

## Chapter 6 - Exercise 4: Women in Science

Cho các dữ liệu *year*, *physical\_sciences*, *computer\_science*, *health*, *education* từ tập tin *women\_in\_science.csv*

- **physical\_sciences** là tỷ lệ phần trăm bằng Khoa học Vật lý được trao cho phụ nữ mỗi năm tương ứng
- **computer\_science** là tỷ lệ phần trăm bằng Khoa học máy tính được trao cho phụ nữ mỗi năm tương ứng
- **health** là tỷ lệ phần trăm phụ nữ tham gia các công việc liên quan đến sức khỏe tương ứng mỗi năm
- **education** là tỷ lệ phần trăm phụ nữ tham gia các công việc liên quan đến giáo dục tương ứng mỗi năm

## Yêu cầu

### Câu 1:

- Vẽ biểu đồ line plot thể hiện tỷ lệ % bằng Khoa học vật lý và Khoa học máy tính được trao cho phụ nữ qua các năm
- Đồ thị có 2 line:
  - line 1: `year, physical_sciences, color='blue';` và
  - line 2: `year, computer_science, color='red'`
- Bạn nhận xét gì về biểu đồ vừa vẽ

### Câu 2:

- Vẽ 2 biểu đồ line plot ở câu 1 nhưng trên 2 vùng:
  - vùng 1: `plt.axes([0.05, 0.05, 0.425, 0.9])`,
  - vùng 2: `plt.axes([0.525, 0.05, 0.425, 0.9])`

### Câu 3:

- Vẽ 2 biểu đồ line plot ở câu 1 nhưng trên 2 subplot:
  - subplot 1: `plt.subplot(1, 2, 1)`,
  - và subplot 2: `plt.subplot(1, 2, 2)`,
- Lưu ý: sử dụng `plt.tight_layout()` trước khi `show()`

### Câu 4:

- Vẽ 4 biểu đồ line với 4 màu khác nhau:
  - Biểu đồ 1: thể hiện tỷ lệ % bằng Khoa học vật lý được trao cho phụ nữ qua các năm (`year - physical_sciences`)
  - Biểu đồ 2: thể hiện tỷ lệ % bằng Khoa học máy tính được trao cho phụ nữ qua các năm (`year - computer_science`)
  - Biểu đồ 3: thể hiện tỷ lệ % phụ nữ tham gia các công việc liên quan đến y tế qua các năm (`year - health`)

- Biểu đồ 4: thể hiện tỷ lệ % phụ nữ tham gia các công việc liên quan đến giáo dục qua các năm (year - education)
- Trên 4 subplot: plt.subplot(2, 2, 1), plt.subplot(2, 2, 2), plt.subplot(2, 2, 3), plt.subplot(2, 2, 4)
- Bạn nhận xét gì về biểu đồ vừa vẽ

## Câu 5:

- Vẽ 2 biểu đồ line plot:
  - Biểu đồ 1: thể hiện tỷ lệ % bằng Khoa học máy tính được trao cho phụ nữ qua các năm (year - computer\_science)
  - Biểu đồ 2: thể hiện tỷ lệ % bằng Khoa học vật lý được trao cho phụ nữ qua các năm (year - physical\_sciences) nhưng giới hạn từ năm 1980-2000 --> có plt.xlim(1980, 2000) và plt.ylim(0, 50)

## Câu 6:

- Vẽ 2 biểu đồ line plot:
  - Biểu đồ 1: thể hiện tỷ lệ % bằng Khoa học máy tính được trao cho phụ nữ qua các năm (year - computer\_science)
  - Biểu đồ 2: thể hiện tỷ lệ % bằng Khoa học vật lý được trao cho phụ nữ qua các năm (year - physical\_sciences) nhưng giới hạn từ năm 1990-2010 --> có plt.xlim(1990, 2010) và plt.ylim(0, 50)
- Bạn nhận xét gì về biểu đồ vừa vẽ

```
In [1]: 1 import numpy as np
        2 import pandas as pd
        3 import matplotlib.pyplot as plt
```

```
In [2]: 1 df = pd.read_csv('data/women_in_science.csv', index_col=0)
```

```
In [3]: 1 df.shape
```

```
Out[3]: (42, 5)
```

```
In [4]: 1 df.head()
```

```
Out[4]:
```

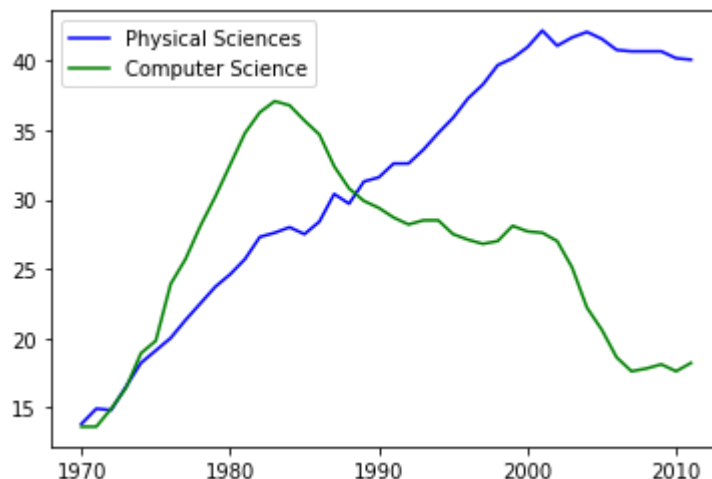
	year	physical_sciences	computer_science	health	education
0	1970	13.8	13.6	77.1	74.535328
1	1971	14.9	13.6	75.5	74.149204
2	1972	14.8	14.9	76.9	73.554520
3	1973	16.5	16.4	77.4	73.501814
4	1974	18.2	18.9	77.9	73.336811

```
In [5]: 1 year = df['year']
        2 physical_sciences = df['physical_sciences']
        3 computer_science = df['computer_science']
        4 health = df['health']
        5 education = df['education']
```

In [6]:

```
1 # Câu 1: Vẽ biểu đồ line plot thể hiện tỷ lệ % bằng Khoa học vật Lý và
2 # Khoa học máy tính được trao cho phụ nữ qua các năm
3 # Đồ thị có 2 line:
4 # Line 1: year, physical_sciences, color='blue'; và
5 # Line 2: year, computer_science, color='red'
6 # Bạn nhận xét gì về biểu đồ vừa vẽ
7
```

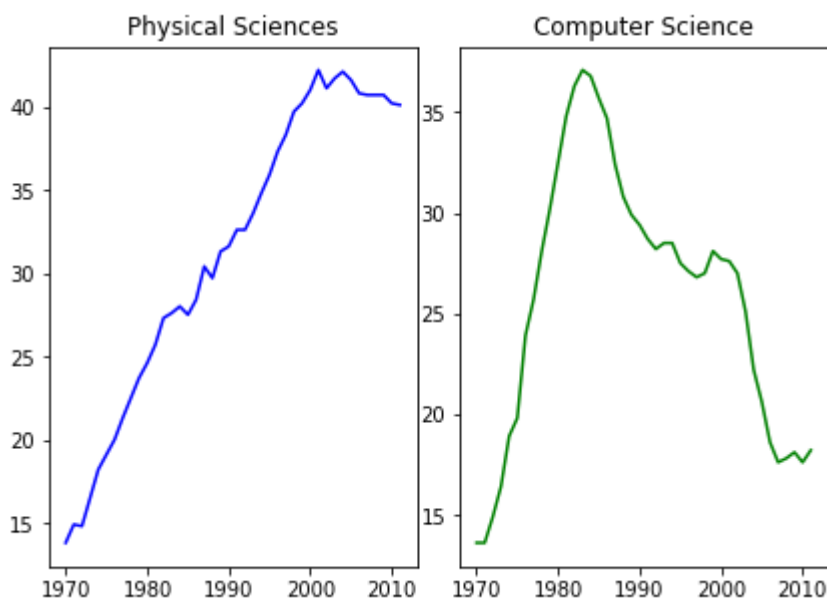
Nhấn vào đây để xem kết quả!



In [7]:

```
1 # Câu 2: Vẽ 2 biểu đồ line plot ở câu 1 nhưng trên 2 vùng:
2 # vùng 1: plt.axes([0.05, 0.05, 0.425, 0.9]),
3 # và vùng 2: plt.axes([0.525, 0.05, 0.425, 0.9])
4
```

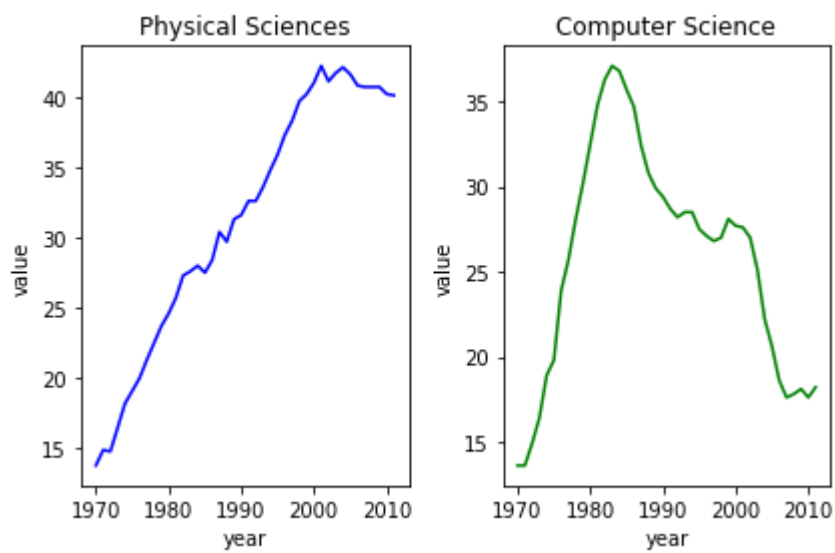
Nhấn vào đây để xem kết quả!



In [8]:

```
1 # Câu 3: Vẽ 2 biểu đồ line plot ở câu 1 nhưng trên 2 subplot:
2 # subplot 1: plt.subplot(1, 2, 1), và subplot 2: plt.subplot(1, 2, 2),
3 # Lưu ý: sử dụng plt.tight_layout() trước khi show()
4
```

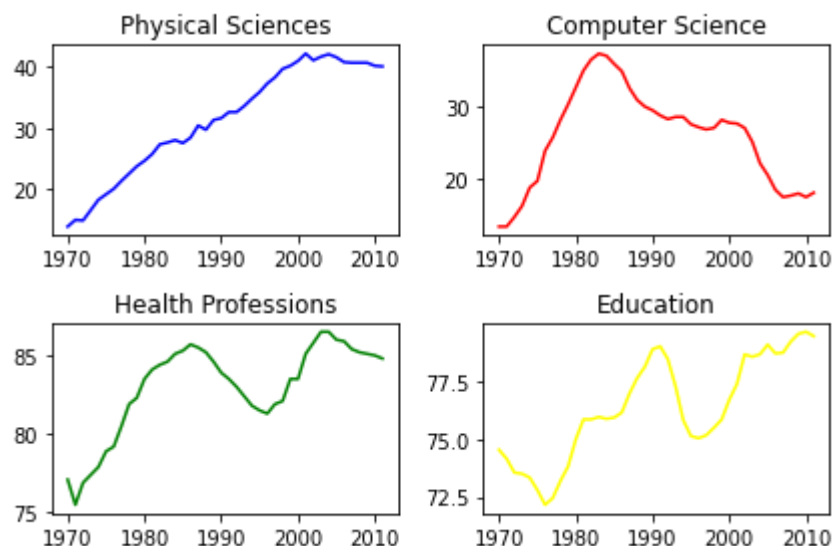
Nhấn vào đây để xem kết quả!



In [9]:

```
1 # Câu 4: Vẽ 4 biểu đồ line với 4 màu khác nhau:
2 # Biểu đồ 1: thể hiện tỷ lệ % bằng Khoa học vật lý
3 # được trao cho phụ nữ qua các năm (year - physical_sciences)
4 # Biểu đồ 2: thể hiện tỷ lệ % bằng Khoa học máy tính
5 # được trao cho phụ nữ qua các năm (year - computer_science)
6 # Biểu đồ 3: thể hiện tỷ lệ % phụ nữ tham gia
7 # các công việc liên quan đến y tế qua các năm (year - health)
8 # Biểu đồ 4: thể hiện tỷ lệ % phụ nữ tham gia
9 # các công việc liên quan đến giáo dục qua các năm (year - education)
10
11 # trên 4 subplot: plt.subplot(2, 2, 1), plt.subplot(2, 2, 2),
12 # plt.subplot(2, 2, 3), plt.subplot(2, 2, 4)
13 # Bạn nhận xét gì về biểu đồ vừa vẽ
14
```

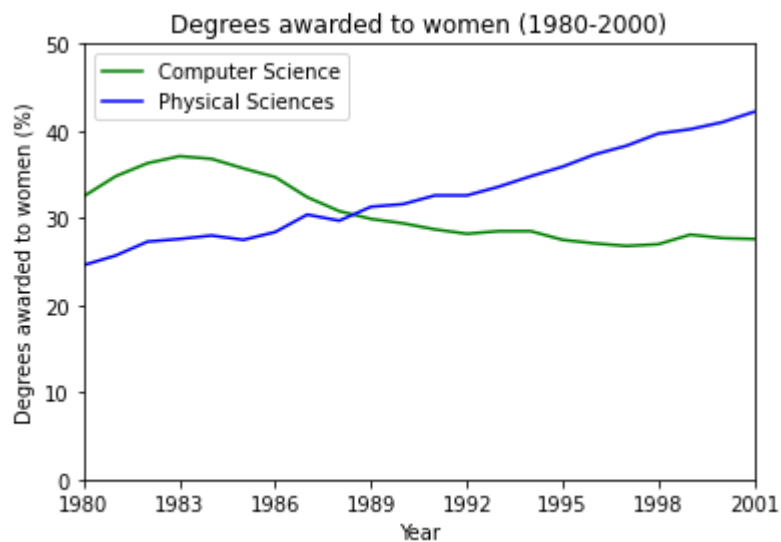
Nhấn vào đây để xem kết quả!



In [10]:

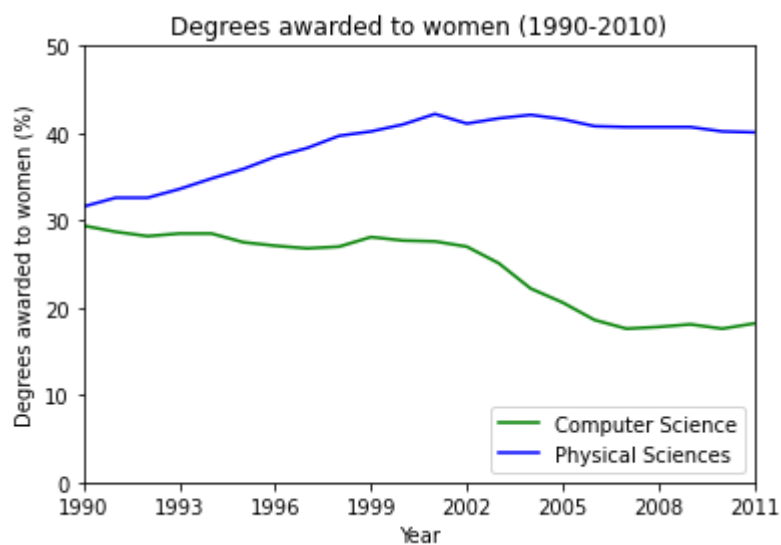
```
1 # Câu 5: Vẽ 2 biểu đồ line plot:
2 # Biểu đồ 1: thể hiện tỷ lệ % bằng Khoa học máy tính
3 # được trao cho phụ nữ qua các năm (year - computer_science)
4 # Biểu đồ 2: thể hiện tỷ lệ % bằng Khoa học vật lý
5 # được trao cho phụ nữ qua các năm (year - physical_sciences)
6 # nhưng giới hạn từ năm 1980-2000 --> có plt.xlim(1980, 2000) và plt.ylim(0, 50)
7
```

Nhấn vào đây để xem kết quả!



```
In [11]: 1 # Câu 6: Vẽ 2 biểu đồ line plot:
2 # Biểu đồ 1: thể hiện tỷ lệ % bằng Khoa học máy tính
3 # được trao cho phụ nữ qua các năm (year - computer_science)
4 # Biểu đồ 2: thể hiện tỷ lệ % bằng Khoa học vật Lý
5 # được trao cho phụ nữ qua các năm (year - physical_sciences)
6 # nhưng giới hạn từ năm 1990-2010 --> có plt.xlim(1990, 2010) và plt.ylim(0, 50)
7 # Bạn nhận xét gì về biểu đồ vừa vẽ
8
```

Nhấn vào đây để xem kết quả!



```
In [ ]: 1
```