

CODES

```
install.packages("vegan")
```

```
library("vegan")
```

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

```
install.packages("ggplot2")
```

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

```
#obtain data from Bold
```

```
Anopheles <-
```

```
read_tsv("http://www.boldsystems.org/index.php/API_Public/specimen?taxon=Anopheles&format=tsv")
```

```
write_tsv(Anopheles, "Anopheles_BOLD_data.tsv")
```

```
Anopheles <- read_tsv("Anopheles_BOLD_data.tsv")
```

```
#subset1 of Data
```

```
Anopheles2 <- Anopheles[, c(1, 8, 10, 12, 14, 16, 18, 20, 22, 39, 40, 41, 47, 48, 55:57)]
```

```
#subset2 of Data
```

```
Anopheles3 <- Anopheles2[, c(2, 9, 13:15)]
```

```
#number of sequences per specie by country
```

```
dfspecies.by.country <- Anopheles2 %>%
```

```
  group_by(country, species_name) %>%
```

```
  count(species_name) %>%
```

```
  filter(!is.na(country)) %>%
```

```
# filtering "number of sequences per specie by country" by malaria causing mosquitoes

dfsubset.species.by.country <- dplyr::filter(dfspecies.by.country, species_name %in%
c("Anopheles philippinensis", "Anopheles crucians", "Anopheles stephensi", "Anopheles
sundaicus", " Anopheles culicifacies", "Anopheles dirus", "Anopheles sinensis", "Anopheles
nivipes", "Anopheles quadrimaculatus", "Anopheles freeborni", "Anopheles gambiae",
"Anopheles funestus"))
```

```
#ploting barchart to show countries where malaria causing species are found

ggplot(dfsubset.species.by.country, aes(n, country, fill = species_name)) +

  geom_bar(stat = "identity")
```

```
#filtering BOLD data by malaria causing species
```

```
dfsubset.map.species <- dplyr::filter(Anopheles3, species_name %in% c("Anopheles
philippinensis", "Anopheles crucians", "Anopheles stephensi", "Anopheles sunaicus", "
Anopheles culicifacies", "Anopheles dirus", "Anopheles sinensis", "Anopheles nivipes",
"Anopheles quadrimaculatus", "Anopheles freeborni", "Anopheles gambiae", "Anopheles
funestus"))
```

```
#histogram showing lon of countries where selected countries lies
```

```
hist(dfsubset.map.species$lon)
```

```
#histogram showing lat of countries where selected countries lies
```

```
hist(dfsubset.map.species$lat)
```

Reference

<https://sparkbyexamples.com/r-programming/select-rows-in-r/>