test 2 –practice

1. Data in R can be stored in **matrices** or **arrays** or other data structures. What are the two main data structures that we have used in class?

Vectors and data frames

1. Suppose you are running code in R, and you receive this error: **Error: could not find function "%>%"** What is the problem, and how can you fix it?

Library(dplyr) or library(tidyverse)

1. Consider these lines of code in R:

**round(53.24) log(53.24) read.csv(file.choose(), na.strings=’’)**

* 1. What is the technical term for **round, log**, and **read.csv**?

Functions?

* 1. What is the technical term for the value **53.24** in round and log, and the **na.strings=’’** in read.csv?

Arguments?

* 1. ggplot2 and dplyr are called \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ libraries
  2. BabyNames, NCHS and S are called \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ databases??

1. Log transformations:
   1. Why/When do we use log transformations with data? Data is skewed
   2. What is the value of log10(100000) ? ??
2. What is the ‘data-ink ratio’ that Tufte discusses? (Define the measure)

??

1. Tufte has one opinion on ‘chartjunk,’ while the blog post you read describes another. Briefly describe each opinion:
   1. Tufte:
   2. Blog:
2. The NCHS data frame (attached) is in your environment, and it is called **N**. (The individual vectors are not available by themselves.) It has over 30,000 lines; only a few are shown below.
   1. For the scatterplot on the last page, identify each variable and aesthetic. Then, **write the ggplot code** for making the graph.

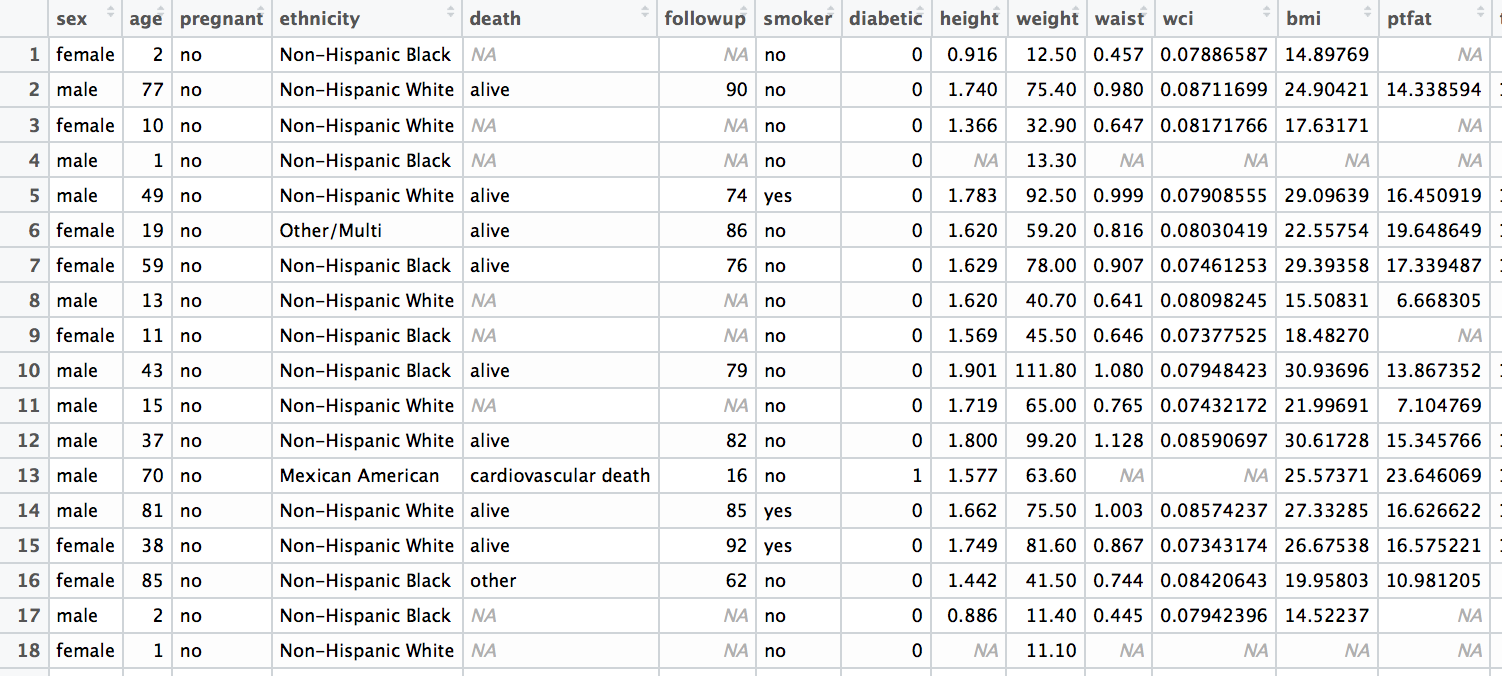
Variable Aesthetic

* 1. Write code that would produce the mean of all of the ptfat values (percentage fat). You may use dplyr OR base package code.
  2. Write **dplyr** code that will calculate the mean ptfat value for each different ethnicity. Use only 3 or so lines of code – don’t create many data frames. (Note: there are no missing values for ethnicity)
  3. Write **dplyr** code to create a new data frame, called **NFem** that includes only females in the data frame.
  4. Write ggplot code that would use data frame **N** to produce **this** bar graph



* 1. Suppose you wanted to create the above graph without the third bar (without the bar that has smoker equal to NA). Show how you could make a data frame, called **N2**, that could be used to make the graph above. (You do not need to make the graph – only create the data frame).

**Use this data for question 7**



**Use this graph for question 7a:**

