Dataset report of a burrow data file

Suzanne Fox

February 2017

## Dataset details

Source data name - testdata  
Source data description - test data csv  
Burrow data created - Sun Feb 05 2017, 13:44:22  
Dataset dimensions - 36 columns (variables) and 500 rows (cases)

## Best Guess at datatype for each variable

Variable1 myData1  
 ID NUMBER  
 first\_name TEXT 95% UNIQUE  
 last\_name TEXT 98% UNIQUE  
 company\_name TEXT 99% UNIQUE  
 address TEXT 99% UNIQUE  
 city TEXT 92% UNIQUE  
 county TEXT 21% UNIQUE  
 postal GEO\_POSTCODE  
 phone1 MAINLY NUMERIC 98% UNIQUE  
 phone2 MAINLY NUMERIC 99% UNIQUE  
 email TEXT 99% UNIQUE  
 web WEB\_URL  
 Household.size CATEGORICAL  
 Likely.to.buy SCALE\_210120  
 likely.to.recommend SCALE\_210120  
 Likely.to.visit SCALE\_210120  
 score1 NUMBER  
 score2 NUMBER  
 score3 NUMBER  
 VisitDate DATE\_DDMMYYYY\_DASHES  
 Tables BIT\_ONEZERO  
 Chairs BIT\_ONEZERO  
 Cabinets BIT\_ONEZERO  
 Beds BIT\_ONEZERO  
 Sofas BIT\_ONEZERO  
 Sideboards BIT\_ONEZERO  
 Kitchens BIT\_ONEZERO  
 Salary CURRENCY\_UK  
 Annual.expenditure CURRENCY\_UK  
 Dist1 NUMBER  
 Dist2 NUMBER  
 Dist3 DECIMALS  
 Dist4 DECIMALS  
 Increase PERCENT  
 Holiday.Cash CURRENCY\_USA  
 Recontact BIT\_YESNO

# print(burrow.report)

# # Visualise Uniques

# p1 <- ggplot(data=burrow.report,

# aes(x=VariableName, y=Unique)) +

# geom\_bar(colour="black", fill="#4C1E9C",

# width=.8, stat="identity") +

# guides(fill=FALSE) +

# xlab("Variables") + ylab("% of Total") +

# ggtitle("Unique Values") +

# ylim(0, 100) +

# theme(panel.grid.major = element\_blank(),

# panel.grid.minor = element\_blank(),

# panel.background = element\_blank(),

# axis.line=element\_blank(),

# axis.text.x=element\_blank(),

# axis.text.y=element\_blank(),

# axis.ticks=element\_blank(),

# legend.position="none")

# 

# 

# # Visualise Missing

# p2 <- ggplot(data=burrow.report,

# aes(x=VariableName, y=Missing)) +

# geom\_bar(colour="black", fill="#4C1E9C",

# width=.8, stat="identity") +

# guides(fill=FALSE) +

# xlab("Variables") + ylab("% of Total") +

# ggtitle("Missing Values") +

# ylim(0, 100) +

# theme(panel.grid.major = element\_blank(),

# panel.grid.minor = element\_blank(),

# panel.background = element\_blank(),

# axis.line=element\_blank(),

# axis.text.x=element\_blank(),

# axis.text.y=element\_blank(),

# axis.ticks=element\_blank(),

# legend.position="none")

# 

# # plot uniques and missings

# plot\_grid(p1, p2, align='h')

# 

# if (sum(burrow.report$Missing)==0) {

# print("There were no missing values for any variable")

# } else {

# print("Dealing with missing data")

# }

# 

# #plotvar <- "mpg"

# #myPlot\_Int(mtcars,plotvar)

# 

# num.histograms <- nrow(burrow.plothist)

# print(paste("There are",num.histograms,"numeric fields."))

# 

# for (i in seq\_len(num.histograms)) {

# plotvar <- burrow.plothist[i,c("VariableName")]

# print(paste("Plotting",plotvar))

# 

# possibleError <- tryCatch(

# myPlot <- myPlot\_Int(burrow.inputdata,plotvar),

# error=function(e) e

# )

# 

# if(inherits(possibleError, "error")) {

# print(paste("error plotting",plotvar))

# next

# }

# 

# print(myPlot)

# }