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 [30]:
```python
most_popular_firstname = (
    presidential_candidates_and_name_popularity
    .sort_values(by = "Count", ascending = False)
    )
most_popular_firstname = most_popular_firstname.head(1)
most_popular_firstname = most_popular_firstname["First Name"]

# put your code to calculate the most popular first name above this line, and output the first
most_popular_firstname
```

Out[30]:
```
28    William
Name: First Name, dtype: object
```

In [31]:
```python
least_popular_firstnames = (
    presidential_candidates_and_name_popularity
    .sort_values(by = "Count", ascending = True)
    )
least_popular_firstnames = least_popular_firstnames.head(1)
least_popular_firstnames = least_popular_firstnames["First Name"]

# put your code to calculate a series with the least popular names above this line
least_popular_firstnames
```

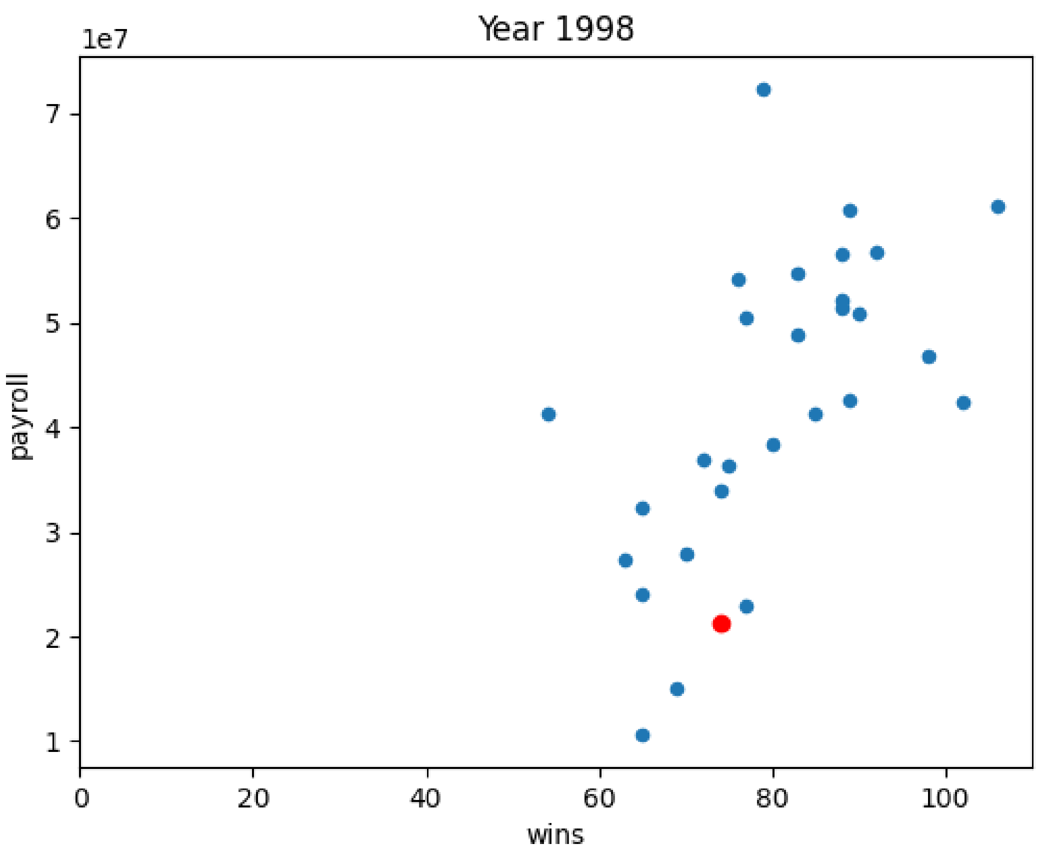
Out[31]:
```
103    Claude
Name: First Name, dtype: object
```

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()
```

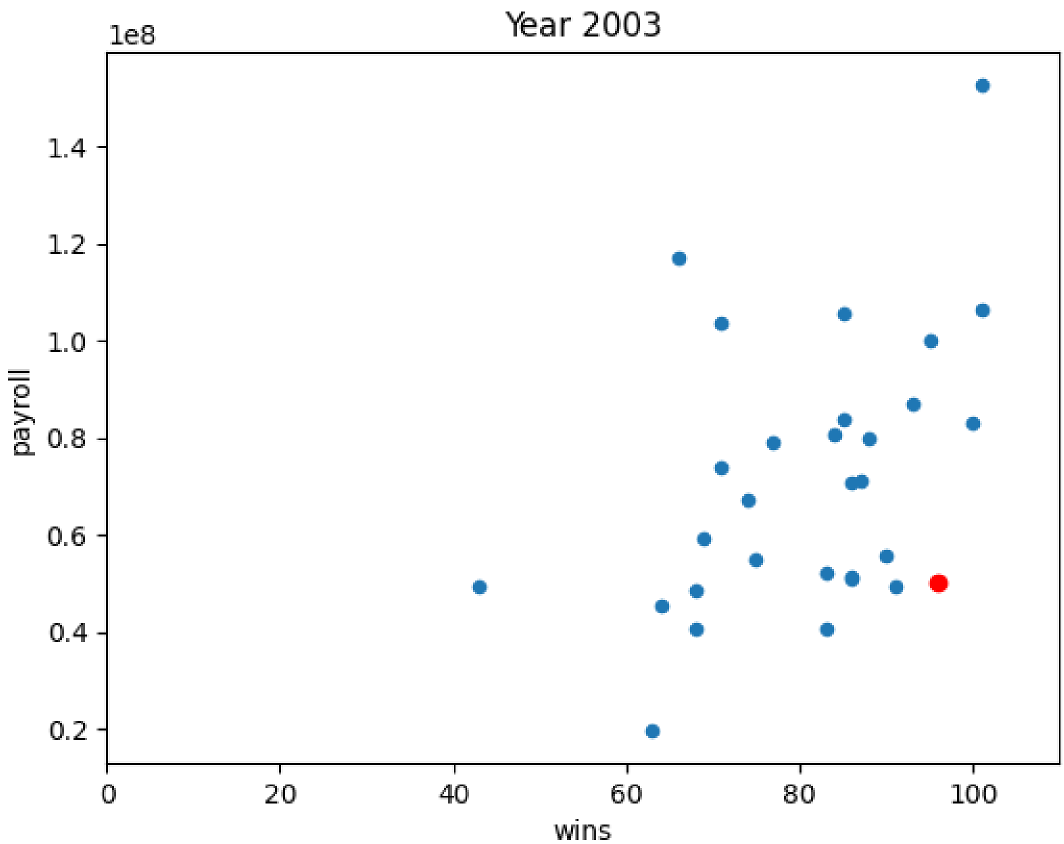


0.0.1 Question 4dii).

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.

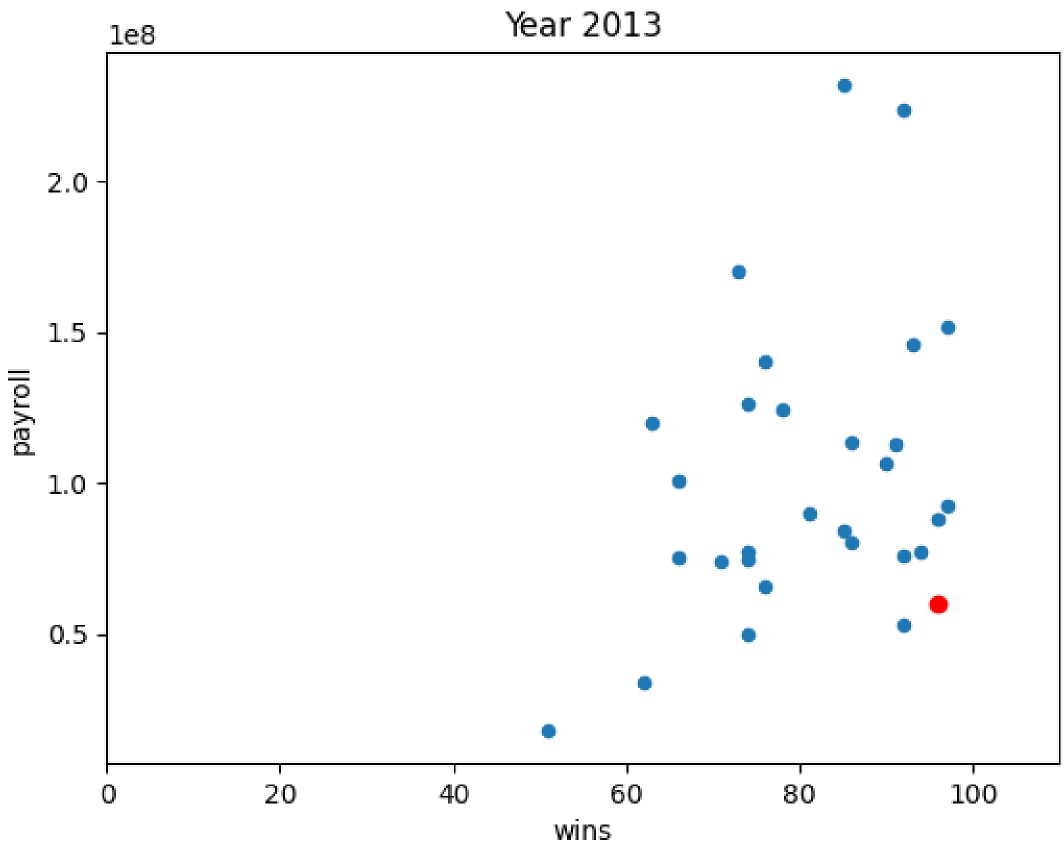
In [46]:
```
mlb = pd.DataFrame(teams_df)
sal = pd.DataFrame(salaries_df)
# your code for 2003 here:
w2003 = (
    mlb.query("yearID == 2003")[["teamID", "name", "W"]]
    .sort_values(by = "teamID", ascending = True)
    .set_index("teamID").rename(columns = {"W": "wins"})
    )
p2003 = (
    sal.query("yearID == 2003")[["teamID", "salary"]]
    .groupby("teamID")["salary"].sum().reset_index()
    .rename(columns = {"salary" : "payroll"}).set_index("teamID")
    )
df_2003 = pd.merge(w2003, 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]: # your code for 2013 here:
         w2013 = (
             mlb.query("yearID == 2013")[["teamID", "name", "W"]]
             .sort_values(by = "teamID", ascending = True)
             .set_index("teamID").rename(columns = {"W": "wins"})
             )
         p2013 = (
             sal.query("yearID == 2013")[["teamID", "salary"]]
             .groupby("teamID")["salary"].sum().reset_index()
             .rename(columns = {"salary" : "payroll"}).set_index("teamID")
             )
         df_2013 = pd.merge(w2013, 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()
```



Year 2013
1e8
2.0
1.5
1.0
0.5
payroll
0
20
40
60
80
100
wins

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).

The Money Ball era that was studied of the Oakland Athletics is a very popular one in baseball. Originally, in 1998, the Oakland Athletics were in the bottom five of payroll and did not win very many games. Upon adopting this new data science strategy, called Money Ball, the general manager Billy Beane hired an analyst away from another team to help get the most 'bang for their buck' with players. This strategy ended up paying dividends for the organization as they maintained the same relative salary cap as they did previously, but won a lot more games.

There is a stark contrast between the wins total of the Oakland Athletics in 1998 and 2003. Upon adopting this new strategy, the Oakland Athletics managed to win a lot more games with close to the same salary cap. This same trend can be observed in 2013, as the A's salary cap hasn't grown very much, but they are still winning a significant number of games.

This strategy was widely adopted by other small market teams, including the Tampa Bay Rays, and it worked for them as well. We can see that the use of data science in this context has a profound affect on a teams performance and can provide a tremendous uptick in production at the major league level.