

101 Graphical Techniques

Dr. Kishore K. Das

Dr. Dibyojyoti Bhallarcharjee

Asian Books Private Limited

101

GRAPHICAL TECHNIQUES

Dr. Kishore K. Das

*Reader, Deptt. of Statistics
Gauhati University, Guwahati*

&

Dr. Dibyojyoti Bhattacharjee

*Senior Lecturer, Deptt. of Statistics
Gauhati University, Gauhati*



Asian Books Private Limited

7/28, Mahavir Lane, Vardan House, Ansari Road,
Darya Ganj, New Delhi-110002

Asian Books Private Limited

Corporate and Editorial office

7/28, Mahavir Lane, Vardan House, Ansari Road, Darya Ganj, New Delhi-110 002.

E-Mail: asian@nda.vsnl.net.in; purobi@asianbooksindia.com

World Wide Web: <http://www.asianbooksindia.com>

Phones: 23287577, 23282098, 23271887, 23259161 Fax: 91-11-23262021

Sales Offices

- Bangalore* 103, Swiss Complex No. 33, Race Course Road, Bangalore-560 001
Phones: 22200438, Fax: 91-80-22256583,
Email: asianblr@airtelbroadband.in
- Chennai* Palani Murugan Building No.21, West Cott Road, Royapettah,
Chennai-600 014, Phones: 28486927, 28486928,
Email: asianmde@vsnl.net
- Delhi* 7/28, Mahavir Lane, Vardan House, Ansari Road, Darya Ganj,
New Delhi-110 002,
Phones: 23287577, 23282098, 23271887, 23259161;
Fax: 91-11-23262021 E-Mail: asian@nda.vsnl.net.in
- Guwahati* 6, G.N.B. Road, Panbazar, Guwahati, Assam-781 001
Phones: 0361-2513020, 2635729 Email: asianghy1@sancharnet.in
- Hyderabad* 3-5-1101/1/B IInd Floor, Opp. Blood Bank, Narayanguda,
Hyderabad-500 029 Phones: 24754941, 24750951, Fax: 91-40-24751152
Email: hydasian@hd2.vsnl.net.in, hydasian@eth.net
- Kolkata* 10 A, Hospital Street, Calcutta-700 072
Phones: 22153040, Fax: 91-33-22159899 Email: calasian@vsnl.com
- Mumbai* Showroom: 3 & 4, Ground Floor, Shilpin Centre, 40, G.D. Ambekar
Marg, Wadala, Mumbai-400 031;
Phones : 22619322, 22623572, Fax: 24159899
- Noida* G-20, Sector 18, Atta Market, Noida
Ph: 9312234916, Email: asiannoida@asianbooksindia.com
- Pune* Shop No. 5-8, G.F. Shaan Brahma Com., Near Ratan Theatre,
Budhwar Peth Pune-02: Phones: 24497208, Fax: 91-20-24497207
Email: asianpune@asianbookindia.com

© Authors

1st Published: 2008

ISBN: 978-81-8412-048-6

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording and/or otherwise, without the prior written permission of the publisher.

Published by **Kamal Jagasia** for Asian Books Pvt. Ltd., 7/28, Mahavir Lane, Vardan House, Ansari Road, Darya Ganj, New Delhi-110 002

Laser Typeset at P.N Computers Shahdara Delhi-110 032

Printed at Yash Printograph.

"This page is Intentionally Left Blank"

Dedicated to our beloved Teachers

**Dr. Jyoti Prasad Medhi,
Emeritus Professor of Gauhati University (Retired)**

**Dr. S. B. Nandi,
Professor of Gauhati University (Retired)**

**Dr. Ashok Kumar Bansal,
Professor of University of Delhi**

**Dr. Salil Dutta
Professor of Arya Vidyapeeth College (Retired), Guwahati**

&

**Indu Bhusan Roy
Professor of G.C. College (Retired), Silchar**

Preface

The title “101 Graphical Techniques”, itself indicates the very purpose of the book. Data visualization or graphical representation of Statistical data is now an emerging subfield of Statistics. Though various graphical techniques existed in statistical literature but some very common graphical techniques are found to replicate themselves in almost all books and journals. But with the advances in the field of technology the use of graphics has increased many times. Today, graphical methods play an important role in all aspects of statistical investigation—it begins with explanatory plots, supports various stages of analysis and helps in the final communication and display of results. These days we see extensive use of graphics in print media, television news, sports coverage, advertisement etc. and this gives us an idea about the development of statistical graphics. Most of the recent versions of statistical software are now producing high-resolution self-explanatory graphics and is adding number of more graphs in every recent version of their product. However, the software packages do not provide sufficient text to support the purpose and interpretation of the relatively less common graphics and this in turn restrict their use. Social scientists can now get their data analyzed by using graphical tools that are handy and easier compared to other statistical methods.

But absence of proper text has hindered the development of graphics and there use is restricted even though software is available for producing the graphs. The graphical techniques, if used can reduce a lot of calculations that is involved with other statistical techniques in reaching to a conclusion. A number of data analyst and research workers are in search of a text which can act as a torch bearer in the world of statistical graphics.

Various universities of Europe and USA have started to develop specialized courses on “Data Visualization” or “Statistical Graphics”. The trend is soon going to enter different parts of the globe and this book can be considered a handy material for “Statistical Graphics”.

The book comprises of 101 statistical plots that can be used for analysis, display and comparison of data. Data sets are provided with most of the plots and relevant calculation if any are also shown. The variables considered along the axes are highlighted and the interpretation of the graph is also discussed. The uses of each of the graphical tool are forwarded along with some related statistical/graphical tools.

The authors like to put into record the support obtained from the software packages like Statistica, Dataplot, DJB Graphics and MS Excel for drawing the plots. Thanks are also due to the web site <http://www.itl.nist.gov/> which encouraged us with the basic idea of writing such a book. We owe indebtedness to all the authors whose work provided us necessary inspiration for completing this arduous task.

Thanks are also due to Mr. Debasish Bhattacharjee, Branch Manager, Asian Books Private Limited, Guwahati and our publisher for considering the matter for publication.

Errors which have inadvertently crept in the book are regretted.

Authors

Contents

Preface	(v)
List of Tables	(xi)
List of Figures	(xv)
1. Age Pyramid	1
2. Andrews Curve.....	3
3. ANOM Plot	6
4. Area Charts	8
5. Association Plot	10
6. Auto Correlation Plot	13
7. Auto Covariance Plot	15
8. Bar Chart (Simple).....	17
9. Bar Chart (Multiple).....	19
10. Bar Chart (Subdivided)	21
11. Bihistogram	23
12. Binomialness Plot	25
13. Block Plot	28
14. Box Plots	31
15. Box-Cox Linearity Plot	34
16. Box-Cox Normality Plot	37
17. Bubble Plot	40
18. Bubble Plot (Categorical).....	43
19. Cartogram	46
20. Categorical Scatter Plot	48
21. <i>c</i> -Chart	51
22. Chernoff Faces	54
23. Chi Plot	57
24. Chigram	60
25. Circles	63

26.	Column Icon Plot	64
27.	Column Plot (Circular Base)	66
28.	Comovement Plot	68
29.	Contour Plot	70
30.	Coplots	72
31.	Correlation Plot	74
32.	Corrogram	76
33.	Cubes	78
34.	CUSUM Control Chart.....	80
35.	Duane Plot	82
36.	Dendrogram	84
37.	Detrended Probability Plot	87
38.	Deviation Plot	90
39.	Digraph	92
40.	DOE Mean Plot	94
41.	DOE Scatter Plot	96
42.	Double Y Axis Plot	98
43.	Doughnut Chart.....	100
44.	Dubey Plot	102
45.	EDF Plot.....	105
46.	EDF Plot (with Doksum Bounds)	108
47.	Error-Bar Plot	112
48.	EWMA Control Chart	114
49.	Extreme Plot	117
50.	Fishbone Chart	119
51.	Four Fold Display	120
52.	Frequency Polygon/Curve	123
53.	Glyph Plot	125
54.	Glyph Plot (Categorical)	127
55.	Hanging Rootogram	129
56.	Hilo Diagrams	131
57.	Histogram	133
58.	Historigrams	135
59.	Homoscedesticity Plot.....	137
60.	I Plot	139
61.	Interaction Plot	141

62.	Jittered plot	143
63.	Kaplan Meier Plot	145
64.	Lag Plot	147
65.	Line Diagram	149
66.	Linear Intercept Plot	151
67.	Linear Slope Plot	153
68.	Lorenz Curve	155
69.	Mosaic Plot	158
70.	Moving Average Plot	162
71.	MSE Plot	164
72.	Normal Probability Plot	167
73.	np -Chart	170
74.	Ogive	173
75.	Ord Plot	175
76.	Parallel Co-ordinate Plot	180
77.	Pareto Plot	183
78.	p -Chart	186
79.	Percent Defective Plot	189
80.	Pictogram	191
81.	Pie Diagram	193
82.	Pie Icon Plot	195
83.	Poissonness Plot	197
84.	Probability Plot	200
85.	Profile Icon Plot	203
86.	Q - Q Plot	205
87.	R - Chart	207
88.	Radar Plot	210
89.	Residual Histogram	212
90.	Residual Plot	214
91.	Rootogram	216
92.	Run Sequence plot	218
93.	Scatter Diagram	220
94.	Scatterplot Matrix	223
95.	Sieve's Diagram	225
96.	Stacked Line Chart	227
97.	Star Icon Plot	229

98. Stem and Leaf Diagram	231
99. Sunflower Plot	233
100. Sunflower Plot (Categorical)	235
101. Sunray Icon Plot	237
References	239
Glossary.....	242

List of Tables

Table No.	Title of the Table	Page No.
1.1.	Population of a town classified by age and sex	1
2.1.	Olympic Decathlon Data	3
3.1.	Birth weight of Poland China Pigs in Pounds	6
4.1.	Demand-supply gap in the power sector of India	8
5.1.	A Contingency Table Showing Age of Respondent and Labour Type	11
6.1.	Table showing papers presented by American authors in International Seminars on Pattern Recognition	13
7.1.	Ten samples of size 25 each drawn from uniform [0, 1] distribution	15
8.1.	Data showing the outlay of expenditure during the second and third five year plans	17
9.1.	Distribution of persons employed in various public sectors in thousand	19
10.1.	Data showing the total outlay of expenditure during the first and second five year plans	21
11.1.	Head length of first and second sons of some families	23
12.1.	Fish Catch Data from David (1971)	26
13.1.	Counts of surviving latherjackets for different Controls and Emulsions each at two levels	28
14.1.	Index of Industrial Production: Sector-wise	31
15.1.	Year-wise Advances (in crores), provided by the various Regional Rural Banks of India	34
16.1.	Head length of sons of some families	37
17.1.	Data of 30 individuals pertaining to their height, weight and age	40

Table No.	Title of the Table	Page No.
18.1.	Data of 30 individuals pertaining to their height, weight, age and sex	43
19.1.	Production of Tea and Coffee in India during 1981-82	46
20.1.	Data corresponding to height, weight and sex of 30 individuals	48
21.1.	Number of defects in the different boxes full of switches	51
22.1.	Protein consumption in European countries	54
23.1.	Some hypothetical values of X and Y	58
24.1.	Lifetime (in hours) of 300 electric lamps from Gupta (1952)	60
25.1.	Sale of a particular product in '000.	63
27.1.	Rate of Advertisement Inflation and Circulation growth in Hindi dailies	66
28.1.	Net availability of cereals and pulses	68
28.2.	Comovement coefficients for different lags	69
32.1.	Random data generated in Excel for 5 variables	76
33.1.	Sale of a particular in '000	78
34.1.	Life in hours of cells after full charging	80
36.1.	Protein consumption in European countries	84
37.1.	A hypothetical data showing the values of X and Y	87
38.1.	Favorite internet cafe of the town	90
42.1.	Literacy rate and population density of India for the various census years	98
44.1.	Number of success in 26306 throws of 12 dice	103
44.2.	Calculation for Dubey plot based on the data in Table 44.1.	103
45.1.	EDF values for the Anderson Data	105
46.1.	EDF and Doksum bounds of the data	109
47.1.	Birth weight of Poland China Pigs in Pounds	112
48.1.	Life in hours of cells after full charging	114
49.1.	10 sub-samples of size 7 each drawn from a U[0, 1] population	117
51.1.	Asthma deaths, and Fenoterol use from Walker and Lanes (1991)	120
52.1.	A frequency distribution of wages of a group of employees	123
55.1.	Data from Jeffers (1978) along with expected frequencies from Poisson distribution	129
56.1.	Different data sets generated in Excel	131

Table No.	Title of the Table	Page No.
58.2.	Population of India for several census years	135
59.1.	Different data sets generated in Excel	137
60.1.	Cork deposits in centigram in trees planted in different directions	139
61.1.	Socio Economic status of parents and mental status of children	141
64.1.	Amount of Deposits in crores of Rupees in RRBs of India	147
65.1.	Population of Assam in different census	149
66.1.	Maximum temperature and relative humidity of 5 cities for 15 days	151
68.1	Data and calculations for Lorenz Curve	155
69.1.	A Contingency Table Showing Age of Respondent and Labour Type	158
70.1.	Yearwise production of an industry in thousand tons.	162
70.2.	3-Yearly moving average based on data in Table 70.1	162
71.1.	GNP and NNP of India for selected years	164
71.2.	MSE for different types of regression based on data in Table 71.1	165
72.1.	Data and calculations for normal probability plot	167
73.1.	Number of defectives in different days	171
74.1.	Frequency distribution of the marks of 95 students	173
75.1.	Clue to the understanding of the type of distribution from Ord Plots	176
75.2.	Calculation data on jeffries 1978 for Ord plot	178
76.1	A hypothetical data of 10 multiple variables	180
77.1.	Distribution of income of urban households of a country in percentage	183
77.2.	Necessary calculations for Pareto Curve and log transformations	184
78.1	Number of defectives in different days	187
79.1.	Number of defectives in different sub-samples by a machine before and after overhaul	189
80.1.	Year-wise female enrollment in a college	191
81.1.	Calculations Related to Pie Diagram	193
83.1.	Count and corresponding frequencies from Student (1906)	197
83.2.	Calculation of expected frequencies, observed and expected	198

(xiv)

Table No.	Title of the Table	Page No.
84.1.	Data and calculations for probability plot	201
87.1.	Life in hours of cells after full charging	207
88.1.	Index of Industrial Production: Sector-wise	210
92.1.	Data related to Coal Mining Disasters	218
93.1.	Percentage of fat in human body and corresponding age	221
96.1.	Actual stock of wheat in million tons	227

List of Figures

Figure No.	Title of the Figure	Page No.
1.1.	Age Pyramid representing the data in Table 1.1.	2
2.1.	An Andrews curve for data in Table 2.1.	5
3.1.	ANOM plot based on data in Table 3.1.	7
4.1.	Area Chart based on data in Table 4.1.	9
5.1.	Cohen's Association Plot based on the data provided in Table 5.1.	11
6.1.	An Autocorrelation plot for the data in Table 6.1.	14
7.1.	An Auto Covariance plot for the data in Table 7.1.	16
8.1.	A multiple bar diagram for the data in Table 8.1.	18
9.1.	A simple Bar Diagram for the data in Table 9.1.	19
10.1.	Sub-divided Bar Diagram for the data in Table 10.1	22
11.1.	A bihistogram for the data in Anderson (1958)	44
12.1.	A binomialness plot for the data in Table 12.1.	27
13.1.	A Block Plot to the data in Table 13.1.	29
14.1.	Box Plot for data provided in Table 14.1.	32
15.1.	A linear fit to the original data	35
16.1.	A Box-Cox Normality Plot	38
17.1.	A bubble plot for the data in Table 17.1	41
18.1.	A categorical bubble plot for the data in Table 18.1	44
19.1.	A cartogram to the data in Table 19.1	47
20.1.	Categorical scatter plot for the data in Table 20.1	49
21.1.	<i>c</i> chart for the data in Table 21.1	52
22.1.	The Chernoff's icon plot for the protein consumption data	55

Figure No.	Title of the Figure	Page No.
23.1.	A chi plot for the data in Table 23.1	58
24.1.	Histogram and chigram based on data in Table 24.1	61
25.1.	Circles corresponding to data in Table 24.1.	63
26.1.	A column icon plot for the protein consumption data	64
27.1.	Column Plot of data provided in Table 27.1.	67
28.1.	An Comovement Plot for the data in Table 28.2	69
29.1.	A contour plot for the above function	70
30.1.	A coplot scanned from Jacoby (1998)	72
31.1.	A Cross-correlation plot for the data in Table 28.1	74
32.1.	A Cross-correlation plot for the data in Table 32.1.	77
33.1.	Cubes corresponding to data in Table 33.1.	78
34.1.	The CUSUM chart drawn for the data provided in Table 34.1.	81
35.1.	Duane Plot corresponding to the data stated above	82
36.1.	A Dendrogram for data provided in Table 36.1.	85
37.1.	A detrended probability plot based on data in Table 37.1.	87
38.1.	A Deviation Plot for the data in Table 38.1.	91
39.1.	A Digraph for transition probability matrix provided above	92
40.1.	DOE Mean plot for data in Table 13.1.	94
41.1.	DOE scatter plot for data in Table 13.1	96
42.1.	Double Y-axis plot used for representing data in Table 42.1	99
43.1.	Doughnut Chart based on data in Table 10.1	100
44.1.	A Dubey plot to the data in Table 44.1.	104
45.1.	EDF plot for the Anderson Data	106
46.1.	An EDF plot along with the Doksum Bounds	110
47.1.	An Error Bar plot for the data in Table 47.1	113
48.1.	EWMA plot for the data provided in Table 48.1.	115
49.1.	An Extreme plot for the data provided in Table 49.1.	117
50.1.	A Fishbone chart showing the problems related to a sample survey	119
51.1.	A four fold display for the raw frequencies give in Table 51.1.	120
52.1.	A frequency polygon representing Table 52.1	123

Figure No.	Title of the Figure	Page No.
53.1.	A glyph plot for the data in Table 17.1.	125
54.1.	A categorical glyph plot for the data in Table 18.1	127
55.1.	A hanging rootogram for the data in Jeffers (1978)	130
56.1.	A HiLo diagram corresponding to the data in Table 56.1	132
57.1.	A histogram to the data in Table 52.1	133
58.1.	A Histogram for data in Table 58.1	136
59.1.	A homoscedasticity plot diagram corresponding to the data in Table 56.1	138
60.1.	An I Plot for the data in Table 60.1	140
61.1.	An Interaction Plot diagram corresponding to the data in Table 56.1.	142
62.1.	A jittered plot for data in Table 62.1.	143
63.1.	The Kaplan Meier Plot for data provided above	147
64.1.	Lag Plot (of lag 1) for data in Table 64.1.	148
65.1.	Line diagram corresponding to the data in Table 65.1	149
66.1.	A Linear intercept plot for the data provided in Table 66.1	152
67.1.	A Linear Slope plot for the data provided in Table 66.1.	153
68.1.	A Lorenz Curve for data in Table 68.1	156
69.1.	A Mosaic display as proposed by Hartigan and Kleiner	159
69.2.	A Mosaic display with the extension proposed by Friendly (1994)	160
70.1.	Moving Average Plot based on data in Table 70.2	163
71.1.	An MSE Plot for the data in Table 71.2	165
72.1.	A normal probability plot to the data in Table 72.1.	169
73.1.	np chart for the data in Table 73.3.	171
74.1.	The less than type Ogive for the data in Table 74.2	174
75.1.	An Ord Plot for the data from Jeffres (1978) along with the fitted lines	179
76.1.	A parallel coordinate plot for data in Table 76.1	181
77.1.	A Pareto curve for data provided in Table 77.1	184
77.2.	Plotting the values of x and y after logarithm transformation	185
78.1.	p chart for the data in Table 78.1	187
79.1.	A Percentage defective plot for the data provided in Table 7.1.	190
80.1.	Pictogram for data in Table 80.1	191

Figure No.	Title of the Figure	Page No.
81.1.	Pie Diagram based on data Tables 81.1	194
82.1.	The pie icon plot for the protein consumption data	195
83.1.	A Posissonness plot for the above data	198
84.1.	A probability plot for the data in Table 84.1	201
85.1.	The profile icon plot for the protein consumption data	203
86.1.	A Q-Q plot for the data in Table 11.1	205
87.1.	R chart for the data in Table 87.1	208
88.1.	Index of Industrial Production: Sector-wise	210
89.1.	Histogram and residual histogram based on data in Table 24.1.	212
90.1.	A Residual Plot for data in Table 37.1	214
91.1.	Histogram and residual rootgram based on data in Table 24.1	216
92.1.	A Run Sequence plot for data in Table 92.1.	218
93.1.	The scatter diagram for various types of correlation	220
93.2.	Scatter Diagram based on data Table 93.1.	221
94.1.	A Scatter plot matrix drawn based on the data in Table 22.1 (abridged)	223
95.1.	A Sieve diagram for data in Table 69.1.	225
96.1.	Stacked Line Chart based on data in Table 96.1	227
97.1.	The star icon plot for the protein consumption data	229
98.1.	A Stem and Leaf diagram to the data provided above	231
99.1.	A sunflower plot for data in Table 17.1	234
100.1.	A categorical sunflower plot for data in Table 18.1.	235
101.1.	The sunray icon plot for the protein consumption data	237

1. AGE PYRAMID

1.1. Definition and Description

For drawing this diagram we use the first quadrant and second quadrant only. Thus the entire X axis and only the positive part of Y axis is considered. Here the variable taken along Y axis (Age in most cases) is divided into several non-overlapping classes. For each class we can consider two related data sets (say Sex). The number of cases of the data sets under each class is then represented by two horizontal bars one along OX and the other along OX' , the length of which are proportional to the value of the observations. For example, in case of sex the number of females (or percentage) under each age group is represented along OX' and the number of males (or percentage) under each age group is represented along OX . This plot can be used for comparison for two related (comparable) data sets for various non overlapping classes. The plot can be used for the comparison of the magnitude of each component of a variable for various age groups. Since with the increase of age groups the number of occurrences keeps on decreasing so the length of each of the bars keep on decreasing with increasing age groups, thus taking the shape of pyramid. In case the values in the various age groups are similar for the different sexes then the pyramid symmetrical about the Y axis.

1.2. Working Data

The data used for drawing the Age pyramid is a hypothetical data showing the population of a city classified by age groups for the different sexes. The figures are expressed in thousands.

Table 1.1: Population of a town classified by age and sex

<i>Age Group</i>	<i>Population in Thousand</i>	
	<i>Female</i>	<i>Male</i>
0 – 10	166.6	178.2
10 – 20	152.3	163.7
20 – 30	195.7	192.2
30 – 40	167.3	167.7
40 – 50	160	161.2
50 – 60	122.3	128.8
60 – 70	122.5	114.3
70 – 80	107.2	69.2
80 – 90	57.2	23.4
90 – 100	9.8	2.3