is the trend for productivity changing with time—long-term or cyclical?

Duration questions: What is the probability an asset or process condition will hold for some time and then what is the probability the condition will end?



- Shows the probability of a condition lasting up to just before an ending event.

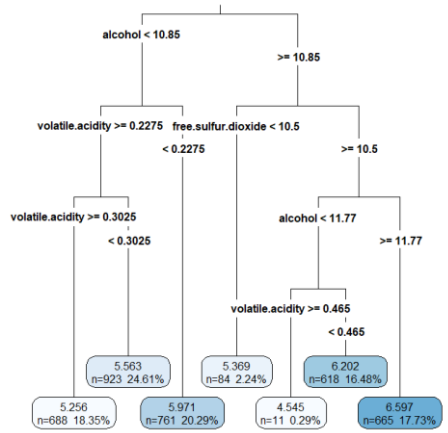
- Shows the probability of an ending event as a function of how long the condition has existed.
- Mathematics of "hazard" make it possible to determine which process variables matter most to the story.

Example question:

Are the shapes of the curves acceptable for each stage along the process and do any show gaming and non-compliance?

Note:
Shown is the Kaplan Meir method. Alternative methods upon life-data are Weibull and Crow-AMSAA.

Apparency questions: Are there hidden predictor variables to the performance of assets and processes?

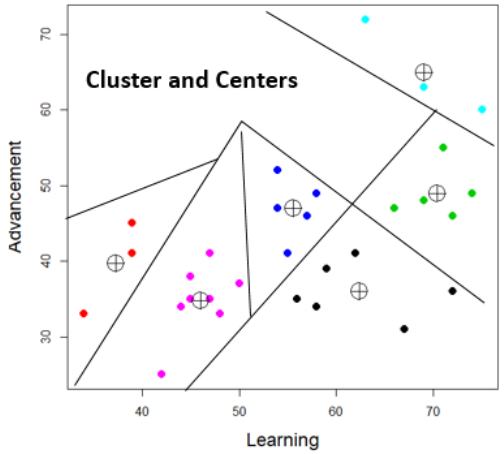


Decision tree:

- Hidden Rules as variables associated with numeric and categorical outcomes.
- An outcome variable is given to the model—supervised (directed).
- The learned rules to score or classify outcomes.

Example question:

Is the process operating to the established rules of conduct?



K-Means:

- Hidden variables are learned by slicing-dicing existing predictor variables until clusters of similarity emerge.
- No outcome variable is given to the model—unsupervised (undirected).
- Shown case: Six hidden variables are teased out of two variables.

Example question:

Do hidden variables point to outcomes and classifications we have never before recognized or recorded?

Within each type of model to a question type, there are subtypes as well as many combinations and configurations of variables

- The choices of models and variations are driven by the nature of the process and the insight deliverables at the places in the process.

- The considerations for variations to each are explained by the following referenced papers:

Question	Paper	Link
Relationship	Find What Matters with Relationship Questions of Operations	https://analytics4strategy.com/relatqstoci
Difference	Know that Improvements Work by Asking Difference Questions	https://analytics4strategy.com/viveladifference
Time series	Explore What Did and May Happen with Time Series Questions	https://analytics4strategy.com/timeseriesqs
Duration	Find the Time That is Money by Asking Duration Questions	https://analytics4strategy.com/tmismnyqstns
Apparency	Dive Below the Surface of Process Functioning with Apparency Questions	https://analytics4strategy.com/apprqsblwfnctng

Agenda:

- The big picture of data-driven operations.
  - Definition and depiction of a data-driven operation.
  - Cost-free “Critical-Mass” strategy for reaching data-drivenness.
  - Essential definitions.
- Structure of methodologies.
  - “R”—as the analytic core of data-driven capability.
  - Gather and join subtables to form and cleanse super tables.
  - Layered charting in contrast to conventional charting.
  - Types of Insight deliverables.
- **Generalized implementation plan.**
- Library of reference papers, presentations and texts.

Your organization needs a plan of stages, steps, tasks and deliverables along which to steer each chosen operation from its current state to being fully data-driven. . .

. . .**BUT** to succeed, the path must also be charted to cause the transfer of knowledge and skills of data-drivenness to the chosen operations.

The organization must make a choice to either incubate or recruit the necessary talents for its data-driven operations

- Polar opposite choices:
  - Upskill the process role holders to be new-age workers.
  - Recruit specialist in the methodologies, leaving the process role holders to continue on with only pre-data-driven skills.

- Drawbacks to the specialist alternative:
  - Is it **fair** to the process role holders to be left in the past?
  - Specialists lack the depth and breadth of domain expertise that process role holders have gained over years.

- The herein charted implementation plan is designed to upskill role process holders as the implementation project unfolds.

Generalized stages, steps and tasks to reaching data-drivenness

Stages and Steps

Stage 1: Set direction and prepare.

1. Form clear understanding with leadership and initial decisions.
2. Prepare nucleus players to participate.
3. Set competitive North Star for data-drivenness.

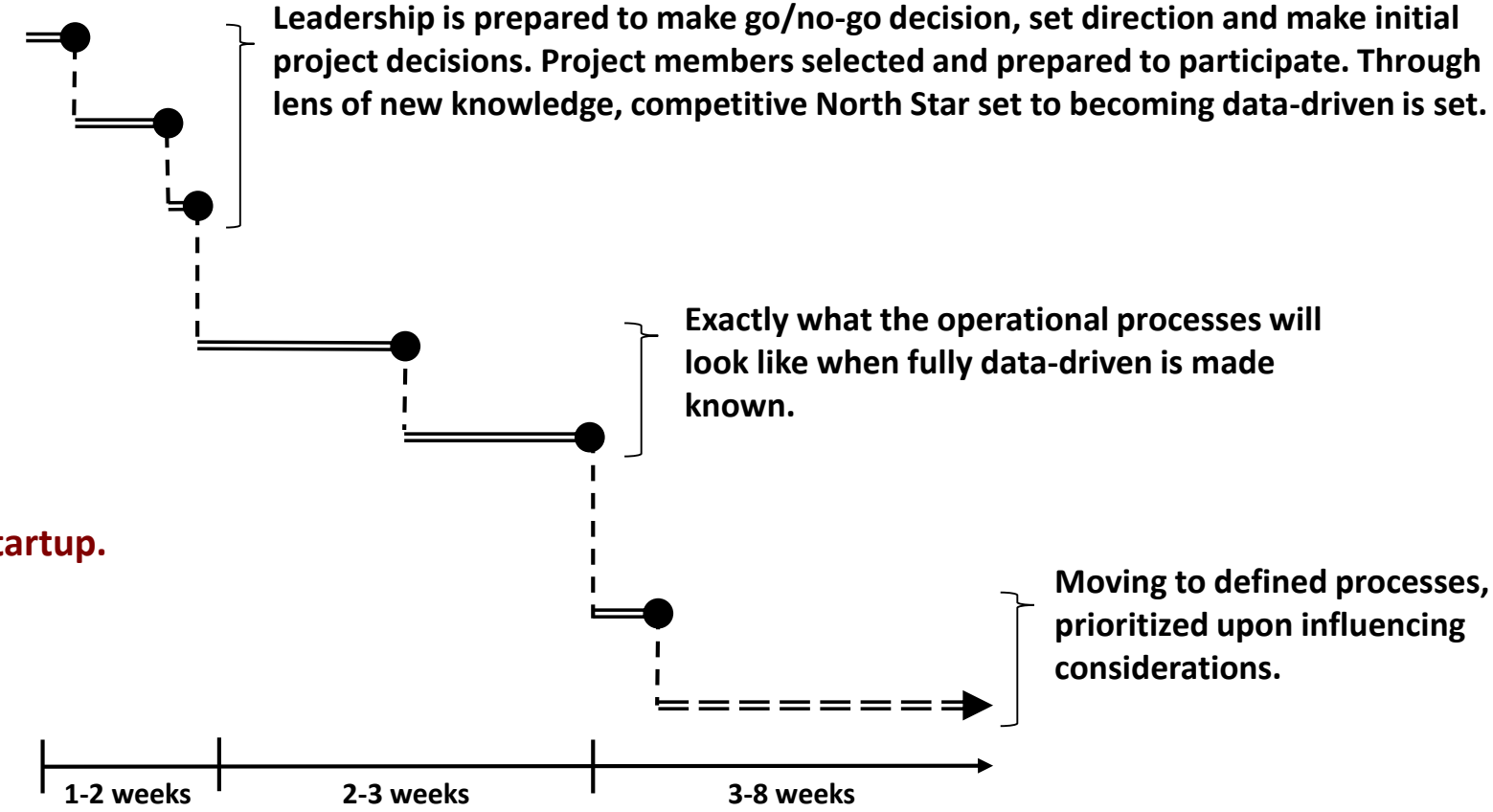
Stage 2: Conduct basic design.

1. Charted, detailed operational processes for data-drivenness.
2. Map deliverables to charted process & form basic designs

Stage 3: Plan and form detailed design and startup.

1. Establish the sequence to implement insight deliverables.
2. Form detailed designs and implement.

Timeline to General Abilities



The previous slide is an overview plot of the generalized project plan and the extent the project stages, steps and deliverables will be presented.

The fully detailed project plan is available as the training slides titled, "Project Plan for Implementation at the Grassroots" at https://analytics4strategy.com/trn-implement

Stage One: Set direction and prepare

Stage prepares the subject organization's leadership to make the initial go/no-go decision for data-drivenness. If "go," and with the preparation of the stage, leadership will set direction and make initial decisions for the project. Upon the decisions, project members will be selected and prepared to participate as project members.

Step 1: Prepare leadership to give direction and make decisions for data-drivenness.

The step is to give leadership a clear, implementable understanding of data drivenness. With the understanding, leadership will set direction and make the initial decisions with respect to go/no-go, geography and project nucleus players.

Activities	Deliverables
1. Present to leadership a clear, implementable explanation of data-drivenness. 2. Bound the organization to be subjected to an initial data-drivenness project. 3. Select the nucleus players from the process of operatives, leaders and analysts to participate in the project.	• Decision-level leadership is conversant and knowledgeable in the principles, practices, software and skills of data-drivenness. • Organization delineated for becoming fully data-driven. • Nucleus of personnel who will participate in building and dispersing the principles, practices and skills to their colleagues within and possibly beyond the targeted geography.

Continued. . .

As the three stages unfold, the knowledge and skills of data-driven operations are transferred to **management** and **nucleus players** by a **drivenness guide**

Stage	Knowledge and Skill Transfer
1. Set direction and prepare.	▪ Management and the nucleus players they select will learn the framework of data-driven operations. ▪ Nucleus players trained hands-on in the skills of building and cleansing super tables, exploring the tables with methods of descriptive statistics and building layered charts.
2. Conduct basic design.	▪ Nucleus players trained in the types, principles and interpretation of ML/AI analytics as mandatory to have the expertise to specify which of all insights deliverables are relevant to the process.
3. Plan and form detailed design and startup.	▪ Nucleus players trained hands-on to build and deploy the insight deliverables they specified in the previous stage.

Agenda:

- The big picture of data-driven operations.
  - Definition and depiction of a data-driven operation.
  - Cost-free “Critical-Mass” strategy for reaching data-drivenness.
  - Essential definitions.
- Structure of methodologies.
  - “R”—as the analytic core of data-driven capability.
  - Gather and join subtables to form and cleanse super tables.
  - Layered charting in contrast to conventional charting.
  - Types of Insight deliverables.
- Generalized implementation plan.
- **Library of reference papers, presentations and texts.**

Library: "What-to" and "how-to" for data-driven development and functioning

Knowledge and skills		Papers, presentations, training sessions	Texts or equivalents
Data-drivenness	Framework	▪ First Step to Becoming a Data-Driven Operation ▪ Data-Driven Maintenance Operations	None available
R	Coding	None available	• R for Dummies, de Vries, Meys, 2015. • Art of Programing R, Matloff, 2011. • Manual at https://r-project.org.
Data tables	Super tables	▪ Build Super Tables from Operational Data ▪ Purge the Fused Spreadsheets That Undermine Data-Drivenness	Access 2016 Bible, Alexander and Kusleika, 2016, Chapters 8 – 13.
Data preparation	Cleansing	▪ Build Super Tables from Operational Data	Rstudio for R Statistical Computing Cookbook. Andrea Cirillo, 2016, Chapter 2
Pivot tables, graphs	Pivot dashboards	None available	Pivot Tables In-Depth for MS Excel 2016, Oesko, 2017.
	Layered charting	None available	ggplot2, Elegant Graphics for Data Analysis, Wickham, 2016

Continued. . .

Library: Continued

Knowledge and skills		Papers and presentations	Texts or equivalents
Five analytic questions	Relationship	Find What Matters with Relationship Questions of Operations	▪ Discovering Statistics Using R, Field and Miles, 2012 ▪ Multilevel Modeling Using R, Holmes, 2014
	Difference	Know that Improvements Work by Asking Difference Questions	
	Time series	Explore What Did and May Happen with Time Series Questions	▪ Introductory Time Series with R, Cowpertwait and Metcalfe, 2009 ▪ R Package "tsoutliers," Javier López-de-Lacalle, 2017
	Duration	Find the Time That is Money by Asking Duration Questions	▪ Event History Analytics with R, Bostrom, 2012 ▪ New Weibull Handbook, Abernathy, 2007 ▪ R Package "WeibullR" Weibull Analysis for Reliability Engineering, Silkworth & Symynck, 2018
	Apparency	Dive Below the Surface of Process Functioning with Apparency Questions	Machine Learning with R, Lantz, 2015
Machine learning, AI	Methodology	None available	