

7350 Assignment 3 Version 2

Jingyu Wang 7701969

April 8, 2019

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.2.1 --
## v ggplot2 3.1.0      v purrr  0.2.5
## v tibble  2.0.1      v dplyr  0.8.0.1
## v tidyr   0.8.2      v stringr 1.3.1
## v readr   1.3.1      v forcats 0.3.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(MASS)
```

```
##
## Attaching package: 'MASS'

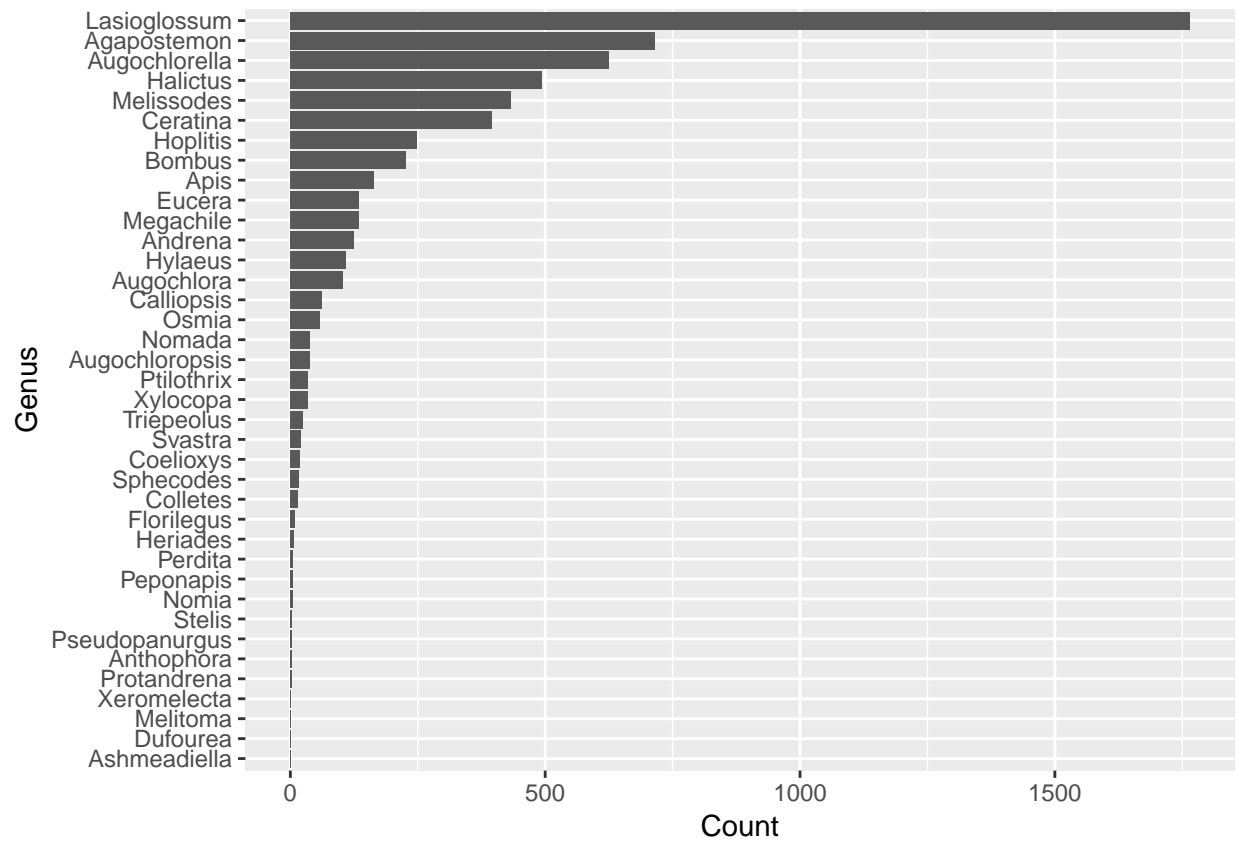
## The following object is masked from 'package:dplyr':
##
##      select
```

```
bees <- read_csv("E:/dropbox/Dropbox/Statistics/graduate course/stat 7350 biostats/Assignments/Assignment3/MASS.csv")
```

```
## Parsed with column specification:
## cols(
##   ID = col_double(),
##   Genus = col_character(),
##   Gender = col_character(),
##   Species = col_character(),
##   State = col_character(),
##   County = col_character(),
##   Refuge = col_character(),
##   Division = col_logical(),
##   `Sample Site` = col_character(),
##   Habitat = col_character(),
##   Collector = col_character(),
##   Month = col_character()
## )
```

```
bees.rare.spe <- bees %>% filter(Species %in% c("sumptuosa", "pumila", "obliqua", "nasonii", "montivaga", "mesilla"))
```

```
bees %>% group_by(Genus) %>% summarise(Count = n()) %>%
  ggplot(aes(x=reorder(Genus, Count), y=Count)) +
  geom_col() +
  coord_flip() +
  labs(x = "Genus")
```

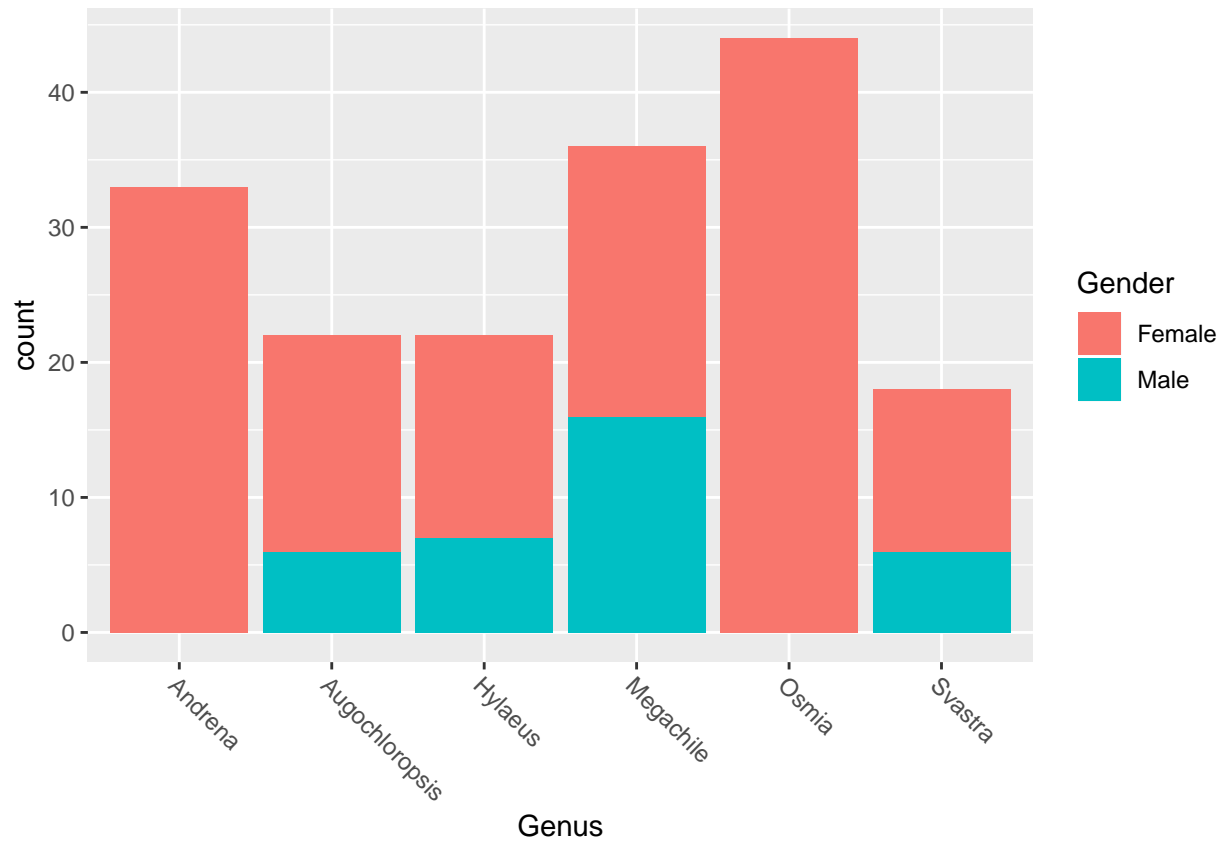


```
genus_count<-bees%>%count(Genus)%>%
  filter(n<=200)

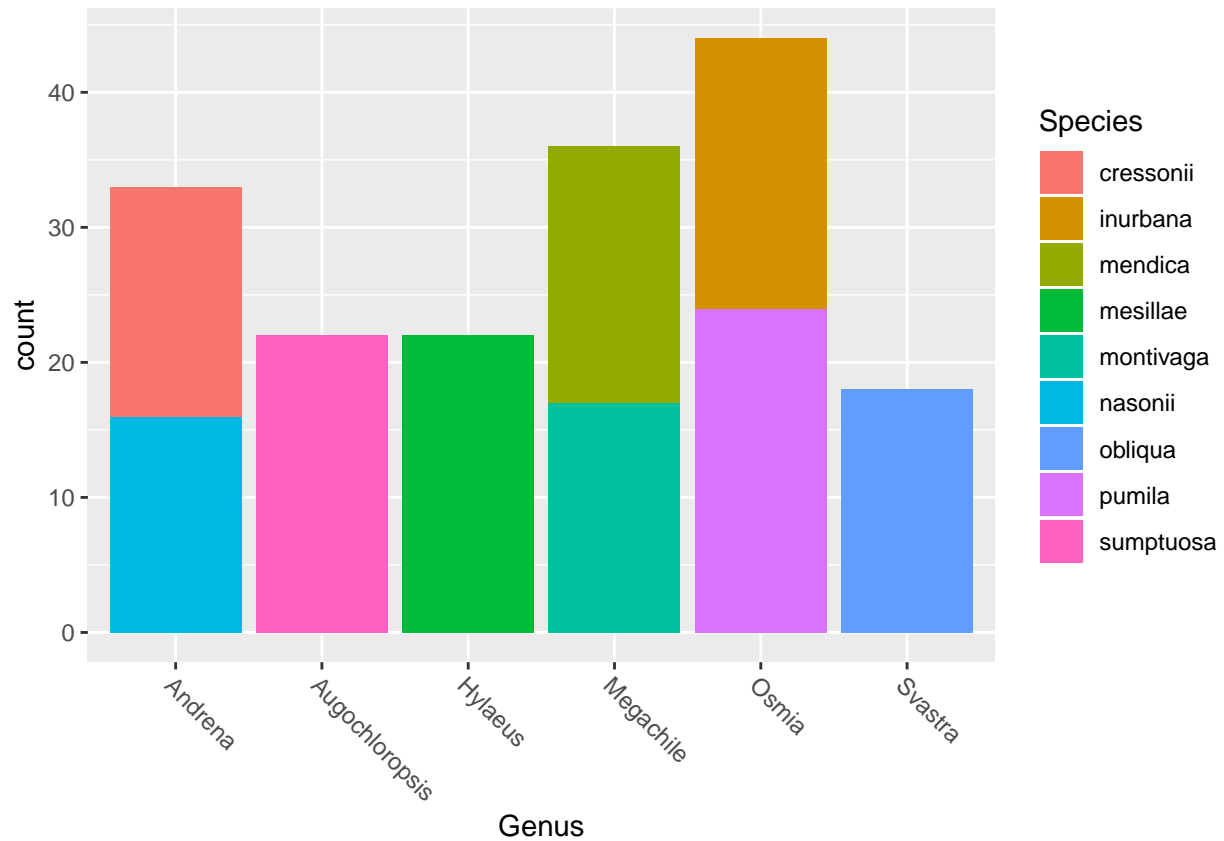
species_counts <- bees%>%
  count(Species)%>%
  filter(n>=15)%>%filter(n<=30)

bees_new<-bees%>%
  filter(Species%in%species_counts$Species)%>%
  filter(Genus%in%genus_count$Genus)

ggplot(data= bees_new, mapping= aes(x=Genus, fill = Gender)) +
  geom_bar()+
  theme(axis.text.x=element_text(angle = -45, hjust = 0))
```

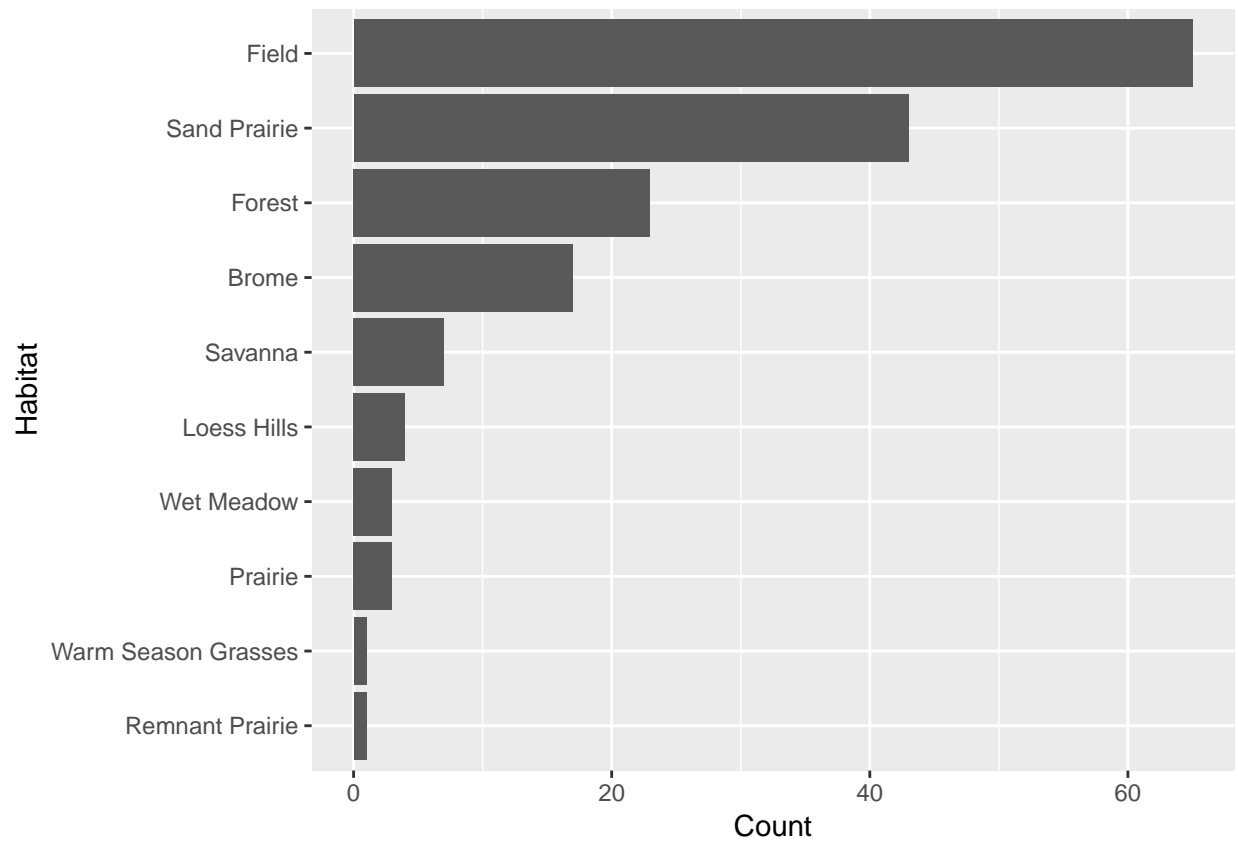


```
ggplot(data= bees_new, mapping= aes(x=Genus, fill = Species)) +  
  geom_bar() +  
  theme(axis.text.x=element_text(angle = -45, hjust = 0))
```



To find the specific habitat for those different species we study

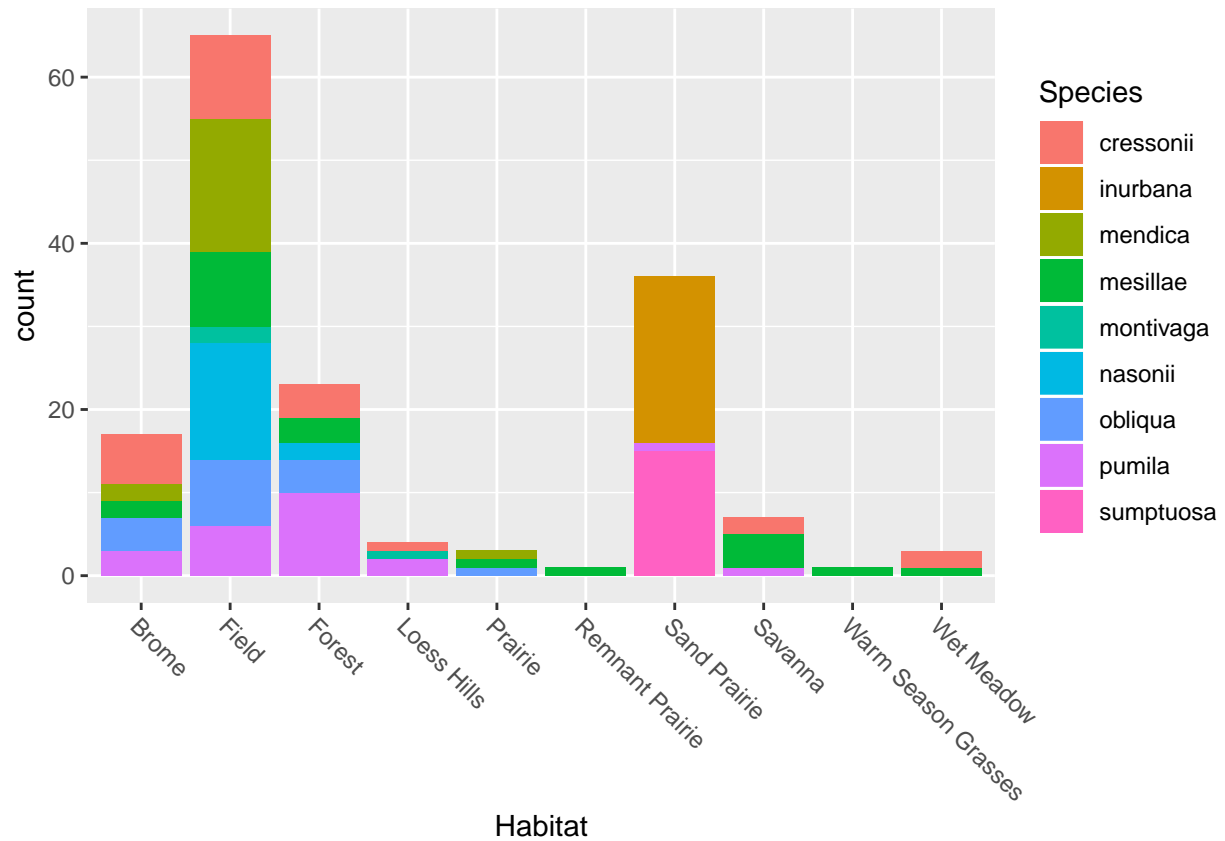
```
bees.rare.spe%>%group_by(Habitat)%>%
  summarise(Count = n())%>%drop_na()%>%
  ggplot(aes(x= reorder(Habitat,Count),y=Count))+
  geom_col()+
  coord_flip()+
  labs(x="Habitat")
```



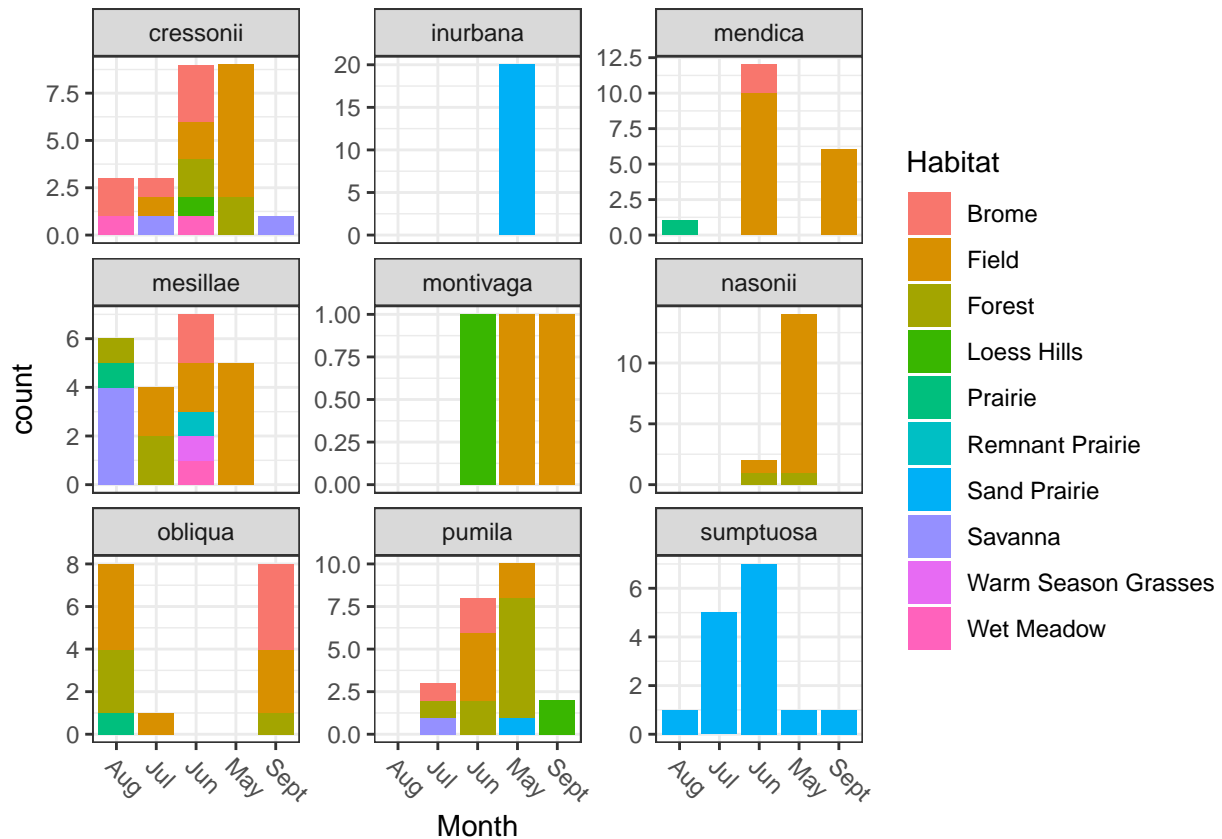
```
habitat_count_f<-bees.rare.spe%>%count(Habitat)%>%drop_na()
month_f<-bees.rare.spe%>%count(Month)%>%drop_na()

bees_new_f1<-bees.rare.spe%>%
  filter(Habitat%in%habitat_count_f$Habitat)%>%
  filter(Month%in%month_f$Month)

ggplot(bees_new_f1, aes(Habitat,fill = Species)) +
  geom_bar()+
  theme(axis.text.x=element_text(angle = -45, hjust = 0))
```



```
ggplot(bees_new_f1, mapping = aes(x=Month, fill = Habitat)) +
  geom_bar() +
  facet_wrap(~ Species, scale = "free_y") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = -45, hjust = 0))
```



Fisher Exact test

Independency between Habitat and Species

```
spec<-bees.rare.spe%>%count(Species)
hab<-bees.rare.spe%>%count(Habitat)
fisher.matr<-cbind(spec$n,hab$n)
fisher.test(fisher.matr,simulate.p.value = TRUE, B = 1e5)

##
## Fisher's Exact Test for Count Data with simulated p-value (based
## on 1e+05 replicates)
##
## data: fisher.matr
## p-value = 1e-05
## alternative hypothesis: two.sided
```

Independency between month and Species

```
spec<-bees.rare.spe%>%count(Species)
mont<-bees.rare.spe%>%count(Month)
```

```
fisher.matr<-cbind(spec$n,mont$n)
fisher.test(fisher.matr,simulate.p.value = TRUE, B = 1e5)

##
## Fisher's Exact Test for Count Data with simulated p-value (based
## on 1e+05 replicates)
##
## data: fisher.matr
## p-value = 1e-05
## alternative hypothesis: two.sided
```