

Foundations of Data Science for Biologists

Wrapping up R

BIO 724D

22-Oct-2023

Instructors: Greg Wray, Paul Magwene, and Jesse Granger

The Bird Cognition Dataset

Danae Diaz

Eastern Bluebirds (*Sialia sialis*)

Question: Does metal contamination affect the speed at which the birds are able to solve a box-opening puzzle?



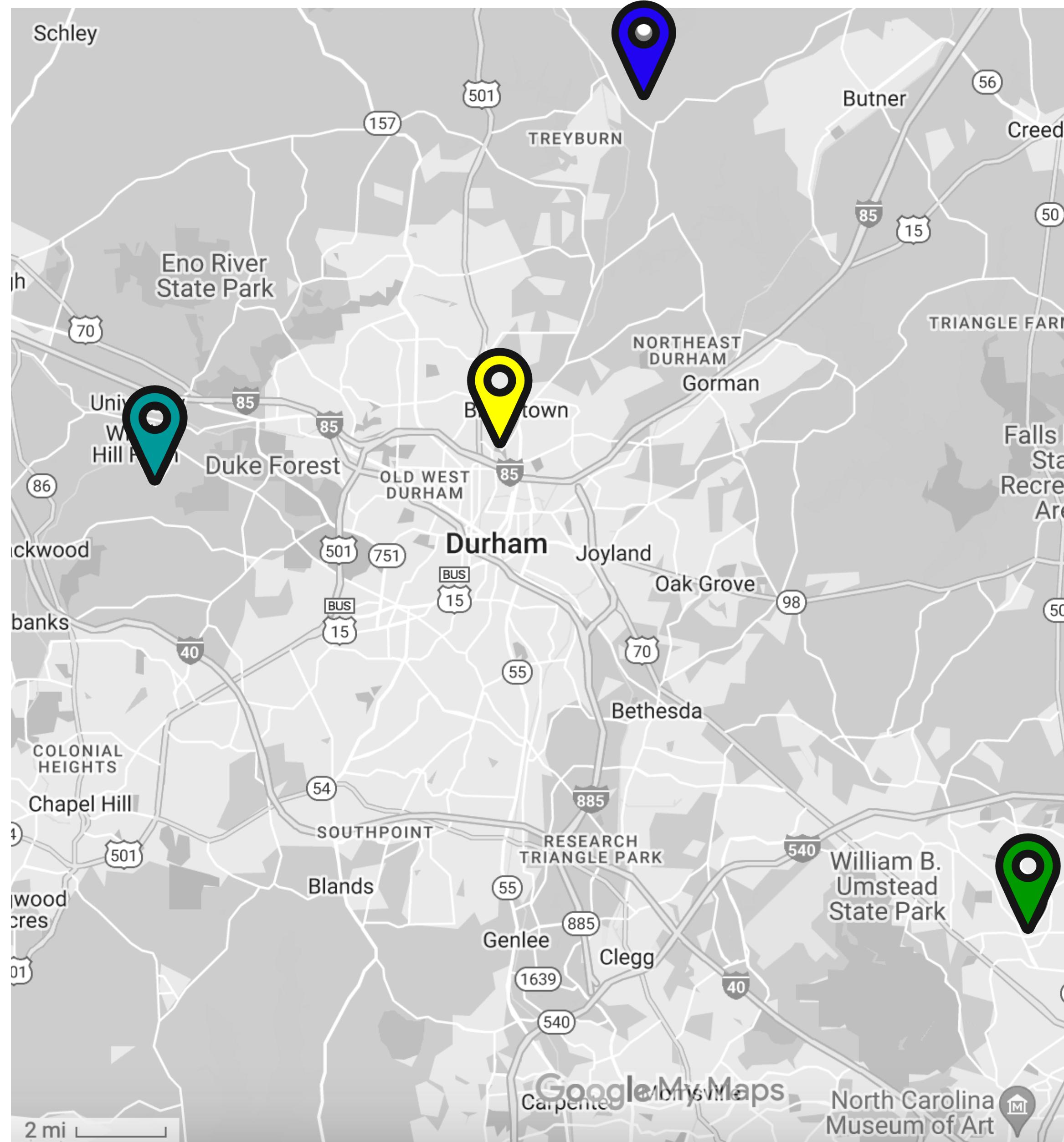
The Puzzle



The Puzzle



The Bird Cognition Dataset – Field Sites



Most Contaminated



Least Contaminated

In Class – Problem Set 07

Complete Problem set 07, posted on the wiki

Separate_wider_position

Separate_wider_position() separates by order (position). It splits a column into n new column where each new column is x characters wide.

The important argument here is widths which should contain a vector of the new column names + the width of each column.

EX: original column was MonthDay where the first two numbers are the month and the last two are the day

MonthDay
0718
0812
1201
0329
1111

Separate_wider_position(MonthDay ,
widths=c(Month=2,Day=2))



Month	Day
07	18
08	12
12	01
03	29
11	11

Separate_wider_position

But what does it do if there are MORE than x number of characters left over in the last column? Or what if there aren't enough characters to complete the task? See example below.

Take a look at the arguments “too_many” and “too_few” to see how you can solve this problem.

MonthDay
0718
0812
121
0329
1111

Separate_wider_position(MonthDay ,
widths=c(Month=2,Day=2))



Month	Day
07	18
08	12
12	???
03	29
11	11

Separate_wider_delim

Separates a column by a specific character (delineator), it splits a column into

new column where each new column contains the characters before or after the delineator character.

The important arguments here are `delim` which tells the function what the separating character is and `names` which should contain a vector of the new column names. The delineator is not kept.

EX: original column was MonthDay where the month and day are separated by

MonthDay
07_18
08_12
12_1
03_29
11_11

```
Separate_wider_delim(MonthDay ,  
                      delim = “_” ,  
                      names = c(“Month” , “Day”))
```



Month	Day
07	18
08	12
12	1
03	29
11	11

Separate_wider_delim

Anything can be used as a delineator. EX:

MonthDay
07D18
08D12
12D1
03D29
11D11

```
Separate_wider_delim(MonthDay,  
                      delim="D",  
                      names=c("Month", "Day"))
```



Month	Day
07	18
08	12
12	1
03	29
11	11

MonthDay
07*18
08*12
12*1
03*29
11*11

```
Separate_wider_delim(MonthDay,  
                      delim="*",  
                      names=c("Month", "Day"))
```



Month	Day
07	18
08	12
12	1
03	29
11	11

Separate_wider_delim

But what does it do if there are sometimes more delineators than new columns? For example, in the table to the right, does it add another column to hold the 23s? Or remove them completely?

Or what if there aren't enough characters to complete the task? See example below.

Again, this is where `too_few` and `too_many`

MonthDay
07
08_12
12_1
03
11_11

Separate_wider_delim(MonthDay ,
delim="_" ,
names=c("Month", "Day"))



MonthDay
07_18_23
08_12
12_1
03_29_23
11_11

Separate_wider_delim(MonthDay ,
delim="_" ,
names=c("Month", "Day"))



?

Month	Day
07	???
08	12
12	01
03	???
11	11