Exploring numerical variables

Robust statistics

Import data

In [1]:

```
import pandas as pd

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")
```

In [2]:

```
%matplotlib inline
import seaborn as sns
import matplotlib.pyplot as plt

# 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]
```

Dot plot

A dot plot of interest rate for the loan50 dataset. The rates have been rounded and the distribution's mean is shown as a red triangle.

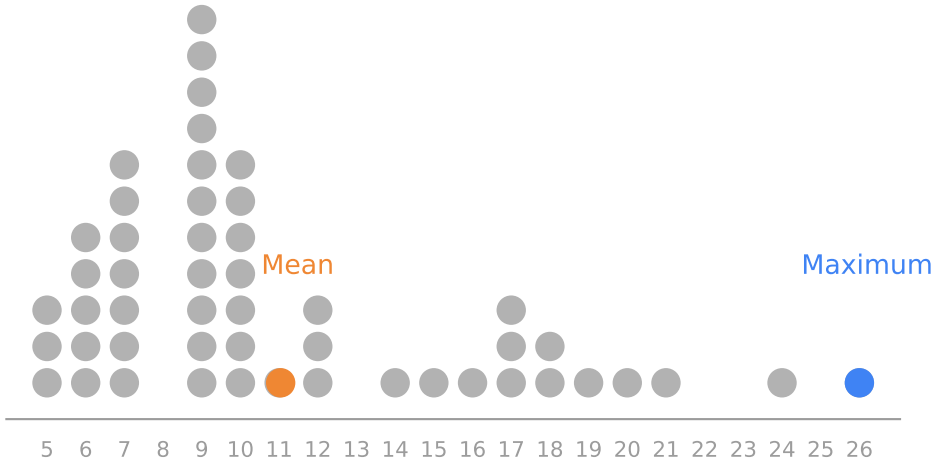
In [3]:

```
import numpy as np

# Prepara data
maximum = df["interest_rate"].max()
mean = df["interest_rate"].mean()
interest_rate = df["interest_rate"]
values, counts = np.unique(interest_rate, return_counts=True)

# Create dot plot
fig, ax = plt.subplots(figsize=(6, 3))
for value, count in zip(values, counts):
    ax.plot([value]*count, list(range(count)), 'o', ms=10, linestyle='', color=grey_light)
for spine in ['top', 'right', 'left']:
    ax.spines[spine].set_visible(False)
# Define optics
ax.yaxis.set_visible(False)
ax.set_ylim(-1, max(counts))
ax.spines['bottom'].set_color(grey)
ax.set_xticks(range(min(values), max(values)+1))
ax.tick_params(axis='x', length=0, pad=8, labelsize=8, colors=grey)
# Anotation
ax.plot(maximum, 0, 'o', ms=10, color=blue)
ax.annotate(text='Maximum', xy=(maximum-1.5, 3), color=blue)
ax.plot(mean, 0, 'o', ms=10, color=orange)
ax.annotate(text='Mean', xy=(mean-0.5, 3), color=orange)

plt.show();
```


Mean
Maximum
5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26