



Discrete  
Structures:  
CMPSC 102

Oliver  
BONHAM-  
CARTER

Let's Discuss

Describing  
Data

Setup VENV

Quantiles and  
Medians

Correletion

Solutions

# Discrete Structures: CMPSC 102

Oliver BONHAM-CARTER

Spring 2024  
Week 12



# Let's Discuss

Discrete  
Structures:  
CMPSC 102

Oliver  
BONHAM-  
CARTER

Let's Discuss

Describing  
Data

Setup VENV

Quantiles and  
Medians

Correletion

Solutions

## Key Questions

How can I describe data using statistical tools such as correlation, variance, standard deviation and others?

## Learning Objectives

To **remember** and **understand** some concepts about **plots**, and the code used to make them from `matplotlib` and other libraries.

# Where Are We Now?

## Saha's Book

Discrete  
Structures:  
CMPSC 102

Oliver  
BONHAM-  
CARTER

Let's Discuss

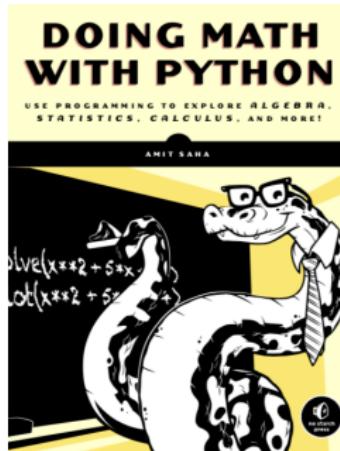
Describing  
Data

Setup VENV

Quantiles and  
Medians

Correletion

Solutions



### Saha, Chapter 3: Describing Data with Statistics

- Basic statistics
- General: Frequencies, Mode, Median, Range, etc
- Complex: Correlation, Quantiles (more to come!)

# Setting Up Virtual Environment

Discrete  
Structures:  
CMPSC 102

Oliver  
BONHAM-  
CARTER

Let's Discuss

Describing  
Data

Setup VENV

Quantiles and  
Medians

Correletion

Solutions

## Installing Software

75%

Create a project directory (outside of lassDocs/)

```
mkdir projects  
cd projects
```

Create virtual environment using Python

```
python3 -m venv myenv
```

Activate *myenv* the virtual environment

```
source myenv/bin/activate # unix  
myenv\scripts\activate # windows
```

Install Dependencies

```
pip install matplotlib numpy pandas plotly
```



# Basic Stats

Check your sandbox/ for example files!

Discrete  
Structures:  
CMPSC 102

Oliver  
BONHAM-  
CARTER

Let's Discuss  
Describing  
Data

Setup VENV

Quantiles and  
Medians

Correletion

Solutions

- See sandbox/ files to experiment with statistical code
  - Examples: Frequencies, Mode, Range, Plotting
  - Be sure to build your virtual environment first to load the libraries!! (See next slide for help)



# Quantiles and Medians

Check your sandbox/ median example: "median.py"

Discrete  
Structures:  
CMPSC 102

Oliver  
BONHAM-  
CARTER

Let's Discuss

Describing  
Data

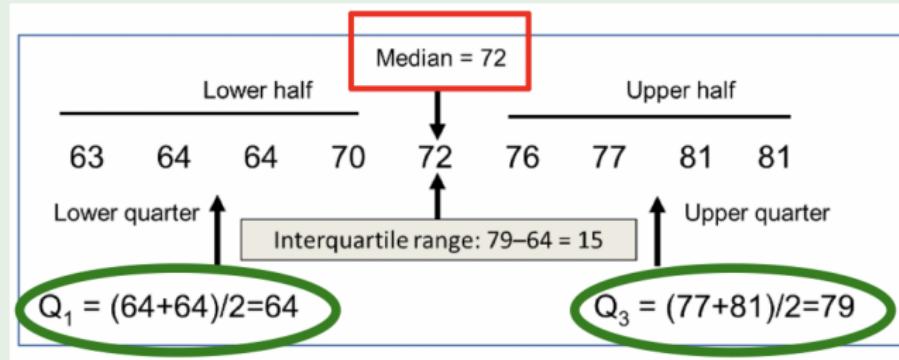
Setup VENV

Quantiles and  
Medians

Correletion

Solutions

- Determine exactly what value “splits” the dataset.
- The  $Q$ ’s denote the four quarters (“Quantiles”)
  - Used to determine specify areas of the sorted dataset
  - $Q_1$  is at first  $\frac{1}{4}$  position of data,  $Q_3$  is the  $\frac{3}{4}$  position in data



# Quantiles and Medians

Check your sandbox/ median example: "median.py"

Discrete  
Structures:  
CMPSC 102

Oliver  
BONHAM-  
CARTER

Let's Discuss

Describing  
Data

Setup VENV

Quantiles and  
Medians

Correletion

Solutions

- You must know if the dataset has an odd or even number of elements

First, arrange the observations in an ascending order.

If the number of observations ( $n$ ) is **odd**:  
the median is the value at position

$$\left( \frac{n+1}{2} \right)$$

If the number of observations ( $n$ ) is **even**:

1. Find the value at position  $\left( \frac{n}{2} \right)$

2. Find the value at position  $\left( \frac{n+1}{2} \right)$

3. Find the average of the two values to get the median.

# Quantiles and Medians

Check your sandbox/ median example: "median.py"

Discrete Structures:  
CMPSC 102

Oliver  
BONHAM-CARTER

Let's Discuss

Describing Data

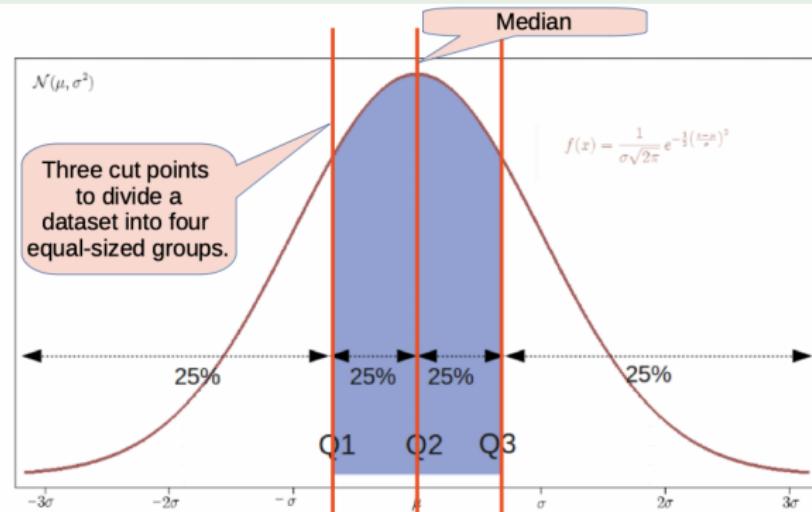
Setup VENV

Quantiles and Medians

Correletion

Solutions

- Determine exactly what value “splits” the dataset.





# Correlation

Check your sandbox/ for example: `correlation.py`

Discrete  
Structures:  
CMPSC 102

Oliver  
BONHAM-  
CARTER

Let's Discuss

Describing  
Data

Setup VENV

Quantiles and  
Medians

Correletion

Solutions

Correlation is a statistical measure that describes the strength and direction of a relationship between two variables. A correlation score ranges from -1 to 1, where:

- -1: Perfect negative correlation
- 0: No correlation
- 1: Perfect positive correlation
- Values between -1 and 1 denote the strength of the correlation, as shown in the example below.

Correlation can be calculated using various methods, one of which is the Pearson correlation coefficient.

# Correlation

Check your sandbox/ for example: `correlation.py`

Discrete  
Structures:  
CMPSC 102

Oliver  
BONHAM-  
CARTER

Let's Discuss

Describing  
Data

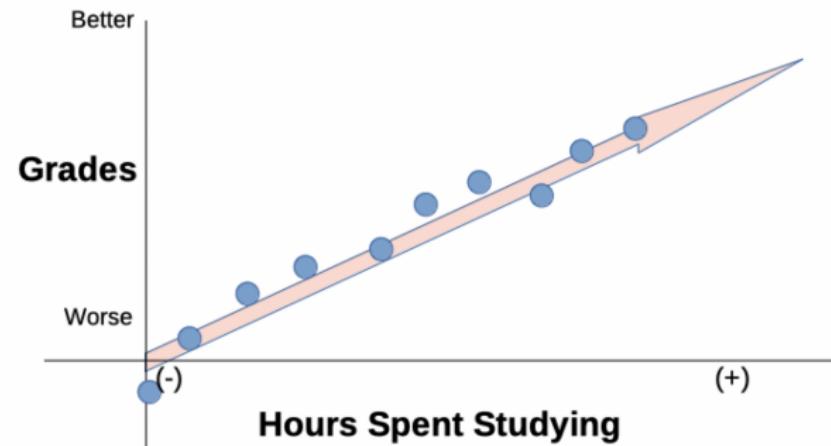
Setup VENV

Quantiles and  
Medians

Correletion

Solutions

- Positive correlation between  $x$  and  $y$



Points lie close to a straight line that has a **positive** gradient.

# Correlation

Check your sandbox/ for example: `correlation.py`

Discrete  
Structures:  
CMPSC 102

Oliver  
BONHAM-  
CARTER

Let's Discuss

Describing  
Data

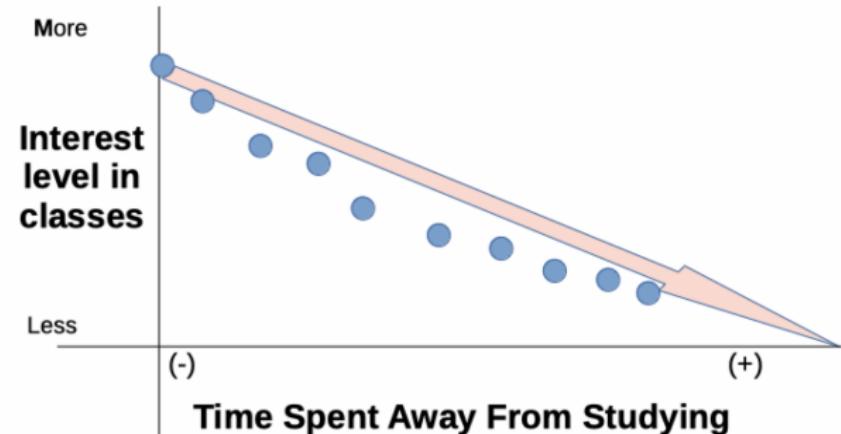
Setup VENV

Quantiles and  
Medians

Correletion

Solutions

- Negative correlation between  $x$  and  $y$



Points lie close to a straight line that has a **negative** gradient.



# Correlation

Check your sandbox/ for example: `correlation.py`

Discrete  
Structures:  
CMPSC 102

Oliver  
BONHAM-  
CARTER

Let's Discuss

Describing  
Data

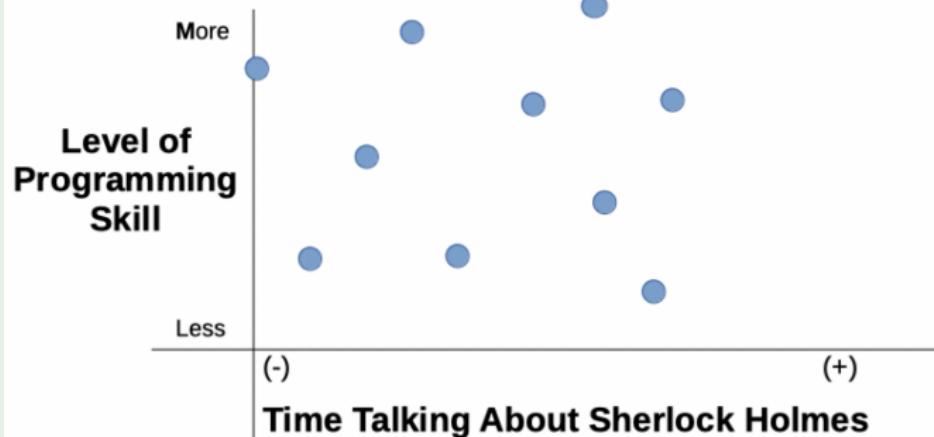
Setup VENV

Quantiles and  
Medians

Correletion

Solutions

- No correlation between  $x$  and  $y$



No pattern exists in the layout of points. :-)



# Correlation: Summary

Check your sandbox/ for example: `correlation.py`

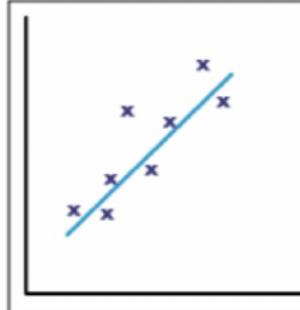
Discrete  
Structures:  
CMPSC 102

Oliver  
BONHAM-  
CARTER

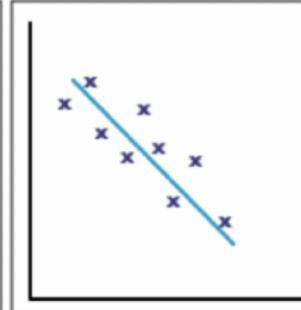
Let's Discuss  
Describing  
Data  
Setup VENV  
Quantiles and  
Medians  
Correletion  
Solutions

## ● In summary

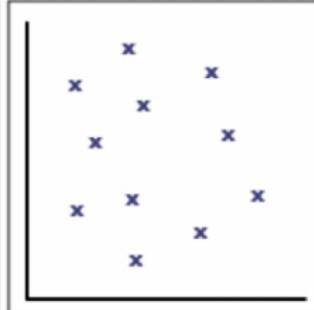
**Positive correlation**



**Negative correlation**



**No correlation**



The points lie close to a straight line, which has a positive gradient.

This shows that as one variable **increases** the other **increases**.

The points lie close to a straight line, which has a negative gradient.

This shows that as one variable **increases**, the other **decreases**.

There is no pattern to the points.

This shows that there is **no connection** between the two variables.

# Creating Solutions

Check your sandbox/ for example files!

Discrete  
Structures:  
CMPSC 102

Oliver  
BONHAM-  
CARTER

Let's Discuss

Describing  
Data

Setup VENV

Quantiles and  
Medians

Correletion

Solutions

