Hughs project

Statistical coding group

21/08/2019

library(tidyverse)

## -- Attaching packages ------------------------------------------------------------------------------------ tidyverse 1.2.1 --

## v ggplot2 3.2.1 v purrr 0.3.2  
## v tibble 2.1.3 v dplyr 0.8.3  
## v tidyr 0.8.3 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

## -- Conflicts --------------------------------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(DT)  
library(plotly)

##   
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':  
##   
## last\_plot

## The following object is masked from 'package:stats':  
##   
## filter

## The following object is masked from 'package:graphics':  
##   
## layout

## Data import

dat <- read\_csv("./data/tracking\_master\_20190708.csv")

## Parsed with column specification:  
## cols(  
## date = col\_character(),  
## pass = col\_character(),  
## tag = col\_double(),  
## lon = col\_double(),  
## lat = col\_double(),  
## depth = col\_character(),  
## flow = col\_double(),  
## br = col\_double(),  
## b = col\_double(),  
## c = col\_double(),  
## p = col\_double(),  
## g = col\_double(),  
## s = col\_double(),  
## silt = col\_double(),  
## rip = col\_double(),  
## aqu = col\_double(),  
## und = col\_double(),  
## leaf = col\_double(),  
## comment = col\_character()  
## )

## Session work

# Anthony  
# date  
#...  
  
#library  
library(tidyverse)  
  
# import   
dat <- read.csv("data/tracking\_master\_20190708.csv")  
  
# data structure  
glimpse(dat)

## Observations: 1,418  
## Variables: 19  
## $ date <fct> 2018.01.17, 2018.01.17, 2018.01.17, 2018.01.17, 2018.0...  
## $ pass <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...  
## $ tag <int> 306750, 396174, 397049, 397419, 400116, 387550, 386869...  
## $ lon <int> 17, 17, 17, 17, 17, 17, 17, 17, 17, 17, 55, 55, 107, 5...  
## $ lat <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ depth <fct> , , , , , , , , , , , , , , , , , , , , , , , ,   
## $ flow <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ br <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ b <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ c <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ p <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ g <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ s <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ silt <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ rip <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ aqu <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ und <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ leaf <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ comment <fct> Release location after tagging, Release location after...

# no of fish  
 count(distinct(dat, tag), na.rm = T) # sums the number of fish we caught (74)

## # A tibble: 1 x 2  
## na.rm n  
## <lgl> <int>  
## 1 TRUE 74

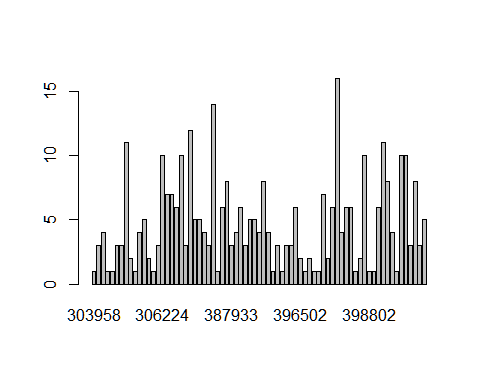
# provides number of "random" sampled points   
 table(dat$tag, exclude = NULL) # counts all the fish and

##   
## 303958 304487 304497 304659 304681 305043 305079 305280 305512 305825   
## 1 3 4 1 1 3 3 11 2 1   
## 305892 305901 305973 305997 306028 306224 306236 306346 306750 385464   
## 4 5 2 1 3 10 7 7 6 10   
## 385725 385905 386103 386790 386869 387435 387550 387726 387779 387898   
## 3 12 5 5 4 3 14 1 6 8   
## 387933 388116 394260 394546 394587 394605 394969 395226 395345 395373   
## 3 4 6 3 5 5 4 8 4 1   
## 395410 395824 396010 396044 396174 396502 396580 396601 396705 396894   
## 3 1 3 3 6 2 1 2 1 1   
## 396941 396966 397048 397049 397419 397443 397557 397744 398343 398543   
## 7 2 6 16 4 6 6 1 2 10   
## 398802 398913 399607 400116 400288 400526 400946 401032 401097 401858   
## 1 1 6 11 8 4 1 10 10 3   
## 402162 402369 403424 <NA>   
## 8 3 5 1075

table(table(dat$tag, exclude = NULL))

##   
## 1 2 3 4 5 6 7 8 10 11 12 14 16 1075   
## 15 6 13 8 6 8 3 4 5 2 1 1 1 1

barplot(table(dat$tag)) # individual fish by tag number by frequency



sum(table(table(dat$tag)))

## [1] 73

# base r language:  
 table(is.na(dat$tag))

##   
## FALSE TRUE   
## 343 1075

# number of fish observations by non-fish observations   
 dat %>% group\_by(is.na(tag)) %>%  
 summarise(n = n())

## # A tibble: 2 x 2  
## `is.na(tag)` n  
## <lgl> <int>  
## 1 FALSE 343  
## 2 TRUE 1075

# base r language: mean(dat$tag)  
# tidyverse: dat %>% mean(tag)  
   
# Do they prefer boulders?  
   
 ##assumed to be in same pass  
 # second pass = distrubed  
   
 table(dat$pass)

##   
## 1 2 night   
## 1075 230 112 1

# hist(dat$br, breaks = 30)  
   
 #NA = zero  
 #hadley drama  
 dat1 <- dat %>%  
 mutate\_at(vars(br:silt),funs(replace(.,is.na(.),0)))

## Warning: funs() is soft deprecated as of dplyr 0.8.0  
## Please use a list of either functions or lambdas:   
##   
## # Simple named list:   
## list(mean = mean, median = median)  
##   
## # Auto named with `tibble::lst()`:   
## tibble::lst(mean, median)  
##   
## # Using lambdas  
## list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))  
## This warning is displayed once per session.

table(dat1$br)

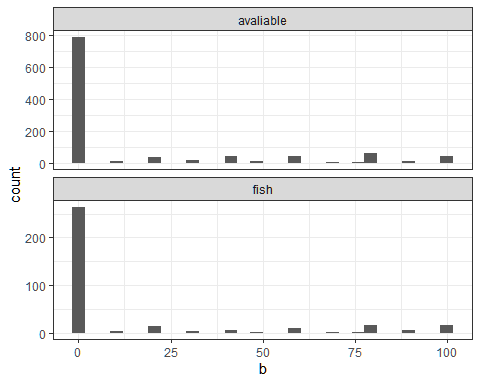
##   
## 0 5 50   
## 1416 1 1

# tidy as  
  
 #no zeros  
glimpse(dat1)

## Observations: 1,418  
## Variables: 19  
## $ date <fct> 2018.01.17, 2018.01.17, 2018.01.17, 2018.01.17, 2018.0...  
## $ pass <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...  
## $ tag <int> 306750, 396174, 397049, 397419, 400116, 387550, 386869...  
## $ lon <int> 17, 17, 17, 17, 17, 17, 17, 17, 17, 17, 55, 55, 107, 5...  
## $ lat <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ depth <fct> , , , , , , , , , , , , , , , , , , , , , , , ,   
## $ flow <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ br <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...  
## $ b <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...  
## $ c <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...  
## $ p <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...  
## $ g <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...  
## $ s <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...  
## $ silt <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...  
## $ rip <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ aqu <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ und <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ leaf <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ comment <fct> Release location after tagging, Release location after...

dat1$fish <- ifelse(is.na(dat1$tag), "avaliable", "fish")  
# histogram of fish   
  
ggplot(dat1, aes(x = b)) +  
 geom\_histogram() +   
 facet\_wrap(~fish, scale = "free\_y", ncol = 1) +  
 theme\_bw()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



#many many zeros  
# change to grouping  
  
str(dat1$b)

## num [1:1418] 0 0 0 0 0 0 0 0 0 0 ...

dat1$b.pa <- ifelse(dat1$silt > 0, 1, 0)  
  
 names(dat1)

## [1] "date" "pass" "tag" "lon" "lat" "depth" "flow"   
## [8] "br" "b" "c" "p" "g" "s" "silt"   
## [15] "rip" "aqu" "und" "leaf" "comment" "fish" "b.pa"

tab1 <- table(dat1$fish, dat1$b.pa)  
tab1

##   
## 0 1  
## avaliable 766 309  
## fish 319 24

#ratios of presence/absents  
tab1[,2]/rowSums(tab1)

## avaliable fish   
## 0.28744186 0.06997085

chisq.test(tab1)

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: tab1  
## X-squared = 67.235, df = 1, p-value = 2.41e-16

ggplot(dat1, aes(x = silt)) +  
 geom\_histogram() +   
 facet\_wrap(~fish, scale = "free\_y", ncol = 1) +  
 theme\_bw()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

