Visual Comparison of **Complex Sets of Quantifiable Forensic** Data by David K. Sorensen, P.E. ISFSEM - July 23, 2015



David Sorensen, P.E., BSME, MSMET, RPE in CA, ID & UT, member SME, IEEE & ASME (IEEE Com. Chair), Patents w GE & Bunker Ramo, Dir. Engr at INL w staff of 600+, Started EG&G Services, grew to \$5 million in 2 yrs w 90% ROI & staff of 125, Bid against Lockheed, Rockwell & Boeing for the NASA Ops Contratc & won (10 yr \$1 billion). Exec VP at Eyring, Founder of RPR Ctr at BYU, Exec Dir of NISt MEP in UT for 15 yrs. Creating over 32,000 jobs in Utah alone.

Data Retrieval is what Google is all about. ... but why do people want to access data

> Data Analysis and Decision Making is what HSG is all about.

What process and approach do people really use?

Decision Making

 A decision involves making a choice between several options.

Each option has a set of attributes.

- Attributes are generally <u>quantifiable</u>.
- <u>Variation</u> in the attributes distinguishes each option from all other options.

The correlation between attributes is an important part of overall decisions.

 Several <u>options</u>, each with quantifiable <u>attributes</u>, are compiled into a spreadsheet.

Spreadsheet Data (Cost of Living Attributes, various cities)

	Groceries	Housing	Utilities	Transportation	Health
US Avg	100	100	100	100	100
Sacramento	104.8	120.3	111.5	111.9	122.8
San Diego	106.4	200.8	78.1	127	125.2
Denver	95.5	117.9	89.9	101.4	113.8
Boise	89.8	101.9	73.7	99.3	87.2
Missoula	103.7	86.4	97.7	100.7	110
Reno	99.1	114.1	97.8	99.5	109.5
Albuquerque	98.1	108.8	99	103	105.1
Seattle	111.2	115.9	63	117	139.7
SLC	94.8	86.9	89.8	101.1	100

Columns and Rows

1 Variable

	Groceries	Housing	Utilities	Transportation	Health
US Avg	100	100	100	100	100
Sacramento	104.8	120.3	111.5	111.9	122.8
San Diego	106.4	200.8	78.1	127	125.2
Denver	95.5	117.9	89.9	101.4	113.8
Boise	89.8	101.9	73.7	99.3	87.2
Missoula	103.7	86.4	97.7	100.7	110
Reno	99.1	114.1	97.8	99.5	109.5
Albuquerque	98.1	108.8	99	103	105.1
Seattle	111.2	115.9	63	117	139.7
SLC	94.8	86.9	89.8	101.1	100

Or





Pie Chart





_

Housing



Utilities



Transportation



Health Care



2 Variables

	Groceries	Housing	Utilities	Transportation	Health
US Avg	100	100	100	100	100
Sacramento	104.8	120.3	111.5	111.9	122.8
San Diego	106.4	200.8	78.1	127	125.2
Denver	95.5	117.9	89.9	101.4	113.8
Boise	89.8	101.9	73.7	99.3	87.2
Missoula	103.7	86.4	97.7	100.7	110
Reno	99.1	114.1	97.8	99.5	109.5
Albuquerque	98.1	108.8	99	103	105.1
Seattle	111.2	115.9	63	117	139.7
SLC	94.8	86.9	89.8	101.1	100



1 - Groceries vs Housing



2 - Groceries vs Utilities



3 - Groceries vs Transportation



4 - Groceries vs Health Care



5 - Housing vs Utilities



6 - Housing vs Transportation



7 - Housing vs Health Care



8 - Utilities vs Transportation



9 - Utilities vs Health Care



10 - Transportation vs Health Care



5 Variables, 10 Cities...

	Groceries	Housing	Utilities	Transportation	Health
US Avg	100	100	100	100	100
Sacramento	104.8	120.3	111.5	111.9	122.8
San Diego	106.4	200.8	78.1	127	125.2
Denver	95.5	117.9	89.9	101.4	113.8
Boise	89.8	101.9	73.7	99.3	87.2
Missoula	103.7	86.4	97.7	100.7	110
Reno	99.1	114.1	97.8	99.5	109.5
Albuquerque	98.1	108.8	99	103	105.1
Seattle	111.2	115.9	63	117	139.7
SLC	94.8	86.9	89.8	101.1	100

....Results in 5 Bar Charts...







	Groceries	Housing	Utilities	Transportation	Health
US Avg	100	100	100	100	100
Sacramento	104.8	120.3	111.5	111.9	122.8
San Diego	106.4	200.8	78.1	127	125.2
Denver	95.5	117.9	89.9	101.4	113.8
Boise	89.8	101.9	73.7	99.3	87.2
Missoula	103.7	86.4	97.7	100.7	110
Reno	99.1	114.1	97.8	99.5	109.5
Albuquerque	98.1	108.8	99	103	105.1
Seattle	111.2	115.9	63	117	139.7
SLC	94.8	86.9	89.8	101.1	100





...and 10 X-Y Scatter Plots







	Groceries	Housing	Utilities	Transportation	Health
US Avg	100	100	100	100	100
Sacramento	104.8	120.3	111.5	111.9	122.8
San Diego	106.4	200.8	78.1	127	125.2
Denver	95.5	117.9	89.9	101.4	113.8
Boise	89.8	101.9	73.7	99.3	87.2
Missoula	103.7	86.4	97.7	100.7	110
Reno	99.1	114.1	97.8	99.5	109.5
Albuquerque	98.1	108.8	99	103	105.1
Seattle	111.2	115.9	63	117	139.7
SLC	94.8	86.9	89.8	101.1	100









Handling More Complex Data

The previous data was from some actual U. S. cities; 10 different cities, compared against 5 decision attributes.

Using traditional graphics, a total of 5 bar charts are required to present comparative information, and 10 more X-Y scatter plots are also required.

<u>#Variables</u>	<u>#Bar Charts</u>	#X-Y Scatt	<u>er Plots</u>	<u>#</u>
Paper Cha	<u>rts</u> 1	0	1	
2	2	1	3	
3	3	3	6	
4	4	6	10	
	5	10	15	
6	<u>6</u>	15	21	
7	7	21	28	
8	8	28	36	
9	9	36	45	
10	10	45	55	

Many Graphics



Many Graphics



One MultiGraf



One MultiGraf



The Problem

Each decision maker must mentally merge a large number of graphics without correlation data to make a decision, which introduces significant error.

TODAY, decision making is where data retrieval was BEFORE.

Today we're choking on the data.



Traditional Problem (Separate Graphics)

Subjective
Inaccurate
Tedious

Traditional Problem (Separate Graphics) HSG Solution (Combined Graphics)

Subjective
Inaccurate
Tedious

Objective
Accurate
Immediate
Summary of HSG's Single Graphic Data Display Features

- Multiple data points (objects) are shown together.
- Each object is represented by only one data point.
- Multiple or all axes (attributes) are shown.
- All data points are referenced to each axis.
- The axes are positioned in the center of the graphic.
- The average for each attribute is graphically shown.

Summary of Features (CONT.)

 The comparative relationship between each attribute and all attributes is graphically displayed.

 The relative correlation between each attribute and all other attributes is accurately maintained.

Standard 2D x-y scatter plot



Copyright © 2008 ECHO Solutions, Inc.



Copyright © 2008 ECHO Solutions, Inc.

















Tabular Data

	Groceries	Housing	Utilities	Transportation	Health
US Avg	100	100	100	100	100
Sacramento	104.8	120.3	111.5	111.9	122.8
San Diego	106.4	200.8	78.1	127	125.2
Denver	95.5	117.9	89.9	101.4	113.8
Boise	89.8	101.9	73.7	99.3	87.2
Missoula	103.7	86.4	97.7	100.7	110
Reno	99.1	114.1	97.8	99.5	109.5
Albuquerque	98.1	108.8	99	103	105.1
Seattle	111.2	115.9	63	117	139.7
SLC	94.8	86.9	89.8	101.1	100

Columns and Rows

HSG's MultiGraf for 10 cities data

- The labeled end of each attribute has the highest numeric value
- The angle between two attributes represents their correlation value
- Objects are placed in 3D relative to all attributes
- Native view (as shown) is "flatest" possible view for all variables displayed

MultiGraf









Unique HSG Features

- The entire data set is displayed objectively.
- No conclusions are required before graphically representing the data.
- The entire data set is analyzed without segmentation.
- Attribute correlations are graphically displayed.
- Relevant data elements are easier to subsequently identify, eliminate and/or explore.
- Numerous traditional graphs are combined into a single MultiGraf.
- Data input is a simple spreadsheet format.

Handling More Complex Data

Following are some actual Consumer Reports' data on 3 different data sets of batteries, compared against 3 decision attributes.

Using traditional bar charts, a total of 9 bar charts are used to present comparative information, and 36 X-Y scatter plot could also be generated.

Consumer Reports' Heavy Duty Battery Test Results

Heavy Duty	Тоу	Heavy Duty	Stereo	Heavy Duty	
Battery	Test	Battery	Test	Battery	Price
Eveready - H	1.2	Eveready - H	4.5	Eveready - H	\$0.57
Mallory - H	1.2	Mallory - H	4.8	Mallory - H	\$0.49
Radio Shack - H	1.0	Radio Shack - H	5.5	Radio Shack - H	\$0.37
Rayovac - H	1.2	Rayovac - H	4.6	Rayovac - H	\$0.56
Sears - H	1.2	Sears - H	4.2	Sears - H	\$0.44

Consumer Reports' Alkaline Battery Test Results

Alkaline	Тоу
Battery	Test
Duracell - A	6.2
Eveready - A	5.3
Kodak - A	5.5
Panasonic - A	5.7
Radio Shack - A	5.1
Rayovac - A	4.5
Sears - A	5.7

Alkaline	Stereo		
Battery	Test		
Duracell - A	14.2		
Eveready - A	14.1		
Kodak - A	13.5		
Panasonic - A	14.1		
Radio Shack - A	13.5		
Rayovac - A	11.2		
Sears - A	14.0		

Alkaline	
Battery	Price
Duracell - A	\$1.08
Eveready - A	\$0.92
Kodak - A	\$0.99
Panasonic - A	\$0.88
Radio Shack - A	\$0.72
Rayovac - A	\$0.98
Sears - A	\$0.70

Consumer Reports' Rechargeable Battery Test Results

Rechargeable	Тоу	Rechargeable	Stereo	Rechargeable	
Battery	Test	Battery	Test	Battery	Price
Eveready - R	1.6	Eveready - R	3.5	Eveready - R	\$3.95
GE - R	2.2	GE - R	5.5	GE - R	\$3.85
Millenium - R	1.7	Millenium - R	4.7	Millenium - R	\$2.99
Radio Shack - R	2.2	Radio Shack - R	3.9	Radio Shack - R	\$2.35

Toy Test







Heavy Duty Price



Toy Test 7.0 6.0 5.0 4.0 ■ Toy Test 3.0 2.0 1.0 0.0

Stereo Test



Alkaline Price





Stereo Test





Heavy Duty	Тоу	Stereo	
Battery	Test	Test	Price
Eveready - H	1.2	4.5	\$0.57
Mallory - H	1.2	4.8	\$0.49
Radio Shack - H	1.0	5.5	\$0.37
Rayovac - H	1.2	4.6	\$0.56
Sears - H	1.2	4.2	\$0.44
Alkaline	Тоу	Stereo	
Battery	Test	Test	Price
Duracell - A	6.2	14.2	\$1.08
Eveready - A	5.3	14.1	\$0.92
Kodak - A	5.5	13.5	\$0.99
Panasonic - A	5.7	14.1	\$0.88
Radio Shack - A	5.1	13.5	\$0.72
Rayovac - A	4.5	11.2	\$0.98
Sears - A	5.7	14.0	\$0.70
Rechargeable	Тоу	Stereo	
Battery	Test	Test	Price
Eveready - R	1.6	3.5	\$3.95
GE - R	2.2	5.5	\$3.85
Millenium - R	1.7	4.7	\$2.99
Radio Shack - R	2.2	3.9	\$2.35

All	Тоу	Stereo		
Batteries	Test	Test	Price	
Eveready - H	1.2	4.5	\$0.57	
Mallory - H	1.2	4.8	\$0.49	
Radio Shack - H	1.0	5.5	\$0.37	
Rayovac - H	1.2	4.6	\$0.56	
Sears - H	1.2	4.2	\$0.44	
Duracell - A	6.2	14.2	\$1.08	
Eveready - A	5.3	14.1	\$0.92	
Kodak - A	5.5	13.5	\$0.99	
Panasonic - A	5.7	14.1	\$0.88	
Radio Shack - A	5.1	13.5	\$0.72	
Rayovac - A	4.5	11.2	\$0.98	
Sears - A	5.7	14.0	\$0.70	
Eveready - R	1.6	3.5	\$3.95	
GE - R	2.2	5.5	\$3.85	
Millenium - R	1.7	4.7	\$2.99	
Radio Shack - R	2.2	3.9	\$2.35	

Example #1: MultiGraf of Comparativ e Battery Data



Some Conclusions From Comparative Battery Data MultiGraf Stereo and toy tests are highly correlated Price is not highly correlated with performance Heavy Duty batteries are a commodity Duracell & Rayovac don't compare in performance Sears alkaline may be the best buy Name brands demand a higher price in new items Buy Duracell, maybe Sears and never Rayovac

	Amount	Мо	APR	Debt	Income	Expenses	Job	Value	Ttl Credit	Neg Credit	Late
Obj1	\$21,356	60	11.5%	0.245	\$4,435	\$1,086	1	\$21,765	11	0	0
Obj2	\$16,491	60	10.5%	0.217	\$3,000	\$650	13	\$16,800	17	0	0
Obj3	\$9,212	48	9.5%	0.273	\$1,300	\$355	8	\$9,434	6	0	0
Obj4	\$16,491	60	10.5%	0.217	\$3,000	\$650	13	\$16,800	17	0	0
Obj5	\$9,212	48	9.5%	0.273	\$1,300	\$355	8	\$9,434	6	0	0
Obj6	\$8,447	60	14.4%	0.286	\$1,000	\$286	3	\$7,495	13	2	5
Obj7	\$15,359	48	13.2%	0.041	\$1,850	\$76	23	\$16,200	6	0	0
Obj8	\$9,415	48	14.5%	1.765	\$4,250	\$7,500	12	\$11,020	17	0	17
Obj9	\$4,531	30	12.5%	0.017	\$1,800	\$30	2	\$7,347	4	0	0
Obj10	\$10,850	48	12.8%	0.428	\$1,600	\$684	12	\$9,150	8	0	0
Obj11	\$10,748	48	9.9%	0.149	\$2,200	\$327	4	\$12,703	5	0	0
Obj12	\$9,766	48	12.5%	0.000	\$1,600	\$0	6	\$11,000	13	0	0
Obj13	\$5,000	42	12.0%	0.334	\$1,790	\$598	10	\$5,500	13	0	0
Obj14	\$2,502	13	14.0%	0.237	\$1,860	\$441	10	\$2,800	13	0	0
Obj15	\$2,389	25	16.2%	0.258	\$1,850	\$477	2	\$2,500	8	0	2
Obj16	\$10,642	60	13.7%	0.196	\$3,300	\$646	10	\$12,611	12	4	0
Obj17	\$5,504	42	13.8%	0.411	\$4,800	\$1,972	11	\$6,500	11	3	0
Obj18	\$11,806	60	11.0%	0.000	\$5,000	\$0	0	\$15,530	10	0	0
Obj19	\$13,878	60	14.3%	0.311	\$2,694	\$837	4	\$14,531	14	0	0
Obj20	\$3,939	24	18.0%	0.650	\$2,000	\$1,300	0	\$3,909	31	7	14
Obj21	\$12,173	60	14.7%	0.286	\$1,750	\$500	7	\$12,350	25	0	10
Obj22	\$5,651	42	13.5%	0.000	\$800	\$0	2	\$4,895	18	0	5
Obj23	\$3,516	24	14.8%	0.117	\$3,000	\$350	11	\$4,495	0	0	15
Obj24	\$10,203	60	13.9%	0.083	\$600	\$50	1	\$11,649	5	0	0
Obj25	\$11,960	60	11.9%	0.189	\$1,322	\$250	2	\$9,800	2	0	2
Obj26	\$9,533	60	18.5%	0.171	\$1,200	\$205	4	\$7,930	3		9
Obj27	\$12,202	60	13.0%	0.122	\$740	\$90	10	\$14,200	1	0	0
Obj28	\$17,703	61	12.8%	0.050	\$2,000	\$100	0	\$15,892	3	0	1
Obj29	\$2,899	18	19.8%	0.119	\$2,100	\$250	1	\$4,064	1	1	0
Obj30	\$10,891	48	15.0%	0.088	\$1,700	\$150	2	\$9,995	3	1	11
Example #2: MultiGraf of Car Loan Data (all attributes)



Example #2: MultiGraf of Car Loan Data

(excludes value, amount, duration, late and APR)



Example #2: MultiGraf of Car Loan Data

(includes only Job, Income, Debt & Late)



Some Conclusions From Car Loan Data MultiGraf

- Attributes range from high positive to high negative correlation, with 3 attributes highly correlated
- Defaulted loans appear in wide range of situations
- Most attributes do not appear to discriminate well on loan repayment predictability
- Eliminating selected attributes increases clarity of loan repayment performance, predictability

Conclusions (cont.)

- Job duration and low late payment experience are key
- Income level is not a significant factor for making loan
- Higher risk loans can be mitigated by requiring higher down payment to lower risk of late payment, etc.



		Heavy	Light		Manage	
	Prep Meals	Housewk	Housewrk	Using Phone	Money	Shopping
Men (65-69)	2.6	9.8	3.5	3	2.6	4.1
Men (70-74)	3.6	13	3.4	4.3	3.1	5.3
Men (75-79)	5.1	14.6	5.2	6.3	5.2	7.6
Men (80-84)	7.8	18.9	8.2	11.3	5.8	13.9
Men (85+)	18.5	33.3	15.2	18.4	19	26.8
Women (65-69)	4.2	21.8	4	1.3	1.4	6.4
Women (70-74)	5.5	27.3	6.2	2.8	2.3	9.4
Women (75-79)	8.3	33.2	8.4	4.1	5.2	14.5
Women (80-84)	14	41.7	14	6	9.3	24.7
Women (85+)	29.5	54.2	27.4	12.1	26.2	41.6



The NEW ENGLAND JOURNAL of MEDICINE

Copyright © 2008 ECHO Solutions,

Example #3: MultiGraf of Aging vs. Capabilities Data



Some Conclusions From Aging vs. Capabilities Data MultiGraf

- All of the attributes are positively correlated
 Three attributes are the same (data distortion)
 Attributes progress from Physical to Social tasks
- Women have more difficulty with physical tasks
- Men have more difficulty with social tasks

Secretary Data

	Type Speed	Accuracy	Hourly \$	WP Emp	Tot Emp	Attend	Ena Prof
Ann	60	92	9.90	4	28	98	85
Beth	65	88	10.50	6	24	97	88
Chris	50	94	10.84	9	30	96	90
Don	45	85	7.90	2	14	99	78
Ellen	30	90	8.24	4	16	98	88
Fay	40	98	10.62	7	60	95	93
Gail	78	95	11.90	12	48	97	94
Holly	70	85	9.70	6	12	88	81
Ida	72	91	10.30	22	72	96	97
Jackie	80	97	13.44	30	84	95	98

Example #4: MultiGraf of Employment Applicants' Performance Data



Some Conclusions From Applicants vs. Performance Data MultiGraf

- All applicants can be evaluated simultaneously relative to all other applicants and all attributes.
- The top candidates can be easily identified.
- Interviewing process is significantly simplified.
- Applicants with limited performance in one or more areas are readily identified.
- Improvement opportunities for training are also readily identified.

Unique HSG Features

The entire data set is displayed objectively.

- No conclusions are required before graphically representing the data.
- The entire data set is analyzed without segmentation.
- Attribute correlations are graphically displayed.
- Relevant data elements are easier to subsequently identify, eliminate and/or explore.
- Numerous traditional graphs are combined into a single MultiGraf.
- Data input is a simple spreadsheet format.

APS Data (Slide 1)



Copyright 2012 ECHO Dat

Month

APS Data (Slide 2)



Copyright 2012 ECHO Data

Hour of Day



APS Data (Slide 4)



Copyright 2012 ECHO Data,

21 Nation Study (slide 1)



Copyright 2012 ECHO Data, Ir

21 Nation Study (slide 2)



21 **Nation** Study (slide 3)



Copyright 2012 ECHO Dat

21 Nation Study (slide 4)



Baseball Data

	Batting	Home				Slugging
	Ave	Runs	Runs	RBIs	Walks	Ave
Aaron	327	47	95	118	71	669
Foxx	364	58	151	169	116	749
Gehrig	373	47	147	175	109	765
Greenberg	315	58	144	146	119	683
Hornsby	403	39	133	143	83	756
Mantle	353	52	132	130	112	705
Maris	269	61	132	142	94	620
Mays	319	51	123	127	79	659
Ruth	378	59	177	171	144	846
Simmons	381	36	152	165	39	708
Williams	406	37	135	120	145	735

