

JMP Topics and Capabilities for
Getting Beyond Spreadsheet Analytics with
Visual Data Discovery

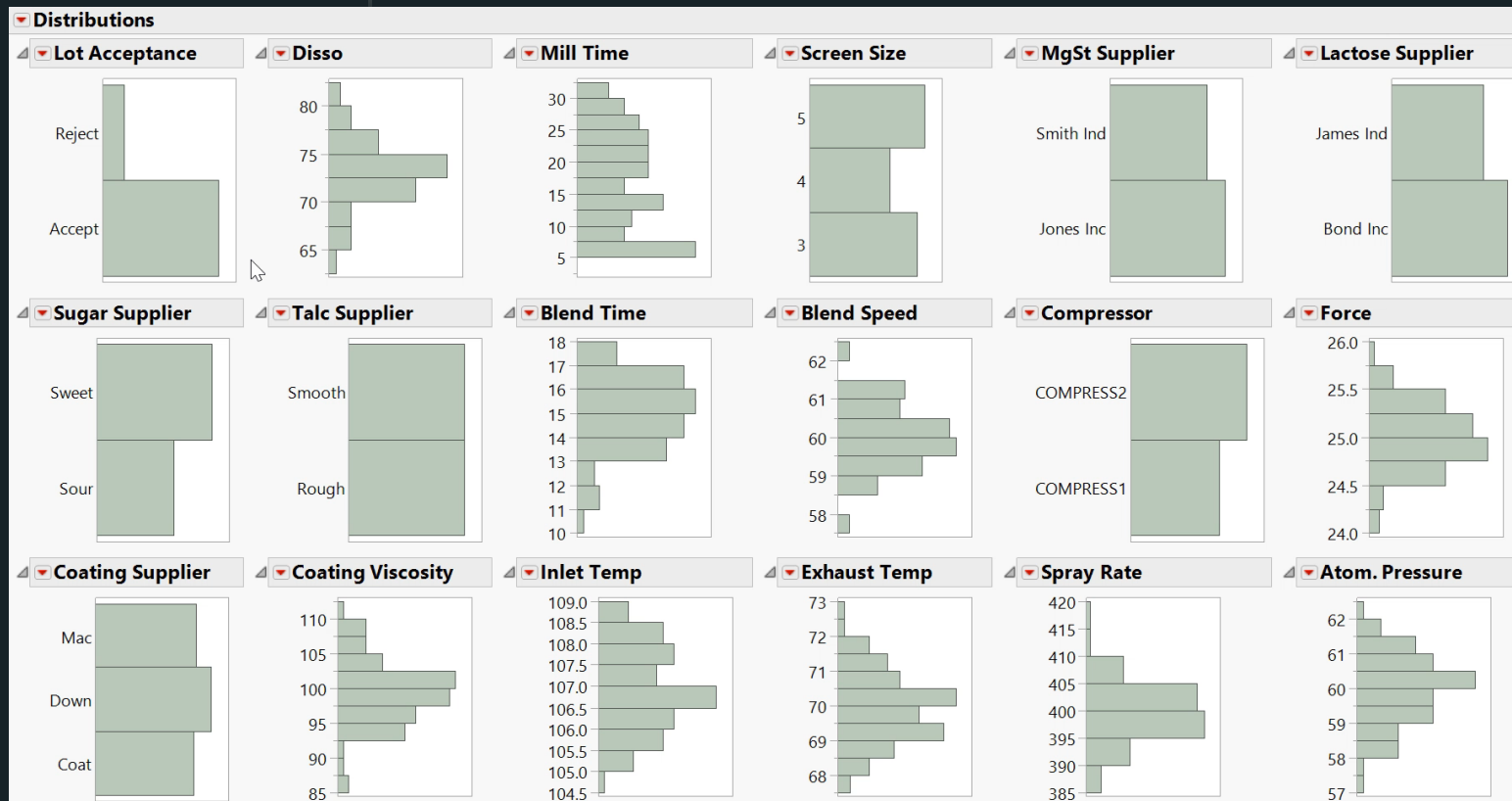
Move from Data to Decision – Faster!

Tom Donnelly, PhD, CAP
Principal Systems Engineer
Defense & Aerospace Team
JMP Statistical Discovery LLC

Desk 919-531-1380
Mobile 302-489-9291
www.jmp.com/fedgov
tom.donnelly@jmp.com

Exploratory Data Analysis

DYNAMIC LINKING AMONG ALL GRAPHS AND DATA TABLE ALLOWS SELECTION OF SUSPICIOUS DATA IN ONE GRAPH AND SEE POSSIBLE CORRELATIONS IN ALL!

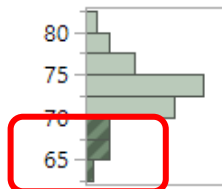


Exploratory Data Analysis

DYNAMIC LINKING AMONG ALL GRAPHS AND DATA TABLE ALLOWS SELECTION OF SUSPICIOUS DATA IN ONE GRAPH AND SEE POSSIBLE CORRELATIONS IN ALL!

Distributions

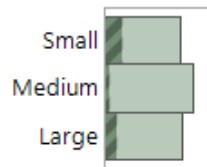
Disso



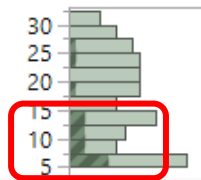
Lot Acceptance



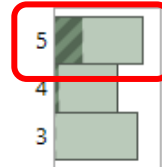
API Particle Size



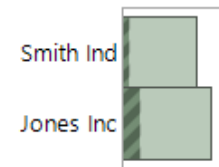
Mill Time



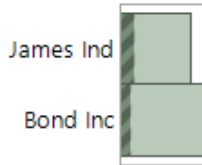
Screen Size



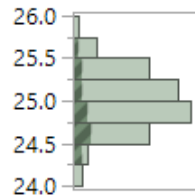
MgSt Supplier



Lactose Supplier

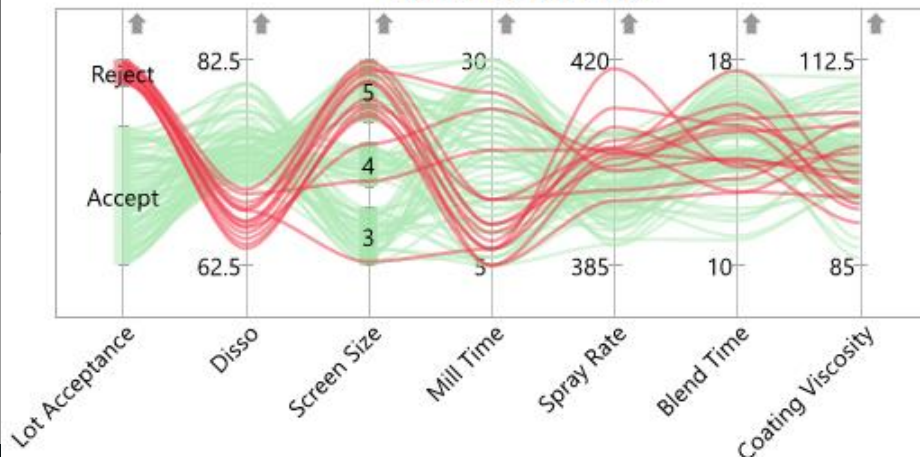


Force

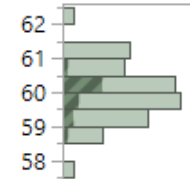


Graph Builder

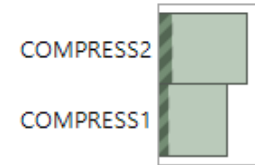
Mill Time & 6 more



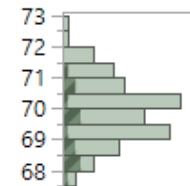
Blend Speed



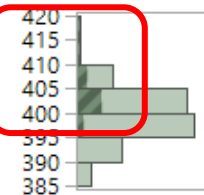
Compressor



Exhaust Temp

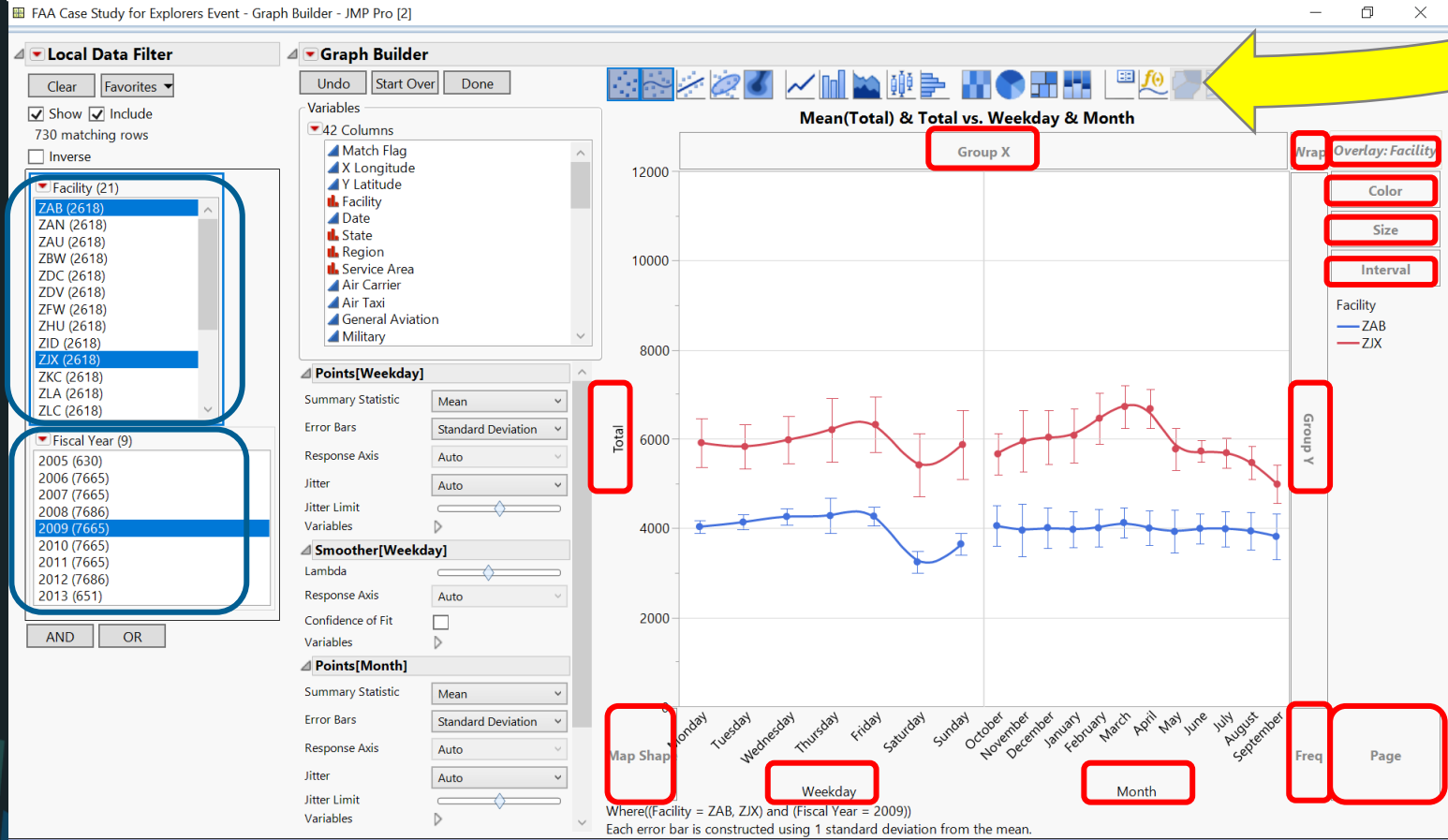


Spray Rate



Highly Interactive Graph Builder

DRAG & DROP VARIABLES TO **DROP ZONES** TO EXPLORE RELATIONSHIPS **FILTER DATA**
TO DRILL DOWN TO FIND INSIGHTS, MORE THAN A DOZEN **GRAPH ELEMENTS**



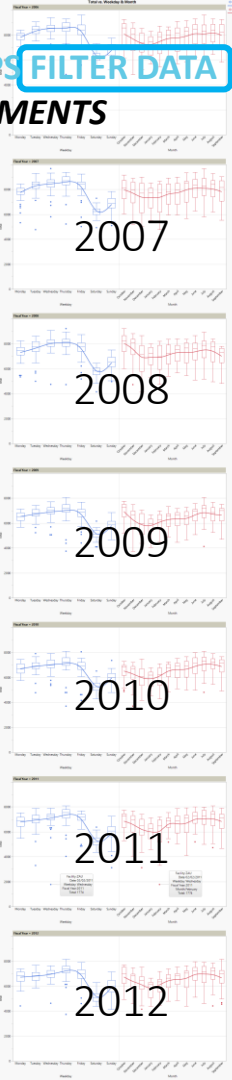
Highly Interactive Graph Builder

DRAG & DROP VARIABLES TO **DROP ZONES** TO EXPLORE RELATIONSHIPS **FILTER DATA**
TO DRILL DOWN TO FIND INSIGHTS, MORE THAN A DOZEN **GRAPH ELEMENTS**

Using **Outlier Box Plot** Graph Element
Filter **Facility** on ZAU (Chicago/Aurora)
Drop **Fiscal Year** in Page Zone to get

Select most extreme outlier

Search on "2/2/2011 & Chicago"
Top hit is "Groundhog Day Blizzard"

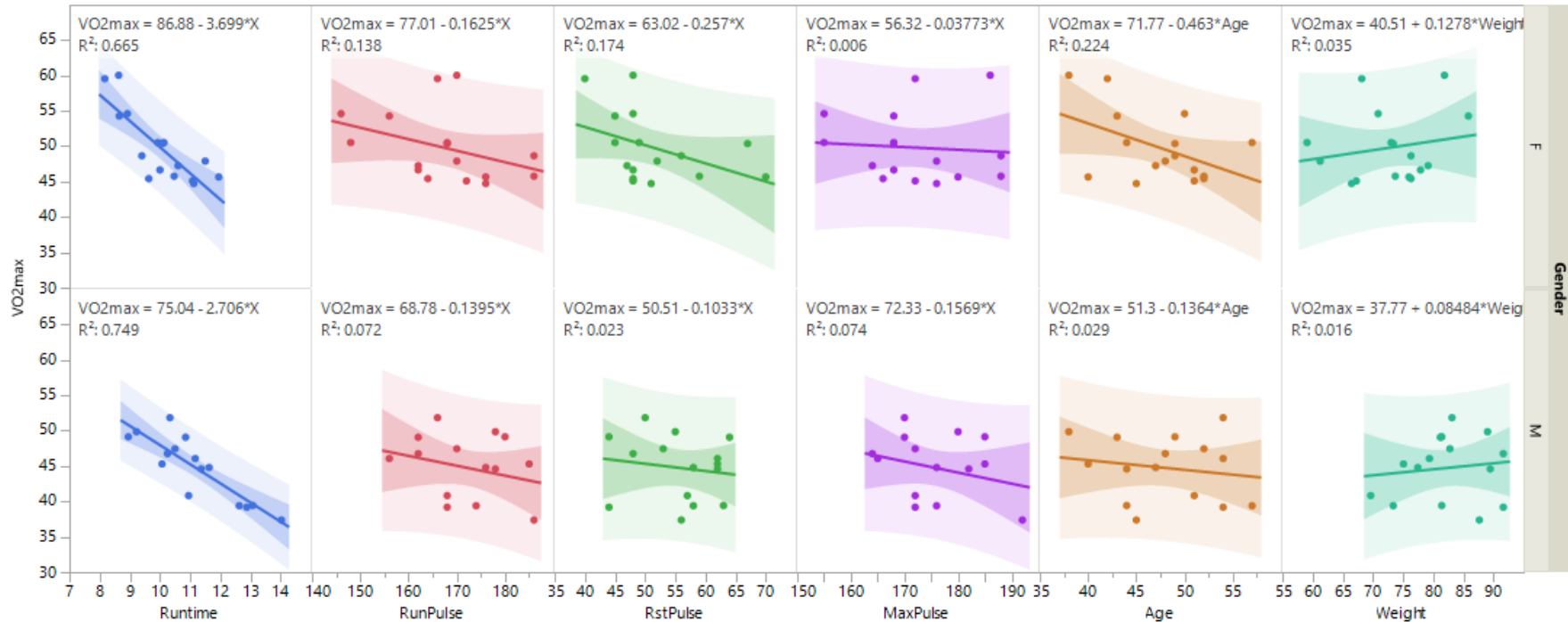


Multivariate Visual Modeling

DRAG & DROP INDIVIDUAL FITS OF EACH CONTINUOUS FACTOR BY GENDER

Graph Builder

VO2max vs. Runtime & 5 more



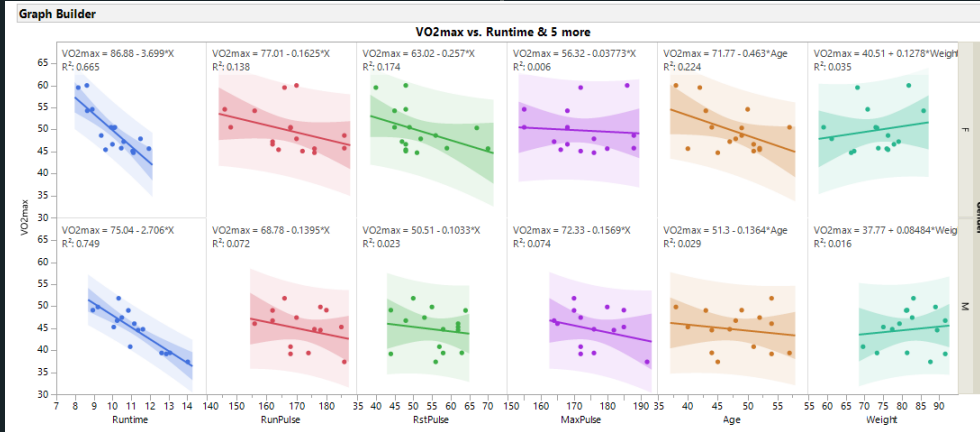
Multivariate Visual Modeling

DRAG & DROP INDIVIDUAL FITS OF EACH CONTINUOUS FACTOR BY GENDER

The screenshot displays the JMP Graph Builder window. On the left, a column list contains variables: Name, V02max, Runtime, RunPulse, RstPulse, MaxPulse, Age, Weight, Gender, and Weight in LBS. Below this is the 'Points' panel with settings for Summary Statistic (None), Error Interval (Auto), Interval Style (Error Bar), Jitter (Auto), Jitter Limit (slider), and Variables (dropdown). The main plot area is empty, with a central text prompt: "Drag variables into drop zones". The plot area is framed by a 'Y' axis on the left and an 'X' axis at the bottom. The title bar above the plot area contains 'Group X' and 'Group Y' labels. On the right side of the plot area, there are control panels for 'Wrap', 'Overlay', 'Color', 'Size', 'Interval', and 'Page'. The 'Interval' panel shows a 'Count' option. The 'Page' panel shows a 'Page' option. The bottom of the window has a status bar with navigation icons.

Multivariate Visual Modeling

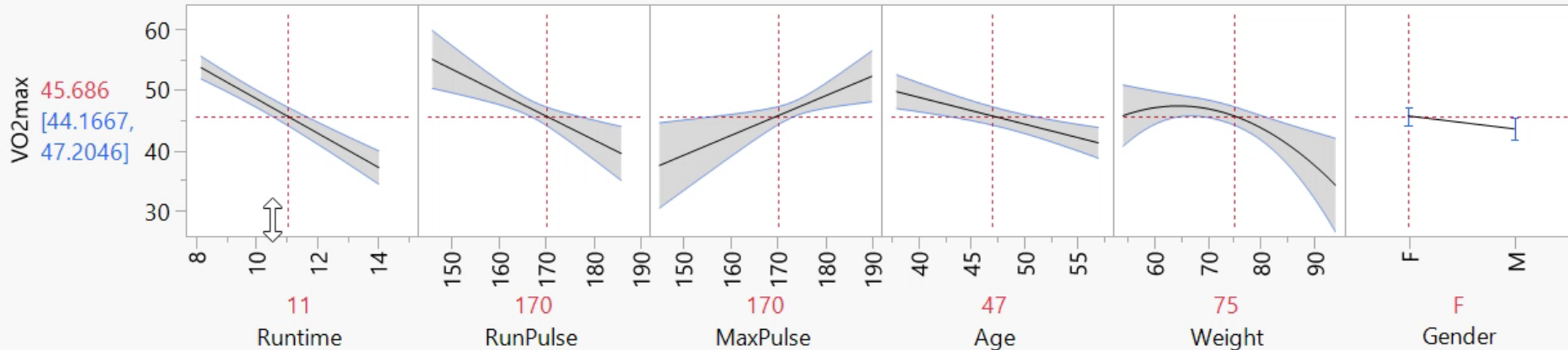
VISUALLY IDENTIFY MOST LIKELY TERMS IN FINAL MODEL



Effect Summary

Source	LogWorth	PValue
Runtime	8.405	0.00000
RunPulse	2.823	0.00150
Weight*Gender	2.741	0.00181
Age	2.198	0.00634
Age*Gender	2.179	0.00663
MaxPulse	1.963	0.01088
Weight*Weight	1.392	0.04056
Weight	0.865	0.13652 ^
Gender	0.161	0.69039 ^

Prediction Profiler

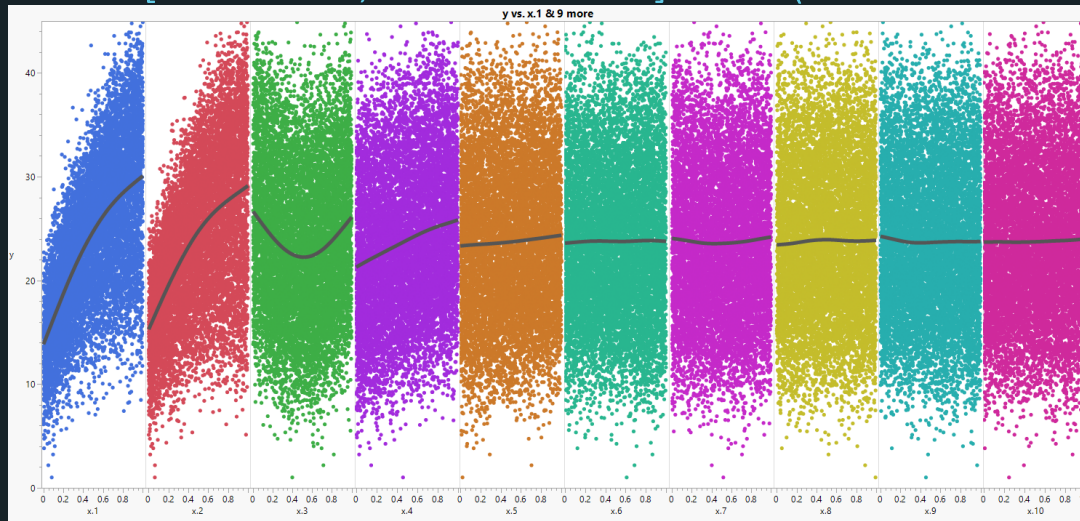


Visual Data Discovery with 200 Continuous Factors!

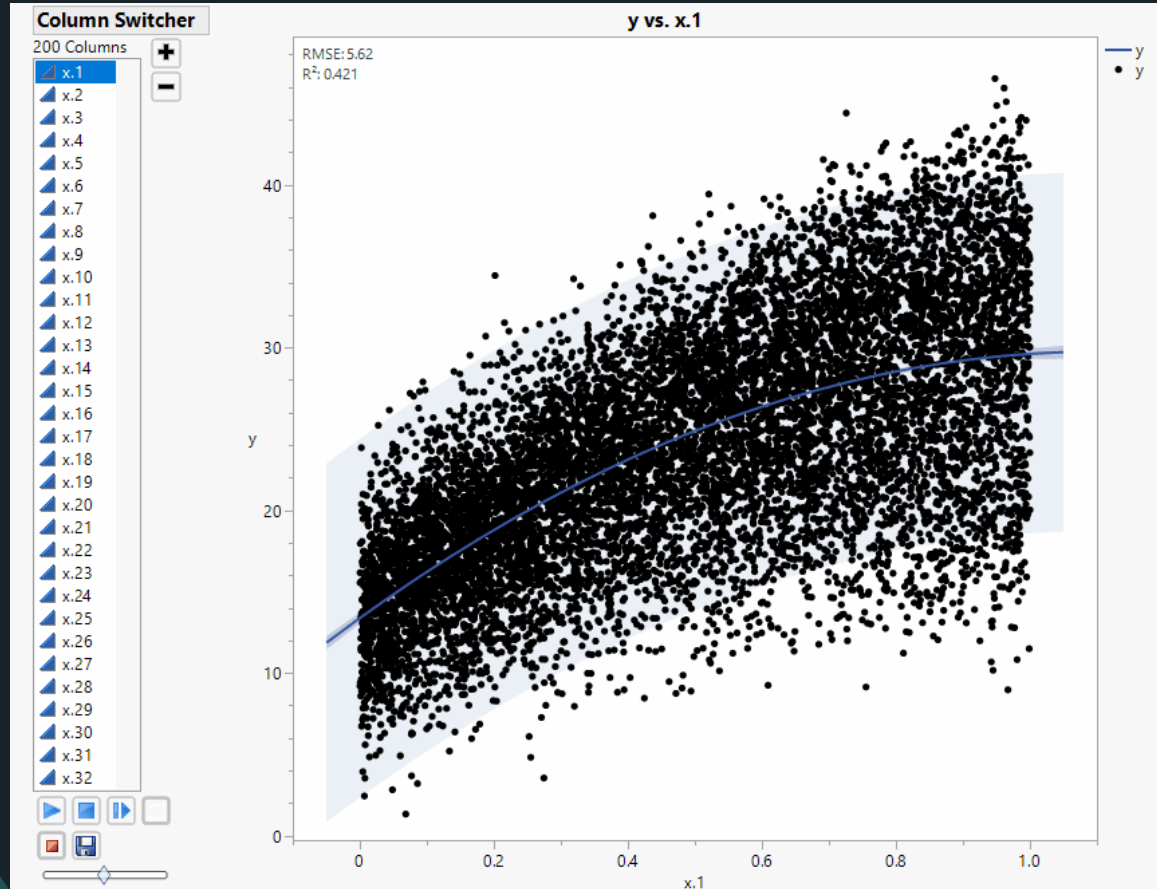
This function has only **5 active factors** ($x_1, x_2, x_3, x_4, & x_5$).

$$y = 7 + 20*x_1 + 17*x_2 + (-20)*x_3 + 10*x_4 + 6*x_5 + (-13)*x_1*x_1 + (-13)*x_2*x_2 + 20*x_3*x_3 + 19*x_1*x_2 + (-10)*x_4*x_5 + \text{Random Normal [Mean = 0, StdDev = 2.8]}$$

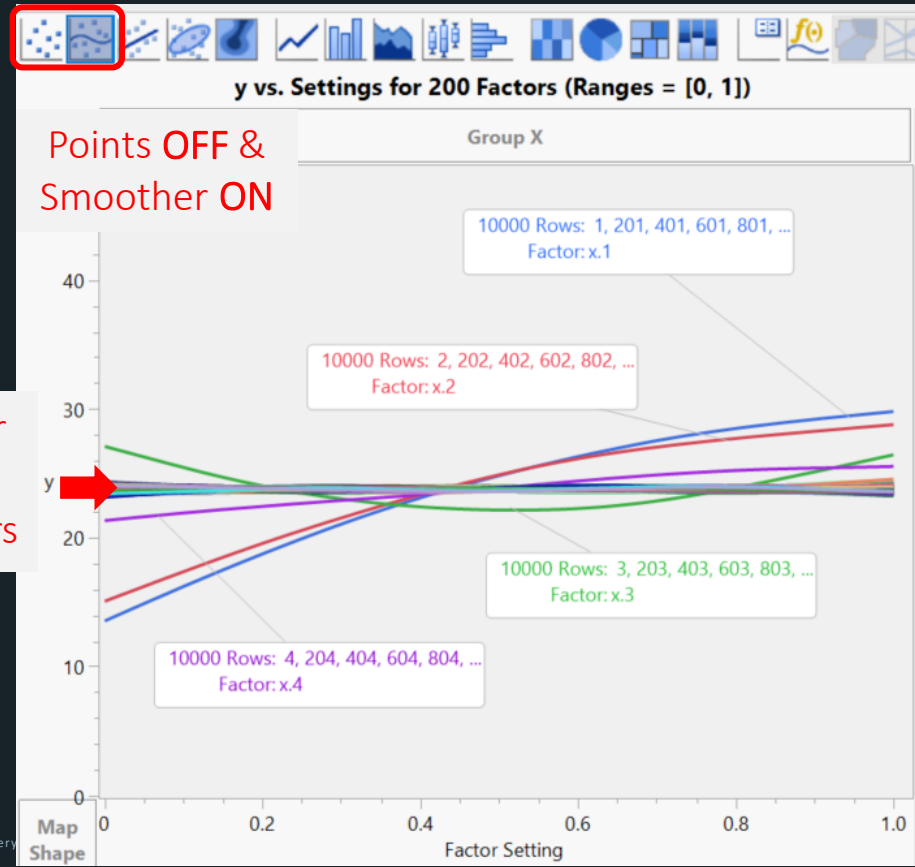
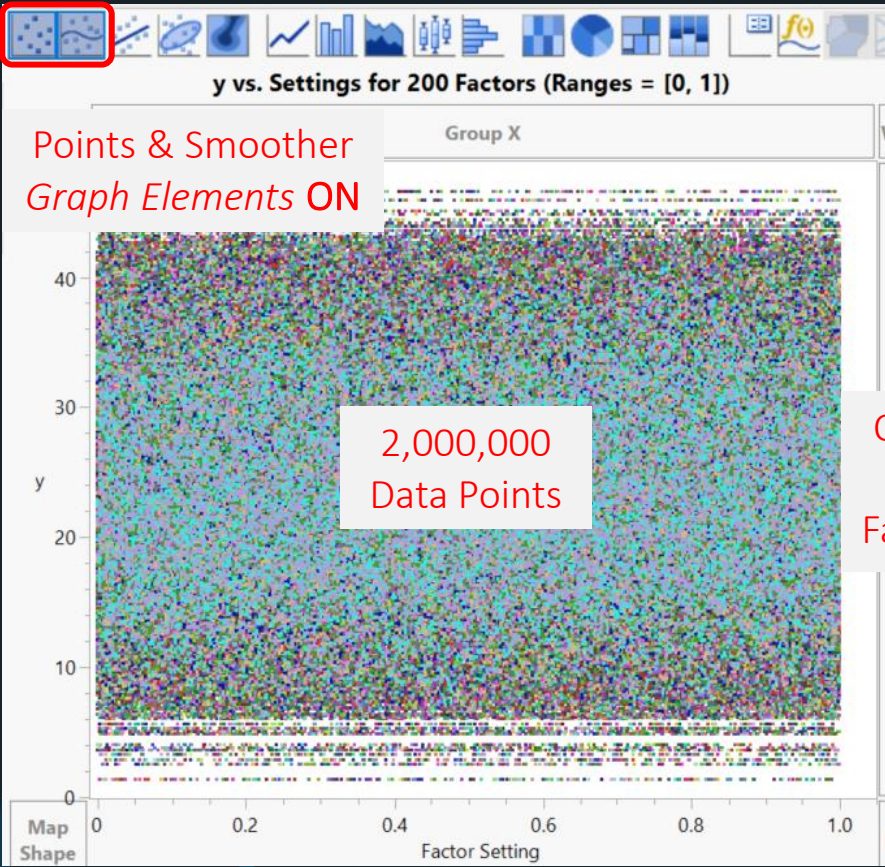
(intercept)
(5 main effects - *slopes*)
(3 squared effects - *curvature*)
(2 interaction effects - *twists*)
(Random error)



Animate Column Switcher to View Each Plot of y vs x_i



Stack ALL Factors (200 X 10K = 2M Rows), Then **Overlay** by Factor, Eliminate the Data Points to View the 200 Smoother, to See the 4 Dominant Factors **x1**, **x2**, **x3**, & **x4**



Further Visual Data Discovery with 200 Continuous Factors! This function has only 5 active factors (x1, x2, x3, x4, & x5).

$$y = 7$$

(intercept)

$$+ 20*x1 + 17*x2 + (-20)*x3 + 10*x4 + 6*x5$$

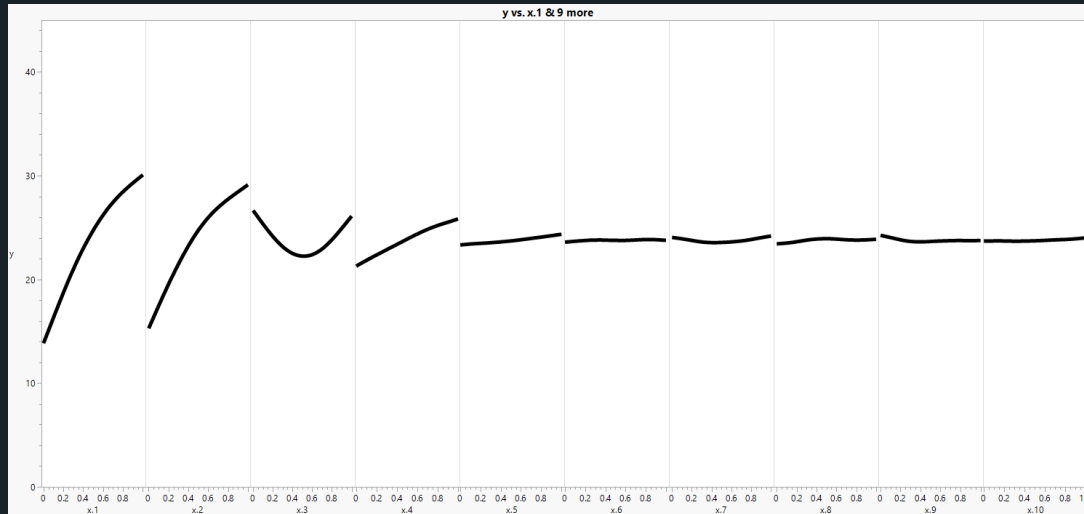
(5 main effects - slopes)

$$+ (-13)*x1*x1 + (-13)*x2*x2 + 20*x3*x3$$

(3 squared effects - curvature)

$$+ 19*x1*x2 + (-10)*x4*x5$$

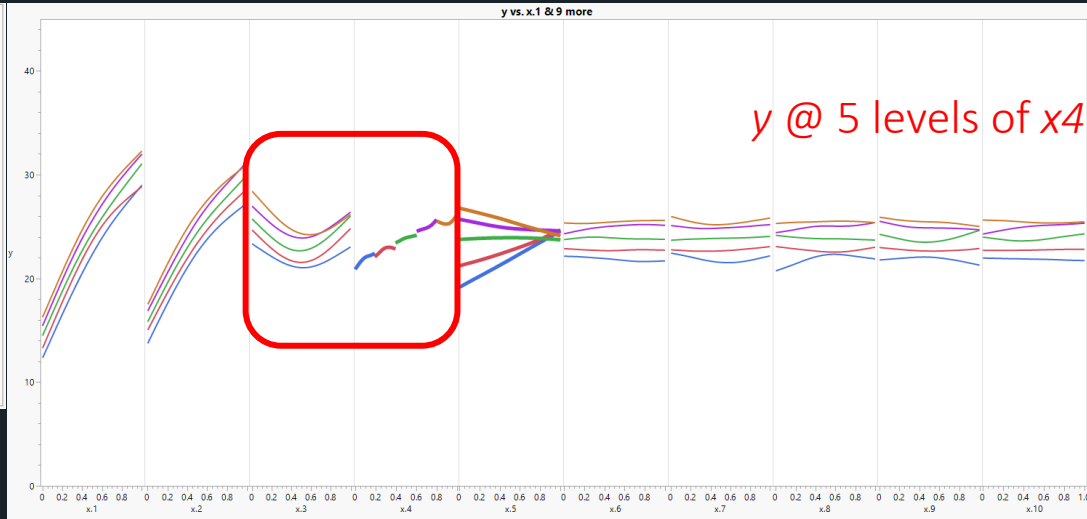
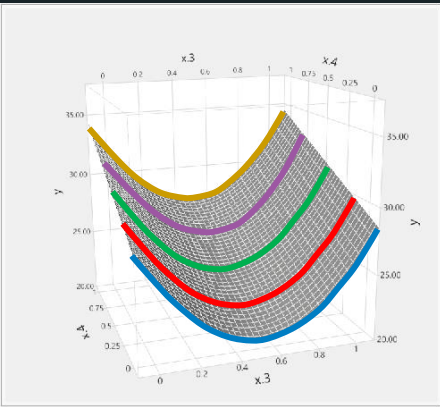
(2 interaction effects - twists)



Overlaying by x_4 to See *Curvature* in x_3 without an Interaction between x_3 & x_4

$$y = 7 + 20x_1 + 17x_2 + (-20)x_3 + 10x_4 + 6x_5 + (-13)x_1^2 + (-13)x_2^2 + 20x_3^2 + 19x_1x_2 + (-10)x_4x_5$$

(intercept)
 (5 main effects - *slopes*)
 (3 squared effects - *curvature*)
 (2 interaction effects - *twists*)

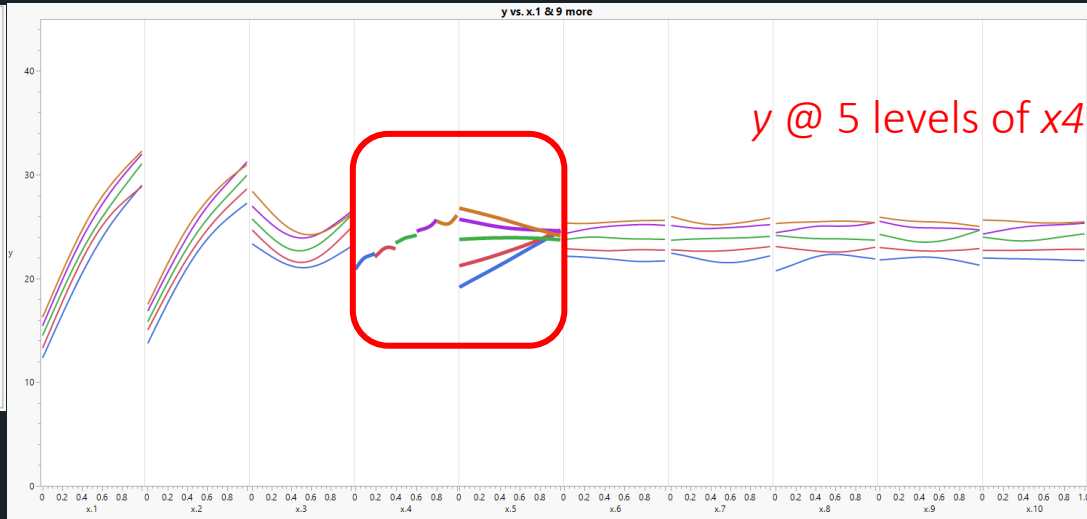
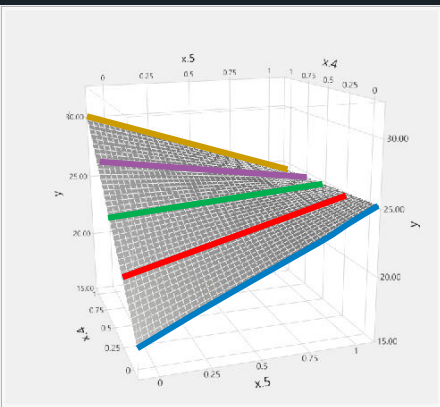


x3	lin	sq	Sum int+lin+sq
0	0	0	0
0.1	-2	0.2	-1.8
0.2	-4	0.8	-3.2
0.3	-6	1.8	-4.2
0.4	-8	3.2	-4.8
0.5	-10	5	-5
0.6	-12	7.2	-4.8
0.7	-14	9.8	-4.2
0.8	-16	12.8	-3.2
0.9	-18	16.2	-1.8
1	-20	20	0

Overlaying by x_4 to See *Interaction* with x_5 without any Curvature in x_4 & x_5

$$y = 7 + 20*x_1 + 17*x_2 + (-20)*x_3 + 10*x_4 + 6*x_5 + (-13)*x_1*x_1 + (-13)*x_2*x_2 + 20*x_3*x_3 + 19*x_1*x_2 + (-10)*x_4*x_5$$

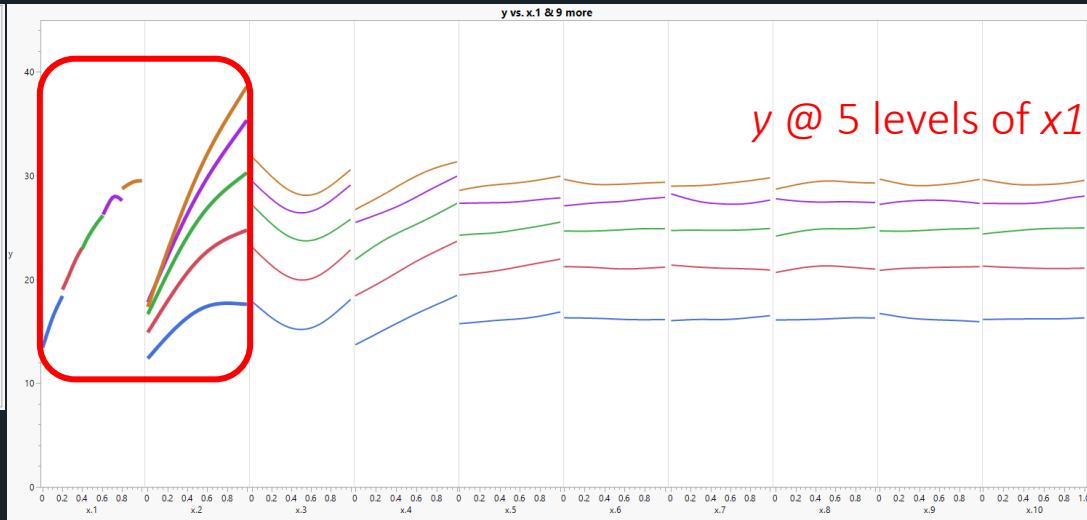
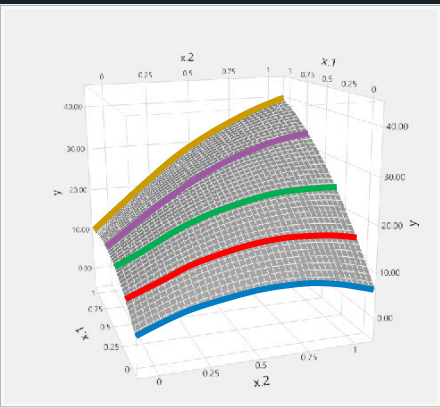
(intercept)
(5 main effects - *slopes*)
(3 squared effects - *curvature*)
(2 interaction effects - *twists*)



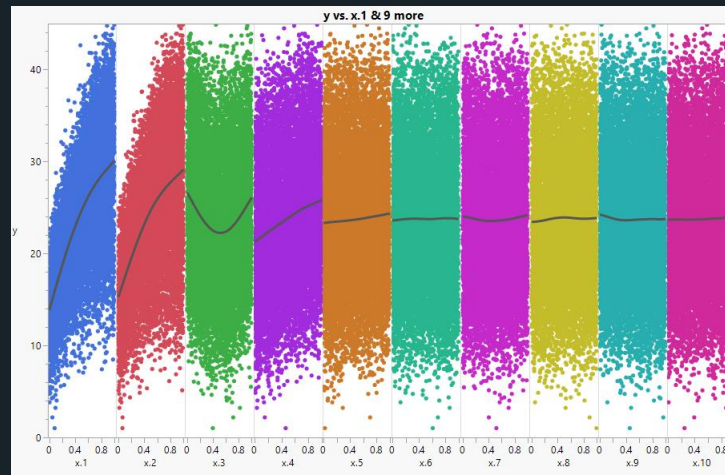
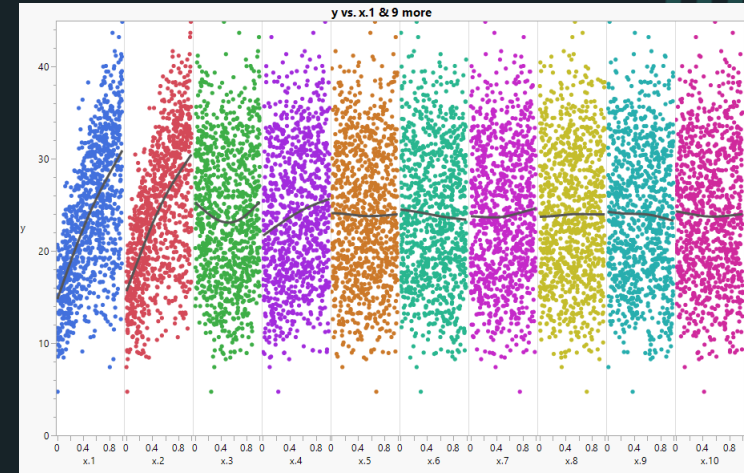
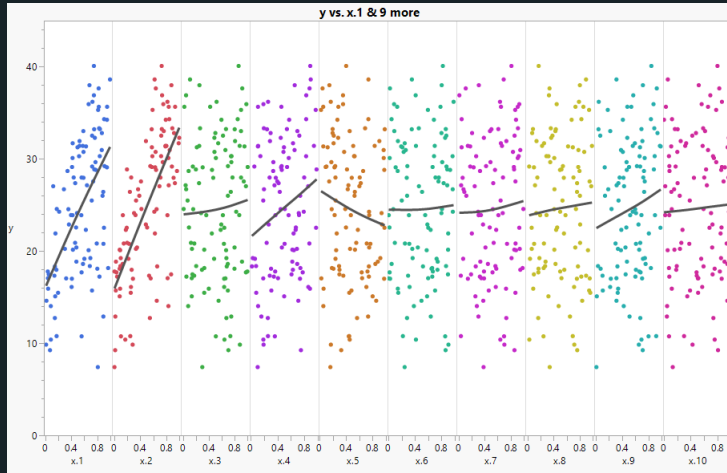
Overlaying by x_1 to See *Interaction* with x_2 and *Curvature* in both x_1 & x_2

$$y = 7 + 20*x_1 + 17*x_2 + (-20)*x_3 + 10*x_4 + 6*x_5 + (-13)*x_1*x_1 + (-13)*x_2*x_2 + 20*x_3*x_3 + 19*x_1*x_2 + (-10)*x_4*x_5$$

(intercept)
(5 main effects - *slopes*)
(3 squared effects - *curvature*)
(2 interaction effects - *twists*)



Compare Results for 100, 1,000, and 10,000 Data Points



The more data used, the better the spline fit approximates the generating function

Generating Function for Data Points:

$$y = 7$$

$$+ 20 * x_1 + 17 * x_2 + (-20) * x_3 + 10 * x_4 + 6 * x_5$$

$$+ (-13) * x_1 * x_1 + (-13) * x_2 * x_2 + 20 * x_3 * x_3$$

$$+ 19 * x_1 * x_2 + (-10) * x_4 * x_5$$

$$+ \text{Random Normal} [\text{Mean} = 0, \text{StdDev} = 2.8]$$

After Doing Visual Data Discovery, Run *Predictor Screening* to Quickly Rank the 200 Factors. There are only 5 active factors (x1, x2, x3, x4, & x5), ALL of which are Correctly Identified.

$$y = 7$$

$$+ 20*x1 + 17*x2 + (-20)*x3 + 10*x4 + 6*x5$$
$$+ (-13)*x1*x1 + (-13)*x2*x2 + 20*x3*x3$$
$$+ 19*x1*x2 + (-10)*x4*x5$$

Predictor	y		Rank ^
	Contribution	Portion	
x.1	143306	0.5326	1
x.2	104031	0.3867	2
x.4	9826	0.0365	3
x.3	6244	0.0232	4
x.5	1246	0.0046	5
x.119	36	0.0001	6
x.168	34	0.0001	7
x.75	31	0.0001	8
x.145	31	0.0001	9
x.22	31	0.0001	10

Top 10 of 200 Shown

With Reduced Factor List, Easy to Use Machine Learning in JMP® Pro

- Partition (Regression/Decision Tree) – very fast
- Bootstrap Forest* used in Predictor Screening Averages Many Trees
 - Fast selection of important factors when you have many
 - Bootstrapping randomly picks factors so dominant ones are sometimes absent so they do not overshadow/obscure less important factors
- Neural Network – one or two layers, and boosted
 - Often best predictor, but tends to overfit
 - More computationally intensive than trees, especially with many factors
- Regression – Ordinary Least Squares, Logistic, Stepwise, or Penalized
 - More computationally intensive than trees, especially with many factors
 - Penalized regression can perform variable selection even with correlated data
 - Analyses provide confidence intervals

*JMP name for Random Forest™

Model Screening with Many Factors (250) and Many Rows of Data (10K)

- Model Screening results (left) with 250 factors and 10,000 rows of data is **too big** a problem for some methods (regression, SVM) or **times out** for others (neural network).
- After reviewing **results for fast tree methods** (e.g., Bootstrap Forest) use **top ranked** factors and rerun Model Screening (right) to **obtain more and better outcomes**.

Test			
Method	N	RSquare	RASE
Boosted Tree	2001	0.8380	2.9795
Bootstrap Forest	2001	0.8039	3.2782
Decision Tree	2001	0.7649	3.5895
Neural Boosted	2001	0.5467	4.9838
K Nearest Neighbors	2001	0.1832	6.6900

[Select Dominant](#) [Run Selected](#) [Save Script Selected](#)

Support Vector Machines avoided when number of rows > 10000. Linear Model methods skipped because the number of parameters, 753, exceeded the limit of 450. Sum Freq and Sum Weight are suppressed when they are the same as N.

Method	Time Limit Exceeded Count
Neural Boosted	1

Test				
Method	Details	N	RSquare	RASE
Fit Stepwise	2FI Quad	2001	0.8541	2.8272
Fit Least Squares	2FI Quad	2001	0.8541	2.8280
Generalized Regression Lasso	2FI Quad	2001	0.8541	2.8280
Neural Boosted		2001	0.8500	2.8666
Boosted Tree		2001	0.8424	2.9382
Bootstrap Forest		2001	0.8332	3.0237
K Nearest Neighbors		2001	0.8217	3.1259
Decision Tree		2001	0.7744	3.5161
Fit Least Squares		2001	0.7218	3.9040
Generalized Regression Lasso		2001	0.7218	3.9046
Fit Stepwise		2001	0.7217	3.9053

[Select Dominant](#) [Run Selected](#) [Save Script Selected](#)

Support Vector Machines avoided when number of rows > 10000. Sum Freq and Sum Weight are suppressed when they are the same as N.

Model Screening

Run 10 selected models all at once (about half a minute in this case) then view ranked performance

Model Screening - JMP EA [2]

Fits many different predictive models and provides summaries of measures of fit.

Select Columns

Cast Selected Columns into Roles

Method

- Decision Tree
- Bootstrap Forest
- Boosted Tree
- K Nearest Neighbors
- Naive Bayes
- Neural
- Support Vector Machines
- Discriminant
- Fit Least Squares
- Fit Stepwise
- Logistic Regression
- Generalized Regression

Modeling Options

- Add Two Way Interactions
- Add Quadratics
- Informative Missing

Operational Options

Set Random Seed

Time Limit Each

- Remove Live Reports
- Show Methods in Log

Folded Crossvalidation

Fit repeatedly with sequenced folds.

- K Fold Crossvalidation K 5
- Nested Crossvalidation K 4 L 5

Repeated K Fold 0

Model Screening for y

Table: 250factors 5sig 10K rows Response: y Validation: Validation 60-20-20

Details

Training

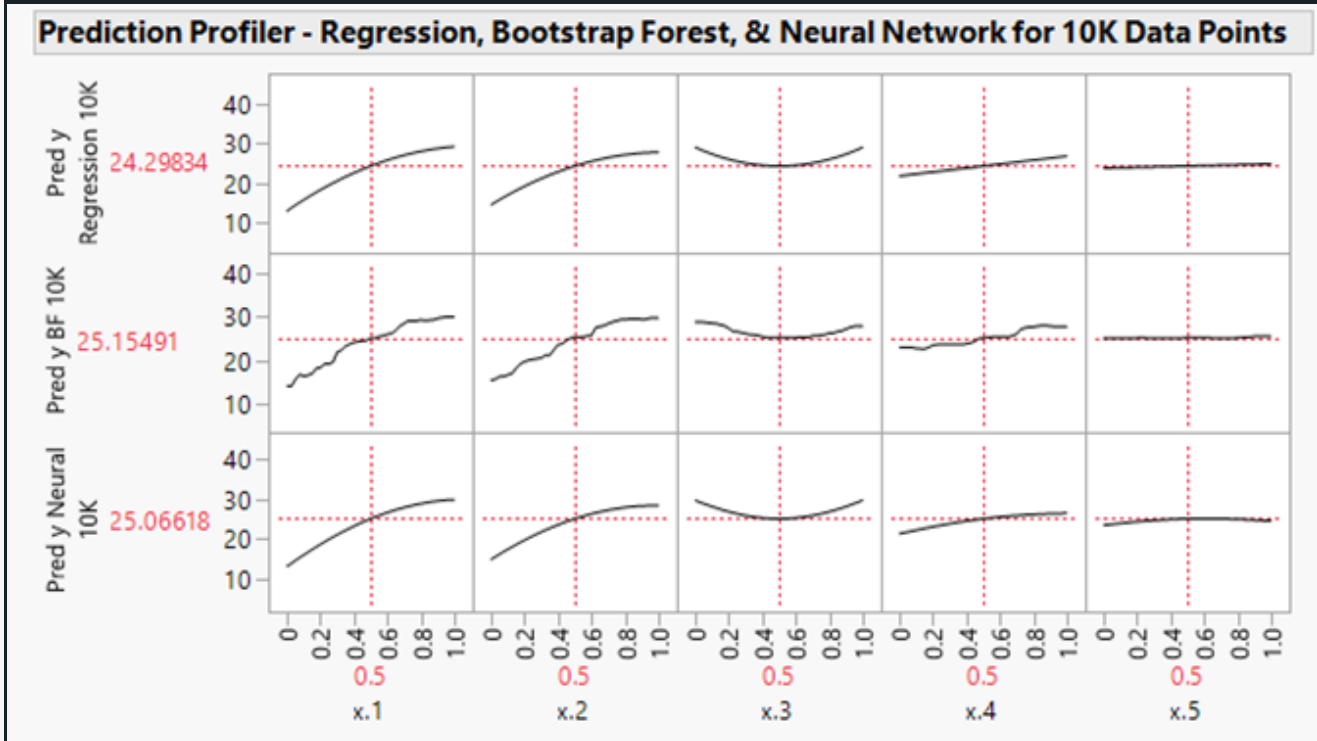
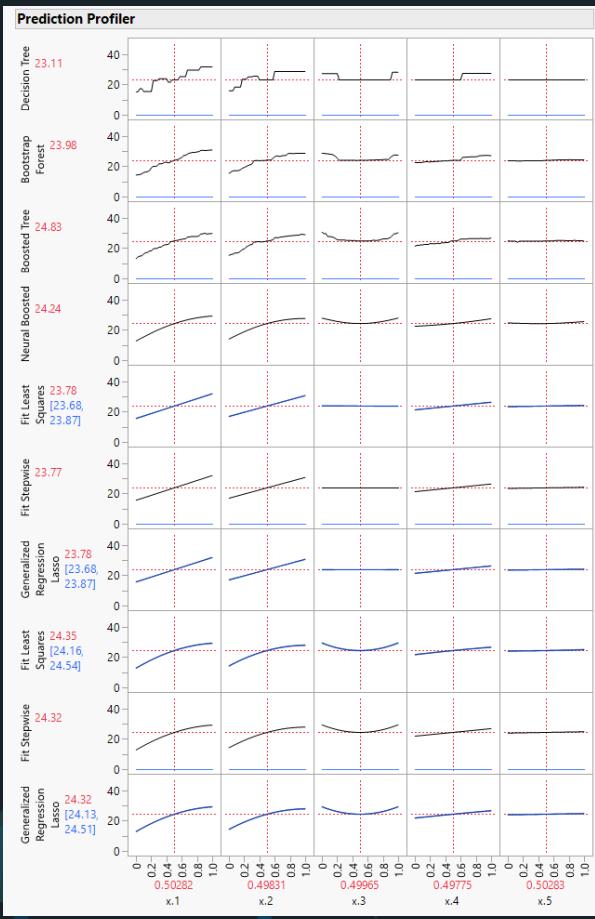
Validation

Method	Details	N	RSquare	RASE
Fit Stepwise	2FI Quad	1991	0.8553	2.7927
Generalized Regression Lasso	2FI Quad	1991	0.8552	2.7935
Fit Least Squares	2FI Quad	1991	0.8549	2.7967
Neural Boosted		1991	0.8510	2.8334
Boosted Tree		1991	0.8387	2.9487
Bootstrap Forest		1991	0.8319	3.0100
K Nearest Neighbors		1991	0.8183	3.1297
Decision Tree		1991	0.7793	3.4486
Generalized Regression Lasso		1991	0.7244	3.8542
Fit Stepwise		1991	0.7242	3.8555
Fit Least Squares		1991	0.7241	3.8558

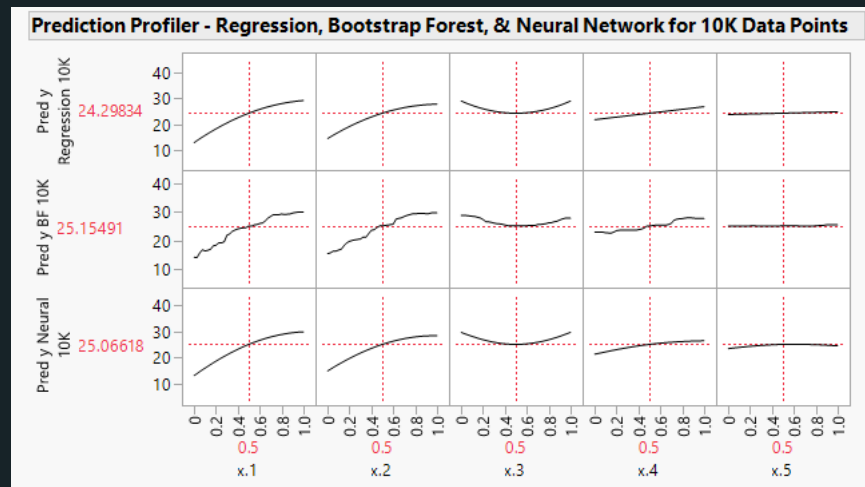
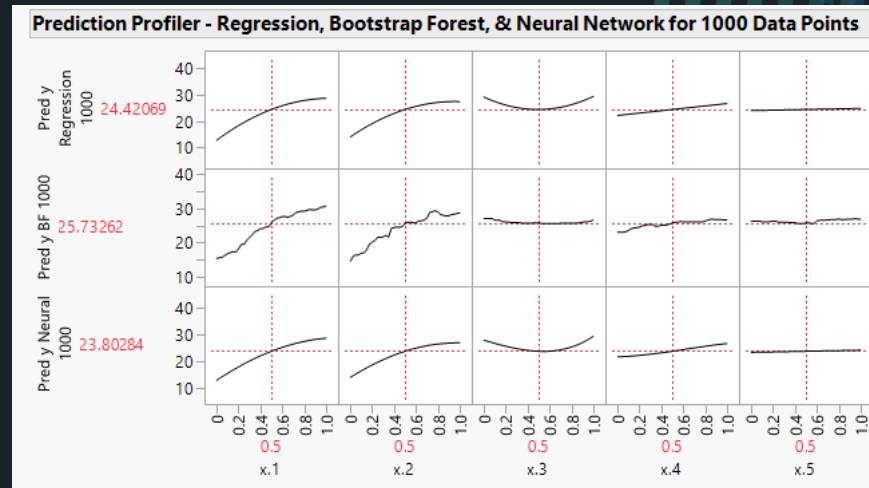
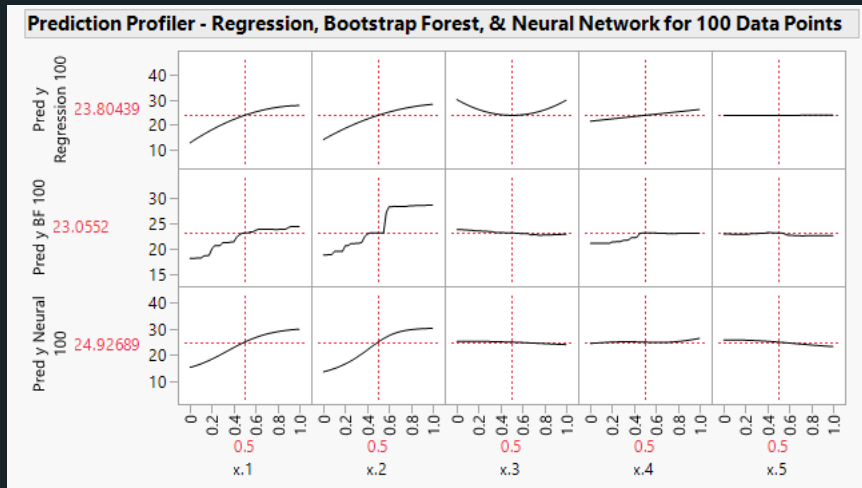
Select Dominant Run Selected Save Script Selected

Publish models in:
Python, C, SAS, SQL, & JavaScript

Profilers for All 10 Models (left) & 3 Selected Models (right)



Compare Results for 100, 1,000, and 10,000 Data Points

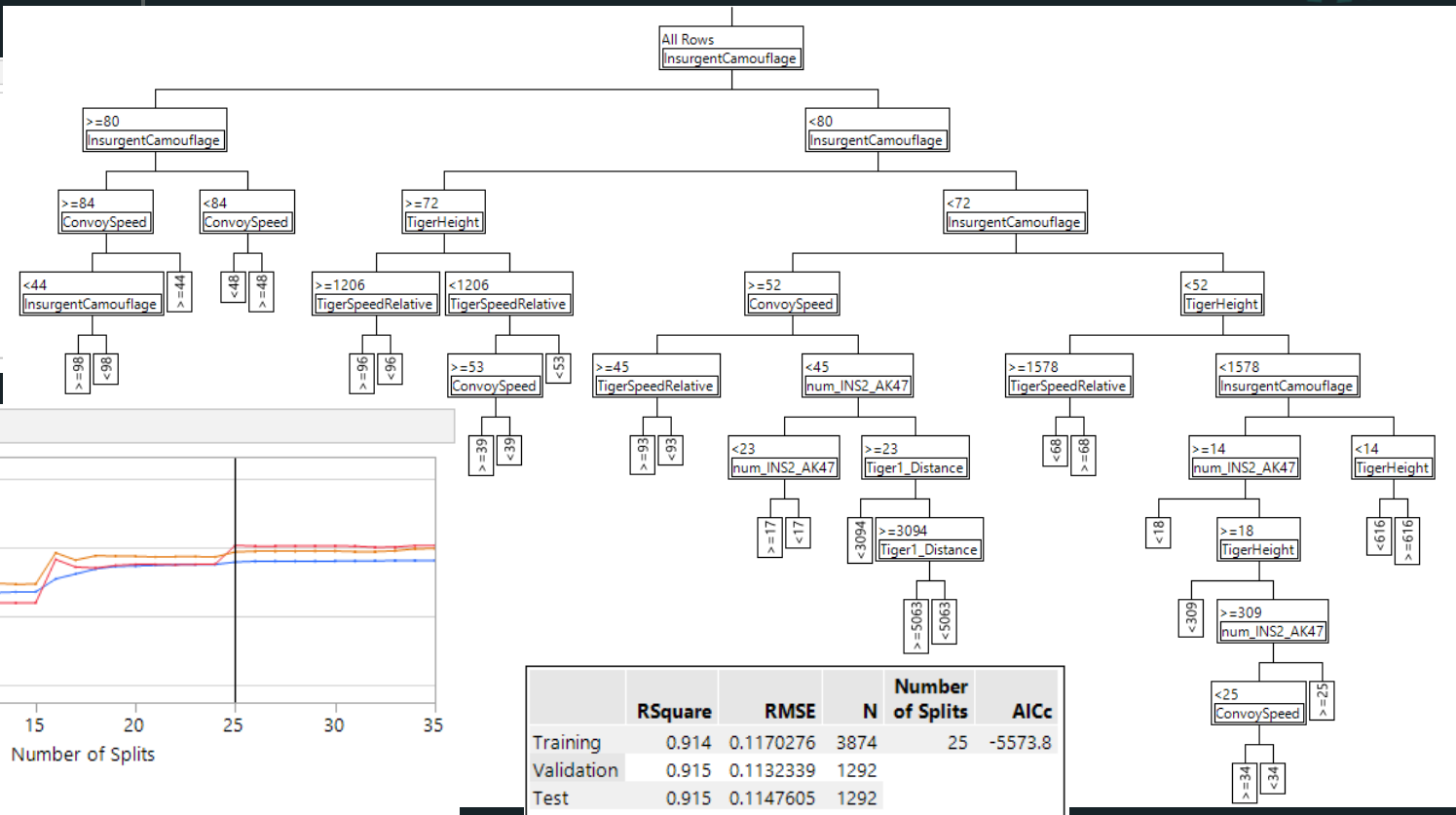
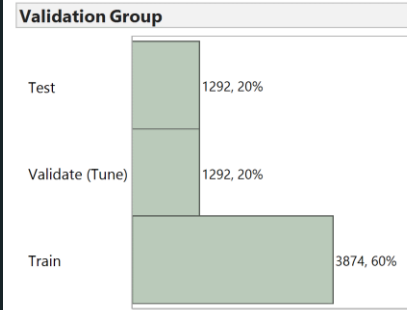


- Three ML methods per group
- Same amount of data in each group

With enough data all three ML methods well approximate the generating function

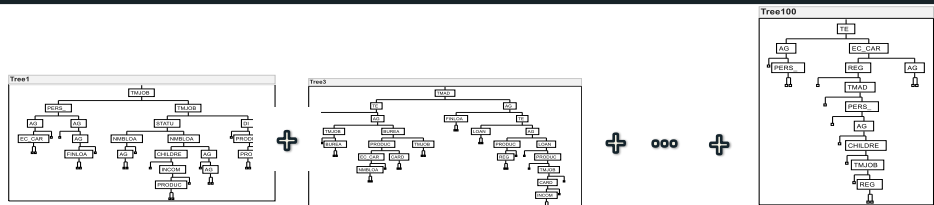
Honest Assessment

SUBSET DATA TO CREATE TRAIN, VALIDATE(TUNE), & TEST GROUPS
USE VALIDATE(TUNE) GROUP TO PREVENT OVER-FITTING DATA MINING MODELS

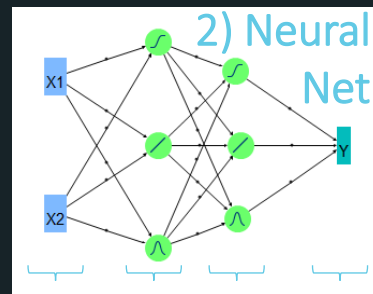


Robust Strategy for Machine Learning

- 1) Bootstrap Forest – FAST even w/many Xs – Unlikely to Miss Less Dominant Xs
- 2) Neural Network – Often Most Flexible & Best Predictor – Tendency to Overfit
- 3) Penalized Regression – Often More Interpretable Model + Confidence Intervals



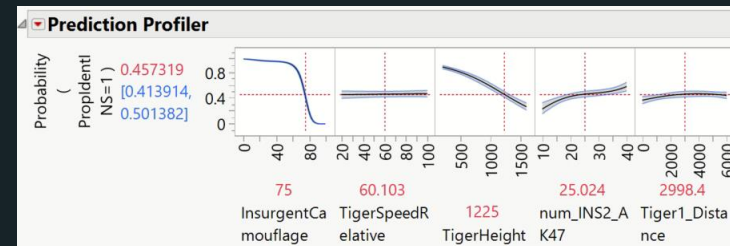
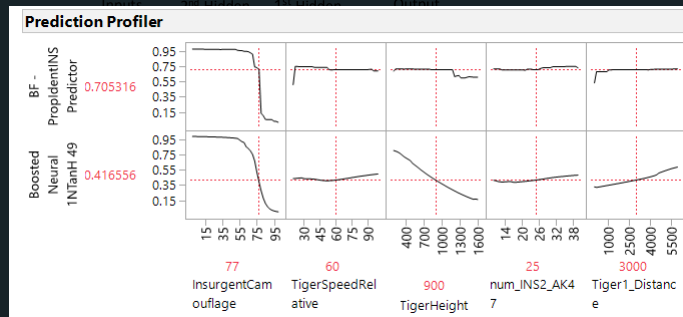
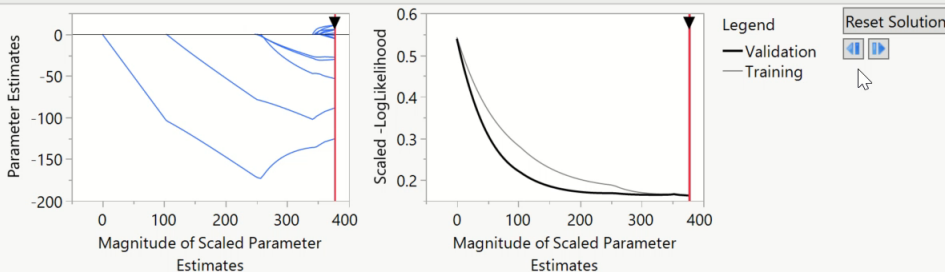
100 1) Bootstrap Forest



Measure	Training	Validation	Test
Generalized RSquare	0.8026309	0.8019205	0.8069096
Lambda Penalty	0.2258932		

3) Penalized Regression

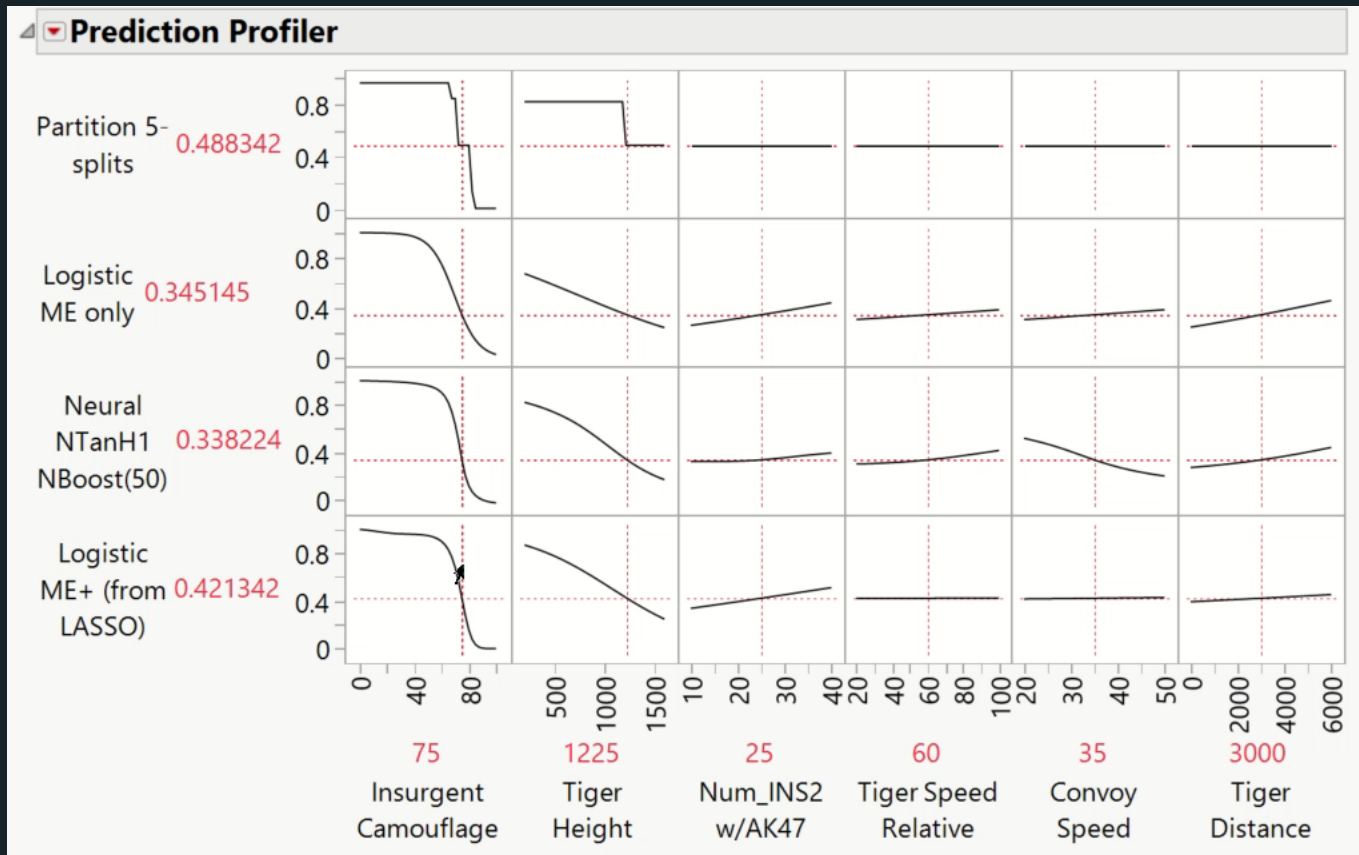
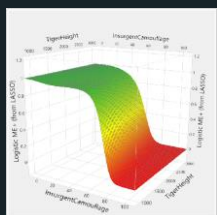
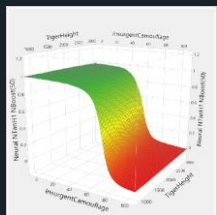
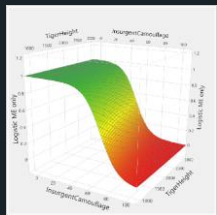
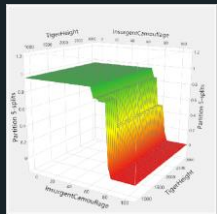
Solution Path



NOT just better prediction, but better understanding!

Visually Compare Multiple Machine Learning Models

Partition with 5-Splits, Logistic Regression, Neural Network, & LASSO (Binomial)



Simplest

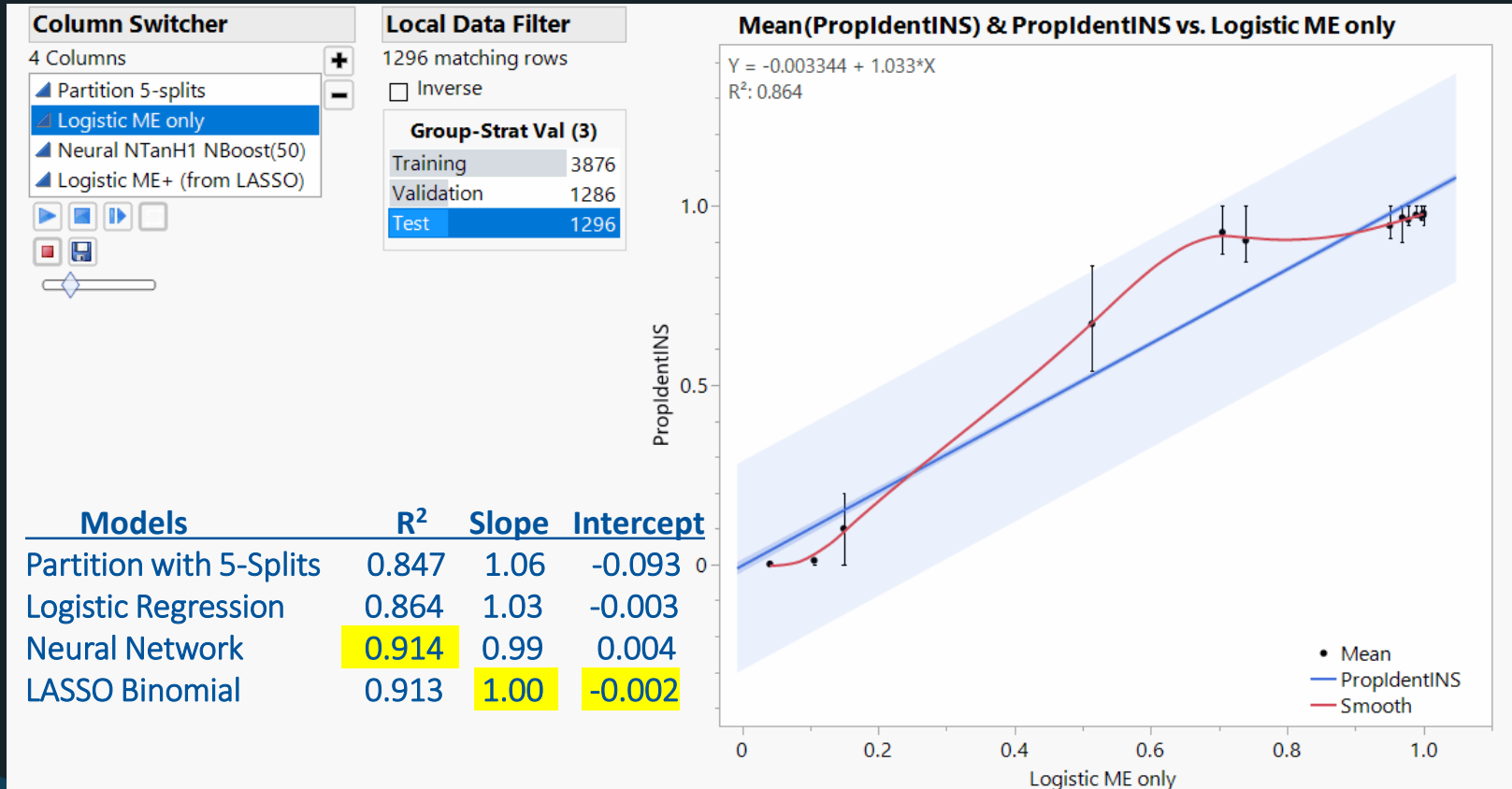
Classic

Best Predictor

More Interpretable

“LASSO” stands for Least Absolute Shrinkage and Selection Operator

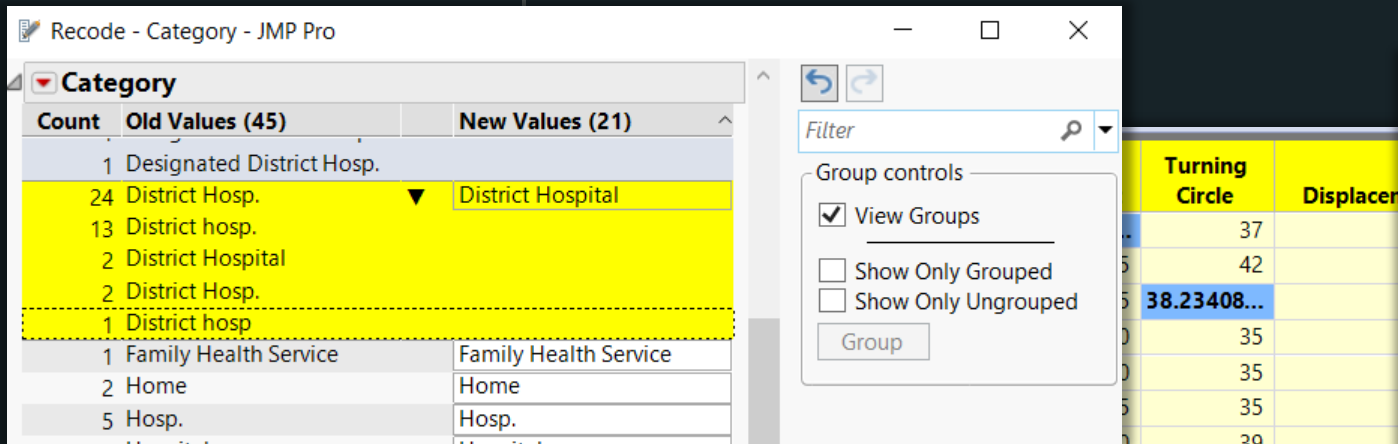
Actual vs. Prediction Plots for 4 Surrogate Models for *Test Data ONLY* (Not used in fitting or tuning the models)



Where(Group-Strat Val = Test)
 Each error bar is constructed using the upper and lower quartiles.

Data Curation

RECODE, OUTLIER DETECTION, AND IMPUTE MISSING VALUES, STACK, SPLIT, ETC.
"WHAT YOU JUST DID IN MINUTES TAKES ME HOURS (DAYS!) IN EXCEL..."



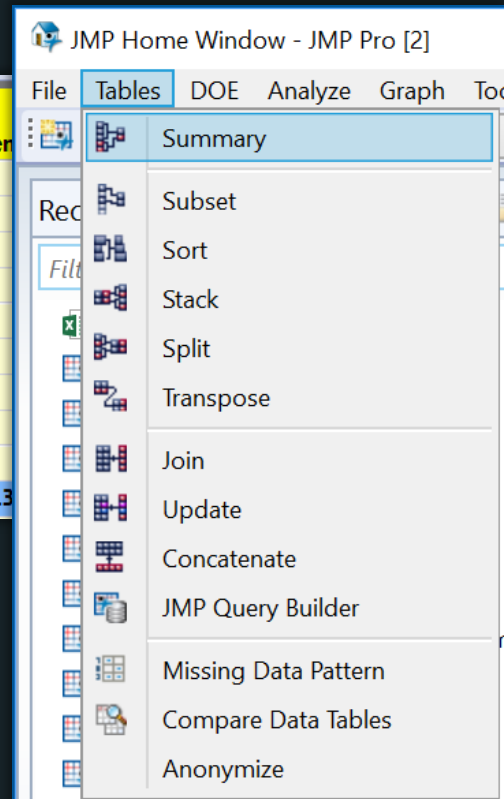
Recode - Category - JMP Pro

Count	Old Values (45)	New Values (21)
1	Designated District Hosp.	
24	District Hosp.	District Hospital
13	District hosp.	
2	District Hospital	
2	District Hosp.	
1	District hosp	
1	Family Health Service	Family Health Service
2	Home	Home
5	Hosp.	Hosp.

Group controls

- View Groups
- Show Only Grouped
- Show Only Ungrouped

Group

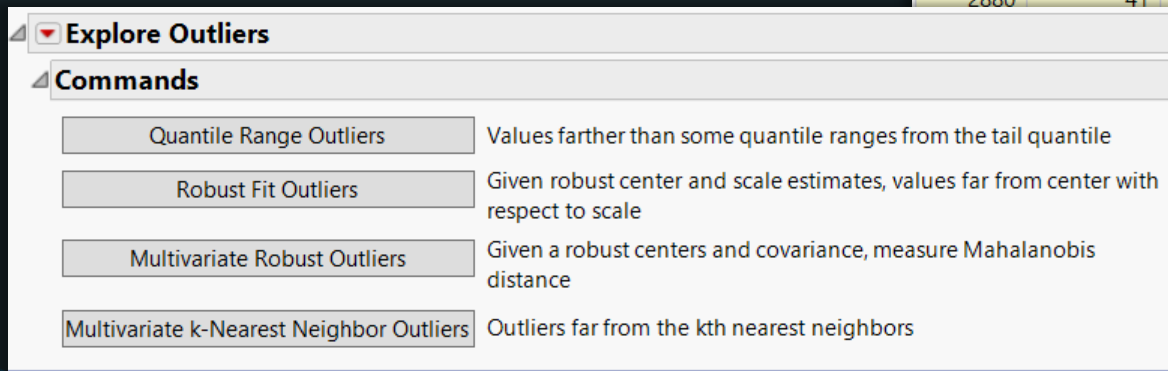


JMP Home Window - JMP Pro [2]

File Tables DOE Analyze Graph To

Summary

- Subset
- Sort
- Stack
- Split
- Transpose
- Join
- Update
- Concatenate
- JMP Query Builder
- Missing Data Pattern
- Compare Data Tables
- Anonymize



Explore Outliers

Commands

- Quantile Range Outliers: Values farther than some quantile ranges from the tail quantile
- Robust Fit Outliers: Given robust center and scale estimates, values far from center with respect to scale
- Multivariate Robust Outliers: Given a robust centers and covariance, measure Mahalanobis distance
- Multivariate k-Nearest Neighbor Outliers: Outliers far from the kth nearest neighbors

Design of Experiments

DERIVE MAXIMUM INFORMATION FROM FEWEST TESTS – YIELDING “INTERACTIVE” OPTIMIZATION AND TRADE-SPACE ANALYSIS

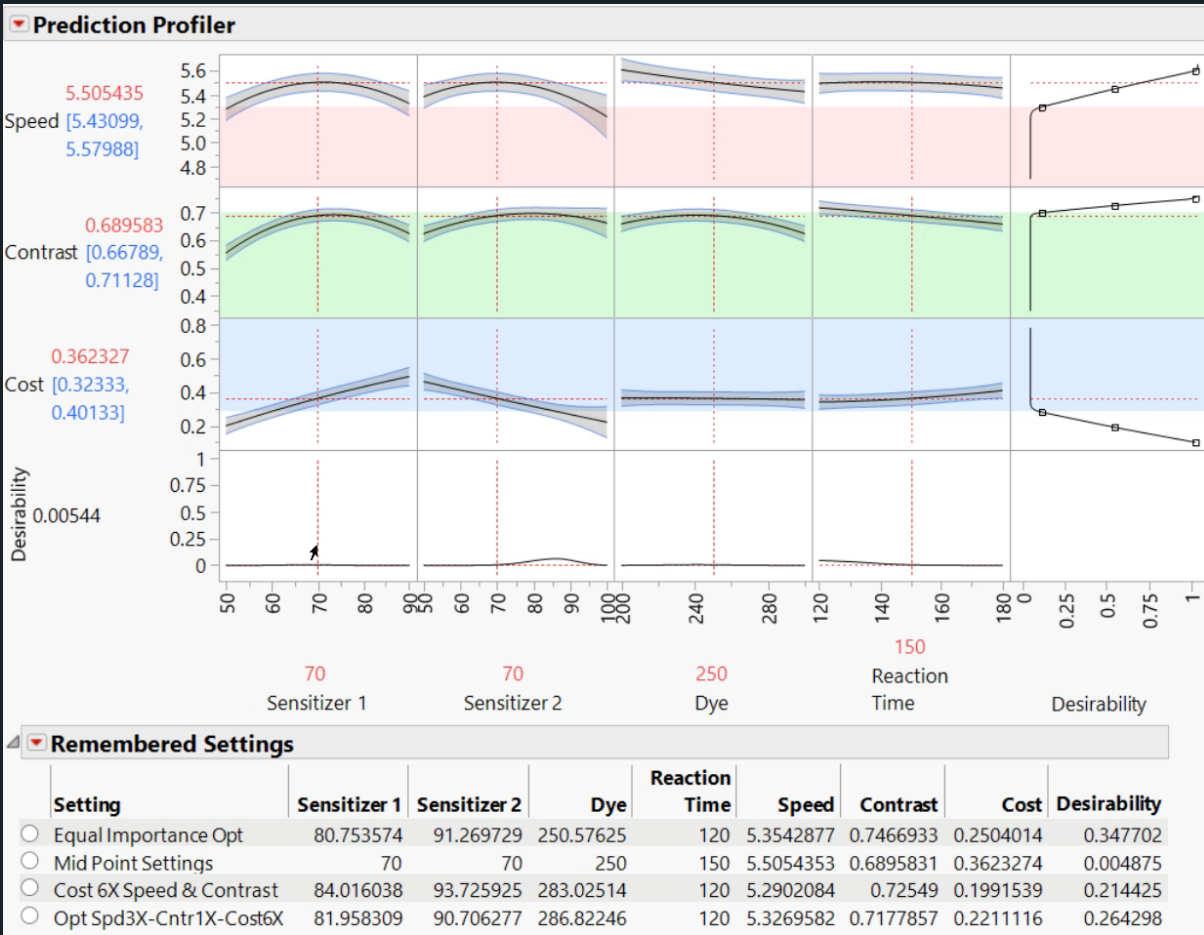
Photo_Cost27 - JMP Pro [3]

File Edit Tables Rows Cols DOE Analyze Graph Tools Add-Ins View Window Help

7/4 Cols

27/0

	Sensitizer 1	Sensitizer 2	Dye	Reaction Time	Speed	Contrast	Cost
1	50	50	250	120	5.36	0.616	0.198
2	50	50	200	180	5.39	0.537	0.175
3	90	70	200	120	5.31	0.623	0.447
4	50	90	200	150	5.13	0.431	0.177
5	70	70	250	180	5.37	0.643	0.445
6	50	90	300	120	4.79	0.375	0.231
7	90	90	200	180	5.45	0.626	0.471
8	90	50	250	150	5.00	0.470	0.670
9	50	50	300	150	5.22	0.478	0.283
10	70	90	200	120	5.41	0.668	0.226
11	90	90	250	120	5.33	0.734	0.310
12	50	50	250	120	5.32	0.574	0.257
13	70	50	200	150	5.49	0.596	0.456
14	50	70	250	180	5.22	0.558	0.166
15	70	70	250	150	5.57	0.689	0.390
16	90	90	300	150	5.26	0.653	0.226
17	70	70	250	150	5.47	0.688	0.356
18	70	70	300	120	5.42	0.657	0.337
19	50	70	200	120	5.43	0.518	0.222
20	50	50	300	150	5.15	0.505	0.287
21	90	70	200	120	5.33	0.661	0.457
22	50	90	300	120	4.97	0.411	0.191
23	90	50	300	120	5.09	0.492	0.588
24	90	50	300	180	5.03	0.358	0.733
25	70	70	250	150	5.59	0.707	0.318
26	70	90	300	180	5.25	0.605	0.290
27	50	90	200	150	5.24	0.476	0.177



Design of Experiments

DERIVE MAXIMUM INFORMATION FROM FEWEST TESTS –
YIELDING “INTERACTIVE” OPTIMIZATION AND TRADE-SPACE ANALYSIS

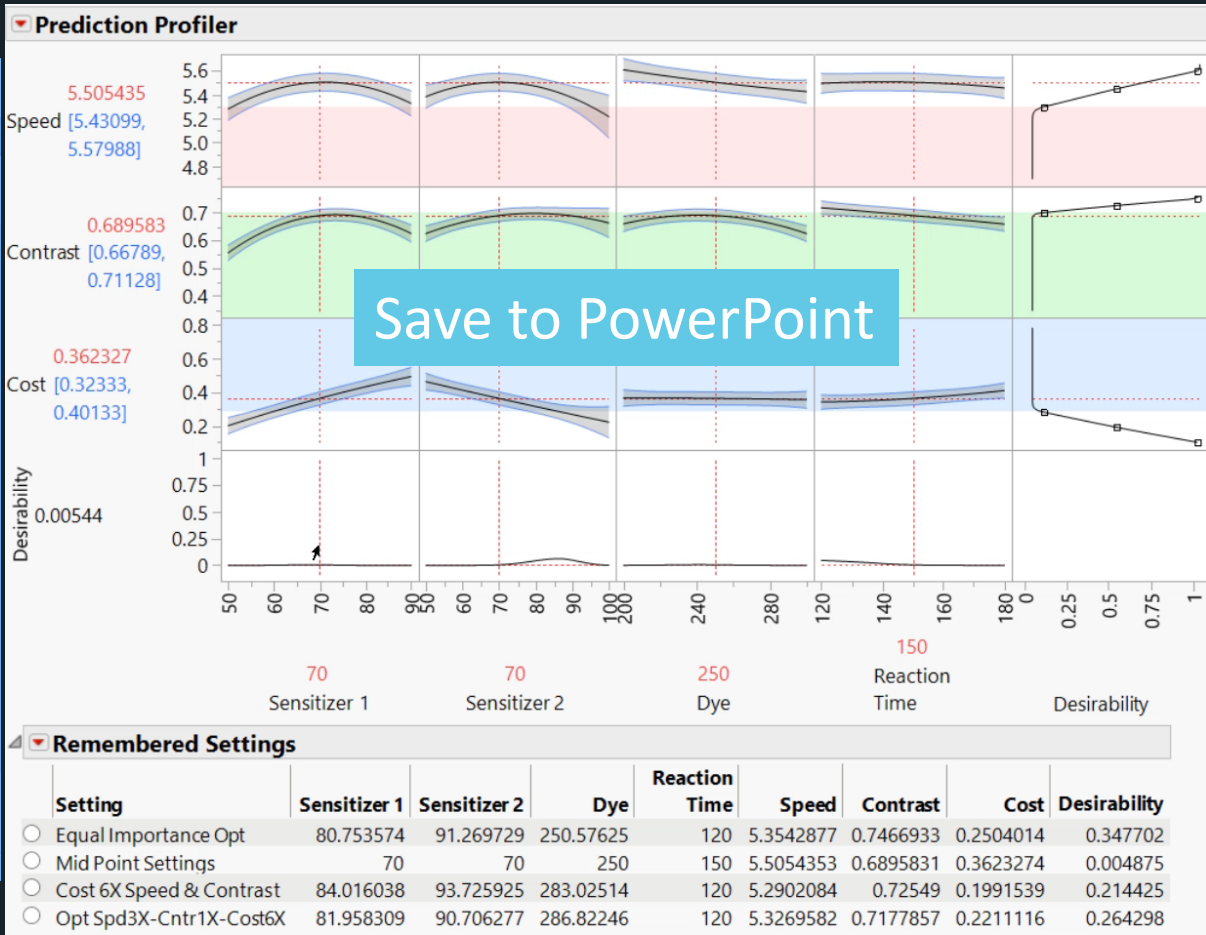
Photo_Cost27 - JMP Pro [3]

File Edit Tables Rows Cols DOE Analyze Graph Tools Add-Ins View Window Help

7/4 Cols

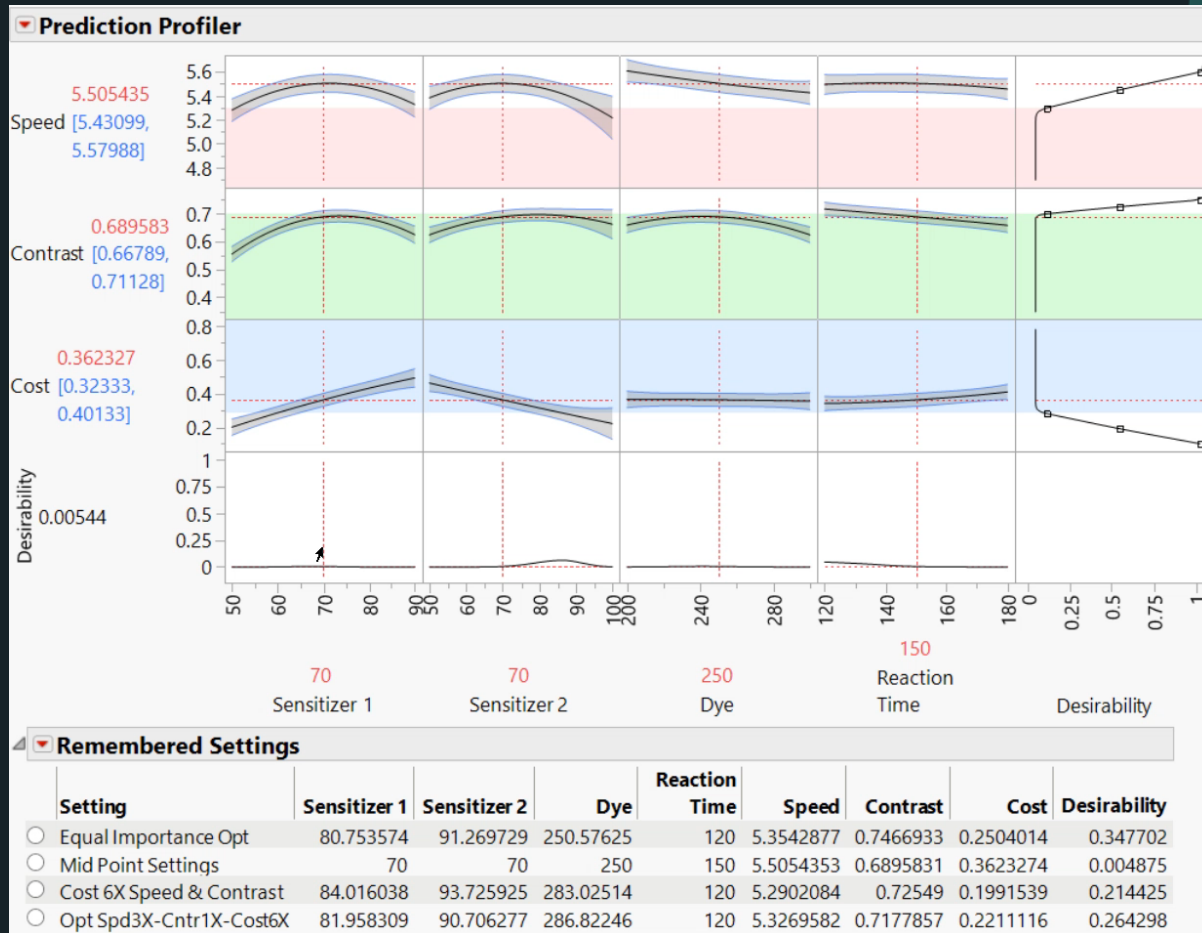
27/0

	Sensitizer 1	Sensitizer 2	Dye	Reaction Time	Speed	Contrast	Cost
1	50	50	250	120	5.36	0.616	0.198
2	50	50	200	180	5.39	0.537	0.175
3	90	70	200	120	5.31	0.623	0.447
4	50	90	200	150	5.13	0.431	0.177
5	70	70	250	180	5.37	0.643	0.445
6	50	90	300	120	4.79	0.375	0.231
7	90	90	200	180	5.45	0.626	0.471
8	90	50	250	150	5.00	0.470	0.670
9	50	50	300	150	5.22	0.478	0.283
10	70	90	200	120	5.41	0.668	0.226
11	90	90	250	120	5.33	0.734	0.310
12	50	50	250	120	5.32	0.574	0.257
13	70	50	200	150	5.49	0.596	0.456
14	50	70	250	180	5.22	0.558	0.166
15	70	70	250	150	5.57	0.689	0.390
16	90	90	300	150	5.26	0.653	0.226
17	70	70	250	150	5.47	0.688	0.356
18	70	70	300	120	5.42	0.657	0.337
19	50	70	200	120	5.43	0.518	0.222
20	50	50	300	150	5.15	0.505	0.287
21	90	70	200	120	5.33	0.661	0.457
22	50	90	300	120	4.97	0.411	0.191
23	90	50	300	120	5.09	0.492	0.588
24	90	50	300	180	5.03	0.358	0.733
25	70	70	250	150	5.59	0.707	0.318
26	70	90	300	180	5.25	0.605	0.290
27	50	90	200	150	5.24	0.476	0.177



OPTIMIZATION AND TRADE-SPACE ANALYSIS

Informing Decision Discussions

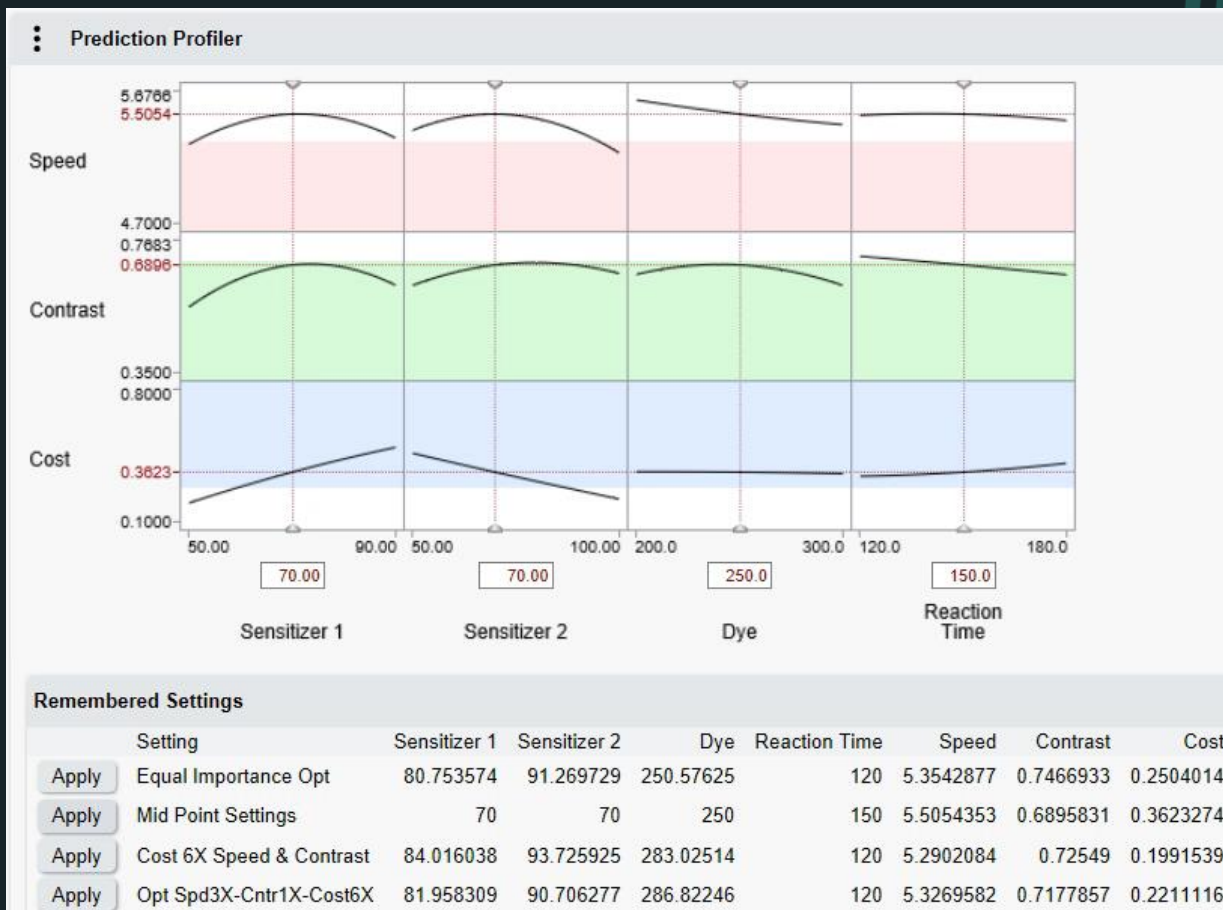


OPTIMIZATION AND TRADE-SPACE ANALYSIS

Share
Analysis
via HTML

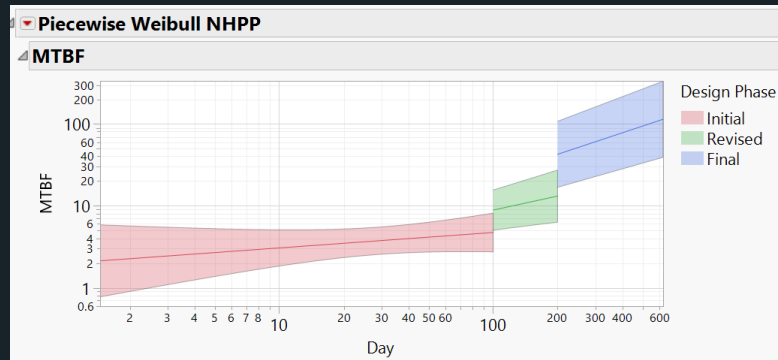
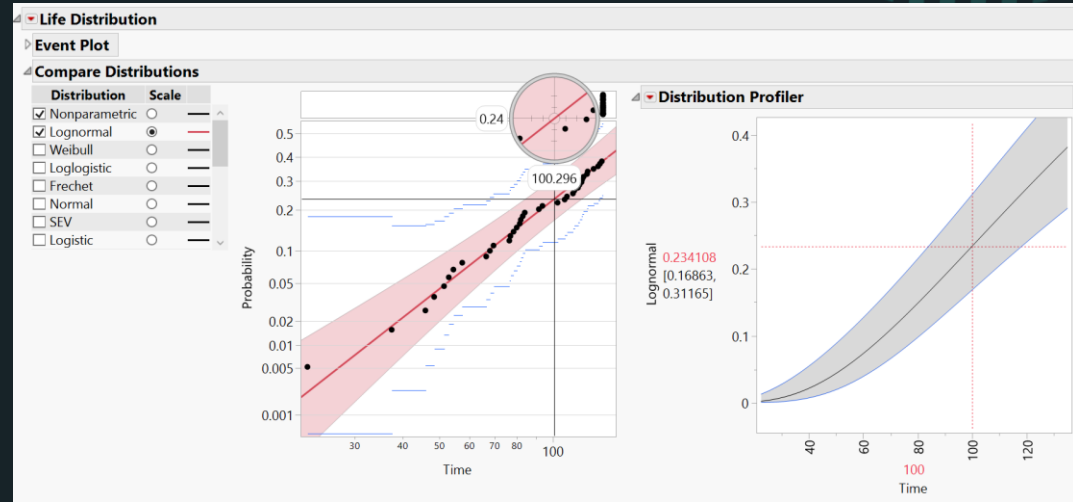
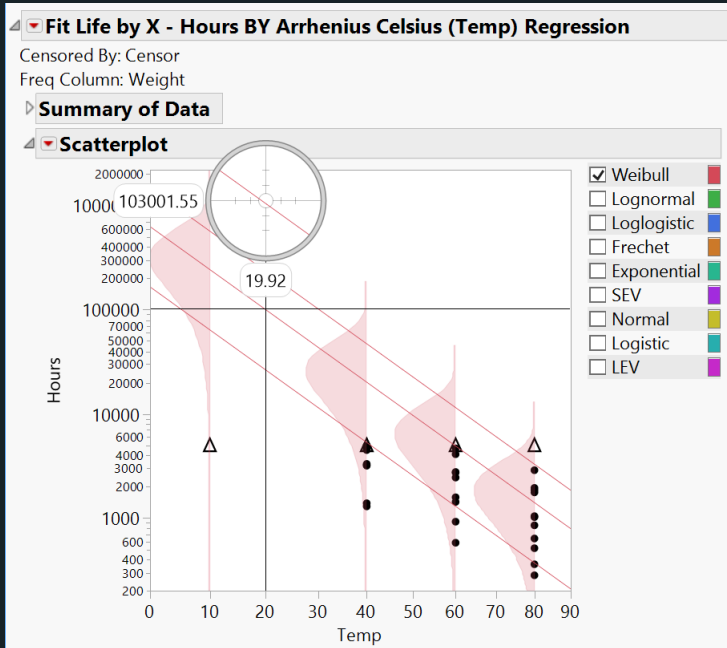
View optimizations
on your phone.

Scan the QR code
to launch browser,
then use finger to
interact with the
Prediction Profiler
and to "Apply"
saved settings.



Reliability

PROBABILITY OF FAILURE, ACCELERATED LIFE TEST, RELIABILITY GROWTH



Fast Surrogate Modeling

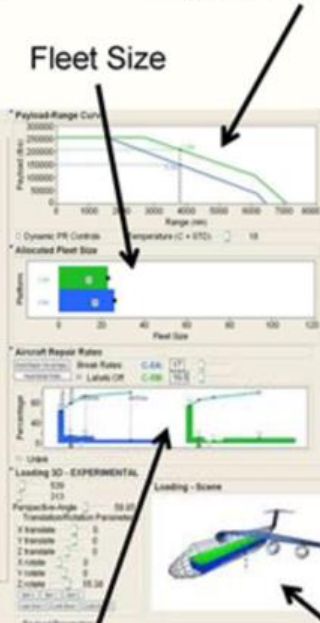
BY MODELING THE "RIGHT NUMBER" OF LONG RUNNING SIMULATIONS (THINK SEQUENTIAL DOE), AN ACCURATE & INSTANTANEOUS PREDICTION CAN BE MADE FOR ANY NEW SCENARIO!

Mission Scenario Inputs



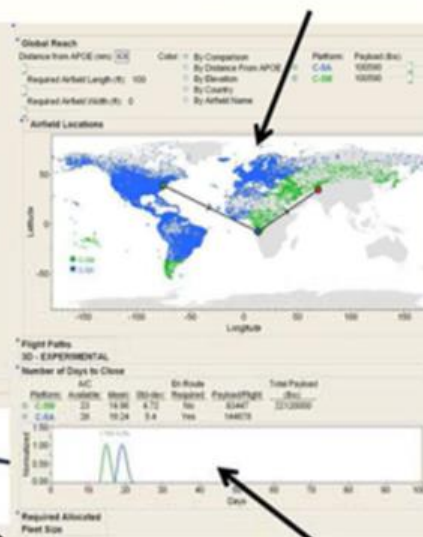
Loading Curve

Payload-Range Curve



Break and Repair Rates

Global Reach and Airfield Locations



Time to Close

Logistic Metrics



Bubble Plot

Download & Recording

- 16 Factors
- 50,000 unique cases
- Each 1,000 times
- 50 Million Simulations
- Neural Network for Surrogate Models

1.6 Comparative Assessment and Decision Support System for Strategic Military Airlift Capability

Comparative Assessment and Decision Support System for Strategic Military Airlift Capability

John Salmon, Curtis Iwata, Dimitri Mavris and Neil Weston

Georgia Institute of Technology

*john.salmon@asdl.gatech.edu, curtis.iwata@asdl.gatech.edu,
dimitri.mavris@aerospace.gatech.edu, neil.weston@ae.gatech.edu*

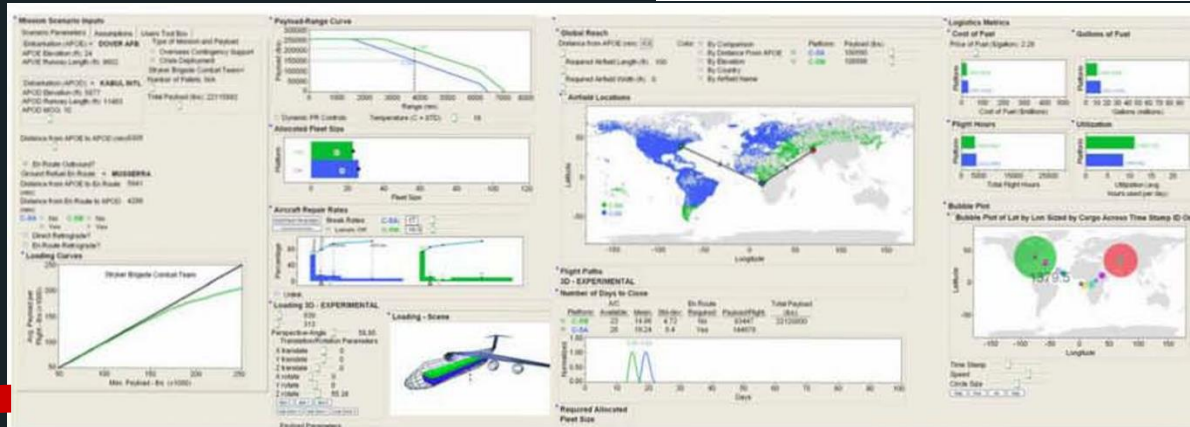
Philip Fahringer

Lockheed Martin Company

philip.fahringer@lmco.com

ABSTRACT

The Lockheed Martin Aeronautics Company has been awarded several programs to modernize the aging C-5 military transport fleet. In order to ensure its continuation amidst budget cuts, it was important to engage the decision makers by providing an environment to analyze the benefits of the modernization program. This paper describes an interface that allows the user to change inputs such as the scenario airfields, take-off conditions, and reliability characteristics. The underlying logistics surrogate model was generated using data from a discrete-event simulation. Various visualizations, such as intercontinental flight paths illustrated in 3D, have been created to aid the user in analyzing scenarios and performing comparative assessments for various output logistics metrics. The capability to rapidly and dynamically evaluate and compare scenarios was developed enabling real-time strategy exploration and trade-offs.



Recording

Figure 2. Strategic Airlift Comparison Tool Layout

Download Document

<https://ntrs.nasa.gov/search.jsp?R=20110012110>

TABULATE

JMP'S DRAG & DROP PIVOT TABLE – EASILY MAKE INTO NEW DATA TABLE

The screenshot displays the JMP Pro interface. The main window is titled "Navy Obligated Money by Congressional District - Tabulate - JMP Pro". The "Tabulate" window shows a pivot table with columns for "award_type", "action_type", and "Total \$ Obligated in \$M". A context menu is open over the pivot table, with "Make Into Data Table" selected. A tooltip explains: "Creates a new data table from the table created in Tabulate." Below the menu, various statistical and data manipulation options are listed, such as "Local Data Filter", "Redo", "Save Script", and various summary statistics like "Sum", "Mean", "Variance", etc.

An inset window titled "Untitled 3 - JMP Pro" shows the resulting data table. The table has the following structure:

	award_type	action_type	Sum(Total \$ Obligated in \$M)	N
1	BPA CALL	ADDITIONAL...	\$4.4	7
2	BPA CALL	CHANGE ...	\$1.0	70
3	BPA CALL	CLOSE OUT	\$2.5	134
4	BPA CALL	EXERCISE ...	\$63.6	40
5	BPA CALL	FUNDING ...	\$115.9	224
6	BPA CALL	LEGAL ...	\$0.0	1
7	BPA CALL	OTHER ...	\$129.8	673
8	BPA CALL	SUPPLEMEN...	\$8.6	140
9	BPA CALL	TERMINATE ...	\$0.1	64
10	DEFINITIVE ...	ADD ...		3
11	DEFINITIVE ...	ADDITIONAL...	\$123,973.4	338

Graph Builder – Visual Pivot Table

CONDITIONAL DATA FILTERING “ACTION TYPE” WITHIN “AWARD TYPE”

Local Data Filter

24898 matching rows

Inverse

[1] award_type (4)

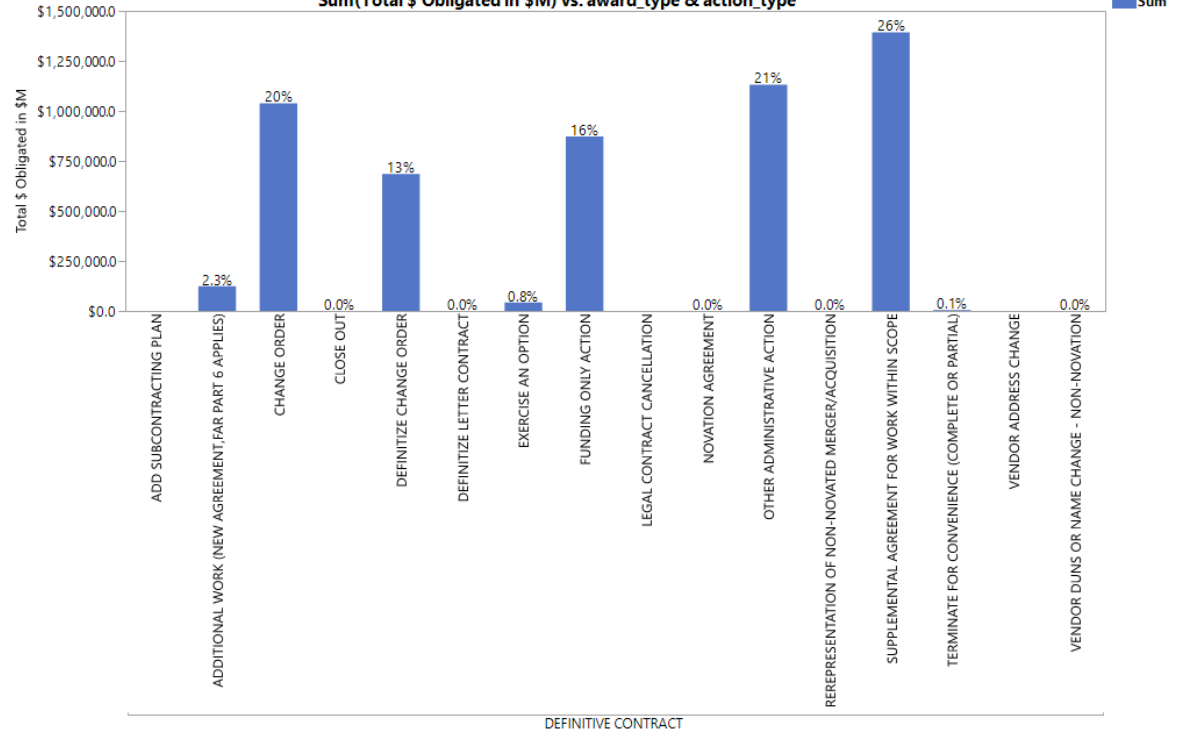
- BPA CALL (6652)
- DEFINITIVE CONTRACT (24898)
- DELIVERY ORDER (111867)
- PURCHASE ORDER (53115)

action_type (17)

- ??? (2561)
- ADD SUBCONTRACTING PLAN (3)
- ADDITIONAL WORK (NEW AGREEMENT,FAR PART 6 APPLIES) (338)
- CHANGE ORDER (1580)
- CLOSE OUT (183)
- DEFINITIZE CHANGE ORDER (419)
- DEFINITIZE LETTER CONTRACT (9)
- EXERCISE AN OPTION (2081)
- FUNDING ONLY ACTION (5602)
- LEGAL CONTRACT CANCELLATION (1)
- NOVATION AGREEMENT (5)
- OTHER ADMINISTRATIVE ACTION (6485)
- REREPRESENTATION OF NON-NOVATED MERGER/ACQUISITION (3)
- SUPPLEMENTAL AGREEMENT FOR WORK WITHIN SCOPE (5566)
- TERMINATE FOR CONVENIENCE (COMPLETE OR PARTIAL) (55)
- VENDOR ADDRESS CHANGE (1)
- VENDOR DUNS OR NAME CHANGE - NON-NOVATION (6)

Graph Builder

Sum(Total \$ Obligated in \$M) vs. award_type & action_type

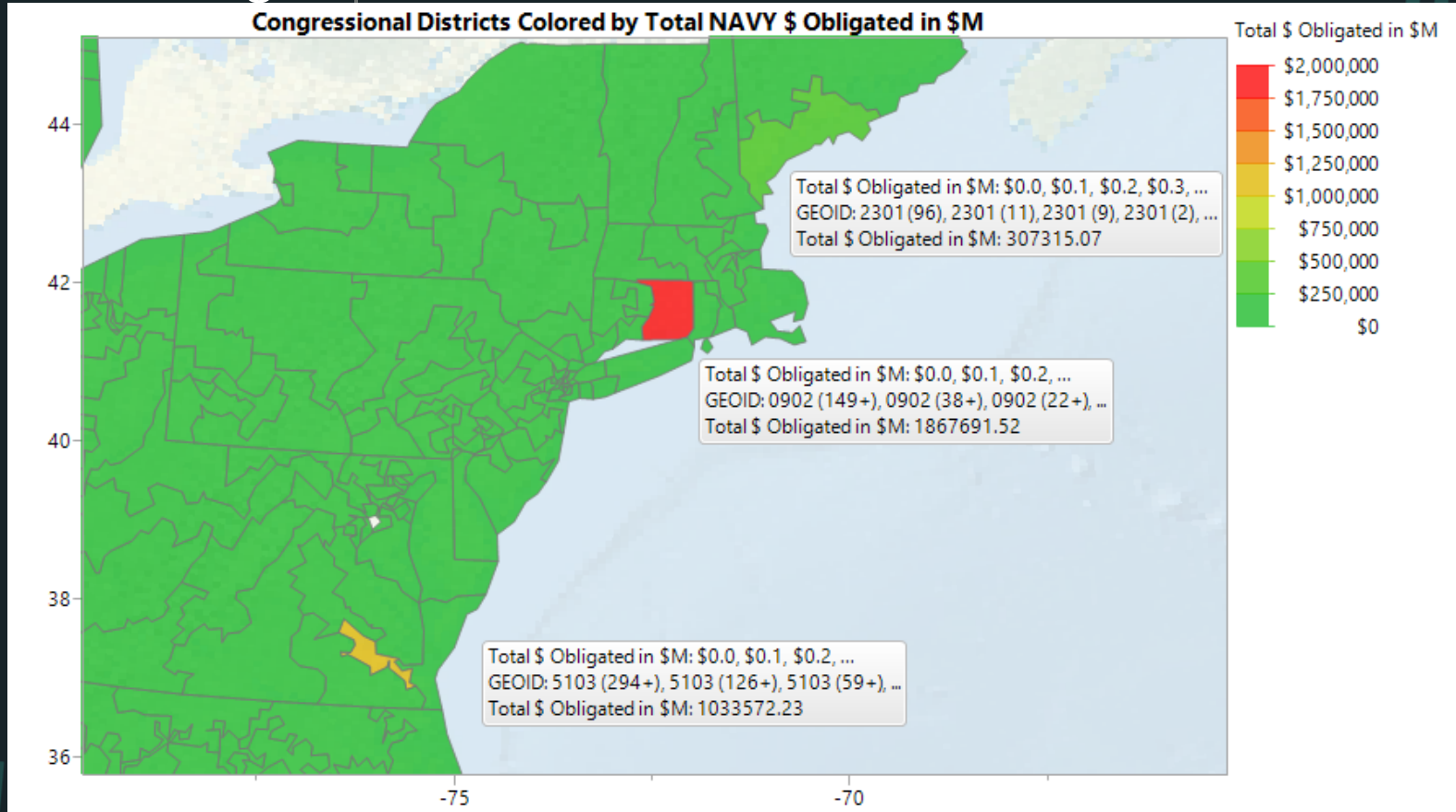


award_type / action_type ordered by total_dollars_obligated (ascending)

Where(award_type = DEFINITIVE CONTRACT)

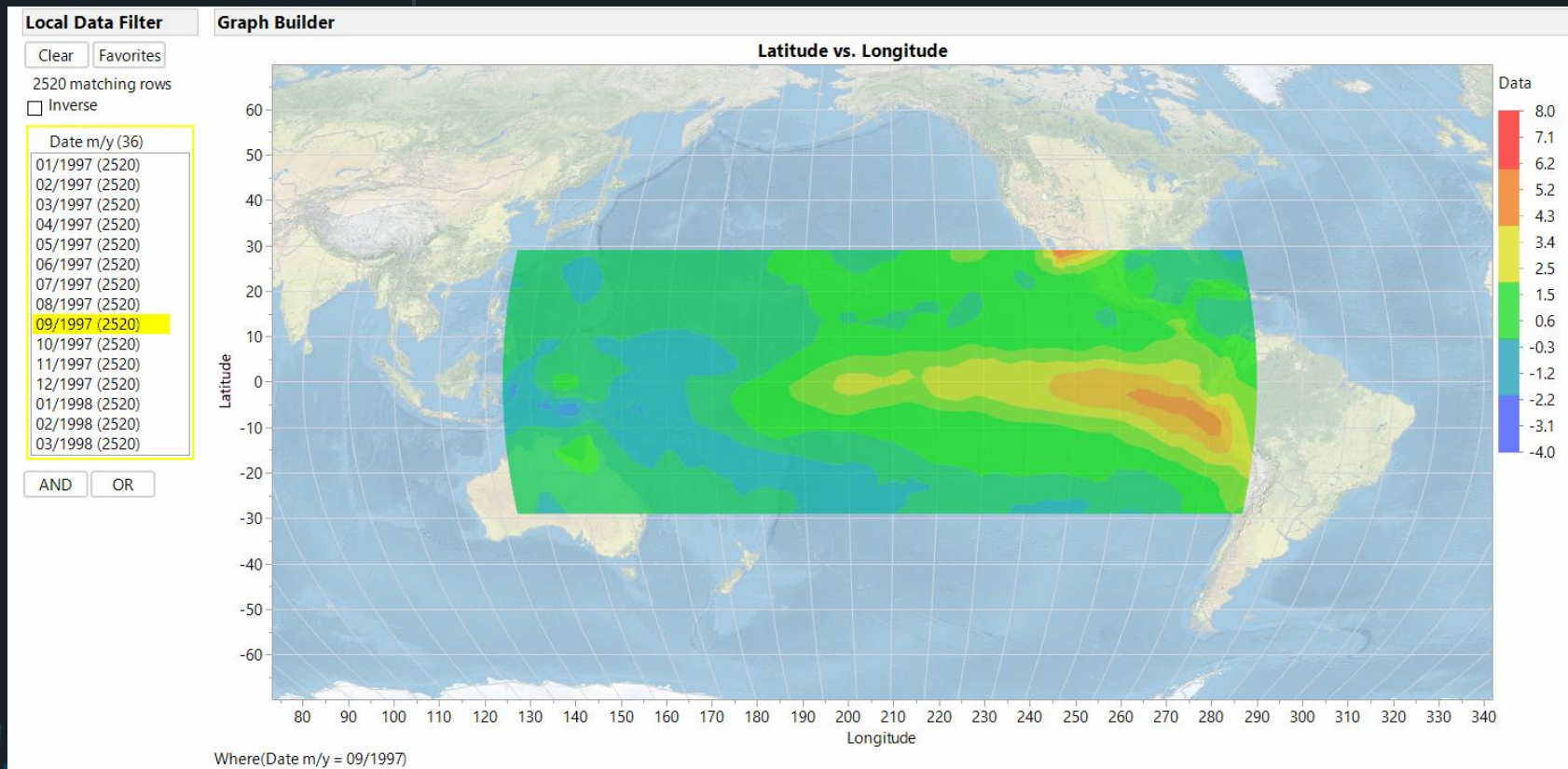
Maps and images

NAVY OBLIGATED SPENDING BY CONGRESSIONAL DISTRICT IN \$M



Maps and images

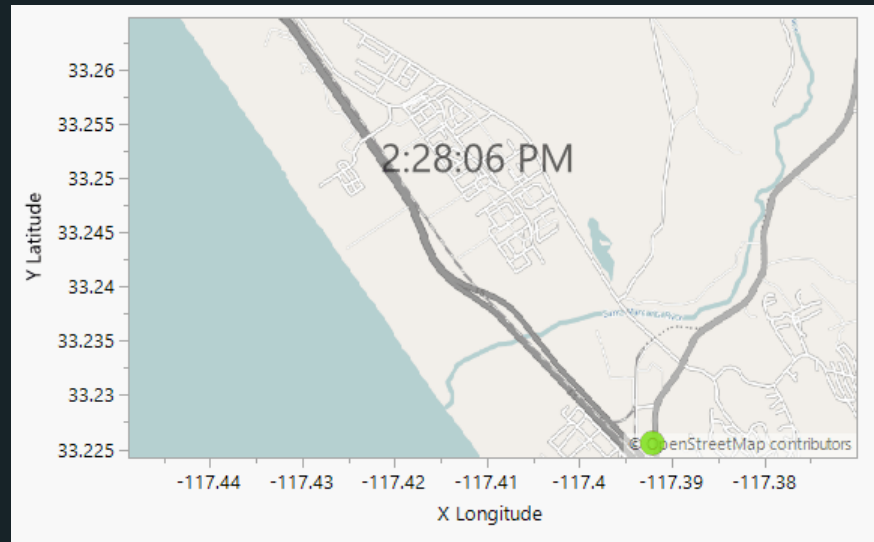
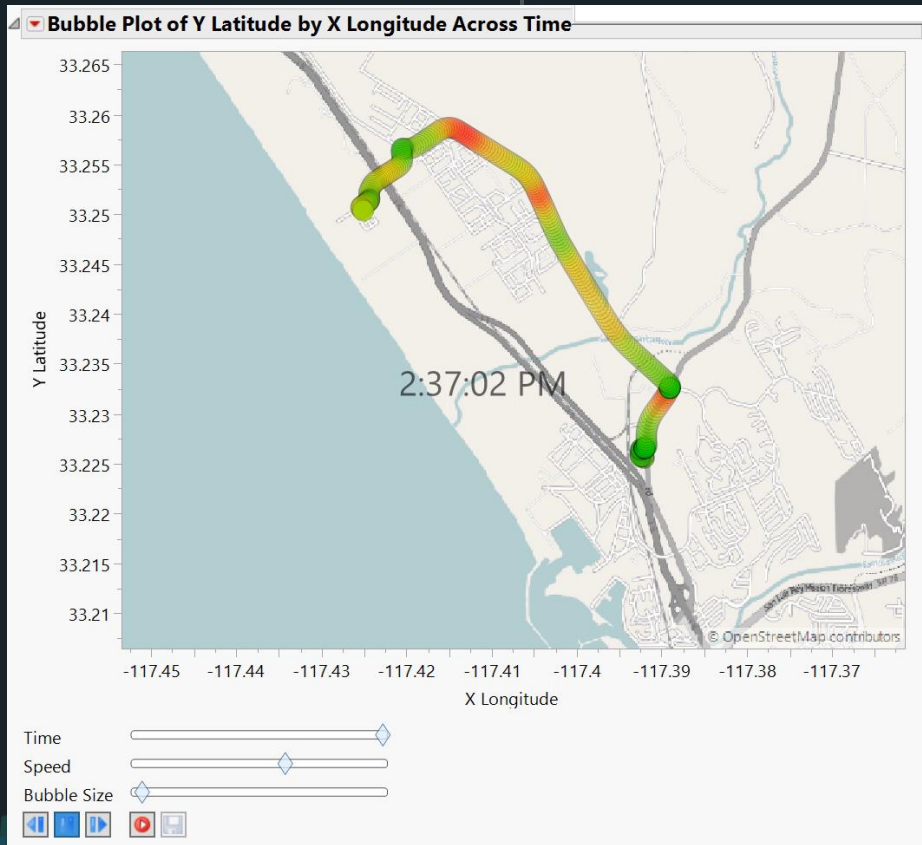
- MAPS ANIMATED OVER TIME AND SAVED AS A .GIF CAN BE DROPPED INTO PPT



DROP GIF HERE

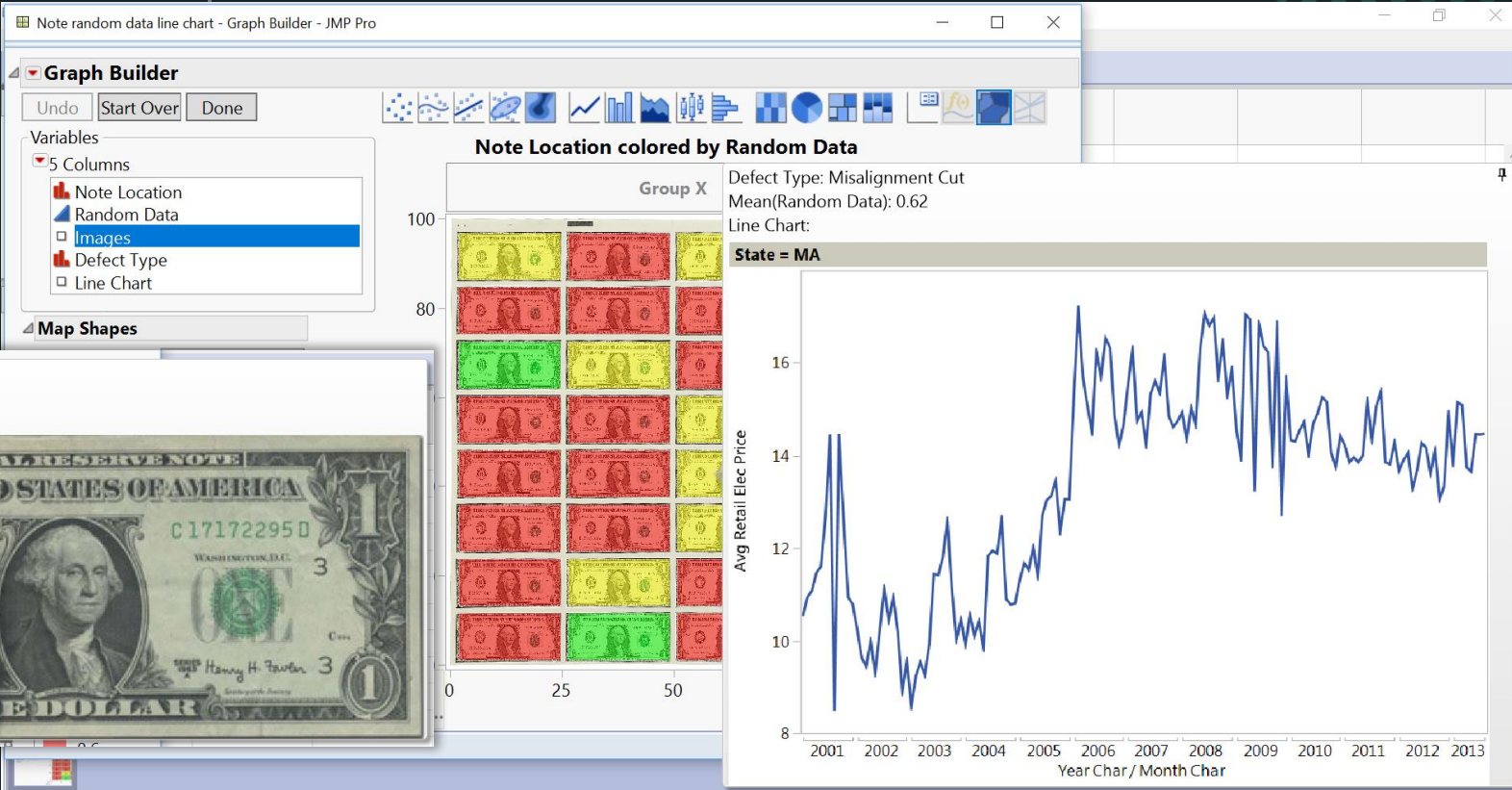
Maps and images

MAPS ANIMATED OVER TIME: SAVED AS PPT VIDEO (LEFT) ANIMATED GIF (RIGHT) OR HTML (USE QR CODE TO VIEW AT PUBLIC.JMP.COM ON YOUR PHONE)



Map of Printing Plate

HOVER TO PULL UP IMAGE OF DEFECT OR PLOT OF SENSOR OVER TIME



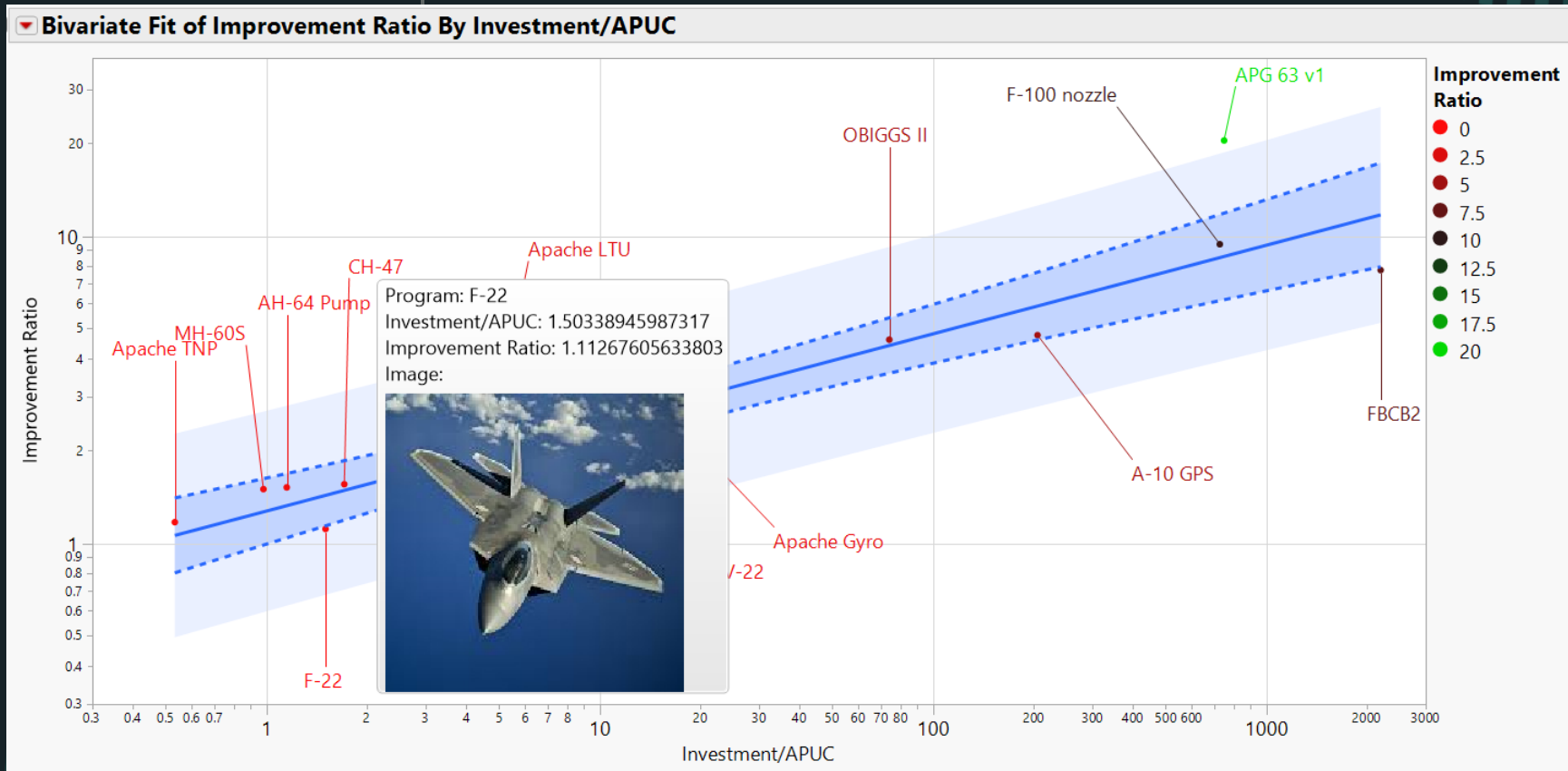
Defect Type: Ink Smear
Mean(Random Data): 0.60
Images:



evaluations done

Maps and images

IMAGES CAN BE IMPORTED IN BULK AND SHOWN BY HOVERING ON A DATA POINT



JMP as Data HUB

EASILY IMPORT DATA FROM EXCEL, TXT, CSV, PDF, SAS, DATA BASES, WEB, &
EASILY EXPORT MODELS (WITH 13 & 14 PRO) TO: SAS, PYTHON, JAVASCRIPT, C, SQL
EASILY PUBLISH REPORTS TO WEB (HTML); SAVE REPORTS TO POWERPOINT

Excel Import Wizard

Data Preview

	Revenue by Territory	Column 2	Column 3
1			
2	Country	Units (1000's)	Revenue ...
3	France	23	216.89
4	Germany	42	396.06
5	Italy	12	113.16
6	UK	32	301.76
7	China	15	141.45
8	Japan	34	320.62
9	Singapore	3	28.29
10	U.S.	104	980.72
11	Canada	12	113.16
12			
13	TOTAL	277	2612.11

Worksheets

Select sheets to open	Custom setting
JAN	
FEB	
MAR	
APR	
MAY	
JUN	
JUL	
AUG	
SEP	
OCT	
NOV	
DEC	

Select all

Individual Worksheet Settings

- Worksheet contains column headers
- 1 Column headers start on row
- 1 Number of rows with column headers
- 2 Data starts on row
- 1 Data starts on column

Preview Pane Refresh

- Update settings on any change
- Update now
- Show all rows

Concatenate worksheets and try to match columns

- Create column with worksheet name when concatenating

Use for all worksheets

Restore Default Settings Back Next Import Cancel Help

JMP Home Window - JMP Pro [2]

File Tables DOE Analyze Graph Tools Add

- New
- Open... Ctrl+O
- Quick Open Alt+Shift+O
- Close Ctrl+W
- Import as Data
- Import Multiple Files...
- Save Ctrl+S
- Save As...
- Revert
- Database
- SAS
- Internet Open...
- Publish...
- Create Excel Workbook...
- Preferences Ctrl+K
- Print... Ctrl+P
- Print Preview
- Page Setup...
- Send...
- Recent Files
- Save Session Script...
- Exit JMP Ctrl+Q

Report: Formula Depot - JMP Pro [3]

Formula Depot

Manage Models and Generate Score

Formula Scripts

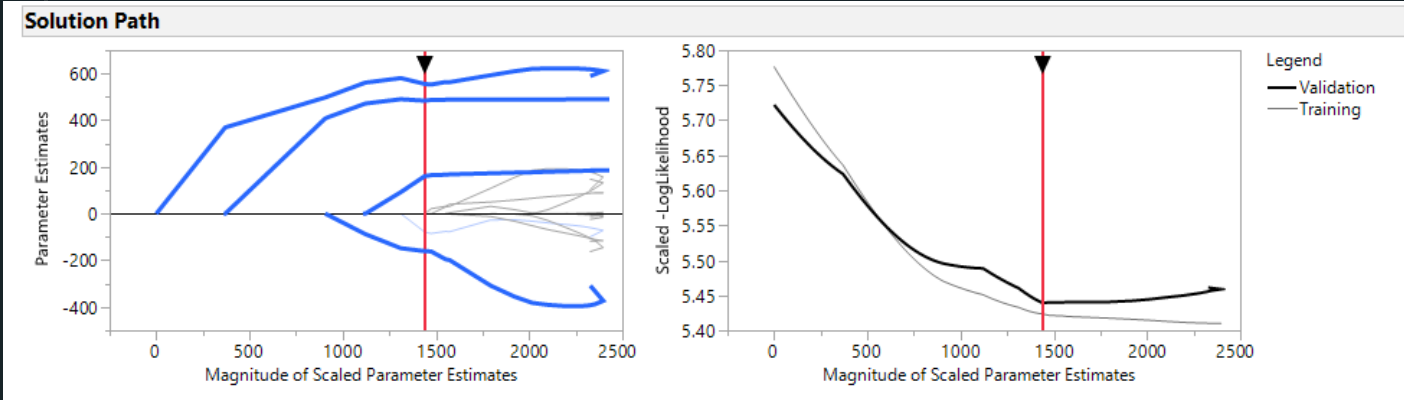
Neural - Y

- Show Script
- Copy Script
- Copy Formula as Function
- Copy Formula as Transform
- Rename New Column
- Generate C Code
- Generate Python Code
- Generate JavaScript Code
- Generate SAS Code
- Generate SQL Code
- Run Script
- Remove

Fast modern regression (JMP 12)

INTERACT WITH ADAPTIVE LASSO MODEL – VARIABLE & MODEL SELECTION FOR DATA THAT’S HIGHLY CORRELATED AND/OR NON-NORMAL

Measure	Training	Validation	Test
Number of rows	294	74	74
Sum of Frequencies	294	74	74
-LogLikelihood	1594.5348	402.5022	399.46519
Number of Parameters	7	7	7
BIC	3228.8547	835.13285	829.05883
AICc	3203.4613	820.70136	814.62735
RSquare	0.506121	0.4320668	0.5061904
RMSE	54.852588	55.703388	53.440629
Lambda Penalty	6030.1572	.	.



Parameter Estimates for Original Predictors

Term	Estimate	Std Error	Wald ChiSquare	Prob > ChiSquare	Lower 95%	Upper 95%
Intercept	-302.705	39.488631	58.761791	<.0001*	-380.1013	-225.3087
BMI	6.3716072	0.8995747	50.167619	<.0001*	4.6084731	8.1347413
BP	0.6863572	0.2729626	6.3225768	0.0119*	0.1513604	1.2213541
Total Cholesterol	-0.253761	0.1183203	4.5997185	0.0320*	-0.485664	-0.021857
HDL	-0.355974	0.28592	1.5500595	0.2131	-0.916367	0.2044185
LTG	61.913643	8.8560351	48.87582	<.0001*	44.556133	79.271153

(BP-94.647)*(Total Cholesterol-189.14)*(LDL-189.14)*(HDL-189.14)*(TCH-189.14)*(HDL-49.7885)*(TCH-4.07025)*(Total Cholesterol-189.14)*(LTG-189.14)

Prediction Profiler

The Prediction Profiler shows four plots for BMI, BP, Total Cholesterol, and LTG. Each plot displays a regression line and a shaded confidence interval. The y-axis is labeled 'Y' with values 0, 100, 200, 300. The x-axis values are 20, 25, 30, 35, 40, 60, 80, 100, 120, 140, 160, 180, 200, 250, 300, 3.0, 3.5, 4.0, 4.5, 5.5, 6.0, 6.5. A vertical red line is at 189.14. A horizontal red line is at 150.4484. The y-axis also has values [142.253, 158.644].

Near Machine Learning accuracy, AND provides a more interpretable model with confidence intervals

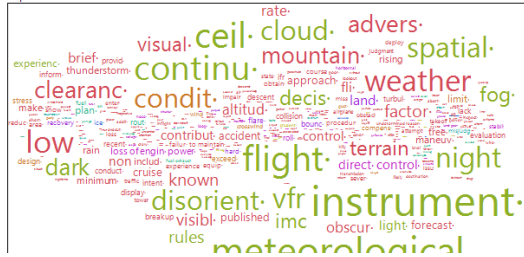


Exploratory Text Analysis (JMP 13)

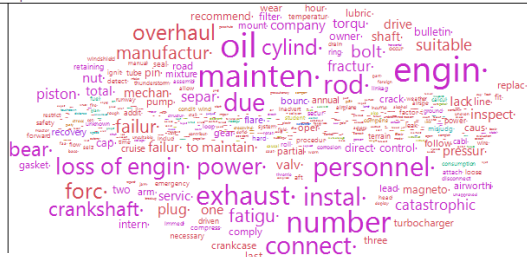
DIMENSION REDUCTION OF SPARSE DOCUMENT TERM MATRIX INTO DOCUMENT AND TERM VECTORS – ALSO CLUSTERING OF DOCUMENTS AND TOPICS

Word Clouds by Topic

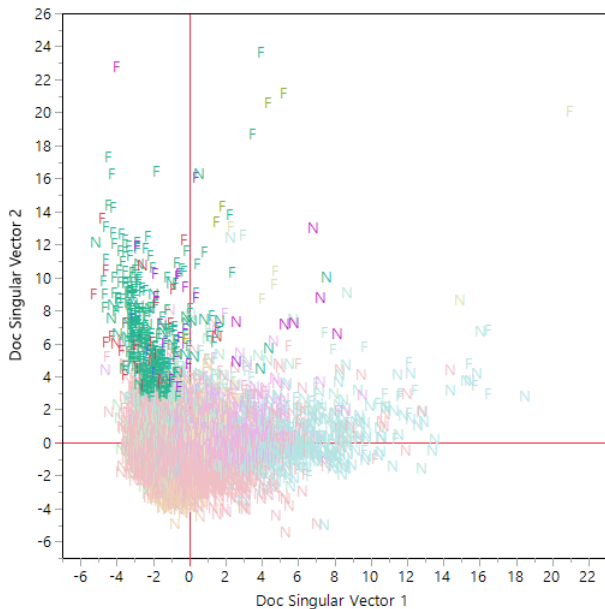
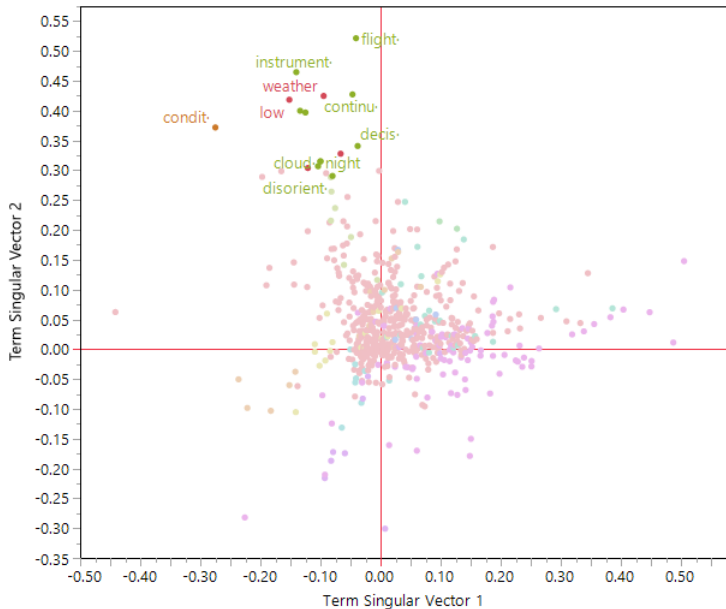
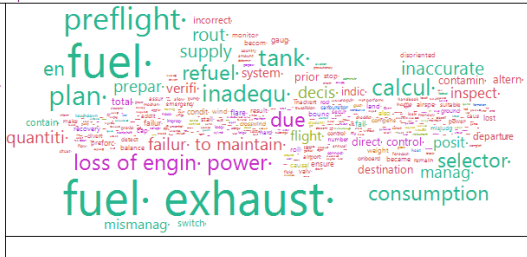
Topic 1



Topic 2



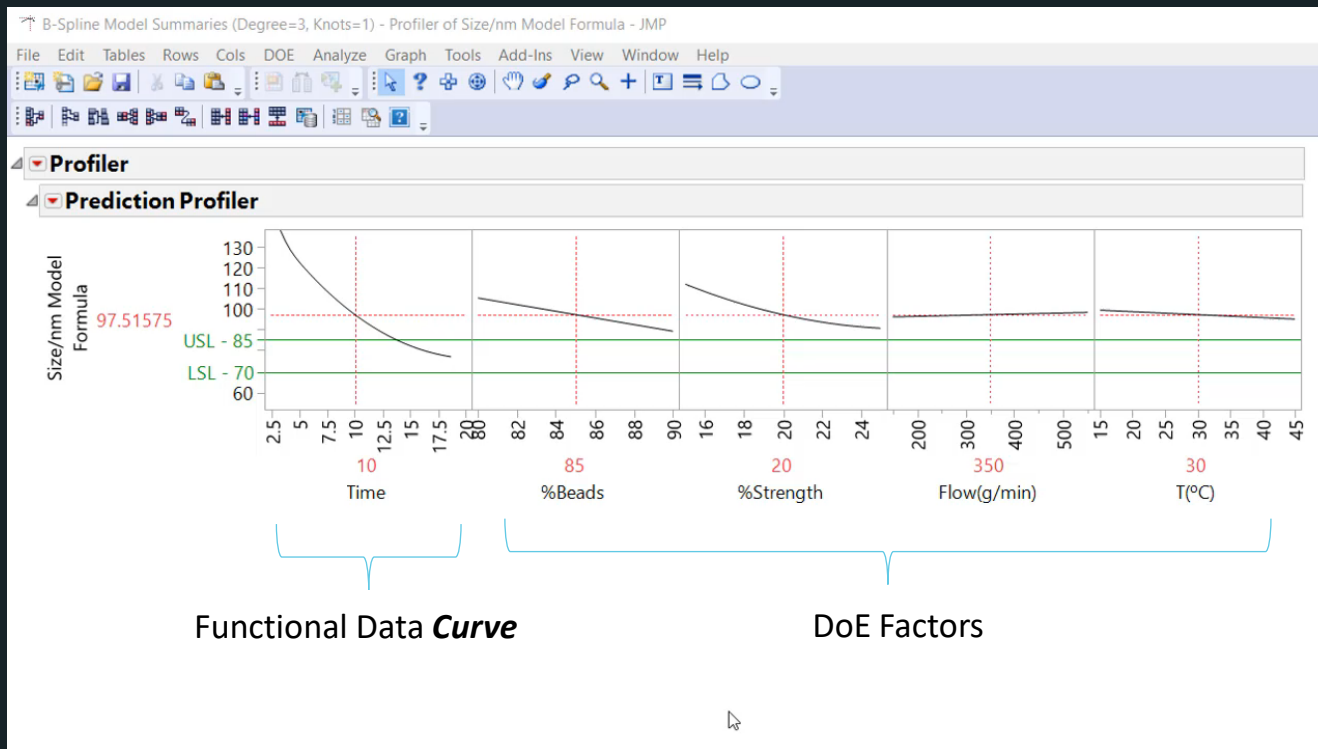
Topic 3




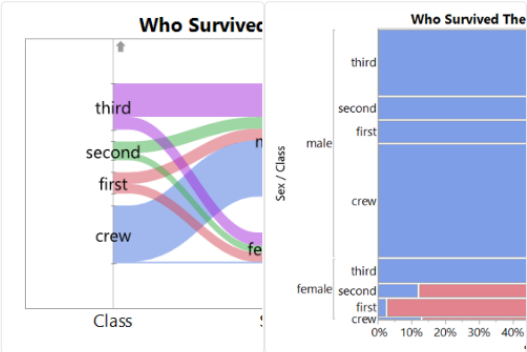
Topic 1

Functional Data Analysis (JMP 14)

MODELING THE "SHAPE" OF A STREAM OF DATA – SHAPE IS THE FUNDAMENTAL UNIT OF OBSERVATION – DIMENSION REDUCTION WITH FUNCTIONAL PCA ABLE TO CONTROL AND PREDICT SHAPE AS FUNCTION OF DOE FACTORS




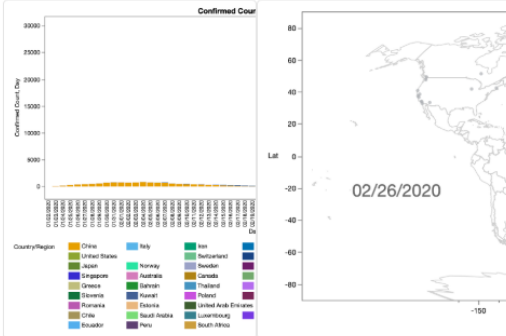
 Who Survived the Titanic?
Phil Kay



Interactive plots visualise Titanic survivor data.


Views: 87 Wed Mar 06, 2019 4:49 am

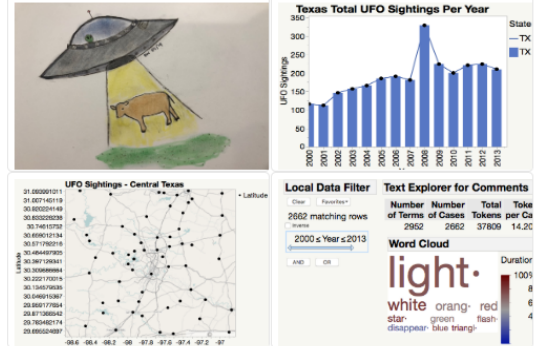
 Coronavirus COVID-19 Global Cases
JMP_Byron



Time series data represented in multiple graph formats

👤 1379 🗨️ 0 Sat Mar 21, 2020 2:54 pm

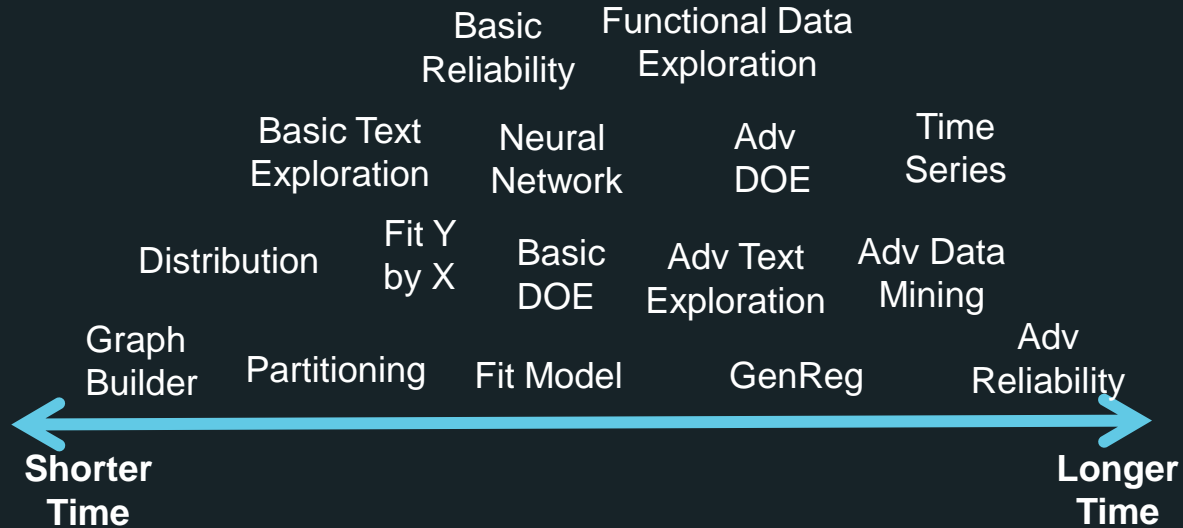
 Texas UFO Sightings
Scott Wise



I visualized the number and location of UFO sightings in Texas from 2000 to 2014. I also explored the text of comments in sighting reports.

Views: 65 Mon Mar 04, 2019 9:57 pm

How long does it take to learn JMP? Depends... on your level of statistical know how



Summary: JMP capabilities that can help you

Move from data to decision – *Faster!*

- Exploratory Data Analysis with Dynamic Linking & Filtering
- Design of Experiments – Real & Simulated Data
- Reliability Analysis, Accelerated Life Test, Reliability Growth
- Surrogate modeling of simulations
- Visual Multivariate Modeling
- Robust Data Mining using Honest Assessment stopping criteria when using Decision Trees, Neural Networks, Regression, Non-Linear modeling
- Maps and Images – animation over time
- Data Curation – recoding, impute missing, find outliers, import, export
- Fast and Modern Penalized Regression – interact with model (12 Pro)
- Exploratory Text Analysis – dimension reduction – Topic Vectors (13 Pro)
- Functional Data Analysis – modeling the “shape” of a data stream (14 Pro)

Resources

My Recorded Tutorials & Slide Decks at www.jmp.com/fedgov

These 9 videos cover predictive analytics (including text exploration), data visualization, and "What's New in JMP 14?" topics.

Building Better Models Overview and Use of Honest Assessment	Neural Networks - Single Layer, Dual Layer, Boosted	All Graphs are Wrong - Some are Useful - Or view Xan Gregg's Original 2015 Discovery Summit Presentation
Regression Linear, Stepwise, Logistic, & All Possible	Generalized Regression Near Machine Learning Accuracy – More Explainable Model	What's New in JMP 14? JMP Learning Resources
Decision Trees Simple Partition, Bootstrap Forest, & Boosted Tree	Text Exploration Analyze Unstructured Free Text	Functional Data Explorer Modeling a "Stream" of Data – New in JMP 14



Questions or comments?

Thank you.



[jmp.com](https://www.jmp.com)

JMP Defense & Aerospace Team

Anna-Christina De La Iglesia

Program Manager

anna-christina.delaiglesia@jmp.com

919-531-2593

Procurement, Upgrades, License Renewals...

Sam Tobin

Senior Account Representative

sam.tobin@jmp.com

919-531-0640

Technical Questions, Getting Started, Tutorials, Mentoring...

Tom Donnelly, PhD, CAP

Principal Systems Engineer

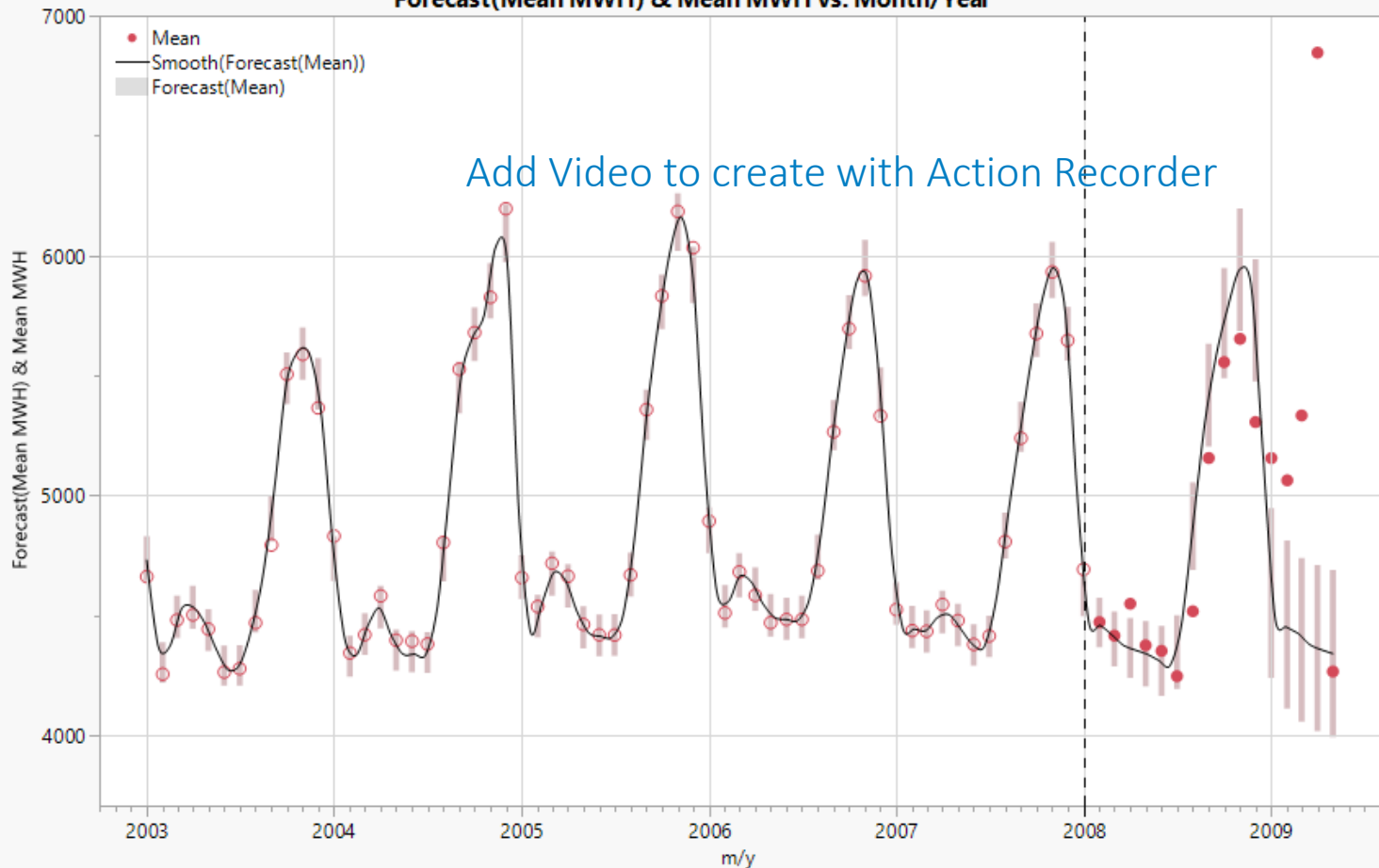
tom.donnelly@jmp.com

302-489-9291

ADDITIONAL TOPICS

- Import WGI data Set
- Time series for Army bases - Action Recorder MFG History
- DOE for Paint coating of shell
- Fuse testing simulation via model
- SFO Crime Dashboard
- Action Recorder MFG History

Forecast(Mean MWH) & Mean MWH vs. Month/Year



Error band is constructed from Lower Forecast(Mean) to Upper Forecast(Mean).