PostgreSQL

NICAR CHICAGO 2018
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, Citus)
- Data types for GIS (via PostGIS)
- Frequent choice for web app backends

PostgreSQL: Installing

- Windows Installers
  - EnterpriseDB.com
  - BigSQL: www.openscg.com/bigsql/
- macOS:
 - Postgres.app: postgresapp.com
- Linux
  - Distribution-specific

PostgreSQL: Managing

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

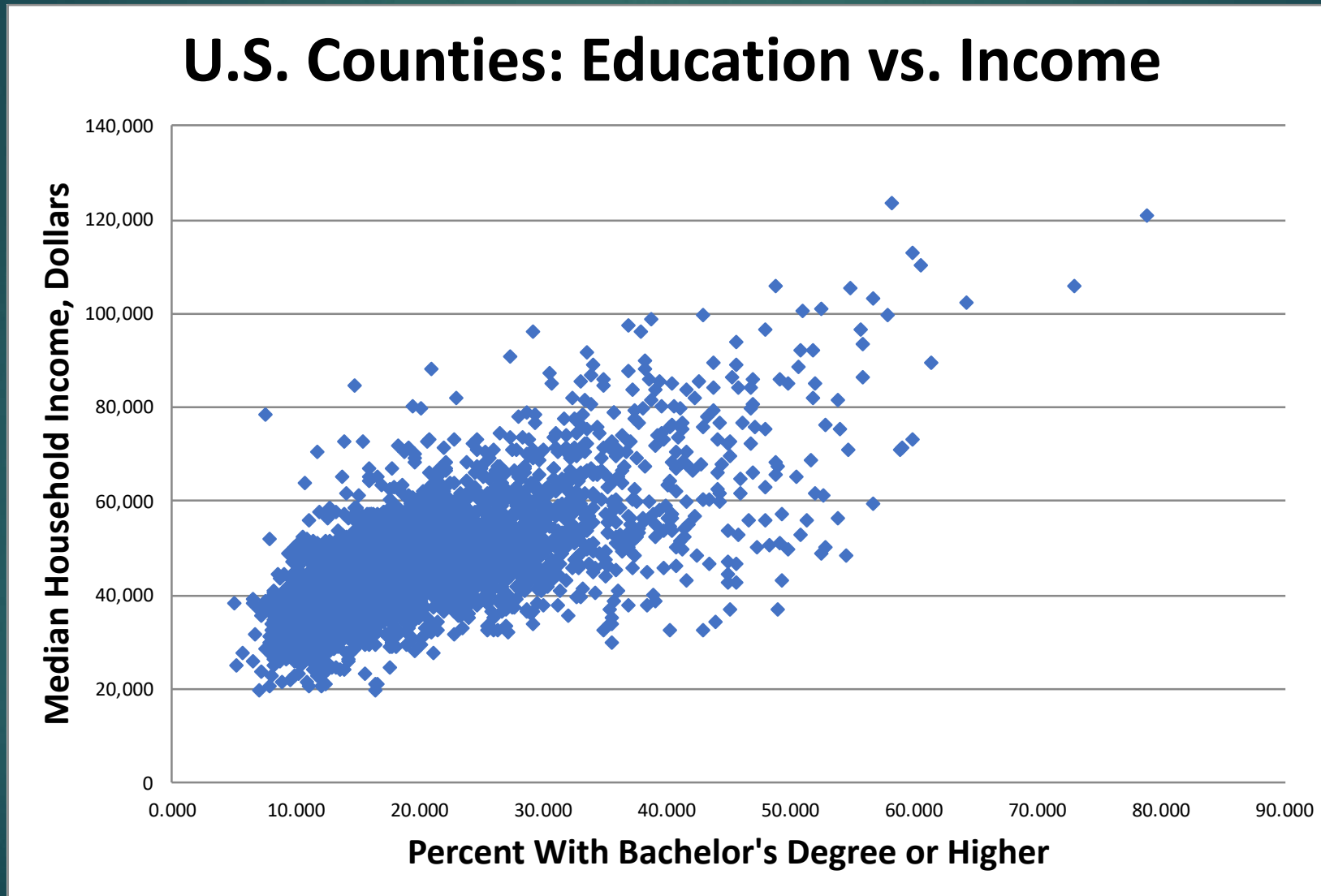# PostgreSQL: Hands-On

Today, we'll cover:

- pgAdmin and some basic queries
- Stats functions
- Spatial queries with PostGIS
- Full-text search
- Creating functions

Shameless Plug ...
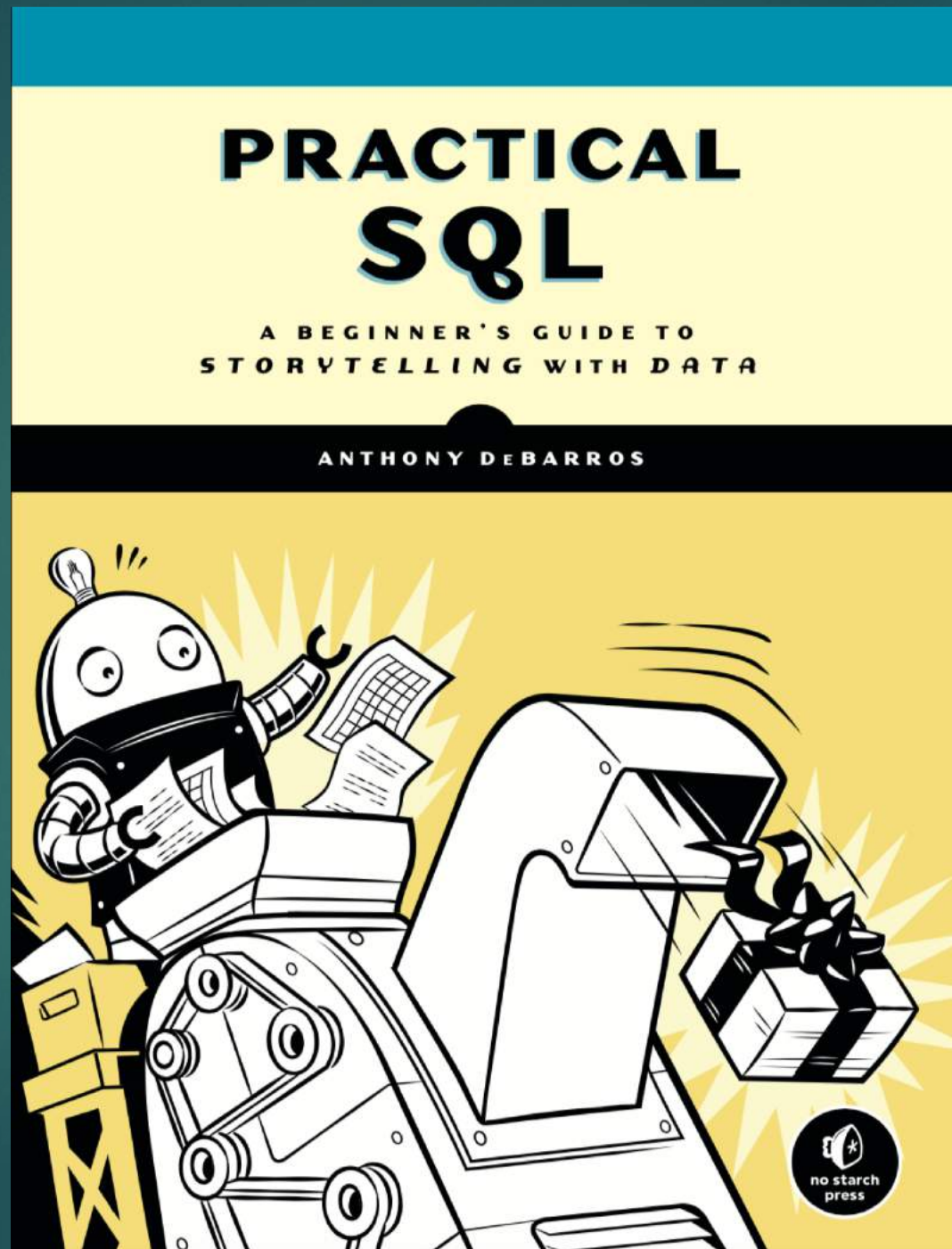
PostgreSQL: Correlation Coefficient

U.S. Counties: Education vs. Income

PostgreSQL: PostGIS

Spatial data types and analysis

- http://postgis.net/
- Data types: geometry and geography
- http://postgis.net/docs/manual-2.4/reference.html

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.