In [1]:
```
import matplotlib.pyplot as plt
import pandas as pd
import statsmodels.formula.api as smf
import seaborn as sns
import numpy as np

%matplotlib inline
```

In [2]:
```
scores=pd.read_csv("2013-2017_School_Math_Results_-_All.csv")
```

In [3]:
```
scores.head()
```

Out[3]:

	DBN	School Name	Grade	Year	Category	Number Tested	Mean Scale Score	Level1_N	Level1_%	Lev
0	01M015	PS 015 ROBERTO CLEMENTE	3	2013	All Students	27	278	16	59.3	11
1	01M015	PS 015 ROBERTO CLEMENTE	3	2014	All Students	18	286	6	33.3	9
2	01M015	PS 015 ROBERTO CLEMENTE	3	2015	All Students	17	280	10	58.8	4
3	01M015	PS 015 ROBERTO CLEMENTE	3	2016	All Students	21	275	13	61.9	4
4	01M015	PS 015 ROBERTO CLEMENTE	3	2017	All Students	29	302	8	27.6	9

In [4]:
```
scores.describe()
```

Out[4]:

	Year	Number Tested
count	23896.000000	23896.000000
mean	2015.019292	161.178607
std	1.410274	182.057290
min	2013.000000	1.000000
25%	2014.000000	62.000000
50%	2015.000000	99.000000
75%	2016.000000	184.000000
max	2017.000000	2349.000000

In [5]:
```
scores.describe(include=["O"])
```

Out[5]:

	DBN	School Name	Grade	Category	Mean Scale Score	Level1_N	Level1_%	Level2_N	Level2
count	23896	23896	23896	23896	23896	23896	23896	23896	23896
unique	1138	1136	7	1	144	458	916	474	671
top	11X083	PS 212	All Grades	All Students	288	14	0.0	20	33.3
freq	35	55	5544	23896	503	434	346	502	401

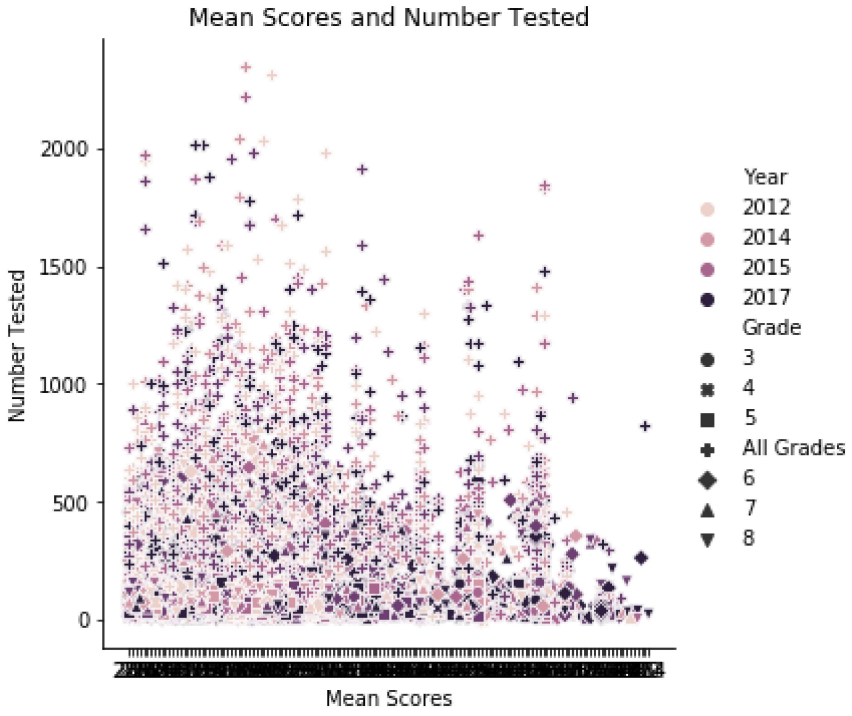
In [6]:
```
dup_filter = scores.duplicated()
scores[dup_filter]
```

Out[6]:

	DBN	School Name	Grade	Year	Category	Number Tested	Mean Scale Score	Level1_N	Level1_%	Level2_N	L

In [7]:
```
sns.relplot(x="Mean Scale Score",y="Number Tested",hue="Year",style="Grade",da
ta=scores)
plt.title("Mean Scores and Number Tested")
plt.xlabel("Mean Scores")
plt.ylabel("Number Tested")
```

Out[7]: Text(34.6663,0.5,'Number Tested')



In [8]:
```
sns.catplot(x = "Year", y = "Number Tested", hue="Grade", kind="bar", data = s
cores)
plt.title("Year and Number Tested")
plt.xlabel("Year")
plt.ylabel("Number Tested")
```

Out[8]: Text(33.1472,0.5,'Number Tested')