

# Visual Statistics

Seeing Data with  
Dynamic Interactive Graphics

Forrest W. Young  
Pedro M. Valero-Mora  
Michael Friendly



WWW.  
LINK AVAILABLE

# Contents

## Part I Introduction

<b>1 Introduction</b>	<b>3</b>
1.1 Visual Statistics	6
1.2 Dynamic Interactive Graphics	7
1.2.1 An Analogy	7
1.2.2 Why Use Dynamic Graphics?	8
1.2.3 The Four Respects	8
1.3 Three Examples	9
1.3.1 Nonrandom Numbers	9
1.3.2 Automobile Efficiency	11
1.3.3 Fidelity and Marriage	14
1.4 History of Statistical Graphics	18
1.4.1 1600–1699: Measurement and Theory	18
1.4.2 1700–1799: New Graphic Forms and Data	19
1.4.3 1800–1899: Modern Graphics and the Golden Age	20
1.4.4 1900–1950: The Dark Ages of Statistical Graphics— The Golden Age of Mathematical Statistics	21
1.4.5 1950–1975: Rebirth of Statistical Graphics	22
1.4.6 1975–2000: Statistical Graphics Comes of Age	23
1.5 About Software	24
1.5.1 XLisp-Stat	25
1.5.2 Commercial Systems	26
1.5.3 Noncommercial Systems	26
1.5.4 ViSta	27
1.6 About Data	29

1.6.1 Essential Characteristics	30
1.6.2 Datatypes	32
1.6.3 Datatype Examples	34
1.7 About This Book	34
1.7.1 What This Book Is—and Isn't	34
1.7.2 Organization	34
1.7.3 Who Our Audience Is—and Isn't	37
1.7.4 Comics	38
1.7.5 Thumb-Powered Dynamic Graphics	39
1.8 Visual Statistics and the Graphical User Interface	40
1.9 Visual Statistics and the Scientific Method	40
1.9.1 A Paradigm for Seeing Data	41
1.9.2 About Statistical Data Analysis: Visual or Otherwise	42
<b>2 Examples</b>	<b>45</b>
2.1 Random Numbers	47
2.2 Medical Diagnosis	52
2.3 Fidelity and Marriage	59
<b>Part II</b>	
<b>See Data—The Process</b>	
<b>3 Interfaces and Environments</b>	<b>73</b>
3.1 Objects	77
3.2 User Interfaces for Seeing Data	78
3.3 Character-Based Statistical Interface Objects	79
3.3.1 Command Line	79
3.3.2 Calculator	80
3.3.3 Program Editor	80
3.3.4 Report Generator	81
3.4 Graphics-Based Statistical Interfaces	81
3.4.1 Datasheets	81
3.4.2 Variable Window	82
3.4.3 Desktop	82
3.4.4 Workmap	83
3.4.5 Selector	87
3.5 Plots	88
3.5.1 Look of Plots	89
3.5.2 Feel of Plots	91

## Contents

3.5.3 Impact of Plot Look and Feel	93
3.6 Spreadplots	94
3.6.1 Layout	96
3.6.2 Coordination	98
3.6.3 SpreadPlots	100
3.6.4 Look of Spreadplots	102
3.6.5 Feel of Spreadplots	104
3.6.6 Look and Feel of Statistical Data Analysis	104
3.7 Environments for Seeing Data	111
3.8 Sessions and Projects	114
3.9 The Next Reality	114
3.9.1 The Fantasy	114
3.9.2 The Reality	116
3.9.3 Reality Check	118
<b>4 Tools and Techniques</b>	<b>119</b>
4.1 Types of Controls	123
4.1.1 Buttons	123
4.1.2 Palettes	125
4.1.3 Menus and Menu Items	125
4.1.4 Dialog Boxes	125
4.1.5 Sliders	126
4.1.6 Control Panels	127
4.1.7 The Plot Itself	127
4.1.8 Hyperlinking	127
4.2 Datasheets	128
4.3 Plots	129
4.3.1 Activating Plot Objects	131
4.3.2 Manipulating Plot Objects	132
4.3.3 Manipulating Plot Dimensions	138
4.3.4 Adding Graphical Elements	141

## Part III Seeing Data—Objects

<b>5 Seeing Frequency Data</b>	<b>145</b>
5.1 Data	148
5.1.1 Automobile Efficiency:	148
5.1.2 Berkeley Admissions Data	148

5.1.3	Tables of Frequency data	150
5.1.4	Working at the Categories Level	151
5.1.5	Working at the Variables Level	153
5.2	Frequency Plots	157
5.2.1	Mosaic Displays	157
5.2.2	Dynamic Mosaic Displays	159
5.3	Visual Fitting of Log-Linear Models	164
5.3.1	Log-Linear Spreadplot	165
5.3.2	Specifying Log-Linear Models and the Model Builder Window	166
5.3.3	Evaluating the Global Fit of Models and Their History	170
5.3.4	Visualizing Fitted and Residual Values with Mosaic Displays	174
5.3.5	Interpreting the Parameters of the Model	176
5.4	Conclusions	179
<b>6</b>	<b>Seeing Univariate Data</b>	<b>181</b>
6.1	Introduction	183
6.2	Data: Automobile Efficiency	185
6.2.1	Looking at the Numbers	186
6.2.2	What Can Unidimensional Methods Reveal?	186
6.3	Univariate Plots	190
6.3.1	Dotplots	190
6.3.2	Boxplots	193
6.3.3	Cumulative Distribution Plots	196
6.3.4	Histograms and Frequency Polygons	199
6.3.5	Ordered Series Plots	208
6.3.6	Namelists	209
6.4	Visualization for Exploring Univariate Data	209
6.5	What Do We See in <i>MPG</i> ?	212
<b>7</b>	<b>Seeing Bivariate Data</b>	<b>215</b>
7.1	Introduction	217
7.1.1	Plots About Relationships	217
7.1.2	Chapter Preview	220
7.2	Data: Automobile Efficiency	221
7.2.1	What the Data Seem to Say	222
7.3	Bivariate Plots	224
7.3.1	Scatterplots	224
7.3.2	Distribution Comparison Plots	233
7.3.3	Parallel-Coordinates Plots and Parallel Boxplots	236
7.4	Multiple Bivariate Plots	236
7.4.1	Scatterplot Plot Matrix	237

7.4.2 Quantile Plot Matrix	238
7.4.3 Numerical Plot-matrix	238
7.4.4 BoxPlot Plot Matrix	239
7.5 Bivariate Visualization Methods	241
7.6 Visual Exploration	242
7.6.1 Two Bivariate Data Visualizations	243
7.6.2 Using These Visualizations	245
7.7 Visual Transformation: Box-Cox	247
7.7.1 The Transformation Visualization	249
7.7.2 Using Transformation Visualization	251
7.7.3 The Box-Cox Power Transformation	255
7.8 Visual Fitting: Simple Regression	256
7.9 Conclusions	260

**8 Seeing Multivariate Data 263**

8.1 Data: Medical Diagnosis	266
8.2 Three Families of Multivariate Plots	270
8.3 Parallel-Axes Plots	272
8.3.1 Parallel-Coordinates Plot	272
8.3.2 Parallel-Comparisons Plot	276
8.3.3 Parallel Univariate Plots	277
8.4 Orthogonal-Axes Plots	279
8.4.1 Spinplot	280
8.4.2 Orbitplot	283
8.4.3 BiPlot	286
8.4.4 Wiggle-Worm (Multivariable Comparison) Plot	291
8.5 Paired-Axes Plots	292
8.5.1 Spinplot Plot Matrix	293
8.5.2 Parallel-Coordinates Plot Matrix	294
8.6 Multivariate Visualization	295
8.6.1 Variable Visualization	295
8.6.2 Principal Components Analysis	296
8.6.3 Fit Visualization	298
8.6.4 Principal Components Visualization	300
8.6.5 One More Step - Discriminant Analysis	302
8.7 Summary	304
8.7.1 What Did We See? Clusters!	304
8.7.2 How Did We See It?	304
8.7.3 How Do We Interpret It? With Diagnostic Groups!	305
8.8 Conclusion	306

<b>9 Seeing Missing Values</b>	<b>309</b>
9.1 Introduction	312
9.2 Data: Sleep in Mammals	314
9.3 Missing Data Visualization Tools	315
9.3.1 Missing Values Bar Charts	316
9.3.2 Histograms and Bar Charts	316
9.3.3 Boxplots	316
9.3.4 Scatterplots	316
9.4 Visualizing Imputed Values	317
9.4.1 Marking the Imputed Values	318
9.4.2 Single Imputation	320
9.4.3 Multiple Imputation	325
9.4.4 Summary of Imputation	327
9.5 Missing Data Patterns	327
9.5.1 Patterns and Number of Cases	328
9.5.2 The Mechanisms Leading to Missing Data	329
9.5.3 Visualizing Dynamically the Patterns of Missing Data	331
9.6 Conclusions	337
<b>References</b>	<b>339</b>
<b>Author Index</b>	<b>351</b>
<b>Subject Index</b>	<b>355</b>