birth-rate-analysis

November 9, 2023

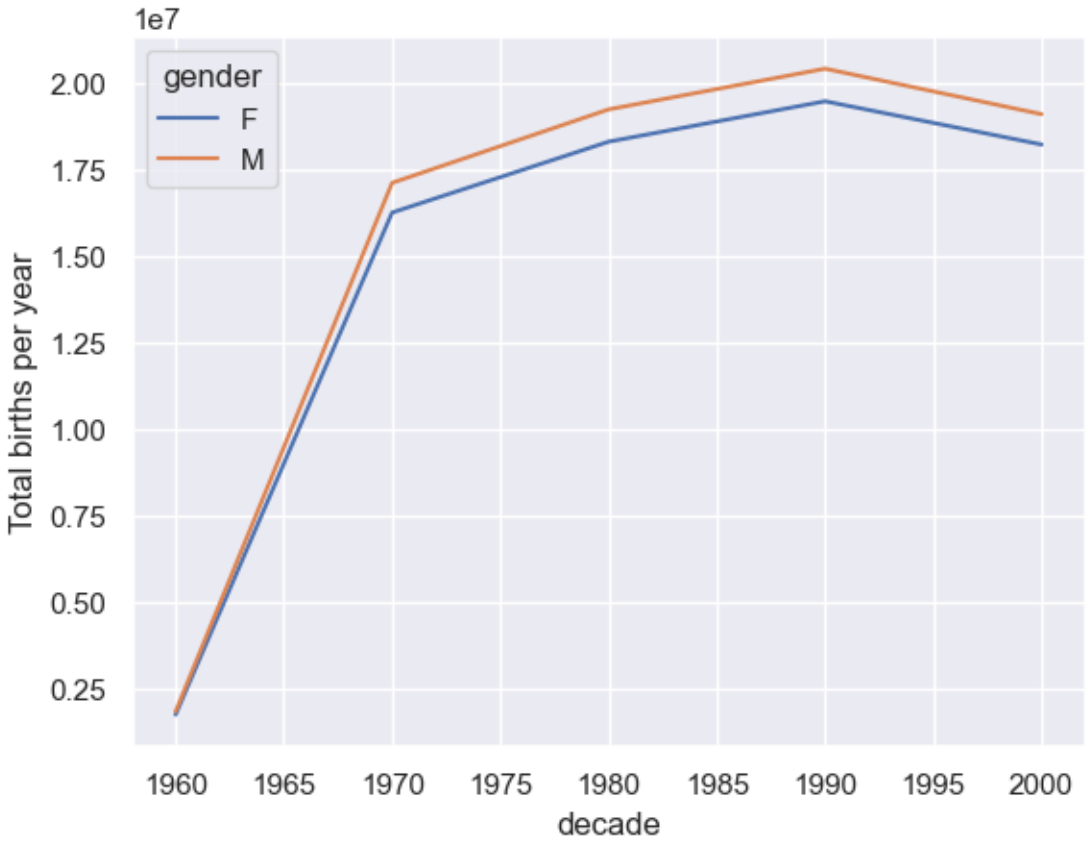
[1]:
```
import pandas as pd
births = pd.read_csv("C:\\Users\\kanch\\OneDrive\\Documents\\Mini␣
 ↪Project\\births.csv")
print(births.head())
births['day'].fillna(0, inplace=True)
births['day'] = births['day'].astype(int)
```

```
   year  month  day gender  births
0  1969      1  1.0      F    4046
1  1969      1  1.0      M    4440
2  1969      1  2.0      F    4454
3  1969      1  2.0      M    4548
4  1969      1  3.0      F    4548
```

[2]:
```
births['decade'] = 10 * (births['year'] // 10)
births.pivot_table('births', index='decade', columns='gender', aggfunc='sum')
print(births.head())
```

```
   year  month  day gender  births  decade
0  1969      1    1      F    4046    1960
1  1969      1    1      M    4440    1960
2  1969      1    2      F    4454    1960
3  1969      1    2      M    4548    1960
4  1969      1    3      F    4548    1960
```

[3]:
```
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()
birth_decade = births.pivot_table('births', index='decade', columns='gender',␣
 ↪aggfunc='sum')
birth_decade.plot()
plt.ylabel("Total births per year")
plt.show()
```



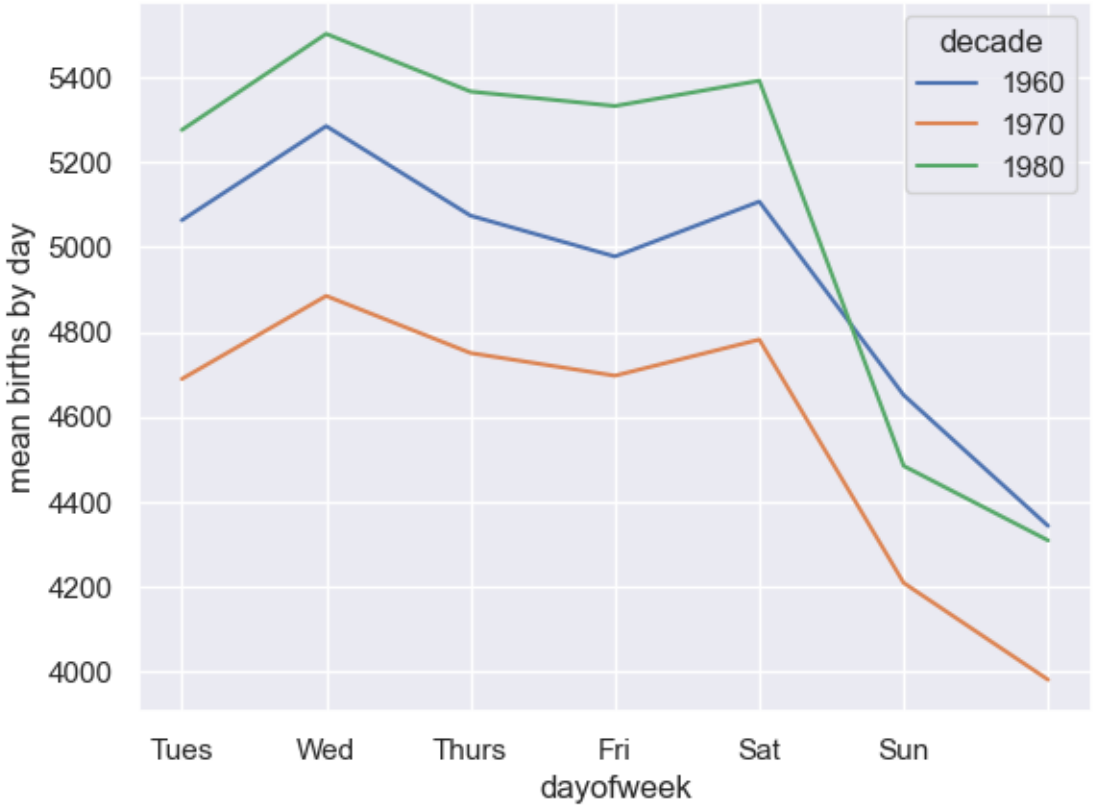
[4]:
```
import numpy as np
quartiles = np.percentile(births['births'], [25, 50, 75])
mu = quartiles[1]
sig = 0.74 * (quartiles[2] - quartiles[0])
```

[5]:
```
births = births.query('(births > @mu - 5 * @sig) & (births < @mu + 5 * @sig)')
births['day'] = births['day'].astype(int)
births.index = pd.to_datetime(10000 * births.year +
                              100 * births.month +
                              births.day, format='%Y%m%d')

births['dayofweek'] = births.index.dayofweek
```

[6]:
```
births.pivot_table('births', index='dayofweek',
                   columns='decade', aggfunc='mean').plot()
plt.gca().set_xticklabels(['Mon', 'Tues', 'Wed', 'Thurs', 'Fri', 'Sat', 'Sun'])
plt.ylabel('mean births by day');
plt.show()
```

```
C:\Users\kanch\AppData\Local\Temp\ipykernel_28932\3504355344.py:3: UserWarning:
```

```
FixedFormatter should only be used together with FixedLocator
  plt.gca().set_xticklabels(['Mon', 'Tues', 'Wed', 'Thurs', 'Fri', 'Sat',
'Sun'])
```



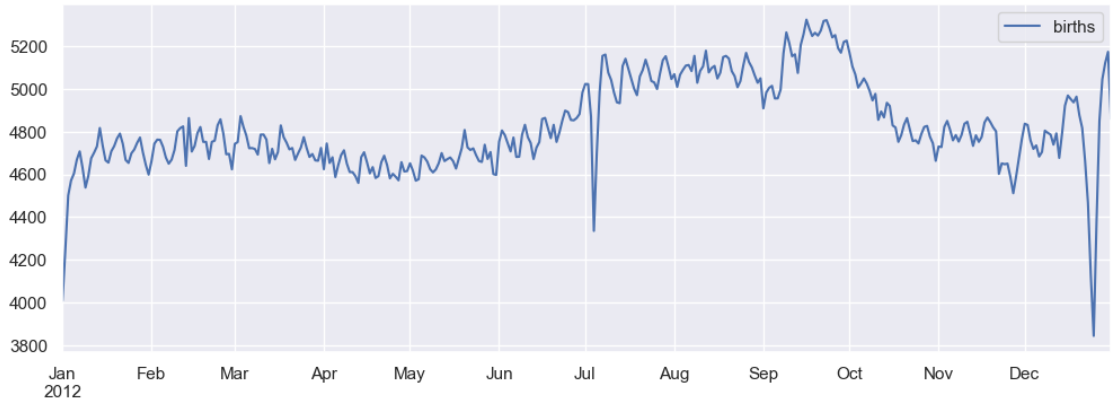
[7]:
```
births_month = births.pivot_table('births', [births.index.month, births.index.
  ↪day])
print(births_month.head())

births_month.index = [pd.datetime(2012, month, day)
                      for (month, day) in births_month.index]
print(births_month.head())
```

```
       births
1 1  4009.225
  2  4247.400
  3  4500.900
  4  4571.350
  5  4603.625
              births
2012-01-01  4009.225
2012-01-02  4247.400
```

```
2012-01-03  4500.900
2012-01-04  4571.350
2012-01-05  4603.625
```

```
C:\Users\kanch\AppData\Local\Temp\ipykernel_28932\3416507701.py:4:
FutureWarning: The pandas.datetime class is deprecated and will be removed from
pandas in a future version. Import from datetime module instead.
  births_month.index = [pd.datetime(2012, month, day)
```

[8]:
```
fig, ax = plt.subplots(figsize=(12, 4))
births_month.plot(ax=ax)
plt.show()
```



[ ]:

[ ]:

[ ]:

[ ]:

[ ]: