**Version:** 1.0

**Date:** 2016-12-20

**Title:** dataprofile()

**Author:**  Jeremy Guinta

**Depends:** gdata, dplyr

December 20, 2016

**Description:** Performs a standard data profile using summary statistics from the summary() function and the structure of the dataset from the str() function. This process returns a data frame object with each column in the data, the data type and summary values for each column. The dataframe object can outputted using write.csv() or similar function to create clean formatted output.

R topics documented:

Contents

[Description 2](#_Toc471390005)

[Usage 2](#_Toc471390006)

[Arguments 2](#_Toc471390007)

[Example 2](#_Toc471390008)

dataprofile() Standard data frame profiling

# Description

This process prepares a standard data profile that shows basic summary statistics for each column along with the structure of the data.

# Usage

dataprofile(dta)

# Arguments

dta a dataframe object

# Example

#Create Data

cll1<-data.frame(

DATE=c("5/11/2010","5/11/2010","5/11/2010","5/11/2010","5/11/2010","5/11/2010","5/11/2010","5/11/2010","5/11/2010","5/11/2010")

,TIME=c("10:58:38","10:58:38","10:58:38","10:59:05","10:59:05","10:59:05","11:00:08","11:00:08","11:00:35", "")

,datetime=c("5/11/2010 10:58:38","5/11/2010 10:58:38","5/11/2010 10:58:38","5/11/2010 10:59:05","5/11/2010 10:59:05","5/11/2010 10:59:05","5/11/2010 11:00:08","5/11/2010 11:00:08","5/11/2010 11:00:08","5/11/2010 11:00:35")

,ANI=c("8003758211","8003758211","8003758211","8003758211","8003758211","8003758211","8003758211","8003758211","8003758211","8003758211")

,DISPOSITION\_CODE=c("No Answer","","Operator Intercept","Left Message","Hangup","No Answer","No Answer","No Answer","No Answer","Dial Error")

,CALL.TYPE=c("Outbound","Outbound","Outbound","Outbound","Outbound"," ","Outbound","Outbound","Outbound","Outbound")

)

dataprofile(cll1)

Var Class st1 Freq1

1 date character factor w/ 1 level "5/11/2010" 1 1 1 1 1 1 1 1 1 1 5/11/2010:10

2 time character factor w/ 5 levels "","105838",.. 2 2 2 3 3 3 4 4 5 1 :1

3 datetime character factor w/ 4 levels "5/11/2010 105838",.. 1 1 1 2 2 2 3 3 3 4 5/11/2010 10:58:38:3

4 ani character factor w/ 1 level "8003758211" 1 1 1 1 1 1 1 1 1 1 8003758211:10

5 disposition\_code character factor w/ 6 levels "","dial error",.. 5 1 6 4 3 5 5 5 5 2 :1

6 call.type character factor w/ 2 levels " ","outbound" 2 2 2 2 2 1 2 2 2 2 :1

Freq2 Freq3 Freq4 Freq5 Freq6

1 <NA> <NA> <NA> <NA> <NA>

2 10:58:38:3 10:59:05:3 11:00:08:2 11:00:35:1 <NA>

3 5/11/2010 10:59:05:3 5/11/2010 11:00:08:3 5/11/2010 11:00:35:1 <NA> <NA>

4 <NA> <NA> <NA> <NA> <NA>

5 dial error :1 hangup :1 left message :1 no answer :5 operator intercept:1

6 outbound:9 <NA> <NA> <NA> <NA>