In [59]:
```
'''
THIS PROJECT IS PART OF THE INTERVIEWING PROCESS FOR OPEN DATA FELLOW (KENYA
) IN 2016
THE DATA USED IN THE PROJECT COMES FROM THE life_expectancy_in_counties.csv
DATASET FOR 2016
The file is life_expectancy.py
'''
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style("whitegrid")
%pylab inline
```

```
Populating the interactive namespace from numpy and matplotlib
```

In [28]:
```
#importing the data
path = '/home/kev/Desktop/'
data = pd.read_csv(path+'Life_expectancy_in_counties.csv')
data['gender'] = data['gender'].astype('category')
data['county'] = data['county'].astype('category')
data.shape
```

Out[28]:
```
(192, 3)
```

In [52]:
```
data.head()
```

Out[52]:

	county	gender	life_expectancy
0	Lamu	males	67
1	Marsabit	males	72
2	Tana River	males	68
3	Mombasa	males	67
4	Kwale	males	65

In [27]:
```
#cleaning the data
clean_data = data[data.county != 'Kenya']
clean_data = data.replace("Murang<U+201F>a", "Murang'a")
```

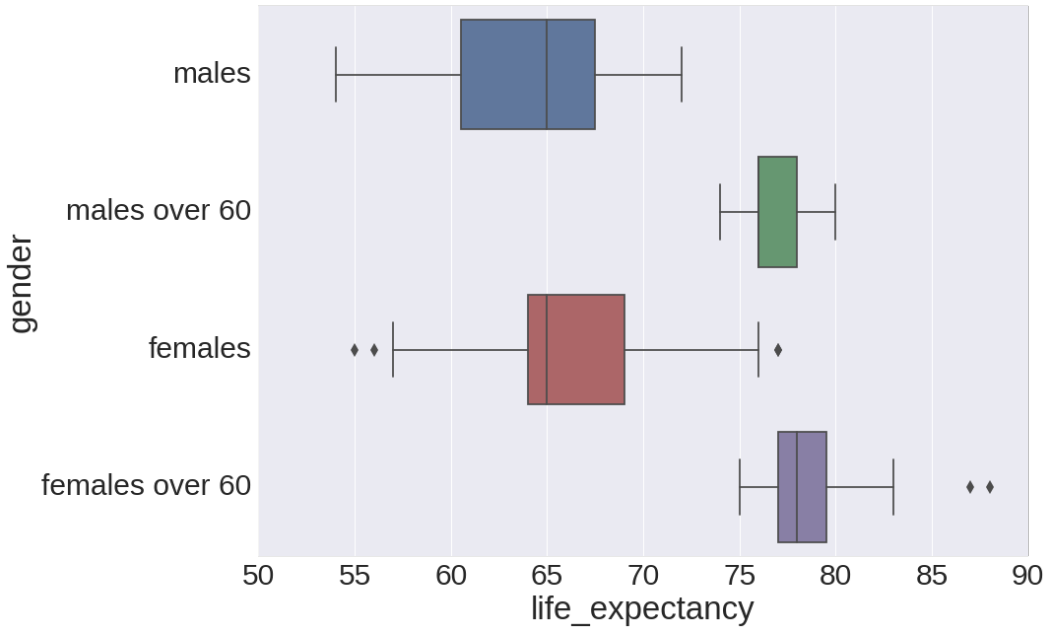
In [48]:
```
#Zooming into the gender categories
p = clean_data.groupby('gender').describe()
p
```

Out[48]:

		life_expectancy
gender		
females	**count**	47.000000
	mean	66.340426
	std	5.247263
	min	55.000000
	25%	64.000000
	50%	65.000000
	75%	69.000000
	max	77.000000
females over 60	**count**	47.000000
	mean	78.531915
	std	2.865470
	min	75.000000
	25%	77.000000
	50%	78.000000
	75%	79.500000
	max	88.000000
males	**count**	47.000000
	mean	64.106383
	std	4.429320
	min	54.000000
	25%	60.500000
	50%	65.000000
	75%	67.500000
	max	72.000000
males over 60	**count**	47.000000
	mean	76.702128
	std	1.627413
	min	74.000000
	25%	76.000000
	50%	76.000000
	75%	78.000000
	max	80.000000

In [45]:
```
plt.figure(figsize=(14,10))
sns.set(font_scale=3)
sns.boxplot("life_expectancy", "gender", data=clean_data,saturation=.6, flie
rsize=10.)
```

Out[45]: <matplotlib.axes._subplots.AxesSubplot at 0x7f8888fe5610>



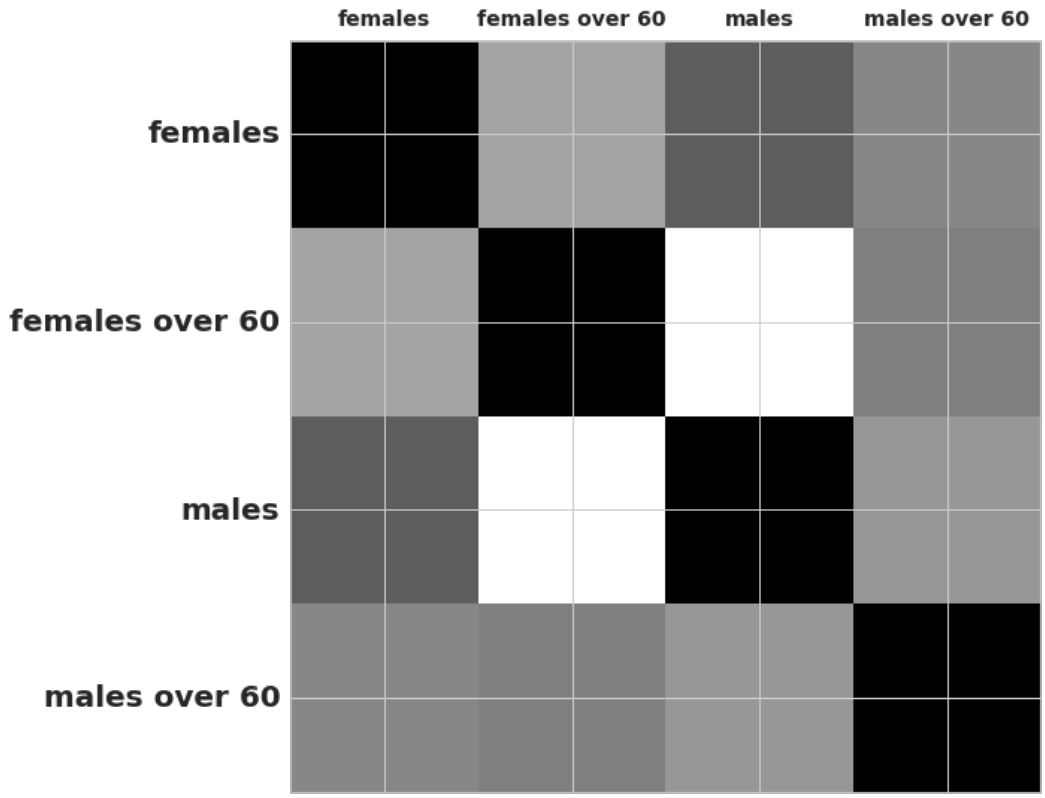
In [51]:
```
#To allow us to conduct further statistical analysis on the data, we need to
 reshape the data
piv_data = clean_data.pivot(index='county', columns='gender', values='life_e
xpectancy')
piv_data
```

Out[51]:

gender	females	females over 60	males	males over 60
county				
Baringo	64	78	60	76
Bomet	65	80	60	76
Bungoma	77	77	67	76
Busia	65	77	63	76
Elgeyo Marakwet	65	76	62	76
Embu	69	77	65	75
Garissa	77	83	68	79
Homa Bay	67	81	59	77
Isiolo	76	82	72	79
Kajiado	69	79	66	77
Kakamega	64	77	63	75
Kericho	65	83	60	76
Kiambu	68	78	69	77
Kilifi	68	78	64	76
Kirinyaga	71	81	69	78
Kisii	67	78	66	78
Kisumu	62	78	61	76
Kitui	74	80	72	79
Kwale	63	77	65	76
Laikipia	61	77	58	76
Lamu	61	75	67	77
Machakos	73	79	68	77
Makueni	73	79	71	78
Mandera	64	77	62	76
Marsabit	68	78	72	79
Meru	70	77	68	77
Migori	65	78	64	78
Mombasa	65	75	67	75
Murang'a	68	78	65	76
Nairobi	68	79	68	77
Nakuru	61	77	59	76
Nandi	61	75	60	75
Narok	72	82	68	80
Nyamira	65	76	65	76
Nyandarua	65	77	66	76
Nyeri	64	78	65	76
Samburu	71	88	62	80
Siaya	58	78	54	75

In [70]:
```
#Correlation table showing the interaction between the four features in the
data
corr = piv_data.corr()
fig, ax = plt.subplots(figsize=(14, 9))
plt.rc('xtick', labelsize=14)
plt.rc('ytick', labelsize=20)
ax.matshow(corr)
plt.xticks(range(len(corr.columns)), corr.columns)
plt.yticks(range(len(corr.columns)), corr.columns)
```

Out[70]:
```
([<matplotlib.axis.YTick at 0x7f8888c0dfd0>,
  <matplotlib.axis.YTick at 0x7f8888c0db50>,
  <matplotlib.axis.YTick at 0x7f8888783490>,
  <matplotlib.axis.YTick at 0x7f888877add0>],
 <a list of 4 Text yticklabel objects>)
```



In [67]:
```
corr
```

Out[67]:

gender	females	females over 60	males	males over 60
gender				
females	1.000000	0.568914	0.767262	0.646018
females over 60	0.568914	1.000000	0.209546	0.668712
males	0.767262	0.209546	1.000000	0.601627
males over 60	0.646018	0.668712	0.601627	1.000000

In [ ]: