In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

In [2]:
```
df = pd.read_csv('./factors-affecting-campus-placement/Placement_Data_Full_Class.csv')
```

In [3]:
```
df['salary']=df['salary'].fillna(0)
```

In [4]:
```
df.head()
```

Out[4]:

	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etes
0	1	M	67.00	Others	91.00	Others	Commerce	58.00	Sci&Tech	No	5
1	2	M	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Yes	8
2	3	M	65.00	Central	68.00	Central	Arts	64.00	Comm&Mgmt	No	7
3	4	M	56.00	Central	52.00	Central	Science	52.00	Sci&Tech	No	6
4	5	M	85.80	Central	73.60	Central	Commerce	73.30	Comm&Mgmt	No	5

In [5]:
```
df.describe()
```

Out[5]:

	sl_no	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
count	215.000000	215.000000	215.000000	215.000000	215.000000	215.000000	215.000000
mean	108.000000	67.303395	66.333163	66.370186	72.100558	62.278186	198702.325581
std	62.209324	10.827205	10.897509	7.358743	13.275956	5.833385	154780.926716
min	1.000000	40.890000	37.000000	50.000000	50.000000	51.210000	0.000000
25%	54.500000	60.600000	60.900000	61.000000	60.000000	57.945000	0.000000
50%	108.000000	67.000000	65.000000	66.000000	71.000000	62.000000	240000.000000
75%	161.500000	75.700000	73.000000	72.000000	83.500000	66.255000	282500.000000
max	215.000000	89.400000	97.700000	91.000000	98.000000	77.890000	940000.000000

In [6]:
```
df_score_salary = df.loc[:, ['degree_p','salary']]
df_gender_salary = df.loc[:, ['gender','salary']]
# df_score_salary
# df_gender_salary
```

In [7]:
```
# df_gender_salary['gender'] = df_gender_salary['gender'].map({'F':1,
  'M':0})
# df.loc[:,'gender'] = df.loc[:,'gender'].map({'F': 1, 'M': 0})

sel_men = df_gender_salary.loc[:,'gender'] == 'M'
sel_women = df_gender_salary.loc[:,'gender'] == 'F'
men_salary_mean  = df_gender_salary.loc[sel_men,:].mean()
women_salary_mean = df_gender_salary.loc[sel_women,:].mean()

# men_salary_mean
# 215043.165468

# women_salary_mean
# 168815.789474

# df_gender_salary.replace('M', 0)
# df_gender_salary.replace('F', 1)
# df.loc[:, ['gender', 'salary']]
# select_M = df['gender'] == 'M'
# select_F = df['gender'] == 'F'
# df_M = df[select_M, salary].mean()
# df_F = df[select_F, salary].mean()



# df_M
# df_F

# df.replace()
```

gender and salary relation

according to scatter plot that's shown below, the salary mean of men is more than womens' salary mean. therefore, this sentence 'women get salaries fewer than men' is true.

Note: some salaries is NaN. they are filled with 0.

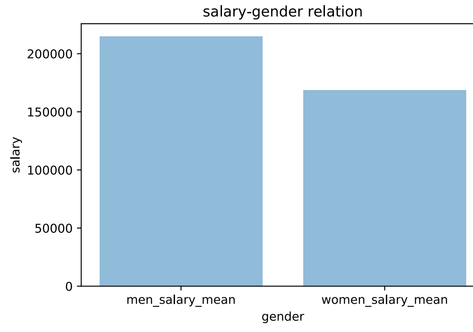
In [8]:
```
# df.plot.scatter(x='', y='salary')

# df.plot.scatter(x='gender', y='salary')

# df_gender_salary.loc[[df_gender_salary.loc[sel_men,:]]
objects = ('men_salary_mean', 'women_salary_mean')
y_pos = np.arange(len(objects))
performance = [float(men_salary_mean), float(women_salary_mean)]

plt.bar(y_pos, performance, align='center', alpha=0.5)
plt.xlabel('gender')
plt.xticks(y_pos, objects)
plt.ylabel('salary')
plt.title('salary-gender relation')

# plt.plot(kind='bar',x=['men_salary_mean', 'women_salary_mean'], y=[men
_salary_mean, women_salary_mean])
plt.show()
# df_gender_salary.plot.scatter(x='gender', y='salary', alpha=1)
# df.plot.scatter(x='gender', y='salary', alpha=0.07)
```



score and salary relation

If we consider degree percentage as a criterion for specialization and salary consequently, then from plot that shown below, it's deduced that people with scores between 65 and 74(approxiamtely) have the most frequency and their salary range varies from 200000 up to 400000. it's clear that high salary is not dependent on high degrees(is's even the most salary is a person with almost 63 degree score). therefore it deduced that 'salary does not depend much on the degree'.

Note: some salaries is NaN. they are filled with 0.

In [13]:
```
df.plot.scatter(x='degree_p', y='salary', alpha=0.5)
```

Out[13]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1e618ecd630>
```