

Research Milestone 2: Clean and Describe Variables

You submission should have two files:

1. PDF file with the following:

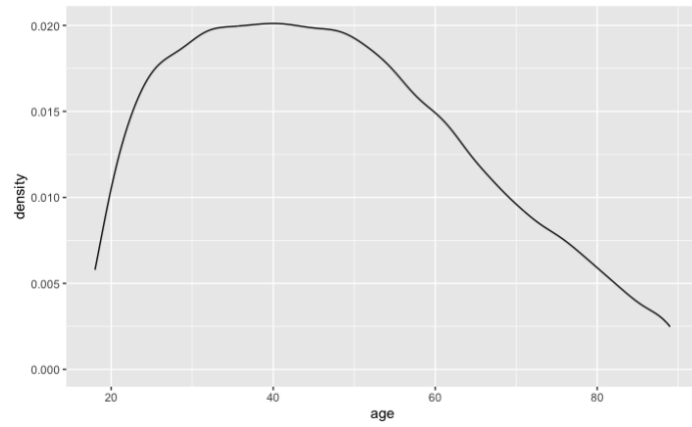
- a) 500-1000 word writeup that contains the following:
 - a. The data source (provide a URL)
 - b. A brief description of the data set (unit of observation, time frame, etc.)
 - c. For each categorical variable, describe how many levels your categorical variable started with and how you collapsed levels if it was needed (5 levels max).
 - d. Summary tables describing your eight variables (see example below)
 - e. Bar plot for every categorical variable, density plot or histogram for every quantitative variable. Histogram is recommended for discrete quantitative variables, density for continuous quantitative variables.
 - f. Descriptions of the variables based on your summary tables and figures:
 - i. Quantitative variables: describe the distribution, missing values
 - ii. Categorical variables: describe the distribution, missing values
 - g. Generative AI acknowledgement:
 - i. If you used generative AI, please explain what you used it for in a few sentences. Please remember that you cannot use it for your writeup.
 - ii. If you did not use generative AI, please include the following statement: *I did not use generative AI for any portion of this assignment. I understand that not disclosing the use of generative AI is a violation of the academic integrity policy.*

2. **R Script.** Upload the R script that loads your data files into R and generates the output included in your writeup.

Summary Table Examples

Quantitative Variables

variable name	N	mean	min	p50	max
age	245	30.5	20	29	90



Categorical Variables

Variable	Levels	Frequency
Race	White	1204
	Black	230
	Other	456

