

# Data Visualization - Assignment 4

Simran Singh

2/4/2022

## Description of Task

For this assignment we are tasked with importing a dataset into a R dataframe. The data I will use is the solar panel installations registry from the New Jersey Board of Public Utilities' Clean Energy Program. This data is publicly available.

This data comes in .csv format so it will be fairly straightforward to import into R. We will use the SRP program registry, as it is the most populated one. The TI program is a more recent program (starting in 2021), but has fewer registrations under it. In a proper analysis, we would consider both programs alongside the legacy ADP program.

I've uploaded the .csv file to this github repository along with this Markdown file:

<https://github.com/psimran-singh/Data-Viz-Seminar-Assignment-4>

## Importing the data

I will import the data into an R dataframe using `read.csv()`.

```
setwd("~/GitHub/Data-Viz-Seminar")  
SRP_Data <- read.csv("SRP_Data.csv")
```

## Aggregating the data

Since this data is very currently very granular (at individual installation level), I will aggregate it by grouping by the county code variable and summing up total system size for county.

I will then arrange the total system sizes in descending order within each county code.

And finally I will rename the columns so they all follow a similar naming convention.

```
SRP_Data_Agg <- SRP_Data %>%  
  group_by(County.....Code) %>%  
  summarise(total_system_size = sum(Calculated.Total.System.Size)) %>%  
  arrange(desc(total_system_size), .by_group=TRUE) %>%  
  rename('county_code' = County.....Code)
```

## Plotting the data

I will now plot the data in a histogram to get an idea of how it looks.

