

# BRIEF CONTENTS

Foreword by Sarah Frostenson . . . . .	xvii
Acknowledgments . . . . .	xxi
Introduction . . . . .	xxiii
Chapter 1: Creating Your First Database and Table . . . . .	1
Chapter 2: Beginning Data Exploration with SELECT . . . . .	11
Chapter 3: Understanding Data Types . . . . .	23
Chapter 4: Importing and Exporting Data . . . . .	39
Chapter 5: Basic Math and Stats with SQL . . . . .	55
Chapter 6: Joining Tables in a Relational Database . . . . .	73
Chapter 7: Table Design That Works for You . . . . .	93
Chapter 8: Extracting Information by Grouping and Summarizing . . . . .	113
Chapter 9: Inspecting and Modifying Data . . . . .	129
Chapter 10: Statistical Functions in SQL . . . . .	155
Chapter 11: Working with Dates and Times . . . . .	171
Chapter 12: Advanced Query Techniques . . . . .	191
Chapter 13: Mining Text to Find Meaningful Data . . . . .	211
Chapter 14: Analyzing Spatial Data with PostGIS . . . . .	241
Chapter 15: Saving Time with Views, Functions, and Triggers . . . . .	267
Chapter 16: Using PostgreSQL from the Command Line . . . . .	291
Chapter 17: Maintaining Your Database . . . . .	313
Chapter 18: Identifying and Telling the Story Behind Your Data . . . . .	325
Appendix: Additional PostgreSQL Resources . . . . .	333
Index . . . . .	337



# CONTENTS IN DETAIL

<b>FOREWORD by Sarah Frostenson</b>	<b>xvii</b>
-------------------------------------	-------------

<b>ACKNOWLEDGMENTS</b>	<b>xxi</b>
------------------------	------------

<b>INTRODUCTION</b>	<b>xxiii</b>
---------------------	--------------

What Is SQL? . . . . .	.xxiv
Why Use SQL? . . . . .	.xxiv
About This Book. . . . .	.xxv
Using the Book's Code Examples . . . . .	.xxvii
Using PostgreSQL. . . . .	.xxviii
Installing PostgreSQL. . . . .	.xxviii
Working with pgAdmin. . . . .	.xxxii
Alternatives to pgAdmin. . . . .	.xxxiii
Wrapping Up . . . . .	.xxxiii

<b>1</b>	
<b>CREATING YOUR FIRST DATABASE AND TABLE</b>	<b>1</b>

Creating a Database . . . . .	3
Executing SQL in pgAdmin. . . . .	3
Connecting to the Analysis Database. . . . .	5
Creating a Table . . . . .	5
The CREATE TABLE Statement. . . . .	6
Making the teachers Table . . . . .	7
Inserting Rows into a Table . . . . .	8
The INSERT Statement . . . . .	8
Viewing the Data . . . . .	9
When Code Goes Bad . . . . .	9
Formatting SQL for Readability . . . . .	10
Wrapping Up . . . . .	10
Try It Yourself. . . . .	10

<b>2</b>	
<b>BEGINNING DATA EXPLORATION WITH SELECT</b>	<b>11</b>

Basic SELECT Syntax . . . . .	12
Querying a Subset of Columns . . . . .	13
Using DISTINCT to Find Unique Values . . . . .	14
Sorting Data with ORDER BY . . . . .	15
Filtering Rows with WHERE . . . . .	17
Using LIKE and ILIKE with WHERE . . . . .	19
Combining Operators with AND and OR. . . . .	20
Putting It All Together . . . . .	21
Wrapping Up . . . . .	21
Try It Yourself. . . . .	22

<b>3</b>	<b>UNDERSTANDING DATA TYPES</b>	<b>23</b>
	Characters . . . . .	24
	Numbers. . . . .	26
	Integers . . . . .	27
	Auto-Incrementing Integers . . . . .	27
	Decimal Numbers . . . . .	28
	Choosing Your Number Data Type . . . . .	31
	Dates and Times . . . . .	32
	Using the interval Data Type in Calculations . . . . .	34
	Miscellaneous Types. . . . .	35
	Transforming Values from One Type to Another with CAST . . . . .	35
	CAST Shortcut Notation . . . . .	36
	Wrapping Up . . . . .	36
	Try It Yourself. . . . .	37
<b>4</b>	<b>IMPORTING AND EXPORTING DATA</b>	<b>39</b>
	Working with Delimited Text Files . . . . .	40
	Quoting Columns that Contain Delimiters . . . . .	41
	Handling Header Rows . . . . .	41
	Using COPY to Import Data. . . . .	42
	Importing Census Data Describing Counties . . . . .	43
	Creating the us_counties_2010 Table . . . . .	44
	Census Columns and Data Types . . . . .	45
	Performing the Census Import with COPY . . . . .	47
	Importing a Subset of Columns with COPY . . . . .	49
	Adding a Default Value to a Column During Import . . . . .	50
	Using COPY to Export Data. . . . .	51
	Exporting All Data. . . . .	51
	Exporting Particular Columns . . . . .	52
	Exporting Query Results . . . . .	52
	Importing and Exporting Through pgAdmin. . . . .	52
	Wrapping Up . . . . .	53
	Try It Yourself. . . . .	54
<b>5</b>	<b>BASIC MATH AND STATS WITH SQL</b>	<b>55</b>
	Math Operators . . . . .	56
	Math and Data Types . . . . .	56
	Adding, Subtracting, and Multiplying . . . . .	57
	Division and Modulo . . . . .	57
	Exponents, Roots, and Factorials. . . . .	58
	Minding the Order of Operations . . . . .	59
	Doing Math Across Census Table Columns . . . . .	60
	Adding and Subtracting Columns . . . . .	60
	Finding Percentages of the Whole. . . . .	62
	Tracking Percent Change . . . . .	63
	Aggregate Functions for Averages and Sums. . . . .	64

Finding the Median . . . . .	65
Finding the Median with Percentile Functions . . . . .	66
Median and Percentiles with Census Data . . . . .	67
Finding Other Quantiles with Percentile Functions . . . . .	67
Creating a median() Function . . . . .	69
Finding the Mode. . . . .	70
Wrapping Up . . . . .	71
Try It Yourself. . . . .	71

## **6 JOINING TABLES IN A RELATIONAL DATABASE 73**

Linking Tables Using JOIN. . . . .	74
Relating Tables with Key Columns . . . . .	74
Querying Multiple Tables Using JOIN. . . . .	77
JOIN Types . . . . .	78
JOIN . . . . .	80
LEFT JOIN and RIGHT JOIN . . . . .	80
FULL OUTER JOIN . . . . .	82
CROSS JOIN . . . . .	82
Using NULL to Find Rows with Missing Values . . . . .	83
Three Types of Table Relationships . . . . .	84
One-to-One Relationship . . . . .	84
One-to-Many Relationship . . . . .	84
Many-to-Many Relationship . . . . .	85
Selecting Specific Columns in a Join. . . . .	85
Simplifying JOIN Syntax with Table Aliases. . . . .	86
Joining Multiple Tables . . . . .	87
Performing Math on Joined Table Columns . . . . .	88
Wrapping Up . . . . .	90
Try It Yourself. . . . .	91

## **7 TABLE DESIGN THAT WORKS FOR YOU 93**

Naming Tables, Columns, and Other Identifiers . . . . .	94
Using Quotes Around Identifiers to Enable Mixed Case . . . . .	94
Pitfalls with Quoting Identifiers . . . . .	95
Guidelines for Naming Identifiers . . . . .	96
Controlling Column Values with Constraints. . . . .	96
Primary Keys: Natural vs. Surrogate . . . . .	97
Foreign Keys . . . . .	102
Automatically Deleting Related Records with CASCADE. . . . .	104
The CHECK Constraint. . . . .	104
The UNIQUE Constraint. . . . .	105
The NOT NULL Constraint . . . . .	106
Removing Constraints or Adding Them Later. . . . .	107
Speeding Up Queries with Indexes. . . . .	108
B-Tree: PostgreSQL's Default Index . . . . .	108
Considerations When Using Indexes . . . . .	111
Wrapping Up . . . . .	111
Try It Yourself. . . . .	112

## **8 EXTRACTING INFORMATION BY GROUPING AND SUMMARIZING 113**

Creating the Library Survey Tables . . . . .	114
Creating the 2014 Library Data Table . . . . .	114
Creating the 2009 Library Data Table . . . . .	116
Exploring the Library Data Using Aggregate Functions . . . . .	117
Counting Rows and Values Using count() . . . . .	117
Finding Maximum and Minimum Values Using max() and min() . . . . .	119
Aggregating Data Using GROUP BY . . . . .	120
Wrapping Up . . . . .	128
Try It Yourself . . . . .	128

## **9 INSPECTING AND MODIFYING DATA 129**

Importing Data on Meat, Poultry, and Egg Producers . . . . .	130
Interviewing the Data Set . . . . .	131
Checking for Missing Values . . . . .	132
Checking for Inconsistent Data Values . . . . .	134
Checking for Malformed Values Using length() . . . . .	135
Modifying Tables, Columns, and Data . . . . .	136
Modifying Tables with ALTER TABLE . . . . .	137
Modifying Values with UPDATE . . . . .	138
Creating Backup Tables . . . . .	139
Restoring Missing Column Values . . . . .	140
Updating Values for Consistency . . . . .	142
Repairing ZIP Codes Using Concatenation . . . . .	143
Updating Values Across Tables . . . . .	145
Deleting Unnecessary Data . . . . .	147
Deleting Rows from a Table . . . . .	147
Deleting a Column from a Table . . . . .	148
Deleting a Table from a Database . . . . .	148
Using Transaction Blocks to Save or Revert Changes . . . . .	149
Improving Performance When Updating Large Tables . . . . .	151
Wrapping Up . . . . .	152
Try It Yourself . . . . .	152

## **10 STATISTICAL FUNCTIONS IN SQL 155**

Creating a Census Stats Table . . . . .	156
Measuring Correlation with corr(Y, X) . . . . .	157
Checking Additional Correlations . . . . .	159
Predicting Values with Regression Analysis . . . . .	160
Finding the Effect of an Independent Variable with r-squared . . . . .	163
Creating Rankings with SQL . . . . .	164
Ranking with rank() and dense_rank() . . . . .	164
Ranking Within Subgroups with PARTITION BY . . . . .	165
Calculating Rates for Meaningful Comparisons . . . . .	167

Wrapping Up . . . . .	169
Try It Yourself. . . . .	169

## **11 WORKING WITH DATES AND TIMES 171**

Data Types and Functions for Dates and Times . . . . .	172
Manipulating Dates and Times. . . . .	172
Extracting the Components of a timestamp Value . . . . .	173
Creating Datetime Values from timestamp Components . . . . .	174
Retrieving the Current Date and Time. . . . .	175
Working with Time Zones . . . . .	177
Finding Your Time Zone Setting . . . . .	177
Setting the Time Zone . . . . .	178
Calculations with Dates and Times . . . . .	180
Finding Patterns in New York City Taxi Data . . . . .	180
Finding Patterns in Amtrak Data . . . . .	186
Wrapping Up . . . . .	189
Try It Yourself. . . . .	190

## **12 ADVANCED QUERY TECHNIQUES 191**

Using Subqueries. . . . .	192
Filtering with Subqueries in a WHERE Clause. . . . .	192
Creating Derived Tables with Subqueries . . . . .	194
Joining Derived Tables. . . . .	195
Generating Columns with Subqueries . . . . .	197
Subquery Expressions . . . . .	198
Common Table Expressions. . . . .	200
Cross Tabulations. . . . .	203
Installing the crosstab() Function . . . . .	203
Tabulating Survey Results. . . . .	203
Tabulating City Temperature Readings. . . . .	205
Reclassifying Values with CASE . . . . .	207
Using CASE in a Common Table Expression . . . . .	209
Wrapping Up . . . . .	210
Try It Yourself. . . . .	210

## **13 MINING TEXT TO FIND MEANINGFUL DATA 211**

Formatting Text Using String Functions . . . . .	212
Case Formatting . . . . .	212
Character Information . . . . .	212
Removing Characters . . . . .	213
Extracting and Replacing Characters . . . . .	213
Matching Text Patterns with Regular Expressions . . . . .	214
Regular Expression Notation. . . . .	214
Turning Text to Data with Regular Expression Functions . . . . .	216
Using Regular Expressions with WHERE. . . . .	228
Additional Regular Expression Functions . . . . .	230

Full Text Search in PostgreSQL . . . . .	231
Text Search Data Types . . . . .	231
Creating a Table for Full Text Search . . . . .	233
Searching Speech Text . . . . .	234
Ranking Query Matches by Relevance . . . . .	237
Wrapping Up . . . . .	239
Try It Yourself . . . . .	239

## 14

### **ANALYZING SPATIAL DATA WITH POSTGIS 241**

Installing PostGIS and Creating a Spatial Database . . . . .	242
The Building Blocks of Spatial Data . . . . .	243
Two-Dimensional Geometries . . . . .	243
Well-Known Text Formats . . . . .	244
A Note on Coordinate Systems . . . . .	245
Spatial Reference System Identifier . . . . .	246
PostGIS Data Types . . . . .	247
Creating Spatial Objects with PostGIS Functions . . . . .	247
Creating a Geometry Type from Well-Known Text . . . . .	247
Creating a Geography Type from Well-Known Text . . . . .	248
Point Functions . . . . .	249
LineString Functions . . . . .	249
Polygon Functions . . . . .	250
Analyzing Farmers' Markets Data . . . . .	250
Creating and Filling a Geography Column . . . . .	251
Adding a GiST Index . . . . .	252
Finding Geographies Within a Given Distance . . . . .	253
Finding the Distance Between Geographies . . . . .	254
Working with Census Shapefiles . . . . .	256
Contents of a Shapefile . . . . .	256
Loading Shapefiles via the GUI Tool . . . . .	257
Exploring the Census 2010 Counties Shapefile . . . . .	259
Performing Spatial Joins . . . . .	262
Exploring Roads and Waterways Data . . . . .	262
Joining the Census Roads and Water Tables . . . . .	263
Finding the Location Where Objects Intersect . . . . .	264
Wrapping Up . . . . .	265
Try It Yourself . . . . .	265

## 15

### **SAVING TIME WITH VIEWS, FUNCTIONS, AND TRIGGERS 267**

Using Views to Simplify Queries . . . . .	268
Creating and Querying Views . . . . .	269
Inserting, Updating, and Deleting Data Using a View . . . . .	271
Programming Your Own Functions . . . . .	275
Creating the percent_change() Function . . . . .	276
Using the percent_change() Function . . . . .	277
Updating Data with a Function . . . . .	278
Using the Python Language in a Function . . . . .	281



Automating Database Actions with Triggers . . . . .	282
Logging Grade Updates to a Table . . . . .	282
Automatically Classifying Temperatures . . . . .	286
Wrapping Up . . . . .	289
Try It Yourself . . . . .	289

## **16 USING POSTGRESQL FROM THE COMMAND LINE 291**

Setting Up the Command Line for psql . . . . .	292
Windows psql Setup . . . . .	292
macOS psql Setup . . . . .	296
Linux psql Setup . . . . .	299
Working with psql . . . . .	299
Launching psql and Connecting to a Database . . . . .	299
Getting Help . . . . .	300
Changing the User and Database Connection . . . . .	300
Running SQL Queries on psql . . . . .	301
Navigating and Formatting Results . . . . .	303
Meta-Commands for Database Information . . . . .	306
Importing, Exporting, and Using Files . . . . .	307
Additional Command Line Utilities to Expedite Tasks . . . . .	310
Adding a Database with createdb . . . . .	310
Loading Shapefiles with shp2psql . . . . .	311
Wrapping Up . . . . .	311
Try It Yourself . . . . .	312

## **17 MAINTAINING YOUR DATABASE 313**

Recovering Unused Space with VACUUM . . . . .	314
Tracking Table Size . . . . .	314
Monitoring the autovacuum Process . . . . .	316
Running VACUUM Manually . . . . .	318
Reducing Table Size with VACUUM FULL . . . . .	318
Changing Server Settings . . . . .	318
Locating and Editing postgresql.conf . . . . .	319
Reloading Settings with pg_ctl . . . . .	321
Backing Up and Restoring Your Database . . . . .	321
Using pg_dump to Back Up a Database or Table . . . . .	321
Restoring a Database Backup with pg_restore . . . . .	322
Additional Backup and Restore Options . . . . .	323
Wrapping Up . . . . .	323
Try It Yourself . . . . .	323

## **18 IDENTIFYING AND TELLING THE STORY BEHIND YOUR DATA 325**

Start with a Question . . . . .	326
Document Your Process . . . . .	326
Gather Your Data . . . . .	326
No Data? Build Your Own Database . . . . .	327

Assess the Data's Origins . . . . .	328
Interview the Data with Queries . . . . .	328
Consult the Data's Owner . . . . .	328
Identify Key Indicators and Trends over Time . . . . .	329
Ask Why. . . . .	331
Communicate Your Findings . . . . .	331
Wrapping Up . . . . .	332
Try It Yourself. . . . .	332

**APPENDIX**  
**ADDITIONAL POSTGRESQL RESOURCES** **333**

PostgreSQL Development Environments . . . . .	333
PostgreSQL Utilities, Tools, and Extensions . . . . .	334
PostgreSQL News . . . . .	335
Documentation. . . . .	335

**INDEX** **337**