12. 作図の基本

プログラミング・データサイエンス I

2021/7/8

1　今日の目的

今日の目的

- 作図の基本
  - 一枚の図を描く
  - 複数の図を描く
- 図形を描く

Python は、データ分析などで活用事例の多いプログラミング言語です。分析結果を図示することで、データから読み取れることを説明することができます。今日は、作図の基本について説明します。`matplotlib` という作図ライブラリを使うことにします。マニュアルは以下の URL にあります。

`https://matplotlib.org/contents.html`

はじめに、一枚の図を描く方法を紹介します。データを点で描く、折れ線で結ぶなどの基本的な方法から始め、図の大きさ、フォントサイズ、作図範囲の指定などの書式設定までを扱います。指定できることは非常に多岐にわたります。上記 URL のマニュアルを見るのが基本ですが、インターネット上にも沢山の例があります。

次に複数の図を描く方法を紹介します。異なる量を横軸や縦軸を揃えて見せたい時に便利です。一枚の図を描く方法とは、コマンド名が少し異なることに注意が必要です。

最後に、図形の描き方を紹介します。

今日のサンプルプログラムをダウンロードしてください。

`https://github.com/first-programming-saga/plotSample`

2　データを一枚の図に作図する

データを一枚の図に作図する

- 折れ線
- 散布図
- 図の書式設定
  - タイトル
  - 軸の名前と範囲
  - 凡例
- 図の表示と保存

作図の技術的内容の前に、基本的なことを確認しておきましょう。

図でデータを示すのは、それを使って読者に何かを伝えることが目的です。不適切な方法で図示すると、誤った情報や印象を伝えることになります。凝った見た目のグラフを作成することで、特定の部分を強調してしまうこともあります。

また、図の種類は、データの性質に応じて選択するべきです。例えば、人口の時間的推移を表すならば横軸に時間を縦軸に人口をとって、折れ線グラフとするべきでしょう。あるいは、企業の従業員数と職員の平均年収の関係を示したいならば、データを点で表示する散布図を使います。各県の人口を比較したいならば、横軸を県名、縦軸を人口とした棒グラフを使います。

縦軸や横軸の範囲の選び方も重要です。人口の変化を表す際に、縦軸の範囲を狭くとると、大きく変動した印象を与えます。範囲を広くとると、変化がほとんどないように見えます。

もちろん、図にはタイトルを付け、x 軸と y 軸にはラベルをつけなければいけません。単位がある量ならば、単位を示さなければいけません。見やすいように、色、線の太さ、点の大きさを調整することも必要です。

それでは、`simplePlot.ipynb` を開いてください。最初の部分でライブラリをインポートしています (ソースコード 2.1)。`matplotlib` が最初に紹介した作図用ライブラリです。`japanize_matplotlib` は、図の中で日本語を使えるようにするライブラリです。

ソースコード 2.1　ライブラリのインポート

```
1 import matplotlib.pyplot as plt
2 import japanize_matplotlib
```

ソースコード 2.2　作図を行う

```
1  def drawData(xList, yList, zList):
2      plt.figure(figsize = (15, 10))#図の大きさ
3      plt.rcParams["mathtext.fontset"] = 'cm'
4      plt.rcParams['mathtext.default'] = 'it'
5      plt.rcParams["font.size"] = 28#フォントサイズ
6      plt.title('グラフタイトル')
7      plt.xlim(−1, 6)#x 軸の範囲
8      plt.ylim(0, 8)#y 軸の範囲
9      plt.xlabel('$x$')#x 軸のラベル
10     plt.ylabel('$y$')#y 軸のラベル
11
12     plt.plot(xList, yList, label = '理論', linewidth = 2)#折れ線
13     plt.scatter(xList, zList, label = 'データ',
14         color = 'red', marker = 's', linewidths = 5)#散布図
15
16     plt.legend(loc = 'upper␣right')#凡例の位置
17     plt.savefig('tmp.pdf')#ファイルへ出力
18     plt.show()
```

ソースコード 2.2 を見てください。2 行目から 8 行目までが、図の書式を定めている部分です。上から、図の大きさ、フォントサイズ、図のタイトル、x 軸の範囲、y 軸の範囲、x 軸のラベル、そして y 軸のラベルです。xは、数式として x を表現することを示しています。3 行目と 4 行目は、数式用のフォントの設定です。

12 行目は、xList というリストを x 軸の値に、yList というリストを y 軸の値として、折れ線でデータを結びます。label という変数を指定すると、凡例にその文字列を表示します。linewidth は線の太さです。

13 行目は、データを点で表します。color は点の色、marker は点の形、linewidth は点の大きさです。使える色と点の形は、以下の URL にあります。

color　https://matplotlib.org/3.1.0/gallery/color/named_colors.html
marker　https://matplotlib.org/3.1.0/api/markers_api.html?highlight=list%20markers

16 行目は、凡例の位置を指定しています。loc='best' とすると、勝手に場所を決めてくれます。

作図した図をファイルに保存し (17 行目)、表示しています (18 行目)。ファイルへの保存の際には、拡張子に合わせた形式になります。ここでは、PDF という形式です。ファイルに出力した結果が、図 1 です。

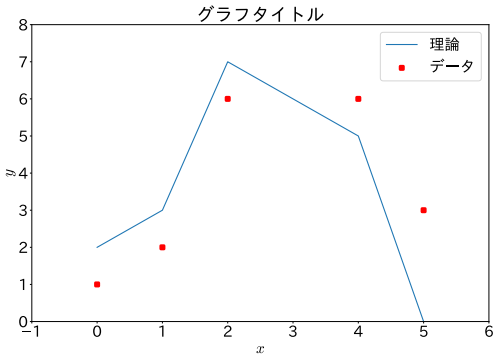

図 1　ファイルへの出力結果

課題 1　ソースコード 2.2 において、以下の変更を行ってみなさい。

1. y 軸の範囲を変更
2. 折れ線の色を設定
3. 散布図の点の形を変更

3　複数の図を描く

複数の図を描く

- 一つの図に、複数のグラフを描く
- x 軸や y 軸を共有する

複数の図を軸を共有しながら描きたいことがあります。その場合は、`subplots()` を使用します。戻り値のうち `ax` は `Axes` という型のリストで、各要素が一つの図の領域を表します。

ソースコード 3.1　二つの図を左右に並べる

```
1 def draw(xList,yList,zList):
2     fig, ax = plt.subplots(1, 2, figsize = (20, 10), sharey = True)
3     plt.rcParams["mathtext.fontset"] = 'cm'
4     plt.rcParams['mathtext.default'] = 'it'
5     plt.rcParams['font.size'] = 28#フォントサイズ
6     plt.yticks([-1 + .5 * i for i in range(6)])#y 軸の目盛り
7     #二つの図に共通の設定
8     for a in ax:
```

```
9           a.set_xlim(−5, 5)
10          a.set_xlabel('$x$')
11          a.tick_params(axis = 'both', length = 15)
12
13      #最初の図
14      ax[0].set_title('$\sin(x)$')
15      ax[0].plot(xList, yList)
16      #2番目の図
17      ax[1].set_title('$\cos(x)$')
18      ax[1].plot(xList, zList)
19
20      plt.savefig('multiPlot.pdf')
21      plt.show()
```

それでは、multiPlot.ipynb を見ましょう。作図をしている部分がソースコード 3.1 です。subplot の最初の二つの引数 1,2 が一行二列の作図領域を作ることを表しています。つまり、横に 2 つ図が並びます。作図領域は ax として返ってきます。今回は、ax[0] と ax[1] です。最後の引数 sharey = True が y 軸を共有することを示しています。

3 行目と 4 行目は、数式フォントの設定です。6 行目は、y 軸の目盛り (tick と言います) とラベル設定です。-1 から 0.5 刻みで、1 まで目盛りを作ります。

8 行目から 11 行目は、二つの図に共通の要素を指定しています。ax がリストであったことを思い出してください。x 軸の範囲、x 軸のラベル、そして目盛りの大きさです。

14 行目と 15 行目で ax[0] に、17 行目と 18 行目で ax[1] に、それぞれタイトルを付け、データを折れ線で描いています。

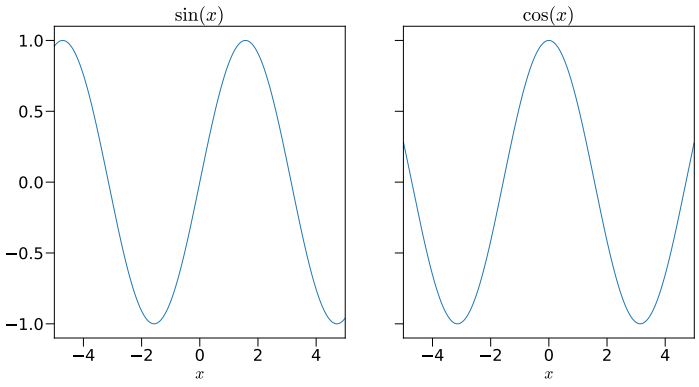

図 2　二つの図を描く

作図結果を図 2 に示します。変数名 x がきれいなイタリック体で、三角関数の名前がローマン体で表示されています。

次は、4 つの図を、x 軸と y 軸を共通として描く例です。multiPlot2.ipynb を開けてください (ソースコード 3.2)。

3 行目の最初の二つの引数 2,2 で、2×2 という四つの図を作ることを指定しています。最後の二つの引数が x 軸と y 軸を共通として描くことの指定です。結果として返ってくる ax も 2×2 のリストです。四つの図に対する共通的設定を 11 行目から 17 行目で行っています。11 行目の最初の for は、行に関する繰り返しです。12 行目の内側の for は、列に関する繰り返しです。繰り返し方が違いますが、理解できますか。16 行目では、行の番号を見て、一番最後の行に対してだけ、x 軸のラベルを描いています。

19 行目から、4 つの図を折れ線で描いています。最後に、25 行目からの二重 for ループで、凡例を作っています。作図結果は図 3 のようになります。

ソースコード 3.2　4 枚の図を描く

```
1  def draw(x, y1, y2, y3, y4):
2      #xy 軸のスケールを揃える
3      fig, ax = plt.subplots(2, 2, figsize = (20, 20),
4          sharex = True, sharey = True)
5      plt.rcParams["mathtext.fontset"] = 'cm'
6      plt.rcParams['mathtext.default'] = 'it'
7      plt.rcParams['font.size'] = 28#フォントサイズ
8
9      fig.suptitle('マルチプロット例')
10     #全ての図に共通の設定
11     for i in range(len(ax)):
12         for b in ax[i]:
13             b.set_xlim(−5, 5)
14             b.set_ylim(−4, 4)
15             b.tick_params(axis = 'both', length = 15)
16             if i == 1:#2行目だけx 軸ラベルを表示
17                 b.set_xlabel('$x$')
18
19     ax[0,0].plot(x, y1, label = '$\sin(x)$', color = 'blue')
20     ax[1,0].plot(x, y2, label = '$\cos(x)$', color = 'green')
21     ax[0,1].plot(x, y3, label = '$\sin(x)␣\cos(x)$', color = 'darkcyan')
22     ax[1,1].plot(x, y4, label = '$x␣\cos(x)$', color = 'red')
23
24     #全ての図の凡例
25     for a in ax:
26         for b in a:
27             b.legend(loc = 'best')
28
29     plt.savefig('multiPlot2.pdf')
30     plt.show()
```

マルチプロット例

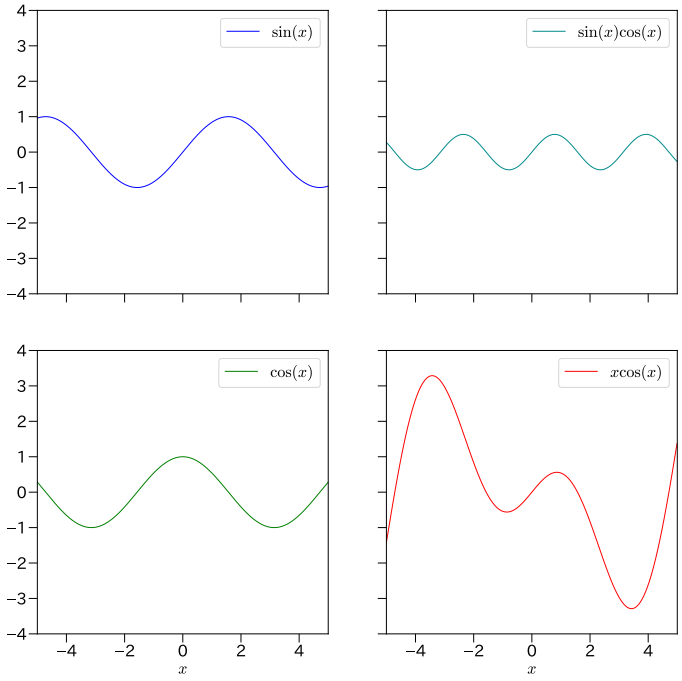

図 3　四つの図を描く

4　その他の作図

その他の作図

- 棒グラフ
- 円グラフ
- ヒストグラム

matplotlib では、様々な型のグラフを描くことができます。ここでは、三種類を示します。

最初は、棒グラフです。plotBars.ipynb を見てください (ソースコード 4.1)。例として、九州各県の人口を棒グラフで表しましょう。

1 行目のリスト population は、九州各県の 2015 年の人口です。棒グラフの高さに相当します。2 行目のリスト prefecture は、九州各県の名前です。棒グラフの横軸に相当します。

9 行目から 12 行目に書けて、軸のラベルや目盛りを設定しています。

14 行目の bar() の最初の引数が横軸、二番目の引数が縦軸です。ただし、横軸は、リスト population の番号を示しているに過ぎません。リスト x は 4 行目で生成しています。横軸の名前は tick_label として渡しています。出力結果が図 4 です。

ソースコード 4.1　棒グラフ

```
1  population = [5120, 847, 1413, 1818, 1191, 1136, 1691]
2  prefecture = ['福岡', '佐賀', '長崎', '熊本',
3                '大分', '宮崎', '鹿児島']
4  x = [i for i in range(len(population))]
5
6  plt.figure(figsize = (15, 15))
7  plt.rcParams['font.size'] = 28
8
9  plt.title('九州の人口, 2015', fontsize = 28)
10 plt.xlabel('県', loc = 'right')
11 plt.ylabel('人口 (千人)', loc = 'top')
12 plt.tick_params(axis = 'both', length = 15)
13
14 plt.bar(x, population, tick_label = prefecture)
15
16 plt.savefig('plotBar.pdf')
17 plt.show()
```

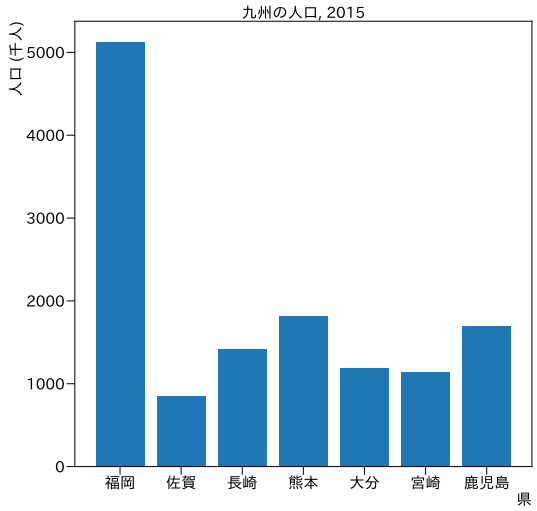

図 4　棒グラフ

次の例は、円グラフです。英語では、パイグラフと言います。お菓子のパイのことです。piPlot.ipynb を開けてください (ソースコード 4.2)。佐賀県の人口構成の変化を表すグラフです。人口を年少 (14 歳以下)、生産年齢 (15 歳以上、65 歳未満)、そして老年 (65 歳以上) に分けて、1995 年と 2015 年で比較します。

ソースコード 4.2　円グラフ

```
1  data2015 = [116122, 483019, 229335]
2  data1995 = [160307, 566671,157329]
3  label = ['年少', '生産年齢', '老年']
4  color = ['y', 'c', 'r']
5
6  fig,ax = plt.subplots(1, 2, figsize = (20, 10))
7  plt.rcParams['font.size'] = 24 #文字の大きさ
8  fig.suptitle('佐賀県の人口') #全体のタイトル
9
10 #1995年の人口比率
11 ax[0].pie(data1995, labels = label, colors = color,
12         startangle = 90, counterclock = False, autopct = '%1.1f%%')
13 ax[0].set_title('1995')
14 #2015年の人口比率
15 ax[1].pie(data2015, labels = label, colors = color,
16         startangle = 90, counterclock = False, autopct = '%1.1f%%')
17 ax[1].set_title('2015')
18
19 plt.savefig('piPlot.pdf')
```

```
20 plt.show()
```

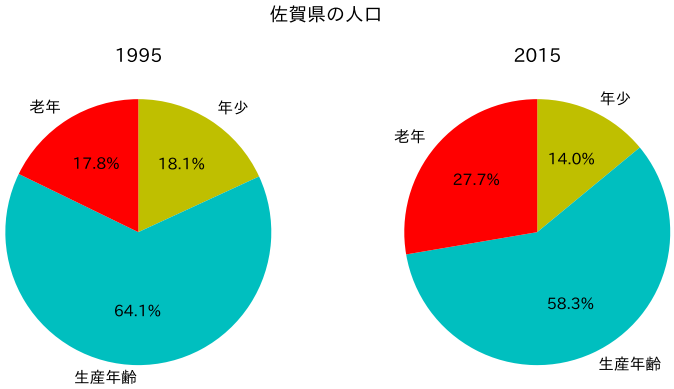

図 5　円グラフ

リスト data に、1995 年と 2015 年の、三つの分類に相当する人口のリストを記述しています。リスト years は、タイトルに使う 1995 年と 2015 年の文字列です。6 行目で、一行二列の図を作っています。

9 行目からの for ループで、二つの図を描いています。10 行目が円グラフを描く命令です。最初の引数が描くデータ、二番目が各データに対応するラベルです。一番最後が、数値を描く書式の指定で、小数以下一桁まで百分率で表すことを指定しています。

結果を図 5 に示します。matplotlib の円グラフのメソッド pie() は、データを与えると、自動的に百分率を計算してくれていることにも注意してください。

最後の例はヒストグラムです。histogram.ipynb を開いてください (ソースコード 4.3)。ヒストグラム (histogram) とは、頻度を表す図のことです。

ソースコード 4.3　ヒストグラム

```
1  #元データ
2  data = [2.5, 4., 5.5, 9.2, 0.7, 8.8, 6.1, 7.1, 4.3, 1.6,
3          3.4, 1.4, 7.7, 4.4, 5.5, 9.6, 6.8, 8.6, 2.1,
4          8.9, 2.9, 1.1, 2.1, 3.7, 2.0, 1.4, 0.5, 5.7]
5  plt.figure(figsize = (15, 15))
6  plt.rcParams['font.size'] = 24
7  plt.title('ヒストグラムの例')
8
9  numBins = 10 #bin の数
10 rWidth = 0.9 #描く幅は、等間隔の幅に対して 0.9倍
11 plt.hist(data, range=(0, 10), bins = numBins,
```

```
12          rwidth = rWidth, color = 'cyan')
13
14  plt.savefig('histogram.pdf')
15  plt.show()
```

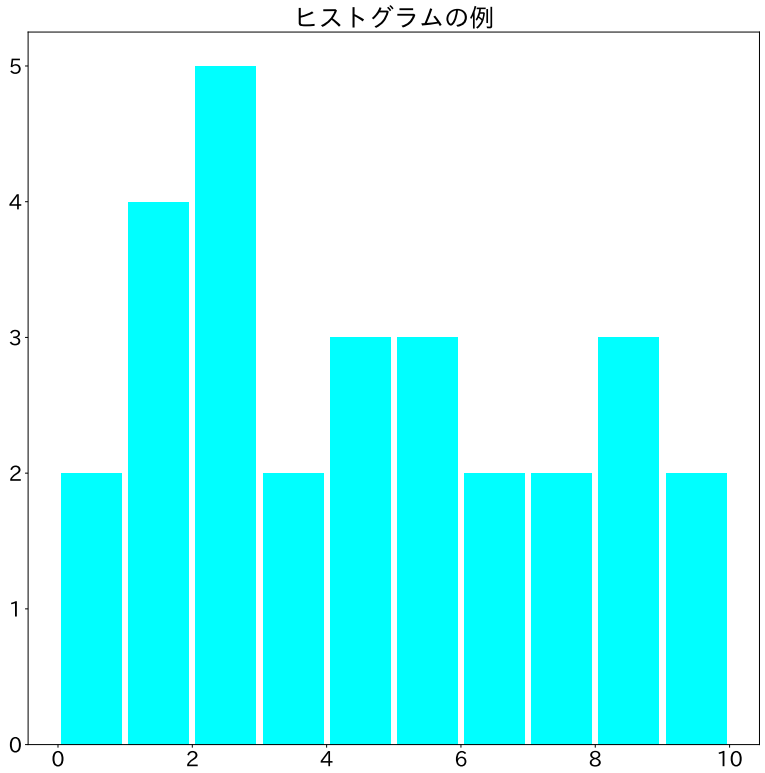

図 6　ヒストグラム

1 行目の `data` には、0 以上 10 未満の小数のデータが入っています。このデータを 0.1 刻みの区間 (bin という) に分けて、各 bin に何個入っているかを図示しましょう。通常では、各 bin に入っている数を数えるプログラムを書く必要がありますが、`matplotlib` を使えば、自動的に数えてくれます。

9 行目の `hist()` メソッドを見てください。最初から、データ、データの区間、bin の数を指定します。`rwidth` は、作図する際に bin の間隔に対する棒グラフの相対的な大きさです。

結果は、図 6 のようになります。

5 図形を描く

図形を描く

- 既定の図を描く
- 点列から図を描く

作図の最後に、図を描くことを考えましょう。drawShapes.ipynb を開けてください。

図を描くには、二つの方法があります。一つ目は、既存の図形を描く方法です。ソースコード 5.1 の 3 行目から 8 行目は、円、四角形、正多角形という既存の図形を描いています。既存の図形の描き方は以下の URL にあります。

https://matplotlib.org/api/patches_api.html

ソースコード 5.1　図形を作図

```
1  def drawPatches(ax):#図形を生成
2      patches=[]
3      #円
4      patches.append(pts.Circle((2,2),1,color='red'))
5      #四角
6      patches.append(pts.Rectangle((4,7),4.,2.,color='blue'))
7      #正多角形
8      patches.append(pts.RegularPolygon((7,3),5,2,color='yellow'))
9      #任意の点の列
10     verts = [(2.,8.),(3,4),(4,9),(6,2),(8,3)]
11     verts.append(verts[0])
12     path = createPath(verts)
13     patches.append(path)
14     #図形を追加
15     for p in patches:
16         ax.add_patch(p)
```

ソースコード 5.2　自分で点列を定義する

```
1  def createPath(vertList):#頂点の列から多角形を生成
2      n = len(vertList)
3      #codes は、各座標への操作
4      codes = [Path.MOVETO]#最初は移動
5      for i in range(n−2):#残りは線を引く
6          codes.append(Path.LINETO)
7      codes.append(Path.CLOSEPOLY)#最後に閉じる
8      path = Path(vertList,codes)#座標と操作を与える
9      return pts.PathPatch(path,facecolor='orange',lw=2)
```

もう一つの方法は、座標の列を使って定義する方法です。ソースコード 5.1 の 10 行目で、座標の列を与えています。注意が必要なのは、最後に出発点の座標を追加しておく必要があることです (11 行目)。

図を定義する過程を説明します。少しイメージしにくいかも知れません。昔、プリンターで図が描けなかった頃、図を描く専用のプロッター (plotter) という機材がありました。

https://ja.wikipedia.org/wiki/%E3%83%97%E3%83%AD%E3%83%83%E3%82%BF%E3%83%BC

プロッターの操作は、

moveto　ペンを、線を描かずに指定した座標に移動する
lineto　ペンを線を引きながら、指定した座標に移動する
close　指定した座標に向かって図形を閉じる

という動作から構成します。これは、今でも、作図用のライブラリで共通に使われています。最後の動作は、最初の点に向けて図形を閉じるというものが多いのですが、matplotlib では、最後の点も指定する必要があります。

ソースコード 5.1 を見てください。点を結んで図形を描くために、9 行目で点のリストを定義します。10 行目では、始点を最後に追加しています。このリストを使って図形を、ソースコード 5.2 で定義します。最初の点へ MOVETO したあと、順次 LINETO し、最後に CLOSEPOLY しています。

図を描くために、matplotlib を使う意味を理解しがたいかもしれません。

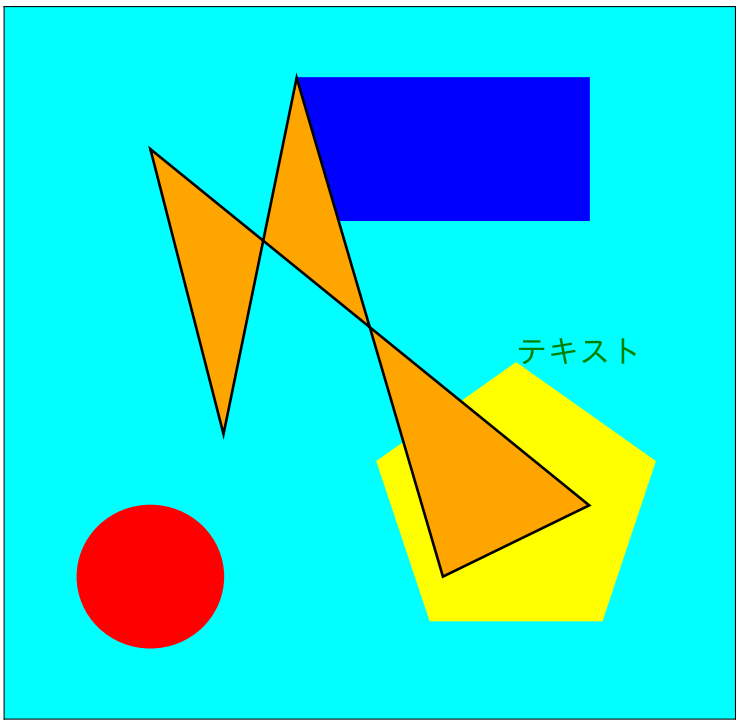

図 7　作図