In [2]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

Подготовка данных

In [4]:

```
ind_data = pd.read_csv('india.csv', delimiter=',')
```

In [5]:

```
ind_data.head()
```

Out[5]:

	stn_code	sampling_date	state	location	agency	type	so2	no2	rspm	spm	location_monitoring_station	pm2_5	dat
0	150	February - M021990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	4.8	17.4	NaN	NaN	NaN	NaN	199(02-0
1	151	February - M021990	Andhra Pradesh	Hyderabad	NaN	Industrial Area	3.1	7.0	NaN	NaN	NaN	NaN	199(02-0
2	152	February - M021990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	6.2	28.5	NaN	NaN	NaN	NaN	199(02-0
3	150	March - M031990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	6.3	14.7	NaN	NaN	NaN	NaN	199(03-0
4	151	March - M031990	Andhra Pradesh	Hyderabad	NaN	Industrial Area	4.7	7.5	NaN	NaN	NaN	NaN	199(03-0

In [9]:

```
ind_data.type.unique()
```

Out[9]:

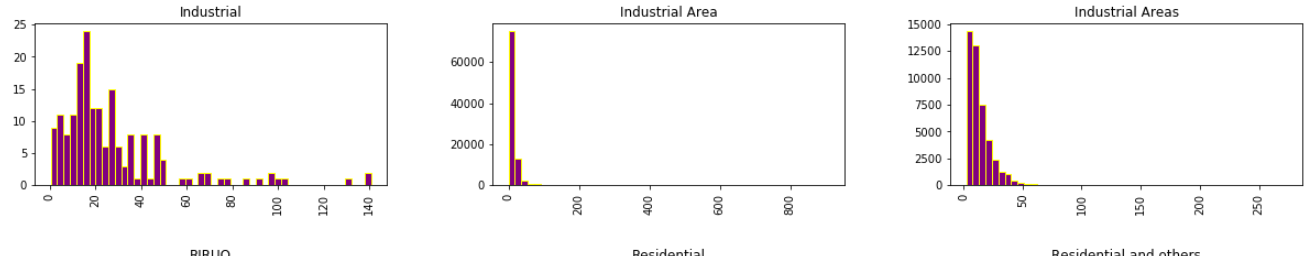
```
array(['Residential, Rural and other Areas', 'Industrial Area', nan,
       'Sensitive Area', 'Industrial Areas', 'Residential and others',
       'Sensitive Areas', 'Industrial', 'Residential', 'RIRUO',
       'Sensitive'], dtype=object)
```

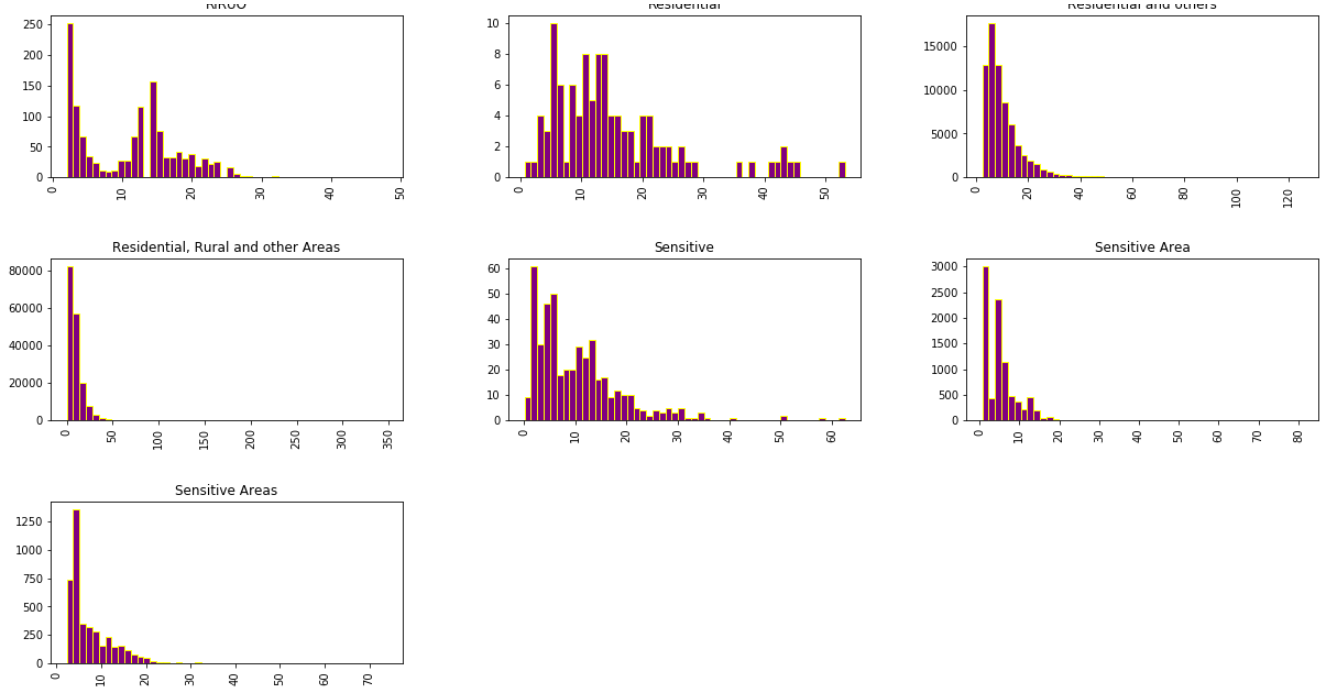
Описательная статистика + гипотезы

1. Попробуем отобразить среднее количество углекислого газа, зафиксированного в конкретных районах

In [17]:

```
ind_data.hist('so2', bins=50, by = ['type'], figsize = [20, 15], facecolor='purple', edgecolor = 'yellow');
```

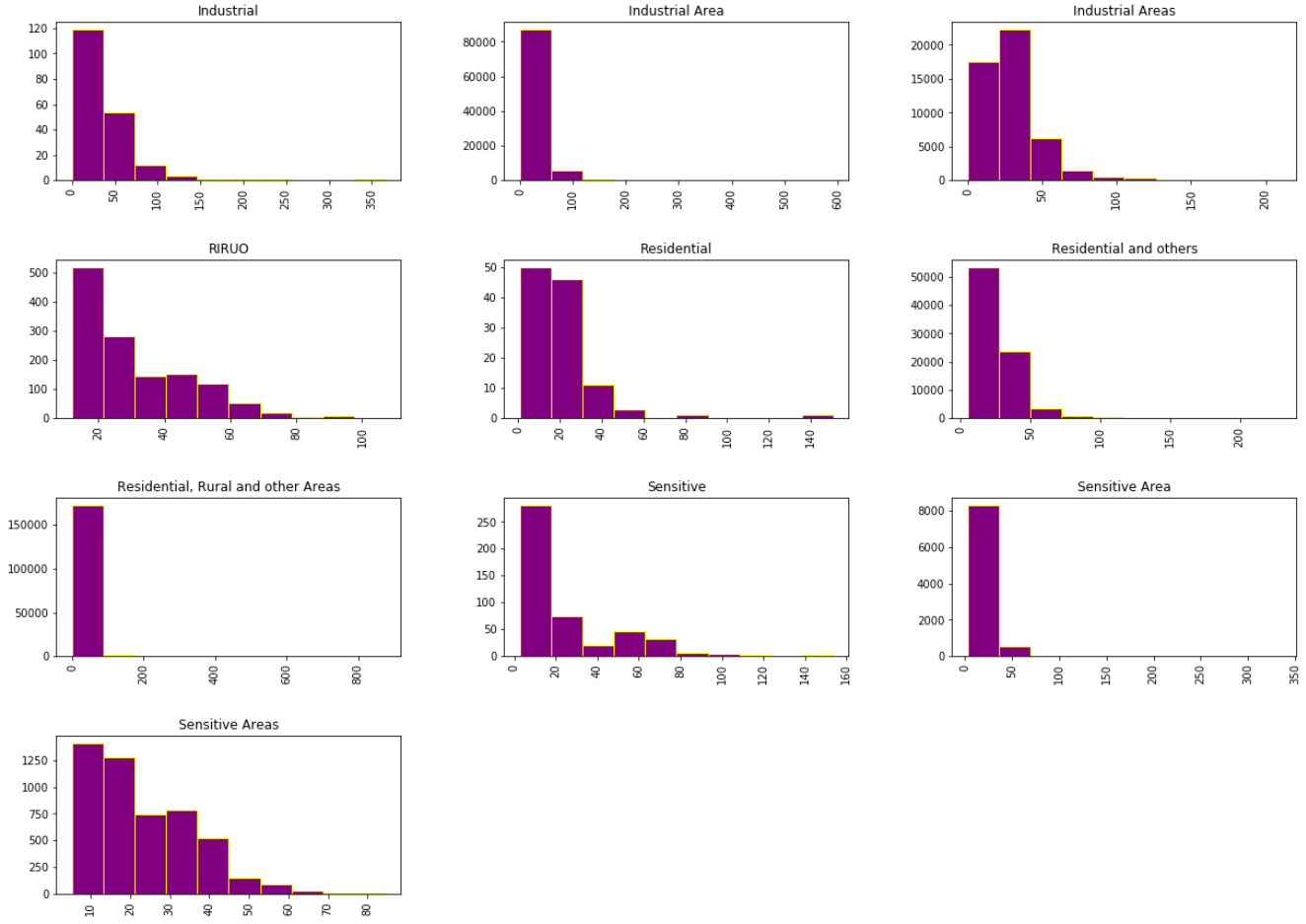




1. Попробуем сделать то же самое для оксида азота

In [6]:

```
ind_data.hist('no2', bins=10, by = ['type'], figsize = [20, 15], );
```



1. Попробуем отобразить все записи о газах в индустриальных районах, чтобы увидеть примерное распределение этих газов

In [33]:

```
so2 = ind_data.loc[ind_data.type == 'Industrial Area', 'so2']
no2 = ind_data.loc[ind_data.type == 'Industrial Area', 'no2']
kwargs = dict(alpha=0.5, bins=500)

plt.hist(so2,**kwargs, color='g', label='so2')
plt.hist(no2,**kwargs, color='b', label='no2')
plt.xlim(0, 160)
```

Out[33]:

```
(0, 160)
```