

pandas_data_description_workbook_answers

September 22, 2020

1 PANDAS DATA DESCRIPTION WORKBOOK

Have a go at the following questions to practice your new found skills.

If you have any questions, go back to the course videos and have another look. One version of the answers is available in the next download. Remember, there are different ways to write code to get the same answer, so your answer can be correct and different to the answer example!

If you feel stuck and want some in person help, then have a look at the events page to join in a workshop <https://swamphen.co.uk/events>.

```
In [ ]: # import pandas
import pandas as pd
```

```
In [ ]: # read data in directly from web link
# https://vincentarelbundock.github.io/Rdatasets/csv/boot/nuclear.csv
# information on data set found at
# https://vincentarelbundock.github.io/Rdatasets/doc/boot/nuclear.html
nuclear = pd.read_csv('https://vincentarelbundock.github.io/Rdatasets/csv/boot/nuclear
```

```
In [ ]: # check you have read the data in ok by printing the head
nuclear.head()
```

```
In [ ]: # print out the last 3 rows
nuclear.tail(3)
```

```
In [ ]: # what is the shape of the pandas data frame?
nuclear.shape
```

```
In [ ]: # do a quick look stats on the data with describe
nuclear.describe()
```

```
In [ ]: # is there any missing data?

# no
```

```
In [ ]: # calculate the mean of the data frame
nuclear.mean()
```

```
In [ ]: # describe the column with the highest mean
nuclear['cap'].describe()
```

```
In [ ]: # calculate the median of the data frame
        nuclear.median()

In [ ]: # group by pt
        by_pt = nuclear.groupby('pt')

In [ ]: # describe the data set grouped by pt
        by_pt.describe()

In [ ]: # was this a sensible thing to do to understand the data set?

        # yes values are quite different between the groups and the full average
```