

Statistical Analysis of Animal Adoption using R

Ridhi Anand, Sherin Miriam Cherian, Sriya Ravi, Vaishnavi Agarwal, Aswin Sampath, S.Shreenidhi, Akash Menon, Srinath K R, Umamaheswari .S

Abstract: Animal shelters and rescue organizations hold the responsibility for catering to the needs of the animals that are abandoned or given up by their owners. The datasets are obtained from adoption centers in Austin, Texas between the years 2013 and 2017 and from Bloomington, Indiana between the years 2004 and 2017 with the aim of finding patterns and conducting data analysis on the different parameters that were involved with these adoption centers. Since 3 primary datasets dealing with different attributes related to the adoption centers are procured, all implementation in this study are purely based on this. Hence, the paper is designed to have several specific objectives and derived inferences for each of the datasets. The data is analysed with the statistical methods and visualization tools like bar plots, pie charts, line graphs, predictive regression models, all executed on R studio, an integrated development environment for R. The statistical model discussed in the data set enable us to make conclusions about why certain animals are euthanized, whether there is any particular preference for adoption, what kind of animals are taken in by adoptions centers, whether the total annual revenue gained by the adoption centers could be predicted and so forth. R has proven to be extremely useful because of its ease of use, excellent visualization tools and robust environment.

Keywords: Adoption, Dataset, Euthanasia, Parameters, Shelter.

I. INTRODUCTION

Animal adoption has seen a solid trend in the society for decades. We have come very far with the way we manage abandoned animals and strays to rehabilitate them. While organizations like the Blue Cross of India (BCI) have operated since the year 1959, with the advancement of technology, we now have the system of online pet adoption where databases of pets taken in by numerous rescue groups and animal shelters are maintained and are openly available for access and utilization by the public. Further, a lot of innovative work has been going into the field of animal welfare.

Revised Manuscript Received on May 07, 2019.

Ridhi Anand, School of Computing Science and Engineering, VIT, Chennai

Sherin Miriam Cherian, School of Computing Science and Engineering, VIT, Chennai

Sriya Ravi, School of Computing Science and Engineering, VIT, Chennai

Vaishnavi Agarwal, School of Computing Science and Engineering, VIT, Chennai

Aswin Sampath, School of Computing Science and Engineering, VIT, Chennai

S.Shreenidhi, School of Computing Science and Engineering, VIT, Chennai

Akash Menon, School of Computing Science and Engineering, VIT, Chennai

Srinath K R, School of Computing Science and Engineering, VIT, Chennai

Umamaheswari .S, School of Advanced Sciences, Mathematics Division, VIT, Chennai.

A good example of this is BCI receiving the Proggy Award from PETA for its interactive computer programmes which serve as an alternate to animal dissection. A good example of this is BCI receiving the Proggy Award from PETA for its interactive computer programmes which serve as an alternate to animal dissection. An important parameter to discuss when we talk about animal welfare is the question of how people deal with unwanted pets. Euthanasia is a controversial option with an open debate on whether putting an animal down is humane in comparison to having it caged and imprisoned for long periods of time.

Some people euthanize pets because of fatal and irrecoverable injuries or illnesses. Others do so even for treatable health problems that they either are unable to financially provide for or are not willing to. Without doubt, many veterinarians consider euthanasia to be a completely unethical procedure. Some other people let go of their pets, into the wild perhaps, or abandon them with the hope that they will be able to care for themselves or that someone will find or adopt them. In reality, most of these pets that are released by their owners perish to harsh weather, hunger, traffic or easily treatable health issues. In addition to this, homes cannot be always found in adoption centres and unless the centre has a no-kill policy, excess animals are also euthanized to make room for new pets. It was estimated by The Humane Society of the US, that every year, 2.4 million adoptable, healthy dogs and cats are euthanized in the United States of America because of unavailability of homes.

II. LITERATURE REVIEW

Paul C. Bartlett, et.al, [1] stated that "In 2003, of the 140,653 dogs that were discharged from Michigan shelters in 2003, it was found that 28% were taken in for adoption and 40% had to undergo euthanasia; this implied that the annual euthanasia rate was 2.6% for the estimated 2003 Michigan population of dogs. The annual rate of euthanasia for cats in 2003 was much higher than that of dogs (3.1% and 2.6%). Of the 134,405 cats discharged from Michigan shelters in 2003, only 24% were adopted and 57% were euthanized. For both dogs and cats, altered animals had much lower rates of euthanasia than unaltered animals. Among government and private owned animal shelters, private shelters had 20% of their dogs euthanized and 29% of their cats compared to 30% of dogs and 50% of cats euthanized in facilities owned by the government. Additionally, the results indicated that the larger the size of the shelter, the higher the rate of euthanasia for both dogs and cats regardless of ownership.



Brown W.P et.al,[2] analysed the effects of age, sex , breed designation, coat color, and coat patterns on the length of stay

(LOS) for cats in a no-kill shelter; and also compared the physical characteristics and LOS of cats adopted at a no-kill shelter and in traditional shelters. His article indicates feline physical characteristics influence LOS, likelihood of adoption, or risk of euthanasia at shelters of all kinds and recommend exploring strategies to increase the adoptability of less desirable cats and kittens.

Rachael Mozes et.al, [11] evaluated the factors associated with euthanasia in an animal shelter in Kitchener-Waterloo, Ontario, Canada. He concluded that 53% of admitted cats were euthanized. The adoption policies used in animal shelters have a great effect on the total number of adoptions and donations that occur. Lord .E et.al, [7] established a stochastic model to simulate the expected costs, revenues, and net income of a hypothetical animal shelter for various alternative management strategies, based on US conditions. Sloane Hawes et.al [12] studied how the quality of life of older animals is impacted by remaining in the care of shelters rather than being euthanized. The sample consisted of 122 dogs and 124 cats that were above the age of 7 years who were taken into a shelter in a one-year period. The article highlights the significance of preventative outreach, behaviour programs and specialized medical, and a powerful foster care system that are equipped to deal with the requirements of older animals in their homes and in a shelter

III. MODEL FORMATION

The article deals with three primary datasets procured from the records of the adoptions centres in Austin, Texas between the years 2013 and 2017 and Bloomington, Indiana between the years 2004 to 2013. The three primary datasets in the model deal with various parameters related to these adoption centres. The objective of the paper is to answer typical questions and gather inferences on each of these datasets. Glimpses of the datasets are provided in the Parameters section of the article. All the datasets went through an initial process of clean up wherein the unwanted parameters (like the names of animals) were removed and data was converted to a more useful form (like the dates converted to proper format and time removed). The first dataset, Austin Animal Intake Dataset, was worked upon to understand the distribution of intake type, different animals adopted, trend in intake and the age of intake for different types. The second dataset worked upon was All Records, where we seek to answer questions about the relation between euthanasia and outcome subtype, and we see how the reproductive state of the animal is related with different parameters. ANOVA testing is used to understand the variation in adoption, return to owner, euthanization and transfer rates based on outcome sex and breed. Both of these datasets deal with animal shelters from Austin, Texas. The final dataset, Animal Care and Control, contains information about the overall summed up numbers of budget, revenue, intake, subtypes etc., for Bloomington, Indiana. This was analysed to visualize and compare trends in the annual budget and the total intake of the adoption centres, the number of adoptions, strays, euthanized animals over the years and to understand how many surrendered animals were returned to owners. A multiple regression analysis was

performed to predict the annual revenue of the adoption centres.

We assume that there are no outliers in the data procured and that the data is accurate. Since there are multiple datasets to work on, we split our analytical objectives based on them. We also make specific inferences pertaining to each individual dataset.

Austin Animal Intake dataset

Austin Animal Intake [3] Austin, Texas

Objectives: To analyse the animal intake dataset and draw conclusions through visualisation.

Work, Output and Inferences

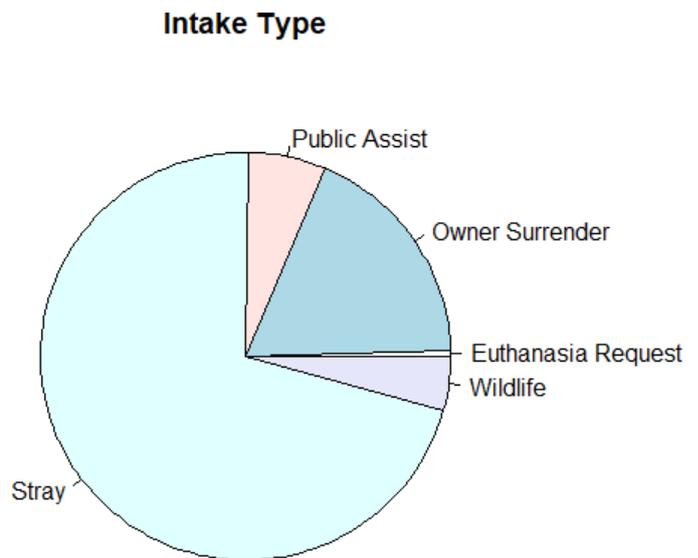


Fig . 1 Distribution of Intake Type

From the above figure, it is clear that stray animals are the ones that are taken by the animal shelters in large numbers.

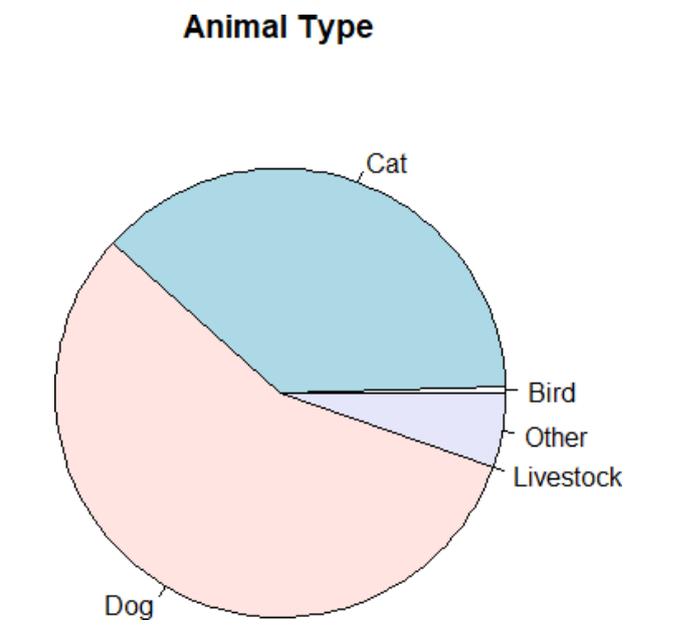


Fig . 2 Distribution of Animal Type

From the pie chart, it is evident that dogs are the animals that are taken by the shelters in large numbers, followed by cats.

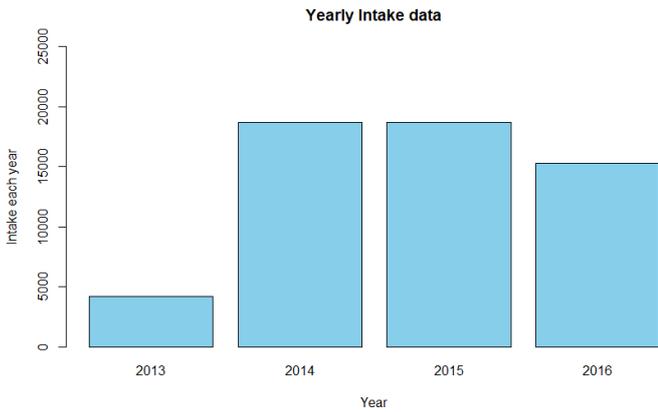


Fig. 3 Yearly Intake of Data

The bar plot reveals that among the given years, 2014 and 2015 witnessed the largest of animal intakes. Chron package is used for manipulating date data type. [6]

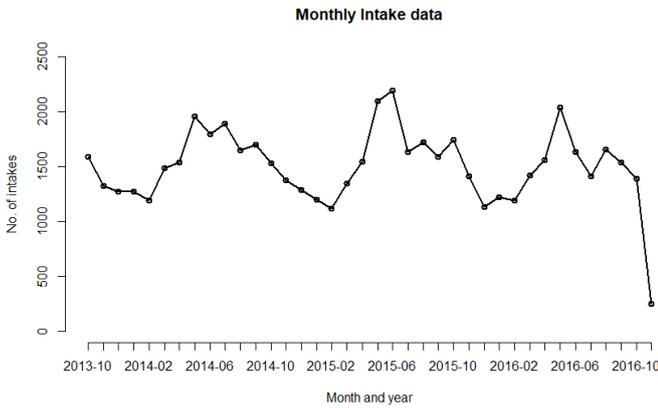


Fig. 4 Monthly Intake of Data

The above plot concludes that June 2015 witnessed the largest animal intakes when compared to other months.

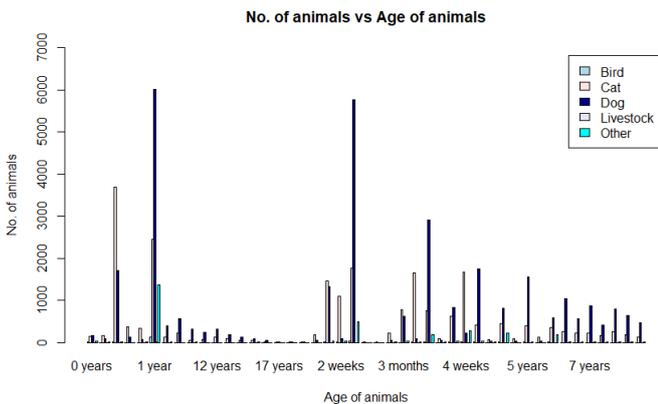


Fig. 5 Number of Animals Vs Age For Each Subtype

The graph shows that the age of dogs corresponding to maximum intake is 1 year. Similarly, for cats it is approx. 5 months.

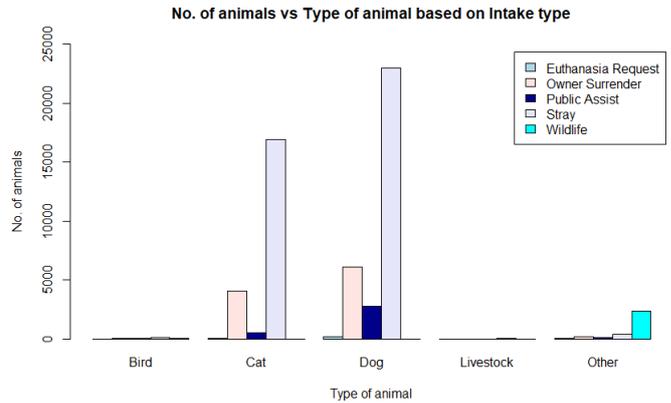


Fig. 6 Segregation by Intake Type

The above figure reveals that euthanasia request for each animal type is negligible; in dogs and cats, stray are given more preference.

All records Dataset

All Records [13] Austin, Texas

Objectives:

- To understand the relation between Euthanasia and Outcome subtype (to look at the reasons for euthanization)
- To check if adoption is preferred for certain genders/neutered and spayed species
- To pictorially understand how Intake sex and Outcome sex relates with a stray dog. (To understand that most strays are neutered to stop reproduction and prevention of more strays).

Work, Output and Inferences:

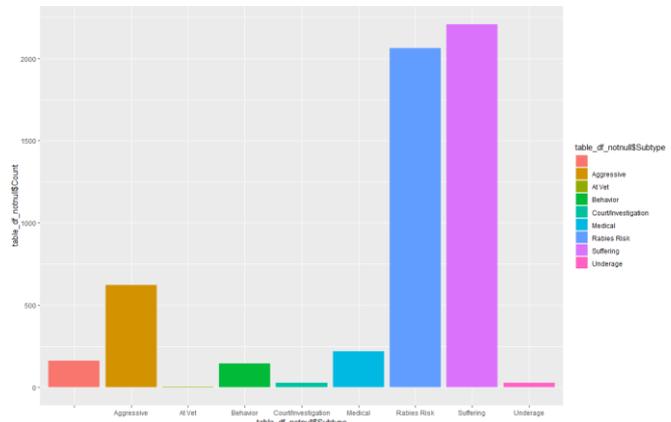


Fig. 7 Euthanasia Vs Outcome Subtype

Suffering is the most common reason to euthanize animals. "Rabies risk" subtype [8]. It can be noted that that risk of rabies is being eliminated effectively. It is also one of the major causes for the animal to be euthanized. Aggressive animals are also euthanized, as there is no other effective method to deal with such animals.

From the above graph, it is noted that a few dogs that are "underage" are euthanized too [8]. This could be attributed to the fact that infant dogs are vulnerable to the living conditions and are at risk. Animal shelters are also not properly equipped to give the personal care that they need.

Nearly a third of dogs originally brought in by owners for euthanasia were determined to be potentially savable [6].

Table 1 . Intake Type and Sex (Intake)

```

all_records.Sex
all_records.Intake_Type Intact Female Intact Male Neutered Male Spayed Female Unknown
Euthanasia Request      51         41          91         76         11
Owner Surrender        2924       2678       5234       5151       135
Public Assist          853        1310       1901       1138       111
Stray                  16585     18268     8140       6832     2673
wildlife                39         57          0          0         2677
    
```

From the above data, it is clear that:

- (i) Male to female ratio is almost equally split.
- (ii) Most of the owners own a neutered or spayed animal.
- (iii) Stray animals are more in number.

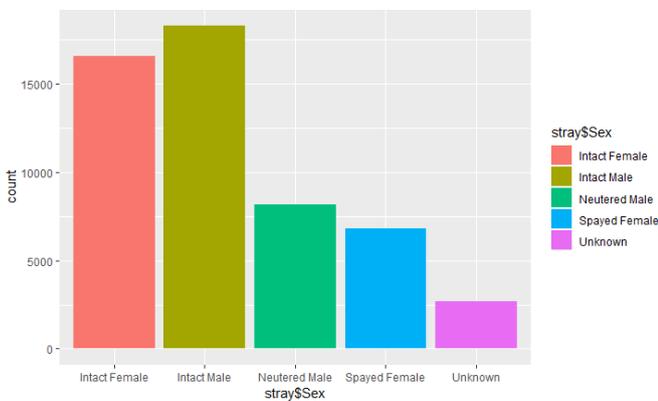


Fig . 8 Intake Type (Stray) And Sex (Intake)

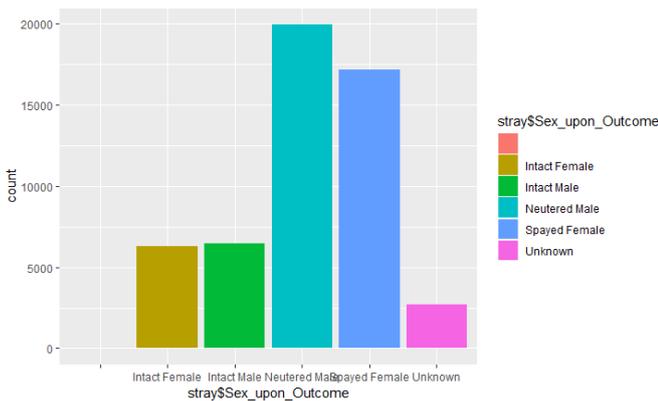


Fig . 9 Intake Type(Stray) And Sex(Outcome)

The graph reveals that (i) Most of the strays are neutered or spayed to limit reproduction. This leads to less number of stray animals. (ii) A lot of effort is put in by the animal shelters to limit reproduction. As we look from other graphs (intake of animal each year), the number of animals that are taken in by the animal shelters also goes down each year.

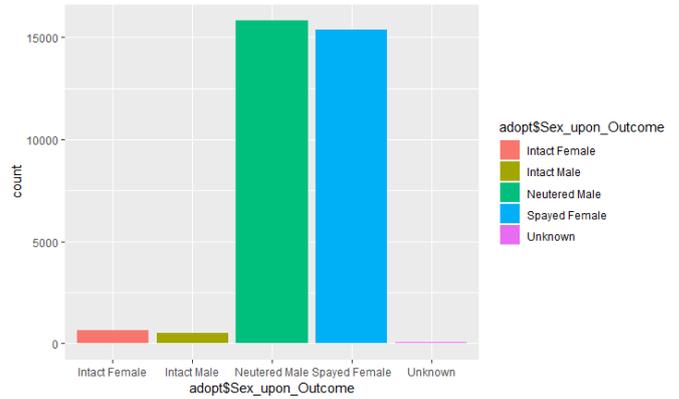


Fig. 10 Adoption and Outcome Sex

It is noted that the most adopted animals are either neutered or spayed. This could be attributed to two reasons:

- (i) By the time the animals are adopted, most animals are neutered or spayed to avoid increase in number of stray animals.
- (ii) Most of the people who adopt animals prefer neutered/spayed animals. This could be attributed to the fact that animal owners have economic and personal time constraints.

All Records dataset
Animal Intake Services Animal Care and Control [6]
Bloomington, Indiana

Objectives:

- To analyse the dataset and to find:
- If there is a significant difference in the adoption rates based on the outcome sex of the animal and its breed
- If there is a significant difference in the euthanasia rates based on the outcome sex of the animal and its breed
- If there is a significant difference in the Return to owner rates based on the outcome sex of the animal and its breed
- If there is a significant difference in the Transfer rates based on the outcome sex of the animal and its breed
- Work, Output and Inferences

A FUNCTION WHICH CREATES A VECTOR OF 0s BASED ON SIZE OF ARGUMENT:

```

create=function(arg){
i=1
vec=c()
while(i<=arg){
vec=append(vec,0,i)
i=i+1
}
return(vec)
}
> vec= create(9)
> vec
[1] 0 0 0 0 0 0 0 0 0
    
```

A FUNCTION WHICH COUNTS THE OCCURRENCE OF A BREED WITH DIFFERENT OUTCOME_SEX TYPES:

```

count=function(br,osex,dataset){
l=nrow(dataset)
i=1
ll=8
    
```



```
sf=create(8)
nm=create(8)
im=create(8)
ife=create(8)
while(i<=l){
j=1
while(j<=ll){
if(br[j]==dataset[i,2]){
if(dataset[i,3]=="Intact Female"){
ife[j]=ife[j]+1
}else if(dataset[i,3]=="Intact Male"){
im[j]=im[j]+1
}else if(dataset[i,3]=="Neutered Male"){
nm[j]=nm[j]+1
}else if(dataset[i,3]=="Spayed Female"){
sf[j]=sf[j]+1
}
}
break
}
j=j+1
}
i=i+1
}
ds=c(sf,nm,im,ife)
dt=data.frame(matrix(data = ds,ncol = 8,byrow = TRUE,dimnames = list(osex,br)))
return(dt)
}
```

It takes a data frame, a vector containing breeds and a vector containing the outcome sex types and returns a data frame which contains the frequency of occurrence of a particular breed type and a particular sex type in the given data frame.

Example:

```
aset=count(breed_req,outcome_sex,adopt_set)
```

Table 2 . Result

	Australian.	Chihuahua	Domestic.	Domestic.	Domestic.	German.Sf	Labrador.F	Pit.Bull.Mix
Intact Fem	327	987	262	397	4014	463	1058	1309
Intact Mal	269	1075	219	437	3840	487	1062	1233
Neutered l	2	32	10	21	165	6	31	25
Spayed Fei	11	27	19	29	266	11	29	30

On saving the data frame as a csv file the following result is obtained. The columns represent the different breed types and the rows contain values for each outcome sex type.

Pre-processing of dataset:

Getting significant breeds:

```
breed=table(dat_f$Breed_intake)
```

```
breed=data.frame(breed)
```

```
breed_refined=subset(breed,breed$Freq>1000)
```

```
breed_req=breed_refined$b
```

```
breed_req=breed_req[breed_req!="Bat Mix"]
```

2) Getting the required outcome sex types:

```
outcome_sex=data.frame(table(dat_f$Sex_upon_Outcome))
```

```
outcome_sex=outcome_sex[3:nrow(outcome_sex)-1,]
```

```
outcome_sex=outcome_sex$Var1
```

3) Creating subsets of the chosen dataset:

```
#creating dataset with just type, sex and breed
```

```
os=dat_f$Sex_upon_Outcome
```

```
type=dat_f$Outcome_Type
```

```
breeds=dat_f$Breed_intake
```

```
complete_set=data.frame(type,breeds,os)
#creating subsets according to outcome type
adopt_set=subset(complete_set,complete_set$type=="Adoption")
euth_set=subset(complete_set,complete_set$type=="Euthanasia")
rto_set=subset(complete_set,complete_set$type=="Return to Owner")
transfer_set=subset(complete_set,complete_set$type=="Transfer")
#refining subsets according to required breeds
adopt_set=subset(adopt_set,adopt_set$breeds %in% breed_req)
euth_set=subset(euth_set,euth_set$breeds %in% breed_req)
rto_set=subset(rto_set,rto_set$breeds %in% breed_req)
transfer_set=subset(transfer_set,transfer_set$breeds %in% breed_req)
#refining subsets according to sex
adopt_set=subset(adopt_set,adopt_set$os %in% outcome_sex)
euth_set=subset(euth_set,euth_set$os %in% outcome_sex)
rto_set=subset(rto_set,rto_set$os %in% outcome_sex)
transfer_set=subset(transfer_set,transfer_set$os %in% outcome_sex)
Example after pre-processing:
> head(transfer_set)
```

Table 3 . Result

	type	breeds	outcome_sex
1	Transfer	Domestic Shorthair Mix	Neutered Male
5	Transfer	Domestic Shorthair Mix	Neutered Male
11	Transfer	Domestic Shorthair Mix	Spayed Female
13	Transfer	Domestic Shorthair Mix	Intact Male
15	Transfer	Pit Bull Mix	Intact Male
23	Transfer	Domestic Shorthair Mix	Intact Male

ANOVA statistics based on Adoption values:

H0: There is no significant difference in outcome sex type

H1: There is a significant difference in outcome sex type

H0: There is no significant difference in Breed type

H1: There is a significant difference in Breed type

Table 4 . Anova Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Sex	3	8744859	2914953	6.669	0.00245 **
Breed	7	11448194	1635456	3.742	0.00874 **
Residuals	21	9178756	437084		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

It concludes that both sex and breed are statistically significant since F values 0.00245 and 0.00874 are smaller than 0.05. So H0 for both outcome sex type and Breed type are rejected. Therefore it leads to a conclusion that there would be a significant difference in the frequency of adoption when these factors are affected.

ANOVA statistics based on Euthanization values:

H0: There is no significant difference in outcome sex type

H1: There is a significant difference in outcome sex type

H0: There is no significant difference in Breed type

H1: There is a significant difference in Breed type

Anova Table:



Table 5 .Anova Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Sex	3	5955	1985	1.368	0.28
Breed	7	139094	19871	13.692	1.5e-06 ***
Residuals	21	30475	1451		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

F value for sex is 0.28 which is >0.05 and F value for Breed is significantly smaller than 0.05. Therefore, H_0 is rejected for breed type and accepted for outcome sex. Hence there is no significant difference in euthanization frequency on outcome sex but there is a difference upon changing the breed.

ANOVA statistics based on Return To owner values:

H_0 : There is no significant difference in outcome sex type

H_1 : There is a significant difference in outcome sex type

H_0 : There is no significant difference in Breed type

H_1 : There is a significant difference in Breed type

Table 6 . Anova Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Sex	3	790341	263447	10.407	0.000211 ***
Breed	7	1572770	224682	8.876	4.35e-05 ***
Residuals	21	531586	25314		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

From this it is clear that both sex and breed are statistically significant since F values for both are smaller than 0.05. So H_0 for both outcome sex type and Breed type are rejected. Therefore there would be a significant difference in the frequency of dogs being returned to owner when these factors are affected.

ANOVA statistics based on Transfer values:

H_0 : There is no significant difference in outcome sex type

H_1 : There is a significant difference in outcome sex type

H_0 : There is no significant difference in Breed type

H_1 : There is a significant difference in Breed type

Table 7. Anova Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Sex	3	127398	42466	0.751	0.534
Breed	7	10829223	1547032	27.370	3.46e-09 ***
Residuals	21	1186978	56523		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

F value for outcome sex is 0.534 which is >0.05 and F value for Breed is significantly smaller than 0.05. Therefore H_0 is rejected for breed type and accepted for outcome sex. So, it infers that there is no significant difference in Transfer of dogs frequency on outcome sex but there is a difference upon changing the breed.

Animal Care and Control dataset

Objectives:

- To visualize the trend in annual budget, total intake of animals.
- To visualize and compare the trend in number of adoptions, strays and euthanized animals.
- To visualize and compare owner surrenders and return to owners.

- To create a regression model and predict the annual adoption Revenue of the Animal centers.

Work, Output and Inferences:

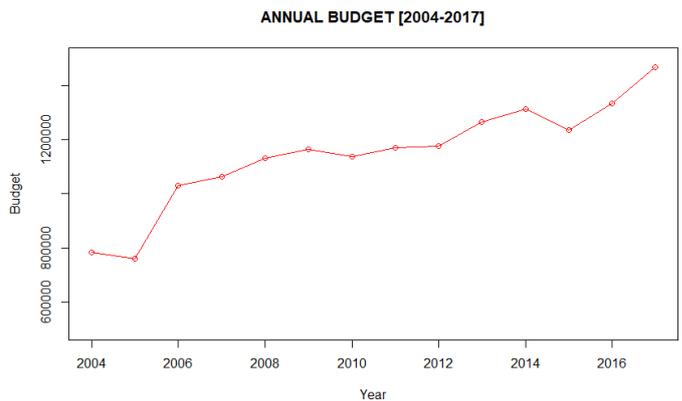


Fig. 11 Annual Budget and Total Intake

IV. CONCLUSION

From the above analysis, it is clear that the stray animals are the largest in number for animal intakes. Suffering is the most common reason to euthanize animals. "Rabies risk" is also one of the common reasons for euthanasia. Aggressive animals can also be expected to be euthanized. Male to female ratio in animals is almost equal. Most of the owners own a neutered or a spayed animal. Stray animals are more in number. Most of the strays are neutered or spayed. Animal shelters probably adopt this method to minimize the number of stray animals. Most adopted animals are also mostly spayed or neutered. There is a significant difference in the frequency of adoption when sex and breed are altered. Euthanization frequency is not affected when outcome sex is altered. It is more dependent on the breed of animals.

A change is noted in the frequency of dogs being returned to the owner when sex/breed value is altered. From this it is clear that both sex and breed are statistically significant since F values for both are smaller than 0.05. So H_0 for both outcome sex type and Breed type are rejected. Therefore there would be a significant difference in the frequency of dogs being returned to owner when these factors are affected. There is also no significant different in the transfer of dogs frequency based on sex, but breed does seem to contribute changes. From the Bloomington, Indiana dataset we see that the annual budget increased significantly from 2004 to 2017. Budget is predicted to increase. Intake of animals has been decreasing throughout the years. This could be due to the consistent efforts of the animal shelters to spay/neuter animals and to encourage people to adopt them. The adoption rates have been increasing; it peaks in the year 2012. Since the number of euthanized and stray animals has been decreasing almost in parallel to the total intake, it appears that they undergo no variation in trend. It goes down by about 20% from 2004-2017. Similarly, the percentage of euthanized animals is decreasing over the years.

The total revenue for the animal adoption centres in Bloomington, Indiana is predicted to be \$144,785.8 given an annual budget of \$1500000 and a total intake of 2000 animals.



This implies the adoption centres on a whole earn about 10% of their annual budget in this hypothetical year.

Hence, the results have been discussed in individual sections and inferences have been made from our work on the datasets. R Studio has proven to be immensely helpful as data analysis and visualization is made simple and it can be used to display easily understandable and aesthetic graphs. R is not just for advanced programmers and includes a variety of packages for prediction algorithms and plots

REFERENCES

1. Bartlett, P. C., Bartlett, A., Walshaw, S., & Halstead, S. (2005). Rates of euthanasia and adoption for dogs and cats in Michigan animal shelters. *Journal of Applied Animal Welfare Science*, 8(2), 97-104.
2. Brown, W. P., & Morgan, K. T. (2015). Age, Breed Designation, Coat Color, and Coat Pattern Influenced the Length of Stay of Cats at a No-Kill Shelter. *Journal of Applied Animal Welfare Science*, 18(2), 169-180
3. Clay, Rebecca (2016): Austin, Texas Animal Shelter Intake and Outtake dataset from <https://data.world/rebeccaclay/austin-tx-animal-center-stats>
4. David James and Kurt Hornik (2015). *chron: Chronological Objects which Can Handle Dates and Times*. R package version 2.3-47.
5. H. Wickham. (2009) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
6. Indiana Government, The United States of America: Animal Care and Control Division Statistics from <https://data.bloomington.in.gov/dataset/9a0fc126-0775-4952-96ff-76ece67ffbc7>
7. Lord, E., Widmar, N. O., & Litster, A. (2014). Economic impacts of adoption and fundraising strategies in animal shelters. *Preventive Veterinary Medicine*, 113, 423-429.
8. La Jeunesse, C. G. (2013). Workable Answers: Expanding Knowledge Around Euthanasia: A Review of Blue Juice: Euthanasia in Veterinary Medicine, by Patricia Morris. *Journal of Applied Animal Welfare Science*, 16(2), 205-208
9. Patronek, G. J., & Crowe, A. (2018). Factors Associated with High Live Release for Dogs at a Large, Open-Admission, Municipal Shelter. *Animals*, 8(4), 45.
10. R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
11. Rachael Mozes, David L Pearl, Lee Niel, J Scott Weese(2017) Epidemiological investigation of euthanasia in an Ontario animal shelter, *Journal of Feline Medicine and Surgery* , 20 (6), 479-486
12. Sloane Hawes, Josephine Kerrigan, and Kevin Morris (2018) Factors Informing Outcomes for Older Cats and Dogs in Animal Shelters, *Animals* , 8(3), 36.
13. Texas government, The United States of America, Data set on Austin Animal shelters [csv files] from <https://data.austintexas.gov/Health-and-Community-Services/Austin-Animal-Center-Outcomes/9t4d-g238>
14. Zar, J. H. (1999). *Biostatistical analysis*. Upper Saddle River, NJ: Prentice-Hall.