Homework 7

Jhanani Ramkumar

2026-01-27

Instructions

Answer the following questions and/or complete the exercises in RMarkdown. Please embed all of your code and push the final work to your repository. Your report should be organized, clean, and run free from errors. Remember, you must remove the `#` for any included code chunks to run.

Load the tidyverse

```
library("tidyverse")
library("janitor")
```

Data

For this assignment, we will use data from a study on elephants and the effects of poaching on tusk size.

Reference: Chiyo, Patrick I., Vincent Obanda, and David K. Korir. "Illegal tusk harvest and the decline of tusk size in the African elephant." Ecology and Evolution 5, 22: 5216–5229 (2015) (https://doi.org/10.1002/ece3.1769). Data deposited at Dryad Digital Repository (https://doi.org/10.5061/dryad.h6t7j).

1. Before starting data analysis, read the abstract of the paper to get an idea of the questions being asked. In 2-3 sentences, describe what the study is testing and the variables involved.
The study involved looking at tusk evolution in African elephants. Researchers obtained data before severe ivory harvesting and after among survivors and among elephants born during population recovery. In general, tusk length and circumference in male and female elephants decreased. This suggested that tusk sizes are hereditary.

2. Load `elephants.csv` and store it as a new object called `elephants`.

```
elephants <- read_csv("C:/Users/jhana/OneDrive/Desktop/BIS15L_W26_JhananiR/lab7/data/elephants.c
sv")
```

```
## Rows: 777 Columns: 7
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (3): Years of sample collection, Elephant ID, Sex
## dbl (4): Estimated Age (years), shoulder Height in  cm, Tusk Length in cm, T...
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

3. Clean the data by converting variable names to lowercase with no spaces or special characters.

```
elephants <- elephants %>%
  clean_names()
```

4. Use one or more of the summary functions you have learned to get an idea of the structure of the data.

```
glimpse(elephants)
```

```
## Rows: 777
## Columns: 7
## $ years_of_sample_collection <chr> "1966-68", "1966-68", "1966-68", "1966-68",…
## $ elephant_id                <chr> "12", "34", "162", "292", "11", "152", "264…
## $ sex                        <chr> "f", "f", "f", "f", "f", "f", "f", "f", "f"…
## $ estimated_age_years        <dbl> 0.080, 0.080, 0.083, 0.083, 0.250, 0.250, 0…
## $ shoulder_height_in_cm      <dbl> 102, 89, 89, 92, 133, 100, 93, 108, 108, 12…
## $ tusk_length_in_cm          <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ tusk_circumference_in_cm   <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
```

5. Use `mutate()` Change the variables `years_of_sample_collection`, `elephant_id`, and `sex` to factors. Be sure to store the output as a new dataframe and use it for the remaining questions.

```
elephants %>%
  mutate(across(c(years_of_sample_collection, elephant_id, sex), as.factor))
```

```
## # A tibble: 777 × 7
##    years_of_sample_collection elephant_id sex   estimated_age_years
##    <fct>                      <fct>       <fct>               <dbl>
##  1 1966-68                    12          f                   0.08
##  2 1966-68                    34          f                   0.08
##  3 1966-68                    162         f                   0.083
##  4 1966-68                    292         f                   0.083
##  5 1966-68                    11          f                   0.25
##  6 1966-68                    152         f                   0.25
##  7 1966-68                    264         f                   0.25
##  8 1966-68                    263         f                   0.5
##  9 1966-68                    266         f                   0.5
## 10 1966-68                    217         f                   1
## # ℹ 767 more rows
## # ℹ 3 more variables: shoulder_height_in_cm <dbl>, tusk_length_in_cm <dbl>,
## #   tusk_circumference_in_cm <dbl>
```

6. From which years were data collected? Show the sample periods below.

```
elephants %>%
  distinct(years_of_sample_collection)
```

```
## # A tibble: 2 × 1
##   years_of_sample_collection
##   <chr>
## 1 1966-68
## 2 2005-13
```

Data was collected from 1966-1968 and from 2005-2013.

7. How many males and females were sampled in this study?

```
elephants %>%
  filter(sex %in% c("f", "m")) %>%
  count(sex)
```

```
## # A tibble: 2 × 2
##   sex       n
##   <chr> <int>
## 1 f       416
## 2 m       361
```

There are 416 female elephants and 361 male elephants.

8. What is the mean, median, and standard deviation for age of males and females included in the study? Separate the results by year of sample collection. Does the sampling look even between years and sexes?

```
elephants %>%
  filter(sex == "m") %>%
  summarize(mean_estimated_age_years_m = mean(estimated_age_years, na.rm = T),
            median_estimated_age_years_m = median(estimated_age_years, na.rm = T),
            sd_estimated_age_years_m = sd(estimated_age_years, na.rm = T))
```

```
## # A tibble: 1 × 3
##   mean_estimated_age_years_m median_estimated_age_years_m sd_estimated_age_yea…¹
##                        <dbl>                        <dbl>                  <dbl>
## 1                       12.1                          8.5                   10.6
## # ℹ abbreviated name: ¹sd_estimated_age_years_m
```

```
elephants %>%
  filter(sex == "f") %>%
  select(years_of_sample_collection, estimated_age_years) %>%
  summarize(mean_estimated_age_years_f=mean(estimated_age_years, na.rm = T),
            median_estimated_age_years_f=median(estimated_age_years, na.rm = T),
            sd_estimated_age_years_f=sd(estimated_age_years, na.rm = T))
```

```
## # A tibble: 1 × 3
##   mean_estimated_age_years_f median_estimated_age_years_f sd_estimated_age_yea…¹
##                        <dbl>                        <dbl>                  <dbl>
## 1                       17.6                           16                   13.0
## # ℹ abbreviated name: ¹sd_estimated_age_years_f
```

Male elephants in this sample are mostly young given the average age is 12-13 years. And most of the male elephants aren't sexually mature yet. Female elephants in this sample are generally older given the average age of around 18 years. So, the sample over represents mature female elephants relative to mature male elephants.

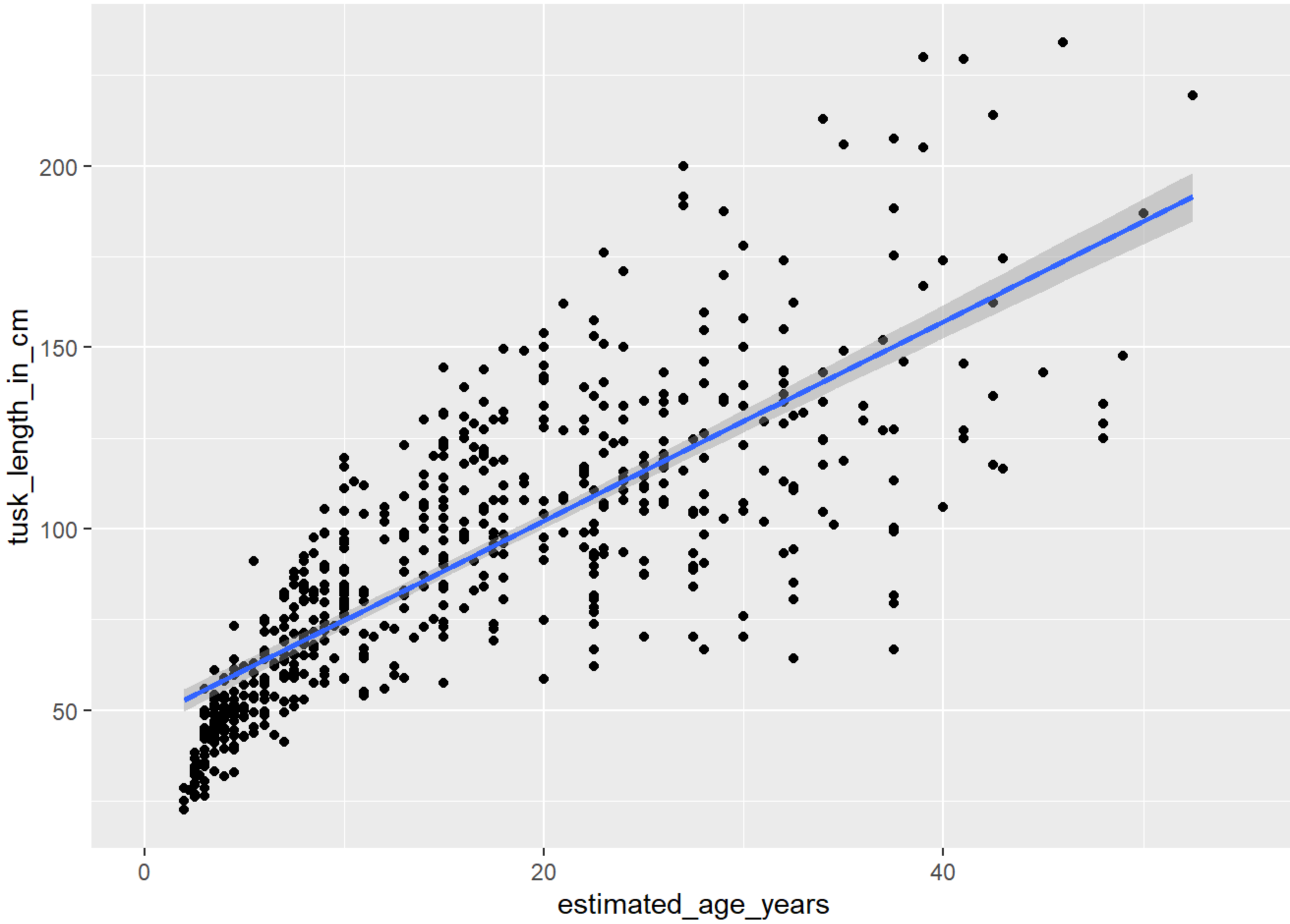
9. Is age (independent variable) a positive predictor of tusk length (dependent variable)? Create a plot that shows the relationship between these variables and add a linear model fit line.

```
elephants %>%
  select(estimated_age_years, tusk_length_in_cm) %>%
  ggplot(mapping=aes(x=estimated_age_years, y=tusk_length_in_cm))+
  geom_point()+
  geom_smooth(method="lm", se = TRUE)
```

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

```
## Warning: Removed 182 rows containing non-finite outside the scale range
## (`stat_smooth()`).
```

```
## Warning: Removed 182 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



Yes, age is a positive predictor for tuck length. As an elephant gets older, generally the tusk length increases.

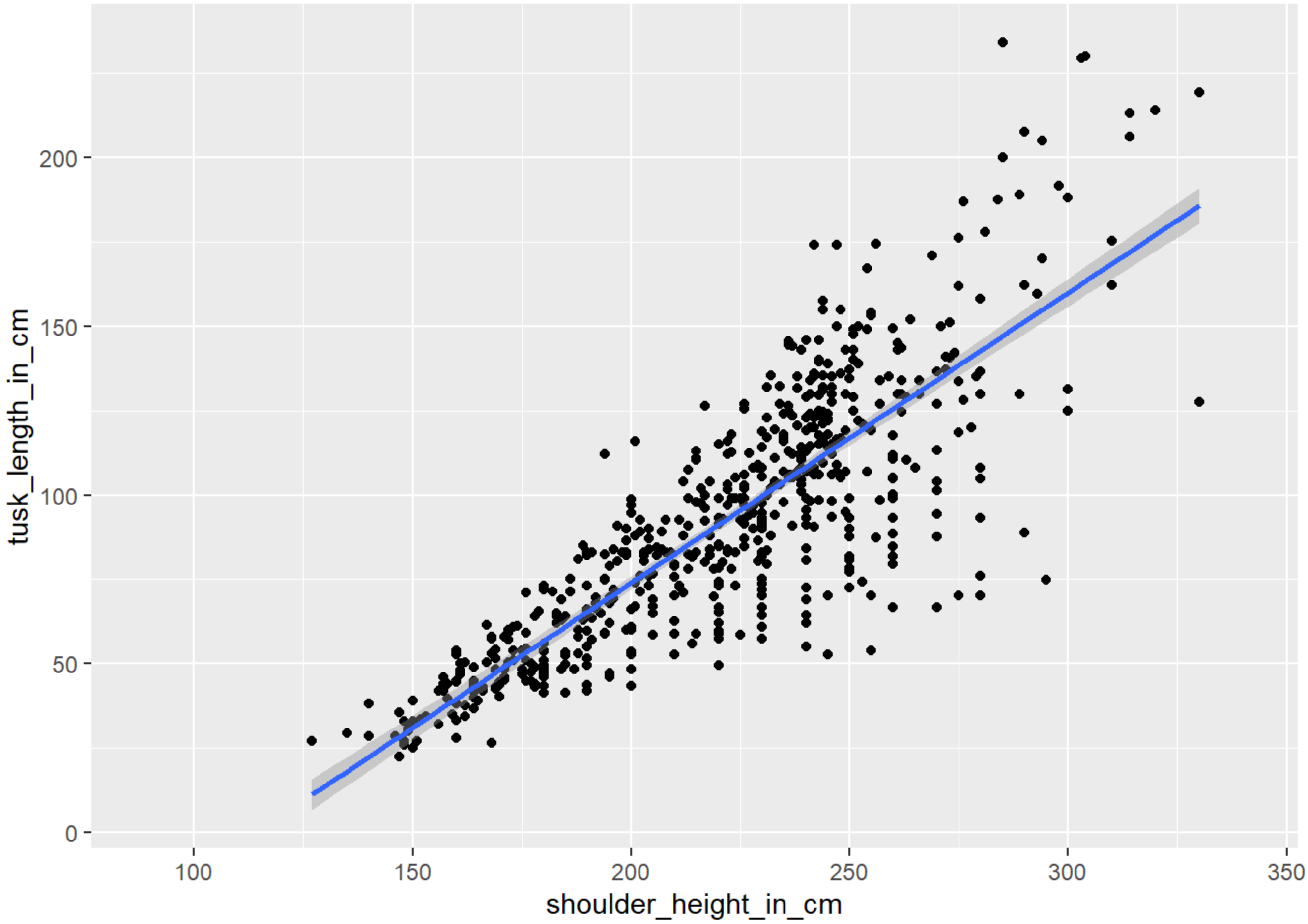
10. Is shoulder height (independent variable) a positive predictor of tusk length (dependent variable)? Create a plot that shows the relationship between these variables and add a linear model fit line.

```
elephants %>%
  select(shoulder_height_in_cm, tusk_length_in_cm) %>%
  ggplot(mapping=aes(x=shoulder_height_in_cm, y=tusk_length_in_cm))+
  geom_point()+
  geom_smooth(method="lm", se = TRUE)
```

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

```
## Warning: Removed 181 rows containing non-finite outside the scale range
## (`stat_smooth()`).
```

```
## Warning: Removed 181 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



Yes, shoulder height is a strong, positive predictor of tusk length.

11. The authors argue that because poachers preferentially target elephants with large tusks, this has resulted in a decrease in average tusk length. Is this supported by the data? Show your code and calculations below.

```
elephants %>%
  filter(years_of_sample_collection == "1966-68", !is.na(tusk_length_in_cm)) %>%
  summarize(mean_tusk = mean(tusk_length_in_cm, na.rm = T),
            median_tusk = median(tusk_length_in_cm))
```

```
## # A tibble: 1 × 2
##   mean_tusk median_tusk
##       <dbl>       <dbl>
## 1      96.9        97.8
```

```
elephants %>%
  filter(years_of_sample_collection == "2005-13", !is.na(tusk_length_in_cm)) %>%
  summarize(mean_tusk = mean(tusk_length_in_cm, na.rm = T),
            median_tusk = median(tusk_length_in_cm))
```

```
## # A tibble: 1 × 2
##   mean_tusk median_tusk
##       <dbl>       <dbl>
## 1      77.2        70.2
```

Yes, there is evidence of the average tusk decrease from 1966-1968 data compared to the 2005-2013 data.

12. Male elephants reach effective sexual maturity at 25 years while females are sexually mature at 12 years. Make a new dataframe that extracts only the males and females at sexual maturity. Then, make a plot that shows the range of tusk length between the two sample periods for these mature elephants.

```
mat_males <- elephants %>%
  filter(sex == "m" & estimated_age_years >= 25, !is.na(tusk_length_in_cm)) %>%
  select(sex, estimated_age_years, tusk_length_in_cm, years_of_sample_collection)
mat_males
```

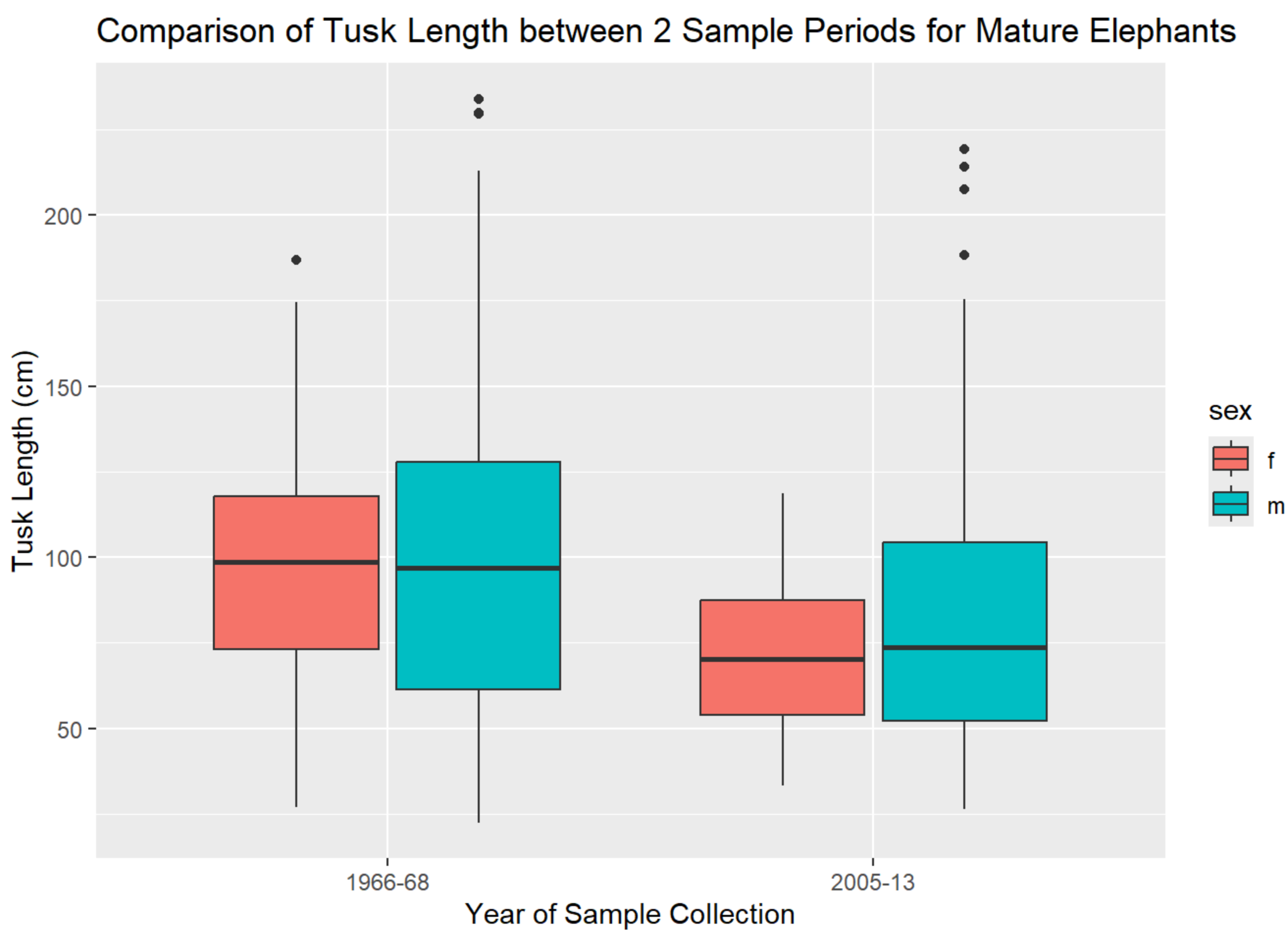
```
## # A tibble: 33 × 4
##    sex   estimated_age_years tusk_length_in_cm years_of_sample_collection
##    <chr>               <dbl>             <dbl> <chr>
##  1 m                      25              120  1966-68
##  2 m                      26              135  1966-68
##  3 m                      27              200  1966-68
##  4 m                      30              178  1966-68
##  5 m                      35              206  1966-68
##  6 m                      39              230  1966-68
##  7 m                      26              137  1966-68
##  8 m                      27              189  1966-68
##  9 m                      27              192. 1966-68
## 10 m                      28              160. 1966-68
## # ℹ 23 more rows
```

```
mat_females <- elephants %>%
  filter(sex == "f" & estimated_age_years >= 12, !is.na(tusk_length_in_cm)) %>%
  select(sex, estimated_age_years, tusk_length_in_cm, years_of_sample_collection)
mat_females
```

```
## # A tibble: 212 × 4
##    sex   estimated_age_years tusk_length_in_cm years_of_sample_collection
##    <chr>               <dbl>             <dbl> <chr>
##  1 f                    12                56   1966-68
##  2 f                    13                91   1966-68
##  3 f                    13                81.5 1966-68
##  4 f                    13                82   1966-68
##  5 f                    13                83   1966-68
##  6 f                    13.5              70   1966-68
##  7 f                    14                84   1966-68
##  8 f                    14                94   1966-68
##  9 f                    14                87   1966-68
## 10 f                    14                73   1966-68
## # ℹ 202 more rows
```

```
ggplot(data=elephants,
       mapping=aes(x=years_of_sample_collection, y=tusk_length_in_cm))+
  geom_boxplot(mapping=aes(fill=sex))+
  labs(title="Comparison of Tusk Length between 2 Sample Periods for Mature Elephants", x = "Yea
r of Sample Collection", y = "Tusk Length (cm)")
```

```
## Warning: Removed 180 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```



In general, there is a decrease in the median tusk length in both female and male elephants when comparing between the 2 sample periods.

Submit the Homework

1. Save your work and knit the .rmd file.
2. Open the .html file and “print” it to a .pdf file in Google Chrome (not Safari).
3. Go to the class Canvas page and open Gradescope.
4. Submit your .pdf file to the homework assignment- be sure to assign the pages to the correct questions.
5. Commit and push your work to your repository.