



Python for Data Analysis

資料視覺化

講者: 楊翔斌
n07061033@mail.ncku.edu.tw

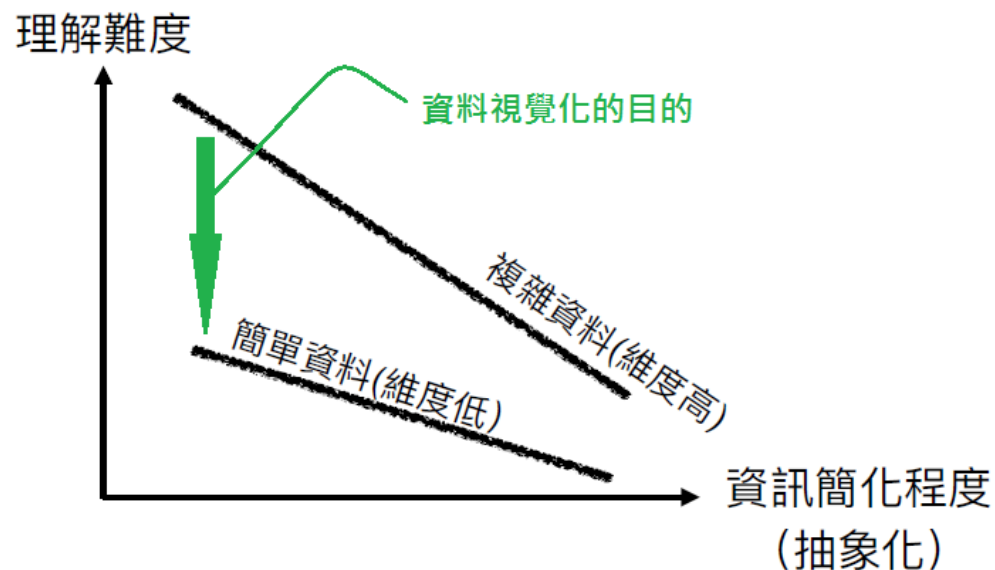
大綱

1. 前言

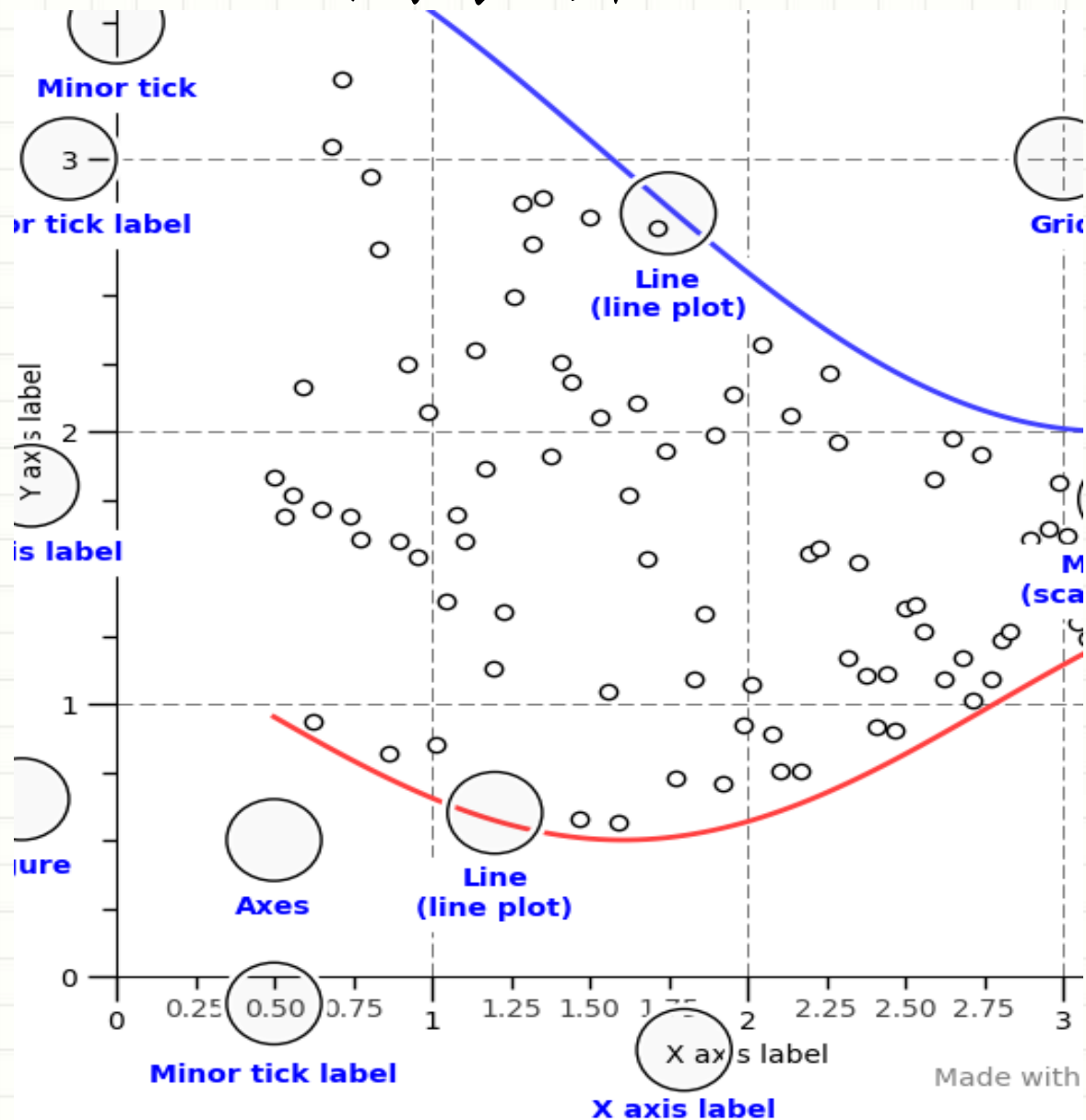
2. 圖表介紹

前言

文不如表 表不如圖



圖表介紹

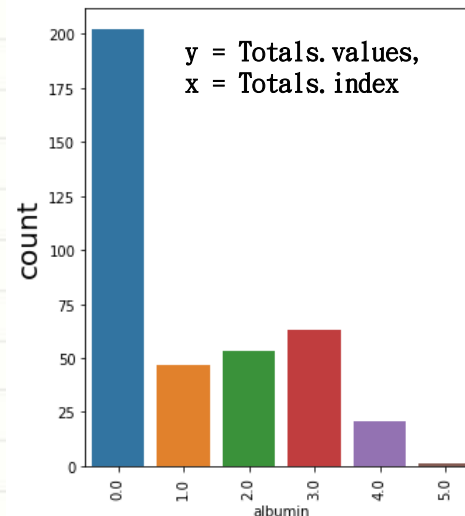
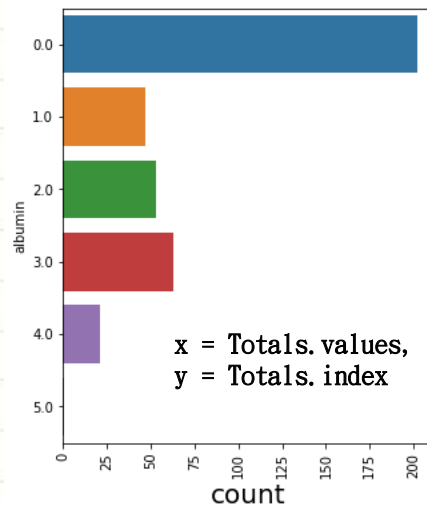


圖表介紹

barplot: 沿用腎臟病數據

Code:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
plt.figure(figsize=(5, 6))#定義圖大小
Totals=data1_new["albumin"].value_counts()#計算albumin各類別的量
sns.barplot(y = Totals.values, x = Totals.index)#繪製barplot
plt.xticks(rotation=90)#可調整tick的角度
plt.ylabel("count", size=20)#y軸名稱及字體大小
plt.xlabel("albumin", size=10)#x軸名稱及字體大小
法2:sns.countplot(x="albumin", data=data1_new)
```

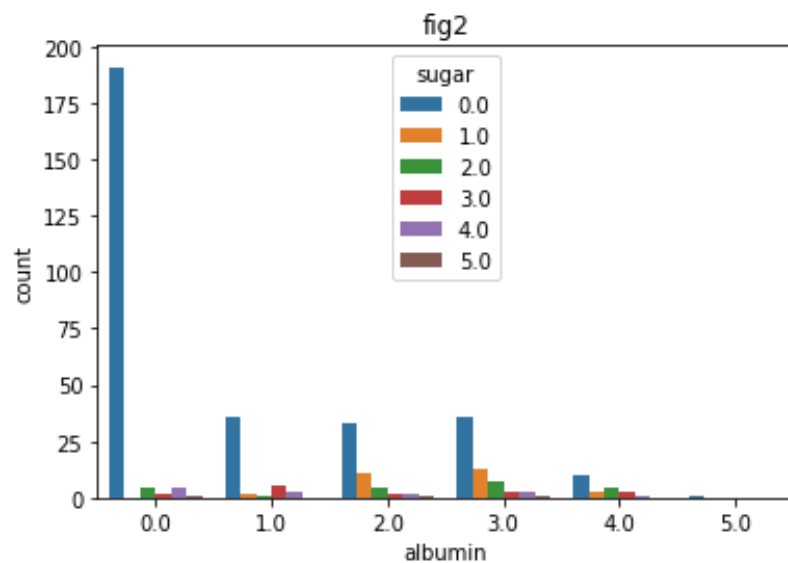
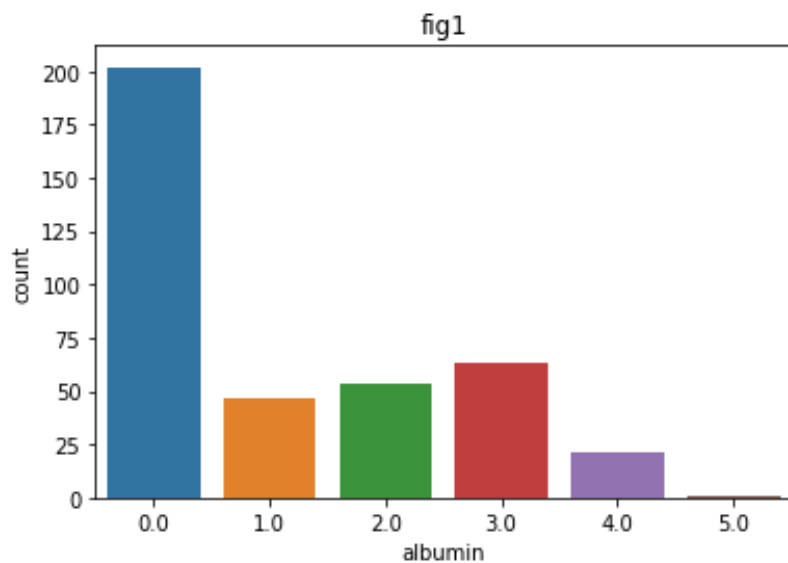


圖表介紹

countplot:

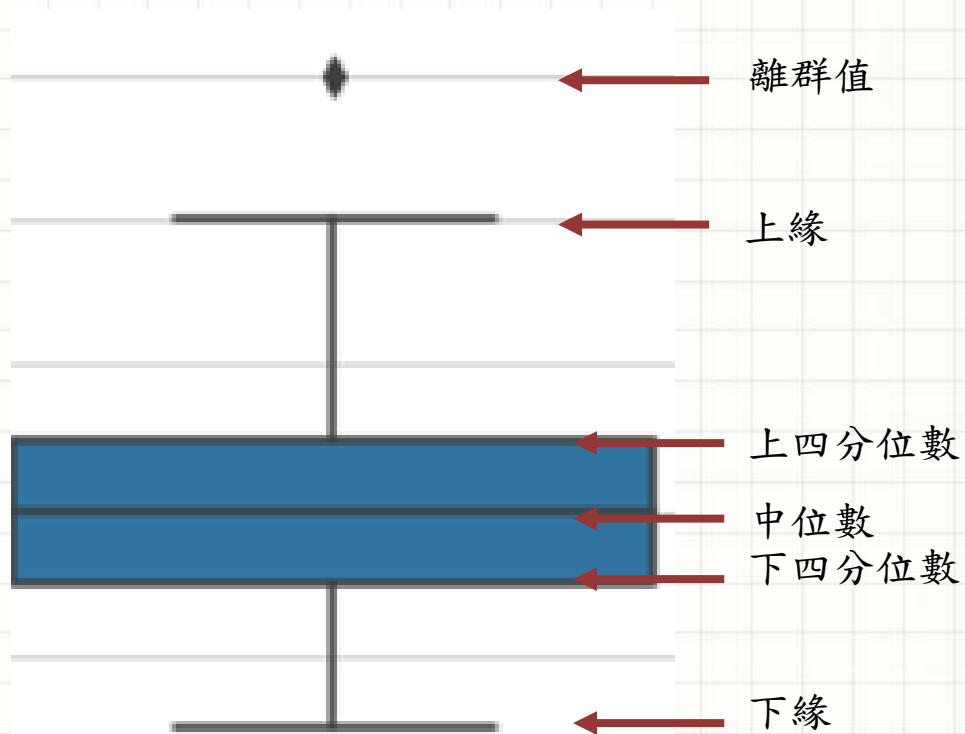
Code:

```
sns.countplot(x="albumin", data=data1_new)#依照albumin計數作圖，如fig1  
sns.countplot(x="albumin", hue="sugar", data=data1_new)  
#barplot 畫好後再根據sugar做分類
```



圖表介紹

boxplot☆☆☆:



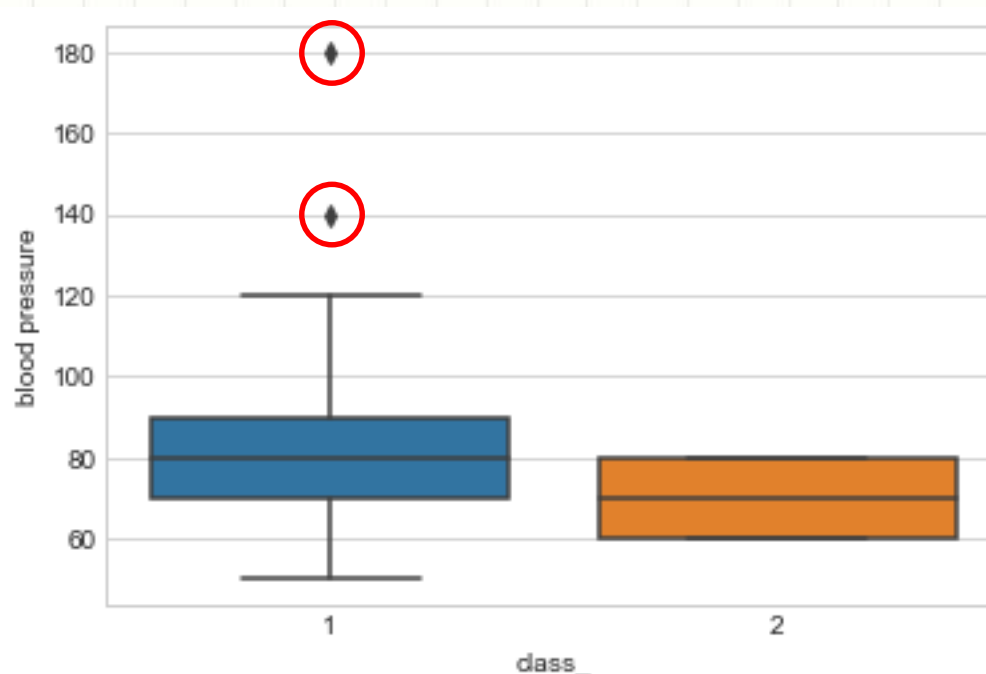
圖表介紹

boxplot:顯示一組數據之分散情況，可用來尋找離群值

Code:

```
sns.boxplot(x="class_", y="blood pressure", data=data1_new)
```

#依照class分類，將blood pressure數據畫出分佈圖



圖表介紹

violinplot: 定量數據在一個維度分類變量的多個層次上的分布

Code:

```
sns.violinplot(x="class_", y="hemoglobin", data=datal_new)#如fig1  
sns.violinplot(x="appetite_", y="hemoglobin", data=datal_new,  
hue="red_blood_cells")#引入第三維度變量進來，fig2  
sns.violinplot(x="appetite_", y="hemoglobin", data=datal_new,  
hue="red_blood_cells", split="1")
```

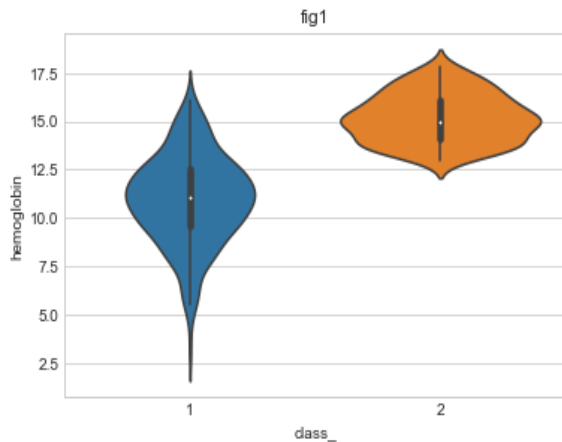


fig1

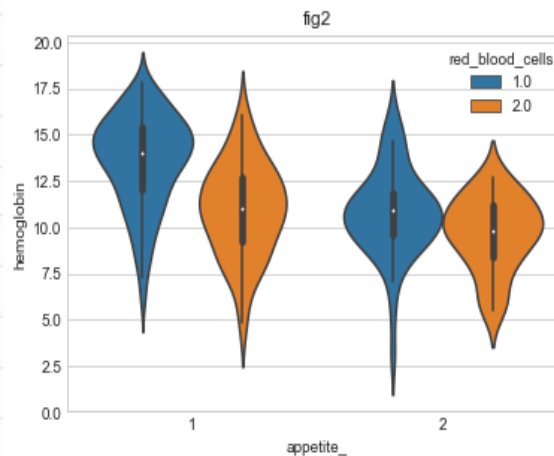


fig2

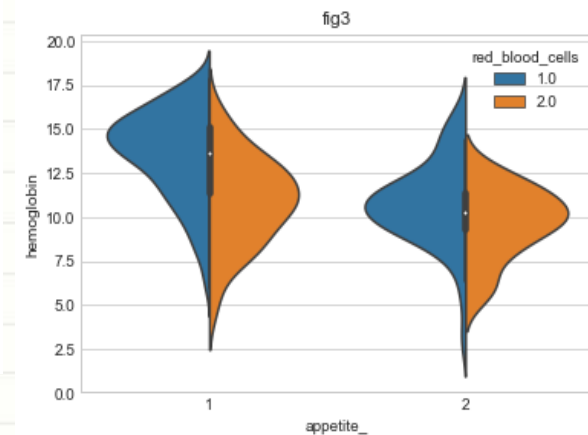


fig3

圖表介紹

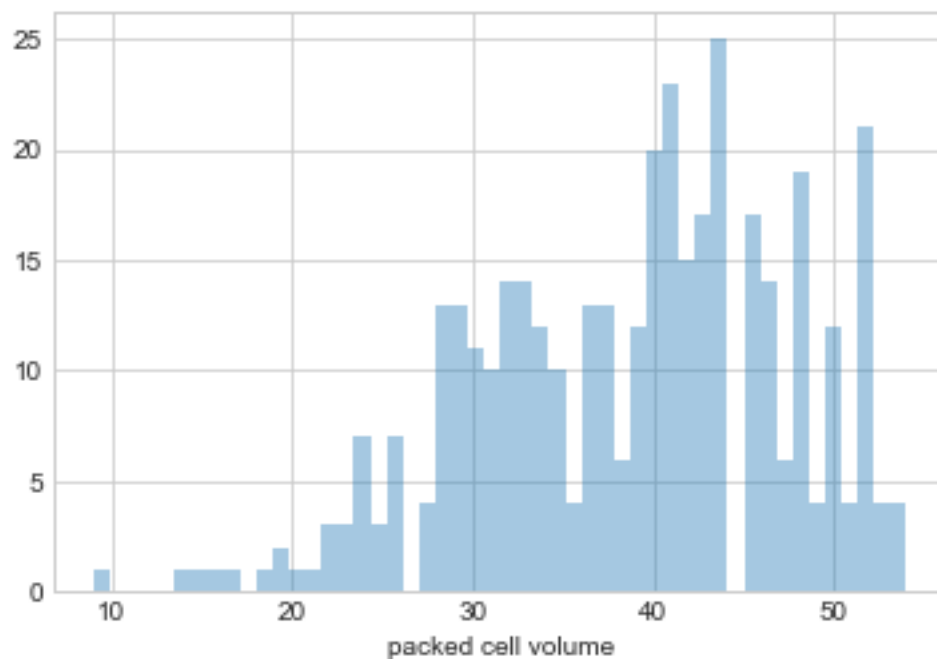
histogram:

Code:

```
sns.distplot(data1_new["packed cell volume"],  
kde=False, bins=50)
```

#kde為False則圖為實際數據，True則圖為機率分佈。

#bin表示柱的數量



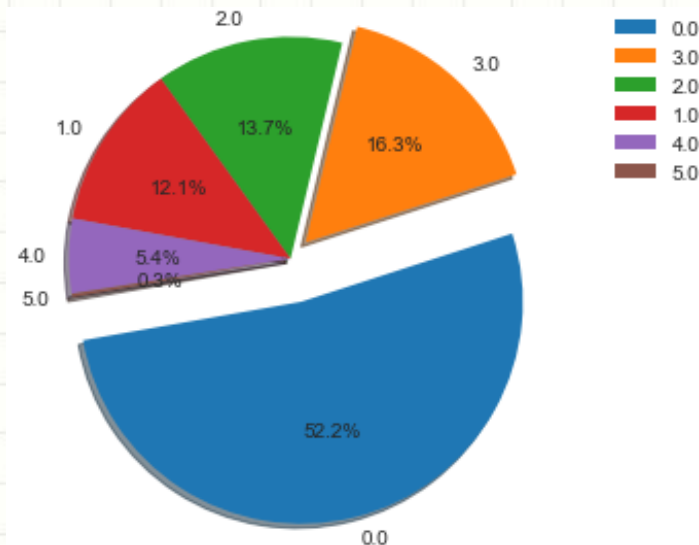
圖表介紹

pie chart:

Code:

```
Totals = data1_new["albumin"].value_counts()  
labels1 =Totals.index#取得label  
explode = (0.5, 0.1, 0, 0, 0, 0)#調整圓餅圖間距
```

```
plt.pie(Totals, explode=explode, labels=labels1 ,  
autopct='%1.1f%%', shadow=True, startangle=190)  
plt.legend( labels1, loc="best")  
plt.axis("equal")  
plt.tight_layout()
```



圖表介紹

Scatter plot:

Code:

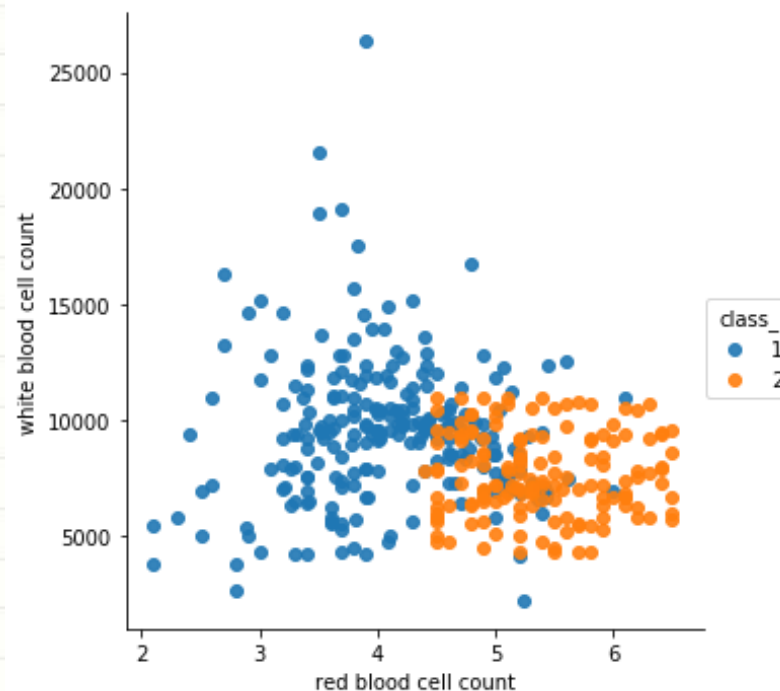
法一:`plt.plot(data1_new["... "], data1_new["... "], 'ro')`

法二:`g=sns.FacetGrid(data1_new, hue="class_", size=5)`

`#要畫散布圖的`

`g.map(plt.scatter, "red blood cell count",
"white blood cell count", alpha=.9)`

`g.add_legend()`



圖表介紹

jointplot:

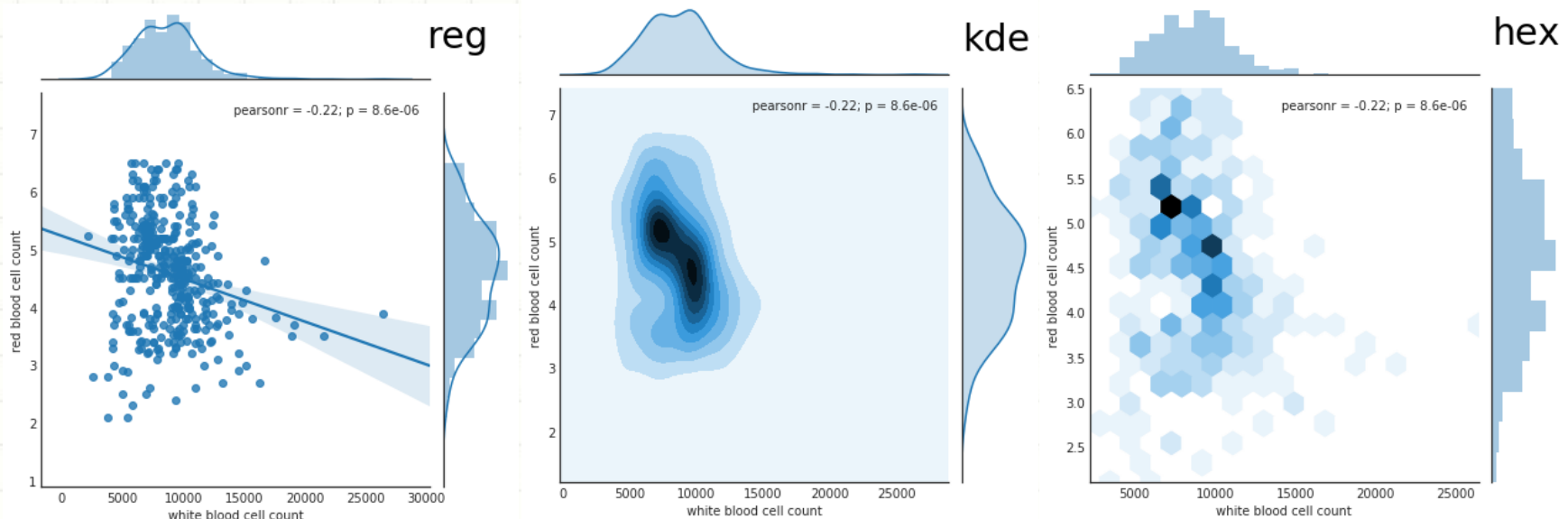
Code:

```
with sns.axes_style("white"):
```

```
    sns.jointplot("white blood cell count", "red blood cell count",  
                  data=datal_new, kind="scatter")
```

```
plt.text(30000, 8.5, "reg", size=30) #size可以調整字體大小
```

```
#kind可以選reg、kde、hex
```



圖表介紹

其他技巧: subplot

Code:

```
fig, axes=plt.subplots(2, 3, figsize=(15, 5))
sns.boxplot(x="sugar", y="blood glucose random",
data=datal_new, ax=axes[0, 1])
sns.boxplot(x="sugar", y="age", data=datal_new, ax=axes[0, 0])
sns.boxplot(x="sugar", y="blood urea", data=datal_new, ax=axes[0, 2])
sns.boxplot(x="sugar", y="serum creatinine",
data=datal_new, ax=axes[1, 0])
sns.boxplot(x="sugar", y="sodium", data=datal_new, ax=axes[1, 1])
sns.boxplot(x="sugar", y="hemoglobin", data=datal_new, ax=axes[1, 2])
```

