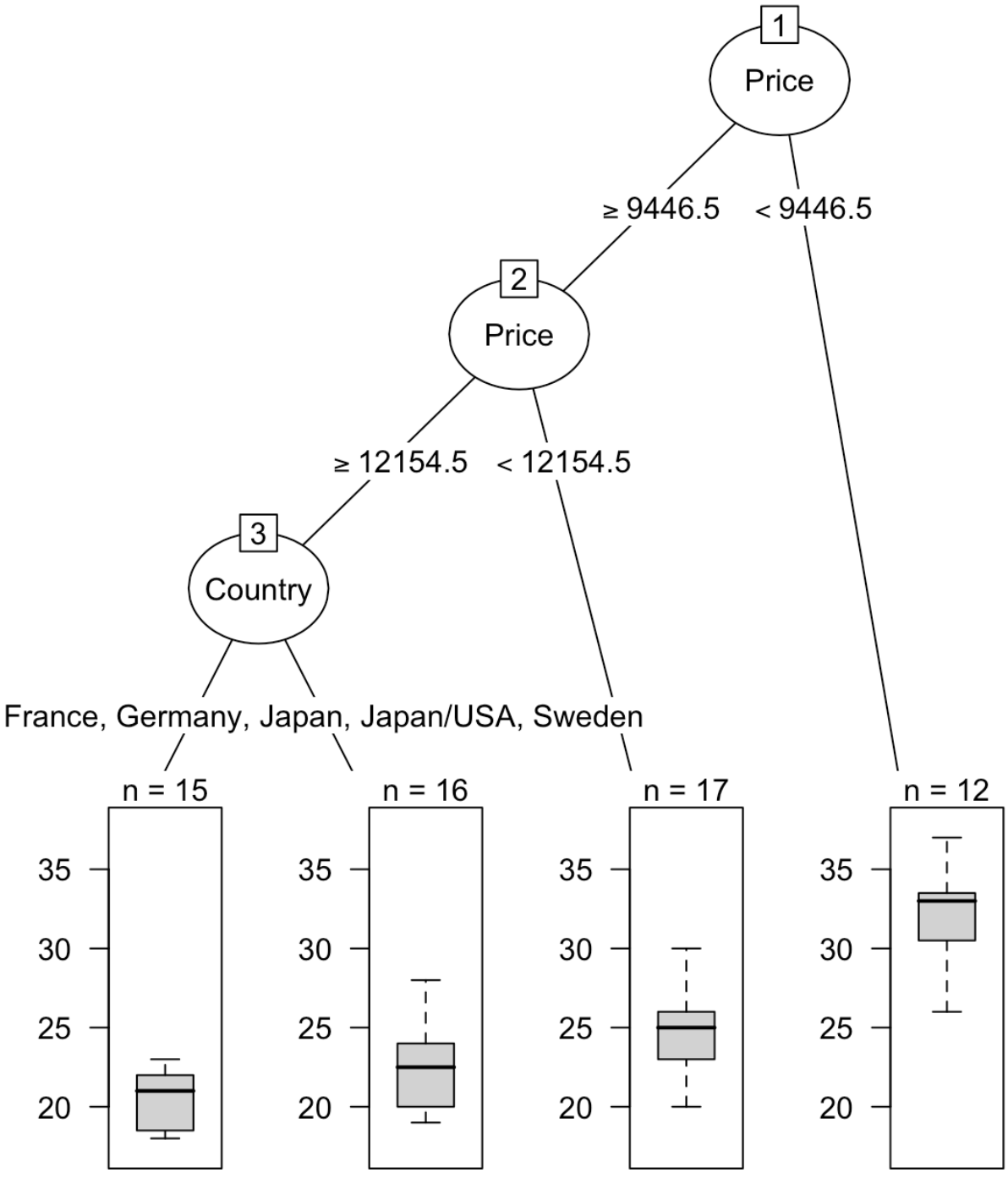
Hw11

20150056 국윤범

Problem 1

(a)



(b)

```
> # (b)
> rfit$variable.importance
      Price     Country Reliability
 962.331018  193.031638    9.032401
```

As you seen above in the graph, **Price** and **Country** variables are used to explain Mileage

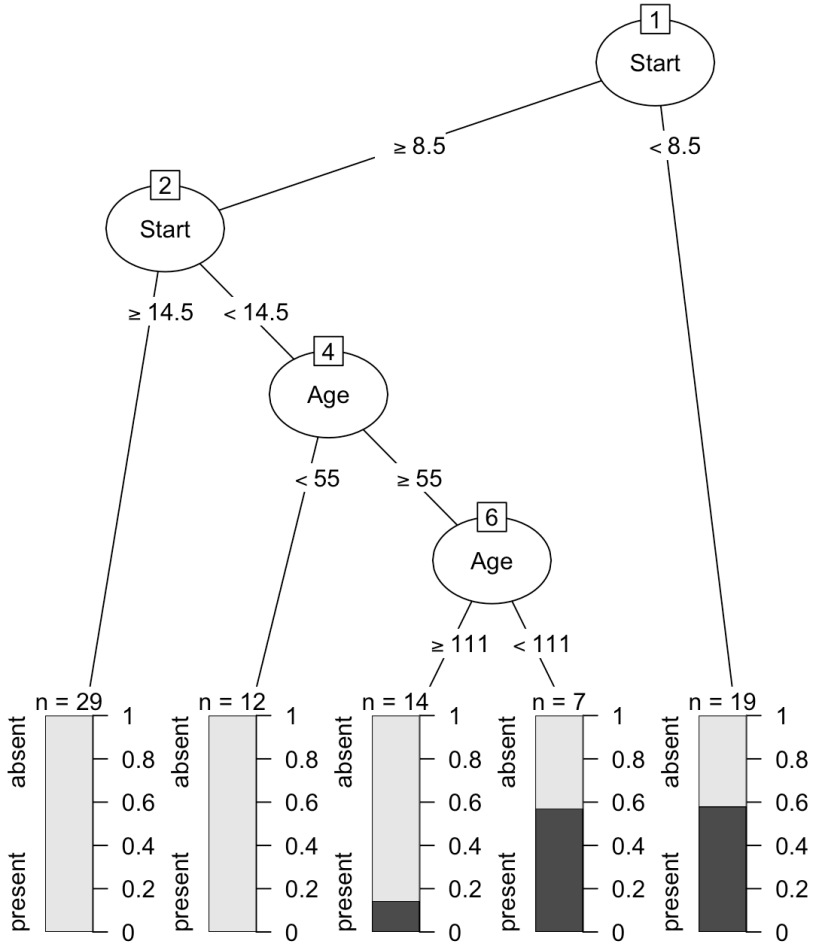
(c)

```
> rfit
n=60 (57 observations deleted due to missingness)

node), split, n, deviance, yval
      * denotes terminal node

1) root 60 1354.58300 24.58333
  2) Price>=9446.5 48  407.91670 22.70833
    4) Price>=12154.5 31  209.35480 21.61290
      8) Country=USA 15   49.73333 20.53333 *
      9) Country=France,Germany,Japan,Japan/USA,Sweden 16  125.75000 22.62500 *
    5) Price< 12154.5 17   93.52941 24.70588 *
  3) Price< 9446.5 12  102.91670 32.08333 *
```

From 1), the average of mileage of whole data is 24.58. From 2), the optimal splitting variable of whole data is Price. The optimal value of Price which yields the most drop in RSS is 9447.5. The group with Price>=9446.5 consists of 48 data and the average of mileage of this group is 22.71. From 3), the group with Price<9446.5 consists of 12 data and the average of mileage of this group is 32.08. Note that this group is no more splitted (i.e. terminal node). Likewise, from 4) and 5), for the data with Price >=9446, the optimal splitting variable is Price once again and the optimal value is 12154. From 8) and 9), for the data with Price>12154, the optimal splitting variable is Country once again and the optimal value is USA or not(France, Germany, Japan, Japan/USA, Sweden).

Problem 2

(a)



(b)

```
> rfit$variable.importance
   Start      Age   Number
8.198442 3.101801 1.521863
```

As you seen above in the graph, **Start** and **Age** variables are used to explain Kyphosis.

(c)

```
> rfit
n= 81

node), split, n, loss, yval, (yprob)
      * denotes terminal node

 1) root 81 17 absent (0.79012346 0.20987654)
   2) Start>=8.5 62  6 absent (0.90322581 0.09677419)
     4) Start>=14.5 29  0 absent (1.00000000 0.00000000) *
     5) Start< 14.5 33  6 absent (0.81818182 0.18181818)
      10) Age< 55 12  0 absent (1.00000000 0.00000000) *
      11) Age>=55 21  6 absent (0.71428571 0.28571429)
        22) Age>=111 14  2 absent (0.85714286 0.14285714) *
        23) Age< 111 7  3 present (0.42857143 0.57142857) *
   3) Start< 8.5 19  8 present (0.42105263 0.57894737) *
```

1) Root : 81 data in total & 64 (79%) of the data is “absent” and 17 (20.9%) is “present”
2) Start>=8.5 : 62 data in total & 56 (90%) are “absent” and 6 (57.8%) are “present”
3) Start<8.5 : 19 data in total & 8 (42%) are “absent” and 11 (57.8%) are “present”
4) Start>=14.5 : 29 data in total & 29 (100%) are “absent” and 0 (0%) are “present”
5) Start<14.5 : 33 data in total & 27 (81.8%) are “absent” and 6 (18.2%) are “present”
6) Start<14.5, Age<55 : 12 data in total & 12 (100%) are “absent” and 0 (0%) are “present”
7) Start<14.5, Age>=55 : 21 data in total & 15 (71.4%) are “absent” and 6 (28.5%) are “present”
8) Age>=111 : 14 data in total & 12 (85.7%) are “absent” and 2 (14.2%) are “present”
9) 55<=Age<111 : 7 data in total & 3 (42.8%) are “absent” and 4 (57.1%) are “present”