Dataset report of a burrow data file

Suzanne Fox

February 2017

## Run the Burrow function to analyse the source data

## [1] "BURROW ... Run started : Tue Feb 07 2017, 10:07:10"  
## [1] "BURROW ... Analysing dataset"  
## [1] "... write\_Datasetinfo ... Writing dataset details"  
## [1] "... write\_Datasetinfo ... Finished."  
## [1] "BURROW ... Calculating correlations"  
## [1] "BURROW ... Analysing field : ID"  
## [1] "BURROW ... Adding Correlation : ID"  
## [1] "BURROW ... Best Guess is : NUMBER"  
## [1] "BURROW ... Analysing field : first\_name"  
## [1] "BURROW ... Best Guess is : TEXT 95% UNIQUE"  
## [1] "BURROW ... Analysing field : last\_name"  
## [1] "BURROW ... Best Guess is : TEXT 98% UNIQUE"  
## [1] "BURROW ... Analysing field : company\_name"  
## [1] "BURROW ... Best Guess is : TEXT 99% UNIQUE"  
## [1] "BURROW ... Analysing field : address"  
## [1] "BURROW ... Best Guess is : TEXT 99% UNIQUE"  
## [1] "BURROW ... Analysing field : city"  
## [1] "BURROW ... Best Guess is : TEXT 92% UNIQUE"  
## [1] "BURROW ... Analysing field : county"  
## [1] "BURROW ... Best Guess is : TEXT 21% UNIQUE"  
## [1] "BURROW ... Analysing field : postal"  
## [1] "BURROW ... Best Guess is : GEO\_POSTCODE"  
## [1] "BURROW ... Analysing field : phone1"  
## [1] "BURROW ... Best Guess is : MAINLY NUMERIC 98% UNIQUE"  
## [1] "BURROW ... Analysing field : phone2"  
## [1] "BURROW ... Best Guess is : MAINLY NUMERIC 99% UNIQUE"  
## [1] "BURROW ... Analysing field : email"  
## [1] "BURROW ... Best Guess is : TEXT 99% UNIQUE"  
## [1] "BURROW ... Analysing field : web"  
## [1] "BURROW ... Best Guess is : WEB\_URL"  
## [1] "BURROW ... Analysing field : Household.size"  
## [1] "BURROW ... Best Guess is : CATEGORICAL"  
## [1] "BURROW ... Analysing field : Likely.to.buy"  
## [1] "BURROW ... Best Guess is : SCALE\_210120"  
## [1] "BURROW ... Analysing field : likely.to.recommend"  
## [1] "BURROW ... Best Guess is : SCALE\_210120"  
## [1] "BURROW ... Analysing field : Likely.to.visit"  
## [1] "BURROW ... Best Guess is : SCALE\_210120"  
## [1] "BURROW ... Analysing field : score1"  
## [1] "BURROW ... Adding Correlation : score1"  
## [1] "BURROW ... Best Guess is : NUMBER"  
## [1] "BURROW ... Analysing field : score2"  
## [1] "BURROW ... Adding Correlation : score2"  
## [1] "BURROW ... Best Guess is : NUMBER"  
## [1] "BURROW ... Analysing field : score3"  
## [1] "BURROW ... Adding Correlation : score3"  
## [1] "BURROW ... Best Guess is : NUMBER"  
## [1] "BURROW ... Analysing field : VisitDate"  
## [1] "BURROW ... Best Guess is : DATE\_DDMMYYYY\_DASHES"  
## [1] "BURROW ... Analysing field : Tables"  
## [1] "BURROW ... Adding Correlation : Tables"  
## [1] "BURROW ... Best Guess is : BIT\_ONEZERO"  
## [1] "BURROW ... Analysing field : Chairs"  
## [1] "BURROW ... Adding Correlation : Chairs"  
## [1] "BURROW ... Best Guess is : BIT\_ONEZERO"  
## [1] "BURROW ... Analysing field : Cabinets"  
## [1] "BURROW ... Adding Correlation : Cabinets"  
## [1] "BURROW ... Best Guess is : BIT\_ONEZERO"  
## [1] "BURROW ... Analysing field : Beds"  
## [1] "BURROW ... Adding Correlation : Beds"  
## [1] "BURROW ... Best Guess is : BIT\_ONEZERO"  
## [1] "BURROW ... Analysing field : Sofas"  
## [1] "BURROW ... Adding Correlation : Sofas"  
## [1] "BURROW ... Best Guess is : BIT\_ONEZERO"  
## [1] "BURROW ... Analysing field : Sideboards"  
## [1] "BURROW ... Adding Correlation : Sideboards"  
## [1] "BURROW ... Best Guess is : BIT\_ONEZERO"  
## [1] "BURROW ... Analysing field : Kitchens"  
## [1] "BURROW ... Adding Correlation : Kitchens"  
## [1] "BURROW ... Best Guess is : BIT\_ONEZERO"  
## [1] "BURROW ... Analysing field : Salary"  
## [1] "BURROW ... Best Guess is : CURRENCY\_UK"  
## [1] "BURROW ... Analysing field : Annual.expenditure"  
## [1] "BURROW ... Best Guess is : CURRENCY\_UK"  
## [1] "BURROW ... Analysing field : Dist1"  
## [1] "BURROW ... Adding Correlation : Dist1"  
## [1] "BURROW ... Best Guess is : NUMBER"  
## [1] "BURROW ... Analysing field : Dist2"  
## [1] "BURROW ... Adding Correlation : Dist2"  
## [1] "BURROW ... Best Guess is : NUMBER"  
## [1] "BURROW ... Analysing field : Dist3"  
## [1] "BURROW ... Adding Correlation : Dist3"  
## [1] "BURROW ... Best Guess is : DECIMALS"  
## [1] "BURROW ... Analysing field : Dist4"  
## [1] "BURROW ... Adding Correlation : Dist4"  
## [1] "BURROW ... Best Guess is : DECIMALS"  
## [1] "BURROW ... Analysing field : Increase"  
## [1] "BURROW ... Best Guess is : PERCENT"  
## [1] "BURROW ... Analysing field : Holiday.Cash"  
## [1] "BURROW ... Best Guess is : CURRENCY\_USA"  
## [1] "BURROW ... Analysing field : Recontact"  
## [1] "BURROW ... Best Guess is : BIT\_YESNO"  
## [1] "BURROW ... Finished. Processing time : 2.9"

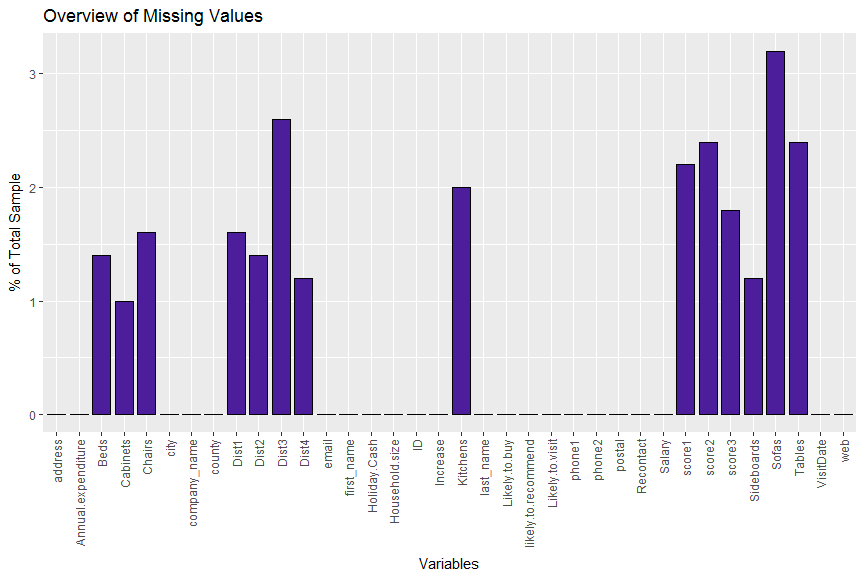
## Dataset details

Source data name - Data.source  
Source data description - synthetic test dataset  
Burrow data created - Tue Feb 07 2017, 10:07:10  
Dataset dimensions - 36 columns (variables) and 500 rows (cases)

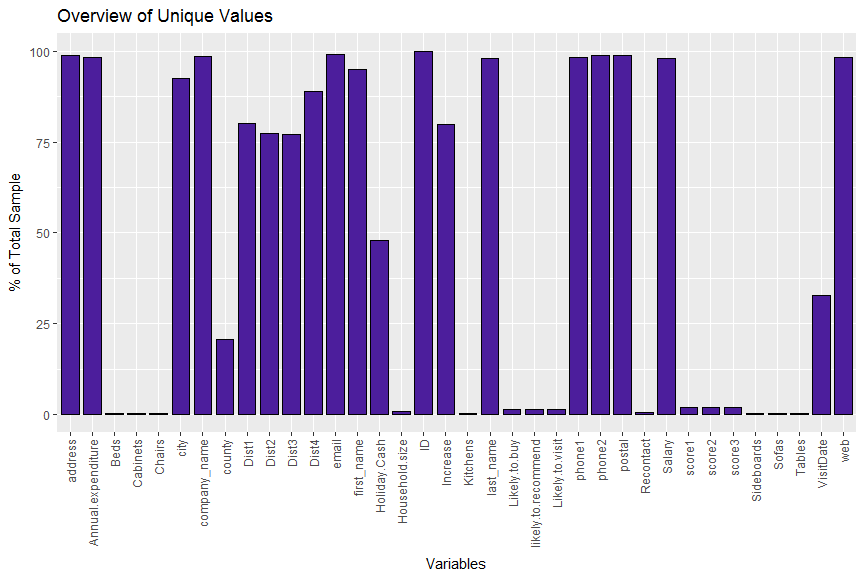
## Best Guess at datatype for each variable

Variable Datatype  
 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

## Variable analysis - level of missing values

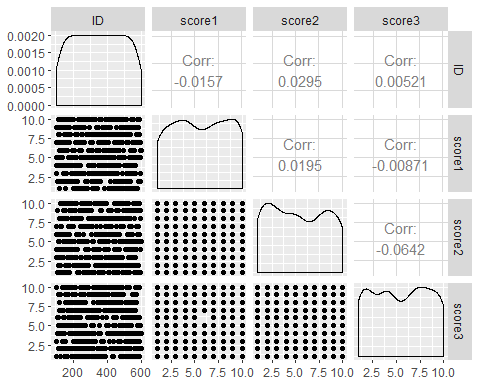


## Variable analysis - level of unique values

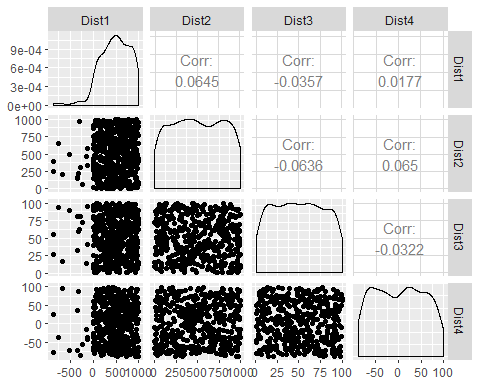


## Variable analysis - correlations

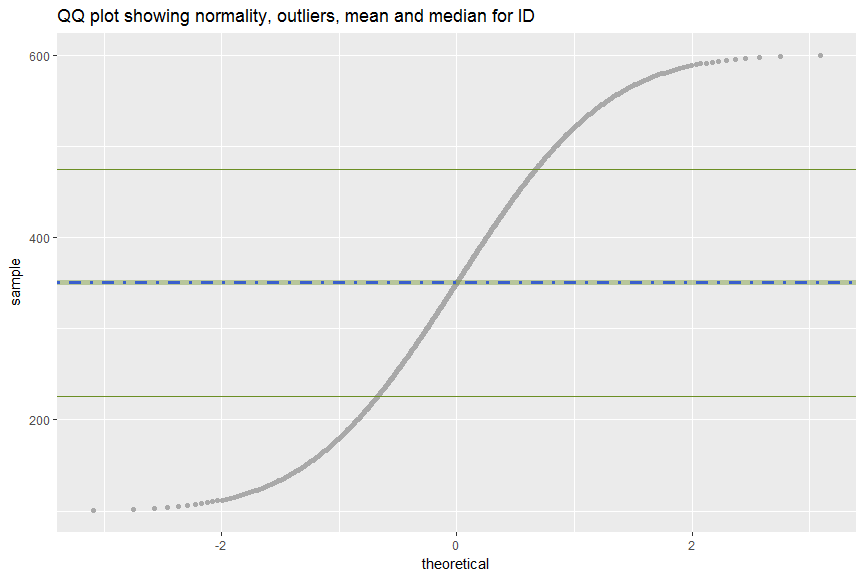
Warning in (function (data, mapping, alignPercent = 0.6, method =  
"pearson", : Removed 11 rows containing missing values  
Warning in (function (data, mapping, alignPercent = 0.6, method =  
"pearson", : Removed 12 rows containing missing values  
Warning in (function (data, mapping, alignPercent = 0.6, method =  
"pearson", : Removed 9 rows containing missing values  
Warning: Removed 11 rows containing missing values (geom\_point).  
Warning: Removed 11 rows containing non-finite values (stat\_density).  
Warning in (function (data, mapping, alignPercent = 0.6, method =  
"pearson", : Removed 23 rows containing missing values  
Warning in (function (data, mapping, alignPercent = 0.6, method =  
"pearson", : Removed 20 rows containing missing values  
Warning: Removed 12 rows containing missing values (geom\_point).  
Warning: Removed 23 rows containing missing values (geom\_point).  
Warning: Removed 12 rows containing non-finite values (stat\_density).  
Warning in (function (data, mapping, alignPercent = 0.6, method =  
"pearson", : Removed 20 rows containing missing values  
Warning: Removed 9 rows containing missing values (geom\_point).  
Warning: Removed 20 rows containing missing values (geom\_point).  
  
Warning: Removed 20 rows containing missing values (geom\_point).  
Warning: Removed 9 rows containing non-finite values (stat\_density).



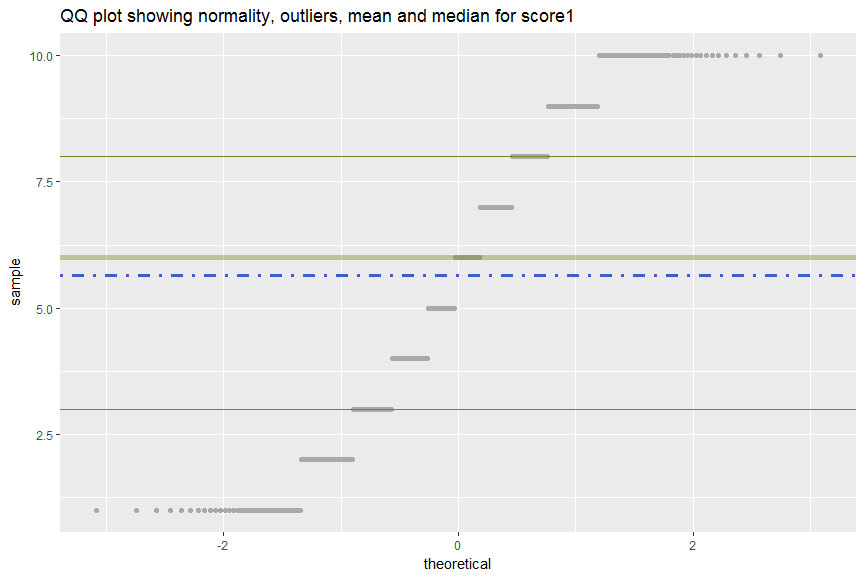
Warning: Removed 8 rows containing non-finite values (stat\_density).  
Warning in (function (data, mapping, alignPercent = 0.6, method =  
"pearson", : Removed 15 rows containing missing values  
Warning in (function (data, mapping, alignPercent = 0.6, method =  
"pearson", : Removed 21 rows containing missing values  
Warning in (function (data, mapping, alignPercent = 0.6, method =  
"pearson", : Removed 14 rows containing missing values  
Warning: Removed 15 rows containing missing values (geom\_point).  
Warning: Removed 7 rows containing non-finite values (stat\_density).  
Warning in (function (data, mapping, alignPercent = 0.6, method =  
"pearson", : Removed 20 rows containing missing values  
Warning in (function (data, mapping, alignPercent = 0.6, method =  
"pearson", : Removed 12 rows containing missing values  
Warning: Removed 21 rows containing missing values (geom\_point).  
Warning: Removed 20 rows containing missing values (geom\_point).  
Warning: Removed 13 rows containing non-finite values (stat\_density).  
Warning in (function (data, mapping, alignPercent = 0.6, method =  
"pearson", : Removed 19 rows containing missing values  
Warning: Removed 14 rows containing missing values (geom\_point).  
Warning: Removed 12 rows containing missing values (geom\_point).  
Warning: Removed 19 rows containing missing values (geom\_point).  
Warning: Removed 6 rows containing non-finite values (stat\_density).

 ## Variable analysis - distributions and outliers for numeric variables

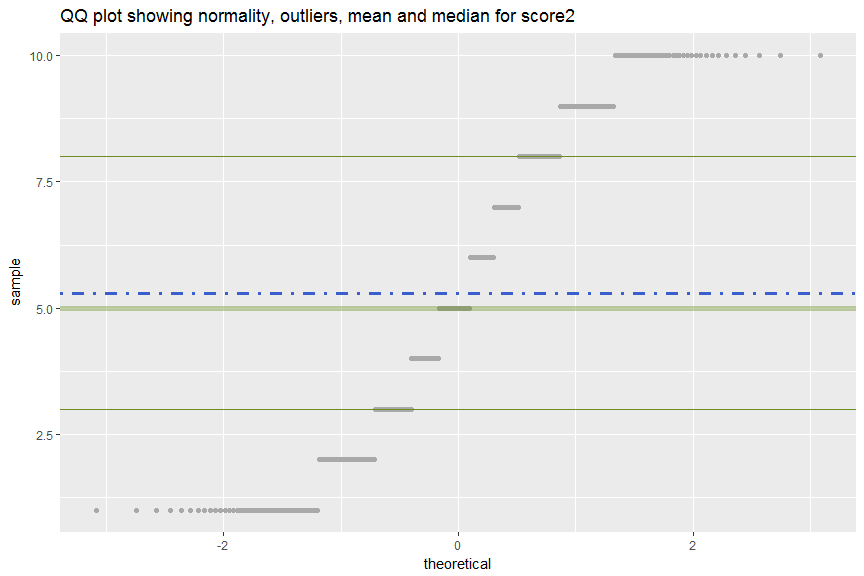
[1] "Charting - ID datatype is NUMBER"



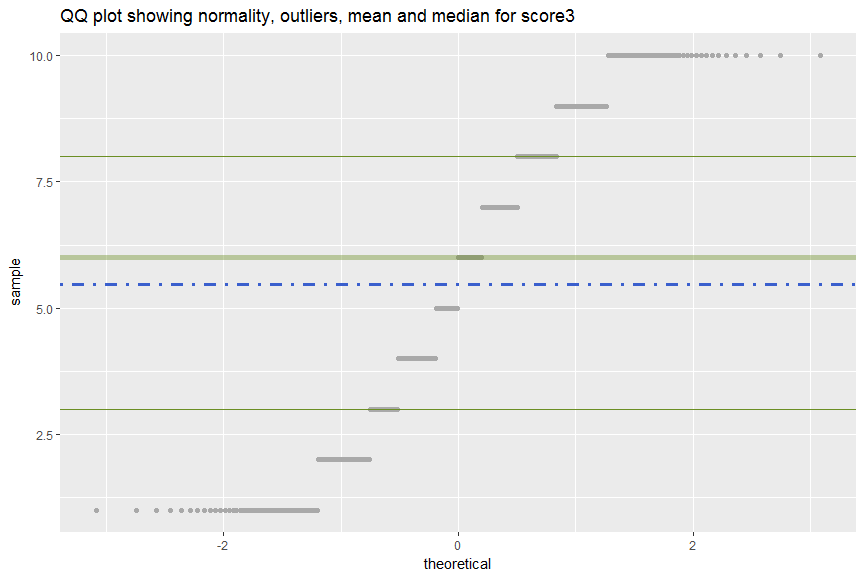
[1] "Charting - score1 datatype is NUMBER"  
Warning: Removed 11 rows containing non-finite values (stat\_qq).



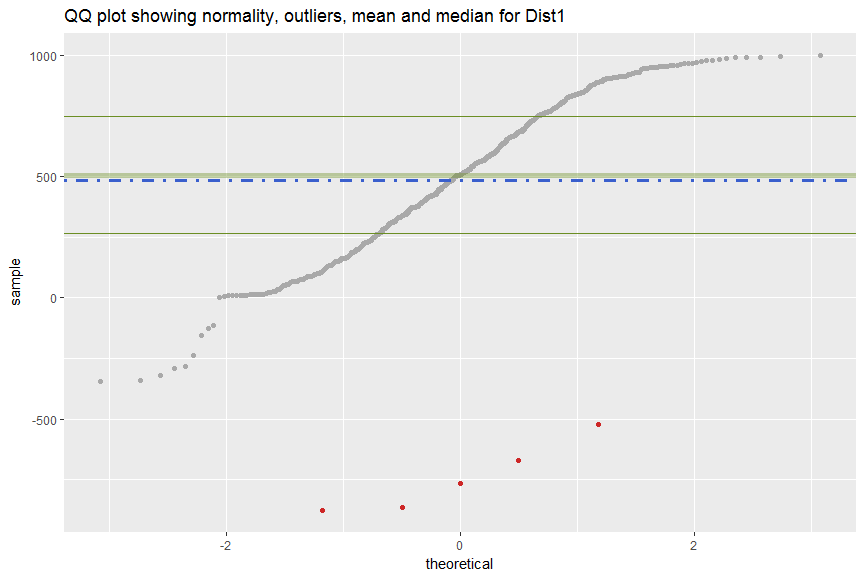
[1] "Charting - score2 datatype is NUMBER"  
Warning: Removed 12 rows containing non-finite values (stat\_qq).



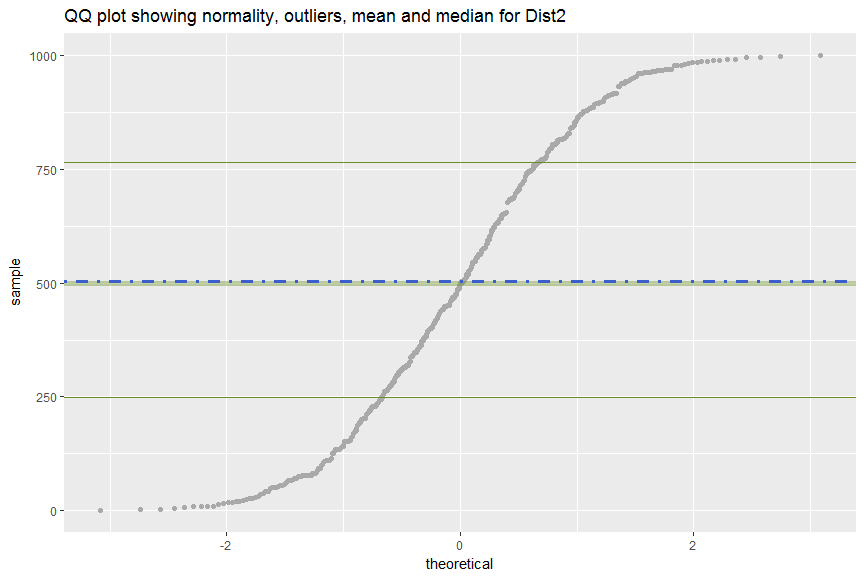
[1] "Charting - score3 datatype is NUMBER"  
Warning: Removed 9 rows containing non-finite values (stat\_qq).



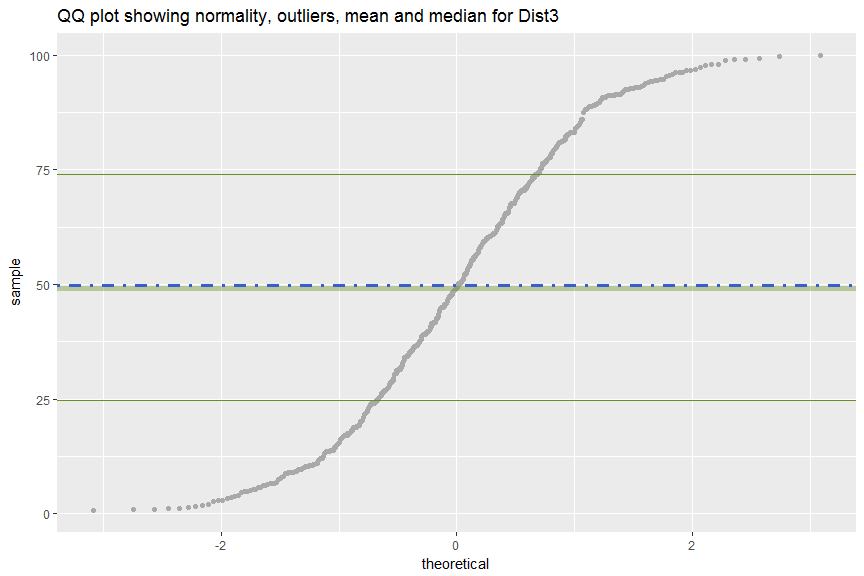
[1] "Charting - Dist1 datatype is NUMBER"  
Warning: Removed 8 rows containing non-finite values (stat\_qq).



[1] "Charting - Dist2 datatype is NUMBER"  
Warning: Removed 7 rows containing non-finite values (stat\_qq).



[1] "Charting - Dist3 datatype is DECIMALS"  
Warning: Removed 13 rows containing non-finite values (stat\_qq).



[1] "Charting - Dist4 datatype is DECIMALS"  
Warning: Removed 6 rows containing non-finite values (stat\_qq).

