

Ans 3 > We are using here Titanic dataset to analyze:

load data:

⇒ `titanic ← read.csv("C:/user/Desktop/titanic.csv", header = TRUE,
Sep = ",")`

Peek your data

⇒ `View(titanic)`

This help us to familiarising with the data set.

⇒ `head(titanic, n) | tail(titanic, n)`

In order to have a quick look at the data, we often use the `head()`/`tail()`.

→ this show top 10 ~~an~~ row of the data set.
and bottom 5 row of data set.

⇒ `names(titanic)`

This help us in checking out all the variables in the data set.

⇒ `str(titanic)`

This help in understanding the structure of the data set.

→ Summary (Titanic)

`summary()` is one of the most important function that help in summarising each attribute in the data set

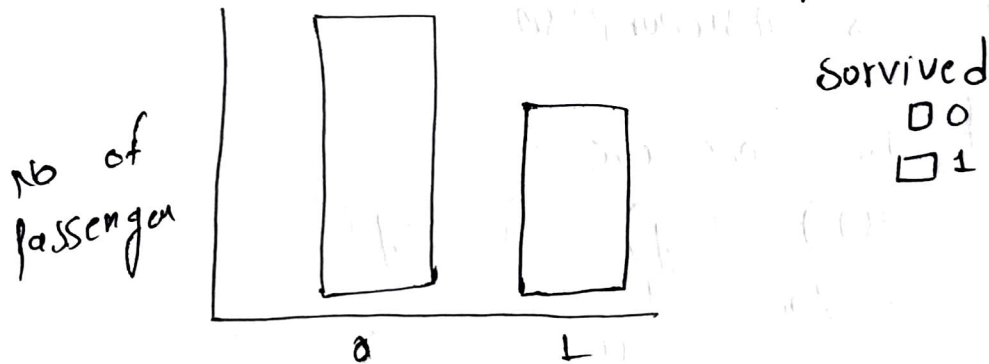
→ as.factor (data Set \$ ColumnName)

In order to such variables treated as factors and not as no. we need explicitly convert them to factor using the function

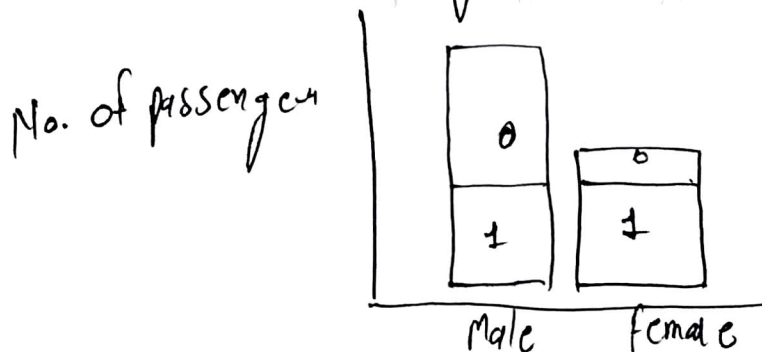
→ Analysis & Visualization

• Survival rate:

```
ggplot (titanic , aes( x= Survived )) + geom - bar()
```



• Survival rate based gender



4 Titanic.aes (x=sex, fill=survived)
theme_bw() + geom_bar() +
labs(y="Number of passengers",
title="Survived rate by gender")

Name (Phone)		Address	City	State	Zip	Phone	Mail	Notes
[1]	James H. Smith	123 Main St	Springfield	MA	01101	555-1234		
[2]	John Doe	456 Oak Ave	Portland	ME	04101	555-5678		

[illegible]

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	ParCh	Ticket	Fare	Cabin	Embarked
886	886	0	3	Aice, Mrs. William (Margaret Norton)	female	39	0	5	382852	29.125		Q
887	887	0	2	Montvila, Rev. Juozas	male	27	0	0	211536	18.000		S
888	888	1	1	Graham, Miss. Margaret Edith	female	19	0	0	112053	30.000	442	S
889	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NA	1	2	W./C. 6607	23.450		S
890	890	1	1	Behr, Mr. Carl Howell	male	26	0	0	111309	30.000	C148	C
891	891	0	3	Dooley, Mr. Patrick	male	32	0	0	370376	7.750		Q

PassengerId			Survived	Pclass	Name		Sex	Age	SibSp			
Min.	: 1.0	Min.	:0.0000	Min.	:1.000	Abbing, Mr. Anthony	: 1	Female: 314	Min.	: 0.42	Min.	:0.000
1st Qu.	:221.5	1st Qu.	:0.0000	1st Qu.	:2.000	Abbott, Mr. Rossmore Edward	: 1	male : 577	1st Qu.	:20.12	1st Qu.	:0.000
Median	:446.0	Median	:0.0000	Median	:3.000	Abbott, Mrs. Stanton (Rosa Hunt)	: 1		Median	:26.00	Median	:0.000
Mean	:446.0	Mean	:0.1816	Mean	:2.109	Abelson, Mr. Samuel	: 1		Mean	:29.70	Mean	:0.521
3rd Qu.	:668.5	3rd Qu.	:1.0000	3rd Qu.	:3.000	Abelson, Mrs. Samuel (Hannah Wozosky)	: 1		3rd Qu.	:38.00	3rd Qu.	:1.000
Max.	:891.0	Max.	:1.0000	Max.	:3.000	Adahl, Mr. Mauritz Nils Martin	: 1		Max.	:80.00	Max.	:8.000
					(other)	:885		NA's : 177				
Parch			Ticket		Fare	Cabin		Embarked				
Min.	:0.0000	1601	: 7	Min.	: 0.00			: 3				
1st Qu.	:0.0000	347082	: 7	1st Qu.	: 7.91	B96 B98	: 4	C1168				
Median	:0.0000	Ca. 23431	: 7	Median	: 14.45	C23 C25 C27	: 4	Q1 77				
Mean	:0.1816	3101295	: 6	Mean	: 32.20	C6	: 4	S1044				
3rd Qu.	:0.0000	347088	: 6	3rd Qu.	: 31.00	C12 C16	: 3					
Max.	:6.0000	Ca. 2144	: 6	Max.	:512.33	D	: 3					
		(other)	:852			(other)	:186					

PassengerId			survived	Pclass	Name		Sex	Age	SibSp			
Min.	: 1.0	Min.	:0.0000	Min.	:1.000	abbing, Mr. Anthony	: 1	Female: 314	Min.	: 0.42	Min.	:0.000
1st Qu.	:221.5	1st Qu.	:0.0000	1st Qu.	:2.000	abbott, Mr. Rossmore Edward	: 1	male : 577	1st Qu.	:20.12	1st Qu.	:0.000
Median	:446.0	Median	:0.0000	Median	:3.000	abbott, Mrs. Stanton (Rosa Hunt)	: 1		Median	:26.00	Median	:0.000
Mean	:446.0	Mean	:0.1818	Mean	:2.109	abelson, Mr. Samuel	: 1		Mean	:29.70	Mean	:0.521
3rd Qu.	:668.5	3rd Qu.	:1.0000	3rd Qu.	:3.000	abelson, Mrs. Samuel (Hannah Wlosky)	: 1		3rd Qu.	:38.00	3rd Qu.	:1.000
Max.	:891.0	Max.	:1.0000	Max.	:3.000	Adah!, Mr. Mauritz Nils Martin	: 1		Max.	:80.00	Max.	:8.000
						(other)	:885		NA's	:177		
Parch			Ticket		Fare		Cabin		Embarked			
Min.	:0.0000	1601	: 7	Min.	: 0.00		:687					: 2
1st Qu.	:0.0000	347082	: 7	1st Qu.	: 7.91	896 898	: 4	C1108				
Median	:0.0000	Ca. 23431	: 7	Median	: 14.45	C23 C25 C27	: 4	Q1 77				
Mean	:0.1818	3101295	: 6	Mean	: 32.20	C6	: 4	S1044				
3rd Qu.	:0.0000	347088	: 6	3rd Qu.	: 31.00	C22 C26	: 3					
Max.	:6.0000	Ca. 2144	: 6	Max.	:512.33	D	: 3					
		(other)	:852			(other)	:186					


```
titanic$survived <- as.factor(titanic$survived)
titanic$Pclass <- as.factor(titanic$Pclass)
titanic$sex <- as.factor(titanic$sex)
titanic$Embarked <- as.factor(titanic$Embarked)
```

```
> str(titanic)
'data.frame':   891 obs. of  12 variables:
 $ PassengerId: int   1 2 3 4 5 6 7 8 9 10 ...
 $ Survived   : Factor w/ 2 levels "0","1": 1 2 2 2 1 1 1 1 2 2 ...
 $ Pclass     : Factor w/ 3 levels "1","2","3": 3 1 3 1 3 3 1 3 3 2 ...
 $ Name       : Factor w/ 891 levels "Abbing, Mr. Anthony"...: 109 191 358 277 16 559 520 629 417 581 ...
 $ Sex        : Factor w/ 2 levels "female","male": 2 1 1 1 2 2 2 2 1 1 ...
 $ Age        : num   22 38 26 35 35 NA 54 2 27 14 ...
 $ SibSp      : int   1 1 0 1 0 0 0 3 0 1 ...
 $ Parch      : int   0 0 0 0 0 0 0 1 2 0 ...
 $ Ticket     : Factor w/ 681 levels "110152","110413"...: 524 597 670 50 473 276 86 396 345 133 ...
 $ Fare       : num   7.25 71.28 7.92 53.1 8.05 ...
 $ Cabin      : Factor w/ 148 levels "", "A10","A14"...: 1 83 1 57 1 1 131 1 1 1 ...
 $ Embarked   : Factor w/ 4 levels "", "C","Q","S": 4 2 4 4 4 3 4 4 4 2 ...
```

PassengerId	Survived	Sex	Name	Age	SibSp	ParCh	Ticket	Fare	Cabin	Embarked
1	0	3	Braund, Mr. Owen Harris	22.00	1	0	A/5 21171	7.2500		S
2	1	1	Cummings, Mrs. John Bradley (Florence Briggs Thayer)	38.00	1	0	PC 17500	71.2833	C85	C
3	1	3	Henderson, Miss. Laina	26.00	0	0	STON/O2. 3101282	7.9250		S
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	35.00	1	0	113803	53.1000	C123	S
5	0	3	Allen, Mr. William Henry	35.00	0	0	373450	8.0500		S
6	0	3	Moran, Mr. James	NA	0	0	530877	8.4500		Q
7	0	1	McCarthy, Mr. Timothy J	54.00	0	0	17463	51.8625	E46	S
8	0	3	Palsson, Master. Gosta Leonard	2.00	3	1	349909	21.0750		S
9	1	3	Johnson, Mrs. Oscar W (Elizabeth Vilhelmina Berg)	27.00	0	2	347742	11.1333		S
10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	54.00	1	0	237736	30.0700		C