Bivariate EDA - Categorical

R Handout

Derek H. Ogle

## Background

The [General Social Survey](http://www.norc.org/Research/Projects/Pages/general-social-survey.aspx) (GSS) is a nationwide survey that has been administered since 1972 to gather data on contemporary American society in an attempt to monitor and explain trends in attitudes, behaviors, and attributes of American society. One part of that survey asked respondents to state their opinion on how their level of agreement with the following statement -- "Animals have rights too". The results from respondents in 1993 and 1994 are recorded in [AnimalRights.csv](https://github.com/droglenc/NCData/blob/master/AnimalRights.csv). Use the data to answer the questions below.

* What type of variable is each variable in the data set?
* What percentage of respondents were from 1993 and strongly agreed with the statement?
* What percentage of the respondents from 1993 strongly disagreed with the statement?
* What percentage of the respondents that were neutral about the statement were from 1993?

## Getting The Data

> library(NCStats)  
> setwd("C:/aaaWork/Web/GitHub/NCMTH107/modules/BivEDA\_Categorical")  
> d <- read.csv("AnimalRights.csv")  
> str(d)

'data.frame': 2793 obs. of 2 variables:  
 $ year : int 1993 1993 1993 1993 1993 1993 1993 1993 1993 1993 ...  
 $ response: Factor w/ 5 levels "Agree","Disagree",..: 4 4 4 4 4 4 4 4 4 4 ...

> levels(d$response)

[1] "Agree" "Disagree" "Neither" "StrAgree" "StrDis"

> d$response <- factor(d$response,  
 levels=c("StrDis","Disagree","Neither","Agree","StrAgree"))  
> levels(d$response)

[1] "StrDis" "Disagree" "Neither" "Agree" "StrAgree"

## Bivariate EDA -- Categorical

> ( freq.tbl <- xtabs(~year+response,data=d) )

response  
year StrDis Disagree Neither Agree StrAgree  
 1993 196 520 294 372 95  
 1994 247 442 272 288 67

> percTable(freq.tbl,margin=1,digits=2)

response  
year StrDis Disagree Neither Agree StrAgree Sum  
 1993 13.27 35.21 19.91 25.19 6.43 100.01  
 1994 18.77 33.59 20.67 21.88 5.09 100.00

> percTable(freq.tbl,margin=2,digits=2)

response  
year StrDis Disagree Neither Agree StrAgree  
 1993 44.24 54.05 51.94 56.36 58.64  
 1994 55.76 45.95 48.06 43.64 41.36  
 Sum 100.00 100.00 100.00 100.00 100.00

> percTable(freq.tbl,digits=2)

response  
year StrDis Disagree Neither Agree StrAgree Sum  
 1993 7.02 18.62 10.53 13.32 3.40 52.89  
 1994 8.84 15.83 9.74 10.31 2.40 47.12  
 Sum 15.86 34.45 20.27 23.63 5.80 100.01

> barplot(freq.tbl,beside=TRUE,legend=TRUE,ylab="Frequency")

