

```
In [ ]: #グラフィックの設定
%matplotlib inline

#ライブラリを読み込む
import pandas as pd #データフレーム
import matplotlib.pyplot as plt #グラフィック
```

```
In [ ]: #ファイルsample.csvを読み込みデータフレームDFに格納
#read_csvはpandasの機能
#ヘッダーは自動で認識される
DF = pd.read_csv("sample.csv", sep=",") #カンマ区切り

#最初の6行を表示
#headはデータフレームに付随する機能
display(DF.head(6))
```

```
In [ ]: #データ型
print(type(DF))
#データの構造 (行数・列数)
#shapeはnumpy配列やデータフレームに付随する機能
print(DF.shape)
```

```
In [ ]: #要約統計量の確認
#describe()はデータフレームに付随する機能
display(DF.describe())
```

```
In [ ]: #原点に近いデータを識別する列を作る
id = DF['varA']**2 + DF['varB']**2 < 1
print(id.head(6))
```

```
In [ ]: #IDという名前で新しい列を作る
#値には全て"steelblue"を入れておく
DF["ID"] = "steelblue"

#原点に近いデータのみIDの値を"darkblue"とする
#locはデータフレームに付随する機能 (行・列の指定)
DF.loc[id==True, "ID"] = "darkblue"

#それぞれのデータがいくつあるかを確認
#value_counts()はデータフレームに付随する機能
DF["ID"].value_counts()
```

```
In [ ]: #データフレームの内容を確認
display(DF.head(6))

#散布図を色分けして表示
#cは色の指定、alphaは透明度、sは点の大きさ
plt.scatter(DF["varA"], DF["varB"],
            c=DF["ID"], alpha=0.6, s=70)
plt.show
```

```
In [ ]:
```