

pandas_data_description_workbook

September 22, 2020

1 PANDAS DATA DESCRIPTION WORKBOOK

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
```

```
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
```

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

```
In [ ]: # print out the last 3 rows
```

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

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

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

```
In [ ]: # calculate the mean of the data frame
```

```
In [ ]: # describe the column with the highest mean
```

```
In [ ]: # calculate the median of the data frame
```

```
In [ ]: # group by pt
```

```
In [ ]: # describe the data set grouped by pt
```

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