MIIN Part 3: Calculate effect sizes

Marissa Lee

June 1, 2015

Filename: MIIN_3_calcEffectSizes.Rmd
This markdown file does the following tasks: 1. Clean dataframes: A) Remove observations that do not fit meta-analysis criteria that it must have least 1 focal exotic invasive species AND at least 1 nonfocal species, B) Remove ancillary soil measurements that will not be used in the meta-analysis, C) Simplify ecosystem type factor, D) Identify the N-fixing species, E) Create a character string to identify the invasive species associated with each observation . . . then, F) rename all the clean dataframes

2. Calculate invasion effect sizes
3. Create a composite dataset for analysis. Look at A) dataset structure, B) distribution of invasion effect size values, C) distribution of unit-standardized soil measurement values, and C) distribution of cwm trait values
4. Export dataframes

```
#knitr::opts_chunk$set(cache=TRUE)
require(plyr)
```

```
## Loading required package: plyr
```

```
require(reshape2)
```

```
## Loading required package: reshape2
```

```
require(ggplot2)
```

```
## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.2.4
```

```
require(metafor)
```

```
## Loading required package: metafor

## Loading required package: Matrix

## Warning: package 'Matrix' was built under R version 3.2.4

## Loading 'metafor' package (version 1.9-8). For an overview
## and introduction to the package please type: help(metafor).
```

```
source('CODE/mytheme.R')
```

```
## Loading required package: grid
```

```
figuresPath<-file.path(getwd()[1], "FIGURES_TABLES", "calcES") #where to put the saved plots
fig.height<-2.5 #inches
fig.width<- 2.5 #inches
fig.res<-300

synthdataPath<-file.path(getwd()[1], "DATA", "DATA_SYNTHESIZED", "calcES") #where to put the clean data

#data synthesized by *MIIN_1_paperData.Rmd*
papers<-read.table("DATA/DATA_SYNTHESIZED/paperData/papers_procd.txt", header=TRUE, sep="\t", quote="")
observations<-read.table("DATA/DATA_SYNTHESIZED/paperData/observations_procd.txt", header=TRUE, sep="\t
cover<-read.table("DATA/DATA_SYNTHESIZED/paperData/cover_procd.txt", header=TRUE, sep="\t")
species<-read.table("DATA/DATA_SYNTHESIZED/paperData/species_procd.txt", header=TRUE, sep="\t")
traits<-read.table("DATA/DATA_SYNTHESIZED/paperData/traits_procd.txt", header=TRUE, sep="\t")
measures<-read.table("DATA/DATA_SYNTHESIZED/paperData/measures_procd.txt", header=TRUE, sep="\t")

#data synthesized by *MIIN_2_cwm.Rmd*
cwm<-read.table("DATA/DATA_SYNTHESIZED/cwm/cwm.txt", header=TRUE, sep="\t")
cwm.quality<-read.table("DATA/DATA_SYNTHESIZED/cwm/cwm_quality.txt", header=TRUE, sep="\t")
spIDcover<-read.table("DATA/DATA_SYNTHESIZED/cwm/spIDcover.txt", header=TRUE, sep="\t")
spIDtraits<-read.table("DATA/DATA_SYNTHESIZED/cwm/spIDtraits.txt", header=TRUE, sep="\t")
numberOfSpecies.cwm<-read.table("DATA/DATA_SYNTHESIZED/cwm/numberOfSpecies_cwm.txt", header=TRUE, sep="

#read-in the list of Frankia and Rhizobia-associated plant genera
nfixGenera<-read.table("DATA/NfixingPlantGenera.txt", header=TRUE, sep="\t")
legumeGenera<-read.table("DATA/Leguminosae.csv", header=TRUE, sep=',')
```

---

1. Clean dataframes

1A. Remove observations that do not fit meta-analysis criteria that they must have least 1 focal exotic invasive species AND at least 1 nonfocal species. Observations will be removed from 'observations' dataframe and the corresponding related data in the following dataframes: cover, species, traits, measures, cwm, papers.

```
summ.spp <- ddply(species,~obsID,summarise,
                  numTotalspp=length(obsID),
                  numInvspp=sum(spInvasive=='invasive' & spExotic=='exotic' & spFocal=='focal'),
                  numNonFocalspp=sum(spFocal=='not focal'))
exclude.1inv<-summ.spp[summ.spp$numInvspp == 0,'obsID']
exclude.1nonfoc<-summ.spp[summ.spp$numNonFocal == 0,'obsID']
exclude.tmp<-c(exclude.1inv,exclude.1nonfoc)
exclude.obsID<-unique(exclude.tmp)
exclude.obsID
```

```
##  [1]   4.01   4.02  39.01  39.02  39.03  42.01  42.02  42.03  42.04  42.05
## [11]  42.06  42.07  50.01  50.02  50.03  62.01  62.02  62.03  62.04  62.05
```

```
## [21]  62.06  62.07  62.08  62.09  62.10  62.11  62.12  62.13  62.14  62.15
## [31]  62.16  62.17  62.18  62.19  62.20  62.21  62.22  62.31  62.32  62.33
## [41]  62.34  62.35  62.36  62.37 106.01 106.02 127.01 150.01 150.02 150.03
## [51] 150.04 150.05 157.01 198.01 198.02 198.03 198.04 275.01 281.01 281.02
## [61] 281.03 281.04 286.01 286.02 533.01 667.01 686.01 712.01 719.01 719.02
## [71] 719.03 719.04  16.01  60.01  60.02  60.03  74.01  75.01  87.04 205.05
## [81] 249.01 249.05 302.01 302.02 327.02 722.03
```

```
paste('Exclude',length(exclude.obsID), 'observations because there is not at least 1 species that is inv
```

```
## [1] "Exclude 86 observations because there is not at least 1 species that is invasive, exotic, AND fo
```

```
#identify the ok obsIDs
observations1<-observations[!observations$obsID %in% exclude.obsID,]
obsOK<-unique(observations1$obsID)
paste('There are',length(obsOK), 'observations remaining in the dataset that have at least 1 species tha
```

```
## [1] "There are 404 observations remaining in the dataset that have at least 1 species that is invasiv
```

```
#subset the remaining dfs from paperData
cover1<-subset(cover, obsID %in% obsOK)
species1<-subset(species, obsID %in% obsOK)
traits1<-subset(traits, obsID %in% obsOK)
measures1<-subset(measures, obsID %in% obsOK)

#subset the dfs from cwm
cwm1<-subset(cwm, obsID %in% obsOK)
cwm.quality1<-subset(cwm.quality, obsID %in% obsOK)
numberOfSpecies.cwm1<-subset(numberOfSpecies.cwm, obsID %in% obsOK)
spTOobs<-function(df){ #first, need to convert the spID to an obsID column in these dfs
  tmp<-ldply(strsplit(as.character(df[,'spID']), ".", fixed=T))
  df[,'obsID']<-paste(tmp[,1],tmp[,2], sep=".")
  return(df)
}
spIDcover<-spTOobs(spIDcover)
spIDtraits<-spTOobs(spIDtraits)
spIDcover1<-subset(spIDcover, obsID %in% obsOK)
spIDtraits1<-subset(spIDtraits, obsID %in% obsOK)

#annotate papers dataframe to reflect removal of observations and thus papers
#identify which observations in the exclude list come from papers that had OK'd observations
exclude.p1<-ldply(strsplit(as.character(exclude.obsID), ".", fixed=T))[,1]
include.p1<-ldply(strsplit(as.character(obsOK), ".", fixed=T))[,1]
reject.p<-unique(exclude.p1[!exclude.p1 %in% include.p1]) #if FALSE, then label these paperIDs as rejec
papers[papers$paperID %in% reject.p,'reject']<-'Yes'
newRationale<-'Not at least 1 species that is invasive, exotic, AND focal'
papers$rejectRationale<-factor(papers$rejectRationale, levels=c(levels(papers$rejectRationale),newRatio
papers[papers$paperID %in% reject.p,'rejectRationale']<-newRationale
papers1<-papers
```

1B. Remove ancillary soil measurements that will not be used in the meta-analysis.

```
summ.meas <- ddply(measures1,~measCat,summarise, numObs=length(unique(obsID)))

removeMeasCats<-c('biom','litterbiom','plantcov',
                  'cn','littercn','percN','litterpercN',
                  'microbcn','ph','soilmoi')
paste('Remove these measurement types:', paste(removeMeasCats, collapse=", "))
```

```
## [1] "Remove these measurement types: biom, litterbiom, plantcov, cn, littercn, percN, litterpercN, mi
```

```
measOK<-summ.meas[!summ.meas$measCat %in% removeMeasCats,'measCat']

#subset the remaining dfs from paperData
measures2<-subset(measures1, measCat %in% measOK)
measures3 <- droplevels(measures2)
```

1C. Simplify ecosystem type factor The 'other' category now consists of studies that took place in a dune system, or some combination of forest, grassland, wetland

```
summ.obs.eco <- ddply(observations1,~ecosystCat,summarise,
                      numObs=length(paperID),
                      numPapers=length(unique(paperID)))
summ.obs.eco
```

```
##                 ecosystCat numObs numPapers
## 1                     dune      2         2
## 2                   forest    123        51
## 3         forest,grassland      1         1
## 4 forest,grassland,wetland      1         1
## 5                grassland    176        53
## 6                shrubland     73        23
## 7                  wetland     28        14
```

```
#limit ecosystem categories to forest, grassland, shrubland, wetland, and other
criteria<-observations1$ecosystCat == 'forest,grassland' |
  observations1$ecosystCat == 'forest,grassland,wetland' |
  observations1$ecosystCat == 'dune'
levels(observations1$ecosystCat) <- c(levels(observations1$ecosystCat), "other")
observations1[criteria,'ecosystCat']<-'other'
```

1D. Identify the N-fixing plant species

```
#pull the unique legume genuses and combine with the full list of Frankia and Rhizobia associated plant
LegGenera<-unique(legumeGenera$Genus)
nfixGenera.sub<-nfixGenera[!nfixGenera$plantFamily =='Leguminosae','plantGenera']
nfixGenera.complete<-c(as.character(LegGenera),as.character(nfixGenera.sub))

#select rows in 'species' dataframe based on whether the species' genus name is present in nfixGenera.c
species1$nfixGenus<-'No' #fill everything in with 'No' first
species1[species1$Genus %in% nfixGenera.complete,'nfixGenus']<-'Yes'
numNfixRows<-dim(species1[species1$Genus %in% nfixGenera.complete,])[1]
numNonNfixRows<-dim(species1[!species1$Genus %in% nfixGenera.complete,])[1]
paste('There are', numNfixRows, 'and',numNonNfixRows, 'species entries (species unique to each observati
```

```
## [1] "There are 192 and 1806 species entries (species unique to each observation) that are N-fixing ar
```

```
#identify observations based on presence/absence of N-fixer as invasive species
selection<-species1$spInvasive == 'invasive' & species1$spExotic == 'exotic' & species1$spFocal == 'foca
df.selection<-species1[selection,]
NfixObsIDs<-unique(df.selection$obsID)
observations1$InvNfix<-'Non-N-fixing' #fill everthing with 'Non-N-fixing' first
observations1[observations1$obsID %in% NfixObsIDs,'InvNfix']<-'N-fixing'
numNfixRows<-sum(observations1$InvNfix == 'N-fixing')
numNonNfixRows<-sum(observations1$InvNfix != 'N-fixing')
paste('There are', numNfixRows, 'and',numNonNfixRows, 'observations that have N-fixing and non-N-fixing
```

```
## [1] "There are 71 and 333 observations that have N-fixing and non-N-fixing invasive species, respecti
```

```
#identify observations based on % native N-fixing species (not cover)
df.notFocal<-species1[species1$spFocal == 'not focal',]
summ.Nfix <- ddply(df.notFocal,~obsID,summarise,
                   NatnumNfix=sum(nfixGenus=='Yes'),
                   NatnumNotNfix=sum(nfixGenus=='No'),
                 NatpercNfix=(NatnumNfix/(NatnumNfix + NatnumNotNfix))*100)
summ.Nfix$NatNfix<-'No N-fixers' #fill everthing with 'No N-fixers' first
summ.Nfix[summ.Nfix$NatnumNfix > 0,'NatNfix']<-'N-fixers present'
numNfixRows<-sum(summ.Nfix$NatNfix == 'N-fixers present')
numNonNfixRows<-sum(summ.Nfix$NatNfix != 'N-fixers present')
paste('There are', numNfixRows, 'and',numNonNfixRows, 'observations that have reference areas with N-fi
```

```
## [1] "There are 58 and 346 observations that have reference areas with N-fixers present and No N-fixe
```

```
#create a composite Nfix factor
observations2<-merge(observations1, summ.Nfix, by='obsID')
observations2$Nfix<-paste(observations2$InvNfix, observations2$NatNfix, sep="_")
unique(observations2$Nfix)
```

```
## [1] "N-fixing_No N-fixers"          "Non-N-fixing_No N-fixers"
## [3] "Non-N-fixing_N-fixers present" "N-fixing_N-fixers present"
```

```
observations2$Nfix<-revalue(observations2$Nfix,
                            c("Non-N-fixing_No N-fixers" = "No N-fixers",
                              "N-fixing_No N-fixers" = "Invasive N-fixers only",
                              "Non-N-fixing_N-fixers present" = "Resident N-fixers only",
                              "N-fixing_N-fixers present" = "Invasive and resident N-fixers"))
```

1E. Create a character string to identify the invasive species associated with each observation

```
#create an obsID x invasive species dataframe
species.tmp<-subset(species1, spInvasive=='invasive' & spExotic=='exotic' & spFocal=='focal')
OBSID<-unique(species.tmp$obsID)
bindedrows<-numeric(0)
i<-0
for(i in 1:length(OBSID)){
  invGenera<-paste(species.tmp[species.tmp$obsID == OBSID[i],'Genus'], collapse='_')
```

```
  nspecies<-length(species.tmp[species.tmp$obsID == OBSID[i],'Genus'])
  if(nspecies > 2){
    invGenera<-'>2spp'
  }
  row<-data.frame(obsID=OBSID[i], invGenera)
  bindedrows<-rbind(bindedrows,row)
}
species.tmp2<-bindedrows
#View(species.tmp2)

#merge by obsID to add invasive species name to observations table
observations3<-merge(observations2, species.tmp2, by='obsID')
```

1F. Rename all the clean dataframes

```
papers.c<-papers1
observations.c<-observations3
cover.c<-cover1
species.c<-species1
traits.c<-traits1
measures.c<-measures3
cwm.c<-cwm1
cwm.quality.c<-cwm.quality1
spIDcover.c<-spIDcover1
spIDtraits.c<-spIDtraits1
numberOfSpecies.cwm.c<-numberOfSpecies.cwm1
```

---

2. Calculate invasion effect sizes and create a composite dataset for analyses (observations, measures, cwm)

To calculate invasion effect sizes, use measurement values that have not been unit-standardized. Calculate effect sizes using the "standard mean difference" (SMD). ALSO – this is where I calculate the difference in CWMs

```
chooseMeasType<-'nonSTD' #decide whether to use standardized/non-standardized soil measurement values
chooseESType<-'SMD' #decide whether to use ROM or SMD to calculate effect sizes

### Data that will be used ###############################################################
#observations.c
#measures.c
#cwm.c

### Calculate measurement ESs ############################################################

#study identifiers
obsID<-measures.c$obsID
measCat<-measures.c$measCat
measQuality<-measures.c$YN
```

```r
#invader impact measures - standardized units
n1i<-measures.c$inv_n
m1i<-measures.c$inv_mean_std
sd1i<-sqrt(measures.c$inv_var_std)
n2i<-measures.c$nat_n
m2i<-measures.c$nat_mean_std
sd2i<-sqrt(measures.c$nat_var_std)

dat.STD<-data.frame(obsID, measCat, measQuality,
                    n1i, m1i, sd1i, n2i, m2i, sd2i)

#invader impact measures - non-standardized units
m1i<-measures.c$inv_mean
sd1i<-sqrt(measures.c$inv_var)
m2i<-measures.c$nat_mean
sd2i<-sqrt(measures.c$nat_var)

dat.nonSTD<-data.frame(obsID, measCat, measQuality,
                       n1i, m1i, sd1i, n2i, m2i, sd2i)


### Combine the selected measurement ES values with the observation ID modifiers ######################
if(chooseMeasType == 'STD'){dat<-dat.STD}
if(chooseMeasType == 'nonSTD'){dat<-dat.nonSTD}

#add obsID factor columns to measures
dat.obs<-merge(dat,observations.c, by='obsID')
#dim(dat);dim(dat.obs) #should more columns, but same number of rows
#colnames(dat.obs) #get rid of unnecessary columns
dat.obs1<-dat.obs[,c('paperID','obsID','measCat',
                 'n1i', 'm1i', 'sd1i', 'n2i', 'm2i', 'sd2i',
                 'measQuality',
                 'ecosystCat','studyType','InvNfix','NatNfix','Nfix','invGenera')]

### Re-organize cwm data and prep for merging ##########################################################
#recast cwm so that type of CWM values are in the same row
cwm.tmp<-cwm.c[,c('obsID','traitCat','invType','qualRank','cwm')]
m.cwm.tmp<-melt(cwm.tmp, id.vars=c('obsID', 'traitCat','invType')) #uses reshape2
c.cwm.tmp<-dcast(m.cwm.tmp, obsID+traitCat~invType+variable)
#View(c.cwm.tmp)

#calculate the raw difference of invaded and native area cwm trait values
c.cwm.tmp$CWMDiff_cwm<-c.cwm.tmp$InvArea_cwm - c.cwm.tmp$NatArea_cwm

#calculate the raw difference of invasive species in invaded area and native area cwm trait values
c.cwm.tmp$CWMDiff2_cwm<-c.cwm.tmp$InvSpInvArea_cwm - c.cwm.tmp$NatArea_cwm

### Combine the selected measurement ES values and obsID info with CWM data ############################
#add cwm data to measures
#View(dat.obs1) #each row is a unique obsID x measCat
#View(c.cwm.tmp) #each row is a unique obsID x traitCat
dat.all<-merge(dat.obs1, c.cwm.tmp, by='obsID', all=TRUE)
#dim(dat.all)
```

```
#paste(length(unique(dat.obs1$obsID)), 'observations') #check to make sure that observations did not ge
#paste(length(unique(dat.all$obsID)), 'observations')

### Calculate the effect sizes ############################################################
dat1 <- escalc(measure=chooseESType, m1i=m1i, sd1i=sd1i, n1i=n1i, m2i=m2i, sd2i=sd2i, n2i=n2i, data=dat

### last, carry over the quality ranks for CWMDiff values
dat1$CWMDiff_qualRank<-dat1$InvArea_qualRank + dat1$NatArea_qualRank
dat1$CWMDiff2_qualRank<-dat1$InvSpInvArea_qualRank + dat1$NatArea_qualRank
```

---

3. Now that we have a composite dataset for analysis. Look at...

3A. Dataset structure

```
#summarize dataset by unique obsID+measCats so that data is not duplicated (multiple traits per obsID+m
summ<-ddply(dat1, ~obsID+measCat, summarize,
      uniqm1i = length(unique(m1i)),
      uniqm2i = length(unique(m2i)),
      uniqyi = length(unique(yi)),
      total = sum(uniqm1i, uniqm2i,uniqyi))
sum(summ$total != 3) # if 0, then obsID + measCat produces all unique rows
```

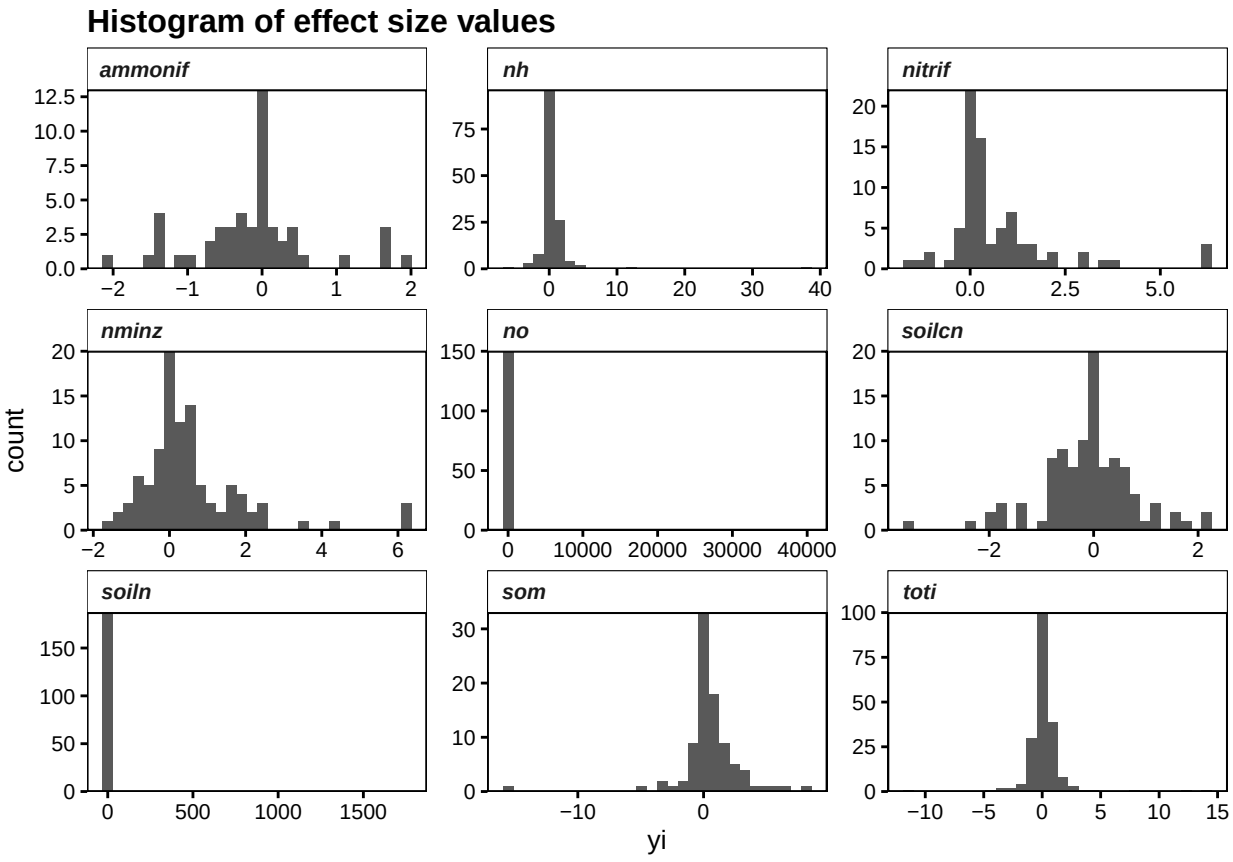
```
## [1] 0
```

```
dat1.meas<-ddply(dat1, ~obsID+measCat, summarize,
                 m1i = unique(m1i),
                 m2i = unique(m2i),
                 yi = unique(yi))
#head(dat1.meas) #each row is a unique obsID x measCat

#summarize dataset by unique obsID+traitCats so that data is not duplicated (multiple measures per obsI
summ<-ddply(dat1, ~obsID+traitCat, summarize,
      uniqInvArea = length(unique(InvArea_cwm)),
      uniqInvSpInvArea = length(unique(InvSpInvArea_cwm)),
      uniqNatArea = length(unique(NatArea_cwm)),
      uniqCWMDiff = length(unique(CWMDiff_cwm)),
      uniqCWMDiff2 = length(unique(CWMDiff2_cwm)),
      total = sum(uniqInvArea, uniqInvSpInvArea, uniqNatArea, uniqCWMDiff, uniqCWMDiff2))
#sum(summ$total != 5) # if 0, then obsID + traitCat produces all unique rows

dat1.tr<-ddply(dat1, ~obsID+traitCat, summarize,
               InvArea = unique(InvArea_cwm),
               InvSpInvArea = unique(InvSpInvArea_cwm),
               NatArea = unique(NatArea_cwm),
               CWMDiff = unique(CWMDiff_cwm),
               CWMDiff2 = unique(CWMDiff2_cwm))
#head(dat1.tr) #each row is a unique obsID x traitCat
```

```
#summarize dataset by unique obsID+traitCats+variable (where variable == c(InvArea, InvSpInvArea, NatAr
m.dat1.tr<-melt(dat1.tr, id.vars=c('obsID', 'traitCat'))
#head(m.dat1.tr) #each row is a unique obsID x traitCat x variable
#sum(is.na(m.dat1.tr$traitCat)) #combination is not fully factorial, so there are NAs here
#sum(is.na(m.dat1.tr$value)) #combination is not fully factorial, so there are NAs here
m.dat1.tr1<-m.dat1.tr[!is.na(m.dat1.tr$value),] #get rid of the NAs
```

3B. Distribution of effect size values

```
dat1.meas_tmp<-dat1.meas[!is.na(dat1.meas$yi),]

pHist_ES<-ggplot(dat1.meas_tmp, aes(x=yi)) + facet_wrap(~measCat, scales='free', ncol=3) + geom_histogra
  mytheme + scale_y_continuous(expand = c(0,0)) + ggtitle('Histogram of effect size values')
pHist_ES
```




```
#get rid of outliers
#nh
dat1[dat1$measCat=='nh' & dat1$yi > 30 & !is.na(dat1$yi),]
```

```
##    obsID paperID measCat n1i  m1i sd1i n2i m2i sd2i measQuality ecosystCat
## 1 706.01     706       nh   5 15.1  0.3   5 4.5  0.2  NoAgg.Conv  grassland
## 2 706.01     706       nh   5 15.1  0.3   5 4.5  0.2  NoAgg.Conv  grassland
##     studyType      InvNfix     NatNfix        Nfix invGenera traitCat
## 1 field study Non-N-fixing No N-fixers No N-fixers  Solidago       cn
## 2 field study Non-N-fixing No N-fixers No N-fixers  Solidago    percN
```

```
##   InvArea_qualRank InvArea_cwm InvSpInvArea_qualRank InvSpInvArea_cwm
## 1                2    28.88740                     2         28.88740
## 2                2     1.67977                     2          1.67977
##   NatArea_qualRank NatArea_cwm CWMDiff_cwm CWMDiff2_cwm      yi      vi
## 1                1   31.950795  -3.0633950   -3.0633950 37.5314 70.8301
## 2                0    1.960199  -0.2804294   -0.2804294 37.5314 70.8301
##   CWMDiff_qualRank CWMDiff2_qualRank
## 1                3                 3
## 2                2                 2
```

```
dat1[dat1$measCat=='nh' & dat1$yi > 30 & !is.na(dat1$yi),'yi']<-NA #replace outlier with NA
#no
dat1[dat1$measCat=='no' & dat1$yi > 30 & !is.na(dat1$yi),]
```

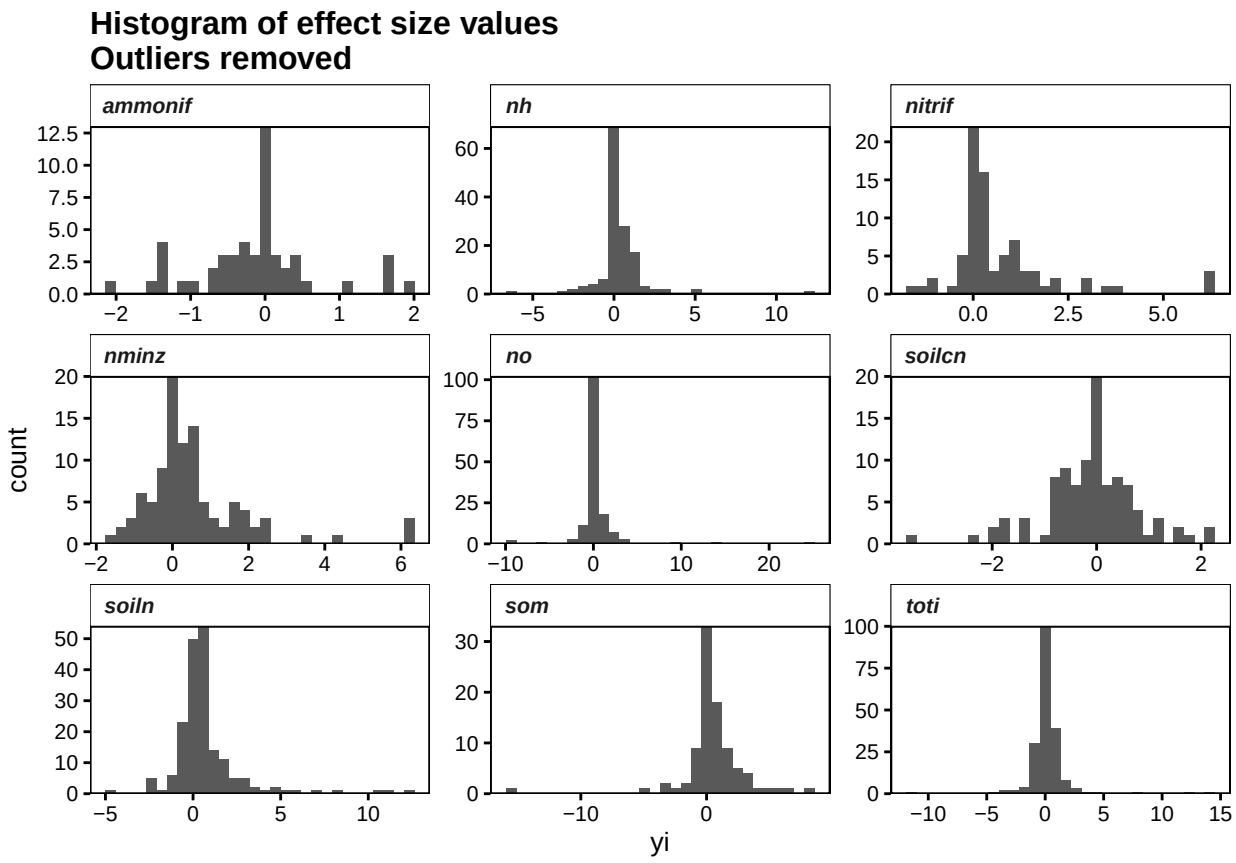
```
##    obsID paperID measCat n1i  m1i         sd1i n2i   m2i         sd2i
## 1  57.02      57      no  20 0.05 4.472136e-05  20 1e-05 4.472136e-05
## 2  57.02      57      no  20 0.05 4.472136e-05  20 1e-05 4.472136e-05
## 3 693.03     693      no   2 2.00 1.414214e-05   2 1e+00 1.414214e-05
##   measQuality ecosystCat   studyType      InvNfix     NatNfix        Nfix
## 1  NoAgg.Conv    wetland field study Non-N-fixing No N-fixers No N-fixers
## 2  NoAgg.Conv    wetland field study Non-N-fixing No N-fixers No N-fixers
## 3  NoAgg.Conv     forest field study Non-N-fixing No N-fixers No N-fixers
##   invGenera traitCat InvArea_qualRank InvArea_cwm InvSpInvArea_qualRank
## 1   Lythrum       cn                3     25.4895                     3
## 2   Lythrum    percN                3      1.7957                     3
## 3 Asparagus    percN                2      2.9475                     2
##   InvSpInvArea_cwm NatArea_qualRank NatArea_cwm CWMDiff_cwm CWMDiff2_cwm
## 1          26.9600                3    25.26340     0.22610      1.69660
## 2           1.6400                3     1.81964    -0.02394     -0.17964
## 3           2.9475                2     2.94750     0.00000      0.00000
##           yi             vi CWMDiff_qualRank CWMDiff2_qualRank
## 1  1095.5776     15003.7280                6                 6
## 2  1095.5776     15003.7280                6                 6
## 3 39894.2280 198943679.8649                4                 4
```

```
dat1[dat1$measCat=='no' & dat1$yi > 30 & !is.na(dat1$yi),'yi']<-NA #replace outlier with NA
# #ph
# dat1[dat1$measCat=='ph' & dat1$yi > 30 & !is.na(dat1$yi),]
# dat1[dat1$measCat=='ph' & dat1$yi < -30 & !is.na(dat1$yi),]
# dat1[dat1$measCat=='ph' & dat1$yi > 30 & !is.na(dat1$yi),'yi']<-NA #replace outlier with NA
# dat1[dat1$measCat=='ph' & dat1$yi < -30 & !is.na(dat1$yi),'yi']<-NA #replace outlier with NA
#soiln
dat1[dat1$measCat=='soiln' & dat1$yi > 30 & !is.na(dat1$yi),]
```

```
##    obsID paperID measCat n1i m1i         sd1i n2i m2i         sd2i
## 1 663.01     663   soiln  32 0.3 5.656854e-05  32 0.2 5.656854e-05
## 2 663.01     663   soiln  32 0.3 5.656854e-05  32 0.2 5.656854e-05
##    measQuality ecosystCat   studyType      InvNfix     NatNfix        Nfix
## 1 NoAgg.NoConv  shrubland field study Non-N-fixing No N-fixers No N-fixers
## 2 NoAgg.NoConv  shrubland field study Non-N-fixing No N-fixers No N-fixers
##   invGenera traitCat InvArea_qualRank InvArea_cwm InvSpInvArea_qualRank
## 1     >2spp       cn                1   27.079636                     1
## 2     >2spp    percN                2    1.437404                     2
```

```
##   InvSpInvArea_cwm NatArea_qualRank NatArea_cwm CWMDiff_cwm CWMDiff2_cwm
## 1         25.53359                1  29.7324372  -2.6528016    -4.198847
## 2          2.15343                2   0.8154332   0.6219708     1.337996
##          yi         vi CWMDiff_qualRank CWMDiff2_qualRank
## 1 1746.2816 23824.2757                2                 2
## 2 1746.2816 23824.2757                4                 4
```

```
dat1[dat1$measCat=='soiln' & dat1$yi > 30 & !is.na(dat1$yi),'yi']<-NA #replace outlier with NA

#update and re-plot
dat1.meas<-ddply(dat1, ~obsID+measCat, summarize,
                 m1i = unique(m1i),
                 m2i = unique(m2i),
                 yi = unique(yi))
pHist_ES_OR<-ggplot(dat1.meas, aes(x=yi)) + facet_wrap(~measCat, scales='free', ncol=3) + geom_histogra
  mytheme + scale_y_continuous(expand = c(0,0)) + ggtitle('Histogram of effect size values\nOutliers re
pHist_ES_OR
```




```
paste('Removed outlier effect size values in nh, no, soiln.')
```

```
## [1] "Removed outlier effect size values in nh, no, soiln."
```

```
paste('Effect size values look normally-distributed-ish')
```

```
## [1] "Effect size values look normally-distributed-ish"
```

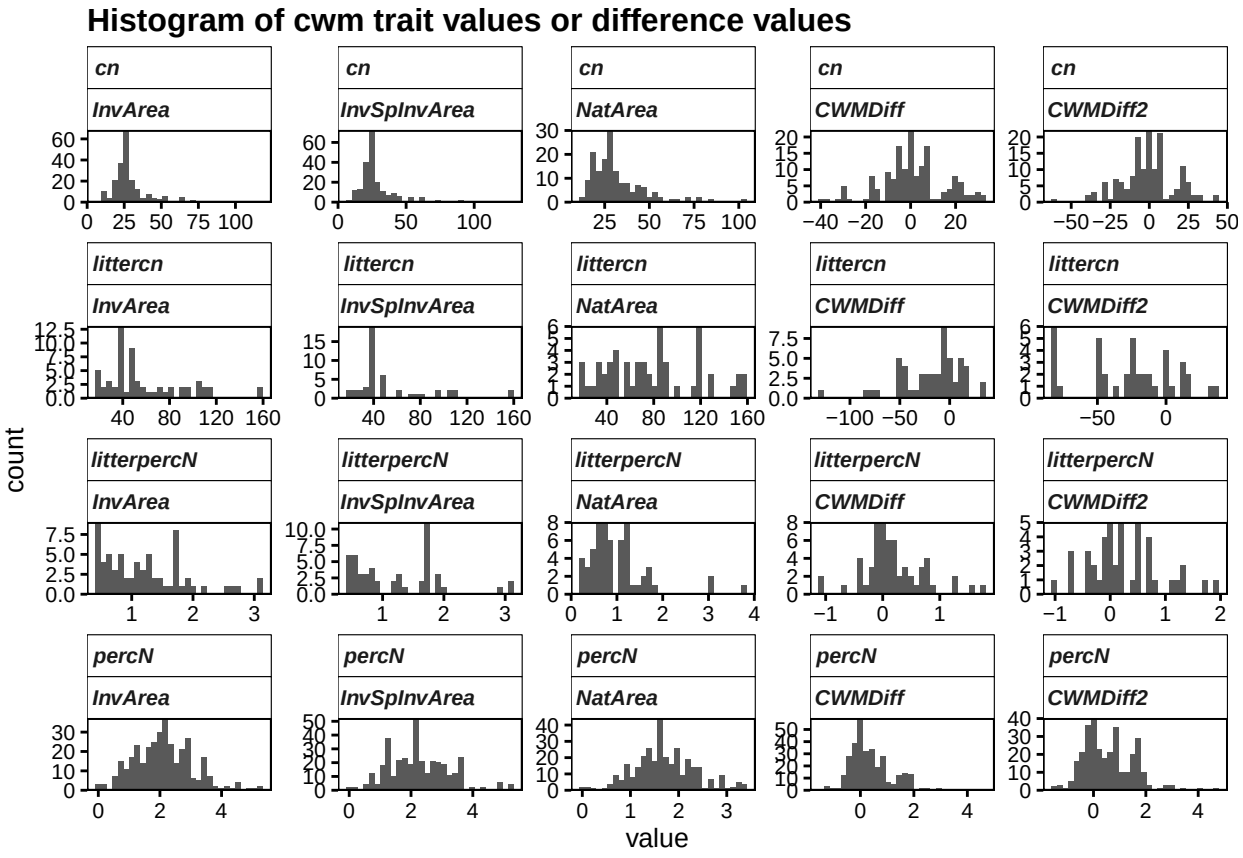
```
newfilename<-'pHist_ES_OR.png'
png(paste(figuresPath,newfilename, sep='/'),
    units='in', width = fig.width*3, height = fig.height*6, res=fig.res)
pHist_ES_OR
dev.off()
```

```
## pdf
##   2
```

NOT RUN – 3C. Distribution of unit-standardized soil measurement values

3D. Distribution of CWM trait values

```
#View(m.dat1.tr1)
pHist_cwm<-ggplot(m.dat1.tr1, aes(x=value)) + facet_wrap(~traitCat+variable, scales='free', ncol=5) + g
  mytheme + scale_y_continuous(expand = c(0,0)) + ggtitle('Histogram of cwm trait values or difference
pHist_cwm
```

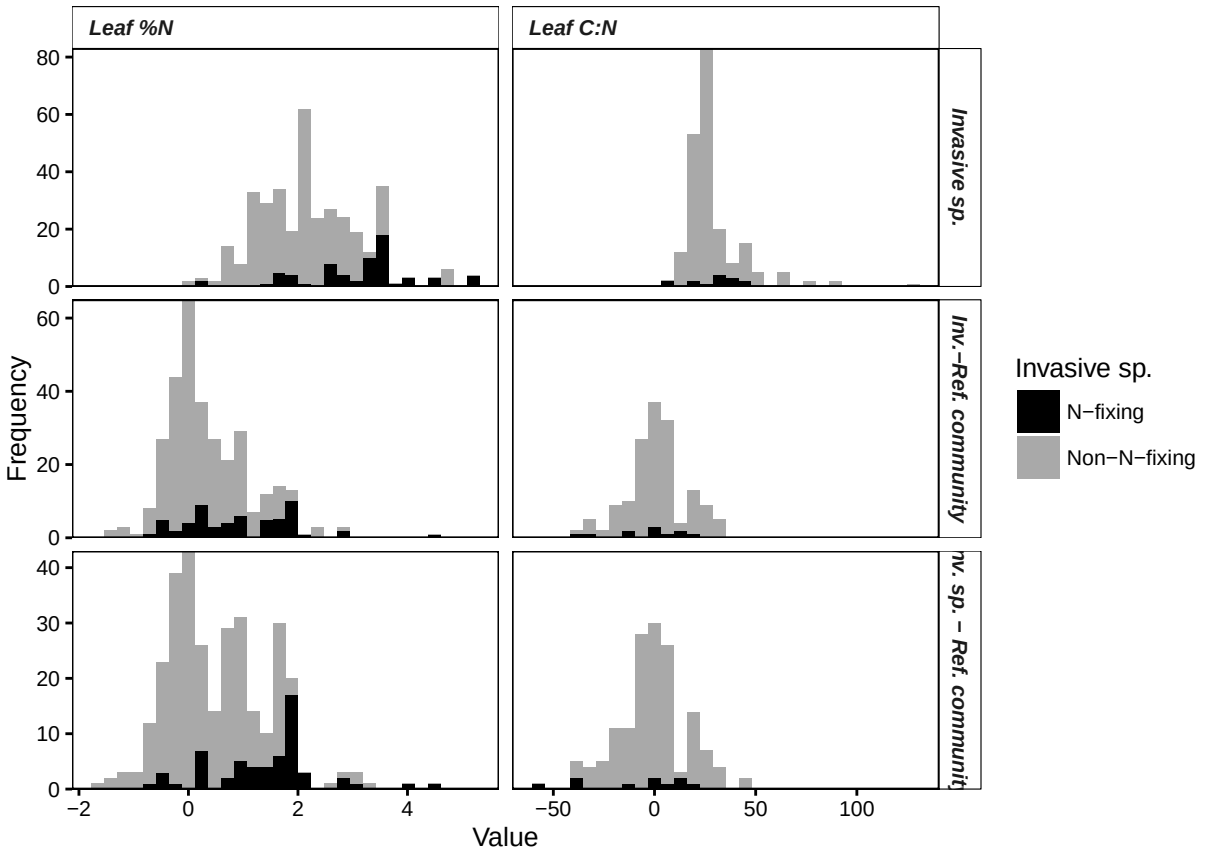



```
#make a pretty set, color coded by if the invader is N-fixing
tmp<-observations.c[,c("obsID","InvNfix")]
m.dat1.tr2<-merge(m.dat1.tr1,tmp)

#just invasive sp
data.invSp<-subset(m.dat1.tr2, variable %in% c("InvSpInvArea","CWMDiff","CWMDiff2"))
data.invSp$traitCat <- factor(data.invSp$traitCat, levels=traitCat_order)
```

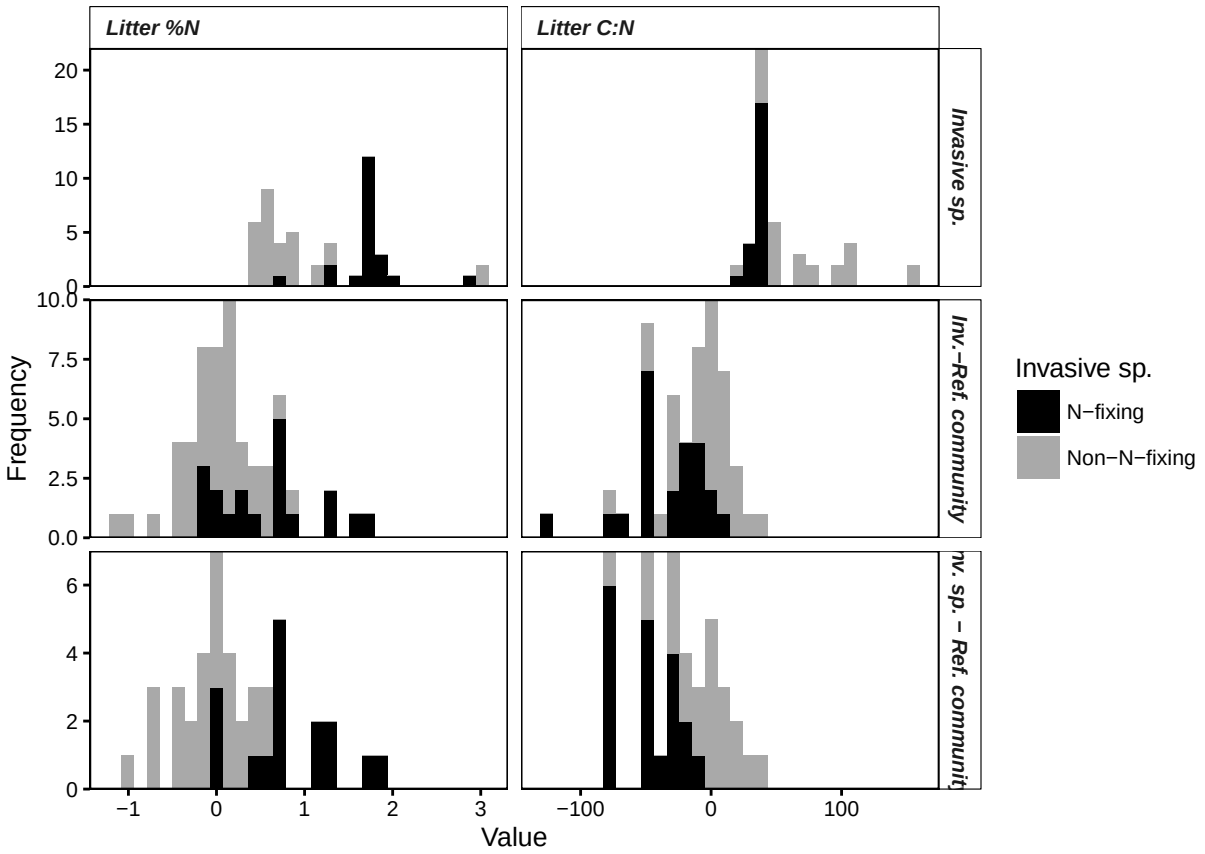
```
data.invSp$traitCat <-mapvalues(data.invSp$traitCat, from = traitCat_order, to = prettylabels.tr)

require(plyr)
summ.traits<-ddply(m.dat1.tr2, ~variable+traitCat, summarize,
      n=sum(!is.na(value)),
      meanTr=mean(value, na.rm=TRUE),
      stdevTr=sd(value, na.rm=TRUE),
      seTr=stdevTr/sqrt(n))
summ.traits
```

```
##        variable    traitCat   n      meanTr    stdevTr       seTr
## 1       InvArea          cn 212  30.2162202 16.1814558 1.11134695
## 2       InvArea    littercn  60  58.4443360 32.9203072 4.24999338
## 3       InvArea litterpercN  66   1.2180689  0.6697569 0.08244138
## 4       InvArea       percN 342   2.1187985  0.9464559 0.05117848
## 5  InvSpInvArea          cn 208  28.7442224 14.4959865 1.00511582
## 6  InvSpInvArea    littercn  47  54.2030222 32.5390965 4.74631504
## 7  InvSpInvArea litterpercN  50   1.2317541  0.7019326 0.09926826
## 8  InvSpInvArea       percN 364   2.2626386  0.9792034 0.05132420
## 9       NatArea          cn 155  31.5742554 14.5227459 1.16649502
## 10      NatArea    littercn  54  77.9183961 38.9843176 5.30509367
## 11      NatArea litterpercN  59   1.0110977  0.6741823 0.08777106
## 12      NatArea       percN 320   1.7069455  0.6372322 0.03562236
## 13      CWMDiff          cn 155   0.8437700 14.2862175 1.14749661
## 14      CWMDiff    littercn  54 -17.4689779 30.3480190 4.12984230
## 15      CWMDiff litterpercN  59   0.1854296  0.5452120 0.07098055
## 16      CWMDiff       percN 318   0.3801104  0.7935267 0.04449877
## 17     CWMDiff2          cn 151  -0.8429423 16.9397172 1.37853463
## 18     CWMDiff2    littercn  41 -26.4167309 33.0985480 5.16912476
## 19     CWMDiff2 litterpercN  43   0.2460547  0.6448076 0.09833223
## 20     CWMDiff2       percN 313   0.5697402  0.9505004 0.05372546
```

```
variable_order<-c("InvSpInvArea","CWMDiff","CWMDiff2")
prettylabels.variable<-c("Invasive sp.","Inv.-Ref. community", "Inv. sp. - Ref. community")
data.invSp$variable <- factor(data.invSp$variable, levels=variable_order)
data.invSp$variable <-mapvalues(data.invSp$variable,
                                from = variable_order, to = prettylabels.variable)

data.invSp.leaf<-subset(data.invSp, traitCat %in% c("Leaf %N","Leaf C:N"))
data.invSp.litter<-subset(data.invSp, traitCat %in% c("Litter %N","Litter C:N"))

pHist_invSp.leaf<-ggplot(data.invSp.leaf, aes(x=value, fill=InvNfix)) +
  geom_histogram() +
  facet_grid(variable~traitCat, scale="free") +
  mytheme +
  scale_y_continuous(expand = c(0,0)) +
  scale_fill_manual(values=c("black","darkgray"),
                    name="Invasive sp.") +
  ylab('Frequency') + xlab('Value')
pHist_invSp.leaf
```




```
pHist_invSp.litter<-ggplot(data.invSp.litter, aes(x=value, fill=InvNfix)) +
  geom_histogram() +
  facet_grid(variable~traitCat, scale="free") +
  mytheme +
  scale_y_continuous(expand = c(0,0)) +
  scale_fill_manual(values=c("black","darkgray"),
                    name="Invasive sp.") +
  ylab('Frequency') + xlab('Value')
pHist_invSp.litter
```




```
newfilename<-'pHist_cwm.png'
png(paste(figuresPath,newfilename, sep='/'),
    units='in', width = fig.width*3, height = fig.height*6, res=fig.res)
pHist_cwm
dev.off()
```

```
## pdf
##   2
```

```
newfilename<-'pHist_leaf.png'
png(paste(figuresPath,newfilename, sep='/'),
    units='in', width = fig.width*2, height = fig.height*2, res=fig.res)
pHist_invSp.leaf
dev.off()
```

```
## pdf
##   2
```

```
newfilename<-'pHist_litter.png'
png(paste(figuresPath,newfilename, sep='/'),
    units='in', width = fig.width*2, height = fig.height*2, res=fig.res)
pHist_invSp.litter
dev.off()
```

```
## pdf
##   2
```

```
paste('No outliers removed at this point, but they were taken out (if necessary) previously in MIIN_2_cu
```

```
## [1] "No outliers removed at this point, but they were taken out (if necessary) previously in MIIN_2_c
```

```
paste('Most cwm values look normally-distributed - except littercn, probably because there was not very
```

```
## [1] "Most cwm values look normally-distributed - except littercn, probably because there was not very
```

---

4. Export dataframes

```
newfilename<-'papers.txt'
write.table(papers.c, file=paste(synthdataPath,newfilename, sep='/'), sep='\t', quote=TRUE, qmethod="dou

newfilename<-'observations.txt'
write.table(observations.c, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')

newfilename<-'cover.txt'
write.table(cover.c, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')

newfilename<-'species.txt'
write.table(species.c, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')

newfilename<-'traits.txt'
write.table(traits.c, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')

newfilename<-'measures.txt'
write.table(measures.c, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')

newfilename<-'cwm.txt'
write.table(cwm.c, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')

newfilename<-'cwm_quality.txt'
write.table(cwm.quality.c, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')

newfilename<-'spIDcover.txt'
write.table(spIDcover.c, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')

newfilename<-'spIDtraits.txt'
write.table(spIDtraits.c, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')

newfilename<-'numberOfSpecies_cwms.txt'
write.table(numberOfSpecies.cwm.c, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')

newfilename<-'metaDataset.txt'
write.table(dat1, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')
```