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:

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:

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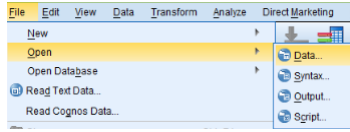
Once 'Open Data' window appears find the file that you want to upload. Make sure that in the 'Files of type:' section, 'All Files' is selected so that we can see all files
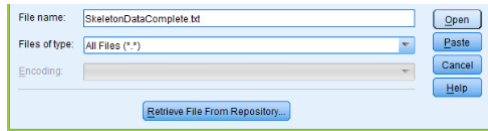

Once the data file is selected, click on the 'Open' button.

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).
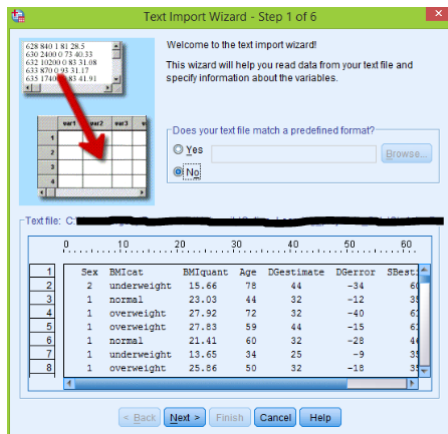

Click **Next** and second window opens. For the 'How are your variable arranged?' question select 'Fixed width' since the data file is in text format (if it is in CSV format, select 'Delimited'). If we have variable names in the text data file (headers) then select 'Yes' for the second question, otherwise select 'No'. In the 'Skeleton' file we have headers so we put 'Yes', then click **Next** button.
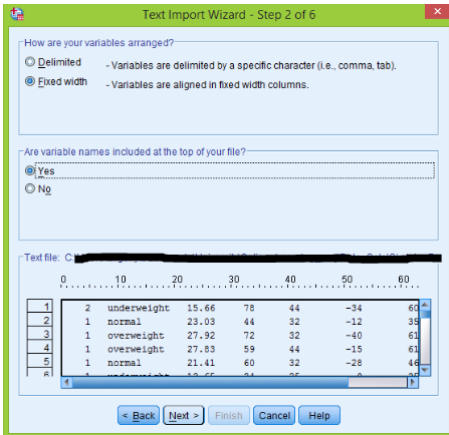
In the third window we do not change anything, just click **Next**
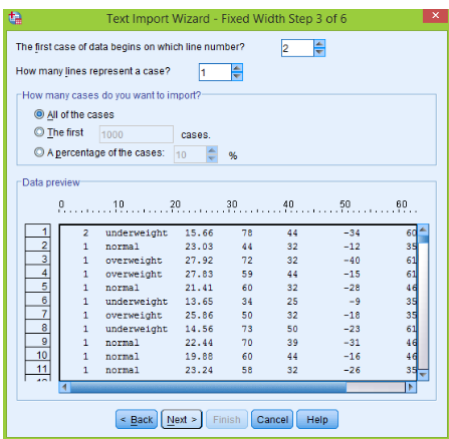
In this window we move break lines (by clicking on them and moving to appropriate positions) in order to specify where each variable begins.

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**
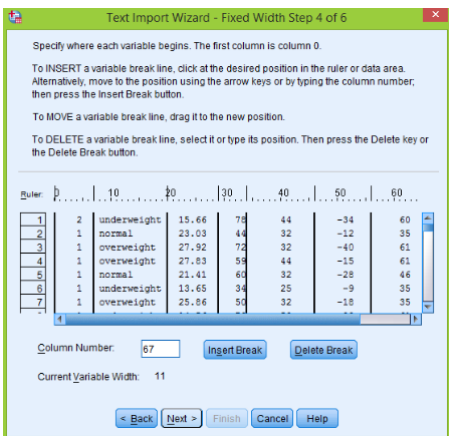

In steps 5 and 6 there is nothing to do, just click **Next** and **Finish** respectively:
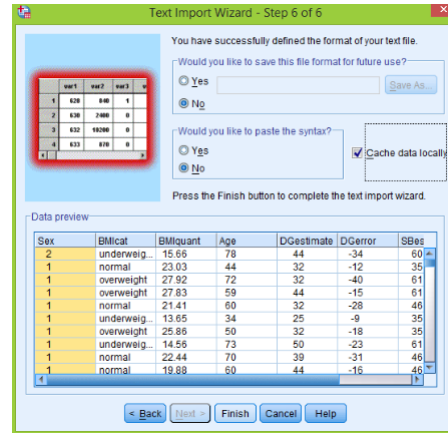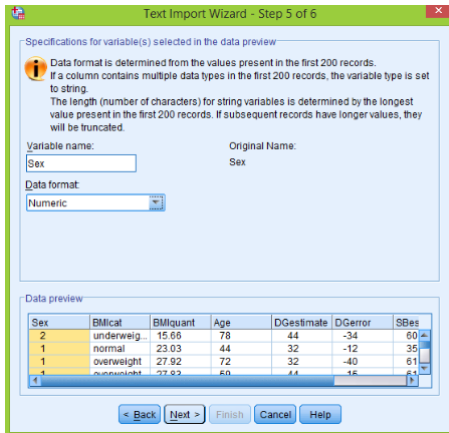
The uploading process is finished and the first 5 observations in the **Data View** section are shown below:
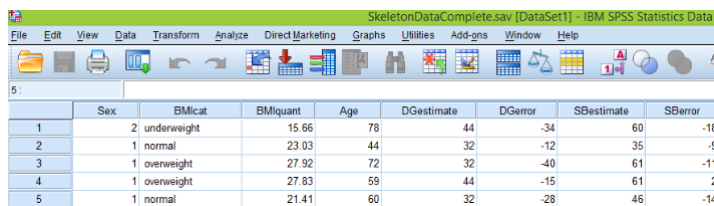

Now we go to the **Variable View** section to indicate characteristics of our variables. First note that 'Sex' variable is categorical and is represented by 1 and 2 corresponding to 'Male' and 'Female' respectively. To indicate that 1 is 'Male' and 2 is 'Female' click on the right side of the 'Values'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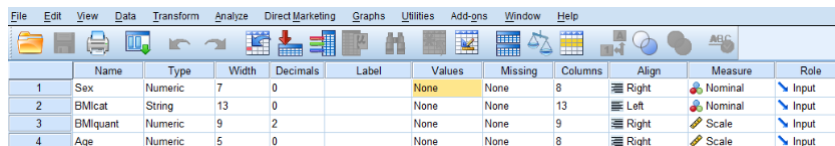
'Value Labels' window appears, in the 'Value' section enter 1 and in the 'Label' section type 'Male':

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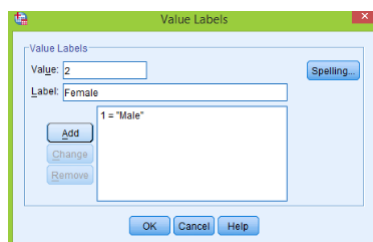
Click **Add** again and **OK**. Now in the outputs instead of 1 and 2 we would see 'Male' and 'Female'. Finally we add some labels and give 'Measures' to our variables, 'Sex' and 'BMIcat' are categorical

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

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Sex	Numeric	7	0		{1, Male}...	None	8	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	Right	Scale	Input
5	DGestimate	Numeric	12	0	Est. Age using D	None	None	12	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	Right	Scale	Input

All the work is done and we suggest to save this data set before analyses (**File** > **Save as** and save in 'sav' format).
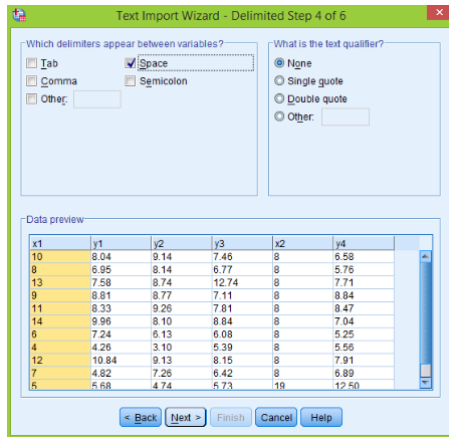
Now we briefly give some useful details for other data sets.

'The Anscombe' data set

In the 2nd step of 'Text Import Wizard' select 'Delimited' and 'Yes' for the first and second questions respectively.
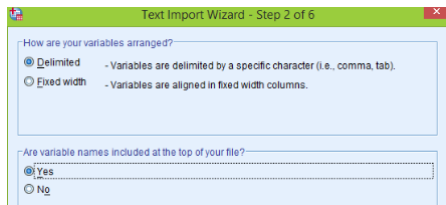
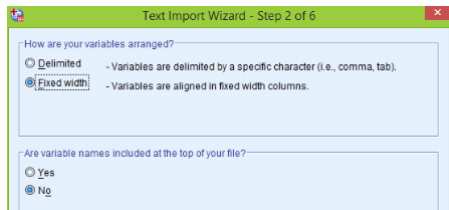In the 4th step select 'Space' as delimiters.

The **Variable View** section looks like that:

	Name	Type	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	y3	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 $0, 10, 28, 37, 48$.
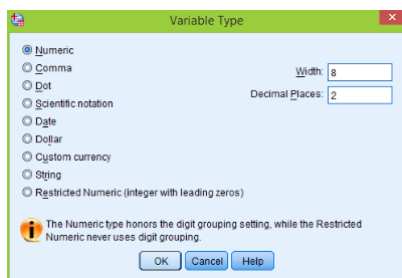The first 5 observations are shown below:

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:

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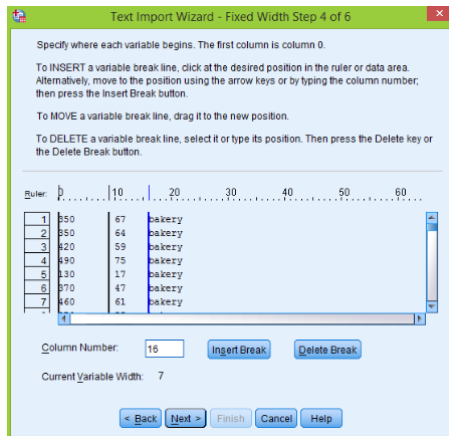

Finally we enter some labels and measures

'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 $0, 9, 16$.

The first 5 observations are displayed below:

In the **Variable View** section we enter Labels and Measures:

'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 $0, 12, 29$.
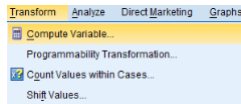The data set looks like that:

Then we give labels and measures:

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':
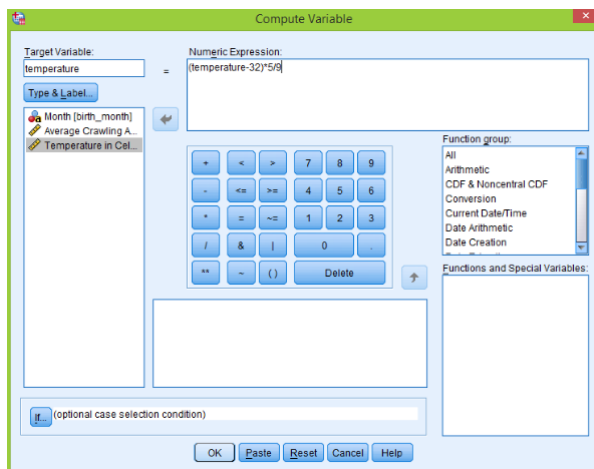
Click **OK**, again click **OK** for 'Change existing variable?' question.
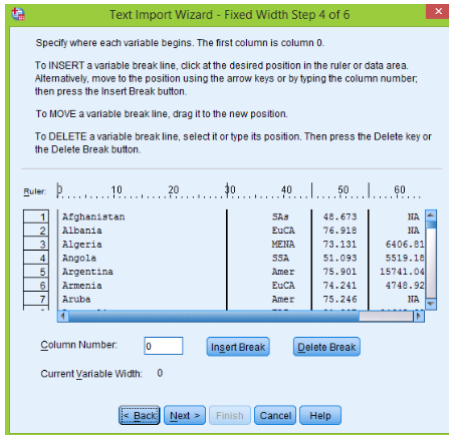
'The Life Expectancy' data set

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

First 5 observations look like that:

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:

Then in the 'Variable Type' select 'Numeric' and enter 15 and 5 for 'Width' and 'Decimal Places' respectively. Click **OK**
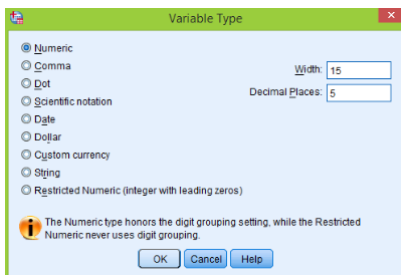
Similar procedure for 'HIV' variable but we enter 7 and 2 for 'Width' and 'Decimal Places'.
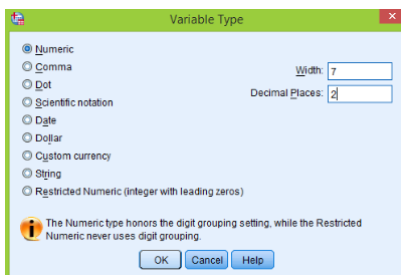
Finally we type some labels and select measures, 'Country' and 'Region' are categorical and so

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

	Name	Type	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										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:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	NYSalary	Numeric	10	2	NY Red Bull Salaries	None	None	10	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)

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	NYSalary	Numeric	10	2	NY Red Bull Salaries	None	None	10	Right	Scale	Input
2	NYSal.in.Th	Numeric	8	2	NY Salaries in Thousands	None	None	13	Right	Scale	Input

Next we need to calculate the values for this variable, to do that go to **Transform > Compute Variable**, then enter the following expression:
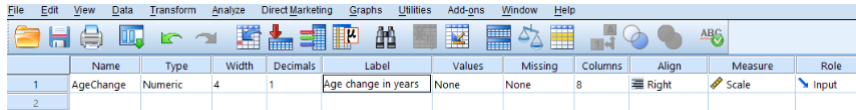
Click **OK** and salaries in thousands are calculated:

	NYSalary	NYSal.in.Th
1	33750.00	33.75
2	33750.00	33.75
3	33750.00	33.75
4	33750.00	33.75
5	44000.00	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:

	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											