جامعة قطر
QATAR UNIVERSITY

Introduction to Datasets & SPSS

March 2013

Ashley Jardina
University of Michigan

SESRI
معهد البحوث الاجتماعية والاقتصادية المسحية
Social & Economic Survey Research Institute

We will be working with the 2010 & 2011 SESRI Omnibus Surveys.

There are three datasets on your computers with a subset of variables from these surveys:

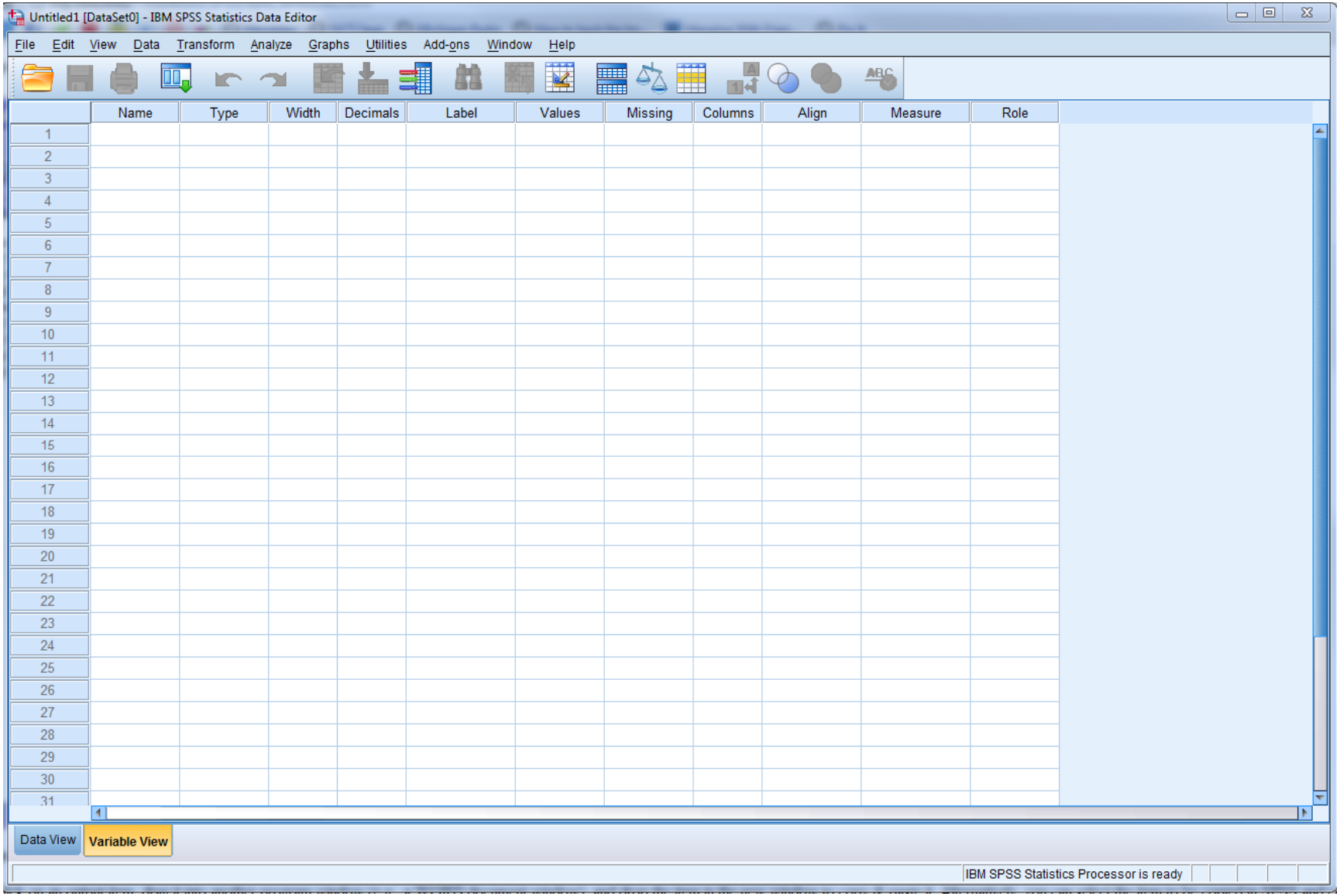
Dataset 1: 2010 Omnibus
Dataset 2: 2010 Omnibus
Dataset 3: 2011 Omnibus

You also have copies of the survey questionnaires as PDF files.

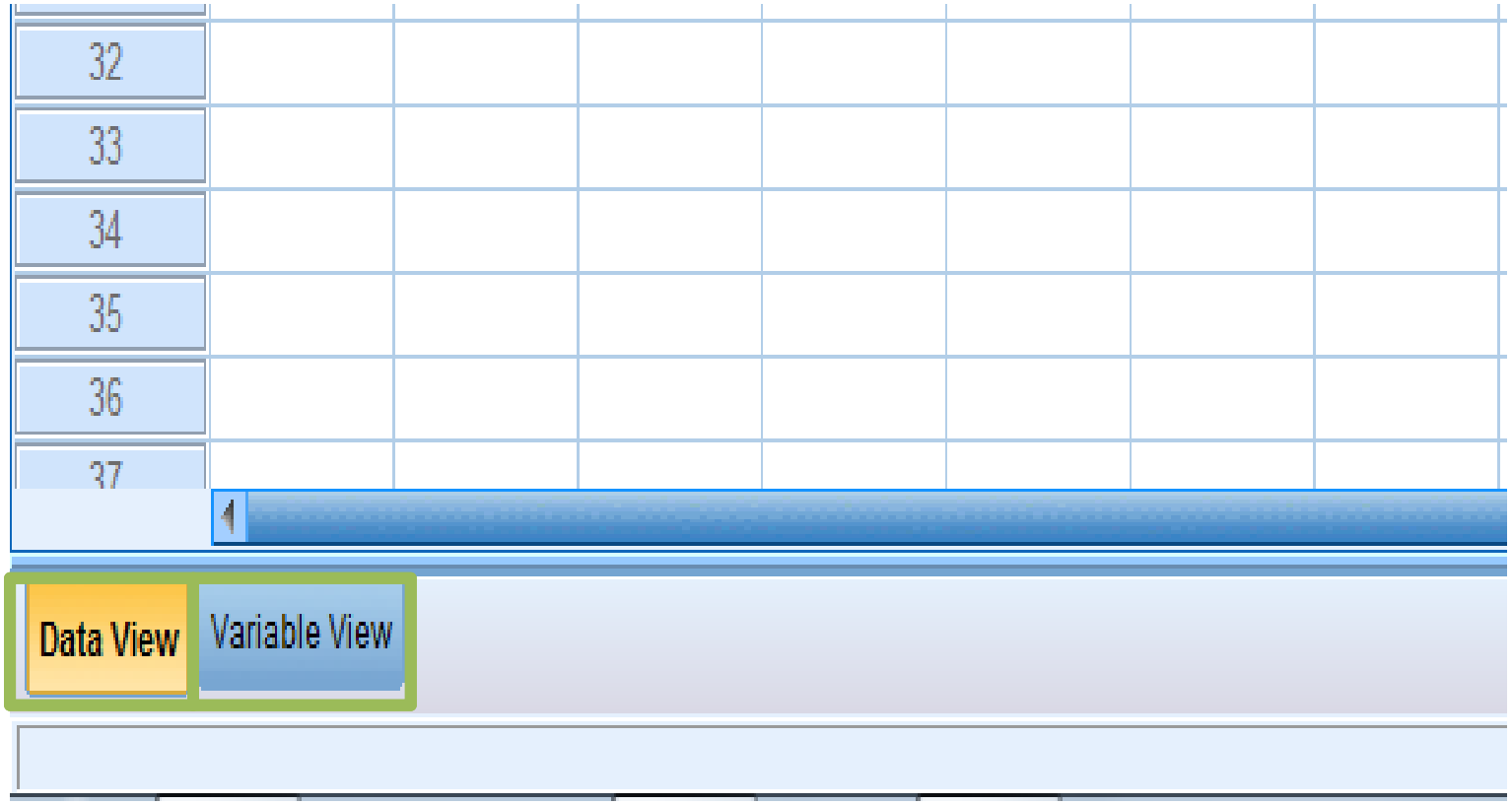
What is SPSS?

- SPSS: Statistical Product and Service Solutions
- SPSS is a software package used for statistical analysis and data management.
- It can be used to analyze data and produce reports, charts, plots, and descriptive statistics for a wide range of data sources.
- The version of SPSS used in this presentation is SPSS 20.

The SPSS Statistics Main Window



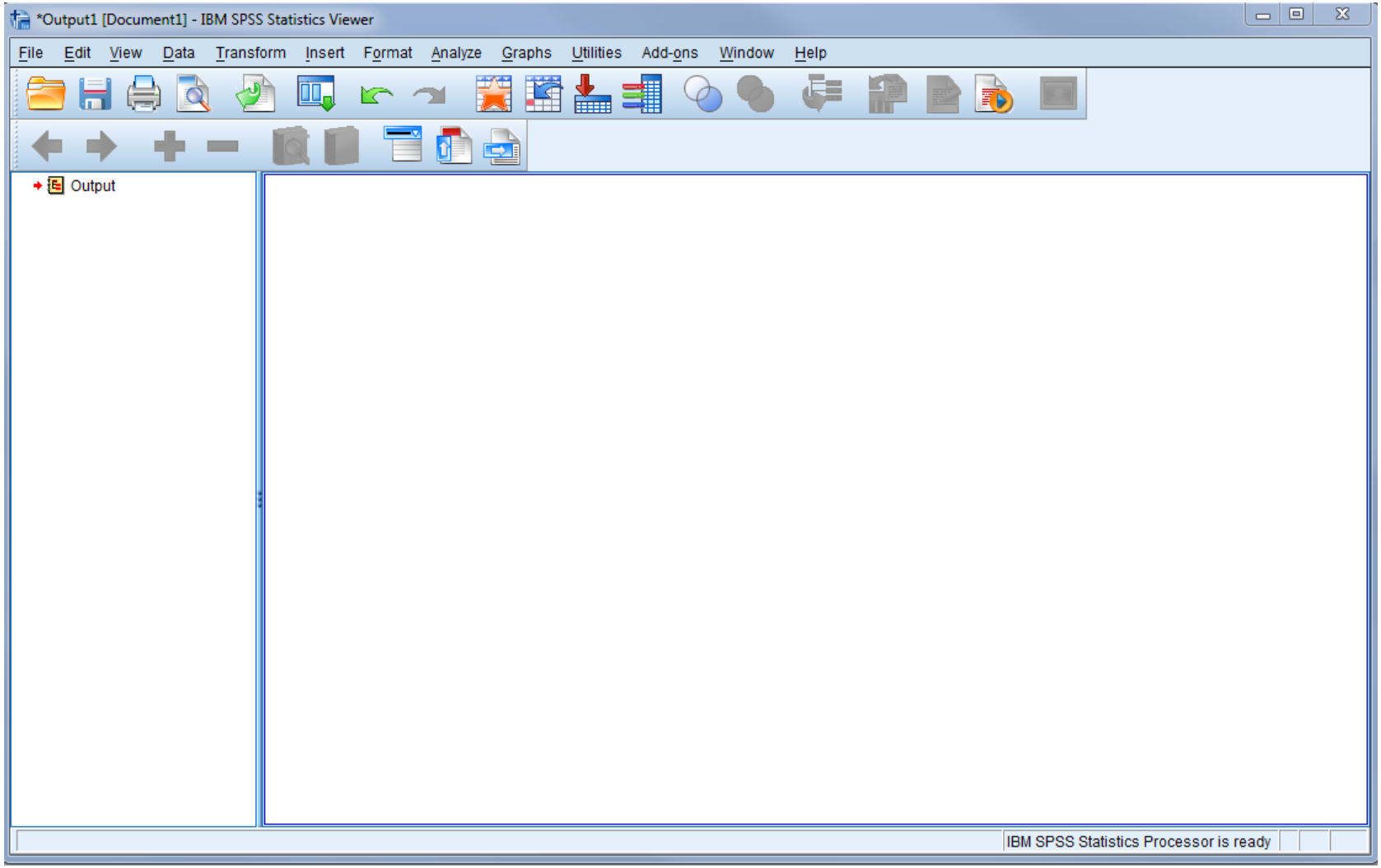
SPSS – Data View and Variable View



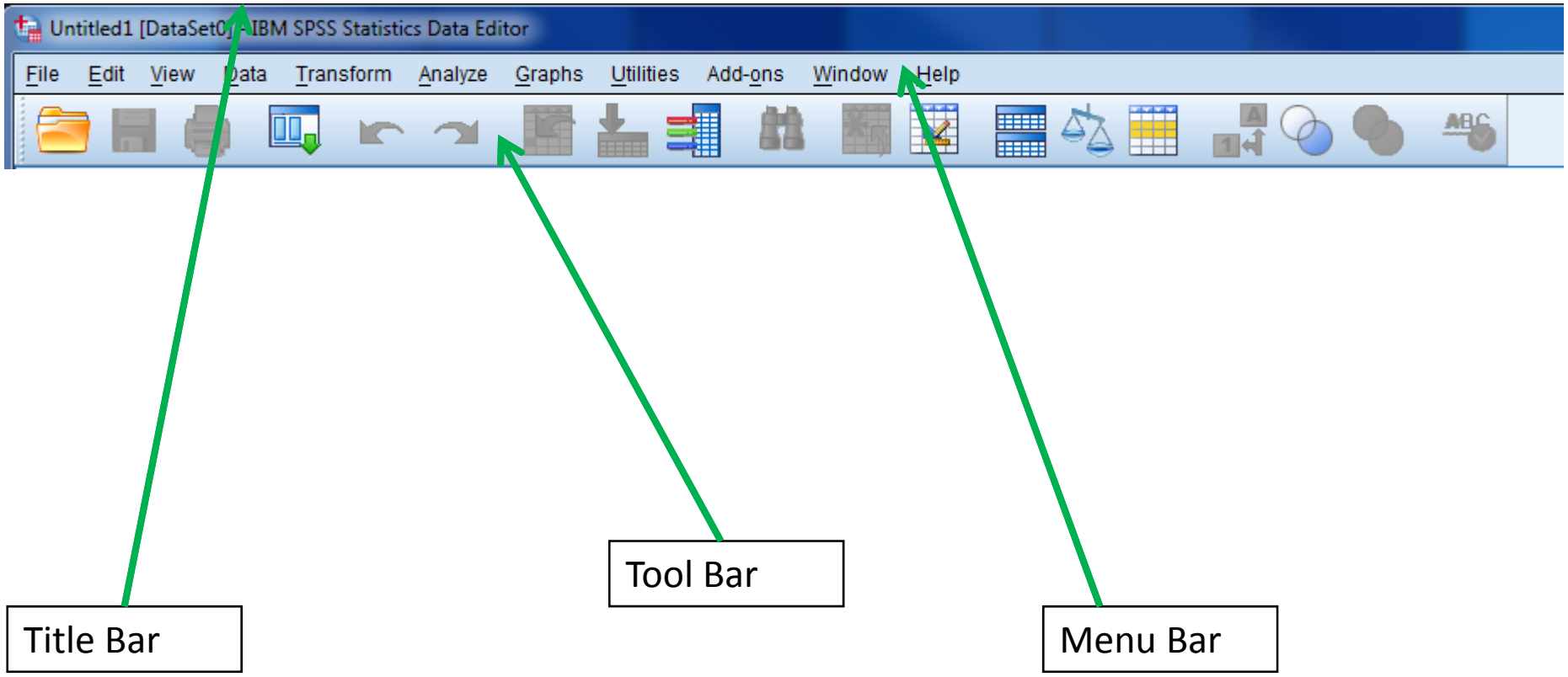
“Data View” displays a spreadsheet of the data, with each row as a case and each column as a variable.

“Variable View” displays information about the variables in the dataset, including variable names, types, and any labels associated with the variables.

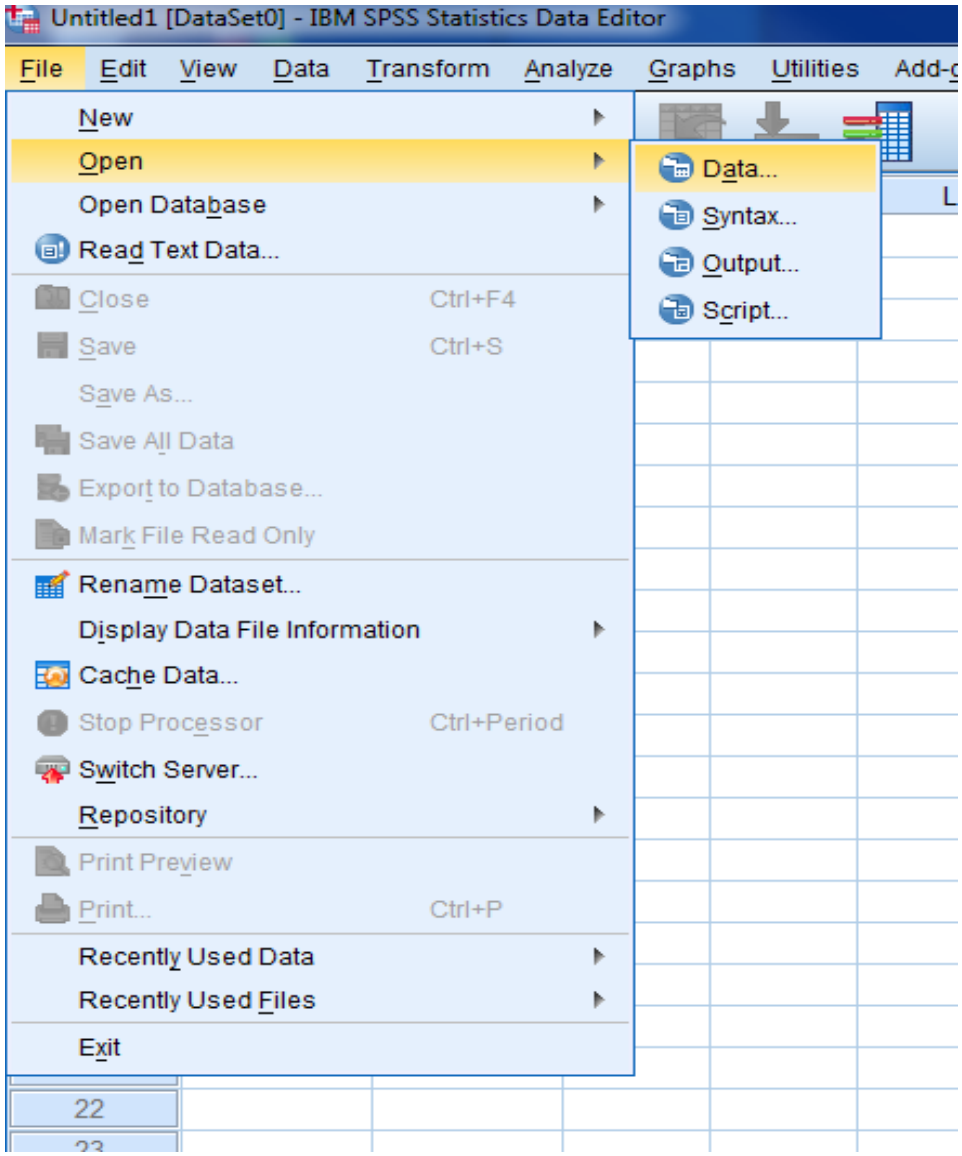
SPSS – Output View



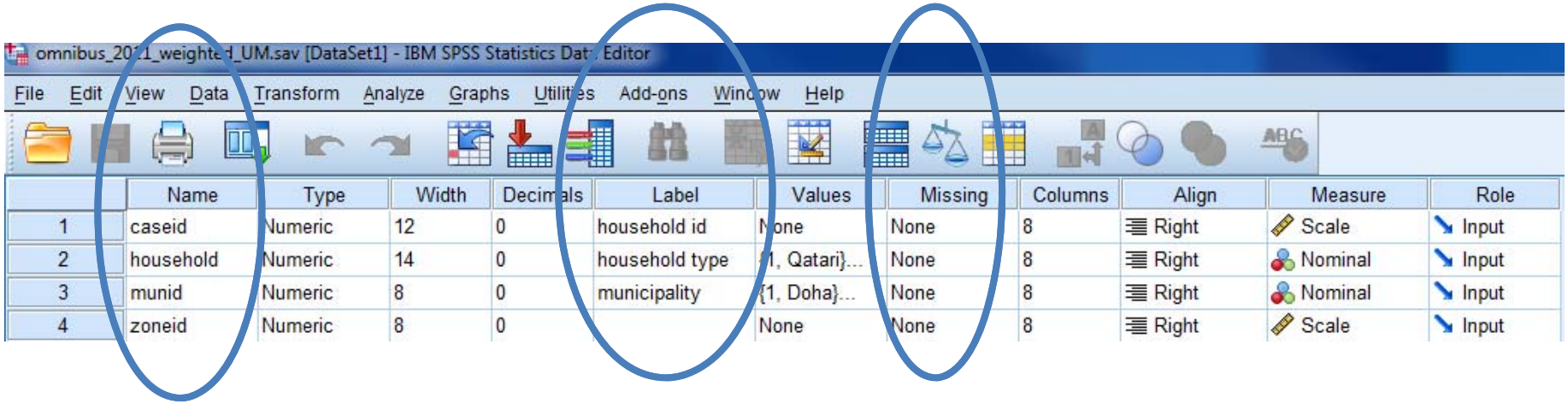
SPSS Menus and Toolbar



Opening a Dataset



Open Dataset in Variable View



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	caseid	Numeric	12	0	household id	None	None	8	Right	Scale	Input
2	household	Numeric	14	0	household type	{1, Qatari}...	None	8	Right	Nominal	Input
3	munid	Numeric	8	0	municipality	{1, Doha}...	None	8	Right	Nominal	Input
4	zoneid	Numeric	8	0		None	None	8	Right	Scale	Input

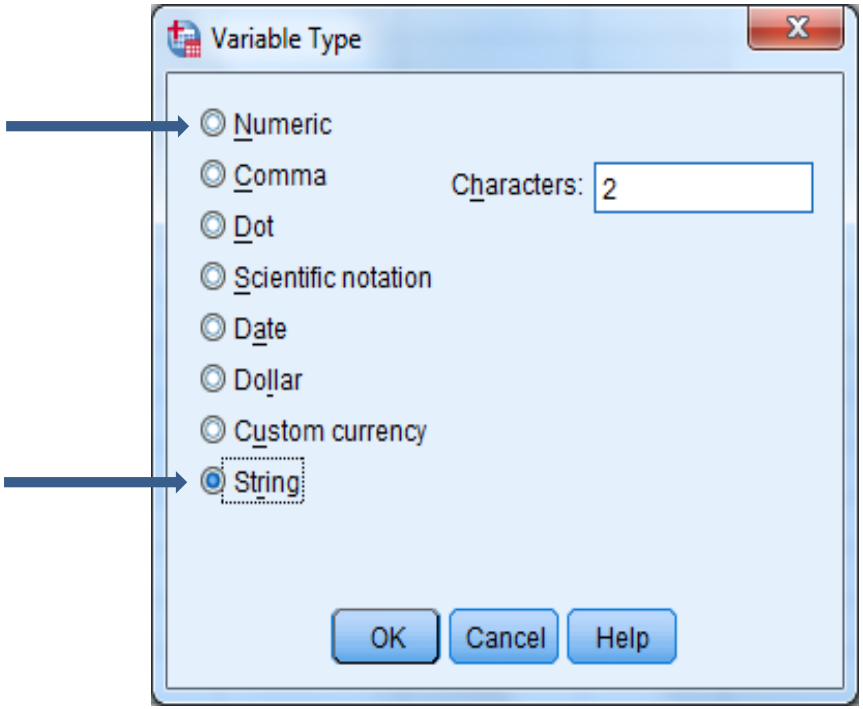
There are several key pieces of information to note about your data under “variable view” in SPSS.

The “name” column specifies the name of the variable.

The “label” column provides a more detailed description of the variable.

The “missing” column indicates which values for each variable SPSS will ignore when it runs statistics or generates tables.

SPSS - types of variables



"Variable View" also indicates the *type* of each variable. If you click on the "type" box in variable view, you will see this window, which allows you to alter a variable's type. The two basic types of variables are **numeric** and **string**. Numeric variables contain only numbers. String variables may contain letters or numbers.

Levels of Measurement

- Nominal: Values represent categories with no intrinsic ranking. E.g., gender, household type

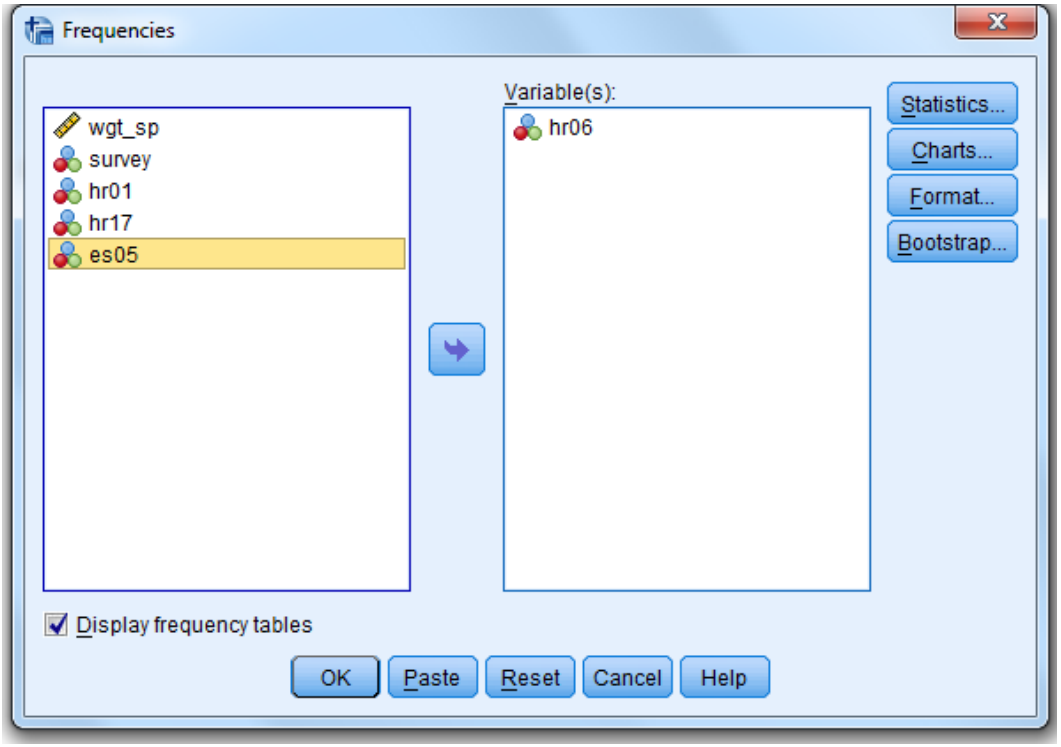
- Ordinal: Values represent categories with some intrinsic ranking. E.g., level of satisfaction

- Scale / Interval: Values represent categories with meaningful metric and ranking. E.g., age in years, income in dollars

Frequencies

- The frequencies command provides a useful summary of a given variable. It indicates the number of responses that correspond to each value of a variable.

- Use the *Analyze / Descriptive Statistics / Frequencies* menu to find the frequencies of a variable.



Using the frequencies command, we can see that there are 1350 male and 789 female survey respondents in our dataset.

gender of hh member

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. male	1350	63.1	63.1	63.1
	2. female	789	36.9	36.9	100.0
	Total	2139	100.0	100.0	

Renaming and labeling a new variable

- Variable es05 in your dataset corresponds to the level of education for each respondent. We will use the "Compute" command to create a new variable equal to es05 with a more intuitive name and a clear label.

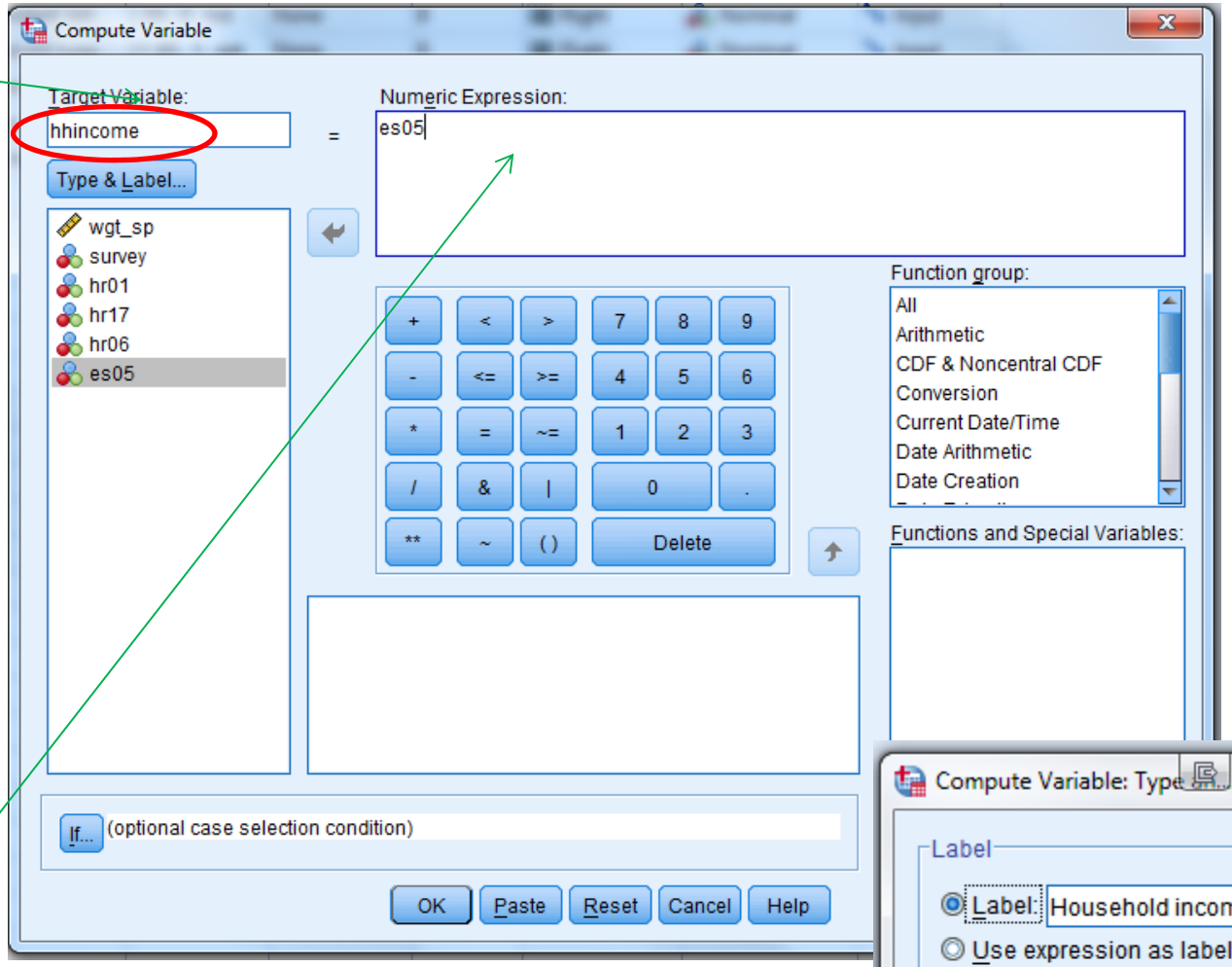
- Use the *Transform / Compute Variable* menu option.

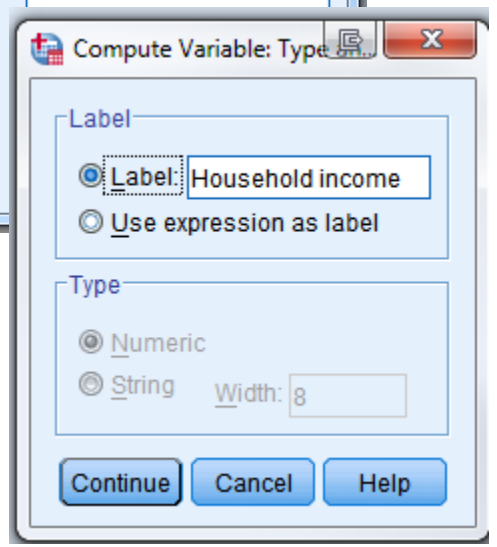
- Note that we could simply rename es05 in the "Variable View" window, but it is good practice to create new variables rather than manipulating the original variables.

The “Target Variable:” box is where you want to input the name of your new variable.

In the example, we used “hhincome” to stand for household income.

Set your new variable equal to the original variable, which in this case is es05.





Select “Type & Label” to give your variable a descriptive label, like “Household income”. This label will appear in “Variable View” as well as in output produced by SPSS when analyzing your data.

Exercise

- Create a new variable that is a copy of variable **hr17**, which corresponds to the survey respondent's level of education. Give your new variable a name that makes it easy for you to identify.
- Run a frequency on your new education variable. Run a frequency on **hr17**. Compare the two variables to make sure they are the same.
- Create a copy of **f1**, which indicates the gender of the survey respondent. Give your new variable a recognizable name.
- Run a frequency on your new gender variable. Run a frequency on **f1**. Compare the two variables to make sure they are the same.

hh member highest level of education

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. primary (1-6)	84	3.9	6.9	6.9
	2. preparatory (7-9)	146	6.8	12.0	18.9
	3. vocational	12	.6	1.0	19.9
	4. secondary (10-12)	409	19.1	33.6	53.5
	5. post secondary	94	4.4	7.7	61.3
	6. university graduate/ba/bcom/bsc	403	18.8	33.1	94.4
	7. masters	49	2.3	4.0	98.4
	8. ph.d.	19	.9	1.6	100.0
	Total	1218	56.9	100.0	
Missing	9. other (specify)	7	.3		
	99. refused	2	.1		
	System	912	42.6		
	Total	921	43.1		
Total		2139	100.0		

gender of hh member

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. male	1350	63.1	63.1	63.1
	2. female	789	36.9	36.9	100.0
	Total	2139	100.0	100.0	