Midterm 1 W24

Kassidy Yu

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?
The column names are "park", "biolyr", "pack", "packcode", "packsize_aug", "mort_yn", "mort_all", "mort_lead", "mort_nonlead", "reprody1", and "persisty1".

```
colnames(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, 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.
The parks represented in the data are "DENA", "GNTP", "VNP", "YNP", and "YUCH".

```
park_factors <- as.factor(wolves$park)
levels(park_factors)
```

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

Problem 4. (4 points) Which park has the largest number of wolf packs?
DENA has the largest number of wolf packs at 69 total.

```
wolves %>%
  group_by(park) %>%
  summarize(num_packs = n_distinct(pack)) %>%
  filter(num_packs == max(num_packs))
```

```
## # A tibble: 1 × 2
##   park  num_packs
##   <chr>     <int>
## 1 DENA         69
```

Problem 5. (4 points) Which park has the highest total number of human-caused mortalities `mort_all`?
The Yukon-Charley Rivers National Preserve ("YUCH") has the highest total number of human-caused mortalities.

```
wolves %>%
  group_by(park) %>%
  summarize(human_mortality_total = sum(mort_all)) %>%
  filter(human_mortality_total == max(human_mortality_total))
```

```
## # A tibble: 1 × 2
##   park  human_mortality_total
##   <chr>                 <int>
## 1 YUCH                    136
```

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?
The average pack size of the Druid Peak Pack for the years represented is 13.93333.

```
druid <- ynp %>%
  filter(pack == "druid") %>%
  arrange(biolyr)
mean(druid$packsize_aug)
```

```
## [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?
The Druid Peak pack had the largest pack size of 37 in 2001.

```
druid %>%
  select(pack, biolyr, packsize_aug)  %>%
  filter(packsize_aug == max(packsize_aug))
```

```
##    pack biolyr packsize_aug
## 1 druid   2001           37
```

It seems like the pack died out in 2010 - there was no human-caused mortality around 2010, so perhaps there was a famine.

```
druid %>%
  select(pack, biolyr, packsize_aug, mort_all) %>%
  arrange(biolyr)
```

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

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?
The Mollie's pack had the highest overall persistence. They have had female alphas with long reigns, which has allowed for a stable pack life. Their unity allows them to excel in hunting down bison, which are the hardest prey for wolves to kill. They use snow, which weakens bison, to their advantage. Sources: Greater Yellowstone Coalition (https://greateryellowstone.org/blog/2020/studyingwolves) and The Spokesman-Review (https://www.spokesman.com/stories/2012/jan/15/hungry-wolf-pack-rearranges-balance-in/).

```
ynp %>%
  group_by(pack) %>%
  filter(persisty1 == 1) %>%
  select(pack, persisty1) %>%
  count(pack) %>%
  arrange(desc(n))
```

```
## # A tibble: 38 × 2
## # Groups:   pack [38]
##    pack            n
##    <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
```

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 selected Yukon-Charley Rivers National Preserve, since in Problem 5, I found that they had the highest total human-caused mortality. Then, I wanted to see which year and packs these mortality numbers, so I selected the variables of interest and sorted by the highest `mort_all`. I found that the highest `mort_all` for one year in YUCH was in 2012, in which 24 wolves from the 70 Mile pack were killed by humans.

```
wolves %>%
  filter(park == "YUCH", mort_all != 0) %>%
  select(biolyr, pack, mort_all) %>%
  arrange(desc(mort_all))
```

```
##    biolyr             pack mort_all
## 1    2012          70 Mile       24
## 2    2005       Cottonwood       14
## 3    2013       Lost Creek       12
## 4    2014      Sheep Bluff       12
## 5    2012       Yukon Fork       10
## 6    2011       Lost Creek        8
## 7    2008  Copper Mountain        5
## 8    2002  Hard Luck Creek        5
## 9    2000          70 Mile        4
## 10   2009   Webber Creek 2        4
## 11   2012      Woodchopper        4
## 12   2000       Cottonwood        3
## 13   2005          70 Mile        2
## 14   2008          70 Mile        2
## 15   2009  Copper Mountain        2
## 16   1996       Cottonwood        2
## 17   1993   Fourth of July        2
## 18   2000   Lower Charley1        2
## 19   2010   Lower Charley2        2
## 20   1995 Washington Creek        2
## 21   2013       Yukon Fork        2
## 22   1994       Cottonwood        1
## 23   2006  Crescent Creek2        1
## 24   1996   Edwards Creek1        1
## 25   2001   Edwards Creek2        1
## 26   2007   Edwards Creek2        1
## 27   2004     Fisher Creek        1
## 28   1996       Flat Creek        1
## 29   2006      Hanna Creek        1
## 30   2005   Step Mountain2        1
## 31   2012   Step Mountain2        1
## 32   1999     Three Finger        1
## 33   2000     Three Finger        1
## 34   1997     Three Finger        1
```