plotting

January 16, 2019

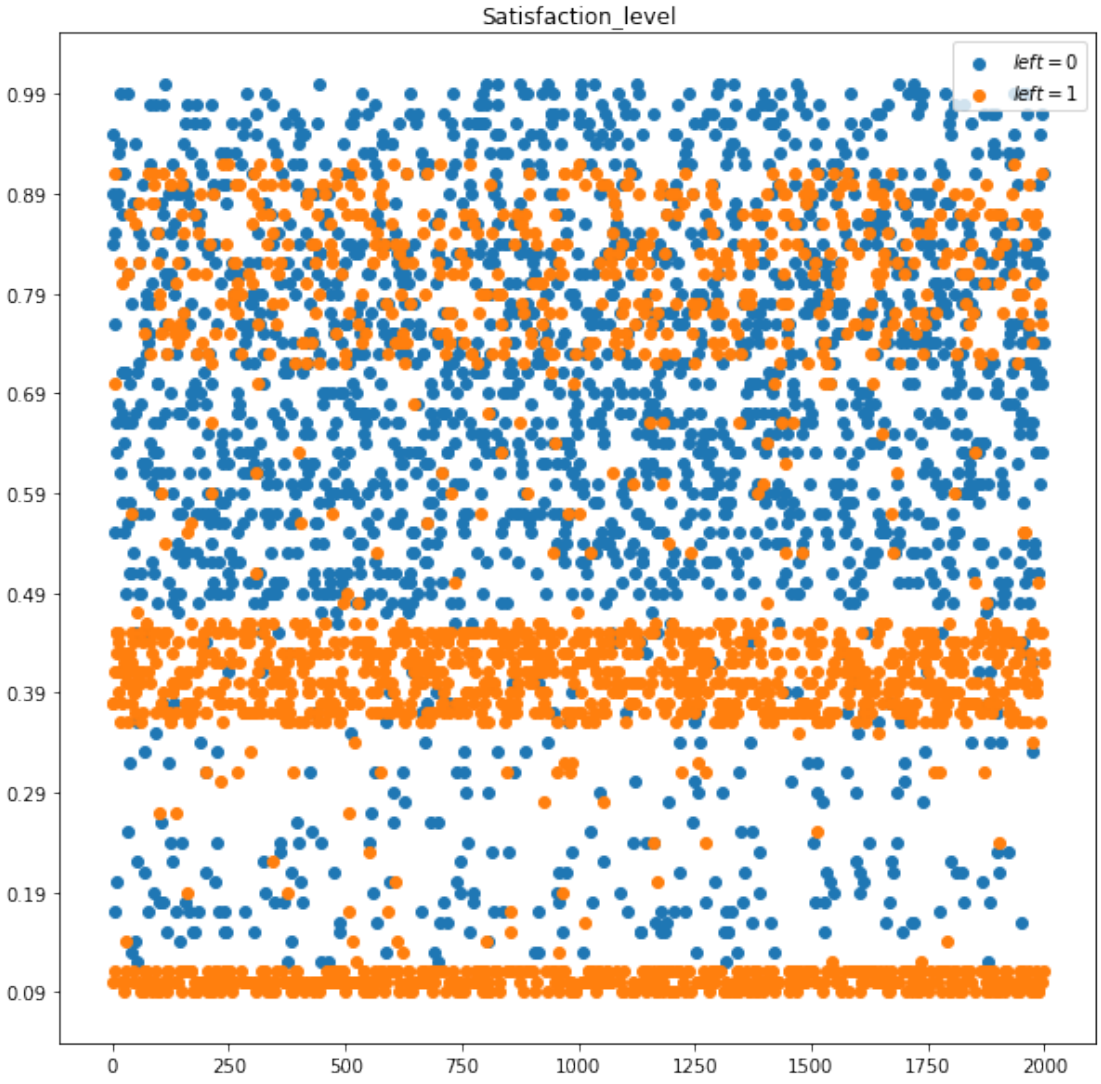
In [6]:
```
import matplotlib
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from pylab import *
```

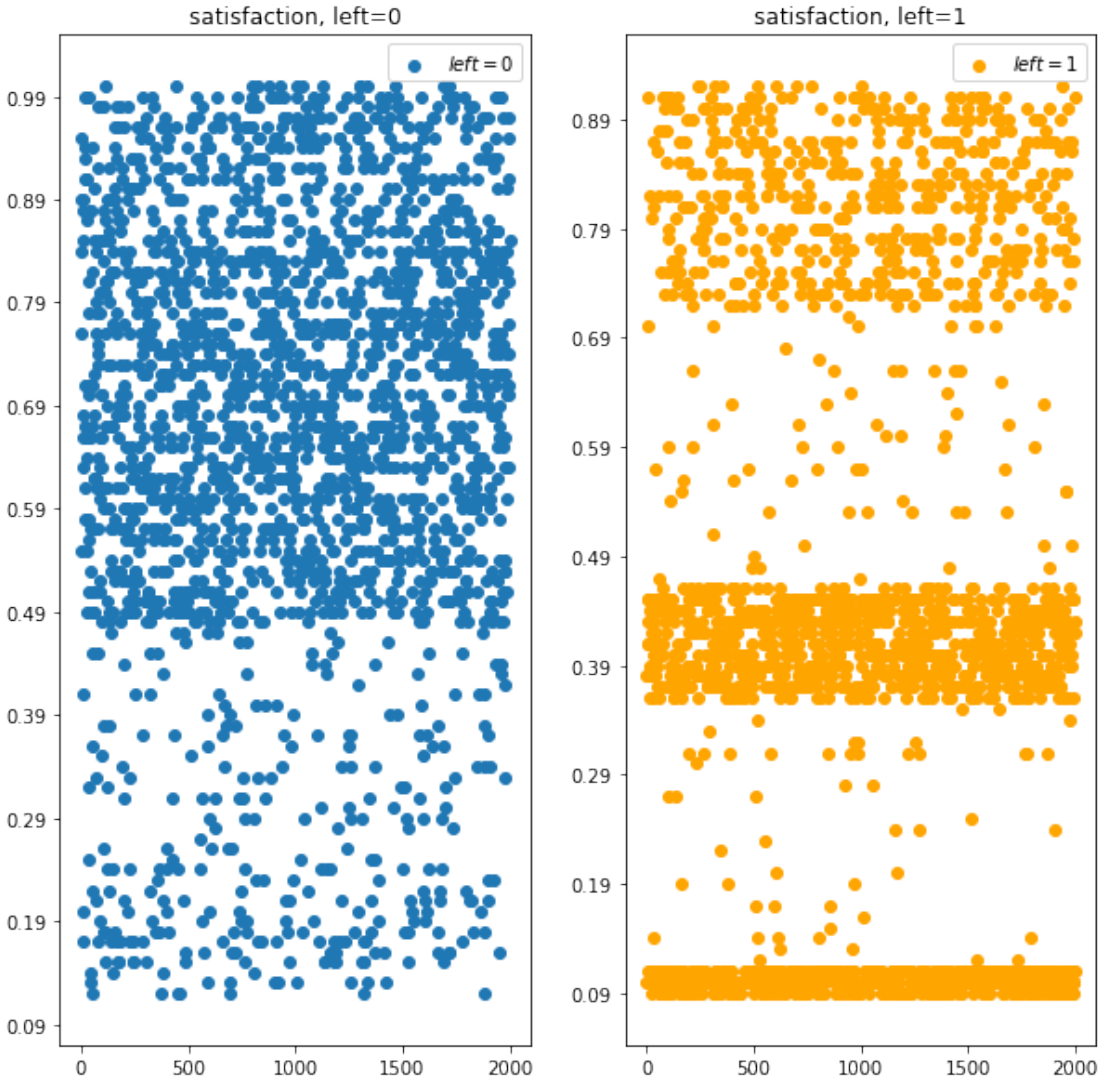
In [80]:
```
filen = "train.csv"
df = pd.read_csv(filen)

#left
left_df = df['left']
# print left_df

left_0 = df[df['left']==0]
left_1 = df[df['left']==1]

work_df0 = left_0['satisfaction_level']
work_df1 = left_1['satisfaction_level']

rows0 = work_df0.shape[0]
print "rows in left=0: ",rows0,
lim=2000
ar0 = work_df0[:lim]
# ar0 = np.split(work_df0, 100)
rows1 = work_df1.shape[0]
print "rows in left=1: ",rows1
ar1 = work_df1[:lim]
print "ar0, ar1 ",ar0.shape[0],ar1.shape[0]

max0=ar0.max()
min0=ar0.min()
max1=ar1.max()
min1=ar1.min()

maxim=max(max1,max0)
minim=min(min0,min1)
print "max, min: ",maxim,minim
```

```
        ar =[]
        i=minim
        while i<maxim:
        #     print i
            ar.append(i)
            i=i+0.1
        print ar
        x = np.linspace(0,lim-1,lim)
        fig, axes = plt.subplots(figsize=(10, 10))
        axes.set_yticks(ar)
        axes.scatter(x,ar0, label=r"$left=0$")
        axes.scatter(x,ar1,label=r"$left=1$")
        legend = axes.legend(loc='best')
        axes.set_title('Satisfaction_level')


        fig2, axes2 = plt.subplots(1, 2, figsize=(10,10))
        axes2[0].set_yticks(ar)
        axes2[1].set_yticks(ar)

        axes2[0].scatter(x,ar0, label=r"$left=0$")
        axes2[1].scatter(x,ar1,label=r"$left=1$",color="orange")
        legend = axes2[0].legend(loc='best')
        legend2 = axes2[1].legend(loc='best')

        axes2[0].set_title('satisfaction, left=0')
        axes2[1].set_title('satisfaction, left=1')
```

```
rows in left=0:  8563 rows in left=1:  2675
ar0, ar1  2000 2000
max, min:  1.0 0.09
[0.09, 0.19, 0.29000000000000004, 0.39, 0.49, 0.59, 0.69, 0.7899999999999999, 0.8899999999999999
```

```
Out[80]: Text(0.5,1,'satisfaction, left=1')
```


Satisfaction_level
left = 0
left = 1
0.99
0.89
0.79
0.69
0.59
0.49
0.39
0.29
0.19
0.09
0
250
500
750
1000
1250
1500
1750
2000



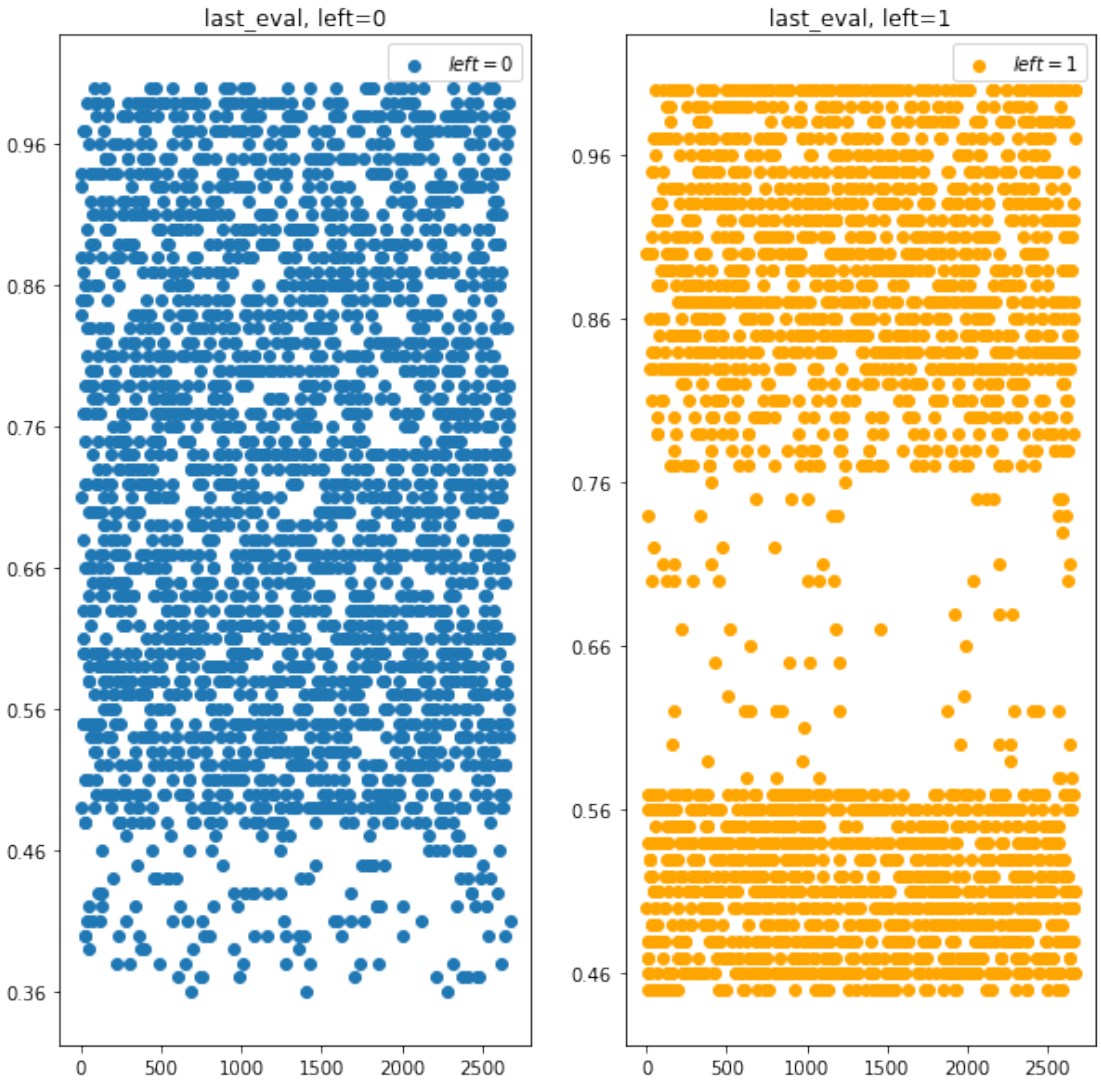
In [79]:
```
filen = "train.csv"
df = pd.read_csv(filen)

#left
left_df = df['left']
# print left_df

left_0 = df[df['left']==0]
left_1 = df[df['left']==1]

work_df0 = left_0['last_evaluation']
work_df1 = left_1['last_evaluation']

rows0 = work_df0.shape[0]
```

```
        print "rows in left=0: ",rows0,
        # ar0 = np.split(work_df0, 100)
        rows1 = work_df1.shape[0]
        print "rows in left=1: ",rows1

        lim=min(rows0-1,rows1-1)
        ar0 = work_df0[:lim]
        ar1 = work_df1[:lim]
        print "ar0, ar1 ",ar0.shape[0],ar1.shape[0]

        max0=ar0.max()
        min0=ar0.min()
        max1=ar1.max()
        min1=ar1.min()

        maxim=max(max1,max0)
        minim=min(min0,min1)
        print "max, min: ",maxim,minim
        ar =[]
        i=minim
        while i<maxim:
        #     print i
            ar.append(i)
            i=i+0.1
        print "ticks: ",ar

        x = np.linspace(0,lim-1,lim)
        fig, axes = plt.subplots(figsize=(10, 10))
        axes.set_yticks(ar)
        axes.scatter(x,ar0, label=r"$left=0$")
        axes.scatter(x,ar1,label=r"$left=1$")
        legend = axes.legend(loc='best')
        axes.set_title('last_evaluated')

        fig2, axes2 = plt.subplots(1, 2, figsize=(10,10))
        axes2[0].set_yticks(ar)
        axes2[1].set_yticks(ar)

        axes2[0].scatter(x,ar0, label=r"$left=0$")
        axes2[1].scatter(x,ar1,label=r"$left=1$",color="orange")
        legend = axes2[0].legend(loc='best')
        legend2 = axes2[1].legend(loc='best')

        axes2[0].set_title('last_eval, left=0')
        axes2[1].set_title('last_eval, left=1')
```

```
rows in left=0:  8563 rows in left=1:  2675
ar0, ar1  2674 2674
```

```
max, min:  1.0 0.36
ticks:  [0.36, 0.45999999999999996, 0.5599999999999999, 0.6599999999999999, 0.7599999999999999,
```

Out[79]: Text(0.5,1,'last_eval, left=1')



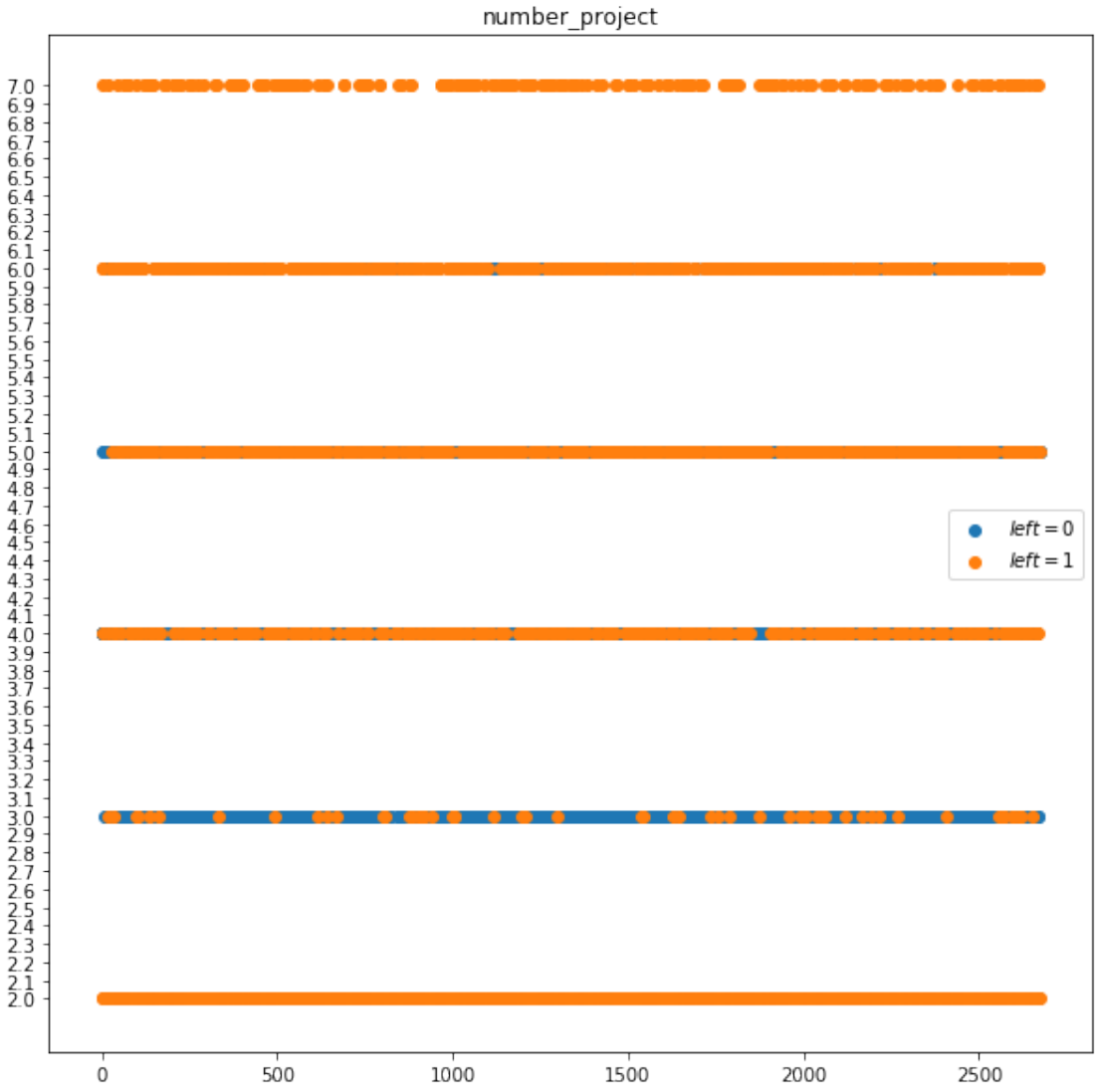
In [78]:
```
filen = "train.csv"
df = pd.read_csv(filen)

#left
left_df = df['left']
# print left_df

left_0 = df[df['left']==0]
left_1 = df[df['left']==1]

work_df0 = left_0['number_project']
work_df1 = left_1['number_project']

rows0 = work_df0.shape[0]
```

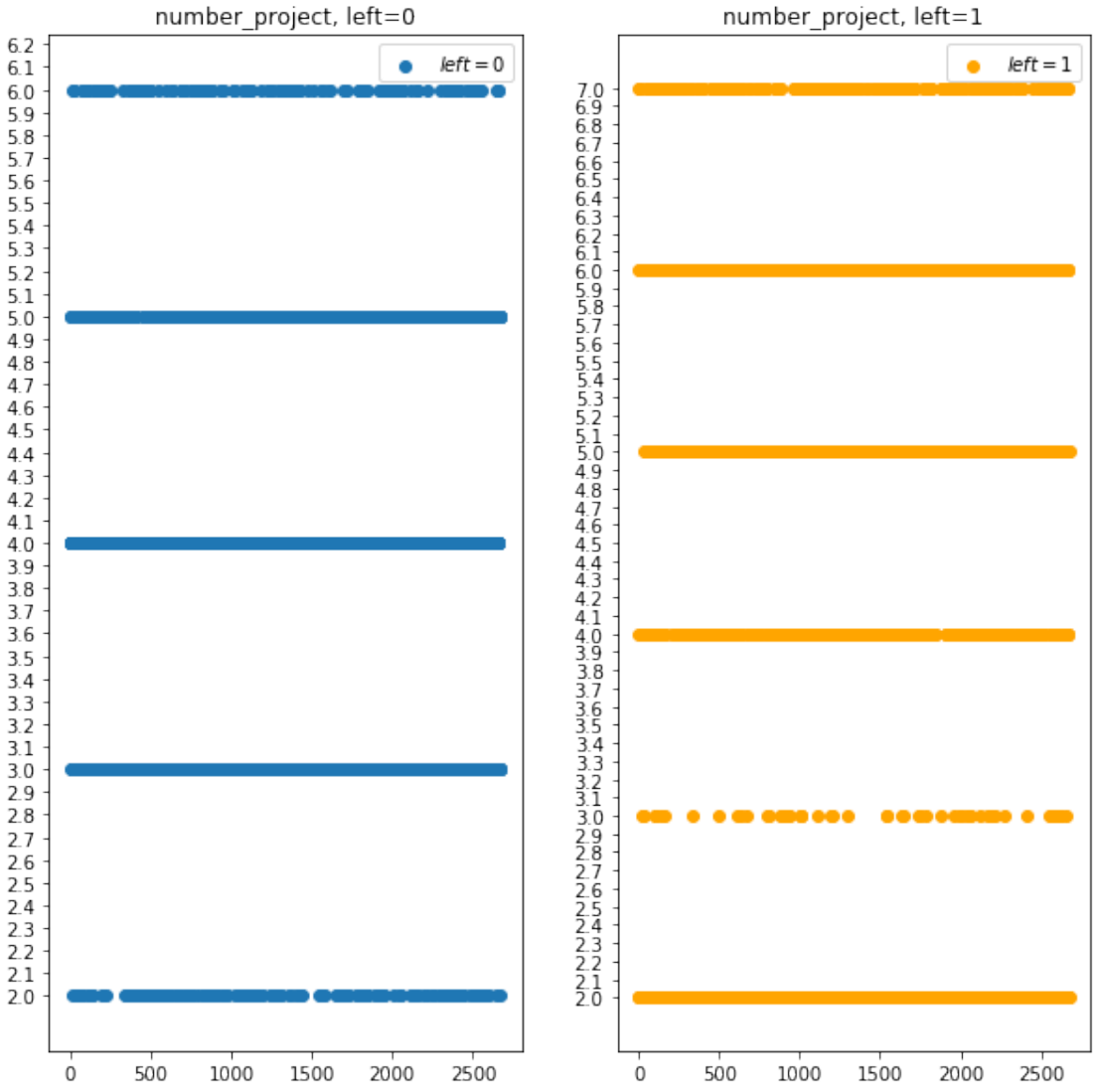
```
        print "rows in left=0: ",rows0,
        # ar0 = np.split(work_df0, 100)
        rows1 = work_df1.shape[0]
        print "rows in left=1: ",rows1

        lim=min(rows0-1,rows1-1)
        ar0 = work_df0[:lim]
        ar1 = work_df1[:lim]
        print "ar0, ar1 ",ar0.shape[0],ar1.shape[0]

        max0=ar0.max()
        min0=ar0.min()
        max1=ar1.max()
        min1=ar1.min()

        maxim=max(max1,max0)
        minim=min(min0,min1)
        print "max, min: ",maxim,minim
        ar =[]
        i=minim
        while i<maxim:
        #     print i
            ar.append(i)
            i=i+0.1
        print "ticks: ",ar

        x = np.linspace(0,lim-1,lim)
        fig, axes = plt.subplots(figsize=(10, 10))
        axes.set_yticks(ar)
        axes.scatter(x,ar0, label=r"$left=0$")
        axes.scatter(x,ar1,label=r"$left=1$")
        legend = axes.legend(loc='best')
        axes.set_title('number_project')

        fig2, axes2 = plt.subplots(1, 2, figsize=(10,10))
        axes2[0].set_yticks(ar)
        axes2[1].set_yticks(ar)

        axes2[0].scatter(x,ar0, label=r"$left=0$")
        axes2[1].scatter(x,ar1,label=r"$left=1$",color="orange")
        legend = axes2[0].legend(loc='best')
        legend2 = axes2[1].legend(loc='best')

        axes2[0].set_title('number_project, left=0')
        axes2[1].set_title('number_project, left=1')
```

```
rows in left=0:  8563 rows in left=1:  2675
ar0, ar1  2674 2674
```

```
max, min:  7 2
ticks:  [2, 2.1, 2.2, 2.3000000000000003, 2.4000000000000004, 2.5000000000000004, 2.600000000000000
```

Out[78]: Text(0.5,1,'number_project, left=1')





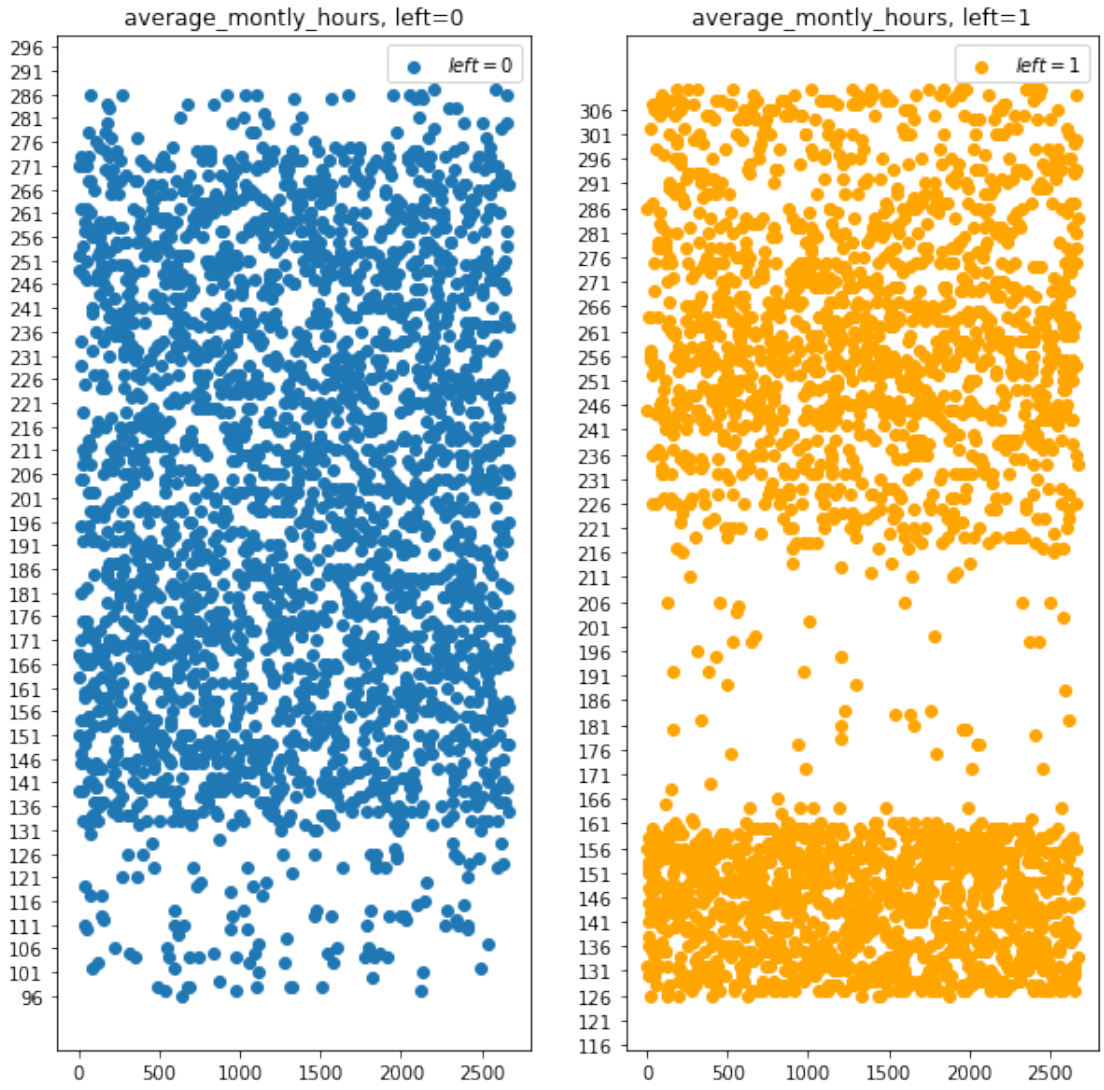

```
In [82]: filen = "train.csv"
         df = pd.read_csv(filen)

         #left
         left_df = df['left']
         # print left_df

         left_0 = df[df['left']==0]
         left_1 = df[df['left']==1]

         work_df0 = left_0['average_montly_hours']
         work_df1 = left_1['average_montly_hours']
```

```
        rows0 = work_df0.shape[0]
        print "rows in left=0: ",rows0,
        # ar0 = np.split(work_df0, 100)
        rows1 = work_df1.shape[0]
        print "rows in left=1: ",rows1

        lim=min(rows0-1,rows1-1)
        ar0 = work_df0[:lim]
        ar1 = work_df1[:lim]
        print "ar0, ar1 ",ar0.shape[0],ar1.shape[0]

        max0=ar0.max()
        min0=ar0.min()
        max1=ar1.max()
        min1=ar1.min()

        maxim=max(max1,max0)
        minim=min(min0,min1)
        print "max, min: ",maxim,minim
        ar =[]
        i=minim
        while i<maxim:
        #     print i
            ar.append(i)
            i=i+5
        print "ticks: ",ar

        x = np.linspace(0,lim-1,lim)
        fig, axes = plt.subplots(figsize=(10, 10))
        axes.set_yticks(ar)
        axes.scatter(x,ar0, label=r"$left=0$")
        axes.scatter(x,ar1,label=r"$left=1$")
        legend = axes.legend(loc='best')
        axes.set_title('average_montly_hours')

        fig2, axes2 = plt.subplots(1, 2, figsize=(10,10))
        axes2[0].set_yticks(ar)
        axes2[1].set_yticks(ar)

        axes2[0].scatter(x,ar0, label=r"$left=0$")
        axes2[1].scatter(x,ar1,label=r"$left=1$",color="orange")
        legend = axes2[0].legend(loc='best')
        legend2 = axes2[1].legend(loc='best')

        axes2[0].set_title('average_montly_hours, left=0')
        axes2[1].set_title('average_montly_hours, left=1')
```

```
rows in left=0:  8563 rows in left=1:  2675
```

```
ar0, ar1  2674 2674
max, min:  310 96
ticks:  [96, 101, 106, 111, 116, 121, 126, 131, 136, 141, 146, 151, 156, 161, 166, 171, 176, 181
```

Out[82]: Text(0.5,1,'average_montly_hours, left=1')



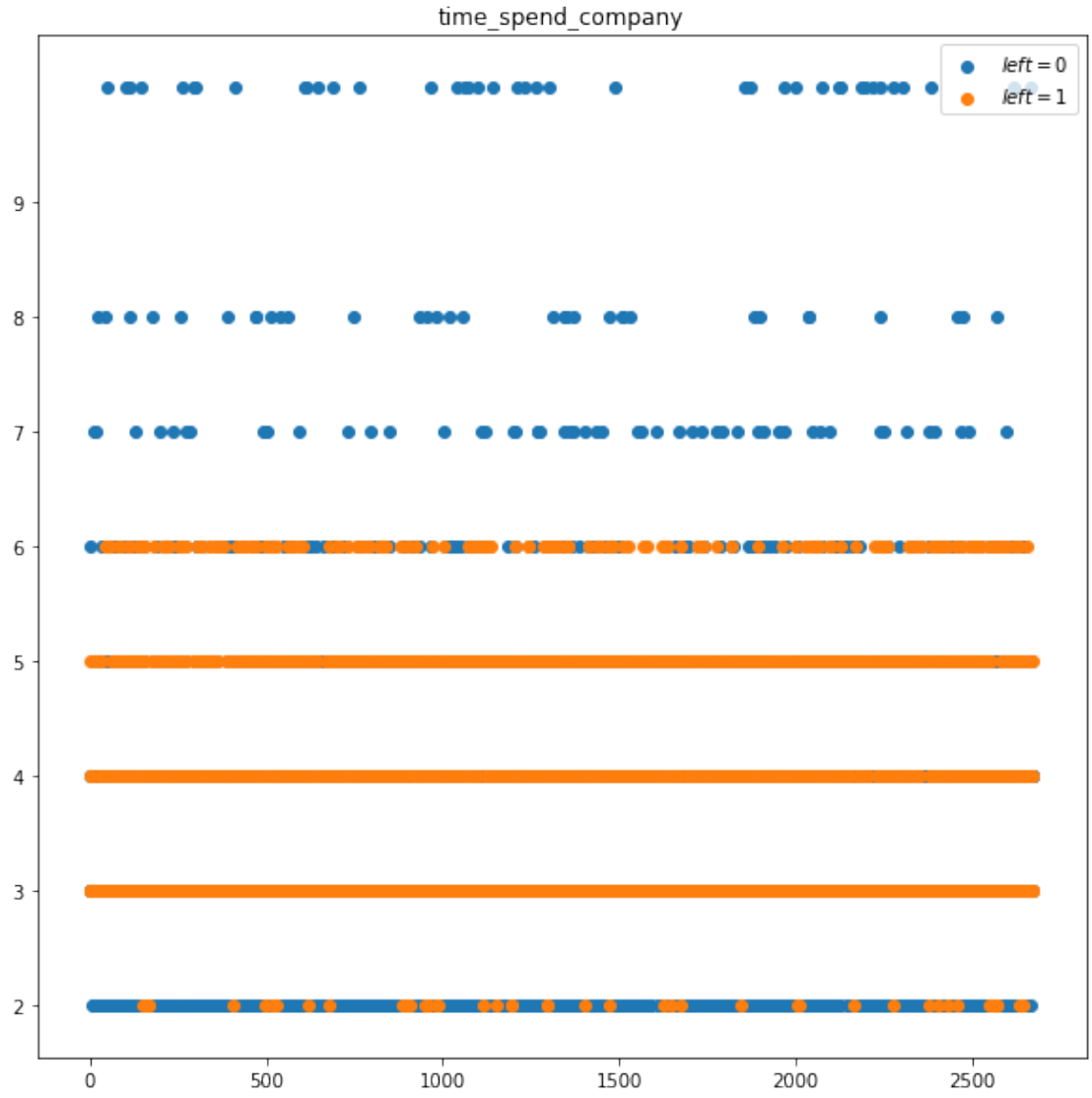
In [84]:
```
filen = "train.csv"
df = pd.read_csv(filen)

#left
left_df = df['left']
# print left_df

left_0 = df[df['left']==0]
left_1 = df[df['left']==1]

work_df0 = left_0['time_spend_company']
work_df1 = left_1['time_spend_company']

rows0 = work_df0.shape[0]
```

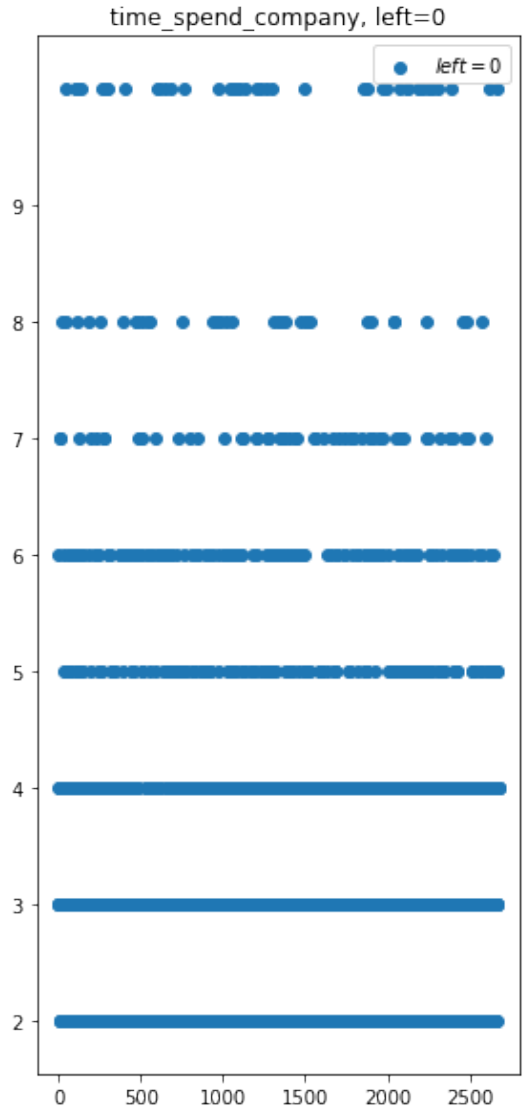
```
        print "rows in left=0: ",rows0,
        # ar0 = np.split(work_df0, 100)
        rows1 = work_df1.shape[0]
        print "rows in left=1: ",rows1

        lim=min(rows0-1,rows1-1)
        ar0 = work_df0[:lim]
        ar1 = work_df1[:lim]
        print "ar0, ar1 ",ar0.shape[0],ar1.shape[0]

        max0=ar0.max()
        min0=ar0.min()
        max1=ar1.max()
        min1=ar1.min()

        maxim=max(max1,max0)
        minim=min(min0,min1)
        print "max, min: ",maxim,minim
        ar =[]
        i=minim
        while i<maxim:
        #     print i
            ar.append(i)
            i=i+1
        print "ticks: ",ar

        x = np.linspace(0,lim-1,lim)
        fig, axes = plt.subplots(figsize=(10, 10))
        axes.set_yticks(ar)
        axes.scatter(x,ar0, label=r"$left=0$")
        axes.scatter(x,ar1,label=r"$left=1$")
        legend = axes.legend(loc='best')
        axes.set_title('time_spend_company')

        fig2, axes2 = plt.subplots(1, 2, figsize=(10,10))
        axes2[0].set_yticks(ar)
        axes2[1].set_yticks(ar)

        axes2[0].scatter(x,ar0, label=r"$left=0$")
        axes2[1].scatter(x,ar1,label=r"$left=1$",color="orange")
        legend = axes2[0].legend(loc='best')
        legend2 = axes2[1].legend(loc='best')

        axes2[0].set_title('time_spend_company, left=0')
        axes2[1].set_title('time_spend_company, left=1')
```

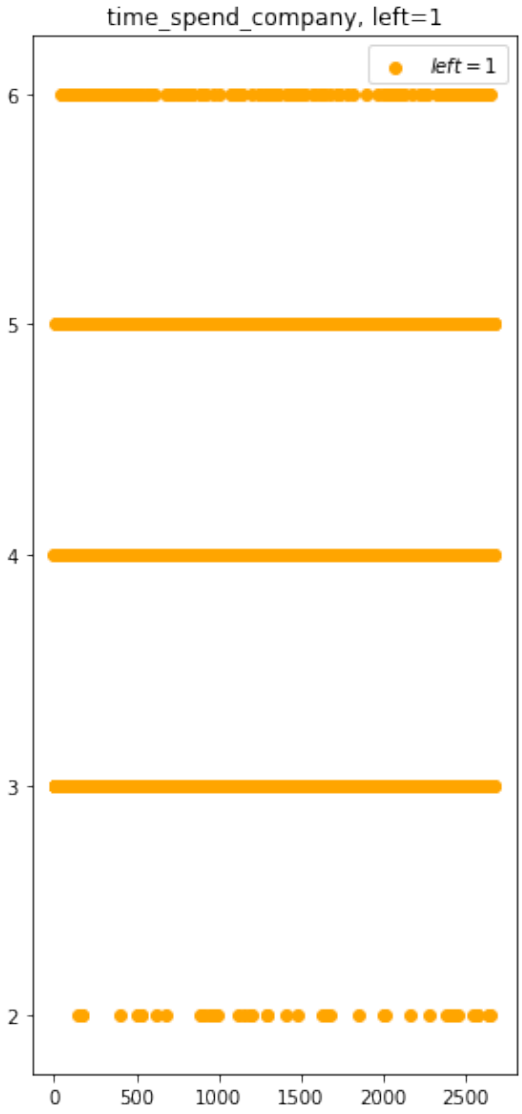
```
rows in left=0:  8563 rows in left=1:  2675
ar0, ar1  2674 2674
```

```
max, min:  10 2
ticks:  [2, 3, 4, 5, 6, 7, 8, 9]
```

Out[84]: Text(0.5,1,'time_spend_company, left=1')




time_spend_company, left=0
left = 0
9
8
7
6
5
4
3
2
0
500
1000
1500
2000
2500


time_spend_company, left=1
left = 1
6
5
4
3
2
0
500
1000
1500
2000
2500