












60080079 Introduction to Statistical Methods
Semester 2 2023-2024
Handout 1_2

Describing Some of the High School and Beyond Data Variables

1. Opening file in SPSS

The **Data View** tab lists the cases (rows) and the variables (columns) in the file.

hsb.sav [DataSet1] - IBM SPSS Statistics Data Editor

File	Edit	View	Data	Transform	Analyze	Graphs	Utilities	Add-ons	Window	Help
										

	ID	SEX	RACE	SES	SCH_TYPE	HSP	LOC_CON	SE
1	1.00	2.00	1.00	1.00	1.00	3.00	.29	
2	2.00	1.00	1.00	1.00	1.00	1.00	-.42	
3	3.00	2.00	1.00	1.00	1.00	1.00	.71	
4	4.00	2.00	1.00	2.00	1.00	3.00	.06	
5	5.00	2.00	1.00	2.00	1.00	3.00	.22	
6	6.00	1.00	1.00	2.00	1.00	1.00	.46	
7	7.00	1.00	1.00	1.00	1.00	2.00	.44	
8	8.00	2.00	1.00	1.00	1.00	2.00	.68	
9	9.00	1.00	1.00	2.00	1.00	2.00	.06	
10	10.00	2.00	1.00	1.00	1.00	2.00	.05	
11	11.00	2.00	1.00	1.00	1.00	2.00	.25	
12	12.00	1.00	1.00	1.00	1.00	1.00	.06	
13	13.00	1.00	1.00	1.00	1.00	2.00	-1.10	
14	14.00	2.00	1.00	2.00	1.00	1.00	.52	
15	15.00	1.00	1.00	2.00	1.00	3.00	-1.28	
16	16.00	2.00	1.00	1.00	1.00	2.00	.06	
17	17.00	2.00	1.00	3.00	1.00	2.00	.26	
18	18.00	1.00	1.00	2.00	1.00	3.00	.21	
19	19.00	1.00	1.00	2.00	1.00	2.00	.06	
20	20.00	1.00	1.00	3.00	1.00	2.00	.28	
21	21.00	1.00	1.00	2.00	1.00	1.00	-.44	

1

Data View Variable View

The **Variable View** tab lists the variables (rows) and their properties (columns).

hsb.sav [DataSet1] - IBM SPSS Statistics Data Editor

	Name	Type	Width	Decimals	Label	Values	Missing	
1	ID	Numeric	8	2	Student Identifi...	None	None	8
2	SEX	Numeric	8	2	Sex	{1.00, Male}...	None	8
3	RACE	Numeric	8	2	Race or ethnicity	{1.00, Hispa...	None	8
4	SES	Numeric	8	2	Socioeconomic ...	{1.00, Low}...	None	8
5	SCH_TYPE	Numeric	8	2	School type	{1.00, Publi...	None	8
6	HSP	Numeric	8	2	High school pro...	{1.00, Gene...	None	8
7	LOC_CON	Numeric	8	2	Locus of control	None	None	8
8	SEL_CON	Numeric	8	2	Self-concept	None	None	8
9	MOTIV	Numeric	8	2	Motivation	None	None	8
10	C_CHOICE	Numeric	8	2	Career choice	{1.00, Cleric...	None	8
11	READING	Numeric	8	2	Reading score	None	None	8
12	WRITING	Numeric	8	2	Writing score	None	None	8
13	MATH	Numeric	8	2	Math score	None	None	8
14	SCIENCE	Numeric	8	2	Science score	None	None	8
15	CIVICS	Numeric	8	2	Civics score	None	None	8
16								
17								
18								
19								
20								
21								
22								

Data View Variable View

2. Pie Chart: Sex

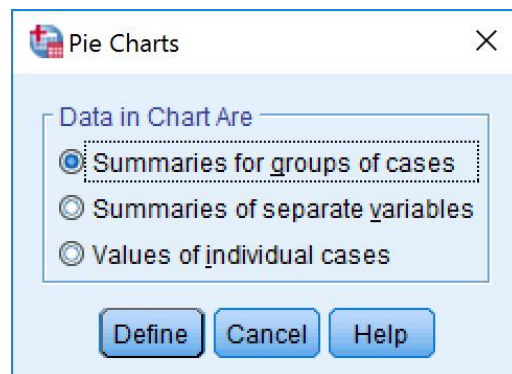
Graphs → Legacy Dialogs → Pie...

hsb.sav [DataSet1] - IBM SPSS Statistics Data Editor

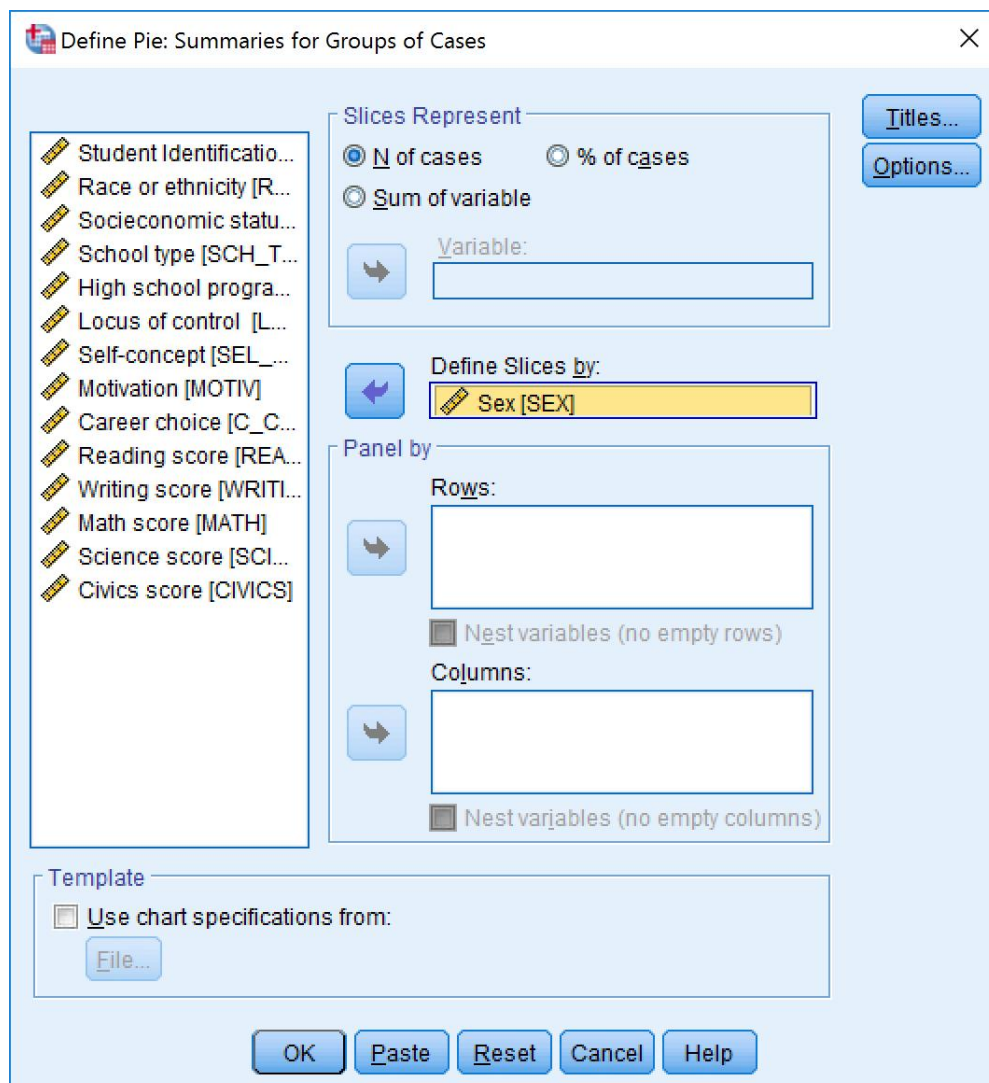
	ID	SEX	RACE	SES	SCH_TYPE	HSP
1	1.00	2.00	1.00	1.00	1.00	3.00
2	2.00	1.00	1.00	1.00	1.00	1.00
3	3.00	2.00	1.00	1.00	1.00	1.00
4	4.00	2.00	1.00	2.00	1.00	3.00
5	5.00	2.00	1.00	2.00	1.00	3.00
6	6.00	1.00	1.00	2.00	1.00	1.00
7	7.00	1.00	1.00	1.00	1.00	2.00
8	8.00	2.00	1.00	1.00	1.00	2.00
9	9.00	1.00	1.00	2.00	1.00	2.00
10	10.00	2.00	1.00	1.00	1.00	2.00
11	11.00	2.00	1.00	1.00	1.00	2.00

Graphs → Legacy Dialogs → Pie...

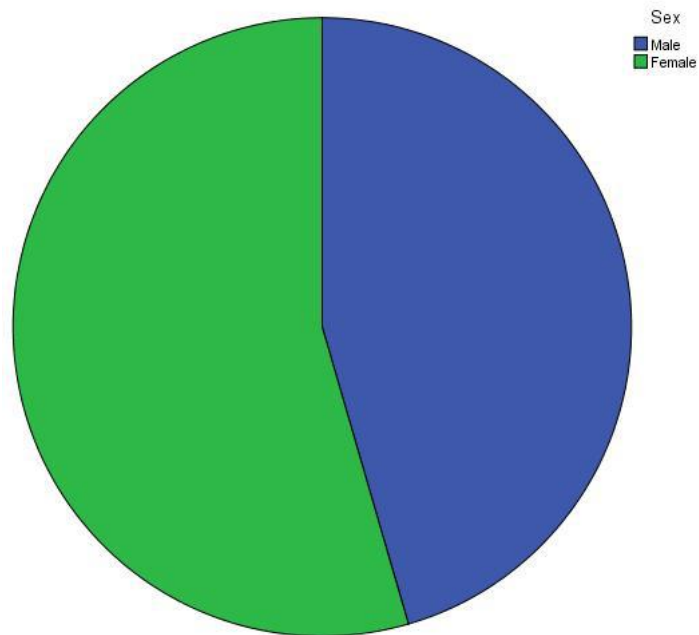
Choose **Summaries for groups of cases** (default) under **Data in Chart Are**, then click **Define**.



Click **Sex** into the **Define Slices by** box, the hit **OK**.



You should get the pie chart below.



3. Bar Graph: Socio-Economic Status

Graphs → Legacy Dialogs → Bar...

hsb.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze **Graphs** Utilities Add-ons Window Help

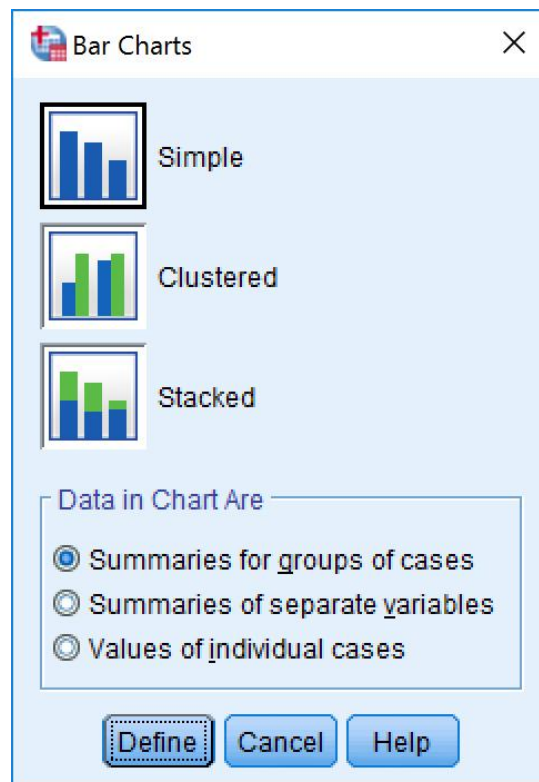
Chart Builder...
Graphboard Template Chooser...

Legacy Dialogs

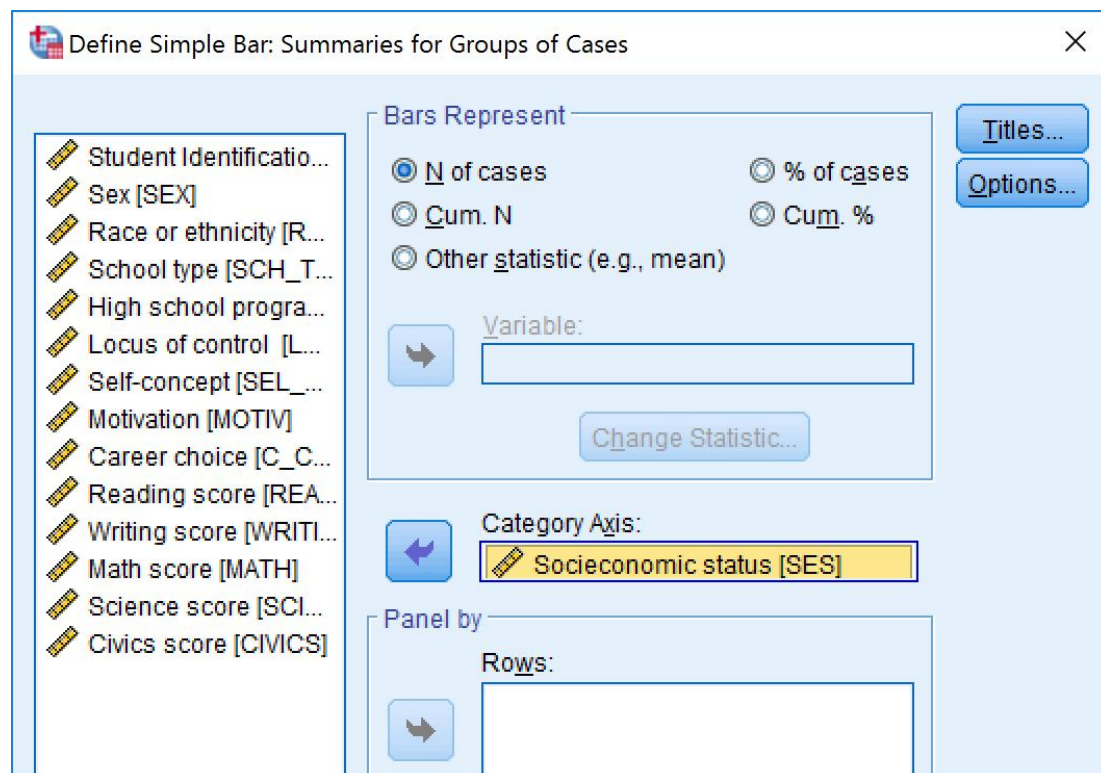
- Bar...**
- 3-D Bar...
- Line...
- Area...
- Pie...
- High-Low...
- Boxplot...
- Error Bar...
- Population Pyramid...
- Scatter/Dot...
- Histogram...

	ID	SEX	RACE	SES	SCH_TYPE	HSP
1	1.00	2.00	1.00	1.00	1.00	3.00
2	2.00	1.00	1.00	1.00	1.00	1.00
3	3.00	2.00	1.00	1.00	1.00	1.00
4	4.00	2.00	1.00	2.00	1.00	3.00
5	5.00	2.00	1.00	2.00	1.00	3.00
6	6.00	1.00	1.00	2.00	1.00	1.00
7	7.00	1.00	1.00	1.00	1.00	2.00
8	8.00	2.00	1.00	1.00	1.00	2.00
9	9.00	1.00	1.00	2.00	1.00	2.00
10	10.00	2.00	1.00	1.00	1.00	2.00
11	11.00	2.00	1.00	1.00	1.00	2.00
12	12.00	1.00	1.00	1.00	1.00	1.00
13	13.00	1.00	1.00	1.00	1.00	2.00
14	14.00	1.00	1.00	1.00	1.00	1.00
15	15.00	1.00	1.00	1.00	1.00	1.00
16	16.00	1.00	1.00	1.00	1.00	1.00
17	17.00	1.00	1.00	1.00	1.00	1.00
18	18.00	1.00	1.00	1.00	1.00	1.00
19	19.00	1.00	1.00	1.00	1.00	1.00
20	20.00	1.00	1.00	1.00	1.00	1.00
21	21.00	1.00	1.00	1.00	1.00	1.00
22	22.00	1.00	1.00	1.00	1.00	1.00
23	23.00	1.00	1.00	1.00	1.00	1.00
24	24.00	1.00	1.00	1.00	1.00	1.00
25	25.00	1.00	1.00	1.00	1.00	1.00
26	26.00	1.00	1.00	1.00	1.00	1.00
27	27.00	1.00	1.00	1.00	1.00	1.00
28	28.00	1.00	1.00	1.00	1.00	1.00
29	29.00	1.00	1.00	1.00	1.00	1.00
30	30.00	1.00	1.00	1.00	1.00	1.00
31	31.00	1.00	1.00	1.00	1.00	1.00
32	32.00	1.00	1.00	1.00	1.00	1.00
33	33.00	1.00	1.00	1.00	1.00	1.00
34	34.00	1.00	1.00	1.00	1.00	1.00
35	35.00	1.00	1.00	1.00	1.00	1.00
36	36.00	1.00	1.00	1.00	1.00	1.00
37	37.00	1.00	1.00	1.00	1.00	1.00
38	38.00	1.00	1.00	1.00	1.00	1.00
39	39.00	1.00	1.00	1.00	1.00	1.00
40	40.00	1.00	1.00	1.00	1.00	1.00
41	41.00	1.00	1.00	1.00	1.00	1.00
42	42.00	1.00	1.00	1.00	1.00	1.00
43	43.00	1.00	1.00	1.00	1.00	1.00
44	44.00	1.00	1.00	1.00	1.00	1.00
45	45.00	1.00	1.00	1.00	1.00	1.00
46	46.00	1.00	1.00	1.00	1.00	1.00
47	47.00	1.00	1.00	1.00	1.00	1.00
48	48.00	1.00	1.00	1.00	1.00	1.00
49	49.00	1.00	1.00	1.00	1.00	1.00
50	50.00	1.00	1.00	1.00	1.00	1.00
51	51.00	1.00	1.00	1.00	1.00	1.00
52	52.00	1.00	1.00	1.00	1.00	1.00
53	53.00	1.00	1.00	1.00	1.00	1.00
54	54.00	1.00	1.00	1.00	1.00	1.00
55	55.00	1.00	1.00	1.00	1.00	1.00
56	56.00	1.00	1.00	1.00	1.00	1.00
57	57.00	1.00	1.00	1.00	1.00	1.00
58	58.00	1.00	1.00	1.00	1.00	1.00
59	59.00	1.00	1.00	1.00	1.00	1.00
60	60.00	1.00	1.00	1.00	1.00	1.00
61	61.00	1.00	1.00	1.00	1.00	1.00
62	62.00	1.00	1.00	1.00	1.00	1.00
63	63.00	1.00	1.00	1.00	1.00	1.00
64	64.00	1.00	1.00	1.00	1.00	1.00
65	65.00	1.00	1.00	1.00	1.00	1.00
66	66.00	1.00	1.00	1.00	1.00	1.00
67	67.00	1.00	1.00	1.00	1.00	1.00
68	68.00	1.00	1.00	1.00	1.00	1.00
69	69.00	1.00	1.00	1.00	1.00	1.00
70	70.00	1.00	1.00	1.00	1.00	1.00
71	71.00	1.00	1.00	1.00	1.00	1.00
72	72.00	1.00	1.00	1.00	1.00	1.00
73	73.00	1.00	1.00	1.00	1.00	1.00
74	74.00	1.00	1.00	1.00	1.00	1.00
75	75.00	1.00	1.00	1.00	1.00	1.00
76	76.00	1.00	1.00	1.00	1.00	1.00
77	77.00	1.00	1.00	1.00	1.00	1.00
78	78.00	1.00	1.00	1.00	1.00	1.00
79	79.00	1.00	1.00	1.00	1.00	1.00
80	80.00	1.00	1.00	1.00	1.00	1.00
81	81.00	1.00	1.00	1.00	1.00	1.00
82	82.00	1.00	1.00	1.00	1.00	1.00
83	83.00	1.00	1.00	1.00	1.00	1.00
84	84.00	1.00	1.00	1.00	1.00	1.00
85	85.00	1.00	1.00	1.00	1.00	1.00
86	86.00	1.00	1.00	1.00	1.00	1.00
87	87.00	1.00	1.00	1.00	1.00	1.00
88	88.00	1.00	1.00	1.00	1.00	1.00
89	89.00	1.00	1.00	1.00	1.00	1.00
90	90.00	1.00	1.00	1.00	1.00	1.00
91	91.00	1.00	1.00	1.00	1.00	1.00
92	92.00	1.00	1.00	1.00	1.00	1.00
93	93.00	1.00	1.00	1.00	1.00	1.00
94	94.00	1.00	1.00	1.00	1.00	1.00
95	95.00	1.00	1.00	1.00	1.00	1.00
96	96.00	1.00	1.00	1.00	1.00	1.00
97	97.00	1.00	1.00	1.00	1.00	1.00
98	98.00	1.00	1.00	1.00	1.00	1.00
99	99.00	1.00	1.00	1.00	1.00	1.00
100	100.00	1.00	1.00	1.00	1.00	1.00

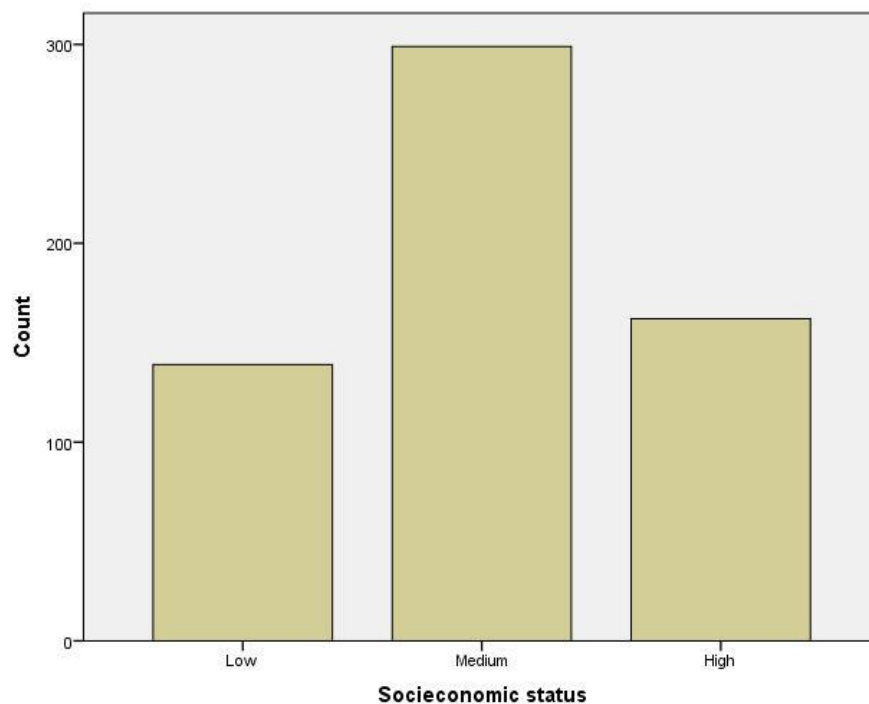
Choose **Simple** (default) under **Bar Charts** and **Summaries for groups of cases** under **Data in the Chart Are**, then click **Define**.



Click **Socioeconomic status** into the **Category Axis** box, then hit **OK**.

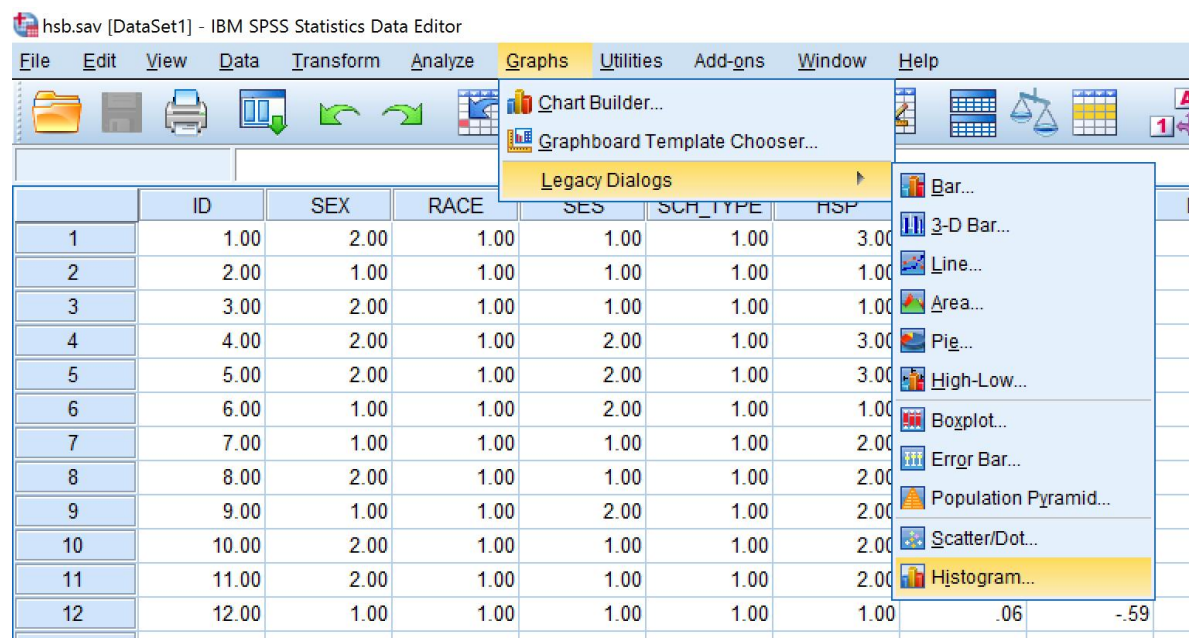


You should get the following bar chart:

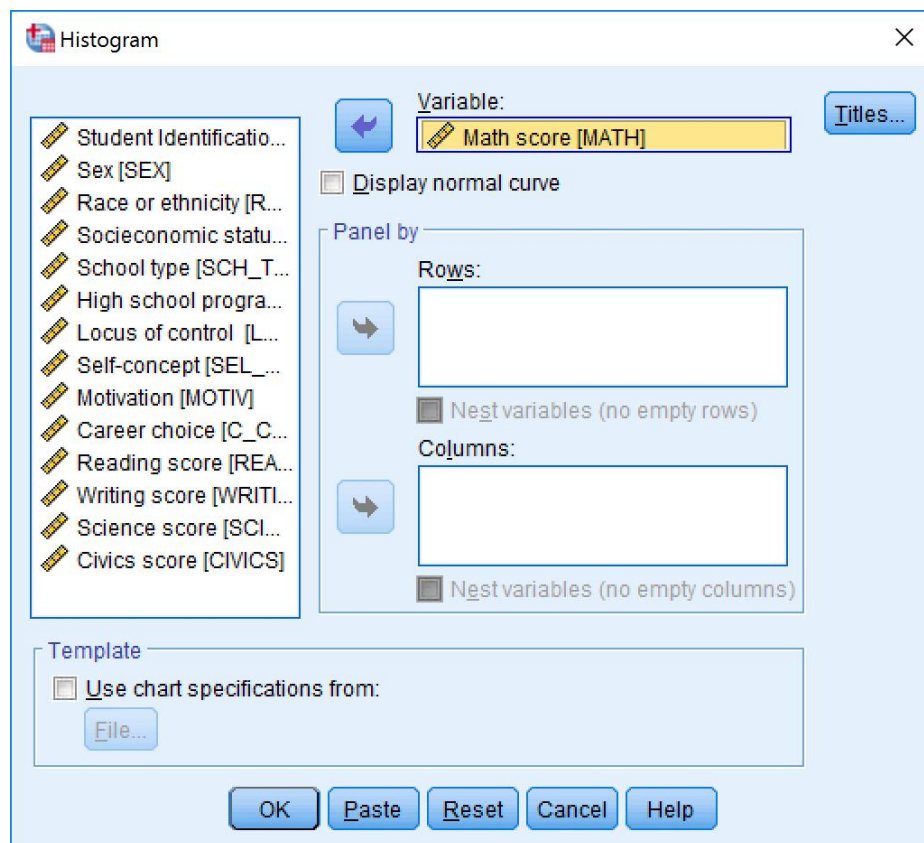


4. Histogram: Math Score

Graphs → Legacy Dialogs → Histogram...

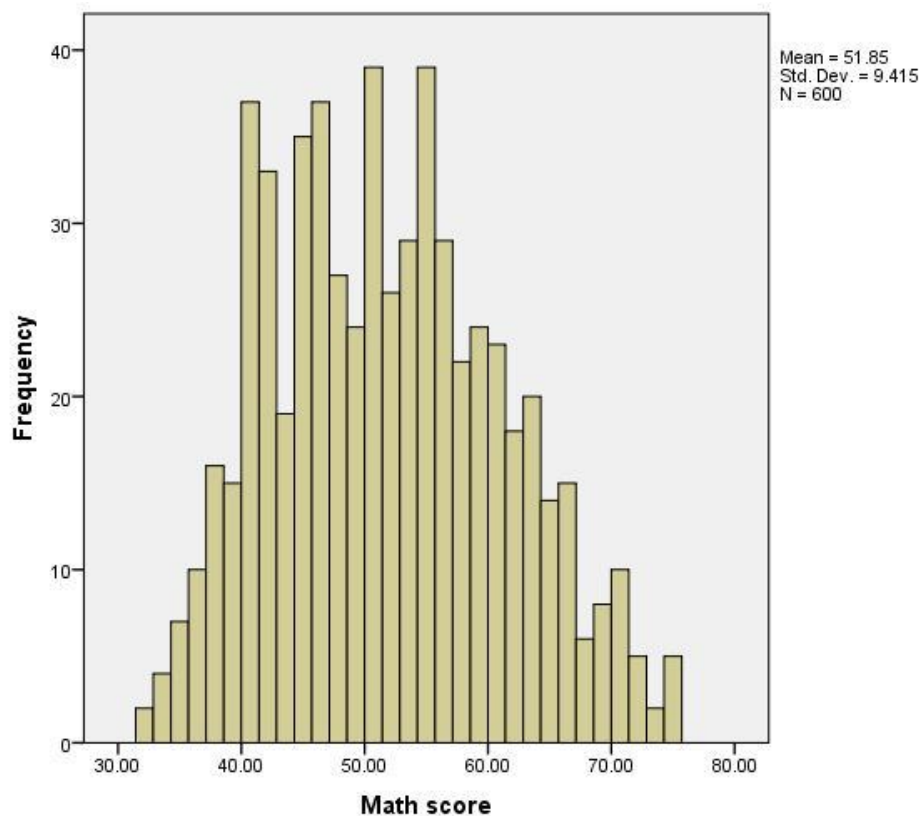


Click **Math score** into the **Variable** box, then hit OK.



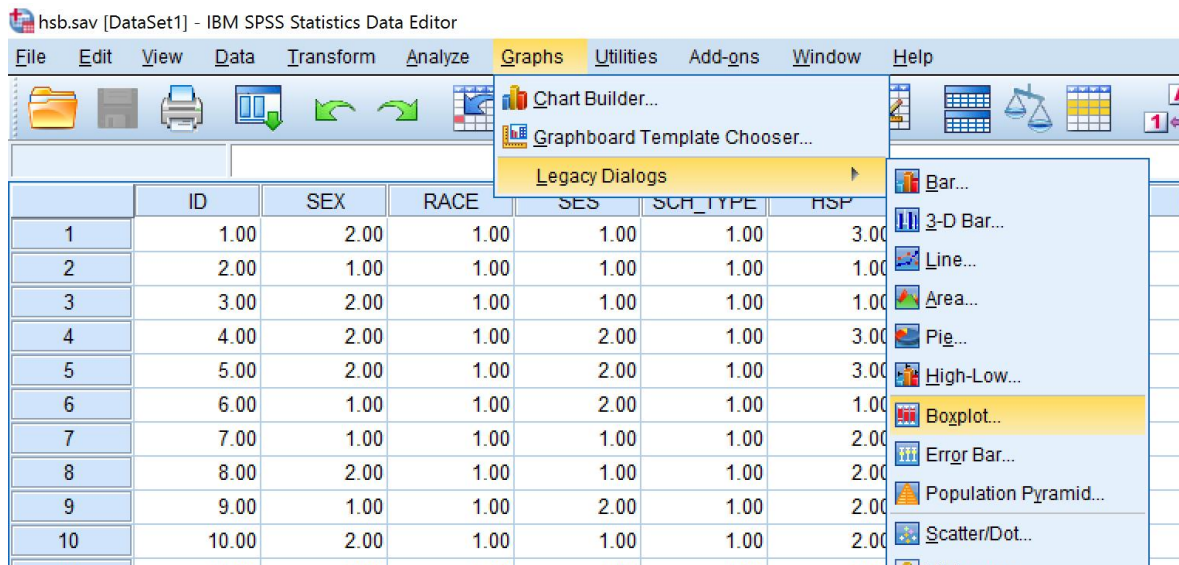
The image shows the SPSS Histogram dialog box. On the left is a list of variables: Student Identification..., Sex [SEX], Race or ethnicity [R...], Socioeconomic statu..., School type [SCH_T..., High school progra..., Locus of control [L..., Self-concept [SEL_..., Motivation [MOTIV], Career choice [C_C..., Reading score [REA..., Writing score [WRITI..., Science score [SCI..., and Civics score [CIVICS]. The 'Variable' box on the right contains 'Math score [MATH]'. There is a 'Titles...' button. Below the variable box is a checkbox for 'Display normal curve'. The 'Panel by' section has 'Rows' and 'Columns' boxes, both empty, with 'Nest variables (no empty rows)' and 'Nest variables (no empty columns)' checkboxes. At the bottom is a 'Template' section with a checkbox for 'Use chart specifications from:' and an 'File...' button. At the very bottom are buttons for 'OK', 'Paste', 'Reset', 'Cancel', and 'Help'.

You should get the following histogram:

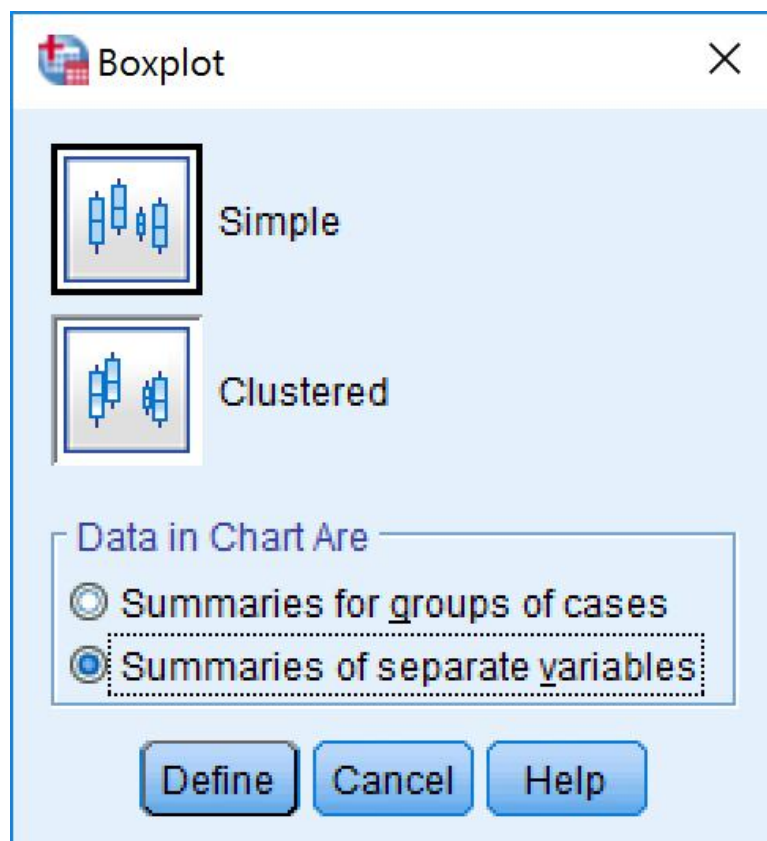


5. Boxplot: Math Score

Graphs → Legacy Dialogs → Boxplot...



Choose **Simple** (default) under **Boxplot** and **Summaries for separate variables** (NOT default) for **Data in Chart Are**, then click **Define**.



Click **Math score** into the **Boxes Represent** box, then hit **OK**.

Define Simple Boxplot: Summaries of Separate Variables

Boxes Represent:

Math score [MATH]

Label Cases by:

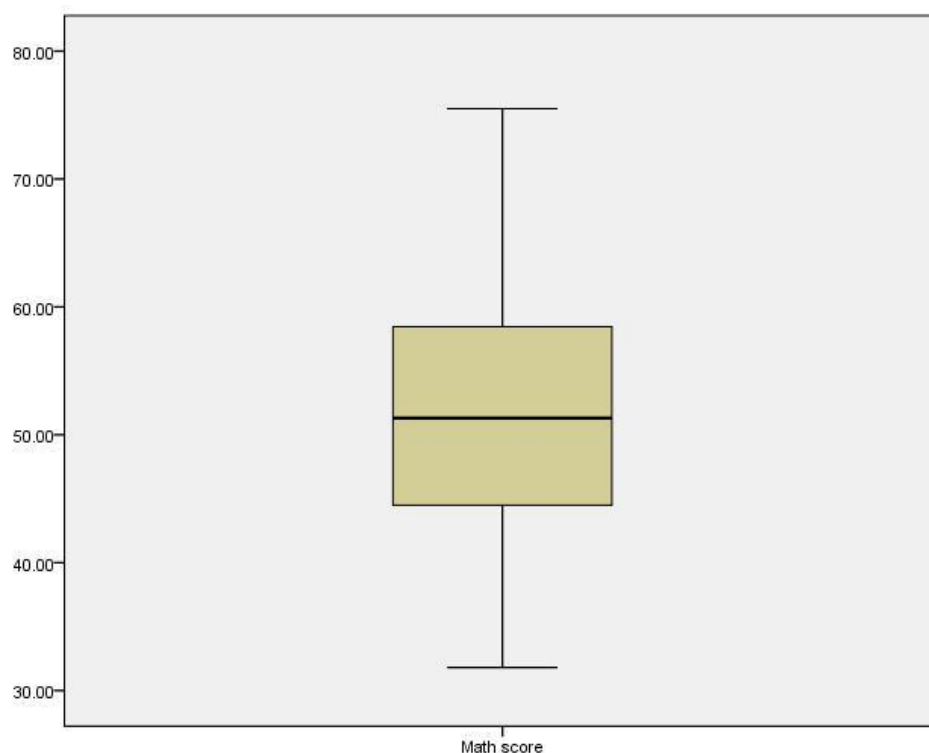
Panel by:

Rows:

Columns:

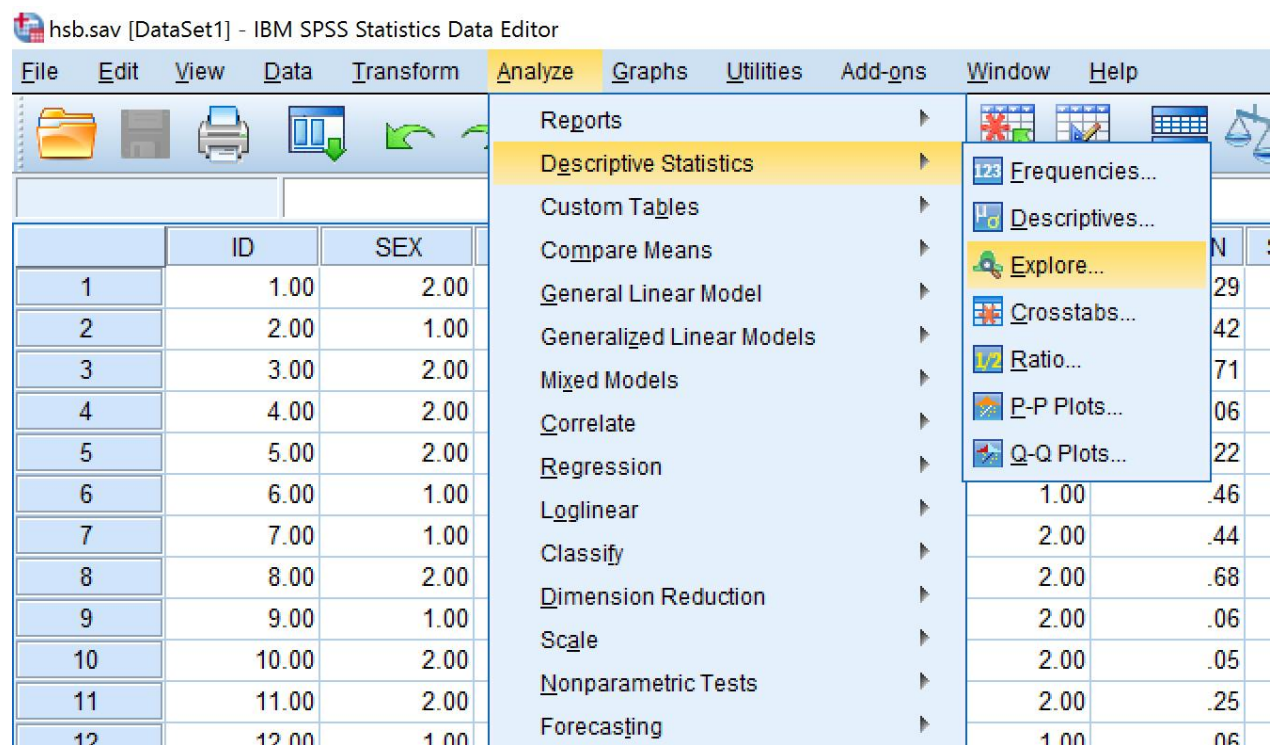
OK Paste Reset Cancel Help

You should get the following boxplot:

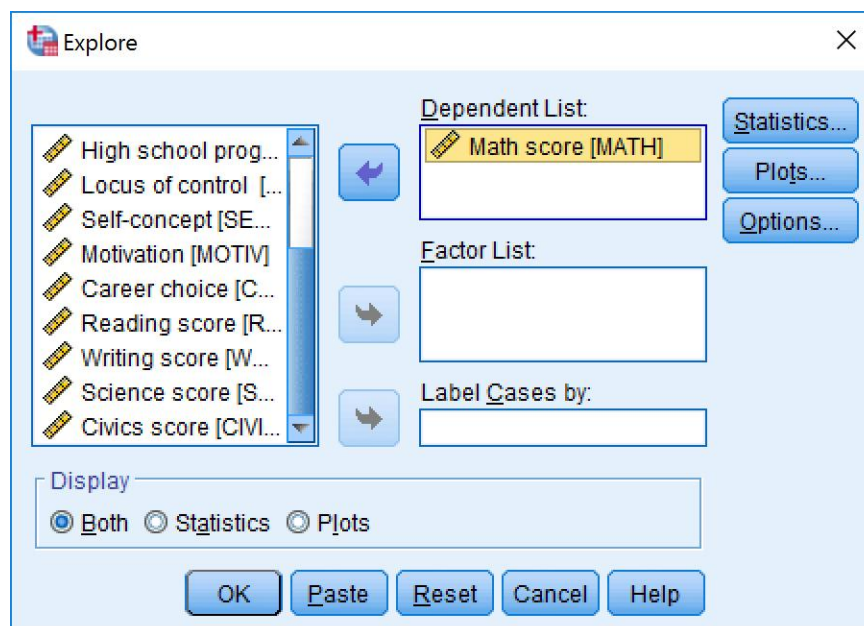


6. Descriptive Statistics: Math Score

Analyze → Descriptive Statistics → Explore...



Click **Math score** into the **Dependent List** box and choose **Both** (default) under **Display**, then hit **OK**.



The first output contains the number of cases that were processed.

Case Processing Summary						
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Math score	600	100.0%	0	0.0%	600	100.0%

The second output contains the descriptive statistics.

Descriptives			
		Statistic	Std. Error
Math score	Mean	51.8490	.38436
	95% Confidence Interval for Mean		
	Lower Bound	51.0942	
	Upper Bound	52.6038	
	5% Trimmed Mean	51.6537	
	Median	51.3000	
	Variance	88.637	
	Std. Deviation	9.41474	
	Minimum	31.80	
	Maximum	75.50	
	Range	43.70	
	Interquartile Range	14.02	
	Skewness	.264	.100
	Kurtosis	-.641	.199

The third output is a stemplot.

Math score Stem-and-Leaf Plot

Frequency	Stem &	Leaf
7.00	3 .	33&
47.00	3 .	5556666677778888888899
105.00	4 .	0000000001111111111111122222222222333333444444444
107.00	4 .	5555555555556666666666777777888888888888999999999
111.00	5 .	00000000011111111111111222222222233333333333444444444
97.00	5 .	555555555555666666666677777788888888889999999
69.00	6 .	000011111111222222233333344444444
35.00	6 .	55566667777888899
20.00	7 .	000112224&
2.00	7 .	5

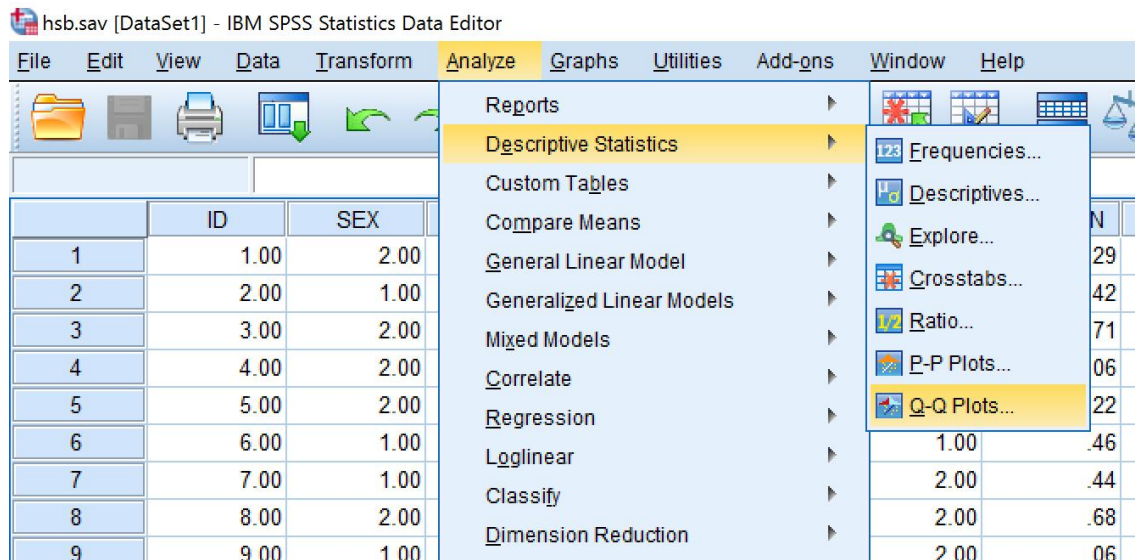
Stem width: 10.00
Each leaf: 2 case(s)

& denotes fractional leaves.

What is the fourth output?

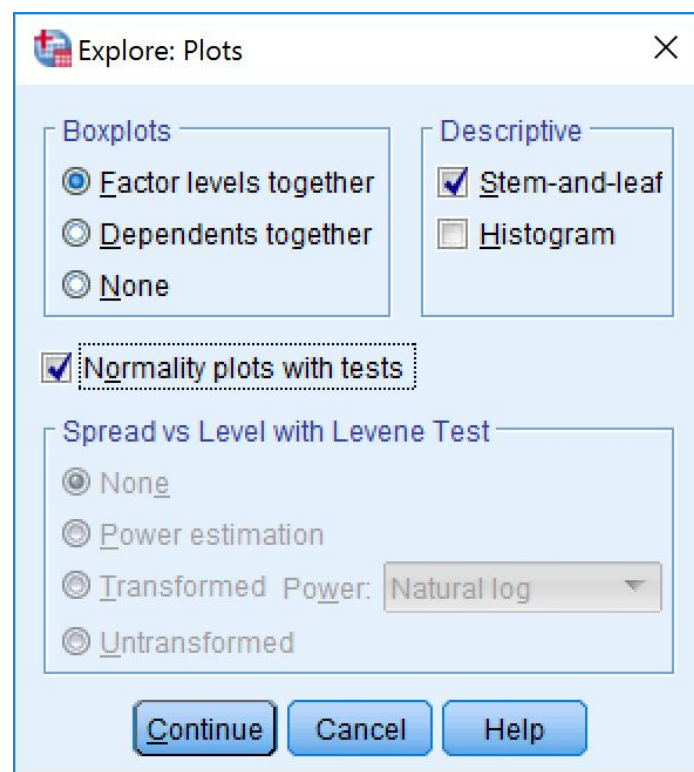
7. Normal Quantile Plot: Math Score

Option 1 (Plot only): **Analyze → Descriptive Statistics → Q-Q Plots...**



Option 2 (Plot and Test): **Analyze → Descriptive Statistics → Explore...**

Click the **Plots** button and then check the **Normality plots with tests** option.

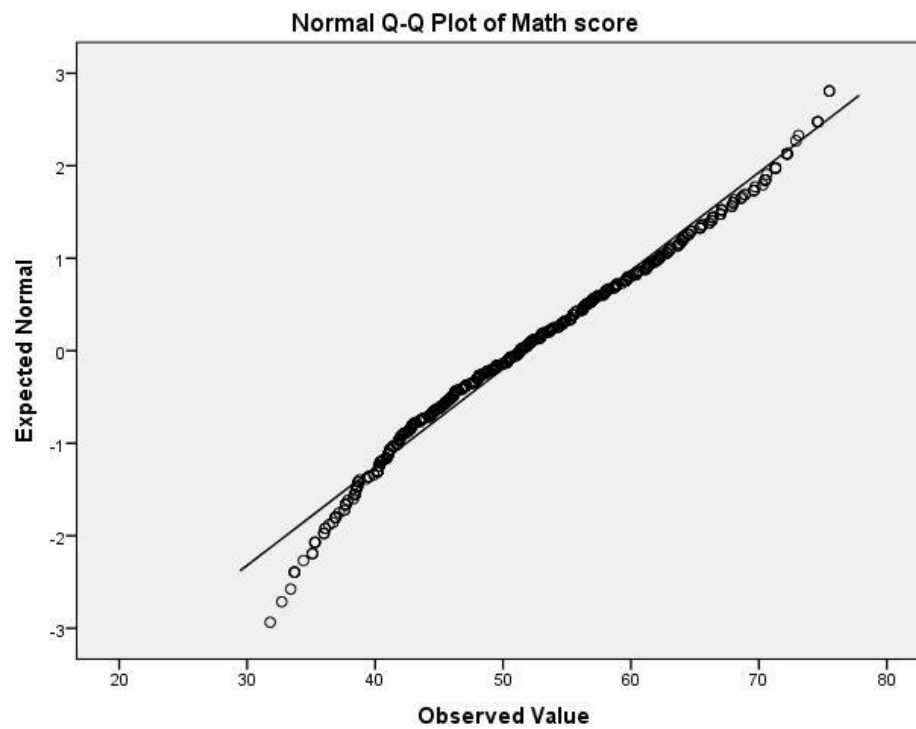


Hit **Continue** then **OK**.

In addition to the regular **Explore** output, you should also get the following:

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Math score	.056	600	.000	.984	600	.000

a. Lilliefors Significance Correction



A Basic Way to Create an SPSS Data File

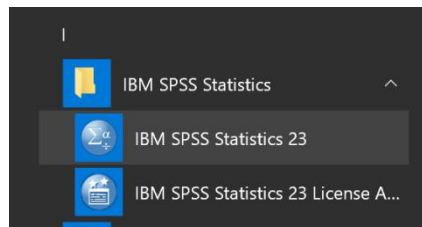
Suppose we want to create a file the following ten cases and four variables (**Data View** tab):

	ID	LOC_CON	SCIENCE	MATH	SEX
1	1	.07	49.50	42.90	2
2	2	-.51	38.90	32.70	1
3	3	1.21	73.30	68.00	2
4	4	.56	44.20	59.80	2
5	5	.03	46.90	42.80	2
6	6	.63	57.40	58.90	1
7	7	.04	49.50	46.20	1
8	8	.51	62.70	48.00	2
9	9	.03	442.00	40.50	1
10	10	.05	38.90	41.80	2

We also have the following information about the variables (**Variable View** tab):

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	ID	Numeric	8	0	ID number	None	None	8	Right	Scale	Input
2	LOC_CON	Numeric	8	2	Locus of control	None	None	8	Right	Scale	Input
3	SCIENCE	Numeric	8	2	Science	None	None	8	Right	Scale	Input
4	MATH	Numeric	8	2	Math score	None	None	8	Right	Scale	Input
5	SEX	Numeric	8	0	Sex	{1, Male}...	None	8	Right	Scale	Input

1. Click the **Windows icon**, and open SPSS. In this example, it is **IBM SPSS Statistics 23**.



2. This will open a blank SPSS file. Go to **Variable View** tab and type **ONLY** the names of the variables (i.e., column 1).

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

SPSS will fill in the default entries for the remaining columns.

3. To change a cell entry, **double click** the cell and **type or choose** the appropriate entry.

For example, we want the variable **SEX** to have **no Decimals**, the Label **Sex**, Values of **1 for Male** and **2 for Female**, and to note that this is a **Nominal Measure**.

Double click Decimals and choose the down button to set the number of decimal points to 0.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	ID	Numeric	8	0	ID number	None	None	8	Right	Nominal	Input
2	LOC_CON	Numeric	8	2	Locus of control	None	None	8	Right	Scale	Input
3	SCIENCE	Numeric	8	2	Science	None	None	8	Right	Scale	Input
4	MATH	Numeric	8	2	Math score	None	None	8	Right	Scale	Input
5	SEX	Numeric	8	0		None	None	8	Right	Scale	Input

Type Sex as the Label.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	ID	Numeric	8	0	ID number	None	None	8	Right	Nominal	Input
2	LOC_CON	Numeric	8	2	Locus of control	None	None	8	Right	Scale	Input
3	SCIENCE	Numeric	8	2	Science	None	None	8	Right	Scale	Input
4	MATH	Numeric	8	2	Math score	None	None	8	Right	Scale	Input
5	SEX	Numeric	8	0	Sex	None	None	8	Right	Scale	Input

After double clicking the Values cell, click the **Option** button to input values and labels.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	ID	Numeric	8	0	ID number	None	None	8	Right	Nominal	Input
2	LOC_CON	Numeric	8	2	Locus of control	None	None	8	Right	Scale	Input
3	SCIENCE	Numeric	8	2	Science	None	None	8	Right	Scale	Input
4	MATH	Numeric	8	2	Math score	None	None	8	Right	Scale	Input
5	SEX	Numeric	8	0	Sex	None	None	8	Right	Scale	Input

To note that 1 corresponds to Male, type **1** in the Value box and **Male** in the Label box, then click **Add**.

Value Labels

Value: 2

Label: Female

Add

Change

Remove

1 = "Male"

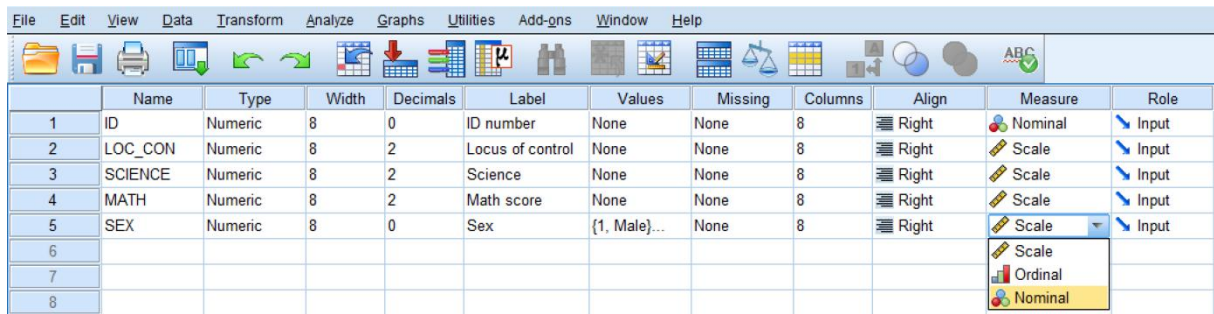
OK

Cancel

Help

Do the same for Female.

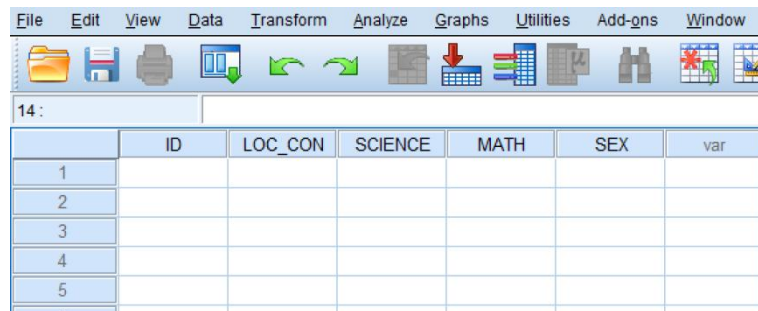
Choose **Nominal** as the Measure for **SEX**.



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	ID	Numeric	8	0	ID number	None	None	8	Right	Nominal	Input
2	LOC_CON	Numeric	8	2	Locus of control	None	None	8	Right	Scale	Input
3	SCIENCE	Numeric	8	2	Science	None	None	8	Right	Scale	Input
4	MATH	Numeric	8	2	Math score	None	None	8	Right	Scale	Input
5	SEX	Numeric	8	0	Sex	{1, Male}...	None	8	Right	Nominal	Input
6											
7											
8											

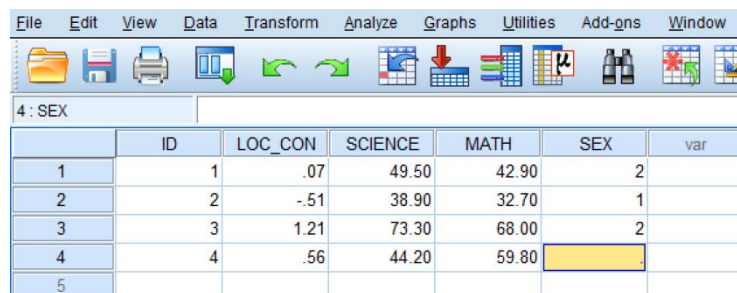
Make sure that the rest of the variables have the correct information.

When you switch to the **Data View** tab, you should get the following:



	ID	LOC_CON	SCIENCE	MATH	SEX	var
1						
2						
3						
4						
5						

We are now ready to enter the data points.



	ID	LOC_CON	SCIENCE	MATH	SEX	var
1	1	.07	49.50	42.90	2	
2	2	-.51	38.90	32.70	1	
3	3	1.21	73.30	68.00	2	
4	4	.56	44.20	59.80		
5						

To check data are correctly entered, do **Analyze → Descriptive Statistics → Explore...**

Use **MATH** in the Dependent List and **SEX** in the Factor List to get the following results:

Descriptives

Sex				Statistic	Std. Error
Math score	Male	Mean		44.5750	5.51866
		95% Confidence Interval for Mean	Lower Bound	27.0122	
			Upper Bound	62.1378	
		5% Trimmed Mean		44.4389	
		Median		43.3500	
		Variance		121.822	
		Std. Deviation		11.03732	
		Minimum		32.70	
		Maximum		58.90	
		Range		26.20	
		Interquartile Range		21.07	
		Skewness		.596	1.014
		Kurtosis		.488	2.619
	Female	Mean		50.5500	4.44115
		95% Confidence Interval for Mean	Lower Bound	39.1337	
			Upper Bound	61.9663	
		5% Trimmed Mean		50.0667	
		Median		45.4500	
		Variance		118.343	
		Std. Deviation		10.87856	
		Minimum		41.80	
		Maximum		68.00	
		Range		26.20	
		Interquartile Range		19.30	
		Skewness		1.050	.845
		Kurtosis		-.603	1.741

