Summary

In this lesson, students analyze multiple years of data from solar cells installed on a residential building in Washington, DC. They will distinguish between two parts of the "story" being told by the data: recurring seasonal change versus year-over-year change. Students will experience how variability can obscure trends and learn methods for exploring "noisy" data. They will produce several distinct plots and continue discussing good and bad data visualization.

Materials

Data file: https://stemcoding.github.io/solar/solar1.csv
Excel desktop (?)
Total time:

Objectives

1. Make predictions and generate questions about the solar panel data
2. Understand how variability/seasonality can obscure larger trends
3. Identify methods and tools for breaking up a large, noisy dataset
4. Reflect on the usefulness of different visualizations
5. Generate new questions based on results of analysis

Key questions

1. What do you expect the solar panel data to look like? Why?
2. What story is this data telling us?
   a. The overt story, seasonal changes vs. the underlying story, decreasing efficiency
3. How can we represent the data in a way that brings out the underlying story?
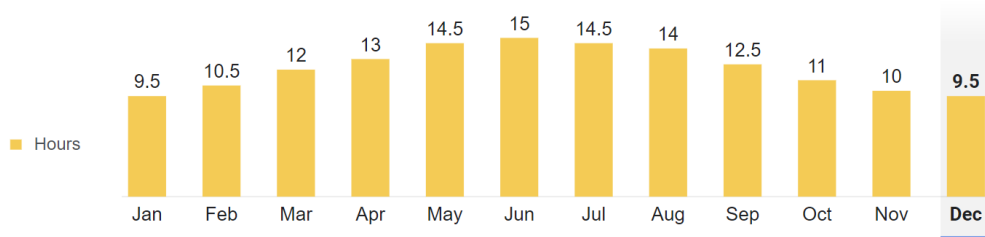4. What do you still wonder about solar panel data?

Statistical concepts, spreadsheet skills, and tools used

1. Sum by month, quarter, and year
2. Trendlines: Simple linear regression
3. Functions: Sum, average
4. Highlighting/selecting data, inserting and formatting charts, copying/cutting and pasting data

PROCEDURE

1.	**Introduce the dataset.** Solar panels convert sunlight into electrical energy. This file contains two columns: Energy produced (in Watt hours, Wh) and time (2017 to 2022, measured once per day) from a science teacher's home solar panels near Washington, DC.	
2.	**Small group discussion: What do we expect the data to look like?** Before downloading the file, have students **make some predictions**. Put the following questions on the board for students to consider: What will you see in the data? What story might it tell us? What do you wonder about it? What might impact the energy produced by solar panels over the months of one year? Over several years? Students will probably think of seasons, cloud cover, temperature, etc. Hopefully they will also consider larger trends. **Encourage them to google climate data** (switch to "graphs" to see more than just temps), info about **solar panel efficiency**, etc. 	<10 min
3.	**Group share/class discussion: What do we expect the data to look like?** Have each group share something they expect or wonder with the class. a. Discussion will probably start with **seasonal changes**. Can show the daylight hours plot:	~10 min

Daylight

Source: NOAA

What months will have least/most energy produced each year?

> **Draw a rough sketch** on board of what the raw line graph might look like based on seasonal changes.

b. Hopefully a student will bring up larger trends (e.g., "I wonder how different it will be from year-to-year and why", "I think they will get worn down over the years"). Discuss **declining efficiency**.

Could be framed as "The science teacher learned that solar panels become less efficient over time, and wants to see if his have significantly declined since 2017".

> "If there is a decline in efficiency, what would the plot look like?" Student can volunteer to **modify the sketch**, or the teacher can do it.

4. <u>Define the goal</u>

We want to investigate the possibility that these solar panels have become less efficient since 2017. We also want to see if our predictions about seasonal changes are accurate.

5. <u>Class discussion: Initial look at the data</u>   10-15 min

a. Have students **download/open the file** and begin to explore using Excel: https://stemcoding.github.io/solar/solar1.csv   and clicking

b. Visualize the data. First, highlight the data (variety of ways to do this), then click insert chart. Can ask students for ideas about what visualization to try first. Basic scatterplot or line graph of **all raw data** as a starting point.

- **Discuss as a class**:
  What do you see in this plot? What can you *not* see in this plot?
  What do you think about this plot?
  Is this a good or bad visualization?
- This is a fine starting point, but can't tell a very convincing story; We can see the seasonal changes, but not much else because there is so much variability. **Discuss the meaning of "noise"** in data.
- **Ask for ideas** on how to proceed. Hopefully students will suggest breaking up the data in some way.
- If a student suggests a linear **trendline**, show how to do this. See that the slope is very close to zero. Is this convincing enough?

6. Got noise? Averaging or summing data can help you see trends	~10 min

Here are data from a DIFFERENT solar array. Look how they summed up the data from each month to make it easier to see the trends.

Multiple Year Energy Report

Columbus, OH

This report provides up to five years energy production ending with the year 2021.

Month	2017	2018	2019	2020	2021
January	338 kWh	392 kWh	365 kWh	318 kWh	337 kWh
February	463 kWh	328 kWh	380 kWh	424 kWh	392 kWh
March	698 kWh	755 kWh	606 kWh	546 kWh	839 kWh
April	759 kWh	730 kWh	699 kWh	710 kWh	775 kWh
May	778 kWh	882 kWh	772 kWh	753 kWh	793 kWh
June	876 kWh	765 kWh	799 kWh	836 kWh	738 kWh
July	817 kWh	863 kWh	905 kWh	869 kWh	784 kWh
August	844 kWh	801 kWh	825 kWh	784 kWh	740 kWh
September	744 kWh	625 kWh	885 kWh	782 kWh	767 kWh
October	616 kWh	623 kWh	701 kWh	522 kWh	548 kWh
November	495 kWh	342 kWh	469 kWh	569 kWh	539 kWh
December	391 kWh	347 kWh	395 kWh	287 kWh	332 kWh
Total	**7,818 kWh**	**7,452 kWh**	**7,801 kWh**	**7,401 kWh**	**7,584 kWh**

What do you notice? What do you wonder?

- Notice that the least solar energy generation is usually January.
- Notice that the most solar energy generation is usually in June or July
- Notice that the total energy from the year fluctuates but there is also a hint of a decline in solar cell efficiency (for example the most solar energy generation is in the first year)
- How much of the seasonal change in energy generation can be explained by changes in the number of hours of sunlight?

7. Summing up spreadsheet data with SUM

Create a new spreadsheet. Don't bother importing the solar data yet, let's try something first.

Learn how to use the SUM function: Create a column of numbers 1-10. In the cell below, use the SUM function to add up all the cells. =SUM(). Highlighting to select cells vs. Typing in range. Go over highlighting and keyboard shortcuts.

Sum up the numbers 1-10 on a calculator or piece of paper and check that your result is the same.

8. Reformat the data

Create a new spreadsheet. Go back to the first spreadsheet with the solar cell data and highlight the first year of data. Press Control + X to cut out this data or Control + C to copy this data to the clipboard. Then use Control + V to paste it into the new spreadsheet. Use Column A for the dates. Use Column B for the first year of data, Columbus C for the second year of data, etc.

This table should have a new row for each day of the year. So each column should be 365 or 366 cells tall if one is using the first cell to give a title to the column.

Note: On Macs and iPads you will use the "Command" key instead of the "Control" key

Question: What is the difference between Cut, and Copy?

9. Use SUM with the solar data

We can try to generate a similar table as the example in 6. using SUM()

Make a decision whether to sum up the energy generation by month, by every two months, or by every three months (in other words every quarter).

What are the advantages / disadvantages of summing over 1 month? 2 months? 3 months?

The procedure will depend on your choice, but for example if you chose to sum by month, you will need to scroll to January 31 and click the leftmost part of the spreadsheet to highlight the entire row for January 31. Then right click and choose "Insert 1 row below". Use this row to calculate the Sum of the energy generation in the previous 31 days. Do this for all years that data is available.

10. Sum up the yearly energy generation

Use the bottom row of the spreadsheet with your table to calculate the total solar energy generation during the one year of data that is highlighted in that column. Bear in mind that SUM(B2:B366) is not going to work because this will now include rows where we have summed up the data. What should we do instead? One option is to use SUM but instead of a range of cells, use it with a few specific cells.

11. Look for evidence that solar panel efficiency is declining or not

It is well known that solar panels decline in efficiency over time. This means that for the same amount of solar energy that is shining on the panels, less electricity is being generated. Is there evidence in this data set for a decline in efficiency?

Create a scatterplot of total energy generation versus year. Don't add a trendline just yet. Look at the data and discuss whether it looks like there is a decline.

12. Use a linear regression to estimate the rate at which the efficiency is declining. Do you get a negative slope?

Click Add Chart Element, at the bottom of the list there will be a trendline option. Click "More Options" to get the full tab of trendline options. Tell students to select linear. Mention that it is usually good to choose the linear option before trying other models.

Towards the bottom of the trendline tab there is an option to show the equation. Click that box to show the best fit equation on the screen.

13. Assuming a linear model, in how many years will the efficiency decrease by 20% compared to what it was during the first year that the solar panels were installed?

14. Knowing what you know about how the efficiency decreases, when should you replace the solar panels? Assume that it costs $15,000 to replace the solar panels and that the cost of electricity is $0.15 per kWh.

15. Compare the month with the largest energy generation to the month with the lowest energy generation. Can the hours per day of sunlight (discussed in steps 2-3) explain the difference in

energy generation? Averaging is important here because the energy generation does fluctuate because of the weather.

16. Choose data from another solar array to compare to solar1.csv You can find all the available data at this link:

https://github.com/stemcoding/stemcoding.github.io/tree/master/solar

Try to choose a data set that no one else near you is looking at.

Note: solar2.csv and solar3.csv are both located Columbus, Ohio

Here are some questions to think about:

Does the solar array data you are looking at now generate more or less energy than the data from solar1.csv?

Is there evidence in the new solar data that the energy generation is decreasing from year to year? Is it decreasing faster or slower than the energy generation from solar1.csv?

Based on your results, how long should you wait before replacing the solar panels? You can assume $15,000 to replace the panels and $0.15 kWh or you can assume something different as long as you describe your assumption clearly.

What other questions might you try to answer?