Otolith_Analysis

Sara Michele Schaal

7/3/2018

```
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 3.4.3
```

```
otolith.data <- read.csv("/Users/saraschaal/Documents/Northeastern/LotterhosLab/Research/Data/CodEcotyp
```

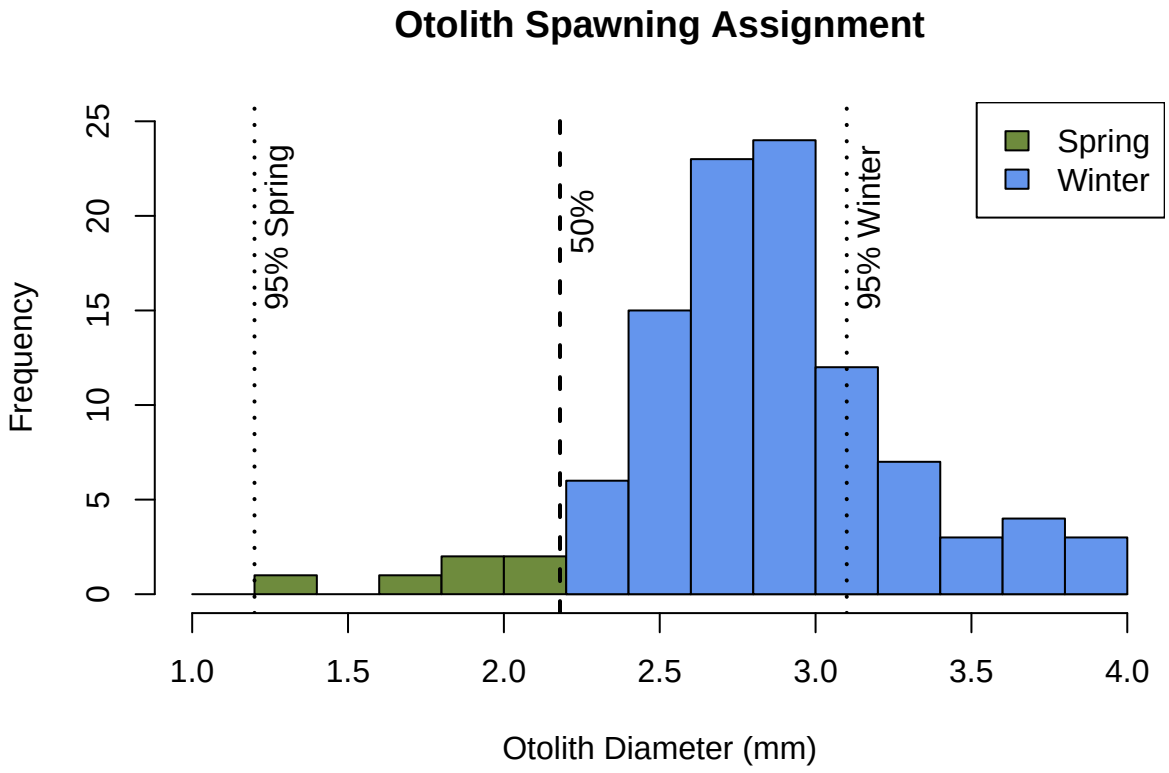
```
head(otolith.data)
```

```
##    SampleID Age_SideA Age_SideB AgeConf Diameter_SideA Diameter_SideB
## 1 16_215_Gm         3         3       5          2.631          2.513
## 2 16_216_Gm         3         3       5          2.479          2.497
## 3 16_217_Gm         3         3       5          2.576          2.580
## 4 16_218_Gm         5         5       5          2.363          2.610
## 5 16_219_Gm         2         2       3          3.073          3.160
## 6 16_220_Gm         3         3       5          3.855          3.670
##   DiaConf BetterSide                                        Notes
## 1       5          B
## 2       5       Both
## 3       5       Both
## 4       5          B
## 5       5       Both     last year band close to edge of otolith
## 6       5          B looks like side B hits closer to the center
```

```
# Calculate final diameter based on two measurements
otolith.data$final.diameter <- NA
for(i in 1:nrow(otolith.data)){
  if(otolith.data$BetterSide[i] == "Both"){
   otolith.data$final.diameter[i] <- mean(c(otolith.data$Diameter_SideA[i], otolith.data$Diameter_SideB
  } else if(otolith.data$BetterSide[i] == "A"){
    otolith.data$final.diameter[i] <- otolith.data$Diameter_SideA[i]
  } else if(otolith.data$BetterSide[i] == "B"){
    otolith.data$final.diameter[i] <- otolith.data$Diameter_SideB[i]
  }
}

# Assign spawning season
otolith.data$season <- NA
for(i in 1:nrow(otolith.data)){
  if(!is.na(otolith.data$final.diameter[i]) & otolith.data$final.diameter[i] > 2.18){
    otolith.data$season[i] <- "Winter"
  } else if(!is.na(otolith.data$final.diameter[i]) & otolith.data$final.diameter[i] <= 2.18){
    otolith.data$season[i] <- "Spring"
  } else if(is.na(otolith.data$final.diameter[i])){
    otolith.data$season[i] <- NA
  }
}
```

Plotting

```r
# Create Histogram with Otolith Diameter and Assignment
  hist(otolith.data$final.diameter[otolith.data$season == "Spring"], col = "darkolivegreen4",
      xlim = c(1, 4), ylim = c(0, 25), breaks = seq(1, 4, 0.2),
      main = "Otolith Spawning Assignment",
      xlab = "Otolith Diameter (mm)")
  hist(otolith.data$final.diameter[otolith.data$season == "Winter"], col = "cornflowerblue",
      add = TRUE)
  legend("topright", legend = c("Spring", "Winter"),
         fill = c("darkolivegreen4", "cornflowerblue"))

# add thresholds for confidence based on Micah's analysis
  abline(v = 2.18, lty = 2, lwd = 2)
  abline(v = 1.2, lty = 3, lwd = 2)
  abline(v = 3.1, lty = 3, lwd = 2)

  text(2.2, 18, pos = 4, labels = "50%", srt = 90)
  text(1.22, 15, pos = 4, labels = "95% Spring", srt = 90)
  text(3.12, 15, pos = 4, labels = "95% Winter", srt = 90)
```

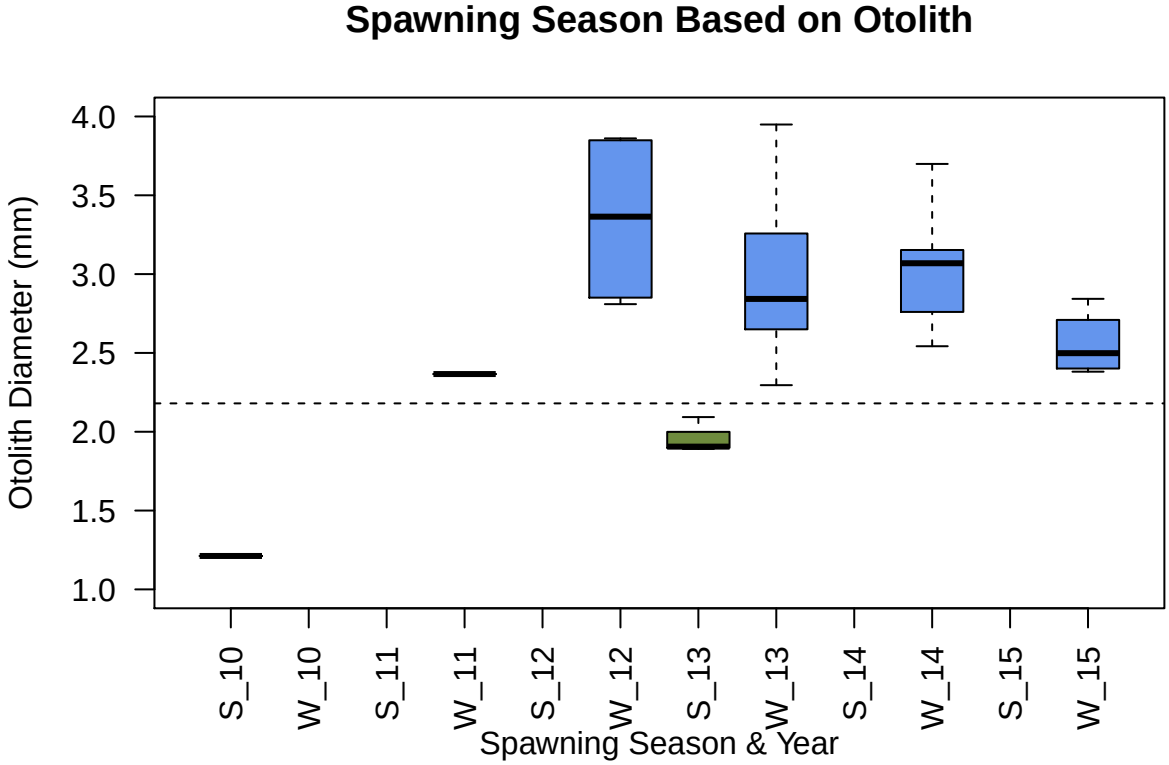



```r
# Save image as pdf
  #pdf(file = "/Users/saraschaal/Documents/Northeastern/LotterhosLab/Research/Data/CodEcotypes/Atlantic
  #write.csv(otolith.data, "/Users/saraschaal/Documents/Northeastern/LotterhosLab/Research/Data/CodEcot
```

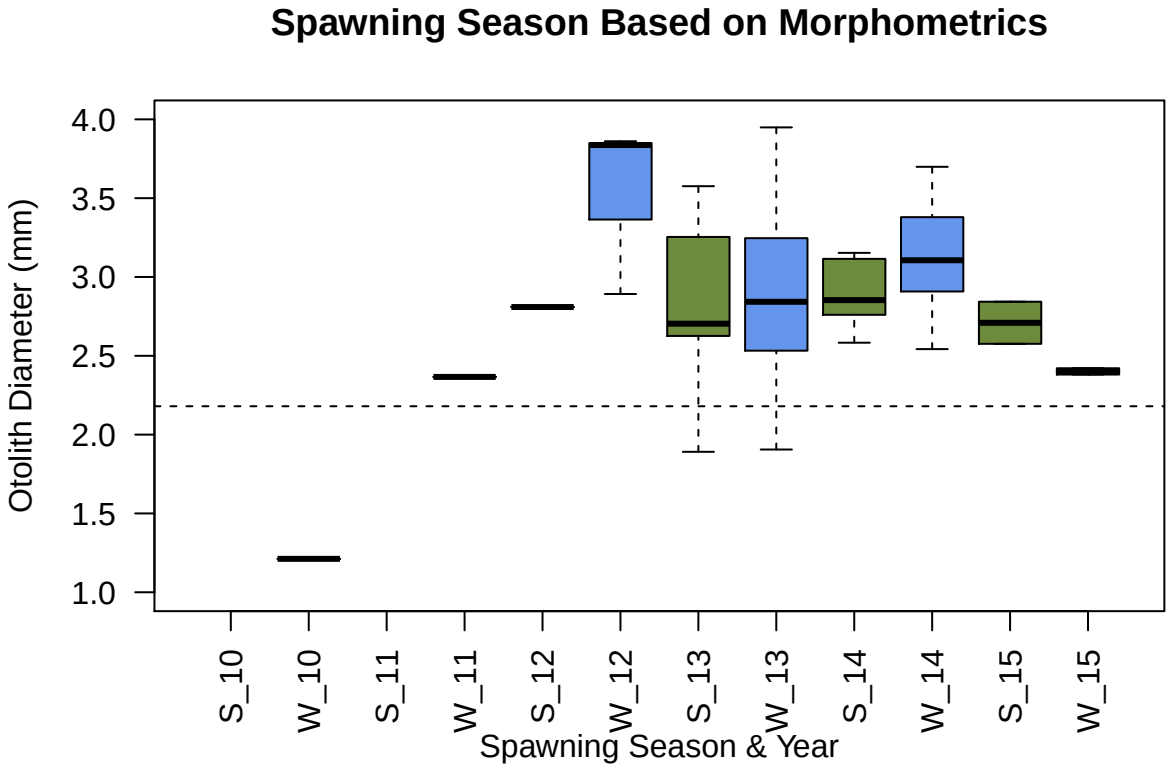
```r
  masterData <- read.csv("/Users/saraschaal/Documents/Northeastern/LotterhosLab/Research/Data/CodEcotype
  colnames(masterData)[1:3] <- c("SampleID", "CollectionDate", "CollectionSite")
  masterOto <- merge(masterData, otolith.data, by = "SampleID", all = TRUE)

  # Add year and month data columns and convert date to actual date object
  masterOto$Year <- as.numeric(substr(masterOto$CollectionDate, 1, 4))
```

```
masterOto$Month <- as.numeric(substr(masterOto$CollectionDate, 5, 6))
masterOto$CollectionDate <- as.Date(as.character(masterOto$CollectionDate), format = ("%Y%m%d"))
masterOto$Year[116] <- 2016
masterOto$Year[117] <- 2017

# combine otolith data and master data
dataAssignSpawn <- read.csv("/Users/saraschaal/Documents/Northeastern/LotterhosLab/Research/Data/CodE
```

```
# Calculate Birth Year
masterOto$birthYear <- as.numeric(masterOto$Year) - masterOto$Age_SideA
masterOtolith <- merge(dataAssignSpawn[,20:23], masterOto, by = "SampleID")

# Plot Otolith Diameter by Spawning Season & Year
#par(mfrow = c(2, 1))
boxplot(final.diameter ~ season + birthYear, data = masterOtolith,
        col = c("darkolivegreen4", "cornflowerblue"), las = 2,
        names = c("S_10", "W_10","S_11", "W_11","S_12", "W_12","S_13",
                  "W_13","S_14", "W_14","S_15", "W_15"),
        ylab = "Otolith Diameter (mm)", xlab = "Spawning Season & Year",
        main = "Spawning Season Based on Otolith", ylim = c(1, 4))
abline(h = 2.18, lty = 2)
```




```
boxplot(final.diameter ~ Spawning + birthYear, data = masterOtolith,
        col = c("darkolivegreen4", "cornflowerblue"), las = 2,
        names = c("S_10", "W_10","S_11", "W_11","S_12", "W_12","S_13",
                  "W_13","S_14", "W_14","S_15", "W_15"),
        ylab = "Otolith Diameter (mm)", xlab = "Spawning Season & Year",
        main = "Spawning Season Based on Morphometrics", ylim = c(1,4))
abline(h = 2.18, lty = 2)
```


Spawning Season Based on Morphometrics
Otolith Diameter (mm)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
S_10
W_10
S_11
W_11
S_12
W_12
S_13
W_13
S_14
W_14
S_15
W_15
Spawning Season & Year

Random Forest - Micah's Data

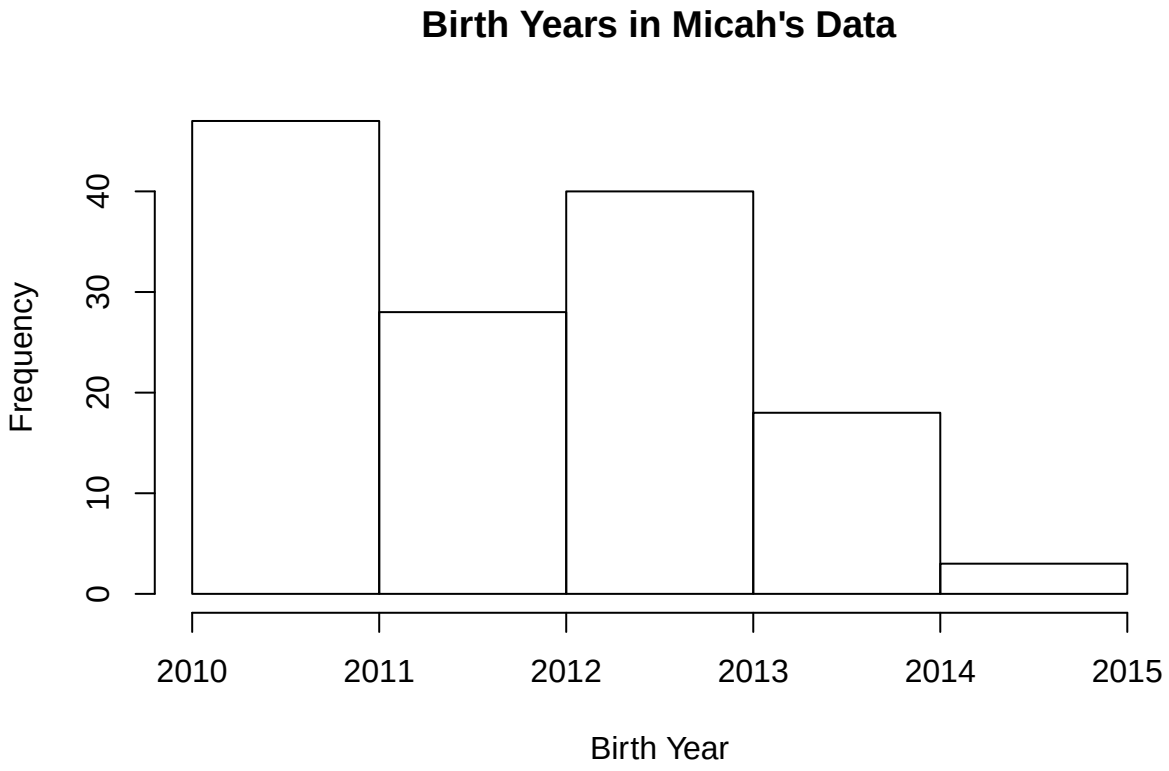
```
library(randomForest)
```

```
## Warning: package 'randomForest' was built under R version 3.4.4
```

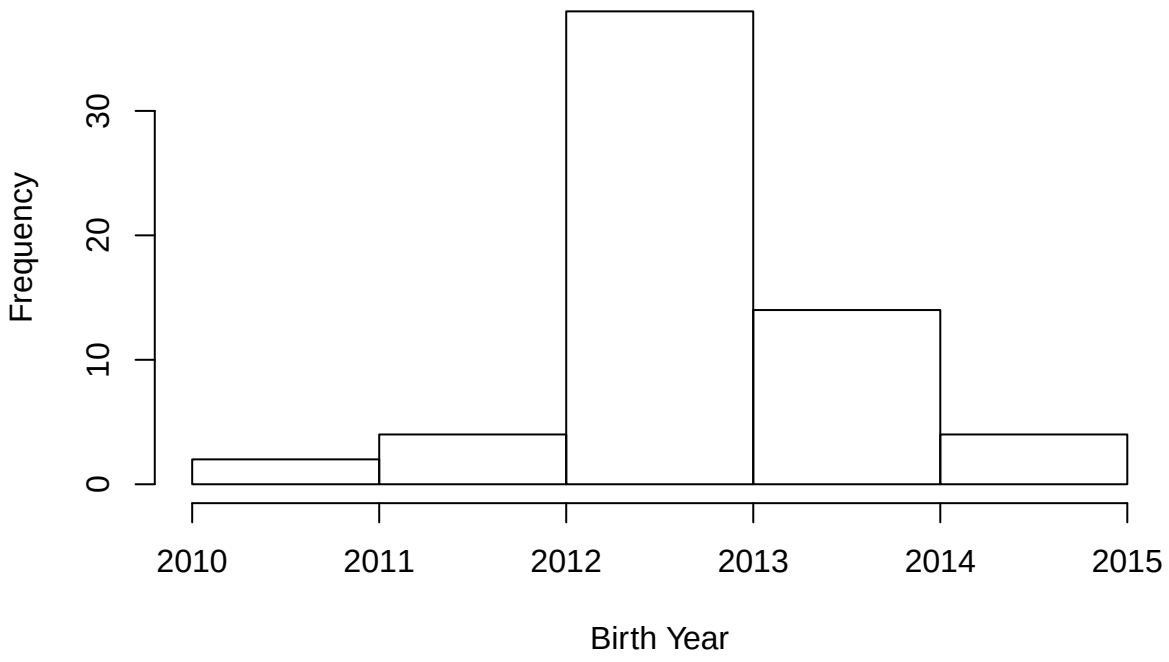
```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
# Input data
MicahData <- read.csv("/Users/saraschaal/Documents/Northeastern/LotterhosLab/Research/Data/CodEcotypes
MicahData$Year <- as.numeric(substr(as.character(MicahData$TOWDATE), 7, 8))+2000
MicahData$Month <- as.numeric(substr(as.character(MicahData$TOWDATE), 1, 2))
MicahData$birthYear <- MicahData$Year - MicahData$AGE
relevantYears <- subset(MicahData, subset = birthYear > 2009)

hist(relevantYears$birthYear, breaks = 6, main = "Birth Years in Micah's Data", xlab = "Birth Year")
```




```
hist(masterOtolith$birthYear, breaks = 6, main = "Birth Years in Sara's Data", xlab = "Birth Year")
```




```
  #Subset data into training and validation datasets to test how well it works on known data
  train <- sample(nrow(relevantYears), 0.7*nrow(relevantYears), replace = FALSE)
  TrainData <- relevantYears[train,]
  ValidData <- relevantYears[-train,]

  #Create RFmodel
  RFModel <- randomForest(SEASON ~ birthYear + A1, data = TrainData, importance = TRUE)
  RFModel
```

```
##
## Call:
##  randomForest(formula = SEASON ~ birthYear + A1, data = TrainData,      importance = TRUE)
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 1
##
##         OOB estimate of  error rate: 20%
## Confusion matrix:
##        SPRING WINTER class.error
## SPRING     21     14  0.40000000
## WINTER      5     55  0.08333333
```

```
  #Run RFmodel on the training dataset
  predTrain <- predict(RFModel, TrainData, type = "class")
  table(predTrain, TrainData$SEASON)
```

```
##
## predTrain SPRING WINTER
##    SPRING     26      0
##    WINTER      9     60
```

```r
  #Run RFmodel on the validation dataset
  predValid <- predict(RFModel, ValidData, type = "class")
  table(predValid, ValidData$SEASON)
```

```
##
## predValid SPRING WINTER
##    SPRING     10      0
##    WINTER      8     23
```

```r
  # Use all data for final analysis
  RFModelAllData <- randomForest(SEASON ~ birthYear + A1, data = relevantYears, importance = TRUE)
  RFModelAllData
```

```
##
## Call:
##  randomForest(formula = SEASON ~ birthYear + A1, data = relevantYears,      importance = TRUE)
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 1
##
##         OOB estimate of  error rate: 19.85%
## Confusion matrix:
##        SPRING WINTER class.error
## SPRING     32     21  0.39622642
## WINTER      6     77  0.07228916
```

```r
  #Run on my unknown data
  RFUnknownData <- masterOtolith[!is.na(masterOtolith$season),]

  colnames(RFUnknownData)[49] <- "A1"
  RFUnknownPred <- predict(RFModelAllData, RFUnknownData)
  RFUnknownPred
```

```
##      1      2      3      4      5      6      7      8      9     10
## WINTER WINTER WINTER WINTER WINTER WINTER WINTER WINTER WINTER WINTER
##     11     12     13     14     15     17     18     19     20     21
## WINTER WINTER WINTER WINTER WINTER WINTER WINTER WINTER WINTER WINTER
##     22     23     24     25     26     27     28     29     30     31
## SPRING WINTER WINTER WINTER WINTER WINTER WINTER WINTER WINTER WINTER
##     32     33     34     35     36     37     38     39     40     41
## WINTER WINTER WINTER WINTER WINTER WINTER WINTER WINTER WINTER WINTER
##     42     43     44     45     46     47     48     50     52     54
## WINTER WINTER WINTER WINTER WINTER WINTER WINTER WINTER WINTER WINTER
##     58     59     60     61     62     63     64     66     67     68
## WINTER WINTER WINTER WINTER WINTER WINTER WINTER WINTER WINTER WINTER
##     69     70
## WINTER WINTER
## Levels: SPRING WINTER
```

```r
  predictedData <- as.data.frame(cbind(RFUnknownData,RFUnknownPred))

  length(which(predictedData$RFUnknownPred == "WINTER"))
```

```
## [1] 61
```

```r
  length(which(predictedData$RFUnknownPred == "SPRING"))
```

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

```
  predictedData[which(predictedData$RFUnknownPred == "SPRING"),]
```

```
##     SampleID     Spring    Winter Spawning CollectionDate
## 22 16_241_Gm 0.09529418 0.9047058   Winter     2016-08-11
##       CollectionSite   D1   D2   D3   D4   D5   D6   D7   D8   D9 D10  D11
## 22 Stellwagon_Bank_B 3.09 7.94 2.96 5.74 5.14 4.39 5.64 7.72 6.72 4.1 7.52
##     D12 D13 D14  D15  D16  D17  D18  D19  D20 D21  D22 TotalLength
## 22 5.41 7.7 6.4 6.86 4.26 4.57 1.41 4.02 5.91 4.6 2.83        26.8
##    WeightTotal WeightGonad WeightLiver Sex LeftGillRaker RightGillRaker
## 22          NA          NA          NA                NA             NA
##    RedGreenRatio Age corrected Age_SideA Age_SideB AgeConf Diameter_SideA
## 22          1.12  NA                  6         6       5          1.212
##    Diameter_SideB DiaConf BetterSide Notes final.diameter season   A1
## 22          1.233       5          A                1.212 Spring 2016
##    Month birthYear RFUnknownPred
## 22     8      2010        SPRING
```

Subset data by years

```
palette(c("darkolivegreen4", "cornflowerblue"))

####################
#### 2011 Data ####
####################
  par(mfrow = c(1,2))

 # Subset data
  SaraYear11 <- subset(RFUnknownData, subset = RFUnknownData$birthYear == 2011)
  MicahYear11 <- subset(MicahData, subset = MicahData$birthYear == 2011)

 # Create a model using Random Forest on known dataset
  RF11 <- randomForest(SEASON~A1, data = MicahYear11)

 # Validate model on known data
  Prob11Micah <- predict(RF11, MicahYear11, type = "prob")
  Assign11Micah <- predict(RF11, MicahYear11)
  colnames(Prob11Micah) <- c("RFSpring", "RFWinter")

 # Use model to predict unknown samples
  Prob11Pred <- predict(RF11, SaraYear11, type = "prob")
  colnames(Prob11Pred) <- c("RFSpring", "RFWinter")
  Assign11Pred <- predict(RF11, SaraYear11)

 # Run LDA
  LDA2011 <- lda(SEASON ~ A1, data = MicahYear11)
  LDA2011MicahPred <- predict(LDA2011)
  LDA2011pred <- predict(LDA2011, newdata = SaraYear11)

 # Bind all data together
  df.2011Micah <- cbind(MicahYear11, Prob11Micah, Assign11Micah, LDA2011MicahPred$x,
                        LDA2011MicahPred$posterior, LDA2011MicahPred$class)
  colnames(df.2011Micah)[c(21, 23:25)] <- c("RFAssign", "LDASpring", "LDAWinter", "LDAAssign")
  df.2011Sara <- cbind(SaraYear11, Prob11Pred, Assign11Pred,
                       LDA2011pred$x, LDA2011pred$posterior, LDA2011pred$class)
```

```
 colnames(df.2011Sara)[c(54, 56:58)] <- c("RFAssign", "LDASpring", "LDAWinter", "LDAAssign")

# Check accuracy of the models
 RF2011correct <- NULL
 RF2011incorrect <- NULL
 for(i in 1:nrow(df.2011Micah)){
   if(df.2011Micah$SEASON[i] == df.2011Micah$RFAssign[i]){
     RF2011correct <- c(RF2011correct, as.character(df.2011Micah$SEASON[i]))
   } else {
     RF2011incorrect <- c(RF2011incorrect, as.character(df.2011Micah$SEASON[i]))
   }
 }
 100-(length(RF2011incorrect)/length(RF2011correct)*100)
```

```
## [1] 100
```

```
 LDA2011correct <- NULL
 LDA2011incorrect <- NULL
 for(i in 1:nrow(df.2011Micah)){
   if(df.2011Micah$SEASON[i] == df.2011Micah$LDAAssign[i]){
     LDA2011correct <- c(LDA2011correct, as.character(df.2011Micah$SEASON[i]))
   } else {
     LDA2011incorrect <- c(LDA2011incorrect, as.character(df.2011Micah$SEASON[i]))
   }
 }
 100-(length(LDA2011incorrect)/length(LDA2011correct)*100)
```

```
## [1] 81.48148
```

```
# Plot data
 par(mfrow = c(1,3))
 nrow(MicahYear11[MicahYear11$SEASON == "SPRING",])
```

```
## [1] 15
```

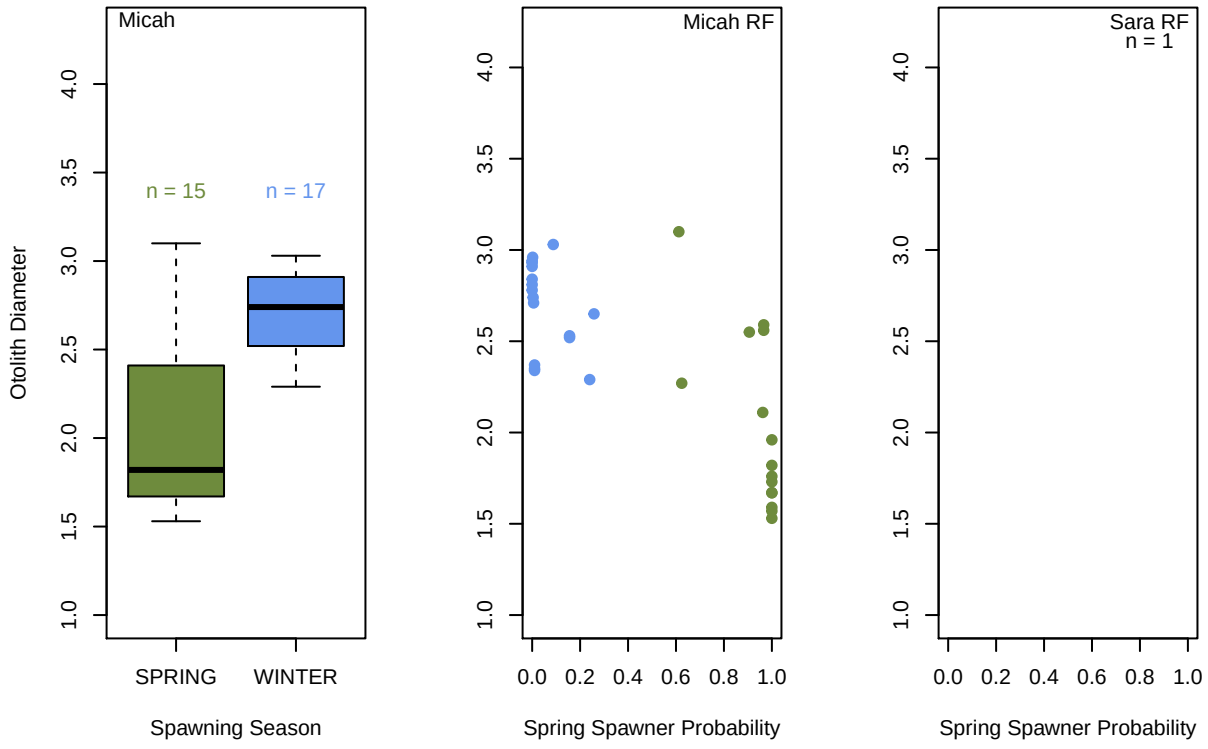
```
 nrow(MicahYear11[MicahYear11$SEASON == "WINTER",])
```

```
## [1] 17
```

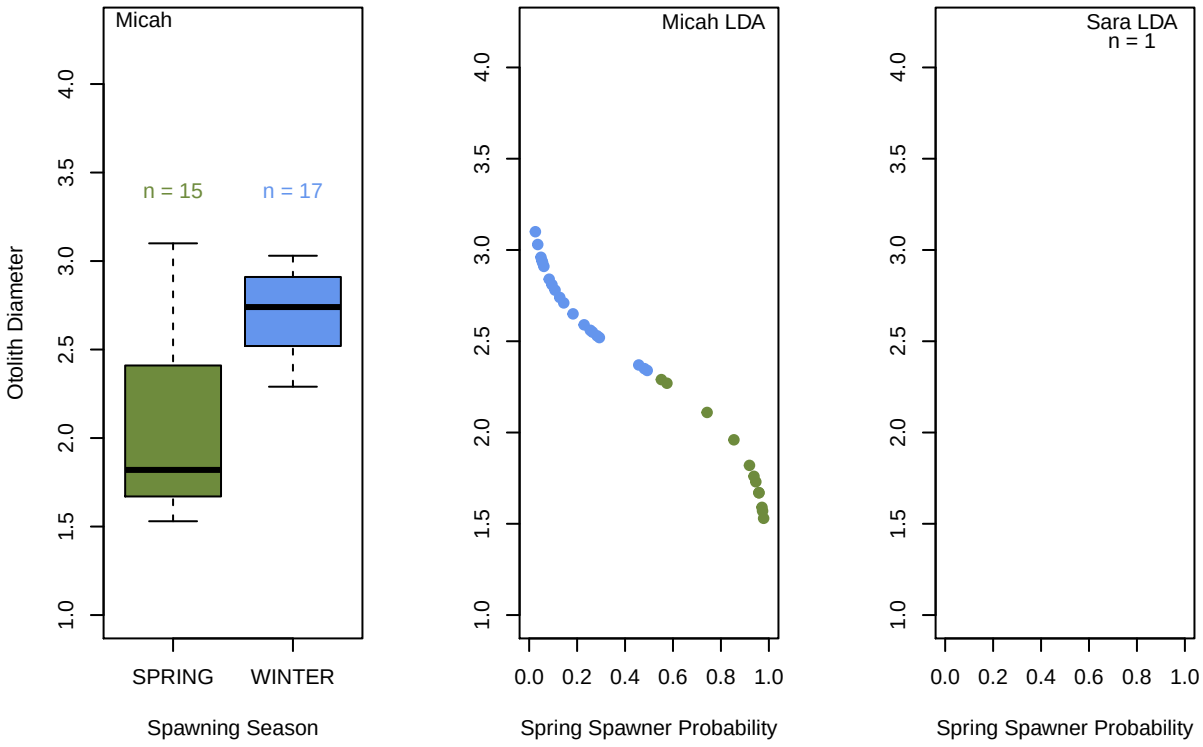
```
 nrow(SaraYear11)
```

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

```
 boxplot(MicahYear11$A1~MicahYear11$SEASON, ylab = "Otolith Diameter", ylim = c(1.0, 4.3),
         col = c("darkolivegreen4", "cornflowerblue"), xlab = "Spawning Season")
 text(1, 3.4, "n = 15", col = "darkolivegreen4")
 text(2, 3.4, "n = 17", col = "cornflowerblue")
 text(1.1, 4.35, pos = 2, "Micah")
 plot(df.2011Micah$A1~df.2011Micah$RFSpring, pch = 19, xlim = c(0, 1), ylim = c(1.0, 4.2),
      xlab = "Spring Spawner Probability", ylab = "",
      col = df.2011Micah$RFAssign)
 text(0.82, 4.25, "Micah RF")
 plot(df.2011Sara$A1~df.2011Sara$RFSpring, pch = 19, xlim = c(0, 1), ylim = c(1.0, 4.2),
      xlab = "Spring Spawner Probability", ylab = "",
      col = df.2011Sara$RFAssign)
 text(0.84, 4.25, "Sara RF")
 text(0.84, 4.15, "n = 1")
```




```
boxplot(MicahYear11$A1~MicahYear11$SEASON, ylab = "Otolith Diameter", ylim = c(1.0, 4.3),
        col = c("darkolivegreen4", "cornflowerblue"), xlab = "Spawning Season")
text(1, 3.4, "n = 15", col = "darkolivegreen4")
text(2, 3.4, "n = 17", col = "cornflowerblue")
text(1.1, 4.35, lwd = 3, pos = 2, "Micah")
plot(df.2011Micah$A1~df.2011Micah$LDASpring, pch = 19, xlim = c(0, 1), ylim = c(1.0, 4.2),
     xlab = "Spring Spawner Probability", ylab = "",
     col = df.2011Micah$LDAAssign)
text(0.77, 4.25, lwd = 3, "Micah LDA")
plot(df.2011Sara$A1~df.2011Sara$LDASpring, pch = 19, xlim = c(0,1), ylim = c(1.0, 4.2),
     xlab = "Spring Spawner Probability", ylab = "",
     col = df.2011Sara$LDAAssign)
text(0.78, 4.25, lwd = 3, "Sara LDA")
text(0.78, 4.15, lwd = 3, "n = 1")
```




```
###################
#### 2012 Data ####
###################

 # Subset Data
  SaraYear12 <- subset(RFUnknownData, subset = RFUnknownData$birthYear == 2012)
  MicahYear12 <- subset(MicahData, subset = MicahData$birthYear == 2012)

 # Create a model using Random Forest on known dataset
  RF12 <- randomForest(SEASON~A1, data = MicahYear12)

 # Validate model on known data
  Prob12Micah <- predict(RF12, MicahYear12, type = "prob")
  Assign12Micah <- predict(RF12, MicahYear12)
  colnames(Prob12Micah) <- c("RFSpring", "RFWinter")

 # Use model to predict unknown samples
  Prob12Pred <- predict(RF12, SaraYear12, type = "prob")
  colnames(Prob12Pred) <- c("RFSpring", "RFWinter")
  Assign12Pred <- predict(RF12, SaraYear12)

 # Run LDA
  LDA2012 <- lda(SEASON ~ A1, data = MicahYear12)
  LDA2012MicahPred <- predict(LDA2012)
  LDA2012pred <- predict(LDA2012, newdata = SaraYear12)

 # Bind all data together
  df.2012Micah <- cbind(MicahYear12, Prob12Micah, Assign12Micah, LDA2012MicahPred$x,
                        LDA2012MicahPred$posterior, LDA2012MicahPred$class)
  colnames(df.2012Micah)[c(21, 23:25)] <- c("RFAssign", "LDASpring", "LDAWinter", "LDAAssign")
```

```
  df.2012Sara <- cbind(SaraYear12, Prob12Pred, Assign12Pred,
                       LDA2012pred$x, LDA2012pred$posterior, LDA2012pred$class)
  colnames(df.2012Sara)[c(54, 56:58)] <- c("RFAssign", "LDASpring", "LDAWinter", "LDAAssign")

 # Check accuracy of the models
  RF2012correct <- NULL
  RF2012incorrect <- NULL
  for(i in 1:nrow(df.2012Micah)){
    if(df.2012Micah$SEASON[i] == df.2012Micah$RFAssign[i]){
      RF2012correct <- c(RF2012correct, as.character(df.2012Micah$SEASON[i]))
    } else {
      RF2012incorrect <- c(RF2012incorrect, as.character(df.2012Micah$SEASON[i]))
    }
  }
  100-(length(RF2012incorrect)/length(RF2012correct)*100)
```

```
## [1] 100
```

```
  LDA2012correct <- NULL
  LDA2012incorrect <- NULL
  for(i in 1:nrow(df.2012Micah)){
    if(df.2012Micah$SEASON[i] == df.2012Micah$LDAAssign[i]){
      LDA2012correct <- c(LDA2012correct, as.character(df.2012Micah$SEASON[i]))
    } else {
      LDA2012incorrect <- c(LDA2012incorrect, as.character(df.2012Micah$SEASON[i]))
    }
  }
  100-(length(LDA2012incorrect)/length(LDA2012correct)*100)
```

```
## [1] 66.66667
```

```
 # Plot data
  par(mfrow = c(1,3))
  nrow(MicahYear12[MicahYear12$SEASON == "SPRING",])
```

```
## [1] 12
```

```
  nrow(MicahYear12[MicahYear12$SEASON == "WINTER",])
```

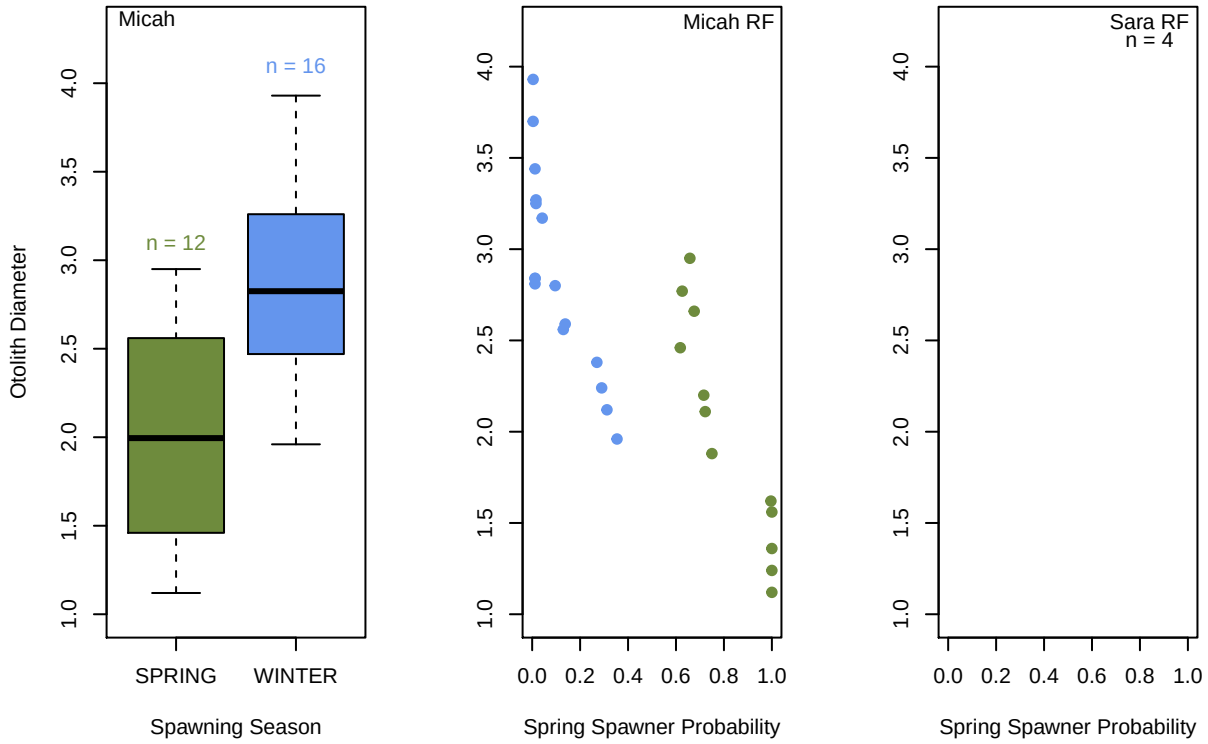
```
## [1] 16
```

```
  nrow(SaraYear12)
```

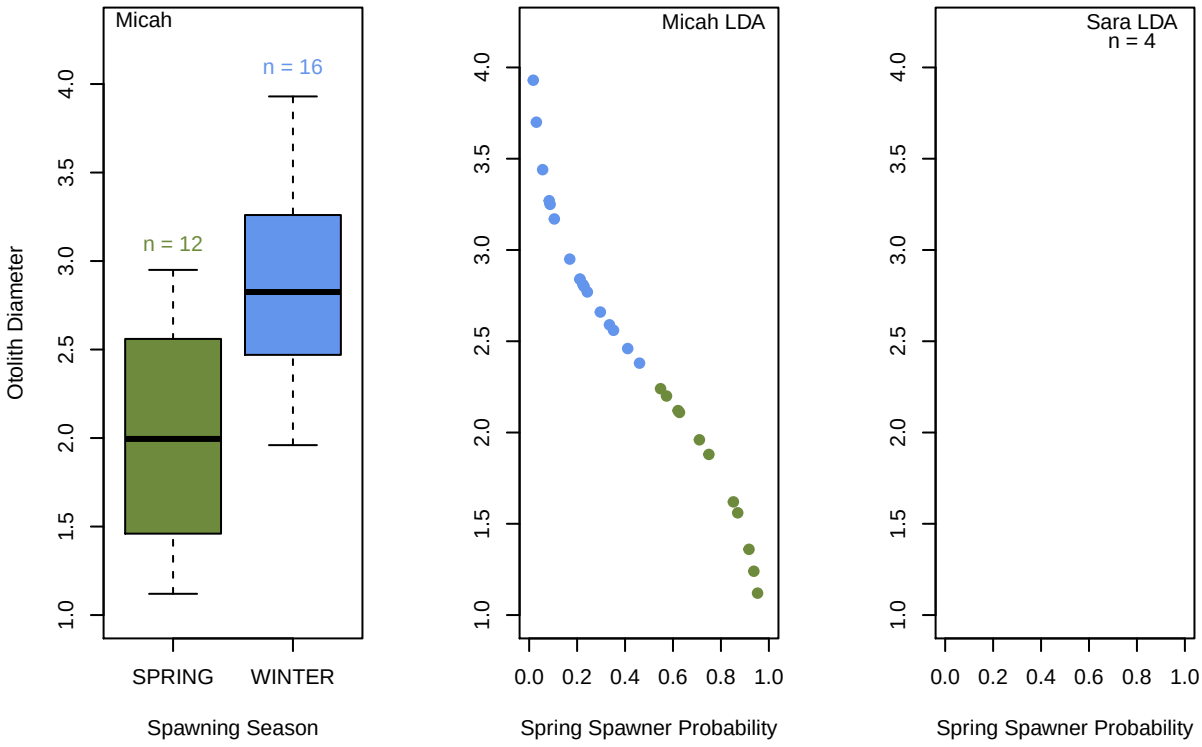
```
## [1] 4
```

```
  boxplot(MicahYear12$A1~MicahYear12$SEASON, ylab = "Otolith Diameter", ylim = c(1.0, 4.3),
          col = c("darkolivegreen4", "cornflowerblue"), xlab = "Spawning Season")
  text(1, 3.1, "n = 12", col = "darkolivegreen4")
  text(2, 4.1, "n = 16", col = "cornflowerblue")
  text(1.1, 4.35, pos = 2, "Micah")
  plot(df.2012Micah$A1~df.2012Micah$RFSpring, pch = 19, xlim = c(0, 1), ylim = c(1.0, 4.2),
       xlab = "Spring Spawner Probability", ylab = "",
       col = df.2012Micah$RFAssign)
  text(0.82, 4.25, "Micah RF")
  plot(df.2012Sara$A1~df.2012Sara$RFSpring, pch = 19, xlim = c(0, 1), ylim = c(1.0, 4.2),
       xlab = "Spring Spawner Probability", ylab = "",
       col = df.2012Sara$RFAssign)
```

```
text(0.84, 4.25, "Sara RF")
text(0.84, 4.15, "n = 4")
```




```
boxplot(MicahYear12$A1~MicahYear12$SEASON, ylab = "Otolith Diameter", ylim = c(1.0, 4.3),
        col = c("darkolivegreen4", "cornflowerblue"), xlab = "Spawning Season")
text(1, 3.1, "n = 12", col = "darkolivegreen4")
text(2, 4.1, "n = 16", col = "cornflowerblue")
text(1.1, 4.35, lwd = 3, pos = 2, "Micah")
plot(df.2012Micah$A1~df.2012Micah$LDASpring, pch = 19, xlim = c(0, 1), ylim = c(1.0, 4.2),
     xlab = "Spring Spawner Probability", ylab = "",
     col = df.2012Micah$LDAAssign)
text(0.77, 4.25, lwd = 3, "Micah LDA")
plot(df.2012Sara$A1~df.2012Sara$LDASpring, pch = 19, xlim = c(0,1), ylim = c(1.0, 4.2),
     xlab = "Spring Spawner Probability", ylab = "",
     col = df.2012Sara$LDAAssign)
text(0.78, 4.25, lwd = 3, "Sara LDA")
text(0.78, 4.15, lwd = 3, "n = 4")
```




```
###################
#### 2013 Data ####
###################
 # Subset Data
  SaraYear13 <- subset(RFUnknownData, subset = RFUnknownData$birthYear == 2013)
  MicahYear13 <- subset(MicahData, subset = MicahData$birthYear == 2013)

 # Random Forest
  RF13 <- randomForest(SEASON~A1, data = MicahYear13)

  # Validate model on known data
  Prob13Micah <- predict(RF13, MicahYear13, type = "prob")
  Assign13Micah <- predict(RF13, MicahYear13)
  colnames(Prob13Micah) <- c("RFSpring", "RFWinter")

 # Use model to predict unknown samples
  Prob13Pred <- predict(RF13, SaraYear13, type = "prob")
  colnames(Prob13Pred) <- c("RFSpring", "RFWinter")
  Assign13Pred <- predict(RF13, SaraYear13)

 # Run LDA
  LDA2013 <- lda(SEASON ~ A1, data = MicahYear13)
  LDA2013MicahPred <- predict(LDA2013)
  LDA2013pred <- predict(LDA2013, newdata = SaraYear13)

 # Bind all data together
  df.2013Micah <- cbind(MicahYear13, Prob13Micah, Assign13Micah, LDA2013MicahPred$x,
                        LDA2013MicahPred$posterior, LDA2013MicahPred$class)
  colnames(df.2013Micah)[c(21, 23:25)] <- c("RFAssign", "LDASpring", "LDAWinter", "LDAAssign")
  df.2013Sara <- cbind(SaraYear13, Prob13Pred, Assign13Pred,
```

```
                        LDA2013pred$x, LDA2013pred$posterior, LDA2013pred$class)
  colnames(df.2013Sara)[c(54, 56:58)] <- c("RFAssign", "LDASpring", "LDAWinter", "LDAAssign")

 # Check accuracy of the models
  RF2013correct <- NULL
  RF2013incorrect <- NULL
  for(i in 1:nrow(df.2013Micah)){
    if(df.2013Micah$SEASON[i] == df.2013Micah$RFAssign[i]){
      RF2013correct <- c(RF2013correct, as.character(df.2013Micah$SEASON[i]))
    } else {
      RF2013incorrect <- c(RF2013incorrect, as.character(df.2013Micah$SEASON[i]))
    }
  }
  100-(length(RF2013incorrect)/length(RF2013correct)*100)
```

```
## [1] 91.89189
```

```
  LDA2013correct <- NULL
  LDA2013incorrect <- NULL
  for(i in 1:nrow(df.2013Micah)){
    if(df.2013Micah$SEASON[i] == df.2013Micah$LDAAssign[i]){
      LDA2013correct <- c(LDA2013correct, as.character(df.2013Micah$SEASON[i]))
    } else {
      LDA2013incorrect <- c(LDA2013incorrect, as.character(df.2013Micah$SEASON[i]))
    }
  }
  100-(length(LDA2013incorrect)/length(LDA2013correct)*100)
```

```
## [1] 70.96774
```

```
 # Plot data
  par(mfrow = c(1,3))
  nrow(MicahYear13[MicahYear13$SEASON == "SPRING",])
```

```
## [1] 8
```

```
  nrow(MicahYear13[MicahYear13$SEASON == "WINTER",])
```

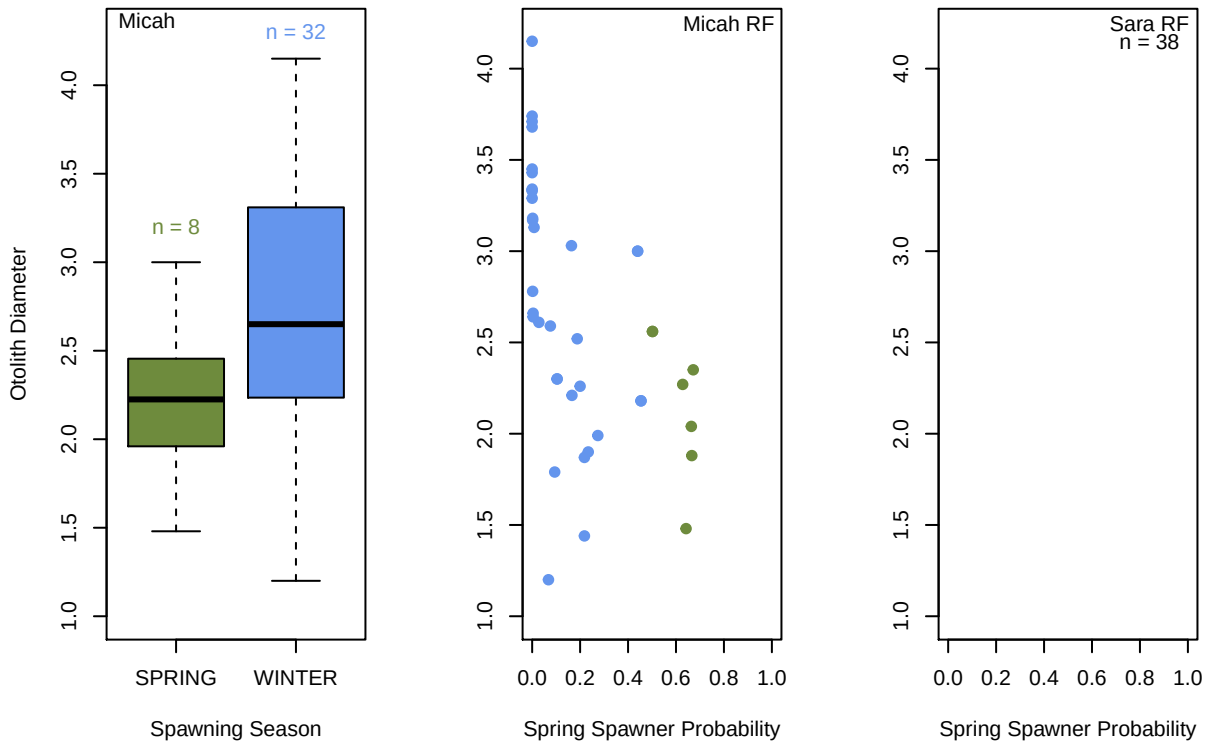
```
## [1] 32
```

```
  nrow(SaraYear13)
```

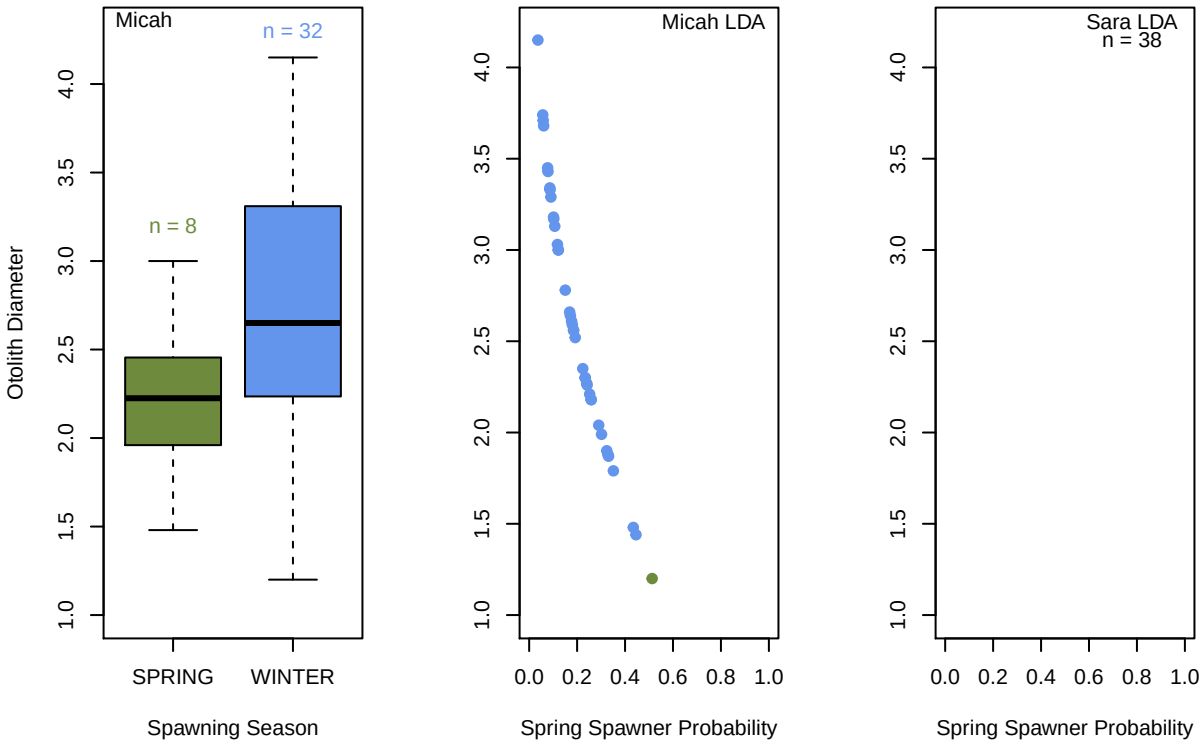
```
## [1] 38
```

```
  # Random Forest Results
  boxplot(MicahYear13$A1~MicahYear13$SEASON, ylab = "Otolith Diameter", ylim = c(1.0, 4.3),
          col = c("darkolivegreen4", "cornflowerblue"), xlab = "Spawning Season")
  text(1, 3.2, "n = 8", col = "darkolivegreen4")
  text(2, 4.3, "n = 32", col = "cornflowerblue")
  text(1.1, 4.35, pos = 2, "Micah")
  plot(df.2013Micah$A1~df.2013Micah$RFSpring, pch = 19, xlim = c(0, 1), ylim = c(1.0, 4.2),
       xlab = "Spring Spawner Probability", ylab = "",
       col = df.2013Micah$RFAssign)
  text(0.82, 4.25, "Micah RF")
  plot(df.2013Sara$A1~df.2013Sara$RFSpring, pch = 19, xlim = c(0, 1), ylim = c(1.0, 4.2),
       xlab = "Spring Spawner Probability", ylab = "",
       col = df.2013Sara$RFAssign)
```

```
text(0.84, 4.25, "Sara RF")
text(0.84, 4.15, "n = 38")
```




```
#LDA Results
boxplot(MicahYear13$A1~MicahYear13$SEASON, ylab = "Otolith Diameter", ylim = c(1.0, 4.3),
        col = c("darkolivegreen4", "cornflowerblue"), xlab = "Spawning Season")
text(1, 3.2, "n = 8", col = "darkolivegreen4")
text(2, 4.3, "n = 32", col = "cornflowerblue")
text(1.1, 4.35, pos = 2, "Micah")
plot(df.2013Micah$A1~df.2013Micah$LDASpring, pch = 19, xlim = c(0, 1), ylim = c(1.0, 4.2),
     xlab = "Spring Spawner Probability", ylab = "",
     col = df.2013Micah$LDAAssign)
text(0.77, 4.25, "Micah LDA")
plot(df.2013Sara$A1~df.2013Sara$LDASpring, pch = 19, xlim = c(0,1), ylim = c(1.0, 4.2),
     xlab = "Spring Spawner Probability", ylab = "",
     col = df.2013Sara$LDAAssign)
text(0.78, 4.25, "Sara LDA")
text(0.78, 4.15, "n = 38")
```




```
###################
#### 2014 Data ####
###################
  # Subset Data
  SaraYear14 <- subset(RFUnknownData, subset = RFUnknownData$birthYear == 2014)
  MicahYear14 <- subset(MicahData, subset = MicahData$birthYear == 2014)

 # Random Forest
  RF14 <- randomForest(SEASON~A1, data = MicahYear14)

  # Validate model on known data
  Prob14Micah <- predict(RF14, MicahYear14, type = "prob")
  Assign14Micah <- predict(RF14, MicahYear14)
  colnames(Prob14Micah) <- c("RFSpring", "RFWinter")

 # Use model to predict unknown samples
  Prob14Pred <- predict(RF14, SaraYear14, type = "prob")
  colnames(Prob14Pred) <- c("RFSpring", "RFWinter")
  Assign14Pred <- predict(RF14, SaraYear14)

 # Run LDA
  LDA2014 <- lda(SEASON ~ A1, data = MicahYear14)
  LDA2014MicahPred <- predict(LDA2014)
  LDA2014pred <- predict(LDA2014, newdata = SaraYear14)

 # Bind all data together
   df.2014Micah <- cbind(MicahYear14, Prob14Micah, Assign14Micah, LDA2014MicahPred$x,
                         LDA2014MicahPred$posterior, LDA2014MicahPred$class)
  colnames(df.2014Micah)[c(21, 23:25)] <- c("RFAssign", "LDASpring", "LDAWinter", "LDAAssign")
  df.2014Sara <- cbind(SaraYear14, Prob14Pred, Assign14Pred,
```

```
                        LDA2014pred$x, LDA2014pred$posterior, LDA2014pred$class)
  colnames(df.2014Sara)[c(54, 56:58)] <- c("RFAssign", "LDASpring", "LDAWinter", "LDAAssign")

  # Check accuracy of the models
  RF2014correct <- NULL
  RF2014incorrect <- NULL
  for(i in 1:nrow(df.2014Micah)){
    if(df.2014Micah$SEASON[i] == df.2014Micah$RFAssign[i]){
      RF2014correct <- c(RF2014correct, as.character(df.2014Micah$SEASON[i]))
    } else {
      RF2014incorrect <- c(RF2014incorrect, as.character(df.2014Micah$SEASON[i]))
    }
  }
  100-(length(RF2014incorrect)/length(RF2014correct)*100)
```

```
## [1] 100
```

```
  LDA2014correct <- NULL
  LDA2014incorrect <- NULL
  for(i in 1:nrow(df.2014Micah)){
    if(df.2014Micah$SEASON[i] == df.2014Micah$LDAAssign[i]){
      LDA2014correct <- c(LDA2014correct, as.character(df.2014Micah$SEASON[i]))
    } else {
      LDA2014incorrect <- c(LDA2014incorrect, as.character(df.2014Micah$SEASON[i]))
    }
  }
  100-(length(LDA2014incorrect)/length(LDA2014correct)*100)
```

```
## [1] 87.5
```

```
 # Plot data
  par(mfrow = c(1,3))
  # Get sample sizes
  nrow(MicahYear14[MicahYear14$SEASON == "SPRING",])
```

```
## [1] 2
```

```
  nrow(MicahYear14[MicahYear14$SEASON == "WINTER",])
```

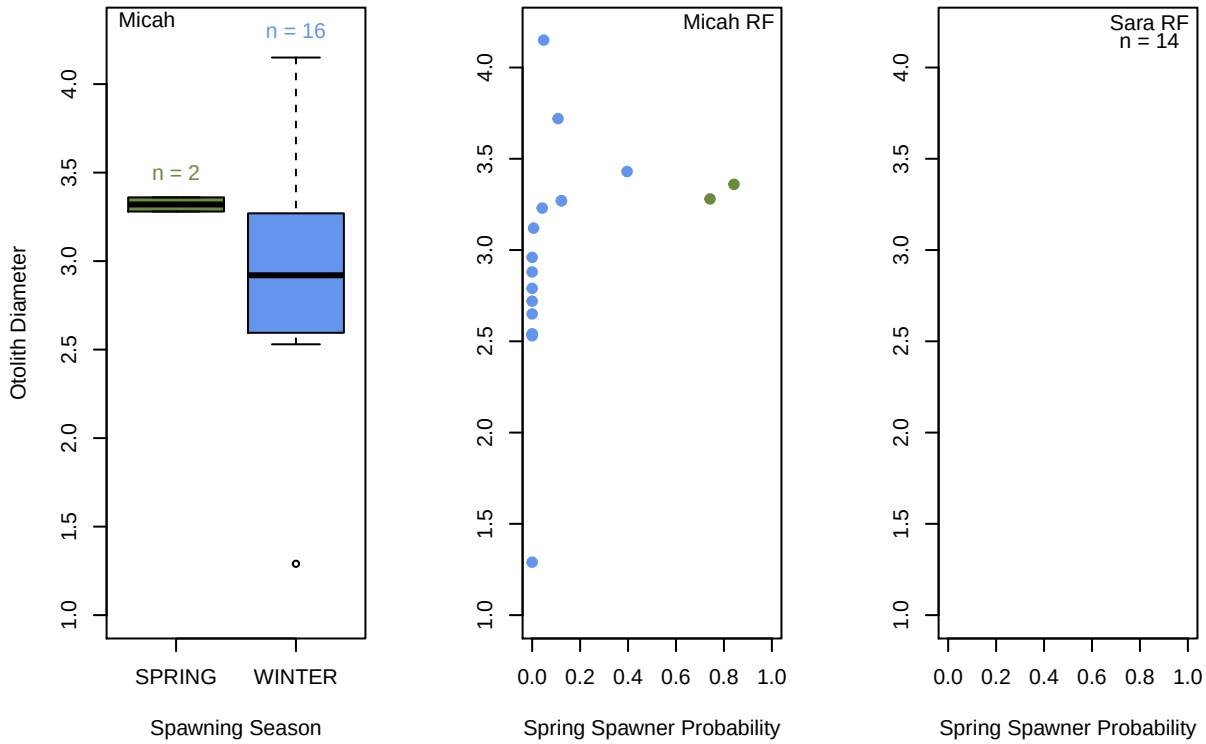
```
## [1] 16
```

```
  nrow(SaraYear14)
```

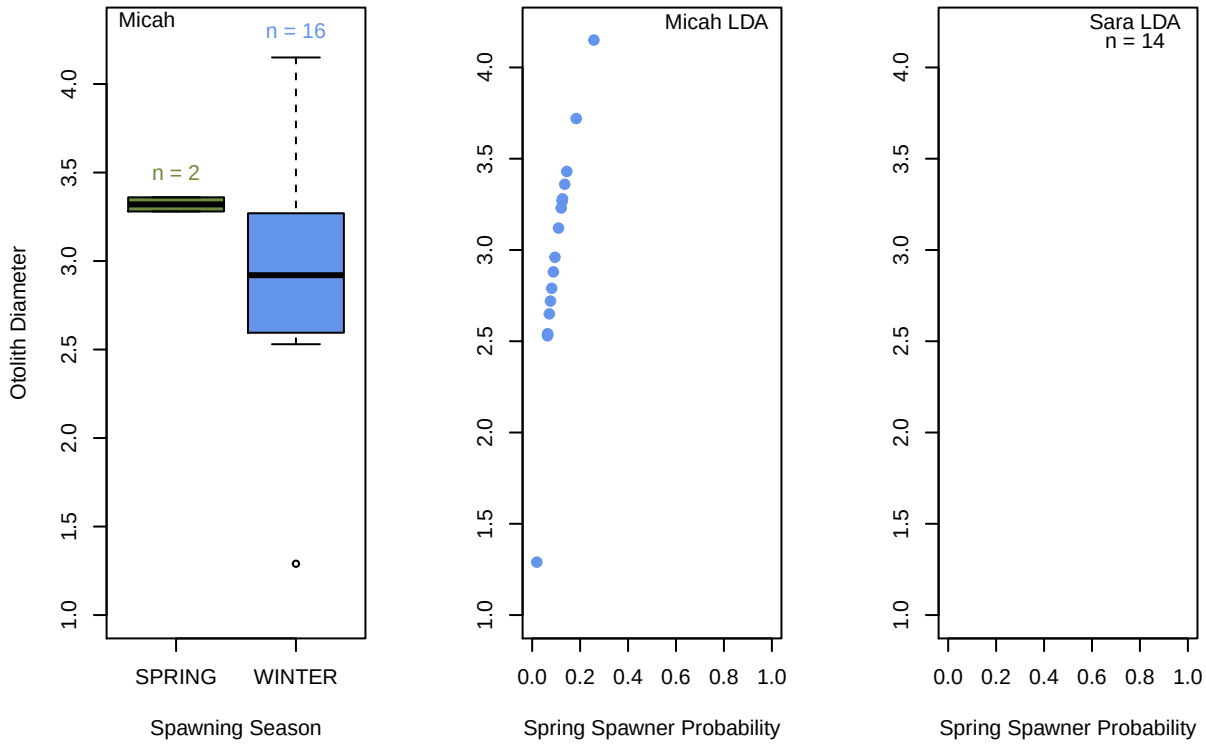
```
## [1] 14
```

```
  # Random Forest Results
  boxplot(MicahYear14$A1~MicahYear14$SEASON, ylab = "Otolith Diameter", ylim = c(1.0, 4.3),
          col = c("darkolivegreen4", "cornflowerblue"), xlab = "Spawning Season")
  text(1, 3.5, "n = 2", col = "darkolivegreen4")
  text(2, 4.3, "n = 16", col = "cornflowerblue")
  text(1.1, 4.35, pos = 2, "Micah")
  plot(df.2014Micah$A1~df.2014Micah$RFSpring, pch = 19, xlim = c(0, 1), ylim = c(1.0, 4.2),
       xlab = "Spring Spawner Probability", ylab = "",
       col = df.2014Micah$RFAssign)
  text(0.82, 4.25, "Micah RF")
  plot(df.2014Sara$A1~df.2014Sara$RFSpring, pch = 19, xlim = c(0, 1), ylim = c(1.0, 4.2),
       xlab = "Spring Spawner Probability", ylab = "",
```

```
      col = df.2014Sara$RFAssign)
text(0.84, 4.25, "Sara RF")
text(0.84, 4.15, "n = 14")
```




```
# LDA Results
boxplot(MicahYear14$A1~MicahYear14$SEASON, ylab = "Otolith Diameter", ylim = c(1.0, 4.3),
        col = c("darkolivegreen4", "cornflowerblue"), xlab = "Spawning Season")
text(1, 3.5, "n = 2", col = "darkolivegreen4")
text(2, 4.3, "n = 16", col = "cornflowerblue")
text(1.1, 4.35, pos = 2, "Micah")
plot(df.2014Micah$A1~df.2014Micah$LDASpring, pch = 19, xlim = c(0, 1), ylim = c(1.0, 4.2),
     xlab = "Spring Spawner Probability", ylab = "",
     col = df.2014Micah$LDAAssign)
text(0.77, 4.25, "Micah LDA")
plot(df.2014Sara$A1~df.2014Sara$LDASpring, pch = 19, xlim = c(0,1), ylim = c(1.0, 4.2),
     xlab = "Spring Spawner Probability", ylab = "",
     col = df.2014Sara$LDAAssign)
text(0.78, 4.25, "Sara LDA")
text(0.78, 4.15, "n = 14")
```




```
###################
#### 2015 Data ####
###################
  # Subset Data
  SaraYear15 <- subset(RFUnknownData, subset = RFUnknownData$birthYear == 2015)
  MicahYear15 <- subset(MicahData, subset = MicahData$birthYear == 2015)

 # Random Forest
  RF15 <- randomForest(SEASON~A1, data = MicahYear15)

  # Validate model on known data
  Prob15Micah <- predict(RF15, MicahYear15, type = "prob")
  Assign15Micah <- predict(RF15, MicahYear15)
  colnames(Prob15Micah) <- c("RFSpring", "RFWinter")

 # Use model to predict unknown samples
  Prob15Pred <- predict(RF15, SaraYear15, type = "prob")
  colnames(Prob15Pred) <- c("RFSpring", "RFWinter")
  Assign15Pred <- predict(RF15, SaraYear15)

 # Run LDA
  LDA2015 <- lda(SEASON ~ A1, data = MicahYear15)
  LDA2015MicahPred <- predict(LDA2015)
  LDA2015pred <- predict(LDA2015, newdata = SaraYear15)

 # Bind all data together
  df.2015Micah <- cbind(MicahYear15, Prob15Micah, Assign15Micah, LDA2015MicahPred$x,
                        LDA2015MicahPred$posterior, LDA2015MicahPred$class)
  colnames(df.2015Micah)[c(21, 23:25)] <- c("RFAssign", "LDASpring", "LDAWinter", "LDAAssign")
  df.2015Sara <- cbind(SaraYear15, Prob15Pred, Assign15Pred,
```

```
                        LDA2015pred$x, LDA2015pred$posterior, LDA2015pred$class)
  colnames(df.2015Sara)[c(54, 56:58)] <- c("RFAssign", "LDASpring", "LDAWinter", "LDAAssign")

  # Check accuracy of the models
  RF2015correct <- NULL
  RF2015incorrect <- NULL
  for(i in 1:nrow(df.2015Micah)){
    if(df.2015Micah$SEASON[i] == df.2015Micah$RFAssign[i]){
      RF2015correct <- c(RF2015correct, as.character(df.2015Micah$SEASON[i]))
    } else {
      RF2015incorrect <- c(RF2015incorrect, as.character(df.2015Micah$SEASON[i]))
    }
  }
  100-(length(RF2015incorrect)/length(RF2015correct)*100)
```

```
## [1] 100
```

```
  LDA2015correct <- NULL
  LDA2015incorrect <- NULL
  for(i in 1:nrow(df.2015Micah)){
    if(df.2015Micah$SEASON[i] == df.2015Micah$LDAAssign[i]){
      LDA2015correct <- c(LDA2015correct, as.character(df.2015Micah$SEASON[i]))
    } else {
      LDA2015incorrect <- c(LDA2015incorrect, as.character(df.2015Micah$SEASON[i]))
    }
  }
  100-(length(LDA2015incorrect)/length(LDA2015correct)*100)
```

```
## [1] 100
```

```
 # Plot data
  par(mfrow = c(1,3))
  # Get sample sizes
  nrow(MicahYear15[MicahYear15$SEASON == "SPRING",])
```

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

```
  nrow(MicahYear15[MicahYear15$SEASON == "WINTER",])
```

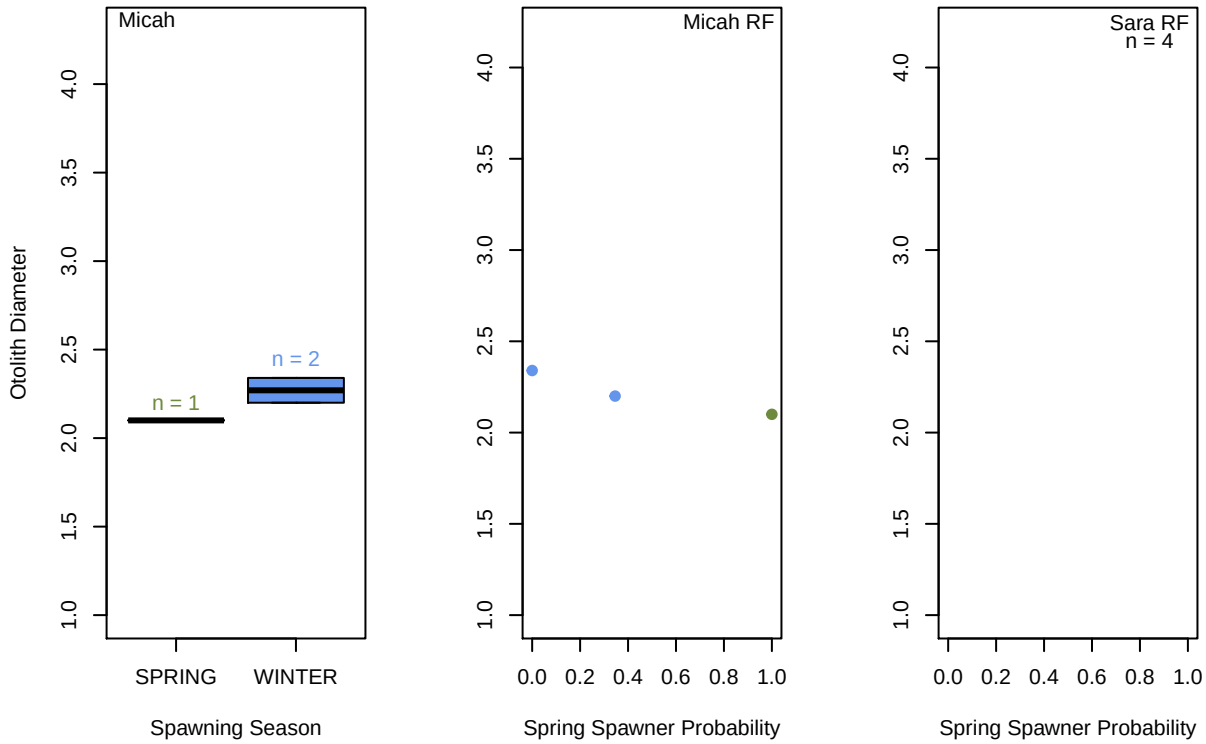
```
## [1] 2
```

```
  nrow(SaraYear15)
```

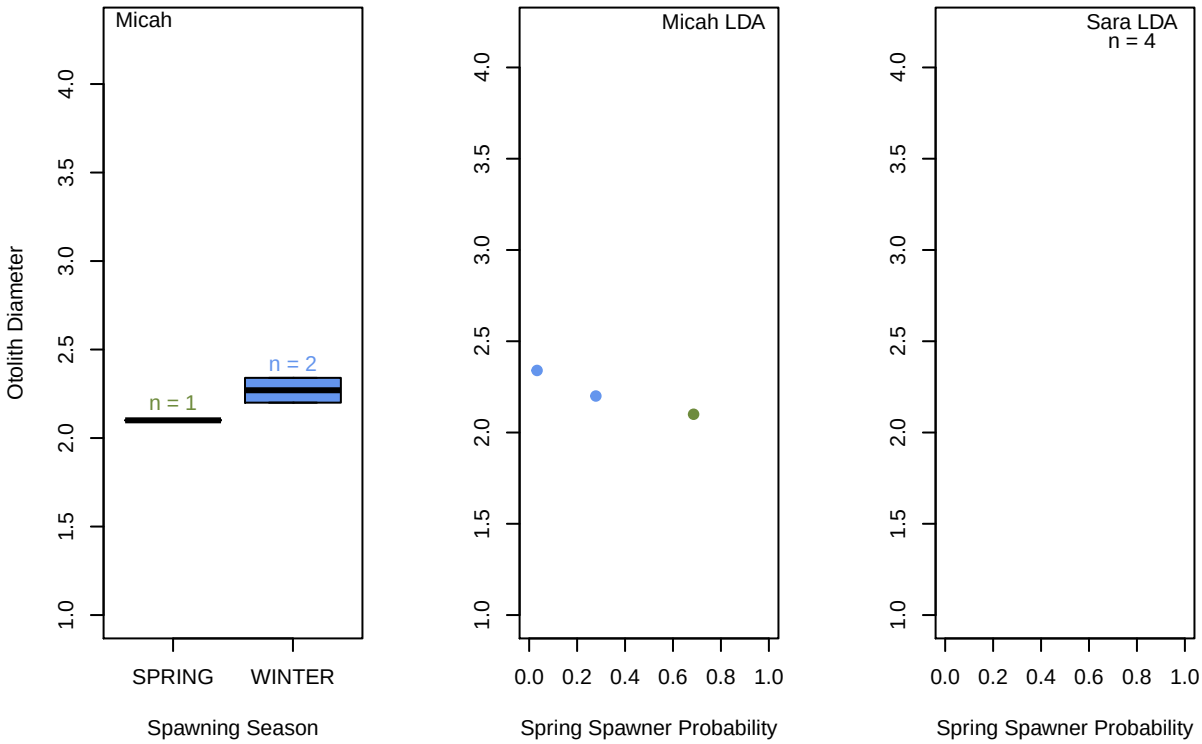
```
## [1] 4
```

```
  # Random Forest Results
  boxplot(MicahYear15$A1~MicahYear15$SEASON, ylab = "Otolith Diameter", ylim = c(1.0, 4.3),
          col = c("darkolivegreen4", "cornflowerblue"), xlab = "Spawning Season")
  text(1, 2.2, "n = 1", col = "darkolivegreen4")
  text(2, 2.45, "n = 2", col = "cornflowerblue")
  text(1.1, 4.35, pos = 2, "Micah")
  plot(df.2015Micah$A1~df.2015Micah$RFSpring, pch = 19, xlim = c(0, 1), ylim = c(1.0, 4.2),
       xlab = "Spring Spawner Probability", ylab = "",
       col = df.2015Micah$RFAssign)
  text(0.82, 4.25, "Micah RF")
  plot(df.2015Sara$A1~df.2015Sara$RFSpring, pch = 19, xlim = c(0, 1), ylim = c(1.0, 4.2),
       xlab = "Spring Spawner Probability", ylab = "",
```

```
      col = df.2015Sara$RFAssign)
text(0.84, 4.25, "Sara RF")
text(0.84, 4.15, "n = 4")
```




```
# LDA Results
boxplot(MicahYear15$A1~MicahYear15$SEASON, ylab = "Otolith Diameter", ylim = c(1.0, 4.3),
        col = c("darkolivegreen4", "cornflowerblue"), xlab = "Spawning Season")
text(1, 2.2, "n = 1", col = "darkolivegreen4")
text(2, 2.42, "n = 2", col = "cornflowerblue")
text(1.1, 4.35, pos = 2, "Micah")
plot(df.2015Micah$A1~df.2015Micah$LDASpring, pch = 19, xlim = c(0, 1), ylim = c(1.0, 4.2),
     xlab = "Spring Spawner Probability", ylab = "",
     col = df.2015Micah$LDAAssign)
text(0.77, 4.25, "Micah LDA")
plot(df.2015Sara$A1~df.2015Sara$LDASpring, pch = 19, xlim = c(0,1), ylim = c(1.0, 4.2),
     xlab = "Spring Spawner Probability", ylab = "",
     col = df.2015Sara$LDAAssign)
text(0.78, 4.25, "Sara LDA")
text(0.78, 4.15, "n = 4")
```


Micah
Otolith Diameter
1.0
1.5
2.0
2.5
3.0
3.5
4.0
n = 1
n = 2
SPRING
WINTER
Spawning Season
Micah LDA
0.0
0.2
0.4
0.6
0.8
1.0
Spring Spawner Probability
Sara LDA
n = 4
Spring Spawner Probability