Dataset report

Suzanne Fox

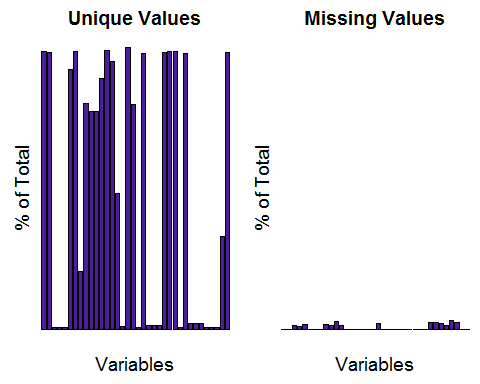
27 November 2016

## Dataset analysis

## Dataset Structure

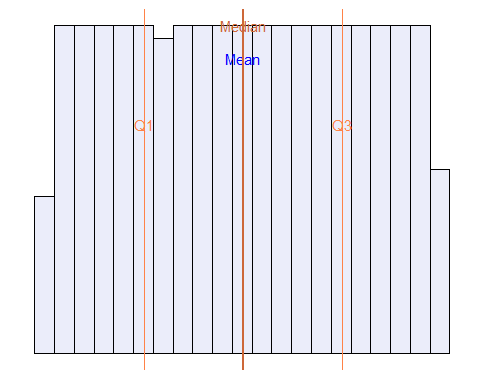
## [1] "This dataset has 36 columns (variables) and 500 rows (cases)"

## VariableName Missing Unique BEST\_GUESS  
## 1 ID 0.0 100.0 NUMBER  
## 2 first\_name 0.0 95.0 TEXT 95% UNIQUE  
## 3 last\_name 0.0 98.0 TEXT 98% UNIQUE  
## 4 company\_name 0.0 98.6 TEXT 99% UNIQUE  
## 5 address 0.0 98.8 TEXT 99% UNIQUE  
## 6 city 0.0 92.4 TEXT 92% UNIQUE  
## 7 county 0.0 20.6 TEXT 21% UNIQUE  
## 8 postal 0.0 98.8 GEO\_POSTCODE  
## 9 phone1 0.0 98.4 MAINLY NUMERIC 98% UNIQUE  
## 10 phone2 0.0 98.8 MAINLY NUMERIC 99% UNIQUE  
## 11 email 0.0 99.0 TEXT 99% UNIQUE  
## 12 web 0.0 98.2 WEB\_URL  
## 13 Household.size 0.0 0.8 CATEGORICAL  
## 14 Likely.to.buy 0.0 1.4 SCALE\_210120  
## 15 likely.to.recommend 0.0 1.4 SCALE\_210120  
## 16 Likely.to.visit 0.0 1.4 SCALE\_210120  
## 17 score1 2.2 2.0 NUMBER  
## 18 score2 2.4 2.0 NUMBER  
## 19 score3 1.8 2.0 NUMBER  
## 20 VisitDate 0.0 32.8 DATE\_DDMMYYYY\_DASHES  
## 21 Tables 2.4 0.4 BIT\_ONEZERO  
## 22 Chairs 1.6 0.4 BIT\_ONEZERO  
## 23 Cabinets 1.0 0.4 BIT\_ONEZERO  
## 24 Beds 1.4 0.4 BIT\_ONEZERO  
## 25 Sofas 3.2 0.4 BIT\_ONEZERO  
## 26 Sideboards 1.2 0.4 BIT\_ONEZERO  
## 27 Kitchens 2.0 0.4 BIT\_ONEZERO  
## 28 Salary 0.0 98.0 CURRENCY\_UK  
## 29 Annual.expenditure 0.0 98.4 CURRENCY\_UK  
## 30 Dist1 1.6 80.2 NUMBER  
## 31 Dist2 1.4 77.4 NUMBER  
## 32 Dist3 2.6 77.2 DECIMALS  
## 33 Dist4 1.2 89.0 DECIMALS  
## 34 Increase 0.0 79.8 PERCENT  
## 35 Holiday.Cash 0.0 48.0 CURRENCY\_USA  
## 36 Recontact 0.0 0.6 BIT\_YESNO

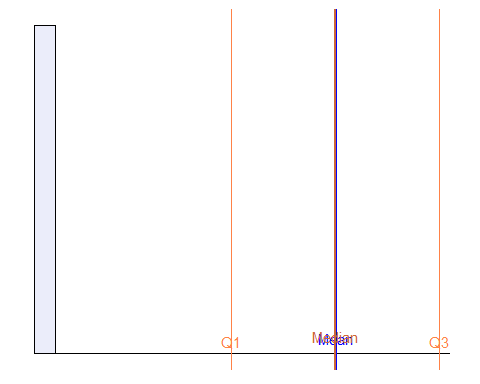


## [1] "Dealing with missing data"

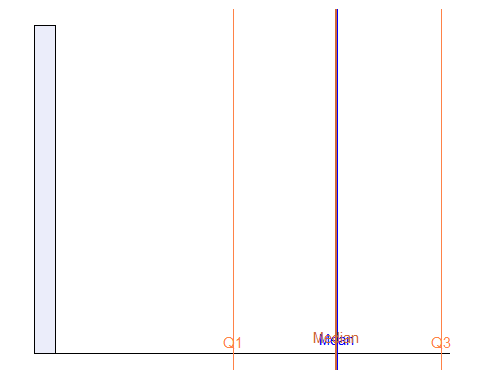
## [1] "Plotting ID"



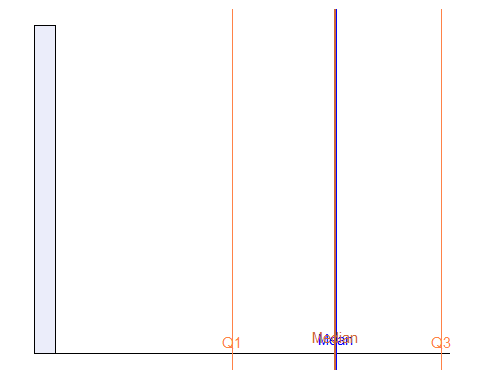
## [1] "Plotting score1"



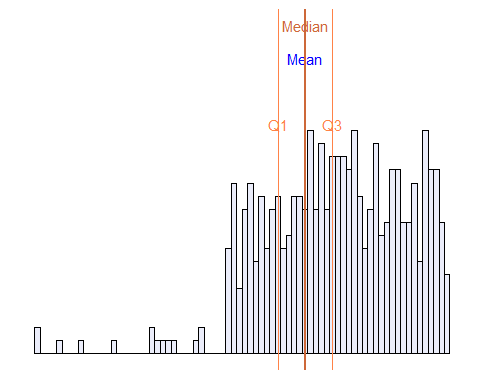
## [1] "Plotting score2"



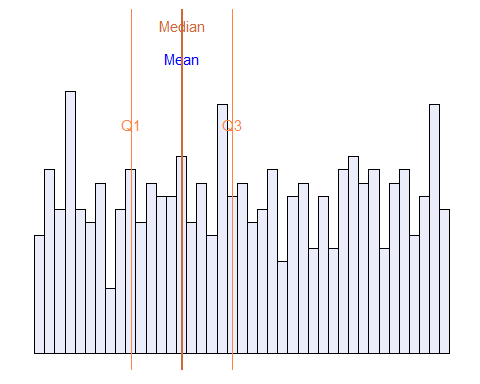
## [1] "Plotting score3"



## [1] "Plotting Dist1"



## [1] "Plotting Dist2"

 ## Variable detail

Best guess for the datatypes of each variable is

## COLUMN\_NAME BEST\_GUESS  
## 1 ID NUMBER  
## 2 first\_name TEXT 95% UNIQUE  
## 3 last\_name TEXT 98% UNIQUE  
## 4 company\_name TEXT 99% UNIQUE  
## 5 address TEXT 99% UNIQUE  
## 6 city TEXT 92% UNIQUE  
## 7 county TEXT 21% UNIQUE  
## 8 postal GEO\_POSTCODE  
## 9 phone1 MAINLY NUMERIC 98% UNIQUE  
## 10 phone2 MAINLY NUMERIC 99% UNIQUE  
## 11 email TEXT 99% UNIQUE  
## 12 web WEB\_URL  
## 13 Household.size CATEGORICAL  
## 14 Likely.to.buy SCALE\_210120  
## 15 likely.to.recommend SCALE\_210120  
## 16 Likely.to.visit SCALE\_210120  
## 17 score1 NUMBER  
## 18 score2 NUMBER  
## 19 score3 NUMBER  
## 20 VisitDate DATE\_DDMMYYYY\_DASHES  
## 21 Tables BIT\_ONEZERO  
## 22 Chairs BIT\_ONEZERO  
## 23 Cabinets BIT\_ONEZERO  
## 24 Beds BIT\_ONEZERO  
## 25 Sofas BIT\_ONEZERO  
## 26 Sideboards BIT\_ONEZERO  
## 27 Kitchens BIT\_ONEZERO  
## 28 Salary CURRENCY\_UK  
## 29 Annual.expenditure CURRENCY\_UK  
## 30 Dist1 NUMBER  
## 31 Dist2 NUMBER  
## 32 Dist3 DECIMALS  
## 33 Dist4 DECIMALS  
## 34 Increase PERCENT  
## 35 Holiday.Cash CURRENCY\_USA  
## 36 Recontact BIT\_YESNO  
## GUESS\_SOURCE  
## 1 REGEX MATCH  
## 2 REGEX ALPHAS 90% MATCH  
## 3 REGEX ALPHAS 90% MATCH  
## 4 REGEX ALPHAS 90% MATCH  
## 5 REGEX ALPHAS 90% MATCH  
## 6 REGEX ALPHAS 90% MATCH  
## 7 REGEX ALPHAS 90% MATCH  
## 8 REGEX MATCH  
## 9 REGEX DIGITS 95% MATCH BUT MAY NOT BE NUMBER FIELD  
## 10 REGEX DIGITS 95% MATCH BUT MAY NOT BE NUMBER FIELD  
## 11 REGEX ALPHAS 90% MATCH  
## 12 REGEX MATCH  
## 13 REGEX ALPHAS 95% MATCH AND < 11 UNIQUES  
## 14 REGEX MATCH  
## 15 REGEX MATCH  
## 16 REGEX MATCH  
## 17 REGEX MATCH  
## 18 REGEX MATCH  
## 19 REGEX MATCH  
## 20 REGEX MATCH  
## 21 REGEX MATCH  
## 22 REGEX MATCH  
## 23 REGEX MATCH  
## 24 REGEX MATCH  
## 25 REGEX MATCH  
## 26 REGEX MATCH  
## 27 REGEX MATCH  
## 28 REGEX MATCH  
## 29 REGEX MATCH  
## 30 REGEX MATCH  
## 31 REGEX MATCH  
## 32 MULTIPLE CONDITIONS REGEX MATCH  
## 33 MULTIPLE CONDITIONS REGEX MATCH  
## 34 REGEX MATCH  
## 35 REGEX MATCH  
## 36 REGEX MATCH