Midterm 1 W24

Christina Chen

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

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

These are the abrievieated names.

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

```
wolves %>%
  group_by(park) %>%
  summarize(distinct_packs = n_distinct(pack))
```

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

It looks like DENA has the largest number of (distinct) packs.

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

```
wolves %>%
  group_by(park) %>%
  summarise(mean_mort_all = mean(mort_all),
            max_mort_all = max(mort_all),
            min_mort_all = min(mort_all))
```

```
## # A tibble: 5 × 4
##   park  mean_mort_all max_mort_all min_mort_all
##   <fct>         <dbl>        <int>        <int>
## 1 DENA          0.188            4            0
## 2 GNTP          0.494            4            0
## 3 VNP           0.229            2            0
## 4 YNP           0.290            4            0
## 5 YUCH          0.901           24            0
```

```
wolves %>%
  select(park, mort_all) %>%
  arrange(desc(mort_all)) %>%
  head()
```

```
##   park mort_all
## 1 YUCH       24
## 2 YUCH       14
## 3 YUCH       12
## 4 YUCH       12
## 5 YUCH       10
## 6 YUCH        8
```

The park YUCH has the largest mean mortality across all years, and the largest mortality for a single year. With this, we can infer that it has the largest total mortality as well. But if you don't want to infer, you can also do the longform way:

```
dena <- wolves %>%
  filter(park == 'DENA')
gntp <- wolves %>%
  filter(park == 'GNTP')
vnp <- wolves %>%
  filter(park == 'VNP')
ynp <- wolves %>%
  filter(park == 'YNP')
yuch <- wolves %>%
  filter(park == 'YUCH')
```

```
sum(dena$mort_all)
```

```
## [1] 64
```

```
sum(gntp$mort_all)
```

```
## [1] 38
```

```
sum(vnp$mort_all)
```

```
## [1] 11
```

```
sum(ynp$mort_all)
```

```
## [1] 72
```

```
sum(yuch$mort_all)
```

```
## [1] 136
```

This also shows that the YUCH park has the highest total mortality.

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')
# I have already created this object, but I am placing the code here for reference.
head(ynp)
```

```
##   park biolyr       pack packcode packsize_aug mort_yn mort_all mort_lead
## 1  YNP   2009 cottonwood       23           12       1        4         1
## 2  YNP   2016      8mile       11           20       1        3         0
## 3  YNP   2017     canyon       20            2       1        3         3
## 4  YNP   2012   junction       33           11       1        3         0
## 5  YNP   2016   junction       33           15       1        3         0
## 6  YNP   2011  642Fgroup        5           10       1        2         1
##   mort_nonlead reprody1 persisty1
## 1            3        0         0
## 2            3        1         1
## 3            0        0         0
## 4            3        1         1
## 5            3        1         1
## 6            1        0         0
```

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') %>%
  summarise(mean_packsize_aug = mean(packsize_aug, na.rm = TRUE))
```

```
##   mean_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?

```
ynp %>%
  filter(pack == 'druid') %>%
  arrange(desc(packsize_aug))
```

```
##    park biolyr  pack packcode packsize_aug mort_yn mort_all mort_lead
## 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
##    mort_nonlead reprody1 persisty1
## 1             0        1         1
## 2             1        1         1
## 3             0        1         1
## 4             0        1         1
## 5             0        1         1
## 6             0        1         1
## 7             0        1         1
## 8             0        1         1
## 9             0        0         0
## 10            0        1         1
## 11            0        1         1
## 12            1        1         1
## 13            0        1         1
## 14            0        1         1
## 15            0        0        NA
```

The pack was largest in 2001, but in 2010, it looks like the pack died out. The pack size was 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?

```
ynp %>%
  group_by(pack) %>%
  filter(persisty1 == 1) %>%
  count() %>%
  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
```

It looks like the mollies have the highest overall persistence, with more 1s than any other pack. This may be caused by their unique habit of group hunting bison when elk migrated out of the area. This pack also seems to have more female alphas and longer lifespans.

First Source Link (https://www.yellowstonewolf.org/yellowstones_wolves.php?pack_id=6)

Second Source Link (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.

Here is the two parks with the largest number of packs, with data only from between 2000 and 2010. The data has been sorted in descending order by pack size, and reduced to the top 6 results.

```
wolves %>%
  filter(park == "DENA" | park == "GNTP") %>%
  filter(between(biolyr, 2000, 2010)) %>%
  arrange(desc(packsize_aug)) %>%
  head()
```

```
##   park biolyr            pack packcode packsize_aug mort_yn mort_all mort_lead
## 1 GNTP   2009         Buffalo      158         26.4       1        1        NA
## 2 DENA   2000       100 Mile       53         23.0       0        0         0
## 3 DENA   2010 McKinley Slough       90         19.0       0        0         0
## 4 GNTP   2007         Buffalo      158         18.0       0        0         0
## 5 DENA   2000     Pinto Creek      102         17.0       0        0         0
## 6 DENA   2001     Pinto Creek      102         17.0       0        0         0
##   mort_nonlead reprody1 persisty1
## 1           NA        0         1
## 2            0        1         1
## 3            0        1         1
## 4            0        1         1
## 5            0        1         1
## 6            0        1         1
```