

基礎演習

第7回 データ解析実習（2）データの解釈

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

Matplotlibを用いて
基本的な図を描ける
ようになる。

前回の内容

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
```

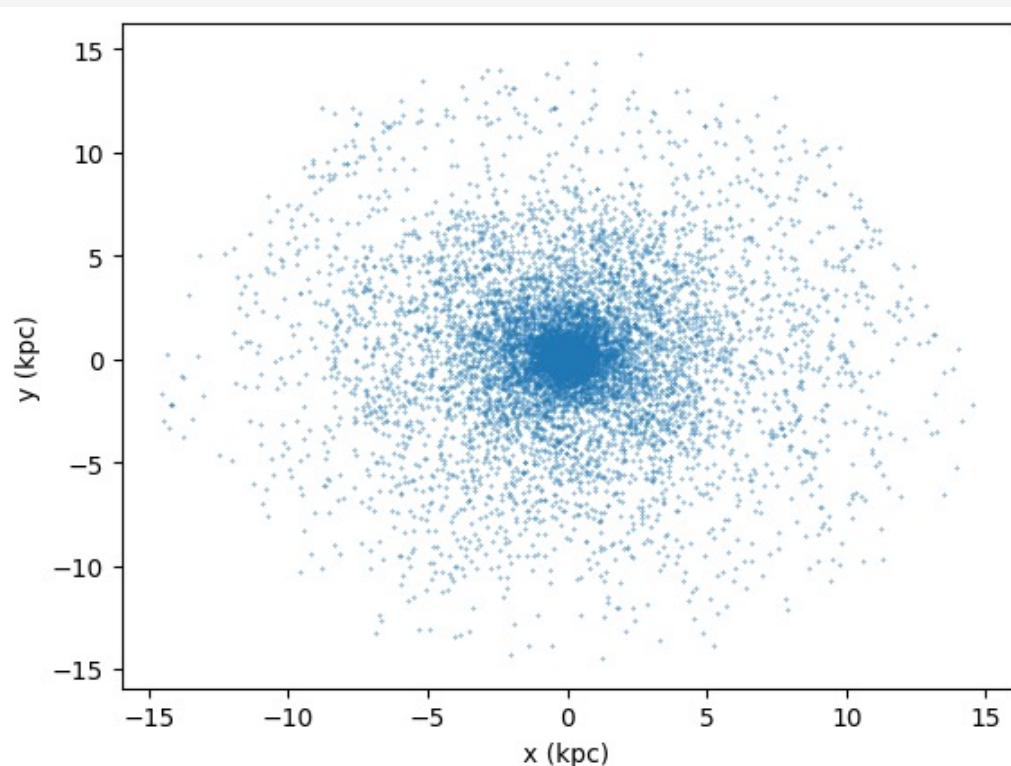
```
import matplotlib.pyplot as plt
```

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
```

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
```



棒グラフ

```
x = [1, 2, 3] #x軸のデータ
```

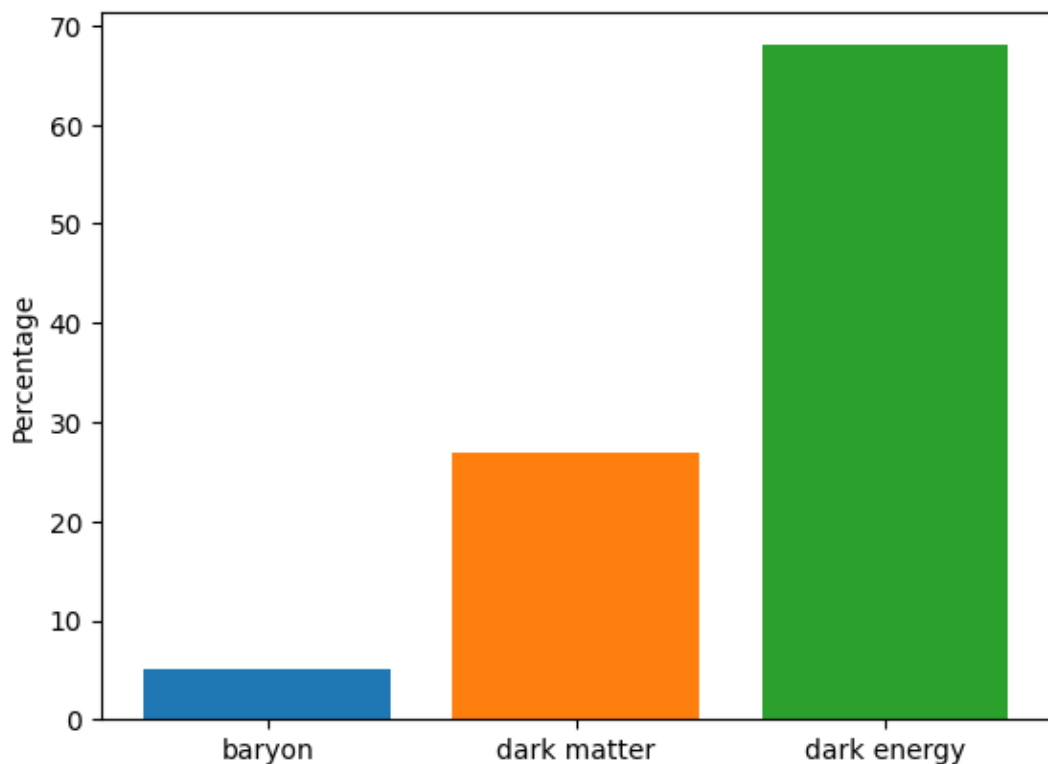
```
y = [5, 27, 68] #y軸のデータ
```

```
labels = ['baryon', 'dark matter', 'dark energy'] #ラベルの設定
```

```
colors = ['tab:blue', 'tab:orange', 'tab:green'] #色を指定
```

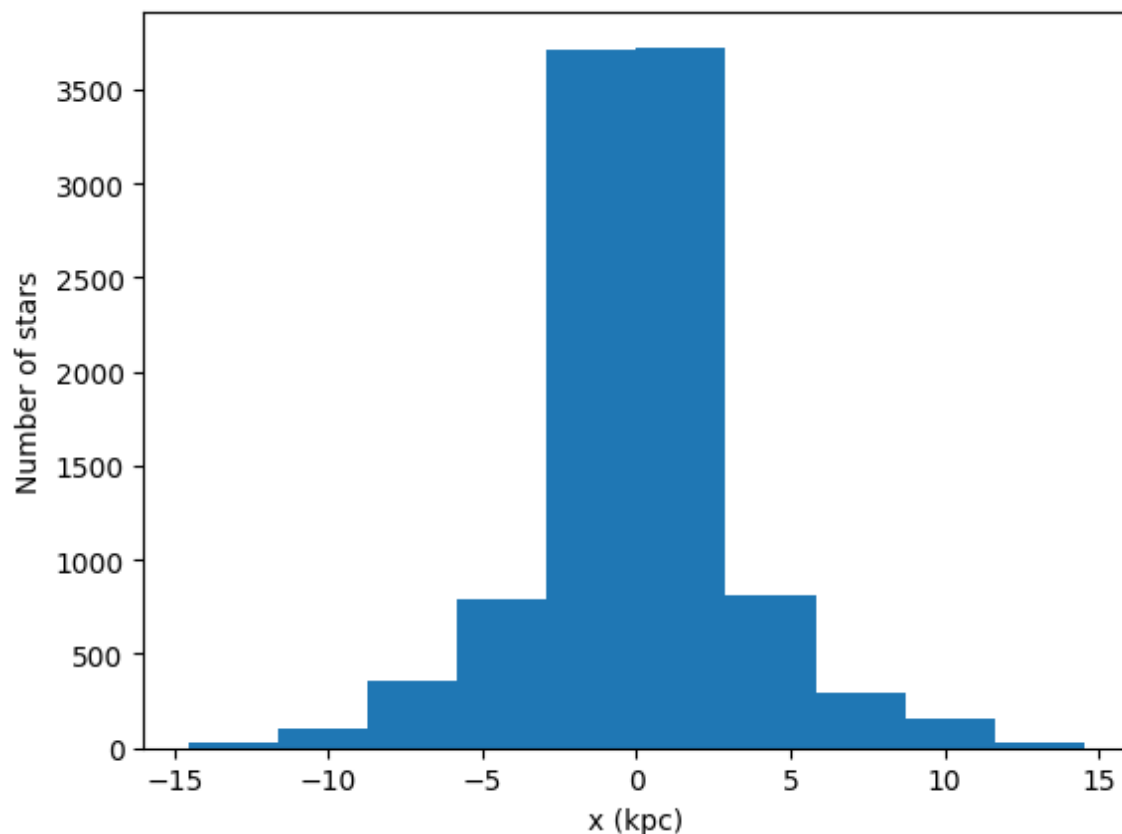
```
plt.ylabel('Percentage') #y軸名
```

```
plt.bar(x, y, tick_label=labels, color=colors) #ラベルと色を指定して表示
```



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

- 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

データ分析した結果
を解釈できるように
なる。

今回の内容

箱ひげ図

円グラフ

グラフのスタイル

データの解釈

今回の内容

箱ひげ図

円グラフ

グラフのスタイル

データの解釈

箱ひげ図

第三四分位数

中央値

第一四分位数



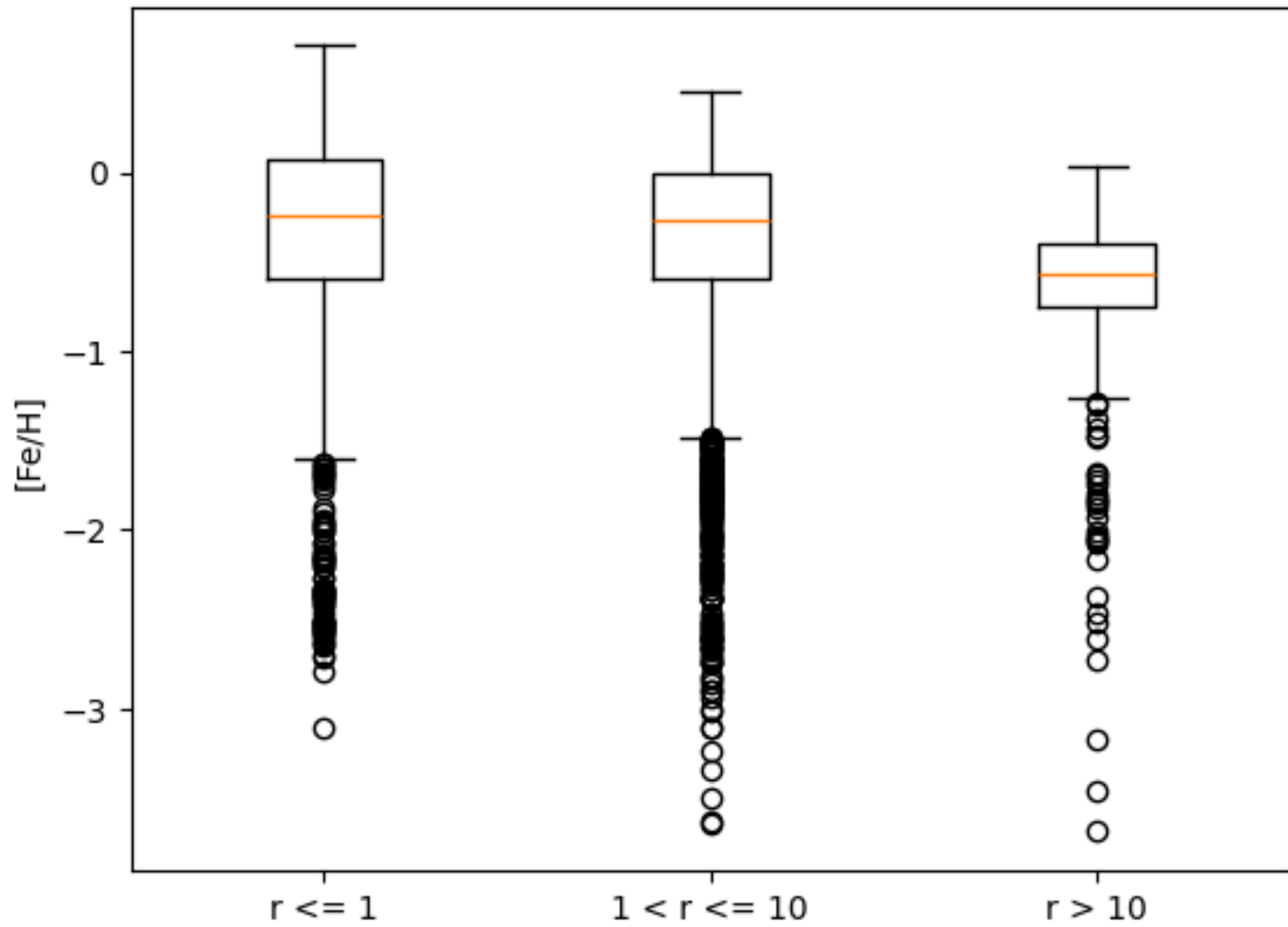
← 第三四分位数+1.5×四分位範囲より小さい
最大値

← 第一四分位数-1.5×四分位範囲より大きい
最小値

外れ値

箱ひげ図

```
#中心からの距離を計算してフィールド 'r' を追加
df = df.assign(r=np.sqrt(df['x']**2 + df['y']**2 + df['z']**2))
df1 = df[df['r'] <= 1] #中心から1kpc以内の星を選択
df2 = df[df['r'] > 1] #中心から1kpcより外側で
df2 = df2[df2['r'] <= 10] #10kpc以内の星を選択
df3 = df[df['r'] > 10] #中心から10kpcより外側の星を選択
plt.ylabel('[Fe/H]') #y軸名
labels = ['r <= 1', '1 < r <= 10', 'r > 10'] #データのラベルを指定
#[Fe/H]分布についての箱ひげ図を描画
plt.boxplot((df1['FeH'], df2['FeH'], df3['FeH']), labels=labels)
```



今回の内容

箱ひげ図

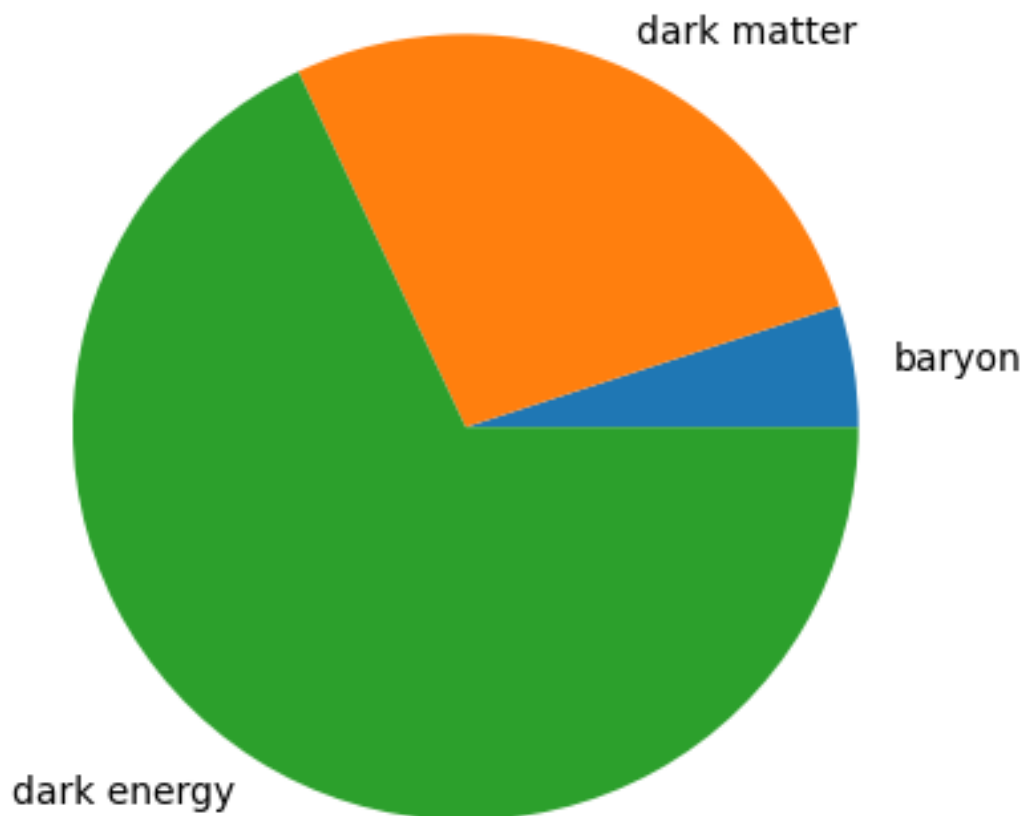
円グラフ

グラフのスタイル

データの解釈

円グラフ

```
data = [5, 27, 68] #円グラフのデータ  
labels = ['baryon', 'dark matter', 'dark energy'] #ラベルの設定  
plt.pie(data, labels=labels) #ラベルを指定して円グラフを表示
```



今回の内容

箱ひげ図

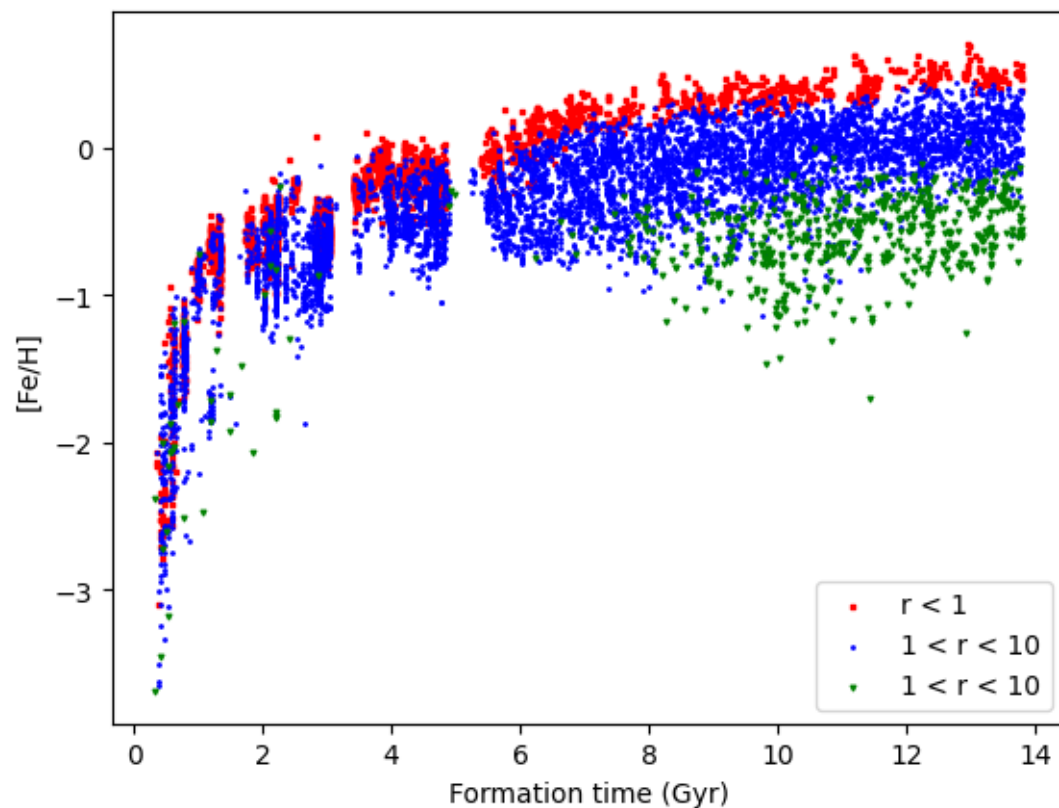
円グラフ

グラフのスタイル

データの解釈

散布図のスタイル

```
plt.xlabel('Formation time (Gyr)') #x軸名
plt.ylabel('[Fe/H]') #y軸名
plt.scatter(df1['t_form'], df1['FeH'], c='red', marker='.', s=2, label='r < 1')
plt.scatter(df2['t_form'], df2['FeH'], c='blue', marker='o', s=1, label='1 < r < 10')
plt.scatter(df3['t_form'], df3['FeH'], c='green', marker='v', s=3, label='1 < r < 10')
plt.legend() #凡例の表示
```



色の指定

“c” (または “color”) で指定

```
plt.scatter(df1['t_form'], df1['FeH'], c='red', marker=',', s=2, label='r < 1')
```

標準色



tab:blue



tab:orange



tab:green



tab:red



tab:purple



tab:brown



tab:pink



tab:gray



tab:olive



tab:cyan

その他の色















 black	 bisque	 forestgreen	 slategrey
 dimgray	 darkorange	 limegreen	 lightsteelblue
 dimgrey	 burlywood	 darkgreen	 cornflowerblue
 gray	 antiquewhite	 green	 royalblue
 grey	 tan	 lime	 ghostwhite
 darkgray	 navajowhite	 seagreen	 lavender
 darkgrey	 blanchedalmond	 mediumseagreen	 midnightblue
 silver	 papayawhip	 springgreen	 navy
 lightgray	 moccasin	 mintcream	 darkblue
 lightgrey	 orange	 mediumspringgreen	 mediumblue
 gainsboro	 wheat	 mediumaquamarine	 blue
 whitesmoke	 oldlace	 aquamarine	 slateblue
 white	 floralwhite	 turquoise	 darkslateblue
 snow	 darkgoldenrod	 lightseagreen	 mediumslateblue
 rosybrown	 goldenrod	 mediumturquoise	 mediumpurple
 lightcoral	 cornsilk	 azure	 rebeccapurple
 indianred	 gold	 lightcyan	 blueviolet
 brown	 lemonchiffon	 paleturquoise	 indigo
 firebrick	 khaki	 darkslategray	 darkorchid
 maroon	 palegoldenrod	 darkslategrey	 darkviolet
 darkred	 darkkhaki	 teal	 mediumorchid
 red	 ivory	 darkcyan	 thistle
 mistyrose	 beige	 aqua	 plum
 salmon	 lightyellow	 cyan	 violet
 tomato	 lightgoldenrodyellow	 darkturquoise	 purple
 darksalmon	 olive	 cadetblue	 darkmagenta
 coral	 yellow	 powderblue	 fuchsia
 orangered	 olivedrab	 lightblue	 magenta
 lightsalmon	 yellowgreen	 deepskyblue	 orchid
 sienna	 darkolivegreen	 skyblue	 mediumvioletred
 seashell	 greenyellow	 lightskyblue	 deeppink
 chocolate	 chartreuse	 steelblue	 hotpink
 saddlebrown	 lawngreen	 aliceblue	 lavenderblush
 sandybrown	 honeydew	 dodgerblue	 palevioletred
 peachpuff	 darkseagreen	 lightslategray	 crimson
 peru	 palegreen	 lightslategrey	 pink
 linen	 lightgreen	 slategray	 lightpink

マーカー（点の種類）の指定

“marker”で指定

```
plt.scatter(df1['t_form'], df1['FeH'], c='red', marker=',', s=2, label='r < 1')
```

"."		point
","		pixel
"o"		circle
"v"		triangle_down
"^"		triangle_up
"<"		triangle_left
">"		triangle_right
"1"		tri_down
"2"		tri_up
"3"		tri_left
"4"		tri_right
"8"		octagon
"5"		square

"p"		pentagon
"P"		plus (filled)
"*"		star
"h"		hexagon1
"H"		hexagon2
"+"		plus
"x"		x
"X"		x (filled)
"D"		diamond
"d"		thin_diamond
" "		vline
"_"		hline
0 (TICKLEFT)		tickleft
1 (TICKRIGHT)		tickright
2 (TICKUP)		tickup

https://matplotlib.org/stable/api/markers_api.html

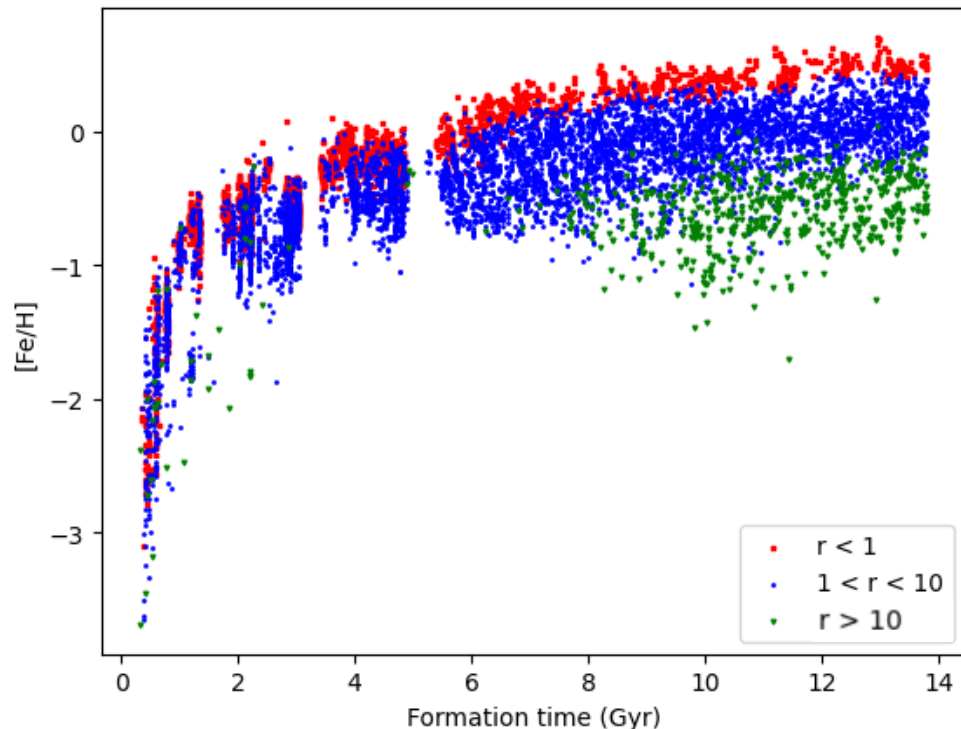
マーカーの大きさの指定

“s”で指定

```
plt.scatter(df1['t_form'], df1['FeH'], c='red', marker=',', s=2, label='r < 1')
```

凡例の表示

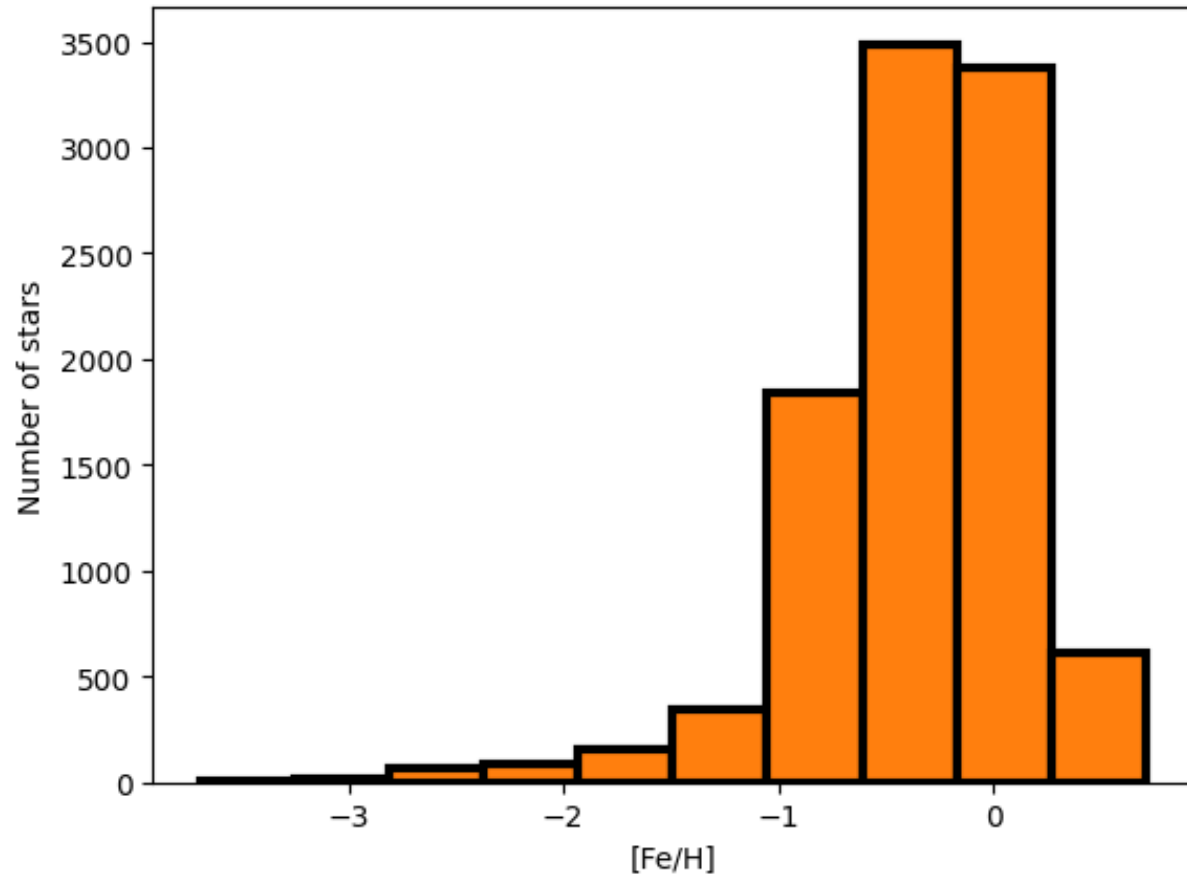
```
plt.xlabel('Formation time (Gyr)') #x軸名
plt.ylabel('[Fe/H]') #y軸名
plt.scatter(df1['t_form'], df1['FeH'], c='red', marker=',', s=2, label='r < 1')
plt.scatter(df2['t_form'], df2['FeH'], c='blue', marker='o', s=1, label='1 < r < 10')
plt.scatter(df3['t_form'], df3['FeH'], c='green', marker='v', s=3, label='r > 10')
plt.legend() #凡例の表示
```



凡例

ヒストグラムのスタイル

```
plt.xlabel('[Fe/H]') #x軸名  
plt.ylabel('Number of stars') #y軸名  
plt.hist(df['FeH'], color='tab:orange', edgecolor='black', linestyle='solid', linewidth=3)
```



ヒストグラムのスタイル

```
plt.xlabel('[Fe/H]') #x軸名  
plt.ylabel('Number of stars') #y軸名  
plt.hist(df['FeH'], color='tab:orange', edgecolor='black', linestyle='solid', linewidth=3)
```

ヒストグラム内
の色

ヒストグラム外
枠の色

ヒストグラム外
枠のスタイル

ヒストグラムのスタイル

```
plt.xlabel('[Fe/H]') #x軸名  
plt.ylabel('Number of stars') #y軸名  
plt.hist(df['FeH'], color='tab:orange', edgecolor='black', linestyle='solid', linewidth=3)
```

ヒストグラム外
枠のスタイル

ヒストグラム外
枠の太さ

“solid” 実線
“dashed” 破線
“dotted” 点線
“dashdot” 一点鎖線

図の保存

```
plt.xlabel('[Fe/H]') #x軸名  
plt.ylabel('Number of stars') #y軸名  
plt.hist(df['FeH'], color='tab:orange', edgecolor='black', linestyle='solid', linewidth=3)  
plt.savefig('histogram.png', format='png')
```

ファイル名

ファイル形式

今回の内容

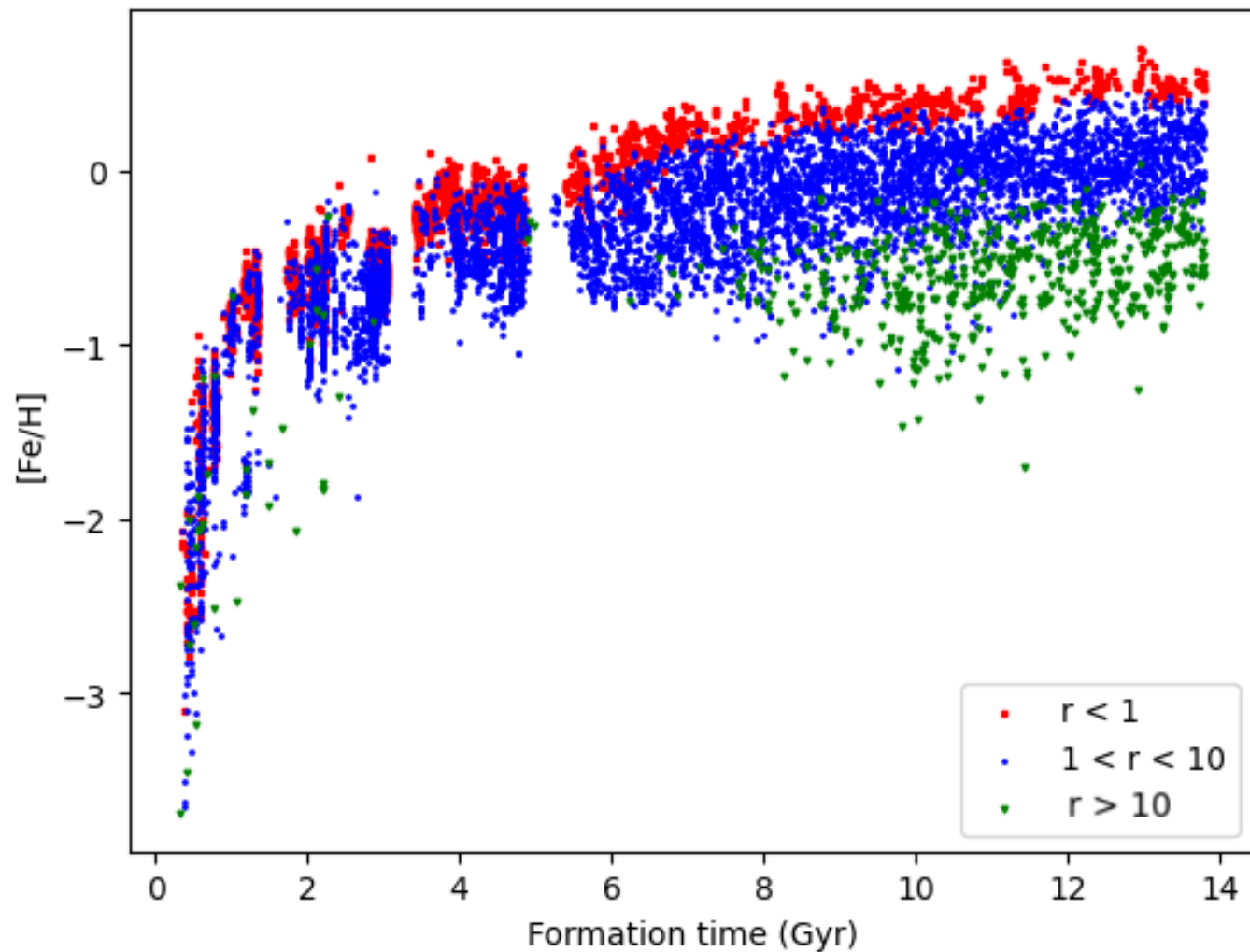
箱ひげ図

円グラフ

グラフのスタイル

データの解釈

データの解釈



この図から何がわかるか？

追加でどんな解析をすればいい？

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)