# Midterm 1 W25

Luc-Tanton Tran

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  
##                                                         
```

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

```
ecosphere %>%
  select(order) %>%
  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? (meaning number of species)

```
ecosphere %>%
  group_by(habitat) %>%
  summarize(n_species = n())
```

```
## # A tibble: 7 × 2
##   habitat   n_species
##   <chr>         <int>
## 1 Grassland        36
## 2 Ocean            44
## 3 Shrubland        82
## 4 Various          45
## 5 Wetland         153
## 6 Woodland        177
## 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_winter_range_area = min(winter_range_area),
    max_winter_range_area = max(winter_range_area),
    mean_winter_range_area = mean(winter_range_area)
  )
```

```
## # A tibble: 1 × 3
##   min_winter_range_area max_winter_range_area mean_winter_range_area
##                   <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(average_mass_g = mean(mass_g)) %>%
  arrange(desc(average_mass_g)) #for legibility/ease of reading
```

```
## # A tibble: 6 × 2
##   migratory_strategy average_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 %>%
  filter(migratory_strategy == "Irruptive") %>%
  group_by(habitat, diet) %>%
  summarize(
    average_population_size = mean(population_size, na.rm=T),
    .groups = 'keep'
  )
```

```
## # A tibble: 7 × 3
## # Groups:   habitat, diet [7]
##   habitat   diet        average_population_size
##   <chr>     <chr>                         <dbl>
## 1 Grassland Vertebrates                   70000
## 2 Shrubland Seed                       31500000
## 3 Various   Seed                         300000
## 4 Woodland  Fruit                      27000000
## 5 Woodland  Omnivore                    3900000
## 6 Woodland  Seed                       21500000
## 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 %>%
  select(common_name,
         scientific_name,
         diet,
         life_expectancy,
         urban_affiliate,
         migratory_strategy,
         mass_g) %>%
  filter(diet == "Vertebrates",
         life_expectancy == "Long",
         urban_affiliate == "No",
         migratory_strategy == "Long") %>%
  filter(mass_g == max(mass_g)) #selecting for the one at highest risk (highest mass)
```

```
## # A tibble: 1 × 7
##   common_name            scientific_name   diet  life_expectancy urban_affiliate
##   <chr>                  <chr>             <chr> <chr>           <chr>
## 1 Black-footed Albatross Phoebastria nigr… Vert… Long            No
## # ℹ 2 more variables: migratory_strategy <chr>, mass_g <dbl>
```

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, "stable", "threatened")) %>%
  arrange(desc(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>
```

```
#Problem: about half of the species listed don't have a population
# size ("NA"), and are unable to be categorized as either "stable" or "threatened".
```

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.

```
# Proposed Question:
#   Within the species with short life spans, which habitat
#   has the highest mean_age_at_sexual_maturity?
#   Assume that the only variables of import are
#   life_expectancy, habitat, and mean_age_at_sexual_maturity,
#   and include n for each habitat.

ecosphere %>%
  select(life_expectancy, habitat, mean_age_at_sexual_maturity) %>%
  filter(life_expectancy == "Short") %>%
  group_by(habitat) %>%
  summarize(
    max_mean_age_at_sexual_maturity = max(mean_age_at_sexual_maturity, na.rm=T),
    n = n(),) %>%
  arrange(-max_mean_age_at_sexual_maturity)
```

```
## # A tibble: 7 × 3
##   habitat   max_mean_age_at_sexual_maturity     n
##   <chr>                               <dbl> <int>
## 1 Ocean                                 4       4
## 2 Wetland                               3      48
## 3 Shrubland                             2.5    61
## 4 Grassland                             2      25
## 5 Various                               2      21
## 6 Woodland                              2     101
## 7 <NA>                                  1       5
```

# 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.