

Fun with Data Problems

Aniket Kesari

August 2, 2016

1. Make a scatterplot of the number of labs and the number of medications someone is on the first 100 patients. Describe the relationship between these two variables.
2. Create a new categorical variable that splits the patient population into two groups, patients who are on 16 or less medications, and those who are more than 16 medications. HINT: Try creating a variable named “med16” using the “ifelse” command.
3. Are African Americans more represented in the 16 or less medications population or in the more than 16 medications population? Create a visual or figure that best supports your answer. HINT: Use the “barplot” command to create your visualization.
4. Create a subset of the data for African American females. Describe the shape and center of the distribution of the number of labs in this population. Compare this distribution to that of African American males. Is there any difference? Can you compare these distributions easily with side-by-side boxplots? Why or why not?