Libraries

In [1]:
```
import numpy as np
import pandas as pd
import os
import us
os.getcwd()
```

Out[1]:
```
'C:\\Users\\nguye\\Documents\\UVA\\Term 3\\Bayesian Machine Learning\\Project\\Scripts'
```

Loading data

In [2]:
```
cdc_lyme = pd.read_csv('C:/Users/nguye/Documents/UVA/Term 3/Bayesian Machine Learning/Project/Data/LD-Case-Coun
```

Data Wrangling

In [3]:
```
# Transforming columns
cdc_lyme['CTYCODE'] = cdc_lyme['CTYCODE'].astype(str).str.zfill(3)
cdc_lyme['STCODE'] = cdc_lyme['STCODE'].astype(str).str.zfill(2)
cdc_lyme['FIPS'] = cdc_lyme['STCODE'] + cdc_lyme['CTYCODE']
cdc_lyme['stabbr'] = cdc_lyme['STCODE'].map(us.states.mapping('fips', 'abbr'))
state_abbr = cdc_lyme['stabbr'].unique()
```

In [4]:
```
# Making variables to feed into dataframe
state_count_2000 = [cdc_lyme[cdc_lyme['stabbr'] == st]['Cases2000'].sum() for st in state_abbr]
state_count_2001 = [cdc_lyme[cdc_lyme['stabbr'] == st]['Cases2001'].sum() for st in state_abbr]
state_count_2002 = [cdc_lyme[cdc_lyme['stabbr'] == st]['Cases2002'].sum() for st in state_abbr]
state_count_2003 = [cdc_lyme[cdc_lyme['stabbr'] == st]['Cases2003'].sum() for st in state_abbr]
state_count_2004 = [cdc_lyme[cdc_lyme['stabbr'] == st]['Cases2004'].sum() for st in state_abbr]
state_count_2005 = [cdc_lyme[cdc_lyme['stabbr'] == st]['Cases2005'].sum() for st in state_abbr]
state_count_2016 = [cdc_lyme[cdc_lyme['stabbr'] == st]['Cases2006'].sum() for st in state_abbr]
state_count_2007 = [cdc_lyme[cdc_lyme['stabbr'] == st]['Cases2007'].sum() for st in state_abbr]
state_count_2008 = [cdc_lyme[cdc_lyme['stabbr'] == st]['Cases2008'].sum() for st in state_abbr]
state_count_2009 = [cdc_lyme[cdc_lyme['stabbr'] == st]['Cases2009'].sum() for st in state_abbr]
state_count_2010 = [cdc_lyme[cdc_lyme['stabbr'] == st]['Cases2010'].sum() for st in state_abbr]
state_count_2011 = [cdc_lyme[cdc_lyme['stabbr'] == st]['Cases2011'].sum() for st in state_abbr]
state_count_2012 = [cdc_lyme[cdc_lyme['stabbr'] == st]['Cases2012'].sum() for st in state_abbr]
state_count_2013 = [cdc_lyme[cdc_lyme['stabbr'] == st]['Cases2013'].sum() for st in state_abbr]
state_count_2014 = [cdc_lyme[cdc_lyme['stabbr'] == st]['Cases2014'].sum() for st in state_abbr]
state_count_2015 = [cdc_lyme[cdc_lyme['stabbr'] == st]['Cases2015'].sum() for st in state_abbr]
state_count_2016 = [cdc_lyme[cdc_lyme['stabbr'] == st]['Cases2016'].sum() for st in state_abbr]
state_count_2017 = [cdc_lyme[cdc_lyme['stabbr'] == st]['Cases2017'].sum() for st in state_abbr]
state_count_2018 = [cdc_lyme[cdc_lyme['stabbr'] == st]['Cases2018'].sum() for st in state_abbr]
state_count_2019 = [cdc_lyme[cdc_lyme['stabbr'] == st]['Cases2019'].sum() for st in state_abbr]
```

In [5]:
```
# Making the data frame
count_summary = pd.DataFrame({'State': state_abbr,
                              '2000_cases': state_count_2000,
                              '2001_cases': state_count_2001,
                              '2002_cases': state_count_2002,
                              '2003_cases': state_count_2003,
                              '2004_cases': state_count_2004,
                              '2005_cases': state_count_2005,
                              '2006_cases': state_count_2016,
                              '2007_cases': state_count_2007,
                              '2008_cases': state_count_2008,
                              '2009_cases': state_count_2009,
                              '2010_cases': state_count_2010,
                              '2011_cases': state_count_2011,
                              '2012_cases': state_count_2012,
                              '2013_cases': state_count_2013,
                              '2014_cases': state_count_2014,
                              '2015_cases': state_count_2015,
                              '2016_cases': state_count_2016,
                              '2017_cases': state_count_2017,
                              '2018_cases': state_count_2018,
                              '2019_cases': state_count_2019})
```

Writing to csv

In [6]:
```
count_summary.to_csv('Lyme Summary.csv', header = True, index = False)
```