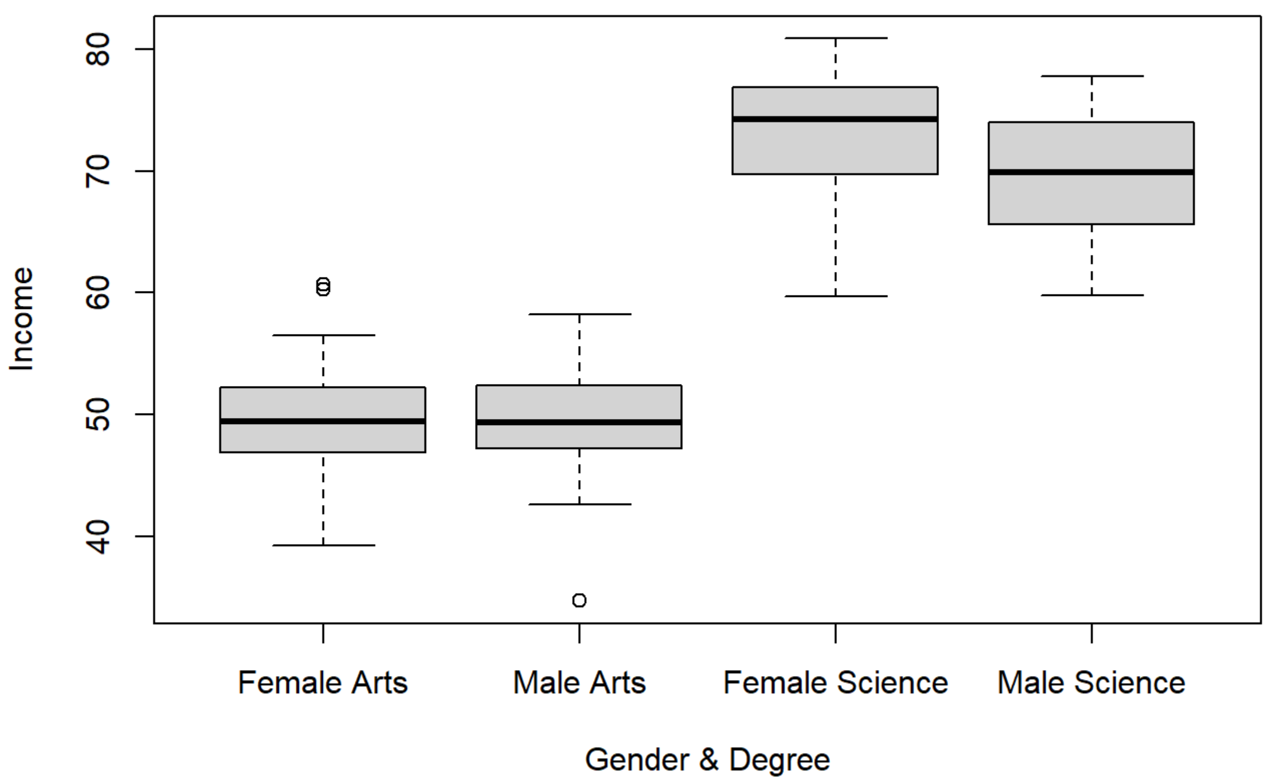
R Concepts Review

Rob Leonard (robleonard@tamu.edu (mailto:robleonard@tamu.edu))

1) Income by Gender and Degree

1a Create Box Plots

```
inc_deg_data <- read.csv("inc_deg_data.csv")
attach(inc_deg_data)
boxplot(income~interaction(gender,degree), data=inc_deg_data,
        names=c("Female Arts","Male Arts","Female Science","Male Science"),
        xlab="Gender & Degree", ylab="Income")
```



1-b Report the mean, median, standard deviation, and first and third quartiles of income.

```
summary(income)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   34.76   49.45   60.04   60.63   71.42   80.84
```

```
cat("Standard Deviation of Income:", sd(income))
```

```
## Standard Deviation of Income: 11.88252
```

1-c Repeat with income expressed in dollars (rather than 1,000's of dollars).

```
summary(income*1000)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   34758   49455   60042   60626   71422   80840
```

```
cat("Standard Deviation of Income:", sd(income*1000))
```

```
## Standard Deviation of Income: 11882.52
```

1-d Repeat excluding the minimum and maximum values

```
inc_deg_data <- inc_deg_data[order(income),]
summary(inc_deg_data$income[2:99]*1000)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   39318   49653   60042   60684   71256   79442
```

```
cat("Standard Deviation of Income excl. min/max:", sd(inc_deg_data$income[2:99]*1000))
```

```
## Standard Deviation of Income excl. min/max: 11532.19
```

```
detach(inc_deg_data)
```

2) Normal Distribution

2a Simulate normal distribution values (100x5) and report the column means

```
set.seed(101)
Mat2 <- matrix(rnorm(500),nrow=100)
Mat2means <- apply(Mat2,2,mean)
Mat2means
```

```
## [1] -0.037191100 -0.042002372  0.002070399 -0.025466513 -0.211317178
```

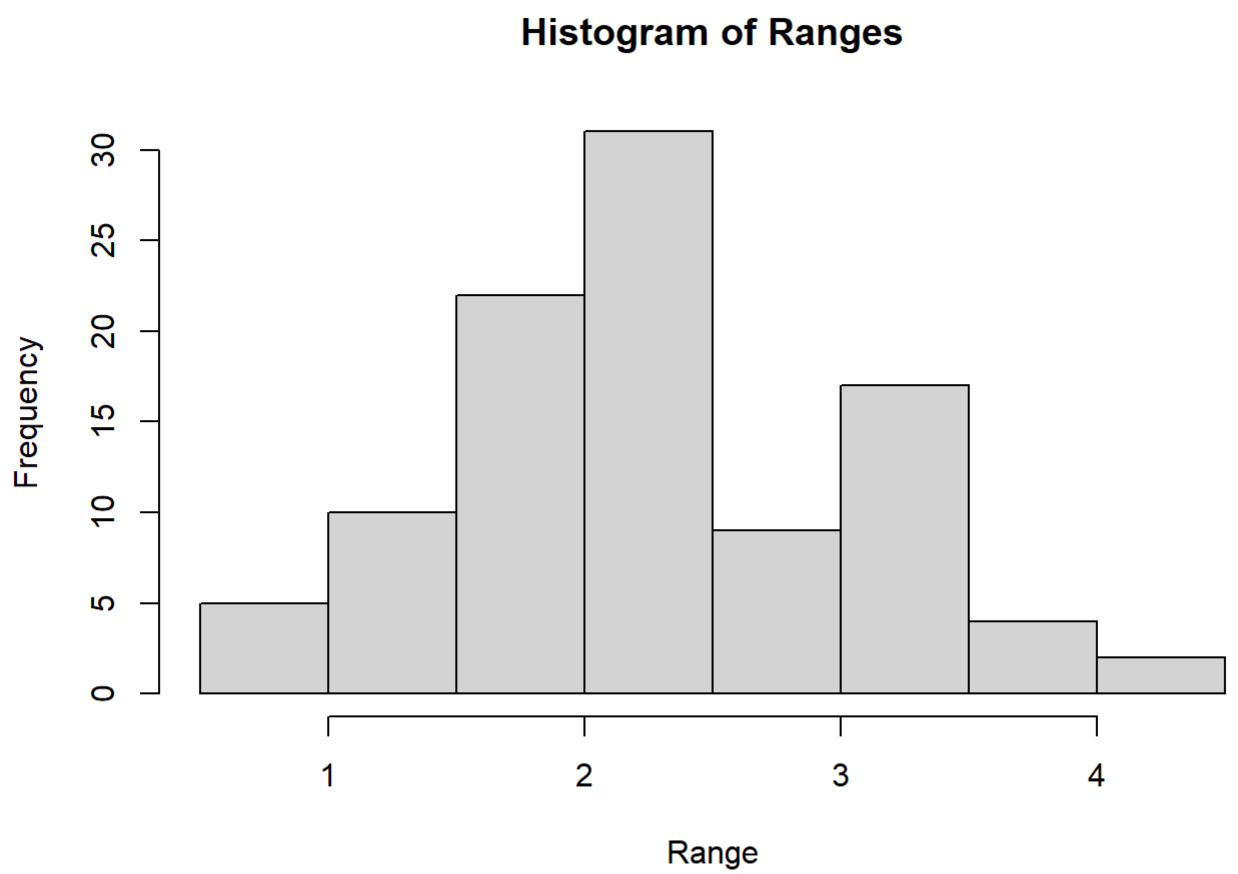
Redo using Matrix/Vector Arithmetic

```
vect1 <- rep(1,100)
Mat2ameans <- (1/100)*(vect1%*%Mat2)
Mat2ameans
```

```
##             [,1]        [,2]        [,3]        [,4]       [,5]
## [1,] -0.0371911 -0.04200237 0.002070399 -0.02546651 -0.2113172
```

2b Histogram of the row ranges

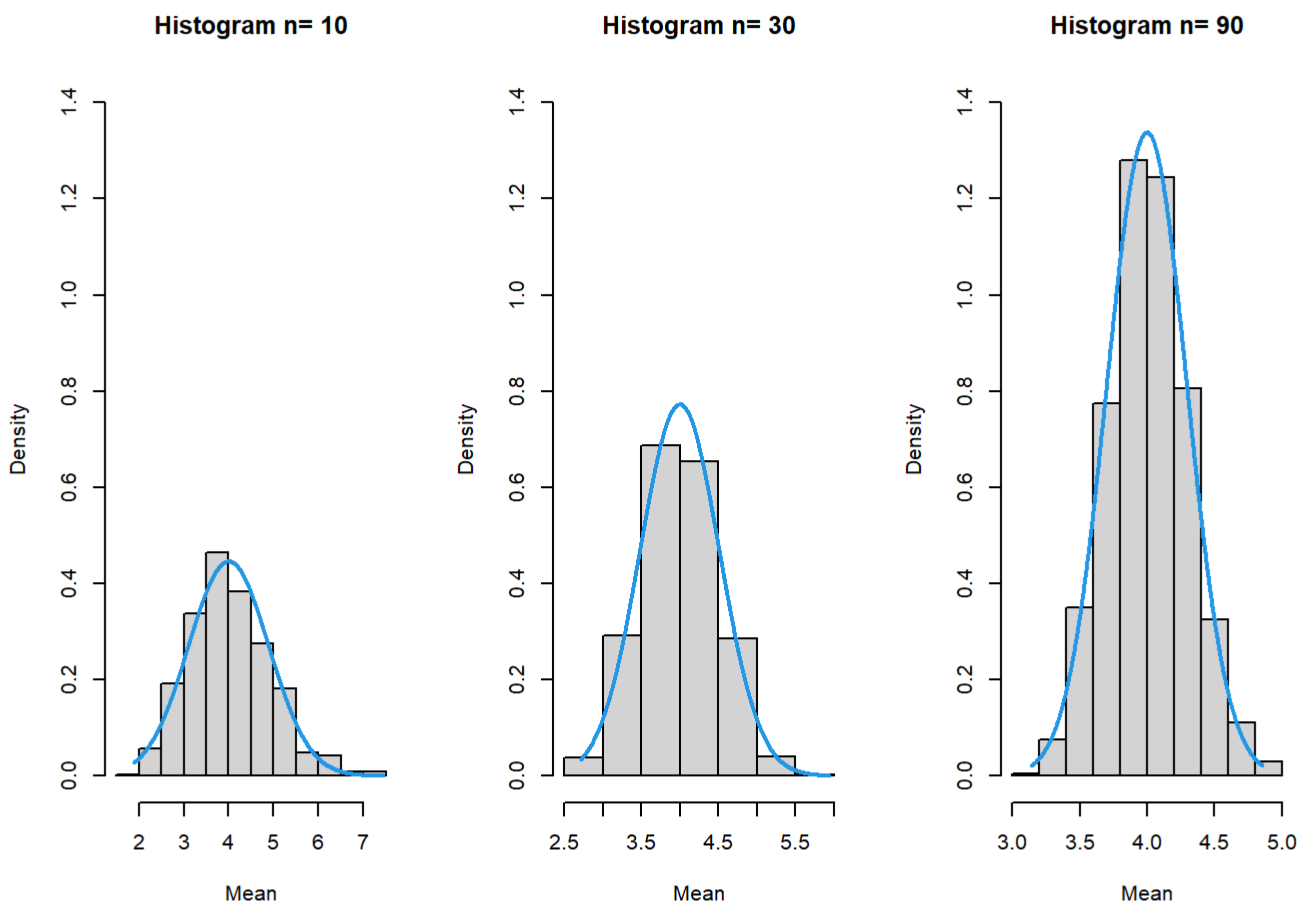
```
Mat2range <- apply(Mat2,1,max)-apply(Mat2,1,min)
hist(Mat2range, main="Histogram of Ranges", xlab="Range")
```



3) Gamma Distribution

Histogram of Gamma Distribution Simulation with Overlayed Normal Distribution

```
par(mfrow=c(1,3))
HistGam <- function(n,nreps=1000,gshape=2,gscale=2) {
  gmean=gshape*gscale
  gvar=gshape*gscale*gscale
  temp<-rep(0,nreps)
  for (i in 1:nreps) temp[i]=mean(rgamma(n,shape=gshape,scale=gscale))
  hist(temp,probability=TRUE,xlab="Mean",main=paste("Histogram n=",n),
       breaks ="Sturges",ylim=c(0,1.4) )
  xnorm <- seq(min(temp),max(temp),length=nreps+1)
  lines(xnorm,dnorm(xnorm,mean=gmean,sd=sqrt(gvar/n)),col=4,lwd=2)
  }
HistGam(n=10)
HistGam(n=30)
HistGam(n=90)
```



Matrix Algebra in R

4a Create matrices A and B and calculate column sums of A

```
set.seed(101)
A <- matrix(c(rnorm(12),runif(8,min=-2,max=2)),nrow=5,ncol=4,byrow=TRUE)
set.seed(102)
B <- matrix(rbeta(20,2,1),nrow=5,ncol=4,byrow=TRUE)
colSums(A)
```

```
## [1]  2.220841  3.323468 -1.551467 -1.851650
```

4b Calculate A+B and display certain elements

```
C<- A+B
C[4,2]
```

```
## [1] 1.368217
```

```
C[4,4]
```

```
## [1] 0.1262414
```

4c Matrix Multiplication and Transpose

```
D <- A%*%t(B)
D[4,2]
```

```
## [1] 0.8355238
```

```
D[4,4]
```

```
## [1] -0.3196251
```

4d Matrix inverse and determinants

```
E <- t(B)%*%A
solve(E)
```

```
##           [,1]     [,2]       [,3]     [,4]
## [1,] -5.099378 2.543412  1.1129536 2.016736
## [2,] -4.033431 3.715907 -0.7478673 1.999762
## [3,] -5.727737 4.837397  0.1813065 1.461192
## [4,] -9.330498 6.454731 -0.4181835 4.349940
```

```
det(E)
```

```
## [1] 0.15468
```