Project #2—Demographic Exploratory Data Analysis

In this project, you will use the R programming language along with any packages you have learned thus far to analyze demographic data. You will demonstrate your understanding of various data transformation functions such as *filter, mutate, summarize, select, arrange, group_by*, etc., to examine the data and discern interesting patterns. Along with the aforesaid transformation functions, you will create plots using ggplot2 to visualize and help interpret the demographic trends and/or patterns in the dataset. The goal is thus to analyze the dataset and provide detailed discussions on what patterns and trends you see given these sample population characteristics. [Note the code must run; otherwise, points will be deducted]

As in Project #1, load the necessary R libraries and read in the provided demographic.csv file. The variables in this dataset (which are fairly self-explanatory) are: *first_name, last_name, gender, age, education, marital_status, occupation_group, credit_score, address, city, state,* and *home_build_year*.

1. Review the data table and look at the data types, distributions, values, etc.
    a. Detail any interesting patterns or issues you see.
2. Perform any necessary data cleaning and/or transformations. This could include, but not limited to, missing data, outliers, generation of new variables, binning, etc.
    a. In addition to the code, explain in detail what you did and why you did it?
    b. Are you making any assumptions? If so, what are they?
3. Summarize and/or aggregate the data table values in various ways with descriptive stats, counts, etc. over the entire dataset and by various groupings.
    a. Explain each summarization step and why you did it. What patterns are you seeing?
4. Leveraging the analyses in steps 1-3, create at least four different plots over variables you finding interesting to include univariate and multivariate (covariation) analyses. Make sure the plots are customized appropriately with labels, titles, colors, and themes.
    a. Explain each visual and why you chose to use this particular plot.
    b. Interpret what each plot is showing and what this could say about the demographics provided (e.g. trends such as age distribution or gender distribution).
    c. What overall patterns and/or trends do you see?
    d. How do the visuals add to your previous review and summarization?
    e. Do the visuals you provided tell a bigger story (when looked at together), i.e. do the individual plots coalesce into a larger narrative about the demographic data?
5. Summarize your interpretation of the overall results of your demographic analysis, discussing any interesting insights or trends you discovered. Posit what could be done with your analysis results—could this demographic data lead to actionable insights?

Grading Rubric:
Data transformations (20 pts)
Data summarization (20 pts)
Visualizations (30 pts)
Results Discussion (30 pts)