Project 4

This is the dataset you will be working with:

```
lemurs <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/20

lemurs
```

```
## # A tibble: 82,609 x 54
##    taxon dlc_id hybrid sex   name     curre~1 stud_~2 dob        birth~3 estim~4
##    <chr> <chr>  <chr>  <chr> <chr>    <chr>   <chr>   <date>       <dbl> <chr>
##  1 OGG   0005   N      M     KANGA    N       <NA>    1961-08-25       8 <NA>
##  2 OGG   0005   N      M     KANGA    N       <NA>    1961-08-25       8 <NA>
##  3 OGG   0006   N      F     ROO      N       <NA>    1961-03-17       3 <NA>
##  4 OGG   0006   N      F     ROO      N       <NA>    1961-03-17       3 <NA>
##  5 OGG   0009   N      M     POOH BE~ N       <NA>    1963-09-30       9 <NA>
##  6 OGG   0009   N      M     POOH BE~ N       <NA>    1963-09-30       9 <NA>
##  7 OGG   0009   N      M     POOH BE~ N       <NA>    1963-09-30       9 <NA>
##  8 OGG   0010   N      M     EEYORE   N       <NA>    1964-05-20       5 <NA>
##  9 OGG   0010   N      M     EEYORE   N       <NA>    1964-05-20       5 <NA>
## 10 OGG   0014   N      F     ROOLETTE N       <NA>    1964-10-27      10 <NA>
## # ... with 82,599 more rows, 44 more variables: birth_type <chr>,
## #   birth_institution <chr>, litter_size <dbl>, expected_gestation <dbl>,
## #   estimated_concep <date>, concep_month <dbl>, dam_id <chr>, dam_name <chr>,
## #   dam_taxon <chr>, dam_dob <date>, dam_age_at_concep_y <dbl>, sire_id <chr>,
## #   sire_name <chr>, sire_taxon <chr>, sire_dob <date>,
## #   sire_age_at_concep_y <dbl>, dod <date>, age_at_death_y <dbl>,
## #   age_of_living_y <dbl>, age_last_verified_y <dbl>, ...
```

Question: In this project, we attempt to answer two questions related to variation in body mass across age categories among White-fronted brown lemurs and Black-and-white ruffed lemurs (EALB and VVV) and the variation of weight w.r.t days before their respective deaths. They are:

- Is there a difference in body weight between White-fronted brown lemurs and Black-and-white ruffed lemurs across different age categories?
- What is the relationship between body weight and remaining days before death among these sub species?

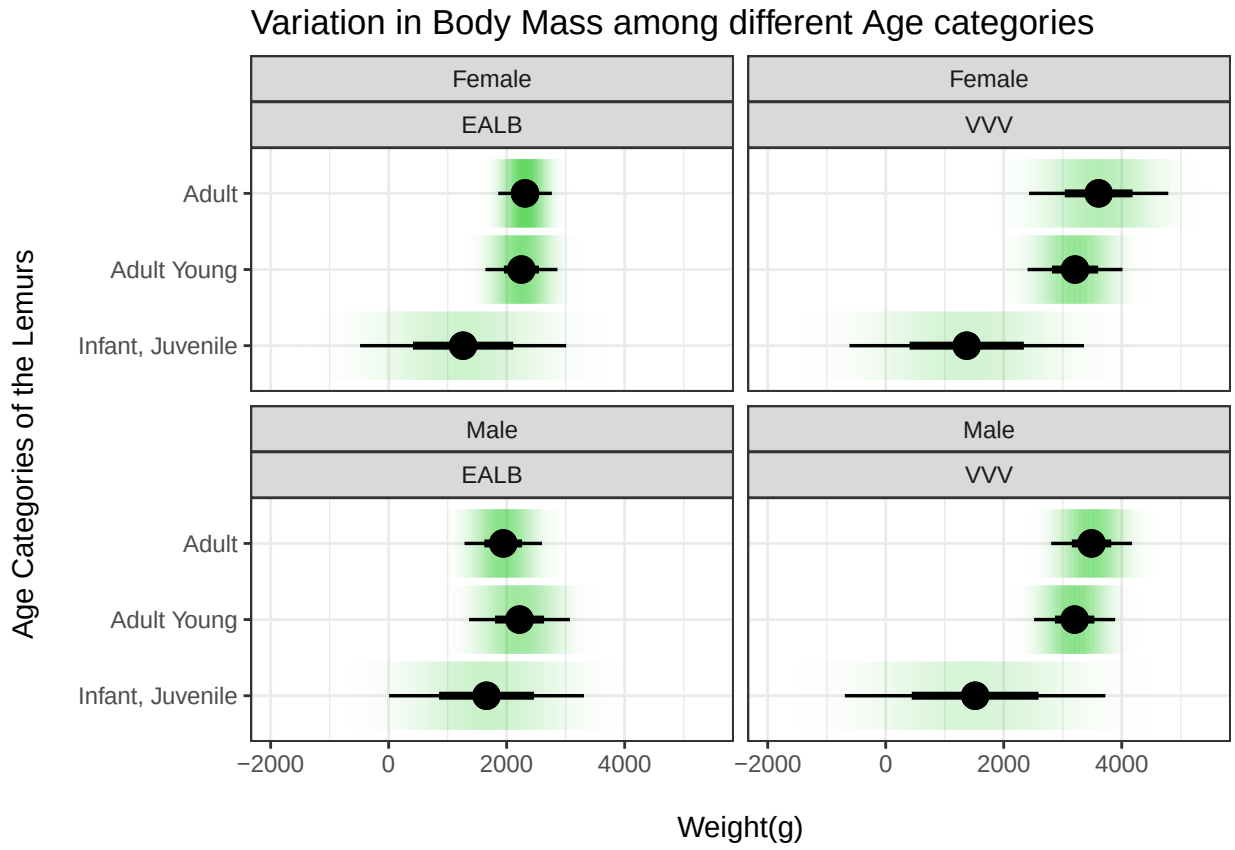
Introduction: The Duke Lemur Center has collected a comprehensive dataset called `Lemurs`, consisting of detailed information on 3,627 individuals from 27 taxonomic groups. The dataset includes 54 columns of information, such as the ID, taxon, name, birth date and location, parental details, body mass, derived metrics, and reproduction information. There are 82,609 rows, with each row representing an observation of a lemur. An observation is considered as a weight measurement on a particular day, and an individual may have multiple observations with all other column values remaining the same.
We will examine the gender (column `sex`), the age category of the lemur (column `age_category`), and the weight (column `weight_g`) for White-fronted brown lemurs and Black-and-white ruffed lemurs. In answering the second question, we will investigate `weight_g` and `days_before_Death` in both subspecies. It is important to include all rows that meet the relevant filters since each row represents a unique record in time.

Approach: To address the first question, we need to perform some data wrangling. Specifically, we will filter the data by EALB and VVV, as well as by `gender`, and select certain columns including `taxon`, `sex`, `age_category`, and `weight_g`. We will then use the `case_when()` function to recode `age_category` into sub categories and M/F into Male/Female. As we are analyzing weight measurements from a sample, we will report uncertainty in a certain form. It will help us compare if there are any significant differences in body mass. We will use a gradient interval to analyze the confidence intervals.
To address the second inquiry, we commence by filtering the dataset based on two sub-species. We then choose to focus on the variables `taxon`, `weight_g`, and `days_before_death` Using the `map()` function, we construct a nested table to fit a linear model of `weight_g` versus `days_before_death` To visually examine the trends, we create a scatter plot, where the `geom_point()` function plots the points, and the regression line is displayed using `geom_smooth()`. Model details obtained using `glance()` are also added i.e, R^2 Value. To improve the visualization, we reverse the x-axis as the "days before death" declines with time.

Analysis:

```
lemurs %>%
  filter(
    taxon %in% c("EALB", "VVV"), sex %in% c("M", "F")) %>%
    select(taxon, sex, age_category, weight_g) %>%
      mutate(
        age = case_when(
          taxon == "EALB" & age_category %in% c("young_adult") ~ "Adult Young",
          taxon == "EALB" & age_category %in% c("adult") ~ "Adult",
          taxon == "EALB" & age_category %in% c("IJ") ~ "Infant, Juvenile",
          taxon == "VVV" & age_category %in% c("young_adult") ~ "Adult Young",
          taxon == "VVV" & age_category %in% c("adult") ~ "Adult",
          taxon == "VVV" & age_category %in% c("IJ") ~ "Infant, Juvenile",
          TRUE ~ NA_character_
        ),
        # Code the age category as specified in the Github link
        # for each sub species
        sex = case_when(sex == "M" ~ "Male",
                        sex == "F" ~ "Female")
        # Detailed names for sex
      ) %>%
      group_by(taxon, sex, age) %>%
      summarize(
        mu = mean(weight_g),
        # Get the mean weight from all measured weights
        sigma = sd(weight_g),
        # Similar for standard deviation
        .groups = "drop"
      ) %>%
```

```
ggplot() +
aes(x = mu, y = fct_rev(age)) +
stat_dist_gradientinterval(
  aes(dist = dist_normal(mu = mu, sigma = sigma)),
  point_size = 4,
  fill = "limegreen",
  fill_type = "segments"
) +
facet_wrap(vars(sex, taxon), nrow = 2) +
labs(title = "Variation in Body Mass among different Age categories",
 x = "\nWeight(g)",
 y = "Age Categories of the Lemurs\n") +
theme_bw(11)
```

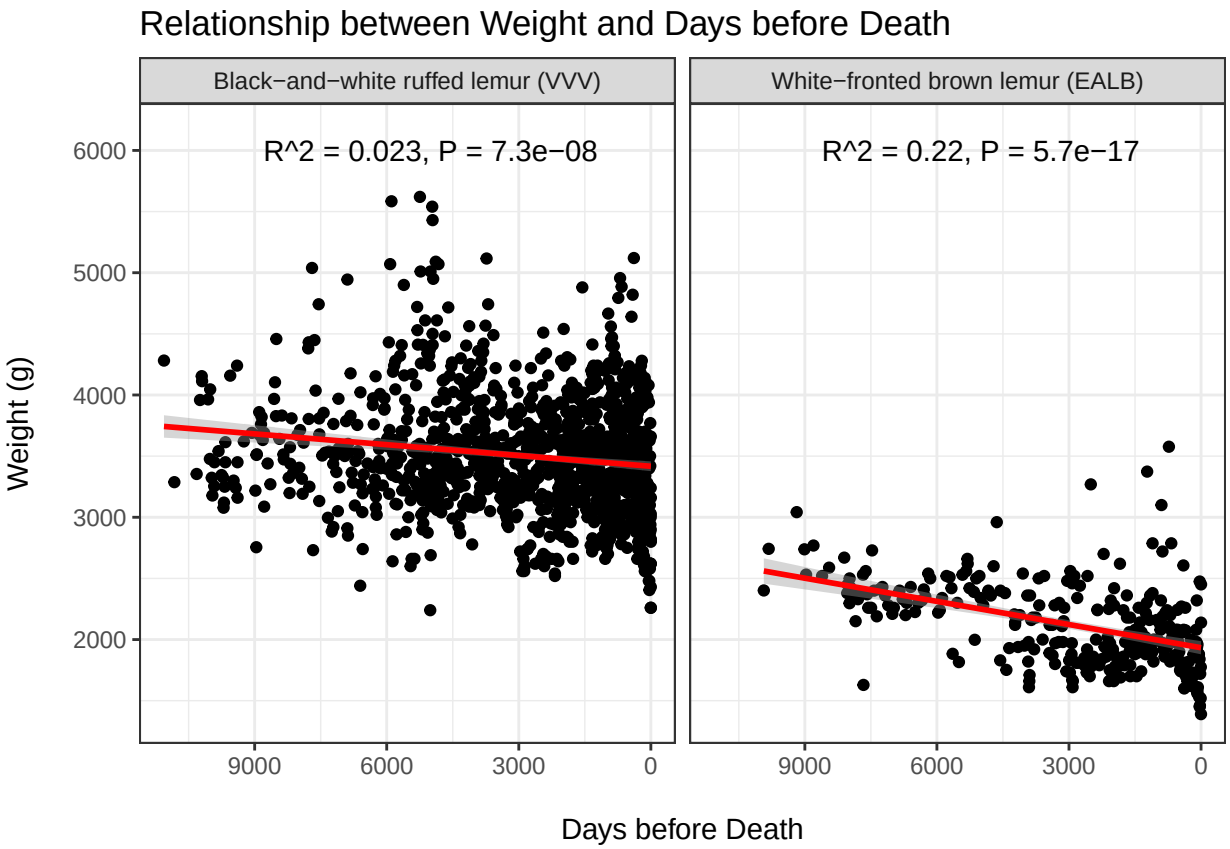


```r
# Get the necessary data
lm_data <- lemurs %>%
  filter(age_category %in% c("adult"),
         taxon %in% c("EALB", "VVV")) %>%
  mutate(
    x_taxon = case_when(
      taxon == "EALB" ~ "White-fronted brown lemur (EALB)",
      taxon == "VVV" ~ "Black-and-white ruffed lemur (VVV)"
    )
  ) %>%
  select(x_taxon, sex, weight_g, days_before_death)

# Fit the data and extract model information
lm_summary <- lm_data %>%
  nest(data = -x_taxon) %>%
  mutate(fit = map(data, ~ lm(weight_g ~ days_before_death, data = .x)),
         glance_out = map(fit, glance)) %>%
  select(x_taxon, glance_out) %>%
  unnest(glance_out)

#Extract relevant model information to be plotted
label_data <- lm_summary %>%
  mutate(
    rsqr = signif(r.squared, 2),
    pval = signif(p.value, 2),
    label = glue("R^2 = {rsqr}, P = {pval}"),
    days_before_death = 5000,
    weight_g = 6000
  ) %>%
  select(x_taxon, label, days_before_death, weight_g)

ggplot(lm_data, aes(days_before_death, weight_g)) +
  geom_point() +
  geom_smooth(method = "lm",
              formula = y ~ x,
              color = "red") +
  geom_text(data = label_data, aes(label = label)) +
  scale_x_reverse() +
  labs(title = "Relationship between Weight and Days before Death",
       x = "\nDays before Death",
       y = "Weight (g)\n") +
  facet_wrap(vars(x_taxon)) +
  theme_bw(11)
```



Discussion:

Plot1 - Variation in Body Mass vs Age Category:

When we look at the variation of body mass among taxon VVV, we see that mean body mass for adults is slightly higher for females compared to the males. While mean body mass for the sexes are approximately the same for Young Adults and Infant, Juveniles. The body mass for adults has more variance for female compared to male. Among EALB, Male adults have lesser mean body mass compared to female adults. Hence, we can conclude that female adult lemurs of both subspecies have a higher mean body mass compared to adult males.

Plot2 - Relationship between Weight and Days before Death:

Intuitively it is known that while nearing death, body mass reduces significantly compared to early years. This plot not only helps to study that relationship quantitatively, but also analyze the strength of said relationship. It can be observed that among EALB, we see a R^2 value of 0.22 demonstrating a strong relationship between the two variables.

The variance is also much smaller and we can observe a decreasing trend in body mass as death nears the lemur. However for VVV, we see a R^2 value of 0.023 indicating a weak positive relationship between the variables. The variance in the early days before death is also small. Despite weak positive relationship, VVV also shows a decreasing trend in weight as death approaches (for lemurs).