Date(s):	Unit 5, Lab 3.5	Cleaning Data

Prerequisite Knowledge: Spreadsheets Lab, Basic spreadsheet skills

Standards:

Computational Thinking - Data Analysis and Visualization

9-12.CT.2 - Collect and evaluate data from multiple sources for use in a computational artifact

9-12.CT.3 - Refine and visualize complex data sets showing how to tell different stories with the same data set.

Learning Objective(s):

- Investigate and analyze a data set.
- Draw and summarize conclusions based on investigation and analysis of a large data set.
- Understand how a graphical visualization can make the data much easier to understand.
- Understand how selecting a row or column of a table gives a useful slice from the data.
- Understand how "cleaning" and filtering data helps narrow and tell a story from a given data set

AGENDA	Q's/CFUs/MISUNDERSTANDINGS

WARM UP (5 MINS):
Aim:
 	
Do Now:
High schools across the world receive data from the CollegeBoard after an administration of the SAT. This information either comes in the form of a report showcasing key information or as a **CSV file** that contains all raw records of each exam. The records typically list the name of the student, demographic information, and their score on the various subskills that make up the test. Which of the following would be an advantage of the CSV file?
    A. Compact and easy to read results
    B. Ability to sort and filter the data
    C. Easily shared with staff and community
    D. Charts display trends over time
 	
Alternate Do Now:
Given this dataset, ask students to find a trend in the raw data. See if they can do it. Ask leading questions about what difficulties they encounter to find a trend. Or give them a question or prompt to find some data and see how they find it.
	Correct answer is B. The CSV file is the raw records of every student taking the test. If this file was opened up in a spreadsheet program or inputted into a computer program, then the editor could use the program to filter or sort the data. The report may be more visibly pleasing, but it would not offer the ability to sort or filter the data.

MINI LESSON (Demo) (5 MINS): Demo: Ask a question: What are the SAT scores for NY colleges? Open spreadsheet and introduce the College SAT scores data sheet. Allow students to explore and answer some prompts which will lead to difficulties/tedious manual searching in finding trend data.	Demo given a goal in mind. How do we answer a question from a student such as what scores do I need to get into a college in NY?
ACTIVITY (5-7 MINS):	Prompt #3 purposefully doesn't start

Students will search for the data to find answers to their own question or to answer a specific prompt. They will explore the data and have to do a lot of data filtering manually. Prompts for students to search through the data 1 - What is the SAT score in the 25h percentile for Massachusetts Institute of Technology. 2 - What is the SAT score in the 75th percentile for NYU? 3 - What is the range for Rochester in NY? 4 - Which college in NY and what value has the highest score for the 75th percentile?	with University of Rochester. This may cause longer time ot search for the school. Students will discover that sometimes data is imperfect.
MINI LESSON (Demo) (5 MINS): Teacher demos the Create Filter option in Google Sheet. - Create a for the headers - Filter for a specific State like NY - Can also filter by a conditional such as <>= - Show how to sort the data in ascending and descending order A data filter is exactly like a conditional. You only want to see certain rows IF it means a certain conditional such as only being in a specific state.	Demo of the built-in filtering function links with conditionals, analyzing and cleaning data. Demo-ing helps the students who don't already know the filter function and learn best by seeing the movement and locations and instructions. Students who already know how to are following the lab document and will move ahead at their own pace.
ACTIVITY (10-15 MINS): Students will now tabulate data by state. Using the filtering function, create a new table that indicates the lowest 25th percentile score and the highest 25th percentile score in each state listed in the data table.	Students will utilize their new skill filtering data to generate a new table of data.
HOMEWORK: No assigned HW.	