

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 x 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
## # ... with 82,599 more rows, and 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>, dod <date>,
## #   age_at_death_y <dbl>, age_of_living_y <dbl>, age_last_verified_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: Do we believe that the weight gain of wild-born, pure-species (not hybrids) has improved over time? Or, to phrase the question another way, do we believe that the **Duke Lemur Center** (DLC) project been a success thus far in improving the *natural* growth of *each species*' weight over time?

Introduction: To answer this question, we will use the **Lemurs** dataset provided directly from the DLC project. The dataset includes over eighty thousand data points on over forty thousand individual lemurs of various species dating all the way back to 1966. The crucial columns we will need from the dataset are the lemurs' unique identifiers (*dlc_id*), their estimated weight (*weight_g*), whether they are a hybrid (*hybrid*), their age at time of recording (*age_at_wt_d*), and whether their birth was natural or in captivity (*birth_type*).

Approach: Because we are specifically concerned with individual species, we will exclude all hybrids. We will produce a subset of data of the lemurs segregated by whether they were natural born or not. To do this, we will first remove all data points with invalid data on any of the five designated columns. We will then filter out all hybrids from the dataset. We can then capture the weight and ages from the dataset to plot.

This process will of course be mapped to each birth type, as we want to compare the rate of growth between captive and wild births. Using linear regressions on each, we may compare the growths of each over time.

Analysis:

```
lemurs <- lemurs %>%
  filter(!is.na(dlc_id),
         !is.na(weight_g),
         !is.na(age_at_wt_d),
         !is.na(birth_type),
         hybrid == "N",
         birth_type != "Unk")

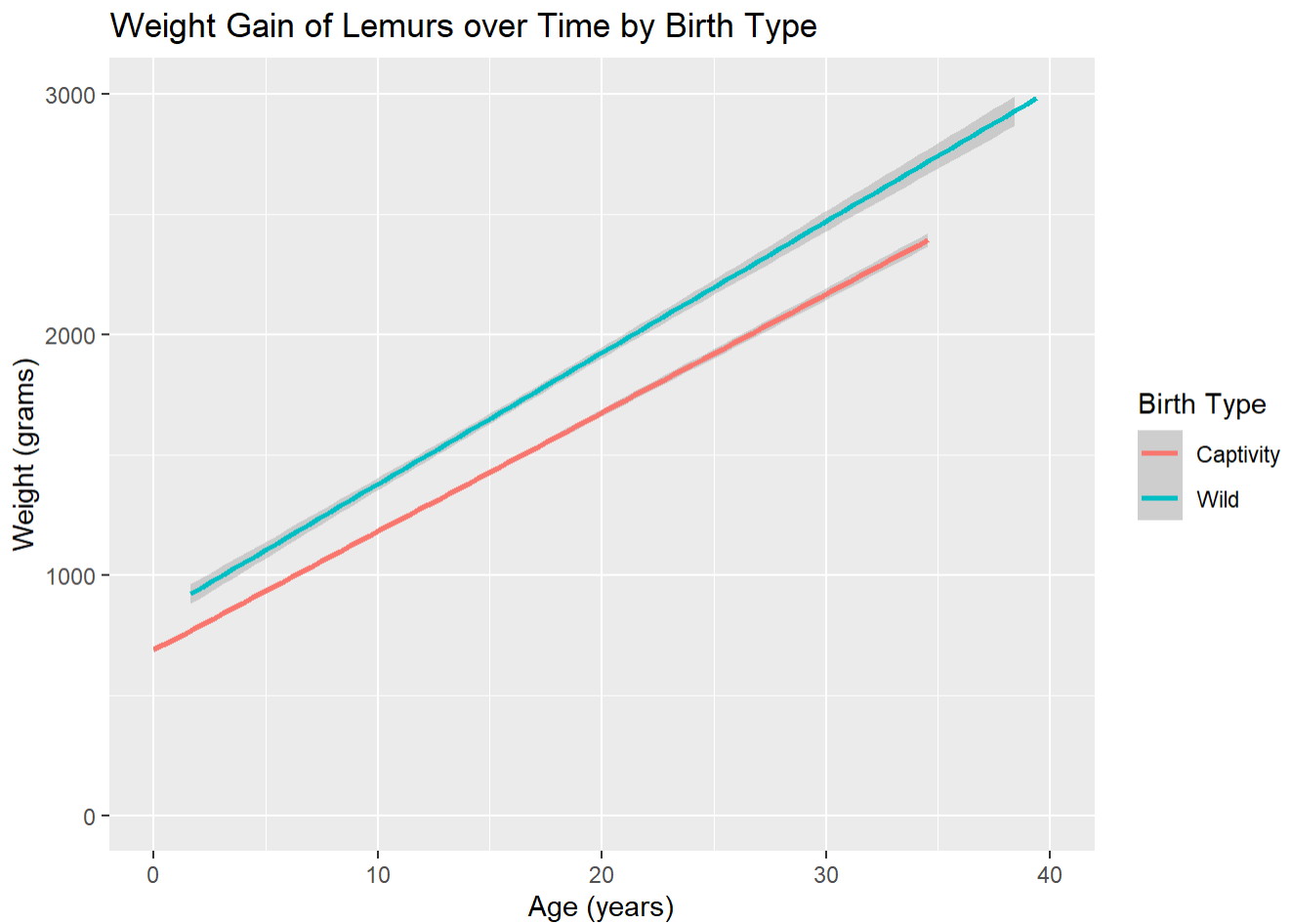
lemurs_lm <- lemurs %>%
  nest(data = -birth_type) %>%
  mutate(
    fit = map(data, ~lm(weight_g ~ age_at_wt_d, data = .x)),
    glance_out = map(fit, glance)
  ) %>%
  select(birth_type, glance_out) %>%
  unnest(cols = glance_out)

lemurs_lm
```

```
## # A tibble: 2 x 13
##   birth_type r.squared adj.r.squared sigma statistic p.value    df logLik
##   <chr>      <dbl>      <dbl> <dbl>    <dbl>    <dbl> <dbl>  <dbl>
## 1 CB        0.0947      0.0947 1259.    7642. 0      1 -624924.
## 2 WB        0.115      0.115  1109.    963. 7.65e-199 1 -62377.
## # ... with 5 more variables: AIC <dbl>, BIC <dbl>, deviance <dbl>,
## #   df.residual <int>, nobs <int>
```

```
lemurs %>%
  ggplot(aes(age_at_wt_d, weight_g, color=birth_type)) +
  geom_smooth(method="lm", formula = y ~ x) +
  scale_x_continuous(
    limits = c(0, 14600),
    breaks = c(0, 3650, 7300, 10950, 14600),
    labels = c("0", "10", "20", "30", "40")
  ) +
  scale_y_continuous(
    limits = c(0, 3000)
  ) +
  scale_color_discrete(
    labels = c("Captivity", "Wild")
  ) +
  labs(
    x = "Age (years)",
    y = "Weight (grams)",
    title = "Weight Gain of Lemurs over Time by Birth Type",
    color = "Birth Type"
  )
)
```

```
## Warning: Removed 11195 rows containing non-finite values (stat_smooth).
```



Discussion: By the fitment summary table shown first, we can see that the **r.squared** result from the *captive* born is less than the *wild* born. By the plot shown below the table, is is shown that while the wild born growth is slightly higher, it is almost negligible. It is however worth noting that wild born lemurs do tend to be heavier as they have a taller intercept. It could be surmised that wild captive born lemurs do grow and are healthy, wild born are generally healthier as they have likely enjoy their natural diet.