

**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





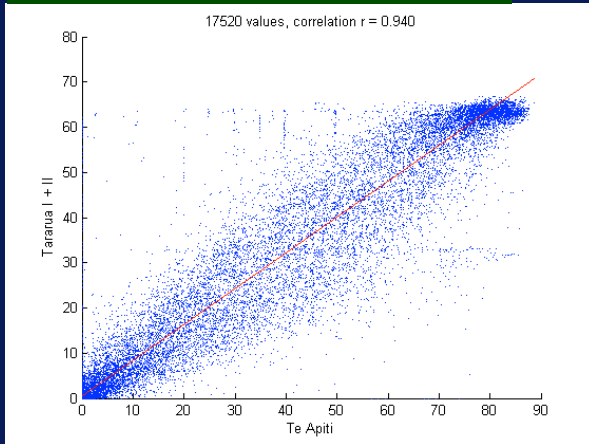
# Why Explore?

Goal: Understand your data



# Why Explore?

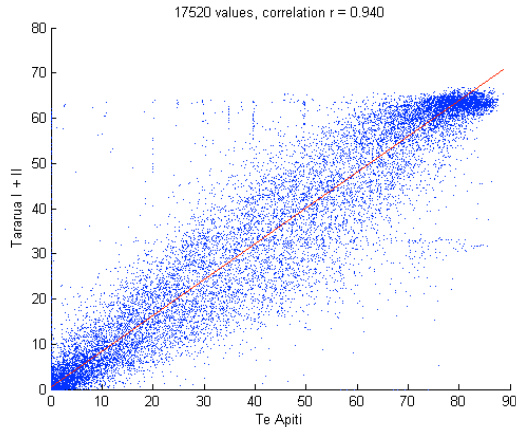
## Correlations



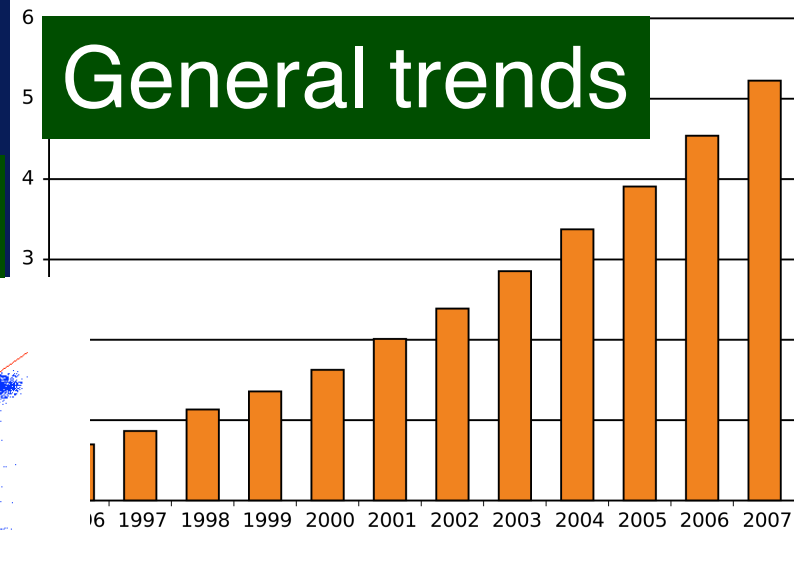


# Why Explore?

## Correlations



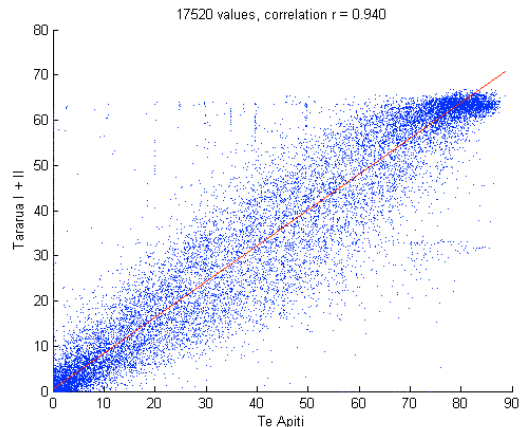
## General trends



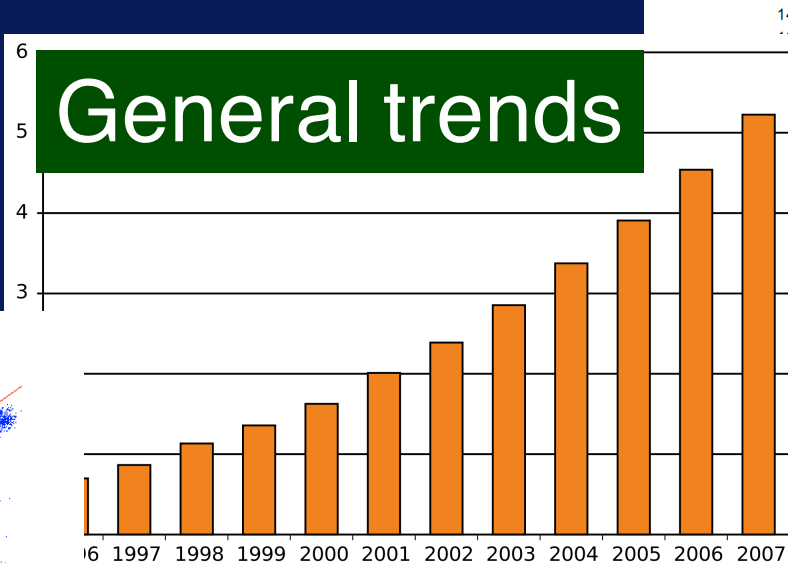


# Why Explore?

## Correlations



## General trends

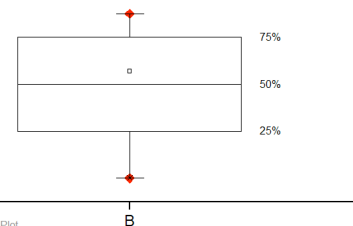


## Outliers

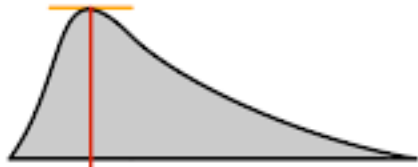
$\{-1, 0, 1, 2, 3, 4, 5, 6, 12\}$



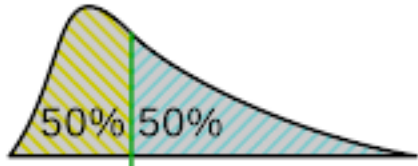
outlier



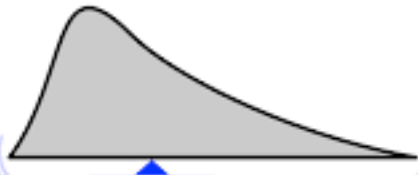
# Describe Your Data



mode



median



mean

Mean	Median
Mode	Range

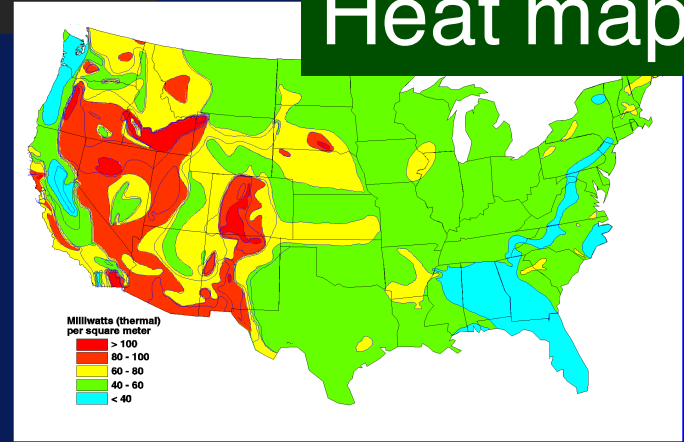
*Statistics*



# Visualize Your Data

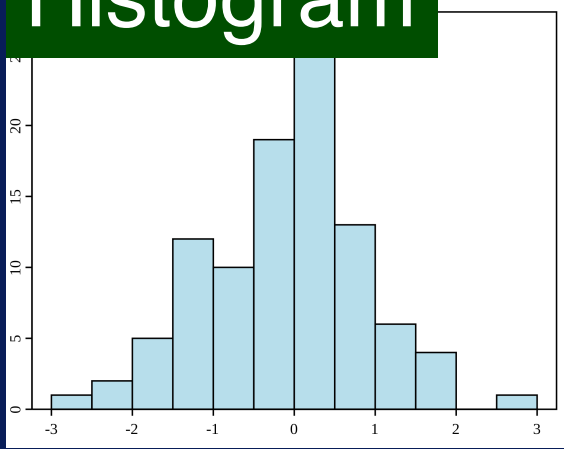
# Visualize Your Data

## Heat maps

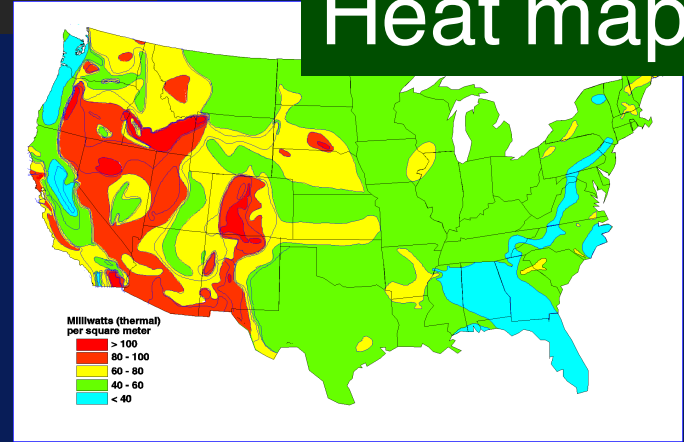


# Visualize Your Data

## Histogram

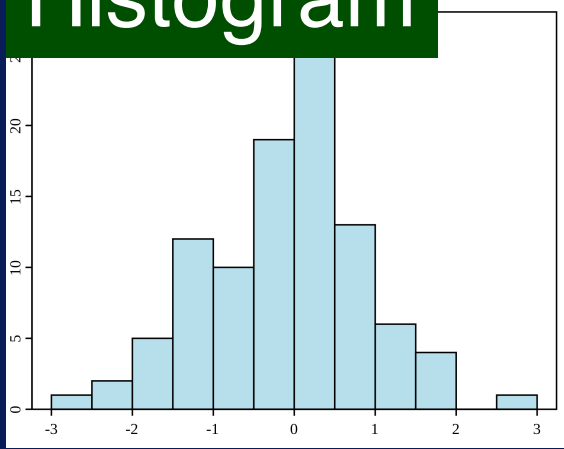


## Heat maps

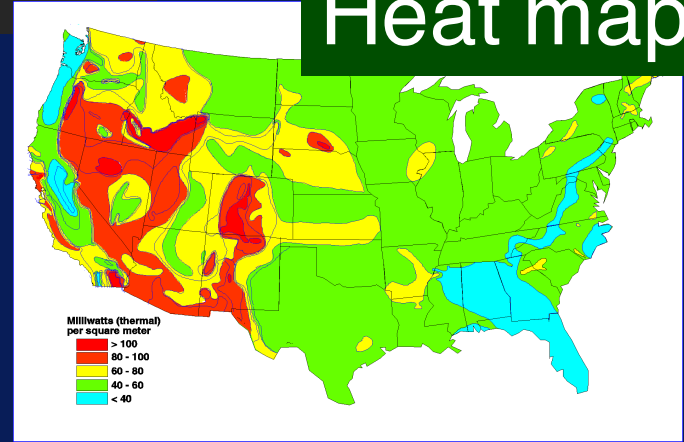


# Visualize Your Data

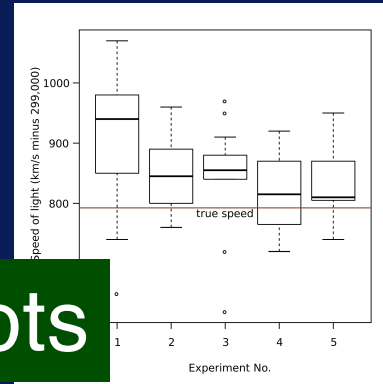
## Histogram



## Heat maps

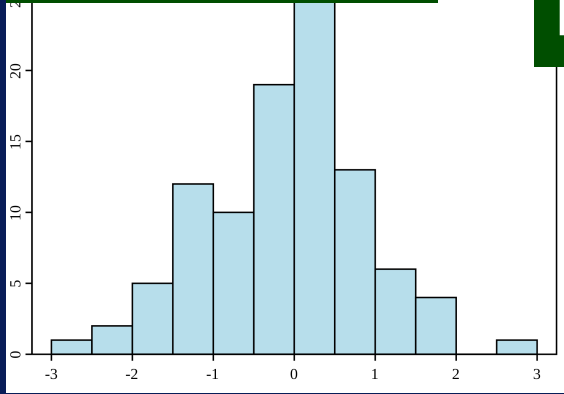


## Boxplots



# Visualize Your Data

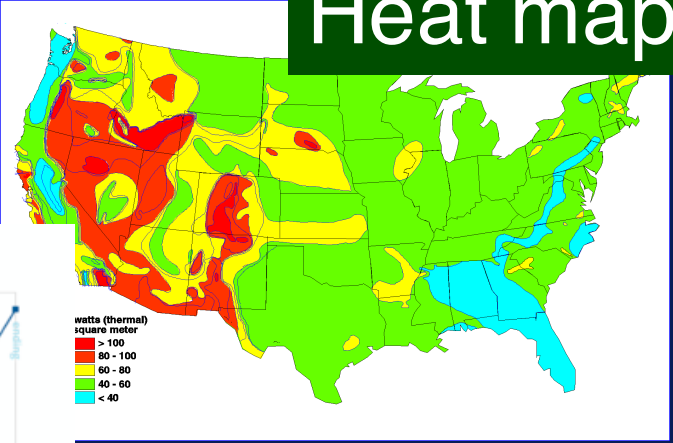
## Histogram



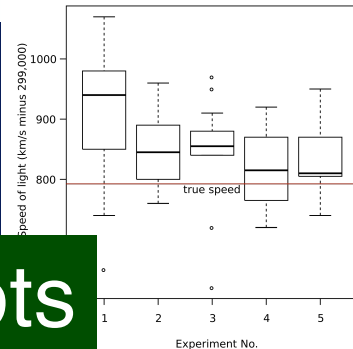
## Line graphs



## Heat maps

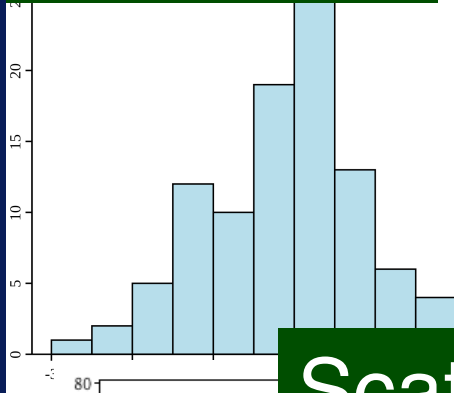


## Boxplots



# Visualize Your Data

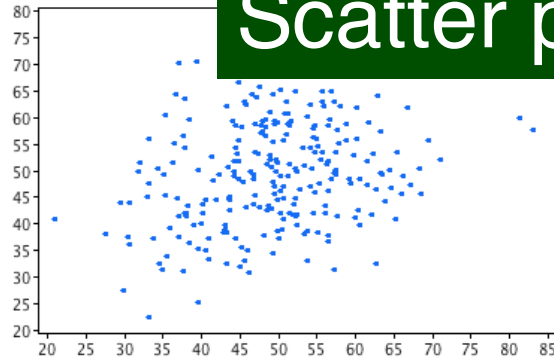
## Histogram



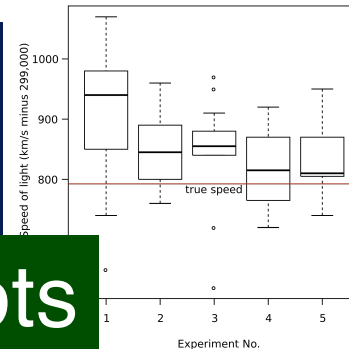
## Line graphs



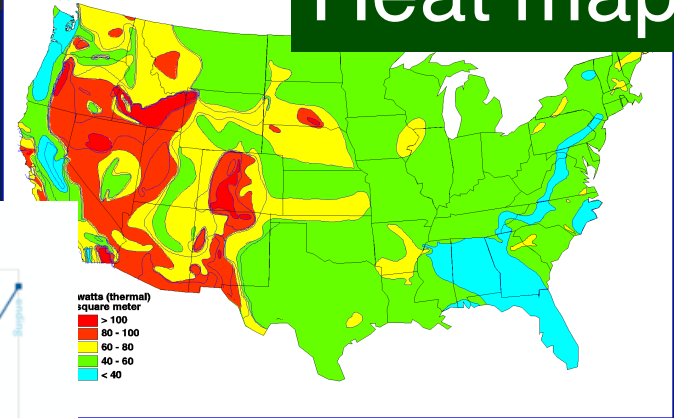
## Scatter plots



## Boxplots



## Heat maps



**Data  
Exploration**



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

A flowchart illustrating a three-step process. It begins with a green rounded rectangle labeled 'Data Exploration' at the bottom left. A blue arrow curves upwards and to the right to a second green rounded rectangle labeled 'Data Understanding' in the center. Another blue arrow curves upwards and to the right from the second rectangle to a third green rounded rectangle labeled 'Informed Analysis' at the top right. The background is a solid dark blue.

**Data  
Understanding**

**Informed  
Analysis**