

Visualizing Multivariate Data and Models in R

Here is where the dedication goes ...

Table of contents

Preface	v
ONE, TWO, MANY	v
Flatland	vi
EUREKA!	viii
What I assume	x
Conventions used in this book	x
1 Introduction	1
1.1 Multivariate vs. multivariable methods	1
1.2 Why use a multivariate design	1
1.3 Linear models: Univariate to multivariate	2
1.4 Visualization is harder	2
1.5 Problems in understanding and communicating MLM results	3
2 Getting Started	5
2.1 Why plot your data?	5
2.1.1 Anscombe's Quartet	5
2.1.2 A real example	8
2.2 Plots for data analysis	11
2.3 Data plots	12
2.4 Model plots	12
2.5 Diagnostic plots	12
2.6 Principles of graphical display	12
3 Plots of Multivariate Data	13
3.1 Bivariate summaries	13
3.1.1 Smoothers	15
3.1.2 Stratifiers	17
3.1.3 Conditioning	20
3.1.4 Data Ellipses	22
3.2 Scatterplot matrices	37
3.2.1 Visual thinning	44
3.2.2 Corrgrams	46
3.3 Generalized pairs plots	48
3.4 Parallel coordinate plots	54
4 PCA and Biplots	59
4.1 <i>Flatland</i> and <i>Spaceland</i>	59
4.1.1 Multivariate juicers	59
4.2 Principal components analysis	61
4.2.1 PCA by springs	62
4.2.2 Mathematics and geometry of PCA	64
4.2.3 Finding principal components	64
4.2.4 Visualizing variance proportions: screeplots	66

4.2.5	Visualizing PCA scores and variable vectors	68
4.3	Biplots	73
4.3.1	Constructing a biplot	73
4.3.2	Biplots in R	74
4.3.3	Example	75
4.3.4	Biplot contributions and quality	78
4.3.5	Supplementary variables	80
4.4	Application: Variable ordering for data displays	83
4.5	Application: Eigenfaces	87
4.6	Elliptical insights: Outlier detection	92
5	Overview of Linear models	95
5.1	Linear combinations	96
5.1.1	Multiple regression	97
5.1.2	Multivariate regression	97
5.1.3	Canonical correlation analysis	97
5.1.4	The General Linear Model	97
5.1.5	Model formulas	97
5.2	Regression	97
5.3	ANOVA	97
5.4	ANCOVA	97
5.5	Regression trees	97
6	Plots for univariate response models	101
6.1	The “regression quartet”	101
6.2	Other Diagnostic plots	106
6.2.1	Spread-level plot	106
6.3	Coefficient plots	106
6.4	Added-variable plots	106
6.5	Marginal plots	106
6.6	Outliers, leverage and influence	106
6.6.1	The leverage-influence quartet	106
6.6.2	Measuring leverage	109
6.6.3	Outliers: Measuring residuals	112
6.6.4	Measuring influence	112
7	Collinearity & Ridge Regression	115
7.1	What is collinearity?	115
7.1.1	Visualizing collinearity	116
7.1.2	Data space and β space	117
7.2	Measuring collinearity	119
7.2.1	Variance inflation factors	119
7.2.2	Collinearity diagnostics	122
7.2.3	Tableplots	123
7.2.4	Collinearity biplots	124
7.3	Remedies for collinearity: What can I do?	126
7.4	Ridge regression	130
7.4.1	What is ridge regression?	130
7.4.2	Univariate ridge trace plots	130
7.4.3	Bivariate ridge trace plots	130
8	Hotelling’s T^2	131
8.1	T^2 as a generalized t -test	131
8.2	T^2 properties	132