

Expository graphics

Rabya Ghafoor

4/21/2020

```
library(tidyverse)
```

```
## -- Attaching packages -----
```

```
## v ggplot2 3.2.1      v purrr    0.3.3
## v tibble   2.1.3      v dplyr    0.8.4
## v tidyr    1.0.2      v stringr  1.4.0
## v readr    1.3.1      vforcats  0.4.0
```

```
## -- Conflicts -----
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()
```

```
library(ggplot2)
library(ggrepel)
```

1.

```
download.file(url="https://ndownloader.figshare.com/files/2292169",
              destfile = "data/portal_data_joined.csv")
```

```
surveys <- read.csv("data/portal_data_joined.csv")
surveys_complete <- surveys %>%
  filter(!is.na(weight),           # remove missing weight
         !is.na(hindfoot_length), # remove missing hindfoot_length
         !is.na(sex))           # remove missing sex
```

```
## Extract the most common species_id
species_counts <- surveys_complete %>%
  count(species_id) %>%
  filter(n >= 50)

## Only keep the most common species
surveys_complete <- surveys_complete %>%
  filter(species_id %in% species_counts$species_id)

write_csv(surveys_complete, path = "data/surveys_complete.csv")
```

```

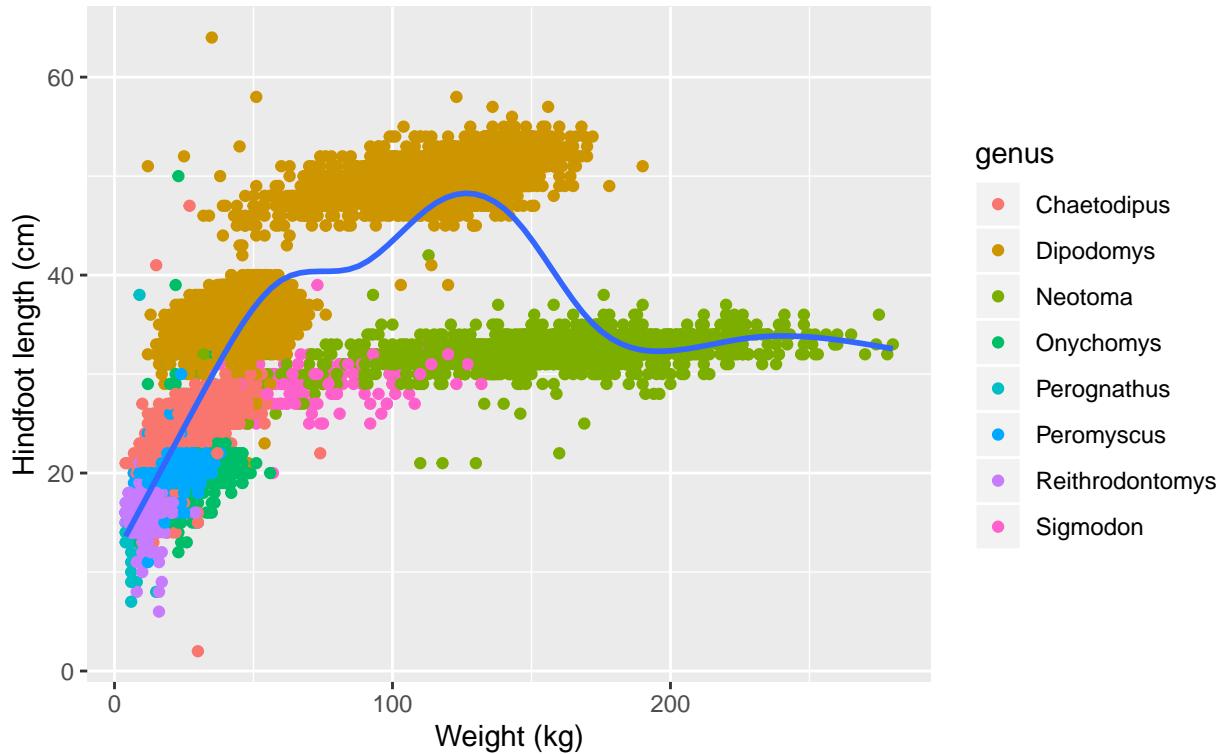
surveys_complete <- read_csv("data/surveys_complete.csv")

## Parsed with column specification:
## cols(
##   record_id = col_double(),
##   month = col_double(),
##   day = col_double(),
##   year = col_double(),
##   plot_id = col_double(),
##   species_id = col_character(),
##   sex = col_character(),
##   hindfoot_length = col_double(),
##   weight = col_double(),
##   genus = col_character(),
##   species = col_character(),
##   taxa = col_character(),
##   plot_type = col_character()
## )
ggplot(surveys_complete, aes(weight, hindfoot_length)) +
  geom_point(aes(color = genus)) +
  geom_smooth(se = FALSE) +
  labs(title = "Hindfoot length vs Weight by Species ID",
       subtitle = "Hindfoot length generally increases with weight with the exception of Neotoma",
       y = "Hindfoot length (cm)",
       x = "Weight (kg)"
  )
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

```

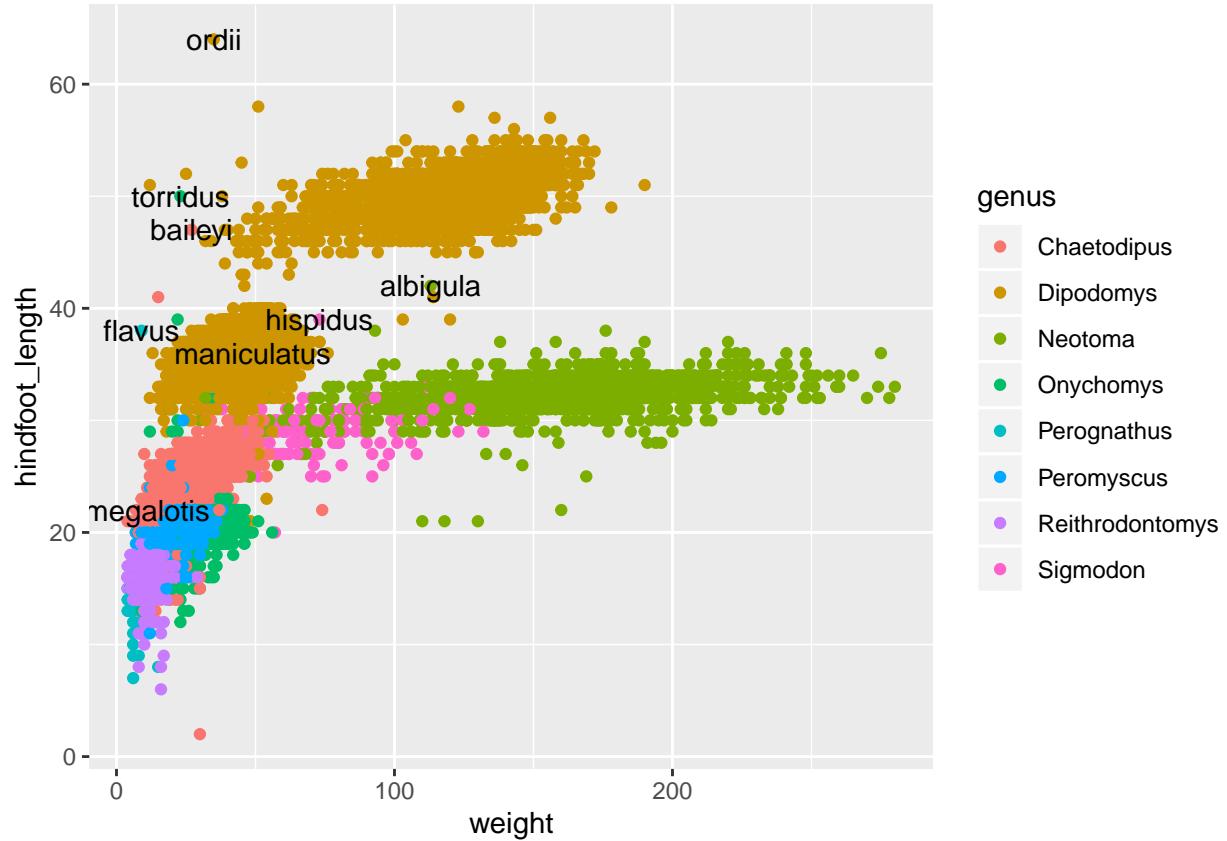
Hindfoot length vs Weight by Species ID

Hindfoot length generally increases with weight with the exception of Neotoma

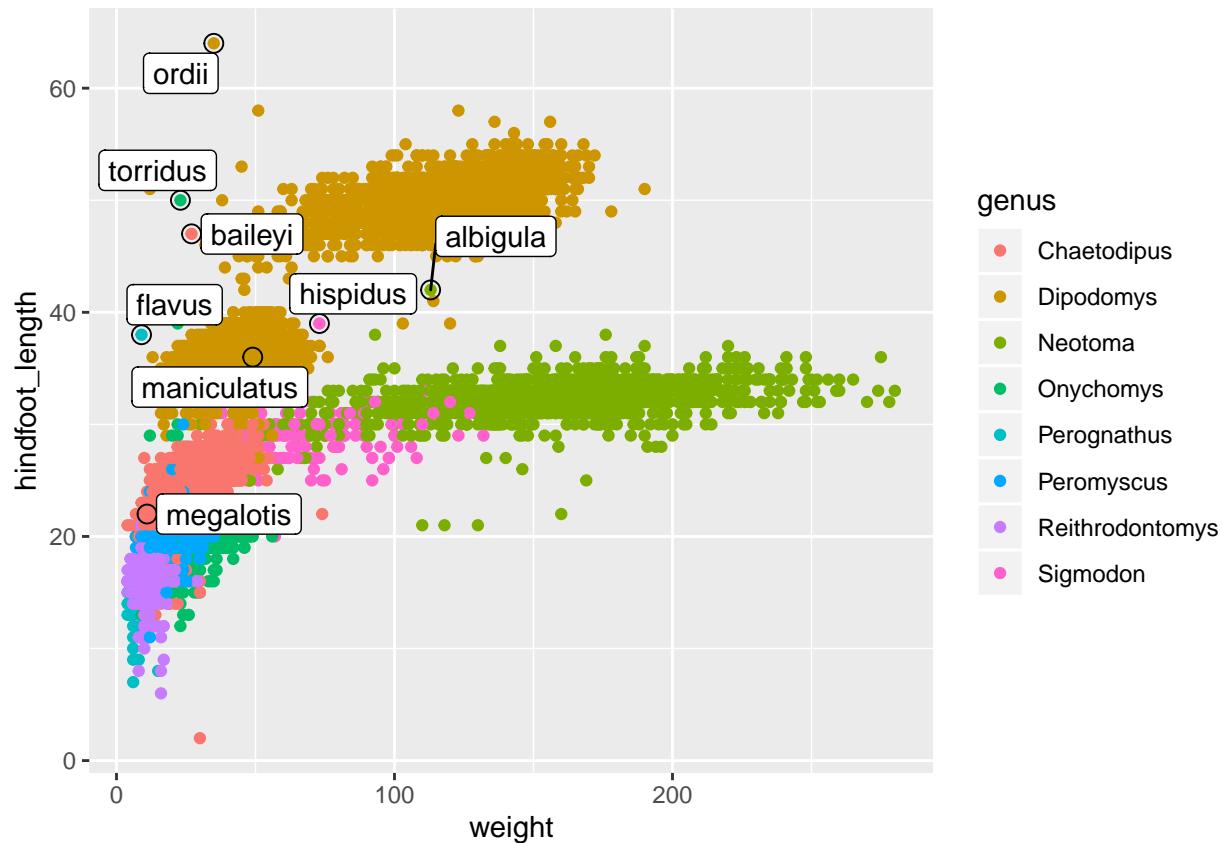


2.

```
best_in_genus <- surveys_complete %>%
  group_by(genus) %>%
  filter(row_number(desc(hindfoot_length)) == 1)
## Warning in rep.default(list_of(integer()), length = nrow(groups)): partial
## argument match of 'length' to 'length.out'
ggplot(surveys_complete, aes(weight, hindfoot_length)) +
  geom_point(aes(colour = genus)) +
  geom_text(aes(label = species), data = best_in_genus)
```

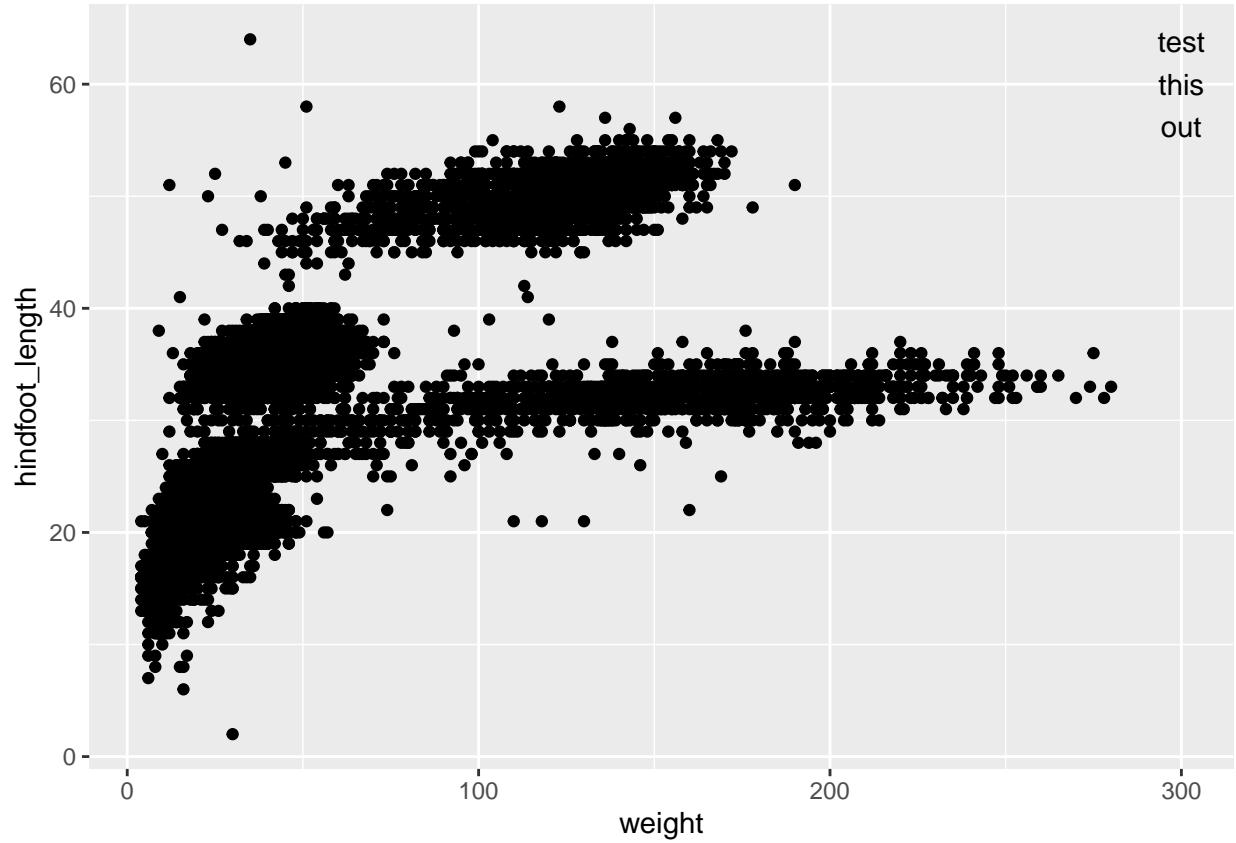


```
ggplot(surveys_complete, aes(weight, hindfoot_length)) +
  geom_point(aes(colour = genus)) +
  geom_point(size = 3, shape = 1, data = best_in_genus) +
  ggrepel::geom_label_repel(aes(label = species), data = best_in_genus)
```



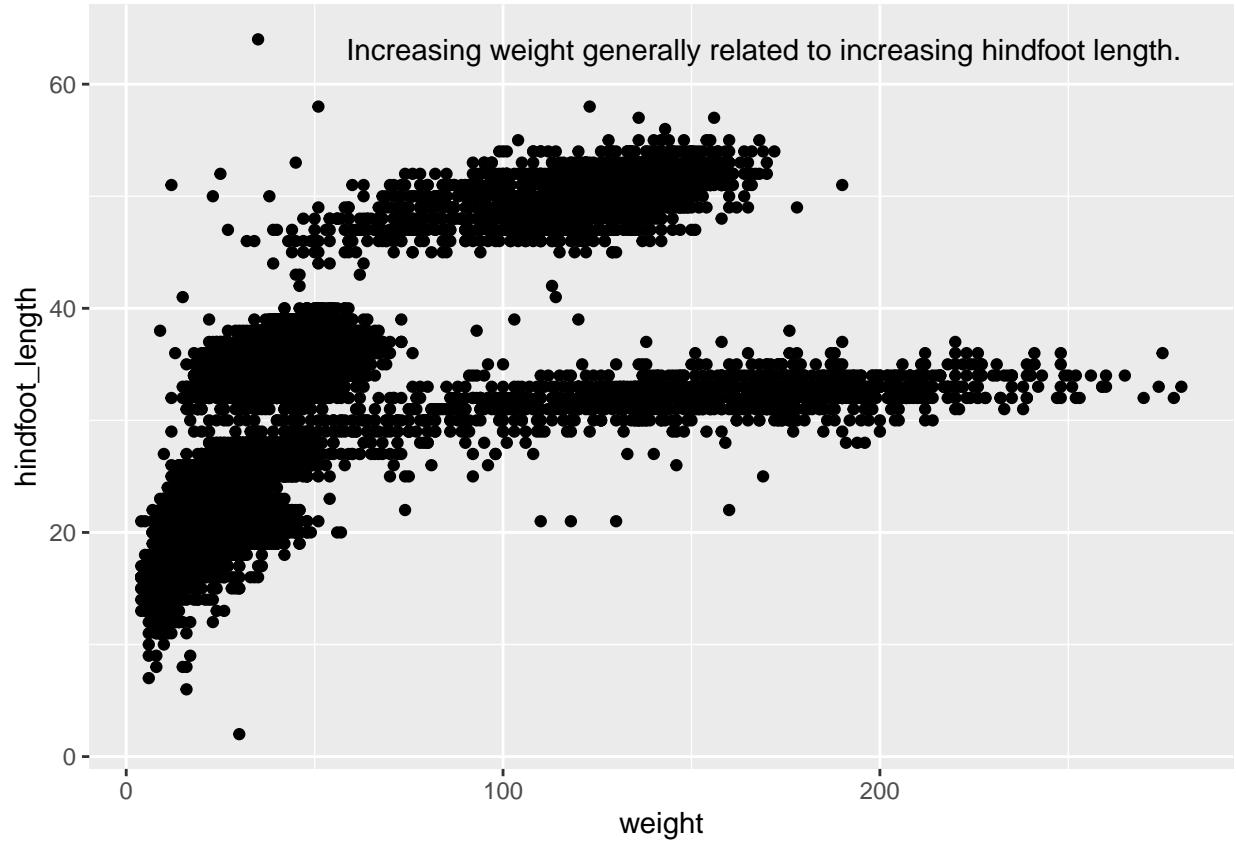
3.

```
text_df <- tibble(text = "test\nthis\nout", x = 300, y = 60)
g1 <- ggplot(surveys_complete, aes(weight, hindfoot_length)) +
  geom_point()
g1 + geom_text(aes(x, y, label = text), data = text_df)
```



```
label <- surveys_complete %>%
  summarise(
    hindfoot_length = max(hindfoot_length),
    weight = max(weight),
    hindfoot_length_min = min(hindfoot_length),
    weight_min = min(weight),
    label = "Increasing weight generally related to increasing hindfoot length."
  )

g1 + geom_text(aes(label = label), data = label,
               vjust = "top", hjust = "right")
```

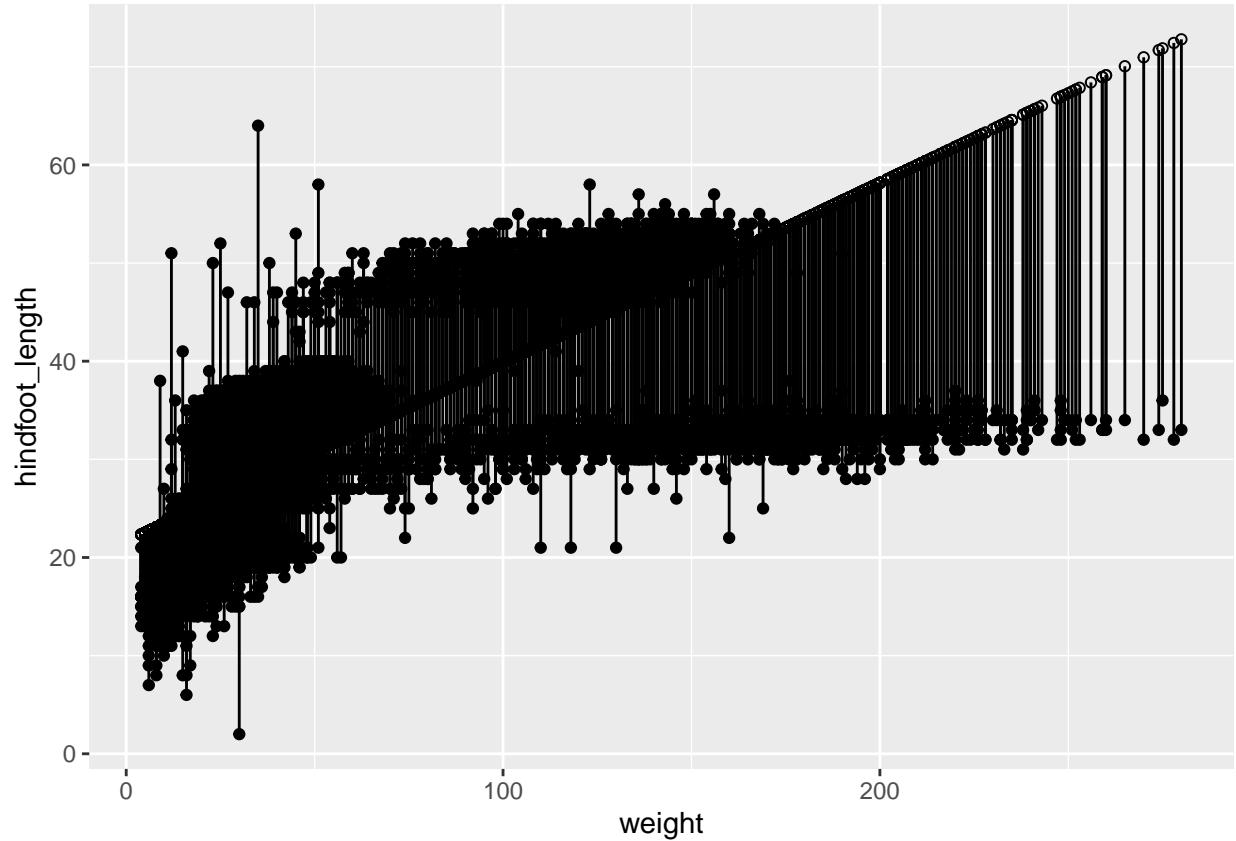


```
fit <- lm(hindfoot_length ~ weight, data = surveys_complete)
surveys_complete$predicted <- predict(fit)
surveys_complete$residuals <- residuals(fit)
```

```
## Warning: Unknown or uninitialized column: 'residuals'.
```

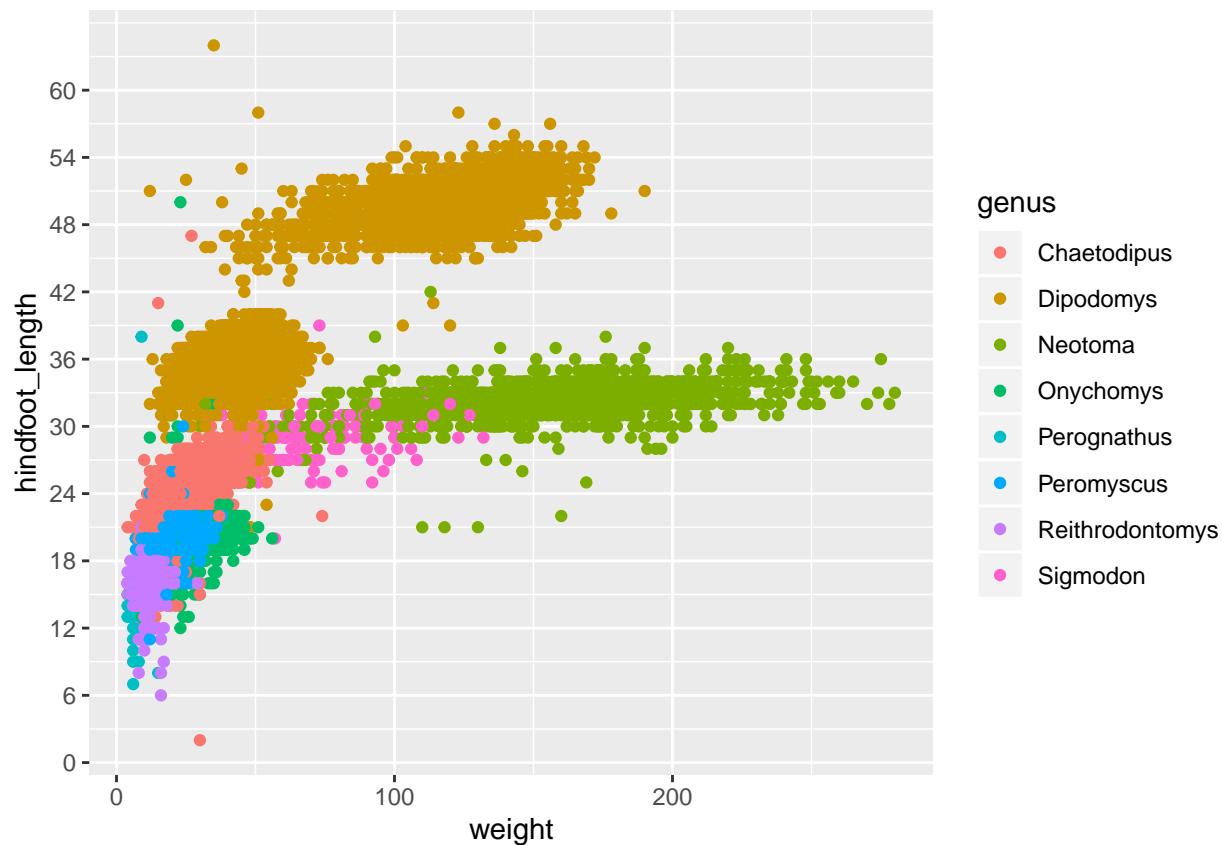
```
## logical(0)
```

```
ggplot(surveys_complete, aes(x = weight, y = hindfoot_length)) +
  geom_segment(aes(xend = weight, yend = predicted)) +
  geom_point() +
  geom_point(aes(y = predicted), shape = 1)
```



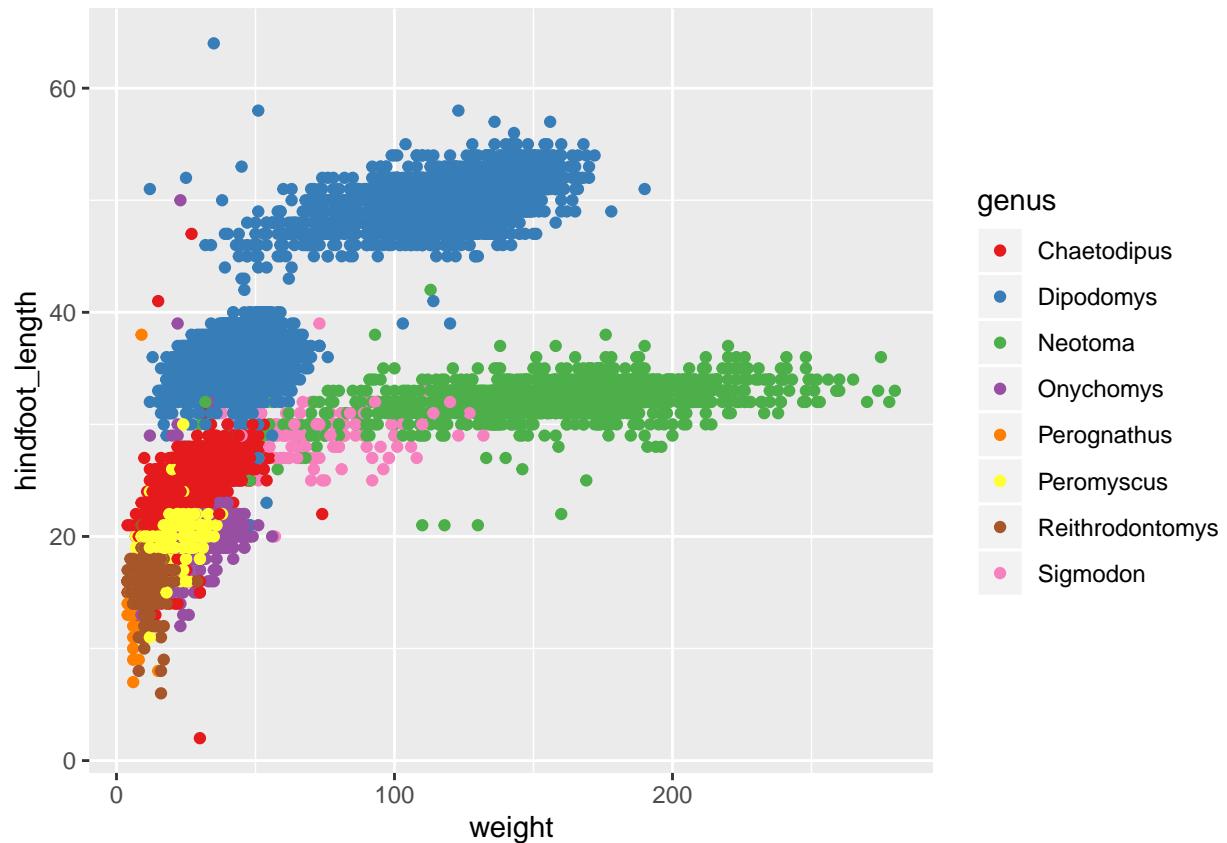
4.

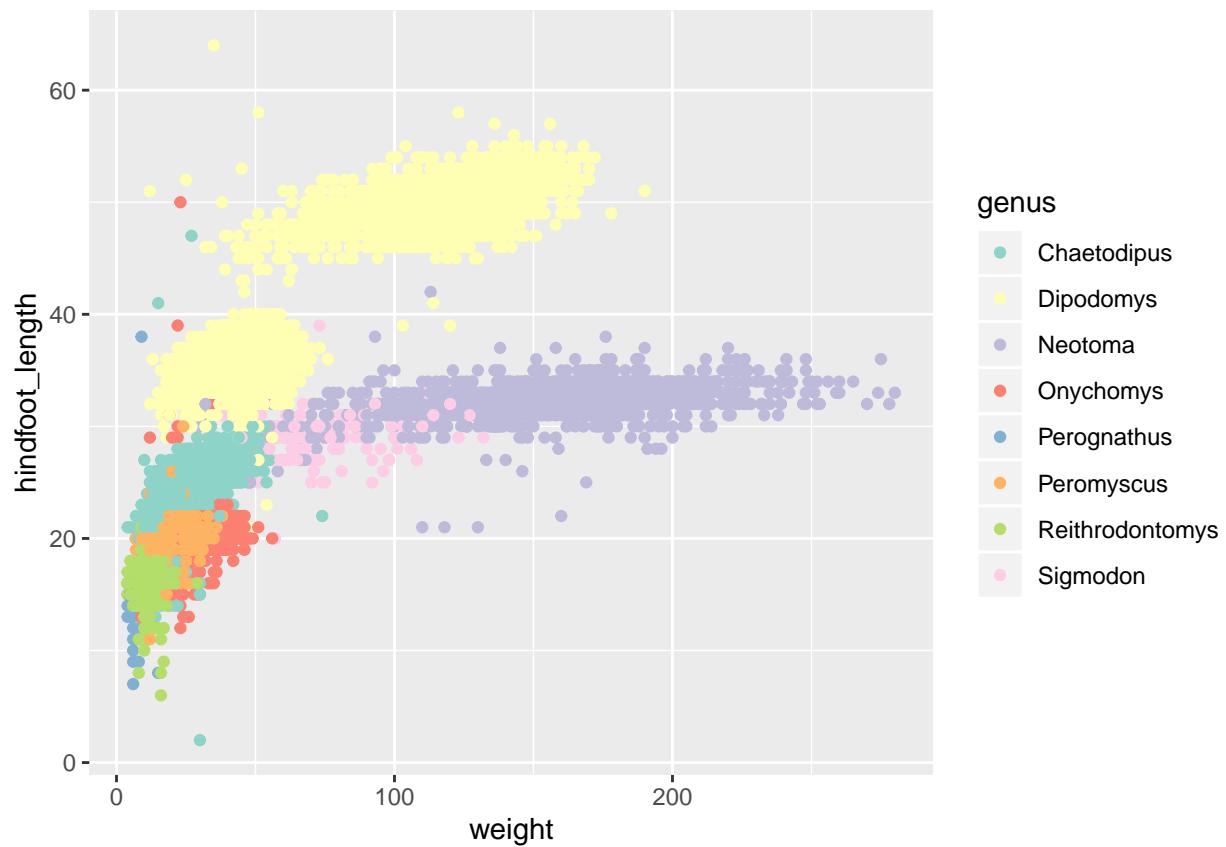
```
ggplot(surveys_complete, aes(weight, hindfoot_length)) +  
  geom_point(aes(colour = genus)) +  
  scale_y_continuous(breaks = seq(0, 60, by = 6))
```



5.

```
ggplot(surveys_complete, aes(weight, hindfoot_length)) +
  geom_point(aes(color = genus)) +
  scale_colour_brewer(palette = "Set1")
```

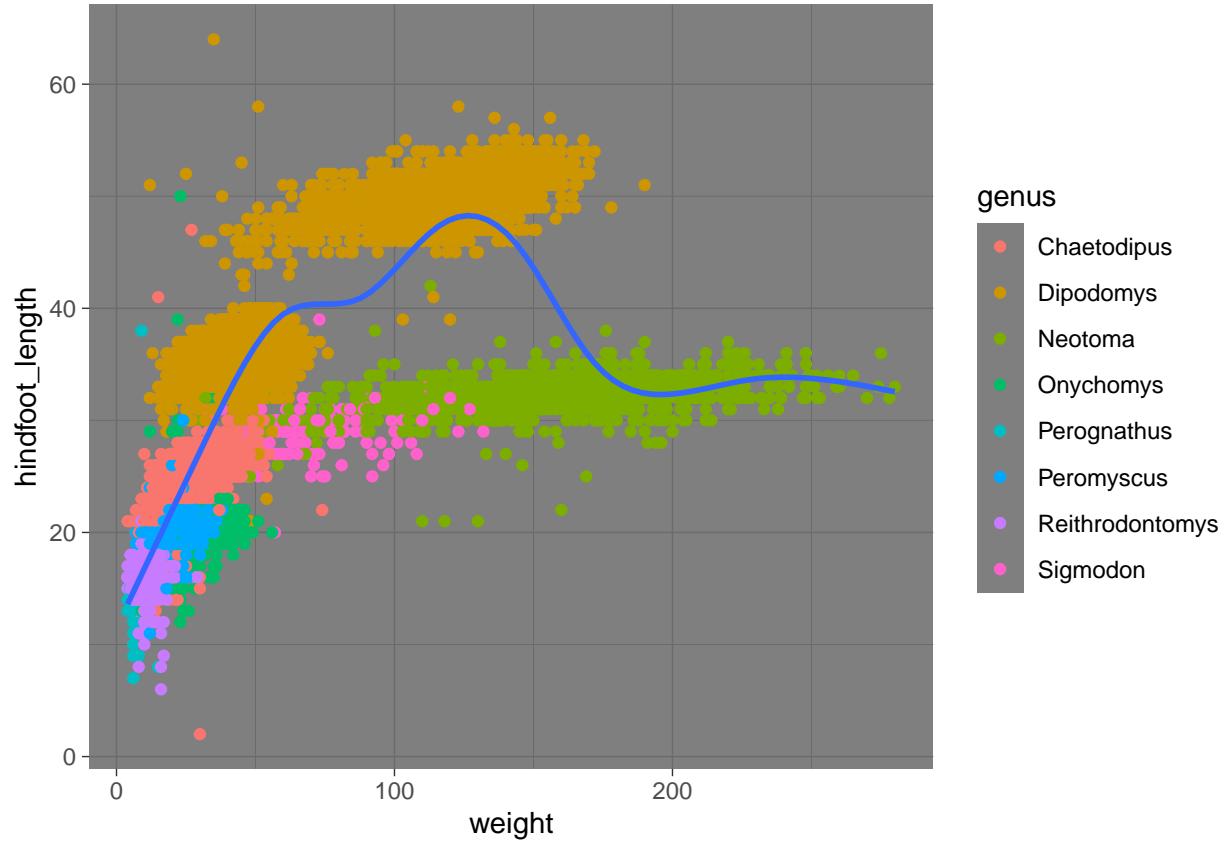




6.

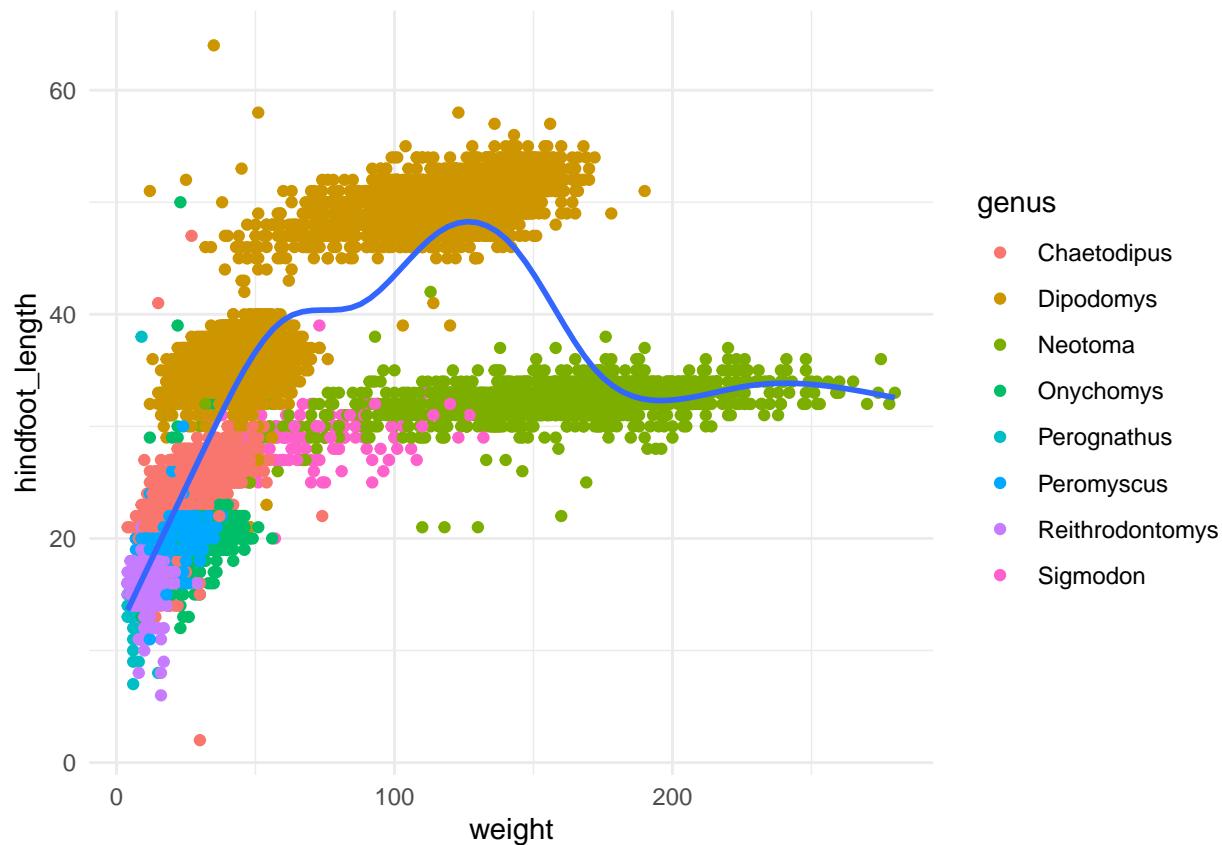
```
ggplot(surveys_complete, aes(weight, hindfoot_length)) +
  geom_point(aes(color = genus)) +
  geom_smooth(se = FALSE) +
  theme_dark()
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



```
ggplot(surveys_complete, aes(weight, hindfoot_length)) +
  geom_point(aes(color = genus)) +
  geom_smooth(se = FALSE) +
  theme_minimal()
```

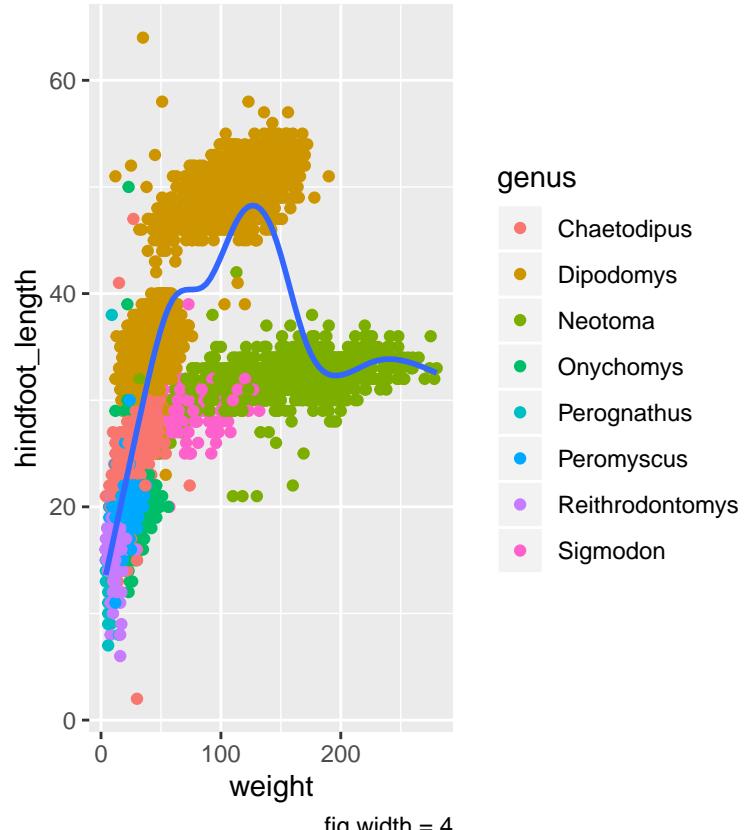
```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



7.

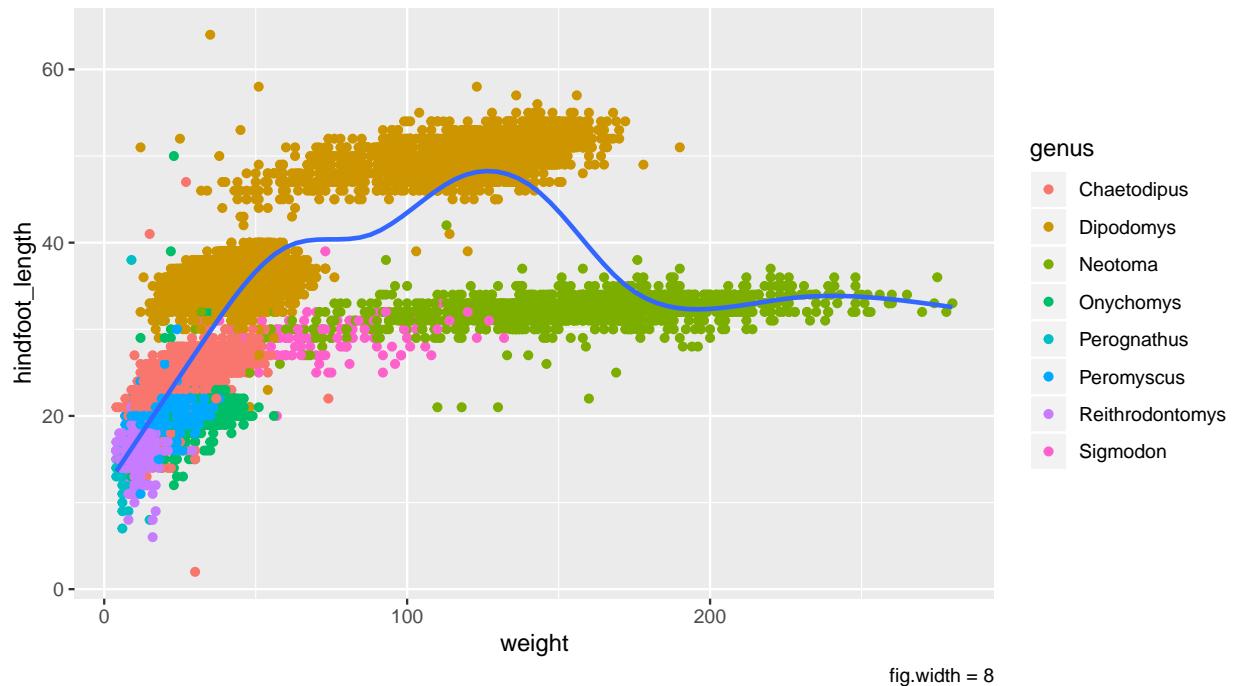
```
ggplot(surveys_complete, aes(weight, hindfoot_length)) +
  geom_point(aes(color = genus)) +
  geom_smooth(se = FALSE) +
  labs(caption = "fig.width = 4")
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



```
ggplot(surveys_complete, aes(weight, hindfoot_length)) +
  geom_point(aes(color = genus)) +
  geom_smooth(se = FALSE) +
  labs(caption = "fig.width = 8")
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



```
ggplot(surveys_complete, aes(weight, hindfoot_length)) +
  geom_point(aes(color = genus)) +
  geom_smooth(se = FALSE) +
  labs(caption = "fig.width = 10") +
  ggsave("plot_weight_hindfoot.jpeg")
```

```
## Saving 10 x 4.5 in image
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

