Lab_3B.Rmd

```
surveys <- read_csv("data/portal_data_joined.csv")
```

```
## Parsed with column specification:
## cols(
##   record_id = col_double(),
##   month = col_double(),
##   day = col_double(),
##   year = col_double(),
##   plot_id = col_double(),
##   species_id = col_character(),
##   sex = col_character(),
##   hindfoot_length = col_double(),
##   weight = col_double(),
##   genus = col_character(),
##   species = col_character(),
##   taxa = col_character(),
##   plot_type = col_character()
## )
```

Challenge-1

Using pipes, subset the surveys data to include animals collected before 1995 and retain only the columns
year , sex , and weight .

```
surveys %>%
  filter(year < 1995) %>%
  select(year , sex , weight)
```

```
## # A tibble: 21,486 x 3
##     year sex    weight
##    <dbl> <chr>   <dbl>
##  1  1977 M          NA
##  2  1977 M          NA
##  3  1977 <NA>       NA
##  4  1977 <NA>       NA
##  5  1977 <NA>       NA
##  6  1977 <NA>       NA
##  7  1977 <NA>       NA
##  8  1978 <NA>       NA
##  9  1978 M         218
## 10  1978 <NA>       NA
## # ... with 21,476 more rows
```

Challenge-2

Create a new data frame from the surveys data that meets the following criteria: contains only the species_id column and a new column called hindfoot_cm containing the hindfoot_length values converted to centimeters. In this hindfoot_cm column, there are no NA s and all values are less than 3. Hint: think about how the commands should be ordered to produce this data frame!

```
new_surveys <- surveys %>%
                mutate("hindfoot_cm"=hindfoot_length/10) %>%
                filter(!is.na(hindfoot_cm)) %>%
                filter(hindfoot_cm < 3)%>%
                select(species_id, hindfoot_cm)
new_surveys
```

```
## # A tibble: 15,371 x 2
##    species_id hindfoot_cm
##    <chr>            <dbl>
##  1 NL                 2.8
##  2 NL                 2.1
##  3 NL                 2.1
##  4 NL                 2.9
##  5 NL                 2.9
##  6 DM                 2.5
##  7 PF                 1.5
##  8 PF                 1.6
##  9 PF                 1.5
## 10 PF                 1.4
## # ... with 15,361 more rows
```

Challenge-3

1. How many animals were caught in each plot_type surveyed?

```
surveys %>%
  count(plot_type)
```

```
## # A tibble: 5 x 2
##   plot_type                    n
##   <chr>                    <int>
## 1 Control                  15611
## 2 Long-term Krat Exclosure  5118
## 3 Rodent Exclosure          4233
## 4 Short-term Krat Exclosure 5906
## 5 Spectab exclosure         3918
```

2. Use group_by() and summarize() to find the mean, min, and max hindfoot length for each species (using species_id). Also add the number of observations (hint: see ?n).

```r
surveys %>%
  #filter(!is.na(hindfoot_length)) %>%
  group_by(species_id) %>%
  summarise(mean_hindfoot=mean(hindfoot_length, na.rm = TRUE),
            min_hindfoot= min(hindfoot_length, na.rm = TRUE),
            max_hindfoot= max(hindfoot_length, na.rm = TRUE),
            number_of_observations= n())
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 48 x 5
##    species_id mean_hindfoot min_hindfoot max_hindfoot number_of_observations
##    <chr>              <dbl>        <dbl>        <dbl>                  <int>
##  1 AB                   NaN          Inf         -Inf                    303
##  2 AH                    33           31           35                    437
##  3 AS                   NaN          Inf         -Inf                      2
##  4 BA                    13            6           16                     46
##  5 CB                   NaN          Inf         -Inf                     50
##  6 CM                   NaN          Inf         -Inf                     13
##  7 CQ                   NaN          Inf         -Inf                     16
##  8 CS                   NaN          Inf         -Inf                      1
##  9 CT                   NaN          Inf         -Inf                      1
## 10 CU                   NaN          Inf         -Inf                      1
## # ... with 38 more rows
```

3. What was the heaviest animal measured in each year? Return the columns year , genus , species_id , and weight .

```r
heaviest_animals_df<-surveys %>%
  group_by(year) %>%
  filter(weight==max(weight, na.rm = TRUE)) %>%
  select(year , genus , species_id , weight)

head(heaviest_animals_df,5)
```

```
## # A tibble: 5 x 4
## # Groups:   year [5]
##    year genus     species_id weight
##   <dbl> <chr>     <chr>       <dbl>
## 1  1979 Neotoma   NL            274
## 2  1986 Neotoma   NL            240
## 3  1987 Neotoma   NL            278
## 4  1988 Neotoma   NL            248
## 5  1992 Neotoma   NL            220
```

spread() function examples

Let's use spread() to transform surveys to find the mean weight of each genus in each plot over the entire survey period. We use filter() , group_by() and summarise() to filter our observations and variables of interest, and create a new variable for the mean_weight .

```
surveys_gw <- surveys %>%
  filter(!is.na(weight)) %>%
  group_by(genus,plot_id) %>%
  summarise(mean_weight=mean(weight))
```

`summarise()` regrouping output by 'genus' (override with '.groups' argument)

```
str(surveys_gw)
```

```
## tibble [196 x 3] (S3: grouped_df/tbl_df/tbl/data.frame)
##  $ genus      : chr [1:196] "Baiomys" "Baiomys" "Baiomys" "Baiomys" ...
##  $ plot_id    : num [1:196] 1 2 3 5 18 19 20 21 1 2 ...
##  $ mean_weight: num [1:196] 7 6 8.61 7.75 9.5 ...
##  - attr(*, "groups")= tibble [10 x 2] (S3: tbl_df/tbl/data.frame)
##   ..$ genus: chr [1:10] "Baiomys" "Chaetodipus" "Dipodomys" "Neotoma" ...
##   ..$ .rows: list<int> [1:10]
##   .. ..$ : int [1:8] 1 2 3 4 5 6 7 8
##   .. ..$ : int [1:24] 9 10 11 12 13 14 15 16 17 18 ...
##   .. ..$ : int [1:24] 33 34 35 36 37 38 39 40 41 42 ...
##   .. ..$ : int [1:24] 57 58 59 60 61 62 63 64 65 66 ...
##   .. ..$ : int [1:24] 81 82 83 84 85 86 87 88 89 90 ...
##   .. ..$ : int [1:23] 105 106 107 108 109 110 111 112 113 114 ...
##   .. ..$ : int [1:24] 128 129 130 131 132 133 134 135 136 137 ...
##   .. ..$ : int [1:24] 152 153 154 155 156 157 158 159 160 161 ...
##   .. ..$ : int [1:19] 176 177 178 179 180 181 182 183 184 185 ...
##   .. ..$ : int [1:2] 195 196
##   .. ..@ ptype: int(0)
##   ..- attr(*, ".drop")= logi TRUE
```

```
surveys_spread <- surveys_gw %>%
  spread(key=genus,value = mean_weight)
```

```
str(surveys_spread)
```

```
## tibble [24 x 11] (S3: tbl_df/tbl/data.frame)
##  $ plot_id        : num [1:24] 1 2 3 4 5 6 7 8 9 10 ...
##  $ Baiomys        : num [1:24] 7 6 8.61 NA 7.75 ...
##  $ Chaetodipus    : num [1:24] 22.2 25.1 24.6 23 18 ...
##  $ Dipodomys      : num [1:24] 60.2 55.7 52 57.5 51.1 ...
##  $ Neotoma        : num [1:24] 156 169 158 164 190 ...
##  $ Onychomys      : num [1:24] 27.7 26.9 26 28.1 27 ...
##  $ Perognathus    : num [1:24] 9.62 6.95 7.51 7.82 8.66 ...
##  $ Peromyscus     : num [1:24] 22.2 22.3 21.4 22.6 21.2 ...
##  $ Reithrodontomys: num [1:24] 11.4 10.7 10.5 10.3 11.2 ...
##  $ Sigmodon       : num [1:24] NA 70.9 65.6 82 82.7 ...
##  $ Spermophilus   : num [1:24] NA NA NA NA NA NA NA NA NA NA ...
```

```
surveys_gw %>%
  spread(key = genus, value = mean_weight, fill = 0) %>%
  head()
```

```
## # A tibble: 6 x 11
##   plot_id Baiomys Chaetodipus Dipodomys Neotoma Onychomys Perognathus Peromyscus
##     <dbl>   <dbl>       <dbl>     <dbl>   <dbl>     <dbl>       <dbl>      <dbl>
## 1       1    7          22.2      60.2    156.      27.7        9.62       22.2
## 2       2    6          25.1      55.7    169.      26.9        6.95       22.3
## 3       3    8.61       24.6      52.0    158.      26.0        7.51       21.4
## 4       4    0          23.0      57.5    164.      28.1        7.82       22.6
## 5       5    7.75       18.0      51.1    190.      27.0        8.66       21.2
## 6       6    0          24.9      58.6    180.      25.9        7.81       21.8
## # ... with 3 more variables: Reithrodontomys <dbl>, Sigmodon <dbl>,
## #   Spermophilus <dbl>
```