Data Wrangling: The Portal Data Set

Luna L Sanchez Reyes

2023-03-02

1. Intro to the Portal data set

Homework: create an intro describing the location of the experiment and the different experimental treatments. Paper here https://esajournals.onlinelibrary.wiley.com/doi/full/10.1890/15-2115.1

2. Load the data

6

6 Short-term Krat Exclosure

There are three different data sets, so to load them we need to create three data frames

```
surveys <- read.csv(file = "../data-raw/surveys.csv")</pre>
head(surveys)
##
     record_id month day year plot_id species_id sex hindfoot_length weight
                     7
## 1
              1
                        16 1977
                                       2
                                                  NL
                                                        М
                                                                         32
                                                                                NA
## 2
              2
                     7
                                       3
                                                        М
                                                                         33
                        16 1977
                                                   NL
                                                                                NA
                                                        F
                                       2
## 3
              3
                     7
                        16 1977
                                                  DM
                                                                         37
                                                                                NA
                     7
                                        7
                                                                         36
                        16 1977
                                                   DM
                                                        М
                                                                                NA
## 5
              5
                     7
                        16 1977
                                        3
                                                  DM
                                                        М
                                                                         35
                                                                                NA
                     7
                        16 1977
                                                        М
                                                                         14
                                                                                NA
species <- read.csv(file = "../data-raw/species.csv")</pre>
head(species)
##
     species_id
                                             species
                             genus
                                                        taxa
## 1
                        Amphispiza
                                           bilineata
                                                        Bird
              AB
## 2
              AH Ammospermophilus
                                             harrisi Rodent
## 3
              AS
                        Ammodramus
                                          savannarum
                                                        Bird
## 4
              BA
                           Baiomys
                                             taylori Rodent
## 5
              CB
                   Campylorhynchus brunneicapillus
                                                        Bird
              CM
                       Calamospiza
                                         melanocorys
                                                        Bird
plots <- read.csv(file = "../data-raw/plots.csv")</pre>
head(plots)
##
     plot_id
                                plot_type
## 1
            1
                       Spectab exclosure
## 2
                                  Control
            3
## 3
               Long-term Krat Exclosure
            4
## 4
                                  Control
## 5
            5
                        Rodent Exclosure
```

3. The dplyr package for data wrangling

Subset columns from a data frame with the function select()

head(surveys) record_id month day year plot_id species_id sex hindfoot_length weight ## ## 1 16 1977 2 32 1 7 NA## 2 2 3 33 7 16 1977 NLΜ NA16 1977 37 ## 3 7 2 DMF 3 NA## 4 4 7 16 1977 7 DMМ 36 NA ## 5 5 7 16 1977 3 DM М 35 NA7 ## 6 6 16 1977 1 PF М 14 NA surveys_subset <- select(surveys, month, day, year)</pre>

Create new variables from existing variables or transform existing variables with mutate()

The hindfoot_length variable is measured in mm. I want a new variable that stores hindfoot length in cm.

```
mutate(head(surveys), hindfoot_length_cm = hindfoot_length/10)
```

```
##
     record_id month day year plot_id species_id sex hindfoot_length weight
## 1
                    7
                       16 1977
                                       2
             1
                                                 NL
                                                       М
## 2
             2
                    7
                                                                       33
                       16 1977
                                       3
                                                 NL
                                                       М
                                                                               NA
## 3
             3
                    7
                       16 1977
                                       2
                                                 DM
                                                       F
                                                                       37
                                                                               NA
                    7
                                      7
             4
                                                                       36
## 4
                       16 1977
                                                 DM
                                                       М
                                                                               NA
             5
## 5
                    7
                       16 1977
                                      3
                                                 DM
                                                       М
                                                                       35
                                                                               NA
## 6
             6
                    7 16 1977
                                       1
                                                 PF
                                                       М
                                                                       14
                                                                               NA
##
     hindfoot_length_cm
## 1
## 2
                     3.3
## 3
                     3.7
## 4
                     3.6
## 5
                     3.5
## 6
                     1.4
```

head(surveys)

##		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	М	36	NA
##	5		5	7	16	1977	3	DM	M	35	NA
##	6		6	7	16	1977	1	PF	М	14	NA

surveys_mutated <- mutate(surveys, hindfoot_length_cm = hindfoot_length/10)
head(surveys_mutated)</pre>

```
record_id month day year plot_id species_id sex hindfoot_length weight
##
## 1
             1
                   7
                       16 1977
                                      2
                                                NL
                                                      М
                                                                             NA
## 2
             2
                    7
                       16 1977
                                      3
                                                NL
                                                     М
                                                                      33
                                                                             NA
## 3
             3
                    7
                       16 1977
                                      2
                                                     F
                                                                      37
                                                DM
                                                                             NA
                   7
                                      7
## 4
             4
                       16 1977
                                                DM
                                                      М
                                                                      36
                                                                             NA
## 5
             5
                   7 16 1977
                                      3
                                                DM
                                                     М
                                                                      35
                                                                             NA
```

```
## 6
             6
                   7 16 1977
                                  1
                                              PF
                                                                  14
                                                                         NA
   hindfoot_length_cm
##
## 1
                    3.2
## 2
                    3.3
## 3
                    3.7
## 4
                    3.6
## 5
                    3.5
## 6
                    1.4
```

Sorting or ordering data with the function arrange()

If we want to order the data frame values based on the weight variable:

```
surveys_arranged <- arrange(surveys, weight)
head(surveys_arranged)</pre>
```

##		record_id	${\tt month}$	day	year	plot_id	species_id	sex	hindfoot_length	weight
##	1	218	9	13	1977	1	PF	М	13	4
##	2	4052	4	5	1981	3	PF	F	15	4
##	3	4290	4	6	1981	4	PF		NA	4
##	4	5346	2	22	1982	21	PF	F	14	4
##	5	7084	11	22	1982	3	PF	F	16	4
##	6	8736	12	8	1983	19	RM	M	17	4

Order values in descendant order with the function desc()

```
surveys_arranged <- arrange(surveys, desc(weight))
head(surveys_arranged)</pre>
```

	record_id	${\tt month}$	day	year	plot_id	species_id	sex	hindfoot_length	weight
1	33049	11	17	2001	12	NL	М	33	280
2	12871	5	28	1987	2	NL	M	32	278
3	15459	1	11	1989	9	NL	M	36	275
4	2133	10	25	1979	2	NL	F	33	274
5	12729	4	26	1987	2	NL	M	32	270
6	13114	7	26	1987	2	NL	M	NA	269
	1 2 3 4 5 6	1 33049 2 12871 3 15459 4 2133 5 12729	1 33049 11 2 12871 5 3 15459 1 4 2133 10 5 12729 4	1 33049 11 17 2 12871 5 28 3 15459 1 11 4 2133 10 25 5 12729 4 26	1 33049 11 17 2001 2 12871 5 28 1987 3 15459 1 11 1989 4 2133 10 25 1979 5 12729 4 26 1987	1 33049 11 17 2001 12 2 12871 5 28 1987 2 3 15459 1 11 1989 9 4 2133 10 25 1979 2 5 12729 4 26 1987 2	1 33049 11 17 2001 12 NL 2 12871 5 28 1987 2 NL 3 15459 1 11 1989 9 NL 4 2133 10 25 1979 2 NL 5 12729 4 26 1987 2 NL	1 33049 11 17 2001 12 NL M 2 12871 5 28 1987 2 NL M 3 15459 1 11 1989 9 NL M 4 2133 10 25 1979 2 NL F 5 12729 4 26 1987 2 NL M	2 12871 5 28 1987 2 NL M 32 3 15459 1 11 1989 9 NL M 36 4 2133 10 25 1979 2 NL F 33 5 12729 4 26 1987 2 NL M 32

Filter values with the function filter()

Filter the data frame to keep rows with weight values that are equal to 4:

```
filter(surveys, weight == 4)
```

##		record_id	month	day	year	plot_id	species_id	sex	hindfoot_	length	weight
##	1	218	9	13	1977	1	PF	М		13	4
##	2	4052	4	5	1981	3	PF	F		15	4
##	3	4290	4	6	1981	4	PF			NA	4
##	4	5346	2	22	1982	21	PF	F		14	4
##	5	7084	11	22	1982	3	PF	F		16	4
##	6	8736	12	8	1983	19	RM	М		17	4
##	7	9790	1	19	1985	16	RM	F		16	4
##	8	9794	1	19	1985	24	RM	М		16	4
##	9	9799	1	19	1985	19	RM	М		16	4
##	10	9823	1	19	1985	23	RM	М		16	4

```
9853
                        19 1985
                                       17
## 11
                                                   RM
                                                                          16
## 12
            9909
                      1
                         20 1985
                                       15
                                                   RM
                                                         F
                                                                          15
                                                                                  4
## 13
                         16 1985
                                                                                  4
            9937
                      2
                                       21
                                                   RM
                                                         М
                                                                          16
           10119
                         17 1985
                                       10
                                                                          16
                                                                                  4
## 14
                      3
                                                   RM
                                                         М
                                                                                  4
## 15
           10439
                      5
                         24 1985
                                        7
                                                   RM
                                                         М
                                                                          16
## 16
           28126
                      6
                         28 1998
                                       15
                                                   PF
                                                         М
                                                                         NA
                                                                                  4
## 17
           29906
                     10
                        10 1999
                                        4
                                                   PP
                                                                          21
                                                                                   4
                                                         М
surveys_filtered <- filter(surveys, weight != 4)</pre>
head(surveys filtered)
##
     record_id month day year plot_id species_id sex hindfoot_length weight
## 1
             63
                     8
                        19 1977
                                       3
                                                  DM
                                                        М
                                                                        35
                                                                                40
## 2
             64
                     8
                        19 1977
                                       7
                                                        М
                                                                        37
                                                                                48
                                                  DM
## 3
                                                        F
                                                                        34
                                                                                29
             65
                     8
                        19 1977
                                       4
                                                  DM
                                                        F
## 4
             66
                     8
                        19 1977
                                       4
                                                  DM
                                                                        35
                                                                                46
## 5
             67
                     8
                       19 1977
                                       7
                                                  DM
                                                        Μ
                                                                        35
                                                                                36
## 6
             68
                     8
                       19 1977
                                       8
                                                  DO
                                                        F
                                                                        32
                                                                                52
surveys_filtered <- filter(surveys, weight > 200)
head(surveys_filtered)
     record_id month day year plot_id species_id sex hindfoot_length weight
##
## 1
            588
                     2
                        18 1978
                                       2
                                                  NL
                                                                               218
                                                        М
                                                                        NA
## 2
            646
                     2
                        20 1978
                                      18
                                                  NL
                                                        М
                                                                        32
                                                                               228
## 3
            655
                     3
                        11 1978
                                       3
                                                  NL
                                                        М
                                                                        32
                                                                               232
## 4
            825
                     4
                        10 1978
                                      18
                                                  NL
                                                        М
                                                                        NA
                                                                               225
## 5
            845
                    5
                         6 1978
                                       2
                                                  NL
                                                        М
                                                                        32
                                                                               204
## 6
            848
                     5
                         6 1978
                                      22
                                                  NL
                                                        М
                                                                        32
                                                                               212
```

Filter with more complex conditions

I want values that have weight larger than 200 AND also are females

```
surveys_filtered <- filter(surveys, weight > 200, sex == "F")
head(surveys_filtered)
```

```
##
     record_id month day year plot_id species_id sex hindfoot_length weight
## 1
           875
                    5
                      17 1978
                                       5
                                                 NL
                                                       F
                                                                       33
                                                                              212
## 2
                       28 1979
                                      12
                                                       F
                                                                       32
                                                                              239
           1731
                    4
                                                  NL
## 3
          2081
                   10
                       24 1979
                                      12
                                                 NL
                                                       F
                                                                       32
                                                                              211
## 4
          2133
                   10
                       25 1979
                                       2
                                                 NL
                                                       F
                                                                       33
                                                                              274
                                      12
                                                       F
## 5
          2247
                   11
                       18 1979
                                                 NL
                                                                       33
                                                                              217
## 6
          2305
                       15 1980
                                      12
                                                  NL
                                                       F
                                                                       32
                                                                              214
```

surveys_filtered <- filter(surveys, weight > 200 & sex == "F")
head(surveys_filtered)

```
##
     record_id month day year plot_id species_id sex hindfoot_length weight
## 1
           875
                    5
                       17 1978
                                      5
                                                 NL
                                                       F
                                                                       33
                                                                              212
                                                       F
## 2
          1731
                    4
                       28 1979
                                     12
                                                 NL
                                                                       32
                                                                              239
## 3
                   10
                                                       F
          2081
                       24 1979
                                      12
                                                 NL
                                                                       32
                                                                              211
          2133
                       25 1979
                                      2
                                                       F
                                                                              274
## 4
                   10
                                                 NL
                                                                       33
## 5
           2247
                   11
                       18 1979
                                      12
                                                 NL
                                                       F
                                                                       33
                                                                              217
                                                       F
## 6
          2305
                       15 1980
                                     12
                                                 NL
                                                                       32
                    1
                                                                              214
```

Now, I want values that have weight larger than 200 OR are also females:

```
surveys_filtered <- filter(surveys, weight > 200 | sex == "F")
head(surveys_filtered)
```

Filtering NA values

NA is a special valie in R. We can't use logical statements with it, we have to use the <code>is.na()</code> function:

```
surveys_filtered <- filter(surveys, !is.na(weight))
head(surveys_filtered)</pre>
```

##		record_id	month	day	year	plot_id	species_id	sex	hindfoot_length	weight
##	1	63	8	19	1977	3	DM	M	35	40
##	2	64	8	19	1977	7	DM	M	37	48
##	3	65	8	19	1977	4	DM	F	34	29
##	4	66	8	19	1977	4	DM	F	35	46
##	5	67	8	19	1977	7	DM	M	35	36
##	6	68	8	19	1977	8	DO	F	32	52