

# Titanic Data Report.

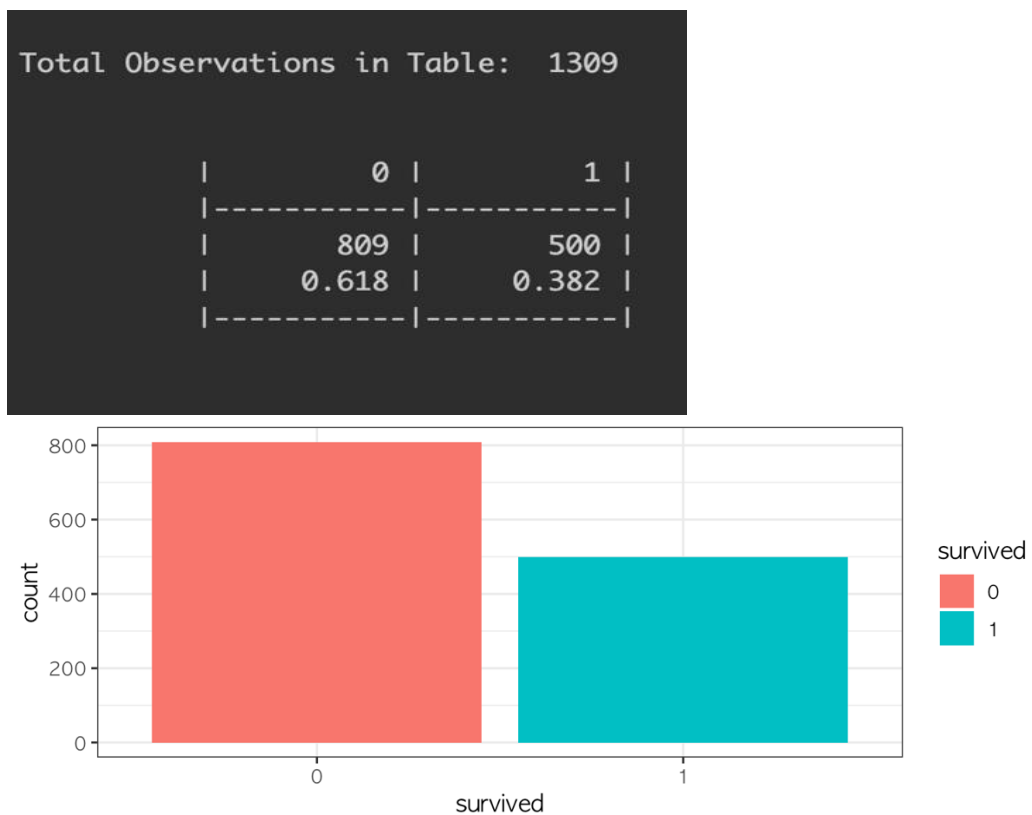
---



제출일	2024.05.19	전공	융합소프트웨어학과
과목	R통계분석	학번	60203042
담당교수	이강선	이름	최원빈

## #Part 1 – 단변량 분석

### 1. 생존자수

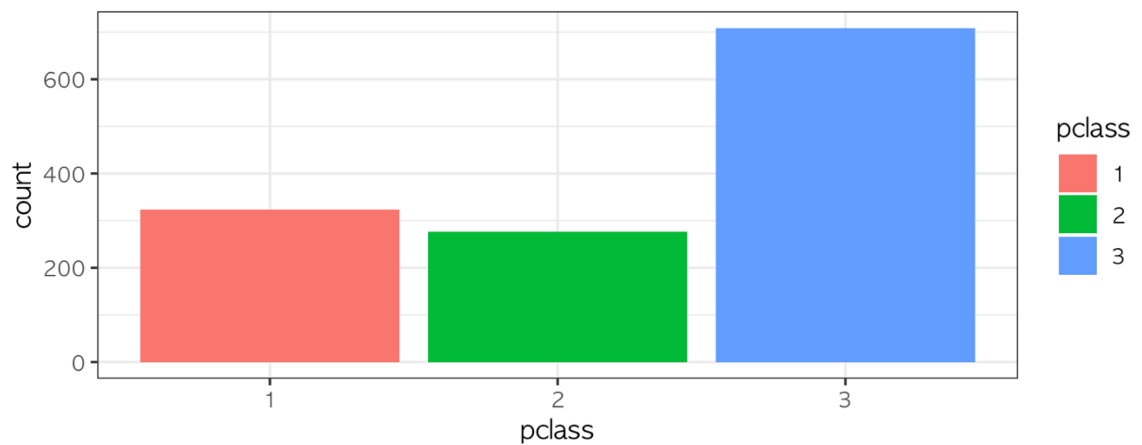


- 데이터에서는 0을 사망자, 1을 생존자로 표기하고 있다.
- 교차테이블에서는 이들의 간략한 수를 나타냈다.
- 사망자가 생존자보다 309명 많으며 백분율로는 61퍼센트를 차지했다.
- 즉 전체적으로는 타이타닉 사고의 사망률이 생존률 보다 높다는것을 알 수 있다.

## 2. Pclass별 탑승객 수

Total Observations in Table: 1309

	1	2	3
count	323	277	709
proportion	0.247	0.212	0.542

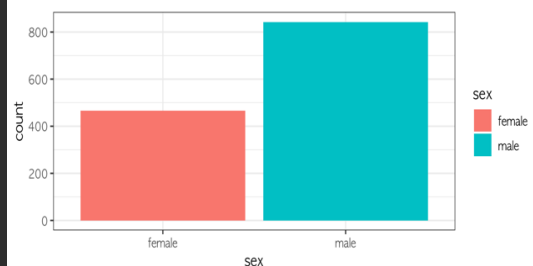


→ 이번엔 등석별 탑승객의 분포를 알아보면 3등석이 54퍼센트로 가장 많은 비율으로 차지한다. 그다음으로 2등석, 1등석으로 순서가 나열된 모습을 볼 수 있다.

## 3. 성별 탑승객 수

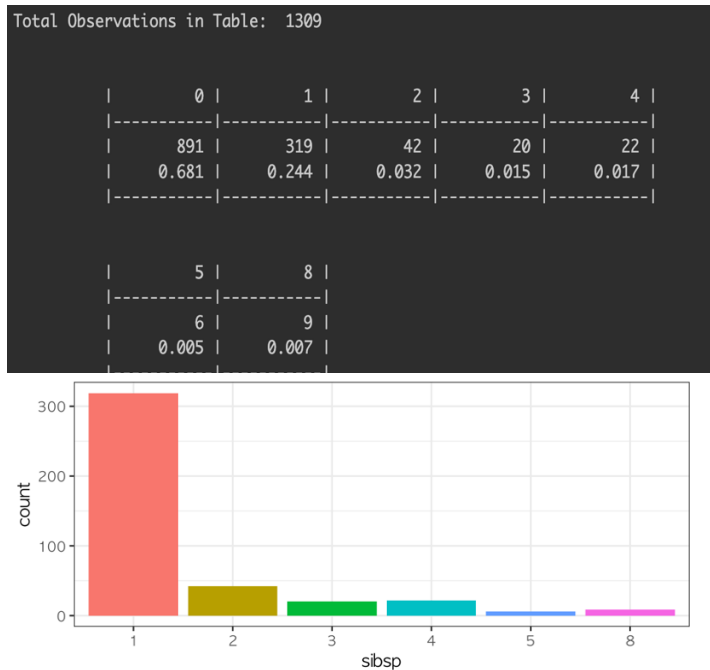
Total Observations in Table: 1309

	female	male
count	466	843
proportion	0.356	0.644



- ➔ 그 다음으로 성별 탑승객 수는 남자가 64퍼센트를 차지했고, 여자가 35퍼센트를 차지했다.
- ➔ 당시의 시대 상 활동이 많은 남자 타이타닉 또한 많이 탑승했다는 것을 알 수 있다.

#### 4. 동반탑승객수 (형제 / 배우자)



- ➔ 동반탑승객수가 0인 혼자 타는 사람이 68퍼센트를 차지했다.
- ➔ 그 다음으로, 1명, 총 둘이 탑승한 승객이 많았다.
- ➔ 혼자 타거나 한명만 동행하는 소규모의 동행이 상당했다는 것을 알 수 있다.

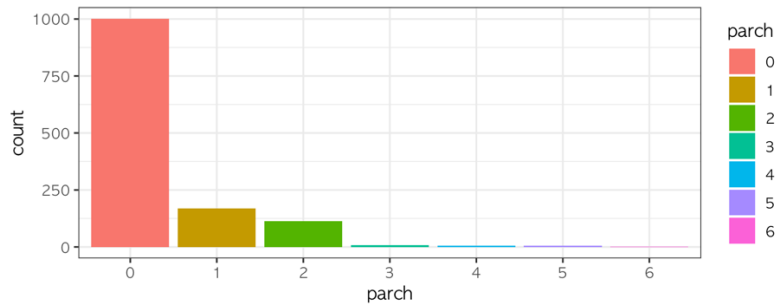
#### 5. 동반탑승객수 ( 부모/자식)

Total Observations in Table: 1309

	0	1	2	3	4
Count	1002	170	113	8	6
Proportion	0.765	0.130	0.086	0.006	0.005

	5	6	9
Count	6	2	2
Proportion	0.005	0.002	0.002

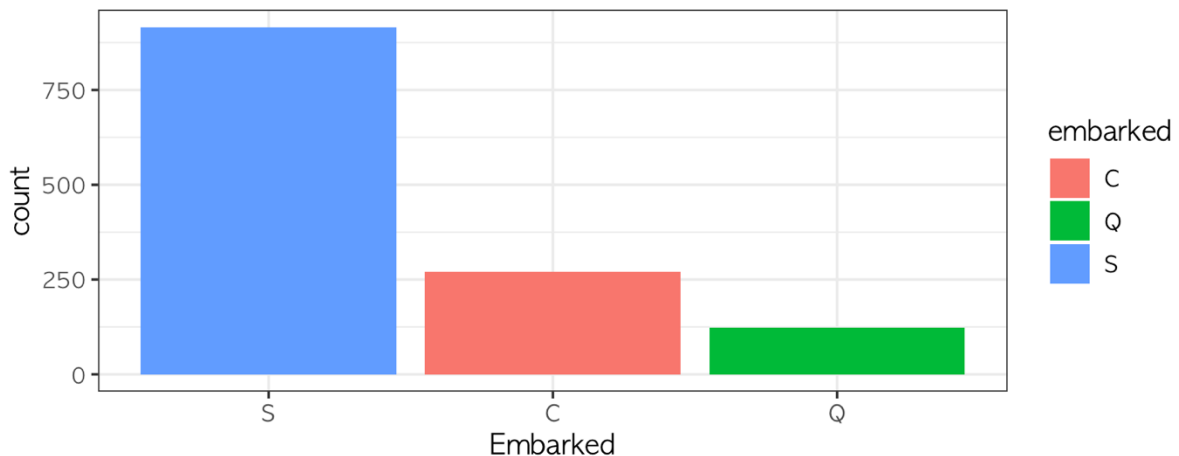


→ 4번의 결과를 보면 알 수 있듯 동반자, 즉, 부모/자식을 동반한 승객에서도 혼자타는 사람이 가장 많았다. 대체적으로 동반자가 적은 승객의 비율이 동반자가 많은 승객의 비율보다 많았다는 것을 알 수 있다.

## 6. 중간경유지 탑승객수

Total Observations in Table: 1309

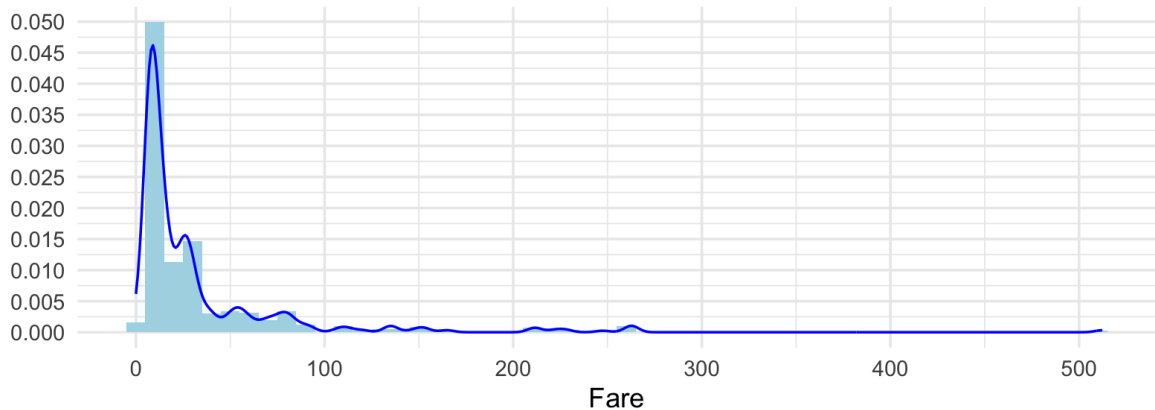
C	Q	S
270	123	916
0.206	0.094	0.700



→ 중간 경유지를 나타낸 지표이다. S -> C -> Q 순으로 해당 지역을 경유해서 간 승객들의 비율이 많다는 것을 알 수 있다.

## 7. 요금

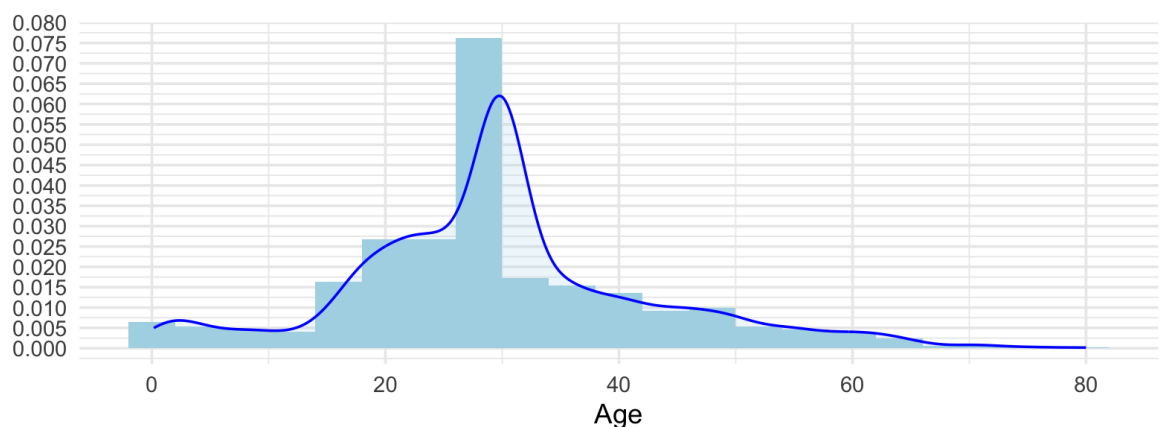
<연속변수이므로, 교차테이블 삽입 x>



- ➔ 요금의 경우, 0 - 100 사이 즉, 비교적 적은 요금을 내고 탄 승객수가 압도적으로 많았다.
- ➔ 흥미로운 건, 우리가 알던 타이타닉호는 초호화 크루즈를 상상하게 되는데 실제로는 그것보다는 비교적 적은 요금을 내고 탔던 서민들이 주를 이뤘다는 추측을 할 수가 있다.

## 8. 나이

<연속변수이므로, 교차테이블 삽입 x>



- ➔ 나이의 경우, 정규분포와 비슷한 형태의 그래프가 나온다.
- ➔ 청년층인 20,30,40대의 비율이 가장 많았고 이 나이층을 기준으로 장년층, 유아층은 비슷하게 적은 비율을 차지한다.

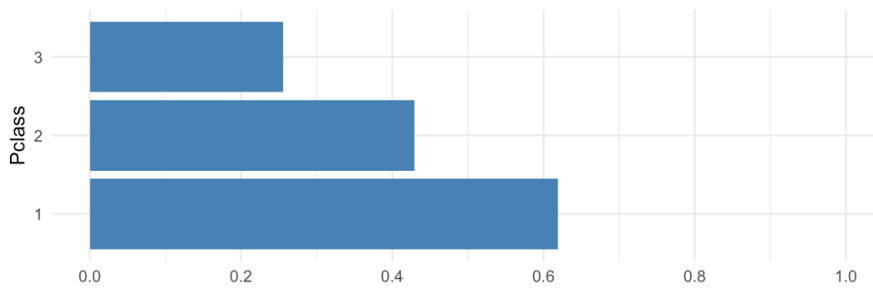
→ 특히 20대 후반, 30대 초반의 나이를 가진 탑승객이 압도적으로 많고, 20대 초중반, 30대 중후반은 그리 많지 않다는 것은 주목할 만하다.

## #Part 2 – 다변량 분석

### 1. 등석별 생존자 비율

Total Observations in Table: 1309

titanic\$class	titanic\$survived		Row Total
	0	1	
1	123	200	323
	29.411	47.587	
	0.381	0.619	0.247
	0.152	0.400	
	0.094	0.153	
2	158	119	277
	1.017	1.645	
	0.570	0.430	0.212
	0.195	0.238	
	0.121	0.091	
3	528	181	709
	18.411	29.788	
	0.745	0.255	0.542
	0.653	0.362	
	0.403	0.138	
Column Total	809	500	1309
	0.618	0.382	



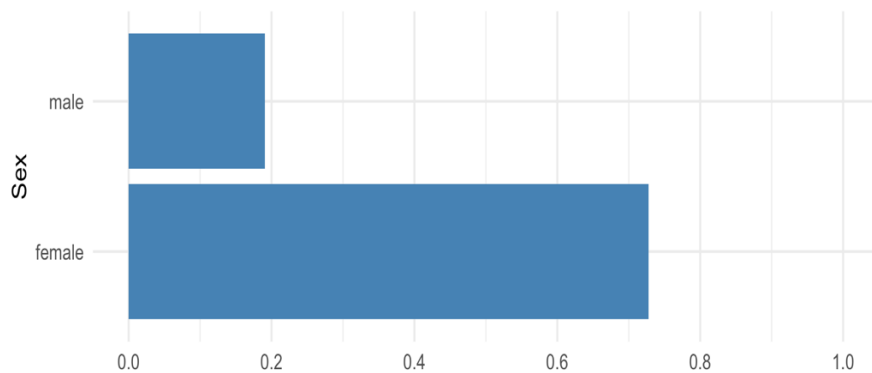
- 그래프에서는 등석별로 생존자의 비율을 나타냈다. 1등석에 탄 승객들의 생존률이 가장 높았고 그 다음으로 2등석, 마지막으로 3등석이다.
- 안전장치나 구조 우선순위, 배의 승무원들이 1등석의 있는 사람들을 우선적으로 챙겼을거라는 추측을 할 수 있다.
- Part 1 : 2번을 참고하면 3등석을 탄 탑승객들이 제일 많고 1등석을 탄 탑승객들이 제일 적는데 1등석의 승객의 생존률이 제일 높고 3등석의 승객의 생존률이 제일 낮다.
- 결론적으로 등석 지표는 생존과 죽음에 있어서 꽤나 유의미한 지표라고 볼 수 있다.

## 2. 성별 생존자 비율

Total Observations in Table: 1309

		titanic\$survived		
titanic\$sex		0	1	Row Total
female		127	339	466
		90.005	145.628	
		0.273	0.727	0.356
		0.157	0.678	
		0.097	0.259	
male		682	161	843
		49.753	80.501	
		0.809	0.191	0.644
		0.843	0.322	
		0.521	0.123	
Column Total		809	500	1309
		0.618	0.382	



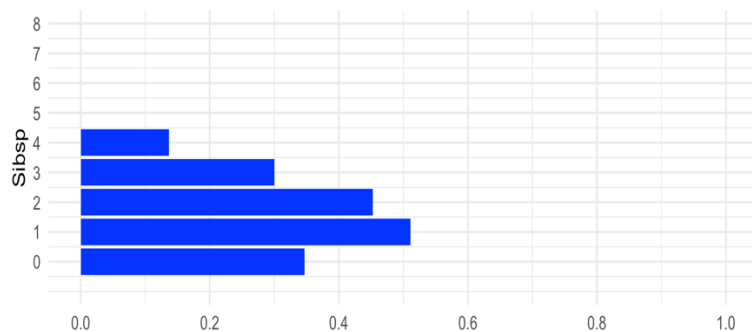


- ➔ 남성의 탑승 비율이 여성보다 높다. 하지만 여성의 생존률이 남성보다 압도적으로 높다. ( part 1. 3번참고)
- ➔ 사고 당시 남성보다는 신체적 약자인 여성들의 구출을 우선시 했을거라는 추측을 할 수 있다.

### 3. 동반탑승객수 별 생존자 (형제/배우자)

Total Observations in Table: 1309

titanic\$sibsp	titanic\$survived		Row Total
	0	1	
0	582	309	891
	1.783	2.885	
	0.653	0.347	0.681
	0.719	0.618	
1	0.445	0.236	
	156	163	319
	8.589	13.898	
	0.489	0.511	0.244
2	0.193	0.326	
	0.119	0.125	
	23	19	42
	0.337	0.545	
	0.548	0.452	0.032
	0.028	0.038	
	0.018	0.015	



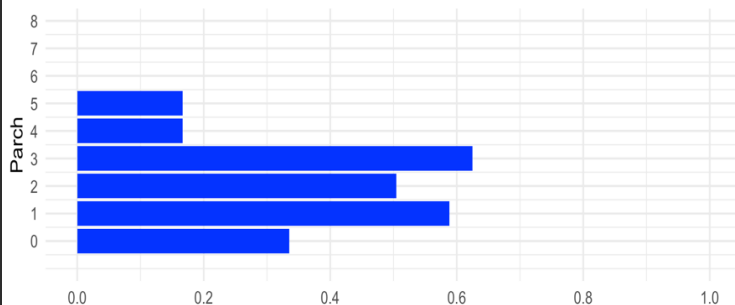
3	14	6	20
	0.217	0.352	
	0.700	0.300	0.015
	0.017	0.012	
	0.011	0.005	
4	19	3	22
	2.147	3.474	
	0.864	0.136	0.017
	0.023	0.006	
	0.015	0.002	
5	6	0	6
	1.416	2.292	
	1.000	0.000	0.005
	0.007	0.000	
	0.005	0.000	
8	9	0	9
	2.125	3.438	
	1.000	0.000	0.007
	0.011	0.000	
	0.007	0.000	
Column Total	809	500	1309
	0.618	0.382	

- ➔ 형제/배우자를 동반한 탑승객의 경우 1명을 동반하거나 2명을 동반한 탑승객의 생존률이 비교적 높았다.
- ➔ 5명 이상의 탑승객을 동반한 탑승객의 생존률은 없다는 것을 볼 수 있다.
- ➔ 하지만 여기서 관과하지 말아야할 점은 동반자수가 5명 이상으로 큰 승객은 거의 없다. ( part 1. 4번 , 5번참고)
- ➔ 5명 이상의 탑승객을 동반한 탑승객의 생존률이 zero지만, 생존률과 동반자 수가 직결된다고 볼 수는 없다.

## 4. 동반탑승객수 별 생존자 (부모/자식)

Total Observations in Table: 1309

titanic\$parch	titanic\$survived		
	0	1	Row Total
0	666	336	1002
	3.527	5.707	
	0.665	0.335	0.765
	0.823	0.672	
	0.509	0.257	
1	70	100	170
	11.703	18.935	
	0.412	0.588	0.130
	0.087	0.200	
	0.053	0.076	
2	56	57	113
	2.742	4.436	
	0.496	0.504	0.086
	0.069	0.114	
	0.043	0.044	



3	3	5	8
	0.765	1.237	
	0.375	0.625	0.006
	0.004	0.010	
	0.002	0.004	
4	5	1	6
	0.450	0.728	
	0.833	0.167	0.005
	0.006	0.002	
	0.004	0.001	
5	5	1	6
	0.450	0.728	
	0.833	0.167	0.005
	0.006	0.002	
	0.004	0.001	
6	2	0	2
	0.472	0.764	
	1.000	0.000	0.002
	0.002	0.000	
	0.002	0.000	
9	2	0	2
	0.472	0.764	
	1.000	0.000	0.002
	0.002	0.000	
	0.002	0.000	
Column Total	809	500	1309
	0.618	0.382	

→ 3번 그래프에서와는 약간 다르게, 부모/자식을 동반한 경우 1명 혹은 3명을 동반한 탑승객의 생존률이 가장 높게 나왔다.

→ 공통점으로는 6명이상 즉, 많은 가족과 동행한 탑승객의 생존률은 없다는 점 또한 주목할만하다.

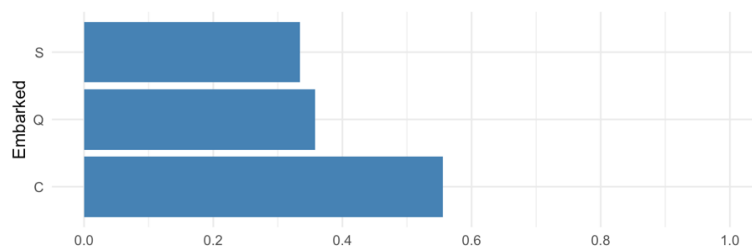
→ 6명이상의 동행자를 지닌 탑승객은 전체 탑승객의 비율에서는 적다. ( part 1. 4번 , 5번참고)

→ 생존률과 동반자 수가 직결된다고 볼 수는 없다. (part2 4,5번 참고)

## 5. 중간경유지 탑승별 생존자 비율

Total Observations in Table: 1309

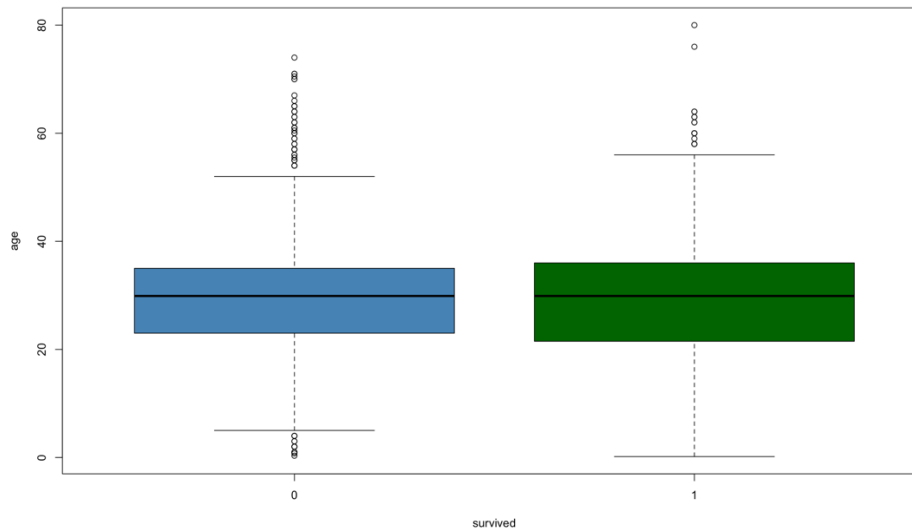
titanic\$embarked	titanic\$survived		Row Total
	0	1	
C	120	150	270
	13.164	21.299	
	0.444	0.556	0.206
	0.148	0.300	
	0.092	0.115	
Q	79	44	123
	0.117	0.189	
	0.642	0.358	0.094
	0.098	0.088	
	0.060	0.034	
S	610	306	916
	3.402	5.504	
	0.666	0.334	0.700
	0.754	0.612	
	0.466	0.234	
Column Total	809	500	1309
	0.618	0.382	



- ➔ 중간경유지에서는 C를 경유한 승객들의 생존률이 가장 높았다. 그 뒤로 Q, S의 생존률의 순서이다.
- ➔ Part 1의 6번을 참고하면, S를 경유한 승객의 비율이 압도적으로 많은데, S를 경유한 승객의 생존률은 가장 낮다.
- ➔ S라는 경유지에 남성이 많았거나, 3등석의 비율이 높았을 수도 있다는 추측이 가능하다.
- ➔ 경유지라는 지표는 생존률에 있어서 어느정도는 유의미하다고 판단된다.

## 6. 나이별 생존자 분포

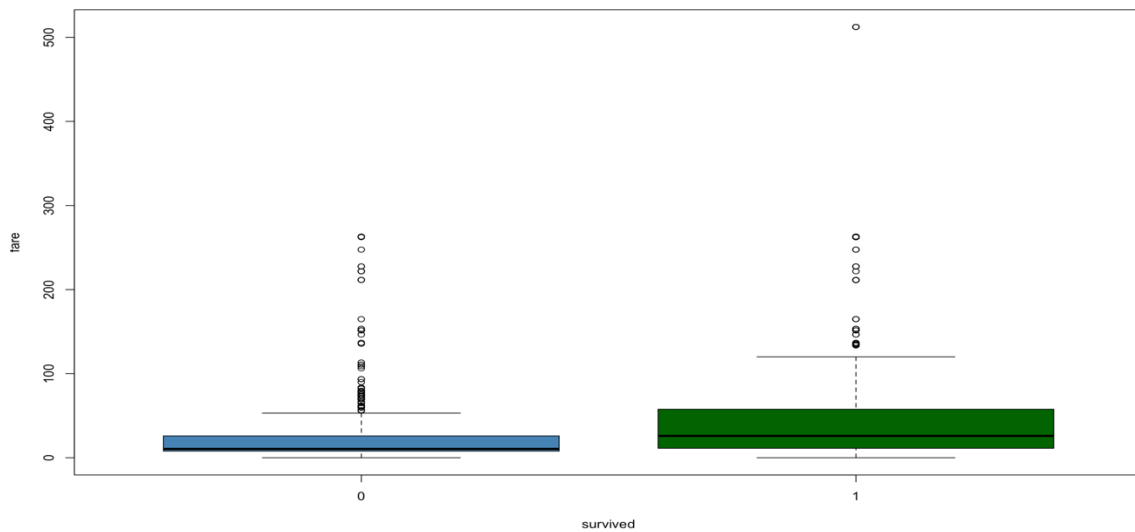
<연속변수이므로, 교차테이블 삽입 x>



- 나이별 생존자는 전체적으로 나이별 전체 탑승객 수와 비슷한 형태를 띄운다
- 다만, 아래 극단치 중 사망자를 표기한 0의 지표에 나이가 어린 즉 10세미만의 어린이들이 많다는 것을 알 수 있다.
- 따라서, 나이라는 기준은 생사를 갈랐던 척도라고 보기엔 힘들다.

## 7.요금별 생존자 분포

<연속변수이므로, 교차테이블 삽입 x>



- ➔ 요금별 생존자 또한 마찬가지로 전체적으로 요금별 전체 탑승객 수와 비슷한 형태를 띠운다.
- ➔ 여러 극단치는 그 구간의 요금을 낸 탑승객이 적음에 따른 것이다.
- ➔ 결론적으로는 요금이라는 기준은 생사를 갈랐던 척도라 보기엔 힘들다.

## 결론.

등석별 : 상당히 유의미한 지표임을 알 수 있었다.

동행자 별 : 그렇게 유의미한 지표는 얻을 수는 없었지만 동행자가 없거나 소규모일때는 생존률이 높다는 것은 주목할만하다.

경유지 별 : S를 경유한 승객들의 생존률이 가장 높은 것은 주목할만하다.

요금과 나이별 : 탑승의 비율과 생존의 비율이 비례하는 점은 주목할만하지만 생사를 갈랐던 척도라고 보기에는 힘들다.