Midterm 1 W24

Catrinel Berevoescu

2024-02-06

Instructions

Answer the following questions and complete the exercises in RMarkdown. Please embed all of your code and push your final work to your repository. Your code must be organized, clean, and run free from errors. Remember, you must remove the `#` for any included code chunks to run. Be sure to add your name to the author header above.

Your code must knit in order to be considered. If you are stuck and cannot answer a question, then comment out your code and knit the document. You may use your notes, labs, and homework to help you complete this exam. Do not use any other resources-including AI assistance.

Don't forget to answer any questions that are asked in the prompt!

Be sure to push your completed midterm to your repository. This exam is worth 30 points.

Background

In the data folder, you will find data related to a study on wolf mortality collected by the National Park Service. You should start by reading the `README_NPSwolfdata.pdf` file. This will provide an abstract of the study and an explanation of variables.

The data are from: Cassidy, Kira et al. (2022). Gray wolf packs and human-caused wolf mortality. Dryad (https://doi.org/10.5061/dryad.mkkwh713f).

Load the libraries.

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

Load the wolves data

In these data, the authors used `NULL` to represent missing values. I am correcting this for you below and using `janitor` to clean the column names.

```
wolves <- read.csv("data/NPS_wolfmortalitydata.csv", na = c("NULL")) %>% clean_names()
```

Questions

Problem 1. (1 point) Let's start with some data exploration. What are the variable (column) names?

The variable (column names) are:

```
names(wolves)
```

```
##  [1] "park"         "biolyr"       "pack"         "packcode"     "packsize_aug"
##  [6] "mort_yn"      "mort_all"     "mort_lead"    "mort_nonlead" "reprody1"
## [11] "persisty1"
```

Problem 2. (1 point) Use the function of your choice to summarize the data and get an idea of its structure.

```
glimpse(wolves) #using the glimpse() function to summarize the data and get an idea of its structure
```

```
## Rows: 864
## Columns: 11
## $ park         <chr> "DENA", "DENA", "DENA", "DENA", "DENA", "DENA", "DENA", "…
## $ biolyr       <int> 1996, 1991, 2017, 1996, 1992, 1994, 2007, 2007, 1995, 200…
## $ pack         <chr> "McKinley River1", "Birch Creek N", "Eagle Gorge", "East …
## $ packcode     <int> 89, 58, 71, 72, 74, 77, 101, 108, 109, 53, 63, 66, 70, 72…
## $ packsize_aug <dbl> 12, 5, 8, 13, 7, 6, 10, NA, 9, 8, 7, 11, 0, 19, 15, 12, 1…
## $ mort_yn      <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, …
## $ mort_all     <int> 4, 2, 2, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, …
## $ mort_lead    <int> 2, 2, 0, 0, 0, 0, 1, 2, 1, 1, 1, 0, 0, 0, 1, 1, 1, 0, 0, …
## $ mort_nonlead <int> 2, 0, 2, 2, 2, 2, 1, 0, 1, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, …
## $ reprody1     <int> 0, 0, NA, 1, NA, 0, 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1…
## $ persisty1    <int> 0, 0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, …
```

Problem 3. (3 points) Which parks/reserves are represented in the data? Don't just use the abstract, pull this information from the data.

```
wolves$park <- as.factor(wolves$park)
is.factor(wolves$park)
```

```
## [1] TRUE
```

```
levels(wolves$park)
```

```
## [1] "DENA" "GNTP" "VNP"  "YNP"  "YUCH"
```

Thus, the 5 parks/reserves represented in the data include "DENA" (Denali National Park and Preserve), "GNTP" (Grand Teton National Park), "VNP" (Voyageurs National Park), "YNP" (Yellowstone National Park), and "YUCH" (Yukon-Charley Rivers National Preserve).

Problem 4. (4 points) Which park has the largest number of wolf packs?

```
wolves %>%
  group_by(park) %>% #grouping by the variable park
  summarize(number_wolf_packs = n_distinct(pack)) %>% #finding the number of distinct wolf packs in each
park
  arrange(desc(number_wolf_packs)) #arranging from the highest to the lowest for ease of viewing
```

```
## # A tibble: 5 × 2
##   park  number_wolf_packs
##   <fct>             <int>
## 1 DENA                 69
## 2 YNP                  46
## 3 YUCH                 36
## 4 VNP                  22
## 5 GNTP                 12
```

Thus, "DENA", Denali National Park and Preserve, has the largest distinct number of wolf packs (69 total).

Problem 5. (4 points) Which park has the highest total number of human-caused mortalities `mort_all`?

```
wolves %>%
  group_by(park) %>% #grouping by the variable park
  summarize(human_caused_mortalities = sum(mort_all)) %>% #finding the total number of human-caused mort
alities at each park
  arrange(desc(human_caused_mortalities)) #arranging from highest to lowest values for ease of viewing
```

```
## # A tibble: 5 × 2
##   park  human_caused_mortalities
##   <fct>                    <int>
## 1 YUCH                       136
## 2 YNP                         72
## 3 DENA                        64
## 4 GNTP                        38
## 5 VNP                         11
```

Thus, the park has the highest total number of human-caused mortalities is "YUCH", Yukon-Charley Rivers National Preserve, with 136 mortalities.

The wolves in Yellowstone National Park (https://www.nps.gov/yell/learn/nature/wolf-restoration.htm) are an incredible conservation success story. Let's focus our attention on this park.

Problem 6. (2 points) Create a new object "ynp" that only includes the data from Yellowstone National Park.

```
ynp <- wolves %>% #creating the new object "ynp"
  filter(park == "YNP") #filtering the wolves data frame for only data from Yellowstone National Park
```

Problem 7. (3 points) Among the Yellowstone wolf packs, the Druid Peak Pack (https://www.pbs.org/wnet/nature/in-the-valley-of-the-wolves-the-druid-wolf-pack-story/209/) is one of most famous. What was the average pack size of this pack for the years represented in the data?

```
ynp %>%
  filter(pack == "druid") %>% #filtering for only the Yellowstone Druid Peak Pack
  group_by(pack) %>% #grouping by the variable pack
  summarize(average_pack_size = mean(packsize_aug, na.rm = T)) #finding the average pack size for the Dr
uid Peak Pack
```

```
## # A tibble: 1 × 2
##   pack  average_pack_size
##   <chr>             <dbl>
## 1 druid              13.9
```

Thus, the average pack size of the Druid Peak pack during the time frame represented in this data is 13.93 individuals (slightly under 14).

Problem 8. (4 points) Pack dynamics can be hard to predict - even for strong packs like the Druid Peak pack. At which year did the Druid Peak pack have the largest pack size? What do you think happened in 2010?

```
ynp %>%
  filter(pack == "druid") %>% #filtering for only the Yellowstone Druid Peak Pack
  group_by(pack) %>% #grouping by the variable pack
  summarize(max_pack_size = max(packsize_aug)) #finding the largest Druid Peak pack size
```

```
## # A tibble: 1 × 2
##   pack  max_pack_size
##   <chr>         <dbl>
## 1 druid            37
```

The maximum pack size for the Druid Peak pack was 37 individuals.

```
ynp %>%
  filter(pack == "druid", packsize_aug == 37) %>% #finding the year with the largest pack size
  select(biolyr, packsize_aug) #selecting variables for ease of viewing
```

```
##   biolyr packsize_aug
## 1   2001           37
```

Thus, the year with the largest pack size for the Druid Peak pack was 2001 with 37 individuals.

What happened in 2010:

```
ynp %>%
  filter(pack == "druid", biolyr == "2010") #filtering for the year 2010
```

```
##   park biolyr  pack packcode packsize_aug mort_yn mort_all mort_lead
## 1  YNP   2010 druid       26            0       0        0         0
##   mort_nonlead reprody1 persisty1
## 1            0        0        NA
```

It appears, based on this data, that there were no individuals in the Druid Peak pack in 2010. This could indicate that there were no wolves in this pack in that year or that there was an issue with the measurements taken (no measurements taken that year). When searching up pack events from this year, it seems that this pack was scavenging carcasses from other packs, leading to a high mortality rate because of violence between wolves. The result of this had a clear deleterious effect on pack numbers, given that this is the last year with data provided for this pack - Druid Peak Pack Yellowstone (https://www.yellowstonewolf.org/yellowstones_wolves.php?pack_id=10#:~:text=In%20early%202010%20the%20pack,legacy%20of%20this%20famous%20pack).

Problem 9. (5 points) Among the YNP wolf packs, which one has had the highest overall persistence `persisty1` for the years represented in the data? Look this pack up online and tell me what is unique about its

behavior- specifically, what prey animals does this pack specialize on?

```
ynp %>%
  select(pack, persisty1) %>% #selecting for variables of interest
  group_by(pack) %>% #grouping by the variable pack
  summarize(overall_mean_persistence = mean(persisty1),
          overall_total_persistence = sum(persisty1)) %>% #finding the total overall persistence
  arrange(desc(overall_total_persistence)) #arranging for ease of viewing
```

```
## # A tibble: 46 × 3
##    pack         overall_mean_persistence overall_total_persistence
##    <chr>                      <dbl>                      <int>
##  1 mollies                    1                           26
##  2 cougar                     1                           20
##  3 yelldelta                  1                           18
##  4 leopold                    0.923                       12
##  5 agate                      0.909                       10
##  6 8mile                      1                            9
##  7 canyon                     0.9                          9
##  8 gibbon/mary                0.9                          9
##  9 junction                   1                            8
## 10 lamar                      1                            8
## # i 36 more rows
```

Among the YNP wolf packs, it appears that the pack with the largest overall persistence was the "Mollies" pack. It appears, according to Mollie Yellowstone Wolves (https://www.yellowstonewolf.org/yellowstones_wolves.php?pack_id=6), that the behavior of this pack is unusual in that they hunt bison (their primary prey due to a limited number of more typical prey, such as elk, as the result of a previous relocation) and have regular interactions with bears.

Problem 10. (3 points) Perform one analysis or exploration of your choice on the `wolves` data. Your answer needs to include at least two lines of code and not be a summary function.

I am interested in finding out which year was the "bloodiest" - in which pack year there was the largest number of human-caused mortalities:

```
wolves %>%
  select(park, biolyr, mort_all) %>% #selecting only for the columns of interest
  group_by(biolyr) %>% #grouping by the variable biolyr
  summarize(total_mortalities = sum(mort_all, na.rm = T)) %>% #adding up the mortalities for each year
  arrange(desc(total_mortalities)) #arranging for ease of viewing
```

```
## # A tibble: 36 × 2
##    biolyr total_mortalities
##    <int>             <int>
##  1  2012                57
##  2  2014                20
##  3  2005                19
##  4  2013                18
##  5  2017                17
##  6  2000                15
##  7  2007                15
##  8  2009                15
##  9  2011                13
## 10  2016                13
## # i 26 more rows
```

Thus, the "bloodiest" year in this data frame, with the largest number of mortalities, was 2012 with 57 deaths.