*Research question: Does the reason a respondent left the parental home, particularly if that reason is getting a job, predict the age at which they first left?*

*Note: no survey weights are used in this analysis*

Program:

1. Clean predictor and outcome variables in the 2017 Family GSS for analyses
2. Clean select covariates
3. Save recode dofile
4. Generate analysis dofile
5. Generate univariate descriptive statistics for the outcome variable
6. Generate multivariate regression

Program was created on September 20th 2022 using…

Operating system: MacOS Monterey Version 12.6; Apple M1 Chip; 8 GB memory

Statistical Software: STATA 13

Data set: 2017 Family Canadian General Social Survey CSV File (Import this dataset into your Do-file before starting part 1, generate new do file)

1. Clean/recode predictor and outcome variable for analyses
	1. Outcome variable: Age at first home leaving (listed in code book as **alhomfc**)
		1. Recode missing values as (.)
		2. Recode **alhomfc** as **a 9-category variable** (0-9) and label each value in accordance to the ages within each bin, and name this new variable **AGELEAVEH** **\*\*\*note:** be mindful of the blue text below… it indicates which ages within the recoded age ranges are missing and therefore do not to be included in your code
			1. Assign a value of 0 to response 15, label as “Under 15”
			2. Assign a value of 1 to responses 15.1 thru 16.9, label as “15 – 16.9”
			3. Assign a value of 2 to responses 17 thru 18.9, label as “17 – 18.9”
			4. Assign a value of 3 to responses 19 thru 20.9, label as “19 – 20.9”
			5. Assign a value of 4 to responses 21 thru 22.9, label as “21 – 22.9”
			6. Assign a value of 5 to responses 23 thru 24.9, label as “23 – 24.9”
			7. Assign a value of 6 Combine 25 thru 26.8 and assign as a value of 6, label as “25 – 26.9”
				1. Omit 26.9 while regrouping the cases into this age range as they have no responses and thus aren’t in the dataset
			8. Assign a value of 7 to responses 27 thru 28.8, label as “27 – 28.9”
				1. Omit 27.2; 27.4; 27.7; 28.2; 28.4; 28.6; 28.9 while regrouping the cases into this age range as they have no responses and thus aren’t in the dataset
			9. Assign a value of 8 to responses 29 thru 30.8, label as “29 – 30.9”
				1. Omit 29.1; 29.2; 29.7; 30; 30.1; 30.6; 30.7; 30.9 while regrouping the cases into this age range as they have no responses and thus aren’t in the dataset
			10. Assign a value of 9 to the response 31, label as “31 Years and over”
		3. Label **AGELEAVEH** as “Age resp. left home”
	2. Predictor variable: Respondent first left parental home for a job (listed in codebook as **realftfb**)
		1. Recode missing values as (.)
		2. Rename realftfb **JOB** and label as **“Move for a Job”**
		3. Label value 1 as “Yes” and value 2 as “No”
2. Clean select covariates
	1. Canadian Citizenship (listed in codebook as **bpr\_19**)
		1. Recode missing values as (.)
		2. Rename bpr\_19 as **CAN**
	2. Age of respondent (listed in codebook as **agegr10**)
		1. Rename agegr10 as **AGE10**
	3. Family living situation in childhood (listed in codebook as **gu\_120**)
		1. Recode missing values as (.)
		2. Rename gu\_120 as **CHILDHOOD**
	4. Highest education obtained by the respondent (listed in codebook as **ehg3\_01b**)
		1. Recode missing values as (.)
		2. Rename ehg3\_01b as **EDUC**
	5. Personal income (listed in codebook as **ttlincg2**)
		1. Rename ttlincg2 as **INCOME**
3. Save recode do-file
	1. Save this dofile/recoded and cleaned data as “FamilyGSSRecode”
4. Generate analysis do-file
	1. Open new dofile and load the recode data “FamilyGSSRecode” to begin your analysis with cleaned data
	2. Save this new dofile as “FamilyGSSAnalysis”
5. Generate univariate descriptive statistics for outcome variable
	1. Generate a frequency table for **AGELEAVEH**
6. Generate multivariate regression
	1. Regress **AGELEAVEH** with **JOB** along with each of the **select** **covariates**