The Giant Database of Dogs

User's Guide, version 1.0
Accompanies dataset release 1.0

The Giant Database of Dogs is a crowdsourced initiative organized by Prof. Leanne Powner. The dataset is explicitly designed for classroom use.

Data Collection Procedures

Respondents were solicited by posts on Powner's professional Twitter profile[1] as well as her personal Facebook profile and in two Facebook groups for political science data and faculty. Multiple posts occurred in both locations; Facebook included extensive review of Friends to tag as many known current or former dog owners as possible. Some respondents on both platforms voluntarily reposted the solicitation on their own feeds and/or in relevant groups to which they had access.

Respondents were directed to a Google form featuring a photo of Powner's current pair of beagles (see Figure 1). Instructions informed respondents that the data was for classroom and educational purposes only, and that no identifying information was being collected. As this research did not involve human subjects or animal testing, and additionally would be exempt as it was for classroom purposes, no human subjects clearance was sought. The Google form was configured not to collect any personally identifying information beyond the minimum needed for successful sending and completion of the form (i.e., an IP address, which it does not save after the form is complete).

Respondents were informed that they could submit multiple dogs and that both current and former dogs were welcome. "Best guess" data was acceptable where precise information was not known; this most likely applies to weight given the clustering of data at 'round' numbers (ending in 5 or 0).

The data set will be updated each semester with new inputs. You are welcome to add your own dog(s) to the data set using the form located at https://forms.gle/aDpw8b8tnUoYGgdU9 .

Data Cleaning

Initial data collection was closed when the sample reached 454 dog entries. Powner then minimally cleaned the data by assigning a unique identifier value (*Obs*) to each case, converting all weights to pounds (multiplying kilogram responses by 2.2), removing all text in the *Weight* and *Age* columns to leave only digits, and removing additional text such as nicknames provided in the *Name* field.

Codebook

Obs – unique observation number for case identification
Name – dog's name as provided by the owner (excess text removed)
Gender – {male, female} closed choice
Fixed – "yes" if spayed or neutered, as appropriate; "no" if not
Color – respondents selected one from {light brown, dark brown, black, white, reddish, yellow/blond, no primary color, other}. Other was not an open-ended response.
Heritage – 1 = single breed, 2 = designer/deliberate mix, 3 = mixed/unknown

---

[1] A similar solicitation on LinkedIn generated only 239 impressions and is unlikely to have driven much traffic.

Age – age in years, as provided by owner; converted if provided in months or fractions of a year.
Weight – approximate weight in pounds, as reported by owner, To convert to kilos, divide by 2.2.
Size – Toy, Small, Medium, Large, Humungous

Figure 1 Data Collection Instrument. Response options collapsed for space reasons.

The Giant Database of Dogs

Building off an idea from Shane Gleason (TAMU-Corpus Christi), I'm building a giant database of dogs to use for classroom statistics exercises. Current or past dogs are welcome. Please provide your best guess for each of the questions below; do a separate form for each dog. The dataset will be made available for public instructional use after it reaches a sufficiently large N for analysis and I have a chance to clean it. Please feel free to share widely!

The data collected will only be used for classroom teaching and assessment purposes. Faculty in a variety of disciplines and institutions will create questions or assignments using this data, which contains no personally identifying data. Students may answer questions like "What is the most common male dog name? Are male dogs heavier on average than female dogs? Does dog size or being fixed (spayed/neutered) affect this conclusion?" All of these can be answered using concepts taught in introductory statistics and research design classes.

What is your good pup's name? [text response]

What gender is your dog? {male, female}

Has your pup been fixed (spayed or neutered, as appropriate)? {yes, no}

What is the primary color of your dog? {light brown, dark brown, black, white, reddish, yellow/blond, no primary color, other [closed response]}

What is your pup's heritage? {Single breed, designer/deliberate mix (e.g., labradoodles), mixed breed/unknown}

How old is/was your pet? [text response]

What is your pup's approximate weight? Please indicate pounds or kilos. [text response]

How big is your dog? {toy, small, medium, large, humungous}