

```
In [1]: 1 import numpy as np
        2 import pandas as pd
        3 %matplotlib inline
        4 import matplotlib.pyplot as plt
        5 import seaborn as sns
        6 from scipy import stats
```

```
In [2]: 1 sns.set(color_codes=True)
```

```
In [3]: 1 births_df = pd.read_csv('births.csv')
```

In [4]:

1 births\_df

Out[4]:

	Year	District Code	District Name	Neighborhood Code	Neighborhood Name	Gender	Number
0	2017	1	Ciutat Vella	1	el Raval	Boys	283
1	2017	1	Ciutat Vella	2	el Barri Gòtic	Boys	56
2	2017	1	Ciutat Vella	3	la Barceloneta	Boys	51
3	2017	1	Ciutat Vella	4	Sant Pere, Santa Caterina i la Ribera	Boys	90
4	2017	2	Eixample	5	el Fort Pienc	Boys	117
5	2017	2	Eixample	6	la Sagrada Família	Boys	207
6	2017	2	Eixample	7	la Dreta de l'Eixample	Boys	185
7	2017	2	Eixample	8	l'Antiga Esquerra de l'Eixample	Boys	176
8	2017	2	Eixample	9	la Nova Esquerra de l'Eixample	Boys	218
9	2017	2	Eixample	10	Sant Antoni	Boys	172
10	2017	3	Sants-Montjuïc	11	el Poble Sec	Boys	156
11	2017	3	Sants-Montjuïc	12	la Marina del Prat Vermell	Boys	4
12	2017	3	Sants-Montjuïc	13	la Marina de Port	Boys	119
13	2017	3	Sants-Montjuïc	14	la Font de la Guatlla	Boys	42
14	2017	3	Sants-Montjuïc	15	Hostafrancs	Boys	78
15	2017	3	Sants-Montjuïc	16	la Bordeta	Boys	81
16	2017	3	Sants-Montjuïc	17	Sants - Badal	Boys	108
17	2017	3	Sants-Montjuïc	18	Sants	Boys	155
18	2017	4	Les Corts	19	les Corts	Boys	199
19	2017	4	Les Corts	20	la Maternitat i Sant Ramon	Boys	101
20	2017	4	Les Corts	21	Pedralbes	Boys	32
21	2017	5	Sarrià-Sant Gervasi	22	Vallvidrera, el Tibidabo i les Planes	Boys	17
22	2017	5	Sarrià-Sant Gervasi	23	Sarrià	Boys	122
23	2017	5	Sarrià-Sant Gervasi	24	les Tres Torres	Boys	74
24	2017	5	Sarrià-Sant Gervasi	25	Sant Gervasi - la Bonanova	Boys	104
25	2017	5	Sarrià-Sant Gervasi	26	Sant Gervasi - Galvany	Boys	206

	Year	District Code	District Name	Neighborhood Code	Neighborhood Name	Gender	Number
26	2017	5	Sarrià-Sant Gervasi	27	el Putxet i el Farró	Boys	127
27	2017	6	Gràcia	28	Vallcarca i els Penitents	Boys	66
28	2017	6	Gràcia	29	el Coll	Boys	26
29	2017	6	Gràcia	30	la Salut	Boys	62
...	...	...	...	...	...	...	...
704	2013	8	Nou Barris	45	Porta	Girls	121
705	2013	8	Nou Barris	46	el Turó de la Peira	Girls	66
706	2013	8	Nou Barris	47	Can Peguera	Girls	8
707	2013	8	Nou Barris	48	la Guineueta	Girls	53
708	2013	8	Nou Barris	49	Canyelles	Girls	16
709	2013	8	Nou Barris	50	les Roquetes	Girls	75
710	2013	8	Nou Barris	51	Verdun	Girls	57
711	2013	8	Nou Barris	52	la Prosperitat	Girls	101
712	2013	8	Nou Barris	53	la Trinitat Nova	Girls	25
713	2013	8	Nou Barris	54	Torre Baró	Girls	19
714	2013	8	Nou Barris	55	Ciutat Meridiana	Girls	37
715	2013	8	Nou Barris	56	Vallbona	Girls	10
716	2013	9	Sant Andreu	57	la Trinitat Vella	Girls	68
717	2013	9	Sant Andreu	58	Baró de Viver	Girls	13
718	2013	9	Sant Andreu	59	el Bon Pastor	Girls	63
719	2013	9	Sant Andreu	60	Sant Andreu	Girls	233
720	2013	9	Sant Andreu	61	la Sagrera	Girls	125
721	2013	9	Sant Andreu	62	el Congrés i els Indians	Girls	56
722	2013	9	Sant Andreu	63	Navas	Girls	81
723	2013	10	Sant Martí	64	el Camp de l'Arpa del Clot	Girls	155
724	2013	10	Sant Martí	65	el Clot	Girls	96
725	2013	10	Sant Martí	66	el Parc i la Llacuna del Poblenou	Girls	64
726	2013	10	Sant Martí	67	la Vila Olímpica del Poblenou	Girls	54
727	2013	10	Sant Martí	68	el Poblenou	Girls	160
728	2013	10	Sant Martí	69	Diagonal Mar i el Front Marítim del Poblenou	Girls	97
729	2013	10	Sant Martí	70	el Besòs i el Maresme	Girls	93
730	2013	10	Sant Martí	71	Provençals del Poblenou	Girls	95
731	2013	10	Sant Martí	72	Sant Martí de Provençals	Girls	97

	Year	District Code	District Name	Neighborhood Code	Neighborhood Name	Gender	Number
732	2013	10	Sant Martí	73	la Verneda i la Pau	Girls	94
733	2013	99	No consta	99	No consta	Girls	2

734 rows × 7 columns

Choose one variable and plot that variable four different ways.

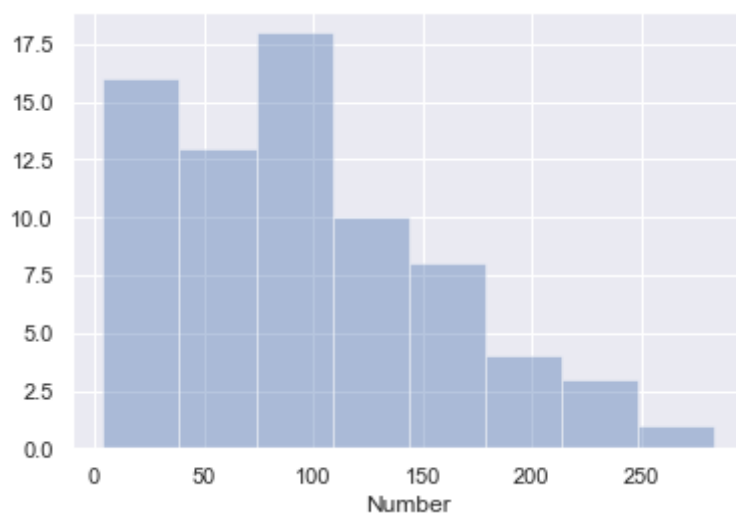
```
In [5]: 1 #before comma means select these rows, after column means select these c
2 #I am placing parenthesis around the boolean statement, not around the c
3 #Do not put the comma inside the boolean statement
4 #Year and gender are two different data types. Because year is an intege
5 #quotes.
6 print ('X axis = number of births in bins of 50. Y axis = frequency, in
7 print ('Frequency of various number ranges of boy births, given the numb
8 num_boy_births = births_df.loc[(births_df['Gender'] == 'Boys') & (births
9 #plt.hist(num_boy_births);
10 sns.distplot(num_boy_births, kde=False, rug=False);
11 plt.show()
```

X axis = number of births in bins of 50. Y axis = frequency, in other words the number of times this number range of births occurred.

Frequency of various number ranges of boy births, given the number of births reported per neighborhood

/Users/katrinajohnson/anaconda3/lib/python3.7/site-packages/scipy/stats/stats.py:1713: FutureWarning: Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.

```
return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```



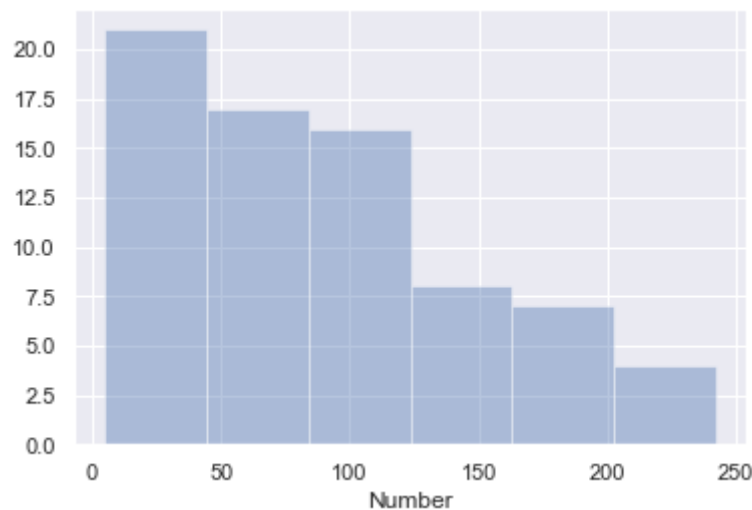
```
In [6]: 1 print ('X axis = number of births in bins of 50. Y axis = frequency, in
2 print ('Frequency of various number ranges of boy births, given the numb
3 num_girl_births = births_df.loc[(births_df['Gender'] == 'Girls') & (birt
4 sns.distplot(num_girl_births, kde=False, rug=False)
5 plt.show()
6
7
```

X axis = number of births in bins of 50. Y axis = frequency, in other words the number of times this number range of births occurred.

Frequency of various number ranges of boy births, given the number of births reported per neighborhood

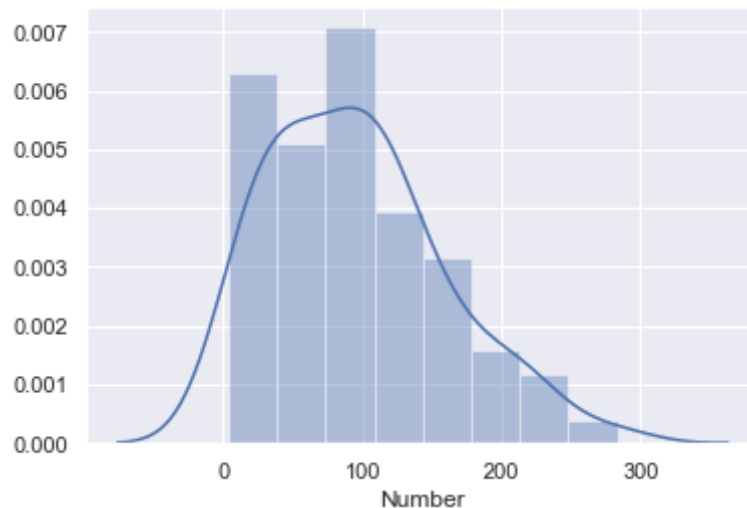
/Users/katrinajohnson/anaconda3/lib/python3.7/site-packages/scipy/stats/stats.py:1713: FutureWarning: Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.

```
return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```



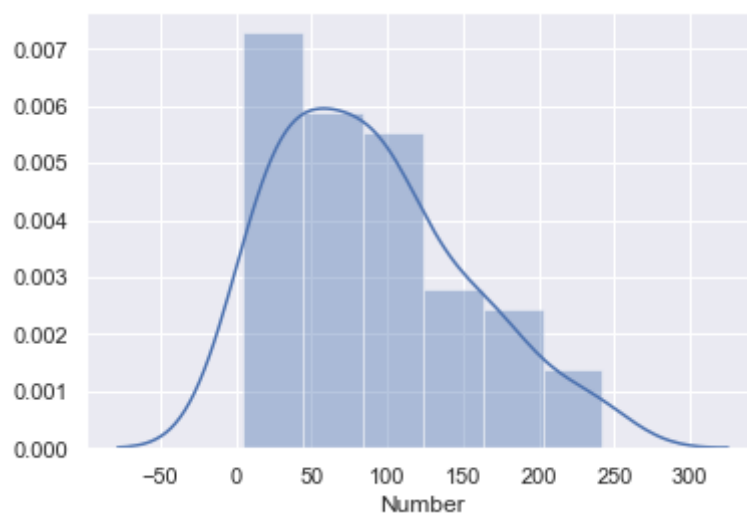
```
In [7]: 1 sns.distplot(num_boy_births, hist=True, rug=False);
```

```
/Users/katrinajohnson/anaconda3/lib/python3.7/site-packages/scipy/stats/s  
tats.py:1713: FutureWarning: Using a non-tuple sequence for multidimensio  
nal indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`.  
In the future this will be interpreted as an array index, `arr[np.array(s  
eq)]`, which will result either in an error or a different result.  
    return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```



```
In [8]: 1 sns.distplot(num_girl_births, hist=True, rug=False);
```

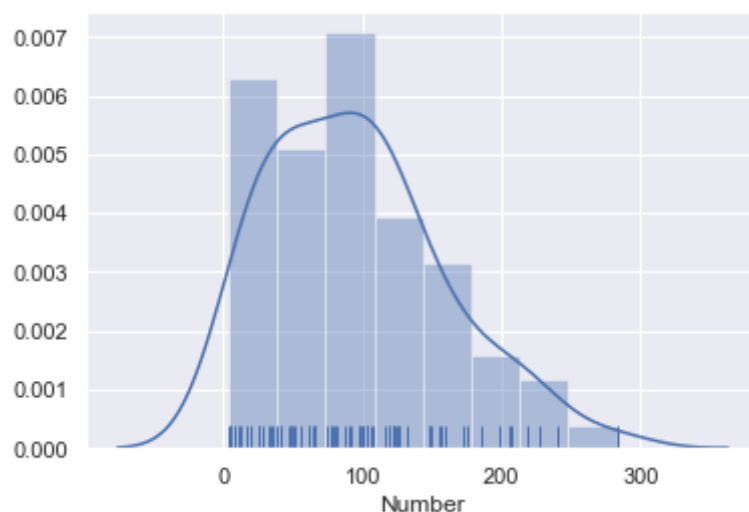
```
/Users/katrinajohnson/anaconda3/lib/python3.7/site-packages/scipy/stats/s  
tats.py:1713: FutureWarning: Using a non-tuple sequence for multidimensio  
nal indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`.  
In the future this will be interpreted as an array index, `arr[np.array(s  
eq)]`, which will result either in an error or a different result.  
    return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```



```
In [18]: 1 sns.distplot(num_boy_births, hist=True, rug=True);
        2 #label='Number of boy births in 2017'
```

/Users/katrinajohnson/anaconda3/lib/python3.7/site-packages/scipy/stats/stats.py:1713: FutureWarning: Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.

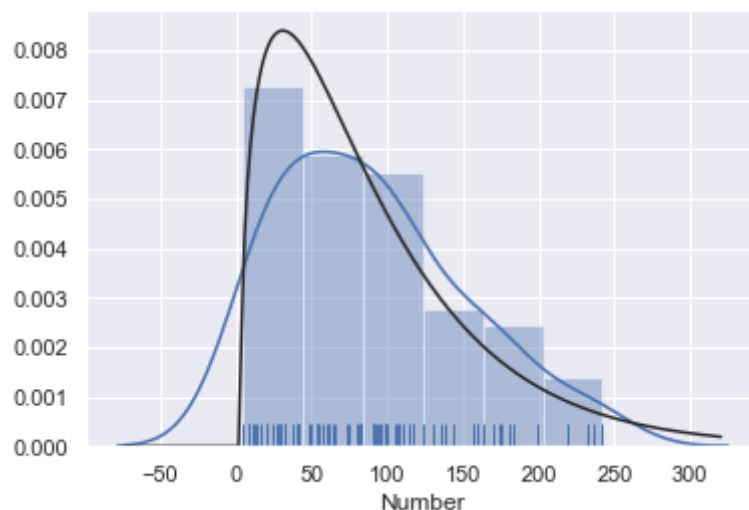
```
return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```



```
In [10]: 1 sns.distplot(num_girl_births, hist=True, rug=True, fit=stats.gamma);
```

/Users/katrinajohnson/anaconda3/lib/python3.7/site-packages/scipy/stats/stats.py:1713: FutureWarning: Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.

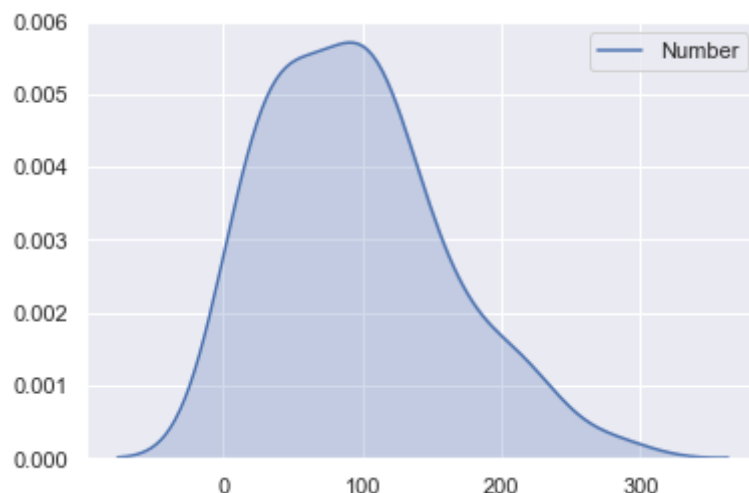
```
return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```



```
In [11]: 1 sns.kdeplot(num_boy_births, shade=True);
```

```
/Users/katrinajohnson/anaconda3/lib/python3.7/site-packages/scipy/stats/s  
tats.py:1713: FutureWarning: Using a non-tuple sequence for multidimensio  
nal indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`.  
In the future this will be interpreted as an array index, `arr[np.array(s  
eq)]`, which will result either in an error or a different result.
```

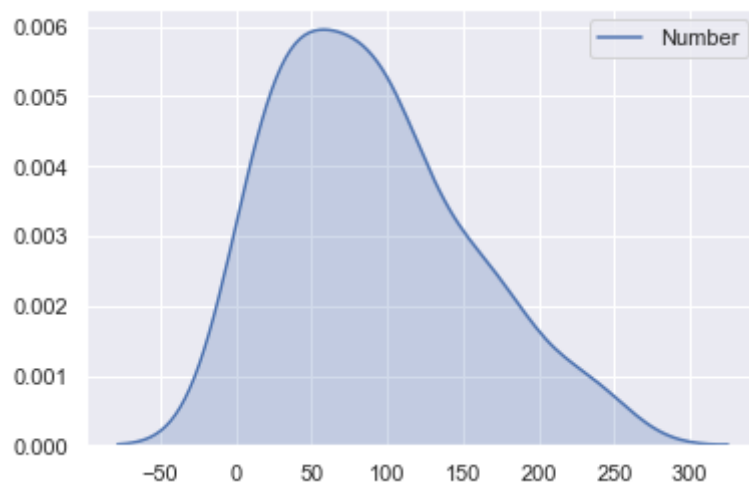
```
return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```



```
In [12]: 1 sns.kdeplot(num_girl_births, shade=True);
```

```
/Users/katrinajohnson/anaconda3/lib/python3.7/site-packages/scipy/stats/s  
tats.py:1713: FutureWarning: Using a non-tuple sequence for multidimensio  
nal indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`.  
In the future this will be interpreted as an array index, `arr[np.array(s  
eq)]`, which will result either in an error or a different result.
```

```
return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```

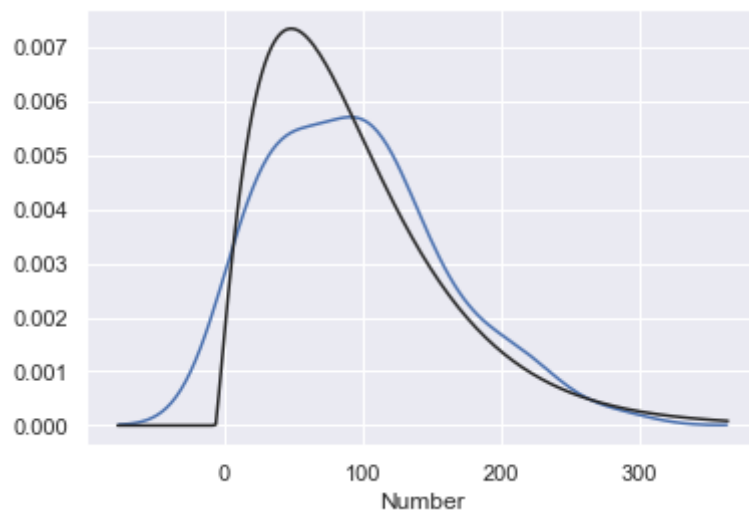




```
In [13]: 1 sns.distplot(num_boy_births, hist=False, fit=stats.gamma);
```

/Users/katrinajohnson/anaconda3/lib/python3.7/site-packages/scipy/stats/stats.py:1713: FutureWarning: Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.

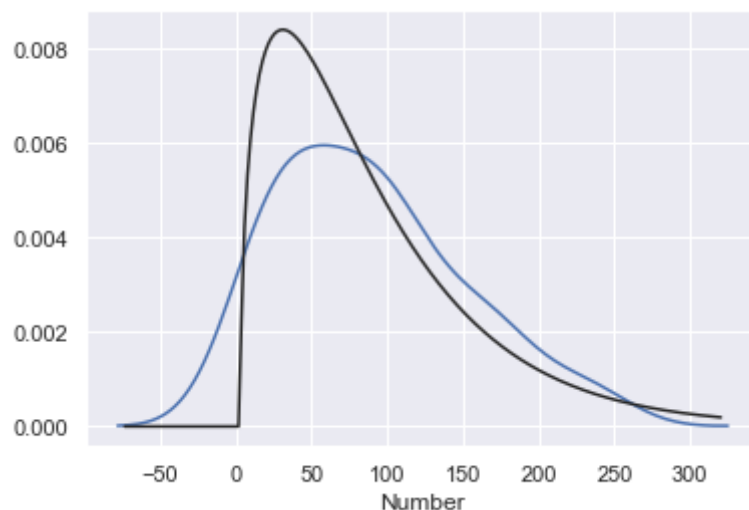
```
return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```



```
In [14]: 1 sns.distplot(num_girl_births, hist=False, fit=stats.gamma);
2
3
```

/Users/katrinajohnson/anaconda3/lib/python3.7/site-packages/scipy/stats/stats.py:1713: FutureWarning: Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.

```
return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```

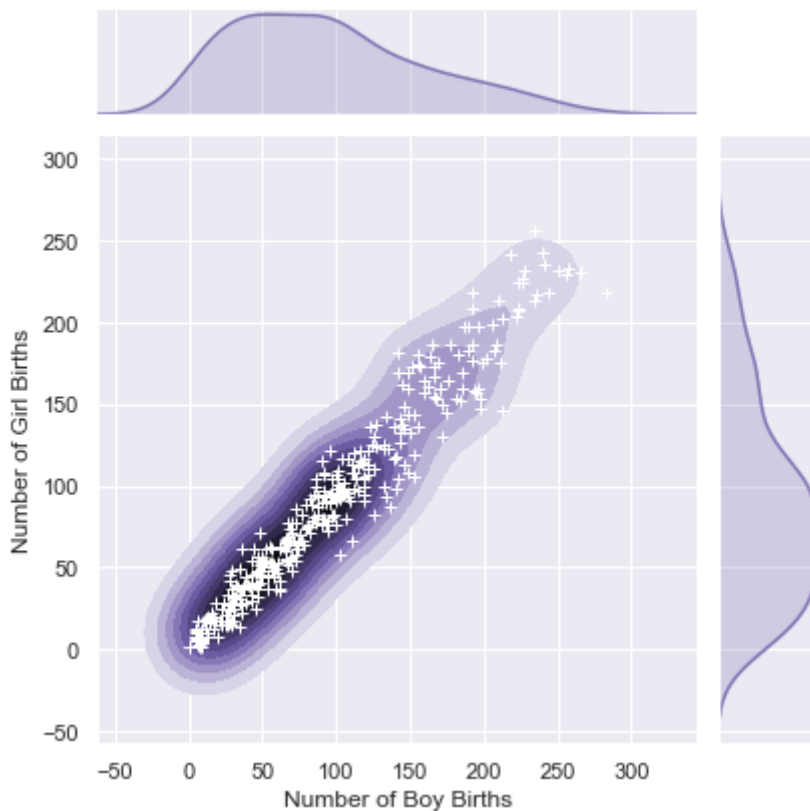


Choose two continuous variables, and plot them three different ways.

```
In [15]: 1 g = sns.jointplot(x=births_df.loc[births_df['Gender'] == 'Boys', 'Number of Boy Births'],
2           y=births_df.loc[births_df['Gender'] == 'Boys', 'Number of Girl Births'],
3           kind='scatter', c="w", s=30, linewidth=1, marker="+")
4           g.ax_joint.collections[0].set_alpha(0)
5           g.set_axis_labels('Number of Boy Births', 'Number of Girl Births');
6           #Here we see the data is shown interacting.
7           #I can see how the numbers of boy and girl births interact.
```

/Users/katrinajohnson/anaconda3/lib/python3.7/site-packages/scipy/stats/stats.py:1713: FutureWarning: Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.

```
return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```



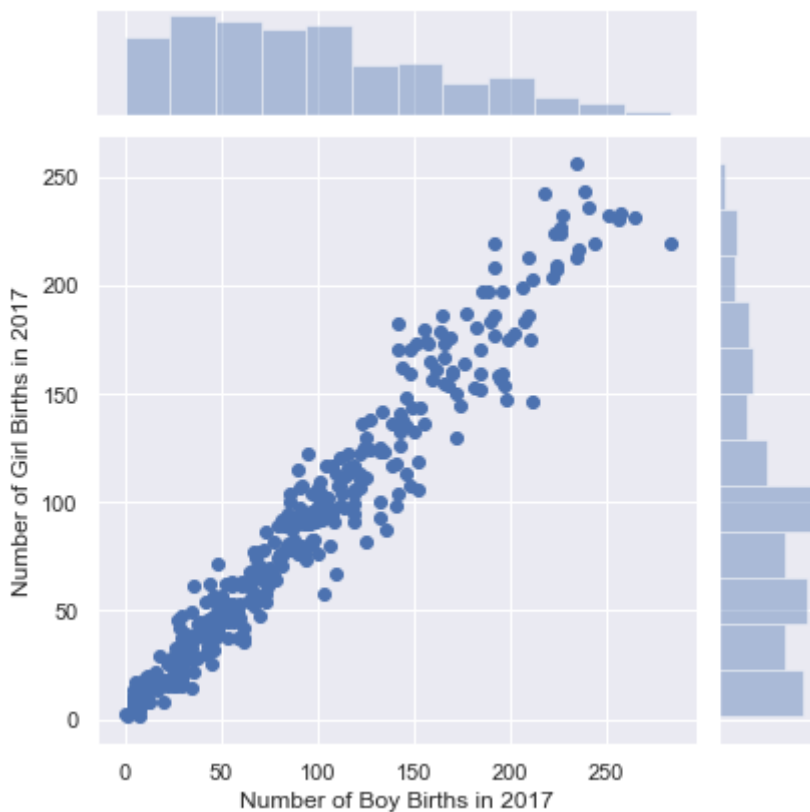
```
In [45]: 1 plt.figure(figsize=(15,3))
2 h = sns.jointplot(x=births_df.loc[births_df['Gender'] == 'Boys', 'Number of Boy Births'], y=births_df.loc[births_df['Gender'] == 'Boys', 'Number of Girl Births'])
3 h.set_axis_labels('Number of Boy Births', 'Number of Girl Births')
4
5 ;
```

/Users/katrinajohnson/anaconda3/lib/python3.7/site-packages/scipy/stats/stats.py:1713: FutureWarning: Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.

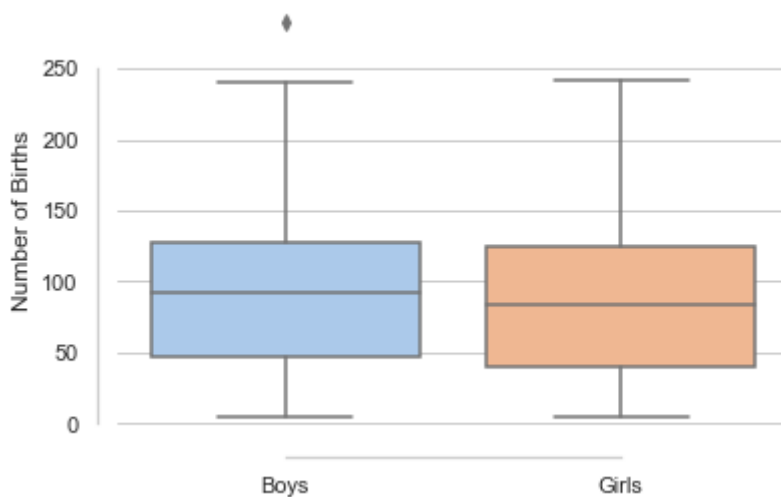
return np.add.reduce(sorted[indexer] \* weights, axis=axis) / sumval

Out[45]: ''

<Figure size 1080x216 with 0 Axes>



```
In [103]: 1 #Comparing groups again, but with a different emphasis.
2
3 # Setting the overall aesthetic.
4 sns.set(style="whitegrid")
5
6 ax = sns.boxplot(x='Gender', y='Number', data=births_df.loc[births_df
7
8 plt.title('')
9
10 sns.despine(offset=10, trim=True)
11 ax.set(xlabel='', ylabel='Number of Births')
12
13 plt.show()
```

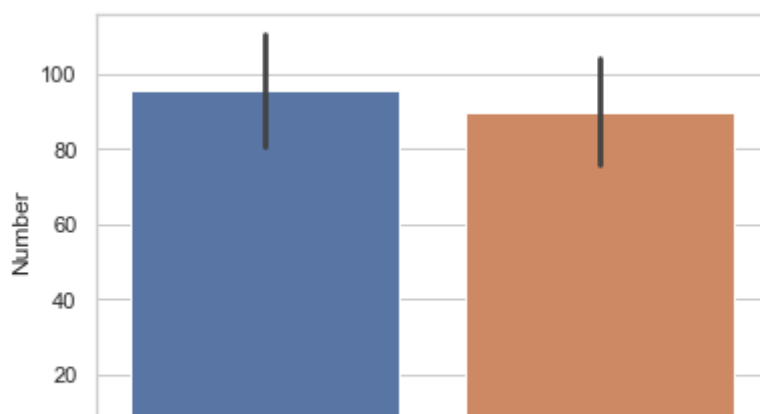


Choose one continuous variable and one categorical variable, and plot them six different ways.

```
In [120]: 1 #num_boy_births = births_df.loc[(births_df['Gender'] == 'Boys') & (bir
2 x = sns.barplot(x='Gender', y='Number', data=births_df.loc[births_df['
3
4
```

/Users/katrinajohnson/anaconda3/lib/python3.7/site-packages/scipy/stats/s  
tats.py:1713: FutureWarning: Using a non-tuple sequence for multidimensio  
nal indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`.  
In the future this will be interpreted as an array index, `arr[np.array(s  
eq)]`, which will result either in an error or a different result.

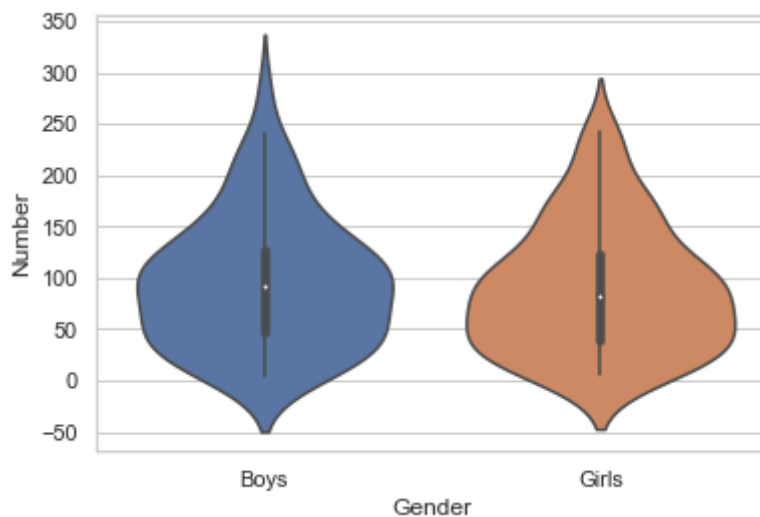
```
return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```



```
In [121]: 1 x = sns.violinplot(x='Gender', y='Number', data=births_df.loc[births_d
```

/Users/katrinajohnson/anaconda3/lib/python3.7/site-packages/scipy/stats/s  
tats.py:1713: FutureWarning: Using a non-tuple sequence for multidimensio  
nal indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`.  
In the future this will be interpreted as an array index, `arr[np.array(s  
eq)]`, which will result either in an error or a different result.

```
return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```



```
In [126]: 1 # Setting the overall aesthetic.
2 sns.set(style="whitegrid")
3
4 g = sns.factorplot(x='Gender', y='Number', hue='Year', data=births_df,
5                  size=6, kind="point", palette="pastel",ci=95,dodge=
6 g.despine(left=True)
7 g.set_ylabels("Number of Births in 2017")
8 g.set_xlabels("")
9 plt.title('')
10 plt.show()
```

/Users/katrinajohnson/anaconda3/lib/python3.7/site-packages/seaborn/categorical.py:3666: UserWarning: The `factorplot` function has been renamed to `catplot`. The original name will be removed in a future release. Please update your code. Note that the default `kind` in `factorplot` (`'point'`) has changed to `strip` in `catplot`.

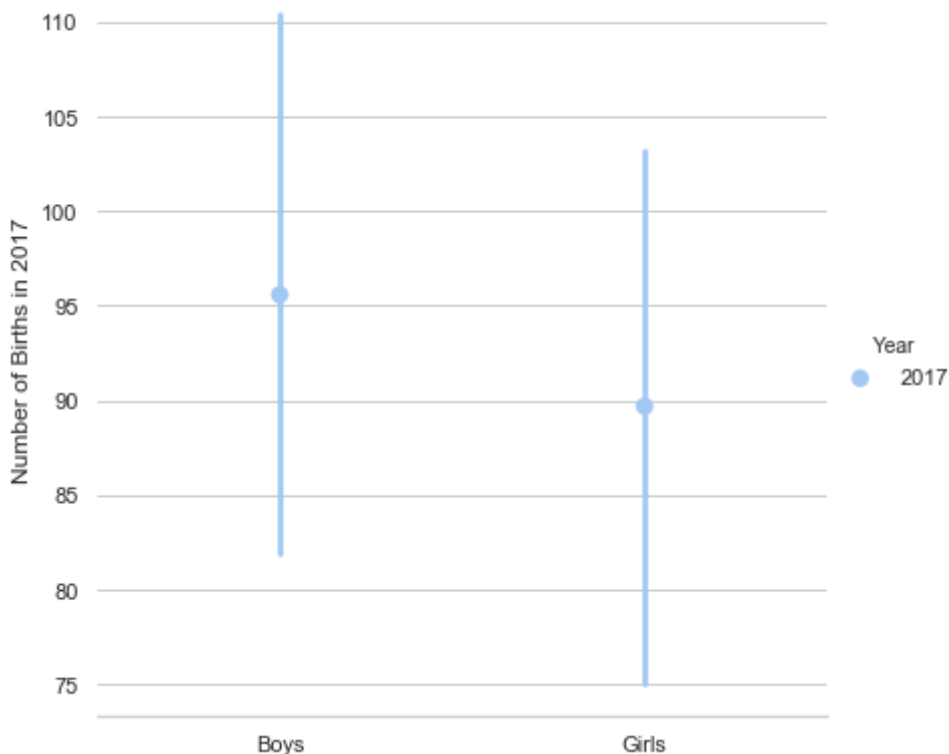
warnings.warn(msg)

/Users/katrinajohnson/anaconda3/lib/python3.7/site-packages/seaborn/categorical.py:3672: UserWarning: The `size` parameter has been renamed to `height`; please update your code.

warnings.warn(msg, UserWarning)

/Users/katrinajohnson/anaconda3/lib/python3.7/site-packages/scipy/stats/stats.py:1713: FutureWarning: Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.

return np.add.reduce(sorted[indexer] \* weights, axis=axis) / sumval



```
In [184]: 1 #x=births_df.loc[births_df['Gender'] == 'Boys', 'Number'], y=births_df
          2
          3 sns.scatterplot(x=births_df.loc[births_df['Gender'] == 'Girls'], y=bi
          4 plt.show()
          5
          6
```

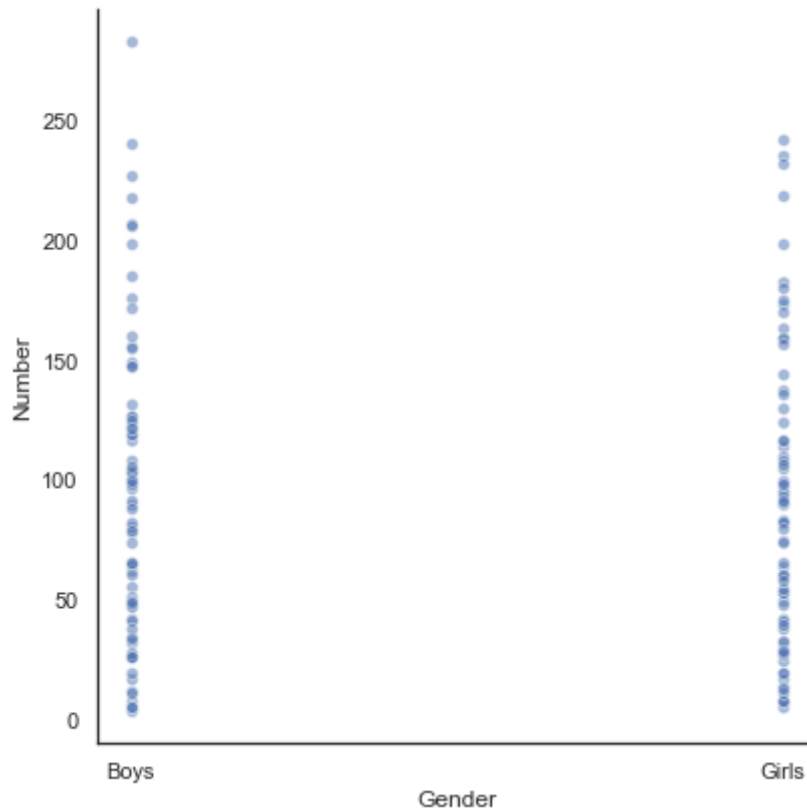
```
-----
--
ValueError                                Traceback (most recent call last)
~/anaconda3/lib/python3.7/site-packages/pandas/core/internals.py in creat
e_block_manager_from_arrays(arrays, names, axes)
    4871     try:
-> 4872         blocks = form_blocks(arrays, names, axes)
    4873         mgr = BlockManager(blocks, axes)

~/anaconda3/lib/python3.7/site-packages/pandas/core/internals.py in form_
blocks(arrays, names, axes)
    4937     if len(items_dict['ObjectBlock']) > 0:
-> 4938         object_blocks = _simple_blockify(items_dict['ObjectBlock'
], np.object_)
    4939         blocks.extend(object_blocks)

~/anaconda3/lib/python3.7/site-packages/pandas/core/internals.py in _simp
le_blockify(tuples, dtype)
    ....
```

```
In [176]: 1
          2
          3 # Plot miles per gallon against horsepower with other semantics
          4 sns.relplot(x='Gender', y='Number', data=births_df.loc[births_df['Year
          5                      sizes=(40, 400), alpha=.5,
          6                      height=6)
```

Out[176]: <seaborn.axisgrid.FacetGrid at 0x1alc295a58>



```
In [ ]: 1
```