

MessyEDA_Allendale_FA

Cass Lee

2025-09-14

! For best results, run these code chunks sequentially in RStudio rather than creating a knitted document.

Preliminary Set Up

```
# import useful libraries  
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --  
## v dplyr      1.1.4      v readr      2.1.5  
## v forcats    1.0.0      v stringr   1.5.1  
## v ggplot2    3.5.1      v tibble    3.2.1  
## v lubridate  1.9.3      v tidyr     1.3.1  
## v purrr      1.0.2  
## -- Conflicts ----- tidyverse_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag()     masks stats::lag()  
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors  
library(mosaic)
```

```
## Registered S3 method overwritten by 'mosaic':  
##   method                from  
##   fortify.SpatialPolygonsDataFrame ggplot2  
##  
## The 'mosaic' package masks several functions from core packages in order to add  
## additional features. The original behavior of these functions should not be affected by this.  
##  
## Attaching package: 'mosaic'  
##  
## The following object is masked from 'package:Matrix':  
##  
##   mean  
##  
## The following objects are masked from 'package:dplyr':  
##  
##   count, do, tally  
##  
## The following object is masked from 'package:purrr':  
##  
##   cross  
##  
## The following object is masked from 'package:ggplot2':  
##
```

```

##      stat
##
## The following objects are masked from 'package:stats':
##
##      binom.test, cor, cor.test, cov, fivenum, IQR, median, prop.test,
##      quantile, sd, t.test, var
##
## The following objects are masked from 'package:base':
##
##      max, mean, min, prod, range, sample, sum
library(ggformula)
library(dplyr)
library(lme4)

##
## Attaching package: 'lme4'
##
## The following object is masked from 'package:mosaic':
##
##      factorize
# set import file path
path_a <- '~/Desktop/ENVR356/LS_FA/allendale_data.csv'
# import full data set
all_data_a <- read_csv(file = path_a)

## Rows: 2602 Columns: 25
## -- Column specification -----
## Delimiter: ","
## chr  (16): line, code, trap type, tags, status, initial bait, rebaited, bait...
## dbl  (7): trap nid, latitude, longitude, easting, northing, nid, strikes
## lgl  (1): Trap sub type
## dtm  (1): date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
glimpse(all_data_a)

## Rows: 2,602
## Columns: 25
## $ line      <chr> "Lower Bamfords", "Lower Bamfords", "Lower Bamfords", ~
## $ `trap nid` <dbl> 7558661, 7558675, 7558661, 7558675, 7558675, 7558661, ~
## $ code      <chr> "DC20 #43", "DC20# 45", "DC20 #43", "DC20# 45", "DC20~
## $ `trap type` <chr> "DOC 200", "DOC 200", "DOC 200", "DOC 200", "DOC 200"~
## $ `Trap sub type` <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ tags      <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ latitude  <dbl> -43.64843, -43.64846, -43.64843, -43.64846, -43.64846~
## $ longitude <dbl> 172.6462, 172.6463, 172.6462, 172.6463, 172.6463, 172~
## $ easting   <dbl> 1571469, 1571479, 1571469, 1571479, 1571479, 1571469, ~
## $ northing  <dbl> 5167113, 5167110, 5167113, 5167110, 5167110, 5167113, ~
## $ nid       <dbl> 7571173, 7571176, 8588029, 8588030, 8857887, 8857909, ~
## $ date      <dtm> 2022-03-17 11:28:00, 2022-03-17 11:29:00, 2022-04-05~
## $ status    <chr> "Still set, bait bad", "Still set, bait OK", "Still s~
## $ `initial bait` <chr> NA, NA, "Whole egg", "None", "None", "Fresh Rabbit", ~

```

```
## $ rebaited      <chr> "Yes", "No", "Yes", "No", "No", "No", "No", "No", "No", "No~
## $ `bait type`   <chr> "Whole egg", "None", "Fresh Rabbit", "None", "None", ~
## $ `Bait details` <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ `recorded by` <chr> "Denis Aldridge", "Denis Aldridge", "Denis Aldridge",~
## $ strikes       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ `species caught` <chr> "None", "None", "None", "None", "None", "None", "None~
## $ sex           <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ maturity      <chr> "Adult", "Adult", "Adult", "Adult", "Adult", "Adult",~
## $ `trap condition` <chr> "OK", "OK", "OK", "OK", "OK", "OK", "OK", "OK", "OK",~
## $ notes         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ Images        <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
```

```
# select relevant cols
```

```
data1_a <- select(all_data_a, `trap nid`, `trap type`, nid, status, `initial bait`, rebaited, `bait type`)
```

```
# rename columns to work better for analysis
```

```
colnames(data1_a) <- c("TrapID", "TrapType", "EntryID", "Status", "InitialBait",
  "Rebaited", "BaitType", "BaitDetails",
  "Strikes", "SpeciesCaught", "Notes")
```

```
glimpse(data1_a)
```

```
## Rows: 2,602
```

```
## Columns: 11
```

```
## $ TrapID      <dbl> 7558661, 7558675, 7558661, 7558675, 7558675, 7558661, 75~
## $ TrapType    <chr> "DOC 200", "DOC 200", "DOC 200", "DOC 200", "DOC 200", "~
## $ EntryID     <dbl> 7571173, 7571176, 8588029, 8588030, 8857887, 8857909, 95~
## $ Status      <chr> "Still set, bait bad", "Still set, bait OK", "Still set,~
## $ InitialBait <chr> NA, NA, "Whole egg", "None", "None", "Fresh Rabbit", "No~
## $ Rebaited    <chr> "Yes", "No", "Yes", "No", "No", "No", "No", "No", "No", ~
## $ BaitType    <chr> "Whole egg", "None", "Fresh Rabbit", "None", "None", "No~
## $ BaitDetails <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ Strikes     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ SpeciesCaught <chr> "None", "None", "None", "None", "None", "None", "None", ~
## $ Notes       <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
```

```
# make sure data looks good
```

```
# TODO vaguely weird, maybe come back to this
```

```
inspect(data1_a)
```

```
##
```

```
## categorical variables:
```

```
##           name      class levels    n missing
## 1      TrapType character         8 2602      0
## 2        Status character         9 2602      0
## 3   InitialBait character        74 2371     231
## 4        Rebaited character         2 2602      0
## 5         BaitType character        79 2602      0
## 6   BaitDetails character         11  82    2520
## 7 SpeciesCaught character         15 2602      0
## 8          Notes character        106 119    2483
##                                     distribution
## 1 DOC 200 (37.5%), Trapinator (21.8%) ...
## 2 Sprung (38.7%) ...
## 3 None (47.8%), Peanut butter (13.9%) ...
## 4 No (56%), Yes (44%)
## 5 None (47%), Peanut butter (14%) ...
## 6 Pams smooth peanut butter (75.6%) ...
```

```

## 7 None (64%), Mouse (14.8%) ...
## 8 5 triggers (4.2%), 4 triggers (2.5%) ...
##
## quantitative variables:
##      name      class      min      Q1      median      Q3      max      mean
## 1  TrapID numeric 7558661 16750633 21849848 23959658 33842395 2.042157e+07
## 2  EntryID numeric 7571173 21192146 24095178 25949082 33907637 2.354799e+07
## 3  Strikes numeric      0      0      0      1      8 3.985396e-01
##      sd      n missing
## 1 4.608310e+06 2602      0
## 2 4.562953e+06 2602      0
## 3 6.218236e-01 2602      0

# select relevant cols for the new type of analysis i'm considering as of 16/10
data2_a <- select(all_data_a, `trap nid`, `trap type`, nid, date, status, `bait type`, `Bait details`,
# rename columns to work better for analysis
colnames(data2_a) <- c("TrapID", "TrapType", "EntryID", "Date", "Status", "BaitType", "BaitDetails",
                      "Strikes", "SpeciesCaught", "Notes")
glimpse(data2_a)

## Rows: 2,602
## Columns: 10
## $ TrapID      <dbl> 7558661, 7558675, 7558661, 7558675, 7558675, 7558661, 75~
## $ TrapType    <chr> "DOC 200", "DOC 200", "DOC 200", "DOC 200", "DOC 200", "~
## $ EntryID     <dbl> 7571173, 7571176, 8588029, 8588030, 8857887, 8857909, 95~
## $ Date        <dtm> 2022-03-17 11:28:00, 2022-03-17 11:29:00, 2022-04-05 15~
## $ Status      <chr> "Still set, bait bad", "Still set, bait OK", "Still set,~
## $ BaitType    <chr> "Whole egg", "None", "Fresh Rabbit", "None", "None", "No~
## $ BaitDetails <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ Strikes     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ SpeciesCaught <chr> "None", "None", "None", "None", "None", "None", "None", ~
## $ Notes       <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~

# make sure data looks good

# EDA for trap type, bait type
tally(~TrapType, data=data1_a)

## TrapType
##      AT220      Cage Trap      D-Rat      DOC 150      DOC 200
##      441      1      534      2      975
## Leg hold trap      Rat trap      Trapinator
##      3      80      566

tally(~BaitType, data=data1_a)

## BaitType
##
##      Blaze
##      1
##      Blaze, Mayo
##      11
##      Blaze, Mayo, Deep Fried Dog Roll, Peanut butter
##      5
##      Blaze, Mayo, Smooth
##      6
##      Blaze, NZAT Lure - Original, Smooth

```

##		2
##	Blaze, Possum Dough	
##		9
##	Blaze, Smooth	
##		4
##	Blaze, Smooth, Mayo	
##		6
##	Blaze, Smooth, NZAT Lure - Original	
##		1
##	Cheese	
##		4
##	Deep Fried Dog Roll	
##		20
##	Deep Fried Dog Roll, Good Nature Meat Lovers	
##		1
##	Deep Fried Dog Roll, Good Nature Meat Lovers, Peanut butter	
##		4
##	Deep Fried Dog Roll, Peanut butter	
##		53
##	Deep Fried Dog Roll, Peanut butter, Good Nature Meat Lovers	
##		3
##	Deep Fried Dog Roll, Peanut butter, Whole egg	
##		1
##	Deep Fried Dog Roll, Possum Dough, Peanut butter	
##		1
##	Dehydrated Rabbit, Peanut butter, Whole egg	
##		1
##	Fake Egg, Mayo	
##		5
##	Fresh meat	
##		25
##	Fresh meat, Fresh Rabbit, Peanut butter	
##		1
##	Fresh meat, Peanut butter	
##		8
##	Fresh meat, Whole egg	
##		1
##	Fresh Rabbit	
##		11
##	Fresh Rabbit, Peanut butter	
##		29
##	Fresh Rabbit, Whole egg	
##		1
##	Good Nature Chocolate	
##		11
##	Good Nature Meat Lovers	
##		5
##	Good Nature Meat Lovers, Mayo	
##		2
##	Good Nature Meat Lovers, NZAT Lure - Original	
##		1
##	Goodnature Cinnamon pre feed	
##		1
##	Mayo	

##		93
##	Mayo, Blaze	
##		6
##	Mayo, Blaze, Possum Dough	
##		2
##	Mayo, Blaze, Smooth	
##		6
##	Mayo, Deep Fried Dog Roll, Good Nature Meat Lovers	
##		2
##	Mayo, Good Nature Meat Lovers	
##		2
##	Mayo, Smooth	
##		32
##	Mayo, Smooth, Blaze	
##		13
##	None	
##		1223
##	NZAT Lure - Original	
##		12
##	NZAT Lure - Original, Blaze	
##		2
##	NZAT Lure - Original, Blaze, Smooth	
##		1
##	NZAT Lure - Original, Smooth	
##		6
##	NZAT Lure - Original, Smooth, Blaze	
##		2
##	Other (please specify)	
##		1
##	Peanut butter	
##		364
##	Peanut butter, Deep Fried Dog Roll, Good Nature Meat Lovers	
##		1
##	Peanut butter, Fresh meat	
##		3
##	Peanut butter, Fresh Rabbit	
##		50
##	Peanut butter, Fresh Rabbit, Possum Dough	
##		1
##	Peanut butter, Fresh Rabbit, Whole egg	
##		1
##	Peanut butter, Good Nature Meat Lovers, Deep Fried Dog Roll	
##		1
##	Peanut butter, Other (please specify)	
##		8
##	Peanut butter, Possum Dough	
##		1
##	Peanut butter, Rat and Possum Lure	
##		1
##	Peanut butter, Salted meat	
##		5
##	Peanut butter, Salted Rabbit	
##		5
##	Peanut butter, Smooth	

```
## 2
## Peanut butter, Smooth, Good Nature Meat Lovers
## 1
## Peanut butter, Whole egg
## 9
## Possum Dough
## 240
## Possum Dough, Blaze
## 1
## Possum Dough, Fresh Rabbit
## 1
## Possum Dough, Mayo
## 3
## Possum Dough, Smooth
## 5
## Possum Dough, Whole egg
## 2
## Rat and Possum Lure
## 1
## Salted meat
## 5
## Salted Rabbit
## 5
## Salted Rabbit, Peanut butter
## 3
## Smooth
## 16
## Smooth, Good Nature Meat Lovers
## 1
## Smooth, Mayo
## 13
## Tinned Sardines
## 10
## Whole egg
## 201
## Whole egg, Peanut butter
## 2
## Whole egg, Peanut butter, Fresh Rabbit
## 1
## Whole egg, Possum Dough
## 2
```

```
# 600 odd just mayo before filtering by just single kill
```

```
# single kill
```

```
# filter to just single kill traps that have mayo or pb as bait
```

```
analysis_data_a <- subset(data1_a, grepl("MAYO", toupper(BaitType)) | grepl("PEANUT BUTTER", toupper(BaitType)) |
  subset(!(grepl("AT220", TrapType) | grepl("Sentinel", TrapType) | grepl("Supervisor Max", TrapType))))
```

```
# how to distinguish single kill traps
```

```
# 27 unspecified
```

```
# single: cage trap, D-Rat, DOC 200, DOC 250, leg hold trap, rat trap, SA cat, Tāwhiti - Smart Cage,
```

```
# cage trap, leg hold trap, Tāwhiti - Smart Cage is live capture
```

```
# rat trap resetting type isn't specified but we can probably pretty safely assume it's single trap
```

```
# self-resetting: AT220, sentinel, supervisor max
```

```
# sentinel, supervisor max assumed to not be by looking up quickly but actually recorded as having
# sentinel trap is specifically for possums, supposed to stop rats from getting bait
```

```
tally(~BaitType, data=analysis_data_a)
```

```
## BaitType
##           Blaze, Mayo, Deep Fried Dog Roll, Peanut butter
##                                           5
## Deep Fried Dog Roll, Good Nature Meat Lovers, Peanut butter
##                                           4
##           Deep Fried Dog Roll, Peanut butter
##                                           53
## Deep Fried Dog Roll, Peanut butter, Good Nature Meat Lovers
##                                           3
##           Deep Fried Dog Roll, Peanut butter, Whole egg
##                                           1
## Deep Fried Dog Roll, Possum Dough, Peanut butter
##                                           1
##           Dehydrated Rabbit, Peanut butter, Whole egg
##                                           1
##           Fresh meat, Fresh Rabbit, Peanut butter
##                                           1
##           Fresh meat, Peanut butter
##                                           8
##           Fresh Rabbit, Peanut butter
##                                           29
##           Mayo
##                                           1
## Mayo, Deep Fried Dog Roll, Good Nature Meat Lovers
##                                           2
##           Peanut butter
##                                           363
## Peanut butter, Deep Fried Dog Roll, Good Nature Meat Lovers
##                                           1
##           Peanut butter, Fresh meat
##                                           3
##           Peanut butter, Fresh Rabbit
##                                           50
## Peanut butter, Fresh Rabbit, Possum Dough
##                                           1
##           Peanut butter, Fresh Rabbit, Whole egg
##                                           1
## Peanut butter, Good Nature Meat Lovers, Deep Fried Dog Roll
##                                           1
##           Peanut butter, Other (please specify)
##                                           8
##           Peanut butter, Possum Dough
##                                           1
##           Peanut butter, Rat and Possum Lure
##                                           1
##           Peanut butter, Salted meat
##                                           5
##           Peanut butter, Salted Rabbit
##                                           5
```



```
## Peanut butter, Smooth
## 2
## Peanut butter, Smooth, Good Nature Meat Lovers
## 1
## Peanut butter, Whole egg
## 9
## Possum Dough, Mayo
## 3
## Salted Rabbit, Peanut butter
## 3
## Whole egg, Peanut butter
## 2
## Whole egg, Peanut butter, Fresh Rabbit
## 1

# look at the self-resetting trap data to see if it can be made to compare to single kill
resetting_a <- subset(data2_a, grepl("MAYO", toupper(BaitType)) | grepl("PEANUT BUTTER", toupper(BaitType)) |
  subset(grepl("AT220", TrapType) | grepl("Sentinel", TrapType) | grepl("Supervisor Max", TrapType))
tally(~BaitType, data=resetting_a)

## BaitType
## Blaze, Mayo Blaze, Mayo, Smooth
## 11 6
## Blaze, Smooth, Mayo Fake Egg, Mayo
## 6 5
## Good Nature Meat Lovers, Mayo Mayo
## 2 92
## Mayo, Blaze Mayo, Blaze, Possum Dough
## 6 2
## Mayo, Blaze, Smooth Mayo, Good Nature Meat Lovers
## 6 2
## Mayo, Smooth Mayo, Smooth, Blaze
## 32 13
## Peanut butter Smooth, Mayo
## 1 13

# 680ish plain mayo, 630ish plain pb
resetting_a[resetting_a$BaitType=="Mayo",]

## # A tibble: 92 x 10
##   TrapID TrapType EntryID Date Status BaitType BaitDetails
##   <dbl> <chr> <dbl> <dtm> <chr> <chr> <chr>
## 1 21865469 AT220 25302129 2024-05-30 14:18:00 Sprung Mayo <NA>
## 2 20226532 AT220 25302496 2024-05-30 14:39:00 Sprung Mayo <NA>
## 3 20226533 AT220 25303440 2024-05-30 15:38:00 Sprung Mayo <NA>
## 4 20395111 AT220 25303772 2024-05-30 16:12:00 Sprung Mayo <NA>
## 5 20395689 AT220 25303895 2024-05-30 16:16:00 Sprung Mayo <NA>
## 6 20396176 AT220 25304030 2024-05-30 16:27:00 Sprung Mayo <NA>
## 7 24362744 AT220 26068237 2024-07-17 09:25:00 Still se~ Mayo <NA>
## 8 24362745 AT220 26069232 2024-07-17 10:04:00 Still se~ Mayo <NA>
## 9 24904091 AT220 26069762 2024-07-17 10:37:00 Sprung Mayo <NA>
## 10 25265493 AT220 26075724 2024-07-17 12:05:00 Sprung Mayo <NA>
## # i 82 more rows
## # i 3 more variables: Strikes <dbl>, SpeciesCaught <chr>, Notes <chr>
```

```
# servicing interval for self-resetting traps
# scatter plot with line drawn through
# x axis is time, strikes is y axis
# entryid is series
gf_line(Strokes~Date, color=~as.factor(TrapID),data=resetting_a)
```

