

IUCN_dat_intro

Sophia Richter

6/8/2021

Read Data

```
data = read_csv('iucn_dat.csv')

# make data set smaller to just columns interested in
imp_data = data %>% select(scientificName, redlistCategory, populationTrend, BodyMass)

imp_data2 = data %>% select(redlistCategory, populationTrend, BodyMass, cat_change, cat_change_bi, syst
```

Data Set Basics

- there are 35451 total species in the data set
- there are 13400 species with body mass data

How many species are in each distinct category?

```
imp_data %>% group_by(redlistCategory) %>% summarize(Sum = n()) %>% arrange(-Sum)
```

```
## # A tibble: 10 x 2
##   redlistCategory      Sum
##   <chr>             <int>
## 1 Least Concern      20473
## 2 Data Deficient     5152
## 3 Vulnerable         2928
## 4 Endangered         2740
## 5 Near Threatened    2333
## 6 Critically Endangered 1521
## 7 Extinct            285
## 8 Extinct in the Wild   15
## 9 Lower Risk/least concern 3
## 10 Lower Risk/conservation dependent 1
```

```
imp_data %>% group_by(populationTrend) %>% summarize(Sum = n()) %>% arrange(-Sum)
```

```
## # A tibble: 5 x 2
##   populationTrend      Sum
##   <chr>             <int>
## 1 Unknown          12754
## 2 Decreasing       11508
```

```
## 3 Stable          9848
## 4 Increasing      900
## 5 <NA>            441
```

```
imp_data2 %>% group_by(cat_change) %>% summarize(Sum = n()) %>% arrange(-Sum)
```

```
## # A tibble: 10 x 2
##   cat_change Sum
##   <dbl> <int>
## 1      NA 32749
## 2       1  1011
## 3      -1   881
## 4       2   352
## 5      -2   311
## 6       3    69
## 7      -3    35
## 8       0    27
## 9       4    10
## 10      -4     6
```

```
imp_data2 %>% group_by(cat_change_bi) %>% summarize(Sum = n()) %>% arrange(-Sum)
```

```
## # A tibble: 3 x 2
##   cat_change_bi Sum
##   <chr>      <int>
## 1 <NA>      32776
## 2 decline   1442
## 3 improve   1233
```

```
imp_data2 %>% group_by(systems) %>% summarize(Sum = n()) %>% arrange(-Sum)
```

```
## # A tibble: 10 x 2
##   systems Sum
##   <chr>    <int>
## 1 Terrestrial 19595
## 2 Terrestrial|Freshwater (=Inland waters) 6712
## 3 Marine 5202
## 4 Freshwater (=Inland waters) 2851
## 5 Terrestrial|Freshwater (=Inland waters)|Marine 494
## 6 Terrestrial|Marine 408
## 7 Freshwater (=Inland waters)|Marine 171
## 8 <NA> 10
## 9 Marine|Marine 7
## 10 Freshwater (=Inland waters)|Marine|Marine 1
```

```
imp_data2 %>% group_by(realm) %>% summarize(Sum = n()) %>% arrange(-Sum)
```

```
## # A tibble: 128 x 2
##   realm Sum
##   <chr> <int>
## 1 Neotropical 10608
```

```
## 2 Afrotropical      5200
## 3 Indomalayan      4719
## 4 <NA>              2789
## 5 Australasian     2288
## 6 Oceanian         1842
## 7 Nearctic         1684
## 8 Palearctic       1680
## 9 Indomalayan|Palearctic 1446
## 10 Nearctic|Neotropical 1003
## # ... with 118 more rows
```

```
imp_data2 %>% group_by(kingdomName) %>% summarize(Sum = n()) %>% arrange(-Sum)
```

```
## # A tibble: 2 x 2
##   kingdomName Sum
##   <chr>      <int>
## 1 animalia   32286
## 2 plantae    3165
```

```
imp_data2 %>% group_by(phylumName) %>% summarize(Sum = n()) %>% arrange(-Sum)
```

```
## # A tibble: 4 x 2
##   phylumName Sum
##   <chr>      <int>
## 1 chordata    27968
## 2 tracheophyta 3165
## 3 arthropoda  2935
## 4 mollusca    1383
```

```
imp_data2 %>% group_by(orderName) %>% summarize(Sum = n()) %>% arrange(-Sum)
```

```
## # A tibble: 106 x 2
##   orderName Sum
##   <chr>      <int>
## 1 passeriformes 6659
## 2 anura         6098
## 3 decapoda      2864
## 4 rodentia      2364
## 5 perciformes  1684
## 6 caryophyllales 1477
## 7 chiroptera    1304
## 8 neogastropoda  633
## 9 caudata       611
## 10 pinales      610
## # ... with 96 more rows
```

```
imp_data2 %>% group_by(className) %>% summarize(Sum = n()) %>% arrange(-Sum)
```

```
## # A tibble: 13 x 2
##   className Sum
##   <chr>      <int>
```

```
## 1 aves      11147
## 2 amphibia  6892
## 3 mammalia  5899
## 4 malacostraca 2935
## 5 actinopterygii 2595
## 6 magnoliopsida 2248
## 7 chondrichthyes 1134
## 8 cephalopoda 750
## 9 gastropoda 633
## 10 pinopsida 610
## 11 cycadopsida 307
## 12 reptilia 225
## 13 myxini 76
```

Number of species in each category of population trend for each category of red list

```
list_pop = imp_data %>% group_by(populationTrend, redlistCategory) %>% summarize(Sum = n())
df = as.data.frame(list_pop)

df %>% pivot_wider(names_from=populationTrend, values_from=Sum)
```

```
## # A tibble: 10 x 6
##   redlistCategory      Decreasing Increasing Stable Unknown  'NA'
##   <chr>              <int>      <int> <int> <int> <int>
## 1 Critically Endangered      1108         19    42   343     9
## 2 Data Deficient           228         NA    72  4834    18
## 3 Endangered              2197         28    81   427     7
## 4 Extinct                  1         NA     1     4   279
## 5 Extinct in the Wild        1         NA    NA    NA    14
## 6 Least Concern           4469        775   9013  6170    46
## 7 Near Threatened          1558         36   314   424     1
## 8 Vulnerable              1946         42   325   552    63
## 9 Lower Risk/conservation dependent      NA         NA    NA    NA     1
## 10 Lower Risk/least concern      NA         NA    NA    NA     3
```

Species in each category for each category of red list and class

```
list_order = imp_data2 %>% group_by(redlistCategory, className) %>% summarize(Sum = n())
df2 = as.data.frame(list_order)

df3 = df2 %>% pivot_wider(names_from=redlistCategory, values_from=Sum)

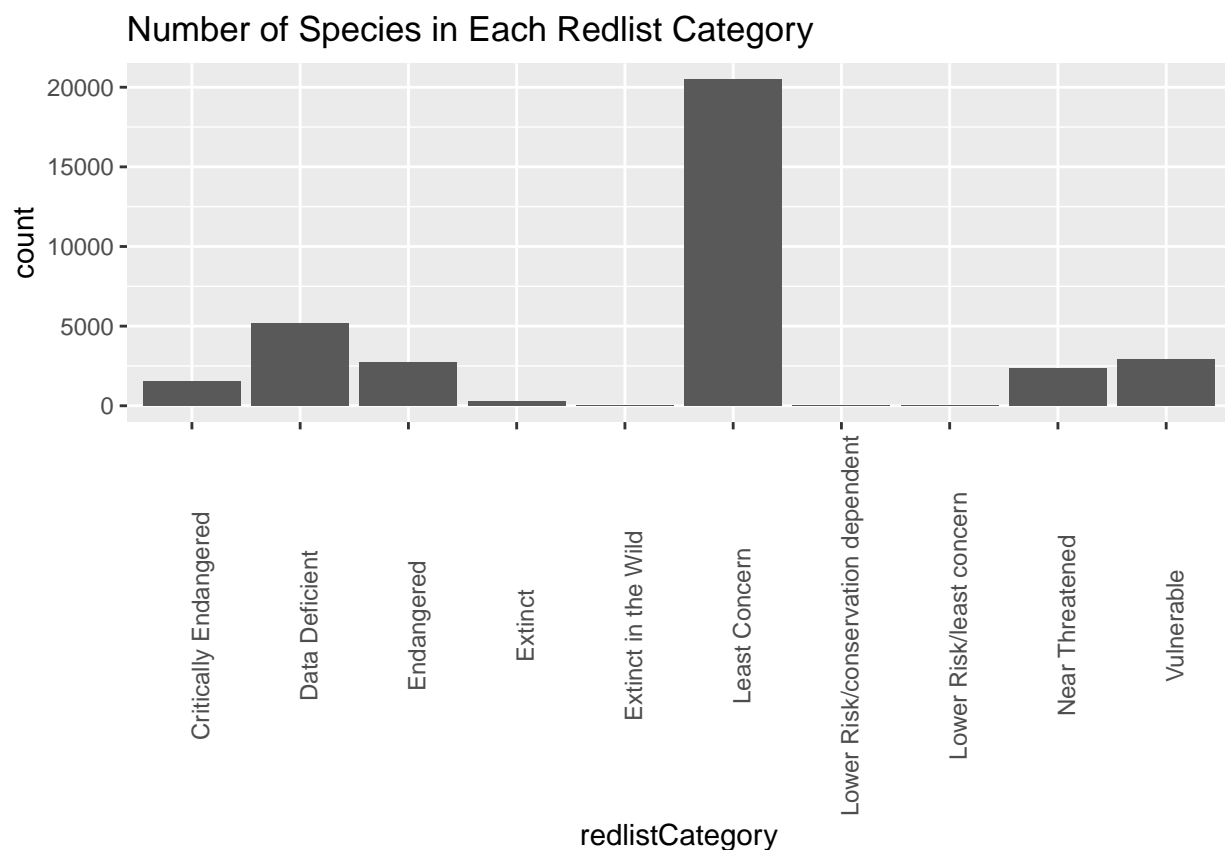
knitr::kable(df3)
```

	Critically Endangered	Data Deficient	Endangered	Extinct	Extinct in the Wild	Least Concern	Lower Risk/conservation dependent	Lower Risk/least concern	Near Threatened	Vulnerable
actinopterygii	10	444	34	NA	NA	1953	NA	NA	49	105
amphibia	610	1336	993	35	2	2856	NA	NA	387	673
aves	225	53	461	159	5	8427	NA	NA	1017	800
cephalopoda	1	419	2	NA	NA	324	NA	NA	2	2
chondrichthyes	43	420	62	NA	NA	382	NA	NA	115	112

className	Critically Endangered	Data Deficient	Endangered	Extinct	Extinct in the Wild	Least Concern	Risk/conservation dependent	Lower Risk/least concern	Near Threatened	Vulnerable
cycadopsida	53	3	65	NA	4	45	NA	NA	63	74
gastropoda	3	88	11	NA	NA	478	NA	NA	26	27
magnoliopsida	185	348	297	NA	2	1107	NA	NA	110	199
malacostraca	122	1124	143	7	NA	1189	NA	NA	71	279
mammalia	222	869	532	84	2	3292	NA	NA	353	545
myxini	1	30	2	NA	NA	35	NA	NA	2	6
pinopsida	29	8	96	NA	NA	298	NA	NA	99	80
reptilia	17	10	42	NA	NA	87	1	3	39	26

Redlist Categories

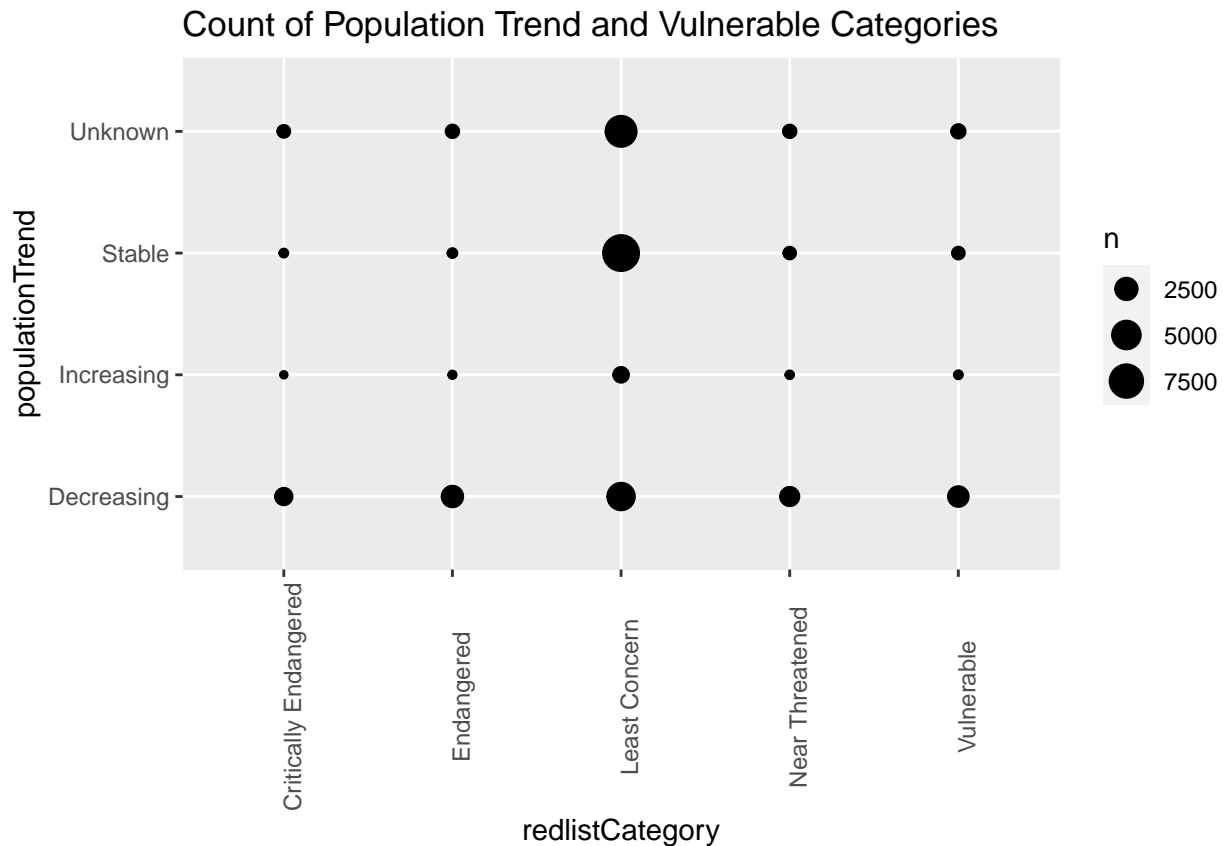
```
# redlist distribution
ggplot(imp_data, aes(x=redlistCategory)) + geom_bar() + theme(axis.text.x = element_text(angle = 90)) +
```



```
# filter data for high risk species
vul_data = imp_data %>% filter(redlistCategory == 'Vulnerable' | redlistCategory == 'Near Threatened' |
```

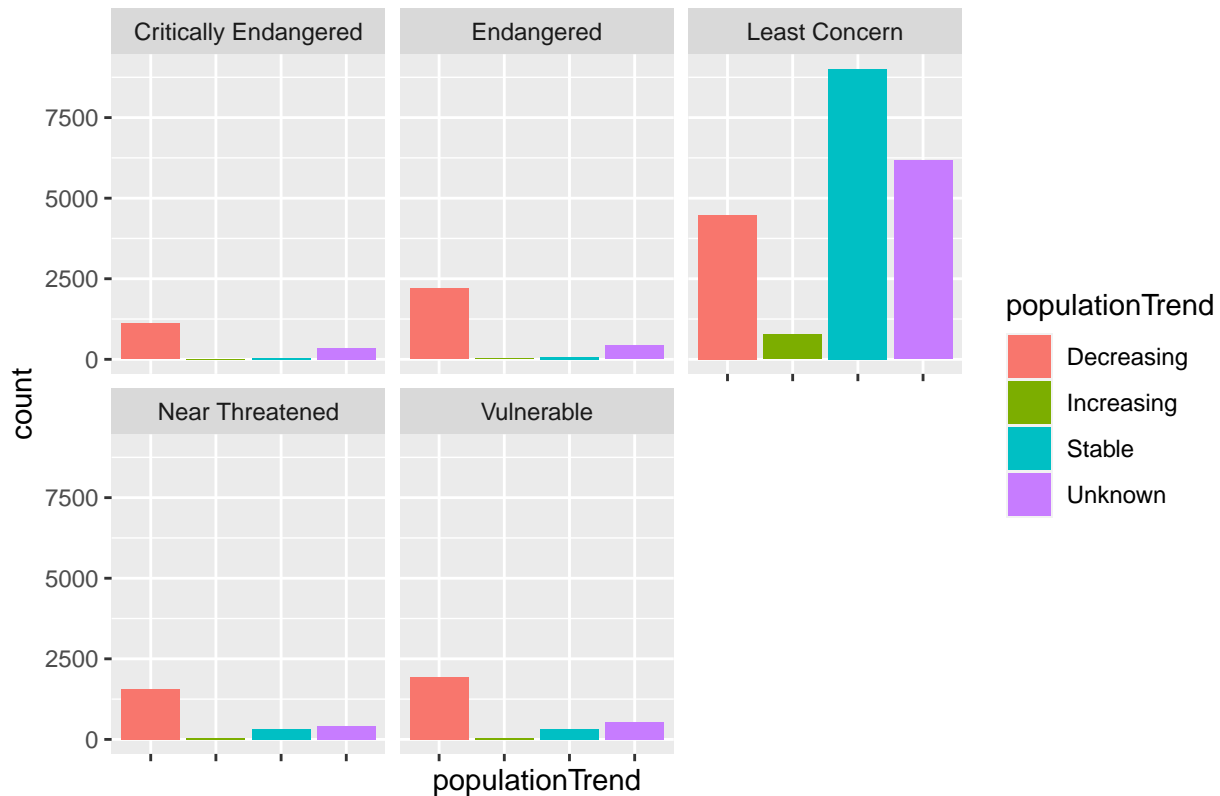
Compare populationTrend and redlistCategory

```
pop = imp_data %>% filter(populationTrend != 'NA') %>% filter(redlistCategory != 'Data Deficient') %>%
ggplot(pop, aes(redlistCategory, populationTrend)) + geom_count() + theme(axis.text.x = element_text(ang
```



```
ggplot(pop, aes(x=populationTrend, fill=populationTrend)) + geom_bar() + facet_wrap(~redlistCategory) +
```

Comparison of Population Trend and Redlist Category



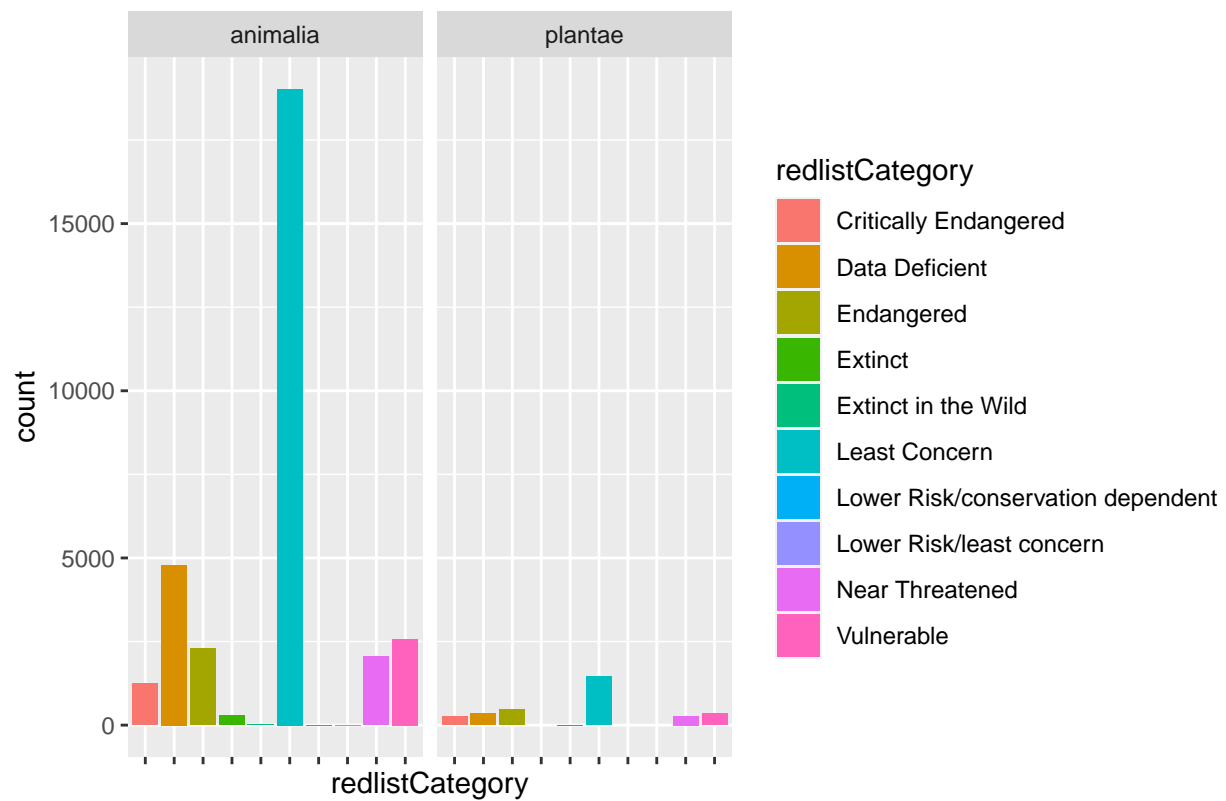
The largest redlistCategory is Least Concern, of which the population trends are more stable or decreasing (unknown excluded). Of the other redlist categories, greater numbers have decreasing population trends than stable/increasing.

Of the populations decreasing, the categories affected in order are least concern, endangered, vulnerable, near threatened, and critically endangered. The majority of populations that are stable or increasing are in the least concern category.

Comparison of redlistCategory and taxonomy

```
# comparison with kingdom
ggplot(imp_data2, aes(x=redlistCategory, fill=redlistCategory)) + geom_bar() + facet_wrap(~kingdomName)
```

Red List Category and Kingdom



comparison with phylum

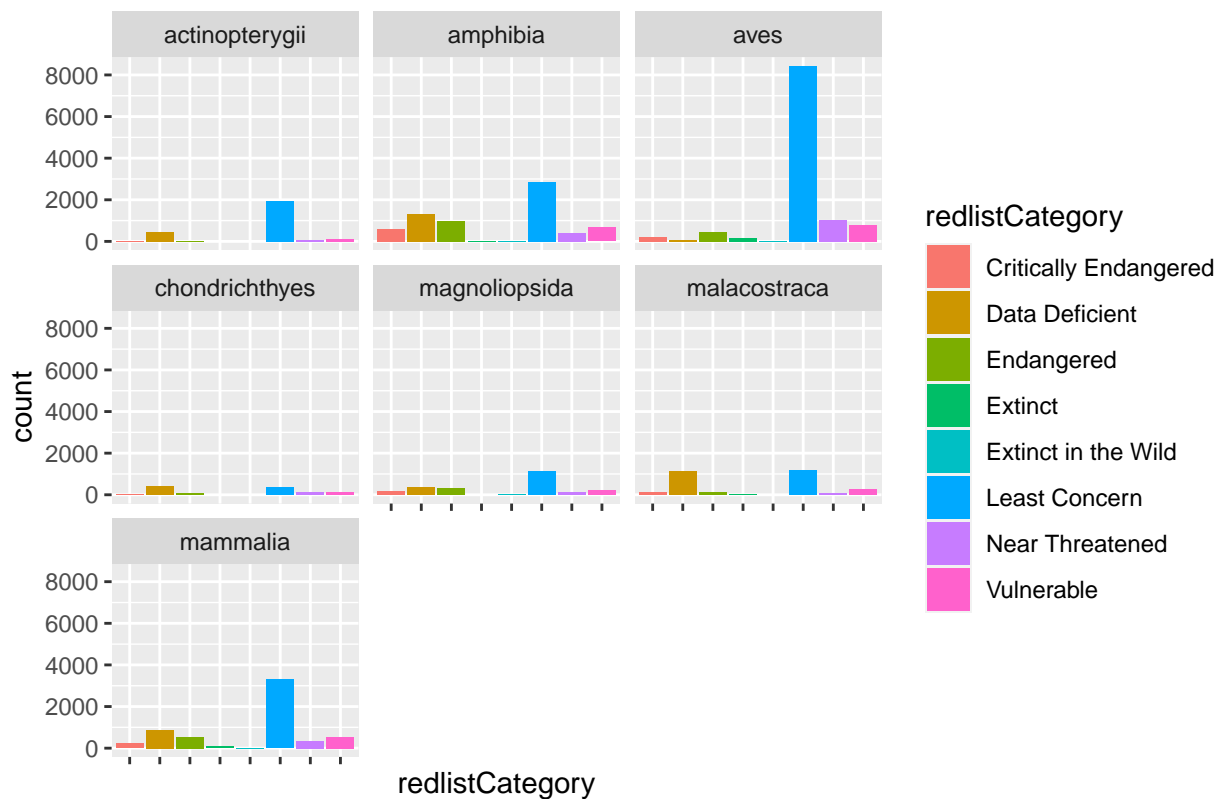
```
ggplot(imp_data2, aes(x=redlistCategory, fill=redlistCategory)) + geom_bar() + facet_wrap(~phylumName)
```


Red List Category and Phylum



```
# comparison with classes that have over 1000 species in the dataset
class_data = imp_data2 %>% filter(className != 'cephalopoda') %>% filter(className != 'gastropoda') %>%
ggplot(class_data, aes(x=redlistCategory, fill=redlistCategory)) + geom_bar() + facet_wrap(~className)
```

Red List Category and Class

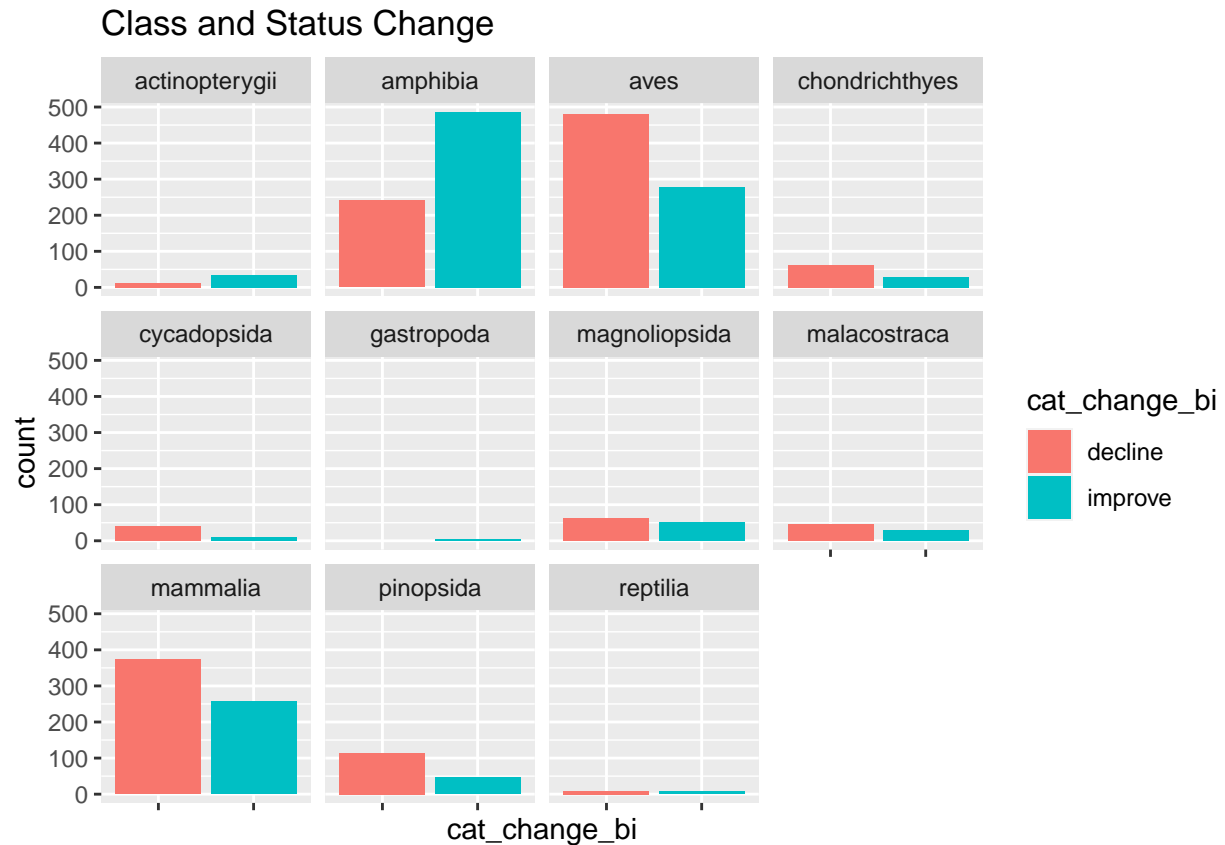


Amphibians (and mammals second) have the highest proportions with critically endangered, endangered, and vulnerable populations. Birds have high proportions of near threatened and vulnerable populations, but they also have the highest count of species in the least concern category.

Status Change variables

```
# binary change (NA removed)
change_bi = imp_data2 %>% filter(cat_change_bi != 'NA')

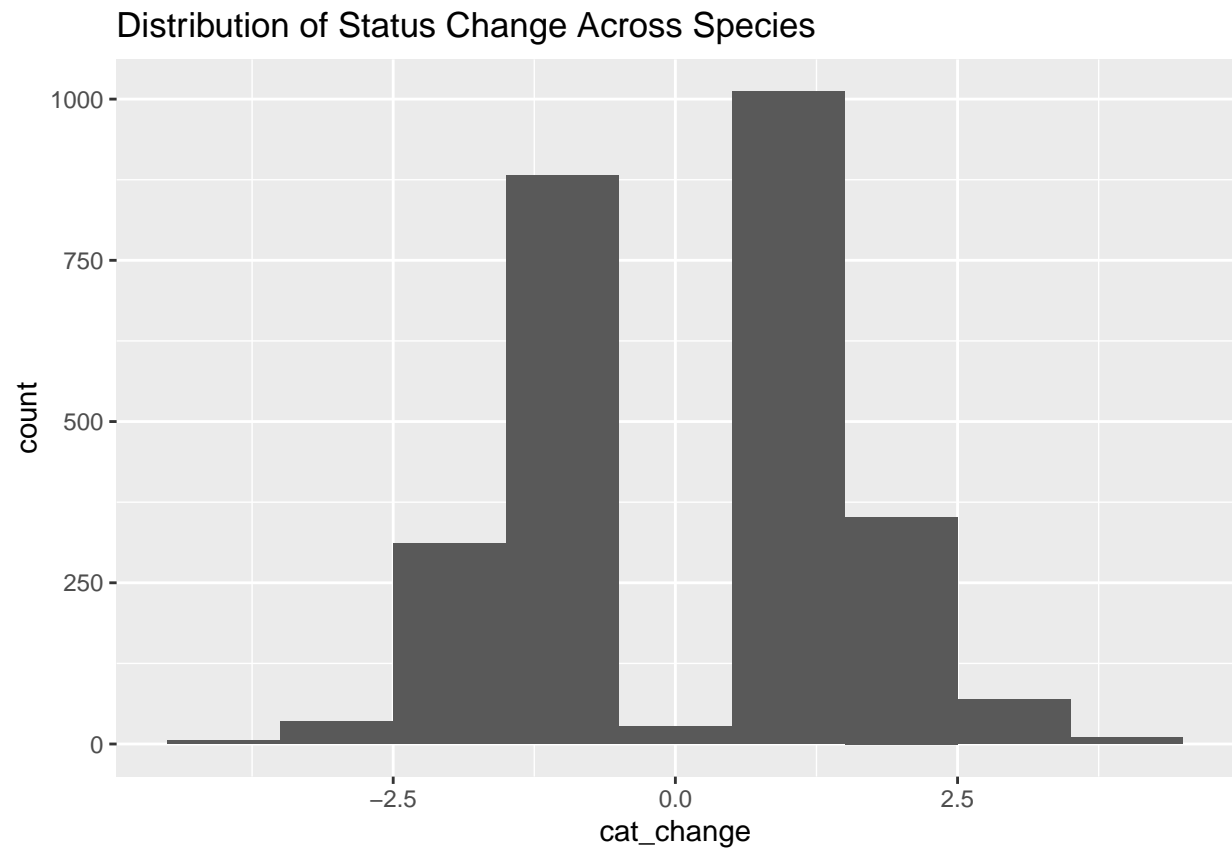
ggplot(change_bi, aes(x=cat_change_bi, fill=cat_change_bi)) + geom_bar() + facet_wrap(~className) + the
```



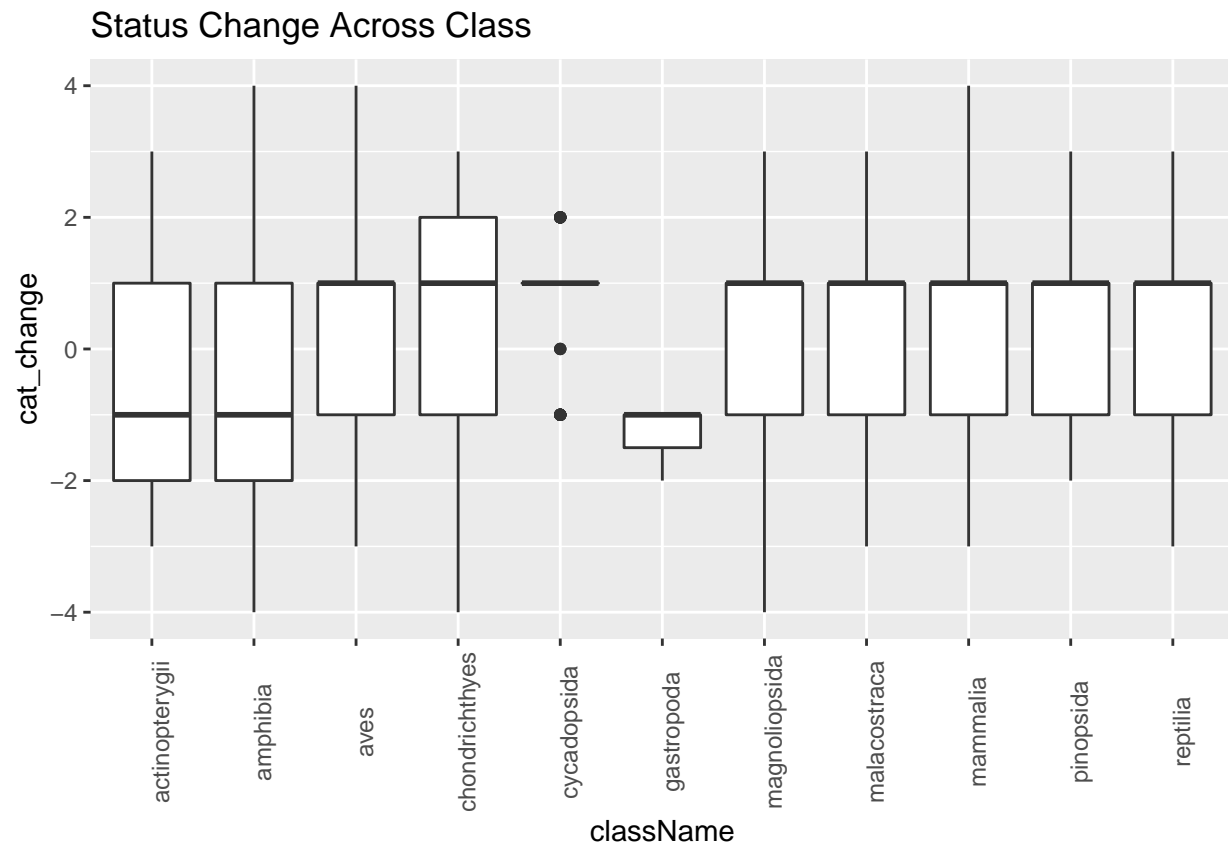
For binary changes, more amphibian and ray-finned fish (actinopterygii) species improved than declined. More bird, mammal, conifer, cycad, and cartilaginous fish species declined than improved (in decreasing order).

```
# non-binary change
change = imp_data2 %>% filter(cat_change != 'NA')

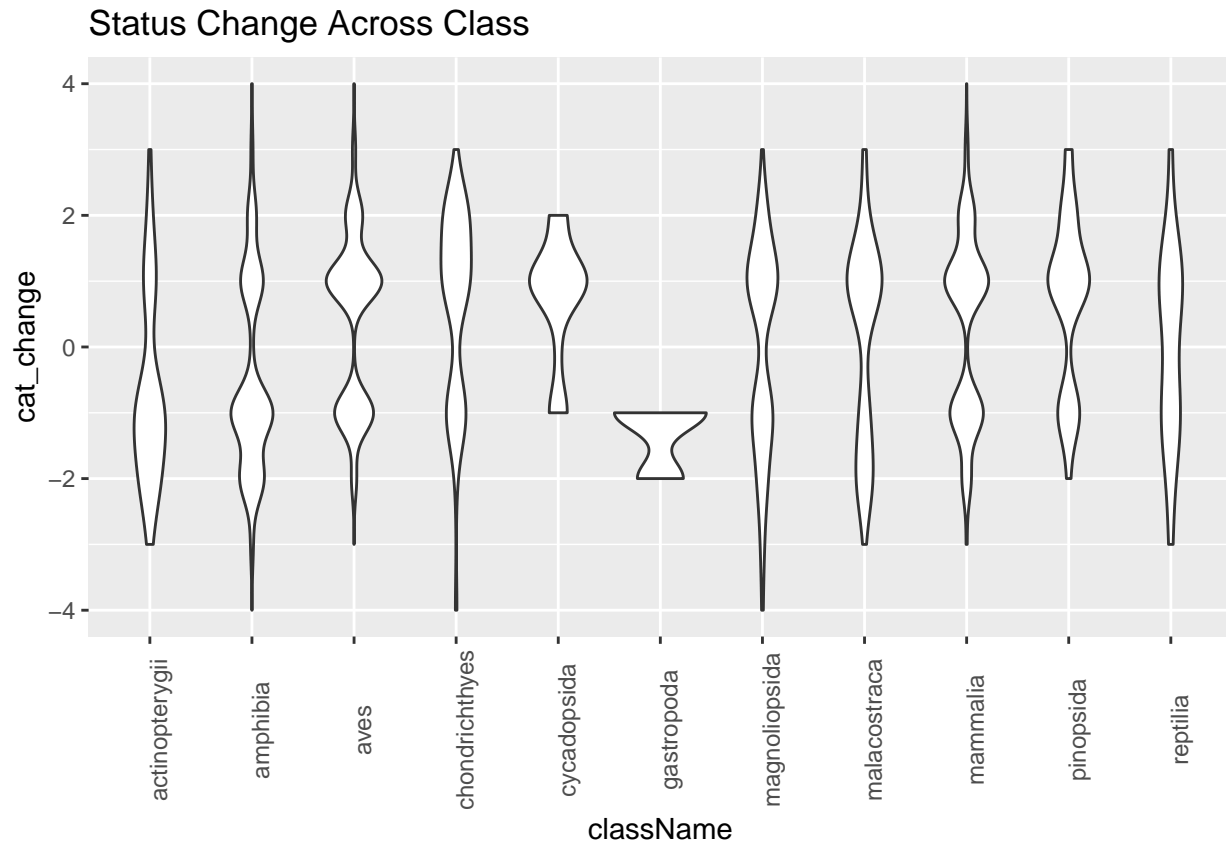
ggplot(change, aes(x=cat_change)) + geom_histogram(binwidth=1) + ggtitle("Distribution of Status Change")
```



```
ggplot(change, aes(x=className, y=cat_change)) + geom_boxplot() + theme(axis.text.x = element_text(angel
```



```
ggplot(change, aes(x=className, y=cat_change)) + geom_violin() + theme(axis.text.x = element_text(angle=
```



For nonbinary changes, the classes with the median `cat_change` below 0 (improve, downlisting) are ray-finned fish, amphibians, and gastropoda. The remaining species have a median change above 0. However, it is notable that all species have an interquartile range that encompasses 0 except for cycadopsida and gastropoda. Birds, mammals, reptiles, and most plants have the median at the upper end of the IQR and above 0 (decline, uplisting).

Body Mass Calculations

```
# average body mass of each red list category and overall
mass = imp_data2 %>% filter(BodyMass != 'NA')
mean(mass$BodyMass)
```

```
## [1] 55947.81
```

```
mass_list = mass %>% group_by(redlistCategory) %>% summarize(AvgMass = mean(BodyMass), MedMass = median(BodyMass))
```

```
## # A tibble: 8 x 3
##   redlistCategory    AvgMass MedMass
##   <chr>             <dbl>   <dbl>
## 1 Endangered       299307.   131
## 2 Data Deficient   116923.   33.4
## 3 Extinct in the Wild 87280. 74531.
## 4 Critically Endangered 86282.   200.
```

```
## 5 Vulnerable      80451.    104.
## 6 Near Threatened 41419.     71
## 7 Least Concern   29065.    36.0
## 8 Extinct         8582.     200
```

```
# same but with population trend
```

```
mass_pop = mass %>% group_by(populationTrend) %>% summarize(AvgMass = mean(BodyMass), MedMass = median(BodyMass))
mass_pop
```

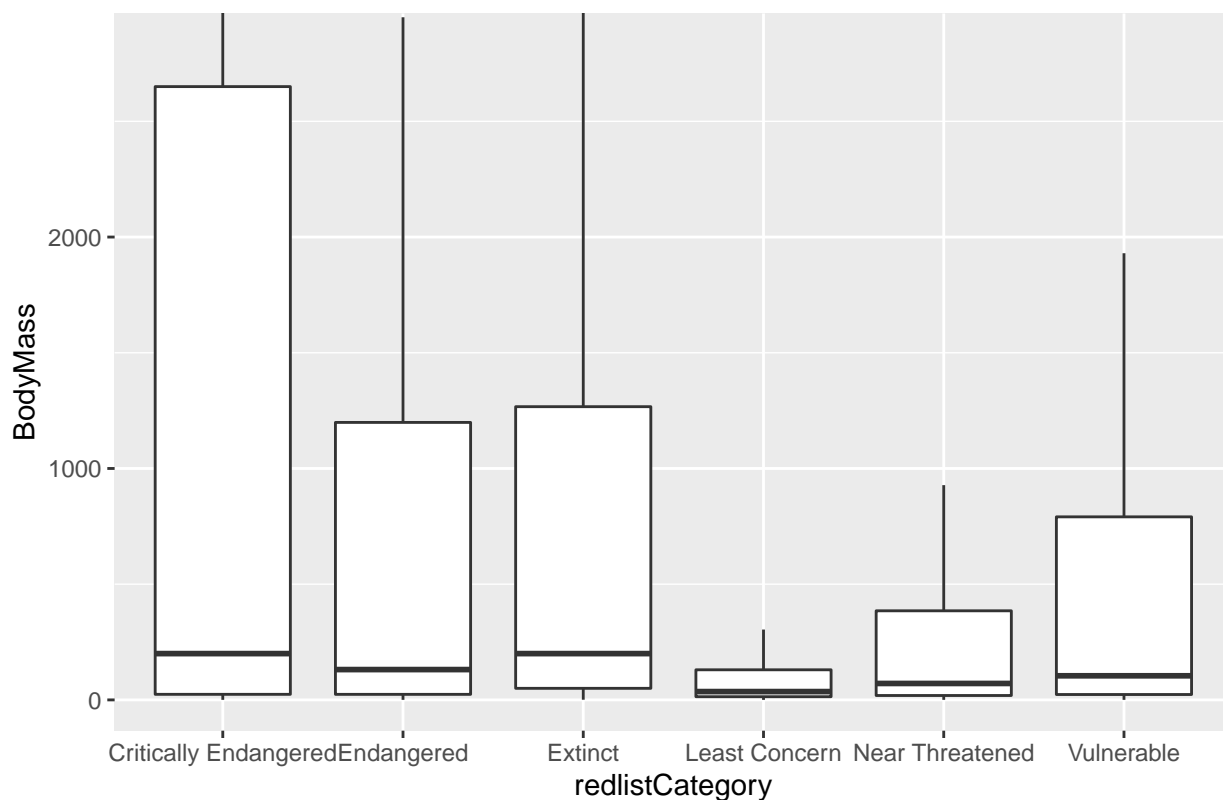
```
## # A tibble: 5 x 3
##   populationTrend AvgMass MedMass
##   <chr>          <dbl>   <dbl>
## 1 Increasing    667111.    78.5
## 2 Unknown       91307.    37.3
## 3 <NA>         13426.    200
## 4 Decreasing    9422.     69
## 5 Stable        8280.    30.0
```

```
# body mass boxplot without outliers
```

```
body_out = imp_data %>% filter(BodyMass != 'NA') %>% filter(redlistCategory != 'Data Deficient') %>% filter(populationTrend != 'Decreasing')
```

```
ggplot(body_out, aes(x=redlistCategory, y=BodyMass)) + geom_boxplot(outlier.shape = NA) + coord_cartesian(ylim=c(0, 2500))
```

Distribution of Body Mass Across Redlist Category (outliers excluded)



The average body mass is generally higher for uplisted species on the red list. The average body mass is generally higher for species who have increasing population trends as compared to those with decreasing or stable trends (this may be in part due to differences in sample size).

The median body mass is lower for species of least concern and near threatened as compared to the median body mass of other categories. (when outliers are ignored)

Systems

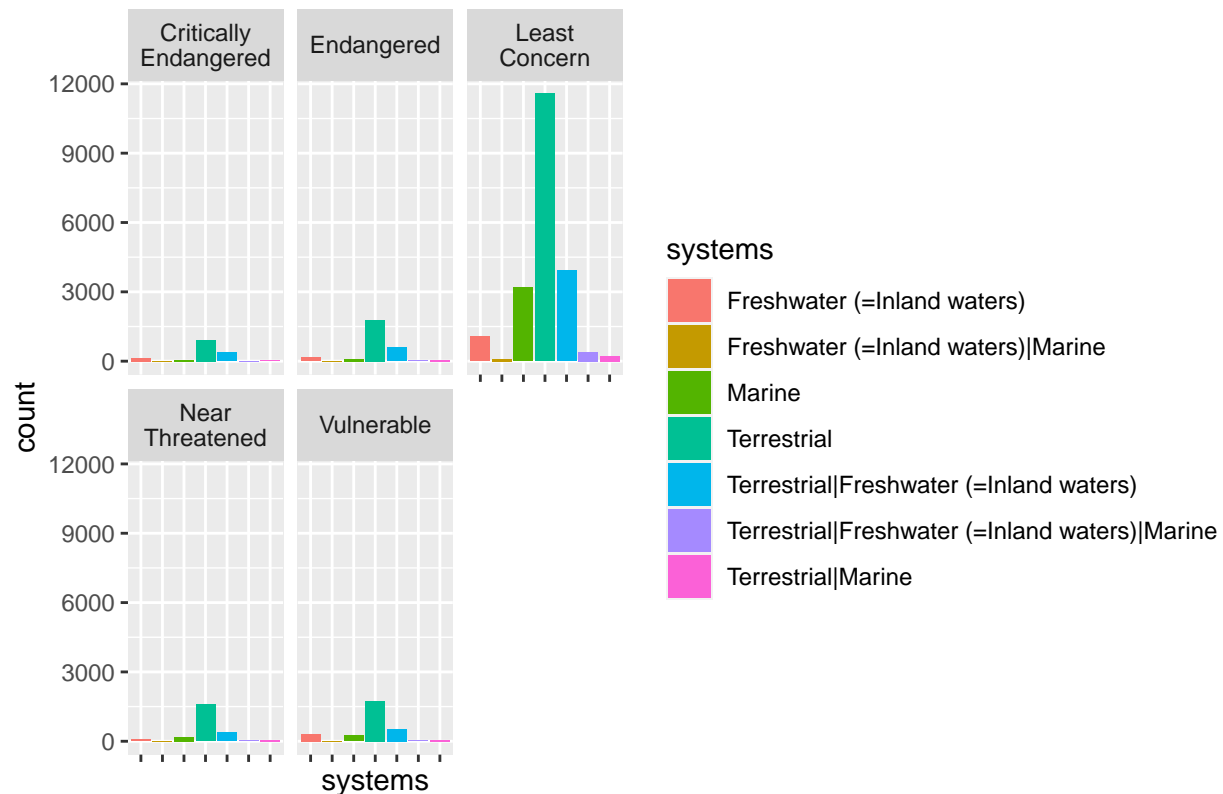
I thought this might be interesting, even though you didn't mention looking at it.

```
# filter out small, NA systems
system = imp_data2 %>% filter(systems != '<NA>') %>% filter(systems != 'Marine|Marine') %>% filter(systems != 'Freshwater (=Inland waters)|Marine')

# filter irrelevant categories
system2 = system %>% filter(redlistCategory == 'Critically Endangered' | redlistCategory == 'Endangered' | redlistCategory == 'Least Concern' | redlistCategory == 'Near Threatened' | redlistCategory == 'Vulnerable')

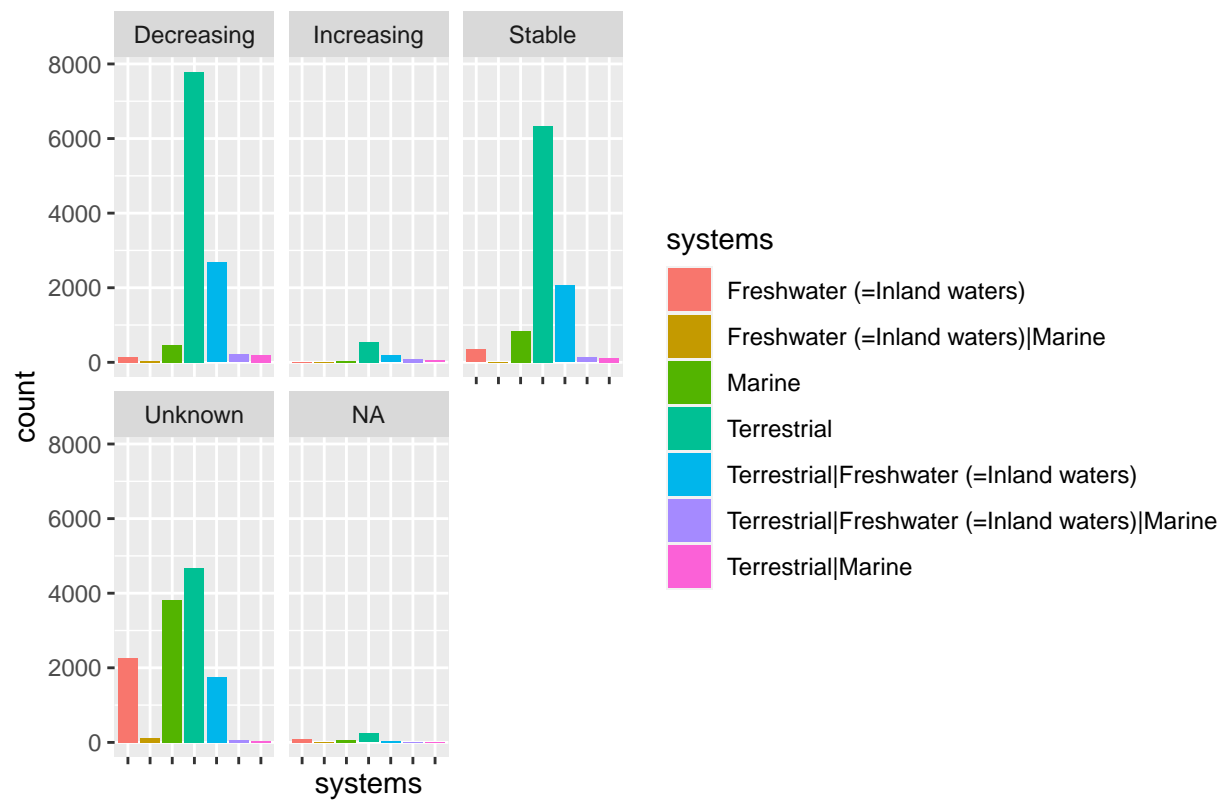
# compare system and redlistCategory
ggplot(system2, aes(x=systems, fill=systems)) + geom_bar() + facet_wrap(~redlistCategory, labeller = label_both)
```

Comparison of Redlist Category and System



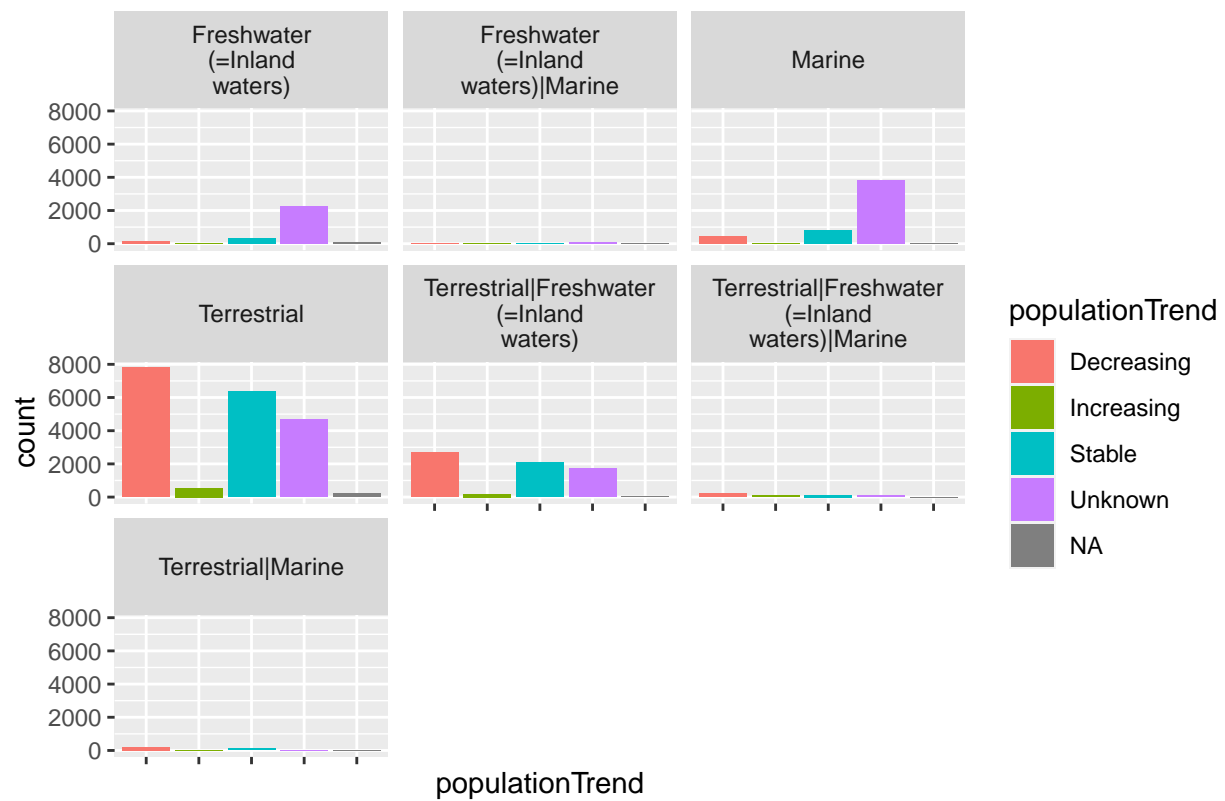
```
# compare system and populationTrend
ggplot(system, aes(x=systems, fill=systems)) + geom_bar() + facet_wrap(~populationTrend) + theme(axis.title.x = 'systems')
```


Comparison of Population Trend and System



```
ggplot(system, aes(x=populationTrend, fill=populationTrend)) + geom_bar() + facet_wrap(~systems, labell
```

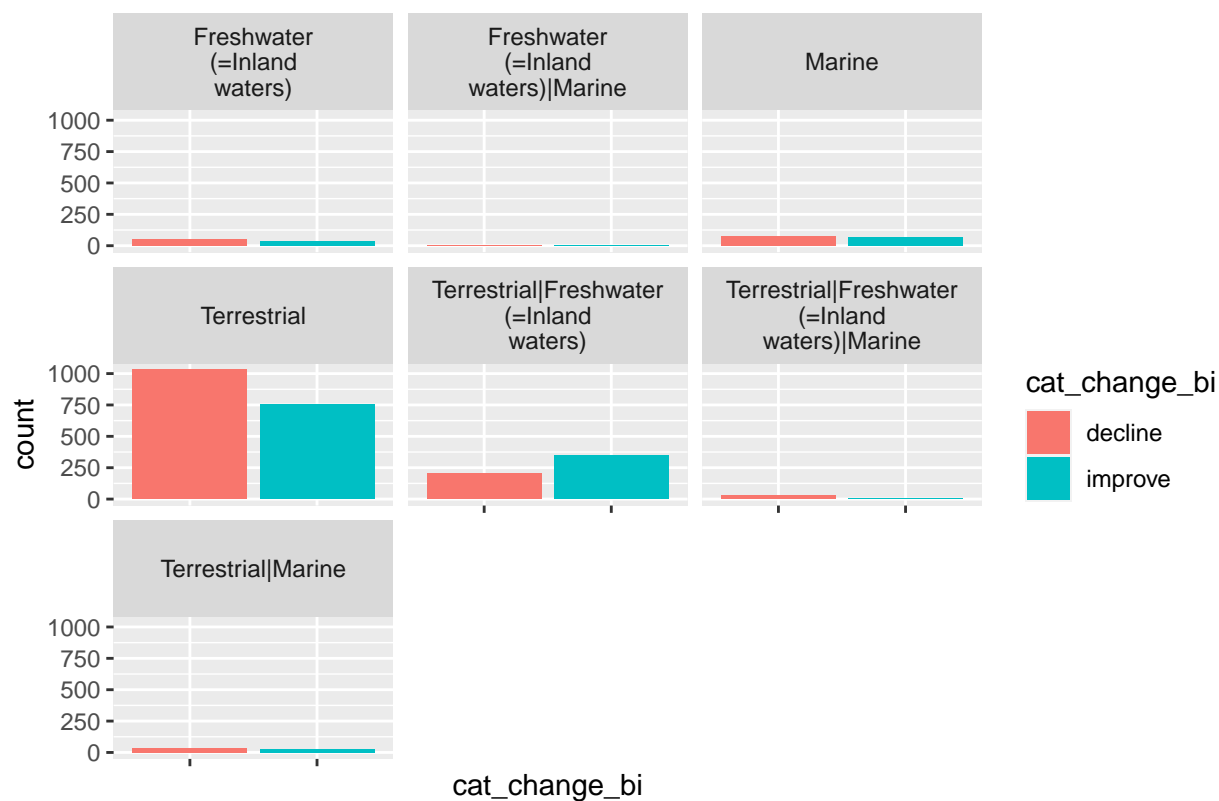
Comparison of Population Trend and System



```
# compare system and cat_change_bi
system_bi = system %>% filter(cat_change_bi != 'NA')

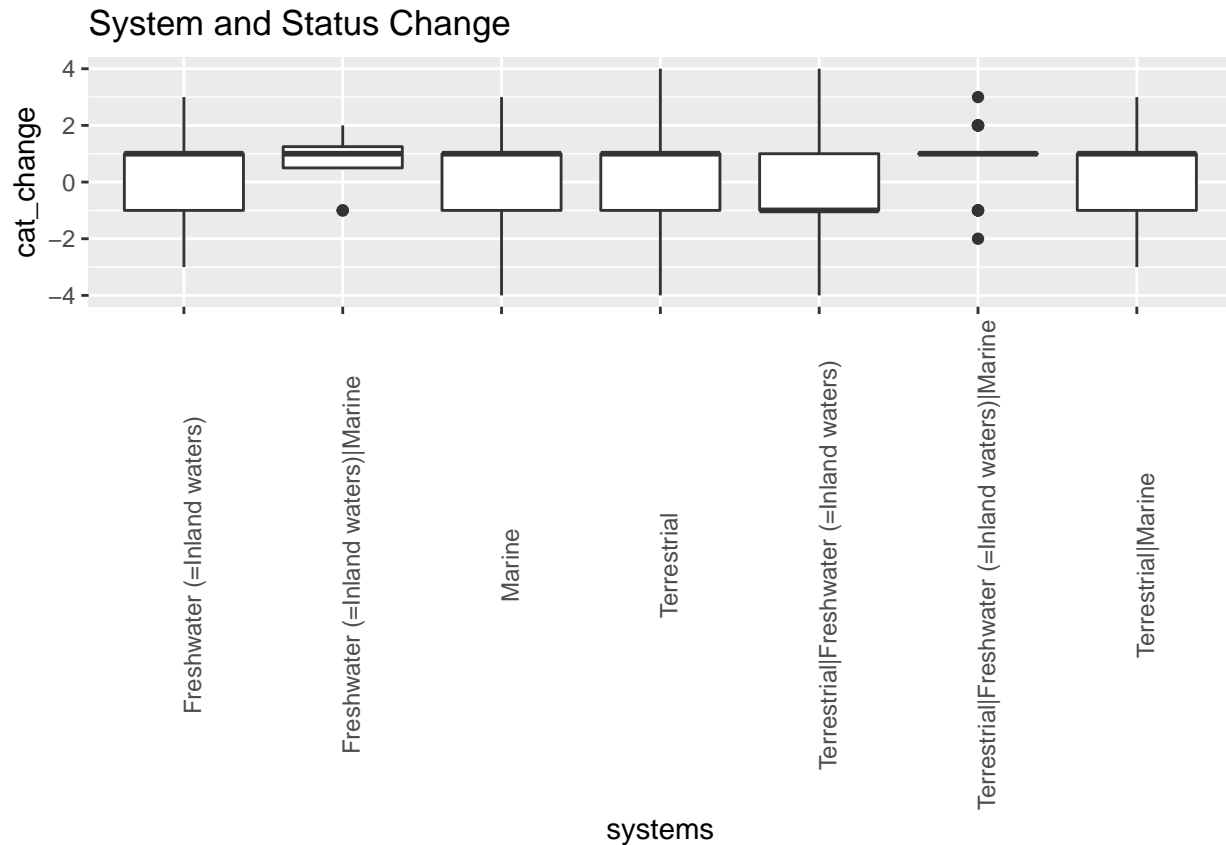
ggplot(system_bi, aes(x=cat_change_bi, fill=cat_change_bi)) + geom_bar() + facet_wrap(~systems, labeller=
```

System and Status Change



```
# system and cat_change
system_change = system %>% filter(cat_change != 'NA')

ggplot(system_change, aes(x=systems, y=cat_change)) + geom_boxplot() + theme(axis.text.x = element_text
```



The majority of systems are in the least concern category. However, of the categories that are not least concern, the systems with the most species are terrestrial and freshwater inland waters. A small portion of Near Threatened and Vulnerable are marine.

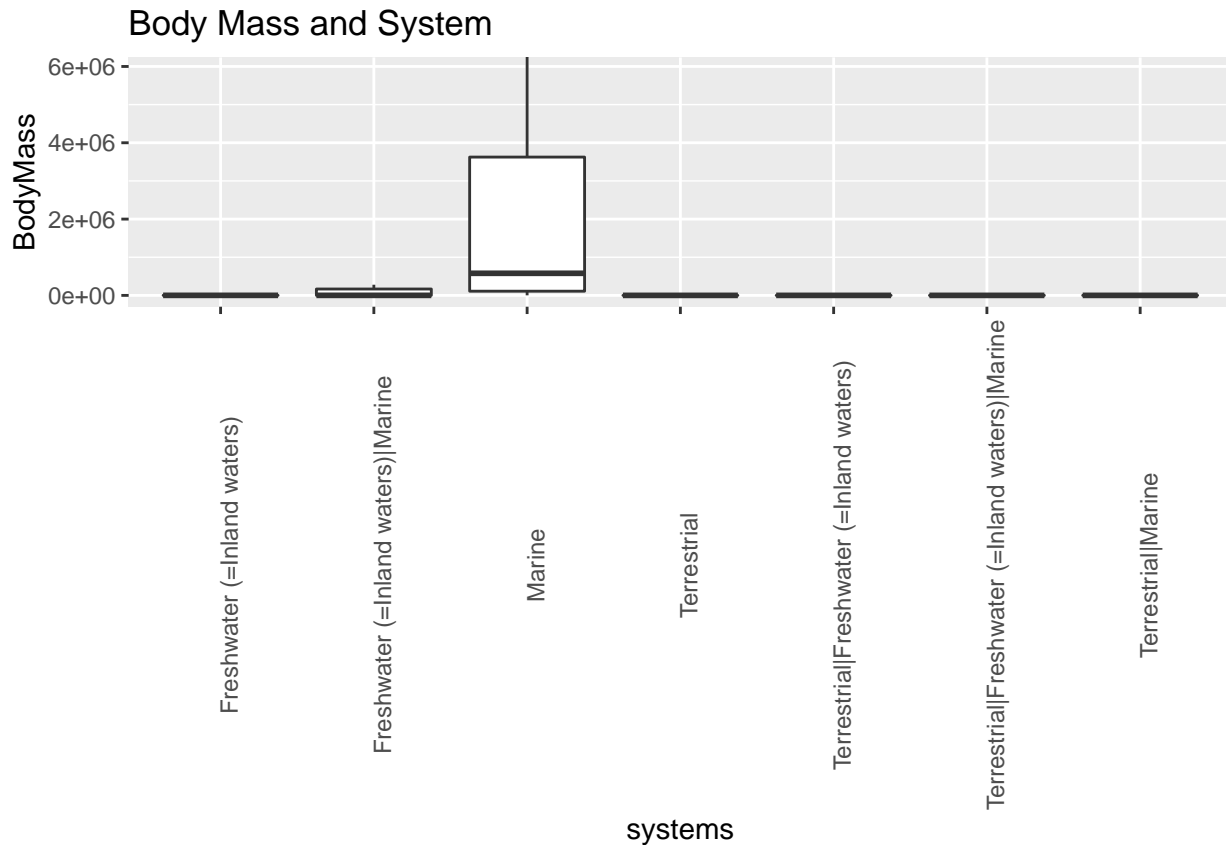
Similarly, terrestrial and freshwater inland water systems have the highest number of species with decreasing population trends. The systems with stable species are in decreasing order: terrestrial, terrestrial freshwater inland, marine, freshwater. Only terrestrial has a noticeable number of species that have increasing trends.

Within the binary change categories, terrestrial had the most change, with more species in decline than improving. Terrestrial freshwater inland had more improved than decline. Marine systems were relatively equal.

Within nonbinary change, the species with median negative -1 change were terrestrial inland freshwater. The remaining systems had a median positive +1 change. ALL IQRs were from -1 to 1 except for freshwater/marine and terrestrial/freshwater/marine, which both had IQRs above 0 cat_change.

For fun

```
# compare system and class, bodymass
ggplot(system, aes(x=systems, y=BodyMass)) + geom_boxplot(outlier.shape = NA) + coord_cartesian(ylim = c(-4, 4))
```



average body mass of each system

```
mass_system = mass %>% group_by(systems) %>% summarize(AvgMass = mean(BodyMass), MedMass = median(BodyMass))
mass_system
```

```
## # A tibble: 7 x 3
```

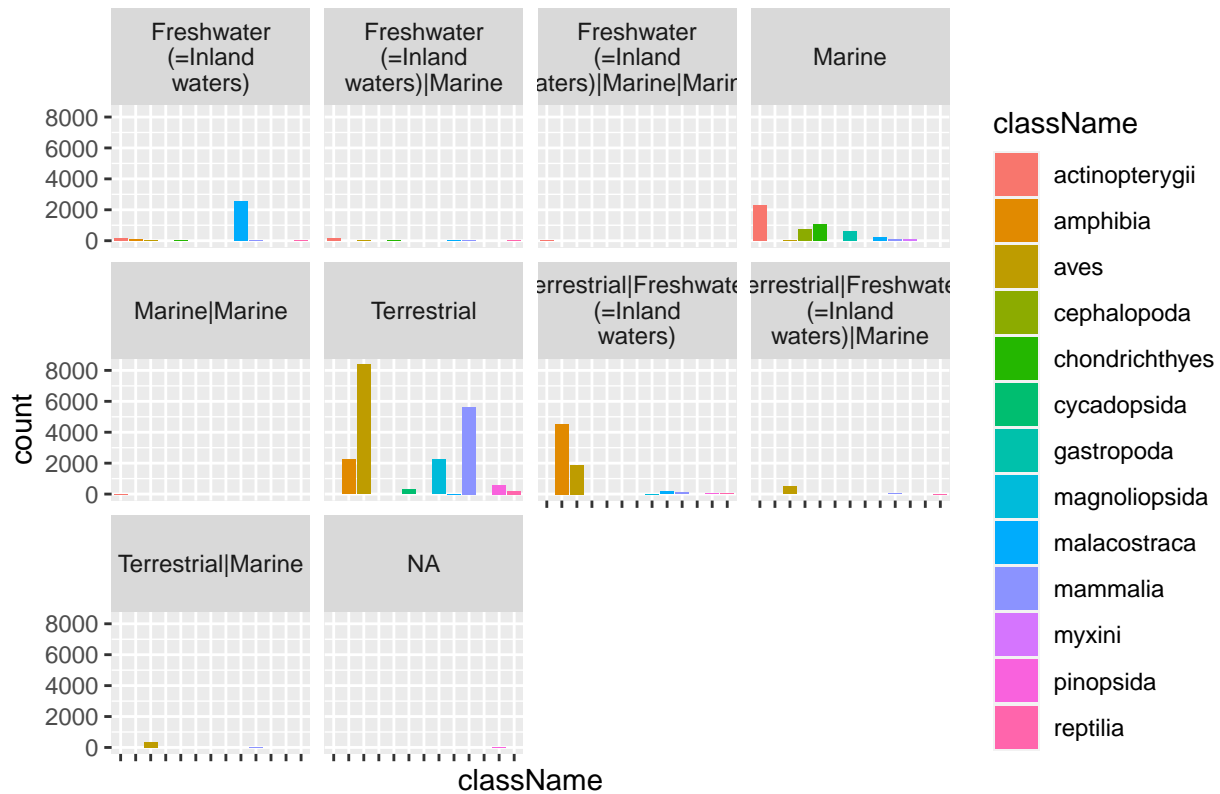
systems	AvgMass	MedMass
<chr>	<dbl>	<dbl>
1 Marine	8676310.	578750
2 Freshwater (=Inland waters) Marine	104567.	1156.
3 Freshwater (=Inland waters)	27856.	681.
4 Terrestrial Marine	25999.	490.
5 Terrestrial Freshwater (=Inland waters) Marine	5061.	323.
6 Terrestrial	3650.	39.9
7 Terrestrial Freshwater (=Inland waters)	2300.	34.9

turns out marine is like wayyyyyy bigger !!

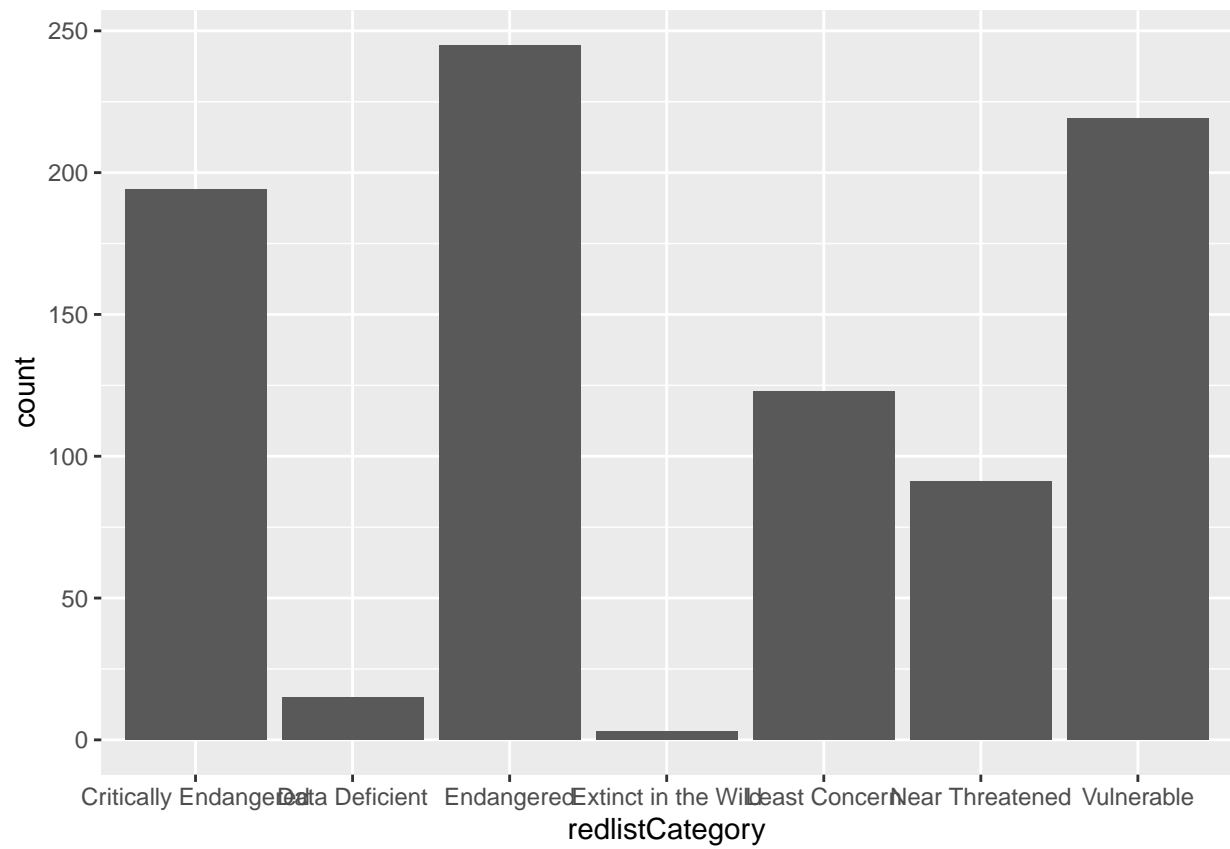
class

```
ggplot(imp_data2, aes(x=className, fill=className)) + geom_bar() + facet_wrap(~systems, labeller = label_both)
```

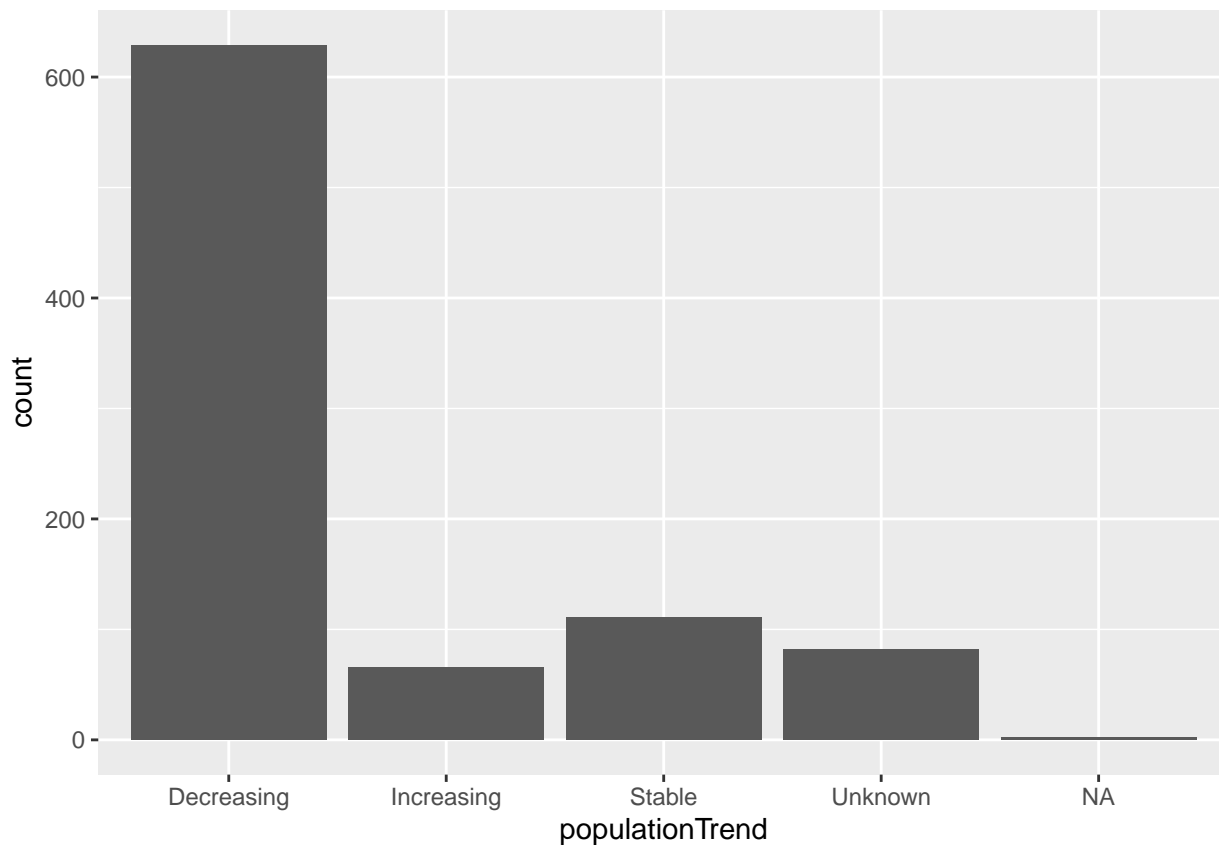
System and Class



```
# education
# it seems education is focused on more threatened species
edu = data %>% filter(education == TRUE)
ggplot(edu, aes(x=redlistCategory)) + geom_bar()
```



```
ggplot(edu, aes(x=populationTrend)) + geom_bar()
```



Amphibian Study

```
amphibian = data %>% filter(className == 'amphibia') %>% filter(cat_change_bi != 'NA')
amphibian %>% group_by(familyName) %>% summarize(Sum = n()) %>% arrange(-Sum)
```

```
## # A tibble: 45 x 2
##   familyName      Sum
##   <chr>          <int>
## 1 craugastoridae  113
## 2 bufonidae       83
## 3 plethodontidae  60
## 4 rhacophoridae   45
## 5 ranidae         36
## 6 microhylidae    33
## 7 hyperoliidae    32
## 8 arthroleptidae  31
## 9 hylidae         31
## 10 mantellidae    30
## # ... with 35 more rows
```

```
amphibian %>% group_by(systems) %>% summarize(Sum = n()) %>% arrange(-Sum)
```



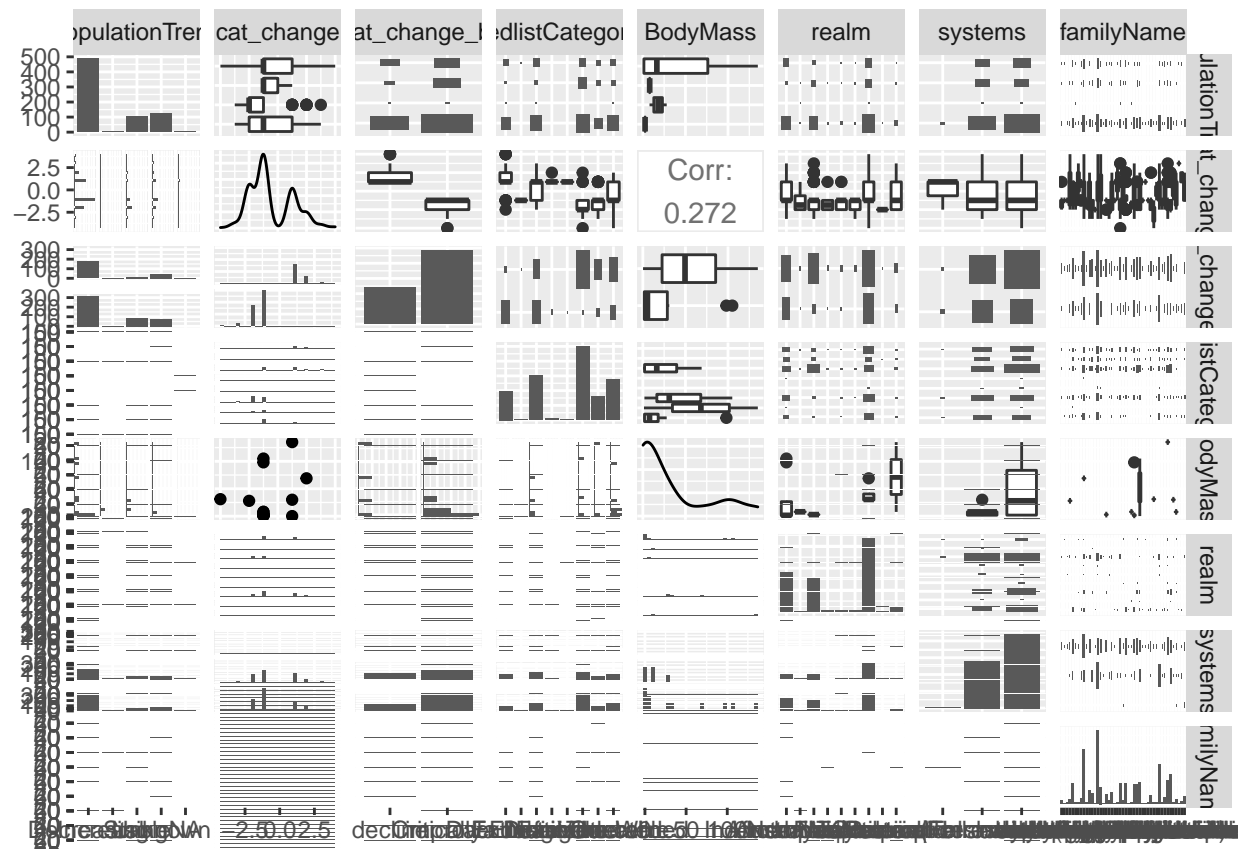
```
## # A tibble: 3 x 2
##   systems      Sum
##   <chr>      <int>
## 1 Terrestrial|Freshwater (=Inland waters)    441
## 2 Terrestrial                               280
## 3 Freshwater (=Inland waters)                6
```

```
amphibian %>% group_by(realm) %>% summarize(Sum = n()) %>% arrange(-Sum)
```

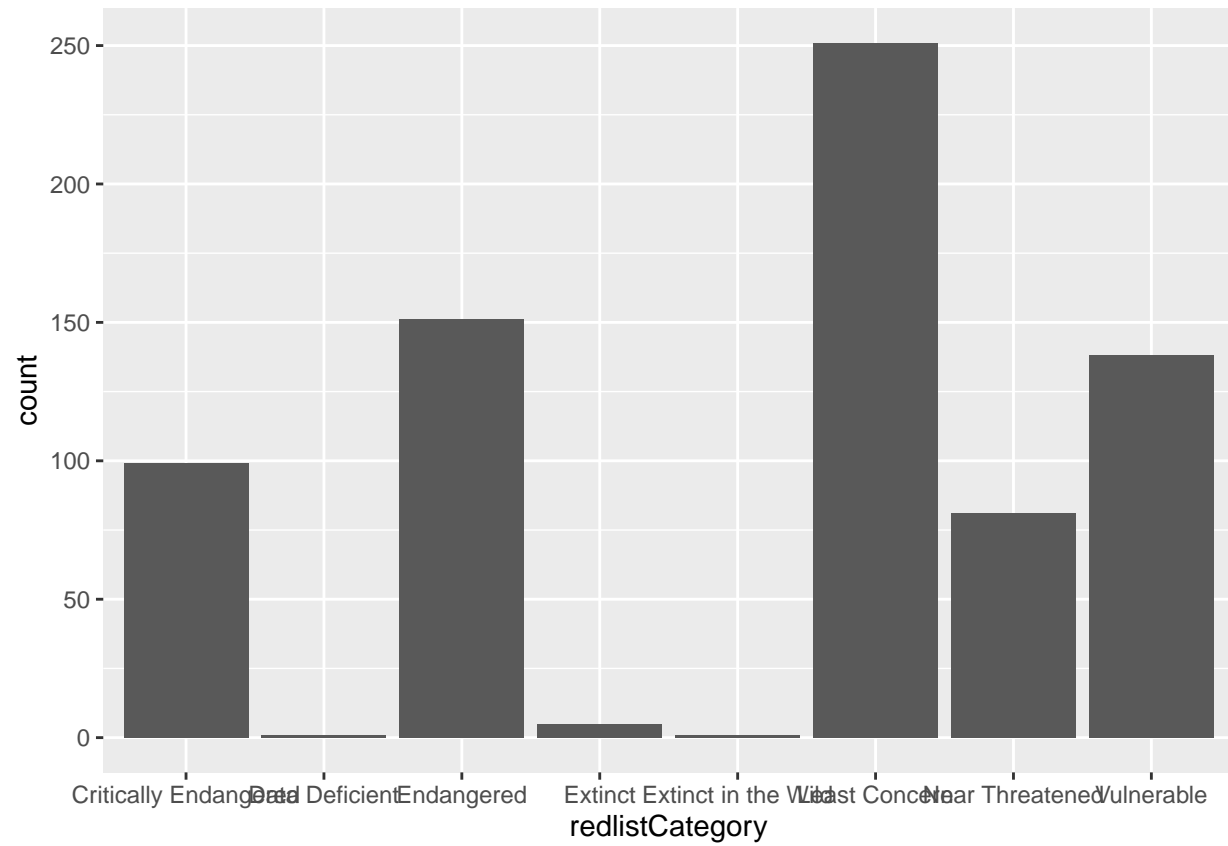
```
## # A tibble: 9 x 2
##   realm      Sum
##   <chr>    <int>
## 1 Neotropical    337
## 2 Afrotropical  181
## 3 Indomalayan   157
## 4 Palearctic     25
## 5 Nearctic        9
## 6 Nearctic|Neotropical    7
## 7 Indomalayan|Palearctic    6
## 8 Australasian    4
## 9 Oceanian        1
```

```
# pull out population trend, cat_change, cat_change_bi, redlistCategory, bodymass, realm, systems, familyName
amphibian_set = amphibian %>% select(populationTrend, cat_change, cat_change_bi, redlistCategory, BodyMass, realm, systems, familyName)

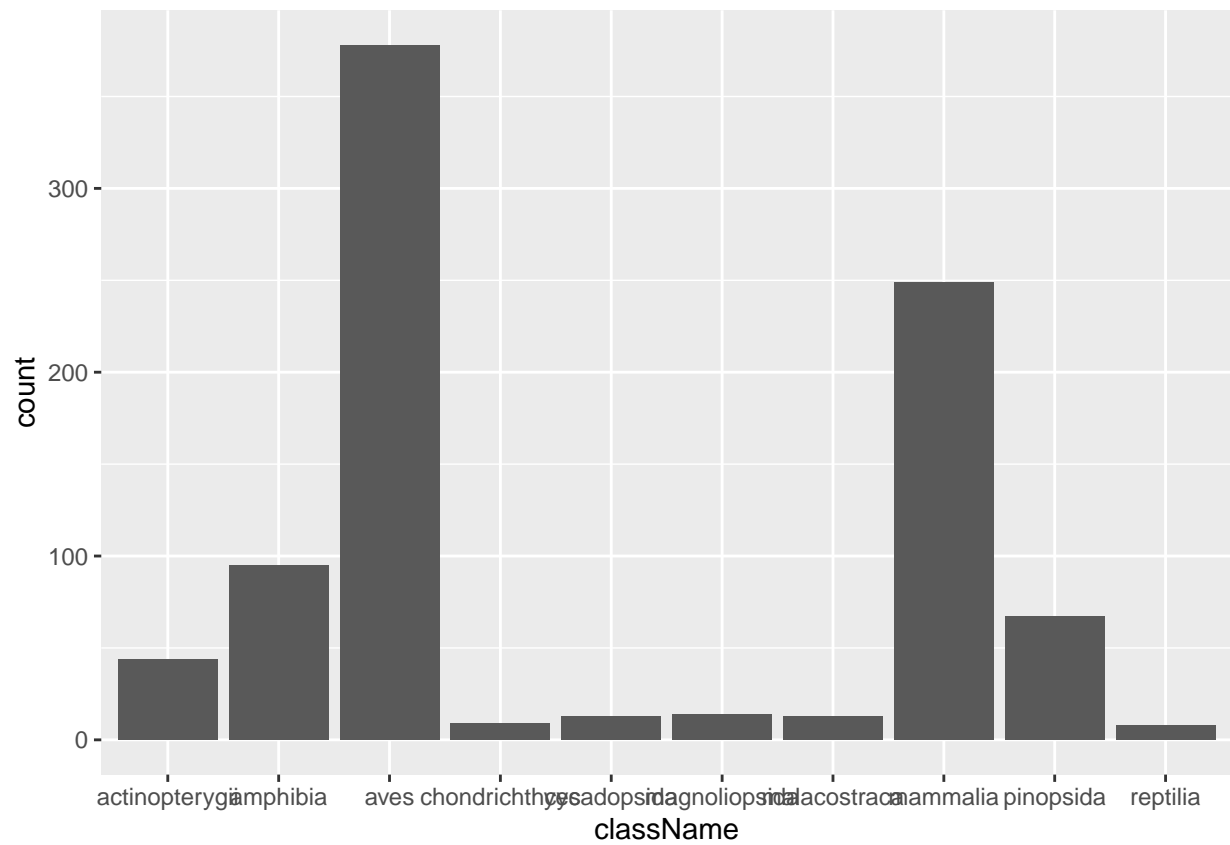
ggpairs(amphibian_set, cardinality_threshold=80)
```



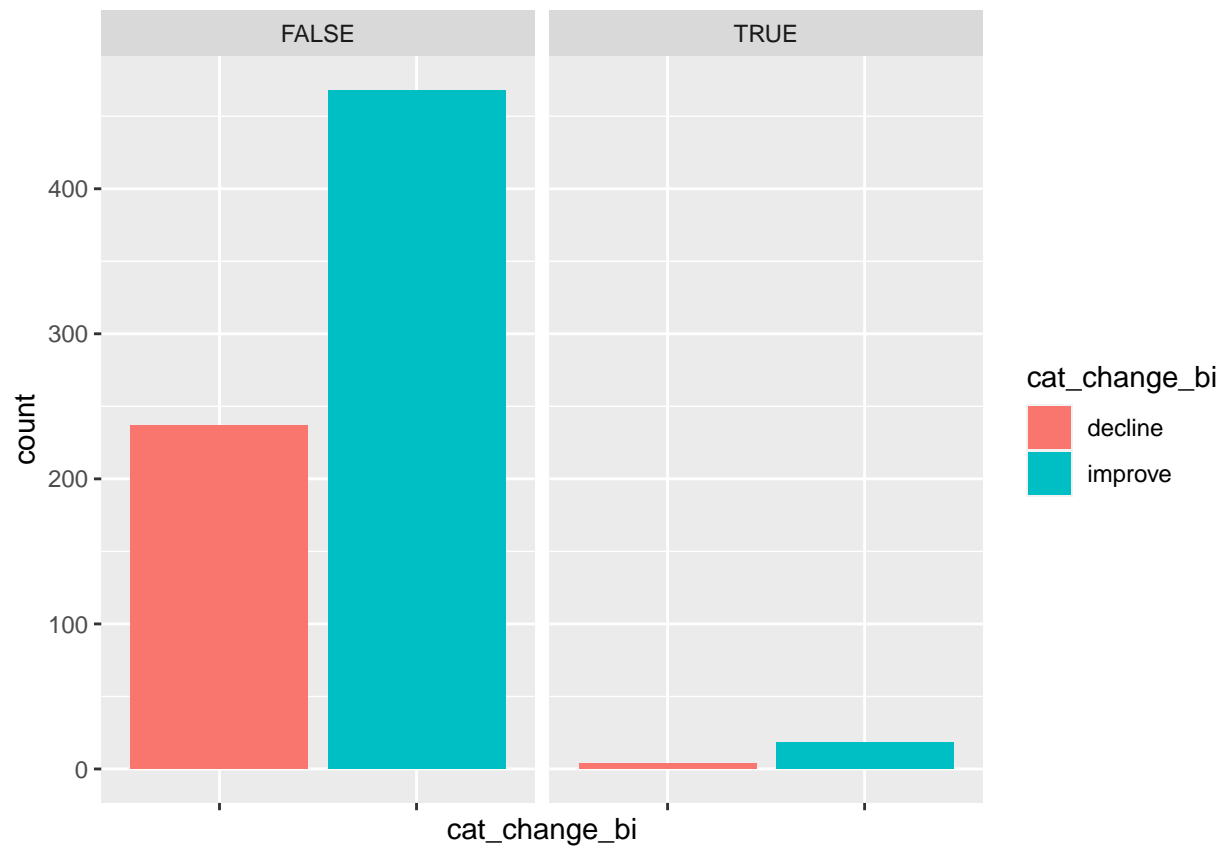
```
ggplot(amphibian, aes(x=redlistCategory)) + geom_bar()
```



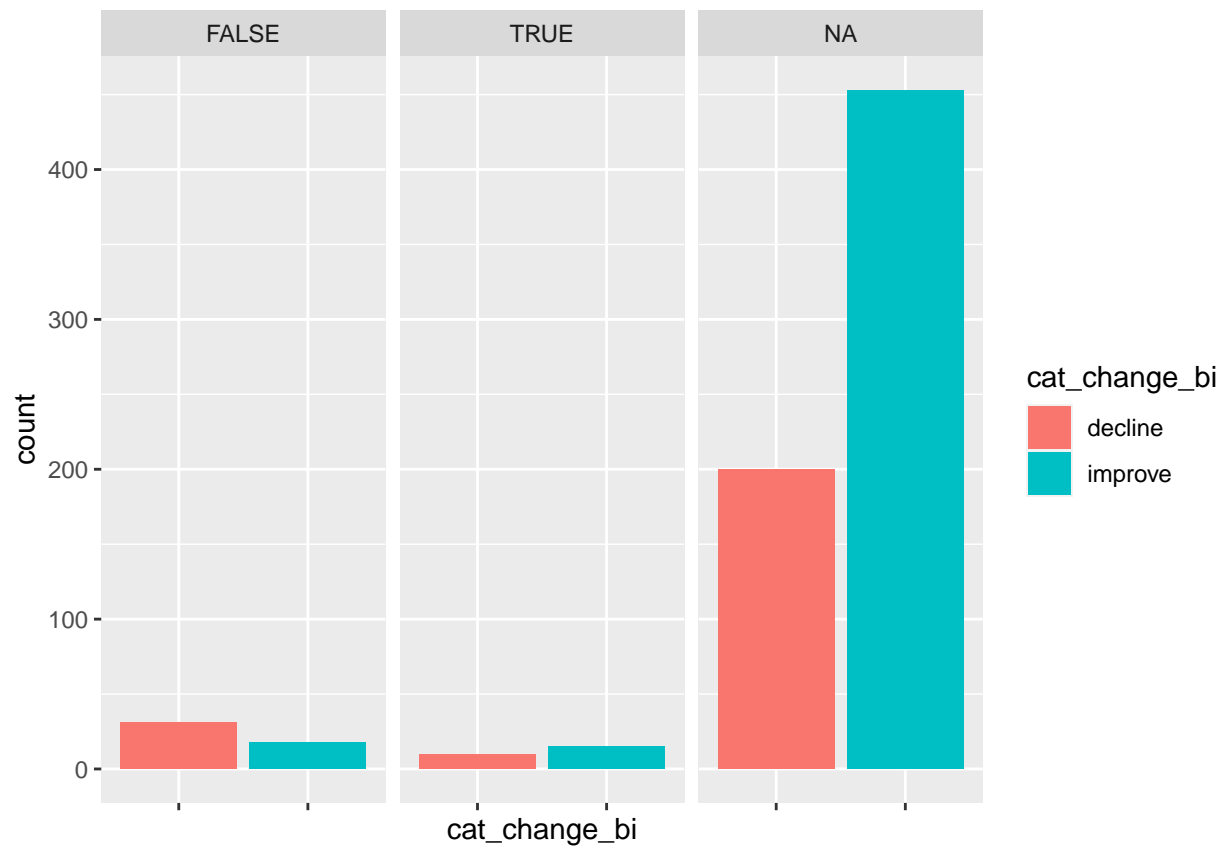
```
# education and cat_change_bi
education = data %>% filter(education == TRUE)
ggplot(education, aes(x=className)) + geom_bar()
```



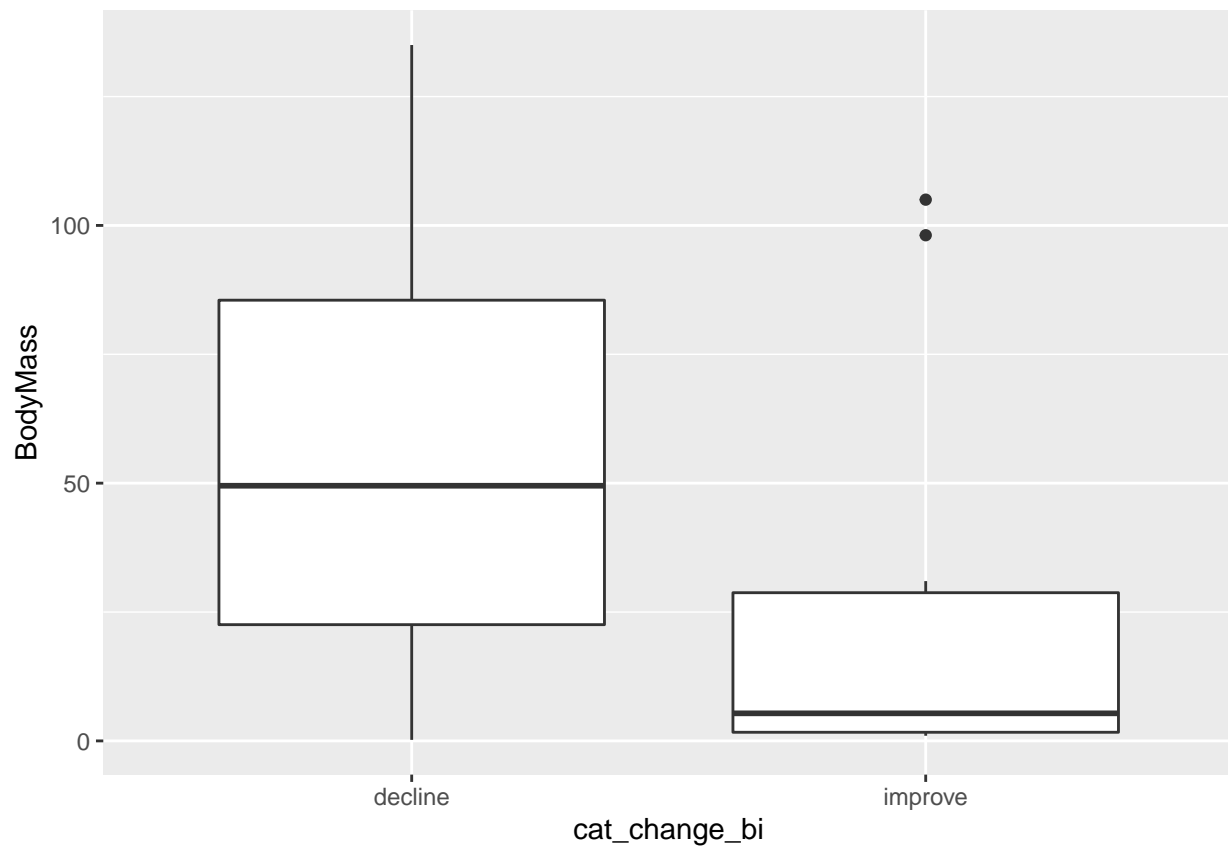
```
ggplot(amphibian, aes(x=cat_change_bi, fill=cat_change_bi)) + geom_bar() + facet_wrap(~education) + theme_minimal()
```



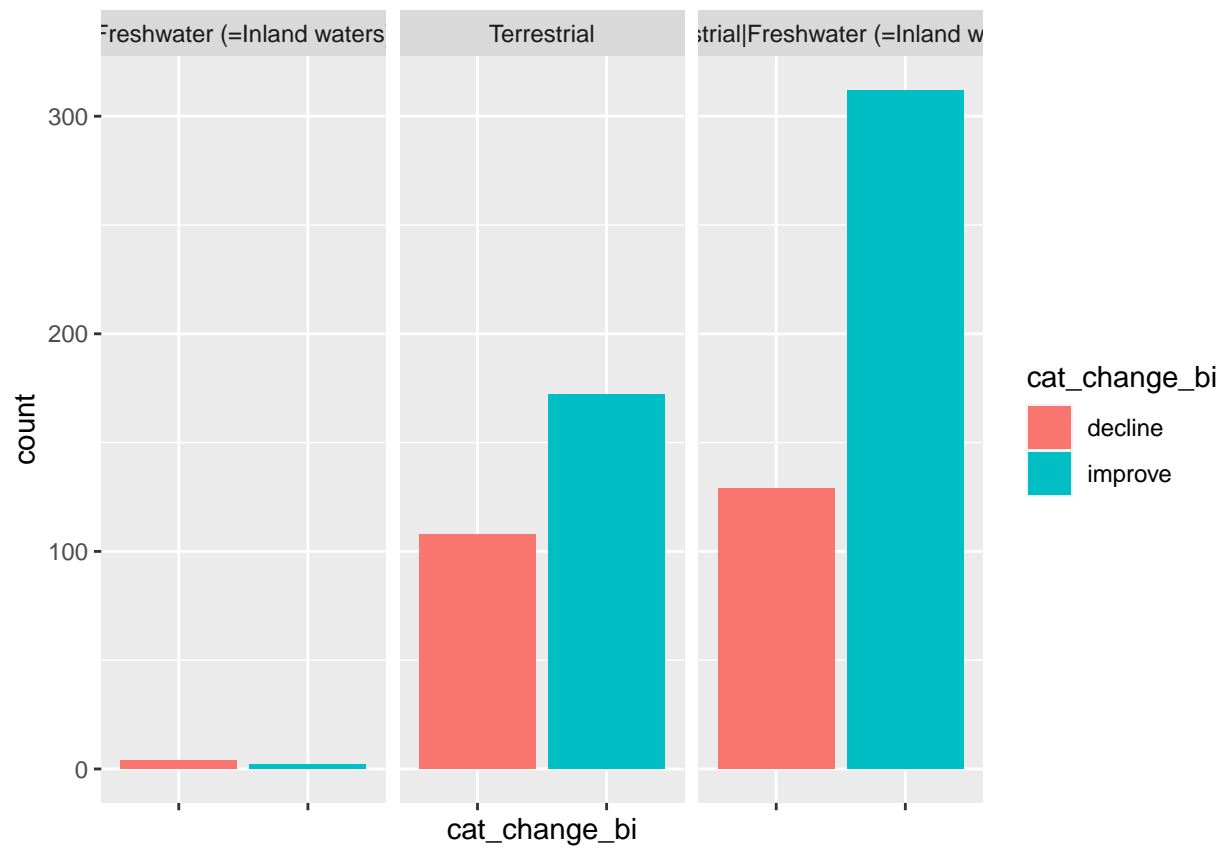
```
# protected area and cat_change_bi
ggplot(amphibian, aes(x=cat_change_bi, fill=cat_change_bi)) + geom_bar() + facet_wrap(~aohPA_prop) + theme_minimal()
```



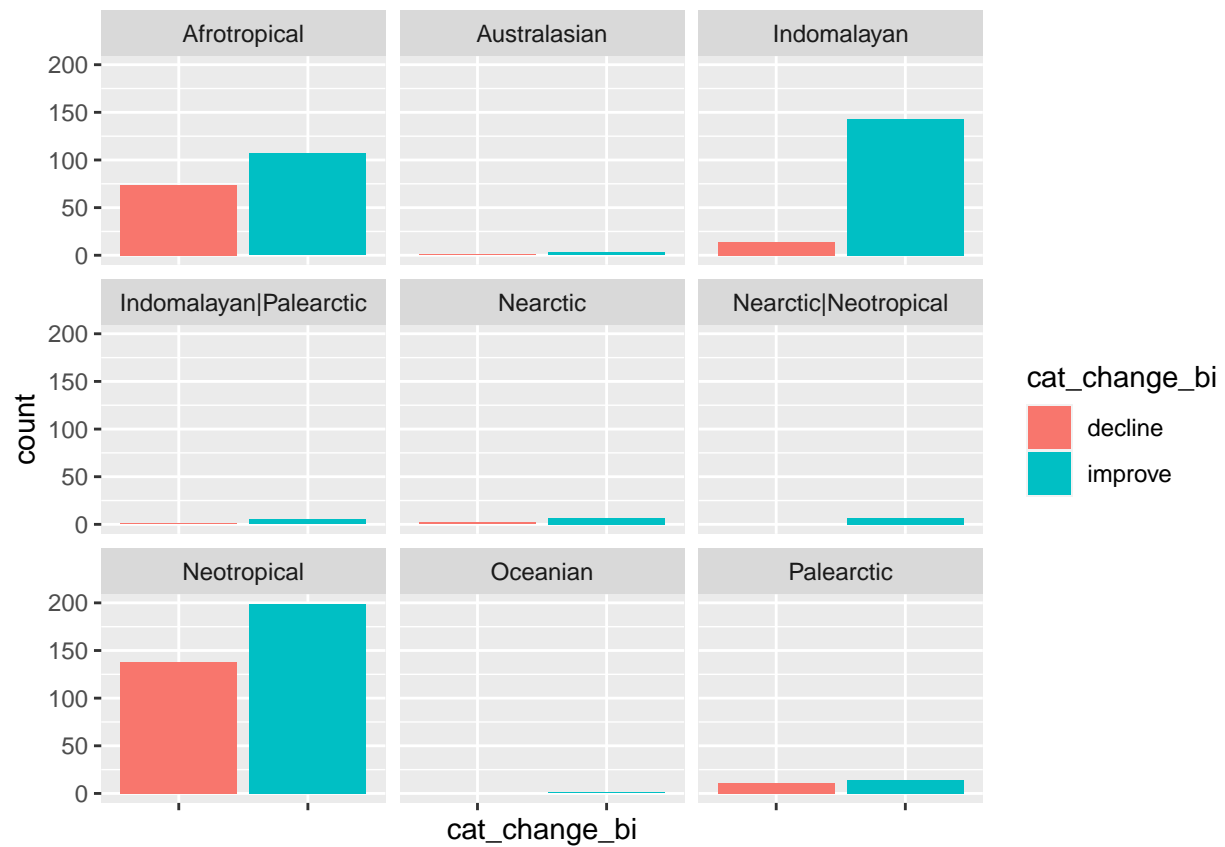
```
# body mass and cat_change_bi  
# improving species have smaller body mass  
ggplot(amphibian, aes(x=cat_change_bi, y=BodyMass)) + geom_boxplot()
```



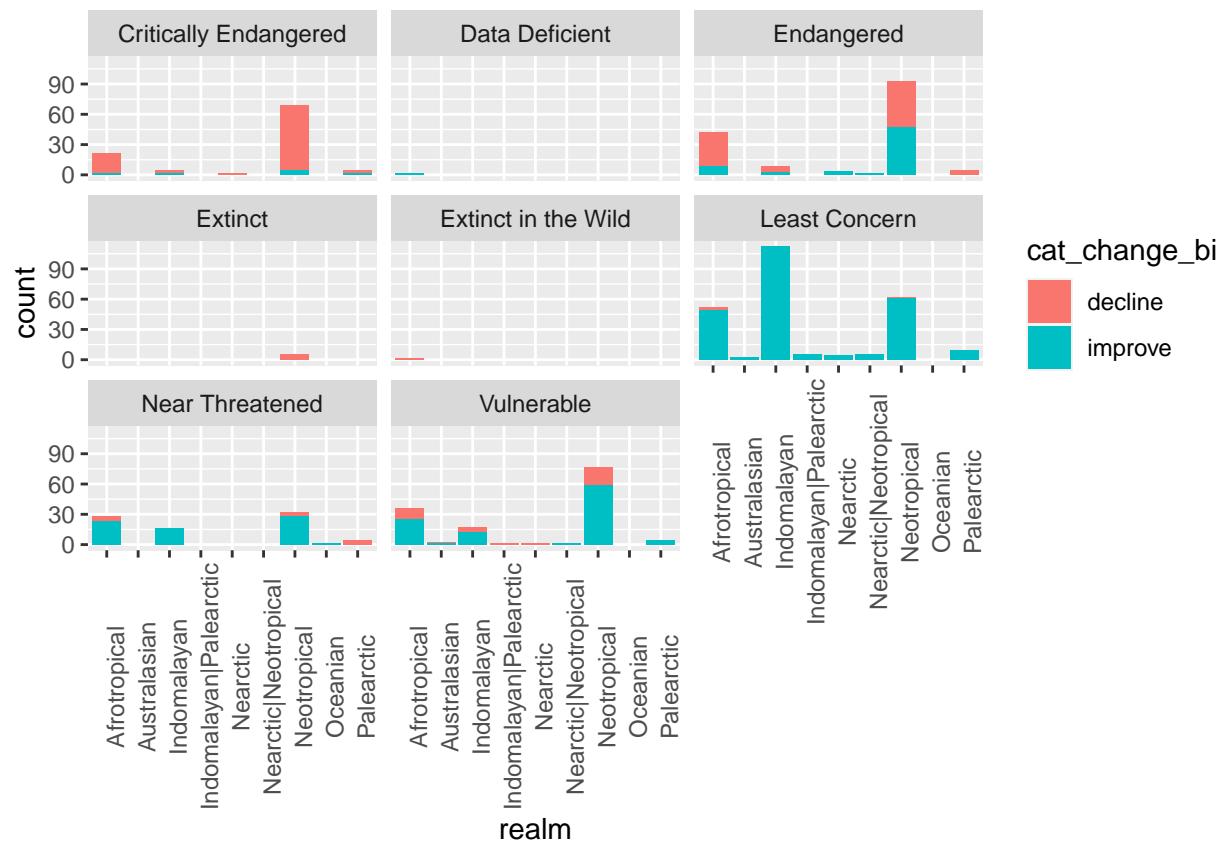
```
# systems
# improving more in terrestrial - associated lands
ggplot(amphibian, aes(x=cat_change_bi, fill=cat_change_bi)) + geom_bar() + facet_wrap(~systems) + theme
```



```
# realm
# doesn't seem to have much of a correlation with realm, although indomalayan notably has a large difference
ggplot(amphibian, aes(x=cat_change_bi, fill=cat_change_bi)) + geom_bar() + facet_wrap(~realm) + theme(as
```



```
# red list category
# critically endangered and endangered are declining
# up listed categories are improving
ggplot(amphibian, aes(x=realm, fill=cat_change_bi)) + geom_bar() + facet_wrap(~redlistCategory) + theme
```

most endangered and critically endangered species are neotropical (which is improving!) and afrotropical
 # largest population in least concern is indomalayan (which are also improving)

Largest factors affecting cat_change_bi seem to be body mass and redlistCategory.