

# Data Reduction and Data Presentation

L4

L4

## Introduction

- ▶ There is a difference between data and information. Data are the raw numbers or facts which must be processed to give useful information.

*Thus 78, 64, 36, 70 and 52 are data which could be processed to give the information that the average mark of five students sitting an exam is 60%.*

L4

2

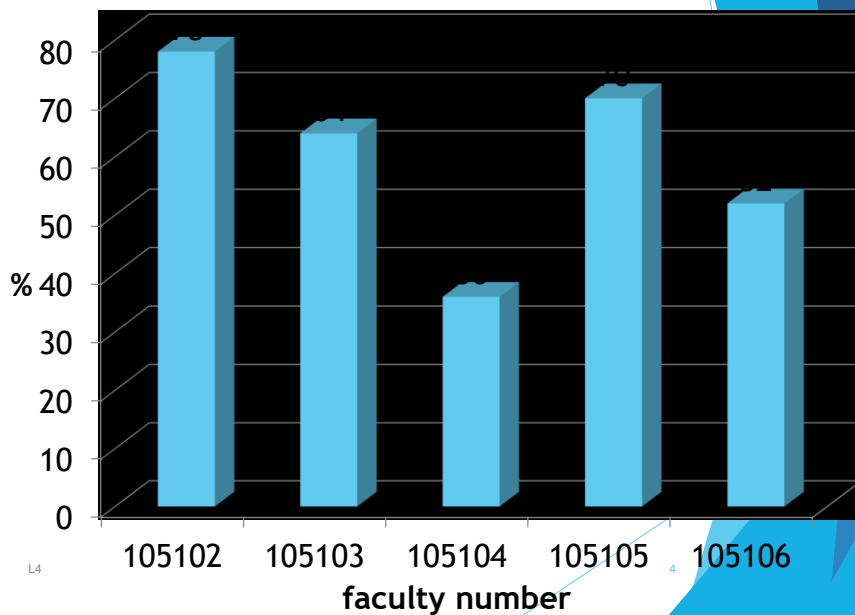
## Introduction

- ▶ The purpose of **data presentation** is to show the characteristics of a set of data and highlight any important patterns. This can either be done **numerically**, or by using **diagrams**.
- ▶ The main **benefit** of using diagrams is that people are good at **recognizing patterns** and can extract a lot of information in a short time.

L4

3

## Results



L4

4

## Data reduction

- ▶ Most people can deal with numerical data. Problems begin when there are a **lot of data** and we are swamped with detail.
- ▶ In most cases we are not interested in the small detail, but really want the **overall picture**.
- ▶ What we need, then, is a way of **identifying general patterns** in data and presenting a **summary** which allows these to be seen.

L4

5

## Data reduction

- ▶ The aim of data reduction is to give a simplified and accurate view of the data which shows the underlying patterns but does not overwhelm us with details.

L4

6

## Data reduction

- ▶ Data reduction has a number of clear **advantages**:
  - + results are shown in a compact form
  - + results are easy to understand
  - + graphical or pictorial representations can be used
  - + overall patterns can be seen
  - + comparisons can be made between different sets of data
  - + quantitative measures can be used

L4

7

## Data reduction

- ▶ Conversely, it has the **disadvantages** that:
  - details of the original data are lost
  - the process is irreversible

L4

8

## Diagrams for presenting data

- ▶ There are several ways in which data can be **summarized in diagrams**, and we shall classify the most important of these as:
  - **tables** of numerical data
  - **graphs** to show relationships between variables
  - **pie charts**, bar charts and pictograms showing relative frequencies
  - **histograms** which show relative frequencies of continuous data

L4

9

## Diagrams for presenting data

- ▶ **Guidelines**
  - select the most suitable format for the purpose
  - present data fairly and honestly
  - make sure any diagram is clear and easy to understand
  - give each diagram a title

L4

10

## Diagrams for presenting data

- state the source of data
- use consistent units and say what these units are
- label axes clearly and accurately
- put a clear scale on axes
- include totals, subtotals and any other useful summaries
- add notes to highlight reasons for unusual or atypical values.

L4

11

## Tables

- ▶ This is perhaps the most widely used method of data presentation.

Week	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Total
1	51	84	49	30	214
2	60	91	44	32	227
3	58	82	41	30	211
4	56	78	45	32	211
5	62	76	38	31	207
6	69	75	28	29	201
7	58	66	37	30	191
8	76	57	40	41	214
9	80	78	42	45	245
10	82	65	22	44	213
11	68	50	25	47	190
12	90	61	26	53	230
13	72	L4 54	21	54	201 <sup>12</sup>
Totals	882	917	458	498	2 755

## Tables

- ▶ In this format, though, the table is still really a presentation of the **raw data** and it is difficult to get a feel for a typical week's sales; there is no indication of minimum or maximum sales; and so on.
- ▶ These defects would be even more noticeable if there were hundreds or thousands of observations.
- ▶ It would be useful to **reduce** the data and emphasize the patterns.

L4

13

## Tables

	A	B	C	D	E	F
1	<b>Week</b>	<b>Quarter 1</b>	<b>Quarter 2</b>	<b>Quarter 3</b>	<b>Quarter 4</b>	<b>Per year</b>
2	1	51	84	49	30	
3	2	60	91	44	32	
4	3	58	82	41	30	
5	4	56	78	45	32	
6	5	62	76	38	31	
7	6	69	75	28	29	
8	7	58	66	37	30	
9	8	76	57	40	41	
10	9	80	78	42	45	
11	10	82	65	22	44	
12	11	68	50	25	47	
13	12	90	61	26	53	
14	13	72	54	21	54	
15	Total per quarter	882	917	458	498	2755
16	max per quarter	90	91	49	54	91
17	min	51	50	21	29	21
18	average	68	71	35	38	53

L4

14

## Frequency table

- ▶ The minimum sales are 21, so we might start by seeing how many weeks had sales in a range of, say, 20 to 29. If we count these, there are six weeks. Then we could count the number of observations in other **ranges**, as follows:

Range of sales	Number of weeks
20 to 29	6
30 to 39	8
40 to 49	10
50 to 59	9
60 to 69	7
70 to 79	6
80 to 89	4
90 to 99	2

- ▶ This table shows how many values are in each **range**, and it is called a **frequency table**

L4

15

## Frequency table

- ▶ The 'ranges' are usually referred to as **classes**. Then we can talk about the '**class of 20 to 29**', where 20 is the **lower** class limit and 29 is the **upper** class limit and the class **width** is  $29 - 20 = 9$ . We arbitrarily chose classes of 20 to 29, 30 to 39, and so on, but could have used any appropriate classes.
- ▶ The only constraint is that there should be enough classes to make any patterns clear, but not so many that they are obscured.

L4

16



## Tables

- ▶ Drawing tables needs a compromise between making them **too long** (where lots of details can be seen, but they are complicated with underlying patterns hidden) and **too short** (where underlying patterns are clear, but most details are lost).
- ▶ The number of classes, in particular, must be a **subjective decision** based on the use of the presentation, but a guideline would set a maximum number at about **ten**.

## Likert scale

- ▶ A Likert scale is a psychometric scale commonly involved in research that employs questionnaires. It is the most widely used approach to scaling responses in survey research, such that the term is often used interchangeably with rating scale, although there are other types of rating scales.
- ▶ The Likert scale is a five (or seven) point scale which is used to allow the individual to express how much they agree or disagree with a particular statement.

## 4 point Likert scale

How often do you re

- Never
- Rarely
- Often
- Every time

- ▶ 4 point Likert scale is basically a forced Likert scale. The reason it is named as such is that the user is forced to form an opinion. There is no safe 'neutral' option. Ideally a good scale for market researchers, they make use of the 4 point scale to get specific responses.

L4

19

## 5 point Likert scale



- ▶ 5 point Likert scale consist of 5 answer options which will contain two extreme poles and a neutral option connected with intermediate answer options.

### + Pros of a 5 Point scale

- ▶ It is relatively easier for respondents to understand.
- ▶ A 5 point scale is ideal for a larger study.
- ▶ 5 point Likert scales tend to produce better distributions of data

L4

20

## 5 point Likert scale

Agreement	Importance	Quality
Strongly Agree	Very Important	Excellent
Agree	Important	Good
Undecided	Moderately Important	Fair
Disagree	Slightly Important	Poor
Strongly Disagree	Unimportant	Very Poor

Frequency	Likelihood	Likelihood
Always	Almost Always True	Definitely
Often	Usually True	Probably
Sometimes	Occasionally True	Possibly
Rarely	Usually Not True	Probably Not
Never	Almost Never True	Definitely Not

## 5 point Likert scale

Response Set	1	2	3	4	5
Frequency	Never	Rarely	Sometimes	Often	Always
Quality	Very poor	Poor	Fair	Good	Excellent
Intensity	None	Very mild	Mild	Moderate	Severe
Agreement	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Approval	Strongly disapprove	Disapprove	Neutral	Approve	Strongly approve
Awareness	Not at all aware	Slightly aware	Moderately aware	Very aware	Extremely aware
Importance	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Familiarity	Not at all familiar	Slightly familiar	Moderately familiar	Very familiar	Extremely familiar
Satisfaction	Not at all satisfied	Slightly satisfied	Moderately satisfied	Very satisfied	Completely satisfied
Performance	Far below standards	Below standards	Meets standards	Above standards	Far above standards

## 7 point Likert scale

- ▶ A 7 point Likert scale offers 7 different answer options related to an agreement that would be distinct enough for the respondents, without throwing them into confusion. Typically, it includes a moderate or neutral midpoint, and 7 point Likert scales are known to be most accurate of the Likert scales.
- ▶ **Pros of a 7 Point scale**
  - ▶ It is the most accurate of the Likert scales
  - ▶ It is easier to use
  - ▶ It gives a better reflection of a respondent's true evaluation.
  - ▶ The best solution for questionnaires such as those used in usability evaluations.

L4

23

## 7 point Likert scale

What's your inclination about US present gun policy?

- Strongly Disagree
- Disagree
- Somewhat Disagree
- Neither Agree nor Disagree
- Somewhat Agree
- Agree
- Strongly Agree

L4

24

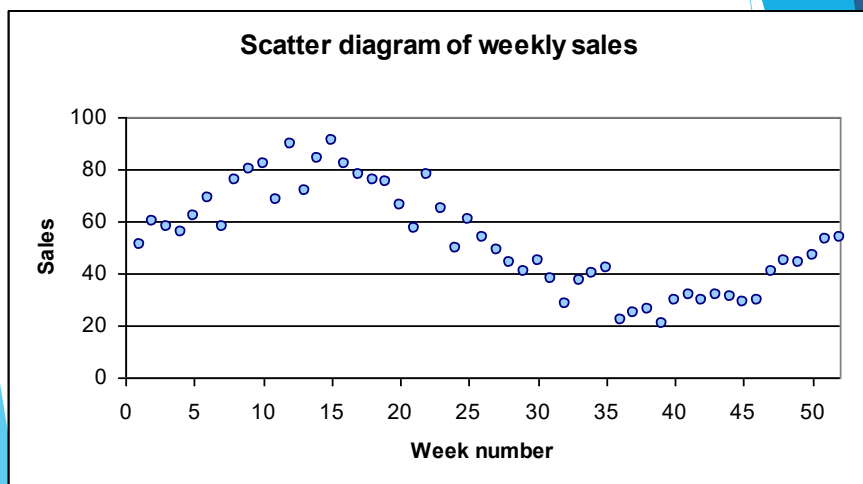
## Graphs

- ▶ A graph shows the relationship between two variables on a pair of rectangular axes, where:
  - the horizontal or **x axis** shows the variable that is responsible for a change (the independent variable)
  - the vertical or **y axis** shows the variable that we are trying to explain (the dependent variable)

L4

25

## Graphs



L4

26

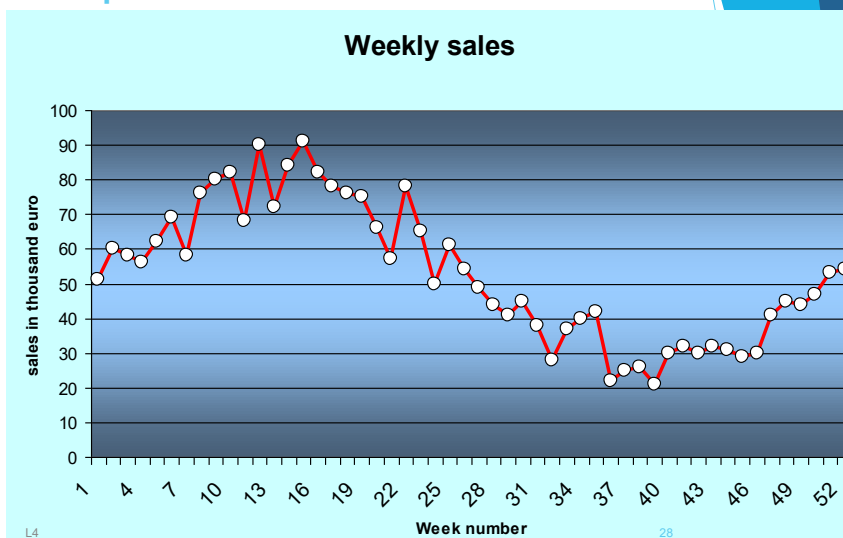
## Graphs

- ▶ The sales clearly follow a seasonal cycle with peak sales around week 12 and lowest sales around week 38. There are small random variations away from this overall pattern, so the graph is not a smooth curve.

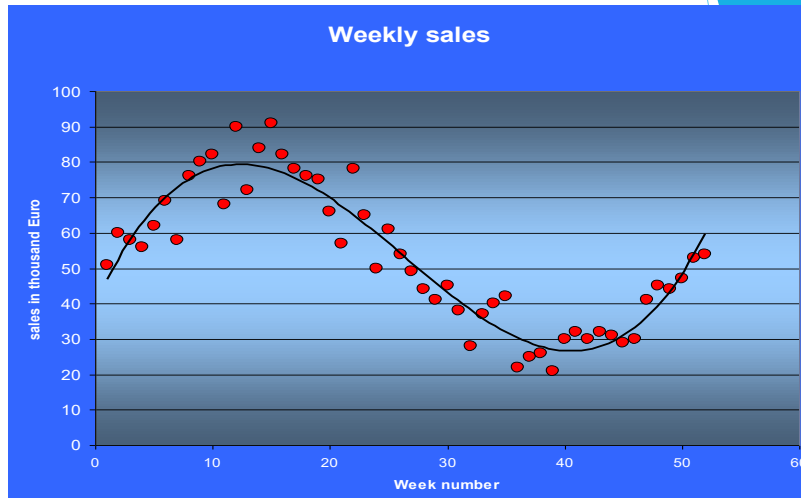
L4

27

## Graphs



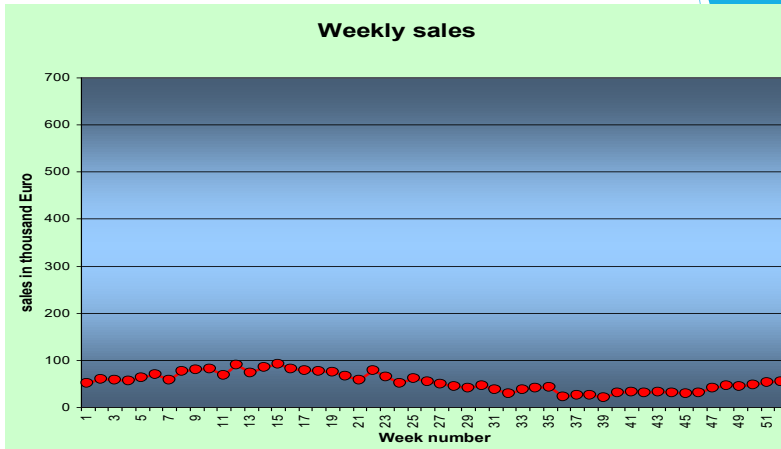
## Graphs



## Graphs

- ▶ As graphs give a very strong initial impact, **the choice of scale** for the axes is clearly important, with a bad choice giving a false view of the data. Although the choice of scale is largely subjective, some guidelines for good practice can be given:
  - ▶ always label the axes clearly and accurately
  - ▶ show the scales on both axes
  - ▶ the maximum of the scale should be slightly above the maximum observation
  - ▶ wherever possible the scale on axes should start at zero.
  - ▶ where appropriate, give the source of data

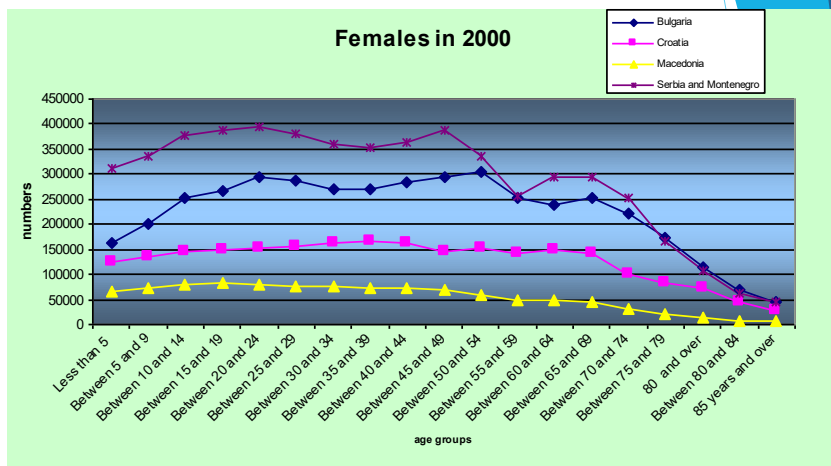
## Graphs



L4

31

## Graphs



L4

32



## Graphs

### ► *IN SUMMARY*

Graphs show clear relationships between two variables. Underlying patterns are easily identified and different sets of data can be compared. Care must be taken in choosing appropriate scales for the axes.

L4

33

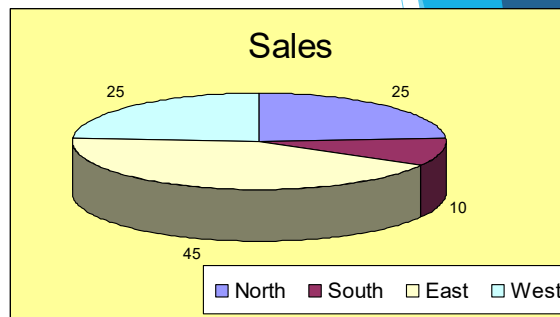
## Pie charts

- **Graphs** are good at showing **relationships** between two variables, but other methods of presenting data rely more directly on pictures. **Pie charts** are simple diagrams that are used for comparisons of limited amounts of information.
- To draw a pie chart the data are first classified into distinct categories. Then a circle is drawn (the pie) which is divided into sectors, each of which represents one category. The area of each sector (and hence the angle at the center of the circle) is **proportional** to the number of observations in the category.

L4

## Pie charts

<i>Region</i>	<i>Sales</i>
North	25
South	10
East	45
West	25
Total	105



L4

35

## Bar charts

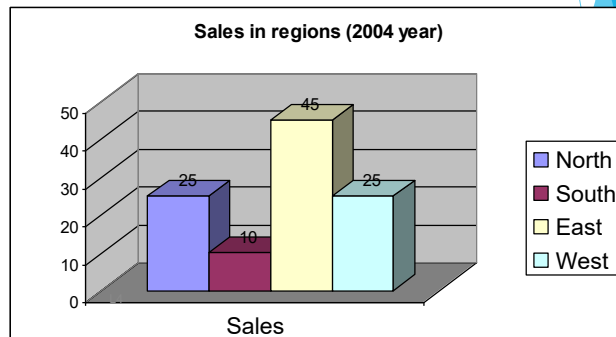
- ▶ Like pie charts, **bar charts** are diagrams that show the number of observations in different categories of data. This time, though, the numbers of observations are shown by lines or bars rather than sectors of a circle.
- ▶ In a bar chart, each category of data is represented by a different bar, and the **length of the bar is proportional to the number of observations**. Bar charts are usually drawn vertically, but they can be horizontal, and there are many adjustments that enhance their appearance. One constant **rule**, however, is that **the scale must start at zero**; any attempt to save space or expand the vertical scale by omitting the lower parts of bars is simply confusing.

L4

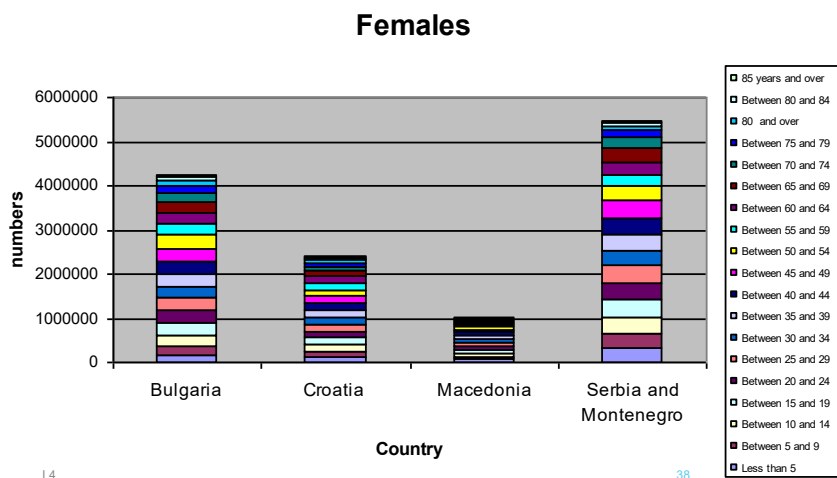
36

## Bar charts

- ▶ There are several different types of bar chart and the most appropriate is, again, a matter of choice. We should, however, remember that the purpose of diagrams is to present the characteristics of the data clearly; it is not necessarily to draw the prettiest picture.



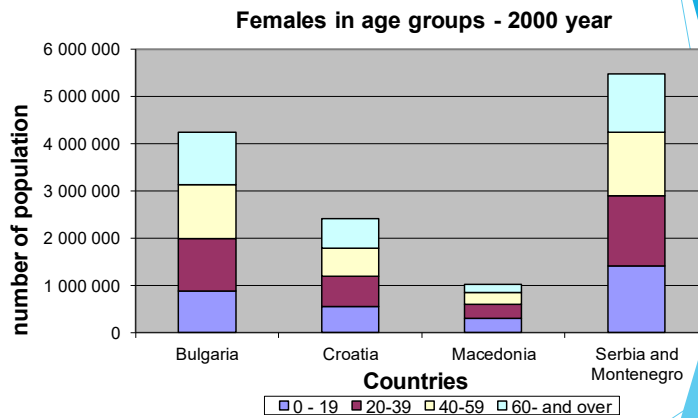
## Bar charts



L4

38

## Bar charts



L4

39

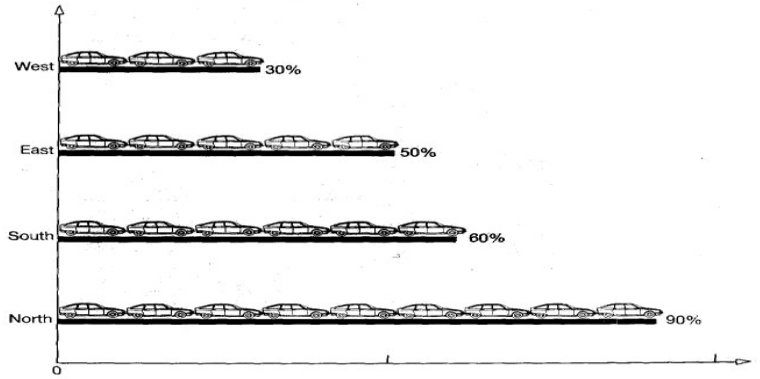
## Pictograms

- ▶ These are similar to bar charts, except that the bars are replaced by sketches of the things being described. Thus the percentage of people owning cars might be represented as in the next slide. In this pictogram, each 10% of people are represented by one car.

L4

40

# Pictograms



Pictogram showing percentage of people with cars.

L4

41

# Pictograms

DISPLAYR Features Use Cases Pricing Resources Login Back Your Demo

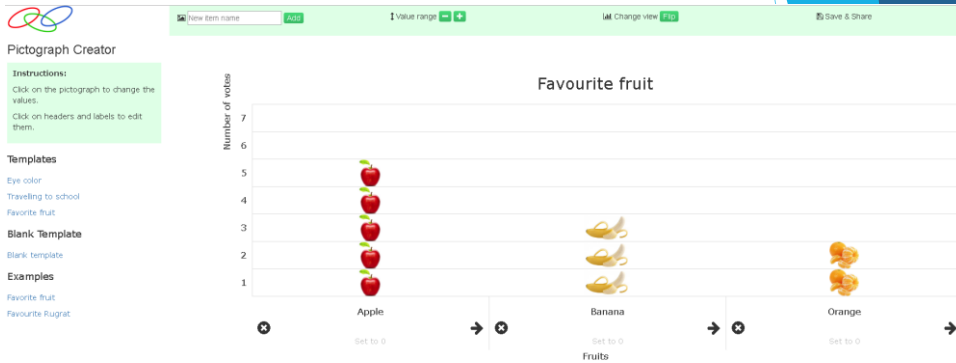
Displayr's pictograph maker has a range of options for you to quickly create your free pictograph.

- 43% approval (with a cartoon man icon)
- Ranking list (with stars and numbers)
- Telecommunication complaints (with icons and numbers)

L4

42

# Pictograms



L4

