Midterm 1 W24

Eric Du

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")
```

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"
```

The names of the variables are: "park", "biolyr", "pack", "packcode", "packsize_aug", "mort_yn", "mort_all", "mort_lead", "mort_nonlead", "reprody1", and "persisty1".

Problem 2. (1 point) Use the function of your choice to summarize the data and get an idea of its structure.

```
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 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%>%
  group_by(park)%>%
  summarise(totalnumber=sum(packsize_aug, na.rm = T))
## # A tibble: 5 × 2
##   park  totalnumber
##   <chr>       <dbl>
## 1 DENA         2500
## 2 GNTP          781.
## 3 VNP            50
## 4 YNP          2731
```

```
## 5 YUCH        1048
```

There are five parks: DENA(Denali National Park and Preserve), GTNP(Grand Teton National Park), VNP(Voyageurs National Park), YNP(Yellowstone National Park), and YUCH(Yukon-Charley Rivers National Preserve).

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

```
wolves%>%
  group_by(park)%>%
  summarise(maxnumber_pack=max(packsize_aug, na.rm = T))
## # A tibble: 5 × 2
##   park  maxnumber_pack
##   <chr>          <dbl>
## 1 DENA            33
## 2 GNTP            26.4
## 3 VNP              7
## 4 YNP             37
## 5 YUCH            24
```

The park YNP has the largest number of pack, which is 37.

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

```
wolves%>%
  group_by(park)%>%
  summarise(highestnum_morall=max(mort_all, na.rm = T))
## # A tibble: 5 × 2
##   park  highestnum_morall
##   <chr>             <int>
## 1 DENA                  4
## 2 GNTP                  4
## 3 VNP                   2
## 4 YNP                   4
## 5 YUCH                 24
```

The park YUCH has the highest total number of human-caused mortalities, which is 24.

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?

```
druid<-filter(ynp,pack=="druid")

mean(druid$packsize_aug)

## [1] 13.93333
```

The average pack size of druid is about 14.

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%>%
  group_by(biolyr)%>%
  arrange(desc(packsize_aug))
## # A tibble: 15 × 11
## # Groups:   biolyr [15]
##    park  biolyr pack  packcode packsize_aug mort_yn mort_all mort_lead
##    <chr>  <int> <chr>    <int>        <dbl>   <int>    <int>     <int>
##  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
## # ℹ 3 more variables: mort_nonlead <int>, reprody1 <int>, persisty1 <int>
```

Since only YNP has the pack Druid, we could use the data from YNP directly. On 2001, the number of Druid pack has the highest number. Since on 2009, the “peristy1” of the pack is 0, which means that the pack is no longer on the territory anymore, causing the packsize on 2010 dropped to 0. The pack may have been dissolved or replaced by other pack.

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)%>%
  summarise(numberofpersist=n())%>%
  arrange(desc(numberofpersist))
## # A tibble: 38 × 2
##    pack        numberofpersist
##    <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 pack mollies has the highest overall persistence of 26 years. From https://www.spokesman.com/stories/2012/jan/15/hungry-wolf-pack-rearranges-balance-in/, we could know that mollies pack was known for being able to hunt down large beasts like bison with the help of deep snow.

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(pack)%>%
  summarise(numberofparks=n_distinct(park))%>%
  arrange(desc(numberofparks))
## # A tibble: 184 × 2
```

```
##    pack       numberofparks
##    <chr>              <int>
##  1 Flat Creek             2
##  2 100 Mile               1
##  3 1118Fgroup             1
##  4 1155Mgroup             1
##  5 642Fgroup              1
##  6 682Mgroup              1
##  7 694Fgroup              1
##  8 70 Mile                1
##  9 755Mgroup              1
## 10 8mile                  1
## # ℹ 174 more rows
```

I want to investigate if there's any pack been to multiple parks and if yes, which pack been to the most. It turns out that most packs have only been in one park. Only the pack "Flat Creek" has been to two parks.