Untitled10

June 25, 2020

```
In [1]: import numpy as np
        import pandas as pd
```

```
In [2]: pd.set_option('display.max_columns', 100)
```

```
In [4]: df = pd.read_csv("nhanes_2015_2016.csv")
```

```
In [5]: df.head()
```

```
Out[5]:     SEQN  ALQ101  ALQ110  ALQ130  SMQ020  RIAGENDR  RIDAGEYR  RIDRETH1  \
        0  83732     1.0     NaN     1.0       1         1        62         3
        1  83733     1.0     NaN     6.0       1         1        53         3
        2  83734     1.0     NaN     NaN       1         1        78         3
        3  83735     2.0     1.0     1.0       2         2        56         3
        4  83736     2.0     1.0     1.0       2         2        42         4

           DMDCITZN  DMDEDUC2  DMDMARTL  DMDHHSIZ   WTINT2YR  SDMVPSU  SDMVSTRA  \
        0       1.0       5.0       1.0         2  134671.37        1       125
        1       2.0       3.0       3.0         1   24328.56        1       125
        2       1.0       3.0       1.0         2   12400.01        1       131
        3       1.0       5.0       6.0         1  102718.00        1       131
        4       1.0       4.0       3.0         5   17627.67        2       126

           INDFMPIR  BPXSY1  BPXDI1  BPXSY2  BPXDI2   BMXWT   BMXHT  BMXBMI  BMXLEG  \
        0      4.39   128.0    70.0   124.0    64.0    94.8   184.5    27.8    43.3
        1      1.32   146.0    88.0   140.0    88.0    90.4   171.4    30.8    38.0
        2      1.51   138.0    46.0   132.0    44.0    83.4   170.1    28.8    35.6
        3      5.00   132.0    72.0   134.0    68.0   109.8   160.9    42.4    38.5
        4      1.23   100.0    70.0   114.0    54.0    55.2   164.9    20.3    37.4

           BMXARML  BMXARMC  BMXWAIST  HIQ210
        0     43.6     35.9     101.1     2.0
        1     40.0     33.2     107.9     NaN
        2     37.0     31.0     116.5     2.0
        3     37.7     38.3     110.1     2.0
        4     36.0     27.2      80.4     2.0
```

```
In [6]: df.columns
```

```
Out[6]: Index(['SEQN', 'ALQ101', 'ALQ110', 'ALQ130', 'SMQ020', 'RIAGENDR', 'RIDAGEYR',
               'RIDRETH1', 'DMDCITZN', 'DMDEDUC2', 'DMDMARTL', 'DMDHHSIZ', 'WTINT2YR',
               'SDMVPSU', 'SDMVSTRA', 'INDFMPIR', 'BPXSY1', 'BPXDI1', 'BPXSY2',
               'BPXDI2', 'BMXWT', 'BMXHT', 'BMXBMI', 'BMXLEG', 'BMXARML', 'BMXARMC',
               'BMXWAIST', 'HIQ210'],
              dtype='object')

In [7]: keep = ['BMXWT', 'BMXHT', 'BMXBMI', 'BMXLEG', 'BMXARML', 'BMXARMC',
               'BMXWAIST']

In [9]: [column for column in df.columns if 'BM' in column]

Out[9]: ['BMXWT', 'BMXHT', 'BMXBMI', 'BMXLEG', 'BMXARML', 'BMXARMC', 'BMXWAIST']

In [12]: df[keep].head(3)

Out[12]:    BMXWT  BMXHT  BMXBMI  BMXLEG  BMXARML  BMXARMC  BMXWAIST
         0   94.8  184.5    27.8    43.3     43.6     35.9     101.1
         1   90.4  171.4    30.8    38.0     40.0     33.2     107.9
         2   83.4  170.1    28.8    35.6     37.0     31.0     116.5

In [13]: df.loc[:, keep].head(3)

Out[13]:    BMXWT  BMXHT  BMXBMI  BMXLEG  BMXARML  BMXARMC  BMXWAIST
         0   94.8  184.5    27.8    43.3     43.6     35.9     101.1
         1   90.4  171.4    30.8    38.0     40.0     33.2     107.9
         2   83.4  170.1    28.8    35.6     37.0     31.0     116.5

In [18]: index_bool = np.isin(df.columns, keep)

In [20]: df.iloc[:, index_bool].head(3)

Out[20]:    BMXWT  BMXHT  BMXBMI  BMXLEG  BMXARML  BMXARMC  BMXWAIST
         0   94.8  184.5    27.8    43.3     43.6     35.9     101.1
         1   90.4  171.4    30.8    38.0     40.0     33.2     107.9
         2   83.4  170.1    28.8    35.6     37.0     31.0     116.5

In [22]: #waist larger than the median
         waist_median = pd.Series.median(df['BMXWAIST'])

In [23]: waist_median

Out[23]: 98.3

In [29]: condition1 = df["BMXWAIST"] > waist_median
         condition2 = df["BMXLEG"] < 32
         df[condition1 & condition2].head()
```

```
Out[29]:      SEQN  ALQ101  ALQ110  ALQ130  SMQ020  RIAGENDR  RIDAGEYR  RIDRETH1  \
        15  83757     1.0     NaN     1.0       2         2        57         2
        27  83785     2.0     1.0     1.0       1         2        60         2
        39  83812     1.0     NaN     2.0       2         2        68         1
        52  83832     2.0     1.0     4.0       2         2        50         1
        55  83837     2.0     2.0     NaN       2         2        45         1

            DMDCITZN  DMDEDUC2  DMDMARTL  DMDHHSIZ  WTINT2YR  SDMVPSU  SDMVSTRA  \
        15       1.0       1.0       4.0         5  11709.11        1       120
        27       1.0       5.0       3.0         4  10495.87        1       128
        39       1.0       3.0       1.0         3  10255.97        1       124
        52       2.0       1.0       4.0         5  11709.11        1       121
        55       1.0       2.0       1.0         7  15415.16        1       133

            INDFMPIR  BPXSY1  BPXDI1  BPXSY2  BPXDI2   BMXWT   BMXHT  BMXBMI  BMXLEG  \
        15      0.77   134.0    68.0   146.0    62.0    80.5   150.8    35.4    31.6
        27      5.00   142.0    74.0   136.0    74.0    75.6   145.2    35.9    31.0
        39      5.00   124.0    56.0   114.0    66.0    63.7   147.9    29.1    26.0
        52      1.41   104.0    76.0     NaN     NaN   105.9   157.7    42.6    29.2
        55      2.18   112.0    68.0   114.0    68.0    77.5   148.3    35.2    30.5

            BMXARML  BMXARMC  BMXWAIST  HIQ210
        15     32.7     33.7     113.5     2.0
        27     33.1     36.0     108.0     2.0
        39     34.0     31.5     110.0     2.0
        52     35.0     40.7     129.1     NaN
        55     34.0     34.4     107.6     2.0

In [30]: df.loc[condition1 & condition2, :].head()

Out[30]:      SEQN  ALQ101  ALQ110  ALQ130  SMQ020  RIAGENDR  RIDAGEYR  RIDRETH1  \
        15  83757     1.0     NaN     1.0       2         2        57         2
        27  83785     2.0     1.0     1.0       1         2        60         2
        39  83812     1.0     NaN     2.0       2         2        68         1
        52  83832     2.0     1.0     4.0       2         2        50         1
        55  83837     2.0     2.0     NaN       2         2        45         1

            DMDCITZN  DMDEDUC2  DMDMARTL  DMDHHSIZ  WTINT2YR  SDMVPSU  SDMVSTRA  \
        15       1.0       1.0       4.0         5  11709.11        1       120
        27       1.0       5.0       3.0         4  10495.87        1       128
        39       1.0       3.0       1.0         3  10255.97        1       124
        52       2.0       1.0       4.0         5  11709.11        1       121
        55       1.0       2.0       1.0         7  15415.16        1       133

            INDFMPIR  BPXSY1  BPXDI1  BPXSY2  BPXDI2   BMXWT   BMXHT  BMXBMI  BMXLEG  \
        15      0.77   134.0    68.0   146.0    62.0    80.5   150.8    35.4    31.6
        27      5.00   142.0    74.0   136.0    74.0    75.6   145.2    35.9    31.0
        39      5.00   124.0    56.0   114.0    66.0    63.7   147.9    29.1    26.0
```

```
52      1.41    104.0    76.0      NaN      NaN   105.9   157.7     42.6     29.2
55      2.18    112.0    68.0    114.0     68.0    77.5   148.3     35.2     30.5


        BMXARML  BMXARMC  BMXWAIST  HIQ210
15         32.7     33.7     113.5     2.0
27         33.1     36.0     108.0     2.0
39         34.0     31.5     110.0     2.0
52         35.0     40.7     129.1     NaN
55         34.0     34.4     107.6     2.0
```

In [31]: df.small = df.head(5)
         df.small

Out[31]:   SEQN ALQ101 ALQ110 ALQ130 SMQ020 RIAGENDR RIDAGEYR RIDRETH1 \
         0 83732   1.0   NaN   1.0   1   1   62   3
         1 83733   1.0   NaN   6.0   1   1   53   3
         2 83734   1.0   NaN   NaN   1   1   78   3
         3 83735   2.0   1.0   1.0   2   2   56   3
         4 83736   2.0   1.0   1.0   2   2   42   4

            DMDCITZN DMDEDUC2 DMDMARTL DMDHHSIZ   WTINT2YR SDMVPSU SDMVSTRA \
         0   1.0   5.0   1.0   2   134671.37   1   125
         1   2.0   3.0   3.0   1   24328.56   1   125
         2   1.0   3.0   1.0   2   12400.01   1   131
         3   1.0   5.0   6.0   1   102718.00   1   131
         4   1.0   4.0   3.0   5   17627.67   2   126

            INDFMPIR BPXSY1 BPXDI1 BPXSY2 BPXDI2 BMXWT BMXHT BMXBMI BMXLEG \
         0   4.39   128.0   70.0   124.0   64.0   94.8 184.5   27.8   43.3
         1   1.32   146.0   88.0   140.0   88.0   90.4 171.4   30.8   38.0
         2   1.51   138.0   46.0   132.0   44.0   83.4 170.1   28.8   35.6
         3   5.00   132.0   72.0   134.0   68.0 109.8 160.9   42.4   38.5
         4   1.23   100.0   70.0   114.0   54.0   55.2 164.9   20.3   37.4

            BMXARML BMXARMC BMXWAIST HIQ210
         0   43.6   35.9   101.1   2.0
         1   40.0   33.2   107.9   NaN
         2   37.0   31.0   116.5   2.0
         3   37.7   38.3   110.1   2.0
         4   36.0   27.2   80.4   2.0

In [34]: df.small.index = ['a', 'b', 'c', 'd', 'e']
         df.small.loc['a', :]

Out[34]: SEQN   83732.00
         ALQ101   1.00
         ALQ110   NaN
         ALQ130   1.00
         SMQ020   1.00
```
```

```
           RIAGENDR          1.00
           RIDAGEYR         62.00
           RIDRETH1          3.00
           DMDCITZN          1.00
           DMDEDUC2          5.00
           DMDMARTL          1.00
           DMDHHSIZ          2.00
           WTINT2YR     134671.37
           SDMVPSU           1.00
           SDMVSTRA        125.00
           INDFMPIR          4.39
           BPXSY1          128.00
           BPXDI1           70.00
           BPXSY2          124.00
           BPXDI2           64.00
           BMXWT            94.80
           BMXHT           184.50
           BMXBMI           27.80
           BMXLEG           43.30
           BMXARML          43.60
           BMXARMC          35.90
           BMXWAIST        101.10
           HIQ210            2.00
           Name: a, dtype: float64

In [38]: df.small.iloc[[1,2], :]

Out[38]:     SEQN  ALQ101  ALQ110  ALQ130  SMQ020  RIAGENDR  RIDAGEYR  RIDRETH1  \
         b  83733     1.0     NaN     6.0       1         1        53         3
         c  83734     1.0     NaN     NaN       1         1        78         3

            DMDCITZN  DMDEDUC2  DMDMARTL  DMDHHSIZ  WTINT2YR  SDMVPSU  SDMVSTRA  \
         b       2.0       3.0       3.0         1  24328.56        1       125
         c       1.0       3.0       1.0         2  12400.01        1       131

            INDFMPIR  BPXSY1  BPXDI1  BPXSY2  BPXDI2  BMXWT  BMXHT  BMXBMI  BMXLEG  \
         b      1.32   146.0    88.0   140.0    88.0   90.4  171.4    30.8    38.0
         c      1.51   138.0    46.0   132.0    44.0   83.4  170.1    28.8    35.6

            BMXARML  BMXARMC  BMXWAIST  HIQ210
         b     40.0     33.2     107.9     NaN
         c     37.0     31.0     116.5     2.0

In [39]: df.small.loc[:, 'BMXBMI']

Out[39]: a    27.8
         b    30.8
         c    28.8
         d    42.4
```

5

```
        e    20.3
        Name: BMXBMI, dtype: float64

In [40]: df.small.loc[:, 'BMXBMI'].values

Out[40]: array([27.8, 30.8, 28.8, 42.4, 20.3])

In [41]: df.small.iloc[:, 'BMXBMI']


        ---------------------------------------------------------------------


        ValueError                                Traceback (most recent call last)

        /opt/conda/lib/python3.6/site-packages/pandas/core/indexing.py in _has_valid_tuple(sel
            222            try:
    --> 223                self._validate_key(k, i)
            224            except ValueError:


        /opt/conda/lib/python3.6/site-packages/pandas/core/indexing.py in _validate_key(self,
           2083            raise ValueError("Can only index by location with "
        -> 2084                            "a [{types}]".format(types=self._valid_types))
           2085


        ValueError: Can only index by location with a [integer, integer slice (START point is


    During handling of the above exception, another exception occurred:


        ValueError                                Traceback (most recent call last)

        <ipython-input-41-b94b562ccb3a> in <module>()
    ----> 1 df.small.iloc[:, 'BMXBMI']


        /opt/conda/lib/python3.6/site-packages/pandas/core/indexing.py in __getitem__(self, key
           1492            except (KeyError, IndexError, AttributeError):
           1493                pass
        -> 1494            return self._getitem_tuple(key)
           1495        else:
           1496            # we by definition only have the 0th axis


        /opt/conda/lib/python3.6/site-packages/pandas/core/indexing.py in _getitem_tuple(self,
           2141    def _getitem_tuple(self, tup):
           2142
```

```
-> 2143              self._has_valid_tuple(tup)
   2144              try:
   2145                  return self._getitem_lowerdim(tup)


   /opt/conda/lib/python3.6/site-packages/pandas/core/indexing.py in _has_valid_tuple(sel
   225                      raise ValueError("Location based indexing can only have "
   226                                          "[{types}] types"
--> 227                                          .format(types=self._valid_types))
   228
   229      def _is_nested_tuple_indexer(self, tup):


   ValueError: Location based indexing can only have [integer, integer slice (START point
```

In [42]: df.small['BMXBMI']

Out[42]: a   27.8
         b   30.8
         c   28.8
         d   42.4
         e   20.3
         Name: BMXBMI, dtype: float64

In [44]: df.small['BMXBMI'] = range(5)
         df.small.BMXBMI

Out[44]: a   0
         b   1
         c   2
         d   3
         e   4
         Name: BMXBMI, dtype: int64

In [46]: df['BMXBMI'] = range(df.shape[0])
         df.small.BMXBMI

Out[46]: a   0
         b   1
         c   2
         d   3
         e   4
         Name: BMXBMI, dtype: int64