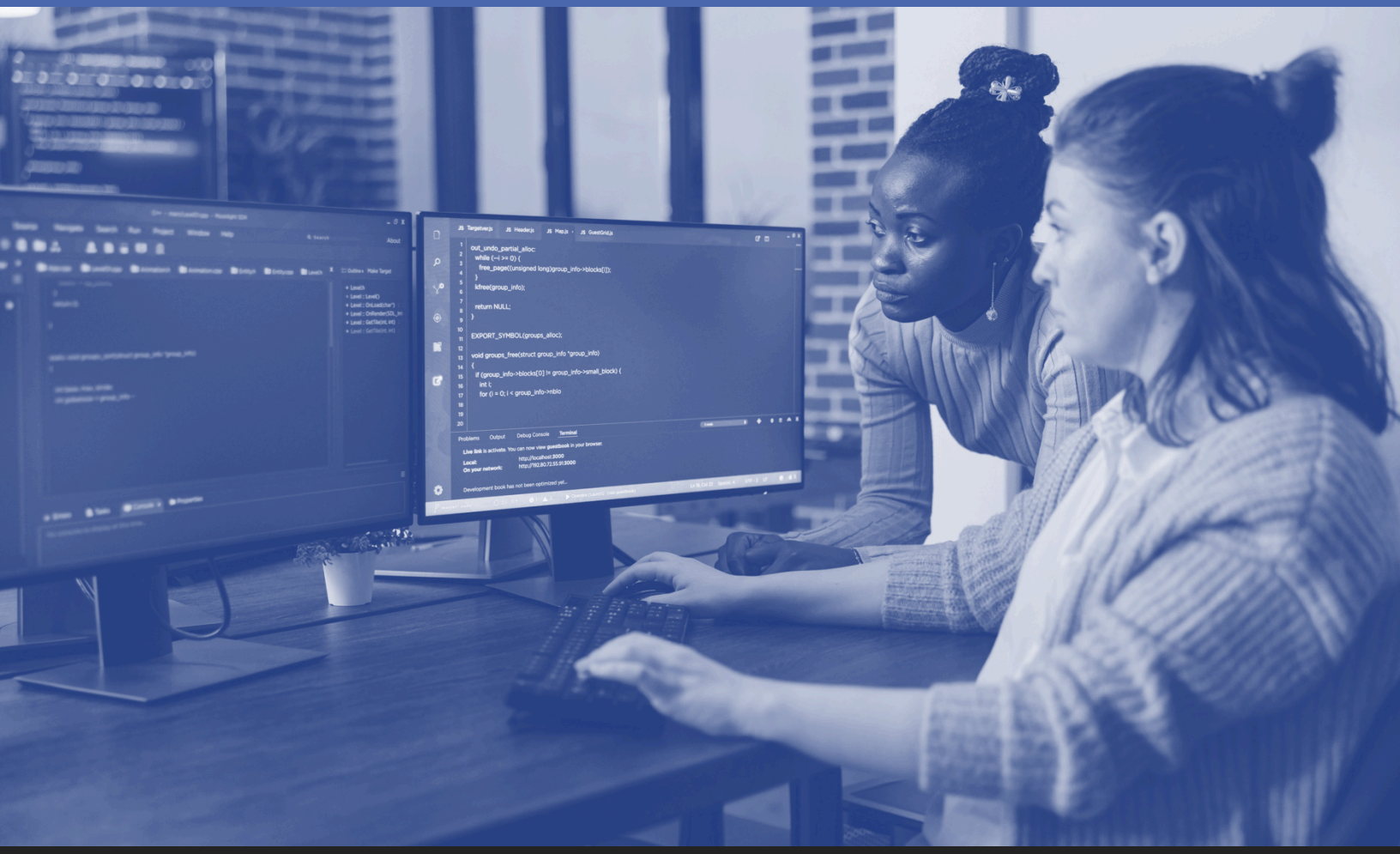
API CAN CODE



Curriculum Overview

November 2024

UNIVERSITY OF MARYLAND 18 56 COLLEGE OF EDUCATION

This work was made possible through generous support from the National Science Foundation (Award # 2141655).

API Can Code

About the Project

API Can Code (APICC) is a research project aimed at situating data science in the lived experiences of today's students. The project explores ways to introduce high-school students to the computational foundations of data science by having them explore meaningful and authentic data that align with their interests. In particular, the learning experiences will center on application programming interfaces (APIs) for popular platforms and services, allowing students to understand the digital worlds surrounding them better.

APICC is part of a research-practice partnership between the University of Maryland and Washington Leadership Academy. As part of this partnership, the APICC team will work with students and teachers to design an interest- and data-driven data science curriculum and iteratively refine and study the curriculum in classrooms over three years.

This document provides an overview of the curricular units that have been developed and tested in the 2024-2025 school year.

For additional information related to API Can Code, visit **apicancode.umd.edu**

APICC Curriculum

Overview

The APICC team has been collaborating with teachers and students from Washington Leadership Academy during this year to test and refine the curricular materials. The curriculum includes three units that introduce students to the computational foundations of data science. Below is a description of these units.

Unit 1 - Data in Learners' Lives

The "Data in Learners' Lives" unit aims to empower students with a comprehensive understanding of data, its significance, and its impact on various aspects of their lives. The unit covers the following goals and learning activities:

- ***Introduction to Data:*** Students will gain a fundamental understanding of data, including its definition, types (qualitative and quantitative), and significance in today's information-driven world. They will also explore and discuss how much data is created every minute and which data they consume and create.
- ***Data Collection and its Purpose:*** Students will explore the diverse data collection entities, such as governments, corporations, educational/research institutions, and non-profit organizations. They will investigate the motivations for collecting data, including its applications in decision-making, research, and problem-solving. Additionally, students will be introduced to the Data-Information-Knowledge-Wisdom (DIKW) model and learn about data transformation into valuable insights and informed decisions.
- ***Sources of Data:*** Students will differentiate between primary and secondary data sources, understanding how information is gathered firsthand or from existing datasets and how these distinctions impact data reliability and accuracy.
- ***Evaluating Datasets:*** Students will be introduced to the 5Vs model of evaluating a dataset's quality and equipped with critical thinking skills to assess datasets. They will ensure they can discern biased, incomplete, or misleading information and verify the data's reliability and trustworthiness.

- ***Data Impact and Equity:*** Students will explore how data can influence social equity and inclusivity. By analyzing real-world situations, students will understand how data can perpetuate biases or contribute to creating a more equitable and just society when used responsibly.

Through this curricular unit, students will become informed and responsible consumers of data, enabling them to make more informed decisions and fostering a deeper awareness of data's role in shaping the world around them. Moreover, they will be encouraged to consider the ethical implications of data use, promoting a sense of responsibility as future citizens and potential data collectors themselves.

Unit 2 - Computational Foundations of Data Science

The "Computational Foundations of Data Science" unit introduces students to the field of data science and helps students gain the essential computational skills required for it. They will learn how to programmatically retrieve and manipulate data from diverse Application Programming Interfaces (APIs) using a block-based platform. This unit covers the following learning goals and activities:

- ***What is Data Science?:*** Students will discuss the different fields contributing to data science and the concepts and practices the field entails.
- ***Accessing Data:*** Students will learn to use APIs to access and save data from government databases and private companies for analysis.
- ***Data Processing:*** Students will learn to clean and preprocess data using programming, specifically in Python, to prepare a dataset for analysis.
- ***Forms of Data:*** Students will learn to recognize and use data in different formats, including JSON and CSV.

Through this curricular unit, students will learn to work with data and manipulate it to meet their analysis requirements. They will also recognize the value of programming techniques in automating efficient workflows for this process.

Unit 3 - Data Science Practices

The "Data Science Practices" unit focuses on the practical aspects of data science, where students will analyze and visualize data to extract meaningful insights. They will be introduced to various data analysis techniques and statistical methods.

- ***Data Visualization with CODAP:*** Students will learn to use the CODAP platform to communicate findings effectively to a chosen audience.
- ***Data Analysis:*** Students will learn to use informal and formal techniques to describe patterns and findings in a dataset.
- ***Graphs and Figures:*** Students will learn to create, customize, and interpret a variety of appropriate summary plots for different kinds of data, including histograms, bar charts, and scatter plots, as well as map-based and time-series visualizations.
- ***Statistical Testing:*** Students will learn to create and use appropriate statistical models, including linear models and confidence intervals.

The unit emphasizes the importance of storytelling with data and conveying insights clearly and compellingly. In addition to the skills of creating data visualizations, students will be asked to think critically about what kinds of data visualizations will best serve the needs of the questions they seek to answer and the stories they seek to tell.

Final Projects

Students will complete a final project on a topic of their choice, drawing data from an API source of their choosing, processing it with EduBlocks, visualizing the data with CODAP, and presenting their findings to the class. The final project is intended to include all three phases covered in the curriculum applied to a subject of interest for the students.

Overall, this data science curriculum is designed to provide learners with a well-rounded understanding of data science. It starts with its foundational concepts and ethics, advances to essential computational skills, and culminates in applying data analysis and visualization techniques to draw meaningful conclusions and effectively communicate their insights. By the end of the curriculum, learners will be equipped with the knowledge and practical skills necessary to tackle real-world data science challenges.

Appendix

Below is a table detailing the subjects taught in each lesson, the datasets, and the main activities. Additionally, the presentation is linked to each lesson.

Unit	Lesson	Topics Covered	Datasets used	Main Activities
Unit 1: Data in Learners' Lives	1.1 Introduction to Data	Nature of data > Numerical/Quantitative Data, Categorical Data; Data Sources > Public Access Data sets; Time Series Data > Date-and-time Variable Formats	Data Science Salaries; Vitamin D; Wind Map; Plastic Bottle Waste	Defining what data is; Identifying data types; Discussing the importance of data science and how data impacts students' lives.
	1.2 Data Collection and its Purpose	Ethics > Privacy, Bias	Data Around the World; TikTok's Privacy Issues; Social Media Polls; Facial Recognition	Exploring and comparing the "Data Never Sleeps" visualization; Discussing data collector and their motivation, privacy issues, and who is represented in data collection and who is not.
	1.3 Using Data: the DIKW Model	Time Series Data > Forecasting	23andme; Cholera Outbreak; DC COVID-19 Dashboard; Heat Sensitivity Exposure Index	Discussing stakeholders and privacy issues; The DIKW model; Local Issue Investigation.
	1.4 Sources of Data	Data Sources > Public Access Data sets; Ethics > Bias	Facebook report; DCPS schools locations; Game scores (API);	Identifying sources of data (primary and secondary sources), exploring Selfiecity, and comparing the findings of various countries; Analyzing three case studies of data collection by stakeholders (GPA scores, Instagram, Starbucks).

	1.5 Evaluating Datasets	Nature of data > Tabular structure Ethics > Bias	Starbucks Yearly Data	Evaluating the Starbucks dataset with the 5V's; Evaluating dataset (student's choice).
	1.6 Data Collection: Impact and Equity	Data Sources > Surveys; Inquiry with Data > Type of Inquiry; Sampling & Simulating > Random Sampling; Generalizability	Students' grades (collected in real-time)	Defining data science questions about local issues and sources to answer them; Discussing messy data and sampling methods, Designing a Google form to collect data on local issues; and Identifying sampling methods in various cases.
Unit 2: Computational Foundations of Data Science	2.1 What is Data Science	Inquiry with Data > Data Cycle	Grocery Store Locations (OpenDataDC)	Defining what data science is; Raising questions about Food Deserts in DC; Identifying metro stops near the grocery stores, calculating and logging their distances; Presenting the findings; Discussing TikTok viral video; Finding interesting dataset in OpenDataDC, Defining the variables and raising questions about it.
	2.2 Manual Data Processing	Inquiry with Data > Data Cycle; Data Manipulation > Filtering; Mutating, Reshaping, Transforming	[TikTok Video]	Analyzing the Streetlights datasets; Identifying missing data; Data calculations - CountIF, Average; Variable transformation; EduBlocks introduction.

	2.3 Intro to Programming with EduBlocks: Filtering & Data Transformation	Nature of data > Tabular Structure; Data Manipulation > Filtering;	Streetlights (OpenDataDC)	EduBlocks and JSON introduction; Understanding the structure of a given program (Mario Kart) and identifying the variables; Modify the code to print the max speed
	2.4 Accessing Data with APIs using RapidAPI	Data Sources > Public Access Datasets; Programming > Logic; Data Manipulation > Cleaning; Filtering; Mutating, Reshaping, Transforming	Mario Kart; IMDB Top 100 Movies API	Introducing APIs and RapidAPI; Connecting EduBlocks to APIs; Modifying a given program (Top 100 movies) and manipulating the data.
	2.5 Preparing Data for Analysis	Data Sources > Public Access Datasets; Programming > Logic; Data Manipulation > Filtering; Mutating, Reshaping, Transforming	IMDB Top 100 Movies API	Modifying a given code (Mario Kart); Understanding the structure of a given program (Billboard), Asking questions about the data and modifying the code accordingly (favorite artist); Sharing the findings.
	2.6 Data Analysis in Practice	Data Sources > Public Access Datasets; Programming > Logic; Data Manipulation > Cleaning; Filtering; Mutating, Reshaping, Transforming	Mario Kart; Billboard Hot 100 API	Exploring the Free NBA API on RapidAPI; Modifying a basic program; Understanding the structure of the response; Modifying the program to investigate favorite team.

Unit 3: Data Science Practices	3.1 Intro to Data Visualization	Data Visualization > Comparing Plots; Ethics > Misleading Information	Napoleon's invasion; The routine of creative people; Hollywood Hits; Data Breach; Country Income and Life Expectancy (Gapminder)	Introducing the Stoop Test and the concept of Visualization, Discussing various infographics; Comparing graphs; Examining Gapminder trends; Choosing the best graph for each scenario; and discussing misleading information.
	3.2 Exploratory Analysis with CODAP	Map Data > Dimensions of Map Data; Location and Region Data; Map-Plotting	Food Nutrition Mammals (CODAP) Roller Coasters (CODAP)	Examining a given graph (Food nutrition); Introducing CODAP; Investigating the Mammals dataset and various graphs; Creating a graph (Students' choice); Discussing Outliers, Exploring the Roller Coasters dataset; Creating a graph (Students' choice).
	3.3 Graphs and Figures: One Variable	Data Visualization > Dotplot; Histogram; Bar Chart; Measures of Center > Mean; Median; Distributions & Variability > Distributional Shape	Prices of eggs graph; Four Seals (CODAP)	Examining a given graph (prices of eggs); Investigating the Four Seals dataset; Discussing various data vis methods and graphs, Choosing the best graph for each scenario (Histogram/bar chart); Creating a new graph.

	3.4 Graphs and Figures: Two Variables	Nature of Data > Frequency Tables; Variable Association > Correlation; Third Variable; Data Visualization > Scatterplot	Choosing a Collage; Roller Coasters (CODAP)	Examining a given graph (Choosing a college); Exploring the Roller Coaster dataset; Understanding and creating scatterplot; Analyzing the correlation between test scores and final grades.
	3.5 Statistical Testing	Distributions & Variability > Standard Deviation (SD); Interquartile Range (IQR); Significance > Tests; Confidence Intervals	Teens Usage of Social Media Platform; Zillow dataset; Ravens and Commanders NFL game scores	Examining a given graph (Teens' usage of social media); Exploring the Zillow dataset and generating a graph that interests the students after watching the Khan Academy video; Comparing data of different states; Modifying the EduBlocks program that uses the Zillow API, Investigating data about DC; Creating interesting graph; Discussing SD, sampled population and biases, Creating a program to investigate students' preferred location; Analyzing the Ravens and Commanders NFL game scores.

	3.6 Linear Models	Variable Association > Linear Models; Time Series Data > Forecasting	Train Derailments; Earthquakes (USGS)	Examining a given graph (Train Derailments); Watching the earthquakes video, Exploring the Earthquakes API, Creating an EduBlocks program and copying the input to CODAP; Creating graphs and presenting the mean, SD, and variance; Describing the graphs; Creating predictions.