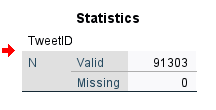
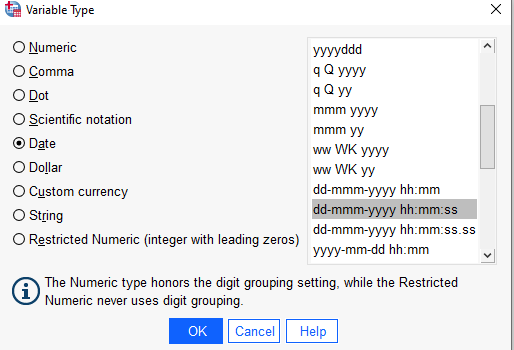
During the Import Task, SPSS should have raised the alert: “Invalid variable names for this application have been found and changed”. Which two variables did this relate to and what were they changed to?

* I noticed that sent- and sent+ got changed to “Sent” and “Sent\_A”.

How many cases were loaded in total?

* Number of cases that was imported onto SPSS = 91303

SPSS initially interprets the DateTime variable as a string type. What pattern did you use to import DateTime correctly?

* I converted the “type” on the variable type to the settings set in the image below
* The reason why I set it to this is show the format 23-Jun-2016 13:00:35 which is equal to dd-mmm-yyyy hh:mm:ss

How many variables has the data dictionary?

* After the import of the data from excel to SPSS, the variable shows 16



1. Variable metadata

Q1: Using the questionnaire-codebook document, determine which SPSS variable name represents the data collected from question: “My tweet has been published in what time zone?”

- When importing the dataset from excel the "Sent+" and "Sent-" variable name contain an illegal character (which is "+" and "-"). So SPSS change it to "Sent" and "Sent\_A" which is not usefull. To combat this issue we can rename "Sent+" to "SentPos" and "Sent-" to "SentNeg"

Q2: What is the size of the sample recorded in this dataset?

- size smaple = 91303

Q3: How many followers has Tweet with id=745607911993270020?

- 89169 followers (userScreenName=pdacosta)

Q4: What is the level of measurement of the variable SentPos?

-The level of measurement is negative

Q5: How many distinct values are associated with the variable SentNeg?

-When analyzing "SentNeg", the minium value= -1 & maximum value= -5 meaning that it contains 5=distinct values

Q6. What is the real name of tweeter who produced the tweet with highest number of Friends?

-Maximum value for "highest number of friends" = 300247, UserScreenName=StewySongs, UserRealName=Stewart Harding

Q7. What is the real author's name of the 10th most popular tweet in the study?

-Retweets=9827 (which is the 10th highest retweet), UserRealName=PaulPopper

Q8. What is the TweetID number of the retweeted post (where the number of retweets is known), for participant in the study located in UTC zone?

Classification

Q9. How many cases belong to the <=20,000 followers' group?

-Selecting the case If condition is satisfied If... UserFollowers <= 20000 and then analysing the data using frequencies. The number of cases that you gain is

N Valid = 87298

Q10. How many cases belong to the 10,000 to 20,000 followers' group?

-Frequency = 5835

Q11. How many cases have no recorded follower data?

-Frequency = 190

Q12: Do users with most of retweets located in UCT or EST zones?

-When UCT=+1 The person with the highest retweet=43361

-When UCT=-4 The person with the highest retweet=35587

-In conclusion UCT has the most retweet based on time zone.

Q13: To one decimal place (1 dp), what is the mean retweeting score for the tweets located in CET zone?

- 20.3

Q14: Amongst tweets located in EST zone, who has the highest number of tweets per week (from Monday to Sunday), on average?

-9.34

Q15: How many tweets per week (from Monday to Sunday) (to 1 dp) do tweeters in CET zone tweet on average?

-3927.4 (1 d.p) tweets on average that week