

Untitled

Scott White

April 5, 2019

```
library(here)

## Warning: package 'here' was built under R version 3.5.3
## here() starts at C:/Users/Scott/Dropbox/Masters/STAT 7350 - Visualization of Biological Data/STAT7350
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.2.1 --
## v ggplot2 3.1.0      v purrr   0.3.0
## v tibble   2.0.1      v dplyr    0.8.0.1
## v tidyr    0.8.2      v stringr  1.4.0
## v readr    1.3.1      vforcats  0.4.0

## Warning: package 'tibble' was built under R version 3.5.2
## Warning: package 'readr' was built under R version 3.5.2
## Warning: package 'purrr' was built under R version 3.5.2
## Warning: package 'dplyr' was built under R version 3.5.2
## Warning: package 'stringr' was built under R version 3.5.2
## Warning: package 'forcats' was built under R version 3.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()
```

Read in data

```
bees <- read_csv(here("Assignments/Assignment3/data_output/cleaned_bees_columns_removed.csv"),
                  col_types = cols(Division = col_character()))

bees <- bees %>% mutate(Month = as_factor(Month))
bees <- bees %>% mutate(Month = fct_relevel(Month, levels = c("May", "Jun", "Jul", "Aug", "Sept")))

## Warning: Outer names are only allowed for unnamed scalar atomic inputs
bees

## # A tibble: 6,071 x 12
##       ID Genus Gender Species State County Refuge Division `Sample Site`
##   <dbl> <chr> <chr>  <chr>  <chr>  <chr>  <chr>  <chr>
## 1     1 Agap~ Female serice~ Miss~ Lafay~ Big M~ <NA>   <NA>
## 2     2 Agap~ Female serice~ Miss~ Lafay~ Big M~ <NA>    BB
## 3     3 Agap~ Female serice~ Miss~ Lafay~ Big M~ <NA>    BB
## 4     4 Agap~ Female serice~ Miss~ Ray    Big M~ <NA>    BB
## 5     5 Agap~ Male   serice~ Miss~ Lafay~ Big M~ <NA>    BB
## 6     6 Agap~ Female serice~ Miss~ Ray    Big M~ <NA>  JAB
## 7     7 Agap~ Female serice~ Miss~ Ray    Big M~ <NA>  JAB
```

```

## 8     8 Agap~ Female serice~ Miss~ Ray    Big M~ <NA>      JAB
## 9     9 Agap~ Female serice~ Miss~ Ray    Big M~ <NA>      JAB
## 10   10 Agap~ Male    serice~ Miss~ Ray    Big M~ <NA>      JAB
## # ... with 6,061 more rows, and 3 more variables: Habitat <chr>,
## #   Collector <chr>, Month <fct>

```

```

strong_states <- c("Iowa", "Minnesota", "Missouri")
weak_states <- c("Illinois", "Wisconsin")

```

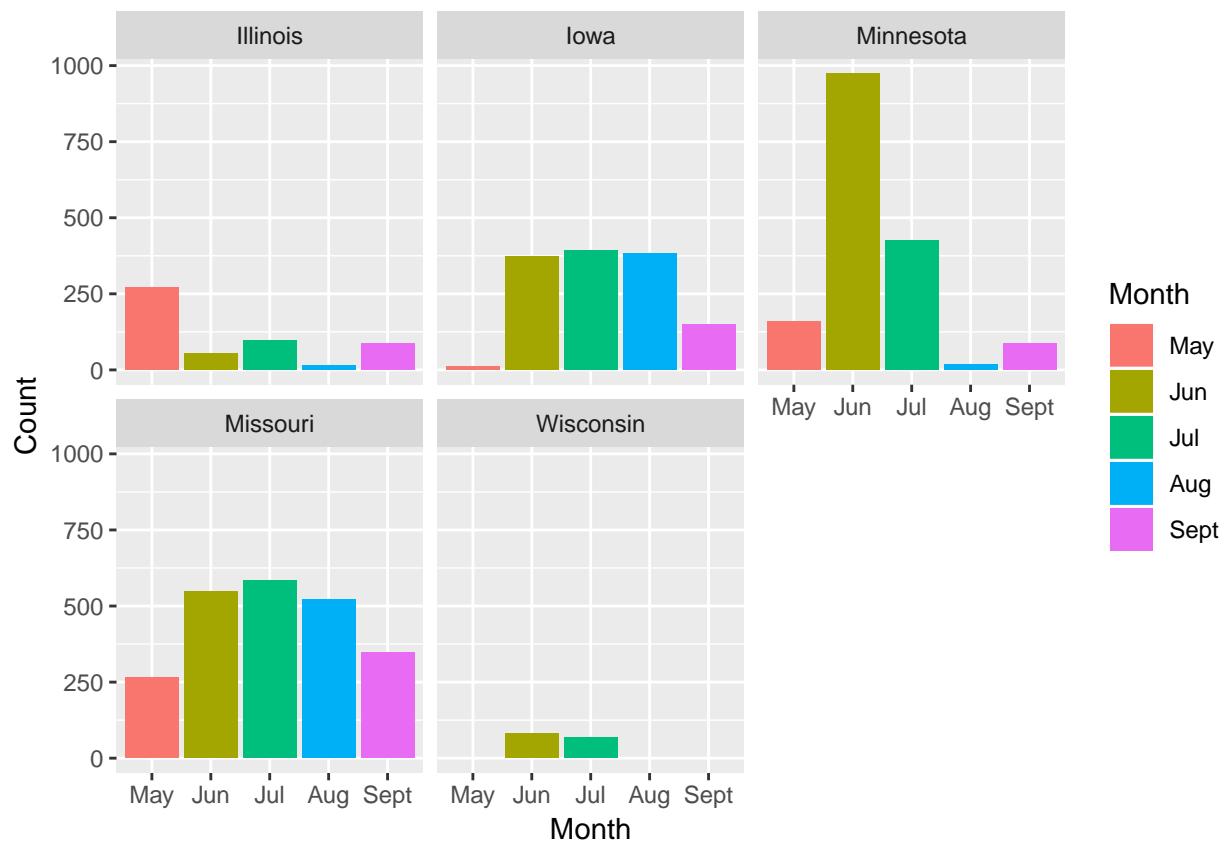
Checking how many monthly counts we have per state

```

state_county_month <- bees %>% group_by(State, Month) %>% summarise(Count = n())

## Warning: Factor `Month` contains implicit NA, consider using
## `forcats::fct_explicit_na`
state_county_month %>% drop_na %>% ggplot(aes(Month, Count)) +
  geom_col(aes(fill = Month), position = "dodge") +
  facet_wrap(.~State)

```



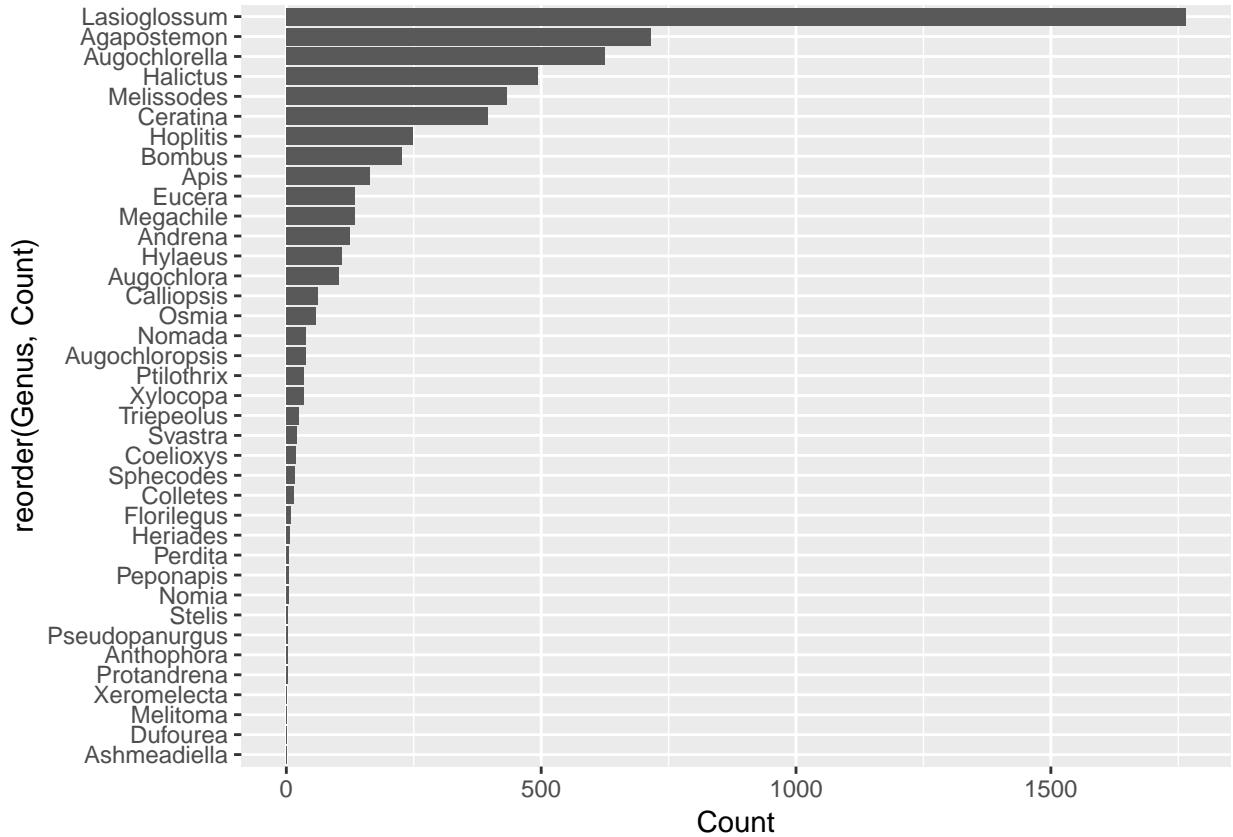
Plot the number of records we have for each genus

```
sort(table(bees$Genus))

##          Ashmeadiella      Dufourea      Melitoma Xeromelecta Protandrena
##                      1                  1                  1                  1                  2
##          Anthophora Pseudopanurgus      Stelis      Nomia Peponapis
##                      3                  3                  3                  4                  4
##          Perdita     Heriades Florilegus     Colletes Sphecodes
##                      4                  7                  8                 14                 16
##          Coelioxys     Svastra Triepeolus     Xylocopa Ptilothrix
##                      18                 21                 25                 34                 35
## Augochloropsis     Nomada      Osmia Calliopsis Augochlora
##                      38                 38                 58                 62                103
##          Hylaeus     Andrena   Megachile     Eucera    Apis
##                      108                125                134                135                163
##          Bombus     Hoplitis   Ceratina Melissodes Halictus
##                      227                248                396                433                494
## Augochlorella Agapostemon Lasioglossum
##                      624                716                1764
```

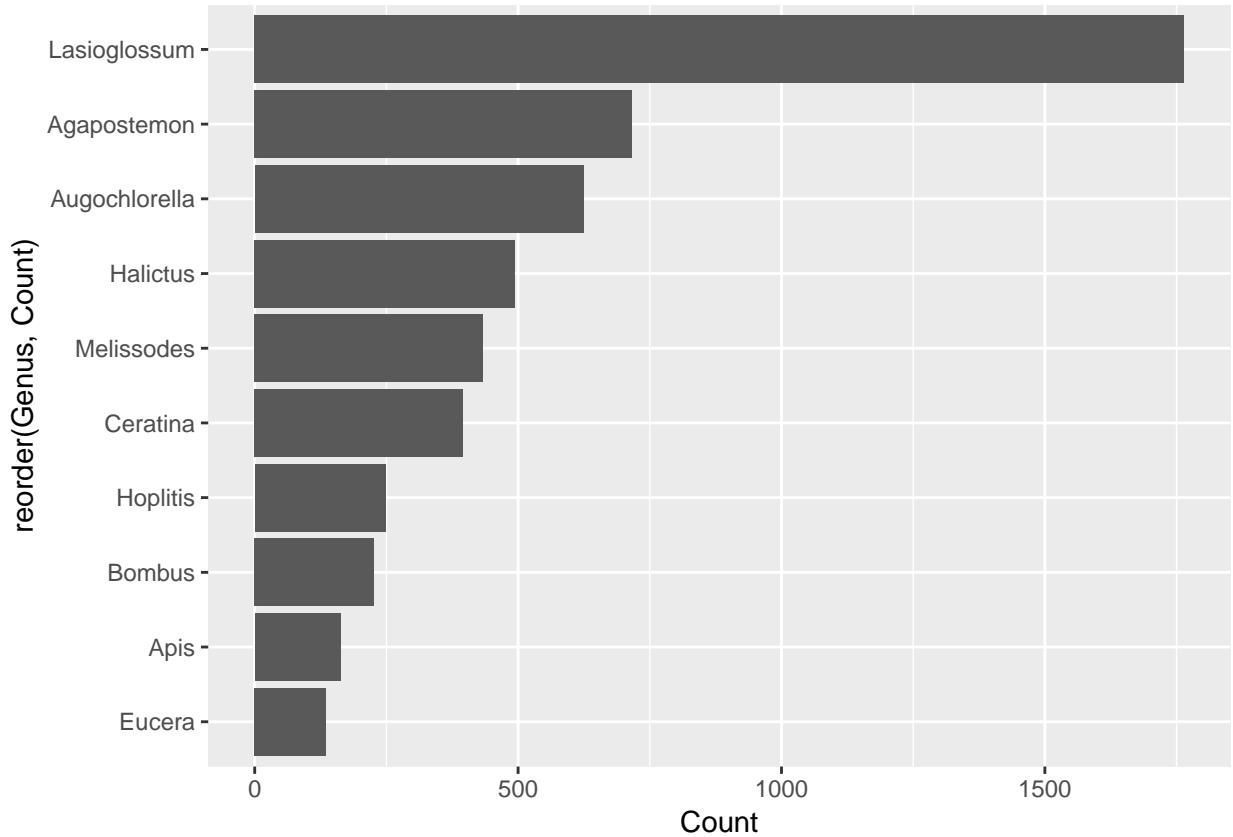
```
genus_count <- bees %>% group_by(Genus) %>% summarise(Count = n())

genus_count %>% ggplot(aes(reorder(Genus, Count), Count)) +
  geom_col() +
  coord_flip()
```



Keep the genus that have more than 100 counts

```
genus_count %>% top_n(10, Count) %>% ggplot(aes(reorder(Genus, Count), Count)) +
  geom_col() +
  coord_flip()
```



Record the names of the top 10 genus counts (in descending order)

```
top10_genus <- genus_count %>% top_n(10, Count) %>% arrange(desc(Count)) %>% select(Genus) %>% .$Genus
```

Plot the number of genus per month

```
month_genus_count <- bees %>% group_by(Month, Genus) %>% summarise(Count = n())
```

```
## Warning: Factor `Month` contains implicit NA, consider using
## `forcats::fct_explicit_na`
month_genus_count
```

```
## # A tibble: 139 x 3
## # Groups:   Month [6]
##   Month Genus      Count
##   <fct> <chr>     <int>
## 1 May   Agapostemon 15
## 2 May   Andrena    49
## 3 May   Apis       23
## 4 May   Augochlora  4
## 5 May   Augochlorella 20
## 6 May   Augochloropsis 1
## 7 May   Bombus     5
```

```

## 8 May Ceratina      84
## 9 May Colletes      1
## 10 May Eucera       12
## # ... with 129 more rows

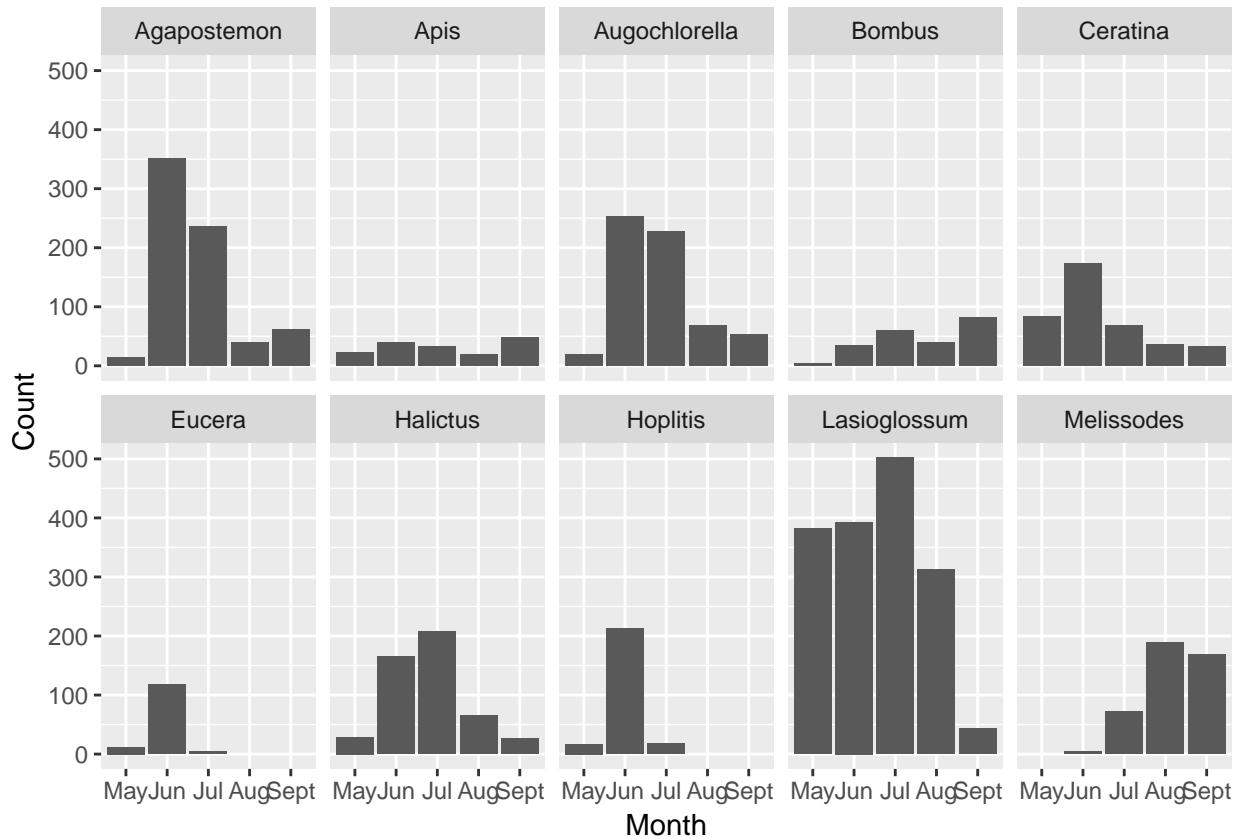
```

When we plot Genus for each month we have to keep in mind we're not counting 130 from Lasioglossum

```

month_genus_count %>% drop_na %>% filter(Genus %in% top10_genus) %>%
  ggplot(aes(Month, Count)) +
  geom_col() +
  facet_wrap(~Genus, nrow = 2)

```



Now lets plot the counts of the top 10 genus by state as well.

```

state_month_genus_count <- bees %>% group_by(State, Month, Genus) %>% summarise(Count = n())

## Warning: Factor `Month` contains implicit NA, consider using
## `forcats::fct_explicit_na`
state_month_genus_count %>% filter(is.na(Month))

## Warning: Factor `Month` contains implicit NA, consider using
## `forcats::fct_explicit_na`

## # A tibble: 8 x 4
## # Groups:   State, Month [1]
##   State   Month Genus      Count

```

```

##   <chr>     <fct> <chr>      <int>
## 1 Illinois <NA> Agapostemon    12
## 2 Illinois <NA> Augochlorella    2
## 3 Illinois <NA> Augochloropsis   7
## 4 Illinois <NA> Bombus       5
## 5 Illinois <NA> Halictus      1
## 6 Illinois <NA> Lasioglossum  130
## 7 Illinois <NA> Megachile      1
## 8 Illinois <NA> Peponapis     1

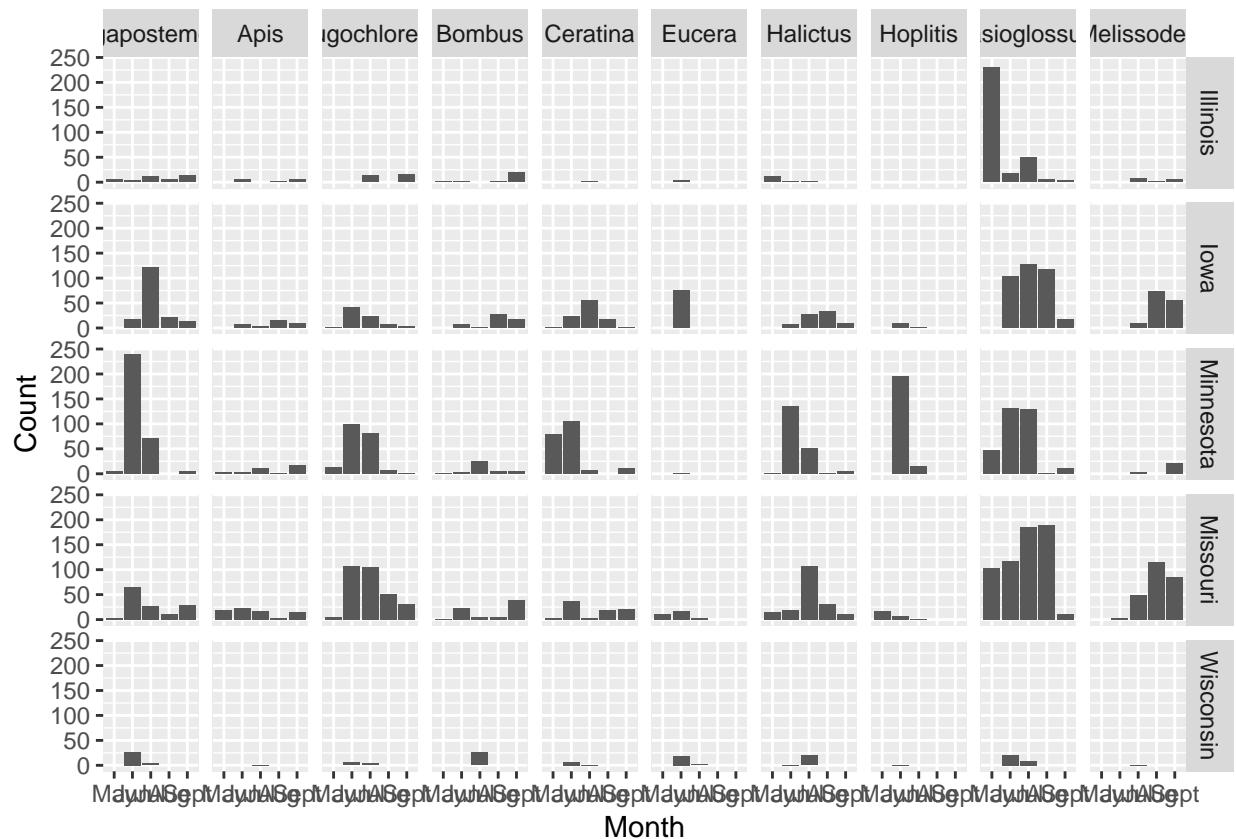
```

A side benefit of comparing tables month_genus_count and state_month_genus_count is that I found out all of the missing Month rows are from Illinois

```

state_month_genus_count %>% drop_na %>% filter(Generus %in% top10_genus) %>%
  ggplot(aes(Month, Count)) +
  geom_col() +
  facet_grid(State~Generus)

```



Investigating the variety of bee species per geographical region

```
bees
```

```

## # A tibble: 6,071 x 12
##   ID Genus Gender Species State County Refuge Division `Sample Site`
##   <dbl> <chr> <chr> <chr> <chr> <chr> <chr> <chr>
## 1 1 Agap~ Female serice~ Miss~ Lafay~ Big M~ <NA>   <NA>
## 2 2 Agap~ Female serice~ Miss~ Lafay~ Big M~ <NA>   BB
## 3 3 Agap~ Female serice~ Miss~ Lafay~ Big M~ <NA>   BB
## 4 4 Agap~ Female serice~ Miss~ Ray    Big M~ <NA>   BB
## 5 5 Agap~ Male   serice~ Miss~ Lafay~ Big M~ <NA>   BB
## 6 6 Agap~ Female serice~ Miss~ Ray    Big M~ <NA>   JAB
## 7 7 Agap~ Female serice~ Miss~ Ray    Big M~ <NA>   JAB
## 8 8 Agap~ Female serice~ Miss~ Ray    Big M~ <NA>   JAB
## 9 9 Agap~ Female serice~ Miss~ Ray    Big M~ <NA>   JAB
## 10 10 Agap~ Male   serice~ Miss~ Ray   Big M~ <NA>   JAB
## # ... with 6,061 more rows, and 3 more variables: Habitat <chr>,
## #   Collector <chr>, Month <fct>
genus_species_state_habitat <- bees %>% group_by(Genus, Species, State, Habitat) %>%
  summarise(Count = n())
genus_species_state_habitat

## # A tibble: 754 x 5
## # Groups:   Genus, Species, State [377]
##   Genus      Species State   Habitat       Count
##   <chr>      <chr>   <chr> <chr>       <int>
## 1 Agapostemon sericeus Illinois <NA>          1
## 2 Agapostemon sericeus Iowa   <NA>          1
## 3 Agapostemon sericeus Iowa   Brome          2
## 4 Agapostemon sericeus Iowa   Savanna         3
## 5 Agapostemon sericeus Minnesota Drawn-Down Wetland Basin 3
## 6 Agapostemon sericeus Missouri <NA>          2
## 7 Agapostemon sericeus Missouri Field          27
## 8 Agapostemon sericeus Missouri Forest          7
## 9 Agapostemon sericeus Missouri Loess Hills     2
## 10 Agapostemon sericeus Missouri Mesic Prairie  1
## # ... with 744 more rows
genus_species_contingency_table <- table(bees$Genus, bees$Species)
str(genus_species_contingency_table)

##  'table' int [1:38, 1:197] 0 0 1 0 0 0 0 0 0 0 ...
##  - attr(*, "dimnames")=List of 2
##    ..$ : chr [1:38] "Agapostemon" "Andrena" "Anthophora" "Apis" ...
##    ..$ : chr [1:197] "abrupta" "addenda" "affinis" "agilis" ...

```

This table gives the number of records of each genus

```
apply(genus_species_contingency_table, 1, sum)
```

##	Agapostemon	Andrena	Anthophora	Apis	Ashmeadiella
##	716	125	3	163	1
##	Augochlora	Augochlorella	Augochloropsis	Bombus	Calliopsis
##	103	624	38	227	62
##	Ceratina	Coelioxys	Colletes	Dufourea	Eucera

##	396	18	14	1	135
##	Florilegus	Halictus	Heriades	Hoplitis	Hylaeus
##	8	494	7	248	108
##	Lasioglossum	Megachile	Melissodes	Melitoma	Nomada
##	1764	134	433	1	38
##	Nomia	Osmia	Peponapis	Perdita	Protandrena
##	4	58	4	4	2
##	Pseudopanurgus	Ptilothrix	Sphecodes	Stelis	Svastra
##	3	35	16	3	21
##	Triepeolus	Xeromelecta	Xylocopa		
##	25	1	34		

This table gives the number of different species in each genus

```
apply(genus_species_contingency_table, 1, function(x){sum(x != 0)})
```

##	Agapostemon	Andrena	Anthophora	Apis	Ashmeadiella
##	4	27	2	1	1
##	Augochlora	Augochlorella	Augochloropsis	Bombus	Calliopsis
##	1	2	3	12	1
##	Ceratina	Coelioxys	Colletes	Dufourea	Eucera
##	6	5	7	1	2
##	Florilegus	Halictus	Heriades	Hoplitis	Hylaeus
##	1	4	4	3	7
##	Lasioglossum	Megachile	Melissodes	Melitoma	Nomada
##	44	12	12	1	12
##	Nomia	Osmia	Peponapis	Perdita	Protandrena
##	2	10	1	2	1
##	Pseudopanurgus	Ptilothrix	Sphecodes	Stelis	Svastra
##	3	1	5	1	2
##	Triepeolus	Xeromelecta	Xylocopa		
##	2	1	1		

This table gives the number of records of each species

```
apply(genus_species_contingency_table, 2, sum)
```

##	abrupta	addenda	affinis
##	1	3	7
##	agilis	albipenne	albitarsis
##	34	41	1
##	americanus	andreniformis	andrenoides
##	1	62	1
##	anomalum	articulata	atripes
##	6	2	3
##	aurata	auricomus	bancrofti
##	542	16	2
##	barbara	barbilabris	bimaculatus
##	1	1	156
##	boltoniae	bombiformis	borealis
##	1	35	2
##	brevis	brunneri	bucconis
##	64	8	1
##	bucephala	calcarata	callidum

##		1	13	1
##	campanulae		carinatus	carlini
##		1		7
##	cinctipes		citrinus	coeruleum
##		4		1
##	collinsiae		coloradensis	commoda
##		1		1
##	communis		comptoides	condignus
##		25		8
##	confusus		conjuncta	coreopsis
##		190		10
##	coriaceum		crataegi	cressoni group
##		75		10
##	cressonii		davisii	desponsa
##		25		26
##	disparile		distincta	dupla
##		15		5
##	ellisiae		enlophi	erigeniae
##		84		5
##	eulophi		fedorense	fervidus
##		1		8
##	forbesii		fraternus	fulgida
##		2		12
##	fuscipenne		gemula	georgica
##		1		1
##	geranii		germana	griseocollis
##		14		55
##	hamata		hartii	helianthi
##		134		2
##	helianthiformis		heterognathum	hippotes
##		1		1
##	hudsoniellum		illinoiensis	imitatrix
##		2		6
##	imitatum		impatiens	inaequalis
##		75		1
##	inermis		integra	interrupta
##		1		1
##	inurbana		kincaidii	labiata
##		20		3
##	labrosus		latimanus leavitii/variolosa	
##		1	10	2
##	leavitti		leucozonium	ligatus
##		3		252
##	lignaria		lustrans	macoupinense
##		2		1
##	maculigera		marginata	mellifera
##		1		163
##	mendica		mesillae	mesillae group
##		19		13
##	metallica		mikmaqi	miranda
##		4		2
##	miserabilis		mitchelli	modestus
##		5		1
##	modestus group		montivaga	nasonii

##		62		17		16
##	nelumbo	nis	nigrae		norton	
##		13		3		3
##	nymphaea	rum	obliqua		obscurum	
##		5		18		2
##	octodentata		octomaculata		paraforbesii	
##		3		3		63
##	parallelus		pectorale		pensylvanicus	
##		41		21		22
##	perplexa		perpunctatum		persimilis	
##		1		4		82
##	personata		petulans		pictum	
##		4		10		6
##	pilosifrons		pilosum		porterae	
##		220		244		1
##	producta		proxima		pruinosa	
##		25		1		4
##	pruinosum		pumila		pura	
##		55		24		103
##	quebecense		quintilus		relativa	
##		3		2		1
##	robertsoni		rosae		rubicundus	
##		1		1		11
##	rudbeckia		rudbeckiae		rufitarsis	
##		5		3		1
##	rufocinctus		rustica		sayi	
##		2		1		8
##	semicaeruleum		sericeus		sigmundi	
##		21		49		8
##	simillima		simulans		sp.	
##		6		2		100
##	sp. (Gnathias)		sp. 1		sp. 2	
##		4		32		8
##	sp. 3		sp. 4		sp. 5	
##		6		1		2
##	sp. 6		sp. A		sp. B	
##		1		2		2
##	sp. C		splendens		spoliata	
##		3		44		3
##	spp.		strenua		subillata	
##		115		10		2
##	subviridatum		succinipenne		sumptuosa	
##		1		66		22
##	swenki		taurea		teglellisiae	
##		137		1		1
##	tegulare		tepaneca		terminalis	
##		137		4		2
##	ternarius		texana		texanus	
##		5		2		135
##	triangulifera		trigeminum		trinodis	
##		1		1		73
##	truncatum		vagans		variolosa	
##		6		8		1
##	versatum		vierecki		violae	

```

##          14          48          1
##    virescens    virginiana    virginica
##          488          1          34
##   wilkella    wilmattae    zephyrum
##          5          7          11
##   ziziae      zonulum
##          5         154

```

This table gives the number of different genus each species belongs to

```
sort(apply(genus_species_contingency_table, 2, function(x){sum(x != 0)}))
```

```

##      abrupta      addenda      affinis
##          1          1          1
##      agilis      albipenne    albitarsis
##          1          1          1
## americanus andreniformis andrenoides
##          1          1          1
## anomalam articulata     atripes
##          1          1          1
## aurata      auricomus    bancrofti
##          1          1          1
## barbara     barbilabris boltoniae
##          1          1          1
## bombiformis borealis     brevis
##          1          1          1
## bruneri      buconis     bucephala
##          1          1          1
## calcarata     callidum campanulae
##          1          1          1
## carinatus      carlini cinctipes
##          1          1          1
## citrinus      coeruleum collinsiae
##          1          1          1
## coloradensis      commoda communis
##          1          1          1
## comptoides      condignus confusus
##          1          1          1
## conjuncta      coriaceum crataegi
##          1          1          1
## cressoni group      davisii desponsa
##          1          1          1
## disparile      distincta dupla
##          1          1          1
## ellisiae      enlophi erigeniae
##          1          1          1
## eulophi      fedorense fervidus
##          1          1          1
## forbesii      fraternus fulgida
##          1          1          1
## fuscipenne      gemula georgica
##          1          1          1
## geranii      germana griseocollis
##          1          1          1

```

##	hamata	hartii	helianthi
##	1	1	1
##	helianthiformis	heterognathum	hippotes
##	1	1	1
##	hudsoniellum	illinoiensis	imitatrix
##	1	1	1
##	imitatum	impatiens	inaequalis
##	1	1	1
##	inermis	integra	interrupta
##	1	1	1
##	inurbana	kincaidii	labiata
##	1	1	1
##	labrosus	latimanus	leavitii/variolosa
##	1	1	1
##	leavitti	leucozonium	ligatus
##	1	1	1
##	lignaria	lustrans	macoupinense
##	1	1	1
##	maculigera	marginata	mellifera
##	1	1	1
##	mendica	mesillae	mesillae group
##	1	1	1
##	metallica	mikmaqi	miranda
##	1	1	1
##	miserabilis	mitchelli	modestus
##	1	1	1
##	modestus group	montivaga	nasonii
##	1	1	1
##	nigrae	norton	nymphaearum
##	1	1	1
##	obliqua	obscurum	octodentata
##	1	1	1
##	octomaculata	paraforbesii	parallelus
##	1	1	1
##	pectorale	pensylvanicus	perplexa
##	1	1	1
##	perpunctatum	persimilis	personata
##	1	1	1
##	petulans	pictum	pilosifrons
##	1	1	1
##	pilosum	porterae	producta
##	1	1	1
##	proxima	pruinosa	pruinosum
##	1	1	1
##	pumila	pura	quebecense
##	1	1	1
##	quintilus	relativa	robertsoni
##	1	1	1
##	rosae	rubicundus	rudbeckia
##	1	1	1
##	rudbeckiae	rufitarsis	rufocinctus
##	1	1	1
##	rustica	sayi	semicaeruleum
##	1	1	1

```

##      sericeus      sigmundi      simillima
##      1             1             1
##      simulans     sp. (Gnathias)    sp. 4
##      1             1             1
##      sp. 5         sp. 6         sp. A
##      1             1             1
##      sp. B         sp. C         splendens
##      1             1             1
##      spoliata     spp.          strenua
##      1             1             1
##      subillata     subviridatum  succinipenne
##      1             1             1
##      sumptuosa     swenki        taurea
##      1             1             1
##      teglellisiae  tegulare      tepaneca
##      1             1             1
##      terminalis    ternarius    texana
##      1             1             1
##      texanus       triangulifera trigeminum
##      1             1             1
##      trinodis      truncatum    vagans
##      1             1             1
##      variolosa     versatum     vierecki
##      1             1             1
##      violae        virescens   virginiana
##      1             1             1
##      virginica     wilkella    wilmattae
##      1             1             1
##      zephyrum      ziziae       zonulum
##      1             1             1
##      bimaculatus   coreopsis   cressonii
##      2             2             2
##      nelumbonis    sp.          sp. 3
##      2             2             2
##      sp. 1         sp. 2
##      3             3

```

Create a histogram of the number of counts of records for each species type

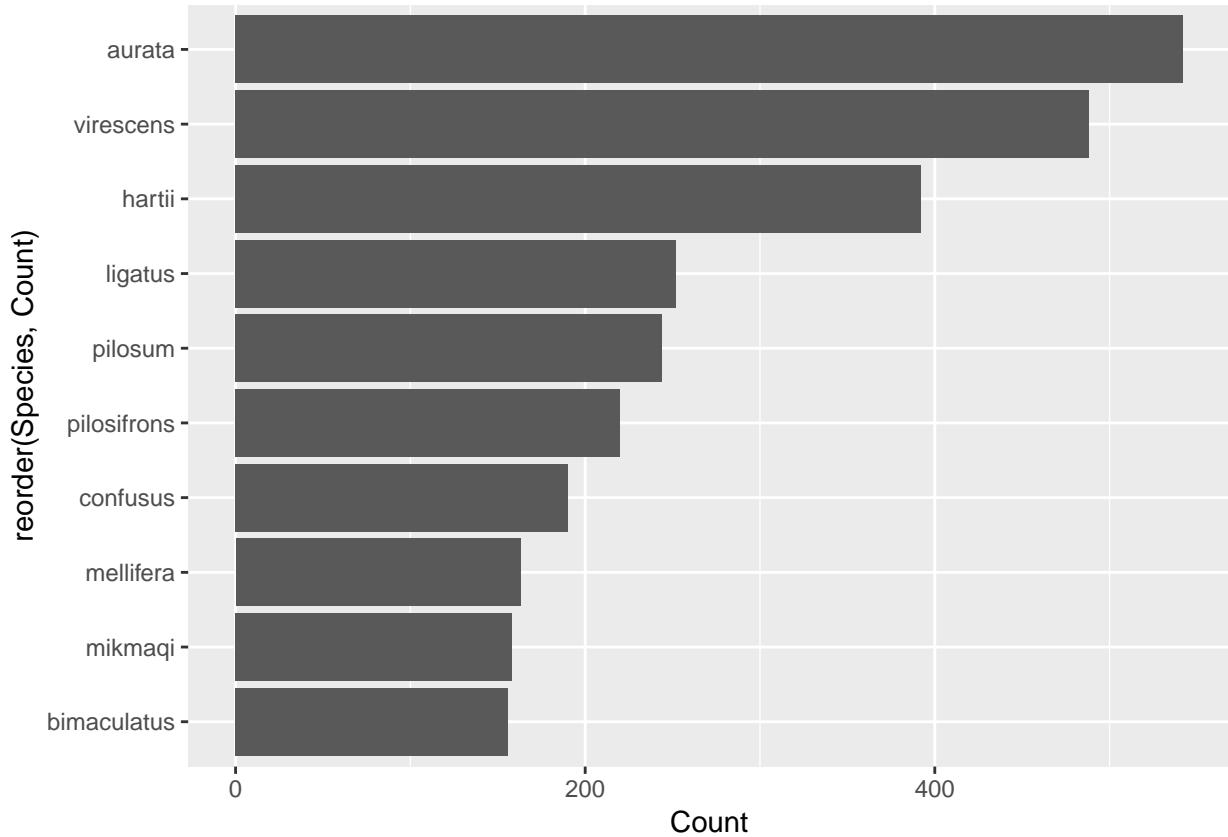
```

species_counts <- bees %>% group_by(Species) %>% summarise(Count = n()) %>% arrange(desc(Count))

top10_species <- species_counts %>% top_n(10, Count) %>% .\$Species

species_counts %>% filter(Species %in% top10_species) %>% ggplot(aes(reorder(Species, Count), Count)) +
  geom_col() + coord_flip()

```



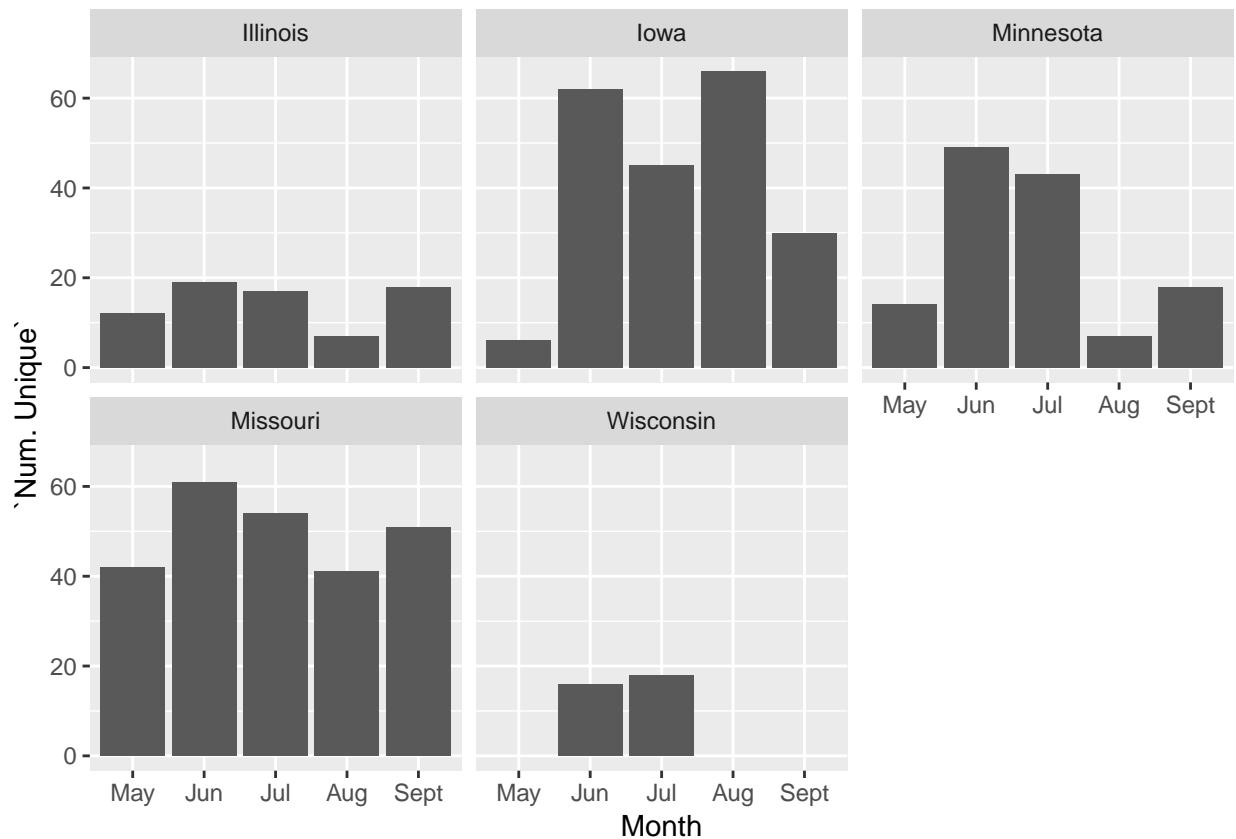
Graph the number of species in each region by month

```
state_month_genus_species <- bees %>% group_by(State, Month, Genus, Species) %>%
  summarise(Count = n(),
            `Num. Unique` = n_distinct(Species))

## Warning: Factor `Month` contains implicit NA, consider using
## `forcats::fct_explicit_na`
state_month_genus_species

## # A tibble: 710 x 6
## # Groups:   State, Month, Genus [316]
##   State    Month Genus      Species     Count `Num. Unique`
##   <chr>   <fct> <chr>      <chr>     <int>       <int>
## 1 Illinois May Agapostemon texanus      5           1
## 2 Illinois May Augochloropsis sumptuosa   1           1
## 3 Illinois May Bombus        fervidus     1           1
## 4 Illinois May Bombus        impatiens    1           1
## 5 Illinois May Halictus       parallelus   12          1
## 6 Illinois May Lasioglossum fedorense    11          1
## 7 Illinois May Lasioglossum nymphaeum    2           1
## 8 Illinois May Lasioglossum pilosum     127         1
## 9 Illinois May Lasioglossum swenki      81          1
## 10 Illinois May Lasioglossum vierecki   9           1
## # ... with 700 more rows
```

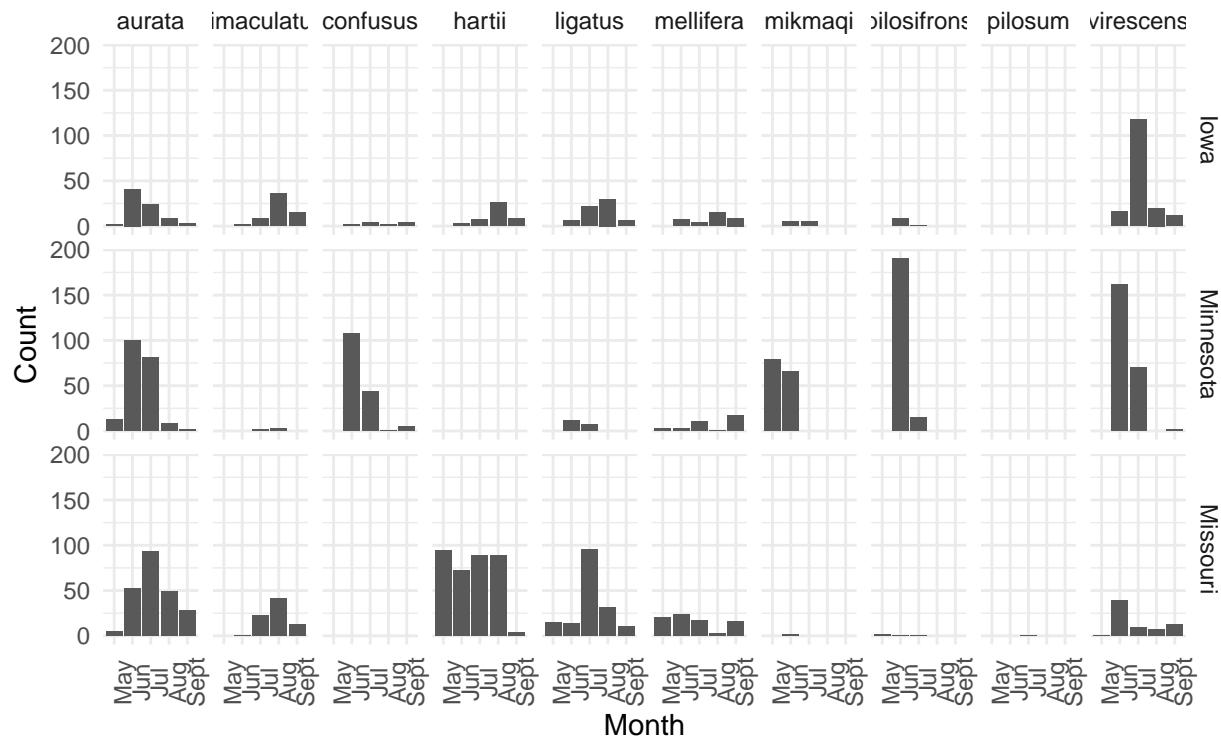
```
state_month_genus_species %>% drop_na() %>% ggplot(aes(Month, `Num. Unique`)) +
  geom_col() +
  facet_wrap(~State)
```



```
fig_state_species_month_strong <- state_month_genus_species %>% filter(Species %in% top10_species) %>% ggplot(aes(Month, Count)) +
  geom_col() +
  facet_grid(State~Species) +
  theme_minimal() +
  ggtitle("Investigating the top ten most recorded species",
          subtitle = "Across the most prevalent states") +
  theme(axis.text.x = element_text(angle = 90))
```

```
fig_state_species_month_strong
```

Investigating the top ten most recorded species Across the most prevalent states



Frequency of counts of the different types of habitats

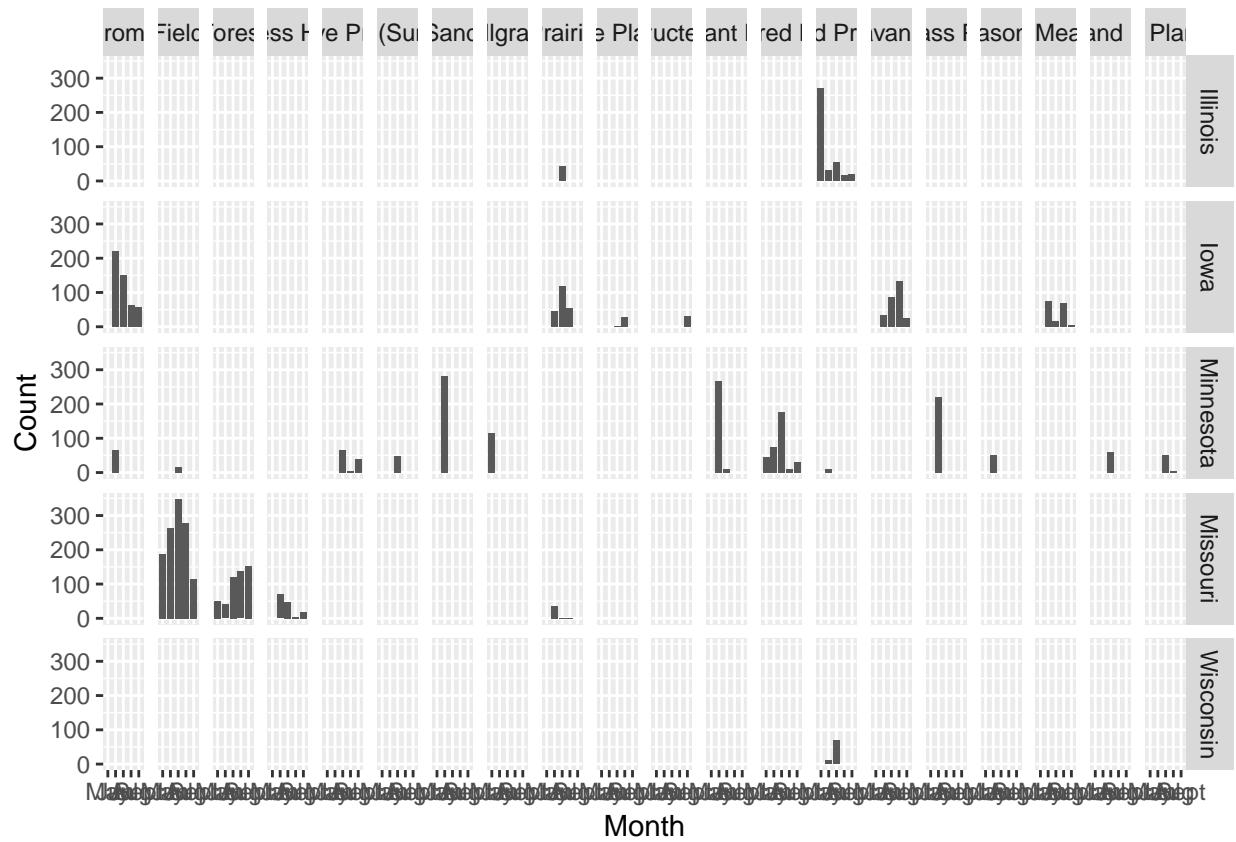
```
habitat_counts <- bees %>% group_by(Habitat) %>% summarise(Count = n()) %>% arrange(desc(Count))
habitat_counts

## # A tibble: 30 x 2
##   Habitat      Count
##   <chr>     <int>
## 1 Field        1209
## 2 Sand Prairie    636
## 3 <NA>         621
## 4 Brome         559
## 5 Forest         500
## 6 Restored Prairie 337
## 7 Prairie        298
## 8 Native Sand Prairie 281
## 9 Savanna        278
## 10 Remnant Prairie 277
## # ... with 20 more rows
top_habitats <- habitat_counts %>% filter(Count >= 30) %>% .$Habitat
```

Lets take a look at how the types and frequency of habitats has changed over time.

```
bees %>% group_by(Month, State, Habitat) %>% summarise(Count=n()) %>% drop_na %>%
  filter(Habitat %in% top_habitats) %>%
  ggplot(aes(Month, Count)) +
  geom_col() +
  facet_grid(State~Habitat)
```

```
## Warning: Factor `Month` contains implicit NA, consider using
## `forcats::fct_explicit_na`
```



State, Habitat, Month, Number of species found in that month

```
bees %>% drop_na() %>% group_by(State, Habitat, Month) %>% summarise(Count = n()) %>% drop_na() %>%
  filter(Habitat %in% top_habitats) %>%
  ggplot(aes(Month, Count)) +
  geom_col() +
  facet_grid(State~Habitat)
```

