Introduction to Data Analysis with SPSS

Oct. 1, 2026

1:30 – 3:30 p.m.

Online

Sherman Centre for Digital Scholarship

Stream

Data Analysis and Visualization

Statistics

			Statistic	Bootstrap[a] Bias	Std. Error	95% Confidence Interval Lower	Upper
Churn within last month	N	Valid	266	0	0	266	266
		Missing	0	0	0	0	0
	Mean		.31	.00	.03	.26	.37
Customer category: Basic service	N	Valid	217	0	0	217	217
		Missing	0	0	0	0	0
	Mean		.27	.00	.03	.21	.34
Customer category: E-service	N	Valid	281	0	0	281	281
		Missing	0	0	0	0	0
	Mean		.16	.00	.02	.11	.20
Customer category: Plus service	N	Valid	236	0	0	236	236
		Missing	0	0	0	0	0
	Mean		.37	.00	.03	.31	.43

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Churn within last month

			Frequency	Percent	Valid Percent	Cumulative Percent	Bootstrap for Percent[a] Bias	Std. Error	95% Confidence Lower
Customer category: Basic service	Valid	No	183	68.8	68.8	68.8	.0	2.8	63.2
		Yes	83	31.2	31.2	100.0	.0	2.8	25.9
		Total	266	100.0	100.0		.0	.0	100.0
Customer category: E-service	Valid	No	158	72.8	72.8				66.4
		Yes	59	27.2	27.2				21.2
		Total	217	100.0	100.0				00.0
Customer category: Plus service	Valid	No	237	84.3	84.3				80.1
		Yes	44	15.7	15.7				11.4
		Total	281	100.0	100.0				00.0
Customer category: Total service	Valid	No	148	62.7	62.7	62.7	-.1	3.1	56.8
		Yes	88	37.3	37.3	100.0	.1	3.1	31.4
		Total	236	100.0	100.0		.0	.0	100.0

Land Acknowledgement

McMaster University is located on the traditional Territories of the Mississauga and Haudenosaunee Nations, and within the lands protected by the “Dish With One Spoon” wampum agreement.

Code of Conduct

The Sherman Centre and the McMaster University Library are committed to fostering a supportive and inclusive environment for its presenters and participants.

As a participant in this session, you agree to support and help cultivate an experience that is collaborative, respectful, and inclusive, as well as free of harassment, discrimination, and oppression. We reserve the right to remove participants who exhibit harassing, malicious, or persistently disruptive behaviour.

Please refer to our code of conduct webpage for more information: **scds.ca/events/code-of-conduct**

Session Recording and Privacy

This session is being recorded with the intention of being shared publicly via the web for future audiences. In respect of your privacy, participant lists will not be shared outside of this session, nor will question or chat transcripts.

Questions asked via the chat box will be read by the facilitator without identifying you. Note that you may be identifiable when asking a question during the session in an audio or visual format.

Book an Appointment with the DASH Team

Receive help from a member of the DASH team! DASH can assist with the following topics:

- Creating data visualizations, including charts, graphs, and scatter plots
- Figuring out which statistical tests to run (e.g., t-test, chi-square, etc.).
- Analyzing data with software including SPSS, Python, R, SAS, ArcGIS, MATLAB, and Excel
- Choosing which software package to use, including free and open-source software
- Troubleshooting problems related to file formats, data retrieval, and download
- Selecting methodology and type of data analysis to use in a thesis project

Book an appointment: **https://library.mcmaster.ca/services/dash**

scds.ca
scds@mcmaster.ca

Certificate Programs

The Sherman Centre for Digital Scholarship Certificate of Attendance

The Sherman Centre's certificate program recognizes attendance at our workshops. It complements degree training, supports the development of critical competencies in data analysis, research data management, and digital scholarship, and formalizes core skills fostered by our workshops.

Participants are invited to attend seven workshops and receive a certificate of attendance. To verify your participation in today's workshop, we will provide a code and additional instructions at the end of the session.

You can learn more about the certificate program at **scds.ca/certificate-program**

The Canadian Certificate for Digital Humanities

This workshop is also eligible for the Canadian Certificate for Digital Humanities. To learn more about the certificate, visit **ccdhhn.ca**. You can also contact local liaison Alexis-Carlota Cochrane at scds@mcmaster.ca

Learning Objectives

- SPSS Background
- SPSS Environment
- Reading and Preparing Data
- Descriptive Statistics

What is SPSS?

- SPSS stands for **Statistical Product and Service Solutions**. It is also known by the name **Predictive Analytics Software (PASW)**
- It is a powerful statistical software with Graphical User Interface (GUI) used for data analysis in business, health care, social science research. SPSS can be used for:
  - Processing Questionnaires
  - Reporting in Tables and Graphs
  - Analyzing Means, Chi-Square, Regression, and much more…

History of SPSS

- Introduced in 1968
- Was originally developed to facilitate statistical analysis in the social sciences.
- Early versions designed to run on mainframe computers.
- The current versions are officially named IBM SPSS Statistics.

SPSS Features

- Import data from many different sources, such as MS Excel and SAS
- Provides analysis tools to generate reports, charts, plots, descriptive statistics, and run advanced statistical analysis.
- Provides a command syntax that can simplify certain things, such as running repetitive tasks.

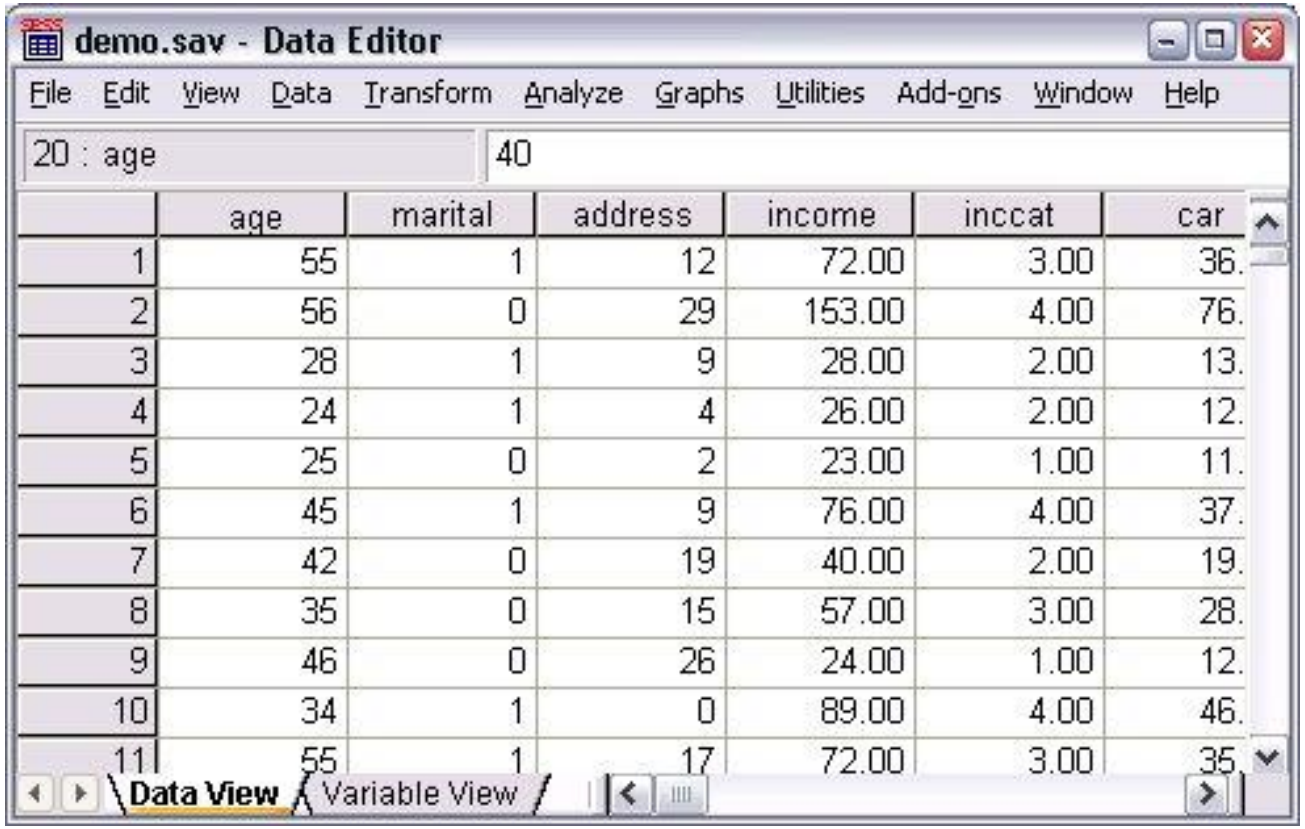
SPSS Environment

- Different type of Windows in SPSS
  - Data View
  - Variable View
  - Output Viewer
  - Pivot Table Editor
  - Chart Editor
  - Text Output Editor
  - Syntax Editor

Reading Data

- Data can be entered directly, or it can be imported from a number of different sources.
  - Spreadsheet applications, such as Microsoft Excel;
  - Database applications, such as Microsoft Access;
  - CSV or Text files

Basic Structure of IBM SPSS Statistics Data Files



	age	marital	address	income	inccat	car
1	55	1	12	72.00	3.00	36.
2	56	0	29	153.00	4.00	76.
3	28	1	9	28.00	2.00	13.
4	24	1	4	26.00	2.00	12.
5	25	0	2	23.00	1.00	11.
6	45	1	9	76.00	4.00	37.
7	42	0	19	40.00	2.00	19.
8	35	0	15	57.00	3.00	28.
9	46	0	26	24.00	1.00	12.
10	34	1	0	89.00	4.00	46.
11	55	1	17	72.00	3.00	35

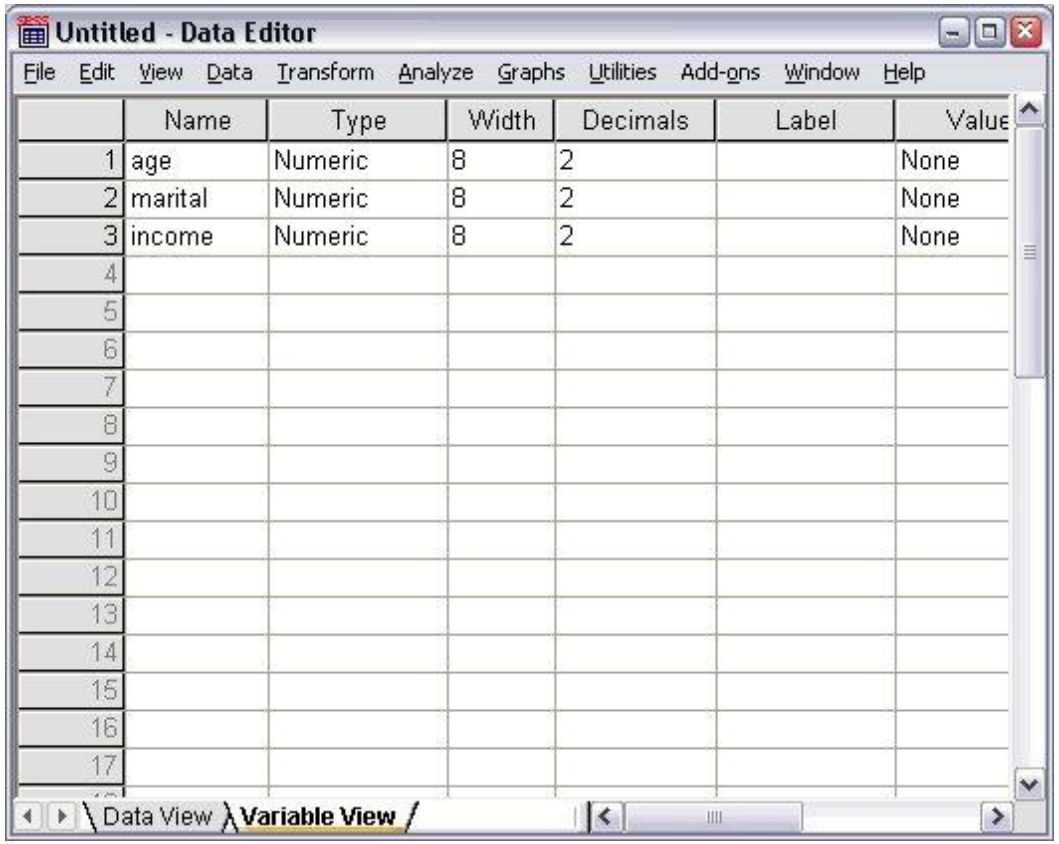
- IBM® SPSS® Statistics data files are organized by **cases (rows)** and **variables (columns)**. In this data file, cases represent individual respondents to a survey. Variables represent responses to each question asked in the survey.

Using the Data Editor

- The Data Editor displays the contents of the active data file. The information in the Data Editor consists of variables and cases.
  - In Data View, columns represent variables, and rows represent cases (observations).
  - In Variable View, each row is a variable, and each column is an attribute that is associated with that variable.

- Variables are used to represent the different types of data that you have compiled. A common analogy is that of a survey. The response to each question on a survey is equivalent to a variable. Variables come in many different types, including numbers, strings, currency, and dates.

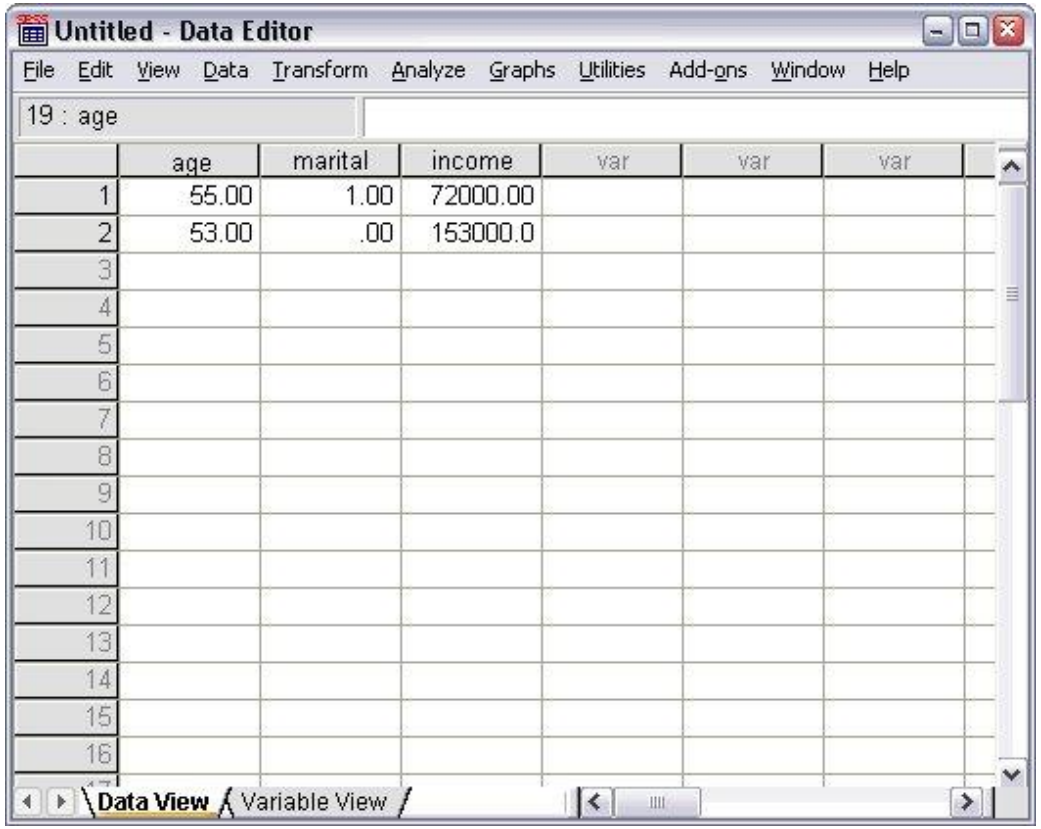
Entering Numeric Data



	Name	Type	Width	Decimals	Label	Value
1	age	Numeric	8	2		None
2	marital	Numeric	8	2		None
3	income	Numeric	8	2		None
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						

- Data can be entered into the Data Editor, which may be useful for small data files or for making minor edits to larger data files.
  - Click the Variable View tab at the bottom of the Data Editor window. You need to define the variables that will be used. In this case, only three variables are needed: *age*, *marital status*, and *income*.
  - New variables are automatically given a Numeric data type.
- If you don't enter variable names, unique names are automatically created. However, these names are not descriptive and are not recommended for large data files.

Entering Numeric Data



	age	marital	income	var	var	var
1	55.00	1.00	72000.00			
2	53.00	.00	153000.0			
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						

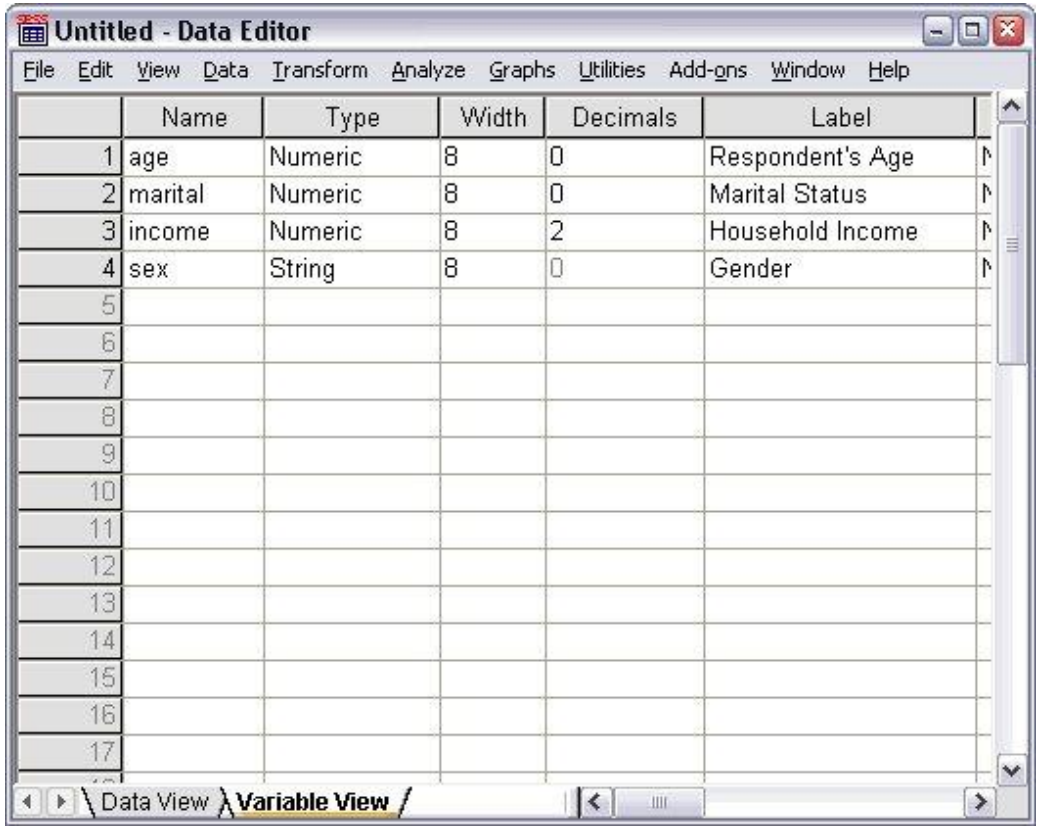
- Click the **Data View** tab to continue entering the data. The names that are entered in Variable View are now the headings for the first three columns in Data View. Begin entering data in the first row, starting at the first column.

- Currently, the *age* and *marital* columns display decimal points, even though their values are intended to be integers. To hide the decimal points in these variables. Click the **Variable View** tab at the bottom of the Data Editor window. In the *Decimals* column of the *age* and *marital* rows, type **0** to hide the decimal

Defining Data

- In addition to defining data types, you can also define descriptive variable labels and value labels for variable names and data values. These descriptive labels are used in statistical reports and charts.

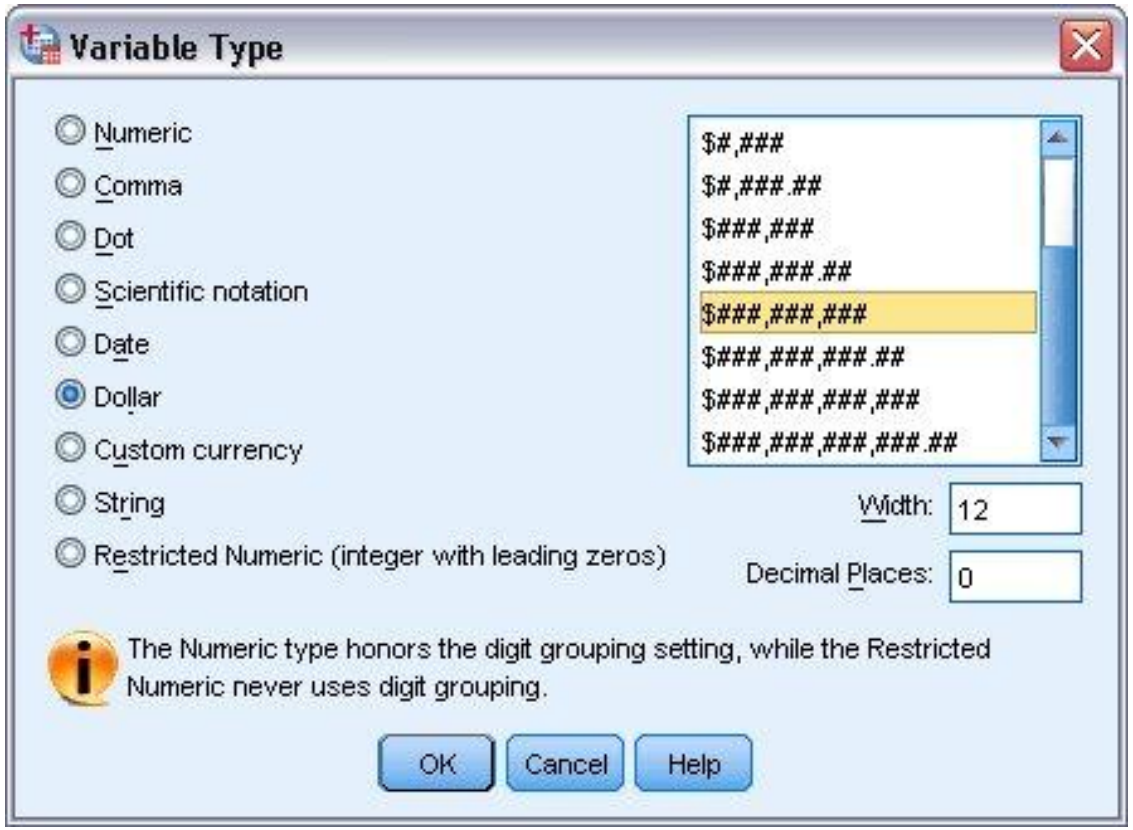
Adding Variable Labels



	Name	Type	Width	Decimals	Label
1	age	Numeric	8	0	Respondent's Age
2	marital	Numeric	8	0	Marital Status
3	income	Numeric	8	2	Household Income
4	sex	String	8	0	Gender

- Labels are meant to provide descriptions of variables. These descriptions are often longer versions of variable names. Labels can be up to 255 bytes. These labels are used in your output to identify the different variables.
  - Click the **Variable View** tab at the bottom of the Data Editor window.

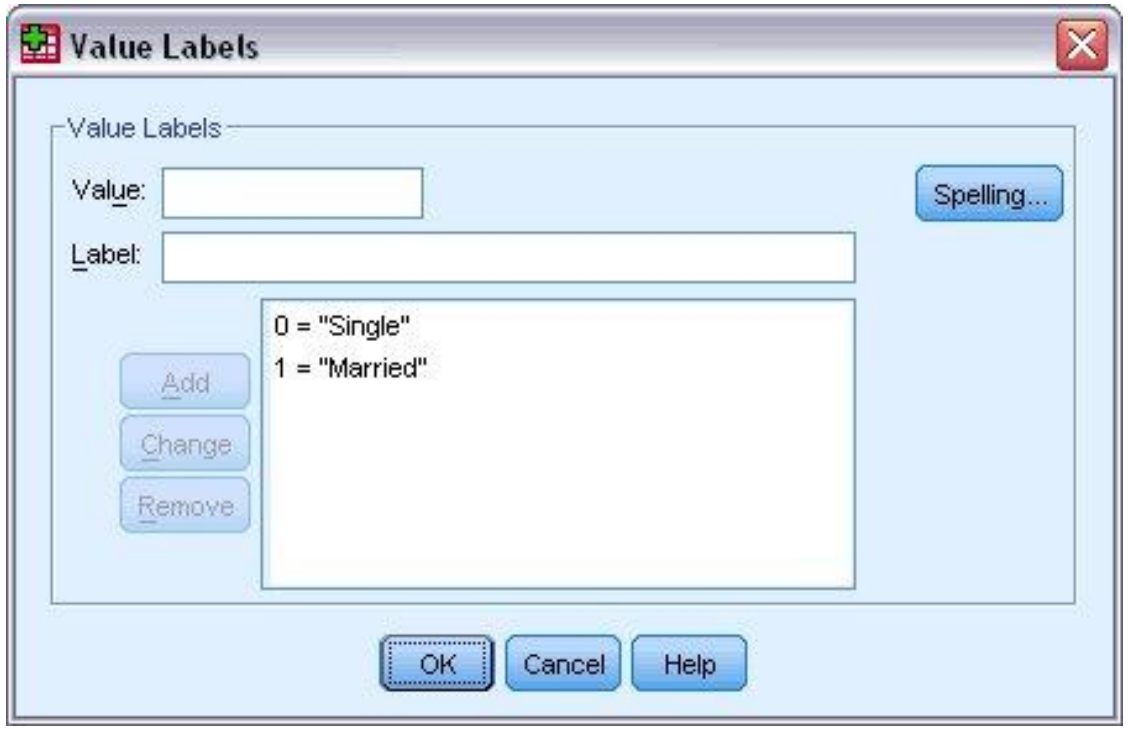
Changing Variable Type and Format



- The *Type* column displays the current data type for each variable. The most common data types are numeric and string, but many other formats are supported. In the current data file, the *income* variable is defined as a numeric type.
  - Click the *Type* cell for the *income* row, and then click the button on the right side of the cell to open the Variable Type dialog box. Select **Dollar**

Adding Value Labels

- Value labels provide a method for mapping your variable values to a string label. In this example, there are two acceptable values for the *marital* variable. A value of 0 means that the subject is single, and a value of 1 means that he or she is married.
  1. Click the *Values* cell for the *marital* row, and then click the button on the right side of the cell to open the Value Labels dialog box. The **value** is the actual numeric value.
  2. The **value label** is the string label that is applied to the specified numeric value.



View > **Value Labels**

- The labels are now displayed in a list when you enter values in the Data Editor.
- This setup has the benefit of suggesting a valid response and providing a more descriptive answer.

If the Value Labels menu item is already active (with a check mark next to it), choosing **Value Labels** again will turn *off* the display of value labels.

Handling Missing Data

- Missing or invalid data are generally too common to ignore. Survey respondents may refuse to answer certain questions, may not know the answer, or may answer in an unexpected format. If you don't filter or identify these data, your analysis may not provide accurate results.
- For numeric data, empty data fields or fields containing invalid entries are converted to system-missing, which is identifiable by a single period.
- The reason a value is missing may be important to your analysis. For example, you may find it useful to distinguish between those respondents who refused to answer a question and those respondents who didn't answer a question because it was not applicable.

Missing Values for a Numeric Variable

- Click the **Variable View** tab at the bottom of the Data Editor window.
- Click the *Missing* cell in the *age* row, and then click the button on the right side of the cell to open the Missing Values dialog box. In this dialog box, you can specify up to three distinct missing values, or you can specify a range of values plus one additional discrete value.

- Select **Discrete missing values**.
- Click **OK** to save your changes and return to the Data Editor.

Now that the missing data value has been added, a label can be applied to that value.

- Type 999 in the first text box and leave the other two text boxes empty.
- Click the *Values* cell in the *age* row, and then click the button on the right side of the cell to open the Value Labels dialog box.
- Type 999 in the Value field.
- Type No Response in the Label field.

Missing Values for a String Variable

- Missing values for string variables are handled similarly to the missing values for numeric variables. However, unlike numeric variables, empty fields in string variables are not designated as system-missing. Rather, they are interpreted as an empty string.
  - Click the **Variable View** tab at the bottom of the Data Editor window.
  - Click the *Missing* cell in the *sex* row, and then click the button on the right side of the cell to open the Missing Values dialog box.
  - Select **Discrete missing values**.
  - Type **NR** in the first text box.
  - Missing values for string variables are case sensitive. So, a value of ***nr*** is not treated as a missing value.

Hands On Time

- Download IBM SPSS
- Exercise Files: http://bit.ly/2MVaTmv

Tell us what you think | Verify your attendance

Provide feedback on this session and verify your participation at

https://u.mcmaster.ca/verification

This session's verification code:

[data-is-cool]

scds.ca
scds@mcmaster.ca

Sherman Centre for Digital Scholarship

SCDS Links

Send SCDS an Email:

scds@mcmaster.ca

Register for a Workshop:

https://u.mcmaster.ca/scds-workshops

Subscribe to our Newsletter:

https://u.mcmaster.ca/sign-up

Schedule a Consultation:

https://libcal.mcmaster.ca/appointments

McMaster University Library