PRACTICAL 10: Practical of Decision Tree

1). Load data set in r_studio, Summary of data set, Display name of column

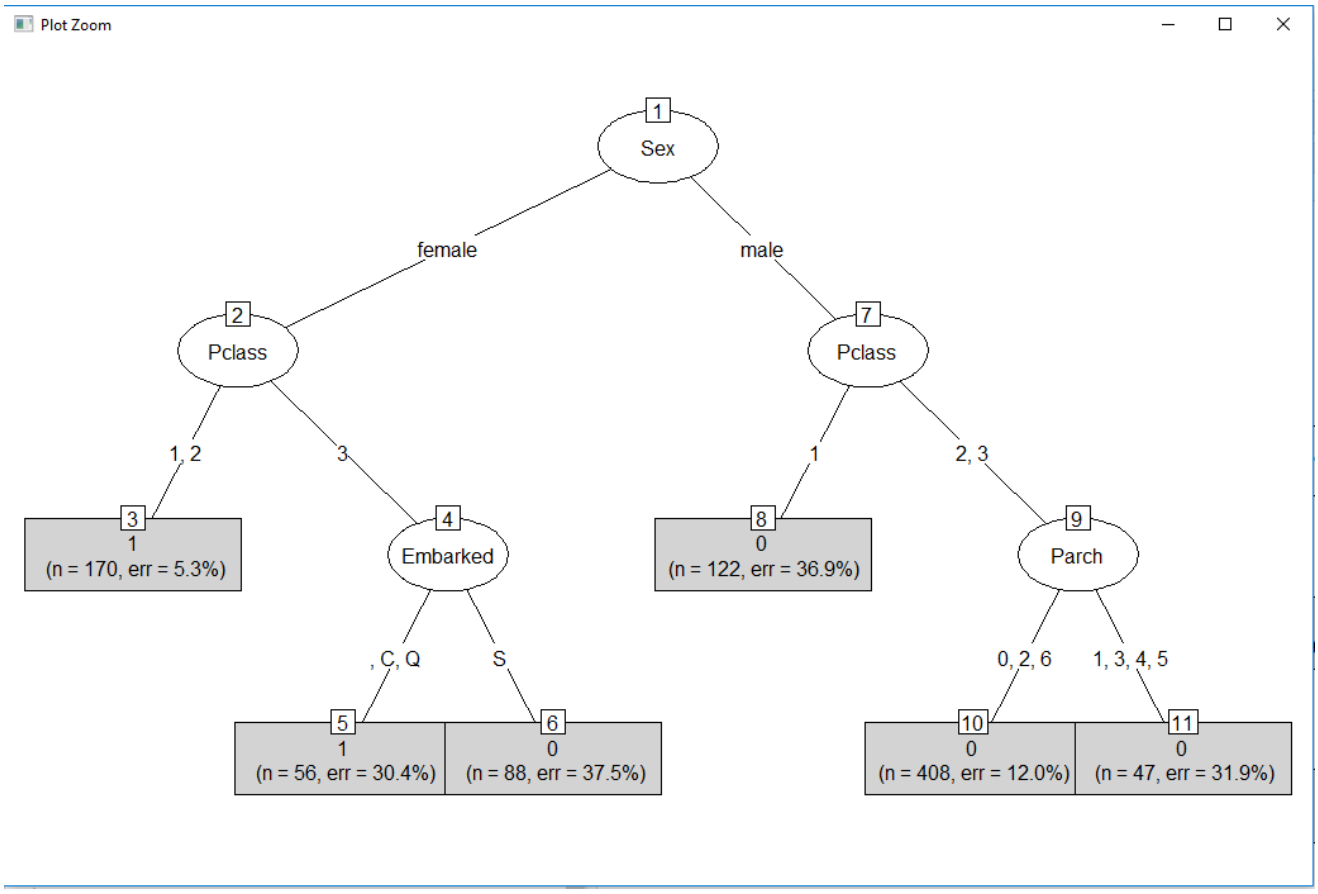
```
> titanic<-read.csv(file.choose(), header = T, sep=",")
> summary(titanic)
  PassengerId       Survived          Pclass
 Min.   :  1.0   Min.   :0.0000   Min.   :1.000
 1st Qu.:223.5   1st Qu.:0.0000   1st Qu.:2.000
 Median :446.0   Median :0.0000   Median :3.000
 Mean   :446.0   Mean   :0.3838   Mean   :2.309
 3rd Qu.:668.5   3rd Qu.:1.0000   3rd Qu.:3.000
 Max.   :891.0   Max.   :1.0000   Max.   :3.000

                                     Name         Sex           Age
 Abbing, Mr. Anthony                   :  1   female:314   Min.   : 0.42
 Abbott, Mr. Rossmore Edward           :  1   male  :577   1st Qu.:20.12
 Abbott, Mrs. Stanton (Rosa Hunt)      :  1                Median :28.00
 Abelson, Mr. Samuel                   :  1                Mean   :29.70
 Abelson, Mrs. Samuel (Hannah Wizosky):  1                3rd Qu.:38.00
 Adahl, Mr. Mauritz Nils Martin        :  1                Max.   :80.00
 (Other)                               :885                NA's   :177
     SibSp           Parch             Ticket         Fare
 Min.   :0.000   Min.   :0.0000   1601     :  7   Min.   :  0.00
 1st Qu.:0.000   1st Qu.:0.0000   347082   :  7   1st Qu.:  7.91
 Median :0.000   Median :0.0000   CA. 2343:  7   Median : 14.45
 Mean   :0.523   Mean   :0.3816   3101295  :  6   Mean   : 32.20
 3rd Qu.:1.000   3rd Qu.:0.0000   347088   :  6   3rd Qu.: 31.00
 Max.   :8.000   Max.   :6.0000   CA 2144  :  6   Max.   :512.33
                                  (Other)  :852
         Cabin     Embarked
            :687     :  2
 B96 B98    :  4   C:168
 C23 C25 C27:  4   Q: 77
 G6         :  4   S:644
 C22 C26    :  3
 D          :  3
 (Other)    :186
> names(titanic)
 [1] "PassengerId" "Survived"    "Pclass"      "Name"        "Sex"
 [6] "Age"         "SibSp"       "Parch"       "Ticket"      "Fare"
[11] "Cabin"       "Embarked"
```

2) Display survived data from data set

install.packages("partykit")
install.packages("CHAID",repos="http://R-Forge.R-project.org",type="source")
library(CHAID) #Chi-Square automatic interaction detection
library(partykit) #Tool used for recursive partitioning

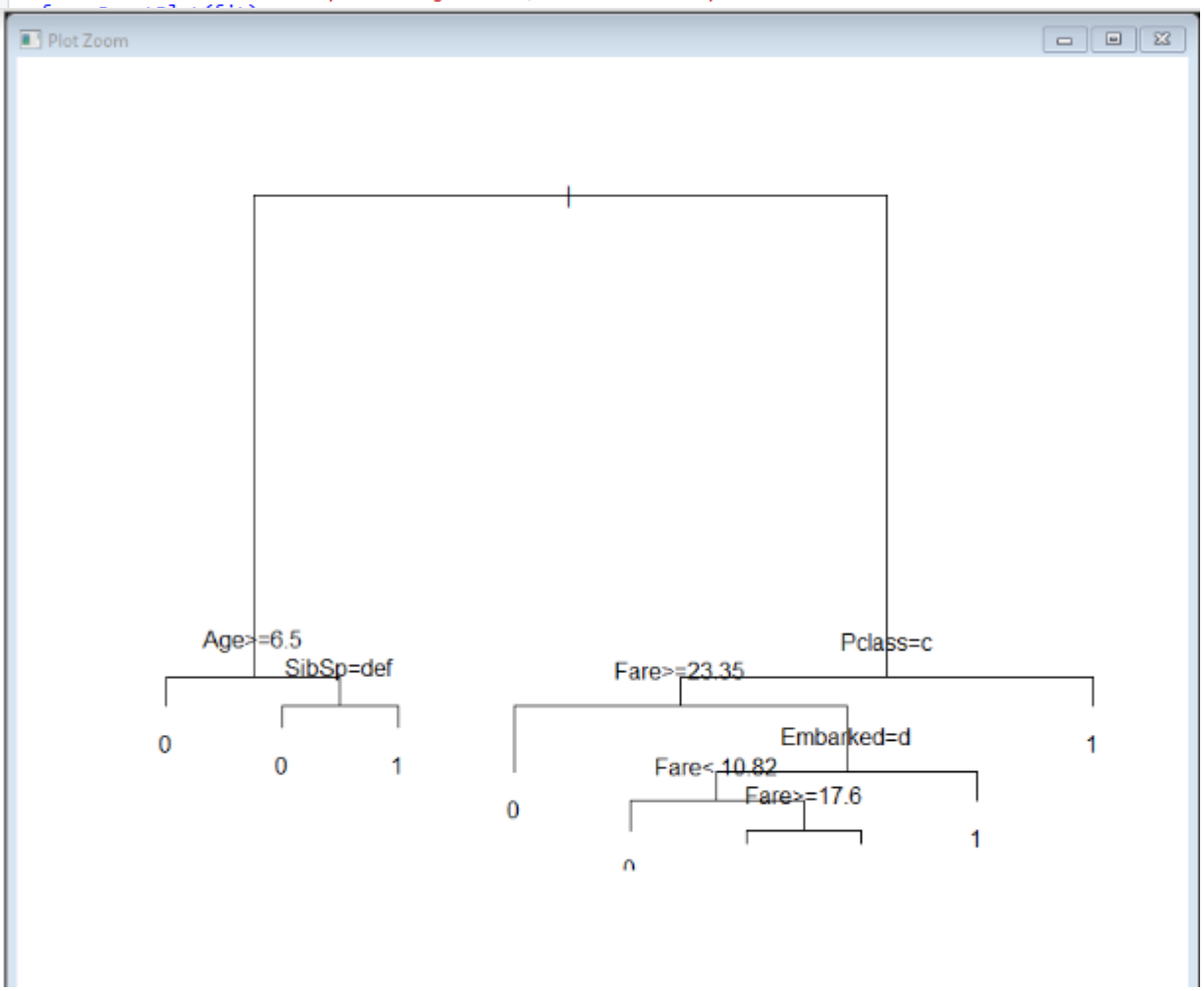
```
> titanic$Survived<-as.factor(titanic$Survived)
> titanic$SibSp<-as.factor(titanic$SibSp)
> titanic$Parch<-as.factor(titanic$Parch)
> titanic$Pclass<-as.factor(titanic$Pclass)
> titanic$Sex<-as.factor(titanic$Sex)
> titanic$Fare<-as.factor(titanic$Fare)
> titanic$Embarked<-as.factor(titanic$Embarked)
> summary(titanic$Survived)
  0   1
549 342
> names(titanic)
 [1] "PassengerId" "Survived"    "Pclass"      "Name"        "Sex"
 [6] "Age"         "SibSp"       "Parch"       "Ticket"      "Fare"
[11] "Cabin"       "Embarked"
> tree<-chaid(formula=Survived~Pclass+Sex+SibSp+Parch+Fare+Embarked,data=titanic)

> tree<-chaid(formula=Survived~Pclass+Sex+SibSp+Parch+Fare+Embarked,data=titanic)
> class(titanic$Survived)
[1] "factor"
> plot(tree, type="simple")
```

Output:



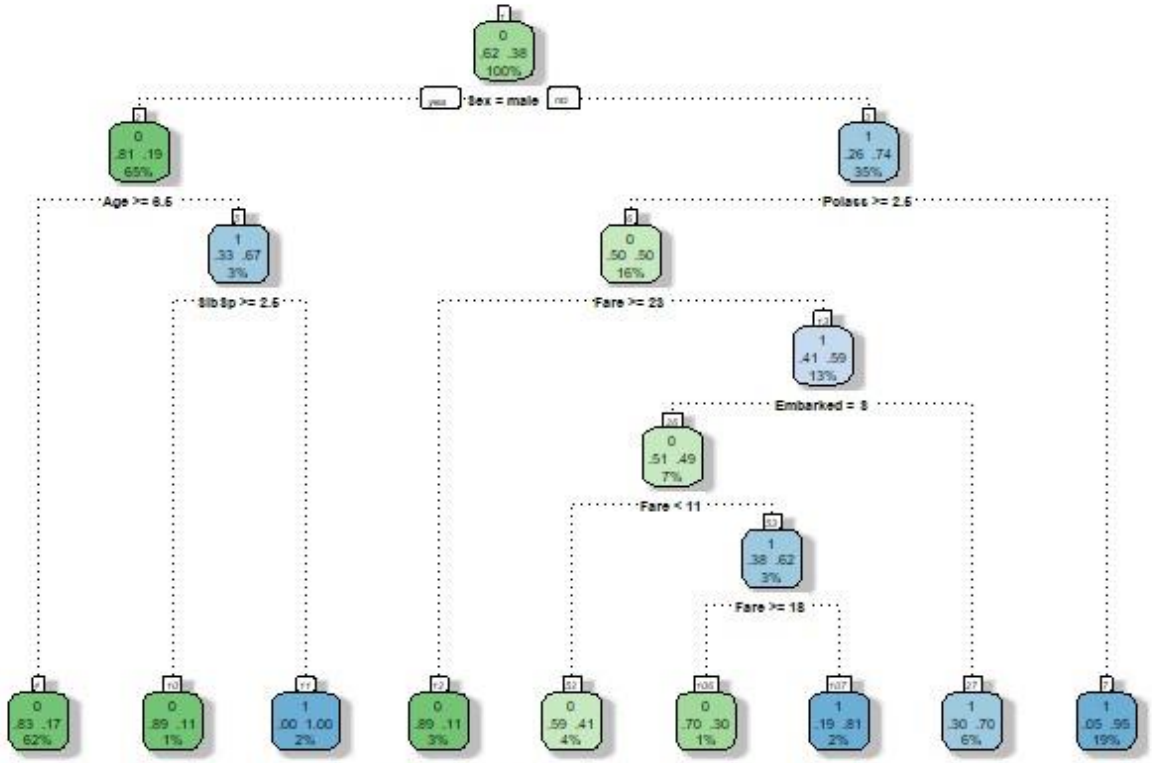
install.packages('rattle')
install.packages('rpart.plot')
install.packages('RColorBrewer')

```
> library(rpart.plot)
> fit <- rpart(Survived ~ Pclass + Sex + Age + SibSp + Parch + Fare + Embarked, data=titanic, method="class")
> plot(fit)
> text(fit)
Warning message:
In labels.rpart(x, minlength = minlength) :
  more than 52 levels in a predicting factor, truncated for printout
```



```
> library(rattle)
> library(rpart.plot)
> library(RColorBrewer)
> fancyRpartPlot(fit)
> Prediction <- predict(fit, titanic, type="class")
> Prediction
  1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19
  0   1   1   1   0   0   0   0   1   1   1   1   0   0   1   1   0   0   0
 20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36  37  38
  1   0   0   1   1   0   1   0   0   1   0   0   1   1   0   0   0   0   0
 39  40  41  42  43  44  45  46  47  48  49  50  51  52  53  54  55  56  57
  0   1   0   1   0   1   1   0   0   1   0   0   0   0   1   1   0   1   1
 58  59  60  61  62  63  64  65  66  67  68  69  70  71  72  73  74  75  76
  0   1   0   0   1   0   0   0   1   1   0   1   0   0   0   0   0   1   0
 77  78  79  80  81  82  83  84  85  86  87  88  89  90  91  92  93  94  95
  0   0   1   1   0   0   1   0   1   1   0   0   1   0   0   0   0   0   0
 96  97  98  99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114
  0   0   1   1   0   0   0   0   0   0   0   1   0   0   0   0   0   0   0
115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133
  0   0   0   0   0   0   0   0   0   1   0   1   0   1   1   0   0   0   1
134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152
  1   0   0   1   0   0   0   1   1   1   0   0   0   0   0   0   0   0   1
153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171
  0   0   0   0   1   0   0   0   0   1   0   0   0   1   1   0   0   1   0
172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190
  0   1   0   0   0   0   0   0   0   0   0   1   1   1   0   1   0   0   0
191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209
  1   0   1   0   1   1   0   0   1   1   0   0   0   0   0   0   0   0   1
210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228
  1   0   1   0   0   0   1   1   0   1   0   0   0   0   0   1   0   0   0
229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247
  0   0   1   0   0   1   0   1   0   1   0   0   0   1   0   0   0   1   0
```

```
  0   0   1   0   1   0   0   1   0   0   1   1   0   0   0   0   0   0   1
609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627
  1   1   0   0   1   0   0   1   0   1   1   0   0   1   1   0   0   0   0
628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646
  1   0   0   1   0   1   0   0   1   0   0   0   0   0   1   0   1   1   1
647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665
  0   1   0   1   0   1   0   1   0   0   0   1   0   0   1   0   0   0   0
666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684
  0   0   0   0   1   1   0   0   0   0   0   0   1   0   1   0   1   0   0
685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703
  1   0   0   0   0   1   1   1   1   0   0   0   0   1   1   0   1   1   0
704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722
  0   0   0   1   1   0   1   1   0   0   0   0   0   1   1   0   0   1   0
723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741
  0   0   0   0   1   1   0   1   1   0   0   0   0   0   0   1   0   0   1
742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760
  0   1   0   0   0   0   1   0   0   1   1   0   0   1   0   0   0   0   1
761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779
  0   0   0   1   0   1   0   1   0   0   0   0   1   0   1   0   0   1   0
780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798
  1   1   1   1   0   0   0   1   0   0   0   0   0   0   0   0   0   1   1
799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817
  0   0   0   1   1   1   0   0   0   0   0   1   0   0   0   0   0   0   1
818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836
  0   0   0   1   0   0   1   0   0   1   0   0   1   0   1   0   0   0   1
837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855
  0   0   1   0   0   0   1   0   0   0   0   0   0   1   0   0   1   1   1
856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874
  1   1   0   1   0   0   0   1   0   0   1   1   0   0   1   0   1   0   0
875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891
  1   1   0   0   0   1   1   0   0   0   0   0   0   1   0   1   0
Levels: 0 1
```



Rattle 2019-Mar-22 10:39:51 15DCS37

The purpose of this dataset is to predict which people are more likely to survive after the collision with the iceberg
The decision tree shows who survived and who did not based on the factors like age, sex, fare, embarked, etc.
The leaf nodes tell us the result:
0 means died
1 means survived.

for detailed explanation: https://www.guru99.com/r-decision-trees.html