Package: nycdogs (via r-universe)

October 19, 2026

Title New York City Dog Licenses and Bite Incidents by Zip Code

Version 0.3.0

Description A data package containing two datasets on dogs in New York City: active dog licenses from the Department of Health and Mental Hygiene, and reported dog bite incidents.

License MIT + file LICENSE

Encoding UTF-8

Roxygen list(markdown = TRUE)

Depends R (>= 4.1.0), nycmaps

Suggests testthat (>= 3.0.0), tibble (>= 3.0.0)

LazyData true

LazyDataCompression xz

Config/testthat/edition 3

URL https://kjhealy.github.io/nycdogs/, https://github.com/kjhealy/nycdogs

BugReports https://github.com/kjhealy/nycdogs/issues

Remotes kjhealy/nycmaps

Config/roxygen2/version 8.1.0

Config/pak/sysreqs libabsl-dev cmake libgdal-dev gdal-bin libgeos-dev libssl-dev libproj-dev libsqlite3-dev libudunits2-dev

Repository https://kjhealy.r-universe.dev

Date/Publication 2026-10-14 17:26:44 UTC

RemoteUrl https://github.com/kjhealy/nycdogs

RemoteRef HEAD

RemoteSha a4a137a52a3b29bb71c7c552869eabd14ea52010

Contents

nyc_bites 2
nyc_license 3

Index **6**

`nyc_bites` *Reported dog bites in New York City*

Description

Dog bite incidents reported to the Department of Health and Mental Hygiene (DOHMH) between January 1st 2015 and December 31st 2025.

Usage

```
nyc_bites
```

Format

`nyc_bites`**:**

A tibble with 39,082 rows and 12 columns:

unique_id Dog bite case identifier. Unique within `year`, but not across the whole table. See Details.

date_of_bite Date of bite.

year Year of bite.

species Species of animal (all "Dog").

breed Breed of dog, as reported. Whitespace and capitalization are standardized. `NA` where no breed was recorded.

breed_rc Recoded breed, coded in the same way as `breed_rc` in nyc_license. See Details.

age Age of dog at time of bite, as reported. Character. Mostly years. Numbers with "M" indicate months, but many other formats appear.

gender Sex of dog: "M" (male), "F" (female), or "U" (unknown).

spay_neuter `TRUE` if the dog was reported to DOHMH as spayed or neutered. `FALSE` if it was not, or if this is unknown.

borough Borough where the bite occurred. "Other" indicates that the bite took place outside New York City.

zip_code Zip code where the bite occurred. Character. `NA` where not available. Same as `zip`.

zip Zip code where the bite occurred.

Details

Section 11.03 of the NYC Health Code requires all animal bites to be reported within 24 hours of the event. Data is collected from reports received online, by mail, by fax, or by phone to 311 or the DOHMH Animal Bite Unit. Each record represents a single dog bite incident. Information on breed, age, gender, and spayed or neutered status has not been verified by DOHMH and is listed only as reported.

`unique_id` is not unique across the table. The source data combines several releases, and the identifier starts again from 1 in each of them: once for 2015–2017, once for 2018–2021, and once for each year from 2022. The combination of `year` and `unique_id` is unique.

`breed_rc` is made with the same lookup table and rules as `breed_rc` in nyc_license, so that breeds can be compared across the two tables. Reports that only describe a mixed-breed dog (e.g. "Mixed/Other", "Large Mixed Breed") are coded "Mixed Breed", and reports with no usable breed (e.g. "Uncertain", "Small Dog") are coded "Unknown". Free-text descriptions of two or more breeds are left as they are, so `breed_rc` has a long tail of rare values that do not appear in nyc_license.

Other than this the data is deliberately lightly cleaned. Values of `age` and `zip_code` are kept as reported, including malformed ones.

Author(s)

Kieran Healy

Source

NYC Open Data https://data.cityofnewyork.us/Health/DOHMH-Dog-Bite-Data/rsgh-akpg, retrieved October 6th 2026.

`nyc_license` *Dogs of New York*

Description

Dog licenses active in New York City, from annual extracts of the Department of Health and Mental Hygiene (DOHMH) Dog Licensing System. Covers extract years 2016–2018, 2022–2024, and 2026.

Usage

```
nyc_license
```

Format

`nyc_license`**:**

A tibble with 819,323 rows and 12 columns:

animal_name Name of dog, as provided by the owner. Sentence case.

animal_gender Sex of dog, as provided by the owner: "M" or "F".

animal_birth_year Year dog was born, as provided by the owner. `NA` where the source has a spreadsheet error in place of a value.

breed_name Dog breed, as provided by the owner. Whitespace and capitalization are standardized. "Unknown" is a reported category.

breed_rc Recoded breed. Variant, abbreviated, and misspelled names are standardized, and varieties are collapsed to their breed. See Details.

zip_code Owner zip code. Same as `zip`.

zip Owner zip code.

license_issued_date Date license issued.

license_expired_date Date license expires.

extract_year Year the record was extracted.

borough Borough of owner, based on zip code. Some zip codes are in more than one borough but here are counted only once. NA for zip codes not in nycmaps::nyc_zip_sf.

city Nominal city (USPS designation), based on zip code.

Details

The data is sourced from the DOHMH Dog Licensing System (https://a816-healthpsi.nyc.gov/DogLicense), where owners can apply for and renew dog licenses. Each record represents a unique dog license that was active during the year, but not necessarily a unique record per dog, since a license that is renewed during the year results in a separate record of an active license period. Licenses are valid for one to five years.

breed_rc is a recoded version of breed_name:

- Inverted names are put in their usual order, e.g. "Bull Dog, French" becomes "French Bulldog".
- Varieties are collapsed to their breed, e.g. "Dachshund Smooth Coat" and "Dachshund, Long Haired Miniature" become "Dachshund"; "Poodle, Toy" and "Poodle, Standard" become "Poodle"; and "Collie, Rough Coat" becomes "Collie". Breeds that are distinct, such as Miniature, Standard, and Giant Schnauzers or Pembroke and Cardigan Welsh Corgis, are kept apart.
- Long, abbreviated, and misspelled names are standardized, e.g. "American Pit Bull Terrier/Pit Bull" becomes "Pit Bull" and "Cav Kc" becomes "Cavalier King Charles Spaniel".
- Crossbreeds are kept distinct from their breed and are labeled consistently, so that e.g. "Terrier Mix", "Terrier X", and "Terrier Crossbreed" are all "Terrier Crossbreed".
- Free-text descriptions of two or more breeds, e.g. "Beagle/Boxer", are left as they are.

The lookup table used is in data-raw/breed_recodes.csv in the package source. breed_rc in nyc_bites is coded in the same way.

Other than this the data is deliberately lightly cleaned. Owner-provided values are kept as reported, including placeholder names (e.g. "Unknown", "Name not provided"), implausible birth years, and malformed zip codes.

Licenses, dogs, and duplicate records:

Two features of the data matter for any analysis.

First, dog licenses expire. Each row is a record of a time-limited license that was issued, not necessarily the record of a unique individual dog. A dog whose license is renewed appears once for each license period.

Second, the table contains several extracts of the licensing data, marked by the extract_year column. A license that was active in more than one extract year appears in each of them, so there are a substantial number of duplicated rows. About 155,000 of the 819,323 rows repeat an earlier record.

Any analysis of the table should try to de-duplicate the records. For example, arrange the data by extract year, find distinct records based on a number of the identifying columns, and keep only one of them (here, the earliest):

```
nyc_license |>
  dplyr::arrange(extract_year) |>
```

```
  dplyr::distinct(
    animal_name,
    animal_gender,
    animal_birth_year,
    breed_name,
    zip,
    license_issued_date,
    license_expired_date,
    .keep_all = TRUE
  )
```

This leaves one row per license. To count dogs and not licenses, drop the two license date columns from the call to `distinct()`. There is no dog identifier in the data, so this is approximate: two dogs with the same name, sex, birth year, and breed in the same zip code cannot be told apart.

Author(s)

Kieran Healy

Source

NYC Open Data https://data.cityofnewyork.us/Health/NYC-Dog-Licensing-Dataset/nu7n-tubp, retrieved October 5th 2026.

Index

∗ **datasets**
    nyc_bites, 2
    nyc_license, 3

nyc_bites, 2, *4*
nyc_license, *2*, 3, *3*
nycmaps::nyc_zip_sf, *4*