Value Labelling the School Register of Needs Survey 2000

Alex Montgomery

March 6, 2012

DataFirst

1 Creating Integer Variables out of String Variables

In the School Register of Needs Survey 2000 (SRNS) dataset, a number of variables that define schools into various categories were incorrectly specified as string variables that only vaguely reflected those categories. In many cases, schools that should probably be grouped together for certain categorical variables are incorrectly defined as being part of different categories because of slight differences in their string values (despite those string values suggesting that they are part of semantically similar groups). For example, some schools were assigned the string value "Mud" for the variable "q2_8_1" (which reflects the material out of which the external wall of the school is built), but others were assigned the value "mud". Because the variable is string valued, most statistical software packages do not recognize that these values for those variables as equivalent (that is, they are invariable case sensitive). Furthermore, integer values for categorical variables are easier objects with which to work when programming. These factors motivate the conversion of numerous variables in the SRNS 2000 from string to integer value variables and assigning them values congruent with those suggested by the questionnaire (the only available source of information on the *true* value labels for categorical variables that respondents would observe when answering the questionnaire).

To return to the example above, the questionnaire indicates that, of the two string values, only "Mud" is a technically valid categorical entry. Furthermore, it suggests that this string value should actually be coded with the integer value of 4. In this example, schools with values "Mud" and "mud", for variable "q2_8_1", are both recoded as "4" and assigned the value label "Mud". The variable is redefined as an integer.

We are agnostic as to the source of these errors, except that they are assumed to be *actual* errors (and not intentional classifications - we assume that there is no difference between "Mud" and "mud", to relate this to the above example). Some assumptions were made to assign as many of these variables into legitimate categories as possible, with some of these assumptions less potentially contentious than others. Given the dearth of metadata on the survey, DataFirst has essentially created a codebook (which is provided in the Appendix) based on the information in the questionnaire. This process separates the latest version of the data from those that were previously available through DataFirst, including the version prepared by the Development Policy Research Unit (DPRU). In total, 157 variables have been redefined as integer variables.

This version includes as much information on the variable and value labels as possible.

A Codebook

Table 1: Variable label - ceiling_material

Definitions	
Value	Label
1	Asbestos
2	Ceiling Board
3	None
4	Other
5	Wood
Variables using this value label	q2_8_3

Table 2: Variable label - former_edu_dpt

	Definitions
Value	Label
1	Department of Education and Training (DET)
2	Department of Education and Culture: House of Assembly (DEC:HoA)
3	Department of Education and Culture: House of Representatives (DEC: HoR)
4	Department of Education and Culture: House of Delegates (DEC: HoD)
5	Bophuthatswana Education Department
6	Ciskei Education Department
7	Gazankulu Department of Education
8	KaNgwane Department of Education
9	KwaNdebele Department of Education
10	KwaZulu Department of Education and Culture
11	Lebowa Department of Education
12	Qwa Qwa Department of Education
13	Transkei Education Department
14	Venda Education Department
15	New Department of Education (1994-2000)
16	Unknown/Other: 'KND'
17	Unknown/Other: 'MPU'
Variables using this value label	q1_19

4

Table 3: Variable label - hi_lo_grades

	Definitions	
Value		Label
1		Grade 1
2		Grade 2
3		Grade 3
4		Grade 4
5		Grade 5
6		Grade 6
7		Grade 7
8		Grade 8
9		Grade 9
10		Grade 10
11		Grade 11
12		Grade 12
13		PM
14		PP
15		R
Variables using this value label		high_grade low_grade

Table 4: Variable label - no_yes_other_welfare

	Definitions	
Value		Label
1		No
2		Yes, another function
3		Yes, function specified as 'welfare'
Variables using this value label		q2_12_1_6

Table 5: Variable label - power_type

	Definitions	
Value		Label
1		Wired and supplied with electricity
2		Wired but not supplied with electricity
3		Generators
4		Solar power
5		Other
Variables using this value label		q4_3_1

Table 6: Variable label - provcode_int

	Definitions
Value	Label
1	EC
2	FS
3	GT
4	KZ
5	LP
6	MP
7	NC
8	NW
9	WC
Variables using this value label	provcode

Table 7: Variable label - q1_12_3_int

	Definitions
Value	Label
1	Afternoon
2	Morning
3	Morning & Afternoon
Variables using this value label	q1_12_3

Table 8: Variable label - q1_13_3_int

	Definitions
Value	Label
1	Afternoon
2	Morning
3	Morning & Afternoon
Variables using this value label	q1_13_3

Table 9: Variable label - q1_15_int

	Definitions
Value	Label
1	Independent
2	Public
Variables using this value label	q1_15

Table 10: Variable label - q1_16_1_int

	Definitions
Value	Label
1	Church
2	Factory
3	Farm
4	Hospital
5	Mine
6	Trust Land
7	Private
8	Private Company
Variables using this value label	q1_16_1

Table 11: Variable label - q1_17_1_int

	Definitions
Value	Label
1	Church
2	Factory
3	Farm
4	Hospital
5	Mine
6	Trust Land
7	Private
8	Private Company
Variables using this value label	q1_17_1

Table 12: Variable label - q2_6_1_1_int

	Definitions
Value	Label
1	Church
Variables using this value label	q2_6_1_1

Table 13: Variable label - q2_6_1_2_int

	Definitions
Value	Label
1	Community Hall
Variables using this value label	q2_6_1_2

Table 14: Variable label - q2_6_1_3_int

	Definitions
Value	Label
1	Another School
Variables using this value label	q2_6_1_3

Table 15: Variable label - q2_6_1_4_int

	Definitions
Value	Label
1	Other
Variables using this value label	q2_6_1_4

Table 16: Variable label - q2_8_1_int

	Definitions
Value	Label
1	Brick
2	Cement
3	Concrete
4	Mud
5	Mud/Clay
6	Other
7	Prefabricated
8	Stone
9	Wood
10	Zinc
Variables using this value label	q2_8_1

Table 17: Variable label - q2_8_2_int

	Definitions
Value	Label
1	Asbestos
2	Clay
3	Other
4	Thatch
5	Tiles
6	Zinc
Variables using this value label	q2_8_2

Table 18: Variable label - q2_8_4_int

	Definitions
Value	Label
1	Cement
2	Ground
3	Mud
4	Other
5	Thatch
6	Tiles
7	Wood
Variables using this value label	q2_8_4

Table 19: Variable label - q2_9_int

Value	Label
	Definitions
Value	Label
1	Good
2	Minor Repairs
3	New
4	Renovated
5	Very Weak
6	Weak
Variables using this value label	q2_9

Table 20: Variable label - q4_2_3_int

Value	Label
	Definitions
Value	Label
1	Dam
2	Other
3	Public Tap
4	Reservoir
5	Reservoir/Bore-hole/Tank
6	River
7	River/Stream
Variables using this value label	q4_2_3

Table 21: Variable label - q4_5_1_int

Value	Label
	Definitions
Value	Label
1	High
2	Low
3	Medium
Variables using this value label	q4_5_1

Table 22: Variable label - q4_8_2_1_int

Definitions	
Value	Label
1	No
2	Taxis
3	Yes
Variables using this value label	q4_8_2_1

Table 23: Variable label - q4_8_2_2_int

Definitions	
Value	Label
1	Buses
2	No
3	Yes
Variables using this value label	q4_8_2_2

Table 24: Variable label - q4_8_2_3_int

Definitions	
Value	Label
1	No
2	Other
3	Yes
Variables using this value label	q4_8_2_3

Table 25: Variable label - road_condition

Definitions	
Value	Label
1	Good tar road
2	Broken tar road
3	Good gravel/dirt road
4	Poor gravel/dirt road
5	Foot path only
6	Other
Variables using this value label	q4_6

Table 26: Variable label - schooltype_int

Definitions	
Value	Label
1	Combined
2	No respon
3	Primary
4	Secondary
Variables using this value label	schooltype

Table 27: Variable label - toilet_system

	Definitions
Value	Label
1	Flush system to sewer
2	Flush system to septic tank
3	Chemical Toilet
4	Pit latrine
5	Bucket system
6	'Flush system' (not sure if septic or sewer)'
Variables using this value label	q4_4_1

Table 28: Variable label - wall_condition

	Definitions
Value	Label
1	Good
2	Needs some repair
3	Bad
4	Other (?)
Variables using this value label	q4_5_3

Table 29: Variable label - wall_material

	Definitions
Value	Label
1	Mesh Wire
2	Barbed wire
3	Brick
4	Concrete/Cement
5	Wood
6	Other
Variables using this value label	q4_5_2

Table 30: Variable label - yes_no

	Definitions
Value	Label
0	No
1	Yes
Variables using this value label	q1_111 q1_14 q1_16 q1_17 q1_18_1 q1_18_10 q1_18_11 q1_18_12 q1_18_13 q1_18_14 q1_18_15 q1_18_2 q1_18_3 q1_18_4 q1_18_5 q1_18_6 q1_18_7 q1_18_8 q1_18_9 q2_1 q2_10 q2_10_5 q2_10_6 q2_10_7 q2_10_8 q2_11 q2_12 q2_12_1_1 q2_12_1_2 q2_12_1_3 q2_12_1_