TidyTuesday World Spider Database

David Jackson

12/7/2021

```
spiders <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2
```

```
## Rows: 49785 Columns: 10
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (7): species_lsid, family, genus, species, subspecies, author, distribution
## dbl (3): speciesId, year, parentheses
```

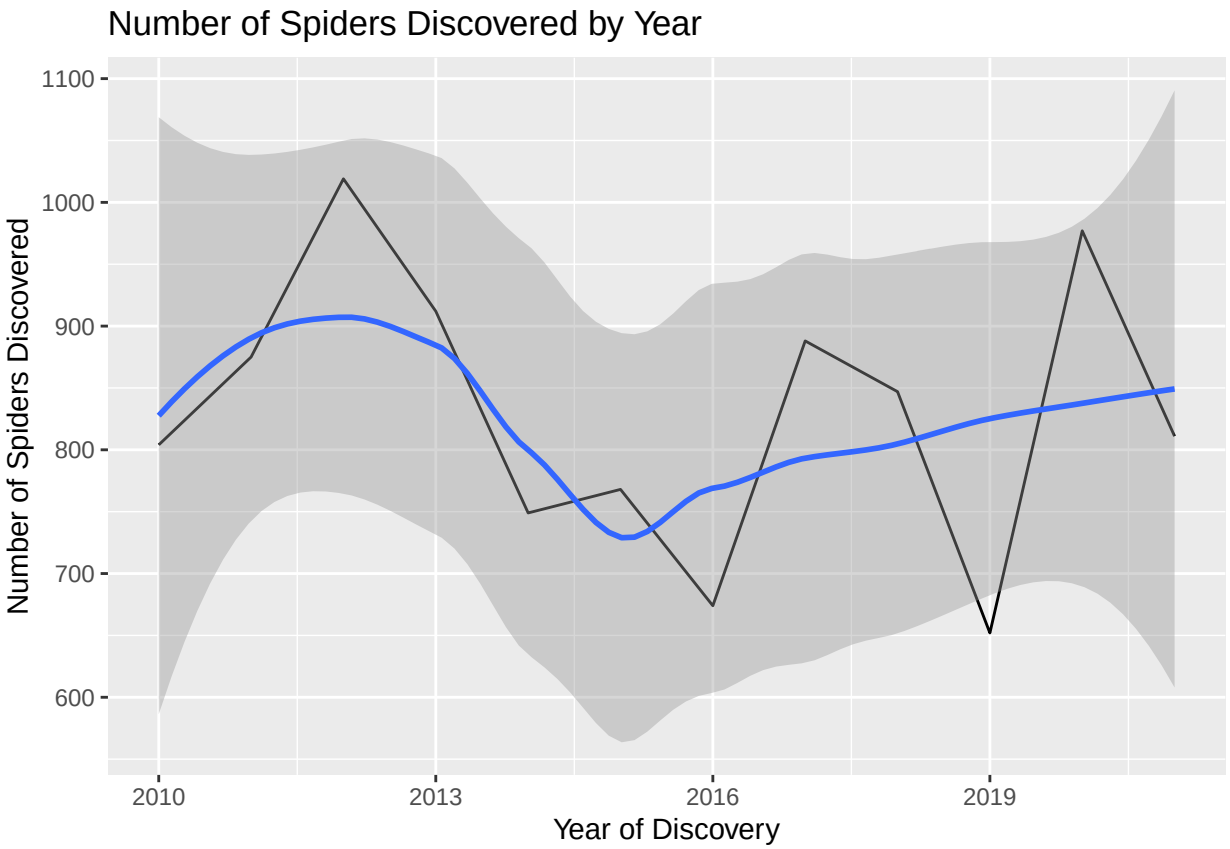
```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
spiders <- spiders %>% janitor::clean_names()
spiders <- spiders %>% janitor::remove_empty(which = c("cols","rows"))
```

Discoveries by Year: 2010 - 2021

```
spiders %>% count(year) %>%
  ggplot(aes(x=year,y=n)) + geom_line() +
  labs(title = "Number of Spiders Discovered by Year",
       x="Year of Discovery",y="Number of Spiders Discovered") +
  geom_smooth(aes(x=year,y=n))
```

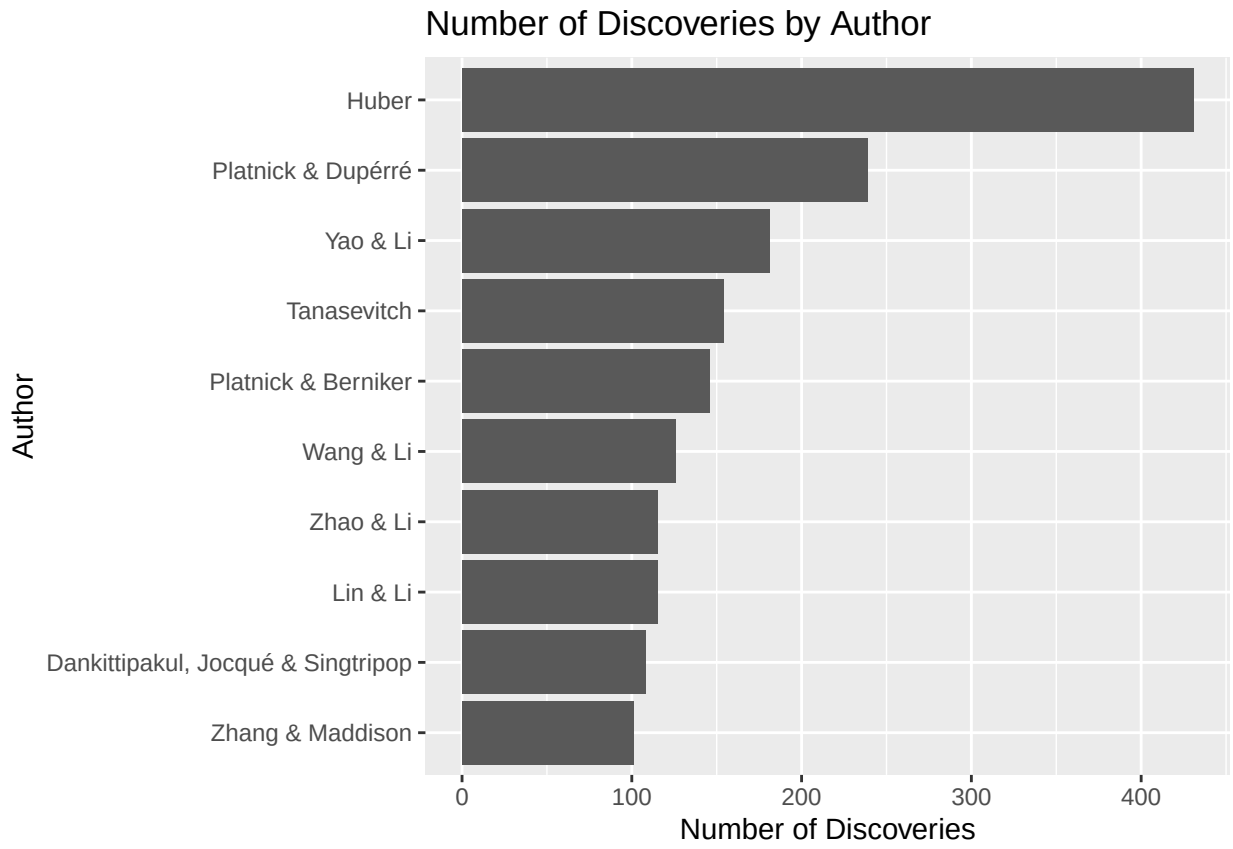
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Count of Spiders Discovered by Author

```
spiders %>% count(author,sort =T) %>% top_n(10) %>% ggplot(aes(x=reorder(author,n),y=n)) +
  geom_col() + coord_flip() +
  labs(title ="Number of Discoveries by Author",
       x="Author",y="Number of Discoveries")
```

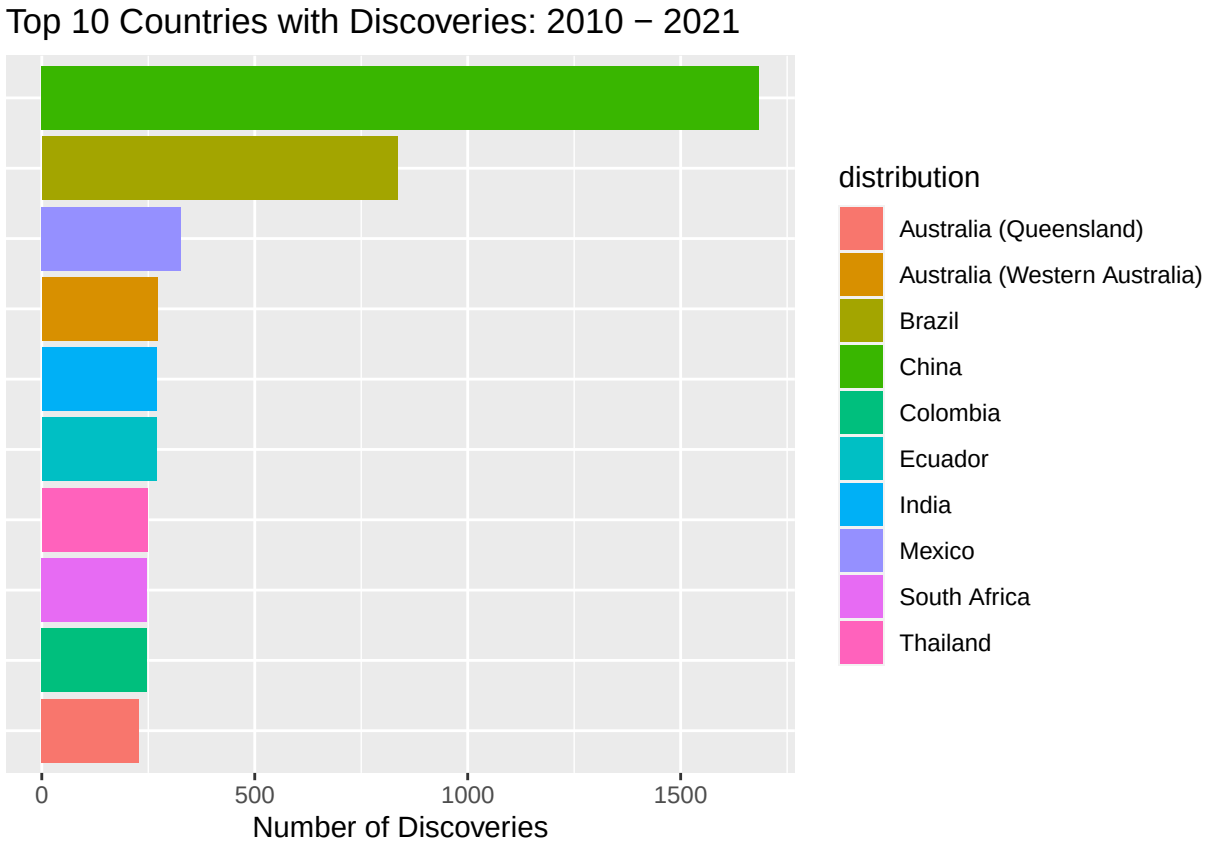
```
## Selecting by n
```



Distribuation by Location: 2010-2021

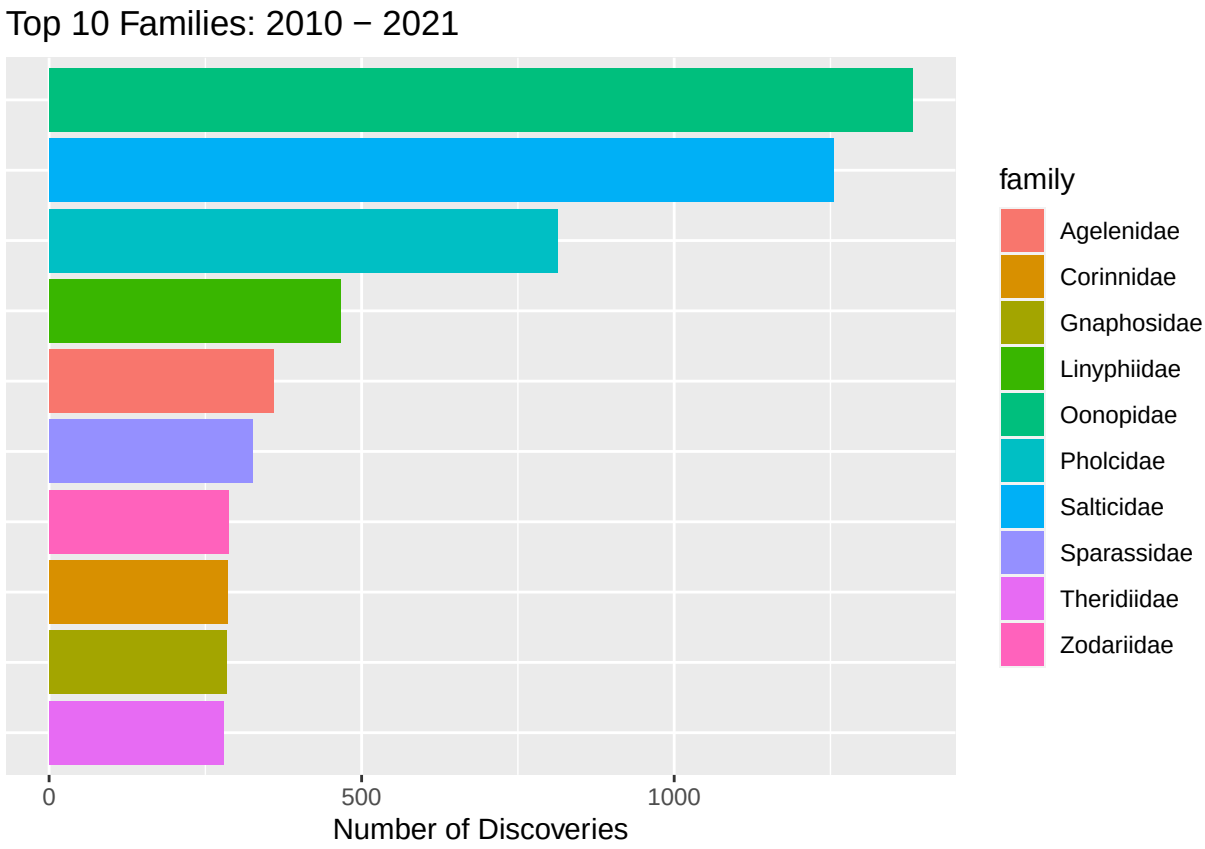
```
spiders %>% count(distribution,sort =T) %>% top_n(10) %>%
  ggplot(aes(x=reorder(distribution,n),y=n)) + geom_col(aes(fill=distribution)) +
  theme(axis.title.y=element_blank(),
    axis.text.y=element_blank(),
    axis.ticks.y=element_blank()) +
  labs(title = "Top 10 Countries with Discoveries: 2010 - 2021",
       y="Number of Discoveries") + coord_flip()
```

```
## Selecting by n
```




```
spiders %>% count(family,sort =T) %>% top_n(10) %>%
  ggplot(aes(x=reorder(family,n),y=n)) + geom_col(aes(fill=family)) +
  theme(axis.title.y=element_blank(),
    axis.text.y=element_blank(),
    axis.ticks.y=element_blank()) +
  labs(title = "Top 10 Families: 2010 - 2021",
       y="Number of Discoveries") + coord_flip()
```

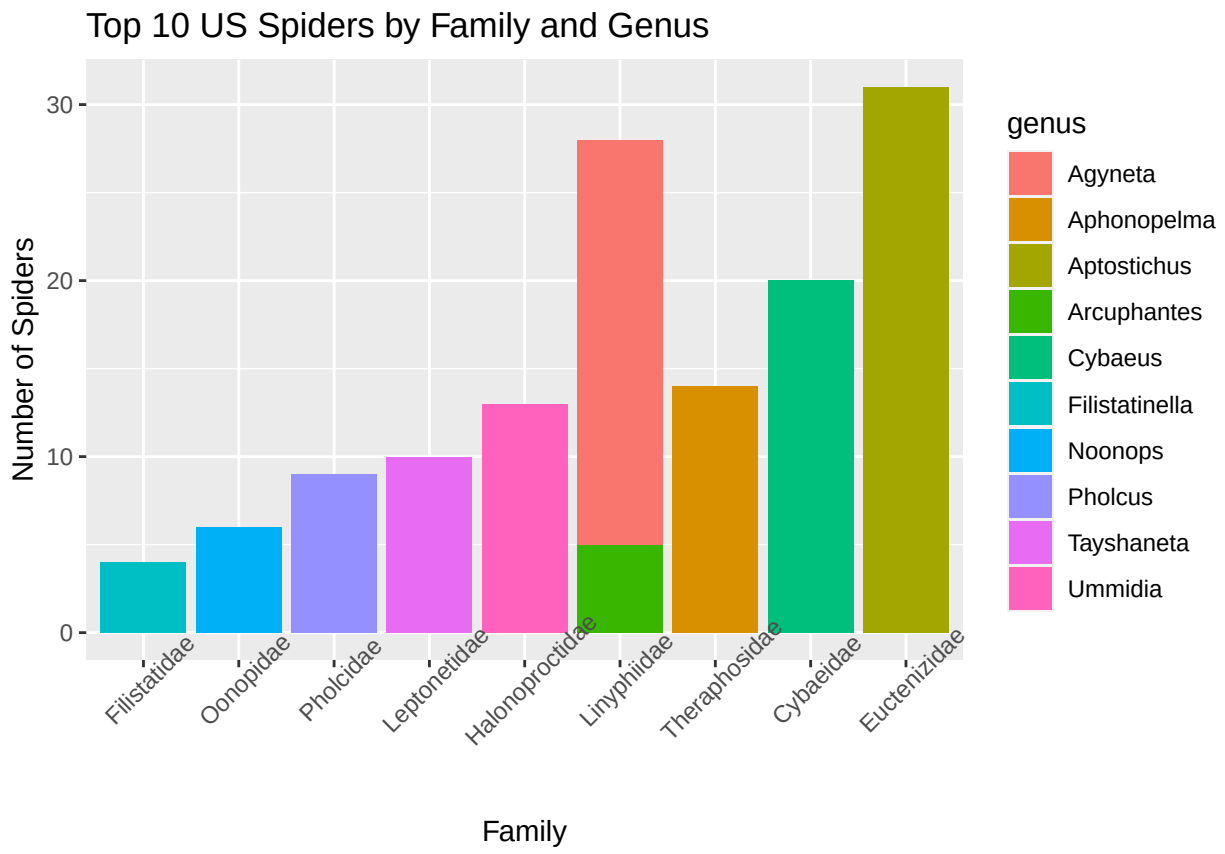
```
## Selecting by n
```



US Discoveries by Family and Genus

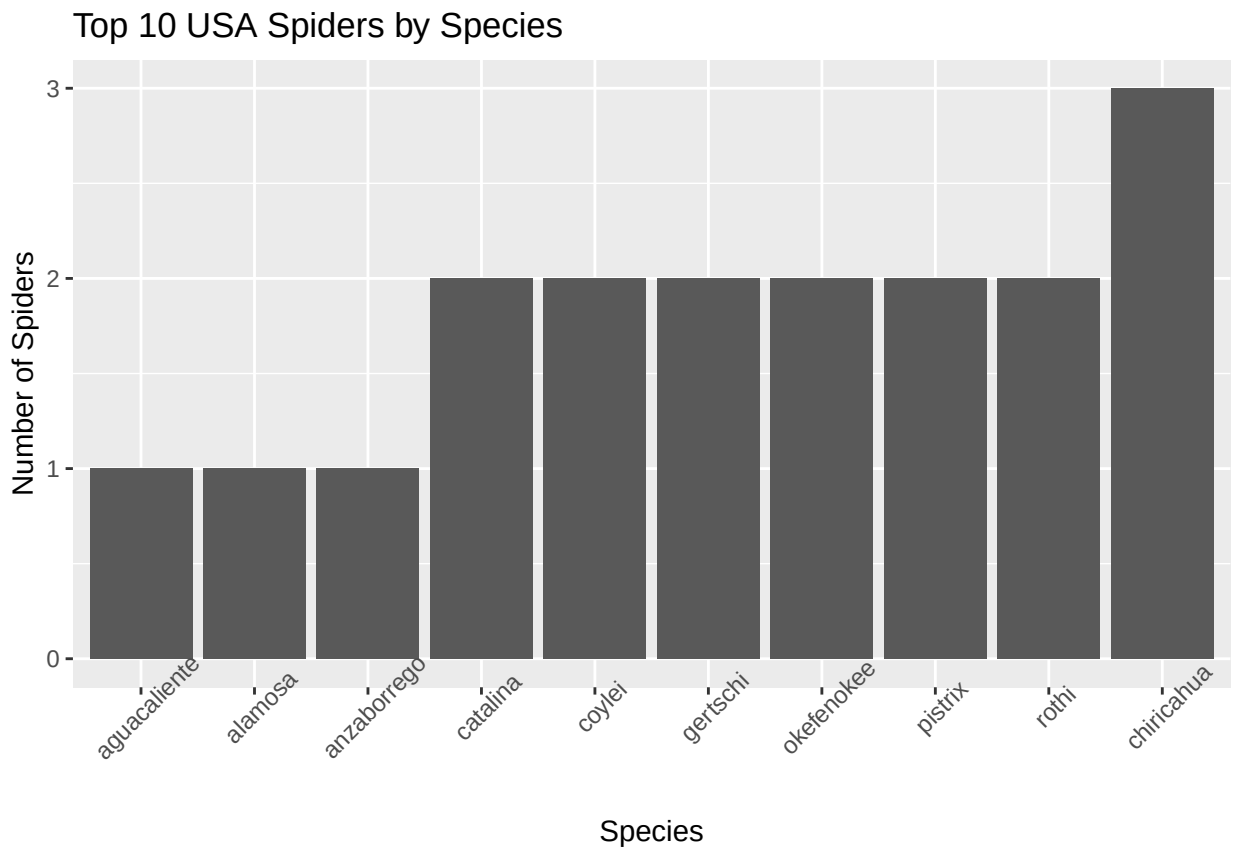
```
spiders %>% filter(distribution =="USA") %>% count(family,genus,sort =T) %>%
  top_n(10) %>%
  ggplot()  + geom_col(aes(x=reorder(family,n),y=n,fill=genus)) +
  theme(axis.text.x = element_text(angle = 45)) +
  labs(title = "Top 10 US Spiders by Family and Genus",x="Family",
       y="Number of Spiders")
```

```
## Selecting by n
```



US Discoveries by Species

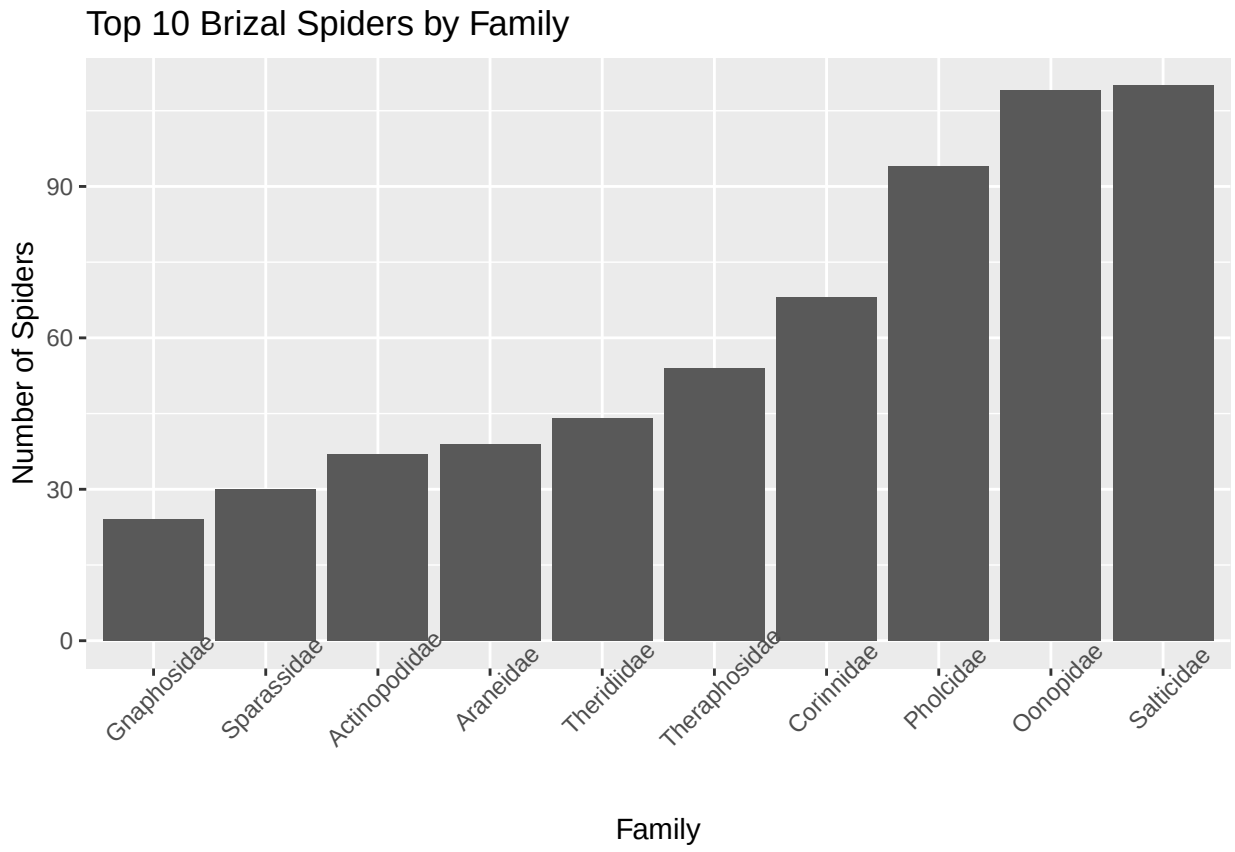
```
spiders %>% filter(distribution =="USA") %>% count(species,sort =T) %>%
  head(10) %>%
  ggplot()  + geom_col(aes(x=reorder(species,n),y=n)) +
  theme(axis.text.x = element_text(angle = 45)) +
  labs(title = "Top 10 USA Spiders by Species",x="Species",
       y="Number of Spiders")
```



Brazil

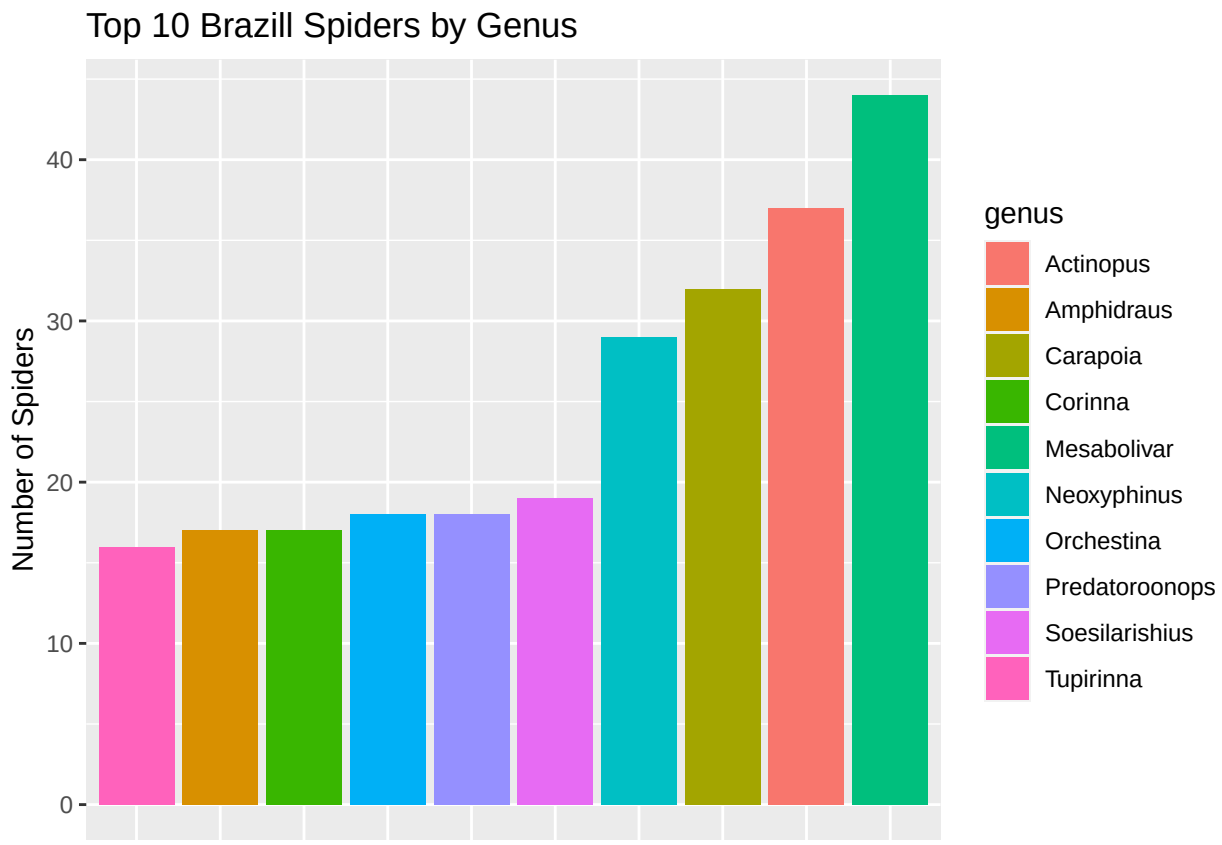
```
spiders %>% filter(distribution =="Brazil") %>% count(family,sort =T) %>%
  top_n(10) %>%
  ggplot()  + geom_col(aes(x=reorder(family,n),y=n)) +
  theme(axis.text.x = element_text(angle = 45)) +
  labs(title = "Top 10 Brizal Spiders by Family",x="Family",
       y="Number of Spiders")
```

```
## Selecting by n
```




```
spiders %>% filter(distribution =="Brazil") %>% count(genus,sort =T) %>%
  top_n(10) %>%
  ggplot()  + geom_col(aes(x=reorder(genus,n),y=n,fill=genus)) +
  theme(axis.text.x = element_text(angle = 45)) +
  labs(title = "Top 10 Brazill Spiders by Genus",x="Genus",
       y="Number of Spiders") +
   theme(axis.title.x=element_blank(),
    axis.text.x=element_blank(),
    axis.ticks.x=element_blank())
```

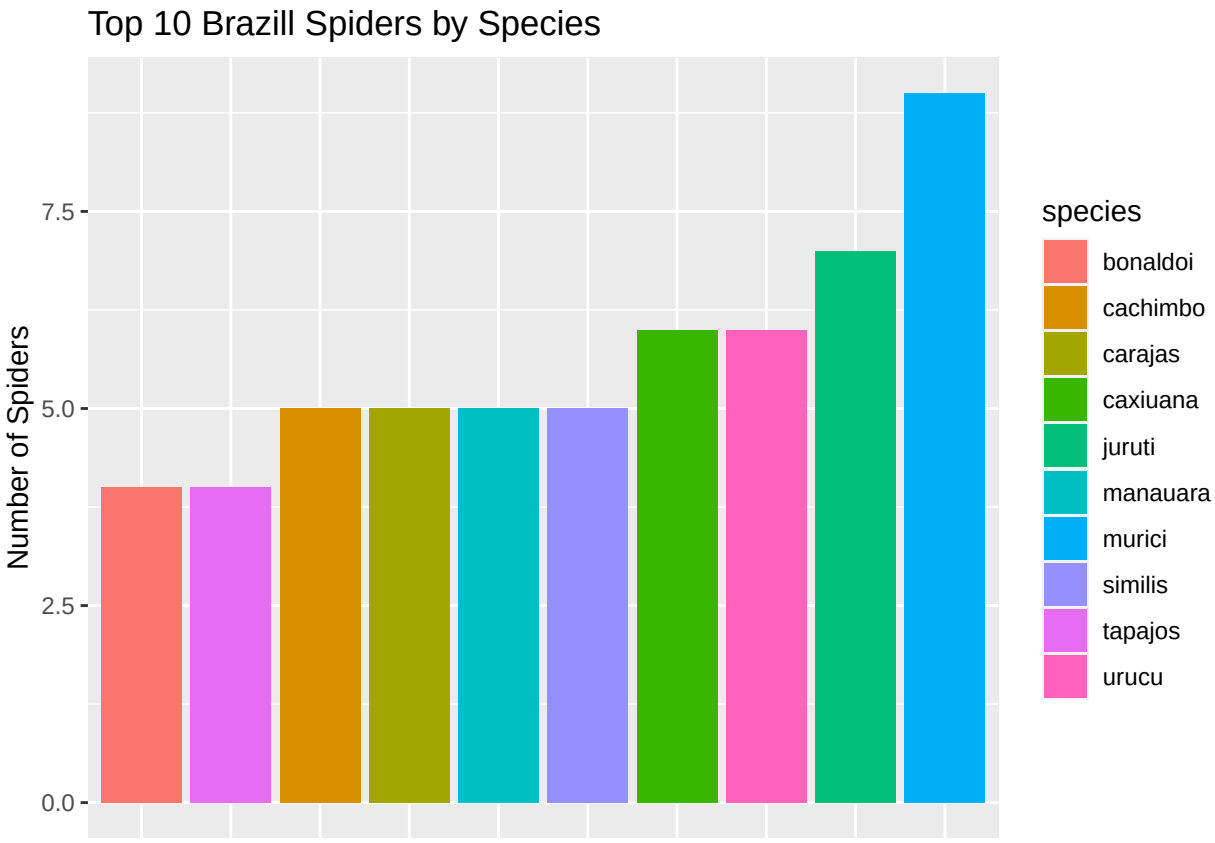
```
## Selecting by n
```



Brazil's Top 10 Spiders by Species

```
spiders %>% filter(distribution =="Brazil") %>% count(species,sort =T) %>%
  top_n(10) %>%
  ggplot()  + geom_col(aes(x=reorder(species,n),y=n,fill=species)) +
  theme(axis.text.x = element_text(angle = 45)) +
  labs(title = "Top 10 Brazill Spiders by Species",x="Species",
       y="Number of Spiders") +
   theme(axis.title.x=element_blank(),
    axis.text.x=element_blank(),
    axis.ticks.x=element_blank())
```

```
## Selecting by n
```


Top 10 Brazill Spiders by Species
Number of Spiders
0.0
2.5
5.0
7.5
species
bonaldoi
cachimbo
carajas
caxiuana
juruti
manauara
murici
similis
tapajos
urucu