HW 02

Stat140-04

Notes on grading

In general, you will receive full points on the question if (1) there are no errors in your solution AND (2) the solution is written in highly articulate Statistical and English language. Point will be taken off if there are errors, your writing of the solution is incomplete, or there are issues with writing organization.

Problem 1

Below are the first few rows of a data frame named `NHANES`. NHANES stands for "National Health and Nutrition Examination Surveys", and the data frame contains information about the health of randomly sampled Americans.

```
##      ID Gender Age Weight Height   BMI BPSysAve BPDiaAve Diabetes
## 1 51624   male  34   87.4  164.7 32.22      113       85       No
## 2 51625   male   4   17.0  105.4 15.30       NA       NA       No
## 3 51630 female  49   86.7  168.4 30.57      112       75       No
## 4 51638   male   9   29.8  133.1 16.82       86       47       No
## 5 51646   male   8   35.2  130.6 20.64      107       37       No
## 6 51647 female  45   75.7  166.7 27.24      118       64       No
```

1. What is each observational unit in this data set?

2. For each of the following variables, is that variable categorical or quantitative?

    - `Gender`
    - `Height`
    - `Diabetes`

Problem 2

The following code creates a contingency table for the gender and diabetes variables in the `NHANES` dataset as in Problem 1. Please answer the following questions with the mathematical symbols for conditional, joint and marginal distribution.

```
table(NHANES$Gender, NHANES$Diabetes)
```

```
##
##             No  Yes
##   female 3088  269
##   male   3013  283
```

1. Calculate the joint distribution of Diabetes and Gender

2. Calculate the marginal distribution of Diabetes

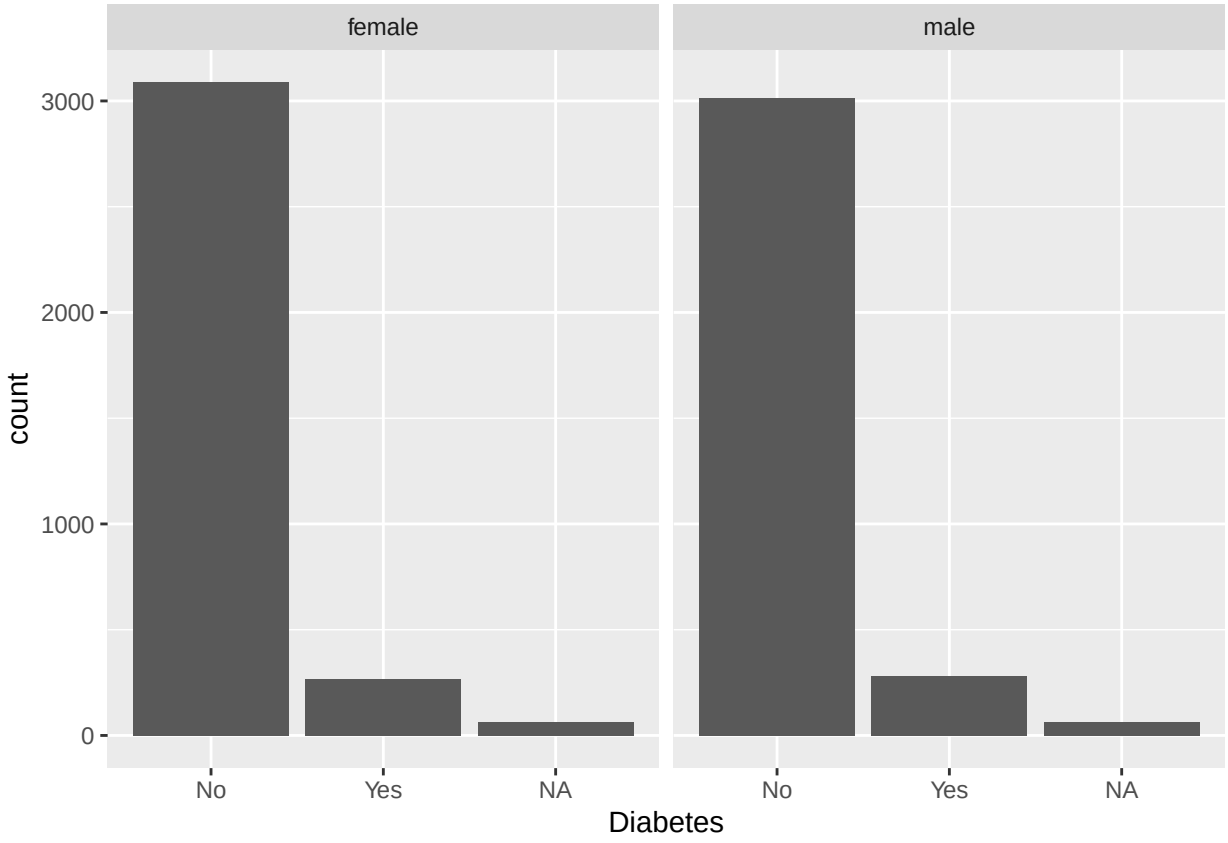
3. Calculate the conditional distribution of Diabetes given that the subject's Gender is male

4. Calculate the conditional distribution of Diabetes given that the subject's Gender is female

Problem 3

Looking at the following bar plots, would you expect that a person's Diabetes status is independent of their Gender? Explain your answer using the definition of independence.

```
sp <- ggplot(NHANES, aes(x=Diabetes)) + geom_bar()
sp + facet_wrap( ~ Gender, nrow=1) +
  labs(y = 'count')
```



Problem 4

Draw a simple diagram to illustrate how a histogram is created.