# R Module 6 Rubric

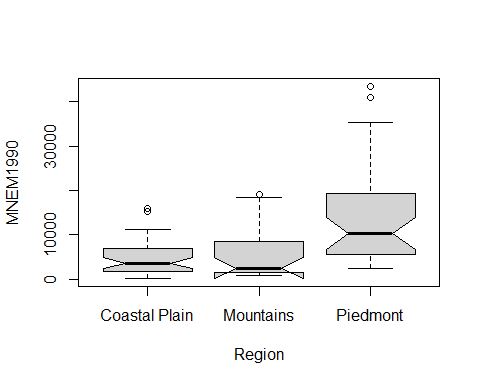
## Question 1: Boxplot

Notched boxplot of your regions

Base R:

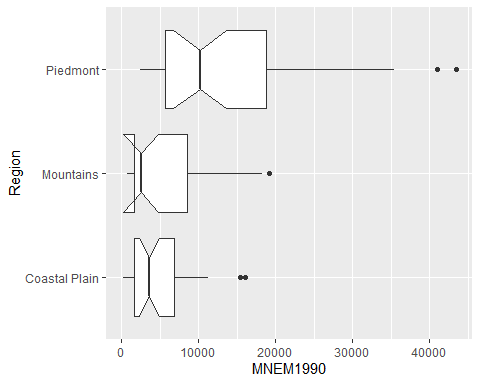
library(sf)  
NC <- read\_sf("data/NC\_REGION.shp")  
  
boxplot(MNEM1990 ~ Region, data = NC, notch = TRUE)

## Warning in (function (z, notch = FALSE, width = NULL, varwidth = FALSE, : some notches went outside hinges ('box'): maybe set notch=FALSE



ggplot2:

library(ggplot2)  
  
ggplot(NC, aes(x = MNEM1990, y = Region)) +   
 geom\_boxplot(notch = TRUE)



## Question 2: ANOVA

Report your F-Statistic and if you should reject or fail to reject the null hypothesis

anova <- aov(MNEM1990 ~ Region, data = NC)  
  
print(anova)

## Call:  
## aov(formula = MNEM1990 ~ Region, data = NC)  
##   
## Terms:  
## Region Residuals  
## Sum of Squares 2127131896 6155559536  
## Deg. of Freedom 2 97  
##   
## Residual standard error: 7966.139  
## Estimated effects may be unbalanced

summary(anova)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Region 2 2.127e+09 1.064e+09 16.76 5.6e-07 \*\*\*  
## Residuals 97 6.156e+09 6.346e+07   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Question 3: oneway.test, Equal Variance

oneway.test(MNEM1990 ~ Region, data = NC, var.equal = TRUE)

##   
## One-way analysis of means  
##   
## data: MNEM1990 and Region  
## F = 16.76, num df = 2, denom df = 97, p-value = 5.6e-07

## Question 4: Bartlett K-Squared

bartlett.test(MNEM1990 ~ Region, data = NC)

##   
## Bartlett test of homogeneity of variances  
##   
## data: MNEM1990 by Region  
## Bartlett's K-squared = 38.195, df = 2, p-value = 5.083e-09

## Question 5: oneway.test, Unequal Variance

oneway.test(MNEM1990 ~ Region, data = NC, var.equal = FALSE)

##   
## One-way analysis of means (not assuming equal variances)  
##   
## data: MNEM1990 and Region  
## F = 11.66, num df = 2.000, denom df = 47.748, p-value = 7.526e-05

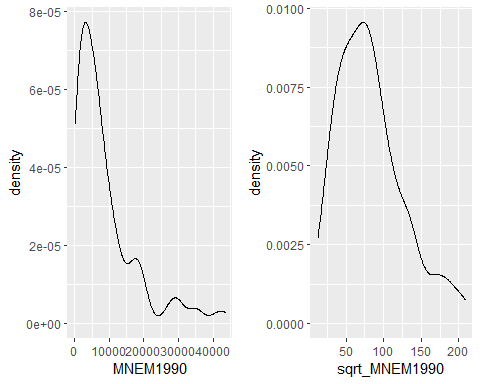
## Question 6: Shapiro-Wilk Normality Test

# Student's don't need to use dplyr or even transform the data, as long as the  
# distribution is normal.  
  
library(dplyr)  
NC <- NC %>%   
 mutate(  
 sqrt\_MNEM1990 = sqrt(MNEM1990)  
 )  
  
shapiro.test(NC$sqrt\_MNEM1990)

##   
## Shapiro-Wilk normality test  
##   
## data: NC$sqrt\_MNEM1990  
## W = 0.94203, p-value = 0.0002569

Illustrate data using density plot for both un-transformed and transformed data

unt <- ggplot(NC, aes(x = MNEM1990)) +   
 geom\_density()  
  
tra <- ggplot(NC, aes(x = sqrt\_MNEM1990)) +   
 geom\_density()  
  
  
# This is just to get two plots side-by-side  
cowplot::plot\_grid(unt, tra)



## Question 7: Tukey Test

aov <- aov(sqrt\_MNEM1990 ~ Region, data = NC)  
  
tukey <- TukeyHSD(aov)  
tukey

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = sqrt\_MNEM1990 ~ Region, data = NC)  
##   
## $Region  
## diff lwr upr p adj  
## Mountains-Coastal Plain 2.87440 -20.29441 26.04321 0.9530922  
## Piedmont-Coastal Plain 49.85867 29.54578 70.17157 0.0000002  
## Piedmont-Mountains 46.98427 23.24429 70.72426 0.0000244

plot(tukey)

