Midterm 1 W24

Elisabeth Sellinger

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")
```

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?

```
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)
```

```
## 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, …
## $ mort_all     <int> 4, 2, 2, 2, 2, 2, 2, 2, 2, 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.

```
table(wolves$park)
```

```
##
## DENA GNTP  VNP  YNP YUCH
##  340   77   48  248  151
```

Using the `table()` function I was able to pull out the names of the parks in this data (it also provides counts). DENA is Denali National Park and Preserve, GNTP is Grand Teton National Parl, VNP is Voyageurs National Park, YNP is Yellowstone National Park, and YUCH is Yukon-Charley Rivers National Preserve.

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

```
wolves %>%
  group_by(park) %>%
  summarize(n_packs = n_distinct(pack))%>%
  arrange(desc(n_packs))
```

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

Denali National Park and Preserve has the largest number of wolf packs at 69.

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

```
wolves %>%
  group_by(park) %>%
  summarize(sum_human_caused_mortalities = sum(mort_all)) %>%
  arrange(desc(sum_human_caused_mortalities))
```

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

Yukon-Charley Rivers National Preserve has the largest number of human-caused mortalities at 136 total over the recorded years.

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 %>%
  filter(park == "YNP")
```

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") %>%
  summarize(druid_mean_pack_size = mean(packsize_aug, na.rm = T))
```

```
##   druid_mean_pack_size
## 1             13.93333
```

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") %>%
  arrange(desc(packsize_aug))
```

```
##    park biolyr  pack packcode packsize_aug mort_yn mort_all mort_lead
## 1   YNP   2001 druid       26           37       0        0         0
## 2   YNP   2000 druid       26           27       1        1         0
## 3   YNP   2008 druid       26           21       0        0         0
## 4   YNP   2003 druid       26           18       0        0         0
## 5   YNP   2007 druid       26           18       0        0         0
## 6   YNP   2002 druid       26           16       0        0         0
## 7   YNP   2006 druid       26           15       0        0         0
## 8   YNP   2004 druid       26           13       0        0         0
## 9   YNP   2009 druid       26           12       0        0         0
## 10  YNP   1999 druid       26            9       0        0         0
## 11  YNP   1998 druid       26            8       0        0         0
## 12  YNP   1997 druid       26            5       1        2         1
## 13  YNP   1996 druid       26            5       0        0         0
## 14  YNP   2005 druid       26            5       0        0         0
## 15  YNP   2010 druid       26            0       0        0         0
##    mort_nonlead reprody1 persisty1
## 1             0        1         1
## 2             1        1         1
## 3             0        1         1
## 4             0        1         1
## 5             0        1         1
## 6             0        1         1
## 7             0        1         1
## 8             0        1         1
## 9             0        0         0
## 10            0        1         1
## 11            0        1         1
## 12            1        1         1
## 13            0        1         1
## 14            0        1         1
## 15            0        0        NA
```

In 2001 the Druid Peak pack had the largest pack size at 37. In 2010 the pack size was 0 and I think this may have occurred due to low reproductive success in previous years since the `reprody1` values are at 0 meaning wolves did not localize and no pups were observed.

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 %>%
  group_by(pack) %>%
  filter(persisty1 == 1) %>%
  summarize(years_of_persistence = sum(persisty1)) %>%
  arrange(desc(years_of_persistence))
```

```
## # A tibble: 38 × 2
##    pack        years_of_persistence
##    <chr>                      <int>
##  1 mollies                       26
##  2 cougar                        20
##  3 yelldelta                     18
##  4 druid                         13
##  5 leopold                       12
##  6 agate                         10
##  7 8mile                          9
##  8 canyon                         9
##  9 gibbon/mary                    9
## 10 nezperce                       9
## # ℹ 28 more rows
```

The Mollies pack had the highest overall persistence (pack persisted with at least two pack members in the same general territory as the previous biological year) at 26 years. This pack specializes in preying on bison. This is because they were displaced from their territory in 1996 by the Druid Peak pack and relocated to Pelican Valley where limited ungulates remained in the interior during the winter. This wolf pack has unique behavior, including hunting bison and regular interactions with bears, which prompted the long-term Pelican Valley study where wolf, bison, and bear interactions are documented.

Source (https://www.yellowstonewolf.org/yellowstones_wolves.php?pack_id=6)

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.

```
wolves %>%
  group_by(park) %>%
  filter(biolyr >= 2000) %>%
  summarize(max_packsize_aug = max(packsize_aug, na.rm = T)) %>%
  arrange(desc(max_packsize_aug))
```

```
## # A tibble: 5 × 2
##   park  max_packsize_aug
##   <chr>            <dbl>
## 1 YNP               37
## 2 GNTP              26.4
## 3 YUCH              24
## 4 DENA              23
## 5 VNP                7
```