# DATA MANAGEMENT IN SPSS

# Practical tips on SPSS set-up:

In the menu click on 'Edit / Options':

🔀 GUSSW 3B_30.sav [DataSet2] - SPSS Data Editor									
File Edit	View Data	Transform Analy	yze Graph	s Utilities Win	idow Help				
	😕 🖬 🖻 🚸 🤲 🐘 🕸 捕 🏛 🏛 🐺 😵 🕥 🌑								
	Name	Туре	Width	Decimals	Label	Values	Missing	Columns	Align ≙
1	ldnumber	Numeric	8	2	Serial number	None	None	8	Right
2	SampType	Numeric	2	0	Sample Type	{1, Birth}	-91	8	Right
3	DcXmnth1	Numeric	2	0	Dc Month of in	{1, January}	-91	10	Right
4	DcXqurt1	Numeric	2	0	Dc Quarter of i	{1, January to	-91	10	Right
5	Outcome	Numeric	3	0	Dc Final outco	{110, Full inter	-91	8	Right
6	McHGrsp03	Numeric	2	0	Mc - Same res	{-9, Refusal}	-91	8	Right
7	DcHGrsp01	Numeric	2	0	Dc - Whether r	{0, Not mentio	-91	8	Right
8	DcHGrsp02	Numeric	2	0	Dc - Whether r	{0, Not mentio	-91	8	Right
9	DcHGnmad	Numeric	2	0	Dc Number of	None	-91	10	Right
10	DcHGnmad	Numeric	2	0	Dc Number of	{0, None}	-91	11	Right
11	DcHGnmkd	Numeric	2	0	Dc Number of	None	-91	10	Right
12	DcHGnmk2	Numeric	2	0	Dc Number of	{1, One}	-91	10	Right
13	DcHGhsiz	Numeric	2	0	Dc Household	None	-91	10	Right
14	DcHGnmsb	Numeric	2	0	Dc - Number of	None	-91	8	Right
15	DcHGnp01	Numeric	2	0	Dc - Number of	{O, Not mentio	-91	8	Right
16	DcHGnp02	Numeric	2	0	Dc - Natural m	{0, Not mentio	-91	8	Right
17	DcHGnp03	Numeric	2	0	Dc - Natural fat	{0, Not mentio	-91	8	Right
18	DcHGnp04	Numeric	2	0	Dc - Responde	{O, Responden	-91	8	Right
19	DcHGrsp04	Numeric	2	0	Dc - Family Ty	{0, Lone Paren	-91	11	Right
20	DcHGrsp05	Numeric	2	0	Dc Whether re	{0, No}	-91	11	Right
< ►	ata View λVa	riable View /	-	-	<	(m. 11.)	· ·	]	>
	SPSS Processor is ready								

The 'Options' dialog box opens at the 'General' tab:

ptions			Þ
Data C	urrency	Sc	ripts
General Viewer Draft Viewer Outp	ut Labels Chart:	Interactive	Pivot Tables
Variable Lists Display labels Display names Alphabetical Eile Session Journal Precord syntax in Journal Append Overwrite F:\SPSSJO*1.JNL Browse Temporary directory:	Output No scientific in tables Viewer Type at S ● Regular Measurement Sy Language: Notification:	notation for small nu tartup: Draft stem: Points English	umbers
C:\DOCUME~1\MIREIL~1\LOCALS~1\T Recently used file list: 9	✓ Raise viewe ✓ Scroll to new Sound: <ul> <li>Na</li> <li>So</li> </ul>	window output one OSystem I und Browse.	beep 
	IK Cance	Apply	Help

Select the option 'Display names' in the 'Variable Lists'

In the 'Viewer' tab tick 'Display commands in the log' at the bottom left: if you run some syntax, it will be displayed in the output screen which opens up automatically when a dataset is opened. This helps spot any errors in the syntax you write.

Options					
Data	Currency	Sci	ripts		
General Viewer Draft Viewer	ewer Output Labels Charts	Interactive	Pivot Tables		
Initial Output State Item: Log V Contents are initially: Shown Hidden Justification: Align left Center Align right V Display commands in the log	Title Font         Arial       14         Text Output Page Size         Width:       Standard (80 characters)         Wide (132 characters)         O Lustom:       255         Text Output Font         Courier New       10	B Z U Length: Standard (59 I Infinite Custom: 59 B Z U Monospace	ines)		
OK Cancel Apply Help					

In the 'Output Labels' tab select the following options in the drop down lists:

ptions					
Data		Currency		Sci	ripts
General Viewer	Draft Viewer	Output Labels	Charts	Interactive	Pivot Tables
Outline Labeling Variables in item labels s	shown as:				
Names and Labels Variable values in item I	abels shown as:				
Pivot Table Labeling Variables in labels show	n as:				
Names and Labels	shown to:	•			
Values and Labels	s shown as.				
		ОК	Cancel	Apply	Help

## Examples of frequencies & cross-tabs via menu:

# 1) Select 'Analyze' in the menu

🖬 GUSS	🖸 GUSSW3B_30.sav [DataSet1] - SPSS Qata Editor								
File Edit	View Data	Transform Anal	yze Gr <mark>a</mark> ph	is Utilities Win	idow Help				
🗁 🔲	≥ 🛯 💩 ↔ ಈ 🏪 🛊 # 推 🖩 Φ 🛒 🎯 🔍								
	Name	Туре	Width	Decimals	Label	Values	Missing	Columns	Align 📤
1	ldnumber	Numeric	8	2	Serial number	None	None	8	Right
2	SampType	Numeric	2	0	Sample Type	{1, Birth}	-91	8	Right
3	DcXmnth1	Numeric	2	0	Dc Month of in	{1, January}	-91	10	Right
4	DcXqurt1	Numeric	2	0	Dc Quarter of i	{1, January to	-91	10	Right
5	Outcome	Numeric	3	0	Dc Final outco	{110, Full inter	-91	8	Right
6	McHGrsp03	Numeric	2	0	Mc - Same res	{-9, Refusal}	-91	8	Right
7	DcHGrsp01	Numeric	2	0	Dc - Whether r	{O, Not mentio	-91	8	Right
8	DcHGrsp02	Numeric	2	0	Dc - Whether r	{O, Not mentio	-91	8	Right
9	DcHGnmad	Numeric	2	0	Dc Number of	None	-91	10	Right
10	DcHGnmad	Numeric	2	0	Dc Number of	{0, None}	-91	11	Right
11	DcHGnmkd	Numeric	2	0	Dc Number of	None	-91	10	Right
12	DcHGnmk2	Numeric	2	0	Dc Number of	{1, One}	-91	10	Right
13	DcHGhsiz	Numeric	2	0	Dc Household	None	-91	10	Right
14	DcHGnmsb	Numeric	2	0	Dc - Number of	None	-91	8	Right
15	DcHGnp01	Numeric	2	0	Dc - Number of	{O, Not mentio	-91	8	Right
16	DcHGnp02	Numeric	2	0	Dc - Natural m	{O, Not mentio	-91	8	Right
17	DcHGnp03	Numeric	2	0	Dc - Natural fat	{O, Not mentio	-91	8	Right
18	DcHGnp04	Numeric	2	0	Dc - Responde	{O, Responden	-91	8	Right
19	DcHGrsp04	Numeric	2	0	Dc - Family Ty	{0, Lone Paren	-91	11	Right
20	DcHGrsp05	Numeric	2	0	Dc Whether re	{0, No}	-91	11	Right 🔜
■ D \ D	ata View λVa	riable View /	10	-		100 11 1		]	>
			S	PSS Processor is	ready				

2) Select 'Descriptive statistics -> Frequencies' in the Analyze drop-down menu: the following dialog box will appear:

Frequencies	
McNMst01 McNMtk01 McNMtk02 McNMtr01 McNMvw01 McNrel03 McObtg01 McObtg02 McObtg02 McObtg02 McObtg02	Variable(s): OK Paste Reset Cancel Help
Statisti	ics Charts Format

3) Browse the list of variables on the left and highlight the one(s) you want and click on the arrow in the middle:

Frequencies			X		
McNMst01  McNMtk01  McNMtk02  McNMtr01  McNMtr01  McNMtr03  McNtrel03  McObtg02  McObtg03  McObtg03  McObtg03	Ţ	Variable(s):	OK Paste Reset Cancel Help		
✓ Display frequency tables					
Statistics Charts Format					

4) Click 'OK': the output will show you the frequency table for that variable

# McObtg01 Mc - Attend group in last 12 months

				Valid	Cumulative
		Frequency	Percent	Percent	Percent
Valid	1 Yes	568	45.6	45.6	45.6
	2 No	678	54.4	54.4	100.0
	Total	1246	100.0	100.0	

5) To do cross-tabs, use the same menu as for frequencies above but select 'Crosstabs' instead: Analyse  $\rightarrow$  Descriptive statistics  $\rightarrow$  Crosstabs

Crosstabs		×
AncB AncC CAssPrb1 CAssPrb2 CAssPrb3 CAssPrb4 CAssPrb5 CAssPrb5 CAssPrb6 PaEthGpM DaEthGpP DathGpP DaMedu01 PaMsec01 DaMsec01	Row(s): Column(s): Layer 1 of 1 Previous Next	OK Paste Reset Cancel Help
Display clustered bar ch	arts Statistics Cells Format	

6) As before browse and highlight the variables you want to cross-reference and move them to the row and column boxes as applicable:

Crosstabs		×
McNMint1         McNMkd01         McNMkd02         McNMkd02         McNMst01         McNMst03         McNMst01         McNMst03         McNMst03         McNst03         McNst03         McObstg03	Row(s): Column(s): Aver 1 of 1 Previous Next	OK Paste Reset Cancel Help
Display clustered bar cha Suppress tables	arts Statistics) Cells Format	

7) Click 'OK': the following output is displayed

McObtg02 Mc - Not group - nothing suitable \* McObtg01 Mc - Attend group in last 12 months Crosstabulation

Count

Oount			
		McObtg0 1 Mc - Attend group in last 12 months	
		2 No	Total
McObtg02 Mc - Not group - nothing	0 Not mentioned	624	624
suitable	1 Mentioned	54	54
Total		678	678

Merging datasets via menu : see separate handout

### Recoding variables via syntax

example: study child's general health between Sw2 and Sw3

1) Check frequencies of the original variables MbHgen01 and McHgen01

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1very good,	898	72.1	72.1	72.1
	2 good,	288	23.1	23.1	95.2
	3 fair,	56	4.5	4.5	99.7
	4 bad,	4	.3	.3	100.0
	Total	1246	100.0	100.0	

### McHgen01 Mc - Childs general health

MbHgen01 Mb - Child's general health

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Very good	829	66.5	67.7	67.7
	2 Good	313	25.1	25.6	93.2
	3 Fair	79	6.3	6.4	99.7
	4 Bad	2	.2	.2	99.8
	5 Very bad	2	.2	.2	100.0
	Total	1225	98.3	100.0	
Missing	System	21	1.7		
Total		1246	100.0		

21 sysmis in MbHgen01 = cases which were at Sweep 3 but not at Sweep 2 (previous sweep = Sweep 1 for those cases)

- 2) Open syntax file: go to menu 'File / Open / Syntax ...'
- 3) Type Recode syntax in the syntax file: we want to group the original variables into answer categories Good (1,2) / Fair (3) and Bad (4,5)

Example for Sweep 3:

RECODE McHgen01 (1 thru 2=1) (3=2) (4 thru 5=3) (else=copy) INTO GenHbdS3. Exe.

4) and run it:

😫 Recode_Etc - SPSS Syntax Editor	×
File Edit View Data Transform Analyze Graphs Utilities Run Window Help	
🗁 🖬 📴 <table-cell-rows> 🥅 🐜 🕼 🗰 🍋 🔍 🖤</table-cell-rows>	
********* RECODE SYNTAX **********	^
* Sweep 3 general health variable:	
RECODE McHgen01 (1 thru 2=1) (3=2) (4 thru 5=3) (else=copy) INTO GenHbdS3. Exe.	
VAR LABS GenHbdS3 'Childs general health - banded'.	
VAL LABS GenHbdS3 1 'Good/Very Good' 2 'Fair' 3 'Bad/Very Bad'. FORMATS GenHbdS3 (F2.0).	Ш
FREQ GenHbdS3.	
* Sweep 2 general health variable:	
Recode MbHgen01 (1 thru 2=1) (3=2) (4 thru 5=3) (else=copy) INTO GenHbdS2. Exe.	
RECODE GenHbdS2 (sysmis=-1). Exe.	
VAR LABS GenHbdS2 'Childs general health - banded'. VAL LABS GenHbdS2 1 'Good/Very Good' 2 'Fair' 3 'Bad/Very Bad' -1 'Not applicable: not at Sw3'. FORMATS GenHbdS2 (F2.0). MISSING VALUES GenHbdS2 (-1).	
FREQ GenHbdS2.	~
SPSS Processor is ready	

5) Check frequencies of the new variable:

## GenHbdS3

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	1186	95.2	95.2	95.2
	2.00	56	4.5	4.5	99.7
	3.00	4	.3	.3	100.0
	Total	1246	100.0	100.0	

6) Tidy up variable label, value labels and output format directly within the variable row in the dataset 'Variable View' or with additional syntax below:

VAR LABS GenHbdS3 'Childs general health - banded'.

VAL LABS GenHbdS3 1 'Good/Very Good' 2 'Fair' 3 'Bad/Very Bad'. FORMATS GenHbdS3 (F2.0).

7) Run the frequencies again:

### GenHbdS3 Childs general health - banded

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Good/Very Good	1186	95.2	95.2	95.2
	2 Fair	56	4.5	4.5	99.7
	3 Bad/Very Bad	4	.3	.3	100.0
	Total	1246	100.0	100.0	

8) If similar Recode syntax run for Sweep 2 variable:

#### GenHbdS2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	1142	91.7	93.2	93.2
	2.00	79	6.3	6.4	99.7
	3.00	4	.3	.3	100.0
	Total	1225	98.3	100.0	
Missing	System	21	1.7		
Total		1246	100.0		

9) So adding a line to recode the system missings into a missing code:

RECODE GenHbdS2 (sysmis=-1). Exe.

#### GenHbdS2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	-1.00	21	1.7	1.7	1.7
	1.00	1142	91.7	91.7	93.3
	2.00	79	6.3	6.3	99.7
	3.00	4	.3	.3	100.0
	Total	1246	100.0	100.0	

10) And running the frequencies again after tidying up variable label, value labels and output format:

VAR LABS GenHbdS2 'Childs general health - banded'. VAL LABS GenHbdS2 1 'Good/Very Good' 2 'Fair' 3 'Bad/Very Bad' -1 'Not applicable: not at Sw3'. FORMATS GenHbdS2 (F2.0). **MISSING VALUES GenHbdS2 (-1).** 

#### GenHbdS2 Childs general health - banded

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Good/Very Good	1142	91.7	93.2	93.2
	2 Fair	79	6.3	6.4	99.7
	3 Bad/Very Bad	4	.3	.3	100.0
	Total	1225	98.3	100.0	
Missing	-1 Not applicable: not at Sw3	21	1.7		
Total		1246	100.0		

### Computing a Derived Variable using syntax

example: evolution of study child's general health between Sw2 and Sw3

- 1) Create a new variable 'GenHS2S3' coded 0 if no change between Sw2 and Sw3, 1 if child's health not as good at Sw3 and 2 if child's health better at Sw3
- 2) Set up and run Compute syntax:

COMPUTE GenHS2S3=-1. IF (GenHbdS3=GenHbdS2) GenHS2S3=0. IF (GenHbdS3<GenHbdS2) GenHS2S3=1. IF (GenHbdS3>GenHbdS2) GenHS2S3=2. Exe.

GenHS2S3

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	-1.00	21	1.7	1.7	1.7
	.00	1119	89.8	89.8	91.5
	1.00	66	5.3	5.3	96.8
	2.00	40	3.2	3.2	100.0
	Total	1246	100.0	100.0	

The same number of Sw2 system missings have been coded as -1 by default

3) Check frequencies, tidy up variable and value labels, output formats:

### 2 ' not as good at Sw3' -1 'Not Applicable: no data at Sw2'. FORMATS GenHS2S3 (F2.0). **MISSING VALUES GenHS2S3 (-1).**

### GenHS2S3 Childs general health evolution Sw2 to Sw3

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 No change	1119	89.8	91.3	91.3
	1 better at Sw3	66	5.3	5.4	96.7
	2 not as good at Sw3	40	3.2	3.3	100.0
	Total	1225	98.3	100.0	
Missing	-1 Not Applicable: no data at Sw2	21	1.7		
Total		1246	100.0		