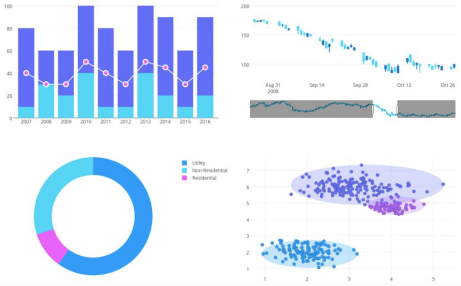
Data Visualization



Aim: How can visualizing data be useful?

Objective: With scaffolds, students will be able to visually present information from source data programmatically

References:

- *Plotly graph documentation: https://plotly.com/python/*
- *Pandas Data Types: https://pbpython.com/pandas_dtypes.html*
- *Billboard Top 100 Singles 2019: https://data.world/kcmillersean/billboard-hot-100-1958-2017*

Lesson Overview:

1. Students examine excel document of data
2. Students program in python to manipulate data (jupyter notebook + pandas)
3. Students program in python to graph data (jupyter notebook + plotly)

Themes:

* What are the benefits and limitations of text data? (excel)
* What is a library? (pandas and plotly)
* Jupyter notebook (maybe we can assume they are already familiar with this)
* pandas
    → data frame
* plotly
    → graph objects (https://plotly.com/python/bar-charts/)
    → figures-
    → graph types (bar, pie, line, etc)

Student Task:

Students will start with a pre-made excel file and a template graphing code. They will complete missing code elements in order to visually present data. Students will create several different plots described in problem prompts and will create their own plot from an open ended prompt as an exit ticket.

Extension:

Sliders, dropdown menus, and interactive elements

Required Resources:

Data Sets: For example, Open Data NYC NYPD CCRB Complaint Data Billboard Top 100

Jupyter Notebook: https://jupyter.org/install
Python Plotly **and Jupyter Notebook extension**: https://plotly.com/python/getting-started/

Class: Intro CS / 9th grade Teacher: Date: September 2020

Unit: Data
Aim: What are the benefits and limitations of text data (excel)?
Objective: SWBAT understand and articulate the benefits and costs of text data.
SWBAT understand the limitations of text data.
SWBAT articulate an example when text and/or visual representations of data would be more useful.
Standards:
NYS K-12 Computer Science and Digital Fluency Standards, Computational Thinking, DATA ANALYSIS AND VISUALIZATION
9-12 CT-3. Refine and visualize a large data set using an appropriate tool in order to persuade an audience.

Materials

Do Now: 5 minutes	Prompt for Stop & Jot / Turn & Talk: **Do Now: Data Explore Your Prior Knowledge** What is data? How/where do we store data? Table partners Turn and Talk to compare your answers. Stop and discuss any differences in answers.
Mini Lesson: 7 minutes	**Definitions:** **Data:** facts and statistics **collected together** for reference or **analysis.** **Data Visualization:** Data visualization how we store data. It is the **graphic representation** of data. It involves producing images that **communicate relationships** among the represented data to viewers of the images. Share Out: **Why would we want to use data?** ● To tell a story ● To give information ● To persuade ● To analyze Pen a prediction: **When would we want to use different types of data visualizations?**

	Navigating data in Excel: How can we do what we want to be able to? Basics: Highlight the Performer Column Yellow Highlight the Song Column Blue Highlight the Weeks on Chart Column Green Filter Demo: Navigate to All the Number One Songs Sort Demo: Sort by Position Double Sort Demo: If I want to find the person who was number one on the week closest to my birthday, I need to take two actions. Filter by number one artists and then by my birth month. Try it with your birth month.
Activity: 20 minutes	Explore the data set to find the answers to the following questions. 1. How many different number one songs were there in 2019? 2. What were the number one songs in 2019? 3. What song was number one the longest? How do you know? 4. What are the top ten songs spending the most weeks on the chart? 5. What are the top ten most featured artists? 6. How would you calculate the percent of time the artist that spent the most time at the #1 spot was there? 7. Attempt to follow a song in the top 10. When did it first appear there? How long was it there? When does it leave? How do you know this?
Closing/Exit Ticket: 5 minutes	What benefits can you see about using spreadsheets? What would you like to be able to do with the data but are currently unable to?
Homework:	Using KtS - Write out an algorithm that you would need to use to announce a song's appearance in the top 25 and the date of appearance, and to announce its departure and the date of departure. Or Choose your own data that you would like to find.

Day 2 Student Materials:

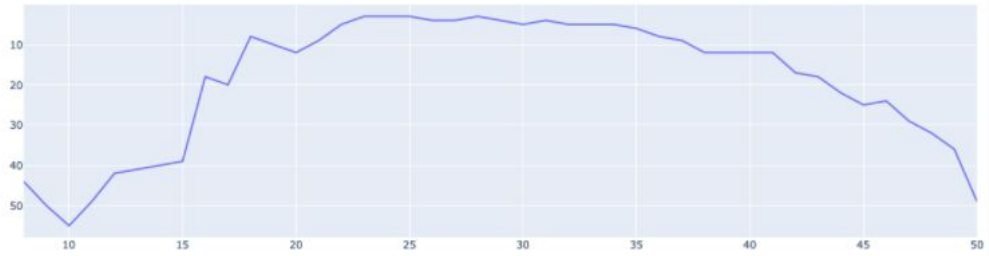
https://docs.google.com/document/d/1XTRFMspGJxOjsxnrQinWsA_ILWIzCvlejn7uuKgJ8aY/edit?usp=sharing

Do Now: __ minutes	Try to describe the the types of information you will find in each column of our 2019 Billboard data set using only one word (Text, String, Integer, Number, Boolean, Float, Date, etc) **Do Now: Data Review (5 minutes/ 5 points)** Try to describe the the types of information you will find in each column of our 2019 Billboard data set using only one word (Text, String, Integer, Number, Boolean, Float, etc) WeekID | Week Number | Week Position | Song | Performer | Previous Week Position | Peak Position | Weeks on Chart 1/5/19 | 1 | 37 | A Lot | 21 Savage | | 37 | 1 1/5/19 | 1 | 40 | Blue Christmas | Elvis Presley | | 40 | 1 1/5/19 | 1 | 57 | Swervin | A Boogie Wit da Hoodie Featuring 6ix9ine | | 57 | 1 1/5/19 | 1 | 58 | Can't Leave Without It | 21 Savage | | 58 | 1 1/5/19 | 1 | 65 | Startender | A Boogie Wit da Hoodie Featuring Offset & Tyga | | 65 | 1 WeekID | Week Number | Week Position | Song | Performer | Previous Week Position | Peak Position | Weeks on Chart Table partners Turn and Talk to compare category descriptions. Stop and discuss reasonings for categories that you described differently
Mini Lesson: __ minutes	**Mini Lesson: Data Analysis with Python (10 minutes/ 5 points)** Python commands | Definitions `import pandas as pd` `import plotly.graph_objects as go` | "Import" is a command that introduces a new set of tools into your environment. Think of your coding space as an empty garage, and each import adds a tool to your garage that wasn't there before (like a tablesaw, or an electric drill) `df = pd.read_csv` | Df stands for "dataframe". pd.read_csv is telling our imported tool "pandas" to read our excel file (the csv) and to store the contents of the data in a variable called "df" Once our excel data is inside our coding environment as the dataframe variable "df", we can use python to interact and manipulate with the data. We'll be using these commands `df['Column Name']=="thing"` | Selects specified values from specific column `df.sort_values(by=['Weeks on Chart'], ascending=False)` | Sorts table by specified column values `df.drop_duplicates(['Column name'])` | Removes duplicate row entries in specified column `smaller_df = df.head(10)` | Creates a new dataframe with the first 10 rows of the original dataframe
Activity: __ minutes	**Data Manipulation**

	Activity: Building dataframes (20 minutes/ 10 points) Given the prompts and starter code below, create and print dataframes from our "Billboard_Top_100.csv" that produce the values described in the prompt. 1. Create and print a dataframe of the top 10 most featured artists descending ``` artist_instances = df['Performer'].value_counts().rename_axis('artist_name').reset_index(name='counts') top_10 = artist_instances.head(____) print(top_10) ``` 2. Create and print a dataframe of the top 10 songs spending the most weeks on the charts ``` sorted_df = df.sort_values(by=['_____'], ascending=False) no_duplicates = sorted_df.drop_duplicates(['Song']) top_10 = no_duplicates.head(___) print(top_10) ``` 3. Pick one song that has been on the chart for at least 25 weeks and create a dataframe of its weekly position ``` song_condition = df['_____']=="_____" song_df = df[_____] print(song_df) ```
Closing/Exit Ticket: __ minutes	Exit ticket previews Day 3 lesson by prompting students to make hypothesis about function of graphing code **Exit Ticket** Create a hypothesis about what you think the code segment below does. Explain your hypothesis by referring to specific sections of the code ``` fig = go.Figure([go.Bar(x=top_25['artist_name'], y=top_25['counts'])]) fig.show() ```
Homework:	N/A

Day 3 Student Materials:

https://docs.google.com/document/d/1woHvlcwWMFa_uuRYD2khZ81Lz8kzijbVBse62GPmvTM/edit?usp=sharing

Do Now: __ minutes	**Do Now: Data Review (5 minutes/ 5 points)** Describe this graph and explain ways that is is useful, not useful, or both 

Mini Lesson: __ minutes	**Mini Lesson: Graphing Data with Python (10 minutes/ 5 points)**
	`fig = go.Figure([go.Bar(x=dataframe['column_name'], y=dataframe['column_name'])`
	Constructing a visual graph in Python requires commands with several arguments. The line above creates a graph object as a variable "fig". Broken down it contains…
	• A creation of a figure object "go.Figure"
	• Designation of the type of figure object "go.Bar"
	• Specification for the data to be used on the X and Y axis
	  ○ In this case, "dataframe['column_name']" indicates that the graph will plot the data contained in 'column_name' of the dataframe as its axis values
	After a graph object is created and stored as a variable "fig", it can then be plotted and displayed with a simple command
	`fig.show()`

Activity: __ minutes	**Activity: Building dataframes (20 minutes/ 10 points)**
	
	Based on yesterday's code for creating dataframes, use those data frames to create and display graphs modeling the prompted scenarios
	
	1. Create a vertical bar graph of the top 10 most featured artists descending (count on Y axis, artist name on X axis)
	
	`fig = go.Figure([go.Bar(x=_____['Song'], y=_____['Weeks on Chart'])])`
	`fig.update_layout(title='Top 25 Songs by Weeks on Chart',xaxis=dict(title='Song Title'), yaxis=dict(title='Weeks on Billboard'))`
	`fig.show()`
	
	2. Create a vertical bar graph of the top 10 songs spending the most weeks on the charts (count on Y axis, artist name on X axis)
	
	`fig = go.Figure([go.Bar(x=_____, y=_____)])`
	`fig.update_layout(title='Top 10 Featured Artists',xaxis=dict(title='Artist Name'), yaxis=dict(title='Songs on Chart'))`
	`fig.show()`
	
	3. Pick one song that has been on the chart for at least 25 weeks and graph its weekly position with a line graph (rank on Y axis, date on X axis)
	
	`fig = _____(data=go.Scatter(x=_____, y=_____))`
	`fig['layout']['yaxis']['autorange'] = "reversed"`
	`fig.update_layout(title='Song Ranking by Week',xaxis=dict(title='_____'), yaxis=dict(title='_____'))`
	`fig.show()`

Closing/Exit Ticket: __ minutes	**Exit Ticket** Think of your own graph you could create from the data that is not listed in the prompts. Describe what you would be visualizing below and attempt to write the data processing and graphing code for it
Homework:	