1. PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, Embarked.
2. PassengerId, Survived, Name, Sex, Ticket, Cabin, Embarked (Pclass is ordinal).
3. Discrete: SibSp, Parch. Continuous: Age, Fare.
4. Ticket, Cabin.
5. Training set: Age, Cabin, Embarked. Test set: Age, Fare, Cabin.
6. Integers: PassengerId, Survived, Pclass, SibSp, Parch. Floats: Age, Fare. Strings (objects): Name, Sex, Ticket, Cabin, Embarked.
7. Training set:

	Age	SibSp	Parch	Fare
count	714.000000	891.000000	891.000000	891.000000
mean	29.699118	0.523008	0.381594	32.204208
std	14.526497	1.102743	0.806057	49.693429
min	0.420000	0.000000	0.000000	0.000000
25%	20.125000	0.000000	0.000000	7.910400
50%	28.000000	0.000000	0.000000	14.454200
75%	38.000000	1.000000	0.000000	31.000000
max	80.000000	8.000000	6.000000	512.329200

Test set:

	Age	SibSp	Parch	Fare
count	332.000000	418.000000	418.000000	417.000000
mean	30.272590	0.447368	0.392344	35.627188
std	14.181209	0.896760	0.981429	55.907576
min	0.170000	0.000000	0.000000	0.000000
25%	21.000000	0.000000	0.000000	7.895800
50%	27.000000	0.000000	0.000000	14.454200
75%	39.000000	1.000000	0.000000	31.500000
max	76.000000	8.000000	9.000000	512.329200

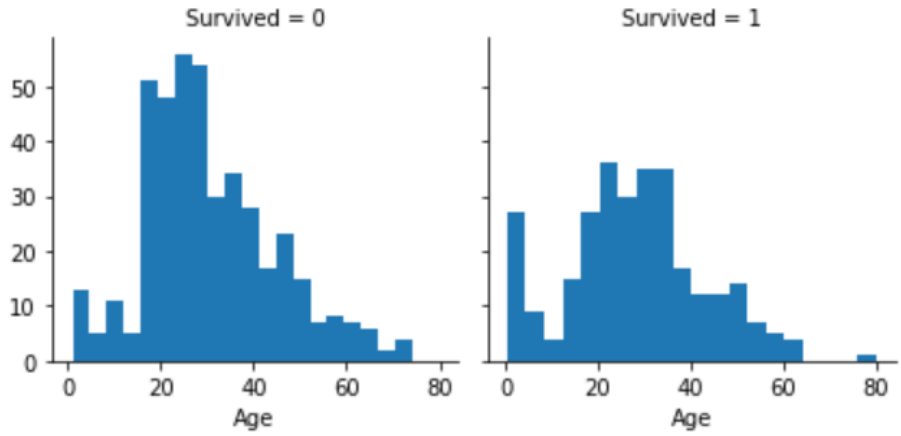
8. Training set:

	PassengerId	Survived	Pclass	Name	Sex	Ticket	Cabin	Embarked
count	891	891	891	891	891	891	204	889
unique	891	2	3	891	2	681	147	3
top	891	0	3	Troutt, Miss. Edwina Celia "Winnie"	male	1601	G6	S
freq	1	549	491	1	577	7	4	644

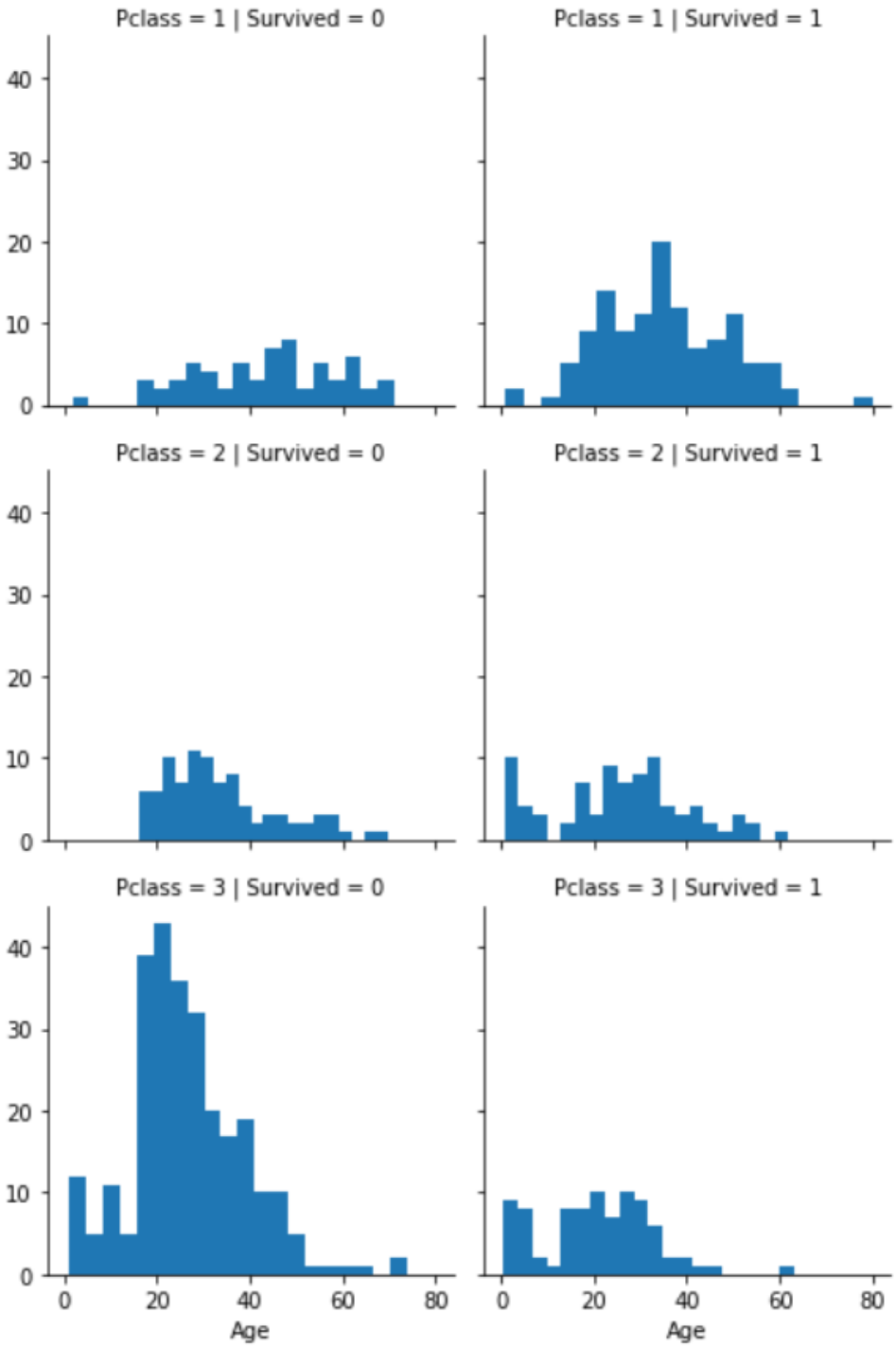
Test set:

	PassengerId	Pclass	Name	Sex	Ticket	Cabin	Embarked
count	418	418	418	418	418	91	418
unique	418	3	418	2	363	76	3
top	1	3	Ryerson, Master. John Borie	male	PC 17608	B57 B59 B63 B66	S
freq	1	236	1	266	5	3	270

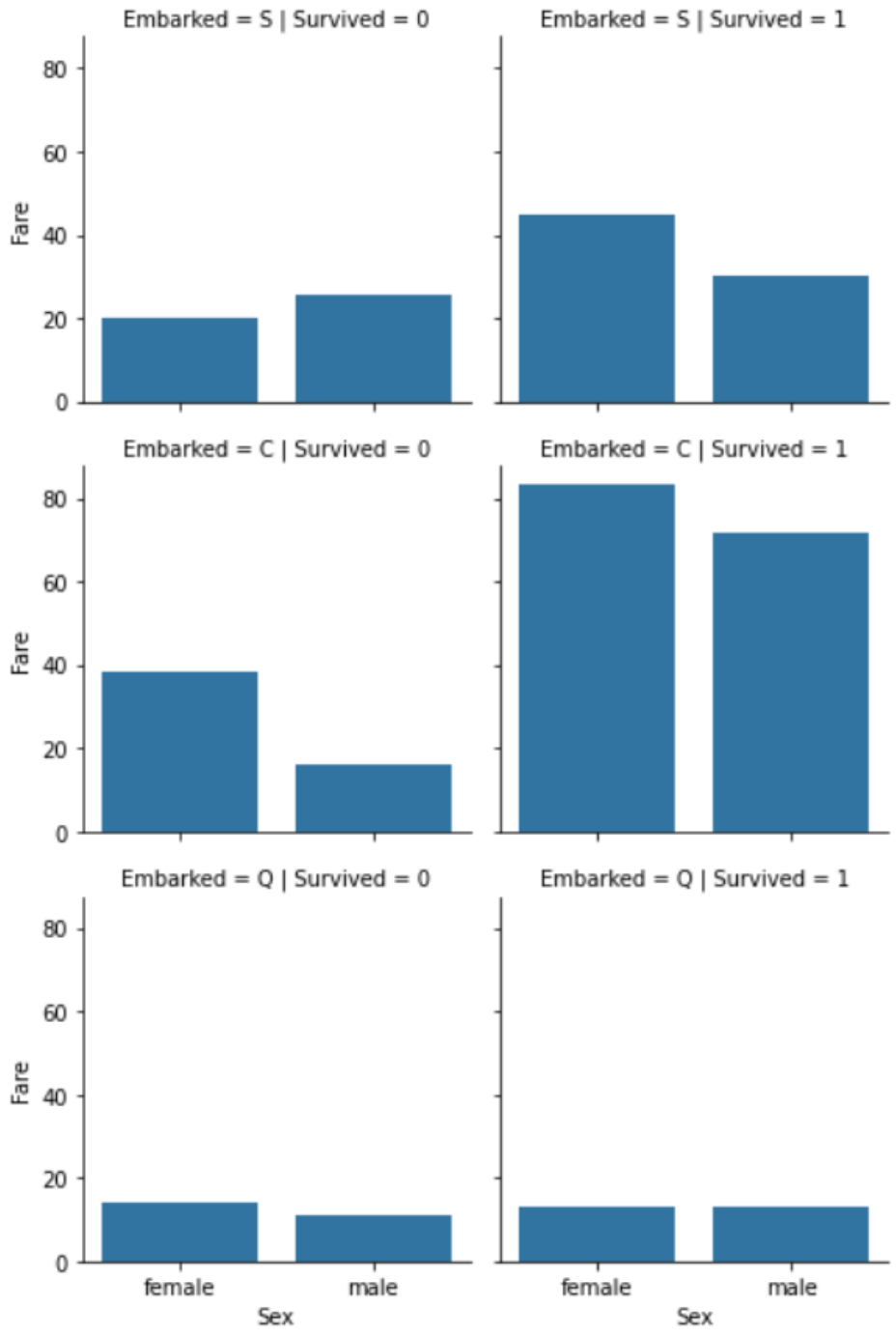
9. Approximately 40% of all survivors were in the first passenger class. This indicates that there is not a significant correlation between the survival rate and passenger class, and so I will not include this feature in the predictive model.
10. Approximately 68% of all survivors were female which indicates that women were more likely to have survived.
11. Infants have a high survival rate. Oldest passengers do survive. Many 15-25-year-olds do not survive. We should include Age in our model training since the distributions shown in the histograms are significantly different from each other, which implies that the information is useful. We should band age groups because the survival rates in each histogram are not equally distributed (most of the people that don't survive, for example, are 15-40 years old).



12. Pclass=3 had the most passengers; however, most did not survive. Infants in Pclass=1 and Pclass=2 mostly survive. Most passengers in Pclass=1 survive. Pclass varies in terms of Age distribution of passengers. We should consider Pclass for model training since there is a strong connection between Pclass and survival rate across different age groups.



13. Higher fare paying passengers have better survival. We should consider banding the Fare feature.



14. From the training set table in number 8 above, we can use the count and unique rows to calculate the rate of duplication for the Ticket feature: $((891 - 681)/891) * 100\% = 23.6\%$. Since there are a lot of duplicates, there is not much of a correlation between Ticket and survival, and so we should drop the Ticket feature.

15. The cabin feature is not complete. There are $891 + 418 - 204 - 91 = 1014$ null values in the Cabin features of the combined dataset of training and test dataset. Since there are so many missing values, we should drop the Cabin feature.

Associated code can be found here.