

# Modeling Titanic Survivability with Binary Logistic Regression



Matthew Melendez

# Data

Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	1	Cavendish, Mrs. Tyrell William (Julia Florence... Wick, Mr. George Dennick	female male	76.0 57.0	1 1	0 1	19877 36928	78.8500 164.8667	C46 NaN	S S
0	2	Nasser, Mr. Nicholas	male	32.5	1	0	237736	30.0708	NaN	C
0	2	Berriman, Mr. William John	male	23.0	0	0	28425	13.0000	NaN	S
0	2	Harris, Mr. Walter	male	30.0	0	0	W/C 14208	10.5000	NaN	S
...	...	...	...	...	...	...	...	...	...	...
1	2	Corey, Mrs. Percy C (Mary Phyllis Elizabeth Mi... Goldsmith, Master. Frank John William "Frankie"	female male	NaN 9.0	0 0	0 2	F.C.C. 13534 363291	21.0000 20.5250	NaN NaN	S S
0	3	Betros, Mr. Tannous	male	20.0	0	0	2648	4.0125	NaN	C
0	3	Pettersson, Miss. Ellen Natalia	female	18.0	0	0	347087	7.7750	NaN	S
0	2	Rogers, Mr. Reginald Harry	male	19.0	0	0	28004	10.5000	NaN	S

# Splitting up the data

TOTAL DATA SET

TEST

INITIAL TRAINING SET

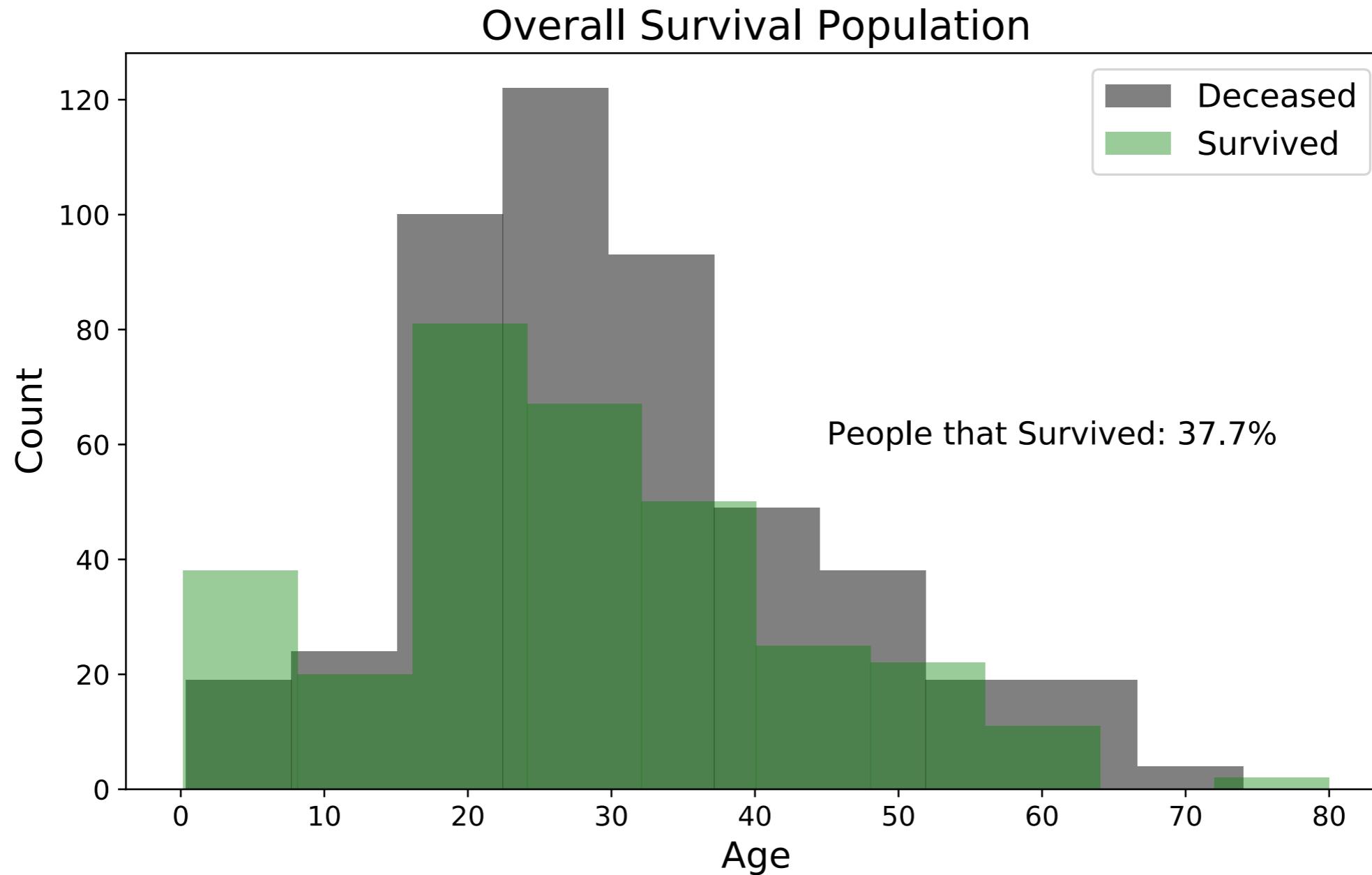
VALIDATE

TEST

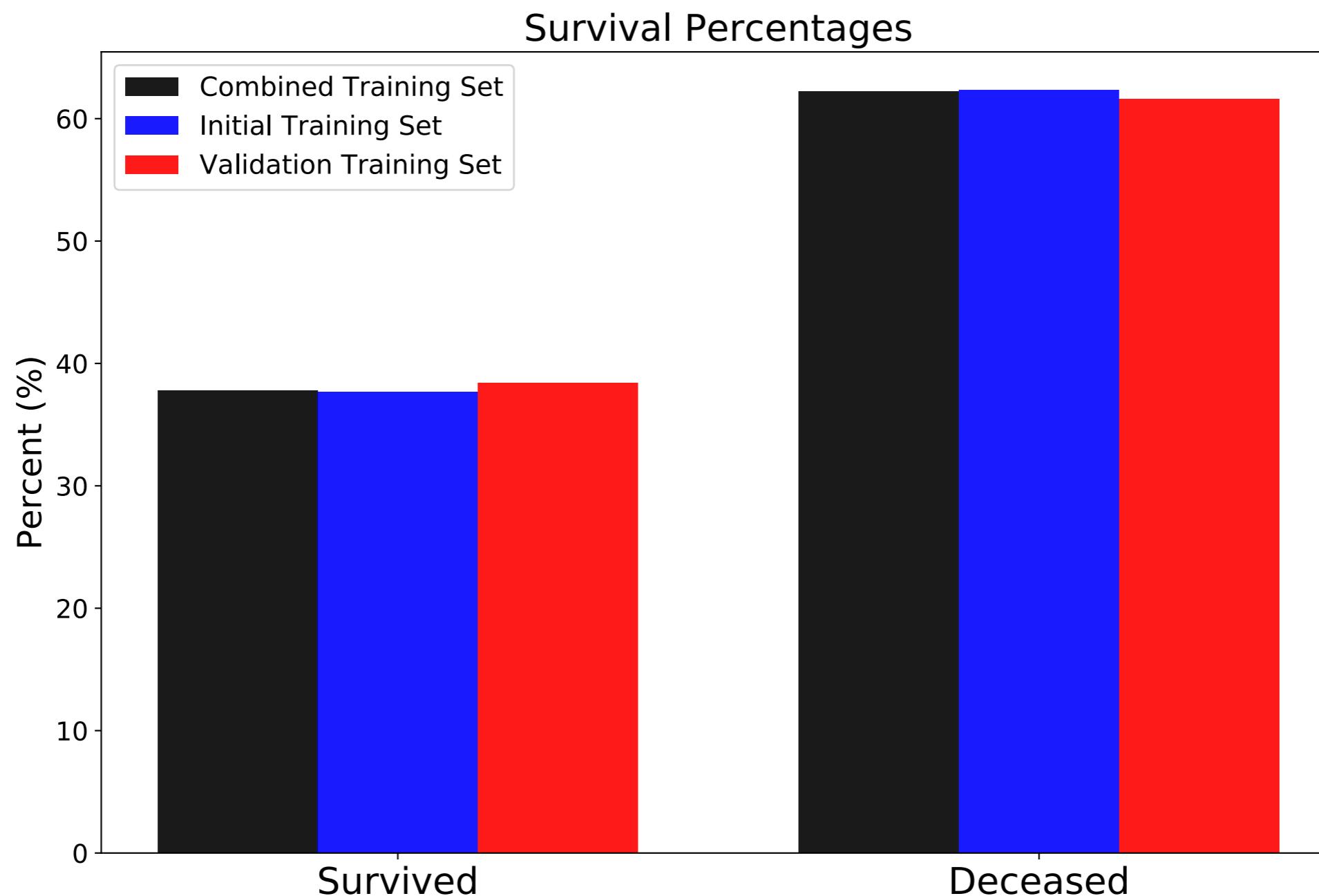
FINAL TRAINING SET

TEST

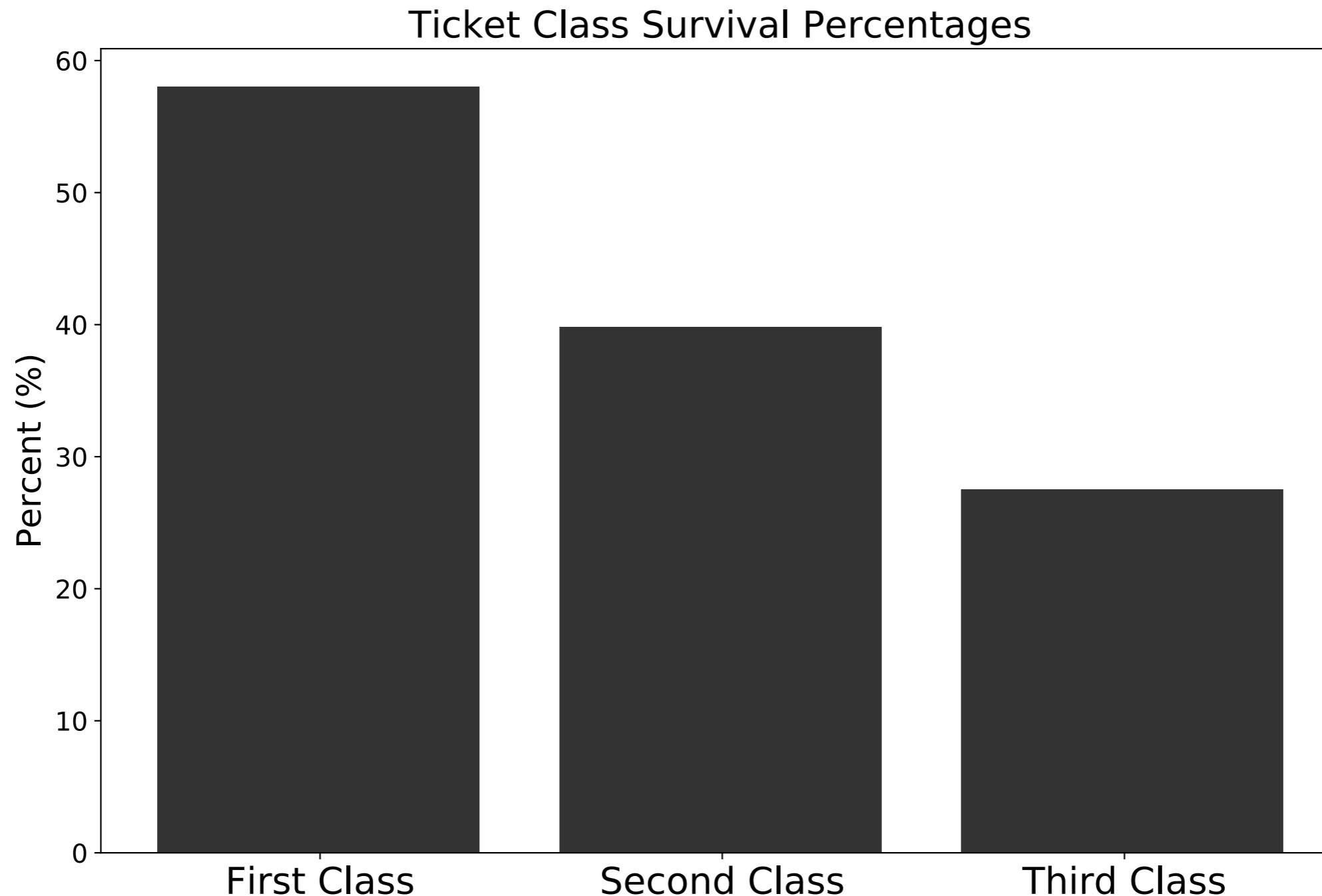
# Data: Survival Rate



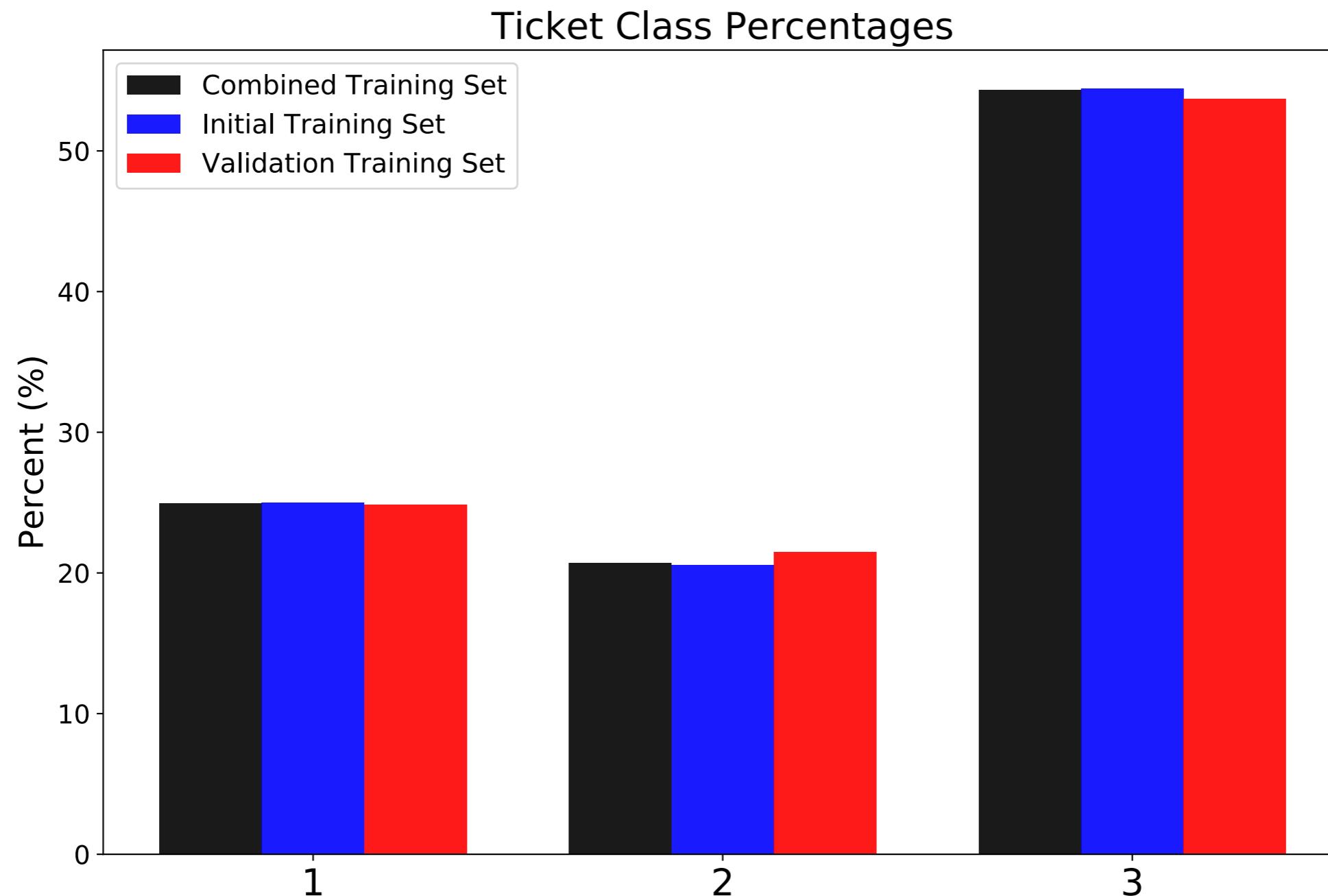
# Data: Survival Rate



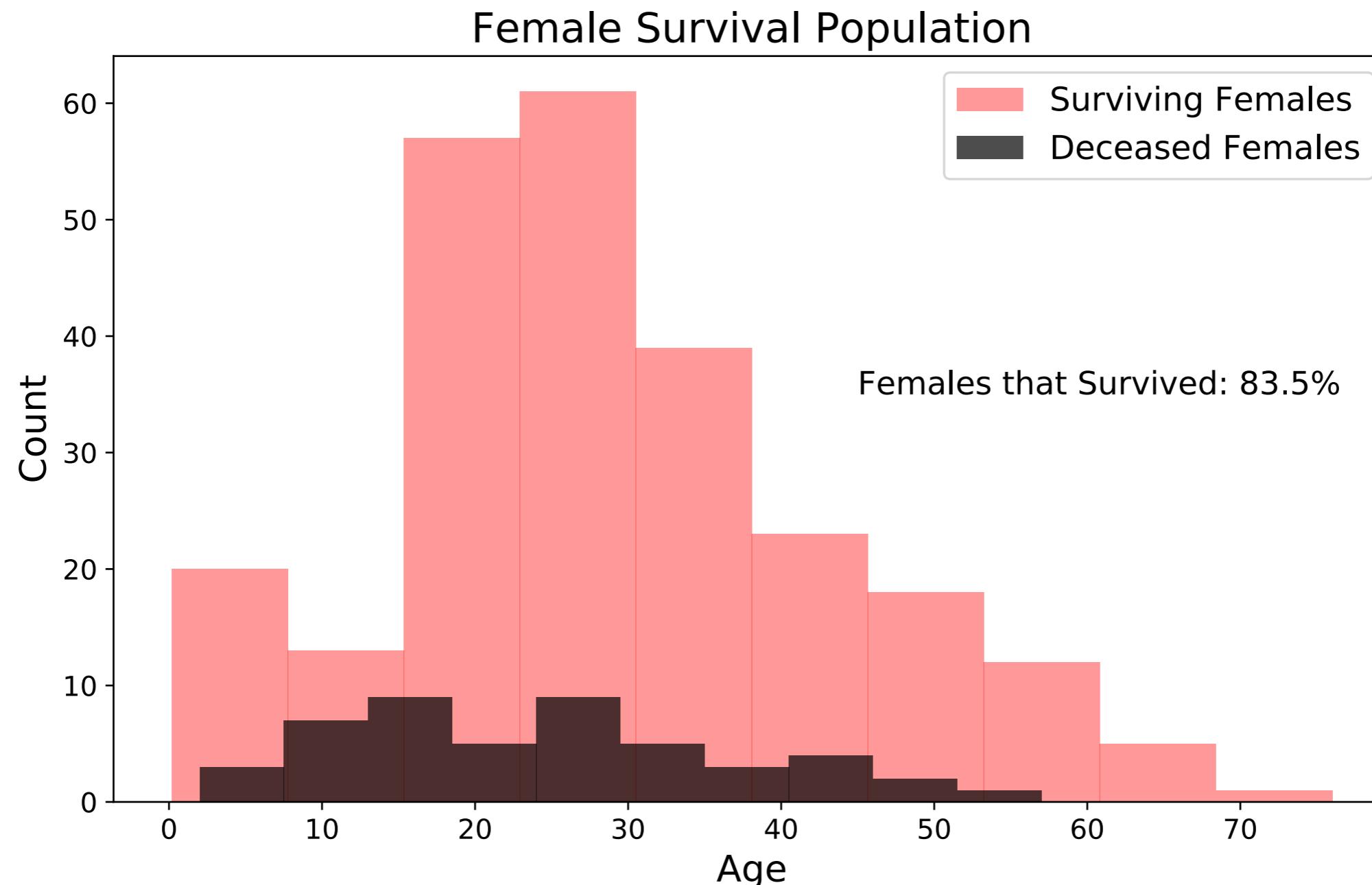
# Data: Ticket Class



# Data: Ticket Class

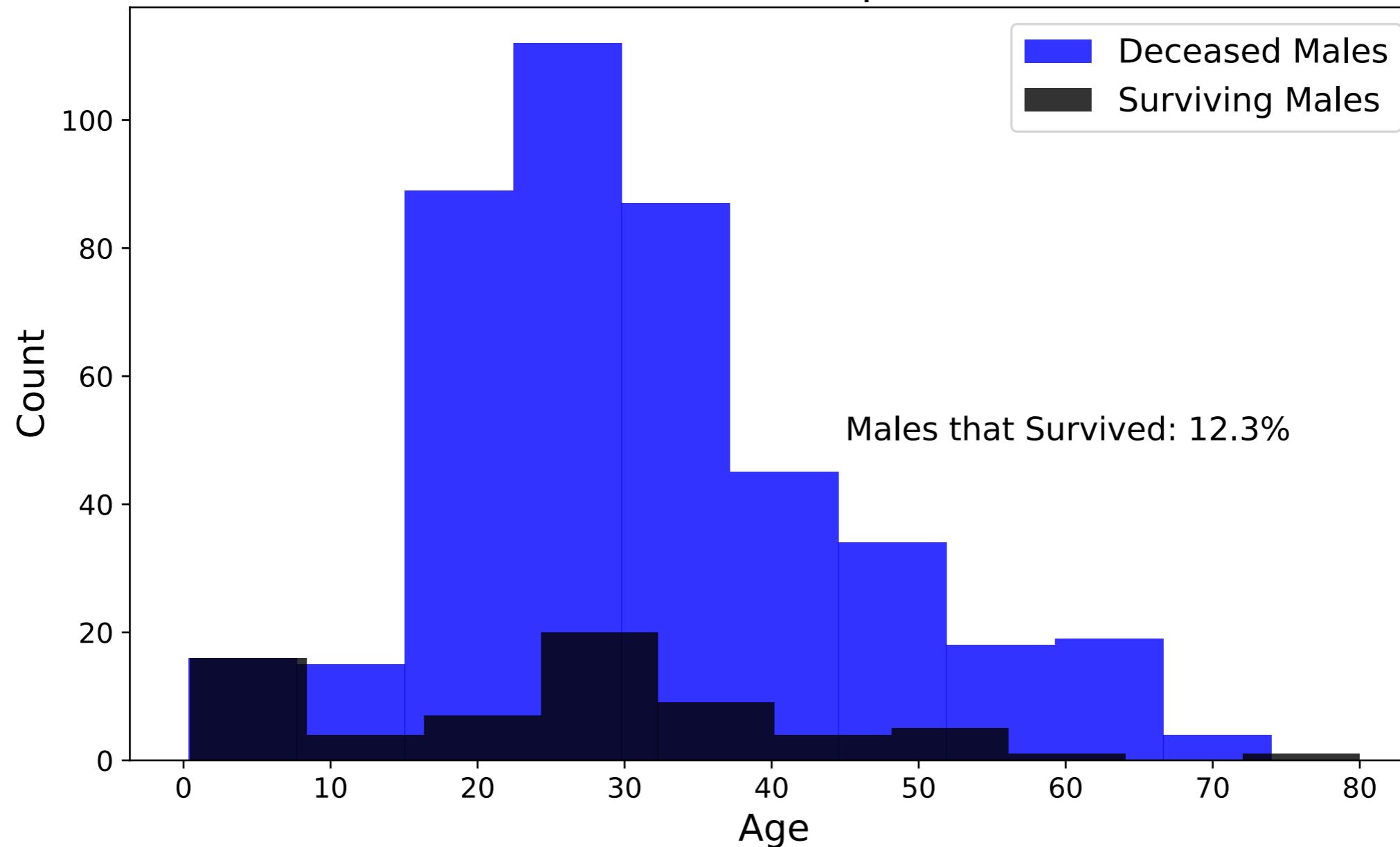


# Data: Gender

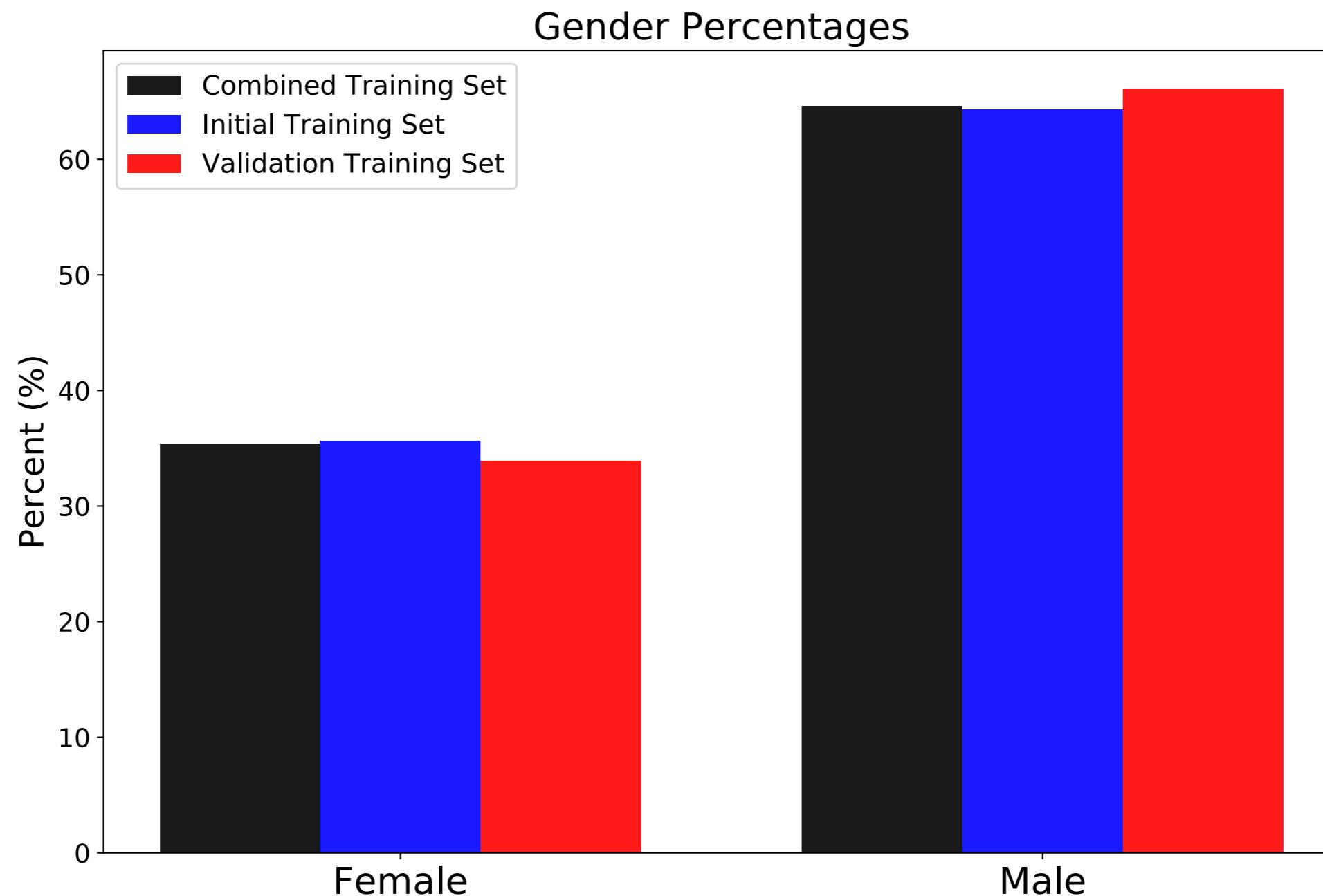


# Data: Gender

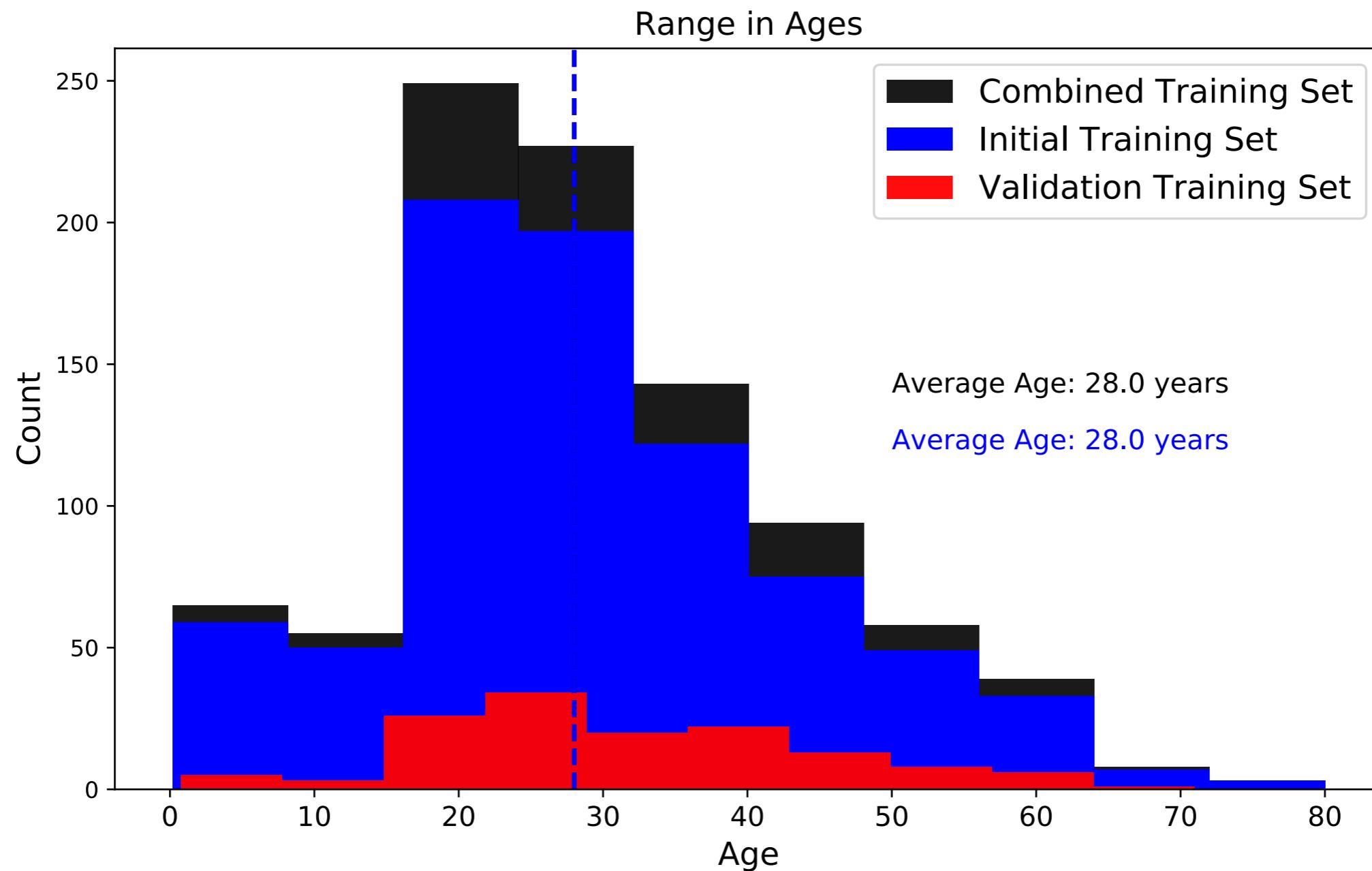
Male Survival Population



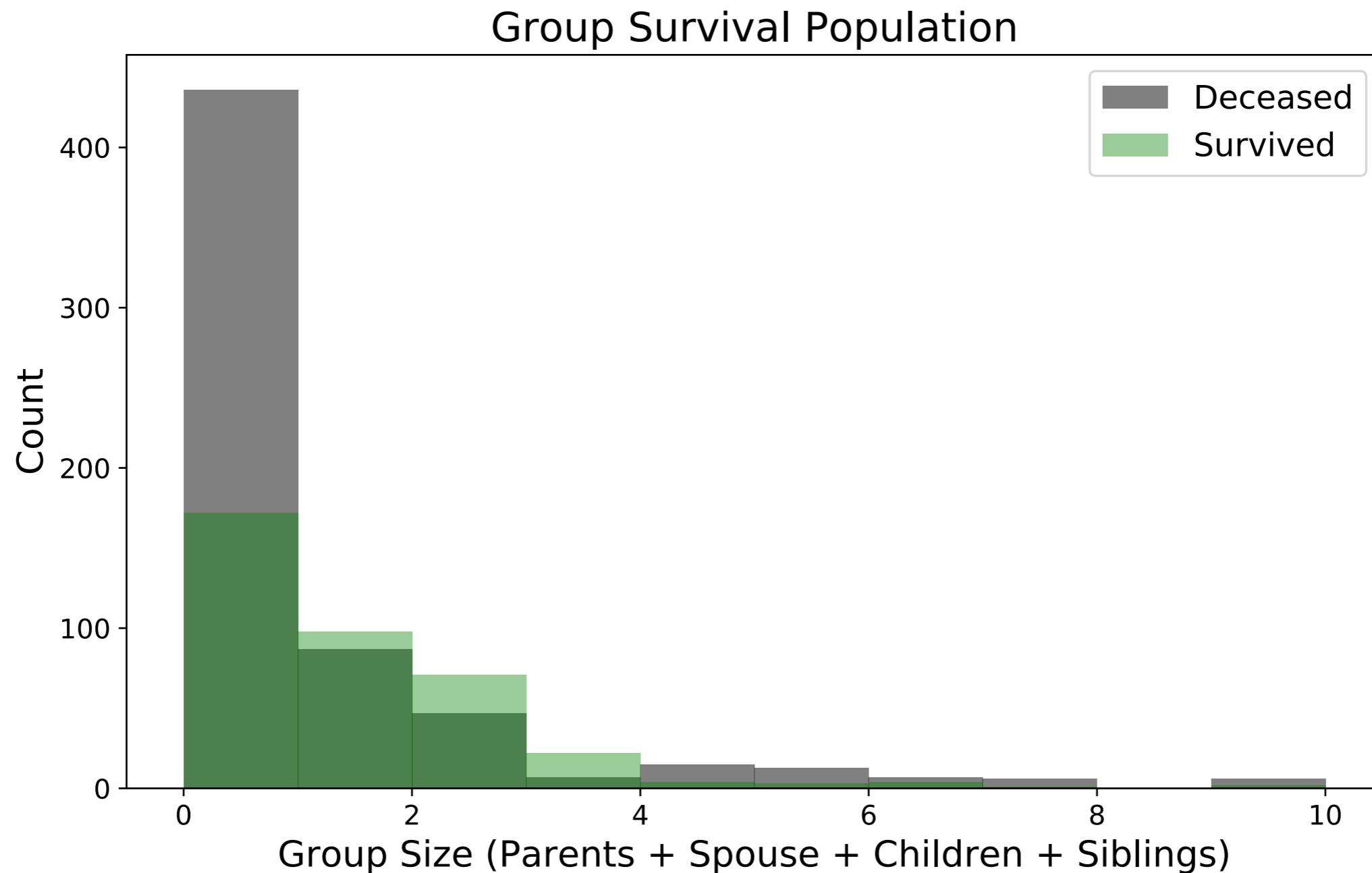
# Data: Gender



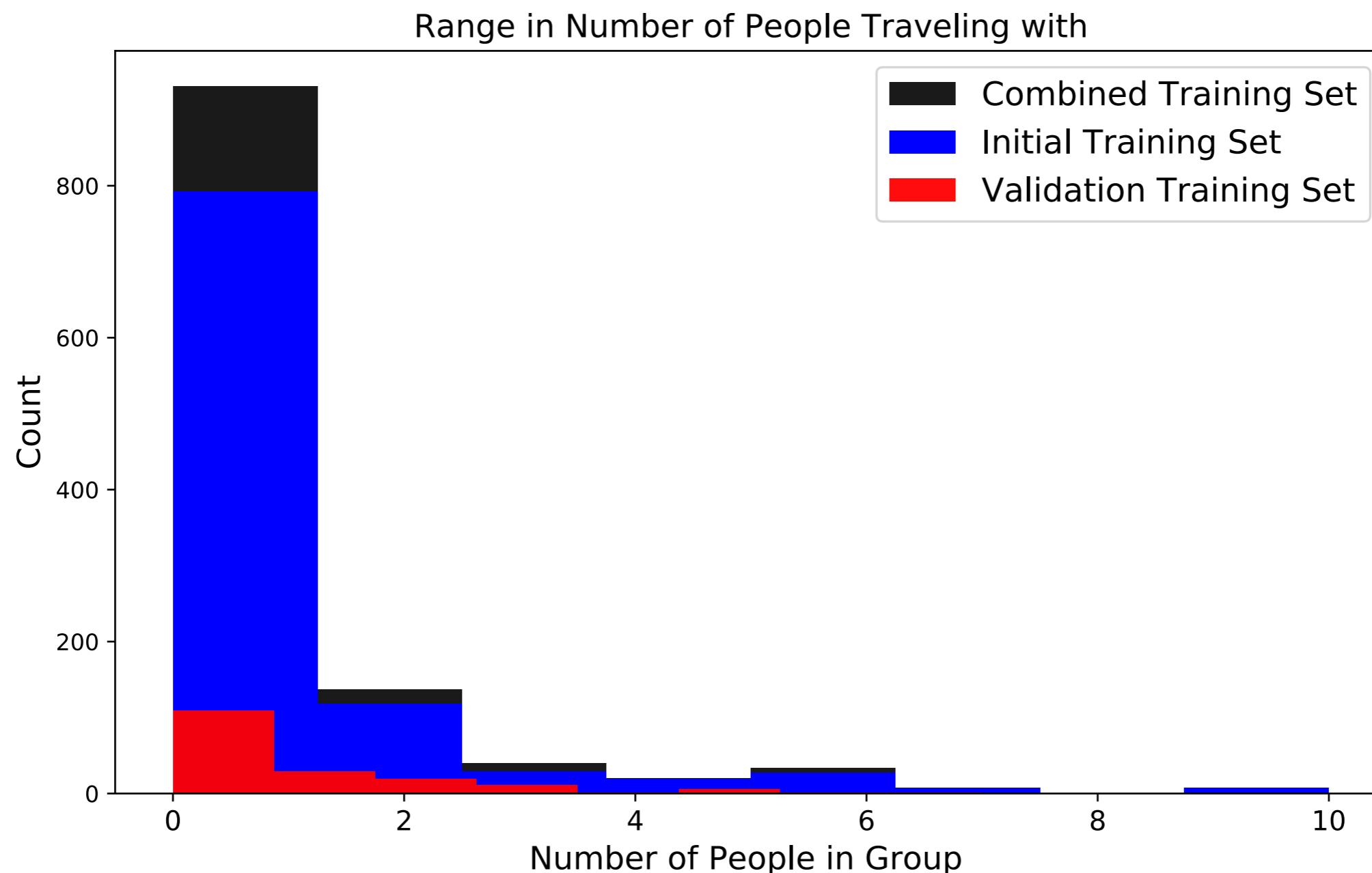
# Data: Age



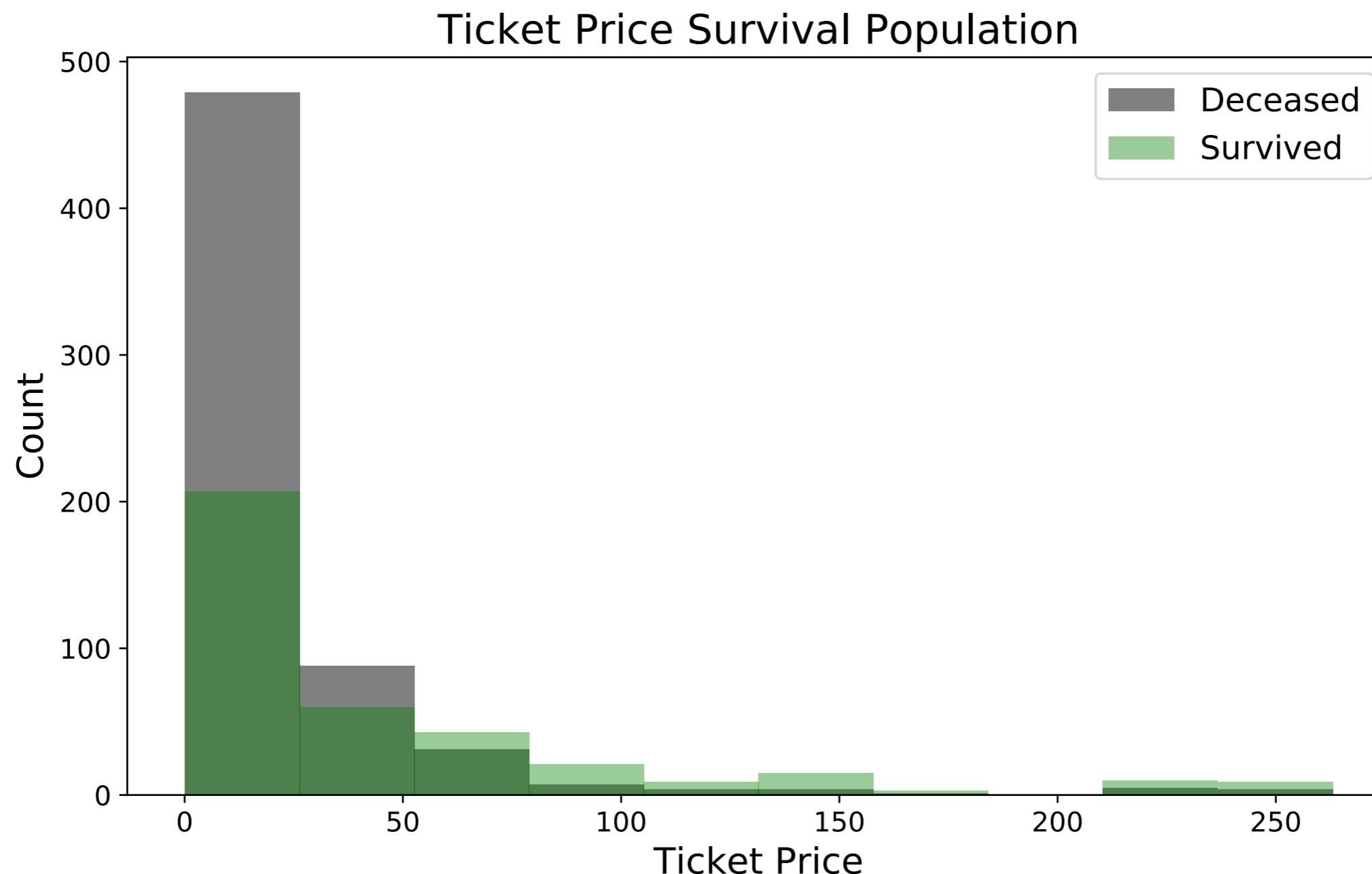
# Data: Group Size



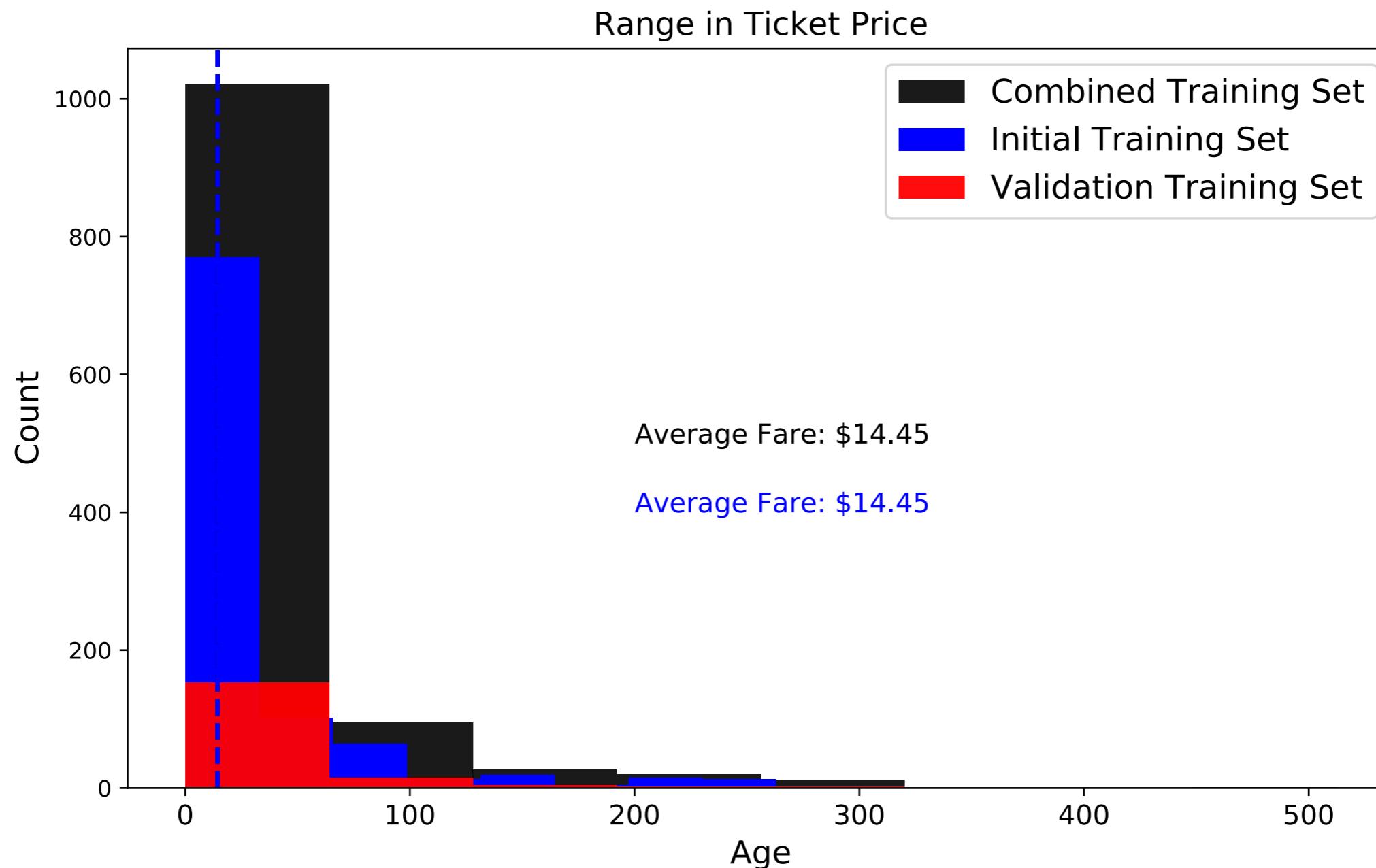
# Data: Group Size



# Data: Ticket Price



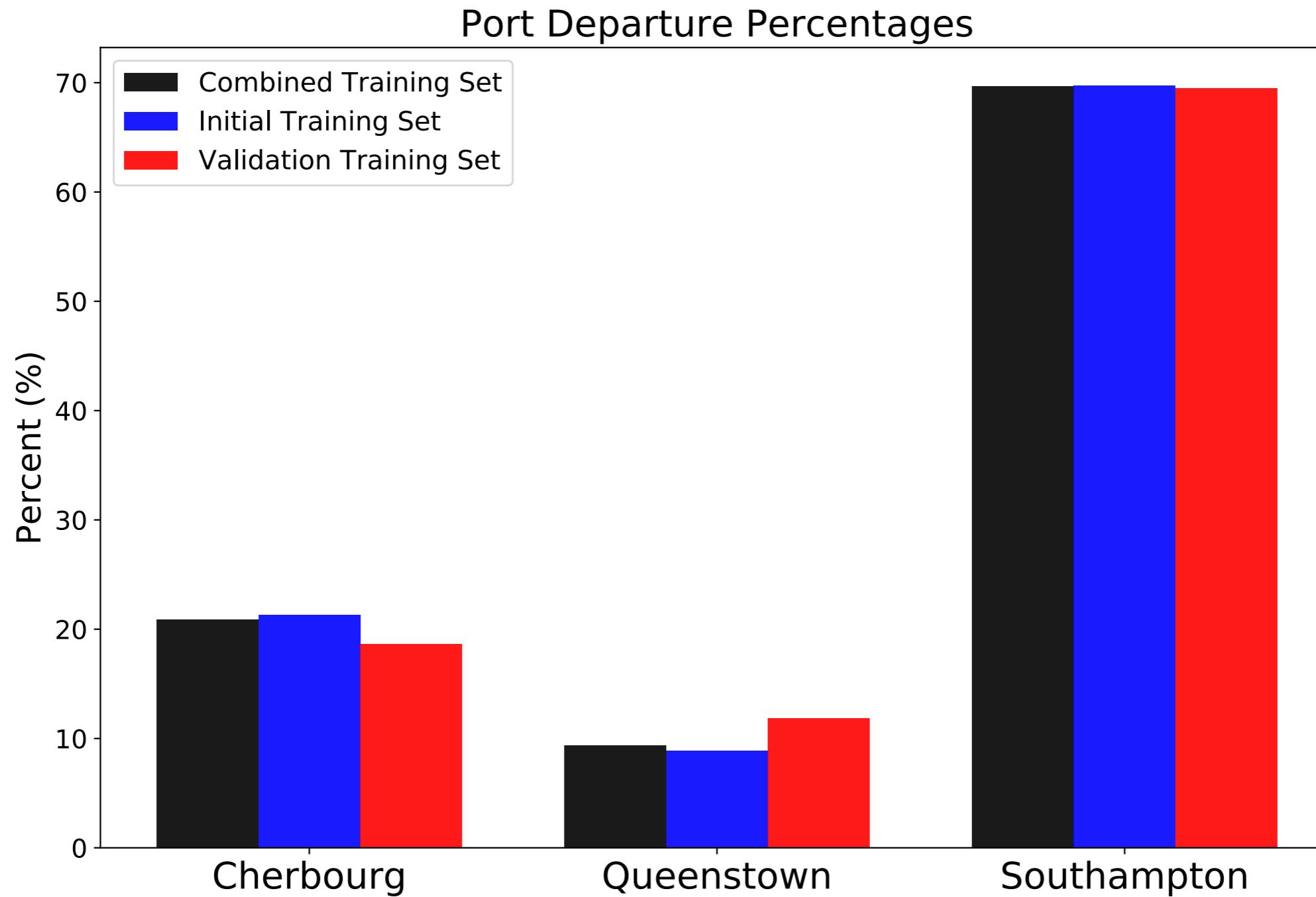
# Data: Ticket Price



# Data: Embarked

- Port C Survival Rate: 50.7%
  - 52.6% of those that survived were in Class 1
  - 12.2% in Class 2
  - 35.2% in Class 3
- Port Q Survival Rate: 46.1%
  - 3.4% of those that survived were in Class 1
  - 4.5% in Class 2
  - 92.1% in Class 3
- Port S Survival Rate: 35.5%
  - 19.2% of those that survived were in Class 1
  - 25.2% in Class 2
  - 55.6% in Class 3

# Data: Embarked



# Data Preparation

Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	1	Cavendish, Mrs. Tyrell William (Julia Florence... Wick, Mr. George Dennick	female male	76.0 57.0	1 1	0 1	19877 36928	78.8500 164.8667	C46 NaN	S S
0	2	Nasser, Mr. Nicholas	male	32.5	1	0	237736	30.0708	NaN	C
0	2	Berriman, Mr. William John	male	23.0	0	0	28425	13.0000	NaN	S
0	2	Harris, Mr. Walter	male	30.0	0	0	W/C 14208	10.5000	NaN	S
...	...	...	...	...	...	...	...	...	...	...
1	2	Corey, Mrs. Percy C (Mary Phyllis Elizabeth Mi... Goldsmith, Master. Frank John William "Frankie"	female male	NaN 9.0	0 0	0 2	F.C.C. 13534 363291	21.0000 20.5250	NaN NaN	S S
0	3	Betros, Mr. Tannous	male	20.0	0	0	2648	4.0125	NaN	C
0	3	Pettersson, Miss. Ellen Natalia	female	18.0	0	0	347087	7.7750	NaN	S
0	2	Rogers, Mr. Reginald Harry	male	19.0	0	0	28004	10.5000	NaN	S

# Data Preparation

Missing Data

Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	1	Cavendish, Mrs. Tyrell William (Julia Florence... Wick, Mr. George Dennick	female male	76.0 57.0	1 1	0 1	19877 36928	78.8500 164.8667	C46 NaN	S S
0	2	Nasser, Mr. Nicholas	male	32.5	1	0	237736	30.0708	NaN	C
0	2	Berriman, Mr. William John	male	23.0	0	0	28425	13.0000	NaN	S
0	2	Harris, Mr. Walter	male	30.0	0	0	W/C 14208	10.5000	NaN	S
...	...	...	...	...	...	...	...	...	...	...
1	2	Corey, Mrs. Percy C (Mary Phyllis Elizabeth Mi... Goldsmith, Master. Frank John William "Frankie"	female male	NaN 9.0	0 0	0 2	F.C.C. 13534 363291	21.0000 20.5250	NaN NaN	S S
0	3	Betros, Mr. Tannous	male	20.0	0	0	2648	4.0125	NaN	C
0	3	Pettersson, Miss. Ellen Natalia	female	18.0	0	0	347087	7.7750	NaN	S
0	2	Rogers, Mr. Reginald Harry	male	19.0	0	0	28004	10.5000	NaN	S

# Data Preparation

Combine



Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	1	Cavendish, Mrs. Tyrell William (Julia Florence... Wick, Mr. George Dennick	female male	76.0 57.0	1 1	0 1	19877 36928	78.8500 164.8667	C46 NaN	S S
0	2	Nasser, Mr. Nicholas	male	32.5	1	0	237736	30.0708	NaN	C
0	2	Berriman, Mr. William John	male	23.0	0	0	28425	13.0000	NaN	S
0	2	Harris, Mr. Walter	male	30.0	0	0	W/C 14208	10.5000	NaN	S
...	...	...	...	...	...	...	...	...	...	...
1	2	Corey, Mrs. Percy C (Mary Phyllis Elizabeth Mi... Goldsmith, Master. Frank John William "Frankie"	female male	NaN 9.0	0 0	0 2	F.C.C. 13534 363291	21.0000 20.5250	NaN NaN	S S
0	3	Betros, Mr. Tannous	male	20.0	0	0	2648	4.0125	NaN	C
0	3	Pettersson, Miss. Ellen Natalia	female	18.0	0	0	347087	7.7750	NaN	S
0	2	Rogers, Mr. Reginald Harry	male	19.0	0	0	28004	10.5000	NaN	S

# Data Preparation

Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	1	Cavendish, Mrs. Tyrell William (Julia Florence... Wick, Mr. George Dennick	female male	76.0 57.0	1 1	0 1	19877 36928	78.8500 164.8667	C46 NaN	S S
0	2	Nasser, Mr. Nicholas	male	32.5	1	0	237736	30.0708	NaN	C
0	2	Berriman, Mr. William John	male	23.0	0	0	28425	13.0000	NaN	S
0	2	Harris, Mr. Walter	male	30.0	0	0	W/C 14208	10.5000	NaN	S
...	...	...	...	...	...	...	...	...	...	...
1	2	Corey, Mrs. Percy C (Mary Phyllis Elizabeth Mi... Goldsmith, Master. Frank John William "Frankie"	female male	NaN 9.0	0 0	0 2	F.C.C. 13534 363291	21.0000 20.5250	NaN NaN	S S
0	3	Betros, Mr. Tannous	male	20.0	0	0	2648	4.0125	NaN	C
0	3	Pettersson, Miss. Ellen Natalia	female	18.0	0	0	347087	7.7750	NaN	S
0	2	Rogers, Mr. Reginald Harry	male	19.0	0	0	28004	10.5000	NaN	S

# Data Preparation

Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	1	Cavendish, Mrs. Tyrell William (Julia Florence... Wick, Mr. George Dennick	female male	76.0 57.0	1 1	0 1	19377 36928	78.8500 164.8667	C46 NaN	S S
0	2	Nasser, Mr. Nicholas	male	32.5	1	0	237736	30.0708	NaN	C
0	2	Berriman, Mr. William John	male	23.0	0	0	28425	13.0000	NaN	S
0	2	Harris, Mr. Walter	male	30.0	0	0	W/C 14208	10.5000	NaN	S
...	...	...	...	...	...	...	...	...	...	...
1	2	Corey, Mrs. Percy C (Mary Phyllis Elizabeth Mi... Goldsmith, Master. Frank John William "Frankie"	female male	NaN 9.0	0 0	0 2	F.C.C. 13534 363291	21.0000 20.5250	NaN NaN	S S
0	3	Betros, Mr. Tannous	male	20.0	0	0	2648	4.0125	NaN	C
0	3	Pettersson, Miss. Ellen Natalia	female	18.0	0	0	347087	7.7750	NaN	S
0	2	Rogers, Mr. Reginald Harry	male	19.0	0	0	28004	10.5000	NaN	S

# Data Preparation

**Dummy Variables**

Survived	Pclass	Sex	Age	Fare	Embarked	Group
1	1	female	76.0	78.8500	S	0
0	1	male	57.0	164.8667	S	1
0	2	male	32.5	30.0708	C	0
0	2	male	23.0	13.0000	S	0
0	2	male	30.0	10.5000	S	0
...	...	...	...	...	...	...
1	2	female	NaN	21.0000	S	0
1	3	male	9.0	20.5250	S	2
0	3	male	20.0	4.0125	C	0
0	3	female	18.0	7.7750	S	0
0	2	male	19.0	10.5000	S	0

# Data Preparation

Normalize

Survived	Pclass	Sex	Age	Fare	Embarked	Group
1	1	female	76.0	78.8500	S	0
0	1	male	57.0	164.8667	S	1
0	2	male	32.5	30.0708	C	0
0	2	male	23.0	13.0000	S	0
0	2	male	30.0	10.5000	S	0
...	...	...	...	...	...	...
1	2	female	NaN	21.0000	S	0
1	3	male	9.0	20.5250	S	2
0	3	male	20.0	4.0125	C	0
0	3	female	18.0	7.7750	S	0
0	2	male	19.0	10.5000	S	0

# Data Preparation

Survived	Age	Fare	GroupSize	Pclass_1	Pclass_2	Pclass_3	Sex_female	Sex_male	Embarked_C	Embarked_Q	Embarked_S
1.0	0.95000	0.299810	0.1	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0
0.0	0.71250	0.626870	0.2	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0
0.0	0.40625	0.114338	0.1	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0
0.0	0.28750	0.049430	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0
0.0	0.37500	0.039924	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0
...	...	...	...	...	...	...	...	...	...	...	...
1.0	0.35000	0.079848	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0
1.0	0.11250	0.078042	0.2	0.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0
0.0	0.25000	0.015257	0.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	0.0
0.0	0.22500	0.029563	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0
0.0	0.23750	0.039924	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0

# Data Preparation

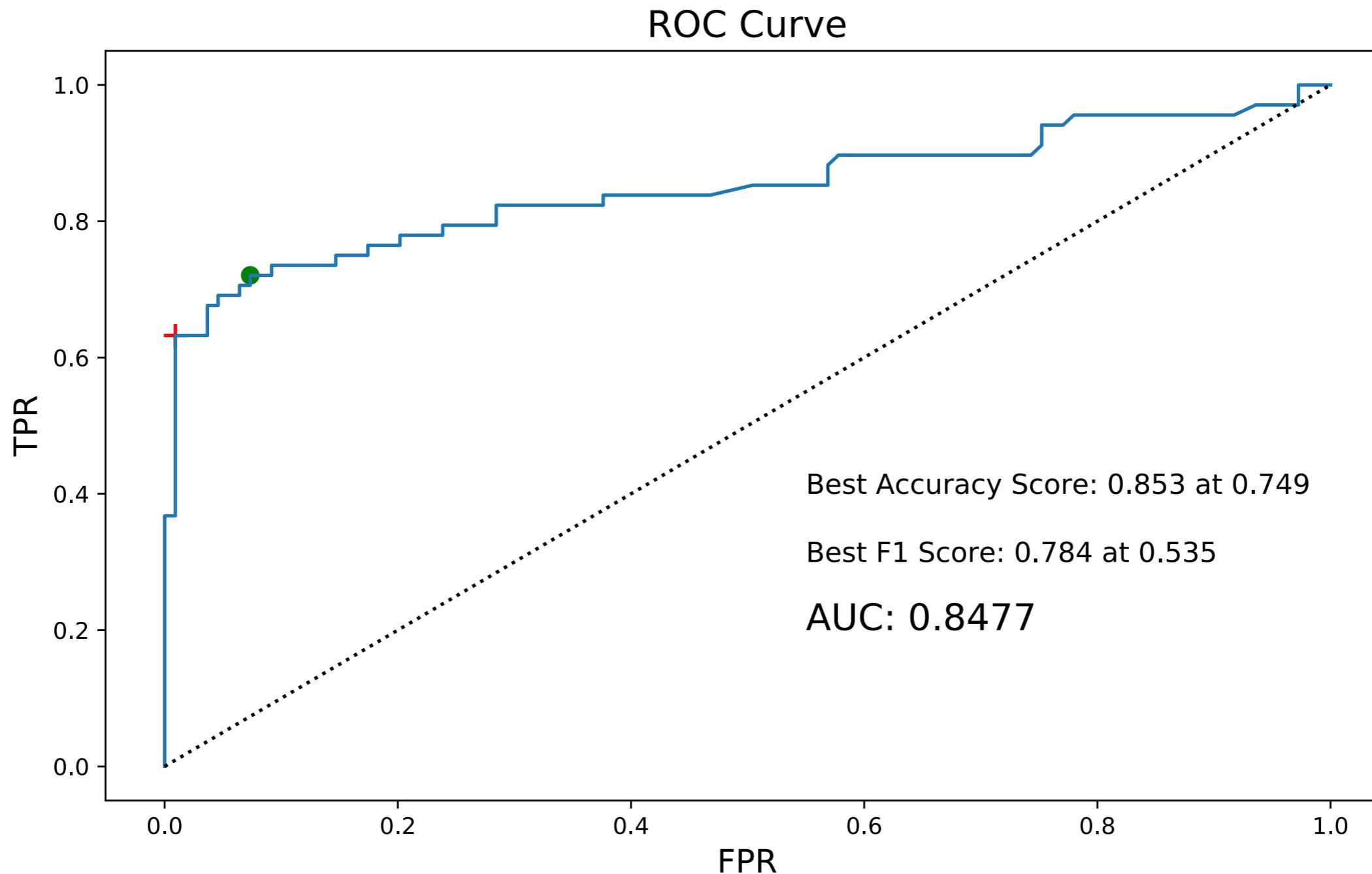
Age	Fare	GroupSize	Pclass_1	Pclass_2	Pclass_3	Sex_female	Sex_male	Embarked_C	Embarked_Q	Embarked_S
0.95000	0.299810	0.1	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0
0.71250	0.626870	0.2	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0
0.40625	0.114338	0.1	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0
0.28750	0.049430	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0
0.37500	0.039924	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0
...	...	...	...	...	...	...	...	...	...	...
0.35000	0.079848	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0
0.11250	0.078042	0.2	0.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0
0.25000	0.015257	0.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	0.0
0.22500	0.029563	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0
0.23750	0.039924	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0

Did this for all Data Sets

# Binary Logistic Regression Model

- 76.5% of the total data was used to train the model (initial training data)
  - 10,000 iterations
  - Initial  $\alpha = 0.0$
  - Initial  $\beta = 0.3$
- Model then used to Predict the 13.5% of the total data that made up the Validation data (validation data)

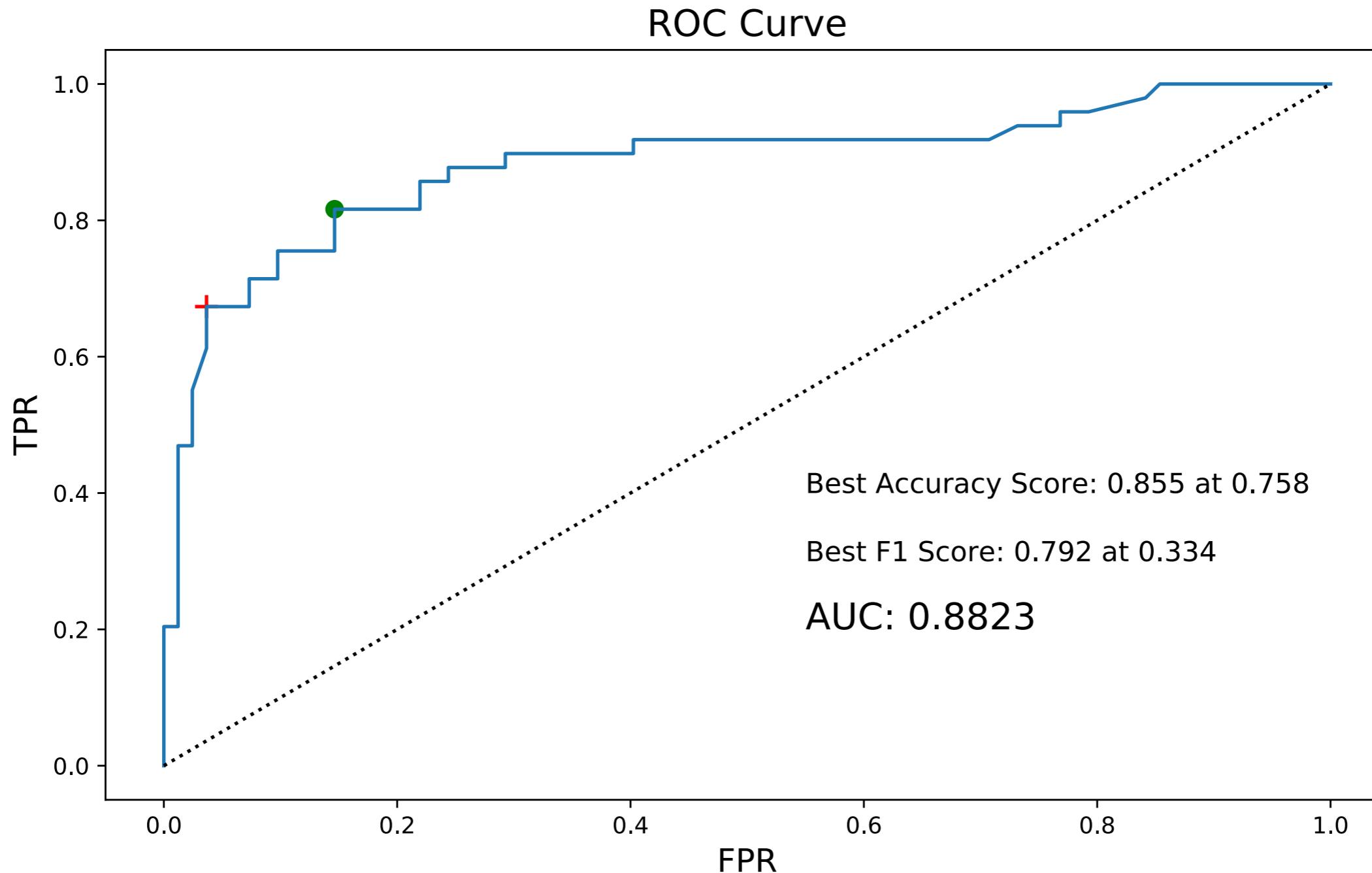
# Results: Validation Data



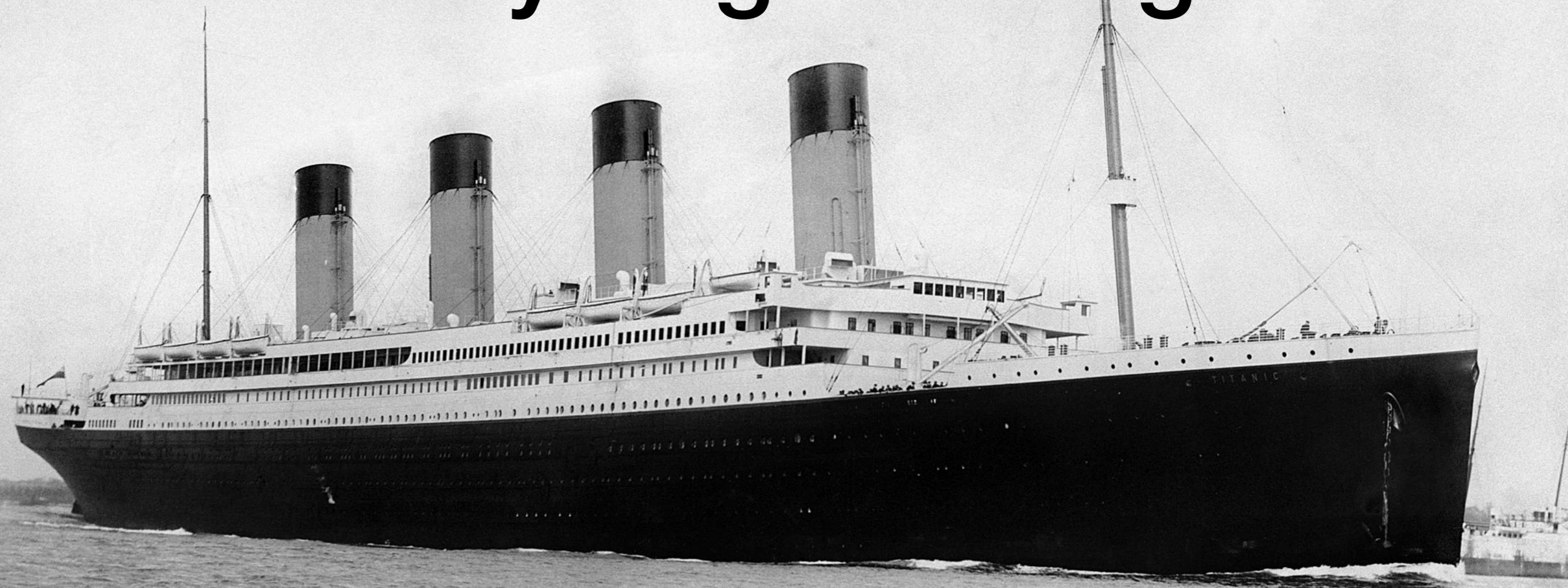
# Binary Logistic Regression Model

- 90% of the total data was used to train the model (combined data)
  - 10,000 iterations
  - Initial  $\alpha = 0.0$
  - Initial  $\beta = 0.3$
- Model then used to Predict the remaining 10% (test data)

# Results: Test Data



# Modeling Titanic Survivability with Binary Logistic Regression



Matthew Melendez