

Overview of Data Analysis Using Statgraphics Centurion

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

Copyright 2011 by StatPoint Technologies, Inc.

Web site: www.statgraphics.com



STATPOINT
TECHNOLOGIES, INC.

What is Statgraphics?

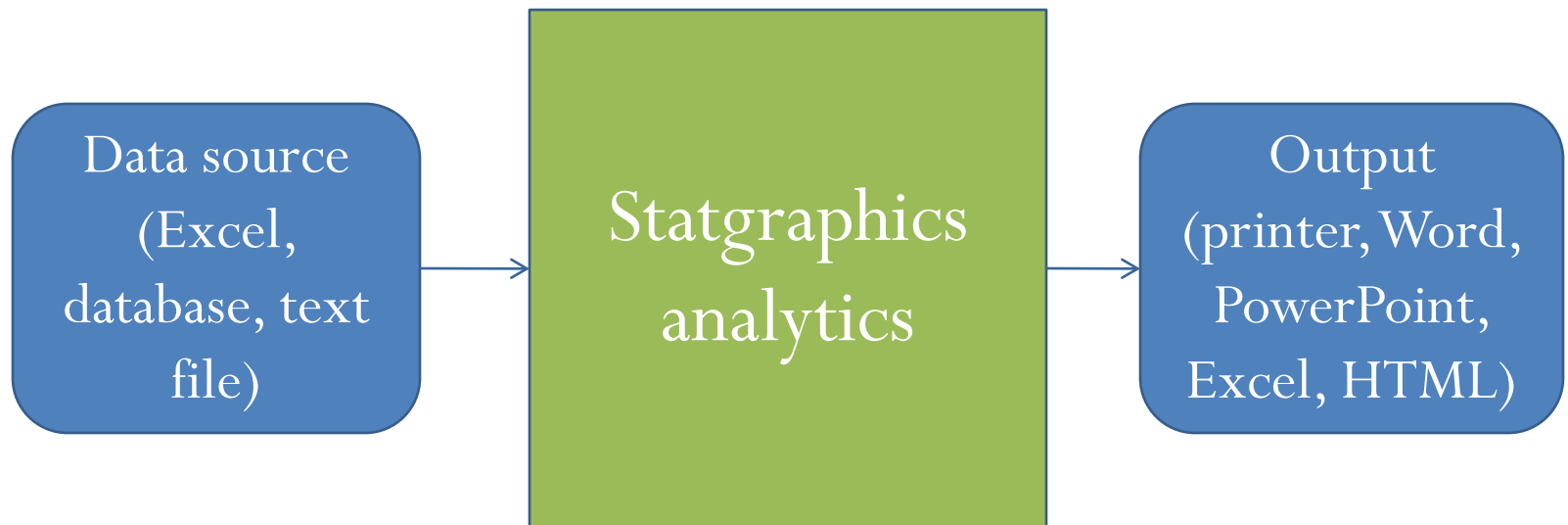
Statgraphics is a data analysis and data visualization program that runs as a standalone application under Microsoft Windows.

It is designed to make it easy to take data from various data sources (such as Excel or databases) and extract the important information from that data.

Short History

- First version of Statgraphics for PC's was released in 1983. Earliest major stat package for PCs and the first to incorporate graphics directly into every statistical procedure.
- Current version is Version 16 with more than 170 procedures.
- Desktop application now called “Statgraphics Centurion” to distinguish it from early versions marketed by Manugistics and from our “Statgraphics Online” product.

Data Analysis Model



Types of Data Analysis

- **Exploratory data analysis (data mining)** – not sure ahead of time what methods will be needed.
- **Repetitive data processing** – repeats a pre-selected set of methods on different data sets or at different times.



Example: Countries of the world

Countries of the world.xlsx - Microsoft Excel

Home Insert Page Layout Formulas Data Review View Add-Ins Acrobat

Clipboard Font Alignment Number Styles Cells Editing

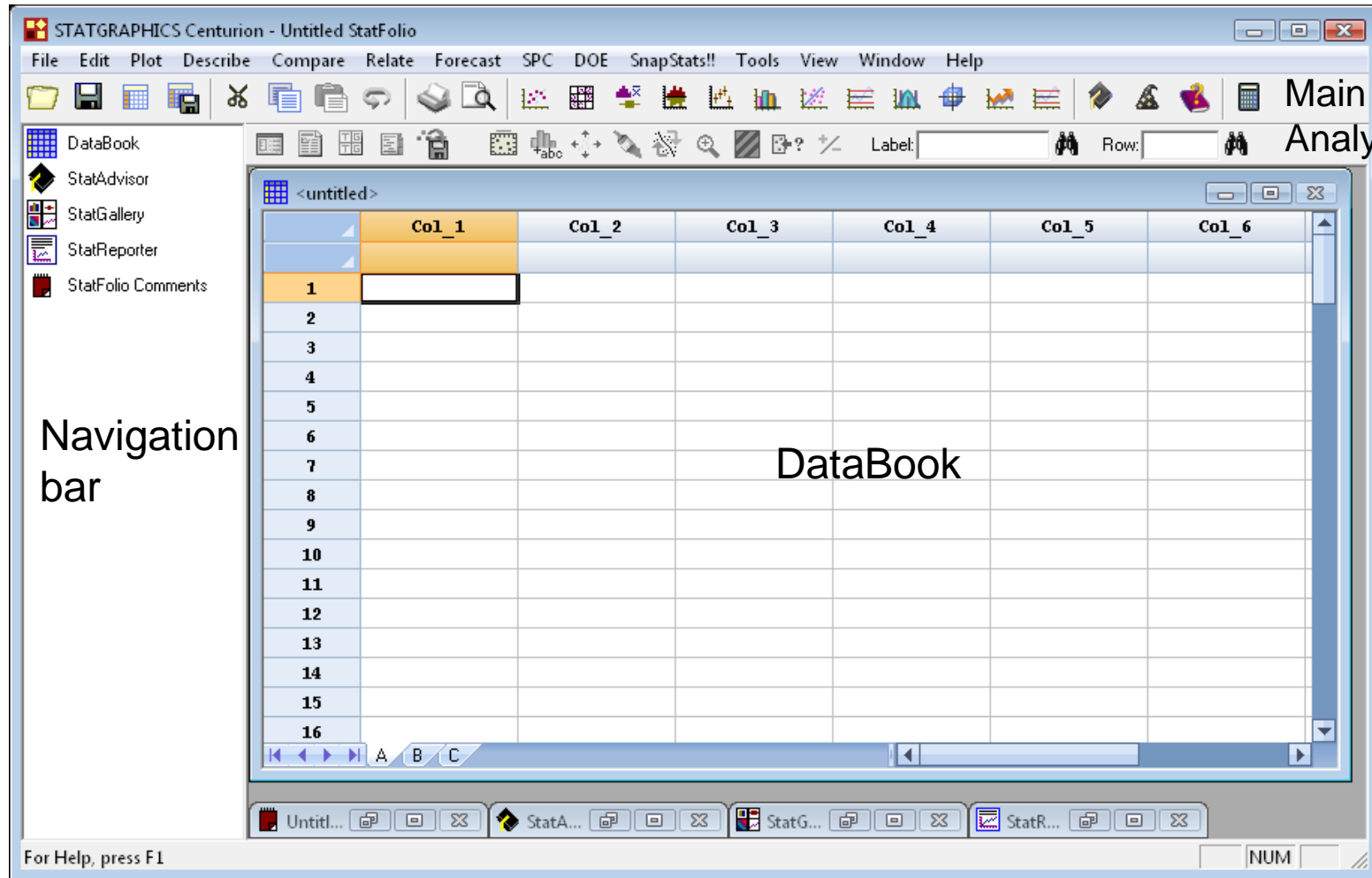
A1 Country

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Country	Region	Population	Area	Pop. Density	Coastline	Net migration	Infant mortality	GDP	Literacy	Phones	Arable	Crops
2				sq. mi.	per sq. mi.	coast/area ratio		per 1000 births	\$ per capita	%	per 1000	%	%
3	Afghanistan	ASIA (EXCLUDING NEA	31056997	647500	48.0	0.00	23.06	163.07	700	36.0	3.2	12.13	0.22
4	Albania	EASTERN EUROPE	3581655	28748	124.6	1.26	-4.93	21.52	4500	86.5	71.2	21.09	4.42
5	Algeria	NORTHERN AFRICA	32930091	2381740	13.8	0.04	-0.39	31	6000	70.0	78.1	3.22	0.25
6	American Samoa	OCEANIA	57794	199	290.4	58.29	-20.71	9.27	8000	97.0	259.5	10	15
7	Andorra	WESTERN EUROPE	71201	468	152.1	0.00	6.6	4.05	19000	100.0	497.2	2.22	0
8	Angola	SUB-SAHARAN AFRICA	12127071	1246700	9.7	0.13	0	191.19	1900	42.0	7.8	2.41	0.24
9	Anguilla	LATIN AMERICA AND	13477	102	132.1	59.80	10.76	21.03	8600	95.0	460.0	0	0
10	Antigua and Barbuda	LATIN AMERICA AND	69108	443	156.0	34.54	-6.15	19.46	11000	89.0	549.9	18.18	4.55
11	Argentina	LATIN AMERICA AND	39921833	2766890	14.4	0.18	0.61	15.18	11200	97.1	220.4	12.31	0.48
12	Armenia	COMMONWEALTH OF	2976372	29800	99.9	0.00	-6.47	23.28	3500	98.6	195.7	17.55	2.3
13	Aruba	LATIN AMERICA AND	71891	193	372.5	35.49	0	5.89	28000	97.0	516.1	10.53	0
14	Australia	OCEANIA	20264082	7686850	2.6	0.34	3.98	4.69	29000	100.0	565.5	6.55	0.04
15	Austria	WESTERN EUROPE	8192880	83870	97.7	0.00	2	4.66	30000	98.0	452.2	16.91	0.86
16	Azerbaijan	COMMONWEALTH OF	7961619	86600	91.9	0.00	-4.9	81.74	3400	97.0	137.1	19.63	2.71
17	Bahamas, The	LATIN AMERICA AND	303770	13940	21.8	25.41	-2.2	25.21	16700	95.6	460.6	0.8	0.4
18	Bahrain	NEAR EAST	698585	665	1050.5	24.21	1.05	17.27	16900	89.1	281.3	2.82	5.63
19	Bangladesh	ASIA (EXCLUDING NEA	147365352	144000	1023.4	0.40	-0.71	62.6	1900	43.1	7.3	62.11	3.07
20	Barbados	LATIN AMERICA AND	279912	431	649.5	22.51	-0.31	12.5	15700	97.4	481.9	37.21	2.33
21	Belarus	COMMONWEALTH OF	10293011	207600	49.6	0.00	2.54	13.37	6100	99.6	319.1	29.55	0.6
22	Belgium	WESTERN EUROPE	10379067	30528	340.0	0.22	1.23	4.68	29100	98.0	462.6	23.28	0.4
23	Belize	LATIN AMERICA AND	287730	22966	12.5	1.68	0	25.69	4900	94.1	115.7	2.85	1.71

Sheet1 Sheet2 Sheet3

Ready 100%

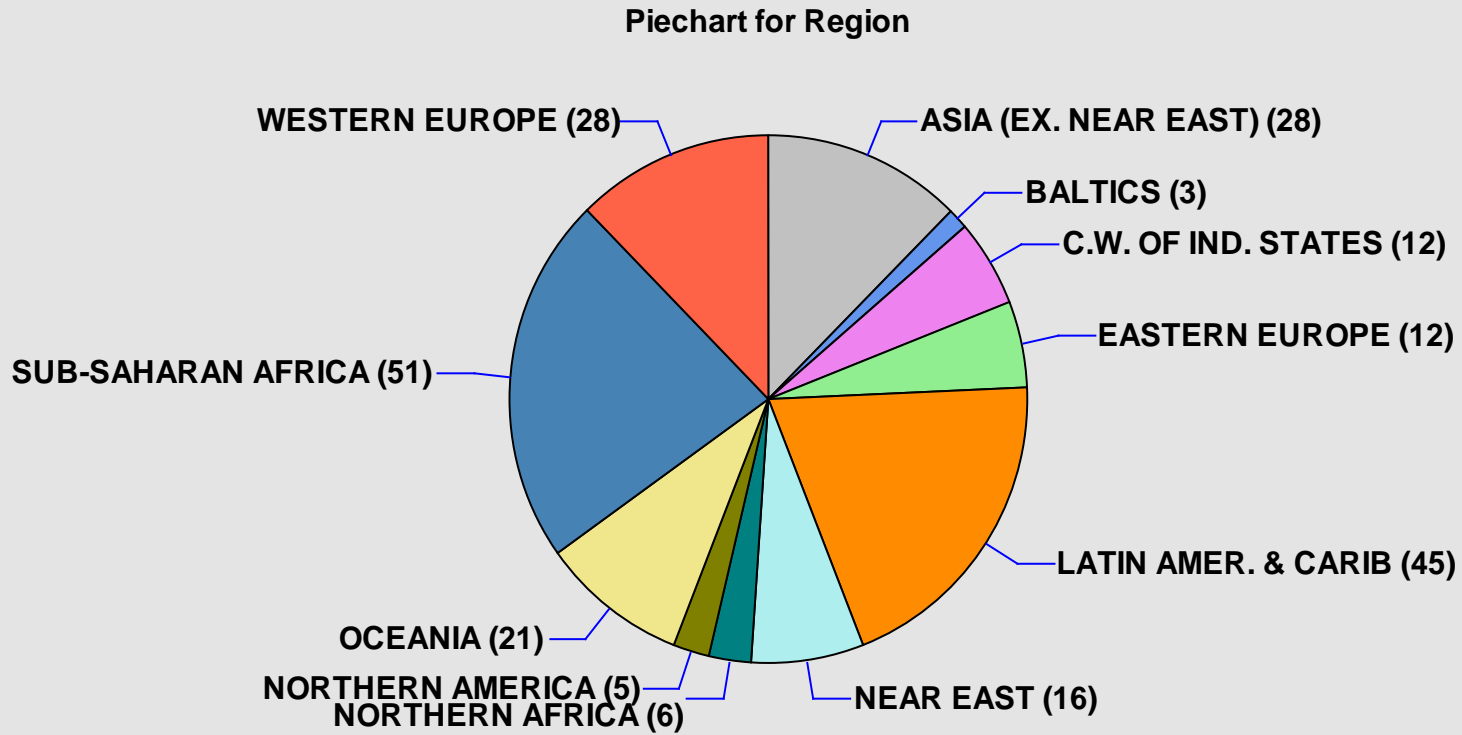
Main Window



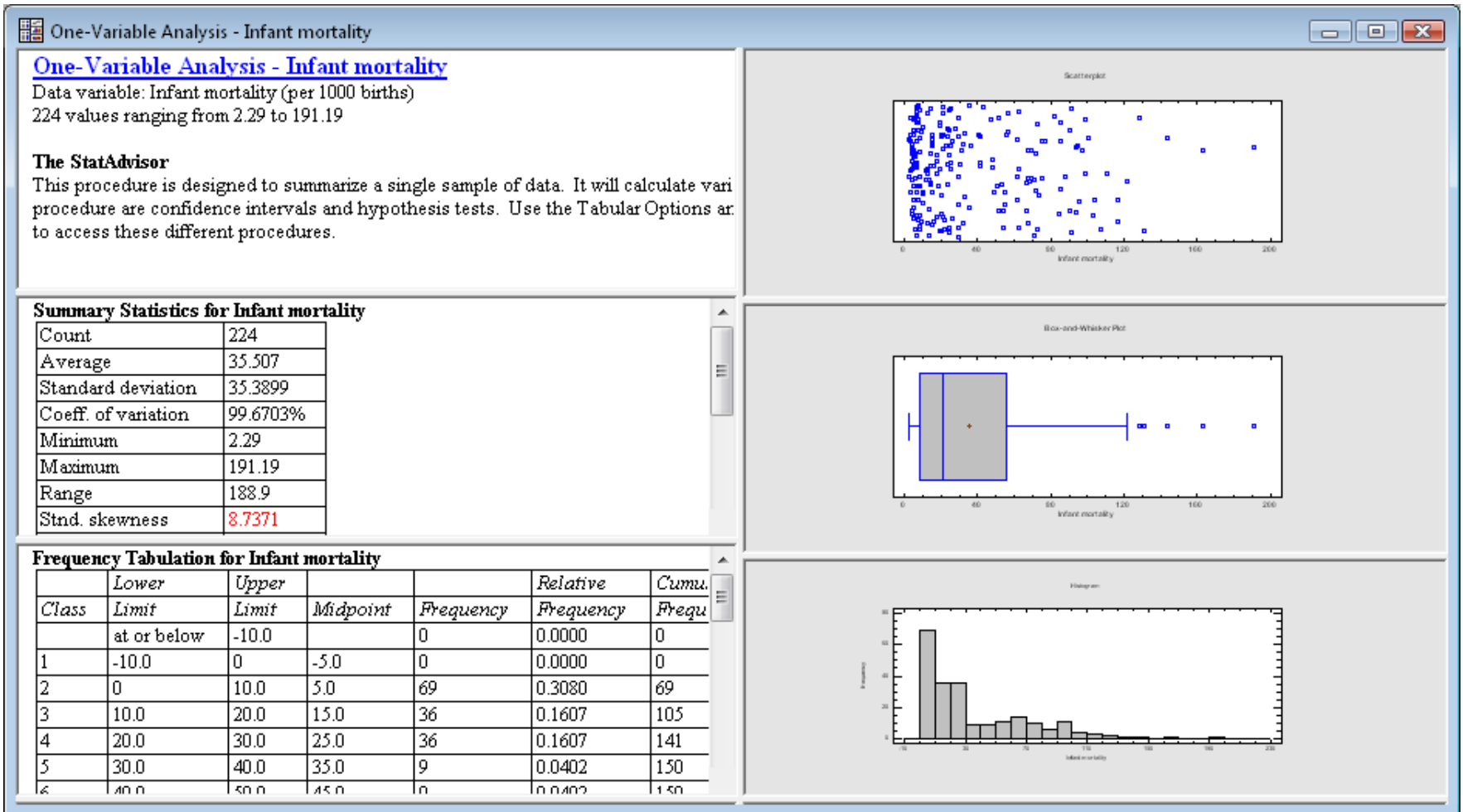
Statgraphics DataBook

	Country	Region	Population	Area	Pop. Density	Coastline	Net migration	Infant mortality
				sq. mi.	per sq. mi.	coast/area ratio		per 1000 births
1	Afghanistan	ASIA (EX. NEAR	31056997	647500	48.0	0.00	23.06	163.07
2	Albania	EASTERN EUROPE	3581655	28748	124.6	1.26	-4.93	21.52
3	Algeria	NORTHERN AFRICA	32930091	2381740	13.8	0.04	-0.39	31
4	American Samoa	OCEANIA	57794	199	290.4	58.29	-20.71	9.27
5	Andorra	WESTERN EUROPE	71201	468	152.1	0.00	6.6	4.05
6	Angola	SUB-SAHARAN AFRICA	12127071	1246700	9.7	0.13	0	191.19
7	Anguilla	LATIN AMER. & CARIBBEAN	13477	102	132.1	59.80	10.76	21.03
8	Antigua & Barbuda	LATIN AMER. & CARIBBEAN	69108	443	156.0	34.54	-6.15	19.46
9	Argentina	LATIN AMER. & CARIBBEAN	39921833	2766890	14.4	0.18	0.61	15.18
10	Armenia	C.W. OF IND. STATES	2976372	29800	99.9	0.00	-6.47	23.28
11	Aruba	LATIN AMER. & CARIBBEAN	71891	193	372.5	35.49	0	5.89
12	Australia	OCEANIA	20264082	7686850	2.6	0.34	3.98	4.69
13	Austria	WESTERN EUROPE	8192880	83870	97.7	0.00	2	4.66
14	Azerbaijan	C.W. OF IND. STATES	7961619	86600	91.9	0.00	-4.9	81.74
15	Bahamas, The	LATIN AMER. & CARIBBEAN	303770	13940	21.8	25.41	-2.2	25.21
16	Bahrain	NEAR EAST	698585	665	1050.5	24.21	1.05	17.27
17	Bangladesh	ASIA (EX. NEAR EAST)	147365352	144000	1023.4	0.40	-0.71	62.6
18	Barbados	LATIN AMER. & CARIBBEAN	279912	431	649.5	22.51	-0.31	12.5
19	Belarus	C.W. OF IND. STATES	10293011	207600	49.6	0.00	2.54	13.37
20	Belgium	WESTERN EUROPE	10379067	30528	340.0	0.22	1.23	4.68

Tabulation

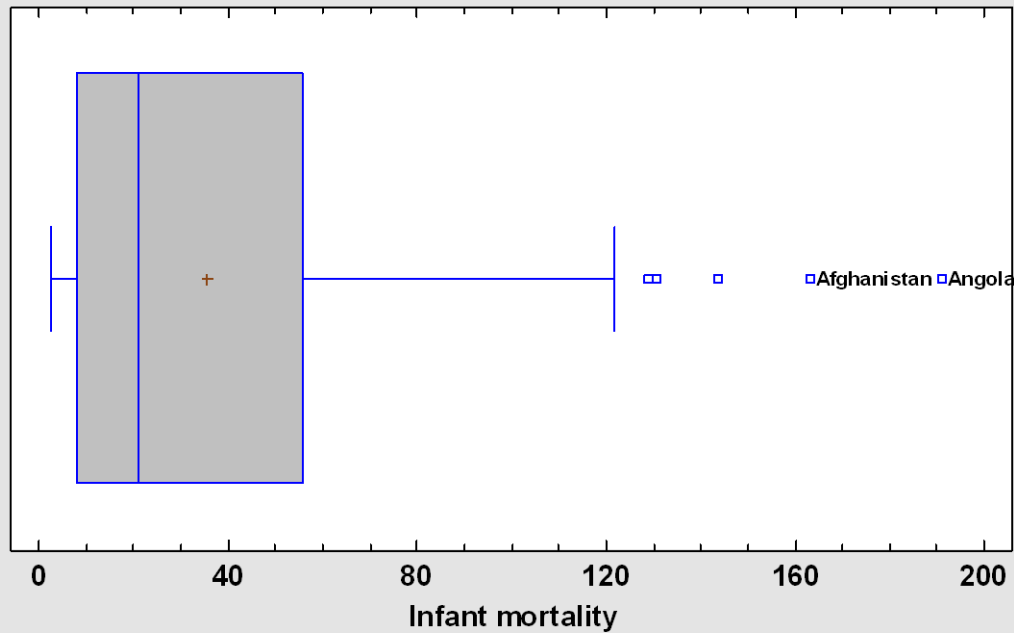


Variable of interest: infant mortality

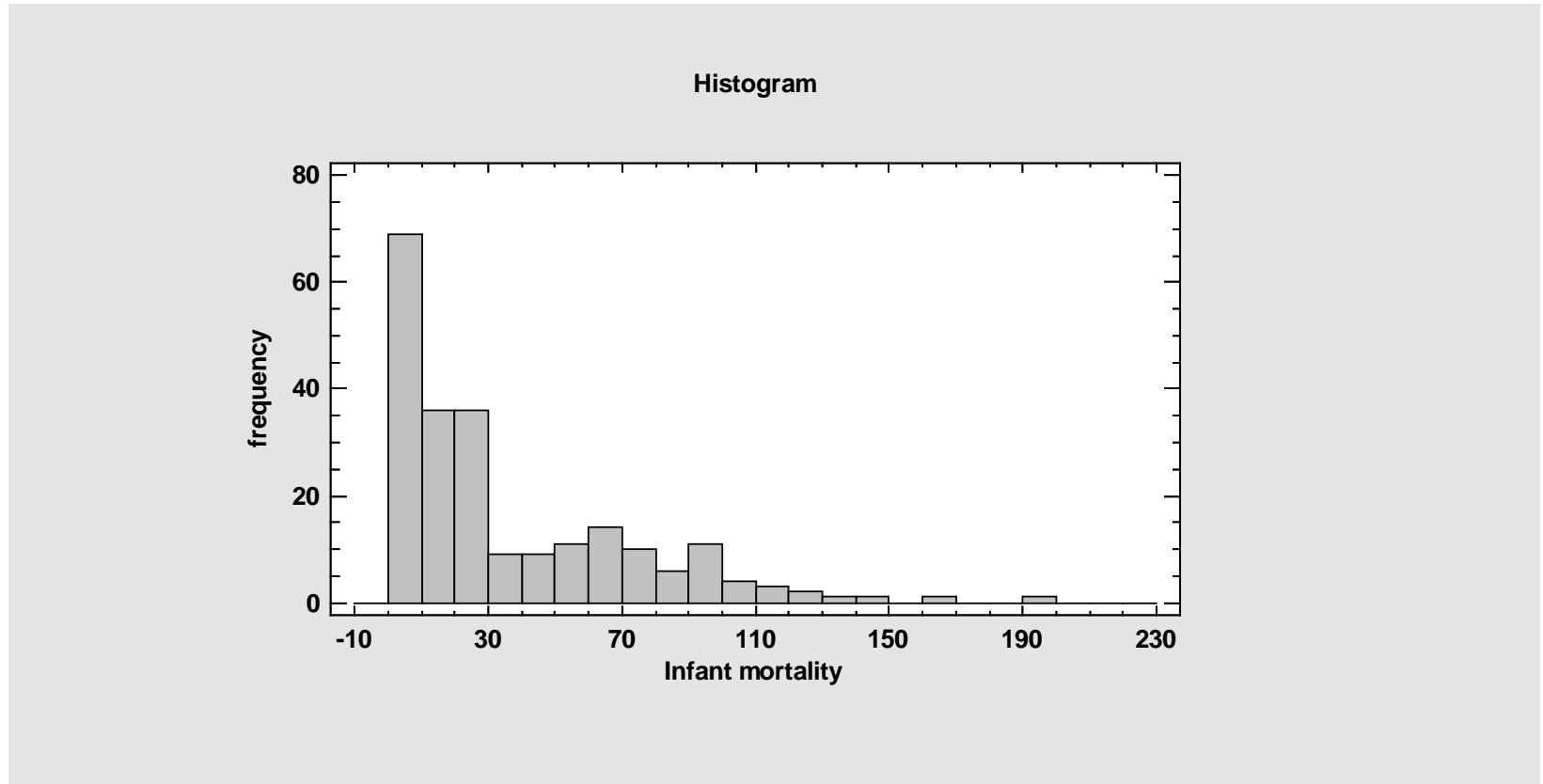


Box-and-whisker plot

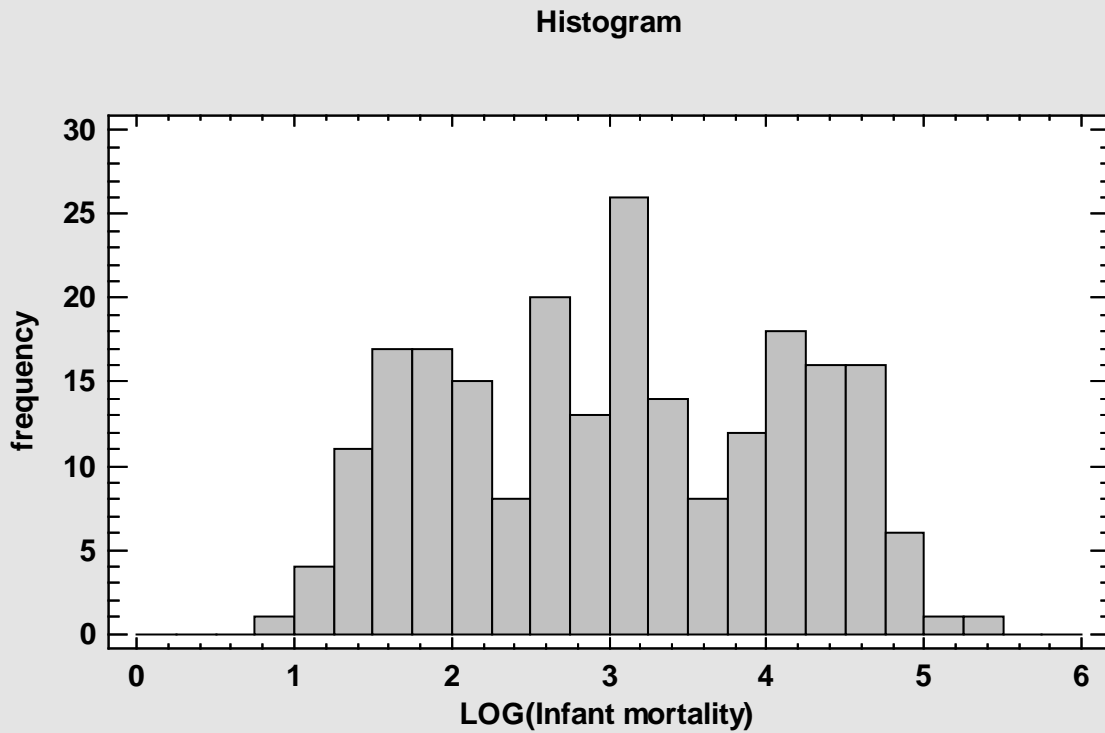
Box-and-Whisker Plot



Frequency Histogram



LOG Transformation



Summary Statistics & StatAdvisor

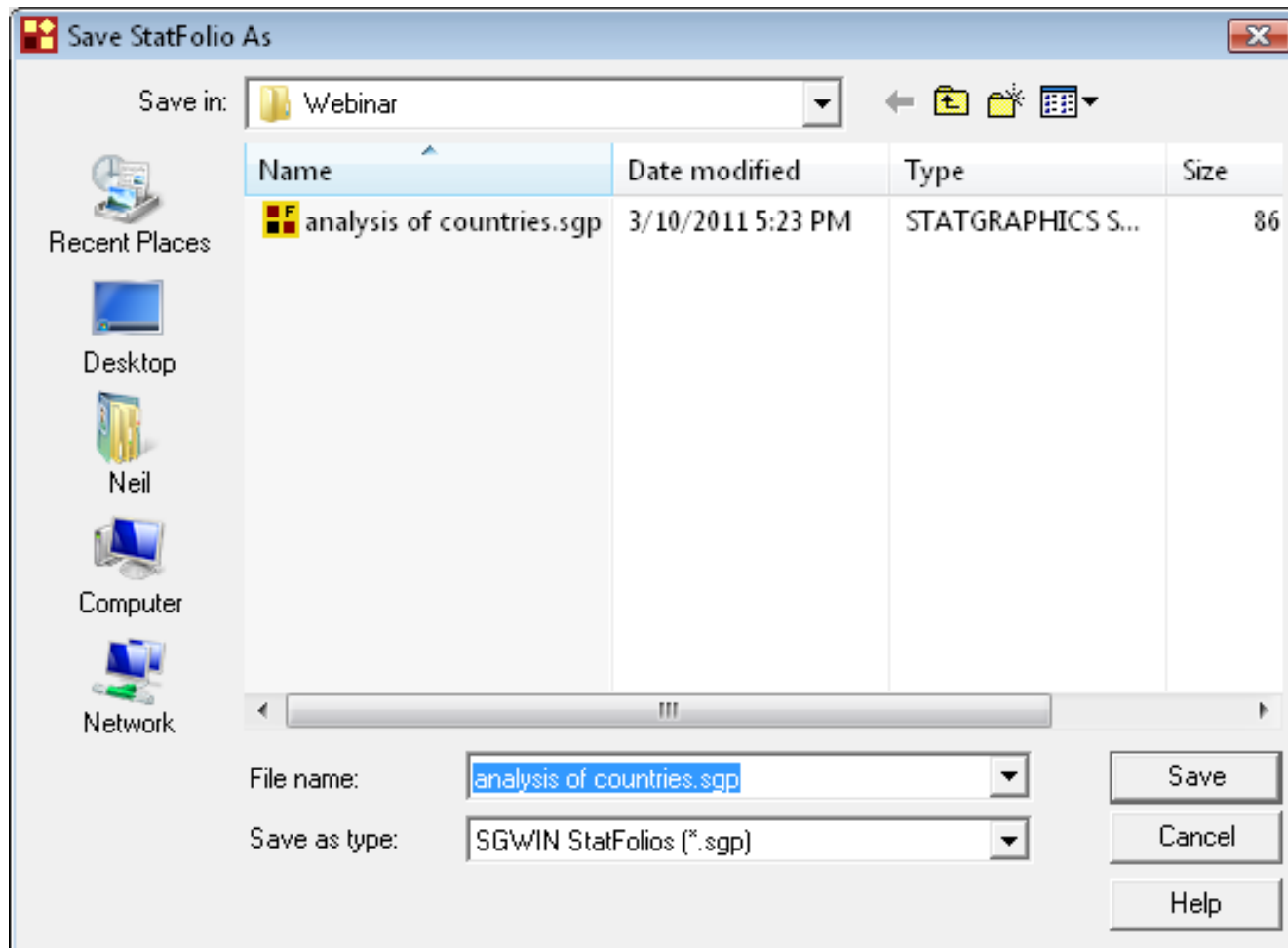
Summary Statistics for Infant mortality

Count	224
Average	35.507
Standard deviation	35.3899
Coeff. of variation	99.6703%
Minimum	2.29
Maximum	191.19
Range	188.9
Std. skewness	8.7371
Std. kurtosis	5.71182

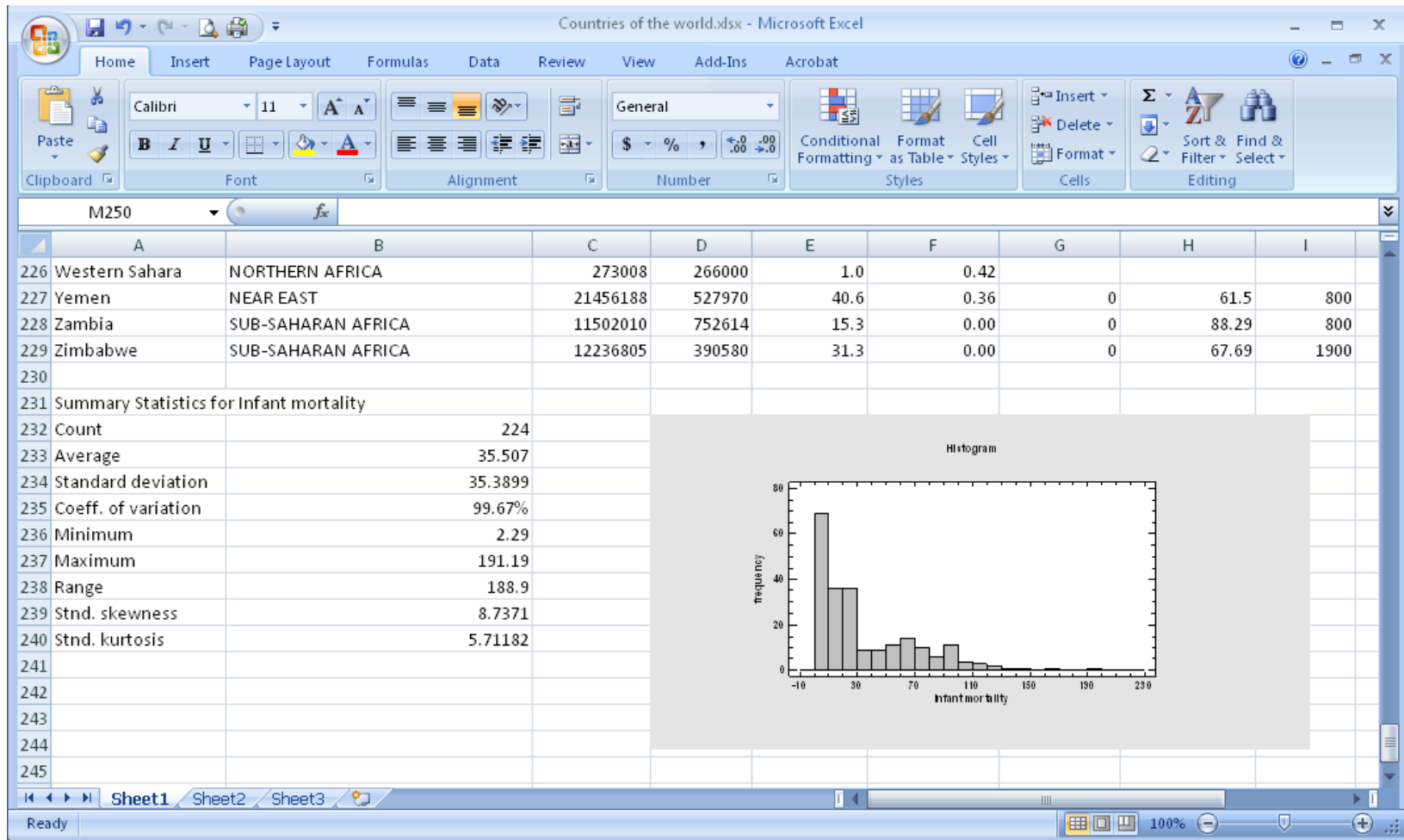
The StatAdvisor

This table shows summary statistics for Infant mortality. It includes measures of central tendency, measures of variability, and measures of shape. Of particular interest here are the standardized skewness and standardized kurtosis, which can be used to determine whether the sample comes from a normal distribution. Values of these statistics outside the range of -2 to +2 indicate significant departures from normality, which would tend to invalidate any statistical test regarding the standard deviation. In this case, the standardized skewness value is not within the range expected for data from a normal distribution. The standardized kurtosis value is not within the range expected for data from a normal distribution.

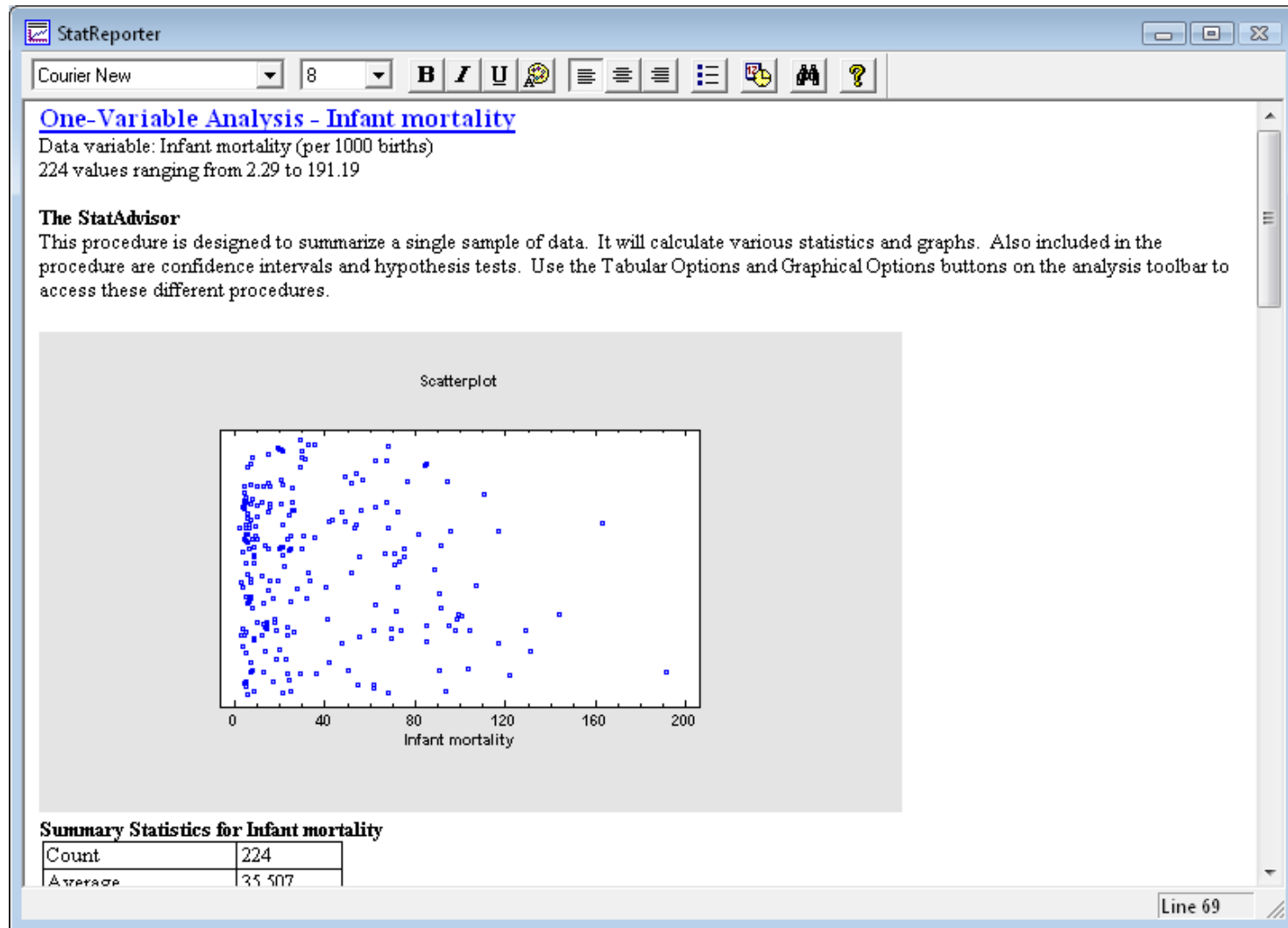
Save Session as a StatFolio



Save Results to Other Applications



Save All Results in StatReporter

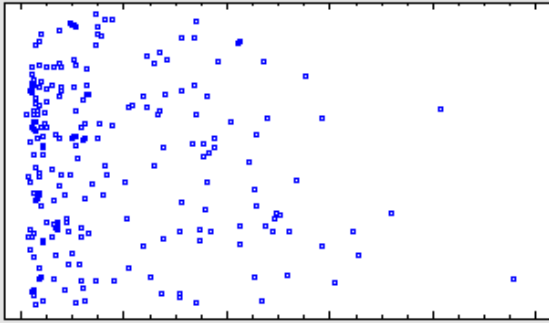


The screenshot shows the StatReporter window with a toolbar at the top containing options for font (Courier New), size (8), and various analysis tools. The main content area displays the following information:

One-Variable Analysis - Infant mortality
Data variable: Infant mortality (per 1000 births)
224 values ranging from 2.29 to 191.19

The StatAdvisor
This procedure is designed to summarize a single sample of data. It will calculate various statistics and graphs. Also included in the procedure are confidence intervals and hypothesis tests. Use the Tabular Options and Graphical Options buttons on the analysis toolbar to access these different procedures.

Scatterplot



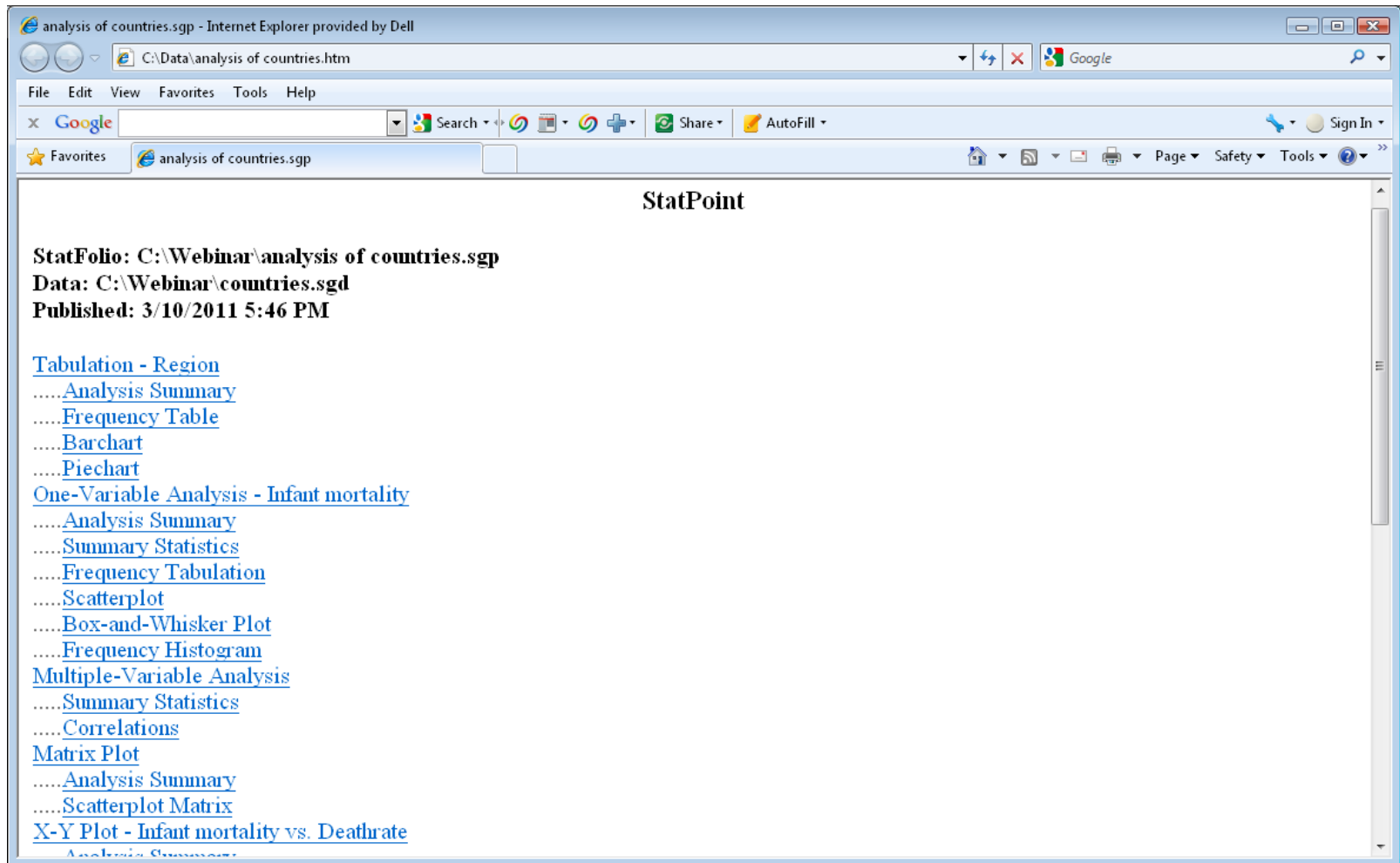
The scatterplot shows a distribution of data points (blue squares) for 'Infant mortality'. The x-axis is labeled 'Infant mortality' and ranges from 0 to 200 with major ticks every 40 units. The y-axis is unlabeled but represents the count of observations. The data points are scattered across the plot, with a higher density between 0 and 40 on the x-axis.

Summary Statistics for Infant mortality

Count	224
Average	35.507

Line 69

Save results as HTML using StatPublish



analysis of countries.sgp - Internet Explorer provided by Dell

C:\Data\analysis of countries.htm

File Edit View Favorites Tools Help

Google Search

analysis of countries.sgp

StatPoint

StatFolio: C:\Webinar\analysis of countries.sgp
Data: C:\Webinar\countries.sgd
Published: 3/10/2011 5:46 PM

[Tabulation - Region](#)
.....[Analysis Summary](#)
.....[Frequency Table](#)
.....[Barchart](#)
.....[Piechart](#)

[One-Variable Analysis - Infant mortality](#)
.....[Analysis Summary](#)
.....[Summary Statistics](#)
.....[Frequency Tabulation](#)
.....[Scatterplot](#)
.....[Box-and-Whisker Plot](#)
.....[Frequency Histogram](#)

[Multiple-Variable Analysis](#)
.....[Summary Statistics](#)
.....[Correlations](#)

[Matrix Plot](#)
.....[Analysis Summary](#)
.....[Scatterplot Matrix](#)

[X-Y Plot - Infant mortality vs. Deathrate](#)
.....[Analysis Summary](#)

Multivariate Analysis

Looking at more than one column simultaneously requires different techniques, such as:

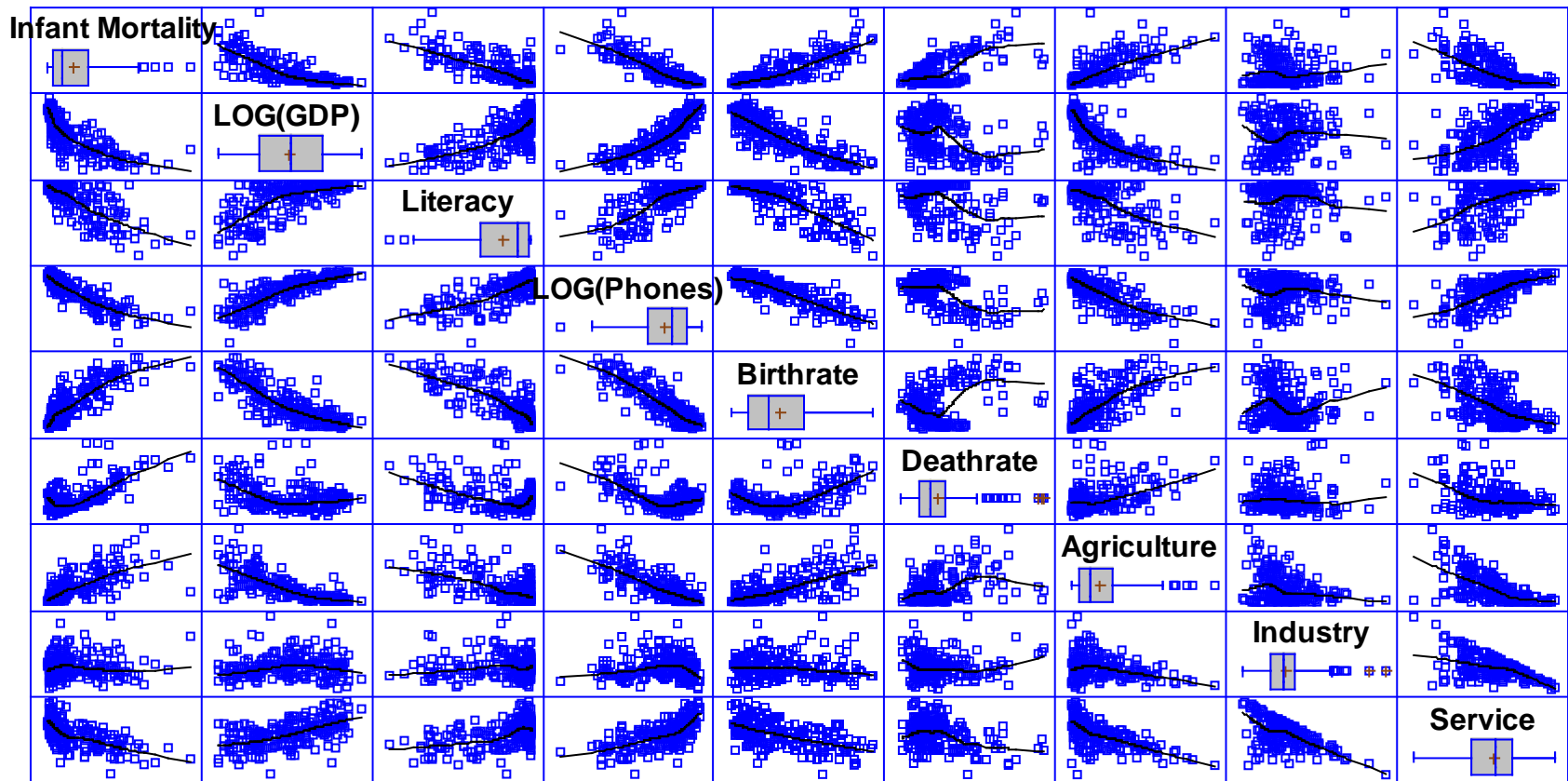
1. Correlation analysis
2. Matrix plot
3. Cluster analysis



Correlations

	Infant mortality	GDP	Literacy	Phones	Arable	Crops	Other
Population	0.0024 (179)	-0.0336 (179)	-0.0382 (179)	-0.0039 (179)	0.2067 (179)	-0.0626 (179)	-0.1373 (179)
	0.9742	0.6551	0.6120	0.9583	0.0055	0.4054	0.0667
Area	0.0029 (179)	0.0684 (179)	0.0007 (179)	0.0779 (179)	-0.0537 (179)	-0.1604 (179)	0.1245 (179)
	0.9690	0.3632	0.9928	0.3001	0.4749	0.0319	0.0967
Pop. Density	-0.1432 (179)	0.1901 (179)	0.0861 (179)	0.1571 (179)	-0.0592 (179)	-0.0366 (179)	0.0668 (179)
	0.0558	0.0108	0.2519	0.0358	0.4310	0.6269	0.3746
Coastline	-0.1060 (179)	0.0358 (179)	0.0996 (179)	0.0984 (179)	-0.0778 (179)	0.3994 (179)	-0.1371 (179)
	0.1581	0.6341	0.1846	0.1902	0.3006	0.0000	0.0673
Net migration	0.0131 (179)	0.3788 (179)	-0.0538 (179)	0.2325 (179)	-0.0658 (179)	-0.4054 (179)	0.2574 (179)
	0.8623	0.0000	0.4745	0.0017	0.3812	0.0000	0.0005
Infant mortality		-0.6391 (179)	-0.7612 (179)	-0.6992 (179)	-0.1230 (179)	-0.0957 (179)	0.1486 (179)
		0.0000	0.0000	0.0000	0.1008	0.2025	0.0471
GDP	-0.6391 (179)		0.5229 (179)	0.8835 (179)	0.0465 (179)	-0.2078 (179)	0.0664 (179)
	0.0000		0.0000	0.0000	0.5368	0.0052	0.3768
Literacy	-0.7612 (179)	0.5229 (179)		0.5920 (179)	0.0865 (179)	0.0607 (179)	-0.1012 (179)
	0.0000	0.0000		0.0000	0.2495	0.4193	0.1778
Phones	-0.6992 (179)	0.8835 (179)	0.5920 (179)		0.1241 (179)	-0.1248 (179)	-0.0386 (179)
	0.0000	0.0000	0.0000		0.0979	0.0959	0.6077

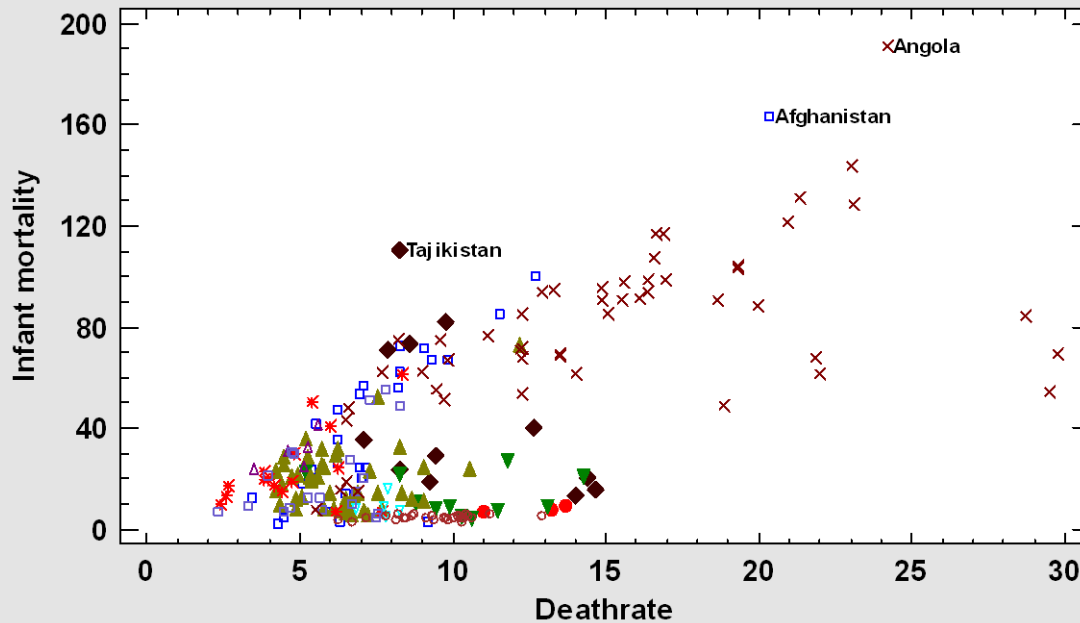
Matrix Plot



2-D Scatterplot

Useful for plotting 2 dimensions.

Plot of Infant mortality vs Deathrate

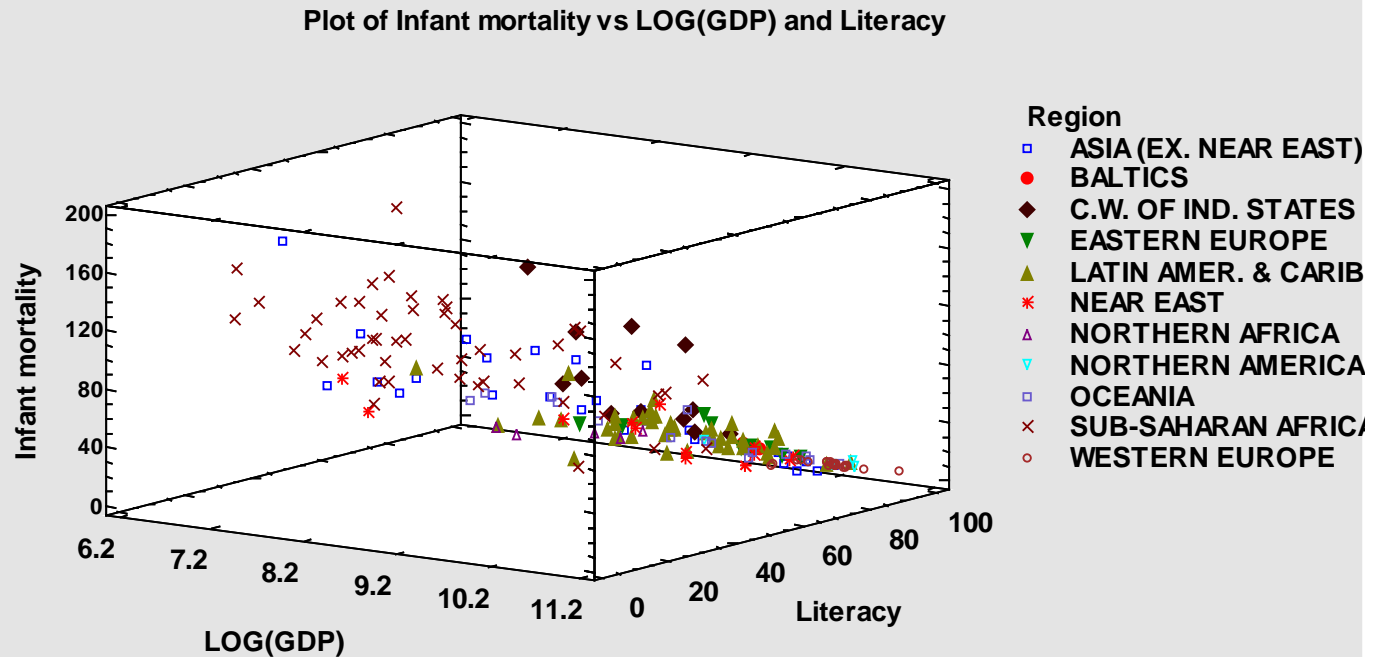


- Region
- ASIA (EX. NEAR EAST)
 - BALTICS
 - ◆ C.W. OF IND. STATES
 - ▼ EASTERN EUROPE
 - ▲ LATIN AMER. & CARIB
 - * NEAR EAST
 - △ NORTHERN AFRICA
 - ▽ NORTHERN AMERICA
 - OCEANIA
 - × SUB-SAHARAN AFRICA
 - WESTERN EUROPE

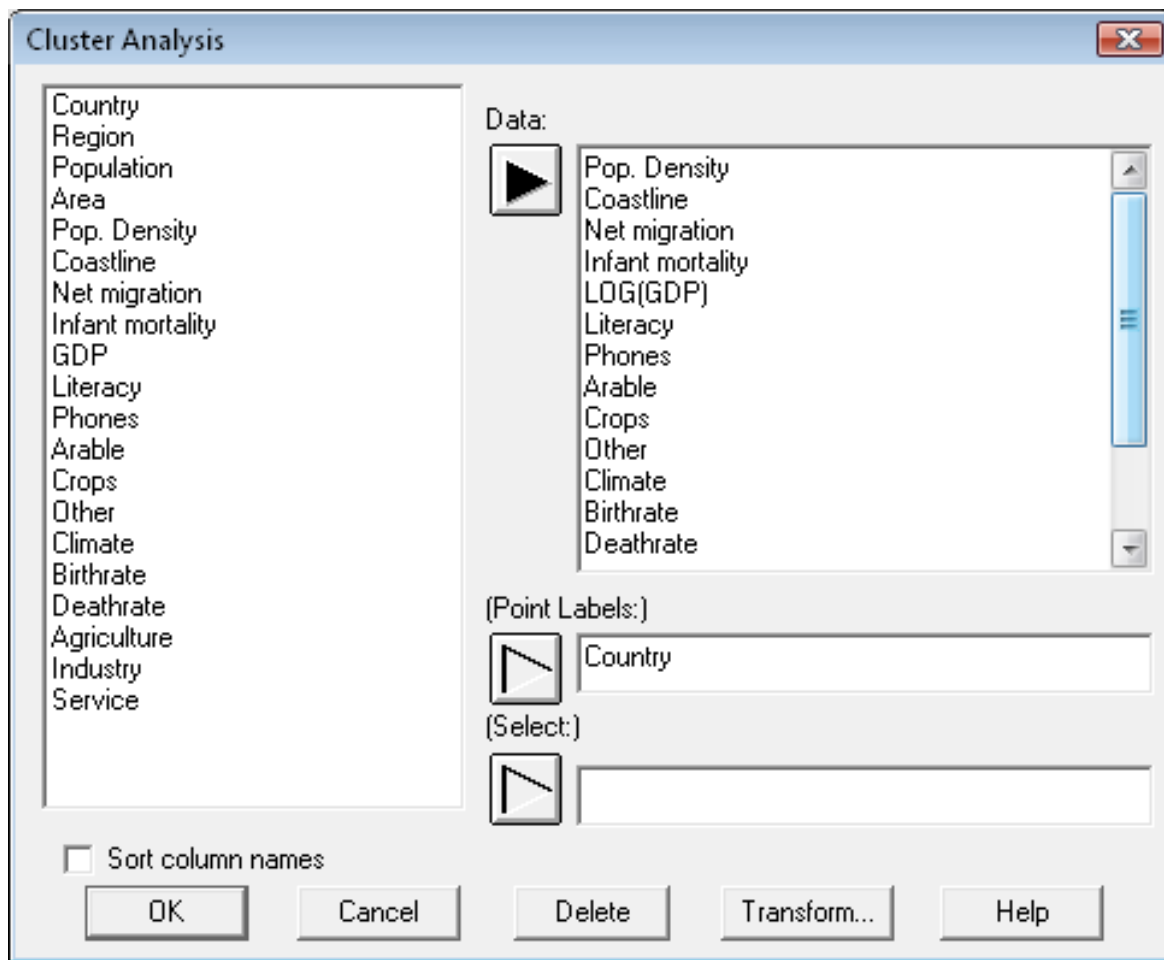


3-D Scatterplot

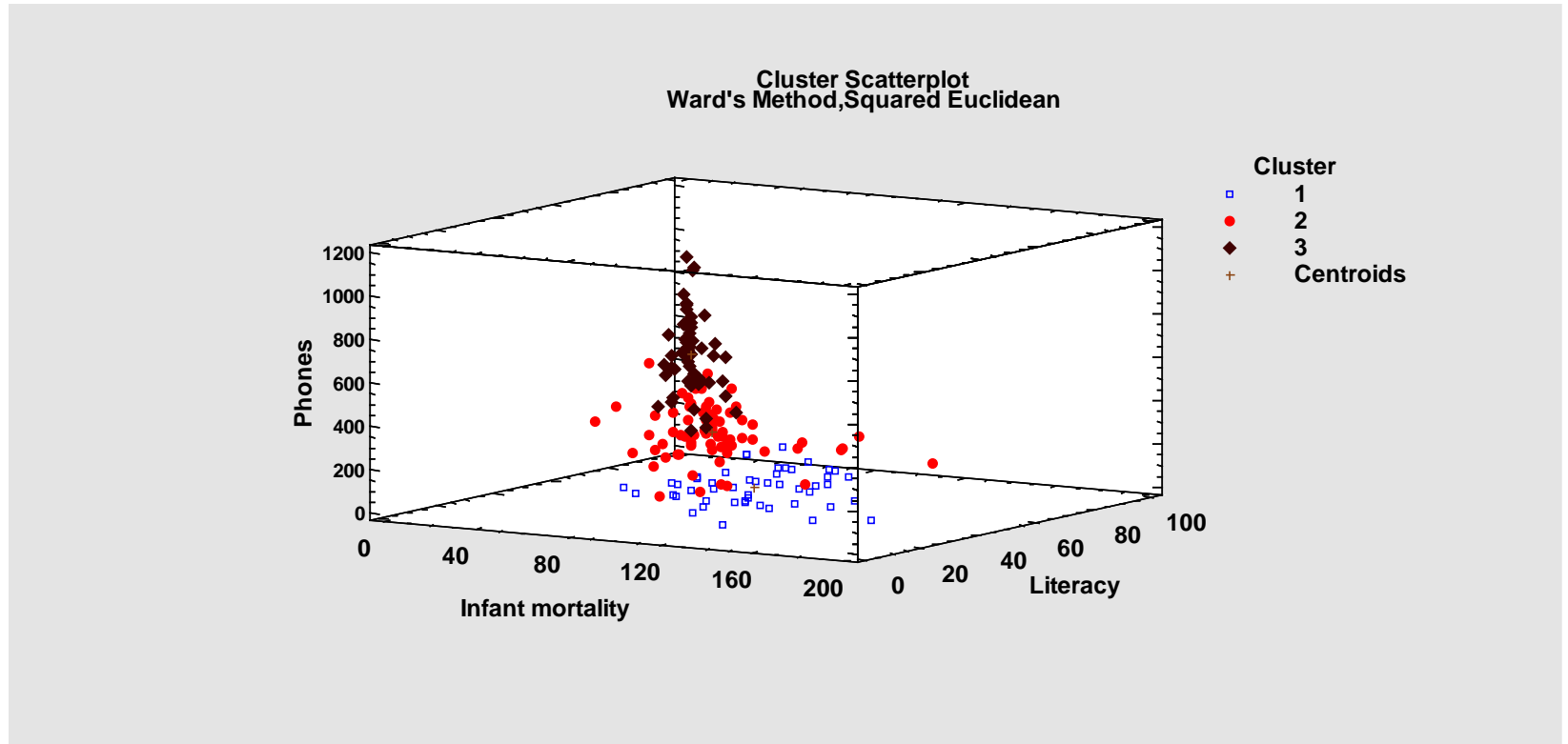
Useful for plotting 3 dimensions.



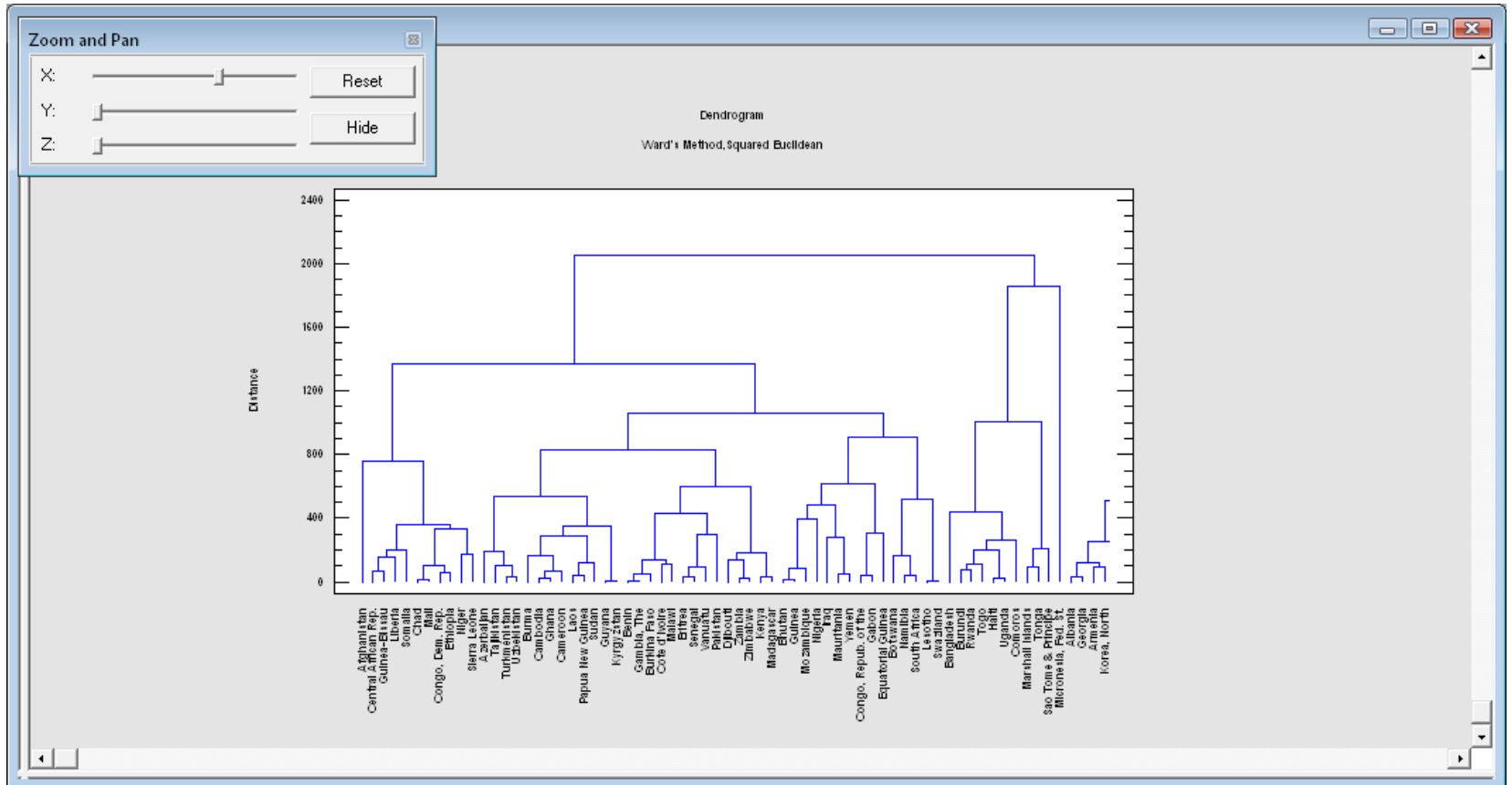
Cluster Analysis



Cluster Plot



Dendrogram



More Information

Go to www.statgraphics.com

Or send e-mail to info@statgraphics.com

