Subsample BlackfootFish

Mark Greenwood

10/6/2020

# Code to reduce file size

In order to avoid RAM limit issues in the shiny environment, the original size of the BlackfootFish data set was reduced for use in the R workshops using this code for reproducibility. The original data set had *n*=18,352. The reduced data set is based on stratified sampling half of the observations taken within each year.

BlackfootFish <- read.csv("data/BlackfootFish.csv", header = TRUE)  
  
BlackfootFish$year.species <- with(BlackfootFish, interaction(year, species))  
levels(BlackfootFish$year.species)

## [1] "1989.Brown" "1990.Brown" "1991.Brown" "1993.Brown" "1996.Brown"  
## [6] "1998.Brown" "2000.Brown" "2002.Brown" "2004.Brown" "2006.Brown"  
## [11] "1989.Bull" "1990.Bull" "1991.Bull" "1993.Bull" "1996.Bull"   
## [16] "1998.Bull" "2000.Bull" "2002.Bull" "2004.Bull" "2006.Bull"   
## [21] "1989.RBT" "1990.RBT" "1991.RBT" "1993.RBT" "1996.RBT"   
## [26] "1998.RBT" "2000.RBT" "2002.RBT" "2004.RBT" "2006.RBT"   
## [31] "1989.WCT" "1990.WCT" "1991.WCT" "1993.WCT" "1996.WCT"   
## [36] "1998.WCT" "2000.WCT" "2002.WCT" "2004.WCT" "2006.WCT"

library(mosaic)  
set.seed(1234)  
BFr <- sample(x=BlackfootFish, size=9000)  
tally(~year.species, data=BFr)

## year.species  
## 1989.Brown 1990.Brown 1991.Brown 1993.Brown 1996.Brown 1998.Brown 2000.Brown   
## 121 117 115 174 113 108 210   
## 2002.Brown 2004.Brown 2006.Brown 1989.Bull 1990.Bull 1991.Bull 1993.Bull   
## 214 181 196 30 19 28 11   
## 1996.Bull 1998.Bull 2000.Bull 2002.Bull 2004.Bull 2006.Bull 1989.RBT   
## 22 21 40 47 23 40 687   
## 1990.RBT 1991.RBT 1993.RBT 1996.RBT 1998.RBT 2000.RBT 2002.RBT   
## 942 922 683 288 375 669 493   
## 2004.RBT 2006.RBT 1989.WCT 1990.WCT 1991.WCT 1993.WCT 1996.WCT   
## 413 548 23 61 68 98 74   
## 1998.WCT 2000.WCT 2002.WCT 2004.WCT 2006.WCT   
## 138 225 146 146 171

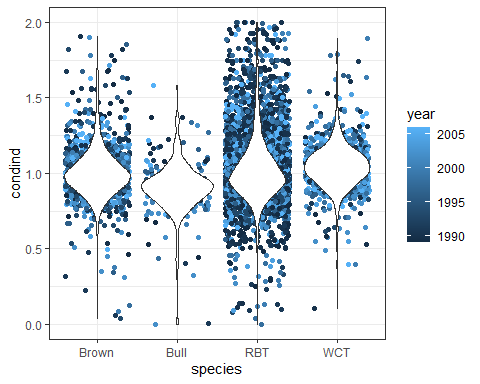
condition\_index <- function(x,y){  
 ifelse(!is.na(x)&!is.na(y), c\_in <- 100000\*(x/(y^3)), NA)  
 return(c\_in)  
}  
  
BFr$condind <- condition\_index(BFr$weight, BFr$length)  
  
summary(BFr$condind)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.000 0.910 1.000 9.936 1.112 19411.084 864

D1 <- subset(BFr, !is.na(condind) & condind <= 2)  
  
summary(D1$condind)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0000 0.9072 0.9957 1.0203 1.1018 2.0000

library(ggplot2)  
  
D1 %>%   
 ggplot(aes(x=species, y=condind)) +  
 geom\_jitter(aes(col=year)) +  
 geom\_violin() +   
 theme\_bw()



library(readr)  
write\_csv(BFr, "data/BlackfootFish2.csv")