Iris example: data loading, basic stats and correlation

In [1]:

```
import pandas as pd
import seaborn as sn
```

Read and examine data

In [2]:

```
df = pd.read_csv('iris.csv')
df.head(10)
```

Out[2]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
5	5.4	3.9	1.7	0.4	setosa
6	4.6	3.4	1.4	0.3	setosa
7	5.0	3.4	1.5	0.2	setosa
8	4.4	2.9	1.4	0.2	setosa
9	4.9	3.1	1.5	0.1	setosa

In [3]:

```
df.dtypes
```

Out[3]:

```
sepal_length    float64
sepal_width     float64
petal_length    float64
petal_width     float64
species          object
dtype: object
```

In [4]:

```
df.describe()
```

Out[4]:

	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

Correlations

In [8]:

```
corrmatrix = df.corr()
print(corrmatrix)
```

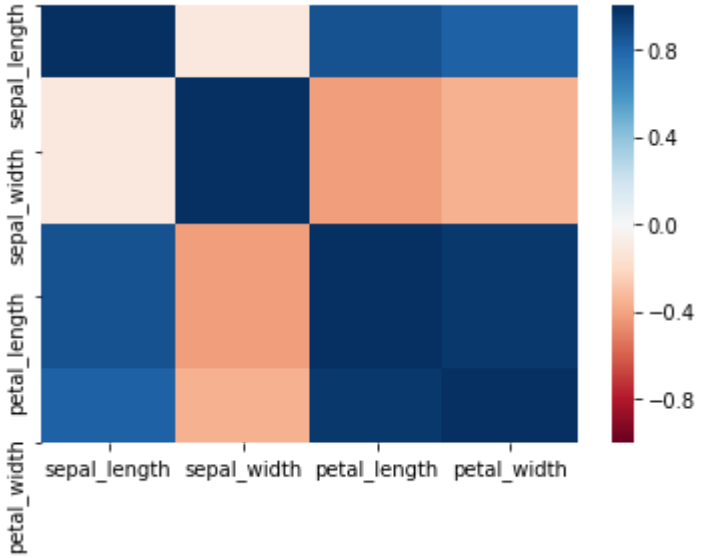
```
              sepal_length  sepal_width  petal_length  petal_width
sepal_length      1.000000    -0.109369      0.871754     0.817954
sepal_width      -0.109369     1.000000     -0.420516    -0.356544
petal_length      0.871754    -0.420516      1.000000     0.962757
petal_width       0.817954    -0.356544      0.962757     1.000000
```

In [9]:

```
sn.heatmap(corrmatrix, vmin=-1.0, vmax=1.0, cmap="RdBu")
```

Out[9]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1b4ce723608>
```



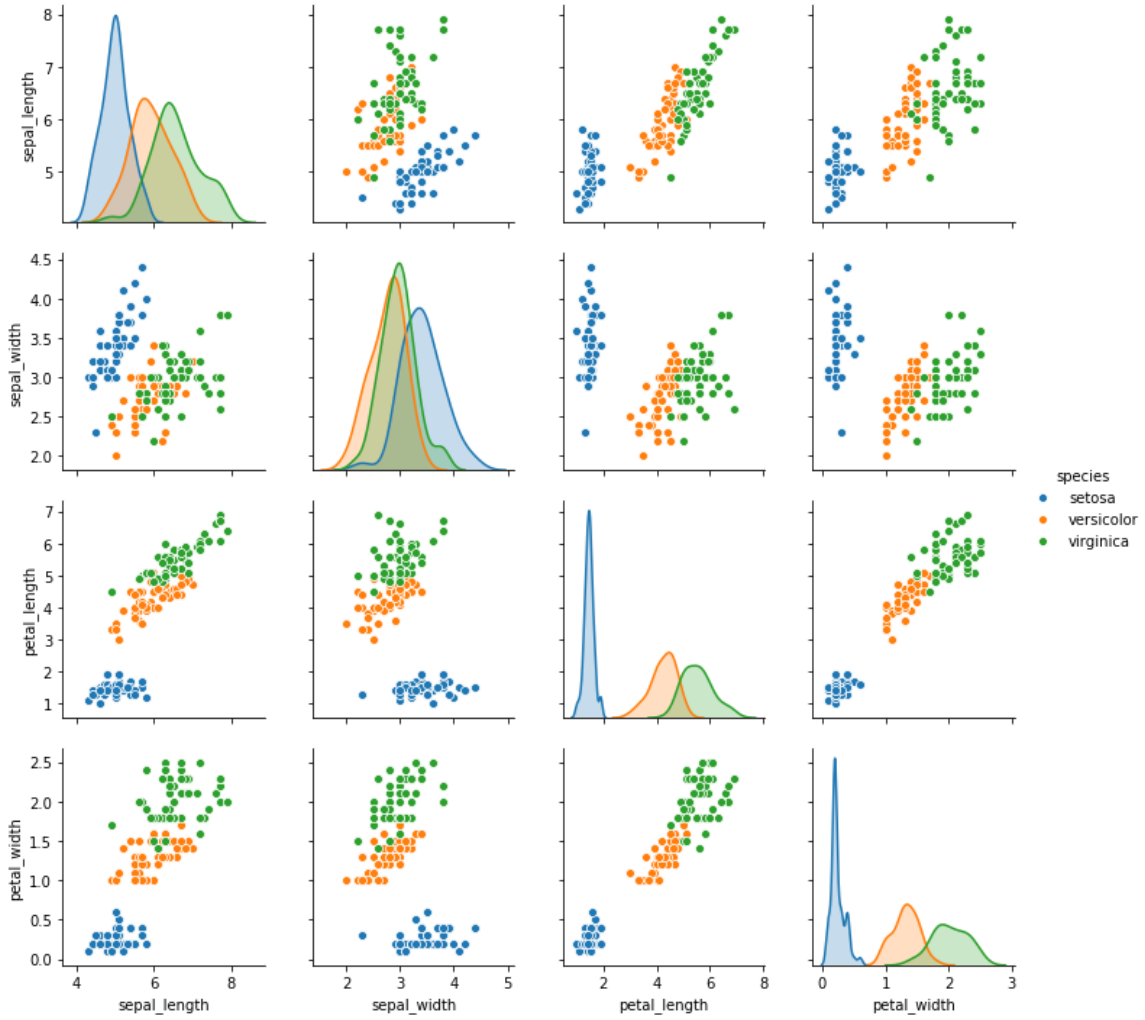
Scatterplots per species

In [7]:

```
sn.pairplot(df, hue="species")
```

Out[7]:

```
<seaborn.axisgrid.PairGrid at 0x1b4ccdf4248>
```



In [ ]: