



Loading Data and Defining Variables in SPSS

Tutorial 1



Introduction

Welcome to this beginner's tutorial on getting started with SPSS!

SPSS, which stands for Statistical Package for the Social Sciences, is a powerful statistical analysis software that is widely used in the fields of psychology, sociology, education, business, and many others.



**Statistical Package
for the Social
Sciences.**

If you are new to SPSS, you may feel overwhelmed by its many features and options. However, this tutorial will guide you step-by-step through the basics of using SPSS. You will learn how to import a data file, how to open it in SPSS, and how to define its variables and their properties.

By the end of this tutorial, you will have a good understanding of how to navigate SPSS and perform some of the most commonly used statistical procedures in readying a data set. So, let's get started and explore the world of SPSS together.

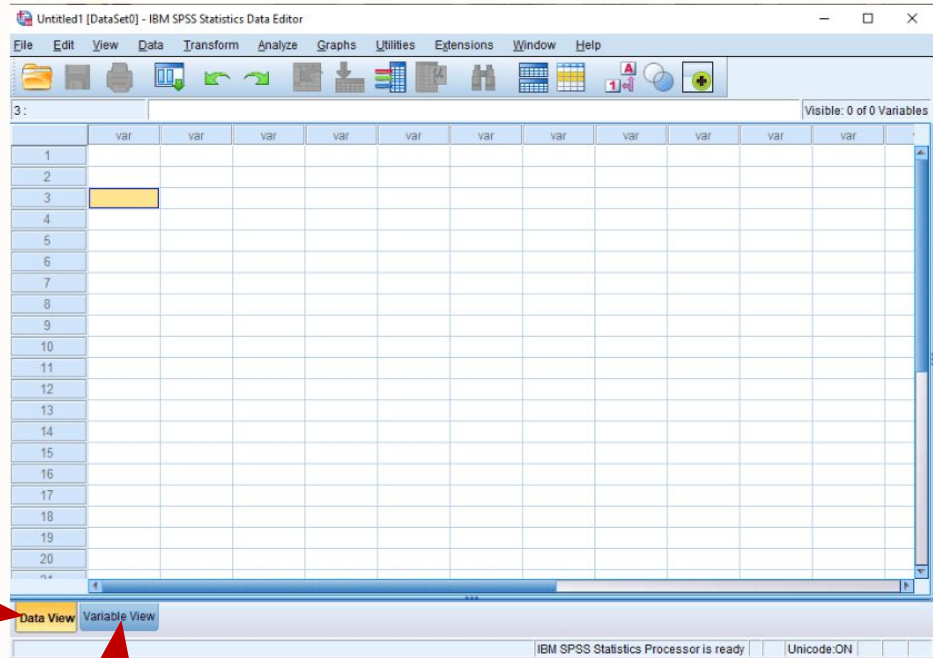
Now, when you first open SPSS, you will see the main interface window or the Data Editor window, as shown below.

The Main Interface

An important feature of the Data Editor window is that it presents two views: **Data View** and **Variable View**.

Data View

The Data View is where you can edit the data you loaded into SPSS.



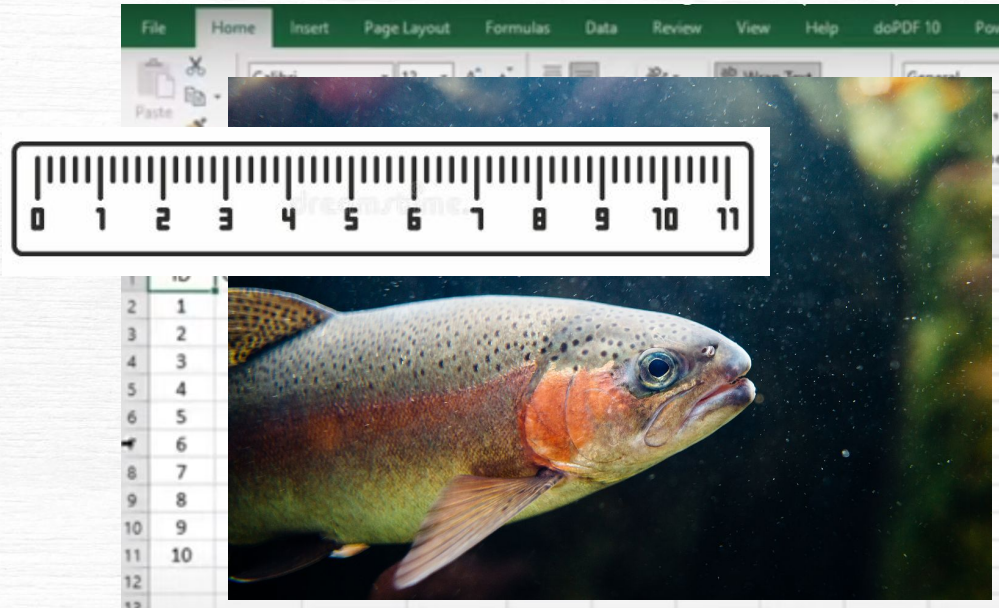
SPSS Data editor window

Variable View

The Variable View is where you define the properties of the variables of your data set.

The Main Interface

Ten fish were caught in Lake Jennings, measured, and released back into the water. The data are shown in the Excel spreadsheet below.

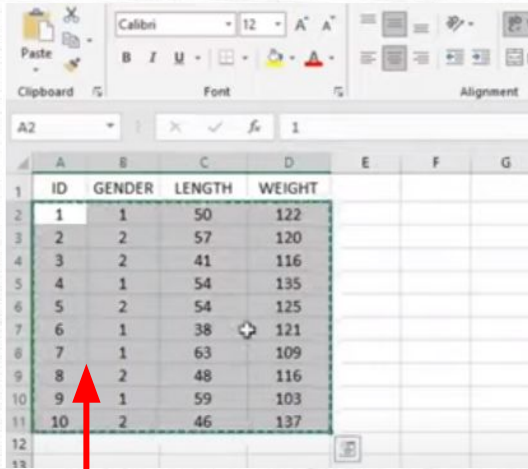




Loading Data into SPSS from Excel: Method 1

Method 1

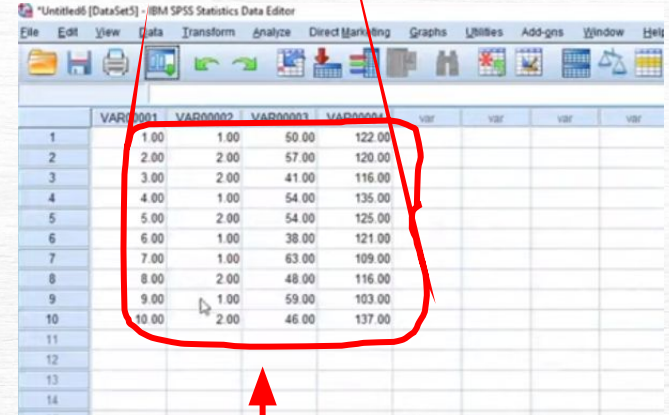
EXCEL



ID	GENDER	LENGTH	WEIGHT
1	1	50	122
2	2	57	120
3	2	41	116
4	1	54	135
5	2	54	125
6	1	38	121
7	1	63	109
8	2	48	116
9	1	59	103
10	2	46	137

In Excel, we highlight all the data **except for the first row, which contains the headings** (variable names).

SPSS – Data View



	VAR00001	VAR00002	VAR00003	VAR00004	VAR	VAR	VAR	VAR
1	1.00	1.00	50.00	122.00				
2	2.00	2.00	57.00	120.00				
3	3.00	2.00	41.00	116.00				
4	4.00	1.00	54.00	135.00				
5	5.00	2.00	54.00	125.00				
6	6.00	1.00	38.00	121.00				
7	7.00	1.00	63.00	109.00				
8	8.00	2.00	48.00	116.00				
9	9.00	1.00	59.00	103.00				
10	10.00	2.00	46.00	137.00				
11								
12								
13								
14								

Now, we go over to SPSS Data View, click on an empty cell, say A1, and simply paste the copied data.

*Untitled6 [DataSet5] - IBM SPSS Statistics Data Editor

	VAR00001	VAR00002	VAR00003	VAR00004		
1	1.00	1.00	50.00	122.00		
2	2.00	2.00	57.00	120.00		
3	3.00	2.00	41.00	116.00		
4	4.00	1.00	54.00	135.00		
5	5.00	2.00	54.00	125.00		
6	6.00	1.00	38.00	121.00		
7	7.00	1.00	63.00	109.00		
8	8.00	2.00	48.00	116.00		
9	9.00	1.00	59.00	103.00		
10	10.00	2.00	46.00	137.00		
11						
12						
13						
14						

When you enter data directly and manually, as we did, you will see that SPSS has generated default names for our variables. In this case, VAR00001, ..., VAR00004.

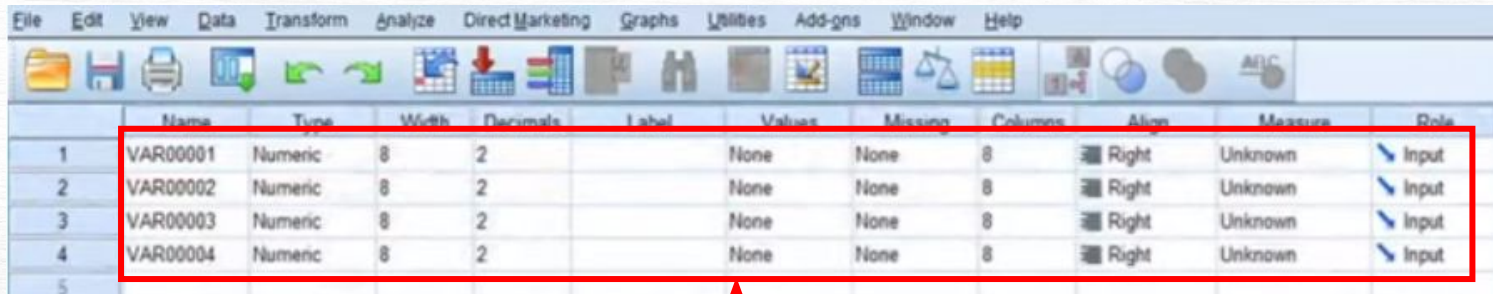
We will now see how to change these default variable names to the true variable names.



Defining Variable Properties

In order to change these default names to the actual names of the variables, we click on the **“Variable View”** tab.

SPSS Variable View

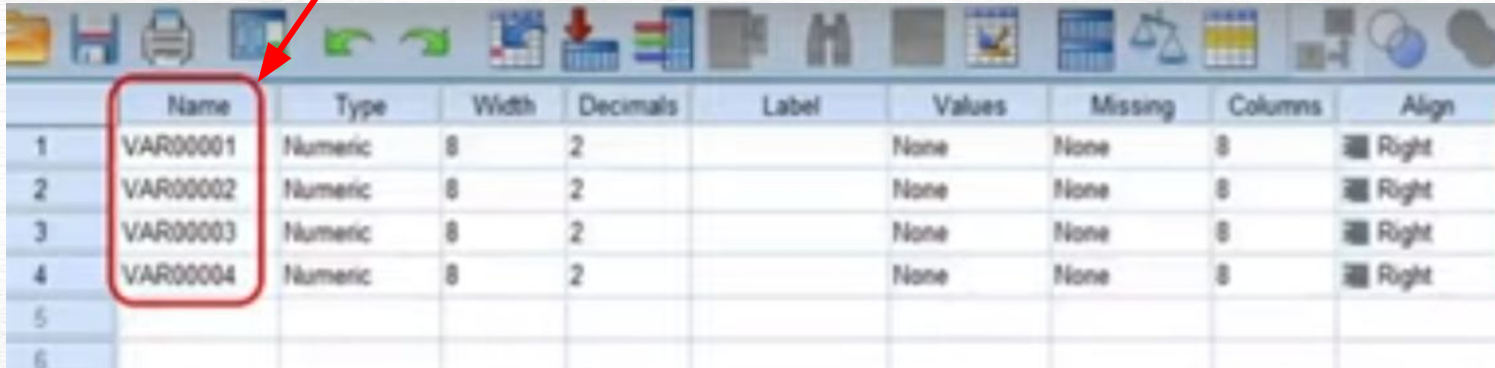


	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	VAR00001	Numeric	8	2		None	None	8	Right	Unknown	Input
2	VAR00002	Numeric	8	2		None	None	8	Right	Unknown	Input
3	VAR00003	Numeric	8	2		None	None	8	Right	Unknown	Input
4	VAR00004	Numeric	8	2		None	None	8	Right	Unknown	Input
5											

There, you will see that SPSS has also assigned default properties to each variable.

This happens whenever we load the data using the direct Copy and Paste method as we did.

The “Name” Column



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align
1	VAR00001	Numeric	8	2		None	None	8	Right
2	VAR00002	Numeric	8	2		None	None	8	Right
3	VAR00003	Numeric	8	2		None	None	8	Right
4	VAR00004	Numeric	8	2		None	None	8	Right
5									
6									

The first Column is “Name.” This is where we give names to the variables instead of “VAR00001.”

The “Name” Column

Before we begin to make name changes in the Variable View, we must familiarize ourselves with some conditions for naming variables in SPSS.

Condition 1: The names given to the variables must not have more than 64 characters.

Condition 2: There should be no spaces between words in the variable name, for example, ***Section Number***. To separate the two words, use the underscore: ***Section_Number***.

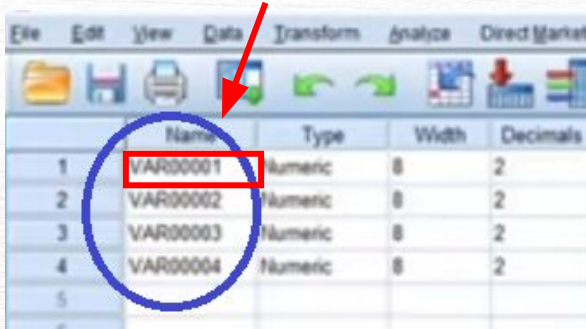
Condition 3: The names must be unique. Two variables cannot have the same name.

Condition 4: Names cannot start with a number or any special character, for example. 3, &, *, #.

With this in mind, let us now change the name of each variable from the default.

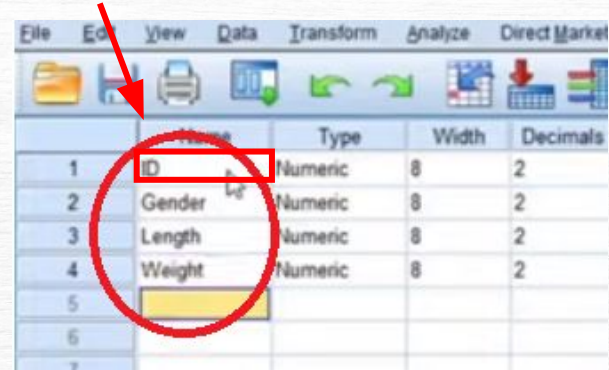
The “Name” Column

To change the first variable's name from VAR00001 to ID, we click in the cell. This will highlight the name. Then, we can type and change it to “ID.”



A screenshot of the SPSS Variable View window. The menu bar includes File, Edit, View, Data, Transform, Analyze, Direct, and Markers. Below the menu is a toolbar with icons for file operations and data manipulation. The main area contains a table with four columns: Name, Type, Width, and Decimals. The first row has the name 'VAR00001' highlighted with a red rectangle. A blue circle is drawn around the first four rows of the table. A red arrow points to the 'Name' column header.

	Name	Type	Width	Decimals
1	VAR00001	Numeric	8	2
2	VAR00002	Numeric	8	2
3	VAR00003	Numeric	8	2
4	VAR00004	Numeric	8	2
5				



A screenshot of the SPSS Variable View window, similar to the previous one. The first row now has the name 'ID' instead of 'VAR00001'. The 'ID' cell is highlighted with a red rectangle. A red circle is drawn around the first four rows of the table. A red arrow points to the 'Name' column header.

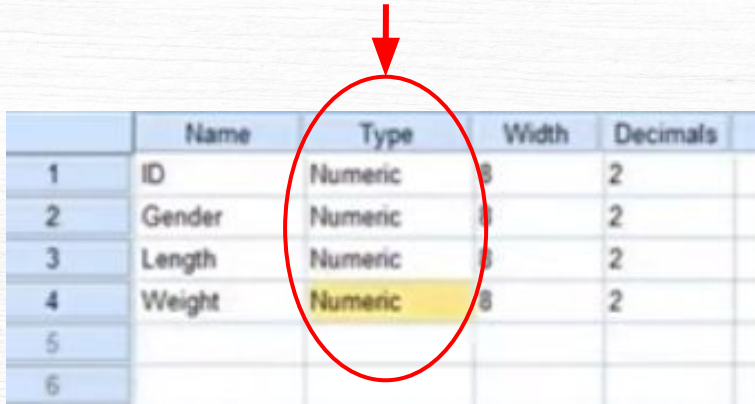
	Name	Type	Width	Decimals
1	ID	Numeric	8	2
2	Gender	Numeric	8	2
3	Length	Numeric	8	2
4	Weight	Numeric	8	2
5				
6				

Change the other names:

From “VAR00002” to “Gender,” from “VAR00003” to “Length,” and from “VAR00004” to “Weight.”

The “Type” Column

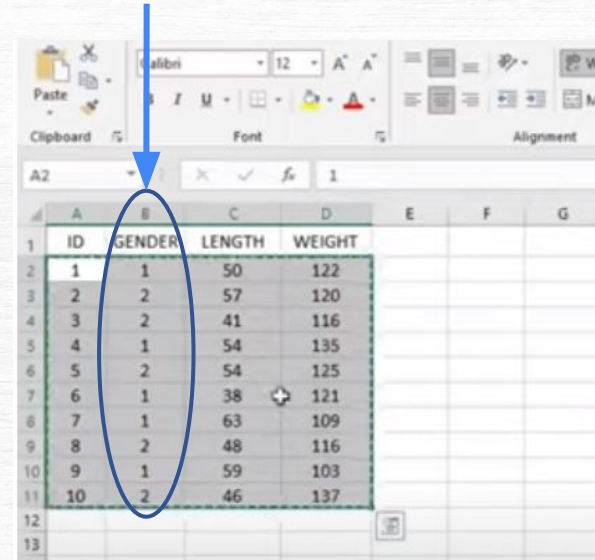
This is where we declare the variable type.
All variables say “Numeric” by default.



	Name	Type	Width	Decimals
1	ID	Numeric	8	2
2	Gender	Numeric	8	2
3	Length	Numeric	8	2
4	Weight	Numeric	8	2
5				
6				

Therefore, we leave all the default “Numeric” types as they are in the SPSS Variable View for now.

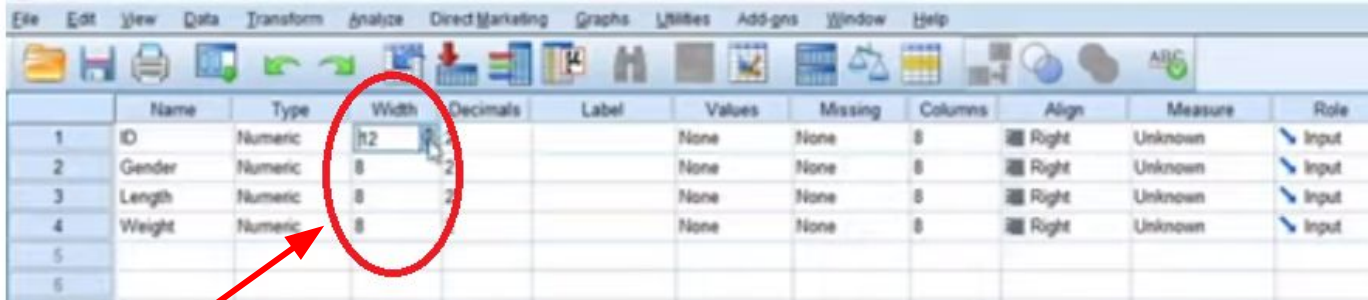
The only non-numeric variable is Gender, but if you recall the Excel data, it is coded as numeric, with 1 = Male and 2 = Female.



	ID	GENDER	LENGTH	WEIGHT
1	1	1	50	122
2	2	2	57	120
3	3	2	41	116
4	4	1	54	135
5	5	2	54	125
6	6	1	38	121
7	7	1	63	109
8	8	2	48	116
9	9	1	59	103
10	10	2	46	137

The “Width” Column

The third column is Width. The width is the number of characters (numbers or letters) in the variable name column.

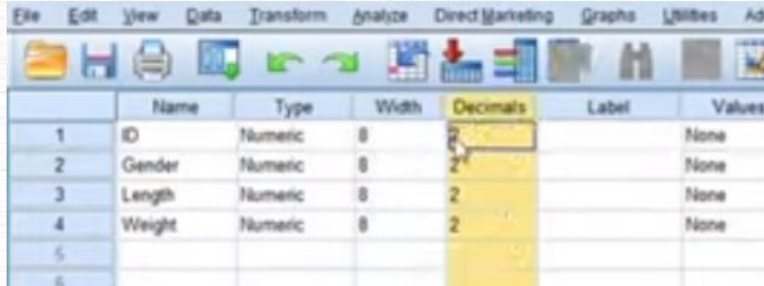


	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	ID	Numeric	8	2		None	None	8	Right	Unknown	Input
2	Gender	Numeric	8	2		None	None	8	Right	Unknown	Input
3	Length	Numeric	8	2		None	None	8	Right	Unknown	Input
4	Weight	Numeric	8	2		None	None	8	Right	Unknown	Input
5											
6											

The default width is 8.

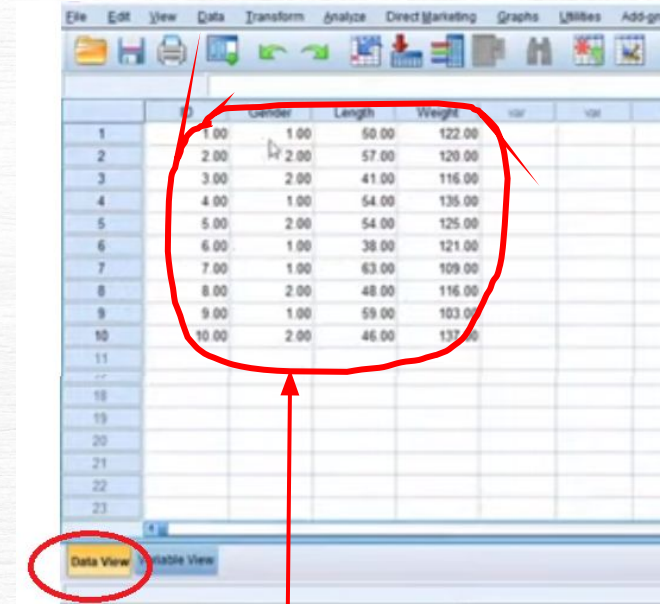
However, we can increase or decrease this value by clicking in the cell and using the down or up arrows that appear to change the values.

The “Decimals” Column



	Name	Type	Width	Decimals	Label	Values
1	ID	Numeric	8	2		None
2	Gender	Numeric	8	2		None
3	Length	Numeric	8	2		None
4	Weight	Numeric	8	2		None
5						
6						

This column gives the number of decimal places for variables that are numeric. The variables are given two decimal places by default.

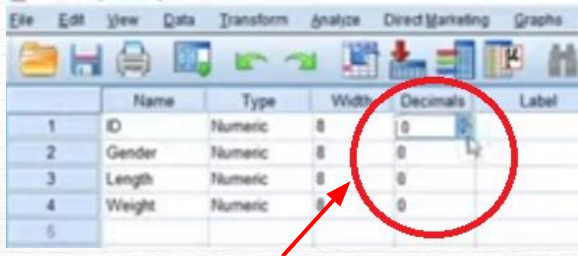


	ID	Gender	Length	Weight	var	var
1	1.00	1.00	50.00	122.00		
2	2.00	2.00	57.00	120.00		
3	3.00	2.00	41.00	116.00		
4	4.00	1.00	54.00	135.00		
5	5.00	2.00	54.00	125.00		
6	6.00	1.00	38.00	121.00		
7	7.00	1.00	63.00	109.00		
8	8.00	2.00	48.00	116.00		
9	9.00	1.00	59.00	103.00		
10	10.00	2.00	46.00	137.00		
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						

If we go to Data View, we will see that all values are expressed in two decimal places.

The “Decimals” Column

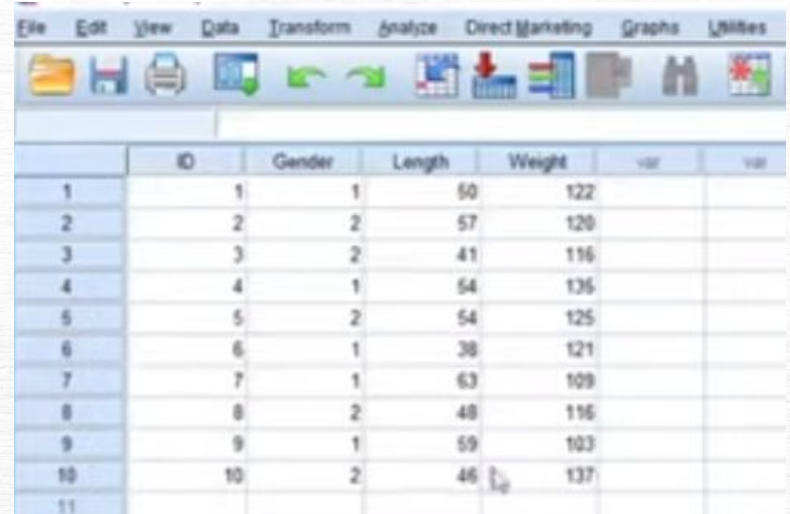
Now, we don't want the values expressed in decimal places, so go back to Variable View.



	Name	Type	Width	Decimals	Label
1	ID	Numeric	8	0	
2	Gender	Numeric	8	0	
3	Length	Numeric	8	0	
4	Weight	Numeric	8	0	
5					

Click on each cell in the Decimal column and change it to zero.

Let us return to Variable View to continue entering the properties of the variables.



	ID	Gender	Length	Weight	var	var
1	1	1	50	122		
2	2	2	57	120		
3	3	2	41	116		
4	4	1	54	135		
5	5	2	54	125		
6	6	1	38	121		
7	7	1	63	109		
8	8	2	48	116		
9	9	1	59	103		
10	10	2	46	137		
11						

If we go to Data View, we can see the results of this change with all numbers being expressed as integers.

The “Label” Column

This is where you enter a description of the variable. This column is one of the most important because what you enter here will show up in all analyses of the data.

	Name	Type	Width	Decimals	Label
1	ID	Numeric	8	0	
2	Gender	Numeric	8	0	
3	Length	Numeric	8	0	
4	Weight	Numeric	8	0	
5					

We change the values as follows:

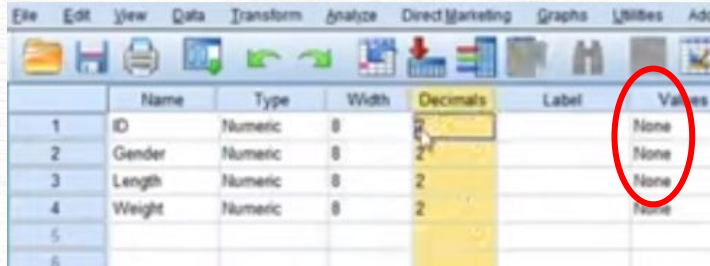
For the first variable, we click in the cell and write “Identification Number” (there can be spaces between words in the “Label”).

For the second variable, we write “Gender.”

We write “Length of fish (cm)” for the third variable and “Weight of fish (kg)” for the fourth.

The screenshot shows the SPSS Data Editor interface. The 'Data' menu is highlighted in the top menu bar. The 'Data' tab is selected, showing a list of variables: ID, Gender, Length, and Weight. The 'Label' column for the 'Weight' variable is highlighted with a red circle. The 'Data' menu is also highlighted in the top menu bar.

The “Values” Column



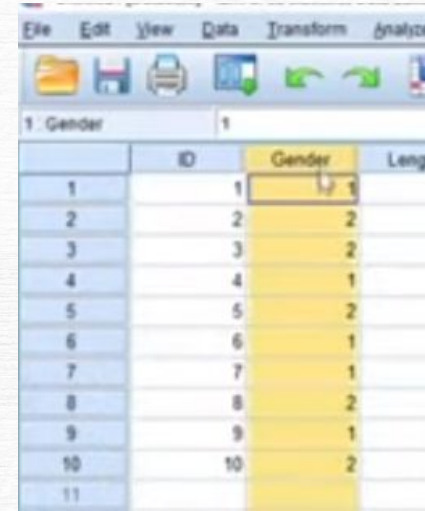
	Name	Type	Width	Decimals	Label	Values
1	ID	Numeric	8	2		None
2	Gender	Numeric	8	2		None
3	Length	Numeric	8	2		None
4	Weight	Numeric	8	2		None
5						
6						

The Values column is where you assign values to categorical and ordinal variables.

The only categorical variable we have is Gender.

So, we go to Data View to see how many groups we have in the Gender Variable.

We see there are only two groups: 1 and 2.

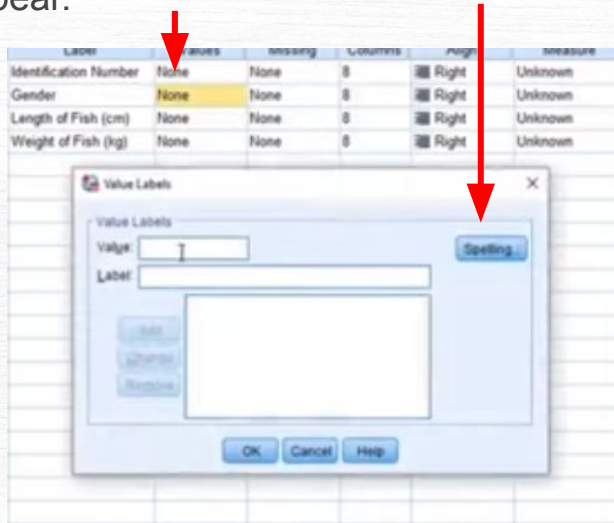


	ID	Gender	Length
1	1	1	
2	2	2	
3	3	2	
4	4	1	
5	5	2	
6	6	1	
7	7	1	
8	8	2	
9	9	1	
10	10	2	
11			

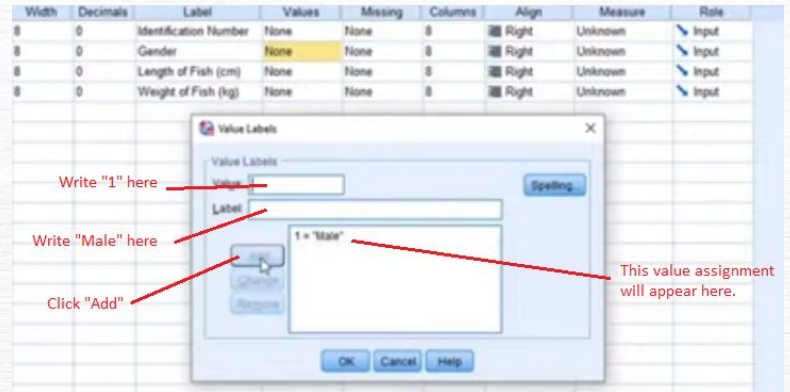
We go back to Variable View and make the changes.

The “Values” Column

We click on the cell of the Gender variable; click on the blue arrow, and a dialog box will appear.



To make the assignments, proceed as follows.



The “Values” Column

Make the second assignment by writing “2” in the Value box and “Female” in the Label box.

Click “OK.” Now, let us go to Data View to see the results of our assignments.

Width	Decimals	Label	Values	Missing	Columns	Align	Measure
8	0	Identification Number	None	None	8	Right	Unknown
8	0	Gender	None	None	8	Right	Unknown
8	0	Length of Fish (cm)	None	None	8	Right	Unknown
8	0	Weight of Fish (kg)	None	None	8	Right	Unknown

Value Labels

Value:

Label:

1 = "Male"

2 = "Female"

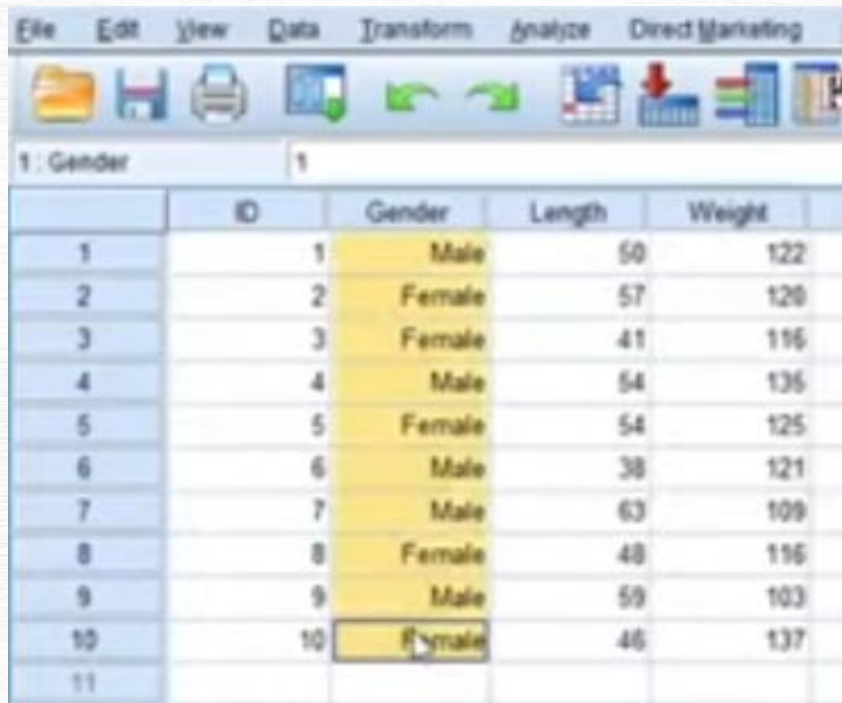
OK

Cancel

Help

The two Gender assignments added

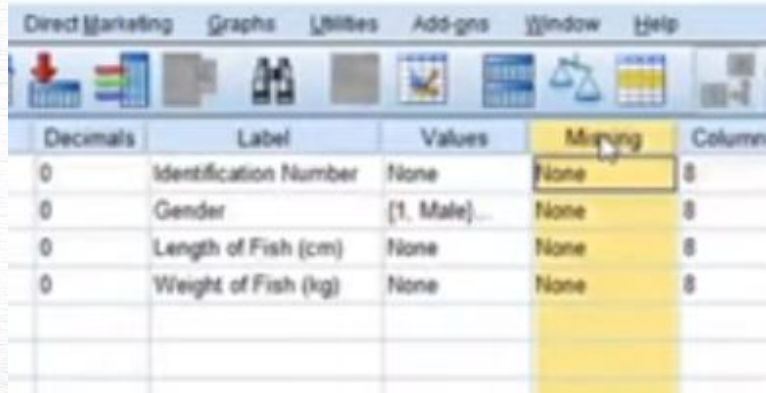
The “Values” Column



1: Gender 1

	ID	Gender	Length	Weight
1	1	Male	60	122
2	2	Female	57	120
3	3	Female	41	116
4	4	Male	64	136
5	5	Female	54	125
6	6	Male	38	121
7	7	Male	63	109
8	8	Female	48	116
9	9	Male	59	103
10	10	Female	46	137
11				

The “Missing” Column

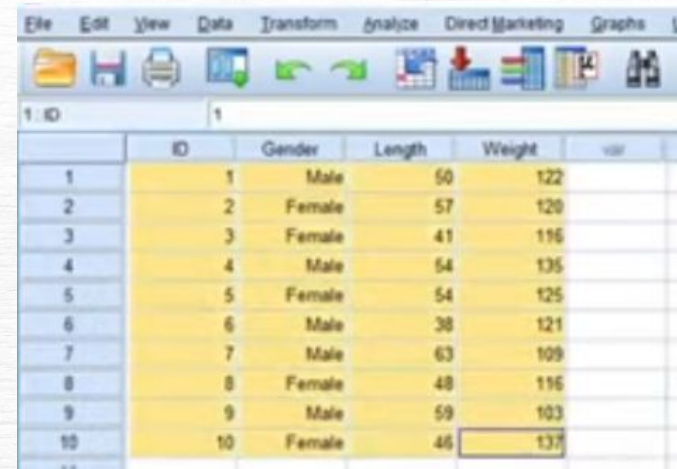


This screenshot shows a software interface with a menu bar (Direct Marketing, Graphs, Utilities, Add-ons, Window, Help) and a toolbar. Below is a table with five columns: Decimals, Label, Values, Missing, and Columns. The 'Missing' column is highlighted in yellow, and a mouse cursor is pointing at the 'None' value in the first row.

Decimals	Label	Values	Missing	Columns
0	Identification Number	None	None	8
0	Gender	{1, Male}...	None	8
0	Length of Fish (cm)	None	None	8
0	Weight of Fish (kg)	None	None	8

This is where we indicate whether there are missing values in any of the variables.

There are no missing values, as we can see from the Data View.

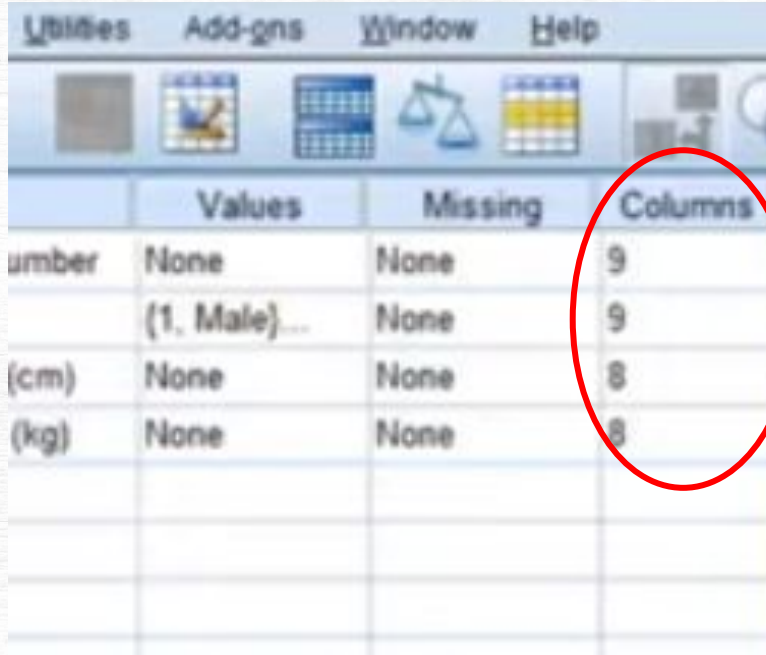


This screenshot shows a software interface with a menu bar (File, Edit, View, Data, Transform, Analyze, Direct Marketing, Graphs) and a toolbar. Below is a table with six columns: ID, Gender, Length, Weight, and var. The table contains 10 rows of data.

	ID	Gender	Length	Weight	var
1	1	Male	50	122	
2	2	Female	57	120	
3	3	Female	41	116	
4	4	Male	54	135	
5	5	Female	54	125	
6	6	Male	38	121	
7	7	Male	63	109	
8	8	Female	48	116	
9	9	Male	59	103	
10	10	Female	46	133	

Therefore, we move on to the next column, “Column.”

The “Columns” Column

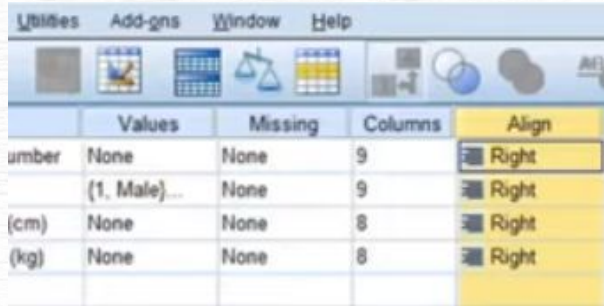


	Values	Missing	Columns
umber	None	None	9
	(1, Male)...	None	9
(cm)	None	None	8
(kg)	None	None	8

You can increase or decrease the width of each variable column in Data View by entering a value in “Column” of Variable View.

The application of the property is self-explanatory, so we move on to the next column.

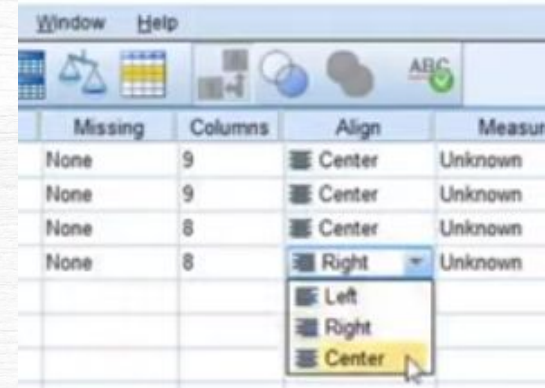
The “Align” Column



This is where you tell SPSS how you want your data arranged in each of the variable columns in Data View. Default is “Right.” This means that all data are right-aligned in Data View.

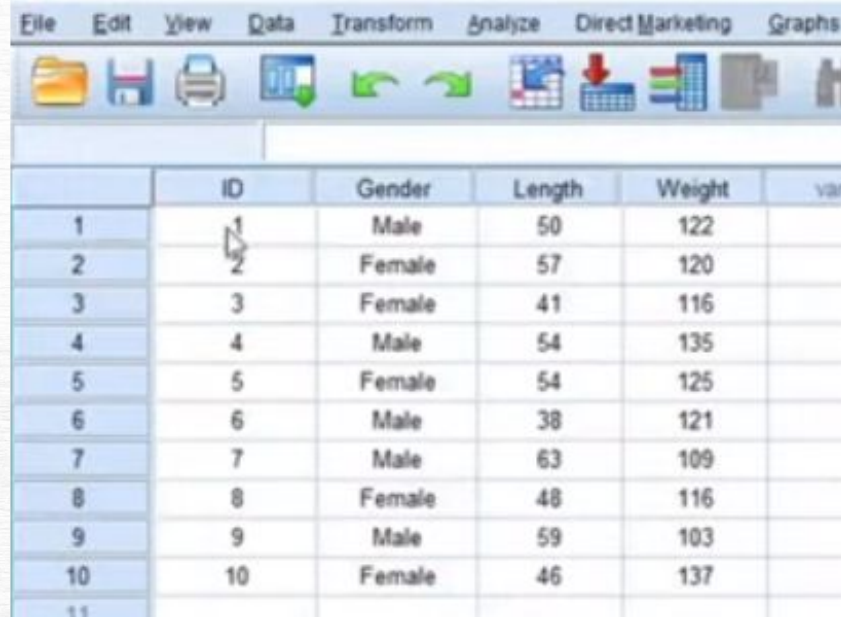
Now, let us change the alignment of the data in all variables to “**Center.**”

To do this, we simply click on the right edge of the cell of each variable and choose “Center” from the drop-down list that appears.



Now, let us go to Data View to see the results of these changes.

The “Align” Column

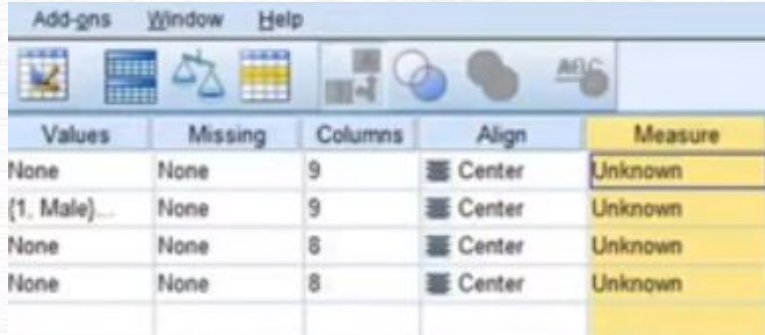


The screenshot shows a software window with a menu bar (File, Edit, View, Data, Transform, Analyze, Direct Marketing, Graphs) and a toolbar with various icons. Below the toolbar is a data table with 6 columns: an unlabeled index column, ID, Gender, Length, Weight, and var. The data is center-aligned.

	ID	Gender	Length	Weight	var
1	1	Male	50	122	
2	2	Female	57	120	
3	3	Female	41	116	
4	4	Male	54	135	
5	5	Female	54	125	
6	6	Male	38	121	
7	7	Male	63	109	
8	8	Female	48	116	
9	9	Male	59	103	
10	10	Female	46	137	

We see that the data are all center-aligned.

The “Measure” Column



Values	Missing	Columns	Align	Measure
None	None	9	Center	Unknown
(1, Male)...	None	9	Center	Unknown
None	None	8	Center	Unknown
None	None	8	Center	Unknown

This is where you declare the level of measurement of each variable, an important property. The default is “Unknown.”

The three levels of measurement used in SPSS are: Scale, Nominal, and Ordinal.

Scale = Numeric variables

Nominal = Categorical unordered (Red, Blue, Purple)

Ordinal = Categorical ordered (Large, Medium, Small)

The “Measure” Column

The first variable, "ID," is Nominal because it is a categorical variable that does not represent any ordering or ranking.

The variable “Gender” is also Nominal because it is a categorical variable that does not represent any ordering or ranking: (Male, Female).

The variables “Length” and “Width” are both numeric. Therefore, we choose “Scale.”

[illegible]

The “Role” Column



Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
ID	Numeric	8	0	Identification N...	None	None	12	Center	Nominal	Input
GENDER	Numeric	8	0	Gender	{1, Male}...	None	12	Center	Nominal	Input
LENGTH	Numeric	8	0	Length of Fish (...)	None	None	12	Center	Scale	Input
WEIGHT	Numeric	8	0	Weight of Fish ...	None	None	12	Center	Scale	Input

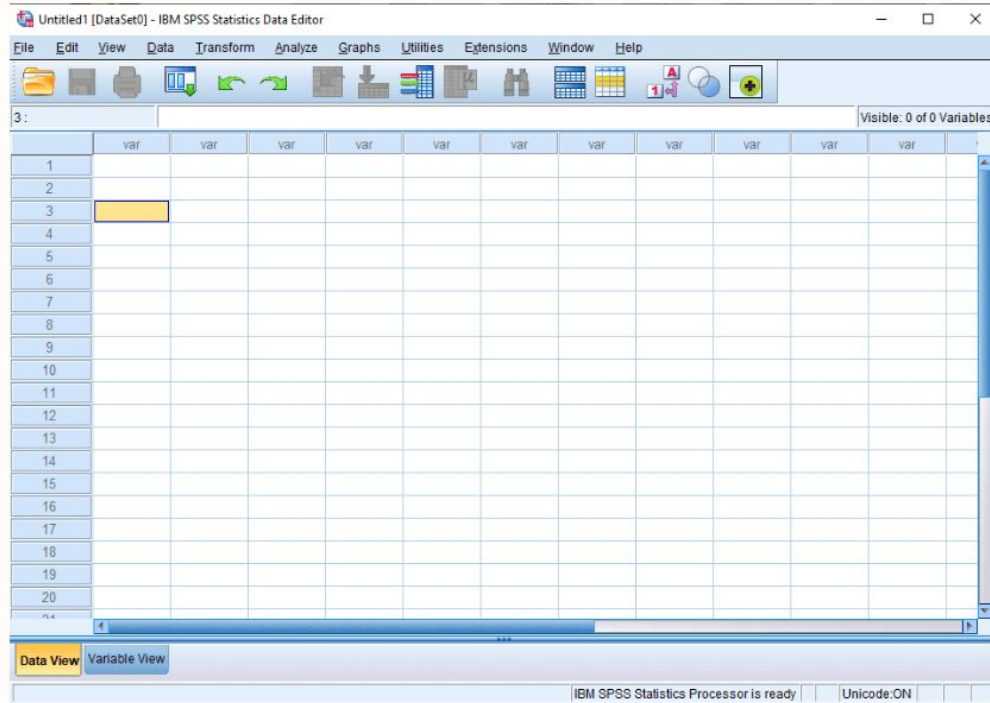
We will leave this property as the default “Input” for all variables.

We have now declared and defined all properties.



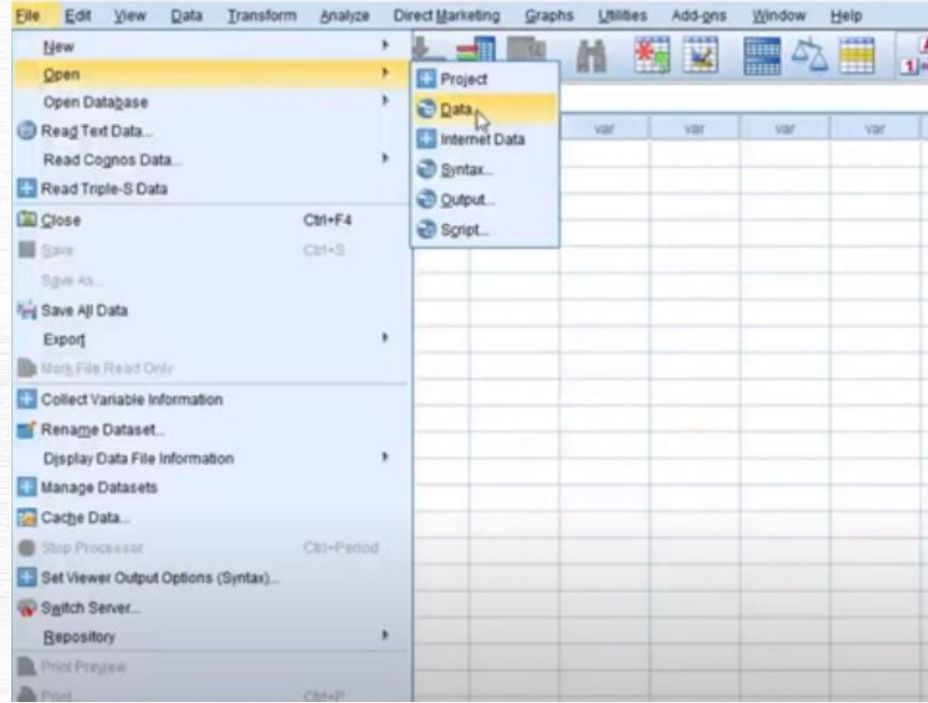
Loading Data into SPSS from Excel: Method 2

We begin with a blank Data Editor window.

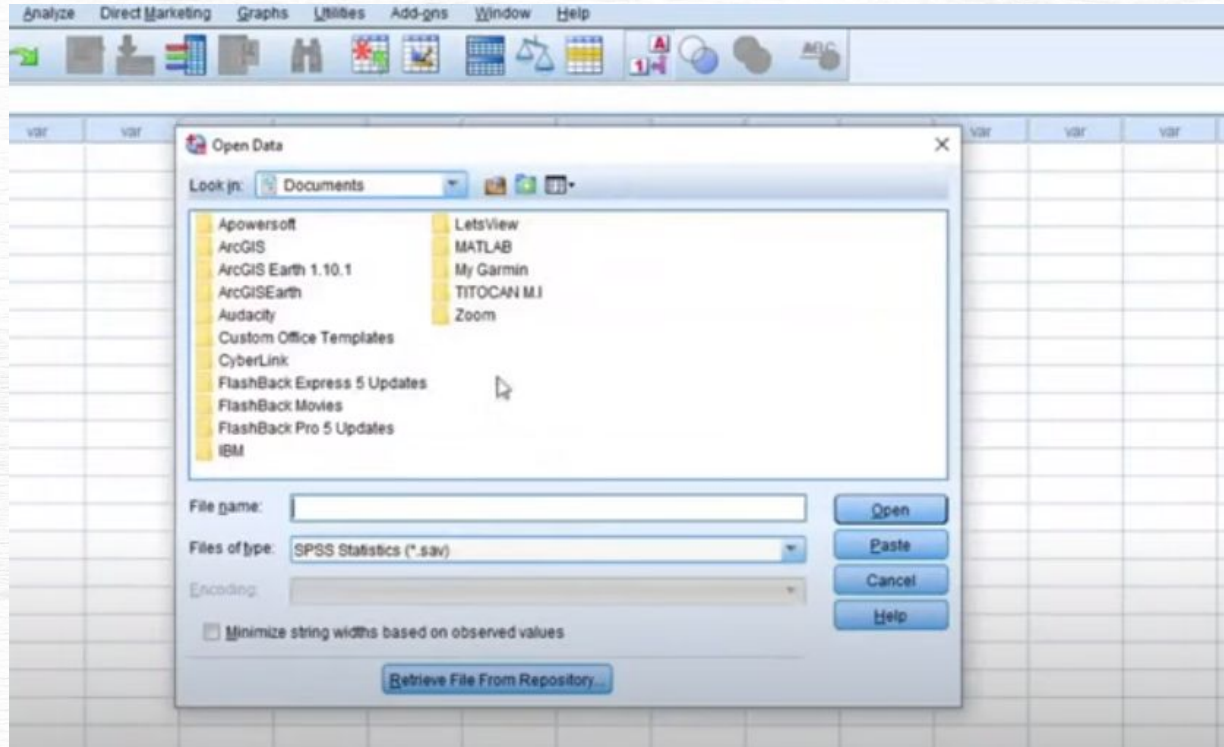


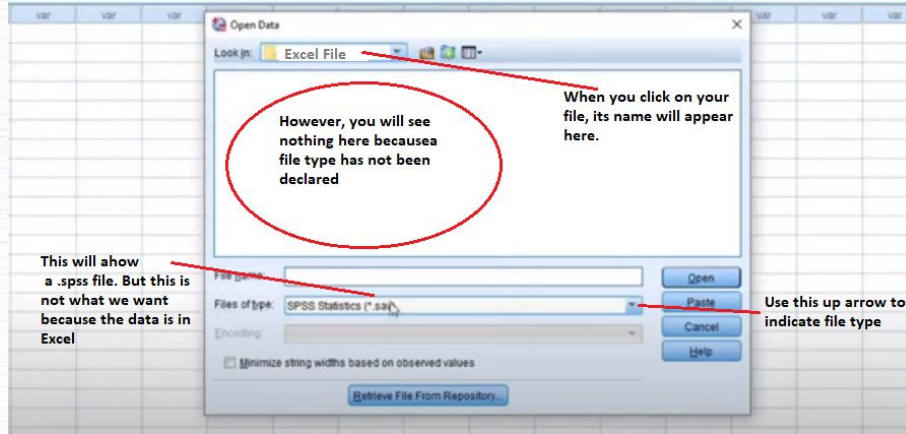
SPSS Data editor window

In the SPSS main interface, click on File Menu > Open > Data.

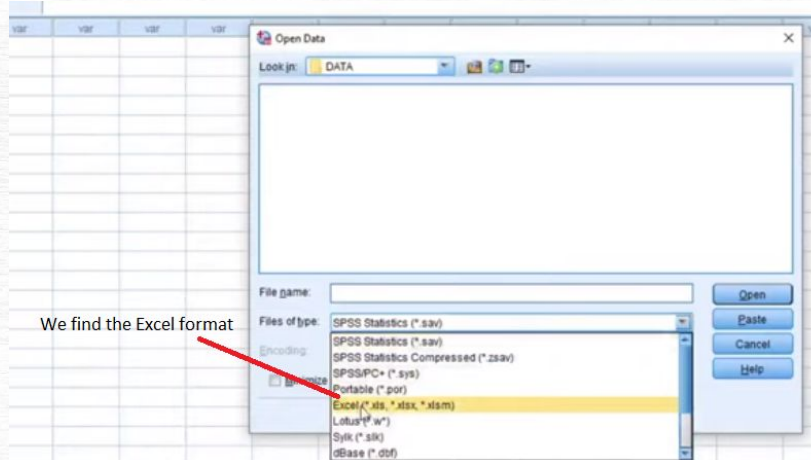


The following dialog box will appear.

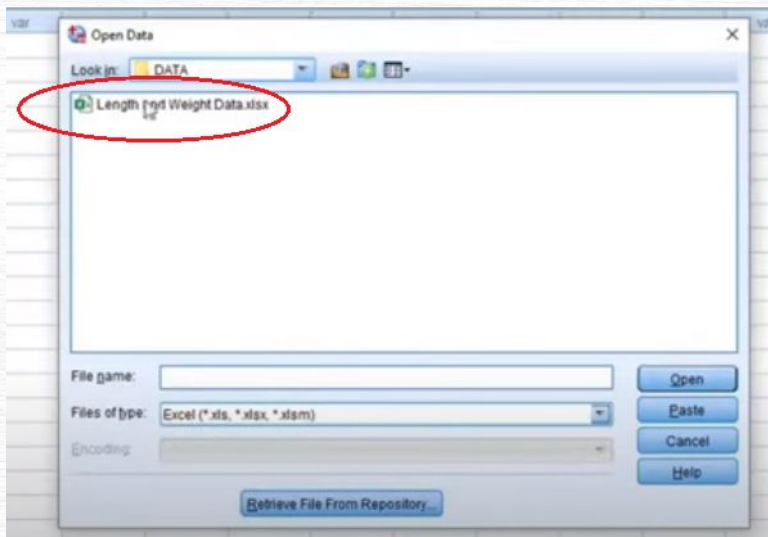




Here, you navigate your computer to the folder in which your Excel file is saved.

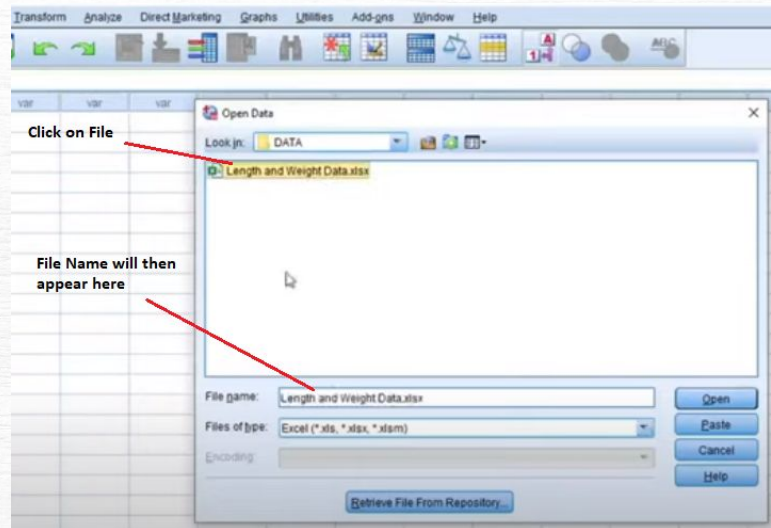


And immediately, our file appears in the open space.

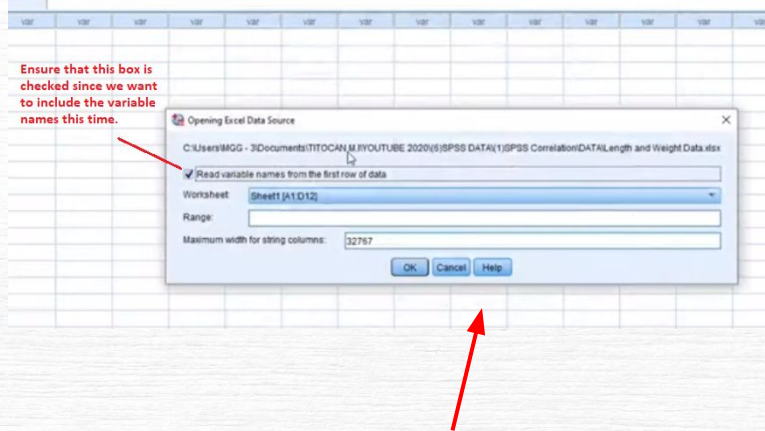


Now, click on the file to highlight it

And its name will appear in the “File Name” box.

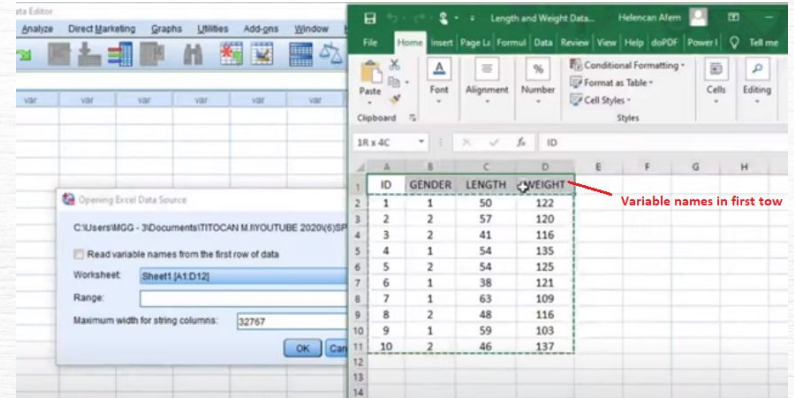


Click "Open," and the following dialog box appears.

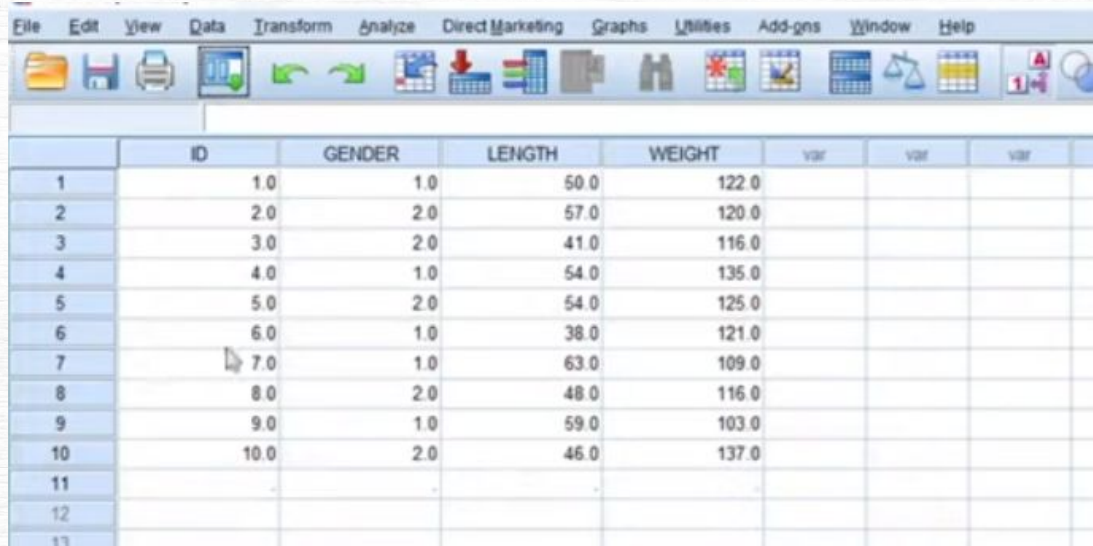


After clicking the checkbox, we click "OK" in the blue dialog box.

Here, we must let SPSS know whether the first row contains the variable names.



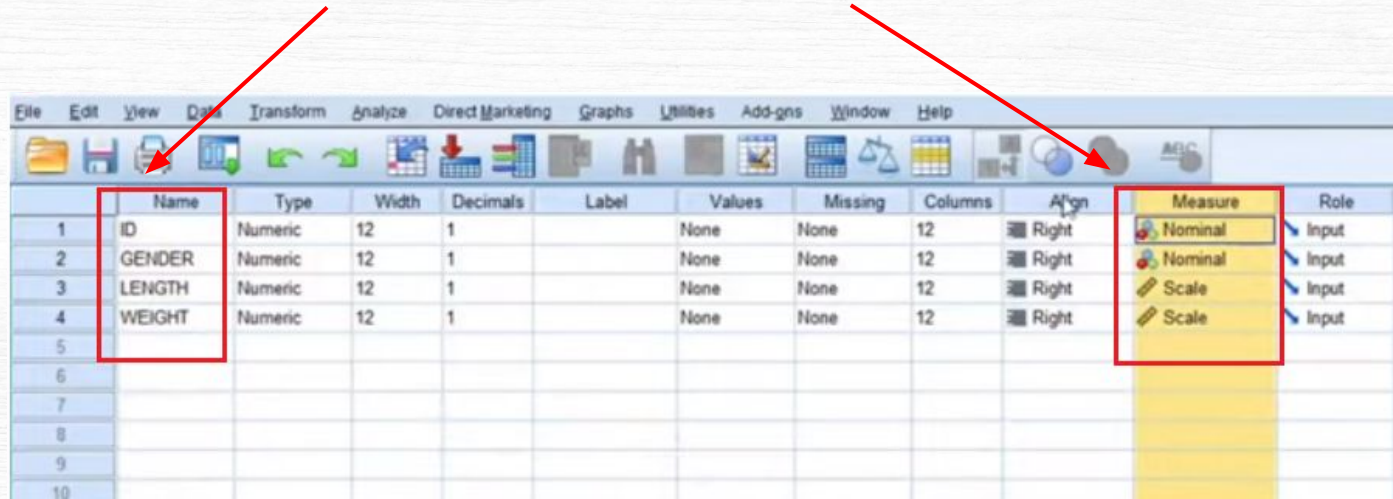
The dataset will be loaded into SPSS with the correct variable names, not the default. This is shown in Data View below.



	ID	GENDER	LENGTH	WEIGHT	var	var	var
1	1.0	1.0	50.0	122.0			
2	2.0	2.0	57.0	120.0			
3	3.0	2.0	41.0	116.0			
4	4.0	1.0	54.0	135.0			
5	5.0	2.0	54.0	125.0			
6	6.0	1.0	38.0	121.0			
7	7.0	1.0	63.0	109.0			
8	8.0	2.0	48.0	116.0			
9	9.0	1.0	59.0	103.0			
10	10.0	2.0	46.0	137.0			
11	-	-	-	-			
12							
13							

Now, let us go into Variable View to examine our data set.

As we can see, the columns for “Name” and “Measure” have already been filled with the correct properties.



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	ID	Numeric	12	1		None	None	12	Right	Nominal	Input
2	GENDER	Numeric	12	1		None	None	12	Right	Nominal	Input
3	LENGTH	Numeric	12	1		None	None	12	Right	Scale	Input
4	WEIGHT	Numeric	12	1		None	None	12	Right	Scale	Input
5											
6											
7											
8											
9											
10											

We, therefore, would now fill in the properties for the remaining columns as we did in Method 1.



**End of
Presentation**