Exploring numerical variables

Boxplots

In [1]:

```
# Setup
%matplotlib inline
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

sns.set_style("white")
# Custom colors
blue = "#3F83F4"
blue_dark = "#062089"
blue_light = "#8DC0F6"
blue_lighter = "#BBE4FA"
grey = "#9C9C9C"
grey_dark = "#777777"
grey_light = "#B2B2B2"
orange = "#EF8733"
colors_blue = [blue, blue_light]
```

Import data

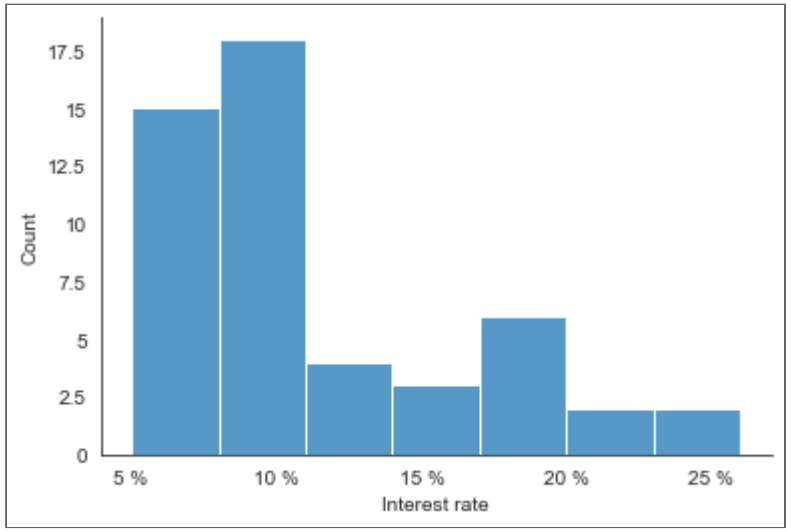
In [2]:

```
ROOT = "https://raw.githubusercontent.com/kirenz/modern-statistics/main/data/"
DATA = "loan50.csv"

df = pd.read_csv(ROOT + DATA)
df["interest_rate"] = df["interest_rate"].astype("int64")
```

Histogram

In [3]:

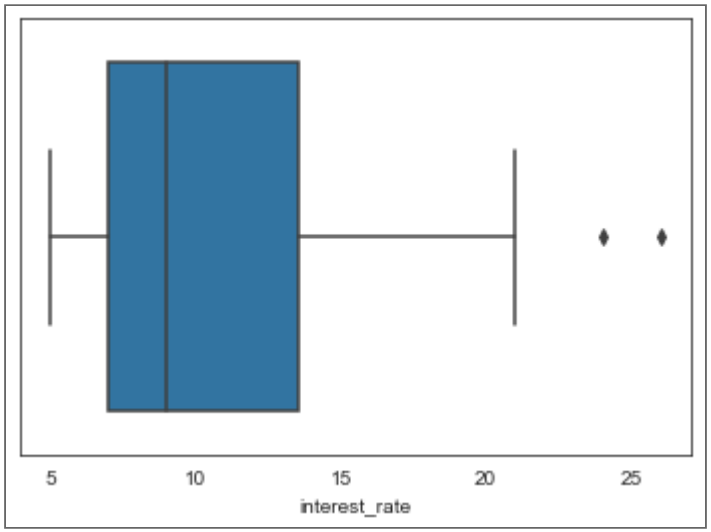
```
import matplotlib.ticker as ticker

fig, ax = plt.subplots()

sns.histplot(data=df,  x= "interest_rate", palette=colors_blue)
ax.xaxis.set_major_formatter(ticker.EngFormatter('%'))
ax.yaxis.set_major_formatter(ticker.EngFormatter())
plt.xlabel("Interest rate")
plt.ylabel("Count")
sns.despine()

plt.show();
```



Boxplot

In [4]:

```
ax = sns.boxplot(data=df, x="interest_rate")
```



Boxplot and Histogram

In [5]:

```
# creating a figure composed of two matplotlib.Axes objects (ax_box and ax_hist)
f, (ax_box, ax_hist) = plt.subplots(2, sharex=True, gridspec_kw={"height_ratios": (.15, .85)})

# assigning a graph to each ax
sns.boxplot(data=df, x="interest_rate", ax=ax_box, palette=colors_blue)
sns.histplot(data=df, x="interest_rate", ax=ax_hist, palette=colors_blue)

ax_box.set(xlabel='')
sns.despine(top=True, right=True, left=True, bottom=True)

plt.show();
```