



# PostgreSQL

NICAR 2019 NEWPORT BEACH  
ANTHONY DEBARROS • @ANTHONYDB

# PostgreSQL: Intro

- Open-source Database Management System
- Closely follows ANSI-SQL standards
- Active for 20+ years; stable
- Strong software ecosystem, support
- Big data variants (Amazon Redshift, Greenplum)
- Data types for GIS (via PostGIS)
- Frequent choice for web app backends

# PostgreSQL: Installing

- Windows Installers
  - PostgreSQL.org: [postgresql.org/download/windows/](https://www.postgresql.org/download/windows/)
    - EnterpriseDB [enterprisedb.com/downloads/postgres-postgresql-downloads](https://enterprisedb.com/downloads/postgres-postgresql-downloads)
    - BigSQL: [openscg.com/bigsql/](https://openscg.com/bigsql/)
- macOS:
  - Postgres.app: [postgresapp.com](https://postgresapp.com)
- Linux
  - Distribution-specific
    - e.g., Ubuntu: [postgresql.org/download/linux/ubuntu/](https://www.postgresql.org/download/linux/ubuntu/)

# PostgreSQL: Managing

- pgAdmin: graphical UI
  - [pgadmin.org](https://pgadmin.org)
- psql: command-line tool
  - Included with PostgreSQL
- Numerous additional tools and add-ons
  - [github.com/dhamaniasad/awesome-postgres/](https://github.com/dhamaniasad/awesome-postgres/)

# PostgreSQL: Hands-On

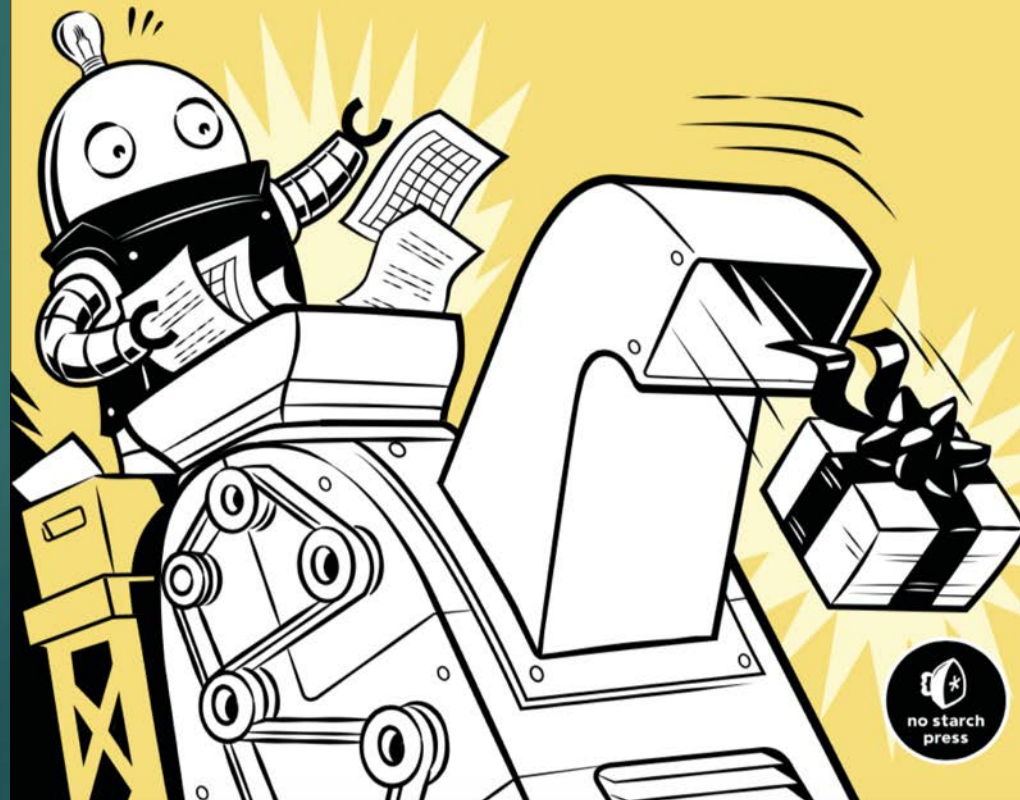
Today, we'll cover:

- pgAdmin, psql and some basic queries
- Creating a function
- Spatial queries with PostGIS
- Full-text search
- Statistics functions

# PRACTICAL SQL

A BEGINNER'S GUIDE TO  
STORYTELLING WITH DATA

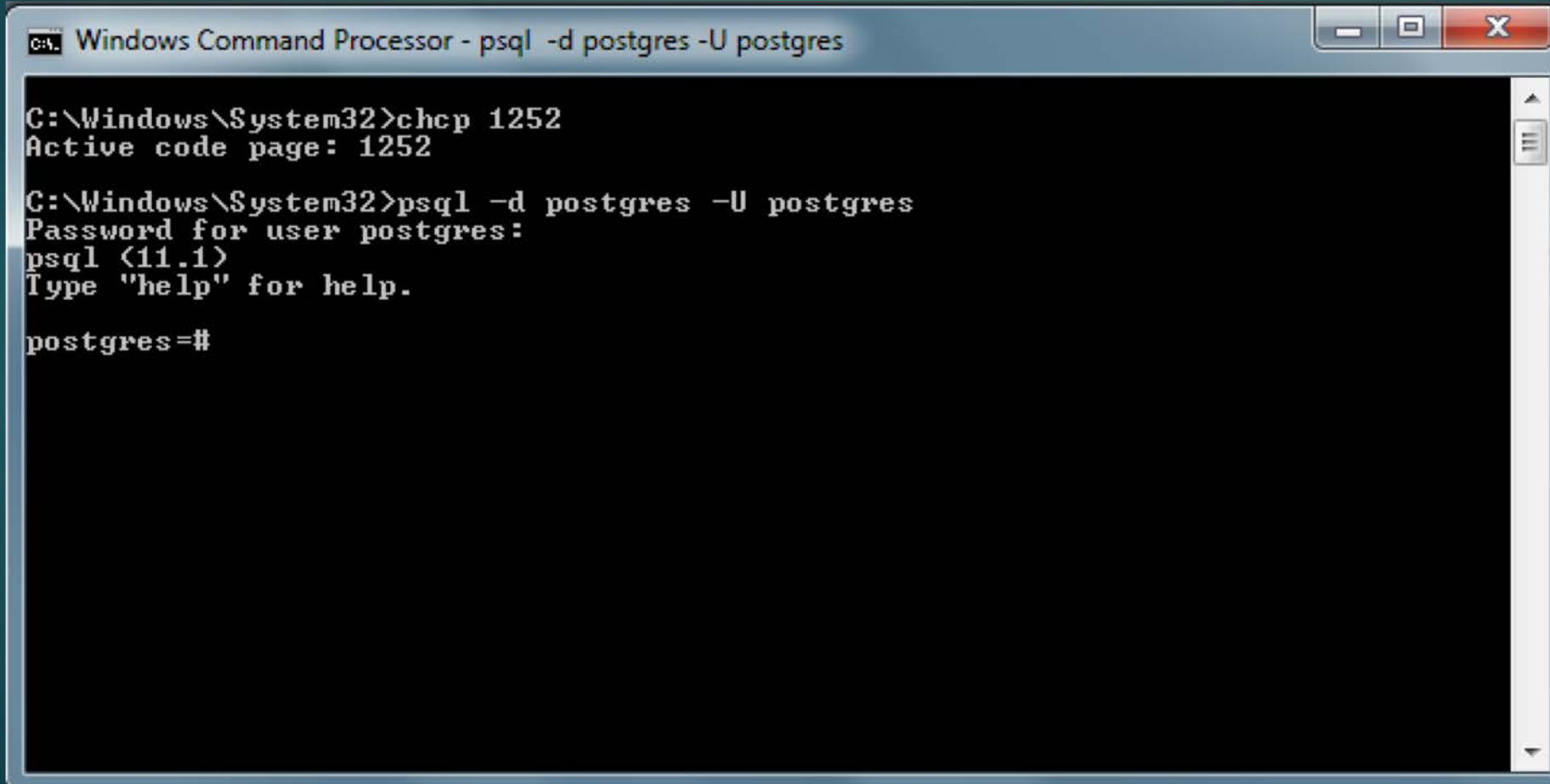
ANTHONY DEBARROS





# Exploring psql

- `C:\ chcp 1252`
- `C:\ psql -d postgres -U postgres`

A screenshot of a Windows Command Processor window. The title bar reads "Windows Command Processor - psql -d postgres -U postgres". The command prompt shows the following sequence of commands and output:  
C:\Windows\System32>chcp 1252  
Active code page: 1252  
  
C:\Windows\System32>psql -d postgres -U postgres  
Password for user postgres:  
psql (11.1)  
Type "help" for help.  
  
postgres=#  
The window has a standard Windows interface with minimize, maximize, and close buttons in the title bar, and a scrollbar on the right side of the command area.

# Exploring pgAdmin

The screenshot displays the pgAdmin 4 web interface. The left sidebar shows a tree view of the database structure under the 'nicar\_2019' database, with 'Tables (7)' expanded. The main panel shows the 'Query Editor' with a SQL query: `SELECT * FROM us_counties_2000;`. Below the editor, the 'Data Output' tab is active, showing a table of results. The table has 7 columns: `geo_name`, `state_us_abbreviation`, `state_fips`, `county_fips`, `p0010001`, and `p0010002`. The results show the first 7 rows of data for counties in Alabama.

	geo_name character varying (90)	state_us_abbreviation character varying (2)	state_fips character varying (2)	county_fips character varying (3)	p0010001 integer	p0010002 integer
1	Autauga County, Alabama	AL	01	001	43671	4326
2	Baldwin County, Alabama	AL	01	003	140415	13894
3	Barbour County, Alabama	AL	01	005	29038	2882
4	Bibb County, Alabama	AL	01	007	20826	2077
5	Blount County, Alabama	AL	01	009	51024	5050
6	Bullock County, Alabama	AL	01	011	11714	1163
7	Butler County, Alabama	AL	01	013	21399	2137



# Exploring PostgreSQL: Basic Queries

- `SELECT * FROM table_name;`
- JOINS
- Calculations

# Exploring PostgreSQL: Features!

- Creating functions
- Spatial queries with PostGIS
- Full-text search
- Stats functions

# PostgreSQL: PostGIS

Spatial data types and analysis

- <http://postgis.net/>
- Data types: geometry and geography
- <http://postgis.net/docs/manual-2.5/>

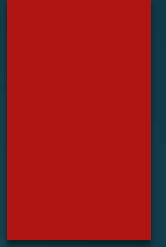
# PostGIS Data Types

Installing PostGIS adds five data types to your database. The two data types we'll use in the exercises are geography and geometry. Both types can store spatial data, such as the points, lines, polygons, SRIDs, and so on you just learned about, but they have important distinctions:

**geography** A data type based on a sphere, using the round-earth coordinate system (longitude and latitude). All calculations occur on the globe, taking its curvature into account. That makes the math complicated and limits the number of functions available to work with the geography type. But because the Earth's curvature is factored in, calculations for distance are more precise; you should use the geography data type when handling data that spans large areas. Also, the results from calculations on the geography type will be expressed in meters.

**geometry** A data type based on a plane, using the Euclidean coordinate system. Calculations occur on straight lines as opposed to along the curvature of a sphere, making calculations for geographical distance less precise than with the geography data type; the results of calculations are expressed in units of whichever coordinate system you've designated.

# PostgreSQL: Full-Text Search



# Stats: Correlation Coefficient

## U.S. Counties: Education vs. Income

