

```
setwd("~/Documents/Working Directory/PNAS")
rm(list=ls())
```

```
#####
```

```
# Libraries #
```

```
#####
```

```
library("tidyverse")
```

```
library("rjags")
```

```
library("coda")
```

```
library("matrixStats")
```

```
library("ggplot2")
```

```
#####
```

```
# #
```

```
# 3. PREPARE AZ REARING DATA #
```

```
# #
```

```
#####
```

```
#####
```

```
# Read in Data #
```

```
#####
```

```
#load rearing data
```

```

df<- read.csv("AZ.csv", header=TRUE, stringsAsFactors = FALSE) ##rearing data

#####

# Fix dates

#####

df$date <- as.Date(df$date, format = "%m/%d/%y")

#Add year, month , day columns to allow for various time analyses

df$year <- as.numeric(format(df$date, "%Y")) # year
df$month <- as.numeric(format(df$date, "%m")) # month
df$day <- as.numeric(format(df$date, "%d")) # day

#Remove data for which there is no date
df<-df[complete.cases(df$date),]
df<- df[order(df$date),]

#Remove coleopteran and non-parasitoid hymenopteran species from database
df <- df[ !(df$insectorder %in% c("Coleoptera","Hymenoptera")), ]
df<-droplevels(df)

#####

# Plot ("plot" or "gencol")

#####

#Designate factor called "collect_method" which has two levels "plot" or "gencol"
#Plot means data was collected from standardized plot
#gencol means data was collected via haphazard collection
df$collect_method<- ifelse(grepl("Plot", df$loc, ignore.case = F), "plot", "gencol")
df$collect_method[grepl("[[:digit:]]",df$Plot)]<-"plot"
df$collect_method[grepl("near plot",df$Plot)]<-"gencol"

#Add category that specifies lep genera

```

```
df$lepgen<-gsub( ".*$", "", df$lepsz )
```

```
#subset and read out df
```

```
df<- df[,c("ID", "loc", "date", "day", "month", "year", "lepsz", "lepfam", "collect_method", "lepgen")]
```

```
write.csv(df, "AZ_workingfile.csv")
```

```
#####
```

```
## Add super family column ##
```

```
#####
```

```
file<-read.csv("AZ_workingfile.csv", header=TRUE)
```

```
file<-file[- grep("Unknown", file$lepfam),]
```

```
file<-file[- grep("unknown", file$lepfam),]
```

```
file<-file[- grep("Micro", file$lepfam),]
```

```
file<-file[- grep("Tenthridinidae", file$lepfam),]
```

```
file<-file[- grep("Diprionidae", file$lepfam),]
```

```
file<-file[- grep("Lamantriidae", file$lepfam),]
```

```
file<-file[- grep("Gelechioid", file$lepfam),]
```

```
file<-droplevels(file)
```

```
file$superfam[file$lepfam=="Bucculatricidae"]<-"Gracillarioidea"
```

```
file$superfam[file$lepfam=="Erebidae" ]<-"Noctuoidea"
```

```
file$superfam[file$lepfam=="Geometridae"]<-"Geometroidea"
```

```
file$superfam[file$lepfam=="Limacodidae"]<-"Zygaenoidea"
```

```
file$superfam[file$lepfam=="Noctuidae"]<-"Noctuoidea"
```

```
file$superfam[file$lepfam=="Oecophoridae"]<-"Gelechioidea"
```

```
file$superfam[file$lepfam=="Pterophoridae"]<-"Pterophoroidea"
```

```
file$superfam[file$lepfam=="Sphingidae"]<-"Bombycoidea"
```

```
file$superfam[file$lepfam=="Acrolophidae"]<-"Tineoidea"
```

```

file$superfam[file$lepfam=="Choreutidae"]<-"Choreutoidea"
file$superfam[file$lepfam=="Euteliidae" ]<-"Noctuoidea"
file$superfam[file$lepfam=="Hesperiidae"]<-"Papilionoidea"
file$superfam[file$lepfam=="Lycaenidae"]<-"Papilionoidea"
file$superfam[file$lepfam=="Nolidae"]<-"Noctuoidea"
file$superfam[file$lepfam=="Papilionidae"]<-"Papilionoidea"
file$superfam[file$lepfam=="Pyralidae"]<-"Pyraloidea"
file$superfam[file$lepfam=="Tortricidae"]<-"Tortricoidea"
file$superfam[file$lepfam=="Apotelodidae"]<-"Bombycoidea"
file$superfam[file$lepfam=="Crambidae"]<-"Pyraloidea"
file$superfam[file$lepfam=="Gelechiidae"]<-"Gelechioidea"
file$superfam[file$lepfam=="Lymantriidae" ]<-"Noctuoidea"
file$superfam[file$lepfam=="Notodontidae" ]<-"Noctuoidea"
file$superfam[file$lepfam=="Pieridae" ]<-"Papilionoidea"
file$superfam[file$lepfam=="Riodinidae"]<-"Papilionoidea"
file$superfam[file$lepfam=="Zygaenidae"]<-"Zygaenoidea"
file$superfam[file$lepfam=="Arctiidae"]<-"Noctuoidea"
file$superfam[file$lepfam=="Depressariidae"]<-"Gelechioidea"
file$superfam[file$lepfam=="Lasiocampidae" ]<-"Lasiocampoidea"
file$superfam[file$lepfam=="Megalopygidae"]<-"Zygaenoidea"
file$superfam[file$lepfam=="Nymphalidae" ]<-"Papilionoidea"
file$superfam[file$lepfam=="Psychidae"]<-"Tineoidea"
file$superfam[file$lepfam=="Saturniidae"]<-"Bombycoidea"
file$superfam<-as.factor(file$superfam)

```

#----Get an idea of how many samples per super family so that we can create minimum threshold to be included in study

```
temp <-
```

```
file%>%  
filter(!is.na(superfam))%>%  
group_by(year, superfam) %>%  
dplyr::summarize(freq = n())
```

```
ggplot(temp, aes(x=year, y=freq, colour=superfam)) +  
  geom_point() +  
  geom_smooth(method=lm, se=FALSE)+  
  facet_wrap(~superfam)
```

#Remove (removed below) the following family bc these superfamilies were samped only 1-3 years:
#Choreutoidea, Gracillarioidea, Tineoidea, Pterophoroidea, Tineaoidea

```
file<- file[!(file$superfam %in% c("Choreutoidea", "Gracillarioidea", "Tineoidea", "Pterophoroidea",  
"Tineaoidea")),]  
file<-droplevels(file)
```

```
year<-file  
month_df<-file
```

```
#####  
# Summarise frequency table  
#####
```

#----ANNUAL SUMMARY

```
a_year <-  
year%>%  
filter(!is.na(superfam))%>%
```

```
group_by(year, superfam) %>%  
dplyr::summarize(freq = n())
```

```
p_year <-  
year%>%  
filter(collect_method=="plot", !is.na(superfam))%>%  
group_by(year, superfam) %>%  
dplyr::summarize(freq = n())
```

```
g_year <-  
year%>%  
filter(collect_method=="gencol", !is.na(superfam))%>%  
group_by(year, superfam) %>%  
dplyr::summarize(freq = n())
```

#----MONTHLY SUMMARY

#----Count of super fam freq per month and collection method

```
month_df <-  
month_df%>%  
filter(!is.na(superfam))%>%  
group_by(year, month, collect_method) %>%  
dplyr::summarize(freq = n())
```

```
all_methods <-  
month_df%>%  
group_by(year, month) %>%  
dplyr::summarize(freq = sum(freq))
```

```
all_methods$collect_method<- "all"
month_df<-rbind(as.data.frame(month_df), as.data.frame(all_methods))
```

```
#####
#                                     #
#           4. BAYESIAN ANALYSES           #
#                                     #
#####
```

```
dat<-a_year
temp<-dat
dat$freq <- as.numeric(scale(log(dat$freq)))
dat$year <- as.numeric(scale(dat$year))
dat<-dat[ which( ! dat$year %in% NA) , ]
```

```
library("rjags")
library("coda")
library("matrixStats")
## put into list format jags wants
dataForJags <- list(freq=dat$freq, year=dat$year, superfam=dat$superfam,
Ncol=length(levels(dat$superfam)), N=length(dat$freq))
```

```
## specify model, priors etc.
model <- "model{
for(i in 1:N){
freq[i] ~ dnorm(alpha[i], tau)
alpha[i] <- mu[superfam[i]] + beta1[superfam[i]]*year[i]
```

```

}
# coefficients for individual collection_method (conditional priors)
for(j in 1:Ncol){
beta1[j] ~ dnorm(beta1mu, beta1tau)
mu[j] ~ dnorm(mumu, mutau)
}
tau ~ dgamma(0.1,0.1)

# hyperpriors for coefficients
beta1mu ~ dnorm(0,0.001)
mumu ~ dnorm(0,0.001)

# hyperpriors for precisions
beta1tau ~ dgamma(0.1,0.1)
mutau ~ dgamma(0.1,0.1)
}"

## compile
mod1 <- jags.model(textConnection(model),data= dataForJags,n.chains=2)

## run it
samples<-jags.samples(model= mod1,variable.names=c("beta1","mu","beta1mu","mumu",
"beta1tau","mutau","tau"),n.iter=1000000)

#take beta1mu (coefficient and 95%CI for highest level of heirarchy) to later ass to caterpillar plot
keyBits <-round(quantile(samples$beta1mu[,,],probs=c(0.5,0.1,0.9)),3)

#Calculate probability of Negative slope
probNeg<-as.vector(samples$beta1mu[,,1])
probNeg <- length(probNeg[probNeg<0]) / length(as.vector(samples$beta1mu[,,1]))

```



```
vec <- data.frame(site= "AZ", superfam= "Overall Beta Coefficient", mid=keyBits[1], lower=keyBits[2],  
upper=keyBits[3], probNeg=probNeg)
```

```
vec
```

```
#### generate a table of genus-specific slopes and intervals
```

```
mid <- NA
```

```
lower <- NA
```

```
upper <- NA
```

```
probNeg <- NA
```

```
res <- samples$beta1
```

```
for(i in 1:length(levels(dat$superfam))){
```

```
  sub <- res[i,,]
```

```
  mid[i] <- quantile(sub,0.5)
```

```
  upper[i] <- quantile(sub,0.9)
```

```
  lower[i] <- quantile(sub,0.1)
```

```
  probNeg[i] <- length(sub[sub<0]) / length(sub)
```

```
  print(i)
```

```
}
```

```
new <- data.frame(levels(dat$superfam),mid,lower,upper, probNeg)
```

```
new <- new[order(new$mid,decreasing=T),]
```

```
colnames(new)[1] <- "superfam"
```

```
new$site<- "AZ"
```

```
new<-rbind(new, vec)
```

```
#hist(new$mid, col="green",xlim=c(-0.2, 0.0001))
```

```
head(new)
```

```
write.csv(new, "new.csv")
```

```
#####
```

```
#
```

```
#
```

```

#           4. GRAPHS           #
#                               #
#####

#####

#       Caterpillar Plot
#
#####

#Use for a_year

new$border_color <- c(rep("#006400", 2),rep("#51B700", 3), rep("#3B94ED", 2),
rep("#3B16A1",2),"#1B1E23")

new$fill_color <- c(rep("#006400", 2),rep("#51B700", 3), rep("#3B94ED", 2), rep("#3B16A1",2),
"#1B1E23")

# Use for p_year

new$border_color <- c( rep("#51B700", 3), rep("#3B94ED", 6), "#1B1E23")
new$fill_color <- c( rep("#51B700", 3), rep("#3B94ED", 6), "#1B1E23")

# Use for g_year

new$border_color <- c( rep("#51B700", 4), rep("#3B94ED", 4), "#3B16A1", "#1B1E23")
new$fill_color <- c( rep("#51B700", 4), rep("#3B94ED", 4), "#3B16A1", "#1B1E23")

#use for g_year

quartz(width=7,height=9)

## plot those point estimates and credible intervals in caterpillar plot
layout(matrix(c(1,1,1,2,1,1,1,3,1,1,1,4,1,1,1,5,1,1,1,6),5,4,byrow=T))
par(mar=c(5,1,0,5))

```

```

par(oma=c(12,0,1,4))

plot(new$mid,10:1,yaxt="n",ylab="",xlab="",xlim=c(-0.9,0.9),ylim=c(1,10),cex.axis=2, cex=3, font=2,
cex.lab=5, pch=21,col=new$border_color, bg= alpha(new$fill_color, 0.5), lwd=2)

mtext("Change in Annual Frequency",1,line=2.5,cex=1.5, font= 2)

arrows(new$lower, 10:1, new$upper, 10:1,length=0,lwd=2, col=new$fill_color)

abline(v=0,lty=2)

axis(4,labels= new$superfam,at=10:1,las=2,cex.axis=2,font=2,tck=-0.02)

grid(nx=10,ny=20)

```

```
#####
```

```
# Box Plot: Monthly Freq Across Years
```

```
#
```

```
#####
```

```
# Box Plot: Monthly Freq Across Years
```

```
# purple: plot
```

```
# blue: gencol
```

```
# forest green= all
```

```
az <-
```

```
month_df%>%
```

```
group_by(year, collect_method) %>%
```

```
dplyr::summarize(mean=mean(freq), sd=sd(freq), n=n_distinct(month))
```

```
az <-
```

```
az%>%
```

```
group_by(year, collect_method ) %>%
```

```
dplyr::mutate( se=(sd/sqrt(n)))
```

```
az$mean<-scale(az$mean)
```

```
az$se<-scale(az$se)
```

```
az$se[is.na(az$se)] <- 0
```

```
az$year<-as.factor(az$year)
```

```
lvls<-levels(az$year)
```

```
ggplot(az, aes(x=as.factor(year), y=mean, color= collect_method))+  
  geom_pointrange(aes(ymin = mean-se, ymax = mean+se, color=  
collect_method),position=position_jitterdodge(-2),alpha=0.7, size=0.75)+  
  scale_x_discrete(limits=lvls,breaks=lvls[seq(1,length(lvls),by=2)])+  
  scale_color_manual(values = c("#CDCDCD", "#878787", "#000000"))+  
  labs( x="", y=expression(paste("Frequency of Encounter")))+  
  theme_classic()+  
  theme(axis.text.x = element_text(face= "bold", size= 20, angle=18, hjust = 1),  
        axis.text.y =element_text(face= "bold", size=20),  
        axis.title.x= element_text( size = rel(2.5)),  
        axis.title.y= element_text( size = rel(2.5)))
```