

**Step 2-A:**

**Exploring Data**

# After this video you will be able to..

- Explain the importance of exploring data
- Identify methods to perform preliminary analysis of your data

Big Data Engineering

Computational Big Data Science

ACQUIRE

**PREPARE**

ANALYZE

REPORT

ACT

Step 2-A: Explore

Step 2-B: Pre-process





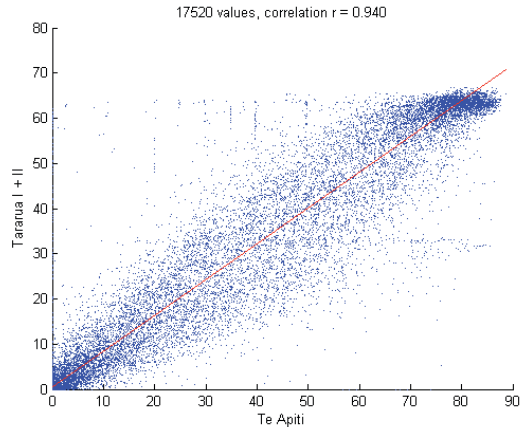
# Why Explore?

Goal: Understand your data



# Why Explore?

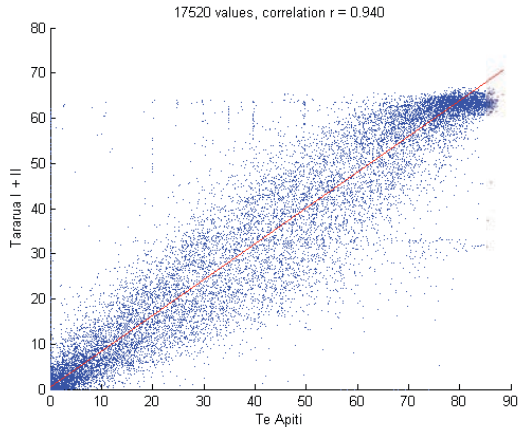
## Correlations



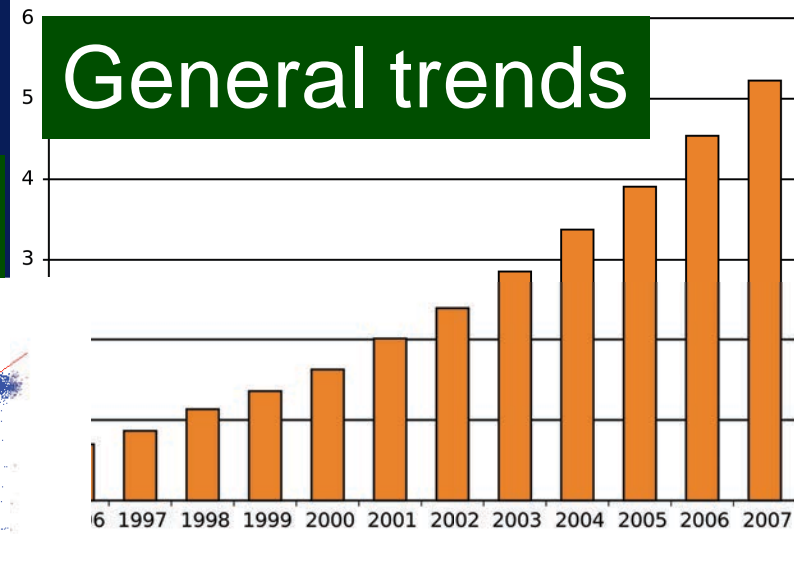


# Why Explore?

## Correlations



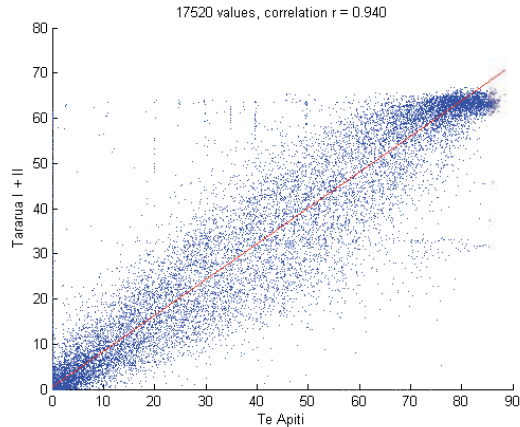
## General trends



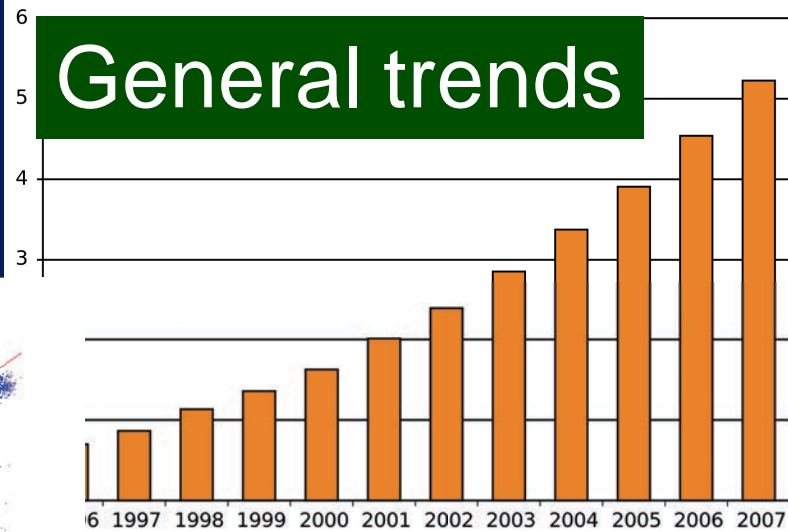


# Why Explore?

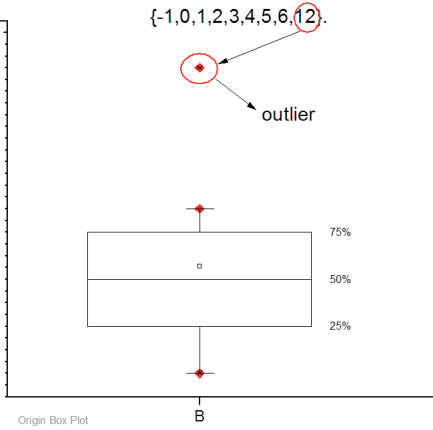
## Correlations



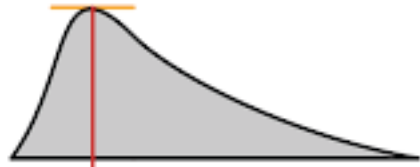
## General trends



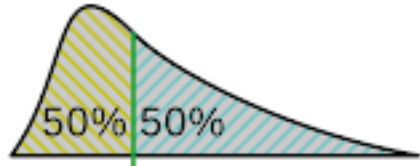
## Outliers



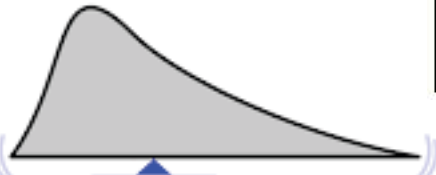
# Describe Your Data



mode



median



mean

Mean	Median
Mode	Range

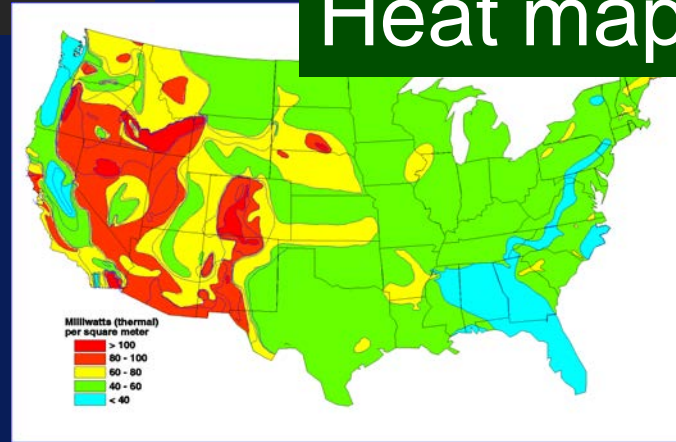
*Statistics*



# Visualize Your Data

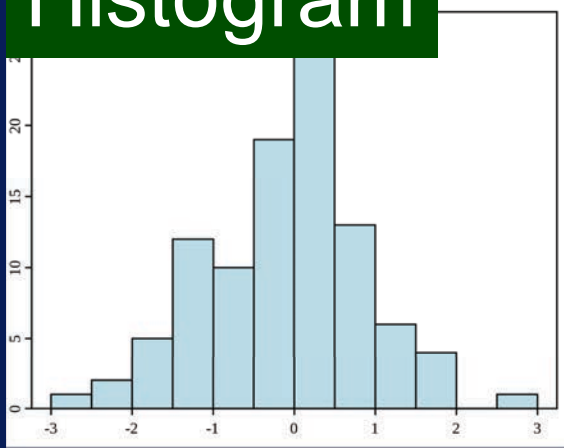
# Visualize Your Data

Heat maps

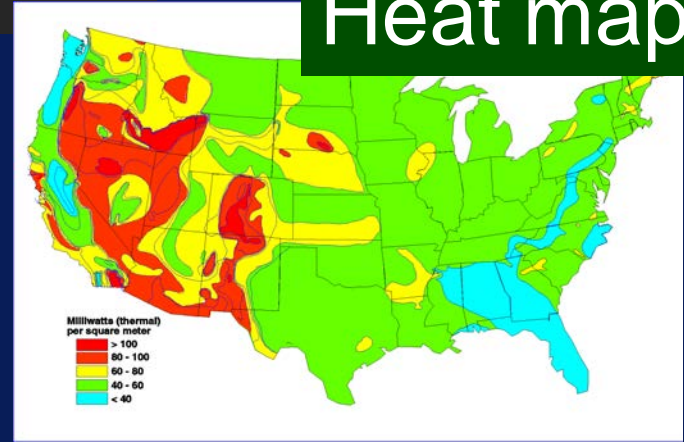


# Visualize Your Data

## Histogram

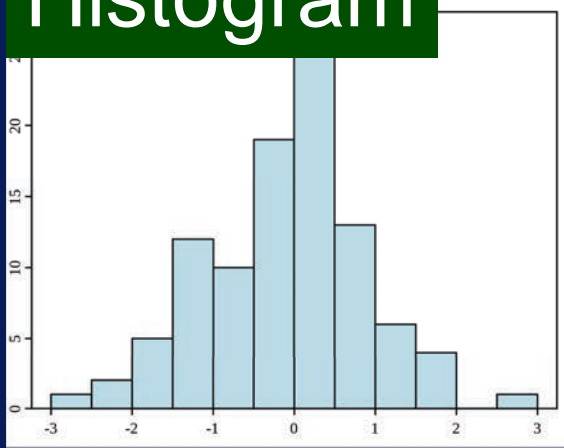


## Heat maps

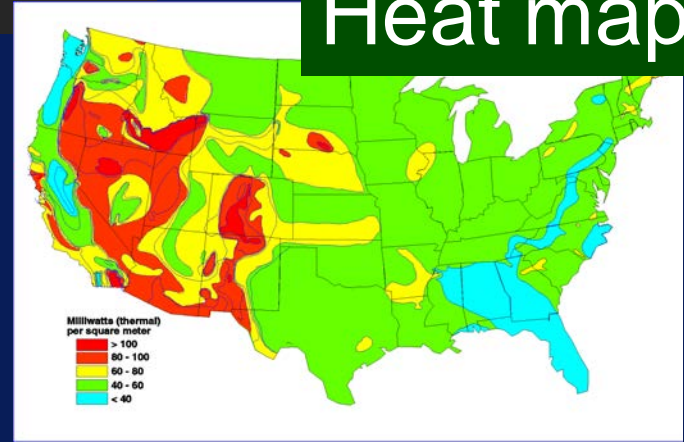


# Visualize Your Data

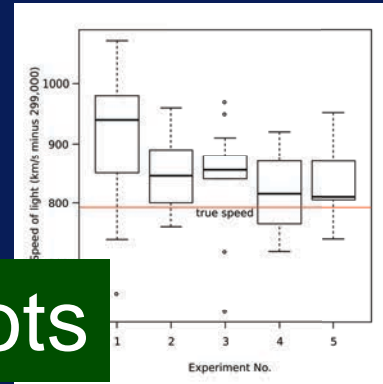
## Histogram



## Heat maps

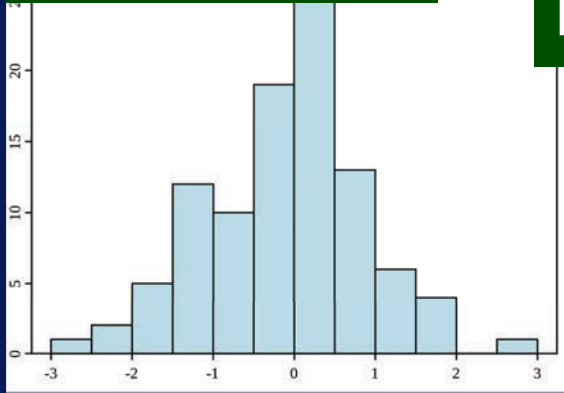


## Boxplots

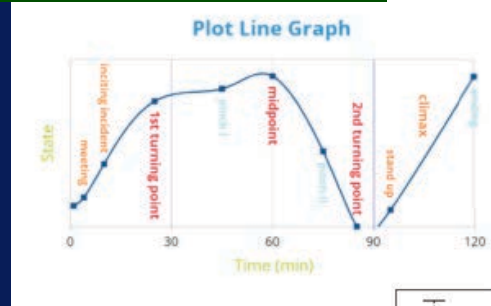


# Visualize Your Data

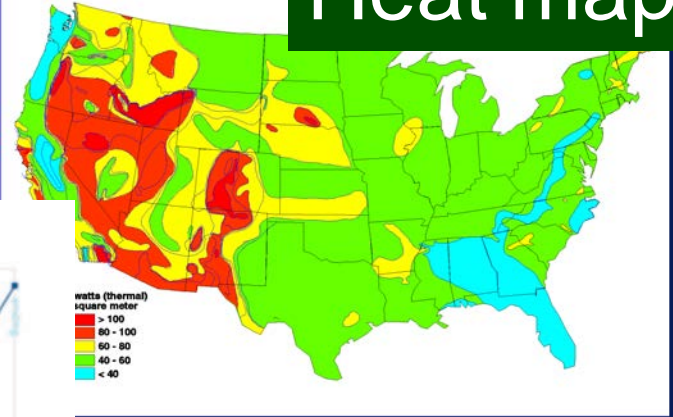
## Histogram



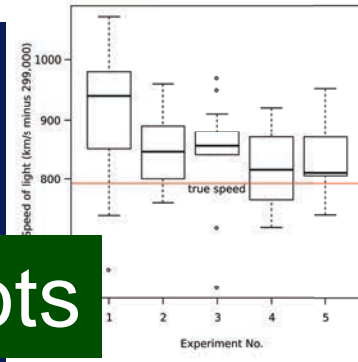
## Line graphs



## Heat maps

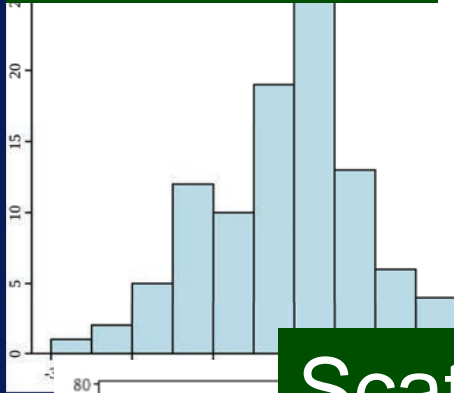


## Boxplots



# Visualize Your Data

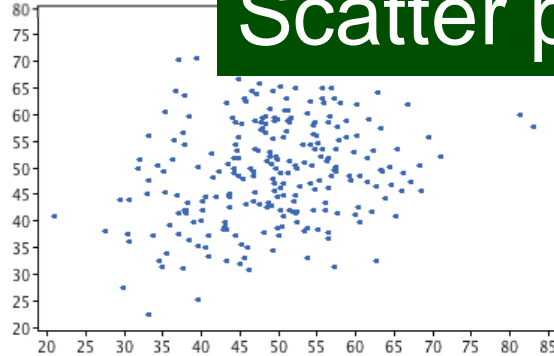
Histogram



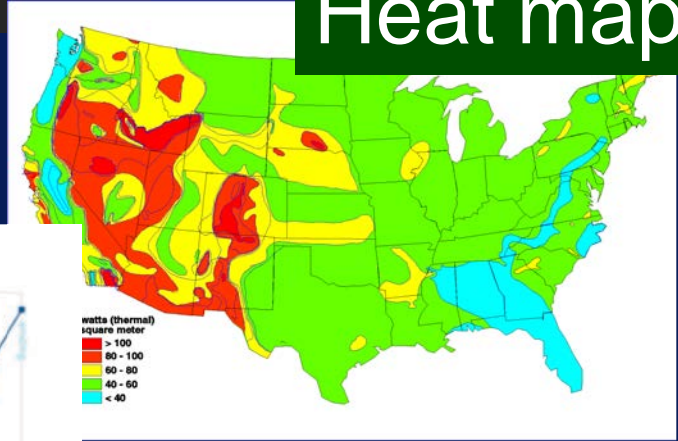
Line graphs



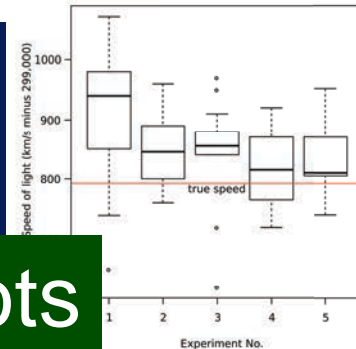
Scatter plots



Heat maps



Boxplots



**Data  
Exploration**



```
graph LR; A[Data Exploration] --> B[Data Understanding]; B --> C[Informed Analysis]
```

A flowchart on a dark blue background showing a three-step process. The steps are represented by green rounded rectangles with white text. The first step, 'Data Exploration', is at the bottom left. A blue arrow curves upwards and to the right to the second step, 'Data Understanding', which is in the center. Another blue arrow curves upwards and to the right from the second step to the third step, 'Informed Analysis', which is at the top right.

**Data  
Understanding**

**Informed  
Analysis**