Beginner's Guide to Data Exploration and Visualisation with R

- Chapter 1: Introduction
  - 1.1 SPEAKING THE SAME LANGUAGE
  - 1.2. GENERAL POINTS
  - 1.3 OUTLINE OF THIS BOOK
- 2. Outliers
  - 2.1 WHAT IS AN OUTLIER?
  - 2.2 BOXPLOT TO IDENTIFY OUTLIERS IN ONE DIMENSION
    - 2.2.1 Simple boxplot
    - 2.2.2 Conditional boxplot
    - 2.2.3 Multi-panel boxplots from the lattice package
  - 2.3 CLEVELAND DOTPLOT TO IDENTIFY OUTLIERS
    - 2.3.1 Simple Cleveland dotplots
    - 2.3.2 Conditional Cleveland dotplots
    - 2.3.3 Multi-panel Cleveland dotplots from the lattice package
  - 2.4 BOXPLOTS OR CLEVELAND DOTPLOTS?
  - 2.5 CAN WE APPLY A TEST FOR OUTLIERS?
    - 2.5.1 Z-score
    - 2.5.2 Grubbs' test
  - 2.6 OUTLIERS IN THE TWO-DIMENSIONAL SPACE
  - 2.7 INFLUENTIAL OBSERVATIONS IN REGRESSION MODELS
  - 2.8 WHAT TO DO IF YOU DETECT POTENTIAL OUTLIERS
  - 2.9 OUTLIERS AND MULTIVARIATE DATA
  - 2.10 THE PROS AND CONS OF TRANSFORMATIONS
- 3 NORMALITY AND HOMOGENEITY
  - 3.1 WHAT IS NORMALITY?
  - 3.2 HISTOGRAMS AND CONDITIONAL HISTOGRAMS
    - 3.2.1 Multipanel histograms from the lattice package
    - 3.2.2 When is normality of the raw data considered?
  - 3.3 KERNEL DENSITY PLOTS
  - 3.4 QUANTILE-QUANTILE PLOTS
    - 3.4.1 Quantile-quantile plots from the lattice package
  - 3.5 USING TESTS TO CHECK FOR NORMALITY
  - 3.6 HOMOGENEITY OF VARIANCE
    - 3.6.1 Conditional boxplots
    - 3.6.2 Scatterplots for continuous explanatory variables
  - 3.7 USING TESTS TO CHECK FOR HOMOGENEITY
    - 3.7.1 The Bartlett test
    - 3.7.2 The F-ratio test
    - 3.7.3 Levene's test
    - 3.7.4 So which test would you choose?
- 4 RELATIONSHIPS
  - 4.1 SIMPLE SCATTERPLOTS
    - 4.1.1 Example: Clam data
    - 4.1.2 Example: Rabbit data
    - 4.1.3 Example: Blow fly data
  - 4.2 MULTIPANEL SCATTERPLOTS
    - 4.2.1 Example: Polychaeta data
    - 4.2.2 Example: Bioluminescence data
  - 4.3 PAIRPLOTS
    - 4.3.1 Bioluminescence data
    - 4.3.2 Cephalopod data
    - 4.3.3 Zoobenthos data
  - 4.4 CAN WE INCLUDE INTERACTIONS?
    - 4.4.1 Irish pH data
    - 4.4.2 Godwit data
    - 4.4.3 Irish pH data revisited
    - 4.4.4 Parasite data
  - 4.5 DESIGN AND INTERACTION PLOTS
- 5 COLLINEARITY AND CONFOUNDING
  - 5.1 WHAT IS COLLINEARITY?
  - 5.2 THE SAMPLE CORRELATION COEFFICIENT
  - 5.3 CORRELATION AND OUTLIERS
  - 5.4 CORRELATION MATRICES
  - 5.5 CORRELATION AND PAIRPLOTS
  - 5.6 COLLINEARITY DUE TO INTERACTIONS
  - 5.7 VISUALISING COLLINEARITY WITH CONDITIONAL BOXPLOTS
  - 5.8 QUANTIFYING COLLINEARITY USING VIFS
    - 5.8.1 Variance inflation factors
    - 5.8.2 Geometric presentation of collinearity
    - 5.8.3 Tolerance
    - 5.8.4 What constitutes a high VIF value?
    - 5.8.5 VIFs in action
  - 5.9 GENERALISED VIF VALUES
  - 5.10 VISUALISING COLLINEARITY USING PCA BIPLOTS
  - 5.11 CAUSES OF COLLINEARITY AND SOLUTIONS
  - 5.12 BE STUBBORN AND KEEP COLLINEAR COVARIATES?
  - 5.13 CONFOUNDING VARIABLES
    - 5.13.1 Visualising confounding variables
    - 5.13.2 Confounding factors in time series analysis
- 6 CASE STUDY: METHANE FLUXES
  - 6.1 INTRODUCTION
  - 6.2 DATA EXPLORATION
    - 6.2.1 Where in the world are the sites?
    - 6.2.2 Working with ggplot2
    - 6.2.3 Outliers
    - 6.2.4 Collinearity
    - 6.2.5 Relationships
    - 6.2.6 Interactions
    - 6.2.7 Where in the world are the sites (continued)?
  - 6.3 STATISTICAL ANALYSIS USING LINEAR REGRESSION
    - 6.3.1 Model formulation
    - 6.3.2 Fitting a linear regression model
    - 6.3.3 Model validation of the linear regression model
    - 6.3.4 Interpretation of the linear regression model
  - 6.4 STATISTICAL ANALYSIS USING A MIXED EFFECTS MODEL
    - 6.4.1 Model formulation
    - 6.4.2 Fitting a mixed effects model
    - 6.4.3 Model validation of the mixed effects model
    - 6.4.4 Interpretation of the linear mixed effects model
  - 6.5 CONCLUSIONS
  - 6.6 WHAT TO PRESENT IN A PAPER
- 7 CASE STUDY: OYSTERCATCHER SHELL LENGTH
  - 7.1 IMPORTING THE DATA
  - 7.2 DATA EXPLORATION
  - 7.3 APPLYING A LINEAR REGRESSION MODEL
  - 7.4 UNDERSTANDING THE RESULTS
  - 7.5 TROUBLE
  - 7.6 CONCLUSIONS
- 8 CASE STUDY: HAWAIIAN BIRD TIME SERIES
  - 8.1 IMPORTING THE DATA
  - 8.2 CODING THE DATA
    - 8.3.1 Attempt 1 using xyplot
    - 8.3.2 Attempt 2 using xyplot
    - 8.3.3 Attempt 3 using xyplot
  - 8.4 MULTI-PANEL GRAPH USING GGPLOT2
  - 8.5 CONCLUSIONS