

Table of Contents

Preface	ix
Chapter 1: Getting Started with R and Social Media Analytics	1
Understanding social media	2
Advantages and significance	4
Disadvantages and pitfalls	6
Social media analytics	7
A typical social media analytics workflow	8
Data access	9
Data processing and normalization	9
Data analysis	10
Insights	11
Opportunities	11
Challenges	12
Getting started with R	13
Environment setup	14
Data types	16
Data structures	18
Vectors	18
Arrays	20
Matrices	21
Lists	22
DataFrames	24
Functions	26
Built-in functions	26
User-defined functions	27
Controlling code flow	28
Looping constructs	28
Conditional constructs	29

Advanced operations	31
apply	31
lapply	33
sapply	33
tapply	34
mapply	35
Visualizing data	36
Next steps	38
Getting help	38
Managing packages	39
Data analytics	39
Analytics workflow	40
Machine learning	42
Machine learning techniques	42
Supervised learning	43
Unsupervised learning	44
Text analytics	44
Summary	45
Chapter 2: Twitter – What's Happening with 140 Characters	47
Understanding Twitter	48
APIs	49
Registering an application	50
Connecting to Twitter using R	53
Extracting sample Tweets	55
Revisiting analytics workflow	56
Trend analysis	56
Sentiment analysis	66
Key concepts of sentiment analysis	67
Subjectivity	67
Sentiment polarity	67
Opinion summarization	68
Features	68
Sentiment analysis in R	69
Follower graph analysis	79
Challenges	86
Summary	87
Chapter 3: Analyzing Social Networks and Brand Engagements with Facebook	89
Accessing Facebook data	91
Understanding the Graph API	91
Understanding Rfacebook	94

Understanding Netvizz	95
Data access challenges	95
Analyzing your personal social network	96
Basic descriptive statistics	97
Analyzing mutual interests	100
Build your friend network graph	102
Visualizing your friend network graph	103
Analyzing node properties	104
Degree	105
Closeness	107
Betweenness	108
Analyzing network communities	109
Cliques	109
Communities	110
Analyzing an English football social network	114
Basic descriptive statistics	116
Visualizing the network	119
Analyzing network properties	120
Diameter	121
Page distances	121
Density	122
Transitivity	123
Coreness	123
Analyzing node properties	124
Degree	125
Closeness	125
Betweenness	126
Visualizing correlation among centrality measures	127
Eigenvector centrality	129
PageRank	130
HITS authority score	131
Page neighbours	132
Analyzing network communities	133
Cliques	133
Communities	134
Analyzing English Football Club's brand page engagements	139
Getting the data	139
Curating the data	140
Visualizing post counts per page	141
Visualizing post counts by post type per page	142
Visualizing average likes by post type per page	143
Visualizing average shares by post type per page	144
Visualizing page engagement over time	145

Visualizing user engagement with page over time	146
Trending posts by user likes per page	148
Trending posts by user shares per page	149
Top influential users on popular page posts	150
Summary	152
Chapter 4: Foursquare – Are You Checked in Yet?	153
<hr/>	
Foursquare – the app and data	154
Foursquare APIs – show me the data	155
Creating an application – let me in	156
Data access – the twist in the story	157
Handling JSON in R – the hidden art	158
Getting category data – introduction to JSON parsing and data extraction	158
Revisiting the analytics workflow	163
Category trend analysis	163
Getting the data – the usual hurdle	163
The required end point	164
Getting data for a city – geometry to the rescue	164
Analysis – the fun part	167
Basic descriptive statistics – the usual	168
Recommendation engine – let's open a restaurant	174
Recommendation engine – the clichés	174
Framing the recommendation problem	174
Building our restaurant recommender	175
The sentimental rankings	180
Extracting tips data – the go to step	180
The actual data	182
Analysis of tips	183
Basic descriptive statistics	183
The final rankings	187
Venue graph – where do people go next?	189
Challenges for Foursquare data analysis	192
Summary	193
Chapter 5: Analyzing Software Collaboration	
Trends I – Social Coding with GitHub	195
<hr/>	
Environment setup	196
Understanding GitHub	197
Accessing GitHub data	200
Using the rgithub package for data access	200
Registering an application on GitHub	201
Accessing data using the GitHub API	203

Analyzing repository activity	206
Analyzing weekly commit frequency	206
Analyzing commit frequency distribution versus day of the week	208
Analyzing daily commit frequency	210
Analyzing weekly commit frequency comparison	211
Analyzing weekly code modification history	213
Retrieving trending repositories	215
Analyzing repository trends	218
Analyzing trending repositories created over time	219
Analyzing trending repositories updated over time	221
Analyzing repository metrics	223
Visualizing repository metric distributions	225
Analyzing repository metric correlations	226
Analyzing relationship between stargazer and repository counts	228
Analyzing relationship between stargazer and fork counts	229
Analyzing relationship between total forks, repository count, and health	232
Analyzing language trends	233
Visualizing top trending languages	233
Visualizing top trending languages over time	235
Analyzing languages with the most open issues	237
Analyzing languages with the most open issues over time	238
Analyzing languages with the most helpful repositories	240
Analyzing languages with the highest popularity score	242
Analyzing language correlations	244
Analyzing user trends	247
Visualizing top contributing users	247
Analyzing user activity metrics	249
Summary	253
Chapter 6: Analyzing Software Collaboration	
Trends II - Answering Your Questions with StackExchange	255
Understanding StackExchange	256
Data access	257
The StackExchange data dump	258
Accessing data dumps	259
Contents of data dumps	259
Quick overview of the data in data dumps	260
Getting started with data dumps	264
Data Science and StackExchange	265
Demographics and data science	274
Challenges	280
Summary	281

Chapter 7: Believe What You See – Flickr Data Analysis	283
A Flickr-ing world	283
Accessing Flickr's data	285
Creating the Flickr app	285
Connecting to R	288
Getting started with Flickr data	291
Understanding Flickr data	292
Understanding more about EXIF	293
Understanding interestingness – similarities	301
Finding K	302
Elbow method	302
Silhouette method	303
Are your photos interesting?	310
Preparing the data	310
Building the classifier	314
Challenges	317
Summary	318
Chapter 8: News – The Collective Social Media!	319
News data – news is everywhere	320
Accessing news data	321
Creating applications for data access	322
Data extraction – not just an API call	323
The API call and JSON monster	324
Sentiment trend analysis	331
Getting the data – not again	332
Basic descriptive statistics – the usual	333
Numerical sentiment trends	336
Emotion-based sentiment trends	339
Topic modeling	343
Getting to the data	344
Basic descriptive analysis	345
Topic modeling for Mr. Trump's phases	348
Cleaning the data	349
Pre-processing the data	349
The modeling part	350
Analysis of topics	351

Summarizing news articles	353
Document summarization	354
Understanding LexRank	354
Summarizing articles with lexRankr	355
Challenges to news data analysis	360
Summary	361
Index	363
