

DSO 545: Statistical Computing and Data Visualization

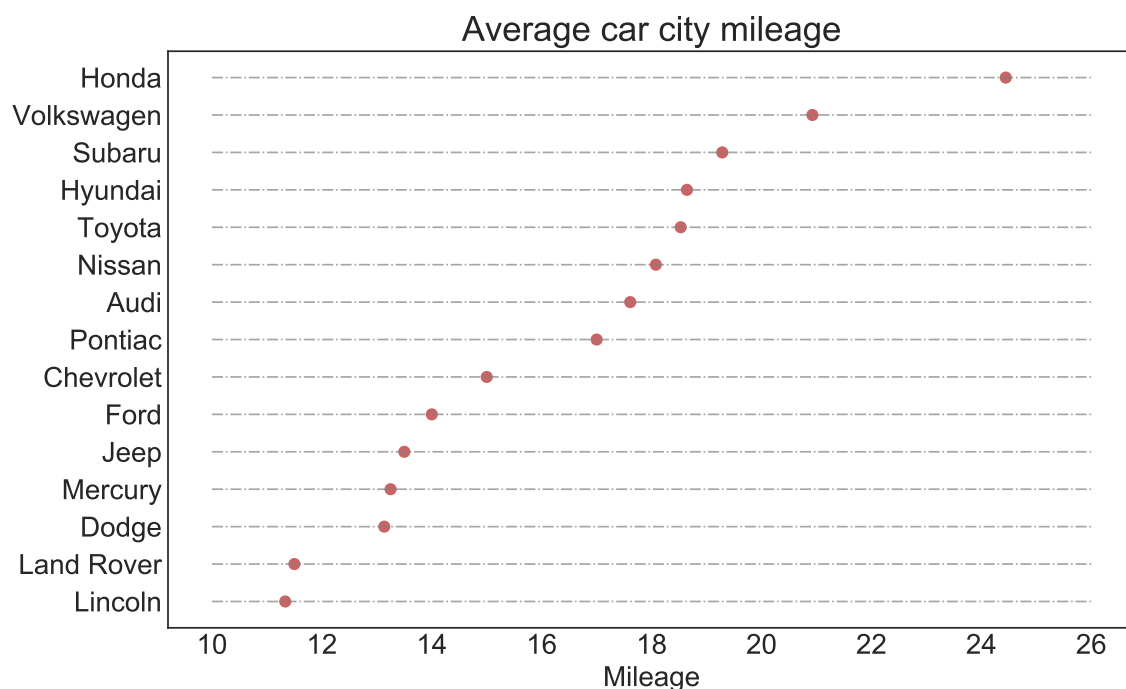
Abbass Al Sharif

Fall 2019

Lab 8: Data Visualization Using Matplotlib (Part2)

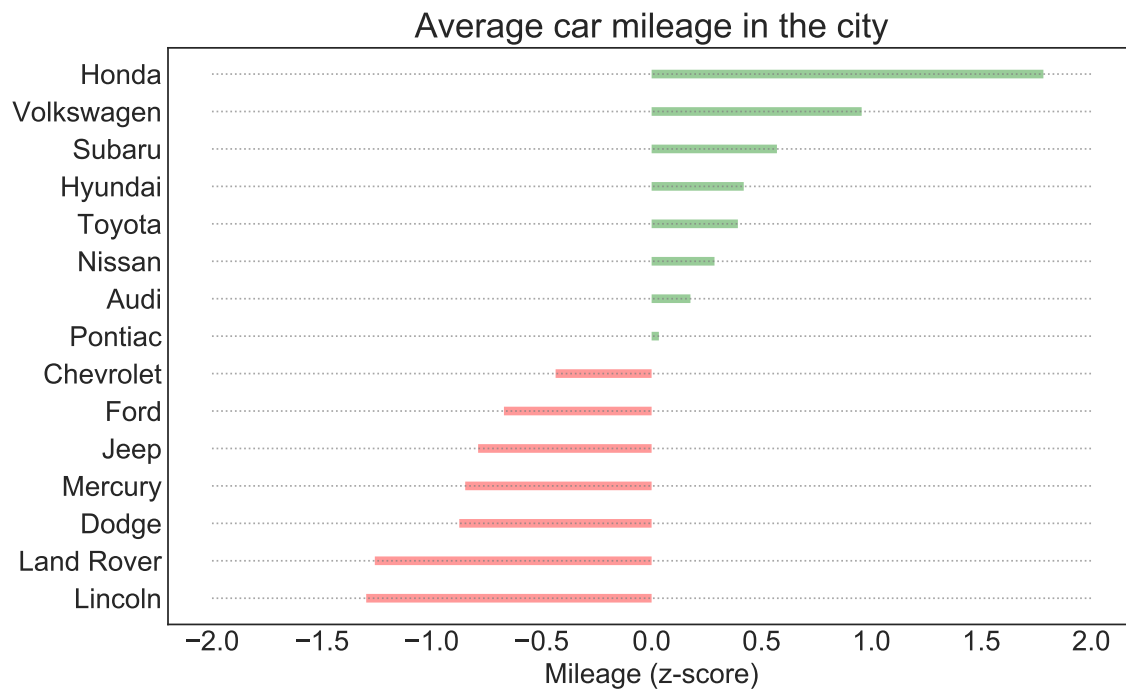
Dot Plot

1. Load the dataset `mpg.csv`, and create the following dot plot using matplotlib. This dotplot looks at the average miles per gallon for each car manufacturer.



Diverging Bars

2. Create the following diverging bar plot using matplotlib. This plot looks at the average miles per gallon (z-scores) for each car manufacturer.

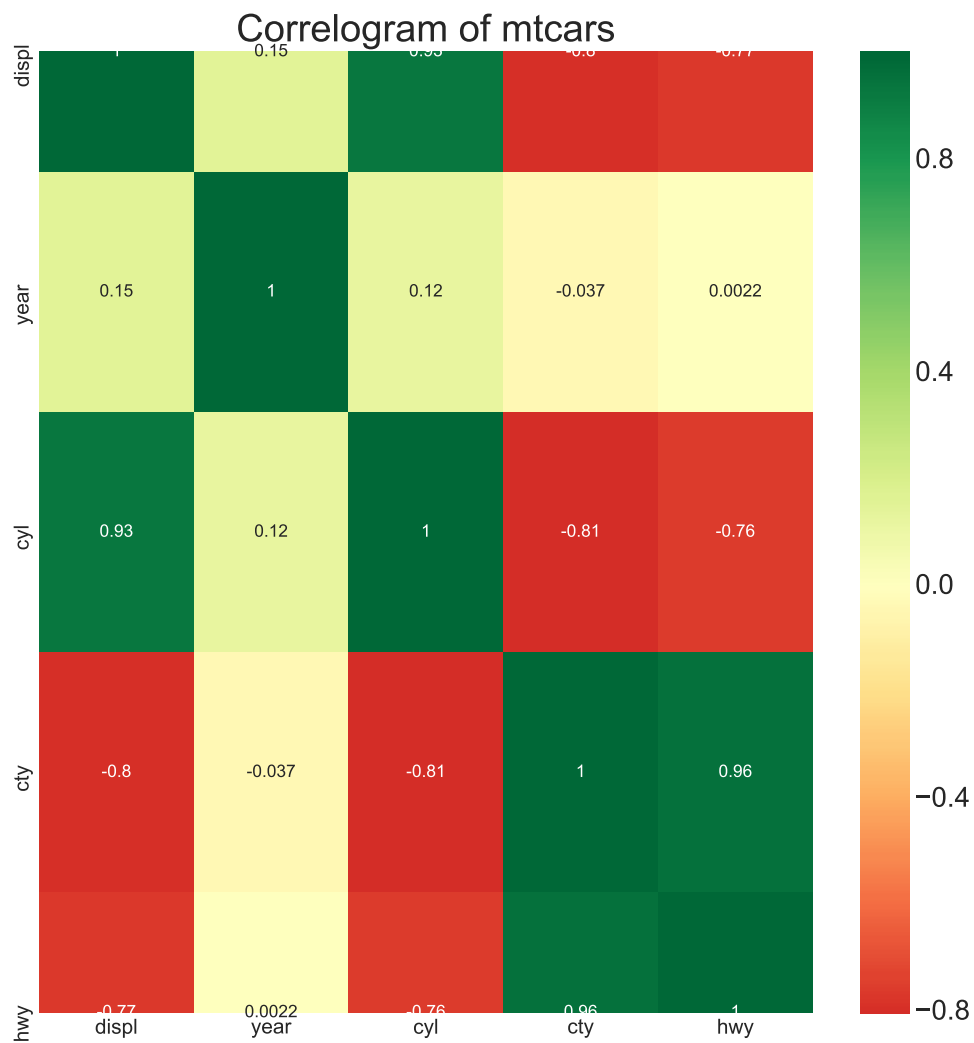


Correllogram

3. Create a correllogram that shows the correlation between all numerical variables in the `mpg.csv` dataset.
(Please note that the aspect ratio of the figure will look differently in Jupyter Notebook)

```
## (array([0.5, 1.5, 2.5, 3.5, 4.5]), <a list of 5 Text xticklabel objects>)
```

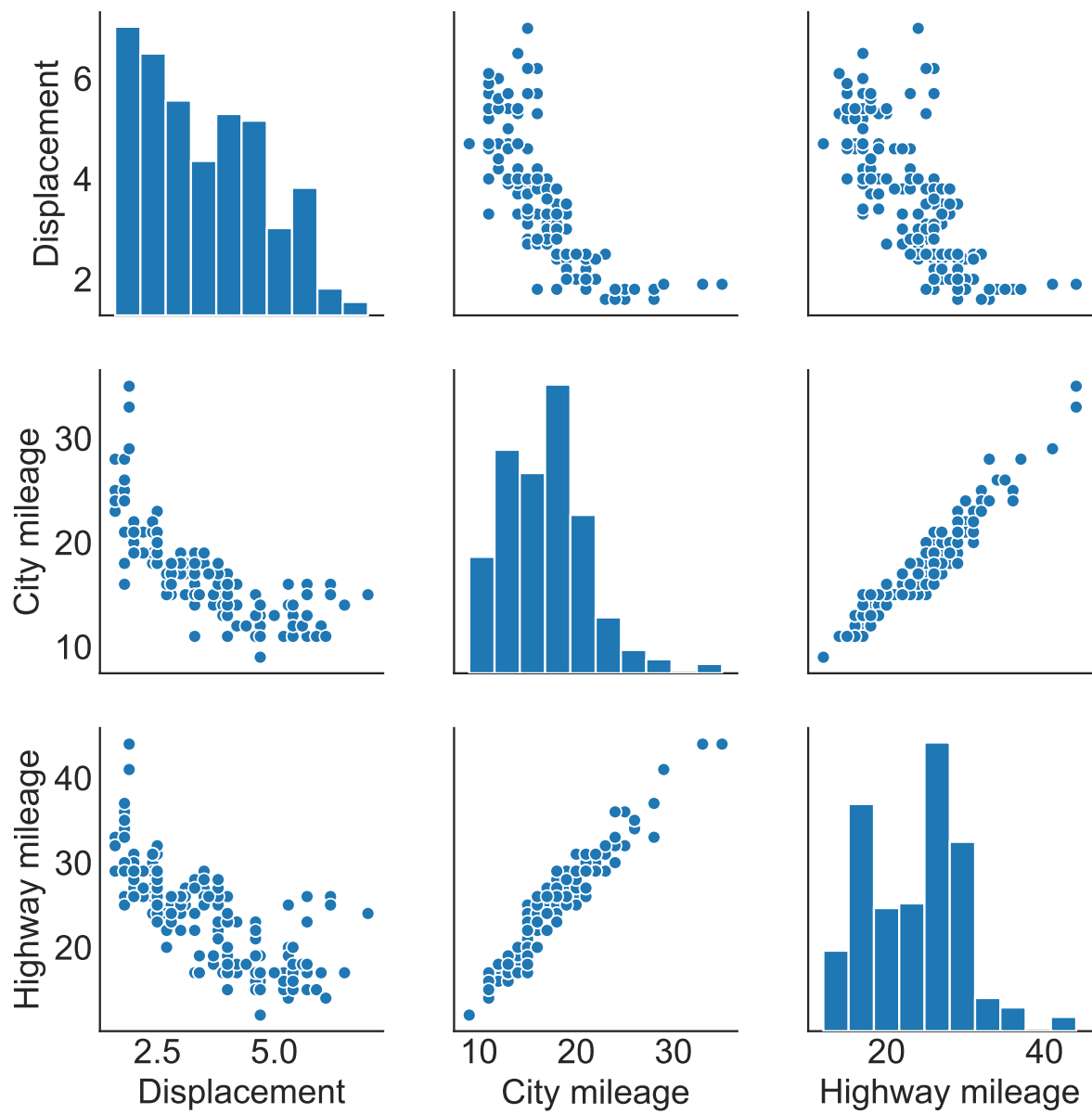
```
## (array([0.5, 1.5, 2.5, 3.5, 4.5]), <a list of 5 Text yticklabel objects>)
```



Pariwise Plots

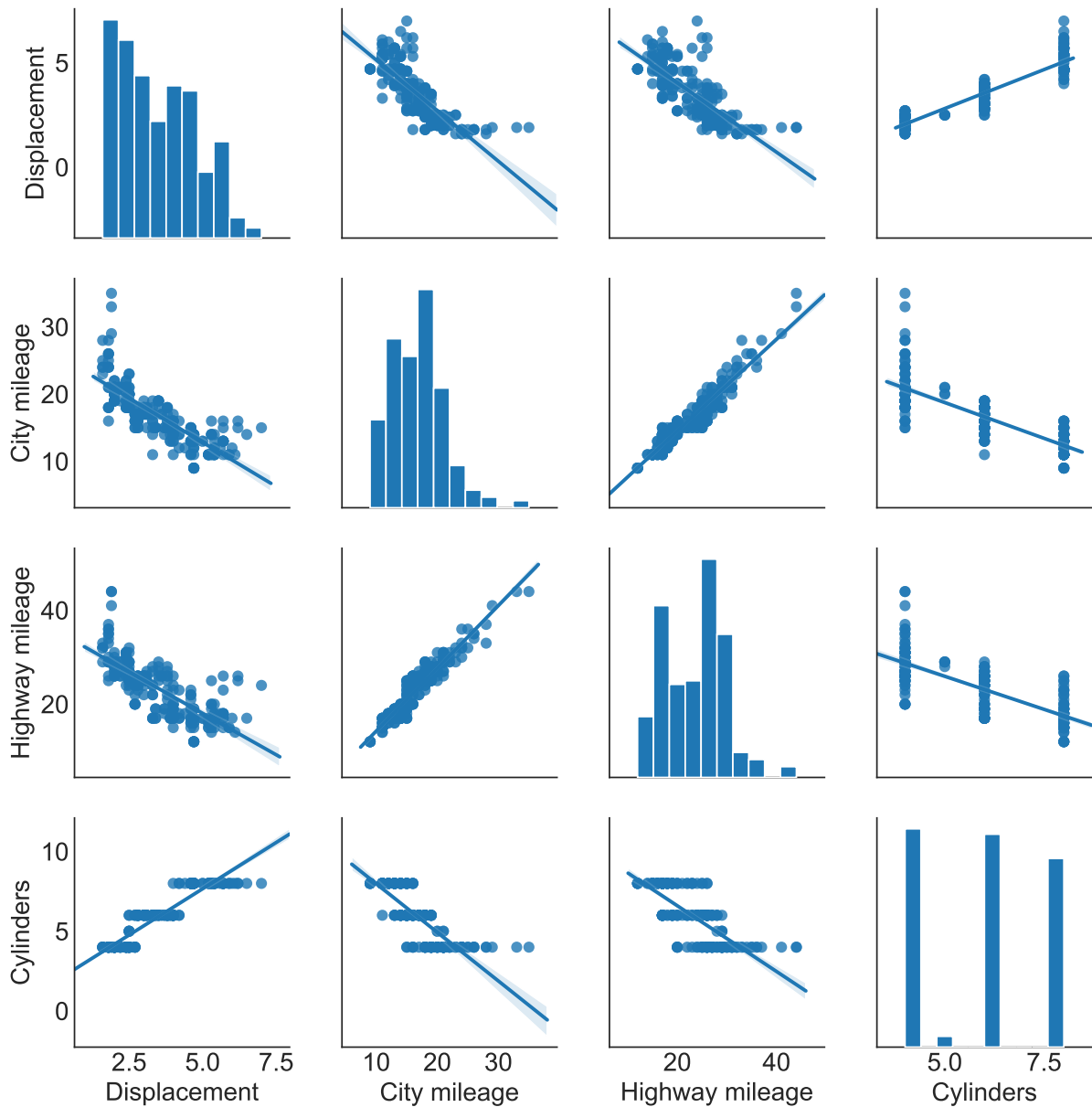
4. Create a pairwise plot to show the relationship between the following numerical variables in the `mpg.csv` dataset.

```
## <seaborn.axisgrid.PairGrid object at 0x122f41550>
```



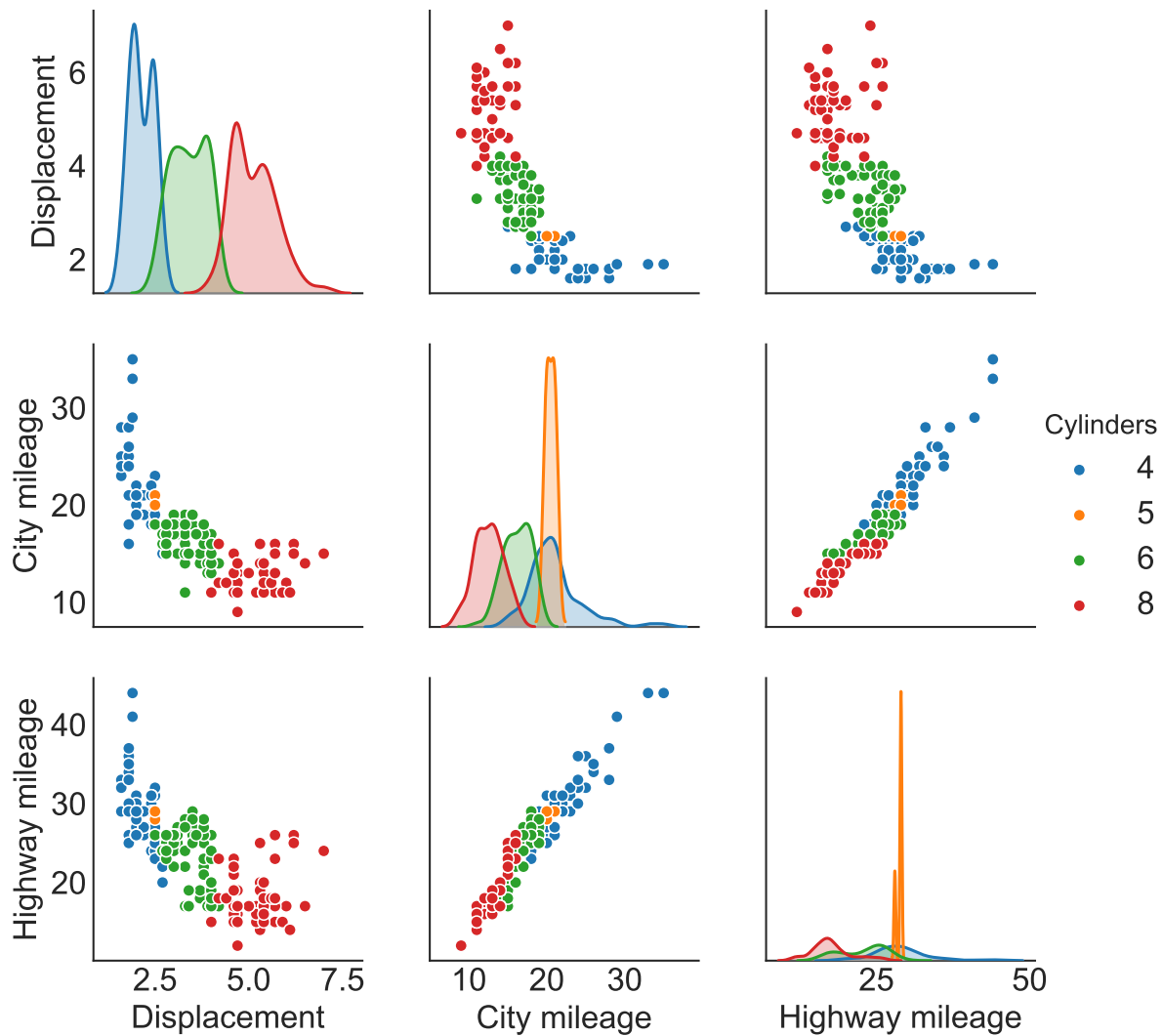
5. Update the previous pairwise plot to the regression lines.

```
## <seaborn.axisgrid.PairGrid object at 0x12495f7d0>
```



6. Update the initial pairwise plot to show the relationships among the different car engine sizes (cyl).

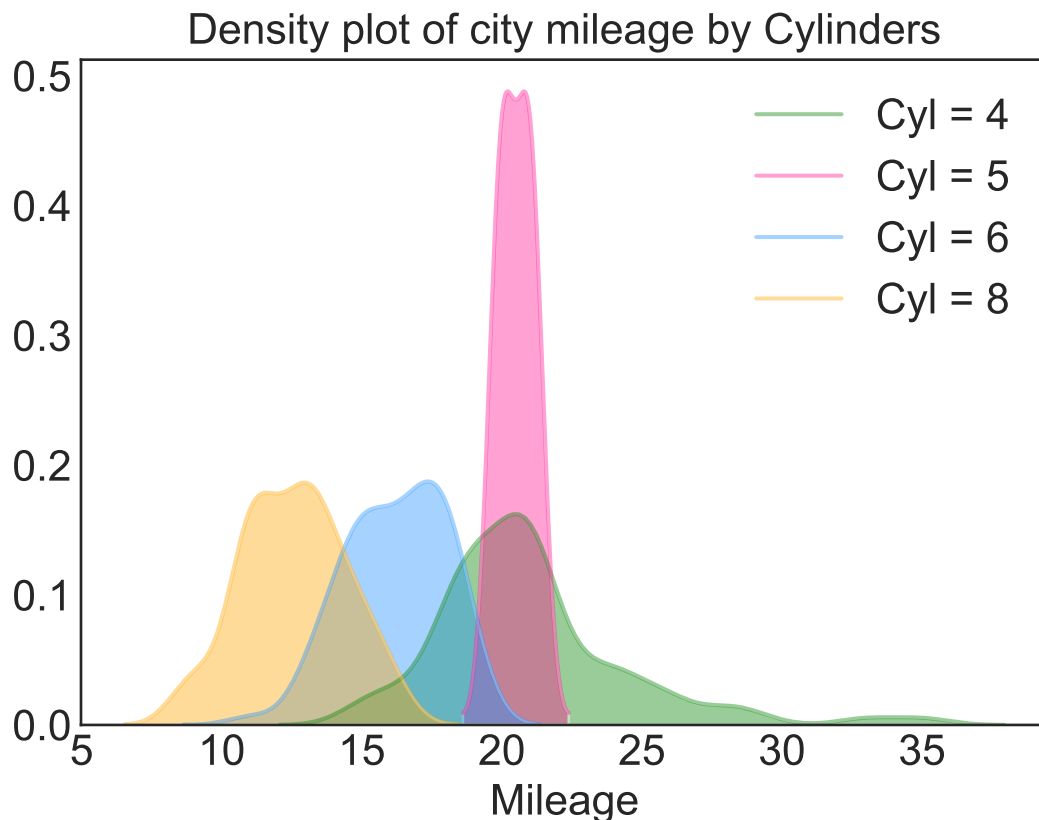
```
## <seaborn.axisgrid.PairGrid object at 0x12694e490>
##
## /anaconda3/envs/r-reticulate/lib/python3.7/site-packages/statsmodels/nonparametric/kde.py:487: RuntimeWarning:
##   binned = fast_linbin(X, a, b, gridsize) / (delta * nobs)
## /anaconda3/envs/r-reticulate/lib/python3.7/site-packages/statsmodels/nonparametric/kdetools.py:34: RuntimeWarning:
##   FAC1 = 2*(np.pi*bw/RANGE)**2
```



Density Plots

7. Create the following density plot to describe how the distribution of city mileage varies with respect the number of cylinders.

```
## <matplotlib.axes._subplots.AxesSubplot object at 0x1268d4910>
## <matplotlib.axes._subplots.AxesSubplot object at 0x1268d4910>
## <matplotlib.axes._subplots.AxesSubplot object at 0x1268d4910>
```



Plotting with different scales using secondary Y axis

8. Use the `economics.csv` dataset to create the following plot which shows the relationship between personal savings rate and # of unemployed from 1967 to 2012.

```
##          date    pce    pop  psavert  uempmed  unemploy
## 0  1967-07-01  507.4 198712    12.5     4.5     2944
## 1  1967-08-01  510.5 198911    12.5     4.7     2945
## 2  1967-09-01  516.3 199113    11.7     4.6     2958
## 3  1967-10-01  512.9 199311    12.5     4.9     3143
## 4  1967-11-01  518.1 199498    12.5     4.7     3066
```

```
## [<matplotlib.axis.XTick object at 0x127eac650>, <matplotlib.axis.XTick object at 0x127e9bd50>, <matplotlib.axis.XTick object at 0x127e9bd50>, <matplotlib.axis.XTick object at 0x127e9bd50>]
```

```
## [Text(0, 0, '1967-07-01'), Text(0, 0, '1972-07-01'), Text(0, 0, '1977-07-01'), Text(0, 0, '1982-07-01'), Text(0, 0, '1987-07-01'), Text(0, 0, '1992-07-01'), Text(0, 0, '1997-07-01'), Text(0, 0, '2002-07-01'), Text(0, 0, '2007-07-01'), Text(0, 0, '2012-07-01')]
```

