Описательная статистика "US Homicide Reports"

Загрузка данных и их описание

Датасет https://www.kaggle.com/murderaccountability/homicide-reports (https://www.kaggle.com/murderaccountability/homicide-reports)

In [1]:

```
import numpy as np
import pandas as pd
from pandas import DataFrame
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [4]:

```
df = pd.read_csv('database.csv')
```

In [7]:

```
df.head(3).info
```

Out[7]:

```
<bound method DataFrame.info of    Record ID Agency Code Agency Name       A
gency Type       City   State  \
0          1     AK00101   Anchorage  Municipal Police  Anchorage  Alaska
1          2     AK00101   Anchorage  Municipal Police  Anchorage  Alaska
2          3     AK00101   Anchorage  Municipal Police  Anchorage  Alaska

   Year    Month  Incident             Crime Type  ... Victim Ethnicity  \
0  1980  January         1  Murder or Manslaughter  ...          Unknown
1  1980    March         1  Murder or Manslaughter  ...          Unknown
2  1980    March         2  Murder or Manslaughter  ...          Unknown

  Perpetrator Sex  Perpetrator Age               Perpetrator Race  \
0            Male               15  Native American/Alaska Native
1            Male               42                          White
2         Unknown                0                        Unknown

  Perpetrator Ethnicity  Relationship         Weapon Victim Count  \
0               Unknown  Acquaintance   Blunt Object            0
1               Unknown  Acquaintance  Strangulation            0
2               Unknown       Unknown        Unknown            0

  Perpetrator Count Record Source
0                 0           FBI
1                 0           FBI
2                 0           FBI

[3 rows x 24 columns]>
```

In [5]:

```
1 df.describe()
```

Out[5]:

	Record ID	Year	Incident	Victim Age	Victim Count	Perpetrat Cou
count	638454.00000	638454.000000	638454.000000	638454.000000	638454.000000	638454.00000
mean	319227.50000	1995.801102	22.967924	35.033512	0.123334	0.18522
std	184305.93872	9.927693	92.149821	41.628306	0.537733	0.58549
min	1.00000	1980.000000	0.000000	0.000000	0.000000	0.00000
25%	159614.25000	1987.000000	1.000000	22.000000	0.000000	0.00000
50%	319227.50000	1995.000000	2.000000	30.000000	0.000000	0.00000
75%	478840.75000	2004.000000	10.000000	42.000000	0.000000	0.00000
max	638454.00000	2014.000000	999.000000	998.000000	10.000000	10.00000

In [6]:

```
1 df
```

Out[6]:

	Record ID	Agency Code	Agency Name	Agency Type	City	State	Year	Month	Incident	Crime Typ
0	1	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	1980	January	1	Murder c Manslaughte
1	2	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	1980	March	1	Murder c Manslaughte
2	3	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	1980	March	2	Murder c Manslaughte
3	4	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	1980	April	1	Murder c Manslaughte
4	5	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	1980	April	2	Murder c Manslaughte
...	...	...	...	...	...	...	...	...	...	

In [7]:

```
1 df.isnull().sum()
```

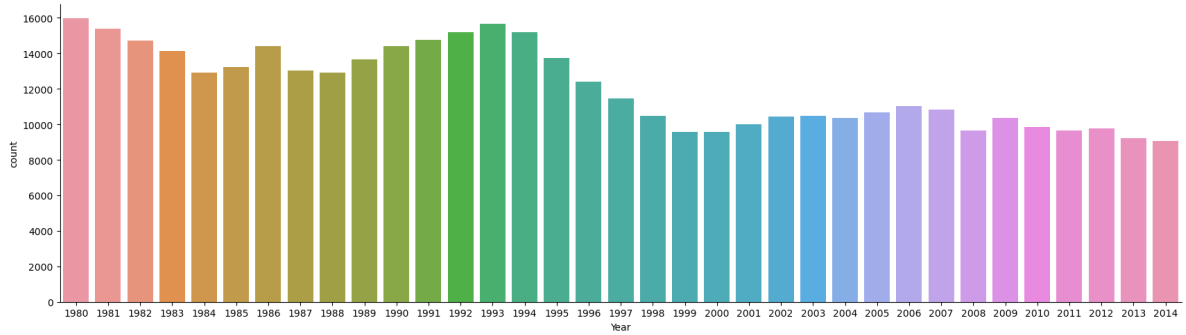
Out[7]:

```
Record ID                0
Agency Code              0
Agency Name              0
Agency Type              0
City                     0
State                    0
Year                     0
Month                    0
Incident                 0
Crime Type               0
Crime Solved             0
Victim Sex               0
Victim Age               0
Victim Race              0
Victim Ethnicity         0
Perpetrator Sex          0
Perpetrator Age          0
Perpetrator Race         0
Perpetrator Ethnicity    0
Relationship             0
Weapon                   0
Victim Count             0
Perpetrator Count        0
Record Source            0
dtype: int64
```

Описательная статистика

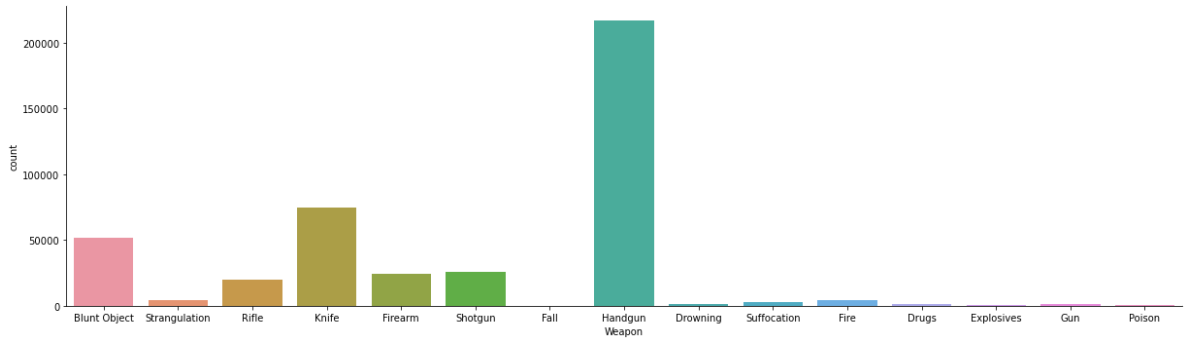
In [114]:

```
1 sns.catplot('Year', data=df, kind='count').fig.set_size_inches(20,5)
```



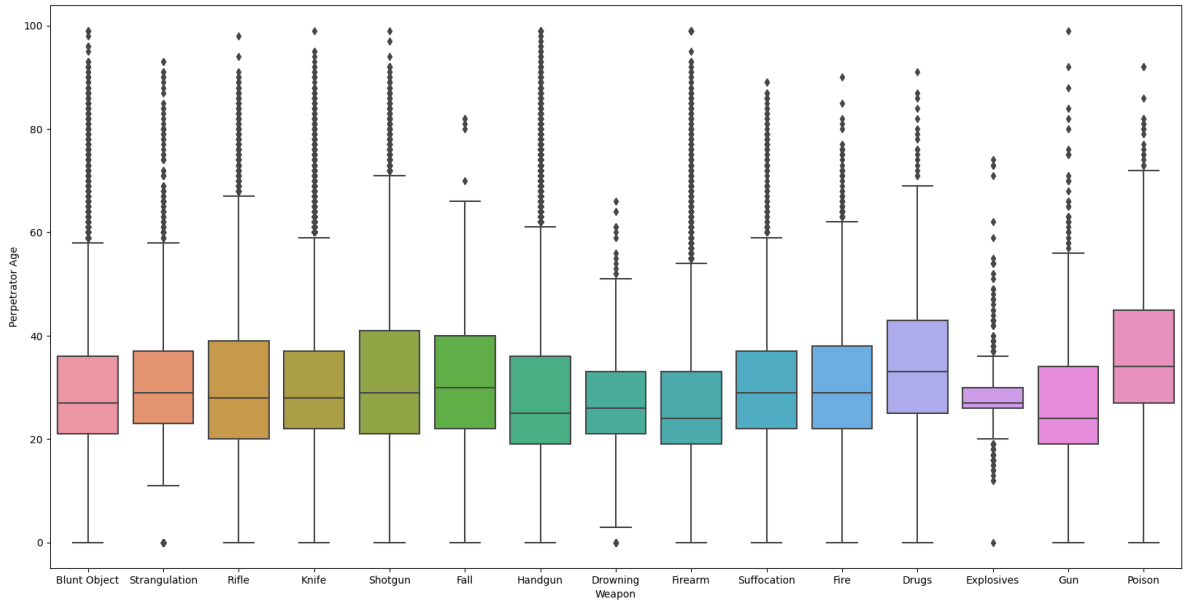
In [9]:

```
df.drop(df[df['Perpetrator Sex'] == 'Unknown'].index, inplace = True)
df.drop(df[df['Weapon'  ] == 'Unknown'].index, inplace = True)
sns.catplot('Weapon', data = df, kind ='count').fig.set_size_inches(20,5)
```



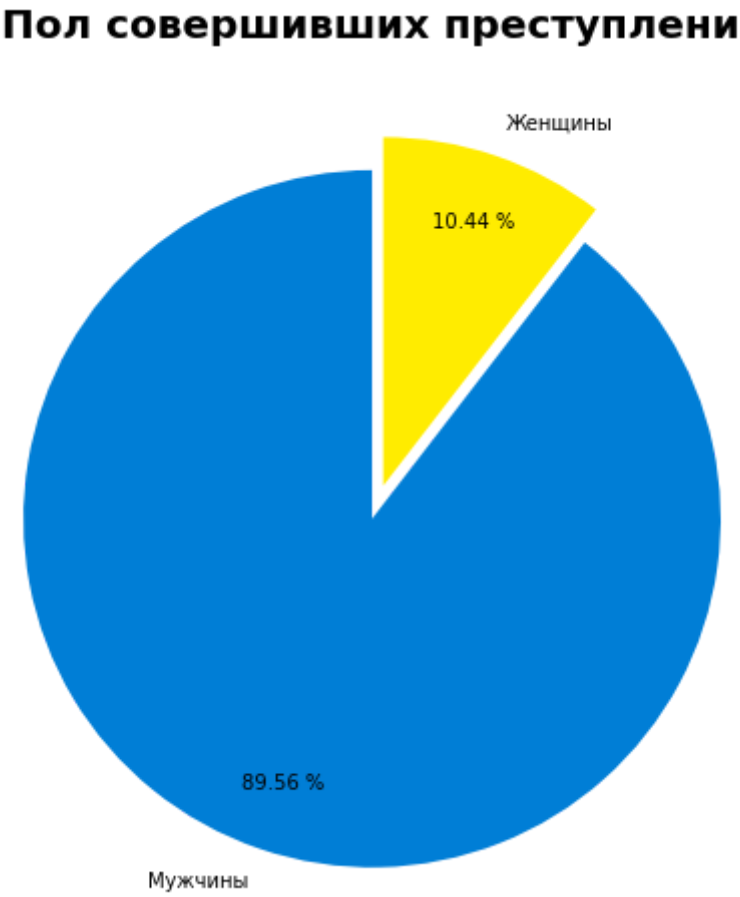
In [133]:

```
df['Perpetrator Age'] = df['Perpetrator Age'].astype(int)

plt.figure(figsize=(20,10))
sns.boxplot(x=df['Weapon'], y = df['Perpetrator Age'])
plt.show()
```



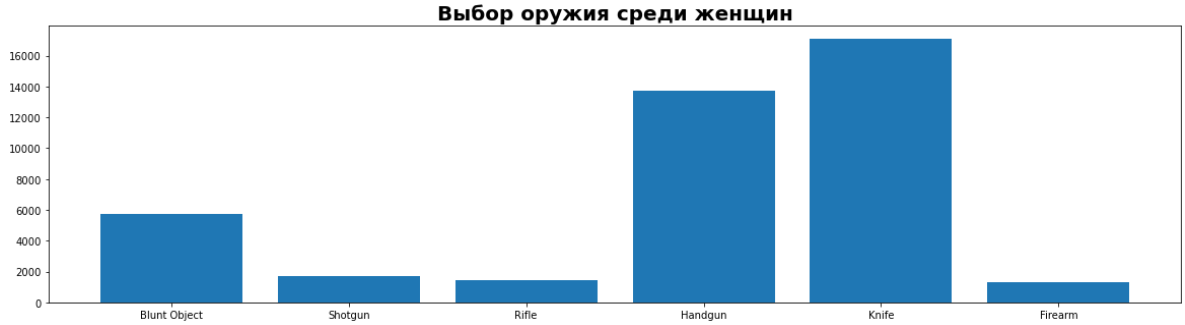
In [10]:

```
male = df.loc[df['Perpetrator Sex'] == 'Male'].count()[0]
female = df.loc[df['Perpetrator Sex'] == 'Female'].count()[0]

labels = ['Мужчины','Женщины']
colors = ['#007ED6','#FFEC00']
psex = [male, female]
explode = (0.1, 0)

plt.figure(figsize=(8,8))
plt.pie(psex, labels = labels, colors = colors, startangle = 90, autopct = '%.2f %%', p
plt.title('Пол совершивших преступление', fontdict = {'fontweight':'bold','fontsize':20
plt.show()
```



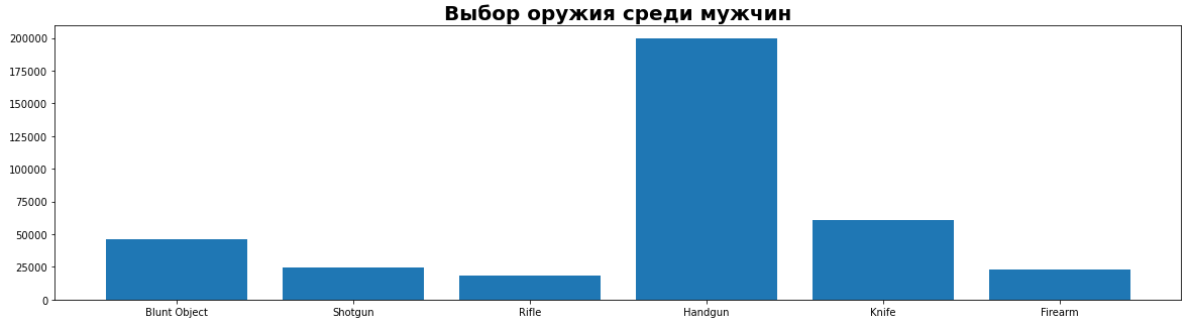
In [19]:

```
fmblunt = df.loc[(df['Perpetrator Sex'] == 'Female') & (df['Weapon'] == 'Blunt Object')
fmshotgun = df.loc[(df['Perpetrator Sex'] == 'Female') & (df['Weapon'] == 'Shotgun')].c
fmrifle = df.loc[(df['Perpetrator Sex'] == 'Female') & (df['Weapon'] == 'Rifle')].count
fmhandgun = df.loc[(df['Perpetrator Sex'] == 'Female') & (df['Weapon'] == 'Knife')].cou
fmknife = df.loc[(df['Perpetrator Sex'] == 'Female') & (df['Weapon'] == 'Handgun')].cou
fmfirearm = df.loc[(df['Perpetrator Sex'] == 'Female') & (df['Weapon'] == 'Firearm')].c

fmcount = [fmblunt, fmshotgun, fmrifle, fmhandgun, fmknife, fmfirearm]
fmtype = ['Blunt Object', 'Shotgun', 'Rifle', 'Handgun', 'Knife', 'Firearm']

plt.figure(figsize=(20,5))
plt.title("Выбор оружия среди женщин", fontdict={'fontweight':'bold','fontsize':20})
plt.bar(fmtype, fmcount)
plt.show()
```



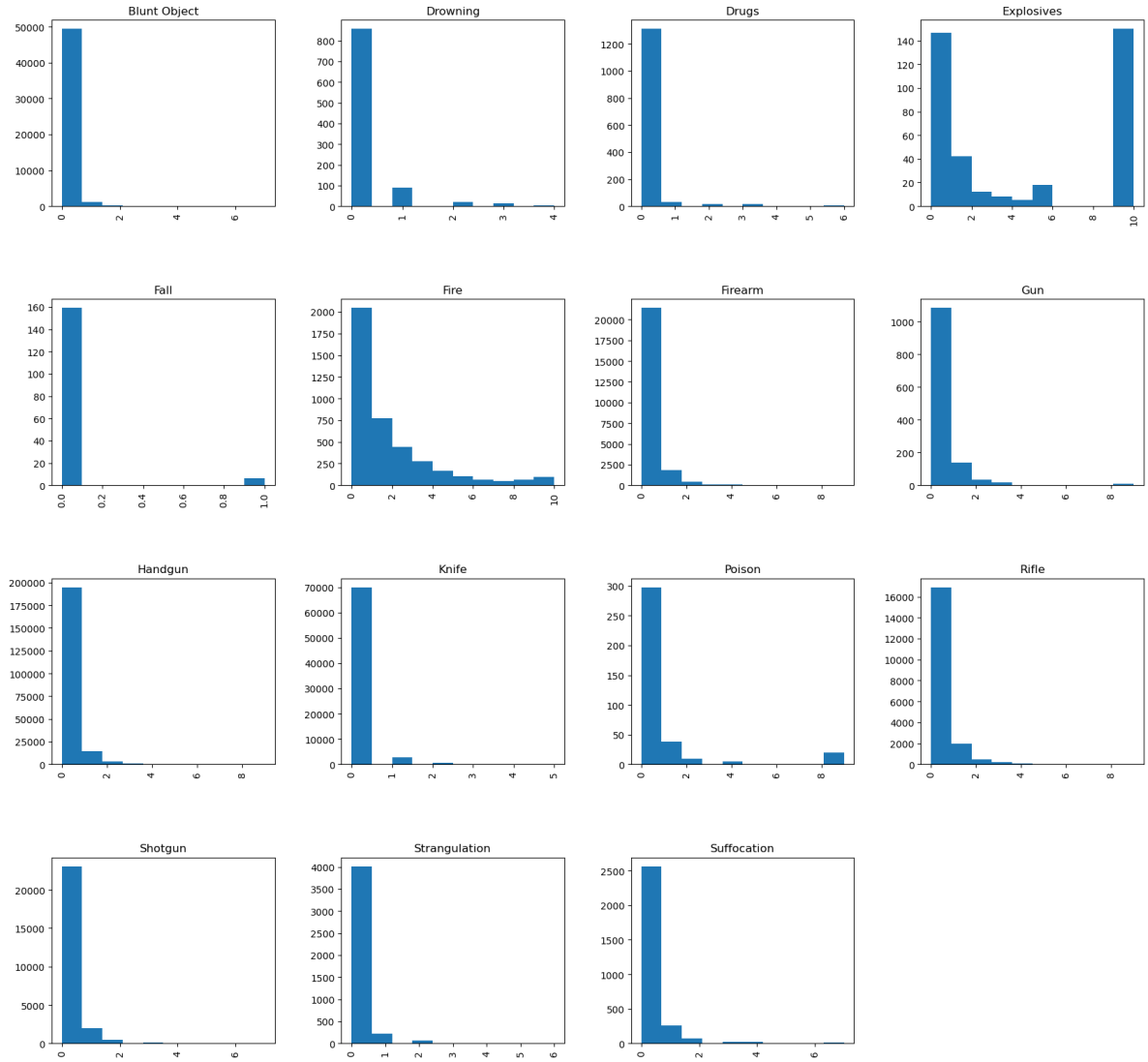
In [20]:

```
mlblunt = df.loc[(df['Perpetrator Sex'] == 'Male') & (df['Weapon'] == 'Blunt Object')].
mlrifle = df.loc[(df['Perpetrator Sex'] == 'Male') & (df['Weapon'] == 'Rifle')].count()
mlknife = df.loc[(df['Perpetrator Sex'] == 'Male') & (df['Weapon'] == 'Knife')].count()
mlhandgun = df.loc[(df['Perpetrator Sex'] == 'Male') & (df['Weapon'] == 'Handgun')].cou
mlshotgun = df.loc[(df['Perpetrator Sex'] == 'Male') & (df['Weapon'] == 'Shotgun')].cou
mlfirearm = df.loc[(df['Perpetrator Sex'] == 'Male') & (df['Weapon'] == 'Firearm')].cou

mlcount = [mlblunt, mlshotgun, mlrifle, mlhandgun, mlknife, mlfirearm]
mltype = ['Blunt Object', 'Shotgun', 'Rifle', 'Handgun', 'Knife', 'Firearm']

plt.figure(figsize=(20,5))
plt.title("Выбор оружия среди мужчин", fontdict = {'fontweight':'bold','fontsize':20})
plt.bar(mltype, mlcount)
plt.show()
```



In [105]:

```
df.hist('Victim Count', by = 'Weapon', figsize = [20,20], bins=10)
plt.show()
```



In [ ]:

```

```