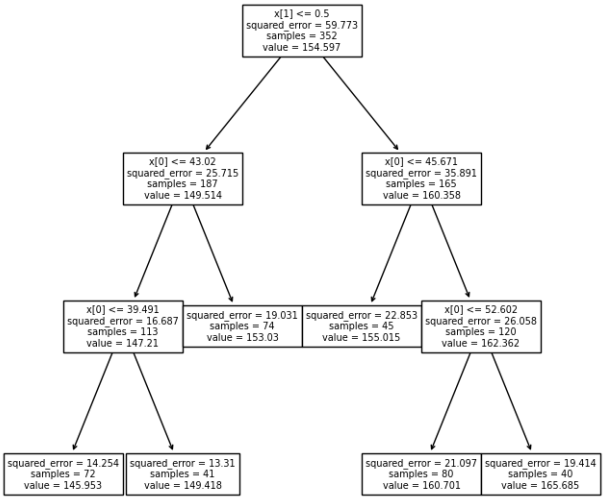
ACTU PS5841 Data Science in Finance and Insurance - Autumn 2023 (Y. Wang)
Assignment - 10
Assigned 11/14/23, Due 11/28/23

The Decision Tree on page 2 is generated in R by the model,

`tree(height ~ weight + as.factor(male))`

fitted with the Howell dataset on those aged 18-years or older (as discussed in class).

[*a*] use `sklearn.tree.DecisionTreeRegressor()` to fit a tree that replicates the R output. Let your code plot the tree

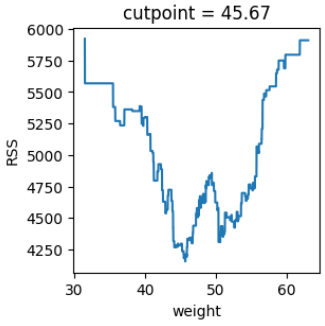


Code in python, but do NOT use `sklearn.tree.DecisionTreeRegressor()` for the following tasks.
[*b*] Based on the recursive binary splitting scheme, show that splitting node 3) by weight is more optimal than splitting it by sex.

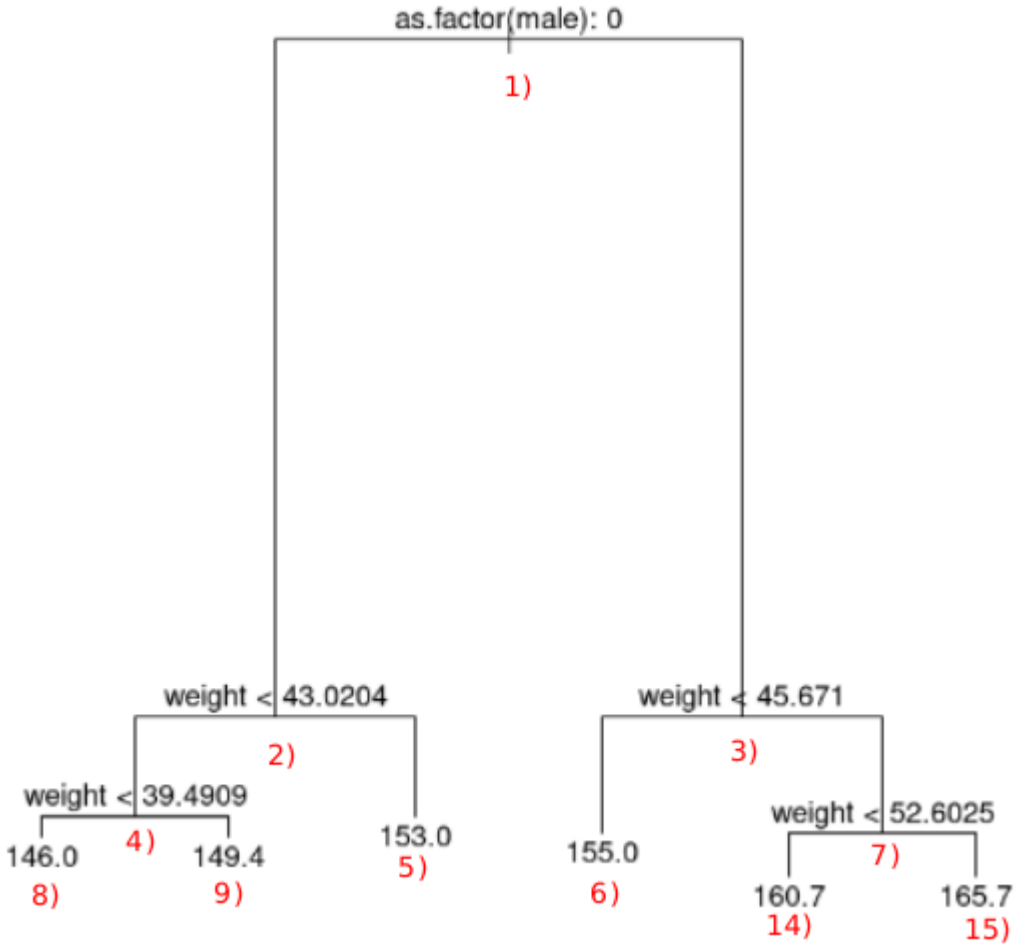
Let your code calculate and report the following.

```
split node 3 by sex:  cutpoints = Male, RSS = 5922.0742

split node 3 by weight:  cutpoint 45.67 RSS = 4155.3382
resulting node 6:  mse = 22.8533, samples = 45, prediction = 155.0149
resulting node 7:  mse = 26.0578, samples = 120, prediction = 162.3623
```



Note: Any cut point in (45.642695, 45.699394) is valid as there are no observations in this interval.


as.factor(male): 0
1)
weight < 43.0204
2)
weight < 45.671
3)
weight < 39.4909
4)
153.0
5)
155.0
6)
weight < 52.6025
7)
146.0
8)
149.4
9)
160.7
14)
165.7
15)