A General Analysis of Black Percentage and Vote Counts ¶

In [53]:
```
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

# find the precinct winner
def precinct_winner_2011(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

# find the precinct winner
def precinct_winner_2013(df,c1,c2,c3,c4,c5,c6,c7,c8):
    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()
    c6_data = df[[c6]].to_numpy()
    c7_data = df[[c7]].to_numpy()
    c8_data = df[[c8]].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
],c6_data[i],c7_data[i],c8_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)
        elif index == 4:
            outcome.append(c5)
        elif index == 5:
            outcome.append(c6)
        elif index == 6:
            outcome.append(c7)
        else:
            outcome.append(c8)
    return outcome
```

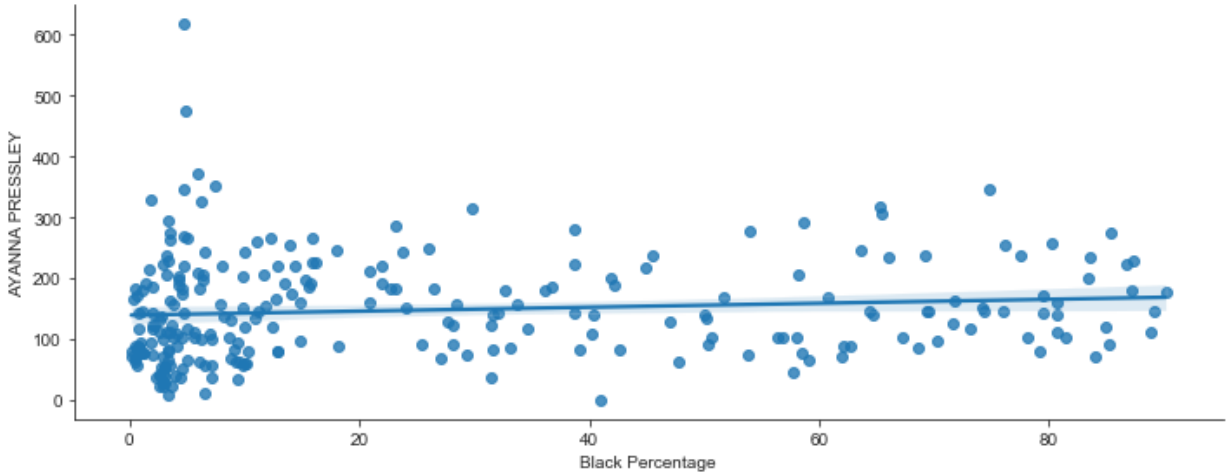
Result Analysis on 2011 Dataset

In [54]:
```
df_2011 = clean_data("2011_CityCouncil_Results_Race_Turnout.csv")
```

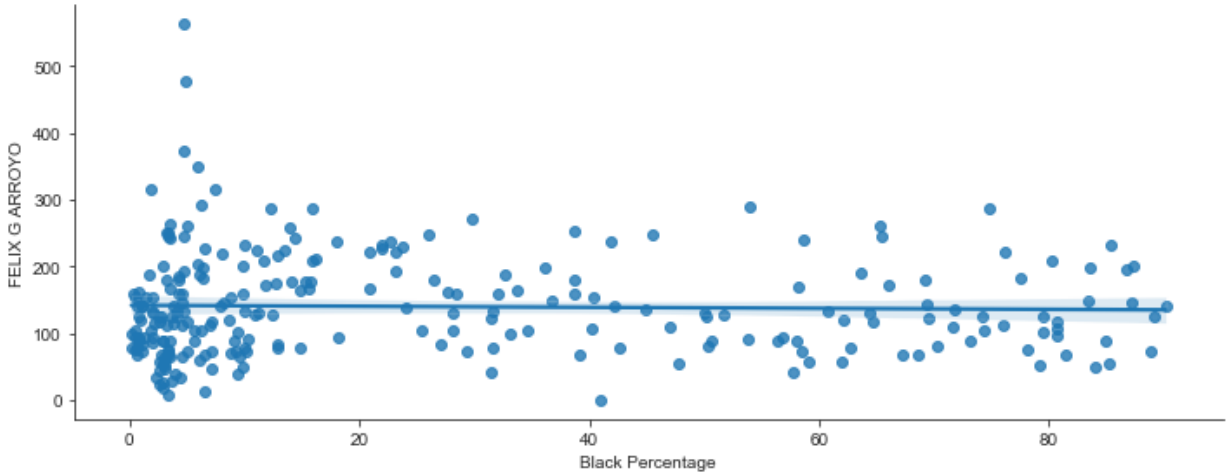
QUESTION1: The correlation of black percentage and candidate's vote count

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

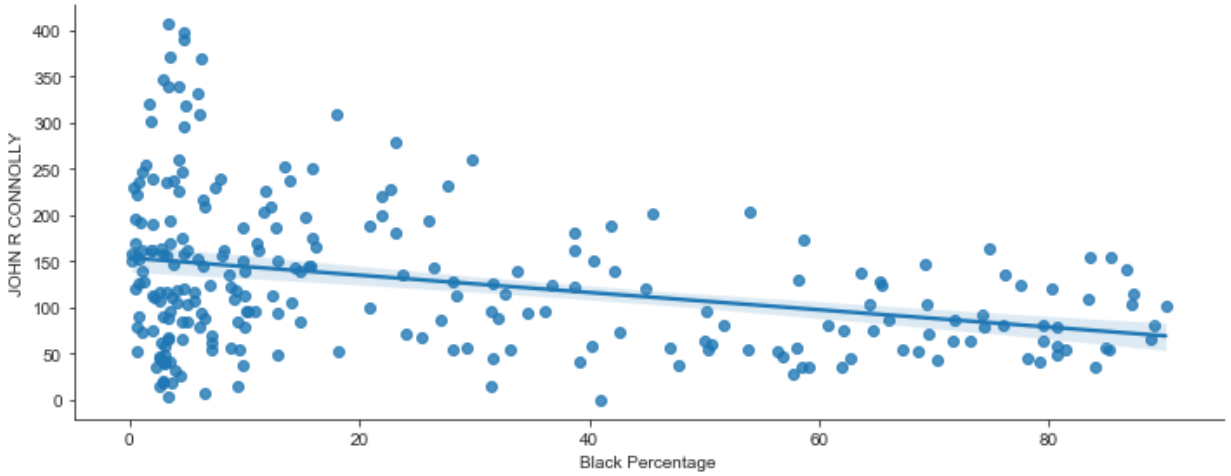
In [55]:
```
# votes for AYANNA PRESSLEY and Black Percentage
sns.lmplot(x="Black Percentage", y="AYANNA PRESSLEY", data=df_2011, he
ight = 4, aspect = 2.5);
```



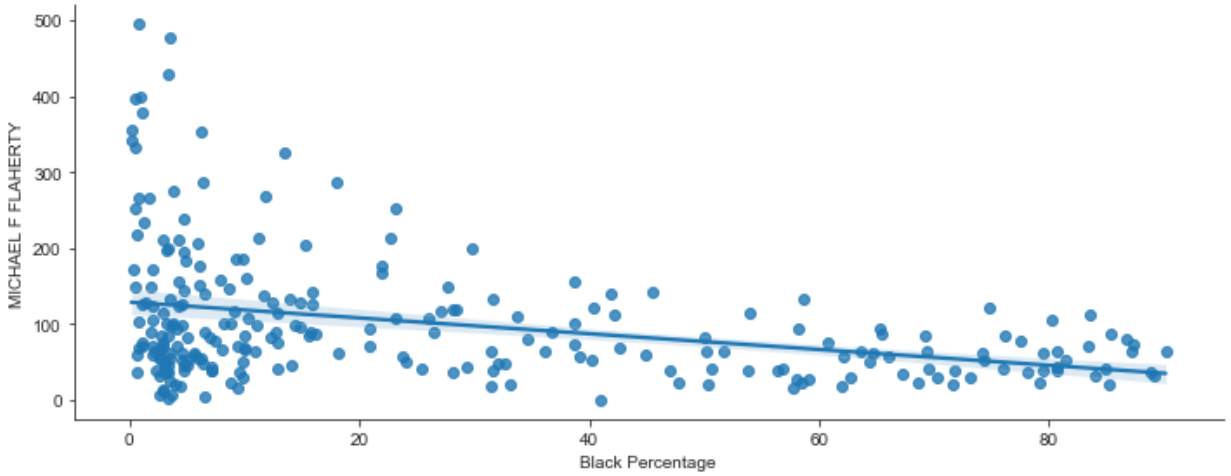
In [56]:
```
# votes for FELIX G ARROYO and Black Percentage
sns.lmplot(x="Black Percentage", y="FELIX G ARROYO", data=df_2011, hei
ght = 4, aspect = 2.5);
```



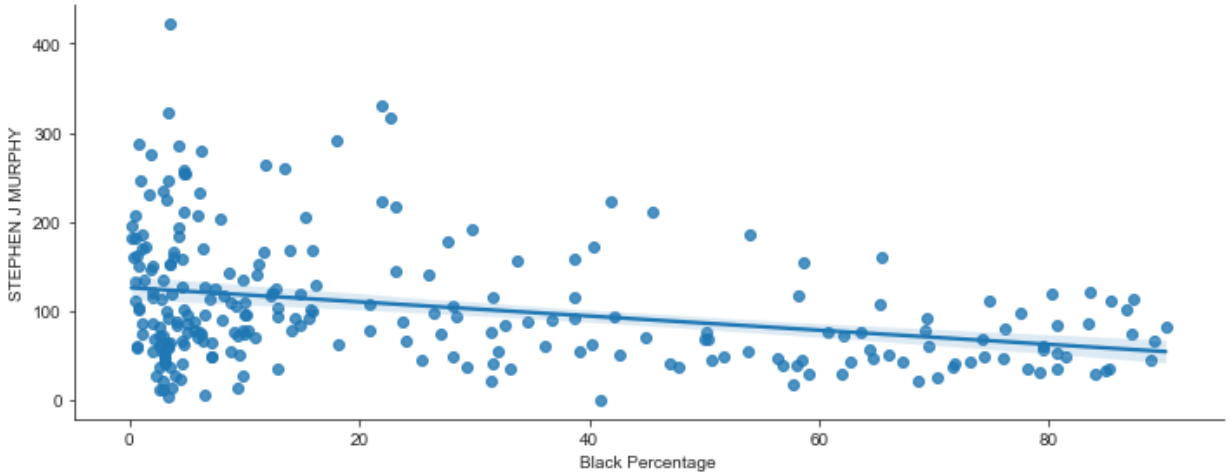
In [57]:
```
# votes for JOHN R CONNOLLY and Black Percentage
sns.lmplot(x="Black Percentage", y="JOHN R CONNOLLY", data=df_2011, he
ight = 4, aspect = 2.5);
```



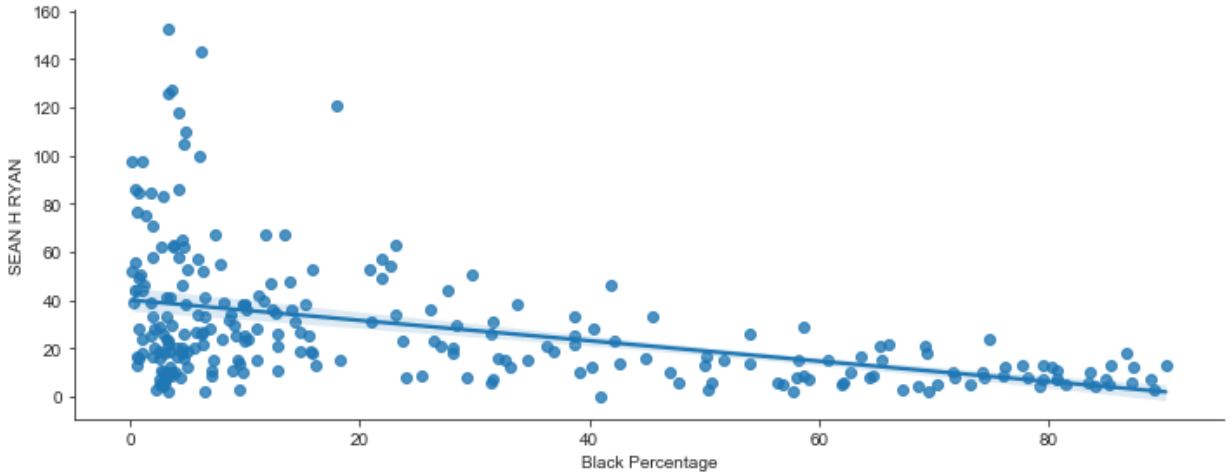
In [58]:
```
# votes for MICHAEL F FLAHERTY and Black Percentage
sns.lmplot(x="Black Percentage", y="MICHAEL F FLAHERTY", data=df_2011,
height = 4, aspect = 2.5);
```



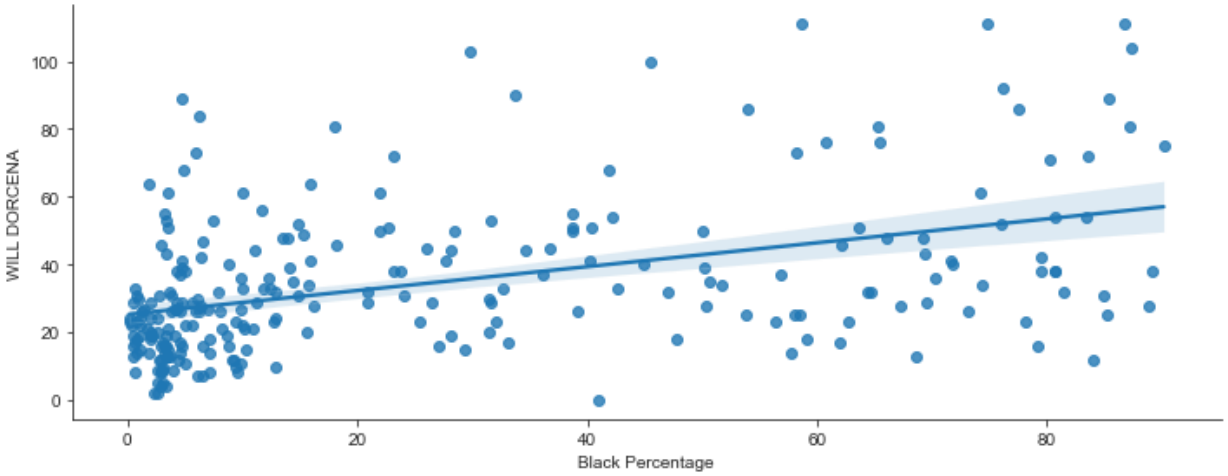
In [59]:
```
# votes for STEPHEN J MURPHY and Black Percentage
sns.lmplot(x="Black Percentage", y="STEPHEN J MURPHY", data=df_2011, h
eight = 4, aspect = 2.5);
```



In [60]:
```
# votes for SEAN H RYAN and Black Percentage
sns.lmplot(x="Black Percentage", y="SEAN H RYAN", data=df_2011, height
= 4, aspect = 2.5);
```



In [61]:
```
# votes for WILL DORCENA and Black Percentage
sns.lmplot(x="Black Percentage", y="WILL DORCENA", data=df_2011, heigh
t = 4, aspect = 2.5);
```



Analysis:

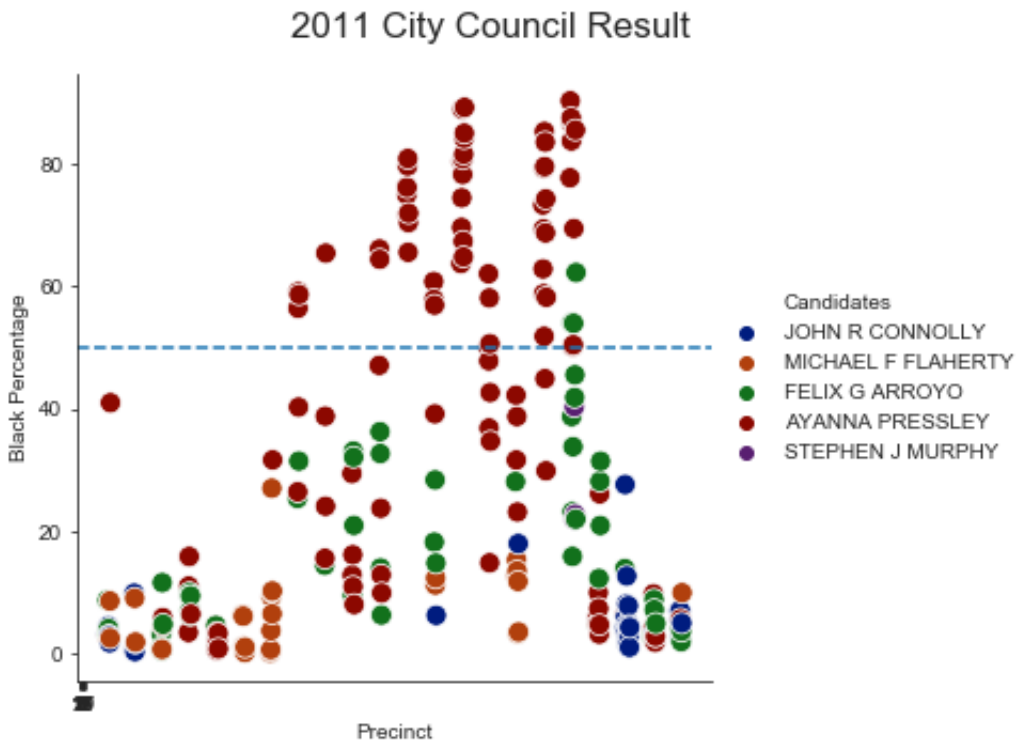
From the plot, we can see that Will Dorcena has the strongest positive correlation between black percentage and vote counts. Ayanna Pressley has a slight positive correlation. JOHN R CONNOLLY, MICHAEL F FLAHERTY and SEAN H RYAN all have negative correlation. According to Wikipedia, Ayanna Pressley is 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 [62]:
```
df_2011['Candidates'] = precinct_winner_2011(df_2011, "AYANNA PRESSLEY
","FELIX G ARROYO","JOHN R CONNOLLY", "MICHAEL F FLAHERTY","STEPHEN J
MURPHY")
sns.set_style("ticks")

g = sns.relplot(data=df_2011, x="Precinct", y="Black Percentage", hue=
"Candidates",palette="dark", kind='scatter', s=100)
g.fig.subplots_adjust(top=0.9) # adjust the Figure in g
g.fig.suptitle('2011 City Council Result',fontsize=17)
g.set(xticks=np.arange(1,23,2))
g.axes[0][0].axhline(50, ls='--')
```

Out[62]: <matplotlib.lines.Line2D at 0x7fe8fda84550>



Analysis

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

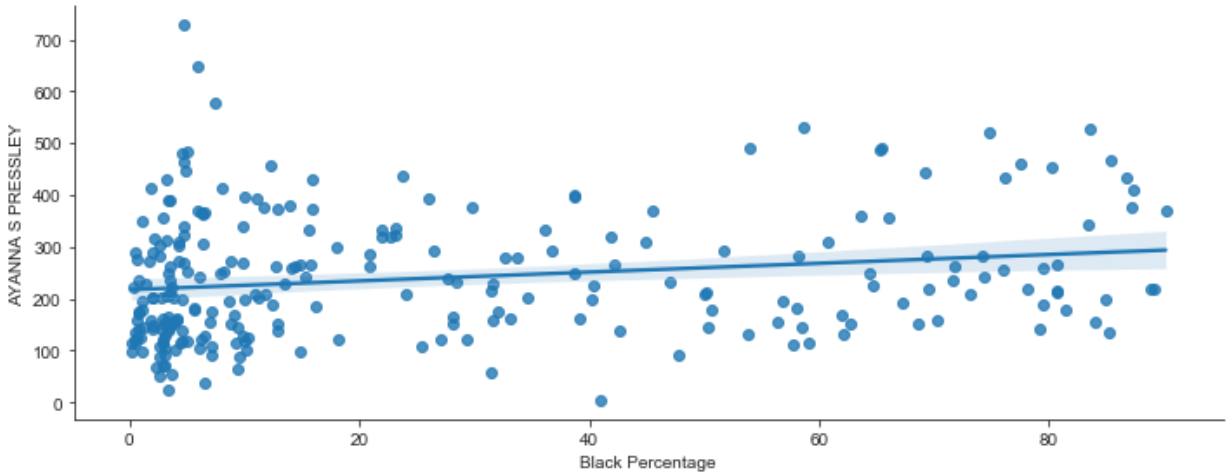
Result Analysis on 2013 Dataset

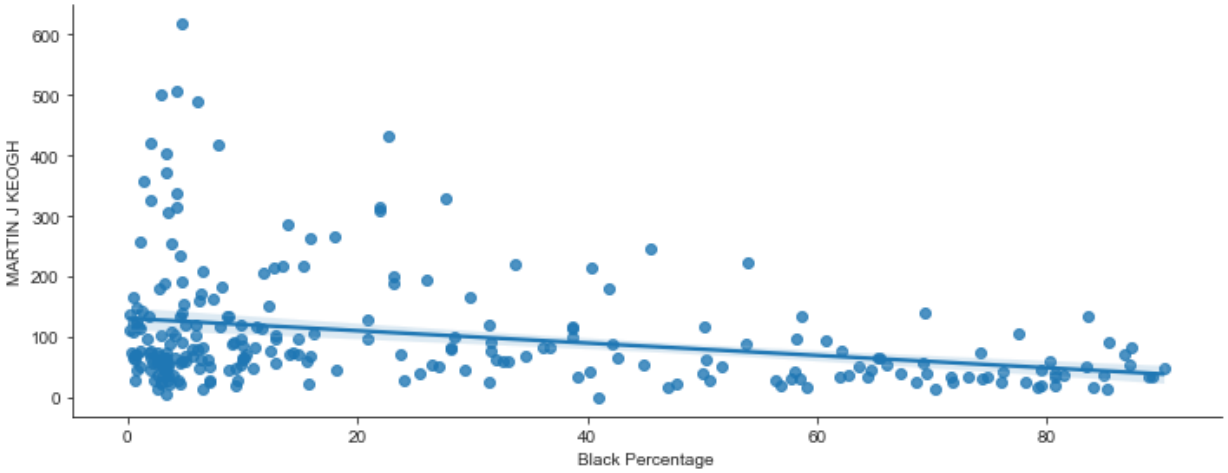
In [63]:
```
df_2013 = clean_data("2013_CityCouncil_Race_Turnout_Results.csv")
```

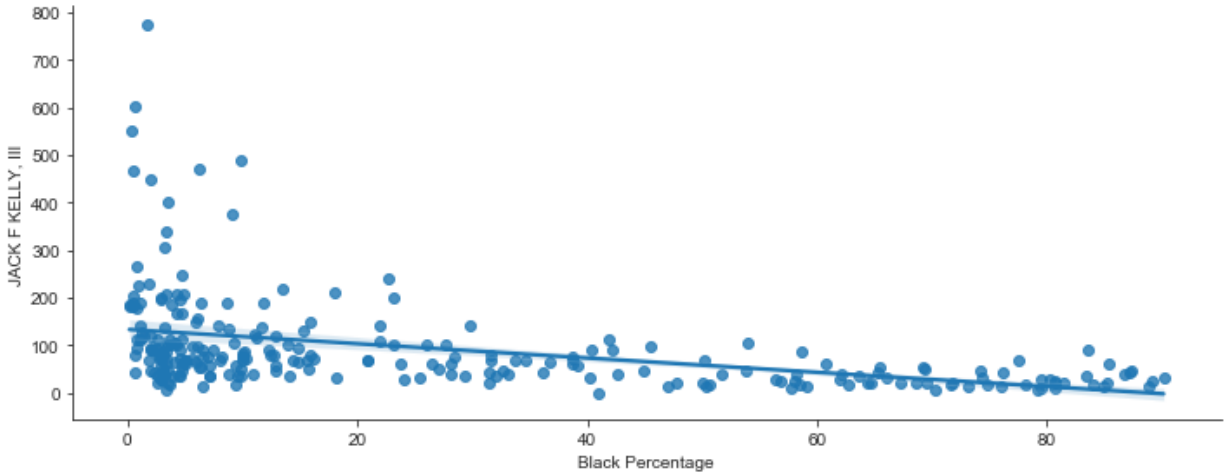
QUESTION1: The correlation of black percentage and candidate's vote count

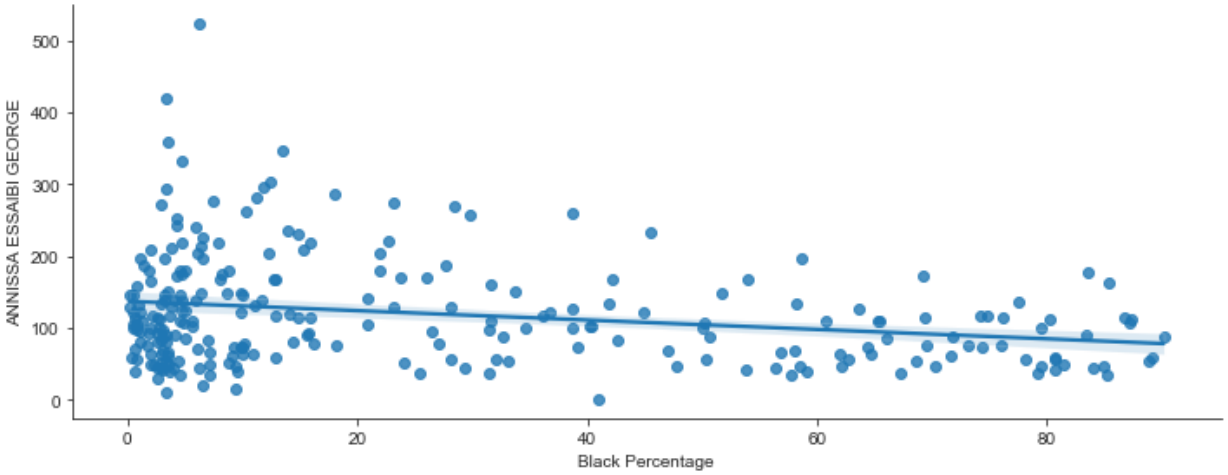
plotting Black Percentage on the x-axis and each candidate on the y-axis for each precinct

In [64]:
```
# votes for AYANNA S PRESSLEY and Black Percentage
sns.lmplot(x="Black Percentage", y="AYANNA S PRESSLEY", data=df_2013,
height = 4, aspect = 2.5);

# votes for MARTIN J KEOGH and Black Percentage
sns.lmplot(x="Black Percentage", y="MARTIN J KEOGH", data=df_2013, hei
ght = 4, aspect = 2.5);

# votes for JACK F KELLY, III and Black Percentage
sns.lmplot(x="Black Percentage", y="JACK F KELLY, III", data=df_2013,
height = 4, aspect = 2.5);

# votes for ANNISSA ESSAIBI GEORGE and Black Percentage
sns.lmplot(x="Black Percentage", y="ANNISSA ESSAIBI GEORGE", data=df_2
013, height = 4, aspect = 2.5);
```

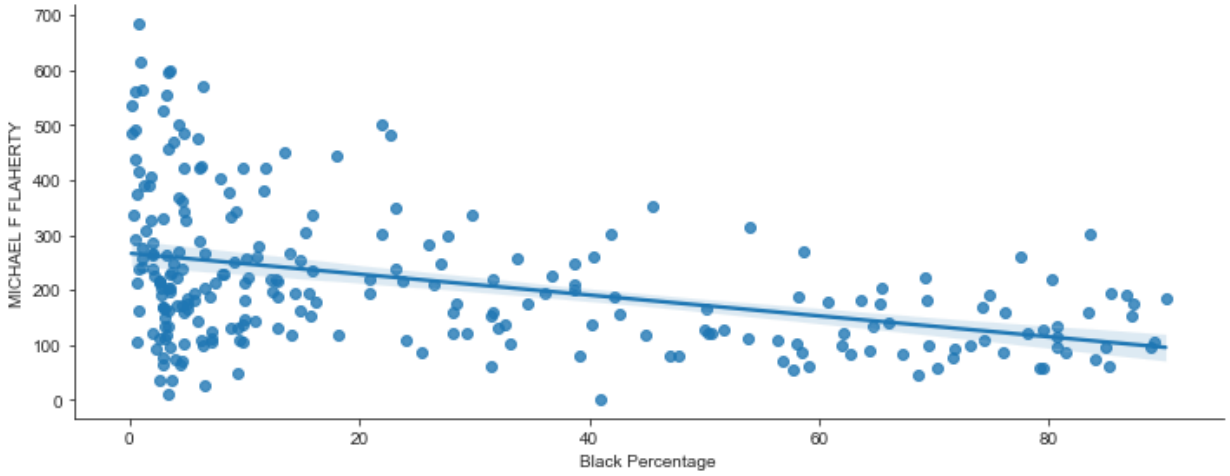


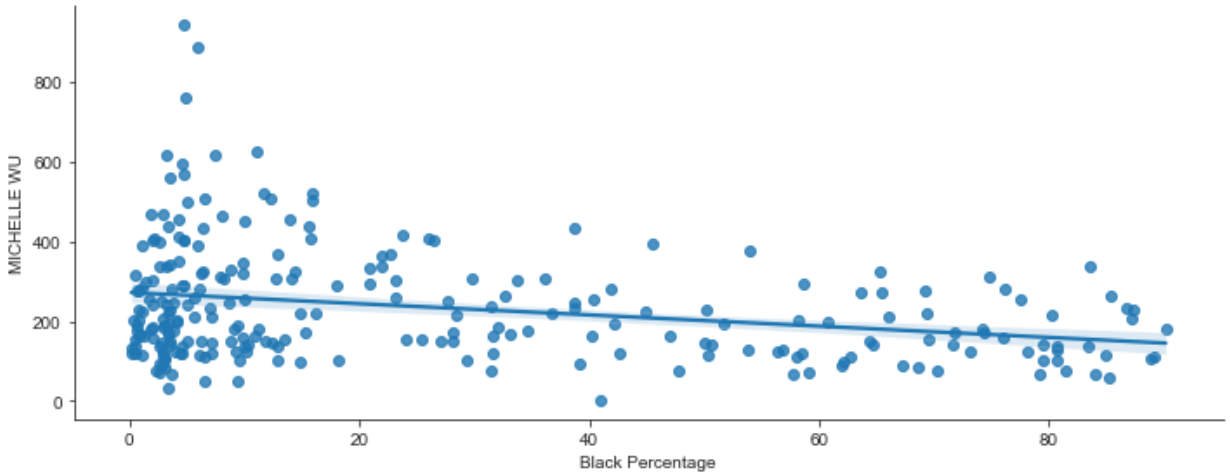

MARTIN J KEOGH
0
100
200
300
400
500
600
0
20
40
60
80
Black Percentage

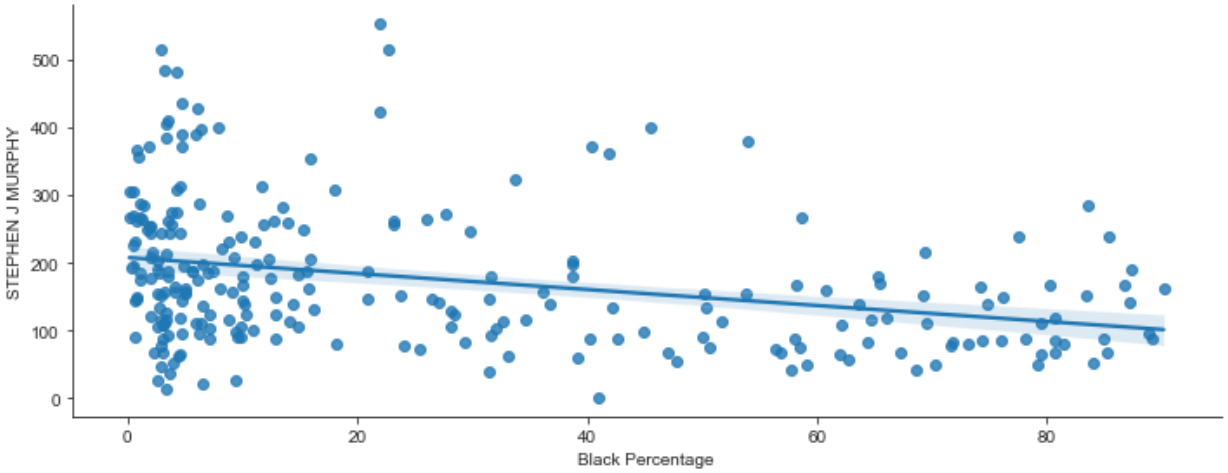

JACK F KELLY, III
0
100
200
300
400
500
600
700
800
0
20
40
60
80
Black Percentage

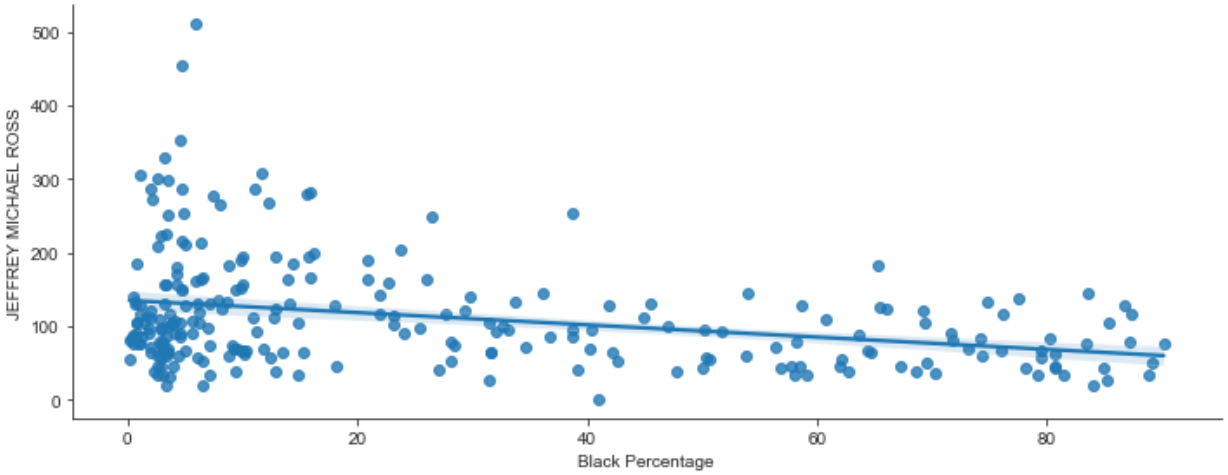

ANNISSA ESSAIBI GEORGE
0
100
200
300
400
500
0
20
40
60
80
Black Percentage

In [65]:
```
# votes for MICHAEL F FLAHERTY and Black Percentage
sns.lmplot(x="Black Percentage", y="MICHAEL F FLAHERTY", data=df_2013,
height = 4, aspect = 2.5);

# votes for MICHELLE WU and Black Percentage
sns.lmplot(x="Black Percentage", y="MICHELLE WU", data=df_2013, height
= 4, aspect = 2.5);

# votes for STEPHEN J MURPHY and Black Percentage
sns.lmplot(x="Black Percentage", y="STEPHEN J MURPHY", data=df_2013, h
eight = 4, aspect = 2.5);

# votes for JEFFREY MICHAEL ROSS and Black Percentage
sns.lmplot(x="Black Percentage", y="JEFFREY MICHAEL ROSS", data=df_201
3, height = 4, aspect = 2.5);
```







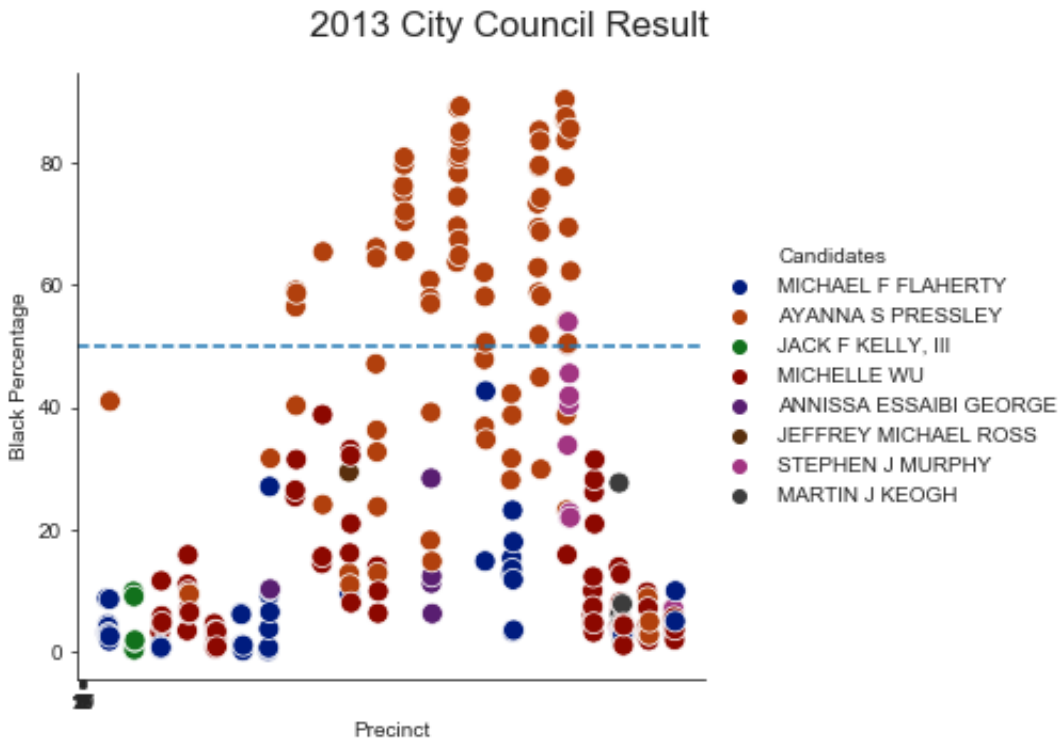


Question2 The Correlation of Black Percentage and Winner in Each Ward

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

In [66]:
```
df_2013['Candidates'] = precinct_winner_2013(df_2013,"AYANNA S PRESSLE
Y","MARTIN J KEOGH","JACK F KELLY, III","ANNISSA ESSAIBI GEORGE","MICH
AEL F FLAHERTY","MICHELLE WU","STEPHEN J MURPHY","JEFFREY MICHAEL ROSS
")
sns.set_style("ticks")

g = sns.relplot(data=df_2013, x="Precinct", y="Black Percentage", hue=
"Candidates",palette="dark", kind='scatter', s=100)
g.fig.subplots_adjust(top=0.9) # adjust the Figure in g
g.fig.suptitle('2013 City Council Result',fontsize=17)
g.set(xticks=np.arange(1,23,2))
g.axes[0][0].axhline(50, ls='--')
```

Out[66]: <matplotlib.lines.Line2D at 0x7fe8ff03cd90>

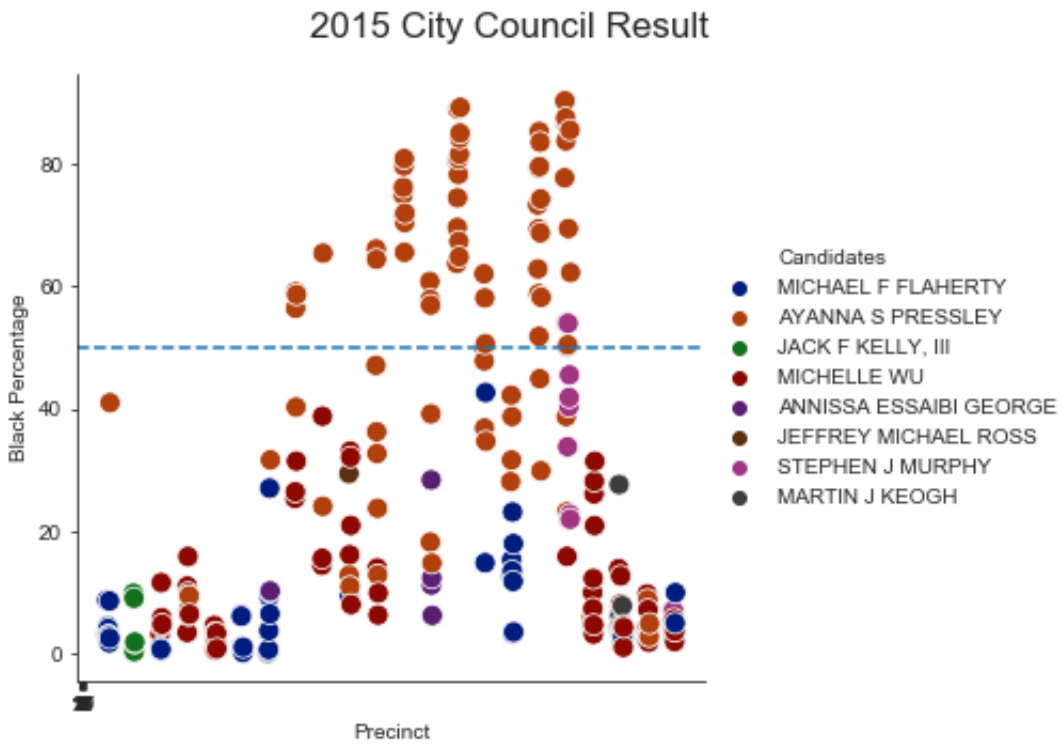


Result Analysis on 2015 Dataset

In [67]:
```
df_2015 = clean_data("2015_city_council.csv")
```

In [68]:
```
df_2015['Candidates'] = precinct_winner_2011(df_2015,"AYANNA S PRESSLE
Y","ANNISSA ESSAIBI GEORGE","MICHAEL F FLAHERTY","MICHELLE WU","STEPHE
N J MURPHY")
sns.set_style("ticks")

g = sns.relplot(data=df_2013, x="Precinct", y="Black Percentage", hue=
"Candidates",palette="dark", kind='scatter', s=100)
g.fig.subplots_adjust(top=0.9) # adjust the Figure in g
g.fig.suptitle('2015 City Council Result',fontsize=17)
g.set(xticks=np.arange(1,23,2))
g.axes[0][0].axhline(50, ls='--')
```

Out[68]: <matplotlib.lines.Line2D at 0x7fe8fe770910>

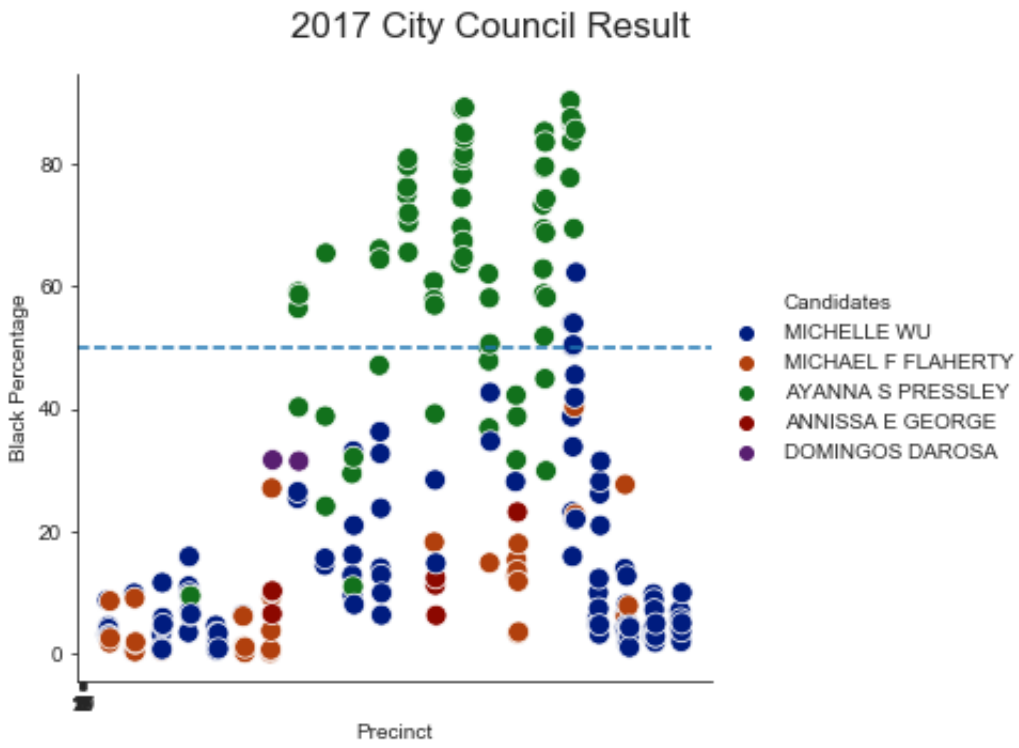


Result Analysis on 2017 Dataset

In [69]:
```
df_2017 = clean_data("2017_CityCouncil_AtLarge_Turnout_Race.csv")
```

In [70]:
```
df_2017['Candidates'] = precinct_winner_2013(df_2017,"AYANNA S PRESSLE
Y","ANNISSA E GEORGE","MICHAEL F FLAHERTY","MICHELLE WU","PAT PAYASO",
"DOMINGOS DAROSA","ALTHEA GARRISON","WILLIAM A KING")
sns.set_style("ticks")

g = sns.relplot(data=df_2017, x="Precinct", y="Black Percentage", hue=
"Candidates",palette="dark", kind='scatter', s=100)
g.fig.subplots_adjust(top=0.9) # adjust the Figure in g
g.fig.suptitle('2017 City Council Result',fontsize=17)
g.set(xticks=np.arange(1,23,2))
g.axes[0][0].axhline(50, ls='--')
```

Out[70]: <matplotlib.lines.Line2D at 0x7fe8fde3bed0>

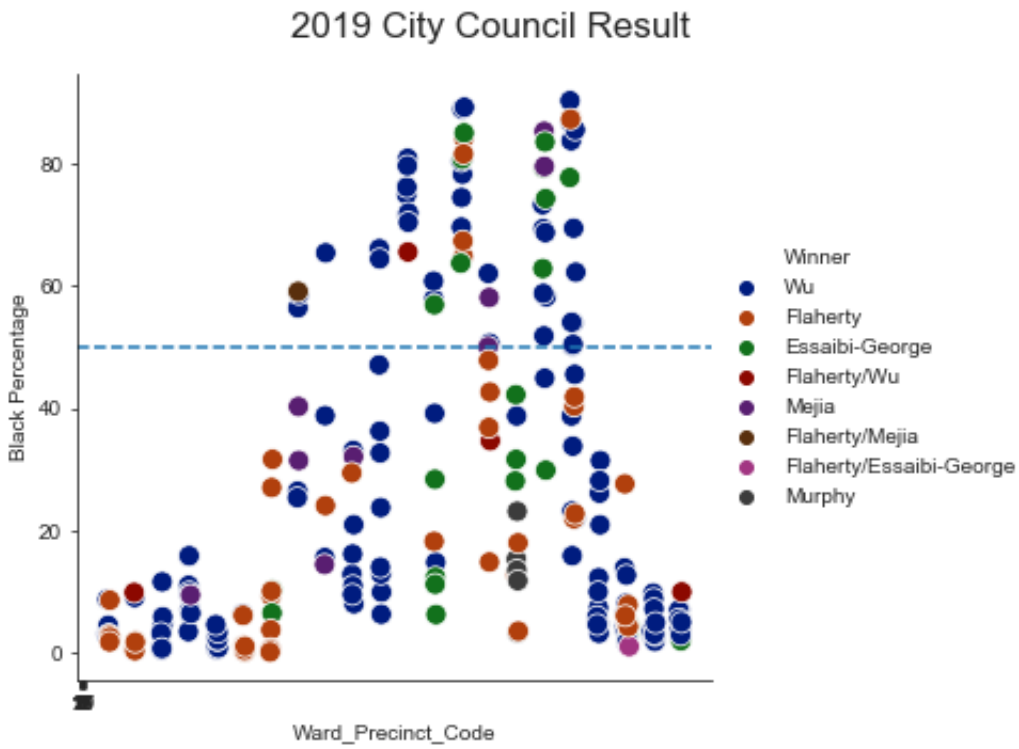


Result Analysis on 2019 Dataset

In [71]:
```
df_2019 = clean_data("2019_CityCouncil_Race Turnout.csv")
```

In [72]:
```
sns.set_style("ticks")

g = sns.relplot(data=df_2019, x="Ward_Precinct_Code", y="Black Percent
age", hue="Winner",palette="dark", kind='scatter', s=100)
g.fig.subplots_adjust(top=0.9) # adjust the Figure in g
g.fig.suptitle('2019 City Council Result',fontsize=17)
g.set(xticks=np.arange(1,23,2))
g.axes[0][0].axhline(50, ls='--')
```

Out[72]: <matplotlib.lines.Line2D at 0x7fe8fde3b890>



In [ ]:
```

```