

Q|U A|D

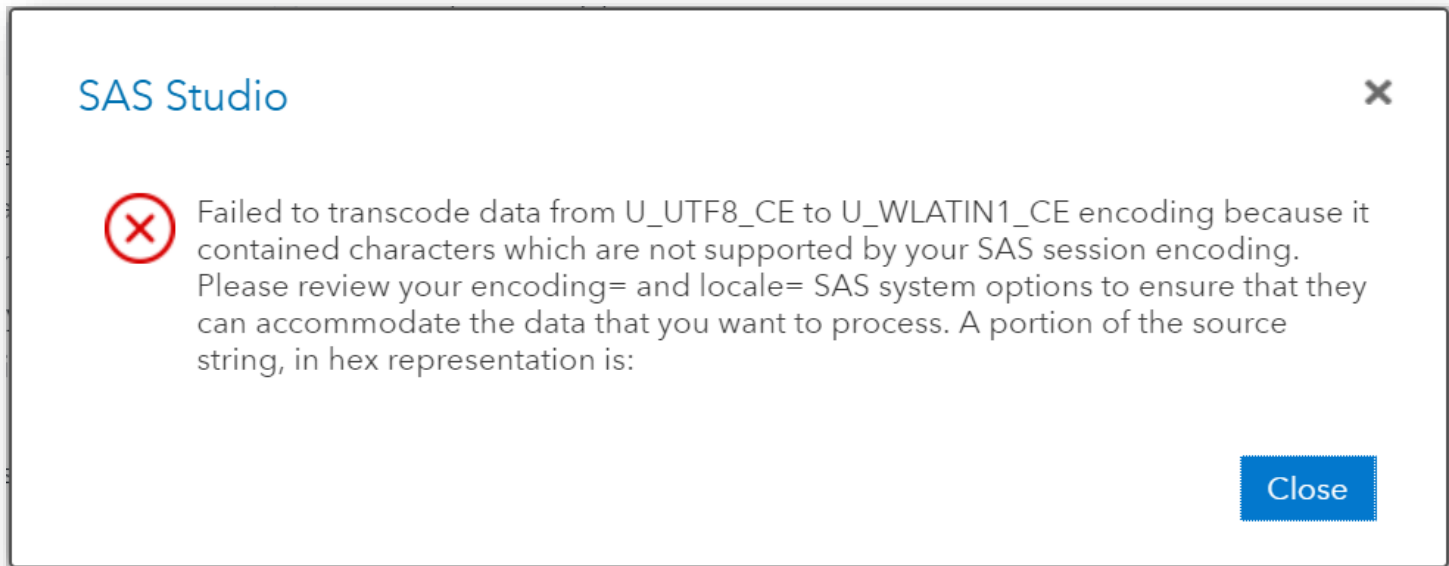
APRIL 28 – MAY 1 | DALLAS, TX

See Your Characters in SAS Studio

Exploring the great unknown



Re store Peace to Your Programs



WLATIN1

	_0	_1	_2	_3	_4	_5	_6	_7	_8	_9	_A	_B	_C	_D	_E	_F
2_	(space)	!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3_	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4_	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5_	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7_	p	q	r	s	t	u	v	w	x	y	z	{		}	~	<u>DEL</u>
8_	€		,	f	„	...	†	‡	^	%o	Š	‹	Œ		Ž	
9_		‘	’	“	”	•	—	—	~	™	š	›	œ		ž	ÿ
A_	<u>NBSP</u>	ı	¢	£	¤	¥	¦	§	¨	©	ª	«	¬	(shy)	®	¯
B_	°	±	²	³	´	µ	¶	·	,	¹	º	»	¼	½	¾	¿
C_	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D_	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
E_	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F_	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

Q | U | A | D

WLATIN1: The ASCII Characters

	_0	_1	_2	_3	_4	_5	_6	_7	_8	_9	_A	_B	_C	_D	_E	_F
2_	(space)	!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3_	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4_	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5_	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7_	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL
8_	€		,	f	„	...	†	‡	^	%o	Š	<	Œ		Ž	
9_		‘	’	“	”	•	—	—	~	™	š	>	œ		ž	Ÿ
A_	NBSP	ı	¢	£	¤	¥	¦	§	¨	©	ª	«	¬	(shy)	®	¯
B_	°	±	²	³	´	µ	¶	·	,	¹	º	»	¼	½	¾	¿
C_	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D_	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
E_	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F_	ö	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

Represented
as **one** byte
in UTF-8

Q | U | A | D

WLATIN1: Not ASCII

	_0	_1	_2	_3	_4	_5	_6	_7	_8	_9	_A	_B	_C	_D	_E	_F
2_	(space)	!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3_	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4_	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5_	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7_	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL
8_	€		,	f	„	...	†	‡	^	%o	Š	◁	Œ		Ž	
9_		‘	’	“	”	•	—	—	~	™	š	▷	œ		ž	Ÿ
A_	NBSP	ı	¢	£	¤	¥	¦	§	¨	©	ª	«	¬	(shy)	®	¯
B_	°	±	²	³	´	µ	¶	·	,	¹	º	»	¼	½	¾	¿
C_	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D_	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
E_	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F_	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

Represented as
two or three
bytes in UTF-8

Q | U | A | D

How UTF-8 characters are different

Hexadecimal representation

	€	1	2	3
WLATIN1	80	31	32	33
UTF-8	E282AC	31	32	33

Q|U|A|D

Suppose to be ASCII

- Keyboard punctuation → typographic characters
 - Hyphen or minus (-) → En dash (–)
 - Apostrophe (') → Quotation marks (' ')
 - Quotation mark (") → Double quotation marks (" ")
- Newline
 - Windows vs. Unix

Q | U | A | D

Looks fine in SAS Studio

Errors in the SAS Log!

- Characters not in SAS Session encoding
- Typographical characters

Q | U | A | D

KPROPDATA

Remove or convert unprintable characters

- Input string
- Output type for unprintable characters
 - UESC – Replace with Unicode escape (\unnnn)
 - BLANK – Replace Space
 - TRIM – Removes
 - *etc*
- Input encoding
- Output encoding

Q | U | A | D

KPROPDATA

```
kpstring = kpropdata(string,  
                      'UESC',  
                      "wlatin1",  
                      "us-ascii");
```

BASECHAR

Convert characters to base type

- Converts characters to the “base” form without accents or other combining characters

Ä → A

Q | U | A | D

