

# Introduction to statistics

Data is usually sorted into a table.

The rows are the **cases** or **individuals** or **participants** or **observations** the data is about.

The columns are the **variables** or characteristics of the observations.

The inside of the table contains the **values** of the variables.

Variables can be **quantitative / numeric** (thought of as a number) or **categorical** (thought of as a category).

name	brand	IsKellogs	calories	sugars	carbo	protein	fat	sodium	fiber	potass	shelf
100%_Bran	Nabisco	no	70	6	5	4	1	130	10	280	Lower
100%_Natural_Bran	Quaker	no	120	8	8	3	5	15	2	135	Lower
All-Bran	Kellogs	yes	70	5	7	4	1	260	9	320	Lower
All-Bran_with_Extra_Fiber	Kellogs	yes	50	0	8	4	0	140	14	330	Lower
Almond_Delight	Ralston	no	110	8	14	2	2	200	1	-1	Lower
Apple_Cinnamon_Cheerios	Gen-Mills	no	110	10	10.5	2	2	180	1.5	70	Upper

# Data Visualization: Types of Graphical Displays

## Single categorical variable

- Bar chart (preferred)
- Pie chart
- Frequency table

## Single quantitative variable

- Histogram (preferred)
- Dotplot
- Boxplot
- Stem-and-leaf display
- Density plot

## Two categorical variables

- Segmented / stacked bar chart (preferred)
- Clustered / grouped bar chart (preferred)
- Side-by-side pie charts
- Contingency table

## Two quantitative variables

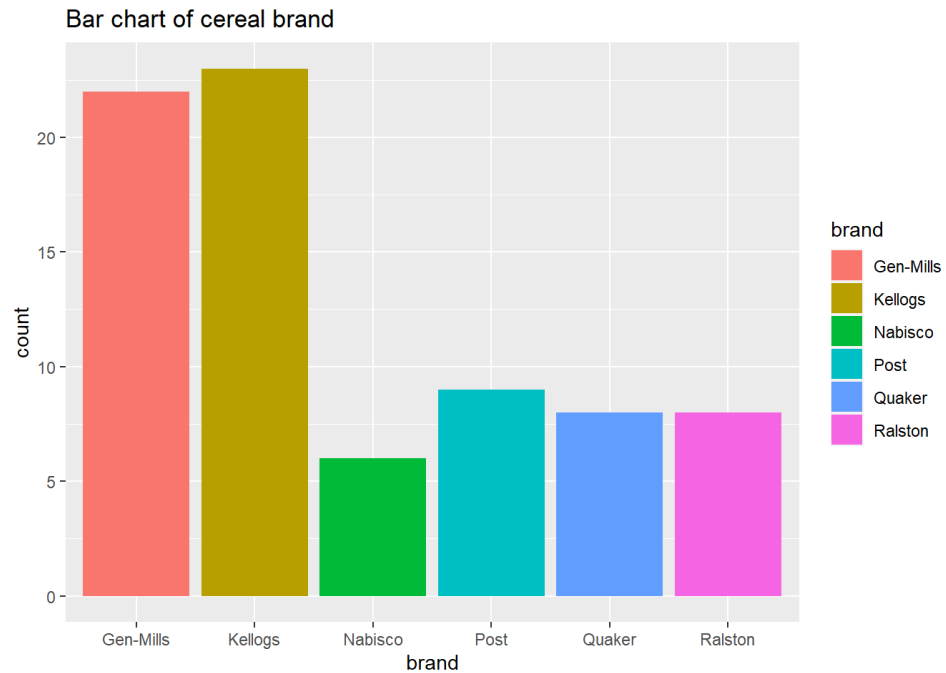
- Scatterplot

## One categorical, one quantitative variable

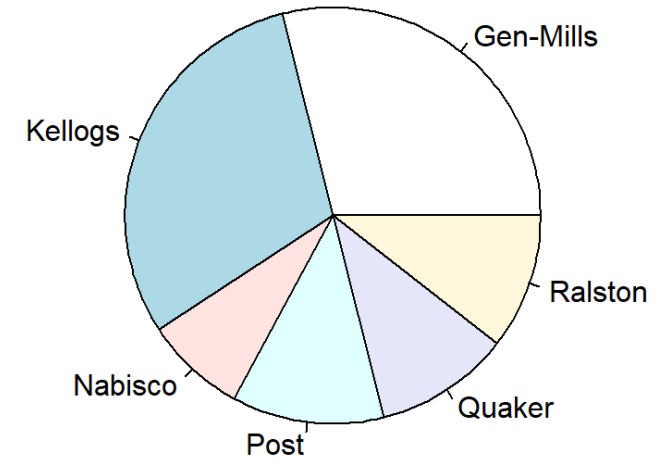
- Side-by-side boxplots (preferred)
- Overlapping histograms

# One categorical variable

## Bar chart



## Pie chart

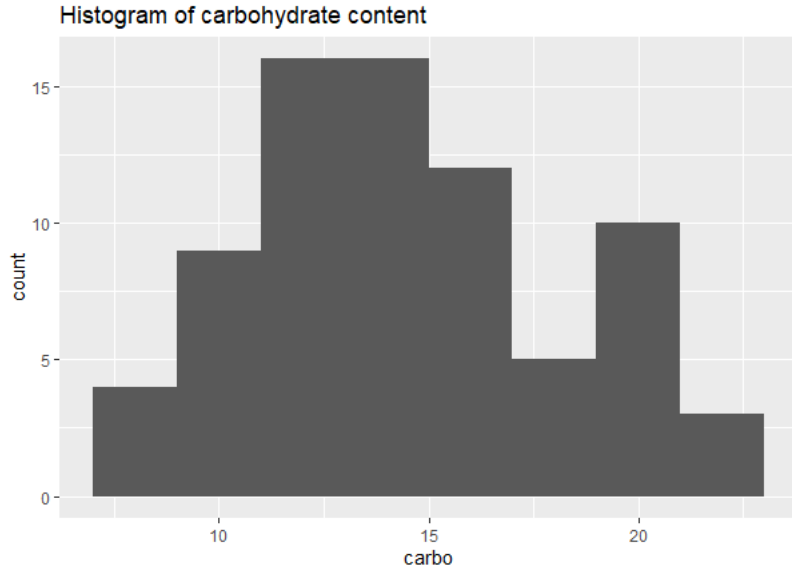


## Frequency table

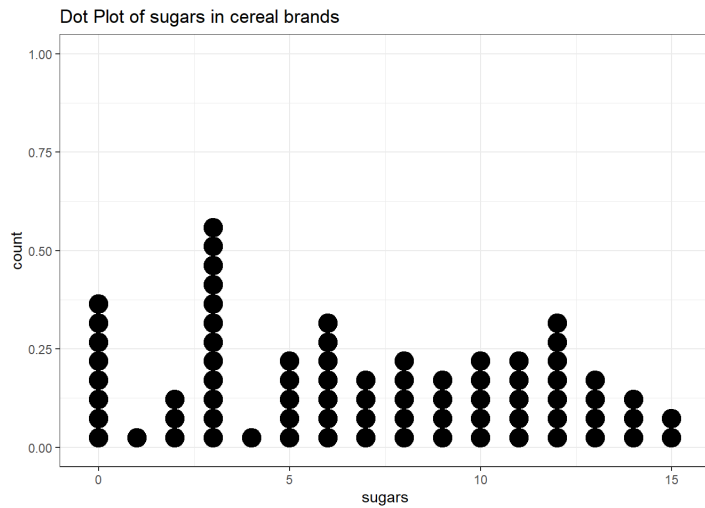
Gen-Mills	Kellogs	Nabisco	Post	Quaker	Ralston
22	23	6	9	8	8

# Single quantitative variable

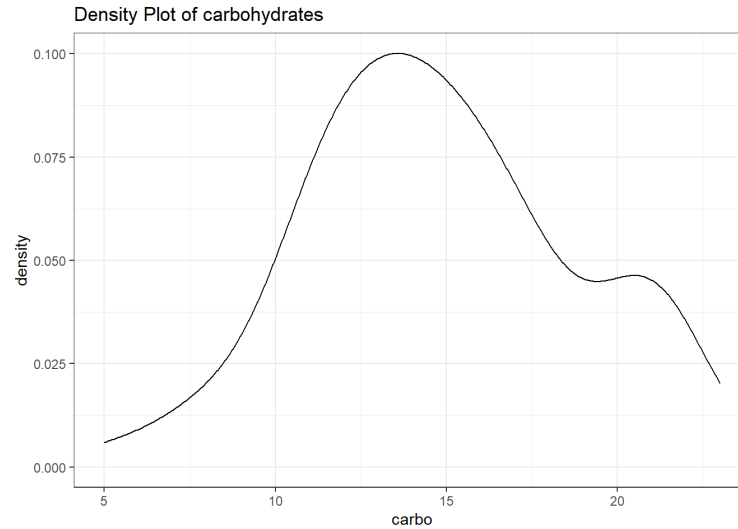
## Histogram



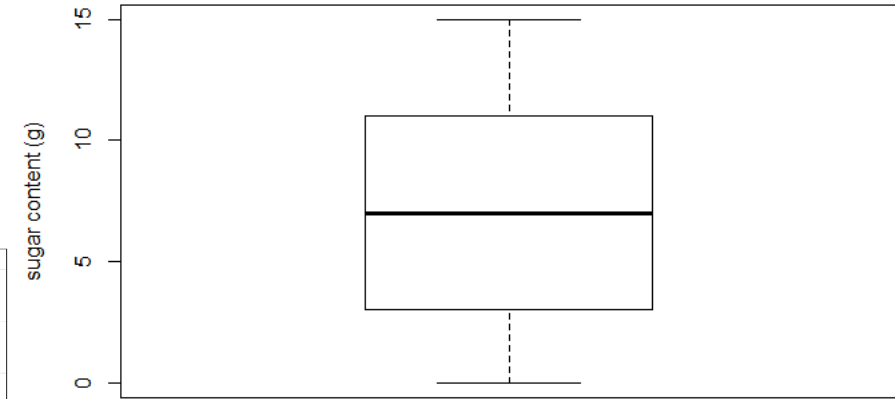
## Dotplot



## Density plot



## Boxplot



## Stem-and-leaf display

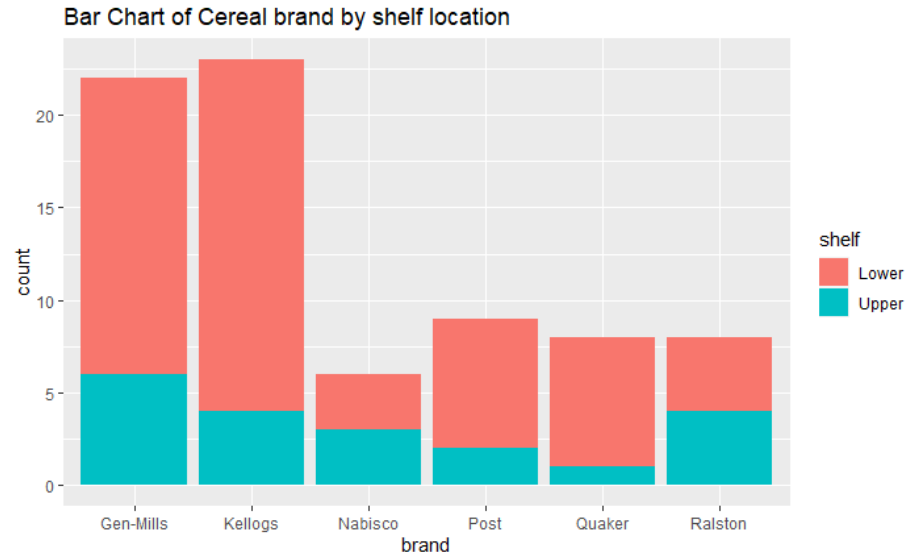
Potassium content

```

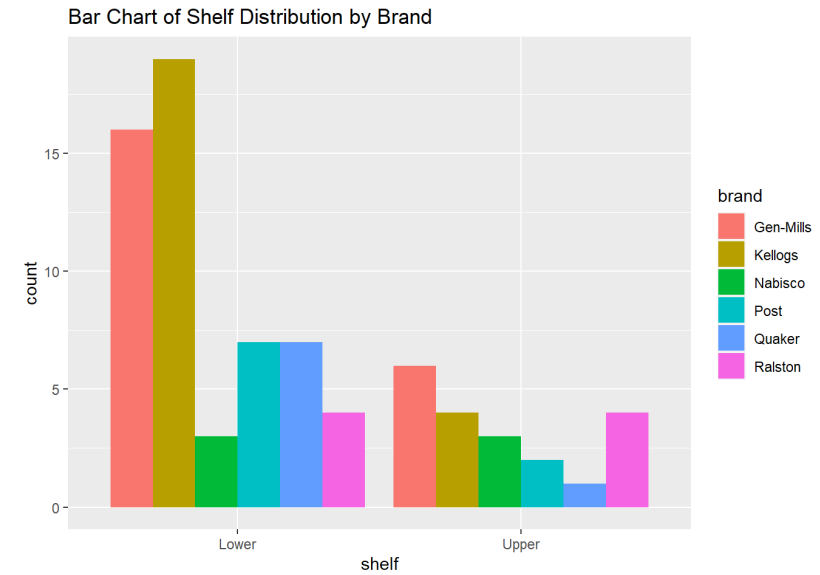
0 | 22333333334444444444
0 | 55555666666778999999
1 | 0000001111111222233444
1 | 667799
2 | 034
2 | 68
3 | 23
    
```

# Two categorical variables

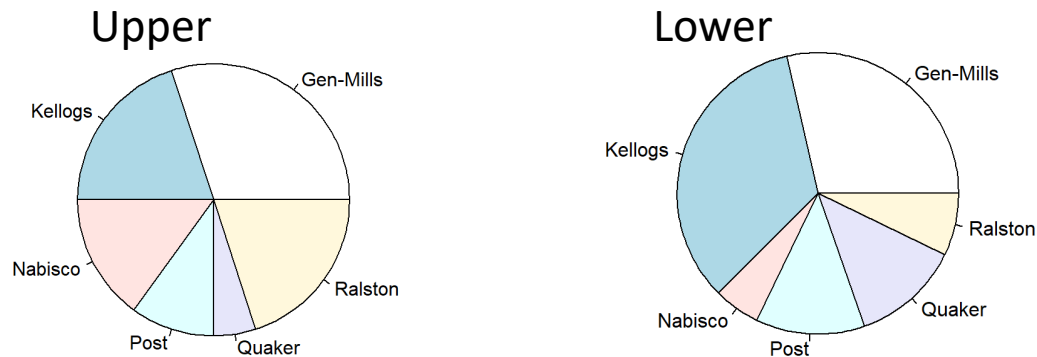
## Segmented / stacked bar chart



## Clustered / grouped bar chart



## Side-by-side pie charts

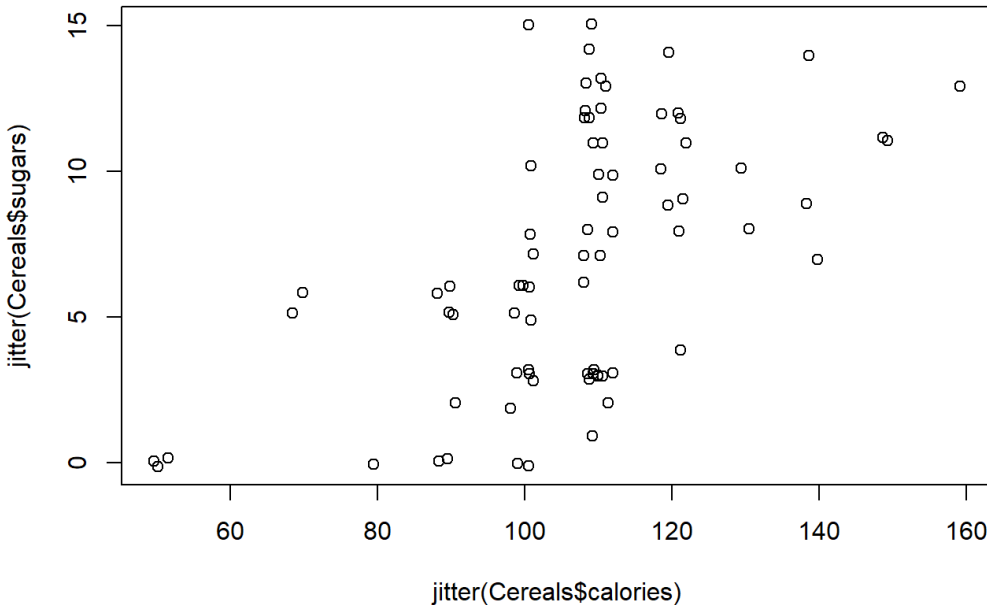


## Contingency table

	Lower	Upper
Gen-Mills	16	6
Kellogs	19	4
Nabisco	3	3
Post	7	2
Quaker	7	1
Ralston	4	4

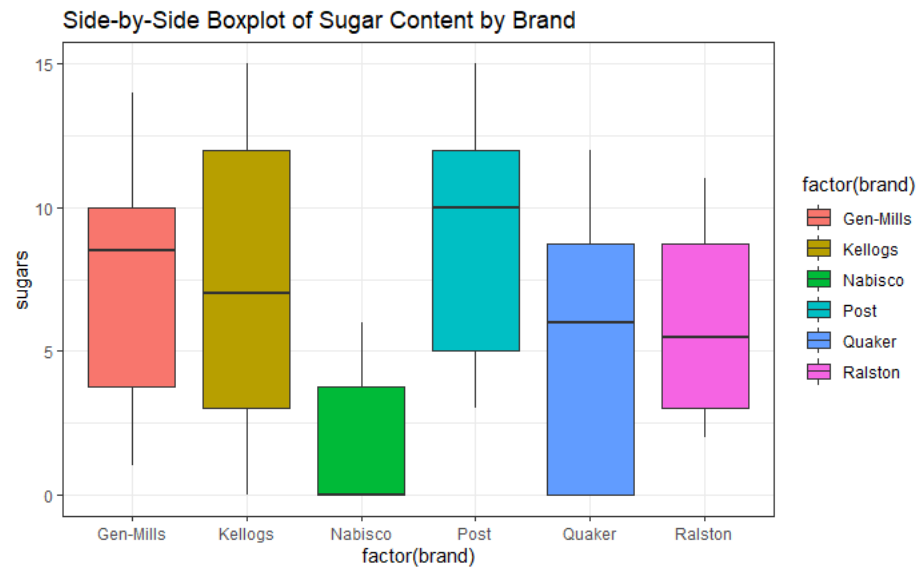
# Two quantitative variables

Scatterplot



# One categorical and one quantitative variable

## Side-by-side boxplots



## Overlapping histograms

