From the plots, I cannot see the clear separation between the three periods. So it is very difficult to distinguish the three periods in this dataset.
1b)

```r
> mvnorm.etest(skull[ skull$period == "1",-5],R=99999)

   Energy test of multivariate normality: estimated parameters
data:  x, sample size 30, dimension 4, replicates 99999
E-statistic = 0.985, p-value = 0.4279

> mvnorm.etest(skull[ skull$period == "2",-5],R=99999)

   Energy test of multivariate normality: estimated parameters
data:  x, sample size 30, dimension 4, replicates 99999
E-statistic = 0.9096, p-value = 0.7617

> mvnorm.etest(skull[ skull$period == "3",-5],R=99999)

   Energy test of multivariate normality: estimated parameters
data:  x, sample size 30, dimension 4, replicates 99999
E-statistic = 1.1716, p-value = 0.03888

#simulation-based approach to test the normality
qvals1 <- testnormality(skull[skull$period == "1",-5],5000)
qvals2 <- testnormality(skull[skull$period == "2",-5],5000)
qvals3 <- testnormality(skull[skull$period == "3",-5],5000)
The q-values for each period from simulation-based approach are 0.768429, 0.8134173, and 0.4627345. So we cannot reject the multivariate normality assumption for all the three groups.

1c)

```r
fit.lm <- lm(cbind(Breadth, basialveolar.height, basialveolar.length,nasal.height)~as.factor(period), data = skull)
fit.manova <- Manova(fit.lm)
summary(fit.manova)
```

Multivariate Tests: as.factor(period)  

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>test stat</th>
<th>approx F</th>
<th>num Df</th>
<th>den Df</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillai</td>
<td>2</td>
<td>0.1722118</td>
<td>2.002148</td>
<td>8</td>
<td>170</td>
<td>0.0489045 *</td>
</tr>
<tr>
<td>Wilks</td>
<td>2</td>
<td>0.8301027</td>
<td>2.049069</td>
<td>8</td>
<td>168</td>
<td>0.0435825 *</td>
</tr>
<tr>
<td>Hotelling-Lawley</td>
<td>2</td>
<td>0.2018820</td>
<td>2.094526</td>
<td>8</td>
<td>166</td>
<td>0.0389623 *</td>
</tr>
<tr>
<td>Roy</td>
<td>2</td>
<td>0.1869691</td>
<td>3.973094</td>
<td>4</td>
<td>85</td>
<td>0.0052784 **</td>
</tr>
</tbody>
</table>
Based on the results from MANOVA, I can conclude that there is significant period effect on at least one of the response variables.

1d)

#contrast

C1<- matrix(c(0, 0,1), ncol = 3, by = T)

fit_contrast1 <- linearHypothesis(model = fit.lm, hypothesis.matrix = C1)

Design matrix is c (0,0,1).

Multivariate Tests:  
  Df test  stat  approx F num Df den Df Pr(>F)  
Pillai   1 0.141918 3.463887       4  84 0.011401 *  
Wilks    1 0.858082 3.463887       4  84 0.011401 *  
Hotelling-Lawley 1 0.1649470 3.463887       4  84 0.011401 *  
Roy      1 0.1649470 3.463887       4  84 0.011401 *  

So there is significant difference between the first and third period.

1e)

C2<- matrix(c(0, 1, 0, 0, -1, 1), nrow = 2, by = T)

fit_contrast2 <- linearHypothesis(model = fit.lm, hypothesis.matrix = C2)

Design matrix is C = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & -1 \end{pmatrix}.

Multivariate Tests:  
  Df test  stat  approx F num Df den Df Pr(>F)  
Pillai   2 0.172118 2.002148       8  170 0.0489045 *  
Wilks    2 0.8201827 2.0948969       8  166 0.0435825 *  
Hotelling-Lawley 2 0.2018820 2.0948969       8  166 0.0389623 *  
Roy      2 0.1869691 3.973094       8  85 0.0052784 **  

So there is a significant difference between the first and second, and second and third periods.

1f)

diff1=mean(skull[31:60,-5])-mean(skull[1:30,-5])
diff2=mean(skull[61:90,-5])-mean(skull[31:60,-5])
t <- qt(0.05/16,87,lower.tail = FALSE)
halfwidth1 <- t*sqrt(diag(fit.manova$SSPE)/87*(1/30+1/30))
c(diff1-halfwidth1,diff1+halfwidth1)
Shown below are 95% Bonferroni confidence intervals for Period 1 – Period 2, and Period 2 – Period 3, for each of the measurements. Since all the intervals contain 0, there is no significant change in any period.

**Period 1 – Period 2**

<table>
<thead>
<tr>
<th></th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>breadth</td>
<td>-4.28</td>
<td>2.28</td>
</tr>
<tr>
<td>bheight</td>
<td>-2.50</td>
<td>4.30</td>
</tr>
<tr>
<td>length</td>
<td>-3.50</td>
<td>3.70</td>
</tr>
<tr>
<td>nheight</td>
<td>-1.95</td>
<td>2.55</td>
</tr>
</tbody>
</table>

**Period 2 – Period 3**

<table>
<thead>
<tr>
<th></th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>breadth</td>
<td>-5.378</td>
<td>1.18</td>
</tr>
<tr>
<td>bheight</td>
<td>-4.503</td>
<td>2.30</td>
</tr>
<tr>
<td>length</td>
<td>-0.566</td>
<td>6.63</td>
</tr>
<tr>
<td>nheight</td>
<td>-2.582</td>
<td>1.92</td>
</tr>
</tbody>
</table>

#check the residuals
```
par(mfrow=c(2,2))
plot(fit.lm$fitted.values[,1],fit.lm$residuals[,1],xlab="fitted",ylab="residuals")
plot(fit.lm$fitted.values[,2],fit.lm$residuals[,2],xlab="fitted",ylab="residuals")
plot(fit.lm$fitted.values[,3],fit.lm$residuals[,3],xlab="fitted",ylab="residuals")
plot(fit.lm$fitted.values[,4],fit.lm$residuals[,4],xlab="fitted",ylab="residuals")
```
From the residual above I find that the residuals are random distributed around zero, thus the covariance matrices assumption is not violated. And from part b, the multivariate normality assumption is also satisfied. So the usual MANOVA assumptions are valid for this data.

3a)

```r

image <- images[images[, 1] == "WINDOW" | images[, 1] == "BRICKFACE" | images[, 1] == "SKY",]

#pooling the variance

image.centered <- image[, -1] - rbind(matrix(apply(image[image[,1] == "WINDOW", -1], 2, mean), nrow = 300, ncol = ncol(image[, -1]), by = T),
                                       matrix(apply(image[image[,1] == "BRICKFACE", -1], 2, mean), nrow = 300, ncol = ncol(image[, -1]), by = T),
                                       matrix(apply(image[image[,1] == "SKY", -1], 2, mean), nrow = 300, ncol = ncol(image[, -1]), by = T))

image.pc <- prcomp(image.centered[, -3], scale=T)

summary(image.pc)
```
I pooled the variances for each of these variables, and shown below are the results of principle components.

**Importance of components:**

<table>
<thead>
<tr>
<th>PC</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
<th>PC6</th>
<th>PC7</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.0664</td>
<td>1.5838</td>
<td>1.10156</td>
<td>1.00703</td>
<td>0.95927</td>
<td>0.86793</td>
<td>0.8314</td>
</tr>
<tr>
<td>0.5224</td>
<td>0.1394</td>
<td>0.06741</td>
<td>0.05634</td>
<td>0.05112</td>
<td>0.04185</td>
<td>0.0384</td>
</tr>
<tr>
<td>0.5224</td>
<td>0.6517</td>
<td>0.72916</td>
<td>0.78550</td>
<td>0.83662</td>
<td>0.87847</td>
<td>0.9169</td>
</tr>
<tr>
<td>0.7213</td>
<td>0.62732</td>
<td>0.48684</td>
<td>0.40419</td>
<td>0.3472</td>
<td>0.24933</td>
<td>0.002833</td>
</tr>
<tr>
<td>0.0289</td>
<td>0.02136</td>
<td>0.01317</td>
<td>0.00908</td>
<td>0.0067</td>
<td>0.00343</td>
<td>0.00000</td>
</tr>
<tr>
<td>0.9450</td>
<td>0.96763</td>
<td>0.98060</td>
<td>0.99998</td>
<td>0.9966</td>
<td>1.00000</td>
<td>1.00000</td>
</tr>
<tr>
<td>4.462e-08</td>
<td>3.399e-08</td>
<td>2.954e-08</td>
<td>2.896e-08</td>
<td>0.000e+00</td>
<td>0.000e+00</td>
<td>0.000e+00</td>
</tr>
<tr>
<td>1.000e+00</td>
<td>1.000e+00</td>
<td>1.000e+00</td>
<td>1.000e+00</td>
<td>1.000e+00</td>
<td>1.000e+00</td>
<td></td>
</tr>
</tbody>
</table>

3b)

```r
plot(cumsum(image.pc$sdev^2)/sum(image.pc$sdev^2),type='b',ylab="cumulative probability")

#formally check how many principle components

for (i in 1:18)
cat(i, PCs.proportion.variation.enuff(image.pc$sdev^2, q = i, 900), "\n")
```

From these probabilities, I think first six components already did a good job since they described around 90% of the variation. Then I run hypothesis tests on whether the first q eigenvalues contain all the variation. The result below shows that the first nine components are adequate enough for representing the variation in the dataset. The screeplot also convinces it.

```
1 4.305412e-84
2 1.688789e-40
3 2.676816e-26
4 4.897679e-17
5 1.488039e-10
6 1.48585e-06
7 0.0007146283
8 0.01889352
9 0.1077151
```
The first PC vector is a contrast between REGION.CENTROID.ROW, EXRED.MEAN, EXGREEN.MEAN, SATURATION.MEAN, HUE.MEAN and INTENSITY.MEAN, RAWRED.MEAN, RAWBLUE.MEAN, RAWGREEN.MEAN, EXBLUE.MEAN, VALUE.MEAN.

The second component is primarily related to SHORT.LINE.DENSITY.2, VEDGE.MEAN, VEDGE.SD, HEDGE.MEAN, HEDGE.SD. And the third one is primarily a contrast between REGION.CENTROID.COL, SHORT.LINE.DENSITY.2, VEDGE.MEAN, VEDGE.SD and SHORT.LINE.DENSITY.5, HEDGE.MEAN, HEDGE.SD.
3d)
source("http://www.public.iastate.edu/~maitra/stat501/Rcode/mmnorm.R")
source("http://www.public.iastate.edu/~maitra/stat501/Rcode/circledraw.R")
source("http://www.public.iastate.edu/~maitra/stat501/Rcode/radviz2d.R")
source("http://www.public.iastate.edu/~maitra/stat501/Rcode/parallelplot.R")
source("http://www.public.iastate.edu/~maitra/stat501/Rcode/combinations.R")
newdata<-as.matrix(image.pc$x) + rbind(matrix(apply(image[1,"WINDOW"],-c(1,4)],2,mean),nrow = 300,ncol = ncol(image[,-c(1,4)],by = T),
   matrix(apply(image[1,"BRICKFACE"],-c(1,4)],2,mean),nrow = 300,ncol = ncol(image[,-c(1,4)],by = T),
   matrix(apply(image[1,"SKY"],-c(1,4)],2,mean),nrow = 300,ncol = ncol(image[,-c(1,4)],by = T)) %*% image.pc$rotation
newdf <- cbind(as.data.frame(newdata),image[,1])
N=ncol(newdf)
colnames(newdf)[N] <- "type"
newdf$ytype <- factor(newdf$ytype, levels=unique(newdf$ytype))
library(Andrews)
Andrews(newdf,clr = 19)
parallelplot(cbind(newdata[,1:9],as.factor(newdf$ytype)),name="image data")
radviz2d(newdf, name = "image data")
pairs((newdata[,1:9],
   panel=function(x,y)panel.smooth(x,y, col = as.numeric(image[,1]) + 2,
     pch = 20, cex = 2)
     abline(lsfit(x,y),lty=2) )}
From these plots, I think these three types can be separated well.
3e)

```
fit.lm <- lm(newdata[,1:9]~as.factor(image[,1]))
M1<- matrix(c(0, 1, 0), nrow = 1, by = T)
fit_contrast1 <- linearHypothesis(model = fit.lm, hypothesis.matrix = M1)
M2<- matrix(c(0, 0, 1), nrow = 1, by = T)
fit_contrast2 <- linearHypothesis(model = fit.lm, hypothesis.matrix = M2)
M3<- matrix(c(0, 1, -1), nrow = 1, by = T)
fit_contrast3 <- linearHypothesis(model = fit.lm, hypothesis.matrix = M3)
```

I ran a one-way MANOVA on the first nine PC vectors, with REGION.TYPE as the group variable. From the results below, I can conclude that there is a significant difference among the three region types.

```
Multivariate Tests: as.factor(image[, 1])
                          Df test stat approx F num Df den Df Pr(>F)
Pillai                  2  1.115676  376.802     6 1792 < 2.22e-16 ***
Wilks                   2  0.034231  1311.137     6 1790 < 2.22e-16 ***
Hotelling-Lawley       23  23.833965  3551.261     6 1788 < 2.22e-16 ***
Roy                    23  23.648788  7063.105     3  896 < 2.22e-16 ***
```

Then I perform the simultaneous pairwise tests for the three image types to see which regions are different. The design vectors are c1 = (0 1 0), c2 = (0 0 1), and c3 = (0 1 -1). All the three p-values are less than 0.001. I compared the p-values to 0.05/3 = 0.0167, using a Bonferroni multiple tests correction. So all three pairs are significantly different from each other.
$H_0: \sum = \sigma^2 I$

\[
\lambda = \frac{\max \mathcal{L}(\lambda, \sigma^2 I)}{\max \mathcal{L}(\lambda_0, I)}
\]

\[
\max \mathcal{L}(\lambda_0, I) = \frac{1}{(2\pi)^{n/2}|\lambda_0|^{1/2}} e^{-\frac{1}{2}}
\]

\[
\Lambda = \frac{1}{n} \sum (x_i - \bar{x}) (x_i - \bar{x})', \quad \bar{x} = \frac{1}{n} \sum x_i
\]

\[
\mathcal{L}(\lambda, \sigma^2 I) = \left(\frac{1}{2\sigma^2}\right)^{n/2} e^{-\frac{n}{2\sigma^2} \lambda} e^{-\frac{1}{2\sigma^2} \sum (x_i - \bar{x}) (x_i - \bar{x})}
\]

\[
\delta^2 = \frac{1}{n} \sum (x_i - \bar{x})(x_i - \bar{x})
\]

\[
\max \mathcal{L}(\lambda_0, \sigma^2 I) = \mathcal{L}(\lambda, \sigma^2 I) \propto \frac{1}{\tau^v(\xi^2)^{\frac{n}{2}}}
\]

\[
\frac{\alpha}{\lambda} \propto \frac{\text{R}^\lambda}{\frac{1}{\lambda} \text{R}^\lambda}
\]

\[
\Lambda^{\text{R}\lambda} \propto \frac{\text{R}^\lambda}{\frac{1}{\lambda} \text{R}^\lambda}
\]