

DATA Step

How do I stop SAS reading a non-printable EOF character in my raw data file?

Using the INFILE and INPUT statements allow many different forms of raw data files to be read into SAS. Likewise there are many different functions we can use to manipulate and cleanse data once it is in a SAS dataset.

The problem arises however when the raw data itself contains non-printable characters, which can cause the INPUT statement to terminate early as if it had encountered an End-of-File marker.

Consider the following raw data file (addr.txt):

```
Mr Alan Rudland Edinburgh
Mr Paul Dubrow Du&#26;kova
Mr Andrew Smith Glasgow
```

Which contains a non-printable character embedded within the fourth field of the second record.

The code:

```
data address ;
  infile 'addr.txt' dlm = '#' ;
  input addr :$40. ;
run ;
```

treats this non-printable character as an End-of-File marker and stops prematurely. Adding the option IGNOREDOEOF on the INFILE statement allows the character to be ignored, and it can then be cleansed before being output.

The code then becomes:

```
data address ;
  infile 'addr.txt' dlm = '#' ignoredoseof ;
  input addr :$40. ;
run ;
```

To cleanse the field of the non-printable characters, use the COMPRESS function with modifiers to K(eep) only W(riteable) characters:

```
data address ;
```

DATA Step

```
infile 'addr.txt' dlm = '#' ignoredoseof ;  
input addr :$40. ;  
addr = compress(addr, , 'kw') ;  
run ;
```

Unique solution ID: #1068

Author: Alan D Rudland

Last update: 2023-04-18 14:54