In [78]:
```python
import pandas as pd
import numpy as np

init_notebook_mode(connected=True)
import plotly.plotly as py
import plotly.graph_objs as go
from plotly.offline import download_plotlyjs, init_notebook_mode, plot, iplot
```

In [79]:
```python
df=pd.read_stata('ps1.dta')
```

In [80]:
```python
df.head()
```

Out[80]:

	state	year	urate	children	nonwhite	finc	earn	age	ed	work	unearn
0	11.0	1991.0	7.6	3	0	7970.792951	7970.792951	39	10	1	0.000000
1	11.0	1991.0	7.6	0	0	31227.973568	14730.176211	48	11	1	16497.797357
2	11.0	1991.0	7.6	0	0	6143.072687	589.207048	36	7	1	5553.865639
3	11.0	1991.0	7.6	0	0	16761.762115	0.000000	44	7	0	16761.762115
4	11.0	1991.0	7.6	1	1	7424.008811	2121.145374	21	9	1	5302.863436

In [81]:
```python
df['employed']=np.where(df['work']==1,1,0)
df['unemployed']=np.where(df['work']==0,1,0)
df['parent']=np.where(df['children']!=0,1,0)
df.head()
```

Out[81]:

	state	year	urate	children	nonwhite	finc	earn	age	ed	work	unearn	employ
0	11.0	1991.0	7.6	3	0	7970.792951	7970.792951	39	10	1	0.000000	
1	11.0	1991.0	7.6	0	0	31227.973568	14730.176211	48	11	1	16497.797357	
2	11.0	1991.0	7.6	0	0	6143.072687	589.207048	36	7	1	5553.865639	
3	11.0	1991.0	7.6	0	0	16761.762115	0.000000	44	7	0	16761.762115	
4	11.0	1991.0	7.6	1	1	7424.008811	2121.145374	21	9	1	5302.863436	

In [82]:
```python
df1=df.groupby(['year', 'parent']).sum()
```

In [83]: `df1.head()`

Out[83]:

		state	urate	children	nonwhite	finc	earn	age	ed	wor
year	parent									
1991.0	0	58933.0	7701.500000	0.0	549.0	2.172756e+07	1.631560e+07	42639.0	9467.0	646
	1	81946.0	10648.700195	3148.0	942.0	1.821520e+07	1.108422e+07	48888.0	13387.0	691
1992.0	0	54641.0	8041.200195	0.0	523.0	1.944554e+07	1.420858e+07	39488.0	8833.0	595
	1	76944.0	11059.299805	2972.0	886.0	1.711481e+07	1.013080e+07	46211.0	12639.0	618
1993.0	0	52770.0	7171.200195	0.0	508.0	1.792198e+07	1.299647e+07	38570.0	8600.0	574

In [84]:
```
df1=df1.reset_index()
df1.head()
```

Out[84]:

	year	parent	state	urate	children	nonwhite	finc	earn	age	ed
0	1991.0	0	58933.0	7701.500000	0.0	549.0	2.172756e+07	1.631560e+07	42639.0	9467.0
1	1991.0	1	81946.0	10648.700195	3148.0	942.0	1.821520e+07	1.108422e+07	48888.0	13387.0
2	1992.0	0	54641.0	8041.200195	0.0	523.0	1.944554e+07	1.420858e+07	39488.0	8833.0
3	1992.0	1	76944.0	11059.299805	2972.0	886.0	1.711481e+07	1.013080e+07	46211.0	12639.0
4	1993.0	0	52770.0	7171.200195	0.0	508.0	1.792198e+07	1.299647e+07	38570.0	8600.0

In [85]:
```
df1['urate']=(df1['employed'])/(df1['employed']+df1['unemployed'])
parent=df1[df1['parent']==1]
nparent=df1[df1['parent']==0]
nparent.head()
```

Out[85]:

	year	parent	state	urate	children	nonwhite	finc	earn	age	ed	work
0	1991.0	0	58933.0	0.583032	0.0	549.0	2.172756e+07	1.631560e+07	42639.0	9467.0	646.0
2	1992.0	0	54641.0	0.571566	0.0	523.0	1.944554e+07	1.420858e+07	39488.0	8833.0	595.0
4	1993.0	0	52770.0	0.571144	0.0	508.0	1.792198e+07	1.299647e+07	38570.0	8600.0	574.0
6	1994.0	0	52178.0	0.590909	0.0	534.0	1.786046e+07	1.343095e+07	38525.0	8519.0	585.0
8	1995.0	0	49851.0	0.574236	0.0	494.0	1.740139e+07	1.326993e+07	35406.0	7791.0	526.0

In [86]:
```python
# Add data
year = parent['year']
parentLMPR= parent['urate']
nparentLMPR = nparent['urate']


# Create and style traces
trace0 = go.Scatter(
    x = year,
    y = parentLMPR,
    name = 'Single Women w/ child mean LMPR',
    line = dict(
        color = ('blue'),
        width = 2)
)
trace1 = go.Scatter(
    x = year,
    y = nparentLMPR,
    name = 'Single Women w/o child mean LMPR',
    line = dict(
        color = ('red'),
        width = 2,)
)


data = [trace0, trace1]

# Edit the layout
layout = dict(title = 'Single Women Labor Market Participation Rates',
              xaxis = dict(title = 'Year'),
              yaxis = dict(title = 'LMPR'),
              )

fig = dict(data=data, layout=layout)
py.iplot(fig, filename='raw-plot')
```
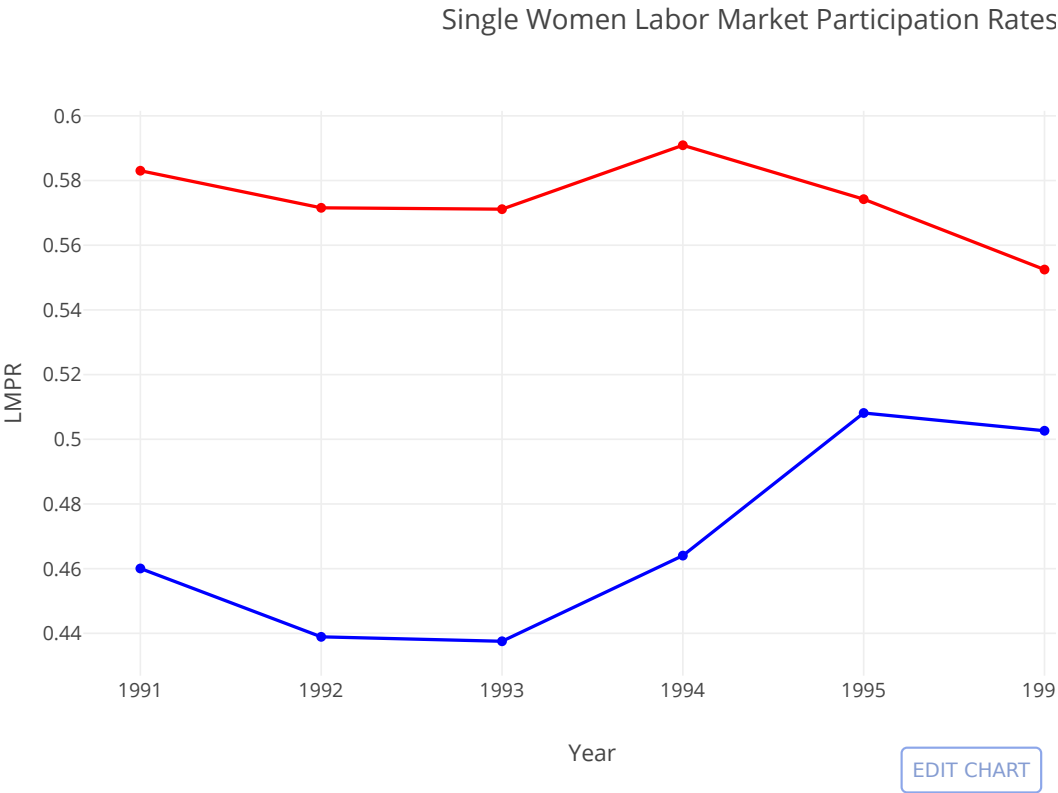
Out[86]:

Single Women Labor Market Participation Rates

In [87]: parent

Out[87]:

	year	parent	state	urate	children	nonwhite	finc	earn	age	ed	won
1	1991.0	1	81946.0	0.460053	3148.0	942.0	1.821520e+07	1.108422e+07	48888.0	13387.0	691.
3	1992.0	1	76944.0	0.438920	2972.0	886.0	1.711481e+07	1.013080e+07	46211.0	12639.0	618.
5	1993.0	1	71450.0	0.437547	2785.0	879.0	1.576093e+07	9.299774e+06	43600.0	12125.0	585.
7	1994.0	1	69276.0	0.464032	2668.0	868.0	1.666913e+07	9.928245e+06	41501.0	11459.0	587.
9	1995.0	1	67747.0	0.508127	2487.0	824.0	1.573574e+07	1.017857e+07	37850.0	10510.0	594.
11	1996.0	1	65650.0	0.502636	2334.0	800.0	1.619950e+07	1.122617e+07	37763.0	10260.0	572.

In [88]:
```
pBaseLevel=parent.iloc[0,3]
nBaseLevel=nparent.iloc[0,3]
nBaseLevel
```

Out[88]: 0.5830324909747292

In [89]:
```
parent['index']=parent['urate']/pBaseLevel
nparent['index']=nparent['urate']/nBaseLevel
```

/home/appertjt/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:1: S
ettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stab
le/indexing.html#indexing-view-versus-copy

/home/appertjt/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:2: S
ettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stab
le/indexing.html#indexing-view-versus-copy

In [90]:
```
parent
```

Out[90]:

	year	parent	state	urate	children	nonwhite	finc	earn	age	ed	wor
1	1991.0	1	81946.0	0.460053	3148.0	942.0	1.821520e+07	1.108422e+07	48888.0	13387.0	691.
3	1992.0	1	76944.0	0.438920	2972.0	886.0	1.711481e+07	1.013080e+07	46211.0	12639.0	618.
5	1993.0	1	71450.0	0.437547	2785.0	879.0	1.576093e+07	9.299774e+06	43600.0	12125.0	585.
7	1994.0	1	69276.0	0.464032	2668.0	868.0	1.666913e+07	9.928245e+06	41501.0	11459.0	587.
9	1995.0	1	67747.0	0.508127	2487.0	824.0	1.573574e+07	1.017857e+07	37850.0	10510.0	594.
11	1996.0	1	65650.0	0.502636	2334.0	800.0	1.619950e+07	1.122617e+07	37763.0	10260.0	572.

In [91]:
```python
# Add data
year = parent['year']
piLMPR= parent['index']
niLMPR = nparent['index']


# Create and style traces
trace0 = go.Scatter(
    x = year,
    y = piLMPR,
    name = 'Single Women w/ child mean LMPR',
    line = dict(
        color = ('blue'),
        width = 2)
)
trace1 = go.Scatter(
    x = year,
    y = niLMPR,
    name = 'Single Women w/o child mean LMPR',
    line = dict(
        color = ('red'),
        width = 2,)
)


data = [trace0, trace1]

# Edit the layout
layout = dict(title = 'Single Women Labor Market Participation Rates, Indexed t
o 1991 rates',
              xaxis = dict(title = 'Year'),
              yaxis = dict(title = 'LMPR'),
              )

fig = dict(data=data, layout=layout)
py.iplot(fig, filename='index-plot')
```
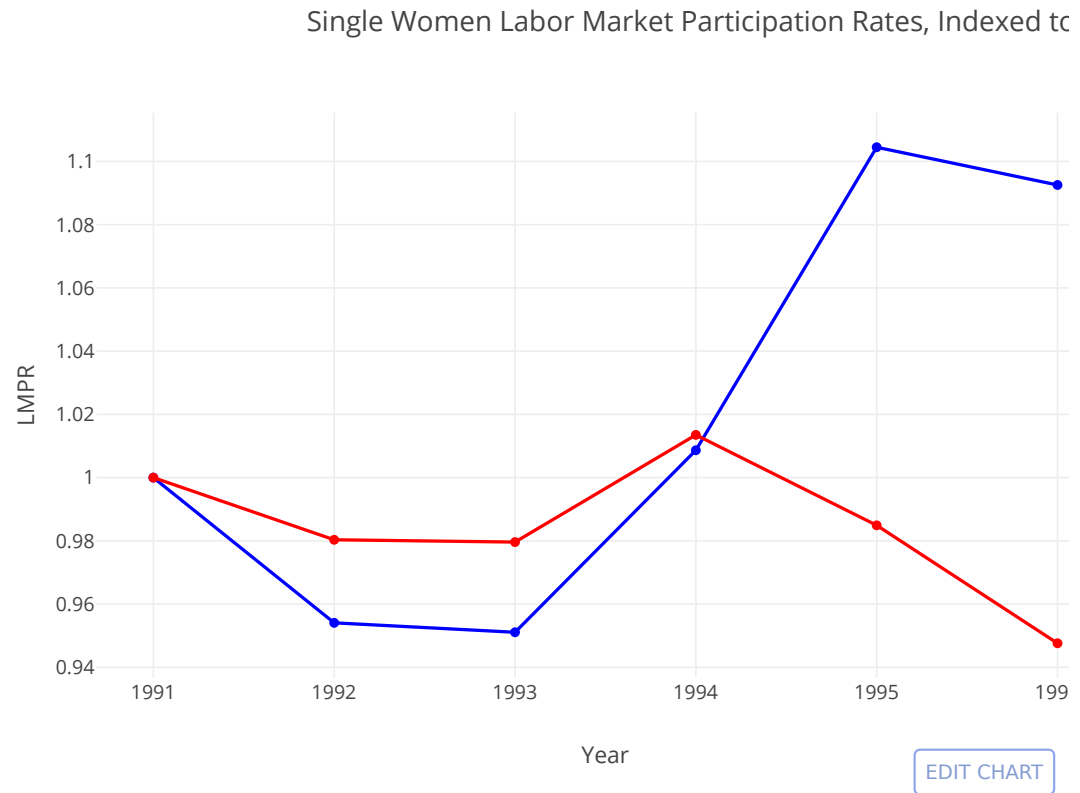
Out[91]:

Single Women Labor Market Participation Rates, Indexed to

EDIT CHART

TODO: Insert comments on validity of using single women with children as a control group.

In this section I calculate the difference in differences for table two. I start back at the original dataframe to get clean averages.

In [92]:
```python
df.head()
parent=df[df['parent']==1]
nparent=df[df['parent']!=1]

#calculate the average of the treatment group pre-1994
tc1=parent[parent['year']<1994]
tc1_empl=tc1['work'].sum()
tc1_mean=tc1_empl/len(tc1)

#calculate the average of the treatment group post-1994
tc2=parent[parent['year']>1993]
tc2_empl=tc2['work'].sum()
tc2_mean=tc2_empl/len(tc1)

#calculate the average of the control group pre-1994
cg1=nparent[nparent['year']<1994]
cg1_empl=cg1['work'].sum()
cg1_mean=cg1_empl/len(cg1)

#calculate the average of the control group post-1994
cg2=nparent[nparent['year']>1993]
cg2_empl=cg2['work'].sum()
cg2_mean=cg2_empl/len(tc1)

#calculate diffs
dif1=tc2_mean-tc1_mean
dif2=cg2_mean-cg1_mean
dif_dif=dif1-dif2

#print (tc1_mean, tc2_mean, cg1_mean, cg2_mean)

l1=["Treatment Group", len(parent), tc1_mean, tc2_mean, dif1, '']
l2=["Control Group", len(nparent), cg1_mean, cg2_mean, dif2, dif_dif]

table=[l1, l2]

headers=['Group', 'Sample Size', 'Pre-1993', 'Post-1993', 'Difference', 'Differ
ence-in-differences']

table2=pd.DataFrame(table, columns=headers)

table2
```

Out[92]:

	Group	Sample Size	Pre-1993	Post-1993	Difference	Difference-in-differences
0	Treatment Group	7819	0.445962	0.412762	-0.033200	
1	Control Group	5927	0.575460	0.374382	-0.201078	0.167878

TODO: Comment on this table

In [ ]: