Lab2

source("http://www.openintro.org/stat/data/cdc.R")

names(cdc)

## [1] "genhlth" "exerany" "hlthplan" "smoke100" "height" "weight"   
## [7] "wtdesire" "age" "gender"

## Exercise 1: How many cases are there in this data set? How many variables? For each variable, identify its data type (i.e., nominal, ordinal, discrete numeric, or continuous numeric).

# Cases: 2000

# Variables: 9. GenHealth (ordinal),Exerany (ordinal), hlthplan (ordinal), smoke100 (discrete numeric), height (continuout numeric), weight (continuous numeric), wtdesire (continuous numeric), age (continuous numeric)

## Exercise 2: Generate 5-number summaries for both height and age (Hint: There should be 5 numbers in each summary, not 6), and compute the interquartile range for each.

# Height

quantile(cdc$height)

## 0% 25% 50% 75% 100%   
## 48 64 67 70 93

# Weight

quantile(cdc$weight)

## 0% 25% 50% 75% 100%   
## 68 140 165 190 500

## Exercise 2: Generate 5-number summaries for both height and age (Hint: There should be 5 numbers in each summary, not 6), and compute the interquartile range for each. Compute and plot the relative frequency distributions for hlthplan and exerany. How many females are in the sample? What percentage of the people in this sample report being in good health?