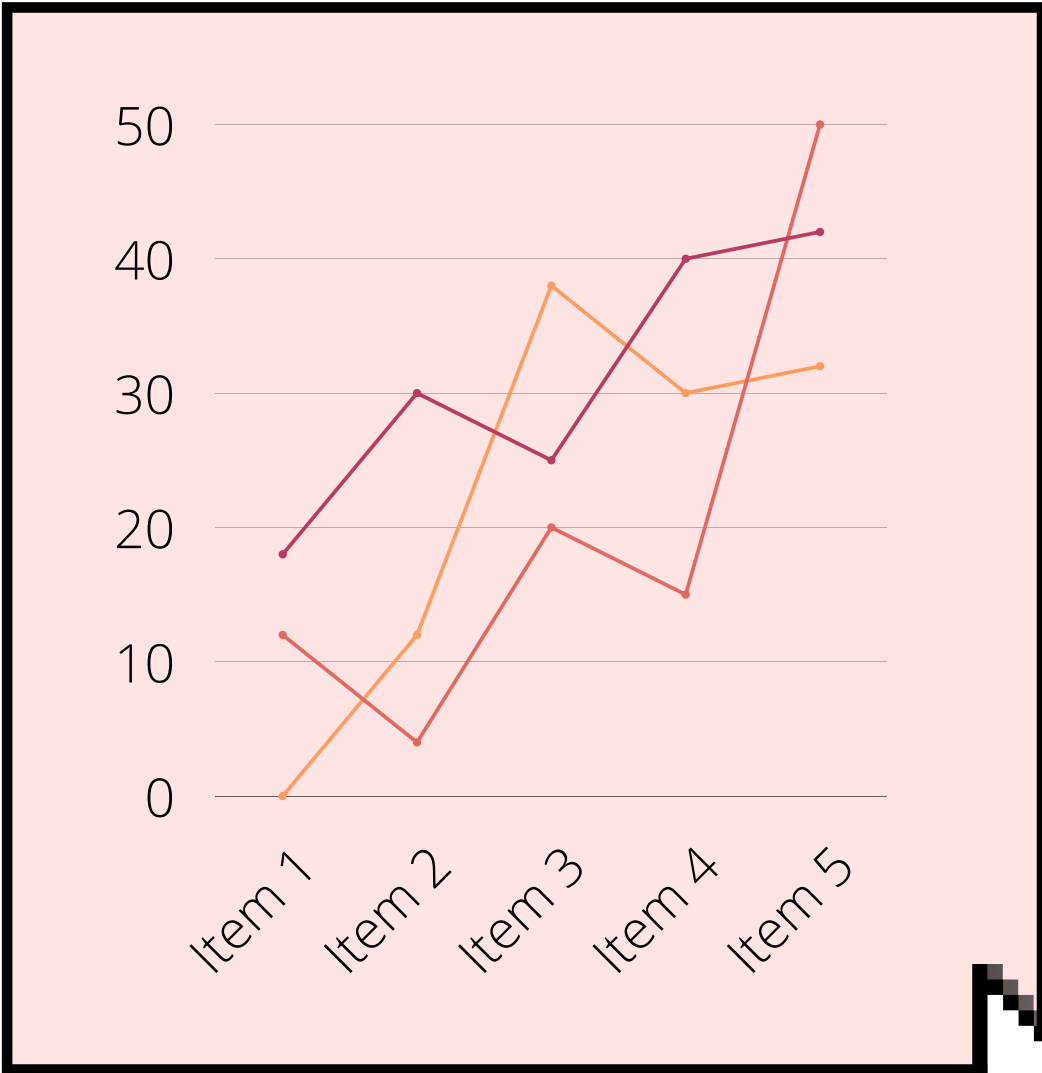
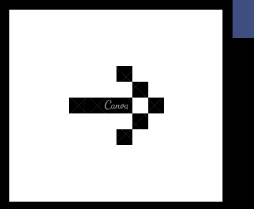


Data Visualisation



CORRELOGRAM

Based on Iris Data Set



Contributors

Astha Mehrotra

180630107029

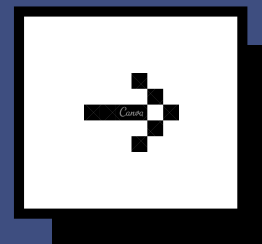
Makwana Krishna Nileshbhai

180630107028

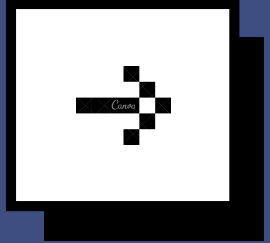
Mahida Dhruvi Hiteshkumar

180630107027

2/17



About Correlogram



Things to note before you jump in

A correlogram or correlation matrix allows to analyse the relationship between each pair of numeric variables of a dataset. The relationship between each pair of variable is visualised through a scatterplot, or a symbol that represents the correlation (bubble, line, number). The diagonal often represents the distribution of each variable, using an histogram or a density plot.

The correlogram is a commonly used tool for checking randomness in a data set.

3/17



Related Charts:

✓ Bubble Chart

for some clarity in 3D

✓ Scatter Plot

Display relationship between 2 numeric variables

✓ Connected Scatterplot

Very close to scatter plot but links data points with segments

✓ Density 2D

best ways to avoid overplotting for big sample size

✓ Line Reference

To keep it free from clutter

Iris flower data set

The Iris flower data set is a multivariate data set

The Data:

The data set consists of 50 samples from each of three species of Iris (Iris setosa, Iris virginica and Iris versicolor). Four features were measured from each sample: the length and the width of the sepals and petals, in centimeters. Based on the combination of these four features, Fisher developed a linear discriminant model to distinguish the species from each other.

The Dataset:

The dataset contains a set of 150 records under five attributes - sepal length, sepal width, petal length, petal width and species.

Data Values :

<https://docs.google.com/spreadsheets/d/1Kc3YI2F6TEF9D2yurz-7gkkTwVMwRkfo8h1ck1kQeAU/edit?usp=sharing>

Code here

HTML

```
<head>
<!-- Load plotly.js into the DOM -->
<script src='https://cdn.plot.ly/plotly-
latest.min.js'></script>
</head>

<body>
<div id='myDiv'><!-- Plotly chart will be
drawn inside this DIV --></div>
</body>
```

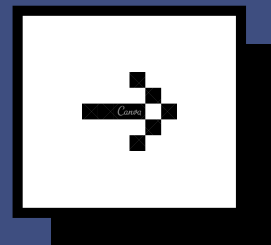
JAVASCRIPT

```
Plotly.d3.csv('https://raw.githubusercontent.com/plotly/datasets/master/iris-data.csv', function(err, rows){

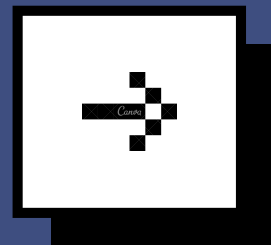
    function unpack(rows, key) {
        return rows.map(function(row) { return row[key.replace('.', ' ')] });
    }

    colors = []
    for (i=0; i < unpack(rows, 'class').length; i++) {
        if (unpack(rows, 'class')[i] == "Iris-setosa") {
            colors.push(0)
        } else if (unpack(rows, 'class')[i] == "Iris-versicolor") {
            colors.push(0.5)
        } else if (unpack(rows, 'class')[i] == "Iris-virginica") {
            colors.push(1)
        }
    }

    var pl_colorscale=[
        [0.0, '#19d3f3'],
        [0.333, '#19d3f3'],
        [0.333, '#e763fa'],
        [0.666, '#e763fa'],
        [0.666, '#636efa'],
        [1, '#636efa']
    ]
```



Code here



JAVASCRIPT (CONT.)

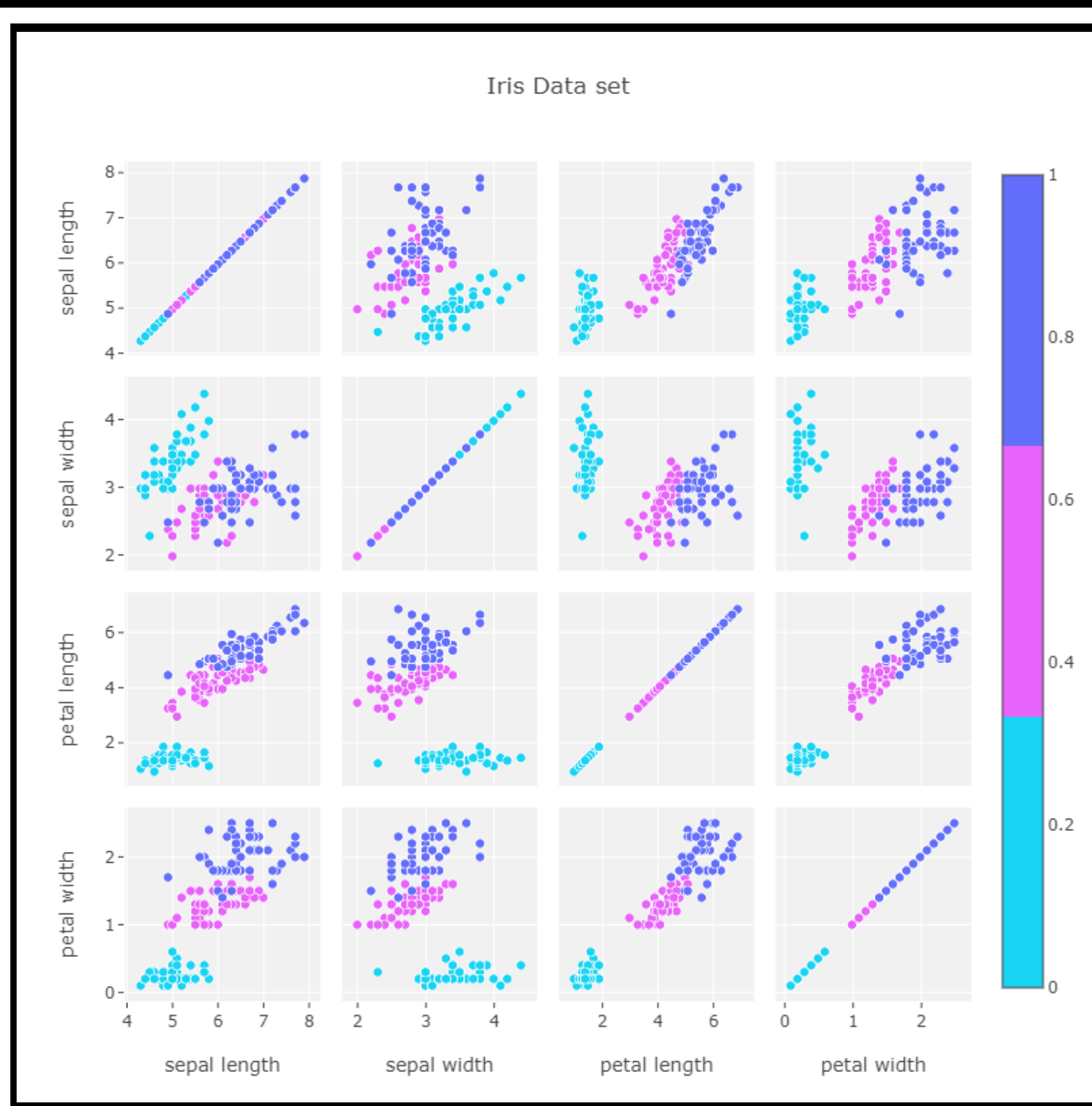
```
var axis = () => ({
  showline:false,
  zeroline:false,
  gridcolor:'#ffff',
  ticklen:4
})

var data = [{
  type: 'splom',
  dimensions: [
    {label:'sepal length',
  values:unpack(rows,'sepal length')},
    {label:'sepal width',
  values:unpack(rows,'sepal width')},
    {label:'petal length',
  values:unpack(rows,'petal length')},
    {label:'petal width',
  values:unpack(rows,'petal width')}
  ]
  text: unpack(rows, 'class'),
  marker: {
    color: colors,
    showscale: true,
```

JAVASCRIPT

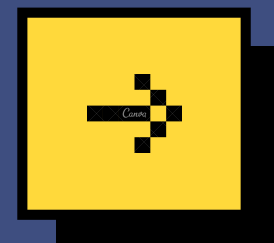
```
colorscale:pl_colorscale,
  size: 7,
  line: {
    color: 'white',
    width: 0.5
  }
}]

var layout = {
  title:'Iris Data set',
  height: 800,
  width: 800,
  autosize: false,
  hovermode:'closest',
  dragmode:'select',
  plot_bgcolor:'rgba(240,240,240, 0.95)',
  xaxis:axis(),
  yaxis:axis(),
  xaxis2:axis(),
  xaxis3:axis(),
  xaxis4:axis(),
  yaxis2:axis(),
  yaxis3:axis(),
  yaxis4:axis(),
  template: {
    data: {
      choroplethmapbox: [{
        marker: {
          colorbar: { ypad: 100},
          showscale: true,
        }
      }]
    }
  }
}
```



THE OUTPUT

Correlogram are really handy for exploratory analysis. It allows to visualize the relationships of the whole dataset in a glimpse. For instance, the linear relationship between petal length and petal width is obvious here, as the one concerning sepal.



GRACIAS

