R-Ladies for PAWS Datathon (2019)

R-Ladies Philadelphia

May 12, 2019

Contents

Executive Summary

Animals

- In 2018, most homeless cats and dogs arrived at PAWS via partner organization, Project MEOW (cats), and ACCT diversions (dogs)
- Peak intake months were March - May for cats and December and January for dogs
- PAWS received noticeably more animals than it placed in adoptive homes in March - June (cats) and July and October (dogs)
- PAWS found adoptive homes for over 92% of homeless cats and dogs that were taken in
- Adoptions were fastest for cats 12 weeks - 1 year old and dogs 6 months - 4 year old
- Slower adoption timelines were observed cats (52 days vs. 20 days for dogs), animals that were strays, tranfers-in, or in PAWS' Foster Program, very young animals (cats under 12 weeks and dogs under 4 weeks) and older dogs (over 10 years)

Applications

- Potential adopters waited longer before adopting cats (19 days) than dogs (8 days)
- PAWS staff spent more time, on average, on each application checklist item for cats (1.2 days) than for dogs (0.9 days)
- Applications that resulted in adoptions from the Foster Program and from Offsite Locations took longer to reach the adoption outcome than applications that resulted in adoptions from other locations
- Checking vet references was the most time consuming item in the application checklist process (approx. 1.8 days)
- Applicants whose applications were processed more quickly tended to live in neighborhoods where fewer people relied on cell data for internet (dogs), where more people 25-34 years were enrolled in school (dogs), where there were more kids enrolled in preschool or nursery school (dogs), and where there were fewer kids in grades 5-8 (cats)
- Red flagged applications required more resources from PAWS staff, as they were sent to a manager for decision, or required more information from the applicant; these applications tended to come from neihborhoods with a lower household median income
- Denied applications were very few, and tended to have no mention of a home pet policy and more reports of potentially problematic incidents with past pets; they also came from neighborhoods with a lower household median income
- Factors positively influencing adoptions included expressing interest in a specific animal, having no children in the home, and applying to adopt a dog
- Cat applicants tended to have a lower monthly pet budget (up to $100), expected to leave the animal home alone for longer hours (8 hours), lived with no other adult in the home, had no children, and lived in areas with a lower median income, and where more people relied on public transit for their commute
- Dog applicants tended to have a higher monthly pet budget ($200-$500), shared their home with 1 other adult, had no children, and lived in areas with a higher median income, and where fewer people relied on public transit for their commute
- PAWS' Twitter activity was uncorrelated with adoption applications

Problem definition

The 2019 R-Ladies for PAWS Datathon aimed to help the Philadelphia Animal Welfare Society (PAWS) improve its adoptions processes. PAWS is a non-profit organization dedicated to saving Philadelphia's homeless and at-risk animals. It is the city's largest rescue partner and provider of low-cost, basic veterinary care for pet owners and rescue organizations that cannot otherwise access or afford it. Through its 3 no-kill shelters, foster care network, and special events, PAWS finds loving homes for thousands of animals each year. For this data challenge, PAWS made de-identified data from 2018 available. We developed analytic approaches to better understand the following topics:

1. An animal's trajectory at PAWS
2. An adoption application's trajectory at PAWS
3. Geographic characteristics that influence adoptions
4. Social media activity that could influence adoptions

Datasets

- **Application data** from online application forms for dogs and cats, received between August - December 2018
- **Animal status data** from PetPoint, including intake date, release date, release location, health status, etc.
- **Application processing data** from Trello, which inlcuded *cards* (information about when the application started being processed by PAWS and what label PAWS determined it should have (e.g. "ready to adopt", "red flag", etc.)) and *actions related to checklist items* (information about the steps in the background check process (e.g. vet check, Pet Point check, landlord check, etc.)).

Timeline and Workflow

Prior to kickoff, data were obtained and preprocessed. After downloading the data from multiple sources, individual entries were matched based on first name, last name, address, and other relevant variables. Once matched, any identifiable information was removed from the dataset. Subsequently, the deidentified data were made available to the group for analysis.

February 12: Kickoff Meetup: At this event, the project was introduced and teams were formed to work on one of the 4 topics outlined above. All participants were encouraged to join the R-Ladies Philly Slack workspace and to set up a GitHub account (which was used as the main collaborative platform). After the kickoff meetup, groups worked together online, getting together on an as-needed basis. Questions were asked and answered via Slack, with an occasional clarification email to PAWS.

March 26: Conclusion Meetup: At this meetup, teams presented their results and discussed their experiences. Individual team reports were finalized in the weeks following this meetup, and then integrated into the present report.

After the March meetup, reports were combined and edited into the present report. The full data and analyses can be found on the RLadies Philly GitHub account.

Results

1. Animal Trajectories

This analysis investigated factors relating to an animal's trajectory in the PAWS system using PetPoint data from 2018. The group operationalized animal trajectory as wait time and outcome (e.g. adoption), with wait time defined as time in days from intake to outcome. We restricted analyses to dogs and cats, as other animals' data points were sparse and compromised statistical power. Our primary factors of interest included animal characteristics (size, breed, health), intake type, and seasonal patterns.

We found that most animals arrived at PAWS via partner transfers-in, followed by Project MEOW (cats) and ACCT diversions (dogs). Furthermore, most animals that arrived at PAWS found an adopter. The 'wait time' from intake to outcome was longer for cats than for dogs, and longer for strays, transfers-in, and animals in the Foster Program. The most quickly adopted animals were 12 week old to 1 year old cats and 6 months old to 4 year old dogs, while the most slowly adopted animals were cats under 12 weeks and dogs under 4 weeks or over 10 years. Peak intake months for cats were March-May and releases were August and November, while dog intakes and releases both peaked in December and January. Increased resources in spring/summer months for young, unhealthy cats may shorten wait times and alleviate staff burden.

Contributors

Alex Lesicko, PhD is a postdoctoral fellow studying auditory coding at the University of Pennsylvania. She recently moved to Philadelphia from Chicago, where she completed her PhD in neuroscience.

Jake Riley is a clinical data analyst at the Children's Hospital of Philadelphia (CHOP). He enjoys developing tools for analytic teams and specializes on data visualization and geospatial information systems (GIS).

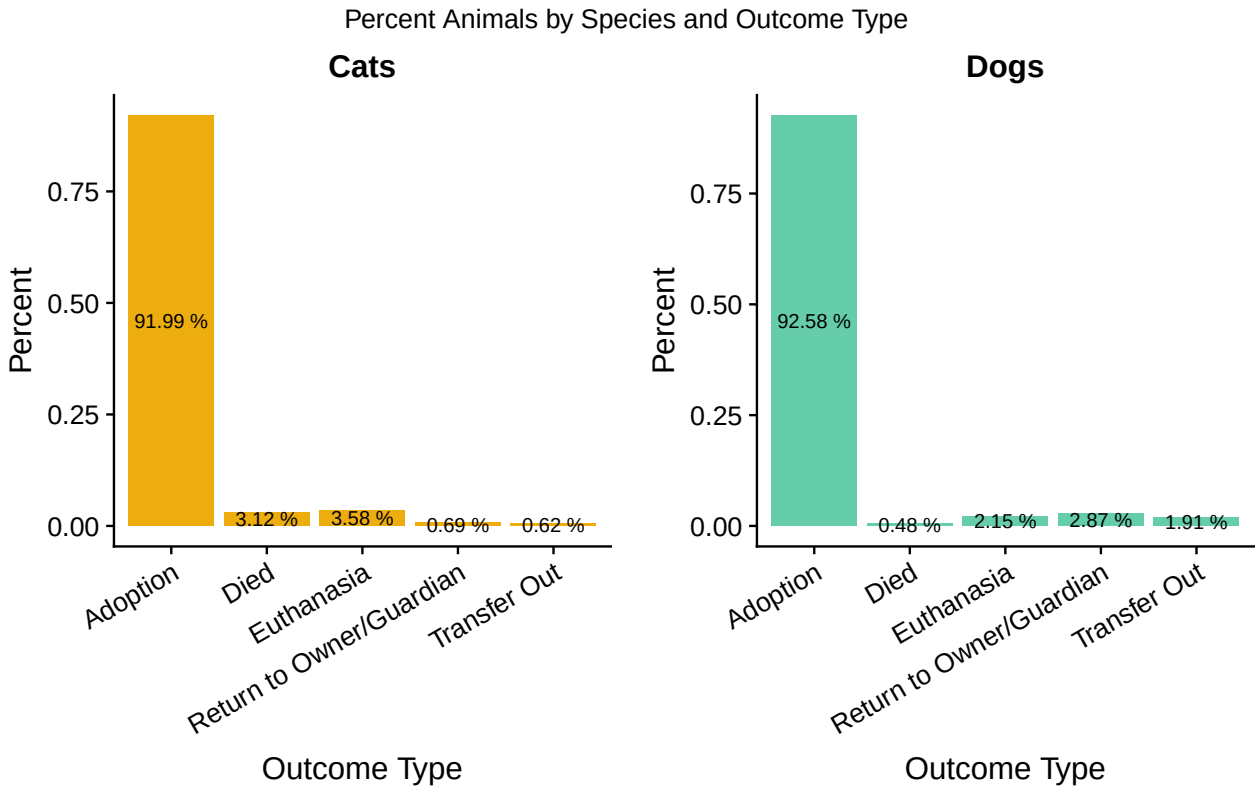
Javier Jasso is a certified speech–language pathologist and a PhD candidate in communication sciences and disorders at the University of Texas at Austin. Javier has expertise in the assessment of culturally/linguistically diverse children, focusing on bilingual language acquisition.

Katerina Placek is a PhD candidate in neuroscience at the University of Pennsylvania and a co-organizer of R-Ladies Philly. She enjoys integrating outreach with teaching and learning in the local data science community.

Overall Animal Outcomes

Among the cats that were taken in at PAWS in 2018, 92% (2,390 cats) were adopted, 1.31% (34 cats) were transferred out or returned to their owner or guardian, and 6.7% (174 cats) died or were euthanized. For dogs, 93% (387 dogs) were adopted, 4.78% (20 dogs) were transferred out or returned to their owners or guardians, and 2.63% (11 dogs) died or were euthanized. In subsequent analyses, we focused only on those animals whose outcome was "Adoption", "Returned to Owner or Guardian" or "Transfer Out".

Across all animals, the median wait time (time in days from intake to release date) was 45 days (51 days for cats, and 18 days for dogs). Cats waited approx. 18.5 days to be transferred out, 4.5 days to be reunited with their owner, and 52 days to find an adopter, while dogs waited 1.5 days to be transferred out, 2 days to be reunited with their owner, and 20 days to be adopted.



Animal Characteristics

For each species, we visualized which PetPoint variables contributed to differences in wait time, and focused our subsequent analyses on the variables with the greatest contributions to wait times: primary breed, age group, and health at intake.

Primary Breed

We found that most cats from PAWS in 2018 were 'domestic short hair', while dogs' breed variability was greater. The most frequent dog breeds were Terrier, Shih Tzu, Pit Bull Terrier, and Chihuahua. Among dogs, Shih Tzus tended to have shorter wait times whereas Terriers tended to have longer wait times. However, given the large number of unique dog breeds in the PetPoint dataset, it was difficult to draw definitive conclusions. Therefore, we classified dogs into 3 size categories based on average weight per breed:

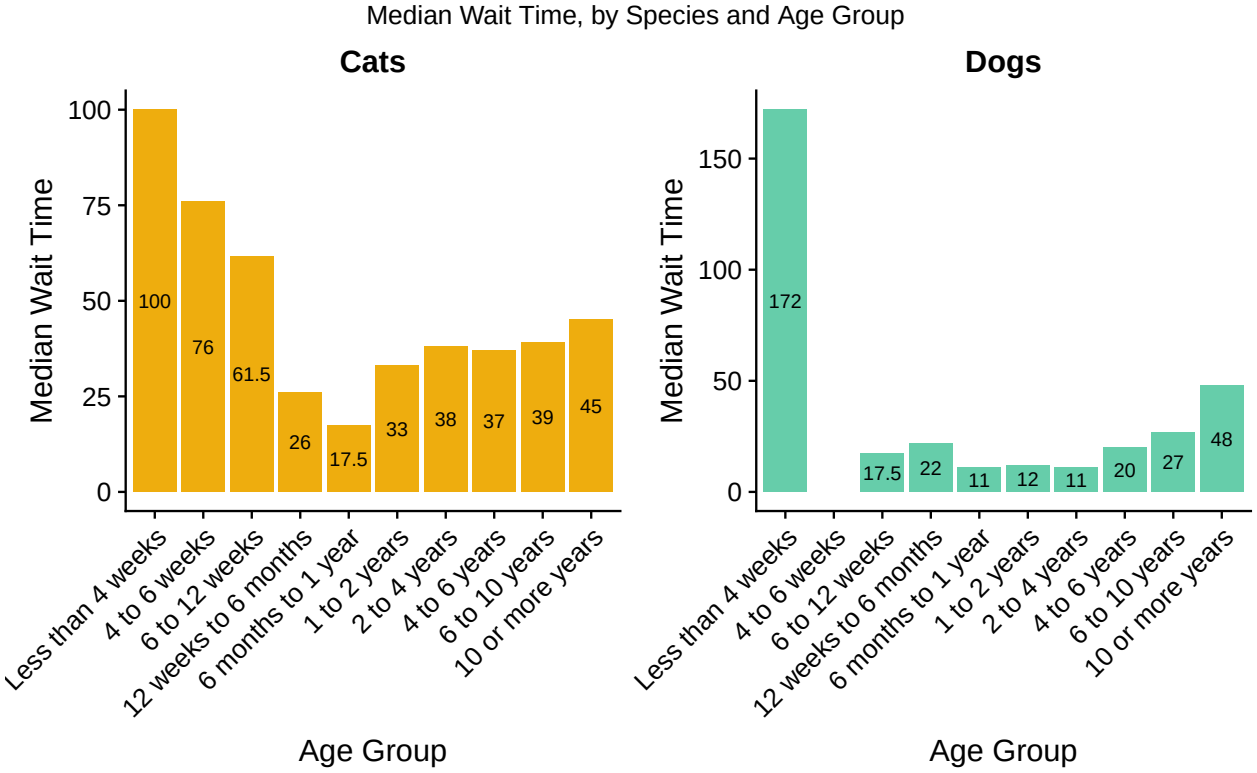
- **small = under 24 lbs**
- **medium = 24-44 lbs**
- **large = 44+ lbs**

While overall our analyses indicated no statistically significant differences in wait time based on dog size category, it is worth mentioning that medium sized dogs' wait time (11 days) trended towards being shorter than the wait time of small breeds (17 days) or large breeds (20 days).

Age Group

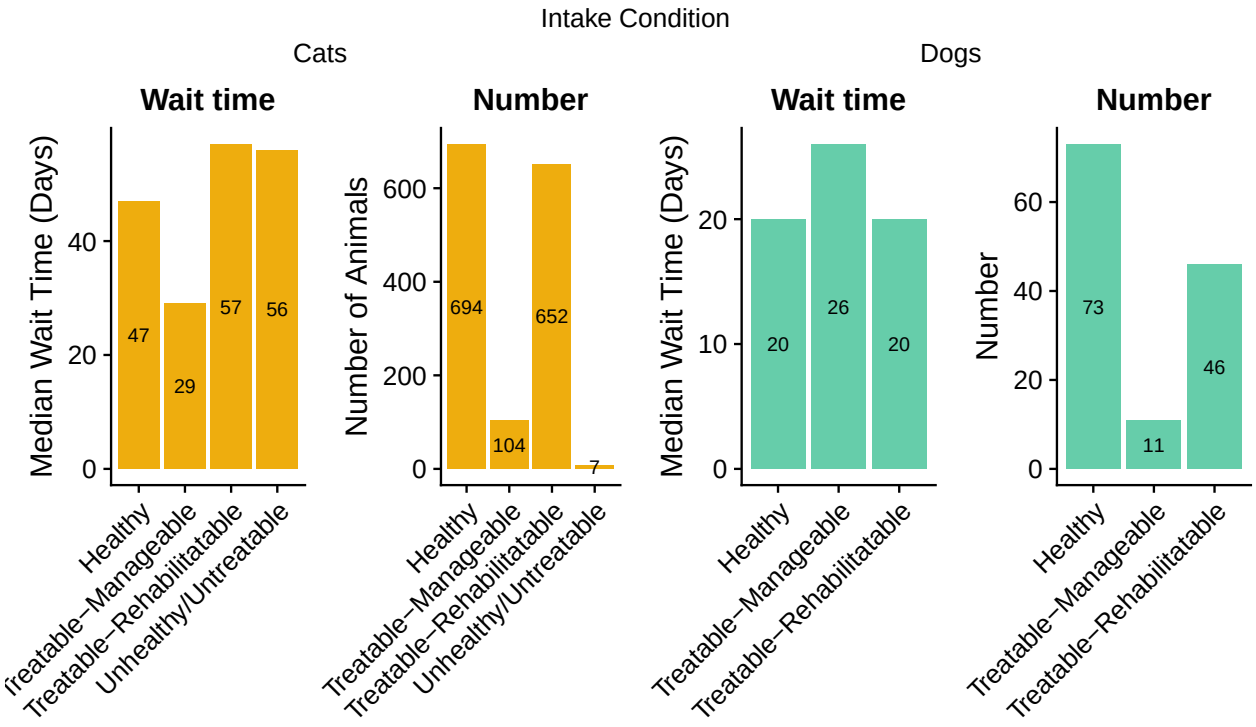
For cats and dogs alike, the longest wait times were in the 0-4 weeks age group. Understandably, these kittens and puppies require more time at PAWS before they can be adopted out. Interestingly, the longer wait time holds for cats up to 12 weeks, after which it decreases from roughly 65 days to 35 days. Cats aged between

12 weeks and 1 year spent the shortest time at PAWS. Dogs aged 6 months to 4 years spent the shortest time at PAWS, while the time at PAWS increased more sharply for dogs 10 or more years of age.



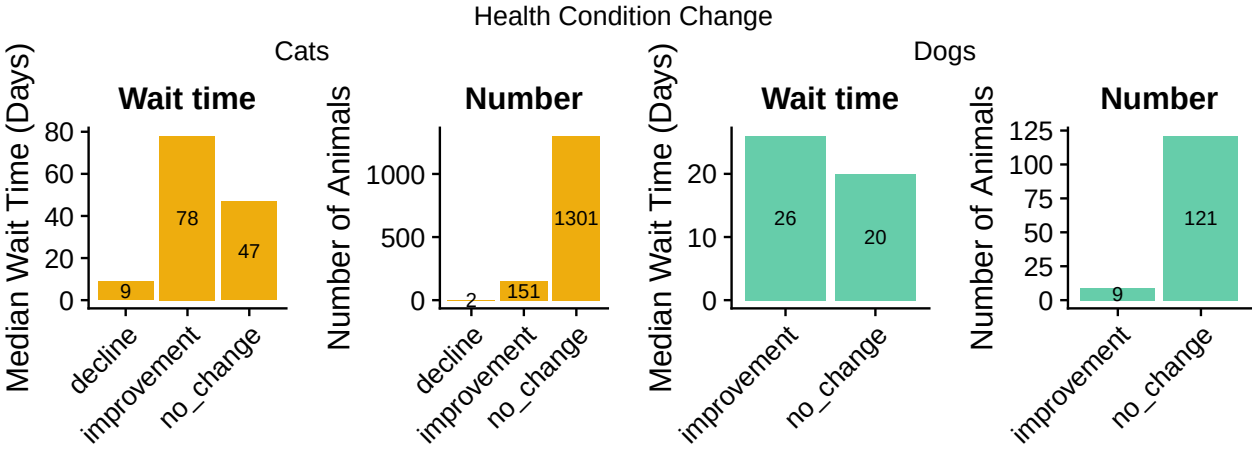
Health Status

We observed that for dogs, the longest wait times were associated with those who came in as "Treatable-Manageable" (however, these dogs represent only a small fraction of the dogs coming in at PAWS, so this effect is unlikely to impact overall wait times for dogs). For cats, the ones classified as 'Treatable-Rehabilitatable' and 'Unhealthy/Untreatable' had longer wait times than 'Healthy' or 'Treatable-Manageable' cats. While cats whose intake status was "Unhealthy/Untreatable" were very few, PAWS took in a very high number of cats labeled as "Treatable-Rehabilitate", and these were also the ones with the longest wait time - thus, reducing the wait time for this intake category could impact overall wait times for cats.



We also looked at health change from intake to outcome. Specifically, we created the following health change categories:

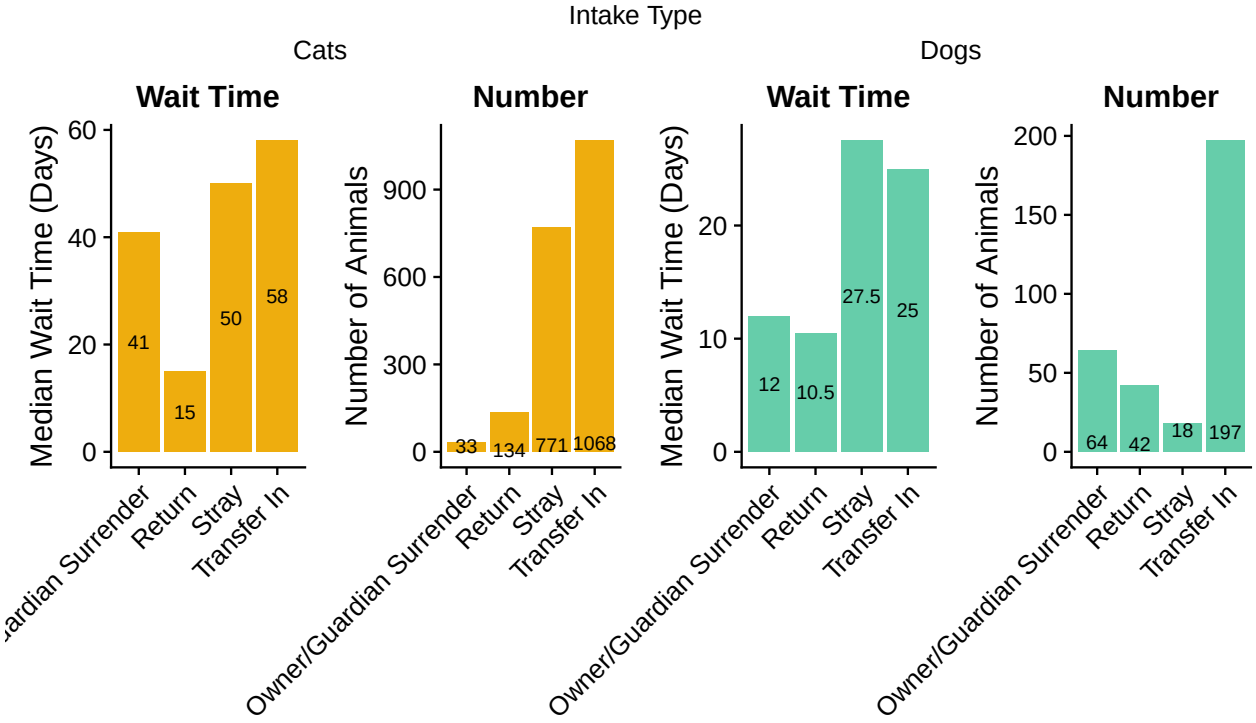
- no change = an animal's intake and outcome ASILOMAR status were the same
- decline = an animal's outcome ASILOMAR status changed from Healthy to Treatable or Unhealthy; or from Treatable to Unhealthy
- improvement = an animal's outcome ASILOMAR status changed from Treatable or Unhealthy to Healthy; or from Unhealthy to Treatable.



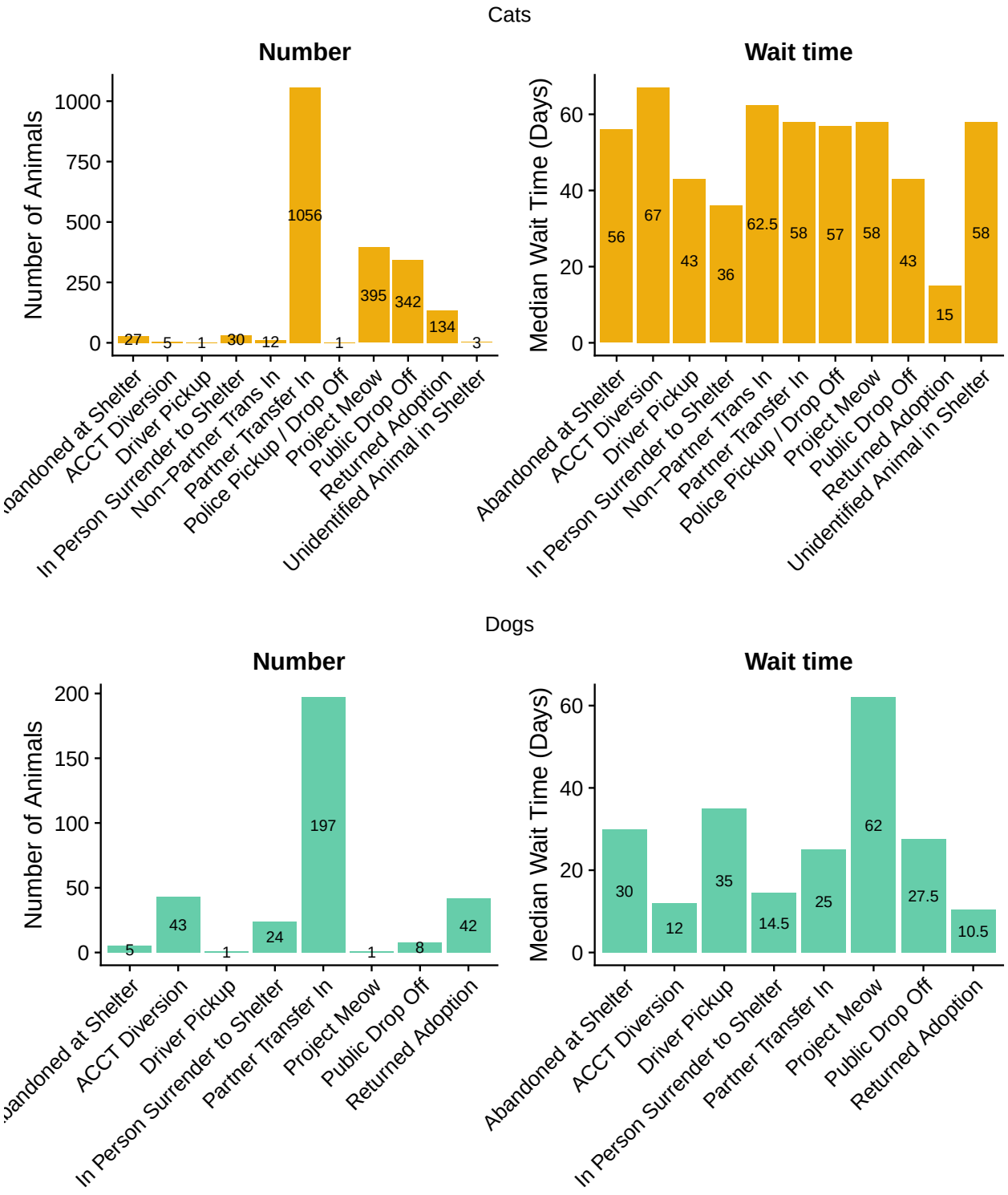
For both cats and dogs, the wait time was longest when they experienced an improvement in overall condition while at PAWS (78 days for cats, 26 days for dogs), although these animals represented only a small fraction of total animals (10% of cats, 7% of dogs). Cats that experienced no change in condition while at PAWS (89%) had a median wait time of 47 days, while dogs that experienced no change in condition while at PAWS (93%) had a median wait time of 20 days.

Intake Characteristics

The majority of animals arriving at PAWS were transfers-in (followed by strays, for cats, and owner surrenders, for dogs). Cats had longer wait times for each intake type relative to dogs. For both cats and dogs, strays and transfers in had the highest wait times.



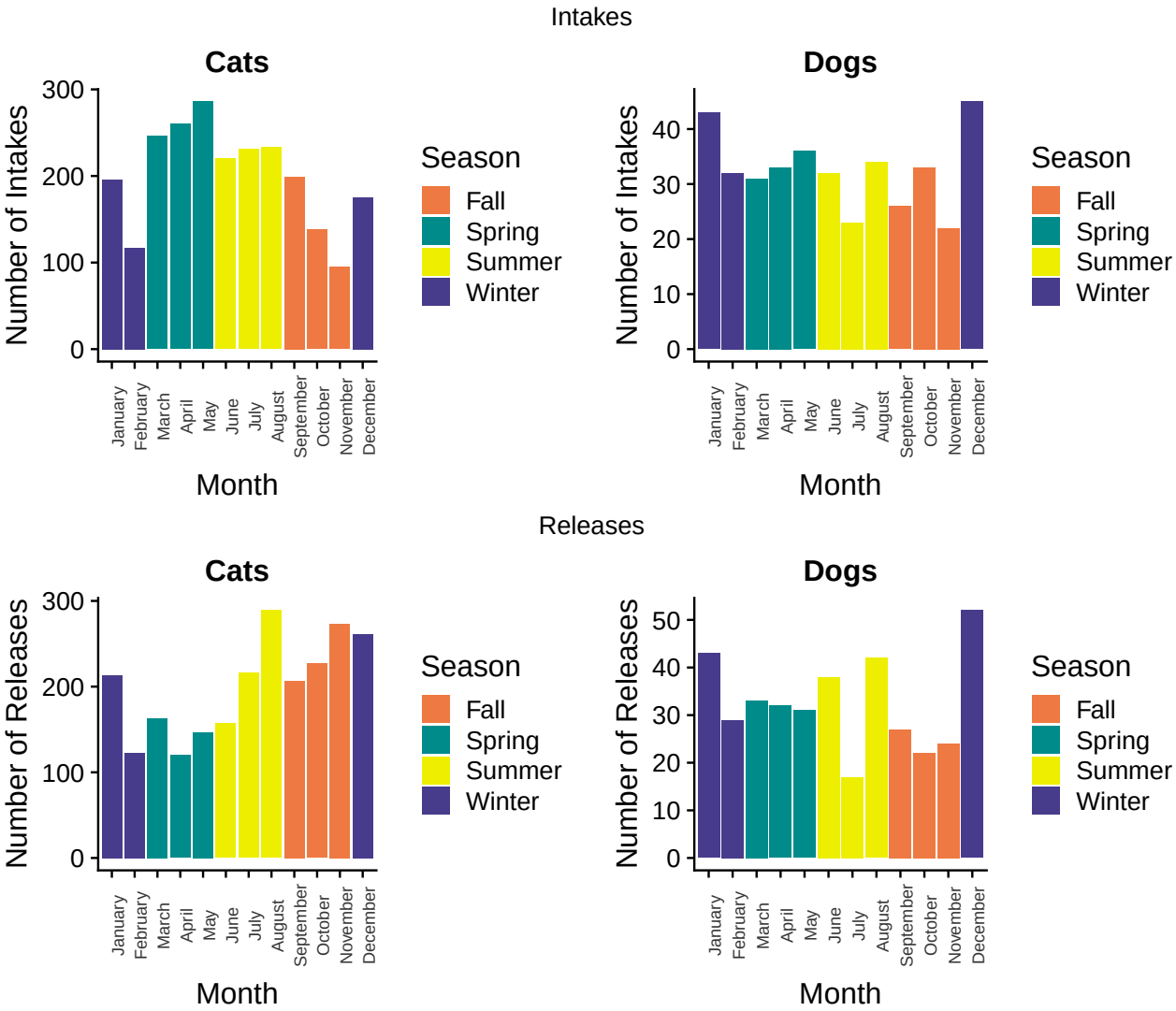
Partner transfers in accounted for the vast majority of intakes. However, differences were observed for the next highest intake subtypes: the next-highest numbers of cats arrived at PAWS via Project Meow (395 cats, or 19.7%), Public Drop-off (342 cats, or 17%)), and Returned Adoption (134 cats, or 7%). Meanwhile, the next-highest numbers of dogs arrived at PAWS as ACCT Diversions (43 dogs or 13%), Returned Adoptions (42 dogs or 13%), and In-Person Surrenders to Shelter (24 dogs or 7%).



It is also worth noting that for both dogs and cats, returned adoptions experienced the speediest re-adoption (15 days median wait for cats, and 10 days median wait for dogs). The next intake subtypes that were associated with shortest wait times for cats were In-Person Surrender to Shelter (36 days), while for dogs these were ACCT Diversion (12 days) and In Person Surrender to Shelter (14.5 days). (Note: we found one record of a dog whose intake subtype was "Project Meow", which may have been an input error - but we included it in the graph nonetheless.)

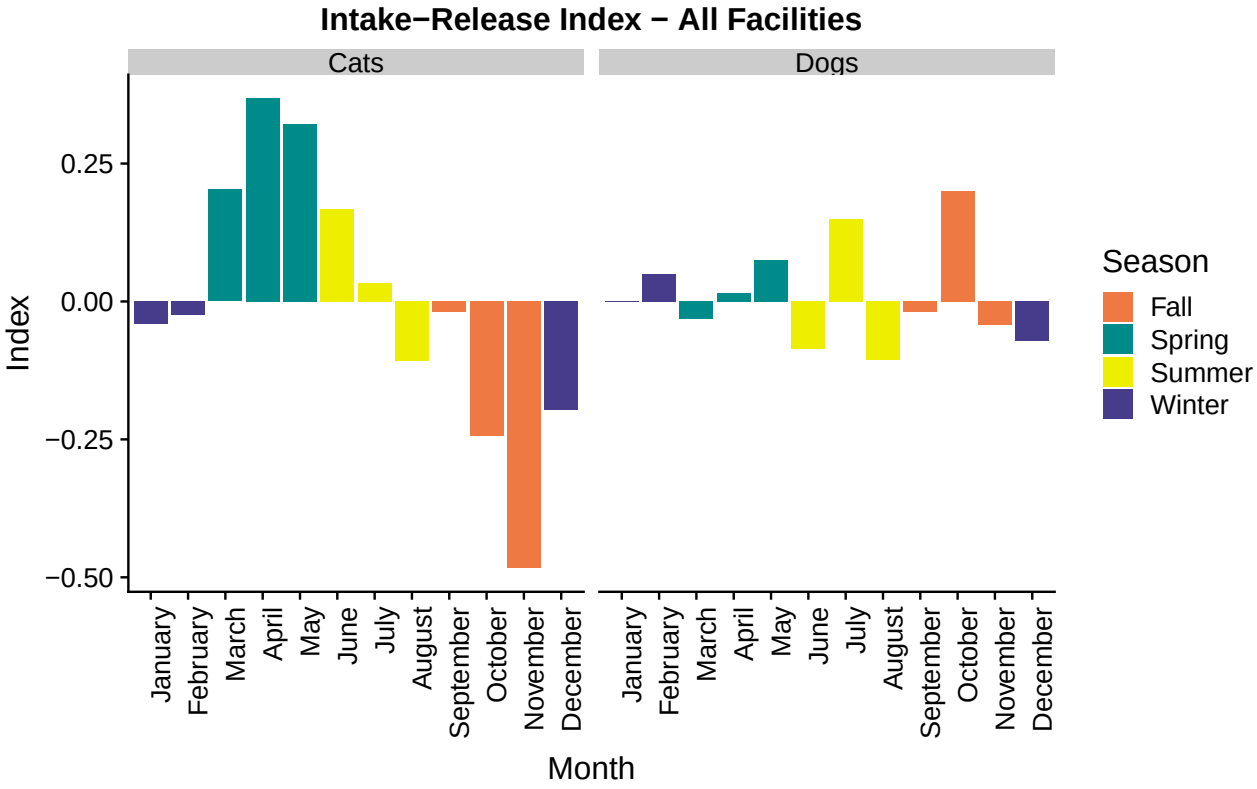
Seasonal and Locational Patterns

It appears that intakes are highest for cats in spring (peaking in May) and summer, and lowest in February and November. Meanwhile for dogs, the highest intakes are in winter (December and January), and the lowest in July and November. On the other hand, for cats, releases from PAWS are highest in August, November and December, while for dogs the highest numbers of releases occur in December, January, August and June.



We also looked at monthly trends in intakes and releases by species. Specifically, the graph below shows the difference between total number of intakes and releases by month, divided by the total number of intakes and releases per month. The resulting bars can be interpreted as follows:

- if >0, the number of intakes is greater than the number of releases
- if <0, the number of releases is greater than the number of intakes
- the closer to 0 a bar, the more evenly distributed intakes and releases were in that month



We found that cats have higher intakes than releases in March - June, while more releases than intakes were observed in October-December. For dogs, June and September represent spikes in intakes compared to releases. These findings may be useful for volunteer resourcing considerations during these times.

We also examined seasonal patterns in intake-release index by location for each species. It is important to note here that these patterns only take into account the location of intake and the location at time of release, and do not track an animal's transfers within PAWS locations. Therefore, the location-specific results should be interpreted with caution.

More cats were taken in than were released in:

- January, for Offsite Adoptions
- March, for Grays Ferry
- April, for Grays ferry and Grant Ave
- May, for Grant Ave, PAC and Foster program
- December, for Offsite Adoptions

It is interesting to observe that the clinic locations received more cats than they released (which could be an effect of cats being relocated) in spring, while offsite adoptions received more cats than they released in January and February.

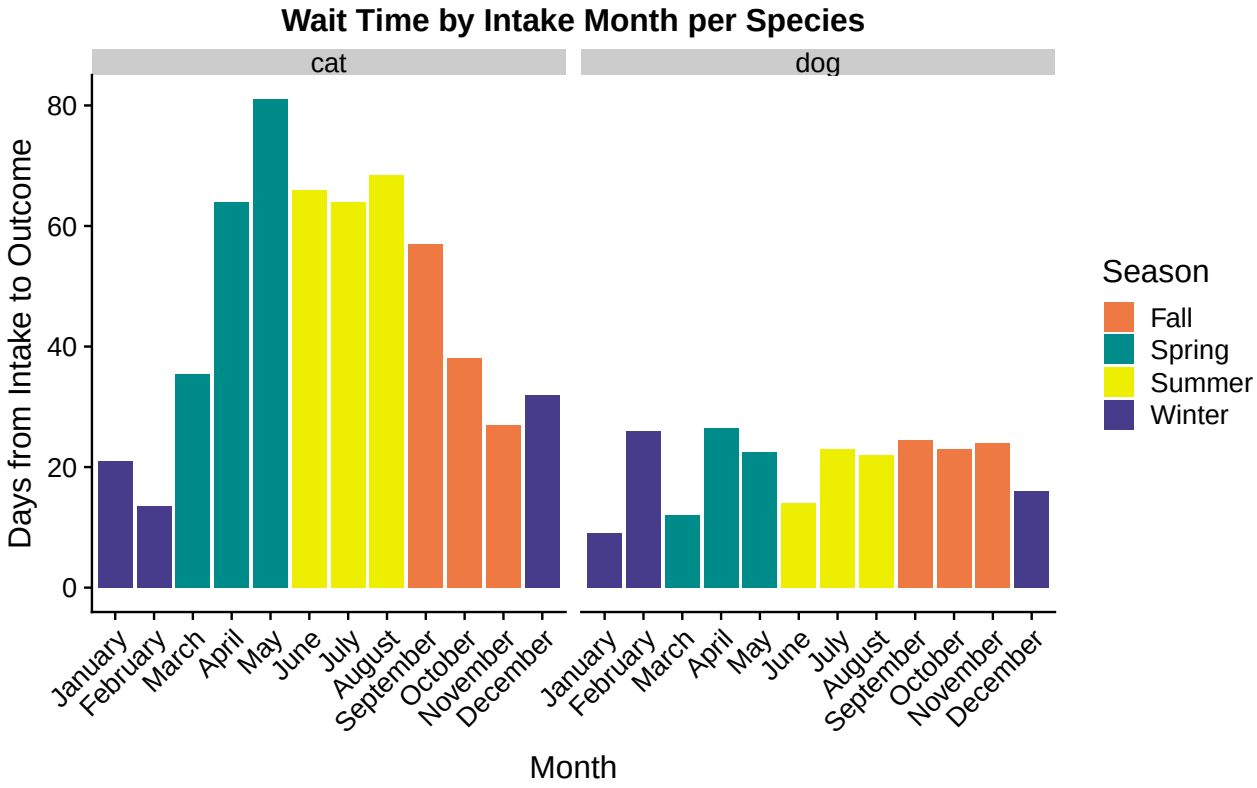
More dogs were taken in than were released in:

- January, for Grant Ave
- February, for Offsite Adoptions
- April and May, for Foster Program
- June, for Grant Ave
- July, for Grays Ferry
- October, for Grays Ferry
- November, for Foster Program

Here it is interesting to observe that the clinic locations received more dogs than they released in summer

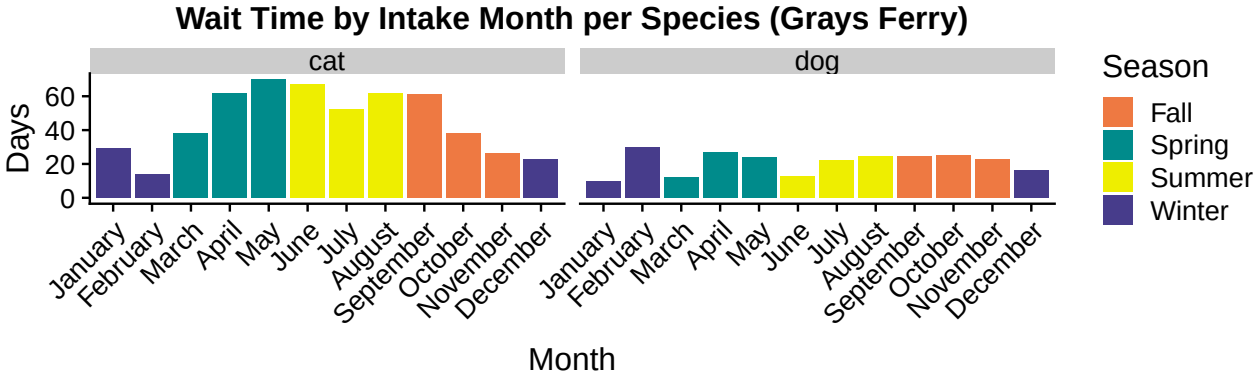
months.

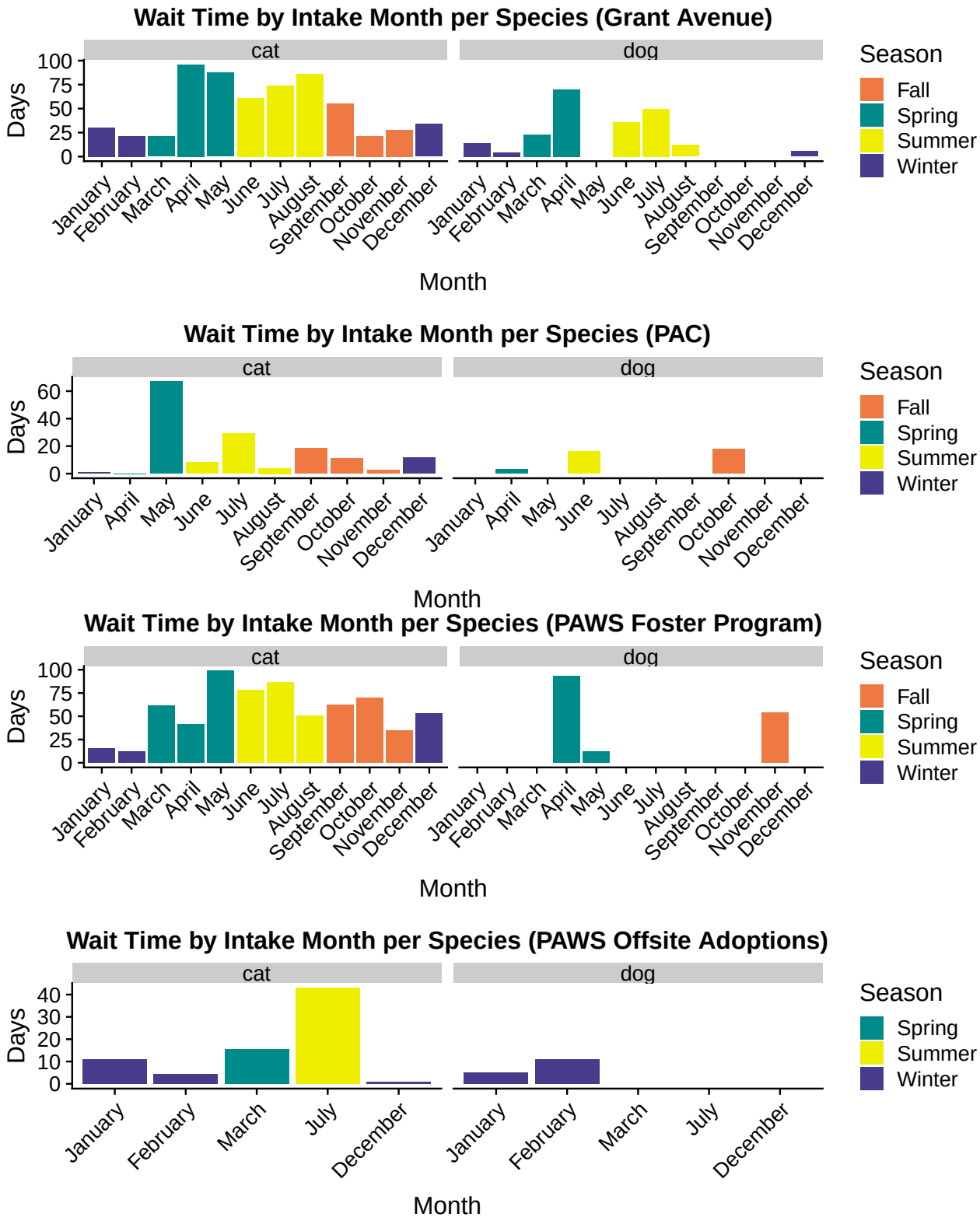
We furthern examined seasonal patterns in wait time, by species:



We observed that for cats that arrived at PAWS between April and September, the wait time was considerably longer than for cats that arrived in colder months. No such trend was seen for dogs. On a location specific basis, we noted the following patterns:

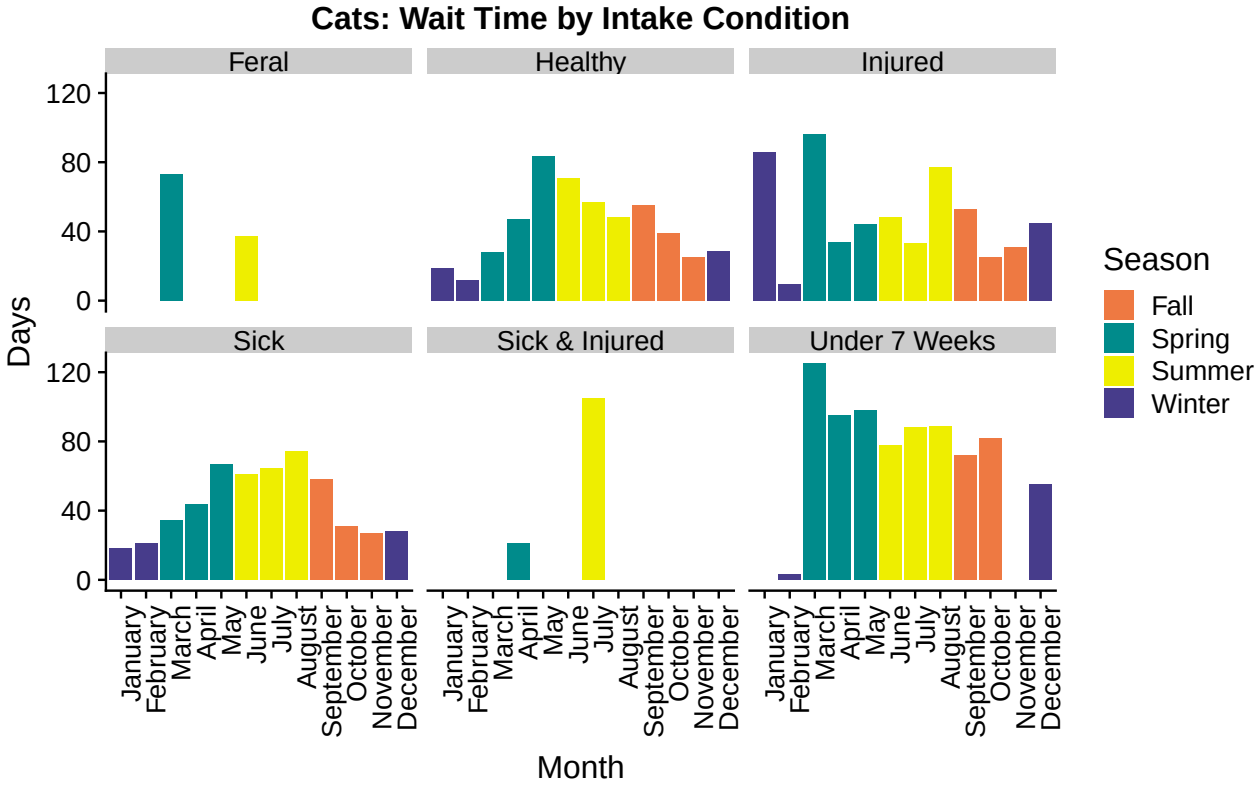
- Cats had longer wait times if their intake location was Grays Ferry and Grant Ave between April and September; PAC in May; Foster Program between March and December; and Offsite Adoptions in July
- Dogs had longer wait times if their intake location was Grant Ave in April, June or July; or Foster Program in April or November.

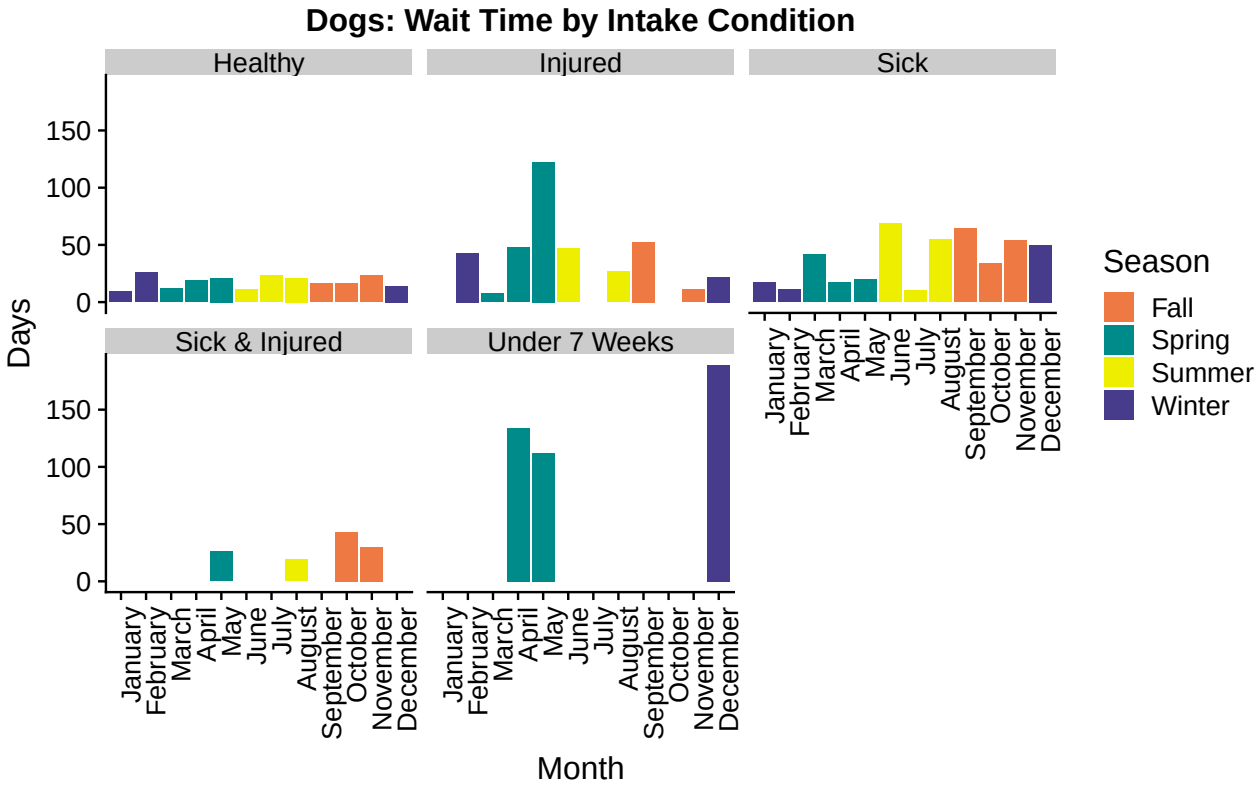




Cats adopted from PAC or from Offsite Adoptions, and dogs adopted from Offsite Adoptions spent the shortest time at PAWS, while cats and dogs in the Foster Program waited longest. Interestingly, for both cats and dogs, animals adopted from Grant Avenue spent longer in the shelter than animals adopted from any other PAWS location.

Wait times also seemed influenced by intake health condition, particularly in spring and summer months (cats arriving 'Sick', 'Injured', and 'Under 7 Weeks', and dogs arriving 'Injured' and 'Sick' tended to have longer wait times).





2. Application Trajectories

This analysis investigated factors relating to an application's trajectory at PAWS - from the time when an applicant submits an application, through the validation checks done by PAWS, to the time the application is marked as complete. The dataset analyzed here included trello IDs (1594 unique trello IDs) and their matched application data from the online application form (applications submitted dates between 2018-08-30 to 2018-12-31).

We found that for applications that resulted in adoption, cat applications took longer to process (19 days) than dog applications (8 days), with the slowest adoptions being at PAWS Foster Program & PAWS Offsite Adoptions locations. We also found that most adopters lived by themselves, or with one other adult in the home, and the majority did not have children in the home. Very few applications were denied in 2018.

Contributors

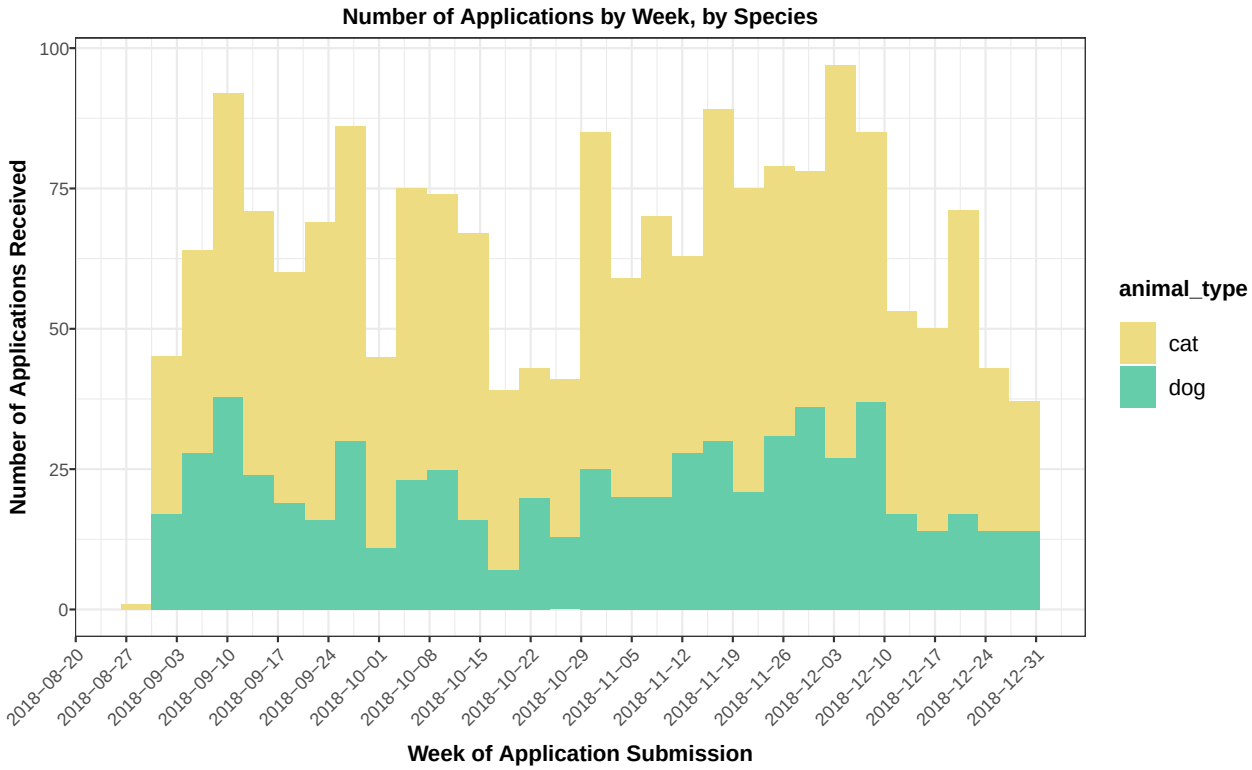
- **Ramaa Nathan** (group leader) is an aspiring data scientist with a PhD in Computer Science and an ongoing masters in Applied Statistics. Her background is in finance and healthcare.
- **Kate Connolly** is a digital analyst at the Philadelphia Inquirer where she helps to maintain the analytics framework and to provide data-driven support and decisions across the organization.
- **Veena Dali** is a senior business intelligence analyst at Comcast working to provide data solutions to support business decisions. Her background is in Neuroscience and Computer Science.
- **Amy Goodwin Davies** is a data scientist with a background in psycholinguistics.
- **Brendan Graham** is a clinical data analyst at The Children's Hospital of Philadelphia with a background in applied statistics.
- **Ambika Sowmyan** heads the Marketing data analytics group at Hartford Funds. Her background is in Finance and Retail and has a graduate degree in Management and Predictive Analytics.

Data Pre-Processing

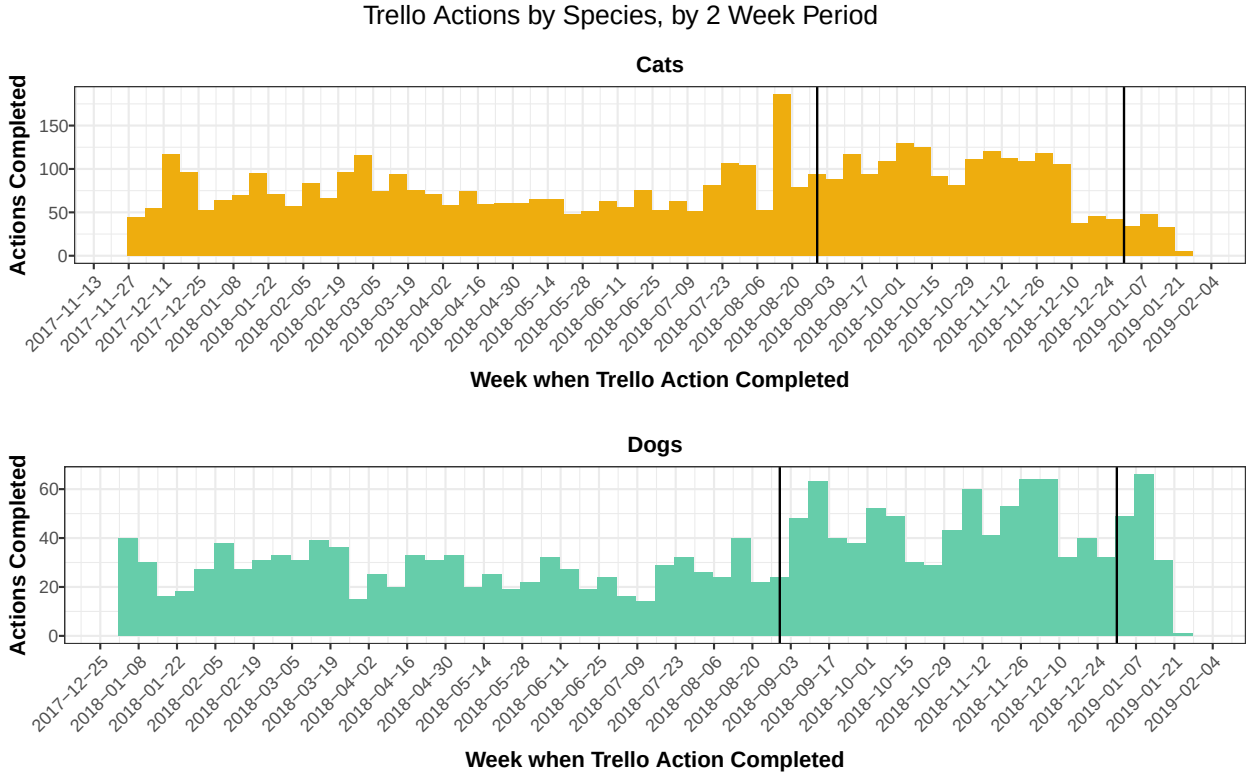
As this group focused on questions about application trajectories, the starting point was an applications dataset comprised of online dog and cat applications (`dog_apps.csv` and `cat_apps.csv`). Several data pre-processing steps are important to highlight:

- We standardized some responses that differed between cat and dog applications but that we believe had the same meaning (e.g. `ideal_adoption_timeline` had responses "next-few-weeks" and "few-weeks" which we standardised as one response ("few-weeks"). (See further discussion of this issue in the *Data Issues affecting Analyses* section)
- We replaced text values for `children_in_home` and `adults_in_home` to our best intepretation of what the responder meant (e.g. "-2" we interpreted as "2") and considered values that were excessive (e.g. 15 children in the home) as missing data.
- Due to a high variability of resonses on monthly and emergency budgets, we capped these responses at $10,000 (monthly) and $20,000 (emergency). Values higher than these we rounded down to the maximum.

Our cleaned applications dataset contained 1906 rows, 1594 unique Trello IDs, and the submitted dates ranged from 2018-08-30 to 2018-12-31.



To this applications dataset, we added fields from Trello actions and cards, and PetPoint outcome (`petpoint.csv`), and took similar data preprocessing steps as with the application dataset. The date range for the applications dataset (123 days) was considerable smaller than the actions and petpoint datasets (417 and 413 days respectively), which was a result of when the online applications were re-initialized in the PAWS website.



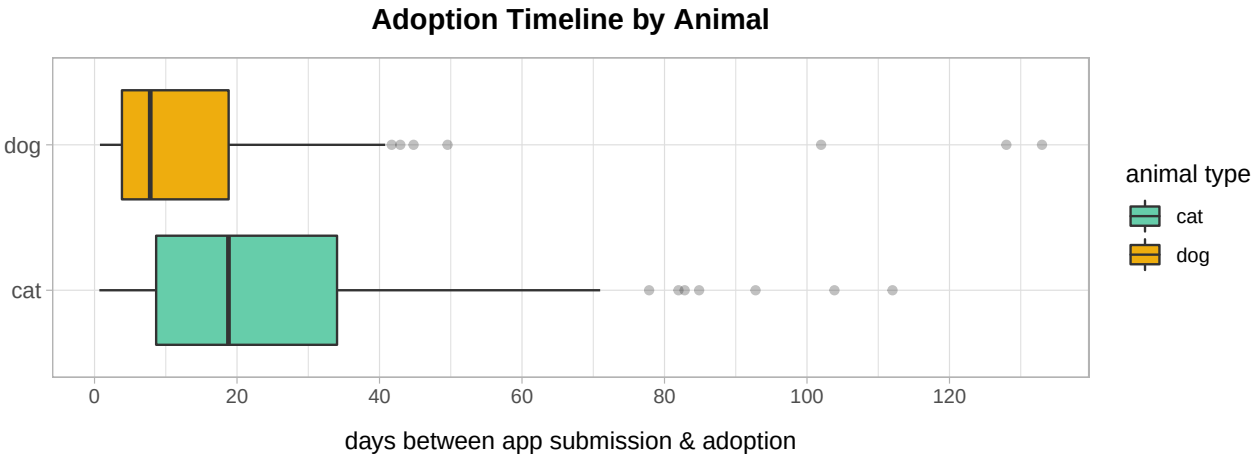
We noticed that for both cat and dog applications, there was an increase in Trello actions that correlated with the re-introduction of online applications at the end of August 2018. At the minimum, this indicates an increase in tracking/recording overhead that coincided with the re-introduction of online applications.

Time to Process Applications

Application Timelines

Application timelines were measured by taking the difference between the time an application was submitted and the time that application resulted in an adoption. Only applications that resulted in adoption were assessed; applications that were denied were not included in the analysis. This is a potential area of further investigation.

In general, cat applications typically take longer than dog applications. The chart below shows that the median adoption timeline for **cats** is approximately **19** days (vertical black line inside red box), while **dog** applications average about **8** to result in an adoption (vertical black line inside blue box).



The chart also illustrates that for longer-than-average application timelines, animal type may influence just *how much longer* those above-average timelines are. Of the longer-than-usual applications, cat ones took between 35 days and 70 days compared to about 18 days to 40 days for dogs.

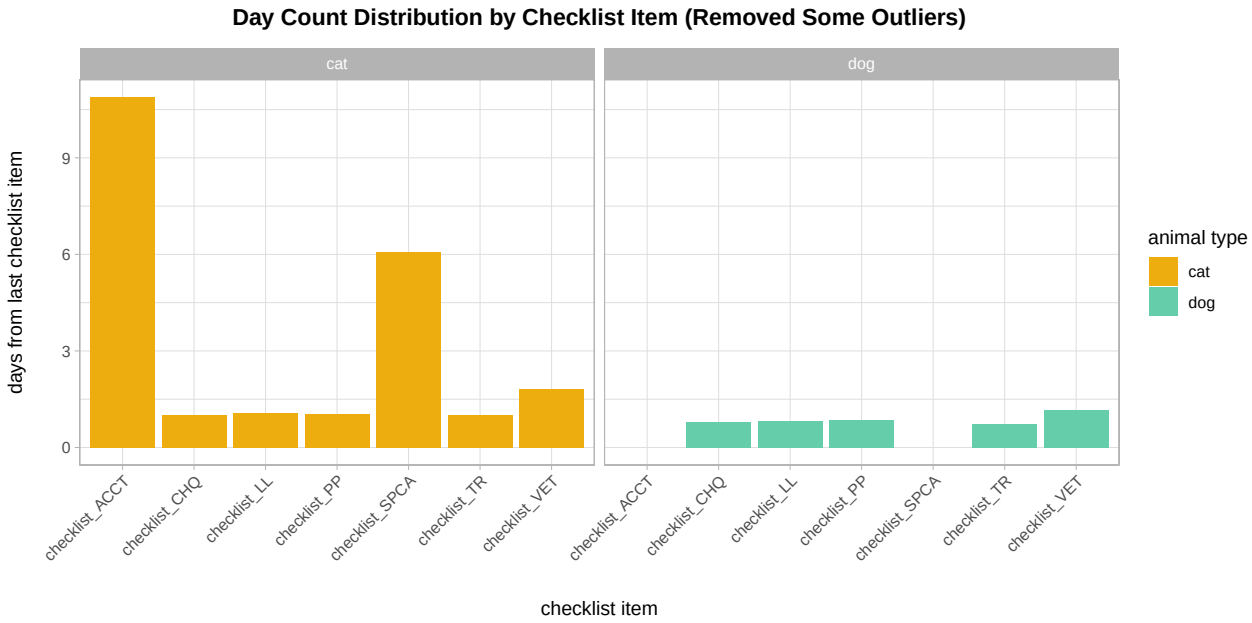
The outcome site for an adoption also influences the timeline of an application. It's important to note that this analysis does not consider all the potential locations that an animal spent its time during the application process; it is strictly based on the animal's outcome site.

outcome_sitename	animal_type	n	median adoption time (days)
Grant Avenue	cat	74	10
	dog	19	6
Grays Ferry Avenue	cat	2	8
	dog	18	13
PAC	cat	70	8
	dog	17	4
PAWS Foster Program	cat	187	25
	dog	20	12
PAWS Offsite Adoptions	cat	44	25
	dog	1	22

Based on median values, here are the fastest & slowest time-to-adoption sites:

- **Cats**
  - Slowest: PAWS Foster Program & Offsite Locations - 25 days
  - Fastest: PAC (Grays Ferry was faster but only had 2 cats in our dataset) - 8 days
- **Dogs**
  - Slowest: Grays Ferry Ave (PAWS Offsite Adoptions was slower but only had 1 dog in our dataset) - 12 days
  - Fastest: PAC - 4 days

Application Checklist Items



Most application items took between one and two days (median) to complete. While the animal type and outcome site didn't significantly impact the individual item times, cat applications generally exhibited slightly longer times between checklist items. Cat applications averaged about **1.2** days between checklist item, compared to **0.9** for dogs (excluding SPCA & ACCT items). The VET checklist item had the greatest difference between cats and dogs, and also was the item that took the longest (besides SPCA & ACCT items). This distinction between animals, while modest, could contribute to longer submission-to-adoption times for cat applications.

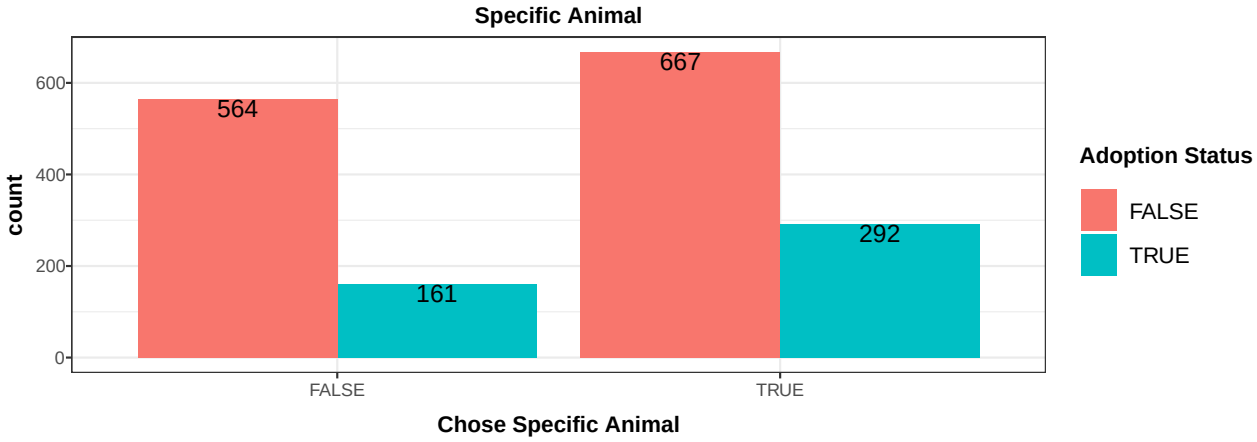
The chart above removed significant outliers, but further inspection of these outliers could be valuable. Understanding what causes certain application steps to take longer could help to streamline parts of the checklist process.

checklist item	n	median days from last item	percent of cards with item checked
checklist_ACCT	1	10.89	0.2%
checklist_SPCA	2	6.07	0.4%
checklist_VET	425	1.80	93.8%
checklist_CHQ	432	0.97	95.4%
checklist_LL	433	1.03	95.6%
checklist_PP	433	1.03	95.6%
checklist_TR	435	0.95	96.0%

The table above shows the exceptions to the average checklist times. The ACCT and SPCA checklist

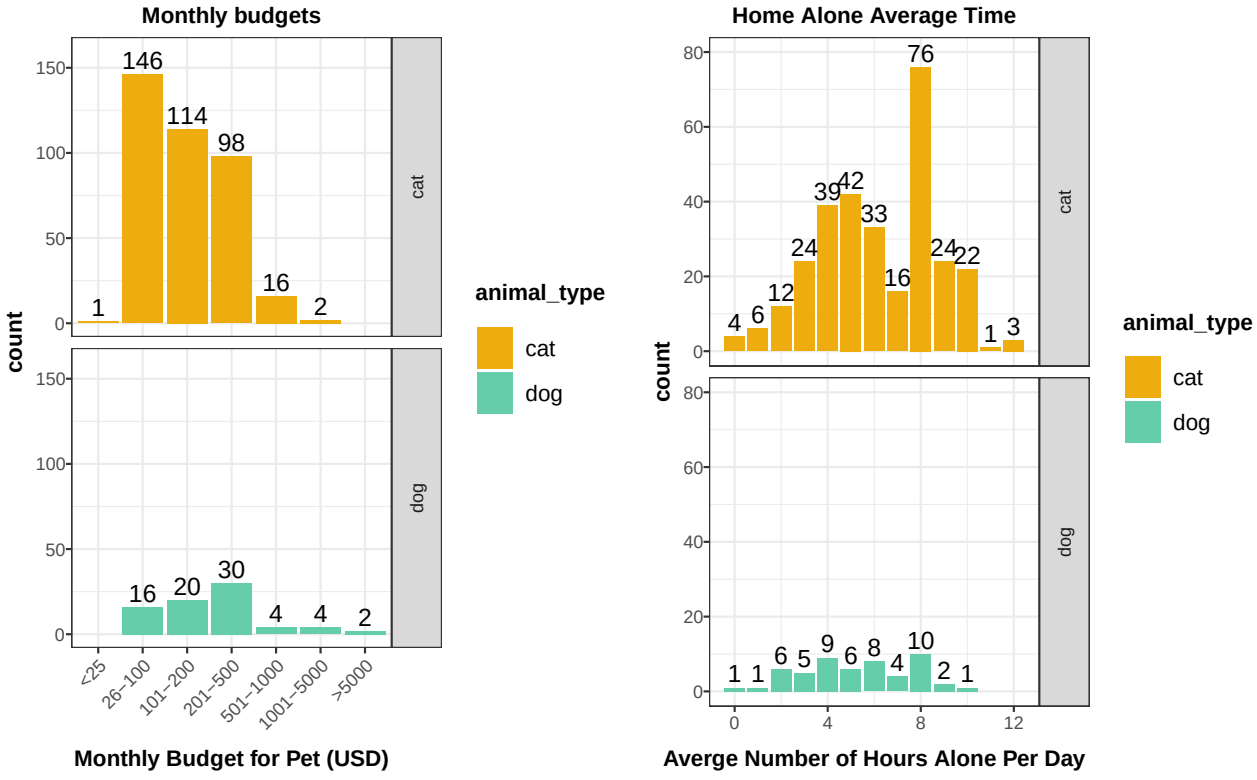
items took considerably longer to complete than other items, but they also were present in less than 1% of applications. This low sample limits any sound conclusions, but does present an area for potential further exploration. It may be valuable to assess if other components of an application—like red flags or particular animal information—lead to this item being more mandatory. But more data would be needed for this analysis.

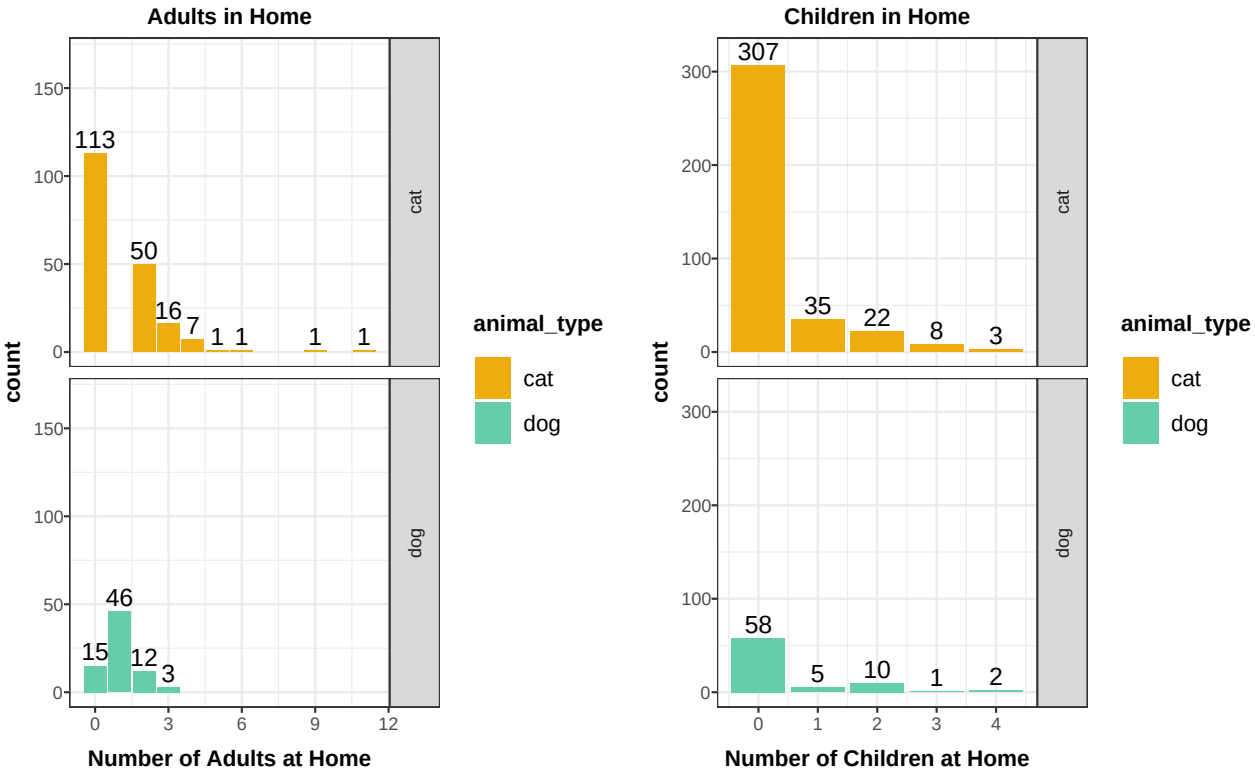
Successful Adoption Application Characteristics



When applicants requested a specific animal, 30% of applications resulted in an adoption vs. only 22% of the applications resulted in an adoption when the applicants did not request a specific animal.

Most of the applicants who adopted a pet had allocated a monthly budget of less than $500. Applicants who expected to leave the animal alone at home for longer hours chose to adopt a cat.

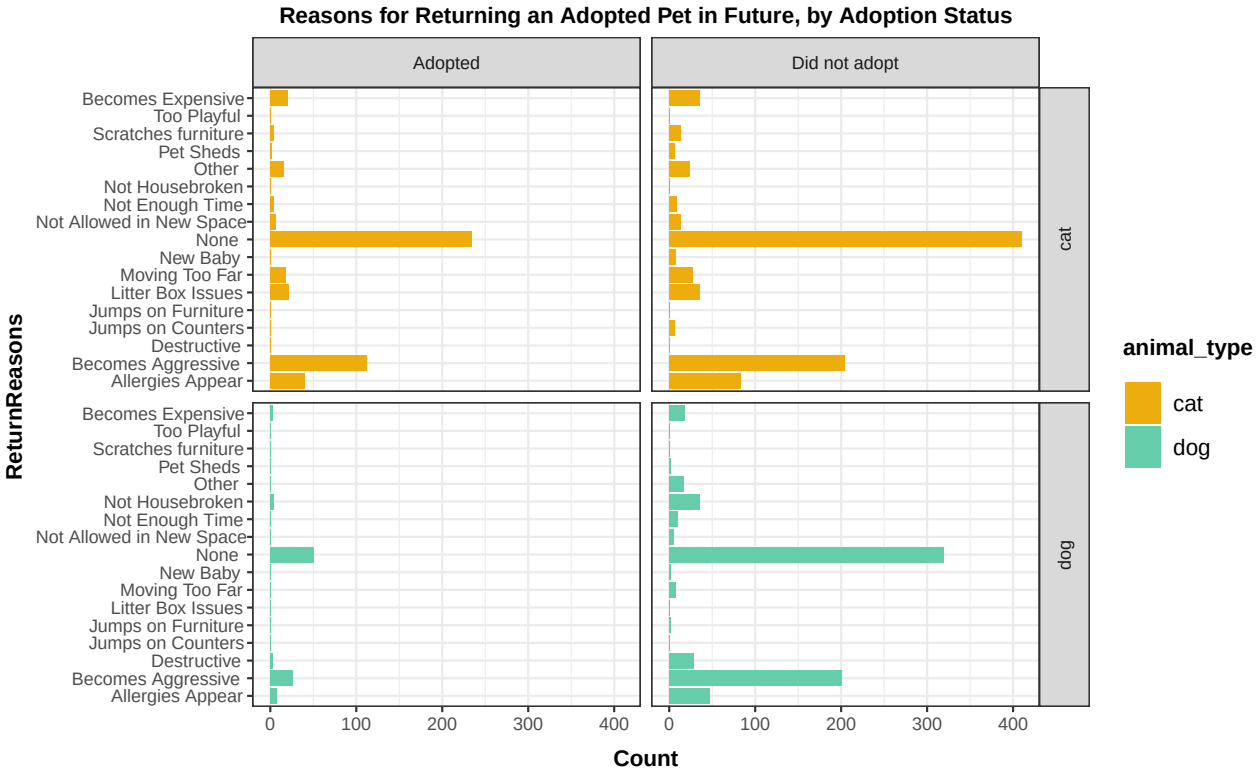




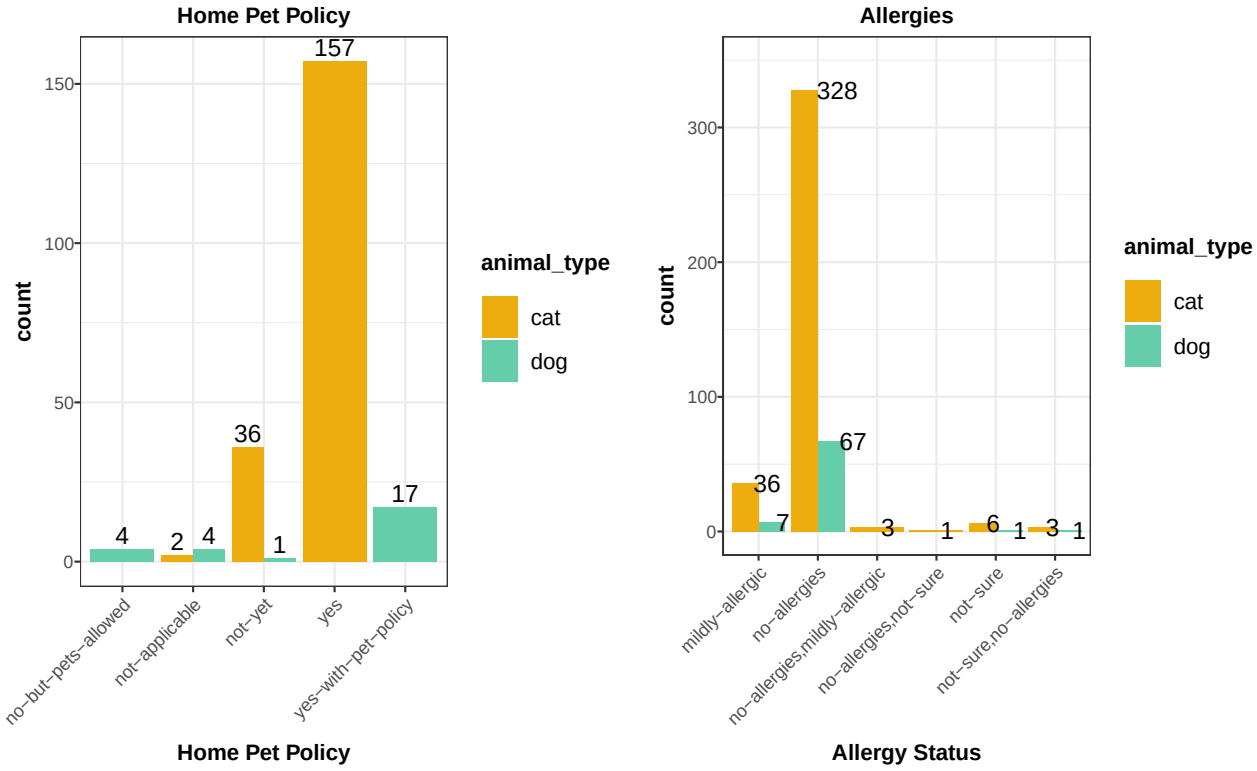
We found that the majority of applicants who lived alone preferred cats, while the the majority of applicants who lived with 1 other adult preferred a dog. Furthermore, individuals with no children at home seemed to make up the largest number of applicants.

When we looked at the type of experiences applicants had with animals, we noticed that a higher number of successful applicants had less experience with animals. Furthermore, the main reason that people would

return a pet in the future is if the pet becomes aggressive or if allergies appeared.



Unsurprisingly, the vast majority of people who adopted a pet from PAWS reported no allergies, and had a home policy that allowed pets.



Denied and Red Flagged Application Characteristics

We further investigated the characteristics of applications that were denied or red flagged. There were 12 applications that were denied, 19 that were withdrawn, and 133 that were red flagged.

Denied Applications We only have data for 12 denied applications so the analysis is limited. In the future when we have more data, we could compare the denied applications to the adopted ones.

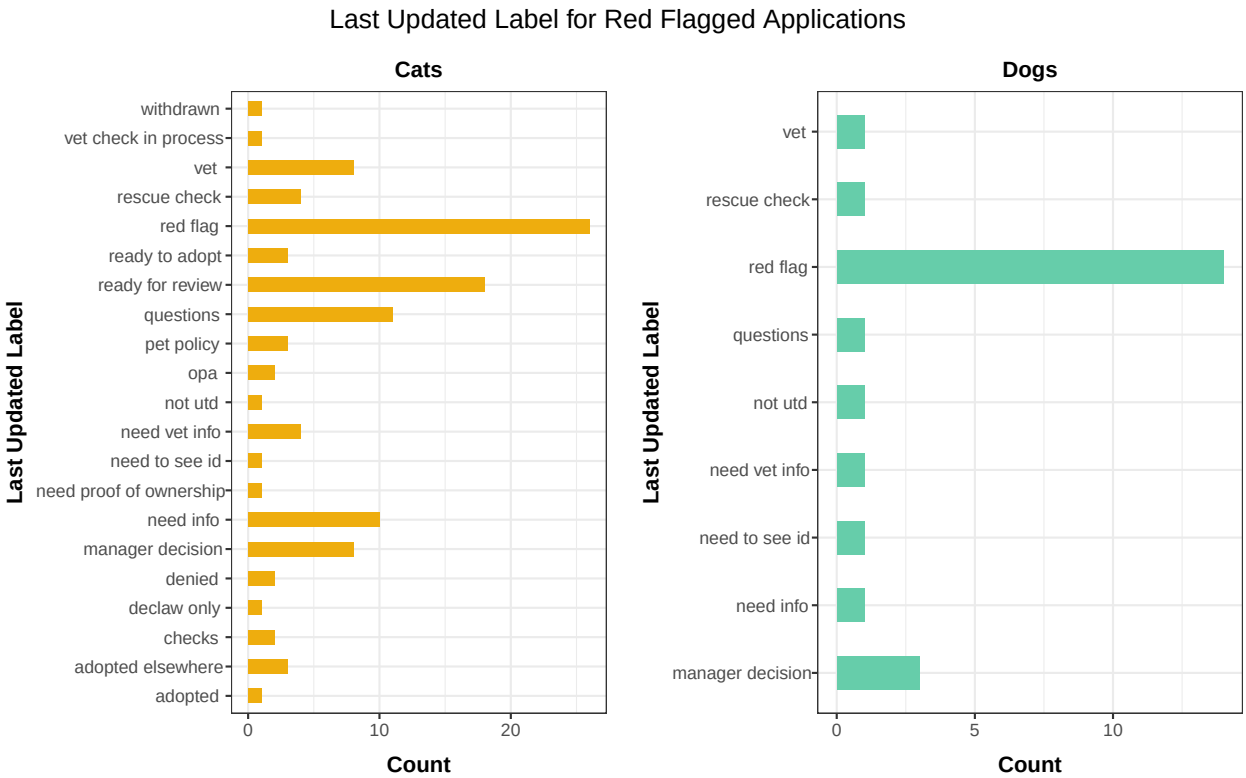
Key takeaways:

- No known allergies for the applicants
- Budget had no impact (same budget range for approved applications)
- All household members agreed to get a pet
- Majority of the applicants did not enter a home pet policy and not everyone is the home owner
- Many applicants had unfortunate incidents with prior pets (e.g. ran away, died in care)

Red Flagged Applications

There were 133 applications that were red flagged.129 of the 133 have not yet resulted in an adoption or are still being procesed. Two of the applications that were flagged were denied but that does not mean that the rest are going to result in adoption. Since the data set for the applications is from the end of 2018, many of the applications are still in progress. We do not have the final status of all the applications so we cannot conclude what happened to the red flagged applications. As a further project, I think it would be interesting to track the final status of the applications that were red flagged.

Below is a visualization that shows the last updated status for applications that were red flagged. After being flagged, the applications were sent to the manager to make a decision or the applicant was requested to provide more information (e.g. in many cases the applicant was required to provide more information about the vet).



Important Features for Prediction

After separately analyzing the different characteristics that affect whether an application results in an adoption, we wanted to further understand how the different features in the dataset could have had a combined effect on the adoption status. We therefore ran a basic Random Forests model on the dataset. A Random Forest is basically a tree-based algorithm where a random subset of predictors (or features) are evaluated at each node and the observed data is split into two regions using one of the predictors and a threshold value for that predictor such that the error in predicting the adoption status is minimized. Starting from the top of the tree with one node, two new nodes are created with each split and the tree is grown recursively till there are only a few observations in each leaf node. Multiple trees are built similarly and the results are combined together to predict the adoption status for any given set of characteristics.

The combined effect of different characteristics on the adoption status can be studied by considering one of the important outputs generated by the Random Forests, the subset of predictor values that are found to be most commonly used as a criteria for splitting the dataset into two smaller regions at each node. We found that the top three characteristics that influenced successful adoptions were the number of children in a home, applying for a dog, and the date the application was submitted.

In future analyses, these variables can be highlighted to better understand adoption patterns.

3. Geographic factors

These analyses examined the data in relation to geographic and population parameters. We sought to understand how the socioeconomic environment of applicants might affect the likelihood of adoption, and whether there were factors contributing to regional differences in application processing speed.

Contributors

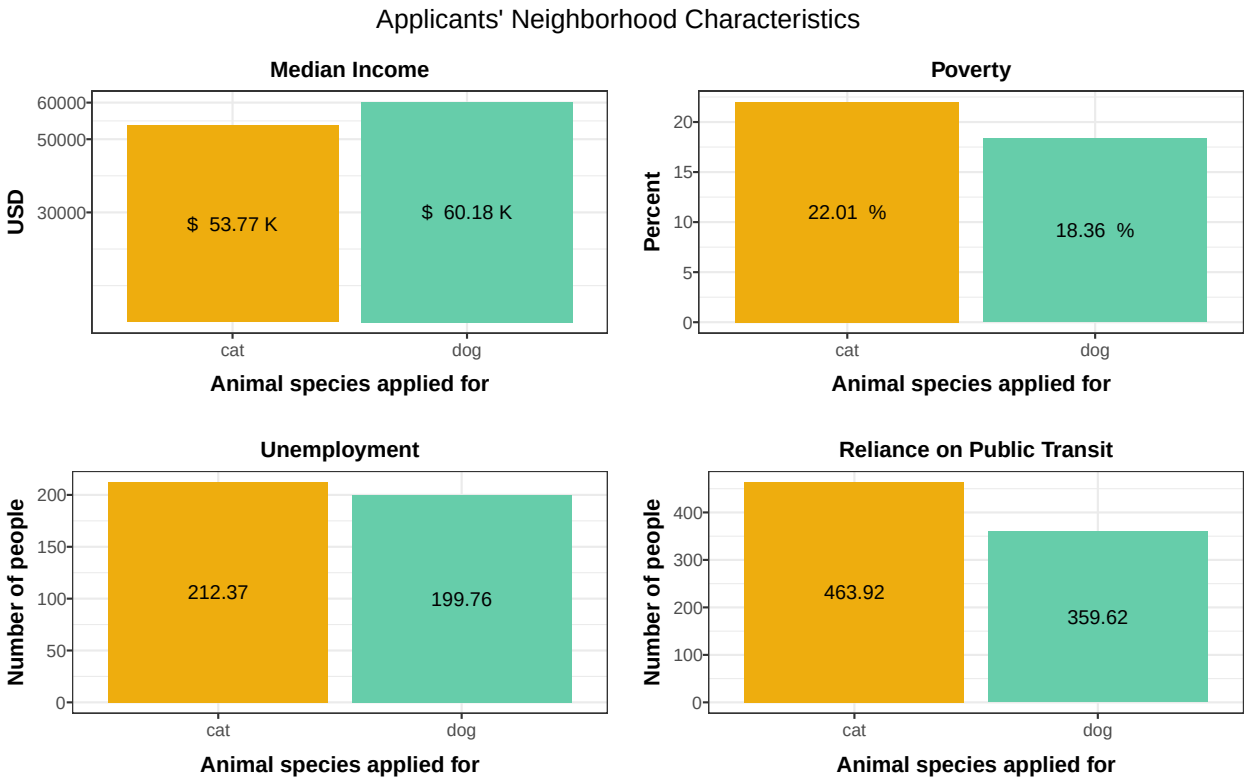
Joy Payton, MS is the Supervisor of Data Education in the Dept. of Biomedical and Health Informatics at the Children's Hospital of Philadelphia. She leads the development and implementation of education and outreach programs.

Karla Fettich, PhD is Head of Algorithm Development at Orchestrall, Inc. She develops data analytics solutions, predictive models and optimization approaches to improve operations and outcomes in long term care facilities.

Data Processing

We extracted addresses and associated them with their respective census tracts (this was handled by a PAWS volunteer prior to de-identified data being made available to the larger group). Census tracts are areas roughly equivalent to a neighborhood established by the Bureau of Census for analyzing populations. We used census data from the 2017 five-year American Community Survey via the American Fact Finder, and included economic characteristics, education characteristics, median rent, and computer and networking characteristics.

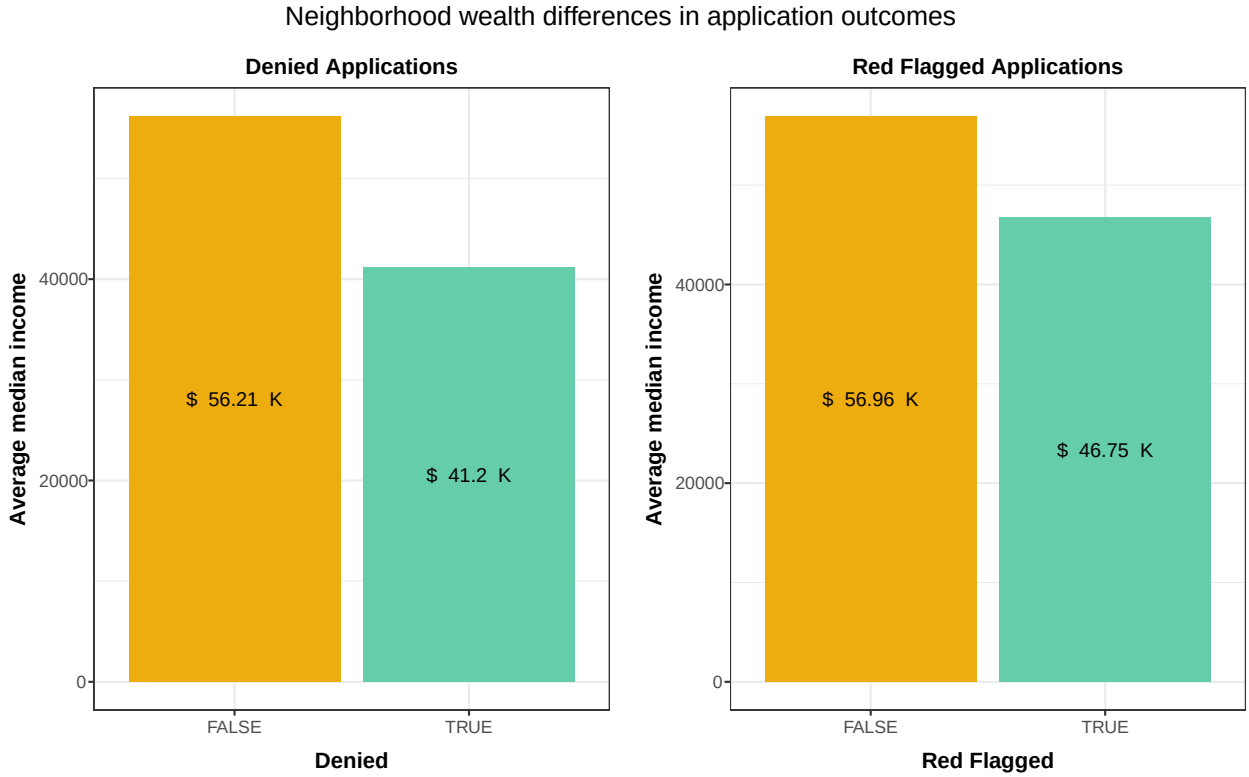
Economic Considerations in Processing Applications



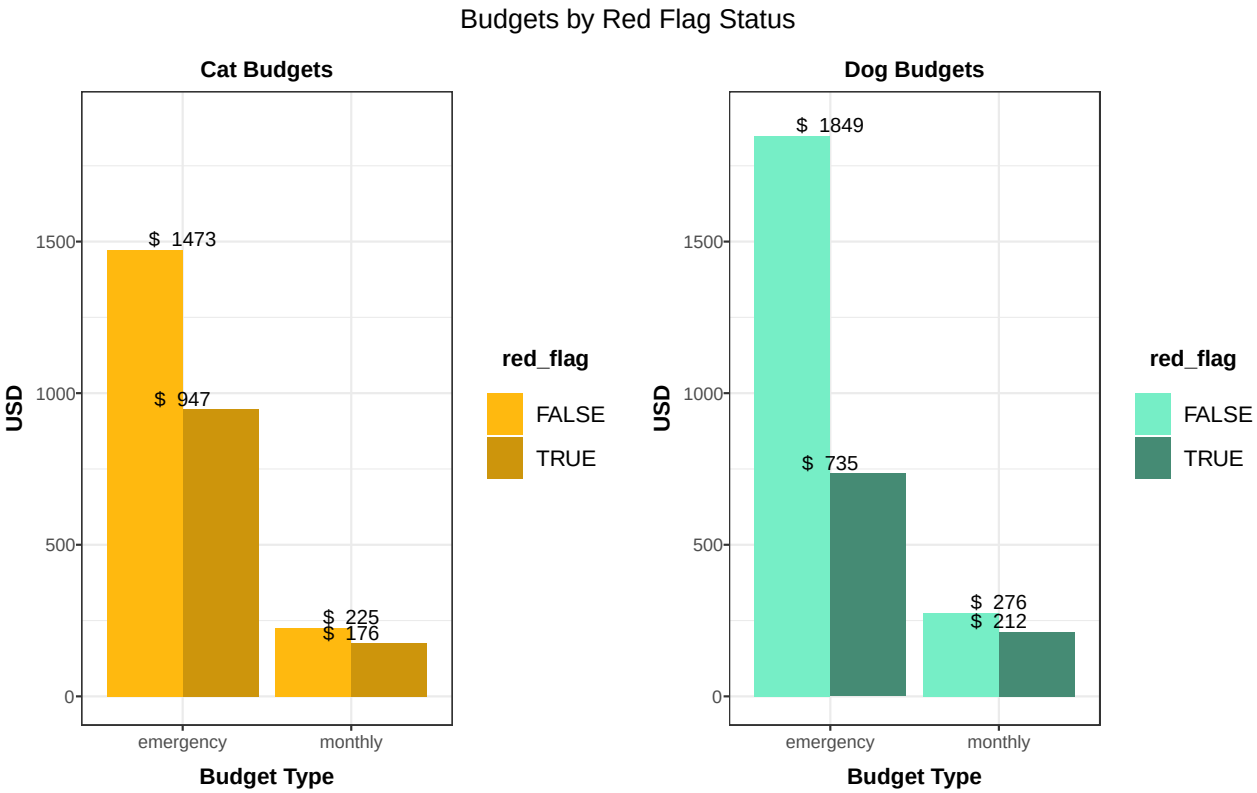
On average, we observed that dog applicants live in areas where the median income is higher compared to cat applicants (around $60,000/year for dog applicants vs. $54,000/year for cat applicants) and where the percent of households living under the poverty level is lower (18% for dog applicants vs. 22% for cat applicants). This suggests that dog applicants are from slightly wealthier neighborhoods. We further observed that dog applicants have more range between lower middle class and upper middle class, while cat applicants tend to skew more toward lower incomes. We also examined whether unemployment and reliance on public transit might play a role in applicants' preferences for dogs vs. cats. We observed no differences in mean neighborhood estimates of number of unemployed people of dog vs. cat applicants, but we did notice that in neighborhoods where cat applicants live, more people tend to rely on public transit for their commute to work.

This finding aligns with the previous group's findings that a greater proportion of cat applicants had a lower monthly budget for pet care, compared to dog applicants. This is also in line with the pet care cost estimates provided by the ASPCA which suggest that the first year total costs of owning a dog ($1,471 - $1,779) exceed those of owning a cat ($1,174).

While we observed neighborhood wealth differences in animal species preference by applicants, we did not observe this factor to be biasing PAWS when processing applications (for this comparison, we used the "complete" status of a trello card at the time when the data were pulled - although this is fairly vague indicator because it does not describe the outcome of an application beyond having been processed).



We further looked into some of the outcomes of application processing, specifically *red flags* and *denied* applications. Applications from neighborhoods with a lower household median income (under $50,000/year) are more likely to be red flagged and denied, compared to those with a higher household median income (over $50,000/year).



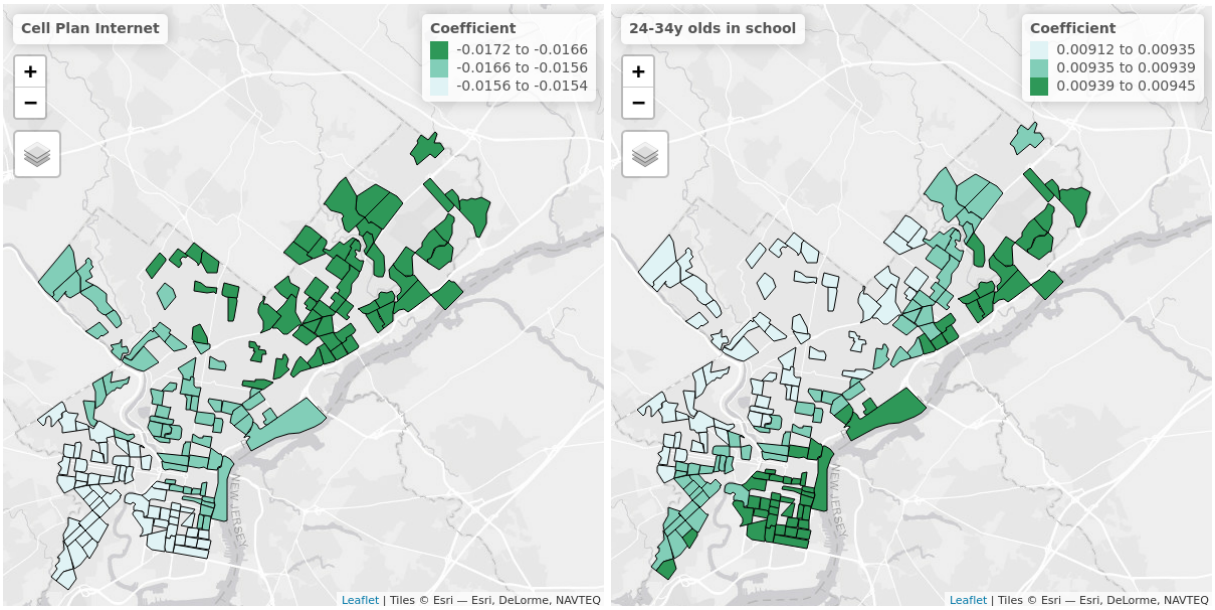
Additionally, red flagged **cat** applicants have a lower estimated monthly budget than their non-red-flagged counterparts ($176 vs. $225). For **dogs**, a similar trend was observed, but it did not reach the statistical significance threshold ($212 vs. $276). This pattern also holds when it comes to emergency budgets: red flagged applicants have a lower estimated emergency budget than their non-red-flagged counterparts ($947 vs. $1,473 for **cats** and $735 vs. $1,849 for **dogs**).

While we found that living in a lower income neighborhood does impact the estimated emergency budget at a statistically significant level, it only accounts for about 7% of the observed pattern. This indicates that there are additional factors that may play a role in how much money an applicant is able to set aside on a regular basis for pet care.

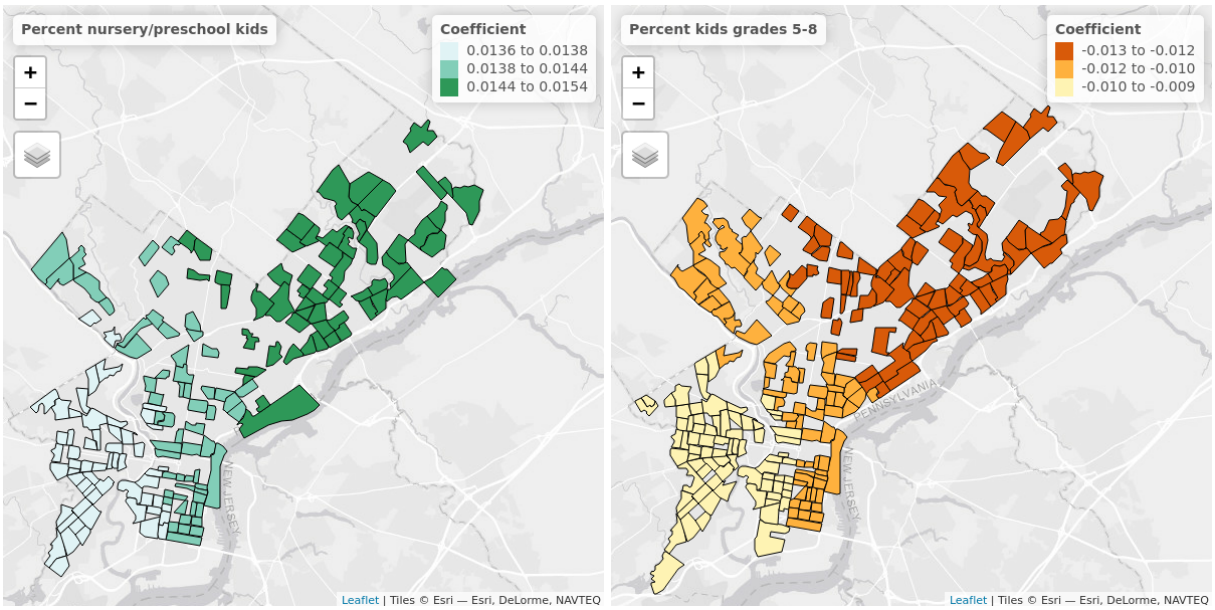
Efficiency Analysis in Philadelphia County

We also looked at applications that were processed within an efficient timeframe (defined here as 10 days), vs. those that did not. An application was considered efficient if it was given a decision label ("denied", "do not follow up", "adopted", "adoption follow up", "approved", "ready to adopt", "ready for review", "reviewed with handouts only", "approved with limitation", "dog meet", "returned", "adopted elsewhere") and the last trello checklist item was checked off 10 days or less from the date of application submission.

We found that in neighborhoods with a higher percentage of people who have a cell data plan and no other type of internet subscription, there was a trend for a lower proportion of efficient **dog** applications, this effect being more pronounced in north and northeast Philadephia. Some reasons could be that these applicants may not be as familiar with filling out online applications (which represents the application dataset we analyzed), or they may have more difficulty finding the information they need, or may be filling out the application form on a mobile device when it was not designed this way.



Additionally, in neighborhoods with a higher percentage of the population 25 to 34 year old enrolled in school, we observed a significantly higher proportion of efficient **dog** applications. We speculate that this may be related to applicants' level of comfort with online applications, access to information, or other factors that are more specific to the life circumstances of individuals enrolled in school. Finally, we observed a trend that **dog** applications were processed more efficiently if applicants lived in neighborhoods with a higher proportion of children enrolled in nursery or preschool, and that this effect was stronger in the northeast.



For **cats** we found that in neighborhoods where there were more middle school children, the proportion of efficient applications was lower, this effect being more pronounced in the north and northeast. While we do not know the reasons for this effect, it may be worth noting that ownership of and interest in pets tend to peak in middle childhood (i.e., 8–12 years). It may be that this effect influences the decision to submit an application, but that other barriers interfere with the application's timely processing (e.g. incomplete information, lack of responsiveness to provide additional information, change of mind).

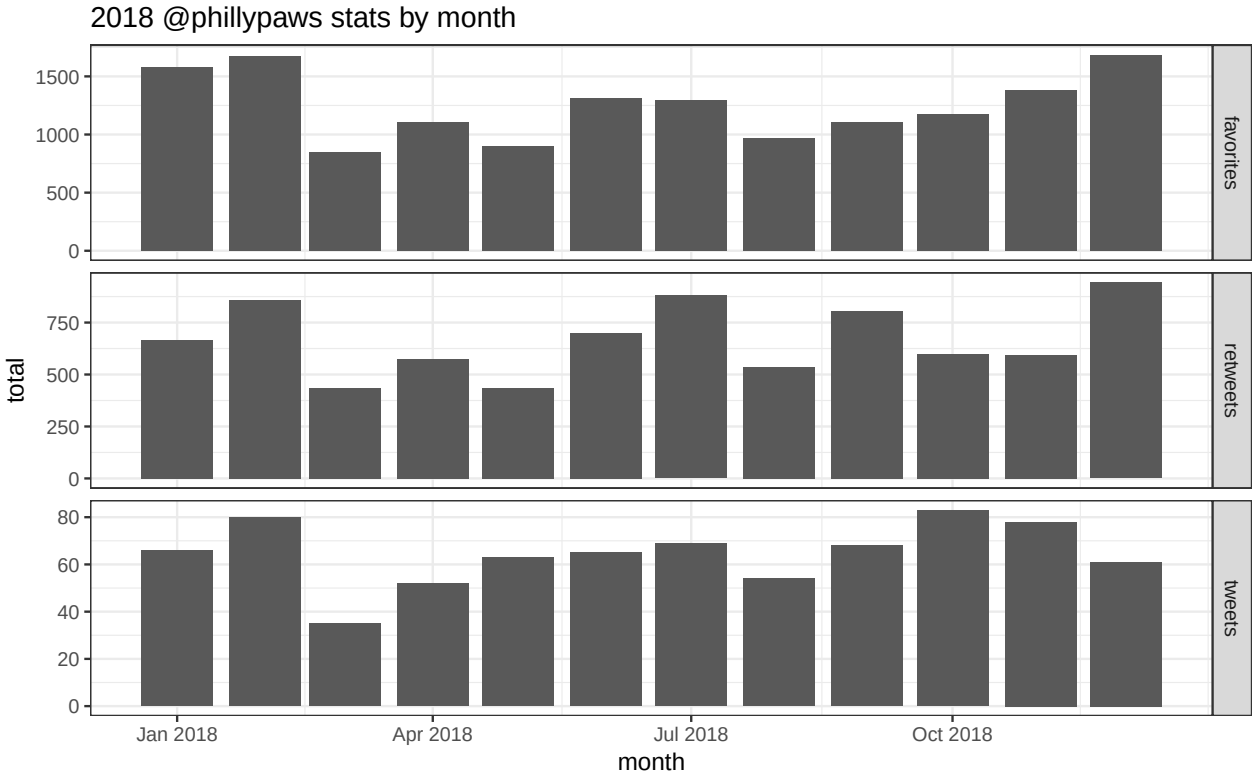
4. Social media factors

In this analysis, we analyzed @phillypaws tweets to understand any patterns in twitter activity and whether that could be linked to application or pet information. We examined the data from the PAWS twitter account, @phillypaws. We used the twitter API to download the most recent 3200 tweets, which included all tweets from 2018. Quotes and retweets were not excluded from the dataset. We did not see a strong association between twitter activity and applications, and we identified the most commonly tweeted words to be "home", "adoption", and "meet". We found that tweets with photos were more often favorited and retweeted.

Contributors

Alice M Walsh, PhD is a computational biologist in the pharmaceutical industry. She enjoys analyzing patient data and trying to make informed decisions using data. She also enjoys walking and training her dog, Pebbles.

Results

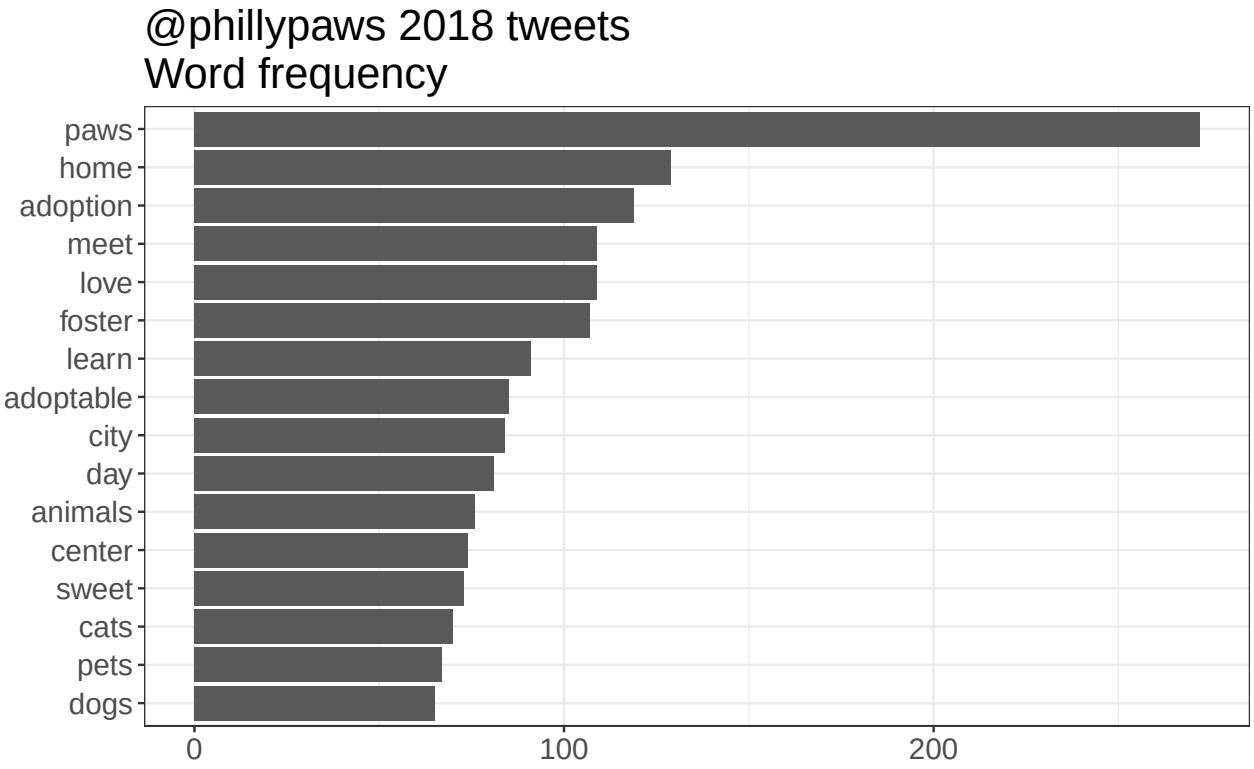


Over the course of 2018, we observed a fairly steady number of tweets per month, with somewhat lower volume in Spring (951 favorites, 50 tweets, 480 retweets) and higher volume in Winter (1475 favorites, 69 tweets, 823 retweets). We also examined tweets per week, and noticed that Sundays had the lowest volume (1649 favorites, 79 tweets, 768 retweets), while Wednesdays had the highest volume of activity (2727 favorites, 122 tweets, 1380 retweets).

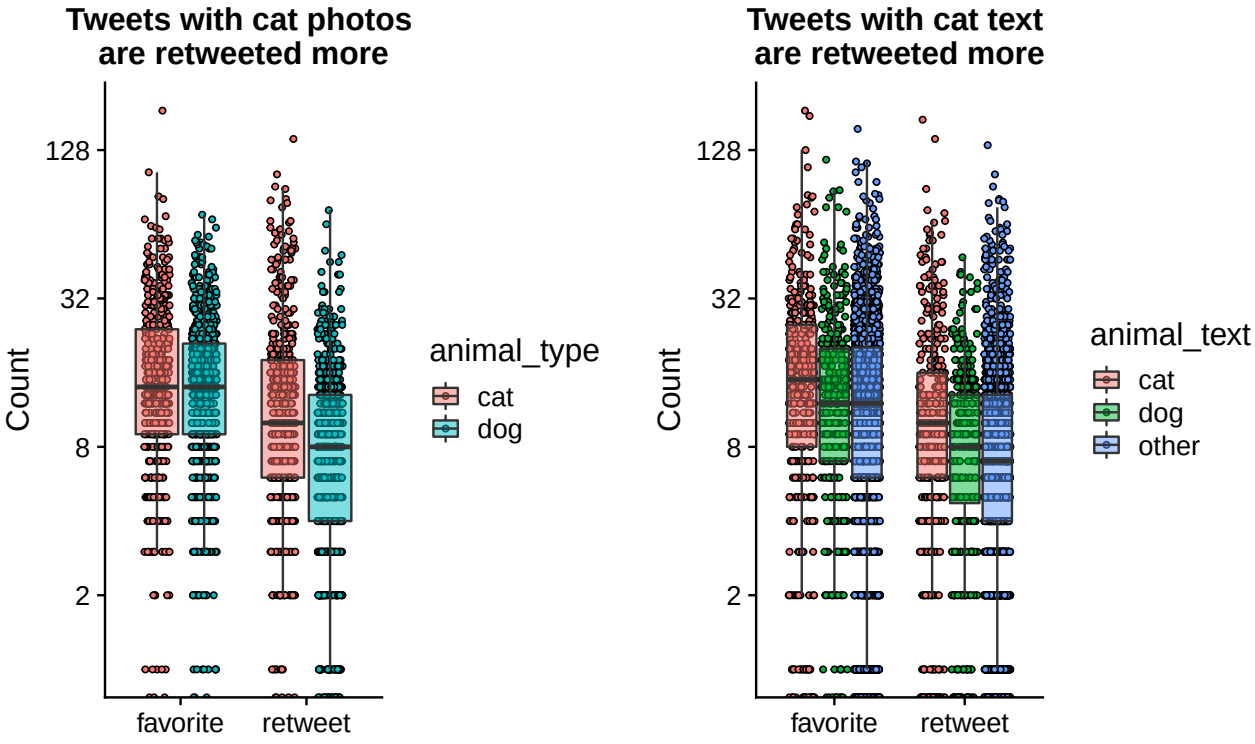
We calculated the ratio of retweets and favorites to PAWS tweets, per day, as an index of how many people see and interact with a tweet. On average in 2018, 1 PAWS tweet was retweeted or liked by 33 people.

We also tested whether the frequency of tweets was related to frequency of applications on a week by week basis. We did not observe a correlation between these two variables.

"PAWS" was the most frequently tweeted word, followed by "home", "adoption", "meet", "love", and "foster".



Tweets with photos, and those with cat text, were retweeted more often.



Data considerations

Over the course of our analyses, we uncovered multiple challenges with the data. We highlight them here in hopes that they will be useful when redesigning the data collection process in the future.

Cross-platform limitations

Every platform used by PAWS (e.g. PetPoint, online forms, Trello, Volgistics, etc.) has its own way of structuring data, and its own limitations as to how the data can be extracted. For instance, the Trello API sets a hard limit on queries for actions, capping a request at 1000 data points. A data point here means every activity (e.g. card creation, checklist added, checklist item checked off, commend added, file attached, etc.) recorded online during the application process, so an animal could have dozens of these. A mass-export of data is therefore more challenging. For a more streamlined data process, it would be ideal if PAWS could reduce the number of platforms used. Additionally, keeping data about the same applicants and the same animals in multiple disconnected databases makes matching these datasets a manual (or nearly manual) process. A better integration of data would result in much deeper insights that could be obtained on a much more regular basis.

Online applications for cats and dogs are not consistent

While the dog and cat applications collect nearly identical information, this is not reflected in the answer options recorded by the online form. For instance, in the question on past experience with pets, answer options were recorded as e.g. "bred-sold-a-pet" and "bred-sold", "euthanized-a-pet" and "euthanized", "lived-wit-previous-housemates_pets" and "lived-with-housemate-pet". The data analysis process could be automated and set up more more efficiently if items that have the same meaning were recorded in identical ways.

Free text entry where a number is desired

Several response options in the online form were set up as free text entries, when in reality the question required a number (e.g. "How long will the cat be left alone?" allows a short paragraph for a response, which leaves it up to whoever processes the application, or whoever analyzes the data, to read the paragraph and make a decision about the number). It would be more efficient for application processing, as well as for analysis, if these fields were set to require a numeric answer. With this kind of data validation in place, PAWS could set up automatic logic *inside* the application form that could immediately identify a response that is problematic and guide the applicant to reconsider.

Missing Data

We were at times limited due to missing data in the applications data set. In the application analysis, the question with the most missing data was the one regarding home pet policy. This seems like an important question, especially for renters, and a non-response here may require manual follow up by PAWS staff. Making this a required question could save time in application processing. More broadly, we recommend that PAWS carefully consider which questions are most valuable, and make those required.

Unlimited Responses and Response Validation

We ran into several challenges as a result of questions having a wide range of possible responses and illogical answers. For example, the 12 different response possibilities below are for the Allergy question:

Response	Count
no-allergies	1,694
mildly-allergic	130
not-sure	38
not-sure,no-allergies	16
very-allergic	10
no-allergies,mildly-allergic	5
no-allergies,not-sure	5

Response	Count
mildly-allergic,no-allergies	3
mildly-allergic,very-allergic	3
mildly-allergic,not-sure	1
very-allergic,mildly-allergic	1
very-allergic,no-allergies	1

In one case the responses conflict with each other: "very-allergic,no-allergies". This makes grouping the data after the fact almost impossible because it's not clear if this applicant has allergies or not. Similar situations were encountered across the dataset. Additionally, there were several items that also required some interpretation and processing decisions. For instance, in the monthly budget question, there were several negative numbers and some extremely large, strange values (i.e $150,159.00). Utilizing some kind of response validation logic (i.e., only allow positive values) and limiting the range of responses to a reasonable size given the question (in this case maybe between 200 and 1,000) would also make future analysis much more efficient and could possibly even help with speed of application processing.

Categorization of response items

For several application questions, it was unclear how these could be classified into broader categories (which could help streamline the online process, as well as future analyses). For instance, the question asking about experience with pets covers a broad range of experiences, some more desirable than others. However, distinguishing problematic answers from non-problematic ones, or classifying someone's experience with a pet (e.g. novice, average, experienced) from these responses is not possible.

Trello label structure

In the Trello card labels we analyzed, we found 37 unique labels, and it was unclear how these were structured. A more deliberate approach to creating and applying labels to cards could help clarify where each application is at a given point without requiring an in-depth read of the comments history.

Conclusions and Next Steps

Based on these analyses, we recommend the following:

- Categorize animals and applications into levels of difficulty for processing so that applications can be matched with staff and volunteers based on their experience level
- Set an expected timeline for animals in foster care, taking into account an animal's species, age and medical needs; incentivize foster parents to adhere to the timeline by creating a clear goal structure (e.g. set number of events to partcipate in, number of social media posts to share, number of applications to process, etc.)
- Identify the reasons for processing speed differences between cat and dog applications
- Reduce time spent on checking vet references by placing some of the processing burden on applicants (e.g. have applicants request a signed statement prior to submitting their application), streamlining the process with vet practices (e.g. set up a shared online system or standardized process), or designing an interview for applicants that can identify the most important information
- Reduce the incidence of red flags by automating some parts of the flagging process (e.g. an insufficient monthly budget) and placing the processing burden back on the applicant before the application reaches PAWS staff (e.g. either by requesting that the applicant attach files with their application); alternatively, early flagging could automatically direct an application to a more experienced staff member who could potentially handle the application more efficiently
- Focus resources on young and sick cats in the spring and summer months in order to reduce wait times
- Develop a "smart" online application, that automatically educates the applicant on the cost of pet ownership when the budget is too low

- Provide applicants with a detailed breakdown of costs for a new pet, and have an adoption counselor go through the itemized list with the applicant to identify how each item could be covered
- Promote sharing or pooling of resources among adopters (e.g. food, treats and toys) to help them financially
- Assess the user-friendliness of the online application form on different platforms (eg web vs. mobile)
- Consider creating programs for families with children to help them complete the adoption process and/or to make adopting an animal more accessible to them
- Keep on tweeting! (for future analyses it may be interesting to look at tweets about certain topics (a specific animal, request for donations) and specific outcomes (did the animal get adopted? did they receive more donations?)