

What are we doing?

Why does it matter?

- Professor Coltfelter is a biologist. He likes to study animal behavior - what they do and why they do it?

How is the data collected?

- By catching birds and putting tags on them. Each tag has an identifying frequency.
- There are 10 special bird feeders. When a bird lands on a feeder, it records the time, the day, and the frequency (which bird is it?).
- Then, we can find out who is dominant to whom.
 - Imagine a bird that is about to start feeding at a feeder. However, this bird leaves because a second bird arrives within a second of the first bird. If the second bird stays at the feeder for 5 seconds after the first bird leaves, then it is dominant.
 - We can then rank the birds through an accumulation of interactions that determine dominance. Look at each one and see which bird “wins”.

What data do we have?

Access?

We have access to three things right now:

- **Feeder Information:**
 - **Antennae** - Binary identifier for which side the bird landed on. (0 or 1)
 - **NOTE- We don't know which sides are 0 and 1.**
 - **Code** - A unique identifier for each bird. The numbers themselves are meaningless, they just come from the manufacturer of the tag.
 - **Time** - Hour, minute, and second. Recorded in military time.
 - **Date** - Formatted as month, day, and year. (EX: 01/01/21)
 - **Format** - Automatically put into .txt files. Each feeder produces its own .txt file. Files are retrieved by physically walking out into the sanctuary and accessing a Wi-Fi hotspot. Then it is automatically put into a Google Drive Folder and then manually input into R.
 - **NOTE - We don't know which .txt file belongs to which bird feeder. This is not included as a data column. We should manually create a column for this.**
 - **NOTE - Birds move around to various feeders.**
 - **NOTE - Manually putting information into R is awful. Could we automate this?**

- **R Code and its accompanying research paper.**

The R code does a lot of the heavy lifting and calculations. It relies on the information from its research paper. All the information below comes DIRECTLY from the R file:

- **Arrive Time** - time of arrival in seconds
- **Depart Time** - time of departure in seconds
- **IDs** - vector of corresponding ids
- **Displace Time** - difference between one individual leaving and another arriving to be considered a displacement
- **Total Time** - total time an individual spends at antennae.
- **Difference Time** - difference in time between an individual departing and an individual arriving.
 - **NOTE - It is recommended that this is calculated prior, and visits by the same individual within a short period are merged.**
- **Sites** - vector of sites
- **Divisions** - vector of by which to divide the data. This can be less computationally intense than calculating a whole site at once.
- **Format** - Running the R script produces a dominance matrix based list of detected interactions.

- **Bird characteristic database.**

- Not much is known about this right now. Professor Clotfelter has to clean up the data before we can see it.
 - However, we know it has information like wingspan, species, and the unique identifier for each caught bird.

Cleaning?

- Tons of different text files. We need to compile all the text files that have ever existed (and will exist) into a single database.
 - **NOTE - Do we want the feeder information and bird characteristics to exist separately? Should we combine them?**
- We need to keep the functionality of the R file somehow.

How much?

Not sure of the exact amount, but the research has been ongoing for a few years.

- Each row of the .txt files records a *second* that the bird was standing on one of the antennae of a specific bird feeder.
 - **NOTE - We should probably find out the year it started.**
- The study year is fall and winter. Essentially, November until nearly the end of march.

What will you deliver to your customer?

Final vision?

- The goal would be to contact IT and see if we could secure a *web hosting service* for a website.
- Essentially, we want to create a website with the functionalities listed below.

Functionalities?

- Automatically collect data from Google Drive, so the .txt files don't have to be manually uploaded.
- A real-time dominance hierarchy as data is input. Who is the top bird? Who is the bottom bird?
- General analysis, such as: Which species is the most dominant? Are certain birds more active at different times of day? Which feeder is the most popular?

Limitations?

- As of now, only members of Amherst College should be allowed to view the website.

Who will use it? What about after we're gone?

- This will be used for years by thesis students, researchers, and his biology class.
- As use expands, there's a possibility that he will want it to continue to grow.

