Mice Behaviour (FY1821): Project Report 1

Cynthia, Josh, Kewei, Srinidhi, Zach

With Prof.Andres Bendesky & Data Science Institute

Contents

Part A: Introduction

1 Goal

The goal of our capstone project is to find the link between the behavior of mice and their genetic makeup, if any. We will construct quantitative features based on the observed behavior of mice during an experiment and using the known genetics of the mice, we aim to find the specific genes that lead to observed behaviors. Before we began work, the lab had already conducted the experiment on the mice, calculated some behavioral features from the data, and linked them to the genetics of the mice. Our task is to explore and expand the feature space, and link the new features we calculate to the genetics of the mice.

2 Mice

The experiment looks at two different species of mouse with noticeably different behaviors in the wild: peromyscus polionotus (oldfield mice) and peromyscus maniculatus (deer mice). The oldfield mice are known to be more exploratory and navigate into open spaces while the deer mice are less prone to be exploratory and prefer closed, hidden spaces. Trials were run with mice from each of these two species, children of these mice, and grandchildren of these mice. The children and grandchildren were created from cross-breeding the mice (Figure 1), so each of the grandchildren has combinations of the genetics of both species. We can determine which gene segments come from which initial species of mouse and use that information to run tests and determine the specific gene segment that each behavior comes from.

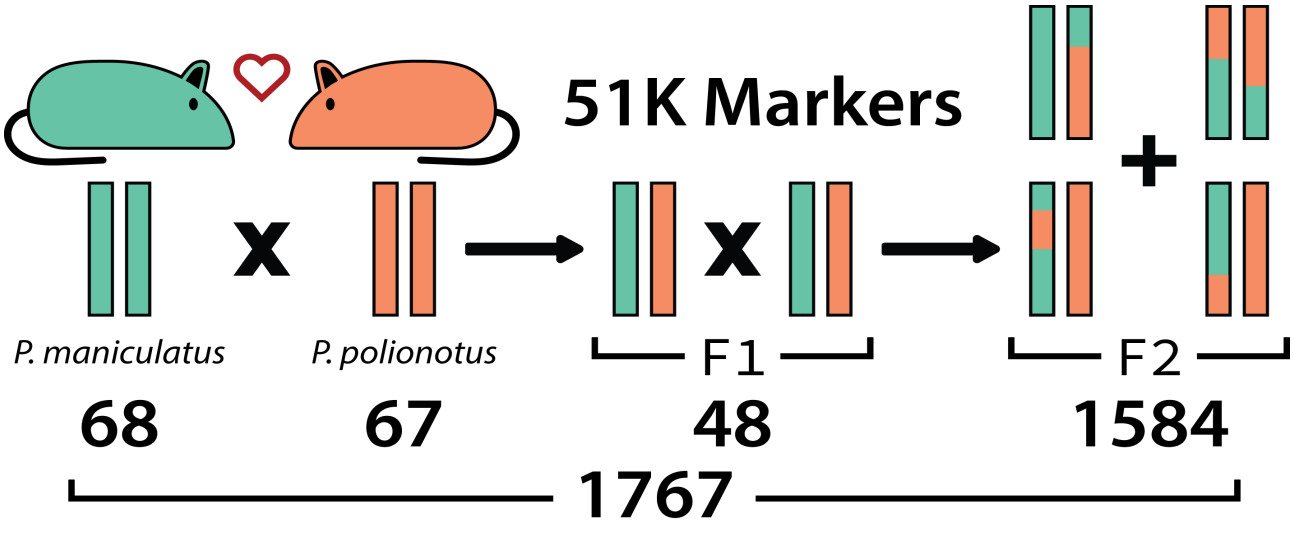

Figure 1: Mice: Pure breed, Generation 1 & 2 hybrids

3 The Experiment

In order to quantify their behavior, each mouse was placed in an elevated plus maze setup as pictured in Figure 2. The arms on the left and right have walls, so we will refer to these as closed arms. The top and bottom arms have no walls, so we will refer to them as open arms. The intersection region is referred to as the 'middle', which is open. We expect the oldfield mice will spend more time in the open than the deer mice.

Each mouse is put into the maze via a small "mouse elevator" and given some time to get used to their new environment. Then, the elevator is removed and the mouse is free to move around the maze. This experiment was run with over 1200 mice and the behavior of each mouse was recorded with a top-down camera for a maximum of 5 minutes. From each video, software is able to determine a bounding box of the location of the mouse at each point in time. We use this location data to derive all the behavioral features we want to study.

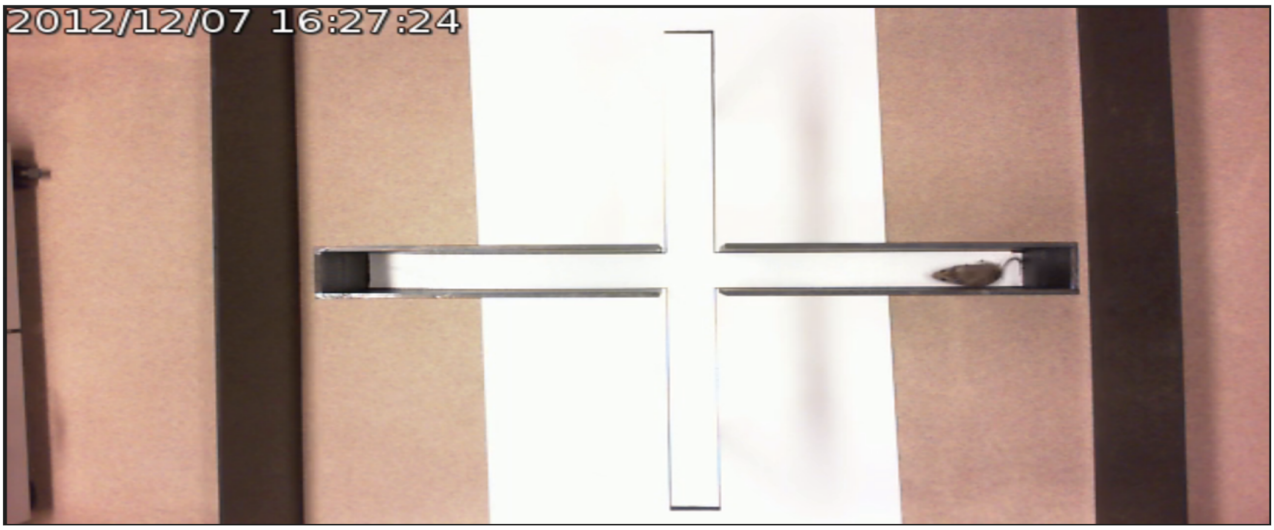

Figure 2: Aerial view of the elevated plus maze. The mouse is located in the closed right arm.

4 Contribution

Much of the work described above was done before our capstone project began. We were given files including video of all the trials, location data of each mouse over time, a script calculating a set of behavioral features, and work done to map those features to genetics of the mice. Our task is to use the raw location data provided to create additional behavioral features and use the same process as done previously to link our new features to mouse genetics, if any.

Part B: Data Exploration

1 The Data

We were provided with access to a shared drive with data in the form of videos and other files containing information extracted from the video. Additionally we were given a script to run that calculates some features from the data provided. After downloading the data and altering the script we were able to run the script on all our data and output a list of initial features to a CSV. When we add more features we'll be able to run the script again and calculate those features for all the mice.

2 Feature Engineering and EDA

· Using the script provided to us to calculate the features that we already had

· Editing the script provided to us to migrate all the features results into a single file (rather than stored separately in their respective files)

Some of the features have been provided from the original data, including:

Mouse Features: Strain, Sex
Behavioral Features (calculated by the script provided to us):

* Fraction of time in each arm
* Median Speed & Smoothed* median speed in each arm
* Total distance & Smoothed* total distance travelled in each arm
* Number of Entries into each arm

*Smoothing is the process of smoothing the data spikes resulting from the imperfection of the tracking software(sometimes, the location may jump around by a frame or two)

In order to explore the data, we altered the script provided to us and used it to create a single CSV file with every mouse and data point that was calculated before we started on the project. Below is a summary of all the mice we have data for

Strain	Count
PO (oldfield mice)	65
BW (deer mice)	57
BWPOF1 (first generation hybrid)	45
BWPOF2 (second generation hybrid)	1014
Other	4

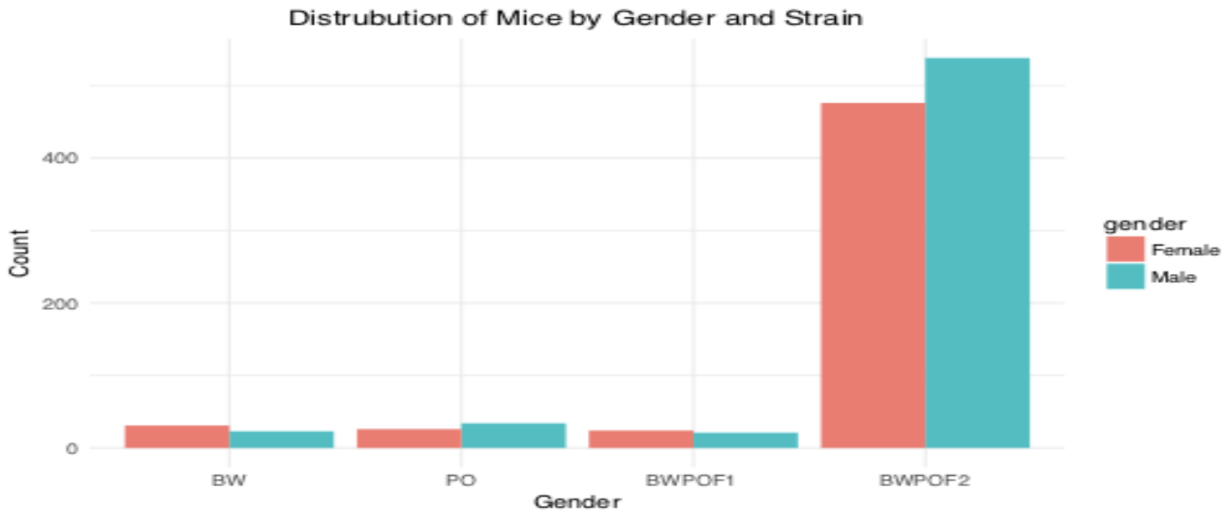

Figure 3: Distribution of mice by gender and strain

Some mice did not conform to the defined naming conventions for strain and gender ("Other" in the tables above). As there are not too many that deviate from the norm, we chose to remove them and were left with 1173 mice with data. Figure 3 shows the distribution of the mice by gender and strain. Most of the observed mice are BWPOF2 and the rest are distributed evenly between the BW, PO and BWPOF1 strains. The distribution of gender is roughly even in each of the strains.

We want to explore behaviors that are demonstrably different between the pure bred mice (BW, PO). Once we've determined behaviors where the breeds differ, we can explore the genetics behind those differences using data from all species.

As an example, let's look at the fraction of time the mice spent in closed vs open arms. Figures 4 and 5 show this result for the pure-bred mice. The distribution by gender does not seem to vary significantly for each strain, suggesting that gender does not not play a significant role in the behavior of the mice. However, we do see a significant difference between the two pure breeds in their behavior. The BW mice spend a larger portion of their time in the closed and middle arms than the PO mice and conversely the PO mice spend a larger portion of their time in the open arms than the BW mice. This is consistent with the expected behavior of the mice as the PO mice are more adventurous and the BW mice prefer closed spaces.

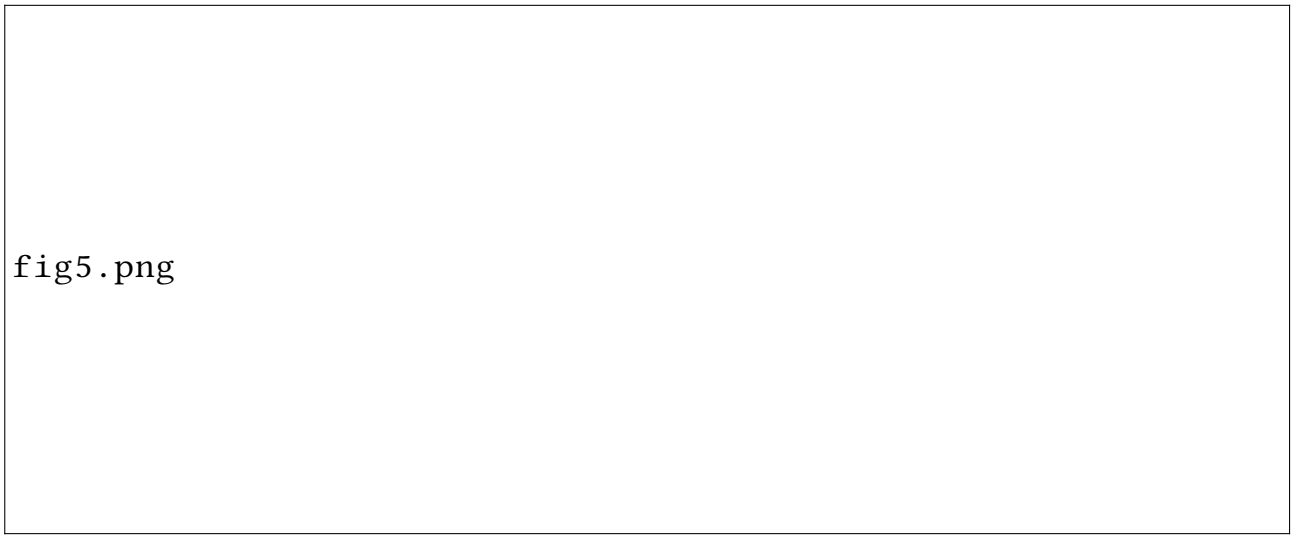

Figure 4: Distribution of the fraction of time spent by each strain in the closed and middle sections of the maze

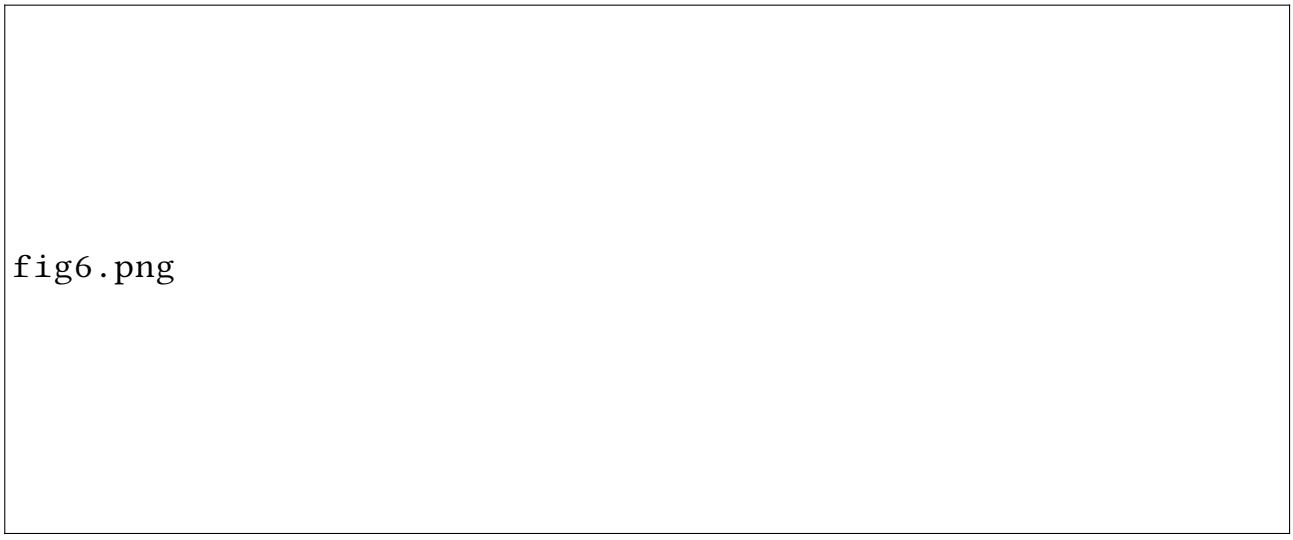

Figure 5: Distribution of the faction of time spent by each strain in the open sections of the maze

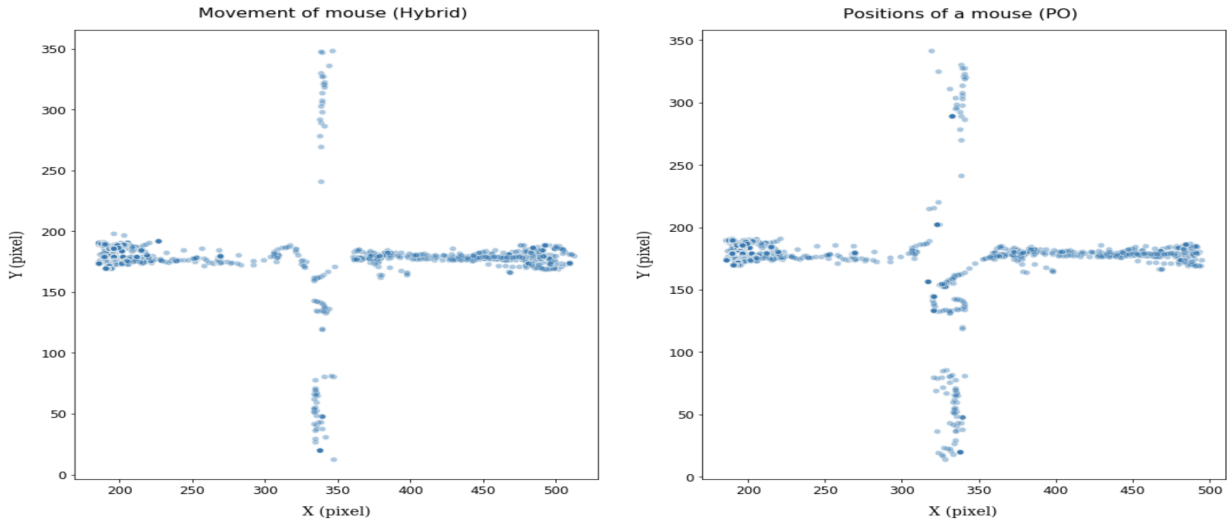

Figure 6: Map of movement of mice: The higher preference of open arms for the PO is visible

Now that we know there is a behavioral difference between pure breeds, we can look at the first and second generation hybrids. In Figures 7 and 8, we see that different generations of the hybrids do not vary too much in the distribution of time spent in each of the sections of the maze. In fact, it appears as if the hybrid strains could be a combination of the distribution of the non-hybrid strains. This is promising, as we can use the behaviors of different mice to find out which genes lead to the exploratory behavior.

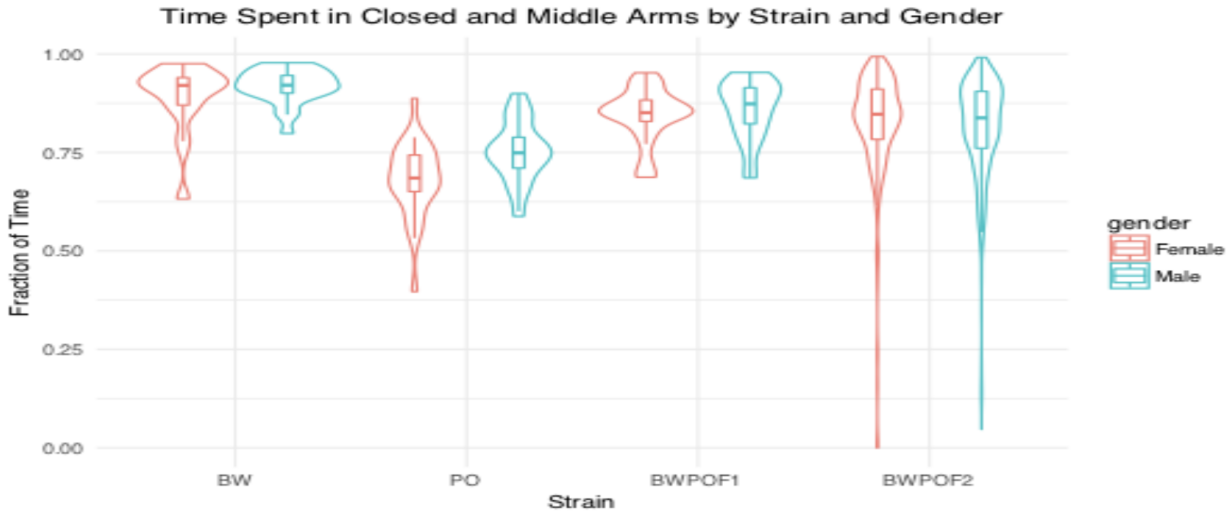

Figure 7: Distribution of the fraction of time spent by the non-hybrid mice and the first and second generation hybrids in the closed and middle sections of the maze. Note: BWPOF2 has outliers that we will look into

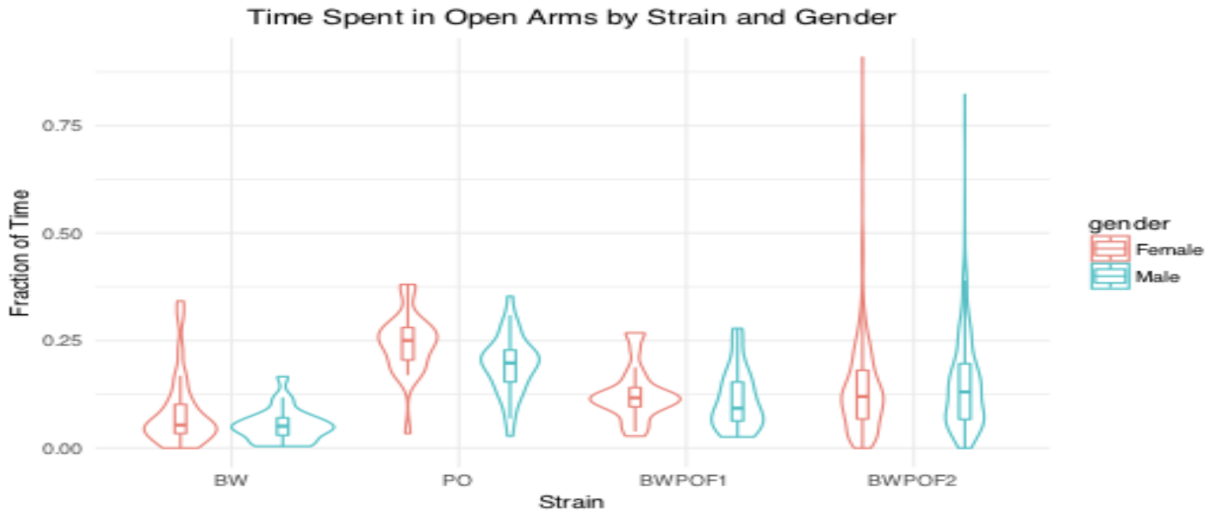

Figure 8: Distribution of the fraction of time spent by the non-hybrid mice and the first and second generation hybrids in the open arms of the maze. Note: BWPOF2 has outliers that we will look into

As a second example, let's look at median speed. As seen in Figure 9, when we look at the median speed of pure mice breeds, we don't see much of a difference. This would leave us to believe that the median speed may not differ much between breeds. If this is the case, we can ignore this feature in the future as we are only looking for behaviors where the pure breeds differ.

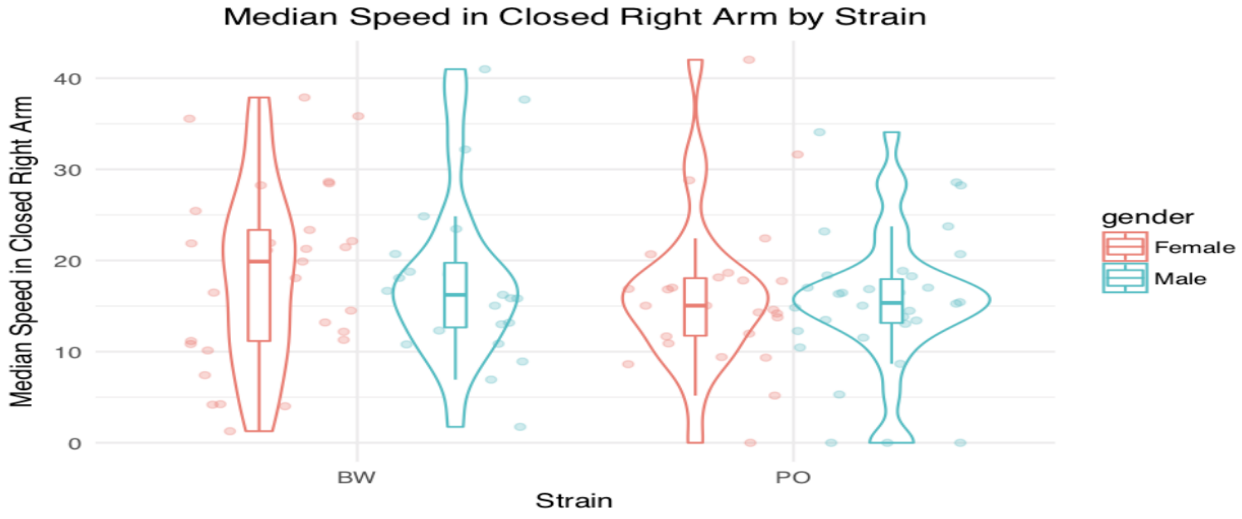

Figure 9: Distribution of the median speed spent by each strain in the closed right arm of the maze

Part C: Related Literature

This is not the first experiment done by the Bendesky lab concerning the mice species in our dataset. Previous work (https://www.nature.com/articles/nature22074) has been done with these animals looking into their parental behavior. Similar to our data on exploratory behavior, the sister species peromyscus polionotus (PO) and peromyscus maniculatus (BW) display drastically different behaviors when it comes to parental care. The previous study explored differences between

the strains in a variety of parental behaviors like licking and huddling. The study concluded that the monogamous BW mice were more parental across observed species compared to the PO mice and they identified genetic regions associated with the observed behaviors. Our capstone work is mirroring the Nature paper, but with a focus on exploratory behaviors. We currently have isolated different features and observed the differences across strains. Future steps will involve extending the work to other features and continuing the template outlined in the paper, as discussed in the next section.

Part D: Future Steps

1 Add New Features

Having looked at the features already provided to us, we will add more features that we hope will capture differences in exploratory behaviors of each strain. Once we have the code implemented for the features and we are able to manually test that code using video files of the mice, we will run the updated script to get a more comprehensive data set for each of the 1173 mice in the study. Below is a list of features we will continue to add(we have added a few already, others just need to be scaled to the entire dataset):

Feature	**Reasoning / Hypothesis**
Median/Average directional speed per arm	Mice that are more fearful of the open areas might be more likely to move into the closed areas quickly and into the open areas slowly
Fraction of time at rest (In closed/open arms)	'Fearful' mice might spend more time at rest inside closed arms and less fearful mice might be more willing to rest in open arms
Fraction of time in safety	The end of the closed arms are considered "safe" areas and fearful mice might spend more time in safety
Fraction of time peeking	We measure the amount of time mice spend peeking from closed arms into the open and hypothesize mice that like to explore may be less cautious
Average time of peeking	We measure the amount of time mice spend peeking from closed arms into the open and hypothesize mice that like to explore may be less cautious
Region to Region Frequency	Count the number of times that a mouse moves from one specific arm to another (including going back to the same arm)
Turn Frequency	To capture the left or right turn preference of Mice
Mouse size	We measure the size of each mouse

2 Testing Features

For each of the features we need to determine if there is a statistically significant difference between the PO mice and BW mice. To do this, we plan to do similar analysis as laid out in the EDA section. We will compare the violin plots by strain and gender for each group of mice. Looking at the distributions(and significance tests results) of the pure bred mice will allow us see if there is a difference between the two distributions. For example, if we see that the BW mice turn right far

more than the PO mice, then we can look at the second generation offspring and determine what gene segments lead to this difference in behavior.

Conclusion

Our project aims to better understand the exploratory behavior of the oldfield and deer mice strains and discover their relationship to genetic markers. We analyze footage of the mice in a maze to create quantifiable behavioral features. Using violin plots to visualize the distribution of the extracted features, we find features where the behavior of the two pure strains of mice are different. Future work will involve adding more features and connecting the observed features to the genetic data of the offspring of oldfield and deer mice. Once we have a set of features unique to each strain, we will analyze the behavior of the first and second generation mice and try to map the behavior back to a genetic marker. This approach will help is to concretely determine if the exploratory characteristics have a genetic basis.

Bibliography

1. Cover Picture: http://www.wildflorida.com/wildlife/mammals/Oldfield_Mouse.php
2. Cover Picture: https://www.youtube.com/watch?v=uXdzuz5Q-hs
3. Smoothing: http://scipy.github.io/old-wiki/pages/Cookbook/SignalSmooth
4. Project GitHub Repo: https://github.com/zachbogart/mice_behavior_capstone
5. Paper by Prof.Bendesky: https://www.nature.com/articles/nature22074