Dog size vs intelligence: my R Notebook

Robert Batzinger

13 July 2017

Steps Taken

1. Read in the dog weight - intelligence data

```r
dogs = read.csv(
  "https://query.data.world/s/9zmgo15g5lvopy98cr4vatfxj")
```

2. Group the weights into discrete bins.

```r
wtLevel = 1+ as.integer(dogs$avgWeight /20)
wtLevel[wtLevel > 6] = 6
weightLevels = c("w:000-19","w:020-39",
    "w:040-59","w:060-79","w:080-99","w:100-200")
wtGroup=weightLevels[wtLevel]
```

3. Simplfy the classification labels

```r
Intelligence = rep("",104)
Intelligence[dogs$Classification ==
  "Above Average Working Dogs"] = "IQ3.AboveAvg"
Intelligence[dogs$Classification ==
  "Average Working/Obedience Intelligence"] = "IQ4.Avg"
Intelligence[dogs$Classification ==
  "Brightest Dogs"] = "IQ1.Bright"
Intelligence[dogs$Classification ==
  "Excellent Working Dogs"] = "IQ2.Excel"
Intelligence[dogs$Classification ==
  "Fair Working/Obedience Intelligence"] = "IQ5.Fair"
Intelligence[dogs$Classification ==
  "Lowest Degree of Working/Obedience Intelligence"] = "IQ6.Lowest"
```

4. Update the dataset with 2 new columns

```r
dogs2 = cbind(dogs, Intelligence,wtGroup)
```

5. Create a cross-tab of intelligence vs weight

```r
wtIQ = xtabs(~ Intelligence + wtGroup, data=dogs2)
```

6 Combine rows to reduce the number of empty bins.

```r
wtIQ2 =  wtIQ[1,] + wtIQ[2,]
wtIQ2 = rbind(wtIQ2, wtIQ[3,])
wtIQ2 = rbind(wtIQ2, wtIQ[4,])
wtIQ2 = rbind(wtIQ2, wtIQ[5,] + wtIQ[6,])

rownames(wtIQ2) = c("Hi", "AboveAvg", "Ave", "Low")
```

Analysis

- Use chi square to determine the probability that there is no significant difference in the intelligence profile of each body size: With `p = 0.05`, the chance of all sizes of dogs having the same distribution of intelligence is very, very low.

```
chisq.test(wtIQ2,simulate.p.value=TRUE)
```

```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  wtIQ2
## X-squared = 24.27, df = NA, p-value = 0.05747
```

- Use a mosiac plot to show the normalized profile of intelligence for each size grouping based on the number of breeds in the respective size group.

```
mosaicplot(t(wtIQ2),
  main="Weight vs IQ of dogs ",
  col=c("red","orange","green","blue"),
  xlab="Weight in lbs",
  ylab="Intelligence classification",
  sub="Based on number of breeds for each size category")
```
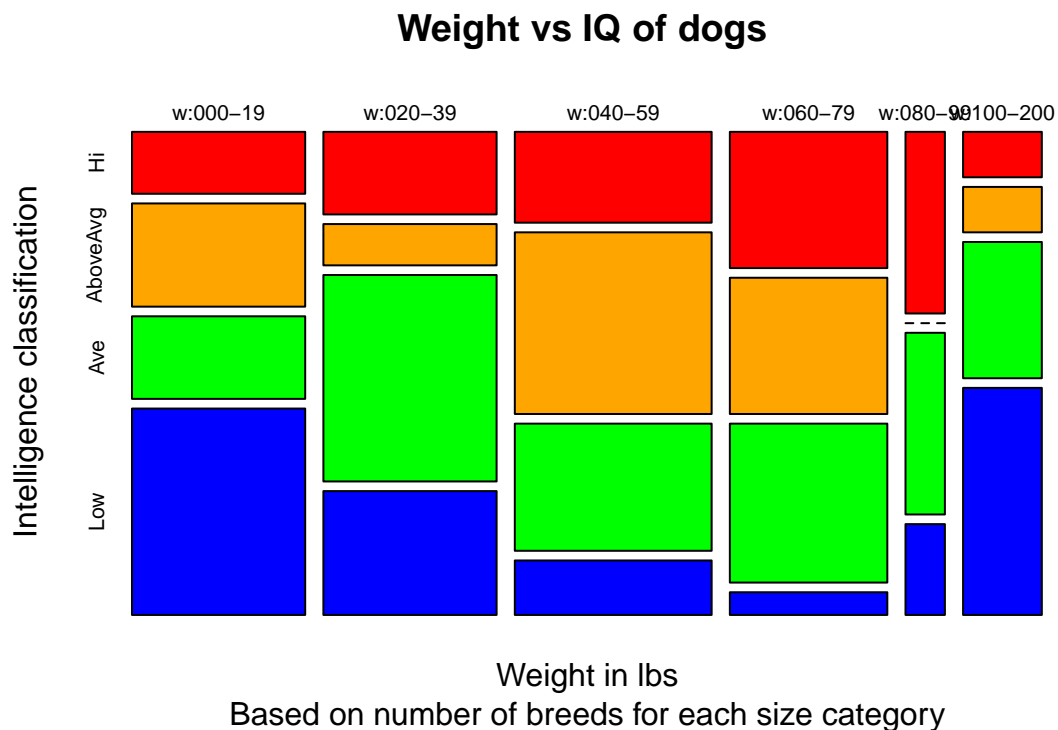
Unanswered questions

- We should look at this question from the point of view that if we were to meet a random dog of a particular size what level of intelligence would we expect knowing only the wieght of the dog. This is similar to asking what level of intelligence can we expect given the gene pool of a dog population.

  This would require taking dog licensing data to estimate the numbers of each breed in order to adjust the previous chart accordingly to the popularity of each breed of dog.

  - It would also be interesting to see the influence of size and/or intelligence on the popularity of dog ownership.