Introduction

Now you are ready to get a deeper understanding of your data.

Run the following cell to load your data and some utility functions (including code to check your answers).

In [ ]:
```
import sys
from pathlib import Path
learntools_dir = Path().absolute().parents[1]
sys.path.append(str(learntools_dir))

import pandas as pd
pd.set_option("display.max_rows", 5)
reviews = pd.read_csv("../../pandas/datasets/winemag-data-130k-v2.csv", inde

from learntools.core import binder; binder.bind(globals())
from learntools.pandas.summary_functions_and_maps import *
print("Setup complete.")

reviews.head()
```

Exercises

1.

What is the median of the `points` column in the `reviews` DataFrame?

In [ ]:
```
median_points = ____

# Check your answer
q1.check()
```

In [ ]:
```
# q1.hint()

# q1.solution()
```

2.

What countries are represented in the dataset? (Your answer should not include any duplicates.)

In [ ]:
```
countries = ____
```

```
# Check your answer
q2.check()
```

In [ ]:
```
# q2.hint()

# q2.solution()
```

3.

How often does each country appear in the dataset? Create a Series `reviews_per_country` mapping countries to the count of reviews of wines from that country.

In [ ]:
```
reviews_per_country = ____

# Check your answer
q3.check()
```

In [ ]:
```
# q3.hint()

# q3.solution()
```

4.

Create variable `centered_price` containing a version of the `price` column with the mean price subtracted.

(Note: this 'centering' transformation is a common preprocessing step before applying various machine learning algorithms.)

In [ ]:
```
centered_price = ____

# Check your answer
q4.check()
```

In [ ]:
```
# q4.hint()

# q4.solution()
```

5.

I'm an economical wine buyer. Which wine is the "best bargain"? Create a variable `bargain_wine` with the title of the wine with the highest points-to-price ratio in the dataset.

In [ ]:
```
bargain_wine = ____
```

```
# Check your answer
q5.check()
```

In [ ]:
```
# q5.hint()

# q5.solution()
```

6.

There are only so many words you can use when describing a bottle of wine. Is a wine more likely to be "tropical" or "fruity"? Create a Series `descriptor_counts` counting how many times each of these two words appears in the `description` column in the dataset. (For simplicity, let's ignore the capitalized versions of these words.)

In [ ]:
```
descriptor_counts = ____

# Check your answer
q6.check()
```

In [ ]:
```
# q6.hint()

# q6.solution()
```

7.

We'd like to host these wine reviews on our website, but a rating system ranging from 80 to 100 points is too hard to understand - we'd like to translate them into simple star ratings. A score of 95 or higher counts as 3 stars, a score of at least 85 but less than 95 is 2 stars. Any other score is 1 star.

Also, the Canadian Vintners Association bought a lot of ads on the site, so any wines from Canada should automatically get 3 stars, regardless of points.

Create a series `star_ratings` with the number of stars corresponding to each review in the dataset.

In [ ]:
```
star_ratings = ____

# Check your answer
q7.check()
```

In [ ]:
```
# q7.hint()

# q7.solution()
```

Keep going

Continue to **grouping and sorting**.