In [10]:
```
import matplotlib.pyplot as plt
import numpy as np
import sqlite3
import pandas as pd
%matplotlib inline

# we will use this dataset.
# source: https://www.kaggle.com/hugomathien/soccer

def get_data():
    cnx = sqlite3.connect('/Users/yiwenzhao/Desktop/Edx 1 Python for DS/Wee
    df = pd.read_sql_query("SELECT * FROM Player_Attributes", cnx)
    return df

df = get_data()
```

In [11]:
```
# Let's see what is in our dataset
df.describe()
```

Out[11]:

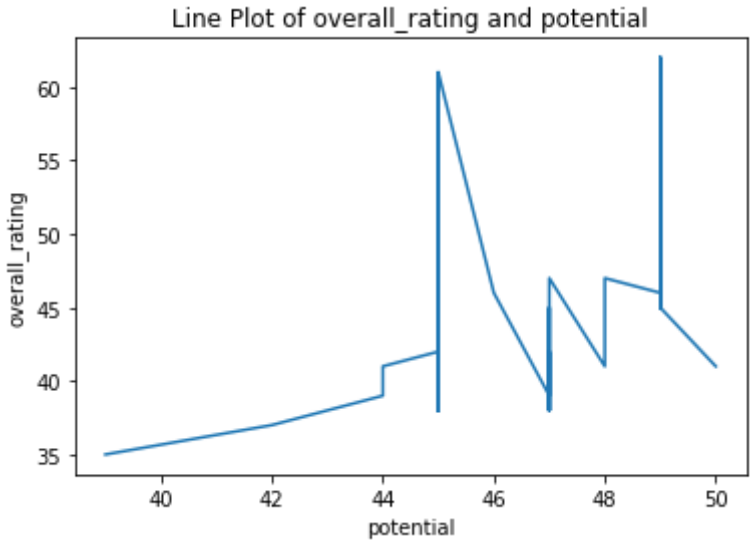
| | id | player_fifa_api_id | player_api_id | overall_rating | potential | crossing |
|---|---|---|---|---|---|---|
| count | 183978.00000 | 183978.000000 | 183978.000000 | 183142.000000 | 183142.000000 | 183142.000000 |
| mean | 91989.50000 | 165671.524291 | 135900.617324 | 68.600015 | 73.460353 | 55.086883 |
| std | 53110.01825 | 53851.094769 | 136927.840510 | 7.041139 | 6.592271 | 17.242135 |
| min | 1.00000 | 2.000000 | 2625.000000 | 33.000000 | 39.000000 | 1.000000 |
| 25% | 45995.25000 | 155798.000000 | 34763.000000 | 64.000000 | 69.000000 | 45.000000 |
| 50% | 91989.50000 | 183488.000000 | 77741.000000 | 69.000000 | 74.000000 | 59.000000 |
| 75% | 137983.75000 | 199848.000000 | 191080.000000 | 73.000000 | 78.000000 | 68.000000 |
| max | 183978.00000 | 234141.000000 | 750584.000000 | 94.000000 | 97.000000 | 95.000000 |

8 rows × 38 columns

In [19]:
```
def line_plot(df, x, y):

    df.sort_values(by = x, inplace = True)
    plt.plot(df[x][:50].values, df[y][:50].values)
    plt.xlabel(x)
    plt.ylabel(y)
    plt.title("Line Plot of %s and %s" % (y, x))

    plt.show()
```

In [20]:

```
line_plot(df, 'potential', 'overall_rating')
```

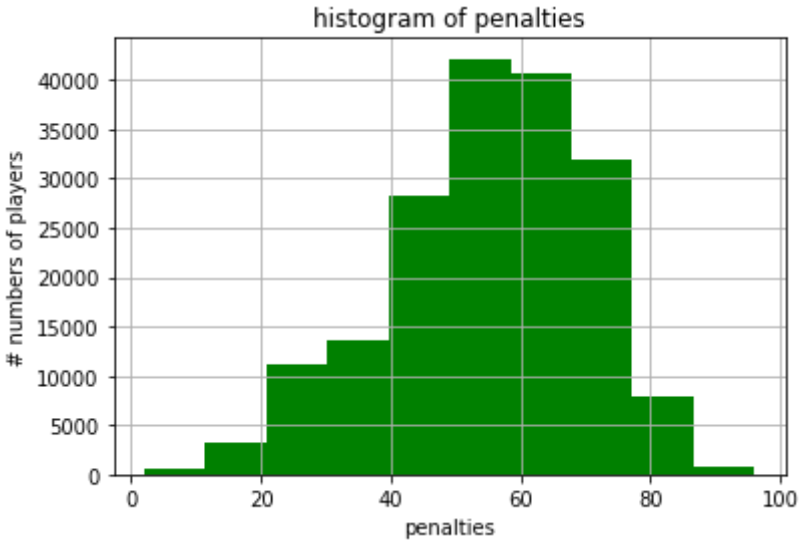



In [44]:

```
def plot_histogram(df, X):

    hist_data = df.dropna()[X].values
    plt.hist(hist_data, 10,
             normed=False, facecolor = 'green')

    plt.xlabel(X)
    plt.ylabel('# numbers of players')
    plt.title('histogram of %s' % X)

    plt.grid(True)
```

In [45]:
```
plot_histogram(df, 'penalties')
```

```
/Users/yiwenzhao/opt/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:7: MatplotlibDeprecationWarning:
The 'normed' kwarg was deprecated in Matplotlib 2.1 and will be removed in 3.1. Use 'density' instead.
  import sys
```



In [53]:
```
# modify this cell

def plot_scatter(df, x, y):
    ### BEGIN SOLUTION
    ### BEGIN SOLUTION
    fig, axis = plt.subplots()
    # Grid lines, Xticks, Xlabel, Ylabel

    axis.yaxis.grid(True)
    axis.set_title('Scatter Plot between %s and %s' % (x, y) ,fontsize=10)
    axis.set_xlabel(x,fontsize=10)
    axis.set_ylabel(y,fontsize=10)

    hori = df[x]
    vert = df[y]

    axis.scatter(hori, vert)
    plt.show()
    ### END SOLUTION




    ### END SOLUTION
```

In [54]:
```
plot_scatter(df, 'gk_diving', 'gk_handling')
```