Eduardo De Paz-Barahona

Data Analyst Sample Work

What conclusions can we draw from examining data on the number of gears in a vehicle and its effect on MPG, specifically ComboMPG (which includes both city MPG and Highway MPG).

In [15]:
```
%matplotlib inline
# Allows for dataset to be shown in jupyter notebook

import pandas as pd
# Package used to import the csv file

df = pd.read_csv ("http://media.sundog-soft.com/SelfDriving/FuelEfficiency.csv")
# Python pulls the data from a specific file path

df.head()
# Display the data
```

Out[15]:

	Mfr Name	Carline	Eng Displ	Cylinders	Transmission	CityMPG	HwyMPG	CombMPG	# Gears
0	aston martin	Vantage V8	4.0	8	Auto(S8)	18	25	21	8
1	Volkswagen Group of	Chiron	8.0	16	Auto(AM-S7)	9	14	11	7
2	General Motors	CORVETTE	6.2	8	Auto(S8)	12	20	15	8
3	General Motors	CORVETTE	6.2	8	Auto(S8)	15	25	18	8
4	General Motors	CORVETTE	6.2	8	Auto(S8)	14	23	17	8

In [6]:
```
len(df)
# Shows the length of the data frame
```

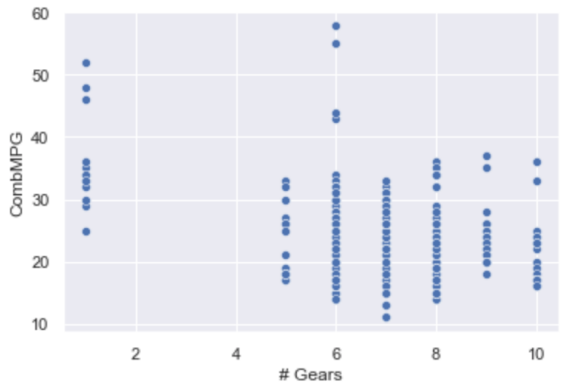
Out[6]: 718

In [9]:
```
import seaborn as sns
sns.set()
# Seaborn is meant to modernize the data visualization!!!

sns.scatterplot (x="# Gears", y= "CombMPG", data=df)

# Asking Python to make a scatterplot, showing the number of gears in the x-axis and combonation of MPG on y-axis
```

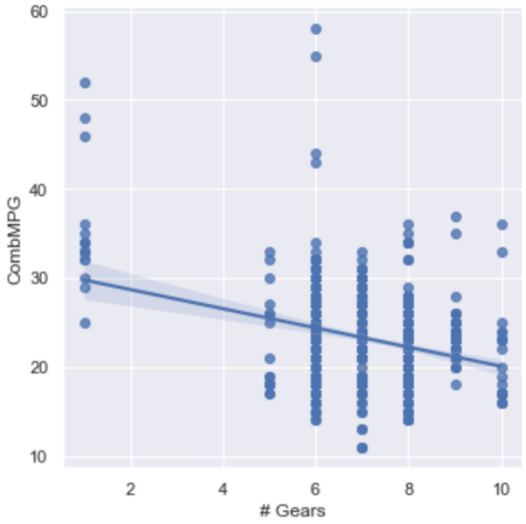
Out[9]: <AxesSubplot:xlabel='# Gears', ylabel='CombMPG'>



Analysis on scatter plot

The data tells us that there's a wide range of MPG values depending on the type of gear box on the y-axis. Also, if you look at where the data points are clustered (gears 6 to 8 especially), you can sort of see a downward trend in MPG as the number of gears increases. But it's subtle.

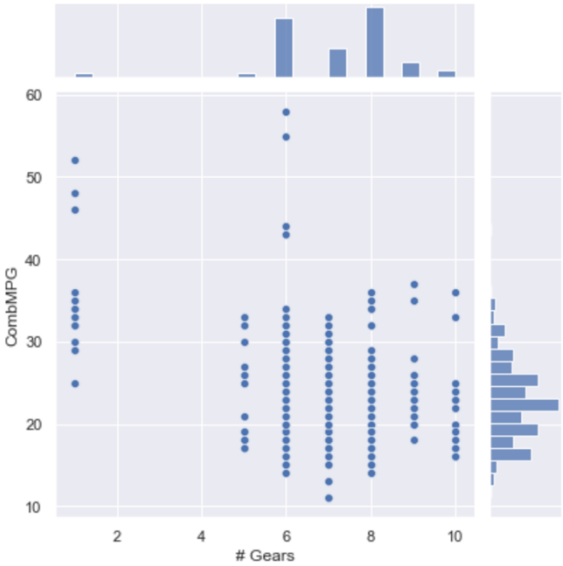
In [9]:
```
sns.lmplot(x="# Gears", y="CombMPG", data=df)
# Told Seaborn to make a linear model showing the number of gears on the x-axis and combonation of MPG on the y-axis
sns.set()
```



Analysis on Linear Regression Model

The linear model plot gives us a linear regression model of our data, which favors the argument of lower MPG's is associated with a greater number of gears. In simple terms, more gears is not better when it comes to efficiency, it seems. We also see the error bars on that regression line, which tells us this trend is probably real and not just the result of randomness in the samples.

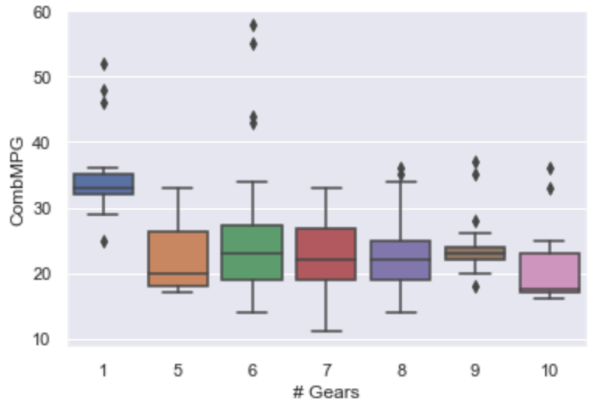
In [11]:
```
sns.jointplot(x= "# Gears", y="CombMPG", data=df)
# Told Seaborn to make a joint plot showing the number of gears on the x-axis and combonation of MPG on the y-axis
sns.set()
```



Analysis on joint plot

The joint plot gives us histograms on each axis, which provides some interesting insights. The most common gear configuration seems to be 8, with 6 closely behind. Also, the MPG ratings seem to roughly follow a bell curve centered at around 22 MPG or so when you examine the histograms on each axis.

In [12]:
```
sns.boxplot(x='# Gears', y="CombMPG", data=df)
# Told Seaborn to make a box plot showing the number of gears on the x-axis and combonation of MPG on the y-axis
sns.set()
```

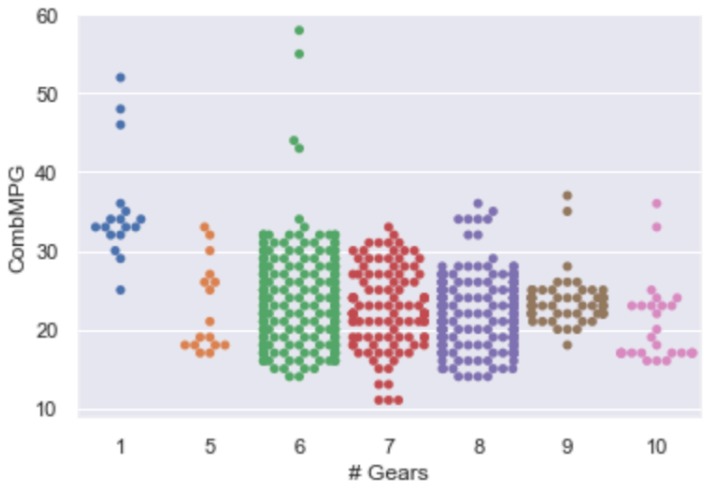


Analysis on boxplot

The box plot shows us that the range of MPG values we see on each gearbox type aren't as crazily distributed as it seemed at first glance; many of the extreme values are in fact outliers that are best discarded when analyzing the trends. This makes the real relationships easier to see; The transmissions with a single gear are really quite good at fuel efficiency (higher MPG's are more efficient vehicles). Gears between 5 and 8 are roughly the same, but from 8-10 the MPG's decline noticeably.

In [10]:
```
sns.swarmplot(x= '# Gears', y='CombMPG', data=df)
# Asking Python to make a swarmplot, showing the number of gears in the x-axis and combonation of MPG on y-axis
```

Out[10]: <AxesSubplot:xlabel='# Gears', ylabel='CombMPG'>



Analysis on swarm plot

The swarm plot makes it even more apparent that the 6 gear vehicles have greater outliers; most 6-gear vehicles have less than 32 MPG. And the overall trend is perhaps easiest to visualize and understand in this representation.

Conclusion: More gears result in worse fuel efficiency.