

Breeds impact on adoption

Matthew Wilcox, Randy Otoo, Anthony Barrios, Arianne PozziBarba

Libraries

```
library(tidyverse)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.2
## Warning: package 'dplyr' was built under R version 4.3.2
## Warning: package 'stringr' was built under R version 4.3.2
## Warning: package 'lubridate' was built under R version 4.3.2

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.4
## v forcats    1.0.0      v stringr    1.5.1
## v ggplot2    3.4.4      v tibble     3.2.1
## v lubridate  1.9.3      v tidyr      1.3.0
## v purrr      1.0.2

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 4.3.2
```

```
library(kableExtra)
```

```
## Warning: package 'kableExtra' was built under R version 4.3.2
```

```
##
## Attaching package: 'kableExtra'
##
## The following object is masked from 'package:dplyr':
##
##   group_rows
```

```
library(Rgraphviz)
```

```
## Loading required package: graph
## Loading required package: BiocGenerics
##
## Attaching package: 'BiocGenerics'
##
## The following objects are masked from 'package:lubridate':
##
##   intersect, setdiff, union
```

```
##
## The following objects are masked from 'package:dplyr':
##
##   combine, intersect, setdiff, union
##
## The following objects are masked from 'package:stats':
##
##   IQR, mad, sd, var, xtabs
##
## The following objects are masked from 'package:base':
##
##   anyDuplicated, aperm, append, as.data.frame, basename, cbind,
##   colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,
##   get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,
##   match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
##   Position, rank, rbind, Reduce, rownames, sapply, setdiff, sort,
##   table, tapply, union, unique, unsplit, which.max, which.min
##
##
## Attaching package: 'graph'
##
## The following object is masked from 'package:stringr':
##
##   boundary
##
## Loading required package: grid
library(data.tree)

## Warning: package 'data.tree' was built under R version 4.3.2
```

Import and Clean Data

```
animal_intake.df <- read_csv("Louisville_Metro_ky_-_Animal_Service_Intake_and_Outcome.csv")

## Rows: 54494 Columns: 18
## -- Column specification -----
## Delimiter: ","
## chr (17): kennel, animalid, jurisdiction, intype, insubtype, indate, surreas...
## dbl (1): ObjectId
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
dog_intake.df <- animal_intake.df %>% filter(animaltype == "DOG") %>%
  mutate(indate = str_remove(indate, ".*"),
         outdate = str_remove(outdate, ".*"),
         indate = as.Date(indate, format = "%Y/%m/%d"),
         outdate = as.Date(outdate, format = "%Y/%m/%d"),
         timeHeld = outdate - indate
  ) %>%
  filter(intype == "STRAY",
         outtype != "DISPOSAL") %>%
  select(outtype, sex, bites, petsize, color, breed, timeHeld)
```

Table 1: Frequency Table of Pitbull and Adoption

	Is Pitbull	Not Pitbull
ADOPTION	1455	3147
EUTH	336	334

Probability

```

pitbull_adopt_euth <- dog_intake.df %>%
  select(outtype, breed) %>%
  filter(
    outtype == "ADOPTION" | outtype == "EUTH"
  ) %>%
  mutate(breed = case_when(
    grepl("PIT BULL", breed) ~ "Is Pitbull",
    TRUE ~ "Not Pitbull"
  )) %>% group_by(outtype, breed) %>%
  summarise(n = n()) %>%
  pivot_wider(names_from = breed, values_from = n) %>%
  column_to_rownames(., var = "outtype")

## `summarise()` has grouped output by 'outtype'. You can override using the
## `.groups` argument.

pitbull_adopt_euth %>% kbl(caption = "Frequency Table of Pitbull and Adoption") %>% kable_classic_2(fu

# if (!require("BiocManager", quietly = TRUE))
#   install.packages("BiocManager")
#
# BiocManager::install("Rgraphviz")
library("Rgraphviz")

total <- sum(pitbull_adopt_euth)

# Probability of a (Prob is pitbull)
a<- sum(pitbull_adopt_euth[1])/total

# Probability (b | a) (Prob Adopted given pitbull)
bGivena<- pitbull_adopt_euth[1,1]/sum(pitbull_adopt_euth[1])

# Probability (b | ~a)
bGivenNota<- pitbull_adopt_euth[1,2]/sum(pitbull_adopt_euth[2])

##### Everything below here will be calculated

# Calculate the rest of the values based upon the 3 variables above
notbGivena<-1-bGivena
notA<-1-a
notbGivenNota<-1-bGivenNota

#Joint Probabilities of a and B, a and notb, nota and b, nota and notb
aANDb<-a*bGivena
aANDnotb<-a*notbGivena

```

```

notaANDb <- notA*bGivenNota
notaANDnotb <- notA*notbGivenNota

# Probability of B
b<- aANDb + notaANDb
notB <- 1-b

# Bayes theorem - probabiliyt of A / B
# (a / b) = Prob (a AND b) / prob (b)
aGivenb <- aANDb / b

# These are the labels of the nodes on the graph
# To signify "Not A" - we use A' or A prime

node1<-"P"
node2<-"Is Pitbull"
node3<-"Not Pitbull"
node4<-"IS Pitbull & Adopted"
node5<-"Is Pitbull & Euthanized'"
node6<-"Not Pitbull & Adopted"
node7<-"Not Pitbull & Euthanized"
nodeName<-c(node1,node2,node3,node4, node5,node6, node7)

rEG <- new("graphNEL", nodes=nodeNames, edgemode="directed")
#Erase any existing plots
# dev.off()

# Draw the "lines" or "branches" of the probability Tree
rEG <- addEdge(nodeNames[1], nodeNames[2], rEG, 1)
rEG <- addEdge(nodeNames[1], nodeNames[3], rEG, 1)
rEG <- addEdge(nodeNames[2], nodeNames[4], rEG, 1)
rEG <- addEdge(nodeNames[2], nodeNames[5], rEG, 1)
rEG <- addEdge(nodeNames[3], nodeNames[6], rEG, 1)
rEG <- addEdge(nodeNames[3], nodeNames[7], rEG, 10)

eAttrs <- list()

q<-edgeNames(rEG)

# Add the probability values to the the branch lines

eAttrs$label <- c(toString(round(a,4)),toString(round(notA,4)),
  toString(round(bGivena,4)), toString(round(notbGivena,4)),
  toString(round(bGivenNota,4)), toString(round(notbGivenNota,4)))
names(eAttrs$label) <- c(q[1],q[2], q[3], q[4], q[5], q[6])
edgeAttrs<-eAttrs

# Set the color, etc, of the tree
attributes<-list(node=list(label="foo", fillcolor="lightgreen", fontsize="15"),
  edge=list(color="red"),graph=list(rankdir="LR"))

#Plot the probability tree using Rgraphvis
plot(rEG, edgeAttrs=eAttrs, attrs=attributes)

```

```
nodes(rEG)
```

```
## [1] "P" "Is Pitbull"
## [3] "Not Pitbull" "IS Pitbull & Adopted"
## [5] "Is Pitbull & Euthanized'" "Not Pitbull & Adopted"
## [7] "Not Pitbull & Euthanized"
```

```
edges(rEG)
```

```
## $P
## [1] "Is Pitbull" "Not Pitbull"
##
## $`Is Pitbull`
## [1] "IS Pitbull & Adopted" "Is Pitbull & Euthanized'"
##
## $`Not Pitbull`
## [1] "Not Pitbull & Adopted" "Not Pitbull & Euthanized"
##
## $`IS Pitbull & Adopted`
## character(0)
##
## $`Is Pitbull & Euthanized`
## character(0)
##
## $`Not Pitbull & Adopted`
## character(0)
##
## $`Not Pitbull & Euthanized`
## character(0)
```

```
#Add the probability values to the leaves of A&B, A&B', A'&B, A'&B'
text(500,420,round(aANDb,4), cex=.8)
```

```
text(500,280,round(aANDnotb,4),cex=.8)
```

```
text(500,160,round(notaANDb,4),cex=.8)
```

```
text(500,30,round(notaANDnotb,4),cex=.8)
```

```
text(340,440,"(Adopted | Is Pitbull)",cex=.8)
```

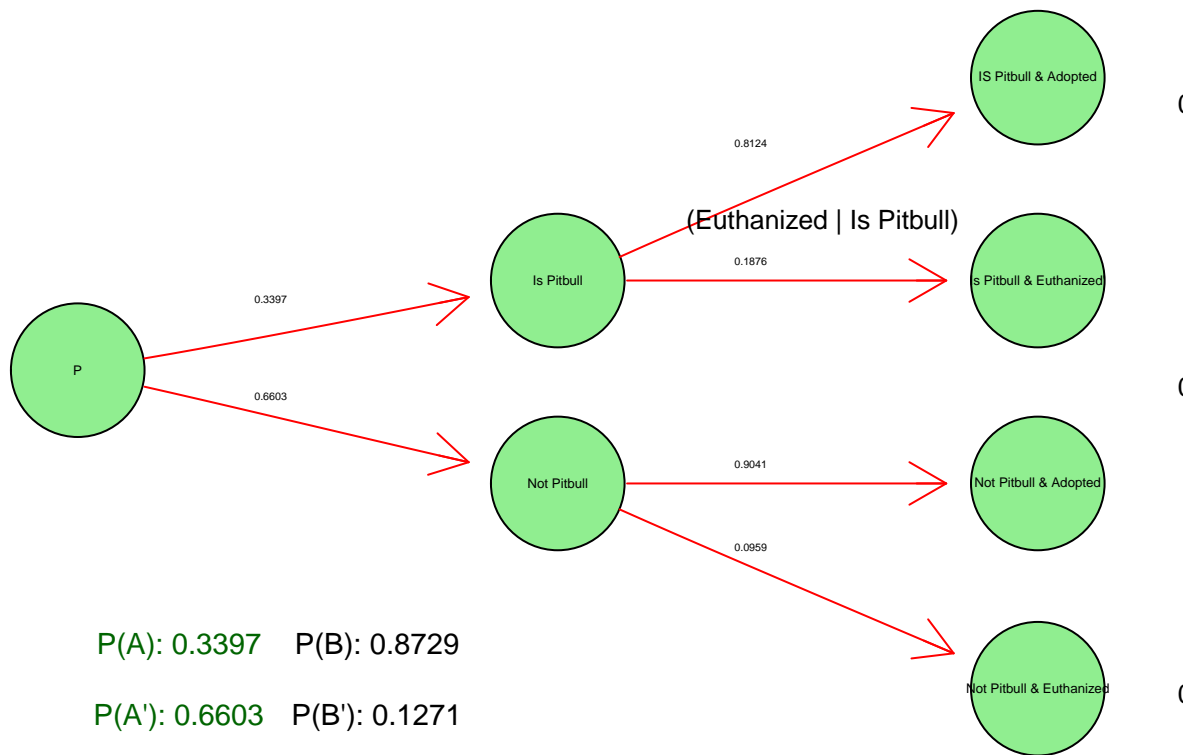
```
text(340,230,"(Euthanized | Is Pitbull)",cex=.8)
```

```
#Write a table in the lower left of the probablites of A and B
text(80,50,paste("P(A):",round(a,4)),cex=.9, col="darkgreen")
text(80,20,paste("P(A'):",round(notA,4)),cex=.9, col="darkgreen")
```

```
text(160,50,paste("P(B):",round(b,digits=4)),cex=.9)
```

```
text(160,20,paste("P(B'):",round(notB, 4)),cex=.9)
```

```
text(80,420,paste("P(A|B): ",round(aGivenb,digits=4)),cex=.9,col="blue")
```



<https://www.harrysurden.com/wordpress/archives/292> referenced to help build the tree diagram.