Mirman School

Computational Data Analysis

Seeing the story through code

Michael Taggart
Director of Technology

Programming is a means, not an end

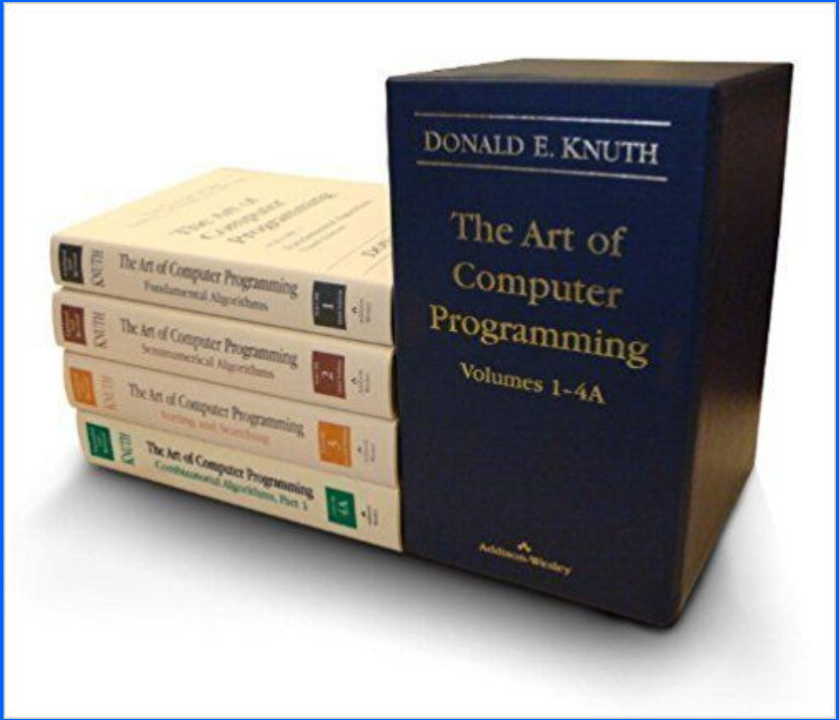
Mirman School

- Learning to “code” shouldn’t just be for apps/games
- Profound way of thinking
- It’s a way of approaching problem solving
- How can we use programming to recontextualize problems in other subject areas?
- How can we make programming a skillset for math/science inquiry?
- Programming is about processing information
  - Information has a shape, aka a “model”
  - So does data in scientific inquiry

Literate Programming

Mirman School

- Donald Knuth
  - “Let us change our traditional attitude to the construction of programs: Instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to human beings what we want a computer to do.”
- Computer instructions as a result of human reasoning



Jupyter

Mirman School

- A tool for easily creating literate programs in one's preferred language
- Mixing text with code snippets to build understanding
- try.jupyter.org

What Literate Programming Looks Like

Mirman School

Line/Scatter plots with more than one variable

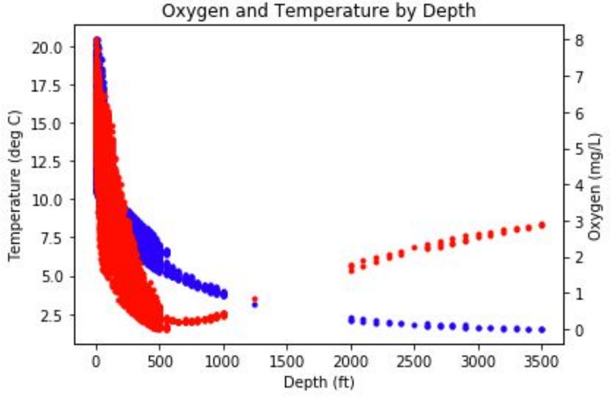
Say you wanted to compare depth, pressure, and oxygen. Your hypothesis might be, "As depth increases, oxygen increases and temperature drops." Given that description, depth would be your x axis, and you'd put pressure and temperature *both* on your y. To do this, we'll need to manually create our plot with the `pyplot` module, used here as `plt`.

We're going to do this as a "scatter" plot, which is like a line plot, but with dots instead of lines. The lines might get messy.

```
plt.clf() # Again, clear all previous plots
fig, axes1 = plt.subplots() # Create the figure and our first set of axes
axes1.plot(df.depth, df.temperature, "b.") # The "b." is a style string that sets the line style to points (.) and in blue (b)
axes1.set_ylabel("Temperature (deg C)")
axes1.set_xlabel("Depth (ft)")

axes2 = axes1.twinx() # Clone the first pair of axes
axes2.plot(df.depth, df.oxygen, "r.")
axes2.set_ylabel("Oxygen (mg/L)")
plt.title("Oxygen and Temperature by Depth")
```

```
Text(0.5,1,'Oxygen and Temperature by Depth')
<matplotlib.figure.Figure at 0x1a23f1beb8>
```



The Essay, Reinvented

Mirman School

“The more sophisticated science becomes, the harder it is to communicate results. Papers today are longer than ever and full of jargon and symbols. They depend on chains of computer programs that generate data, and clean up data, and plot data, and run statistical models on data. These programs tend to be both so sloppily written and so central to the results that it’s contributed to a replication crisis, or put another way, a failure of the paper to perform its most basic task: to report what you’ve actually discovered, clearly enough that someone else can discover it for themselves.”

James Somers, “The Scientific Paper is Obsolete”. The Atlantic, 4/5/2018

DCB

Mirman School

- 7th grade STEM class
- Intended to combine computer science with physical design/building projects
- This year's theme: The Ocean
- Integrated with 7th grade science

Model → Simulation → Analysis

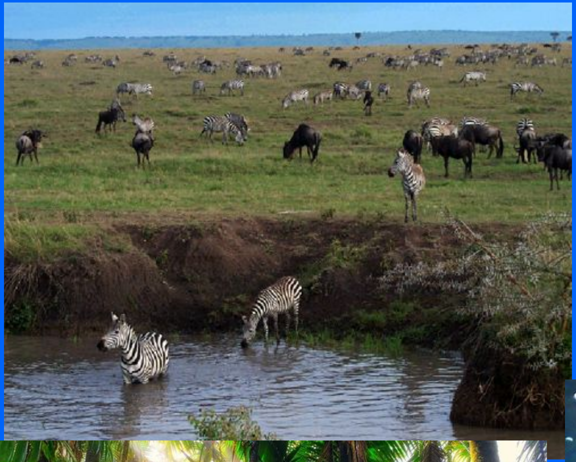
Mirman School

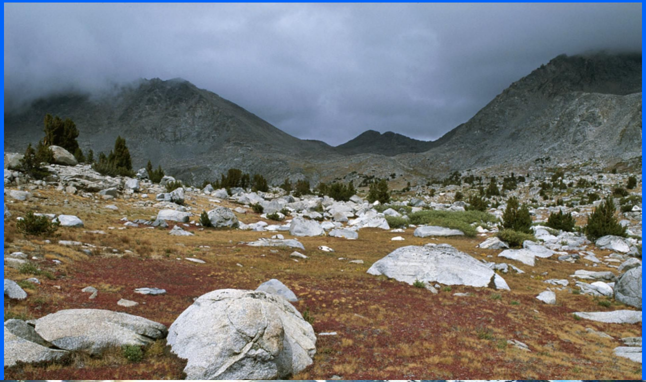
1) Choose Variables
2) Assign values to variables (research)
3) Run your model (simulation)
4) Analyze the results → draw conclusions

Ecosystems

Mirman School

- Biotic and Abiotic Factors
- Species Interactions
- Food Webs
- Biodiversity



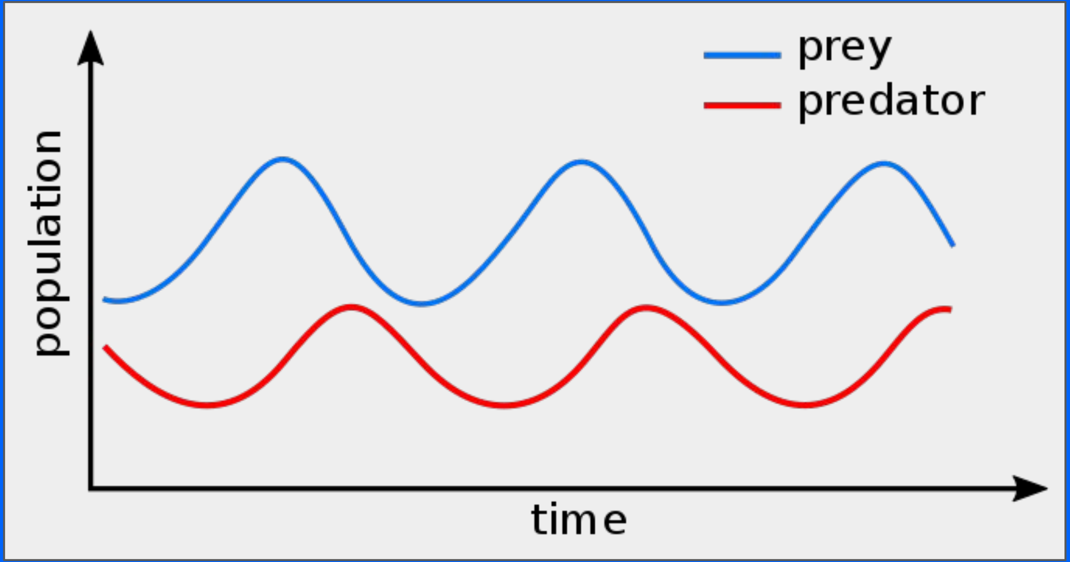


Predator-Prey Relationships

Mirman School

- Population level hard to measure/conceptualize
- Lotka-Volterra model
  - Defines four variables
    - Prey Growth Rate
    - Predation Rate
    - Predator Growth Rate
    - Predator Death Rate

$$\frac{dx}{dt} = \alpha x - \beta x y$$
$$\frac{dy}{dt} = \delta x y - \gamma y$$



Mirman School

```python
def lotka_volterra(state,t):
  prey_pop = state[0] # DON'T TOUCH
  pred_pop = state[1] # DON'T TOUCH
  prey_growth_rate = 0.1 # Change
  predation_rate =  0.1 # Change
  pred_growth_rate = 0.1 # Change
  pred_death_rate = 0.1 # Change
  prey_change = prey_pop*(prey_growth_rate - predation_rate*pred_pop) # DON'T TOUCH
  pred_change = -pred_pop*(pred_death_rate - pred_growth_rate*prey_pop) # DON'T TOUCH
  return [prey_change,pred_change] # DON'T TOUCH

t = arange(0,500,1) # Change the middle number for longer/shorter simulations
state0 = [0.5,0.5] # Change if you want to alter the starting situation
state = odeint(lotka_volterra,state0,t) # DON'T TOUCH
pl.figure(1)
pl.plot(t,state)
pl.ylim([0,3]) # Change if you want the y axis to grow/shrink
pl.xlabel('Time')
pl.ylabel('Population Size')
pl.legend(('Prey','Predator'))
pl.title('Lotka-Volterra equations')
pl.show(1)
```

Lotka-Volterra in Python

Lesson Design - Model

Mirman School

- Scaffold coding skills needed → Jupyter notebook
    - Introduce assigning variables, running basic models, and graphing results
- Introduce Lotka-Volterra model
- Student choice in predator and prey to model within ecosystem
- Students use research to justify a value for variables in model

Predator Death rate (value) - 0.6

Justification: *I noticed that the jaguar population was decreasing extremely quickly, mainly due to habitat loss. Thousands of jaguars were dying each year, and over 50% were dying over time, so I rounded and got 60%.*

Lesson Design - Simulation

Mirman School

- After defining initial parameters, students chose an environmental factor that could change one or more of the variables

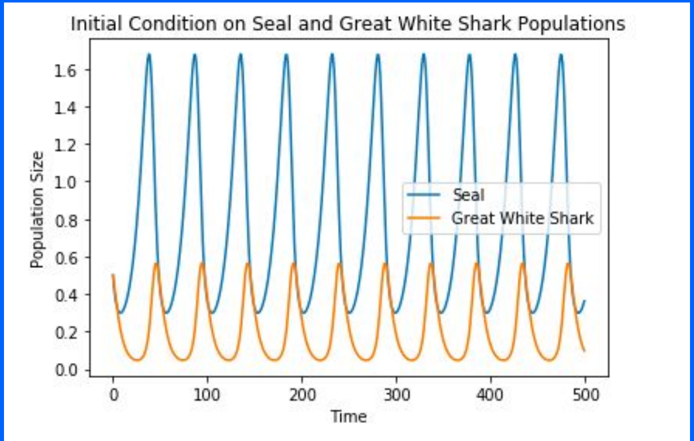
Abiotic Factor (name or description) - *Disease / Leishmania tropica*
I chose this because this disease is carried by 40 percent of all Jackals. Also, this disease is very fatal for canines without treatment which will raise the death rate of the Jackals a lot.

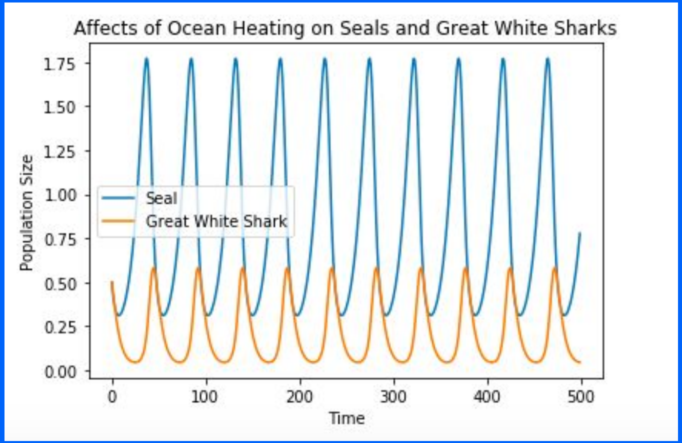
- Students formulate their research question and hypothesis
- Students run both simulations (initial and alternative) and graph results

Lesson Design - Analysis

Mirman School

- Students analyze model outputs (graphs) to answer their research question





I changed my Predator Death rate by 0.01 because of the effects of Ocean Warming. The Ocean Warming will kill the sharks because all marine animals are used to a very certain temperature. When the temperature changes too much, the animals will die. This did not change the population of the Great White substantially, because the temperature has not changed enough to do any real damage yet, but the seal population grew almost 10%! This is because there are less great white sharks to eat them. From these models, you can see that when a predator population death rate goes up, the prey's population will grow.

Next Steps

Mirman School

- This year: Use coding skills to answer other data questions
  - Collecting Primary Datasets
  - Practice on NOAA Datasets
  - Collating and data analysis in Python
  - Generating graphs to visualize data
  - Write a report on findings

- Next year: Repeat activity, with earlier and further emphasis on the formulation of research questions and how coding and models enable us to answer these questions
  - Intro to Python in science itself

Mirman School

DCB Students working on ROVs for data collection

Thanks!

Mirman School

- mtaggart@mirman.org
- github.com/mirman-school
- github.com/mttaggart
- @mttaggart