

INDEX

Symbols

- + (addition operator), 56, 57
- & (ampersand operator), 232, 236
- * (asterisk)
 - as multiplication operator, 56, 57
 - as wildcard in SELECT, 12
- \ (backslash), 42–43, 215
 - escaping characters with, 219
- , (comma), 40
- ||/ (cube root operator), 56, 58
- { } (curly brackets), 215
 - denoting an array, 68
- <-> (distance operator), 232, 236
- @@ (double at sign match operator), 232
- :: (double-colon CAST operator), 36
- \$\$ (double-dollar quoting), 280
- || (double-pipe concatenation operator), 143, 225
- " (double quote), 41, 94
- = (equals comparison operator), 18
- ! (exclamation point)
 - as factorial operator, 56, 59
 - as negation, 228, 232, 236
- ^ (exponentiation operator), 56, 58
- / (forward slash)
 - as division operator, 56, 57
 - in macOS file paths, 42
- > (greater than comparison operator), 18
- >= (greater than or equals comparison operator), 18
- (hyphen subtraction operator), 56, 57
- < (less than comparison operator), 18
- <= (less than or equals comparison operator), 18
- != (not equal comparison operator), 18
- <> (not equal comparison operator), 18
- () (parentheses), 6, 8
 - to designate order of operations, 20
 - to specify columns for importing, 50
- % (percent sign)
 - as modulo operator, 56, 57
 - wildcard for pattern matching, 19
- | (pipe character)
 - as delimiter, 26, 43
 - to redirect output, 311
- ; (semicolon), 3
- ' (single quote), 8, 42
- |/ (square root operator), 56, 58
- ~* (tilde-asterisk case-insensitive matching operator), 228
- ~ (tilde case-sensitive matching operator), 228
- _ (underscore wildcard for pattern matching), 19

A

- adding numbers, 57
 - across columns, 60
- addition operator (+), 56, 57
- aggregate functions, 64, 117
 - avg(), 64
 - binary (two-input), 158
 - count(), 117–119, 131
 - filtering with HAVING, 127
 - interviewing data, 131
 - max(), 119–120
 - min(), 119–120
 - PostgreSQL documentation, 117
 - sum(), 64, 124–125
 - using GROUP BY clause, 120–123
- aliases for table names, 86, 125
- ALTER COLUMN statement, 107
- ALTER TABLE statement, 137
 - ADD COLUMN, 137, 252
 - ADD CONSTRAINT, 107
 - ALTER COLUMN, 137
 - DROP COLUMN, 137, 148
 - table constraints, adding and removing, 107

- American National Standards Institute (ANSI), xxiv
- ampersand operator (&), 232, 236
- ANALYZE keyword
 - with EXPLAIN command, 109
 - with VACUUM command, 317
- AND operator, 20
- ANSI (American National Standards Institute), xxiv
- antimeridian, 46
- array, 68
 - array_length() function, 230
 - functions, 68
 - notation in query, 224
 - passing into ST_MakePoint(), 250
 - returned from regexp_match(), 219, 224
 - type indicated in results grid, 224
 - unnest() function, 68
 - with curly brackets, 68, 220
- array_length() function, 230
- AS keyword
 - declaring table aliases with, 86, 90
 - renaming columns in query results with, 60, 61, 205
- ASC keyword, 15
- asterisk (*)
 - as multiplication operator, 56, 57
 - as wildcard in SELECT statement, 12
- attribute, 5
- auto-incrementing integers, 27
 - as surrogate primary key, 101
 - gaps in sequence, 28
 - identity column SQL standard, 27
- autovacuum, 316
 - editing server setting, 319
 - time of last vacuum, 317
- average, 64
 - vs. median, 65, 194
- avg() function, 64, 195

B

- backslash (\), 42–43, 215
 - escaping characters with, 219
- backups
 - column, 140
 - improving performance when updating tables, 151–152
 - restoring from copied table, 142
 - tables, 139
- BETWEEN comparison operator, 18, 198
 - inclusive property, 19
- bigint integer data type, 27
- bigserial integer data type, 6, 27, 101
 - as surrogate primary key, 102
- binary aggregate functions, 158
- BINARY file format, 42
- birth data, U.S., 330
- Boolean value, 74
- B-Tree (balanced tree) index, 108

C

- camel case, 10, 94
- caret symbol (^) exponentiation operator, 58
- carriage return, 43
- Cartesian Product
 - as result of CROSS JOIN, 82
- CASCADE keyword, 104
- case sensitivity
 - with ILIKE operator, 19
 - with LIKE operator, 19
- CASE statement, 207
 - ELSE clause, 208
 - in Common Table Expression, 209–210
 - in UPDATE statement, 226
 - syntax, 207
 - WHEN clause, 208, 288
 - with trigger, 286
- CAST() function, 35
 - shortcut notation, 36
- categorizing data, 207
- char character string type, 24
- character set, 16
- character string types, 24–26
 - char, 24
 - functional difference from number types, 26
 - performance in PostgreSQL, 25
 - text, 25
 - varchar, 24
- character varying data type. *See* varchar data type
- char_length() function, 212
- CHECK constraint, 104–105
- classify_max_temp() user function, 287
- clock_timestamp() function, 176
- Codd, Edgar F., xxiv, 73
- coefficient of determination. *See* *r*-squared

- collation setting, 16
- column, 5
 - adding numbers in, 64
 - alias, 60
 - alter data type, 137
 - averaging values in, 64
 - avoiding spaces in name, 95
 - deleting, 148
 - indexes, 110
 - naming, 94
 - populating new during backup, 151
 - retrieving in queries, 13
 - updating values, 138
- comma (,), 40
- comma-delimited files. *See* CSV
 - (comma-separated values)
- command line, 291
 - advantages of using, 292
 - createdb command, 310
 - psql application, 299
 - setup, 292
 - macOS, 296
 - PATH environment variable, 292, 296
 - Windows, 292
 - shell programs, 296
- comma-separated values (CSV). *See* CSV
- comments in code, xxvii
- COMMIT statement, 149
- Common Table Expression (CTE), 200
 - advantages, 201
 - CASE statement example, 209
 - definition, 200
- comparison operators, 18
 - combining with AND and OR, 20
- concatenation, 143
- conditional expression, 207
- constraints, 6, 96–97
 - adding and removing, 107
 - CHECK, 104–105, 157
 - column vs. table, 97
 - CONSTRAINT keyword, 76
 - foreign key, 102–103
 - NOT NULL, 106–107
 - PRIMARY KEY, 99
 - primary keys, 75, 97
 - UNIQUE, 76, 105–106
 - violations when altering table, 138
- constructor, 68
- Coordinated Universal Time (UTC), 33
- COPY statement
 - DELIMITER option, 43
 - description of, 39
 - exporting data, 25, 51–52
 - FORMAT option, 42
 - FROM keyword, 42
 - HEADER option, 43
 - importing data, 42–43
 - naming file paths, 25
 - QUOTE option, 43
 - specifying file formats, 42
 - TO, 51, 183
 - WITH keyword, 42
- correlated subquery, 192, 199
- corr() function, 157
 - correlation vs. causation, 163
- count() function, 117, 131, 196
 - distinct values, 118
 - on multiple columns, 123
 - values present in a column, 118
 - with GROUP BY, 122
- counting
 - distinct values, 118
 - missing values displayed, 133
 - rows, 117
 - using pgAdmin, 118
- CREATE DATABASE statement, 3
- createdb utility, 310
- CREATE EXTENSION statement, 203
- CREATE FUNCTION statement, 276
- CREATE INDEX statement, 108, 110
- CREATE TABLE statement, 6
 - backing up a table with, 139
 - declaring data types, 24
 - TEMPORARY TABLE, 50
- CREATE TRIGGER statement, 285
- CREATE VIEW statement, 269
- CROSS JOIN keywords, 82, 202
- crosstab() function, 203, 205, 207
 - with tablefunc module, 203
- cross tabulations, 203
- CSV (comma-separated values), 40
 - header row, 41
- CTE. *See* Common Table Expression (CTE)
- cube root operator (||/), 58
- curly brackets ({}), 215
 - denoting an array, 68
- current_date function, 175
- current_time function, 175
- current_timestamp function, 176
- cut points, 66

D

- data
 - identifying and telling stories
 - in, 325
 - spatial, 241
 - structured and unstructured, 211
- database
 - backup and restore, 321
 - connecting to, 4, 5
 - create from command line, 310
 - creation, 1, 3–5
 - importing data with COPY, 42–43
 - maintenance, 313
 - server, 3
 - using consistent names, 94
- database management system, 3
- data dictionary, 23
- data types, 5, 23
 - bigint, 27
 - bigserial, 6, 101
 - char, 24
 - character string types, 24–26
 - date, 5, 32, 172
 - date and time types, 32–34
 - decimal, 29
 - declaring with CREATE TABLE, 24
 - double precision, 29
 - full text search, 231
 - geography, 247
 - geometry, 247
 - importance of using appropriate type, 23, 46
 - integer, 27
 - interval, 32, 172
 - modifying with ALTER COLUMN, 137
 - number types, 26–31
 - numeric, 6, 28
 - real, 29
 - returned by math operations, 56
 - serial, 12, 101
 - smallint, 27
 - smallserial, 101
 - text, 25
 - time, 32, 172
 - timestamp, 32, 172
 - transforming values with CAST(), 35–36
 - tsquery, 232
 - tsvector, 231
 - varchar, 6, 24
- date data types
 - date, 5, 32, 172
 - interval, 32, 172
 - matching with regular expression, 217
- date_part() function, 173, 207
- dates
 - input format, 5, 8, 33, 173
 - setting default style, 320
- daylight saving time, 178
- deciles, 67
- decimal data types, 28
 - decimal, 29
 - double precision, 29
 - numeric, 28
 - real, 29
- decimal degrees, 46
- DELETE statement, 50
 - removing rows matching criteria, 147
 - with subquery, 194
- DELETE CASCADE statement
 - with foreign key constraint, 104
- delimited text files, 39, 40–41
- delimiter character, 40
- DELIMITER keyword
 - with COPY statement, 43
- dense_rank() function, 164
- derived table, 194
 - joining, 195–197
- DESC keyword, 15
- direct relationship in correlation, 158
- dirty data, 11, 129
 - cleaning, 129
 - foreign keys help to avoid, 103
 - when to discard, 137
- distance operator (<->), 232, 236
- DISTINCT keyword, 14, 118
- division, 57
 - finding the remainder, 58
 - integer vs. decimal, 57, 58
- documenting code, 23
- double at sign match operator (@@), 232
- double-colon CAST operator (::), 36
- double-dollar quoting (\$\$), 280
- double-pipe concatenation operator (||), 143, 225
- double quote ("), 41, 94
- DROP statement
 - COLUMN, 148
 - INDEX, 111
 - TABLE, 148

- duplicate data
 - created by spelling variations, 132
 - guarding against with constraints, 76

E

- Eastern Standard Time (EST), 33
- ELSE clause, 208, 227
- entity, 2
- environment variable, 292
- epoch, 174, 189
- equals comparison operator (=), 18
- error messages, 9
 - CSV import failure, 47, 49
 - foreign key violation, 103
 - out of range value, 27
 - primary key violation, 99, 101
 - relation already exists, 95
 - UNIQUE constraint violation, 106
 - when using CAST(), 36
- escaping characters, 219
- EST (Eastern Standard Time), 33
- exclamation point (!)
 - as factorial operator, 56, 59
 - as negation, 228, 232, 236
- EXISTS operator
 - in WHERE clause, 139
 - with subquery, 199
- EXPLAIN statement, 109
- exponentiation operator (^), 56, 58
- exporting data
 - all data in table, 51–52
 - from query results, 52
 - including header row, 43
 - limiting columns, 52
 - to BINARY file format, 42
 - to CSV file format, 42, 183–184
 - to TEXT file format, 42
 - using command line, 307
 - using COPY statement, 51–52
 - using pgAdmin wizard, 52–53
- expressions, 34, 192
 - conditional, 207
 - subquery, 198
- extract() function, 174

F

- factorials, 58
- false (Boolean value), 74
- Federal Information Processing Standards (FIPS), 259, 269

- field, 5
- file paths
 - import and export file locations, 42
 - naming conventions for operating systems, 25, 42
- filtering rows
 - HAVING clause, 127
 - WHERE clause, 17, 192
 - with subquery, 192
- findstr Windows command, 134
- FIPS (Federal Information Processing Standards), 259, 269
- fixed-point numbers, 28
- floating-point numbers, 29
 - inexact math calculations, 30
- foreign key
 - creating with REFERENCES keyword, 102
 - definition, 76, 102
- formatting SQL for readability, 10
- forward slash (/)
 - as division operator, 56, 57
 - in macOS file paths, 42
- FROM keyword, 12
 - with COPY, 42
- FULL OUTER JOIN keywords, 82
- full text search, 231
 - adjacent words, locating, 236–237
 - data types, 231–233
 - functions to rank results, 237–239
 - highlighting terms, 235
 - lexemes, 231–232
 - multiple terms in query, 236
 - querying, 234
 - setting default language, 320
 - table and column setup, 233
 - to_tsquery() function, 232
 - to_tsvector() function, 231
 - ts_headline() function, 235
 - ts_rank_cd() function, 237
 - ts_rank() function, 237
 - using GIN index, 234
- functions, 267
 - creating, 275, 276–277
 - full text search, 231
 - IMMUTABLE keyword, 277
 - RAISE NOTICE keywords, 280
 - RETURNS keyword, 277
 - specifying language, 276
 - string, 212
 - structure of, 276
 - updating data with, 278–280

G

- generate_series() function, 176, 207, 315
- geography data type, 247
- GeoJSON, 243
- geometry data type, 247
- GIN (Generalized Inverted Index), 108
 - with full text search, 234
- GIS (Geographic Information System), 241
 - decimal degrees, 46
- GiST (Generalized Search Tree) index, 108, 252
- greater than comparison operator (>), 18
- greater than or equals comparison operator (>=), 18
- grep Linux command, 134
- GROUP BY clause
 - eliminating duplicate values, 120
 - on multiple columns, 121
 - with aggregate functions, 120
- GUI (graphical user interface), 257, 291
 - list of tools, 333

H

- HAVING clause, 127
 - with aggregate functions, 127, 132
- HEADER keyword
 - with COPY statement, 43
- header row
 - found in CSV file, 41
 - ignoring during import, 41
- hyphen subtraction operator (-), 56, 57

I

- identifiers
 - avoiding reserved keywords, 95
 - enabling mixed case, 94–95
 - naming, 10, 94, 96
 - quoting, 95
- identifying and telling stories in data, 325
 - asking why, 331
 - assessing the data's origins, 328
 - building your own database, 327
 - communicating your findings, 331
 - consulting the data's owner, 328

- documenting your process, 326
- gathering your data, 326
- identifying trends over time, 329
- interviewing the data with queries, 328
 - starting with a question, 326
- ILIKE comparison operator, 18, 19–20
- importing data, 39, 42–43
 - adding default column value, 50
 - choosing a subset of columns, 49
 - from non-text sources, 40
 - from TEXT file format, 42
 - from CSV file format, 42
 - ignoring header row in text files, 41, 43
 - using command line, 307
 - using COPY statement, 39
 - using pgAdmin import wizard, 52–53
- IN comparison operator, 18, 144, 198
 - with subquery, 198
- indexes, 108
 - B-Tree, 108
 - considerations before adding, 111
 - creating on columns, 110
 - dropping, 111
 - GIN, 108
 - GiST, 108, 252
 - measuring effect on performance, 109
 - not included with table backups, 140
 - syntax for creating, 108
- initcap() function, 212
- INSERT statement, 8–9
- inserting rows into a table, 9–10
- Institute of Museum and Library Services (IMLS), 114
- integer data types, 27
 - auto-incrementing, 27
 - basic math operations, 57
 - bigint, 27
 - bigserial, 27
 - difference in integer type capacities, 27
 - integer, 27
 - serial, 27
 - smallint, 27
 - smallserial, 27
- International Date Line, 46

- International Organization for Standardization (ISO), xxiv, 33, 243
- interval data type, 32, 172
 - calculations with, 34, 187
 - cumulative, 188
 - value options, 34
- interviewing data, 11, 131–132
 - across joined tables, 124
 - artificial values as indicators, 120, 124
 - checking for missing values, 13, 132–134
 - correlations, 157–159
 - counting rows and values, 117–119
 - determining correct format, 13
 - finding inconsistent values, 134
 - malformed values, 135–136
 - maximum and minimum values, 119–120
 - rankings, 164–167
 - rates calculations, 167–169
 - statistics, 155
 - summing grouped values, 124
 - unique combinations of values, 15
- inverse relationship, 158
- ISO (International Organization for Standardization), xxiv, 33, 243
 - time format, 172

J

- JOIN keyword, 74
 - example of using, 80
 - in FROM clause, 74
- joining tables, 73
 - derived tables, 195–197
 - inequality condition, 90
 - multiple-table joins, 87
 - naming tables in column list, 85, 125
 - performing calculations across tables, 88
 - spatial joins, 262, 263
 - specifying columns to link tables, 77
 - specifying columns to query, 85
 - using JOIN keyword, 74, 77

- join types
 - CROSS JOIN, 82–83
 - FULL OUTER JOIN, 82
 - JOIN (INNER JOIN), 80, 125
 - LEFT JOIN, 80–81
 - list of, 78
 - RIGHT JOIN, 80–81
- JSON, 35

K

- key columns
 - foreign key, 76
 - primary key, 75
 - relating tables with, 74

L

- latitude
 - in U.S. Census data, 46
 - in well-known text, 245
- least squares regression line, 161
- LEFT JOIN keyword, 80–81
- left() string function, 213
- length() string function, 135, 213
- less than comparison operator (<), 18
- less than or equals comparison operator (<=), 18
- lexemes, 231
- LIKE comparison operator, 18
 - case-sensitive search, 19
 - in UPDATE statement, 143
- LIMIT clause, 48
- limiting number of rows query
 - returns, 48
- linear regression, 161
- linear relationship, 158
- Linux
 - file path declaration, 26, 42
 - Terminal setup, 299
- literals, 8
- locale setting, 16
- localhost, xxxii, 4
- localtime function, 176
- localtimestamp function, 176
- longitude
 - in U.S. Census data, 46
 - in well-known text, 245
 - positive and negative values, 49
- lower() function, 212

M

macOS

- file path declaration, 25, 42
- Terminal, 296
 - .bash_profile*, 296
 - bash shell, 296
 - entering instructions, 297
 - setup, 296, 297
 - useful commands, 298

make_date() function, 175

make_time() function, 175

make_timestamptz() function, 175

many-to-many table relationship, 85

map

- projected coordinate system, 245
- projection, 245

math

- across joined table columns, 88
- across table columns, 60–64
- median, 65–70
- mode, 70
- order of operations, 59
- with aggregate functions, 64–65

math operators, 56–59

- addition (+), 57
- cube root ($\sqrt[3]{}$), 58
- division (/), 57
- exponentiation (^), 58
- factorial (!), 58
- modulo (%), 57
- multiplication (*), 57
- square root ($\sqrt{\quad}$), 58
- subtraction (-), 57

max() function, 119

median, 65

- definition, 65
- vs. average, 65, 194
- with percentile_cont() function, 66

median() user function

- creation, 69
- performance concerns, 70
- vs. percentile_cont(), 70

Microsoft Access, xxiv

Microsoft Excel, xxiv

Microsoft SQL Server, xxviii, 94, 203

Microsoft Windows

- Command Prompt
 - entering instructions, 295
 - setup, 292, 294
 - useful commands, 295
- file path declaration, 25, 42
- folder permissions, xxvii

min() function, 119

mode, 70

mode() function, 70

modifying data, 136–137

- for consistency, 142
- updating column values, 141

modulo operator (%), 56, 57–58

multiplying numbers, 57

MySQL, xxviii

N

naming conventions

- camel case, 94
- Pascal case, 94
- snake case, 94, 96

National Center for Education Statistics, 327

National Center for Health Statistics, 330

natural primary key, 97, 131

New York City taxi data, 180

- calculating busiest hour of day, 182
- creating table for, 180
- exporting results, 183–184
- importing, 181
- longest trips, 184–185

normal distribution of data, 194

NOT comparison operator, 18

- with EXISTS, 200

not equal comparison operator

- != syntax, 18
- <> syntax, 18

NOT NULL keywords

- adding to column, 137
- definition, 106
- removing from column, 107, 138

now() function, 33, 176

NULL keyword

- definition, 83
- ordering with FIRST and LAST, 133
- using in table joins, 83

number data types, 26

- decimal types, 28
 - double precision, 29
 - fixed-point type, 28
 - floating-point types, 29
 - numeric data type, 6, 28
 - real, 29

- integer types, 27
 - bigint, 27
 - integer, 27
 - serial types, 27
 - smallint, 27
- usage considerations, 31

O

- OGC (Open Geospatial Consortium), 243
- ON keyword
 - used with DELETE CASCADE, 104
 - used with JOIN, 74
- one-to-many table relationship, 84
- one-to-one table relationship, 84
- operators
 - addition (+), 56, 57
 - comparisons with, 17
 - cube root ($\sqrt[3]{}$), 56, 58
 - division (/), 56, 57
 - exponentiation (^), 56, 58
 - factorial (!), 56, 58
 - modulo (%), 56, 57
 - multiplication (*), 56, 57
 - precedence, 59
 - prefix, 58
 - square root ($\sqrt{\quad}$), 56, 58
 - subtraction (-), 56, 57
 - suffix, 59
- OR operator, 20
- Oracle, xxiv
- ORDER BY clause, 15
 - ASC, DESC options, 15
 - on multiple columns, 16
 - specifying columns to sort, 15
 - specifying NULLS FIRST or LAST, 133
- OVER clause, 164

P

- Pacific time zone, 33
- padding character columns with spaces, 24, 26
- parentheses (), 6, 8
 - to designate order of operations, 20
 - to specify columns for importing, 50
- Pascal case, 94
- pattern matching
 - using LIKE and ILIKE, 19
 - with regular expressions, 214
 - with wildcards, 19

- Pearson correlation coefficient (r), 157
- percent sign (%)
 - as modulo operator, 56, 57
 - wildcard for pattern matching, 19
- percentage
 - of the whole, 62
 - percent change, 63
 - formula, 63, 89, 276
 - function, 276
- percent_change() user function, 276
 - using with Census data, 277
- percentile, 66, 192
 - continuous vs. discrete values, 66
- percentile_cont() function, 66
 - finding median with, 185
 - in subquery, 193
 - using array to enter multiple values, 68
- percentile_disc() function, 66
- pgAdmin, xxxi
 - connecting to database, 4, 5, 242
 - connecting to server, xxxii, 4
 - executing SQL, 3
 - importing and exporting data, 52–53
 - installation
 - Linux, xxxi
 - macOS, xxxi, xxxii
 - Windows, xxix, xxxi
 - keyword highlighting, 95
 - localhost, xxxii, 4
 - object browser, xxxii, 5, 7
 - Query Tool, xxxiii, 4, 243
 - text display in results grid, 218
 - viewing data, 9, 75, 118
 - viewing tables, 45
 - views, 269
- pg_ctl utility, 321
- pg_dump utility, 321
- pg_restore utility, 322
- pg_size_pretty() function, 315
- pg_total_relation_size() function, 315
- pipe character ($|$)
 - as delimiter, 26, 43
 - to redirect output, 311
- pivot table. *See* cross tabulations
- PL/pgSQL, 276, 279
 - BEGIN ... END block, 280, 284
 - IF ... THEN statement, 284
- PL/Python, 281
- point, 46
- position() string function, 213

- PostGIS, xxviii, 242
 - creating spatial database, 242–243
 - creating spatial objects, 247
 - data types, 247
 - geography, 247
 - geometry, 247
 - displaying version, 243
 - functions
 - ST_AsText(), 260
 - ST_DFullyWithin(), 254
 - ST_Distance(), 254
 - ST_DWithin(), 253
 - ST_GeogFromText(), 248, 254
 - ST_GeometryType(), 262
 - ST_GeomFromText(), 247
 - ST_Intersection(), 264
 - ST_Intersects(), 263
 - ST_LineFromText(), 250
 - ST_MakeLine(), 250
 - ST_MakePoint(), 249
 - ST_MakePolygon(), 250
 - ST_MPolyFromText(), 250
 - ST_PointFromText(), 249
 - ST_PolygonFromText(), 250
 - installation, 242–243
 - Linux, xxxi
 - macOS, xxx
 - troubleshooting, xxx
 - Windows, xxix–xxx
 - loading extension, 243
 - shapefile
 - loading, 257, 258, 311
 - querying, 259
 - spatial joins, 262, 263
- Postgres.app, xxx–xxxi, 4
- PostgreSQL
 - advantages of using, xxviii
 - backup and restore, 321
 - pg_dump, 321
 - pg_restore, 322
 - collation setting, 16
 - command line usage, 291
 - comparison operators, 18
 - configuration, 313
 - creating functions, 275
 - default postgres database, 3
 - description of, 3
 - documentation, 335
 - functions, 267
 - GUI tools, 333
 - importing from other database managers, 40
 - installation, xxviii
 - Linux, xxxi
 - macOS, xxx–xxxi
 - troubleshooting, xxx
 - Windows, xxix–xxx
 - locale setting, xxix, 16
 - maintenance, 313
 - news websites, 335
 - postgresql.conf* settings file, 319
 - recovering unused space, 314
 - settings, 318
 - spatial data analysis, 241, 253, 254
 - starting and stopping, 321
 - statistics collector, 317
 - table size, 314
 - triggers, 267, 282
 - utilities, tools, and extensions, 334
 - views, 267
- postgresql.conf* settings file, 178, 319
 - editing, 319
 - reloading settings, 321
- precision argument
 - with numeric and decimal types, 28
- primary key, 2, 12
 - composite, 100–101
 - definition of, 75, 97
 - natural, 97, 131
 - surrogate, 97, 98
 - auto-incrementing, 101–102
 - creating, 102
 - data types for, 101
 - syntax, 98–100
 - uniqueness, 76
 - using auto-incrementing serial type, 28
 - using Universally Unique Identifier, 98
 - violation, 99, 101
- Prime Meridian, 46, 246
- procedural language, 276
- projection (map), 245
 - Albers, 246
 - Mercator, 245
- psql command line application, 3, 292
 - connecting to database, 299, 300
 - displaying table info, 306
 - editing queries, 303
 - executing queries from a file, 309
 - formatting results, 303, 304
 - help commands, 300
 - importing and exporting files, 307

- meta-commands, 306
- multiline queries, 302
- paging results, 303
- parentheses in queries, 302
- running queries, 301
- saving query output, 308
- setup
 - Linux, 299
 - macOS, 296–298
 - Microsoft Windows, 293–295
 - superuser prompt, 300
- Public Libraries Survey, 114
- Python programming language,
 - xxv, 335
 - creating PL/Python extension, 281
 - in PostgreSQL function, 277, 281

Q

- quantiles, 66
- quartiles, 67
- query
 - choosing order of columns, 13
 - definition, 1
 - eliminating duplicate values, 14
 - execution time, 109–110
 - exporting results of, 52
 - limiting number of rows
 - returned, 48
 - measuring performance with
 - EXPLAIN, 109
 - order of clauses, 21
 - retrieving a subset of columns, 13
 - selecting all rows and columns, 12
- quintiles, 67
- quotes, single vs. double, 8

R

- rank() function, 164
- ranking data, 164
 - by subgroup, 165–167
 - rank() and dense_rank() functions, 164–165
- rates calculations, 167, 196
- record_if_grade_changed() user
 - function, 284
- REFERENCES keyword, 103
- referential integrity, 97
 - cascading deletes, 104
 - foreign keys, 102
 - primary key, 99

- regexp_match() function, 219
 - extracting text from result, 224
- regexp_matches() function, 220
- regexp_replace() function, 230
- regexp_split_to_array() function, 230
- regexp_split_to_table() function, 230
- regr_intercept() function, 162
- regr_r2() function, 163
- regr_slope() function, 162
- regular expressions, 214
 - capture group, 215, 221
 - escaping characters, 219
 - examples, 216
 - in WHERE clause, 228–229
 - notation, 214–216
 - parsing unstructured data, 216, 222
 - regexp_match() function, 219
 - regexp_matches() function, 220
 - regexp_replace() function, 230
 - regexp_split_to_array()
 - function, 230
 - regexp_split_to_table()
 - function, 230
 - with substring() function, 216
- relational databases, 2, 73
 - join types
 - CROSS JOIN, 82–83
 - FULL OUTER JOIN, 82
 - JOIN (INNER JOIN), 80, 125
 - LEFT JOIN, 80–81
 - list of, 78
 - RIGHT JOIN, 80–81
 - querying, 77
- relating tables, 74–77
- relational model, 73, 84
 - reducing redundant data, 77
 - table relationships
 - many-to-many, 85
 - one-to-many, 84
 - one-to-one, 84
- replace() string function, 214
- reserved keywords, 95
- RIGHT JOIN keywords, 80–81
- right() string function, 213
- ROLLBACK statement, 149
- roots, square and cube, 58
- round() function, 64, 160
- row
 - counting, 117
 - definition, 73
 - deleting, 147–148

- row (*continued*)
 - in a CSV file, 40
 - inserting, 8
 - recovering unused, 314
 - updating specific, 141
- r* (Pearson correlation coefficient), 157
- r*-squared, 163
- R programming language, xxv

S

- scalar subquery, 192
- scale argument
 - with numeric and decimal types, 29
- scatterplot, 158, 159
- search. *See* full text search
- SELECT statement
 - definition, 11
 - order of clauses, 21
 - syntax, 12
 - with DISTINCT keyword, 14–15
 - with GROUP BY clause, 120
 - with ORDER BY clause, 15–17
 - with WHERE clause, 17–20
- selecting all rows and columns, 12
- semicolon (;), 3
- serial, 27, 101
- server
 - connecting, 4
 - localhost, 4
 - postgresql.conf* file, 178
 - setting time zone, 178
- SET keyword
 - clause in UPDATE, 138, 192
 - timezone, 178
- shapefile, 256
 - contents of, 256–257
 - loading into database, 257
 - shp2pgsql command line utility, 311
 - U.S. Census TIGER/Line, 258, 262
- SHOW command
 - config_file*, 319
 - data_directory*, 321
 - timezone, 177
- shp2pgsql command line utility, 311
- significance testing, 163
- simple feature standard, 243
- single quote ('), 8, 42
- slope-intercept formula, 161
- smallint data type, 27
- smallserial data type, 27, 101

- snake case, 10, 94, 96
- sorting data, 15
 - by multiple columns, 16
 - dependent on locale setting, 16
 - on aggregate results, 123
- spatial data, 241
 - area analysis, 260
 - building blocks, 243
 - distance analysis, 253, 254
 - finding location, 261
 - geographic coordinate system, 243, 245, 246
 - geometries, 243
 - constructing, 245, 247
 - LineString, 243, 249–250
 - MultiLineString, 244
 - MultiPoint, 244
 - MultiPolygon, 244
 - Point, 243, 249
 - Polygon, 243, 250
 - intersection analysis, 264
 - joins, 262, 263
 - projected coordinate system, 245
 - projection, 245
 - shapefile, 256
 - simple feature standard, 243
 - Spatial Reference System Identifier (SRID), 244, 246
 - well-known text (WKT), 244
 - WGS 84 coordinate system, 246
- Spatial Reference System Identifier (SRID), 244, 246
 - setting with ST_SetSRID(), 252
- SQL
 - comments in code, xxvii
 - history of, xxiv
 - indenting code, 10
 - math operators, 56
 - relational model, 73
 - reserved keywords, 95
 - standards, xxiv
 - statistical functions, 155
 - style conventions, 6, 10, 36, 94
 - using with external programming languages, xxv
 - value of using, xxiv
- square root operator (/), 56, 58
- SRID (Spatial Reference System Identifier), 244, 246
 - setting with ST_SetSRID(), 252

- statistical functions, 155
 - correlation with `corr()`, 157–159
 - dependent and independent variables, 158
 - linear regression, 160
 - `regr_intercept()` function, 162
 - `regr_r2()` function, 163
 - `regr_slope()` function, 162
 - rates calculations, 167
- string functions, 135, 212
 - case formatting, 212
 - character information, 212
 - `char_length()`, 212
 - extracting and replacing characters, 213
 - `initcap()`, 212
 - `left()`, 213
 - `length()`, 135, 213
 - `lower()`, 212
 - `position()`, 213
 - removing characters, 213
 - `replace()`, 214
 - `right()`, 213
 - `to_char()`, 187
 - `trim()`, 213
 - `upper()`, 212
- subquery
 - correlated, 192, 199
 - definition, 192
 - expressions, 198
 - generating column with, 197–198
 - in `DELETE` statement, 194
 - in `FROM` clause, 194
 - in `IN` operator expression, 198–199
 - in `UPDATE` statement, 139, 192
 - in `WHERE` clause, 192–194
 - scalar, 192
 - uncorrelated, 192
 - with `crosstab()` function, 205
- `substring()` function, 216
- subtracting numbers, 57
 - across columns, 60
- `sum()` function, 64
 - example on joined tables, 124
 - grouping by column value, 125
- summarizing data, 113
- surrogate primary key, 98
 - creating, 102

T

- tab character
 - as delimiter, 42–43
 - as regular expression, 215
- table
 - add column, 137, 140
 - aliases, 86, 195
 - alter column, 137
 - autovacuum, 316
 - backup, 94
 - constraints, 6
 - creation, 5–7
 - definition of, 1
 - deleting columns, 137, 148
 - deleting data, 147–149
 - deleting from database, 148–149
 - derived table, 194
 - design best practices, 93
 - dropping, 148
 - holds data on one entity, 73
 - indexes, 108
 - inserting rows, 8–9
 - key columns, 74
 - modifying with `ALTER` statement, 137–138
 - naming, 94, 96
 - querying multiple tables using joins, 77
 - relationships, 1
 - size, 314
 - temporary tables, 50
 - viewing data, 9
- tablefunc module, 203
- table relationships
 - many-to-many, 85
 - one-to-many, 84
 - one-to-one, 84
- temporary table
 - declaring, 50
 - removing with `DROP TABLE`, 51
- text data types, 24–26
 - `char`, 24
 - `text`, 25
 - `varchar`, 6, 24
- text operations
 - case formatting, 212
 - concatenation, 143
 - escaping characters, 219

- text operations (*continued*)
 - extracting and replacing
 - characters, 213–214
 - formatting as timestamp, 173
 - formatting with functions, 212–214
 - matching patterns with regular expressions, 214
 - removing characters, 213
 - sorting, 16
- text files, delimited. *See* delimited text files
- text qualifier
 - ignoring delimiters with, 41
 - specifying with QUOTE option in COPY, 43
- tilde-asterisk case-insensitive matching operator (~*), 228
- tilde case-sensitive matching operator (~), 228
- time data types
 - interval, 32, 172
 - matching with regular expression, 215
 - time, 32, 172
 - timestamp, 32, 172
- timestamp, 32, 172
 - calculations with, 180
 - creating from components, 174–175, 225
 - extracting components from, 173–174
 - finding current date and time, 175–176
 - formatting display, 187
 - subtracting to find interval, 187
 - timestampz shorthand, 172
 - with time zone, 32, 172
 - within transactions, 176
- time zones
 - AT TIME ZONE keywords, 179
 - automatic conversion of, 173, 175
 - finding server setting, 177–178
 - including in timestamp, 32, 173, 226
 - setting, 178–180
 - setting server default, 320
 - standard name database, 33
 - viewing names of, 177
 - working with, 177
- to_char() function, 187
- to_tsquery() function, 232
- to_tsvector() function, 231
- transaction blocks, 149–151
 - COMMIT, 149
 - definition, 149
 - ROLLBACK, 149
 - START TRANSACTION, 149
 - visibility to other users, 151
- transactions, 149
 - with time functions, 176
- triggers, 267, 282
 - BEFORE INSERT statement, 288
 - CREATE TRIGGER statement, 285
 - FOR EACH ROW statement, 285
 - FOR EACH STATEMENT statement, 285
 - NEW and OLD variables, 284
 - RETURN statement, 285
 - testing, 285, 288
- trim_county() user function, 281
- trim() function, 213
- true (Boolean value), 74
- ts_headline() function, 235
- tsquery data type, 232
- ts_rank_cd() function, 237
- ts_rank() function, 237
- tsvector data type, 231

U

- uncorrelated subquery, 192
- underscore wildcard for pattern matching (_,) 19
- UNIQUE constraint, 76, 105–106
- Universally Unique Identifier (UUID), 35, 98
- unnest() function, 68
- unstructured data, 211
 - parsing with regular expressions, 216, 222
- UPDATE statement
 - definition, 138
 - PostgreSQL syntax, 139
 - SET clause, 138
 - using across tables, 138, 145, 192
 - with CASE statement, 226
- update_personal_days() user function, 279
- upper() function, 212
- USA TODAY, xxiii
- U.S. Census
 - 2010 Decennial Census data, 43
 - calculating population change, 89
 - county shapefile analysis, 259

- description of columns, 45–47
- finding total population, 64
- importing data, 43–44
- racial categories, 60
- short form, 60
- 2011–2015 American Community Survey
 - description of columns, 156
 - estimates and margin of error, 157
 - importing data, 156
- apportionment of U.S. House of Representatives, 44
- methodologies compared, 157, 328
- U.S. Department of Agriculture, 130
 - farmers' market data, 250
- U.S. Federal Bureau of Investigation (FBI) crime report data, 167
- UTC (Coordinated Universal Time), 33, 174
 - UTC offset, 33, 179, 187
- UTF-8, 16
- UUID (Universally Unique Identifier), 35, 98

V

- VACUUM command, 314
 - ANALYZE option, 317
 - autovacuum process, 316
 - editing server setting, 319
 - FULL option, 318
 - monitoring table size, 314
 - pg_stat_all_tables view, 317
 - running manually, 318
 - time of last vacuum, 317
 - VERBOSE option, 318
- VALUES clause with INSERT, 8
- varchar data type, 6, 24
- views, 267
 - advantage of using, 268
 - creating, 269–271
 - deleting data with, 275
 - dropping, 269
 - inserting data with, 273–274
 - inserting, updating, deleting data, 271
 - LOCAL CHECK OPTION, 272, 273

- materialized, 268
- pg_stat_all_tables, 317
- queries in, 269
- retrieving specific columns, 271
- updating data with, 274

W

- well-known text (WKT), 244
 - extended, 248
 - order of coordinates, 245
- WHEN clause, 208
 - in CASE statement, 227
- WHERE clause, 17
 - in UPDATE statement, 138
 - filtering rows with, 17–19
 - with DELETE FROM statement, 147
 - with EXISTS clause, 139, 192
 - with ILIKE operator, 19–20
 - with IS NULL keywords, 133
 - with LIKE operator, 19–20, 143
 - with regular expressions, 228
- whole numbers, 27
- wildcard
 - asterisk (*) in SELECT statement, 12
 - percent sign (%), 19
 - underscore (_), 19
- window functions
 - definition of, 164
 - OVER clause, 164
 - PARTITION BY clause, 165
- WITH
 - as Common Table Expression, 200
 - options with COPY, 42
- WKT (well-known text), 244
 - extended, 248
 - order of coordinates, 245
- working tables, 148

X

- XML, 35

Z

- ZIP Codes, 135
 - loss of leading zeros, 135
 - repairing botched, 143