The
First Step to Becoming a Data-Driven Operation
is to
Get a Clear, Implementable Understanding of Data-Drivenness

Training Session for Competency

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

Agenda:

- ❑ Purpose of the training session.
- ❑ The big picture of data-driven operations.
  - ➢ Definition and depiction of data-driven operational processes.
  - ➢ The 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.
  - ➢ Eliminate fused data from operations
  - ➢ Layered charting in contrast to conventional charting.
  - ➢ The primary types of Insight deliverables—system reports, know-thy-data, recountive and modeled.
- ❑ Generalized implementation plan.
  - ➢ Big picture of implementation.
  - ➢ People—the central issue for development.
  - ➢ Stages, steps, actions and deliverables.
- ❑ Library of what-to and how-to papers, presentations and texts.

Training purpose:

The first step to becoming a fully data-driven operation is that process role holders must reach a clear, implementable understanding of data-drivenness. With the understanding, the audience organization will be able to layout their implementation plan to reach the vision and as the plan unfolds be ready to absorb, hands-on, the introduced methodologies.

The purpose of the half-day training session is to be the first step.

Agenda:

- ❑ Purpose of the training session.
- ❑ The big picture of data-driven operations.
  - ➢ **Definition and depiction of data-driven operational processes.**
  - ➢ The 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.
  - ➢ Eliminate fused data from operations
  - ➢ Layered charting in contrast to conventional charting.
  - ➢ The primary types of Insight deliverables—system reports, know-thy-data, recountive and modeled.
- ❑ Generalized implementation plan.
  - ➢ Big picture of implementation.
  - ➢ People—the central issue for development.
  - ➢ Stages, steps, actions and deliverables.
- ❑ Library of what-to and how-to papers, presentations and texts.

Data-driven defined:

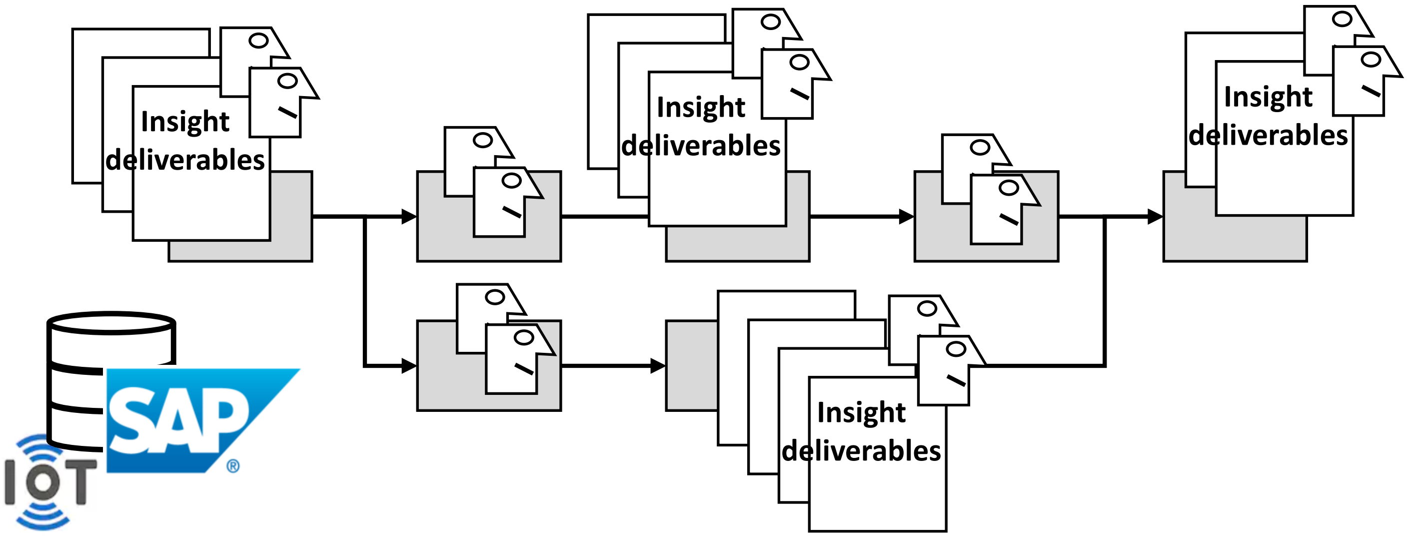
A "Data-driven" operation is defined as one that harnesses its operational data to augment the experience and judgement of operatives, managers, analysts and engineers as they plan, organize, conduct and control their processes.

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 many places, the "best outcomes" can only be realized when the experience and judgement of the process role players are augmented with "insight deliverables"—system reports, know-thy-data, recountive and modeled.

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



Agenda:

- ❑ Purpose of the training session.
- ❑ The big picture of data-driven operations.
  - ➢ Definition and depiction of data-driven operational processes.
  - ➢ **The 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.
  - ➢ Eliminate fused data from operations
  - ➢ Layered charting in contrast to conventional charting.
  - ➢ The primary types of Insight deliverables—system reports, know-thy-data, recountive and modeled.
- ❑ Generalized implementation plan.
  - ➢ Big picture of implementation.
  - ➢ People—the central issue for development.
  - ➢ Stages, steps, actions and deliverables.
- ❑ Library of what-to and how-to papers, presentations and texts.

Operations can take a “critical-mass” strategy because data-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 are largely inclusive; travel to up-teched and up-scaled strategies.

  - Up-teching from critical-mass software will not practicably increase the power of the insight that is extracted from the operation’s data.

  - Can be the grassroot of change culminating in global organizational ability.

There are two dimensions of organizational ability that emerge as operations select processes to make data-driven

Knowledge and skills. . .

Are gained as the first and subsequently selected processes are brought to be data-driven.

Note: Any implementation plan must include very visible steps that will transfer the knowledge and skills to the implementation team and to the operatives, managers and experts along the subject operational processes.

Software. . .

The triad of R, MS Excel and MS Access collective possess the functionality to conduct all work actions of data-drivenness.

Note: Throughout the training session, the knowledge and skills of data-driven will be demonstrated in the context of the triad.

All organizations have immediately available to them—without issues and cost—the triad of software to becoming data-driven

- The triad of MS Excel, MS Access and R as critical-mass" was previously introduced.
- There are no system issues and constraints to software because all are currently standard to operations (Excel, Access) and one is a free to download open-systems (R).
- Every seat along the operational process can be armed with the triad, rather than be bottlenecked by financial barriers.
- Skills of engaging the triad at every seat. . .
  - Many are already ubiquitous from standard operating systems and Excel in almost every job.
  - Ubiquitous skills predispose individuals to snap to new skills.
  - New skills are gained as projects to build data-driven operations unfold.

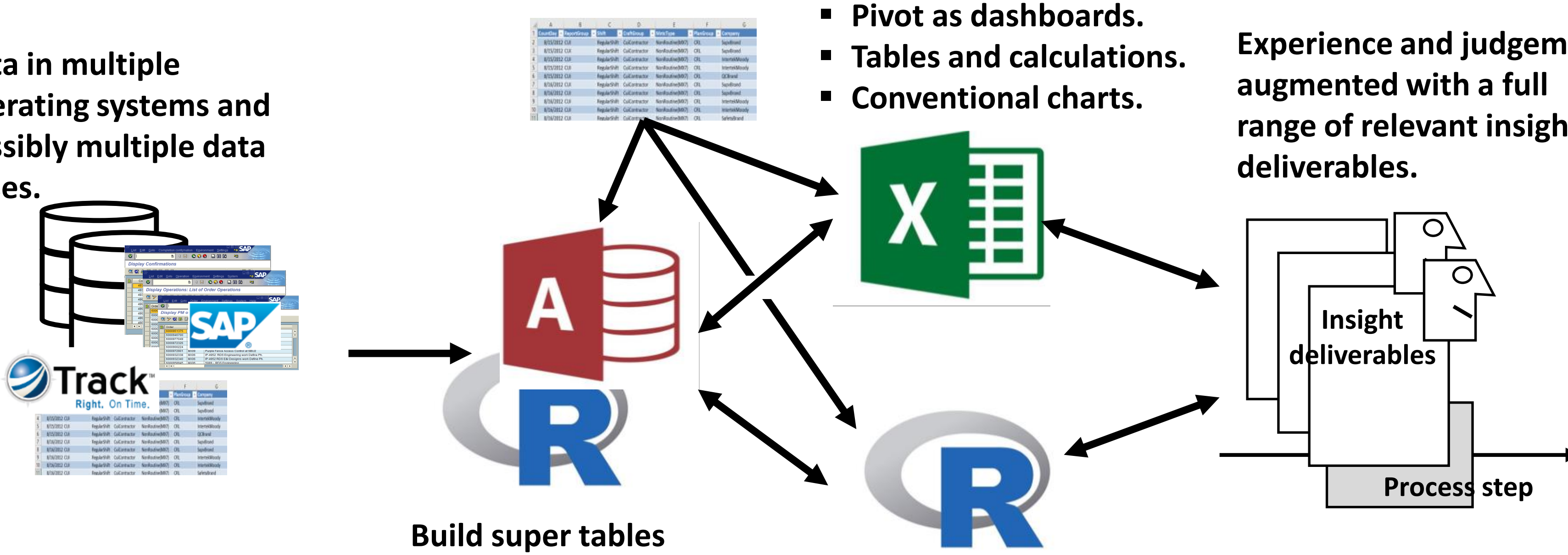
The triad maps to the stages from sourcing data and building super tables through to functioning insight deliverables

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 is not a choice of technology and IT, but a choice to take the most direct path to becoming data-driven and most widely so

- **Immediately starts the organization toward becoming data-driven without the typical delays of organizational impediments natural to up-teching and new capabilities.**

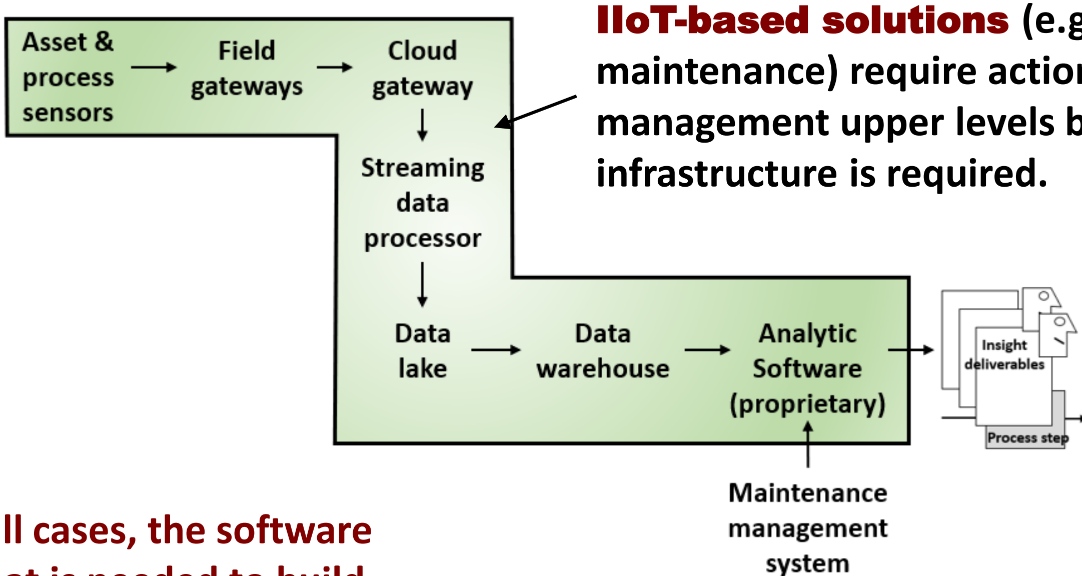
- **Universality of knowledge and skills in design and function—gained beginning at the first touch by role holders—transfer to up-teched choices.**

- **Ubiquity of the critical-mass software will surround the always limited feasible dispersal of up-teched elements of the data-driven capability; retaining capability at all seats.**

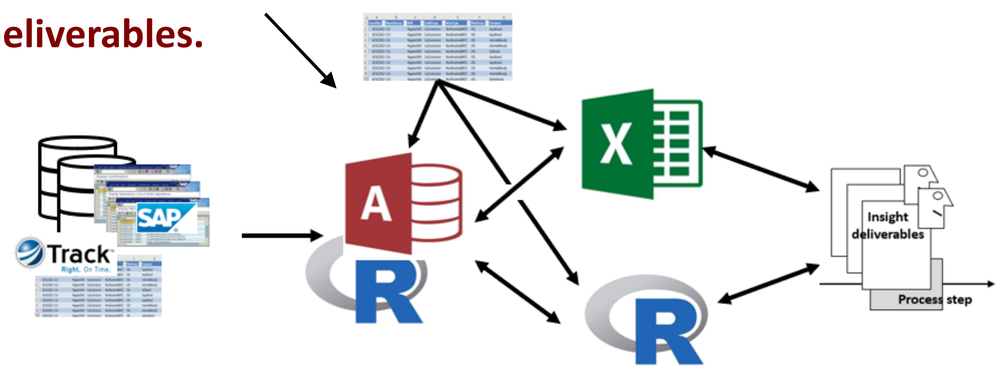
- **R is a software we will see elsewhere: corporate IT strategies and commercial software are interfacing to R for the strongest of analytics (e.g., Tableau) and layered charting (e.g., Power BI).**

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

IIoT-based solutions (e.g., predictive maintenance) require action down from management upper levels because new IT infrastructure is required.



For almost all cases, the software triad is all that is needed to build envisioned insight deliverables.



The critical-mass triad is not limited in its power for insight, but reasons for targeted up-teching may emerge

- **Globalization:**
  - Parent firm wishes to converge on a particular software from all reaches.
  - What is learned and built locally is to be adopted globally.
- **Functionality:**
  - E.g., mobile-friendly interactive dashboards, real-time alerts, non-structured data, etc.

Note:
When targeted up-teching is the case, expect that the triad will still dominate.

Agenda:

- ❑ Purpose of the training session.
- ❑ The big picture of data-driven operations.
  - ➢ Definition and depiction of data-driven operational processes.
  - ➢ The 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.
  - ➢ Eliminate fused data from operations
  - ➢ Layered charting in contrast to conventional charting.
  - ➢ The primary types of Insight deliverables—system reports, know-thy-data, recountive and modeled.
- ❑ Generalized implementation plan.
  - ➢ Big picture of implementation.
  - ➢ People—the central issue for development.
  - ➢ Stages, steps, actions and deliverables.
- ❑ Library of what-to and how-to papers, presentations and texts.

The typically over-hyped jargon of data-science does not tell the story of data-drivenness because they are the pieces to the whole or a big vision

- **Jargon needing to be demystified:**
  - Data and big data
  - Machine learning
  - Artificial intelligence
  - Algorithms
  - Models
  - IoT
  - Visualization
  - Platforms
  - Digitalization
  - Intelligence
- **All reside behind the curtains of data-drivenness.**
- **It is good to frame the meaning of all terms.**
  - Data and big data—huge distinction.
  - Analytic core to data-drivenness—machine learning, artificial intelligence and algorithm as models.
  - Visualization—a rename.
  - IoT, Digitalization and intelligence—open-ended, umbrella-like visions.
  - Platform—leave it to the geeks.

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.
- Media tends to report stories of "data analytics" as stories of "big data", but most are actually stories of analytics-augmented breakthroughs.
- "Big data" is the case in which data is massive or disparate (e-mail, document, video, photo, audio, webpage) to structured data.
  - Litmus test—the data or the analytics of the subject data cannot be processed on our notebook computers.
  - Entails high-tech systems, specialized skills and substantial organizational costs.
- The data analytics conducted in either arena are 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 one type of model—linear regression

- Cases (rows) are selected from a table of prepared data.
- One or more variables are selected as "predictors" and another as "outcome" to a model of a type to answer some particular types of questions.
- The cases and variables are fed to the model and its gut **algorithm** conducts a trial-and-error calculation until "learning" the best fit.
- Because our computer is a machine, the trial-and-error to arrive at the fit constitutes "**machine learning**."
- 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.
- "Learning" may be a better word because learning is still a human and machine iteration.
  - Human: Determine the variables to include, make model selection and design decisions, and evaluate the fit for legitimacy.
  - Machine: Algorithm as calculation to a model, converges on a fit to the chosen type of model.

“**Artificial intelligence**” is to feed a fresh set of cases to the learned model and be “advised” of the outcome to be expected for each case

- **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.**

- **When AI is to be the case, the learning process should take an additional layer.**
  - **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—or augment human judgement.**

- **Values for cases upon which we wish to be advised, are fed to the learned model and we receive a predicted or forecasted outcomes.**

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

Definitions for the remaining jargon items of data science

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.
Platforms	Places in the information system at which take place the aspects of functioning such as connectivity, data management, computational software (e.g., MS Office) and analytics (models).

Agenda:

- ❑ Purpose of the training session.
- ❑ The big picture of data-driven operations.
  - ➢ Definition and depiction of data-driven operational processes.
  - ➢ The 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.
  - ➢ Eliminate fused data from operations
  - ➢ Layered charting in contrast to conventional charting.
  - ➢ The primary types of Insight deliverables—system reports, know-thy-data, recountive and modeled.
- ❑ Generalized implementation plan.
  - ➢ Big picture of implementation.
  - ➢ People—the central issue for development.
  - ➢ Stages, steps, actions and deliverables.
- ❑ Library of what-to and how-to papers, presentations and texts.

Section training purpose:

To establish a foundational understanding of "R" as the analytic core to data-driven operations—where to download it, what it looks like and how to work it.

Note:

Along the way, all players in the implementation and functioning of data-driven operations will become competent in R, as well as, all other methodologies (see generalized implementation plan).

The foundation technology to being data-driven is an analytic software

- **Only through an analytic software can there be descriptive statistics, layered charting, data cleansing and ML/AI-based models.**

- **Without a data analytic software in the structure for data-driven operations; the organization will never qualify as data-driven.**

- **There are two regarded system leaders:**
  - **R—open system.**
  - **SAS—commercial system.**

- **As mentioned before, the critical-mass strategy is built upon the openness of "R."**

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

[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

The R Project for Statistical Computing

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 ;

Some aspects of R that you should know that makes it such a powerful force to achieving data-drivenness

- **Each type of analytic is a package (currently over 10,000) of functions with arguments:**
  - **Developed and managed by 1,000s of individuals and institutions around the globe.**
  - **All packages are required to be accompanied with a full explanation and examples with data.**
- **Online support is highly evolved:**
  - **Online detail to the packages and their functions—"R" pages and blogs.**
  - **Huge online community provides help and examples—blogs.**
- **Texts on the many data and analytic methodologies are plentiful in which R is used to demonstrate them.**
- **Through R's interface functions, some commercial software (e.g., Tableau) build top-tier analytics in their offerings and some firms install analytics in their corporate IT system.**

“R” scares us (did me) at first because we are accustomed to GUI while “R” is coded interface— actually making it much more process- and task-friendly

- **As tasks along a process, GUI requires prohibitively long documented instructions.**
  - **Writing process instructions can be a major task, thus, invitation to failure.**
  - **Users must work with “one finger” on the instructions switching eyes back and forth from instruction to monitor.**

- **Code-based software only requires the user to open “R,” open or paste a txt-type file (called script) to the ”R” interface and press the run command—often the only required skill level for most process tasks.**

- **Explanations can be coded into the script to explain:**
  - **Where the task or tasks of the code takes place in the data-driven process.**
  - **What the code does or will generate.**
  - **How to interpret the generated insight deliverable.**
  - **Whatever.**

- **Scripts can be located in a shared place and role holders can open the script from their personal computers.**

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

- Our "biggest" initial challenge is to identify the functions to our desired insight deliverable—working the functions is not a significant skill issue.

- 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, ...)
```

- Each function and its arguments are explained in depth via the "R" website or the on-line community.
  - Coding "?lm" pulls up an R page.
  - Entering "lm() r" on the internet pulls multiple pages to chose from.
  - In both sources, examples are provided along with demo datasets.

- Code (as shown above) at the "R" website or on-line community pages can be cut and pasted to our scripts and modified as we write them—if we wish to.

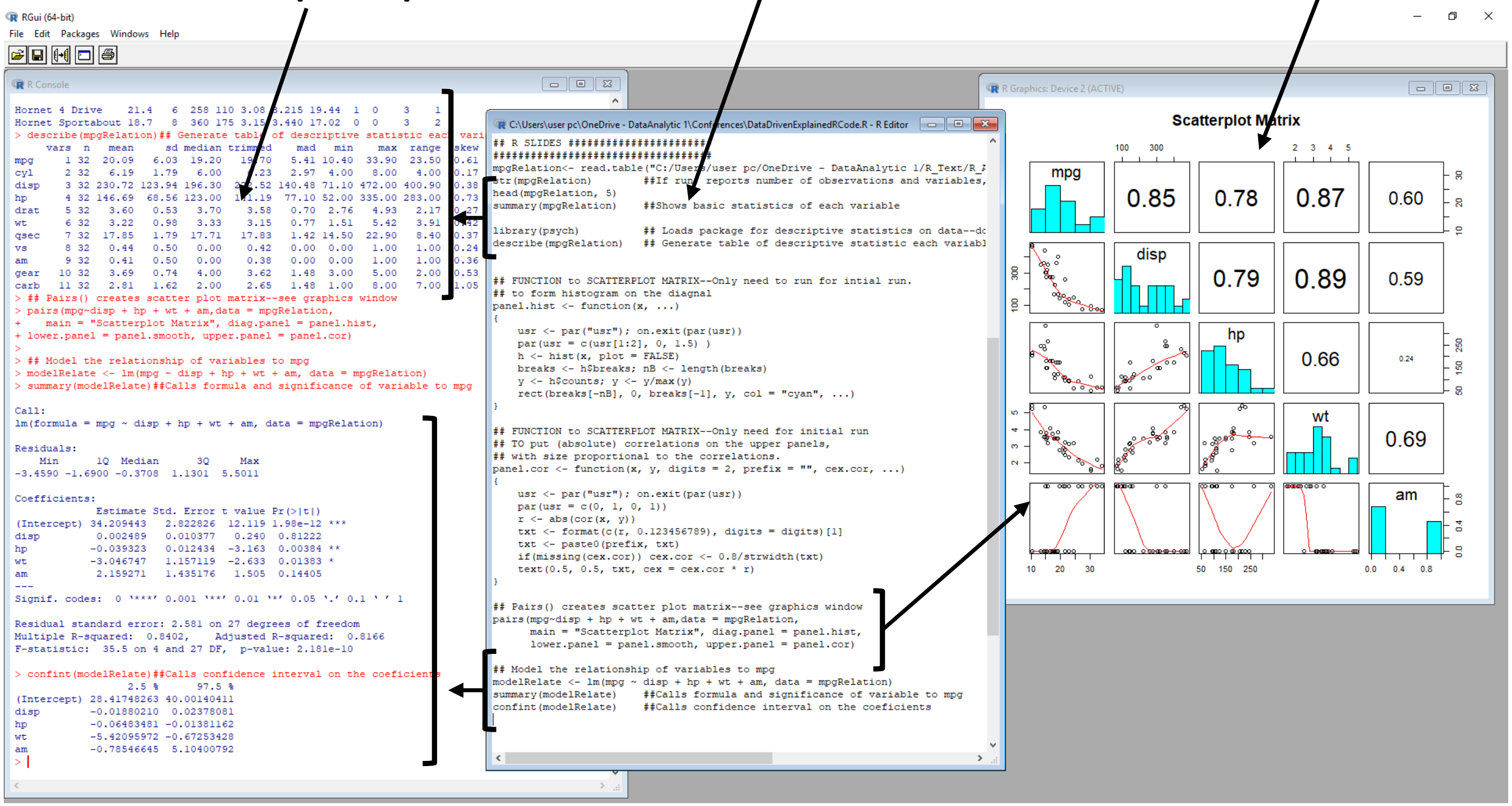
The frontend view of the "R" software

Run button (ctrl R)

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.



We see the coded elements of an insight deliverable—running the code imports the super table, explores the data and generates the insight

Data imported from source, e.g., super table built in Access.

Functions to explore the super table—know-thy-data

Background functions to the pairs() function to follow—of no concern to the user other than must be run initially.

Note: The functions were cut-and-paste from the R site page that explained the pairs() function.

The lines preceded by # are explanations to the code's purpose, etc.

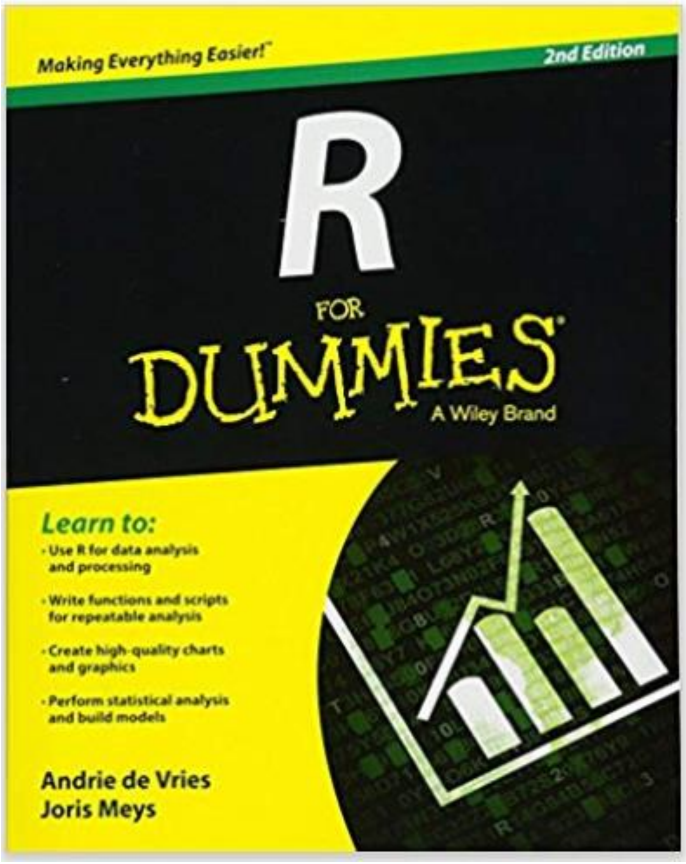
Function to generate the scatter-plot-matrix for know-thy-data.

Functions that generate the insight deliverable—a linear regression.

C:\Users\user pc\OneDrive - DataAnalytic 1\Conferences\DataDrivenExplainedRCode.R - R Editor

```
## R SLIDES ######################
##################################
mpgRelation<- read.table("C:/Users/user pc/OneDrive - DataAnalytic 1/R_Text/R_A
str(mpgRelation)         ##If run, reports number of observations and variables,
head(mpgRelation, 5)
summary(mpgRelation)     ##Shows basic statistics of each variable

library(psych)           ## Loads package for descriptive statistics on data--dc
describe(mpgRelation)    ## Generate table of descriptive statistic each variabl


## FUNCTION to SCATTERPLOT MATRIX--Only need to run for intial run.
## to form histogram on the diagnal
panel.hist <- function(x, ...)
{
    usr <- par("usr"); on.exit(par(usr))
    par(usr = c(usr[1:2], 0, 1.5) )
    h <- hist(x, plot = FALSE)
    breaks <- h$breaks; nB <- length(breaks)
    y <- h$counts; y <- y/max(y)
    rect(breaks[-nB], 0, breaks[-1], y, col = "cyan", ...)
}

## FUNCTION to SCATTERPLOT MATRIX--Only need for initial run
## TO put (absolute) correlations on the upper panels,
## with size proportional to the correlations.
panel.cor <- function(x, y, digits = 2, prefix = "", cex.cor, ...)
{
    usr <- par("usr"); on.exit(par(usr))
    par(usr = c(0, 1, 0, 1))
    r <- abs(cor(x, y))
    txt <- format(c(r, 0.123456789), digits = digits)[1]
    txt <- paste0(prefix, txt)
    if(missing(cex.cor)) cex.cor <- 0.8/strwidth(txt)
    text(0.5, 0.5, txt, cex = cex.cor * r)
}

## Pairs() creates scatter plot matrix--see graphics window
pairs(mpg~disp + hp + wt + am,data = mpgRelation,
      main = "Scatterplot Matrix", diag.panel = panel.hist,
      lower.panel = panel.smooth, upper.panel = panel.cor)

## Model the relationship of variables to mpg
modelRelate <- lm(mpg ~ disp + hp + wt + am, data = mpgRelation)
summary(modelRelate)     ##Calls formula and significance of variable to mpg
confint(modelRelate)     ##Calls confidence interval on the coeficients
```

If you are a self-directed learner who wants to get started, this will get you there



The easy-read-and-do text will make you competent enough to work with the texts to the methodologies explained in the subsequent subsections to methodologies.

Final point. . .

The analytic core will arise repeatedly in the methodology subsections to follow.

Agenda:

- ❑ Purpose of the training session.
- ❑ The big picture of data-driven operations.
  - ➢ Definition and depiction of data-driven operational processes.
  - ➢ The 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.**
  - ➢ Eliminate fused data from operations
  - ➢ Layered charting in contrast to conventional charting.
  - ➢ The primary types of Insight deliverables—system reports, know-thy-data, recountive and modeled.
- ❑ Generalized implementation plan.
  - ➢ Big picture of implementation.
  - ➢ People—the central issue for development.
  - ➢ Stages, steps, actions and deliverables.
- ❑ Library of what-to and how-to papers, presentations and texts.

Training Purpose for Section:

In a data-driven operation, not everyone needs to have the hands-on skills in building tables, but almost everyone must be able to participate in the discussion of data in routine operations and ad hoc needs.

According the purpose of the section is to give all attendees a conversational understanding of how data tables are built.

Training to the depth of hands-on skills is the purpose of the session titled, "Build Super Tables from Operational Data" (https://analytics4strategy.com/train-builddatatables)

There are three truths that, once you know of them, will send you down the path to build the super tables you always wanted but could never have

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

  When not,—e.g., status history in most computerized maintenance management systems—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 variable they have in common.**

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

- **Bad data is rarely a deal killer:**
  - **"Cleansing" the data often neutralizes the flaws.**
  - **Bad data is most often the result of compliance failures in the source operational process—immediate enforcement is the fix.**

    The first day of collecting good data soon becomes weeks, months and years of good data.

Why you would use **MS Access** to do your data work

- **Your firm already has rights to Access by virtue of its MS Office license—you only need to request it, if not already installed on all computers.**

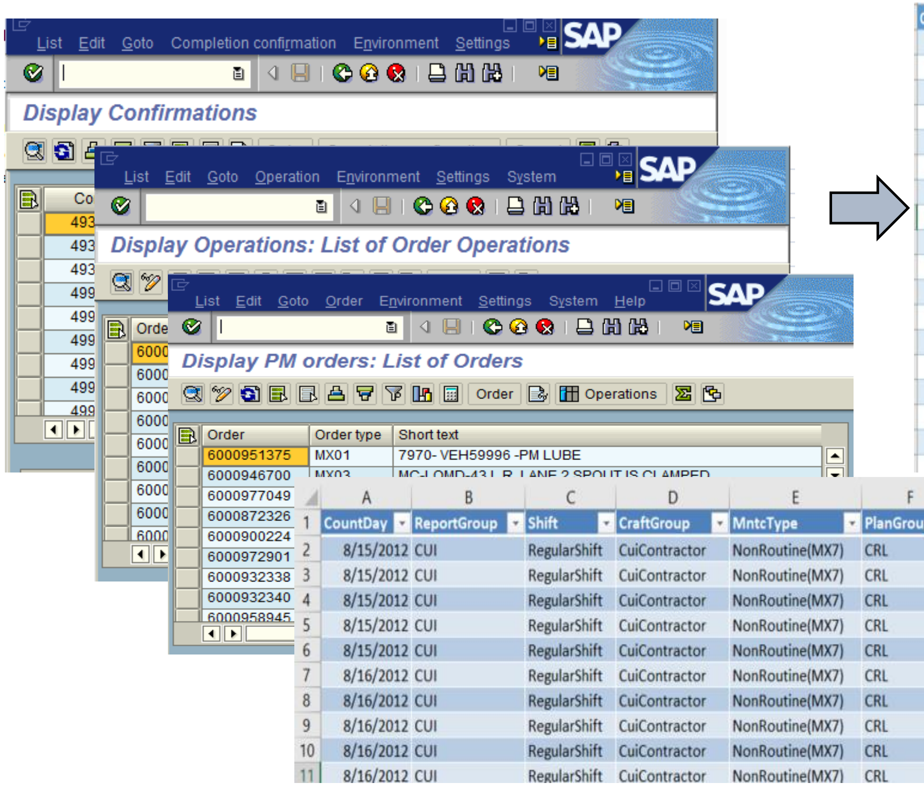
- **MS Access has the functionality to build super tables the same way as all table-building software—but has the shallowest of learning curves.**

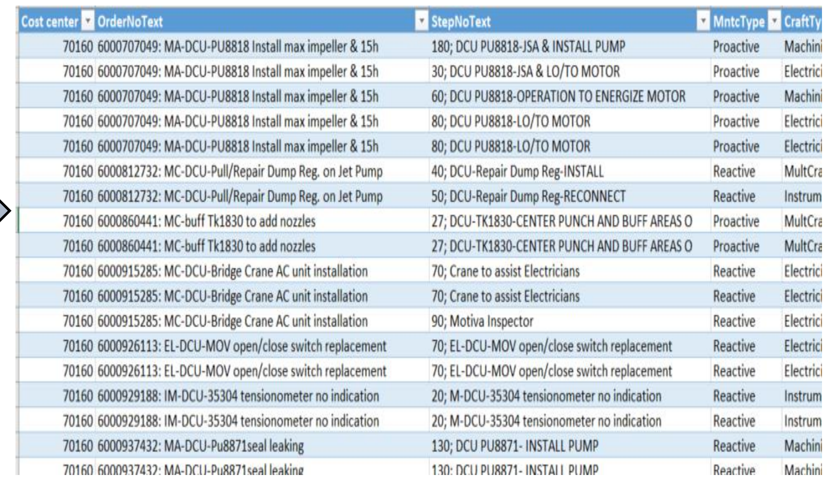
- **The knowledge and skills learned to build super tables in Access are universal, thus, transfer to other software (e.g., Tableau, Power BI).**

- **Because standard query language (SQL) codes in the background (query by example) of Access. . .**
  - **The need for SQL skills has been eliminated as an obstacle to incubating table-building skills across an organization.**
  - **But we still stay close to the bone of SQL—as its elements remain visible to and touched by us.**
  - **The necessary skills are largely those we all have from doing our current work with Excel and operating systems.**

Imagine everyone up and down your halls who normally works with Excel also becoming able to work with data—talk about a power jump!!

The goal is to extract topic-specific data from sources and fabricate a **super table** as required to build one or more specified insight deliverables





- 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 takes 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 →

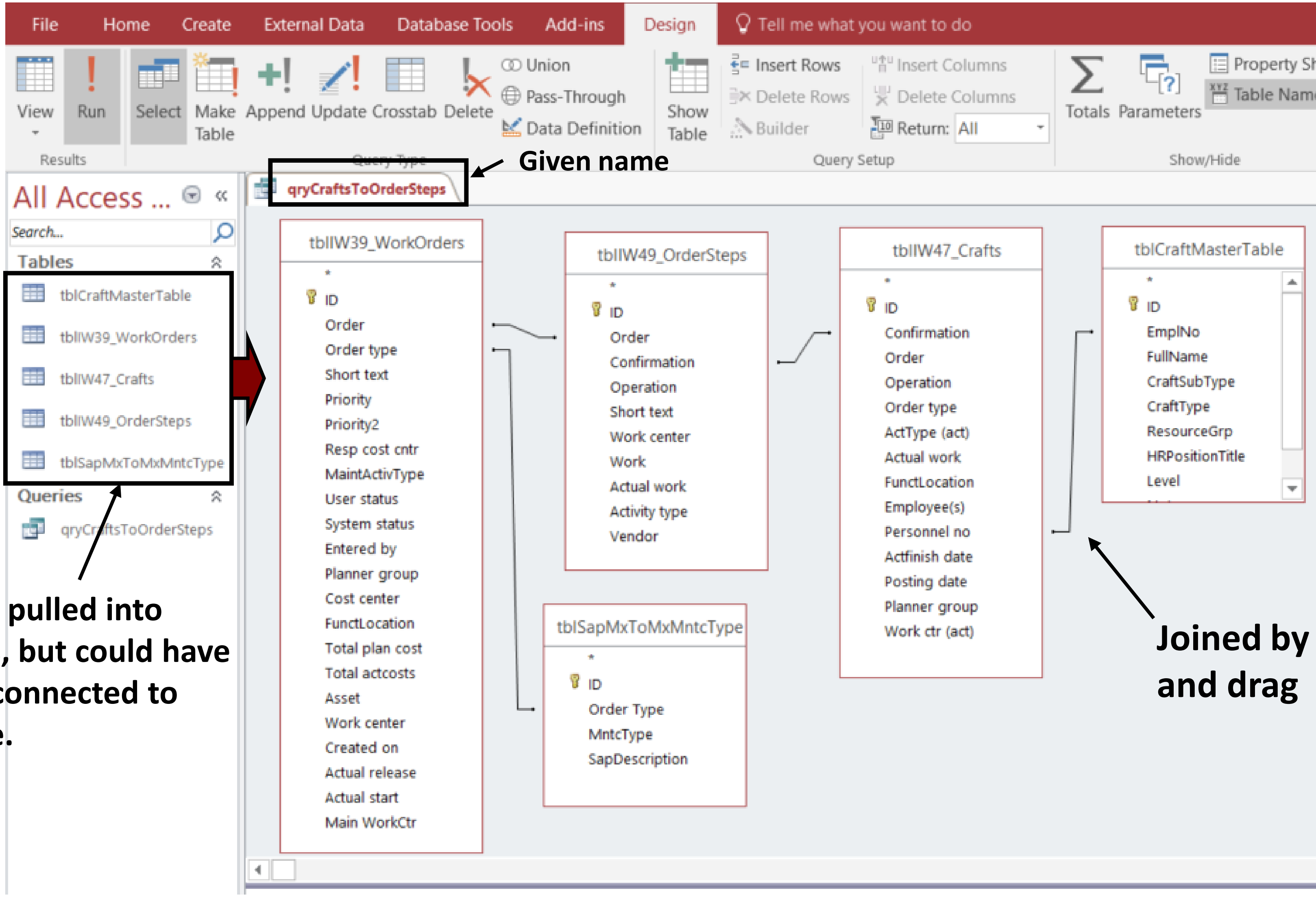
Step 4: Build super table with query functionality →

Step 5: Explore, cleanse the data →

Step 6: Build aggregate variables into super table →

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

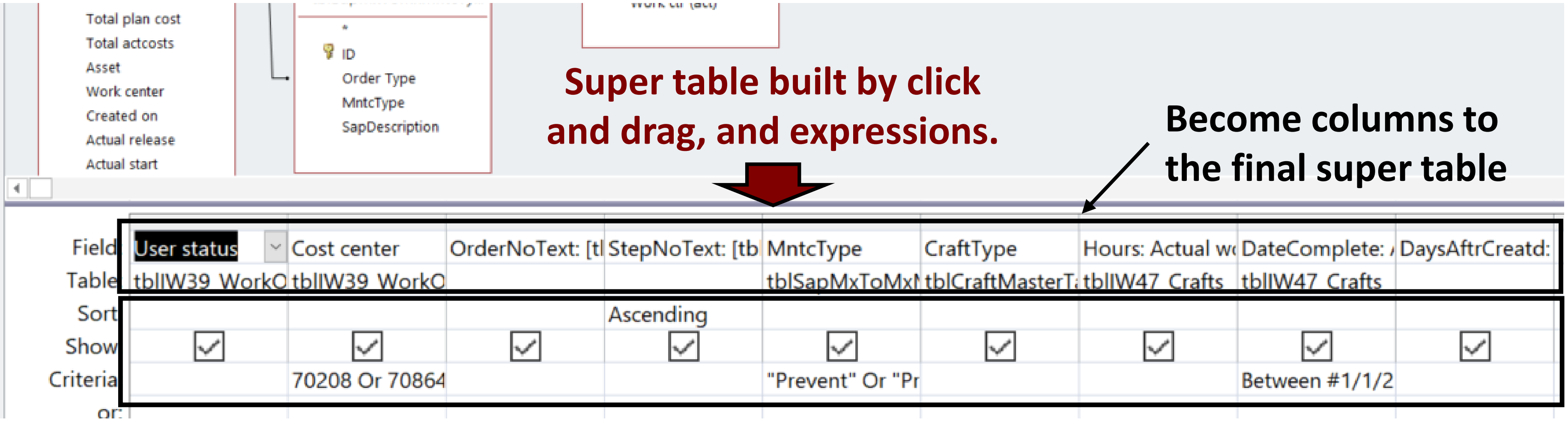
With a query, the subtables are joined by the variables they have in common; creating a grand table with all the shown imported variables



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

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

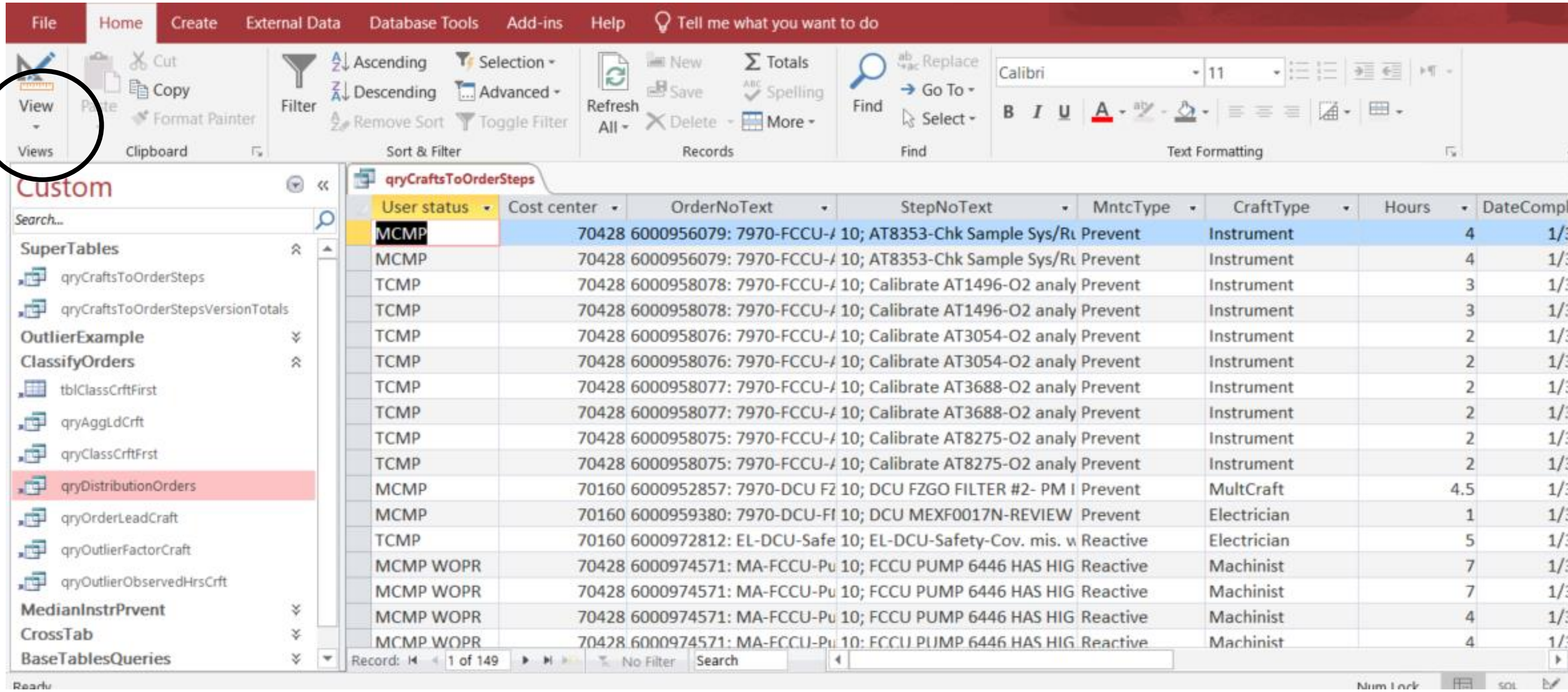


Super table built by click and drag, and expressions.

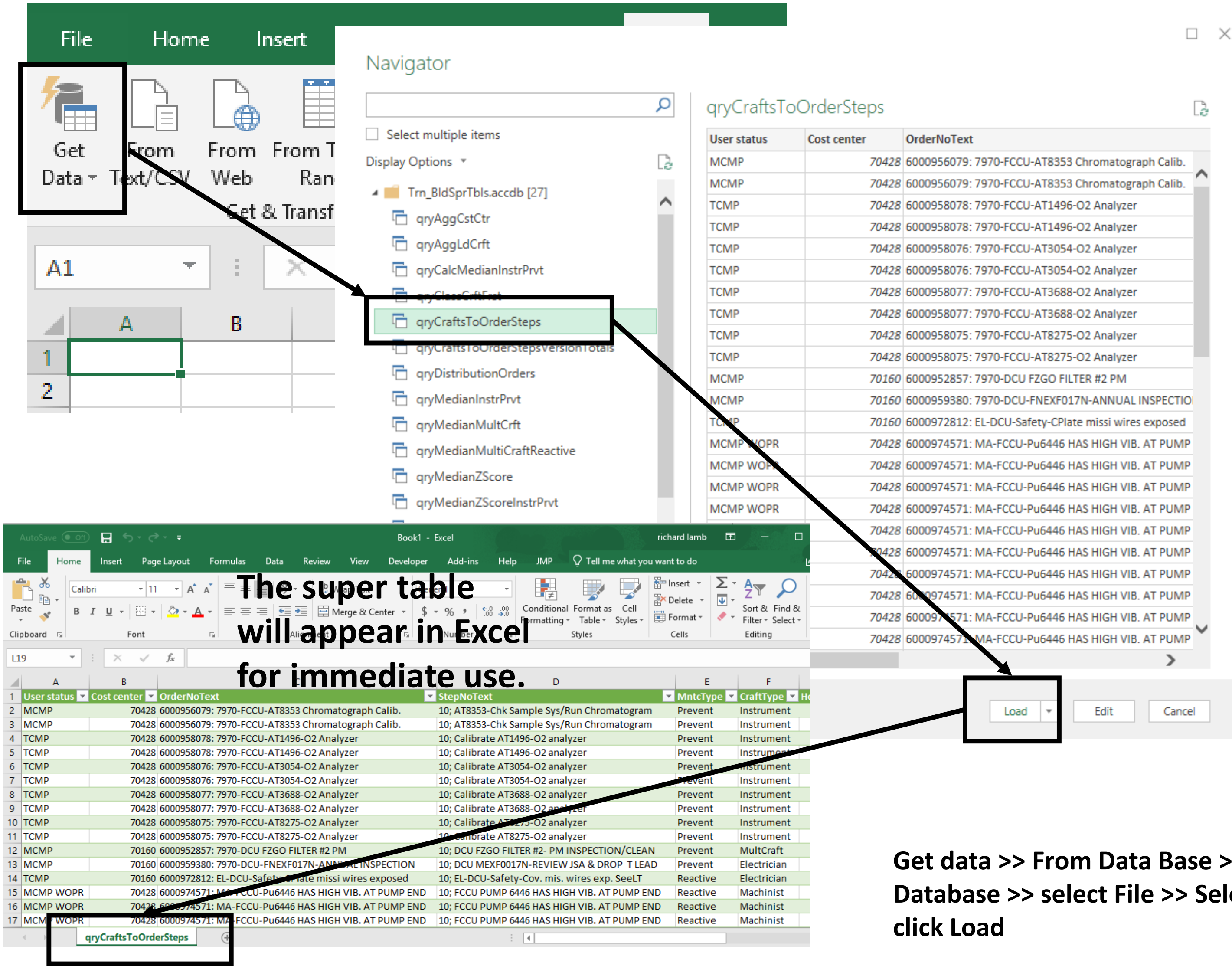
Become columns to the final super table

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

The materializing super table is viewed and explored back and forth between “Design” and “Table” view—a hugely insightful process in its own right



Bring the query into Excel by clicking the Excel "Get Data" button and following the path to select the query from Access

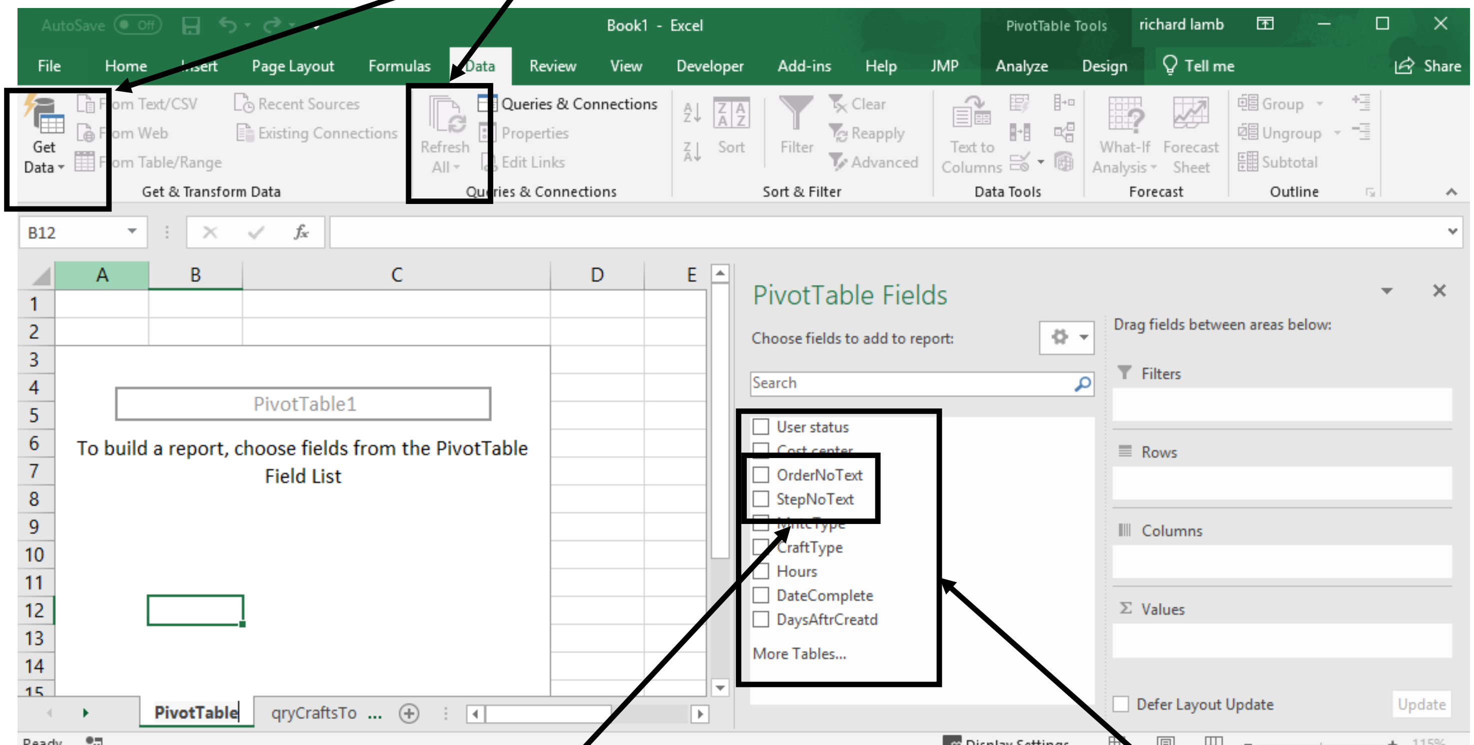


Note: Could have connected to the super table rather than extracted.

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

The super table can be made available to any insight deliverable—Pivots and data analytics—by connection or import

If make connection direct to query in Access, then can update the Pivot when source tables are updated or query is updated or upgraded.



- **A power of a super table is to give Pivots multiple pieces of information as a single-line field.**
- **In this case, order and step ID with their description—as a result of using the concatenation criteria, "&."**

The fields of the super table appear in the list of fields, to be dragged to the pivot areas for interactive slice-dice, drill-down and formatting

Comments on bad data and cleansing

- **The cleansing step is also a de facto evaluation of the source operational processes for compliance and weaknesses. Accordingly, cleansing reveals opportunities for impactful process improvements.**

- **The occurrence of bad data may be fading as firms update their operating systems and they, in turn, better control for work flow, format, and omission.**

- **When flow, format and omission are not enforced by old systems, replaced by new, the cleansing process may be a one-off exercise to the pre-modern era data.**

- **In maintenance operations, currently, two types of bad data are almost universal, because they are not automatically enforced by our systems.**

  - **Accurate allocation of craft hours to work assigned work—fix the timesheet process because hours are the life's blood to maintenance management.**

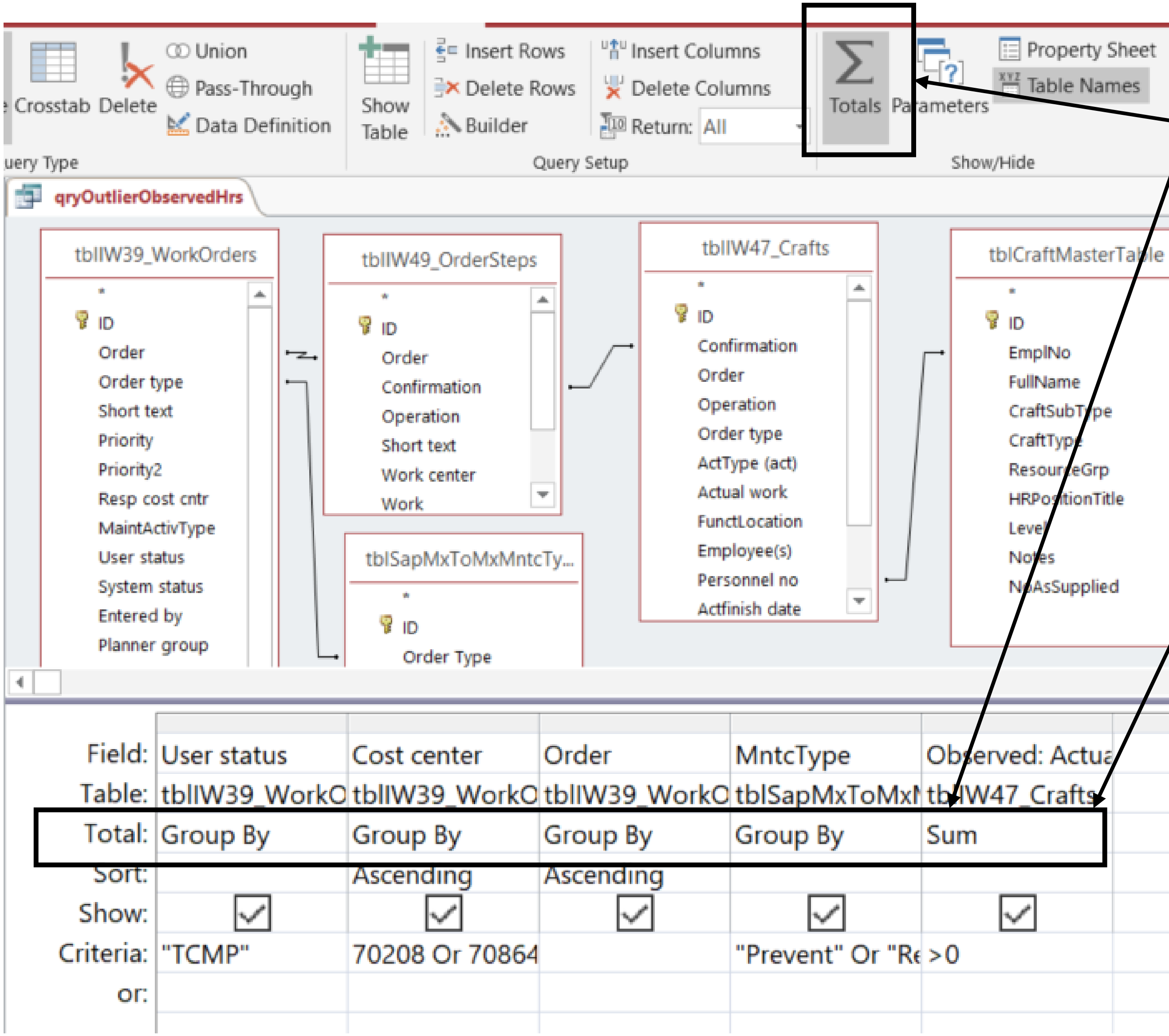
  - **Failure modes—enforce the maintenance work order process and reconstruct the past with insight deliverables.**

There are five types of bad data in a table—the good news is that there are methods to deal with each if it is poison to the final insight deliverable

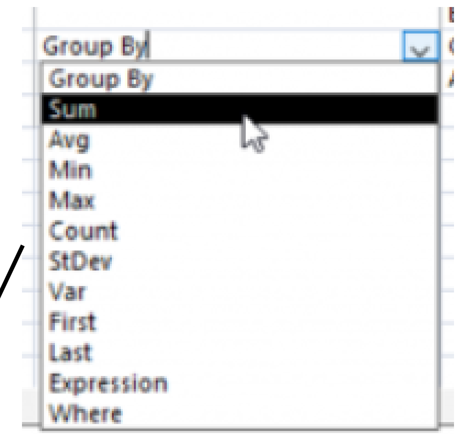
Type	Strategy
Duplicate cases	Seek cases with duplicate query.
Empty cells	▪ Form table with all permutation of empty cells—use "or" rows in query grid and use Pivots to form table of cases and counts. ▪ Evaluate the ramifications of loss of information to insight deliverable of remaining empty. ▪ Decide 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 modeled insight deliverables in the section titled, "The primary 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 joining tables, 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"

Access versus “R” for building super tables

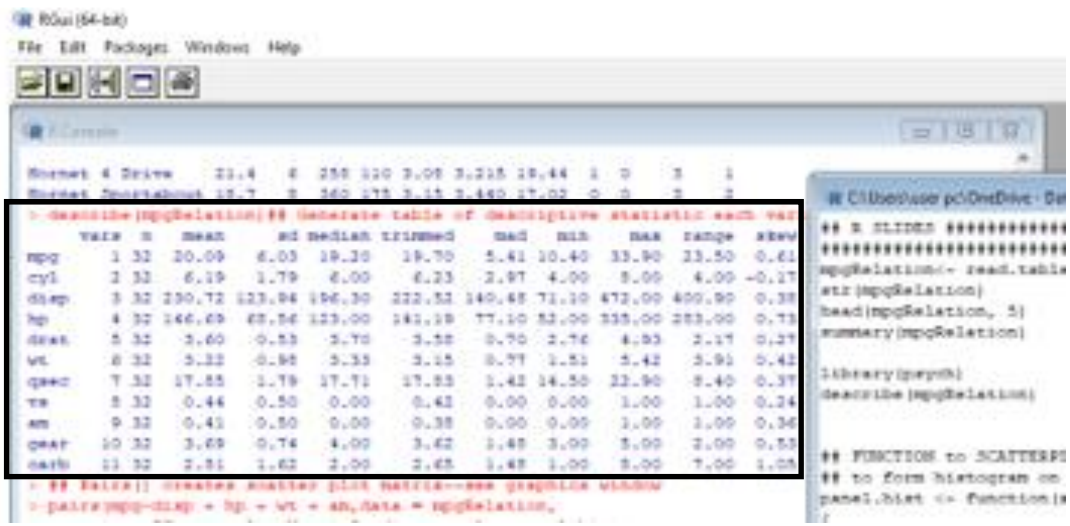
- Access does most of anything we could ever want and ask of any full-power software such as “R,” and typically all we want.

- The offset to the power gap is that Access is extremely easy to learn such that everyone in an organization can have the skill.

  Imagine everyone up and down your halls who normally work with Excel also becoming able to work with data—talk about a power jump!!

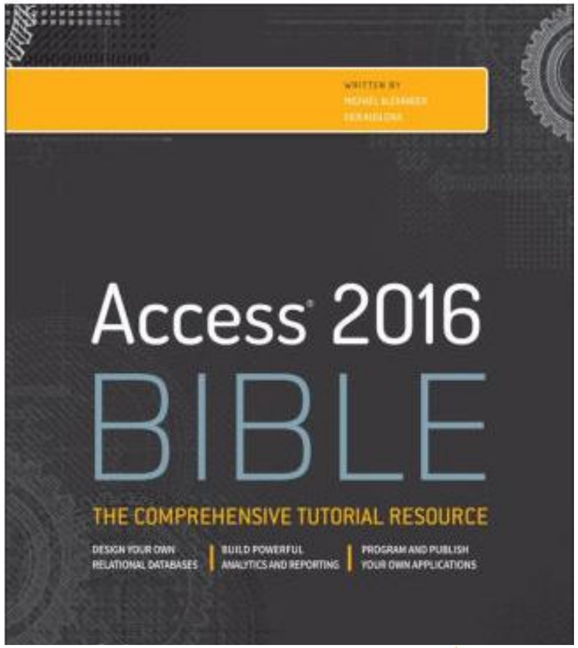
- When more is needed the Access-built table or raw data can be pulled into “R” and powered up—with a script such that anyone can execute all future cycles.

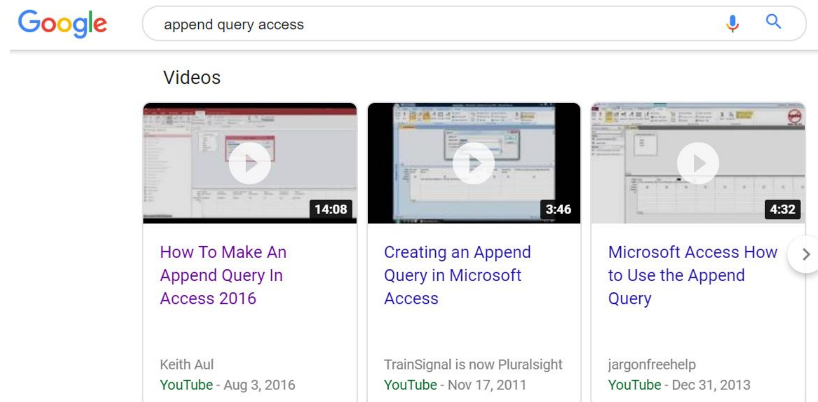
- Note: Specialized tables often generate in the process of running an analytic in “R”—they can be imported into Access and joined with conventional data tables.

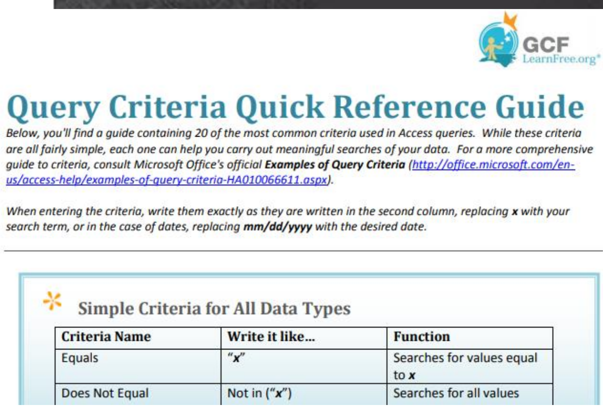


Another advantage of Access is that there is an immense support community

- **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

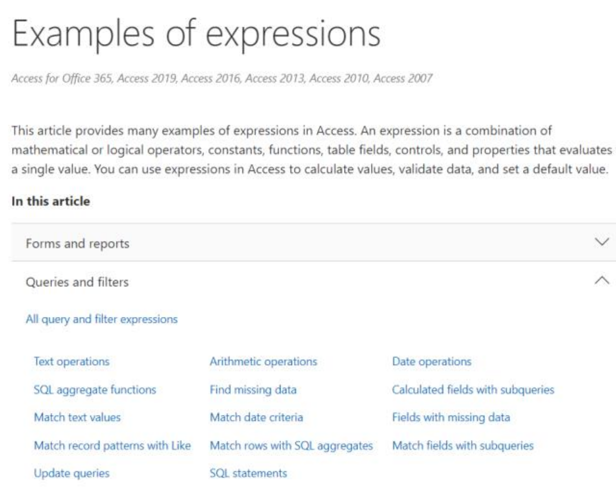
Query Criteria Quick Reference Guide

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	

Agenda:

- ❑ Purpose of the training session.
- ❑ The big picture of data-driven operations.
  - ➢ Definition and depiction of data-driven operational processes.
  - ➢ The 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.
  - ➢ **Eliminate fused data from operations.**
  - ➢ Layered charting in contrast to conventional charting.
  - ➢ The primary types of Insight deliverables—system reports, know-thy-data, recountive and modeled.
- ❑ Generalized implementation plan.
  - ➢ Big picture of implementation.
  - ➢ People—the central issue for development.
  - ➢ Stages, steps, actions and deliverables.
- ❑ Library of what-to and how-to papers, presentations and texts.

What is meant by **fused data** and why its elimination from operational processes has immense ramification for data-drivenness

- **Some tasks along operational processes are managed outside the standard systems—usually with Excel as the "IT object."**
- **The data processed in these case is often some of the most valuable and expensive to capture, manage and disseminate, as well as, most easily lost to the operation.**
- **To qualify as fully data-driven, these cases must be located and restructured to be data-driven compatible.**

Now that we have explained super tables, we can easily grasp the issue of fused data and why it is disastrous to operational excellence.

It may be obvious from the previous section, but let's state directly what format qualifies as a table rather than a spreadsheet

Each column is one piece of information, with the name it has been given.

Order	Confirmation	Operation	Short text	Work
6000689624	3862969	10	CUI 2009 ALKY PKG #5 - SEE LONG TEXT	CRAK-
6000689624	3879730	20	AB Clean to remove stockpiled sand from	CRAK-
6000689624	3899247	30	WASTE MANAGEMENT SERVICES	CRAK-
6000689624	3920227	40	ALLIED WASTE RR20090394	CRAK-
6000689624	3924998	50	7970- nonhazardous trash disposal	CRAK-
6000689624	3989116	60	ICU to test or monitor	HSE--C
6000689624	4038490	70	Allied Waste RR20090740	CRAK-
6000689624	4072292	80	COASTAL WELDING - CUI PROJECT SPO	CRAK-
6000689624	4086802	90	HPP - RECYCLE BLAST MEDIA	CRAK-
6000689624	4104129	100	ICU-TCLP (ABRASIVE DISPOSAL)	HSE--C
6000689624	4104717	110	COASTAL WELDING - CUI PROJECT - HPG GAS	CRAK-
6000689624	4112364	120	HAGEMEYER - MONITOR RENTAL - ALKY UNIT	CRAK-
6000689624	4171719	130	PAY INVOICE 611708	CRAK-
6000689624	4191435	140	ALLIED WASTE - RR20100071 - DIB TWR	CRAK-
6000707049	3933686	10	DCU PU8818-SHOP ORDER IMPELLER	MSPU-
6000707049	3939602	20	DCU PU8818-JSA & WORK SCOPE	COKE-
6000707049	3939603	30	DCU PU8818-JSA & LO/TO MOTOR	COKE-
6000707049	3939604	40	DCU PU8818-PULL PUMP	COKE-
6000707049	3939605	50	DCU PU8818-INSPECT COOLER & BASE	COKE-
6000707049	3939606	60	DCU PU8818-OPERATION TO ENERGIZE MOTOR	COKE-

What is *not* along the side also defines a data table:

- **There are no row titles.**
- **There are no subtotals and totals.**
- **There are no empty rows**

It can't be shown, but data tables are continuous; they are not subdivided in separate tables with respect to any one variable—e.g., day and week.

Spreadsheets violate every rule of a data table, rendering its data unable to serve anything other then what it is to report—gold reduced to garbage

	A	H	I	J	L	M	N	O	Q	R	S	T	V	W	X	Y	AA
1	Date: 04/16/2012 Monday	16/2012 Monday			Date: 04/16/2012 Monday				Date: 04/16/2012 Monday				Date: 04/16/2012 Monday				04/16/2012
2		de Lubes			Utilities				Logistics				Maintenance/Bldg/Roads/Grounds				Total
3	Work Order Type	MX07	MX09	MX12	MX01,2,3 & 4	MX07	MX09	MX12	MX01,2,3 & 4	MX07	MX09	MX12	MX01,2,3 & 4	MX07	MX09	MX12	
4																	
5	**Routine Maintenance**																
6	**Total Motiva Maintenance**	0	0	0	23	9	0	0	21	0	0	0	6	0	0	0	109
7																	
8	**Routine Contractors**																
9	Echo	0	0	0	5	0	0	0	10	0	4	0	2	0	0	0	30
10	Brand Carpenters	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14
11	Brand Insulators	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	4
12	Brand Painters	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	USIS	0	0	0	1	0	0	0	2	0	0	0	0	0	0	0	14
14	Evergreen	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	3
15	Newtron	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	6
16	Corrpro	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
17	**Subtotal**	0	0	0	12	0	3	0	17	0	4	0	2	0	0	0	72
18																	
19	**Other**																
20	USIS (Abclean) **(Environmental Cleanup)**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	Dashiel	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	2
22	GP (8Train,#981089,dh)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
23	Echo(8Train,#981089,dh)	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	6
24	Echo(steam leak,#972719,dh)	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	3
25	Ems Line work on row suspected leak	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	2
26	ICT(HTU3,#988499,fc)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
27	Secho(VPS4 fin fan,#986680,fc)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
28	ECHo(2VPS EX 82,#2464100,fc)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
29	Deep South(2VPS EX 82,#2464100,fc)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
30	Total Crane (jg)	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	4
31	Sulzer (jg)	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	8
32	PAL (jg)	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	2
33	GP (jg)	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	2
34		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
35	Echo(PAT Levee Gates, #990633,tm)	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	3
36	USIS(Alky DIB Ex 40-43,#1002385,tm)	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
37	Brand Scaffold (Alky DIB Ex 40-43,#1002385,tm)	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	3
38	Ohmstead (Alky DIB Ex 40-43,#1002385,tm)	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	8
39		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40	**Subtotal**	0	0	0	0	24	0	0	4	2	15	0	0	0	0	0	55
41																	
42	**Tank Contractors**																

Sheet1 Sheet2 Sheet3

Ready

This tedious daily report could have been created in a few seconds by Pivot with the six columns of data.

The content of the spreadsheet could have been easily captured as a six-column continuous data table, making it available to any imagined Excel Pivot and super data table.

Date	Unit	MaintType	Count	LaborGrp	Contractor

To become fully data-driven, the operation must find all fused cases and restructure them as two layers

- **Layer 1:**

  The data to the report—henceforth redesign the process to collect them in Excel as a continuous data table.

- **Layer 2:**

  Upon the data of layer 1, generate the same report with a pivot software (e.g., Excel Pivot, Tableau) or other software.

Agenda:

- ❑ Purpose of the training session.
- ❑ The big picture of data-driven operations.
  - ➢ Definition and depiction of data-driven operational processes.
  - ➢ The 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.
  - ➢ Eliminate fused data from operations
  - ➢ **Layered charting in contrast to conventional charting.**
  - ➢ The primary types of Insight deliverables—system reports, know-thy-data, recountive and modeled.
- ❑ Generalized implementation plan.
  - ➢ Big picture of implementation.
  - ➢ People—the central issue for development.
  - ➢ Stages, steps, actions and deliverables.
- ❑ Library of what-to and how-to papers, presentations and texts.

You need to know that the ggplot2 package of "R" allows us to breakthrough to charting data and information in layered perspectives

Charting as two generations

Conventional

Chart types are named staples since the 1600s

Layered

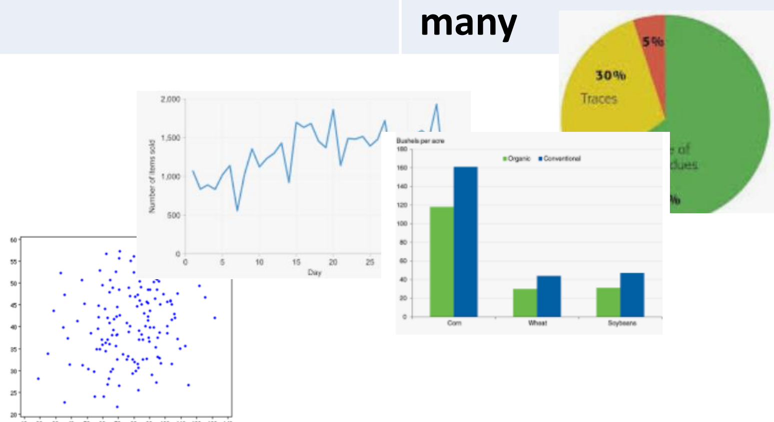
Slip the bounds of centuries by placing perspectives as layers in single charts

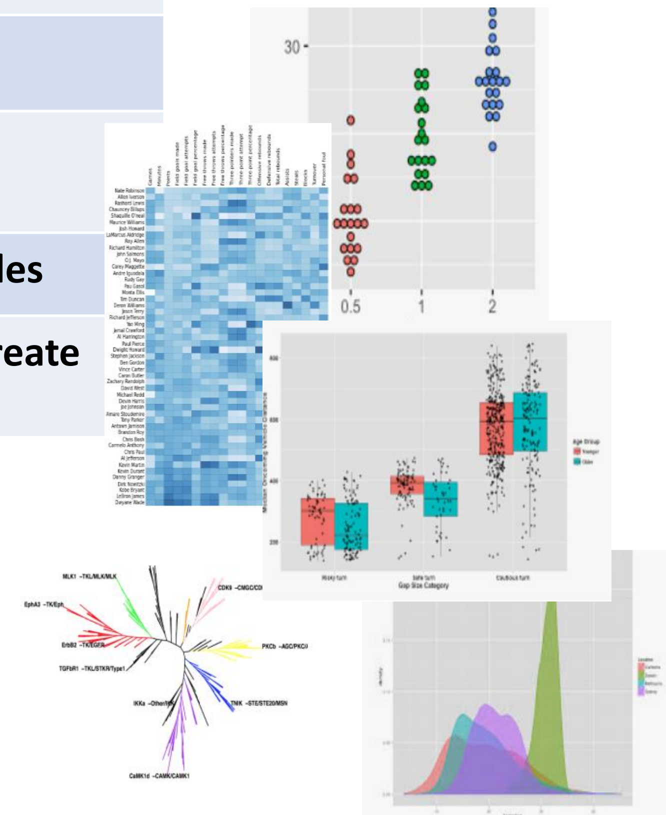
Note

Layered charting allows creativeness so extreme that some people have taken to creating artwork.

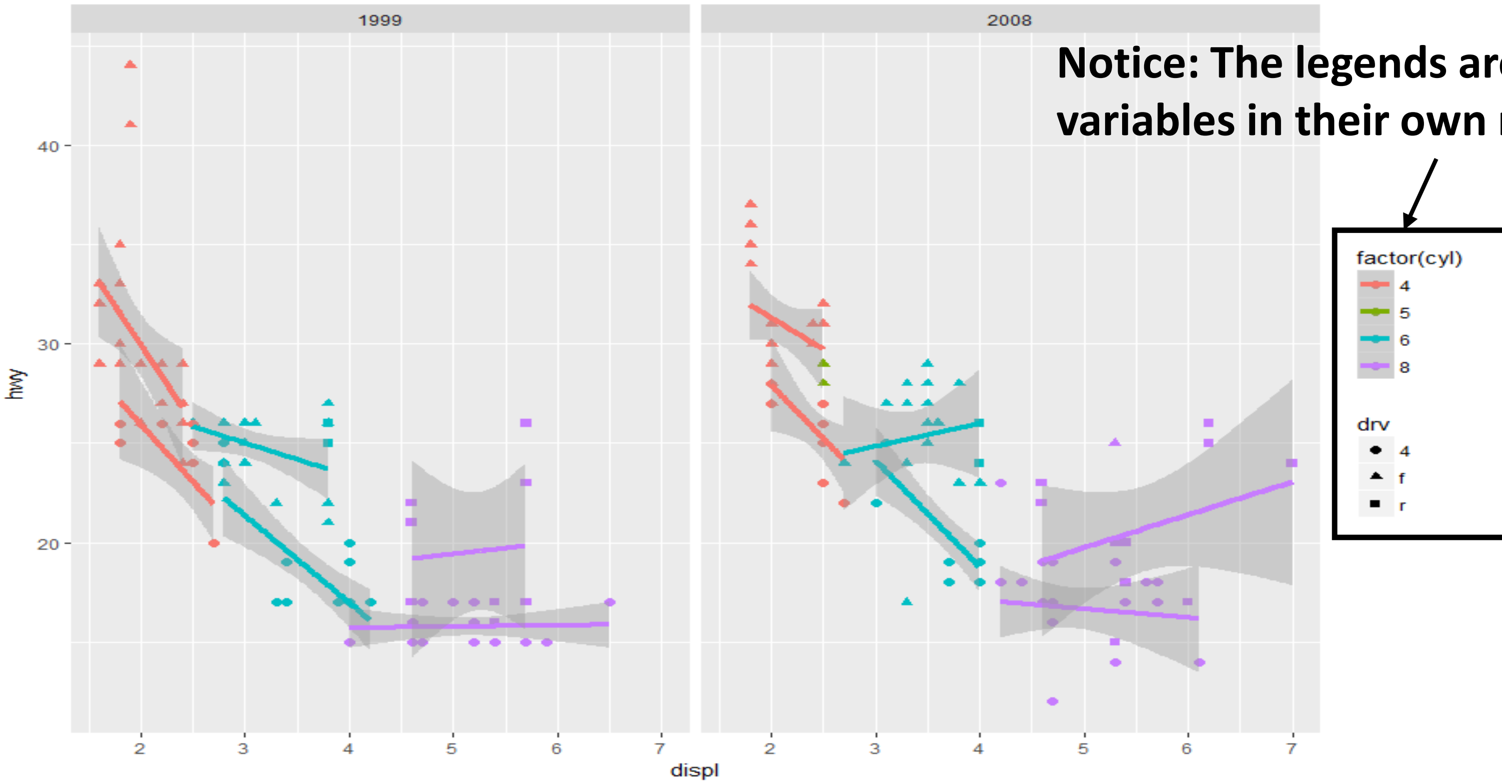
Conventional charts have been the staple since they were invented as early as the 1600s—now layered charting allows perspectives we have never had before

Characteristic	Conventional	Layered
Types	limited	Unlimited
Naming	Per chart type	Per purpose and content
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





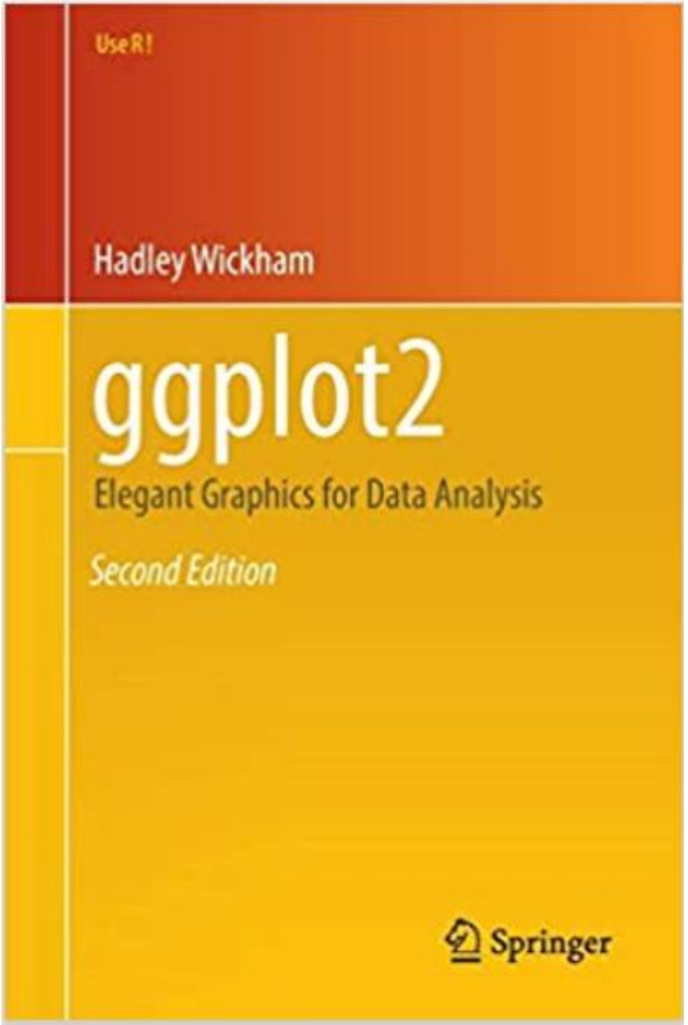
“**Layered**” charts place perspectives as insights layered in a single chart named by their purpose and content—as long has they have common axis



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**
- **Not shown: How distinctiveness is preserved if there were many points as shown in two of the layered charts of the previous slide.**

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



Go 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.

You will need to know the basics of R—see the section for recommendation for self-directed learning.

You will have fun!!

Agenda:

- ❑ Purpose of the training session.
- ❑ The big picture of data-driven operations.
  - ➢ Definition and depiction of data-driven operational processes.
  - ➢ The 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.
  - ➢ Eliminate fused data from operations
  - ➢ Layered charting in contrast to conventional charting.
  - ➢ **The primary types of Insight deliverables—system reports, know-thy-data, recountive and modeled.**
- ❑ Generalized implementation plan.
  - ➢ Big picture of implementation.
  - ➢ People—the central issue for development.
  - ➢ Stages, steps, actions and deliverables.
- ❑ 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 sectors, paths and steps along an operational process

Category	Description
System reports	Taken from operating systems as standard reports.
Know-thy-data	Data is presented in descriptive, graphic and statistical perspectives as a window into how the process is working compliance-wise in contrast to how it should be working.
Recountive	Insight direct from data—without processing through analytics—to ask and answer questions of who, what, when, where and how much, and metrics formed across them.
Modeled	Insight gained upon data flowing through ML/AI models that ask and answer questions of relationship, difference, time series, duration and apparency.

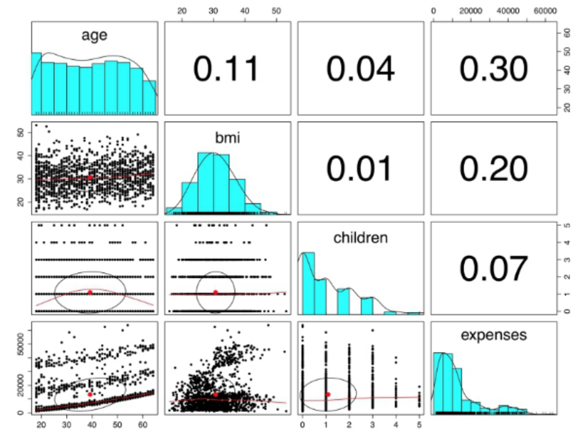
Next

Know-thy-data, recountive and modeled 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	Some is possible with Access. However, the best of KTD insight is generated through various R functions specifically developed for that purpose.
Recountive	R will come into play via the "R" functions for layered charting.
Modeled	All depend upon "R" functions.

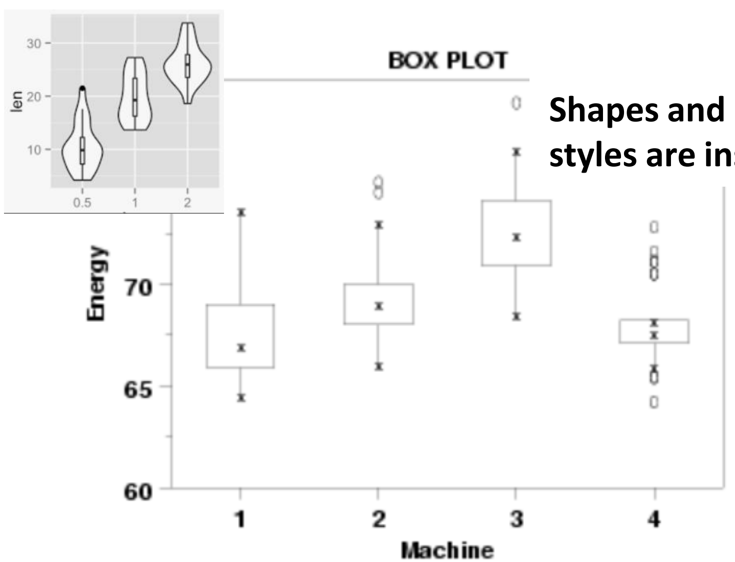
The many tools to "**Know-thy-data**" are combined as insight deliverables to probe, monitor and control that processes are functioning as intended



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

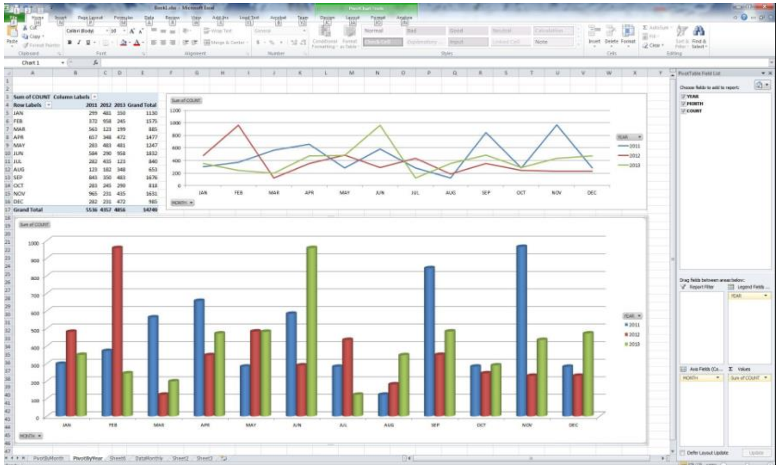


Shapes and plotting styles are insightful.

- **KTD 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

- **Insight is largely limited to what happened with respect to who, what, where, when, how much and metrics.**
- **As compared to modeled insight, recountive insight can be packaged and disseminated through Pivot software directly from the super table.**
- **Recountive insight can be powered-up greatly with layered charting.**



Note

Recountive insight is currently the most often observed degree—if any—of accomplishment in operational processes.

"Indicative" insight deliverables are either analytic models or constructed on models to know that a process is becoming and remaining effective and efficient

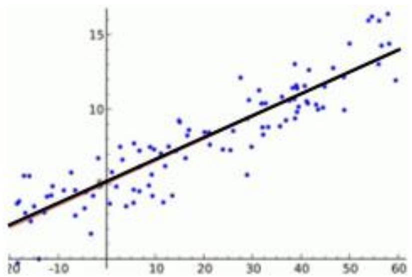
- In contrast to recountive insight direct from the super table, modeled cases pass the same data through ML/AI-based models.
- The models ask and answer five types of questions of the subject operational process—relationship, difference, time series, duration and apparency.
- Each type of question entails different types of models.
- In sharp contrast, recountive insight deliverables cannot deal with the five types of questions.
- There are two categories of modeled insight deliverables.
  - The model is chosen, designed and worked to be an insight deliverable in its own right.
  - The variables and issues that matter are revealed by model; thence insight is built upon the associated variables.

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 and proportional hazard regression and Weibull regression.
Apparency	Are there hidden predictor variables to the performance of assets and processes?	Decision tree, regression tree, model tree and K-mean.

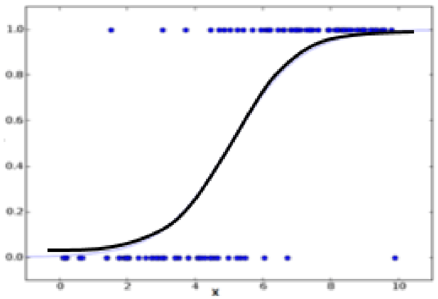
(1) A set of explanatory papers to the models can be downloaded at https://analytics4strategy.com/new-age-five-questions

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



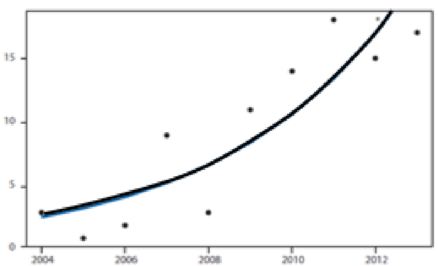
Linear regression:

- 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:

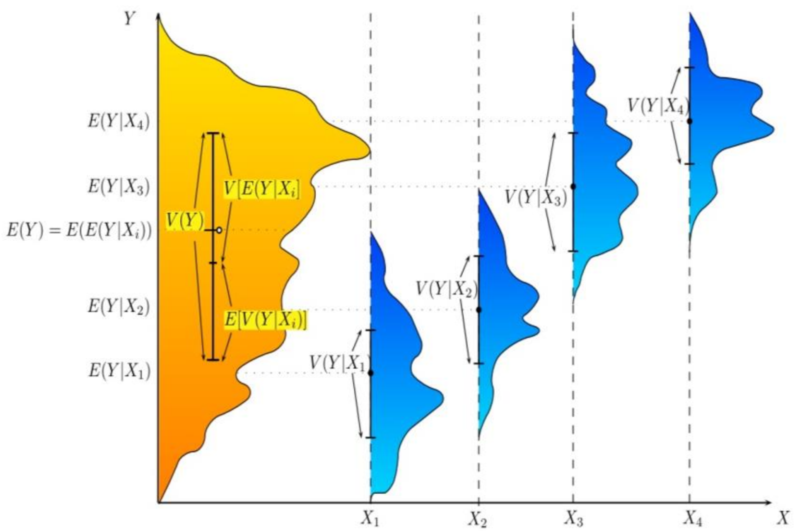
- Classifies between possible categorical outcomes.
- Binomial: Yes/No and multinomial: yes, no, maybe.
- Classification based on the probability of an outcome.
- Example question: Which asset and process variables most foretell the probability of finding crews engaged in non-value work?



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?

<u>Difference questions</u>: How do combinations of asset and process variables comparatively effect a performance of interest?



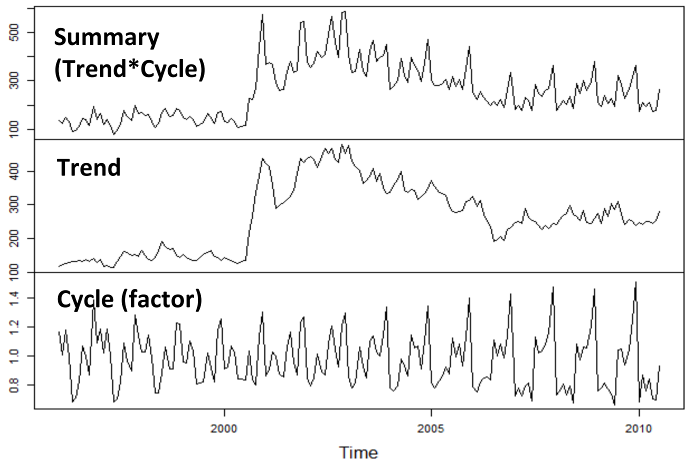
Explanation:

- 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 Questions:

- Are costs, hours, productivity, various occurrences and KPIs practicably different from previous periods?
- In what situations are there the largest gap (non-value) between planned and actual craft hours for work orders?

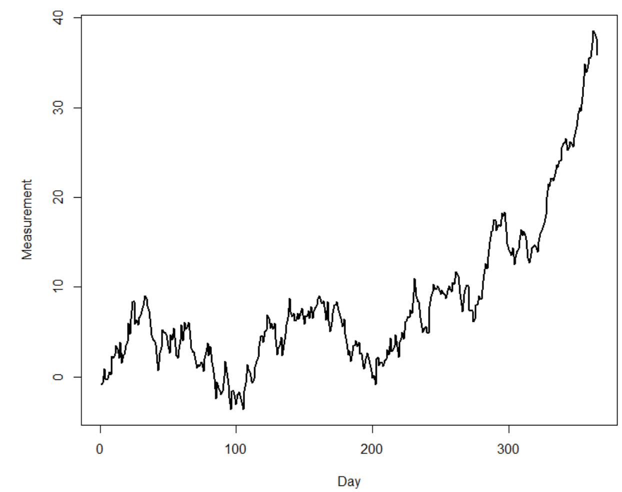
Time series questions: What are the components that underlie the summary-level-only history that operating systems are limited to providing?



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?**
- **Most of us would assume this trend is deterministic—making us very happy.**
- **Times series modeling would find the trend is a random walk—we are in fools' paradise.**

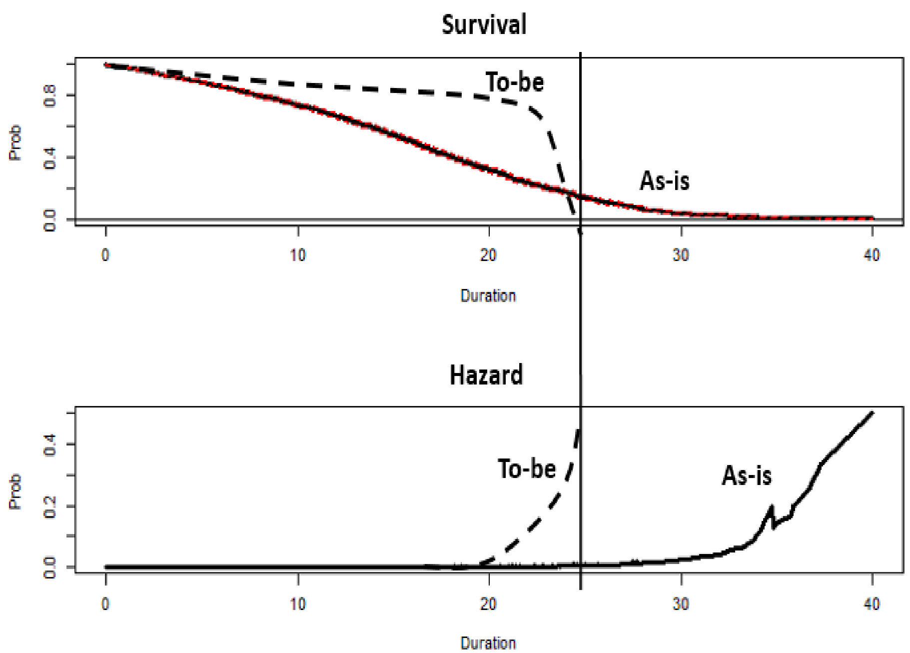


Example questions:

- **Is the trend for productivity changing with time?**
- **Are there noteworthy trends in compliance to process policies—toward or away—along the process?**

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

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.
- The shown as-is case is FIFO with a considerable degree of gaming the process and some orders stagnate in the condition.

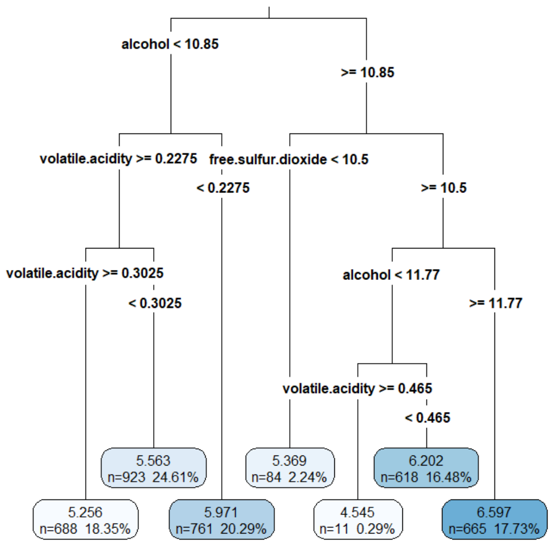
- Shows the probability of an ending event as a function of how long the condition has existed.
- The shown as-is case is long then need be and, late on, exits increase at decreasing rate
- Mathematics of "hazard" make it possible to determine which process variables matter most to the story.

Example questions:

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

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?

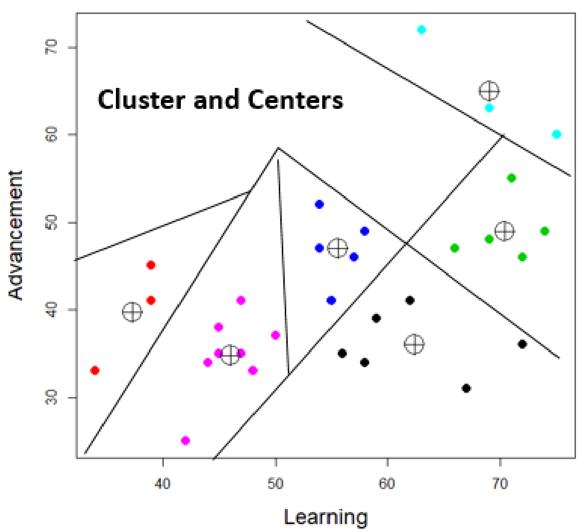


Decision tree:

- Hidden Rules as variables associated with numeric and categorical outcomes.
- An outcome variable is given to the model--supervised.
- The learned rules 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.
- Shown case: Six hidden variables are teased out of two variables, but no outcome variable was given to the model.

Example question:

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

The five modeled insights are depicted in the simplest form to their model type—one predictive variable and one outcome variable without interaction

- Reality is that within each type of model, there are subtypes as well as many combinations and configurations of variables.
- The complexities 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 papers:

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

Comment on AI:

For all of the media attention to AI; logistic regression and rule trees are most frequently the models of choice—because they classify outcomes for decision-making.

Agenda:

- ❑ Purpose of the training session.
- ❑ The big picture of data-driven operations.
  - ➢ Definition and depiction of data-driven operational processes.
  - ➢ The 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.
  - ➢ Eliminate fused data from operations
  - ➢ Layered charting in contrast to conventional charting.
  - ➢ The primary types of Insight deliverables—system reports, know-thy-data, recountive and modeled.
- ❑ Generalized implementation plan.
  - ➢ **Big picture of implementation.**
  - ➢ People—the central issue for development.
  - ➢ Stages, steps, actions and deliverables.
- ❑ Library of what-to and how-to papers, presentations and texts.

Your organization now needs an implementation plan to put the knowledge, skills and methodologies in play

- **A path of stages, steps, tasks and deliverables along which to steer each chosen operation from its current state to being fully data-driven.**

- **Generalized—not intended to fit every case—so that you can see what needs to happen and modify the plan to fit the nature of your organization.**

- **Structured to cause the transfer of knowledge, skills and methodologies to the role holders in implementation and subsequent functioning.**

Generalized progression of stages, steps and accomplished organizational abilities

Stages and Steps | **Timeline to General Abilities**

Stage 1: Set direction and prepare.

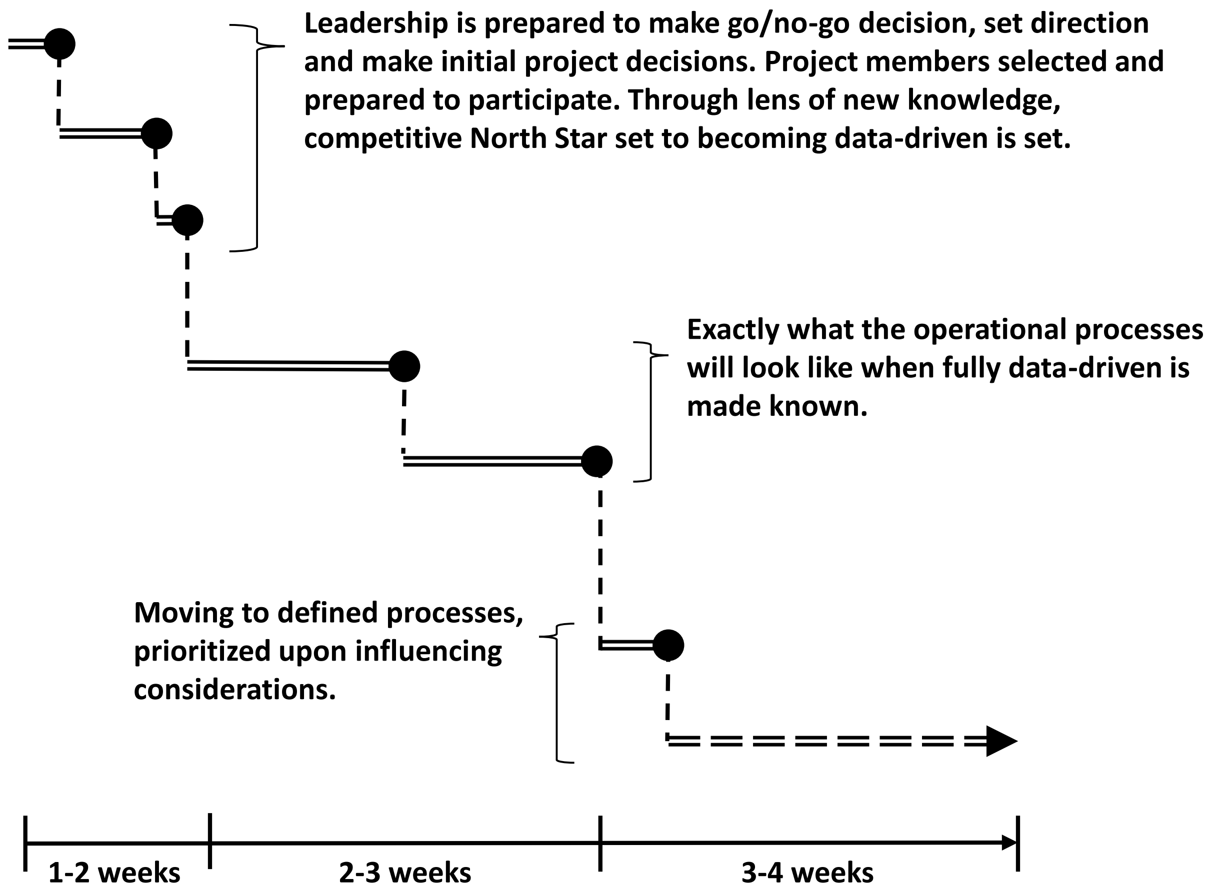
1. **Form clear understanding with leadership and initial decisions.**
2. **Prepare nucleus players to participate.**
3. **Set competitive North Star for data-drivenness.**

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.**



Agenda:

- ❑ Purpose of the training session.
- ❑ The big picture of data-driven operations.
  - ➢ Definition and depiction of data-driven operational processes.
  - ➢ The 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.
  - ➢ Eliminate fused data from operations
  - ➢ Layered charting in contrast to conventional charting.
  - ➢ The primary types of Insight deliverables—system reports, know-thy-data, recountive and modeled.
- ❑ Generalized implementation plan.
  - ➢ Big picture of implementation.
  - ➢ **People—the central issue for development.**
  - ➢ Stages, steps, actions and deliverables.
- ❑ 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 set a strategy to either incubate or recruit the necessary human talents for data-driven operations

- **Polar opposite choices:**
  - **Upskill the process role holders to be new-age workers.**
  - **Recruit specialist professionals in data, charting and analytics to conduct data-driven tasks, leaving the process players to continue on largely with only pre-data-driven skills.**
- **Drawbacks to the specialist alternative:**
  - **Is it fair to the process role holders to be essentially left in the past?**
  - **Specialists lack the depth and breadth of domain expertise that process role holders have gained over years.**
- **The charted implementation plan is designed to upskill role process holders as the implementation project unfolds.**
- **The plan is constructed upon three categories of participants.**
  - **Drivenness guide.**
  - **Nucleus players.**
  - **Management advocates.**

Drivenness guide: Currently the most difficult resource to find and probably does not yet lurk in most organizations

- **Strategy:**

  Engage a drivenness guide, once found, in a manner such that many of those they work with will evolve to being able to take on the guide role—one step behind.

- **Primary qualifications:**
  - Considerable history in operational excellence work.
  - Has learned to incorporate the knowledge and skills of data, charting and analytics into their long-gained, advisory-grade acumen.

- **Role:**
  - Off-load the skills of data-drivenness to all who are involved in the team and roles in bringing the subject operation to data-drivenness.
  - Participate, by collaboration and mentoring, as a team member to identify, design and build the insight deliverables.
  - Collaboration is the platform from which to train the engaged nucleus players and role holders in the principles, practices and tools of data, charting and analytics.

Nucleus players: First individuals to be engaged in the design, build, dissemination and use of all insight deliverables

- **Role as strategy:**
  - **Master the hands-on skills of data-drivenness.**
  - **Disperse the skills to everyone along the subject processes for which working with data and insight deliverables will become part and parcel to their roles.**
  - **One or more of the nucleus players will take on the role of project manager to seed the skills for managing data-drivenness projects across the organization—local and global.**
- **Qualifications: Process operatives, managers, experts and engineers in the subject processes and operations.**

Management advocates: Always an issue to assure that the vision of becoming data-driven will become the reality

- **Data-drivenness will cut across the bounded subprocesses.**
- **Managers, downward from the pinnacle of the involved processes.**
  - **Advocate data-drivenness by word and demonstration.**
  - **Encourage individuals to strive to become new-age employees.**
- **Managers have skin in the game as one of the greatest beneficiaries of the insight deliverables.**

As the stages unfold, the knowledge and skills of data-drivenness are transferred to managers and nucleus players by the 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:

- ❑ Purpose of the training session.
- ❑ The big picture of data-driven operations.
  - ➢ Definition and depiction of data-driven operational processes.
  - ➢ The 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.
  - ➢ Eliminate fused data from operations
  - ➢ Layered charting in contrast to conventional charting.
  - ➢ The primary types of Insight deliverables—system reports, know-thy-data, recountive and modeled.
- ❑ Generalized implementation plan.
  - ➢ Big picture of implementation.
  - ➢ People—the central issue for development.
  - ➢ **Stages, steps, actions and deliverables.**
- ❑ Library of what-to and how-to papers, presentations and texts.

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. . .

Continued: Stage One to set direction and prepare

<table>
<tr><th colspan="2">Step 2: Prepare the chosen nucleus players to participate in the design, build and function of data-drivenness.</th></tr>
<tr><td colspan="2">The chosen nucleus players will receive the same training given to leadership. They will be additionally trained hands-on in the skills of building and cleansing super tables, exploring the tables with methods of descriptive statistics and building layered charts.</td></tr>
<tr><th>Activities</th><th>Deliverables</th></tr>
<tr><td>1. Train the nucleus players to be conversant in the principles, practices, software and skills of data-drivenness.
2. Train hands-on the nucleus players to extract, join and cleanse the data with MS Access and R, and subject the tabled data to descriptive statistical analysis with R and layered charting with ggplot2.
3. Act upon discoveries during the step that have obvious, immediate ramifications for operational effectiveness and efficiency.</td><td>• Nucleus of personnel with the working skills to extract, join and work with data, as well as, mentor others in the skills.
• Library of “cookbook” materials such as slides, training materials, articles and texts.
• Installed, or underway, developments from the discoveries of the step.
• Informal and probable: Self-directed actions by individuals to upgrade the tasks of their position with newly gained skills for insight deliverables.</td></tr>
</table>

Continued. . .

Continued: Stage One to set direction and prepare

Step 3: Set the competitive North Star for data-drivenness.	
It is necessary to establish the top-down competitive framework by which the progression to full data-drivenness can set its direction, make its choices and set its priorities along the way. The step is placed third in order to conduct a competitive assessment through the lens of what is possible through data-drivenness and conversely view data-drivenness with respect to begetting competitiveness.	
Activities	**Deliverables**
1. Frame how the enterprise competes and wins and how that is reflected in return on investment through the financial statements. 2. Generalize how the subject organizational processes variously affect competitiveness. 3. Establish qualitative and quantitative measures downward from returns and through the financial statements by which the ramifications of data-drivenness will be judged and insight deliverables will be valued. 4. Review the initial process geographic decisions for necessary revision.	• Statement of competitiveness for the firm and how the acts to plan, organize, conduct and control the subject operations play into competitiveness. • Structure of categorical and numeric metrics by which the ramifications of data-drivenness and insight deliverables will be evaluated. • Sharpened original direction and geographic decisions—if revealed as necessary by the step.

End of Stage One. . .

Stage Two: Conduct basic design

The stage will assess how the selected operational processes work and, in turn, how insight deliverables would practicably enhance their effectiveness and efficiency. Ultimately, a basic design will be formulated for each insight deliverable.

The nucleus players will be trained in the types and capabilities of the range of machine-learning/artificial-intelligence-based (ML/AIB) analytics upon which modeled insight deliverables depend. Before it is possible to conduct basic design, it is necessary for the team to learn which and how the ML/AIB analytics variously enable five types of questions for insight to be asked and answered of the processes—relationship, difference, time series, duration and apparency. In contrast to the training of the step, hands-on training in the ML/AIB analytics will take place in the third stage as the recognized insight deliverable are built and made functional.

Continued. . .

Continued: Stage Two to conduct basic design

Step 1: Detail the operational processes for becoming data-driven.	
The step is to fully understand the subject processes, identify external threats to their performance and kick off any discovered obviously mandatory remedial actions along the critical path to reaching data-drivenness.	
Activities	**Deliverables**
1. Chart the processes as they are intended to work, understand their operational systems, find all cases of process steps conducted outside of the systems with Excel or other software, catalog the data captured and generated along the processes, and identify all existing insight deliverables. 2. Cluster the process sectors, paths and steps with common ramifications to the dimensions of enterprise competiveness (e.g., production, inventory, and variable and fixed costs). 3. Search out local and global threats from other processes to the effectiveness and efficiency of the process clusters. 4. Identify and spin off for immediate action any discovered cases of process and data failures for which remediation is obviously mandatory to becoming data-driven.	• Process charts overlaid with existing operating systems, tasks conducted with Excel and other software, data along the processes, existing insight deliverables and would-be fatal flaws to data-drivenness. • Documented cases of where, when and how surrounding local and global operations present threats to realizing the enterprise-level competitive ramifications of the subject processes. • Proactive initiatives underway to remedy fatal processes, behaviors and data.

Continued. . .

Continued: Stage Two to conduct basic design

Step 2: Map proposed insight deliverables along the charted operational processes and form basic designs.	
Whereas the previous step lays out the processes and external threats to their effectiveness and efficiency, this step will lay out the operational processes with respect to what they will be when fully data-driven; including data-driven strategies to deal with external threats.	
Activities	**Deliverables**
1. Train the project team in the ML/AIB analytics such that they can finally assess and recognize the ramifications of any of all types of insight deliverables to a process and then form the basic-design for each. 2. Identify two cases for insight. a. Along the charted processes where an improved ability to plan, organize, conduct and control will practicably influence enterprise competitiveness. b. Outside the charted processes where non-compliance to related processes are a threat to the subject operational processes. 3. Determine the set of one or more insight deliverables at each location of influence along the processes and to deal with external threats to the influence. 4. Specify the format and content for each proposed insight deliverable. 5. Identify the source data, software, knowledge and skills to generate and utilize each insight deliverable.	• Nucleus of personnel who have extended their knowledge and skills to being fully conversant in the ML/AIB analytics as insight deliverables as well as mentor others in the knowledge and skills—arriving at a fully understanding all available insight deliverables. • Insight deliverables mapped to the charted processes and threats to their influential effectiveness and efficiency. • Table of basic designs for each insight deliverable: title, users, preparers, content, source data, implementing software and skills. • Informal and probable: Self-directed actions by individuals to upgrade the tasks of their position in response to the findings and new knowledge and skills of the stage.

End of Stage Two. . .

Stage Three: Plan and conduct detailed design, build and startup

The stage starts by determining the order that individual or groups of insight deliverables will be designed and put permanently into play. The stage then prepares and progresses through micro-projects to conduct the detailed designs and implement them. The training of nucleus players continues to completion by virtue of doing the hands-on building work of the stage. They, in turn, will train process role holders as there are start-up activities.

<table>
<tr><th colspan="2">Step 1: Establish the sequence by which the proposed insight deliverables are to be implemented.</th></tr>
<tr><td colspan="2">Not all insight deliverables have equal ramifications and homogenous issues. The step deals with arriving at an optimal, practical sequence for bringing the insight deliverables on line.</td></tr>
<tr><td>Activities</td><td>Deliverables</td></tr>
<tr><td>1. Develop a set of criteria for classifying the proposed insight deliverables for their relative ramifications and issues.
2. Rate and rank the insight deliverables for implementation against the criteria.
3. Make decisions for sequential implementation along single or multiple paths.</td><td>• All propositions classified for ramifications and issues.
• Decisions for order of implementation—subject to reconsideration as implementation unfolds.</td></tr>
</table>

Continued. . .

Continued: Stage Three to plan and conduct detailed design, build and startup

Step 2: Form detailed design and chart cyclical process for each insight deliverable and implement.	
As micro-projects, the step will conduct the detailed design of each insight deliverable according to the sequence plan. The cyclical process for each or set of insight deliverables will also be charted and detailed. The nucleus players will learn to build the ML/AIB analytics for modeled insight deliverables in addition to reinforcing the earlier hands-on trained skills in building cleansed super tables, layered charting, and know-thy-data and recountive insight deliverables. Process role holders, not on the project team, will receive the skills for their roles from the nucleus players.	
Activities	**Deliverables**
1. Develop and approve micro-projects for detailed design and startup—including tasks for upskilling non-nucleus-player process operatives. 2. Form and approve the full functional detailed designs for the subject insight deliverables and their cyclical processes. 3. Roll out and bring each data-driven proposition to full function.	• New-age operations: Built insight deliverables embedded in the fully functioning processes to prepare, conduct initial analytics, disseminate and consume insight. • New-age people: Rapidly and organically growing number of operational role holders who are trained in the thought, design and functioning of insight deliverables.

End of Stage Three and project plan. . .

Agenda:

- ❑ Purpose of the training session.
- ❑ The big picture of data-driven operations.
  - ➢ Definition and depiction of data-driven operational processes.
  - ➢ The 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.
  - ➢ Eliminate fused data from operations
  - ➢ Layered charting in contrast to conventional charting.
  - ➢ The primary types of Insight deliverables—system reports, know-thy-data, recountive and modeled.
- ❑ Generalized implementation plan.
  - ➢ Big picture of implementation.
  - ➢ People—the central issue for development.
  - ➢ Stages, steps, actions and deliverables.
- ❑ **Library of what-to and how-to papers, presentations and texts.**

The data-driven operation will make a library of "what-to" and "how-to" materials readily available to its process role holders and advisors

- **All that has been explained to this point has considerable instructive materials in the public domain.**
- **The library (next slides) has three layers:**
  - **Areas of knowledge and skills.**
  - **Papers and presentations of overview and training-session materials.**
  - **Texts and equivalents to methodologies.**
    - **Full explanation of the subject method—principles and practices.**
    - **Method being built in the triad of software—Excel Pivot, Access and R.**
    - **Validating analytic methods for legitimacy and offering an alternative if there is an invalidating issue.**
    - **Interpreting the output and how to present the output.**
- **Will grow as training sessions occur and actual designs become go-bys.**
- **Role holders will become proficient in utilizing the library as implementations progress and subsequently into functioning operations.**

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

Knowledge and skills		Papers and presentations	Texts or equivalents
Data-drivenness	Framework	Training session: First Step to Becoming a Data-Driven Operation	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	▪ Building the Super Tables Behind Data-Driven Operations ▪ Purge the Fused Spreadsheets That Undermine Data-Drivenness	Access 2016 Bible, Alexander and Kusleika, 2016, Chapters 8 – 13.
Data preparation	Cleansing	None available	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
ML and AI	Methodology	None available	