

Machine Learning Methods and Applications

Weeks-1

Mutlu Kazak

04 03 2021

DATA PROFILING REPORT

```
# Task :Get your data from OpenML dataset by using R and Python and upload the descriptive statistics report to the folder.
```

Needed Packages and Library

```
#install.packages("ggpubr")
#install.packages("ggplot2")
#install.packages("mlbench")
library(ggpubr) ; library(mlbench) ; library(ggplot2) ; library(dplyr)
```

```
## Warning: package 'ggpubr' was built under R version 4.0.4
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 4.0.4
```

```
## Warning: package 'mlbench' was built under R version 4.0.4
```

```
## Warning: package 'dplyr' was built under R version 4.0.4
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

See the Data

```
data(Zoo)

#A data frame with 17 columns: hair, feathers, eggs, milk, airborne, aquatic, predator, toothed, backbone, b
reathes, venomous, fins, legs, tail, domestic, catsize, type.

#16 columns are logical and indicate whether the corresponding animal has the corresponding characteristic
or not. The only 1 exceptions is legs variable takes values 0, 2, 4, 5, 6 and 8.

glimpse(Zoo)
```

```
## Rows: 101
## Columns: 17
## $ hair      <lgl> TRUE, TRUE, FALSE, TRUE, TRUE, TRUE, TRUE, FALSE, FALSE, TRUE~
## $ feathers <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE~
## $ eggs      <lgl> FALSE, FALSE, TRUE, FALSE, FALSE, FALSE, FALSE, FALSE, TRUE, TRUE, F~
## $ milk      <lgl> TRUE, TRUE, FALSE, TRUE, TRUE, TRUE, TRUE, TRUE, FALSE, FALSE, TRUE~
## $ airborne <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE~
## $ aquatic   <lgl> FALSE, FALSE, TRUE, FALSE, FALSE, FALSE, FALSE, FALSE, TRUE, TRUE, F~
## $ predator <lgl> TRUE, FALSE, TRUE, TRUE, TRUE, FALSE, FALSE, FALSE, FALSE, TRUE, FAL~
## $ toothed   <lgl> TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, T~
## $ backbone <lgl> TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, T~
## $ breathes <lgl> TRUE, TRUE, FALSE, TRUE, TRUE, TRUE, TRUE, FALSE, FALSE, TRUE~
## $ venomous  <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE~
## $ fins      <lgl> FALSE, FALSE, TRUE, FALSE, FALSE, FALSE, FALSE, FALSE, TRUE, TRUE, F~
## $ legs      <int> 4, 4, 0, 4, 4, 4, 0, 0, 4, 4, 2, 0, 0, 4, 6, 2, 4, 0, 0, 2~
## $ tail      <lgl> FALSE, TRUE, TRUE, FALSE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, FALSE~
## $ domestic  <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, TRUE, TRUE, FALSE, ~
## $ catsize   <lgl> TRUE, TRUE, FALSE, TRUE, TRUE, TRUE, TRUE, FALSE, FALSE, FALSE, FALS~
## $ type      <fct> mammal, mammal, fish, mammal, mammal, mammal, mammal, mammal, fish, f~
```

Basic Statistics

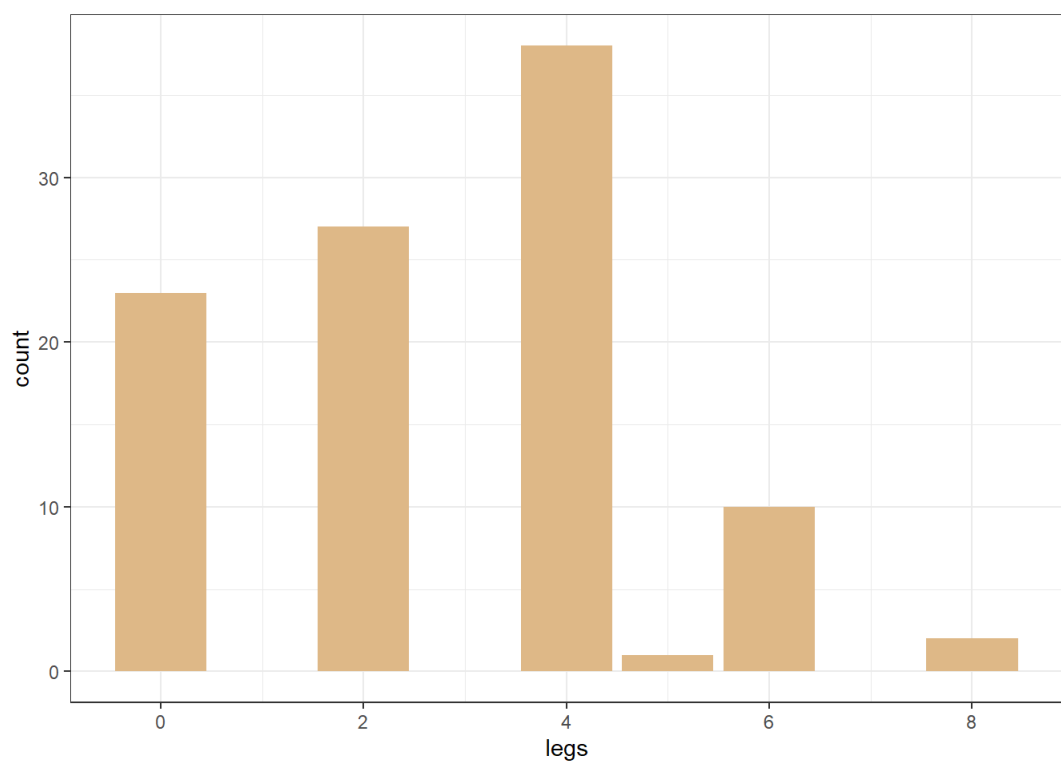
```
#Row Counts
Rows <- length(rownames(Zoo))
Columns <- length(colnames(Zoo))
Mis.val <- sum(is.na(Zoo))
Dis.col <- 16
Con.col <- 1
Tot.obs <- length(row(Zoo))
Names<-c("Rows",
         "Columns",
         "Missing Observations",
         "Discrete Columns",
         "Continuous Columns",
         "Total Observation")

Values<-c(Rows,Columns,Mis.val,Dis.col,Con.col,Tot.obs)
#As we can see in the table , data has 101 rows , 17 columns , 0 missing value , 16 discrete Column , 1 Continuous
df <- data.frame(Names,Values) ; df
```

```
##           Names Values
## 1           Rows    101
## 2          Columns     17
## 3 Missing Observations     0
## 4   Discrete Columns    16
## 5 Continuous Columns     1
## 6   Total Observation   1717
```

Visualization

```
ggplot(Zoo , aes(x=legs))+geom_bar(fill="burlywood") + theme_bw()
```



```

a<-ggplot(Zoo , aes(x=hair))+
  geom_bar(fill="burlywood") + theme_bw() + coord_flip()

b<-ggplot(Zoo , aes(x=feathers)) +
  geom_bar(fill="burlywood") + theme_bw() + coord_flip()

c<-ggplot(Zoo , aes(x=feathers)) +
  geom_bar(fill="burlywood") + theme_bw() + coord_flip()

d<-ggplot(Zoo , aes(x=eggs)) +
  geom_bar(fill="burlywood") + theme_bw() + coord_flip()

e <- ggplot(Zoo , aes(x=milk)) +
  geom_bar(fill="burlywood") + theme_bw() + coord_flip()

f<- ggplot(Zoo , aes(x=airborne)) +
  geom_bar(fill="burlywood") + theme_bw() + coord_flip()

g<-ggplot(Zoo , aes(x=aquatic)) +
  geom_bar(fill="burlywood") + theme_bw() + coord_flip()

h<-ggplot(Zoo , aes(x=predator)) +
  geom_bar(fill="burlywood") + theme_bw() + coord_flip()

j<-ggplot(Zoo , aes(x=toothed)) +
  geom_bar(fill="burlywood") + theme_bw() + coord_flip()

k<-ggplot(Zoo , aes(x=backbone)) +
  geom_bar(fill="burlywood") + theme_bw() + coord_flip()

l<-ggplot(Zoo , aes(x=breathes)) +
  geom_bar(fill="burlywood") + theme_bw() + coord_flip()

m<-ggplot(Zoo , aes(x=venomous)) +
  geom_bar(fill="burlywood") + theme_bw() + coord_flip()

n<-ggplot(Zoo , aes(x=fins)) +
  geom_bar(fill="burlywood") + theme_bw() + coord_flip()

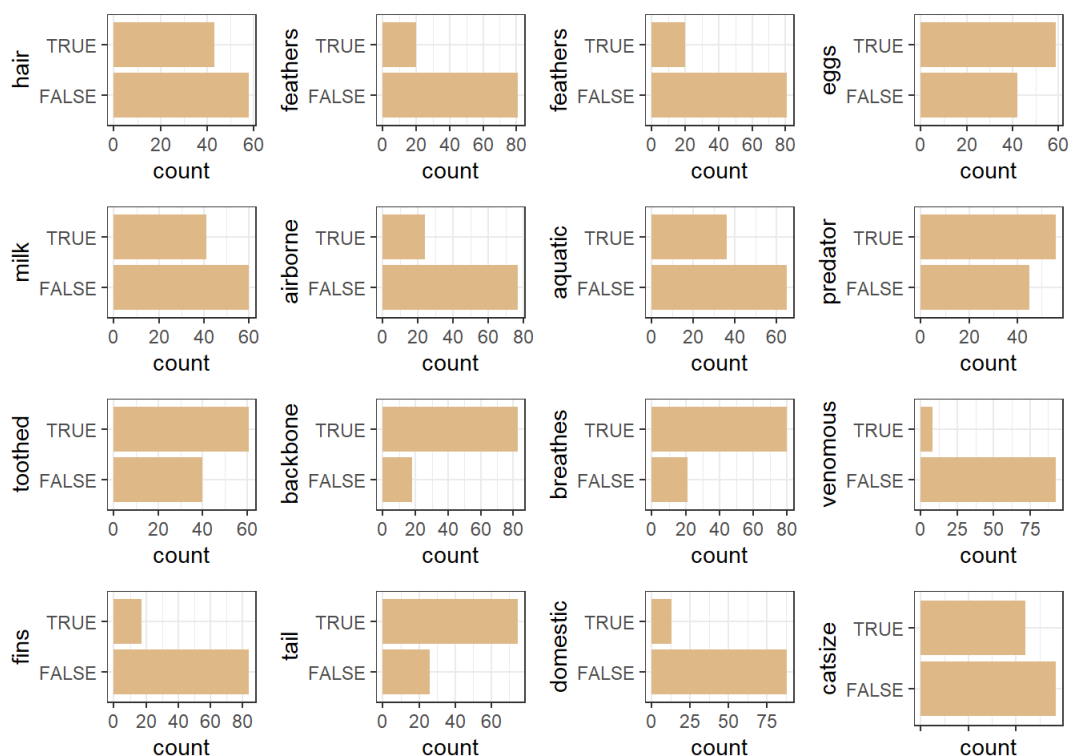
o<-ggplot(Zoo , aes(x=tail)) +
  geom_bar(fill="burlywood") + theme_bw() + coord_flip()

p<-ggplot(Zoo , aes(x=domestic)) +
  geom_bar(fill="burlywood") + theme_bw() + coord_flip()

r<-ggplot(Zoo , aes(x=catsize)) +
  geom_bar(fill="burlywood") + theme_bw() + coord_flip()

# When we search the plots below , we can comment as
#- hair variable has 43 True value , it means 43 of 101 animals have hair .
#- domestic variable has 13 True values , it means 13 of 101 animals are domestic rest of them unteatable ani
mal.
#- tail variable has 75 True values , it means 75 of 101 animals have tail on their body
#-
ggarrange(a,b,c,d,e,f,g,h,j,k,l,m,n,o,p,r + rremove("x.text"),
  ncol = 4, nrow = 4)

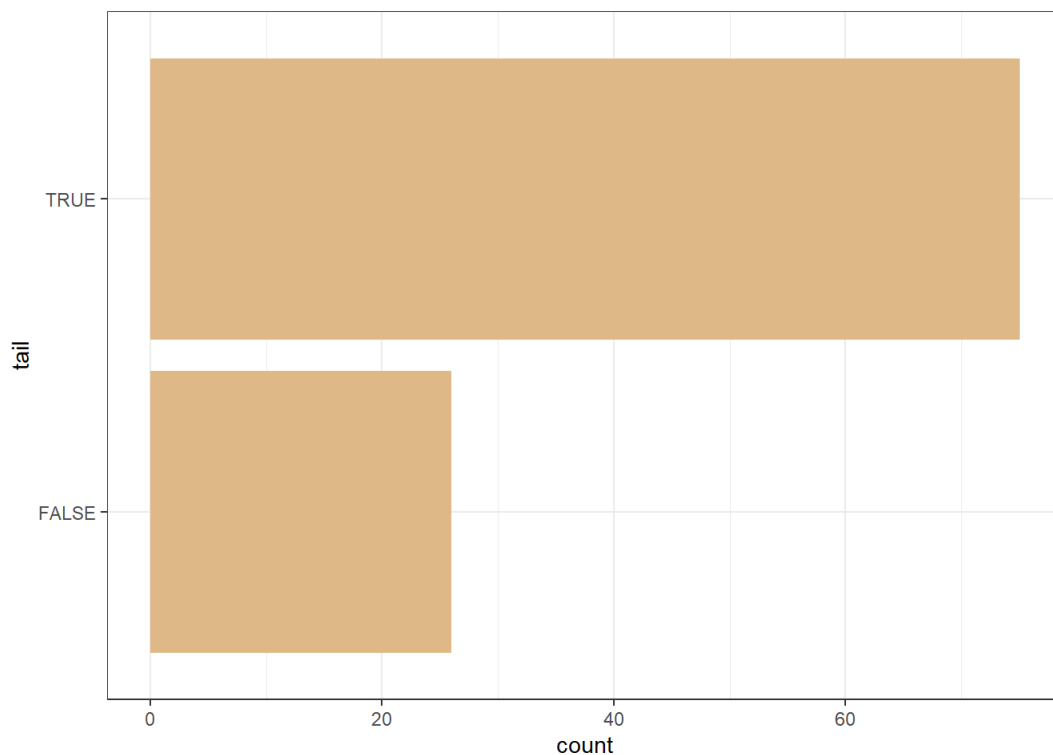
```



Spesific1

```
# we can see that in the output which animals have tail [TRUE options says the animal has tail]
tapply(rownames(Zoo), Zoo$tail, function(x) x); ggplot(Zoo, aes(x=tail)) + geom_bar(fill="burlywood") + theme_bw() + coord_flip()
```

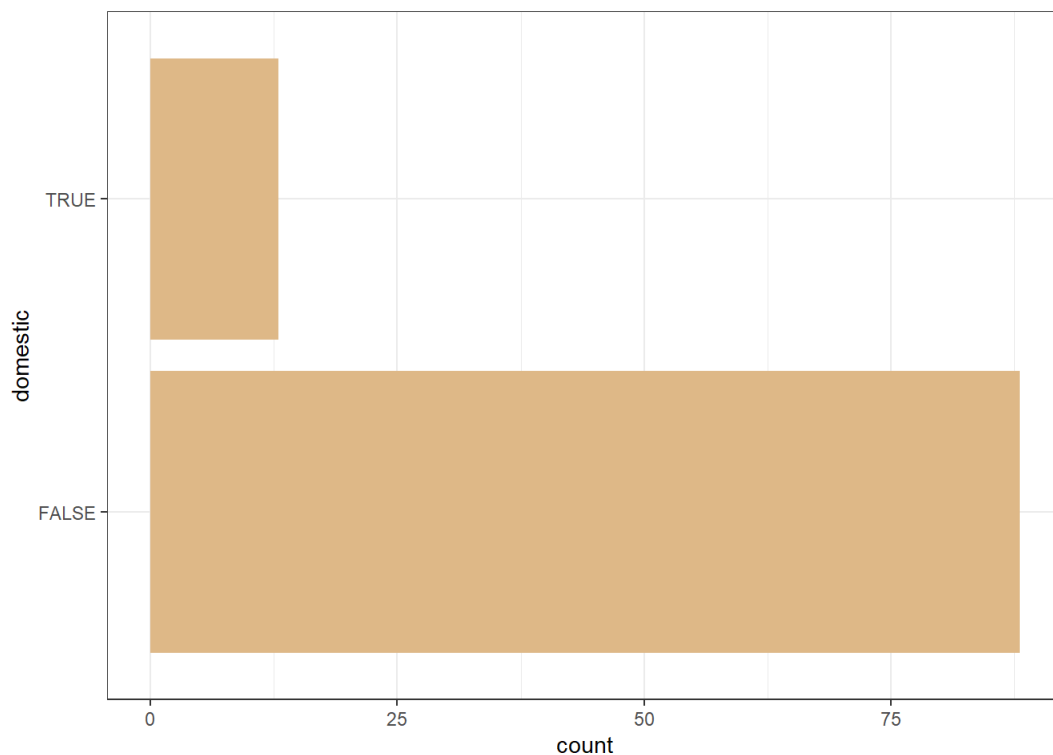
```
## $`FALSE`
## [1] "aardvark" "bear" "cavy" "clam" "crab" "crayfish"
## [7] "flea" "frog.1" "frog.2" "girl" "gnat" "gorilla"
## [13] "honeybee" "housefly" "ladybird" "lobster" "moth" "octopus"
## [19] "seal" "seawasp" "slug" "starfish" "termite" "toad"
## [25] "wasp" "worm"
##
## $`TRUE`
## [1] "antelope" "bass" "boar" "buffalo" "calf" "carp"
## [7] "catfish" "cheetah" "chicken" "chub" "crow" "deer"
## [13] "dogfish" "dolphin" "dove" "duck" "elephant" "flamingo"
## [19] "fruitbat" "giraffe" "goat" "gull" "haddock" "hamster"
## [25] "hare" "hawk" "herring" "kiwi" "lark" "leopard"
## [31] "lion" "lynx" "mink" "mole" "mongoose" "newt"
## [37] "opossum" "oryx" "ostrich" "parakeet" "penguin" "pheasant"
## [43] "pike" "piranha" "pitviper" "platypus" "polecat" "pony"
## [49] "porpoise" "puma" "pussycat" "raccoon" "reindeer" "rhea"
## [55] "scorpion" "seahorse" "sealion" "seasnake" "skimmer" "skua"
## [61] "slowworm" "sole" "sparrow" "squirrel" "stingray" "swan"
## [67] "tortoise" "tuatara" "tuna" "vampire" "vole" "vulture"
## [73] "wallaby" "wolf" "wren"
```



Spesific2

```
# we can see that which animals are domestic.[TRUE options says the animal is domestic]
tapply(rownames(Zoo), Zoo$domestic, function(x) x) ; ggplot(Zoo, aes(x=domestic)) + geom_bar(fill="burlywood")
) + theme_bw() + coord_flip()
```

```
## $`FALSE`
## [1] "aardvark" "antelope" "bass"      "bear"      "boar"      "buffalo"
## [7] "catfish"  "cheetah"  "chub"      "clam"      "crab"      "crayfish"
## [13] "crow"     "deer"     "dogfish"   "dolphin"   "duck"      "elephant"
## [19] "flamingo" "flea"     "frog.1"    "frog.2"    "fruitbat"  "giraffe"
## [25] "gnat"     "gorilla"  "gull"      "haddock"   "hare"      "hawk"
## [31] "herring"  "housefly" "kiwi"      "ladybird"  "lark"      "leopard"
## [37] "lion"     "lobster"  "lynx"      "mink"      "mole"      "mongoose"
## [43] "moth"     "newt"     "octopus"   "opossum"   "oryx"      "ostrich"
## [49] "penguin"  "pheasant" "pike"      "piranha"   "pitviper"  "platypus"
## [55] "polecat"  "porpoise" "puma"      "raccoon"   "rhea"      "scorpion"
## [61] "seahorse" "seal"     "sealion"   "seasnake"  "seawasp"   "skimmer"
## [67] "skua"     "slowworm" "slug"      "sole"      "sparrow"   "squirrel"
## [73] "starfish" "stingray" "swan"      "termite"   "toad"      "tortoise"
## [79] "tuatara"  "tuna"     "vampire"   "vole"      "vulture"   "wallaby"
## [85] "wasp"     "wolf"     "worm"      "wren"
##
## $`TRUE`
## [1] "calf"      "carp"      "cavy"      "chicken"   "dove"      "girl"
## [7] "goat"     "hamster"   "honeybee"  "parakeet"  "pony"      "pussycat"
## [13] "reindeer"
```



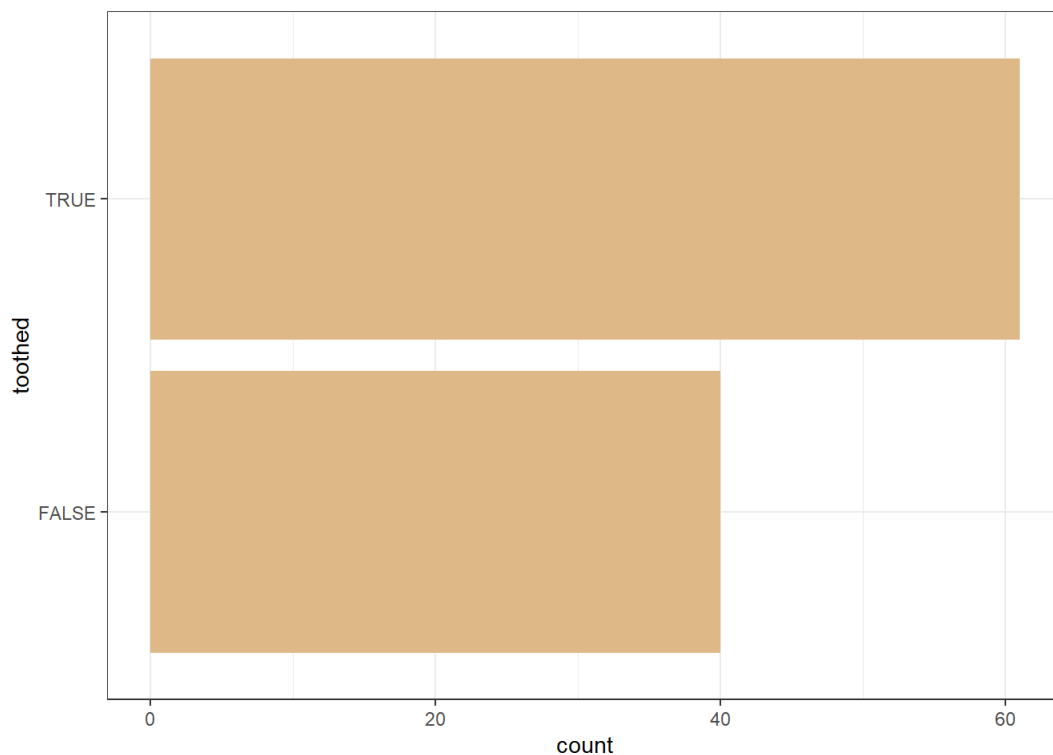
```
# This codes give us directly which animals domestic
rownames(Zoo)[Zoo$domestic]
```

```
## [1] "calf"      "carp"      "cavy"      "chicken"   "dove"      "girl"
## [7] "goat"      "hamster"   "honeybee"  "parakeet"  "pony"      "pussycat"
## [13] "reindeer"
```

Spesific3

```
# we can see that which animals have teeth.[TRUE options says the animal has teeth]
tapply(rownames(Zoo), Zoo$toothed, function(x) x) ; ggplot(Zoo, aes(x=toothed)) + geom_bar(fill="burlywood")
+ theme_bw() + coord_flip()
```

```
## $`FALSE`
## [1] "chicken" "clam"    "crab"    "crayfish" "crow"    "dove"
## [7] "duck"    "flamingo" "flea"    "gnat"     "gull"    "hawk"
## [13] "honeybee" "housefly" "kiwi"    "ladybird" "lark"    "lobster"
## [19] "moth"    "octopus" "ostrich" "parakeet" "penguin" "pheasant"
## [25] "platypus" "rhea"    "scorpion" "seawasp"  "skimmer" "skua"
## [31] "slug"    "sparrow" "starfish" "swan"     "termite" "tortoise"
## [37] "vulture" "wasp"    "worm"    "wren"
##
## $`TRUE`
## [1] "aardvark" "antelope" "bass"     "bear"     "boar"     "buffalo"
## [7] "calf"     "carp"     "catfish"  "cavy"     "cheetah"  "chub"
## [13] "deer"     "dogfish"  "dolphin"  "elephant" "frog.1"   "frog.2"
## [19] "fruitbat" "giraffe"  "girl"     "goat"     "gorilla"  "haddock"
## [25] "hamster"  "hare"     "herring"  "leopard"  "lion"     "lynx"
## [31] "mink"     "mole"     "mongoose" "newt"     "opossum"  "oryx"
## [37] "pike"     "piranha"  "pitviper" "polecat"  "pony"     "porpoise"
## [43] "puma"     "pussycat" "raccoon"  "reindeer" "seahorse" "seal"
## [49] "sealion"  "seasnake" "slowworm" "sole"     "squirrel" "stingray"
## [55] "toad"     "tuatara"  "tuna"     "vampire"  "vole"     "wallaby"
## [61] "wolf"
```



Spesific4

by looking at the output, we can understand which animal is which type of animal. , there is 41 mammal type animals , 20 bird , 5 reptile , 13 fish , 4 amphibian , 8 insect , 10 mollusc.etal animals in the data.

```
tapply(rownames(Zoo), Zoo$type, function(x) x)
```

```
## $mammal
## [1] "aardvark" "antelope" "bear"      "boar"      "buffalo"   "calf"
## [7] "cavy"     "cheetah"  "deer"     "dolphin"   "elephant"  "fruitbat"
## [13] "giraffe"  "girl"     "goat"     "gorilla"   "hamster"   "hare"
## [19] "leopard"  "lion"     "lynx"     "mink"      "mole"      "mongoose"
## [25] "opossum"  "oryx"     "platypus" "polecat"   "pony"      "porpoise"
## [31] "puma"     "pussycat" "raccoon"  "reindeer"  "seal"      "sealion"
## [37] "squirrel" "vampire"  "vole"     "wallaby"   "wolf"
##
## $bird
## [1] "chicken"  "crow"     "dove"     "duck"      "flamingo"  "gull"
## [7] "hawk"     "kiwi"     "lark"     "ostrich"   "parakeet"  "penguin"
## [13] "pheasant" "rhea"     "skimmer"  "skua"      "sparrow"   "swan"
## [19] "vulture"  "wren"
##
## $reptile
## [1] "pitviper" "seasnake" "slowworm" "tortoise"  "tuatara"
##
## $fish
## [1] "bass"     "carp"     "catfish"  "chub"     "dogfish"   "haddock"
## [7] "herring"  "pike"     "piranha"  "seahorse" "sole"      "stingray"
## [13] "tuna"
##
## $amphibian
## [1] "frog.1" "frog.2" "newt"     "toad"
##
## $insect
## [1] "flea"     "gnat"     "honeybee" "housefly"  "ladybird"  "moth"     "termite"
## [8] "wasp"
##
## $mollusc.etal
## [1] "clam"     "crab"     "crayfish" "lobster"   "octopus"   "scorpion"
## [7] "seawasp"  "slug"     "starfish" "worm"
```