In [1]:
```
import pandas as pd
```

In [2]:
```
# for visualizations, don't worry about these for now
import seaborn as sns
import matplotlib.pyplot as plt
from matplotlib.patches import Rectangle
```

Datasets

Age of First Marriage

In [3]:
```
age_marriage = pd.read_csv("age_at_mar.csv", index_col=0).squeeze("columns")

age_marriage.head()
```

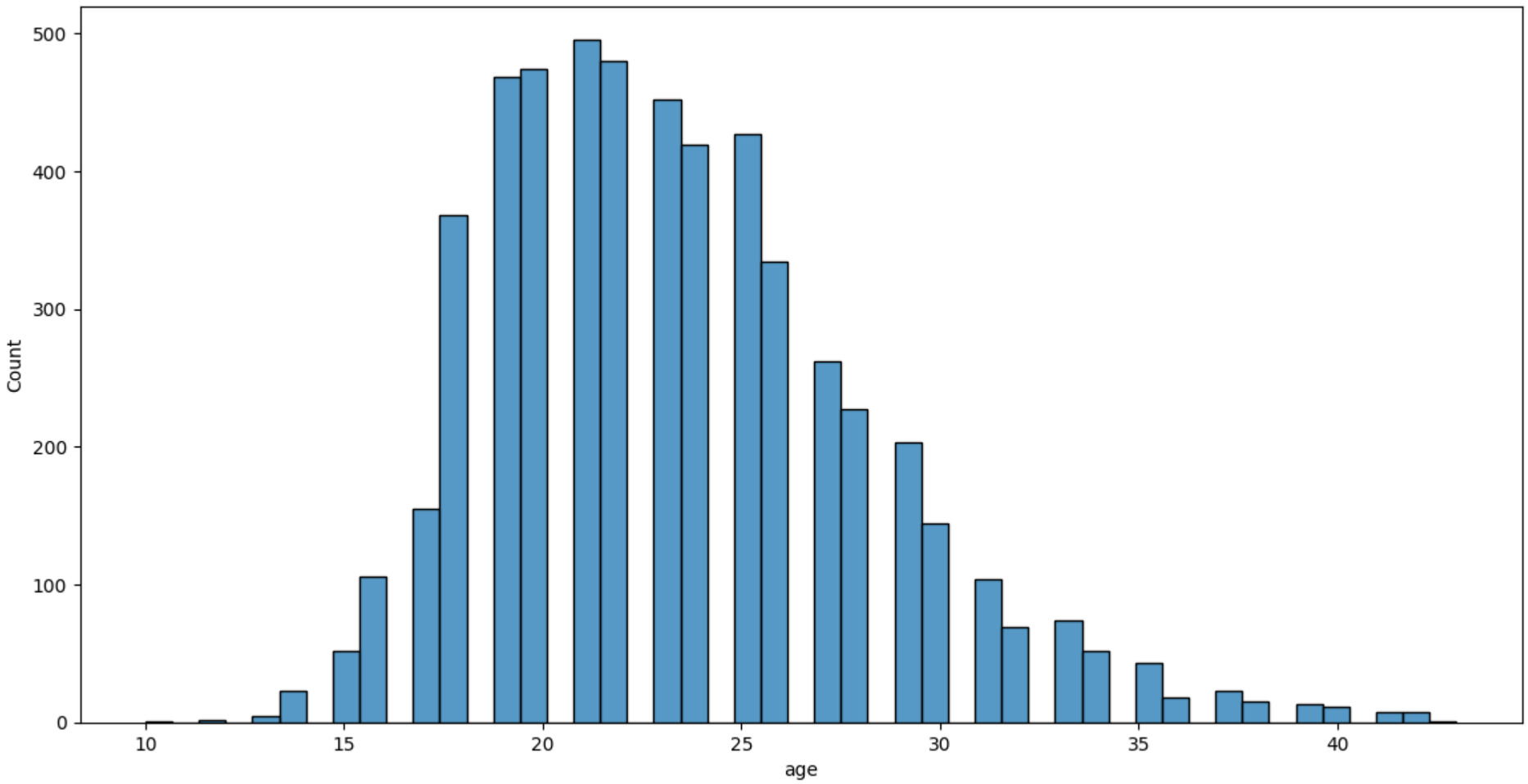
Out[3]:
```
1    32
2    25
3    24
4    26
5    32
Name: age, dtype: int64
```

In [4]:
```
age_marriage.shape
```

Out[4]:
```
(5534,)
```

In [5]:
```
fig, ax = plt.subplots(figsize=(14, 7))
sns.histplot(age_marriage, ax=ax)
```

Out[5]: <Axes: xlabel='age', ylabel='Count'>



S&P Returns 1990's

In [6]:
```
sp500 = pd.read_csv('SP500.csv', index_col=0).squeeze("columns")
sp500.head()
```

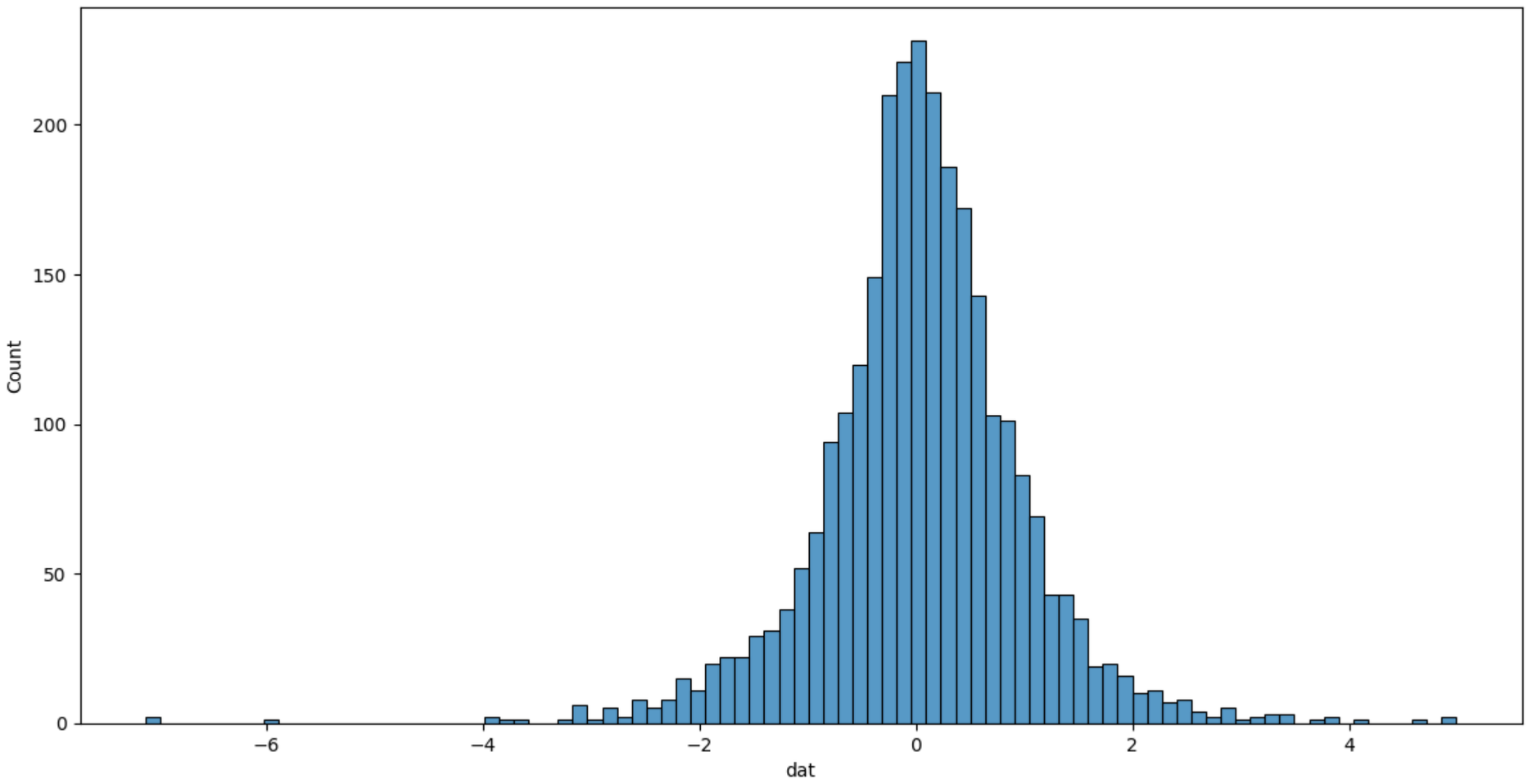
Out[6]:
```
1   -0.258891
2   -0.865031
3   -0.980414
4    0.450432
5   -1.185667
Name: dat, dtype: float64
```

In [7]:
```
sp500.shape
```

Out[7]: (2780,)

In [8]:
```
fig, ax = plt.subplots(figsize=(14, 7))
sns.histplot(sp500, ax=ax)
```

Out[8]: <Axes: xlabel='dat', ylabel='Count'>



Activities

1. Rename the series

In [11]:
```
age_marriage.name = "Age of First Marriage"
age_marriage
```

Out[11]:
```
1       32
2       25
3       24
4       26
5       32
        ..
5530    28
5531    19
5532    24
5533    21
5534    19
Name: Age of First Marriage, Length: 5534, dtype: int64
```

In [12]:
```
sp500.name = "S&P500 Returns 90s"
sp500
```

Out[12]:
```
1      -0.258891
2      -0.865031
3      -0.980414
4       0.450432
5      -1.185667
          ...
2776    0.705040
2777    1.038544
2778    0.398027
2779   -1.050302
2780   -2.843233
Name: S&P500 Returns 90s, Length: 2780, dtype: float64
```

Basic Analysis

2. What's the maximum Age of marriage?

In [14]:
```
age_marriage.max()
```

Out[14]:
```
43
```

3. What's the median Age of Marriage?

In [15]:
```
age_marriage.mean()
```

Out[15]:
```
23.440187929165162
```

4. What's the minimum return from S&P500?

In [16]:
```
sp500.min()
```

Out[16]:
```
-7.11274461287603
```

Simple Selection and Filtering

5. How many Women marry at age 21?

In [20]:
```
age_marriage.value_counts()
```

Out[20]:
```
Age of First Marriage
21    495
22    480
20    474
19    468
23    452
25    427
24    419
18    368
26    334
27    262
28    227
29    203
17    155
30    144
16    106
31    104
33     74
32     69
15     52
34     52
35     43
37     23
14     23
36     18
38     15
39     13
40     11
41      7
42      7
13      5
12      2
10      1
43      1
Name: count, dtype: int64
```

6. How many Women marry at 39y/o or older?

In [22]:
```
age_marriage_39 = age_marriage >= 39
age_marriage_39.sum()
```

Out[22]:
```
39
```

7. How many positive S&P500 returns are there?

The following visualization shows a red vertical line at the point `0`, we're looking for everything at the right of that line:

In [ ]:
```
ax = sns.histplot(sp500)
ax.axvline(0, color='red')
```

In [24]:
```
sp500_positive = sp500 > 0
sp500_positive.sum()
```

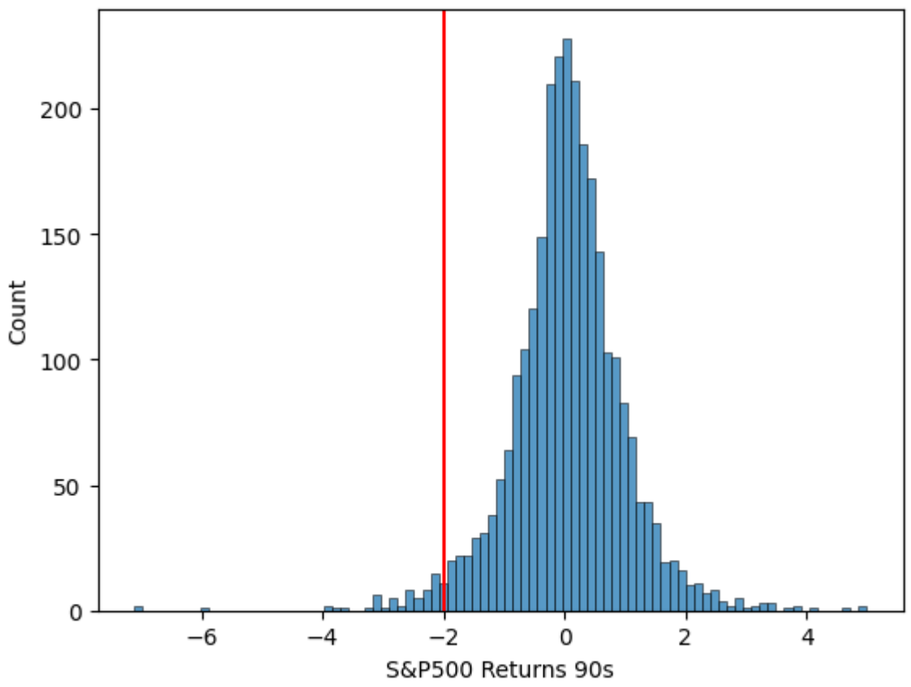
Out[24]: 1474

8. How many returns are less or equals than -2?

(Left to the red line)

In [23]:
```
ax = sns.histplot(sp500)
ax.axvline(-2, color='red')
```

Out[23]: <matplotlib.lines.Line2D at 0x7f04f6a1aa90>



In [25]:
```
sp500_neg = sp500 <= -2
sp500_neg.sum()
```

Out[25]: 63

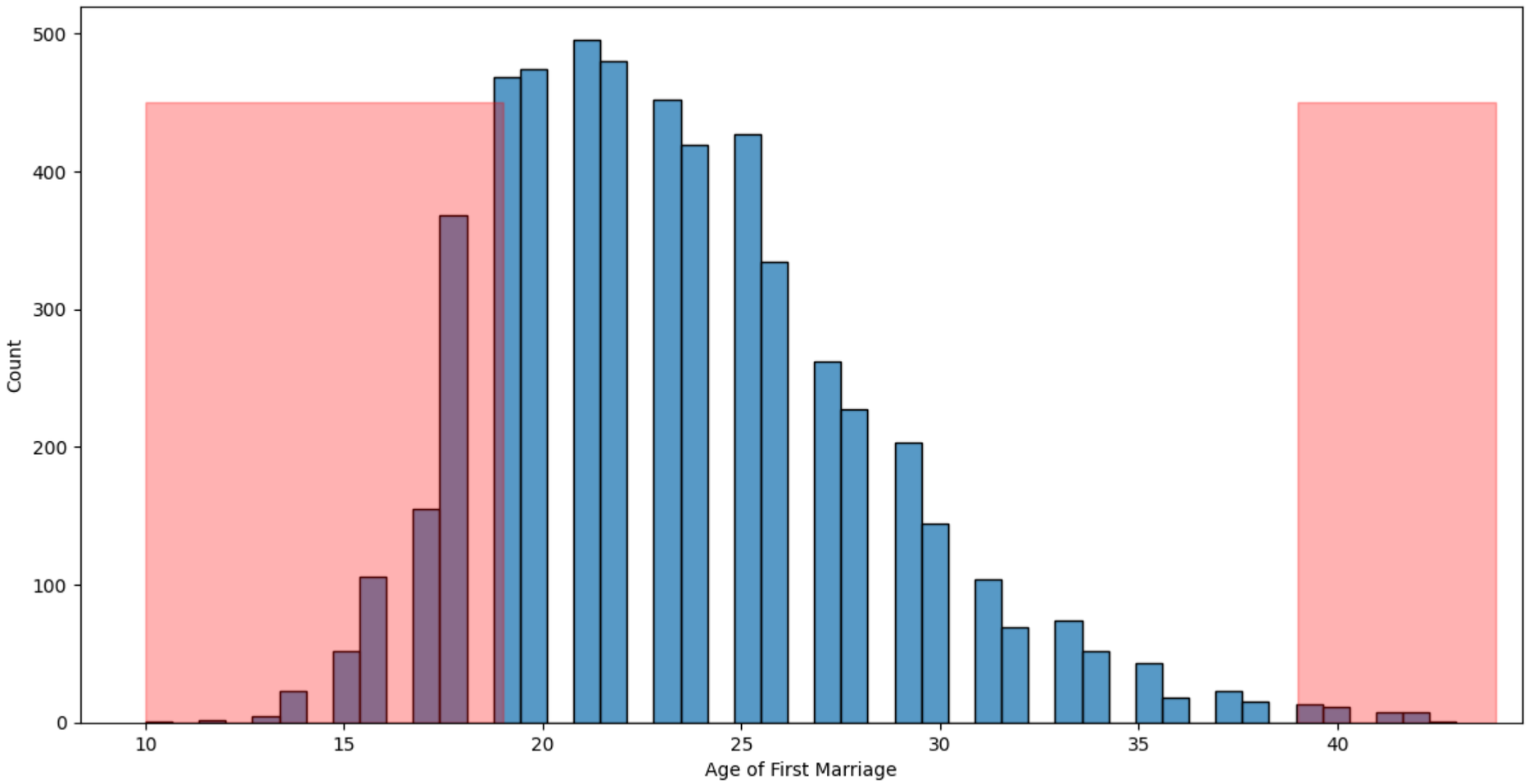
Advanced Selection with Boolean Operators

9. Select all women below 20 or above 39

The segments depicted below:

In [26]:
```
fig, ax = plt.subplots(figsize=(14, 7))
sns.histplot(age_marriage, ax=ax)
ax.add_patch(Rectangle((10, 0), 9, 450, alpha=.3, color='red'))
ax.add_patch(Rectangle((39, 0), 5, 450, alpha=.3, color='red'))
```

Out[26]:
```
<matplotlib.patches.Rectangle at 0x7f04f6968c90>
```



In [29]:
```
age_20_39 = age_marriage.loc[(age_marriage < 20) | (age_marriage > 39)]
age_20_39
```

Out[29]:
```
14      40
35      19
74      19
76      17
84      19
        ..
5517    16
5520    18
5527    19
5531    19
5534    19
Name: Age of First Marriage, Length: 1206, dtype: int64
```

10. Select all women whose ages are even, and are older than 30 y/o

In [41]:
```
age_30_even = age_marriage.loc[(age_marriage >30) & (age_marriage %2==0)]
age_30_even
```

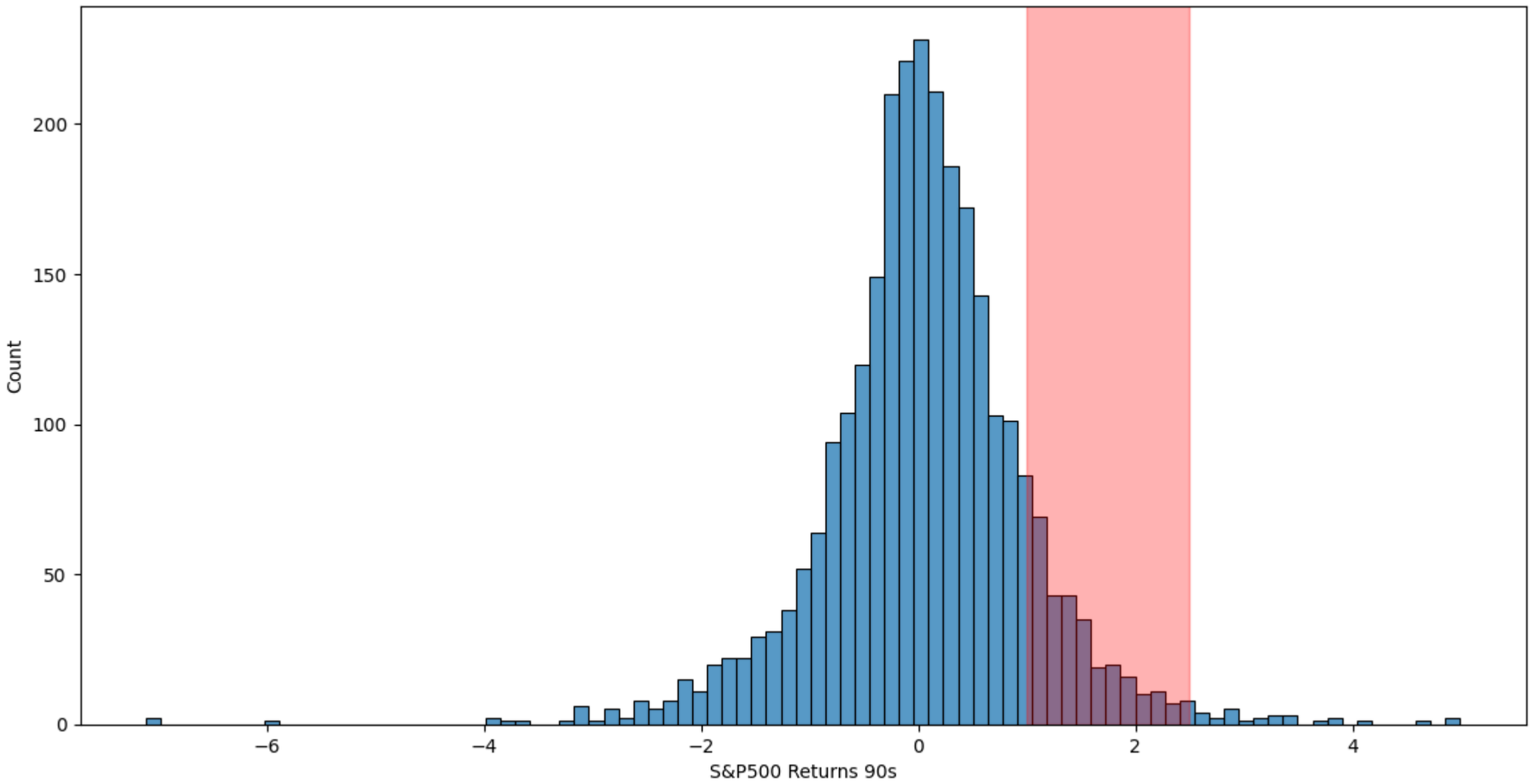
Out[41]:
```
1       32
5       32
14      40
24      34
55      32
        ..
5477    32
5488    40
5510    32
5516    34
5528    34
Name: Age of First Marriage, Length: 172, dtype: int64
```

10. Select the S&P500 returns between 1.5 and 3

The ones depicted below:

In [34]:
```
fig, ax = plt.subplots(figsize=(14, 7))
sns.histplot(sp500, ax=ax)
ax.add_patch(Rectangle((1, 0), 1.5, 250, alpha=.3, color='red'))
```

Out[34]: <matplotlib.patches.Rectangle at 0x7f04f6906050>



In [35]:
```
sp_15_to_3 = sp500.loc[(sp500 >=1.5) & (sp500 <=3)]
sp_15_to_3
```

Out[35]:
```
21      1.871048
91      2.351291
102     1.697397
188     1.673790
189     2.863366
          ...
2715    2.199155
2738    2.174014
2748    2.318141
2765    1.941508
2775    2.409438
Name: S&P500 Returns 90s, Length: 123, dtype: float64
```

In [ ]: