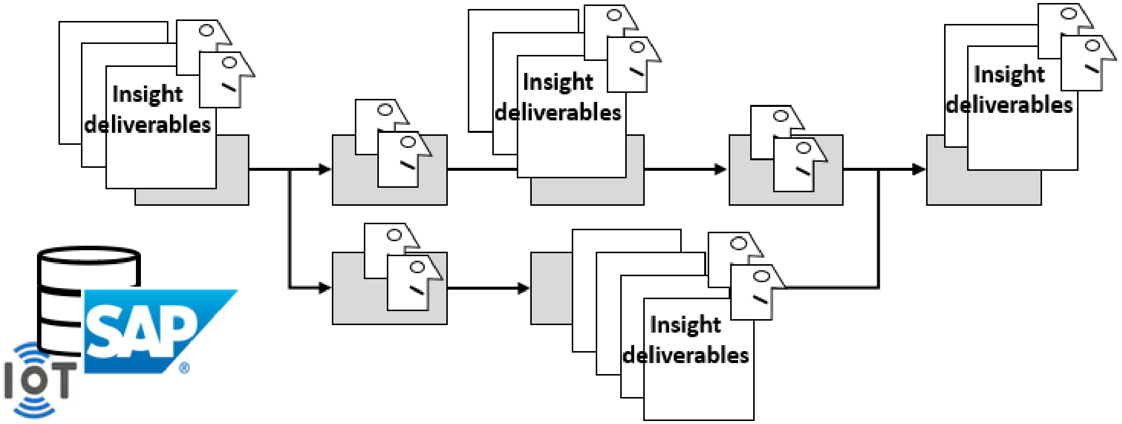
It Is All About Data-Driven Operations

A Clear, Implementable Understanding of Data-Drivenness



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

Download slides at: https://analytics4strategy.com/datadriventrc2019
Unabridged version: analytics4strategy.com >> Training Sessions >> Scroll to "First Step to. . ." >> Download slides at bottom

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

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

We are here:

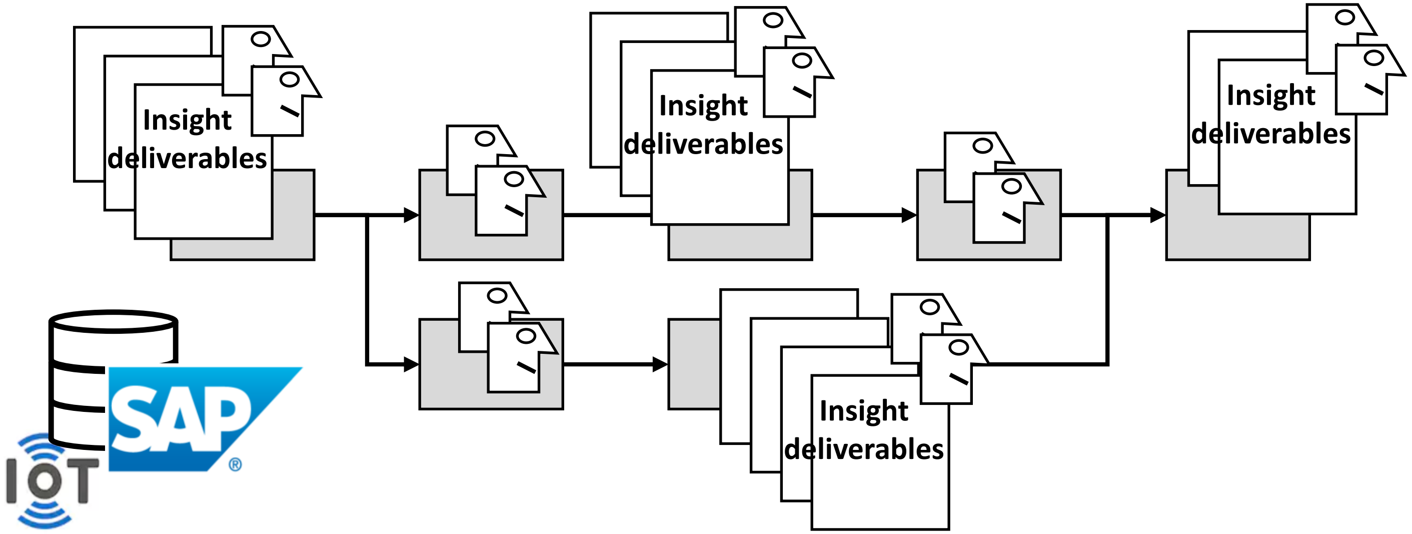
- The big picture of data-driven operations.
  - Definition and depiction of a data-driven operation.
  - Cost-free "Critical-Mass" strategy for reaching data-drivenness.
  - Jargon of data science reduced to relevance to data-drivenness.
- Structure of methodologies.
  - "R"—as the analytic core of data-driven capability.
  - Gather, join, cleanse and form data into super tables.
  - Layered charting in contrast to conventional charting.
  - Types of Insight Deliverables.
- Generalized implementation plan.
- Library of what-to and how-to 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

- Operational processes are sectors, paths and steps at which there are acts to plan, organize, conduct or control.
- At some places, the "**best outcomes"** can only be realized when the experience and judgement of the process role players are augmented with "**insight deliverables.**"
- At each identified such place, a set of system reports, tables, charts and models is built to realize the best of outcomes.



We are here:

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

Your operation has the option of taking a "critical-mass" strategy to reach data-drivenness because drivenness is not high-tech or new-tech as much as it is 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 exercised across **critical-mass-improved** operations transfer to up-teched and up-scaled strategies.
  - Up-teching and up-scaling will not practicably increase the power of the insight that would otherwise be extracted from the operation's data.

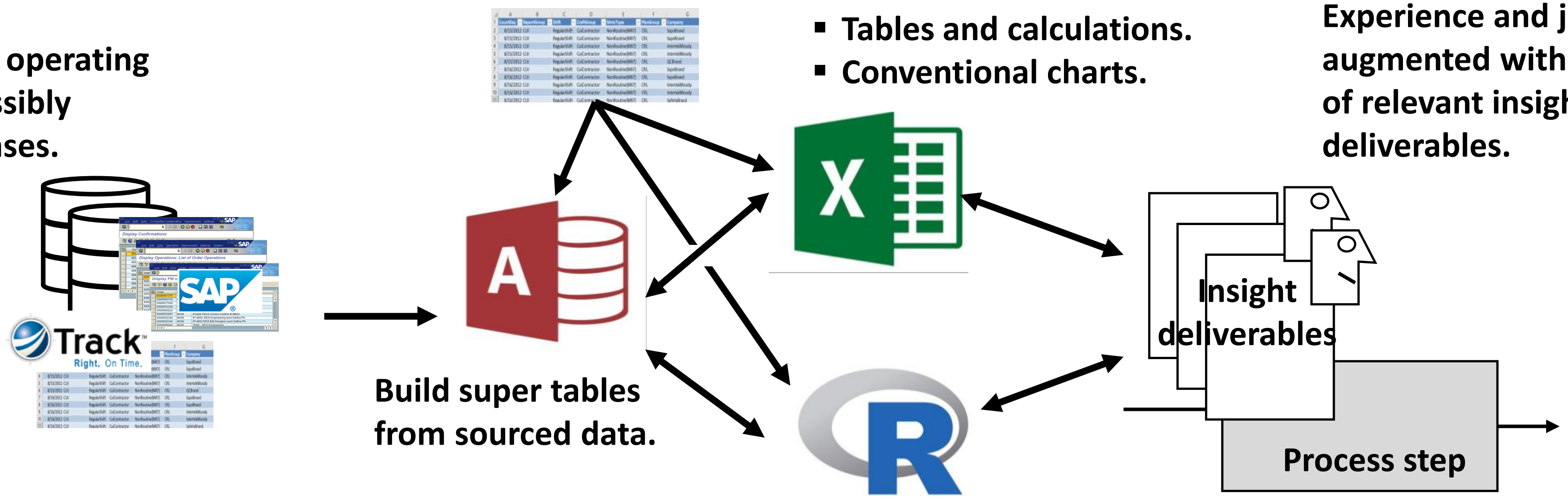
What sets the threshold to critical-mass is determined by a triad of software that maps along the progression from source data through to augmenting insight deliverables

Data in multiple operating systems and possibly multiple data bases.

Data from process tasks conducted with Excel.

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

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



Build super tables from sourced data.

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

The triad has immense ramifications for grassroot initiatives to build data-driven operations

- **It is a full-powered alternative—R being best of class.**
- **There are no organizational and system issues.**

  Two are standard to MS office (Excel, Access) and one is an open source, free to download (R).
- **Every seat along the operational process can be armed with the triad.**
- **R is a software we will see elsewhere:**

  Corporate IT strategies and commercial software are interfacing to R for analytics and charting.

We are here:

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

The over-mystified jargon of data-science does not tell the story of data-drivenness—only causes us to conclude that the story is beyond our reality

- Jargon needing to be demystified:
  - Data and big data
  - Machine learning
  - Artificial intelligence
  - Algorithms
  - Models
  - Visualization
  - IoT
  - Digitalization
  - Intelligence
- All reside behind the curtain of data-drivenness.
- It is useful to frame the meaning of the terms.
  - Data and big data—an important distinction.
  - Analytic tasks—machine learning, artificial intelligence, algorithm and models.
  - Visualization—charting renamed.
  - IoT, digitalization and intelligence—open-ended, aspirational.

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 (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 of the same.

- **What you need to know is that the data of operational processes rarely cross over the line from data to big data.**

Think of **"machine learning"** as demonstrated by your own experience with linear regressions

- Variables from a super table are selected—one or more as "predictors" and another as the "outcome."
- The data are fed to the model and its gut **algorithm** conducts a trial-and-error calculation until by "**machine learning**" arrives at the best fit to a multidimensional linear line.
- What is learned is returned as an expression with a coefficient for each variable telling us how much, if any, the variable plays in predicting the outcome.

Extending the linear regression example, "**Artificial intelligence**" is to put the learned model to the work of "advising" us of the outcomes to be expected

- **Often times we do not "learn" for the purpose of predicting and forecasting, but to get a clear window into how things work—we stop at machine learning.**

- **Artificial intelligence (AI) takes a step farther to predict or forecast outcomes upon the "learned" model.**

- **In the example of a linear regression, values to the variables are fed to the learned model and we receive a predicted or forecasted outcome.**

- **AI does not distinguish the model—all model types entail machine learning and most can be deployed as AI.**

Definitions for the remaining listed jargon

Jargon	Definition
Visualization	Rename of charting along with renaming content as "patterns."
IoT (internet of things)	System of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers (UIDs) and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction.
Digitalization	Use of digital technologies to change a business model and provide new revenue and value-producing opportunities.
Intelligence	Umbrella term that includes the applications, infrastructure and tools, and to access and analyze information to improve and optimize decisions and performance.

We are here:

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

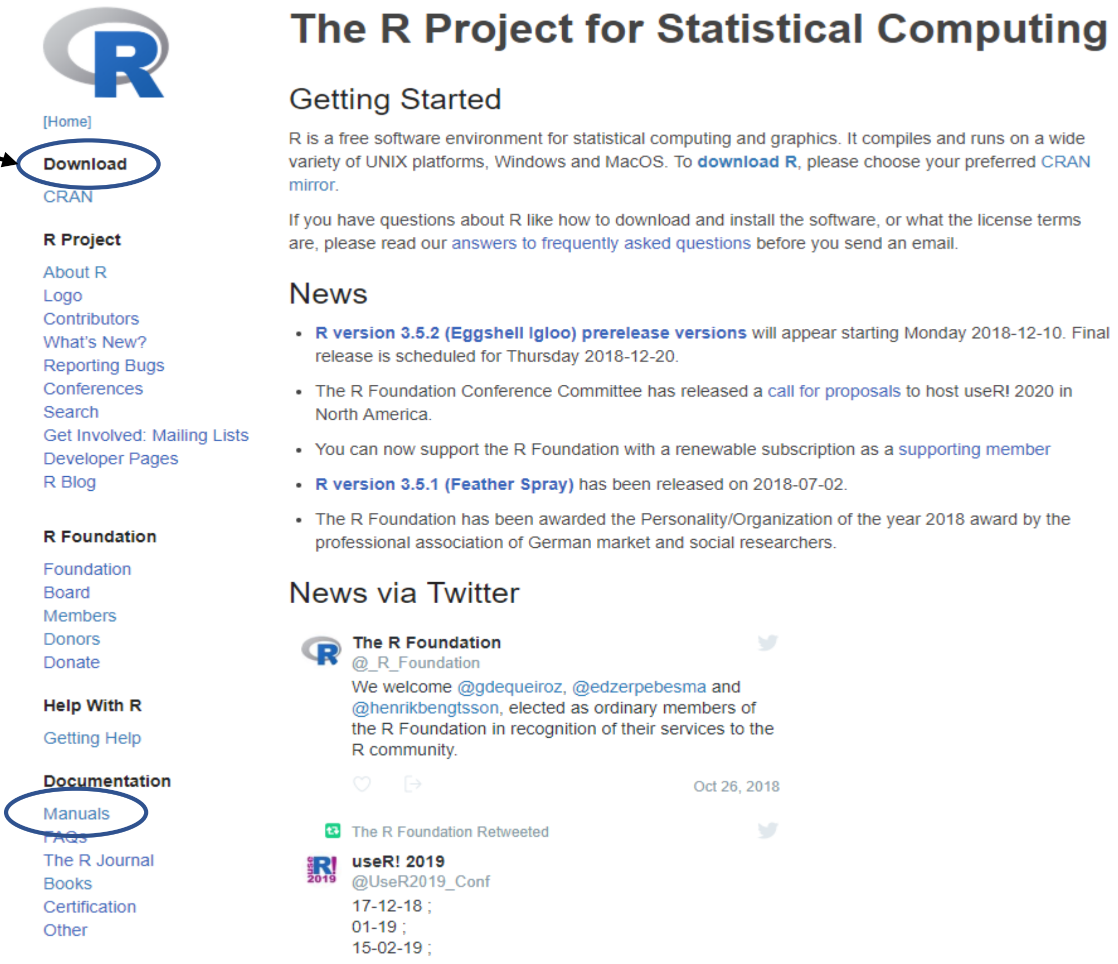
The foundation technology to being data-driven is an analytic software because only through analytics can there be. . .

- **Descriptive statistics.**
- **Layered charting.**
- **Maximal data cleansing.**
- **ML/AI-based models.**

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

Download

Manuals



The R Project for Statistical Computing

[Home]

Download

CRAN

R Project

About R
Logo
Contributors
What's New?
Reporting Bugs
Conferences
Search
Get Involved: Mailing Lists
Developer Pages
R Blog

R Foundation

Foundation
Board
Members
Donors
Donate

Help With R

Getting Help

Documentation

Manuals
FAQs
The R Journal
Books
Certification
Other

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred CRAN mirror.

If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

News

- **R version 3.5.2 (Eggshell Igloo) prerelease versions** will appear starting Monday 2018-12-10. Final release is scheduled for Thursday 2018-12-20.
- The R Foundation Conference Committee has released a call for proposals to host useR! 2020 in North America.
- You can now support the R Foundation with a renewable subscription as a supporting member
- **R version 3.5.1 (Feather Spray)** has been released on 2018-07-02.
- The R Foundation has been awarded the Personality/Organization of the year 2018 award by the professional association of German market and social researchers.

News via Twitter

The R Foundation
@_R_Foundation
We welcome @gdequeiroz, @edzerpebesma and @henrikbengtsson, elected as ordinary members of the R Foundation in recognition of their services to the R community.
Oct 26, 2018

The R Foundation Retweeted
useR! 2019
@UseR2019_Conf
17-12-18 ;
01-19 ;
15-02-19 ;

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

- **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.**
- **Texts on data and analytics are plentiful in which R is used.**

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.

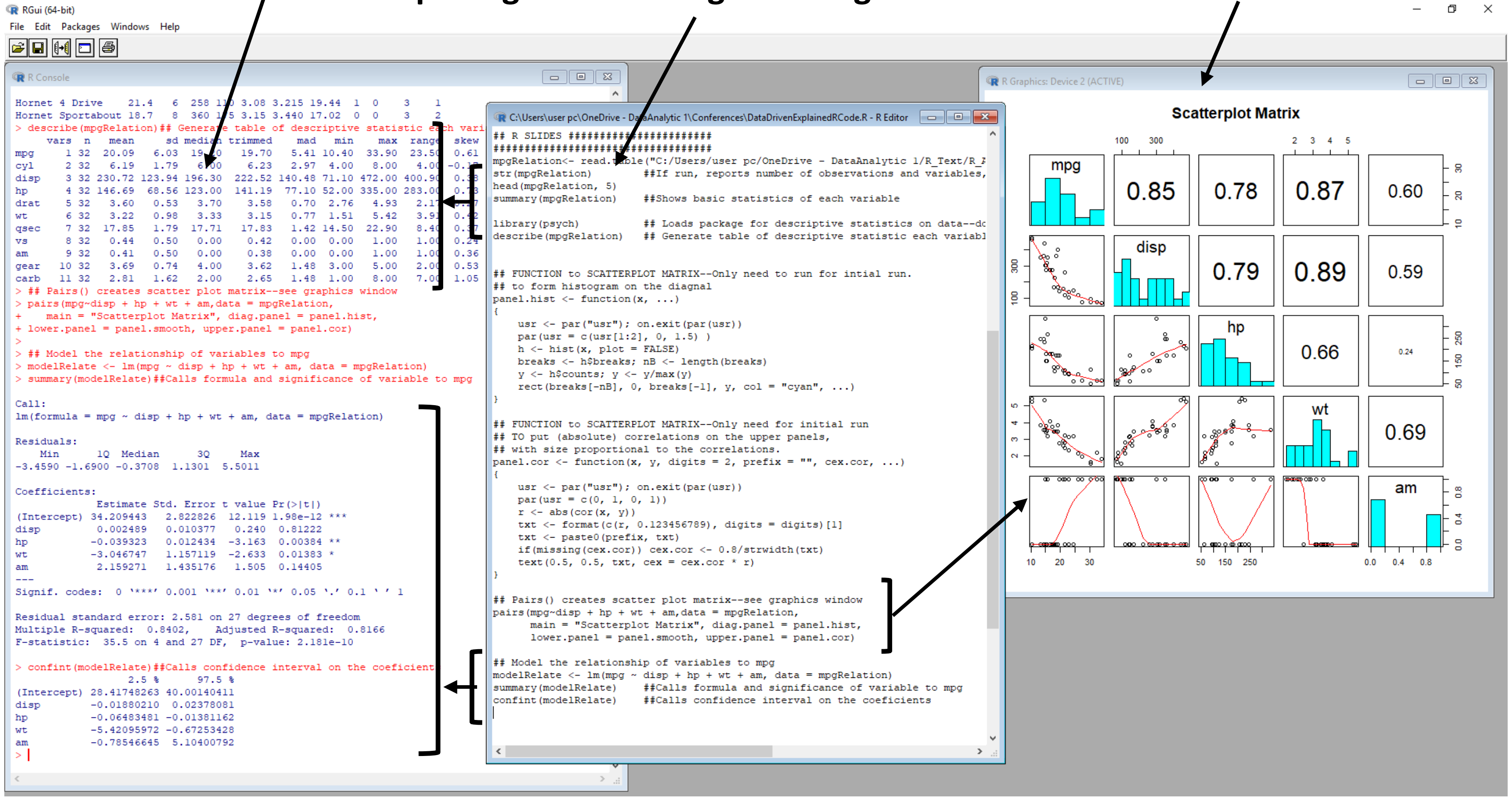
```
lm(formula, data, subset, weights, na.action, method = "qr",
   model = TRUE, x = FALSE, y = FALSE, qr = TRUE,
   singular.ok = TRUE, contrasts = NULL, offset, ...)
```

What it looks like, the frontend view of the "R" software

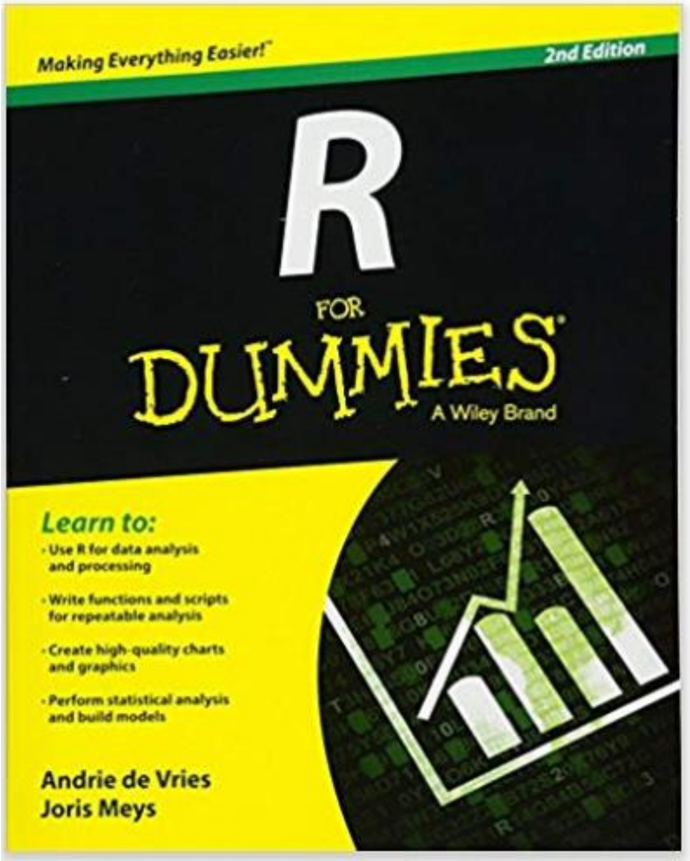
Textual results from running entire script or specific lines

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

Graphic results from running associated lines in the script.



If you are a self-directed learner, this will get you there quickly



The easy-read-and-do text will make you competent enough to work with the methodologies of data-drivenness.

We are here:

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

In a data-driven operation, not everyone needs to have hands-on skills in building super tables, but almost everyone must be conversant—the purpose of this section.

Training slides to the depth of hands-on skills is the purpose of the session titled, "Build Super Tables from Operational Data," >> analytics4strategy.com >> Training Sessions >> Scroll to "Build Super. . ." >> Download slides at bottom

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.**

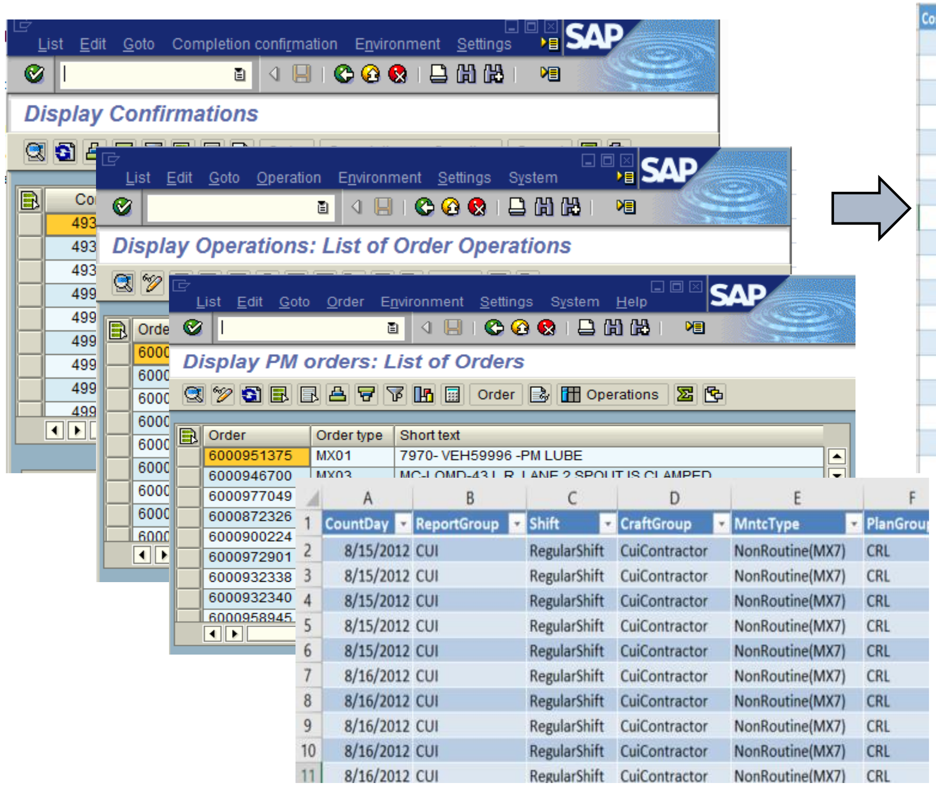
  When not, the IT data specialists can give us an on-demand tool to do so.

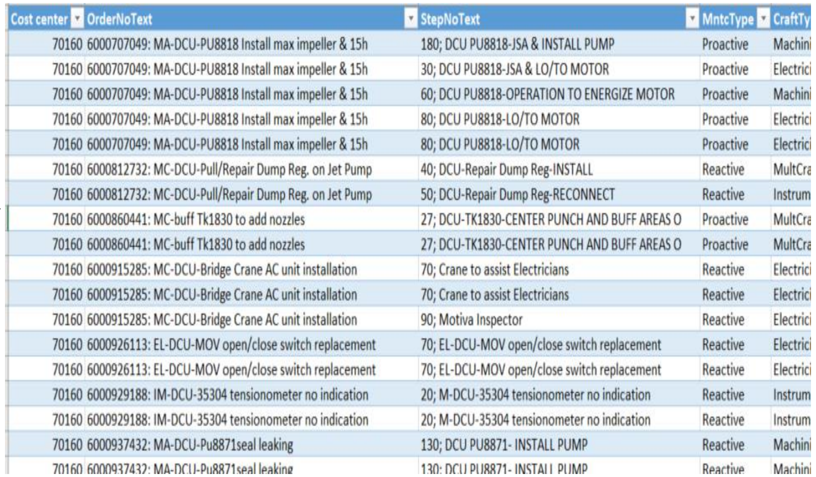
- **Individual data tables from any one or more systems or sources can be joined into one table by any identifying variable they have in common.**

  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: Extract topic-specific data from sources and fabricate a **super table** as required to build one or more intended 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.

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

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

Step 1: Identify relevant tables →

Step 2: Extract tables from source systems →

Step 3: Import or connect source tables into a query software →

Step 4: Build super table with query functionality →

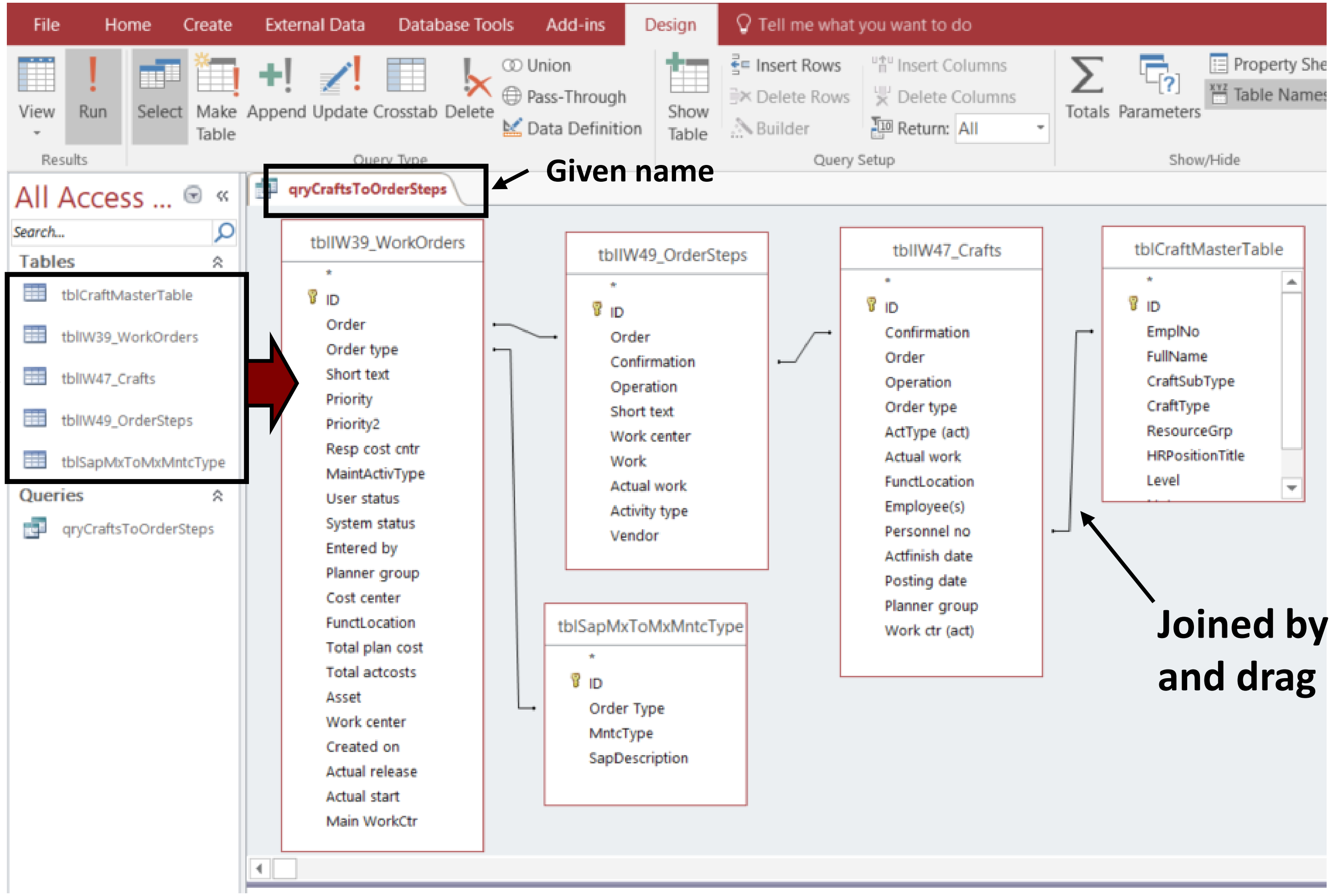
Step 5: Explore, cleanse the data →

Step 6: Build aggregate variables into super table →

Step 7: Administer, update and upgrade, and disseminate super table

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.



Given name

Joined by click and drag

The created “raw” table can serve many insight deliverables; but we typically mold super tables with respect to the particular insights we seek

Super table built by click and drag, and expressions.

Become columns to the final super table

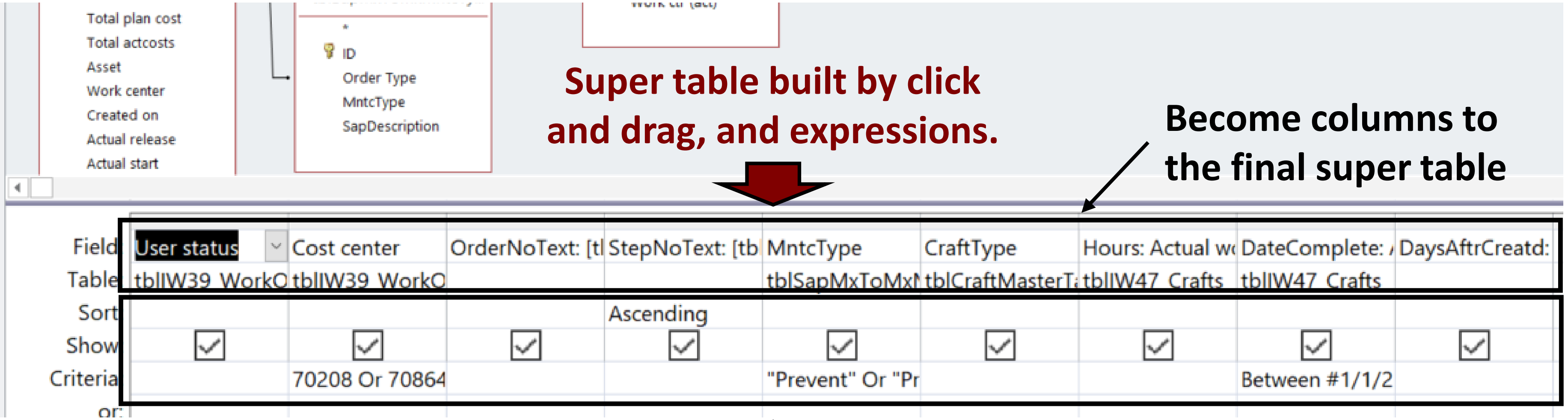
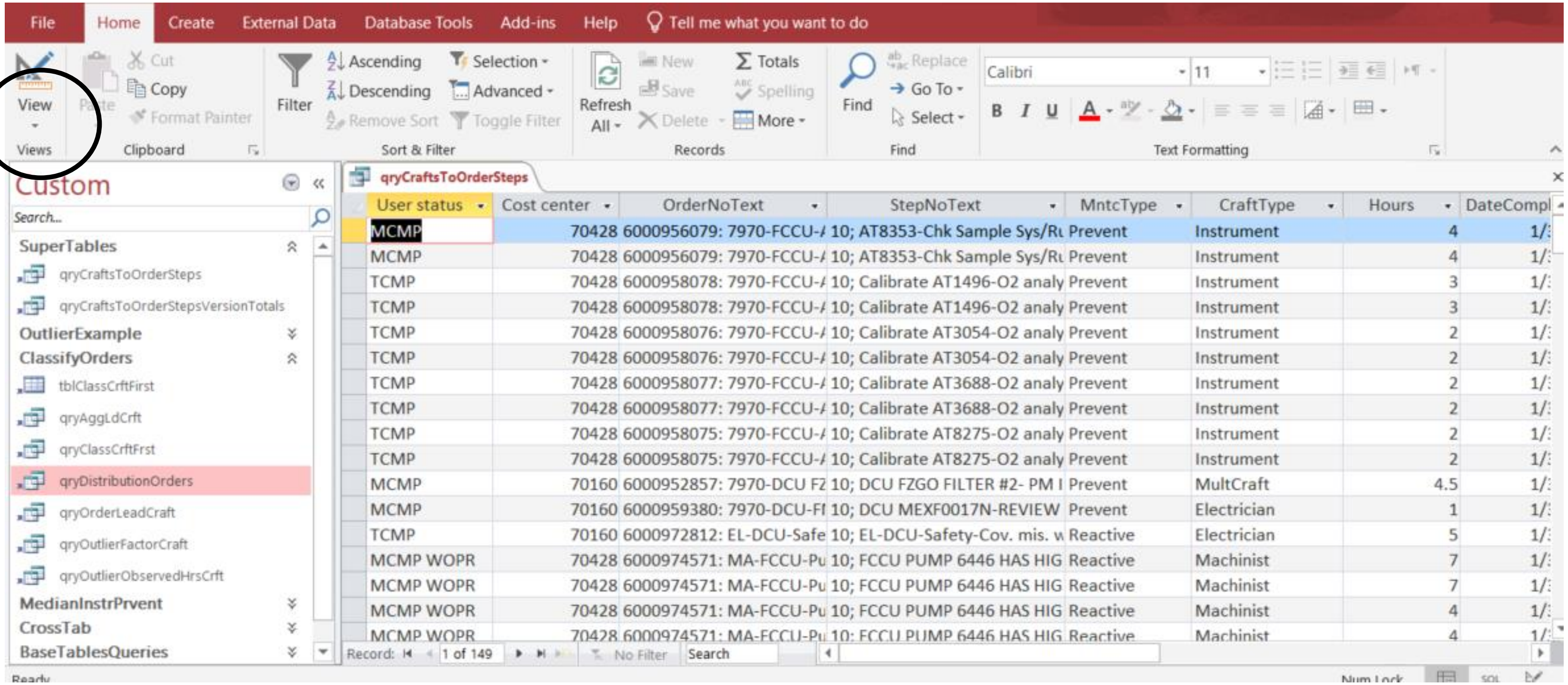
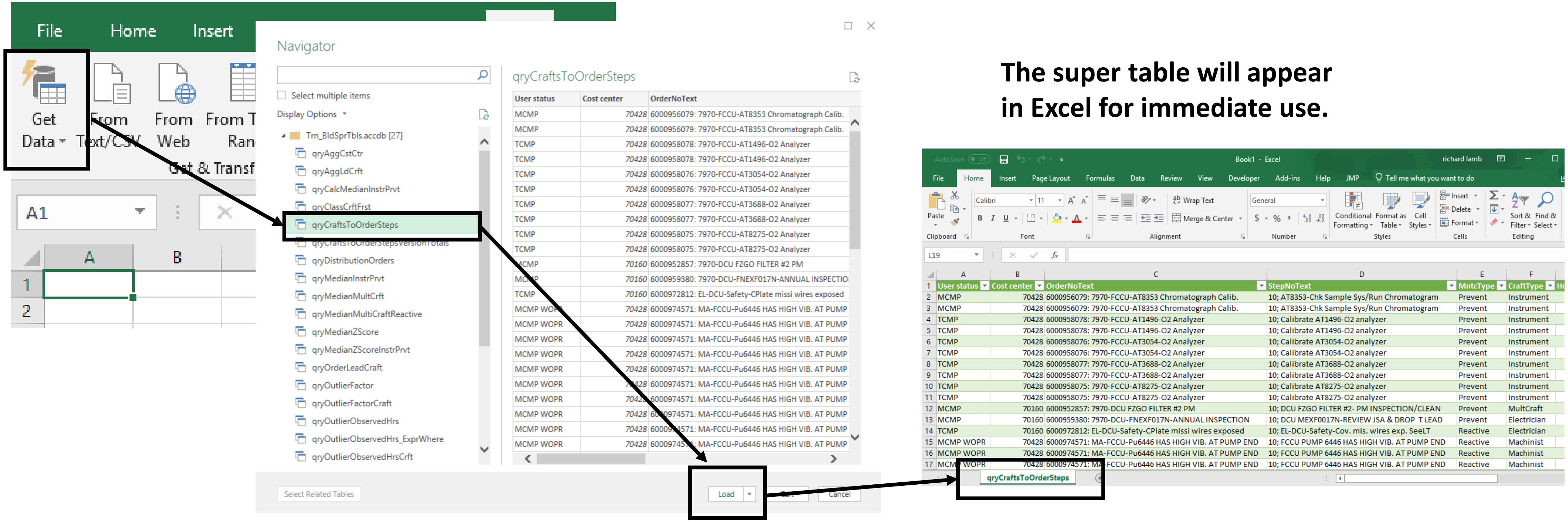


Table is further molded by choices for sorting and criteria (filtering)

The materializing super table is viewed and explored back and forth between the "Design" and "Table" view to see what has been wrought



Bring the query into Excel 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.

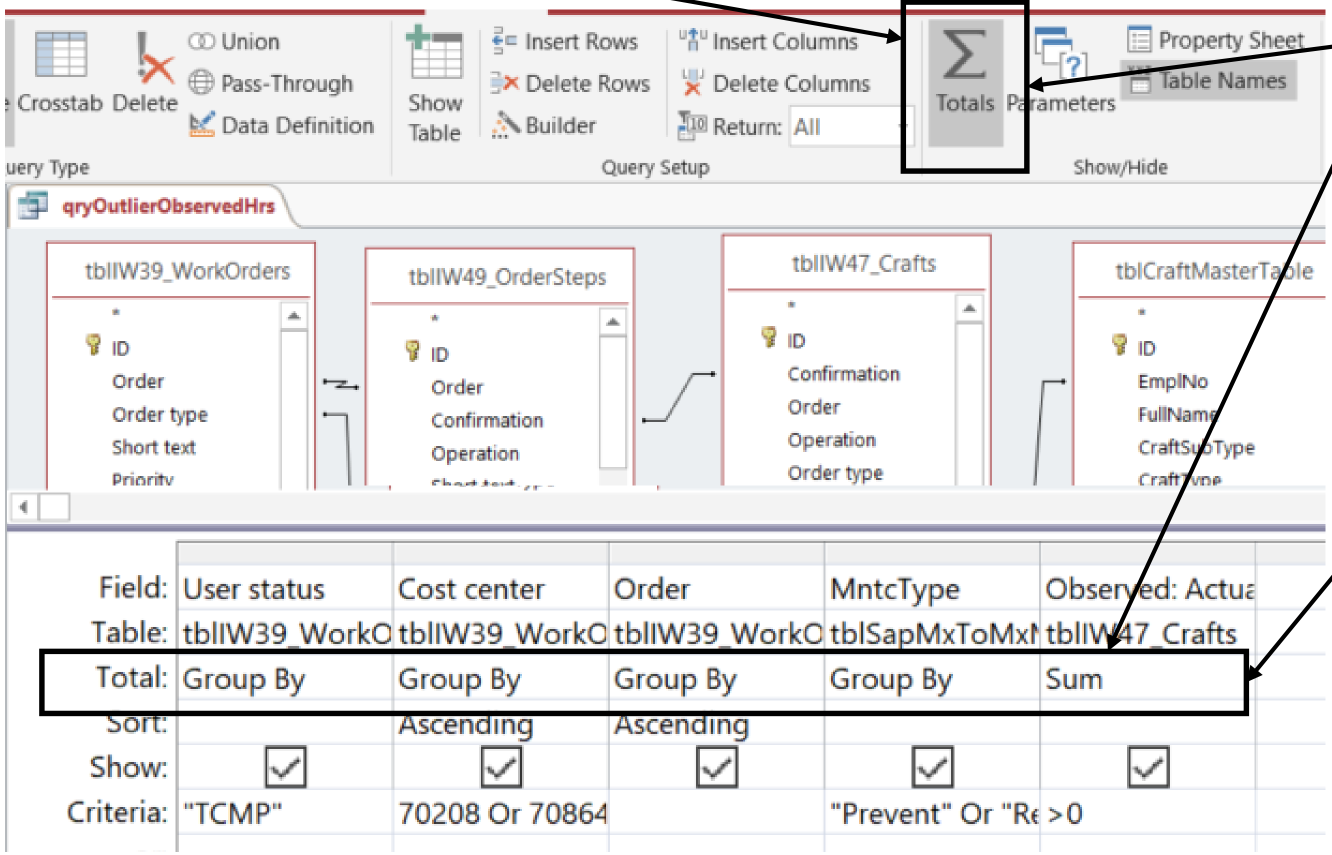
Get data >> From Data Base >> From Microsoft Access Database >> select File >> Select table from list >> click Load

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

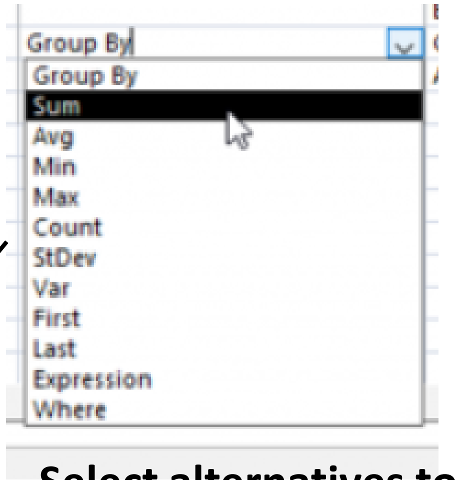
Type	Strategy
Duplicate cases	Seek cases with duplicate query.
Empty cells	▪ Form table with all permutation of empty cells—use "or" rows in query grid and use Pivots to form table of cases and counts. ▪ Evaluate the ramifications of loss of information to insight deliverable of remaining empty. ▪ Decide a strategy. ○ Decide to ignore for various statistical logic or. . . ○ use ML/AI models applied to good cases to predict or classify what should be. ▪ Likely models are one of three regressions (linear, logistic, Poisson), rule trees, naïve Bayes and K-Means. (1)
Misclassifications	▪ Essentially an equivalent case of empty cells for categorical variables. ▪ Likely models are logistic regression, rule trees, naïve Bayes and K-Means. (1)
Misformatted	▪ Build translation tables for each bad-data case to a variable. (2) ▪ Attach to super tables and use translated, rather than source dirty variable.
Outliers (numerical)	▪ Use aggregate functionality of the "Total" row (introduced in later slide) and build an outlier test variable into the super table. (2) ▪ Locate outliers filtered on test and determine if interesting or bad. . . ○ If interesting, is new insight and retained in super table. ○ If bad, remove case or impute as equivalent case to empty cells.

(1) See discussion of indicative insight deliverables in the section titled, "Types of insight deliverable."
(2) See slides to training session titled, "Build Super Tables from Operational data"

Aggregation functionality is activated within a select query, thence creating fields and criteria are the same—Except



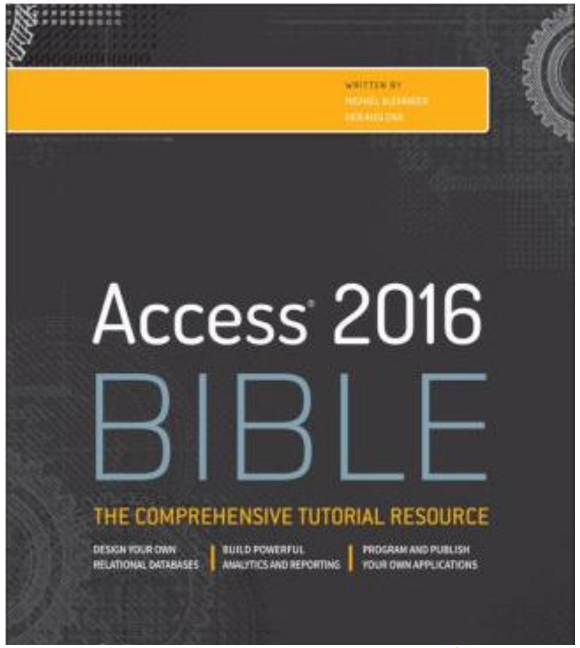
Clicking the icon adds the "Total" row of the query grid.



Select alternatives to the default "Group By"

Another advantage of Access is that there is an immense support community—including referenced slides

- Chapters 8 through 16 explain most of everything there is to know about building and exploring super tables.
- On line, every subject in the book can be found explained and demonstrated as a YouTube video, blog or article.



GCF LearnFree.org

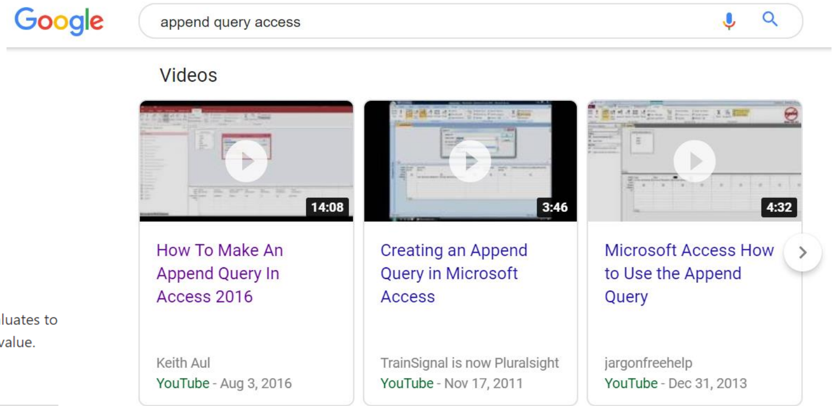
Query Criteria Quick Reference Guide

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



Examples of expressions

Access for Office 365, Access 2019, Access 2016, Access 2013, Access 2010, Access 2007

This article provides many examples of expressions in Access. An expression is a combination of mathematical or logical operators, constants, functions, table fields, controls, and properties that evaluates to a single value. You can use expressions in Access to calculate values, validate data, and set a default value.

In this article

Forms and reports

Queries and filters

All query and filter expressions

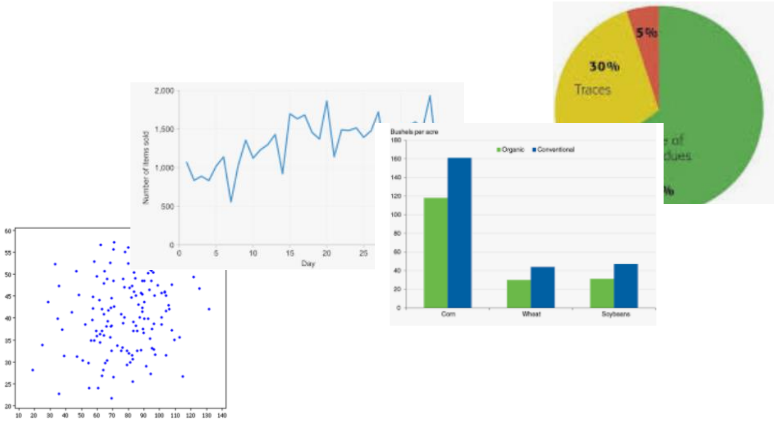
Text operations	Arithmetic operations	Date operations
SQL aggregate functions	Find missing data	Calculated fields with subqueries
Match text values	Match date criteria	Fields with missing data
Match record patterns with Like	Match rows with SQL aggregates	Match fields with subqueries
Update queries	SQL statements	

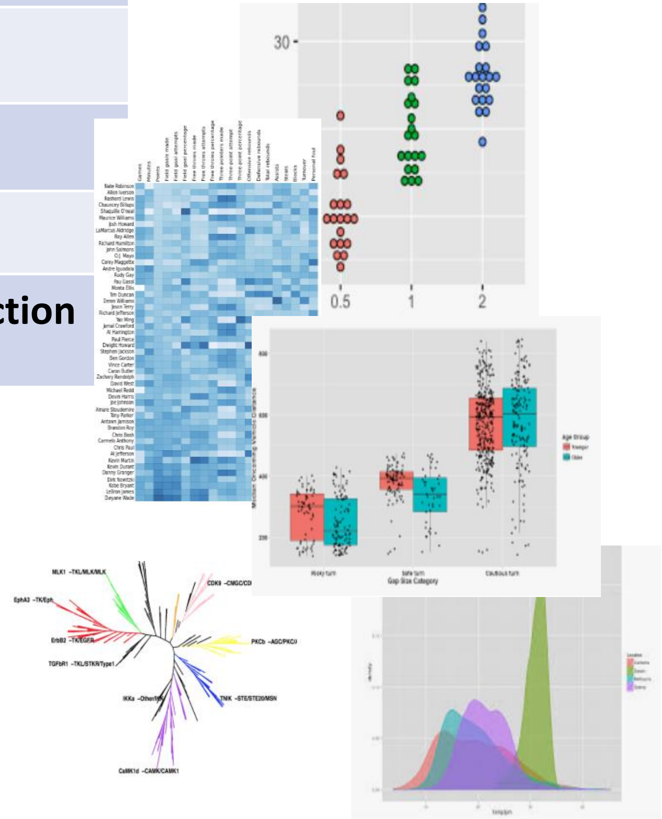
We are here:

- ❑ The big picture of data-driven operations.
  - ➢ Definition and depiction of a data-driven operation.
  - ➢ Cost-free "Critical-Mass" strategy for reaching data-drivenness.
  - ➢ Jargon of data science reduced to relevance to data-drivenness.
- ❑ Structure of methodologies.
  - ➢ "R"—as the analytic core of data-driven capability.
  - ➢ Gather, join, cleanse and form data into super tables.
  - ➢ **Layered charting in contrast to conventional charting.**
  - ➢ Types of Insight deliverables.
- ❑ Generalized implementation plan.
- ❑ Library of what-to and how-to papers, presentations and texts.

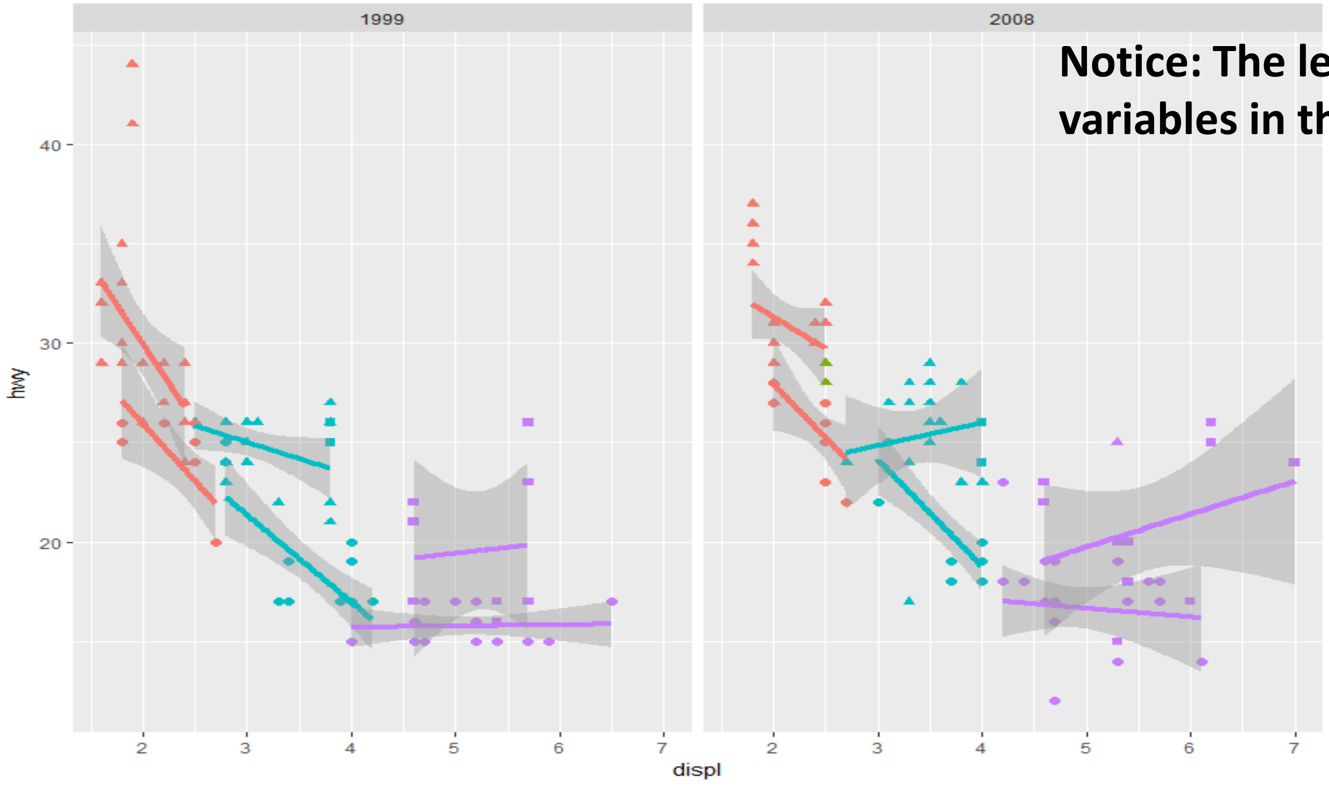
Conventional charts have been the staple since they were invented as early as the 1600s—now layered charting with ggplot2 in R allows perspectives we could not have before

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

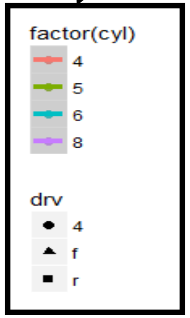




With this example you can begin to imagine how you would explore the interplay of KPIs—to name one idea



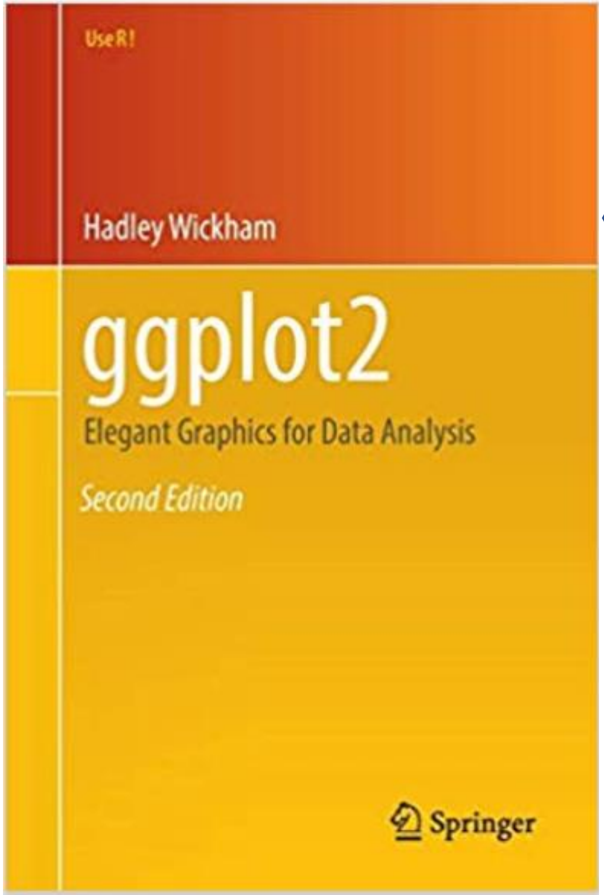
Notice: The legends are variables in their own right.



In this case:

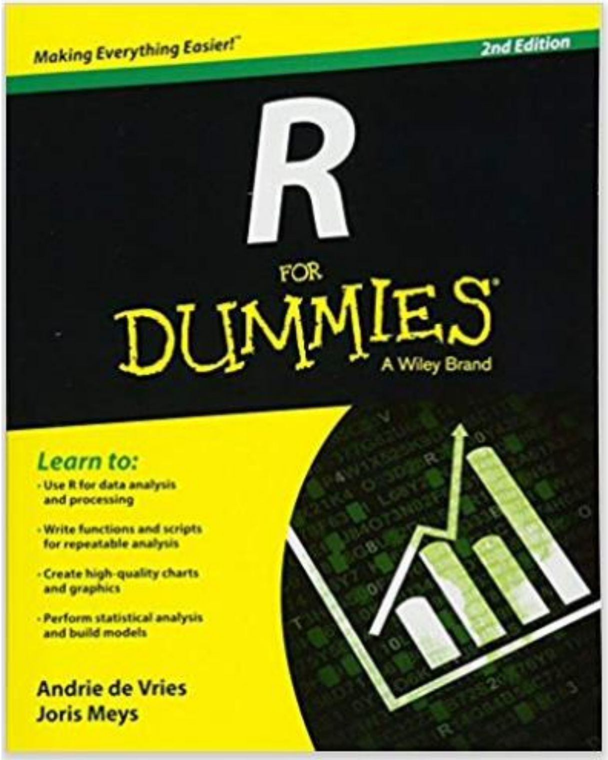
- ➢ **Interplay of five variables: highway mileage, displacement, cylinders, drive and years.**
- ➢ **Three perspectives are layered on the five variables: cross plot, linear fit and confidence intervals.**

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.

But first, you will need the basic skills to code with R—thence, all recipes will make sense.



We are here:

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

None, one or more of four types of insight deliverables may be relevant to 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.
Indicative	Insight gained upon data flowing through ML/AI models that ask and answer questions of relationship, difference, time series, duration and apparency.

Next

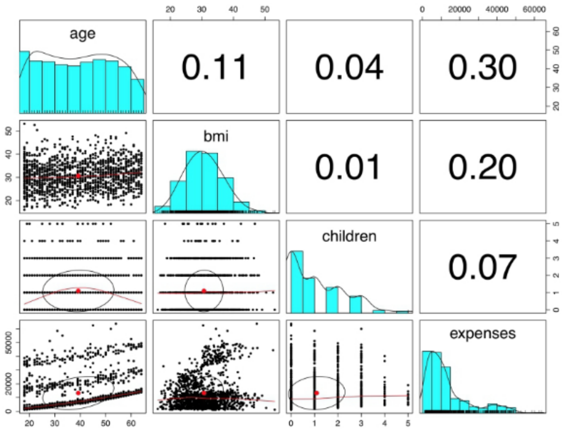
Know-thy-data, recountive and indicative insight deliverables will be explained. System reports require no further explanation.

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.
Indicative	All are generated with “R” functions.

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

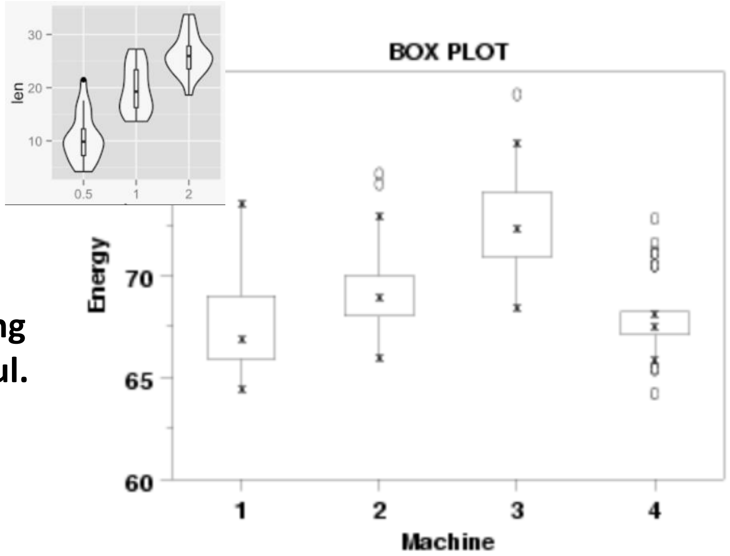
Unified view of seven insights into the nature and cross relationship of variables.




```
> 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
```

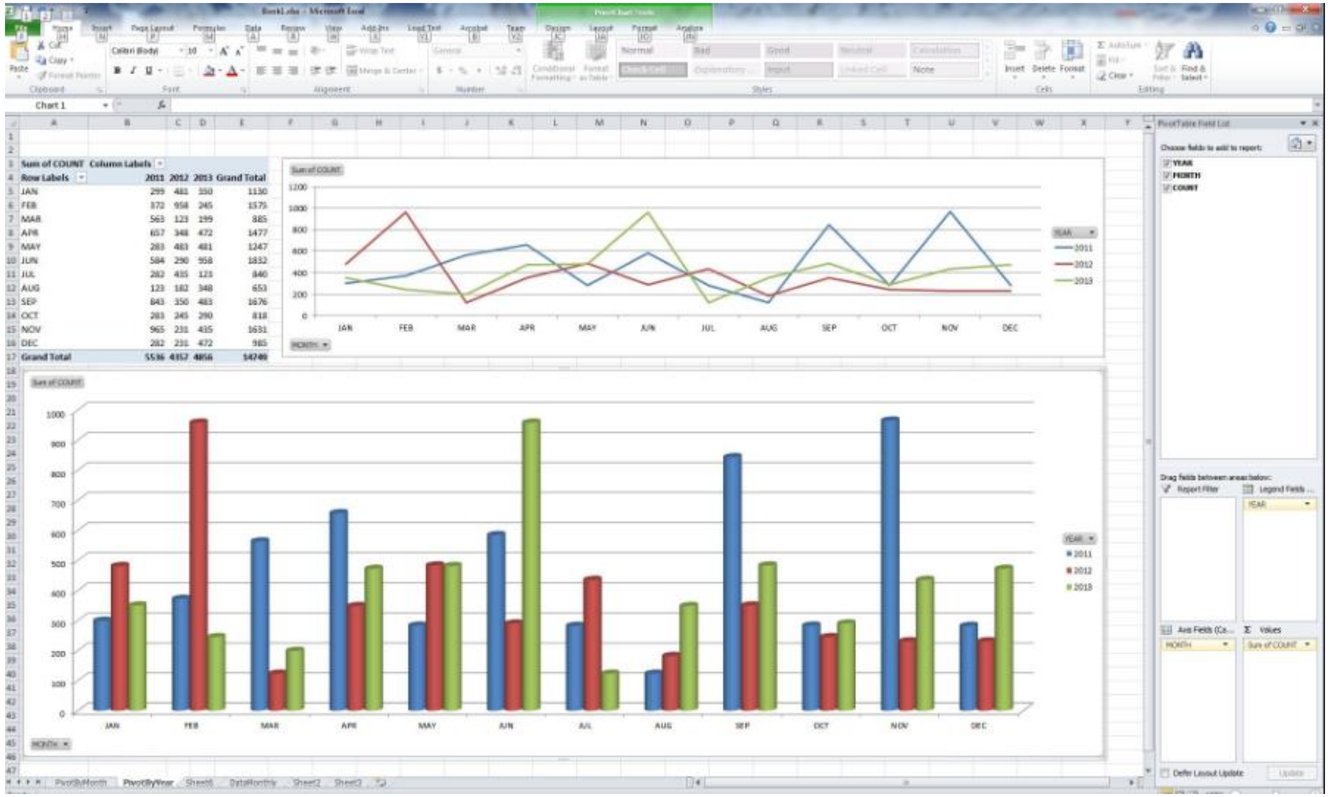
Min/max, median, mean, quartiles, categories and counts, and missing



Shapes and plotting styles are insightful.

- Know-thy-data is standard procedure to building and cleansing super tables.
- Shown are only a portion of possibilities; including aggregate variables.

"**Recountive**" insight deliverables report what happened, but without analytical perspectives



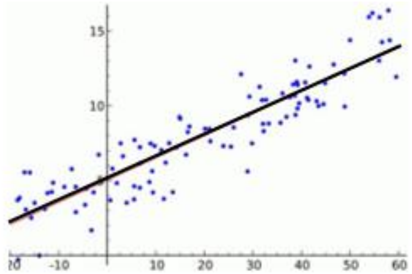
For most operations this is were they are, at best, if they are not practicing the framework and methodologies of this presentation.

For "Indicative" insight deliverables, each of the five types of questions map to one or more ML/AI analytic model

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 and K-mean.

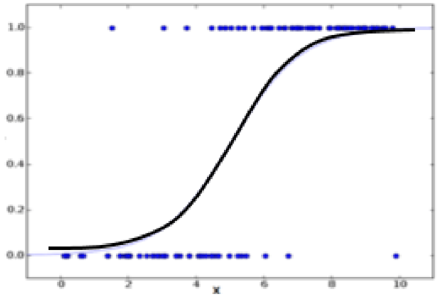
(1) A set of explanatory papers to the models can be downloaded at https://analytics4strategy.com/new-age-five-questions

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



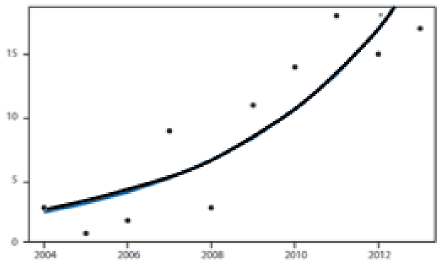
Linear regression:

- 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:

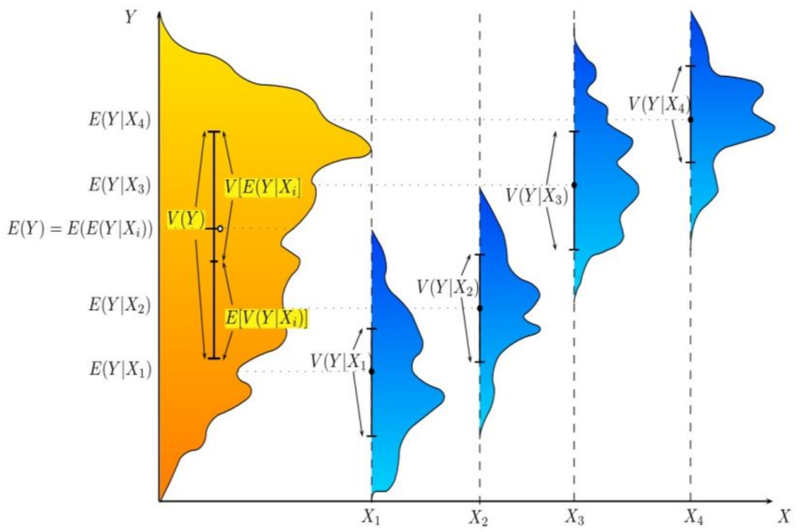
- Classifies between possible categorical outcomes.
- Binomial: Yes/No and multinomial: yes, no, maybe.
- Classification based on the probability of an outcome.
- Example question: Which asset and process variables most foretell the probability of finding crews engaged in non-value work?



Poisson regression:

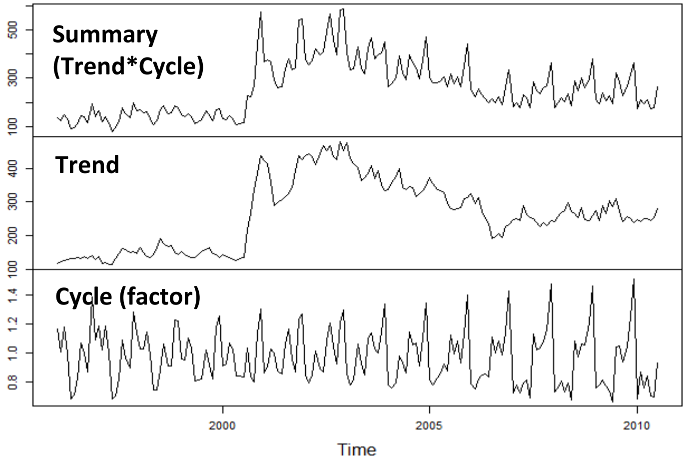
- Outcomes of interest are occurrences—e.g., failures and emergency orders.
- Scored as counts (failures) or rates (failures per month).
- Example question: Which variables are most related to the count and rate of specific process non-compliances?

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



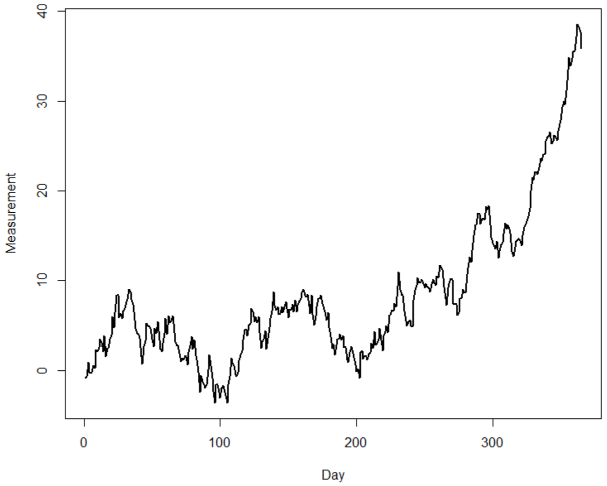
- **Explanation:**
  - A data set made up of all variables in a model has a mean (expected value)—left-most.
  - Grouping one or more variables results in a mean for each group.
  - Difference models are necessary because simple pair-wise statistical comparisons are misinformation due to the math of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

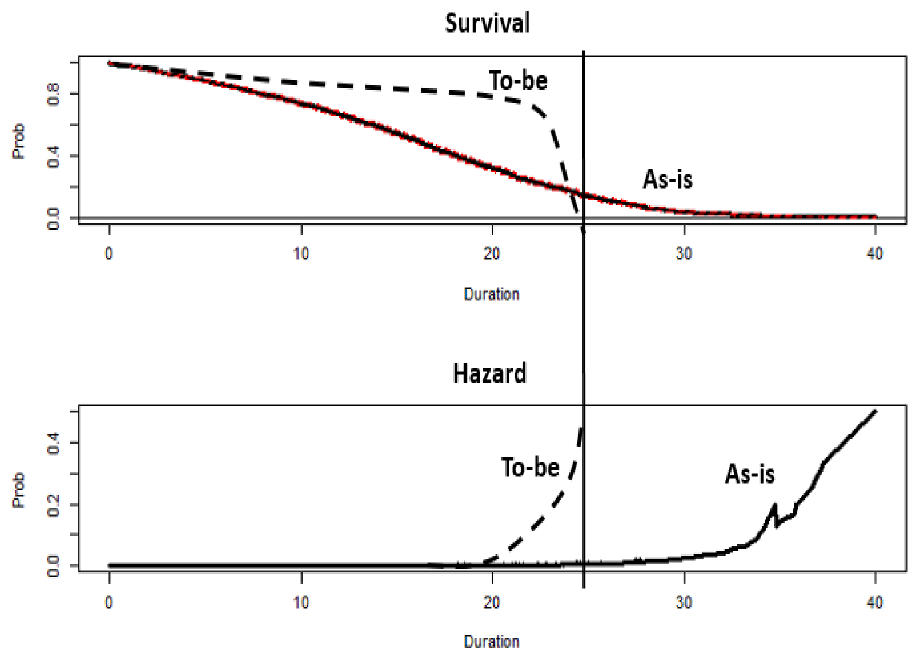


Just as important, what is the trend—deterministic, random or random walk?

Note:
Time series analytics play heavily in analytics upon asset readings.

Example question:
Is the trend for productivity changing with time?

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?



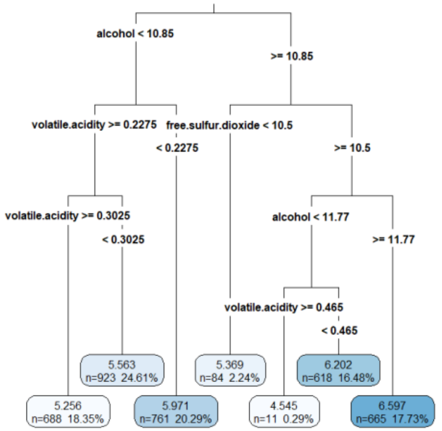
- Shows the probability of a condition lasting up to just before an ending event.
- Shows the probability of an ending event as a function of how long the condition has existed.
- Mathematics of "hazard" make it possible to determine which process variables matter most to the story.

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?

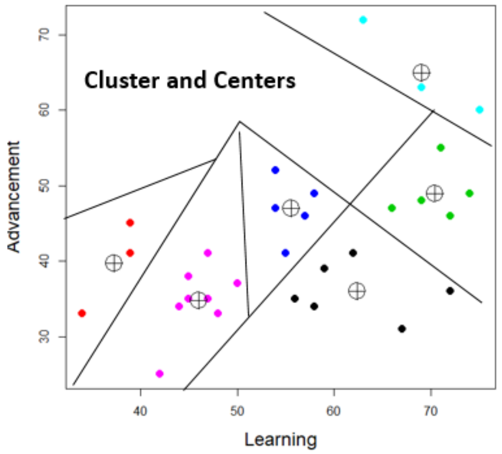


Decision tree:

- 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 mobile-friendly papers located on the website page https://analytics4strategy.com/new-age-five-questions

Question	Paper (linked)
Relationship	Find What Matters with Relationship Questions of Operations
Difference	Know that Improvements Work by Asking Difference Questions
Time series	Explore What Did and May Happen with Time Series Questions
Duration	Find the Time That is Money by Asking Duration Questions
Apparency	Dive Below the Surface of Process Functioning with Apparency Questions

To download pdf: Scroll to bottom of paper

We are here:

- The big picture of data-driven operations.
  - Definition and depiction of a data-driven operation.
  - Cost-free “Critical-Mass” strategy for reaching data-drivenness.
  - Jargon of data science reduced to relevance to data-drivenness.
- Structure of methodologies.
  - “R”—as the analytic core of data-driven capability.
  - Gather, join and cleanse data, and form super tables.
  - Layered charting in contrast to conventional charting.
  - Types of Insight deliverables.
- **Generalized implementation plan.**
- Library of what-to and how-to 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 | **Timeline to General Abilities**

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.

Leadership is prepared to make go/no-go decision, set direction and make initial project decisions. Project members selected and prepared to participate. Through lens of new knowledge, competitive North Star set to becoming data-driven is set.

Stage 2: Conduct basic design.

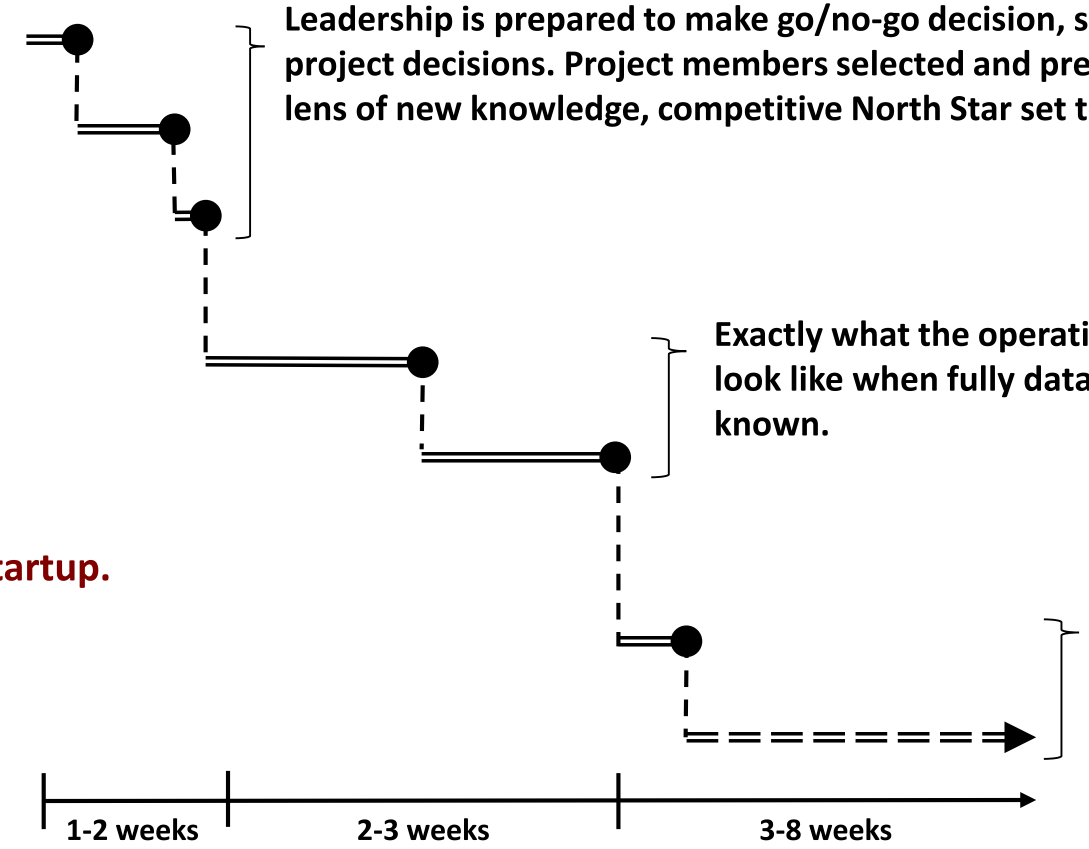
1. Charted, detailed operational processes for data-drivenness.
2. Map deliverables to charted process & form basic designs

Exactly what the operational processes will look like when fully data-driven is made known.

Stage 3: Plan and form detailed design and startup.

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

Moving to defined processes, prioritized upon influencing considerations.



NOTE: Project details to each stage are available from the unabridged version of this session (analytics4strategy.com >> Training Sessions >> "First Step to. . .")

Sample

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.

We are here:

- ❑ The big picture of data-driven operations.
  - ➢ Definition and depiction of a data-driven operation.
  - ➢ Cost-free "Critical-Mass" strategy for reaching data-drivenness.
  - ➢ Jargon of data science reduced to relevance to data-drivenness.
- ❑ Structure of methodologies.
  - ➢ "R"—as the analytic core of data-driven capability.
  - ➢ Gather, join and cleanse data, and form super tables.
  - ➢ Layered charting in contrast to conventional charting.
  - ➢ Types of Insight deliverables.
- ❑ Generalized implementation plan.
- ❑ **Library of what-to and how-to 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 hands-on	None available	

Richard G. Lamb, PE, CPA, ICBB
Tel: 832-710-0755; Email: rchrd.lamb@gmail.com

Website (educational): https://analytics4strategy.com/

Download slides at: https://analytics4strategy.com/datadriventrc2019

Unabridged version: analytics4strategy.com >> Training Sessions >> Scroll to "First Step to. . ." >> Download slides at bottom