Automated Insights Report

Diagnosic & Exploratory Data Report

16 November 2017

Data Profile

Dataset name: iris.csv

Dimensions: 150 rows and 5 columns.

Column names: Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, Species

Data Structure:

```
## 'data.frame':    150 obs. of  5 variables:
##  $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
##  $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
## NULL
```

Number of Records

Column.Name	Num.Records	Unique.Records
Sepal.Length	150	35
Sepal.Width	150	23
Petal.Length	150	43
Petal.Width	150	22
Species	150	3

Missing Values

There are no missing (NA) values in the dataset

Data Preview

Preview of the top 5 rows:

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5	3.6	1.4	0.2	setosa

Preview of the bottom 5 rows:

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
146	6.7	3	5.2	2.3	virginica
147	6.3	2.5	5	1.9	virginica
148	6.5	3	5.2	2	virginica
149	6.2	3.4	5.4	2.3	virginica
150	5.9	3	5.1	1.8	virginica

Unique Levels:

```
## Species has unique levels: setosa versicolor virginica
```

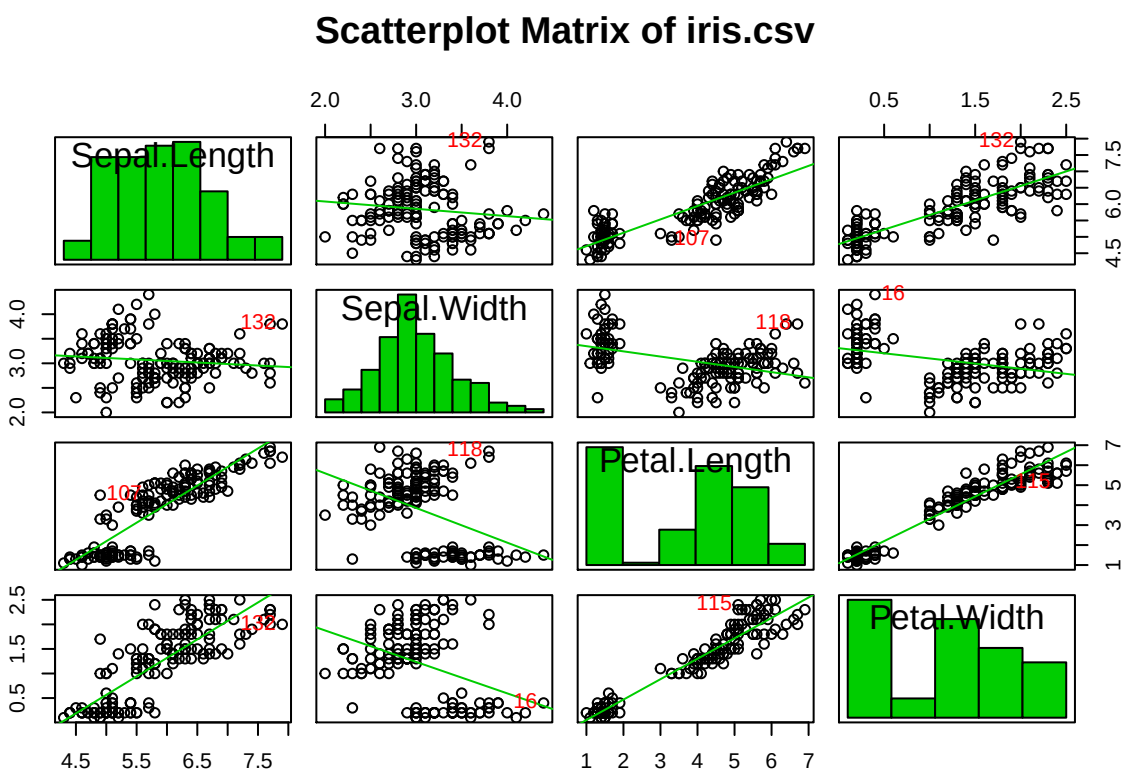
Statistics

Summary Statistics

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
Min. :4.300	Min. :2.000	Min. :1.000	Min. :0.100	setosa :50
1st Qu.:5.100	1st Qu.:2.800	1st Qu.:1.600	1st Qu.:0.300	versicolor:50
Median :5.800	Median :3.000	Median :4.350	Median :1.300	virginica :50
Mean :5.843	Mean :3.057	Mean :3.758	Mean :1.199	NA
3rd Qu.:6.400	3rd Qu.:3.300	3rd Qu.:5.100	3rd Qu.:1.800	NA
Max. :7.900	Max. :4.400	Max. :6.900	Max. :2.500	NA

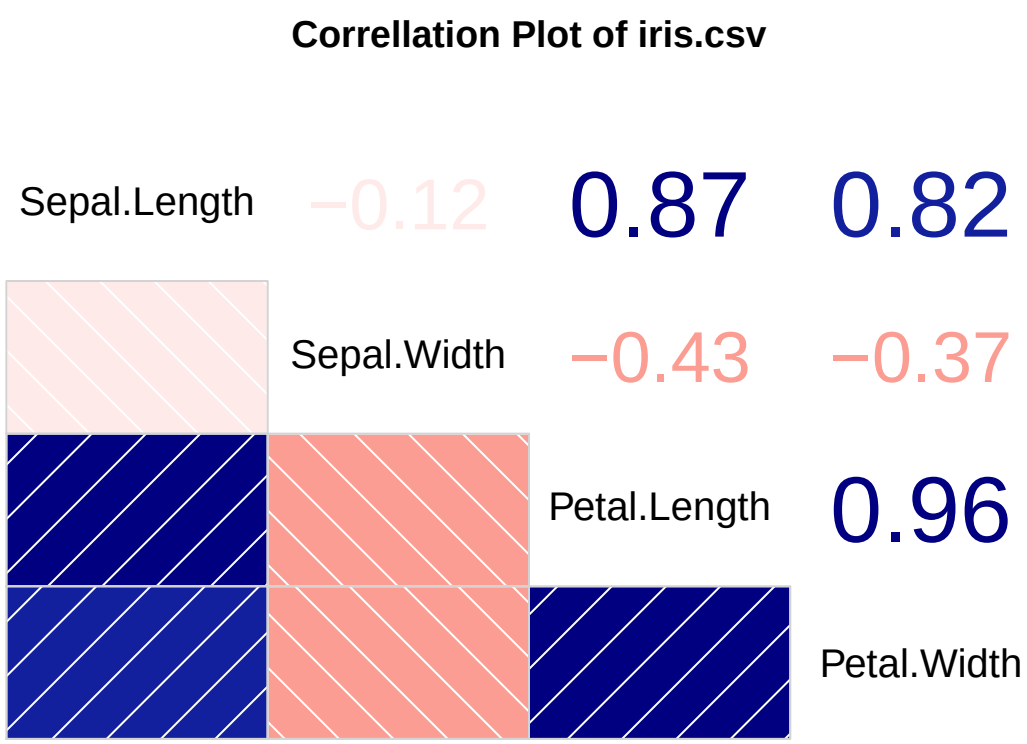
Visualisation

Scatterplot Matrix

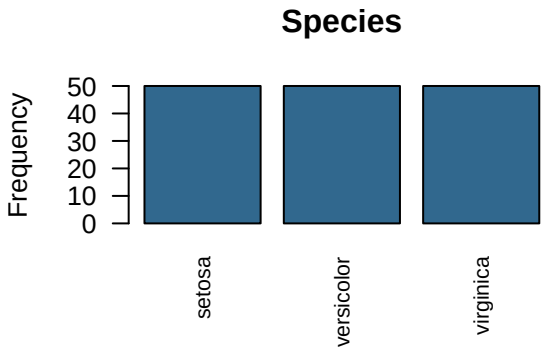


Only a maximum of six numeric variables are presented

Correlation Plot

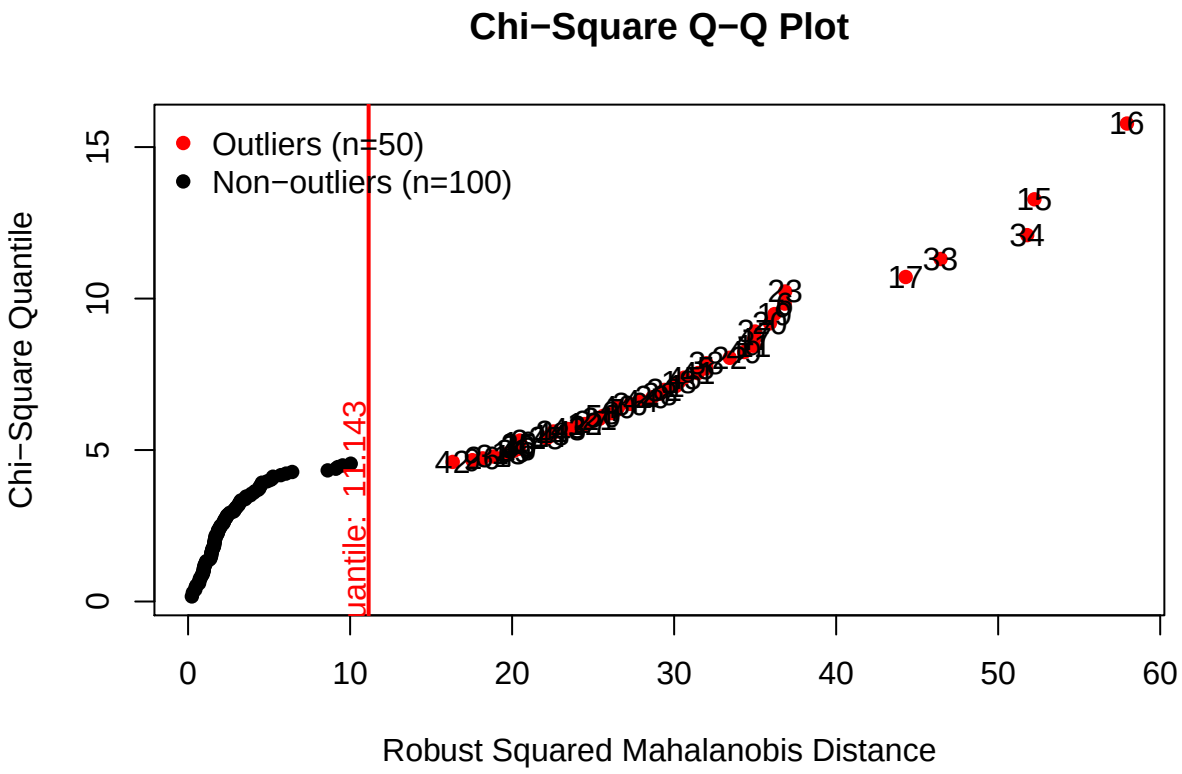


Factor Plot


Species
Frequency
50
40
30
20
10
0
setosa
versicolor
virginica

Multivariate Outlier Detection

Assessment for multivariate outliers using a Mahalanobis Chi-Square QQ Plot. An observation is an outlier if it's Mahalanobis distance is greater than the chi-square quantile value.



	Observation	Mahalanobis Distance	Outlier
16	16	57.948	TRUE
15	15	52.236	TRUE
34	34	51.790	TRUE
33	33	46.447	TRUE
17	17	44.290	TRUE
23	23	36.853	TRUE
6	6	36.821	TRUE
19	19	36.195	TRUE
20	20	35.916	TRUE
37	37	34.996	TRUE
47	47	34.938	TRUE
11	11	34.918	TRUE
49	49	34.326	TRUE
22	22	33.433	TRUE
38	38	31.993	TRUE
5	5	31.972	TRUE
41	41	31.465	TRUE
45	45	30.618	TRUE
18	18	30.304	TRUE
1	1	30.116	TRUE

	Observation	Mahalanobis Distance	Outlier
32	32	29.399	TRUE
28	28	29.133	TRUE
29	29	28.670	TRUE
44	44	27.961	TRUE
36	36	27.291	TRUE
40	40	26.550	TRUE
7	7	26.440	TRUE
8	8	26.119	TRUE
50	50	25.614	TRUE
21	21	25.402	TRUE
27	27	24.912	TRUE
12	12	24.485	TRUE
3	3	24.026	TRUE
43	43	23.470	TRUE
24	24	22.551	TRUE
48	48	22.466	TRUE
14	14	22.312	TRUE
25	25	22.128	TRUE
2	2	20.473	TRUE
10	10	20.461	TRUE
35	35	20.416	TRUE
30	30	20.403	TRUE
46	46	20.398	TRUE
39	39	20.107	TRUE
13	13	19.816	TRUE
4	4	19.604	TRUE
31	31	18.866	TRUE
26	26	18.182	TRUE
9	9	17.546	TRUE
42	42	16.333	TRUE