THE DATA SCIENCE LABS:

Hands-On Mathematics Education

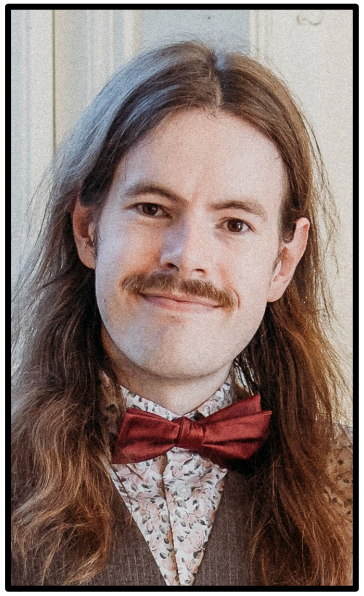
KAITLYN HOOD, PURDUE UNIVERSITY

JOINT MATH MEETINGS 2024

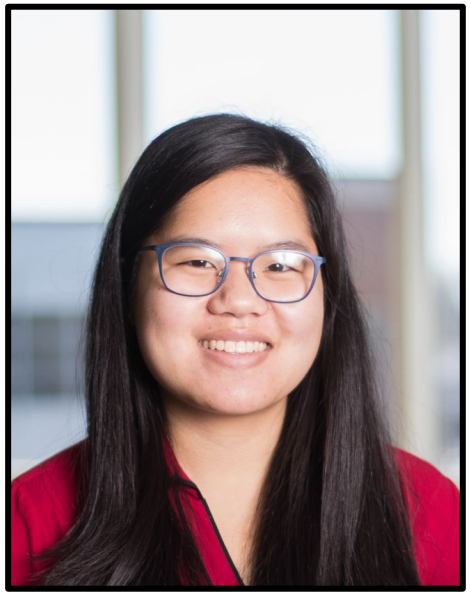
SAN FRANCISCO, CA

TEACHING DATA SCIENCE THROUGH MATH LABORATORIES

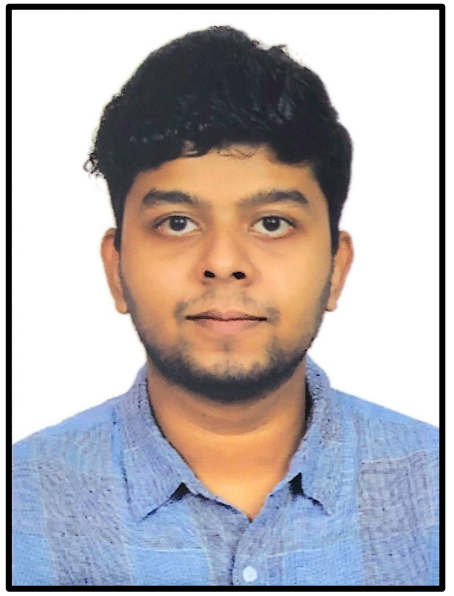
Mimi Boutin



Alden Bradford



Kindyl King



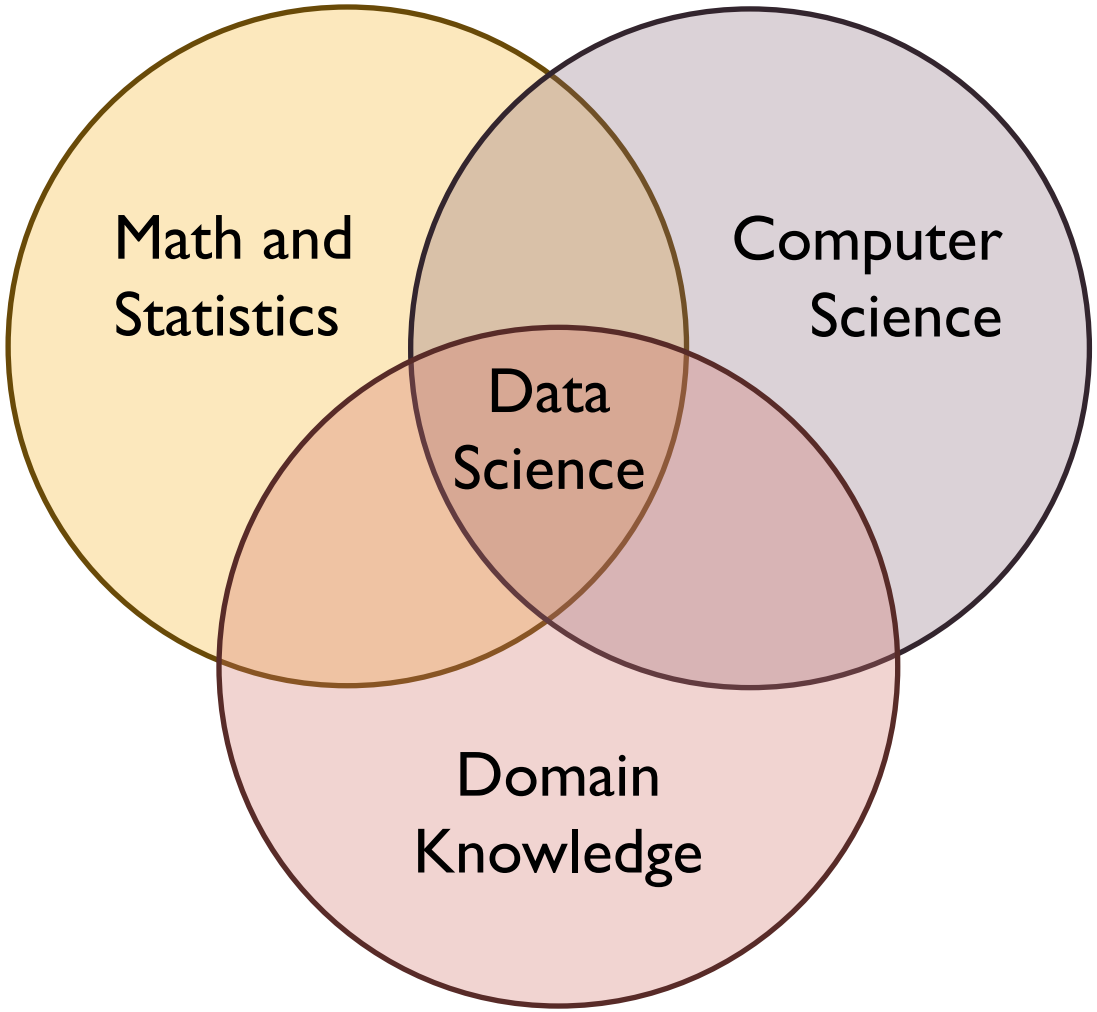
Adharsh Sabukumar

Department of Mathematics

Elmore Family School of Electrical and Computer Engineering

WHAT IS DATA SCIENCE?

- "The study of extracting value from data" (Wing, 2019)
- The term *data science* was driven by industry:
  - "Postgraduate degrees in … Statistics and Computer Science, did not guarantee the expertise needed to successfully complete these projects." (Irizarry, 2020)

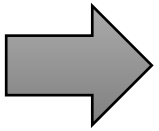


(Danyluk & Leidig, 2021)

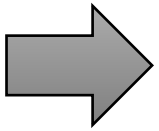
WHAT IS DATA SCIENCE?

- Data science "is not even a single discipline by itself" (Meng 2019)
  - increasing consensus that the term data scientist is only useful as an umbrella term

Backend Track:
"the part that deals with hardware, efficient computing, and data storage infrastructure, or what is often referred to as data engineering." (Irizarry, 2020)



Frontend track:
"The data analysts wrangle, explore, quality assess, fit models to data, perform statistical inference, and develop protypes." (Irizarry, 2020)



Machine Learning Track:
"build and assess prediction algorithms and make the solution scalable and robust for many users." (Irizarry, 2020)

HOW DO WE TRAIN NEW DATA SCIENTISTS?

- *"Bring applications to the forefront."* (Hicks & Irizarry, 2018)
- *Real-world experience.* (National Academies of Sciences, Engineering, and Medicine, 2018)
- *Practical programming skills.* (Irizarry, 2020)

DATA SCIENCE LABS: CONCEPT

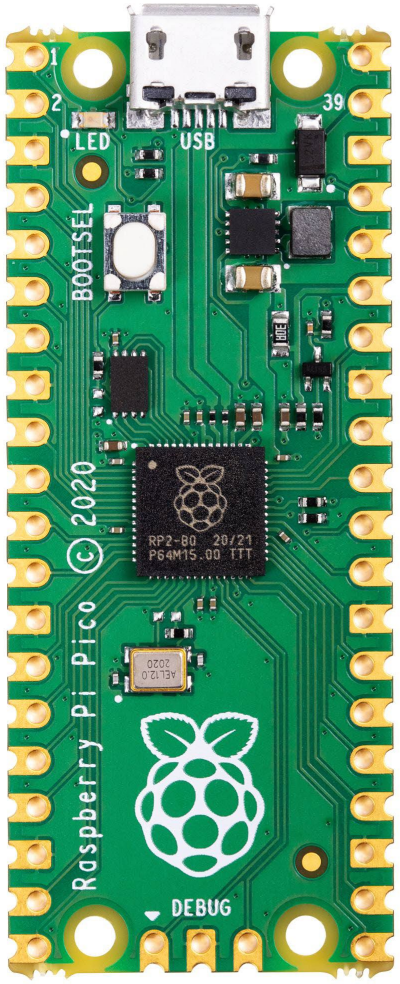
Hardware

Introduce students to hardware:

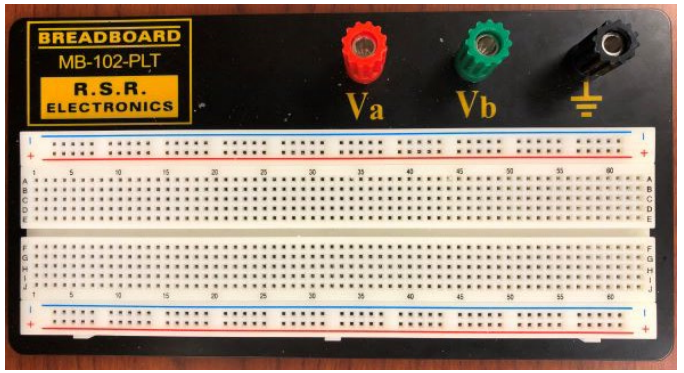
- Arduino hardware
- Raspberry Pi Pico: a low-cost, high-performance microcontroller board

Prepare students for Purdue Semiconductors Program

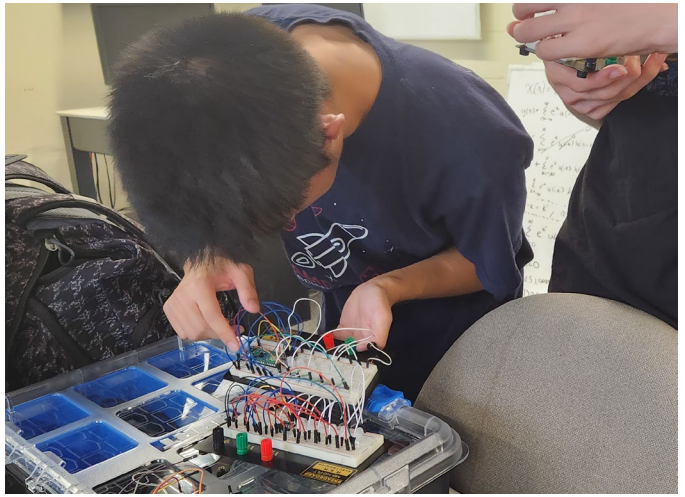
Train workforce funded by the CHIPS and Science Act (2022)



Raspberry Pi Pico



Breadboard



Student building a circuit

DATA SCIENCE LABS: CONCEPT

Hardware

Introduce students to hardware:

- Arduino
- Raspberry Pi Pico: a low-cost, high-performance microcontroller board

Prepare students for Purdue Semiconductors Program

Train workforce funded by the CHIPS and Science Act (2022)

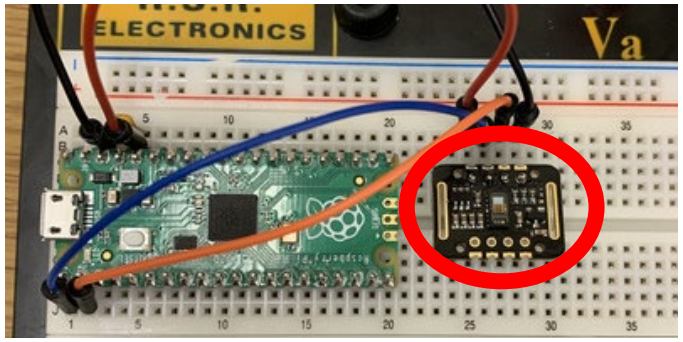
Data Collection

Students collect data using sensors:

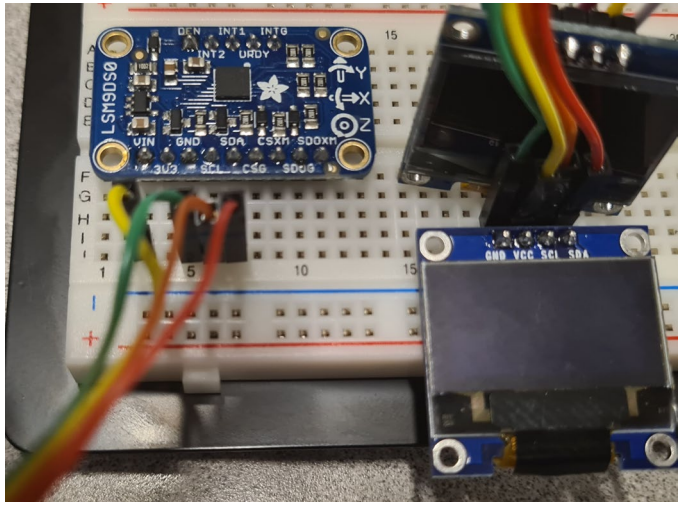
- PPG sensor
- Inclinometer
- Ultrasonic sensor

Students practice working with real data sets

Students use Python in Jupyter to store and analyze data



Photoplethysmography (PPG) sensor



Inclinometer

DATA SCIENCE LABS: CONCEPT

Hardware

Introduce students to hardware:

- Arduino
- Raspberry Pi Pico: a low-cost, high-performance microcontroller board

Prepare students for Purdue Semiconductors Program

Train workforce funded by the CHIPS and Science Act (2022)

Data Collection

Students collect data using sensors:

- PPG sensor
- Inclinometer
- Ultrasonic sensor

Students practice working with real data sets

Students use Python in Jupyter to store and analyze data

Mathematical Analysis

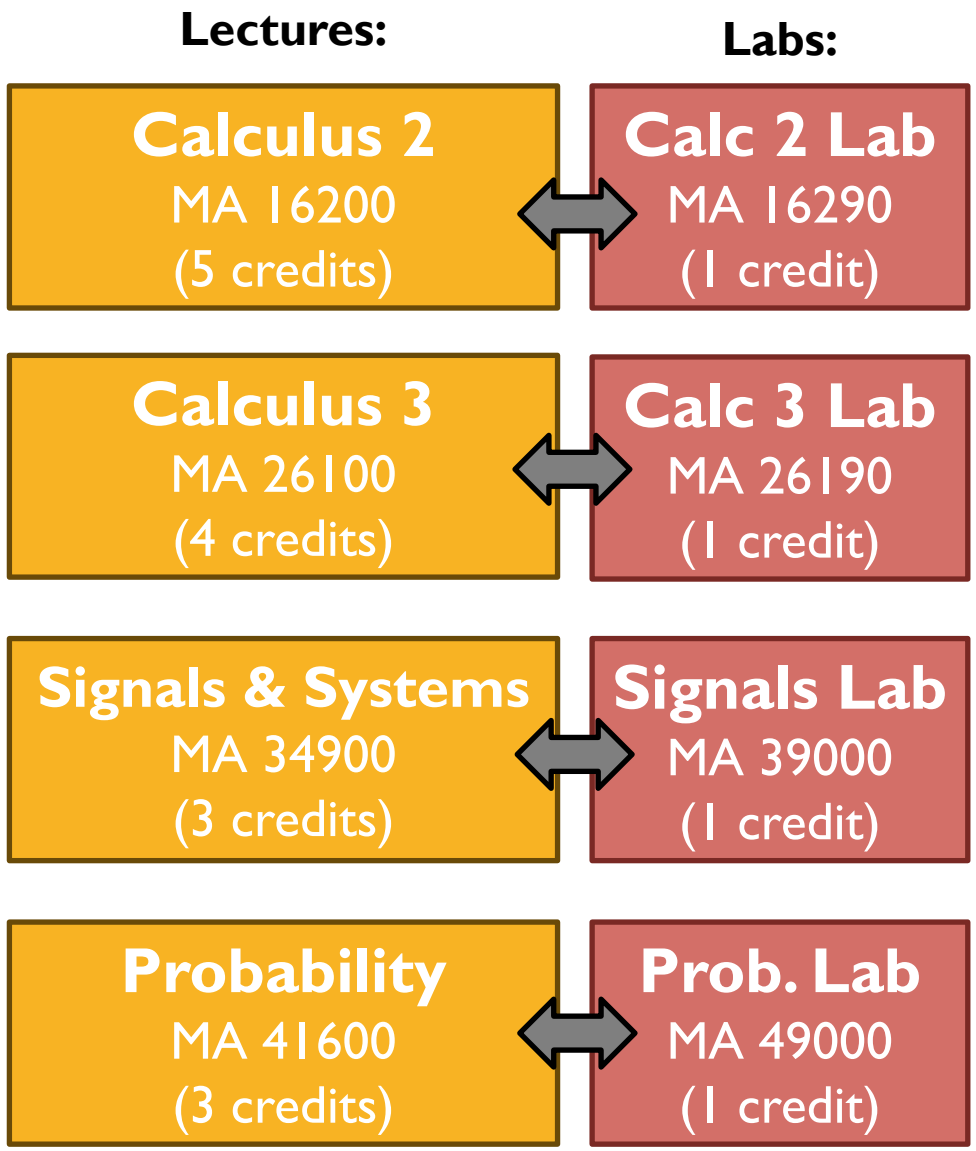
Students use math to analyze the data they collected

- Calculus
- Fourier Analysis
- Probability

Mathematical concepts applied to real applications

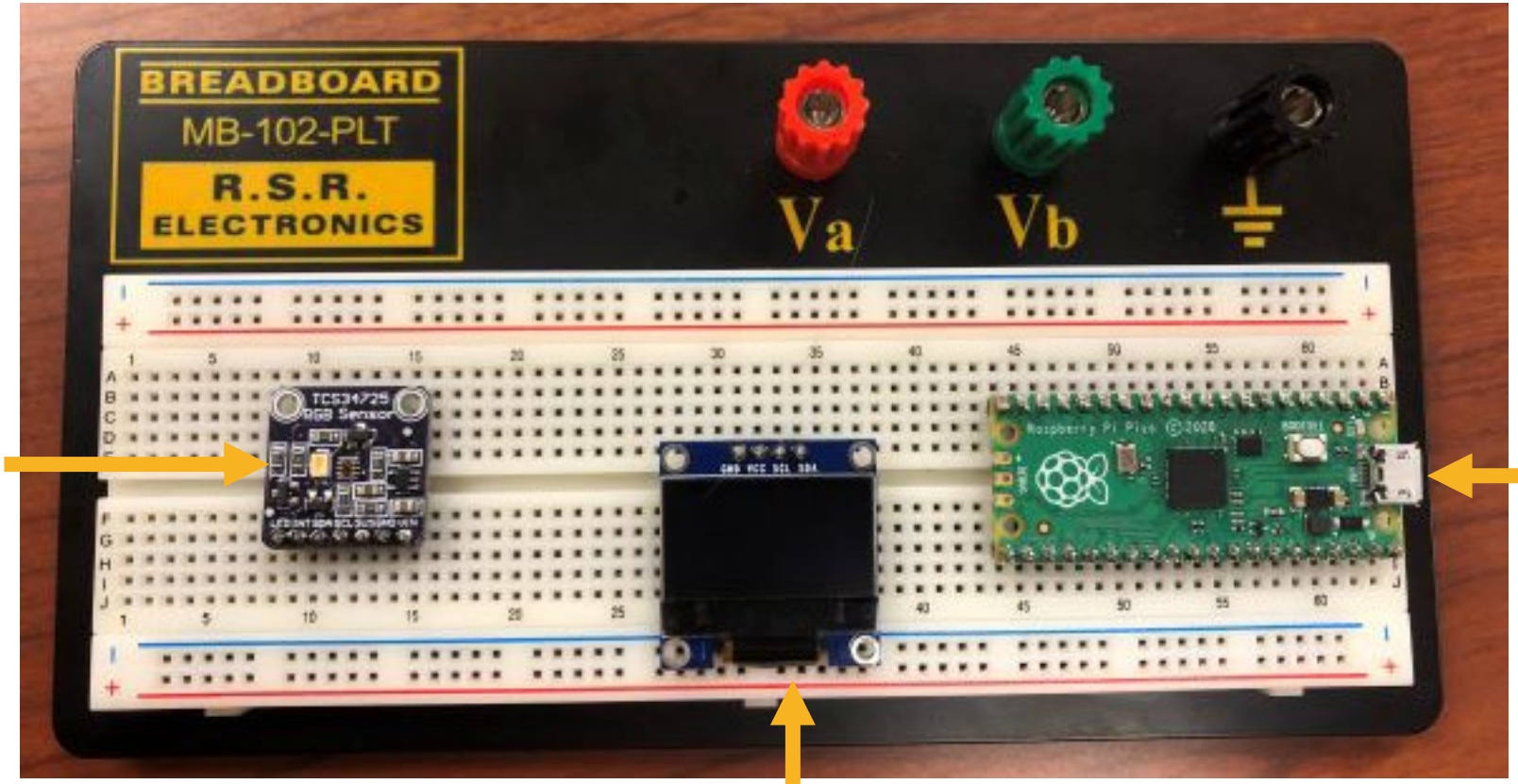
Students learn to code "in context"

DATA SCIENCE LAB CURRICULUM

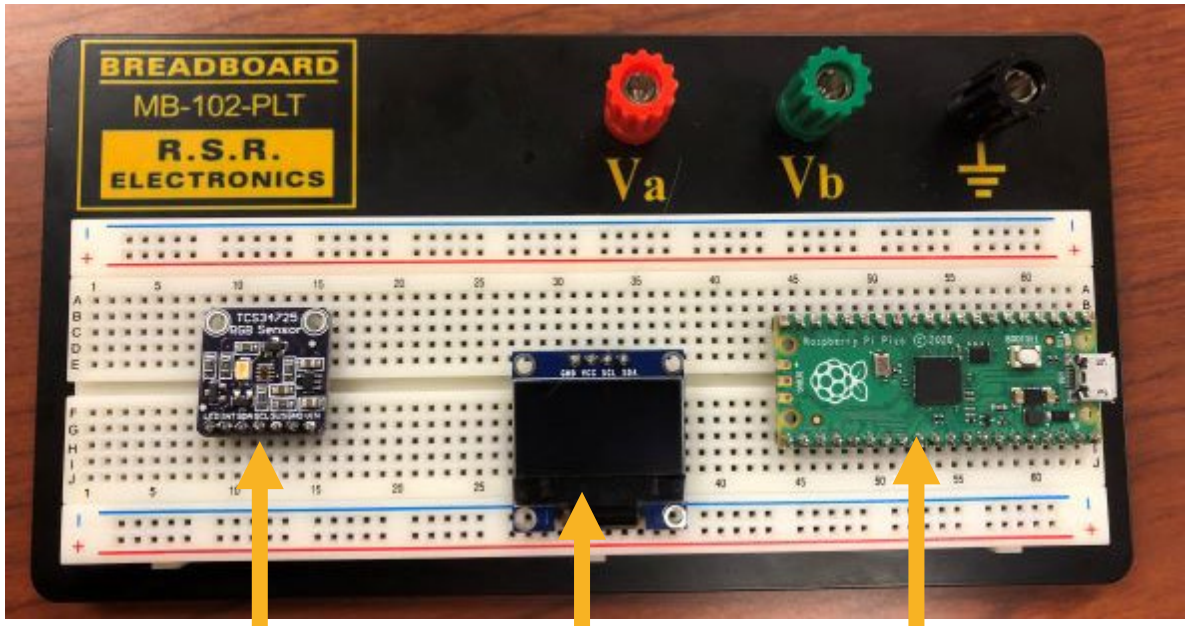


- 1 credit = 150 minutes in a computer lab once a week
- Associated to an existing "companion course" in math curriculum
- At Purdue, taking the lab upgrades the course to "honors" designation
- For example, …

EX1: DETECT COLOR OF AN M&M (CALC 3 LAB)



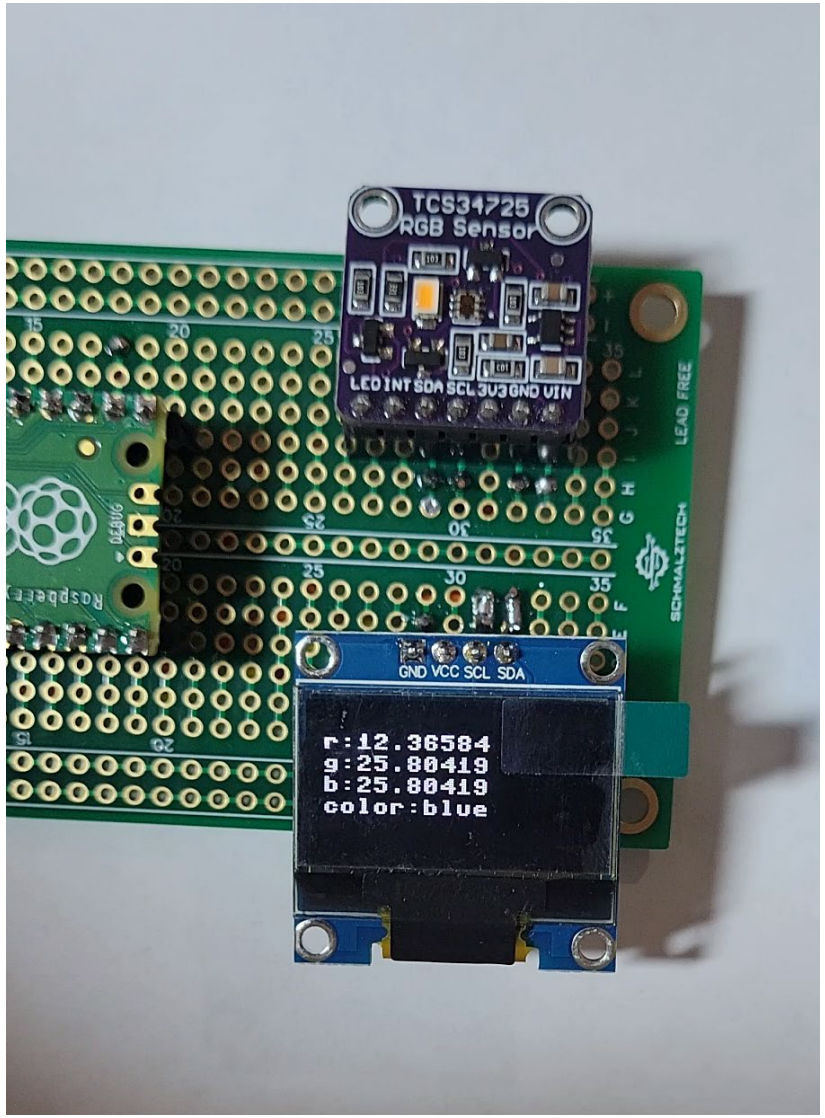
EX1: DETECT COLOR OF AN M&M (CALC 3 LAB)



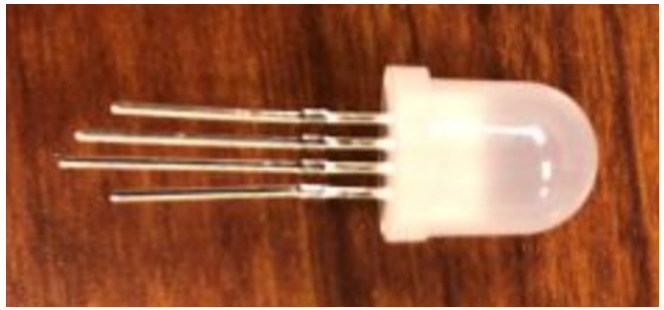
RGB Sensor

OLED Display

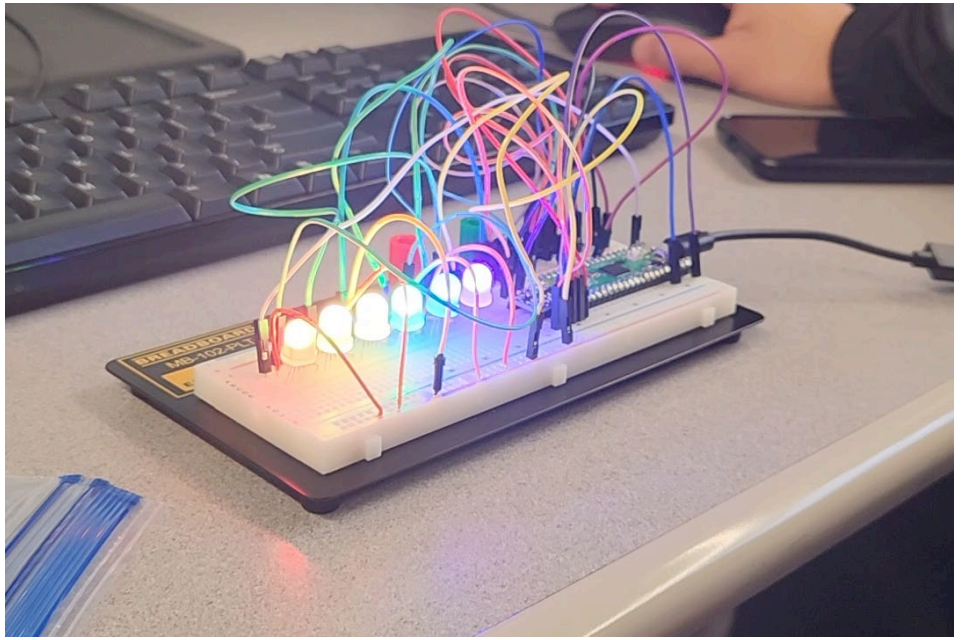
Raspberry Pi Pico

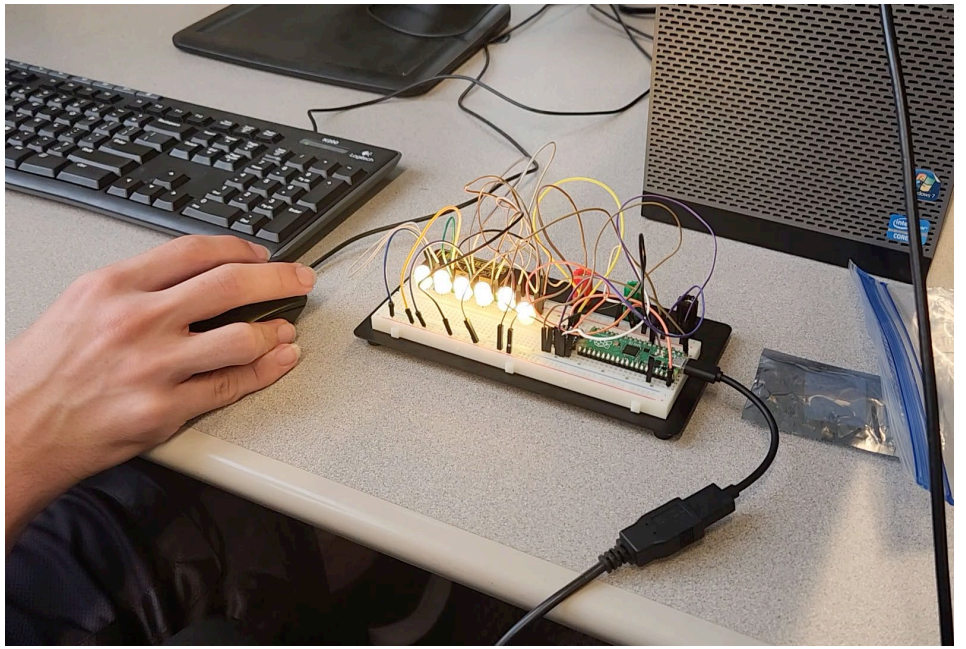


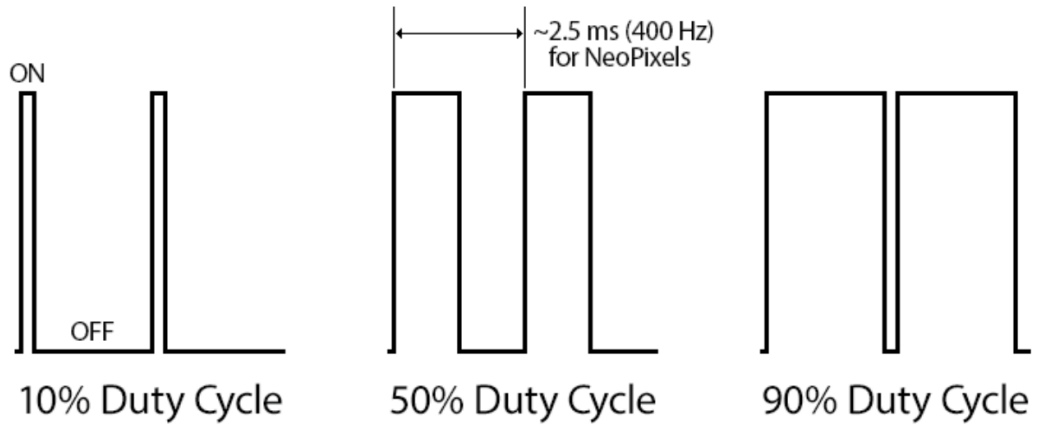
EX2: COLOR LED (CALC 3 LAB)



NeoPixel: individually addressable LED



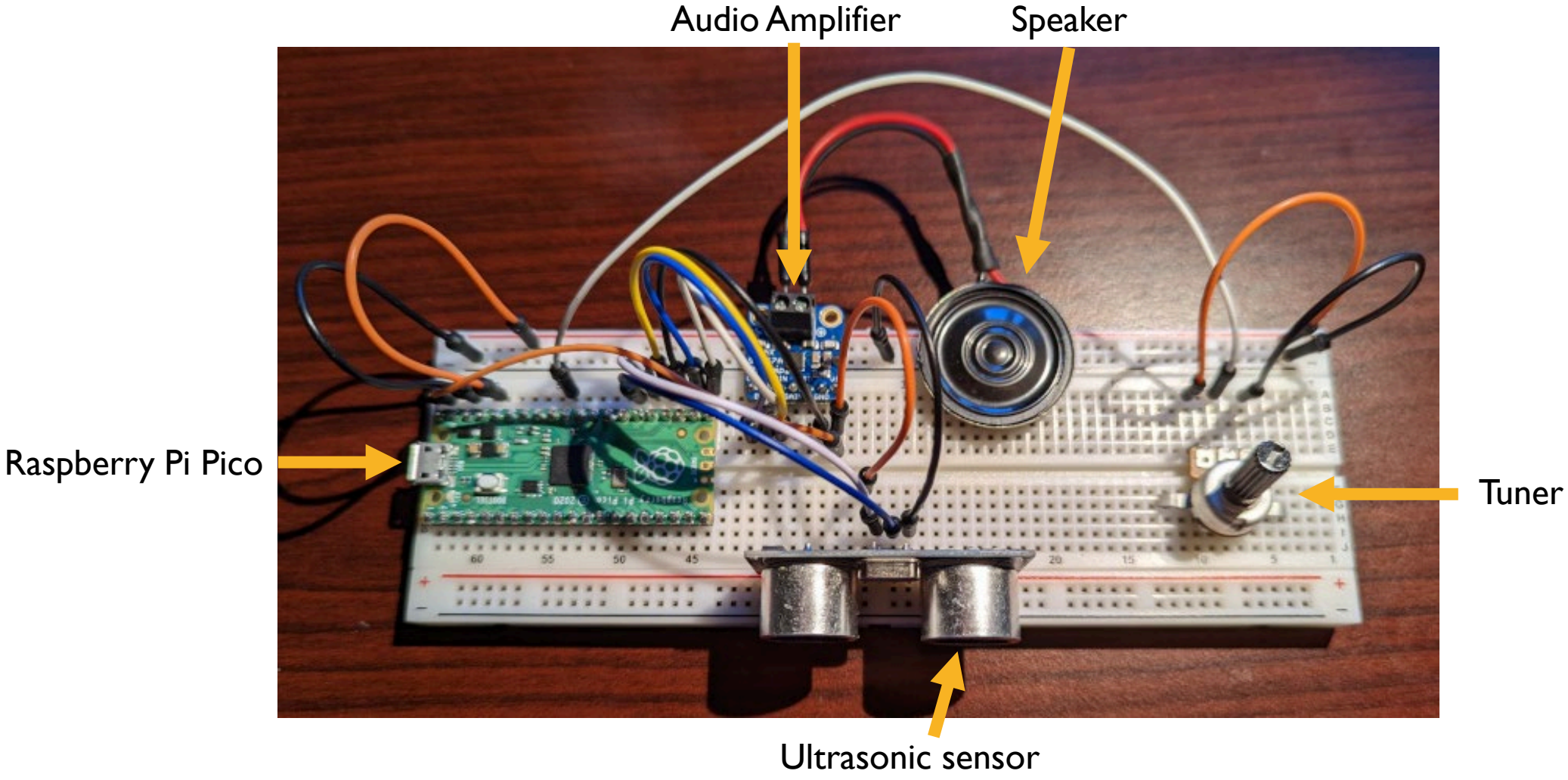




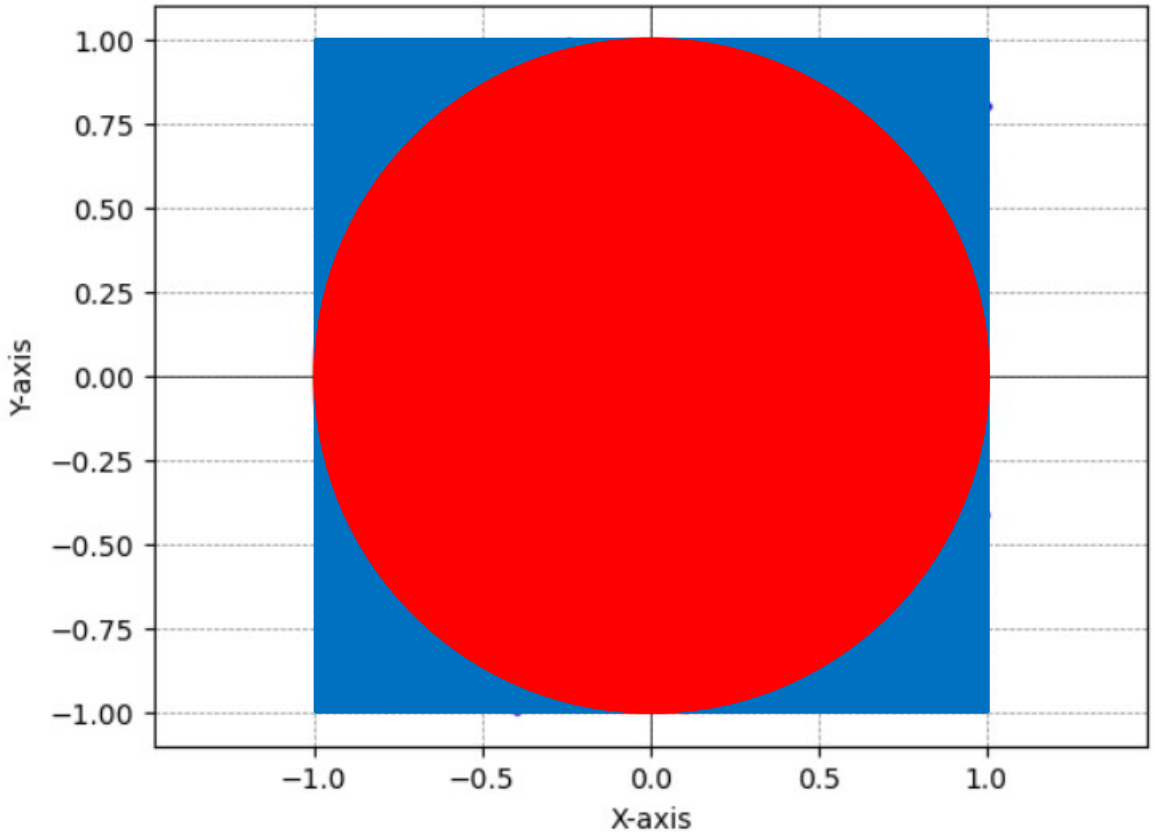
EX3: BUILD A THEREMIN (SIGNALS LAB)

- Electronic musical instrument controlled without physical contact by the performer
- Named after Leon Theramin, who patented the device in 1928
- Alexandra Stepanoff playing the theremin on NBC Radio, 1930

EX3: BUILD A THEREMIN (SIGNALS LAB)



EX4: USE A RNG TO ESTIMATE PI (PROBABILITY LAB)



Area of blue square $[-1,1] \times [-1,1] = 4$

Area of red circle of radius $1 = \pi$

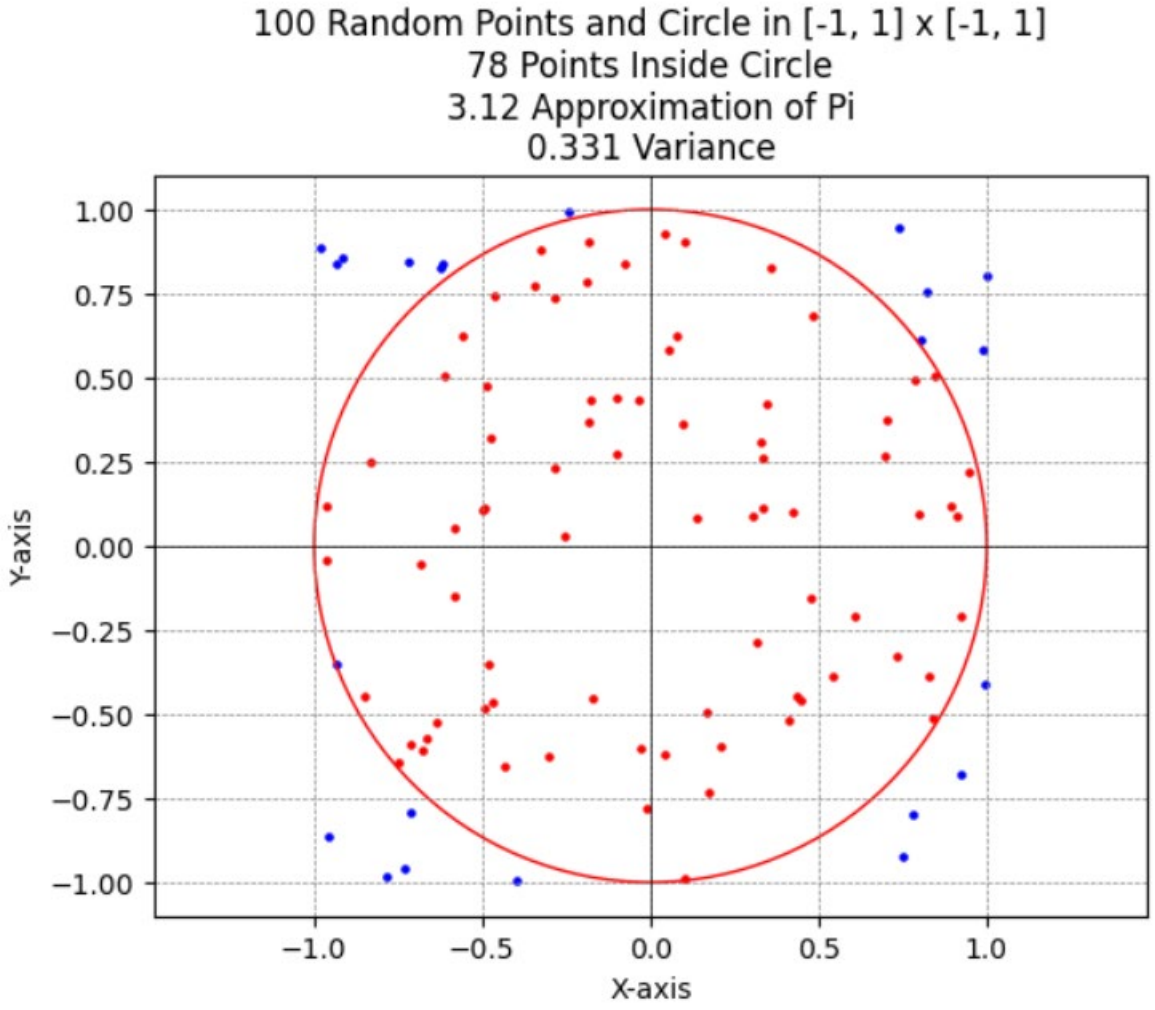
Monte Carlo:

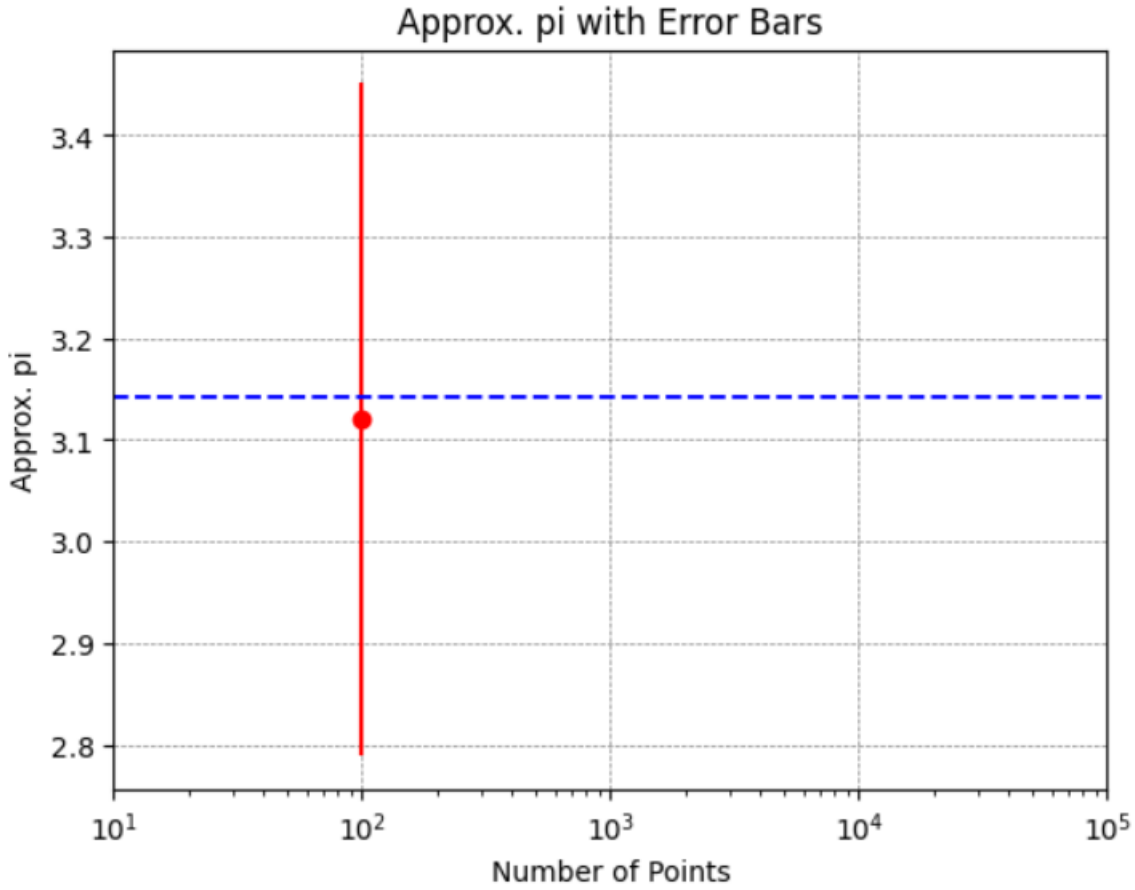
- Approximate area with a uniformly random distribution of points in the square

$$\frac{\#\text{ red points}}{\#\text{ blue points}} = \frac{\text{area of circle}}{\text{area of square}} = \frac{\pi}{4}$$

EX4: USE A RNG TO ESTIMATE PI (PROBABILITY LAB)

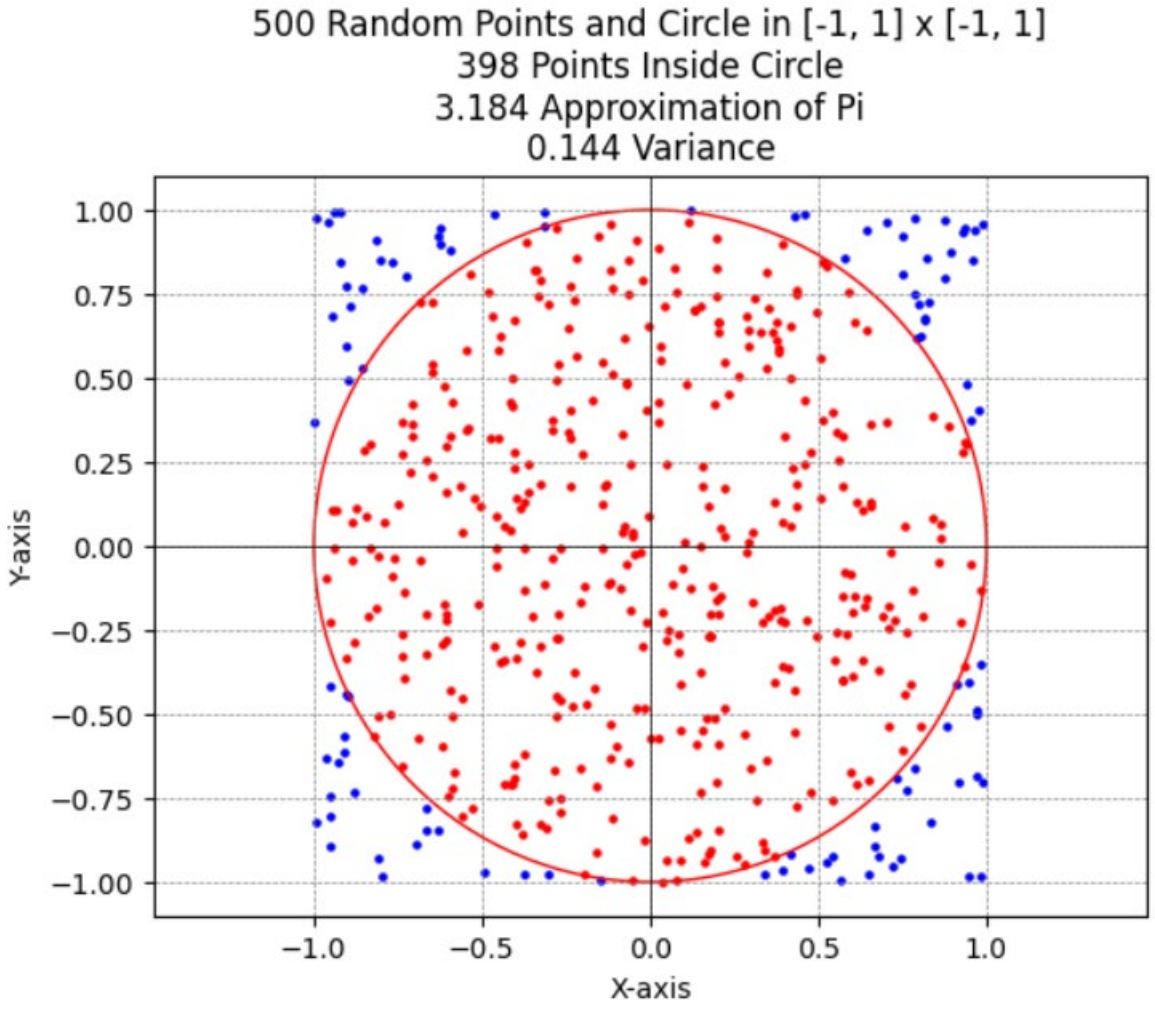
$$\frac{\#\text{ red points}}{\#\text{ blue points}} = \frac{\text{area of circle}}{\text{area of square}} = \frac{\pi}{4}$$

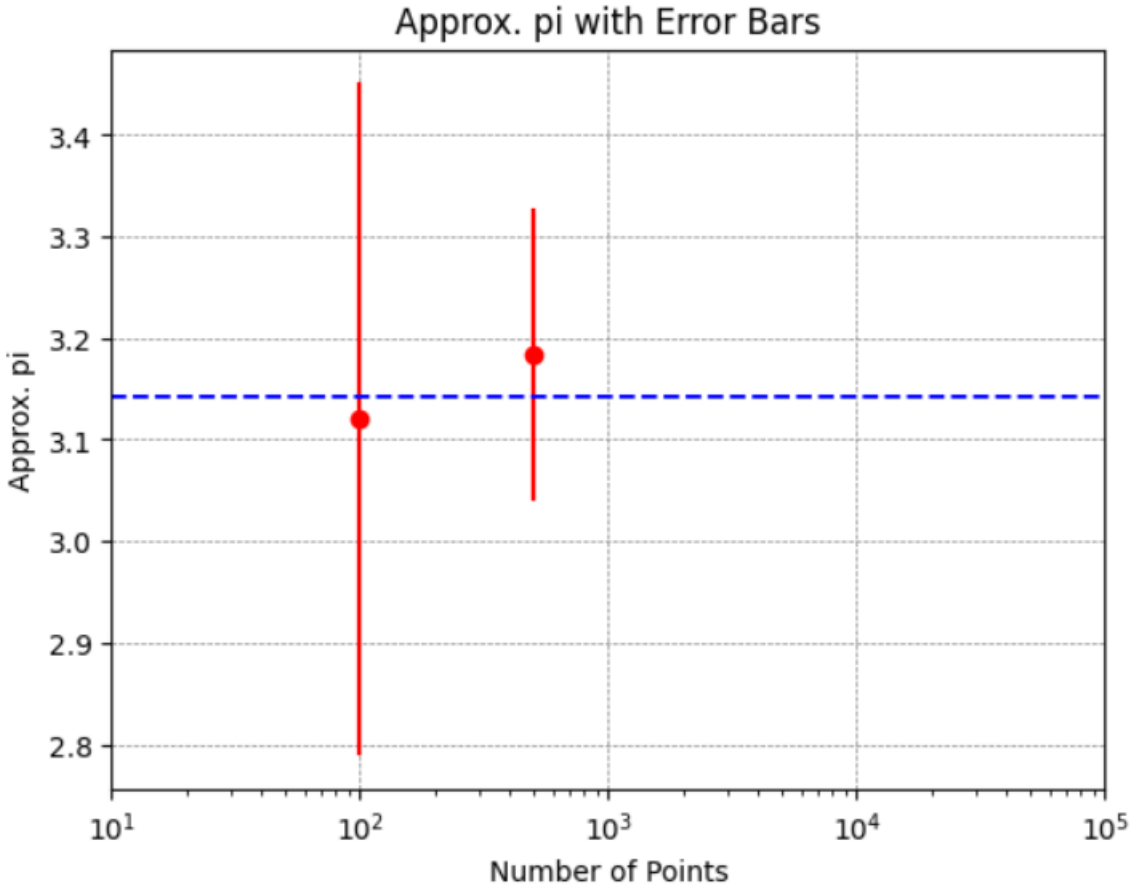




EX4: USE A RNG TO ESTIMATE PI (PROBABILITY LAB)

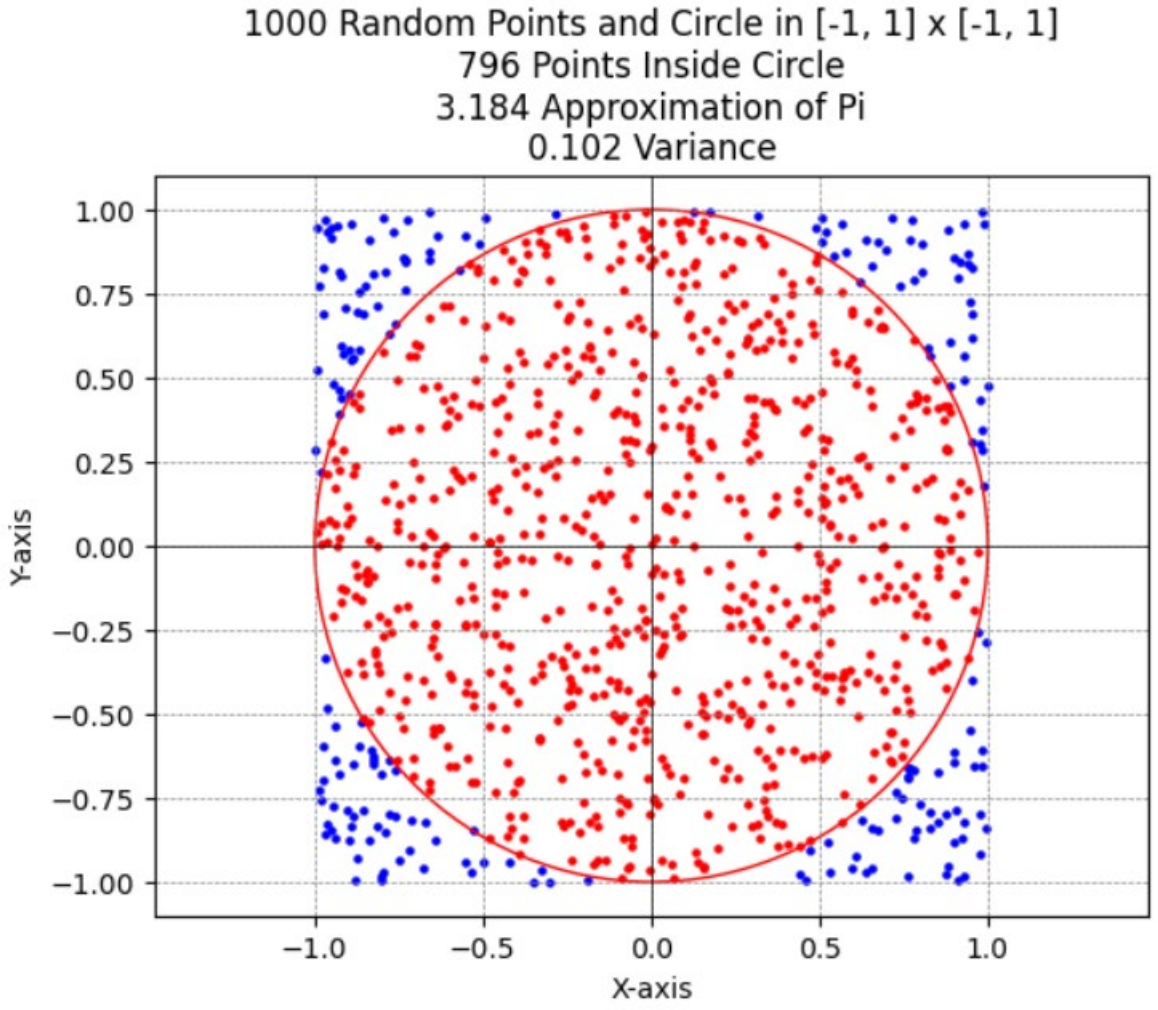
$$\frac{\#\text{ red points}}{\#\text{ blue points}} = \frac{\text{area of circle}}{\text{area of square}} = \frac{\pi}{4}$$

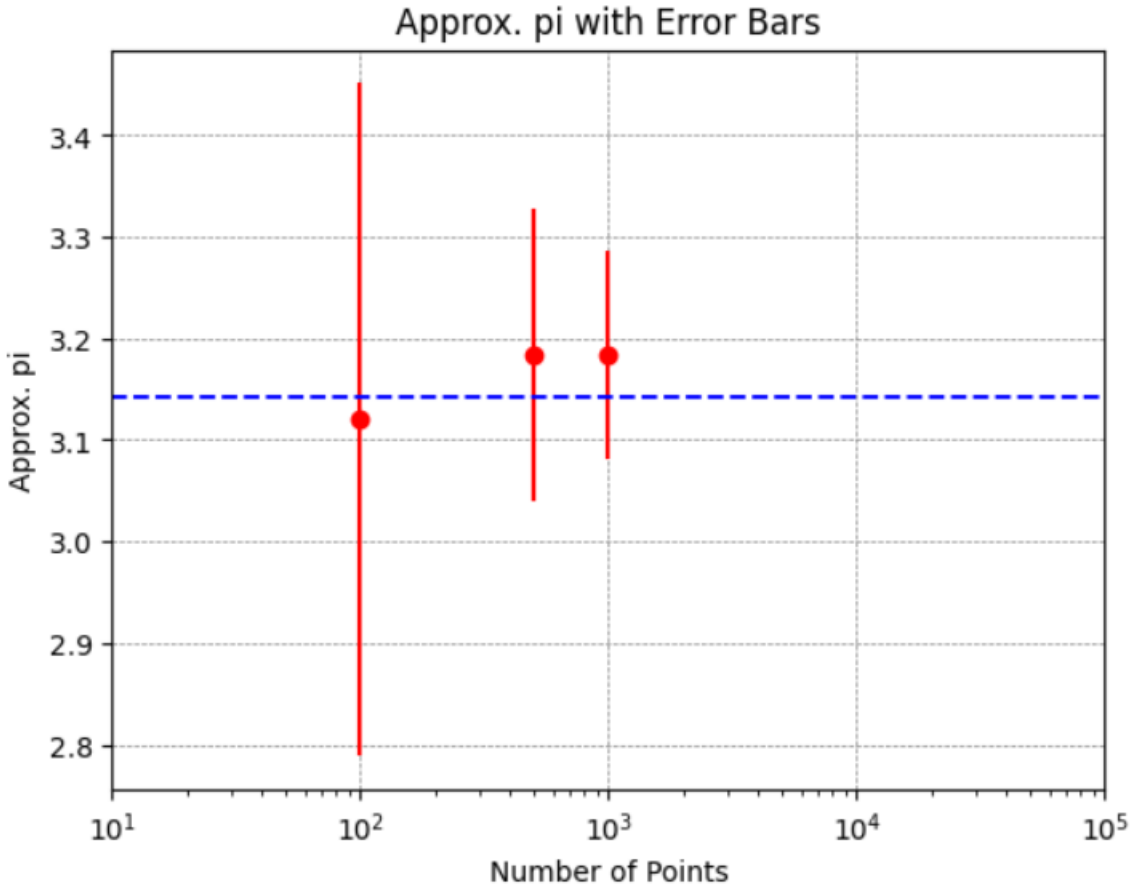




EX4: USE A RNG TO ESTIMATE PI (PROBABILITY LAB)

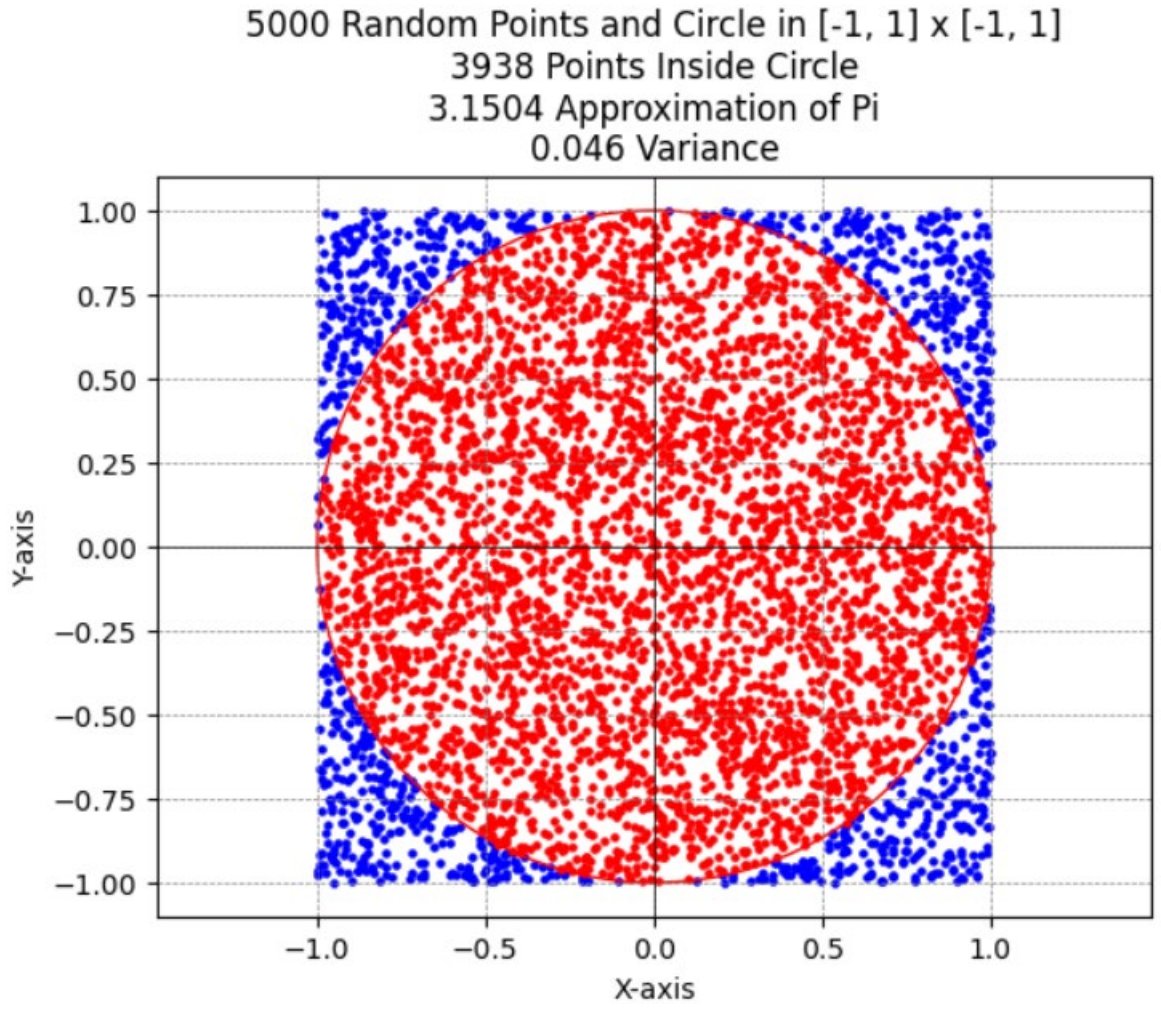
$$\frac{\#\text{ red points}}{\#\text{ blue points}} = \frac{\text{area of circle}}{\text{area of square}} = \frac{\pi}{4}$$

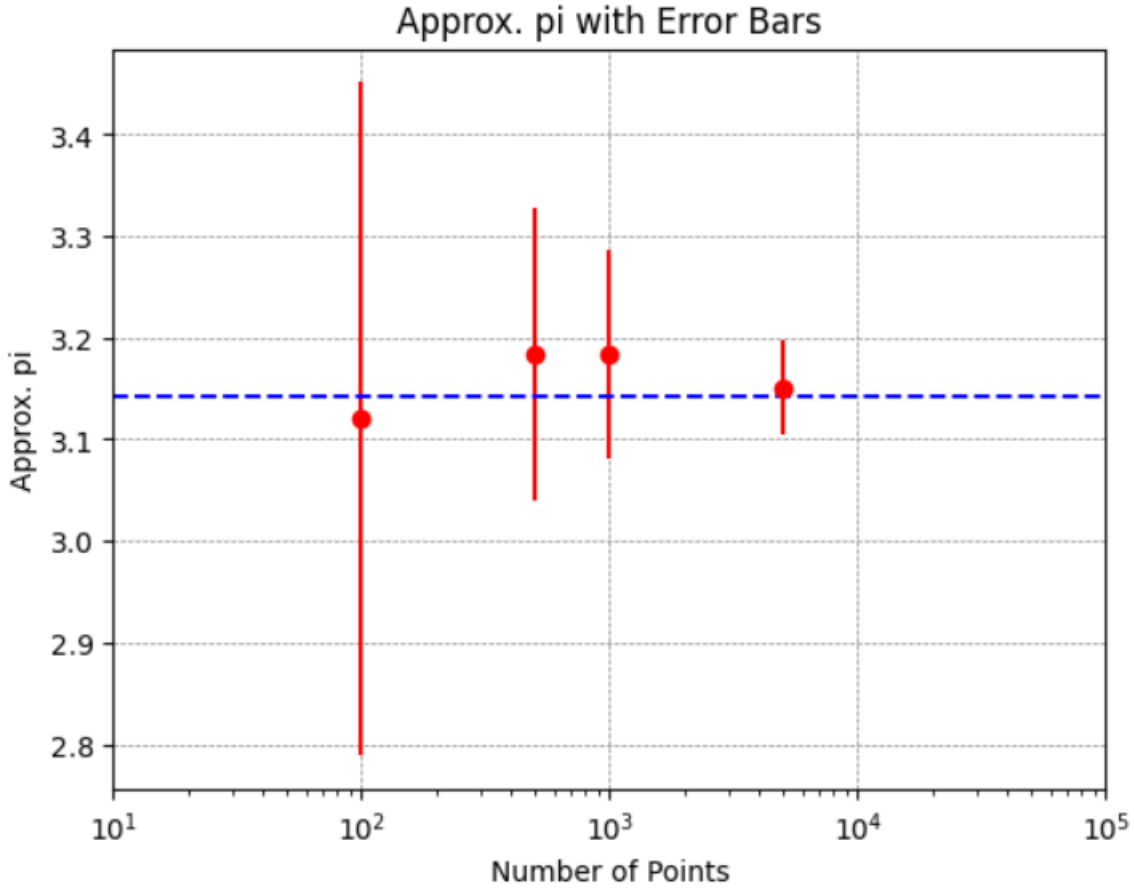




EX4: USE A RNG TO ESTIMATE PI (PROBABILITY LAB)

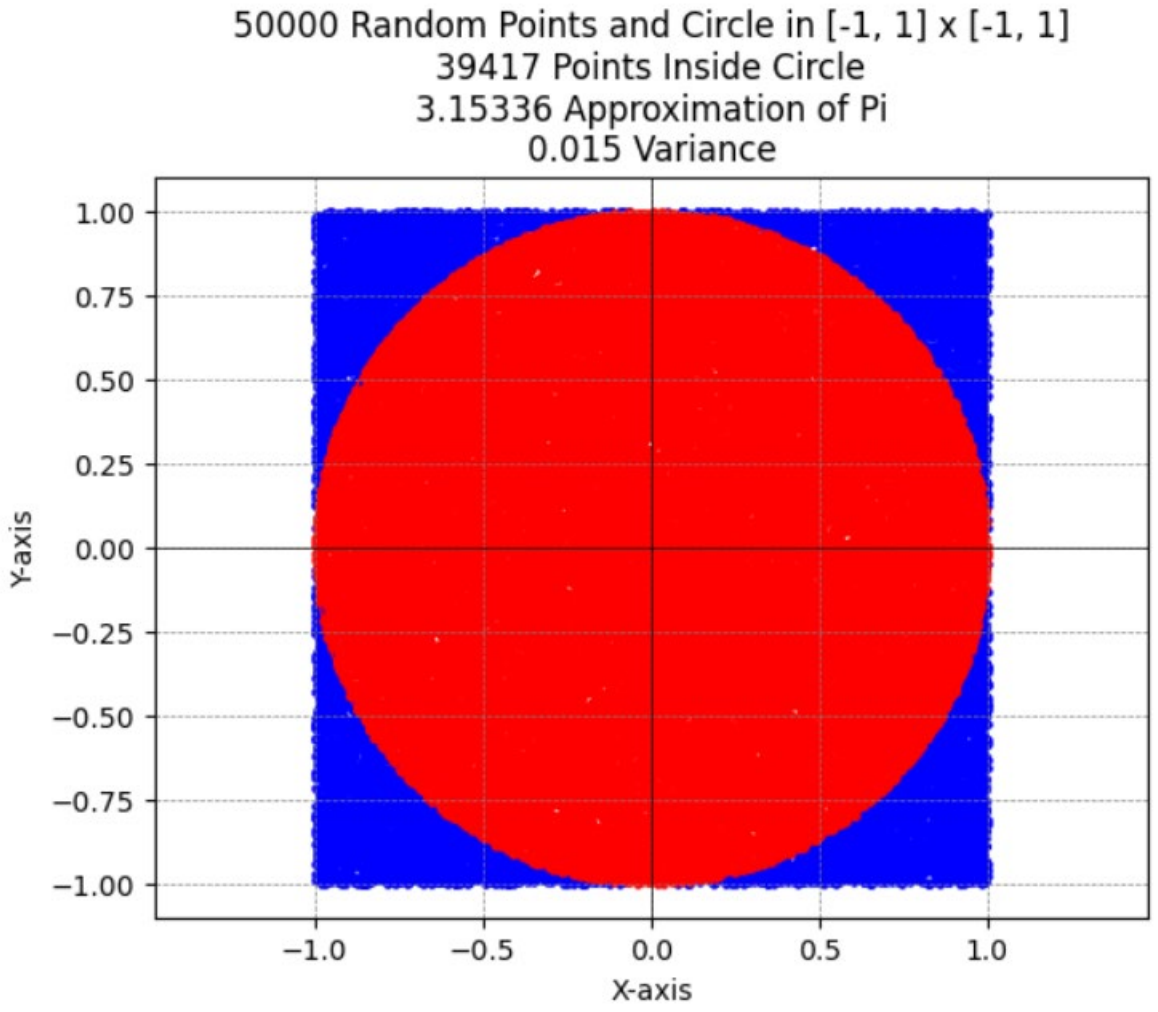
$$\frac{\#\text{ red points}}{\#\text{ blue points}} = \frac{\text{area of circle}}{\text{area of square}} = \frac{\pi}{4}$$

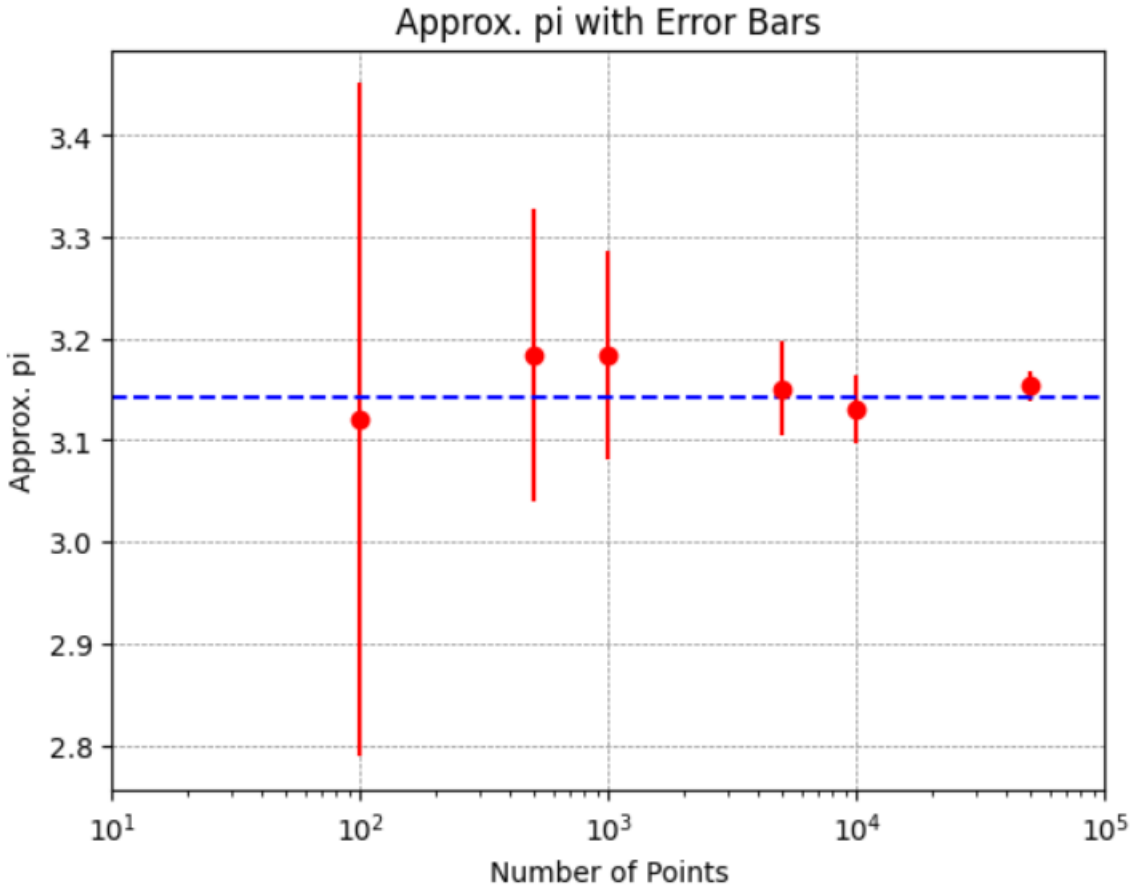




EX4: USE A RNG TO ESTIMATE PI (PROBABILITY LAB)

$$\frac{\#\text{ red points}}{\#\text{ blue points}} = \frac{\text{area of circle}}{\text{area of square}} = \frac{\pi}{4}$$





EX5: PLANIMETER TO FIND AREAS (CALC 3 LAB)

Green's theorem:

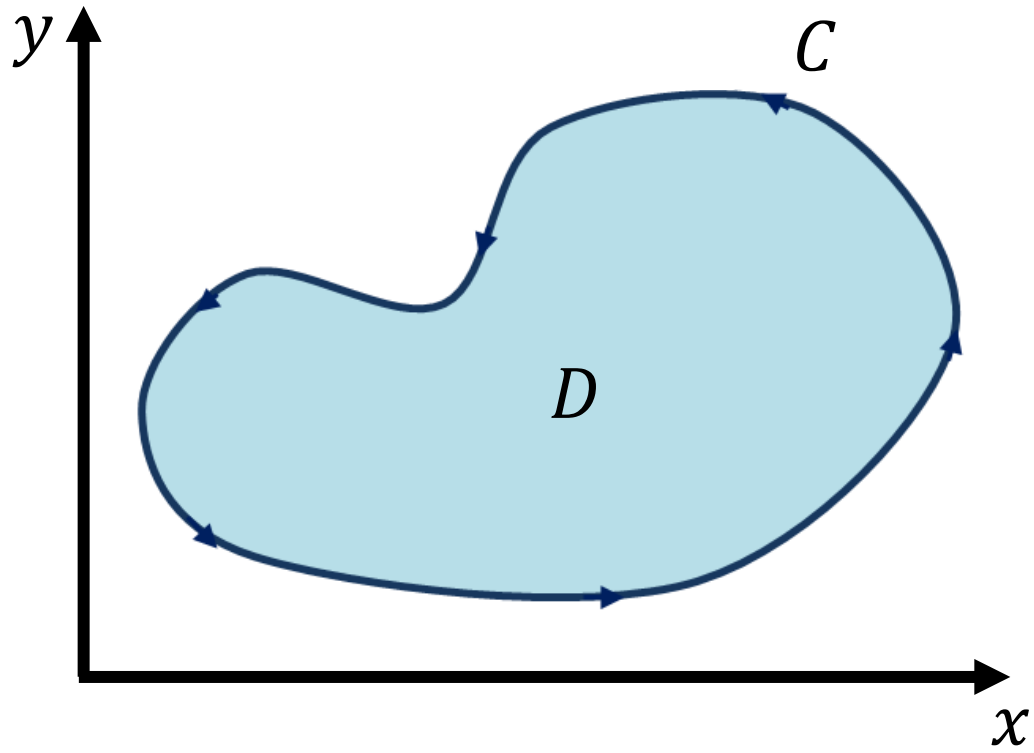
$$\oint_C P\,dx + Q\,dy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA$$

To find the area of region D:

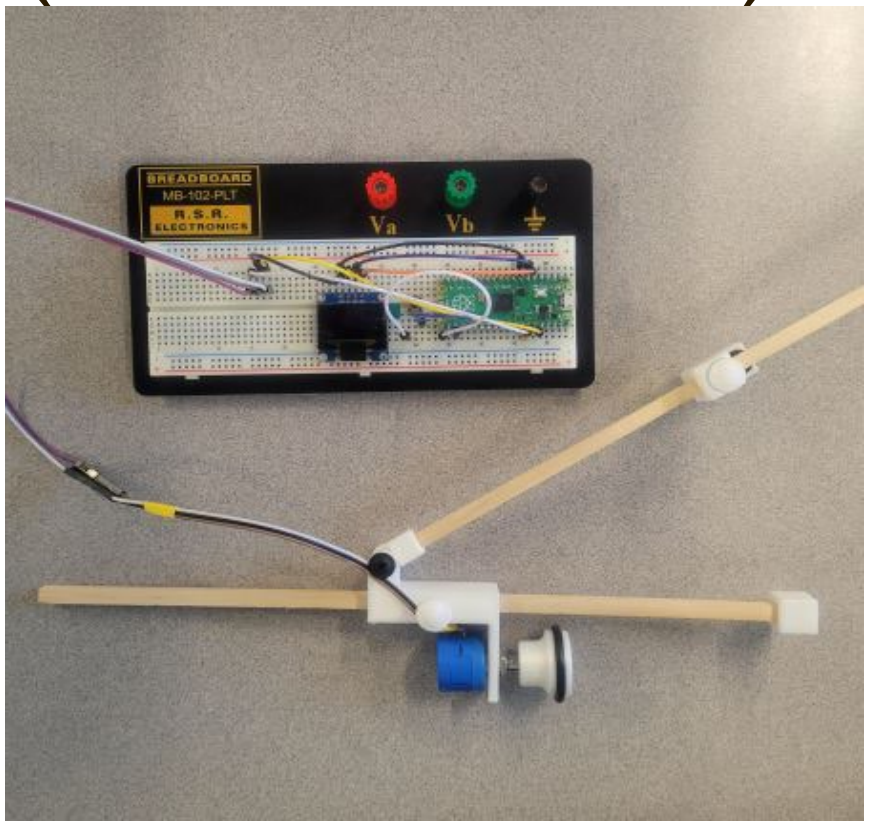
$$\text{area}(D) = \iint_D 1\,dA$$

If we choose $P = 0$ and $Q = x$ and apply Green's theorem, then:

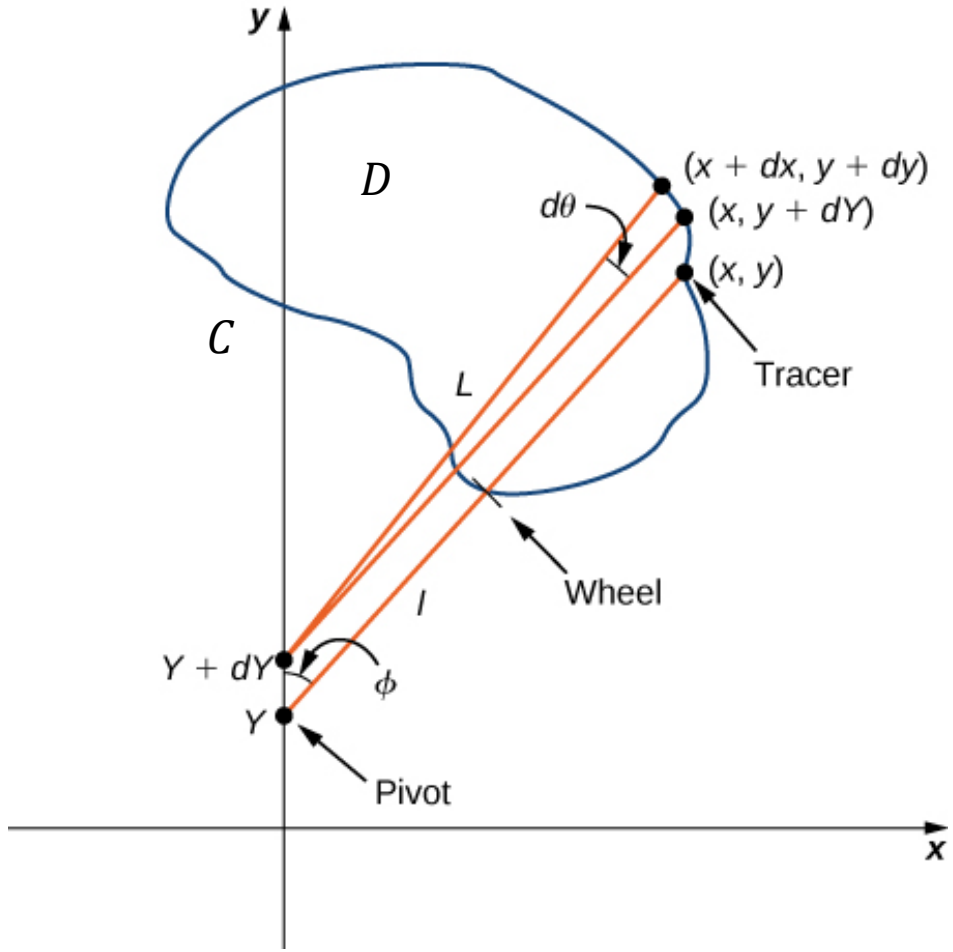
$$\text{area}(D) = \oint_C x\,dy$$



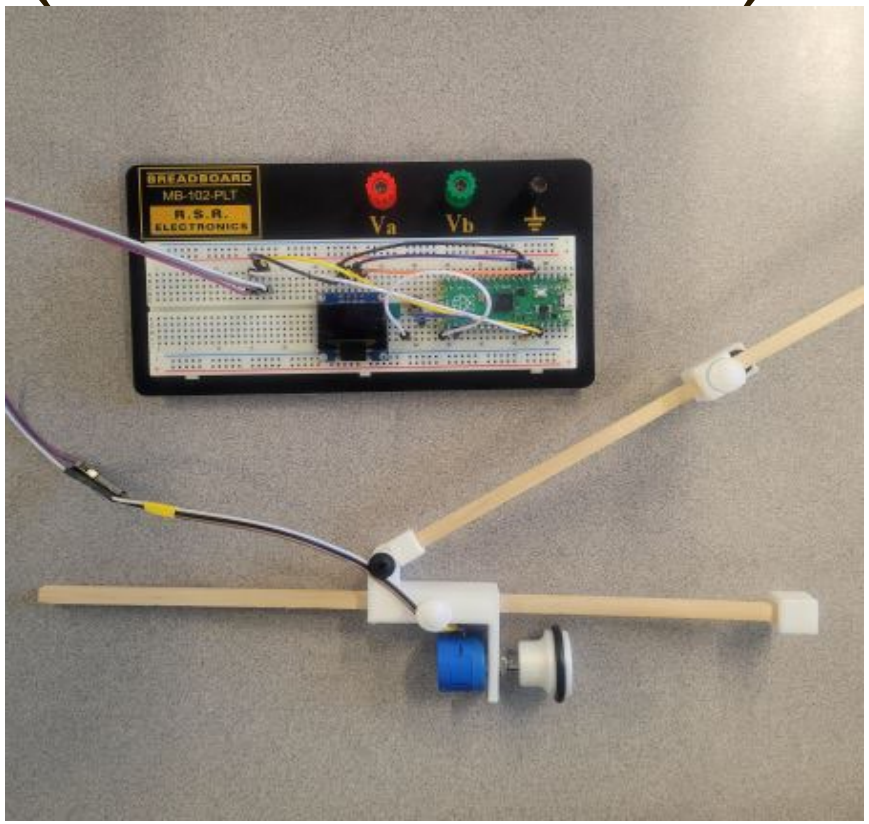
EX5: PLANIMETER TO FIND AREAS (CALC 3 LAB)



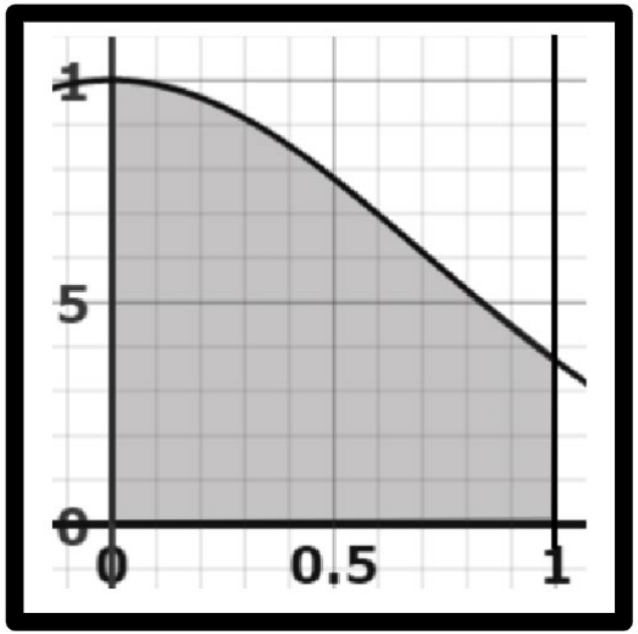
Planimeter to find the area of a 2D shape by using a potentiometer and Green's Theorem
(Calc 3 Lab)



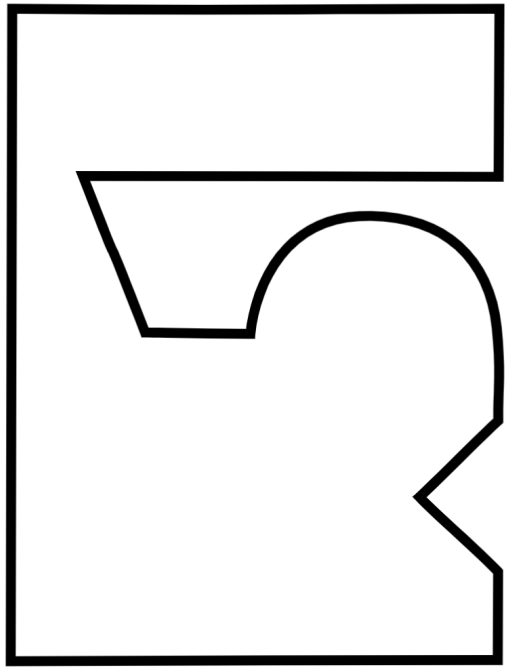
EX5: PLANIMETER TO FIND AREAS (CALC 3 LAB)



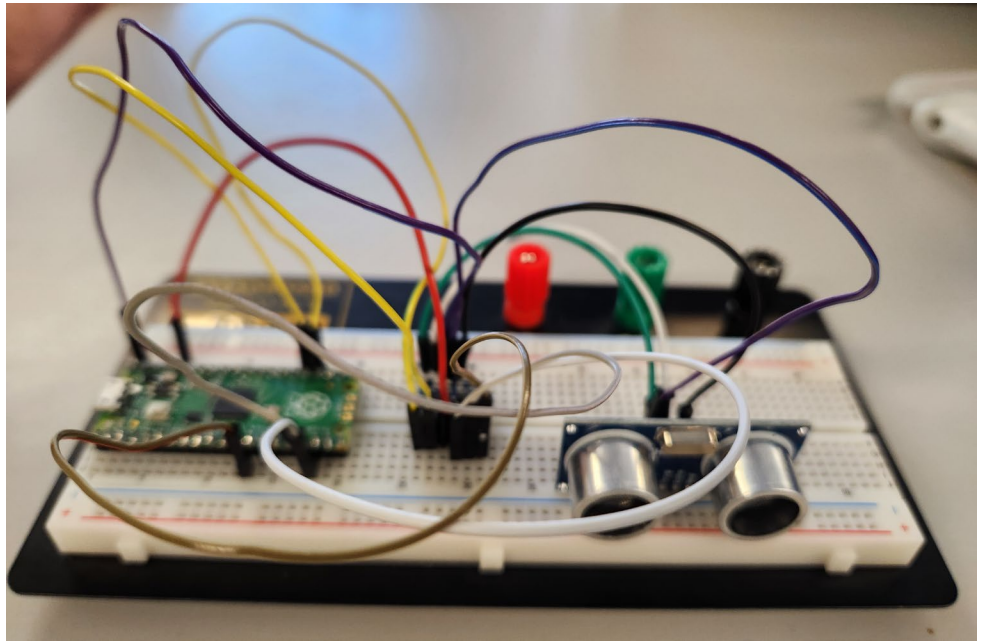
Planimeter to find the area of a 2D shape by using a potentiometer and Green's Theorem
(Calc 3 Lab)



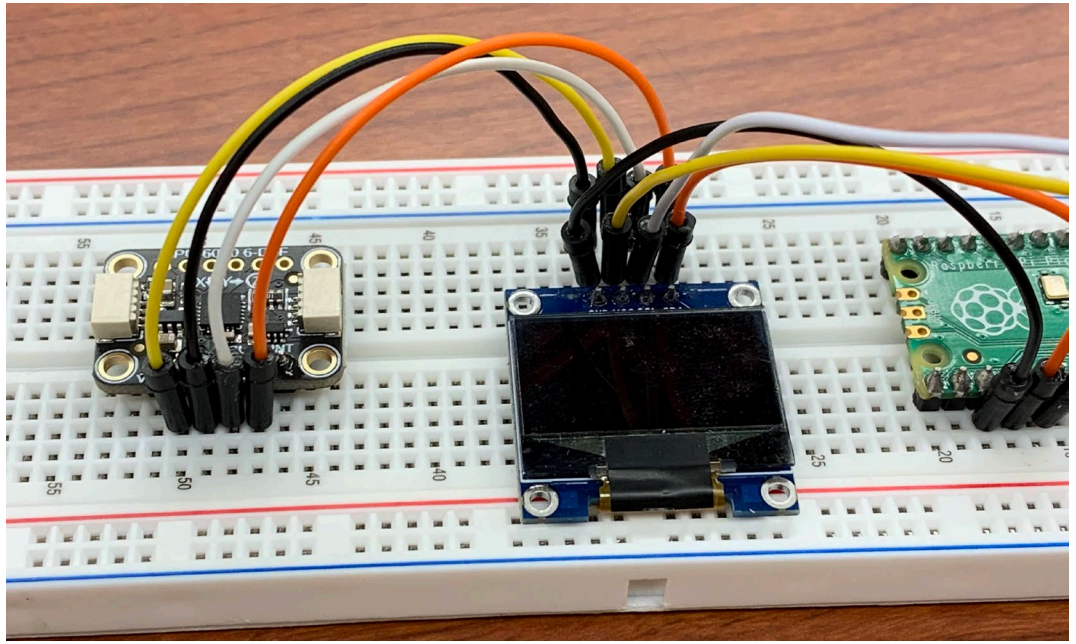
Graph of e^{-x^2} between 0 and 1



A FEW MORE EXAMPLES

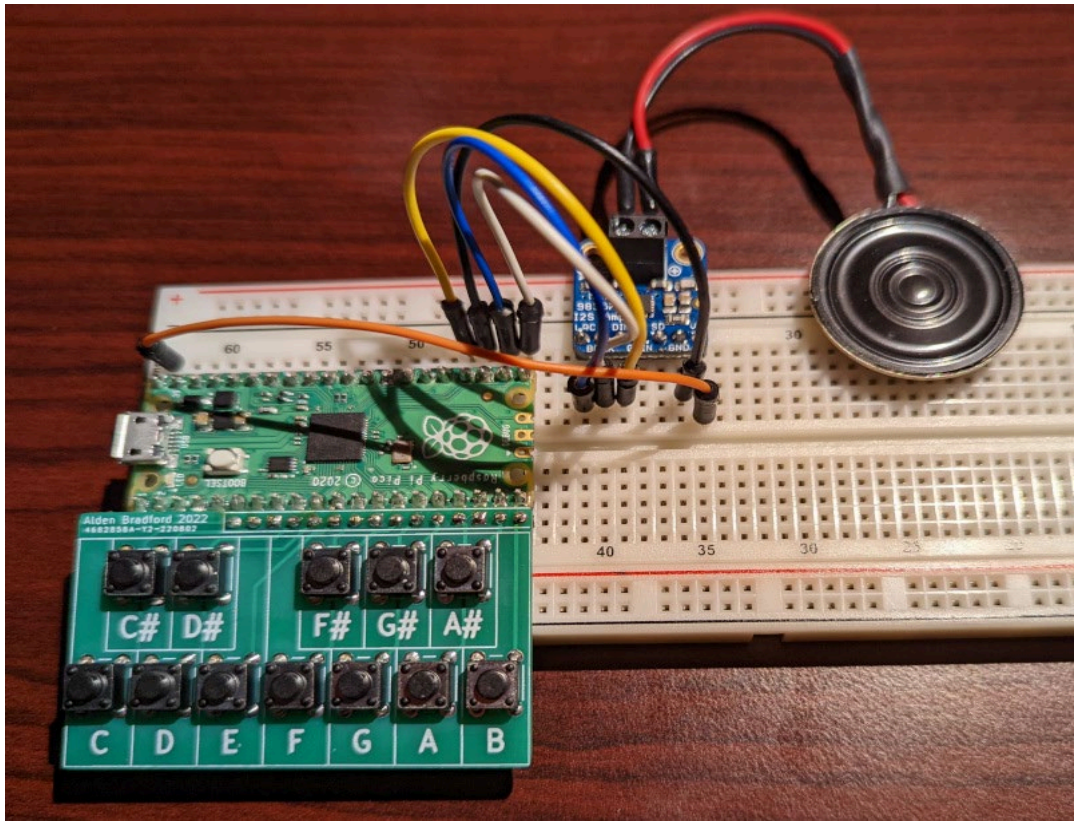


Ultrasonic sensor for estimating distances to nearby objects (Calc 2 Lab)

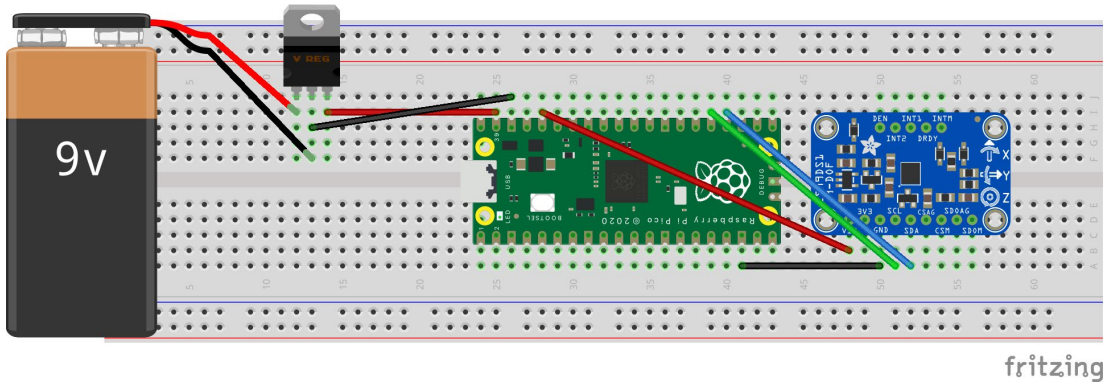


Inclinometer for estimating heights of campus landmarks (Calc 2 Lab)

A FEW MORE EXAMPLES



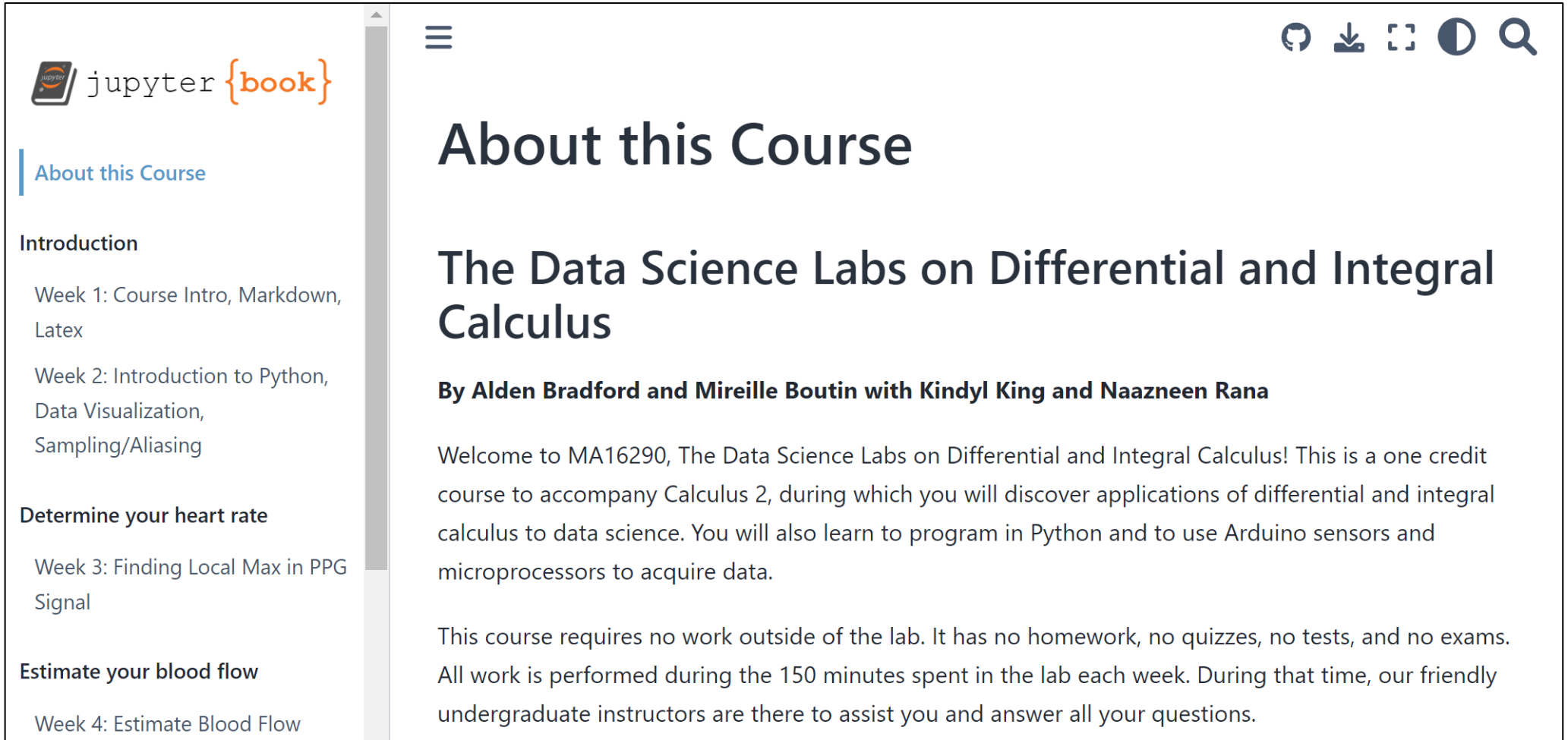
Build a synthesizer using I2S protocol
(Signals & Systems Lab)



Schematics to build a pedometer using an accelerometer
(Signals & Systems Lab)

FREE ONLINE TEXTBOOKS

- MA16290: "DS Labs on Differential and Integral Calculus"
  - https://thedatasciencelabs.github.io/DSLab_Calculus/



FREE ONLINE TEXTBOOKS

- MA16290: "DS Labs on Differential and Integral Calculus"
  - https://thedatasciencelabs.github.io/DSLab_Calculus/
- MA 26190: "DS Labs on Multivariable Calculus"
  - https://thedatasciencelabs.github.io/DataLab_Multivariate_Calculus
- MA 39000: "DS Labs on Signals and Systems (Fourier)"
  - https://thedatasciencelabs.github.io/DSLab_Fourier/
- MA 49000: "DS Labs on Probability"
  - https://thedatasciencelabs.github.io/DSLab_Probability/

STUDENT EVALUATION (CALC 2 LAB)

"I wanted to share the impact of this class on me. I was excited to take this class because of its interdisciplinary element and practicality. I think it's so cool how we used math, programming, and hardware for these labs. In high school I liked my classes, but I was always thinking when will I ever use this stuff? This extends particularly to calculus. This was the first time I applied calculus concepts and it was a very rewarding experience."

– Email from Student

STUDENT EVALUATION (CALC 2 LAB)

"I took the MA16290 pilot class last semester as an FYE student. It was not a large time commitment (only a 1 credit hour class), and **provided a low-stress, collaborative environment where I could practically apply some of calculus 2 knowledge** [*emphasis mine*]. This class was also much more hands on than many of my other typical FYE classes, which I loved. The TAs (as well as professor Mimi) were always more than willing to explain concepts and answer questions. I came into the class without any knowledge about raspberry pi's or arduinos and was able to keep up/learn what I needed as the class went along. Regardless of your major, I recommend taking this class if it's something that interests you, since it probably won't take up much space or overlap with anything else in your schedule. Thanks prof Mimi!"

– Reddit comment

ACKNOWLEDGEMENTS

The development of the data science labs was supported by Purdue's College of Engineering, Department of Mathematics and the Elmore Family School of Electrical and Computer Engineering. We thank Prof. Alina Alexeenko, Prof. Eric Nauman , Prof. Kristina Bross, Prof. Milind Kulkarni, Prof. Uli Walther and Dr. Natasha Duncan for their invaluable input and support. Special thanks also to Julia Long and Ben Manning.

Thank you to Prof. Mimi Boutin who planned, developed, and coordinated the Data Science Labs.

Thank you to Adharsh Sabukumar who built the demonstrations.

For more info: https://engineering.purdue.edu/~mboutin/Data_Science_labs.html
Or email kthood@purdue.edu

THE RESEARCH

- The need to train a larger number of data scientists to meet the predicted market growth has long been established
  - Miller, Steven, and Debbie Hughes. "The quant crunch: How the demand for data science skills is disrupting the job market." Burning Glass Technologies (2017).
- Recently, there has been a proliferation of new data science (DS) programs. But questions remain about the success and relevancy of existing curricula.
  - Aimee Schwab-McCoy, Catherine M. Baker & Rebecca E. Gasper (2021) Data Science in 2020: Computing, Curricula, and Challenges for the Next 10 Years, Journal of Statistics and Data Science Education, 29:sup1, S40-S50
- Asamoah et al. [3] state that DS is an interdisciplinary field and highlight two successful interdisciplinary teaching formats.
  - Asamoah, Daniel Adomako, Derek Doran, and Shu Schiller. "Interdisciplinarity in data science pedagogy: a foundational design." Journal of Computer Information Systems 60.4 (2020): 370-377
- Irizzary [4] argues that "data science is not a discipline but rather an umbrella term used to describe a complex process involving not one data scientist possessing all the necessary expertise, but a team of data scientists with nonoverlapping complementary skills." Based on this, they suggest that DS programs should bring applications to the forefront, offer different tracks, and develop programming skills.
  - Irizarry, R. A. (2020). The Role of Academia in Data Science Education . Harvard Data Science Review, 2(1).
- Other research suggests that assigning relevant and meaningful work is a major factor of retention in computing.
  - L. Barker, C. L. Hovey and L. D. Thompson, "Results of a large-scale, multiinstitutional study of undergraduate retention in computing," 2014 IEEE Frontiers in Education Conference (FIE) Proceedings, 2014, pp. 1-8.

COSTS:
CALC 2 LAB MATERIAL

Equipment	Price	Quantity
Raspberry Pi Pico	5	1
micro USB to USB connector cable	14.99	12
Jumper wires (20 male to male)	1.16	1
Breadboard	19.99	10
Pulse sensor (MAX30102)	3.25	1
Ceramic capacitor 0.1 uF	5.48	25
Tactile switch (TL1105)	22.93	100
OLED display (SSD1306)	15.99	4
Accelerometer (LSM9DS0)	9.99	3
9V battery	12.99	8
9V battery clip	3.69	1
Voltage regulator (7805)	7.99	25
Ultrasonic distance sensor (SR04)	9.19	5
Level shifter 3.3V to 5V	7.99	10
Soldering iron and wire		
Rope		
Measuring tape		
Cart for supplies		
Average Cost per Student	**$ 28.70**	

COSTS:
CALC 3 LAB MATERIAL

Equipment	Price	Quantity
Raspberry Pi Pico	5	1
micro USB to USB connector cable	14.99	12
Jumper wires (20 male to male)	1.16	1
Breadboard	19.99	10
Neopixels (WS2812) (5 pack)	3.5	1
Level shifter 3.3V to 5V	7.99	10
RGB color sensor (TCS3472)	12.88	3
OLED display (SSD1306)	15.99	4
Camera (Arducam HM01B0)	12.99	1
Ceramic capacitor 0.1 uF	5.48	25
Tactile switch (TL1105)	22.93	100
Potentiometer	9.79	3
Wood dowel (2)	9.49	15
Screws (100)	7.72	1
Screw knobs (50)	4.99	1
Rubber o-ring	2.92	10
PLA for 3D printed parts (STL files free)		
Soldering iron and wire		
M&Ms		
Cart for Supplies		
Average Cost per Student	**$ 52.33**	