Python Lab 7 Solutions

Question 1. *Many possible answers:*

- Pandas not limited to 1 million rows (Microsoft Excel).
- Pandas code is repeatable (automate the data processing).
- Pipe data queries/analysis into plots.
- Import data from many different sources.

Further reading:
https://www.cbtnuggets.com/blog/certifications/microsoft/why-pandas-is-a-better-data-analysis-tool-than-excel

Question 2.

(a) Many possibilities (any one of these would do)

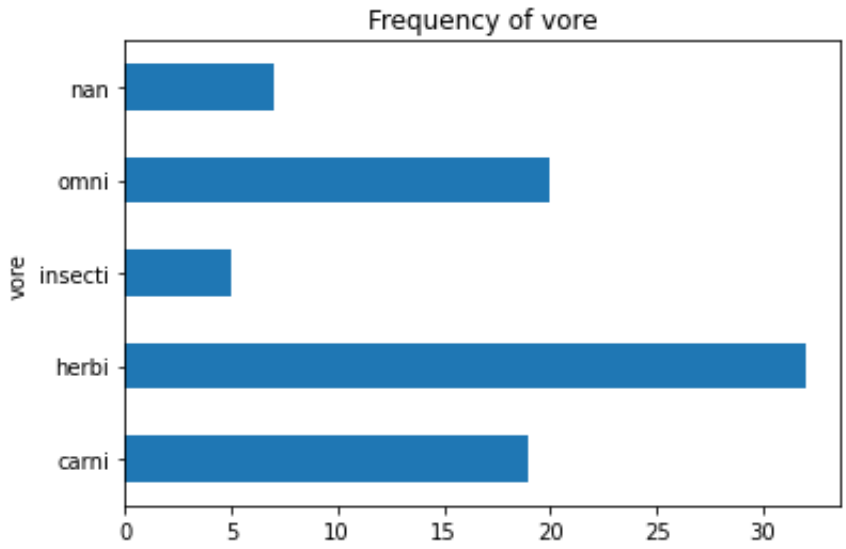
```
msleep.shape
(len(msleep),len(msleep.columns))
msleep.info
```

(b) Any two of these:

```
msleep.order.nunique()
msleep.order.describe()
len(msleep.order.unique())
```

(c) *Carnivore etc*

```
table = msleep.vore.value_counts(dropna=False)
# or ... table = msleep.groupby('vore',dropna=False).name.count()
import matplotlib.pyplot as plt
table.plot.barh()
plt.title('Frequency of vore')
```



(d) *Rodents*

```
rodents = msleep.query("order=='Rodentia'")
rodents.boxplot(column=['sleep_cycle','sleep_rem'])
# optional ... plt.title('Distribution of sleep_cycle and sleep_rem')
rodents.sort_values('brainwt',ascending=False)
```