

Contents

Part I Microanalysis

1	R Basics	3
2	First Foray into Text Analysis with R	11
3	Accessing and Comparing Word Frequency Data	25
4	Token Distribution Analysis	29
5	Correlation	47

Part II Mesoanalysis

6	Measures of Lexical Variety	59
7	<i>Hapax</i> Richness	69
8	Do It KWIC	73
9	Do It KWIC (Better)	81
10	Text Quality, Text Variety, and Parsing XML	89

Part III Macroanalysis

11	Clustering	101
12	Classification	119
13	Topic Modeling	135

A	Variable Scope Example	161
B	The LDA Buffet	163
C	Start up Code	167
D	R Resources for Further Reading	171
	Practice Exercise Solutions	173
	Index	193