Data Exploration Salafsky edited

Annabelle

June 15, 2020

added edited to file name to keep seperate from first round of networks if that is something want to go back to instead of knitting over (could be fine because had doc that was sent to Gwen)

note on scaling - can over ride the default with “rescale = F” in final plot command can manually manipulate by setting which every layout wanted to l and then multiplying See multi for example of this All layouts are currently set to circle to increase readability

* need to fix scaling and color

## Set up file structure

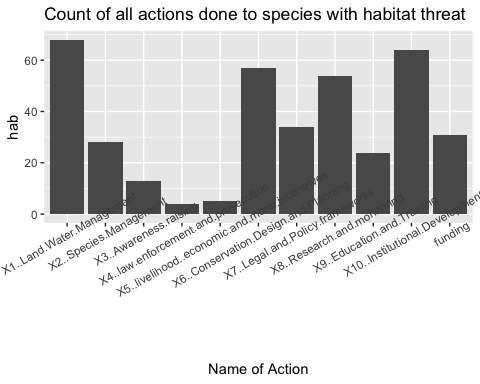
## load in different datasets

Group bar chart species-threat - note, M not removed

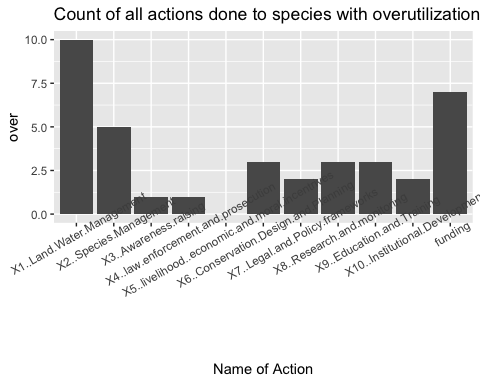
#doesn't like that I'm not specifying melt variable names (even though there is only one) so removing warnings from knitting   
alldf <- sdata #change variable name   
  
species <- alldf[,c(3,11:21)] #note, had to change index because some extra column (X1) had added itself   
  
eachsp <- species %>% 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.

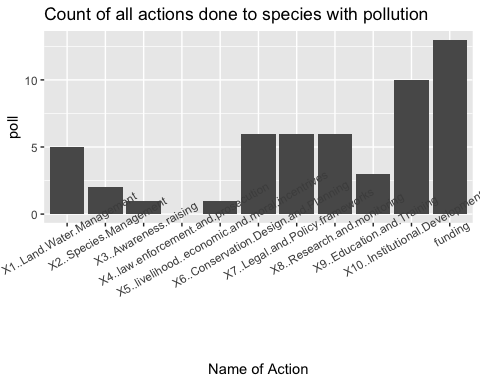
allspindf <- eachsp  
new <- tdata #need to join with tdata (renamed here)  
###new$scientific\_name  
#select relevant columns  
tthreats <- select(new, scientific\_name, hab\_x\_x:threats\_addressed\_by\_conservation\_x\_x)  
  
###tthreats$scientific\_name  
#need to join with tdata   
#to do so need 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)  
  
  
############### make graph for each threat  
  
#habitat  
hab <- threats %>% filter\_at(vars(hab\_x\_x), any\_vars(. %in% c(1)))   
hab <- hab[,-c(1,13:20)]  
hab\_m <- hab %>% melt(value.name="hab")  
ggplot(hab\_m, aes(x=variable, y=hab)) + geom\_bar(stat = "identity") + theme(axis.text.x = element\_text(angle = 30)) + ggtitle("Count of all actions done to species with habitat threat") + scale\_x\_discrete(name ="Name of Action")



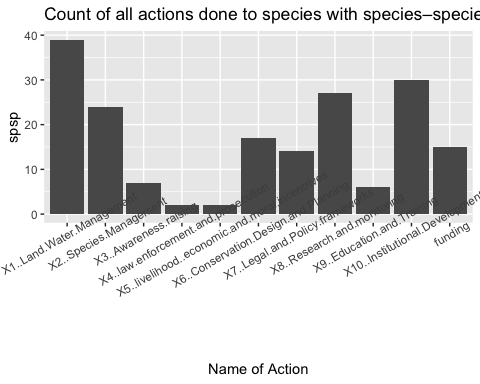
#overutilization  
over <- threats %>% filter\_at(vars(over\_x\_x), any\_vars(. %in% c(1)))   
over <- over[,-c(1,13:20)]  
over\_m <- over %>% melt(value.name="over")  
ggplot(over\_m, aes(x=variable, y=over)) + geom\_bar(stat = "identity") + theme(axis.text.x = element\_text(angle = 30)) + ggtitle("Count of all actions done to species with overutilization") + scale\_x\_discrete(name ="Name of Action")



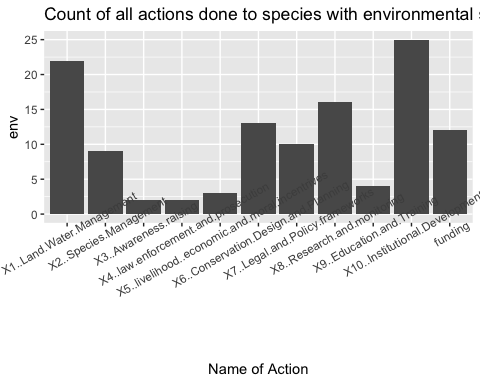
#pollution  
poll <- threats %>% filter\_at(vars(poll\_x\_x), any\_vars(. %in% c(1)))   
poll <- poll[,-c(1,13:20)]  
poll\_m <- poll %>% melt(value.name="poll")  
ggplot(poll\_m, aes(x=variable, y=poll)) + geom\_bar(stat = "identity") + theme(axis.text.x = element\_text(angle = 30)) + ggtitle("Count of all actions done to species with pollution") + scale\_x\_discrete(name ="Name of Action")



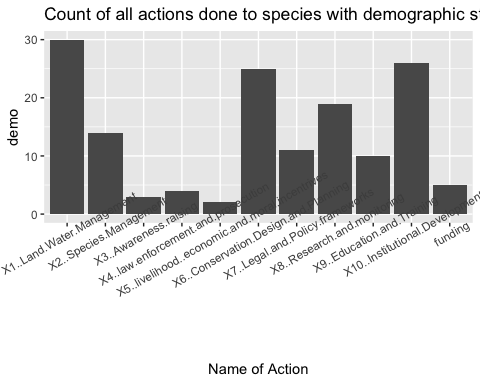
#species–species  
spsp <- threats %>% filter\_at(vars(spsp\_x\_x), any\_vars(. %in% c(1)))   
spsp <- spsp[,-c(1,13:20)]  
spsp\_m <- spsp %>% melt(value.name="spsp")  
ggplot(spsp\_m, aes(x=variable, y=spsp)) + geom\_bar(stat = "identity") + theme(axis.text.x = element\_text(angle = 30)) + ggtitle("Count of all actions done to species with species–species threat") + scale\_x\_discrete(name ="Name of Action")



#environmental stochasticity  
env <- threats %>% filter\_at(vars(env\_x\_x), any\_vars(. %in% c(1)))   
env <- env[,-c(1,13:20)]  
env\_m <- env %>% melt(value.name="env")  
ggplot(env\_m, aes(x=variable, y=env)) + geom\_bar(stat = "identity") + theme(axis.text.x = element\_text(angle = 30)) + ggtitle("Count of all actions done to species with environmental stochasticity threat") + scale\_x\_discrete(name ="Name of Action")



#demographic stochasticity  
demo <- threats %>% filter\_at(vars(demo\_x\_x), any\_vars(. %in% c(1)))   
demo <- demo[,-c(1,13:20)]  
demo\_m <- demo %>% melt(value.name="demo")  
ggplot(demo\_m, aes(x=variable, y=demo)) + geom\_bar(stat = "identity") + theme(axis.text.x = element\_text(angle = 30)) + ggtitle("Count of all actions done to species with demographic stochasticity threat") + scale\_x\_discrete(name ="Name of Action")



#combine into one dataset   
### Should have just used group\_by - found the total number for each based on threat type and then joined   
#threat\_actions <- join\_all(list(hab\_m, demo\_m, spsp\_m, env\_m, poll\_m, over\_m), by='variable', type='left')  
#I think this is replicating information when joining, unsure how to fix  
#this would be the next step of putting onto 1 dataframe  
#tidythreat <- threat\_actions %>% pivot\_longer(c(hab, demo), names\_to = 'threat', values\_to = 'count')  
  
#unsure how to put on same graph- this is from stackoverflow   
require(gridExtra)

## Warning in library(package, lib.loc = lib.loc, character.only = TRUE,  
## logical.return = TRUE, : there is no package called 'gridExtra'

plot1 <- qplot(1)  
plot2 <- qplot(1)  
#grid.arrange(plot1, plot2, ncol=2)

plot hists = M

use facet\_wrap

### Histogram of number of partners per species

Histogram of Ms take 2 - # of partners per species - # of species per partner - actpartsum - actsum - how many partners are participating in each action - number of species with each action - number of partners that apply action to at least 1 species

scrap work

* no. of partners per species

## I think would have made sense to melt before combining into one table   
  
  
 dataA <- sdata %>% group\_by(Scientific.name) %>% count('Scientific.name') #count the number of times rows for each scientific name  
 dataA$All <- dataA$freq

## Warning: Unknown or uninitialised column: 'freq'.

AA <- melt(dataA)

## Using Scientific.name, "Scientific.name" as id variables

#the following was repeated for each category   
 #dataA <- dataA %>% add\_count(freq) %>% distinct(freq, n) %>% arrange(freq)  
 dataB <- sdata %>% filter(type.of.partners == "M") %>% group\_by(Scientific.name) %>% count('Scientific.name')   
 dataB$onlyM <- dataB$freq

## Warning: Unknown or uninitialised column: 'freq'.

BB <- melt(dataB)

## Using Scientific.name, "Scientific.name" as id variables

#dataB <- dataB %>% add\_count(freq) %>% distinct(freq, n) %>% arrange(freq)  
 dataC <- sdata %>% filter(type.of.partners != "M") %>% group\_by(Scientific.name) %>% count('Scientific.name')  
 dataC$noM <- dataC$freq

## Warning: Unknown or uninitialised column: 'freq'.

CC <- melt(dataC)

## Using Scientific.name, "Scientific.name" as id variables

joint <- AA %>% full\_join(BB, by = "Scientific.name") %>% full\_join(CC, by = "Scientific.name")  
  
  
joint2 <- joint[-which(joint$variable.x == "freq"),]  
joint3 <- joint2[-which(joint2$variable.y == "freq"),]  
joint4 <- joint3[-which(joint3$variable == "freq"),]  
  
joint4 <- joint4[,-c(1)]  
  
five <- melt(joint4)

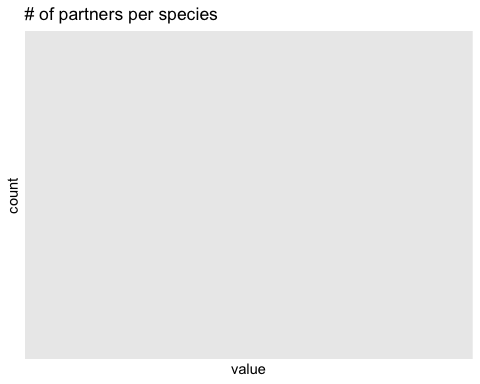
## Using "Scientific.name".x, variable.x, "Scientific.name".y, variable.y, "Scientific.name", variable as id variables

whoops <- dataA %>% full\_join(dataB, by = "Scientific.name") %>% full\_join(dataC, by = "Scientific.name")  
right <- melt(whoops)

## Using Scientific.name, "Scientific.name".x, "Scientific.name".y, "Scientific.name" as id variables

right <- right[-which(right$variable == "freq"),]  
right <- right[-which(right$variable == "freq.x"),]  
right <- right[-which(right$variable == "freq.y"),]  
  
right <- right[-which(is.na(right$value)),]  
  
right <- right[,-c(1)]  
  
ggplot(data = right, aes(x = value, fill = factor(variable))) +  
geom\_histogram(position = "dodge", stat="count") + ggtitle("# of partners per species")

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



* no. of species per partner

dataA <- sdata %>% group\_by(partner.in.agreement) %>% count('partner.in.agreement') #count the number of times rows for each scientific name  
 dataA$All <- dataA$freq

## Warning: Unknown or uninitialised column: 'freq'.

dataB <- sdata %>% filter(type.of.partners == "M") %>% group\_by(partner.in.agreement) %>% count('partner.in.agreement')   
 dataB$onlyM <- dataB$freq

## Warning: Unknown or uninitialised column: 'freq'.

dataC <- sdata %>% filter(type.of.partners != "M") %>% group\_by(partner.in.agreement) %>% count('partner.in.agreement')  
 dataC$noM <- dataC$freq

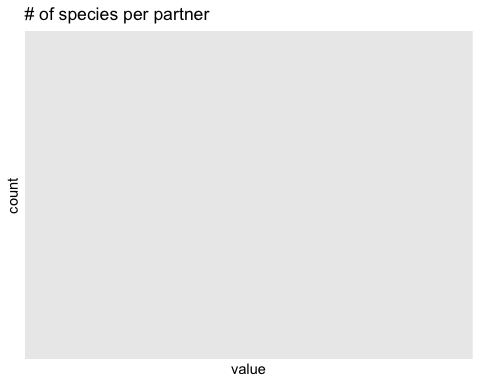
## Warning: Unknown or uninitialised column: 'freq'.

joint <- dataA %>% full\_join(dataB, by = "partner.in.agreement") %>% full\_join(dataC, by = "partner.in.agreement")  
  
  
  
right <- melt(joint)

## Using partner.in.agreement, "partner.in.agreement".x, "partner.in.agreement".y, "partner.in.agreement" as id variables

right <- right[-which(right$variable == "freq"),]  
right <- right[-which(right$variable == "freq.x"),]  
right <- right[-which(right$variable == "freq.y"),]  
  
right <- right[-which(is.na(right$value)),]  
  
right <- right[,-c(1)]  
  
  
ggplot(data = right, aes(x = value, fill = factor(variable))) +  
geom\_histogram(position = "dodge", stat="count") + ggtitle("# of species per partner")

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



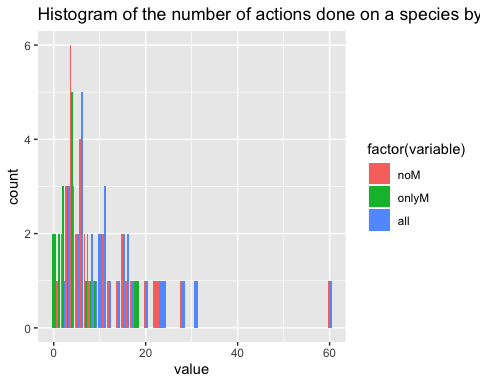
* actpartsum

species <- sdata  
  
#all   
speciesall <- species[,c(3,11:21)]  
eachspall <- speciesall %>% group\_by(Scientific.name) %>% summarise\_each(funs(sum))   
noactionswithpartnerall <- eachspall %>% mutate("all" = rowSums(eachspall[,c(2:12)]))  
noactionswithpartnerall <- noactionswithpartnerall[,-c(2:12)]  
  
#onlyM  
  
onlyM <- sdata[which(sdata$type.of.partners == "M"),]  
  
speciesonlyM <- onlyM  
speciesonlyM <- speciesonlyM[,c(3,11:21)]  
  
eachsponlyM <- speciesonlyM %>% group\_by(Scientific.name) %>% summarise\_each(funs(sum))   
noactionswithpartneronlyM <- eachsponlyM %>% mutate("onlyM" = rowSums(eachsponlyM[,c(2:12)]))  
noactionswithpartneronlyM <- noactionswithpartneronlyM[,-c(2:12)]  
  
#no M  
noM <- sdata[-which(sdata$type.of.partners == "M"),]  
  
speciesnoM <- noM  
speciesnoM <- speciesnoM[,c(3,11:21)]  
  
eachspnoM <- speciesnoM %>% group\_by(Scientific.name) %>% summarise\_each(funs(sum))   
noactionswithpartnernoM <- eachspnoM %>% mutate("noM" = rowSums(eachspnoM[,c(2:12)]))  
noactionswithpartnernoM <- noactionswithpartnernoM[,-c(2:12)]  
  
  
#join all df  
  
actpartsum <- noactionswithpartnernoM %>% full\_join(noactionswithpartneronlyM, by = "Scientific.name") %>% full\_join(noactionswithpartnerall, by = "Scientific.name")  
  
  
  
moltactpartsum <- melt(actpartsum)

## Using Scientific.name as id variables

moltactpartsum <- moltactpartsum[-which(is.na(moltactpartsum$value)),]  
  
moltactpartsum <- moltactpartsum[,-c(1)]  
  
  
  
ggplot(data = moltactpartsum, aes(x = value, fill = factor(variable))) +  
geom\_histogram(position = "dodge", stat="count") + ggtitle("Histogram of the number of actions done on a species by all partners")

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



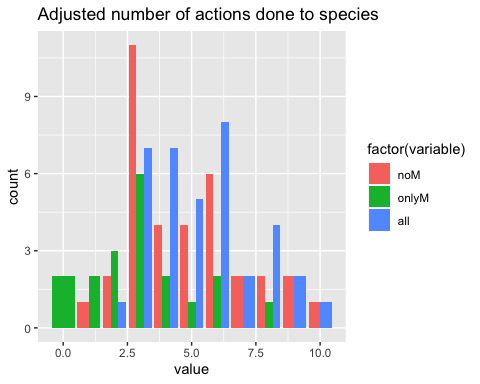
* actsum

species <- sdata  
  
  
#all   
speciesall <- species[,c(3,11:21)]  
eachspall <- speciesall %>% group\_by(Scientific.name) %>% summarise\_each(funs(sum))   
  
onesandzerosall <- eachspall %>% mutate\_if(is.numeric, ~1 \* (. > 0)) #changed all values back to ones and zeros  
totalnoall <- onesandzerosall %>% mutate("all" = rowSums(onesandzerosall[,c(2:12)]))   
totalnoall <- totalnoall[,-c(2:12)]  
  
  
#onlyM  
  
eachsponlyM <- speciesonlyM %>% group\_by(Scientific.name) %>% summarise\_each(funs(sum))   
  
onesandzerosonlyM <- eachsponlyM %>% mutate\_if(is.numeric, ~1 \* (. > 0)) #changed all values back to ones and zeros  
totalnoonlyM <- onesandzerosonlyM %>% mutate("onlyM" = rowSums(onesandzerosonlyM[,c(2:12)]))   
totalnoonlyM <- totalnoonlyM[,-c(2:12)]  
  
  
#noM   
eachspnoM <- speciesnoM %>% group\_by(Scientific.name) %>% summarise\_each(funs(sum))   
  
onesandzerosnoM <- eachspnoM %>% mutate\_if(is.numeric, ~1 \* (. > 0)) #changed all values back to ones and zeros  
totalnonoM <- onesandzerosnoM %>% mutate("noM" = rowSums(onesandzerosnoM[,c(2:12)]))   
totalnonoM <- totalnonoM[,-c(2:12)]  
  
  
#join all df  
  
actionsum <- totalnonoM %>% full\_join(totalnoonlyM, by = "Scientific.name") %>% full\_join(totalnoall, by = "Scientific.name")  
  
  
  
moltactionsum <- melt(actionsum)

## Using Scientific.name as id variables

moltactionsum <- moltactionsum[-which(is.na(moltactionsum$value)),]  
  
moltactionsum <- moltactionsum[,-c(1)]  
  
  
ggplot(data = moltactionsum, aes(x = value, fill = factor(variable))) +  
geom\_histogram(position = "dodge", stat="count") + ggtitle("Adjusted number of actions done to species")

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



Other histograms to complete - - how many partners are participating in each action - number of species with each action - number of partners that apply action to at least 1 species

Make Network for each type of org

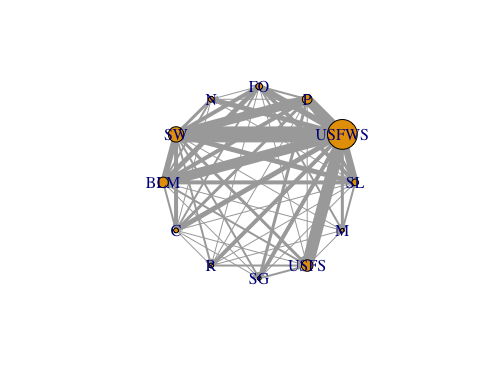
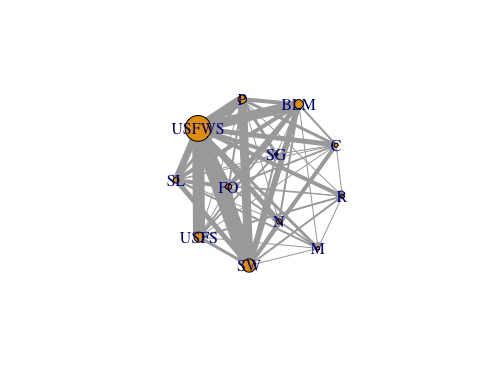
start from - managed to get df where all cells were lists (even though some said NULL?) was trying to unlist everything and then went pear shaped.. try and pick up from there

want to duplicate rows that are have cells labeled SB

orgtyp\_base2 <- orgtyp\_base  
  
dummy <- 2  
  
#subset and duplicate rows with SB as orgtype   
 orgtyp\_base3 <- orgtyp\_base2 %>%   
 filter(type\_of\_org=="SB") %>%   
 uncount(dummy) %>%   
 rownames\_to\_column() #this is the variable can index on because the rest are the same   
#change one of each rows to SL and the other to SW   
   
orgtyp\_base3[which(orgtyp\_base3$rowname == 1),9] <- "SW"  
orgtyp\_base3[which(orgtyp\_base3$rowname == 1.1),9] <- "SL"  
orgtyp\_base3[which(orgtyp\_base3$rowname == 2),9] <- "SW"   
orgtyp\_base3[which(orgtyp\_base3$rowname == 2.1),9] <- "SL"   
orgtyp\_base3[which(orgtyp\_base3$rowname == 3),9] <- "SW"   
orgtyp\_base3[which(orgtyp\_base3$rowname == 3.1),9] <- "SL"   
#then need to remove first column so has same dimensions as original df   
orgtyp\_base3 <- orgtyp\_base3[,-c(1)]  
   
#now remove old dataset with SB and join with this new one (rbind should work bc order doesn't matter)  
  
orgtyp\_base2 <- orgtyp\_base2[-which(orgtyp\_base2$type\_of\_org == "SB"),]  
  
#orgtyp\_base2 <- orgtyp\_base2 %>%   
 # filter(type\_of\_org !="SB") #removes all rows that have SB   
# so this doesn't actually work because need NAs for other parts of code to work   
  
#then join   
orgtyp\_base <- rbind(orgtyp\_base2, orgtyp\_base3)

## Warning: Values in `Scientific.name` are not uniquely identified; output will contain list-cols.  
## \* Use `values\_fn = list(Scientific.name = list)` to suppress this warning.  
## \* Use `values\_fn = list(Scientific.name = length)` to identify where the duplicates arise  
## \* Use `values\_fn = list(Scientific.name = summary\_fun)` to summarise duplicates

## [1] 35 12

Note - node labels seem to move every time for this one so need to check might actually be shifting as edit csv and sort various columns in that process  Caption: node = org type node size = # of species that all parnters categorized as that org works on edge width = # of species two orgs work on together Note: now where there use to be a SB (state both), this category has been duplicated into both SL and SW

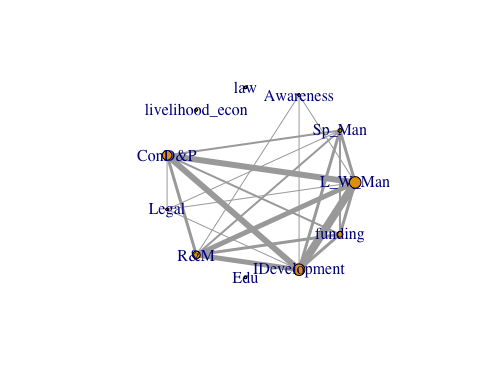
## Network for each org

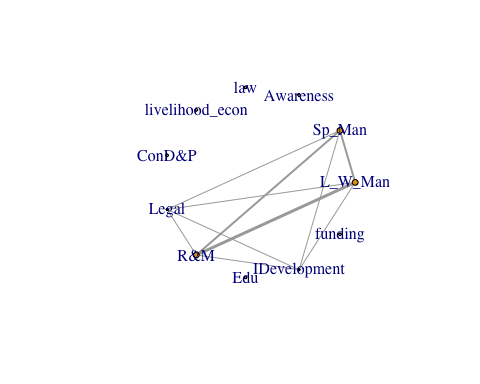
base <- orgtyp\_base  
base <- base[,-c(4:7,9:12)]  
  
multi <- base[which(is.na(base$type\_of\_org)),] #vector for multi-partner strings  
  
  
base <- base[-which(is.na(base$type\_of\_org)),] #remove rows with multi-partner strings  
  
#would be super easy with a loop   
  
  
#general process  
## subset by unique character in type\_of\_org column   
  
#so manual process of c&p didn't work and would be easier to change row names here in parent df than going individually and trying to fix vertex.label   
#now trying to change rownames   
  
colnames(base)[5:15] <- c("L\_W\_Man" ,"Sp\_Man", "Awareness", "law" , "livelihood\_econ", "ConD&P","Legal","R&M" ,"Edu","IDevelopment","funding") #brackets indicate which columns want to change the names of

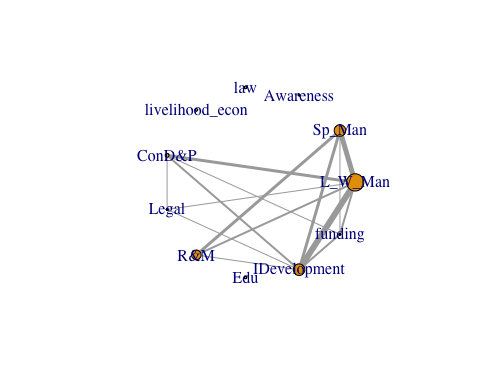
Caption: Network for each org where nodes are actions node size - number of times action is done Edge - number of partnerships that apply both types of actions

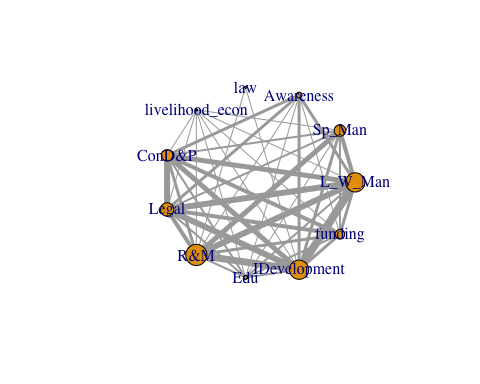
Node names (full action name -> shortened version for label)

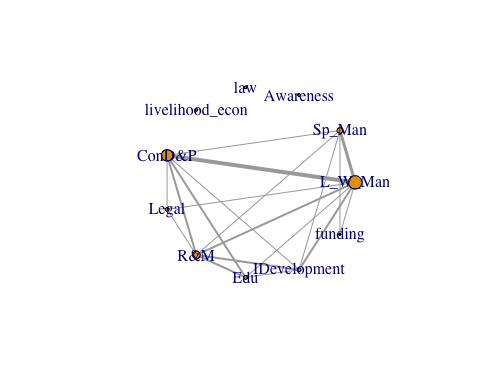
X1..Land.Water.Management“, –>”L\_W\_Man" “X2..Species.Management”, –> “Sp\_Man” “X3..Awareness.raising”, –> “Awareness” “X4..law.enforcement.and.prosecution” , –> “law” “X5..livelihood..economic.and.moral.incentrives”, –> “livelihood\_econ” “X6..Conservation.Design.and.Planning”, –> “ConD&P” “X7..Legal.and.Policy.frameworks”, –> “Legal” “X8..Research.and.monitoring”, –> “R&M” “X9..Education.and.Training”, –> “Edu” “X10..Institutional.Development”, –>“IDevelopment” “funding –>”funding"

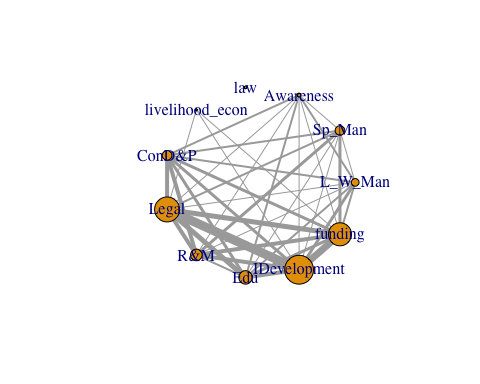
C = Corporation no. species = 6 no. partners = 11 

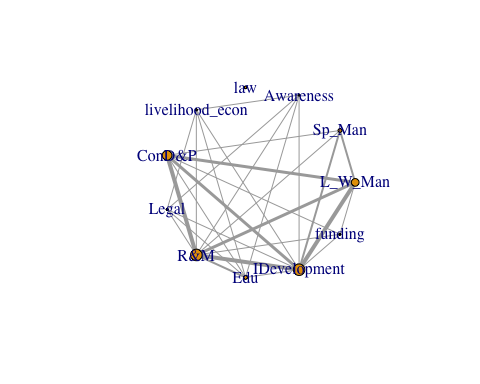
M = Military no. species = 4 no. partners = 7 

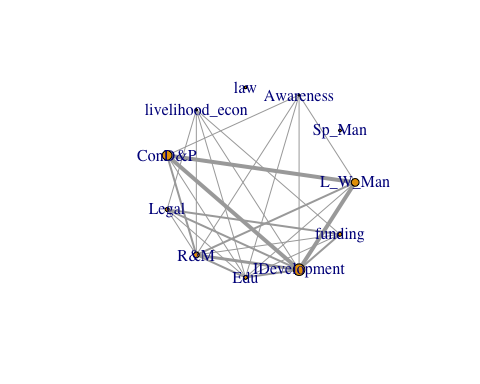
P = Private landowner no. species = 11 no. partners = 11 

SW = State Wildlife (Agency) no. species = 14 no. partners = 17 

BLM no. species = 11 no. partners = 11 

USFWS no. species = 31 no. partners = 34 

FO = Federal Other no. species = 7 no. partners = 9 

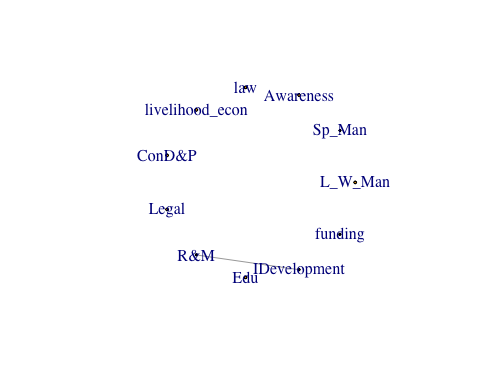
N = NGO no. species = 6 no. partners = 9 

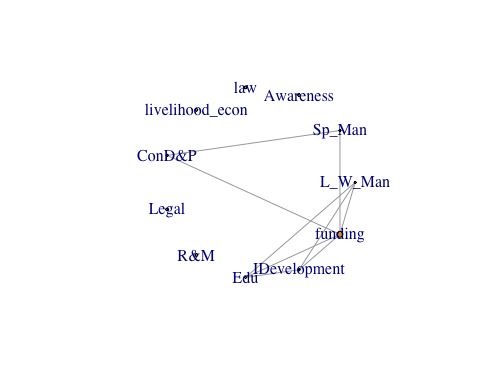
#### Skipping SB because there are only 2 state agencies in this category -

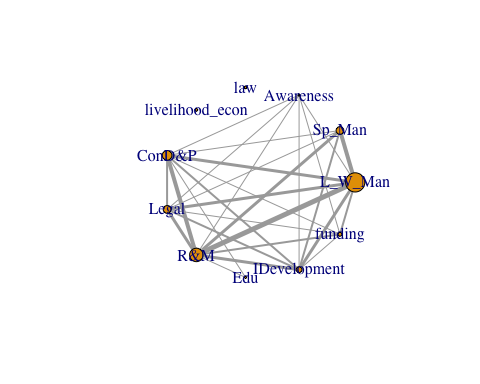
* Probably makes more sense to add them in both SW and SL

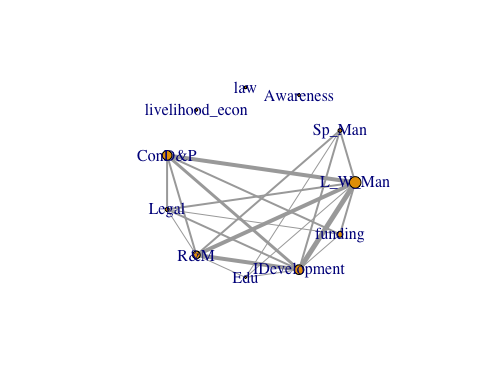
##### Other category removed

O = Other

R = Reserach no. species = 5 no. partners = 5 

SG = State Governmnet no. species = 4 no. partners = 5 

USFS no. species = 12 no. partners = 12 

SL no. species = 8 no. partners = 11 

For multi-partner strings no. species = 23 no. partners = 52 