This is an extract from an exploratory exercise we undertook as part of the ESRC eBook (ES/K007246/1) project…
(see http://www.bristol.ac.uk/cmm/research/ebooks/): breaking down elements of an early workflow being developed by Ian Brunton-Smith…
...and Daniel McCarthy in the preliminary stages of a project investigating young people's participation in online crime.
Please note this resource is for exemplar purposes only: it is not designed to be a definitive.

Stage	Category	Sub-category	Question / objective	How achieved with software	Datasets / other objects	Other objects passed	Blocks	Decision points
	Hypotheses / Design		Lit review; research questions: Offending, Crime & Justice Survey (OCJS) provides potential opportunity to explore these. OCJS downloaded & initial exploration. Check details of OCJS tech reports.					
1	Data prep	Save/load	Start with the first cohort	Upload 2003 dataset	2003 v.1			Use another cohort? (there were 5); how about longitudinal analysis?
2	Data exploration	Table	What do the (potential) dependent variables look like? Are they suitable for modelling as is?	1-way tabulation (needs options) of each dependent variable of interest			A	
3	Data prep	Generate new / overwrite variables	No, need to recode some of the categories	Create new dependent variables (loops thru them) based on old with recoding of certain values, including missing values	2003 v.2			
4	Data exploration	Table	What do the recoded dependent variables look like? Are they suitable for modelling now?	1-way tabulation (needs options) of new dependent variables				
5			Better, but some variables have a lot of refusals; how does this break down by age?	2-way tabulation (needs options) of new dependent variables vs age				

6	Data prep	Save/load	Reveals too many refusals, so explore another cohort…	Upload 2004 dataset	2004 v.1			Use another cohort? (there were 5); how about longitudinal analysis?
7	Data exploration	Table	What do the (potential) dependent variables look like? Are they suitable for modelling as is?	1-way tabulation (needs options) of each dependent variable of interest			A	
8	Data prep	Generate new / overwrite variables	No, need to recode some of the categories	Create new dependent variables based on old with recoding of certain values	2004 v.2			
9	Data exploration	Table	What do the recoded dependent variables look like? Are they suitable for modelling now?	1-way tabulation (needs options) of new dependent variables				
10			How does this break down by age?	2-way tabulation (needs options) of new dependent variables vs age				
			Happier: no problems with high refusal; slightly high "Don't know"s; poss consider MI in future					Deal with missing data via another means: multiple imputation?
11	Data prep	Generate new / overwrite variables	Generate new dependent variable, based on old, but with more intuitive name & with different missing value code	Create new dependent variable based on old with recoding of certain values	2004 v.3		B	
12			Generate new dependent variable, based on values in two old variables, with more intuitive name	Create new dependent variable based on values in two old variables with recoding of certain values	2004 v.4			
	Data exploration	Table	What does new dependent variable look like: e.g. any apparent errors in what I've done?	1-way tabulation (needs options) of new dependent variable				

…here we break off from this initial stage of the workflow…

…and look at a later section…

37	Data exploration	Table	What do the family relations-related variables look like? Are they suitable for submitting to factor analysis as is?	1-way tabulation (needs options) of each variable of interest				
	Data prep	Generate new / overwrite variables	No, need to recode some of the categories	Create new dependent variables (loops thru them) based on old with recoding of certain values	2004 v.22		B	
	Data exploration	Table	What do my new variables look like?	1-way tabulation (needs options) of independent variables of interest				
38	Model fit	Correlation	I'm going to use polychoric correlation to facilitate factor analysis (FA) with my binary variables. I'll first look at a straightforward correlation matrix	Correlate variables wish to submit to FA				
39	Model fit	Correlation	Now I'll generate a matrix of polychoric correlations	Generate a matrix of polychoric correlations			C	
40	Post-process model			Display this polychoric correlation matrix		Using model output from #39		
41				Save polychoric correlation matrix under different name		Using model output from #39		
42			Need to know the sample size for factormat function	Display sample size of polychoric matrix		Using model output from #39		
				Assign this sample size to a global setting		Using model output from #39		

43	Model fit	Factor Analysis	Run factor analysis	Run factor analysis (using factormat function)		Using model output from #39		
	Post-process model		How did the FA go?	Rotate FA & inspect results		Using model output from #39		
44	Model fit		Heywood case detected; also some low loadings; will drop two variables and run again	Generate a matrix of polychoric correlations (with two fewer variables)			C	
	Post-process model			Display this polychoric correlation matrix		Using model output from #44a		
				Save polychoric correlation matrix under different name		Using model output from #44a		
				Display sample size of polychoric matrix		Using model output from #44a		
				Assign this sample size to a global setting		Using model output from #44a		
	Model fit			Run factor analysis (using factormat function)		Using model output from #44a		
	Post-process model			Rotate FA & inspect results		Using model output from #44a		
			Create factors of interest	Using predict function				

…etc. (the workflow continues, and undergoes further revisions as the project progresses.)