Stat 408 **Name:** _____
In class midterm
3/1/2018

Choose 8 of the questions to answer. Place an X on the number of the question you don't want graded, if you do not do this the first 8 that you answer will be graded.

1. (5 points) Describe a strategy for calculating the average number of points in football games MSU lost using the data frame below. Pseudocode is preferred, prose is acceptable - but be specific.

```
msu.football <- data.frame( msu.opponent = c('Washington State', 'South Dakota State','North Dakota',
'Weber State','Portland State'),
msu.points = c(0, 27, 49, 17, 30),
msu.outcome = c('Loss', 'Loss', 'Win', 'Loss', 'Win'))
```

2. (5 points) Suppose you are working on dataset that contains: daily snow fall totals in inches, day of the week as a three letter string, and the number of cars in the Bridger Bowl parking lot. You hope to estimate the average number of cars for four types of days: weekend with more than 2 inches of snow, weekday with more than 2 inches of snow, weekend with 2 inches of snow or less, and weekday with 2 inches of snow or less. Describe how you would do this and include important functions if you are not using pseudocode.

3. (5 points) Describe a strategy to merge to two data frames defined below and write the output you'd expect to see.

```
df1 <- data.frame(school = c('MSU','VT','Mines', 'Luther'), state= c('MT','VA','CO','IA'))
df2 <- data.frame(college = c('Mines','MSU','VT'), enrollment = c(5794,15688,30598))
```

4. (5 points) Describe at least two principles of good data visualization and include a sketch to demonstrate - you will not be graded on your artistic ability, but keep it neat.

5. (5 points) Describe a way or sketch out R code to find the mean of the cost vector below, note mean(cost) will give an error.

```
cost <- c('$1100','$700.21', '$310')
```

6. (5 points) Create the resultant plot based on code below.

```
num.sims <- 1000
dice <- rep(0, 1000)
for (i in 1:num.sims){
  dice[i] <- sum(sample(6, size = 5, replace = T))
}
hist(dice, main= 'Distribution for sum of 5 dice', xlab='Sum of 5 dice', xlim=c(5,30),
ylab='Frequency of Occurrence')
text(x=25, y=150, 'Most results are \n between 15 and 20')
```

7. (5 points) Write the output from the code below.

```
msu.football <- data.frame(msu.points = c(0, 27, 49, 17, 30),
                           msu.outcome = c('Loss', 'Loss', 'Win', 'Loss', 'Win'))
for (i in 1:nrow(msu.football)){
     msu.football[i,2]
}
```

8. (5 points) Write the output from the code below.

```
msu.football2 <- data.frame(msu.points = c(0, 27, 49, 17, 30),
msu.outcome = c('Loss', 'Loss', 'Win', 'Loss', 'Win'))
library(dplyr)
msu.football2 %>% group_by(msu.outcome) %>% summarize(MaxPoints = max(msu.points))
```

9. (5 points) Assume you write a function in R, what elements are necessary for documenting this function? Give an example.