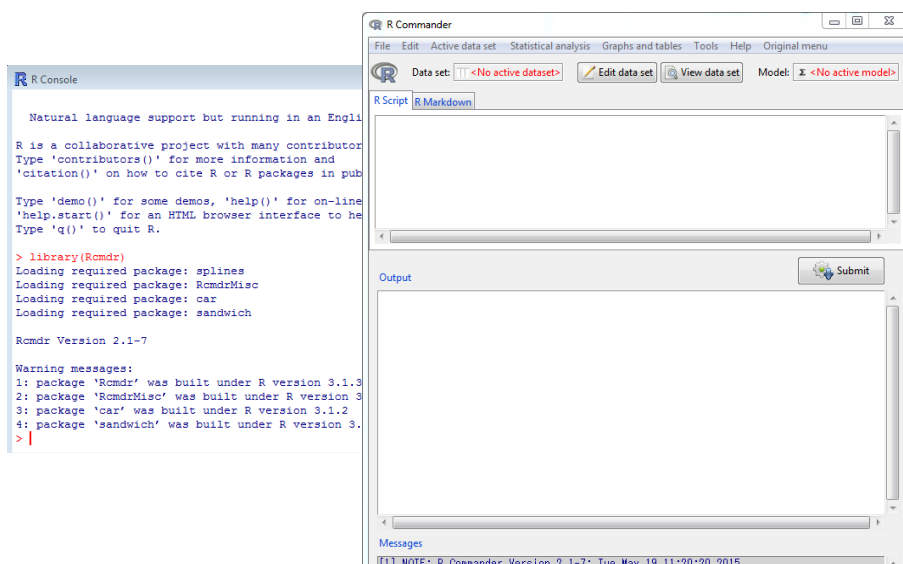


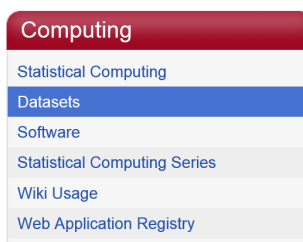
EZR as a R commander plugin



1

Dataset: Titanic

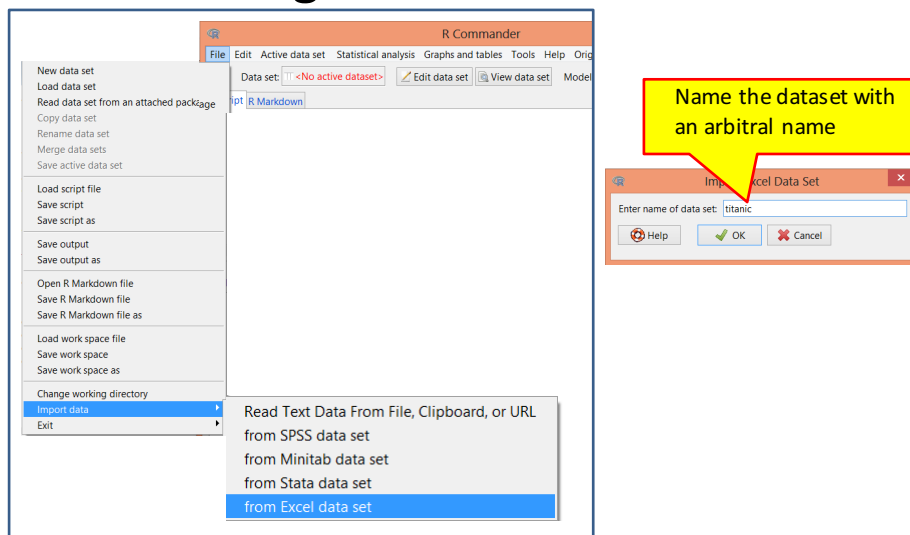
- Go to <http://biostat.mc.vanderbilt.edu/wiki>
- Click on Dataset



- Select titanic.xls

Data for Titanic passengers				
titanic.html	titanic.sav	titanic.sdd	NA	titanic.txt
NA	titanic2.sav	titanic2.sdd	NA	NA
NA	titanic3.sav	titanic3.sdd	titanic3.xls	titanic3.csv

Reading data into Rcmdr-EZR



Passenger Characteristics

	Died (N=809)	Survived (N=500)
Age, Mean (SD)	30.5 (13.9)	28.9 (15.1)
Fare, Mean (SD)	23.3 (34.1)	49.4 (68.6)
Passenger Class, N (%)		
1 st	123 (15.2%)	200 (40%)
2 nd	158 (19.5%)	119 (23.8%)
3 rd	528 (65.3%)	189 (36.2%)

Descriptive Statistics with EZR(R)

The screenshot shows the R Commander interface. The 'Statistical analysis' menu is open, and 'Numerical summaries' is selected. The 'Output' window displays the following text:

```
#####Numerical summaries#####  
mean    sd    0% 25% 50% 75% 100%  n NA  
29.88113 14.4135 0.1667 21 28 39 80 1046 263
```

The 'Numerical Summaries' dialog box is also visible, showing a list of variables (age, fare, parch, pclass, sibsp, survived) and options for 'Show graph', 'Mean', 'Standard Deviation', and 'Quantiles'.

Describing data by groups (pclass)

The screenshot shows the R Commander interface. The 'Variables' menu is open, and 'Convert numeric variables to factors' is selected. A yellow callout box points to this menu item with the text: "Convert pclass to a factor variable".

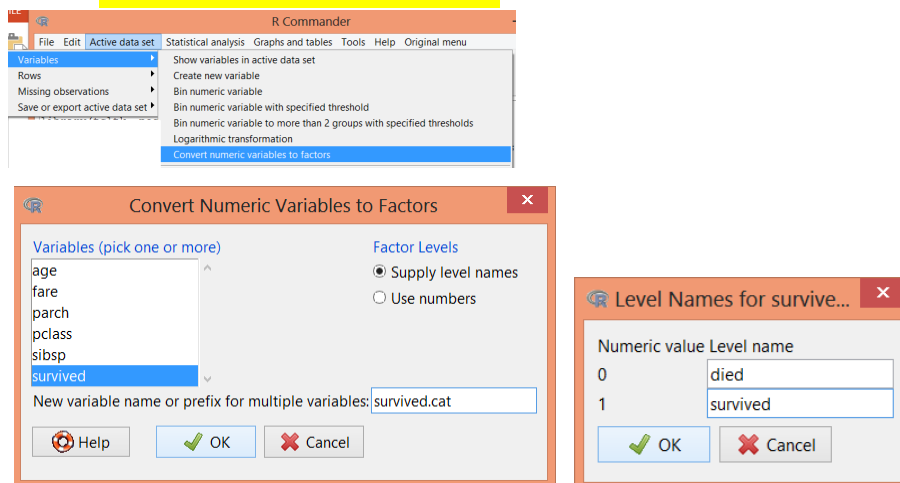
The 'Convert Numeric Variables to Factors' dialog box is open, showing a list of variables (age, fare, parch, pclass, sibsp, survived) and options for 'Factor Levels' (Supply level names, Use numbers). The 'New variable name or prefix for multiple variables' field is set to 'pclass.fac'.

The 'Level Names for pclass.f...' dialog box is also open, showing a table with numeric values and level names:

Numeric value	Level name
1	1st class
2	2nd class
3	3rd class

Describing data by groups (pclass)

Convert "survived" to a factor variable



The image shows two screenshots from the R Commander interface. The first screenshot shows the 'Statistical analysis' menu with 'Frequency distributions' selected. The second screenshot shows the 'Frequency Distributions' dialog box with 'survived.cat' selected in the 'Variables (pick one or more)' list, 'Chi-square goodness-of-fit test (for one variable only)' checked, 'Show missing data' checked, 'Show percent' checked, and 'Show graph' unchecked.

```
> #####Frequency Distributions#####

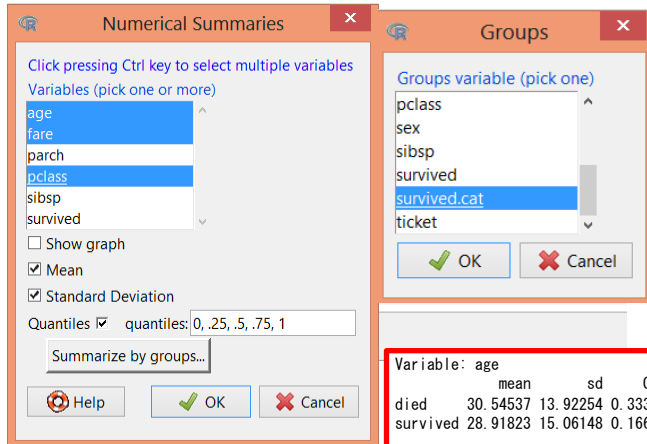
> (Table <- table(titanic$survived.cat,
exclude=NULL))
+ # counts for survived.cat

died survived <NA>
809 500 0

> round(100*Table/sum(Table), 2) #
percentages for survived.cat

died survived <NA>
61.8 38.2 0.0
```

Describing data by groups (survived)



The 'Numerical Summaries' dialog box shows the following options:

- Variables (pick one or more): age, fare, parch, pclass, sibsp, survived
- Show graph: ☐
- Mean: ☒
- Standard Deviation: ☒
- Quantiles: ☒ quantiles: 0, .25, .5, .75, 1
- Summarize by groups...: ☐

The 'Groups' dialog box shows the following options:

- Groups variable (pick one): pclass, sex, sibsp, survived, survived.cat, ticket

The following summary statistics are displayed for the variable 'age' (grouped by 'survived'):

Variable: age	mean	sd	0%	25%	50%	75%	100%	n	NA
died	30.54537	13.92254	0.3333	21	28	39	74	619	190
survived	28.91823	15.06148	0.1667	20	28	38	80	427	73

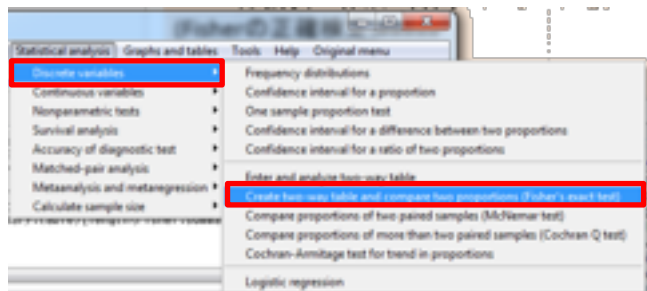
The following summary statistics are displayed for the variable 'fare' (grouped by 'survived'):

Variable: fare	mean	sd	0%	25%	50%	75%	100%	n	NA
died	23.35383	34.1451	0	7.8542	10.5	26.00	263.0000	808	1
survived	49.36118	68.6488	0	11.2146	26.0	57.75	512.3292	500	0

The following summary statistics are displayed for the variable 'pclass' (grouped by 'survived'):

Variable: pclass	mean	sd	0%	25%	50%	75%	100%	n	NA
died	2.500618	0.7448251	1	2	3	3	3	809	0
survived	1.962000	0.8729720	1	1	2	3	3	500	0

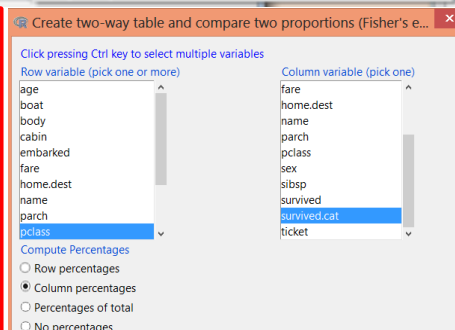
Cross-tabulation table



The menu path is: Statistical analysis > Graphs and tables > Tools > Help > Original menu > Discrete variables > Create two-way table and compare two proportions (Fisher's exact test).

```
> .Table
      survived. cat
pclass died survived
1      123      200
2      158      119
3      528      181

> colPercents(.Table) # 列のパーセント表示
      survived. cat
pclass died survived
1      15.2      40.0
2      19.5      23.8
3      65.3      36.2
Total 100.0 100.0
Count 809.0 500.0
```



The dialog box shows the following options:

- Row variable (pick one or more): age, boat, body, cabin, embarked, fare, home.dest, name, parch, pclass
- Column variable (pick one): fare, home.dest, name, parch, pclass, sex, sibsp, survived, survived.cat, ticket
- Compute Percentages: ☒ Row percentages, ☐ Column percentages, ☐ Percentages of total, ☐ No percentages

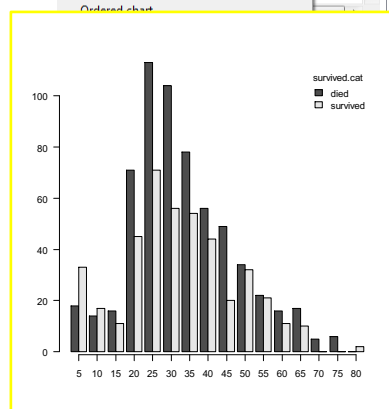
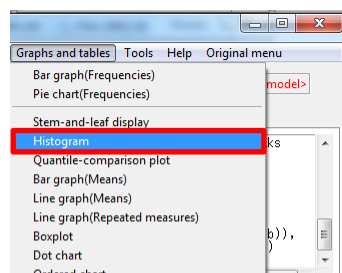
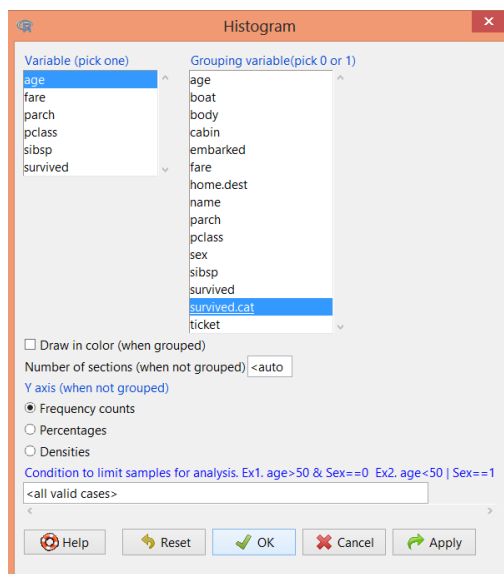
Passenger Characteristics

	Died (N=809)	Survived (N=500)
Age, Mean (SD)	30.5 (13.9)	28.9 (15.1)
Fare, Mean (SD)	23.3 (34.1)	49.4 (68.6)
Passenger Class, N (%)		
1 st	123 (15.2%)	200 (40%)
2 nd	158 (19.5%)	119 (23.8%)
3 rd	528 (65.3%)	189 (36.2%)

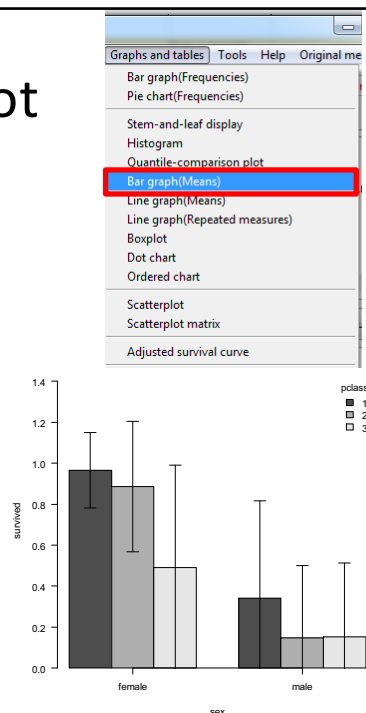
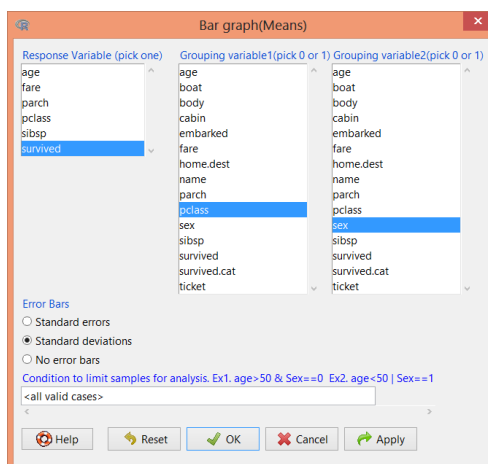


Is there a better way to show the data?

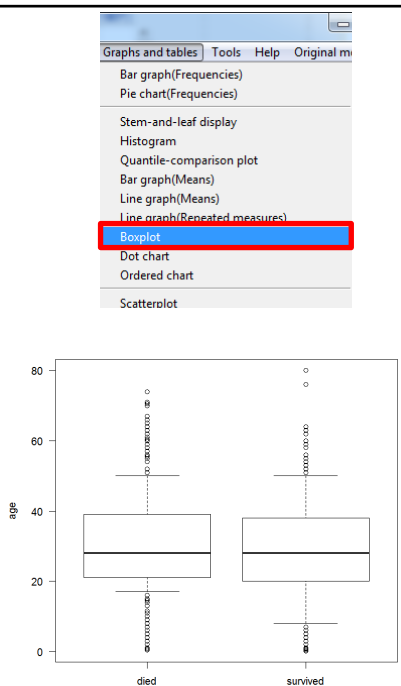
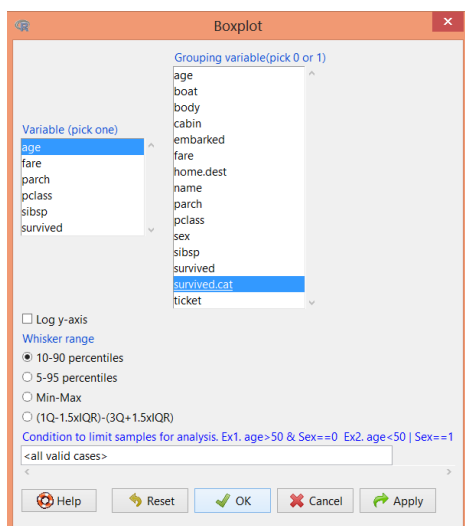
Graph



Bar-plot



Box-plot



Scatter-plot

