データベース演習 （Advanced Database Exercise）

１１．種々のグラフ（Python の pandas）

URL: http://www.kkaneko.jp/cc/dbenshu/index.html

**概要 Abstract**

今日の授業では、データのビジュアライズに関する演習を行う．今日は、CSVファイルを使う．

Today's class is exercise on data visualization.

**演習 11 (Exercises 11)**

* **ステップ1 (Step 1) データファイルの準備 (prepare data files)**

 USB メモリを使い、CSVデータファイル**iris.csv** と**planets.csv** を配る．各自、ファイルを **C:\iris.csv と C:\planets.csv にコピー.**

**※昔の C:\iris.csv が残っている場合には、新しい iris.csv で上書きしなさい**

 Two data files named '**iris.csv**' and '**planets.csv**' will be distributed using a USB memory. Everyone is expected to copy the two files to '**C:\iris.csv**' and '**C:\planets.csv**'. Please remove the old iris.csv (pleause use the new iris.csv)



 もし、**USB メモリが使えない**場合には、Webブラウザを開き、

　　[http://133.5.18.161/rinkou/addb/iris.csv](http://www.kkaneko.com/rinkou/addb/iris.csv)　と [http://133.5.18.161/rinkou/addb/planets.csv](http://www.kkaneko.com/rinkou/addb/iris.csv)

 **をダウンロードして、C:\iris.csv と C:\planets.csv にコピー.**

　 ( If you can not use USB memory, open [http://133.5.18.161/rinkou/addb/iris.csv](http://www.kkaneko.com/rinkou/addb/iris.csv)

　　 and [http://133.5.18.161/rinkou/addb/planets.csv](http://www.kkaneko.com/rinkou/addb/iris.csv) using a Web browser, and copy it to **C:\iris.csv and C:\planets.csv**)

**◇ iris.csv (Microsoft Excel の画面）**



**◇ planets.csv (Microsoft Excel の画面）**



* **ステップ2 (Step 2) Python のパッケージ seaborn のインストール (Install seaborn packaghe)**

**ステップ2を行うには、インターネット接続が必要である. いまインターネット接続ができない人は、ステップ2以降を宿題とする (Internet connection is necessary. If you can not do internetc connection, the step 2 and later are your homework)**

1. Python のパッケージをインストールしたいので、IPython(sh) を起動 (execute the "IPython (sh)" to install python packages)



2. IPython(sh) で次のコマンドを実行 (execute the following commands in IPython(sh)'s console)

 **easy\_install seaborn**

 ※ このコマンドを実行したときに、エラーメッセージが出た場合には、「pip install seaborn」を実行

**＜実行結果の例の一部分＞**

****

**ステップ５以降は、ステップ２が必要**

* **ステップ3 (Step 3)**

**課題：以下の手順で Spyderの設定を行いなさい　(Setup the Spyder software)**

1. Spyder の起動 (execute the Spyder)



1. Interpreters で 「**Open an IPython Console**」を選ぶ. (Select Interpreters, and Select “Open an IPython console”)



1. 「IPython console」をクリック. (Click “IPython console”)



* **ステップ4 (Step 4)**

**CSVファイルを Pandasのデータフレームに格納する (read CSV file and store into Pandas data frame)**

1. まず CSVファイルのファイル名を確認する (Examine the correct CSV file name)



　以下、ファイル名が **C:\planets.csv** と **C:\iris.csv** であるとして説明を続ける

 (In this exercise, the file names are c:\planets.csv and c:\iris.csv)

2. データフレームに CSVファイルを読み込む.　(read CSV file into dataframe).



3. planets の確認 (print planets)

　 head() は先頭行の表示 ('head()' is to print the top 5 lines)



4. iris の確認 (print iris)

　 head() は先頭行の表示 ('head()' is to print the top 5 lines)



* **ステップ5 (Step 5)**

ビジュアライズ (visualize)

**(1) 準備 (import packages)**



**(2) 相関行列　(correlation matrix)**

 各列間の相関係数が表示される (correlation coefficients are displayed)





**(3) 散布図　(scatter plot)**

◇ データフレーム: planets, 列: distance, mass ※ number を使って色付け



　◇ データフレーム: planets, 列: distance, mass ※ number を使って色付け



◇ データフレーム: iris, 列: petal\_length, petal\_width ※ species を使って色付け



**(4) ペア・プロット (pair plot)**

◇ データフレーム: iris



**(5) ヒストグラム　(histogram)**

◇ データフレーム: iris ※ species を使って色付け



◇ データフレーム: planets, 列: year

 ※ 横軸は 1985, 1986, …, 2014．縦横比は 2.5



　◇ データフレーム: planets, 列: year ※ number を使って色付け

 ※ 横軸は 1985, 1986, …, 2014．縦横比は 2.5



 ◇ データフレーム: iris, 列: sepal\_length

 ※ 横軸は 4.0, 4.1, …, 7.9．縦横比は 1.8



 ◇ データフレーム: iris, 列: sepal\_length ※ species を使って色付け

 ※ 横軸は 4.0, 4.1, …, 7.9．縦横比は 1.8



**(6) １変数の分布 (Univariate distribution)**

データの分布や変化を可視化する. ヒストグラム，カーネル密度推定などを使用．

(Visualizing univariate distributions or time-changes using histogram, kernel density estimation).

　◇ データフレーム: planets, 列: year



　◇ データフレーム: iris, 列: sepal\_length





カーネル密度推定

(kernel density estimation)



カーネル密度推定

(kernel density estimation)

**(7) 2変数の分布 (Bivariate distribution)**

データの分布や変化を可視化する. ヒストグラム，カーネル密度推定などを使用．

(Visualizing bivariate distributions using histogram, kernel density estimation).

　◇ データフレーム: iris, 列: petal\_length, petal\_width



散布図とヒストグラム

 (Scatter plot and histogram)



カーネル密度推定

(kernel density estimation)



回帰分析 (regression)