Ten Data Analysis Tools You Can't Afford to be Without

Neil W. Polhemus, CTO, StatPoint Technologies, Inc.

George H. Dyson, Director of Six Sigma Services

STATPOINT TECHNOLOGIES, INC.

Business Improvement Objectives

- Businesses must create value. This output must be greater than the inputs needed to produce it.
- If the output meets the customers' needs, the business is effective.
- If the business creates added value with minimum resources, the business is efficient.

The Role of Six Sigma is to Help a Business Produce the Maximum Value While Using Minimum Resources

(Pyzdek 2003)

STATPOINT TECHNOLOGIES, INC.

Six Sigma Business Successes

- **Cost reductions**
- **Market - share growth**
- **Defect reductions**
- **Culture changes**
- **Productivity improvements**
- **Customer relations improvements**
- **Product and service improvements**
- **Cycle - time reductions**

(CSSBB Primer, 2001) / (Pande, 2000)

All these Successes have a common thread….

STATPOINT TECHNOLOGIES, INC.

DATA!!!

- Data drives analysis.
- Analysis uses statistical models & tools of all kinds.
  - To extract meaningful information
  - To uncover signals in the presence of noise
  - To understand the past
  - To monitor the present
  - To forecast the future
- This understanding results in reduced cost and saving money.
- Deming: "Doesn't anyone care about Profit?"

STATPOINT TECHNOLOGIES, INC.

Examples

- Product comparisons
- Survey analysis
- Distribution fitting
- Comparison of multiple samples
- Outlier detection
- Curve fitting
- Response surface modeling
- Time series forecasting
- Event rate modeling
- Interactive maps

Problem #1 – Product Comparisons

- Consumer Reports 2010: Sedans (Family, Upscale, Luxury)

- 7 variables
  - Price (dollars)
  - Road-test score (0 – 100)
  - Predicted reliability (1 – 5)
  - Owner satisfaction (1 – 5)
  - Owner cost (1 – 5)
  - Safety (1 – 5)
  - Fuel economy (overall mpg)

STATPOINT TECHNOLOGIES, INC.

Data file: cars.sgd (n=30)

cars.sgd

	Model	Class	Price	Road Test	Reliability	Owner Satisfaction	Owner Cost	Safety	MPG
1	Nissan Altima	Family	28465	89	5	4	4	4	23
2	Honda Accord	Family	28695	88	4	4	4	3	21
3	Toyota Camry XLE	Family	29839	87	5	4	4	4	23
4	Volkswagen Passat	Family	27440	87	2	5	4	3	24
5	Toyota Camry Hybrid	Family	29720	84	3	3	3	4	34
6	Hyundai Sonata	Family	26435	83	4	4	4	4	22
7	Chevrolet Malibu	Family	28045	83	5	4	4	3	20
8	Ford Fusion	Family	28400	80	3	4	4	3	22
9	Mercury Milan	Family	28025	80	3	4	4	3	22
10	Nissan Altima	Family	28225	78	3	4	4	4	32
11	Ford Taurus	Family	27510	75	5	5	4	3	18
12	Kia Optima	Family	24640	73	3	2	4	4	22
13	Chevrolet Impala	Family	26840	64	5	2	4	5	20
14	Lexus ES	Upscale	38615	91	4	4	2	4	23
15	Toyota Avalon	Upscale	33070	89	4	4	4	3	22
16	Acura TL	Upscale	35715	85	4	4	5	3	23
17	Hyundai Azera	Upscale	30075	83	5	4	5	4	19
18	Lincoln MKZ	Upscale	32675	77	4	4	5	3	20
19	Buick Lucerne	Upscale	38935	73	4	5	2	4	17
20	Volvo S60	Upscale	34980	58	4	5	5	4	22

cars | B | C

STATPOINT TECHNOLOGIES, INC.

Multivariate Visualization

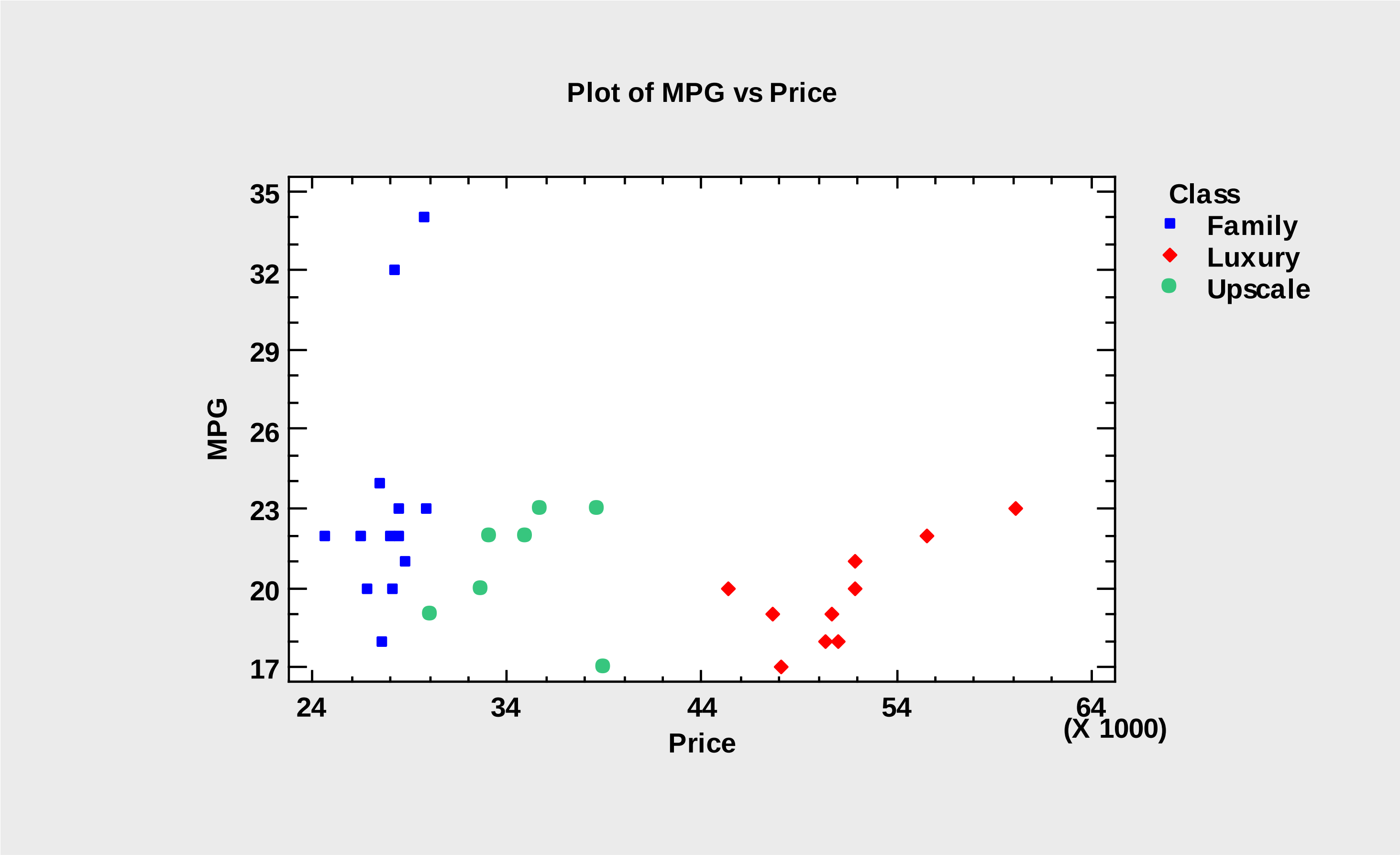
- The data consist of 7 quantitative variables plus one categorical factor. Each car is thus a point in 7-dimensional space with one other feature. How can we visualize what's going on?

1. 2-D scatterplot
2. 3-D scatterplot
3. Scatterplot matrix
4. Bubble chart
5. Parallel coordinates plot
6. Star glyphs
7. Chernoff faces
8. Radar/spider plot

STATPOINT TECHNOLOGIES, INC.

2-D Scatterplot

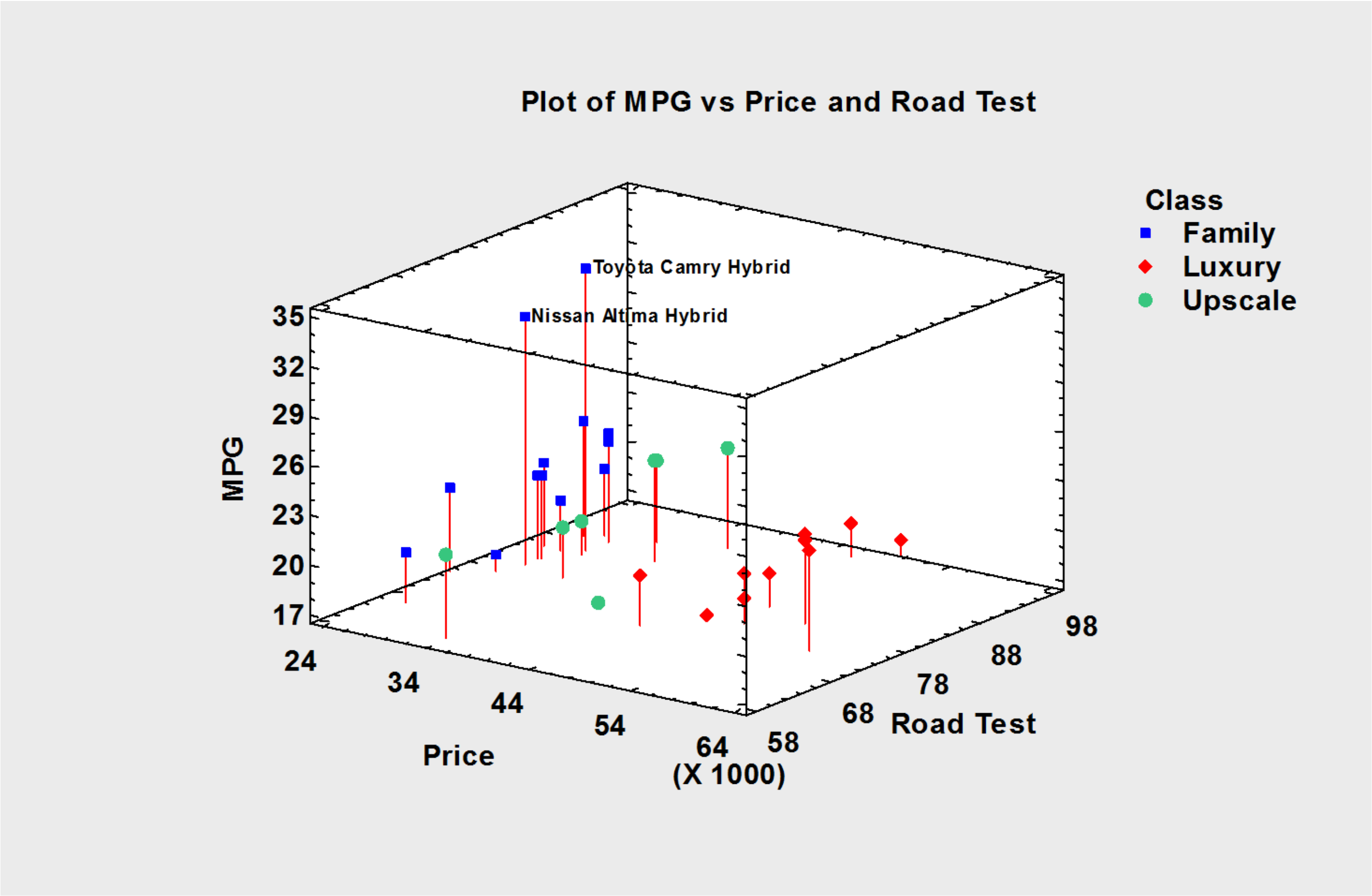
Useful for plotting 2 dimensions.



STATPOINT TECHNOLOGIES, INC.

3-D Scatterplot

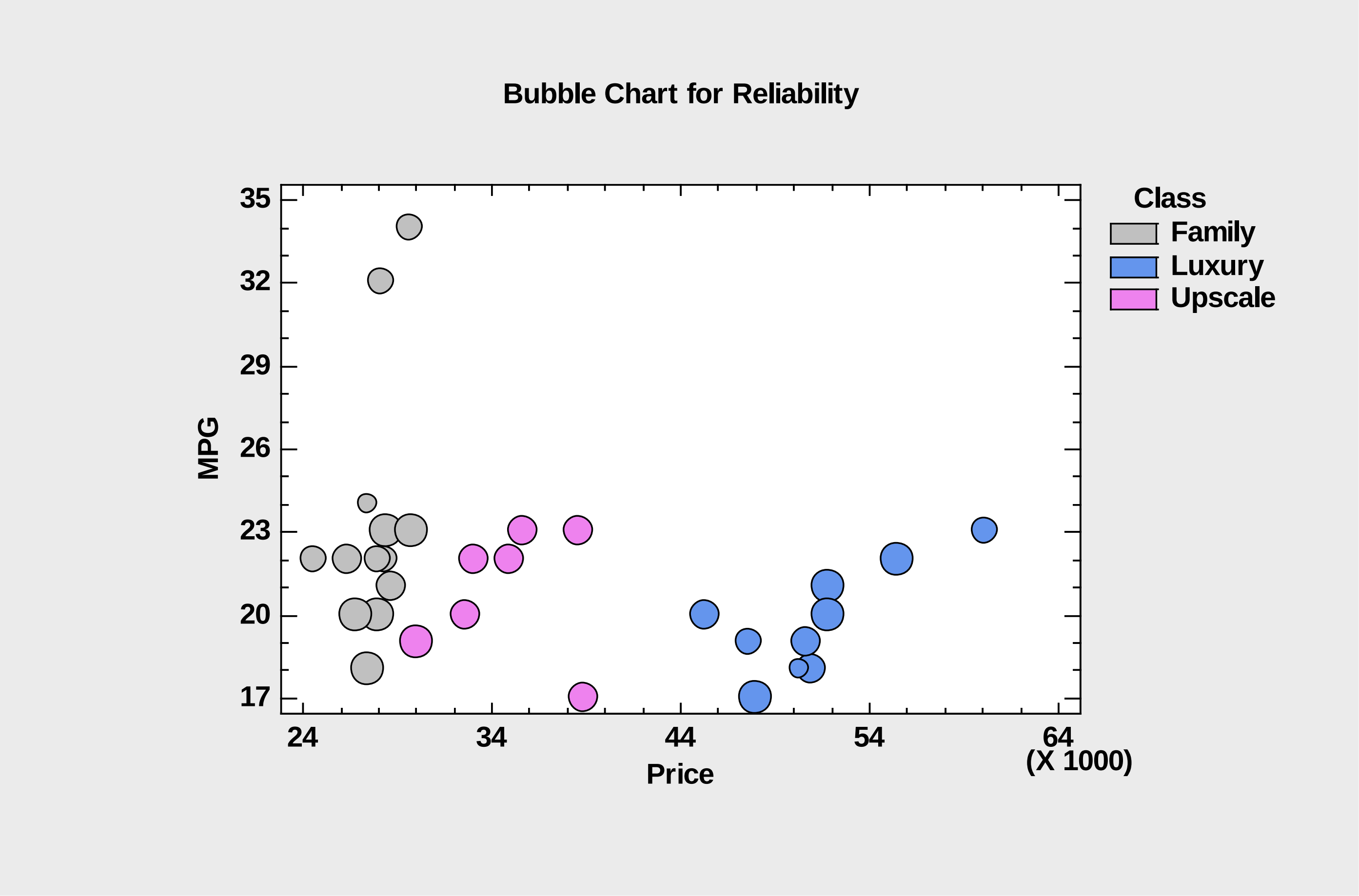
Useful for plotting 3 dimensions.



STATPOINT TECHNOLOGIES, INC.

Bubble Chart

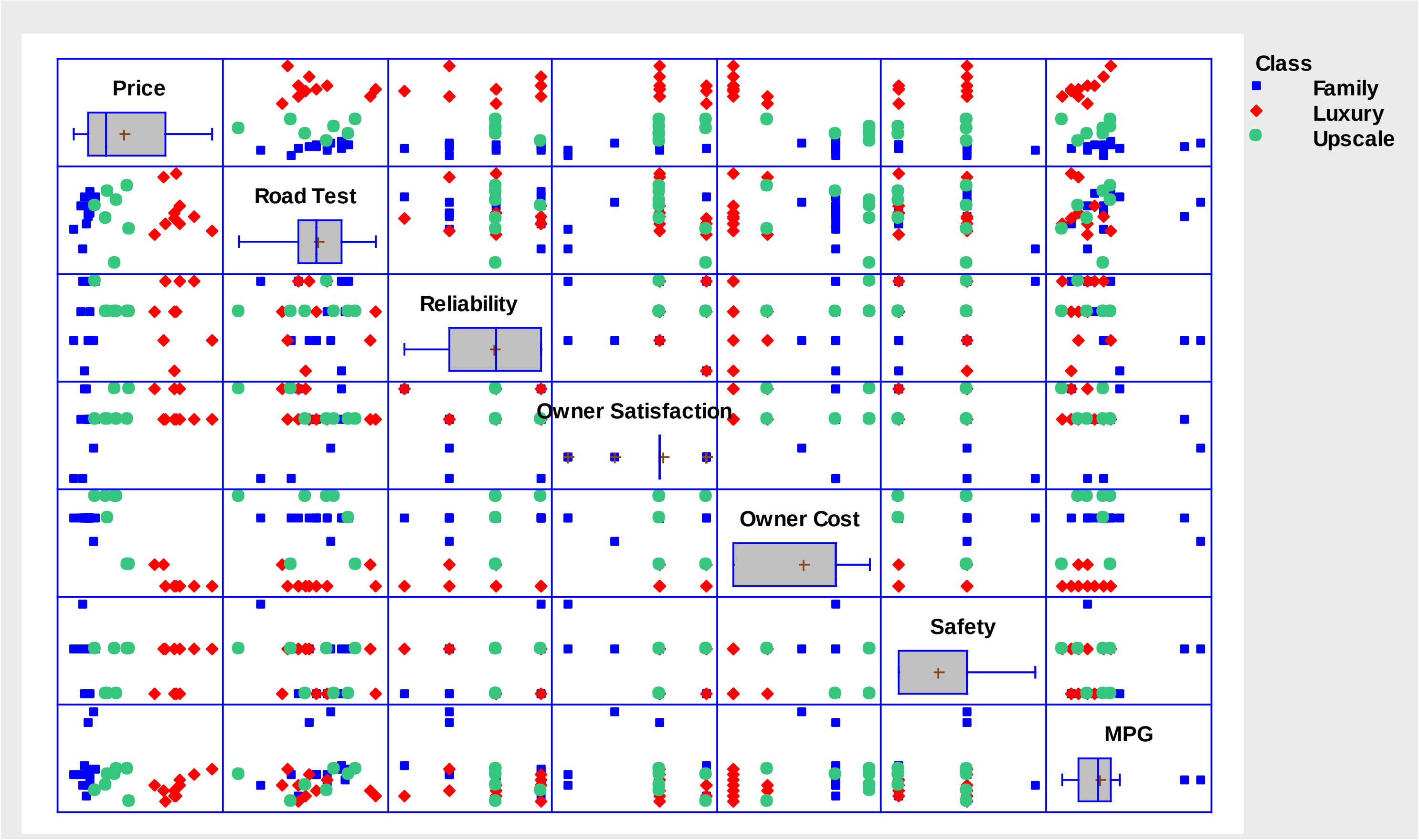
Uses size of bubble to illustrate a third dimension.



STATPOINT TECHNOLOGIES, INC.

Scatterplot Matrix

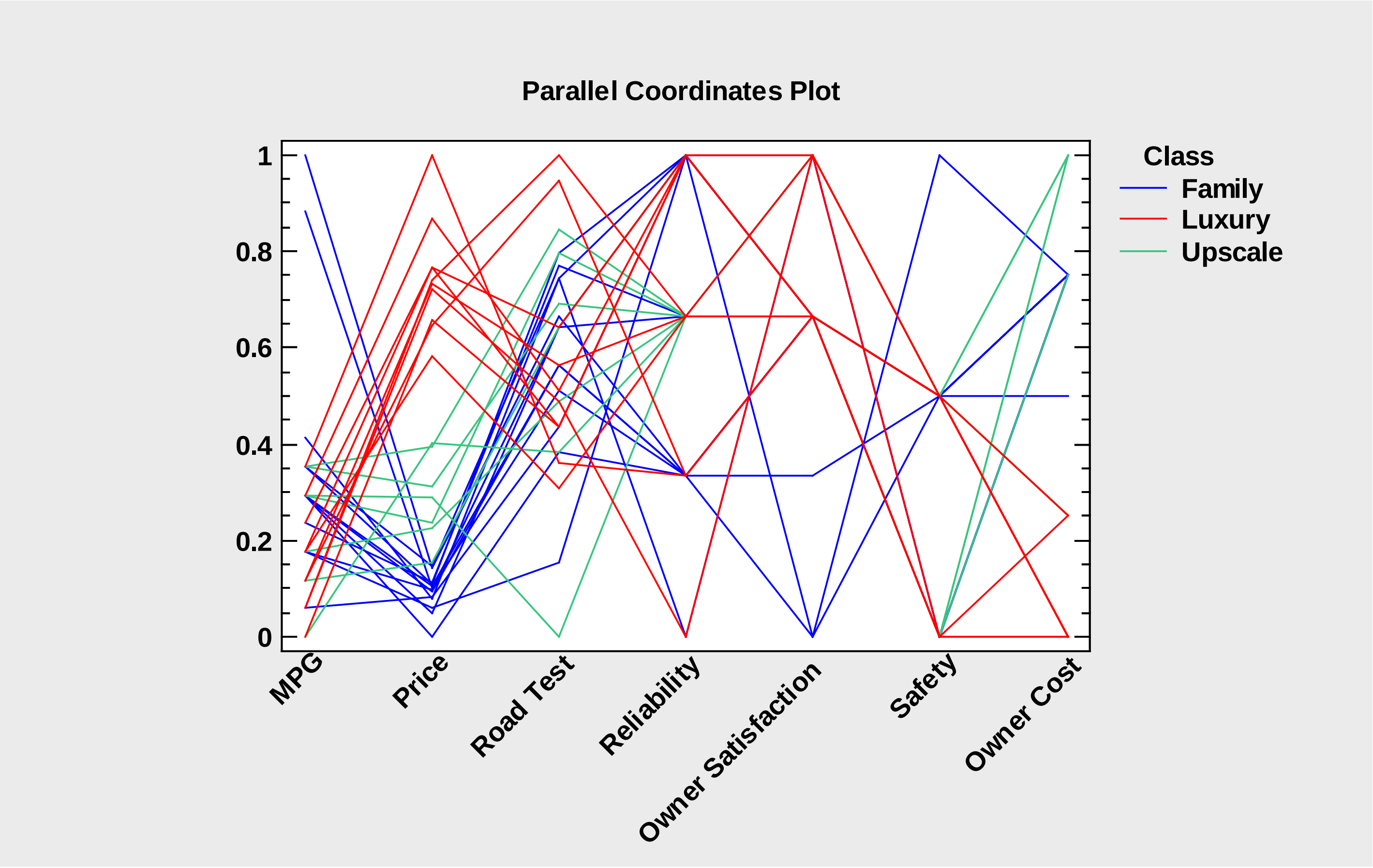
Plots all pairs of variables.



STATPOINT TECHNOLOGIES, INC.

Parallel Coordinates Plot

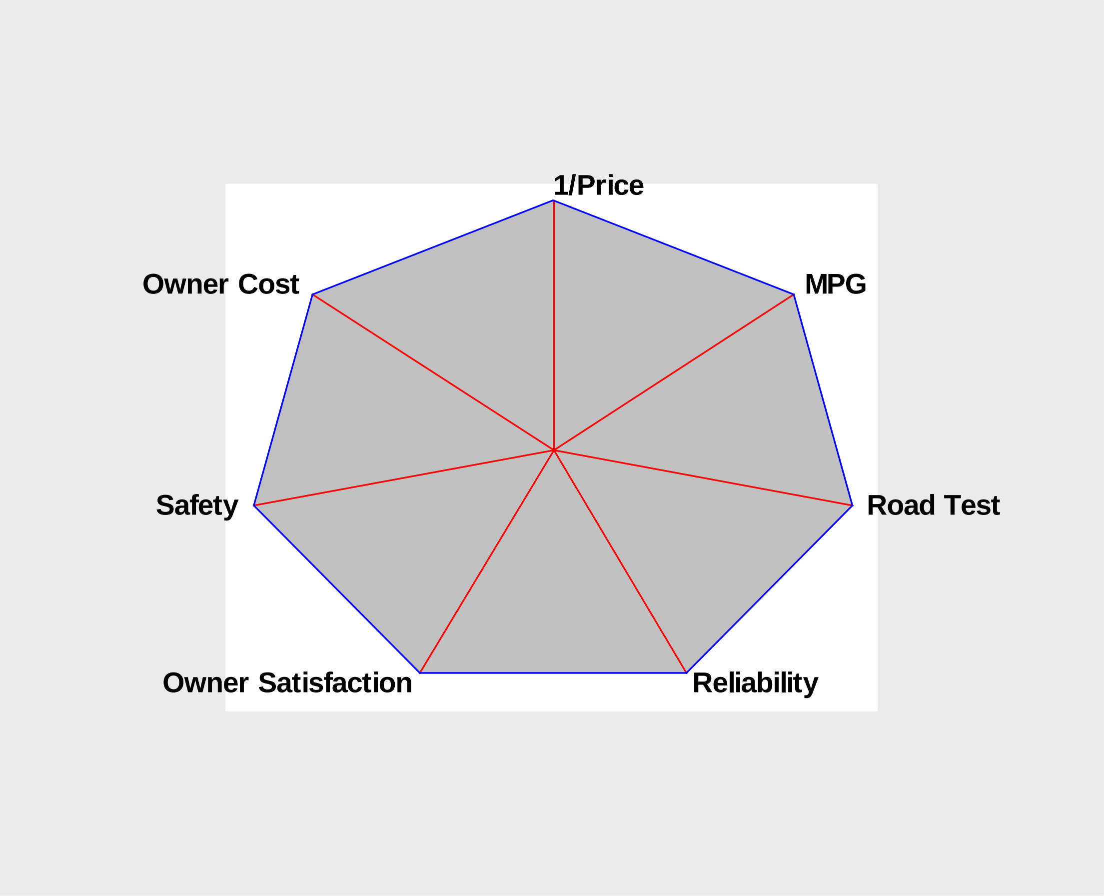
Each case is shown as a line connecting the values of the variables.



STATPOINT TECHNOLOGIES, INC.

Star Glyphs

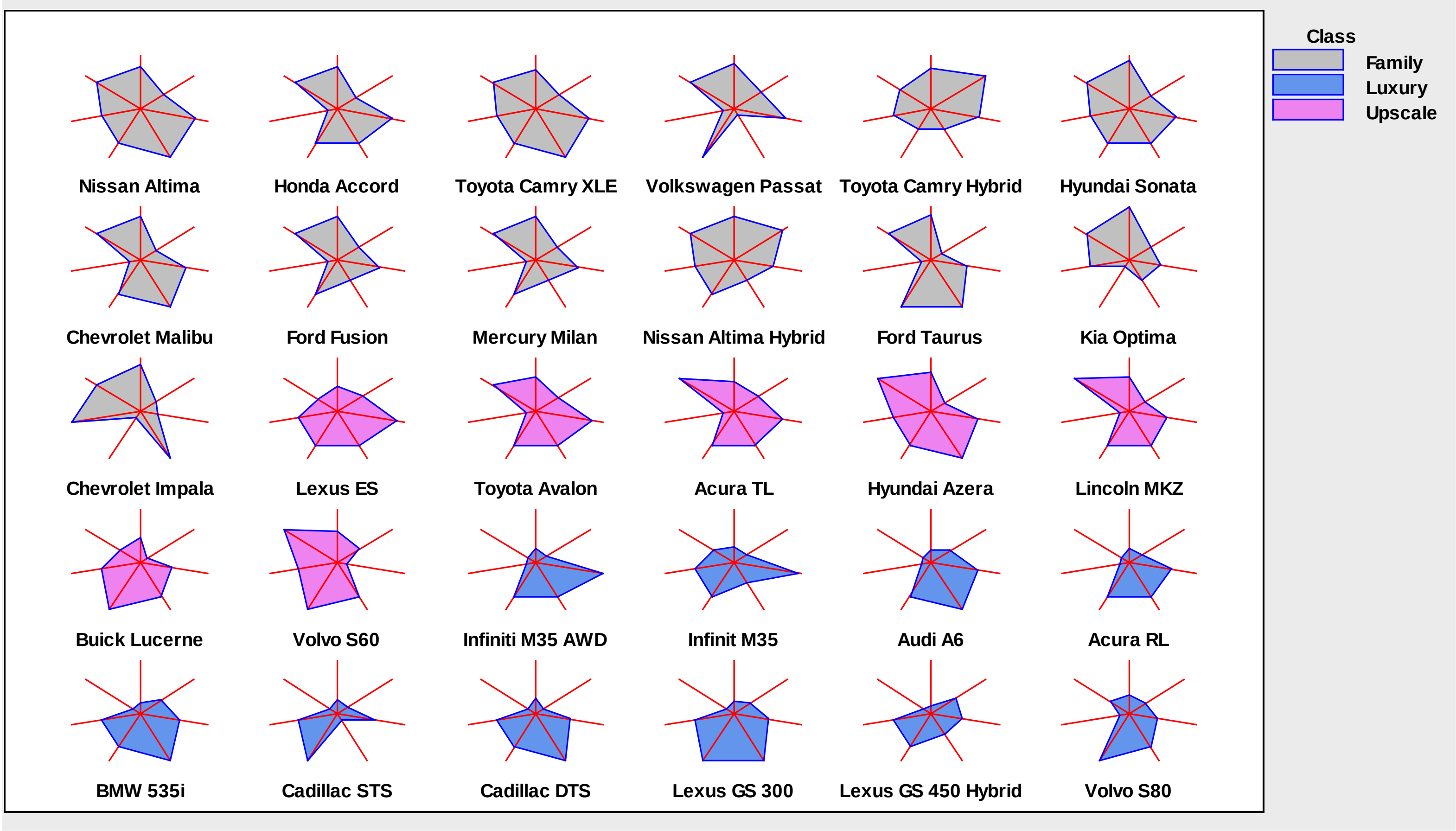
Each case is shown as a polygon with vertices scaled by the variables.



STATPOINT TECHNOLOGIES, INC.

Star Glyphs

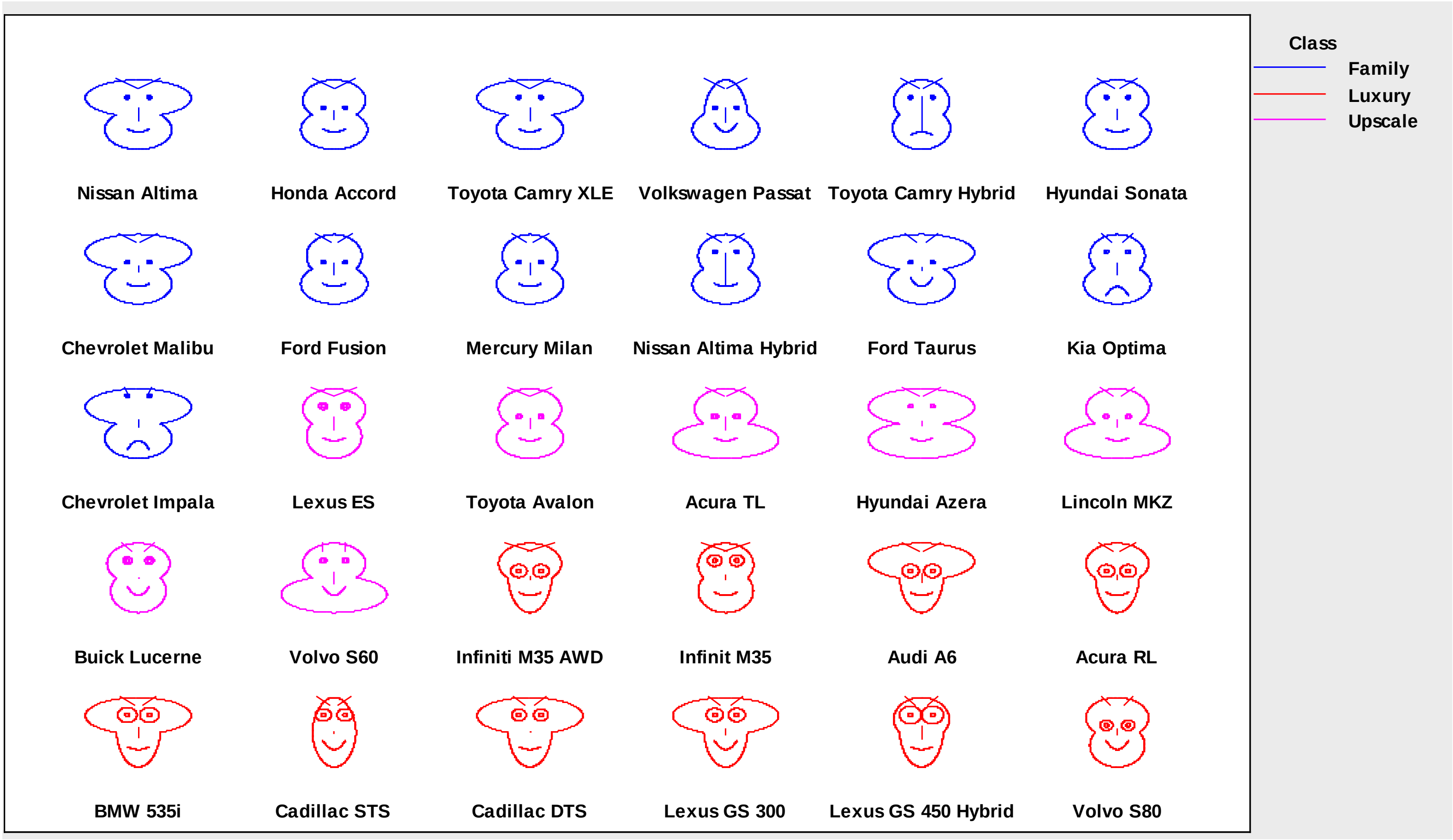
Variables are scaled so that larger area is better.



STATPOINT TECHNOLOGIES, INC.

Chernoff Faces

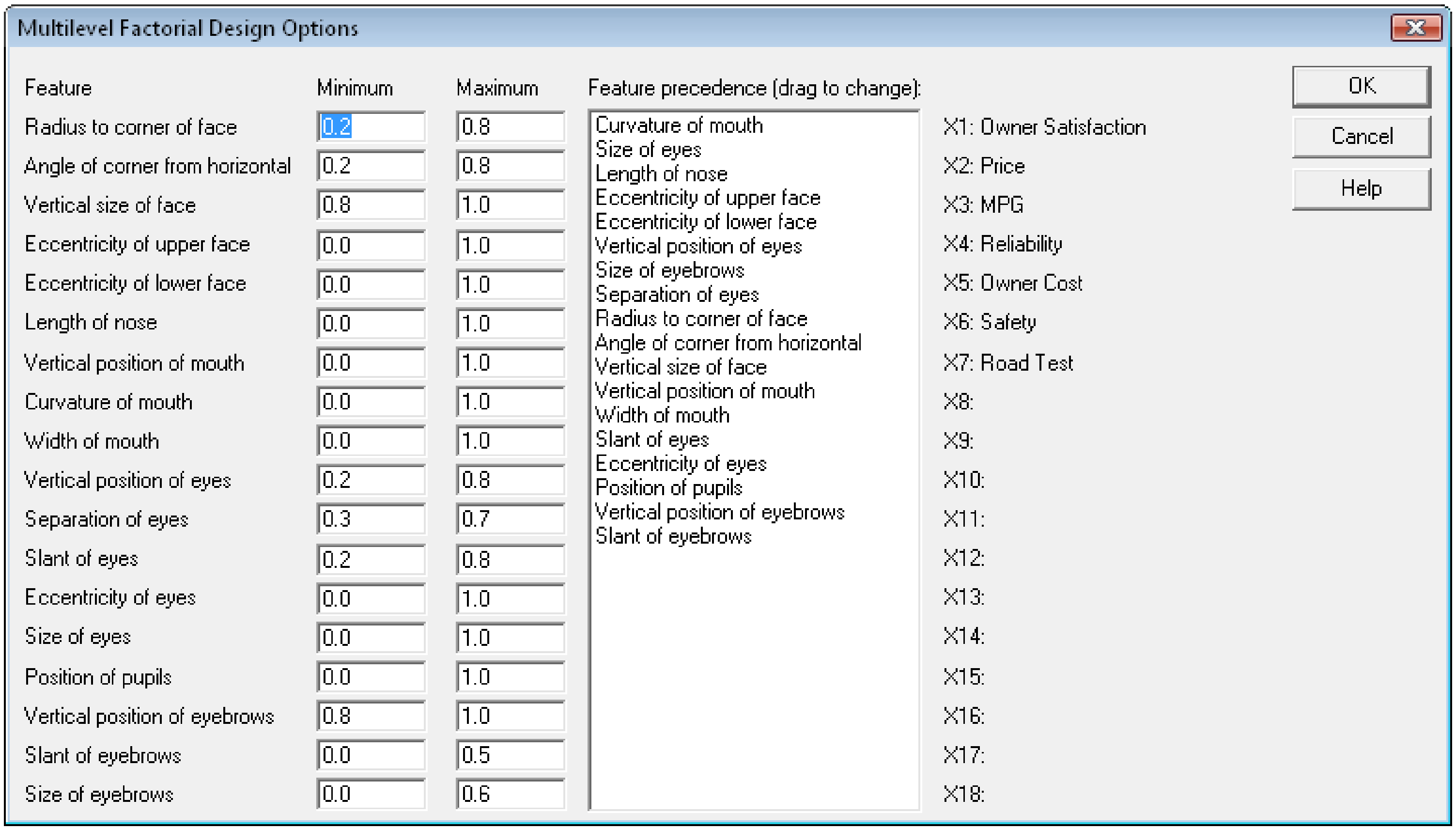
Each variable is assigned to a different feature of the faces.



STATPOINT TECHNOLOGIES, INC.

Feature Assignments

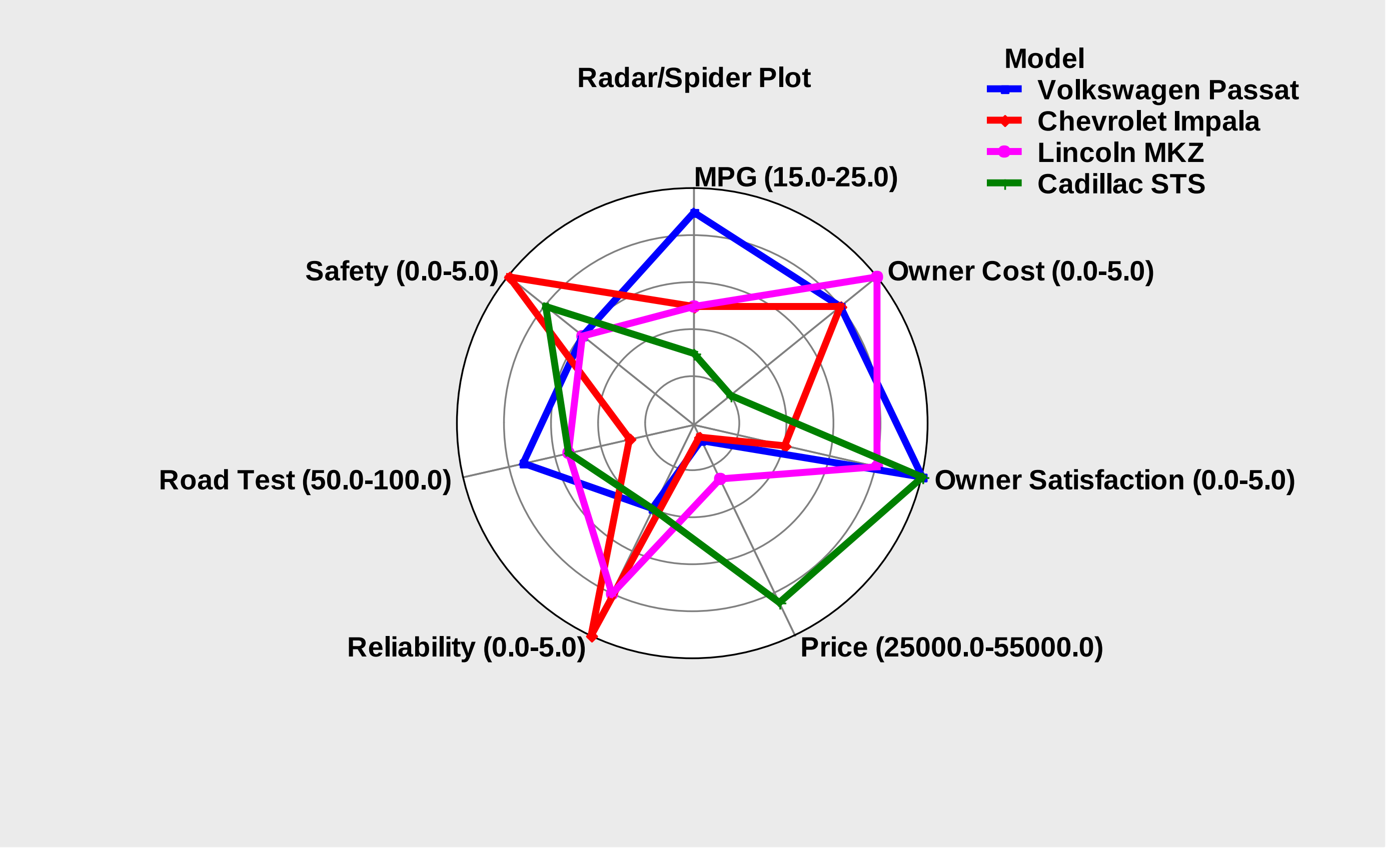
Unfortunately, some features have a greater impact than others.



STATPOINT TECHNOLOGIES, INC.

Radar/Spider Plot

Good for comparing a small number of cases.



STATPOINT TECHNOLOGIES, INC.

Problem #2: Survey Analysis

- The most commonly used statistical procedure is the calculation of a two-way table (a tabulation of responses that can be classified in 2 ways).

- Such *crosstabulations* result in contingency tables that provide much useful information.

- Example: On January 18-19, 2010, Rasmussen Reports asked 1000 likely voters: "How would your rate the U.S. healthcare system?"

Data file: healthcare.sgd

healthcare.sgd

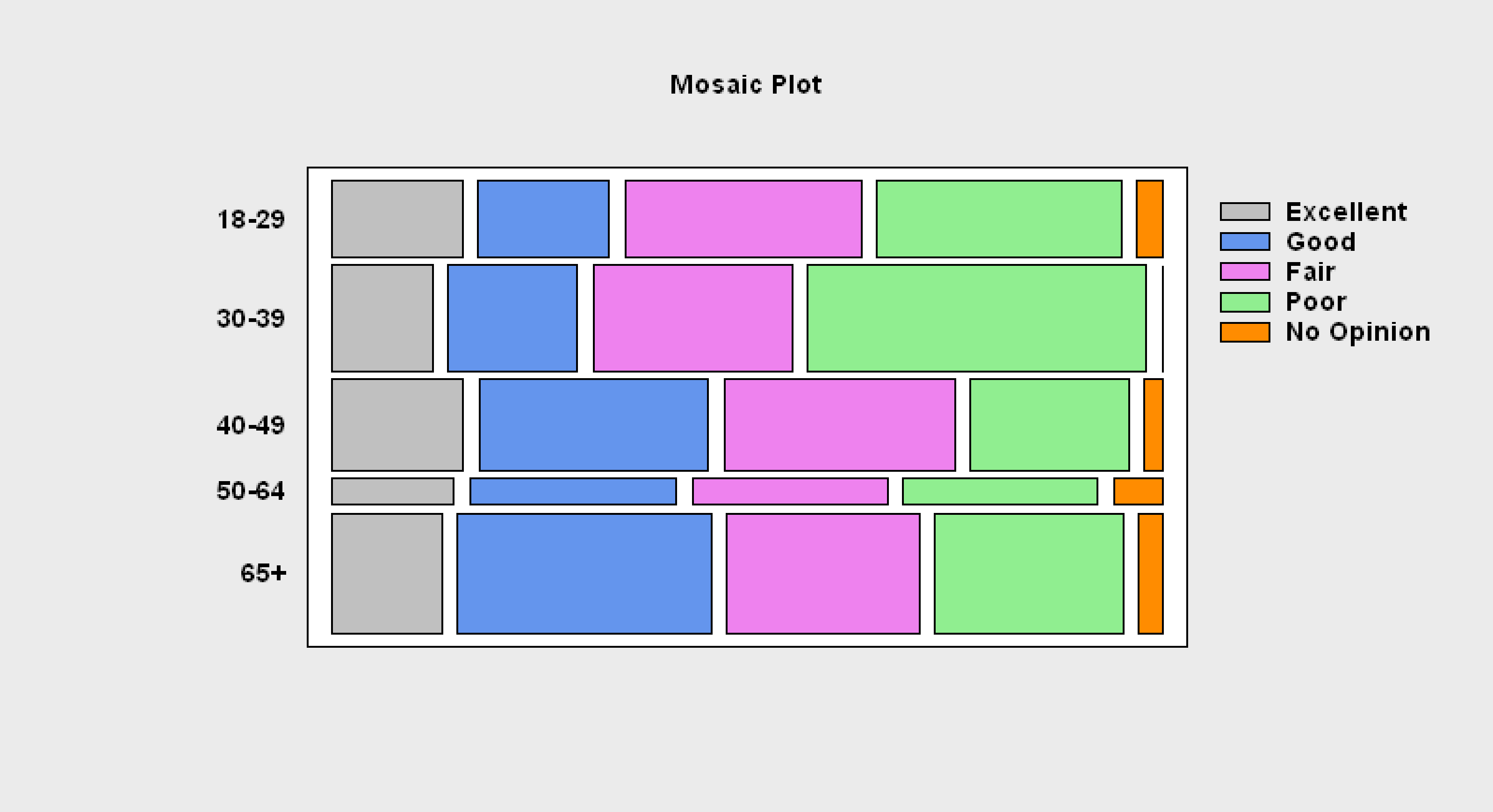
	Age group	Excellent	Good	Fair	Poor	No Opinion
1	18-29	31	31	56	58	6
2	30-39	33	43	66	112	0
3	40-49	37	65	65	45	5
4	50-64	10	17	16	16	4
5	65+	41	95	72	70	9
6						
7						

healthcare B C D E F G H I J K L M

STATPOINT TECHNOLOGIES, INC.

Mosaic Plot

Scales the area of each bar according to the counts in the table.



Chi-square Test

Tests for lack of independence between row and column classification.

Contingency Tables

Tests of Independence

Test	*Statistic*	*Df*	*P-Value*
Chi-Square	63.149	16	0.0000

Warning: some expected cell counts < 5.

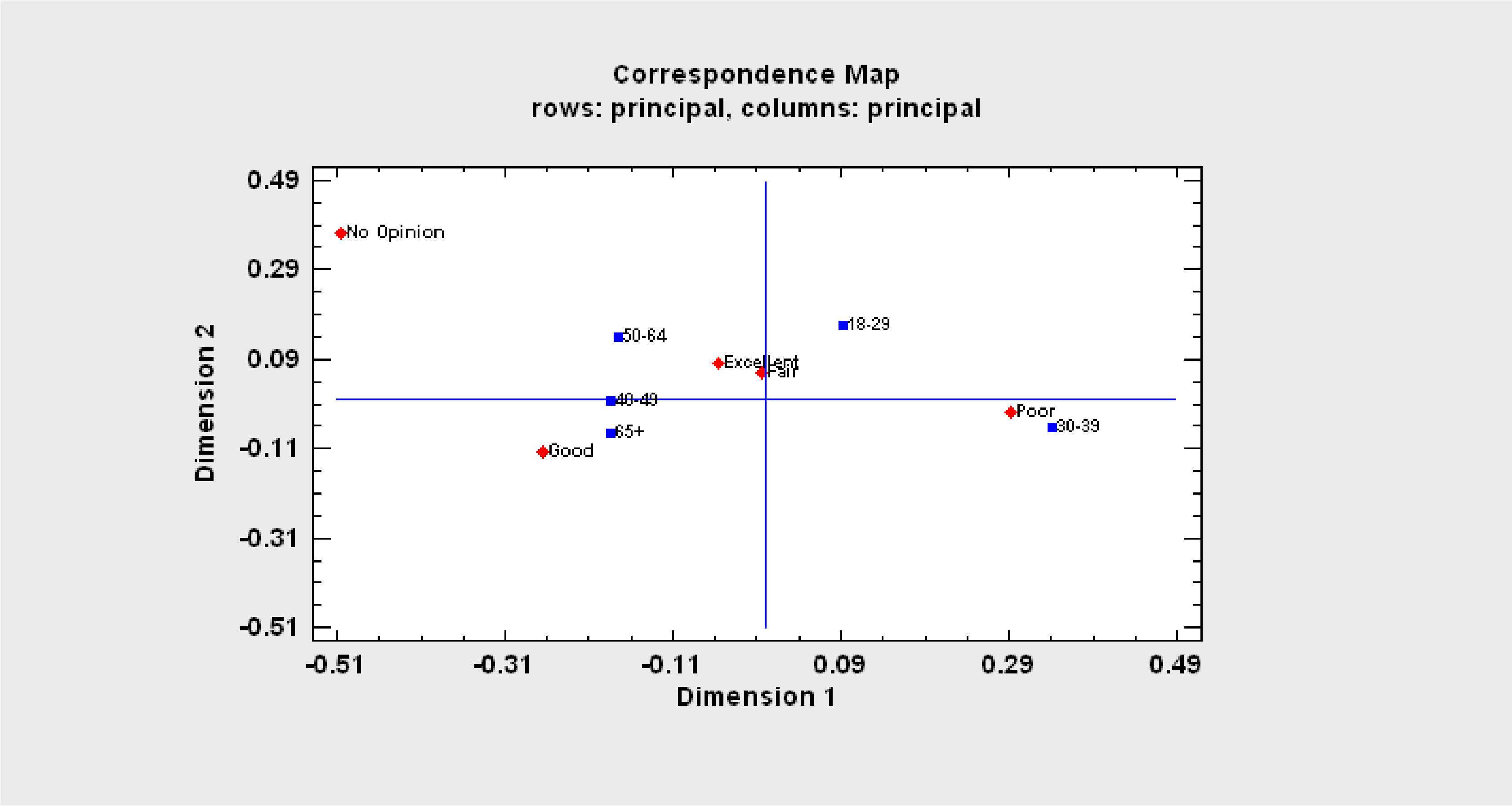
The StatAdvisor

This table shows the results of a hypothesis test run to determine whether or not to reject the idea that the row and column classifications are independent. Since the P-value is less than 0.05, we can reject the hypothesis that rows and columns are independent at the 95.0% confidence level. Therefore, the observed row for a particular case is related to its column.

STATPOINT TECHNOLOGIES, INC.

Correspondence Analysis

Used to help visualize the important information in two-way tables.



STATPOINT TECHNOLOGIES, INC.

Problem #3: Distribution Fitting

- In many studies, determining the distribution of a quantitative variable is critical.
- Common examples covered in Six Sigma include capability studies.
- Distribution fitting is also quite critical in many design problems.

Data file: waves.sgd (n=26,304)

Source: www.iahr.net

waves.sgd

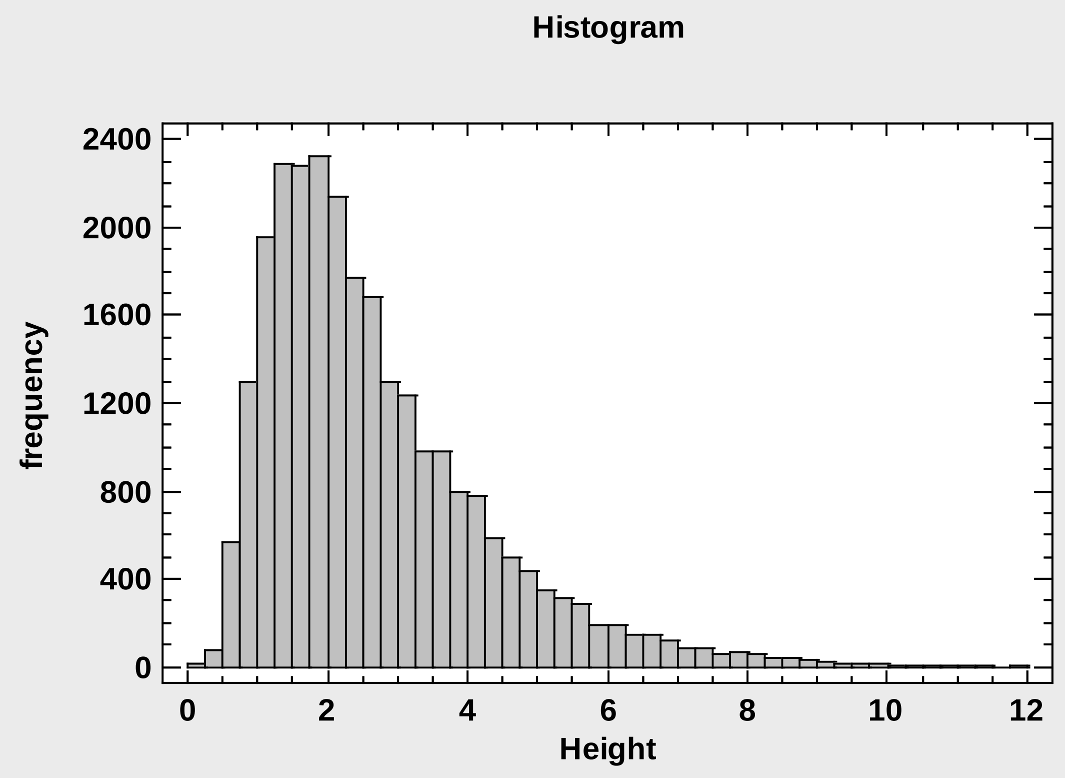
	Height	Col_2
	meters	
1	5.09	
2	4.51	
3	3.93	
4	3.73	
5	3.52	
6	3.37	
7	3.22	
8	3.17	
9	3.12	
10	3.02	
11	2.92	
12	2.68	
13	2.43	
14	2.24	
15	2.05	
16	1.87	
17	1.69	
18	1.56	

waves B C

STATPOINT TECHNOLOGIES, INC.

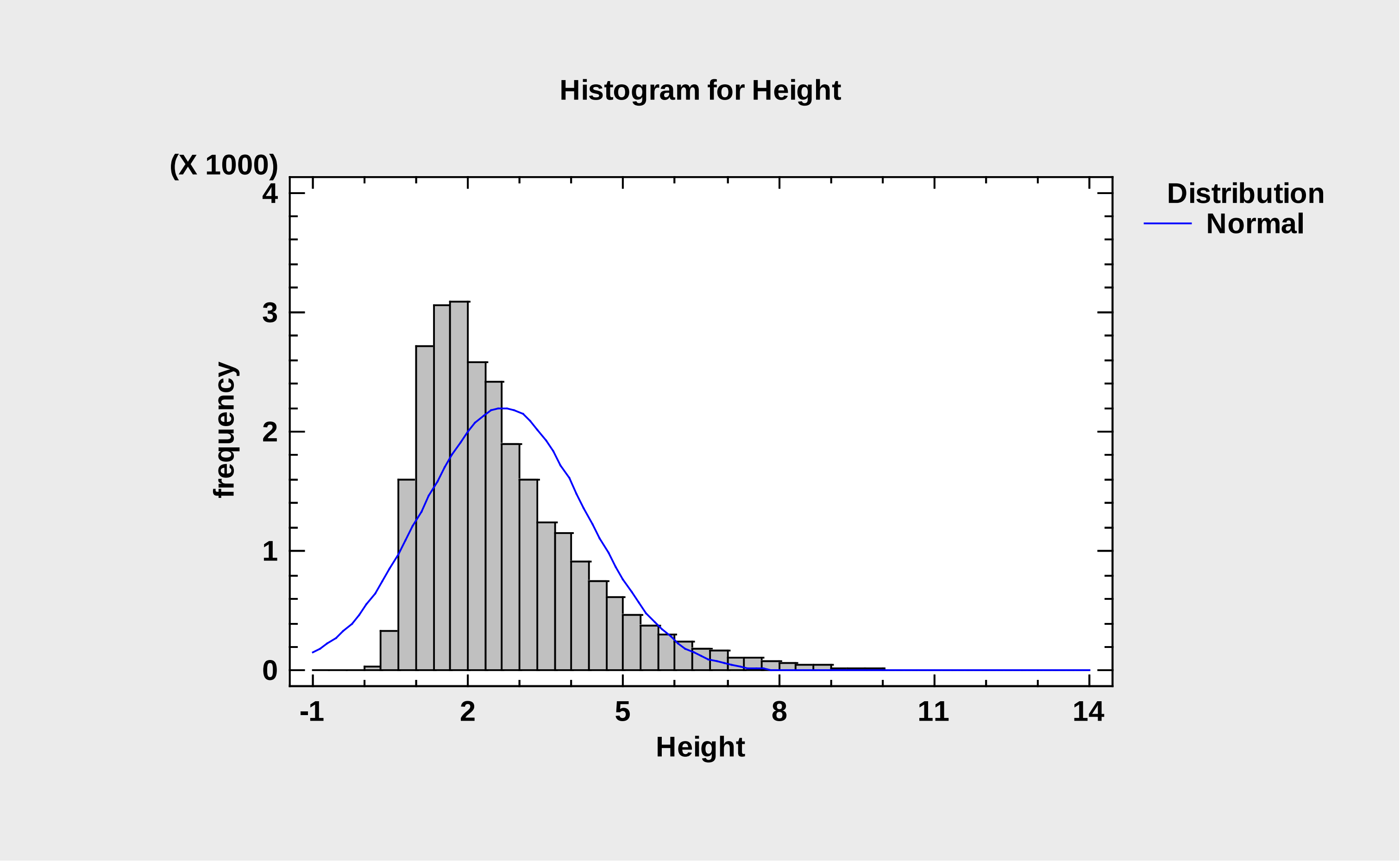
Frequency Histogram

Shows the number of observations in non-overlapping intervals.



Normal Distribution

The normal distribution is a poor model for this data.



Comparison of Distributions

Fits many distributions and sorts them by goodness of fit.

Uncensored Data - Height

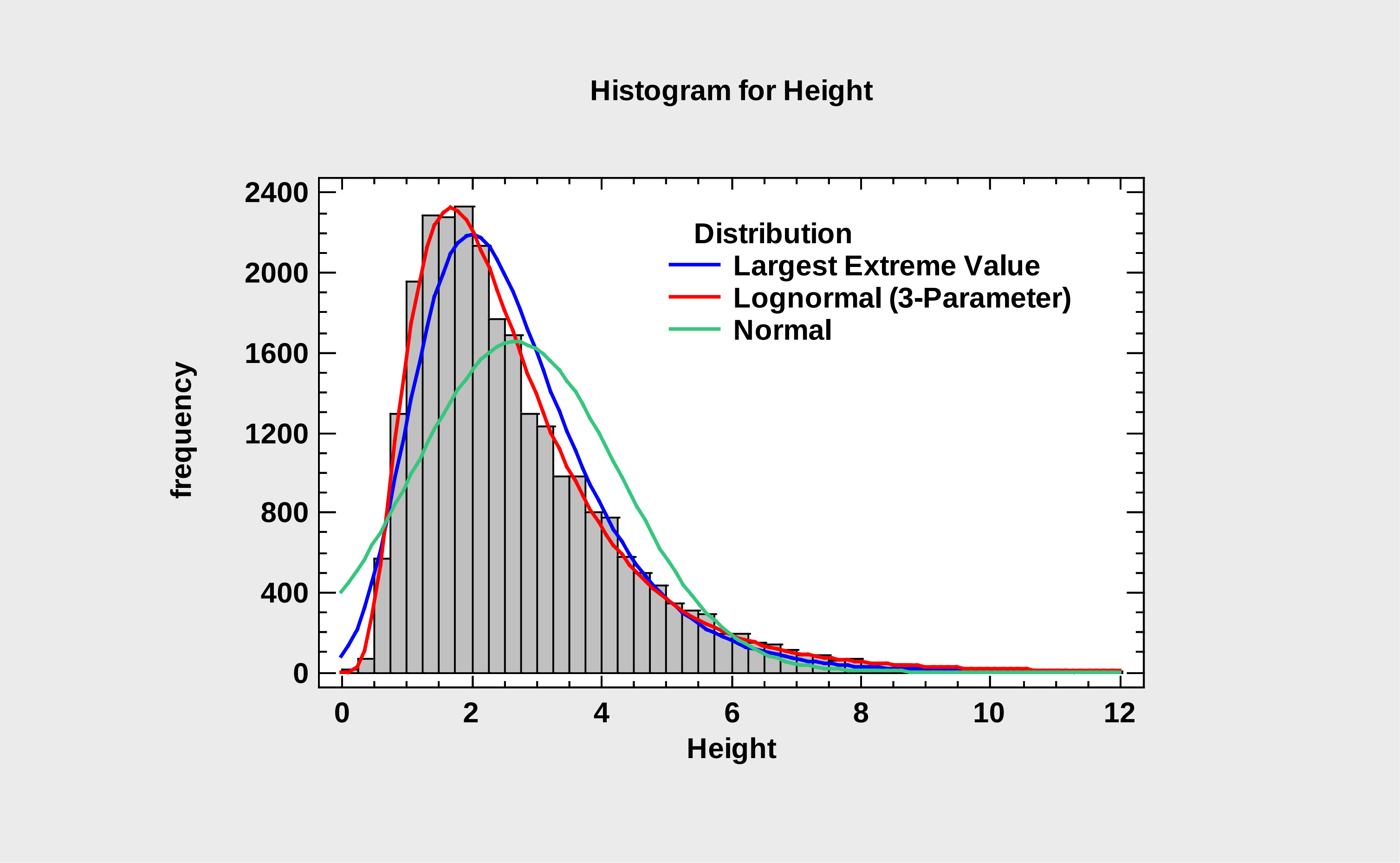
Comparison of Alternative Distributions

Distribution	*Est. Parameters*	*Log Likelihood*	*KS D*	*A^2*
Lognormal (3-Parameter)	3	-44492.3	0.0158875	8.90066
Largest Extreme Value	2	-45244.7	0.0493346	129.004
Logistic	2	-48435.3	0.0917526	459.214
Laplace	2	-48657.4	0.112977	590.784
Normal	2	-49533.0	0.111481	712.055
Exponential	1	-52321.9	0.238928	
Uniform	2	-66457.8	0.518133	
Lognormal	2	-1.E9	0.202882	
Weibull	2	-1.E9	0.442693	
Gamma	2	-1.E9	0.28725	
Pareto	1	-1.E9	0.435193	
Loglogistic	2	-1.E9	0.343987	
Inverse Gaussian	2	-2.6304E13	0.999962	
Birnbaum-Saunders	<no fit>			
Smallest Extreme Value	<no fit>			

STATPOINT TECHNOLOGIES, INC.

Lognormal Distribution

The 3-parameter lognormal distribution is much better.



Problem #4: Multiple Samples

- Data are frequently obtained from more than one sample.
- Asserting a significant difference between the samples (or lack thereof) is an important application of data analysis.

STATPOINT TECHNOLOGIES, INC.

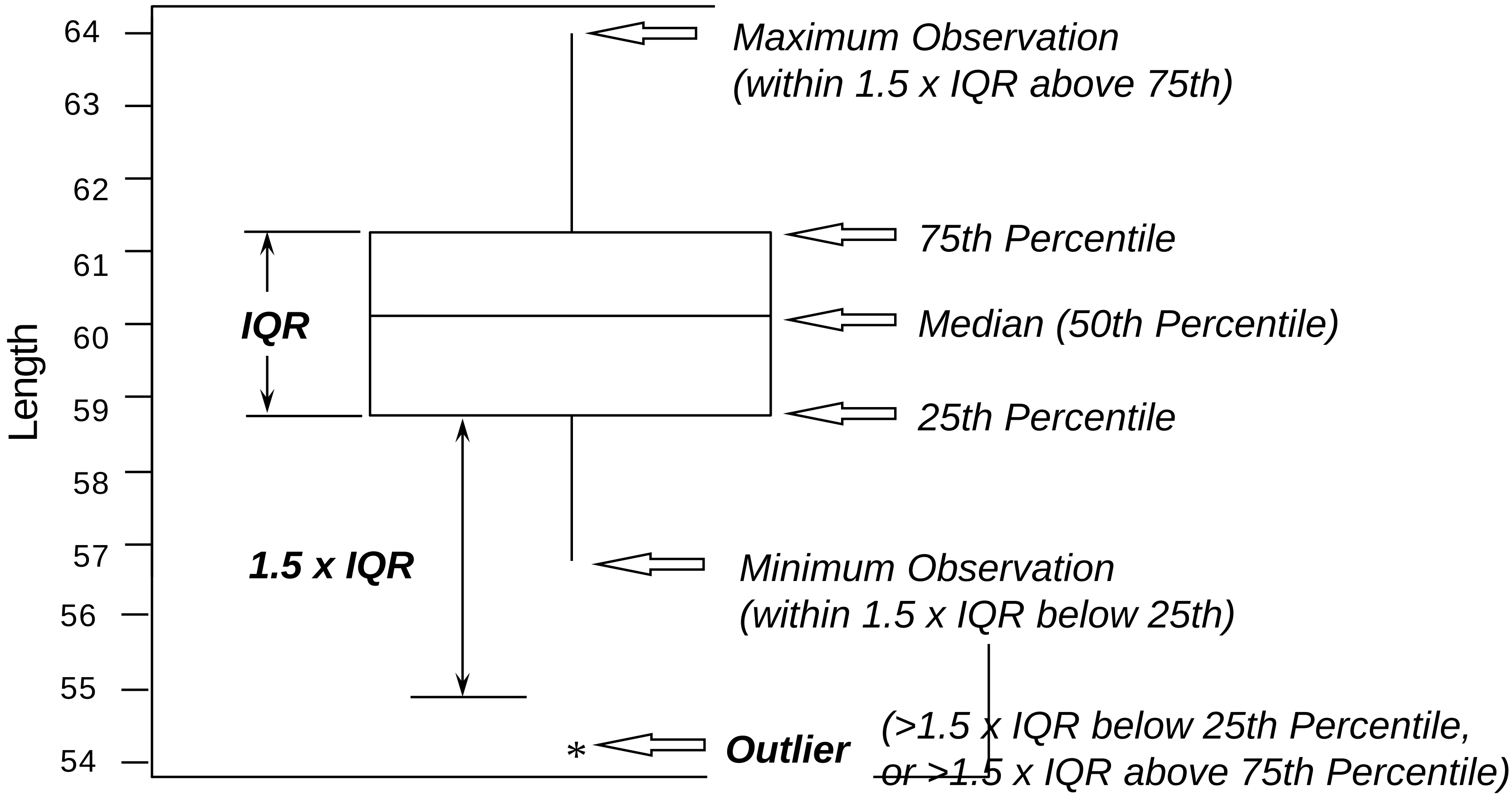
Data file: thickness.sgd (n=480)

	Location	Thickness
1	TOP	2.4
2	TOP	2.5
3	TOP	2.4
4	TOP	2.4
5	TOP	2.5
6	TOP	2.2
7	TOP	2.7
8	TOP	2.3
9	MIDDLE	2.4
10	MIDDLE	2.1
11	MIDDLE	2.5
12	MIDDLE	2.2
13	MIDDLE	2.4
14	MIDDLE	2.4
15	MIDDLE	2.6
16	MIDDLE	2.3
17	BOTTOM	2.7
18	BOTTOM	2.2
19	BOTTOM	2.8
20	BOTTOM	2.3

STATPOINT TECHNOLOGIES, INC.

Box-and-Whisker Plots

A very useful plot for comparing samples (from John Tukey).

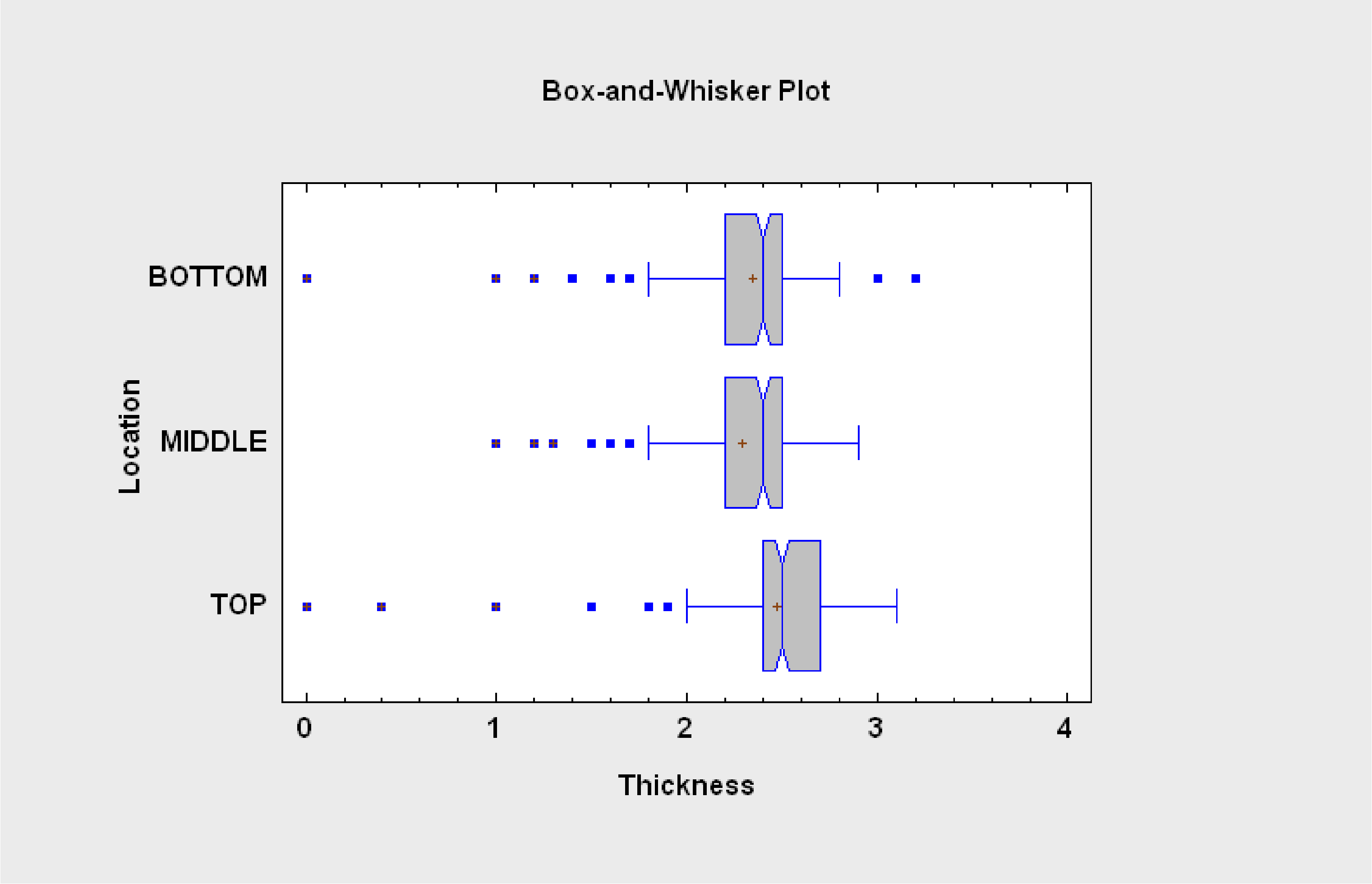


Plus sign may be added to show sample mean.

STATPOINT TECHNOLOGIES, INC.

Notched Box-and-Whisker Plots

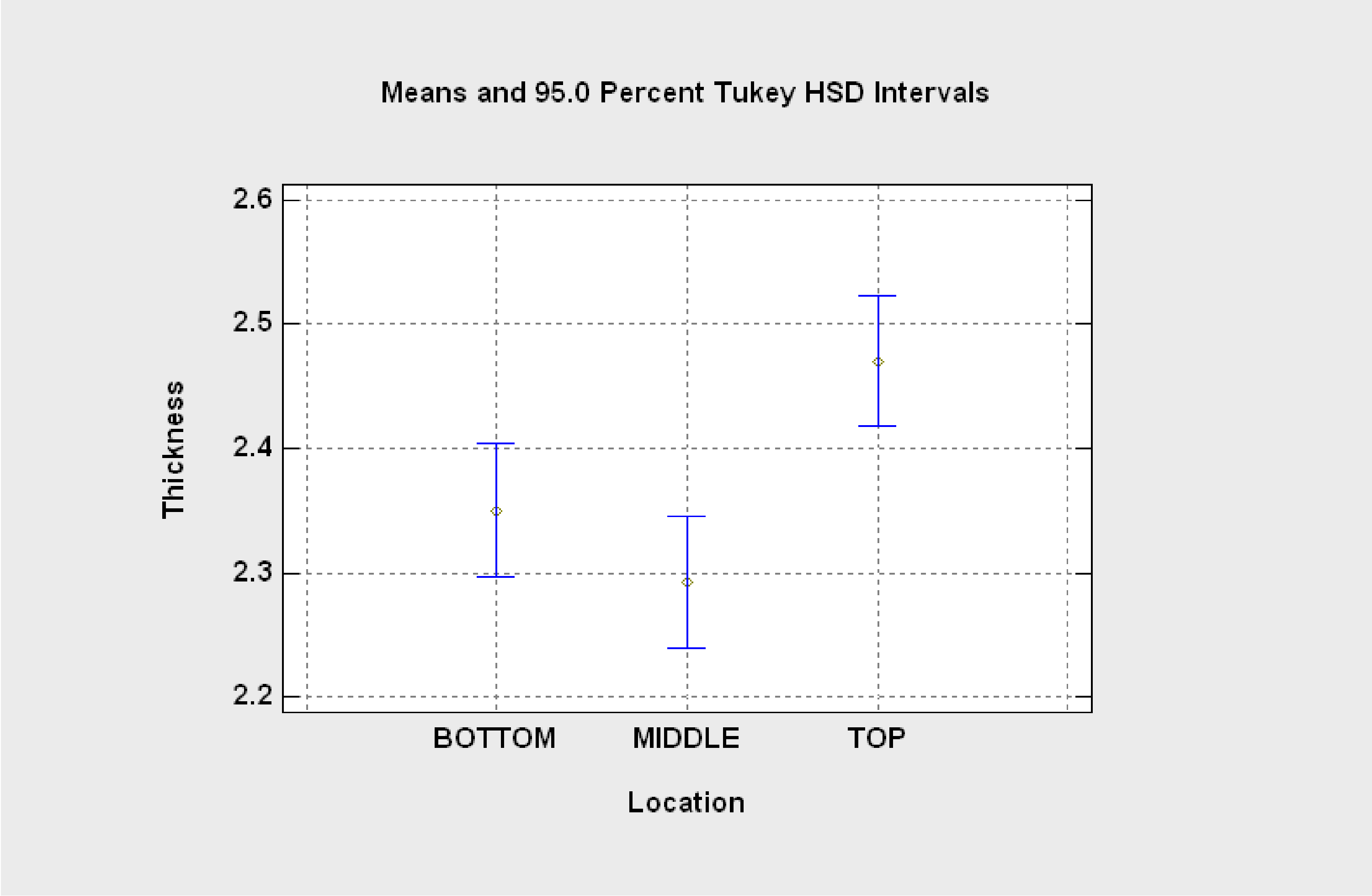
Non-overlapping notches indicate significantly different medians.



STATPOINT TECHNOLOGIES, INC.

HSD Intervals

Allow pairwise comparison of all level means.



Problem #5: Outlier Detection

- Many data sets contain aberrant observations that don't come from the same distribution as the others.

- Identifying outliers and treating them separately often results in better models.

Data file: bodytemp.sgd (n=130)

bodytemp.sgd

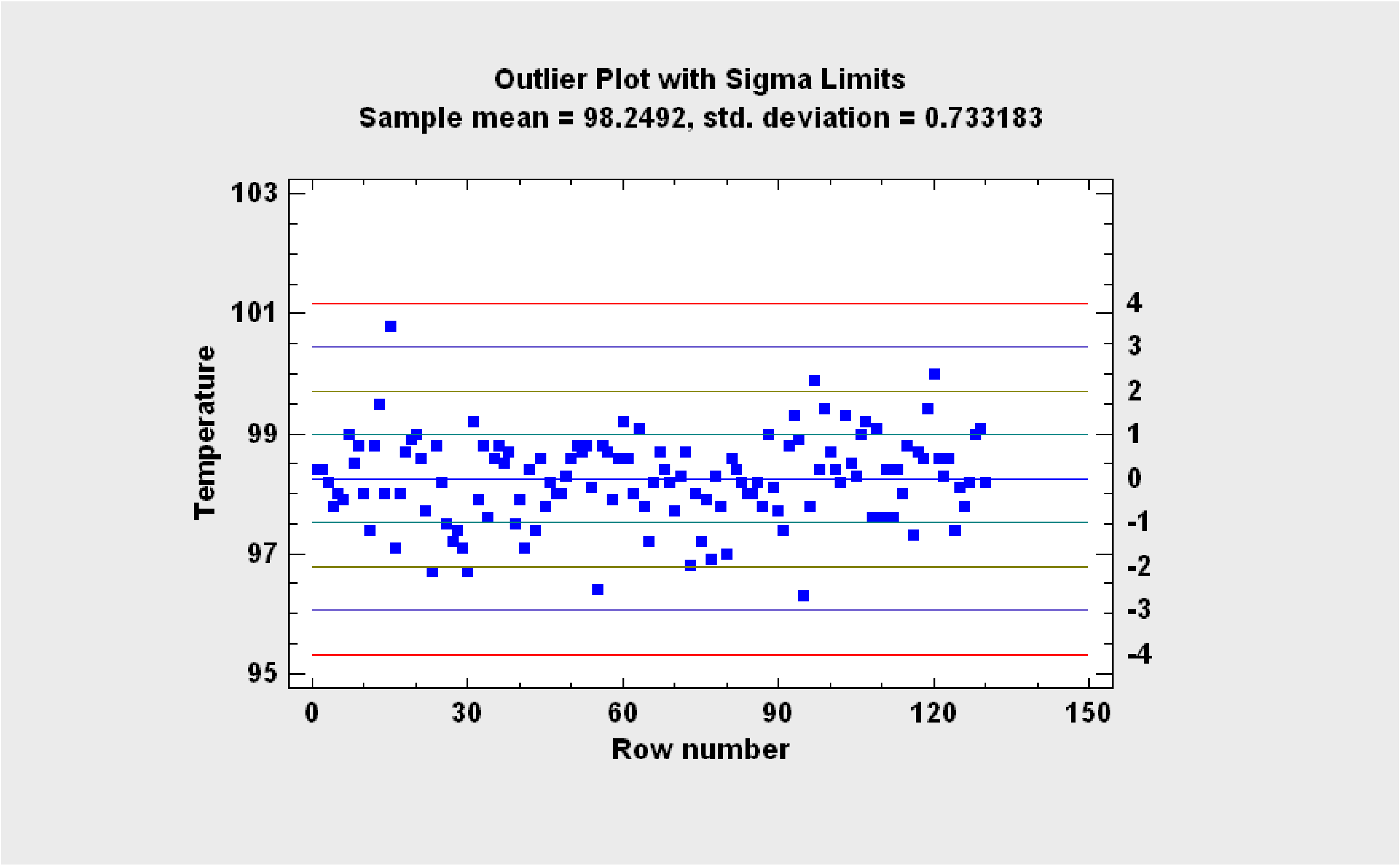
	Temperature degrees	Gender	Heart Rate beats per minute
1	98.4	Male	84
2	98.4	Male	82
3	98.2	Female	65
4	97.8	Female	71
5	98	Male	78
6	97.9	Male	72
7	99	Female	79
8	98.5	Male	68
9	98.8	Female	64
10	98	Male	67
11	97.4	Male	78
12	98.8	Male	78
13	99.5	Male	75
14	98	Female	73
15	100.8	Female	77
16	97.1	Male	75
17	98	Male	71
18	98.7	Female	72
19	98.9	Male	80
20	99	Male	75
21	98.6	Male	77

bodytemp / B / C / D / E

STATPOINT TECHNOLOGIES, INC.

Outlier Plot

Shows each data value with lines at 1, 2, 3 and 4-sigma.



STATPOINT TECHNOLOGIES, INC.

Grubbs' Test

Small P-value indicates that the extreme Studentized deviate (ESD) is highly unusual.

Outlier Identification - Temperature

Sorted Values

		Studentized Values	*Studentized Values*	*Modified*
Row	*Value*	*Without Deletion*	*With Deletion*	*MAD Z-Score*
95	96.3	-2.65859	-2.74567	-2.698
55	96.4	-2.52219	-2.59723	-2.5631
23	96.7	-2.11302	-2.15912	-2.1584
30	96.7	-2.11302	-2.15912	-2.1584
73	96.8	-1.97663	-2.01521	-2.0235
...				
99	99.4	1.56955	1.59096	1.4839
13	99.5	1.70594	1.7323	1.6188
97	99.9	2.25151	2.30628	2.1584
120	100.0	2.3879	2.45231	2.2933
15	100.8	3.47903	3.67021	3.3725

Grubbs' Test (assumes normality)

Test statistic = 3.47903

P-Value = 0.0484379

STATPOINT TECHNOLOGIES, INC.

Problem #6: Curve Fitting

- A common data analysis problem involves determining the relationship between a response variable Y and a predictor variable X.

- If we can estimate a model where Y = f(X), then we can use that model to make predictions.

STATPOINT TECHNOLOGIES, INC.

Data file: chlorine.sgd (n=44)

C:\Data\Six Sigma\chlorine.sgd

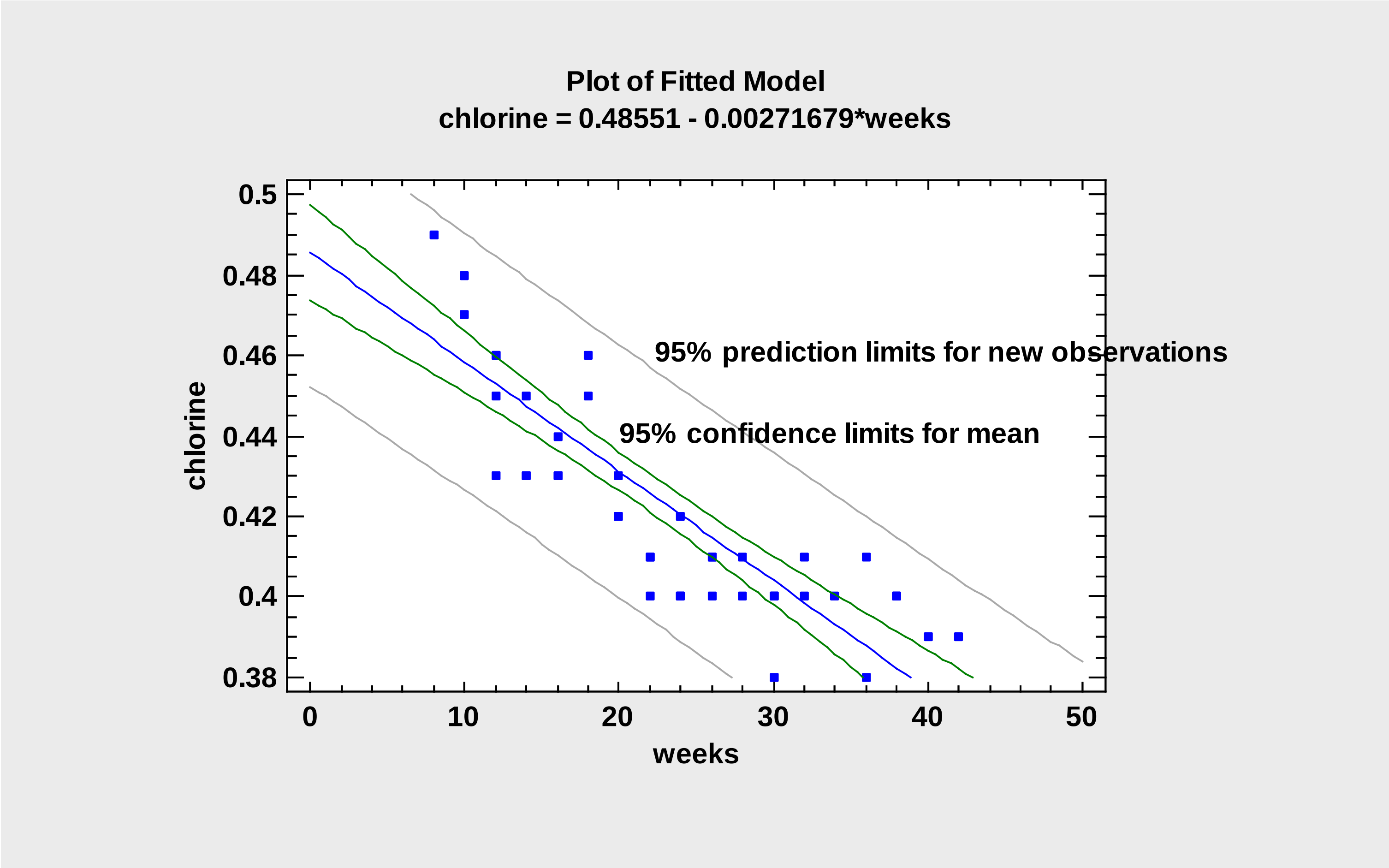
	weeks	chlorine
	weeks since production	percent available
1	8	0.49
2	8	0.49
3	10	0.48
4	10	0.47
5	10	0.48
6	10	0.47
7	12	0.46
8	12	0.46
9	12	0.45
10	12	0.43
11	14	0.45
12	14	0.43
13	14	0.43
14	16	0.44
15	16	0.43
16	16	0.43
17	18	0.46
18	18	0.45
19	20	0.42
20	20	0.42
21	20	0.43
22	22	0.41

chlorine B C

STATPOINT TECHNOLOGIES, INC.

Simple Linear Regression

Fits a linear model of the form Y = mX + b.



Comparison of Alternative Models

Fits many transformable nonlinear models and sorts by R-Squared.

Simple Regression - chlorine vs. weeks

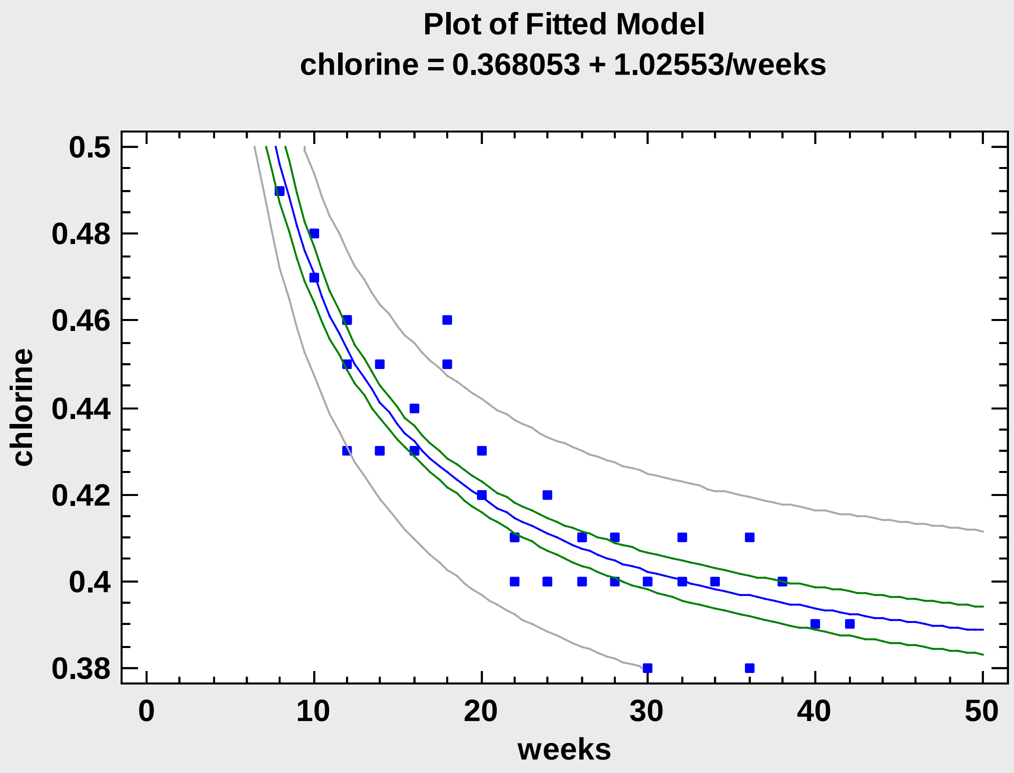
Comparison of Alternative Models

Model	*Correlation*	*R-Squared*
Squared-Y reciprocal-X	0.9367	87.75%
Reciprocal-X	0.9333	87.11%
Square root-Y reciprocal-X	0.9312	86.71%
S-curve model	0.9288	86.27%
Double reciprocal	-0.9233	85.25%
Reciprocal-Y logarithmic-X	0.9219	84.99%
Multiplicative	-0.9218	84.98%
Logarithmic-X	-0.9207	84.77%
Squared-Y logarithmic-X	-0.9185	84.36%
Reciprocal-Y square root-X	0.9038	81.69%
Logarithmic-Y square root-X	-0.9012	81.21%
Square root-X	-0.8974	80.54%
Squared-Y square root-X	-0.8926	79.68%
Reciprocal-Y	0.8759	76.73%
Exponential	-0.8710	75.87%
Square root-Y	-0.8682	75.37%
Logistic	-0.8665	75.08%
Log probit	-0.8662	75.03%
Linear	-0.8651	74.83%
Squared-Y	-0.8581	73.63%
Reciprocal-Y squared-X	0.8023	64.37%
Logarithmic-Y squared-X	-0.7941	63.05%
Square root-Y squared-X	-0.7896	62.34%
Squared-X	-0.7849	61.60%
Double squared	-0.7748	60.04%
Double square root	<no fit>	
Square root-Y logarithmic-X	<no fit>	

STATPOINT TECHNOLOGIES, INC.

Reciprocal X Model

A nonlinear model of the form Y = m/X + b is much better.



Problem #7: Response Surfaces

- The MISSION: air dominance at the lowest possible price.
- Use optimization models, built from performance data, to design the best aircraft engine.
- Note: the data have been altered and are for demonstration purposes only.

STATPOINT TECHNOLOGIES, INC.

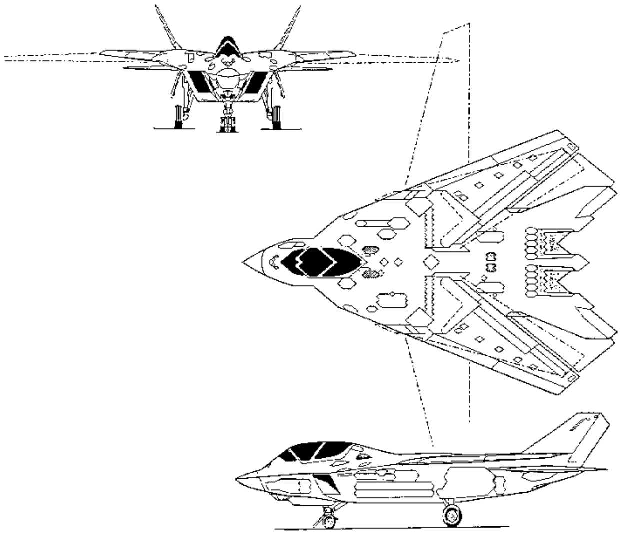
Problem Statement

- Optimize 3 response variables:

  Minimize total fleet acquisition cost (Y1)

  Maximize climb rate (Y2)

  Maximize launch rate (Y3)



- Input factors:

  X1: Fan Pressure Ratio: 3.9 to 4.7

  X2: Overall Pressure Ratio : 34 to 40

  X3: Inlet airflow : 240 to 270 pps

STATPOINT TECHNOLOGIES, INC.

Data file: engines.sgx (n=584)

engines.sgx

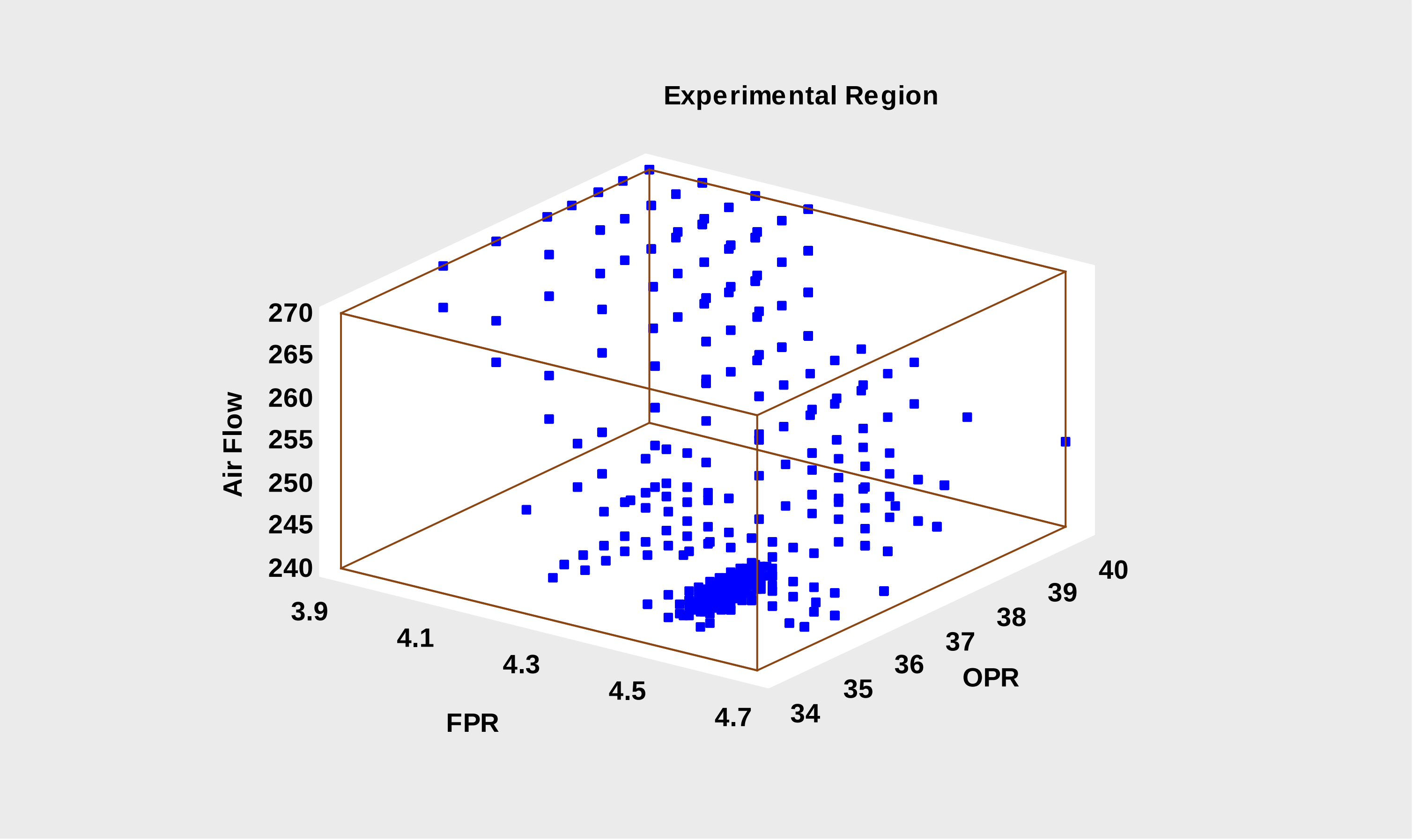
	BLOCK	FPR	OPR	Air Flow	Y1 Cost	Y2 Ps	Y3 Launch
1	1	4.46	34.6	241.0	5943.8	150.4	2071.42
2	1	4.46	34.8	242.0	5954.3	151.3	2085.9
3	1	4.46	35.0	243.0	5964.6	152.1	2100.39
4	1	4.3	37.0	260.0	6164.3	162.7	2275.9
5	1	4.3	37.0	260.0	6164.3	162.7	2275.9
6	1	4.48	35.0	242.0	5948.9	151.6	2092.56
7	1	4.48	34.8	241.0	5937.6	150.7	2078.14
8	1	4.5	34.8	240.0	5921.3	150.1	2069.9
9	1	4.5	35.0	241.0	5932.0	151.0	2084.42
10	1	4.34	34.2	246.0	6040.4	152.3	2100.5
11	1	4.34	34.2	246.0	6040.4	152.3	2100.5
12	1	4.38	35.4	250.0	6080.4	156.6	2173.03
13	1	4.38	35.4	250.0	6080.4	156.6	2173.03
14	1	4.5	35.2	242.0	5974.9	151.9	2099.51
15	1	4.48	35.2	243.0	5991.7	152.5	2107.57
16	1	3.9	36.0	270.0	6350.8	154.9	2169.72
17	1	3.9	36.0	270.0	6350.8	154.9	2169.72
18	1	4.48	35.4	243.0	5992.1	152.3	2105.71
19	1	4.5	35.4	242.0	5975.3	151.7	2097.43
20	1	4.5	35.4	242.0	5975.3	151.7	2097.43
21	1	4.5	35.4	242.0	5975.3	151.7	2097.43

engines | B | C

STATPOINT TECHNOLOGIES, INC.

Design Plot

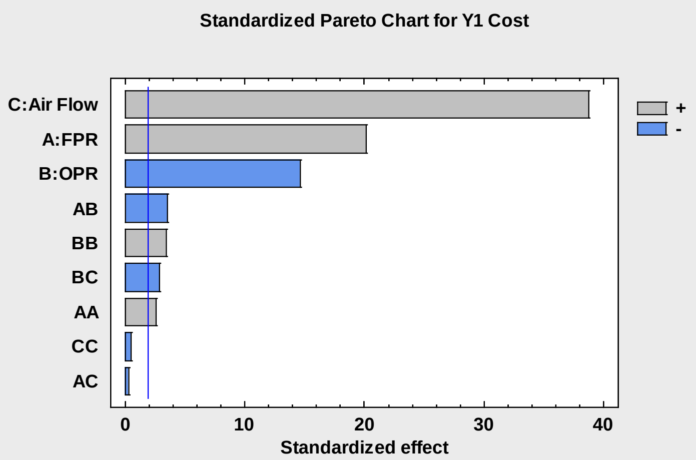
Shows the location of the historical data within the factor space.

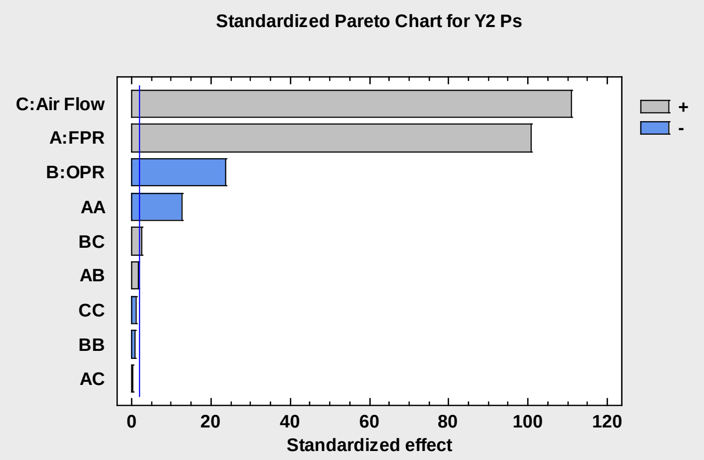


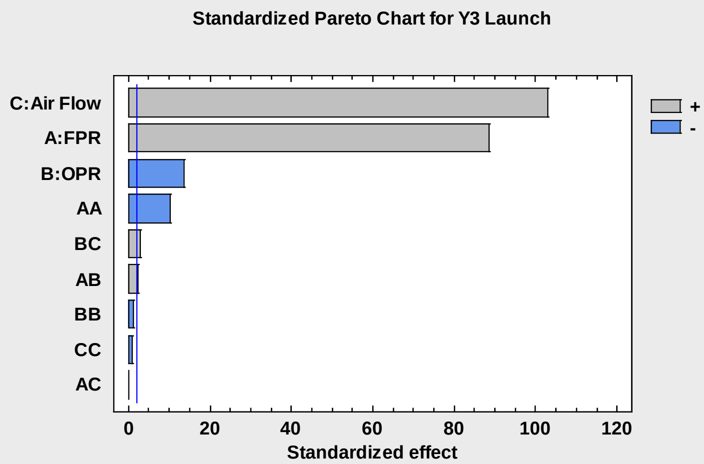
STATPOINT TECHNOLOGIES, INC.

Standardized Pareto Charts

Show the significant factors affecting each response.







STATPOINT TECHNOLOGIES, INC.

Desirability Function

Quantifies the desirability of a joint response (Y1, Y2, Y3).

For responses to be minimized:

$$d = \begin{cases} 1 & \hat{y} < low \\ \left(\dfrac{\hat{y} - high}{low - high} \right)^s, & low \le \hat{y} \le high \\ 0 & \hat{y} > high \end{cases}$$

For responses to be maximized:

$$d = \begin{cases} 0 & \hat{y} < low \\ \left(\dfrac{\hat{y} - low}{high - low} \right)^s, & low \le \hat{y} \le high \\ 1 & \hat{y} > high \end{cases}$$

Combined desirability:

D=d(Y1)*d(Y2)*d(Y3)

STATPOINT TECHNOLOGIES, INC.

Optimal Conditions

Found at the levels shown below:

Experimental Design Wizard

Step 1:Define responses | Step 3:Select design | Step 5:Select runs | Step 7:Save experiment | Step 9:Optimize responses | Step 11:Augment design

Step 2:Define exp. factors | Step 4:Specify model | Step 6:Evaluate design | Step 8:Analyze data | Step 10: Save results | Step 12:Extrapolate

Step 8: Analyze the experimental results

Model	*Y1 Cost*	*Y2 Ps*	*Y3 Launch*
Transformation	none	none	none
Model d.f.	9	9	9
P-value	0.0000	0.0000	0.0000
Error d.f.	574	574	574
Stnd. error	29.7747	0.447015	7.47643
R-squared	93.60	98.68	98.73
Adj. R-squared	93.50	98.66	98.71

Step 9: Optimize the responses

Response Values at Optimum

Response	*Optimized*	*Prediction*	*Lower 95.0% Limit*	*Upper 95.0% Limit*	*Desirability*
Y1 Cost	yes	6121.18	6108.08	6134.29	0.541166
Y2 Ps	yes	160.0	159.803	160.197	0.999998
Y3 Launch	yes	2243.17	2239.88	2246.46	0.886339

Optimized desirability = 0.782786

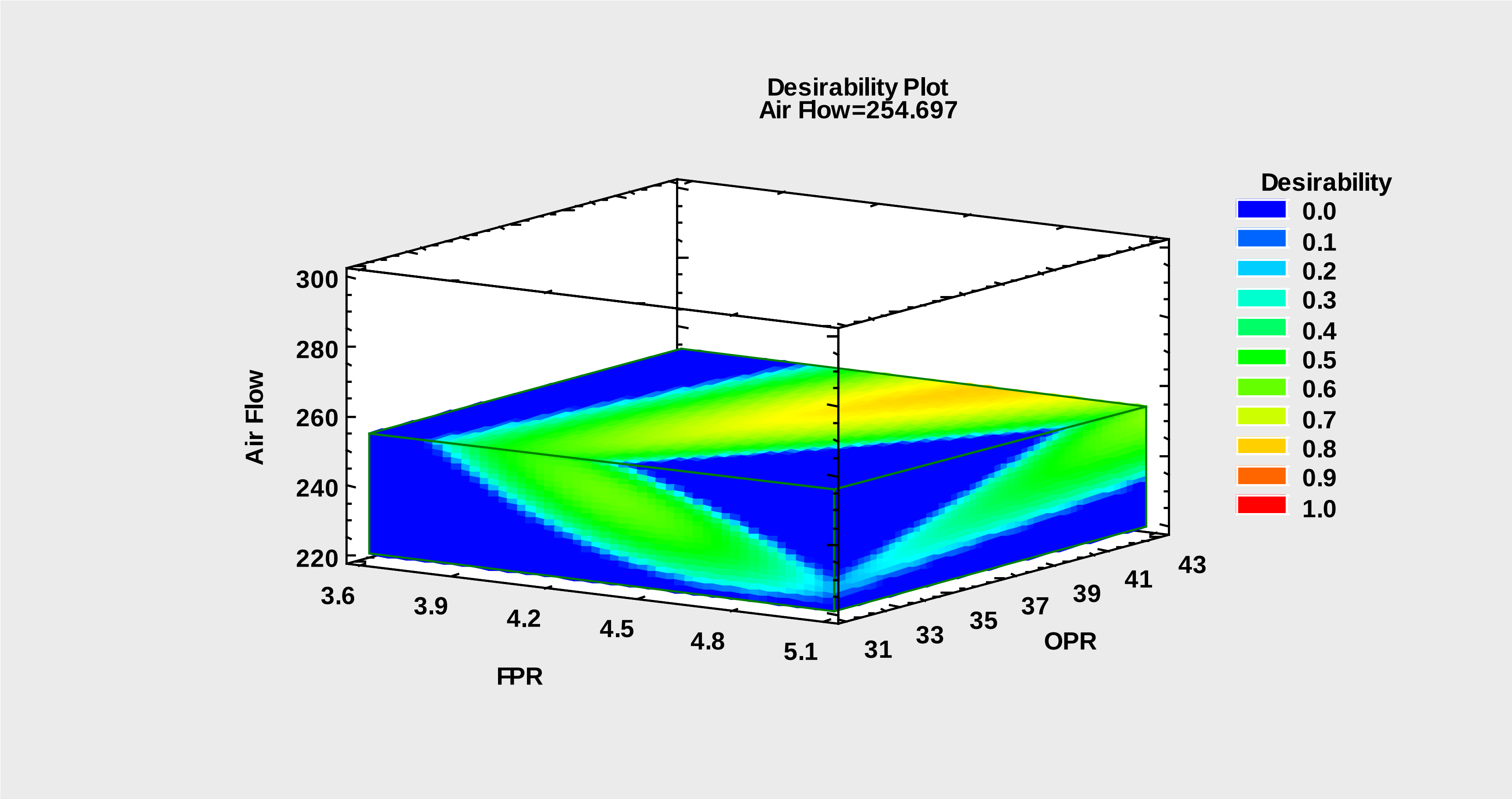
Factor Settings at Optimum

Factor	*Setting*
FPR	4.49812
OPR	40.0
Air Flow	254.697

STATPOINT TECHNOLOGIES, INC.

Response Surface

Show the estimated desirability throughout the experimental region.



STATPOINT TECHNOLOGIES, INC.

Problem #8: Time Series Data

- Data recorded at equally spaced points in time is called a *time series*.

- Time series models are used for various purposes:
  - Analysis of trends and seasonal effects
  - Forecasting
  - Control

- Autocorrelation between adjacent observations requires special models.

STATPOINT TECHNOLOGIES, INC.

Data file: customers.sgd (n=168)

customers.sgd

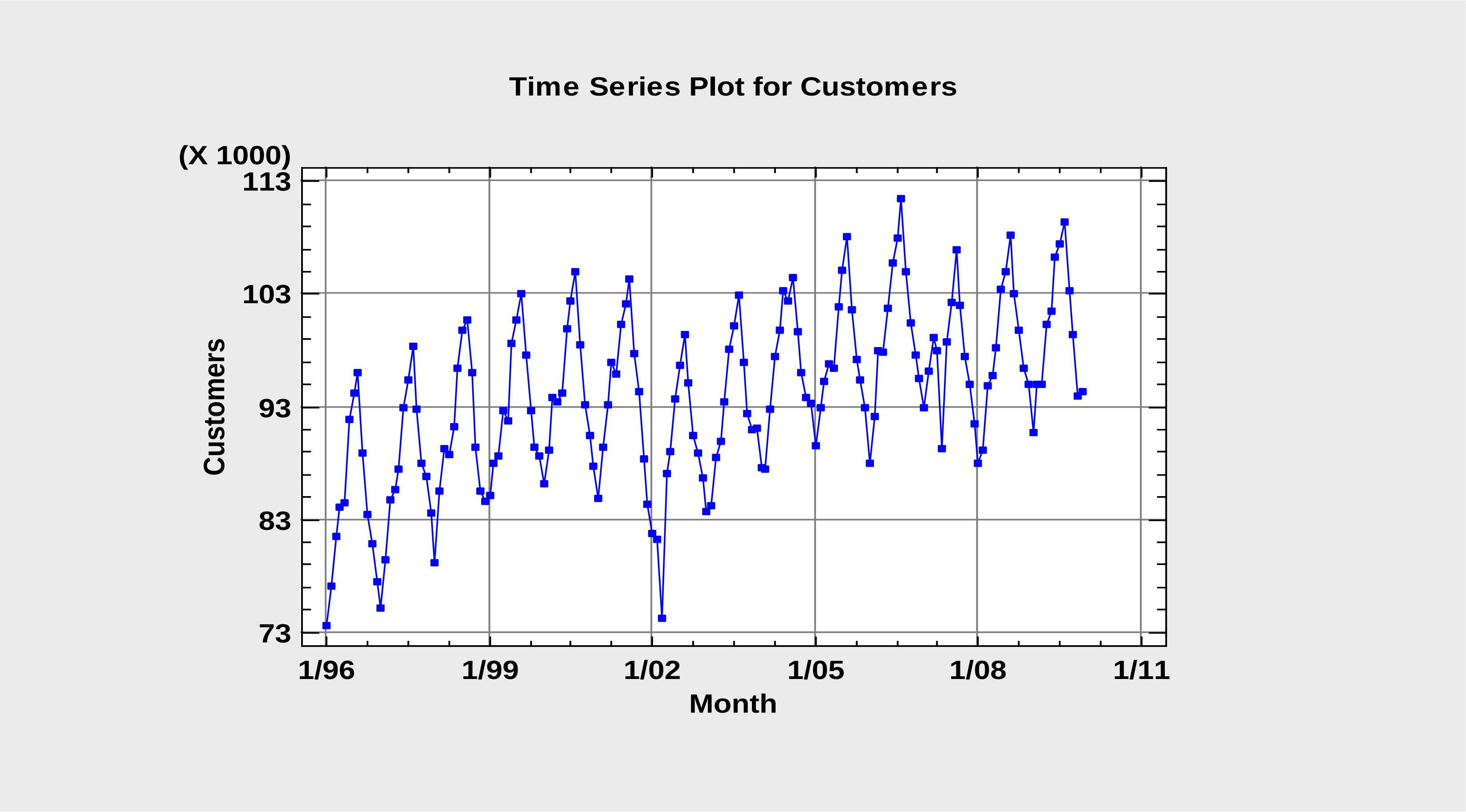
	Month	Customers	Col_3	Col_4
1	1/96	73637		
2	2/96	77136		
3	3/96	81481		
4	4/96	84127		
5	5/96	84562		
6	6/96	91959		
7	7/96	94174		
8	8/96	96087		
9	9/96	88952		
10	10/96	83479		
11	11/96	80814		
12	12/96	77466		
13	1/97	75225		
14	2/97	79418		
15	3/97	84813		
16	4/97	85691		
17	5/97	87490		

customers / B / C / D / E / F / G / H / I

STATPOINT TECHNOLOGIES, INC.

Time Sequence Plot

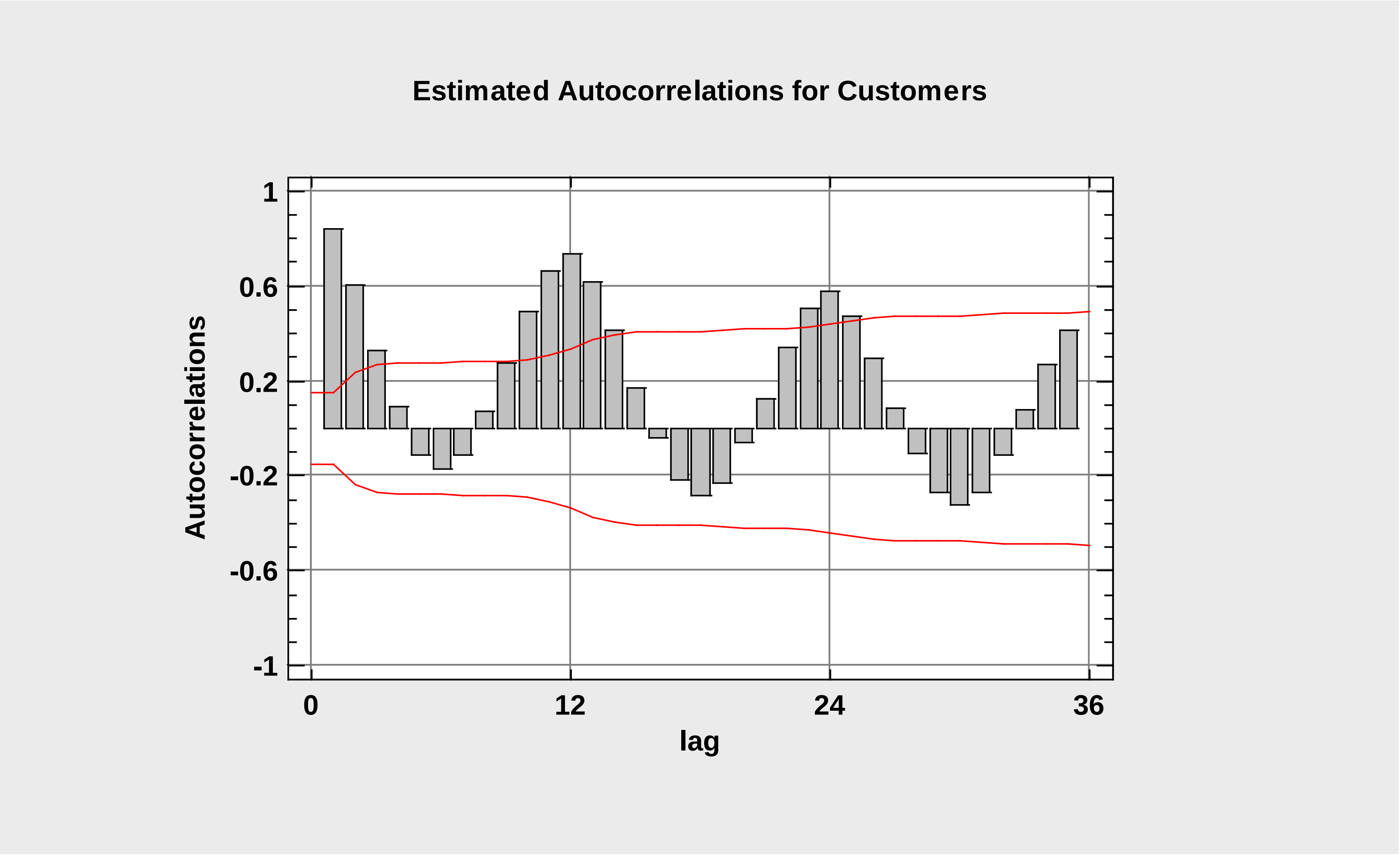
Plots the data versus time.



STATPOINT TECHNOLOGIES, INC.

Autocorrelation Function

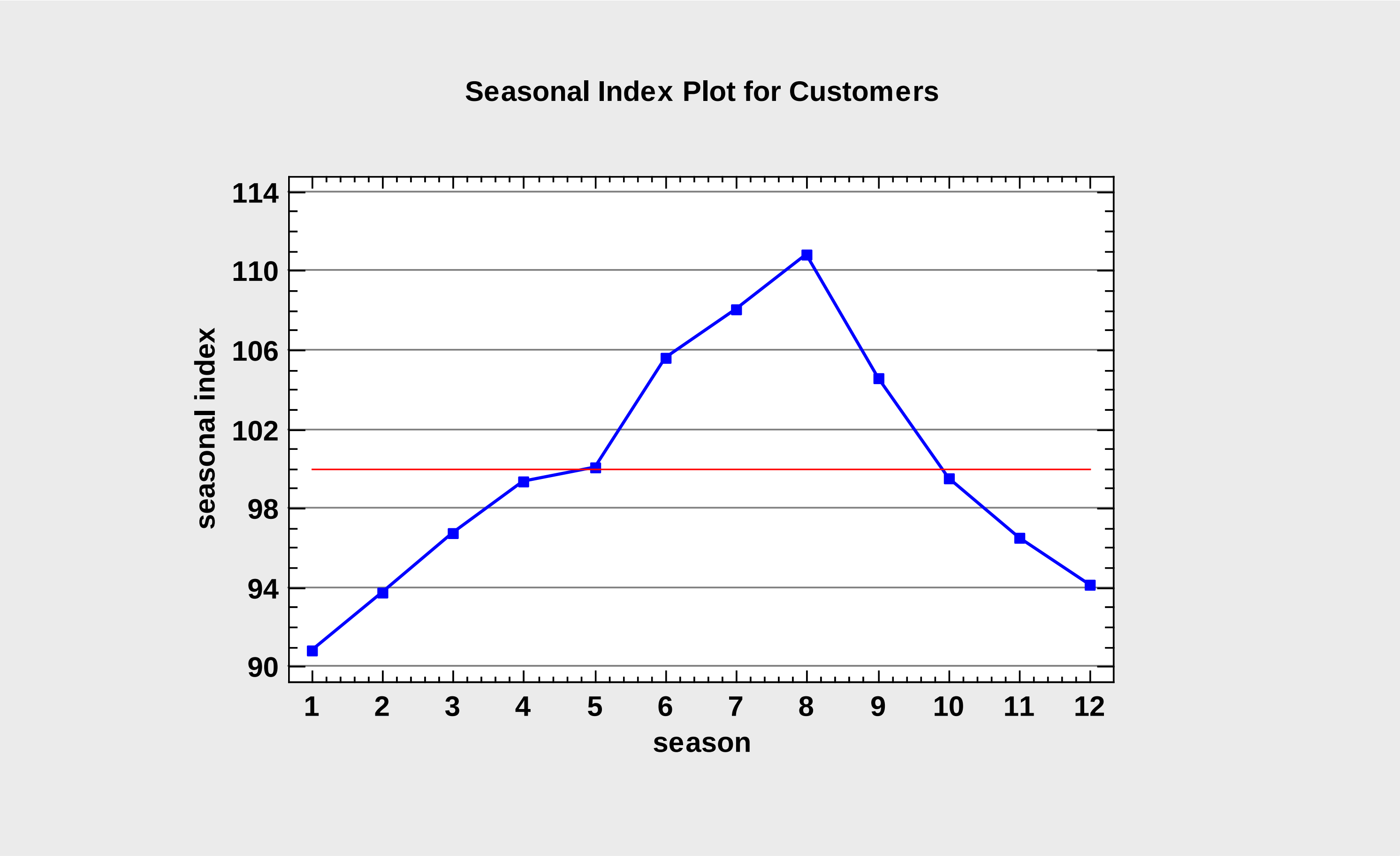
Estimates the correlation between observations at different lags.



STATPOINT TECHNOLOGIES, INC.

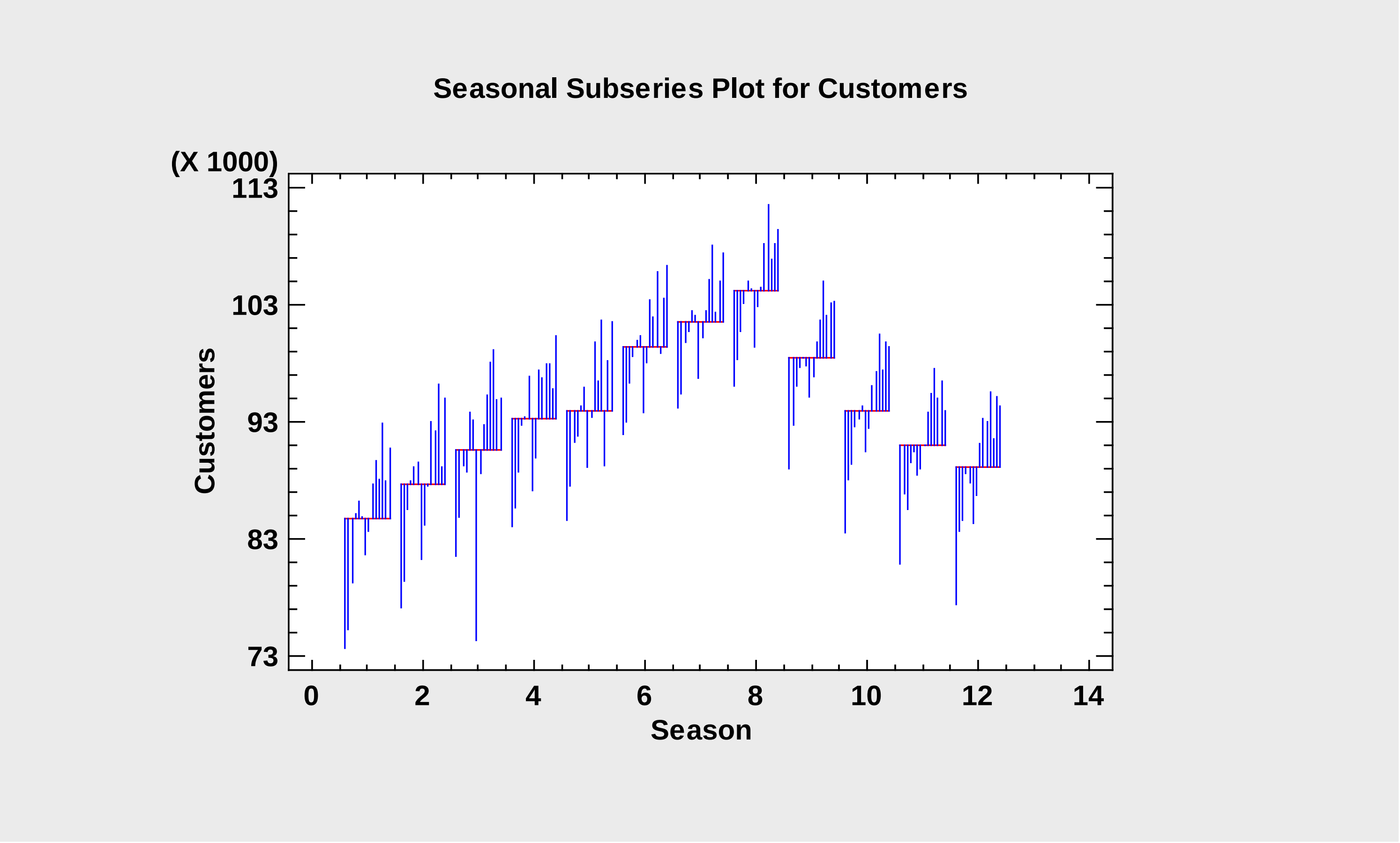
Seasonal Decomposition

Shows the average value during each season (scaled to 100).



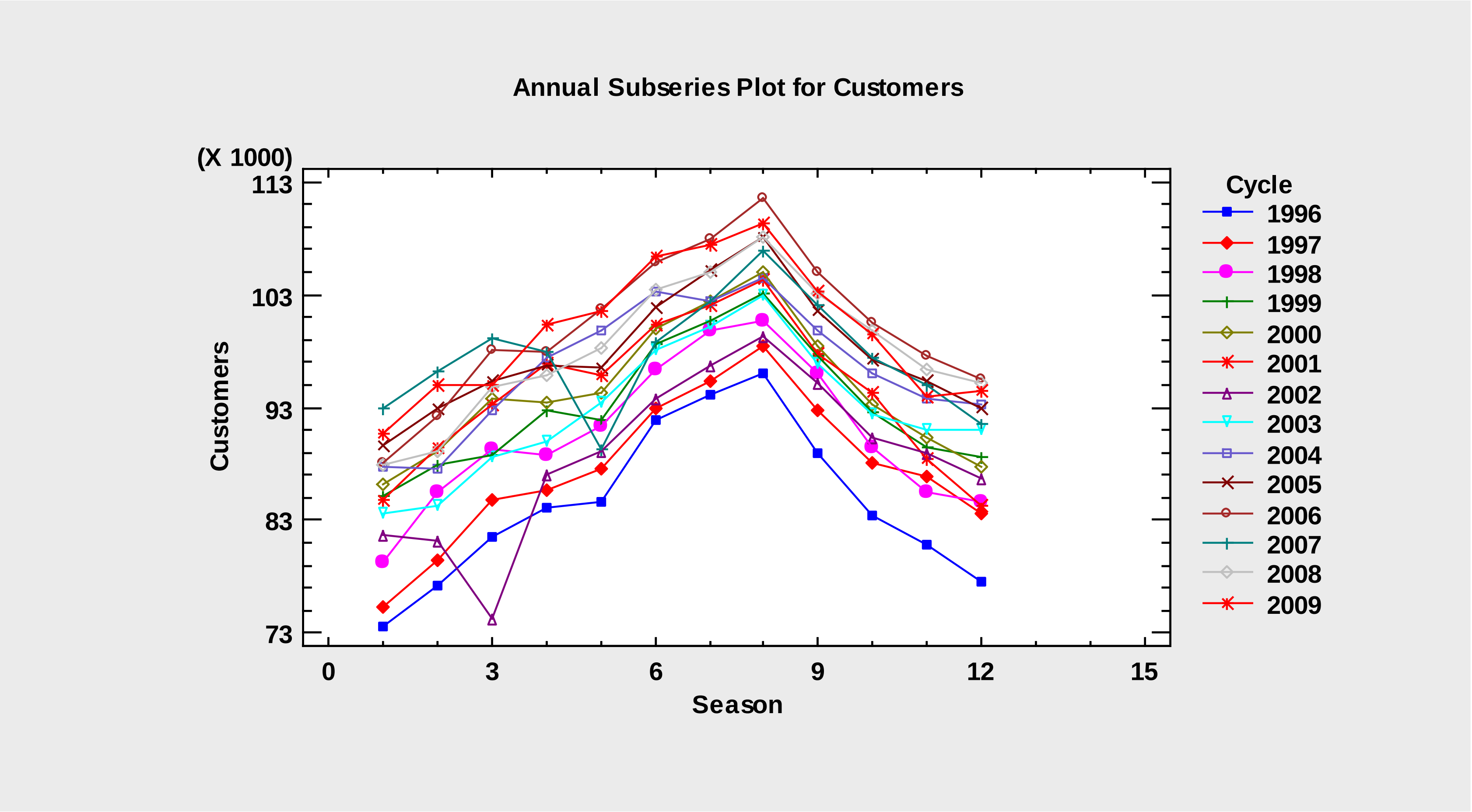
Seasonal Subseries Plot

Shows the seasonal averages and trend within each season.



Annual Subseries Plot

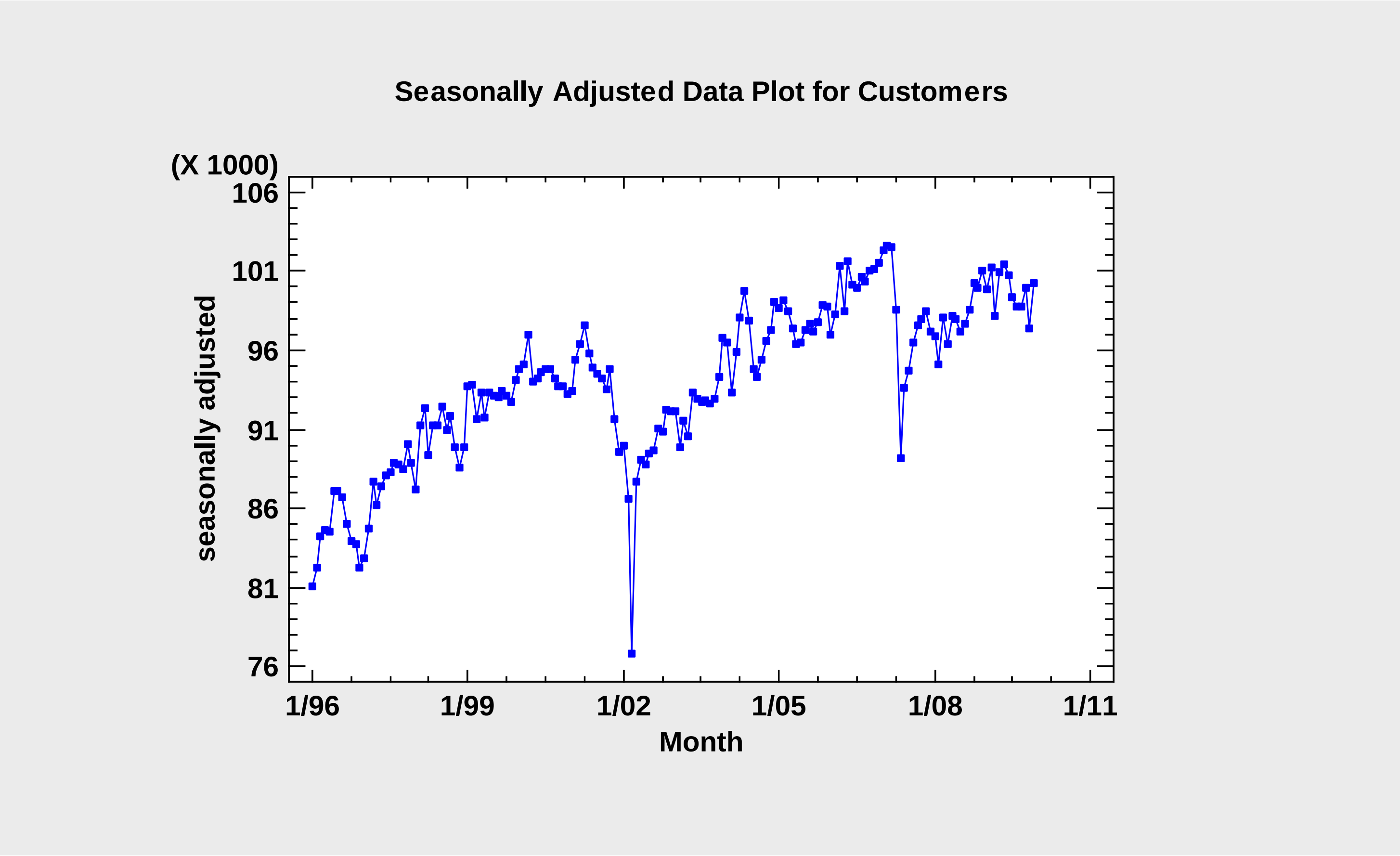
Shows the seasonal effect separately for each cycle.



STATPOINT TECHNOLOGIES, INC.

Seasonally Adjusted Data

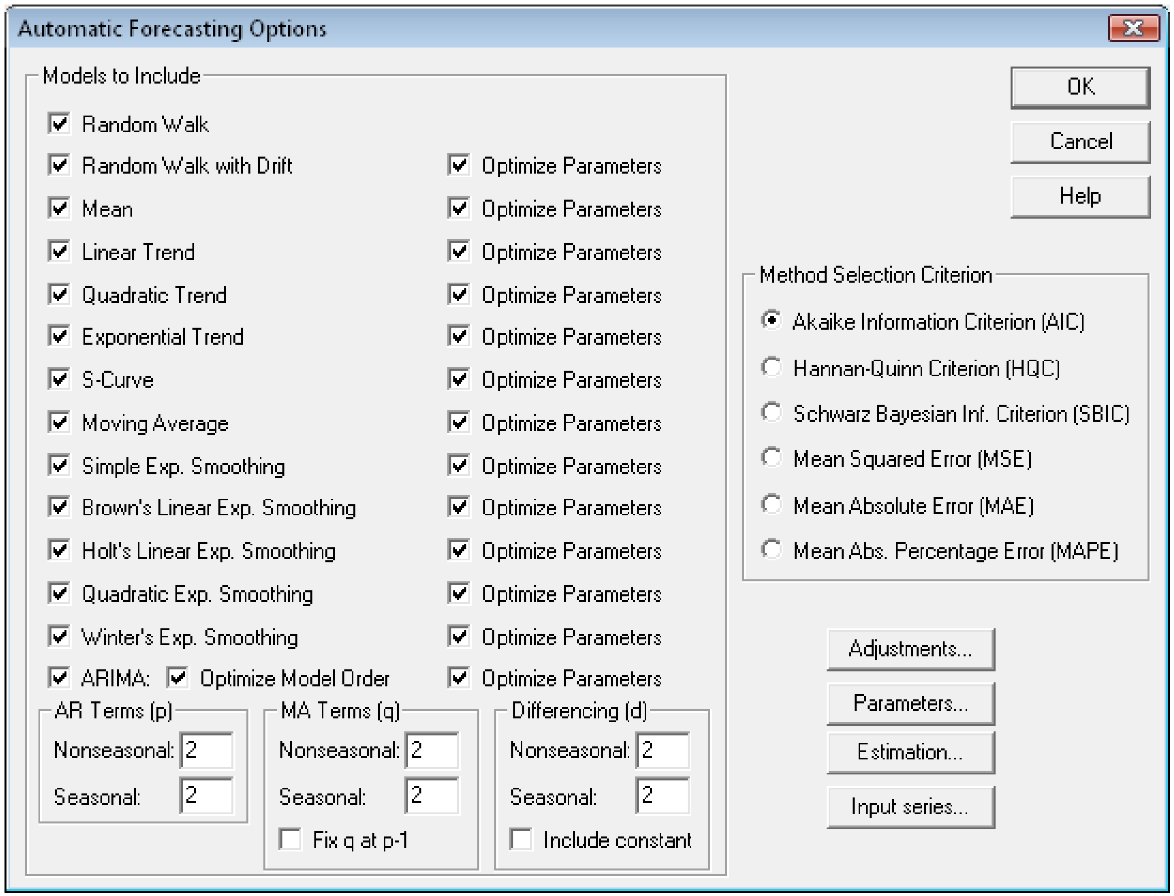
Removes the seasonal effects from the data.



STATPOINT TECHNOLOGIES, INC.

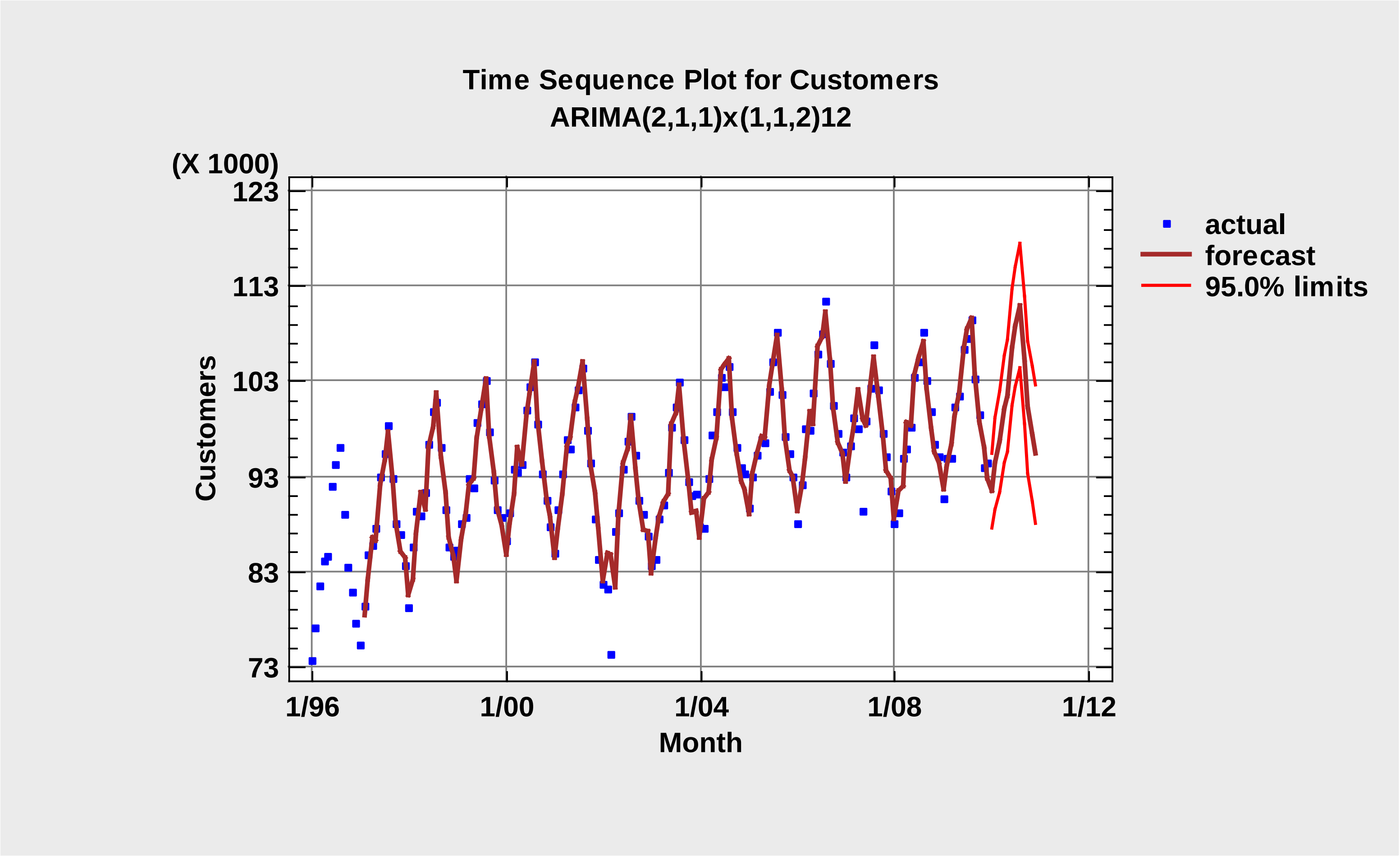
Automatic Forecasting

Fits many models and automatically selects the best.



ARIMA Model

Selected model is a rather complicated seasonal ARIMA model.



STATPOINT TECHNOLOGIES, INC.

Problem #9: Event Rate Modeling

- When the data to be analyzed consist of the time at which events occur, the process by which those events are generated is called a *point process*.

- The most common type of point process is a *Poisson process*, in which the times between events are independent and follow a negative exponential distribution.

- If the event rate is constant, the process is called *homogeneous*. If the event rate changes over time, then the process is called *nonhomogeneous*.

STATPOINT TECHNOLOGIES, INC.

Data file: earthquakes.sgd (n=52)

C:\Data\Six Sigma\earthquakes.sgd

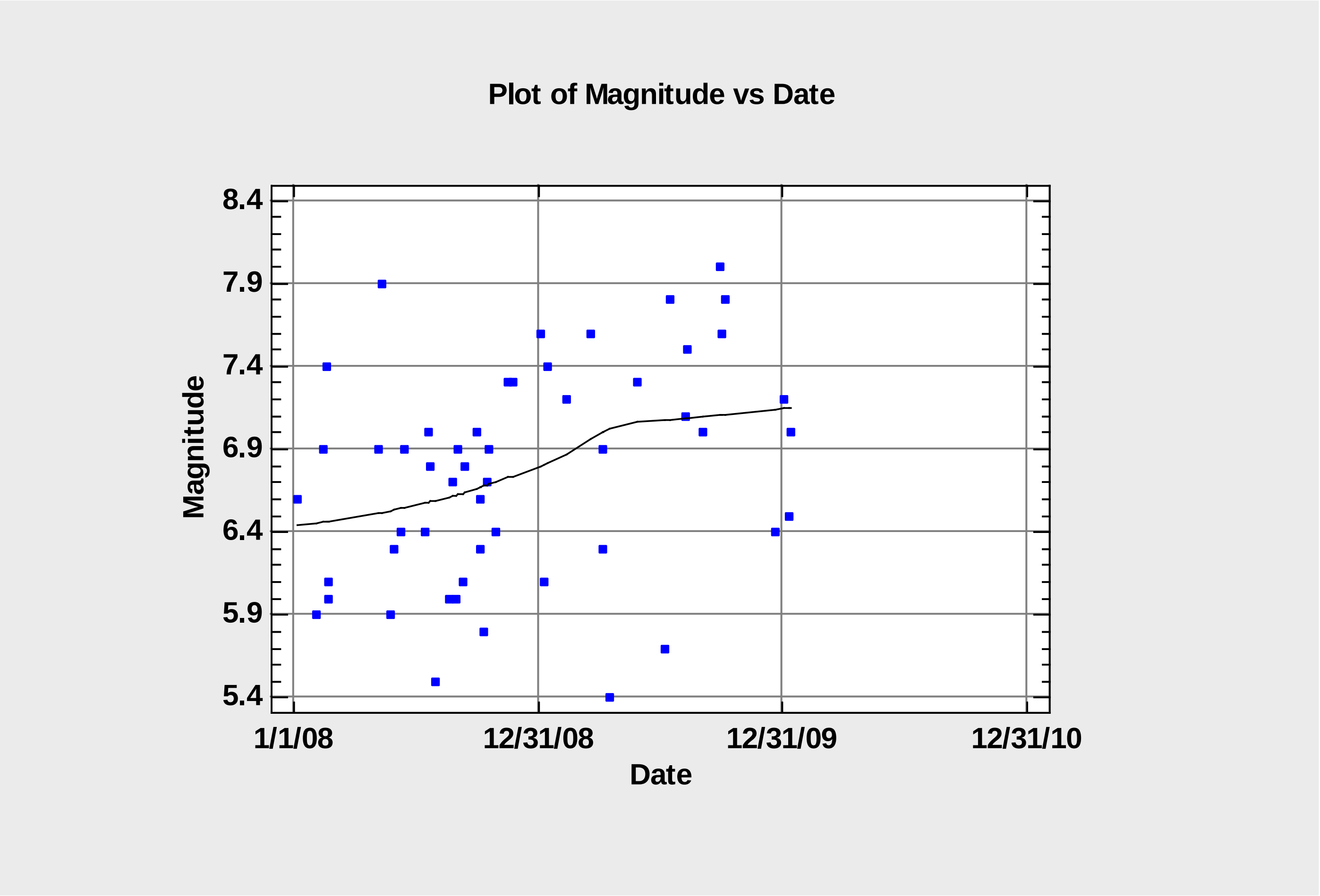
	Date	Location	Magnitude	C
36	3/19/09	Tonga	7.6	
37	4/6/09	Italy	6.3	
38	4/7/09	Kuril Islands	6.9	
39	4/16/09	Afganistan	5.4	
40	5/28/09	Honduras	7.3	
41	7/9/09	China	5.7	
42	7/15/09	New Zealand	7.8	
43	8/9/09	Japan	7.1	
44	8/10/09	Indian Ocean	7.5	
45	9/2/09	Indonesia	7.0	
46	9/29/09	Samoa	8.0	
47	9/30/09	Indonesia	7.6	
48	10/7/09	Vanuatu	7.8	
49	12/19/09	Taiwan	6.4	
50	1/3/10	Solomon Islands	7.2	
51	1/10/10	California	6.5	
52	1/12/10	Haiti	7.0	

earthquakes B C

STATPOINT TECHNOLOGIES, INC.

Plot of Magnitude Versus Date

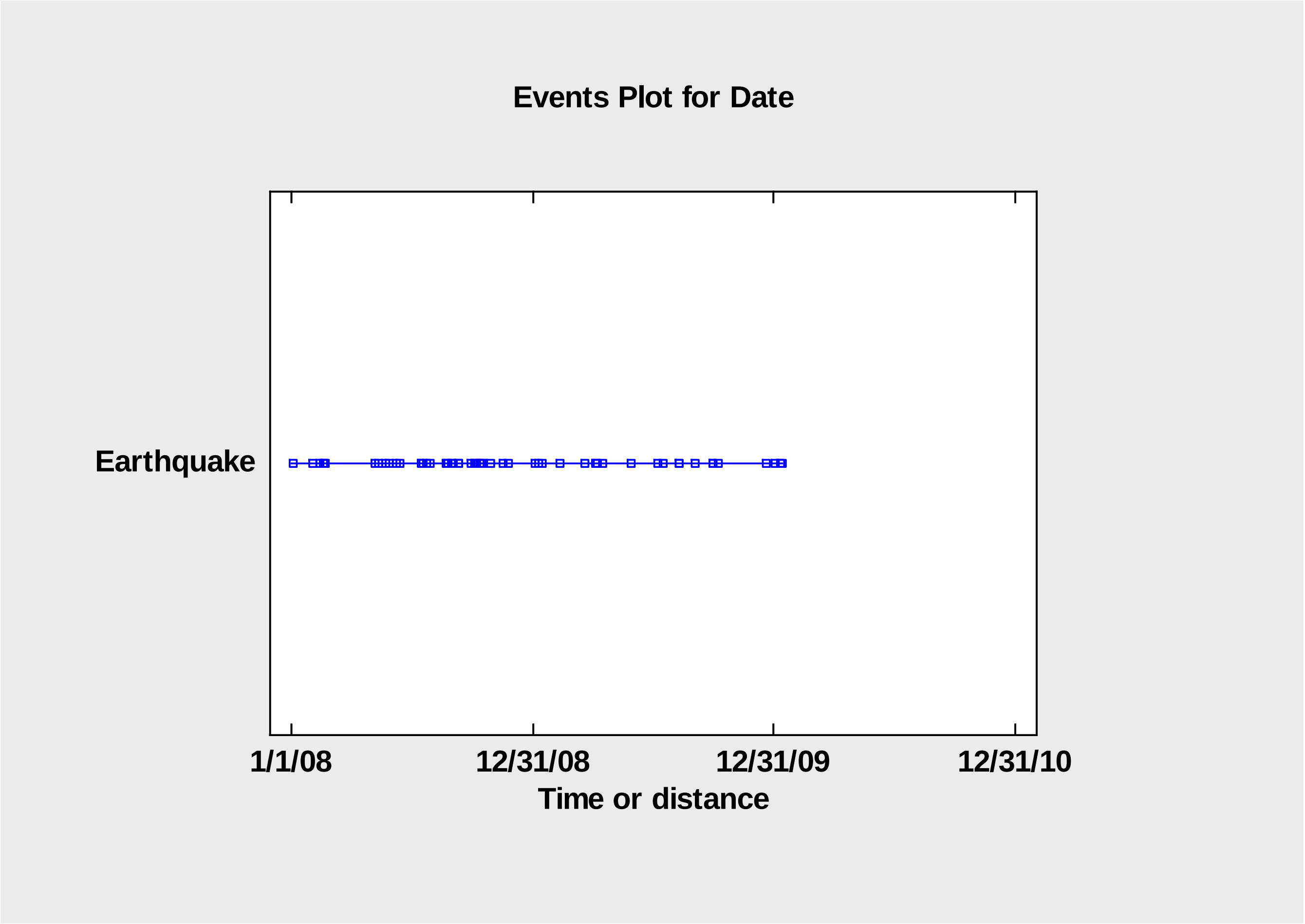
Shows an apparent increase in magnitude over the sampling period.



STATPOINT TECHNOLOGIES, INC.

Point Process Plot

Shows the dates of occurrence only.



Event Rate

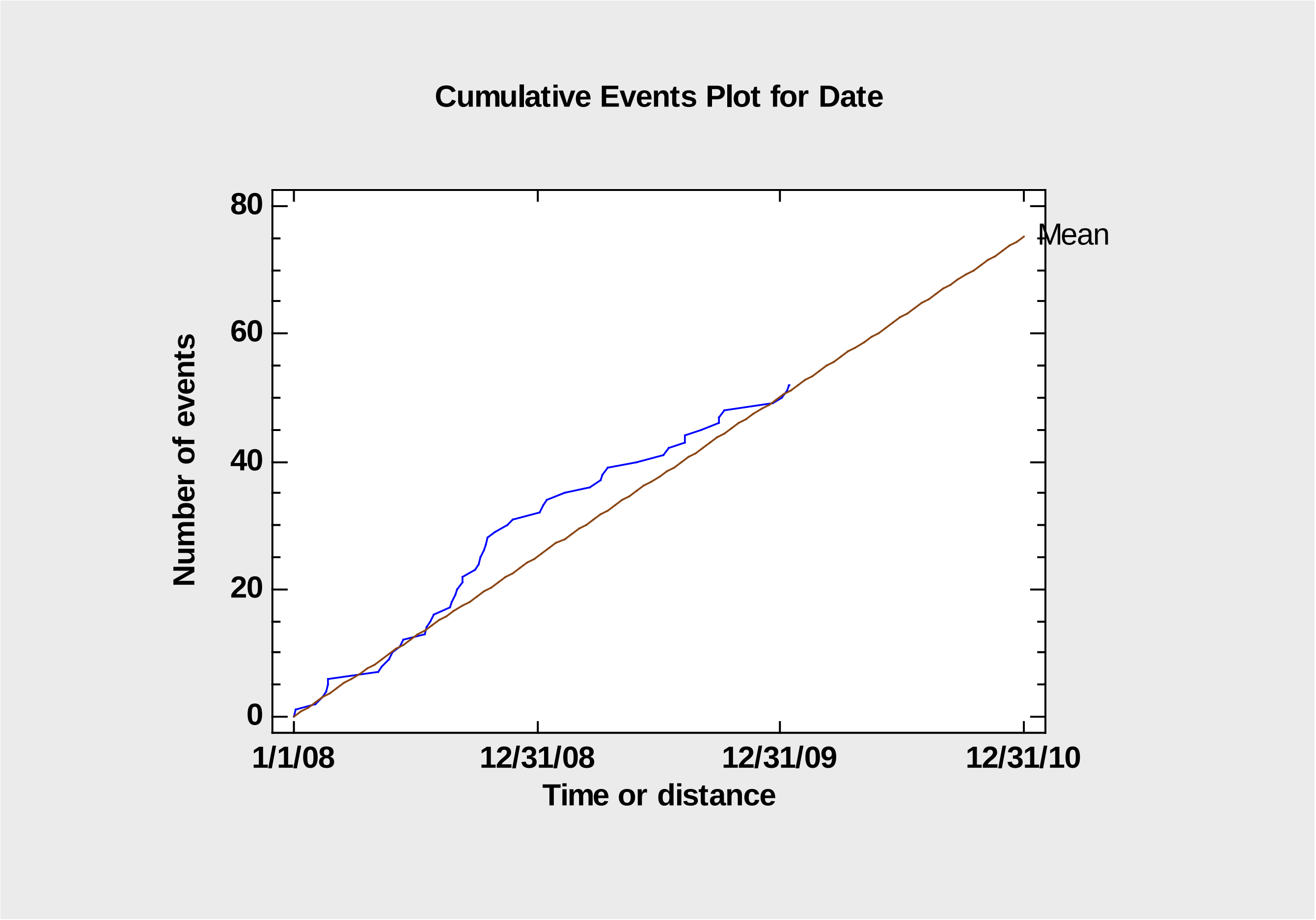
- The critical parameter in a point process is the rate of events per unit time (such as earthquakes per year). This parameter is usually called λ.

- The rate parameter is also related to the mean time between events, with MTBE = $1 / \lambda$.

- λ may be constant (a homogeneous process) or vary over time (a nonhomogenous process).

Cumulative Events Plot

The slope of the line is related to the event rate.



Trend Test

A small P-value would indicate a significant trend.

One Dimensional Point Processes

Trend Test

Laplace Centroid Test

	Test statistic	*P-Value*
Date	-0.962298	0.335899

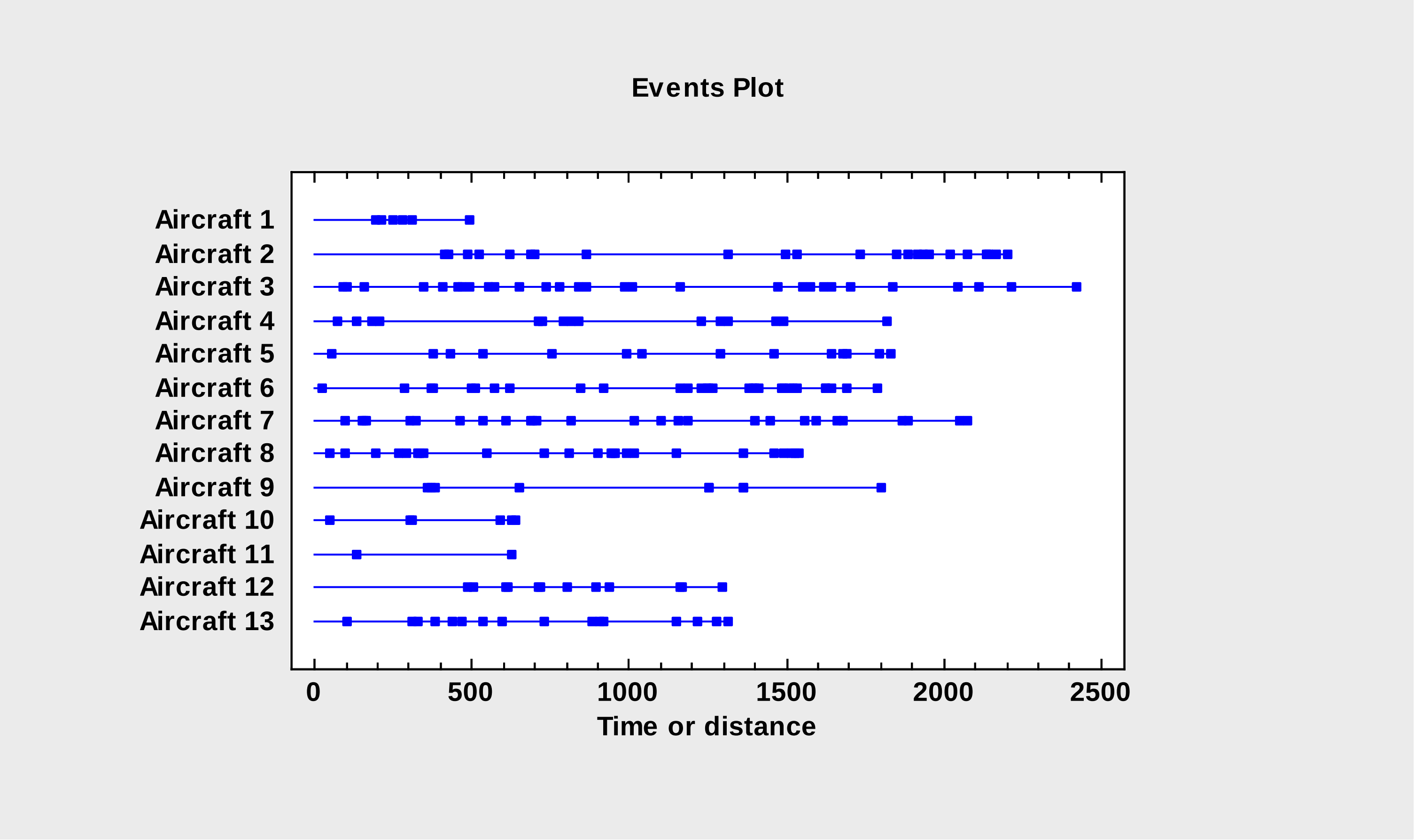
The StatAdvisor

The above table displays a test for trend. The Laplace Centroid Test tests whether the event times are uniformly distributed over the sampling interval. A small P-value indicates the presence of a trend. For example, since the P-value calculated when testing Date is greater than or equal to 0.05, there is not a statistically significant trend at the 5.0% significance level.

STATPOINT TECHNOLOGIES, INC.

Other Uses

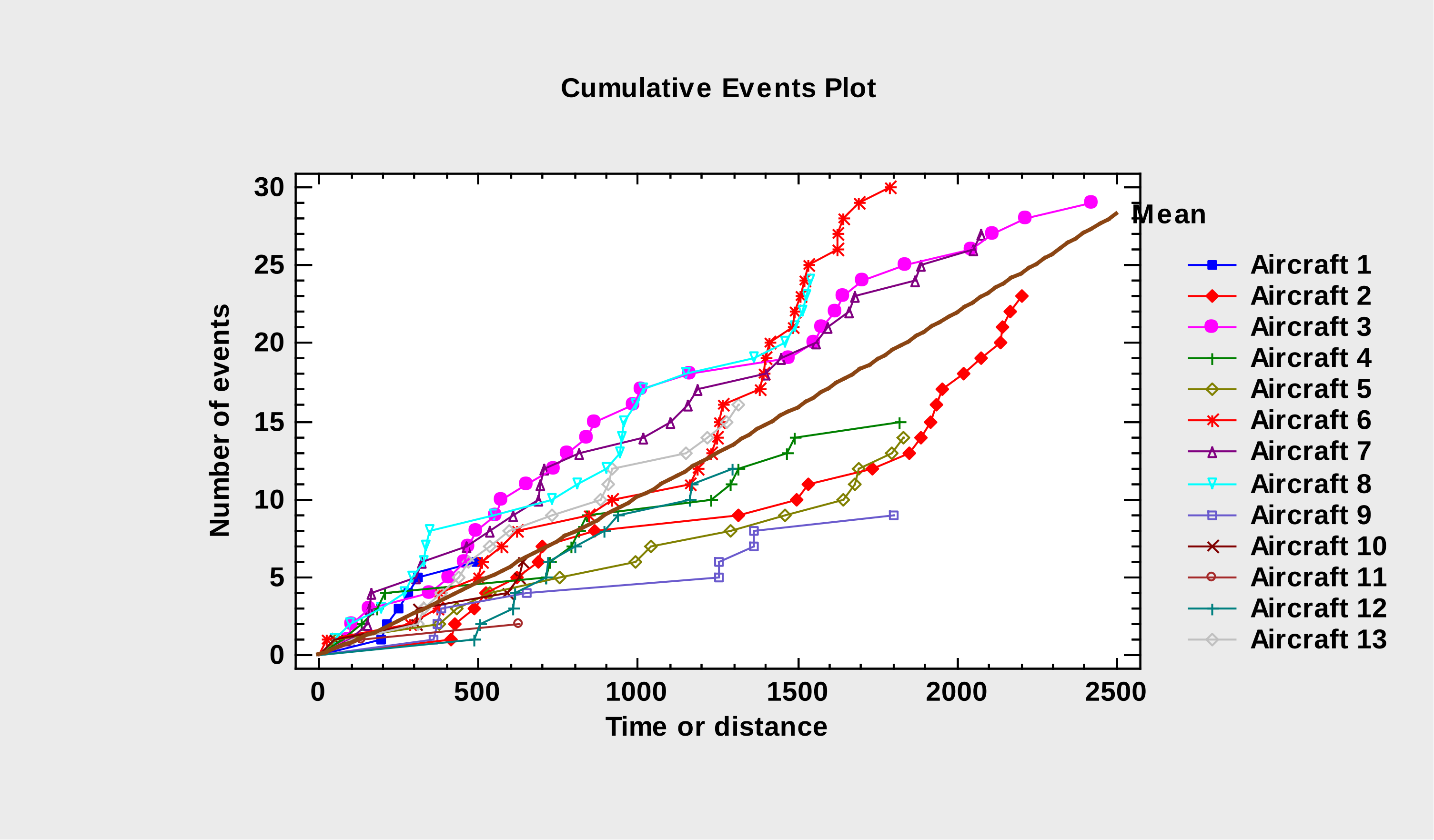
Point process models are also very useful for estimating failure rates.



STATPOINT TECHNOLOGIES, INC.

Between Group Comparisons

Tests can be made to determine whether there are significant differences.



Problem #10: Interactive Maps

- Graphics that allow the user to interact with the data are extremely useful.

- Maps are one important example.

STATPOINT TECHNOLOGIES, INC.

Data file: census2000.sgd (n=51)

C:\DocData16\census2000.sgd

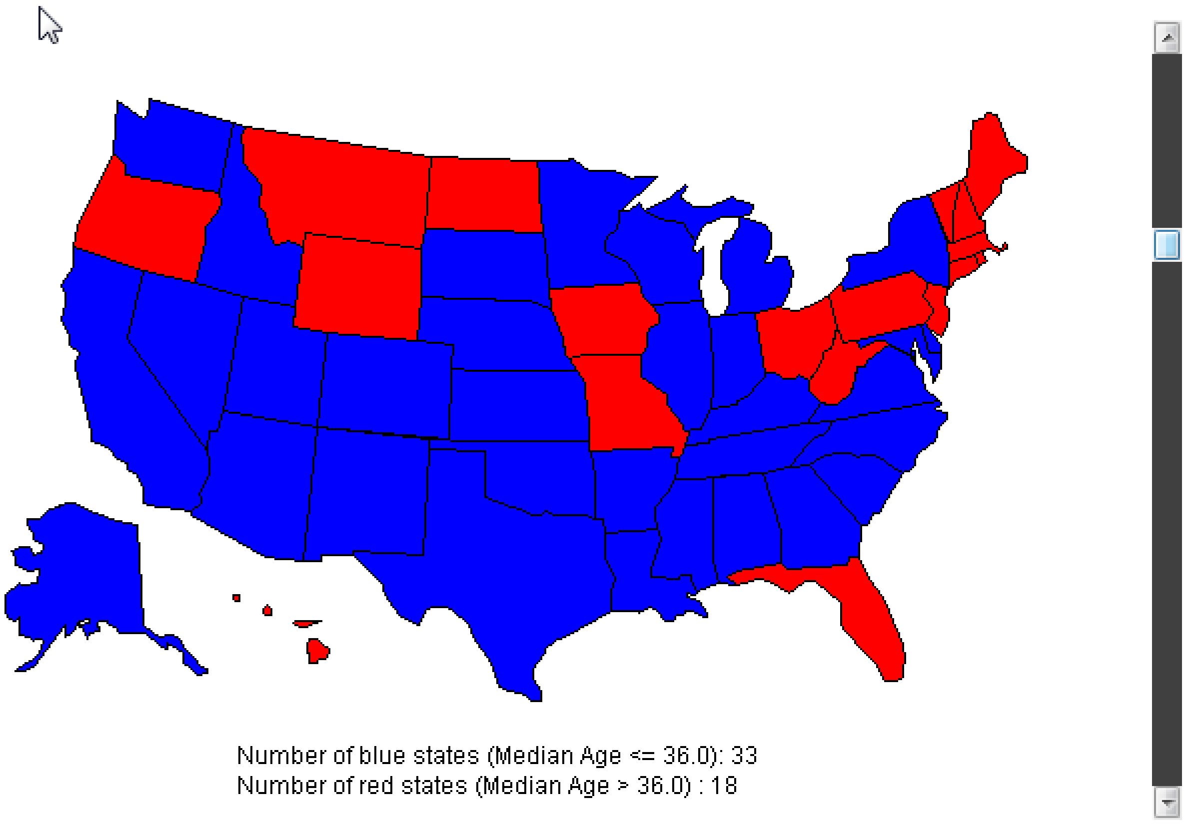
	State	Population	Median Age	Percent Female	Per Capita Income
1	Alabama	4447100	35.8	51.7	18819
2	Alaska	626932	32.4	48.3	22660
3	Arizona	5130632	34.2	50.1	20275
4	Arkansas	2673400	36	51.2	16904
5	California	33871648	33.3	50.2	22711
6	Colorado	4301261	34.3	49.6	24049
7	Connecticut	3405565	37.4	51.6	28766
8	Delaware	783600	36	51.4	23305
9	D.C.	572059	34.6	52.9	28659
10	Florida	15982378	38.7	51.2	21557
11	Georgia	8186453	33.4	50.8	21154
12	Hawaii	1211537	36.2	49.8	21525
13	Idaho	1293953	33.2	49.9	17841
14	Illinois	12419293	34.7	51	23104
15	Indiana	6080485	35.2	51	20397
16	Iowa	2926324	36.6	50.9	19674
17	Kansas	2688418	35.2	50.6	20506
18	Kentucky	4041769	35.9	51.1	18093
19	Louisiana	4468976	34	51.6	16912
20	Maine	1274923	38.6	51.3	19533
21	Maryland	5296486	36	51.7	25614
22	Massachusetts	6349097	36.5	51.8	25952

census2000 B C

STATPOINT TECHNOLOGIES, INC.

U.S. Map Statlet

The slider changes the cutoff between the red and blue states.



STATPOINT TECHNOLOGIES, INC.