Assignment 1: Data visualization

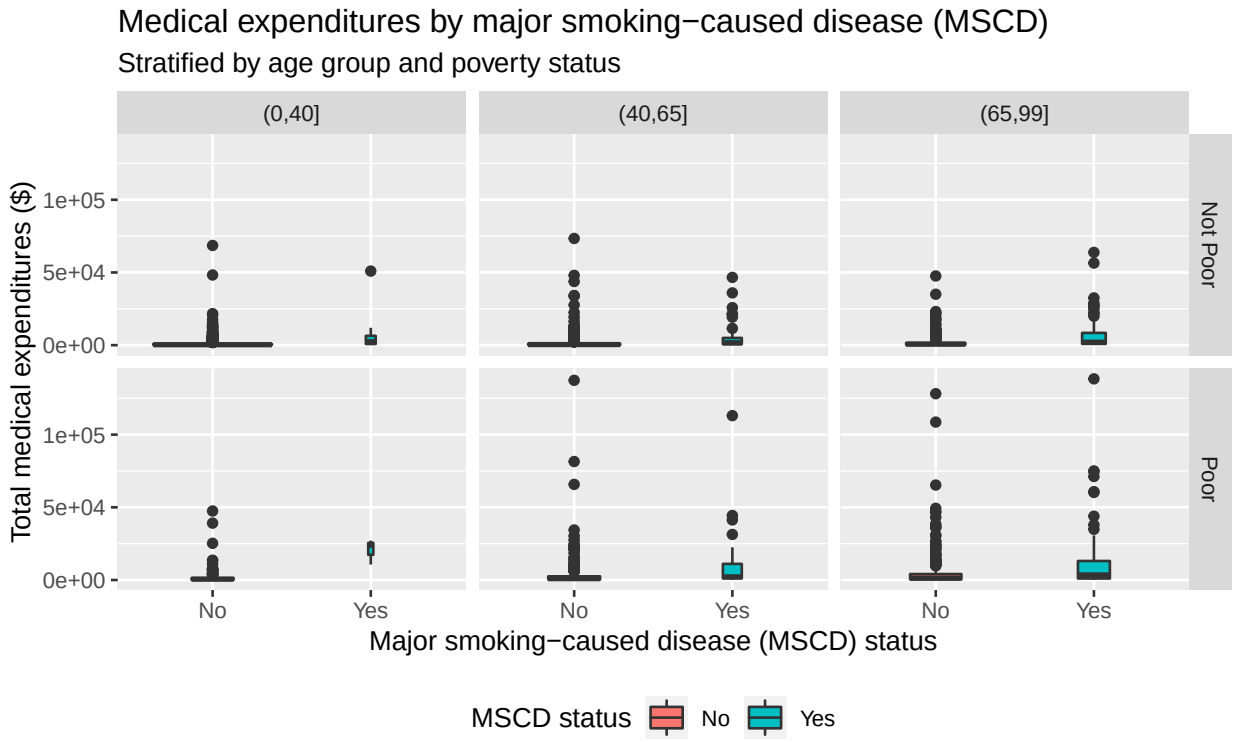
Due Sunday, September 25, 2022 @ 11:59pm

Public Health 280.346, Fall 2022 Your assignment for this module is to create graphs using the full National Medical Expenditures Survey dataset, rather than just the small subset of data we used in class. This full dataset can be found in the `nmes2018.rda` file in this repository on Github.

Use the `Assignment1_template.Rmd` file in this repository to complete your assignment. This template contains code to load the `tidyverse` and the data file.

Your assignment has two parts:

(1) The boxplot below shows the relationship between total medical expenditures (`totalexp`) and presence of a major smoking-caused disease (`mscd`) within different categories defined by age (`ageCat`) and poverty (`poor`). The first part of your assignment is to reproduce this graphic **exactly** using the `ggplot2` package in `R`. By **exactly**, I mean You should match every detail of this graphic, including labels, titles, and legends.



Note: This isn't the **best** plot to look at this relationship, as you will see in Module 2 in Public Health Biostatistics. We will improve this plot in the next homework assignment once we learn how to transform variables in `R`.

(2) Choose three variables from the NMES data set and make a graph that you think shows an interesting relationship between the variables. Be sure your graph includes meaningful labels, titles, and legends. Write a couple of sentences describing what your graph shows and why you think it's interesting.