Exploring Monthly Birth Data

The number of live births (in thousands) per month in the U.S. was collected for the past 31 years (`data/birth.txt`) starting in January 1980 and ending December, 2010. We will be exploring this time series using various methods and predict the birth counts for 2011.

Download the data here (https://learn.galvanize.com/content/gSchool/dsi-curriculum/master/time-series/data).

1. Load the data into a pandas dataframe.

2. Using `pandas.date_range()` to create a `dates` variable (from January 1980 and ending December, 2010).

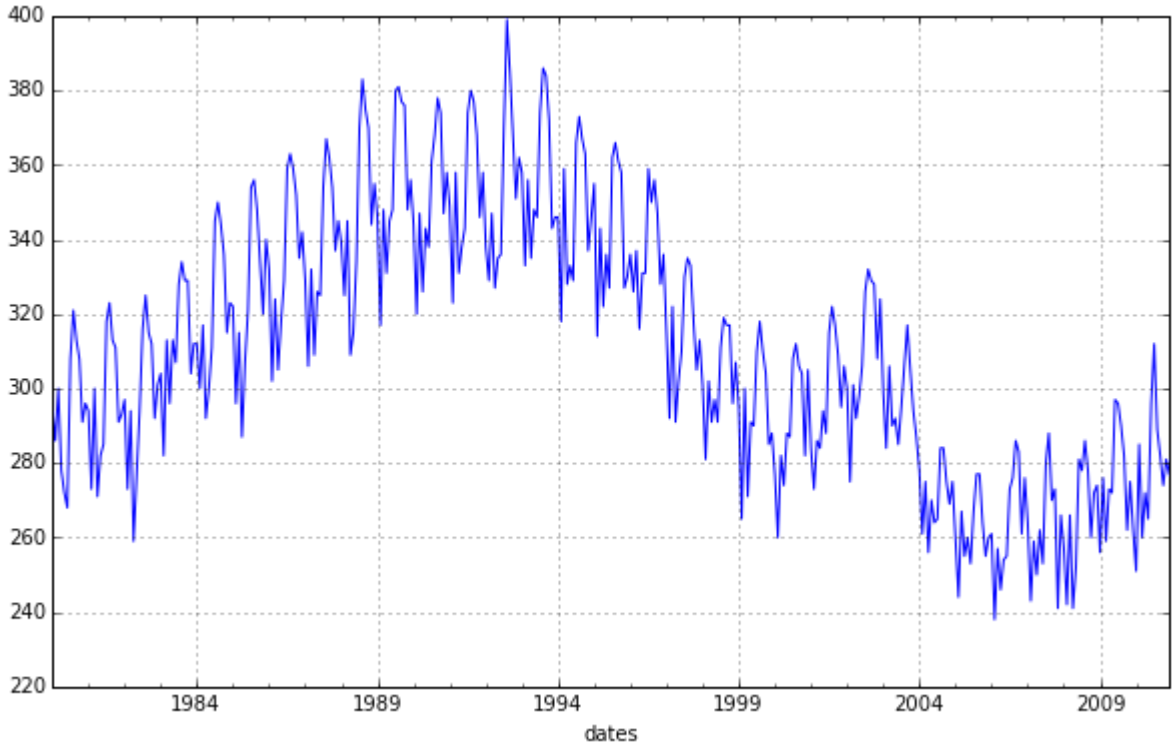
3. Create a `time` variable (range: 1-372) to be used later in the regressions and both a `month` and `year` variable (use `pd.DatetimeIndex` to strip these values from your dates).

4. Set the `dates` variable as the index of your dataframe.

5. Calculate some aggregated statistics by month and year. What months have the highest birthrates? Any intuition as to why?

6. Turn the `num_births` into a time series using `pd.Series()` .

7. Plot the overall data. What are your thoughts about the general pattern and or seasonal variation?



8. Plot the data for 2006-2010, is the seasonal pattern more apparent?

9. Use `df.resample('Q-NOV')` to get quarterly means that follow the seasons of the year (spring, summer, fall, winter).

10. Superimpose the yearly averages and the seasonal averages onto the monthly data.