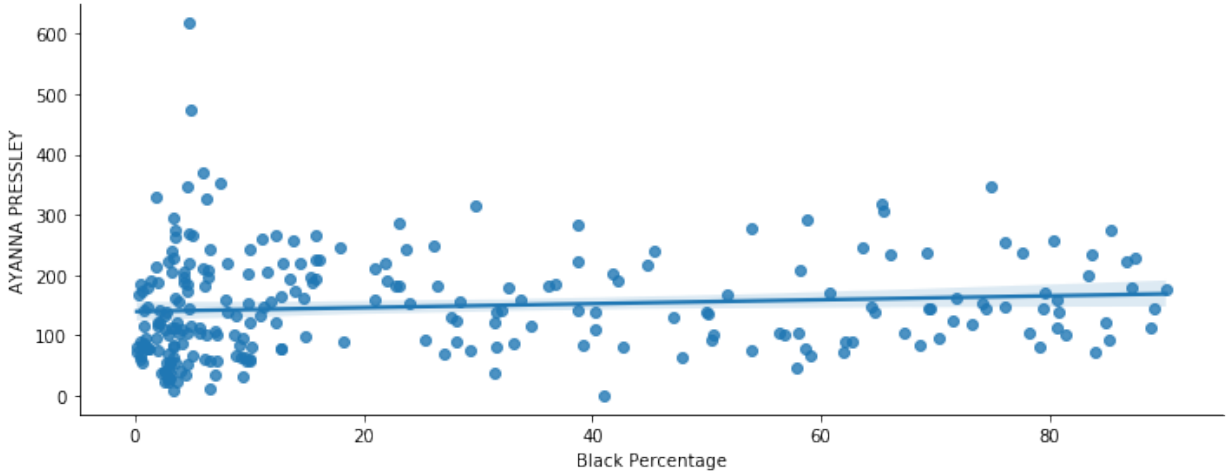
A General Analysis of Black Percentage and Vote Counts

In [38]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

# processing data
def clean_data(filename): # removes any lines with nan values
    df = pd.read_csv(filename)
    #replacing empty entries with nan
    df.replace(r'', np.nan)
    #cleaning data
    cleaned_df = df[~pd.isnull(df).any(axis = 1)]
    return cleaned_df

df = clean_data("2011_CityCouncil_Results_Race_Turnout.csv")
```

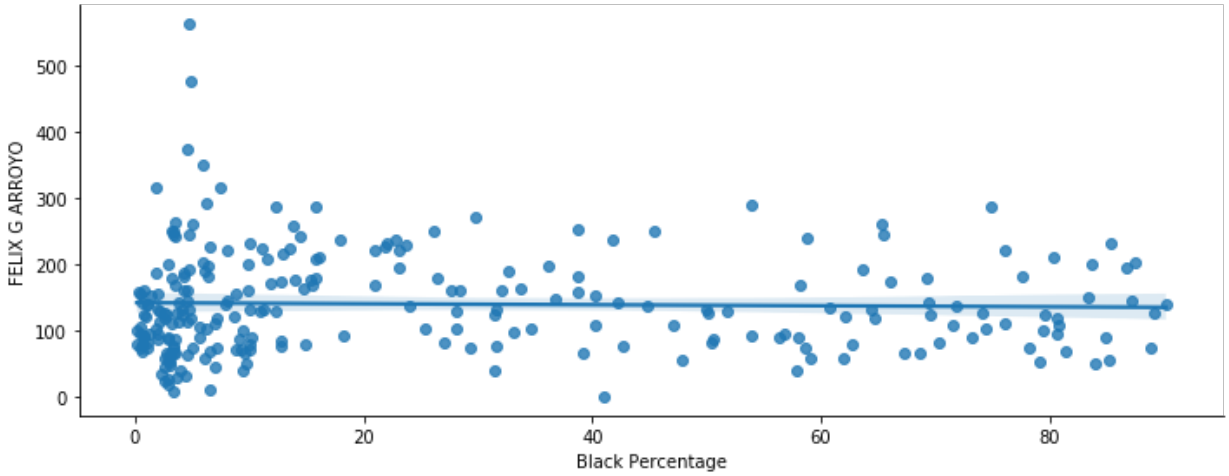
QUESTION1: The correlation of black percentage and candidate's vote count

plotting Black Percentage on the x-axis and each candidate on the y-axis for each ward

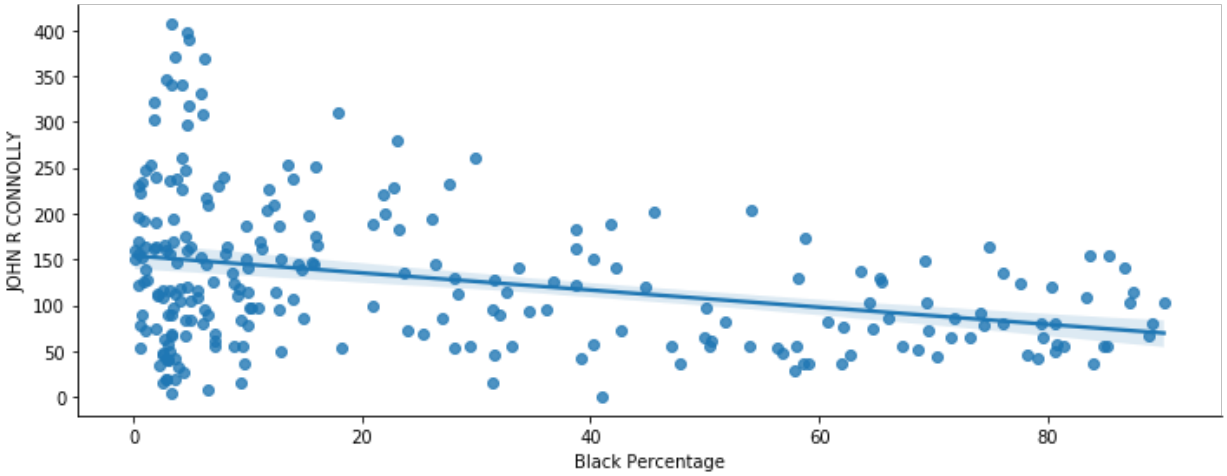
In [39]:
```
# votes for AYANNA PRESSLEY and Black Percentage
sns.lmplot(x="Black Percentage", y="AYANNA PRESSLEY", data=df, height
= 4, aspect = 2.5);
```



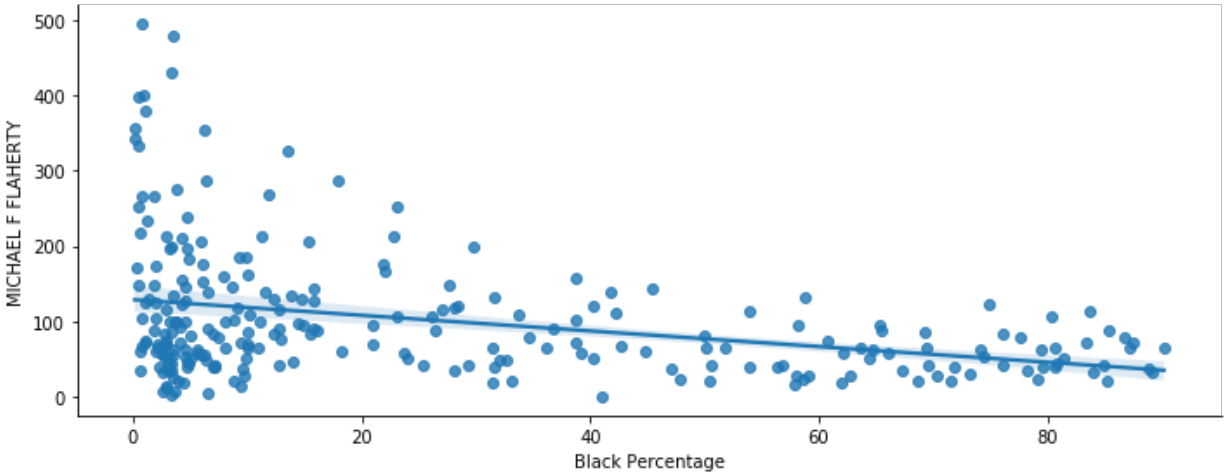
In [40]:
```
# votes for FELIX G ARROYO and Black Percentage
sns.lmplot(x="Black Percentage", y="FELIX G ARROYO", data=df, height =
4, aspect = 2.5);
```



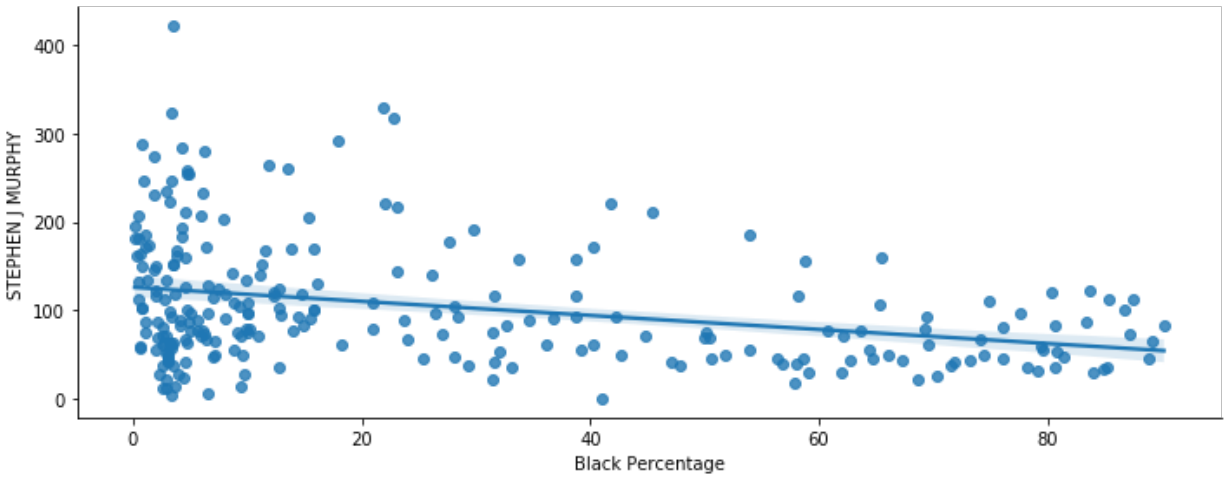
In [41]:
```
# votes for JOHN R CONNOLLY and Black Percentage
sns.lmplot(x="Black Percentage", y="JOHN R CONNOLLY", data=df, height
= 4, aspect = 2.5);
```



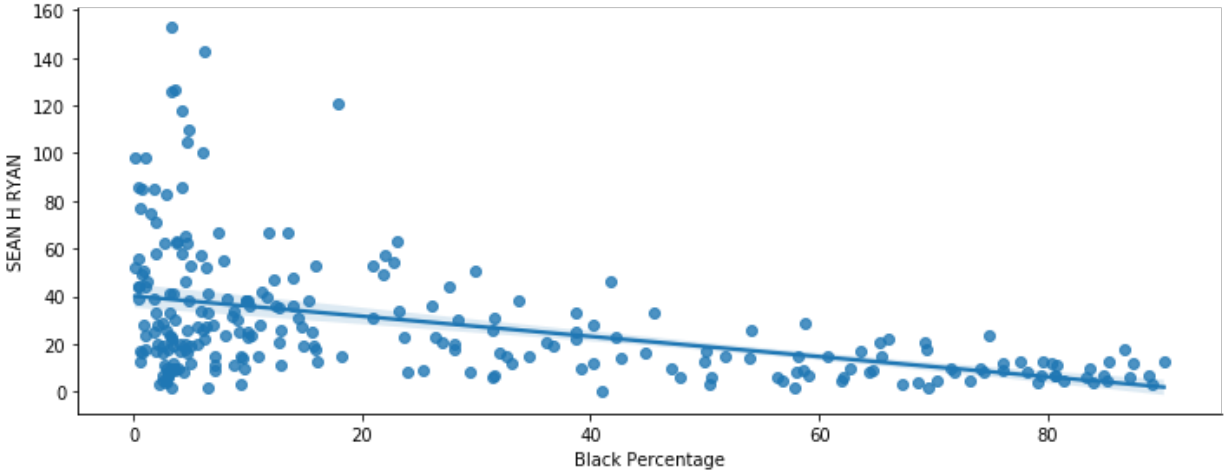
In [42]:
```
# votes for MICHAEL F FLAHERTY and Black Percentage
sns.lmplot(x="Black Percentage", y="MICHAEL F FLAHERTY", data=df, heig
ht = 4, aspect = 2.5);
```



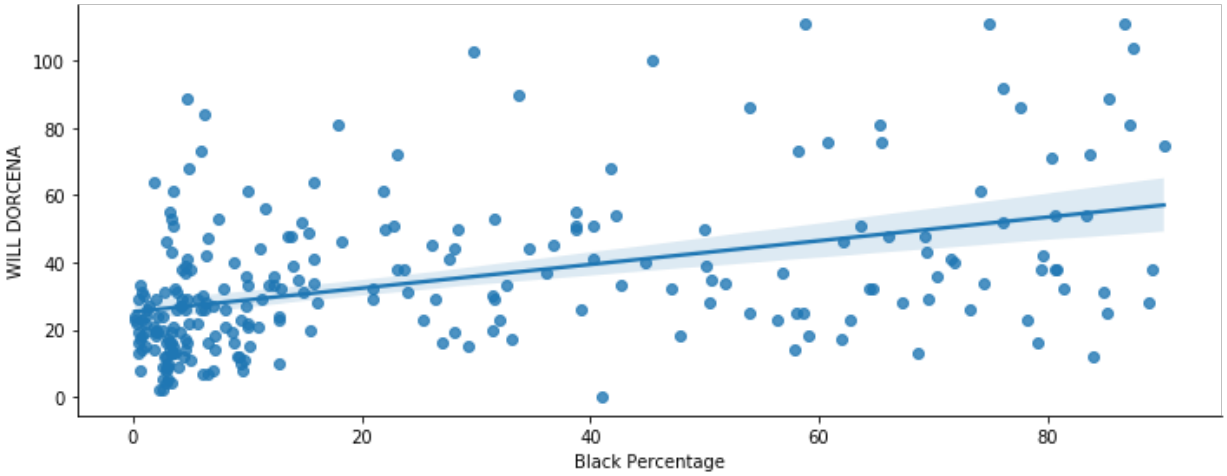
In [43]:
```
# votes for STEPHEN J MURPHY and Black Percentage
sns.lmplot(x="Black Percentage", y="STEPHEN J MURPHY", data=df, height
= 4, aspect = 2.5);
```



In [44]:
```
# votes for SEAN H RYAN and Black Percentage
sns.lmplot(x="Black Percentage", y="SEAN H RYAN", data=df, height = 4,
aspect = 2.5);
```



In [45]:
```
# votes for WILL DORCENA and Black Percentage
sns.lmplot(x="Black Percentage", y="WILL DORCENA", data=df, height = 4
, aspect = 2.5);
```



Analysis:

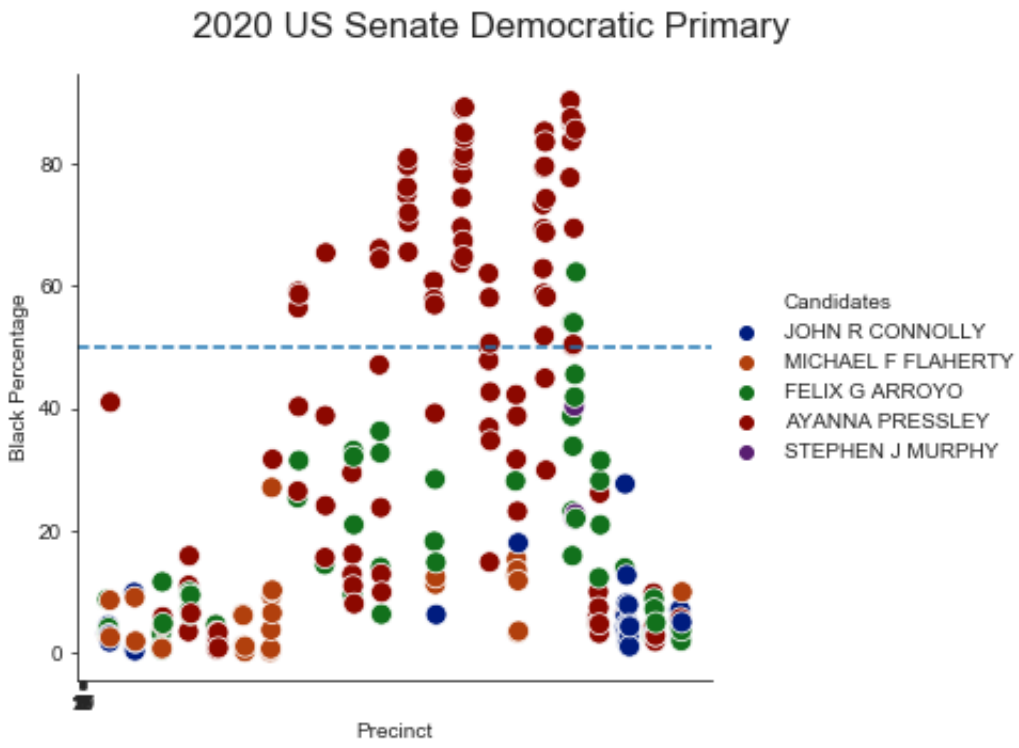
From the plot, we can see that Will Dorcena has the strongest positive correlation between black percentage and vote counts. Ayanna Pressley has a slight positive correlation. JOHN R CONNOLLY, MICHAEL F FLAHERTY and SEAN H RYAN all have negative correlation. According to Wikipedia, Ayanna Pressley is first black woman elected to the Boston City Council. Will Dorcena is a black man and the rest of candidates are white men. Thus, the finding is not surprising and it comfirms the hypothesis that black voters are more likely to vote black candidates.

Question2 The Correlation of Black Percentage and Winner in Each Ward

plotting ward on the x-axis and black percentage on the y-axis and hue is the winner

In [55]:
```
def ward_winner(df,c1,c2,c3,c4,c5):
    outcome = []
    c1_data = df[[c1]].to_numpy()
    c2_data = df[[c2]].to_numpy()
    c3_data = df[[c3]].to_numpy()
    c4_data = df[[c4]].to_numpy()
    c5_data = df[[c5]].to_numpy()
    for i in range(len(c1_data)):
        ltemp = [c1_data[i],c2_data[i],c3_data[i],c4_data[i],c5_data[i
]]
        index = ltemp.index(max(ltemp))
        if index == 0:
            outcome.append(c1)
        elif index == 1:
            outcome.append(c2)
        elif index == 2:
            outcome.append(c3)
        elif index == 3:
            outcome.append(c4)
        else:
            outcome.append(c5)
    return outcome

df['Candidates'] = ward_winner(df, "AYANNA PRESSLEY","FELIX G ARROYO",
"JOHN R CONNOLLY", "MICHAEL F FLAHERTY","STEPHEN J MURPHY")
sns.set_style("ticks")

g = sns.relplot(data=df, x="Precinct", y="Black Percentage", hue="Cand
idates",palette="dark", kind='scatter', s=100)
g.fig.subplots_adjust(top=0.9) # adjust the Figure in g
g.fig.suptitle('2020 US Senate Democratic Primary',fontsize=17)
g.set(xticks=np.arange(1,23,2))
g.axes[0][0].axhline(50, ls='--')
```

Out[55]: <matplotlib.lines.Line2D at 0x7fef1175c910>



Analysis

From the graph we can see that Ayanna Pressley is leading in precincts that has large black percentage.

In [ ]: