Code for Overview Doc

Annabelle Stanley, Gwen Iacona, Becky Epanchin-Niell

June 12, 2020

## Set up file structure

## load in different datasets

checking against Detailed Methods that all species have correct docs

# Information from Partner dataset (previous seen in thesis and from larger dataset that salafsky is a subset of)

## 1) what are the characteristics of partnership for species that are precluded from listing?

### a. How many partners are there?

#### - i) Total number of partners

## [1] 197

#### - ii) Histogram of # partners/species

Graph thought - most species have very few partners working on them with one outlier in this dataset

#### - iii) Max, min, median, and 1st and 3rd quantiles across species

summary(PartnerData$TotalPartners)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.000 2.000 3.500 6.019 7.000 51.000

### b. How does number of partners relate to species/recovery programs characteristics?

#### - i) Thesis tested: Area, %publicland, taxa (1 = flowering plant), employment, total number of threats (a few other predictors looked at and removed due to high correlation were human footprint, threats = habitat as threat (1/0 as y/n))

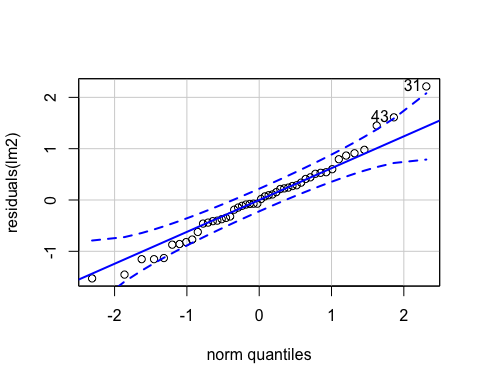
#### - ii) Test performing regression of form

• log(# of partners) = B1(taxa) + B2(%public land) + B3(area) + B4(sum of relevant employment sectors) + B5(total number of threats)

lm2 <- lm(log(total) ~ taxa + percentpublic + area\_x + nsumemploy + total\_x\_x, data=RegData)  
summary(lm2)

##   
## Call:  
## lm(formula = log(total) ~ taxa + percentpublic + area\_x + nsumemploy +   
## total\_x\_x, data = RegData)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.5279 -0.4195 -0.0260 0.4178 2.2137   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.906e-01 3.600e-01 2.474 0.0175 \*  
## taxa -3.463e-01 2.647e-01 -1.308 0.1980   
## percentpublic 1.517e-02 4.471e-01 0.034 0.9731   
## area\_x 4.853e-12 3.607e-12 1.345 0.1857   
## nsumemploy 6.771e-07 3.084e-06 0.220 0.8273   
## total\_x\_x 2.543e-01 9.482e-02 2.682 0.0104 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.8096 on 42 degrees of freedom  
## (3 observations deleted due to missingness)  
## Multiple R-squared: 0.24, Adjusted R-squared: 0.1495   
## F-statistic: 2.653 on 5 and 42 DF, p-value: 0.03585

#plot(lm2) #don't need so many plots  
  
qqPlot(residuals(lm2))



## 31 43   
## 30 41

AIC(lm2)

## [1] 123.5299

qqPlot looks normal! Most of this code ^ was taken from regressionmodelforthesis.Rmd

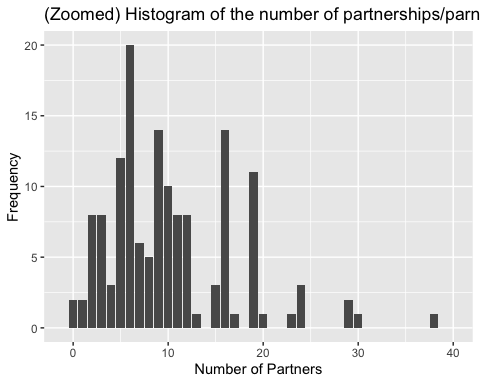
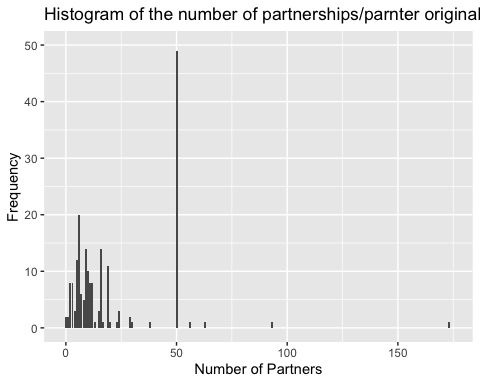
## b) who are the organizations partnering with FWS?

### a. List of all the organizations involved

### b. Which partners are most commonly working with others

#### - i) Histogram of partnerships/partner

## X1 rssum  
## 1 USFWS 173  
## 2 BLM 93  
## 3 USFS 63  
## 4 Natural Resources Conservation Service 56

 Graph takeaway – most partners are working with a number of other partner, but not an excessive amount (couple of outliers here) (second graphi is just zoomed in on the x-axis)

#### - ii) Max, min, median, and 1st and 3rd quantiles partners/partner

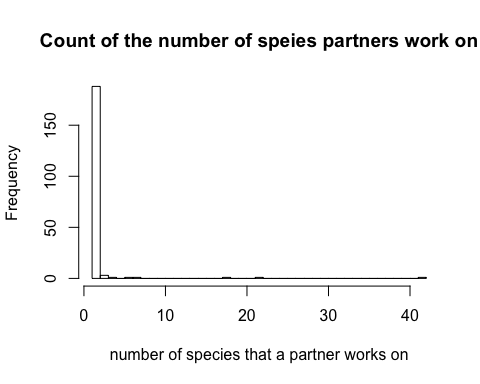
summary(explore)

## X1 rssum   
## Length:198 Min. : 0.00   
## Class :character 1st Qu.: 6.00   
## Mode :character Median : 12.00   
## Mean : 21.77   
## 3rd Qu.: 50.00   
## Max. :173.00

### c. How many species each partner is involved in projects on

#### - i) Histogram of species /partner

#### - ii) Max, min, median, and 1st and 3rd quantiles species/partner



##   
## 1 2 3 4 6 7 18 22 42   
## 171 17 3 1 1 1 1 1 1

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.000 1.000 1.000 1.589 1.000 42.000

Caption - most partners only work on one species and the outliers tend to be federal agency (biggest outlier is FWS)

# We use this data to answer

## All the following is applicable for Salafsky

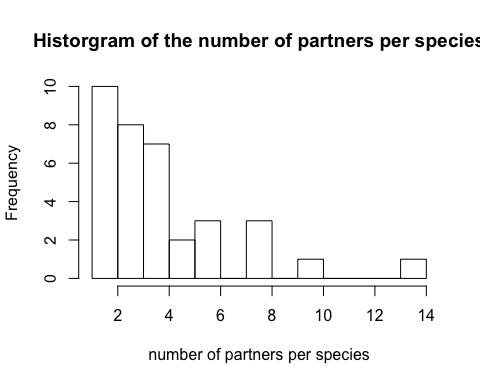
\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

The following information is calculated without using the multi partner stings

#### - i) Salafsky Total number of partners

## [1] 89

#### - ii) Salafsky Histogram of # partners/species

 Caption: Still highly skewed – most species don’t have more than 6 partners with two minor outliers

#### - iii) Salafsky Max, min, median, and 1st and 3rd quantiles across species

summary(nopartners\_noM)

## Scientific.name tally   
## Length:35 Min. : 1.0   
## Class :character 1st Qu.: 2.0   
## Mode :character Median : 3.0   
## Mean : 4.2   
## 3rd Qu.: 5.0   
## Max. :14.0

### b. How does number of partners relate to species/recovery programs characteristics?

* regression for salafsky found below

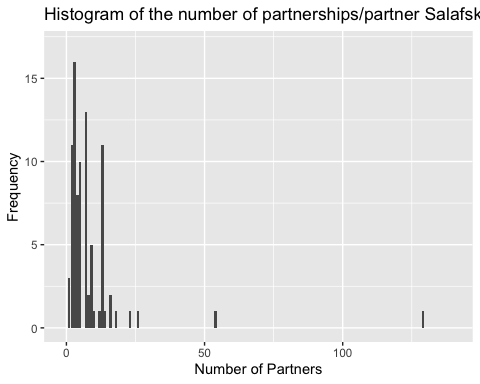
## b) Salafsky who are the organizations partnering with FWS?

### a. Salafsky List of all the organizations involved

### b. Salafsky Which partners are most commonly working with others

#### - \*\* Salafsky i) Histogram of partnerships/partner\*\*

## [1] "Aliciella caespitosa" "Allium gooddingii"   
## [3] "Astragalus anserinus" "Castilleja christii"   
## [5] "Cicindela albissima" "Cimicifuga arizonica"   
## [7] "Erigeron basalticus" "Euphydryas anicia cloudcrofti"   
## [9] "Fallicambarus gordoni" "Horkelia hendersonii"   
## [11] "Iotichthys phlegethontis" "Lepidium papilliferum"   
## [13] "Lithobates onca" "Lupinus aridus ssp. ashlandensis"  
## [15] "Moxostoma" "Nerodia erythrogaster neglecta"   
## [17] "Oncorhynchus clarkii virginalis" "Oncorhynchus mykiss aquilarum"   
## [19] "Oncorhynchus mykiss ssp." "Opuntia X multigeniculata"   
## [21] "Pediocactus paradinei" "Phacelia stellaris"   
## [23] "Phrynosoma mcallii" "Potentilla basaltica"   
## [25] "Pseudanophthalmus catorycetes" "Pseudanophthalmus pholeter"   
## [27] "Pyrgulopsis morrisoni" "Sceloporus arenicolus"   
## [29] "Solidago plumosa" "Sonorella macrophallus"   
## [31] "Sylvilagus transitionalis" "Thymallus arcticus"   
## [33] "Urocitellus endemicus" "Urocitellus washingtoni"   
## [35] "Zaitzevia thermae"

 Caption: This is a histogram of the number of partnerships per partner. Most partners don’t have very many other partners with the exception of FWS and BLM

#### - ii) Salafsky Max, min, median, and 1st and 3rd quantiles partners/partner

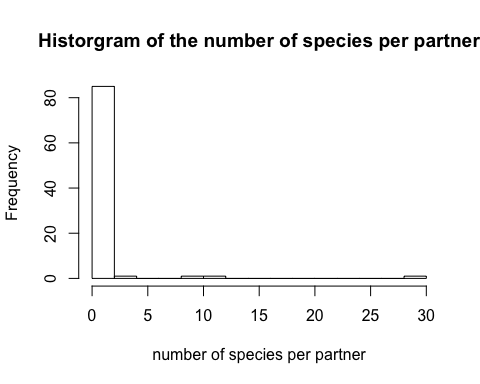
summary(modno\_orglist\_noparnters\_noM)

## partner.in.agreement nopartersperpartner  
## Length:89 Min. : 1.000   
## Class :character 1st Qu.: 3.000   
## Mode :character Median : 5.000   
## Mean : 8.629   
## 3rd Qu.: 9.000   
## Max. :129.000

### c. Salafsky How many species each partner is involved in projects on

#### - i) Histogram of species /partner

#### - ii) Max, min, median, and 1st and 3rd quantiles species/partner

 Caption: The number of species a partner is working on. Similarly to bigger dataset, most partners only work on one species

### d. Network representation of who works with who. Partners as nodes (weighted by number of species they work on), edges between partners weighted by number of species they work on together

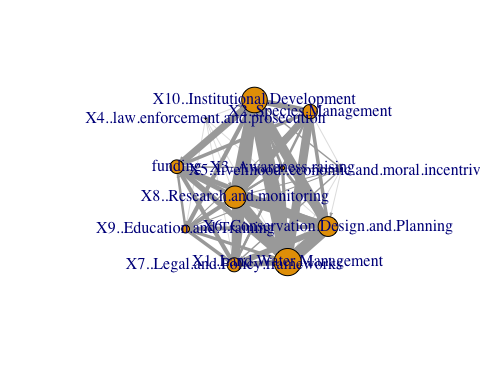
Gwen’s note - or use choard diagram - I don’t think this lends itself to a high number of “nodes” \*\* Need to look at network documentation to find out how to set up from matrix (everything I’ve read so far has been based on lists)

current code source -<https://www.mjdenny.com/Preparing_Network_Data_In_R.html>

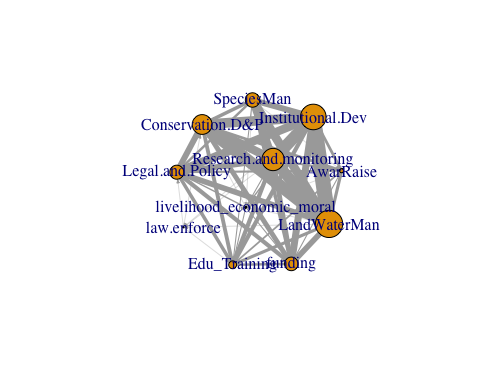
condense <- sdata[,c(1,2,6,10:20)] #select relevant columns  
  
condense <- condense[-which(condense$type.of.partners == "M"),] #remove ones with multiple partners in string   
#this is avoiding the data issue instead of addressing it  
  
condense <- condense %>% select(Scientific.name, partner.in.agreement, X1..Land.Water.Management:funding) %>% group\_by(Scientific.name)  
#losing scientific\_name so will add with join   
#con <- condense2 %>% left\_join(condense, by = c("X1..Land.Water.Management" , "X2..Species.Management" , "X3..Awareness.raising" , "X4..law.enforcement.and.prosecution" , "X5..livelihood..economic.and.moral.incentrives","X6..Conservation.Design.and.Planning" , "X7..Legal.and.Policy.frameworks" , "X8..Research.and.monitoring" , "X9..Education.and.Training" , "X10..Institutional.Development" , "funding"))  
  
  
#actions network   
#does not include multipartner strings   
  
## note, need to go back and check when vectors were created   
  
rawdf <- condense # take datafram of 0s and 1s without multi species strings   
#remove extra columns   
rawdf <- rawdf[,-c(1,2)] #remove non numeric columns  
trn <- t(rawdf) #transform  
rawdf <- t(trn) #rawdf was coming up as non numeric so tranformed again to overwrite  
  
trial1 <- trn %\*% rawdf  
  
#double check numbers   
colSums(rawdf) #checks out!!

## X1..Land.Water.Management   
## 66   
## X2..Species.Management   
## 35   
## X3..Awareness.raising   
## 10   
## X4..law.enforcement.and.prosecution   
## 2   
## X5..livelihood..economic.and.moral.incentrives   
## 4   
## X6..Conservation.Design.and.Planning   
## 48   
## X7..Legal.and.Policy.frameworks   
## 34   
## X8..Research.and.monitoring   
## 54   
## X9..Education.and.Training   
## 19   
## X10..Institutional.Development   
## 62   
## funding   
## 33

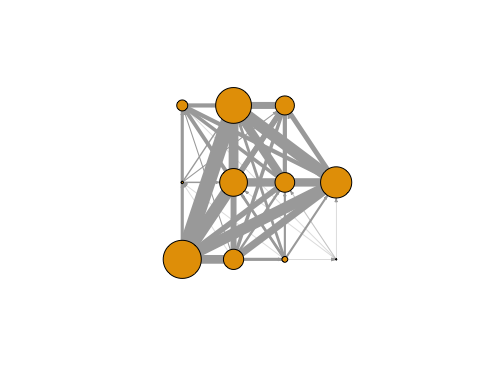
########### put into network analysis code   
  
  
tactadjmat <- trial1  
  
tactadjmat[lower.tri(tactadjmat)] <- NA  
diag(tactadjmat) <- NA  
  
actlist <- tactadjmat %>% melt()   
actlist <- actlist[-which(is.na(actlist$value)),]  
  
#nsize <- tolist %>% filter((Var1 == Var2)) #set new vector with diag values  
#tolist <- tolist %>% filter((Var1 != Var2))   
  
#allthreat <- tolist %>% right\_join(nsize, by = c("Var1"))  
#allthreat <- allthreat[,-c(4)]  
  
actnet <- graph\_from\_data\_frame(d=actlist, directed=T)   
  
  
#plot(net, edge.arrow.size=.2, edge.curved=0,  
  
 # vertex.color="orange", vertex.frame.color="#555555",  
  
 # vertex.label=V(net)$Var1, vertex.label.color="black",  
  
 # vertex.label.cex=.7, vertex.size = net$value.y)  
  
E(actnet)$arrow.size <- .2 #determine size of arrows (this makes the heads real small so they are more like lines)  
#V(net)$names <- c("a","b","c","d","e","f","g","h") #naming the nodes -- try this later so don't over write what actal names are   
#V(df)$names <- c("hab\_x\_x", "over\_x\_x", "poll\_x\_x", "spsp\_x\_x", "env\_x\_x", "demo\_x\_x")  
node.size<-setNames(c(66,35,10,2,4,48,34,54,19,62,33), c("X1..Land.Water.Management" ,"X2..Species.Management", "X3..Awareness.raising", "X4..law.enforcement.and.prosecution" , "X5..livelihood..economic.and.moral.incentrives", "X6..Conservation.Design.and.Planning","X7..Legal.and.Policy.frameworks","X8..Research.and.monitoring" ,"X9..Education.and.Training","X10..Institutional.Development","funding"))  
   
  
#so not setting in the same order   
  
  
E(actnet)$width <- E(actnet)$value\*.35   
  
plot(actnet, vertex.size=node.size\*.5)



#plot(net, edge.arrow.size=0.01,vertex.label.color = "black", vertex.size=node.size)  
  
  
node.size<-setNames(c(66,35,10,2,4,48,34,54,19,62,33), c("LandWaterMan" ,"SpeciesMan", "AwarRaise", "law.enforce" , "livelihood\_economic\_moral", "Conservation.D&P","Legal.and.Policy","Research.and.monitoring" ,"Edu\_Training","Institutional.Dev","funding")) #tried resetting names but nothing happened   
   
  
#changed label names and layout  
l <- layout\_with\_fr(actnet)  
plot(actnet, vertex.label = c("LandWaterMan" ,"SpeciesMan", "AwarRaise", "law.enforce" , "livelihood\_economic\_moral", "Conservation.D&P","Legal.and.Policy","Research.and.monitoring" ,"Edu\_Training","Institutional.Dev","funding"), layout = l, vertex.size=node.size\*.5)



#trying to show edge lengths without distraction of labels  
l <- layout\_on\_grid(actnet)  
plot(actnet, vertex.label=NA, layout = l, vertex.size=node.size\*.75)



### Network for threats

### set up (taken from lower in script)  
#this is species specific so ==M not relevant   
  
alldf <- sdata #change variable name   
  
species <- alldf[,c(2,10:20)] #unable to knit with select so indexed   
  
eachsp <- species %>% group\_by(Scientific.name) %>% summarise\_each(funs(sum))

## 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.

allspindf <- eachsp  
new <- tdata #need to join with tdata (renamed here)  
###new$scientific\_name  
#select relevant columns  
tthreats <- select(new, scientific\_name, hab\_x\_x:threats\_addressed\_by\_conservation\_x\_x)  
  
###tthreats$scientific\_name  
#need to join with tdata   
#to do so need to change col name so match   
  
tthreats <- rename(tthreats, Sciname = scientific\_name)  
allspindf <- rename(allspindf, Sciname = Scientific.name)  
  
threats <- left\_join(allspindf, tthreats, join\_by = Sciname)

## Joining, by = "Sciname"

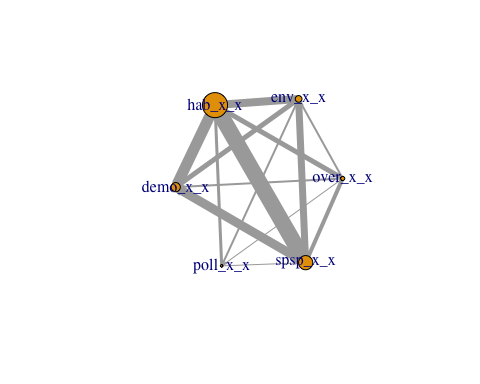
################### for threats   
  
tdf <- threats #df made somewhere else in script   
#take away excess columns   
tdf <- tdf[,c(13:18)]  
tdf <- tdf[-which(is.na(tdf$env\_x\_x)),] #remove nas  
trantdf <- t(tdf)  
is.numeric(tdf)

## [1] FALSE

tdf <- t(trantdf)  
  
threatmat <- trantdf %\*% tdf  
  
  
#double check numbers  
colSums(tdf) #yay! All good

## hab\_x\_x over\_x\_x poll\_x\_x spsp\_x\_x env\_x\_x demo\_x\_x   
## 30 5 3 17 8 11

########### make network   
  
  
tadjmat <- threatmat  
  
tadjmat[lower.tri(tadjmat)] <- NA  
diag(tadjmat) <- NA  
  
tolist <- tadjmat %>% melt()   
tolist <- tolist[-which(is.na(tolist$value)),]  
  
#nsize <- tolist %>% filter((Var1 == Var2)) #set new vector with diag values  
#tolist <- tolist %>% filter((Var1 != Var2))   
  
#allthreat <- tolist %>% right\_join(nsize, by = c("Var1"))  
#allthreat <- allthreat[,-c(4)]  
  
net <- graph\_from\_data\_frame(d=tolist, directed=T)   
  
  
#plot(net, edge.arrow.size=.2, edge.curved=0,  
  
 # vertex.color="orange", vertex.frame.color="#555555",  
  
 # vertex.label=V(net)$Var1, vertex.label.color="black",  
  
 # vertex.label.cex=.7, vertex.size = net$value.y)  
  
E(net)$arrow.size <- .2 #determine size of arrows (this makes the heads real small so they are more like lines)  
#V(net)$names <- c("a","b","c","d","e","f","g","h") #naming the nodes -- try this later so don't over write what actal names are   
#V(df)$names <- c("hab\_x\_x", "over\_x\_x", "poll\_x\_x", "spsp\_x\_x", "env\_x\_x", "demo\_x\_x")  
node.size<-setNames(c(30,5,3,17,8,11), c("hab\_x\_x", "over\_x\_x", "poll\_x\_x", "spsp\_x\_x", "env\_x\_x", "demo\_x\_x"))  
 #so not setting in the same order   
E(net)$width <- E(net)$value  
  
plot(net, vertex.size=node.size)



#plot(net, edge.arrow.size=0.01,vertex.label.color = "black", vertex.size=node.size)

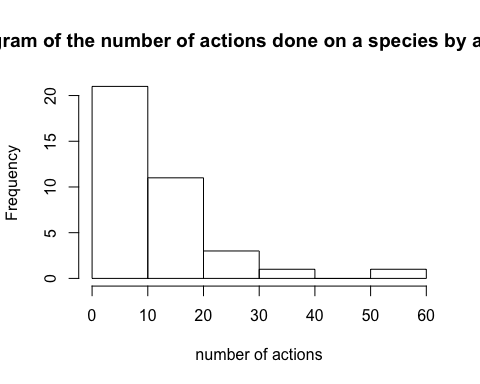
Caption: While not super clear, there is actually a link missing between two points (location changes everytime I print it out) - should be missing btw poll and demo

## c) Does the number of actions relate to species characteristics?

### a. Same as the number of partners analysis, but with number of actions as the response variable

* note to self - this might not knit if vector is produced lower in script so repeating code here to find no actions per species
* decided to try two new predictors (the addition is the number of actions increased by each partner doing that action (might make sense to change this to a weighted value later on))

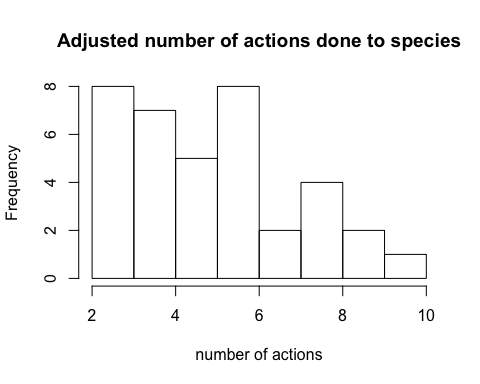
####copied text to find new variable   
alldf <- sdata #change variable name   
  
#species <- select(alldf, Scientific.name, X1..Land.Water.Management ,X2..Species.Management, X3..Awareness.raising,X4..law.enforcement.and.prosecution,X5..livelihood..economic.and.moral.incentrives, X6..Conservation.Design.and.Planning,X7..Legal.and.Policy.frameworks,X8..Research.and.monitoring, X9..Education.and.Training, X10..Institutional.Development,funding) #selecting all relevant columns  
  
  
species <- alldf[,c(2,10:20)] #unable to knit with select so indexed   
  
eachsp <- species %>% group\_by(Scientific.name) %>% summarise\_each(funs(sum))   
  
  
onesandzeros <- eachsp %>% mutate\_if(is.numeric, ~1 \* (. > 0)) #changed all values back to ones and zeros  
  
noactionswithpartner <- eachsp %>% mutate("actpartsum" = rowSums(eachsp[,c(2:12)]))  
  
totalno <- onesandzeros %>% mutate("actionsum" = rowSums(onesandzeros[,c(2:12)])) #added column "actionsum" that took the row sums for each species to give count of how many actions each species receives   
  
hist(noactionswithpartner$actpartsum,  
 xlab ="number of actions",  
 main = "Histogram of the number of actions done on a species by all partners")



#### Caption: calculated taking each species and adding up any actions the partners working on that species were doing. (So if two partners were working on land management and a third on funding, then values would be 2 and 1 for those column). All columns for a species were added by each species (so for the example above the row sum would be three). This graph represents the number of actions being done by all partners per species.  
  
summary(noactionswithpartner$actpartsum)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.00 6.00 9.00 12.05 15.00 60.00

hist(totalno$actionsum,  
 xlab ="number of actions",  
 main = "Adjusted number of actions done to species")

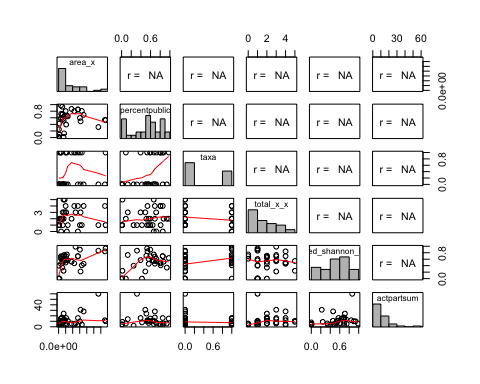


###Caption: The calculation here is slightly different from previous one as it changes all values back to 1s and 0s. (So for previous example instead of columns having a 2 and a 1 they would be converted back to 1 and 1). These 1s and 0s were then added by each species so the rowsum is the total number of actions done per species. No species had all actions done to it. Also no less than 2 actions per species even though there could have been  
  
summary(totalno$actionsum)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 4.000 5.000 5.351 6.000 10.000

changes made to the regression - removed employment variable - adding vert (1 = y, 0 = no) predictor from tdata - Need access to fed land manager

tdata3 <- tdata2  
  
tdata3 <- tdata3[,-c(1:4,7,8,12,14:16)]  
  
#reorder so that predictors are at end of df   
tdata3 <- tdata3[,c(3:7,1,2)]   
  
  
regdf1 <- tdata3[,-c(7)] #so this is with actionpartsum  
  
###############   
  
  
#set up  
PredictorsOnlyPixel <- regdf1[,c(6)]  
PredictAndResponsePixel <- regdf1  
PredictAndResponseGrid <- regdf1  
   
# put histograms on the diagonal panel   
panel.hist <- function (x,...) # define a function that says what we want to plot in the diagonal  
{  
 usr <- par("usr"); on.exit(par(usr)) # not sure what usr is for?  
 par(usr = c(usr[1:2],0,1.5))  
 h <- hist(x, plot = FALSE)  
 breaks <- h$breaks; nB <- length(breaks) # make the hist   
 y <- h$counts; y <- y/max(y)  
 rect(breaks[-nB], 0, breaks[-1], y, col="grey", ...) # defines what the histogram is going to look like  
}  
  
# put correlations on the upper panels,  
panel.cor <- function(x, y, digits=2, prefix="", cex.cor, ...)  
{  
 usr <- par("usr"); on.exit(par(usr))  
 par(usr = c(0, 1, 0, 1))  
 r <- cor(x, y,use="everything")   
 txt <- format(c(r, 0.123456789), digits=digits)[1]  
 prefix <- "r = "  
 rc <- cor.test(x,y,method = c("pearson")) ## calculate pearsons rho for upper grid  
 txt <- paste(prefix,txt,sep="")  
 text(0.5, 0.5, txt, cex = 1)  
}  
  
## plot a correlation matrix plot that uses the functions specified above to say what to plot where  
 ## this was taken directly from website and still not plotting r values for all   
#pairs(PredictAndResponsePixel[1:6], lower.panel=panel.smooth, cex = .8, diag.panel=panel.hist, cex.labels = 1.2, font.labels=2, upper.panel=panel.cor)  
  
pairs(PredictAndResponsePixel,lower.panel = panel.smooth, diag.panel=panel.hist,upper.panel=panel.cor)



#double check cor  
  
cor(regdf1, method = "pearson", use = "complete.obs")

## area\_x percentpublic taxa total\_x\_x fed\_shannon\_h  
## area\_x 1.0000000 0.230686358 0.1294415 0.11470758 0.54897691  
## percentpublic 0.2306864 1.000000000 0.3944271 0.10493003 0.55041962  
## taxa 0.1294415 0.394427060 1.0000000 -0.20716180 0.29269717  
## total\_x\_x 0.1147076 0.104930033 -0.2071618 1.00000000 0.04493631  
## fed\_shannon\_h 0.5489769 0.550419620 0.2926972 0.04493631 1.00000000  
## actpartsum 0.4529198 0.003214887 -0.2611172 0.15073209 0.34119916  
## actpartsum  
## area\_x 0.452919771  
## percentpublic 0.003214887  
## taxa -0.261117163  
## total\_x\_x 0.150732090  
## fed\_shannon\_h 0.341199161  
## actpartsum 1.000000000

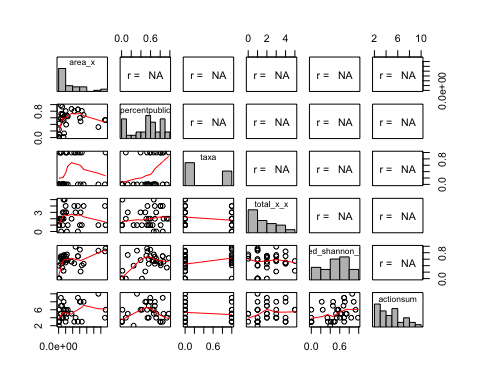
# so still having some issues with getting r values to print out   
  
######### VIFs  
  
vif(lm(actpartsum ~ area\_x +percentpublic + taxa + total\_x\_x + fed\_shannon\_h,data = regdf1))

## area\_x percentpublic taxa total\_x\_x fed\_shannon\_h   
## 1.468117 1.640516 1.291973 1.108227 1.986167

#looks good   
  
  
######## Run the model for actpartsum   
  
lm\_actpartsum <- lm(actpartsum ~ area\_x +percentpublic + taxa + total\_x\_x + fed\_shannon\_h,data = regdf1)  
summary(lm\_actpartsum)

##   
## Call:  
## lm(formula = actpartsum ~ area\_x + percentpublic + taxa + total\_x\_x +   
## fed\_shannon\_h, data = regdf1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -17.929 -3.895 -2.855 2.890 31.703   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.981e+00 4.329e+00 1.613 0.1180   
## area\_x 1.043e-10 5.497e-11 1.898 0.0680 .  
## percentpublic -4.380e+00 7.156e+00 -0.612 0.5455   
## taxa -7.706e+00 3.855e+00 -1.999 0.0554 .  
## total\_x\_x 2.974e-01 1.246e+00 0.239 0.8131   
## fed\_shannon\_h 1.230e+01 8.338e+00 1.475 0.1513   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 9.611 on 28 degrees of freedom  
## (5 observations deleted due to missingness)  
## Multiple R-squared: 0.3598, Adjusted R-squared: 0.2455   
## F-statistic: 3.148 on 5 and 28 DF, p-value: 0.02228

#lose significance of threats  
#plot(lm\_actionsum)  
  
  
############### ############### second regression with new predictors   
  
  
  
regdf2 <- tdata3[,-c(6)] #so this is with actionsum  
  
###############   
  
  
#set up  
PredictorsOnlyPixel <- regdf2[,c(6)]  
PredictAndResponsePixel <- regdf2  
PredictAndResponseGrid <- regdf2  
   
# put histograms on the diagonal panel   
panel.hist <- function (x,...) # define a function that says what we want to plot in the diagonal  
{  
 usr <- par("usr"); on.exit(par(usr)) # not sure what usr is for?  
 par(usr = c(usr[1:2],0,1.5))  
 h <- hist(x, plot = FALSE)  
 breaks <- h$breaks; nB <- length(breaks) # make the hist   
 y <- h$counts; y <- y/max(y)  
 rect(breaks[-nB], 0, breaks[-1], y, col="grey", ...) # defines what the histogram is going to look like  
}  
  
# put correlations on the upper panels,  
panel.cor <- function(x, y, digits=2, prefix="", cex.cor, ...)  
{  
 usr <- par("usr"); on.exit(par(usr))  
 par(usr = c(0, 1, 0, 1))  
 r <- cor(x, y,use="everything")   
 txt <- format(c(r, 0.123456789), digits=digits)[1]  
 prefix <- "r = "  
 rc <- cor.test(x,y,method = c("pearson")) ## calculate pearsons rho for upper grid  
 txt <- paste(prefix,txt,sep="")  
 text(0.5, 0.5, txt, cex = 1)  
}  
  
## plot a correlation matrix plot that uses the functions specified above to say what to plot where  
 ## this was taken directly from website and still not plotting r values for all   
#pairs(PredictAndResponsePixel[1:6], lower.panel=panel.smooth, cex = .8, diag.panel=panel.hist, cex.labels = 1.2, font.labels=2, upper.panel=panel.cor)  
  
pairs(PredictAndResponsePixel,lower.panel = panel.smooth, diag.panel=panel.hist,upper.panel=panel.cor)



#double check cor  
  
cor(regdf2, method = "pearson", use = "complete.obs")

## area\_x percentpublic taxa total\_x\_x fed\_shannon\_h  
## area\_x 1.0000000 0.2306864 0.1294415 0.11470758 0.54897691  
## percentpublic 0.2306864 1.0000000 0.3944271 0.10493003 0.55041962  
## taxa 0.1294415 0.3944271 1.0000000 -0.20716180 0.29269717  
## total\_x\_x 0.1147076 0.1049300 -0.2071618 1.00000000 0.04493631  
## fed\_shannon\_h 0.5489769 0.5504196 0.2926972 0.04493631 1.00000000  
## actionsum 0.3050782 0.1663274 -0.1214172 0.19504162 0.44355778  
## actionsum  
## area\_x 0.3050782  
## percentpublic 0.1663274  
## taxa -0.1214172  
## total\_x\_x 0.1950416  
## fed\_shannon\_h 0.4435578  
## actionsum 1.0000000

# so still having some issues with getting r values to print out   
  
######### VIFs  
  
vif(lm(actionsum ~ area\_x +percentpublic + taxa + total\_x\_x + fed\_shannon\_h,data = regdf2))

## area\_x percentpublic taxa total\_x\_x fed\_shannon\_h   
## 1.468117 1.640516 1.291973 1.108227 1.986167

#looks good   
  
######## Run the model for actpartsum   
  
lm\_actionsum <- lm(actionsum ~ area\_x +percentpublic + taxa + total\_x\_x + fed\_shannon\_h,data = regdf2)  
summary(lm\_actionsum)

##   
## Call:  
## lm(formula = actionsum ~ area\_x + percentpublic + taxa + total\_x\_x +   
## fed\_shannon\_h, data = regdf2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.7228 -1.1771 -0.1263 1.4118 3.6488   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.554e+00 8.809e-01 4.035 0.000383 \*\*\*  
## area\_x 3.329e-12 1.118e-11 0.298 0.768197   
## percentpublic -2.978e-01 1.456e+00 -0.205 0.839413   
## taxa -1.004e+00 7.845e-01 -1.280 0.211220   
## total\_x\_x 1.841e-01 2.536e-01 0.726 0.473865   
## fed\_shannon\_h 3.744e+00 1.697e+00 2.207 0.035687 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.956 on 28 degrees of freedom  
## (5 observations deleted due to missingness)  
## Multiple R-squared: 0.2834, Adjusted R-squared: 0.1555   
## F-statistic: 2.215 on 5 and 28 DF, p-value: 0.08087

#lose significance of threats  
#plot(lm\_actionsum)

Old regression

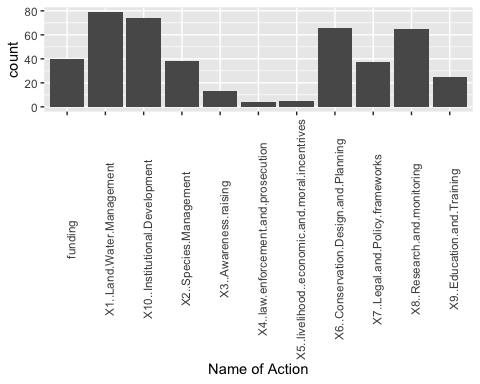
## d) How are actions distributed across partners?

From a planning perspective, we care about actions because decision makers need to know a) what needs to be done for species recovery, and b) which actors can best do it. Thus, we want to know “what actions are different partners doing and how do they contribute to the sum total of what needs to be done for different species?”

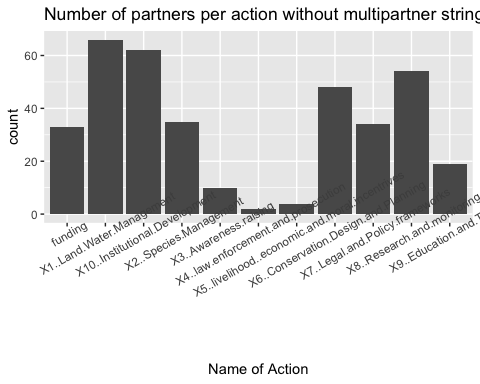
### Data Exploration

### a. Frequency distribution (histogram) of how many partners are participating in each type of action

df <- as\_tibble(sdata)  
df[,c(10:20)] <- sapply(df[ ,c(10:20)], as.numeric)  
  
#rowSums((code2[,c(11:21)]))  
  
colsum <- (as.data.frame(colSums(df[,c(10:20)])) #creating dataframe so can plot  
 %>% rownames\_to\_column()) #making sure that dataframe has rownames to set as x and y   
   
   
colsum <- colsum %>% rename(count = `colSums(df[, c(10:20)])`) #renaming column produced by colsums  
   
  
#ggplot(colsum) + geom\_point(mapping = aes(x = rowname, y = count))  
  
ggplot(colsum) + geom\_bar(mapping = aes(x = rowname, y = count), stat = "identity")+ theme(axis.text.x = element\_text(angle = 90)) + scale\_x\_discrete(name ="Name of Action")



#same graph, just expanded with text shifted  
#ggplot(colsum) + geom\_bar(mapping = aes(x = rowname, y = count), stat = "identity")+ theme(axis.text.x = element\_text(angle = 30)) + scale\_x\_discrete(name ="Name of Action") + ggtitle("Number of partners per action")  
  
#### caption: The number of partners participating in each action (not this is including multi species partner strings as an additional count  
  
#compared to noM   
  
noM[,c(10:20)] <- sapply(noM[ ,c(10:20)], as.numeric)  
  
colsum\_noM <- (as.data.frame(colSums(noM[,c(10:20)])) #creating dataframe so can plot  
 %>% rownames\_to\_column()) #making sure that dataframe has rownames to set as x and y   
   
   
colsum\_noM <- colsum\_noM %>% rename(count = `colSums(noM[, c(10:20)])`) #renaming column produced by colsums  
   
  
#ggplot(colsum) + geom\_point(mapping = aes(x = rowname, y = count))  
  
ggplot(colsum\_noM) + geom\_bar(mapping = aes(x = rowname, y = count), stat = "identity")+ theme(axis.text.x = element\_text(angle = 30)) + scale\_x\_discrete(name ="Name of Action") +ggtitle("Number of partners per action without multipartner strings")



#made with noM so doesn't have partner strings included in the calculation

Caption: This is the graph without the additional partner strings which seems to be a similar distribution

### b. For each species

#### - i) For each species: how many different actions are applied (and how many of each action).

#### - ii) Action richness and diversity for each species?

Source - <http://www.colby.edu/biology/BI131/Lab/Lab09CalcBiodivers.pdf> - D, the Menhinick’s index = s/sqrt(N) - So higher number = more richness

#done with summed actions  
noactionswithpartner <- eachsp %>% mutate("actpartsum" = rowSums(eachsp[,c(2:12)]))  
  
richness1 <- noactionswithpartner %>%  
 mutate\_at(vars(actpartsum), funs(./sqrt(sum(.)))) ##D = actionsum colum (unsure how to change the name)  
#quick check of numbers   
##colSums(noactionswithpartner[,13] )  
##15/sqrt(446)  
#math worked!  
  
#done with all partner actions  
totalno <- onesandzeros %>% mutate("actionsum" = rowSums(onesandzeros[,c(2:12)])) #added column "actionsum" that   
richness2 <- totalno %>%  
 mutate\_at(vars(actionsum), funs(./sqrt(sum(.)))) ##D = actionsum colum (unsure how to change the name)

Other options for richness: Margalef’s index is the number of species (n) minus 1 divided by the natural logarithm of the total number of individuals (N). D=n−1/lnN

Shannon Diversity Index Calc

#done with summed actions  
noactionswithpartner <- eachsp %>% mutate("actpartsum" = rowSums(eachsp[,c(2:12)]))  
pi <- noactionswithpartner %>%  
 mutate\_at(vars(actpartsum), funs(./sum(.))) ##pi = actionsum colum (unsure how to change the name)  
lnpi <- pi %>% mutate("log\_pi" = log(actpartsum))  
pilnpi <- lnpi %>% mutate("pilnpi" = log\_pi\*actpartsum)  
  
H <- -1 \* colSums(pilnpi[,15])  
H

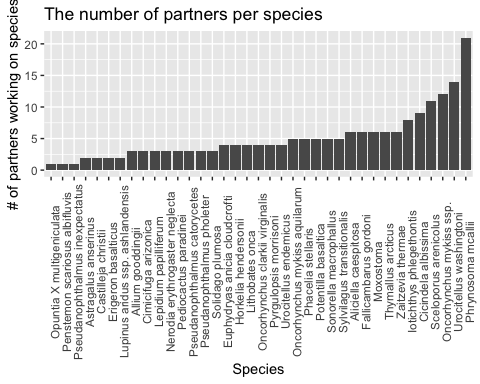
## pilnpi   
## 3.315919

#done with all partner actions  
totalno <- onesandzeros %>% mutate("actionsum" = rowSums(onesandzeros[,c(2:12)])) #added column "actionsum" that   
  
pi2 <- totalno %>%  
 mutate\_at(vars(actionsum), funs(./sum(.))) ##pi = actionsum colum (unsure how to change the name)  
lnpi2 <- pi2 %>% mutate("log\_pi2" = log(actionsum))  
pilnpi2 <- lnpi2 %>% mutate("pilnpi" = log\_pi2\*actionsum)  
  
H2 <- -1 \* colSums(pilnpi2[,15])  
H2

## pilnpi   
## 3.541426

#### - iii) How many partners does each species have and do partners conduct the same or different actions from each other

nopartners <- (alldf %>% group\_by(Scientific.name) #selecting each species  
%>% distinct(partner.in.agreement) #count how many partners are distinct   
%>% summarise(n())) # Count the number of distinct   
  
#kable(nopartners)  
  
nopartners <- rename(nopartners, tally = 'n()')  
  
#reset as bar plot with each species name as value on x-axis   
  
ordered <- nopartners %>% arrange(tally)  
#ordered$Scientific.name  
  
ggplot(ordered) + geom\_bar(mapping = aes(x = reorder(Scientific.name, tally), y = tally), stat = "identity") + theme(axis.text.x = element\_text(angle = 90)) + scale\_x\_discrete(name ="Species") + scale\_y\_continuous(name ="# of partners working on species") + ggtitle("The number of partners per species")

 Caption: The number of distinct partners working on a species

- Can find number of partners but unsure how to answer second part of Q  
 - "Do partners conduct same or different actions.." for same partner working on each species or across different species?

Dissimilarity <https://www.researchgate.net/post/What_is_the_difference_between_Bray-Curtis_Similarity_Sorensen_Distance_and_Bray-Curtis_Index> - So Bray-Curtis index seems most relevant here [https://www.statisticshowto.com/bray-curtis-dissimilarity/#](https://www.statisticshowto.com/bray-curtis-dissimilarity/):~:text=The%20Bray%2DCurtis%20dissimilarity%20is,don’t%20share%20any%20species.&text=But%20the%20Bray%20Curtis%20index,%C2%B7%20100%2C%20or%2061%25. - for two sites i and j BCij = 1 - ((2Cij)/(Si + Sj)) Where: - i & j are the two sites, - Si is the total number of specimens counted on site i, - Sj is the total number of specimens counted on site j, - Cij is the sum of only the lesser counts for each species found in both sites. To compare multiple sites, seems like need to put together a matrix and then eyeball which values differ? <http://www.econ.upf.edu/~michael/stanford/maeb5.pdf>

#### - iv) How many threats does each species face

- information found in regression prep script

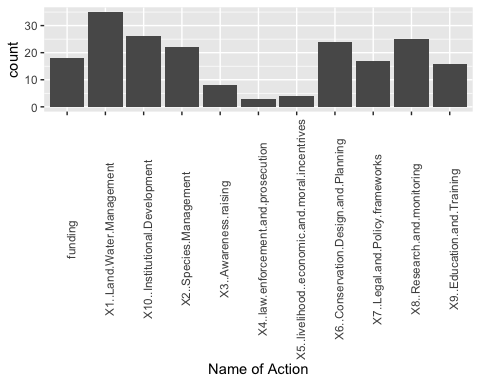
### c. Graphs

#### - i) For each action – how many species receive that action

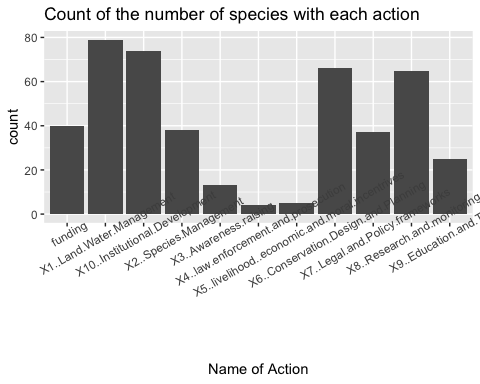
# take column sum of onesandzeros dataset created above (chunk b. For each species - i.. )  
actionsums <- colSums(onesandzeros[,c(2:12)]) #add all species for which that action happened  
  
actionsums <- (as.data.frame(actionsums) #had to change to df  
 %>% rownames\_to\_column() #moving rownames to columns  
 %>% rename(count = actionsums)) #renaming column produced by colsums  
  
kable(actionsums)

|  |  |
| --- | --- |
| rowname | count |
| X1..Land.Water.Management | 35 |
| X2..Species.Management | 22 |
| X3..Awareness.raising | 8 |
| X4..law.enforcement.and.prosecution | 3 |
| X5..livelihood..economic.and.moral.incentrives | 4 |
| X6..Conservation.Design.and.Planning | 24 |
| X7..Legal.and.Policy.frameworks | 17 |
| X8..Research.and.monitoring | 25 |
| X9..Education.and.Training | 16 |
| X10..Institutional.Development | 26 |
| funding | 18 |

ggplot(actionsums) + geom\_bar(mapping = aes(x = rowname, y = count), stat = "identity") + theme(axis.text.x = element\_text(angle = 90)) + scale\_x\_discrete(name ="Name of Action")

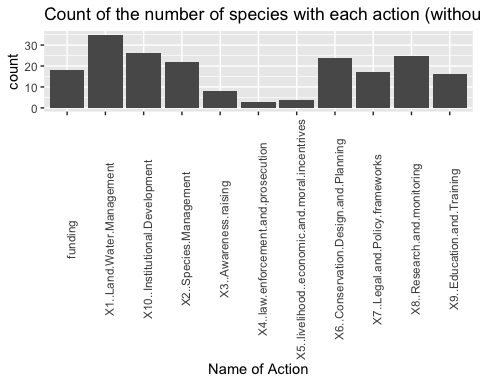


#same graph, just expanded with text shifted  
ggplot(colsum) + geom\_bar(mapping = aes(x = rowname, y = count), stat = "identity")+ theme(axis.text.x = element\_text(angle = 30)) + scale\_x\_discrete(name ="Name of Action") + ggtitle("Count of the number of species with each action")

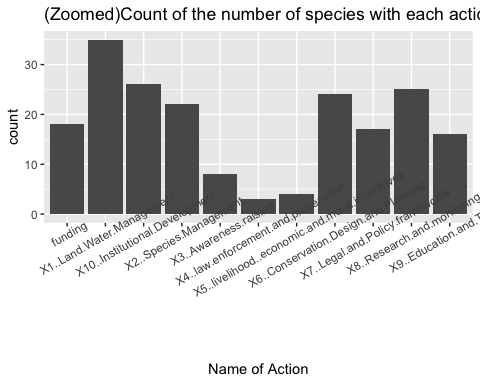
 Caption: Number of species that are part of each action

#### - ii) For each action – how many times is it applied total (counting each speciesXpartner separately) information contained in csum output

## for all speciesXpartners csum but this includes some repetition ("partners" which are actually multiple partners working on same action)  
csum <- (colSums(sdata[,c(10:20)]))  
  
#check csum  
check <- colSums(eachsp[,c(2:12)])  
#yes get the same values   
  
#redo csum calculation and take out = M  
  
##  
noM <- sdata[-which(sdata$type.of.partners == "M"),]  
newcsum <- (colSums(noM[,c(10:20)]))  
  
newcsum <- as.data.frame(actionsums) #had to change to df  
   
ggplot(newcsum) + geom\_bar(mapping = aes(x = rowname, y = count), stat = "identity") + theme(axis.text.x = element\_text(angle = 90)) + scale\_x\_discrete(name ="Name of Action") + ggtitle("Count of the number of species with each action (without multiparnter strings)")



#same graph, just expanded with text shifted  
ggplot(newcsum) + geom\_bar(mapping = aes(x = rowname, y = count), stat = "identity")+ theme(axis.text.x = element\_text(angle = 30)) + scale\_x\_discrete(name ="Name of Action")+ ggtitle("(Zoomed)Count of the number of species with each action (without multiparnter strings)")

 Caption: Same as above but not counting multipartner strings as partners

#### - iii )For each action – how many partners apply that action to at least 1 species

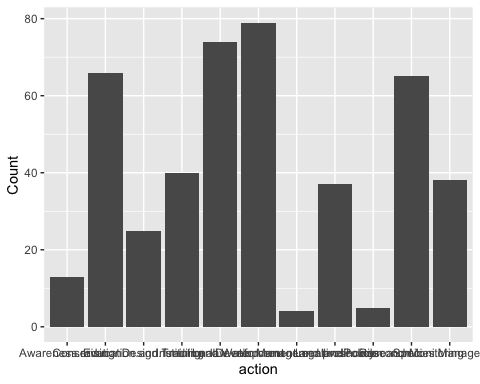
partners <- select(alldf, partner.in.agreement, Scientific.name, X1..Land.Water.Management ,X2..Species.Management, X3..Awareness.raising,X4..law.enforcement.and.prosecution,X5..livelihood..economic.and.moral.incentrives, X6..Conservation.Design.and.Planning,X7..Legal.and.Policy.frameworks,X8..Research.and.monitoring, X9..Education.and.Training, X10..Institutional.Development,funding) #selecting all relevant columns  
  
class(partners$funding)

## [1] "numeric"

#so summarise will only accept 1 value per group so going to try and do for each group   
#there is definetly a more elegant/better way to do this but this works  
x1 <- partners %>% group\_by(partner.in.agreement) %>% summarise(sum(partners[,c(3)])) #84  
x2 <- partners %>% group\_by(partner.in.agreement) %>% summarise(sum(partners[,c(4)])) #45  
x3 <- partners %>% group\_by(partner.in.agreement) %>% summarise(sum(partners[,c(5)])) #13  
x4 <- partners %>% group\_by(partner.in.agreement) %>% summarise(sum(partners[,c(6)])) #4  
x5 <- partners %>% group\_by(partner.in.agreement) %>% summarise(sum(partners[,c(7)])) #5   
x6 <- partners %>% group\_by(partner.in.agreement) %>% summarise(sum(partners[,c(8)])) #68  
x7 <- partners %>% group\_by(partner.in.agreement) %>% summarise(sum(partners[,c(9)])) #38  
x8 <- partners %>% group\_by(partner.in.agreement) %>% summarise(sum(partners[,c(10)])) #72  
x9 <- partners %>% group\_by(partner.in.agreement) %>% summarise(sum(partners[,c(11)])) #26  
x10 <- partners %>% group\_by(partner.in.agreement) %>% summarise(sum(partners[,c(12)])) #75  
x11 <- partners %>% group\_by(partner.in.agreement) %>% summarise(sum(partners[,c(13)])) #41  
## (for each row, value is to the right)  
  
partnerapptosp <- as.data.frame(c(x1, x2, x3, x4,x5, x6,x7,x8,x9,x10,x11))  
  
#condense into one df   
partnerapptosp <- partnerapptosp[,-c(1,3,5,7,9,11,13,15,17,19,21)] #remove relplicated columns (and don't need to know partner ids)  
#colnames(partnerapptosp)  
setnames(partnerapptosp, old = c('sum.partners...c.3...', 'sum.partners...c.4...', 'sum.partners...c.5...', 'sum.partners...c.6...', 'sum.partners...c.7...', 'sum.partners...c.8...', 'sum.partners...c.9...', 'sum.partners...c.10...', 'sum.partners...c.11...', 'sum.partners...c.12..', 'sum.partners...c.13..'),skip\_absent=TRUE, new = c('Land.Water.Management' , 'Species.Management', 'Awareness.raising','law.enforcement.and.prosecution', 'livelihoodeconomic', 'ConservationDesign', 'LegalandPolicy', 'ResearchMonitoring', 'Education.and.Training', 'InstitutionalDevelopment','funding')) #for some reason this wouldn't over write c12 or c13 (institutional development and funding) so setting manually below   
  
partnerapptosp <- partnerapptosp %>% rename(InstitutionalDevelopment = 'sum.partners...c.12...')  
partnerapptosp <- partnerapptosp %>% rename(funding = 'sum.partners...c.13...')  
  
partnerapptosp <- partnerapptosp[2,] #only need to select 1 row   
action <- t(partnerapptosp)   
#trying not to loose rownames when convert  
  
actiondf <- data.frame(action = row.names(action),action) #changed to df and set column names   
rownames(actiondf) <- NULL  
  
kable(actiondf)

|  |  |
| --- | --- |
| action | X2 |
| Land.Water.Management | 79 |
| Species.Management | 38 |
| Awareness.raising | 13 |
| law.enforcement.and.prosecution | 4 |
| livelihoodeconomic | 5 |
| ConservationDesign | 66 |
| LegalandPolicy | 37 |
| ResearchMonitoring | 65 |
| Education.and.Training | 25 |
| InstitutionalDevelopment | 74 |
| funding | 40 |

ggplot(actiondf) + geom\_bar(mapping = aes(x = action, y = X2), stat = "identity")+ scale\_y\_continuous(name ="Count")



#same graph with names rotated   
ggplot(actiondf) + geom\_bar(mapping = aes(x = action, y = X2), stat = "identity") + theme(axis.text.x = element\_text(angle = 90)) + scale\_x\_discrete(name ="Name of Action") + scale\_y\_continuous(name ="Count") + ggtitle("Number of partners that apply an action to at least one species")



#col\_list <- partners[,c(3:13)]  
#for(coln in col\_list){  
# partners %>% group\_by(partner.in.agreement) %>% summarise(sum(partners[,coln]))  
#} ## Got error

Caption: Number of partners that apply an action to at least one species. This is a modification of previous chart).

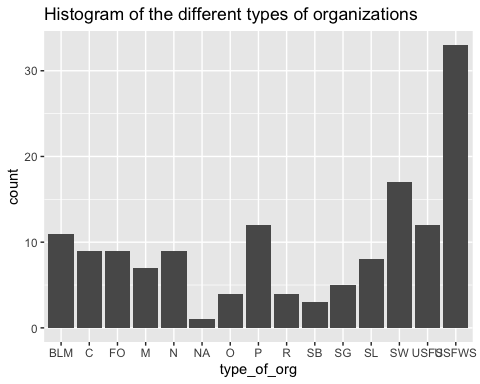
### d. Thoughts about typologies/groupings

#### - i) Can we group partners into types based on the actions that they do? (numbers and type of actions?)

#### - ii) Can we group partners base on type of organization (fed agency, wildlife agency, private landowner, ngo, researchers?)

#putting draft of excel sheet in here straight from wd beacuse will need to update  
  
orgtyp <- read.csv(paste0("/usr/local/bin/store/partner\_rff/data/tableofPandAmanedited.csv"), stringsAsFactors = FALSE, na = "") #issue was that "" is coming up instead of NA  
  
#filter out all rows that don't have values   
orgtyp <- orgtyp[-which(is.na(orgtyp$type\_of\_org)),] #this isn't doing anything, lets try again  
  
ggplot(orgtyp) + geom\_histogram(mapping = aes(x = type\_of\_org), stat = "count") + ggtitle("Histogram of the different types of organizations")

## Warning: Ignoring unknown parameters: binwidth, bins, pad

 Key (note, currently unsure how to reorder but will keep trying to this may change) \* BLM = Bureau of Land Management \* C = Cooporation \* FO = Federal Other \* M = Military \* N = NGO \* NA \* O = Other \* P = Private landowner \* R = Reserach \* SB = State Both (land and wildlife) \* SG = State Governmnet \* SL = State Land Managemetn \* SW = State Wildlife \* USFS = US Forest Service \* USFWS = US Fish and Wildlife Service

#### - iii) Can we group species based on the sets of threats that they face?

insert network here?

#### - iv) East coast v west coast (makes most sense to focus on west coast split as there are more species here)

primary exploration attached below in spatial analysis section

#### - v) Taxa

#code taken from regression section, just need to pull out the taxa variable  
  
  
modtdata <- tdata  
newregdata <- RegData #doesn't work because nothing to join by   
  
#first going to join totalno and noactionswithpartner  
  
predictors <- noactionswithpartner %>% full\_join(totalno, by = "Scientific.name") #join two new predictors together  
#predictors1 <- predictors %>% select(Scientific.name, actionsum, actpartsum)  
predictors1 <- predictors[,c(1,13,25)] #unable to knit with select so indexed   
predictors2 <- rename(predictors1, scientific\_name = Scientific.name) #renamed predictors vector so would have something to join by   
  
#so if select values that don't change from tdata with somes that weren't modified in RegData, should be able to join  
#modtdata <- modtdata %>% select(scientific\_name, total\_x\_x, area\_x) #select relevant predictors from tdata   
modtdata <- modtdata[,c(3,19,26,32)] #Added column 19 for Plants, Inverts and verts   
newregdata <- newregdata %>% left\_join(modtdata, by = c("total\_x\_x", "area\_x")) #note 1 name is missing, but will see if that is an issue based on names in action dataset   
  
# now going to join with predictors   
  
regdf <- predictors2 %>% left\_join(newregdata, by = "scientific\_name")  
#so issue now isn't the missing name, but the number of species that have missing na values   
  
#now sort through 1s and 0s for plants, verts and inverts   
  
ggplot(regdf) + geom\_histogram(mapping = aes(x = vertebrate\_invertebrate\_plant\_x\_x), stat = "count")+ ggtitle("Histogram of the taxa: vertebrate\_invertebrate\_plant")

## Warning: Ignoring unknown parameters: binwidth, bins, pad



## Questions we want to answer

### e. How are total actions distributed across partners (e.g., what proportion of land management is enacted by FWS?)

#### Annabelle Spatial analysis

Background to analysis

* map of centers and ranges (load in pdf) Map of all species Informative caption Map: All species, for which I have partner data, have their ranges colored in orange and the average center of their range indicated with a green dot. Ranges may be overlapping and were calculated at a county scale.

west coast v east coast #### variable okay produced in RegressionPredictoExploration range chunk removing all 0s columns not working

combo <- read.csv(paste0("/usr/local/bin/store/partner\_rff/OutputAndInputsForThesis/combodf.csv"))  
center <- read\_excel(paste0(DataSource, "/mean\_center\_range\_subsetofprecluded\_with\_partnerdata.xlsx"))  
  
#### correct values here   
ugh <- combo  
#ugh <- ugh[,c(2,6)] #had to re calibrate when reloaded the data  
okay <- ugh %>% left\_join(center, by = "scientific\_name")  
#so have slight issue where Chorizanthe parryi var fernandina is splet Chorizanthe parryi var fernandina and Chorizanthe parryi var. fernandina (period after var.)  
#okay[53,2] <- 2  
  
okay <- okay[-which(is.na(okay$XCoord)),]  
  
  
#plot(okay$x, okay$y)  
  
## divide by xcoord value to make east coast v westcoast column   
  
okay <- okay %>% mutate(geo = x < -95)  
summary(okay$geo) #adams cave beetle have same point which is why getting 10 but only see 9 points  
okay[which(okay$geo == FALSE),9] <- 0  
okay[which(okay$geo == TRUE),9] <- 1  
## true or 1 means species is westcoast  
  
west <- okay %>% filter(geo == TRUE)  
fullwest <- west %>% left\_join(full)  
  
east <- okay %>% filter(geo == FALSE)  
fulleast <- east %>% left\_join(full)  
  
west <- okay %>% filter(geo == TRUE) #### \*\*\*need to find what geo is - I think this was where made e/w divide but idk where in previous code that was   
west <- west [,c(1,3)]  
west <- west %>% left\_join(PartnerData, by = "scientific\_name")  
nzwest <- west[,-c(1:4)][,colSums(west[,-c(1:4)])>=1] #removed columns with all 0s  
dim(west)  
nzwest <- as.matrix(nzwest) #prep for making adj matrix  
wajmat <- t(nzwest) %\*% nzwest #new adjmat  
tidyweajmat <- as\_tibble(wajmat, rownames = "id")  
  
  
east <- okay %>% filter(geo == FALSE)  
east <- east [,c(1,3)]  
east <- east %>% left\_join(PartnerData, by = "scientific\_name")   
nzeast <- east[,-c(1:4)][,colSums(east[,-c(1:4)])>=1]  
dim(nzeast)  
nzeast <- as.matrix(nzeast) #prep for making adj matrix  
eajmat <- t(nzeast) %\*% nzeast #new adjmat  
#need to change from matrix to tidyr   
#need to change to tidyverse for later code  
tidyeajmat <- as\_tibble(eajmat, rownames = "id")  
  
  
#this is just the diagonal matrix from adj matrix   
  
  
graphme <- wajmat  
graphme <- graphme %>% melt()  
graphme <- graphme %>% filter(Var1 == Var2)  
graphme <- graphme[,-c(1)] #removed duplicate names  
#all partners  
ggplot(data = graphme) + geom\_col(mapping = aes(x=Var2, y=value))  
g\_one <- graphme %>% filter(value >1)  
ggplot(data = g\_one, mapping = aes(x=Var2, y=value)) + geom\_col() + theme(axis.text.x = element\_text(angle = 90)) + scale\_y\_continuous(name ="Number of Species", limits=c(0,45)) + scale\_x\_discrete(name ="Name of Partner")

* variograms

combo <- read.csv(paste0("/usr/local/bin/store/partner\_rff/OutputAndInputsForThesis/combodf.csv"))  
 center <- read\_excel(paste0(DataSource, "/mean\_center\_range\_subsetofprecluded\_with\_partnerdata.xlsx"))  
  
   
#make changes so can join   
combo$common\_name[10] <- "CHRISTS PAINTBRUSH" #the accent on the o gets messed up in this species so set to same

## Warning in `[<-.factor`(`\*tmp\*`, 10, value = structure(c(NA, 18L, 31L,  
## 1L, : invalid factor level, NA generated

center$common\_name[9] <- "CHRISTS PAINTBRUSH"  
combo$common\_name[1] <- "San Fernando Valley spineflower" #loaded without common name

## Warning in `[<-.factor`(`\*tmp\*`, 1, value = structure(c(NA, 18L, 31L, 1L, :  
## invalid factor level, NA generated

#now join the two df and drop species with data missing   
  
full <- center %>% left\_join(combo, by = "common\_name")

## Warning: Column `common\_name` joining character vector and factor, coercing  
## into character vector

#remove ones without partner data  
  
full <- full[-which(is.na(full$total)),]  
  
#remove extra scientific name column   
full <- full[,-c(7,8)]  
  
#note, only one species was removed due to inadequate information about geographic range (artic greyling)

Generate variogram with raw data - L10 was unable to run as part of markdown script (previously in .r so that might be source of problem) current error is “object logp not found”

Head <- full #made in chunk above  
Head <- Head[,c(5,6,9)] #select only necessary columns   
Head <- as.data.frame(Head)  
#  
Head$logp = log(Head$total)  
#  
# remove NA values   
### Head <- Head[-which(is.na(Head$total)),]  
#  
library(sp)  
#data(meuse)  
#  
coordinates(Head) = c("x", "y")  
#  
library(RColorBrewer)  
library(classInt)  
library(gstat)  
 # hscat(logp~1,data=Head, breaks=c(0,5,10,15,20,25,30,200))  
hscat(logp~1,data=Head, breaks=c(0,1,2,3,5,10,15,20,25,30,35)) ## axes are wrong? Plots look square..   
#  
# How many points are there?  
#  
length(Head$logp)  
#  
# How many pairs of points?  
#  
choose(49,2)  
#  
# Scatter plot of squared differences  
#  
plot(variogram(logp~1,Head,cloud=T))  
#  
# Differences averaged for distance increments  
# to show empirical variogram  
#  
logzinc.vario = variogram(logp~1,Head)  
#  
# Take a look at the contents of logzinc.vario  
#  
logzinc.vario  
#  
# Now plot the empirical variogram  
# and identify the number of pairs of points  
# going into each estimate  
#  
plot(gamma~dist,data=logzinc.vario,xlim=c(0,30))  
#text(logzinc.vario$dist+100,logzinc.vario$gamma,logzinc.vario$np)  
#  
# Sometimes outliers can have a big effect  
# Note that for this data set one outlier will go into the   
# computation of the squared difference 155 times.  
# Cressie suggests a robust measure of the variogram.  
#  
logzinc.vario.robust = variogram(logp~1,Head,cressie=T)  
#  
par(mfrow=c(1,2))  
plot(gamma~dist,data=logzinc.vario,type="b")  
#title("Classical Variogram")  
plot(gamma~dist,data=logzinc.vario.robust,type="b")   
#title("Robust Variogram")  
par(mfrow=c(1,1))  
#  
#  
# \*\*\* this part of the code doesn't work right yet.. don't think the lines would be super relevant for this data anyway?   
#  
#  
# Do a permutation test by shifting the values associated with  
# each location around at random (i.e. permute them)  
# and compute the variogram and plot it each time this is done.  
#  
# First plot the original variogram estimate  
#  
plot(gamma~dist,data=logzinc.vario,type="n")  
#  
# Now do 100 permutations and plot the resulting "pure nugget" effect  
# Where our variogram lies outside this envelope show significant  
# correlation.  
#  
x=data.frame(Head)$x  
y=data.frame(Head)$y  
logzinc=Head$logp  
id0=seq(length(Head$logp))  
for(i in 1:100)  
{  
 id = sample(id0)  
 hold.perm = data.frame(x=x,y=y,logp=logp[id])  
 coordinates(hold.perm)=c("x","y")  
 hold.vario = variogram(logp~1,hold.perm)  
 lines(gamma~dist,data=hold.vario,col=2)  
}  
lines(gamma~dist,data=logzinc.vario,col=1,type="b")  
#  
# Examining for Anisotropy  
#  
logzinc.vario.dir = variogram(logp~1,Head,alpha=c(0,45,90,135))  
plot(logzinc.vario.dir)

* regression exploration
* Pat Sullivan’s notes - have c&p into word document

Putting together map of species in salafsky subset

center <- read\_excel(paste0(DataSource, "/mean\_center\_range\_subsetofprecluded\_with\_partnerdata.xlsx"))  
sub <- sdata %>% select(Scientific.name) %>% rename("scientific\_name" = Scientific.name)  
  
  
sub <- sub %>% left\_join(center, by = c("scientific\_name"))  
  
plot(sub$YCoord ~ sub$XCoord)

