Dallas Animal Analysis

Ace

2025-08-25

# Executive Summary

This report analyzes intake and outcome data for dogs and cats at the Dallas Animal Shelter (Week 18). The goal is to understand trends in intake, outcomes, and breed-specific patterns, identify areas of concern, and provide actionable insights.

**Key findings:** - Dogs are more likely than cats to be returned to owners or adopted. - Cats have higher rates of euthanasia. - Certain dog breeds (Pitbulls, Labradors, German Shepherds) are disproportionately represented among euthanized animals. - Stray animals are more likely to be euthanized than owner surrenders. - Council districts with the highest intake counts can inform outreach and resource allocation.

# readxl installed because data sheet is excel and not csv  
 #janitor for cleaning data  
   
  
# Load packages  
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.2 ✔ tibble 3.3.0  
## ✔ lubridate 1.9.4 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.4   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(readxl)  
library(janitor)

##   
## Attaching package: 'janitor'  
##   
## The following objects are masked from 'package:stats':  
##   
## chisq.test, fisher.test

library(lubridate)

# Data Cleaning & Preparation

Steps taken: - Filtered dataset to dogs and cats only. - Removed duplicate outcomes (outcome\_type == "DUPLICATE"). - Converted Excel numeric dates (intake\_date, outcome\_date) into R Date objects. - Created derived columns: intake\_year, outcome\_year, days\_in\_shelter. - Filtered out rows with missing outcome dates only for time-based analyses, preserving other rows for categorical analyses.

**Rationale:** Ensures analysis is based on clean, accurate data without losing valuable records unnecessarily.

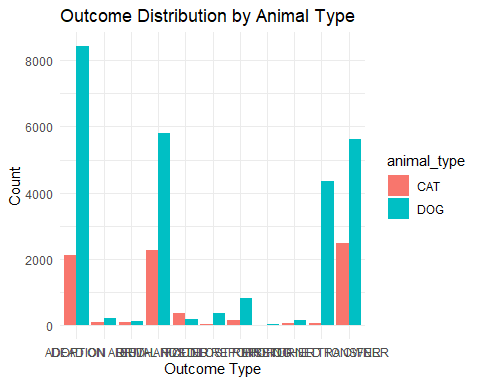
# Load dataset  
sh\_raw <- read\_excel("C:/Users/HOME/Desktop/CSV files/data-analysis-journey/Dallas\_Animals/week18\_dallas\_animals.xlsx", sheet = "simple")  
  
# Clean and prepare main dataset  
sh <- sh\_raw %>%  
 clean\_names() %>%  
 filter(animal\_type %in% c("DOG", "CAT")) %>%  
 filter(outcome\_type != "DUPLICATE") %>%  
 mutate(  
 intake\_date = as.Date(as.numeric(intake\_date), origin = "1899-12-30"),  
 outcome\_date = suppressWarnings(as.Date(as.numeric(outcome\_date), origin = "1899-12-30"))  
 )  
  
# Create subset for time-based analyses  
sh\_with\_dates <- sh %>%  
 filter(!is.na(outcome\_date)) %>%  
 mutate(  
 intake\_year = year(intake\_date),  
 outcome\_year = year(outcome\_date),  
 days\_in\_shelter = as.numeric(outcome\_date - intake\_date)  
 )  
  
 #)  
  
view(sh\_with\_dates)

# Overall Outcome Distribution

**Objective:** Compare outcomes between dogs and cats.

**Observations:** Dogs dominate adoptions and returns to owners. Cats have higher euthanasia rates. Very few animals die on arrival.

sh %>%  
 count(animal\_type, outcome\_type) %>%  
 ggplot(aes(x = outcome\_type, y = n, fill = animal\_type)) +  
 geom\_col(position = "dodge") +  
 labs(  
 title = "Outcome Distribution by Animal Type",  
 x = "Outcome Type", y = "Count"  
 ) +  
 theme\_minimal()



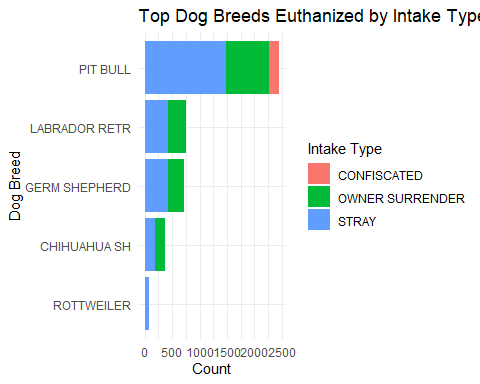
# Breed-Specific Euthanasia Patterns

**Objective:** Identify dog breeds most frequently euthanized.

**Observations:** Pitbulls, Labradors, German Shepherds, and Chihuahuas are the most frequently euthanized. Intake type contributes: strays and owner surrenders show different patterns.

**Interpretation:** Potential breed bias or resource allocation issues at the shelter. Useful for policy review and breed-specific outreach programs.

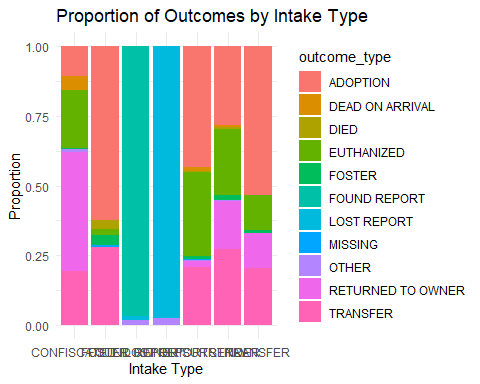
# Filter euthanized dogs  
euth\_dogs <- sh %>%  
 filter(animal\_type == "DOG", outcome\_type == "EUTHANIZED")  
  
# Top 10 breeds  
euth\_breed\_intake <- euth\_dogs %>%  
 count(animal\_breed, intake\_type, sort = TRUE) %>%  
 slice\_max(n, n = 10)  
  
# Bar chart rotated  
ggplot(euth\_breed\_intake, aes(x = reorder(animal\_breed, n), y = n, fill = intake\_type)) +  
 geom\_col() +  
 coord\_flip() +  
 labs(  
 title = "Top Dog Breeds Euthanized by Intake Type",  
 x = "Dog Breed",  
 y = "Count",  
 fill = "Intake Type"  
 ) +  
 theme\_minimal()

 # Outcomes by Intake Type

**Objective:** Understand how intake circumstances relate to outcomes.

**Observations:** Stray animals are disproportionately euthanized. Owner surrenders have higher adoption rates. Foster-intake animals: very few “missing” cases observed (documented anomalies).

sh %>%  
 count(intake\_type, outcome\_type) %>%  
 ggplot(aes(x = intake\_type, y = n, fill = outcome\_type)) +  
 geom\_col(position = "fill") +  
 labs(  
 title = "Proportion of Outcomes by Intake Type",  
 y = "Proportion", x = "Intake Type"  
 ) +  
 theme\_minimal()

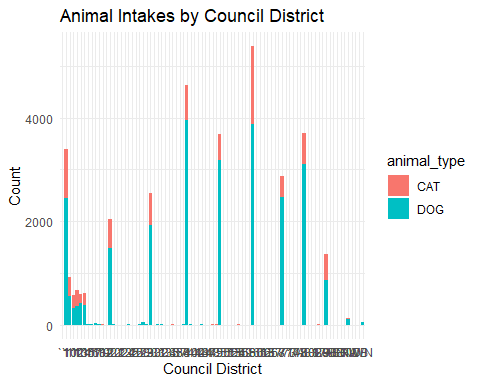


# Intakes by Council District

**Objective:** Geographic distribution of shelter intake.

**Observations:** Certain districts have consistently high intakes. Useful for community outreach or low-cost spay/neuter programs.

# Intakes by Council District -------------------------------  
  
sh %>%  
 count(council\_district, animal\_type) %>%  
 ggplot(aes(x = council\_district, y = n, fill = animal\_type)) +  
 geom\_col() +  
 labs(  
 title = "Animal Intakes by Council District",  
 x = "Council District", y = "Count"  
 ) +  
 theme\_minimal()



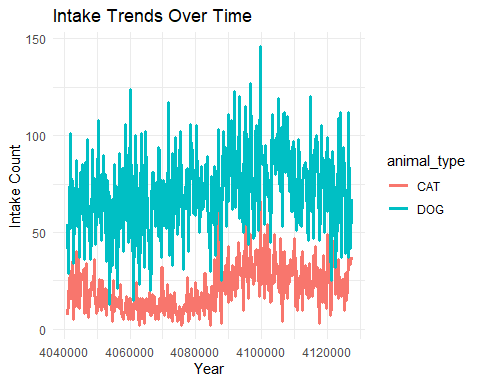
# Time Trends

**Objective:** Explore annual intake trends.

**Observations:** Dogs consistently have higher intake counts than cats. Trends indicate increasing/decreasing intake for specific years (based on cleaned subset).

# Time Trend of Intakes -------------------------------------  
  
sh\_with\_dates %>%  
 count(intake\_year, animal\_type) %>%  
 ggplot(aes(x = intake\_year, y = n, color = animal\_type)) +  
 geom\_line(size = 1.2) +  
 labs(  
 title = "Intake Trends Over Time",  
 x = "Year", y = "Intake Count"  
 ) +  
 theme\_minimal()

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
## ℹ Please use `linewidth` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.

 # Notes on Data Quality / Anomalies

* Some Excel dates required conversion due to non-standard formats.
* Rows with missing outcome dates were excluded only when calculating time-based metrics.
* Foster animals marked “MISSING” included duplicates; retained in dataset as possible distinct cases.

# Recommendations

* **Microchipping campaigns:** Increase return-to-owner rates.
* **Breed-specific interventions:** Review euthanasia policies for high-risk breeds.
* **Geographic outreach:** Focus on high-intake council districts.
* **Time-based monitoring:** Track trends to anticipate resource needs.