

FunctionHw(3/21)

Anita Dhillon

2023-03-23

Function for the Portal Data Set

1. If surveys.csv, species.csv, and plots.csv are not available in your workspace download them, and load them into R using read.csv().

```
surveys <- read.csv(file = "../data-raw/surveys.csv")
species <- read.csv(file = "../data-raw/species.csv")
plots <- read.csv(file = "../data-raw/plots.csv")
intersect(colnames(surveys), colnames(species))
```

```
## [1] "species_id"
```

2. Combine the surveys and species tables into a single data frame.

There are several ways to create data frames within R. Depending on the data we want we as a data frame, there are different functions that we can use to create a data frame. - 'read.csv()': If we already have a file (either CSV or TSV) with data that we want to read into R as a data frame - "data.frame()": if we want to create data frame from vectors - '_join()': if we already have some data frames in R that we want to combine - There are many functions that create a data frame as an output: - 'summarize()': it creates a new data frame summarizing information from another data frame - but there are many other packages that can generate data frames with different types of data

```
combined_table <- inner_join(surveys, species, by = "species_id")
head(combined_table)
```

```
##   record_id month day year plot_id species_id sex hindfoot_length weight
## 1         1     7  16 1977      2        NL   M             32      NA
## 2         2     7  16 1977      3        NL   M             33      NA
## 3         3     7  16 1977      2        DM   F             37      NA
## 4         4     7  16 1977      7        DM   M             36      NA
## 5         5     7  16 1977      3        DM   M             35      NA
## 6         6     7  16 1977      1        PF   M             14      NA
##           genus species  taxa
## 1    Neotoma albigula Rodent
## 2    Neotoma albigula Rodent
## 3  Dipodomys merriami Rodent
## 4  Dipodomys merriami Rodent
## 5  Dipodomys merriami Rodent
## 6 Perognathus  flavus Rodent
```

3. Write a function that: Takes three arguments - a data frame (the combined table created before), a genus name, and a species name, Uses functions from the dplyr package to produce a data frame with two columns: year and count, where count is the number of individuals (i.e., the number of rows) for the species indicated by genus and species in that year, Returns the resulting data frame. Note that this data frame is a time series by year.

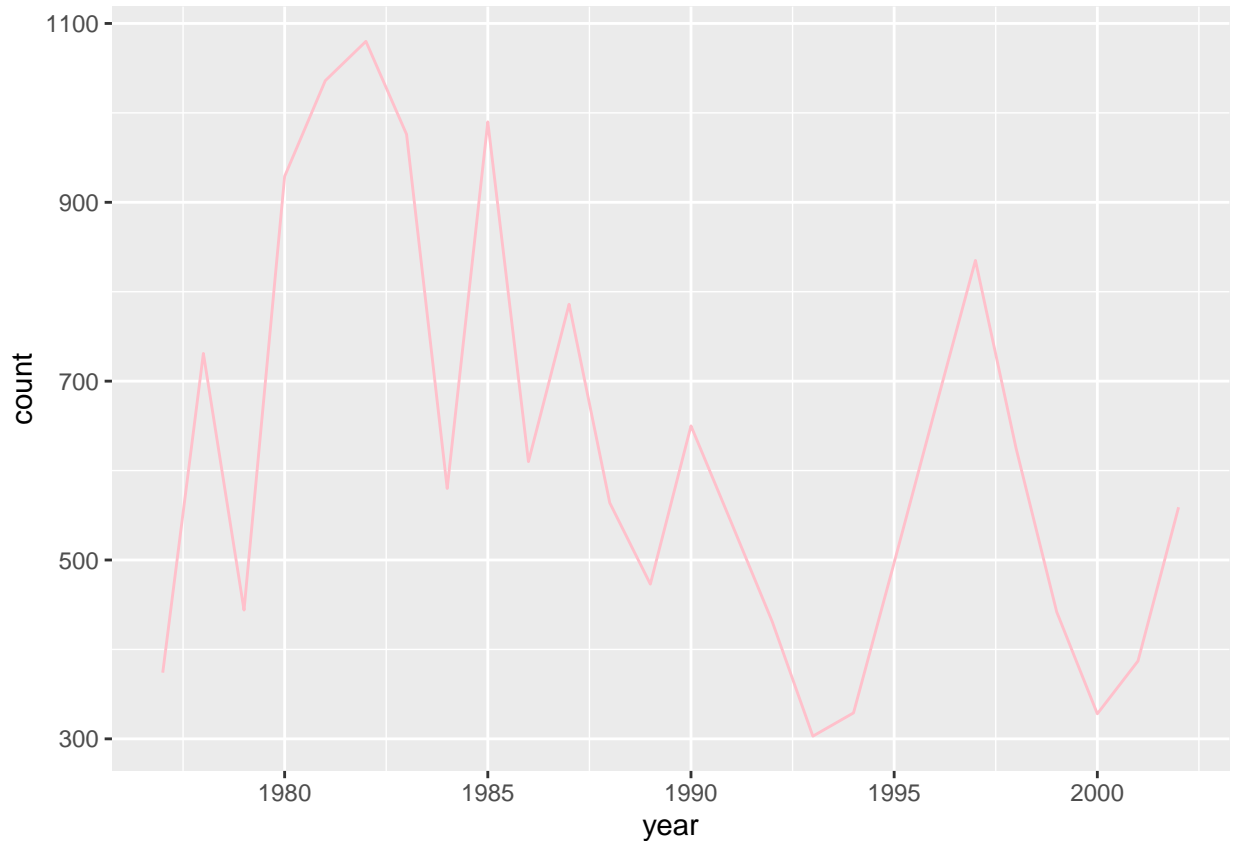
- As opposed to data frames: there is one way to create a variable : the function 'function()':

```
my_function1 <- function(a, genus_name = "Dipodomys", species_name = "merriami") {  
  combined_table %>%  
    group_by(year) %>%  
    filter(genus == genus_name | species == species_name) %>%  
    summarize(count = n()) -> output  
  return(output)  
}  
my_function1(a = a)
```

```
## # A tibble: 26 x 2  
##   year count  
##   <int> <int>  
## 1  1977   374  
## 2  1978   731  
## 3  1979   444  
## 4  1980   929  
## 5  1981  1036  
## 6  1982  1080  
## 7  1983   976  
## 8  1984   580  
## 9  1985   990  
## 10 1986   610  
## # ... with 16 more rows
```

4. Use your function to get the data frame of time series by year for the genus Dipodomys and the species merriami and then make a graph of this time series using ggplot that has year on the x axis, count on the y axis, and displays the data as points.

```
my_function1(a =a) %>%  
  ggplot(mapping = aes(x = year, y = count)) +  
  geom_line(color = "pink")
```



5. Use your function to get the data frame time series by year for the genus *Chaetodipus* and species *penicillatus* and then make a graph of this time-series using ggplot that has year on the x axis, count on the y axis, and displays the data as blue points (with size = 1) connected by blue lines (with size = 2). Change the x axis label to Year and the y axis label to Number of Individuals.

```
my_function1(a = a, "Chaetodipus", "penicillatus") %>%
  ggplot(mapping = aes(x = year, y = count)) +
  geom_point(size = 1, color = "yellow" ) +
  geom_line(size = 2, color = "red") +
  labs(x = "year", y = "Number of Individuals")
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
```

