Exploring categorical variables

Pie charts

Import data

In [1]:

```
import pandas as pd

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

df = pd.read_csv(ROOT + DATA)

df.head()
```

Out[1]:

	emp_title	emp_length	state	homeownership	annual_income	verified_income	debt_to_income	annual_income_joint	verification_income_joint	debt_to_income_joint	...	sub_grade	issue_month	loan_stat
0	global config engineer	3.0	NJ	mortgage	90000.0	Verified	18.01	NaN	NaN	NaN	...	C3	Mar-2018	Curren
1	warehouse office clerk	10.0	HI	rent	40000.0	Not Verified	5.04	NaN	NaN	NaN	...	C1	Feb-2018	Curren
2	assembly	3.0	WI	rent	40000.0	Source Verified	21.15	NaN	NaN	NaN	...	D1	Feb-2018	Curren
3	customer service	1.0	PA	rent	30000.0	Not Verified	10.16	NaN	NaN	NaN	...	A3	Jan-2018	Curren
4	security supervisor	10.0	CA	rent	35000.0	Verified	57.96	57000.0	Verified	37.66	...	C3	Mar-2018	Curren

5 rows × 55 columns

Bar plots with two variables

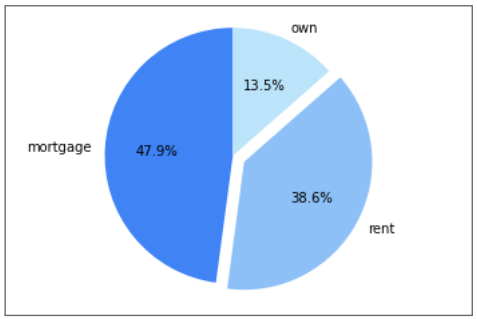
In [3]:

```
%matplotlib inline
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_light, blue_lighter, grey_dark, grey, grey_light]
```

Pie chart

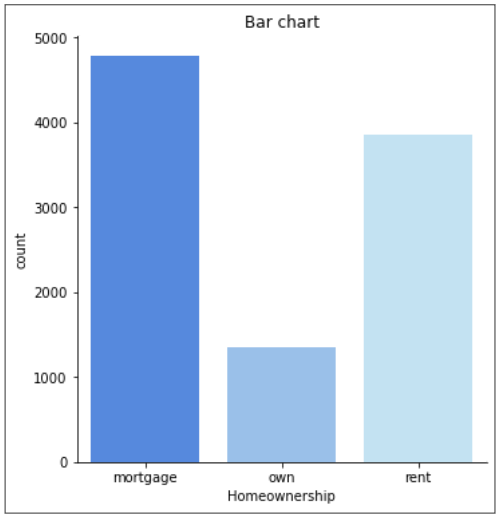
In [4]:

```
# Pie chart, where the slices will be ordered and plotted counter-clockwise:
# Data
homeownership = df.homeownership.value_counts()
labels = list(df['homeownership'].value_counts().index)
# Offsetting a slice with "explode"
explode = (0, 0.1, 0)  # only "explode" the 2nd slice
# Define plot
fig, ax = plt.subplots()
# Draw pie chart
ax.pie(homeownership, explode=explode, labels=labels, autopct='%1.1f%%',
        shadow=False, startangle=90, colors=colors)
# Equal aspect ratio ensures that pie is drawn as a circle.
ax.axis('equal')
plt.show()
```



In [5]:

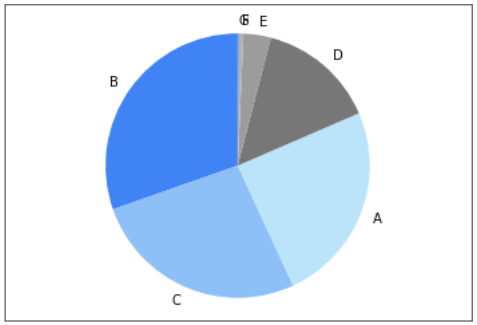
```
import seaborn as sns

# Counts of values of the homeownership variable.
sns.catplot(x="homeownership", kind = "count", palette=colors, data=df)

plt.title("Bar chart")
plt.xlabel("Homeownership")

plt.show();
```



Pie chart with too many levels

In [6]:

```
## Pie chart of loan grades
# Data
grade = df.grade.value_counts()
labels = list(df['grade'].value_counts().index)
# Define plot
fig, ax = plt.subplots()
# Draw pie chart
ax.pie(grade, labels=labels,shadow=False, startangle=90, colors=colors)
# Equal aspect ratio ensures that pie is drawn as a circle.
ax.axis('equal')
plt.show()
```



Bar plot

In [7]:

```
# Counts of values of the homeownership variable.
sns.catplot(x="grade", kind = "count", palette=colors, data=df)

plt.title("Bar chart")
plt.xlabel("Loan grade")

plt.show();
```