Stat 408 **Name:** _____
In class midterm
10/18/2017

Short Answer Questions

Choose 5 of the short answer questions to answer. Circle the number of the questions you'd like graded, if you do not do this the first 5 that you answer will be graded. For the short answer questions please try to keep your answers to a maximum of 3-4 sentences.

1. (4 points) Describe a strategy for indexing elements in the data frame defined below.

```
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. (4 points) Assume you write a function in R, what elements are necessary for documenting this function?

3. (4 points) Define a vectorized operation in R and give an example code demonstrating a vectorized procedure.

4. (4 points) Describe a strategy to merge to two data frames defined below.

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

5. (4 points) Describe at least two principles of good data visualization.

6. (4 points) Describe the similarity and differences between matrices and data frames.

Coding and Graphical Questions:

1. (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')
```

2. (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'))
for (i in 1:nrow(msu.football2)){
     msu.football[i,2]
}
```

3. (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 %>% filter(msu.outcome == 'Loss') %>% summarize(MaxPoints = max(msu.points))
```

4. (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')
```