Spaceship Titanic Project

The objective of this project is to conduct an exploratory data analysis of the Spaceship Titanic data set using statistics summary and visualization approaches to explore the data and to identify interesting patterns and significant information within the data.

Check data structure

```
'data.frame':   8693 obs. of  14 variables:
 $ PassengerId : chr  "0001_01" "0002_01" "0003_01" "0003_02" ...
 $ HomePlanet  : chr  "Europa" "Earth" "Europa" "Europa" ...
 $ CryoSleep   : chr  "False" "False" "False" "False" ...
 $ Cabin       : chr  "B/0/P" "F/0/S" "A/0/S" "A/0/S" ...
 $ Destination : chr  "TRAPPIST-1e" "TRAPPIST-1e" "TRAPPIST-1e" "TRAPPIST-1e" ...
 $ Age         : num  39 24 58 33 16 44 26 28 35 14 ...
 $ VIP         : chr  "False" "False" "True" "False" ...
 $ RoomService : num  0 109 43 0 303 0 42 0 0 0 ...
 $ FoodCourt   : num  0 9 3576 1283 70 ...
 $ ShoppingMall: num  0 25 0 371 151 0 3 0 17 0 ...
 $ Spa         : num  0 549 6715 3329 565 ...
 $ VRDeck      : num  0 44 49 193 2 0 0 NA 0 0 ...
 $ Name        : chr  "Maham Ofracculy" "Juanna Vines" "Altark Susent" "Solam Susent" ...
 $ Transported : chr  "False" "True" "False" "False" ...
```

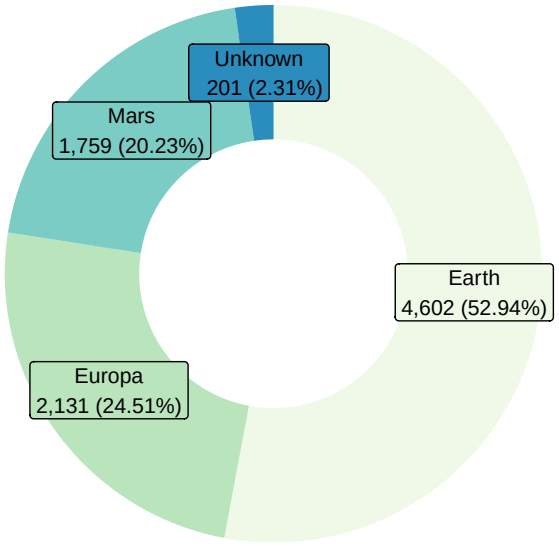
Getting insights from the variables

1. Where were the home planets of those passengers departed from?

The passengers departed from three main home planets Earth, Europa, and Mars. About half of those passengers (53%) were from Earth. There're small number (2%) of passengers whose their home planets information are missing.

Passenger count by HomePlanet:

HomePlanet	Count	Total %
Earth	4602	53
Europa	2131	25
Mars	1759	20
Unknown	201	2
Sum	8693	100



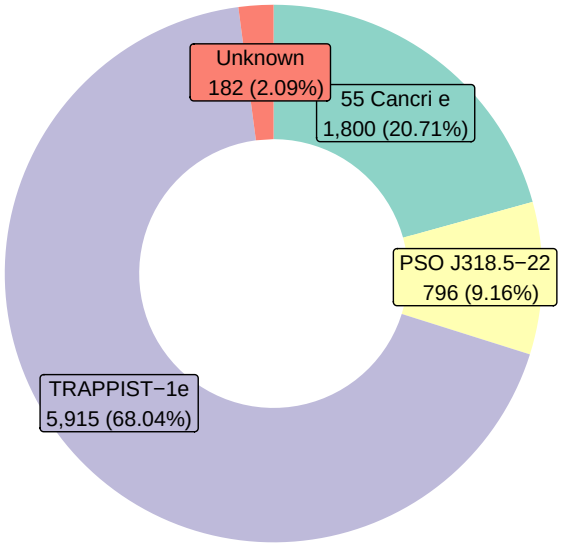
2. Which planet destination most passengers would be debarking to?

There are three destinations the passengers would be debarking to 55 Cancri e, PSO J318.5-22, and TRAPPIST-1e. TRAPPIST-1e appears to be the top destination - 68% of the passengers were to emigrate there.

Passenger count by Destination:

Destination	Count	Total %
55 Cancri e	1800	21

PSO J318.5-22	796	9
TRAPPIST-1e	5915	68
Unknown	182	2
Sum	8693	100



3. Top 3 HomePlanet and Destination pairs

The majority of passengers were emigrating from Earth to TRAPPIST-1e (36% of the passengers). The 2nd and 3rd were Mars to TRAPPIST-1e and Europa to TRAPPIST-1e respectively. Though there were higher number of passengers from Earth who were being transported to TRAPPIST-1e. Those from Mars shows higher likelihood to emigrate to TRAPPIST-1e than those from Earth (84% of passengers from Mars opted to emigrate to TRAPPIST-1e vs. 67% of those from Earth). Passengers from Europa were the only group that shows higher interest in emigrating to 55 Cancri e (42% of Europa Passengers were heading there). PSO J318.5-22 appears to be the least appealing destination among the passengers.

HomePlanet & Destination pairs (count):

	HomePlanet	Destination	n	percent_total
1	Earth	TRAPPIST-1e	3101	36
2	Mars	TRAPPIST-1e	1475	17

3	Europa	TRAPPIST-1e	1189	14
4	Europa	55 Cancri e	886	10
5	Earth	PSO J318.5-22	712	8
6	Earth	55 Cancri e	690	8
7	Mars	55 Cancri e	193	2
8	Unknown	TRAPPIST-1e	150	2
9	Earth	Unknown	99	1
10	Mars	PSO J318.5-22	49	1
11	Mars	Unknown	42	0
12	Europa	Unknown	37	0
13	Unknown	55 Cancri e	31	0
14	Europa	PSO J318.5-22	19	0
15	Unknown	PSO J318.5-22	16	0
16	Unknown	Unknown	4	0

Crosstab of HomePlanet & Destination (row %):

Destination	55 Cancri e	PSO J318.5-22	TRAPPIST-1e	Unknown	Sum
HomePlanet					
Earth	14.99	15.47	67.38	2.15	100.00
Europa	41.58	0.89	55.80	1.74	100.00
Mars	10.97	2.79	83.85	2.39	100.00
Unknown	15.42	7.96	74.63	1.99	100.00

4. Who got transported based on HomePlanet and Destination?

Again, despite the higher number of passengers from Earth on board comparing to the other two home planets, there was a higher percentage among passengers from Europa that got transported to their destination (66% of passengers from Europa were transported vs. 42% of passengers from Earth). Earth was the only home planet that shows a higher likelihood of those passengers who were not transported.

55 Cancri e destination has a higher portion of transported passengers (61% of passengers got transported to 55 Cancri e comparing to 47% of those transported to Earth).

Crosstab of HomePlanet & Transported:

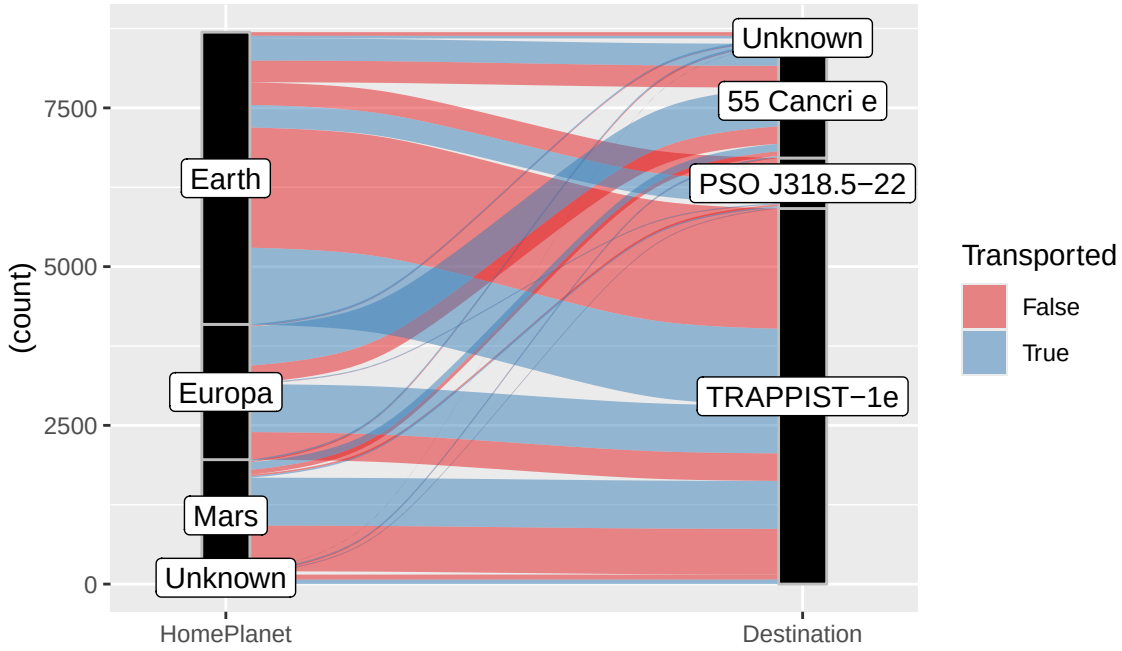
	Count		Row %	
Transported	False	True	False	True
HomePlanet				
Earth	2651	1951	58	42

Europa	727	1404	34	66
Mars	839	920	48	52
Unknown	98	103	49	51

Crosstab of Destination & Transported:

	Count		Row %	
Transported	False	True	False	True
Destination				
55 Cancri e	702	1098	39	61
PSO J318.5-22	395	401	50	50
TRAPPIST-1e	3128	2787	53	47
Unknown	90	92	49	51

Transported passengers by HomePlanet and Destination:



5. What does the age groups look like among the passengers?

Average age of the passengers are 29, youngest passengers are 0 (assuming they were newborns who hadn't reach 1 year old), and the oldest passengers are 79. The majority of the passengers

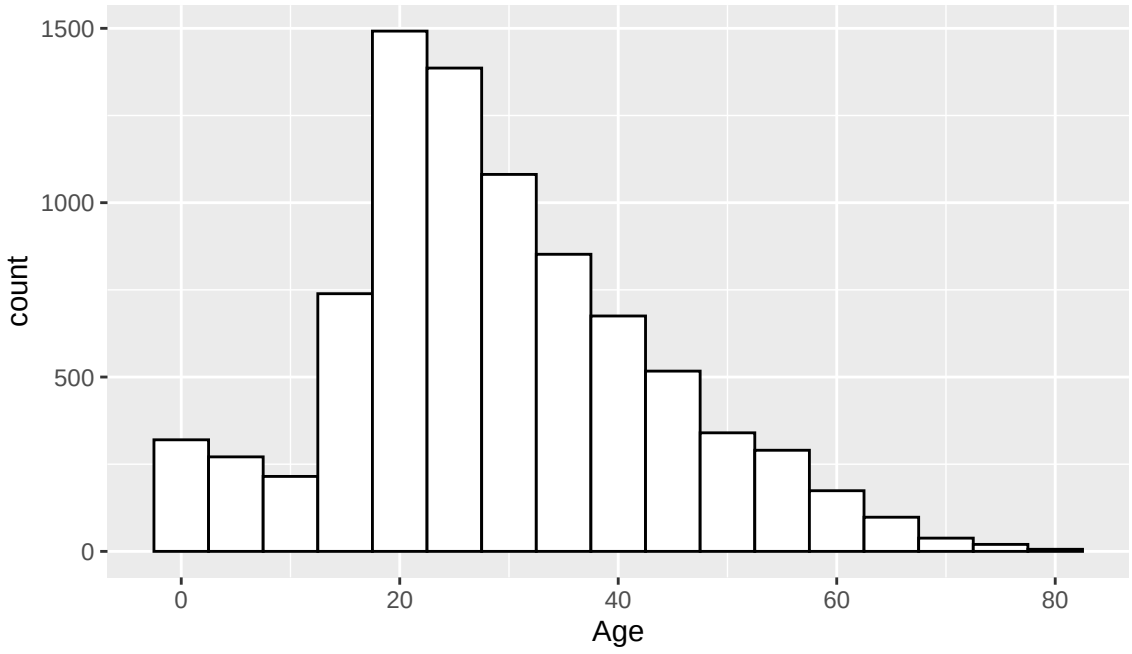
(53%) were adult between 20-39. There were much smaller passengers among those children 0-12 (9%) and senior 60+ (3%).

Among the transported passengers, children 0-12 and teens 13-19 shows the higher likelihood of being transported (70% and 52% respectively) than the older age groups.

Summary statistics (Age):

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   0.00   19.00   27.00   28.83   38.00   79.00     179
```

Age distribution of the passengers:



Age group (count):

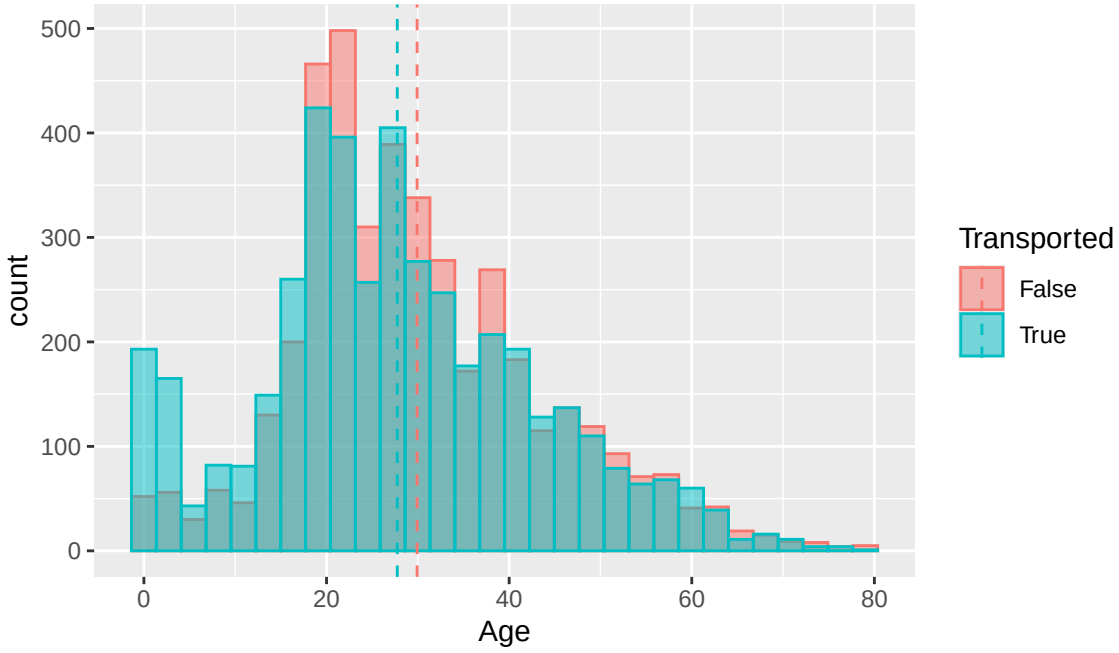
Age_group	Count	Total %
Children 0-12	806	9
Teen 13-19	1352	16
Adult 20-39	4497	53
Adult 40-59	1605	19

Senior 60+	254	3
Sum	8514	100

Crosstab of Age group & Transported:

	Count			Row %		
Transported	False	True	Sum	False	True	Sum
Age_group						
Children 0-12	242	564	806	30	70	100
Teen 13-19	645	707	1352	48	52	100
Adult 20-39	2405	2092	4497	53	47	100
Adult 40-59	799	806	1605	50	50	100
Senior 60+	135	119	254	53	47	100

Age distribution of the passengers by Transported groups:



6. Who used CryoSleep during the voyage?

More passengers (64%) opted to not be put into CryoSleep. Younger passengers (children 0-12 and teens 13-19) appears to have slightly higher percentage to use CryoSleep than older groups. There's significant percentage indicating those who opted to use CryoSleep had a

higher chance (82%) being transported comparing to those who didn't opt CryoSleep with more likelihood 67% for not being transported.

Passenger opting CryoSleep (count):

CryoSleep	Count	Total %
False	5439	64
True	3037	36
Sum	8476	100

CryoSleep vs Avg passenger age:

CryoSleep	Avg_age
False	30
True	27
NA	28

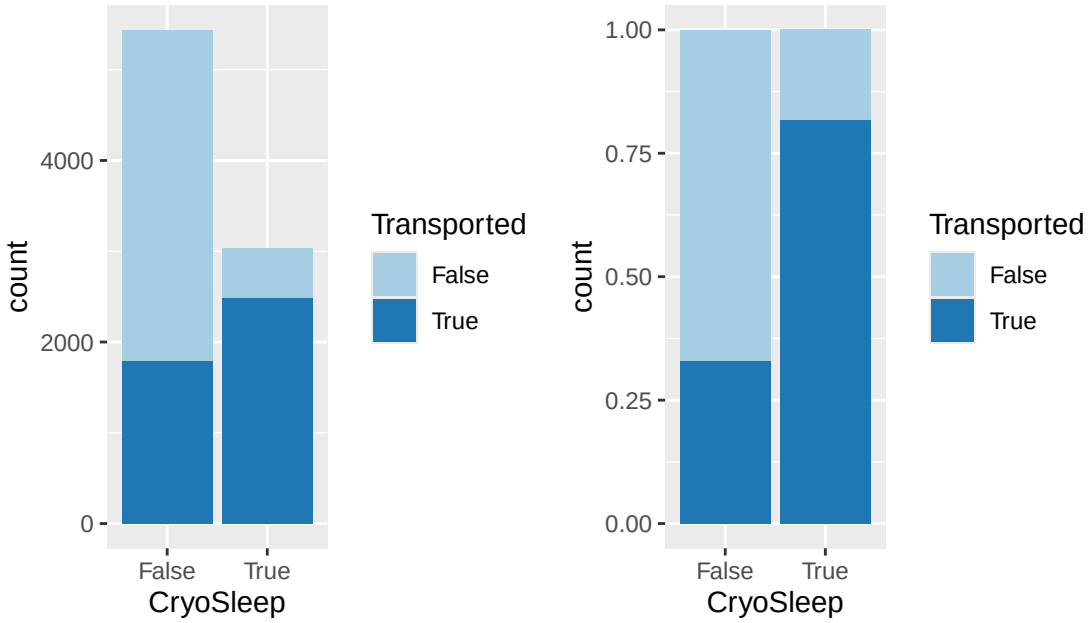
Crosstab of Age group & CryoSleep:

	Count			Row %		
CryoSleep	False	True	Sum	False	True	Sum
Age_group						
Children 0-12	406	377	783	52	48	100
Teen 13-19	758	555	1313	58	42	100
Adult 20-39	2973	1416	4389	68	32	100
Adult 40-59	1042	524	1566	67	33	100
Senior 60+	164	83	247	66	34	100

Crosstab of CryoSleep & Transported:

	Count			Row %		
Transported	False	True	Sum	False	True	Sum
CryoSleep						
False	3650	1789	5439	67	33	100
True	554	2483	3037	18	82	100

Passengers in cryosleep vs Transported:



CryoSleep vs HomePlanet and Destination:

HomePlanet	55 Cancri e	PSO J318.5-22	TRAPPIST-1e	Unknown	Total
Earth	205	355	809	13	1382
Europa	441	9	447	14	911
Mars	76	11	561	21	669
Unknown	17	8	50	0	75
Total	739	383	1867	48	3037

7. Who were those VIP passengers? Did many of them get transported?

There was only a small group (2%) of the passengers who paid for VIP service during the voyage. Among those VIP passengers were adult and senior age groups. Interestingly, there were no passengers from Earth paid for VIP - only those from Europa and Mars paid for the service.

More than half (62%) of those VIP passengers didn't get transported to their destination.

VIP passengers (count):

VIP	Count	Total %
False	8291	98
True	199	2
Sum	8490	100

VIP passengers vs Avg age:

	VIP	Avg_age
	<chr>	<dbl>
1	False	29
2	True	37
3	<NA>	28

VIP passengers vs Age group:

	Count			Row %		
VIP	False	True	Sum	False	True	Sum
Age_group						
Children 0-12	782	0	782	100	0	100
Teen 13-19	1317	2	1319	100	0	100
Adult 20-39	4270	131	4401	97	3	100
Adult 40-59	1515	54	1569	97	3	100
Senior 60+	235	11	246	96	4	100

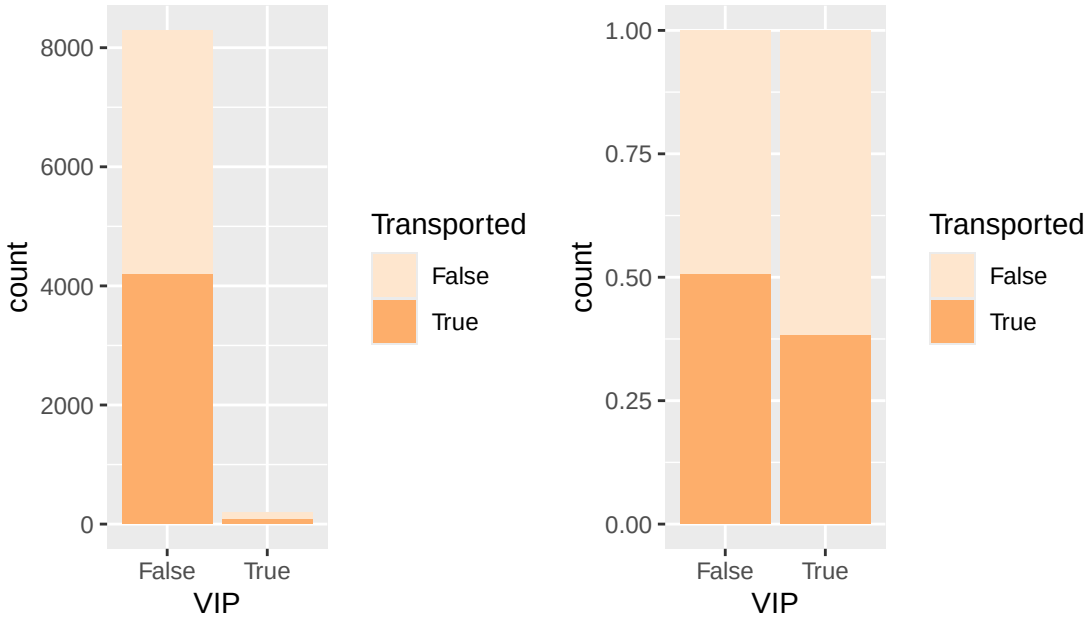
VIP passengers vs Transported:

	Count			Row %		
Transported	False	True	Sum	False	True	Sum
VIP						
False	4093	4198	8291	49	51	100
True	123	76	199	62	38	100

VIP passengers vs HomePlanet & Destination:

HomePlanet	55 Cancri e	PSO J318.5-22	TRAPPIST-1e	Unknown	Total
Europa	96.9% (63)	55.6% (10)	49.1% (56)	100.0% (2)	65.8 (131)
Mars	0.0% (0)	44.4% (8)	48.2% (55)	0.0% (0)	31.7% (63)
Unknown	3.1% (2)	0.0% (0)	2.6% (3)	0.0% (0)	2.5% (5)
Total	100.0% (65)	100.0% (18)	100.0% (114)	100.0% (2)	100.0 (199)

VIP passengers vs Transported:



8. Who were those big spenders on board?

It appears Europa passengers were the biggest spender on the luxury amenities - 6.8 million in total bill with an average of 3.5 thousand per passenger. Earth passengers appears to be the second in spending by total billed at 2.8 million - but that's mainly because they were the largest group on board. They were actually the least spenders with an average of 688 per passenger. Adult 20-39 group was the top spender by total billed (6.8 million) while Senior 60+ was the top spender by average spending per passenger (around 2 thousand per passenger).

Top 5 spenders by total bill:

```
                Name Age HomePlanet Destination   VIP Transported Total_bill
1  Markar Radisiouss  68     Europa 55 Cancri e False       False      35987
2 Pulchib Quidedbolt  41     Europa 55 Cancri e  True        True      31076
3     Scharab Conale  31     Europa 55 Cancri e  True        True      31074
4   Mirfar Optionful  18     Europa TRAPPIST-1e False       False      30478
5     Maiam Oilloody  36     Europa 55 Cancri e False       False      29608
```

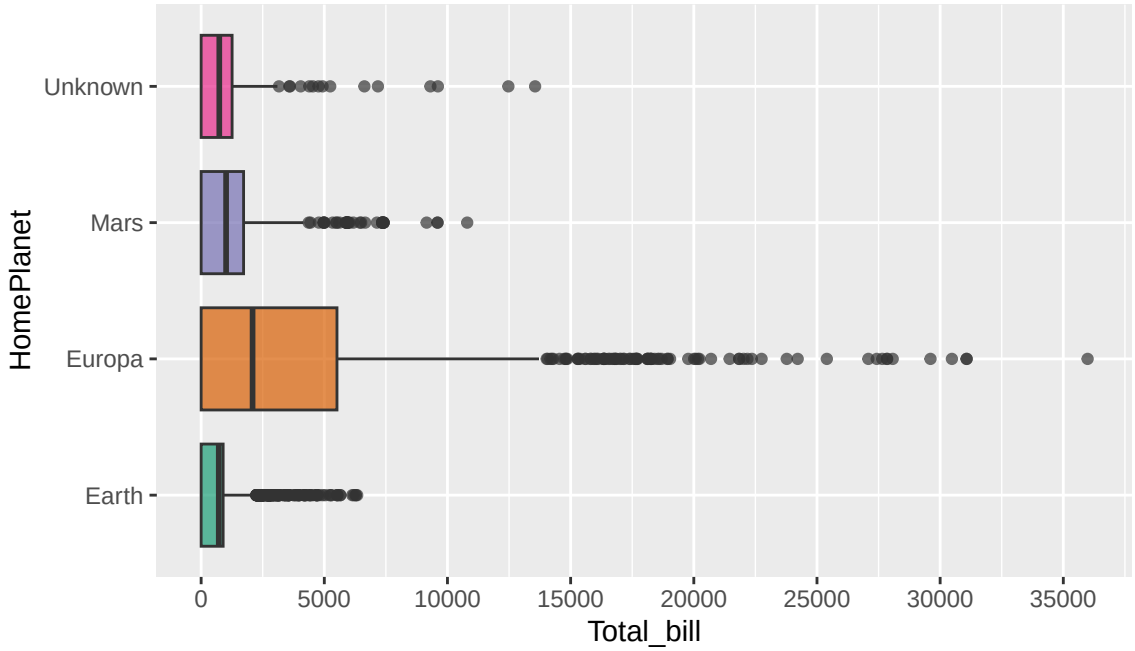
Sum of total bill by HomePlanet:

```
# A tibble: 4 x 2
  HomePlanet Sum_Total_bill
  <chr>               <dbl>
1 Europa            6828385
2 Earth             2823306
3 Mars              1696039
4 Unknown            209893
```

Average total bill by HomePlanet:

```
# A tibble: 4 x 2
  HomePlanet Avg_Total_bill
  <chr>               <dbl>
1 Europa              3553.
2 Unknown             1147.
3 Mars                1074.
4 Earth                688.
```

Boxplot of HomePlanet vs total bill:



Sum of total bill by Destination:

```
# A tibble: 4 x 2
  Destination   Sum_Total_bill
  <chr>                  <dbl>
1 TRAPPIST-1e          7316054
2 55 Cancri e          3606677
3 PSO J318.5-22         416911
4 Unknown               217981
```

Average total bill by Destination:

```
# A tibble: 4 x 2
  Destination   Avg_Total_bill
  <chr>                 <dbl>
1 55 Cancri e           2235.
2 TRAPPIST-1e           1382.
3 Unknown               1337.
4 PSO J318.5-22          582.
```

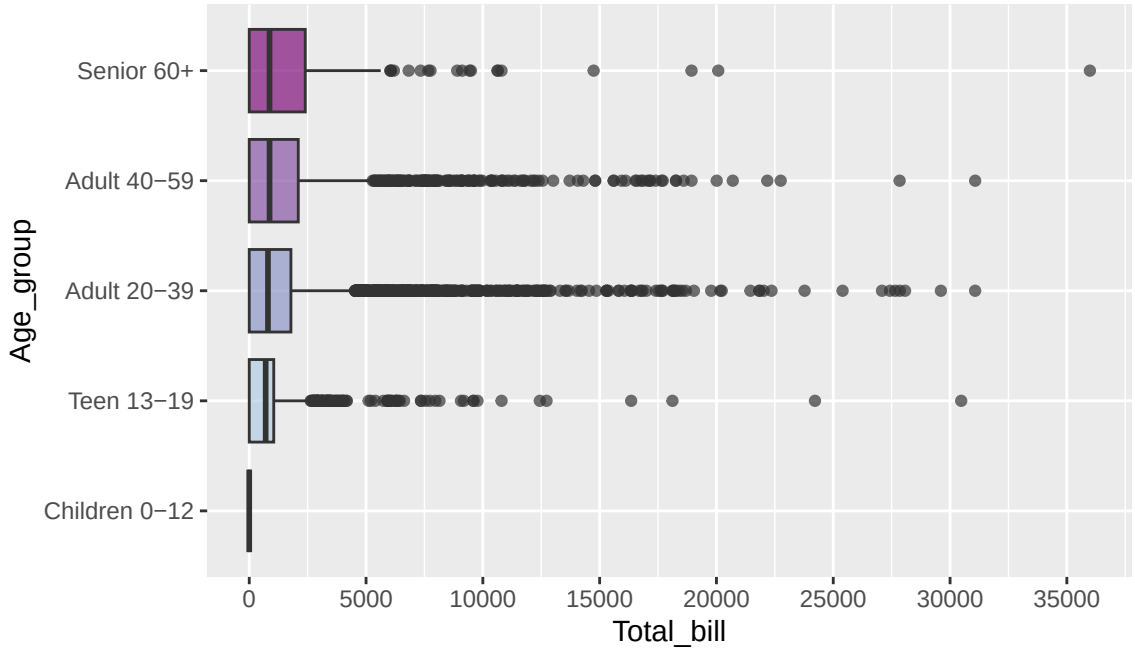
Sum of total bill by Age group:

```
# A tibble: 6 x 2
  Age_group     Sum_Total_bill
  <fct>                  <dbl>
1 Adult 20-39          6896604
2 Adult 40-59          2841561
3 Teen 13-19           1154290
4 Senior 60+            460733
5 <NA>                  204435
6 Children 0-12              0
```

Average total bill by Age group:

```
# A tibble: 6 x 2
  Age_group     Avg_Total_bill
  <fct>                  <dbl>
1 Senior 60+             2012.
2 Adult 40-59            1971.
3 Adult 20-39            1717.
4 <NA>                   1239
5 Teen 13-19              952.
6 Children 0-12             0
```

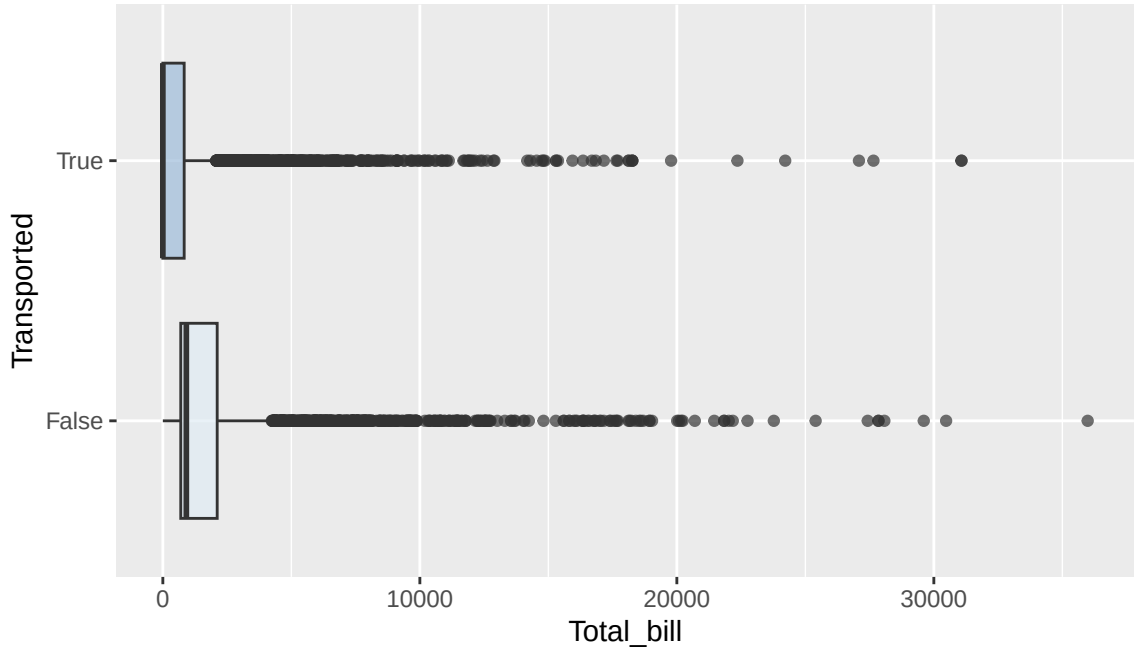
Boxplot of Age group vs total bill:



Sum of total bill by Transported:

```
# A tibble: 2 x 2
  Transported Sum_Total_bill
  <chr>                <dbl>
1 False              7938955
2 True               3618668
```

Boxplot of Transported vs total bill:



9. How many passengers got transported?

Passengers count by Transported:

Transported	Count	Total %
False	4315.0	49.6
True	4378.0	50.4
Sum	8693.0	100.0

Conclusion

This Spaceship Titanic data has records of 8,693 passengers (observations) who were on the spaceship emigrating to new habitable planets. The dataset collects the passenger information including their Name, HomePlanet, Destination, Age, amenity usage like CryoSleep, their spending on luxury amenities, VIP status, and whether they were transported to another dimension.

After analyzing the data, it was found that the passengers were from Earth, Europa, and Mars emigrating to these three new destinations 55 Cancri e, PSO J318.5-22, and TRAPPIST-1e. Earth passengers were the largest group among all passengers (4,602 / 53%). TRAPPIST-1e was the destination where the majority of the passengers were debarking to (5,915 / 68%).

The passenger age ranges from youngest (less than 1 year old) to oldest 79 years old. The largest age group were those adult age 20-39 (4,497 / 53%). More passengers (5,439 / 64%) opted to not be put into CryoSleep for the duration of the voyage. There was very small group of passengers (199 / 2%) who paid for VIP service. The passengers from Europa were dominating the VIP service and they were also the biggest spenders on the luxury amenities.

Among the passengers who were transported to an alternate dimension, a combination of the following factors appear to show higher chance of being transported. The passengers from Europa (61%), those who would be debarking to 55 Cancri e (61%), those children age group between 0-12 (70%), those who opted to use CryoSleep (82%), and those who spent less on amenities and non-VIP.