

Midterm 1 W24

Spencer Bryan

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.

```
wolves %>%
  tabyl(park)
```

```
## park    n    percent
## DENA  340 0.39351852
## GNT  77 0.08912037
## VNP   48 0.05555556
## YNP  248 0.28703704
## YUCH 151 0.17476852
```

Denali National Park and Preserve (DENA), Grand Teton National Park (GTNP), Voyageurs National Park (VNP), Yellowstone National Park (YNP), and Yukon-Charley Rivers National Preserve (YUCH) are represented in this data.

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

```
wolves %>%
  group_by(park) %>%
  select(park, packcode) %>%
  summarise(number_packs=n_distinct(packcode))%>%
  arrange(desc(number_packs))
```

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

Denali Nat'l Park and preserve has the largest number of unique wolfpacks

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

```
wolves %>%
  group_by(park) %>%
  select(park, mort_all) %>%
  summarise(total_mort= sum(mort_all)) %>%
  arrange(desc(total_mort))
```

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

Yukon-Rivers Charlie Nat'l Preserve (YUCH) has the highest total human-caused 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 %>%
  filter(pack == "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 %>%
  select(pack, packsize_aug) %>%
  filter(pack == "druid") %>%
  summarise(avg_packsize= mean(packsize_aug))
```

```
##   avg_packsize
## 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?

This filters to find the largest August pack size recorded for the "Druid" pack

```
ynp %>%
  select(pack, packsize_aug, biolyr) %>%
  filter(pack == "druid") %>%
  summarise(largest_packsize = max(packsize_aug))
```

```
##   largest_packsize
## 1                37
```

This displays the pertinent related data for the maximum August pack size found.

```
ynp %>%
  select(pack, biolyr, packsize_aug) %>%
  filter(pack == "druid") %>%
  filter(packsize_aug == 37)
```

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

View data for 2010 for the "druid" pack

```
ynp %>%
  select(pack, biolyr, packsize_aug) %>%
  filter(pack == "druid") %>%
  filter(biolyr == 2010)
```

```
##      pack biolyr packsize_aug
## 1 druid   2010             0
```

It is likely that the pack died off during at some point between the 2010 and 2009 August pack size counting (the 2009 biological year). There is no data observed after 2010, so a likely conclusion is that the "druid" pack died off. The data shown by the code chunk below shows the data to support this.

```
ynp %>%
  select(pack, biolyr, packsize_aug) %>%
  filter(pack == "druid") %>%
  arrange(desc(biolyr))
```

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

Problem 9. (5 points) Among the YNP wolf packs, which one has had the highest overall persistence `persistyl` 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?

```
# coded based on total persistence over time by summing the amount of times the pack  
has remained in it's same spot.
```

```
ynp %>%  
  select(pack, packcode, persistyl) %>%  
  group_by(packcode) %>%  
  summarise(total_pack_persisty= sum(persistyl)) %>%  
  arrange(desc(total_pack_persisty))
```

```
## # A tibble: 46 × 2  
##   packcode total_pack_persisty  
##   <int>         <int>  
## 1      38             26  
## 2      24             20  
## 3      52             18  
## 4      36             12  
## 5      15             10  
## 6      11              9  
## 7      20              9  
## 8      29              9  
## 9      33              8  
## 10     34              8  
## # i 36 more rows
```

```
# to determine the name of pack 38
```

```
ynp %>%  
  select(pack, packcode) %>%  
  filter(packcode == 38)
```

```
##      pack packcode
## 1  mollies      38
## 2  mollies      38
## 3  mollies      38
## 4  mollies      38
## 5  mollies      38
## 6  mollies      38
## 7  mollies      38
## 8  mollies      38
## 9  mollies      38
## 10 mollies      38
## 11 mollies      38
## 12 mollies      38
## 13 mollies      38
## 14 mollies      38
## 15 mollies      38
## 16 mollies      38
## 17 mollies      38
## 18 mollies      38
## 19 mollies      38
## 20 mollies      38
## 21 mollies      38
## 22 mollies      38
## 23 mollies      38
## 24 mollies      38
## 25 mollies      38
## 26 mollies      38
```

Pack 38, or "mollies" has displayed the most overall persistence in their territory over time. This pack specializes in hunting bison.

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 going to find the largest August pack size recorded in each park.

```
wolves %>%
  group_by(park) %>%
  summarise(largest_pack = max(packsize_aug, na.rm = T)) %>%
  arrange(desc(largest_pack))
```

```
## # A tibble: 5 × 2
##   park   largest_pack
##   <chr>         <dbl>
## 1 YNP             37
## 2 DENA            33
## 3 GNTTP          26.4
## 4 YUCH            24
## 5 VNP              7
```

#SAVE AS PDF AND SAVE AND SUBMIT TO GRADESCOPE