plotting-data

October 23, 2019

0.1 Import modules

[1]:

```
import pandas as pd
import matplotlib.pyplot as plt
from pandas.plotting import register_matplotlib_converters
register_matplotlib_converters()
import numpy as np
```

[2]:

```
pypl_df = pd.read_csv('../data/pypl.csv')
```

[3]:

```
type(pypl_df)
```

[3]:

```
pandas.core.frame.DataFrame
```

[4]:

```
pypl_df.dtypes[:10]
```

[4]:

```
Year                int64
Month               int64
Day                 int64
Java              float64
Python            float64
PHP               float64
C#                float64
Javascript        float64
C/C++             float64
Objective-C       float64
dtype: object
```

```
[5]: pypl_df.dtypes[10:]
```

```
[5]: R              float64
     Swift          float64
     Matlab         float64
     Ruby           float64
     VBA            float64
     VisualBasic    float64
     Scala          float64
     Perl           float64
     Lua            float64
     Delphi         float64
     Go             float64
     Haskell        float64
     Rust           float64
     TypeScript     float64
     Kotlin         float64
     Julia          float64
     dtype: object
```

```
[6]: pypl_df.head(10)
```

```
[6]:    Year  Month  Day    Java  Python     PHP      C#  Javascript   C/C++  \
     0  2004      6    1  0.2956  0.0263  0.1897  0.0491      0.0900  0.1008
     1  2004      7    1  0.2930  0.0278  0.1910  0.0526      0.0902  0.0976
     2  2004      8    1  0.2919  0.0289  0.1922  0.0537      0.0899  0.0961
     3  2004      9    1  0.2942  0.0304  0.1920  0.0551      0.0868  0.0931
     4  2004     10    1  0.2948  0.0298  0.1930  0.0548      0.0869  0.0951
     5  2004     11    1  0.2992  0.0283  0.1936  0.0544      0.0864  0.0946
     6  2005      0    1  0.2988  0.0303  0.1961  0.0544      0.0854  0.0958
     7  2005      1    1  0.3036  0.0298  0.1968  0.0538      0.0830  0.0965
     8  2005      2    1  0.3058  0.0287  0.1969  0.0550      0.0807  0.0979
     9  2005      3    1  0.3071  0.0287  0.1986  0.0559      0.0818  0.0999

        Objective-C  ...   Scala    Perl     Lua  Delphi   Go  Haskell    Rust  \
     0       0.0016  ...  0.0001  0.0766  0.0024  0.0292  0.0   0.0023  0.0012
     1       0.0015  ...  0.0001  0.0752  0.0018  0.0282  0.0   0.0023  0.0011
     2       0.0017  ...  0.0001  0.0748  0.0018  0.0281  0.0   0.0020  0.0011
     3       0.0021  ...  0.0001  0.0746  0.0019  0.0288  0.0   0.0022  0.0013
     4       0.0021  ...  0.0002  0.0739  0.0018  0.0290  0.0   0.0025  0.0013
     5       0.0023  ...  0.0003  0.0735  0.0019  0.0289  0.0   0.0026  0.0014
     6       0.0019  ...  0.0003  0.0729  0.0017  0.0277  0.0   0.0024  0.0014
     7       0.0018  ...  0.0003  0.0717  0.0021  0.0276  0.0   0.0025  0.0016
     8       0.0016  ...  0.0005  0.0703  0.0022  0.0272  0.0   0.0028  0.0015
     9       0.0009  ...  0.0005  0.0694  0.0022  0.0264  0.0   0.0025  0.0010
```

```
   TypeScript  Kotlin  Julia
0      0.0000     0.0    0.0
1      0.0000     0.0    0.0
2      0.0000     0.0    0.0
3      0.0000     0.0    0.0
4      0.0000     0.0    0.0
5      0.0000     0.0    0.0
6      0.0000     0.0    0.0
7      0.0000     0.0    0.0
8      0.0001     0.0    0.0
9      0.0001     0.0    0.0

[10 rows x 26 columns]
```

```
[7]: pypl_df.tail(10)
```

```
[7]:      Year  Month  Day    Java  Python     PHP      C#  Javascript   C/C++  \
     174  2019      0    1  0.2142  0.2595  0.0737  0.0762      0.0826  0.0631
     175  2019      1    1  0.2115  0.2627  0.0732  0.0756      0.0823  0.0626
     176  2019      2    1  0.2084  0.2667  0.0726  0.0757      0.0837  0.0617
     177  2019      3    1  0.2059  0.2707  0.0718  0.0756      0.0841  0.0606
     178  2019      4    1  0.2047  0.2736  0.0718  0.0750      0.0842  0.0592
     179  2019      5    1  0.2030  0.2786  0.0699  0.0744      0.0846  0.0582
     180  2019      6    1  0.1998  0.2832  0.0676  0.0742      0.0849  0.0580
     181  2019      7    1  0.1984  0.2880  0.0663  0.0743      0.0852  0.0574
     182  2019      8    1  0.1974  0.2918  0.0653  0.0738      0.0839  0.0576
     183  2019      9    1  0.1957  0.2949  0.0634  0.0735      0.0840  0.0587

          Objective-C  ...   Scala    Perl     Lua  Delphi      Go  Haskell  \
     174       0.0315  ...  0.0119  0.0058  0.0039  0.0025  0.0104   0.0031
     175       0.0310  ...  0.0118  0.0059  0.0037  0.0026  0.0110   0.0031
     176       0.0302  ...  0.0118  0.0059  0.0035  0.0026  0.0111   0.0031
     177       0.0294  ...  0.0119  0.0060  0.0036  0.0028  0.0113   0.0032
     178       0.0293  ...  0.0117  0.0059  0.0036  0.0028  0.0118   0.0032
     179       0.0281  ...  0.0119  0.0059  0.0036  0.0028  0.0119   0.0031
     180       0.0276  ...  0.0119  0.0059  0.0037  0.0029  0.0120   0.0029
     181       0.0267  ...  0.0120  0.0057  0.0038  0.0027  0.0121   0.0029
     182       0.0263  ...  0.0116  0.0058  0.0037  0.0027  0.0122   0.0030
     183       0.0260  ...  0.0115  0.0057  0.0037  0.0025  0.0125   0.0029
```

```
       Rust  TypeScript  Kotlin   Julia
174  0.0041      0.0157  0.0114  0.0031
175  0.0043      0.0158  0.0115  0.0032
176  0.0047      0.0159  0.0120  0.0030
177  0.0050      0.0162  0.0123  0.0029
178  0.0055      0.0165  0.0127  0.0027
179  0.0056      0.0170  0.0135  0.0029
180  0.0059      0.0177  0.0145  0.0027
181  0.0064      0.0183  0.0147  0.0025
182  0.0065      0.0184  0.0154  0.0027
183  0.0064      0.0187  0.0161  0.0028

[10 rows x 26 columns]
```

[8]:

```
pypl_df[:10]
```

[8]:

```
   Year  Month  Day    Java  Python     PHP      C#  Javascript   C/C++  \
0  2004      6    1  0.2956  0.0263  0.1897  0.0491      0.0900  0.1008
1  2004      7    1  0.2930  0.0278  0.1910  0.0526      0.0902  0.0976
2  2004      8    1  0.2919  0.0289  0.1922  0.0537      0.0899  0.0961
3  2004      9    1  0.2942  0.0304  0.1920  0.0551      0.0868  0.0931
4  2004     10    1  0.2948  0.0298  0.1930  0.0548      0.0869  0.0951
5  2004     11    1  0.2992  0.0283  0.1936  0.0544      0.0864  0.0946
6  2005      0    1  0.2988  0.0303  0.1961  0.0544      0.0854  0.0958
7  2005      1    1  0.3036  0.0298  0.1968  0.0538      0.0830  0.0965
8  2005      2    1  0.3058  0.0287  0.1969  0.0550      0.0807  0.0979
9  2005      3    1  0.3071  0.0287  0.1986  0.0559      0.0818  0.0999

   Objective-C  ...   Scala    Perl     Lua  Delphi   Go  Haskell    Rust  \
0       0.0016  ...  0.0001  0.0766  0.0024  0.0292  0.0   0.0023  0.0012
1       0.0015  ...  0.0001  0.0752  0.0018  0.0282  0.0   0.0023  0.0011
2       0.0017  ...  0.0001  0.0748  0.0018  0.0281  0.0   0.0020  0.0011
3       0.0021  ...  0.0001  0.0746  0.0019  0.0288  0.0   0.0022  0.0013
4       0.0021  ...  0.0002  0.0739  0.0018  0.0290  0.0   0.0025  0.0013
5       0.0023  ...  0.0003  0.0735  0.0019  0.0289  0.0   0.0026  0.0014
6       0.0019  ...  0.0003  0.0729  0.0017  0.0277  0.0   0.0024  0.0014
7       0.0018  ...  0.0003  0.0717  0.0021  0.0276  0.0   0.0025  0.0016
8       0.0016  ...  0.0005  0.0703  0.0022  0.0272  0.0   0.0028  0.0015
9       0.0009  ...  0.0005  0.0694  0.0022  0.0264  0.0   0.0025  0.0010
```

```
   TypeScript  Kotlin  Julia
0      0.0000     0.0    0.0
1      0.0000     0.0    0.0
2      0.0000     0.0    0.0
3      0.0000     0.0    0.0
4      0.0000     0.0    0.0
5      0.0000     0.0    0.0
6      0.0000     0.0    0.0
7      0.0000     0.0    0.0
8      0.0001     0.0    0.0
9      0.0001     0.0    0.0

[10 rows x 26 columns]
```

[9]:
```
pypl_df[-10:]
```

[9]:
```
     Year  Month  Day    Java  Python     PHP      C#  Javascript   C/C++  \
174  2019      0    1  0.2142  0.2595  0.0737  0.0762      0.0826  0.0631
175  2019      1    1  0.2115  0.2627  0.0732  0.0756      0.0823  0.0626
176  2019      2    1  0.2084  0.2667  0.0726  0.0757      0.0837  0.0617
177  2019      3    1  0.2059  0.2707  0.0718  0.0756      0.0841  0.0606
178  2019      4    1  0.2047  0.2736  0.0718  0.0750      0.0842  0.0592
179  2019      5    1  0.2030  0.2786  0.0699  0.0744      0.0846  0.0582
180  2019      6    1  0.1998  0.2832  0.0676  0.0742      0.0849  0.0580
181  2019      7    1  0.1984  0.2880  0.0663  0.0743      0.0852  0.0574
182  2019      8    1  0.1974  0.2918  0.0653  0.0738      0.0839  0.0576
183  2019      9    1  0.1957  0.2949  0.0634  0.0735      0.0840  0.0587

     Objective-C  ...   Scala    Perl     Lua  Delphi      Go  Haskell  \
174       0.0315  ...  0.0119  0.0058  0.0039  0.0025  0.0104   0.0031
175       0.0310  ...  0.0118  0.0059  0.0037  0.0026  0.0110   0.0031
176       0.0302  ...  0.0118  0.0059  0.0035  0.0026  0.0111   0.0031
177       0.0294  ...  0.0119  0.0060  0.0036  0.0028  0.0113   0.0032
178       0.0293  ...  0.0117  0.0059  0.0036  0.0028  0.0118   0.0032
179       0.0281  ...  0.0119  0.0059  0.0036  0.0028  0.0119   0.0031
180       0.0276  ...  0.0119  0.0059  0.0037  0.0029  0.0120   0.0029
181       0.0267  ...  0.0120  0.0057  0.0038  0.0027  0.0121   0.0029
182       0.0263  ...  0.0116  0.0058  0.0037  0.0027  0.0122   0.0030
183       0.0260  ...  0.0115  0.0057  0.0037  0.0025  0.0125   0.0029
```

```
       Rust  TypeScript  Kotlin   Julia
174  0.0041      0.0157  0.0114  0.0031
175  0.0043      0.0158  0.0115  0.0032
176  0.0047      0.0159  0.0120  0.0030
177  0.0050      0.0162  0.0123  0.0029
178  0.0055      0.0165  0.0127  0.0027
179  0.0056      0.0170  0.0135  0.0029
180  0.0059      0.0177  0.0145  0.0027
181  0.0064      0.0183  0.0147  0.0025
182  0.0065      0.0184  0.0154  0.0027
183  0.0064      0.0187  0.0161  0.0028

[10 rows x 26 columns]
```

[10]:
```
pypl_df[20:30]
```

[10]:
```
    Year  Month  Day    Java  Python     PHP      C#  Javascript   C/C++  \
20  2006      2    1  0.3073  0.0392  0.1968  0.0646      0.0818  0.0903
21  2006      3    1  0.3102  0.0382  0.1976  0.0668      0.0818  0.0894
22  2006      4    1  0.3099  0.0379  0.2000  0.0679      0.0817  0.0869
23  2006      5    1  0.3074  0.0388  0.2008  0.0685      0.0834  0.0836
24  2006      6    1  0.3021  0.0379  0.2036  0.0702      0.0841  0.0839
25  2006      7    1  0.3012  0.0382  0.2053  0.0711      0.0851  0.0828
26  2006      8    1  0.3022  0.0392  0.2057  0.0715      0.0859  0.0807
27  2006      9    1  0.2984  0.0406  0.2033  0.0721      0.0874  0.0806
28  2006     10    1  0.2979  0.0405  0.2045  0.0718      0.0880  0.0829
29  2006     11    1  0.2994  0.0406  0.2025  0.0720      0.0863  0.0855

    Objective-C  ...   Scala    Perl     Lua  Delphi   Go  Haskell    Rust  \
20       0.0008  ...  0.0001  0.0589  0.0027  0.0239  0.0   0.0024  0.0009
21       0.0009  ...  0.0001  0.0579  0.0028  0.0219  0.0   0.0028  0.0011
22       0.0010  ...  0.0001  0.0566  0.0031  0.0211  0.0   0.0029  0.0012
23       0.0012  ...  0.0000  0.0567  0.0034  0.0203  0.0   0.0029  0.0010
24       0.0011  ...  0.0000  0.0571  0.0035  0.0201  0.0   0.0028  0.0011
25       0.0011  ...  0.0000  0.0569  0.0036  0.0195  0.0   0.0029  0.0009
26       0.0010  ...  0.0002  0.0568  0.0036  0.0200  0.0   0.0028  0.0009
27       0.0010  ...  0.0003  0.0566  0.0035  0.0208  0.0   0.0026  0.0007
28       0.0010  ...  0.0003  0.0552  0.0036  0.0203  0.0   0.0028  0.0006
29       0.0008  ...  0.0004  0.0549  0.0037  0.0198  0.0   0.0029  0.0007
```

```
    TypeScript  Kotlin  Julia
20         0.0     0.0    0.0
21         0.0     0.0    0.0
22         0.0     0.0    0.0
23         0.0     0.0    0.0
24         0.0     0.0    0.0
25         0.0     0.0    0.0
26         0.0     0.0    0.0
27         0.0     0.0    0.0
28         0.0     0.0    0.0
29         0.0     0.0    0.0

[10 rows x 26 columns]
```

[11]:

```
pypl_df['Month'][:10]
```

[11]:

```
0     6
1     7
2     8
3     9
4    10
5    11
6     0
7     1
8     2
9     3
Name: Month, dtype: int64
```

[12]:

```
pypl_df.Month[:10]
```

[12]:

```
0     6
1     7
2     8
3     9
4    10
5    11
6     0
7     1
8     2
9     3
Name: Month, dtype: int64
```

```
[13]: x = 1
      x += 1
      print(x)
```

```
2
```

```
[14]: pypl_df.Month[:5]
```

```
[14]: 0     6
      1     7
      2     8
      3     9
      4    10
      Name: Month, dtype: int64
```

```
[15]: pypl_df.Month += 1
```

```
[16]: pypl_df.Month[:5]
```

```
[16]: 0     7
      1     8
      2     9
      3    10
      4    11
      Name: Month, dtype: int64
```

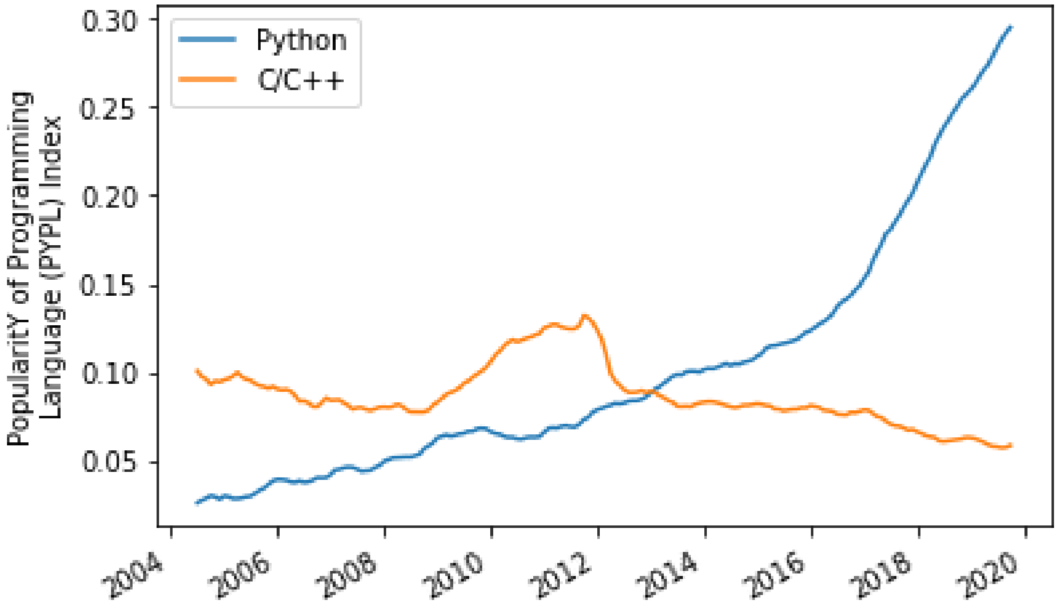
```
[17]: pypl_df['Date'] = pd.to_datetime(dict(year=pypl_df.Year, month=pypl_df.Month, day=pypl_df.
        ↪Day))
```

```
[18]: pypl_df[['Date', 'Year', 'Month', 'Day']].head(10)
```

```
[18]:         Date  Year  Month  Day
      0 2004-07-01  2004      7    1
      1 2004-08-01  2004      8    1
      2 2004-09-01  2004      9    1
      3 2004-10-01  2004     10    1
      4 2004-11-01  2004     11    1
      5 2004-12-01  2004     12    1
      6 2005-01-01  2005      1    1
      7 2005-02-01  2005      2    1
      8 2005-03-01  2005      3    1
      9 2005-04-01  2005      4    1
```

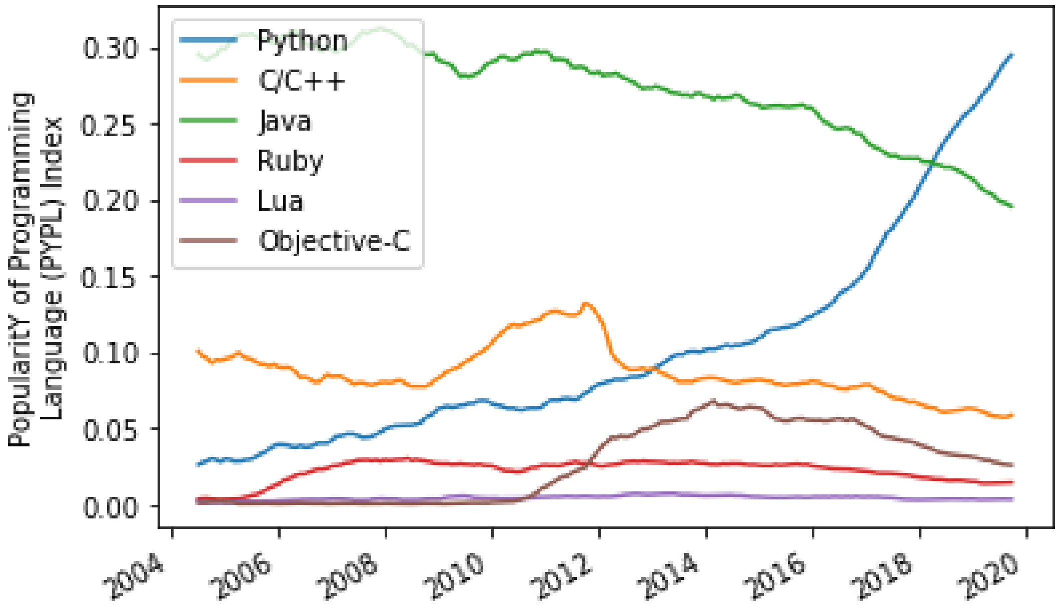
[19]:

```
plt.plot(pypl_df.Date, pypl_df['Python'], label='Python')
plt.plot(pypl_df.Date, pypl_df['C/C++'], label='C/C++')
# beautify the x-labels
plt.gcf().autofmt_xdate()
plt.ylabel('PopularitY of Programming\nLanguage (PYPL) Index')
plt.legend()
plt.show()
```



```
[20]: def plot_pypl(pypl_df, names):
          for name in names:
              plt.plot(pypl_df.Date, pypl_df[name], label=name)
          # beautify the x-labels
          plt.gcf().autofmt_xdate()
          plt.ylabel('PopularitY of Programming\nLanguage (PYPL) Index')
          plt.legend()
          plt.show()
```

```
[21]: plot_pypl(pypl_df, ['Python', 'C/C++', 'Java', 'Ruby', 'Lua', 'Objective-C'])
```



```
[22]: names=np.array(pypl_df.columns[3:-1])
      names
```

```
[22]: array(['Java', 'Python', 'PHP', 'C#', 'Javascript', 'C/C++',
             'Objective-C', 'R', 'Swift', 'Matlab', 'Ruby', 'VBA',
             'VisualBasic', 'Scala', 'Perl', 'Lua', 'Delphi', 'Go', 'Haskell',
             'Rust', 'TypeScript', 'Kotlin', 'Julia'], dtype=object)
```

```
[23]: popularity=np.array(pypl_df[names][-1:].values[0])
      popularity
```

```
[23]: array([0.1957, 0.2949, 0.0634, 0.0735, 0.084 , 0.0587, 0.026 , 0.0382,
             0.0257, 0.0187, 0.0147, 0.0139, 0.0099, 0.0115, 0.0057, 0.0037,
             0.0025, 0.0125, 0.0029, 0.0064, 0.0187, 0.0161, 0.0028])
```

```
[24]: date=pd.to_datetime(pypl_df.Date[-1:].values[0])
      date
```

```
[24]: Timestamp('2019-10-01 00:00:00')
```

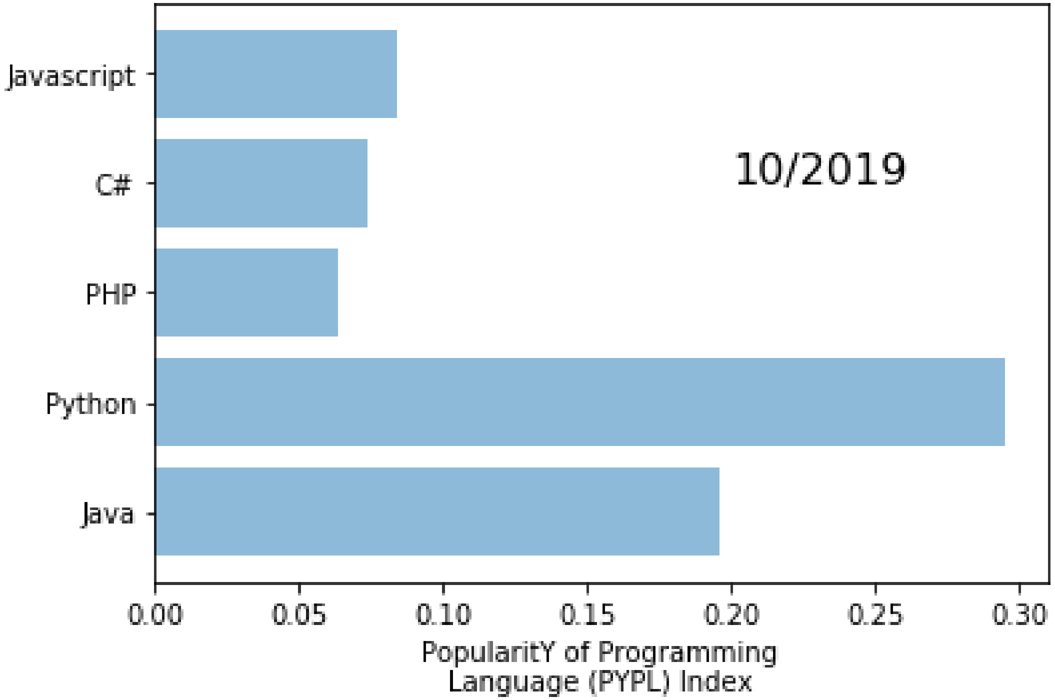
```
[25]: date.strftime("%m/%d/%Y")
```

```
[25]: '10/01/2019'
```

[26]:
```
y_pos = np.arange(len(names))
y_pos
```

[26]:
```
array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16,
       17, 18, 19, 20, 21, 22])
```

[27]:
```
plt.barh(y_pos[0:5], popularity[0:5], align='center', alpha=0.5)
plt.yticks(y_pos[0:5], names)
plt.xlabel('PopularitY of Programming\nLanguage (PYPL) Index')
plt.text(.2, 3, date.strftime("%m/%Y"), fontsize=16)
plt.show()
```


Javascript
C#
PHP
Python
Java
10/2019
0.00
0.05
0.10
0.15
0.20
0.25
0.30
PopularitY of Programming
Language (PYPL) Index

[28]:
```
pypl_df.iloc[[10]]
```

[28]:
```
    Year  Month  Day    Java  Python     PHP      C#  Javascript   C/C++  \
10  2005      5    1  0.3085  0.0291  0.1987  0.0569       0.083  0.0975

    Objective-C  ...    Perl     Lua  Delphi   Go  Haskell    Rust  \
10       0.0009  ...  0.0683  0.0023  0.0252  0.0   0.0023  0.0007

    TypeScript  Kotlin  Julia       Date
10      0.0001     0.0    0.0 2005-05-01

[1 rows x 27 columns]
```

[29]:
```
date = pd.to_datetime(pypl_df.iloc[[10]].Date.values[0])
date
```

[29]:
```
Timestamp('2005-05-01 00:00:00')
```

[30]:
```
popularity = pypl_df[names].iloc[[10]].values[0]
popularity
```

[30]:
```
array([3.085e-01, 2.910e-02, 1.987e-01, 5.690e-02, 8.300e-02, 9.750e-02,
       9.000e-04, 4.300e-03, 0.000e+00, 2.450e-02, 4.600e-03, 1.540e-02,
       7.740e-02, 3.000e-04, 6.830e-02, 2.300e-03, 2.520e-02, 0.000e+00,
       2.300e-03, 7.000e-04, 1.000e-04, 0.000e+00, 0.000e+00])
```

```
[31]: df = pd.DataFrame({'names': names,
                         'popularity': popularity})
      df.nlargest(5, 'popularity')
```

```
[31]:          names  popularity
      0          Java      0.3085
      2           PHP      0.1987
      5         C/C++      0.0975
      4    Javascript      0.0830
      12  VisualBasic      0.0774
```

[32]:
```
import random
```

[33]:
```
number_of_colors = len(names)

color = ["#"+''.join([random.choice('0123456789ABCDEF') for j in range(6)])
             for i in range(number_of_colors)]
```

[34]:
```
color
```

[34]:
```
['#38F133',
 '#9D38A5',
 '#EBED1C',
 '#8E63A8',
 '#7CC336',
 '#E63A2D',
 '#FCDFD2',
 '#502D89',
 '#ABC1D7',
 '#DEFF3F',
 '#C68577',
 '#268A3A',
 '#29BBCE',
 '#062AC7',
 '#97CB25',
 '#001A3A',
 '#4EBD76',
 '#C7A5EC',
 '#9BE32F',
 '#D47FA6',
 '#9CBF6F',
 '#76484D',
 '#8968DD']
```

[35]:

```
def get_top_five(pypl_df, i, color):
    date = pd.to_datetime(pypl_df.iloc[[i]].Date.values[0])
    popularity = pypl_df[names].iloc[[i]].values[0]
    df = pd.DataFrame({'names': names,
                       'popularity': popularity,
                       'color': color})
    return date, df.nlargest(5, 'popularity').sort_values(by=['popularity'], ascending=1)
```

[36]:

```
date, df = get_top_five(pypl_df,10, color)
print(date.strftime("%m/%Y"))
print(df)
```

```
05/2005
          names  popularity    color
12  VisualBasic      0.0774  #29BBCE
4    Javascript      0.0830  #7CC336
5         C/C++      0.0975  #E63A2D
2           PHP      0.1987  #EBED1C
0          Java      0.3085  #38F133
```

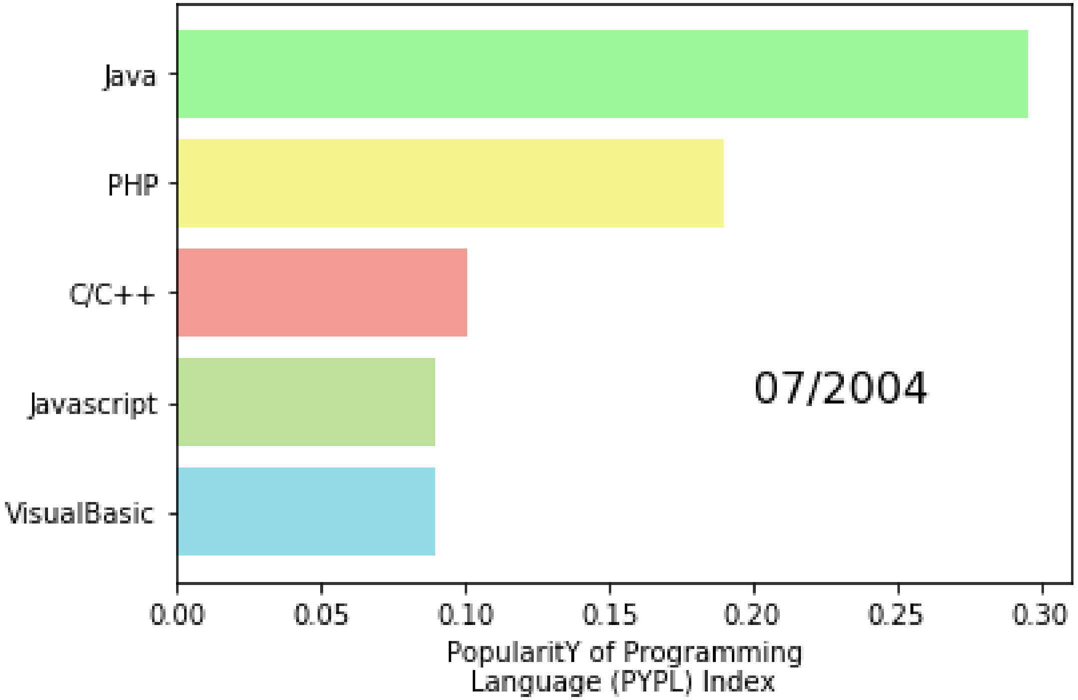
[37]:

```
date, df = get_top_five(pypl_df,50, color)
print(date.strftime("%m/%Y"))
print(df)
```

```
09/2008
        names  popularity    color
5       C/C++      0.0776  #E63A2D
```

```
3          C#      0.0812  #8E63A8
4  Javascript      0.0833  #7CC336
2         PHP      0.2035  #EBED1C
0        Java      0.2970  #38F133
```

```
[38]: date, df = get_top_five(pypl_df,0, color)
      y_pos = np.arange(len(df))
      plt.barh(y_pos, df['popularity'], align='center', alpha=0.5, color=df['color'])
      plt.yticks(y_pos, df['names'])
      plt.xlabel('PopularitY of Programming\nLanguage (PYPL) Index')
      plt.text(.2, 1, date.strftime("%m/%Y"), fontsize=16)
      plt.show()
```



```
[39]: str(1).zfill(3)
```

```
[39]: '001'
```

```
[40]: str(11).zfill(3)
```

```
[40]: '011'
```

```
[41]: str(111).zfill(3)
```

```
[41]: '111'
```

```
[42]: my_dpi=96
      for i in range(len(pypl_df)):
          fig = plt.figure(figsize=(680/my_dpi, 480/my_dpi), dpi=my_dpi)
          date, df = get_top_five(pypl_df, i, color)
          y_pos = np.arange(len(df))
          plt.xlabel('PopularitY of Programming\nLanguage (PYPL) Index')
          plt.barh(y_pos, df['popularity'], align='center', alpha=0.5, color=df['color'])
          plt.yticks(y_pos, df['names'])
          plt.xlim([0,.35])
          plt.text(.2, 1, date.strftime("%m/%Y"), fontsize=16)
          filename='popularity_step_'+str(i).zfill(3)+'.png'
          plt.savefig(filename, dpi=96)
          plt.gca()
          plt.close(fig)
```

Now run the following command (in bash) to generate an animated gif from the 184 png files:

```
convert -delay 20 popularity_step_*.png language_popularity_animated.gif
```