assignment 5

March 8, 2022

1 Filtering and univariate analysis on lifexpectancy dataset

[7]:
```
import pandas as pd
```

[8]:
```
lifeExpectancy = pd.read_csv('lifeExpectancyAtBirth.csv', sep=',')
```

[9]:
```
lifeExpectancy.head()
lifeExpectancy
```

[9]:
```
         Location  Period                         Indicator        Dim1  \
0     Afghanistan    2019  Life expectancy at birth (years)  Both sexes
1     Afghanistan    2019  Life expectancy at birth (years)        Male
2     Afghanistan    2019  Life expectancy at birth (years)      Female
3     Afghanistan    2015  Life expectancy at birth (years)  Both sexes
4     Afghanistan    2015  Life expectancy at birth (years)        Male
…             …     …                                  …           …
2192     Zimbabwe    2010  Life expectancy at birth (years)        Male
2193     Zimbabwe    2010  Life expectancy at birth (years)      Female
2194     Zimbabwe    2000  Life expectancy at birth (years)  Both sexes
2195     Zimbabwe    2000  Life expectancy at birth (years)        Male
2196     Zimbabwe    2000  Life expectancy at birth (years)      Female

      First Tooltip
0             63.21
1             63.29
2             63.16
3             61.65
4             61.04
…             …
2192          49.58
2193          53.21
2194          46.57
2195          45.15
2196          48.12

[2197 rows x 5 columns]
```

[10]:
```
filter = (lifeExpectancy['Dim1'] == 'Both sexes') & (lifeExpectancy['Period']␣
↪== 2019)
filteredDf = lifeExpectancy[filter]
filteredDf
```

[10]:

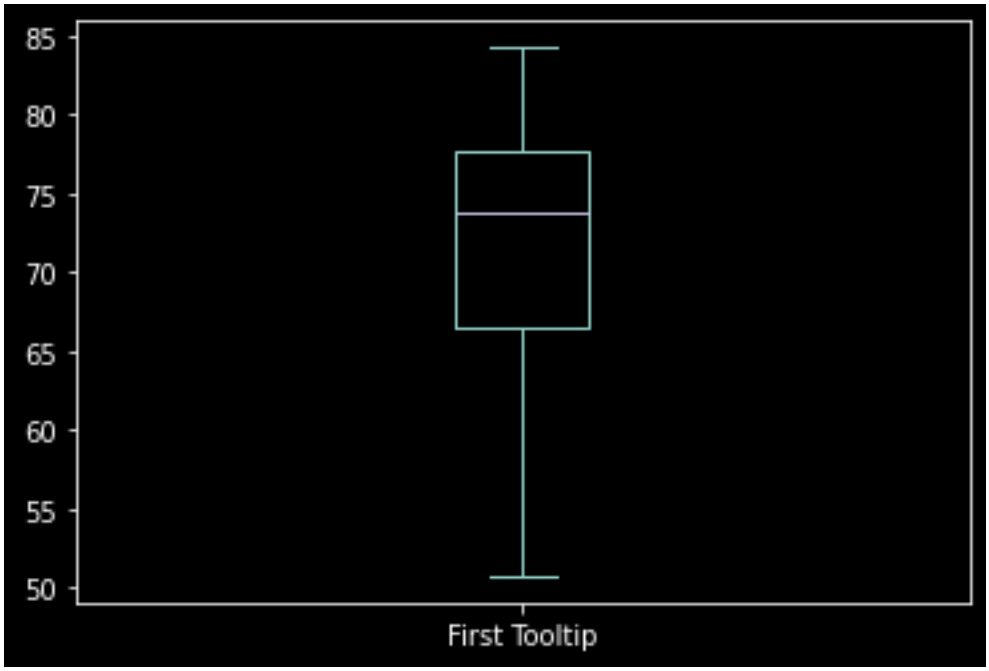
	Location	Period	\
0	Afghanistan	2019	
12	Albania	2019	
24	Algeria	2019	
36	Angola	2019	
48	Antigua and Barbuda	2019	
…	…	…	
2137	Venezuela (Bolivarian Republic of)	2019	
2149	Viet Nam	2019	
2161	Yemen	2019	
2173	Zambia	2019	
2185	Zimbabwe	2019	

	Indicator	Dim1	First Tooltip
0	Life expectancy at birth (years)	Both sexes	63.21
12	Life expectancy at birth (years)	Both sexes	78.00
24	Life expectancy at birth (years)	Both sexes	77.13
36	Life expectancy at birth (years)	Both sexes	63.06
48	Life expectancy at birth (years)	Both sexes	76.45
…	…	…	…
2137	Life expectancy at birth (years)	Both sexes	73.95
2149	Life expectancy at birth (years)	Both sexes	73.74
2161	Life expectancy at birth (years)	Both sexes	66.63
2173	Life expectancy at birth (years)	Both sexes	62.45
2185	Life expectancy at birth (years)	Both sexes	60.68

[183 rows x 5 columns]

[11]:
```
filteredDf['First Tooltip'].plot(kind='box')
```

[11]:
```
<AxesSubplot:>
```



1.1 The countries with the highest and lowest life expectancy

[17]:
```
filteredDf.sort_values(by='First Tooltip',ascending= False)
```

[17]:

	Location	Period	Indicator	\
997	Japan	2019	Life expectancy at birth (years)	
1897	Switzerland	2019	Life expectancy at birth (years)	
1573	Republic of Korea	2019	Life expectancy at birth (years)	
1837	Spain	2019	Life expectancy at birth (years)	
1753	Singapore	2019	Life expectancy at birth (years)	
…	…	…	…	
1333	Mozambique	2019	Life expectancy at birth (years)	
661	Eswatini	2019	Life expectancy at birth (years)	
1801	Somalia	2019	Life expectancy at birth (years)	
373	Central African Republic	2019	Life expectancy at birth (years)	
1117	Lesotho	2019	Life expectancy at birth (years)	

	Dim1	First Tooltip
997	Both sexes	84.26
1897	Both sexes	83.45
1573	Both sexes	83.30
1837	Both sexes	83.22
1753	Both sexes	83.22
…	…	…
1333	Both sexes	58.14

```
661   Both sexes          57.73
1801  Both sexes          56.47
373   Both sexes          53.10
1117  Both sexes          50.75

[183 rows x 5 columns]
```

[16]:
```
print(filteredDf.describe())
```

```
       Period  First Tooltip
count   183.0     183.000000
mean   2019.0      72.540492
std       0.0       7.129956
min    2019.0      50.750000
25%    2019.0      66.550000
50%    2019.0      73.740000
75%    2019.0      77.730000
max    2019.0      84.260000
```