While the class variable is limited to a manageable 3 values, the fare is again a continuous variable that needs to be reduced to something that can be easily tabulated. Let’s bin the fares into less than $10, between $10 and $20, $20 to $30 and more than $30 and store it to a new variable:

> train$Fare2 <- '30+'

> train$Fare2[train$Fare < 30 & train$Fare >= 20] <- '20-30'

> train$Fare2[train$Fare < 20 & train$Fare >= 10] <- '10-20'

> train$Fare2[train$Fare < 10] <- '<10'

Now let’s run a longer aggregate function to see if there’s anything interesting to work with here:

> aggregate(Survived ~ Fare2 + Pclass + Sex, data=train, FUN=function(x) {sum(x)/length(x)})

Fare2 Pclass Sex Survived

1 20-30 1 female 0.8333333

2 30+ 1 female 0.9772727

3 10-20 2 female 0.9142857

4 20-30 2 female 0.9000000

5 30+ 2 female 1.0000000

6 <10 3 female 0.5937500

7 10-20 3 female 0.5813953

8 20-30 3 female 0.3333333 \*\*

9 30+ 3 female 0.1250000 \*\*

10 <10 1 male 0.0000000

11 20-30 1 male 0.4000000

12 30+ 1 male 0.3837209

13 <10 2 male 0.0000000

14 10-20 2 male 0.1587302

15 20-30 2 male 0.1600000

16 30+ 2 male 0.2142857

17 <10 3 male 0.1115385

18 10-20 3 male 0.2368421

19 20-30 3 male 0.1250000

20 30+ 3 male 0.2400000