

# Learning from the CBARQ Survey

...

Lindsey Schaeffer

# **The Canine Behavioral Assessment and Research Questionnaire**

# Problem:

Are certain dog breeds more likely to exhibit certain behaviors?

Millions of dog bites per year in the United States

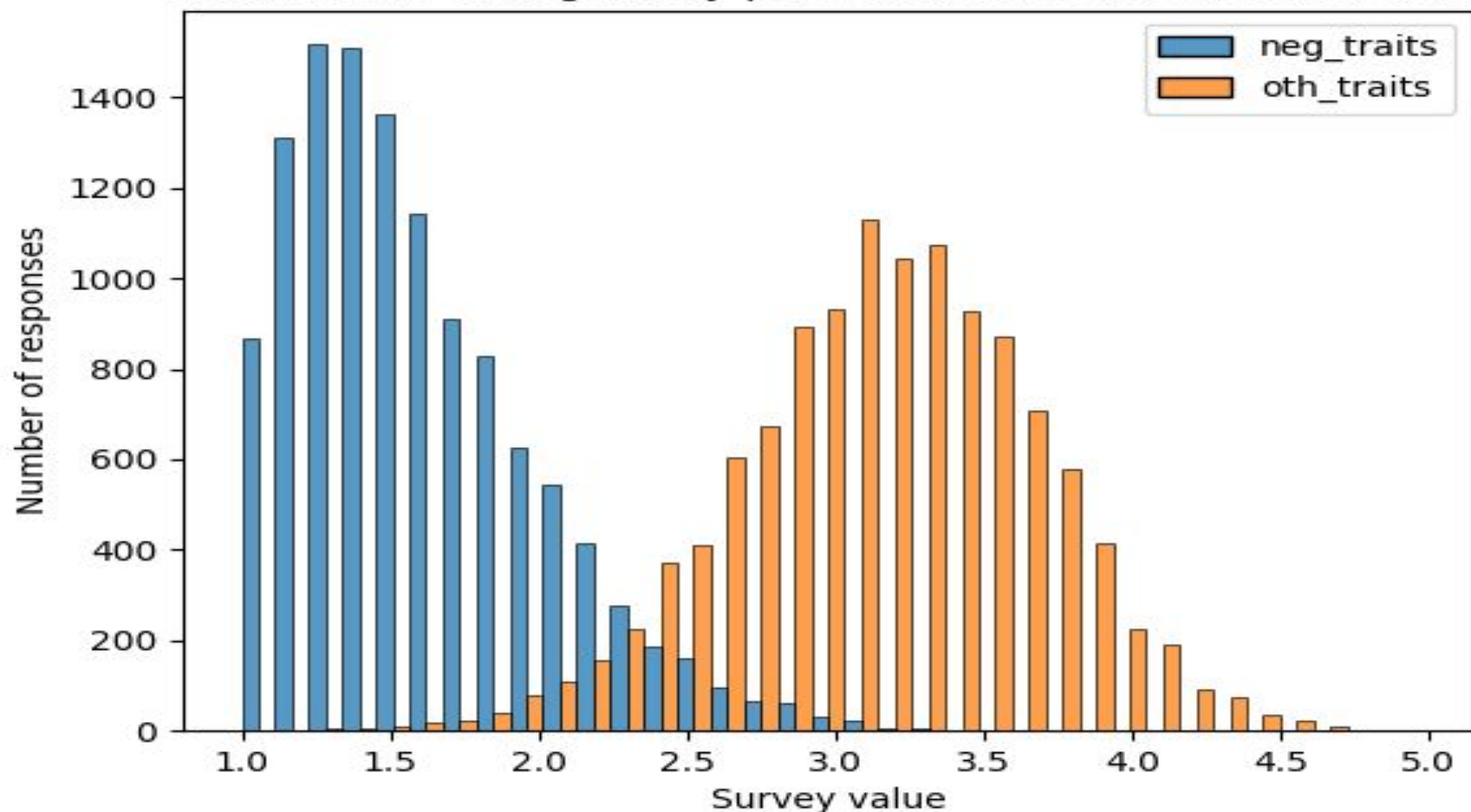
Millions of dogs in shelters

Cause and effect? Or simply correlation?

# Categories

- stranger-directed aggression
- owner-directed aggression
- dog-directed aggression
- dog-directed fear
- stranger-directed fear
- nonsocial fear
- separation-related behavior problems
- dog rivalry
- touch sensitivity
- attachment and attention-seeking behavior
- energy level
- excitability
- trainability
- chasing

Distributions of negatively-perceived traits versus other traits

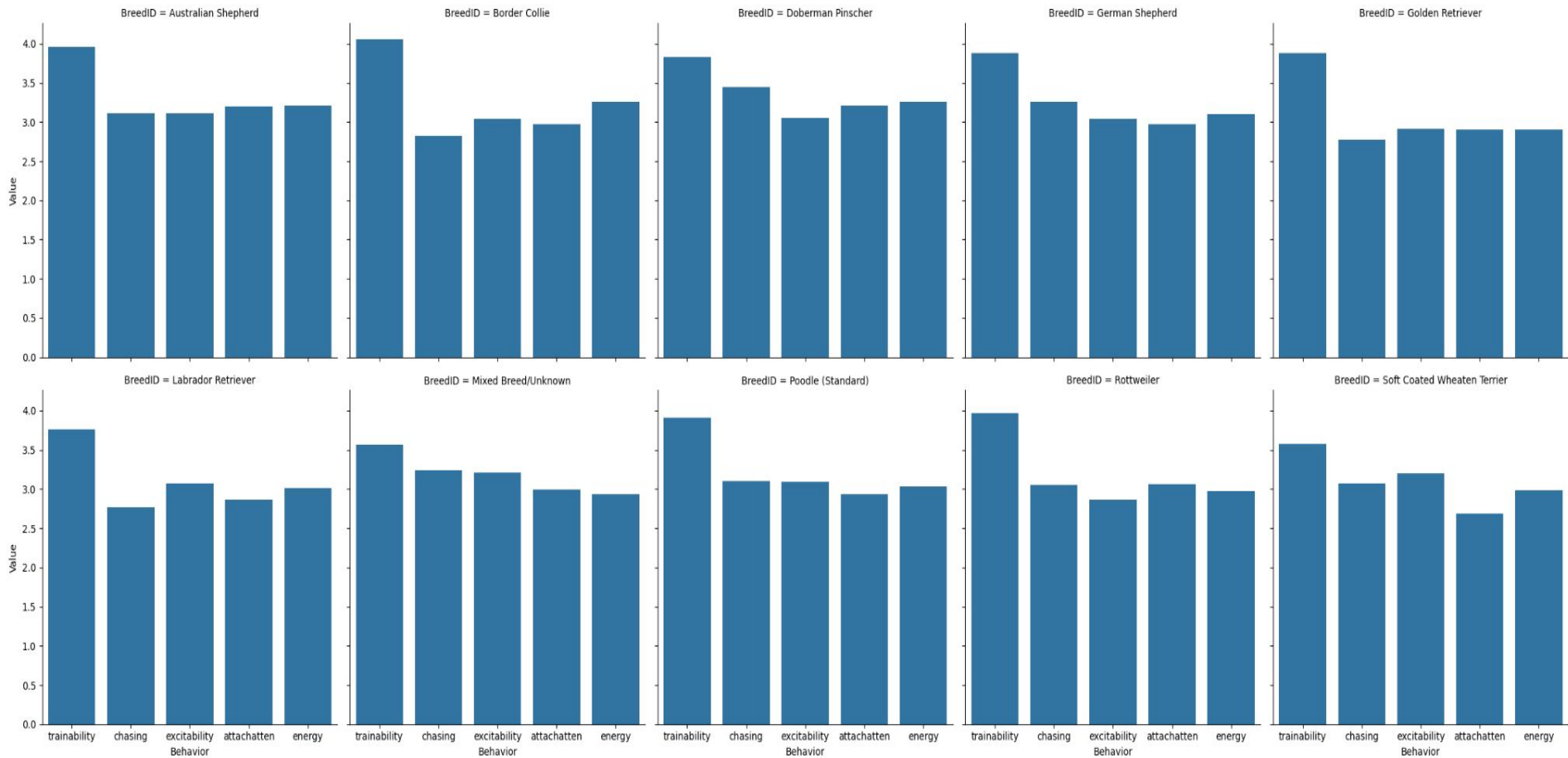


- trainability
- chasing
- excitability
- energy
- attachment/attention-seeking behaviors

# Breeds

- Australian Shepherd
- Border Collie
- Doberman Pinscher
- German Shepherd
- Golden Retriever
- Labrador Retriever
- Mixed Breed/Unknown
- Poodle (Standard)
- Rottweiler
- Soft Coated Wheaten Terrier

# Mean of Trait Values Per Breed





# ANOVA Results

- Trainability:  $2.79230972e-100$
- Chasing:  $1.24067542e-042$
- Excitability:  $8.95360550e-018$
- Attachment/Attention-Seeking Behaviors:  $5.57490138e-019$
- Energy:  $5.13445541e-010$

# Models

- Linear Regression
- Lasso Regression
- XGBoost
- Random Forest Regressor

# Linear Regression

- Each point corresponds with a dog's breed and its behavior score
- Fits data to a straight line of best fit using all points
- Simplest model

# Lasso Model

- Least Absolute Shrinkage and Selection Operator
- Adds a penalty for too many variables
- If certain variables don't help much in predictions, it shrinks their coefficients to 0
- Prevents overfitting

# XGBoost Model

- Builds decision trees in an ensemble
- Can handle more complex relationships
- Does not assume a straight line of relation
- Prevents overfitting and generalizes well to new data

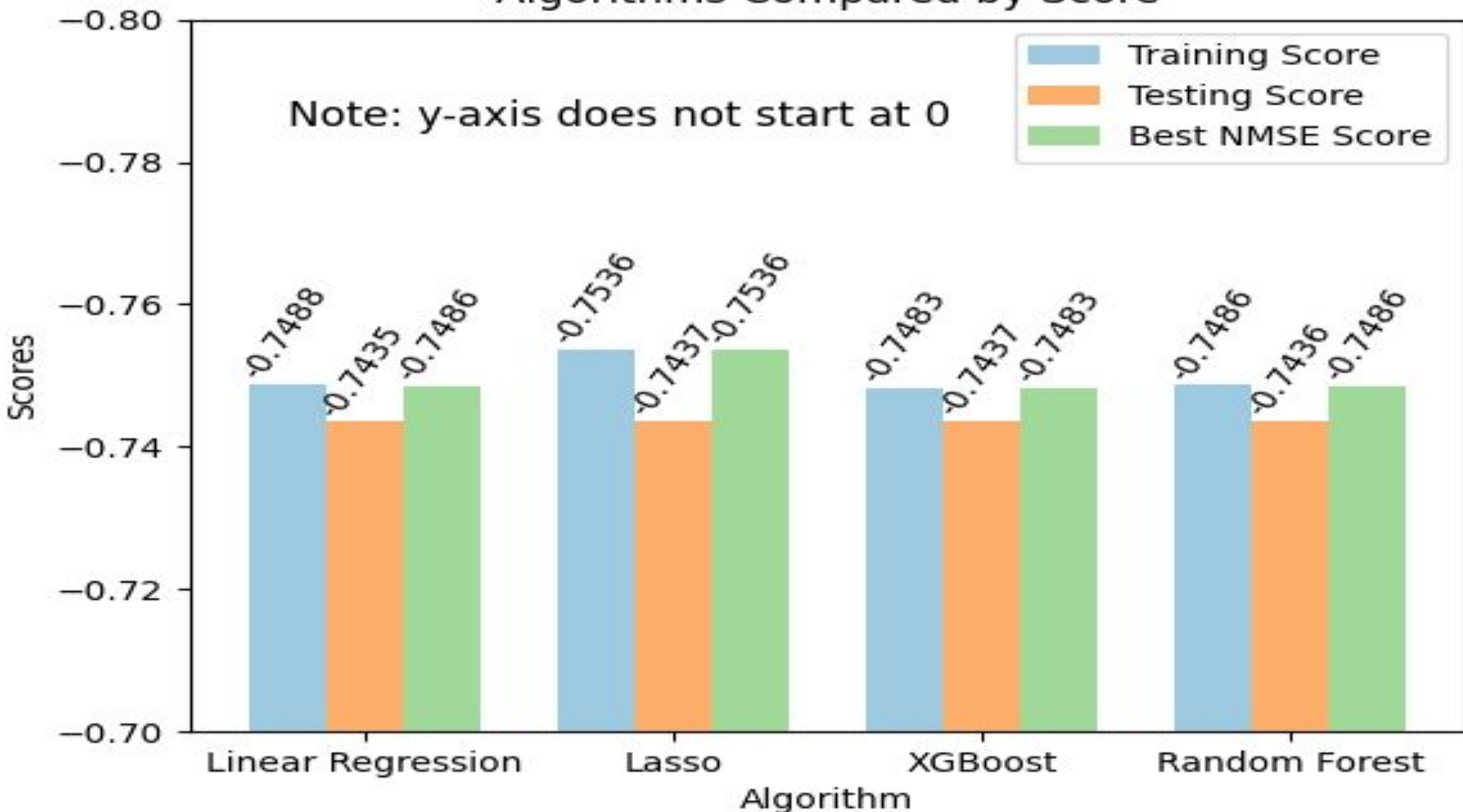
# Random Forest Regressor

- Non-linear relationships
- Can handle greater number of breeds
- Can provide feature importance scores
- Lacks the boosting and regularization that XGBoost uses

# Algorithms Compared by Score

Note: y-axis does not start at 0

Training Score  
Testing Score  
Best NMSE Score



# Final Model

## Parameters:

- `copy_X = True`
- `Fit_intercept = False`
- `positive = False`

## Scoring:

- Best NMSE score:  
-0.7485746361602028
- NMSE training score:  
-0.747773582924948
- NMSE testing score:  
-0.7426515099102272



# Final Model

- The simplicity of the data (breeds and behaviors) meant that the linear regression worked fine
- With more (non-behavior) variables, the more complicated models may have worked better
- Random Forest Regressor may be better if more breeds are included

# More Research

## Additional Behaviors:

- Barking
- Howling

## Additional Traits:

- Age
- Sex
- Neutered/Spayed
- Background