Fieldwork and coding internship experience

Yuki Kawahara（mesa college, computer science）

My name is **Yuki Kawahara**. I am a second year student at **Mesa College** majoring in **computer science**.
I've always loved the natural environment, and as a child I wanted to be a scientist, but various things prevented me from making that dream come true. However, with this internship opportunity, I was very excited to find out that the computer science I'm currently studying can be involved in research.

In this blog, I will summarize the steps and my impressions for each of the three research stages.

1, fieldwork.

2，compiling data.

3，data mining using programming.

4，Using R to analyze data

I did fieldwork with my mentors Xavius Boone and Aydan in three rivers in San Diego: San Elijo, San Dieguito, and Los Penasquitos.

We set up underwater cameras with squid bait on the bottom of these rivers and recorded them.

We investigated the relationship between the number of fish captured on the camera and the water temperature.

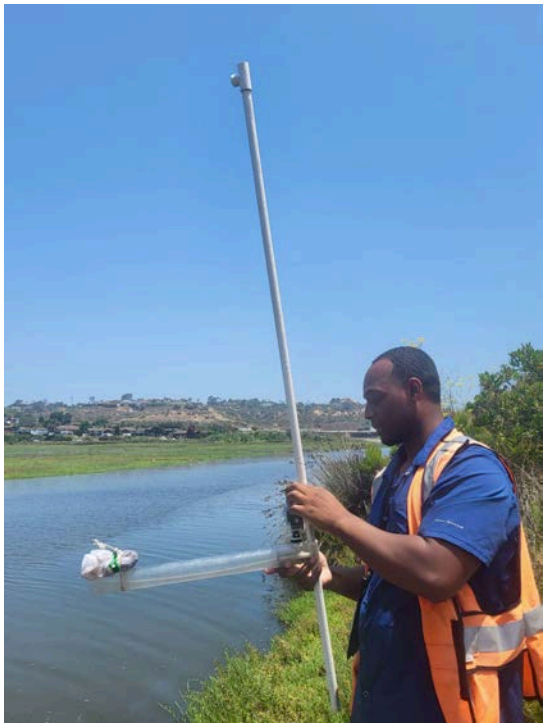
1，fieldwork

Fieldwork steps

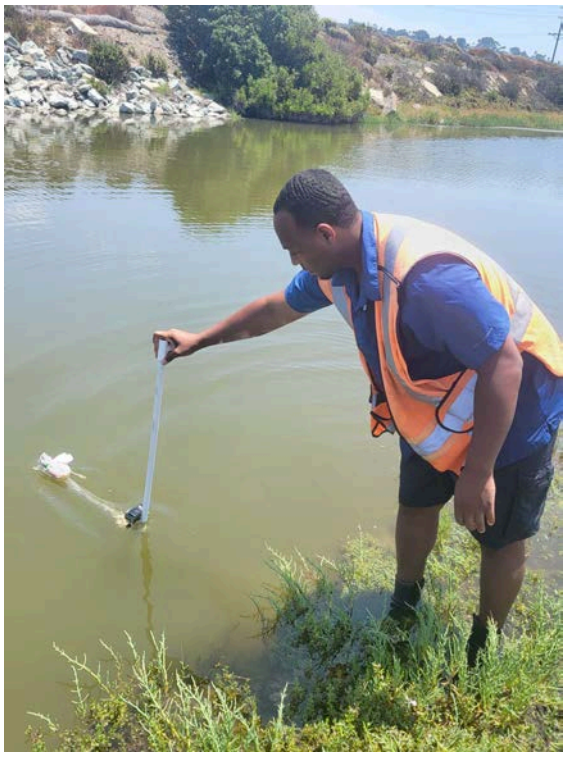
1. Attach a squid as bait to the end of the tube.

2. Install the camera

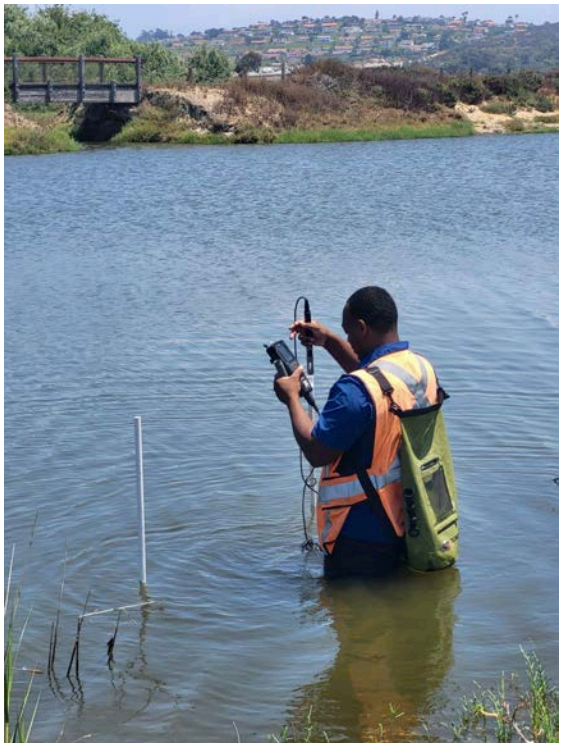
3. Attach a camera and other equipment to a pole that will be fixed to the riverbed

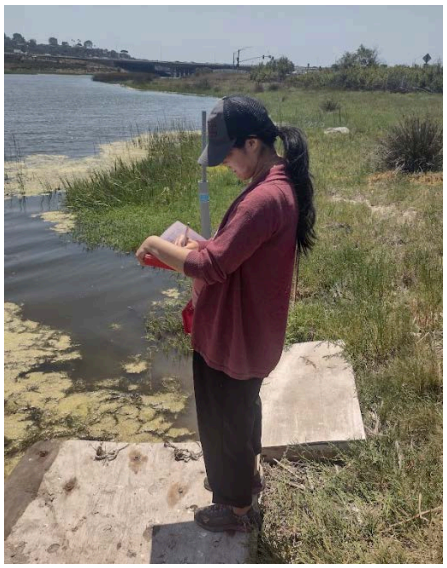


4. Install on the riverbed



5. Measure and record the temperature of the riverbed and the river surface





6. The camera will be collected after an hour and a half

Impressions:
It was so much fun to go into places that I wouldn't be able to access unless it was for research, and see the natural scenery, fish, and birds. It was exciting because it felt like I was on an adventure with my mentor and the other interns.

2. Summary of data

After the fieldwork was over, I checked the video and counted the number of fish.

I used Excel or Google Spreadsheets.

I recorded the number of fish, the type, and the number of fish that were on the screen.

It was quite a pain to look for fish when there were no fish.

But it was a really great experience to see fish in the water that you don't usually see.

It was fun to watch the fish approach and move away from the squid bait,

and to watch the fish school back and forth.

3. Data mining

After the data is collected, the task is to shape the data. This is what programming requires. What I did was to write a code that uses python to automate data mining.

The code I wrote is as follows

1. connected the code to a Google Drive folder

```
# connect your drive to this code
from google.colab import drive
drive.mount('/content/drive')
```

2. Modules were imported.
The modules used were pandas, os, glob, and openxl

```
# import modules
import pandas as pd
import os
from glob import glob
import openpyxl as op
```

3. Import the folder

```
# import folda
folda_path = r'/content/drive/MyDrive/学校/intern/derectory_files'
```

```
# check files in the folda
files = os.listdir(folda_path)
files
```

```
# check first file in the folda
df = pd.read_excel(rf"{folda_path}/{files[0]}")
df.head(5)
```

4, Create a new folder (this is where you will put the completed file)

```
# make new folda
upper_folda_name = os.path.dirname(folda_path)
# print(new_files)
new_files_path = f"{upper_folda_name}/new_files"
new_folda = os.mkdir(new_files_path)
```

5. Create a new Excel file (contain the files you want to mine in this file)

```
# make new excel file
wb = op.Workbook()
# save excel file
wb.save(f"{upper_folda_name}/new_files/new_excel.xlsx")
```

```
# read new excel file by pandas
df_base = pd.read_excel(f"{new_files_path}/new_excel.xlsx")
```

6. The files in the folder will be automatically mined and the mined files will be merged into one

```
for file_index in range(len(files)):
    df = pd.read_excel(f"{folda_path}/{files[file_index]}")
    # Delete rows with missing values in the rep column
    df.dropna(subset=["rep"], inplace=True)
    print(df.head(5))
    new_df = df[["estuary", "station", "month", "maxns",
"total.ns", "species.richness", "scientific.name"]]
    df_base = pd.concat([df_base, new_df])
```

```
# check new excel file
print(df_base)
```

7. After data mining, combine the files into one new file and place it in the (df_base) folder

```
# save result excel file
df_base.to_excel(f"{new_files_path}/new_file.xlsx")
```

8. The Excel file new_excel.xlsx containing the data is no longer needed, so delete it

```
# delete new_excel.xlsx
os.remove(f"{new_files_path}/new_excel.xlsx")
```

Good points:
I learned what kinds of problems people face in research and what kind of code is needed.
I liked that I got to actually practice writing code.

Challenges:
I realized that I didn't have enough knowledge about data mining.
I felt the importance of creating applications that are easier to understand for people who don't use Python.

4, Using R to analyze data

In R programming, I learned how to create graphs using a library called ggplot.

import libraries

```
library(dplyr)
library(tidyverse)
library(googlesheets4)
library(readxl)
library(esquisse)
```

```
library(readxl)
```

set directory

```
setwd("###derectry path")
```

import file path

```
fish = read_xlsx("###file path")
head(fish)
```

check type

```
str(fish)
```

change station(num) to character

(The station column is a numeric type, so unlike Python, in R the graph won't be displayed properly unless the type is a string type, so I changed it to a string type.)

```
fish$station = as.character(fish$station)
```

chaeck type again

```
str(fish)
```

write graphs

```
fish%>%
  ggplot() +
  aes(x=station, y = `sal.bottom (ppt)`, fill = estuary) +
geom_boxplot()
```

It was my first time learning R programming, but thanks to my mentor's guidance, I was able to create graphs in just two months, which I was thrilled about.
R programming is similar to Python in some ways.
There are many commonalities in the writing methods, etc.
Python can be used for a wider range of purposes, but I think R programming has many advantages over Python.
Especially when creating graphs, there is less code than Python, and it is simpler and easier to understand.
I was also happy to learn that the R language is often used in research.

I am still a beginner when it comes to the R language, so I would like to continue studying.

Lastly,

It's rare to have the opportunity to work in a research setting, so this was a valuable experience. I'd like to thank the staff at SDSU's Coastal Marine Institute, the staff at Mesa College, Xavius, and Aydan for this valuable experience.