Midterm 1 W24

Jacques Mak

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.

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?

```
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.

```
summary (wolves)
```

```
##      park               biolyr         pack              packcode
##  Length:864         Min.   :1986   Length:864         Min.   :  2.00
##  Class :character   1st Qu.:1999   Class :character   1st Qu.: 48.00
##  Mode  :character   Median :2006   Mode  :character   Median : 86.50
##                     Mean   :2005                      Mean   : 91.39
##                     3rd Qu.:2012                      3rd Qu.:133.00
##                     Max.   :2021                      Max.   :193.00
##
##   packsize_aug       mort_yn          mort_all         mort_lead
##  Min.   : 0.000   Min.   :0.0000   Min.   : 0.0000   Min.   :0.00000
##  1st Qu.: 5.000   1st Qu.:0.0000   1st Qu.: 0.0000   1st Qu.:0.00000
##  Median : 8.000   Median :0.0000   Median : 0.0000   Median :0.00000
##  Mean   : 8.789   Mean   :0.1956   Mean   : 0.3715   Mean   :0.09552
##  3rd Qu.:12.000   3rd Qu.:0.0000   3rd Qu.: 0.0000   3rd Qu.:0.00000
##  Max.   :37.000   Max.   :1.0000   Max.   :24.0000   Max.   :3.00000
##  NA's   :55                                          NA's   :16
##   mort_nonlead        reprody1         persisty1
##  Min.   : 0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.: 0.0000   1st Qu.:1.0000   1st Qu.:1.0000
##  Median : 0.0000   Median :1.0000   Median :1.0000
##  Mean   : 0.2641   Mean   :0.7629   Mean   :0.8865
##  3rd Qu.: 0.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :22.0000   Max.   :1.0000   Max.   :1.0000
##  NA's   :12        NA's   :71       NA's   :9
```

```
glimpse (wolves)
```

```
## Rows: 864
## Columns: 11
## $ park         <chr> "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, …
```

```
str(wolves)
```

```
## 'data.frame':    864 obs. of  11 variables:
##  $ park        : chr  "DENA" "DENA" "DENA" "DENA" ...
##  $ biolyr      : int  1996 1991 2017 1996 1992 1994 2007 2007 1995 2003 ...
##  $ pack        : chr  "McKinley River1" "Birch Creek N" "Eagle Gorge" "East Fork" ...
##  $ packcode    : int  89 58 71 72 74 77 101 108 109 53 ...
##  $ packsize_aug: num  12 5 8 13 7 6 10 NA 9 8 ...
##  $ mort_yn     : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ mort_all    : int  4 2 2 2 2 2 2 2 2 1 ...
##  $ mort_lead   : int  2 2 0 0 0 0 1 2 1 1 ...
##  $ mort_nonlead: int  2 0 2 2 2 2 1 0 1 0 ...
##  $ reprody1    : int  0 0 NA 1 NA 0 0 1 0 1 ...
##  $ persisty1   : int  0 0 1 1 1 1 0 1 0 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)
```

```
wolves %>%
  summarize(distinct_park = n_distinct(park))
```

```
##   distinct_park
## 1             5
```

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

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

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

```
wolves%>%
  group_by(park)%>%
  summarize(num_wolf=sum(packsize_aug, na.rm=T))%>%
  arrange(desc(num_wolf))
```

```
## # A tibble: 5 × 2
##   park  num_wolf
##   <fct>    <dbl>
## 1 YNP       2731
## 2 DENA      2500
## 3 YUCH      1048
## 4 GNTP       781.
## 5 VNP         50
```

Park YNP has the largest size of wolf packs.

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

```
wolves%>%
  group_by(park)%>%
  summarize(mort_human=sum(mort_all, na.rm=T))%>%
  arrange(desc(mort_human))
```

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

YUCH park has the highest total number of human-caused mortalities.

The wolves in Yellowstone National Park 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 <- filter(wolves, park=="YNP")
```

Problem 7. (3 points) Among the Yellowstone wolf packs, the Druid Peak Pack is one of most famous. What was the average pack size of this pack for the years represented in the data?

```
mean_druid_peak_pack<-filter(ynp, pack=="druid")%>%
  summarize(mean_druid=mean(packsize_aug, na.rm = T))
```

The mean is 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?

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

At year 2001 they have the most. The packs of wolves either migrated to another area or they all died out at 2010.

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(total_persisty=sum(persisty1, na.rm = T)) %>%
  arrange(desc(total_persisty))
```

```
## # A tibble: 46 × 2
##    pack        total_persisty
##    <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 36 more rows
```

Mollies has the highest overall persistence.

Deep snow helps weaken bison, making it harder for them to feed and move. Wolves chase bison into deep snow where they're easier to attack without the wolves being kicked or charged.

The 19-member Mollie's pack – the largest in the park – has become well-known for its adaptations to killing bison. To bring down the 1,000- to 2,000-pound beasts, Mollie's has nurtured some of the park's largest wolves.

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.
What was the average pack size of Druid Peak Pack between year 1999 and 2002 represented in the data? The mean from year 1999 to 2002 is 22.25.

```
secmean_druid_peak_pack<-filter(ynp, pack=="druid")%>%
  filter(between(biolyr, 1999,2002))%>%
  summarize(mean_druid=mean(packsize_aug, na.rm = T))
```