

```
df.columns
Index(['Year', 'Number of Accidents', 'Death', 'Number of Serious
Injuries',
      'Number of Minor Injuries', 'Number of Moderate Injuries',
      'Number of Severe Injuries'],
      dtype='object')
```

```
%%time
from gc import collect;
from warnings import filterwarnings;

import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
import seaborn as sns
plt.style.use("fivethirtyeight")

from datetime import datetime

from scipy import stats

filterwarnings('ignore');
from IPython.display import display_html, clear_output;
```

```
clear_output();
print();
collect();
```

```
CPU times: user 2.24 s, sys: 226 ms, total: 2.46 s
Wall time: 3.18 s
```

```
0
```

```
%%time
```

```
df.columns = df.columns.str.lower().str.replace(' ', '_')
```

```
print("\nNew column names:")
print(df.columns)
```

```
print();
collect();
```

```
New column names:
```

```
Index(['year', 'number_of_accidents', 'death',
      'number_of_serious_injuries',
      'number_of_minor_injuries', 'number_of_moderate_injuries',
      'number_of_severe_injuries'],
      dtype='object')
```

```
CPU times: user 78.9 ms, sys: 966 µs, total: 79.9 ms
Wall time: 79.6 ms
```

```
df.isna().sum()
```

```
year                0
number_of_accidents 0
death               0
number_of_serious_injuries 0
number_of_minor_injuries 0
number_of_moderate_injuries 0
number_of_severe_injuries 0
dtype: int64
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 44 entries, 0 to 43
```

```
Data columns (total 7 columns):
```

| # | Column | Non-Null Count | Dtype |
|---|-----------------------------|----------------|-------|
| 0 | year | 44 non-null | int64 |
| 1 | number_of_accidents | 44 non-null | int64 |
| 2 | death | 44 non-null | int64 |
| 3 | number_of_serious_injuries | 44 non-null | int64 |
| 4 | number_of_minor_injuries | 44 non-null | int64 |
| 5 | number_of_moderate_injuries | 44 non-null | int64 |
| 6 | number_of_severe_injuries | 44 non-null | int64 |

```
dtypes: int64(7)
```

```
memory usage: 2.5 KB
```

```
duplicate_values=df.duplicated().sum()
```

```
print(f'The data contains {duplicate_values} duplicate values')
```

```
The data contains 0 duplicate values
```

```
print(f'The dataset contains {df.shape[0]} rows and {df.shape[1]} columns')
```

```
The dataset contains 44 rows and 7 columns
```

```
df.describe().style.background_gradient(cmap='ocean')
<pandas.io.formats.style.Styler at 0x79535ad45840>

features = ['number_of_accidents', 'death',
            'number_of_serious_injuries',
            'number_of_minor_injuries', 'number_of_moderate_injuries',
            'number_of_severe_injuries']

fig, axs = plt.subplots(2, 3, figsize=(18, 14), facecolor='gray')

axs = axs.flatten()

subplot_color = 'gray'

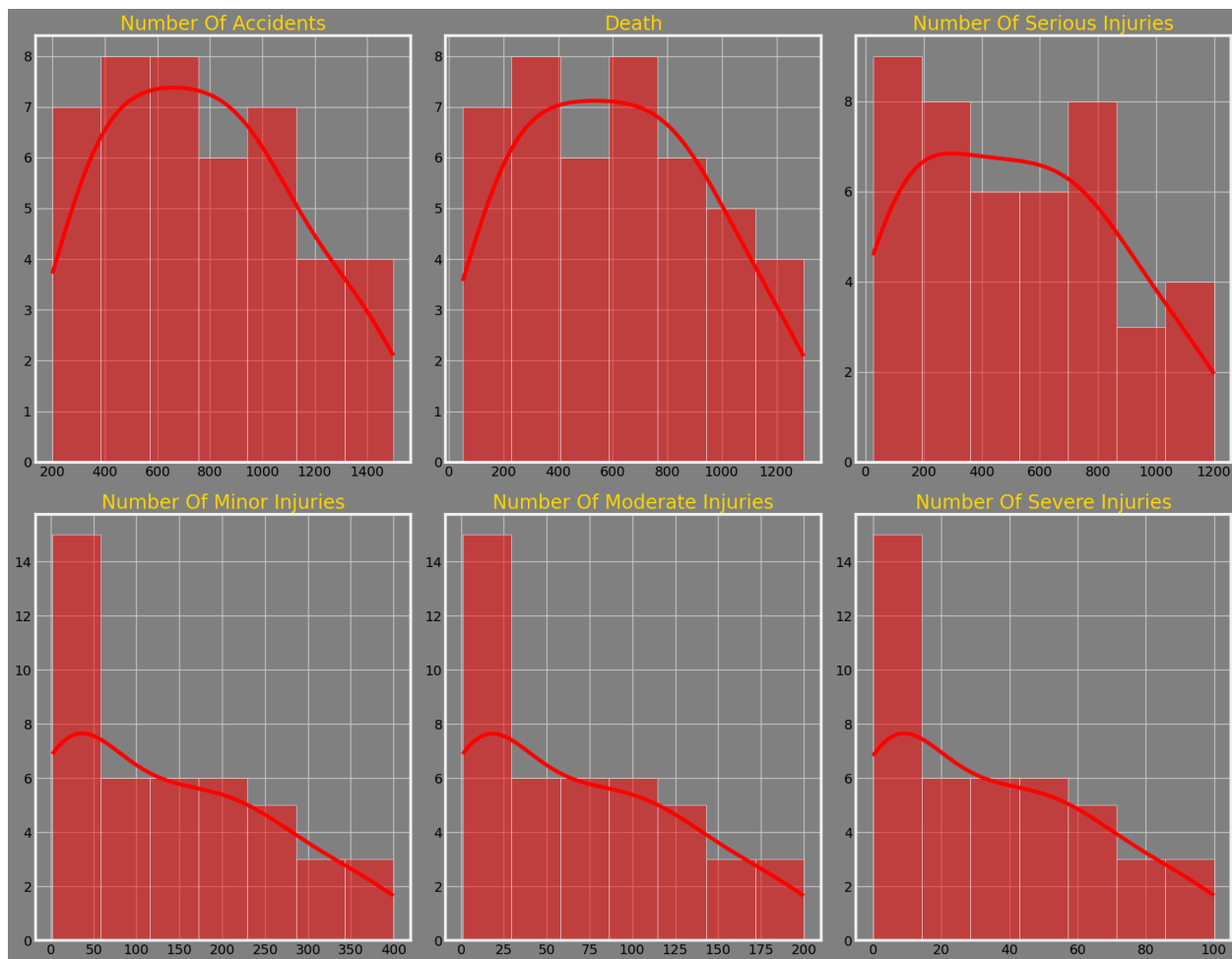
for i, col in enumerate(features):
    sns.histplot(data=df, x=col, color='red', kde=True, ax=axs[i])
    axs[i].set_facecolor(subplot_color)

    title = col.replace('_', ' ').title()
    axs[i].set_title(title, color='gold', fontname='Latin Modern Roman
Light')

    axs[i].set_xlabel('')
    axs[i].set_ylabel('')

plt.tight_layout()

plt.show()
```



```
plt.figure(figsize=(12, 8), facecolor='gray')
plt.gca().set_facecolor('gray')

features = ['number_of_accidents', 'death',
            'number_of_serious_injuries',
            'number_of_minor_injuries', 'number_of_moderate_injuries',
            'number_of_severe_injuries']

palette = sns.color_palette("husl", len(features))
for i, feature in enumerate(features):
    sns.scatterplot(data=df, x='year', y=feature,
                    label=feature.replace('_', ' ').title(),
                    color=palette[i])
plt.xlabel('Accidental Year')
plt.ylabel('Value')
plt.title('Trends of Various Features Over Time')
```

```
plt.xticks(rotation=45)
plt.legend(title='Feature', bbox_to_anchor=(1.05, 1), loc='upper
left', facecolor='gray')
plt.show()
```

