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')

lemurs
```

```
## # A tibble: 82,609 × 54
##    taxon dlc_id hybrid sex   name       current_resident stud_book dob
##    <chr> <chr>  <chr>  <chr> <chr>      <chr>            <chr>     <date>
##  1 OGG   0005   N      M     KANGA      N                <NA>      1961-08-25
##  2 OGG   0005   N      M     KANGA      N                <NA>      1961-08-25
##  3 OGG   0006   N      F     ROO        N                <NA>      1961-03-17
##  4 OGG   0006   N      F     ROO        N                <NA>      1961-03-17
##  5 OGG   0009   N      M     POOH BEAR  N                <NA>      1963-09-30
##  6 OGG   0009   N      M     POOH BEAR  N                <NA>      1963-09-30
##  7 OGG   0009   N      M     POOH BEAR  N                <NA>      1963-09-30
##  8 OGG   0010   N      M     EEYORE     N                <NA>      1964-05-20
##  9 OGG   0010   N      M     EEYORE     N                <NA>      1964-05-20
## 10 OGG   0014   N      F     ROOLETTE   N                <NA>      1964-10-27
## # ℹ 82,599 more rows
## # ℹ 46 more variables: birth_month <dbl>, estimated_dob <chr>,
## #   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>, …
```

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: Does the age of a lemur's mother when she conceives affect the lemur's lifespan? If so, how strong is this relationship, and does it change depending on the number of siblings the lemur has?

Introduction:

Lemurs are fascinating primates that are native to the island of Madagascar. They are known for their large eyes, long tails, and playful nature. However, there is still much that we do not know about lemurs, including how their lifespan is affected by their environment and genetics.By understanding how these factors affect a lemur's lifespan, we can learn more about the biology and conservation of the most threatened group of mammals on the planet.

This project will utilize a dataset named as 'lemur' containing information about lemurs housed at the Duke Lemur Center. The dataset encompasses life history records for all individuals that resided at the center from 1966 to 2014, serving as a valuable resource for lemur conservation efforts. Despite its extensive size of over 82,000

records and 54 columns, we will focus on a subset of the dataset containing the following fields: dlc_id (animal identification), sex, name, date of birth (dob), date of death (dod), age at death, maximum age attained, birth type, litter size, dam's identification (dam_id), dam's name, and dam's age at conception.

In this project, we will investigate the relationship between the age of a lemur's mother when she conceives and the lemur's lifespan. We will also examine whether this relationship is affected by the number of siblings the lemur has.

Key fields that will aid in addressing the research question include:

-*age_at_death_y*: The age of the animal at the time of their verifiable death.

-*litter_size*: The total number of infants in the litter the focal animal was born into, including itself.

-*dam_age_at_concep_y*: The estimated age of the female parent at the time of conception of the focal animal.

Approach:

Before embarking on our analysis, we must first clean the dataset. This involves selecting the necessary fields from the original dataset and removing records with missing values or "n/a" entries. The resulting cleaned dataset will be named "lemur_filtered."

To address the first part of the research question, we will employ a straightforward linear regression approach using the 'lm' function in R. This function allows us to model the relationship between the variables *age_at_death_y* and *dam_age_at_concep_y* using the formula Y ~ X, where Y is the outcome variable (age at death) and X is the predictor variable (dam's age at conception).By applying the 'summary' function, we can obtain summary statistics, including coefficients, intercept, and slope information. This will enable us to determine the strength and direction of the association between these two variables in the filtered dataset.

To address the second part of the research question, we delve deeper into the analysis. Our goal is to determine the point estimates, including statistics, p-values, and standard deviations, for the fitted model corresponding to each *litter_size.* We will utilize functions we have already employed, such as *lm* and *summary*, along with additional functions like *map*, *fit*, *nest*, and *unnest*. This comprehensive approach will facilitate the extraction of point estimates for the predictive model based on the *litter_size* variable.

Visualization serves as the final step of our analysis, culminating in a scatter plot for each research question. Scatter plots, which represent values for two different numeric variables using dots, are employed to visualize the relationships between variables. Trend or regression lines within these plots further enhance our understanding of the variables' associations.

To address the second question, we utilize the 'facet' approach, which partitions a plot into a matrix of panels. Each panel displays a distinct subset of the data based on the 'litter_size' variable. This visualization technique effectively illustrates the relationship between 'dam_age_at_concep_y' and 'age_at_death_y' across different litter sizes.

Analysis:

Firstly,we must extract the required fields from original data set and clean the same.

```
lemur_filtered <- lemurs %>%
  select(dlc_id, sex, name, dob,dod,age_at_death_y,age_max_live_or_dead_y,birth_type,litter_siz
e,dam_id,dam_name,dam_age_at_concep_y)  %>% #selection of required fields
  filter(!is.na(dob), !is.na(age_at_death_y)) %>% # Filter rows w/o DOB and age at death
  filter(!is.na(litter_size))%>%
  filter(!is.na(dam_age_at_concep_y)) %>%
  filter(!is.na(age_max_live_or_dead_y)) %>%
  filter(!is.na(age_at_death_y)) %>%
  distinct(dlc_id, .keep_all=TRUE)

lemur_filtered
```

```
## # A tibble: 1,181 × 12
##    dlc_id sex   name      dob        dod        age_at_death_y
##    <chr>  <chr> <chr>     <date>     <date>              <dbl>
##  1 0005   M     KANGA     1961-08-25 1977-02-07           15.5
##  2 0006   F     ROO       1961-03-17 1974-10-15           13.6
##  3 0009   M     POOH BEAR 1963-09-30 1974-02-13           10.4
##  4 0010   M     EEYORE    1964-05-20 1977-11-02           13.5
##  5 0026   M     BARNABY   1971-06-10 1973-02-07            1.67
##  6 0028   F     TOOTSIE   1972-10-04 1973-06-11            0.68
##  7 0270   F     LILLO     1971-09-10 1972-07-01            0.81
##  8 0277   M     CARLOS    1975-05-02 1978-12-13            3.62
##  9 0417   M     HERMES    1968-09-16 1969-06-02            0.71
## 10 0418   M     ALPHABET  1969-06-02 1972-03-24            2.81
## # i 1,171 more rows
## # i 6 more variables: age_max_live_or_dead_y <dbl>, birth_type <chr>,
## #   litter_size <dbl>, dam_id <chr>, dam_name <chr>, dam_age_at_concep_y <dbl>
```

Then, we will employ a straightforward linear regression approach using the 'lm' function.This will be supportive to tackle the first part of research question.

```
lm_out = lm(age_at_death_y ~ dam_age_at_concep_y, data = lemur_filtered)
summary(lm_out)
```

```
##
## Call:
## lm(formula = age_at_death_y ~ dam_age_at_concep_y, data = lemur_filtered)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.0295  -8.5980  -0.5343   6.6090  24.3291
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           10.68471    0.46482  22.987   <2e-16 ***
## dam_age_at_concep_y    0.06056    0.05895   1.027    0.304
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.886 on 1179 degrees of freedom
## Multiple R-squared:  0.0008943,  Adjusted R-squared:  4.69e-05
## F-statistic: 1.055 on 1 and 1179 DF,  p-value: 0.3045
```

Based on the outcome of above statistics,we can say that there is an association between these two variables.We will be able to understand much clear from the associated plots given in the end of analysis.

Now,Let us obtain the information about the model fits for each litter size.This will be helpful to research on the second part of the question .

```
library(tidyr)
library(broom)

lm_summary <- lemur_filtered %>%
  nest(data = -litter_size) %>% # nest all data except one column
  mutate(
    # apply linear model to each nested data frame
    fit = map(data, ~lm(age_at_death_y ~ dam_age_at_concep_y, data = .x)),
    glance_out = map(fit, glance)
  )%>%
  select(litter_size, glance_out) %>%
  unnest(cols = glance_out)

lm_summary
```

```
## # A tibble: 4 × 13
##   litter_size r.squared adj.r.squared sigma statistic p.value      df logLik   AIC
##         <dbl>     <dbl>         <dbl> <dbl>     <dbl>    <dbl>  <dbl>   <dbl> <dbl>
## ## 1           1 0.00494       0.00324  8.95      2.91   0.0884      1  -2126. 4257.
## ## 2           2 0.0234        0.0207   8.93      8.56   0.00365     1  -1294. 2595.
## ## 3           4 0.0000163    -0.0204   6.96      0.000801 0.978     1   -170.  347.
## ## 4           3 0.0256        0.0202   8.65      4.72   0.0311      1   -650. 1306.
## # ℹ 4 more variables: BIC <dbl>, deviance <dbl>, df.residual <int>, nobs <int>
```

We have gathered enough information to make the plots.Label data can be prepared as below to make the plots much easier with info on regression lines.

```
library(glue)  # for easy text formatting
label_data <- lm_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_death_y = 40, dam_age_at_concep_y = 10# label position in plot
  ) %>%
  select(litter_size, label, age_at_death_y,dam_age_at_concep_y)
label_data
```

```
## # A tibble: 4 × 4
##    litter_size label                        age_at_death_y dam_age_at_concep_y
##          <dbl> <glue>                                <dbl>               <dbl>
## 1            1 R^2 = 0.0049, P = 0.088                  40                  10
## 2            2 R^2 = 0.023, P = 0.0036                  40                  10
## 3            4 R^2 = 1.6e-05, P = 0.98                  40                  10
## 4            3 R^2 = 0.026, P = 0.031                   40                  10
```

Finally ,It is time to make plots.

Below plot will clearly explain the relationship between the the age of a lemur's mother when she conceives affect the lemur's lifespan.

```
#dam_age_at_concep_y vs age_max_live_or_dead_y

ggplot(lemur_filtered, aes(dam_age_at_concep_y,age_at_death_y))+
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  scale_y_continuous(
    name = "Maximum Life Span (in years)"
  )+
   scale_x_continuous(
    name = "Age of female parent at conception(in years)"
  )+
  ggtitle("Model Fit: Lemur's Longevity and mother's age at conception")
```

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

Model Fit: Lemur's Longevity and mother's age at conception
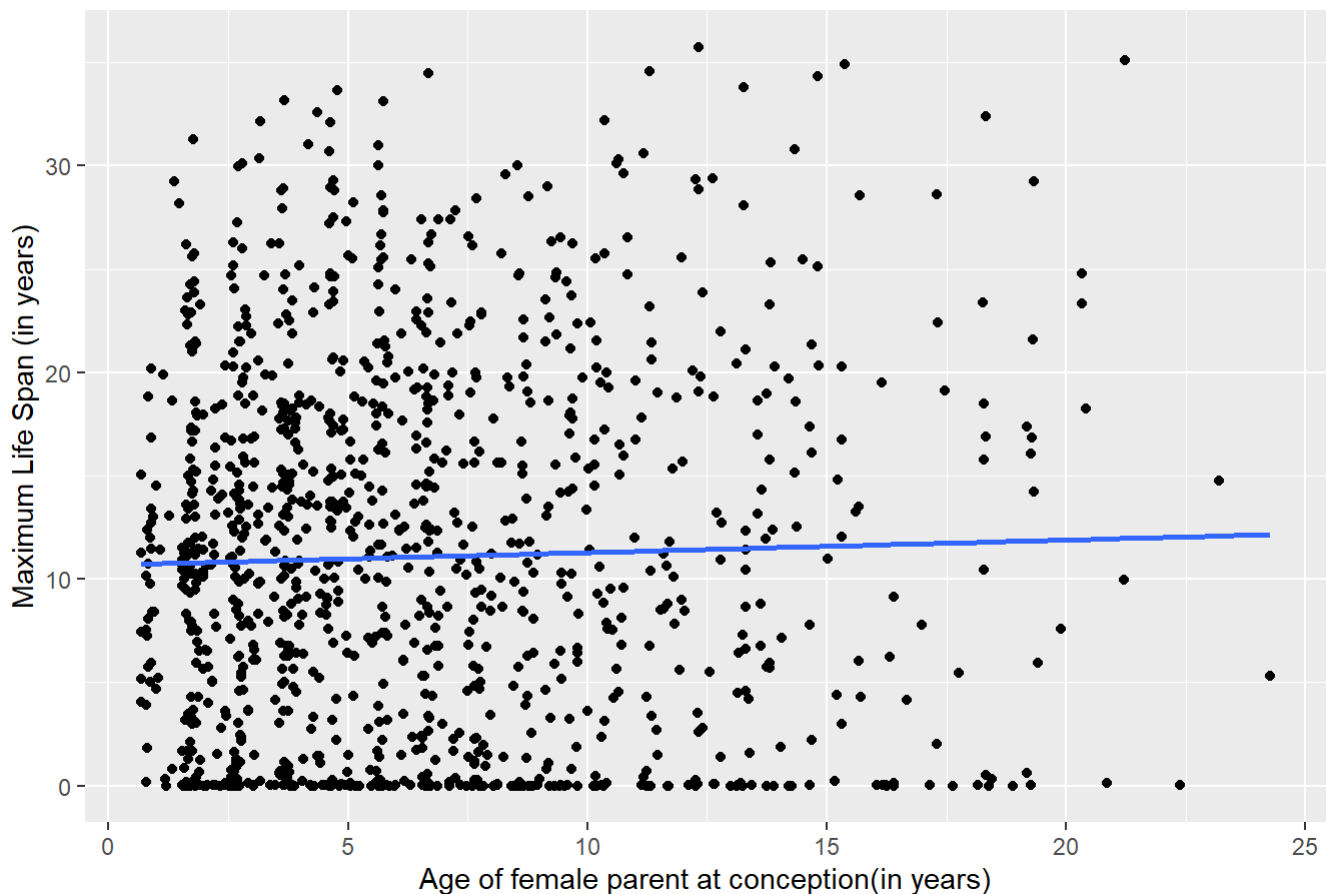
Variation in relationship between Lemur's longevity and mother's age at conception for each of the litter size can be ploted using facet approach as below.This would assist us to answer second part of the question.

```
ggplot(lemur_filtered, aes(dam_age_at_concep_y,age_at_death_y))+ geom_point() +
  geom_text(
    data = label_data, aes(label = label),
    size = 10/.pt, hjust = 1  # 10pt, right-justified
  ) +
  geom_smooth(method = "lm", se = FALSE)+
   scale_y_continuous(
    name = "Maximum Life Span (in years)"
  )+
   scale_x_continuous(
    name = "Age of female parent at conception(in years)"
  )+
  ggtitle("Influence of Litter Size-Female Parent Age Vs Lemur Lifespan")+
 facet_wrap(vars(litter_size))
```

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

Influence of Litter Size-Female Parent Age Vs Lemur Lifespan
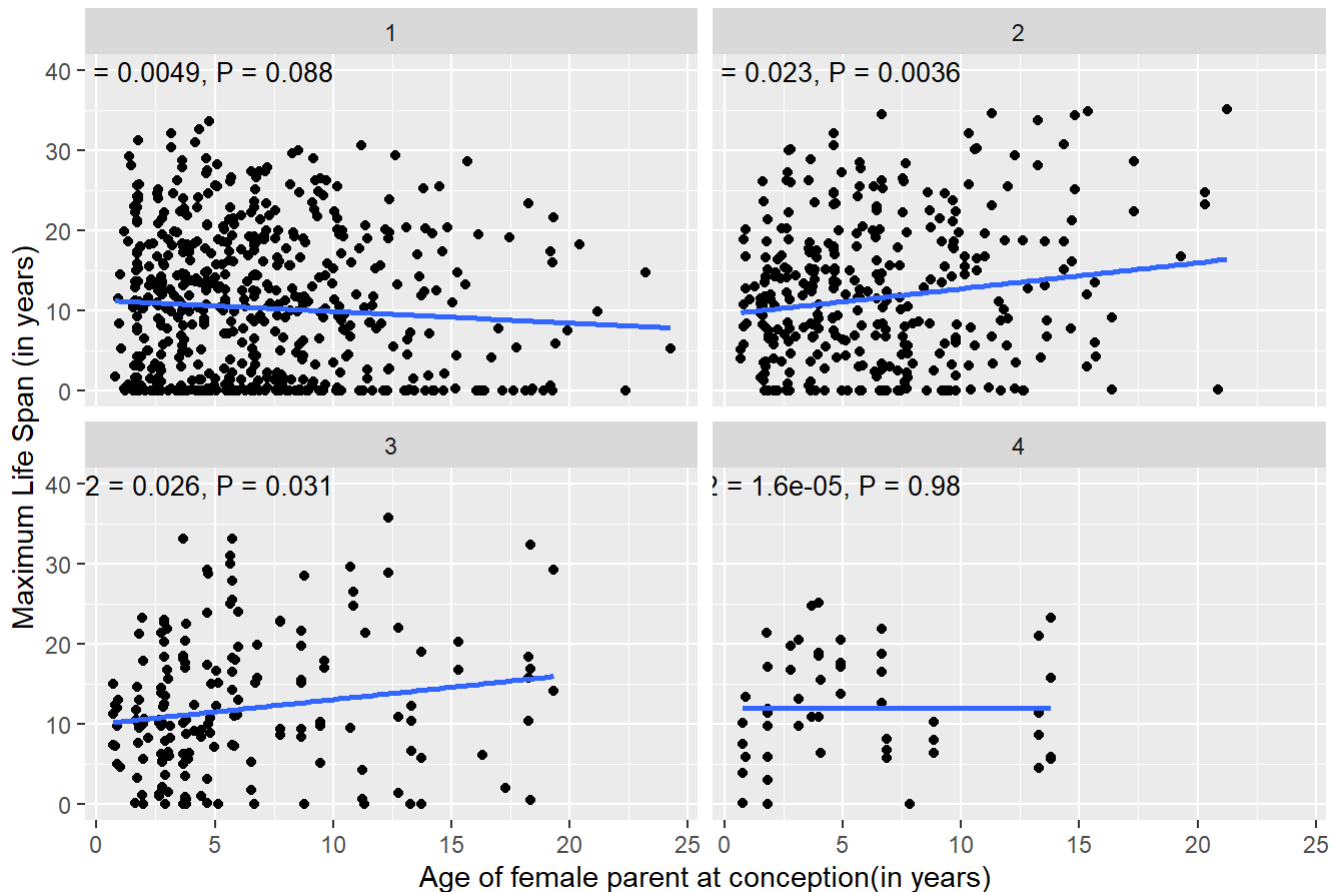
Discussion:

From the analysis above ,It is clear that the influence of female parent age on lemur lifespan is subtle yet consistently positive. While the impact is not statistically significant, it appears that lemurs born to older mothers tend to live slightly longer than those born to younger ones.The scatter plot reveals a concerning trend of higher mortality rates among younger lemurs. This suggests that lemur mortality is not evenly distributed throughout their lifespan, with a greater risk of death occurring in the early years.

As female lemurs age, their conception rate diminishes significantly. This biological decline in fertility is consistent with observations across various species.The data indicates that female lemurs can retain their reproductive capacity until around 25 years of age, though the maximum life expectancy is approximately 40 years.The regression line's near-parallelism to the x-axis implies that the relationship between female parent age and lemur lifespan exhibits an extremely low growth rate. Extremely less p-value implies that the impact of age on lifespan is relatively small.

The analysis of lemur lifespan takes a more intriguing turn when we consider the statistical estimates for each litter size. The faceted scatter plots effectively reveal the variation in the trend across different litter size panels.

The distribution of lemurs across litter sizes reveals that a significant portion are born with no siblings or one sibling. Panels 1 and 2 illustrate this concentration, highlighting the prevalence of these litter sizes.Interestingly, a negative trend emerges for lemurs born with no siblings or four siblings, including themselves. This indicates that lemurs born as singletons or with three or more siblings tend to live shorter lives. Conversely, a positive trend surfaces for lemurs born with two siblings, suggesting that those born with one sibling live longer.

In summary, the age of the female parent at conception significantly impacts lemur longevity. While lemurs born with no siblings or four siblings tend to live shorter lives, those born with one or two siblings exhibit a longer lifespan. This result underscores the complex interplay of litter size and maternal age in influencing lemur

longevity.