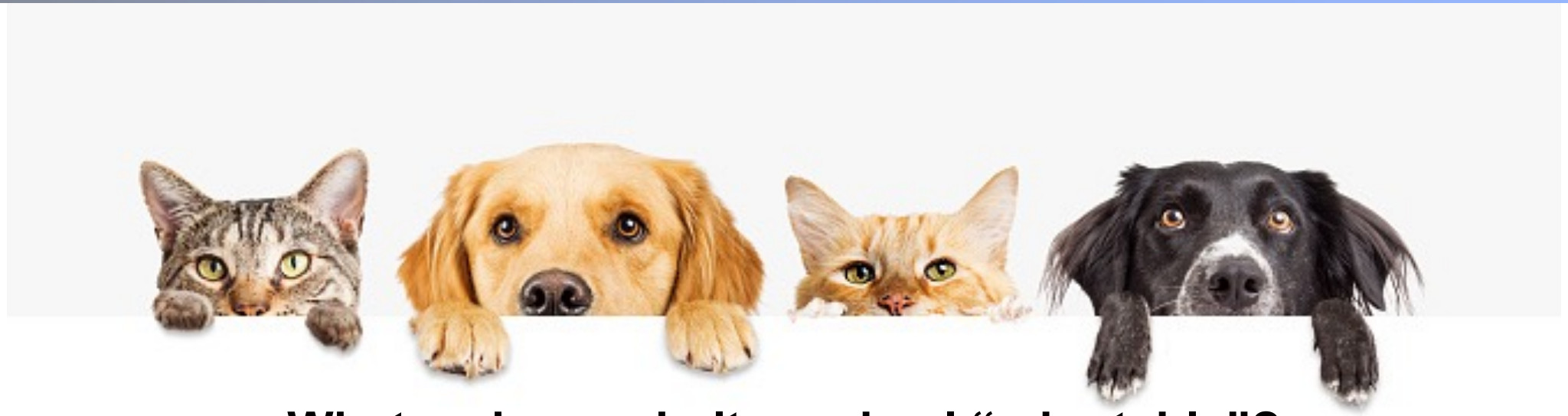


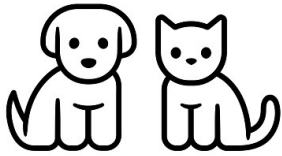
What makes a shelter animal “adoptable”?

Helping shelters evaluate pet features to increase rehoming

DSL, Group 27:
Brandon Hayes
Steven Phan
Vasiliki Tassopoulou

28 April 2023

Motivation: Why model animal adoption rates?



Amount of space for both cats and dogs in shelters is becoming extremely limited.

Philly is currently facing an adoption crisis, with an increase to 13% euthanasia rate

'Filled to the brim': Philly's only open-intake animal shelter is in desperate need of adopters, fosters

"It's the first time in ages we are being left with decisions of euthanizing for space," said Sarah Barnett, "which is something we never thought would really happen again."

By Tom MacDonald · December 28, 2022

ACCT Philly closed to public while it works to contain canine flu outbreak

By clearing out, the nonprofit can isolate dogs who are sick or have been exposed to canine flu.

By 6abc Digital Staff and Beccah Hendrickson

Monday, April 24, 2023 5:57AM

Some reasons why this problem has become prominent in 2022-2023:

- Intake rate is higher than adoption rate
- Potential fosters/adopters having financial constraints
- Ongoing employment worries pushes potential adopters away
- Increased eviction rates post-pandemic
- Reduced budget for shelter organizations since they are "not essential"

https://www.shelteranimalscount.org/as-animal-intake-outpaces-adoption-philly-shelters-and-rescue-groups-are-in-crisis-mode/

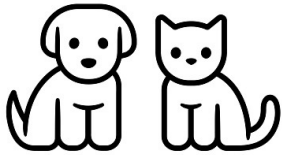
Motivation: Why model animal adoption rates?

How we do go about modeling adoption? Compile the animals' features and also track how long it takes for them to get adopted.

This is Athena:
Clearly cute
Young tabby, brown fur
Smaller-than-avg size
Loving & playful
Independent
Slightly wobbly (neuro. condition)

This is Christopher:
Handsome, squishy, & sleepy
Older ginger cat
Average size
Loving
Shy but loving
FIV+ requiring regular health checkups

Motivation: Why model animal adoption rates?



Data source:

- PetFinder.my is a Malaysian welfare platform that has created a database of 150,000 animals up for adoption since 2008.
- PetFinder.my has made publicly available datasets for data scientists to assess adoption rates, build AI tools to see how an animal's cuteness affects homing, and more.

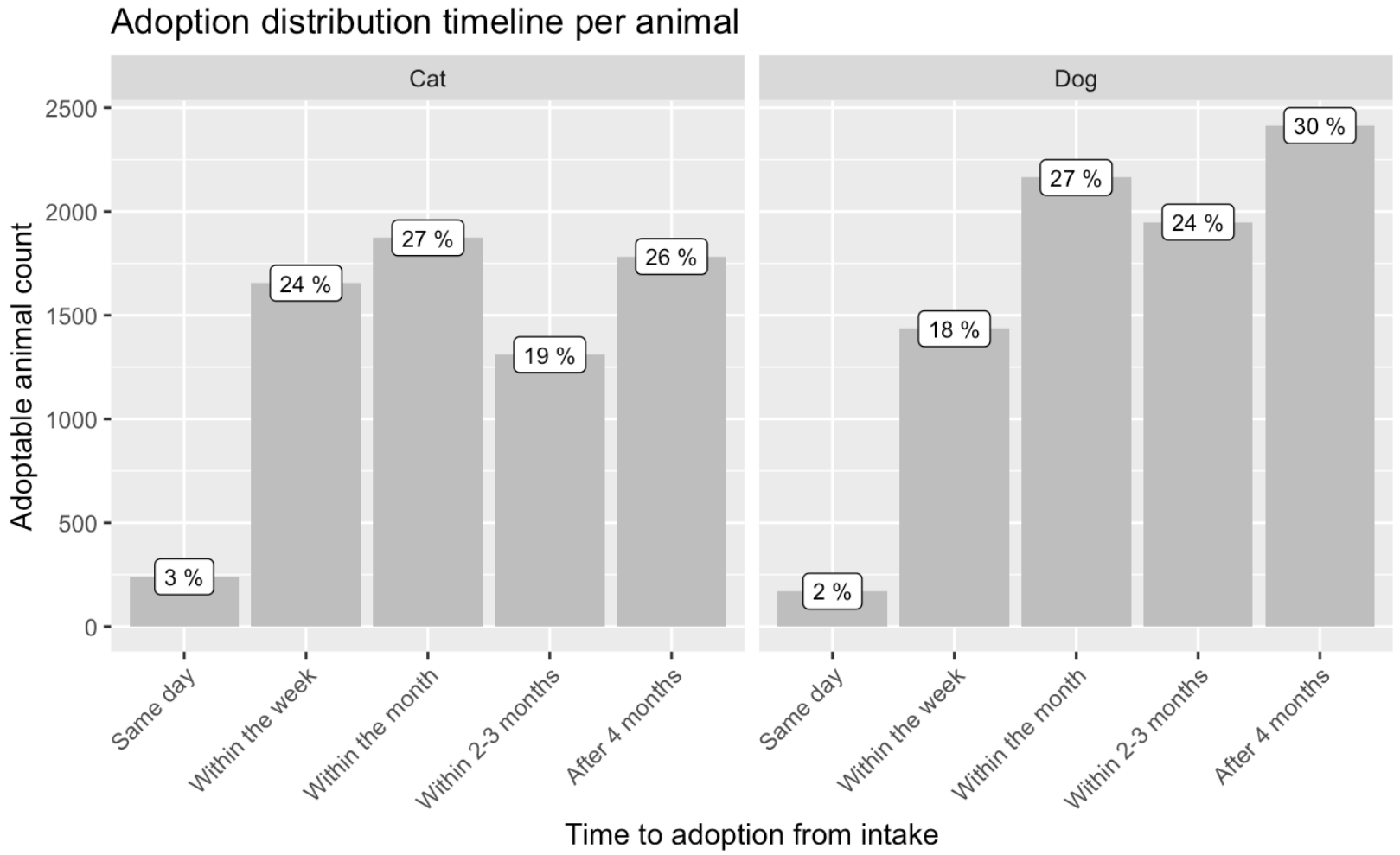
What data are we given?

- **PetID**
- **AdoptionSpeed** - Categorical speed of adoption
  Type - Type of animal
- **Name**
- **Age** - Age of pet when listed, in months
- **Breed1** - Primary breed
- **Breed2** - Secondary breed
- **Gender**
- **Color1**
- **Color2**
- **Color3**
- **MaturitySize** - Size at maturity
- **FurLength**
- **Vaccinated** - Pet has been vaccinated
- **Dewormed** - Pet has been dewormed
- **Sterilized** - Pet has been spayed / neutered
- **Health Quantity** - Number of pets represented in profile
- **Fee** - Adoption fee
- **State** - State location in Malaysia
- **RescuerID** - Unique hash ID of rescuer
- **VideoAmt** - Total uploaded videos for this pet
- **PhotoAmt** - Total uploaded photos for this pet
- **Description** - Profile write-up for this pet

EDA

What does the distribution of adoption speed look like?

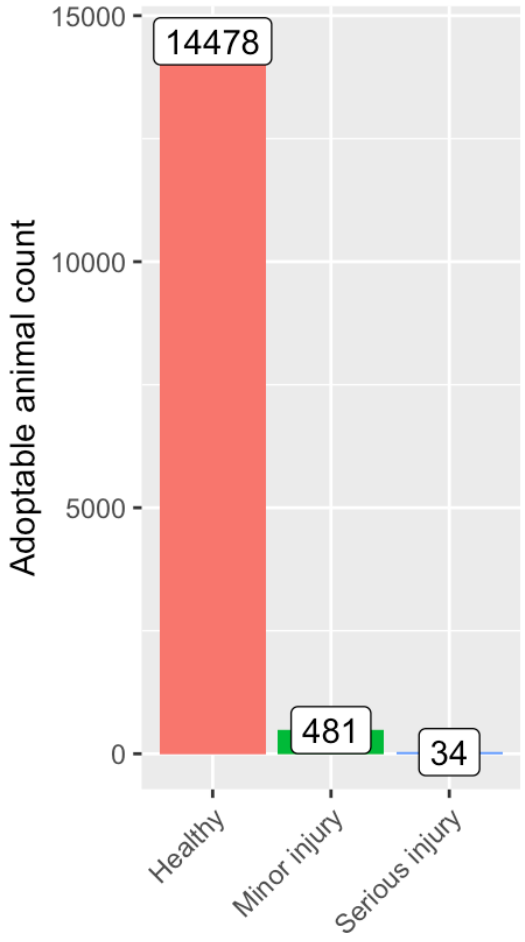
- Overall, more animals get adopted than not (within 3 months after being listed)
- Immediate adoption is rare
- Roughly similar adoption speeds for all other timing bins

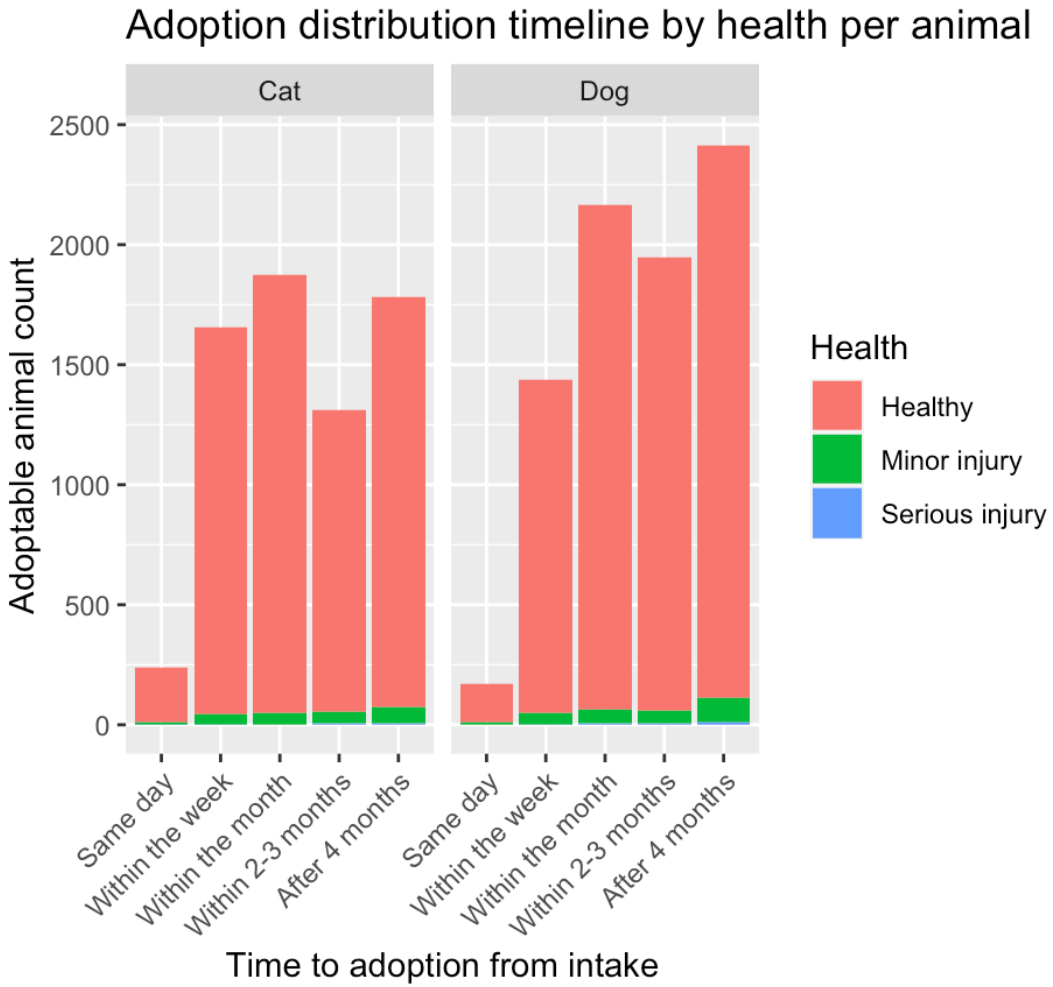


EDA

What does the distribution of adoption speed look like when we consider health?

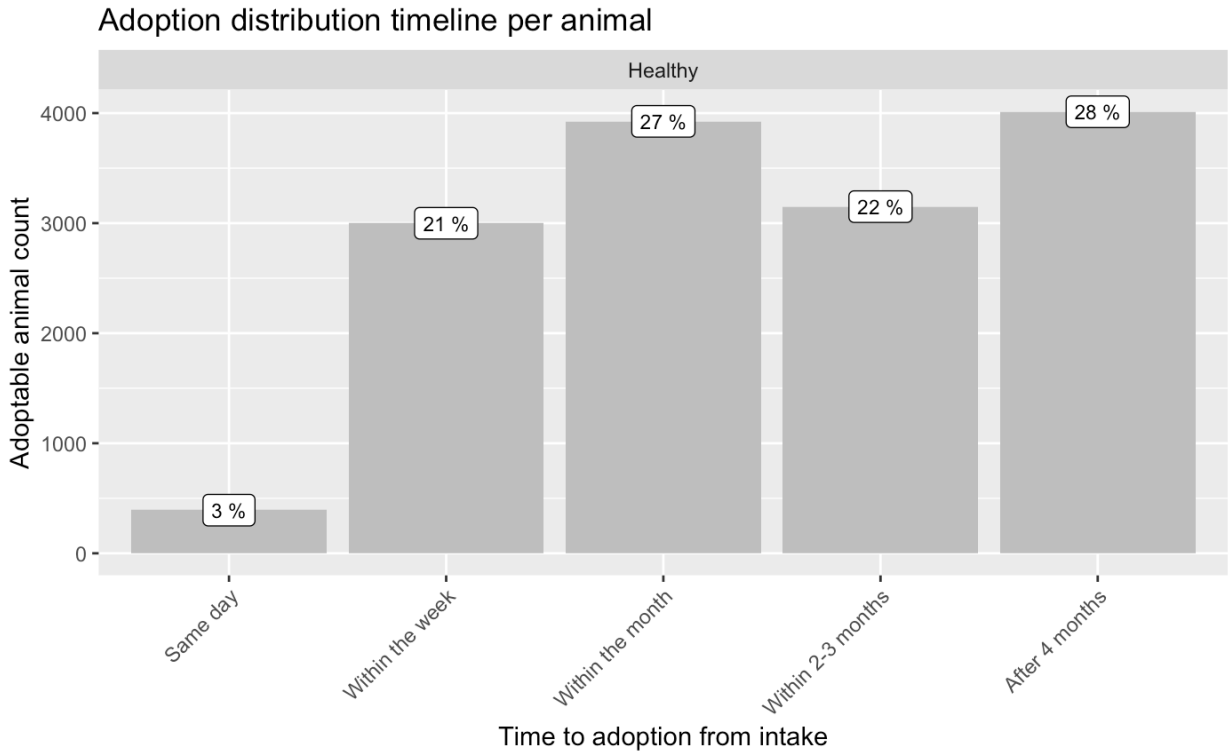
- Imbalanced – few entries for serious injury, not a very large amount for minor injury

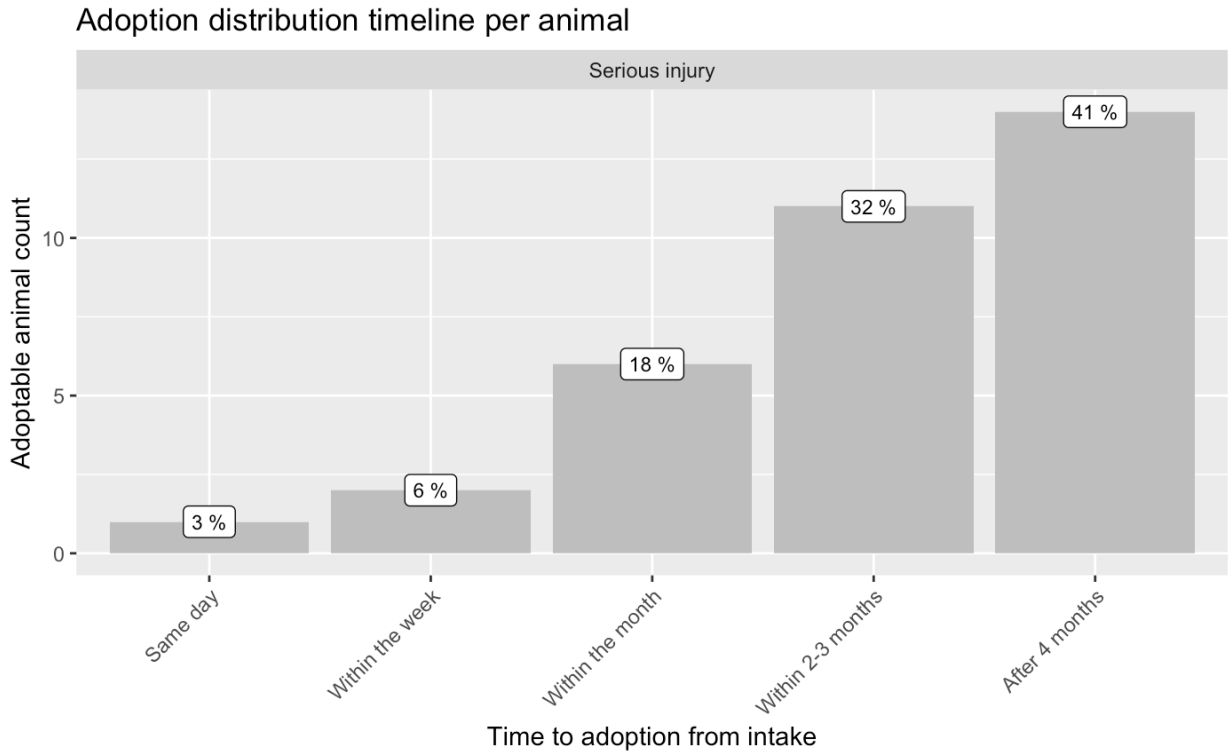




EDA

Closer look into injured animals

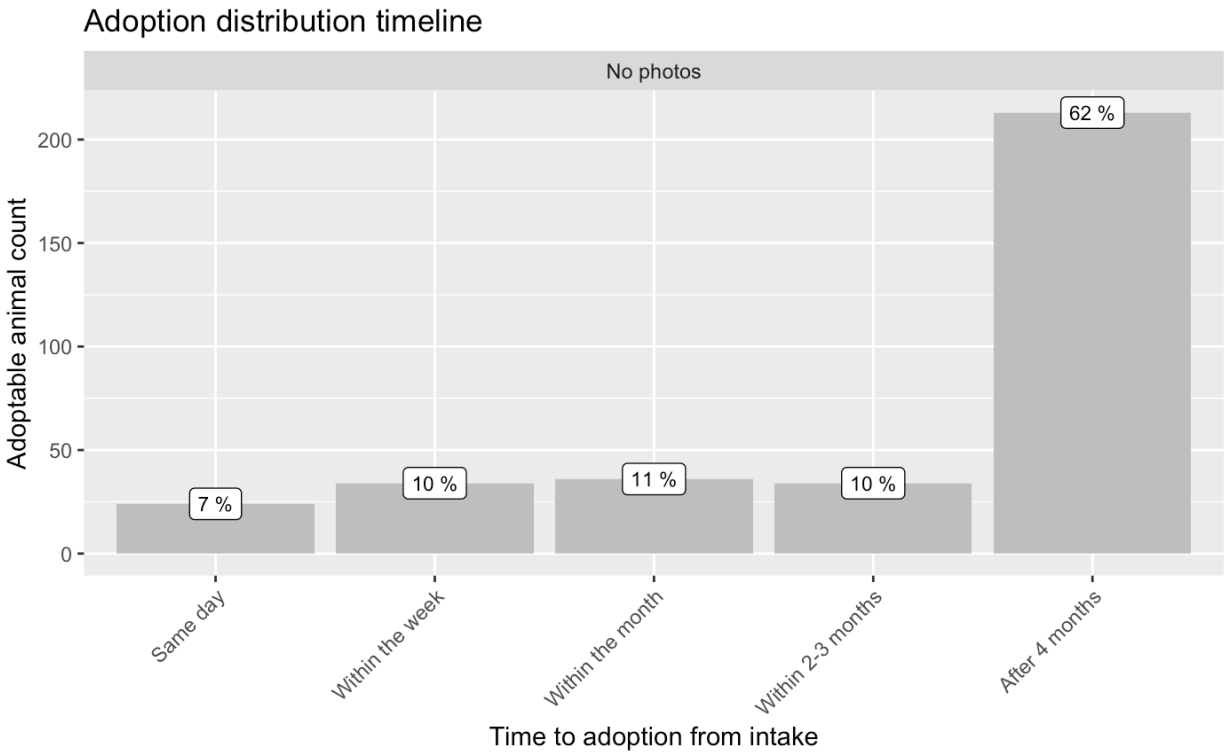


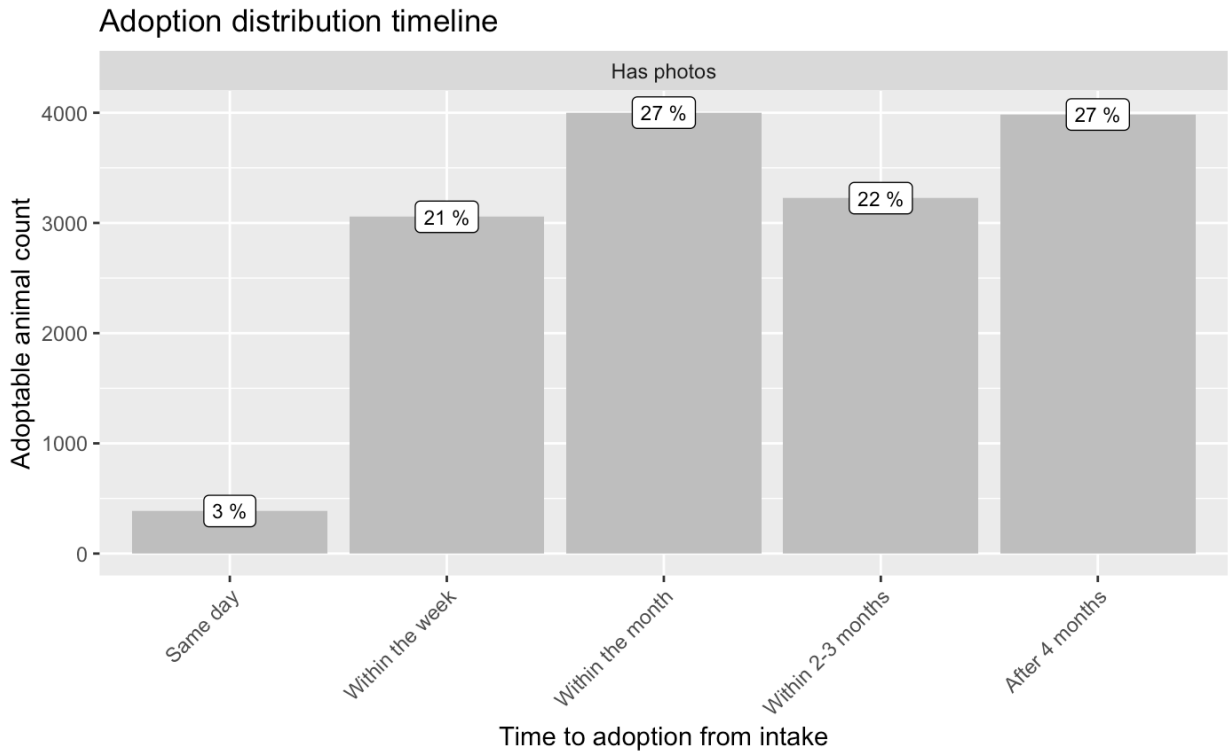


Injuries do show a trend of increasing length to adoption

EDA

How does social media presence affect adoption speed?





Photos of adoptable animal tend to increase adoption speed

Motivation: Why model animal adoption rates?

Key questions:

- Based on animal characteristics given, can we develop a model that predicts how quickly an animal can be adopted?
- Can EDA and the model give us insights in what features may lead to slower adoption rates so we can help more at-risk animals?
- If we struggle to build a successful model, what possible recommendations can we think of for the data collection to someday build a better model?

Model: LASSO

Input Predictors:

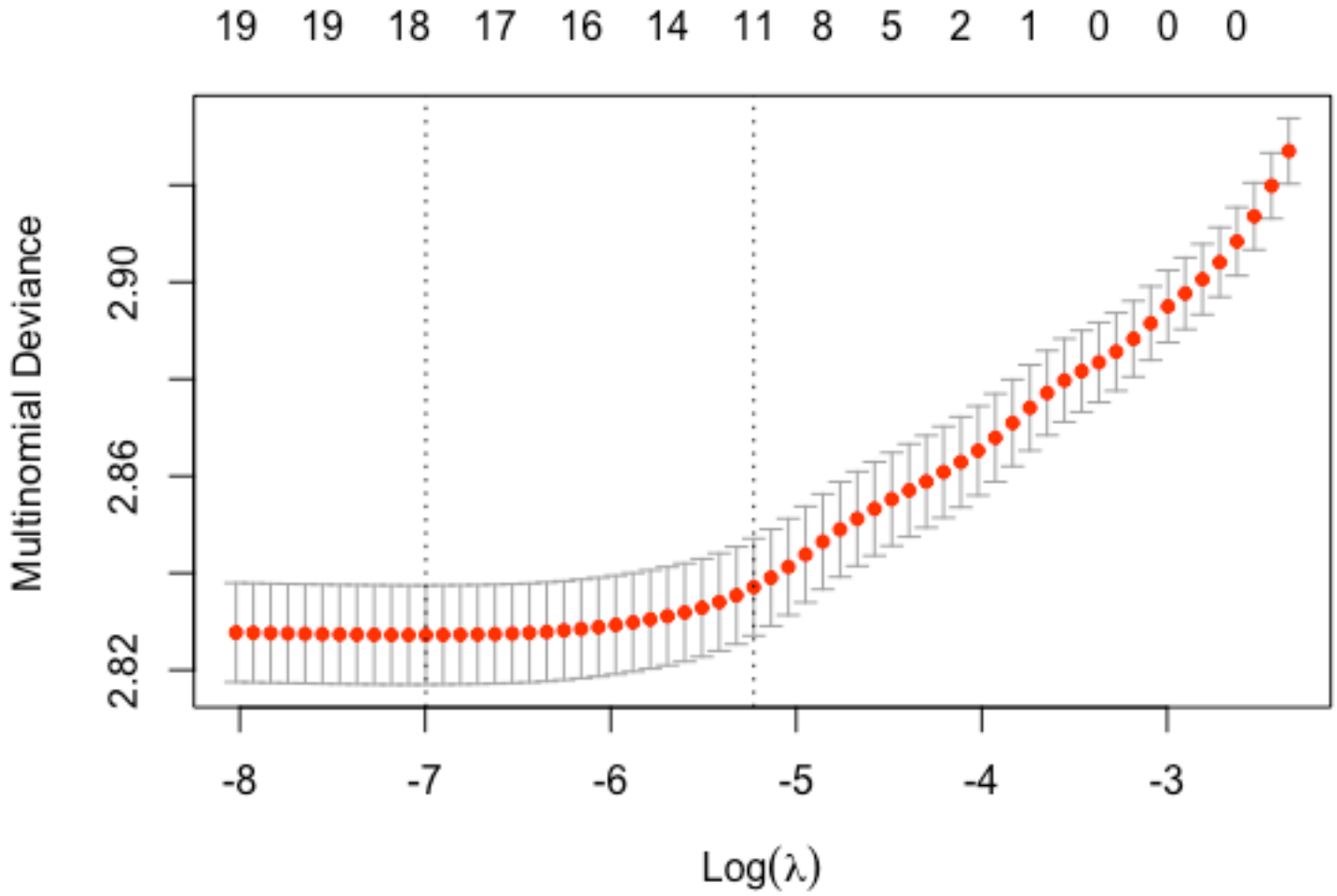
- Type, Age, Gender, Color1, MaturitySize, FurLength, Vaccinated, Dewormed, Sterilized, Health, Fee, PhotoAmt+VideoAmt

LASSO Parameters:

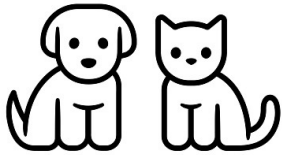
- alpha=1
- nfolds=10
- family=multinomial
- lambda.1se (for prediction)

Results:

Test accuracy = 34.5%



Model: Multinomial Logistic Regression



Input Predictors (from LASSO):

- Age, Gender, Color1, MaturitySize, FurLength, Vaccinated, Dewormed, Sterilized, Health, Fee, PhotoAmt+VideoAmt

Results:

Test accuracy = 34.6%

$$\begin{aligned}\log odds = {} & 2.36 + 0.012 * Age + 0.212 * GenderF + 0.630 * GenderMix \\ & -0.329 * Color1Brown - 0.238 * Color1Golden - 0.826 * Color1Cream \\ & -0.727 * Color1Gray + 0.566 * MaturitySizeMed - 0.928 * MaturitySizeXL \\ & -0.387 * FurLengthMed - 1.85 * FurLengthLong - 0.204 * VaccinatedN \\ & +0.094 * DewormedN - 0.593 * SterilizedN - 0.069 * HealthMildInjury \\ & +0.565 * FeeY + 0.114 * Photo\&VideoY\end{aligned}$$

Model: Neural network

Attempt #1:

NN architecture:

- Input: 15 units, ReLU activation
- Hidden: 10 units, ReLU activation
- Output: 5 units, Softmax

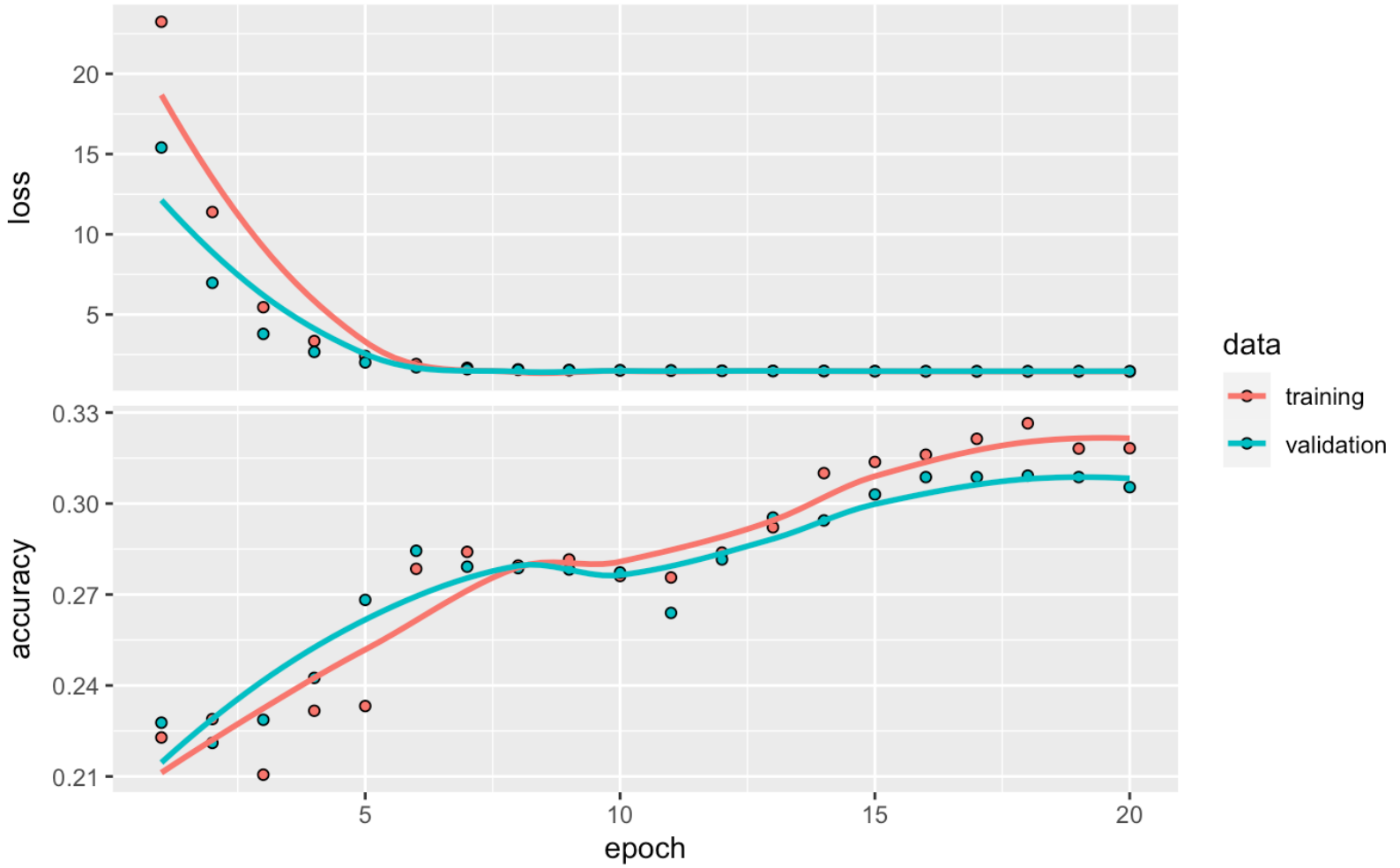
Note: more hidden layers, only slightly approves accuracy

Optimized NN parameters:

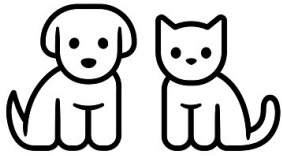
- 20 epochs
- 500 batch

Results:

Test accuracy = 28.89%



Model: Neural network



Attempt #1:

NN architecture:

- Input: 15 units, ReLU activation
- Hidden: 10 units, ReLU activation
- Output: 5 units, Softmax

Note: more hidden layers, only slightly approves accuracy

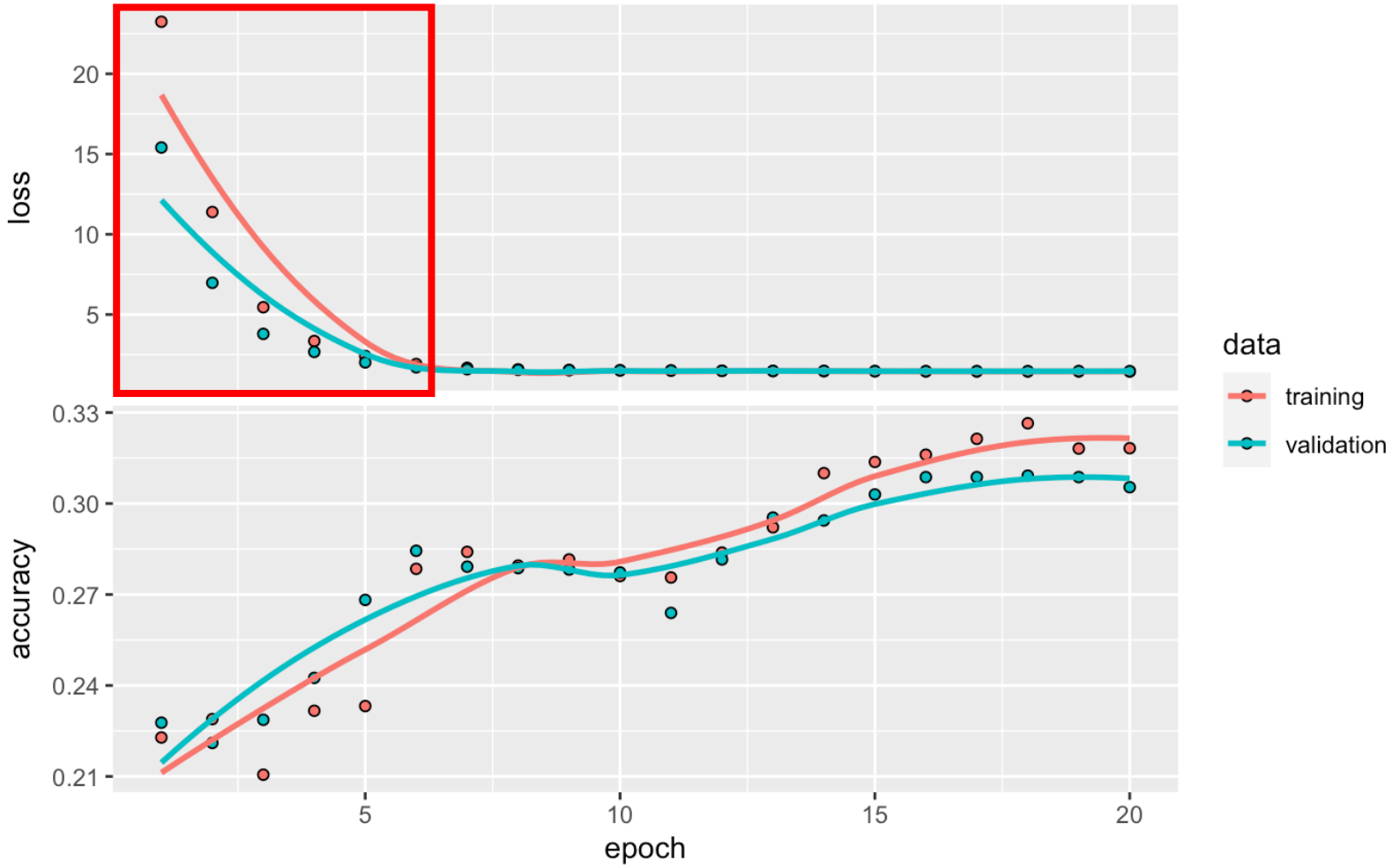
Optimized NN parameters:

- 20 epochs
- 500 batch

Results:

Test accuracy = 28.89%

Validation below training at early epochs. Possible indication that validation is not representative of training. Validation data was partitioned from the same dataset though, leading us to hypothesize data is noisy. Incorrigible regardless of NN architecture.



Model: Neural network

Attempt #2:

NN architecture:

- Input: 15 units, ReLU activation
- Hidden: 10 units, ReLU activation
- Output: 5 units, Softmax

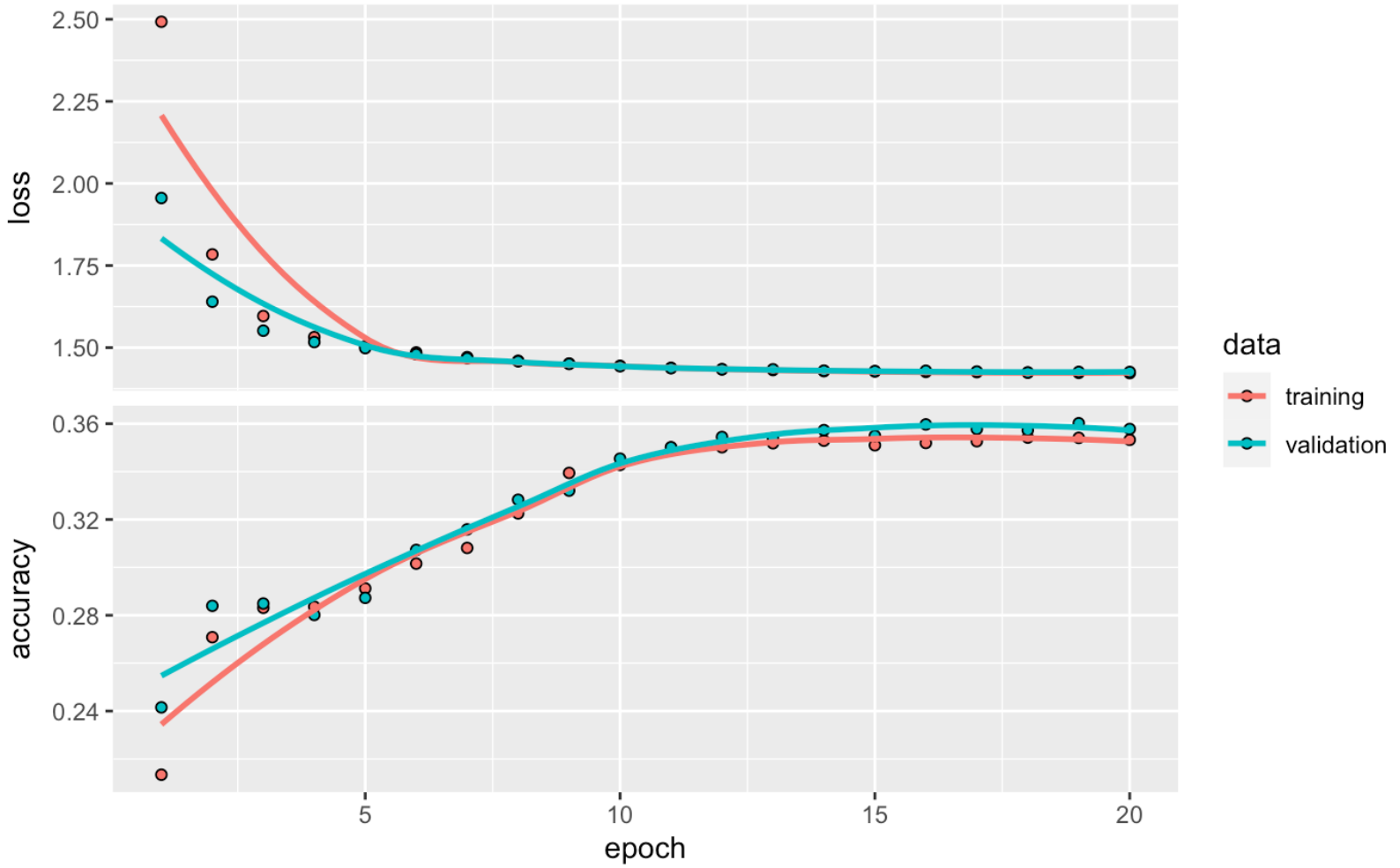
Note: more hidden layers, only slightly approves accuracy

Optimized NN parameters:

- 20 epochs
- 500 batch

Results:

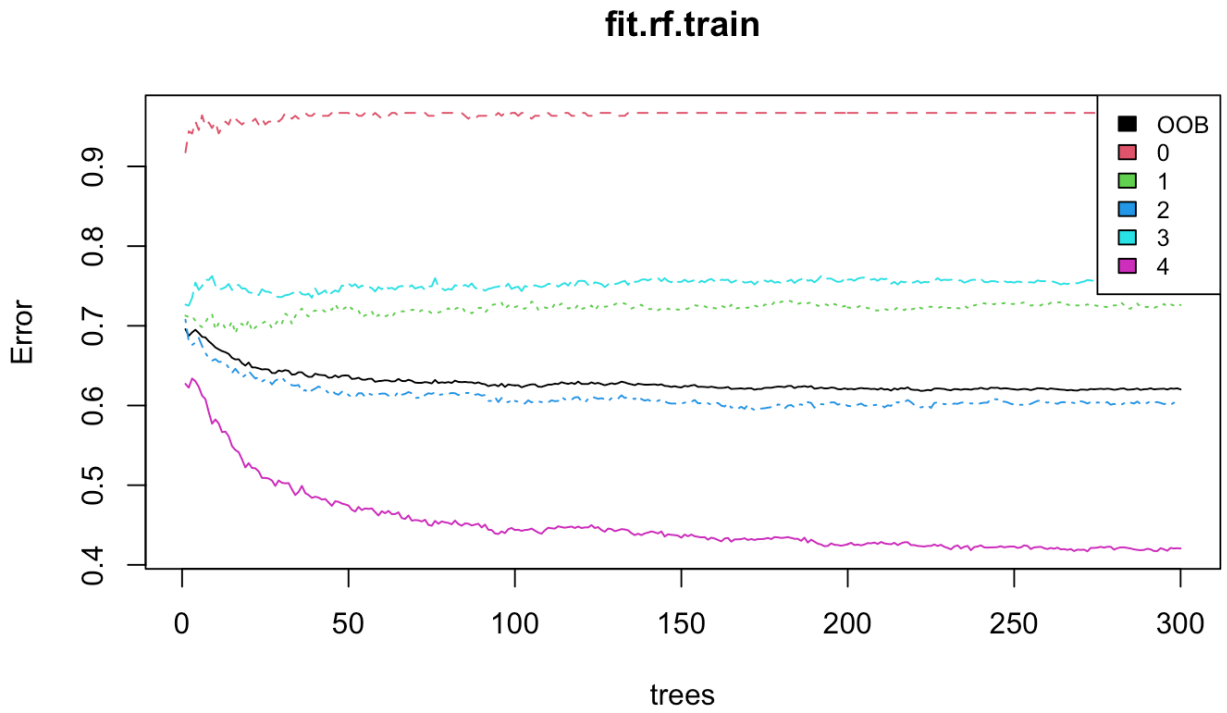
Test accuracy = 35.53%

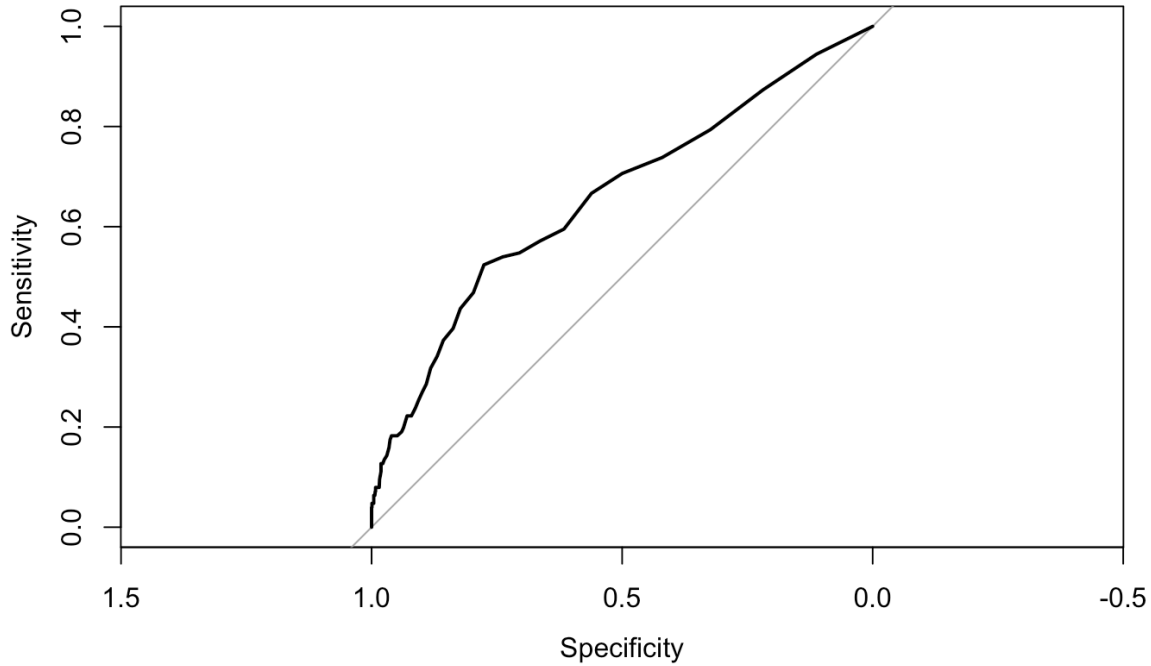


Model: Random Forest

Random Forest parameters:

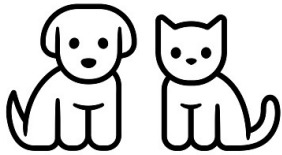
- mtry: 5
- Number of trees: 300





Testing accuracy = 37.88%
AUC = 0.5579

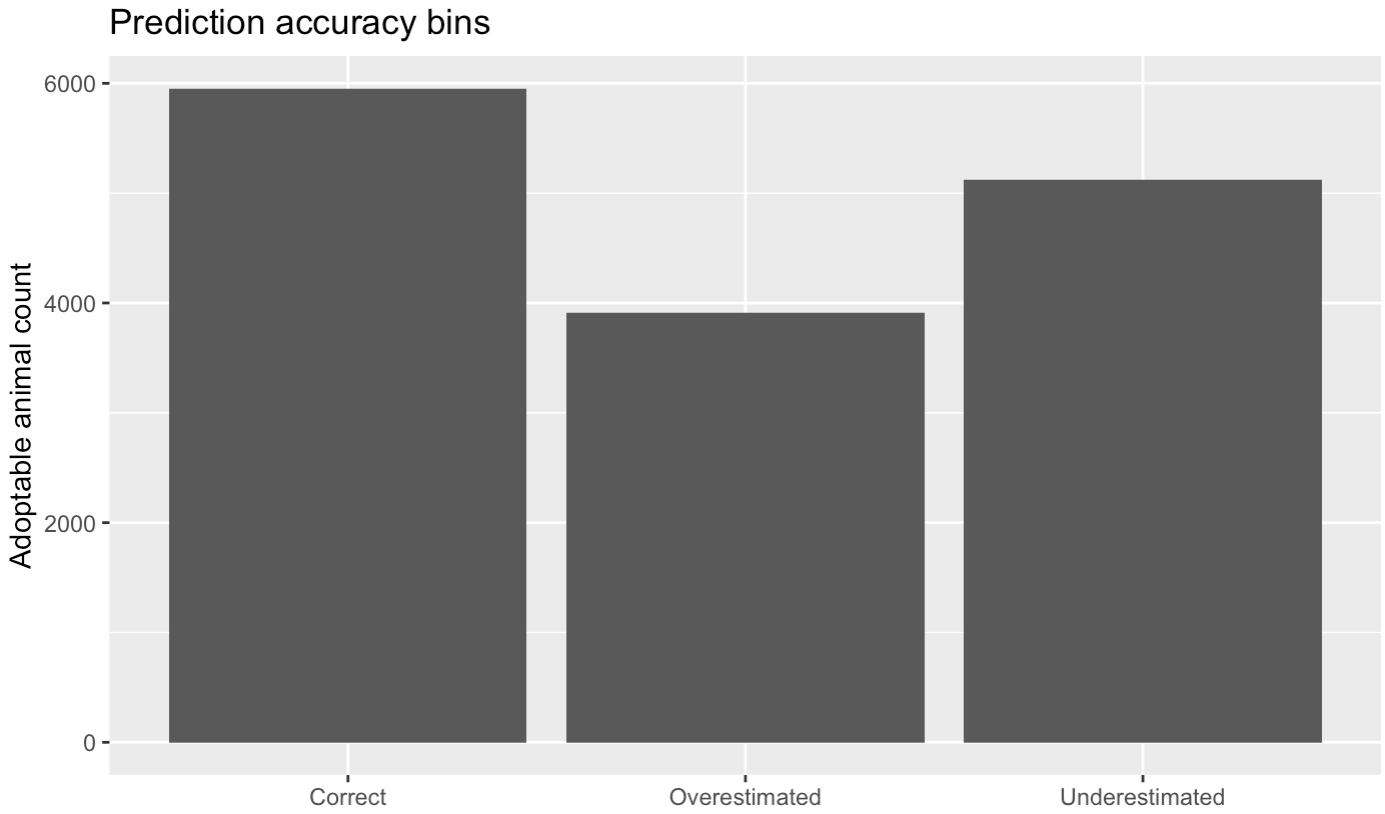
Summary of all attempted models



Modeling method	Testing accuracy
LASSO	34.47%
Multinomial logistic regression	34.60%
Neural network	35.53%
Convolutional neural network	30.96%
Random Forest	37.88%
XGBoost	37.58%

Summary of all attempted models

Given the low accuracy, what can we do?



Pooling together correct predictions and those that overestimate how long it takes for an animal gets adopted, 2/3 of these predictions can be useful.

Overestimates can be conservative approaches in prediction adoption rate.

The major problem is that is 1/3 of the estimates will underestimate adoption rate.

Recommendations

Why was this dataset challenging?

- Lots of “uncertain” entries regarding vaccination, deworming & spay
- Poor adoption rate classification, categories cover days range, weeks range, month range. More quantitative or standardized intervals could be useful
- No information on animal behavior or sociability
- No information on adopter socioeconomic status – this could affect willingness to adopt a “less healthy” animal
- Some class imbalance
- Data limited to Malaysia – this can affect data entry & standardization

What can be done to improve future data entries?

- Include information on animal sociability (with humans & other animals)
- Standardize adoption rate information
- Consider including information on the need for referrals for adoption
- Shelter information could be useful as a possible confound factor
- Are animals sick with a transmissible disease (ex. FIV)?