

基礎演習

第6回 データ解析実習（1）図の作成

メディア情報コース
平居 悠（ひらい ゆたか）

到達目標

- データ解析の基本的なスキルを身につける。
- 適切に文献を引用し、自分の意見を記述したレポートを作成できる。
- 調べてわかったことをわかりやすく他者にわかりやすく伝えることができる。

前回

- 第1回 授業の概要、自己紹介 (4/13)
- 第2回 天文学入門 138億年の宇宙の歴史 (4/20)
- 第3回 図書館ツアー、データベース検索 (4/27)
- 第4回 データ解析基礎 (1) プログラミング言語Python入門 (5/11)
- 第5回 データ解析基礎 (2) 基本統計量 (5/18)**
- 第6回 データ解析実習 (1) 図の作成 (5/25)
- 第7回 データ解析実習 (2) データの解釈 (6/1)
- 第8回 レポートの作成 (1) 文章の読み方と要約、レポートの構成と表現 (6/8)
- 第9回 レポートの作成 (2) 著作権 (6/15)
- 第10回 プレゼンテーションの方法 (6/22)
- 第11回 発表準備 (1) 発表資料の作成 (6/29)
- 第12回 発表準備 (2) 発表練習 (7/6)
- 第13回 発表会 (7/13)

前回の目標

プログラミング言語Python
でデータを読み込んで基本
統計量を算出できるように
なる。

前回の内容

解析するデータについて
データの読み込み
基本統計量

解析するデータ

天の川銀河形成シミュレーション

論文：

Hirai et al. (2022), MNRAS, 517, 4856

<https://academic.oup.com/mnras/article/517/4/4856/6808594>

Hirai et al. (2024), ApJ, 970, 105

<https://iopscience.iop.org/article/10.3847/1538-4357/ad500c>

Hirai et al. (2025), ApJ, 9990, 125

<https://iopscience.iop.org/article/10.3847/1538-4357/adf10a>

基本的物理量

元素量

太陽と比べてどれくらいの元素量か？

鉄の場合

[Fe/H]

$= \log_{10}(\text{鉄の数密度} / \text{水素の数密度})$

$-\log_{10}(\text{太陽の鉄の数密度} / \text{太陽の水素の数密度})$

[Fe/H] = 0なら太陽と同じ、[Fe/H] = -1なら
太陽の1/10の鉄の量

レコードとフィールド

レコード：全ての項目が揃った1単位のデータ

フィールド：レコードの中にある項目

レコード

ID3	x	y	z	vx	vy	vz	t_form	MgH	FeH	MgFe
96282195	0.0908529	0.647045	1.096310000	-121.438	16.5403	66.8084	3.618454351	-0.6094579	-0.5087503	-0.1007076

フィールド

CSVファイルについて

複数のフィールドごとにカンマ
(,)で区切り、1レコードを1行に
おさめたファイル形式

ファイル内のデータの説明

ID3: 星のID

x, y, z: 位置 (kpc)

v_x, v_y, v_z: 速度 (km/s)

t_{form}: 星の形成時刻 (Gyr)

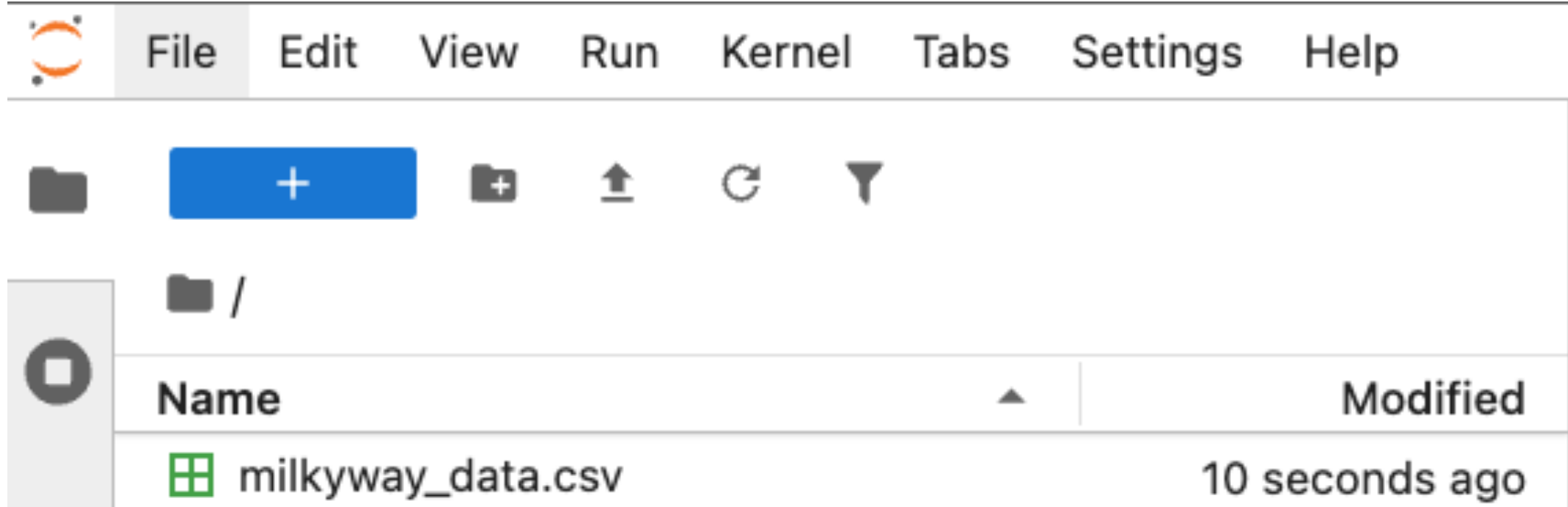
MgH: [Mg/H]

FeH: [Fe/H] (金属量)

MgFe: [Mg/Fe]

新たなノートブックを生成

1. JupyterLabを開き、左上の+ボタンから新たなノートブックを作成。
2. ファイル名は“milkyway.ipynb”とする。



必要なライブラリのインポート

Pandas, Numpy, Matplotlibをインポートする。

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

データの読み込み

read_csv()を用いる

```
df = pd.read_csv("milkyway_data.csv")
```

平均値: mean メソッド

個々の数値を全て足し合わせた後、
数値の個数で割った値

星の形成時刻の平均値 星の金属量の平均値

```
df["t_form"].mean()
```

```
6.673826215048475
```

```
df["FeH"].mean()
```

```
-0.35654102618473477
```

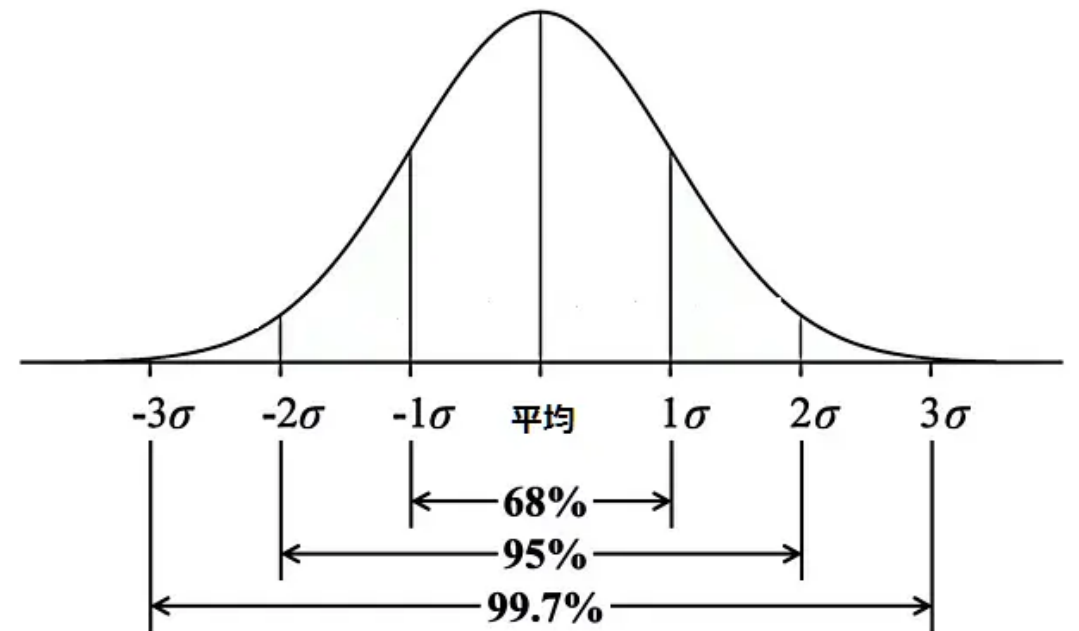
データのばらつき

標準偏差：

分散の平方根をとった値

標準偏差の意味：

平均値50、標準偏差10の場合、平均値50に対して、各データが標準的に10離れている。50 ± 10の範囲に全データの約68%が含まれる。



<https://avilen.co.jp/personal/knowledge-article/standard-deviation/>

中央値: median メソッド

データを小さい順に並べたときにちょうど真ん中に来る値

星の形成時刻の中央値

```
df["t_form"].median()
```

6.47567872935

星の金属量の中央値

```
df["FeH"].median()
```

-0.2815630059364782

四分位数: quantile() メソッド

データを小さい順に並べたとき、データの数で4等分した時の区切り値

第一四分位数：小さい方から25%

第二四分位数：中央値

第三四分位数：小さい方から75%

四分位範囲 = 第三四分位数 - 第一四分位数

今回

- 第1回 授業の概要、自己紹介 (4/13)
- 第2回 天文学入門 138億年の宇宙の歴史 (4/20)
- 第3回 図書館ツアー、データベース検索 (4/27)
- 第4回 データ解析基礎 (1) プログラミング言語Python入門 (5/11)
- 第5回 データ解析基礎 (2) 基本統計量 (5/18)
- 第6回 **データ解析実習 (1) 図の作成 (5/25)**
- 第7回 データ解析実習 (2) データの解釈 (6/1)
- 第8回 レポートの作成 (1) 文章の読み方と要約、レポートの構成と表現 (6/8)
- 第9回 レポートの作成 (2) 著作権 (6/15)
- 第10回 プレゼンテーションの方法 (6/22)
- 第11回 発表準備 (1) 発表資料の作成 (6/29)
- 第12回 発表準備 (2) 発表練習 (7/6)
- 第13回 発表会 (7/13)

今回の目標

Matplotlibを用いて
基本的な図を描ける
ようになる。

今回の内容

Matplotlibについて

折れ線グラフ

散布図

棒グラフ

ヒストグラム

今回の内容

Matplotlibについて

折れ線グラフ

散布図

棒グラフ

ヒストグラム

Matplotlib

グラフ描画用ライブラリ
データの可視化に便利

公式サイト：

<https://matplotlib.org/>

なぜMatplotlibか？

- 多種多様なグラフ作成
- 高いカスタマイズ性
- JupyterLabとの連携
- 表計算ソフトでは描けな
い美しい図を作成できる

Matplotlibのインポート

```
import pandas as pd
```

```
import numpy as np
```

```
import matplotlib.pyplot as plt
```

描画オブジェクト

```
ax = plt.subplots() #プロットの枠を生成
```

今回の内容

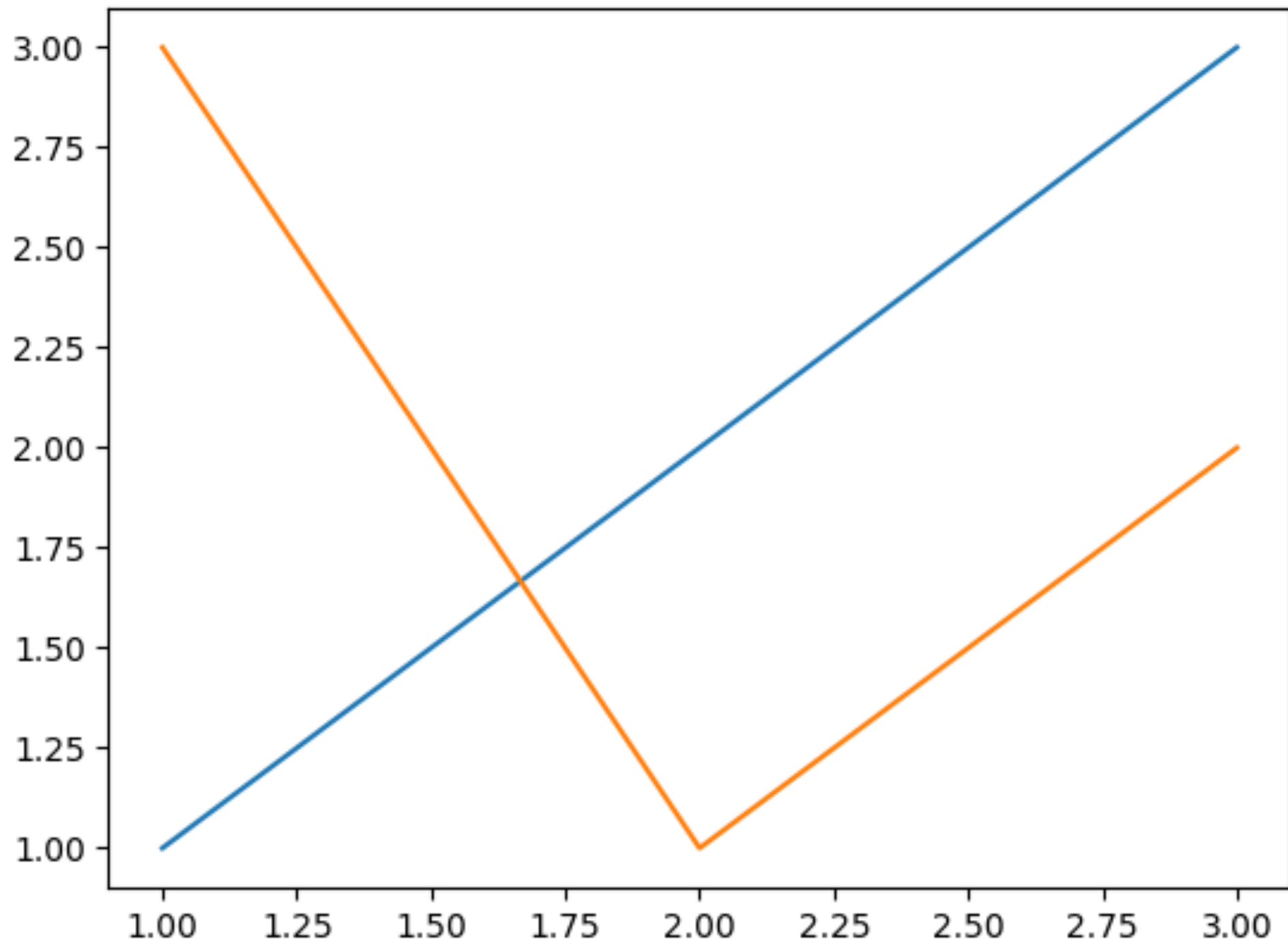
Matplotlibについて

折れ線グラフ

散布図

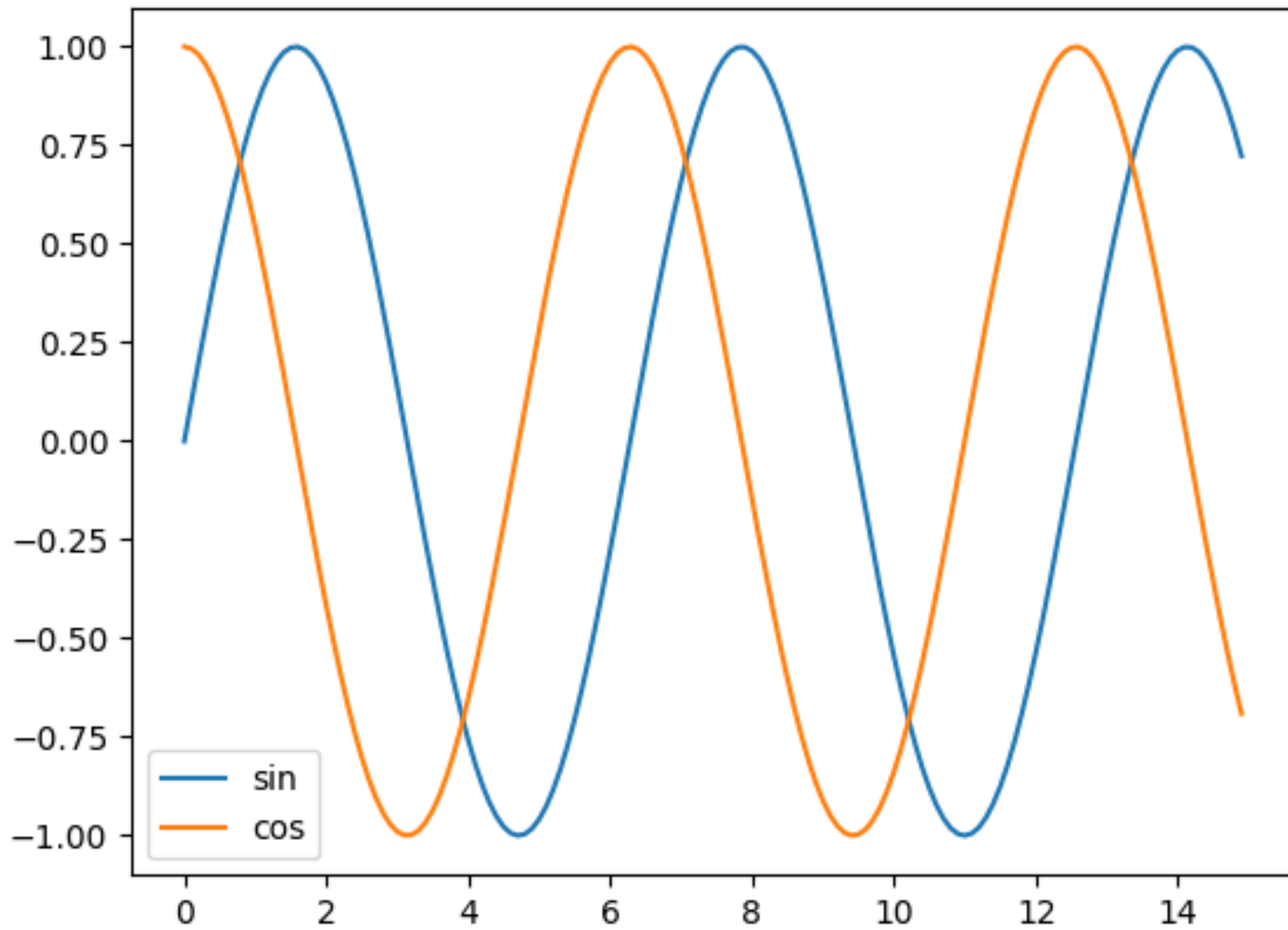
棒グラフ

ヒストグラム



折れ線グラフ

```
x = [1, 2, 3]      #グラフ描画用のリスト
y1 = [1, 2, 3]
y2 = [3, 1, 2]
plt.plot(x, y1)   #xとy1のデータについての折れ線グラフ
plt.plot(x, y2)   #xとy2のデータについての折れ線グラフ
```



折れ線グラフ

```
x=np.arange(0.0, 15.0, 0.1) #0から15までの数値を0.1刻みで生成
y1 = np.sin(x)
y2 = np.cos(x)

plt.plot(x, y1, label='sin')
plt.plot(x, y2, label='cos')
plt.legend() #凡例の表示
```

今回の内容

Matplotlibについて

折れ線グラフ

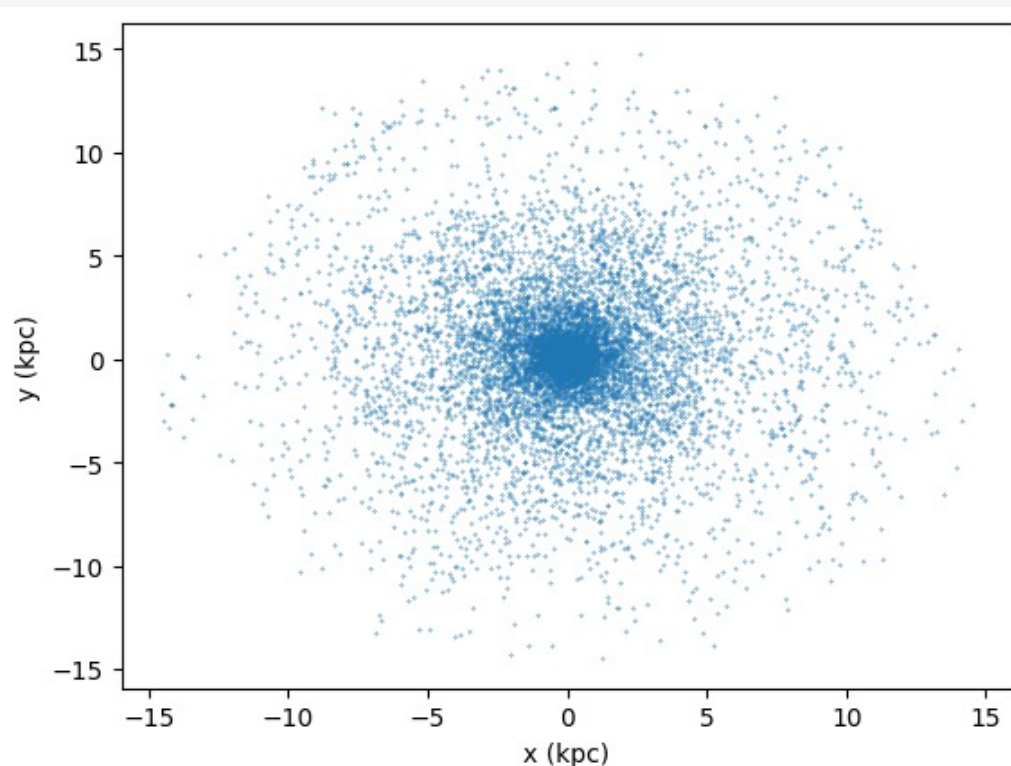
散布図

棒グラフ

ヒストグラム

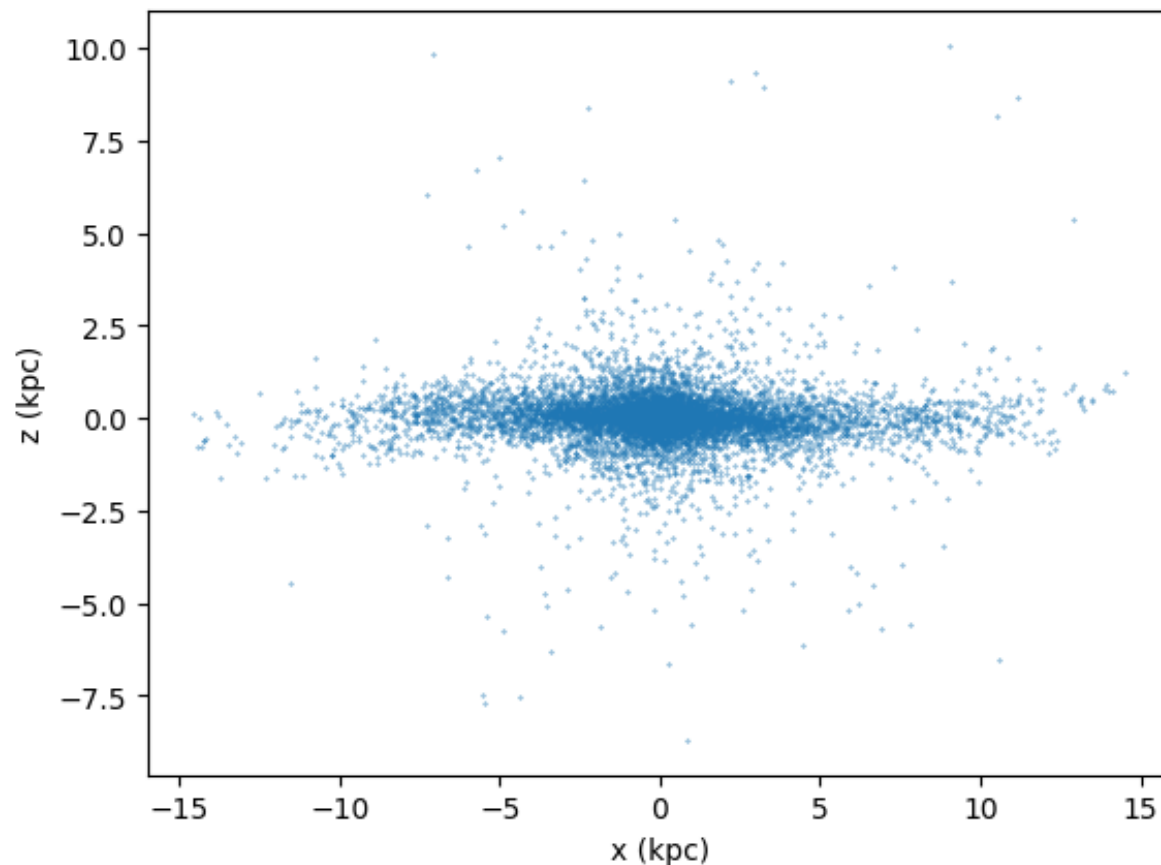
散布図

```
df = pd.read_csv('milkyway_data.csv') #データ読み込み  
plt.xlabel('x (kpc)') #x軸名  
plt.ylabel('y (kpc)') #y軸名  
plt.scatter(df['x'], df['y'], s=0.1) #星の空間分布を表示
```



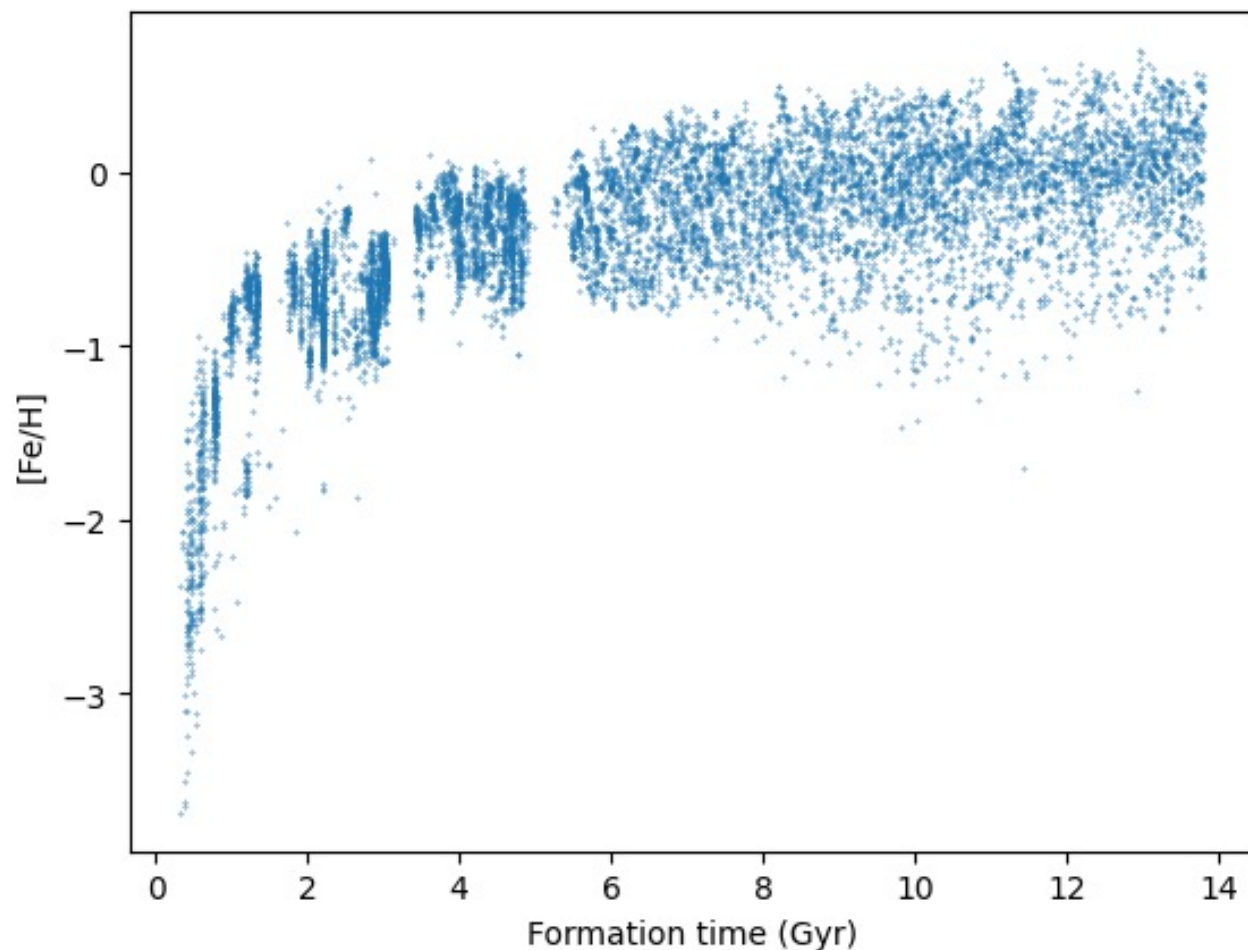
散布図

```
plt.xlabel('x (kpc)') #x軸名  
plt.ylabel('z (kpc)') #y軸名  
plt.scatter(df['x'], df['z'], s=0.1) #星の空間分布を表示
```



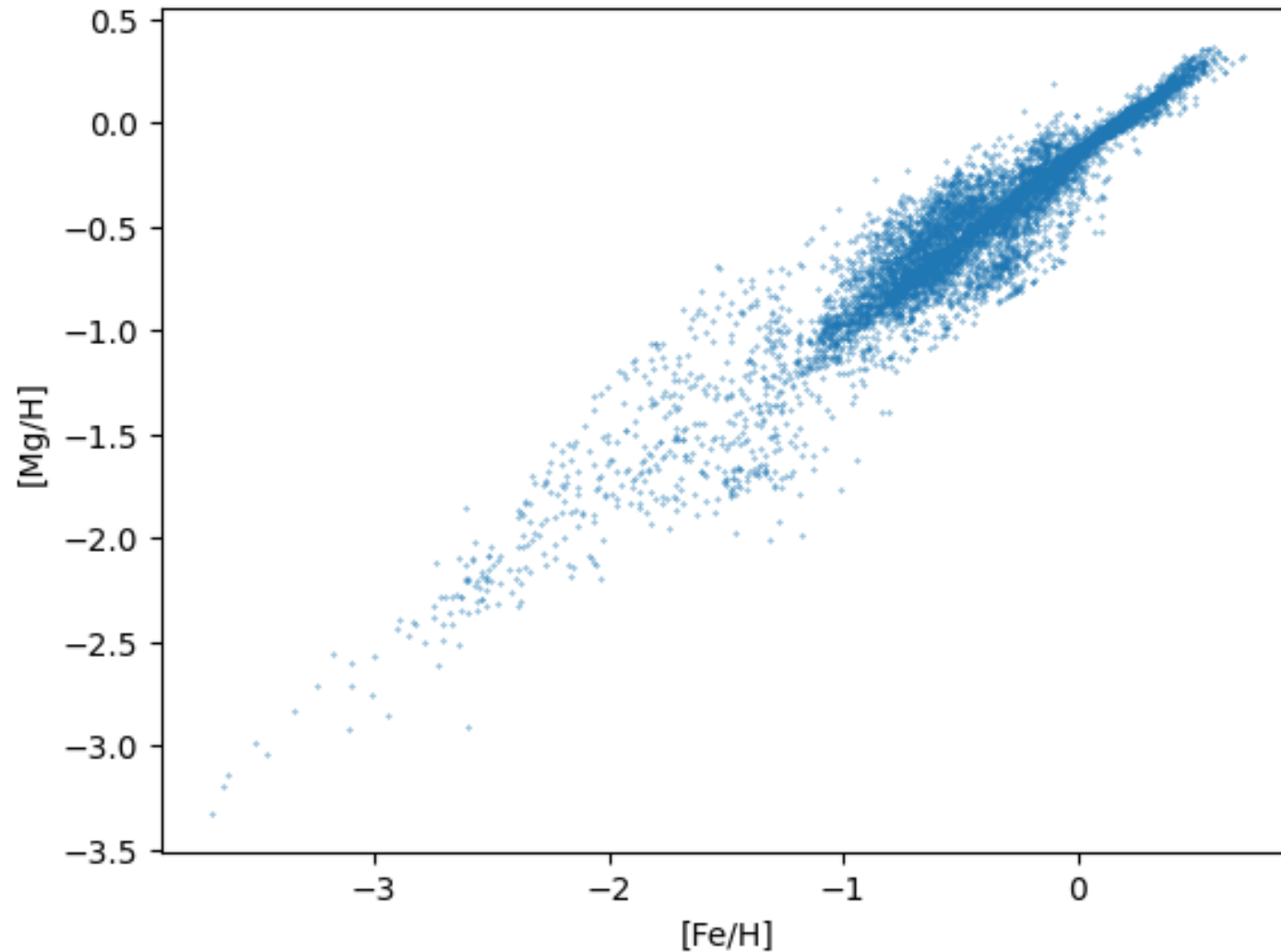
散布図

```
plt.xlabel('Formation time (Gyr)') #x軸名  
plt.ylabel('[Fe/H]') #y軸名  
plt.scatter(df['t_form'], df['FeH'], s=0.1) #[Fe/H]と星の形成時刻の関係を表示
```



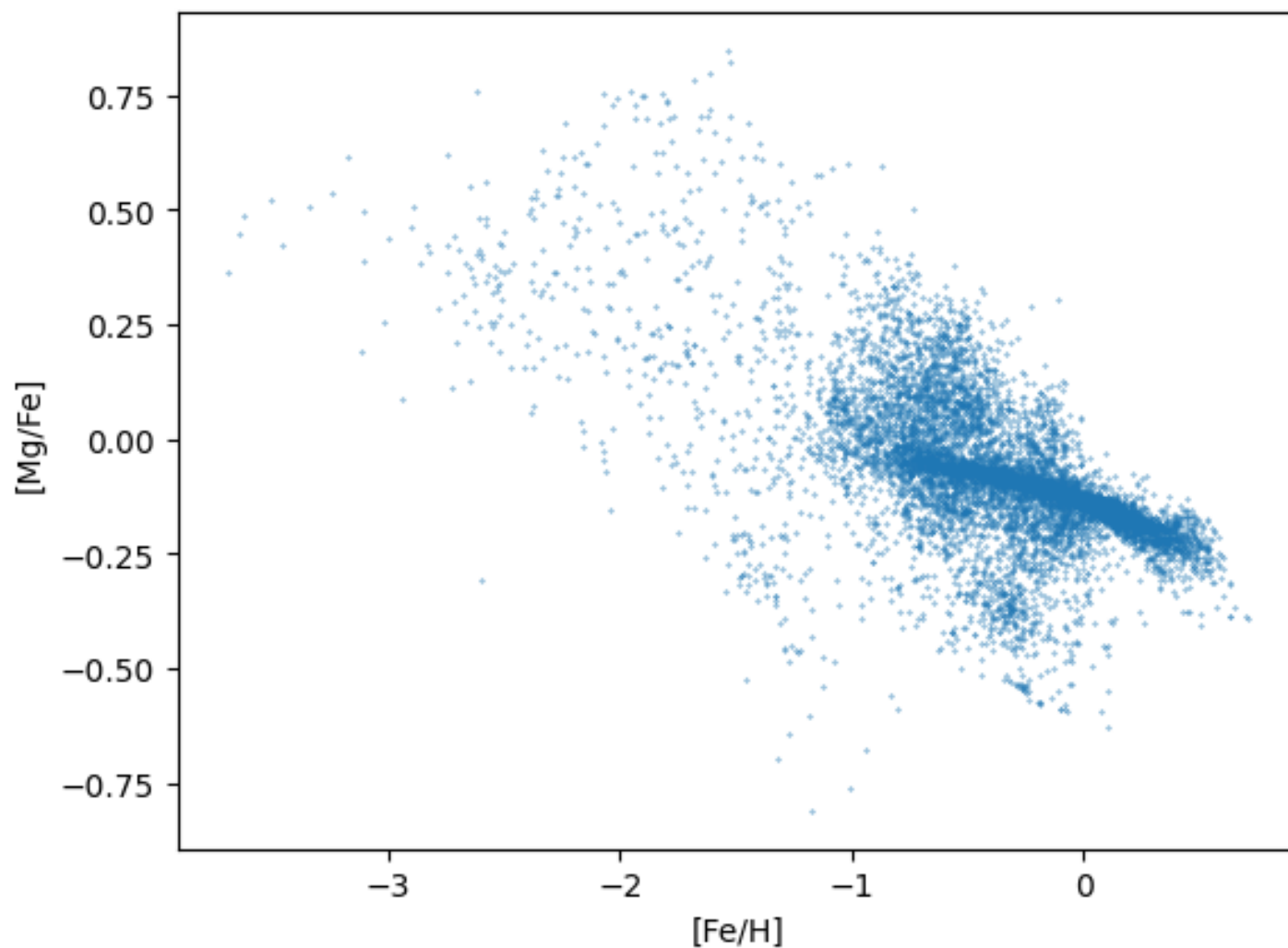
散布図

```
plt.xlabel('[Fe/H]') #x軸名  
plt.ylabel('[Mg/H]') #y軸名  
plt.scatter(df['FeH'], df['MgH'], s=0.1) #[Mg/H]と[Fe/H]と星の形成時刻の関係を表示
```



散布図

```
plt.xlabel('[Fe/H]') #x軸名  
plt.ylabel('[Mg/Fe]') #y軸名  
plt.scatter(df['FeH'], df['MgFe'], s=0.1) #[Mg/Fe]と[Fe/H]と星の形成時刻の関係を表示
```



今回の内容

Matplotlibについて

折れ線グラフ

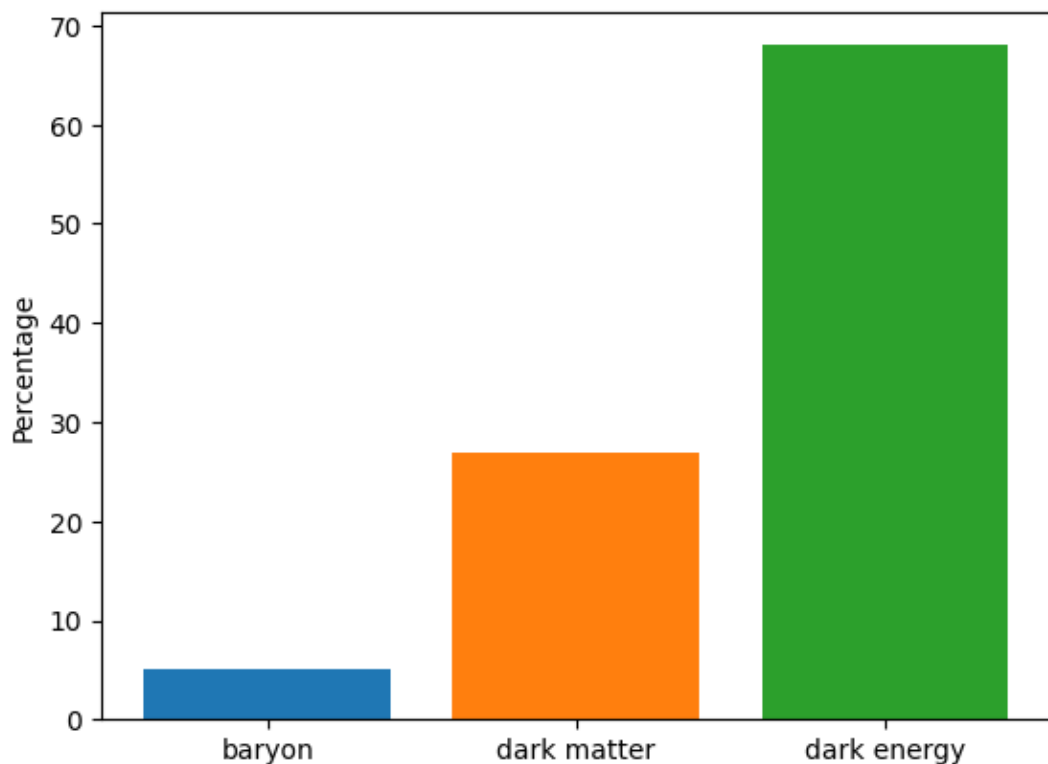
散布図

棒グラフ

ヒストグラム

棒グラフ

```
x = [1, 2, 3] #x軸のデータ  
y = [5, 27, 68] #y軸のデータ  
labels = ['baryon', 'dark matter', 'dark energy'] #ラベルの設定  
colors = ['tab:blue', 'tab:orange', 'tab:green'] #色を指定  
plt.ylabel('Percentage') #y軸名  
plt.bar(x, y, tick_label=labels, color=colors) #ラベルと色を指定して表示
```



今回の内容

Matplotlibについて

折れ線グラフ

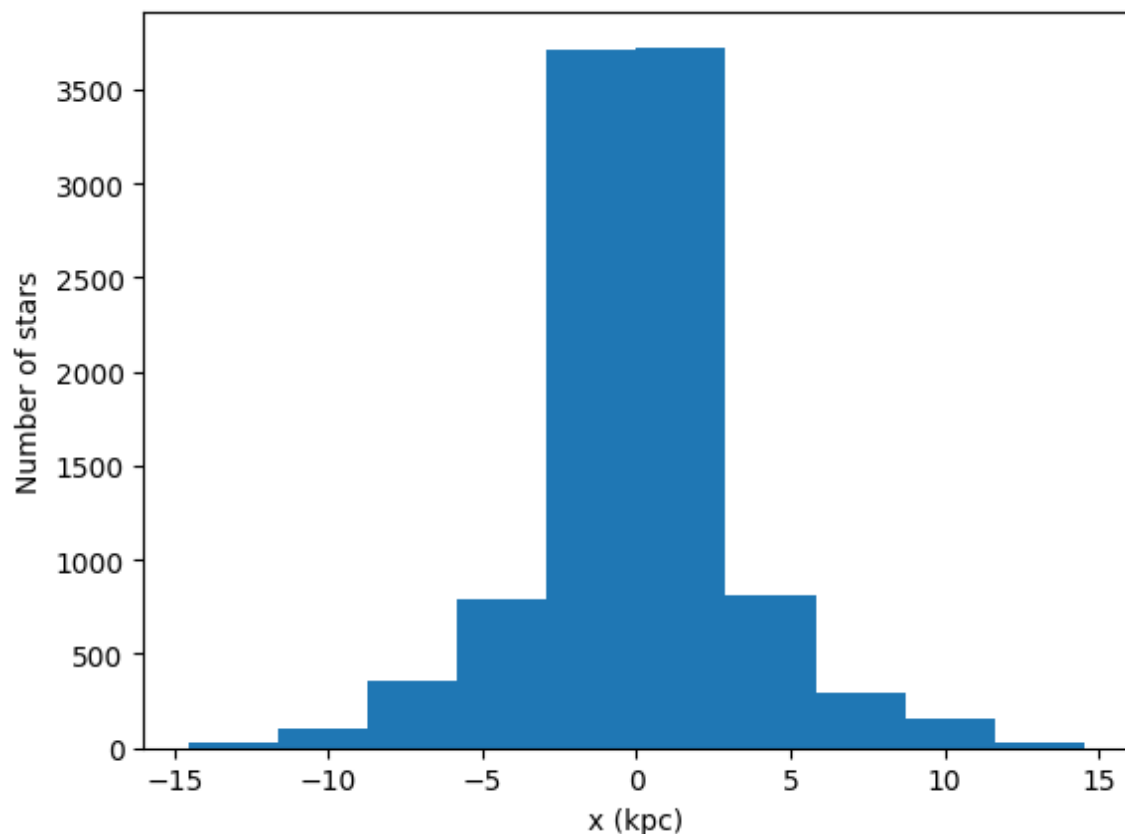
散布図

棒グラフ

ヒストグラム

ヒストグラム

```
plt.xlabel('x (kpc)') #x軸名  
plt.ylabel('Number of stars') #y軸名  
plt.hist(df['x']) #位置xについてのヒストグラム
```



速度 (v_x)、星の形成時刻、 $[\text{Fe}/\text{H}]$ についてのヒストグラムも描いてみよう

今回の目標

Matplotlibを用いて
基本的な図を描ける
ようになる。

今回の内容

Matplotlibについて

折れ線グラフ

散布図

棒グラフ

ヒストグラム

次回

- 第1回 授業の概要、自己紹介 (4/13)
- 第2回 天文学入門 138億年の宇宙の歴史 (4/20)
- 第3回 図書館ツアー、データベース検索 (4/27)
- 第4回 データ解析基礎 (1) プログラミング言語Python入門 (5/11)
- 第5回 データ解析基礎 (2) 基本統計量 (5/18)
- 第6回 データ解析実習 (1) 図の作成 (5/25)
- 第7回 **データ解析実習 (2) データの解釈 (6/1)**
- 第8回 レポートの作成 (1) 文章の読み方と要約、レポートの構成と表現 (6/8)
- 第9回 レポートの作成 (2) 著作権 (6/15)
- 第10回 プレゼンテーションの方法 (6/22)
- 第11回 発表準備 (1) 発表資料の作成 (6/29)
- 第12回 発表準備 (2) 発表練習 (7/6)
- 第13回 発表会 (7/13)