



VISUALIZATION

SYRACUSE UNIVERSITY
School of Information Studies

DATA VISUALIZATION

Complementary approach for data analysis in a visual way

Basic tools for numeric variables

Histogram: Show distribution of one variable.

Box plot: Use five key values to show distribution.

Scatter plot: Plot relationship between two variables.

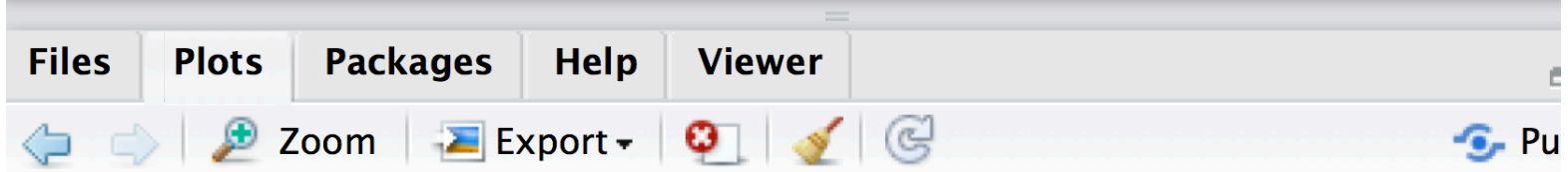
Basic tools for nominal variables

Pie chart or bar chart to show frequencies

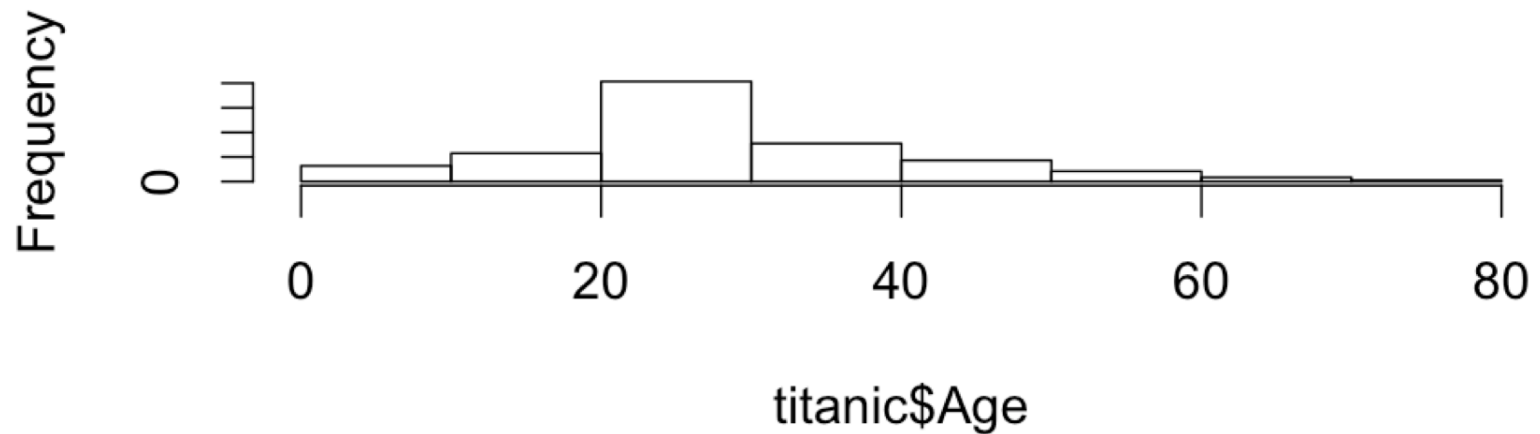
Cross-tab to show relationship between two variables

HISTOGRAM

```
> hist(titanic$Age)
>
```



Histogram of titanic\$Age



BOX PLOT

Five values of box plot

Bottom of box: Q1

Top of box: Q3

Band near the middle of box: Median

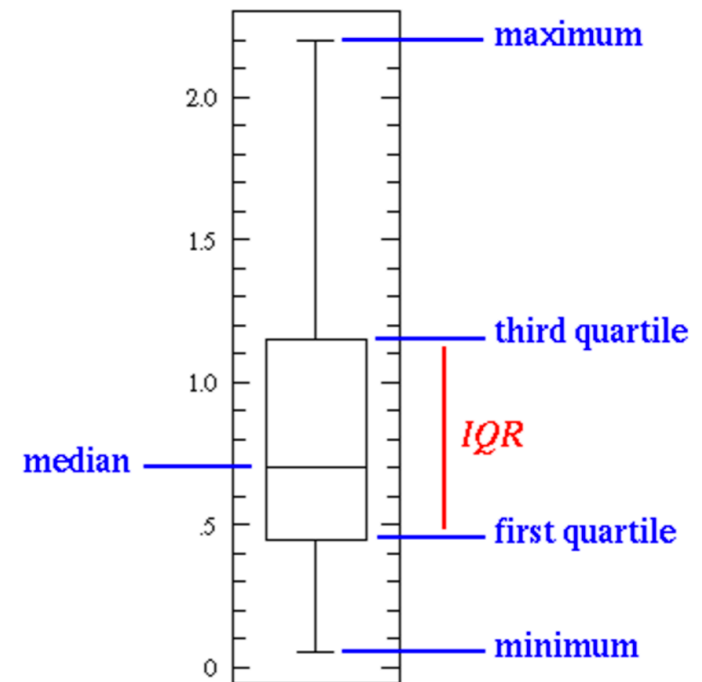
Upper whisker = $\min(\max, Q3 + 1.5IQR)$

Lower whisker = $\max(\min, Q1 - 1.5IQR)$

Use visualization for outlier detection

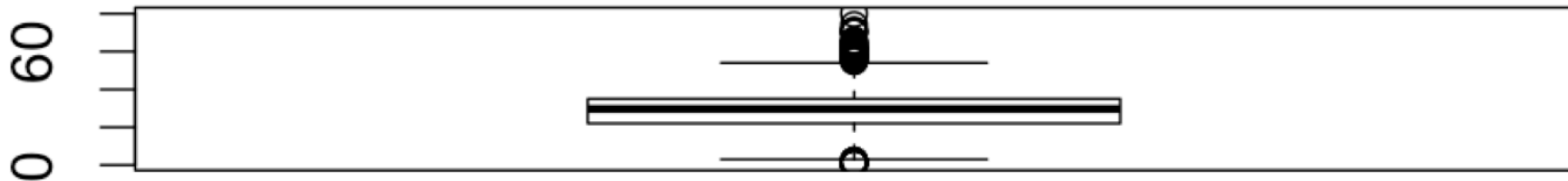
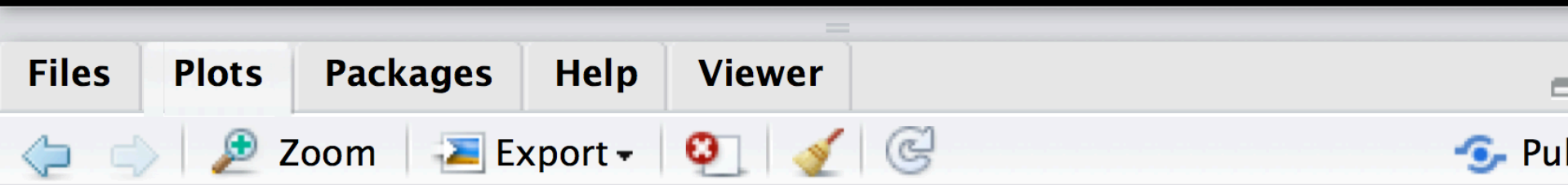
<http://www.stat.wmich.edu/s160/book/node8.html>

<http://www.r-bloggers.com/about-boxplot/>



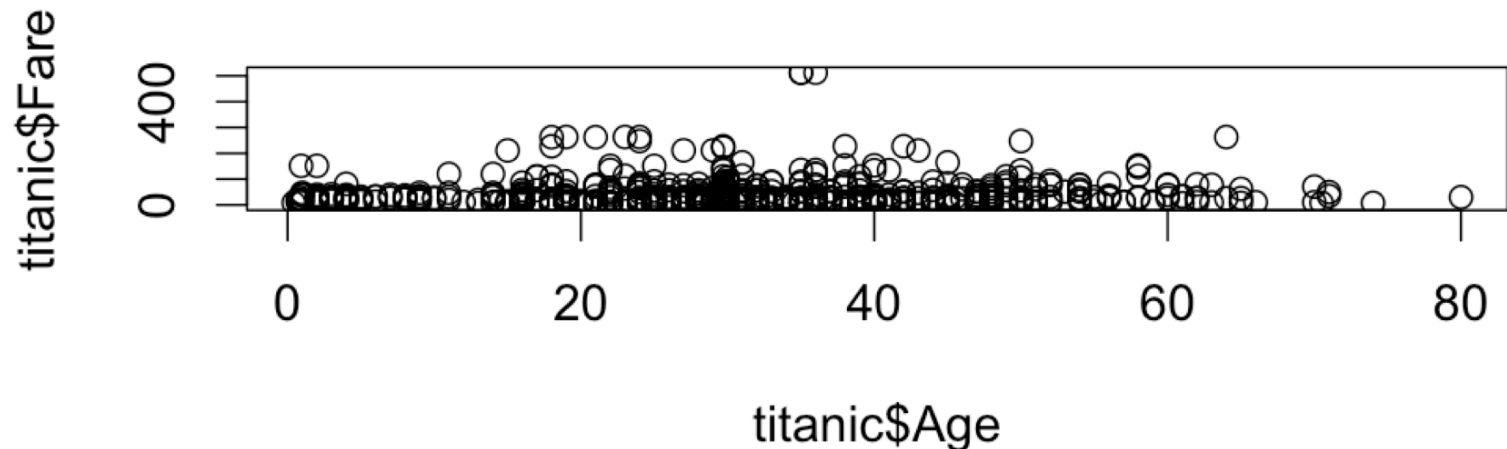
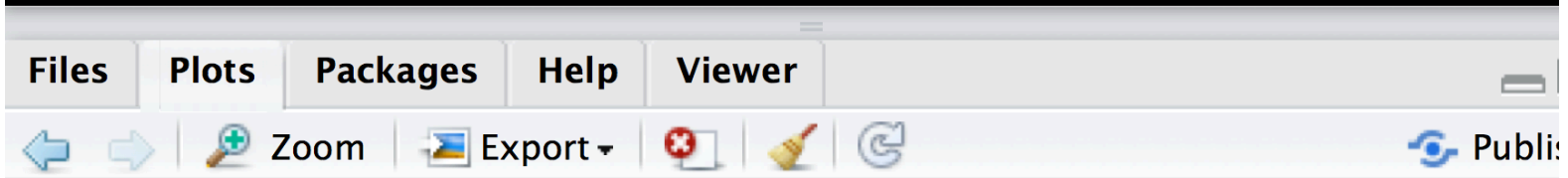
BOX PLOT

```
> boxplot(titanic$Age)
> |
```



SCATTER-PLOT TWO NUMERIC VARIABLES

```
> plot(titanic$Age, titanic$Fare)
```



CROSS-TAB TWO NOMINAL VARIABLES

```
> table(titanic$Sex, titanic$Survived)
```

	0	1
female	81	233
male	468	109

RELATIONSHIP BETWEEN A NOMINAL AND A NUMERIC VARIABLE

```
> male=titanic[titanic$Sex=='male',]  
> mean(male$Fare)  
[1] 25.52389  
> female=titanic[titanic$Sex=='female',]  
> mean(female$Fare)  
[1] 44.47982
```


CORRELATION VS. CAUSATION

Did ice cream cause drowning?

<https://www.youtube.com/watch?v=8B271L3NtAw>

“Cargo Cult Science” (Richard Feynman)

<http://calteches.library.caltech.edu/51/2/CargoCult.htm>

Search “a man named Young” in the article, and read that story.