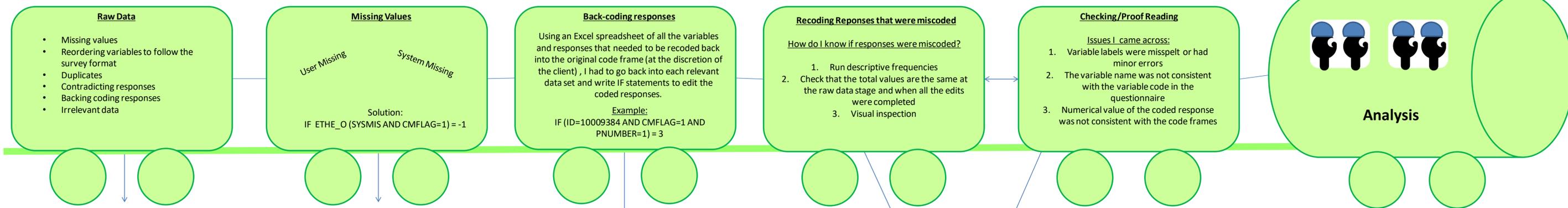


'Cooking' Raw Data: The Process of Data Cleansing

Outlining and dealing with the issues that arose when data cleansing the Millennium Cohort Study Sixth Sweep (MCS6)
By Nicole Amaning

What is Data Cleansing?

"Data cleaning, also called data cleansing or scrubbing, deals with detecting and removing errors and inconsistencies from data in order to improve the quality of data" – (Kumar and Chadrasekaran, 2011:1)
 "...can be tedious, but it can often prevent totally misleading results later on"
 "requires real detective work to uncover all possible errors that might be present"
 - (Albright et al, 2008:189)



Case Study 1: An Isolated Incident
A father of one of the cohort members whom did not live in the household but was present whilst the interviews were taking place requested to be interviewed as well. Although he was interviewed, he did not live in the household and was not supposed to be interviewed. His responses were therefore included in the raw data and we had to decide what to do with this data.

Case Study 2: Frequent Occurrence
Respondents would select 'Other – please specify response', and write their response. Often this response can be back coded into one of the pre coded responses, or if there are multiple responses with the same response the coding team will make a new coded response or keep it as an other response (not ideal). I have to then mimic their decision and carry out the edits on all the relevant data.

Reasons:
 •No data recorded (.)
 •Incomplete interview
 •"Not Applicable"
 •"Refusal"
 •"Don't Know"
 •"Vague Irrelevant Answer"

Example:
 Question: "Since the last interview has CM suffered any injuries?
 Pre coded Responses: "Open cut wound", "Head injury", "Broken bone", "Other – please specify".
Response when selected Other:
 Response 1: "Sophie hit her head and feel unconscious"
 Response 2: "David fainted in class and injured his head"
 Solution: Create a new coded variable called "Loss of consciousness" with clients approval and back code response 2 into the newly created code as well as "head injury" (if multi-coded question)
 Response 3: Sally had a paper cut
 Solution: Back code response into "open cut wound"
 Response 4: Suffered a nose bleed
 Solution: Keep coding as "other"
 Reflection: What about those who had a nose bleed but did not consider it an 'injury' hence selected 'no' to begin with?

– MAIN Reason relationship with other parent ended MC2

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Relationship reasons	7	.1	4.7	4.7
Financial reasons (eg unemployment)	39	.3	26.0	30.7
Health reasons (depressed, sick etc)	16	.1	10.7	41.3
Alcohol/drug problems	26	.2	17.3	58.7
Violent abusive behaviour	32	.2	21.3	80.0
Other (specify)	29	.2	19.3	99.3
Spontaneous: Relationship not ended	1	.0	.7	100.0
Total	150	1.1	100.0	
Missing Not applicable	13707	98.9		
Total	13857	100.0		

Source: Centre for Longitudinal Studies (2015), Millennium Cohort Study Fifth Sweep*

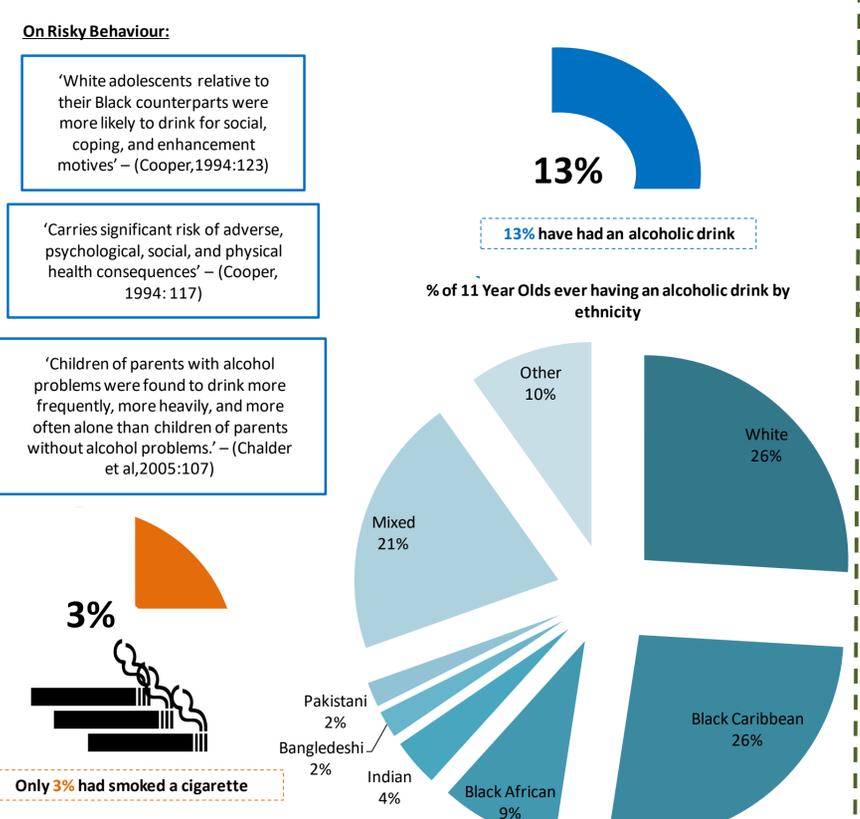
Bibliography:
 Albright, C.S. Winston, W. Zappe, C. (2008). Getting the data right: Cleansing the data. In: Data analysis and decision making with Microsoft excel, revised. London. Cengage learning. pp189-196
 Chalder, M. Elgar, F.J. Bennett, P. (2005). 'Drinking and motivations to drink among adolescent children of parents with alcohol problems'. *Alcohol and Alcoholism*. 41(1). 107-113
 Cooper, M.L (1994). 'Motivations for Alcohol Use Among Adolescents: Development and Validation of a Four Factor Model'. *Psychological Assessment*. 6(2). 117-128
 Kumar, R.K and Chadrasekaran, R.M.(2011). Attribute correction – Data cleaning using association rule and clustering methods. *International Journal of Data Mining and Knowledge Management Process*. 1(2). 22-32
 Meltzer, H and Gatward, R.(1999). *The mental health of children and adolescents in Great Britain*. London: Stationary Office
 ONS(2015)- *Insights into children's mental health and well-being*. [Online]. Available from: http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/dcp171766_420239.pdf. [Accessed on: 02/11/16]
 ONS(2015a)- *Families and Households*. [Online]. Available from: <http://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/bulletins/familiesandhouseholds/2015-11-05>. [Accessed on: 03/11/16]
 ONS(2013). *Divorces in England and Wales*. [Online]. Available from: <http://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/divorce/bulletins/divorcesinenglandandwales/2013>. [Accessed on: 02/11/16]
 ONS(2011)- *Stepfamilies in 2011*. [Online]. Available from: <http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/rel/family-demography/stepfamilies/2011/stepfamilies-rpt.html>. [Accessed on: 03/11/16]
 Platt, L(2014). *Millennium Cohort Study: Initial findings from the Age 11 survey*. London: Centre for Longitudinal Studies
 University of London. Institute of Education. Centre for Longitudinal Studies. (2015). *Millennium Cohort Study: Fifth Survey, 2012*. [data collection]. 2nd Edition. UK Data Service. SN: 7464. <http://dx.doi.org/10.5255/UKDA-SN-7464-2>.

MCS Data Outcome: On Family Structure:
 Most common family type was a married or civil partner couple family with dependent children (ONS,2015a)
 'Around 42% of marriages end up in divorce with the average marriage in 2013 lasting 11.7 years before a divorce was granted' (ONS,2015:1)
 11% of families with dependent children were stepfamilies (ONS,2011)



On Mental Health:
 'World Happiness cites research that found over half of children who have a mental health problem will suffer from mental ill-health as adults' – (ONS,2015:2)
 'Mental health of children and young people in GB (2004) found children with mental health were more likely to have time off school, unauthorised absences and less likely to have a network of friends/family' – (ONS,2015:2)
 10% of 11 year olds experienced a mental health problem in 2012.

Family structure: Children with cohabiting parents were more likely than those with married parents to have a mental health issue
 1 in 5 boys in single parent households had a mental health problem compared with 1 in 8 with girls
 Social Class/Poverty: Children with parents in unskilled occupations were 3x more likely to have a mental disorder (Meltzer, and Gatward, 1999)
 Mental health issues are TWICE as common in boys than girls



On Technology:
 'Nearly three quarters (72%) had their own mobile phone. More than a third of those with a mobile phone being able to access the internet from it (37%)' – (Platt,2014:18)

*Data from MCS5 due to public unavailability of MCS6 data