

Computer Science Education in WA High Schools for 2017-2018

A Code and Cognition Lab project by Dr. Andy Ko and Zoshua Colah to identify high schools in Washington State which teach computer science.

About this Repository

This repository contains the files and instructions to process high school student performance data provided by the Office of Superintendent of Public Instruction to generate a map of Washington state high schools which teach computer science.

The goals of this project are:

1. To prepare a dataset which labels courses as computer science or non-computer science based off a definition
2. To merge and combine various datasets to create one single source of truth.
3. To create a map visualization which highlights which schools teach computer science and don't and which schools do we have no information on.
4. To embed the map into Google sites.

Getting Started

Knowledge Pre-reqs

Before you get started it is important to know the following technologies. If you do not, we recommend using the links we have provided below.

Python 3 & Jupyter Notebook:

Python is a scripting language for data science used by programmers that want to delve into data analysis or apply statistical technique. It is the most popular language for data science on Stack Overflow and on Kaggle - this ensures there is always more than sufficient community support and help. In addition to this there are many data science packages created for python.

We will be using python in this project to clean, prepare (wrangle, re-shape, merge), and generate a

dataset for an interactive data visualization report which will be made in Microsoft Power BI

- [Install Python 3](#)
- **Learn:** There are many online Python courses. Our recommendation are the [Microsoft Python for Data Science edX courses](#).
- [Documentation](#)

Python Data Science Packages

- **Pandas:** for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. The course recommended above discusses how to use this package in detail.
 - [Installation Guide](#)
 - [CheatSheet](#): We highly recommended reading this cheatsheet and using it as a reference point when manipulating dataframes.

Python Jupyter Notebooks

The Jupyter Notebook is a very powerful tool for creating and presenting data science projects. It can combine code with visualizations, narrative text, mathematical equations and other media. Data scientists use it to write their code, document it, run it, look at the outcome and visualize data in the same environment!

- [Installation Guide](#)
- [CheatSheet](#)

Microsoft Power BI Online

Power BI, is a data visualization web application (similar to Tableau) which is used to build data driven reports and dashboards which contain interactive visualizations such as pie charts, bar charts, bubble maps, chloropleth maps, and more.

In this project we will be using Power BI to create an interactive data map visualization which allows a person to see which schools teach and don't teach computer science in Washington State. Furthermore there will also be a page allowing them to explore the schools for each state.

- [Sign In and Start Using Power BI](#):
 - **For UW Students and Faculty:** All active UW Net ID students and faculty members have access to Power BI. To use it, click on the link and sign in with your active UW Net ID (e.g. zcolah@uw.edu) and

password.

- **For Non-UW Students:** You can sign in to Power BI online with your organization email (e.g. zubair@techmentor.ml). *Note: you cannot use a public domain email service ID such as zoshua@gmail.com or kashmira@yahoo.co.in*
- [Learn](#): Microsoft's own website does a fabulous job of documenting and teaching you how to use their product. [We recommend that you complete the guided learning section of visualization with Power BI.](#)

Google Sites

Google Sites is a structured wiki- and Web page-creation tool offered by Google. The goal of Google Sites is for anyone to be able to create simple web sites that support collaboration between different editors.

In this project we will use Google Sites, to embed our interactive data visualizations made in Power BI into the CS for All Google Website

- [Sign In and Use](#)
- [Documentation](#)

Setup

To setup this project on your local computer you will need to:

1. [Install Git](#): a free and open source distributed version control system designed to handle everything from small to very large projects with speed and efficiency.
2. [Install Gitbash Bash \(only if you have Windows\)](#)
3. [Create a Github Account](#)
4. **Install Anaconda**
5. **Open your terminal(Mac) or Gitbash(Windows)**
 1. Type `cd desktop` to change your directory to the desktop
 2. Type `git clone https://github.com/zcolah/CS_High_School_Education_WA.git` to clone this repository.
 3. Type `cd CS_High_School_Education_WA` to change your active directory to the repository folder you just downloaded
 4. Type `jupyter notebook` to launch Jupyter Notebook. You can use it to explore and modify the python notebooks in this repository and to create a new notebook in this repository.

6. Open and Read the following notebooks in the following order

a. Part 1: Labelling Courses as CS and Non-CS Courses

b. Part 2: Preparing Dataset for Visualization

c. Part 3: Making Visualization in Power BI

d. Part 4: Embedding Visualization in Google Sites

To learn more about any dataset read the data schema notebook. We recommended referring to this while reading the Part 2 Notebook.