Research Replicability and Workflow Management
Data Activity
Yue Teng

Goal:

1. Using binary logistic regression model to examine how do age, sex, marital status, minority status and education level influence whether people have child or not.
2. Independent variable: age, sex, marital status, minority status and education level;
3. Dependent variable: Having children or not.

Data and variable used in the analysis:

- GSS_family_2017
- SEX (sex of the respondent), AGEGR10 (age group of 10), MARSTST (marital status), VISMIN(visible minority status), EHG3_01B (highest certificate, diploma or degree), TOTCHDC (children number reported by the respondent).

Data cleaning:

1. Subset missing values as “NA”
2. Recode sex into dummy variable (Female=1, Male =0)
3. Recode Minority status into dummy variable (Minority=1, Non-minority=0)
4. Recode marital status into dummy variable (Have partner = 1, No partner=0)
5. Recode education level into dummy variable (Low education (1-3) = 0, High education (4-7)=1)
6. Recode Children number reported into dummy variable (No child= 0, more than one child =1)

Descriptive statistics:

1. Use base R to run descriptive statistics for each categorical variable included in the analysis

Logistic regression model:

1. Run logistic model and summarize the model

Table 1 Descriptive statistics

Age group	15-24	0.0746
	25-34	0.1364
	35-44	0.1551
	45-54	0.1489
	55-64	0.1995
	65-74	0.1753
	>75	0.1103
Have child or not	Childless	0.3024
	More than one child	0.6976
Partner	No partner	0.5387
	Have partner	0.4613
Minority	Non-minority	0.8751
	Minority	0.1249
Female	Male	0.4562
	Female	0.5438
Education level	Low education	0.4623
	High education	0.5377

Table 2 Logistic regression results

	Coefficients:	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.00334	0.05541	-36.156	< 2e-16 ***
Female	0.44977	0.037	12.155	< 2e-16 ***
Minority	-0.3051	0.05383	-5.668	1.44e-08 ***
Partner	1.80349	0.04056	44.467	< 2e-16 ***
High_edu	-0.29584	0.03807	-7.77	7.82e-15 ***
AGEGR10	0.54022	0.01104	48.919	< 2e-16 ***
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				