Homework 1
Ben Matern

1. **2.1** The diff function gives the difference between adjacent values. Since there are 8 values in miles, there are 7 differences between them.
   x = [313 284 311 280 322 324 302]
   max(x) = 324
   mean(x) = 305.1429
   min(x) = 280
   2.2
   max = 24
   mean=18.9
   min=15
   commute = replace(commute, commute==24,18)
   4 values >= 20
   sum(commute < 17) / length(commute) = .3
   2.3
   sum(bill) = 473
   min(bill) = 30
   length(bill[bill>40]) = 5
   length(bill[bill>40])/length(bill) = 4.167..

2. **2.5**
   1) (2,4,6,8,10)
   2) (4,6,10,14,22,26)
   3) 5, 6
   4) I think this doesn't work with different size arrays.
   5) 2, and 16
   6) 3
   7) 5
   8) undefined?
   9) uses values of x as indices of y. There are no indces 7 and 9 in y.
   10) (7 11 13)

3. **2.6**
   1) sum(x)/10 = 4.8
   2) log(x,10) = log10(x) = (0.0000000 0.9030900 0.3010300 0.7781513 0.4771213 0.9030900 0.6989700 0.6989700 0.6989700 0.6989700)
   3) (x-4.4)/2.875 = (-1.1826087 1.2521739 -0.8347826 0.5565217 -0.4869565 1.2521739 0.2086957 0.2086957 0.2086957 0.2086957)
   4) max(x) - min(x) = 7

4. c(rep(1,5), rep(2,4), rep(3,3), rep(4,2), 5)

5. mat <- matrix(c(1:13,NA,14:15),nrow=4)
   mat

```
badIndices = !is.na(mat[,4]) & mat[,4] < 14
badIndices
result = mat[-badIndices,]
result
```

6. **a)**
```
gadgets <- 1:600
thingies <- 301:900
objects  <- 401:1000
superobjects <- objects * 3
```
b)
```
myNewData <- data.frame(y=gadgets,x1=thingies,x2=objects,x3=superobjects)
str(myNewData)
write.table(myNewData,
file="/Users/bmatern/school/Fall2015/Stats4HumGenomics/Hwk1/benoutput.txt", row.names=F,
quote=F, sep="\t")
reloadedData <-
read.csv("/Users/bmatern/school/Fall2015/Stats4HumGenomics/Hwk1/benoutput.txt",sep="\t")
str(reloadedData)
```
c)
```
save(myNewData,
file="/Users/bmatern/school/Fall2015/Stats4HumGenomics/Hwk1/newdata.RData")
load("/Users/bmatern/school/Fall2015/Stats4HumGenomics/Hwk1/newdata.RData")
str(myNewData)
```

7. allele frequency = portion of a specific allele to all alleles
   heterozygous = an individual has two separate alleles for a gene
   genotype = the type of gene or genes an individual has
   haplotype = a set of polymorphisms that are usually inherited together.
   SNP marker = the physical location on a gene where a single nuc polymorphism is
   gene expression = an organism expresses a gene when it is translated into a protein, and presented
   in the organisms phenotype. gene expression refers to what genes are being used in the individual.

8. **a)**
```
mat <- matrix(runif(100000000, min=0, max=1),ncol=10)
beginTime <- proc.time()
for(i in 1:10)
{
  meanValue <- mean(mat[,i])
}
endTimeMean <- proc.time() - beginTime
```

Total Time = 1.1 seconds

b)
```
mat <- matrix(runif(100000000, min=0, max=1),ncol=10)
beginTime <- proc.time()
for(i in 1:10)
```

```
{
  means <- apply(mat,2,mean)
}
endTimeApply <- proc.time() - beginTime
```

Total Time = 19.6 seconds

9. **a)**
```
nucs<-c("A","G","C","T")
dna<-sample(nucs, 10000, replace=T)
```
b)
```
table(dna)
   A   C   G   T
2544 2490 2513 2453
```
c)

Code in script. Results:

"A A A" "A A G" "A A C" "A A T" "A G A" "A G G" ……..
"171"   "163"   "143"   "153"   "199"   "151"   ………..

10. `factorialFunction<-function(upperValue)`
```
{
  if(upperValue < 1)
  {
    return(0)
  }
  results <- 1
  for(x in 2:upperValue)
  {
    results <- results * x
  }
  return(results)
}
```

`factorialFunction(10)`

results = 3628800