Midterm 1 W25

Joyce

2025-02-04

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 or other students' work.

Don't forget to answer any questions that are asked in the prompt! Each question must be coded; it cannot be answered by a sort in a spreadsheet or a written response.

Be sure to push your completed midterm to your repository and upload the document to Gradescope. This exam is worth 30 points.

Please load the following libraries.

```
library(tidyverse)
library(janitor)
```

Disable scientific notation.

```
options(scipen=999)
```

In the midterm 1 folder there is a second folder called `data`. Inside the `data` folder, there is a .csv file called `ecs21351-sup-0003-SupplementS1.csv`. These data are from Soykan, C. U., J. Sauer, J. G. Schuetz, G. S. LeBaron, K. Dale, and G. M. Langham. 2016. Population trends for North American winter birds based on hierarchical models. Ecosphere 7(5):e01351. 10.1002/ecs2.1351. This study uses the CBC (Christmas Bird Count) data to estimate population trends for North American winter birds.

Please load these data as a new object called `ecosphere`. In this step, I am providing the code to load the data, clean the variable names, and remove a footer that the authors used as part of the original publication.

```
ecosphere <- read_csv("data/ecs21351-sup-0003-SupplementS1.csv", skip=2) %>%
  #load the data and skip the first two rows
  clean_names() %>%
  #clean the variable names
  slice(1:(n() - 18))
  #remove the footer
```

Questions

Problem 1. (1 point) What are the variable names?

```
names(ecosphere)
```

```
##  [1] "order"                       "family"
##  [3] "common_name"                 "scientific_name"
##  [5] "diet"                        "life_expectancy"
##  [7] "habitat"                     "urban_affiliate"
##  [9] "migratory_strategy"          "log10_mass"
## [11] "mean_eggs_per_clutch"        "mean_age_at_sexual_maturity"
## [13] "population_size"             "winter_range_area"
## [15] "range_in_cbc"                "strata"
## [17] "circles"                     "feeder_bird"
## [19] "median_trend"                "lower_95_percent_ci"
## [21] "upper_95_percent_ci"
```

Problem 2. (1 point) Use the function of your choice to provide a data summary.

```
summary(ecosphere)
```

```
##     order              family           common_name        scientific_name   
##  Length:551         Length:551         Length:551         Length:551        
##  Class :character   Class :character   Class :character   Class :character  
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character  
##                                                                             
##                                                                             
##                                                                             
##                                                                             
##      diet           life_expectancy      habitat          urban_affiliate   
##  Length:551         Length:551         Length:551         Length:551        
##  Class :character   Class :character   Class :character   Class :character  
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character  
##                                                                             
##                                                                             
##                                                                             
##                                                                             
##  migratory_strategy   log10_mass    mean_eggs_per_clutch
##  Length:551         Min.   :0.480   Min.   : 1.000      
##  Class :character   1st Qu.:1.365   1st Qu.: 3.000      
##  Mode  :character   Median :1.890   Median : 4.000      
##                     Mean   :2.012   Mean   : 4.527      
##                     3rd Qu.:2.685   3rd Qu.: 5.000      
##                     Max.   :4.040   Max.   :17.000      
##                                                         
##  mean_age_at_sexual_maturity population_size     winter_range_area  
##  Min.   : 0.200              Min.   :    15000   Min.   :       11  
##  1st Qu.: 1.000              1st Qu.:  1100000   1st Qu.:   819357  
##  Median : 1.000              Median :  4900000   Median :  2189639  
##  Mean   : 1.592              Mean   : 18446745   Mean   :  5051047  
##  3rd Qu.: 2.000              3rd Qu.: 18000000   3rd Qu.:  6778598  
##  Max.   :12.500              Max.   :300000000   Max.   :185968946  
##                              NA's   :273                            
##   range_in_cbc        strata          circles       feeder_bird       
##  Min.   :  0.00   Min.   :  1.00   Min.   :   2.0   Length:551        
##  1st Qu.:  2.35   1st Qu.:  3.00   1st Qu.:  46.5   Class :character  
##  Median : 30.30   Median : 11.00   Median : 184.0   Mode  :character  
##  Mean   : 38.48   Mean   : 32.43   Mean   : 558.9                     
##  3rd Qu.: 72.95   3rd Qu.: 42.00   3rd Qu.: 661.0                     
##  Max.   :100.00   Max.   :159.00   Max.   :3202.0                     
##                                                                       
##   median_trend   lower_95_percent_ci upper_95_percent_ci
##  Min.   :0.739   Min.   :0.5780      Min.   :    0.798  
##  1st Qu.:0.993   1st Qu.:0.9675      1st Qu.:    1.011  
##  Median :1.009   Median :0.9930      Median :    1.027  
##  Mean   :1.016   Mean   :0.9857      Mean   :   33.709  
##  3rd Qu.:1.030   3rd Qu.:1.0140      3rd Qu.:    1.055  
##  Max.   :1.396   Max.   :1.3080      Max.   :18000.000  
##                                                         
```

```
glimpse(ecosphere)
```

```
## Rows: 551
## Columns: 21
## $ order                       <chr> "Anseriformes", "Anseriformes", "Anserifor…
## $ family                      <chr> "Anatidae", "Anatidae", "Anatidae", "Anati…
## $ common_name                 <chr> "American Black Duck", "American Wigeon", …
## $ scientific_name             <chr> "Anas rubripes", "Anas americana", "Buceph…
## $ diet                        <chr> "Vegetation", "Vegetation", "Invertebrates…
## $ life_expectancy             <chr> "Long", "Middle", "Middle", "Long", "Middl…
## $ habitat                     <chr> "Wetland", "Wetland", "Wetland", "Wetland"…
## $ urban_affiliate             <chr> "No", "No", "No", "No", "No", "No", "No", …
## $ migratory_strategy          <chr> "Short", "Short", "Moderate", "Moderate", …
## $ log10_mass                  <dbl> 3.09, 2.88, 2.96, 3.11, 3.02, 2.88, 2.56, …
## $ mean_eggs_per_clutch        <dbl> 9.0, 7.5, 10.5, 3.5, 9.5, 13.5, 10.0, 8.5,…
## $ mean_age_at_sexual_maturity <dbl> 1.0, 1.0, 3.0, 2.5, 2.0, 1.0, 0.6, 2.0, 1.…
## $ population_size             <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA…
## $ winter_range_area           <dbl> 3212473, 7145842, 1812841, 360134, 854350,…
## $ range_in_cbc                <dbl> 99.1, 61.7, 69.8, 53.7, 5.3, 0.5, 17.9, 72…
## $ strata                      <dbl> 82, 124, 37, 19, 36, 5, 26, 134, 145, 103,…
## $ circles                     <dbl> 1453, 1951, 502, 247, 470, 97, 479, 2189, …
## $ feeder_bird                 <chr> "No", "No", "No", "No", "No", "No", "No", …
## $ median_trend                <dbl> 1.014, 0.996, 1.039, 0.998, 1.004, 1.196, …
## $ lower_95_percent_ci         <dbl> 0.971, 0.964, 1.016, 0.956, 0.975, 1.152, …
## $ upper_95_percent_ci         <dbl> 1.055, 1.009, 1.104, 1.041, 1.036, 1.243, …
```

Problem 3. (2 points) How many distinct orders of birds are represented in the data?

```
ecosphere$order %>% unique()
```

```
##  [1] "Anseriformes"      "Apodiformes"       "Caprimulgiformes"
##  [4] "Charadriiformes"   "Ciconiiformes"     "Columbiformes"
##  [7] "Coraciiformes"     "Cuculiformes"      "Falconiformes"
## [10] "Galliformes"       "Gaviiformes"       "Gruiformes"
## [13] "Passeriformes"     "Piciformes"        "Podicipediformes"
## [16] "Procellariiformes" "Psittaciformes"    "Strigiformes"
## [19] "Trogoniformes"
```

```
ecosphere %>%
  summarize(n_order=n_distinct(order))
```

```
## # A tibble: 1 × 1
##   n_order
##     <int>
## 1      19
```

Problem 4. (2 points) Which habitat has the greatest species diversity?

```
ecosphere %>%
  group_by(habitat) %>%
  summarize(n_species=n_distinct(scientific_name)) %>%
  arrange(-n_species)
```

```
## # A tibble: 7 × 2
##   habitat   n_species
##   <chr>         <int>
## 1 Woodland        177
## 2 Wetland         153
## 3 Shrubland        82
## 4 Various          45
## 5 Ocean            44
## 6 Grassland        36
## 7 <NA>             14
```

Problem 5. (2 points) For species associated with urban environments, what is the min, max, and mean winter range area?

```
ecosphere %>%
  filter(urban_affiliate=="Yes") %>%
  summarize(min_wra=min(winter_range_area, na.rm=T), max_wra=max(winter_range_area, na.r
m = T), mean_wra=mean(winter_range_area, na.rm=T))
```

```
## # A tibble: 1 × 3
##   min_wra  max_wra mean_wra
##     <dbl>    <dbl>    <dbl>
## 1     193 26419123 5969323.
```

Problem 6. (2 points) As part of our analysis, we need `mass_g` as a new variable. Please convert `log10_mass` to mass in grams (hint: mass_g=10^log10_mass) and store the output as part of the `ecosphere` data.

```
ecosphere <- ecosphere %>%
  mutate(mass_g=10^log10_mass)
```

Problem 7. (4 points) Which migratory strategy has the highest average mass (mass_g)?

```
ecosphere %>%
  group_by(migratory_strategy) %>%
  summarize(mean_mass_g=mean(mass_g, na.rm=T)) %>%
  arrange(-mean_mass_g)
```

```
## # A tibble: 6 × 2
##   migratory_strategy mean_mass_g
##   <chr>                    <dbl>
## 1 Moderate                  523.
## 2 Short                     493.
## 3 Withdrawal                480.
## 4 Resident                  435.
## 5 Irruptive                 371.
## 6 Long                      306.
```

Problem 8. (4 points) Irruptive migratory behavior is characterized by unpredictable movements in response to food availability. What is the average population size for species with irruptive migratory behavior, grouped by habitat and diet?

```
ecosphere %>%
  group_by(habitat, diet) %>%
  filter(migratory_strategy=="Irruptive") %>%
  summarize(ave_popu_size=mean(population_size, na.rm=T), .groups = "keep") %>%
  arrange(-ave_popu_size)
```

```
## # A tibble: 7 × 3
## # Groups:   habitat, diet [7]
##   habitat   diet        ave_popu_size
##   <chr>     <chr>               <dbl>
## 1 Shrubland Seed             31500000
## 2 Woodland  Fruit            27000000
## 3 Woodland  Seed             21500000
## 4 Woodland  Omnivore          3900000
## 5 Various   Seed               300000
## 6 Grassland Vertebrates         70000
## 7 Woodland  Vertebrates         60000
```

Problem 9. (4 points). Diet, life expectancy, urban_affiliate, and migratory_strategy are all variables associated with extinction risk or population decline. Which species have a combination of vertebrate diet, long life expectancy, no urban affiliation, and are long-distance migrants? Assuming that the bird with the highest mass is the most at risk, which is the species of greatest concern?

```
ecosphere %>%
  filter(diet=="Vertebrates" & life_expectancy=="Long" & urban_affiliate=="No" & migrato
ry_strategy=="Long") %>%
  select(common_name, scientific_name, mass_g) %>%
  arrange(-mass_g)
```

```
## # A tibble: 3 × 3
##   common_name              scientific_name       mass_g
##   <chr>                    <chr>                  <dbl>
## 1 Black-footed Albatross   Phoebastria nigripes   2818.
## 2 Sooty Shearwater         Puffinus griseus        794.
## 3 Short-tailed Shearwater Puffinus tenuirostris   562.
```

Problem 10. (4 points). Make a new column `conservation_status` that labels species with a population size less than 300,000 as “threatened” and species with a population size greater than 300,000 as “stable”. Make sure your results are sorted in descending order. How many species are threatened vs. stable? Based on the results, do you see a problem with this analysis?

```
ecosphere %>%
  mutate(conservation_status=ifelse(population_size<300000,"threatened", "NA")) %>%
  mutate(conservation_status=ifelse(population_size>300000,"stable", conservation_statu
s)) %>%
  arrange(-population_size)
```

```
## # A tibble: 551 × 23
##    order        family common_name scientific_name diet  life_expectancy habitat
##    <chr>        <chr>  <chr>       <chr>           <chr> <chr>           <chr>
##  1 Passeriform… Turdi… American R… Turdus migrato… Fruit Middle          Woodla…
##  2 Passeriform… Ember… Chipping S… Spizella passe… Seed  Short           Woodla…
##  3 Passeriform… Ember… Dark-eyed … Junco hyemalis  Seed  Middle          Woodla…
##  4 Passeriform… Ember… Savannah S… Passerculus sa… Omni… Short           Grassl…
##  5 Passeriform… Ember… White-thro… Zonotrichia al… Seed  Short           Woodla…
##  6 Passeriform… Ember… Song Sparr… Melospiza melo… Omni… Middle          Various
##  7 Passeriform… Parul… Yellow-rum… Dendroica coro… Inve… Short           Woodla…
##  8 Passeriform… Icter… Red-winged… Agelaius phoen… Omni… Middle          Various
##  9 Passeriform… Icter… Brown-head… Molothrus ater  Omni… Middle          Various
## 10 Passeriform… Polio… Blue-gray … Polioptila cae… Inve… Short           Woodla…
## # ℹ 541 more rows
## # ℹ 16 more variables: urban_affiliate <chr>, migratory_strategy <chr>,
## #   log10_mass <dbl>, mean_eggs_per_clutch <dbl>,
## #   mean_age_at_sexual_maturity <dbl>, population_size <dbl>,
## #   winter_range_area <dbl>, range_in_cbc <dbl>, strata <dbl>, circles <dbl>,
## #   feeder_bird <chr>, median_trend <dbl>, lower_95_percent_ci <dbl>,
## #   upper_95_percent_ci <dbl>, mass_g <dbl>, conservation_status <chr>
```

```
ecosphere %>%
  mutate(conservation_status=ifelse(population_size<300000,"threatened", "NA")) %>%
  mutate(conservation_status=ifelse(population_size>300000,"stable", conservation_statu
s)) %>%
  arrange(-population_size) %>%
  group_by(conservation_status) %>%
  summarize(n_speices=n_distinct(scientific_name))
```

```
## # A tibble: 4 × 2
##   conservation_status n_speices
##   <chr>                   <int>
## 1 NA                          6
## 2 stable                    244
## 3 threatened                 28
## 4 <NA>                      273
```

```
ecosphere %>%
  mutate(conservation_status=ifelse(population_size<300000,"threatened", "NA")) %>%
  mutate(conservation_status=ifelse(population_size>300000,"stable", conservation_statu
s)) %>%
  filter(population_size==300000 | is.na(population_size))
```

```
## # A tibble: 279 × 23
##    order        family common_name scientific_name diet  life_expectancy habitat
##    <chr>        <chr>  <chr>       <chr>           <chr> <chr>           <chr>
##  1 Anseriformes Anati… "American … Anas rubripes   Vege… Long            Wetland
##  2 Anseriformes Anati… "American … Anas americana  Vege… Middle          Wetland
##  3 Anseriformes Anati… "Barrow's … Bucephala isla… Inve… Middle          Wetland
##  4 Anseriformes Anati… "Black Bra… Branta bernicla Vege… Long            Wetland
##  5 Anseriformes Anati… "Black Sco… Melanitta amer… Inve… Middle          Wetland
##  6 Anseriformes Anati… "Black-bel… Dendrocygna au… Vege… Short           Wetland
##  7 Anseriformes Anati… "Blue-wing… Anas discors    Vege… Middle          Wetland
##  8 Anseriformes Anati… "Bufflehea… Bucephala albe… Inve… Middle          Wetland
##  9 Anseriformes Anati… "Cackling … Branta hutchin… Vege… Middle          Wetland
## 10 Anseriformes Anati… "Canvasbac… Aythya valisin… Vege… Middle          Wetland
## # ℹ 269 more rows
## # ℹ 16 more variables: urban_affiliate <chr>, migratory_strategy <chr>,
## #   log10_mass <dbl>, mean_eggs_per_clutch <dbl>,
## #   mean_age_at_sexual_maturity <dbl>, population_size <dbl>,
## #   winter_range_area <dbl>, range_in_cbc <dbl>, strata <dbl>, circles <dbl>,
## #   feeder_bird <chr>, median_trend <dbl>, lower_95_percent_ci <dbl>,
## #   upper_95_percent_ci <dbl>, mass_g <dbl>, conservation_status <chr>
```

Problem 11. (4 points) Use the `ecosphere` data to perform one exploratory analysis of your choice. The analysis must have a minimum of three lines and two functions. You must also clearly state the question you are attempting to answer.

What is the average population size within the woodland habitat, grouped by diet and life_expectancy?

```
ecosphere %>%
  filter(population_size!="NA" & habitat=="Woodland") %>%
  group_by(diet, life_expectancy) %>%
  summarize(ave_pop_size=mean(population_size, na.rm=T), .groups="keep") %>%
  arrange(-ave_pop_size)
```

```
## # A tibble: 13 × 3
## # Groups:   diet, life_expectancy [13]
##    diet          life_expectancy ave_pop_size
##    <chr>         <chr>                  <dbl>
##  1 Fruit         Middle            102033333.
##  2 Seed          Short              44845455.
##  3 Seed          Middle             41495000
##  4 Fruit         Short              26550000
##  5 Invertebrates Short              20492727.
##  6 Invertebrates Middle             13610667.
##  7 Nectar        Short               9250000
##  8 Omnivore      Short               8255333.
##  9 Omnivore      Middle              6754118.
## 10 Nectar        Middle              5000000
## 11 Vertebrates   Long                2005000
## 12 Vertebrates   Middle              1073750
## 13 Vertebrates   Short                 60000
```

Submit the Midterm

1. Save your work and knit the .rmd file.
2. Open the .html file and “print” it to a .pdf file in Google Chrome (not Safari).
3. Go to the class Canvas page and open Gradescope.
4. Submit your .pdf file to the midterm assignment- be sure to assign the pages to the correct questions.
5. Commit and push your work to your repository.