

## 04-az-scatter

2024-09-01

```
[ ]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
sns.set_style('whitegrid')
import dateutil

[ ]: df = pd.read_csv('../data/BGH-XI-helpers/mitverschulden-abb-2.csv', sep=';')
df['Datum'] = [dateutil.parser.parse(x, dayfirst=True) for x in df['Datum']]
df['Anlage_Boolean'] = df['Anlage'] == 'Anlage'
df = df.rename({'Az.': 'Aktenzeichen'}, axis=1)
df = df.rename({x:x.lower() for x in df.columns}, axis=1)
df['color'] = ['k' if x else 'grey' for x in df.anlage_boolean]
df['y'] = [(x.month-1) * 31 + x.day for x in df.datum]
df.head()

[ ]: plt.rcParams['figure.figsize'] = (9,6)
plt.rcParams['font.serif'] = 'BitstreamVeraSans Roman'
plt.rcParams['font.style'] = 'normal'
plt.rcParams['font.size'] = 6.5

[ ]: def plot_figure(df, save_path):
    minsize, maxsize = 5, 20
    maxmatches = max(df.treffer.values)
    fontsizes = np.linspace(minsize, maxsize, maxmatches)
    txt = df.aktenzeichen.values
    colors = df.color.values

    plt.scatter(df.datum, df.y, color=colors,
                marker='', s=df.treffer)
    plt.yticks(range(0, 31*12+1, 31),
                labels=['Jan', 'Feb', 'Mär', 'Apr', 'Mai', 'Jun', 'Jul', 'Aug', '
↳ Sep', 'Okt',
                                'Nov', 'Dez', ''])

    plt.ylim(0, 31*12)
    plt.xlim(dateutil.parser.parse('1988-01-01'), dateutil.parser.
↳ parse('2020-01-01'))
```

```

plt.xticks([dateutil.parser.parse(f'{x}-01-01') for x in range(1988,2021)],  

↳labels=range(1988,2021))
plt.xlabel('Jahr')
plt.ylabel('Monat und Tag')
for i, text in enumerate(txt):
    plt.text(df.datum[i], df.y[i], text,  

            fontdict=dict(fontsize=fontsizes[df.  

↳treffer[i]-1],color=colors[i],  

                        verticalalignment='baseline',  

↳horizontalalignment='left'))
plt.tight_layout()
plt.savefig(save_path)

```

```

[ ]: df_no_xizr = pd.DataFrame(df)
df_no_xizr['aktenzeichen'] = [x.split(' ')[-1] for x in df['aktenzeichen']]

```

```

[ ]: for ending in ['svg', 'png', 'eps', 'ps', 'pdf']:
    plot_figure(df, f'../figures/abbildung-2-no-xizr.{ending}')

```

```

[ ]:

```