Healthcare Analysis:
Tori Green
Data Analysis with R

1. As determined by Graph 1.1 and 1.2, the age category of people who frequent the hospital and has the maximum expenditure is 9 years old; followed by the 3 and 5-year-old age groups.

   The mean and median values of hospital costs for the 9-year-old age group is $10,573.50 with the number of patients in this age group being 2 as compared to the overall average total charge per patient of $3,888.41.

2. The diagnosis-related group that has the highest overall average is "911" with an average cost of $48,388 followed by "602" at $29,188, and "421" at $26,356.

   The diagnosis-related group that has the highest overall length of stay is "602" with 41 days, followed by "421" at 39 days and "863" at 24 days.

3. There is not sufficient evidence to conclude that the race of a patient is related to the hospitalization costs.
   Null Hypothesis: $B_1 = 0$ (No Relationship between hospital costs and race)
   Alternate Hypothesis: $B_1 \neq 0$ (There is a relationship between hospital costs and race)

   The Multiple R-squared value is 0.0003299 which signifies that the regression relationship is not very strong. The p-value of 0.686 is greater than the 0.05 significance value, therefore, we fail to reject the null hypothesis that the total charge is not related to race.

4. As signified by the p-values of the multiple regression model, age and gender together are both statistically significant to total hospital costs. The p-values are 0.000763 and 0.034967 which are both less than the significant value of 0.05. However, the small multiple R-squared value of 0.0261 shows that only 2.6% of the variability in the model can be explained by these variables. For every increase in age (holding all other variables constant), total hospital costs are expected to increase by $86.28 and being male (all other variables constant) causes a decrease in hospital costs of $748.19.

   As previously shown, the age groups of 3, 5, and 9 have the highest overall mean expenditures. Table 4.1 illustrates that majority of the 3-year old patients consist of females. The 5-year old patients are only females and the 9-year-old patients are only males.

   Holding all other variables constant, gender is not statistically significant to the model. However, being a male patient, reduces the total cost of the hospital stay by $467.80.

On the other hand, age is statistically significant to the total cost as determined by the p-value of 0.00318. For every year change in age, the total cost increases by $73.68.

5. None of the variables are statistically significant to the model because all of the p-values are greater than the significant value of 0.05. From this, we can conclude that there is not enough statistical evidence to conclude that age, gender, and race are significant to In addition the R-squared value is 0.007898. Therefore, from the table in figure 6.1, we can conclude that 0.78% of the variability in the length of stay can be attributed to the variables of race, gender, and age. In other words, the regression relationship is not very strong.

6. Age (AGE), length of stay (LOS), and diagnosis related groups (APRDRG) are all statistically significant to the model when all explanatory variables are included as determined by their p-values in model6a.

   The variable that mainly affects hospital costs is “Length of stay in days” (LOS). Of the 55.36% R-squared value, LOS contributes 39.87%.