

Project 4

This is the dataset you will be working with:

```
lemurs <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2021/2021-08-24/lemur_data.csv')
```

More information about the dataset can be found here:

<https://github.com/rfordatascience/tidytuesday/tree/master/data/2021/2021-08-24>

(<https://github.com/rfordatascience/tidytuesday/tree/master/data/2021/2021-08-24>) and

<https://www.nature.com/articles/sdata201419> (<https://www.nature.com/articles/sdata201419>).

Question:

Is there any association between weight (g) and age (years) of various lemur species? Does the weight depend on the sex of the various lemur species?

Introduction: Here we will be working with the `lemurs` dataset that contains information on over 200 lemurs housed at the Duke Lemur Center (from 1996-2014), such as weight, gender, number of offspring, age etc. There are total 54 variables and 82,609 observations in this dataset. To answer the first question we will work with three variables: `weight_g` (weights of lemurs given in grams, data type: `dbl`), `age_at_wt_y` (age in years when weight was taken, data type: `dbl`) and `taxon` (taxonomic code, data type: `chr`).

For the second question we will work with three variables: `weight_g` (weights of lemurs given in grams, data type: `dbl`), `taxon` (taxonomic code, data type: `chr`) and `sex` (M = male, F = female, ND = not determined, data type: `chr`).

Approach:

To answer the first question we fitted a linear model on weight wrt age (years) and then applied `glance()` to provide model-wide summary estimates in tidy format, and finally unnested and saved it in `lemur_summary`. We plotted age and weight as a scatter plot faceted over `taxon` to view an association trend. This gave us the opportunity to compare the trends between different taxons.

For the second question, we simply plotted weight and sex as a bar chart faceted over `taxon` to view any association between sex and weight across the taxons. It is helpful to view the graphs side-by-side for comparison.

Analysis:

```
# For the association between weight and age

#fit a model
lemur_summary <- lemurs %>%
  nest(data=-taxon) %>%
  mutate(
    fit = map(data, ~lm(weight_g ~ age_at_wt_y, data=.x)),
    glance_output = map(fit, glance)
  ) %>%
  select(taxon, glance_output) %>%
  unnest(cols=glance_output)

head(lemur_summary)
```

```
## # A tibble: 6 x 13
##   taxon r.squared adj.r.squared sigma statistic  p.value    df logLik    AIC
##   <chr>   <dbl>         <dbl> <dbl>      <dbl>    <dbl> <dbl> <dbl> <dbl>
## 1 OGG     0.305           0.305   306.      492. 1.02e- 90     1 -8014. 16034.
## 2 EMON     0.198           0.198   335.      900. 7.78e-177     1 -26360. 52726.
## 3 HGG      0.0677          0.0672  250.      154. 4.00e- 34     1 -14706. 29418.
## 4 VVV      0.202           0.202  1032.      692. 4.03e-136     1 -22860. 45725.
## 5 EUL      0.344           0.344   636.      830. 3.93e-147     1 -12465. 24935.
## 6 ERUF     0.347           0.347   510.      818. 1.57e-144     1 -11771. 23547.
## # ... with 4 more variables: BIC <dbl>, deviance <dbl>, df.residual <int>,
## #   nobs <int>
```

#make a label data with the fitted model

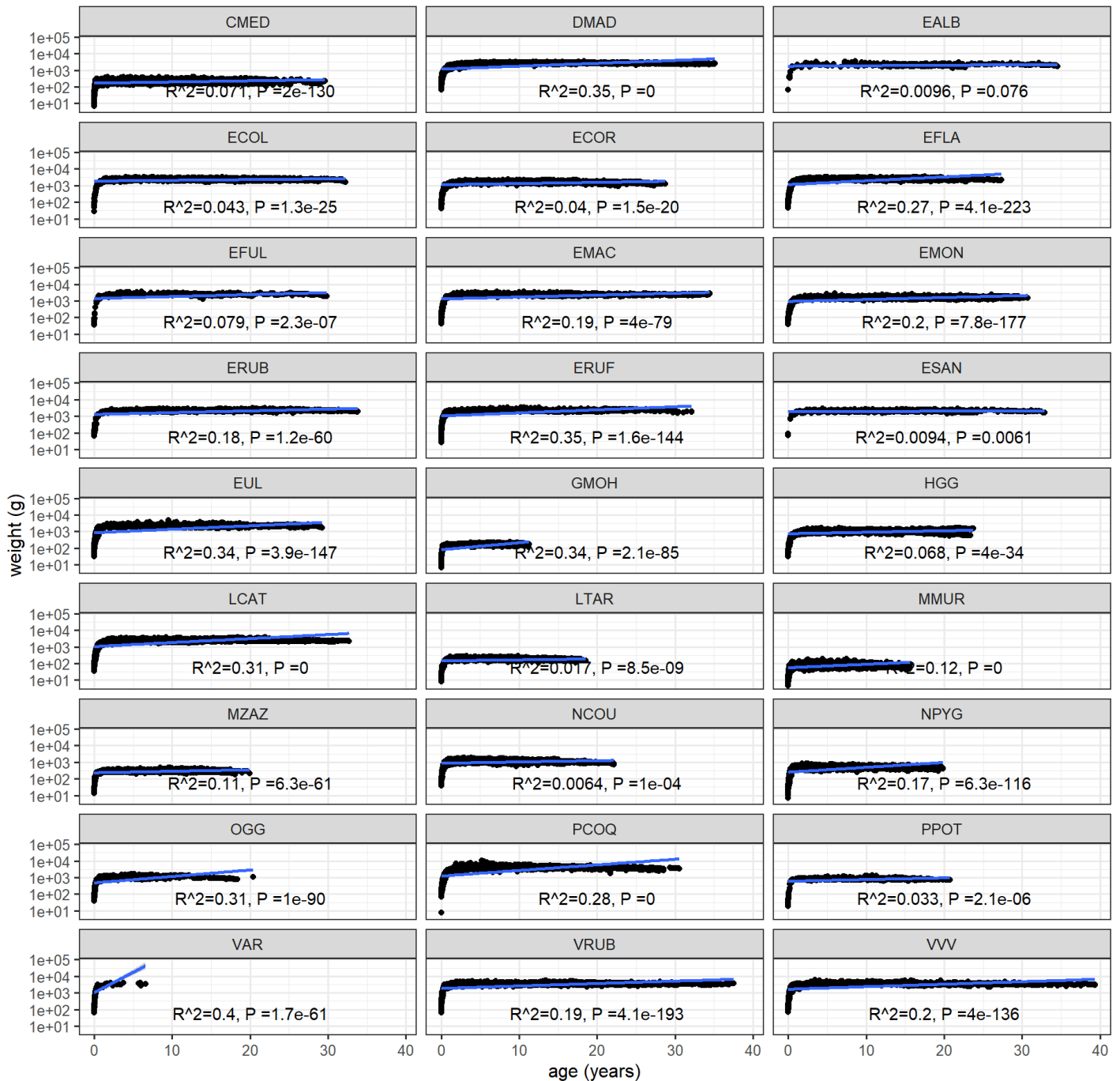
```
label_data <- lemur_summary %>%
  mutate(
    rsqr = signif(r.squared,2), # round to 2 significant digits
    pval = signif(p.value,2),
    label = glue("R^2={rsqr}, P ={pval}"),
    age_at_wt_y = 20, weight_g = 60 # label position
  ) %>%
  select(taxon, label, age_at_wt_y, weight_g)
head(label_data)
```

```
## # A tibble: 6 x 4
##   taxon label                age_at_wt_y weight_g
##   <chr> <glue>                  <dbl>    <dbl>
## 1 OGG   R^2=0.31, P =1e-90          20      60
## 2 EMON  R^2=0.2, P =7.8e-177       20      60
## 3 HGG   R^2=0.068, P =4e-34       20      60
## 4 VVV   R^2=0.2, P =4e-136       20      60
## 5 EUL   R^2=0.34, P =3.9e-147      20      60
## 6 ERUF  R^2=0.35, P =1.6e-144      20      60
```

#plot age in years vs weight

```
ggplot(lemurs, aes(age_at_wt_y, weight_g))+
  geom_point(na.rm=TRUE)+
  scale_y_log10(name = "weight (g)")+
  scale_x_continuous(name = "age (years)")+
  geom_text(
    data= label_data, aes(label=label),
  )+
  geom_smooth(method="lm",
    formula= y~x,
    na.rm = TRUE)+
  facet_wrap(~taxon,
    ncol=3)+
  ggtitle("Weight distribution of various taxonomy lemurs wrt age")+
  theme_bw(12)
```

Weight distribution of various taxonomy lemurs wrt age



```
#association between sex and weight of various lemur species
ggplot(lemurs, aes(sex, weight_g, fill = sex))+
  geom_col(na.rm = TRUE)+
  scale_y_log10(name="weight(g)")+
  facet_wrap(~taxon)+
  ggtitle("Weight distribution of various taxonomy lemurs wrt sex")+
  theme_classic(12)
```

Weight distribution of various taxonomy lemurs wrt sex



Discussion:

When we look at the linear model fit plot between age and weight of various taxon of lemurs, we mostly see a slight positive association between the variables across all the taxons. So, as the age increases, the weight increases in lemurs. The VAR taxon has a compact dataset between age 0-6 years (it is the youngest lemur species in the dataset) and the trendline looks skewed for it.

From the bar plots we can see weight is associated to sex differently for different taxons. For DMAD, ESAN, HGG, LCAT, VRUB taxons male lemurs weighed less. For ECOL, EFLA, EMAC, EMON, ERUF, EUL, MMUR, VVV taxons female lemurs weighed less. For the rest of the taxons the difference between sexes was very marginal.

From the plots we observed one more interesting fact that some taxons have data for only very young animals, such as VAR, GMOH (both <10 years of age). We can do more analysis to see if the longevity of lemurs varies over various taxons in future.