Challenge 1 Instructions

Meredith Rolfe

08/15/2022

```
#| label: setup
#| warning: false
#| message: false

library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --v ggplot2 3.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.0
## v readr   2.1.3      v forcats 0.5.1 -- Conflicts ------------------------------------------ tidyvers
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(dplyr)
library("readxl")

knitr::opts_chunk$set(echo = TRUE, warning=FALSE, message=FALSE)
```

Challenge Overview

Today's challenge is to

1) read in a dataset, and
2) describe the dataset using both words and any supporting information (e.g., tables, etc)

Read in the Data

Read in one (or more) of the following data sets, using the correct R package and command.

Find the `_data` folder, located inside the `posts` folder. Then you can read in the data, using either one of the `readr` standard tidy read commands, or a specialized package such as `readxl`.

```
#Read the railroad_2012_clean_county csv file
clean_county_data = read.csv('_data/railroad_2012_clean_county.csv')

#Get dimensions
dim(clean_county_data)
```

```
## [1] 2930    3
```

```
#The dataset has 2930 rows and 3 columns

#Get column names
colnames(clean_county_data)
```

```
## [1] "state"           "county"          "total_employees"
```

```
#Displaying top 5 columns of dataframe
head(clean_county_data, n=5)
```

```
##   state               county total_employees
## 1    AE                  APO               2
## 2    AK            ANCHORAGE               7
## 3    AK FAIRBANKS NORTH STAR               2
## 4    AK               JUNEAU               3
## 5    AK    MATANUSKA-SUSITNA               2
```

```
#Grouping dataframe by state and county, to get total employees employed per state per county
#Displaying only county having highest number of employees
suppressWarnings(
clean_county_data %>%
  group_by(state, county) %>%
  summarise_each(funs(sum)) %>%
  arrange(state, county, desc(total_employees)) %>%
  slice(1))
```

```
## # A tibble: 53 x 3
## # Groups:   state [53]
##    state county        total_employees
##    <chr> <chr>                   <int>
##  1 AE    APO                         2
##  2 AK    ANCHORAGE                   7
##  3 AL    AUTAUGA                   102
##  4 AP    APO                         1
##  5 AR    ARKANSAS                   11
##  6 AZ    APACHE                    270
##  7 CA    ALAMEDA                   346
##  8 CO    ADAMS                     553
##  9 CT    FAIRFIELD                 486
## 10 DC    WASHINGTON DC             279
## # ... with 43 more rows
```

Add any comments or documentation as needed. More challenging data sets may require additional code chunks and documentation.

Describe the data

Using a combination of words and results of R commands, can you provide a high level description of the data? Describe as efficiently as possible where/how the data was (likely) gathered, indicate the cases and variables (both the interpretation and any details you deem useful to the reader to fully understand your chosen data).

The clean_county_data dataset looks to maintain information about the number of individuals employed per county in the United States within the railroad department, for the year 2012.
The dataset contains a total of 2930 rows, with 3 columns, namely 'state', 'county' and 'total_employees'.
This data was perhaps gathered via a census carried out in the year 2012 of all employees in the railroad department throughout the United States, or could also be gathered via historical data maintained by the department, as a subspace of the particular year (i.e. 2012).

```
#| label: summary
#Read the wild_bird_data xlsx file
```

```
wild_bird_data = read_excel('_data/wild_bird_data.xlsx', skip = 1)

#Get dimensions
dim(wild_bird_data)
```

```
## [1] 146   2
```

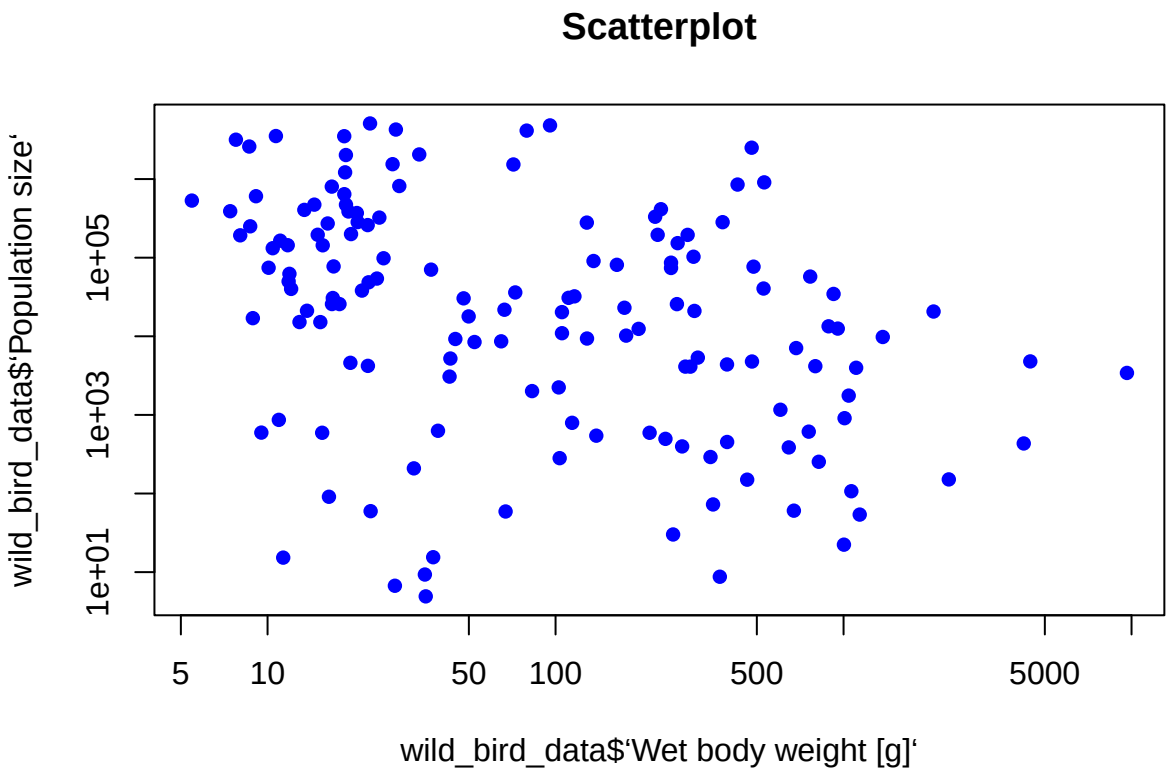
```
#Get column names
colnames(wild_bird_data)
```

```
## [1] "Wet body weight [g]" "Population size"
```

```
#Displaying top 5 columns of dataframe
head(wild_bird_data, n=5)
```

```
## # A tibble: 5 x 2
##   `Wet body weight [g]` `Population size`
##                   <dbl>             <dbl>
## 1                  5.46           532194.
## 2                  7.76          3165107.
## 3                  8.64          2592997.
## 4                 10.7           3524193.
## 5                  7.42           389806.
```

```
#Plot weight vs population size to check relationship
plot(wild_bird_data$`Wet body weight [g]`, wild_bird_data$`Population size`, log='xy', col='blue', main=
```



The wild_bird_data dataset seems to maintain information about body wieght of birds (when wet) alongwith the population size in a certain region. The dataset consists of 146 rows and 2 columns, namely 'wet body weight' and the 'population'. The body weight measures is maintained in grams.
This data was probably gathered as an effort to get the weight estimates of bird populations as a part of a biomass study.

```r
#Read the wild_bird_data xlsx file
bird_data= read.csv('_data/birds.csv')

#Get dimensions
dim(bird_data)
```

```
## [1] 30977    14
```

```r
#Get column names
colnames(bird_data)
```

```
##  [1] "Domain.Code"      "Domain"           "Area.Code"        "Area"
##  [5] "Element.Code"     "Element"          "Item.Code"        "Item"
##  [9] "Year.Code"        "Year"             "Unit"             "Value"
## [13] "Flag"             "Flag.Description"
```

```r
#Displaying top 5 columns of dataframe
head(bird_data, n=5)
```

```
##   Domain.Code       Domain Area.Code        Area Element.Code Element Item.Code
## 1          QA Live Animals         2 Afghanistan         5112  Stocks      1057
## 2          QA Live Animals         2 Afghanistan         5112  Stocks      1057
## 3          QA Live Animals         2 Afghanistan         5112  Stocks      1057
## 4          QA Live Animals         2 Afghanistan         5112  Stocks      1057
## 5          QA Live Animals         2 Afghanistan         5112  Stocks      1057
##       Item Year.Code Year      Unit Value Flag Flag.Description
## 1 Chickens      1961 1961 1000 Head  4700    F     FAO estimate
## 2 Chickens      1962 1962 1000 Head  4900    F     FAO estimate
## 3 Chickens      1963 1963 1000 Head  5000    F     FAO estimate
## 4 Chickens      1964 1964 1000 Head  5300    F     FAO estimate
## 5 Chickens      1965 1965 1000 Head  5500    F     FAO estimate
```

```r
#Distinct regions within the dataset
unique(bird_data[c("Area")])
```

```
##                                                 Area
## 1                                        Afghanistan
## 59                                           Albania
## 291                                          Algeria
## 523                                   American Samoa
## 581                                           Angola
## 639                              Antigua and Barbuda
## 697                                        Argentina
## 929                                          Armenia
## 983                                            Aruba
## 1012                                       Australia
## 1186                                         Austria
## 1418                                      Azerbaijan
## 1472                                         Bahamas
## 1530                                         Bahrain
## 1588                                      Bangladesh
## 1704                                        Barbados
## 1820                                         Belarus
## 1901                                         Belgium
## 1977                            Belgium-Luxembourg
## 2133                                          Belize
```

```
## 2307                                        Benin
## 2365                                      Bermuda
## 2475                                       Bhutan
## 2533          Bolivia (Plurinational State of)
## 2707                    Bosnia and Herzegovina
## 2815                                     Botswana
## 2873                                       Brazil
## 3047                         Brunei Darussalam
## 3163                                     Bulgaria
## 3395                                 Burkina Faso
## 3453                                      Burundi
## 3539                                   Cabo Verde
## 3597                                     Cambodia
## 3713                                     Cameroon
## 3771                                       Canada
## 4003                                Cayman Islands
## 4056                      Central African Republic
## 4172                                         Chad
## 4230                                        Chile
## 4346                     China, Hong Kong SAR
## 4625                         China, Macao SAR
## 4683                          China, mainland
## 4857                     China, Taiwan Province of
## 5089                                     Colombia
## 5147                                      Comoros
## 5205                                        Congo
## 5263                                 Cook Islands
## 5372                                   Costa Rica
## 5430                               Côte d'Ivoire
## 5516                                     Croatia
## 5624                                        Cuba
## 5682                                      Cyprus
## 5959                                     Czechia
## 6063                               Czechoslovakia
## 6191     Democratic People's Republic of Korea
## 6307          Democratic Republic of the Congo
## 6365                                     Denmark
## 6597                                    Dominica
## 6655                          Dominican Republic
## 6713                                     Ecuador
## 6945                                       Egypt
## 7235                                 El Salvador
## 7293                           Equatorial Guinea
## 7409                                     Eritrea
## 7435                                     Estonia
## 7543                                    Eswatini
## 7601                                    Ethiopia
## 7627                                 Ethiopia PDR
## 7659          Falkland Islands (Malvinas)
## 7717                                        Fiji
## 7891                                     Finland
## 8057                                      France
## 8347                                French Guyana
## 8463                             French Polynesia
```

```
## 8579                                    Gabon
## 8637                                   Gambia
## 8695                                  Georgia
## 8749                                  Germany
## 8981                                    Ghana
## 9039                                   Greece
## 9329                                  Grenada
## 9387                               Guadeloupe
## 9541                                     Guam
## 9599                                Guatemala
## 9657                                   Guinea
## 9715                            Guinea-Bissau
## 9773                                   Guyana
## 9831                                    Haiti
## 10063                                Honduras
## 10121                                Hungary
## 10353                                Iceland
## 10411                                  India
## 10527                              Indonesia
## 10643          Iran (Islamic Republic of)
## 10875                                   Iraq
## 10933                                Ireland
## 11165                                 Israel
## 11377                                  Italy
## 11493                                Jamaica
## 11551                                  Japan
## 11667                                 Jordan
## 11915                             Kazakhstan
## 11969                                  Kenya
## 12027                               Kiribati
## 12085                                 Kuwait
## 12143                             Kyrgyzstan
## 12251       Lao People's Democratic Republic
## 12425                                 Latvia
## 12479                                Lebanon
## 12571                                Lesotho
## 12629                                Liberia
## 12745                                  Libya
## 12803                          Liechtenstein
## 12861                              Lithuania
## 12969                             Luxembourg
## 12988                             Madagascar
## 13220                                 Malawi
## 13278                               Malaysia
## 13394                                   Mali
## 13452                                  Malta
## 13616                             Martinique
## 13776                             Mauritania
## 13834                              Mauritius
## 14066                                 Mexico
## 14240       Micronesia (Federated States of)
## 14296                               Mongolia
## 14354                             Montenegro
## 14367                             Montserrat
```

```
## 14425                                      Morocco
## 14541                                   Mozambique
## 14773                                      Myanmar
## 15063                                      Namibia
## 15179                                        Nauru
## 15237                                        Nepal
## 15353                                  Netherlands
## 15519                  Netherlands Antilles (former)
## 15577                               New Caledonia
## 15635                                  New Zealand
## 15867                                    Nicaragua
## 15925                                        Niger
## 15983                                      Nigeria
## 16041                                         Niue
## 16099                               North Macedonia
## 16126                                       Norway
## 16300                                         Oman
## 16386                 Pacific Islands Trust Territory
## 16446                                     Pakistan
## 16562                                    Palestine
## 16620                                       Panama
## 16794                             Papua New Guinea
## 16968                                     Paraguay
## 17200                                         Peru
## 17258                                  Philippines
## 17490                                       Poland
## 17722                                     Portugal
## 17853                                  Puerto Rico
## 17911                                        Qatar
## 17969                           Republic of Korea
## 18197                          Republic of Moldova
## 18251                                     Réunion
## 18408                                     Romania
## 18640                           Russian Federation
## 18748                                       Rwanda
## 18876    Saint Helena, Ascension and Tristan da Cunha
## 18934                          Saint Kitts and Nevis
## 18992                                   Saint Lucia
## 19050                       Saint Pierre and Miquelon
## 19142             Saint Vincent and the Grenadines
## 19200                                       Samoa
## 19258                          Sao Tome and Principe
## 19432                                  Saudi Arabia
## 19521                                       Senegal
## 19579                                        Serbia
## 19631                          Serbia and Montenegro
## 19687                                    Seychelles
## 19803                                  Sierra Leone
## 19919                                     Singapore
## 20035                                      Slovakia
## 20139                                      Slovenia
## 20247                                Solomon Islands
## 20305                                       Somalia
## 20363                                  South Africa
```

```
## 20595                                        South Sudan
## 20602                                              Spain
## 20778                                          Sri Lanka
## 20894                                              Sudan
## 20901                                      Sudan (former)
## 20952                                           Suriname
## 21068                                             Sweden
## 21184                                        Switzerland
## 21416                                Syrian Arab Republic
## 21706                                         Tajikistan
## 21733                                           Thailand
## 21907                                        Timor-Leste
## 21965                                               Togo
## 22023                                            Tokelau
## 22081                                              Tonga
## 22139                                Trinidad and Tobago
## 22197                                            Tunisia
## 22290                                             Turkey
## 22522                                       Turkmenistan
## 22576                                             Tuvalu
## 22634                                             Uganda
## 22692                                            Ukraine
## 22800                              United Arab Emirates
## 22858 United Kingdom of Great Britain and Northern Ireland
## 23090                         United Republic of Tanzania
## 23206                            United States of America
## 23380                       United States Virgin Islands
## 23438                                            Uruguay
## 23670                                               USSR
## 23732                                         Uzbekistan
## 23786                                            Vanuatu
## 23844                  Venezuela (Bolivarian Republic of)
## 23902                                           Viet Nam
## 24018                         Wallis and Futuna Islands
## 24076                                              Yemen
## 24134                                       Yugoslav SFR
## 24258                                             Zambia
## 24316                                           Zimbabwe
## 24490                                              World
## 24780                                             Africa
## 25070                                     Eastern Africa
## 25302                                     Middle Africa
## 25476                                    Northern Africa
## 25766                                    Southern Africa
## 26056                                    Western Africa
## 26172                                           Americas
## 26404                                   Northern America
## 26636                                    Central America
## 26810                                          Caribbean
## 27042                                    South America
## 27274                                             Asia
## 27564                                     Central Asia
## 27672                                     Eastern Asia
## 27962                                     Southern Asia
```

```
## 28194                                South-eastern Asia
## 28484                                      Western Asia
## 28774                                            Europe
## 29064                                    Eastern Europe
## 29296                                   Northern Europe
## 29528                                   Southern Europe
## 29818                                    Western Europe
## 30108                                           Oceania
## 30340                       Australia and New Zealand
## 30572                                         Melanesia
## 30746                                        Micronesia
## 30862                                         Polynesia
```

```
#Distinct livestock information within the dataset
unique(bird_data[c("Item")])
```

```
##                        Item
## 1                  Chickens
## 117                   Ducks
## 175  Geese and guinea fowls
## 233                Turkeys
## 4520   Pigeons, other birds
```

The birds dataset looks to maintain information about the livestock information (specifically different birds) between the years 1961 to 2018 (minimum and maximum year present in the dataset). The dataset contains 30977 rows and 14 columns.
The dataset records information about 248 regions across the world and different poultry such as chickens, ducks, geese, etc. It uniquely records information about 5 distinct poultry animals.
This data was probably gathered as a measure to get information about the livestock maintained by different regions of the world for the different poultry.