Midterm 1 W24

Lejla Becirevic

2024-02-06

```
getwd()
```

```
## [1] "/Users/lbecirev/Desktop/BIS15W2024_lbecirevic/midterm1"
```

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.

```
wolves %>%
  count(park) #DENA, GNTP, VNP, YNP, YUCH are represented in the data
```

```
##   park   n
## 1 DENA 340
## 2 GNTP  77
## 3  VNP  48
## 4  YNP 248
## 5 YUCH 151
```

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

```
wolves %>%
  group_by(park) %>%
  summarize(n_pack=n_distinct(pack)) %>%
  arrange(desc(n_pack)) #DENA has the largest number of wolf packs
```

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

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

```
wolves %>%
  group_by(park) %>%
  summarize(sum_mort=sum(mort_all)) %>%
  arrange(desc(sum_mort)) #YUCH has the highest number of human-caused mortalities
```

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

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(mean_pack_size=mean(packsize_aug))
```

```
##   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 %>%
  group_by(biolyr) %>%
  filter(pack=="druid") %>%
  summarize(max_pack_size=max(packsize_aug)) %>%
  arrange(desc(max_pack_size)) #The Druid Peak pack largest pack size was 2001
```

```
## # A tibble: 15 × 2
##     biolyr max_pack_size
##      <int>         <dbl>
##  1   2001            37
##  2   2000            27
##  3   2008            21
##  4   2003            18
##  5   2007            18
##  6   2002            16
##  7   2006            15
##  8   2004            13
##  9   2009            12
## 10   1999             9
## 11   1998             8
## 12   1996             5
## 13   1997             5
## 14   2005             5
## 15   2010             0
```

Based off the wolves data abstract, anthropocentric variables impacts the survival of wolf packs in various parks. In 2010 there was no observation of the Druid Peak pack possibly due to human involvement.

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) %>%
  summarize(sum_persistence=sum(persisty1)) %>%
  arrange(desc(sum_persistence)) #The pack with the highest overall persistence was Moll
ie's pack
```

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

Mollie's pack consist of size alpha females and is considered a "girl-power" group. The pack preys upon bison after the elk migrated to lower elevation.

Reference: Reed E. (2020). Studying Yellowstone's iconic wolves.
(https://greateryellowstone.org/blog/2020/studyingwolves
(https://greateryellowstone.org/blog/2020/studyingwolves))

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, pack) %>%
  filter(park=="DENA") %>%
  filter(mort_lead==1 & reprody1==1)
```

```
## # A tibble: 12 × 11
## # Groups:   park, pack [10]
##    park  biolyr pack          packcode packsize_aug mort_yn mort_all mort_lead
##    <chr>  <int> <chr>            <int>        <dbl>   <int>    <int>     <int>
##  1 DENA    2003 100 Mile           53            8       1        1         1
##  2 DENA    1992 East Fork          72           15       1        1         1
##  3 DENA    2004 East Fork          72           12       1        1         1
##  4 DENA    2005 East Fork          72           14       1        1         1
##  5 DENA    1992 McKinley River1    89            7       1        1         1
##  6 DENA    2008 Mt Margaret        95            2       1        1         1
##  7 DENA    2013 Nenana River       98            9       1        1         1
##  8 DENA    1994 Savage1           109            9       1        1         1
##  9 DENA    1991 Stampede          112            2       1        1         1
## 10 DENA    2006 Starr Lake        113            3       1        1         1
## 11 DENA    2007 Hauke Creek        76            4       1        1         1
## 12 DENA    1993 Headquarters       77            9       1        1         1
## # ℹ 3 more variables: mort_nonlead <int>, reprody1 <int>, persisty1 <int>
```