

CONTENTS IN DETAIL

PREFACE TO THE SECOND EDITION	xvii
--------------------------------------	-------------

ACKNOWLEDGMENTS	xxi
------------------------	------------

INTRODUCTION	xxiii
---------------------	--------------

What Is SQL?	xxiv
Why SQL?	xxv
Who Is This Book For?	xxvi
What You'll Learn.	xxvi

1	
SETTING UP YOUR CODING ENVIRONMENT	1

Installing a Text Editor	2
Downloading Code and Data from GitHub	3
Installing PostgreSQL and pgAdmin	3
Windows Installation	4
macOS Installation	8
Linux Installation	9
Working with pgAdmin.	11
Launching pgAdmin and Setting a Master Password	11
Connecting to the Default postgres Database	12
Exploring the Query Tool	14
Customizing pgAdmin	15
Alternatives to pgAdmin	15
Wrapping Up	16

2	
CREATING YOUR FIRST DATABASE AND TABLE	17

Understanding Tables	17
Creating a Database	19
Executing SQL in pgAdmin.	20
Connecting to the analysis Database	22
Creating a Table	22
Using the CREATE TABLE Statement	22
Making the teachers Table	23
Inserting Rows into a Table	25
Using the INSERT Statement	25
Viewing the Data	26
Getting Help When Code Goes Bad	26
Formatting SQL for Readability	27
Wrapping Up	27

3		
BEGINNING DATA EXPLORATION WITH SELECT		29
Basic SELECT Syntax		30
Querying a Subset of Columns		31
Sorting Data with ORDER BY		32
Using DISTINCT to Find Unique Values		33
Filtering Rows with WHERE		35
Using LIKE and ILIKE with WHERE		37
Combining Operators with AND and OR		38
Putting It All Together		39
Wrapping Up		39
4		
UNDERSTANDING DATA TYPES		41
Understanding Characters		42
Understanding Numbers		44
Using Integers		45
Auto-Incrementing Integers		46
Using Decimal Numbers		47
Choosing Your Number Data Type		50
Understanding Dates and Times		51
Using the interval Data Type in Calculations		53
Understanding JSON and JSONB		54
Using Miscellaneous Types		55
Transforming Values from One Type to Another with CAST		55
Using CAST Shortcut Notation		56
Wrapping Up		56
5		
IMPORTING AND EXPORTING DATA		59
Working with Delimited Text Files		60
Handling Header Rows		61
Quoting Columns That Contain Delimiters		61
Using COPY to Import Data		62
Importing Census Data Describing Counties		63
Creating the us_counties_pop_est_2019 Table		64
Understanding Census Columns and Data Types		65
Performing the Census Import with COPY		67
Inspecting the Import		67
Importing a Subset of Columns with COPY		68
Importing a Subset of Rows with COPY		70
Adding a Value to a Column During Import		71
Using COPY to Export Data		72
Exporting All Data		72
Exporting Particular Columns		73
Exporting Query Results		73
Importing and Exporting Through pgAdmin		74
Wrapping Up		75

6		
BASIC MATH AND STATS WITH SQL		77
Understanding Math Operators and Functions		78
Understanding Math and Data Types		78
Adding, Subtracting, and Multiplying		79
Performing Division and Modulo		79
Using Exponents, Roots, and Factorials		80
Minding the Order of Operations		81
Doing Math Across Census Table Columns		82
Adding and Subtracting Columns		82
Finding Percentages of the Whole		84
Tracking Percent Change		85
Using Aggregate Functions for Averages and Sums		86
Finding the Median		87
Finding the Median with Percentile Functions		88
Finding Median and Percentiles with Census Data		89
Finding Other Quantiles with Percentile Functions		90
Finding the Mode		91
Wrapping Up		92
7		
JOINING TABLES IN A RELATIONAL DATABASE		93
Linking Tables Using JOIN		94
Relating Tables with Key Columns		94
Querying Multiple Tables Using JOIN		97
Understanding JOIN Types		98
JOIN		100
LEFT JOIN and RIGHT JOIN		101
FULL OUTER JOIN		102
CROSS JOIN		103
Using NULL to Find Rows with Missing Values		104
Understanding the Three Types of Table Relationships		105
One-to-One Relationship		105
One-to-Many Relationship		105
Many-to-Many Relationship		105
Selecting Specific Columns in a Join		106
Simplifying JOIN Syntax with Table Aliases		107
Joining Multiple Tables		107
Combining Query Results with Set Operators		109
UNION and UNION ALL		109
INTERSECT and EXCEPT		111
Performing Math on Joined Table Columns		112
Wrapping Up		114
8		
TABLE DESIGN THAT WORKS FOR YOU		117
Following Naming Conventions		118
Quoting Identifiers Enables Mixed Case		118
Pitfalls with Quoting Identifiers		119
Guidelines for Naming Identifiers		120

Controlling Column Values with Constraints	120
Primary Keys: Natural vs. Surrogate	121
Foreign Keys	128
How to Automatically Delete Related Records with CASCADE	129
The CHECK Constraint	130
The UNIQUE Constraint	131
The NOT NULL Constraint	132
How to Remove Constraints or Add Them Later	132
Speeding Up Queries with Indexes	133
B-Tree: PostgreSQL's Default Index	134
Considerations When Using Indexes	137
Wrapping Up	137

9

EXTRACTING INFORMATION BY GROUPING AND SUMMARIZING

139

Creating the Library Survey Tables	140
Creating the 2018 Library Data Table	140
Creating the 2017 and 2016 Library Data Tables	141
Exploring the Library Data Using Aggregate Functions	142
Counting Rows and Values Using count()	143
Finding Maximum and Minimum Values Using max() and min()	145
Aggregating Data Using GROUP BY	146
Wrapping Up	154

10

INSPECTING AND MODIFYING DATA

157

Importing Data on Meat, Poultry, and Egg Producers	158
Interviewing the Dataset	159
Checking for Missing Values	160
Checking for Inconsistent Data Values	162
Checking for Malformed Values Using length()	162
Modifying Tables, Columns, and Data	164
Modifying Tables with ALTER TABLE	165
Modifying Values with UPDATE	166
Viewing Modified Data with RETURNING	167
Creating Backup Tables	167
Restoring Missing Column Values	168
Updating Values for Consistency	170
Repairing ZIP Codes Using Concatenation	172
Updating Values Across Tables	173
Deleting Unneeded Data	175
Deleting Rows from a Table	175
Deleting a Column from a Table	176
Deleting a Table from a Database	176
Using Transactions to Save or Revert Changes	177
Improving Performance When Updating Large Tables	179
Wrapping Up	180

11		
STATISTICAL FUNCTIONS IN SQL		183
Creating a Census Stats Table		184
Measuring Correlation with corr(Y, X)		185
Checking Additional Correlations		188
Predicting Values with Regression Analysis		189
Finding the Effect of an Independent Variable with r-Squared		191
Finding Variance and Standard Deviation		192
Creating Rankings with SQL		193
Ranking with rank() and dense_rank()		193
Ranking Within Subgroups with PARTITION BY		195
Calculating Rates for Meaningful Comparisons		196
Finding Rates of Tourism-Related Businesses		197
Smoothing Uneven Data		198
Wrapping Up		202
12		
WORKING WITH DATES AND TIMES		203
Understanding Data Types and Functions for Dates and Times		204
Manipulating Dates and Times		204
Extracting the Components of a timestamp Value		205
Creating Datetime Values from timestamp Components		207
Retrieving the Current Date and Time		208
Working with Time Zones		209
Finding Your Time Zone Setting		209
Setting the Time Zone		210
Performing Calculations with Dates and Times		212
Finding Patterns in New York City Taxi Data		213
Finding Patterns in Amtrak Data		218
Wrapping Up		222
13		
ADVANCED QUERY TECHNIQUES		223
Using Subqueries		223
Filtering with Subqueries in a WHERE Clause		224
Creating Derived Tables with Subqueries		225
Joining Derived Tables		226
Generating Columns with Subqueries		228
Understanding Subquery Expressions		229
Using Subqueries with LATERAL		231
Using Common Table Expressions		234
Performing Cross Tabulations		236
Installing the crosstab() Function		237
Tabulating Survey Results		237
Tabulating City Temperature Readings		239
Reclassifying Values with CASE		241
Using CASE in a Common Table Expression		242
Wrapping Up		244

14
MINING TEXT TO FIND MEANINGFUL DATA **245**

Formatting Text Using String Functions	246
Case Formatting	246
Character Information	246
Removing Characters.	247
Extracting and Replacing Characters.	247
Matching Text Patterns with Regular Expressions	247
Regular Expression Notation.	248
Using Regular Expressions with WHERE.	250
Regular Expression Functions to Replace or Split Text.	251
Turning Text to Data with Regular Expression Functions	253
Full-Text Search in PostgreSQL	265
Text Search Data Types	265
Creating a Table for Full-Text Search	267
Searching Speech Text.	268
Ranking Query Matches by Relevance.	271
Wrapping Up	273

15
ANALYZING SPATIAL DATA WITH POSTGIS **275**

Enabling PostGIS and Creating a Spatial Database	276
Understanding the Building Blocks of Spatial Data	276
Understanding Two-Dimensional Geometries	277
Well-Known Text Formats	278
Projections and Coordinate Systems	279
Spatial Reference System Identifier	279
Understanding PostGIS Data Types.	280
Creating Spatial Objects with PostGIS Functions	281
Creating a Geometry Type from Well-Known Text	281
Creating a Geography Type from Well-Known Text	283
Using Point Functions.	283
Using LineString Functions	284
Using Polygon Functions	284
Analyzing Farmers' Markets Data	285
Creating and Filling a Geography Column.	285
Adding a Spatial Index	286
Finding Geographies Within a Given Distance	287
Finding the Distance Between Geographies	289
Finding the Nearest Geographies	290
Working with Census Shapefiles	291
Understanding the Contents of a Shapefile	292
Loading Shapefiles	292
Exploring the Census 2019 Counties Shapefile.	295
Examining Demographics Within a Distance	298
Performing Spatial Joins	300
Exploring Roads and Waterways Data	300
Joining the Census Roads and Water Tables	300
Finding the Location Where Objects Intersect	301
Wrapping Up	302

16		
WORKING WITH JSON DATA		305
Understanding JSON Structure		306
Considering When to Use JSON with SQL		307
Using json and jsonb Data Types		308
Importing and Indexing JSON Data		309
Using json and jsonb Extraction Operators		310
Key Value Extraction		311
Array Element Extraction		312
Path Extraction		314
Containment and Existence		315
Analyzing Earthquake Data		318
Exploring and Loading the Earthquake Data		318
Working with Earthquake Times		319
Finding the Largest and Most-Reported Earthquakes		321
Converting Earthquake JSON to Spatial Data		323
Generating and Manipulating JSON		327
Turning Query Results into JSON		327
Adding, Updating, and Deleting Keys and Values		329
Using JSON Processing Functions		331
Finding the Length of an Array		331
Returning Array Elements as Rows		332
Wrapping Up		333
17		
SAVING TIME WITH VIEWS, FUNCTIONS, AND TRIGGERS		337
Using Views to Simplify Queries		338
Creating and Querying Views		338
Creating and Refreshing a Materialized View		341
Inserting, Updating, and Deleting Data Using a View		342
Creating Your Own Functions and Procedures		346
Creating the percent_change() Function		347
Using the percent_change() Function		348
Updating Data with a Procedure		349
Using the Python Language in a Function		352
Automating Database Actions with Triggers		354
Logging Grade Updates to a Table		354
Automatically Classifying Temperatures		358
Wrapping Up		360
18		
USING POSTGRESQL FROM THE COMMAND LINE		363
Setting Up the Command Line for psql		364
Windows psql Setup		364
macOS psql Setup		368
Linux psql Setup		370
Working with psql		370
Launching psql and Connecting to a Database		370
Running SQL Queries on psql		374
Navigating and Formatting Results		375

Meta-Commands for Database Information	379
Importing, Exporting, and Using Files	380
Additional Command Line Utilities to Expedite Tasks	383
Adding a Database with createdb	383
Loading Shapefiles with shp2pgsql	383
Wrapping Up	384

19
MAINTAINING YOUR DATABASE **387**

Recovering Unused Space with VACUUM	388
Tracking Table Size	388
Monitoring the Autovacuum Process	390
Running VACUUM Manually	392
Reducing Table Size with VACUUM FULL	392
Changing Server Settings	392
Locating and Editing postgresql.conf	393
Reloading Settings with pg_ctl	395
Backing Up and Restoring Your Database	395
Using pg_dump to Export a Database or Table	395
Restoring a Database Export with pg_restore	396
Exploring Additional Backup and Restore Options	397
Wrapping Up	397

20
TELLING YOUR DATA’S STORY **399**

Start with a Question	400
Document Your Process	400
Gather Your Data	400
No Data? Build Your Own Database	401
Assess the Data’s Origins	402
Interview the Data with Queries	402
Consult the Data’s Owner	403
Identify Key Indicators and Trends over Time	403
Ask Why.	405
Communicate Your Findings	405
Wrapping Up	406

APPENDIX
ADDITIONAL POSTGRESQL RESOURCES **407**

PostgreSQL Development Environments	407
PostgreSQL Utilities, Tools, and Extensions	408
PostgreSQL News and Community	409
Documentation	410

INDEX **411**