

exo_ggplot

Diego Ayala

11/4/2020

```
library(ggplot2)
library(forcats)
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## v tibble 3.0.4      v purrr 0.3.4
## v tidyr 1.1.2       v dplyr 1.0.2
## v readr 1.3.1       v stringr 1.4.0
```

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

Dataset

```
pantheria <- readr::read_delim(here::here("data", "pantheria-traits", "PanTHERIA_1-0_WR05_Aug2008.txt"))
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   MSW05_Order = col_character(),
##   MSW05_Family = col_character(),
##   MSW05_Genus = col_character(),
##   MSW05_Species = col_character(),
##   MSW05_Binomial = col_character(),
##   References = col_character()
## )
```

```
## See spec(...) for full column specifications.
```

```
dat <- pantheria %>%
  mutate(
    order = as_factor(MSW05_Order),
    family = as_factor(MSW05_Family)
  ) %>%
  rename(
    # Conversion de type
    # Nom des colonnes
```

```

    adult_bodymass = '5-1_AdultBodyMass_g',
    dispersal_age   = '7-1_DispersalAge_d',
    gestation       = '9-1_GestationLen_d',
    homerange       = '22-2_HomeRange_Indiv_km2',
    litter_size     = '16-1_LittersPerYear',
    longevity       = '17-1_MaxLongevity_m'
  )

dat <- dat %>%
  select(                                # Sélection de colonnes
    order,
    family,
    adult_bodymass,
    dispersal_age,
    gestation,
    homerange,
    litter_size,
    longevity
  ) %>%
  na_if(-999)

dat %>% View()

```

Explore

```

dat %>%
  count(order)

```

```

## # A tibble: 29 x 2
##   order      n
##   <fct>    <int>
## 1 Artiodactyla    240
## 2 Carnivora      286
## 3 Primates       376
## 4 Didelphimorphia  87
## 5 Cingulata      21
## 6 Rodentia      2277
## 7 Cetacea        84
## 8 Erinaceomorpha  24
## 9 Diprotodontia   143
## 10 Chiroptera    1116
## # ... with 19 more rows

```

```

dat %>%
  count(family)

```

```

## # A tibble: 153 x 2
##   family      n
##   <fct>    <int>

```

```
## 1 Camelidae      4
## 2 Canidae       35
## 3 Bovidae      143
## 4 Pitheciidae   40
## 5 Didelphidae   87
## 6 Dasypodidae   21
## 7 Muridae      730
## 8 Spalacidae    36
## 9 Cricetidae   681
## 10 Neobalaenidae 1
## # ... with 143 more rows
```

```
dat %>%
  filter(!is.na(homerange)) %>%
  summarise(mean(homerange))
```

```
## # A tibble: 1 x 1
##   'mean(homerange)'
##           <dbl>
## 1           327.
```

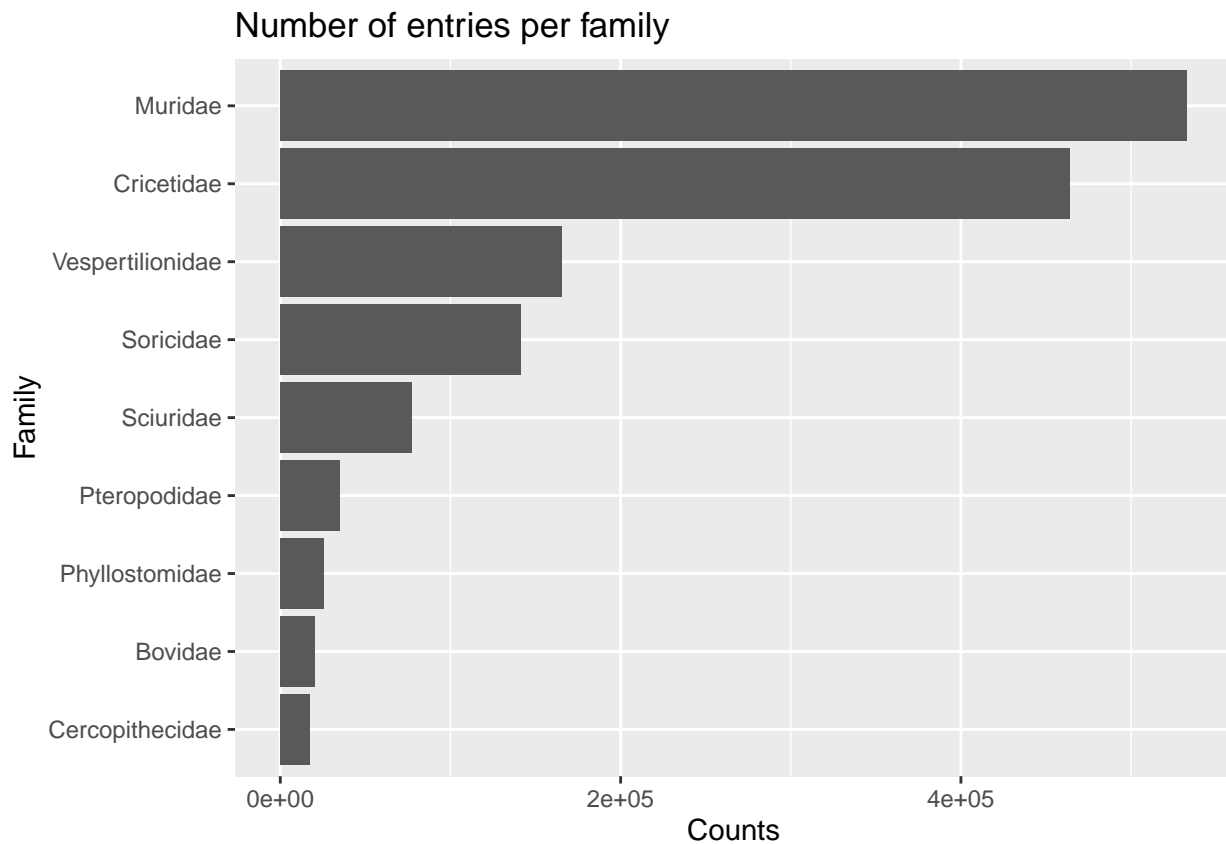
```
dat %>%
  filter(!is.na(homerange)) %>%
  group_by(family) %>%
  summarise(m = mean(homerange), sd = sd(homerange), n = n())
```

```
## 'summarise()' ungrouping output (override with '.groups' argument)
```

```
## # A tibble: 93 x 4
##   family      m      sd      n
##   <fct>    <dbl>  <dbl> <int>
## 1 Camelidae  0.13    NA      1
## 2 Canidae   49.6   176.    21
## 3 Bovidae   44.6   159.    24
## 4 Pitheciidae 0.8    1.27     7
## 5 Didelphidae 0.0704  0.160   10
## 6 Dasypodidae 1.38    2.14     4
## 7 Muridae    0.00336 0.00839 35
## 8 Spalacidae 0.0000732 NA      1
## 9 Cricetidae 0.00305  0.00572 62
## 10 Erinaceidae 0.0255  0.0306   4
## # ... with 83 more rows
```

Including Plots

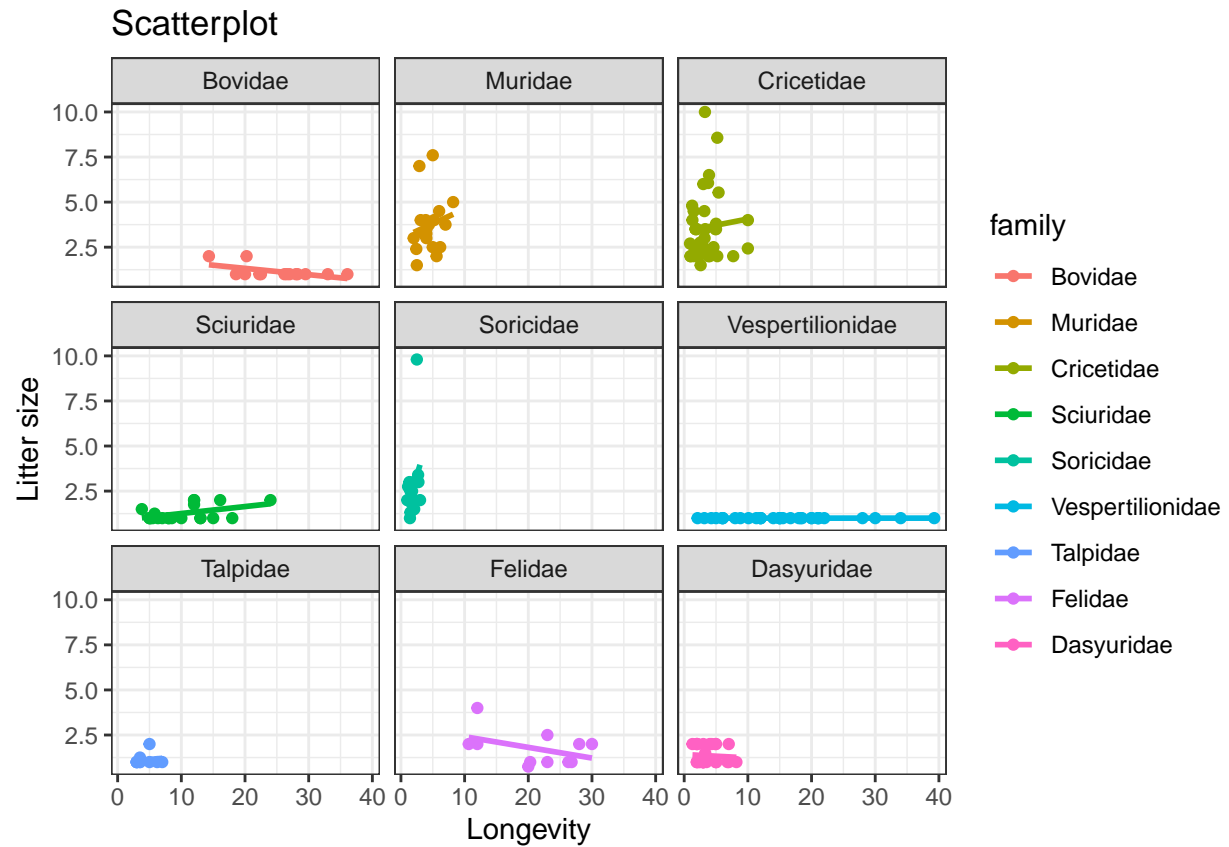
Plot 1



Plot 2

```
theme_set(theme_bw()) # play around with theme
dat %>%
  filter(!is.na(litter_size), !is.na(longevity)) %>%
  group_by(family) %>% # group by family
  mutate(n = n()) %>% # count the number of entries per family
  mutate(longevity = longevity / 12) %>% # Change month to year
  filter(n > 10) %>% # select only those families with more than 50 entries
  ggplot() +
  aes(x = longevity, y = litter_size, col = family) + # scatter plot
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) + # se = TRUE to add confidence intervals
  xlab("Longevity") + # add label for X axis
  ylab("Litter size") + # add label for Y axis
  ggtitle("Scatterplot") + # add title
  facet_wrap(~ family, nrow = 3) # split in several panels,
```

```
## 'geom_smooth()' using formula 'y ~ x'
```



```
# one for each family
# remove scale = 'free' for
# same scale for all plots
```