## Title: Solar energy installations

Please refer to the online lesson plan on the DT Hub to access all website links and additional resources.



Source: Pixabay

Year 7-8: This lesson uses data about solar energy installations to investigate data analysis. The dataset shows how many solar systems were installed, in each postcode, from 2001 to 2016. It is a useful way to understand how to explore and characterise datasets as well as to explore the use of data in the media. The media reporting of this dataset makes a classic mistake, doing a simple analysis that gives a misleading conclusion. This is a rich dataset that offers a range of options for exploration. This lesson was devised by Linda McIver, Australian Data Science Education Institute.

Learning intentions

* Analyse data about solar energy installations.
* Save, store and access datasets and use data cleaning techniques
* Sort and filter datasets to make sense of data to answer particular inquiry questions
* Visualise data using online mapping software or as a chart such as a histogram or column graph.

## Suggested steps

## Learning hook

Solar installations seem to be popular in domestic housing. Why do people install solar systems, and is there an upward trend in installation across Australia?

View the video Renewable Energy 101: National Geographic (3 minutes) and discuss how renewable energy can address climate change.

Alternatively, view the video How Do Solar Panels Work? (5 minutes) and discuss the impact of solar panels on household CO2 emissions.

Provide students with a dataset found on Postcode data for small-scale installations.Note: there is a range of datasets of this web page to accommodate students with differing levels of spreadsheeting skills.

In particular, use this dataset provided in csv format:

* Postcode data for small-scale installations – SGU-Solar

### Step 1: Dataset download and data cleaning

Download the dataset and open the csv file in a spreadsheet.

Look at the postcodes.

In most spreadsheets this data will show postcodes with one, three and four digits, although postcodes in Australia all have four digits. What has happened to these postcodes?

|  |
| --- |
| This is an example of a spreadsheet being programmed to modify the data (hiding things the programmer believes the user does not need to know about – in this case, leading 0s). Mathematically speaking, there’s no difference between 0, 00, 000, and 0000. They all just mean 0. So spreadsheets (and other software) tend to remove the leading 0s, which means postcode 0 is actually 0000, 200 is actually 0200 and so on.  This is an important part of data cleaning. Sometimes you have to convert the data back to its original form to fix errors that spreadsheets and other software introduce in an attempt to be ‘helpful’. Format cells and select ‘custom’ from the drop-down menu and type in 0000 to allow for a 4-digit postcode. |

### Step 2: Is this dataset useful?

Now let’s look at the first two columns in the dataset. The first is historical installations from 2001 to 2016. The dataset doesn’t seem to have any data for installations prior to 2001, but that’s not because there were no installations. What might be the reason?

|  |
| --- |
| It turns out it’s because 2001 was when the Government introduced the mandatory renewable energy target and began tracking renewable energy. |

Students may ask, ‘How many solar panels are actually operating currently?’

|  |
| --- |
| We can’t exactly tell that from this dataset. This data tracks installations. It doesn’t track people getting rid of their solar panels, or the panels ceasing to work. Installations are a reasonable, but not perfect, measure of how much solar we have.  This question can springboard a useful conversation about the data we want, versus the data we have, and how many data studies work with flawed or missing data, simply because it’s all that is available. |

**Class discussion:** How might we find out how much solar is actually operating right now in Australia? What organisations might have that information?

### Step 3: Sorting the data

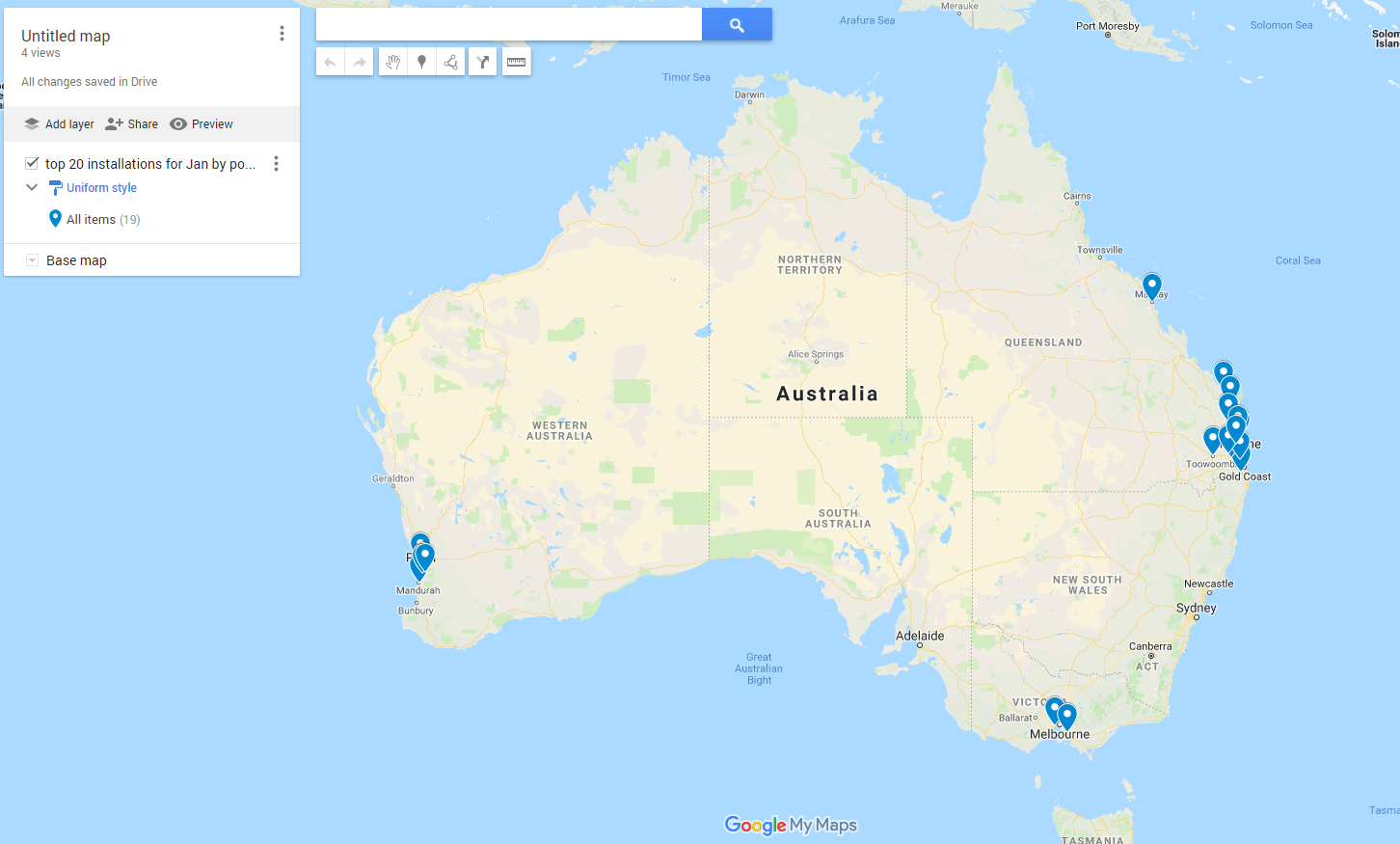
Look at the first column in the dataset. Having it sorted by postcode is logical, but not terribly interesting. Let’s look at the top 20 postcodes – to do that, we can sort the entire table by the second column (how many installations happened between 2001 and 2016), in descending order. In other words, put the largest values up the top.

#### Solar Installations sorted by Previous Years column, largest to smallest



|  |
| --- |
| A quick glance shows us that the majority of the top 20 postcodes start with a 4, indicating they’re in Queensland. (To check postcodes refer to a postcode site.)  The top postcode, 4670, covers 53 regions, including Bundaberg. There’s a surprisingly large gap between the top postcodes and the bottom of the top 20, which is interesting. Most of the postcodes in this list that aren’t in Queensland are in Western Australia, except for 3029, which is west of Melbourne, around Hoppers Crossing, and 3977, which is south-east of Melbourne, in the Cranbourne area.  There’s a rich conversation to be had around why these suburbs have so much more solar than other places in Victoria. Toorak, for example, a famously wealthy suburb, comes in at 1701 on the list. Areas with a lot of new housing are more likely to have solar, as it gets put in when the house is built as a way to increase the energy rating of the house.  This is a topic worth exploring! You don’t have to know all the answers, as it’s an opportunity for the students to research and explore, and come up with their own theories for why it might be the case. |

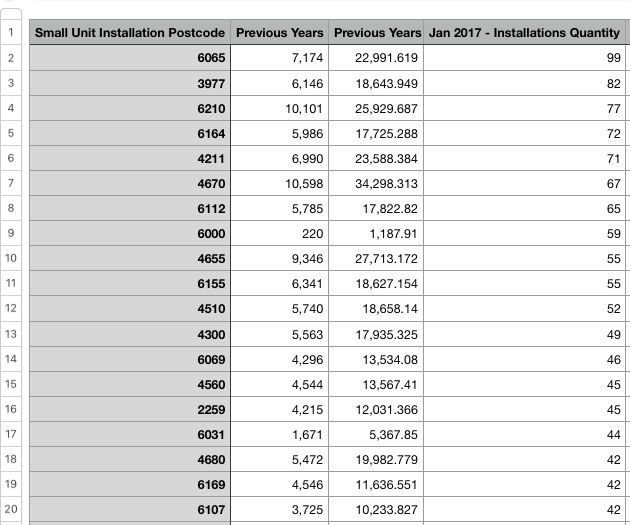
#### Data visualised in Google MyMaps: 20 postcodes with highest number of installations



### Step 4: Data analysis

Let’s look at column 4: solar installations in January 2017. How different are the top 20 if you sort the whole table by this column?

#### Solar Installations sorted by January Installations



|  |
| --- |
| Now WA scores better, and the rest is still largely over to Queensland, except for one Victorian postcode (Cranbourne area again), and this time one NSW representative.  Why do WA and Queensland do so well on both historic and recent measures? This is an opportunity to explore the politics and have your students find out what incentives there are to install solar in those states. Could it be due to solar feed-in tariffs, government incentives, or home energy rating requirements? |

### Step 5: Data calculations

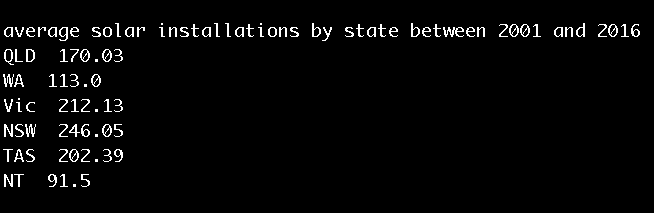
Calculate the average solar installations per postcode per state. You can do this by sorting the data by postcode, and manually copying and pasting each state into a separate sheet, or you can write a Python script to sort the data into a separate file for each state using the first digit of the postcode.

|  |
| --- |
| The media misreported this data by using a naive analysis technique – just sorting the data. It is an easy mistake to make. Despite WA and Queensland dominating the top 20 postcodes list, when you calculate the average for each state, you get a quite different picture.  For examples of the media ‘analysis’ go to:   * Australia’s top 10 solar-friendly postcodes revealed * Queensland city tops national solar panel list   ‘Queensland is leading Australia’s rooftop solar boom with eight of the country’s top 10 postcodes for installations in the Sunshine State, according to the new Clean Energy Australia 2018 report.’  The Sydney Morning Herald, June 16, 2018 |

Compare the media analysis with the averages by state.

Check that students’ results come out the same as in the example below: (Note: in the example ACT has been included in NSW postcode data as it makes the postcode handling easier).

#### Python program output: Average solar installations by state



|  |
| --- |
| Interestingly, this shows that the state that dominates the top 20 doesn’t perform as well when you average over all of its postcodes, so there is another rich conversation to be had about different ways of ranking data outcomes, and how you can characterise data in accurate but misleading ways.  The media reported this data saying that Queensland and WA were the best for solar. However, although they had the highest-ranking postcodes, as states they ranked very low. |

### Step 6: Data – interrelated

Students can continue to explore the different columns. Alternatively, as a further challenge look at how the columns are related. For example, are postcodes with a lot of historical solar installations also likely to have a lot of recent ones? You can do that roughly by eye, simply by looking at whether the top twenty, when sorted by those two columns, is similar or very different, or you can use the correlation function to find out whether the columns are correlated.

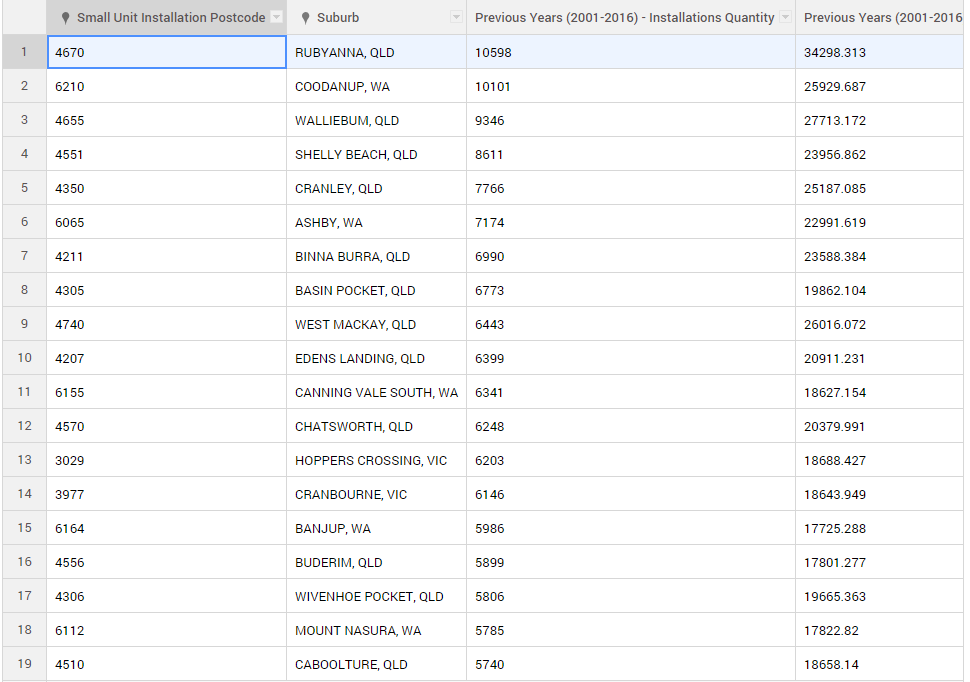
### Step 7: Visualise and aggregate data

Students can visualise their data. Students first consider the aim of their visualisation.

Use question prompts such as:

* What is it you want to highlight about the data?
* For example, do you want to highlight January Installations, the states that do best with solar installations in total, or the change over time?
* Which column or summary statistic shows that point best?
* What is the best graph type for that information?
* This is a useful tutorial on graph types.
* How much data do you want to display?
* Trying to graph all of the information will make your graph cluttered and your graph elements hard to read.
* Can you aggregate the data in some way – for example, by state instead of by postcode, or show the number of postcodes above 20,000, the number between 15,000 and 20,000, the number between 10,000 and 15,000. This is a histogram.
* How would the data display on an online map?
* The data could be plotted on an online map such as Google MyMaps.
* Data could be collated in tabs on a csv spreadsheet for each state and territory and uploaded. The data can be selected and deselected to visualise data.
* In order to effectively upload data using online mapping software, the location needs to be identified in some way, often using latitude and longitude. In our dataset we have postcode data. However, for that data to be plotted more accurately, adding the suburb information helps [Note: this can be time consuming for a large dataset.]

#### Image: Dataset with suburb name and state/territory added



Another technique would be to colour a map by number of solar installations, say, bright red for >9000 and becoming paler for each drop of 1000. This would be rather time-consuming given that there are 2795 postcodes listed, so this is an opportunity to consider aggregating your data and colouring by state. It’s a great example of not needing complex technical skills to explore a dataset. This is called a choropleth map. You can use National Map to do this.

## Resources

* Australian Government Clean Energy Regulator: Postcode data for small-scale installations
* Solar PV Maps and Tools

Understand the Australian solar PV market with live generation data, historical maps and animations, and tools to explore rooftop PV potential and per-postcode market data.

* National Map

National Map is an online map-based tool to allow easy access to spatial data from Australian government agencies. Mapping data can be added, but needs to include latitude and longitude data.

* Find a postcode

This postcode finder from Australia Post is a quick and easy way to search and check postcodes for all suburbs and locations around Australia.

* Postcodes in Australia

Provides postcode ranges for Australian states and territories.

* Graphing tutorial

Useful tutorial on graph types

## Curriculum links:

|  |  |
| --- | --- |
| Links with Digital Technologies Curriculum Area. | |
| **Year** | **Strand** |
| **7-8** | Acquire data from a range of sources and evaluate authenticity, accuracy and timeliness (ACTDIP025)  Analyse and visualise data using a range of software to create information, and use structured data to model objects or events (ACTDIP026) |

|  |  |
| --- | --- |
| Links with Science Curriculum Area. | |
| **Year** | **Strand** |
| **7** | Science Inquiry: Analyse data and information to describe patterns, trends and relationships and identify anomalies (AC9S7I05 )  Science Inquiry: Write and create texts to communicate ideas, findings and arguments for specific purposes and audiences, including selection of appropriate language and text features, using digital tools as appropriate (AC9S7I08 ) |

|  |  |
| --- | --- |
| Links with Mathematics Curriculum Area. | |
| **Year** | **Strand** |
| **7** | Statistics: Create different types of numerical data displays including stem-and-leaf plots using software where appropriate; describe and compare the distribution of data, commenting on the shape, centre and spread including outliers and determining the range, median, mean and mode (AC9M7ST02 ) |