MIIN Part 4: Meta-dataset overview

Marissa Lee

June 1, 2015

Filename: MIIN_4_datasetOverview.Rmd
This markdown file does the following tasks: 1. Article selection statistics

2. Number of papers and observations
3. Types of observations
4. Plant species statistics
5. Cover data statistics
6. Trait data statistics
7. Soil measurement statistics
8. Effect size statistics
9. CWM trait value statistics

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

require(plyr)
```

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

```
require(doBy)
```

```
## Loading required package: doBy

## Loading required package: survival
```

```
require(ggplot2)
```

```
## Loading required package: ggplot2

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

```
require(reshape2)
```

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

```
require(gridExtra)
```

```
## Loading required package: gridExtra

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

```r
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).
```

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

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

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

#from MIIN_3_calcEffectSizes.Rmd
papers<-read.table("DATA/DATA_SYNTHESIZED/calcES/papers.txt", sep="\t")
vilaRefs<-read.table("DATA/Vila_references.txt", header=TRUE, sep="\t")
castroRefs<-read.table("DATA/CastroDiez_references.txt", header=TRUE, sep="\t")
observations<-read.table("DATA/DATA_SYNTHESIZED/calcES/observations.txt", header=TRUE, sep="\t")
cover<-read.table("DATA/DATA_SYNTHESIZED/calcES/cover.txt", header=TRUE, sep="\t")
species<-read.table("DATA/DATA_SYNTHESIZED/calcES/species.txt", header=TRUE, sep="\t")
traits<-read.table("DATA/DATA_SYNTHESIZED/calcES/traits.txt", header=TRUE, sep="\t")
measures<-read.table("DATA/DATA_SYNTHESIZED/calcES/measures.txt", header=TRUE, sep="\t")
cwm<-read.table("DATA/DATA_SYNTHESIZED/calcES/cwm.txt", header=TRUE, sep="\t")
cwm.quality<-read.table("DATA/DATA_SYNTHESIZED/calcES/cwm_quality.txt", header=TRUE, sep="\t")
spIDcover<-read.table("DATA/DATA_SYNTHESIZED/calcES/spIDcover.txt", header=TRUE, sep="\t")
spIDtraits<-read.table("DATA/DATA_SYNTHESIZED/calcES/spIDtraits.txt", header=TRUE, sep="\t")
numberOfSpecies.cwm<-read.table("DATA/DATA_SYNTHESIZED/calcES/numberOfSpecies_cwms.txt", header=TRUE, s
metaDataset<-read.table("DATA/DATA_SYNTHESIZED/calcES/metaDataset.txt", header=TRUE, sep="\t")
```

---

1. Article selection statistics

```r
### Number of papers detected by source ###
summ.papers <- ddply(papers,~source,summarise,
                     numPapers=length(read),
                     numAcceptedPapers=sum(reject=='No'))
summ.papers<-orderBy(~-numPapers, summ.papers)
summ.papers
```

```
##                                source numPapers numAcceptedPapers
## 19                     search2_111714       388                36
## 18                     search1_111714       219                46
## 12                          Liao2007        94                47
## 11 independent search for plant traits         3                 0
## 3                         cited by 249         2                 2
## 8                         cited by 368         2                 2
## 10                        cited by 626         2                 1
## 1                         cited by 155         1                 1
## 2                         cited by 229         1                 0
## 4                          cited by 25         1                 1
## 5                         cited by 256         1                 0
## 6                          cited by 29         1                 1
## 7                         cited by 317         1                 1
## 9                         cited by 455         1                 1
## 13                        ReferencedBy         1                 1
## 14                related record 181         1                 1
## 15                related record 188         1                 1
## 16                  related record 4         1                 0
## 17                related record 570         1                 1
```

```
### Number of unique number of papers detected ###
summ.papers2 <- ddply(papers,~source+rejectRationale,summarise,
                      numPapers=length(read),
                      numAcceptedPapers=sum(reject=='No'))
summ.papers2<-orderBy(~-numPapers, summ.papers2)
totalNumReturned<-sum(summ.papers$numPapers) #total number of papers detected
numAlreadyFound<-sum(summ.papers2[summ.papers2$rejectRationale == 'alreadyFound' & !is.na(summ.papers2$
numUnique<-totalNumReturned - numAlreadyFound #total number of unique papers detected
paste(numUnique, 'unique papers identified by search criteria and their references')
```

```
## [1] "483 unique papers identified by search criteria and their references"
```

```
paste(sum(summ.papers$numAcceptedPapers), 'papers were accepted')
```

```
## [1] "143 papers were accepted"
```

```
length(unique(metaDataset$paperID)) #this should be the same number
```

```
## [1] 143
```

```
### Identify which papers were also detected by Vila 2011 and Castro-Diez 2014
papers.a<-papers[papers$reject == 'No',]
dim(papers.a)
```

```
## [1] 143  14
```

```
length(unique(papers.a$paperID))
```

```
## [1] 143
```

```
#vila
vilaRefs1<-vilaRefs[,2:3]
colnames(vilaRefs1)<-c('author1','year')
vilaRefs1$vila<-rep('Yes',dim(vilaRefs1)[1])
papers.a.tmp<-merge(papers.a, vilaRefs1, all.x=TRUE)

#problem: merge added rows without different info
dim(papers.a.tmp) #there are extra rows in here
```

```
## [1] 145  15
```

```
length(unique(papers.a.tmp$paperID)) #same number of unique paperIDs
```

```
## [1] 143
```

```
# tmp<-ddply(papers.a.tmp, ~paperID, summarise,
#            nRows=length(author1))
# tmp[tmp$nRows>1,]
# papers.a.tmp[papers.a.tmp$paperID %in% c(78,715),] # I have no idea why this happened
#solution: pull out unique rows
papers.a.new<-ddply(papers.a.tmp, ~paperID, summarise,
           nRows=length(author1),
           author1.new=unique(author1),
           year.new=unique(year),
           title.new=unique(title),
           journal.new=unique(journal),
           source.new=unique(source),
           vila.new=unique(vila))
#papers.a.new

#castro-diez
castroRefs1<-castroRefs[,2:3]
colnames(castroRefs1)<-c('author1.new','year.new')
castroRefs1$castroDiez<-rep('Yes',dim(castroRefs1)[1])
papers.a.tmp1<-merge(papers.a.new, castroRefs1, all.x=TRUE)

#problem: merge added rows without different info
dim(papers.a.tmp1) #there are extra rows in here
```

```
## [1] 147   9
```

```
length(unique(papers.a.tmp1$paperID)) #same number of unique paperIDs
```

```
## [1] 143
```

```
#solution: pull out unique rows
papers.a.new1<-ddply(papers.a.tmp1, ~paperID, summarise,
           author1=unique(author1.new),
           year=unique(year.new),
           title=unique(title.new),
```

```
            journal=unique(journal.new),
            source=unique(source.new),
            vila=unique(vila.new),
            castroDiez.new=unique(castroDiez))
#papers.a.new1

#liao
papers.a.new1[papers.a.new1$source == 'Liao2007','liao']<-'Yes'
#View(papers.a.new1)
colnames(papers.a.new1)[colnames(papers.a.new1)=='castroDiez.new']<-'castroDiez'
papers.a<-papers.a.new1[,c('paperID','author1','year','title','journal','liao','vila','castroDiez')]

#number of papers that overlap across metas
nOverlap.liao<-sum(!is.na(papers.a$liao))
nOverlap.vila<-sum(!is.na(papers.a$vila))
nOverlap.castroDiez<-sum(!is.na(papers.a$castroDiez))
paste('This dataset has', nOverlap.liao, 'papers in common with Liao')
```

```
## [1] "This dataset has 47 papers in common with Liao"
```

```
paste(nOverlap.vila, 'papers in common with Vila')
```

```
## [1] "49 papers in common with Vila"
```

```
paste(nOverlap.castroDiez, 'papers in common with Castro-Diez')
```

```
## [1] "43 papers in common with Castro-Diez"
```

```
#make a new column to indicate if the paper is new to this study
papers.a$sourceOverlap<-'PaperInPreviousMeta'
papers.a[is.na(papers.a$liao) & is.na(papers.a$vila) & is.na(papers.a$castroDiez),'sourceOverlap']<-'New

#attached the liao, vila, castroDiez columns to the metaDataset if it isn't there already
if(sum(colnames(metaDataset) %in% c('sourceOverlap'))==0){
  temp_indx<-papers.a[,c('paperID','sourceOverlap')]
  metaDataset<-merge(metaDataset, temp_indx, by='paperID')

  #re-write the metaDataset file so that it has the source2 column
  newfilename<-'metaDataset.txt'
  synthdataPath<-file.path(getwd()[1], "DATA", "DATA_SYNTHESIZED")
  write.table(metaDataset, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')
}
```

---

2. Number of papers and observations

```r
#how many observations?
paste(length(unique(observations$obsID)), 'observations in the full dataset')
```

```
## [1] "404 observations in the full dataset"
```

```r
length(unique(metaDataset$obsID)) #these should be the same
```

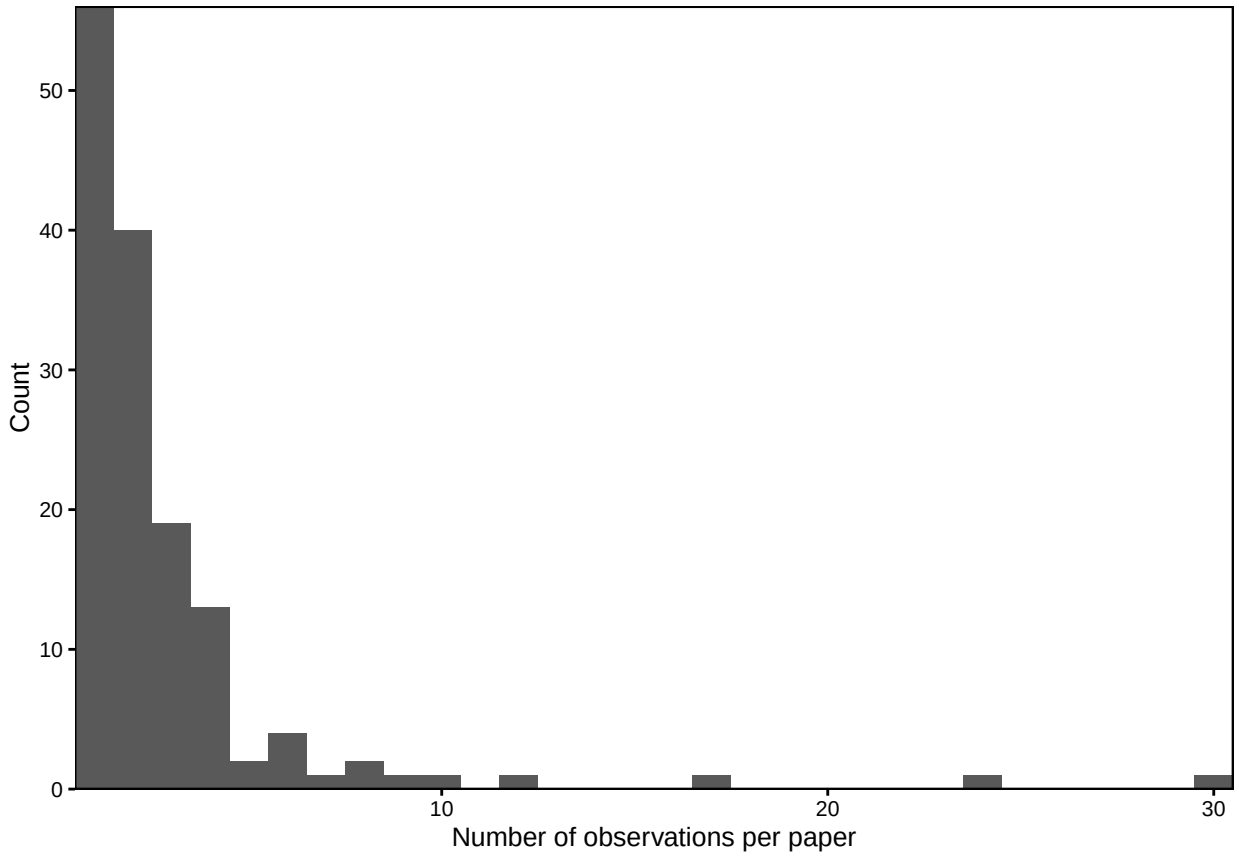
```
## [1] 404
```

```r
#how many observations per paper?
summ.obs <- ddply(observations,~paperID,summarise, numObs=length(paperID))
median(summ.obs$numObs); range(summ.obs$numObs)
```

```
## [1] 2
```

```
## [1]  1 30
```

```r
pHist_obs<-ggplot(summ.obs, aes(x=numObs)) +
  scale_y_continuous(expand=c(0,0)) + scale_x_continuous(expand=c(0,0)) +
  geom_histogram() + mytheme +
  ylab('Count') + xlab('Number of observations per paper')
pHist_obs
```



3. Types of observations

```
tmp <- ddply(metaDataset,~obsID,summarise,
             ecosyst=unique(ecosystCat),
             stud=unique(studyType),
             nfix=unique(Nfix),
             sourceOverlap=unique(sourceOverlap))
summ.obs.eco<-ddply(tmp, ~ecosyst+sourceOverlap, summarise, numObs=length(obsID))
summ.obs.st<-ddply(tmp, ~stud+sourceOverlap, summarise, numObs=length(obsID))
summ.obs.nfix<-ddply(tmp, ~nfix+sourceOverlap, summarise, numObs=length(obsID))
factorlist<-list(summ.obs.eco, summ.obs.st, summ.obs.nfix)
factortab<-ldply(factorlist)

newfilename<-'numObsTable_categorical.txt'
write.table(factortab, file=paste(figuresPath,newfilename, sep='/'), sep='\t')

tmp<-ddply(metaDataset,~obsID+traitCat,summarise,
      InvSpInvArea=unique(InvSpInvArea_cwm),
      NatArea=unique(NatArea_cwm),
      CWMDiff=unique(CWMDiff_cwm),
      sourceOverlap=unique(sourceOverlap))
tmp1<-tmp[!is.na(tmp$traitCat),]
summ.obs.continuous<-ddply(tmp1, ~traitCat+sourceOverlap, summarise,
                           nObs.Inv=sum(!is.na(InvSpInvArea)),
                           nObs.Ref=sum(!is.na(NatArea)),
                           nObs.Diff=sum(!is.na(CWMDiff)))
summ.obs.continuous
```

```
##      traitCat       sourceOverlap nObs.Inv nObs.Ref nObs.Diff
## 1          cn            NewPaper      104       76        76
## 2          cn PaperInPreviousMeta      104       79        79
## 3    littercn            NewPaper       25       26        26
## 4    littercn PaperInPreviousMeta       22       28        28
## 5 litterpercN            NewPaper       16       20        20
## 6 litterpercN PaperInPreviousMeta       34       39        39
## 7       percN            NewPaper      197      171       171
## 8       percN PaperInPreviousMeta      167      149       147
```

```
newfilename<-'numObsTable_continuous.txt'
write.table(summ.obs.continuous, file=paste(figuresPath,newfilename, sep='/'), sep='\t')
```

---

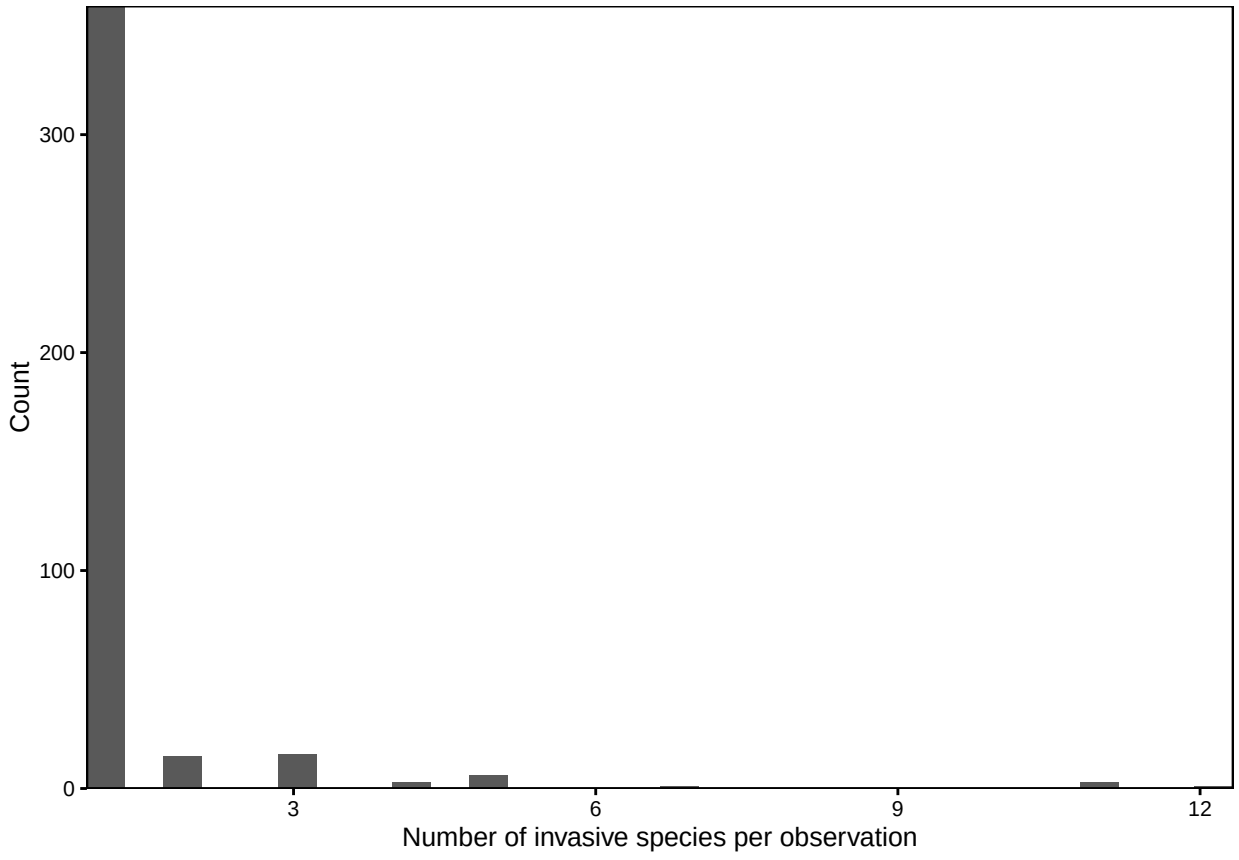
4. Plant species statistics

What is the distribution of invasive species per observation? Native species? Are certain invasive species over-represented?

```
#how many total species x study
paste('There are a total of', dim(species)[1], 'species x study')
```

```
## [1] "There are a total of 1998 species x study"
```

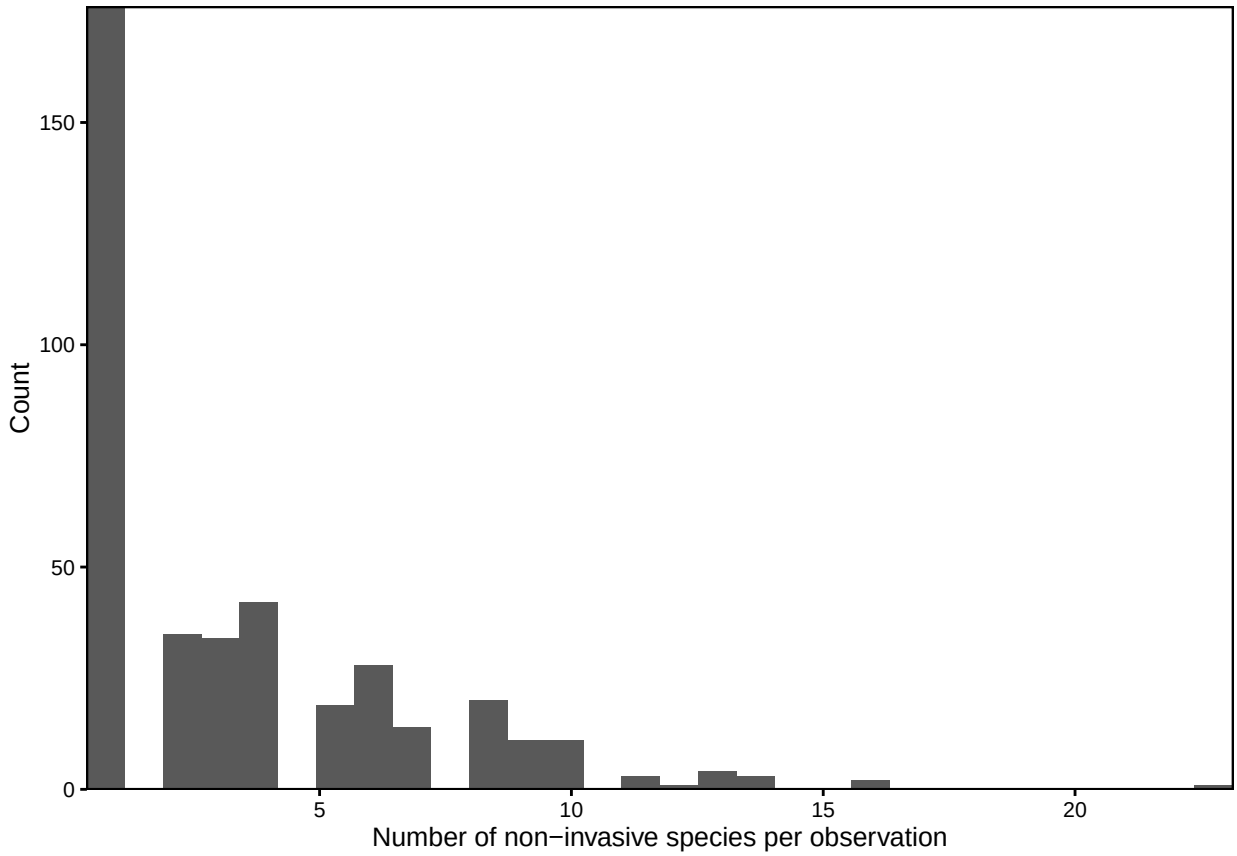
```
#what is the distribution of invasive species per observation?
summ.spp <- ddply(species,~obsID,summarise,
                  numTotalspp=length(obsID),
                  numInvspp=sum(spInvasive=='invasive' & spExotic=='exotic' & spFocal=='focal'),
                  numNonInvspp=sum(spInvasive=='not invasive'),
                  numOthers=numTotalspp-(numInvspp + numNonInvspp))

hist_Inv<-ggplot(summ.spp, aes(x=numInvspp)) + geom_histogram() +
  scale_y_continuous(expand=c(0,0)) + scale_x_continuous(expand=c(0,0)) +
  mytheme +
  ylab('Count') + xlab('Number of invasive species per observation')
hist_Inv; median(summ.spp$numInvspp); range(summ.spp$numInvspp)
```




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

```
## [1]  1 12
```

```
hist_Nat<-ggplot(summ.spp, aes(x=numNonInvspp)) + geom_histogram() +
  scale_y_continuous(expand=c(0,0)) + scale_x_continuous(expand=c(0,0)) +
  mytheme +
  ylab('Count') + xlab('Number of non-invasive species per observation')
hist_Nat; median(summ.spp$numNonInvspp); range(summ.spp$numNonInvspp)
```




```
## [1] 2
```

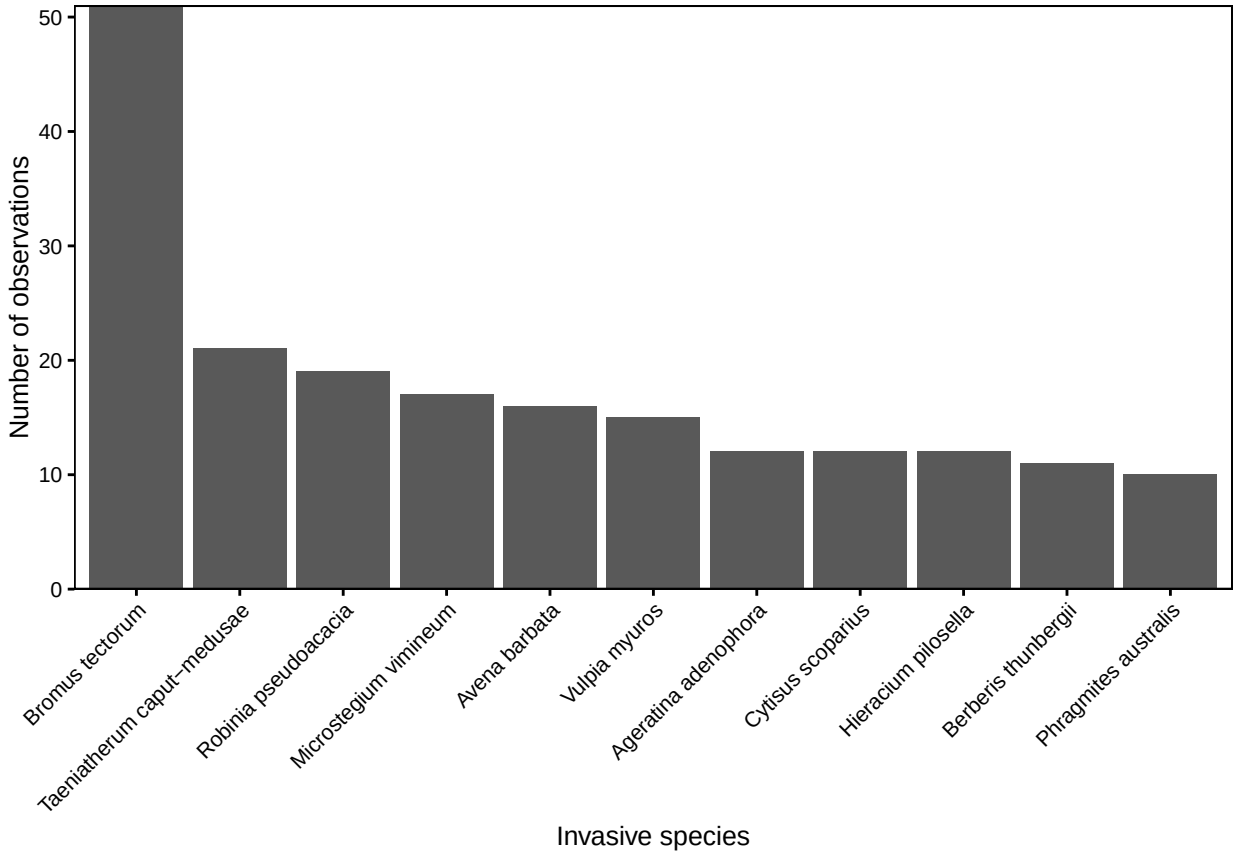
```
## [1]  1 23
```

```
#number of observations per species
summ.spp.nam <- ddply(species,~spName+spFocal+spExotic,summarise,
                  numObs=length(obsID),
                  numPapers=length(unique(paperID)))
spp.many<-summ.spp.nam[which(summ.spp.nam$numObs > 9 & summ.spp.nam$spFocal == 'focal'),] #more than 9
spp.many.o<-orderBy(~-numObs, spp.many)
spp.many.o
```

```
##                         spName spFocal spExotic numObs numPapers
## 130            Bromus tectorum   focal   exotic     51        19
## 616 Taeniatherum caput-medusae   focal   exotic     21         3
## 551       Robinia pseudoacacia   focal   exotic     19         7
## 405      Microstegium vimineum   focal   exotic     17         7
## 92               Avena barbata   focal   exotic     16         5
```

```
## 660             Vulpia myuros   focal   exotic     15         4
## 29        Ageratina adenophora   focal   exotic     12         1
## 216          Cytisus scoparius   focal   exotic     12         6
## 314        Hieracium pilosella   focal   exotic     12         5
## 103        Berberis thunbergii   focal   exotic     11         4
## 463       Phragmites australis   focal   exotic     10         5
```

```
positions<-spp.many.o$spName
pHist_spp<-ggplot(spp.many.o, aes(x=spName, y=numObs)) + geom_bar(stat='identity') +
  scale_y_continuous(expand=c(0,0)) + scale_x_discrete(limits = positions) +
  mytheme +  theme(axis.text.x=element_text(angle=45, hjust=1)) +
  ylab('Number of observations') + xlab('Invasive species')
pHist_spp
```




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

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

```
#which species appear both as exotic and native species in the dataset?
summ.spp <- ddply(species,~spName+spExotic, summarise,
                  numObs=length(obsID),
                  numPapers=length(unique(paperID)))
summ.spp.nam2 <- ddply(summ.spp,~spName,summarise,
                  numInvNat=length(spExotic))
summ.spp.nam2[summ.spp.nam2$numInvNat==2,] # if the length of spInvasive col==2, then there is native a
```

```
##                     spName numInvNat
## 5       Acacia longifolia         2
## 10           Acer negundo         2
## 61  Anthoxanthum odoratum         2
## 115          Briza maxima         2
## 118     Bromus hordeaceus         2
## 208     Cytisus scoparius         2
## 209    Dactylis glomerata         2
## 269   Festuca arundinacea         2
## 305        Holcus lanatus         2
## 451  Phragmites australis         2
## 477         Poa pratensis         2
## 498       Prunus serotina         2
## 553   Schedonorus phoenix         2
## 615          Trifolium sp         2
```

---

5. Cover data statistics

What percent of observations have measured cover data?

```
summ.cov.obs <- ddply(cover,~obsID,summarise,
                  numMeasured= sum(covQuality=='measured'))
numMeasured.obs<-sum(summ.cov.obs$numMeasured > 0) #number of observations with cover measurement value
numtotal.obs<-length(summ.cov.obs$numMeasured > 0) #total number of observations
cov.obs.perc<-round((numMeasured.obs / (numtotal.obs) ) *100, digits=2)
paste(cov.obs.perc, '% of observations have any cover data at all that was measured in the original pap
```

```
## [1] "27.97 % of observations have any cover data at all that was measured in the original paper"
```

```
#What is the frequency of cover observations for each cover measure type?
summ.cov <- ddply(cover,~covCat,summarise,
                  numMeas = length(obsID),
                  numObs=length(unique(obsID)),
                  numPapers=length(unique(paperID)))
orderBy(~-numMeas, summ.cov)
```

```
##        covCat numMeas numObs numPapers
## 3 sp_plantcov    1141    404       143
## 1  sp_biomass      81     32        13
## 2      sp_ind      80     22         6
```

```
#What units are commonly reported for each cover measure type?
summ.covUnit <- ddply(cover,~covCat+covUnit,summarise,
                      numMeas = length(obsID),
                      numObs=length(unique(obsID)),
                      numPapers=length(unique(paperID)))
COVCAT<-unique(summ.covUnit$covCat)
covUnitList<-list()
i<-0
for(i in 1:length(COVCAT)){
  subdf<-summ.covUnit[summ.covUnit$covCat==COVCAT[i],]
  covUnitList[[as.character(COVCAT[i])]]<-orderBy(~-numMeas, subdf)
}
covUnitList
```

```
## $sp_biomass
##       covCat covUnit numMeas numObs numPapers
## 2 sp_biomass    g/m2      38     13         6
## 5 sp_biomass  ind/ha      12      6         1
## 7 sp_biomass   m2/ha      11      3         1
## 4 sp_biomass   g/pot       6      3         1
## 6 sp_biomass   kg/m2       6      3         2
## 1 sp_biomass       %       4      2         1
## 3 sp_biomass  g/m2*y       4      2         1
##
## $sp_ind
##    covCat     covUnit numMeas numObs numPapers
## 11 sp_ind notReported      34      8         1
## 9  sp_ind    ind/30m2      19      1         1
## 13 sp_ind    stems/m2      11      5         1
## 12 sp_ind   plants/m2      10      5         1
## 8  sp_ind    ind/10m2       4      1         1
## 10 sp_ind      ind/ha       2      2         1
##
## $sp_plantcov
##         covCat covUnit numMeas numObs numPapers
## 14 sp_plantcov       %    1141    404       143
```

---

6. Trait data statistics

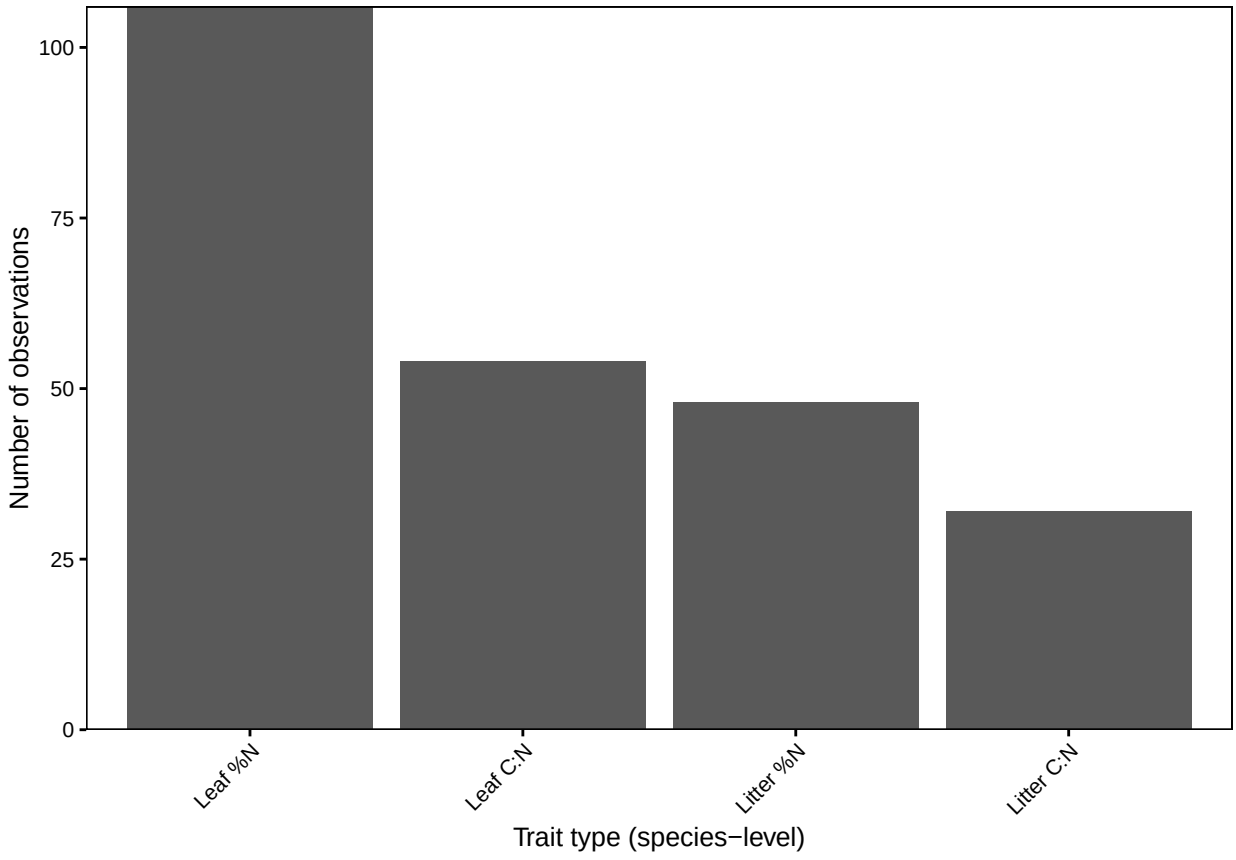
```
# what percent of observations had trait data reported within the original article?
n.ot<-length(unique(traits$obsID)) # number of observations with trait data
n.o<-length(unique(observations$obsID)) # total number of observations
tr.obs.perc<-round((n.ot/n.o) *100, digits=2) # percent of observations with trait data
paste(tr.obs.perc, '% of observations with species-level trait data from the original paper',collapse='
```

```
## [1] "34.41 % of observations with species-level trait data from the original paper"
```

```
summ.tr <- ddply(traits,~traitCat,summarise,
                 numObs = length(unique(obsID)),
                 numPapers = length(unique(paperID)))
summ.tr.o<-orderBy(~-numObs, summ.tr)
summ.tr.o
```

```
##         traitCat numObs numPapers
## 4       sp_percN    106        40
## 1          sp_cn     54        21
## 3 sp_litterpercN     48        22
## 2    sp_littercn     32        18
```

```
positions<-summ.tr.o$traitCat
pBar.tr<-ggplot(summ.tr.o, aes(x=traitCat, y=numObs)) + geom_bar(stat='identity') +
  scale_y_continuous(expand=c(0,0)) +
  scale_x_discrete(limits = positions,
                   labels = c("sp_percN" = "Leaf %N",
                              "sp_cn" = "Leaf C:N",
                              "sp_litterpercN" = "Litter %N",
                              "sp_littercn" = "Litter C:N")) +
  mytheme +  theme(axis.text.x=element_text(angle=45, hjust=1)) +
  ylab('Number of observations') + xlab('Trait type (species-level)')
pBar.tr
```



```
#What units and methods are commonly reported for each measurement?
summ.traitUnit <- ddply(traits,~traitCat+traitUnit,summarise,
                  numMeas = length(obsID),
                  numObs=length(unique(obsID)),
                  numPapers=length(unique(paperID)))
summ.traitUnit
```

```
##          traitCat traitUnit numMeas numObs numPapers
## 1           sp_cn         %      16      4         2
## 2           sp_cn     %C/%N      11      5         5
## 3           sp_cn molC/molN     128     45        14
## 4     sp_littercn     %C/%N       6      2         2
## 5     sp_littercn molC/molN      81     30        16
## 6  sp_litterpercN         %      85     29        14
## 7  sp_litterpercN      g/kg       6      4         2
## 8  sp_litterpercN      mg/g      38     15         6
## 9        sp_percN         %     160     58        23
## 10       sp_percN      g/kg      50     21         3
## 11       sp_percN      mg/g      66     23        11
## 12       sp_percN   mmol/kg       4      1         1
## 13       sp_percN      ug/g      14      2         1
## 14       sp_percN     ug/mg       2      1         1
```

```
TRAITCAT<-unique(summ.traitUnit$traitCat)
traitUnitList<-list()
i<-0
for(i in 1:length(TRAITCAT)){
  subdf<-summ.traitUnit[summ.traitUnit$traitCat==TRAITCAT[i],]
  traitUnitList[[as.character(TRAITCAT[i])]]<-orderBy(~-numMeas, subdf)
}
traitUnitList
```

```
## $sp_cn
##   traitCat traitUnit numMeas numObs numPapers
## 3    sp_cn molC/molN     128     45        14
## 1    sp_cn         %      16      4         2
## 2    sp_cn     %C/%N      11      5         5
##
## $sp_littercn
##      traitCat traitUnit numMeas numObs numPapers
## 5 sp_littercn molC/molN      81     30        16
## 4 sp_littercn     %C/%N       6      2         2
##
## $sp_litterpercN
##         traitCat traitUnit numMeas numObs numPapers
## 6 sp_litterpercN         %      85     29        14
## 8 sp_litterpercN      mg/g      38     15         6
## 7 sp_litterpercN      g/kg       6      4         2
##
## $sp_percN
##    traitCat traitUnit numMeas numObs numPapers
## 9  sp_percN         %     160     58        23
## 11 sp_percN      mg/g      66     23        11
```

```
## 10 sp_percN      g/kg      50     21         3
## 13 sp_percN      ug/g      14      2         1
## 12 sp_percN   mmol/kg       4      1         1
## 14 sp_percN     ug/mg       2      1         1
```

– Not Run – # 7. Soil measurement statistics

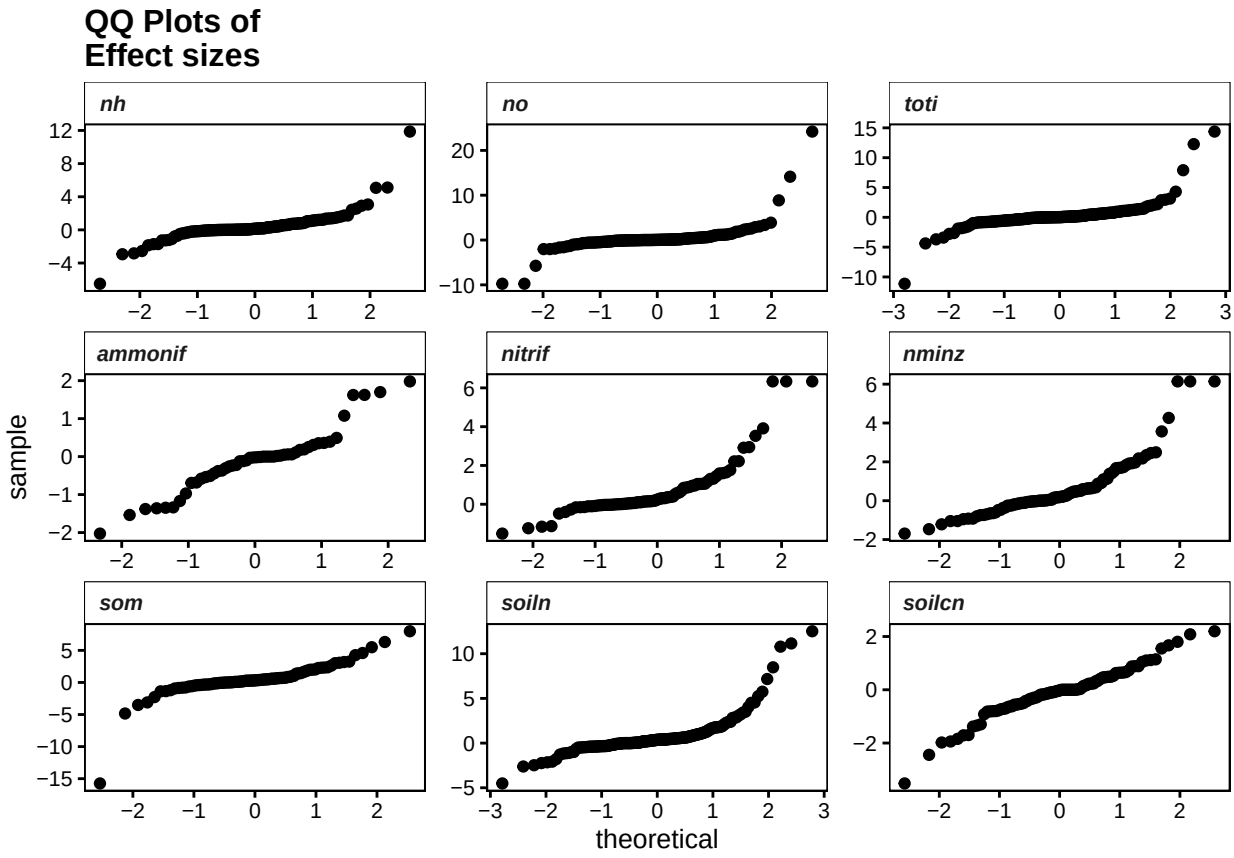
---

8. Effect size statistics

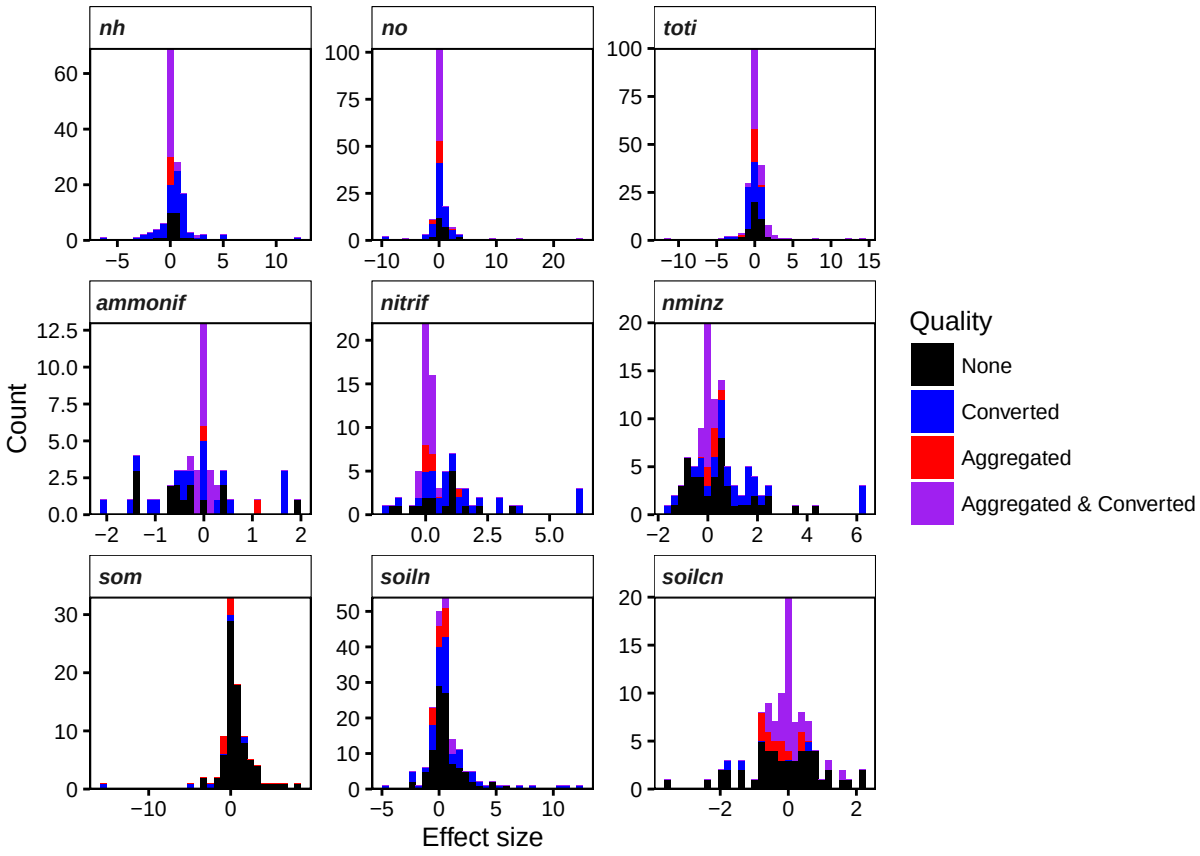
```
#what is the mean percent change in soil measures from reference to invaded?
#re-shape measures so that inv and nat are in the same column temporarily
tmp<-ddply(metaDataset, ~obsID+measCat, summarize,
      m1i = unique(m1i), #inv
      m2i = unique(m2i)) #ref
tmp$Diff<- tmp$m1i - tmp$m2i
tmp$percDiff<- ((tmp$Diff)/ abs(tmp$m2i))*100 #((inv - ref) / ref)*100
tmp2<-ddply(tmp, ~measCat, summarize,
            meanPercDiff = mean(percDiff, na.rm=T))
tmp2 #mean percent change in soil measures from reference to invaded
```

```
##   measCat meanPercDiff
## 1      nh    48.542117
## 2      no  4305.392775
## 3    toti    94.267813
## 4 ammonif  -976.086065
## 5  nitrif   370.881252
## 6   nminz   104.815257
## 7     som    36.523570
## 8   soiln    65.472844
## 9  soilcn    -2.224086
```

```
#Shapiro Test
# ddply(measures, ~measCat, summarise,
#       shapTest=shapiro.test(value)$p.value,
#       shapTest.Ln=shapiro.test(log(value+1))$p.value)
#none are normal according to Shapiro test

# Q-Q plots
tmp<-ddply(metaDataset, ~obsID+measCat, summarize,
      yi = unique(yi),
      measQuality = unique(measQuality))
qq<-ggplot(tmp, aes(sample=yi)) +
  facet_wrap(~measCat, scales='free', ncol=3) +
  stat_qq() + mytheme + ggtitle('QQ Plots of \nEffect sizes')
qq
```

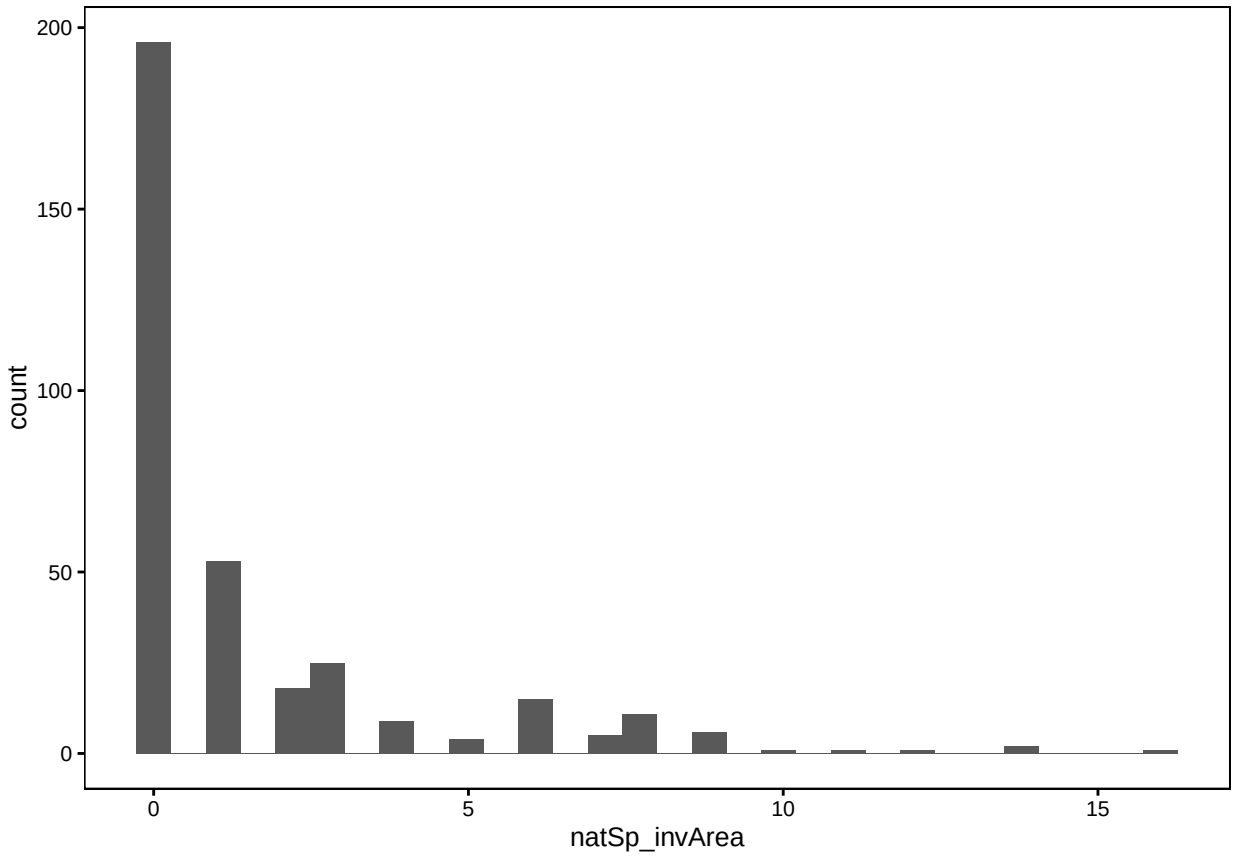



```
#Plot Quality Histograms
#re-order measQuality levels
tmp$measQuality <- factor(tmp$measQuality, levels = c('NoAgg.NoConv','NoAgg.Conv','Agg.NoConv','Agg.Conv

pHist_ESmeasQual<-ggplot(data=tmp, aes(x=yi,fill=measQuality)) + mytheme +
  facet_wrap(~measCat, scales='free', ncol=3) + geom_histogram() +
  scale_y_continuous(expand = c(0,0)) +
  scale_fill_manual(name = "Quality",
                    labels = c("Agg.Conv"="Aggregated & Converted",
                               "Agg.NoConv"="Aggregated",
                               "NoAgg.Conv"="Converted",
                               "NoAgg.NoConv"="None"),
                    values=c("Agg.Conv" = "purple",
                             "Agg.NoConv" = "red",
                             "NoAgg.Conv" = "blue",
                             "NoAgg.NoConv" = "black")) +
  ylab('Count') + xlab('Effect size')
pHist_ESmeasQual
```



9. CWM trait value statistics

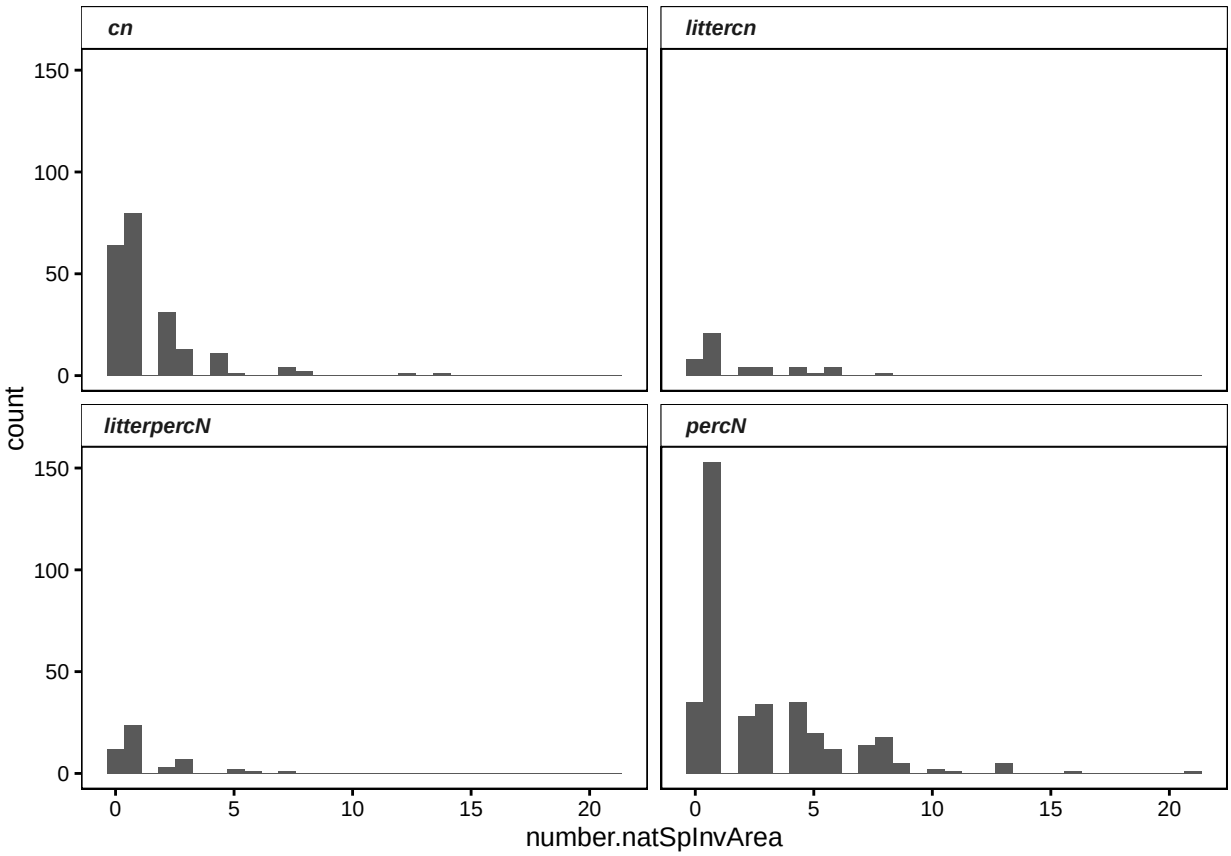
```
#in each observation, how many invasive and native species are present per invaded and native areas?
#just look at calculated datasets
cwm.calc<-subset(cwm, qualityCWMcalc == 'calculated')
summ.cwm1 <- ddply(cwm.calc,~obsID,summarise,
                   invSp_invArea=unique(n_invSp_invArea),
                   invSp_natArea=unique(n_invSp_natArea),
                   natSp_invArea=unique(n_natSp_invArea),
                   natSp_natArea=unique(n_natSp_natArea))
summ.cwm1$invSp<-summ.cwm1$invSp_invArea + summ.cwm1$invSp_natArea
summ.cwm1$natSp<-summ.cwm1$natSp_invArea + summ.cwm1$natSp_natArea
summ.cwm1$invArea<-summ.cwm1$invSp_invArea + summ.cwm1$natSp_invArea
summ.cwm1$natArea<-summ.cwm1$natSp_natArea + summ.cwm1$invSp_natArea
#what is the distribution of native species in invaded areas
p<-ggplot(summ.cwm1, aes(x=natSp_invArea)) +
  geom_histogram() + mytheme
p
```




```
#which studies have 1 or more native species in an invaded area?
n_studies_natSpInvArea<-length(which(summ.cwm1$natSp_invArea > 0))
n_studies_calcCWM<-dim(summ.cwm1)[1]
perc_studies_natSpInvArea<-round((n_studies_natSpInvArea / n_studies_calcCWM)*100, digits=2)
paste(n_studies_natSpInvArea,'studies have 1 or more non-invasive species in the invaded area, which is
```

```
## [1] "152 studies have 1 or more non-invasive species in the invaded area, which is  43.68 %"
```

```
#which cwms have native species in the invaded area?
#look at the distribution of native species in invaded areas
numberOfSpecies.cwm$number.natSpInvArea<-numberOfSpecies.cwm$number.spInvArea - numberOfSpecies.cwm$numb
weird.obs<-numberOfSpecies.cwm[numberOfSpecies.cwm$number.natSpInvArea<0,'obsID'] #check out these obse
#View(metaDataset[metaDataset$obsID %in% weird.obs,]) #ok it looks like I wasn't able to find trait dat
#just exclude those observations for now
numberOfSpecies.cwm1<-subset(numberOfSpecies.cwm, number.natSpInvArea>=0)
p<-ggplot(numberOfSpecies.cwm1, aes(x=number.natSpInvArea)) +
  facet_wrap(~traitCat)+
  geom_histogram() + mytheme
p
```




```
#how many obsIDs x traitCats have 0 native species in the invaded area?
summ.numberOfSp <- ddply(numberOfSpecies.cwm1,~traitCat, summarise,
                         numberObsIDs.natInInvArea=sum(number.natSpInvArea==0),
                         totalObsIDs=sum(!is.na(number.invSp)),
                         perc.sameVals= round((numberObsIDs.natInInvArea/totalObsIDs)*100, digits=2),
                         perc.diffVals=100-perc.sameVals)
summ.numberOfSp
```

```
##      traitCat numberObsIDs.natInInvArea totalObsIDs perc.sameVals
## 1          cn                        64         208         30.77
## 2    littercn                         8          47         17.02
## 3 litterpercN                        12          50         24.00
## 4       percN                        35         364          9.62
##   perc.diffVals
## 1         69.23
## 2         82.98
## 3         76.00
## 4         90.38
```

```
#column name key:
#numberObsIDs.natInInvArea = # of studies with no non-invasive species in the invaded area
#totalObsIDs = # of studies in total that have this type of trait data and are acceptable for inclusion
#perc.sameVals = % of studies where the invaded area CWM and invasive species' trait values will be exa
#perc.diffVals = % of studies where the invaded area CWM and invasive species' trait values will be uni

#how many CWMs are calculated vs reported by invType and traitType?
```

```
summ.cwm.calc <- ddply(cwm,~traitCat+invType,summarise,
                  nCWMs=sum(!is.na(cwm)),
                  nCWMs_calculated=sum(qualityCWMcalc=='calculated' & !is.na(cwm)),
                  nCWMs_reported=sum(qualityCWMcalc=='reported' & !is.na(cwm)),
                  total.check=sum(nCWMs_calculated,nCWMs_reported))
summ.cwm.calc
```

```
##       traitCat      invType nCWMs nCWMs_calculated nCWMs_reported
## 1           cn      InvArea   212              198             14
## 2           cn InvSpInvArea   208              198             10
## 3           cn      NatArea   155              141             14
## 4     littercn      InvArea    60               40             20
## 5     littercn InvSpInvArea    47               40              7
## 6     littercn      NatArea    54               34             20
## 7  litterpercN      InvArea    66               42             24
## 8  litterpercN InvSpInvArea    50               42              8
## 9  litterpercN      NatArea    59               35             24
## 10       percN      InvArea   342              316             26
## 11       percN InvSpInvArea   364              316             48
## 12       percN      NatArea   320              295             25
##    total.check
## 1          212
## 2          208
## 3          155
## 4           60
## 5           47
## 6           54
## 7           66
## 8           50
## 9           59
## 10         342
## 11         364
## 12         320
```

```
newfilename<-'cwm_calc.txt'
write.table(summ.cwm.calc, file=paste(figuresPath,newfilename, sep='/'), sep='\t')


#of the CWMs that are calculated, how many CWMs have each level of quality?
summ.cwm.qual <- ddply(cwm,~traitCat+invType+
                         qualityCWMcalc+ #reported or calculated
                         binMeasCov+bin1spCov+ #cover quality
                         binOrigTr+binTryGS, #trait quality
                       summarise,
                  nCWMs=sum(!is.na(cwm)))
summ.cwm.qual
```

```
##        traitCat      invType qualityCWMcalc        binMeasCov
## 1            cn      InvArea     calculated greaterThan25perc
## 2            cn      InvArea     calculated greaterThan25perc
## 3            cn      InvArea     calculated greaterThan25perc
## 4            cn      InvArea     calculated greaterThan25perc
## 5            cn      InvArea     calculated greaterThan25perc
```

```
## 6           cn      InvArea     calculated    lessThan25perc
## 7           cn      InvArea     calculated    lessThan25perc
## 8           cn      InvArea     calculated    lessThan25perc
## 9           cn      InvArea     calculated    lessThan25perc
## 10          cn      InvArea     calculated    lessThan25perc
## 11          cn      InvArea     calculated    lessThan25perc
## 12          cn      InvArea       reported             <NA>
## 13          cn InvSpInvArea     calculated greaterThan25perc
## 14          cn InvSpInvArea     calculated greaterThan25perc
## 15          cn InvSpInvArea     calculated greaterThan25perc
## 16          cn InvSpInvArea     calculated greaterThan25perc
## 17          cn InvSpInvArea     calculated    lessThan25perc
## 18          cn InvSpInvArea     calculated    lessThan25perc
## 19          cn InvSpInvArea     calculated    lessThan25perc
## 20          cn InvSpInvArea     calculated    lessThan25perc
## 21          cn InvSpInvArea     calculated    lessThan25perc
## 22          cn InvSpInvArea       reported greaterThan25perc
## 23          cn InvSpInvArea       reported    lessThan25perc
## 24          cn InvSpInvArea       reported    lessThan25perc
## 25          cn InvSpInvArea       reported             <NA>
## 26          cn      NatArea     calculated greaterThan25perc
## 27          cn      NatArea     calculated greaterThan25perc
## 28          cn      NatArea     calculated greaterThan25perc
## 29          cn      NatArea     calculated greaterThan25perc
## 30          cn      NatArea     calculated    lessThan25perc
## 31          cn      NatArea     calculated    lessThan25perc
## 32          cn      NatArea     calculated    lessThan25perc
## 33          cn      NatArea     calculated    lessThan25perc
## 34          cn      NatArea     calculated    lessThan25perc
## 35          cn      NatArea     calculated    lessThan25perc
## 36          cn      NatArea     calculated             <NA>
## 37          cn      NatArea       reported             <NA>
## 38    littercn      InvArea     calculated greaterThan25perc
## 39    littercn      InvArea     calculated greaterThan25perc
## 40    littercn      InvArea     calculated greaterThan25perc
## 41    littercn      InvArea     calculated    lessThan25perc
## 42    littercn      InvArea     calculated    lessThan25perc
## 43    littercn      InvArea     calculated    lessThan25perc
## 44    littercn      InvArea     calculated    lessThan25perc
## 45    littercn      InvArea     calculated    lessThan25perc
## 46    littercn      InvArea       reported             <NA>
## 47    littercn InvSpInvArea     calculated greaterThan25perc
## 48    littercn InvSpInvArea     calculated greaterThan25perc
## 49    littercn InvSpInvArea     calculated    lessThan25perc
## 50    littercn InvSpInvArea     calculated    lessThan25perc
## 51    littercn InvSpInvArea     calculated    lessThan25perc
## 52    littercn InvSpInvArea     calculated    lessThan25perc
## 53    littercn InvSpInvArea       reported    lessThan25perc
## 54    littercn InvSpInvArea       reported             <NA>
## 55    littercn      NatArea     calculated greaterThan25perc
## 56    littercn      NatArea     calculated greaterThan25perc
## 57    littercn      NatArea     calculated greaterThan25perc
## 58    littercn      NatArea     calculated greaterThan25perc
## 59    littercn      NatArea     calculated    lessThan25perc
```

```
## 60     littercn      NatArea     calculated    lessThan25perc
## 61     littercn      NatArea     calculated    lessThan25perc
## 62     littercn      NatArea     calculated    lessThan25perc
## 63     littercn      NatArea     calculated    lessThan25perc
## 64     littercn      NatArea     calculated    lessThan25perc
## 65     littercn      NatArea     calculated    lessThan25perc
## 66     littercn      NatArea     calculated              <NA>
## 67     littercn      NatArea       reported              <NA>
## 68  litterpercN      InvArea     calculated greaterThan25perc
## 69  litterpercN      InvArea     calculated greaterThan25perc
## 70  litterpercN      InvArea     calculated    lessThan25perc
## 71  litterpercN      InvArea     calculated    lessThan25perc
## 72  litterpercN      InvArea     calculated    lessThan25perc
## 73  litterpercN      InvArea     calculated    lessThan25perc
## 74  litterpercN      InvArea       reported              <NA>
## 75  litterpercN InvSpInvArea     calculated greaterThan25perc
## 76  litterpercN InvSpInvArea     calculated greaterThan25perc
## 77  litterpercN InvSpInvArea     calculated    lessThan25perc
## 78  litterpercN InvSpInvArea     calculated    lessThan25perc
## 79  litterpercN InvSpInvArea     calculated    lessThan25perc
## 80  litterpercN InvSpInvArea     calculated    lessThan25perc
## 81  litterpercN InvSpInvArea       reported    lessThan25perc
## 82  litterpercN InvSpInvArea       reported    lessThan25perc
## 83  litterpercN InvSpInvArea       reported              <NA>
## 84  litterpercN      NatArea     calculated greaterThan25perc
## 85  litterpercN      NatArea     calculated greaterThan25perc
## 86  litterpercN      NatArea     calculated    lessThan25perc
## 87  litterpercN      NatArea     calculated    lessThan25perc
## 88  litterpercN      NatArea     calculated    lessThan25perc
## 89  litterpercN      NatArea     calculated    lessThan25perc
## 90  litterpercN      NatArea     calculated    lessThan25perc
## 91  litterpercN      NatArea     calculated              <NA>
## 92  litterpercN      NatArea       reported              <NA>
## 93        percN      InvArea     calculated greaterThan25perc
## 94        percN      InvArea     calculated greaterThan25perc
## 95        percN      InvArea     calculated greaterThan25perc
## 96        percN      InvArea     calculated greaterThan25perc
## 97        percN      InvArea     calculated greaterThan25perc
## 98        percN      InvArea     calculated greaterThan25perc
## 99        percN      InvArea     calculated greaterThan25perc
## 100       percN      InvArea     calculated    lessThan25perc
## 101       percN      InvArea     calculated    lessThan25perc
## 102       percN      InvArea     calculated    lessThan25perc
## 103       percN      InvArea     calculated    lessThan25perc
## 104       percN      InvArea     calculated    lessThan25perc
## 105       percN      InvArea     calculated    lessThan25perc
## 106       percN      InvArea     calculated              <NA>
## 107       percN      InvArea       reported              <NA>
## 108       percN InvSpInvArea     calculated greaterThan25perc
## 109       percN InvSpInvArea     calculated greaterThan25perc
## 110       percN InvSpInvArea     calculated greaterThan25perc
## 111       percN InvSpInvArea     calculated greaterThan25perc
## 112       percN InvSpInvArea     calculated    lessThan25perc
## 113       percN InvSpInvArea     calculated    lessThan25perc
```

```
## 114       percN InvSpInvArea     calculated    lessThan25perc
## 115       percN InvSpInvArea     calculated    lessThan25perc
## 116       percN InvSpInvArea     calculated    lessThan25perc
## 117       percN InvSpInvArea     calculated    lessThan25perc
## 118       percN InvSpInvArea     calculated             <NA>
## 119       percN InvSpInvArea       reported greaterThan25perc
## 120       percN InvSpInvArea       reported    lessThan25perc
## 121       percN InvSpInvArea       reported    lessThan25perc
## 122       percN InvSpInvArea       reported             <NA>
## 123       percN      NatArea     calculated greaterThan25perc
## 124       percN      NatArea     calculated greaterThan25perc
## 125       percN      NatArea     calculated greaterThan25perc
## 126       percN      NatArea     calculated greaterThan25perc
## 127       percN      NatArea     calculated greaterThan25perc
## 128       percN      NatArea     calculated greaterThan25perc
## 129       percN      NatArea     calculated greaterThan25perc
## 130       percN      NatArea     calculated greaterThan25perc
## 131       percN      NatArea     calculated    lessThan25perc
## 132       percN      NatArea     calculated    lessThan25perc
## 133       percN      NatArea     calculated    lessThan25perc
## 134       percN      NatArea     calculated    lessThan25perc
## 135       percN      NatArea     calculated    lessThan25perc
## 136       percN      NatArea     calculated    lessThan25perc
## 137       percN      NatArea     calculated    lessThan25perc
## 138       percN      NatArea     calculated    lessThan25perc
## 139       percN      NatArea     calculated             <NA>
## 140       percN      NatArea       reported             <NA>
##             bin1spCov         binOrigTr          binTryGS nCWMs
## 1   greaterThan25perc greaterThan25perc    lessThan25perc    12
## 2   greaterThan25perc    lessThan25perc greaterThan25perc     1
## 3   greaterThan25perc    lessThan25perc    lessThan25perc     6
## 4      lessThan25perc greaterThan25perc    lessThan25perc     1
## 5      lessThan25perc    lessThan25perc    lessThan25perc     4
## 6   greaterThan25perc greaterThan25perc    lessThan25perc    34
## 7   greaterThan25perc    lessThan25perc greaterThan25perc    55
## 8   greaterThan25perc    lessThan25perc    lessThan25perc    55
## 9      lessThan25perc greaterThan25perc    lessThan25perc     3
## 10     lessThan25perc    lessThan25perc greaterThan25perc    23
## 11     lessThan25perc    lessThan25perc    lessThan25perc     4
## 12               <NA>              <NA>              <NA>    14
## 13  greaterThan25perc greaterThan25perc    lessThan25perc    11
## 14  greaterThan25perc    lessThan25perc greaterThan25perc     1
## 15  greaterThan25perc    lessThan25perc    lessThan25perc     6
## 16     lessThan25perc    lessThan25perc    lessThan25perc     4
## 17  greaterThan25perc greaterThan25perc    lessThan25perc    39
## 18  greaterThan25perc    lessThan25perc greaterThan25perc    60
## 19  greaterThan25perc    lessThan25perc    lessThan25perc    58
## 20     lessThan25perc    lessThan25perc greaterThan25perc    15
## 21     lessThan25perc    lessThan25perc    lessThan25perc     4
## 22  greaterThan25perc    lessThan25perc    lessThan25perc     4
## 23  greaterThan25perc greaterThan25perc    lessThan25perc     1
## 24  greaterThan25perc    lessThan25perc    lessThan25perc     5
## 25               <NA>              <NA>              <NA>     0
## 26  greaterThan25perc greaterThan25perc    lessThan25perc    12
```

```
## 27  greaterThan25perc    lessThan25perc    lessThan25perc     5
## 28     lessThan25perc    lessThan25perc greaterThan25perc     1
## 29     lessThan25perc    lessThan25perc    lessThan25perc     4
## 30  greaterThan25perc greaterThan25perc    lessThan25perc    23
## 31  greaterThan25perc    lessThan25perc greaterThan25perc     9
## 32  greaterThan25perc    lessThan25perc    lessThan25perc    47
## 33     lessThan25perc greaterThan25perc    lessThan25perc    14
## 34     lessThan25perc    lessThan25perc greaterThan25perc    21
## 35     lessThan25perc    lessThan25perc    lessThan25perc     5
## 36               <NA>              <NA>              <NA>     0
## 37               <NA>              <NA>              <NA>    14
## 38  greaterThan25perc greaterThan25perc    lessThan25perc     7
## 39     lessThan25perc    lessThan25perc greaterThan25perc     1
## 40     lessThan25perc    lessThan25perc    lessThan25perc     1
## 41  greaterThan25perc greaterThan25perc    lessThan25perc    11
## 42  greaterThan25perc    lessThan25perc greaterThan25perc    13
## 43  greaterThan25perc    lessThan25perc    lessThan25perc     4
## 44     lessThan25perc greaterThan25perc    lessThan25perc     2
## 45     lessThan25perc    lessThan25perc greaterThan25perc     1
## 46               <NA>              <NA>              <NA>    20
## 47  greaterThan25perc greaterThan25perc    lessThan25perc     6
## 48  greaterThan25perc    lessThan25perc greaterThan25perc     2
## 49  greaterThan25perc greaterThan25perc    lessThan25perc    14
## 50  greaterThan25perc    lessThan25perc greaterThan25perc    13
## 51  greaterThan25perc    lessThan25perc    lessThan25perc     4
## 52     lessThan25perc greaterThan25perc    lessThan25perc     1
## 53  greaterThan25perc greaterThan25perc    lessThan25perc     7
## 54               <NA>              <NA>              <NA>     0
## 55  greaterThan25perc greaterThan25perc    lessThan25perc     6
## 56     lessThan25perc greaterThan25perc    lessThan25perc     1
## 57     lessThan25perc    lessThan25perc greaterThan25perc     1
## 58     lessThan25perc    lessThan25perc    lessThan25perc     1
## 59  greaterThan25perc greaterThan25perc greaterThan25perc     2
## 60  greaterThan25perc greaterThan25perc    lessThan25perc     6
## 61  greaterThan25perc    lessThan25perc    lessThan25perc     6
## 62     lessThan25perc greaterThan25perc greaterThan25perc     1
## 63     lessThan25perc greaterThan25perc    lessThan25perc     4
## 64     lessThan25perc    lessThan25perc greaterThan25perc     2
## 65     lessThan25perc    lessThan25perc    lessThan25perc     4
## 66               <NA>              <NA>              <NA>     0
## 67               <NA>              <NA>              <NA>    20
## 68  greaterThan25perc greaterThan25perc    lessThan25perc     6
## 69     lessThan25perc    lessThan25perc    lessThan25perc     2
## 70  greaterThan25perc greaterThan25perc    lessThan25perc    15
## 71  greaterThan25perc    lessThan25perc    lessThan25perc    10
## 72     lessThan25perc greaterThan25perc    lessThan25perc     7
## 73     lessThan25perc    lessThan25perc    lessThan25perc     2
## 74               <NA>              <NA>              <NA>    24
## 75  greaterThan25perc greaterThan25perc    lessThan25perc     6
## 76     lessThan25perc    lessThan25perc    lessThan25perc     2
## 77  greaterThan25perc greaterThan25perc    lessThan25perc    19
## 78  greaterThan25perc    lessThan25perc    lessThan25perc    10
## 79     lessThan25perc greaterThan25perc    lessThan25perc     3
## 80     lessThan25perc    lessThan25perc    lessThan25perc     2
```

```
## 81  greaterThan25perc greaterThan25perc    lessThan25perc     6
## 82  greaterThan25perc    lessThan25perc    lessThan25perc     2
## 83               <NA>              <NA>              <NA>     0
## 84  greaterThan25perc greaterThan25perc    lessThan25perc     5
## 85     lessThan25perc greaterThan25perc    lessThan25perc     1
## 86  greaterThan25perc greaterThan25perc    lessThan25perc    12
## 87  greaterThan25perc    lessThan25perc    lessThan25perc     5
## 88     lessThan25perc greaterThan25perc    lessThan25perc     8
## 89     lessThan25perc    lessThan25perc greaterThan25perc     1
## 90     lessThan25perc    lessThan25perc    lessThan25perc     3
## 91               <NA>              <NA>              <NA>     0
## 92               <NA>              <NA>              <NA>    24
## 93  greaterThan25perc greaterThan25perc greaterThan25perc     1
## 94  greaterThan25perc greaterThan25perc    lessThan25perc    15
## 95  greaterThan25perc    lessThan25perc greaterThan25perc    26
## 96  greaterThan25perc    lessThan25perc    lessThan25perc     7
## 97     lessThan25perc greaterThan25perc greaterThan25perc     4
## 98     lessThan25perc    lessThan25perc greaterThan25perc     9
## 99     lessThan25perc    lessThan25perc    lessThan25perc     2
## 100 greaterThan25perc greaterThan25perc    lessThan25perc    37
## 101 greaterThan25perc    lessThan25perc greaterThan25perc   133
## 102 greaterThan25perc    lessThan25perc    lessThan25perc    30
## 103    lessThan25perc greaterThan25perc greaterThan25perc     1
## 104    lessThan25perc greaterThan25perc    lessThan25perc     5
## 105    lessThan25perc    lessThan25perc greaterThan25perc    46
## 106              <NA>              <NA>              <NA>     0
## 107              <NA>              <NA>              <NA>    26
## 108 greaterThan25perc greaterThan25perc    lessThan25perc    14
## 109 greaterThan25perc    lessThan25perc greaterThan25perc    26
## 110 greaterThan25perc    lessThan25perc    lessThan25perc    12
## 111    lessThan25perc    lessThan25perc greaterThan25perc     6
## 112 greaterThan25perc greaterThan25perc greaterThan25perc     2
## 113 greaterThan25perc greaterThan25perc    lessThan25perc    43
## 114 greaterThan25perc    lessThan25perc greaterThan25perc   144
## 115 greaterThan25perc    lessThan25perc    lessThan25perc    48
## 116    lessThan25perc greaterThan25perc    lessThan25perc     2
## 117    lessThan25perc    lessThan25perc greaterThan25perc    19
## 118              <NA>              <NA>              <NA>     0
## 119 greaterThan25perc    lessThan25perc greaterThan25perc     4
## 120 greaterThan25perc greaterThan25perc    lessThan25perc     4
## 121 greaterThan25perc    lessThan25perc greaterThan25perc    40
## 122              <NA>              <NA>              <NA>     0
## 123 greaterThan25perc greaterThan25perc greaterThan25perc     2
## 124 greaterThan25perc greaterThan25perc    lessThan25perc    14
## 125 greaterThan25perc    lessThan25perc greaterThan25perc    23
## 126 greaterThan25perc    lessThan25perc    lessThan25perc     5
## 127    lessThan25perc greaterThan25perc greaterThan25perc     3
## 128    lessThan25perc greaterThan25perc    lessThan25perc     1
## 129    lessThan25perc    lessThan25perc greaterThan25perc     9
## 130    lessThan25perc    lessThan25perc    lessThan25perc     4
## 131 greaterThan25perc greaterThan25perc greaterThan25perc     3
## 132 greaterThan25perc greaterThan25perc    lessThan25perc    29
## 133 greaterThan25perc    lessThan25perc greaterThan25perc    94
## 134 greaterThan25perc    lessThan25perc    lessThan25perc    25
```

```
## 135    lessThan25perc greaterThan25perc greaterThan25perc     1
## 136    lessThan25perc greaterThan25perc    lessThan25perc    14
## 137    lessThan25perc    lessThan25perc greaterThan25perc    59
## 138    lessThan25perc    lessThan25perc    lessThan25perc     9
## 139              <NA>              <NA>              <NA>     0
## 140              <NA>              <NA>              <NA>    25
```

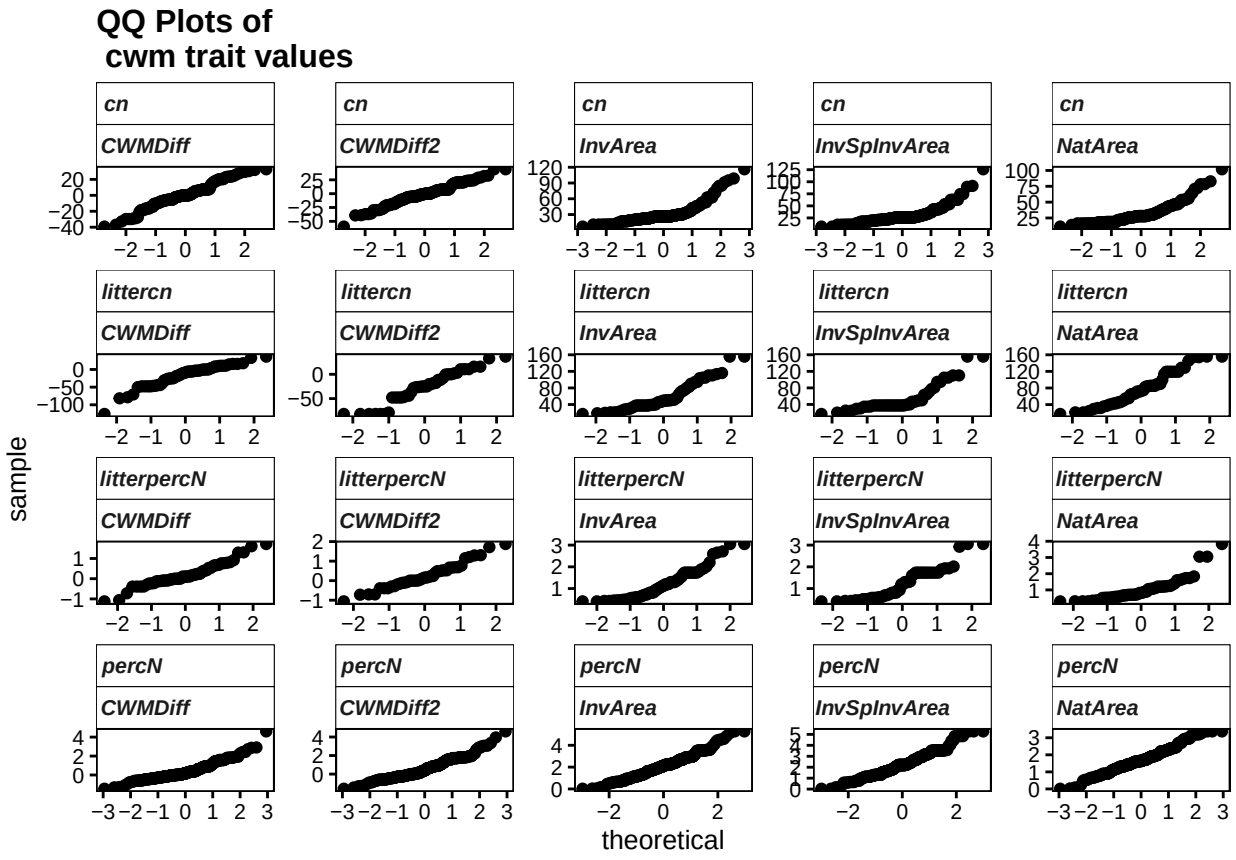
```
c.summ<-dcast(summ.cwm.qual, invType+qualityCWMcalc+binMeasCov+bin1spCov+binOrigTr+binTryGS~traitCat, va
newfilename<-'cwm_qual.txt'
write.table(c.summ, file=paste(figuresPath,newfilename, sep='/'), sep='\t')

#by qualRank
summ.cwm.rank <- ddply(cwm,~traitCat+invType+qualRank, summarise,
                  nCWMs=sum(!is.na(cwm)))
summ.cwm.rank
```

```
##       traitCat      invType qualRank nCWMs
## 1           cn      InvArea        0     4
## 2           cn      InvArea        1    85
## 3           cn      InvArea        2    96
## 4           cn      InvArea        3    13
## 5           cn      InvArea        4    14
## 6           cn InvSpInvArea        0     4
## 7           cn InvSpInvArea        1    77
## 8           cn InvSpInvArea        2   105
## 9           cn InvSpInvArea        3    12
## 10          cn InvSpInvArea        4    10
## 11          cn      NatArea        0     5
## 12          cn      NatArea        1    86
## 13          cn      NatArea        2    38
## 14          cn      NatArea        3    12
## 15          cn      NatArea        4    14
## 16    littercn      InvArea        1     8
## 17    littercn      InvArea        2    25
## 18    littercn      InvArea        3     7
## 19    littercn      InvArea        4    20
## 20    littercn InvSpInvArea        1     5
## 21    littercn InvSpInvArea        2    27
## 22    littercn InvSpInvArea        3     8
## 23    littercn InvSpInvArea        4     7
## 24    littercn      NatArea        0     4
## 25    littercn      NatArea        1    14
## 26    littercn      NatArea        2    10
## 27    littercn      NatArea        3     6
## 28    littercn      NatArea        4    20
## 29 litterpercN      InvArea        0     2
## 30 litterpercN      InvArea        1    19
## 31 litterpercN      InvArea        2    15
## 32 litterpercN      InvArea        3     6
## 33 litterpercN      InvArea        4    24
## 34 litterpercN InvSpInvArea        0     2
## 35 litterpercN InvSpInvArea        1    15
## 36 litterpercN InvSpInvArea        2    19
## 37 litterpercN InvSpInvArea        3     6
```

```
## 38 litterpercN InvSpInvArea        4     8
## 39 litterpercN      NatArea        0     3
## 40 litterpercN      NatArea        1    14
## 41 litterpercN      NatArea        2    13
## 42 litterpercN      NatArea        3     5
## 43 litterpercN      NatArea        4    24
## 44       percN      InvArea        0     0
## 45       percN      InvArea        1    84
## 46       percN      InvArea        2   190
## 47       percN      InvArea        3    42
## 48       percN      InvArea        4    26
## 49       percN InvSpInvArea        0     0
## 50       percN InvSpInvArea        1    69
## 51       percN InvSpInvArea        2   207
## 52       percN InvSpInvArea        3    40
## 53       percN InvSpInvArea        4    48
## 54       percN      NatArea        0     9
## 55       percN      NatArea        1   103
## 56       percN      NatArea        2   144
## 57       percN      NatArea        3    39
## 58       percN      NatArea        4    25
```

```
c.summ2<-dcast(summ.cwm.rank, invType+qualRank~traitCat, value.var='nCWMs')
newfilename<-'cwm_qual_rank.txt'
write.table(c.summ2, file=paste(figuresPath,newfilename, sep='/'), sep='\t')

#look at whether the cwm values are normally distributed
#re-shape measures so that inv and nat are in the same column temporarily
tmp<-ddply(metaDataset, ~obsID+traitCat, summarize,
      InvArea_cwm = unique(InvArea_cwm),
      InvSpInvArea_cwm = unique(InvSpInvArea_cwm),
      NatArea_cwm = unique(NatArea_cwm),
      CWMDiff_cwm = unique(CWMDiff_cwm),
      CWMDiff2_cwm = unique(CWMDiff2_cwm),
      InvArea_qualRank = unique(InvArea_qualRank),
      InvSpInvArea_qualRank = unique(InvSpInvArea_qualRank),
      NatArea_qualRank = unique(NatArea_qualRank),
      CWMDiff_qualRank = unique(CWMDiff_qualRank),
      CWMDiff2_qualRank = unique(CWMDiff2_qualRank))
tmp$obsID<-as.factor(tmp$obsID)
m.tmp<-melt(tmp, idcols=c('obsID','traitCat'))
m.tmp$dataType<-rep(NA,length(dim(m.tmp)[1])) #dataType
m.tmp[grepl('_qualRank', m.tmp$variable),'dataType']<-'qualRank'
m.tmp[grepl('_cwm', m.tmp$variable),'dataType']<-'cwm'
m.tmp$invType<-rep(NA,length(dim(m.tmp)[1])) #invType
m.tmp[grepl('InvArea', m.tmp$variable),'invType']<-'InvArea'
m.tmp[grepl('InvSpInvArea', m.tmp$variable),'invType']<-'InvSpInvArea'
m.tmp[grepl('NatArea', m.tmp$variable),'invType']<-'NatArea'
m.tmp[grepl('CWMDiff', m.tmp$variable),'invType']<-'CWMDiff'
m.tmp[grepl('CWMDiff2', m.tmp$variable),'invType']<-'CWMDiff2'
c.tmp<-dcast(m.tmp, obsID+traitCat+invType~dataType)
c.tmp<-c.tmp[!is.na(c.tmp$cwm),]
#View(c.tmp)
```

```
# #Shapiro Test
# ddply(cwm, ~traitCat, summarise,
#       shapTest=shapiro.test(cwm)$p.value,
#       shapTestLn=shapiro.test(log10(cwm))$p.value)
# #none are normal according to Shapiro test

# Q-Q plots
qq<-ggplot(c.tmp, aes(sample=cwm)) +
  facet_wrap(~traitCat+invType, scales='free', ncol=5) +
  stat_qq() + mytheme + ggtitle('QQ Plots of \n cwm trait values')
qq
```



---

10. Table of articles for supplementary info

```
#subset accepted papers and the relevant columns
#papers.a is from chunk1, above

#add info about the number of observations per paper
nObs<-ddply(observations, ~paperID, summarise,
      nObs=length(obsID))
papers.a1<-merge(papers.a, nObs)
```

```
#add info about the number and type of measurements per paper
paperID<-ldply(strsplit(as.character(measures$obsID), '.', fixed=T))[,1]
measures$paperID<-paperID
tmp<-ddply(measures, ~paperID+measCat, summarise,
      nObs=length(obsID))
tmp2<-dcast(tmp, paperID ~ measCat, value.var='nObs', fill=0)
papers.a2<-merge(papers.a1, tmp2)

newfilename<-'paperSummary.txt'
write.table(papers.a2, file=paste(figuresPath,newfilename, sep='/'), sep='\t')
```

---

11. How much missing species info was there for the observations that needed CWM values calculated?

```
### 1: Determine which spIDs have cover data, which don't

#Invasive species
#spID that are missing cover data
invSp.cover.na<-cover[is.na(cover$stdmeanInv) &
                      cover$covInvasive=='invasive' & cover$covFocal=='focal', c('spID')]
invSp.cover.nSp.na<-length(unique(invSp.cover.na[!is.na(invSp.cover.na)]))
#spIDs that have cover data
invSp.cover<-cover[cover$stdmeanInv>0 & !is.na(cover$stdmeanInv) &
                      cover$covInvasive=='invasive' & cover$covFocal=='focal', c('spID')]

#Invaded area
#spID that are missing cover data
inv.cover.na<-cover[is.na(cover$stdmeanInv),c('spID')]
inv.cover.nSp.na<-length(unique(inv.cover.na[!is.na(inv.cover.na)]))
#spIDs that have cover
inv.cover<-cover[cover$stdmeanInv>0 & !is.na(cover$stdmeanInv),c('spID')]

#Reference area
#spID that are missing cover data
ref.cover.na<-cover[is.na(cover$stdmeanNat), c('spID')]
ref.cover.nSp.na<-length(unique(ref.cover.na[!is.na(ref.cover.na)]))
#spIDs that have cover data
ref.cover<-cover[cover$stdmeanNat>0 & !is.na(cover$stdmeanNat),c('spID')]

#summary
missingCov.tab<-data.frame(communityType=c('InvSp','Inv','Ref'),
                           nSp.missingCover=c(invSp.cover.nSp.na, inv.cover.nSp.na, ref.cover.nSp.na))
missingCov.tab
```

```
##   communityType nSp.missingCover
## 1         InvSp               41
## 2           Inv              100
## 3           Ref               92
```

```
### 2: Of the spIDs that have cover data, which have trait data?
IdentifyspID.traits<-function(sptrait.df, spID.vec, nSp.missingCoverData){

  #calculate the number of unique species entries in the spID.vec
  nSp<-length(unique(spID.vec[!is.na(spID.vec)]))

  #identify spIDs in the species x trait dataset
  tmp.spIDs<-sptrait.df[sptrait.df$spID %in% spID.vec,]

  #pull out the spIDs that have data for each trait type
  percN.spIDs<-tmp.spIDs[!is.na(tmp.spIDs$mean_percN),c('spID')]
  cn.spIDs<-tmp.spIDs[!is.na(tmp.spIDs$mean_cn),c('spID')]
  litterpercN.spIDs<-tmp.spIDs[!is.na(tmp.spIDs$mean_litterpercN),c('spID')]
  littercn.spIDs<-tmp.spIDs[!is.na(tmp.spIDs$mean_littercn),c('spID')]

  listthing<-list(percN=percN.spIDs, cn=cn.spIDs, litterpercN=litterpercN.spIDs, littercn=littercn.spIDs
  result.tab<-ldply(lapply(listthing, length))
  colnames(result.tab)<-c('traitCat','nSp.withCoverTraitData')
  result.tab$nSp.withCoverData<-rep(nSp, dim(result.tab)[1])
  result.tab$nSp.missingCoverData<-rep(nSp.missingCoverData, dim(result.tab)[1])

  return(result.tab)
}

#Invasive species
tmp<-missingCov.tab[missingCov.tab$communityType=='InvSp','nSp.missingCover']
result.invSp<-IdentifyspID.traits(sptrait.df=spIDtraits,
                            spID.vec=invSp.cover,
                            nSp.missingCoverData=tmp)
#Invaded area
tmp<-missingCov.tab[missingCov.tab$communityType=='Inv','nSp.missingCover']
result.inv<-IdentifyspID.traits(sptrait.df=spIDtraits,
                            spID.vec=inv.cover,
                            nSp.missingCoverData=tmp)
#Reference area
tmp<-missingCov.tab[missingCov.tab$communityType=='Ref','nSp.missingCover']
result.ref<-IdentifyspID.traits(sptrait.df=spIDtraits,
                            spID.vec=ref.cover,
                            nSp.missingCoverData=tmp)
#summary
summ.list<-list(invSp=result.invSp, inv=result.inv, ref=result.ref)
summ.tab<-ldply(summ.list)
summ.tab
```

```
##      .id    traitCat nSp.withCoverTraitData nSp.withCoverData
## 1  invSp       percN                    376              405
## 2  invSp          cn                    204              405
## 3  invSp litterpercN                     46              405
## 4  invSp    littercn                     46              405
## 5    inv       percN                    537              596
## 6    inv          cn                    285              596
## 7    inv litterpercN                     85              596
## 8    inv    littercn                     78              596
## 9    ref       percN                    476              560
```

```
## 10   ref          cn                   249              560
## 11   ref litterpercN                    95              560
## 12   ref    littercn                    63              560
##    nSp.missingCoverData
## 1                    41
## 2                    41
## 3                    41
## 4                    41
## 5                   100
## 6                   100
## 7                   100
## 8                   100
## 9                    92
## 10                   92
## 11                   92
## 12                   92
```

```
newfilename<-'traitcoverageSummary.txt'
write.table(summ.tab, file=paste(figuresPath,newfilename, sep='/'), sep='\t')
```