

Ohio: BBS bird abundance

Olivia Burge

5 October 2014

Task one: the traits

I've saved a csv of common birds in steep decline from Kelly's xlsx. The only change I made in excel was deleting the apostrophes as they looked *nasty* when read in. Otherwise, nada. I haven't set a working directory for this and used full paths because I like that. But if you guys give me a stern talking to I will. Slash I will once I put all this stuffs in one place.

```
traits_common_steep <- read.csv("raw_data/commonbirds_steepdecline_traits.csv")
str(traits_common_steep)

#subsetting just the birds that are in Ohio and eat insect
traits_common_steep_insectohio <- as.data.frame(traits_common_steep %>%
  filter(ohio_presence > 0 & diet == "insects"))

str(traits_common_steep_insectohio)

#because we use it later
traits_common_steep_insectohio$common_name <- factor(toupper(traits_common_steep_insectohio$common_name))

#which birds are we looking at? :
levels(traits_common_steep_insectohio$common_name)
```

Task two - AOU codes for bird species

Next I also had to import the AOU codes because those are used in BBS, so we have a common 'key'.

```
AOU_codes <- read.csv("raw_data/AOU_codes.csv")

# merge the AOU codes into the ohio data
traits_common_steep_insectohio2 <- as.data.frame(inner_join(AOU_codes, traits_common_steep_insectohio))

## Joining by: "common_name"

head(traits_common_steep_insectohio2)
```

Task three: BBS data

I got the BBS data from: ftp://ftpext.usgs.gov/pub/er/md/laurel/BBS/DataFiles//50-StopData/1997ToPresent_SurveyWide/Fifty7.zip. The metadata is available: <ftp://ftpext.usgs.gov/pub/er/md/laurel/BBS/DataFiles//FiftySt.txt>.

```

ohio_BBS <- read.csv("raw_data/fifty7.csv")
head(ohio_BBS)

# merge with the traits data
ohio_BBS_insect <- as.data.frame(inner_join(ohio_BBS, traits_common_steep_insectohio2))

## Joining by: "AOU"

ohio_BBS_insect$common_name<-factor(ohio_BBS_insect$common_name)
head(ohio_BBS_insect)

#now we need to locate the routes in each county. I read in the file Tyson had already located for us:
ohio_BBS_routes_countiesraw <- read.csv("raw_data/BBSdata.by.county.csv")
head(ohio_BBS_routes_countiesraw)
ohio_BBS_routes_countiesraw2 <- ohio_BBS_routes_countiesraw[!duplicated(ohio_BBS_routes_countiesraw$Route), ]

#and then joined the two
ohio_BBS_birds <- inner_join(ohio_BBS_insect, ohio_BBS_routes_countiesraw2)

## Joining by: "Route"

```

```

head(ohio_BBS_birds)
names(ohio_BBS_birds)

# created a column of sums for all the stops on each route:
ohio_BBS_birds$route_abundance <- rowSums(ohio_BBS_birds[grepl("^Stop[0-9]+", names(ohio_BBS_birds))])
#using grep means even if the columns change position, we will still get the right ones for rowsums.

```

Next up: a summary. Some counties have more than one route. Thought on this?

Getting a version of the neonics

```

#this is from Tyson's work.
filenames <- list.files(path = "raw_data/raw_pesticides/", recursive = TRUE, full.names = TRUE)
#I've amended the line above so that we can get the file paths as well
#(the recursive and full.names options).

pesticides_raw <- do.call("rbind", lapply(filenames[grepl("high", filenames)],
                                         read.table, header = TRUE, sep = "\t"))

pesticides_raw_ohio <- pesticides_raw[which(pesticides_raw$STATE_FIPS_CODE == 39), ]

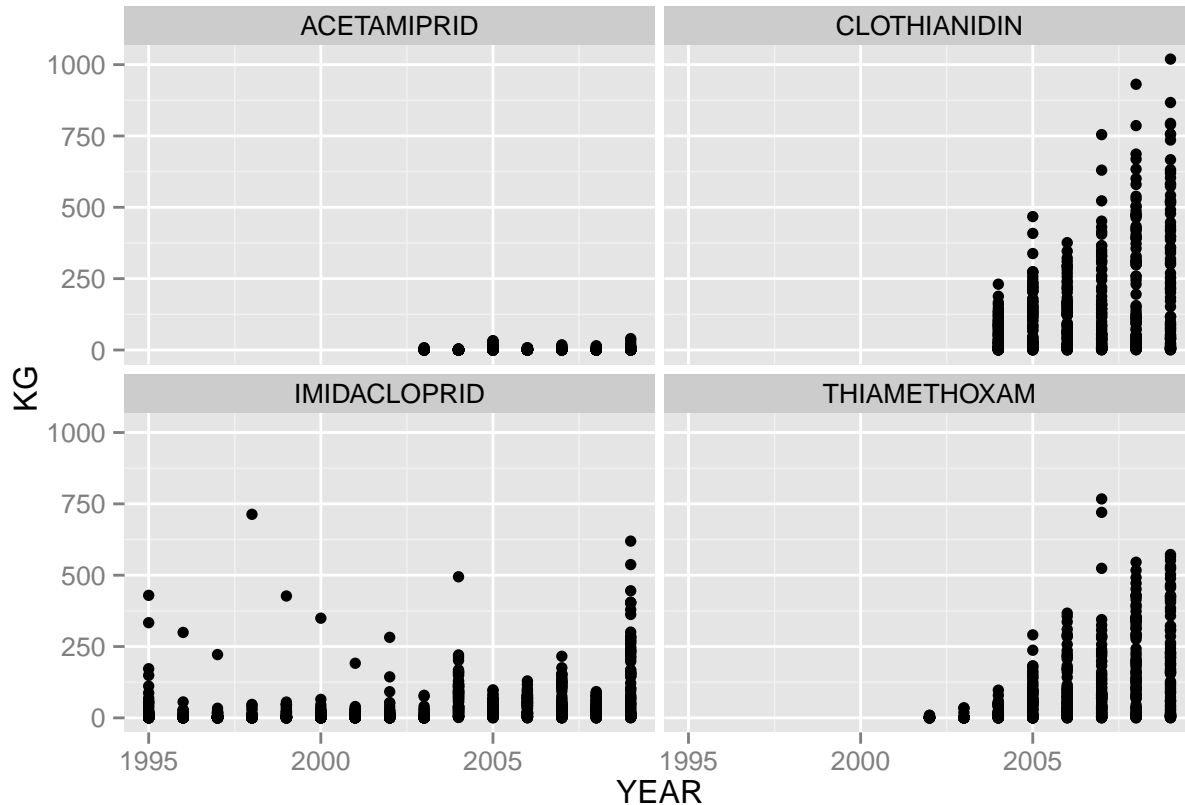
rm(pesticides_raw) #cos it's massive

names(pesticides_raw_ohio)

#getting the neonics out
neonics_raw_ohio <- pesticides_raw_ohio %>%
  filter(COMPOUND %in% c("ACETAMIPRID", "CLOTHIANIDIN", "IMIDACLOPRID", "THIAMETHOXAM"))

```

```
ggplot(neonics_raw_ohio, aes(x = YEAR, y = KG))+
  geom_point()+
  facet_wrap(~COMPOUND)
```



#similar to Tyson's. Phew.

Controlling for county area

Metadata: GCT-PH1 population, Housing Units, Area, and Density: 2010 - State – County / County Equivalent

NOTE: For information on confidentiality protection, nonsampling error, and definitions, see <http://www.census.gov/prod/cen2010/doc/sf1.pdf>.

Available from: <http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk>

```
ohio_facts<-read.csv("raw_data/DEC_10_SF1_GCTPH1/DEC_10_SF1_GCTPH1.ST05_with_ann.csv")

head(ohio_facts)
names(ohio_facts)
ohio_facts_r <- ohio_facts[ , c(2, 3, 5, 7, 10:14)]
head(ohio_facts_r)
names(ohio_facts_r) <- c("state_fips", "state", "GEOID", "county",
  "total_area", "total_water", "total_land",
```

```

"popdens_sqmile", "housedens_sqmile")

#FIPS codes and tidying the double header row!

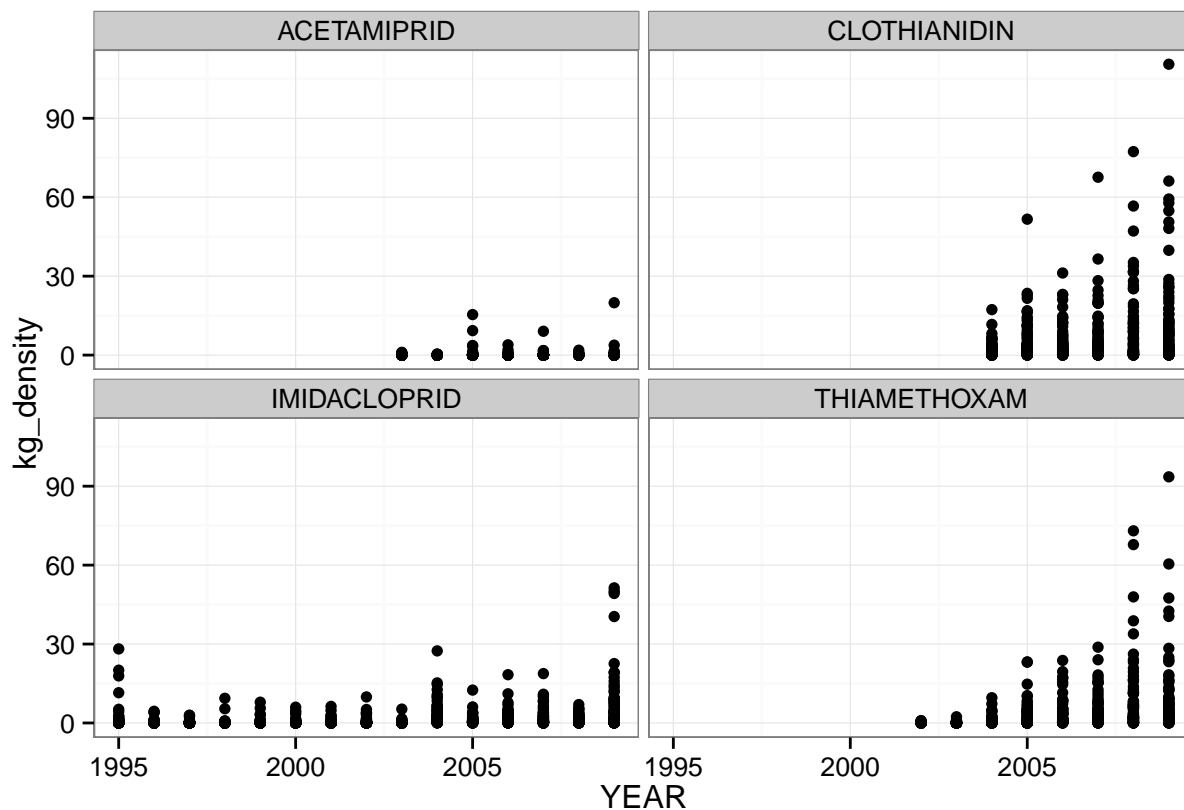
#now to merge
neonics_raw_ohio$county_id<-str_pad(neonics_raw_ohio$COUNTY_FIPS_CODE, 3, "left", pad = "0")
neonics_raw_ohio$GEOID <- paste(neonics_raw_ohio$STATE_FIPS_CODE, neonics_raw_ohio$county_id, sep = "")

ohio_facts_r <- ohio_facts_r[-c(1,2),]
neonics_facts_ohio <- merge(neonics_raw_ohio, ohio_facts_r)

neonics_facts_ohio$total_land<-as.numeric(neonics_facts_ohio$total_land)
neonics_facts_ohio$kg_density <- neonics_facts_ohio$KG/neonics_facts_ohio$total_land

#not quite the same as Tyson's:
ggplot(neonics_facts_ohio, aes(x = YEAR, y = kg_density)) +
  geom_point()+
  facet_wrap(~COMPOUND)+
  theme_bw()

```

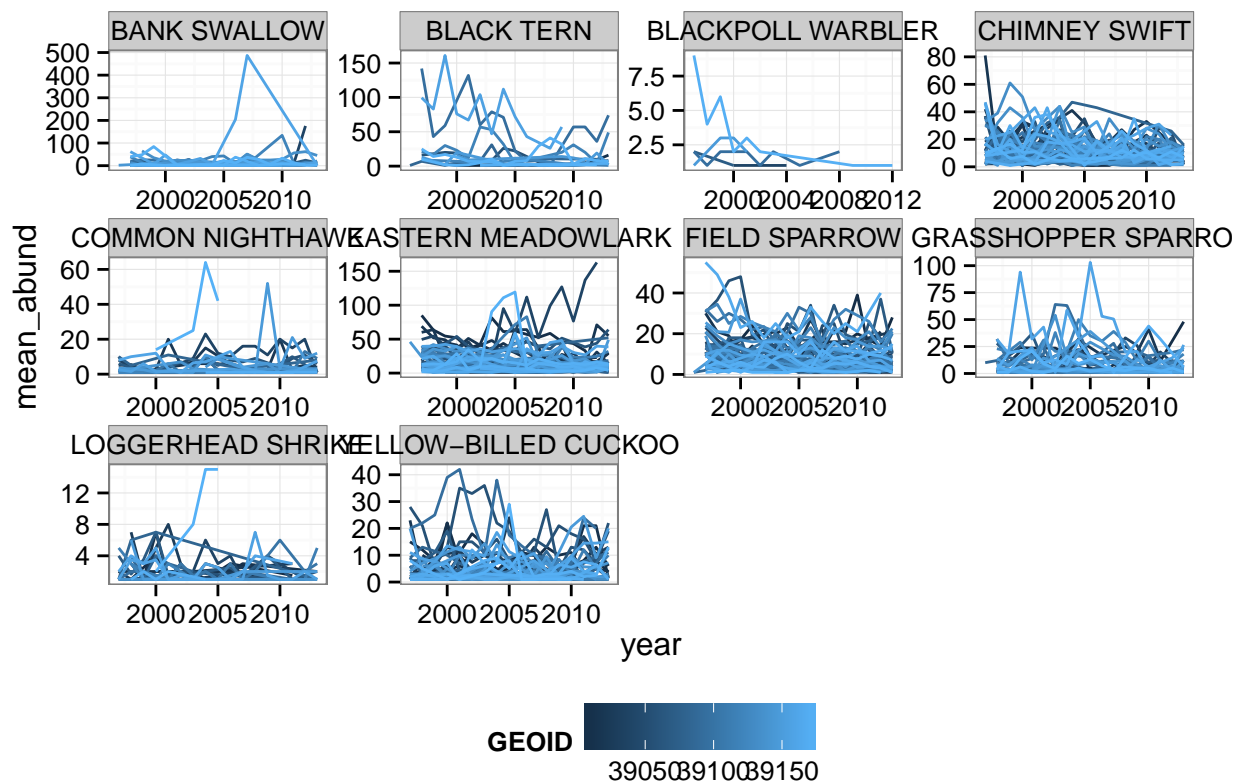


Data reading done (almost)

```
ohio_BBS_birds_summary <- data.frame(ohio_BBS_birds %>%  
  group_by(year, GEOID, common_name) %>%  
  summarise(mean_abund = mean(route_abundance),  
            sd = sd(route_abundance),  
            n = n()) %>%  
  mutate(sem_abund = sd/sqrt(n)))
```

Summarising data

```
#which(ohio_BBS_birds_summary$n>1)  
#names(ohio_BBS_birds_summary)  
  
ggplot(ohio_BBS_birds_summary, aes(x = year, y = mean_abund))+  
  geom_line(aes(colour = GEOID, group = GEOID))+  
  facet_wrap(~common_name, scales = "free") +  
  theme_bw()+  
  theme(legend.position = "bottom")
```



```
names(ohio_BBS_birds_summary)<-toupper(names(ohio_BBS_birds_summary))
neonics_facts_ohio$GEOID<-as.numeric(neonics_facts_ohio$GEOID)

head(neonics_facts_ohio)
```

```
##   GEOID      COMPOUND YEAR STATE_FIPS_CODE COUNTY_FIPS_CODE   KG county_id
## 1 39001  ACETAMIPRID 2003           39           1 0.6         001
## 2 39001  THIAMETHOXAM 2008           39           1 34.7        001
## 3 39001  CLOTHIANIDIN 2006           39           1 8.3         001
## 4 39001  IMIDACLOPRID 1998           39           1 35.5        001
## 5 39001  IMIDACLOPRID 2001           39           1 20.7        001
## 6 39001  IMIDACLOPRID 2003           39           1 19.9        001
##   state_fips state      county total_area total_water total_land
## 1      39 Ohio Adams County    586.25      2.38         80
## 2      39 Ohio Adams County    586.25      2.38         80
## 3      39 Ohio Adams County    586.25      2.38         80
## 4      39 Ohio Adams County    586.25      2.38         80
## 5      39 Ohio Adams County    586.25      2.38         80
## 6      39 Ohio Adams County    586.25      2.38         80
##   popdens_sqmile housedens_sqmile kg_density
## 1      48.9      22.2      0.00750
## 2      48.9      22.2      0.43375
## 3      48.9      22.2      0.10375
## 4      48.9      22.2      0.44375
## 5      48.9      22.2      0.25875
## 6      48.9      22.2      0.24875
```

```
head(ohio_BBS_birds_summary)
```

```
##   YEAR GEOID      COMMON_NAME MEAN_ABUND  SD N SEM_ABUND
## 1 1996 39083      BLACK TERN         1 NaN 1         NaN
## 2 1996 39083      FIELD SPARROW         1 NaN 1         NaN
## 3 1996 39083 GRASSHOPPER SPARROW        10 NaN 1         NaN
## 4 1996 39143      BANK SWALLOW         2 NaN 1         NaN
## 5 1996 39143 EASTERN MEADOWLARK        46 NaN 1         NaN
## 6 1996 39143      FIELD SPARROW         1 NaN 1         NaN
```

```
ohio_insects <- merge(neonics_facts_ohio[neonics_facts_ohio$YEAR >= 1996 & neonics_facts_ohio$YEAR <= 2009],
                      ohio_BBS_birds_summary, by = "GEOID")

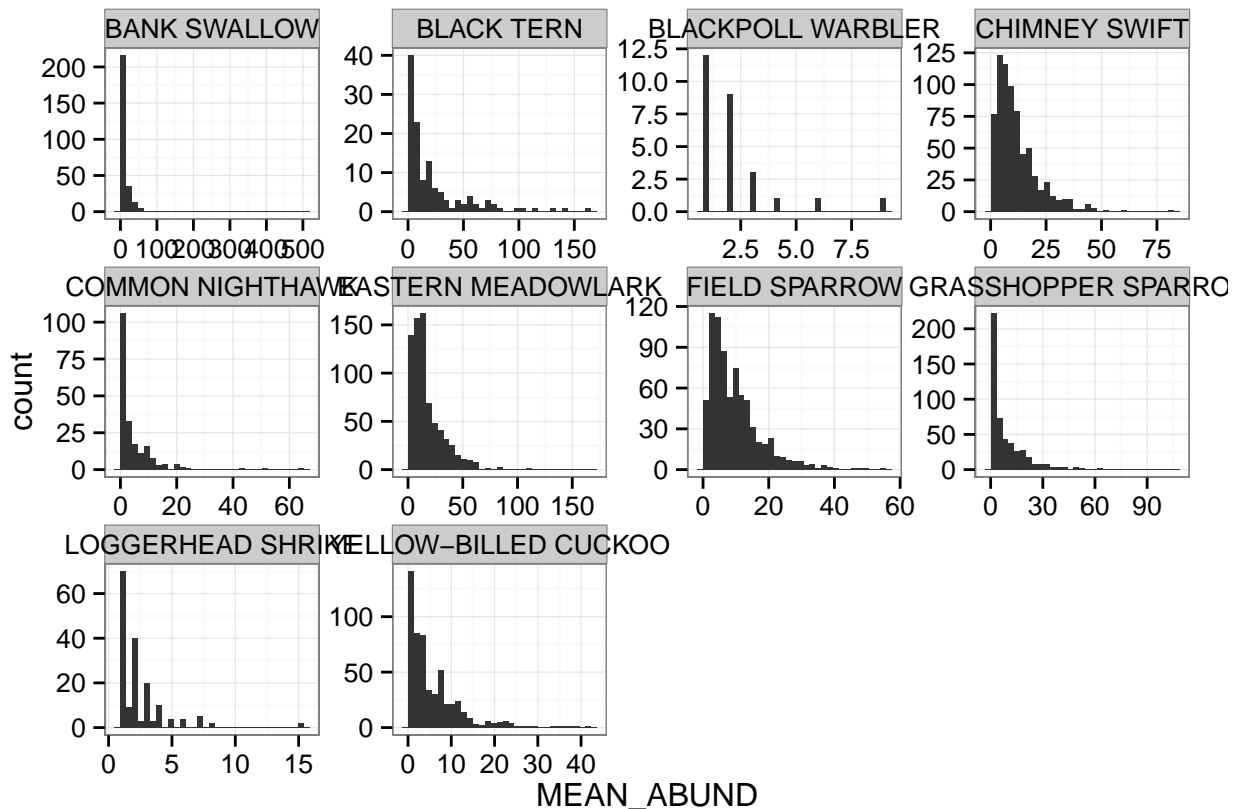
unique(ohio_insects$YEAR)
```

```
## [1] 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
```

```
# without doing it for each county, response looks pretty poisson
ggplot(ohio_BBS_birds_summary, aes(x = MEAN_ABUND))+
  geom_histogram()+
  theme_bw()+
  facet_wrap(~COMMON_NAME, scales = "free")
```

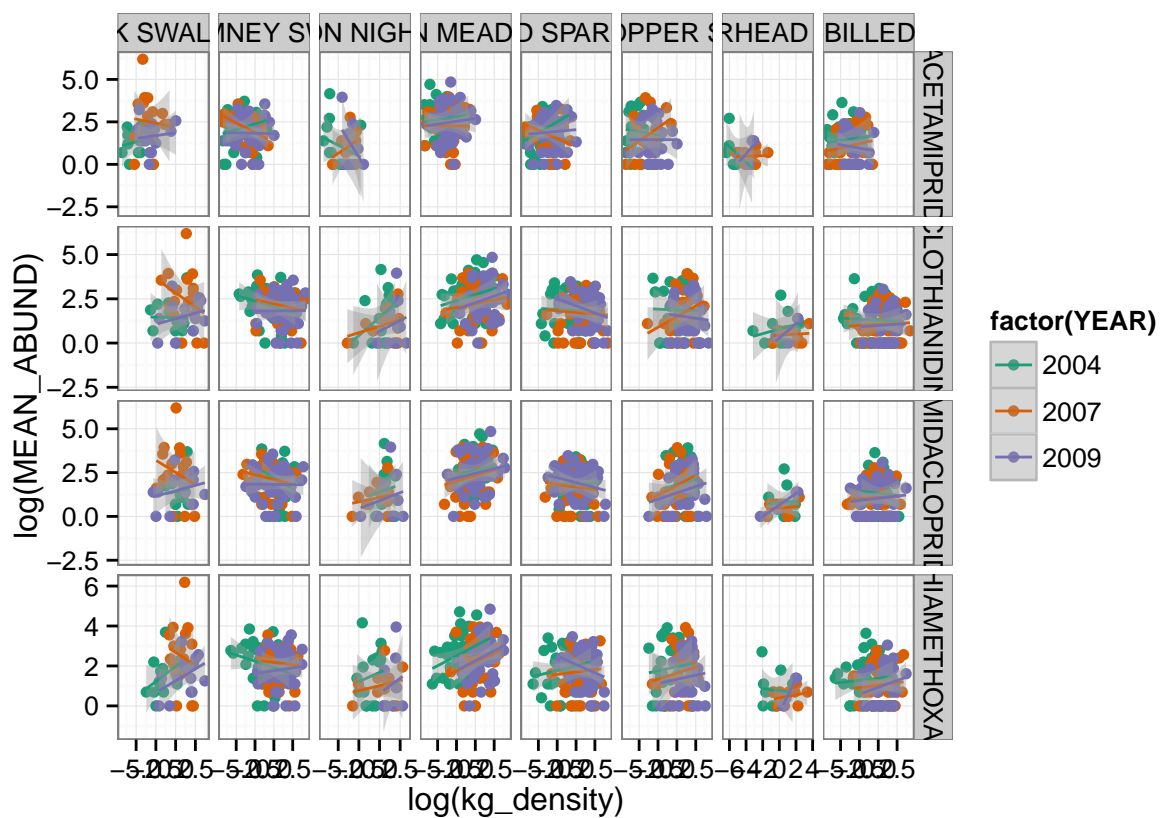
```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```

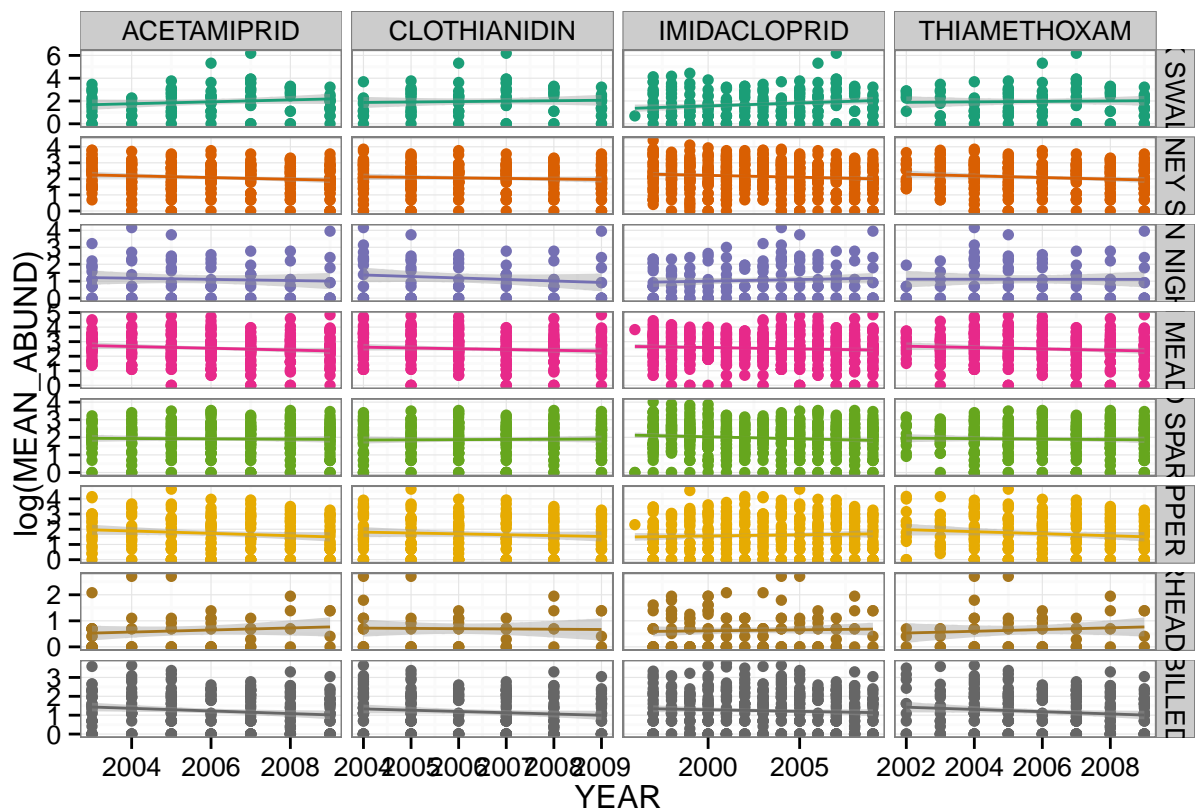


#more graphs. I had a play. Excluded the two species as very low occurrence.

```
ggplot(subset(ohio_insects,ohio_insects$YEAR %in% c(2004,2007, 2009) & !ohio_insects$COMMON_NAME %in% c
  geom_point(aes(colour = factor(YEAR)))+
  scale_colour_brewer(type = "qual", palette = "Dark2")+
  geom_smooth(aes(colour = factor(YEAR)), method = lm)+
  theme_bw()+
  facet_grid(COMPOUND~COMMON_NAME, scales = "free")+
  coord_fixed()
```



```
ggplot(subset(ohio_insects, !ohio_insects$COMMON_NAME %in% c("BLACK TERN", "BLACKPOLL WARBLER")), aes(y
  geom_point(aes(colour = factor(COMMON_NAME)))+
  scale_colour_brewer(type = "qual", palette = "Dark2", guide = FALSE)+
  geom_smooth(aes(colour = factor(COMMON_NAME)), method = lm)+
  theme_bw()+
  facet_grid(COMMON_NAME~COMPOUND, scales = "free")+
  coord_fixed()
```

First attempt at modelling

```
ohio_insects_wide <- spread(ohio_insects, key = COMPOUND, value = kg_density)

ohio_insects_wide$GEOID<-as.factor(ohio_insects_wide$GEOID)
ohio_insects_wide$county<-as.factor(ohio_insects_wide$county)
ohio_insects$MEAN_ABUND <- round(ohio_insects$MEAN_ABUND) #poisson requires integers
ohio_insects_wide$MEAN_ABUND <- round(ohio_insects_wide$MEAN_ABUND)

#fails. I had one working
try(mod1 <- glm(data=ohio_insects_wide, MEAN_ABUND ~ COMMON_NAME:(CLOTHIANIDIN + ACETAMIPRID + IMIDACLOPRID + THIAMETHOXAM),
  family = "poisson"))
try(summary(mod1))

### i think i need to go back and put in wide levels of compound and put zero where not recorded#####
try(mod1_pred <- predict(mod1, type = "response"))

#works. But this is not the wide version and only one species, otherwise the interactions get messy.
mod2<- glm(data= ohio_insects[ohio_insects$COMMON_NAME == "CHIMNEY SWIFT",], MEAN_ABUND ~ kg_density:COMMON_NAME,
  family = "poisson")
summary(mod2)

##
## Call:
```

```
## glm(formula = MEAN_ABUND ~ kg_density:COMPOUND, family = "poisson",
##      data = ohio_insects[ohio_insects$COMMON_NAME == "CHIMNEY SWIFT",
##      ])
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.0696  -2.2201  -0.5614   1.1829  13.2037
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      2.463391   0.008844 278.533 < 2e-16 ***
## kg_density:COMPOUNDACETAMIPRID -0.362782   0.136210  -2.663 0.007735 **
## kg_density:COMPOUNDCLOTHIANIDIN -0.008021   0.002103  -3.814 0.000137 ***
## kg_density:COMPOUNDIMIDACLOPRID -0.008787   0.004114  -2.136 0.032708 *
## kg_density:COMPOUNDTHIAMETHOXAM -0.011073   0.003047  -3.634 0.000279 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 8913.4  on 1336  degrees of freedom
## Residual deviance: 8875.0  on 1332  degrees of freedom
## AIC: 14195
##
## Number of Fisher Scoring iterations: 5
```

```
save.image(file = "BBS_pest.Rdata")
```

```
#setwd("/Users/Liv/Documents/NCEAS_GIT/NCEAS-RENCI_2014/BBS_data")
landcover_all <- read.csv("../Landcover/landcover_all.csv")
head(landcover_all)
```

```
##      GEOID reclass  total year
## 1 18001         1    2482 1992
## 2 18001         2   78987 1992
## 3 18001         3     207 1992
## 4 18001         4   45232 1992
## 5 18001         5   14413 1992
## 6 18001         6  831727 1992
```

```
head(ohio_insects)
```

```
##      YEAR GEOID      COMPOUND STATE_FIPS_CODE COUNTY_FIPS_CODE  KG county_id
## 1 1996 39083 IMIDACLOPRID           39           83 5.5      083
## 2 1996 39083 IMIDACLOPRID           39           83 5.5      083
## 3 1996 39083 IMIDACLOPRID           39           83 5.5      083
## 4 1996 39143 IMIDACLOPRID           39          143 28.3     143
## 5 1996 39143 IMIDACLOPRID           39          143 28.3     143
## 6 1996 39143 IMIDACLOPRID           39          143 28.3     143
##      state_fips state      county total_area total_water total_land
## 1      39 Ohio      Knox County    529.63      4.14      69
## 2      39 Ohio      Knox County    529.63      4.14      69
```

```
## 3      39 Ohio      Knox County      529.63      4.14      69
## 4      39 Ohio Sandusky County      417.71      9.25      24
## 5      39 Ohio Sandusky County      417.71      9.25      24
## 6      39 Ohio Sandusky County      417.71      9.25      24
##      popdens_sqmile housedens_sqmile kg_density      COMMON_NAME
## 1      115.9      47.8 0.07971014 GRASSHOPPER SPARROW
## 2      115.9      47.8 0.07971014      FIELD SPARROW
## 3      115.9      47.8 0.07971014      BLACK TERN
## 4      149.2      64.6 1.17916667      BANK SWALLOW
## 5      149.2      64.6 1.17916667      EASTERN MEADOWLARK
## 6      149.2      64.6 1.17916667      FIELD SPARROW
##      MEAN_ABUND      SD      N      SEM_ABUND
## 1      10      NaN      1      NaN
## 2      1      NaN      1      NaN
## 3      1      NaN      1      NaN
## 4      2      NaN      1      NaN
## 5      46      NaN      1      NaN
## 6      1      NaN      1      NaN
```

```
unique(ohio_insects$YEAR)
```

```
## [1] 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
```

```
unique(landcover_all$year)
```

```
## [1] 1992 2001 2006 2011
```

```
sort(unique(neonics_facts_ohio$YEAR))
```

```
## [1] 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008
## [15] 2009
```

```
sort(unique(neonics_facts_ohio$YEAR))
```

```
## [1] 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008
## [15] 2009
```

```
sort(unique(ohio_BBS_birds_summary$YEAR))
```

```
## [1] 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
## [15] 2010 2011 2012 2013
```

```
sort(unique(ohio_BBS_birds$year))
```

```
## [1] 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
## [15] 2010 2011 2012 2013
```

```
sort(unique(neonics_raw_ohio$YEAR))
```

```
## [1] 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008  
## [15] 2009
```

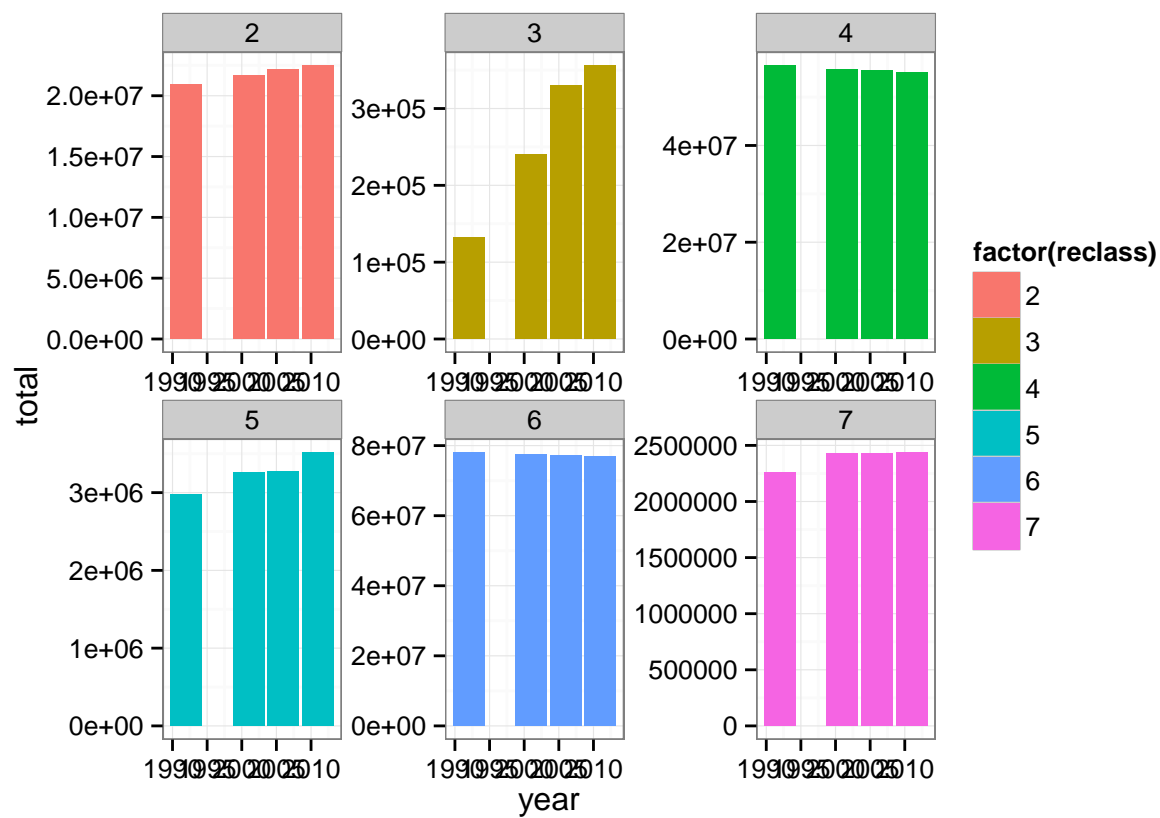
```
class(landcover_all$year)
```

```
## [1] "integer"
```

```
head(landcover_all)
```

```
##   GEOID reclass  total year  
## 1 18001      1   2482 1992  
## 2 18001      2  78987 1992  
## 3 18001      3    207 1992  
## 4 18001      4  45232 1992  
## 5 18001      5  14413 1992  
## 6 18001      6 831727 1992
```

```
landcover_sum1 <- data.frame(landcover_all %>%  
                             group_by(year, reclass) %>%  
                             summarise(  
                               landcover = sum(total)))  
landcover_sum2 <- aggregate(total ~ year + reclass, landcover_all, FUN = sum)  
  
ggplot(landcover_sum2[landcover_sum1$year>1992, ], aes(x = year, y = total)) +  
  geom_bar(aes(fill = factor(reclass)), stat = "identity")+  
  theme_bw() +  
  facet_wrap(~ reclass, scales = "free")
```



landcover-1.pdf