Question 3b). i). Which historical presidential candidate first name was the most popular in 2020?

ii). What 3 historical presidential candidate first names were tied for the least popular in 2020 according to this presidential_candidates_and_name_popularity table?

Note: Here you'll observe a common problem in data science – one of the least popular names is actually due to the fact that one recent president was so commonly known by his nickname that he appears named as such in the database from which you pulled election results.

In [82]:
```
...
most_popular_firstname = presidential_candidates_and_name_popularity.groupby("First Name").sum
# put your code to calculate the most popular first name above this line, and output the first
most_popular_firstname
```

Out[82]:

First Name	Count
William	151716

In [85]:
```
...
least_popular_firstnames = presidential_candidates_and_name_popularity.groupby("First Name").ma
least_popular_firstnames
```

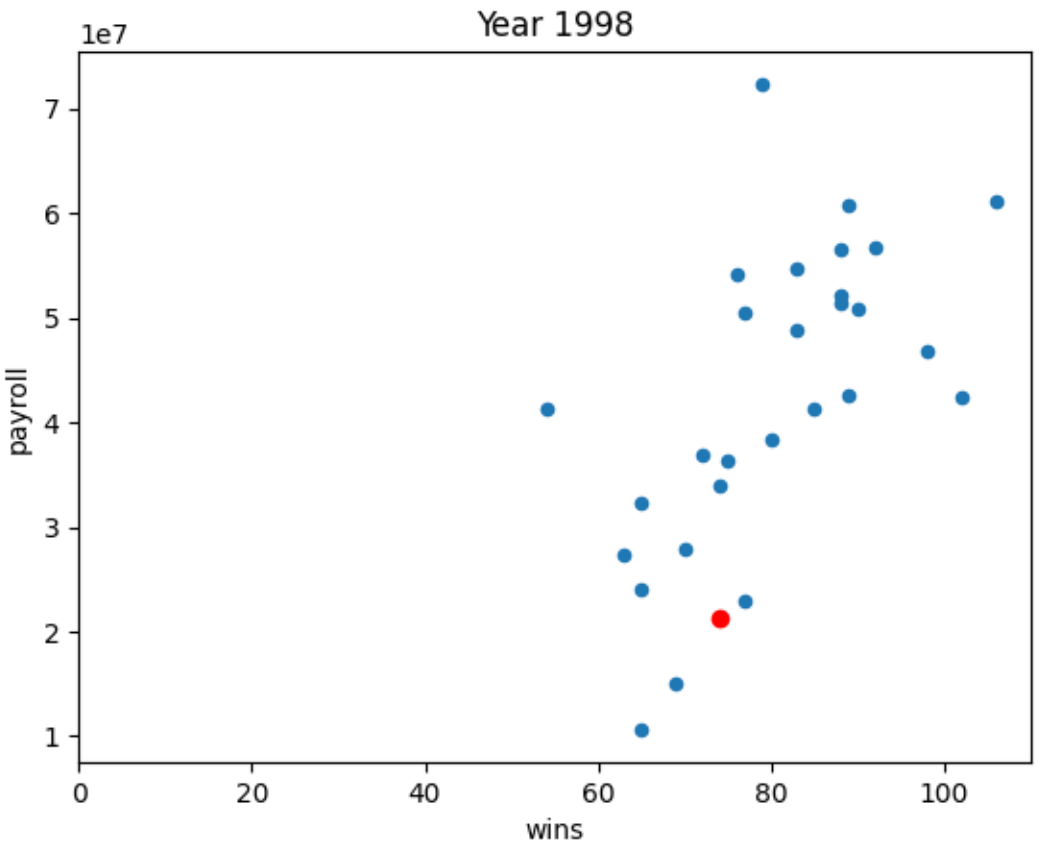
Out[85]:

First Name	Count
Hubert	5
Bill	5
Claude	5

Question 4di). What's up with Oakland? In this problem, you're going to produce scatter plots to confirm the intuition that the data science approach that Oakland adopted changed their efficiency (wins per dollar spent).

Run the code below to produce a scatter plot of the payroll (y-axis) *vs* the number of Wins (x-axis) for all teams during the year 1998, using your dataframe `df_1998`. Notice the code below also highlights the datapoint for Oakland in red.

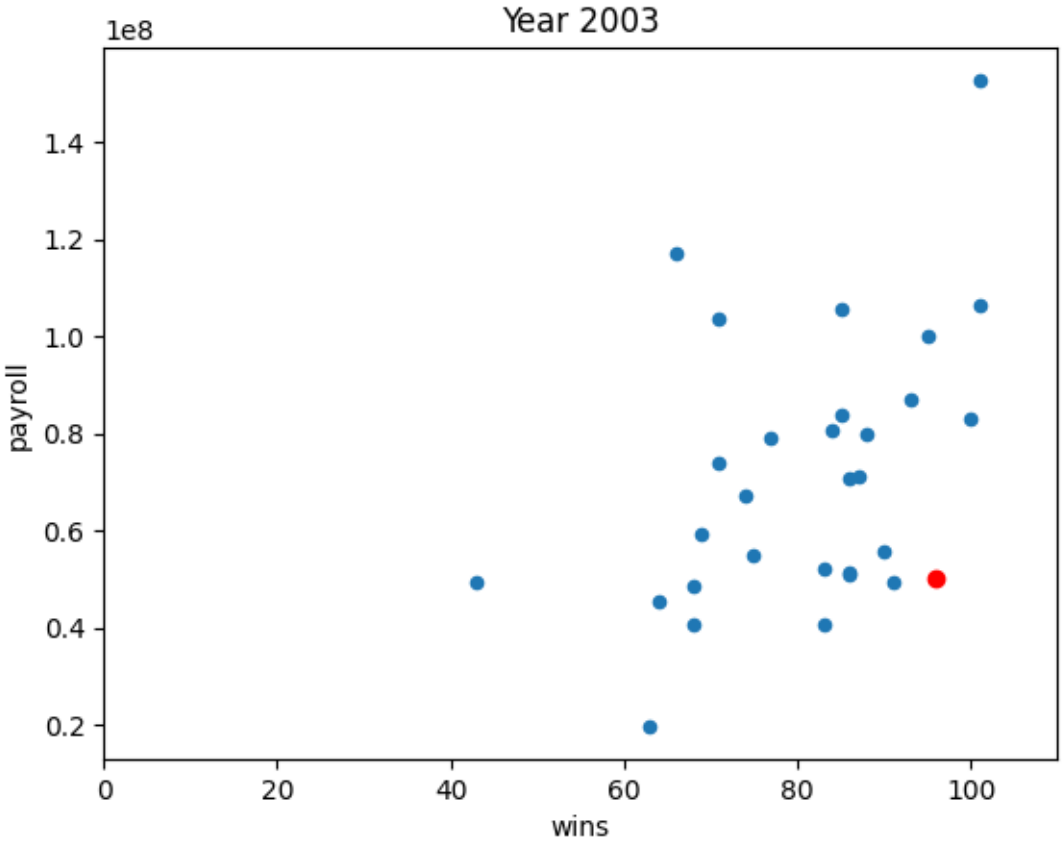
```
In [45]: df_1998.plot.scatter('wins', 'payroll')

         plt.title('Year 1998')
         plt.plot(df_1998.loc['OAK','wins'],df_1998.loc['OAK','payroll'], 'ro')
         plt.xlim(0,110)
         plt.show()
```



0.0.1 Question 4dii).

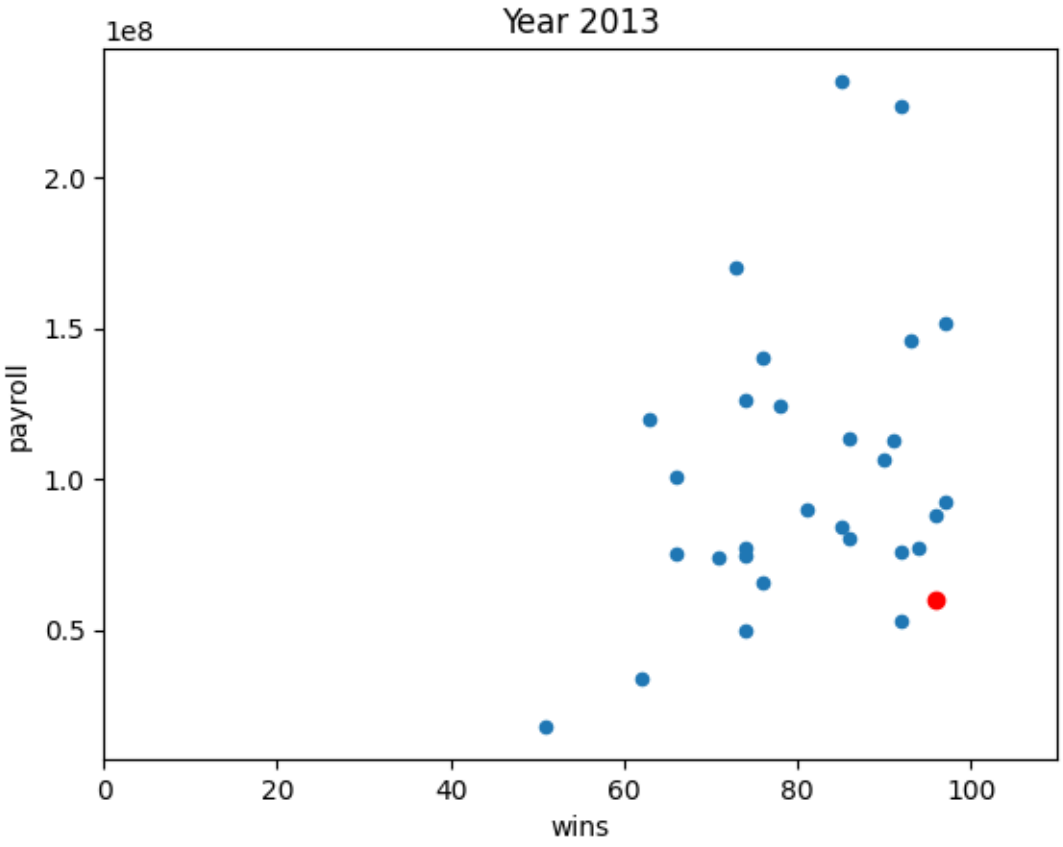
Create two more of these scatterplots (one for 2003 and one for 2013) of wins vs payroll for all teams, and highlight Oakland in red.

```
In [46]: w2003 = T[T["yearID"] == 2003][["teamID","name","W"]].sort_values("teamID",ascending = True).r
         p2003 = S[S["yearID"] == 2003][["teamID","salary"]].groupby("teamID").sum().rename(columns={"s
         df_2003 = pd.merge(left = w2003, right = p2003, left_on = "teamID", right_on = "teamID")
         df_2003.plot.scatter('wins', 'payroll')
         plt.title('Year 2003')
         plt.plot(df_2003.loc['OAK','wins'],df_2003.loc['OAK','payroll'], 'ro')
         plt.xlim(0,110)
         plt.show()
```




```
In [47]: w2013 = T[T["yearID"] == 2013][["teamID","name","W"]].sort_values("teamID",ascending = True).r
         p2013 = S[S["yearID"] == 2013][["teamID","salary"]].groupby("teamID").sum().rename(columns={"s
```

```
df_2013 = pd.merge(left = w2013, right = p2013, left_on = "teamID", right_on = "teamID")
df_2013.plot.scatter('wins', 'payroll')
plt.title('Year 2013')
plt.plot(df_2013.loc['OAK','wins'],df_2013.loc['OAK','payroll'], 'ro')
plt.xlim(0,110)
plt.show()
```



0.0.2 QUESTION 4e).

Examining your scatterplots above, what was the effect of introducing statistics and data science in selecting players for the Oakland A's? (i.e. comment on what trend you notice from the graphs regarding the Oakland A's between 1998, 2003 and 2013).

As the payroll increases, the win rate increases. Although there are some outliers, such as when the payroll is above the 2.0 line, there is less wins compared to ones with 0.5.