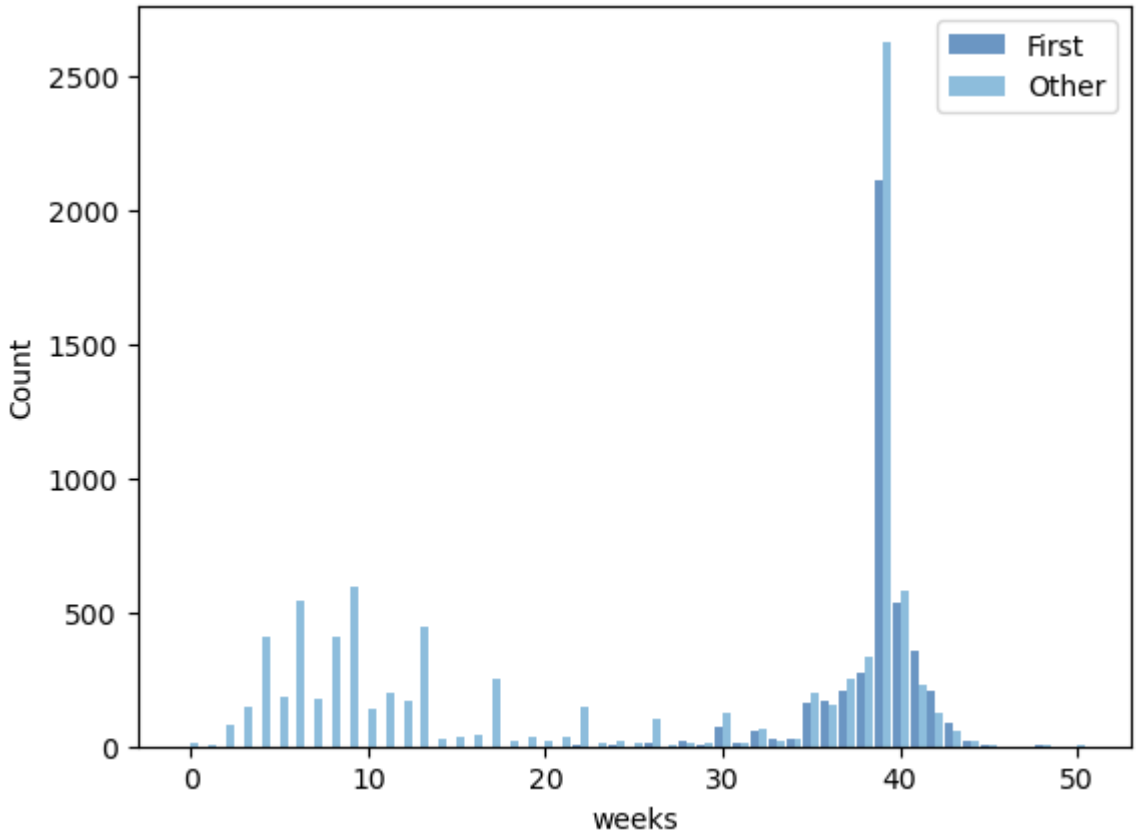
WEEK 3 - Assignment 2.1

In [1]:
```
import nsfg
import pandas as pd
import thinkstats2
import thinkplot
from os.path import basename, exists


def download(url):
    filename = basename(url)
    if not exists(filename):
        from urllib.request import urlretrieve

        local, _ = urlretrieve(url, filename)
        print("Downloaded " + local)

download("https://github.com/AllenDowney/ThinkStats2/raw/master/code/thinkstats2.py
download("https://github.com/AllenDowney/ThinkStats2/raw/master/code/thinkplot.py")
```

In [4]:
```
preg = nsfg.ReadFemPreg()
```

In [6]:
```
# live= preg[preg.outcome == 1]
firsts = preg[preg.birthord == 1]
others = preg[preg.birthord != 1]
```

In [12]:
```
# live= preg[preg.outcome == 1]
firsts = preg[preg.birthord == 1]
others = preg[preg.birthord != 1]
first_hist = thinkstats2.Hist(firsts.prglngth, label='First')
other_hist = thinkstats2.Hist(others.prglngth, label='Other')
width = 0.45
thinkplot.PrePlot(2)
thinkplot.Hist(first_hist, align='right', width=width)
thinkplot.Hist(other_hist, align='left', width=width)
thinkplot.Config(xlabel='weeks', ylabel='Count')
```



In [11]:
```
firstborn= firsts[['caseid', 'prglngth']]
otherborn = others[['caseid', 'prglngth']]

dump = pd.merge(firstborn,otherborn, on = ['caseid'])
dump
```

Out[11]:

	caseid	prglngth_x	prglngth_y
0	1	39	39
1	2	39	39
2	2	39	39
3	6	38	40
4	6	38	42
...	...	...	...
8215	12569	34	17
8216	12571	39	6
8217	12571	39	5
8218	12571	39	39
8219	12571	39	39

8220 rows × 3 columns

In [23]:
```
dump.drop_duplicates()
```

Out[23]:

	caseid	prglngth_x	prglngth_y
0	1	39	39
1	2	39	39
3	6	38	40
4	6	38	42
5	7	39	35
...	...	...	...
8214	12568	39	3
8215	12569	34	17
8216	12571	39	6
8217	12571	39	5
8218	12571	39	39

6577 rows × 3 columns

In [26]:
```
dumplist["val"] = dumplist['prglngth_x'] > dumplist['prglngth_y']
```

In [24]:
```
dumplist = dumplist.drop_duplicates()
dumplist
```

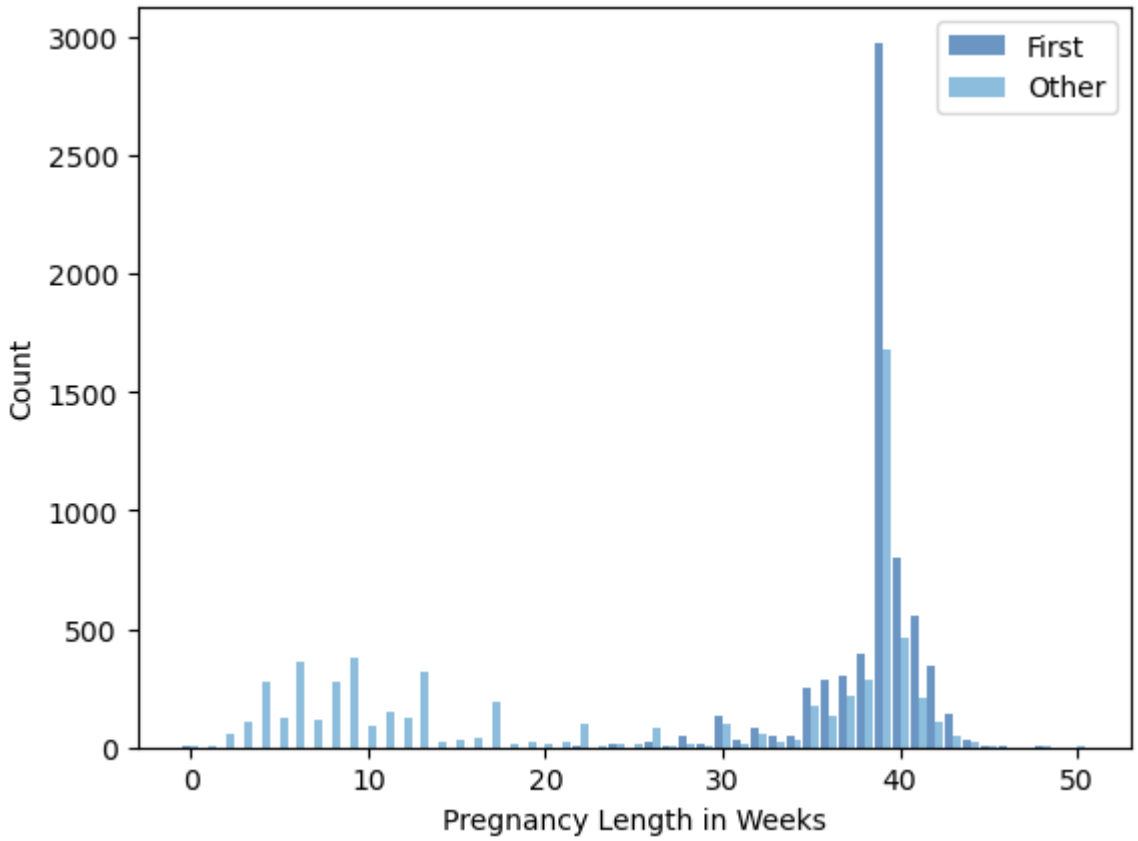
Out[24]:

	caseid	prglngth_x	prglngth_y	val
0	1	39	39	False
1	2	39	39	False
3	6	38	40	False
4	6	38	42	False
5	7	39	35	True
...	...	...	...	...
8214	12568	39	3	True
8215	12569	34	17	True
8216	12571	39	6	True
8217	12571	39	5	True
8218	12571	39	39	False

6577 rows × 4 columns

In [25]:
```
import thinkstats2
import thinkplot
```

```
first_hist = thinkstats2.Hist(dumplist.prglngth_x, label='First')
other_hist = thinkstats2.Hist(dumplist.prglngth_y, label='Other')

width = 0.45
thinkplot.PrePlot(2)
thinkplot.Hist(first_hist, align='right', width=width)
thinkplot.Hist(other_hist, align='left', width=width)
thinkplot.Config(xlabel='Pregnancy Length in Weeks', ylabel='Count')
```



In [27]:
```
dumplist.val.value_counts()
```

Out[27]:
```
True     4107
False    2470
Name: val, dtype: int64
```

In [ ]:
```
#Looking at the values and the plot for pregnancy length greater for first born
# vs others, it appears first born babies arrive later than other.
```