Midterm

Question 1

```
set.seed(2864)
# a
x <- sample(c(1:100),1000, replace=TRUE)

# b
sum((x %% 10) == 0)
```

```
## [1] 88
```

```
# c
y <- x[c(TRUE,FALSE)]
length(y)
```

```
## [1] 500
```

```
head(y,10)
```

```
##  [1]  92  64  63  74  75  90 100  25  65  89
```

```
# d
z <- x[x%%2 == 1]
head(z,10)
```

```
##  [1] 63 81 75 95 15 25 65 89  3 75
```

Question 2

```
(2 ^ c(1:50)) > (c(1:50) ^ 3)
```

```
##  [1]  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE
## [13]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [25]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [37]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [49]  TRUE  TRUE
```

Question 3

```
# a
dieRolls <- sample(1:6,3000,replace=TRUE)
head(dieRolls,20)
```

```
##  [1] 1 1 5 2 5 3 3 1 4 6 3 5 2 2 2 3 3 5 5 2
```

```
# b
breaks <- c(0,1,2,3,4,5,6)
```

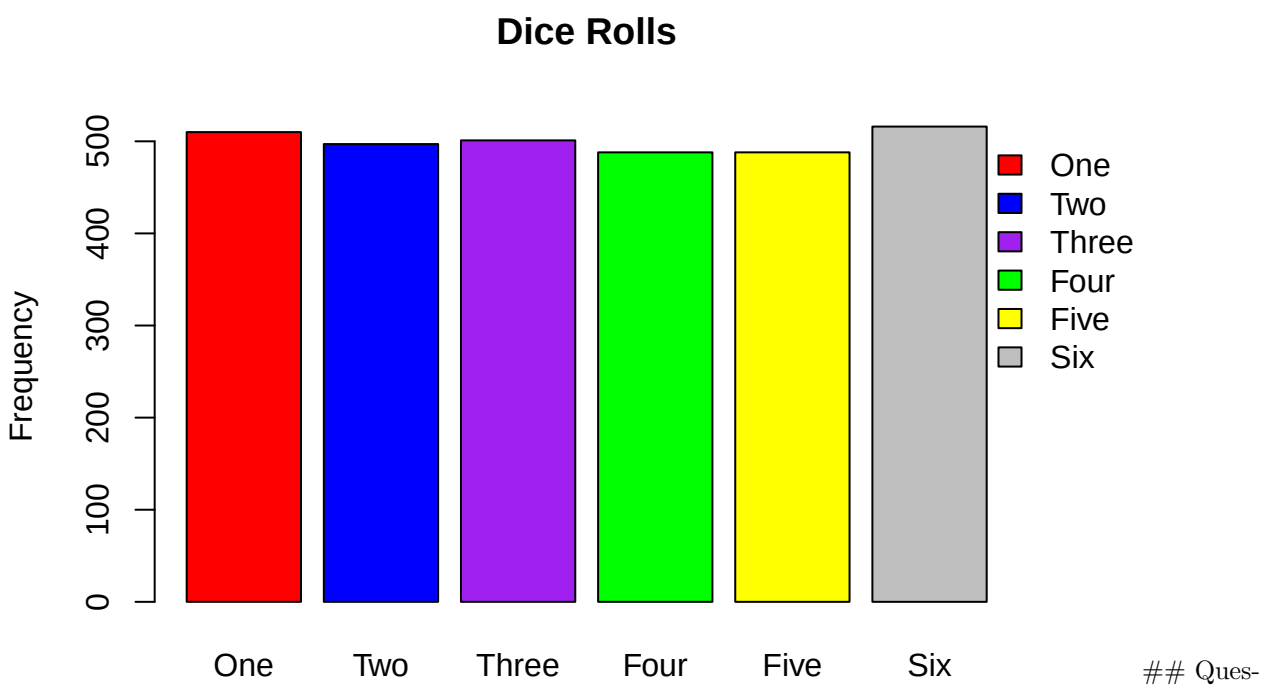
```r
tags <- c('One','Two','Three', 'Four', 'Five', 'Six')
dieRolls.f <- cut(dieRolls,breaks=breaks,labels=tags)

# c
sum(dieRolls <= 3)/length(dieRolls)
```

```
## [1] 0.5026667
```

```r
# yes this matches our expectation because we expect around half of the dice rolls
# to be 1,2 or 3 in a fair die

# d
par(mar=c(5,5,5,5))
barplot(table(dieRolls.f), main='Dice Rolls', ylab='Frequency', legend=TRUE,
        args.legend = list(x = "topright", bty = "n", inset=c(-0.15, 0)),
        col=c('red','blue','purple', 'green','yellow','grey'))
```



Question 4

```r
set.seed(2864)
x <- rnorm(1000,1,2)
# we know this has a skew of 0

numerator <- sum((x - mean(x))^3)/length(x)
denominator <- (sum((x - mean(x))^2)^(3/2))/length(x)

numerator/denominator
```

```
## [1] -0.000498571
```

Question 5

```
# a
names(mtcars)
```

```
##  [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"
## [11] "carb"
```

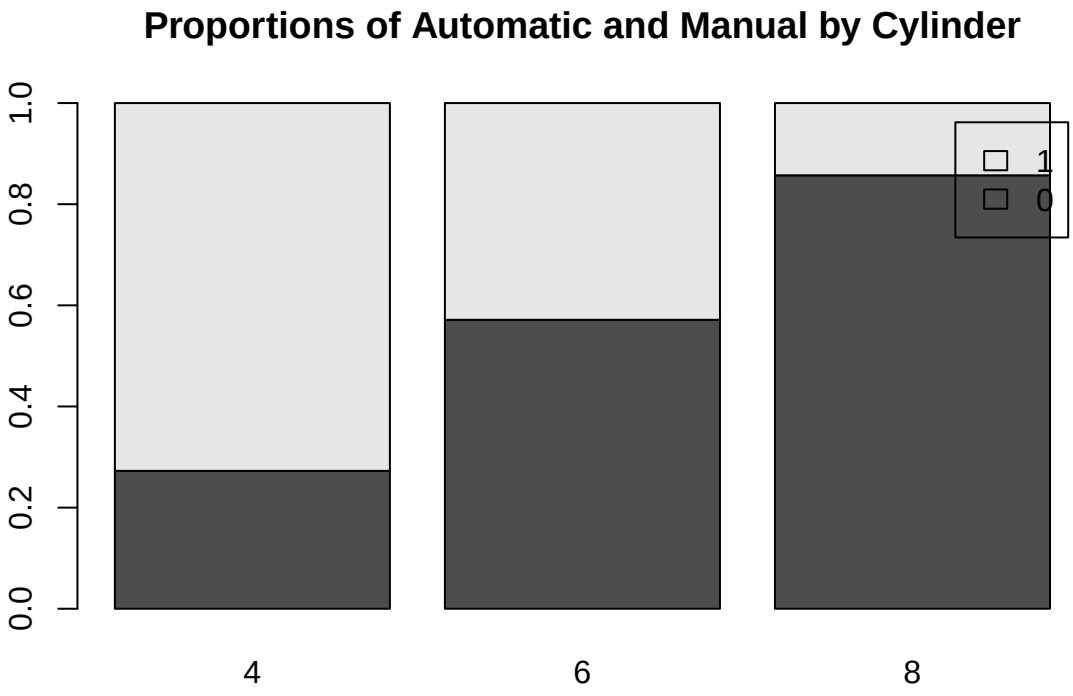
```
# b
df <- mtcars[mtcars$am==0 & mtcars$wt > mean(mtcars$wt),c('mpg', 'cyl', 'wt','am')]

# c
table(mtcars$cyl,mtcars$am)
```

```
##
##      0  1
##   4  3  8
##   6  4  3
##   8 12  2
```

```
# yes vehicles with more cylinders are more likely to have a manual transmission

# d
#barplot(table(mtcars$am, mtcars$cyl))
barplot(prop.table(table(mtcars$am,mtcars$cyl),margin=2), main='Proportions of Automatic and Manual by C
```




```
# yes the graph of the proportions confirms the answer in part c
```

Question 6

```
# 1
set.seed(2864)
```

```r
x <- rnorm(100,0,1)
y <- rnorm(100,0.5,1)

head(abs(x - y),10)
```

```
##  [1] 2.5581520 0.3587171 0.8697184 0.1936351 1.2192332 2.4886765 1.7598093
##  [8] 0.1268201 0.9070288 1.2022257
```

```r
head(sign(x - y),10)
```

```
##  [1] -1 -1 -1 -1 -1  1 -1 -1  1 -1
```

```r
# 2
rank(-abs(x - y))
```

```
##   [1]   7  73  56  83  43  11  20  86  54  44  42  63  96  23  64  36  95  46
##  [19]  15  17  99  77   1  69  27   4  61  59  47  76  72  81  88  85  25  70
##  [37]  14  38  82  62  98  28  91  68  80  74  52  40   2  87  13  89  90  94
##  [55]  21  50   6   5  93  78  34  97  67  18  12  35   9  32  30  60  19  92
##  [73]  49  57 100  53  26  71  45  29  22  75  79  65  16  55  24  58   8  66
##  [91]  31  10  33  51  37  39  48  41  84   3
```

```r
ri <- rank(-abs(x - y))

# 3
dfWilcox <- data.frame(xi=x,yi=y,Ri=rank(-abs(x - y)))

# 4
abs(sum(apply(dfWilcox, 1, function(x) sign(x['xi'] - x['yi']) * x['Ri'])))
```

```
## [1] 678
```

```r
abs(sum(sign(x-y)* ri))
```

```
## [1] 678
```