Homework 7

Alejandro Acevedo

2026-01-28

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 was running test on elephants that were victims of the ivory trade to see if selectio
n would lead their descendants to grow smaller tusk as a response to make them less desirable fo
r the ivory trade. They measured several size related variables, including tusk circumference, t
usk size, and overall animal size by measuring shoulder height
```

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

```
elephants <- read.csv("data/elephants.csv")
```

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

```
elephants_new <- clean_names(elephants)
glimpse(elephants_new)
```

```
## 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,…
```

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

```
glimpse(elephants_new)
```

```
## 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,…
```

```
names(elephants_new)
```

```
## [1] "years_of_sample_collection" "elephant_id"
## [3] "sex"                        "estimated_age_years"
## [5] "shoulder_height_in_cm"      "tusk_length_in_cm"
## [7] "tusk_circumference_in_cm"
```

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_newest <- elephants_new %>%
  mutate(years_of_sample_collection = as.factor(years_of_sample_collection),
         elephant_id = as.factor(elephant_id),
         sex = as.factor(sex))
```

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

```
elephants_newest %>%
  select(years_of_sample_collection) %>%
  distinct(years_of_sample_collection)
```

```
##   years_of_sample_collection
## 1                    1966-68
## 2                    2005-13
```

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

```
names(elephants_newest)
```

```
## [1] "years_of_sample_collection" "elephant_id"
## [3] "sex"                        "estimated_age_years"
## [5] "shoulder_height_in_cm"      "tusk_length_in_cm"
## [7] "tusk_circumference_in_cm"
```

```
elephants_newest %>%
  count(sex)
```

```
##   sex   n
## 1   f 416
## 2   m 361
```

```
#416 Females and #361 Males were sampled
```

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_newest %>%
  filter(years_of_sample_collection=="1966-68" & sex=="f") %>%
  summarise(mean_age_f_1966_68=mean(estimated_age_years, na.rm=T),
            sd_age_f_1966_68=sd(estimated_age_years, na.rm=T),
            median_age_f_1966_68=median(estimated_age_years, na.rm = T))
```

```
##   mean_age_f_1966_68 sd_age_f_1966_68 median_age_f_1966_68
## 1           17.57244         13.56243                   15
```

```
#The sampling is not even, it appears as though the mean age of a male elephant went up noticeab
ly from the 60's to the 2000's. The female age however is virtually unchanged.
```

```
elephants_newest %>%
  filter(years_of_sample_collection=="1966-68" & sex=="m") %>%
  summarise(mean_age_m_1966_68=mean(estimated_age_years, na.rm=T),
            sd_age_m_1966_68=sd(estimated_age_years, na.rm=T),
            median_age_m_1966_68=median(estimated_age_years, na.rm = T))
```

```
##   mean_age_m_1966_68 sd_age_m_1966_68 median_age_m_1966_68
## 1           10.79044          9.18845                    8
```

```
elephants_newest %>%
  filter(years_of_sample_collection=="2005-13" & sex=="f") %>%
  summarise(mean_age_f_2005_13=mean(estimated_age_years, na.rm=T),
            sd_age_f_2005_13=sd(estimated_age_years, na.rm=T),
            median_age_f_2005_13=median(estimated_age_years, na.rm = T))
```

```
##   mean_age_f_2005_13 sd_age_f_2005_13 median_age_f_2005_13
## 1           17.85484         10.95137                 17.5
```

```
elephants_newest %>%
  filter(years_of_sample_collection=="2005-13" & sex=="m") %>%
  summarise(mean_age_m_2005_13=mean(estimated_age_years, na.rm=T),
            sd_age_m_2005_13=sd(estimated_age_years, na.rm=T),
            median_age_m_2005_13=median(estimated_age_years, na.rm = T))
```

```
##   mean_age_m_2005_13 sd_age_m_2005_13 median_age_m_2005_13
## 1           16.70779          13.9046                    9
```

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.

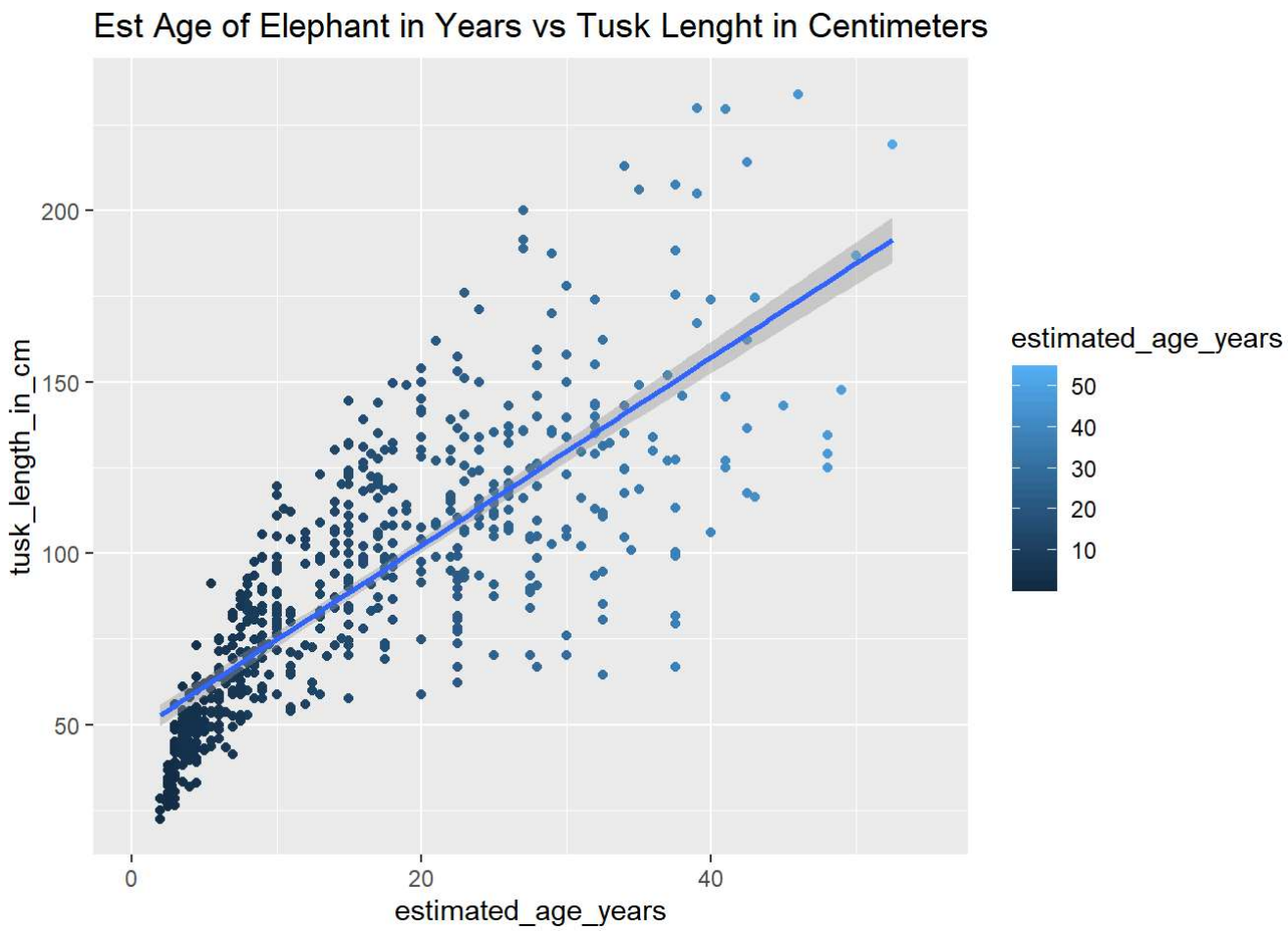
```
elephants_newest %>%
  ggplot(mapping=aes(x=estimated_age_years, y=tusk_length_in_cm, color=estimated_age_years))+
  geom_point()+
  geom_smooth(method=lm)+
  labs(title="Est Age of Elephant in Years vs Tusk Lenght in Centimeters")
```

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

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

```
## Warning: The following aesthetics were dropped during statistical transformation:
## colour.
## ℹ This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## ℹ Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```

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



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.

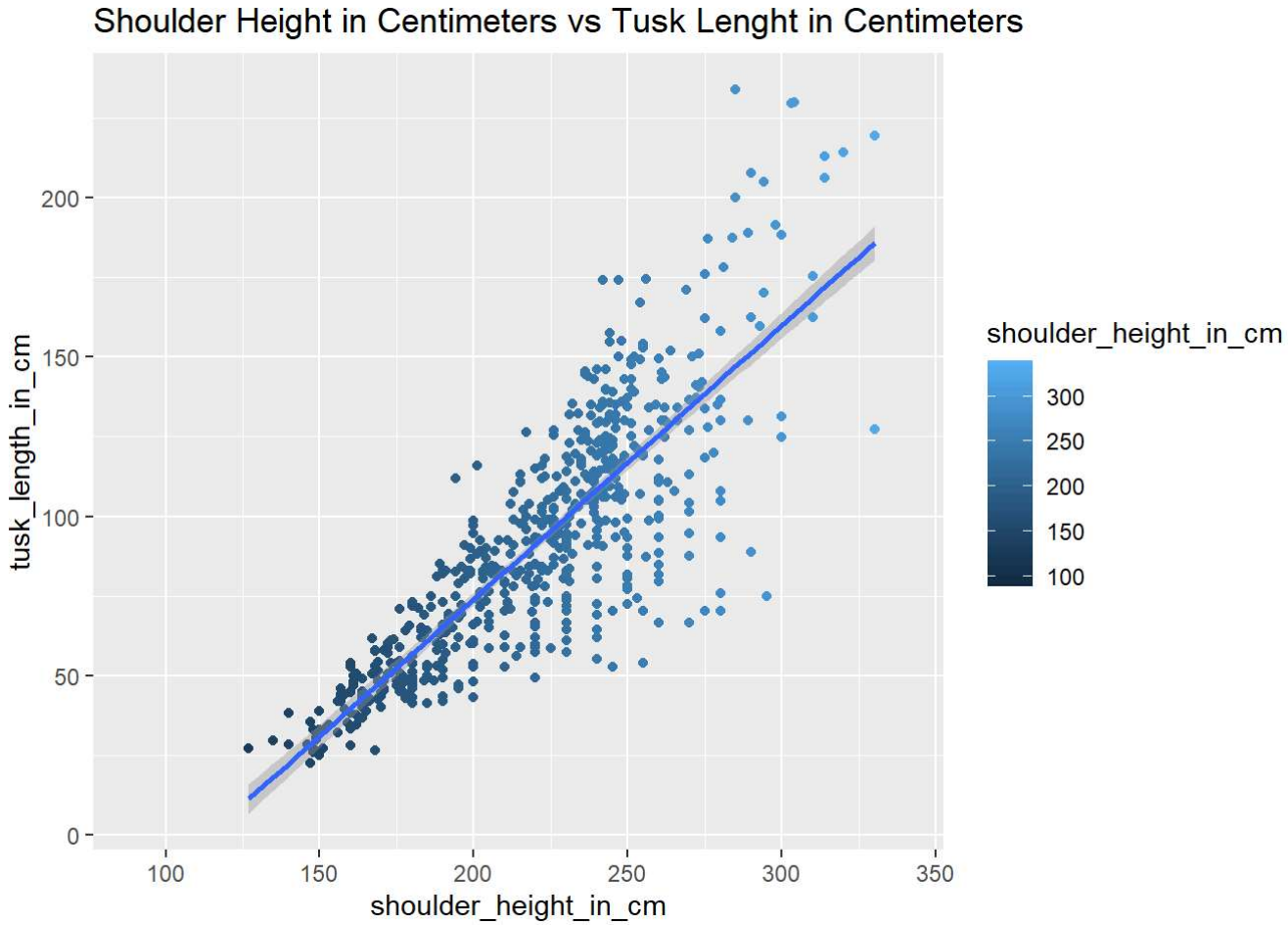
```
elephants_newest %>%
  ggplot(mapping=aes(x=shoulder_height_in_cm, y=tusk_length_in_cm, color=shoulder_height_in_cm))
+
  geom_point()+
  geom_smooth(method=lm)+
  labs(title="Shoulder Height in Centimeters vs Tusk Lenght in Centimeters")
```

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

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

```
## Warning: The following aesthetics were dropped during statistical transformation:
## colour.
## ℹ This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## ℹ Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```

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



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_newest %>%
  filter(years_of_sample_collection=="1966-68") %>%
  summarize(mean_tusk_length_cm=mean(tusk_length_in_cm, na.rm=T),
            mean_tusk_circumfrence_cm=mean(tusk_circumference_in_cm, na.rm=T))
```

```
##   mean_tusk_length_cm mean_tusk_circumfrence_cm
## 1            96.90436                  20.77743
```

```
elephants_newest %>%
  filter(years_of_sample_collection=="2005-13") %>%
  summarize(mean_tusk_length_cm=mean(tusk_length_in_cm, na.rm=T),
            mean_tusk_circumfrence_cm=mean(tusk_circumference_in_cm, na.rm=T))
```

```
##   mean_tusk_length_cm mean_tusk_circumfrence_cm
## 1             77.2056                  21.59942
```

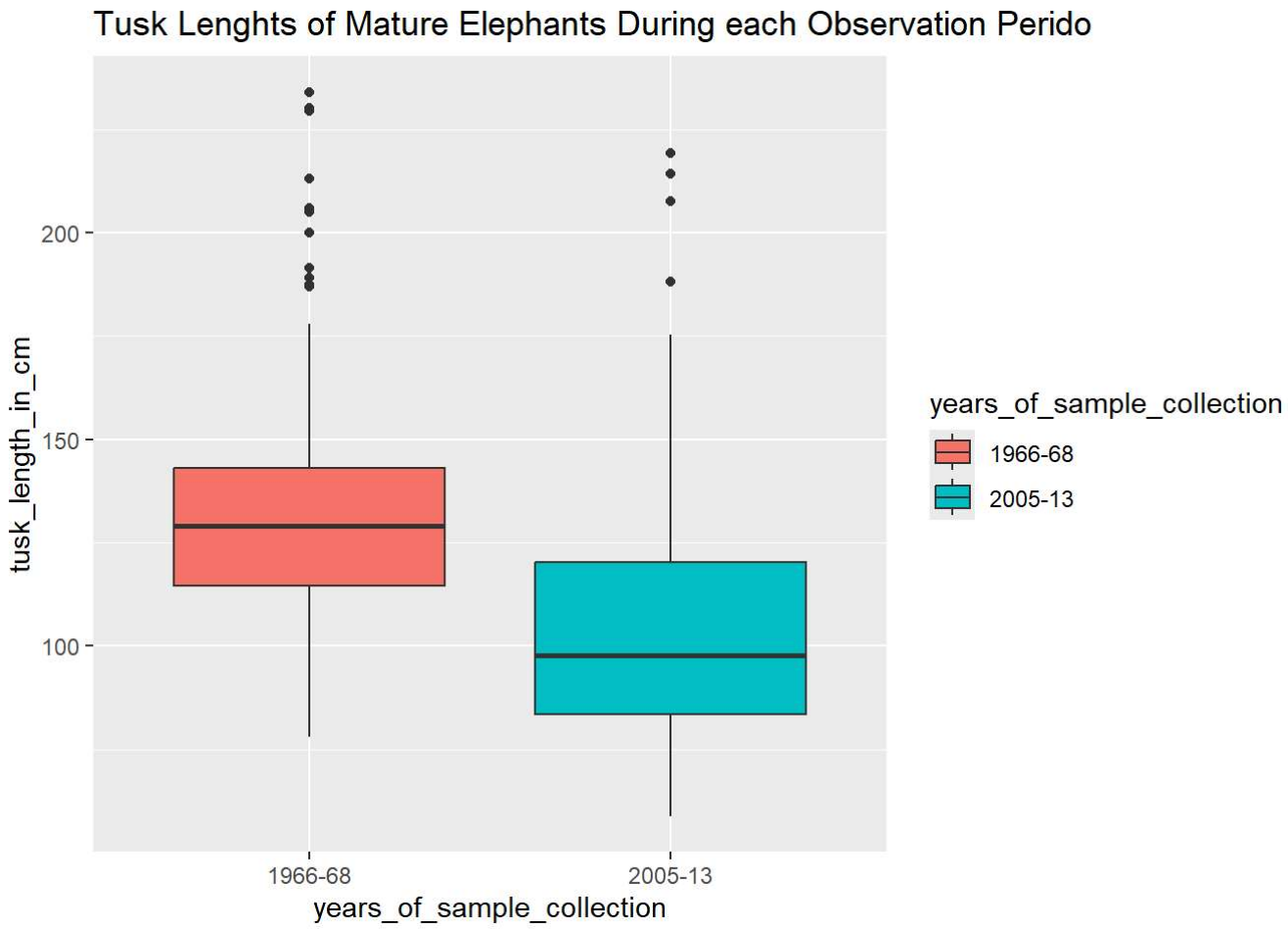
```
  #The data supports the elephant tusk shrinkage hypothesis as you can clearly see that the aver
age tusk length decreased by abnout 20 centimeters from the 60's to the newer observed group in
the 2000's
```

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.

```
elephants_mature <- elephants_newest %>%
  filter(estimated_age_years>=25 & sex=="f" | estimated_age_years>=12 & sex=="m")
```

```
elephants_mature %>%
  ggplot(mapping=aes(x=years_of_sample_collection, y=tusk_length_in_cm, fill=years_of_sample_col
lection))+
  geom_boxplot()+
  labs(title="Tusk Lenghts of Mature Elephants During each Observation Perido")
```

```
## Warning: Removed 46 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```



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.