HW 2, STAT 450

Due: Monday, September 9

Directions: The assignment should be completed using Quarto and submitted to Canvas as a self-contained HTML or PDF file.

The exercises in this assignment will use the CDC data set discussed in lecture 5. Run the following command to read this data set into R:

```
cdc <- read.csv("https://ericwfox.github.io/data/cdc1000.csv")
```

Exercise 1

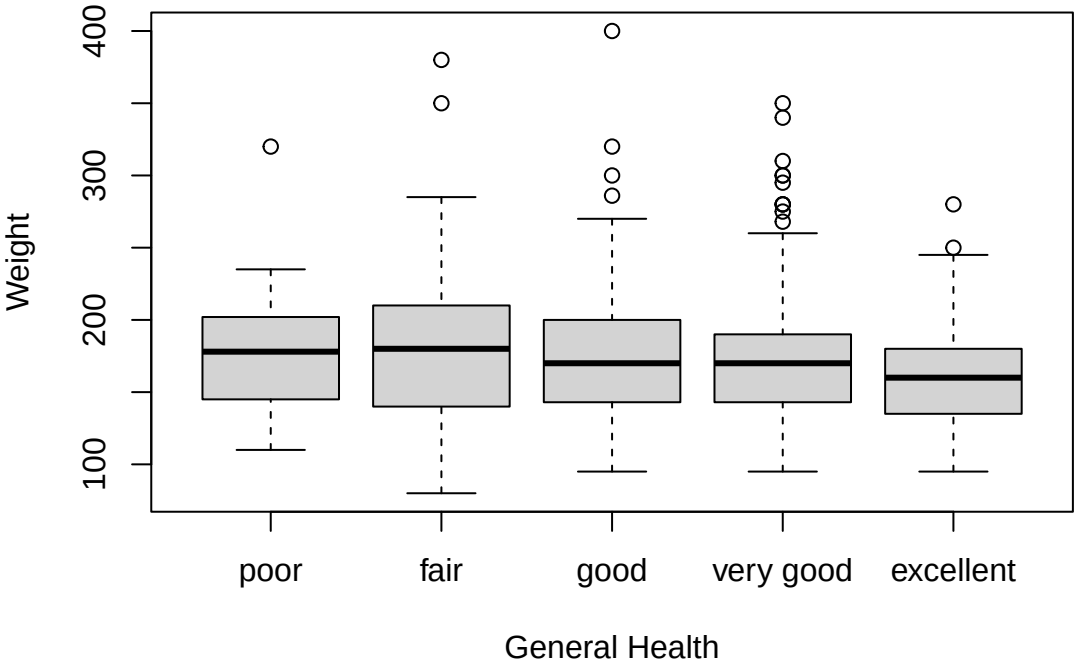
Make a bar plot of the variable `exerany`. Label the x-axis "Exercised in the last month" and label the bars "no" and "yes".

Exercise 2

Make a scatter plot using two numerical variables of your choosing. Describe the association between the variables in your scatter plot.

Exercise 3

Make side-by-side box plots with `genhlth` on the x-axis and `weight` on the y-axis. In the plot the categories of `genhlth` should be ordered as poor, fair, good, very good, excellent (hint: use the `factor()` function to specify the correct ordering). This is what the plot should look like:



Exercise 4

Make a new variable called `wtdiff` which is the difference between each person's desired weight, `wtdesire`, and current weight, `weight` (that is, desired weight − current weight). Compute some summary statistics, and plot a histogram and box plot of `wtdiff`. Comment on the shape and center of the distribution. Are there any outliers?