**UNIT PLAN**

**Title**

Introduction to Data Visualization using Python.

**Author**

Daniel Moscoe

dmoscoe@gmail.com

**Relevant courses**

New York Algebra I

New York High School Computer Science

New York High School Data Science

AP Computer Science Principles

**Essential questions and enduring understandings**

EQ: How can we summarize large datasets?

EU: We can summarize large datasets with descriptive statistics. Descriptive statistics are useful because they summarize a large amount of information with just a few numbers.

EQ: How can we compare large datasets using visualizations?

EU: We can compare some aspects of large datasets using histograms. Histograms help us describe and compare the center, shape, and spread of univariate datasets. In general, data visualizations are images that help us see patterns in datasets, and they help us compare datasets with each other. Data visualizations are a powerful, widely used tool for telling stories with data.

EQ: How can programming help us find meaning in data?

EU: Programming helps us work with data by streamlining repetitive computations and procedures. We can create our own programs, and we can modify programs created by others to develop powerful and specialized tools that help us understand our data.

**Introduction**

[pass until lesson plans are complete]

**Standards**

*NY State Computer Science and Digital Fluency Learning Standards*

* 9-12.CT.1: Create a simple digital model that makes predictions of outcomes.
  + [How do students meet this standard? Where is this described in the unit plan?]
* 9-12.CT.2: Collect and evaluate data from multiple sources for use in a computational artifact.
  + [How do students meet this standard? Where is this described in the unit plan?]
* 9-12.CT.3: Refine and visualize complex data sets to tell different stories with the same data set.
  + [How do students meet this standard? Where is this described in the unit plan?]

*NY State Next Generation Mathematics Learning Standards*

* AI-S.ID.1. Represent data with plots on the real number line (dot plots, histograms, and box plots).
  + [How do students meet this standard? Where is this described in the unit plan?]
* AI-S.ID.2. Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, sample standard deviation) of two or more different datasets.
  + [How do students meet this standard? Where is this described in the unit plan?]
* AI-S.ID.3. Interpret differences in shape, center, and spread in the context of the datasets, accounting for possible effects of extreme data points (outliers).
  + [How do students meet this standard? Where is this described in the unit plan?]
* AII-S.IC.2. Determine if a value for a sample proportion or sample mean is likely to occur based on a given simulation.
  + [How do students meet this standard? Where is this described in the unit plan?]
* AII-S.IC.4. Given a simulation model based on a sample proportion or mean, construct the 95% interval centered on the statistic (+/- two standard deviations) and determine if a suggested parameter is plausible.
  + [How do students meet this standard? Where is this described in the unit plan?]

*AP Computer Science Principles Learning Objectives*

[Does what is contained in this unit really match the intention of these standards? Would this unit help someone be successful on the AP Exam?]

* DAT-2.A. Describe what information can be extracted from data.
  + [How do students meet this standard? Where is this described in the unit plan?]
* DAT-2.C. Identify the challenges associated with processing data.
  + [How do students meet this standard? Where is this described in the unit plan?]
* DAT-2.D. Extract information from data using a program.
  + [How do students meet this standard? Where is this described in the unit plan?]
* DAT-2.E. Explain how programs can be used to gain insight and knowledge from data.
  + [How do students meet this standard? Where is this described in the unit plan?]

**Student performance portfolio**

[Come back to these descriptions after you create the assignments and rubrics.]

* *Designing Histograms*. Given a dataset as CSV, students create histograms that reveal the shape of their datasets. Students choose axis labels, colors, bin widths, and titles that make their plots informative summaries of the data they describe. Students explain their choices in the context of the data. Code and annotations accompany the plots.
* *Histograms and Descriptive Statistics*. Students compute descriptive statistics for their chosen datasets. Students add elements to their histograms that show appropriate measures of center and spread. Code and annotations accompany the updated plots.
* *Using Histograms to Compare Groups*. Students analyze how the properties of their dataset may vary by group. Students visually compare the shape, center, and spread of groups within their dataset by creating histograms in stacked subplots (“small multiples”). Students interpret their findings in the context of the data. Code and annotations accompany the plots.

**Lesson plans**

In progress…