Московский государственный технический университет имени Н.Э.Баумана

Кафедра «Системы обработки информации и управления»

О Т Ч Е Т

Лабораторная работа №1

по курсу «Методы машинного обучения»

«Разведочный анализ данных. Исследование и визуализация данных.»

Исполнитель: **Соболева Е.Д.**

группа ИУ5-11М

Проверил: **Гапанюк Ю.Е.**

Москва, 2020

In [1]:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
sns.set(style="ticks")
```

In [2]:

```
data = pd.read_csv('african_crises.csv', sep=",")
```

In [3]:

```
data.head()
```

Out[3]:

	case	cc3	country	year	systemic_crisis	exch_usd	domestic_debt_in_default	sov
0	1	DZA	Algeria	1870	1	0.052264	0	
1	1	DZA	Algeria	1871	0	0.052798	0	
2	1	DZA	Algeria	1872	0	0.052274	0	
3	1	DZA	Algeria	1873	0	0.051680	0	
4	1	DZA	Algeria	1874	0	0.051308	0	

In [5]:

```
data.shape
```

Out[5]:

```
(1059, 14)
```

In [8]:

```
total_count = data.shape[0]
```

In [10]:

```
print('Всего строк: {}'.format(total_count))
```

```
Всего строк: 1059
```

In [11]:

```
data.columns
```

Out[11]:

```
Index(['case', 'cc3', 'country', 'year', 'systemic_crisis', 'exch_us
d',
       'domestic_debt_in_default', 'sovereign_external_debt_defaul
t',
       'gdp_weighted_default', 'inflation_annual_cpi', 'independenc
e',
       'currency_crises', 'inflation_crises', 'banking_crisis'],
      dtype='object')
```

In [12]:

```
data.dtypes
```

Out[12]:

```
case                                int64
cc3                                object
country                            object
year                                int64
systemic_crisis                     int64
exch_usd                          float64
domestic_debt_in_default            int64
sovereign_external_debt_default     int64
gdp_weighted_default              float64
inflation_annual_cpi              float64
independence                        int64
currency_crises                     int64
inflation_crises                    int64
banking_crisis                     object
dtype: object
```

In [14]:

```
#Пустые значения
for col in data.columns:
    temp_null_count = data[data[col].isnull()].shape[0]
    print('{} - {}'.format(col, temp_null_count))
```

```
case - 0
cc3 - 0
country - 0
year - 0
systemic_crisis - 0
exch_usd - 0
domestic_debt_in_default - 0
sovereign_external_debt_default - 0
gdp_weighted_default - 0
inflation_annual_cpi - 0
independence - 0
currency_crises - 0
inflation_crises - 0
banking_crisis - 0
```

In [15]:

```
data.describe()
```

Out[15]:

	case	year	systemic_crisis	exch_usd	domestic_debt_in_defau
count	1059.000000	1059.000000	1059.000000	1059.000000	1059.00000
mean	35.613787	1967.767705	0.077432	43.140831	0.03966
std	23.692402	33.530632	0.267401	111.475380	0.19525
min	1.000000	1860.000000	0.000000	0.000000	0.00000
25%	15.000000	1951.000000	0.000000	0.195350	0.00000
50%	38.000000	1973.000000	0.000000	0.868400	0.00000
75%	56.000000	1994.000000	0.000000	8.462750	0.00000
max	70.000000	2014.000000	1.000000	744.306139	1.00000

caseA - number which denotes a specific country

cc3A three letter country code

countryThe name of the country

yearThe year of the observation

systemic_crisis"0" means that no systemic crisis occurred in the year and "1" means that a systemic crisis occurred in the year.

exch_usdThe exchange rate of the country vis-a-vis the USD

domestic_debt_in_default"0" means that no sovereign domestic debt default occurred in the year and "1" means that a sovereign domestic debt default occurred in the year

sovereign_external_debt_default"0" means that no sovereign external debt default occurred in the year and "1" means that a sovereign external debt default occurred in the year

gdp_weighted_defaultThe total debt in default vis-a-vis the GDP

inflation_annual_cpiThe annual CPI Inflation rate

independence"0" means "no independence" and "1" means "independence"

currency_crises"0" means that no currency crisis occurred in the year and "1" means that a currency crisis occurred in the year

inflation_crises"0" means that no inflation crisis occurred in the year and "1" means that an inflation crisis occurred in the year

banking_crisis"no_crisis" means that no banking crisis occurred in the year and "crisis" means that a banking crisis occurred in the year

In [46]:

```
data['sovereign_external_debt_default'].unique()
```

Out[46]:

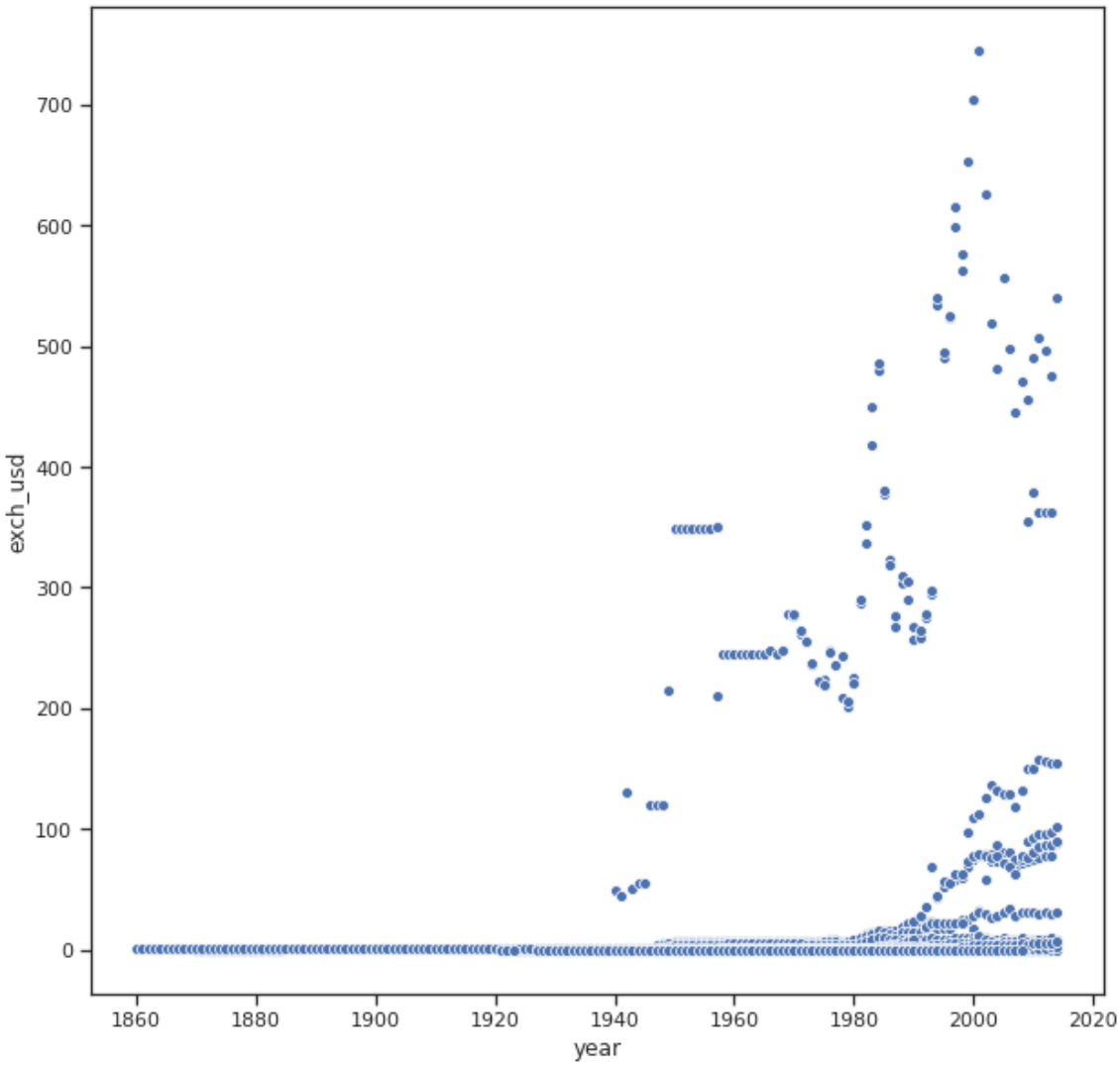
```
array([0, 1])
```

In [34]:

```
fig, ax = plt.subplots(figsize=(10,10))
sns.scatterplot(ax=ax, x='year', y='exch_usd', data=data)
```

Out[34]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f51f7de5ed0>
```

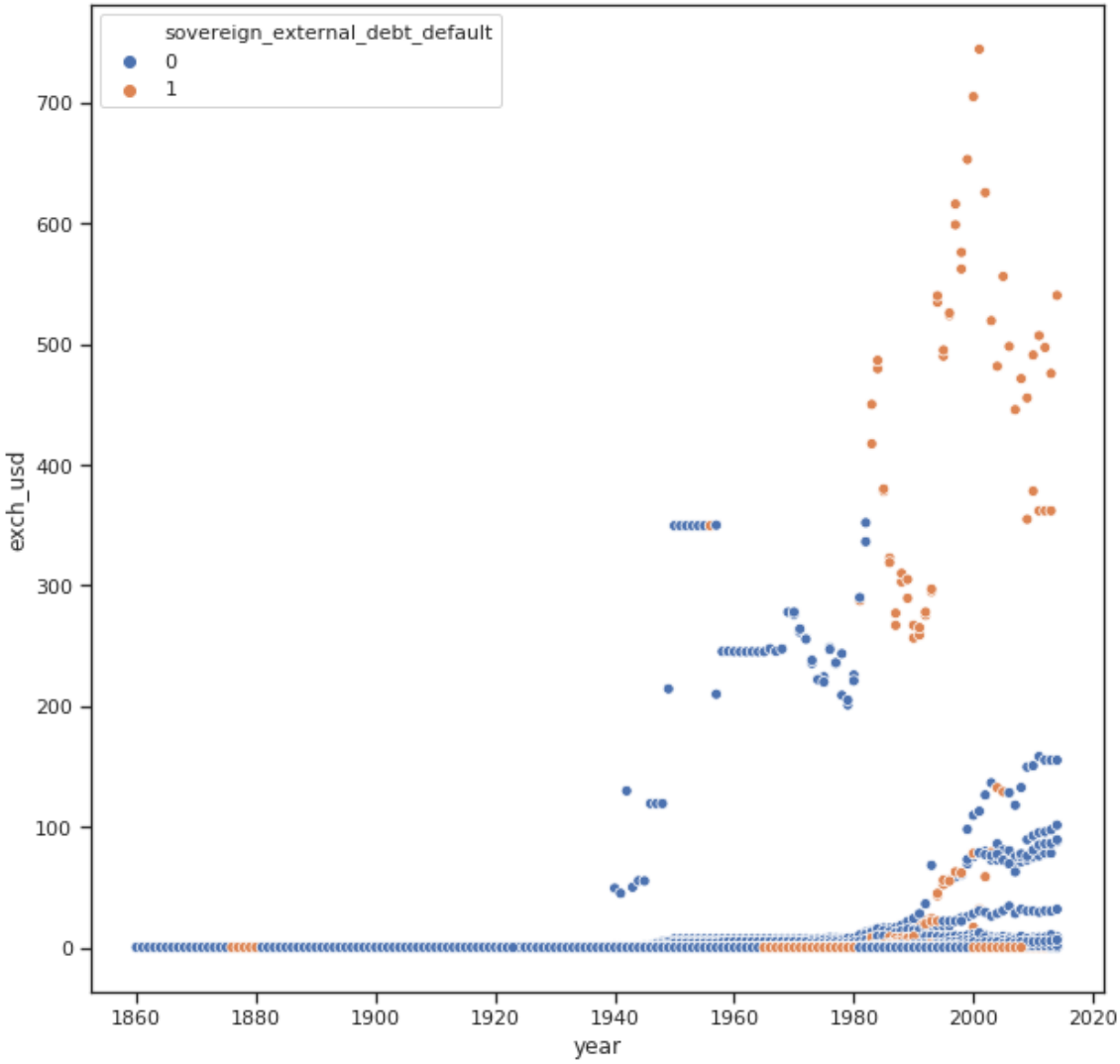


In [45]:

```
fig, ax = plt.subplots(figsize=(10,10))
sns.scatterplot(ax=ax, x='year', y='exch_usd', data=data, hue='sovereign_external_debt_default')
```

Out[45]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f51f7818a10>
```

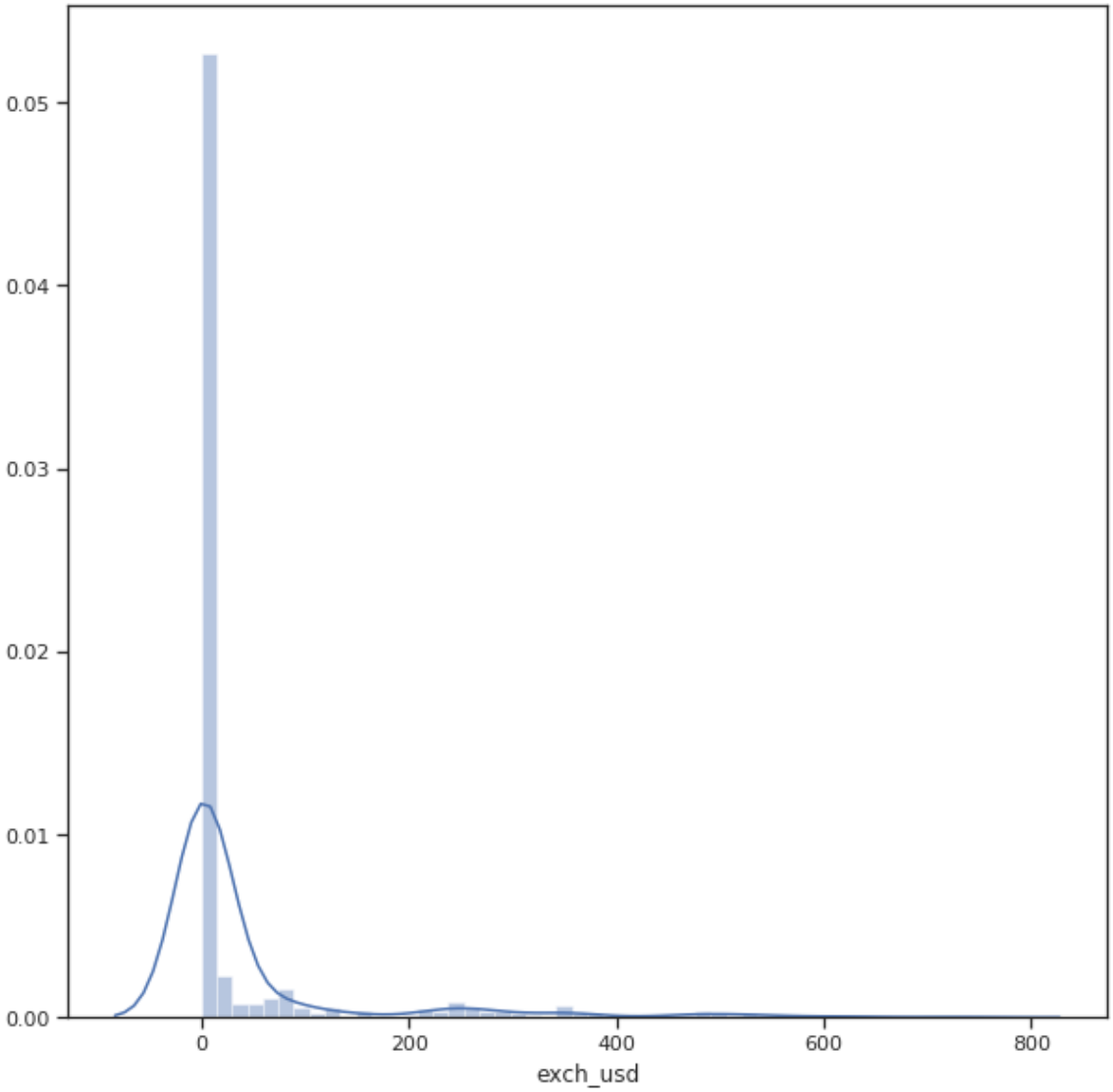


In [38]:

```
fig, ax = plt.subplots(figsize=(10,10))
sns.distplot(data['exch_usd'])
```

Out[38]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f51f7b13410>
```

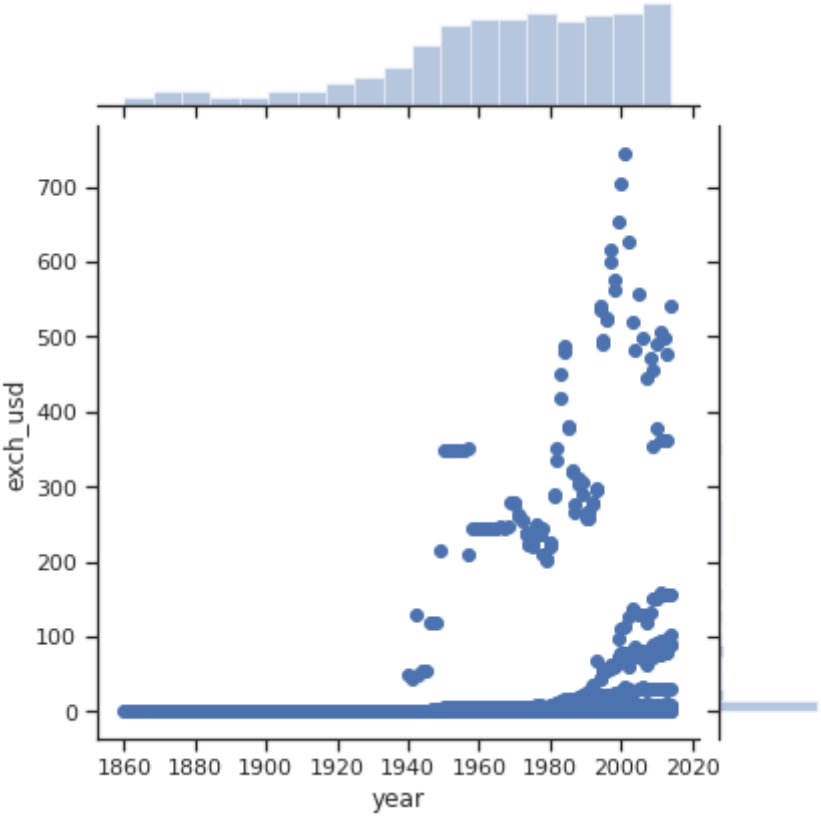


In [49]:

```
sns.jointplot(x='year', y='exch_usd', data=data)
```

Out[49]:

```
<seaborn.axisgrid.JointGrid at 0x7f51f7686b50>
```

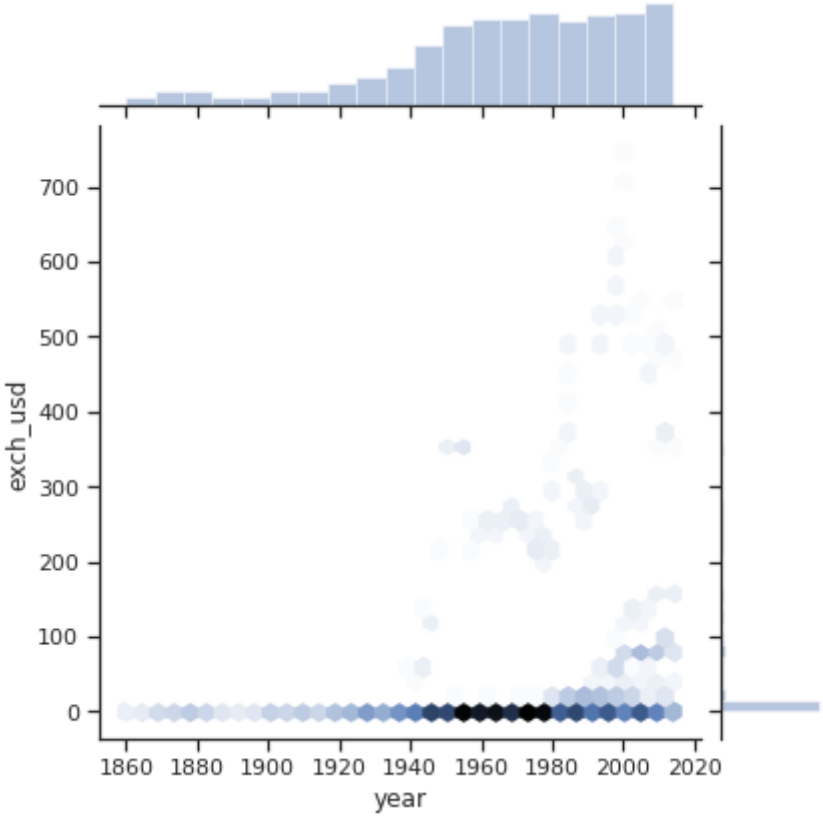


In [51]:

```
sns.jointplot(x='year', y='exch_usd', data=data, kind="hex")
```

Out[51]:

```
<seaborn.axisgrid.JointGrid at 0x7f51f72b1890>
```

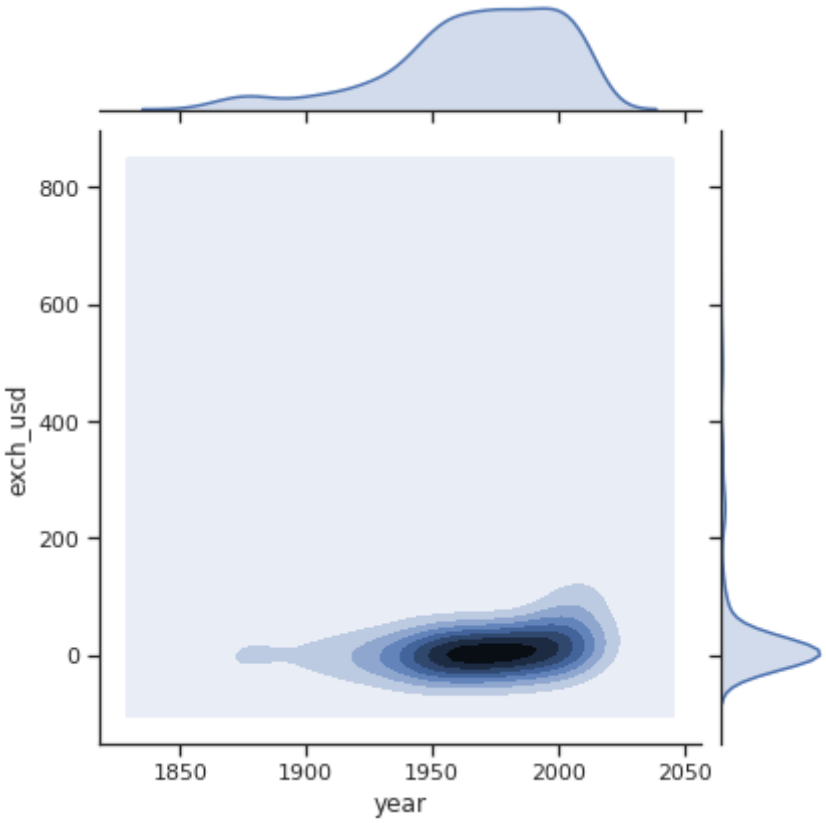


In [52]:

```
sns.jointplot(x='year', y='exch_usd', data=data, kind="kde")
```

Out[52]:

```
<seaborn.axisgrid.JointGrid at 0x7f51f7a41b10>
```

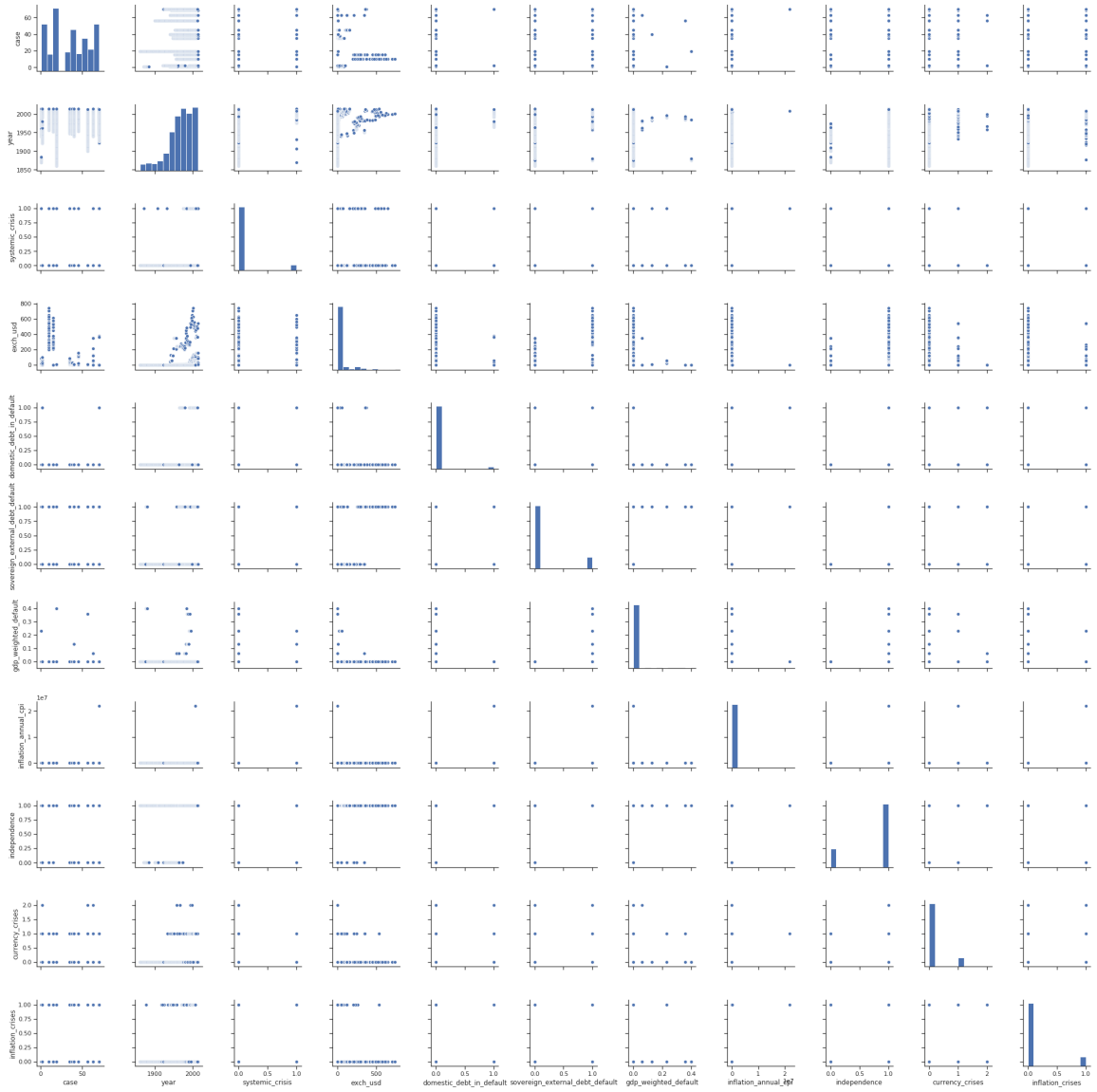


In [54]:

```
sns.pairplot(data)
```

Out[54]:

```
<seaborn.axisgrid.PairGrid at 0x7f51f3385c90>
```



In [ ]:

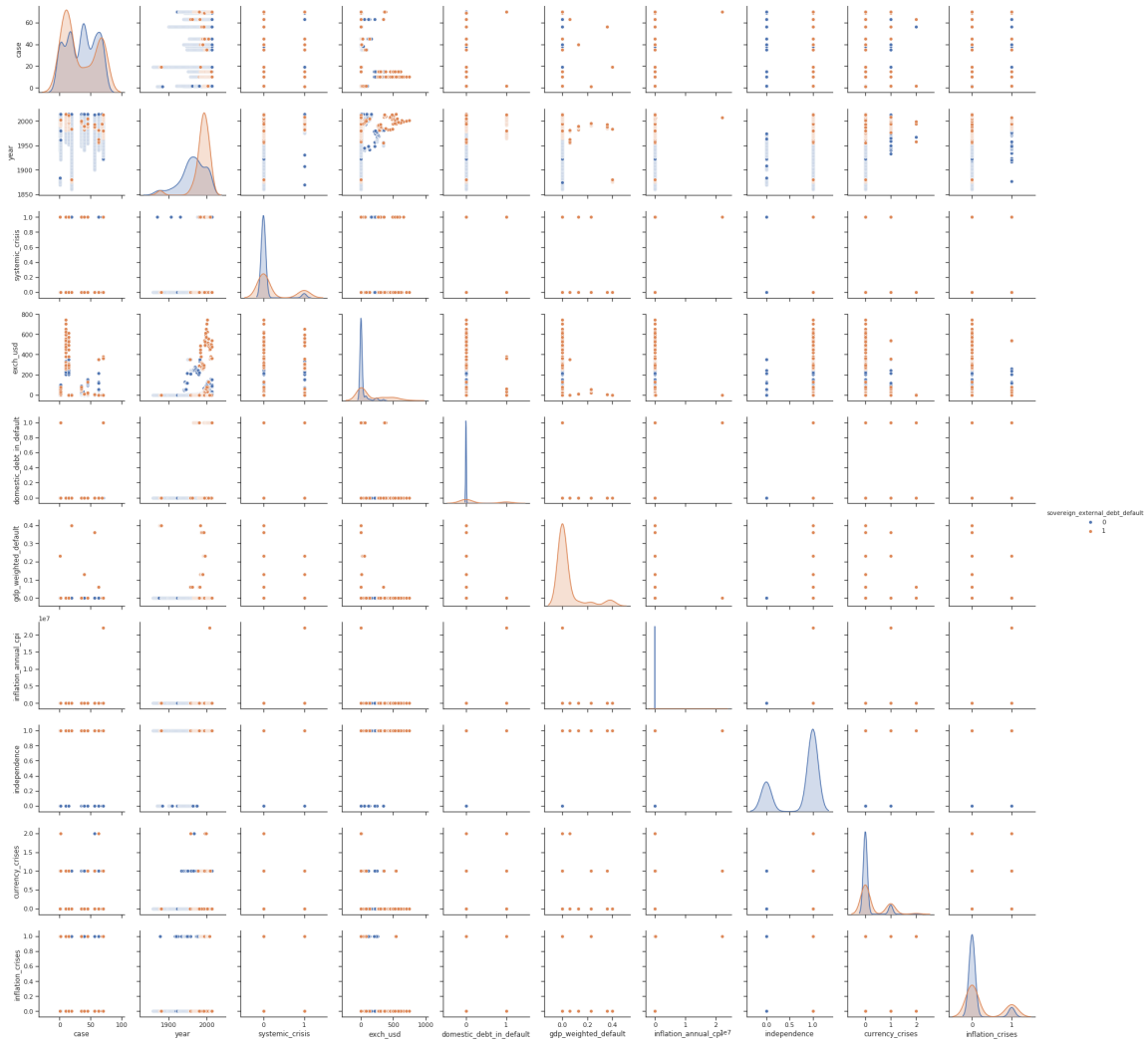
In [55]:

```
sns.pairplot(data, hue="sovereign_external_debt_default")
```

```
/home/lisobol/tensorflow_env/my_tensorflow/lib/python3.7/site-packag
es/seaborn/distributions.py:288: UserWarning: Data must have varianc
e to compute a kernel density estimate.
  warnings.warn(msg, UserWarning)
/home/lisobol/tensorflow_env/my_tensorflow/lib/python3.7/site-packag
es/seaborn/distributions.py:288: UserWarning: Data must have varianc
e to compute a kernel density estimate.
  warnings.warn(msg, UserWarning)
```

Out[55]:

```
<seaborn.axisgrid.PairGrid at 0x7f51e928cfd0>
```

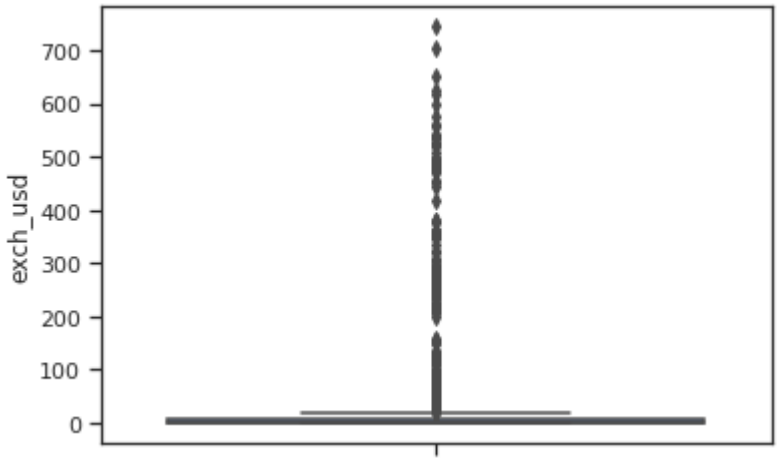


In [57]:

```
sns.boxplot(y=data['exch_usd'])
```

Out[57]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f51e4321e50>
```

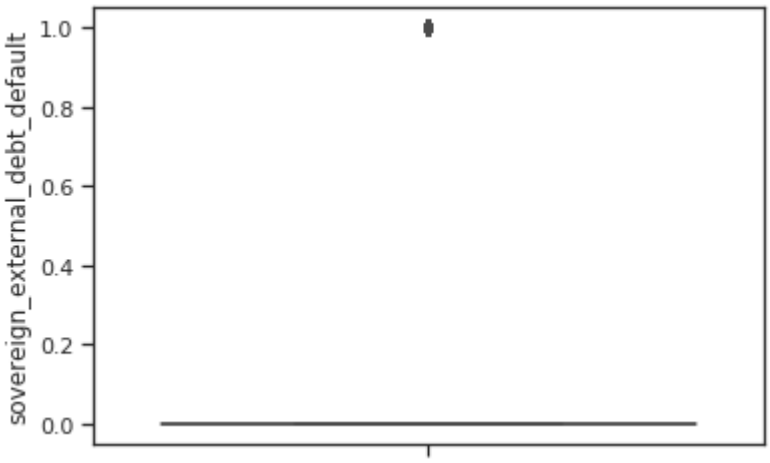


In [58]:

```
sns.boxplot(y=data['sovereign_external_debt_default'])
```

Out[58]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f51e42d9550>
```

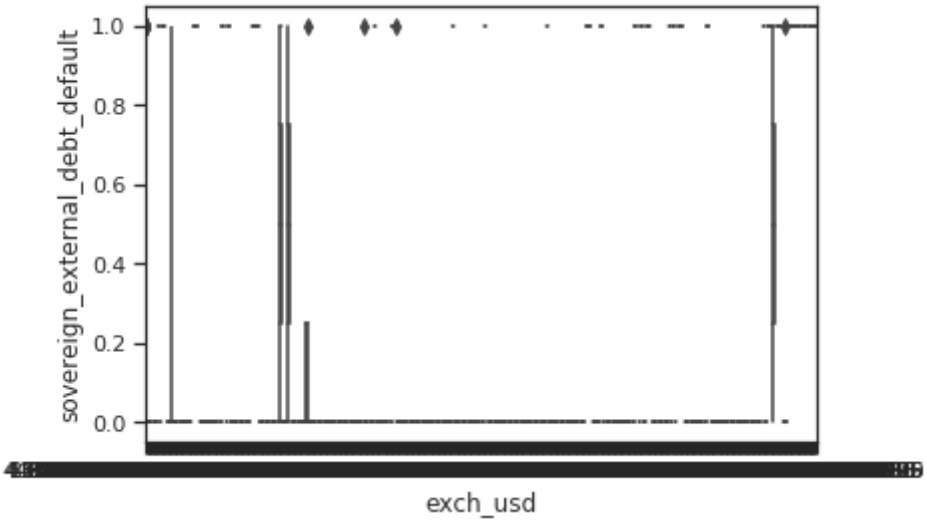


In [59]:

```
sns.boxplot(x='exch_usd', y='sovereign_external_debt_default', data=data)
```

Out[59]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f51e34e75d0>
```

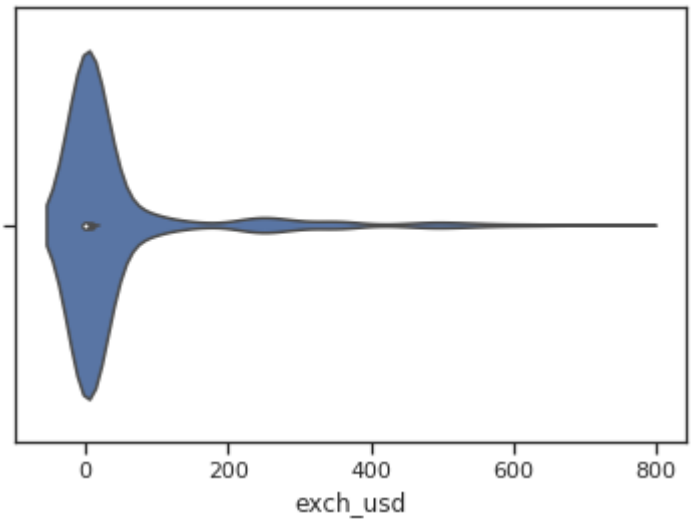


In [60]:

```
sns.violinplot(x=data['exch_usd'])
```

Out[60]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f51df1c9a10>
```

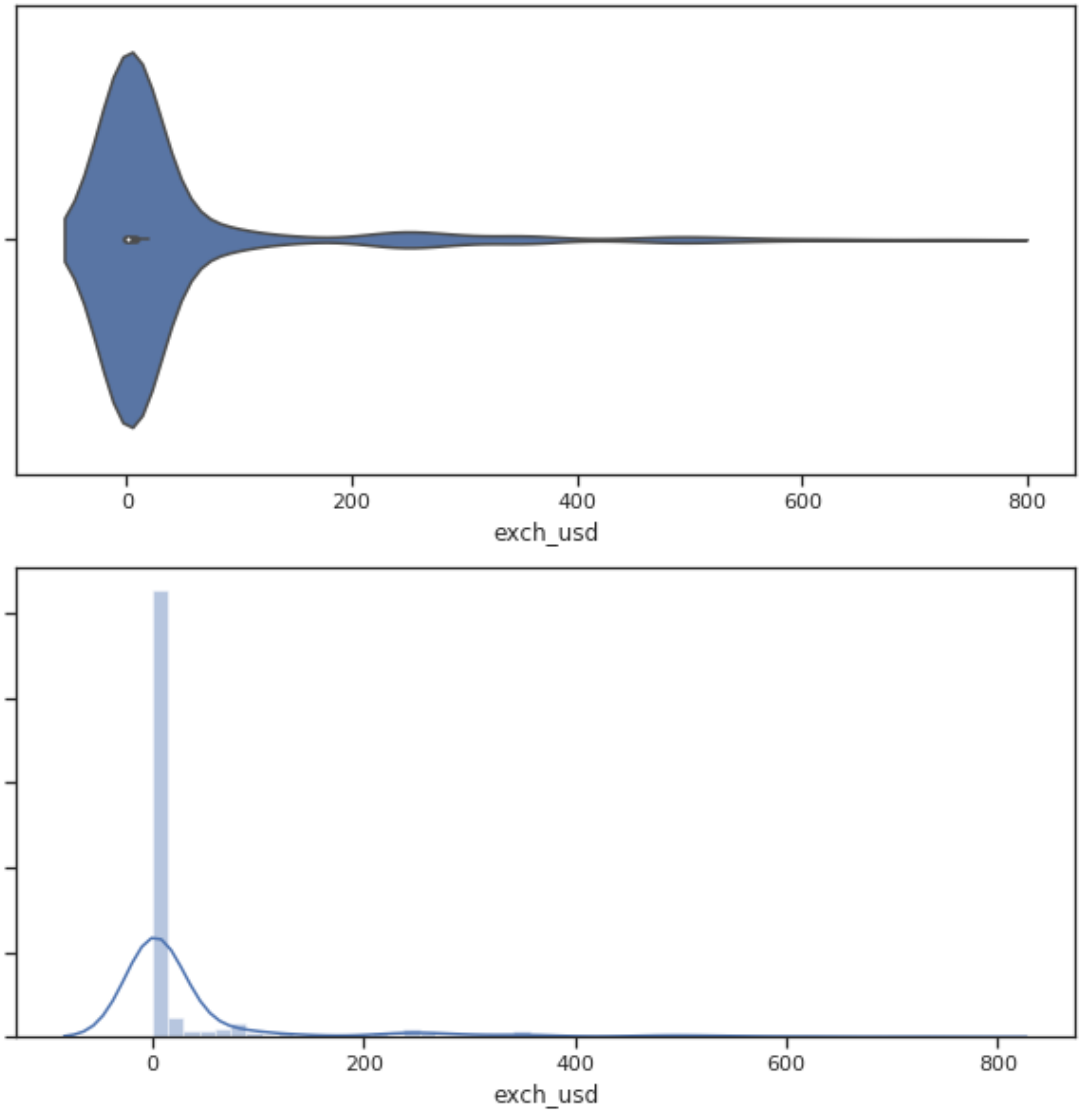


In [61]:

```
fig, ax = plt.subplots(2, 1, figsize=(10,10))
sns.violinplot(ax=ax[0], x=data['exch_usd'])
sns.distplot(data['exch_usd'], ax=ax[1])
```

Out[61]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f51deb8e150>
```

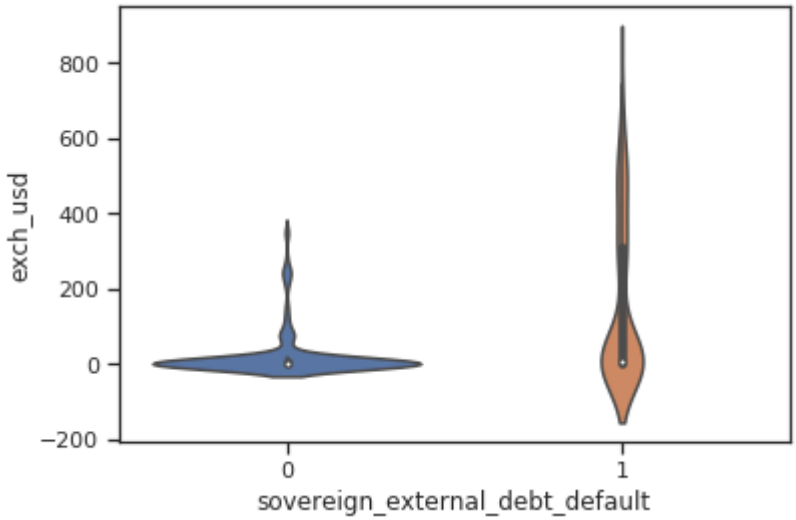


In [62]:

```
# Распределение параметра exch_usd сгруппированные по sovereign_external_debt_default.
sns.violinplot(x='sovereign_external_debt_default', y='exch_usd', data=data)
```

Out[62]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f51e34b5290>
```

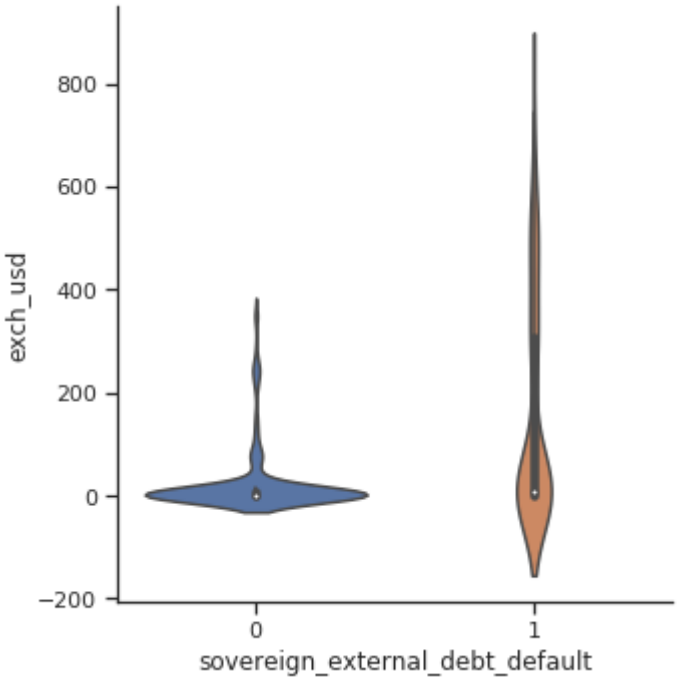


In [63]:

```
sns.catplot(y='exch_usd', x='sovereign_external_debt_default', data=data, kind=
"violin", split=True)
```

Out[63]:

```
<seaborn.axisgrid.FacetGrid at 0x7f51dead7890>
```



In [64]:

```
data.corr()
```

Out[64]:

	case	year	systemic_crisis	exch_usd	dome
case	1.000000	0.115574	0.010991	-0.231976	
year	0.115574	1.000000	0.197450	0.248757	
systemic_crisis	0.010991	0.197450	1.000000	0.202687	
exch_usd	-0.231976	0.248757	0.202687	1.000000	
domestic_debt_in_default	0.128358	0.136828	0.122158	0.005253	
sovereign_external_debt_default	-0.039262	0.271890	0.249850	0.422890	
gdp_weighted_default	-0.032981	-0.054670	0.005274	-0.040726	
inflation_annual_cpi	0.044762	0.037035	0.106452	-0.011947	
independence	0.021858	0.407360	0.147083	0.126034	
currency_crises	0.095339	0.189390	0.112751	-0.056472	
inflation_crises	0.006405	0.098630	0.172562	-0.063783	

In [65]:

```
data.corr(method='pearson')
```

Out[65]:

	case	year	systemic_crisis	exch_usd	dome
case	1.000000	0.115574	0.010991	-0.231976	
year	0.115574	1.000000	0.197450	0.248757	
systemic_crisis	0.010991	0.197450	1.000000	0.202687	
exch_usd	-0.231976	0.248757	0.202687	1.000000	
domestic_debt_in_default	0.128358	0.136828	0.122158	0.005253	
sovereign_external_debt_default	-0.039262	0.271890	0.249850	0.422890	
gdp_weighted_default	-0.032981	-0.054670	0.005274	-0.040726	
inflation_annual_cpi	0.044762	0.037035	0.106452	-0.011947	
independence	0.021858	0.407360	0.147083	0.126034	
currency_crises	0.095339	0.189390	0.112751	-0.056472	
inflation_crises	0.006405	0.098630	0.172562	-0.063783	

In [66]:

```
data.corr(method='kendall')
```

Out[66]:

	case	year	systemic_crisis	exch_usd	domes
case	1.000000	0.026923	0.009063	-0.187706	
year	0.026923	1.000000	0.186655	0.368117	
systemic_crisis	0.009063	0.186655	1.000000	0.142943	
exch_usd	-0.187706	0.368117	0.142943	1.000000	
domestic_debt_in_default	0.130510	0.121203	0.122158	-0.114600	
sovereign_external_debt_default	-0.030077	0.260867	0.249850	0.098473	
gdp_weighted_default	-0.000727	0.017020	0.014215	0.038681	
inflation_annual_cpi	0.019151	0.217394	0.126022	0.101890	
independence	0.021411	0.416064	0.147083	0.247936	
currency_crises	0.087497	0.158605	0.119988	0.013929	
inflation_crises	0.012142	0.072022	0.172562	-0.023109	

In [67]:

```
data.corr(method='spearman')
```

Out[67]:

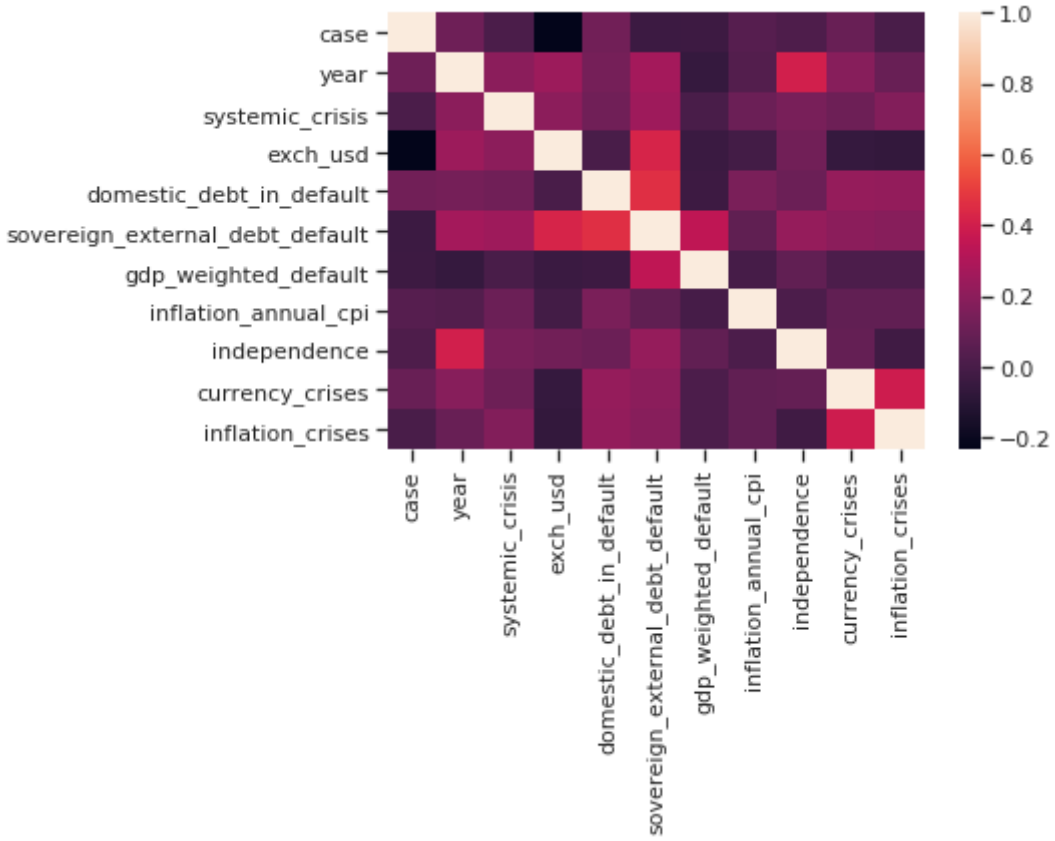
	case	year	systemic_crisis	exch_usd	domes
case	1.000000	0.039273	0.010663	-0.250869	
year	0.039273	1.000000	0.227475	0.512162	
systemic_crisis	0.010663	0.227475	1.000000	0.174649	
exch_usd	-0.250869	0.512162	0.174649	1.000000	
domestic_debt_in_default	0.153555	0.147710	0.122158	-0.140019	
sovereign_external_debt_default	-0.035389	0.317916	0.249850	0.120315	
gdp_weighted_default	-0.000585	0.019954	0.014295	0.047506	
inflation_annual_cpi	0.026830	0.309844	0.154250	0.154363	
independence	0.025191	0.507054	0.147083	0.302930	
currency_crises	0.103034	0.193589	0.120208	0.017115	
inflation_crises	0.014286	0.087773	0.172562	-0.028235	

In [68]:

```
sns.heatmap(data.corr())
```

Out[68]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f51dea2f050>
```

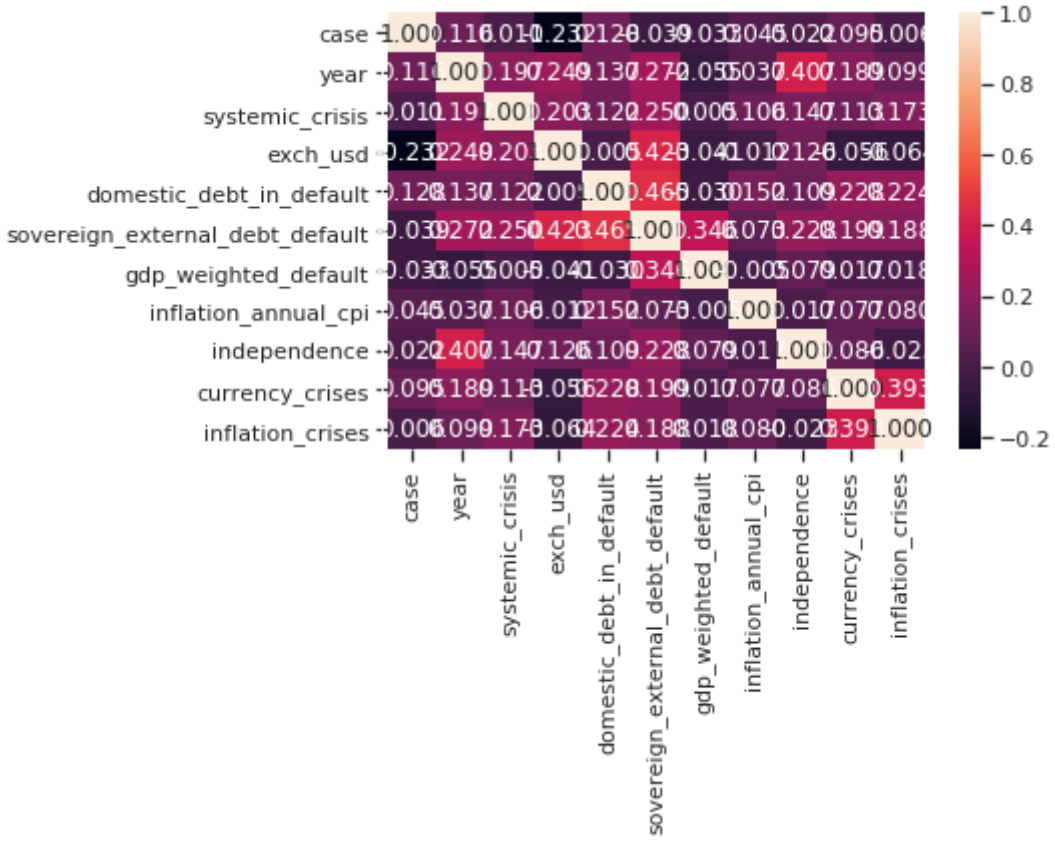


In [69]:

```
sns.heatmap(data.corr(), annot=True, fmt='.3f')
```

Out[69]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f51fbb5ae50>
```

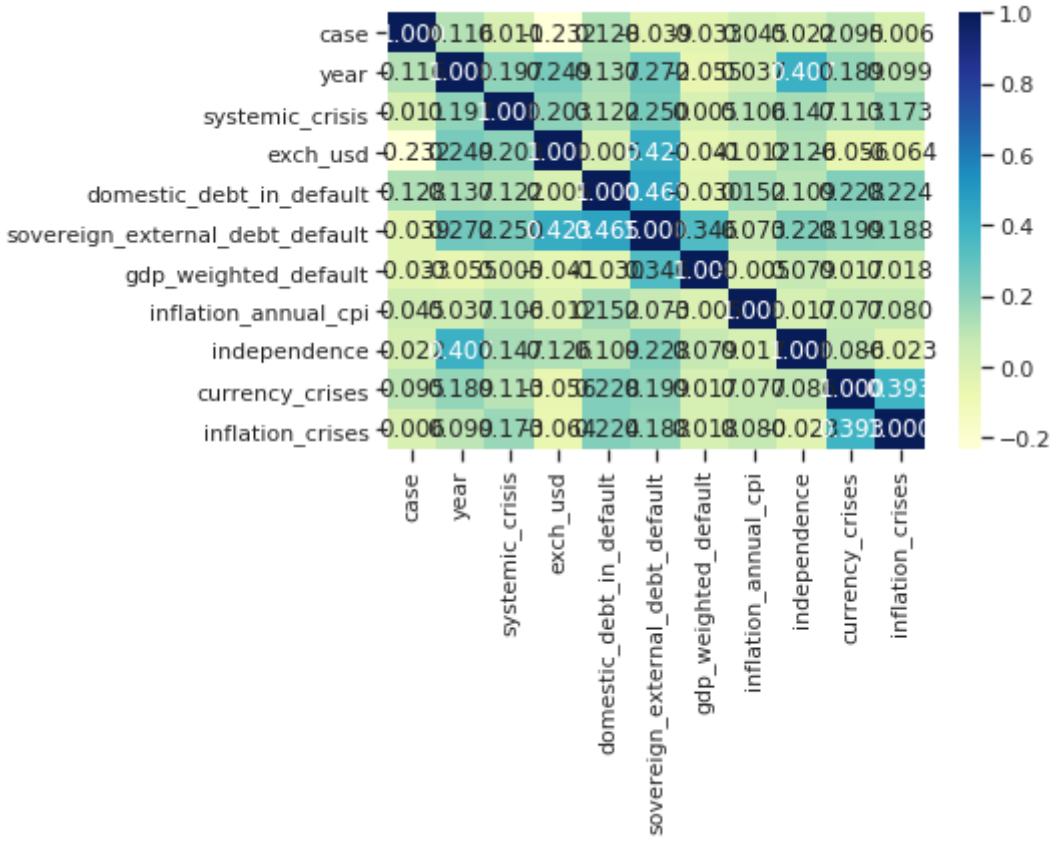


In [70]:

```
sns.heatmap(data.corr(), cmap='YlGnBu', annot=True, fmt='.3f')
```

Out[70]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f51de6a9a10>
```

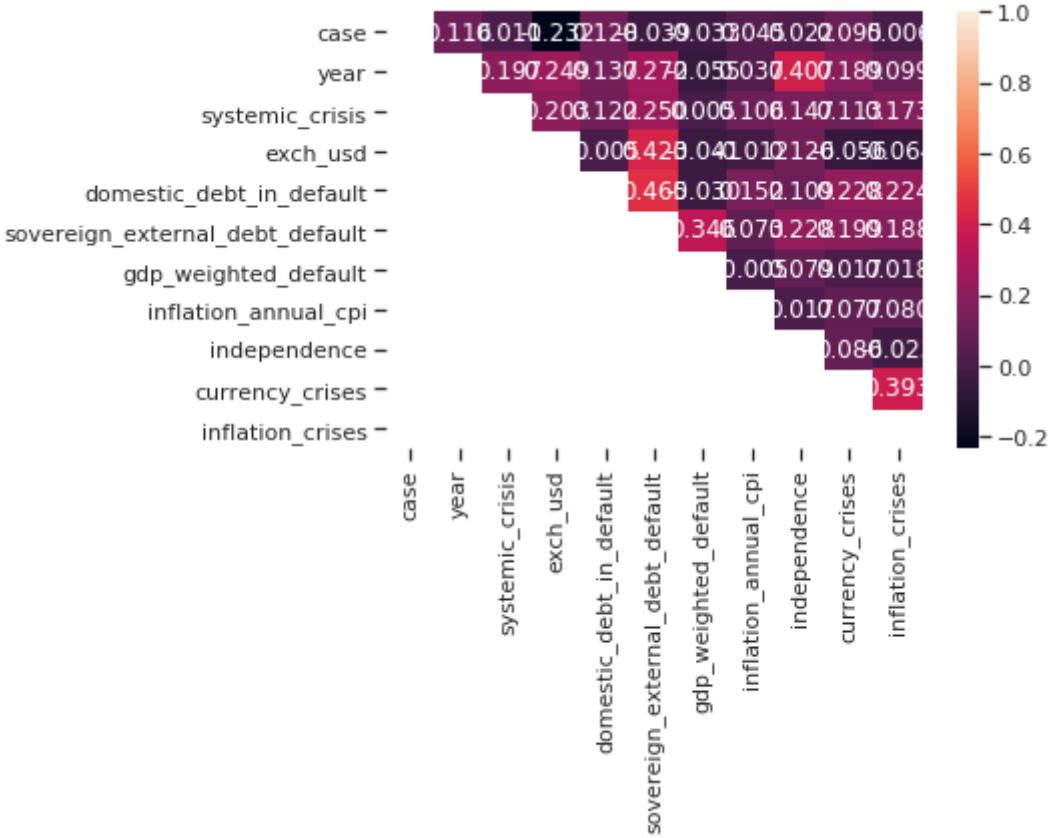


In [73]:

```
# Треугольный вариант матрицы
mask = np.zeros_like(data.corr(), dtype=np.bool)
# чтобы оставить нижнюю часть матрицы
# mask[np.triu_indices_from(mask)] = True
# чтобы оставить верхнюю часть матрицы
mask[np.tril_indices_from(mask)] = True
sns.heatmap(data.corr(), mask=mask, annot=True, fmt='.3f')
```

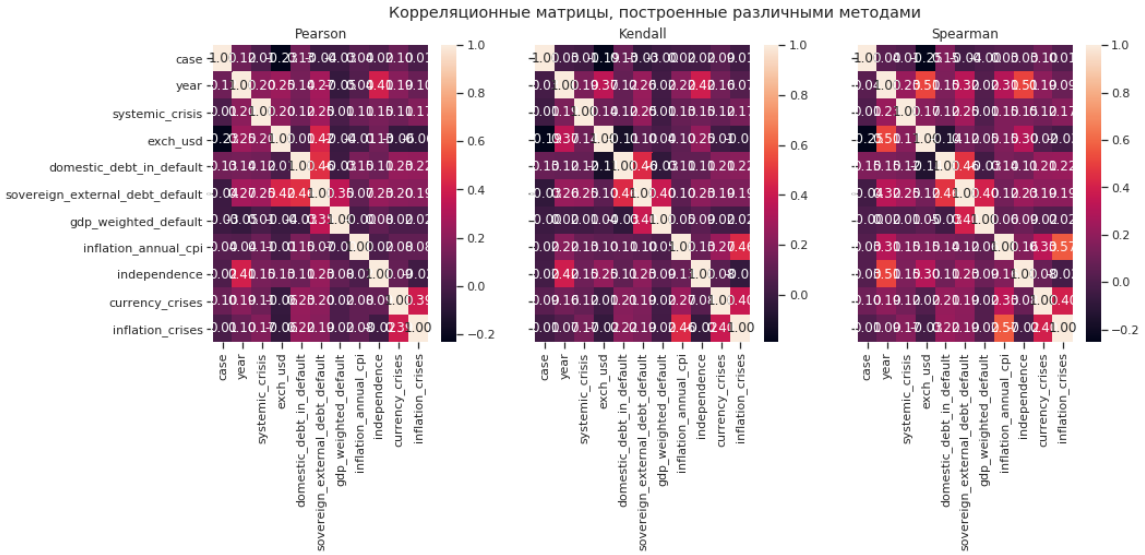
Out[73]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f51de3a63d0>
```



In [74]:

```
fig, ax = plt.subplots(1, 3, sharex='col', sharey='row', figsize=(15,5))
sns.heatmap(data.corr(method='pearson'), ax=ax[0], annot=True, fmt='.2f')
sns.heatmap(data.corr(method='kendall'), ax=ax[1], annot=True, fmt='.2f')
sns.heatmap(data.corr(method='spearman'), ax=ax[2], annot=True, fmt='.2f')
fig.suptitle('Корреляционные матрицы, построенные различными методами')
ax[0].title.set_text('Pearson')
ax[1].title.set_text('Kendall')
ax[2].title.set_text('Spearman')
```



In [ ]: