

P3_fr

December 24, 2018

```
In [1]: from urllib import request
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()
```

```
In [2]: request.urlretrieve ("https://s3-eu-west-1.amazonaws.com/static.oc-static.com/prod/cou
hubble = pd.read_csv("hubble.csv")
```

```
In [3]: hubble.describe()
```

```
Out[3]:
```

	distance	recession_velocity
count	24.000000	24.000000
mean	0.911375	373.125000
std	0.645496	371.254666
min	0.032000	-220.000000
25%	0.406250	165.000000
50%	0.900000	295.000000
75%	1.175000	537.500000
max	2.000000	1090.000000

```
In [4]: from sklearn.linear_model import LinearRegression
```

```
X = hubble.distance.values.reshape(-1,1)
Y = hubble.recession_velocity
```

```
lr = LinearRegression()
lr.fit(X, Y)
```

```
Out[4]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
normalize=False)
```

```
In [5]: BIGGER_SIZE = 15
```

```
plt.rc('font', size=BIGGER_SIZE)           # taille de texte par défaut
plt.rc('axes', titlesize=BIGGER_SIZE)      # taille des titres des axes
plt.rc('axes', labelsizsize=BIGGER_SIZE)   # taille des labels des axes
plt.rc('xtick', labelsizsize=BIGGER_SIZE)  # taille des ticks des ascisses
```

```

plt.rc('ytick', labelsize=BIGGER_SIZE) # taille des ticks des ordonnées
plt.rc('legend', fontsize=BIGGER_SIZE) # taille de la légende
plt.rc('figure', titlesize=BIGGER_SIZE) # taille du titre

plt.figure()
plt.title("Données de Hubble")
plt.scatter(X, Y, color='black', label="données")
plt.plot(X, lr.predict(X), color='red', linewidth=3, label="prédiction")
plt.xlabel("Distance")
plt.ylabel("Vitesse de récession")
plt.legend()
plt.show()

```

