

# Data sets import in SPSS

In this document we explain how to upload different data sets that are used in the course to SPSS. We assume that data sets are in the text format (for CSV format similar procedures are used).

When we open SPSS, we would see the following spreadsheet:

<u> </u>		🛄 🖛										-
-		mental m				14	*, 🖬		s 🛄 👔	14	<b>6</b> 48	
	var	var	var	var	var	var	var	var	var	var	var	var
1												
2												
3												
4												
5												
6												
7												
8												
9												
10												
11												
12												
13												
14												
15												
16												
17												
18												
19												
20												
21												
22												
	4		_	_	_	_	_				_	

This is a **Data View** section. Here each column corresponds to a variable and rows correspond to observations. So in this section the whole data are stored. In the bottom we see two buttons: one is **Data View** and the other is **Variable View**. If we click on **Variable View** button we would see another spreadsheet:

3			з 📕		P	A I		4	2	ा जन	M 45	
	Name	Туре	Width	Decimals	Label	Val	lues	Missing	Columns	Align	Measure	Role
1												
2												
3												
4												
5												
6												
7												
8												
9												
10												
13												
14												
15												
16												
17												
18												
19												
20												
21												
22												
23												
24												
25												
_	4		_	_		_	_		_			

In this section, each row corresponds to a variable. Here we give different characteristics of our variables. The most important for us is 'Name' which is a short name for a variable, 'Type' can be numeric, string, date, etc., 'Decimal' is a number of decimal places shown in the **Data View** section, 'Label' is description of a variable, 'Values' are used for categorical variables and 'Measure' can be scale, ordinal or nominal depending on a kind of a variable.

It is important to distinguish between these two sections. For example if we want to enter some data manually, we first go to **Variable View** give names and characteristics of variables and then click on the **Data View** where we enter the data. When we have a data file in a text format for example then we first upload the data into **Data View** section and then in **Variable View** we give characteristics to variables.

When we do any analysis in SPSS (including graphs, tables, etc.) results will appear in the third window which is called 'Output' window. To return to the **Data View** section just close or

minimize the 'Output' window.

To upload the data set, select  $\mathbf{File} > \mathbf{Open} > \mathbf{Data}$ 

<u>F</u> ile	<u>E</u> dit	View	Data	Transform	Analyze	Di	ect <u>M</u> arketing
N	ew					۶.	Ł
0	pen					۴.	🔁 Data
0	pen Da	ta <u>b</u> ase				۲.	Syntax
🕕 Re	ea <u>d</u> Te	t Data					Output
R	ad Co	gnos Dat	ta				Script

Once 'Open Data' window appears find the file that you want to upload. Make sure that in the 'Files of type:' section, 'All Files' is selected so that we can see all files

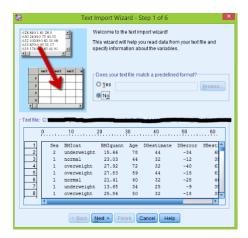
File name:	SkeletonDataComplete.txt	Open
Files of type:	All Files (*.*)	Paste
Encoding:	*	Cancel
		Help
	Retrieve File From Repository	

Once the data file is selected, click on the 'Open' button.

Now we give detailed instructions for 'Skeleton' data set, other data sets will be similar.

#### 'The Skeleton' Data Set

Once you select 'Skeleton' data set in the 'Open Data' window and click 'Open', SPSS opens 'Text Import Wizard' window. Here for 'Does your text file match a predefined format?' question, 'No' should be selected (by default).



Click **Next** and second window opens. For the 'How are your variable arranged?' question select 'Fixed width' since the data file is in text format (if it is in CSV format, select 'Delimited'). If we have variable names in the text data file (headers) then select 'Yes' for the second question, otherwise select 'No'. In the 'Skeleton' file we have headers so we put 'Yes', then click **Next** button.

How are you	ur variables arr	anged?					
O Delimite	d - Variabi	les are deli	mited by a	a specific	character (	i.e., comma, ta	ib).
Eixed wide	tth - Variab	les are aligi	ned in fixe	ed width o	columns.		
Are variable Yes	names include	ed at the top	o of your f	ile?			
© N <u>o</u> Text file: C∣				30	40	50	60
Text file: C;	10 2 underw		5.66	30	40	-34	6D
Text file: C;		eight 1			(5)		
Text file: C	2 underw	eight 1 2	5.66	78	44	-34	60 📤
-Text file: C; 0	2 underw 1 normal	eight 1 2 ight 2	5.66	78 44	44 32	-34 -12	60 <b>4</b> 35
Text file: C1	2 underw 1 normal 1 overwe	eight 1 2 ight 2 ight 2	5.66 3.03 7.92	78 44 72	44 32 32	-34 -12 -40	60 35 61
Text file: C; 0	2 underw 1 normal 1 overwe 1 overwe	eight 1 2 ight 2 ight 2	5.66 3.03 7.92 7.83	78 44 72 59	44 32 32 44	-34 -12 -40 -15	60 35 61

In the third window we do not change anything, just click **Next** 

tà	Text Import	Wizard - I	Fixed Wi	idth Step 3	of 6	>
The first ca	se of data begins on wh	ich line num	iber?	2		
How many	ines represent a case?	1				
-How man	cases do you want to i	mport?				
() All o	the cases					
O The		1				
_		cases.				
© A ⊵e	rcentage of the cases:	10 🌲	%			
Data and						
Data previ	ew					
0	10	20	30	40	50	60
1	2 underweight	15.66	78	44	-34	60
2	1 normal	23.03	44 72	32 32	-12	35
	1 overweight		59	32	-40	61
4	1 overweight 1 normal	27.83	59	44	-15	61
			34	25	-28	46
		13.65 25.86	50	25	-18	35 35
			73	50	-10	
2 3 4 5 6 7 8 9	1 underweight 1 normal			39	-23	61
10		22.44	70		-31	46
		19.88	60	44	-16	46
11	1 normal	23.24	58	32	-26	35 👻
1						4
	< Back	vext > Fi	nish C:	ancel He	in	
	C Back (	VEAL +		пе	ip .	

In this window we move break lines (by clicking on them and moving to appropriate positions) in order to specify where each variable begins.

We suggest to move them to the following positions 0, 7, 20, 29, 34, 46, 56, 67. Of course other break positions are possible, we just need to separate columns. Click **Next** 

ta Text Imp	ort Wizard	- Fixed V	Vidth Step	4 of 6	×
Specify where each variabl To INSERT a variable break	-				ta area
Alternatively, move to the po then press the Insert Break	sition using th				
To MOVE a variable break I	ine, drag it to th	ne new po	sition.		
To DELETE a variable brea the Delete Break button.	k line, select it	or type its	position. Th	en press the	Delete key or
<u>Buler:</u> ∮	‡0	30		50	
1 2 underweig		78	44	-34	60 🕋
2 1 normal	23.03	4.4	32	-12	35
3 1 overweight		72	32	-40	61
4 1 overweigh 5 1 normal	27.83	59 60	44 32	-15	61 46
6 1 underweig		34	25	-9	35
7 1 overweigh		50	32	-18	35
Column Number: 67	· Ir	n <u>s</u> ert Breal	k Dele	te Break	
Current <u>V</u> ariable Width:	11				
< Back	Next >	Finish	Cancel H	elp	

In steps 5 and 6 there is nothing to do, just click **Next** and **Finish** respectively:

Data f If a co to strii The le value	length (number of e present in the fir be truncated.	ined from the nultiple data f f characters)	values pres ypes in the fli for string vari Is. If subsequ	ent in the first 2 rst 200 records ables is detern	, the variable nined by the le	ongest	2 63	yar2         yar3         y           28         840         1         1           38         2480         0         1           32         18240         0         1           33         870         0         1	Would O <u>Y</u> es ONO	you like to s	ave this file formation of the syntax?	at for future (	
ex			Sex						@ <u>N</u> o				
ata format	at												
ata format Numeric	at	*					r Data previ	iew	Press th	ie Finish bu	tton to complete t	he text impo	ort wizard.
	at	×					Data previ	BMIcat	Press th BMIquant	Age	tton to complete t		sBes
	at	*											SBes
	at							BMIcat	BMIquant	Age	DGestimate	DGerror	SBes
	at							BMIcat underweig	BMIquant 15.66	Age 78	DGestimate	DGerror -34	SBes 60
lumeric								BMIcat underweig normal	BMIquant 15.66 23.03	Age 78 44	DGestimate 44 32	DGerror -34 -12	SBes 60 35
		×						BMIcat underweig normal overweight	BMIquant 15.66 23.03 27.92	Age 78 44 72	DGestimate 44 32 32	DGerror -34 -12 -40	SBes 60 35 61
lumeric	ew	innet	Âre	DGestimate	DCerror	SBac		BMIcat underweig normal overweight overweight	BMIquant 15.66 23.03 27.92 27.83	Age 78 44 72 59	DGestimate 44 32 32 44	DGerror -34 -12 -40 -15	SBes 60 35 61 61
lumeric ata previe Sex	ew	BMIquant	Age 78	DGestimate		SBes		BMIcat underweig normal overweight overweight normal	BMIquant 15.66 23.03 27.92 27.83 21.41	Age 78 44 72 59 60	DGestimate 44 32 32 44 32	DGerror -34 -12 -40 -15 -28 -9 -18	SBes 60 35 61 61 46
lumeric	ew BMIcat underweig	BMIquant 15.66	78	44	-34	60 📥		BMIcat underweig normal overweight normal underweig	BMIquant 15.66 23.03 27.92 27.83 21.41 13.65	Age 78 44 72 59 60 34	DGestimate 44 32 32 44 32 25	DGerror -34 -12 -40 -15 -28 -9	SBes 60 35 61 61 46 35
lumeric ata previe Sex	ew	BMIquant						BMIcat underweig normal overweight normal underweig overweight	BMIquant 15.66 23.03 27.92 27.83 21.41 13.65 25.86	Age 78 44 72 59 60 34 50	DGestimate 44 32 32 44 32 25 32	DGerror -34 -12 -40 -15 -28 -9 -18	SBes 60 35 61 61 46 35 35

The uploading process is finished and the first 5 observations in the **Data View** section are shown below:

ta 🛛					Sk	eletonDataCon	nplete.sav [DataS	et1] - IBM SPSS S	tatistics Data E
<u>F</u> ile <u>E</u> dit	View	<u>D</u> ata	<u>T</u> ransform <u>A</u> n	alyze Direct <u>M</u> ark	eting <u>G</u> raph	s <u>U</u> tilities Ad	d- <u>o</u> ns <u>W</u> indow	<u>H</u> elp	
( <b>a</b> )			<b>i</b> n a	- 🖹 🕌		H 👪 🛛	🖬 🔚 🖧	. 🛄 🔒 🤇	) <b>•</b> 4
5:									
	Sex	c	BMIcat	BMIquant	Age	DGestimate	DGerror	SBestimate	SBerror
1		2	underweight	15.66	78		44 -3	4 60	-18
2		1	normal	23.03	44		32 -1	2 35	-9
3		1	overweight	27.92	72		32 -4	0 61	-11
4		1	overweight	27.83	59		44 -1	5 61	2
5		1	normal	21.41	60		32 -2	8 46	-14

Now we go to the **Variable View** section to indicate characteristics of our variables. First note that 'Sex' variable is categorical and is represented by 1 and 2 corresponding to 'Male' and 'Female' respectively. To indicate that 1 is 'Male' and 2 is 'Female' click on the right side of the 'Values' cell of 'Sex' variable:

<u>File E</u> dit	<u>V</u> iew <u>D</u> ata	Transform	<u>A</u> nalyze [	Direct <u>M</u> arketin	g <u>G</u> raphs <u>U</u> t	lilities Add- <u>o</u> n	s <u>W</u> indow	<u>H</u> elp			
2	🖨 🔟		¥ 🏼	<b>*</b> =			- 4		M 🕢 🎙	AB6	
	Name	Туре	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Sex	Numeric	7	0		None	None	8	酒 Right	🚓 Nominal	S Input
2	BMIcat	String	13	0		None	None	13	📰 Left	\delta Nominal	🔪 Input
3	BMIquant	Numeric	9	2		None	None	9	🔳 Right	🛷 Scale	🔪 Input
4	Age	Numeric	5	0		None	None	8	置 Right	🔗 Scale	S Input

'Value Labels' window appears, in the 'Value' section enter 1 and in the 'Label' section type 'Male':

<b>ta</b>	Value Labels	×
-Value La Val <u>u</u> e: 1 Label: N		Spelling
Cha	ad nge	
	OK Cancel Help	

Click Add, then in the 'Value' section put 2 and in the 'Label' section type 'Female':

<b>6</b>	Value Labels	×
Value Labels Value: 2 Label: Female Change Remove	Male"	Spelling
	OK Cancel Help	

Click Add again and OK. Now in the outputs instead of 1 and 2 we would see 'Male' and 'Female'. Finally we add some labels and give 'Measures' to our variables, 'Sex' and 'BMIcat' are categorical

<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>D</u> ata	Transform	<u>A</u> nalyze D	Direct <u>M</u> arketi	ng <u>G</u> raphs <u>U</u> tilities A	dd- <u>o</u> ns <u>W</u> ind	ow <u>H</u> elp					
😑 H												
	Name	Туре	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role	
1	Sex	Numeric	7	0		{1, Male}	None	8	酒 Right	🚓 Nominal	💊 Input	
2	BMIcat	String	13	0	The body mass ind. cat.	None	None	13	≣≣ Left	🚓 Nominal	🔪 Input	
3	BMIquant	Numeric	9	2	The body mass ind. quant.	None	None	9	🔳 Right	🛷 Scale	💊 Input	
4	Age	Numeric	5	0	Actual Age	None	None	8	) I Right	🛷 Scale	💊 Input	
5	DGestimate	Numeric	12	0	Est. Age using D	None	None	12	遭 Right	🛷 Scale	💊 Input	
6	DGerror	Numeric	10	0	Est Act. using D (years)	None	None	10	Right	🛷 Scale	🔪 Input	
7	SBestimate	Numeric	11	0	Est. Age using SB	None	None	11	🔳 Right	🔗 Scale	💊 Input	
8	SBerror	Numeric	9	0	Est Act. using SB (years)	None	None	9	)置 Right	🔗 Scale	💊 Input	

and we select 'Nominal' measure, all the other variables have 'Scale' measures.

All the work is done and we suggest to save this data set before analyses (File > Save as and save in 'sav' format).

Now we briefly give some useful details for other data sets.

## 'The Anscombe' data set

In the 2nd step of 'Text Import Wizard' select 'Delimited' and 'Yes' for the first and second questions respectively.

ta 🛛	Text Import Wizard - Step 2 of 6	×
How are your val	iables arranged? - Variables are delimited by a specific character (i.e., comma, tab). - Variables are aligned in fixed width columns.	
Are variable nam (Yes No	es included at the top of your file?	

In the 4th step select 'Space' as delimiters.

Comm Othe <u>r</u> :	a 🗾	Semicolon		<ul> <li>None</li> <li>Singl</li> <li>Doub</li> <li>Other</li> </ul>	le quote	
ata previ	ew y1	y2	y3	x2	¥4	_
-		y2 9.14	y3 7.46	x2 8	y4 6.58	4
x1 10	y1					4
x1	y1 8.04	9.14	7.46	8	6.58	
x1 10 8	y1 8.04 6.95	9.14 8.14	7.46	8	6.58 5.76	
x1 10 8 13	y1 8.04 6.95 7.58	9.14 8.14 8.74	7.46 6.77 12.74	8 8 8	6.58 5.76 7.71	
x1 10 8 13 9 11	y1 8.04 6.95 7.58 8.81	9.14 8.14 8.74 8.77	7.46 6.77 12.74 7.11	8 8 8	6.58 5.76 7.71 8.84	
x1 10 8 13 9 11 14	y1 8.04 6.95 7.58 8.81 8.33	9.14 8.14 8.74 8.77 9.26	7.46 6.77 12.74 7.11 7.81	8 8 8 8	6.58 5.76 7.71 8.84 8.47	
x1 10 8 13 9 11 14 6	y1 8.04 6.95 7.58 8.81 8.33 9.96	9.14 8.14 8.74 8.77 9.26 8.10	7.46 6.77 12.74 7.11 7.81 8.84	8 8 8 8 8	6.58 5.76 7.71 8.84 8.47 7.04	
x1 10 8 13 9	y1 8.04 6.95 7.58 8.81 8.33 9.96 7.24	9.14 8.14 8.74 8.77 9.26 8.10 6.13	7.46 6.77 12.74 7.11 7.81 8.84 6.08	8 8 8 8 8 8 8	6.58 5.76 7.71 8.84 8.47 7.04 5.25	
x1 10 8 13 9 11 14 6 4	y1 8.04 6.95 7.58 8.81 8.33 9.96 7.24 4.26	9.14 8.14 8.74 9.26 8.10 6.13 3.10	7.46 6.77 12.74 7.11 7.81 8.84 6.08 5.39	8 8 8 8 8 8 8 8 8	6.58 5.76 7.71 8.84 8.47 7.04 5.25 5.56	

The Variable View section looks like that:

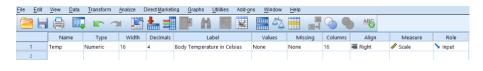
<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>D</u> ata	Transform	<u>A</u> nalyze E	Direct <u>M</u> arketin	ig <u>G</u> raphs <u>U</u> t	ilities Add- <u>o</u> n	s <u>W</u> indow	Help			
😑 H		5	• 📓	<b>k</b> =	E H		- 4		H 🕢 🌑	ABG	
	Name	Туре	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	x1	Numeric	2	0		None	None	8	🗃 Right	🛷 Scale	🔪 Input
2	y1	Numeric	5	2		None	None	8	🗃 Right	🛷 Scale	🔪 Input
3	y2	Numeric	4	2		None	None	8	🗃 Right	🛷 Scale	🔪 Input
4	уз	Numeric	5	2		None	None	8	🔳 Right	🛷 Scale	🔪 Input
5	x2	Numeric	2	0		None	None	8	🗃 Right	🛷 Scale	🔪 Input
6	y4	Numeric	5	2		None	None	8	🔳 Right	🛷 Scale	🔪 Input

## 'The Body Temperature' data set

In the 2nd step of 'Text Import Wizard' select 'No' for the second question.



In the 4th step there is only 1 break at position 0. In the **Variable View** section we enter Labels and Measures:



#### 'CFC-11 concentration' data set

In the 2nd step of 'Text Import Wizard' select 'Yes' for the second question. In the 4th step we put break positions at 0,10,28,37,48.

The first 5 observations are shown below:

File	Edit	View	Data 1	Transform	<u>A</u> n	alyze Direct <u>M</u> a	arketing (
			<b>II</b> ,	5 3		<b>i</b>	=
323 :	year						
		)	rear	month		time	cfc11
	1		1977		1	1977.00	139.90
	2		1977		2	1977.08	139.50
	3		1977		3	1977.17	139.00
	4		1977		4	1977.25	134.10
	5		1977		5	1977.33	135.00

In the **Variable View** section, SPSS gave 'String' type to 'cfc11' variable which we need to change to numeric. Click on the right side of the 'Type' cell in 'cfc11' variable:

File	Edit	<u>V</u> iew <u>D</u> ata	Transform	<u>A</u> nalyze E	Direct Marketin	g <u>G</u> raphs	Utilities Add-on	s <u>W</u> indow	<u>H</u> elp			
6	• 🗄		5 3	- 📓	<b>!</b>	r H		<u>-</u>		- 0	ABG	
		Name	Туре	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
	1	year	Numeric	28	0		None	None	12	🚟 Right	🛷 Scale	🔪 Input
	2	month	Numeric	9	0		None	None	9	雇 Right	\delta Nominal	🔪 input
	3	time	Numeric	11	2		None	None	11	🚟 Right	🛷 Scale	🔪 Input
	4	cfc11	String	8	0		None	None	8	📑 Left	\delta Nominal	ゝ Input

In the 'Variable Type' window select 'Numeric' and enter 8 and 2 in 'Width' and 'Decimal Places' respectively, click **OK**:

tariabl Variabl	е Туре 🛛 🗙
<u>Numeric</u> <u>Comma</u> <u>Dot</u> <u>O</u> t	<u>W</u> idth: 8 Decimal Places: 2
String     Restricted Numeric (integer with leadin     The Numeric type honors the digit gr     Numeric never uses digit grouping.     OK Car	

Finally we enter some labels and measures

<u>File</u> Edit	<u>V</u> iew <u>D</u> ata	Transform	<u>A</u> nalyze [	Direct <u>M</u> arketi	ng <u>G</u> raphs	Utilities Add-ons	<u>N</u> indow <u>H</u> elp					
😂 H		5 3		<b>*</b> =	TF H			A A	•	€		
	Name	Туре	Width	Decimals		Label	Values	Missing	Columns	Align	Measure	Role
1	year	Numeric	28	0	Year		None	None	12	🚟 Right	🛷 Scale	🔪 Input
2	month	Numeric	9	0	Month		None	None	9	🗃 Right	\delta Nominal	🔪 Input
3	time	Numeric	11	2	Time		None	None	11	🗃 Right	I Scale	🔪 Input
4	cfc11	Numeric	8	2	CFC-11 concer	tration (parts per trillio	) None	None	8	🗃 Right	🛷 Scale 🛛 🔻	🔪 Input
5											🛷 Scale	
6											I Ordinal	
7											\delta Nominal	

# 'The Coffee Shop' data set

In the 2nd step of 'Text Import Wizard' select 'Yes' for the second question. In the 4th step there are 3 break lines at positions 0, 9, 16.

ta 🛛	Text I	mport Wizard - Fixed Width Step 4 of 6	×
Specify when	e each vari	able begins. The first column is column 0.	
	move to the	reak line, click at the desired position in the ruler or data area. position using the arrow keys or by typing the column number; eak button.	
To MOVE a v	ariable bre	ak line, drag it to the new position.	
To DELETE : the Delete Br		reak line, select it or type its position. Then press the Delete key	or
<u>Buler:</u> þ		20 30 40 50 60	
1 850	67	bakery	4
	64	bakery	
2 850 3 420	59	bakery	
4 490	75	bakery	
5 130	17	bakery	
4 490 5 130 6 370 7 460	47	bakery	
7 460	61	bakery	Ŧ
1		•	
<u>C</u> olumn Nu	mber:	16 Ingert Break Delete Break	
Current Var	iable Width	c 7	
	< E	ack Next > Finish Cancel Help	

The first 5 observations are displayed below:

<u>F</u> ile <u>E</u> dit	View	<u>D</u> ata	Transform	<u>A</u> nalyze	Dire
		10.			4
		i	3		
	calor	ies	carb	type	
1		350		bakery	
2		350	64	bakery	
3		420	59	bakery	
4		490	75	bakery	
5		130	17	bakery	

In the Variable View section we enter Labels and Measures:

<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>D</u> ata	Transform	<u>A</u> nalyze D	Direct <u>M</u> arketii	ng <u>G</u> raphs <u>U</u> t	ilities Add- <u>o</u> n:	s <u>W</u> indow	<u>H</u> elp			
😑 H		<b>E</b> 7	- 📓	<b>!</b>	РА				- Co	ABS	
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	calories	Numeric	9	0	Calories	None	None	9	🚟 Right	🛷 Scale	🔪 Input
2	carb	Numeric	7	0	Carbohydrates (g)	None	None	8	🗃 Right	🛷 Scale	🔪 Input
3	type	String	12	0	Туре	None	None	12	🗮 Left	💑 Nominal	🔪 Input

## 'The Crawling' data set

In the 2nd step of 'Text Import Wizard' select 'Yes' for the second question. In the 4th step we put break positions at 0, 12, 29. The data set looks like that:

The data set looks like that:

<u>F</u> ile	Edit	View	<u>D</u> ata	Transform	Analyze	Direct M	arketing	G
6				<b>F</b>	<b>1</b>			Ч
13 : bi	rth_mo	nth						
		birth	n_month	avg_c	rawling_age	ter	mperature	e
1	1	Januar	1		29	.84		66
2	2	Februa	ry		30	.52		73
3	3	March			29	.70		72
4	1	April			31.	.84		63
5	5	May			28	.58		52
6	5	June			31.	.44		39
7	7	July			33.	.64		33
8	3	August			32	.82		30
ç	)	Septen	nber		33.	.83		33
1	0	Octobe	er		33.	.35		37
1	1	Novem	iber		33.	.38		48
1	2	Decem	ber		32	.32		57

Then we give labels and measures:

<u>F</u> ile <u>E</u> d	t <u>V</u> iew <u>D</u> ata	Transform	<u>A</u> nalyze (	Direct <u>M</u> arketi	ng <u>G</u> raphs <u>U</u> tilities	Add-ons	Window Help	)			
😑 l											
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	birth_month	String	12	0	Month	None	None	12	📰 Left	\delta Nominal	🔪 Input
2	avg_crawling	Numeric	17	2	Average Crawling Age	None	None	17	🔳 Right	🛷 Scale	🔪 Input
3	temperature	Numeric	2	0	Temperature in Celsius	None	None	11	🗃 Right	🛷 Scale	🔪 Input

One problem still remains the temperature in the data set is in Fahrenheit but we need in Celsius. To make transformation go to **Transform** > **Compute Variable** 



For the 'Target Variable' put 'temperature' and type the conversion equation into 'Numerical Expression':

<b>5</b>	Compute Variable	×
Tryset Variable: temperature Type & Label. Month (pint_month) Average Crawing A Temperature in Cel. (cptional case select)	Numetic Expression:           Imperature 32)*56           Imperature 32)*56	
	OK Paste Reset Cancel Help	

Click  $\mathbf{OK},$  again click  $\mathbf{OK}$  for 'Change existing variable?' question.

## 'The Life Expectancy' data set

In the 2nd step of 'Text Import Wizard' select 'Yes' for the second question. In the 4th step we put break positions at 0, 30, 45, 55, 70.

ta Text Import Wizard - F	ixed Width Step	4 of 6	×							
Specify where each variable begins. The firs										
To INSERT a variable break line, click at the desired position in the ruler or data area. Alternatively, move to the position using the arrow keys or by typing the column number, then press the Insert Break button.										
To MOVE a variable break line, drag it to the new position.										
To DELETE a variable break line, select it or type its position. Then press the Delete key or the Delete Break button.										
<u>Ruler:</u> [2	10		.60							
1 Afghanistan	SAs	48.673	NA 📥							
2 Albania 3 Algeria	EuCA	76.918	NA							
3 Algeria	MENA		6406.81							
4 Angola 5 Argentina	SSA Amer		5519.18 5741.04							
4 Angola 5 Argentina 6 Armenia	EuCA		4748.92							
7 Aruba	Amer	75.246	NA -							
Column Number: 0 Ingert Break Delete Break										
Current Variable Width: 0										
Sext > Fir	ish Cancel H	elp								

First 5 observations look like that:

File	Edit	View	Data	Transform	Analyze	Direct M	arketing	Graph	s <u>U</u> tiliti	es	Add- <u>o</u> ns	Window	Help	
(			1	<b>E</b> 1	7 🖁	1		ч	AA 🚦	5		- 43 - 13		1
198 : 0	Country													
				Country			Regio	n	LifeE	φ		GDP	н	IV
1	1	Afghan	istan				SAs		4	8.673	NA		NA	
2	2	Albania	a				EuCA		7	6.918	NA		NA	
3	3	Algeria					MENA		7	3.131	6406.8	1662	0.10	)
4	4	Angola					SSA		5	1.093	5519.1	8318	2.00	D
5	5	Argent	ina				Amer		7	5.901	15741.	04577	0.50	)

SPSS gave 'String' type for 'GDP' and 'HIV' variables, but we clearly need them to be 'Numeric', so click on the right side of the 'Type' cell of 'GDP' variable:

<u>F</u> ile	Edit	<u>V</u> iew <u>D</u> ata	Transform	<u>A</u> nalyze [	Direct <u>M</u> arketii	ng <u>G</u> raphs <u>U</u> tilities	Add- <u>o</u> ns	Window Help	p			
6			5	• 📓	<b>*</b> =					🌏 🌑 4	**	
		Name	Туре	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
	1	Country	String	30	0		None	None	30	📰 Left	\delta Nominal	🔪 Input
	2	Region	String	15	0		None	None	15	📑 Left	💑 Nominal	🔪 input
	3	LifeExp	Numeric	10	3	Life Expectancy (years)	None	None	10	🚟 Right	🛷 Scale	🔪 Input
	4	GDP	String	15	0		None	None	15	📑 Left	💑 Nominal	🔪 Input
	5	HIV	String	7	0		None	None	7	📰 Left	\delta Nominal	🔪 Input

Then in the 'Variable Type' select 'Numeric' and enter 15 and 5 for 'Width' and 'Decimal Places' respectively. Click  $\mathbf{OK}$ 



Similar procedure for 'HIV' variable but we enter 7 and 2 for 'Width' and 'Decimal Places'.

Variable Type	×
Bytumeric     Comma     Comma     Ded     Det     Cogtom     Currency     Stipn     Restricted Numeric (integer with leading zeros)	×
The Numeric type honors the digit grouping setting, while the Restricted Numeric never uses digit grouping.	
Numeric never uses digit grouping.     OK Cancel Help	

Finally we type some labels and select measures, 'Country' and 'Region' are categorical and so

have 'Nominal' measures while all other variables have 'Scale' measure.

<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>D</u> ata	Transform	<u>A</u> nalyze D	irect <u>M</u> arketii	ng <u>G</u> raphs <u>U</u> tilities	Add- <u>o</u> ns	<u>W</u> indow <u>H</u> elp	)			
😑 H											
	Name	Туре	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Country	String	30	0		None	None	30	📰 Left	🗞 Nominal	🔪 Input
2	Region	String	15	0		None	None	15	📑 Left	🗞 Nominal	🔪 Input
3	LifeExp	Numeric	10	3	Life Expectancy (years)	None	None	10	🖷 Right	🛷 Scale	🔪 Input
4	GDP	Numeric	15	5		None	None	15	🗃 Right	🛷 Scale	🔪 Input
5	HIV	Numeric	7	2		None	None	7	🗃 Right	💰 Nominal 🚿	🔪 Input
6										🛷 Scale	
7										J Ordinal	
8										💑 Nominal	

### 'The New York Red Bulls Salary' data set

In the 2nd step of 'Text Import Wizard' select 'No' for the second question. In the 4th step there is only 1 break at position 0.

In the Variable View section we give labels and measures:

<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>D</u> ata	Transform	<u>A</u> nalyze (	Direct <u>M</u> arketin	ig <u>G</u> raphs <u>U</u> til	ities Add- <u>o</u> ns	Window H	lelp			
😑 t	🖨 🛙		<b>1</b>	<b>*</b>	r H		- 4	<u>م</u>		ABC	
	Name	Туре	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	NYSalary	Numeric	10	2	VY Red Bull Salarie	s None	None	10	置 Right	🛷 Scale	🔪 Input
2											

It will be convenient to have another variable which is salary but in thousands of dollars (will be useful for box-plots). So we introduce a new variable which we call 'NYSal.in.Th' (NY Salaries in Thousands)

File	Edit	View	Data	Transform	<u>A</u> nalyze I	Direct <u>M</u> arketi	ng <u>G</u> raphs <u>U</u> tilities	Add-o	ns <u>W</u> indow	Help				
	😂 🖩 🖨 💷 🖛 🛥 🎬 🚣 🗐 📴 🔥 🏢 🖾 📰 📲 🏠 🚟 📑 🎯 🧠 🤏													
		Nan	ne	Type	Width	Decimals	Label		Values	Missing	Columns	Align	Measure	Role
	1	NYSalary	/ 1	Numeric	10	2	NY Red Bull Salaries		None	None	10	🗮 Right	🛷 Scale	🔪 Input
:	2	NYSal.in.	Th I	Numeric	8	2	NY Salaries in Thousand	s	None	None	13	🔳 Right	🛷 Scale	ゝ Input

Next we need to calculate the values for this variable, to do that go to **Transform** > **Compute Variable**, then enter the following expression:

<b>ta</b>	Compute Variable	×
Target Vanable: NYSalin, Th Type & Label. WYR-ed Bull Salarie. NY Salaries in Thou	Imagic Expression:           =           NVSstay1000           Imagic Expression:           Imagic Expressinter           Imagic Expr	
(optional case selec	ion condition)	
L	OK Paste Reset Cancel Help	-

Click **OK** and salaries in thousands are calculated:

<u>F</u> ile	Edit	View	<u>D</u> ata	Transform	n <u>A</u> naly
6			1.	2	2
1:					
		NYS	alary	NYSal.	in.Th
	1	33	3750.00		33.75
	2	3	3750.00		33.75
	3	3	3750.00		33.75
	4	3	3750.00		33.75
	5	4	4000.00		44.00

# 'The Plastic Surgery' data set

In the 2nd step of 'Text Import Wizard' select 'No' for the second question. In the 4th step there is only 1 break at position 0. Finished **Variable View** looks like that:

<u>File</u>	it <u>V</u> iew <u>D</u> al	a <u>T</u> ransform	Analyze	Direct Market	ing <u>G</u> raphs <u>U</u> tilitie	s Add- <u>o</u> ns	Window Help	2			
	🚬 H 🖨 💷 🖛 🛥 📰 💺 📰 🌇 🏙 📰 🖾 📰 🛣 📰 🖓 🕒 🤲										
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	AgeChange	Numeric	4	1	Age change in years	None	None	8	🖀 Right	🛷 Scale	🔪 Input
2											