Introduction

The problem statement at hand is to analyse data for three species of fishes which were tagged and observed over the course of 2013-2021. The observation is primarily done through two arrays over the Dolores River between Utah and Colorado.

The species in question are Bluehead Sucker, Flannelmouth Sucker and Roundtail Chub. Each fish is tagged with it's attributes recorded and is then re-released into the river. The river has two main arrays of gates, namely, Rio Mesa and Disappointment. Both of these have upstream and downstream subarrays which are in-line with the riverflow. Each subarray has numbered antennas which are pinged when the fishes pass nearby.

The data at hand is split into three parts:

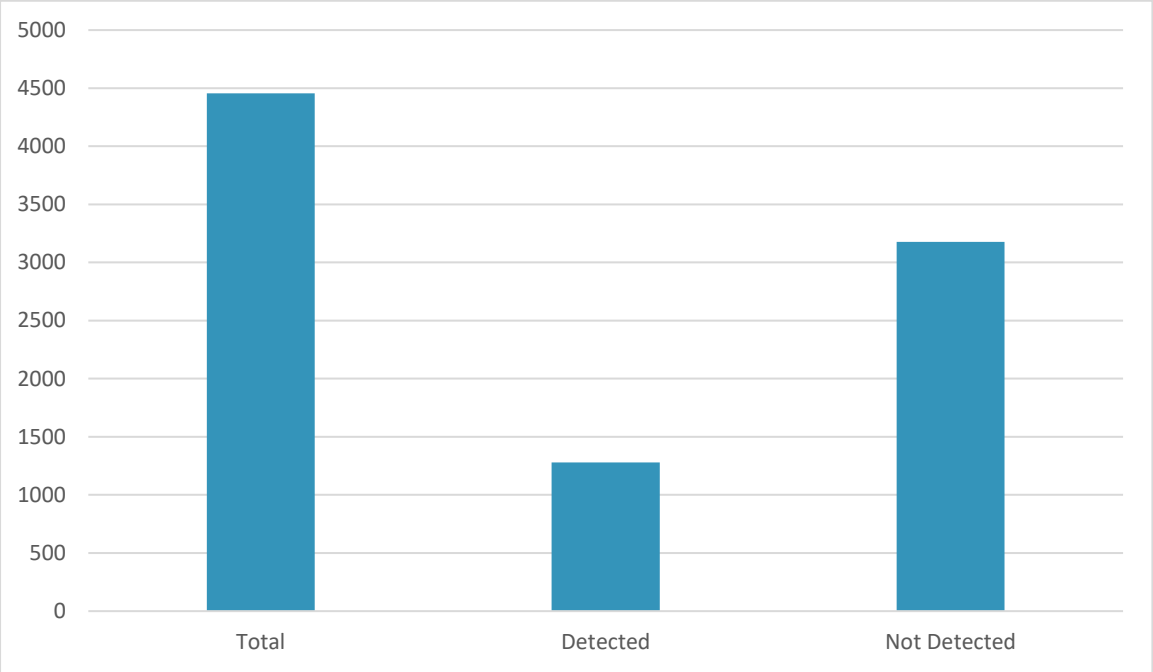
- Rio Mesa Detection (3844 x 4)
- Disappointment Detection (1146 x 4)
- Pit Data (Fish Attributes) (4457 x 11)

Below are some questions we're trying to answer:

- Site Fidelity: which sites get more year-on-year repetition
- Fidelity by Species: which species are more likely to return
- Is there a distinct migration pattern?

Percentage of Fishes Never Detected

The first thing we'll look at is that how many fishes were tagged but never detected. This would give us a basic place to start from and help us gauge the data at hand before me make assumptions about behaviour.



Out of a total of 4457 fishes tagged, 3178 or 71.3% were not detected at all. We're going to use the 1279 of the detected fishes for our analysis further. While this data is way less than the overall tagged fishes, to include the data of fishes never detected will skew our findings further. Therefore, we're only going to consider fishes that were detected.

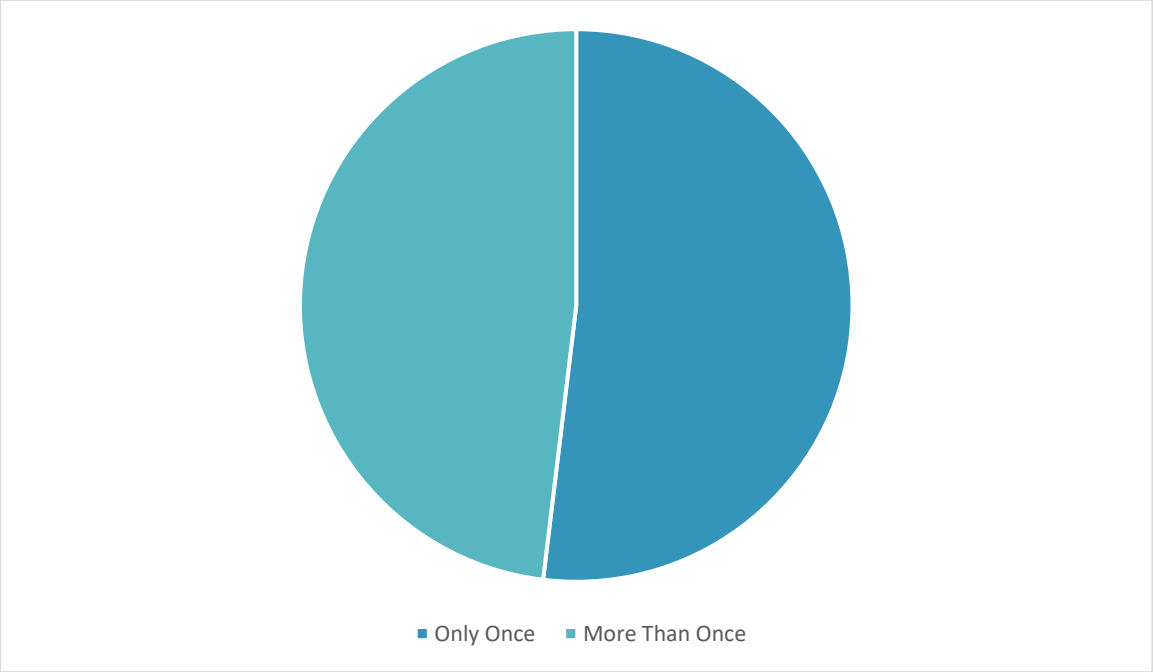
The reason for fishes not being detected could simply be them dying or them diverting to another river such as the San Miguel River.

Repetitive Behaviour

For this section, we'll only consider whether a fish was detected on more than one year. The difference between these years is not considered for this section. This analysis will give us a general sense of the repetitive behaviour. We will also explore streak behaviour further in this report.

Repetitive Behaviour (Overall)

Out of the 1279 fishes that were detected, 48% or 615 were seen in multiple years. The rest were seen only once and never detected again.

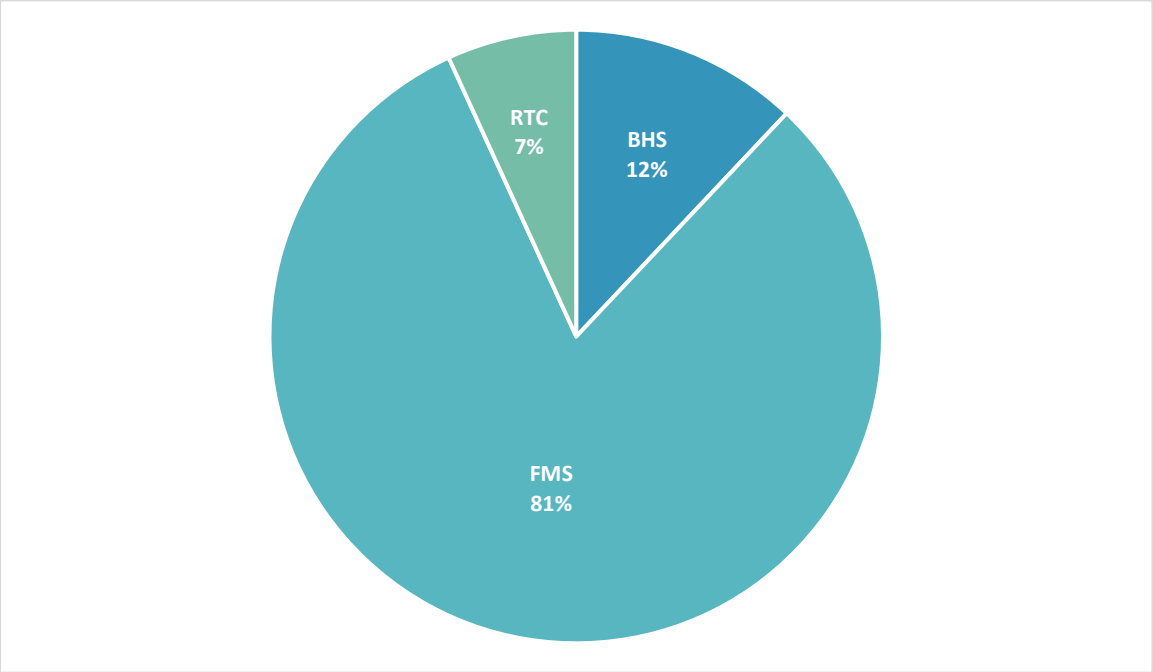


Next, we'll break this down further into Species and PIA. Then, we'll look at both together to understand if there is a pattern between specific species and PIA combinations. This could potentially reveal if a certain species prefers one segment of the Dolores among other things.

Repetitive Behaviour (Species)

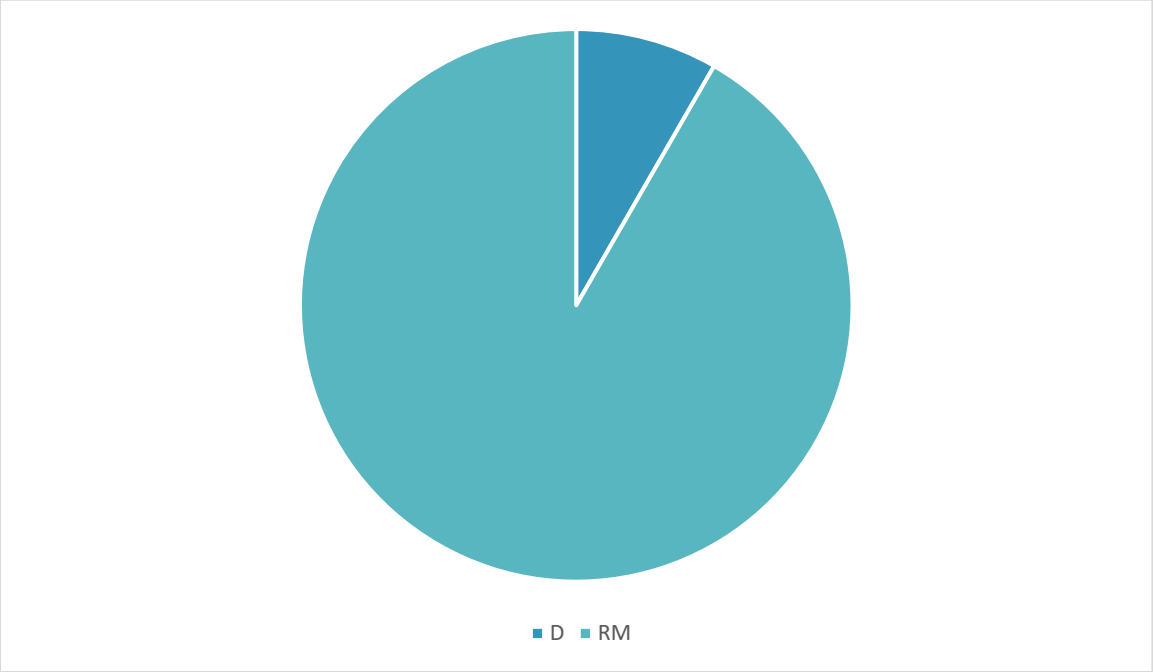
When we break the above into species, we see that the FMS show the most amount of repetitive behaviour. That is, most of the FMS species are seen in multiple years, followed by BHS with a large difference between both of them.

Numerically, the FMS species accounts for 499 out of the 615 while the BHS and RTC sit at 74 and 42 respectively.



Repetitive Behaviour (PIA)

Here, we'll break the data by the PIA or the arrays that the fishes were detected on. Interestingly, when we break the same data down by the PIA, we see that the fishes detected which was 615 is now reduced to 589. Since there are no NULLs in the data, we have reason to assume this is because of overlap (or the opposite) between the two arrays.

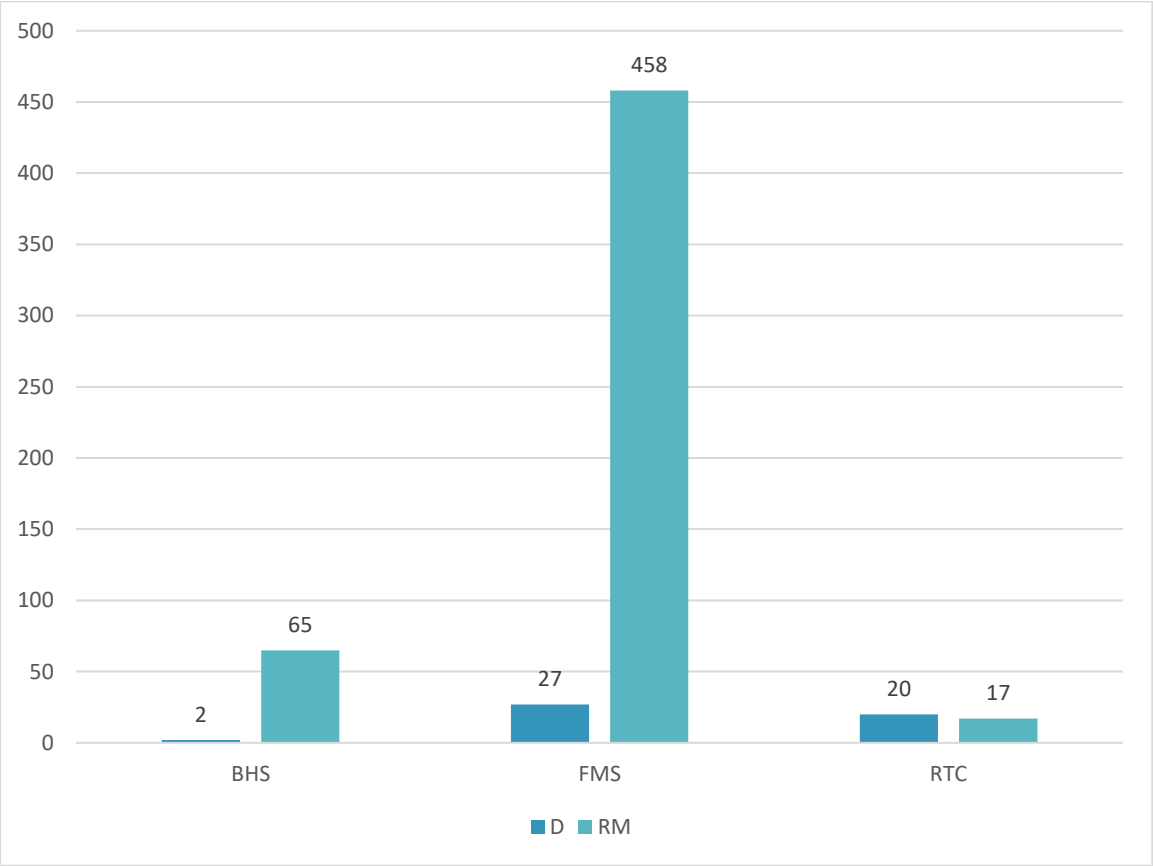


The Disappointment array only sees 8% of the overall detected fishes while the Rio Mesa array has 92% of the detections. The Rio Mesa array shows up significantly higher repeat behaviour when compared to the Disappointment array.

We will now explore both Species and PIA together.

Repetitive Behaviour (PIA and Species)

Exploring the data for repetitive behaviour on both PIA and Species would help us understand if there is a significant combination for when it comes to repetitive behaviour.



FMS on the Rio Mesa array shows a significantly high repetitive behaviour. It's important to note, we're not looking at consecutive repetitions as of now.

Mean & Median Lag After Tagging

How long does it take for a fish to be detected after it is tagged? That is the question we will try to answer in this section. We'll dive into whether this behaviour differs by species and PIA.

By Species

On average and by median, all three species are comparable when it comes to the lag after tagging. That is, almost all species are detected first after a year of tagging. The difference between the means could simply be owed to sample sizes. The RTC has an overall lesser share as we've seen above.

Species	Mean Lag	Median Lag
BHS	1.39	1
FMS	1.32	1
RTC	0.94	1

By PIA

In terms of the PIA, we notice that the fishes have a higher chance of being detected on the same year as their tagging. This could or could not explained by the site where fishes are released after tagging and which flow the fishes are put into.

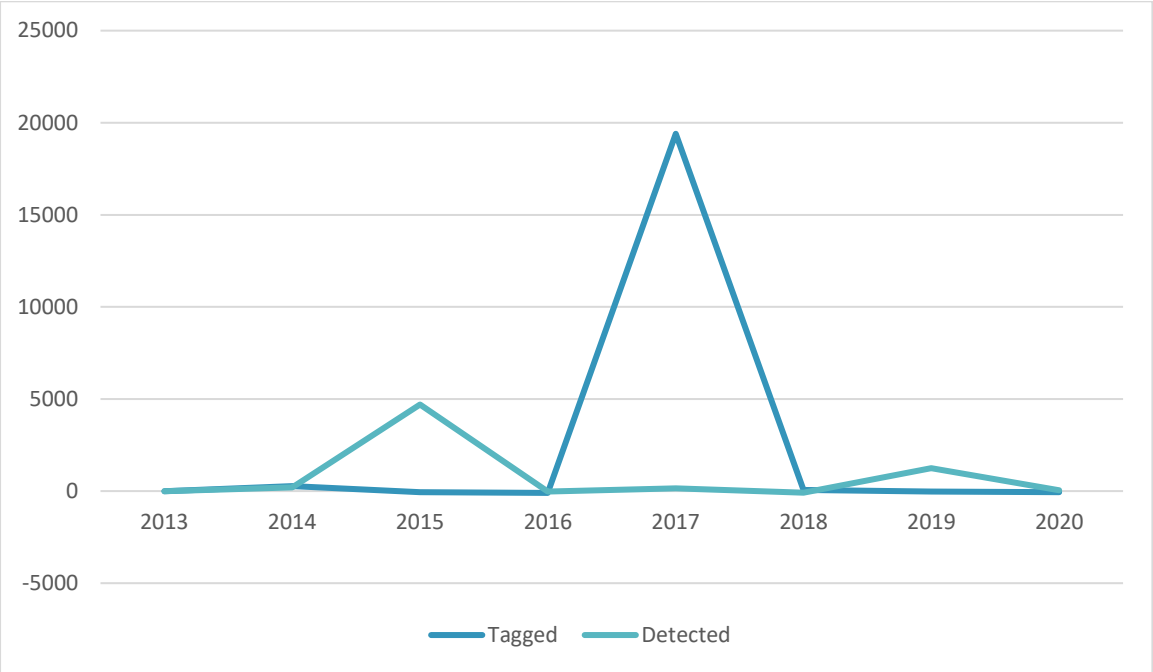
PIA	Mean Lag	Median Lag
D	0.785	0
RM	1.39	1

Percentage Change in Year-on-Year Detection

Here, we'll look at the per cent change in a trend-like manner for year-on-year detection of fishes. We'll bifurcate this on PIA and Species.

Overall

First of all, we'll look at a per cent change in the # of fishes tagged and detected over the years.



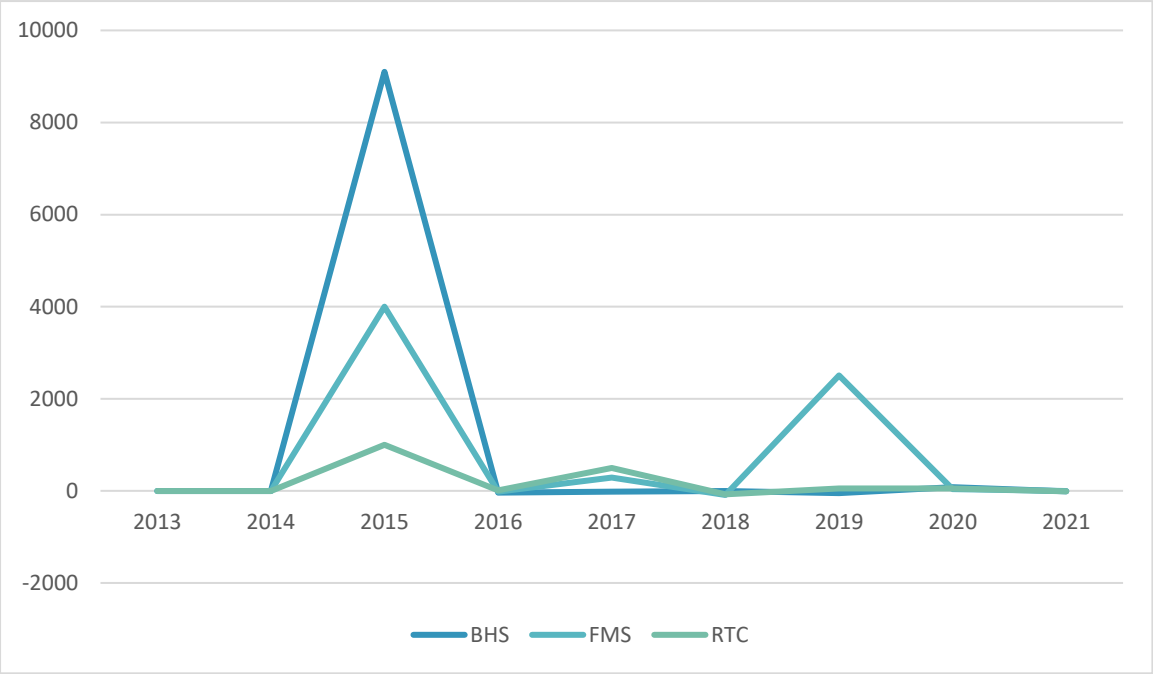
As we can see, the % change in tagging of fishes between 2016 and 2017 saw a huge spike going from 2 fishes to 390 fishes which is a per cent increase of 19400%. While this data does not give us much of a direct insight, it shows us that there is barely any correlation between tagging and detection of fishes. In fact, numerically, there is a correlation of only 8% between the raw values of tagging and detection.

The same data is visualised as a table below.

Year	Total Tagged	% Change	Total Detected	% Change
2013	88	0	1	0
2014	323	267.05	3	200
2015	113	-65.02	144	4700
2016	2	-98.23	107	-25.69
2017	390	19400	258	141.12
2018	610	56.41	33	-87.21
2019	509	-16.56	444	1245.45
2020	180	-64.64	636	43.24

By Species

When looking at the Species, we've only considered the Detections (per cent change). Overall, we can see that the RTC species has a more consistent detection rate in comparison to the others. While this doesn't add much value, we can conclude that RTC has the most consistent detection.

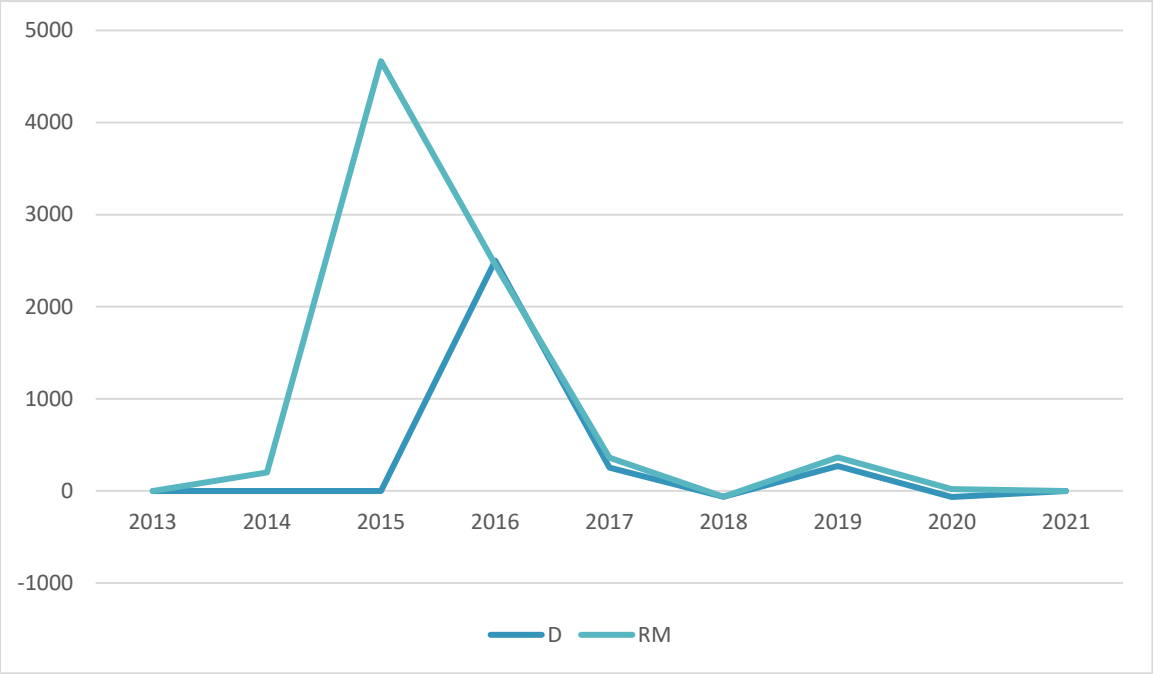


Here's a tabulation of the same data,

Year	Species	Total	% Change
2014	BHS	1	0
2015	BHS	92	9100
2016	BHS	60	-34.78
2017	BHS	49	-18.33
2019	BHS	25	-48.97
2020	BHS	45	80
2021	BHS	40	-11.11
2013	FMS	1	0
2014	FMS	1	0
2015	FMS	41	4000
2016	FMS	35	-14.63
2017	FMS	137	291.42
2018	FMS	15	-89.05
2019	FMS	391	2506.66
2020	FMS	549	40.40
2021	FMS	512	-6.73
2014	RTC	1	0
2015	RTC	11	1000
2016	RTC	12	9.09
2017	RTC	72	500
2018	RTC	18	-75
2019	RTC	28	55.55
2020	RTC	42	50
2021	RTC	37	-11.90

By PIA

Conducting the same analysis using the PIA, we see that the trends are more or less similar. So, we can conclude that the per cent change in detections does not depend much on the array where the detections take place.



The data is tabulated below,

Year	PIA	Total	% Change
2015	D	1	0
2016	D	26	2500
2017	D	92	253.84
2018	D	33	-64.13
2019	D	122	269.69
2020	D	41	-66.39
2013	RM	1	0
2014	RM	3	200
2015	RM	143	4666.66
2016	RM	81	-43.35
2017	RM	166	104.93
2019	RM	322	93.97
2020	RM	595	84.78
2021	RM	589	-1.01

Year-on-Year Repetition (Streaks)

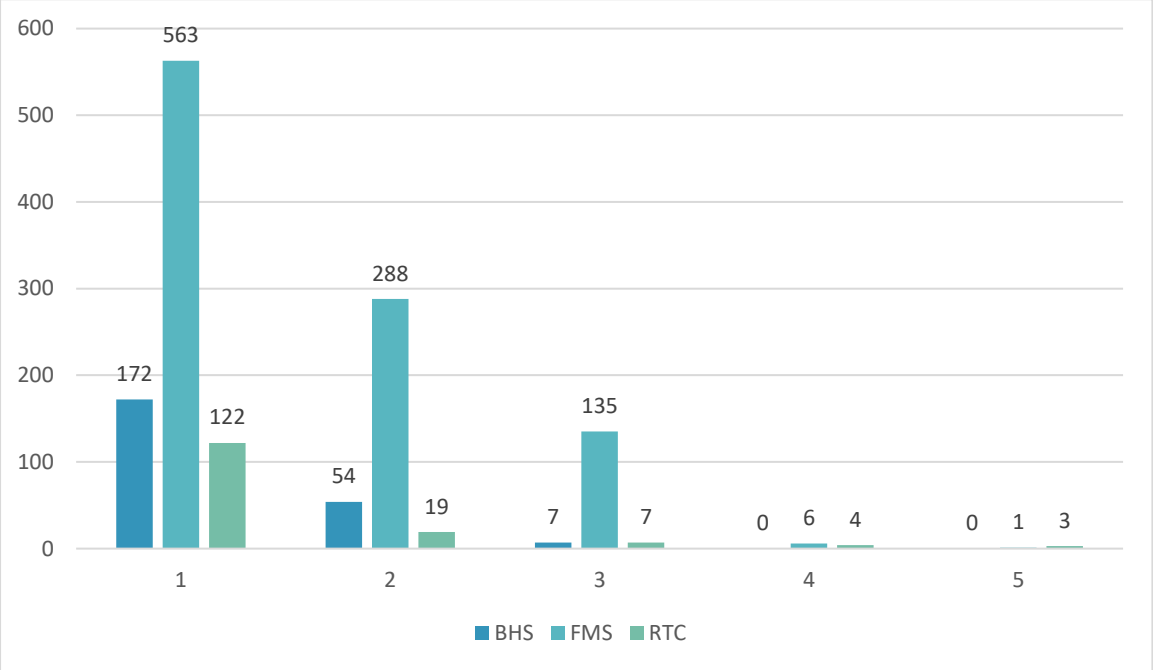
The last analysis conducted here is based on the idea of streaks. This is a way to see whether fishes show up for consecutive years and are detected. We will consider the maximum consecutive streak for each fish and analyse it on a mean/median level, as well as counts for each maximum streak.

Streaks By Species

From a species perspective, the maximum streak is comparable for all the species.

Species	Mean Streak	Median Streak
BHS	1.29	1
FMS	1.58	1
RTC	1.37	1

Below is a plot of the distribution of streaks and species,



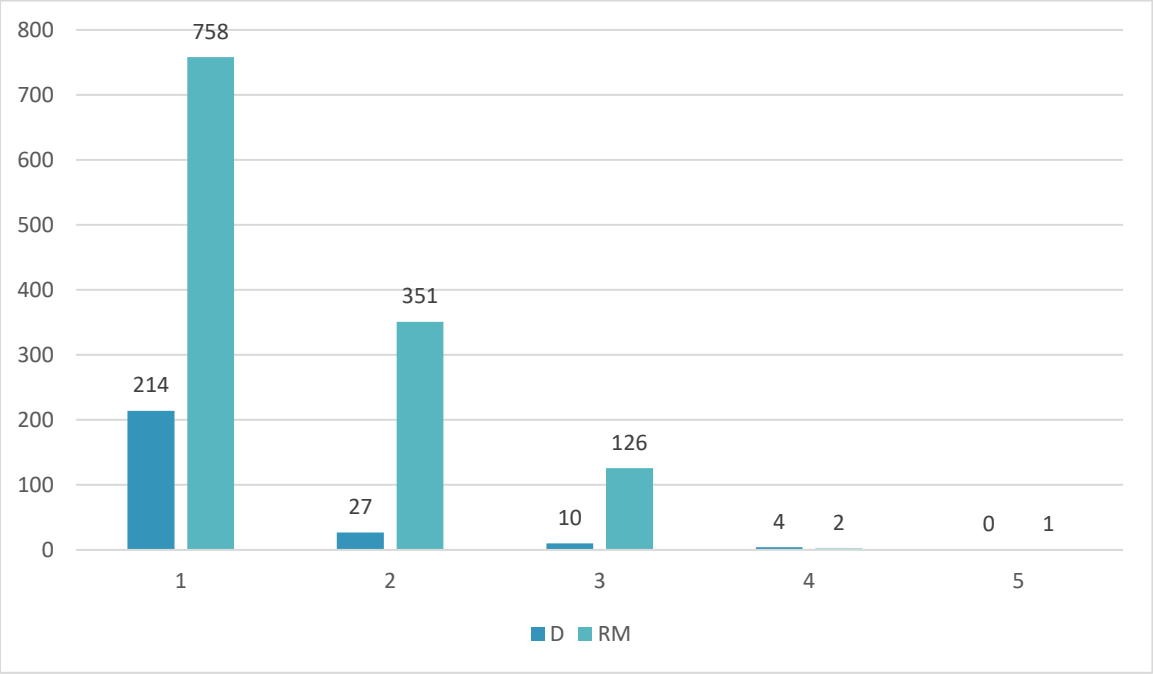
By comparison, FMS has the highest number of fishes detected with 1, 2 and 3 streaks. For fishes appearing in a streak of 4 and 5 years, FMS takes the lead closely followed by RTC.

Streaks By PIA

The maximum streak is comparable for all the arrays as well.

PIA	Mean Streak	Median Streak
D	1.23	1
RM	1.50	1

Below is a plot of the distribution of streaks and species,



It's clear that the Rio Mesa array sees higher number of fishes for all 1, 2 and 3 maximum streaks. On the contrary, the Disappointment array has more fishes which appear for 4 years in a row.

Streaks By Species & PIA

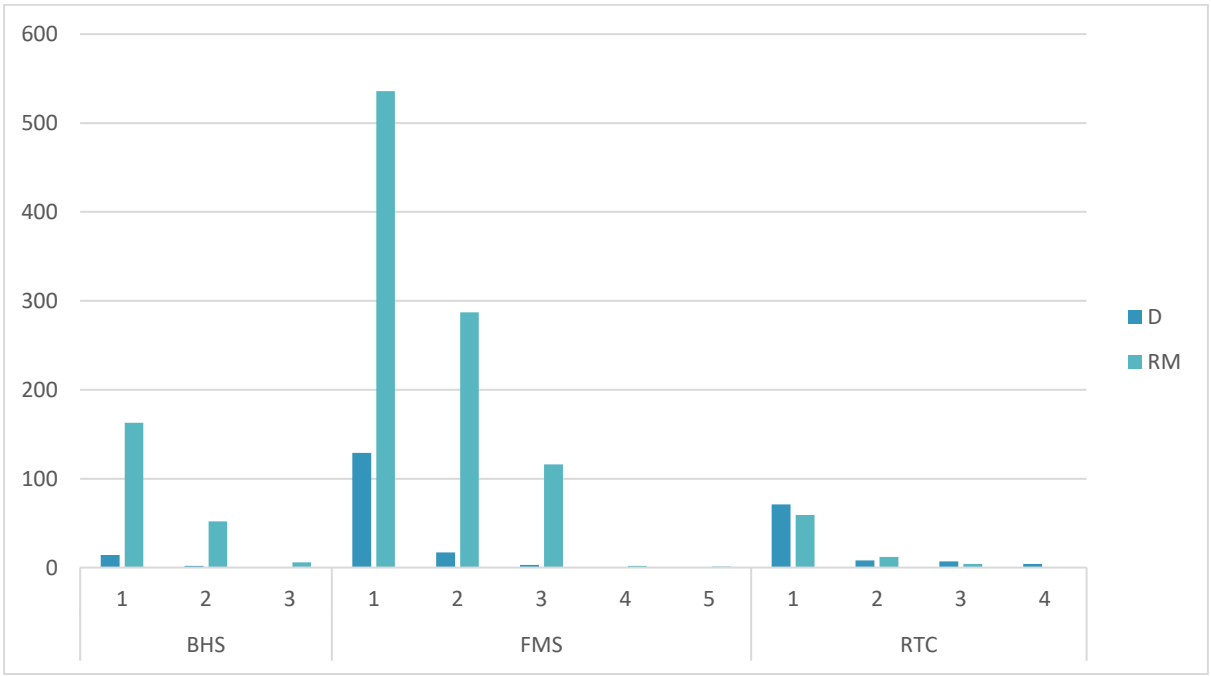
For consecutive behaviour, we're going to look at both Species and PIA together. This will help us understand if particular combinations change the consecutive repeat behaviour for the fishes.

From a measure of mean and median, all combinations are comparable and the central streak seems to be that of 1 consecutive appearance only.

Species	PIA	Mean Streak	Median Streak
BHS	D	1.13	1
BHS	RM	1.29	1
FMS	D	1.15	1
FMS	RM	1.56	1
RTC	D	1.38	1
RTC	RM	1.27	1

Below is a plot of the distribution of streaks by Species and PIA. From here, we can conclude that the combination of FMS with Rio Mesa sees maximum fishes returning for at least 3 years consecutively.

In other words, fishes tend to revisit the areas around Rio Mesa array, and those fishes tend to be FMS.



Conclusion

There are some closing points which are evident. There is a clear relationship between the FMS species and the Rio Mesa array when it comes to repetitive behaviour. We've seen this in both ways we've explored the data for multiple detections. There may be a pattern here but without significant knowledge and data, it is difficult to conclude the same.

Further, the data loss with the fishes which were never detected does make this analysis a bit off but if only the detections were considered, there is a significant repetition for some species and arrays in combination.

Also, more ways to explore this data could be to group the fishes by their physical attributes and see if their sizes or weights influence such behaviours. The dataset still has a lot of what we're not using from it and can be explored further in due time.