Autor: Bernhard Jacobs, Medienzentrum der Philosophischen Fakultät der Universität des Saarlandes

Experimental Analyses of the Graphical Presentation of Data in Line Graphs and Bar Charts in Superposition and Juxtaposition

Summarization of my research on graphical perception of data from

http://www.uni-saarland.de/z-einr/MZ/graph/titel.html

Main Table of Contents

General Comments on All the Experiments

Point of Departure

As Jacobs (1990) was able to prove, graphical displays are clearly superior to tabular presentations for many tasks [1]. Thus the basic question of "**Should a graph be used?**" posed by Bertin (1974) can, at least for certain tasks, be answered in the affirmative. The more favorable visual perception in a graph makes complex relations easier to recognize and enables more demanding questions concerning a data set to be answered.

The amount of data becomes more complex when several data sets are to be displayed in one presentation. In this case graphical presentations are often necessary in order to make it possible for certain relations to be recognized within a reasonable amount of time. Whereas earlier research often concentrated on the analysis of relatively simple tasks using different graph formats, (e.g. Croxton & Stein 1932, Culbertson & Powers 1963, Feliciano et al. 1963), the current experiments are aimed at examining the trends of several data sets at an advanced experimental level. This involves testing the influence of graph arrangement and graph type on how well graphs are understood in order to find answers based on empirical findings to a further question posed by Bertin (1974), "**What type of graph should be used?**".

What Kinds of Questions Should Be Asked in Connection with Graphical Presentations?

There is a large number of questions that could be asked when extensive amounts of data are presented in a graphical display and they undoubtedly cannot all be examined in an experimental situation. One disadvantage of many studies is that only a few more or less arbitrary tasks are taken into account at all, which, at the end of the experiment, can lead to the suggestion that certain graph formats are better than others. What is really important is that the "right questions" are asked in connection with the presentation of data and in particular that the interaction between the graphical presentation and the question asked is taken into consideration.

In this experiment we shall above all examine questions (or tasks) that appear to be particularly suitable for graphical presentations (as opposed to tables). As Coll, Thyagaranjan & Chopra (1991), Coll (1992) and Coll, Coll & Thakur (1994) have shown, graphs are consistently superior to tables in "relational info questions". Bertin (1974) distinguishes three groups of questions in connection with graphical presentations:

1. Questions that only relate to one element of a data set, e.g.: What is the value of level x on the x-axis?
2. Questions that relate to a group of elements, e.g.: Do the values increase in the first three levels of X?
3. Questions that relate to a data set in its entirety, e.g.: What kind of trend does the data set display?

One of the main concerns of this research is finding questions concerning graphical presentations that are interesting from a theoretical point of view and relevant in practice. To this end tasks will be developed that relate to groups of elements (2) and in particular to data sets in their entirety (3). Answering these questions requires the use of a special type of perceptual approach (graphical chunks) that is rendered easier or more difficult depending on the graph type and the graph display. We are mainly interested in establishing whether tasks are solved more quickly due to a better perception of the individual conditions in a graph and what reason can be given for the improved perception. In the long term this could lead to new approaches to the problem, which would allow applied research to be linked to "pure research", e.g.: "How are certain visual variables grouped together, e.g. bars added up, or bar sizes compared? Which perceptual processes occur automatically and which require further selective searching? How can the process of answering a question be divided into steps?

Should Several Tasks Be Presented Using One Graph Type or Each Task Tested Using a Different Graph Type?

If there were no kind of interaction between the graph type and the task, one graphical form of presentation would suffice. But in practice there are many different kinds of graphical presentations in use and any graphics program that wishes to be taken seriously will include the appropriate options. Thus it is desirable to find out which is the ideal graphical presentation for a particular task.

Nevertheless, often one would also like to consider a data set from different perspectives and ask several questions concerning the same data. In the computer age this problem could, of course, be solved easily by using diagnostics, whereby according to the task concerned, the data are displayed in their ideal presentation form.

Another strategy would be to choose a graph format that allows most of the questions asked to be answered satisfactorily and which would thus, on average, be better than other presentations. Occasionally it is difficult to know what kind of task to set in connection with a particular graph, and it is the graph itself that sparks off ideas for possible questions. These graphs in particular must allow several tasks to be solved. Above all, this requires the skill on the part of the viewers to be able to recognize relations that they would not have noticed in the absence of a graph.

The present research is interested in both questions, in the strengths and weaknesses of the individual graph formats as well as the all-round qualities of certain graph formats. This approach means that tasks are set that cannot necessarily be answered in an optimal fashion using the graph types displayed.

The Graphical Presentation Forms Used in the Experiments

If several data sets are to be displayed at one time, each set can be presented in a separate graph (juxtaposition) or all sets can be shown in one single graph (superposition). For these two types of graph display, bar charts and line graphs as graph types were a logical choice. (see also: the graph variants in the experiments)

Graph Types

A particular graph type can appear in many different forms and the question remains as to what extent all these variants fulfill the basic definition of the graph type concerned. When exactly can we speak of a bar graph? Can one still refer to a graph as a simple line graph when the values of the individual levels of X are specifically indicated next to each point on the line, or should this kind of graph be called a "point-line graph"? The graphs examined in these experiments have been kept simple and should correspond to a large extent to the basic form of each graph type. In addition, they have been designed to allow several tasks to be solved as well as possible. Numbering the levels of X is, for instance, vital if particular levels are to be compared. We endeavored to ensure the comparability of the graph variants because "unless the design of all displays is optimal conclusions drawn from the comparisons have to be regarded as tentative only" (Meyer Shinar & Leiser 1996).

Examples of the graphs used for each task are always shown (usually the originals). These examples all fulfill the particular experimental requirements and represent, strictly speaking, possible examples because the graphs shown in the experiments were generated according to certain rules, including random processes, in order to be able to test as many different graphs as possible.

Graph Display

The basic types of graph display (c.f. Bertin 1974, p. 109) are as follows:

- A) Several data sets are displayed in one graph (superposition; multiple line diagram, Schutz 1961b) or
- B) Each data set is displayed in a separate graph (juxtaposition; multiple graph diagram, Schutz 1961b).

B allows several display variations. Occasionally juxtaposition was tested using two special forms of arrangement: horizontal and vertical arrangement. These two types of arrangement were then compared. However, this distinction was not made for most tasks for pragmatic reasons and usually "normal arrangement", as seen below, was chosen:

Different Forms of Arrangement for Juxtaposition

(The figures in the columns showing the different forms of arrangement correspond to the number and position of each graph (e.g. 1 2 = first graph on the left, second graph on the right))

Number of data sets	normal arrangement	horizontal arrangement	vertical arrangement
2	1 2	1 2	1 2
4	1 2 3 4	1 2 3 4	1 2 3 4
8	1 2 3 4 5 6 7 8	-	-

To ensure that the data sets displayed in superposition could be easily viewed and distinguished from each other, a visual variable enabling the best possible viewing was to be used. We thus chose the medium of color, which is seen by many in the field as one of the best ways of coding several data sets (c.f. Bertin 1974, p. 99, Schutz 1961b, Casali and Gaylin 1988, p. 41, Travis 1991, p. 122). The software and hardware at our disposal not allowing us to choose colors freely, we did not endeavor to select the optimal colors from a physiological point of view. Instead, certain aspects were simplified and standardized (e.g. the first data set was always blue, the second always red, c.f. the many examples).

One particular problem, which cannot be discussed further at this point, is enabling a fair comparison between superposition and juxtaposition. In our view it was absolutely vital to maintain the same relationship between the ordinate and the abscissa for all the graphs.

Experimental design

All experiments are based on the principle of repeated measurement, in which each subject works through all the experimental conditions. The order of trials within each task was always determined at random, but the tasks themselves always appeared in the same order.

The construction principle of the experimental conditions is highly complex and can only be outlined here:

- In order to increase the reliability of the experiment, the subjects were required to view several graphs for each particular experimental condition.
- To ensure that the subjects' memory did not interfere with the results, they did not receive exactly the same data in the graphs used to compare the experimental conditions.

- In order to increase the external validity of the independent variables, the individual subjects had to work with different requirements and different data in the graphs.

This approach enabled many different groups of data to be tested in the experiment, forming a sound basis on which to make generalizations based on the findings. At the same time, the internal validity of the experimental design was guaranteed by the repeated measurements and the series of randomization processes that were carried out. In contrast, we accepted that our chosen procedure meant doing without a reduction of the variance of errors and would lead to a loss of statistical efficacy of the experimental design.

General Procedure

Before the actual test phase began, the experimental subjects were given sufficient instructions as to the particular type of task to be completed and were shown the corresponding graphs (c.f. the sample instructions). Then the subjects were required to complete at least two examples, although they could try out as many examples as they liked and were always instructed not to begin the actual experiment until they were quite sure what was required of them.

The subjects first saw a specific question appear on the screen, e.g.: "What kind of trend does the red data set display?" The next instruction told them to answer the question **as fast as possible, while ensuring that the answer was also correct**. Pressing the space bar made the question disappear and the graphical presentation appear in its place. Pressing the space bar again, which made the graphics disappear, indicated when the subjects had decided on their answer. Then they answered, usually by clicking on the mouse or typing in a short answer.

Overview: Procedure for each trial

Steps	specific (implicit) instruction	remarks
1	Read the question without seeing the graph	Memorize the question, e.g.: "Do two data sets have the same trend?"
2	Press the space bar to make the graph appear	Start timing response time
3	Look at the graph	Try to answer the question while viewing the graph
4	Press the space bar to remove the graph from the screen	Stop timing response time
5	Indicate the answer	Click or write the answer

Please note: the experimental subjects did not have to label or name any of the data sets or graphs. Thus questions such as "Which data set does the blue line represent or what data set does graph 2 show?" were not asked. We ask direct questions about such things as "the blue line" or "graph 1" because visual perception, not cognitive orientation, was our main concern.

---

[1] There are also studies that support the superiority of tables over bar graphs and line graphs in certain tasks (c.f. Lohse 1993 and Meyer, Shinar & Leiser 1996, submitted for publication). Whether tables deliver better results than graphs depends on a number of factors, e.g. type of task, the complexity of the task, the number of data elements and data sets or how difficult it is to assign labels to the data. (See also: Coll (1992), Coll, Coll & Thakur (1994))

Experiment 1

The Influence of Graph Type and Graph Display on the Perception and Comprehension of Trends

Abstract

If several data sets are to be illustrated, it is possible to present each data set in a separate diagram ***(juxtaposition)*** or to present all data sets in one diagram ***(superposition)*** . Both ***bar graphs*** and ***line graphs*** were logical choices as graph types. According to the hypothesis that differences between the graph variants would become more evident the more difficult the tasks became, the number of data sets was varied. The aim of this investigation was to find an empirical answer to the question of what influence the above three factors have on the recognition of trends. The main focus was the experimental subject's ability to detect changes in the slope of a curve. A complex computer experiment was developed for this purpose. Twenty experimental subjects took part.

It has already been shown that bar graphs and line graphs are equally suitable for representing the trend of a single data set. This may be the reason why no differences between the two graph types occur when they are displayed in juxtaposition, irrespective of how many data sets are represented. Only when superposition is used does the advantage of line graphs over bar graphs become more and more evident as the number of data sets represented grows. However, superposition as a choice of graph display is overtaxed at the very latest when eight trends are represented. Superline can compete on an equal footing with juxtaposition in the display of up to four data sets, but is more likely to produce errors in the case of eight data sets.

This investigation is the first part of a larger test series centering around the specification of interactions between questions and graph variants.

Introduction to the Topic:

As shown in an earlier experiment, graphical presentations are superior to tabular presentations for many tasks. In a total of ten tasks, the clear advantage of graphical presentations over tables was shown to be the better recognition of trends (see Results from Jacobs 1990). Thus the basic question of "**Should a graph be used?**" posed by Bertin (1974) can, at least for certain tasks, be answered in the affirmative.

Whereas data displayed in a table must be analyzed sequentially and, to a certain extent, assessed cognitively, trends represented by graphical means can, in certain cases, be recognized at one glance. The more favorable visual perception in a graph makes complex relations easier to recognize and enables more demanding questions concerning a data set to be answered.
The amount of data becomes more complex when several data sets are to be displayed in one presentation. In this case graphical presentations are often necessary in order to make it possible for certain relations to be recognized within a reasonable amount of time. It is surprising that hardly any research has been carried out in the areas in which graphical displays are second to none in their ability to display data - in order to determine the optimal graph types or displays. This and the following experiments aim to redress this deficit, limiting the analysis to bar graphs and line graphs, being the most well-known and probably also the most widely used graph types.

Hypotheses

Before the beginning of the experiment, the hypotheses (in a more detailed form than shown here) were handed over to a witness in a sealed envelope.

1. Line graphs and bar graphs are equally suitable for enabling the trend of one data set to be recognized.

2. The graph types displayed in juxtaposition are equally suitable for enabling the identification of trends, irrespective of the number of data sets.

3. Where two or more trends are concerned, superline is superior to superbar (e.g. it requires less time). The difference between line graphs and bar graphs grows with a rising number of trends.

4. We have no clear expectations as regards the performance of the two graph types displayed in juxtaposition and superline. If differences do appear, they will most likely occur in the display of eight data sets, whereby juxtaposition will probably be at an advantage.

The Independent Variables

Faktors (Steps)

- A) **Number of Trends** (1, 2, 4, 8 Trends)
- B) **Graph Display** (Superposition, Juxtaposition)
- C) **Graph Type** (Bar Graph, Line Graph)

All factors are repeated measurement factors.

The Dependent Variables

- a) Time taken to answer the question (Time)
- b) Percentage of right answers (Accuracy)

One very important dependent variable is the time taken for the subject to reach a decision once the graphical presentation has appeared. The person was asked to answer the question as quickly, but also as accurately as possible. The whole course of the experiment was controlled by the experimental subject. He or she read the question, started the graphical presentation by pressing a key and ended it again in the same manner. Then he or she answered the question.

Construction of the Experimental Conditions

The graphs for the individual experimental conditions were made parallel according to certain rules. Each subject worked with different data. The data in the graphs were generated on the basis of certain functions and variable parts of these functions were varied at random within a particular range. Thereby, it was possible to test a wide variety of data constellations and to improve the external validity considerably.

You can download the exact experimental procedure, the explication of the experiment, a simulation of possible trends, examples of graphs for each experimental condition and two experiments from Graph Experiment 1.(**Instructions regarding Data Transfer)**

Experimental subjects

Twenty people served as experimental subjects. Sixteen of them were students of education. They had done a course taken by the author in the winter semester 1993/94 "An Introduction to Planning Experiments". The remaining four subjects were students employed by the university as research or technical assistants.

Course of Experiment

The students completed the experiment in the computer laboratory of the Arts Faculty in groups of three to seven (at individual computers). The assistants completed it alone in their studies.

The whole course of the experiment was controlled by the computer program. The experimental subjects had to solve sample tasks for each experimental condition and were told only to start the presentation once they had understood the question completely. The experiment took between 70 and 100 minutes to complete.

Questions asked in the Experiment

Identifying a trend was limited to identifying the sequence of signs (positive or negative) of the trend's gradient. So the experimental subject was asked to identify the sequence of signs of a particular trend.

Two questions were asked in order to analyze the perception of trends in two different ways:

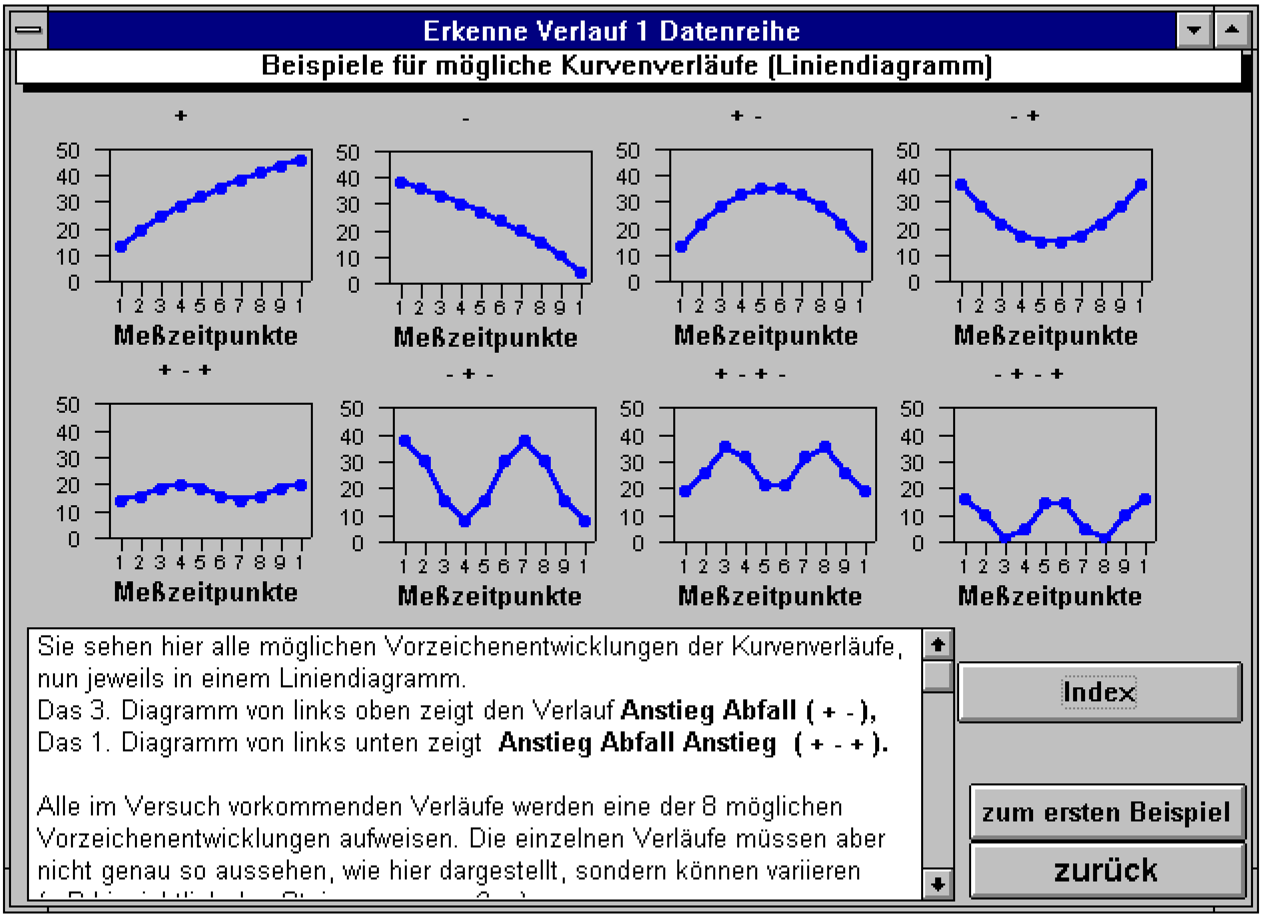
1. **Identify the Trend**
   A data set was shown and the person was asked to identify the sequence of signs.
2. **Determine Which Data Set..**
   A particular sequence of signs (i.e. +, -) was shown and the person was asked to determine which of the data sets displayed the corresponding sequence.

It was only possible to test the conventional line graph and bar graph using the question "Identify the trend" because at least two trends have to be presented for the question "Determine which data set..."

Identify the Sign or the Sequence of Signs of the Gradient Displayed by One Data Set

A trend in the form of a line graph or a bar graph was shown to the experimental subjects and they were asked to identify its sign or its sequence of signs:

Example shown before the Experiment



Only one trend was shown in the experiment at a time. These trends were based on particular functions: each function was varied at random within a certain range as regards its height and gradient. Thereby it was possible to test a wide variety of trends. For example, the sine function of a trend with a sequence of four signs and a high amplitude can fill up the whole graph, but also cause a barely noticeable oscillation at the upper edge of a graph. This experimental design meant that the subjects were shown different forms of one trend

Results

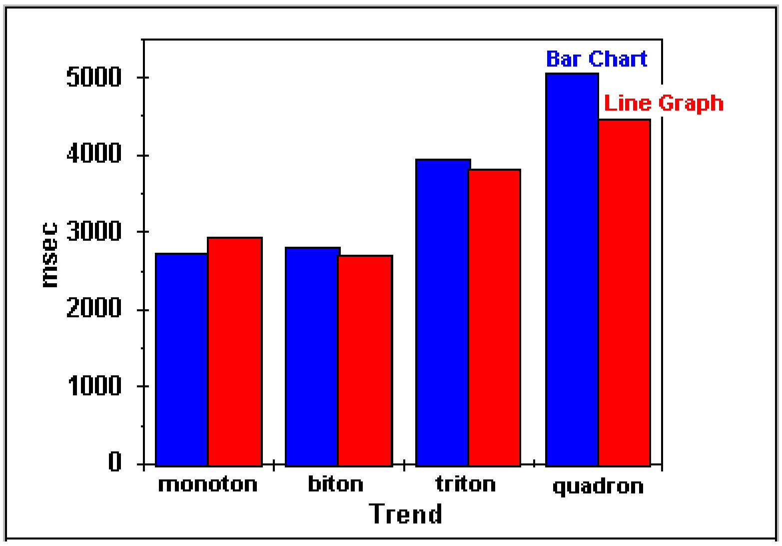
Percentage of Errors

```
      bar chart   3%
                       (z = .9129, 2-Tailed, p =.36, p>.05)
      line graph 2%
```

Time (Arithmetic mean in sec)needed to Identify the Sequence of Signs of a Data Set (mean from all trend types)

Variable	Corr	Mean	SD	t	df	Sign. (two tailed)	
bar chart		3.6	2.1				
	.92			.81	19	.428	p>.01
line graph		3.5	2.2				

Time to Identify the Sequence of Signs of a Data Set (for the particular trends)



trend		sequence of signs		
monoton:	mean of	+	and	-
biton:	mean of	+ -	and	- +
triton:	mean of	+ - +	and	- + -
quadron:	mean of	+ - + -	and	- + - +

Important Results

1. Line graphs and bar graphs do not differ significantly. This holds for both time and accuracy.
2. The interaction between the kind of trend and the graph type is not significant.
3. The results of Jacobs (1990) have thus been clearly confirmed.

In a Presentation of Several Data Sets, Identify the Sign or the Sequence of Signs of the Gradient Displayed by One Particular Data Set.

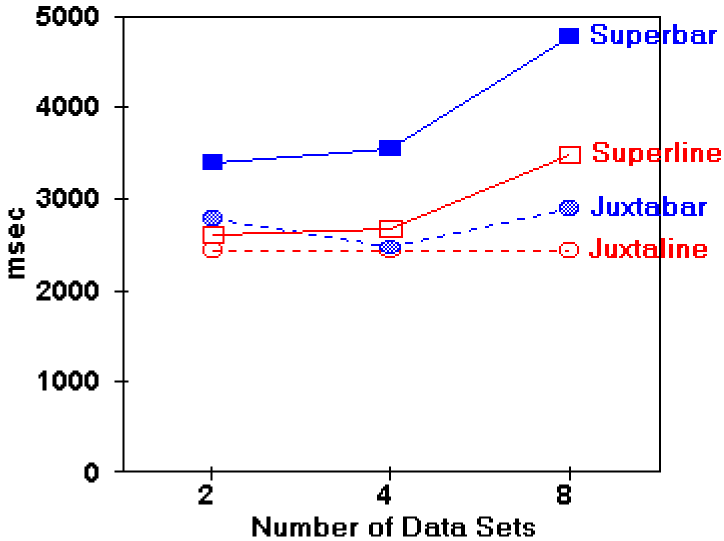
Two, four and eight data sets were shown to the experimental subjects. Before they could see the diagrams, they were instructed to memorize a particular data set because they would be asked to identify its trend in a later presentation. **What kind of trend does a particular data set display?** The data set whose trend is to be identified will be called the target data set.

With superposition as the choice of graph display, the target data set was specified using its color (i.e. **What kind of trend does the red data set display?**) With juxtaposition, the diagram number specified the target data set (i.e. **What kind of trend is displayed by the data set in diagram no. 2?**)

Completing this task depended on the experimental subjects' ability to isolate the target data set from the presentation as a whole and not to let themselves be distracted by irrelevant trends.

Results

Time Needed to Identify the Sequence of Signs of a Given Data Set
(Identify the trend of a particular data set)




```
Examples of Tasks:   What kind of trend does the red data set display?
                     What kind of trend does the data set in diagram no.
4 display?
```

- Juxtaline = Line graph in Juxtaposition
- Juxtabar = Bar chart in Juxtaposition
- Superline = Line Graph in Superposition (as shown in the diagram above)
- Superbar = Bar Graph in Superposition

Important Results:

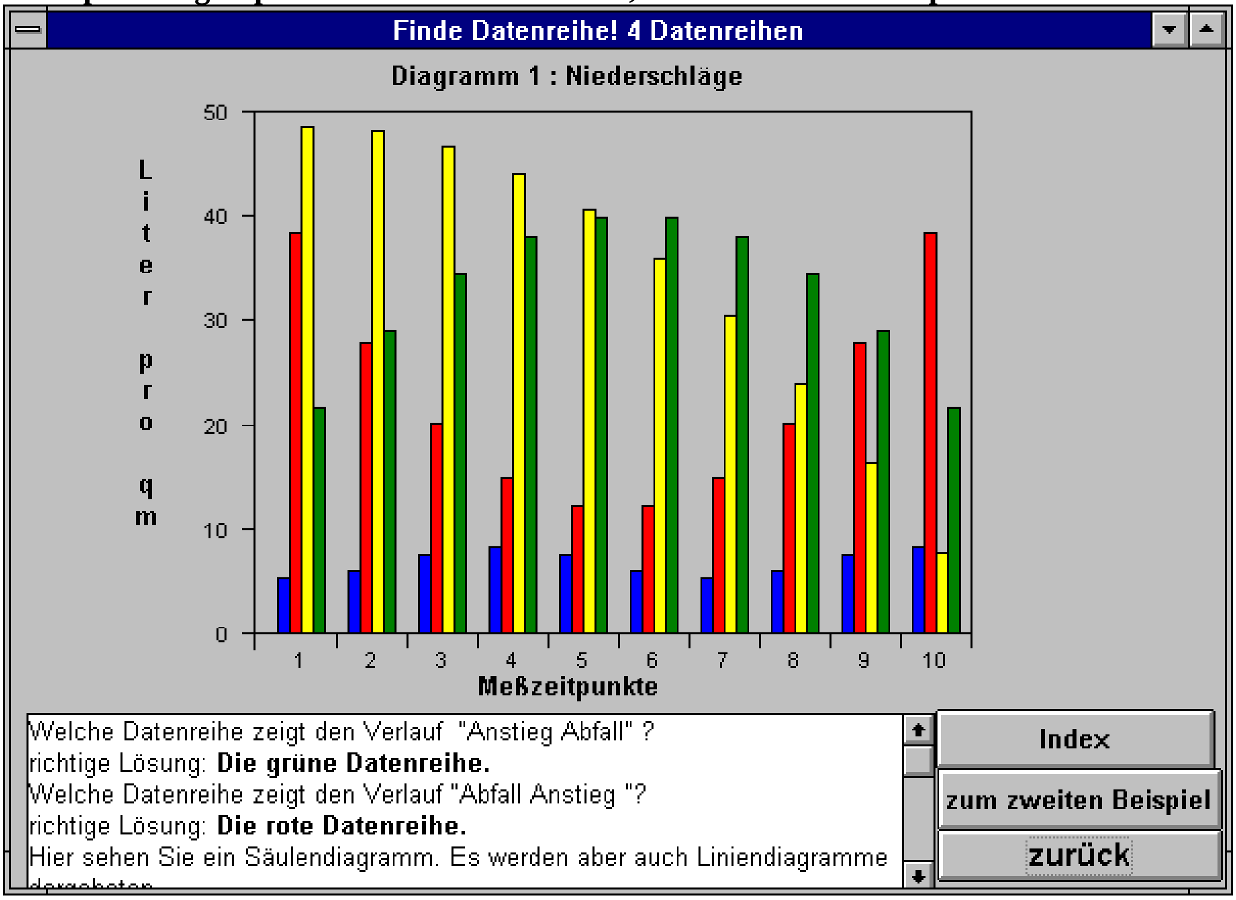
- Superbar requires significantly more time for each step of the factor "number of data sets" than the remaining graph variants.
- In contrast to superline, juxtaline delivers significantly shorter times and significantly higher accuracy with eight data sets.

Determine which of the Data Sets Has a Trend whose Gradient Displays a Particular Sequence of Signs

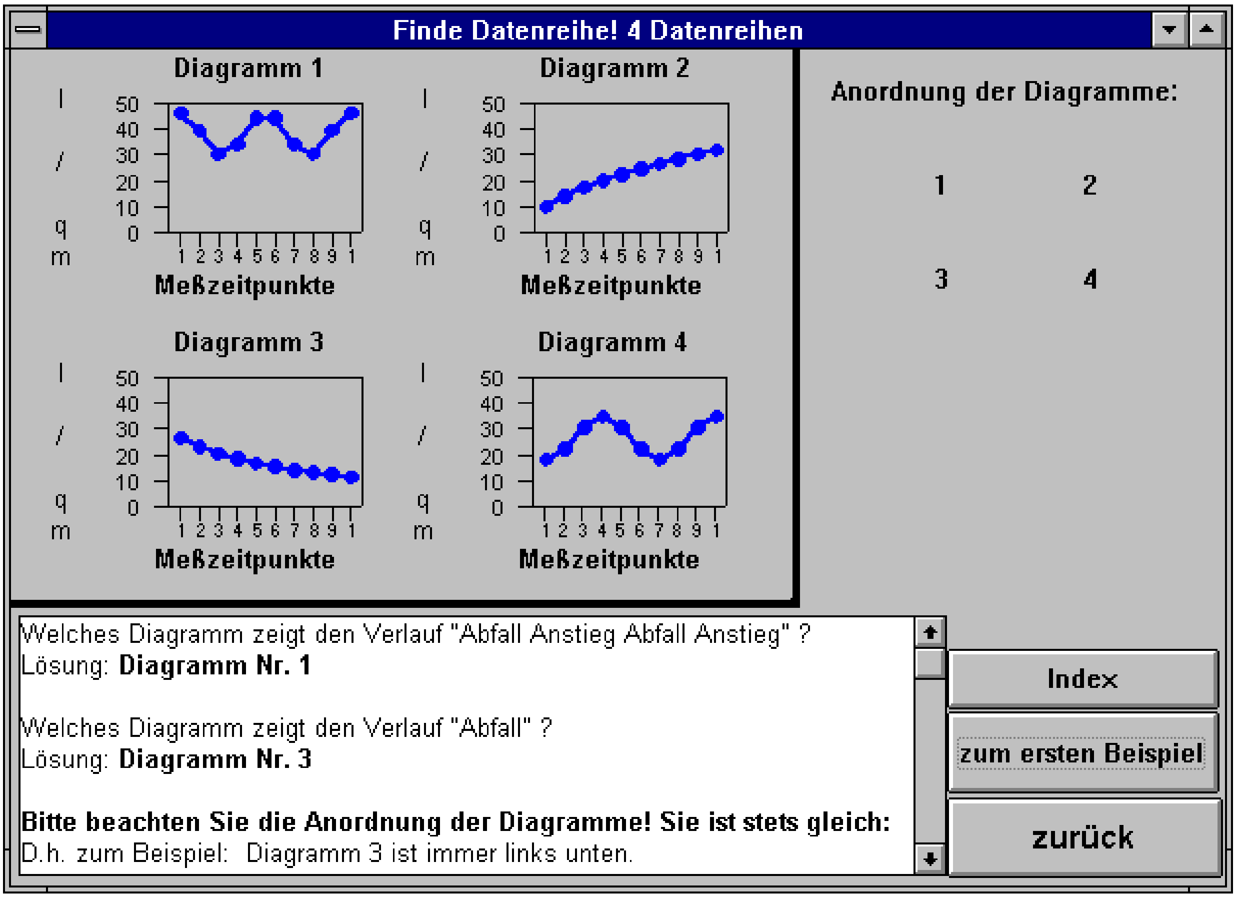
The graphical presentation contained two, four or eight data sets. A particular trend (= target trend) was described to the experimental subject before the graphical presentation began (**Which data set shows the trend "rise, fall" ?**) . Immediately after the presentation the data set showing the target trend was to be identified by that person. The experimental subject was asked to identify the correct color for data sets displayed in superposition and the correct diagram number for those shown in juxtaposition.

Completing this task depended on the experimental subject's ability to locate the matching trend from amongst several possible trends.

Example Using Superbar with Four Data Sets, Shown before the Experiment

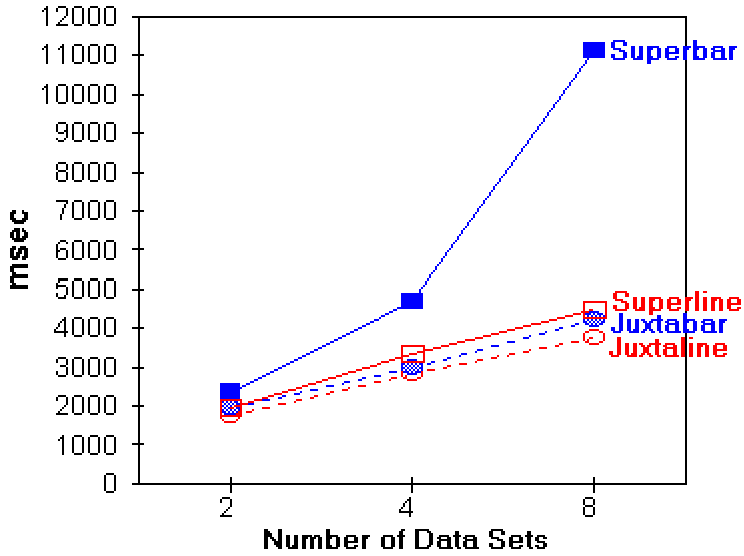


Example Using Juxtaline with Four Data Sets, Shown before the experiment



Results:

Time needed to identify the Data Set with a Particular Trend




```
Examples of Tasks: Which data set shows the trend "rise, fall"?
                   Which diagram shows the trend "fall, rise, fall?"
```

Important Results:

- Superbar takes significantly more time for every step of the factor "number of data sets" than the remaining graph variants, which in turn do not differ significantly from each other in terms of the time required under every experimental condition.
- With eight data sets juxtaposition produces significantly fewer errors than superposition. Juxtaline delivers a significantly higher percentage of correct answers than does superline and the same holds for juxtabar in respect to superbar.

Conclusions

1. Line graphs and bar graphs are equally suitable for representing the trend of one data set.
2. Where two or more trends are concerned, superbar is inferior to the other graph variants in terms of the time needed to answer the question. For the display of up to four data sets, superline is just as suitable as the two graph types displayed in juxtaposition. Superline can also be expected to show clear advantages in other questions.
3. For the display of eight or more data sets, juxtaposition is to be recommended. It is of little importance which graph type is displayed in juxtaposition. The second type of graph display, superposition, is overtaxed by eight trends as it leads to poor accuracy. This also holds for line graphs displayed in superposition.

Experiment 2

Experimental Analyses of How Trends in Curves are Perceived and Compared in Line Graphs and Bar Graphs Displayed in Superposition and Juxtaposition.

Abstract

- Introduction to the Topic
- Variables, Experimental Subjects and Course of Experiment

Particular Tasks:

- What type of trend does a particular data set follow?
- Locate a data set with a particular trend from among several sets.
- Does a particular data set show any kind of deviation from an ideal trend?
- Which data set displays some kind of deviation from its ideal trend?
- Do several data sets show the same type of trend?

Abstract

The continuation of Experiment 1 consisted of the examination of the influences of **graph type** (line graph, bar graph), **graph display** (superposition, juxtaposition) and **the number of data sets** (=trends) on the speed and accuracy of experimental subjects when answering particular questions. In this case we were mainly interested in questions concerning all the data of a particular data set. Such questions require the comprehensive analysis of the relations involved. They deal with the type of trend displayed by a data set, deviations from an expected trend and similarities or dissimilarities of the trends of several data sets. The advantages of a particular graphical presentation should mainly be easier visual perception.

A complex computer program was developed in order to test the questions. It included a data generator to generate a wide range of artificial data. Twenty-five students served as experimental subjects.

The results offer convincing statistical material suggesting that line graphs are at an advantage as far as these kinds of tasks are concerned.

Introduction to the Topic

Graphical presentations of data prove to be especially fruitful when a tabular depiction does not show the relations between the data at all or only in a manner that requires a lot of effort on the part of the viewer. This is particularly the case when several data series have to be analyzed at once. If this is so, the relevant question is not whether a graph is helpful, but what kind of graph and what kind of display are most suitable for each particular task.

The main advantages of a particular graphical presentation should be better visual perception. **Essentially, the focus of our research is the problem of how best to represent the relations between numbers as spatial relations so that anyone who**

knows what relations he/she has to concentrate on will be able to recognize them quickly and accurately. Another important aspect is to review the question of how trends can be recognized particularly easily when their characteristics as a whole have to be assessed. This investigation takes up Experiment 1, modifies a central question and adds several new questions to the analysis. Again the graph types bar graph and line graph in superposition and juxtaposition are analyzed.

The Independent Variables

Factors (steps)

- A) Number of trends (1, 2, 4, 8 data sets)
- B) Graph display (superposition, juxtaposition)
- C) Graph type (bar graph, line graph)

All the factors are repeated measurement factors.

The Dependent Variables

- a) Time needed to answer the question (time)
- b) Percentage of right answers (accuracy)

The Data in the Diagrams (Graphs)

The data in the data sets were generated on the basis of certain functions. Variable parts of the functions were varied by chance within a certain range. A positive linear function, for example, always produces a straight line with a positive gradient. In this experiment the gradient and the intercept of the y-axis were subject to certain random processes, which produced steep and flat trends at different positions in the graph.
The diagrams for the individual experimental conditions were not identical. They were made parallel according to certain rules. Each subject worked on different data. Thereby, it was possible to test a wide variety of data combinations and to improve the external validity considerably.

Experimental Subjects

Twenty-five students took part in this investigation. Half of them were students (mostly mathematical/scientific) from a pedagogical seminar taught by author, the rest were students of educational science taking part in a statistics seminar.

Course of Experiment

The students completed the experiment (at one computer each) in groups of 3 to 8 in the computer laboratory of the Arts Faculty. The whole course of the experiment was run by the computer program. Before the students were allowed to answer a question, they were instructed to solve sample tasks concerning each of the experimental conditions to make sure that they had understood the question. Several graphs were tested for each experimental condition. At the end all the results were grouped together and analyzed.

The whole course of the experiment was controlled by the student. He/she read each question, started and stopped the graphical presentation using the keyboard. Afterwards, he/she answered by clicking on the mouse.

Identifying a Certain Trend

Experiment Toolbook 1.5 _ Please send me the experiment * * * * _You will need Toolbook 1.5 Runtime._

In Experiment 1, as well as in an earlier experiment Jacobs (1990), recognizing a graph's trend was simplified and expressed as the sign of gradient. So a u-shaped trend had to be identified as a pattern of "-,+". In both experiments it was agreed that line graphs and bar graphs achieve comparable results with regard to accuracy and the time needed. In contrast, Schutz (1961a), using a method for recognizing trends that cannot be described here, delivered proof of significant advantages of line graphs over bar graphs.

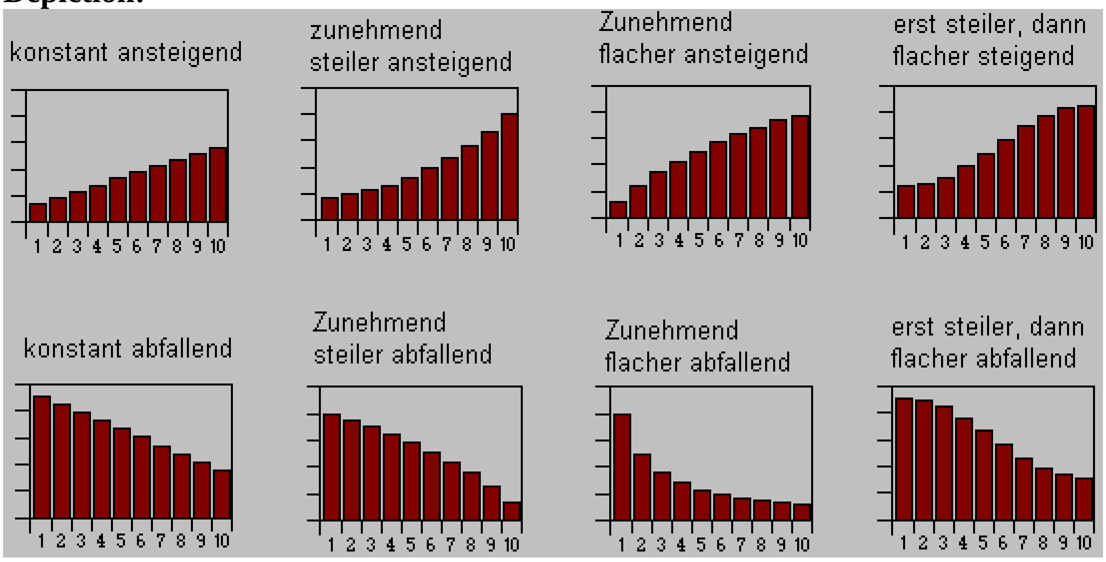
The task in Experiment 1 may promote sequential searching and confirmation and prevent the graph from being viewed as a whole. To test a more integral perception of trends another method was selected. The main task of the subjects was to identify the type of function shown in a graphical display.

Intuitively, the line graph seemed to be more favorable because a line represents a trend as a whole more concisely than a bar, whose relationship to the rest of the trend may first have to be established mentally.

Experimental Procedure

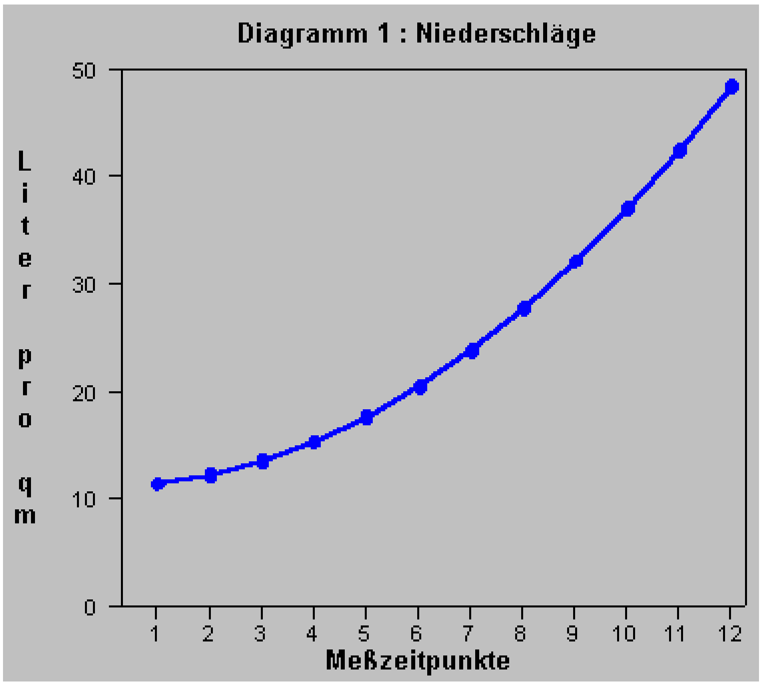
First, the students were presented with the trend types shown, both in the form of line graphs and bar graphs. Each trend type was described in words above the graphs so that the subjects had two ways by which to remember each type.

Possible Types of Trends in Prototypical Depiction:



After this stage of instruction the actual experiment began. A graphical display was presented to the experimental subject and the task was to identify its matching function. The graphs in the form of a line graph or a bar graph corresponded to one of the trend types listed above, but they varied randomly in respect of particular features (gradient, intercept of the y-axis).

Example
Decide as quickly, but as accurately as possible, which type of trend is displayed in the graph



`After decision the student had to click on the corresponding function type above`
`correct answer: zunehmend steiler ansteigend`

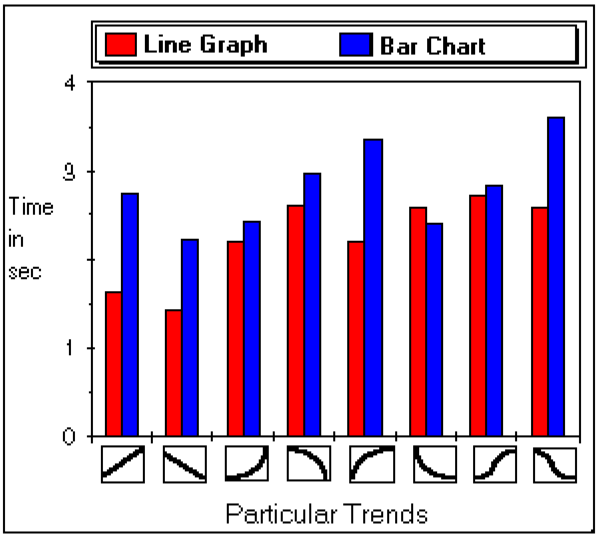
Results for Identifying a Certain Trend

```
Error as a Percentage (Mean value of all trends)
      Bar chart     10%
                          (z= 2.19, 2-Tailed, p = .0284, p<.05)
      Line graph     3%
```

```
time (Arithmetic mean in sec taken for the function type of a trend
      to be recognized (Mean value of all trends))

     Bar chart      2.8
                           (F(1,24)= 21.72,  p<.01.)
     Line graph     2.2
```

Results for Identifying the Particular Trends



Summary of the Most Important Results:

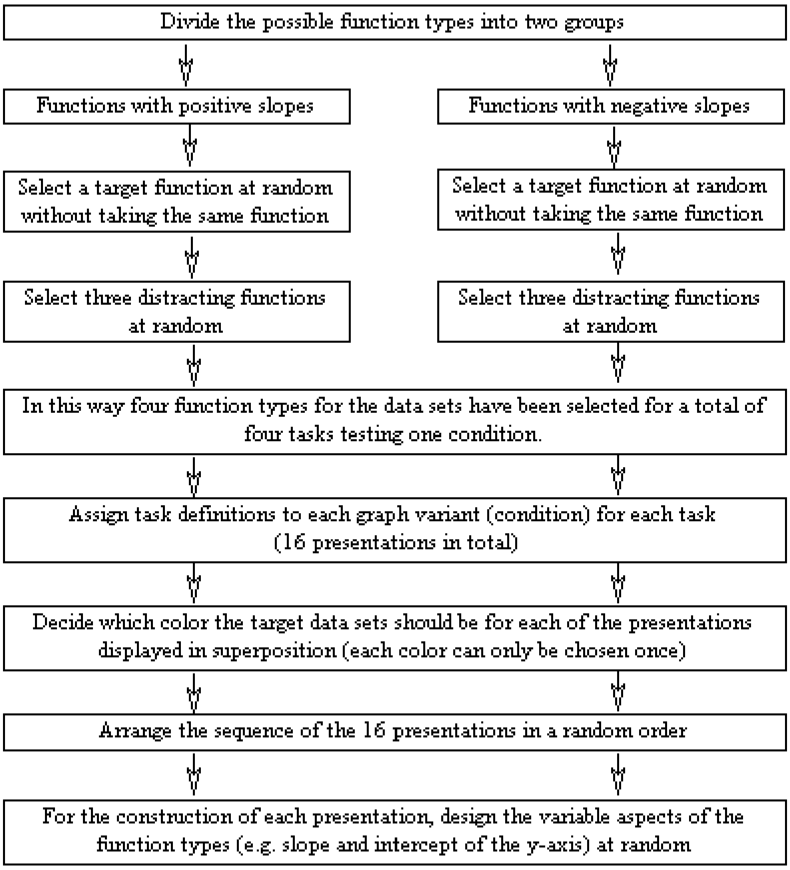
- Line graphs deliver significantly more accurate results than bar graphs.
- Line graphs produce significantly faster decisions than bar graphs. (see diagram above).
- The interaction between graph types and trend types is insignificant.

Locate a Data Set with a Particular Trend

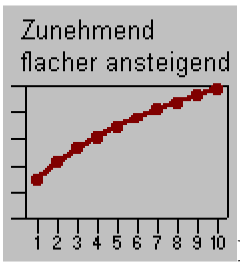
2, 4 or 8 data sets (=data serie = trends) were presented in one display. A prototype of a particular trend was shown to the students before they saw the actual display. Then they were instructed to identify the data set whose trend was an example of the prototype. Superbar, superline, juxtabar and juxtaline served as graph variants. Similar results were expected to those in Experiment 1 with the difference that juxtaline was expected to be superior to juxtabar because line graphs had proved to be more favorable for identifying trends.

What follows is an example of how the presentations were constructed for four data sets:

Construction Principles of the Experimental Conditions for each Subject:

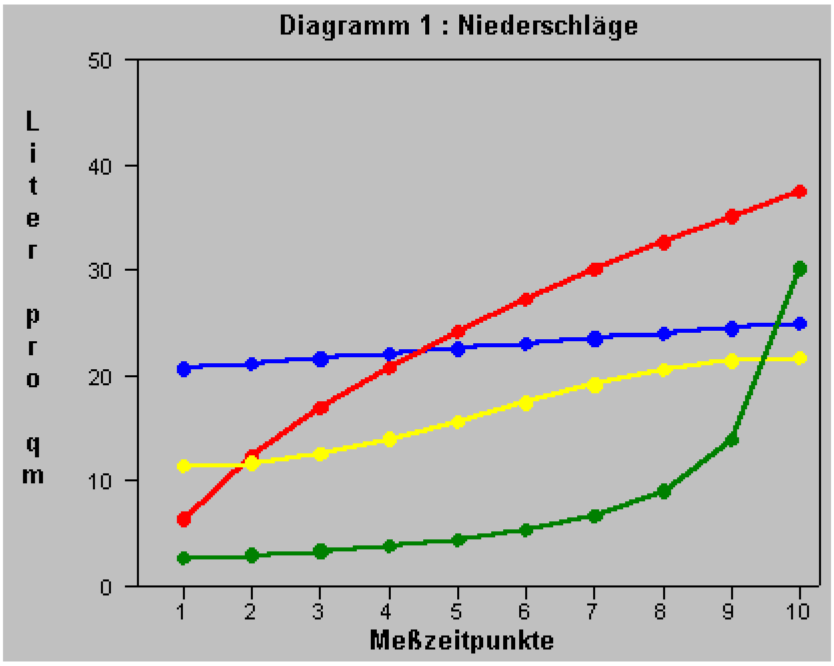


Example



Identify the Data Set with the this trend

You will see a graphical display. When the graph appears, identify the data set showing the trend above as quickly and accurately as possible.



Now the student had to identify the color of the trend or, in the case of juxtaposition, the number of the diagram.
Correct answer: the red line

Results

Errors

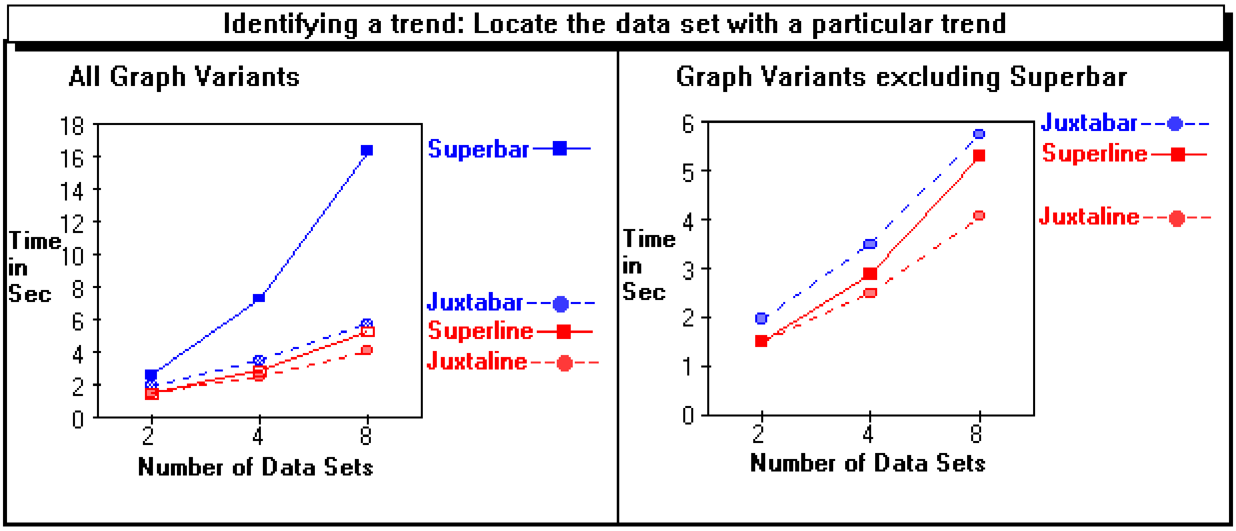
When up to four data sets were tested, very few errors were made on the whole (1-11%) and no significant differences between the graph variants were found. With eight data sets, juxtaline (4%) and superline (8%) produced significantly fewer errors (p<.05) than superbar (27%), but no significant differences were found between them and juxtabar (11%).

Time

Arithmetic mean and *(standard deviation)* **Time** (in sec)
to answer: **Locate a data set with a particular function**

number of Data Sets	**Graph Variants** Juxtaline	Superline	Juxtabar	Superbar
2	**1.50** *(0.68)*	**1.50** *(0.78)*	**1.97** *(0.89)*	**2.63** *(1.40)*
4	**2.49** *(1.22)*	**2.88** *(1.56)*	**3.49** *(1.84)*	**7.30** *(3.92)*
8	**4.06** *(2.73)*	**5.30** *(3.43)*	**5.74** *(3.61)*	**16.33** *(8.19)*

The mean values in each row that are not underlined differ from each other significantly; p<.05 according to the Tukey-Test.

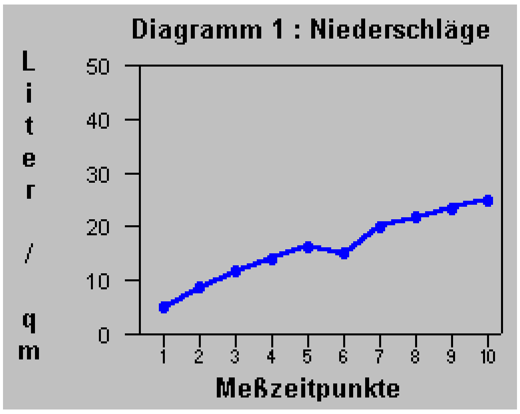


Summary of the Most Important Results:

- Where 2 or more trends are concerned superline is more favorable than superbar.
- In the case of juxtaposition line graphs are always superior to bar graphs. Here the difference between graph types, as established in the experiment on identifying trends, was confirmed several times.
- Superline can compete with juxtaline when up to 4 trends are displayed. Only from 8 trends upwards does the superline graph come off worse, as shown by the statistics.

Recognizing a Deviation from an Ideal Trend

An essential advantage of graphical displays is the immediate perception of particular phenomena as striking features. In this case, an obvious disturbance within an otherwise ideal trend was seen as a striking feature.



In the graph above the trend could well be expected to become increasingly flat, but this appears to be disturbed at level 6. The students were asked to recognize any deviations, but not to locate them. They were not told what kind of function the diagram displayed. So they had to recognize the deviation spontaneously.

Construction of the Experimental Conditions:

The trend prototypes (c.f. Identifying a Particular Trend) functioned as trend types. However, instead of the s-shaped trends, a u-shaped or n-shaped function type was used because it appeared to be difficult to define clear deviations in the case of s-shaped trends. Of course, in the experiment one particular function type could vary in appearance as a result of the randomization of its variable features. The individual tasks (=presentations) were designed in the same way both for line graphs and bar graphs as far as most aspects were concerned (such as the function type used), but some aspects were kept variable (e.g. as regards the position at which a disturbance might appear or the direction of the deviation). What follows is an example of the construction principles.

Construction Principles of the Tasks for each Subject	
Number of tasks for both conditions:	4
Selection of the trends for the 4 tasks:	4 from 8 possible functions selected at random (once selected, the functions were not returned to the pool of functions)
Kind of trends:	2 with and 2 without a deviation from an ideal trend, selected at random
Kind of deviation:	Position, direction and extent of the deviation determined at random while respecting certain rules and limits

Results:

According to the percentage of right answers, line graphs and bar graphs produced comparable results in terms of accuracy. The decisions were made faster with the help of a line graph, as the following table clearly illustrates:

Time in seconds:

Variable	M	s	t	df	Sign.N.	Effect-size
Line graph	1.02	0.45	4.42	23	p<.01	0.68
Bar chart	1.40	0.61				

Which Data Set Shows a Deviation from its Ideal Trend?

2, 4 or 8 data sets (=data series = trends) were presented in one display. As in an earlier task (Recognizing a Deviation from an Ideal Trend), one of the data sets showed some kind of deviation from its ideal trend. The student was told to find it and to identify the trend by its color or the corresponding diagram number. Superbar, superline, juxtabar and juxtaline served as graph variants.

As in all the other experiments, the presentations were designed for each subject according to various randomization processes (e.g. which special functions the data set could take on, which of those functions should display a deviation, and where and in what direction (positive or negative) the deviation should occur).

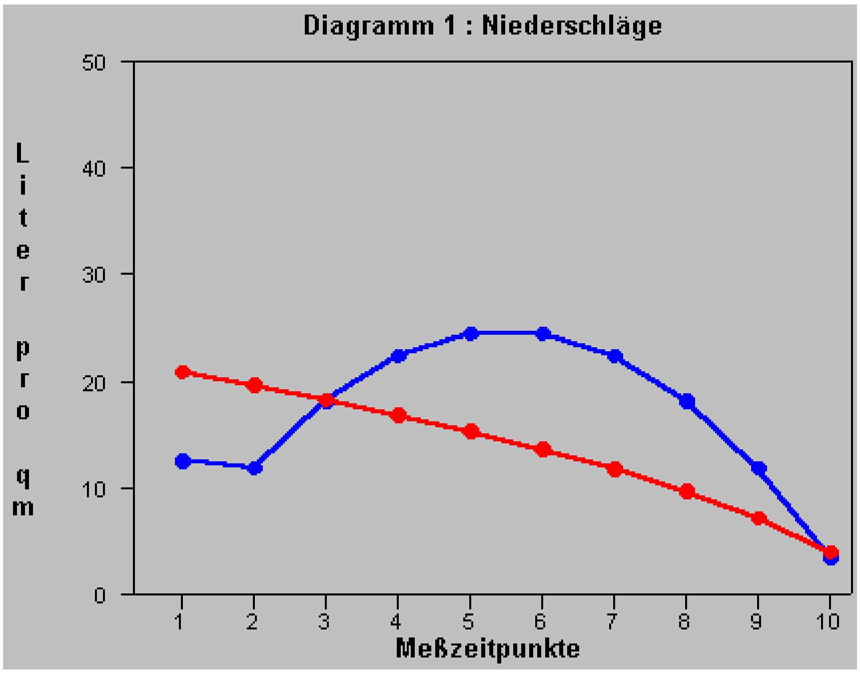
You can download the experiment with four data series.

Experiment
Toolbook 1.5 Please send me the experiment * * * * You will need Toolbook 1.5 Runtime.

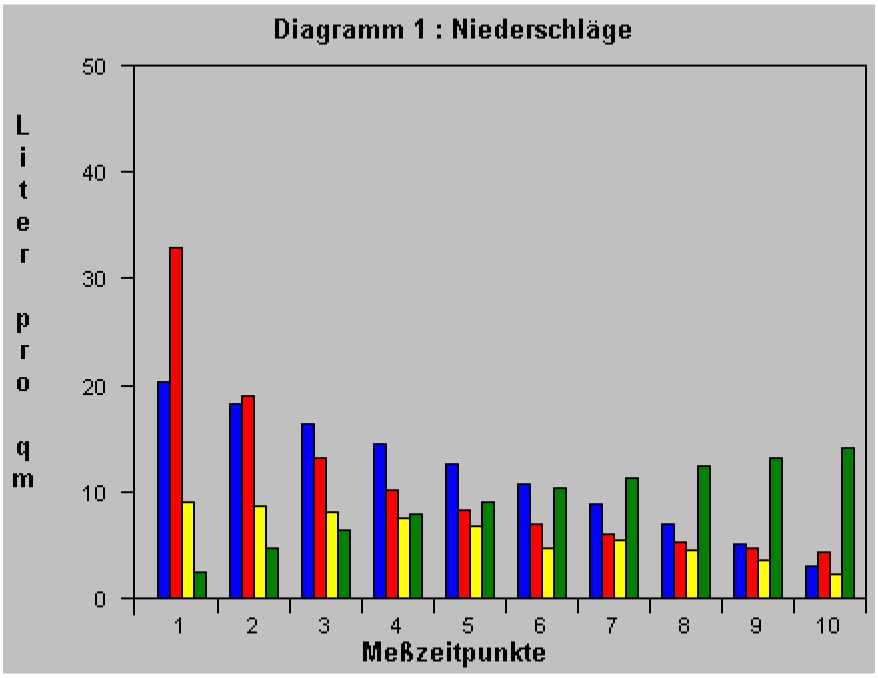
Clear advantages were expected for the line graph and the more trends were displayed, the more this held true.

Examples of tasks

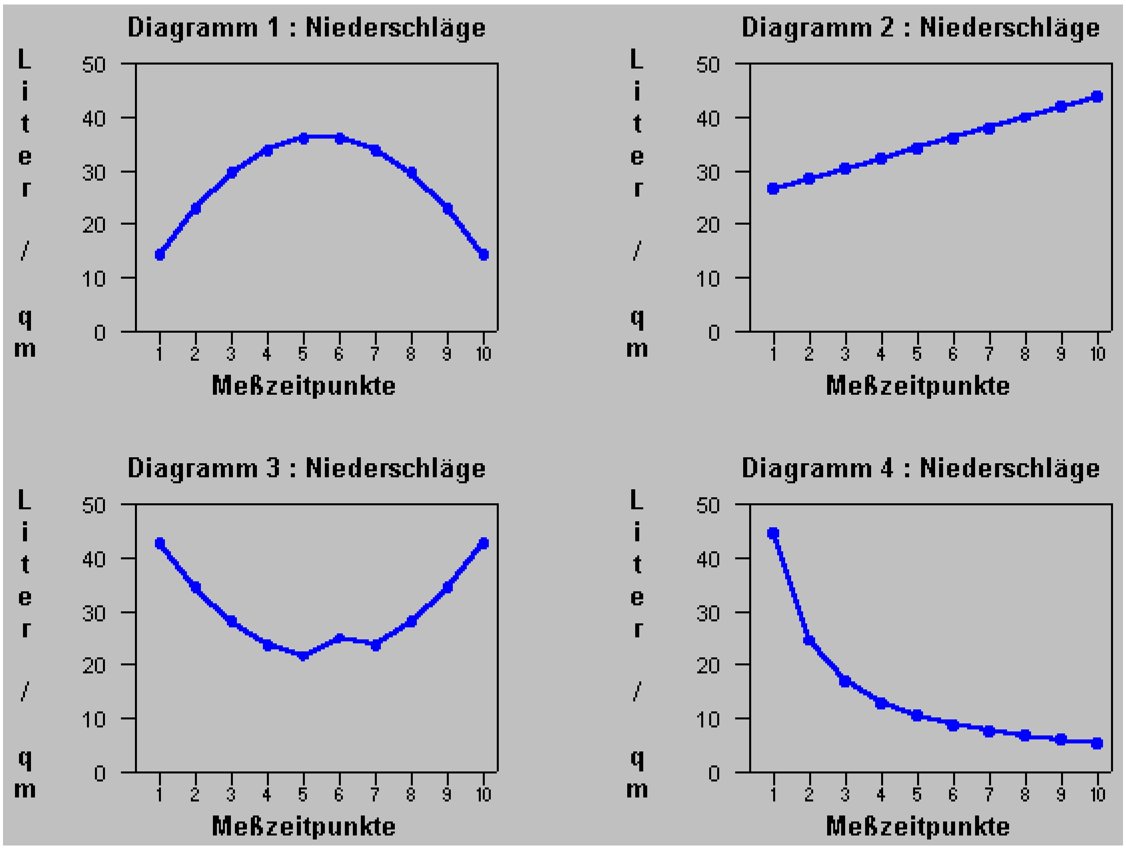
Identify the data set displaying a deviation from its ideal trend as quickly and accurately as possible.



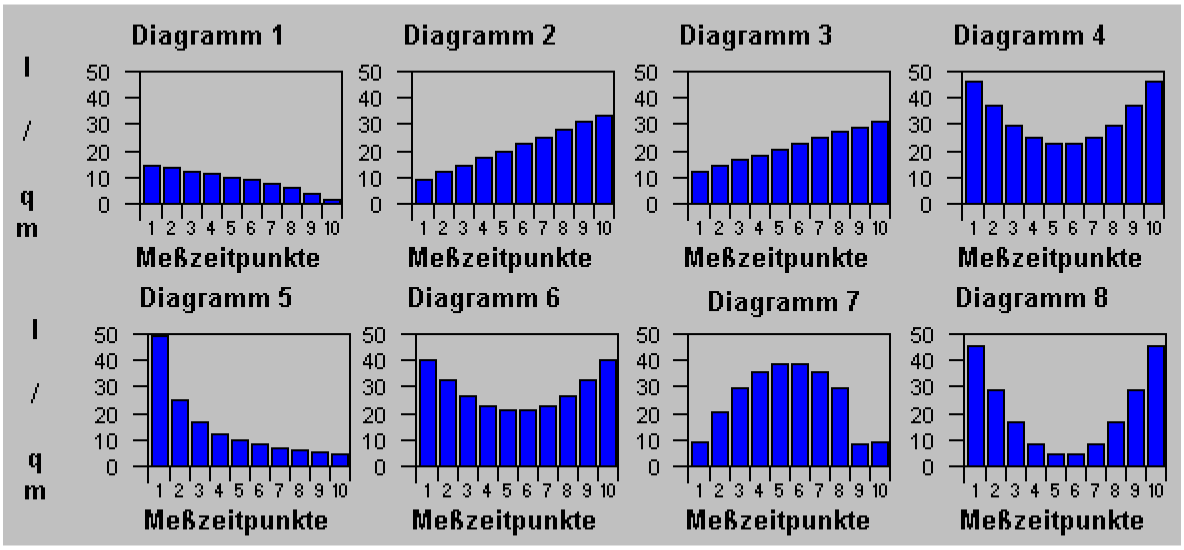
What is the color of the data series displaying a deviation from its ideal trend?



What is the color of the data series displaying a deviation from its ideal trend?
(Yellow is the correct answer)



Which diagram displays a data set containing a deviation from its ideal trend?
(Diagram 3 is the correct answer).



Which Diagram Shows a Deviation from its Ideal Trend? Now the student had to indicate the correct diagram number (here: 7). In the case of superposition it was the color of the trend that the student had to indicate.

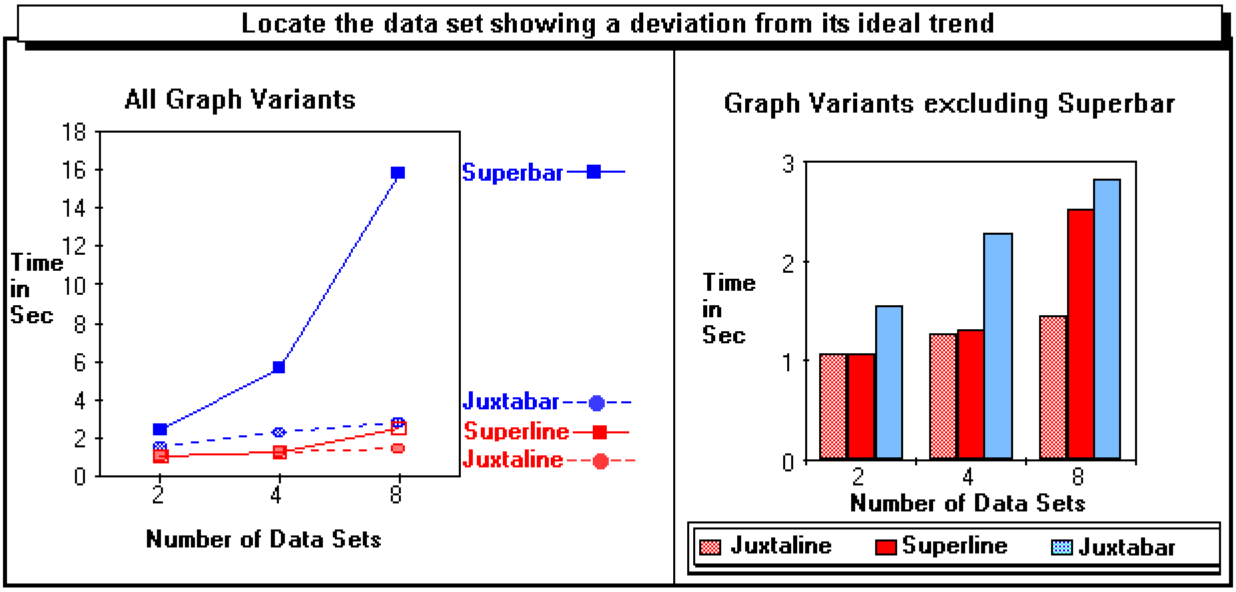
Results for Which Data Set Shows a Deviation from its Ideal Trend?

In total 97.5% of the tasks were answered correctly. No statistical significant differences in accuracy between the graph variants was delivered. This holds true regardless of how many data sets were displayed at once. In contrast, the response times deliver conclusive results.

Arithmetic mean and *(standard deviation)* Response **time** (in sec) : **Locate the data set containing a deviation from its ideal trend**

Number of Data Sets	**Graph Variants**			
	Juxtaline	Superline	Juxtabar	Superbar
2	**1.07** *(0.46)*	**1.07** *(0.49)*	**1.56** *(0.56)*	**2.42** *(0.86)*
4	**1.26** *(0.47)*	**1.31** *(1.53)*	**2.27** *(0.67)*	**5.69** *(1.65)*
8	**1.44** *(0.51)*	**2.53** *(2.17)*	**2.82** *(1.00)*	**15.77** *(6.14)*

The mean values in each row that are not underlined differ from each other significantly; p<.05 according to the Tukey-Test.



Summary of the Most Important Results:

- Where 2 or more trends are concerned superbar is significantly inferior to the other graph variants.
- As expected, juxtaline is constantly superior to juxtabar.
- Superline can compete with juxtaline when up to 4 trends are displayed. Only from 8 trends upwards does superline come off worse. This is proved by the statistics.

Do n Data Sets Show the Same Trend or Not?

The comparison of trends is an important and demanding task for graphical displays and only very few empirical investigations of the subject exist. Casali and Gaylin (1988) found no differences in graph types between line graphs and bar graphs when they tested the comparison of trends. However, in an earlier experiment of my own, it was found that it two trends could be compared significantly faster (ca. 1 second) in a line graph than in a bar graph ($t(39)=3.69$, $p<.001$, effect size=.58 according to Jacobs 1994, p. 82). In the case at hand, a similar problem was examined by another method and an additional comparison was drawn between the graph displays of superposition and juxtaposition. The line graph was expected, especially with growing complexity, to be the more favorable graph type.

Design of the Experiment

2, 4 or 8 data series (=trends) were presented in one display. The following experimental conditions were laid down:

- **2 Data Sets = low Complexity**; task: Do the 2 data-sets show the same kind of function?
- **4 Data Sets = moderate Complexity**; task: Do 2 data-sets show the same kind of function?
- **8 Data Sets = high Complexity**; task: Do 3 data-sets show the same kind of function?

Superbar, superline, juxtabar and juxtaline served as graph variants, but for "high complexity" only line graphs were used. Each experimental condition was tested using four tasks, whereby the questions given above could twice be answered with "yes" and twice with "no". Two trends have equivalent functions if they are based on an identical function type. A sharply increasing straight line and a slightly increasing straight line are both based on a positive linear function because they both have a constantly increasing trend. A slightly increasing straight line and a slightly increasing non-linear function may look more similar, but are based on different functions.

The complexity of the task is increased on the one hand by the variation of the number of data sets in each presentation and on the other hand by the question asked. The levels of complexity vary according to the number of data sets that would have to be taken into account by the subjects in order to solve the task if they were to check the data sets sequentially (e.g. from top to bottom).

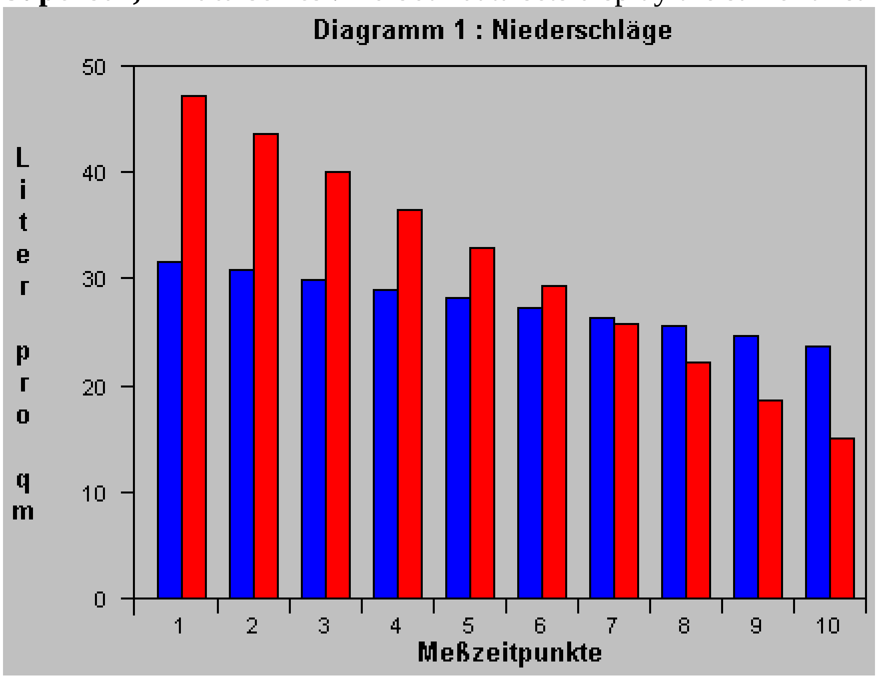
In the first level of complexity (two data sets displayed) two data sets must always be viewed, in the level of moderate complexity (four data sets displayed) the trend of at least two or three data sets must be ascertained. In the last level of complexity (eight data sets) the number of data sets that must be examined in order to solve the task ranges from three to eight. In this case each presentation was designed so that the graph(s) either showed three trends with the same function (if yes is the correct answer we shall refer to the question as a true-question) or four pairs with the same function (each, not all four; if no is the correct answer it will be called a false-question). So if "Are there three data sets displaying the same trend?" is a false-question, it is particularly difficult to answer (which has been proved empirically). For false-questions all eight data sets must be examined, whereas true-questions only require the examination of three to eight data sets (assuming the subject takes a sequential approach).

Like in all the other experiments, it was decided at random which of the possible functions would be tested on each subject.

Examples of tasks

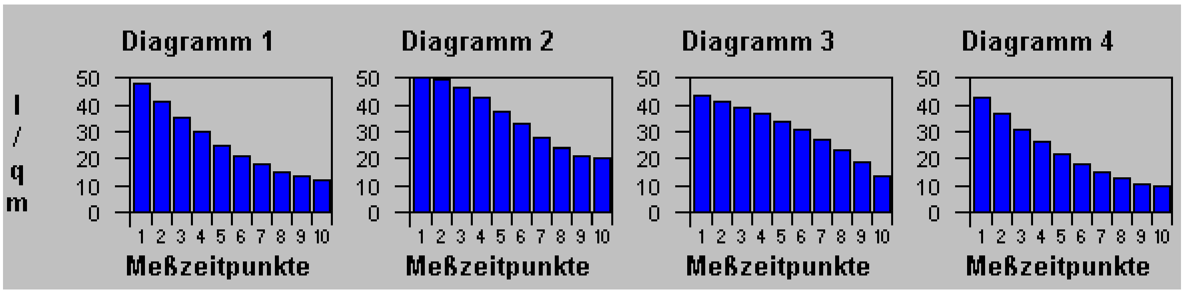
If you click on the following examples, you will see a graphical display. Find out in a quick and accurate manner if the presentation shows n data sets with the same kind of function.

- **Superbar, 2 Data series** : Do both data sets display the same function type?



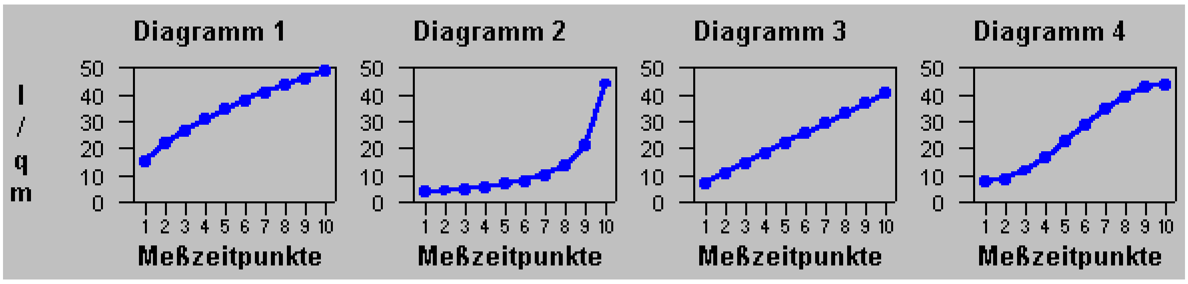
Yes, both data sets display a negative linear function.

- **Juxtabar, 4 Data series** : Do two data sets show the same function type?



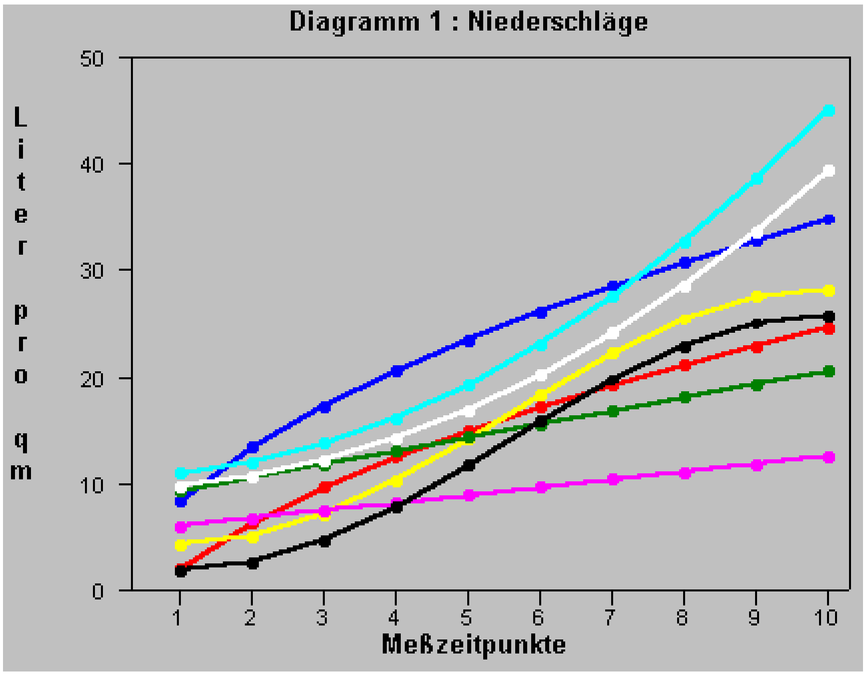
Yes, diagram 1 and diagram 4 show the same kind of function.

- **Juxtaline, 4 Data series** : Do two data sets show the same function type?



No, every diagram shows a different kind of function.

- **Superline, 8 Data series** : Do three data sets show the same function type?



No, there are four pairs of data sets, each with the same function type, but none of the trends occurs three times. As a false-question requiring "no" as an answer, this task has been shown to be the most difficult.

Results for "Do n Data Sets Show the Same Trend or Not?"

Errors

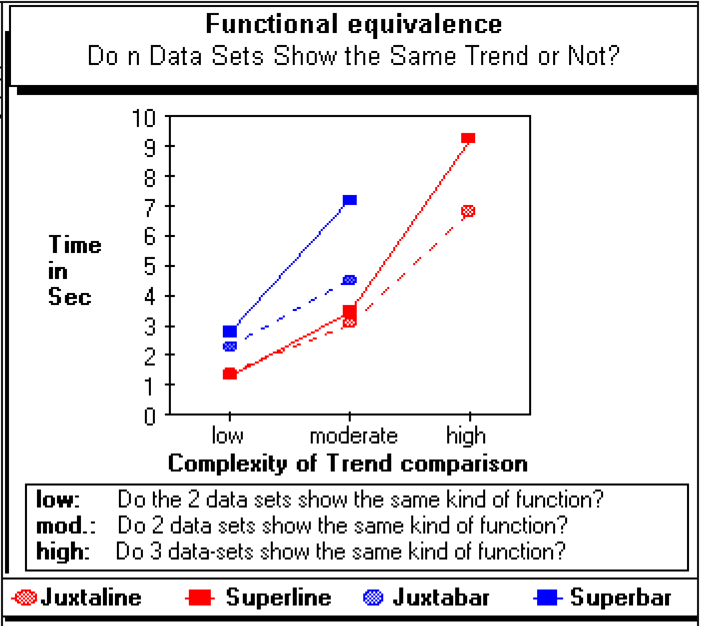
The tasks were solved correctly on the whole. Errors varied between 2% and 15%. For tasks of moderate complexity, (e.g. difficulty as is proven empirically) juxtaline delivered significantly more favorable results for accuracy than juxtabar and superbar.

Time

Arithmetic mean and (*standard deviation*): **Time** in sec: **Do n Data Sets Show the Same Trend or Not?**

complexity, difficulty	**Graph Variant**			
	Juxtaline	Superline	Juxtabar	Superbar
low	**1.41** *(0.60)*	**1.32** *(0.49)*	**2.29** *(0.90)*	**2.77** *(1.12)*
moderate	**3.07** *(0.69)*	**3.44** *(0.99)*	**4.49** *(1.14)*	**7.17** *(1.89)*
high	**6.78** (1.20)	**9.23** *(2.75)*		

The mean values in each row that are not underlined differ from each other significantly; p<.05 according to the Tukey-Test.

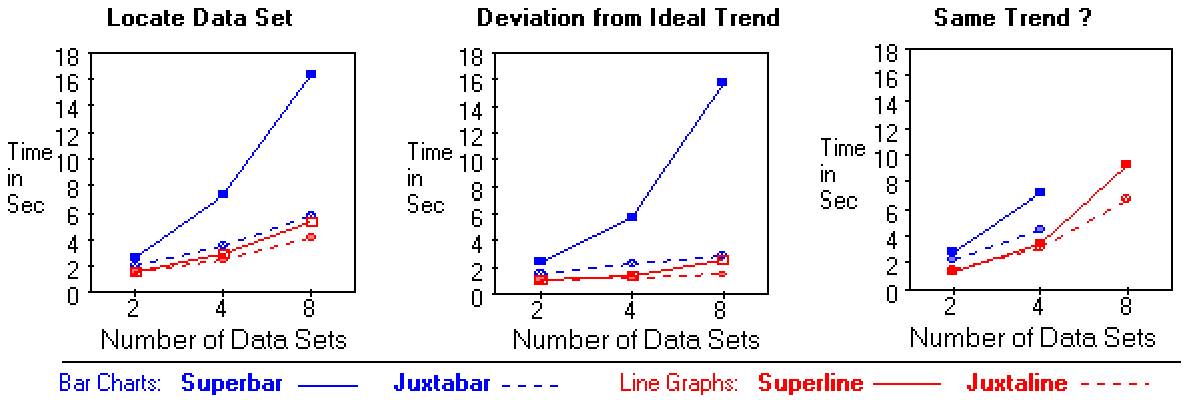


Summary of the Most Important Results:

- Even at the lowest level of complexity, the difference between superline and superbar is more than effect size =1.
- Juxtaline is always superior to juxtabar.
- For simple comparisons of trends in general both varieties of line graph are just as suitable. With rising complexity juxtaline is superior to superline (interaction between the graph variants (superline, juxtaline) and complexity (low, moderate,high) for decision time : $F(2.44)= 10.23$, $p<.01$). At the highest level of complexity the advantages of juxtaline over superline become very clear.

Conclusion

The task results show an amazing consistency:



- **Line graphs clearly proved to be more favorable for the perception of trends.**
  This superiority becomes apparent when:
  - trends have to be assessed in general
  - particular trend characteristics have to be identified (like trend type or any deviations)
  - several trends have to be compared

- **In the case of only one trend, line graphs also proved to be more favorable than bar graphs.**
  Thus line graphs are the most appropriate graph type when typical characteristics and nuances of trends have to be identified.

- **The results for superbar are by far worse than those of the remaining graph variants.**
  This is comprehensible in the light of the principles of gestalt psychology. The difference rises rapidly with a growing number of trends. If a bar graph is to be taken into consideration for the type of tasks tested here at all, juxtabar is preferable to superbar.

- **The only real alternatives when it comes to deciding which graph variants are most suitable for displaying several data sets in tasks such as those used here are: "Juxtaline or Superline"**.
  When up to 4 trends are displayed, superline and juxtaline achieve similar results. It is not until 8 trends are displayed that superline consistently begins to loose ground to juxtaline.

Experiment 4 - The Visual Perception of Particular Relationships in Line Graphs and Bar Charts in Superposition and Juxtaposition.

Abstract

- Introduction to the Topic
- Variables, Subjects and Procedure
- Construction Principles and Data Generation

Special tasks (Questions asked in the Experiment)

- Where is the maximum or minimum value within one data set?
- Where is the maximum or minimum value within two data sets?
- Where is the greatest ascent or descent within one data set?
- Where is the greatest ascent or descent within two data sets?
- Which of the four data sets exhibits the greatest ascent within a particular area?
- Where is the biggest difference between the two data sets?
- Where is the point of intersection of the two data sets?
- Where is the sum of the four data sets highest or lowest?
- What is the proportion between two data points?

Conclusions

Abstract:

This study marks the end of a project begun by Jacobs in 1994 entitled "Experimental Analyses of the Graphical Representation of Data in Line Graphs and Bar Charts in Superposition and Juxtaposition". The main focus of this project is the visual perception of aspects that can be considered to be special features within a set of data. These include such questions as the maximum or minimum value, the biggest difference between two data points, the greatest ascent or descent within a data set or the greatest difference in ascent between several data sets. Where the constellation of data is favorable such features leap to the eye, but even where the data are complex, spatial perception facilitates the search process greatly and leads to a faster result than when tables are used (see Coll, Thyagaranjan & Chopra (1991), Coll (1992) and Coll, Coll & Thakur (1994)). The main concern of the experiment was to establish empirically which graph type (line graph, bar chart) in which type of graph display (superposition, juxtaposition) is most suited to each particular question.

As in earlier experiments a computer program controlled the whole procedure and provided a wide variety of different data constellations for testing. Forty-six students volunteered to take part as experimental subjects without payment. The quantity of results fulfills our expectation that the advantage of a particular graph type depends on the question asked. Statistical proof was delivered of the effects of graph display (where superposition was at a clear advantage), of moderate differences between graph types in both directions and of multiple interactions between graph type and graph display.

Introduction to the Topic

The use of graphics is particularly useful when the spatial arrangement of the data enables the relevant relationships between them to be determined more easily than in a table. In a graph more data can be viewed at the same time. Some data can be put into data groups or ordered according to particular features so that the human perceptual system is relieved and remains open for complex comparisons. In this experiment we shall analyze particular relationships that in some way indicate interesting points or features:

For example, we shall ask for the maximum or minimum value which, as extreme values in a data set, often offer points of reference and which together indicate the range of the data. Where several data sets are used, the subjects are asked to identify which level of the x-axis displays the highest or lowest sum of values, which requires an estimation of the extreme values of all of the data sets. Within one data set the greatest ascent or descent are prominent points of development. Where two data sets are used, the biggest difference between the two sets can be particularly important. The point of intersection (if there is one) of these two data sets can be of particular significance as it indicates a reversal of the size relationship between them. **These questions are demanding in that they require, as a rule, the analysis of all data elements as well as comprehensive local comparisons and, to a certain extent, comparisons on a global scale**. Some questions require an x-axis with time series information or at least an ordered x-axis. Where other questions are concerned, the scale of the x-axis is not of importance.

One of the focal points of this research is **the question of favorable visual perception through the use of graphical representations**. Thus line graphs and bar charts in superposition and juxtaposition once again enter into competition against each other, whereby the arrangement of the individual graphs in juxtaposition is subject to variation.

Problems concerning cognitive orientation are not of interest to us here. The meaning of the individual parts of the graph is taken to be understood. Nor is it decisive what an individual thinks or what comparisons he or she makes when suddenly confronted with a graph (c.f. Maichle 1994). What is important is how well the person concerned can answer a clear and precise question with the help of each particular graph.

The independent Variables

Depending on the question, different variables were tested. They will be explained in detail in the pertinent section. The following relevant **factors** were examined:

- **Graph typ** (bar chart, line graph)
- **Graph display** (superposition, juxtaposition)
- **Arrangement of the graphs in juxtaposition** (horizontal (all graphs placed next to each other), vertical (all graphs placed one below the other))
- **Grid** (yes,no)
- **Search direction** (maximum value, minimum value)
- **Difficulty of task** (easy/hard)

All the factors tested are repeated measurement factors. When juxtaposition and superposition were compared the mean of the two types of graph arrangement was used.

The Dependent Variables

- **time** (taken for the question to be answered)
- **accuracy:**
  - percentage of correct answers. For the graphical representation of the data the percentage of incorrect answers is used because both dependent variables then point in the same theoretical direction.
  - for proportion estimation the deviation from the correct result is recorded.

A graph variant is only considered superior to another if it shows shorter problem solving times with equal or better accuracy.

Subjects

Forty-six students volunteered to take part in the experiment without payment. Most of them were first year students who had just started their university course and who had chosen to do a statistics seminar as part of an educational science or psychology course.

Procedure

All the subjects completed the experiment in the computer laboratory of the Arts Faculty of the University of the Saarland in groups of 3 to 9, with each sitting at his or her own computer. Depending on how fast each participant worked, the instrument took approximately 1.5 to 2 hours to complete. The whole procedure of the experiment was controlled by a computer program. The subjects were required to complete a number of sample tasks before beginning each test series and were instructed not to start the test until the tasks required of them were quite clear. In each case, the order of appearance of the graphics was determined randomly. For each experimental condition several graphs were tested and grouped as test variables for the purpose of analysis.

In each case, the participants were to answer the particular question **as fast as possible, while still answering it correctly**. The whole procedure is controlled by the actual subject: he or she reads the particular question, determines when the graph presentation is to begin by pressing a key (space bar) and ends it again by pressing a key (space bar). Then the subject answers the question, usually by typing in a number

The Construction of the Data in the Graphics

In a similar manner to that used in Experiment 3 artificial data were generated. They are based on certain functions that are subject to various random processes concerning ascent or descent and the intercept. In addition, each value of the function was altered up or down by a random deviation. The data generated in this fashion do not follow an ideal pattern, but nor do they produce complete chaos. What they should do is produce results that are empirically feasible and that have certain structural characteristics.

Construction of the Experimental Conditions

The construction of the experimental conditions is based in principle on the method used in the previous experiments. Unlike in usual group experiments, the participants did not receive the same material as regards the data in the graphs when all other experimental conditions were identical. Each subject was required to complete parallel tasks in order to ensure the comparability of the experimental conditions. However, 'subject 2' worked partly with data constellations producing a different structure than 'subject 1' under the same experimental condition (e.g. each were shown different function types). In this way, if the internal validity (of the results of each individual) was sufficient, the external validity was able to be improved greatly because so many variable data constellations were able to be tested in the experiment as a whole.

Comparability of the Requirements

All the questions in this study require certain estimates that were standardized in a particular way for experimental reasons. For example, it was decided that the highest value must 1/25 of the length of the axis of ordinate higher than the second-highest value. In order to be able to deliver proof of differences between the individual graph variants the tasks must not be too trivial, but they must also be solvable. Self-tests were used to try to determine the critical differences. Due to these specifications graphs are shown here for each of the questions, since, as concrete examples, they fulfill the experimental conditions that were laid down.

Locate the Maximum or Minimum Value within a Data Set!

Experiment
Toolbook 1.5
Please send me the experiment * * * * You will need Toolbook 1.5 Runtime

Besides enabling the visual perception of the trend of a curve and the comparison of groups of data, graphical presentations are particularly suitable for comparing and ordering values. In an earlier study Jacobs (1990) impressive proof was delivered in numerous experiments that, where the recognition of the relation "bigger/smaller" is concerned, graphical presentation (bar charts) was clearly and consistently superior to tables. This suggests that it is worthwhile using graphics for this type of question, especially, if there a many levels on the x-axis in the presentation.

But the table has advantage about the graphs, if there are many data series in the presentation, a few levels on the x-axis pro data serie and a lot of time is required to identify the labels of the graphs, as were shown by Meyer, J. Shinar, D. & Leiser (1996) and Meyer (1996).

In a graph maximum and minimum values mark interesting points of reference and in this part of the experiment we wanted to find out whether line graphs or bar charts enable extreme values to be identified more easily.

Hypotheses

Our expectation is that the maximum or minimum will be located in both graph types at more or less the same speed, but that bar charts will allow the level of the x-axis corresponding to the extremum to be found faster. That means that the extremum can be located at the same speed whether it is represented as a dot or as the end of a bar, but that in a bar chart the level of the x-axis can be determined faster because the bottom end of the bar leads directly to it, whereas with a line graph the viewer must first imagine lines of reference, which is in any case more time-consuming. Thus we can expect the principle advantage of bar charts over line graphs to be the identification of the maximum value. This advantage should disappear when a grid is used in line graphs to improve reference.

Design of the Experiment

The design of the experiment can be specified as a 3-factor repeated measurement design with the factors **graph type** (bar chart, line graph), **grid** (with/without) and **search direction** (minimum value, maximum value), as shown here:

	bar chart		line graph	
	with grid	without grid	with grid	without grid
maximmum value?	2	2	2	2
minimum value?	2	2	2	2

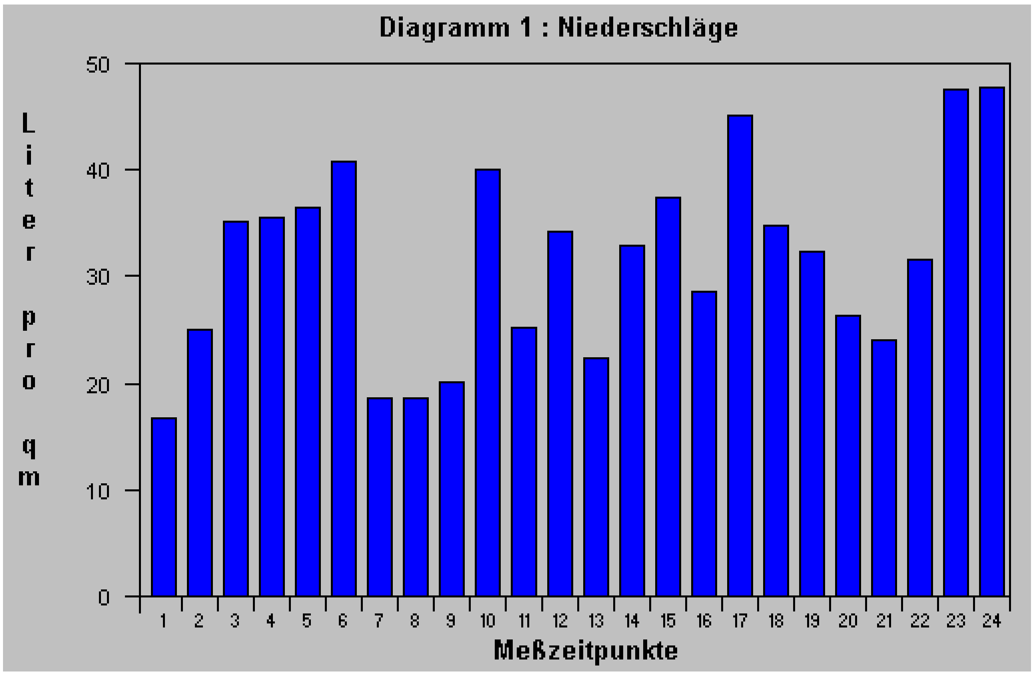
The figures correspond to the number of presentations per condition.

All the presentations were constructed so that the extreme value was 1/25 of the axis of ordinate more extreme than the value closest to it.

Examples of Tasks

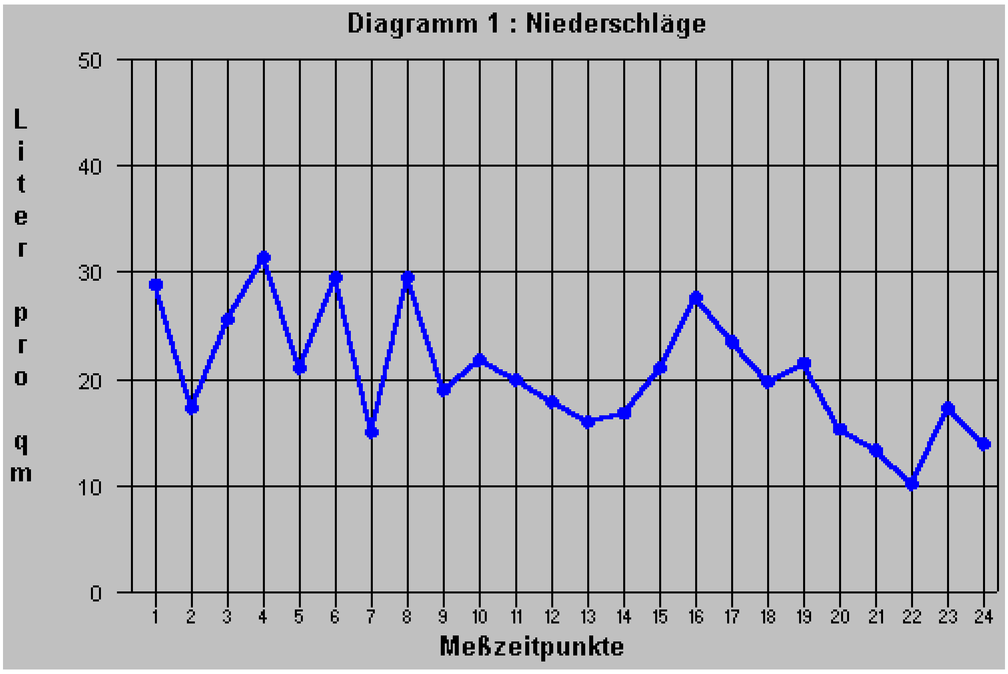
Decide as fast as possible, while still answering correctly, which level of the x-axis displays the extremum requested.

- **Bar chart without a grid:** where is the minimum value?



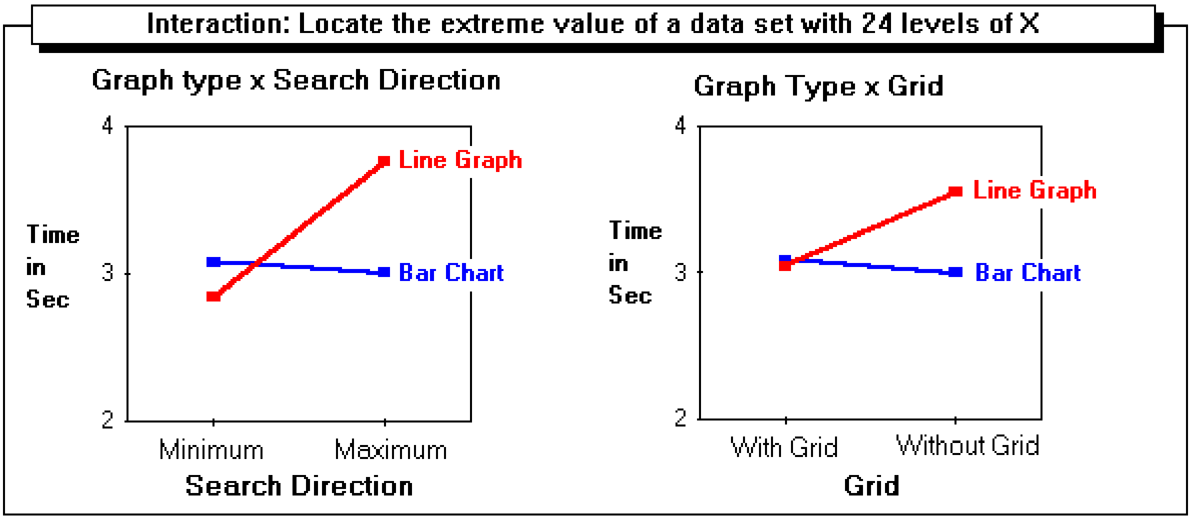
correct answer: 1

- **Line graph with a grid:** where is the maximum value?

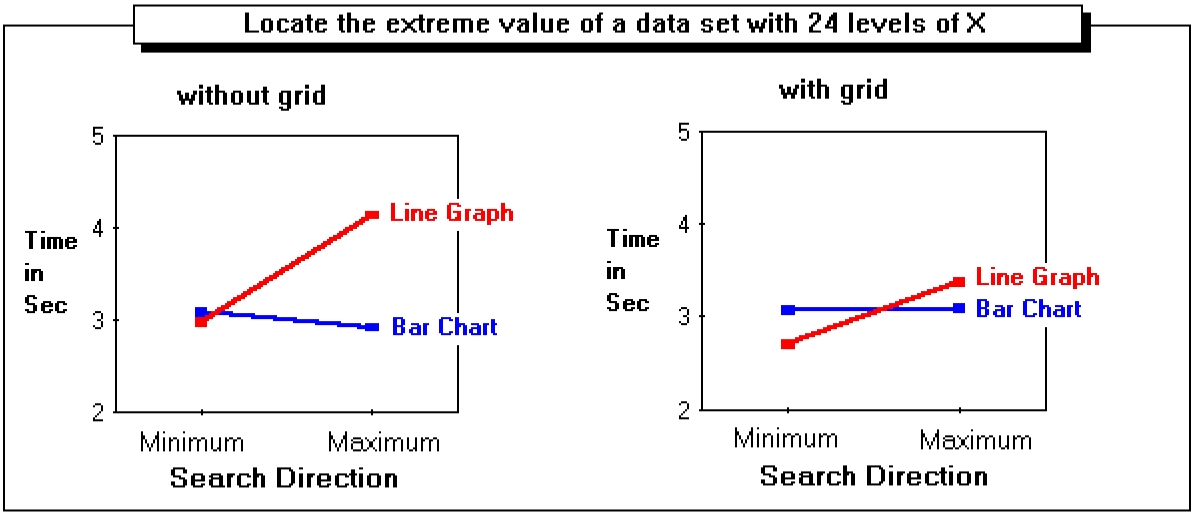


correct answer: 4

Results: arithmetic means (N=46)



- Both graph types achieved comparably high accuracy with over 95% of the answers correct.
- Only where the identification of the maximum value is concerned do bar charts produce better times than line graphs (significant interaction between graph types and search direction; $F(1,45)= 40.88$, $p<.01$).
- When a grid is used, the advantage of bar charts over line graphs disappears (significant interaction between graph types and grids; $F(1,45)= 14.15$ $p<.01$).



Locate the Maximum or Minimum Value within two Data Sets

In this case the identification of the extreme values was also required. This time, however, two data sets, each having 12 levels of X, were used in order to test not only differences between graph types, but also the differences between the graph displays (superposition, juxtaposition). The four graph variants 'superbar', 'superline', 'juxtabar' and 'juxtaline' were compared amongst each other. Each time the maximum or minimum value in the presentation as a whole was to be located. Thus the subjects had to ignore what set a particular data point belonged to and instead concentrate on the extreme value (greatest precipitation in the presentation as a whole). The participants were only required to state the level of the x-axis that contained the extreme value and not the data set to which the value belonged.

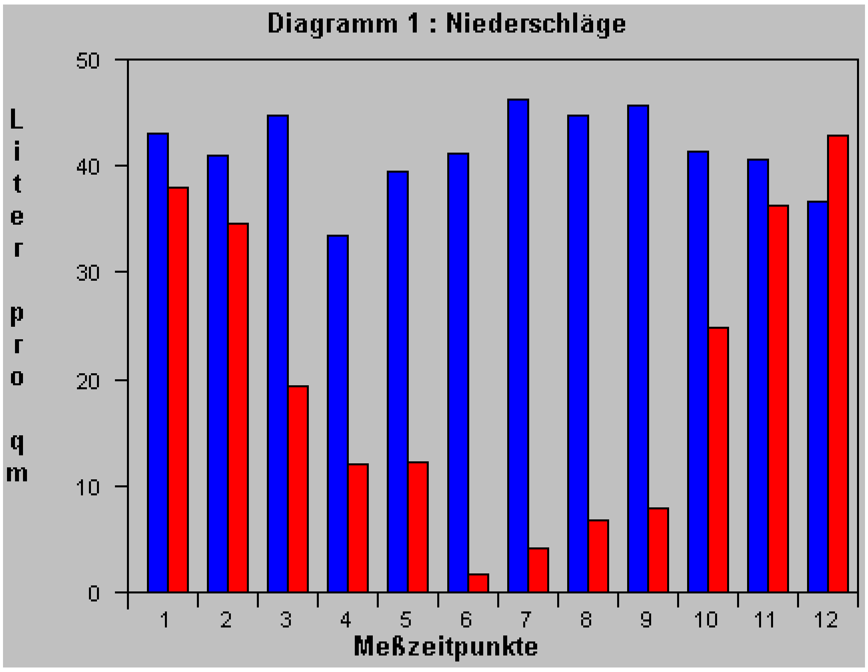
Design of the Experiment

The design of the experiment is a 3-factor repeated measurement design with the factors **graph type** (bar chart, line graph), **graph display** (superposition, juxtaposition) and **search direction** (minimum value, maximum value). Juxtaposition, belonging to the factor of graph display, was varied in terms of the arrangement of the individual graphs (horizontal, vertical). All the presentations were constructed so that the extreme value was 1/25 of the axis of ordinate more extreme than the value closest to it.

Examples of Tasks

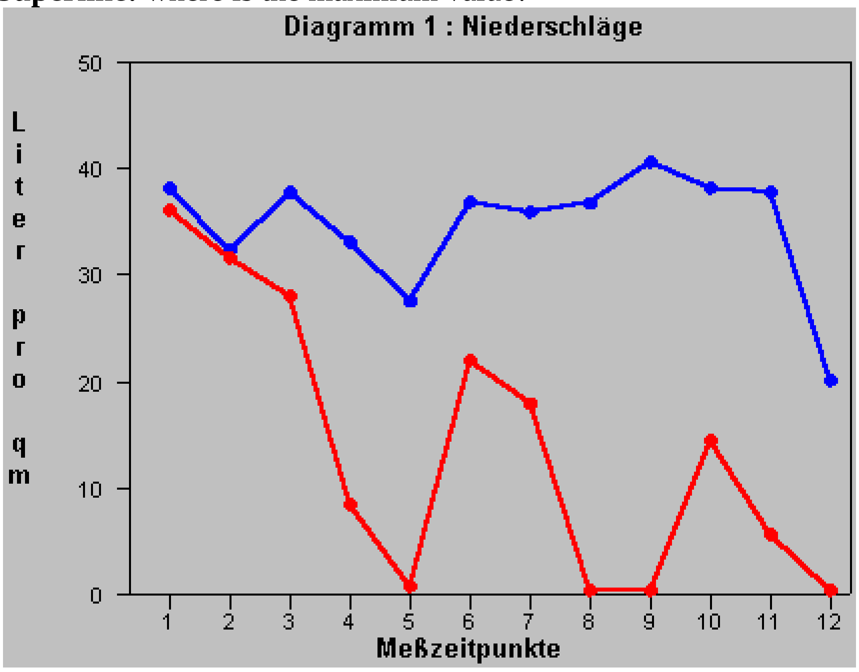
Click on an example and when the graph appears decide as fast as possible, while still answering correctly, which level of the x-axis displays the extremum requested. · Superbar: where is the minimum value? · Superline: where is the maximum value? · Juxtabar (graphs arranged vertically): where is the minimum value? · Juxtaline (graphs arranged horizontally): where is the maximum value?

Superbar: where is the minimum value?



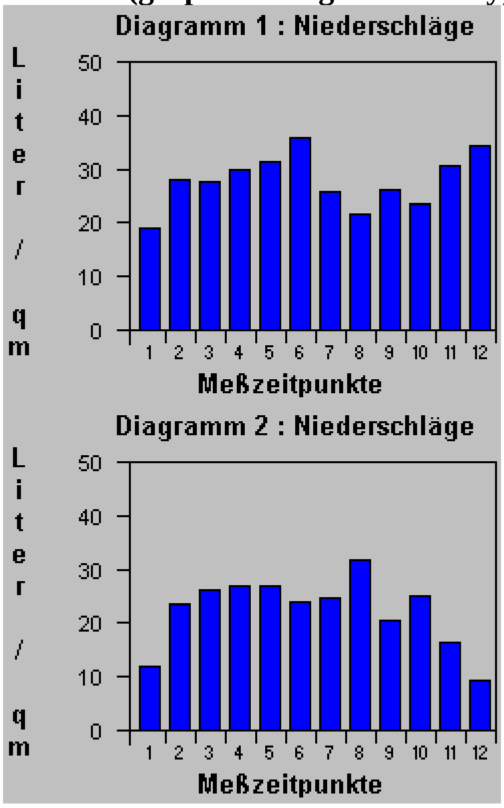
Level 6 is the right answer in this case because the data set displayed in red shows the least precipitation of all the values in the presentation at this point.

Superline: where is the maximum value?



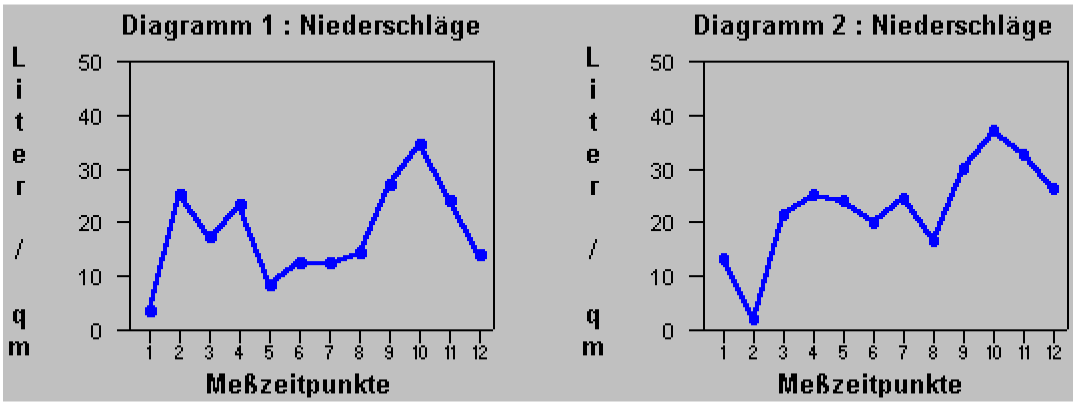
The correct answer here is 9 because at this level the data set displayed in blue shows the maximum value of all the data points.

Juxtabar (graphs arranged vertically): where is the minimum value?



12 is the correct answer because level 12 in the bottom graph displays the least precipitation of all the bars.

Juxtaline (graphs arranged horizontally): where is the maximum value?



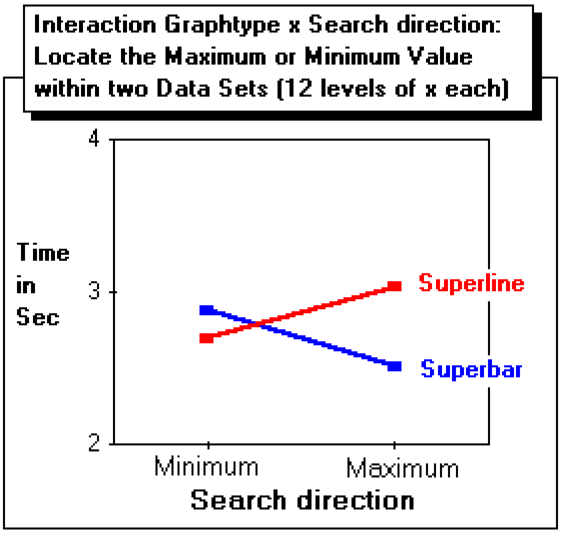
10 is correct because level 10 in the right-hand graph displays the greatest precipitation.

Results for all the graph variants: (time in seconds) (N=46)

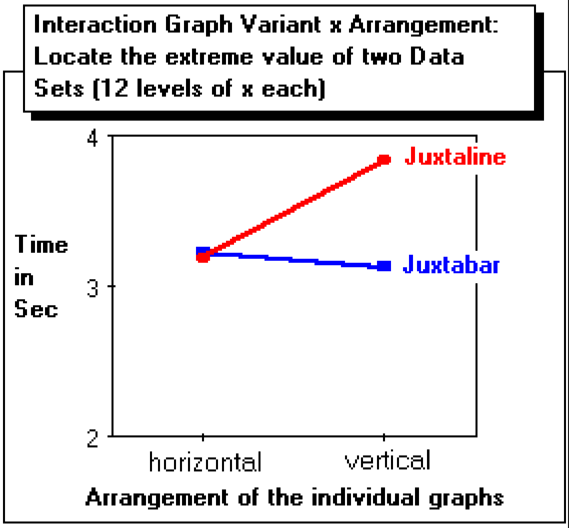
		Graph type	
		bar	line
Graph display	Super	2.7	2.9
	Juxta	3.2	3.5

- All the graph variants achieved comparably high accuracy with over 95% of the answers correct.
- Of all the possible factors, only the factor of graph display had a definite, significant effect on response times (superposition being the clear winner; time: $F(1,45)=50,64$; $p<.01$). The advantage of superposition can probably be accounted for by the closer spatial proximity of the data points to be compared.
- For superposition the interaction between graph type and search direction (established previously for one data set) was confirmed clearly (time: $F(1,45)=20,15$; $p<.01$). In this case it is also easier to identify the maximum value in a bar chart than in a line graph ($t(45)=4.48$; $p<.05$). This result corresponds to the outcome of Meyer, Shinar und Leiser (submitted for publication; also based on a comparison of data sent by Meyer via email on 6. Nov. 1996).
- For juxtaposition there was no significant confirmation of the interaction between graph type and search direction (previously for one data set the interaction was significant), although the numeric data clearly suggest that the maximum value can be identified faster in a bar chart (3.2 s) than in a line graph (3.8 s).

Results for Superposition (N=46)



Results for Graph Arrangement in Juxtaposition (N=46)



Only in the case of line graphs does graph arrangement have a positive effect in favor of horizontal arrangement (*significant interaction between graph variant and graph arrangement ;* time: $F(1,45)= 7.89$ $p<.01$).

Which section of the graph displays the greatest ascent or descent?

Questions concerning the greatest rise or the greatest fall refer to special features in a graph and require the analysis of all the data points of a set of data. Such questions are primarily (but not exclusively) useful for finding out information about trends. In practice they are used to determine such things as the decisive collapse in an economic cycle. The greatest rise in a line graph corresponds to the greatest difference between a data point and the one immediately preceding it in a bar chart.

The difference in the way both graph types are viewed can be summarized as follows: in line graphs rises or falls are compared directly as visible sections of the curve, whereas in bar charts the difference in height of consecutive bars is assessed.

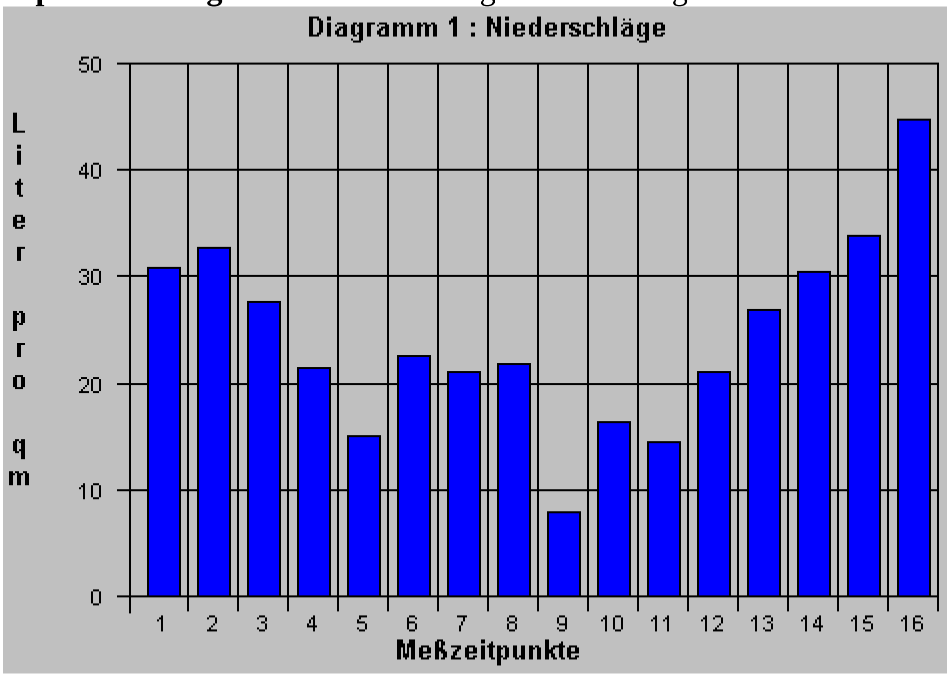
Design of the Experiment

Even though the constellation of data was subject to variation, we wanted a clear criterion (which would be comparable for all the tasks) concerning the maximum rise or fall of a segment of the curve. So it was decided that the maximum rise or fall of a segment of the curve should be 20 per cent more extreme than the maximum rise or fall of any other segment of the curve. The subjects had to identify the level of X of the segment of the curve that marked either the beginning of the greatest rise or the greatest fall ("Where does the greatest rise begin?"). The graph types used were bar charts and line graphs. The experiment was designed as a 2x2x2 factor experiment with the factors **graph type** (bar chart, line graph), **grid** (yes,no) and **search direction** (greatest rise, greatest fall).

Examples of Tasks

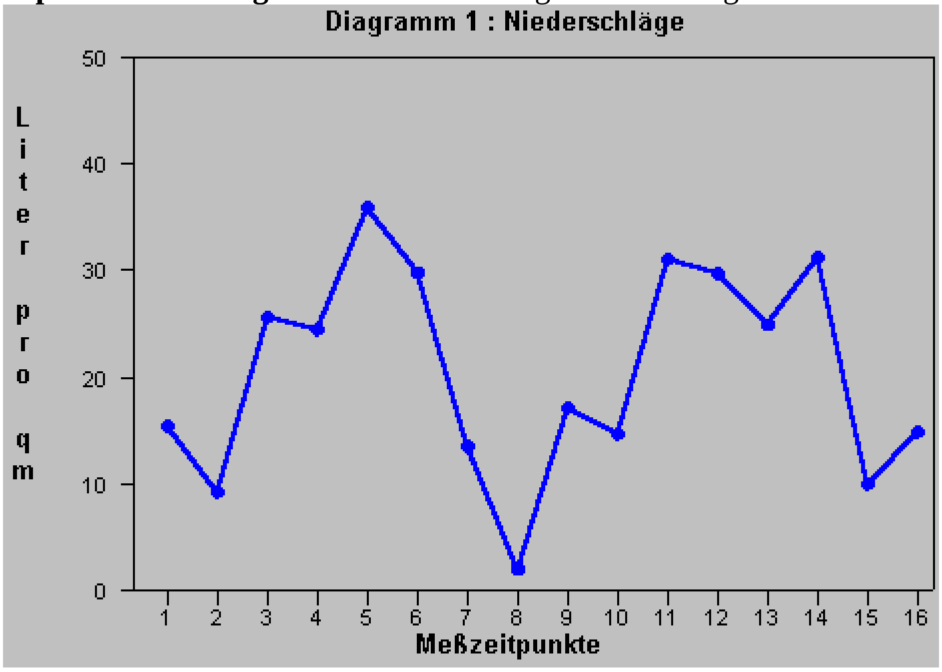
Decide as fast as possible, while still answering correctly, at which level of the x-axis the greatest rise or the greatest fall begins.

- **Superbar with grid**: where does the greatest rise begin?



The correct answer is 15

- **Superline without grid:** Where does the greatest fall begin?



The correct answer is 14

Results

Mean (**M**) and standard deviation (s); (N=46) for the question
'Which section of the graph displays the greatest ascent or descent?'

	error in %		time in Sec	
	M	s	**M**	s
Bar Chart:	**6**	10.5	**5.7**	2.6
Line Graph:	**6**	10.8	**5.2**	1.9
			p = .008	

- Of all the possible factors only the graph type produced a significant effect on response time. In line graphs the greatest rise or fall can be identified faster.

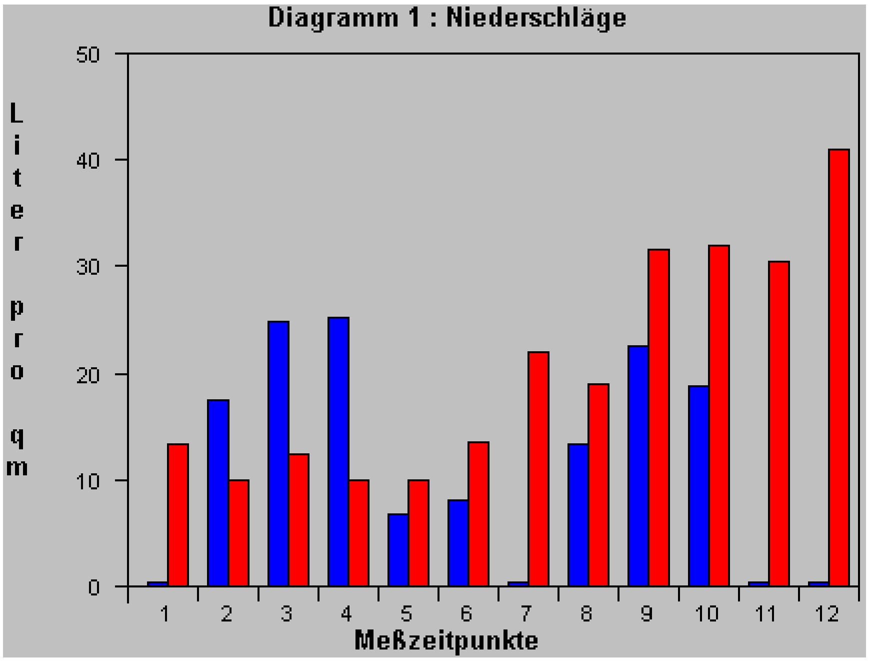
Locate the greatest rise or fall within two data sets!

This task requires a longitudinal analysis of the ascending or descending segments in both data sets and is thus more difficult than when just one data set is analyzed. The greatest change was to be located taking all the segments of the curve and of the presentation as a whole into account. For instance, if the annual production figures of two automobile manufacturers are portrayed, we are interested in finding the segment in which, say, the greatest annual increase took place, an increase that may have taken place in company A or B. The graph variants tested were superbar, superline, juxtabar and juxtaline. In this case it was also decided that the greatest rise or fall should be 20 per cent more extreme than the next biggest rise or fall. The experimental design is a 2x2 factor design with the factors **graph type** (line graph, bar chart) and **graph display** (superposition, juxtaposition).

Examples of tasks

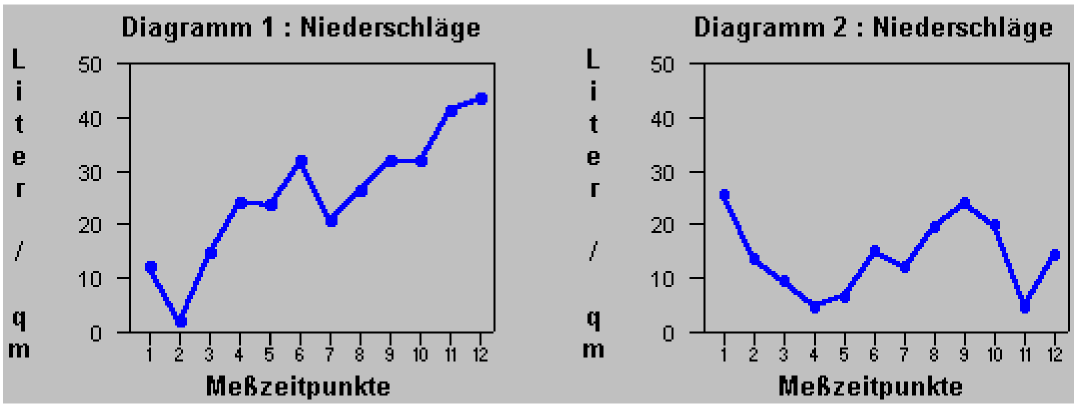
Decide as fast as possible, while still answering correctly, at which level of the x-axis the greatest rise begins.

Superbar Where is the greatest rise?



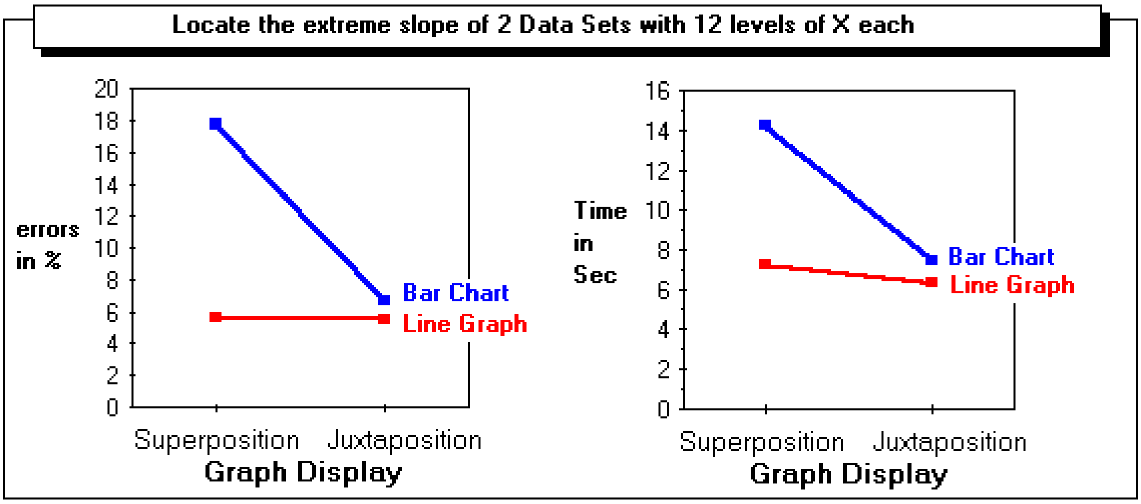
Where does the greatest rise begin?
In this case, all the rises in the data set displayed in red and all those in the data set portrayed in blue must be examined. If this is done, it will be seen that of all the ascents in both data sets, the one from level 1 to 2 of the data set displayed in blue is the greatest. Thus the greatest rise begins at level 1. The subjects had to enter '1' as the answer, but did not need to indicate the color of the data set.

Juxtaline, horizontal graph arrangement: Where is the greatest rise?



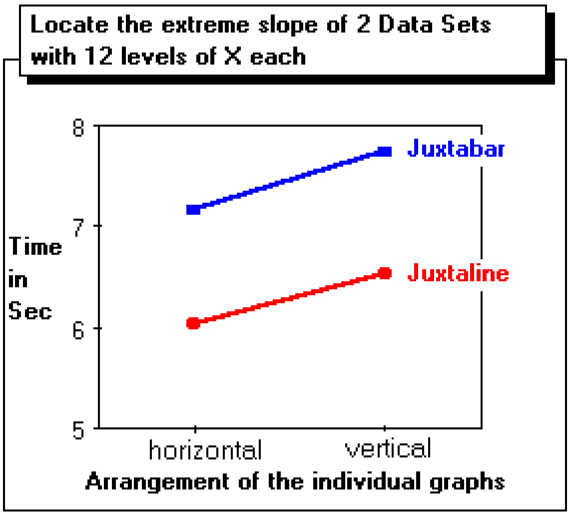
Where does the greatest rise begin?
2 is the correct answer because the greatest rise takes place from level 2 to level 3. The greatest rise is located in diagram 1.

Results for all graph variants: (N=46)



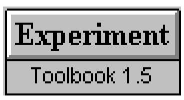
- Superbar is significantly inferior to all other graph variants in terms of accuracy and response times.
- Superline, juxtaline and juxtabar all produce a comparable number of errors. Juxtaline enables tasks to be solved fastest.

Results for Graph Arrangement in Juxtaposition (N=46)



- Juxtaline produces significantly more favorable response times than juxtabar (time: $F(1,44)= 8.53$; $p<.01$).
- This confirms the superiority of line graphs (established for one data set).
- The differences in graph arrangement are not significant. It seems that it is of no importance whether the individual graphs in juxtaposition are arranged horizontally or vertically

Which Data Set Displays the Greatest Rise within a Particular Area of Data?



Please send me the experiment * * * * You will need Toolbook 1.5 Runtime.

One of the previous questions was concerned with comparing slopes within one data set. Here we are interested in differences in slope between various data sets within a particular area of data.
In practice such comparisons are often made, for instance, when a comparative analysis is to be made of different economic sectors in order to find out which sector recorded the highest growth during a boom. To answer the question the relevant area of data (here the relevant levels of the x-axis) must first be isolated. If the levels of X are located directly next to each other, the comparison of slopes can be carried out in the usual fashion because the levels of X will be connected to each other by lines like in a linear function. If other levels of X are located between those to be compared, then there will not be any straight lines, as a rule (at least not in the case of empirical data). Instead the segments of the curve will often be quite different from each other and will no longer display a uniform slope. Strictly speaking, this can no longer be called a simple comparison of slopes. Therefore, what is important to us is the comparison of differences. The question can be defined as follows:

D = (value of level x + n) - (value of level x)
In which of the four data sets is D the highest?

Hypotheses

The results from Experiment 2 (recognition, identification and comparison of trends) clearly showed that superbar was inferior to all the other graph variants, which stems essentially from the fact that it is difficult to isolate the data sets from the presentation as a whole. Isolating works best with juxtaposition and least well with the superbar graph variant. But with juxtaposition, in comparison to superposition, the comparison of data sets is more difficult due to the greater spatial distance and the fact that, even though the scales on the x-axis are identical, they are nonetheless separate. It is difficult to estimate to what extent each part of the visualization task affects the solution of the task as a whole, but the arguments speak for the superiority of superline over superbar. However, this very expectation was not confirmed by the findings of Casali and Gaylin (1988). As part of their investigation into trend comparison, Casali and Gaylin also analyzed questions that required the estimation of slopes and their ratios in several data sets in an area encompassing two or more data points. This question is described in brackets as "differences in line slopes" and is thus very similar to the question being examined here.

Design of the Experiment

Each data set contained eight levels on the x-axis. All the graph variants served as experimental conditions. For juxtaposition both horizontal and vertical graph arrangement was tested, meaning that in total there were six experimental conditions. Four tasks were presented for each condition. For task 1 the levels on the x-axis to be compared were situated directly next to each other; for task 2 differences had to be compared in the area from the first to the last level on the x-axis. In the remaining two tasks the area of data to be compared ranged from three to five data points. The first two tasks are considered to be relatively easy, the last two to be relatively difficult because the identification of the area of data is more difficult for the last two tasks, which was clearly proved by the empirical results obtained.

The greatest rise must be easy to identify and as such must fulfill certain conditions:

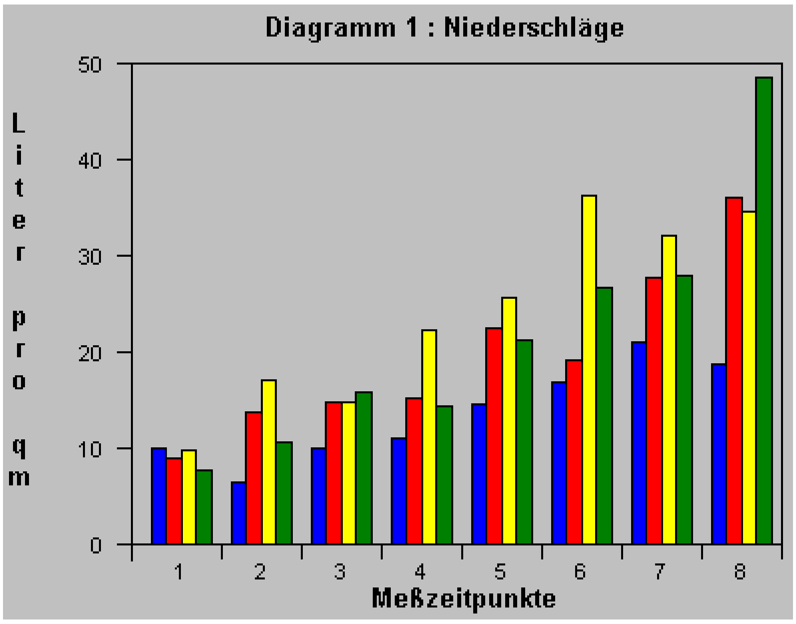
1. The second greatest rise must display a rise of 1/25 of the y-axis, i.e. it must be clearly identifiable as a positive slope.

2. The greatest rise should be exactly 50% steeper than the second-greatest rise.

Example of tasks

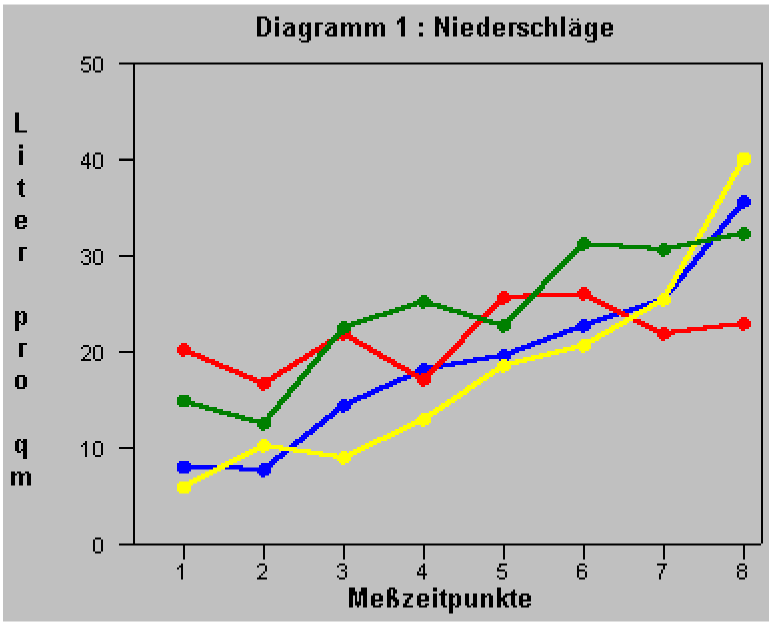
Decide as fast as possible, while still answering correctly, which data set displays the greatest rise within the area of data indicated.

- **Superbar:** Which data set displays the greatest rise as a whole in the area from level 1 to 8?



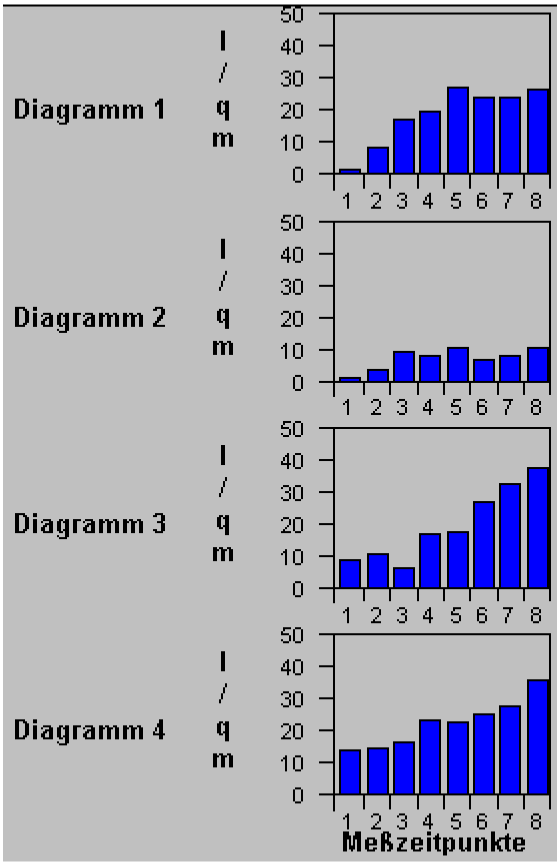
correct answer: the dataset in green

- **Superline:** Which data set displays the greatest rise in the area from level 3 to 7 on the x-axis



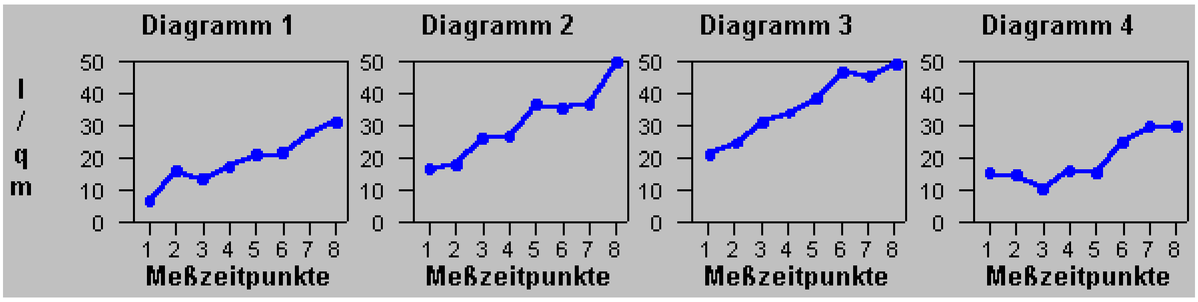
correct answer: the dataset in yellow

- **Juxtabar (arranged vertically):** Which diagram displays the greatest rise as a whole in the area from level 3 to 4?



correct answer: diagramm 3

- **Juxtaline (arranged horizontally):** Which diagram displays the greatest rise as a whole in the area from level 3 to 4?

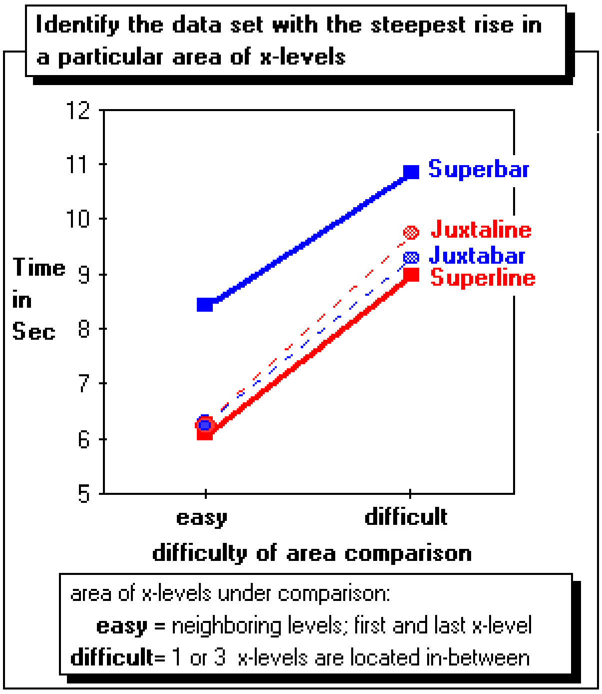


correct answer: diagramm 2

Results

In terms of accuracy there are no differences amongst the graph variants, as over 90% of the tasks were answered correctly at all times.

Results for all Graph Variants: (N=46)



- Superbar requires significantly more time than all the other graph variants, which themselves display no differences.
- The arrangement of the individual diagrams in juxtaposition had no effect. Thus it makes no difference whether graphs are arranged next to each other or one above the other.

Detailed Results:
Comparison of rises in an area of consecutive levels of the x-axis; task 1 *(times; means: N=46)

Super-bar	Super-line	Juxta-bar	Juxta-line
8,6	5,2	6,4	6,4

- Superline produces significantly more favorable times that both graph variants using juxtaposition.
- For juxtabar and juxtaline the vertical arrangement of the individual graphs leads to faster response times (in each case approx. one second) than when horizontal graph arrangement is used.

Where is the Greatest Difference Between the Levels of X of Two Data Sets?

When the trends of two data sets are analyzed, the point at which the two sets show the greatest divergence is often of particular interest. A union leader, for instance, could argue that the divergent development of company profits and wages at a given point in time has reached unknown proportions and suggest that the next wage rounds be used to level out the discrepancy. But the question is not restricted to trends. In order to verify the maximum difference, all the data points in the presentation must be viewed.

In bar charts displayed in superposition, the bars of both data sets are arranged directly next to each other at each level of the x-axis and the difference between them can be seen in the difference in height. With superline the difference must be estimated using the vertical distance between the two data points. The task concerning juxtaposition as the type of graph display is more difficult than superposition for a number of reasons. When the graphs are arranged horizontally, the level of X must be identified by the subject twice (once in each graph), the distance between the data points to be compared is greater and between them are other data points that can have a distracting effect. With vertical graph arrangement it is also more difficult to estimate differences because the data points in question are not located on the same scale and are spatially separated from each other. Thus it was quite clear that juxtaposition would not perform as well as superposition, but it was less obvious which graph type could be expected to produce more favorable results.

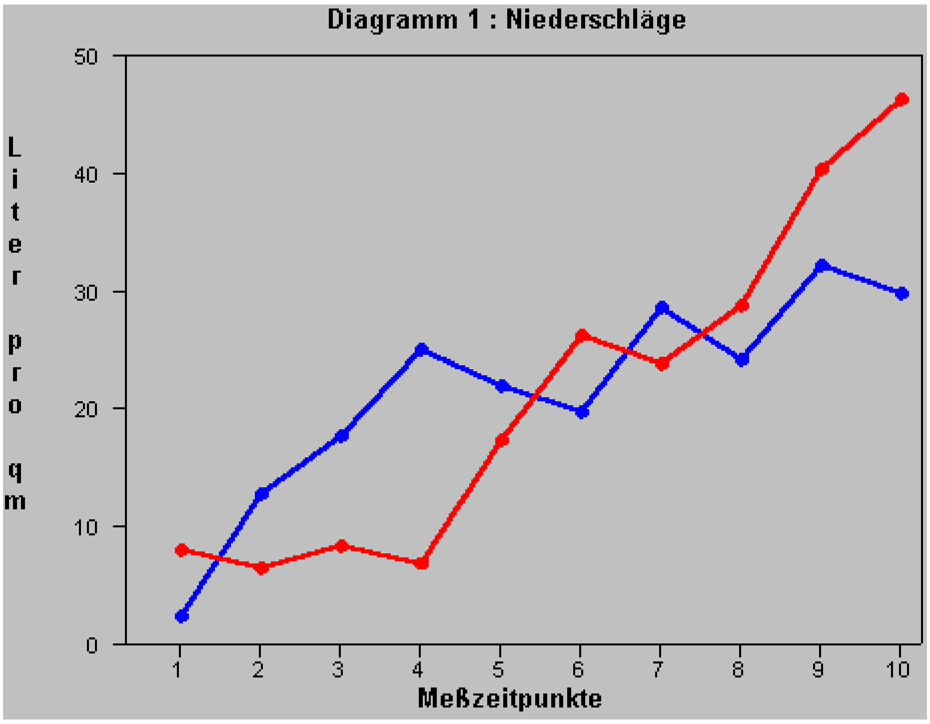
Experimental Design and Procedure

The graph variants used were superbar, juxtabar, superline and juxtaline. For juxtaposition the individual graphs were arranged vertically. The subjects were asked to identify the biggest difference (in terms of the amount) between the two data sets in all 10 levels of X ("Where is the biggest difference between the two data sets?"). Each subject worked on different presentations, which each displayed two data sets based on different functions. For example, task 1 showed two data sets with a positive function, task 2 two data sets with a negative function and task 3 two data sets, one with a positive and one with a negative function. The functions for the individual tasks shown to each subject were sometimes different, but the experimental conditions remained the same. All the presentations were constructed so that the biggest difference was 20% greater than the second-biggest difference. Both the biggest and second-biggest difference appeared in different places and the computer program ensured that there was a sufficient variety of data in order to guarantee reliable results.*

Example of tasks

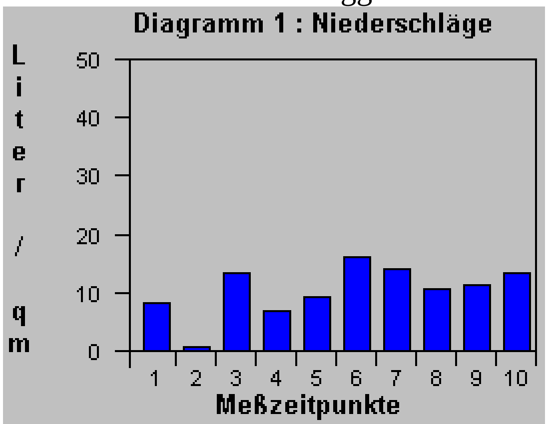
Decide as fast as possible, while still answering correctly, which level of X displays the biggest difference (in terms of the amount).

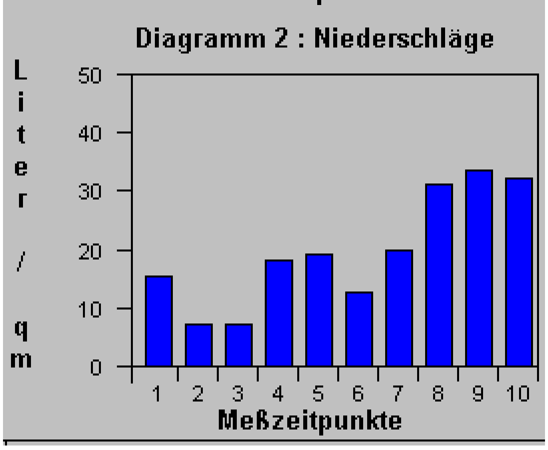
- **Superline:** Where is the biggest difference?



The correct answer is 4 because at level 4 the difference between the data set in blue and that in red is biggest. It does not matter whether the blue line or the red line is located above the other or vice versa. What is important is the amount of difference.

- **Juxtabar:** Where is the biggest difference?

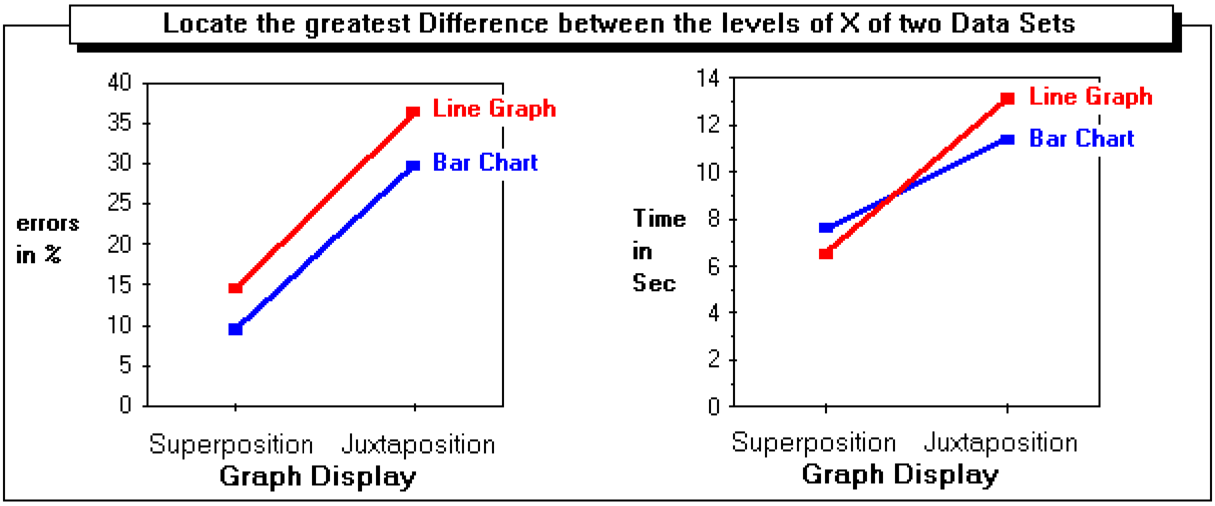




At level 9 because the difference between the bar in diagram 1 and the bar in diagram 2 at level 9 is greater than at all the other levels. At level 9 the difference between diagram 1 and diagram 2 is negative. However, it is of no importance whether the difference is positive or negative. What is important is the absolute amount of difference.

Results

Results for all the Graph Variants: (N=46)



- In terms of the number of errors and response times, superposition performs quite significantly better than juxtaposition.
- The main factor of graph type is not significant in terms of the number of errors and response times.
- As far as response times are concerned, there is a clearly significant disordinal interaction between the graph type and graph display.

The analysis has shown clearly that for the estimation of differences, displays using superposition are definitely preferable to those using juxtaposition. The differences found in this case are considerable and are of great practical significance. In contrast, the differences amongst the graph types are less important.

Locate the Point of Intersection of Two Data Sets

Imagine two sets of data moving towards each other. At some point the time will come when the data set that had the higher x-values at the outset will meet the data set with the lower x-values, cross over it and the relationship between the two will have changed inasmuch as the data set that started off with the lower x-values will now have the higher ones. This point, at which the relationship between the two data sets is reversed, marks a prominent part of a graph and one that is important for some fields. For instance, an employee affected by inflation is hoping for a falling rate of inflation and is looking for the point at which inflation falls below the consistently small wage rises of a few percentage points granted in the last few years.

Hypotheses

The graph variant superline appears to be particularly suitable for this kind of question because the intersection of the two data sets is a obvious feature that can be identified directly in the graph. In superbar graphs the point of intersection itself is not visible and must be determined cognitively. As shown earlier, various trends displayed in juxtaposition can be identified at least as well as when the superline graph variant is used. However, the former type of display does not show the point of intersection. Instead, depending on the slopes in each data set, the viewer has to estimate the location of the point of intersection by comparing the two data sets repeatedly within a realistic area. This is a tiring task and is likely to produce errors because even the comparison of two data sets at one particular level of X is more difficult with juxtaposition. It is difficult to decide whether juxtaline or juxtabar will perform better because the findings to date suggest that the comparison of trends is somewhat more successful when juxtaline is used, but that the comparison of differences is slightly easier using juxtabar. In addition, it appears that answering the question requires skills needed for both of the above examples.

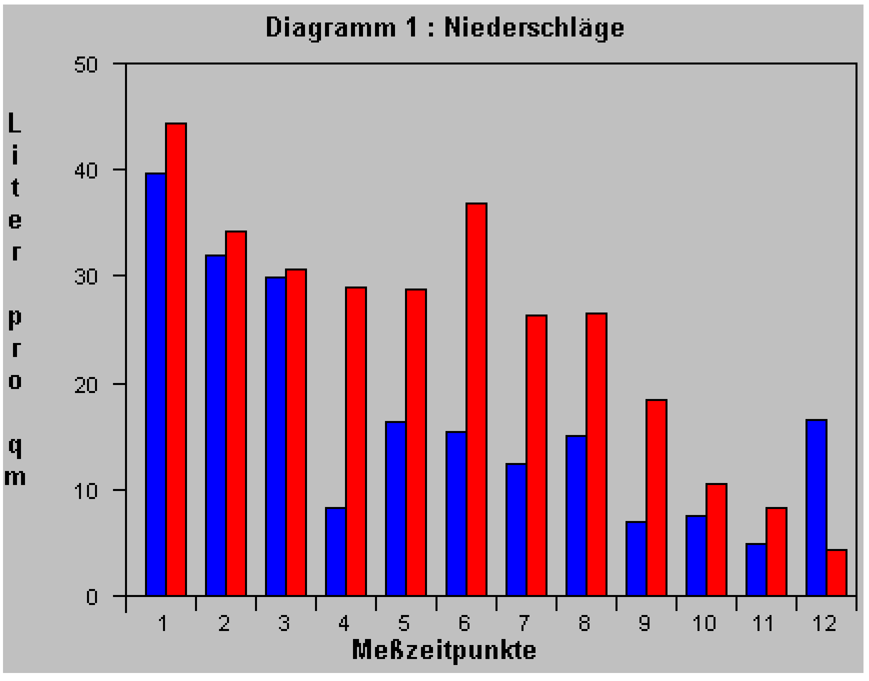
Design of the Experiment

The graph variants used were superbar, juxtabar, superline and juxtaline. In the case of juxtaposition the individual graphs were arranged vertically. The experimental subjects were asked to determine the level of X following directly after the point of intersection of the two data sets. ("Which level of X follows directly after the point of intersection of the two data sets?"). Contrary to the usual principles of construction, this time no criterion was laid down stating that, for instance, all the tasks had to display the same difference (e.g. the same difference in slope at the point of intersection between the two data sets).

Examples of tasks

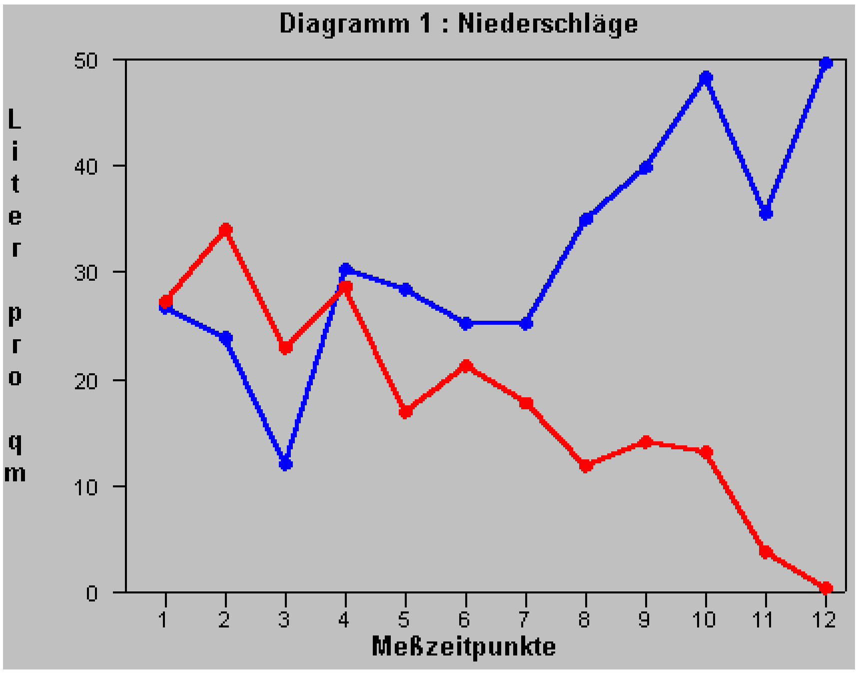
Decide as fast as possible, while still answering correctly, which level of X follows directly after the point of intersection of the two data sets.

- **Superbar:** Which level of X follows directly after the point of intersection of the two data sets?



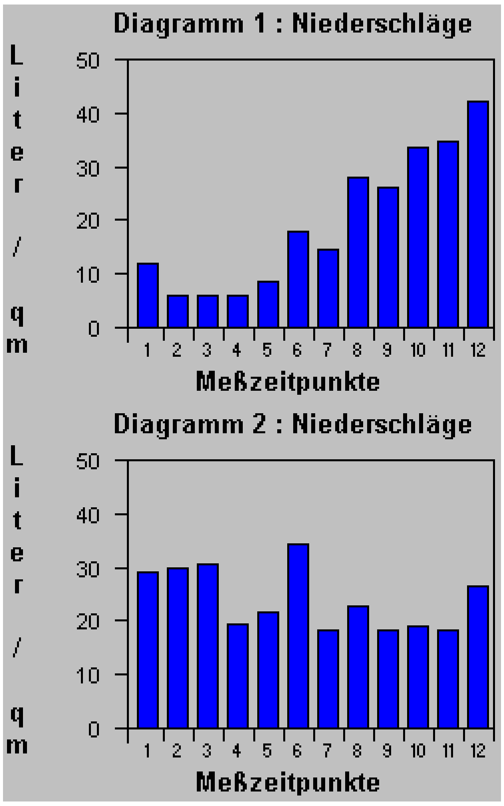
correct answer: 12

Superline: Which level of X follows directly after the point of intersection of the two data sets?



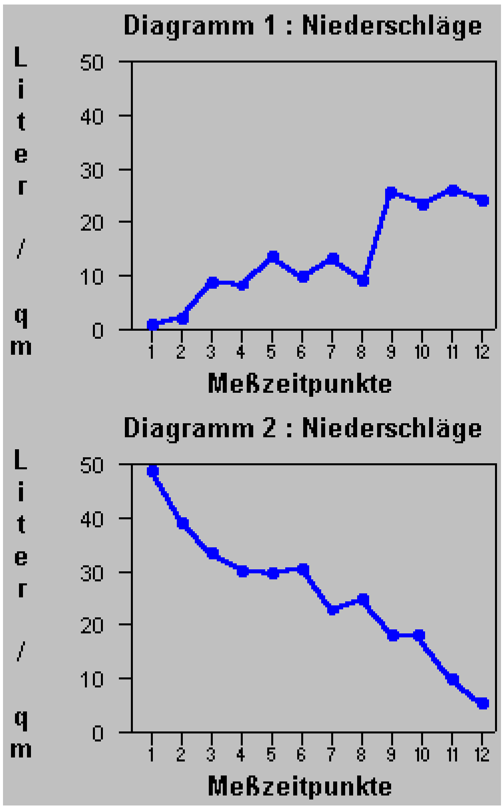
correct answer: 4

Juxtabar : Which level of X follows directly after the point of intersection of the two data sets?



correct answer: 8

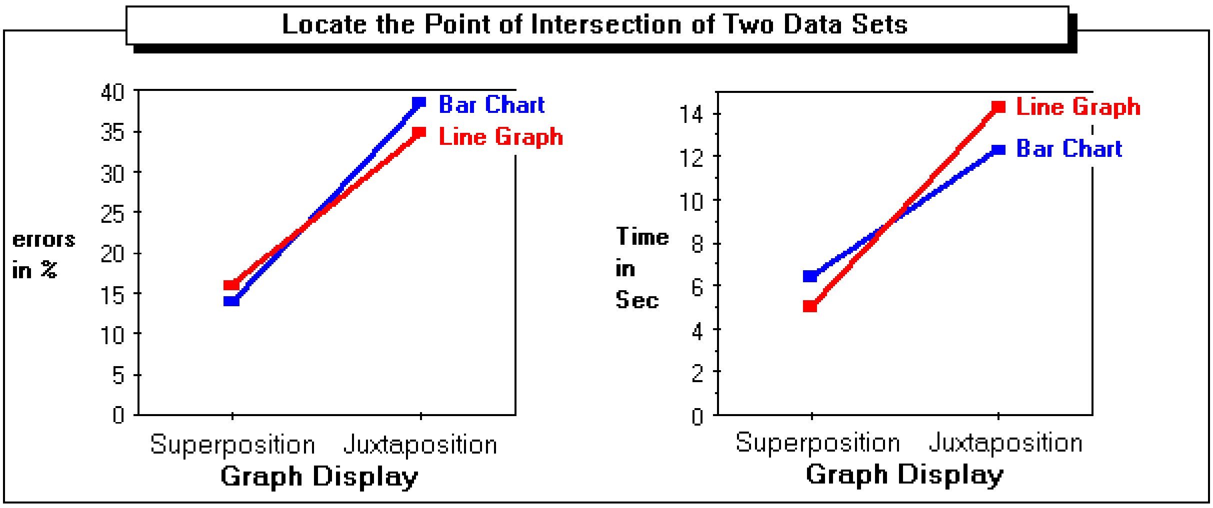
Juxtaline : Which level of X follows directly after the point of intersection of the two data sets?



correct answer: 9

Results

Results for all the Graph Variants: (N=46)



- In terms of the number of errors and response times, superposition performs quite significantly better than juxtaposition.
- The main factor of graph type is not significant in terms of the number of errors and response times.

- As far as response times are concerned, there is a clearly significant disordinal interaction between the graph type and graph display. The statistics clearly show that the question can be answered faster using superline than using superbar ($p<.001$). In contrast, the advantage of juxtabar over juxtaline cannot be proved so conclusively ($p<.05$).

As with the estimation of the maximum difference, the analysis in this case has shown clearly that for the identification of the point of intersection of two data sets, displays using superposition are definitely preferable to those using juxtaposition. Within superposition, superline is the better graph type.

Locate the Level of X with the Greatest Sum of Values (or the Greatest Mean)

In a display representing four data sets, each level of X displays four data points. If the levels are compared, the four data points belonging to the various data sets can be added up and the sum or mean used as a basis for comparing the levels of X. For instance, if the data sets represent four different countries and the levels of X display different products, it may be interesting to identify the product with the best sales figures overall. This task is different from the comparison of mean values required in Experiment 3 in that, in this case, the data points of the four sets of data must be grouped together. In order to do so, the subject has to ignore the data set variable and instead compare the sum of the values of each level of X.

Hypotheses

Superposition appears to be more suitable for this task because the data points to be compared are located closer to each other and can be compared using the same scale. Thus it will be of greater interest to see the results delivered by the individual graph types, particularly of those using superposition. How can you tell what is easier - comparing the sums of 10 levels of X, each made up of 4 differently colored bars, or comparing the estimated mean values of 10 levels of X, made up of 4 differently colored little circles connected by lines? We had no clear expectations.

Design of the Experiment

The graph variants used were superbar, superline, juxtabar and juxtaline. For juxtaposition horizontal and vertical arrangement of the individual graphs was used. The subjects were asked to identify the level of X with the highest or lowest sum of values or the highest or lowest mean value. In half of the tasks testing each experimental condition, the level of X with the highest sum of values (or the highest mean value) had to be identified from a total of 10 levels of X. In the other half, the lowest sum of values or mean value was required. All the displays were constructed so that the highest or lowest sum was 20% more extreme than the next closest sum.

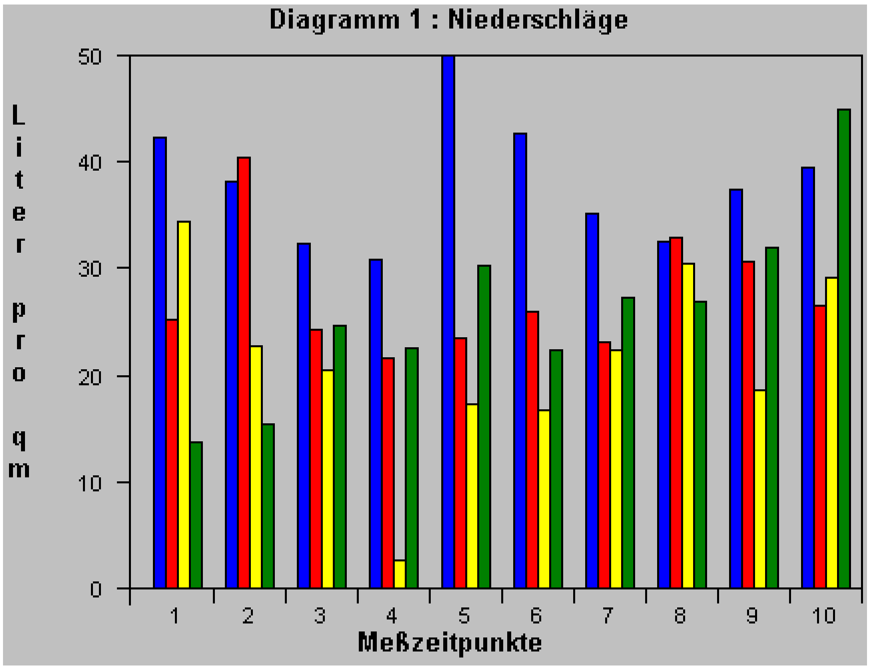
Examples of Tasks

Click on an example and when the graph appears decide as fast as possible, while still answering correctly, at what level of X the most or the least precipitation was recorded.

Superline: where did the most precipitation occur? Juxtabar (horizontal arrangement): where did the most precipitation occur? Juxtaline (horizontal arrangement): where did the least precipitation occur?

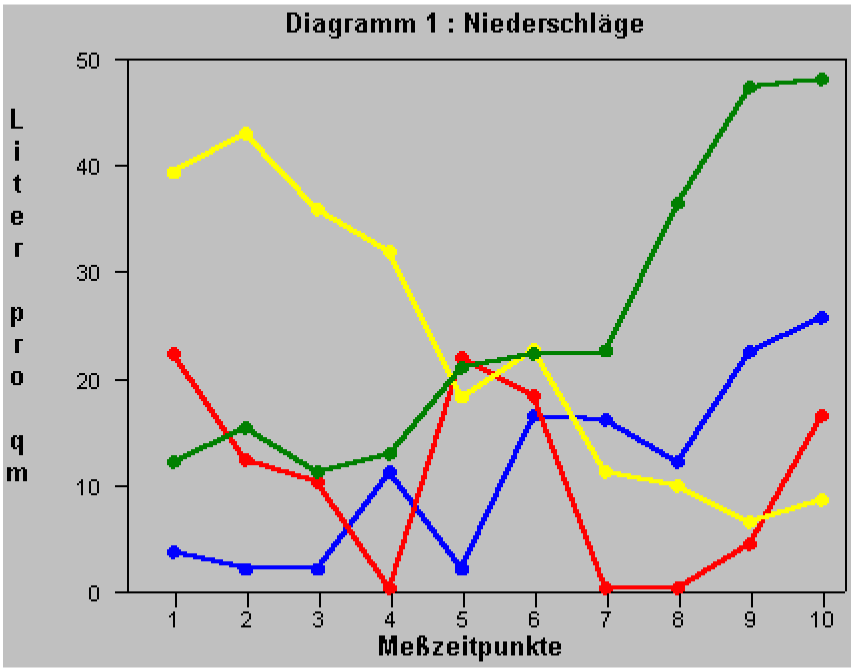
If you are not sure, you will find the correct answers to all the examples here: Ein Beispiel anklicken und beim Erscheinen der Graphik möglichst schnell, aber auch korrekt entscheiden, zu welchem Meßzeitpunkt insgesamt am meisten bzw. am wenigsten Niederschlag fiel!

- **Superbar:** Where did the least precipitation occur ?



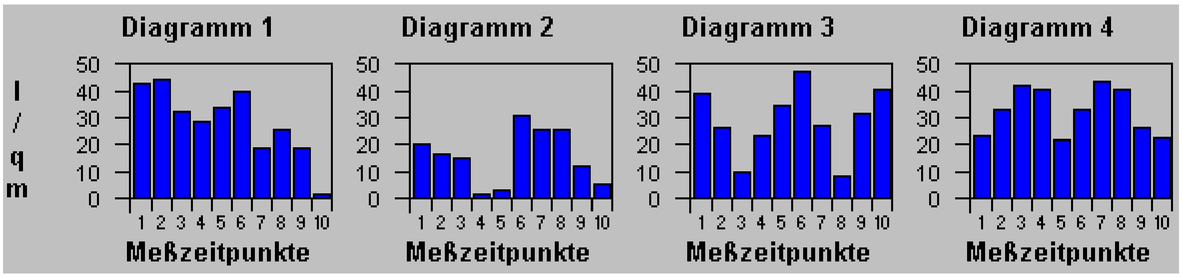
correct answer: the least precipitation (level 4)

- **Superline:** Where did the most precipitation occur ?



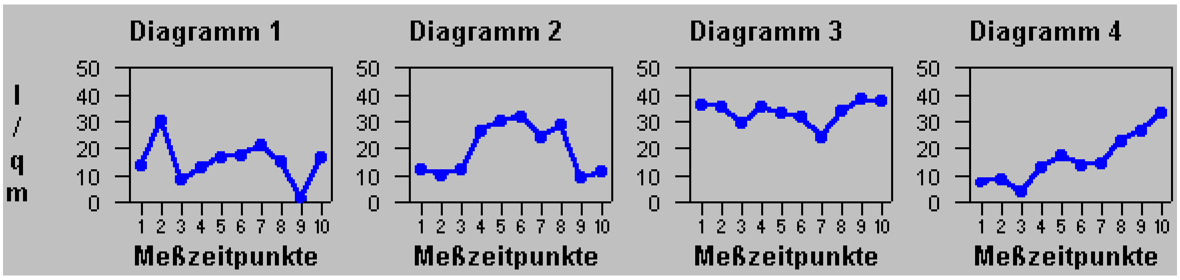
correct answer: the most precipitation (level 10)

- **Juxtabar: (horizontal arrangement):** Where did the most precipitation occur?



correct answer: the most precipitation (level 6)

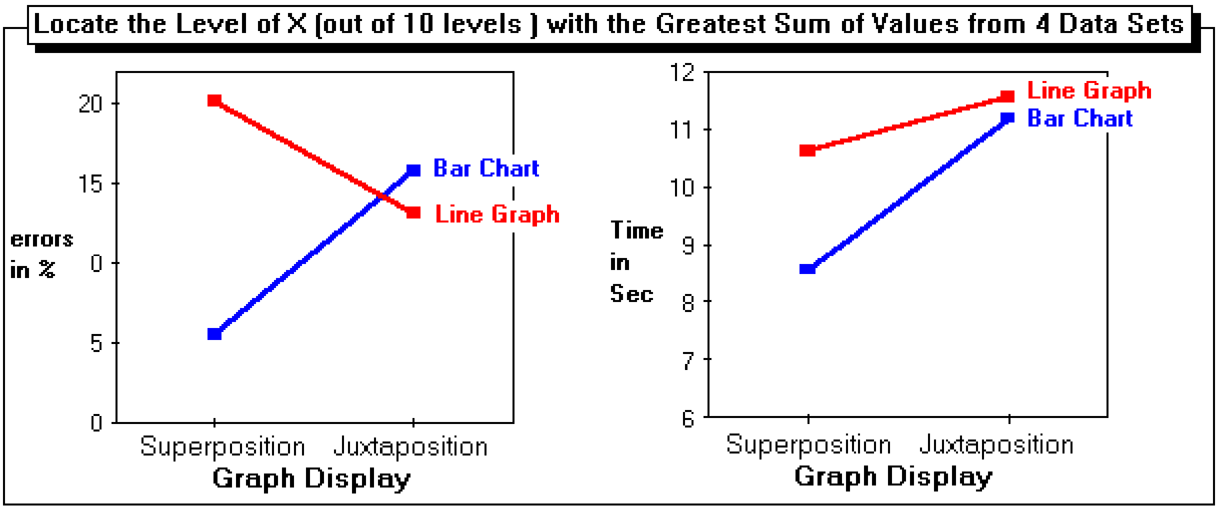
- **Juxtaline: (horizontal arrangement):** Where did the least precipitation occur?



correct answer: the most precipitation (level 3)

Results

Results for all the Graph Variants



- As the graphs indicate, superbar is the clear winner. Superbar is significantly superior to all other graph variants both in terms of accuracy and response times.
- These interactions, for which statistical proof has been delivered, show essentially that the advantage of superposition only holds for bar charts.
- The arrangement of the individual graphs in juxtaposition (results not shown here) has no effect. Thus it does not matter whether the graphs are arranged next to each other or one on top of the other.

Estimating One Value as a Percentage of Another (Ratios)

Estimating ratios is one of the elementary tasks in connection with graphical displays of data and investigations were carried out very early on in the history of the subject. We distinguish between the relationship of one value to all the other values (part-to-whole relationship) and the relationship of one value to another value (local comparison), which is our main interest in this case. The question was tested using only superposition as the type of graph display, for the simple reason that we did not wish to place excessive demands on the experimental subjects. The subjects were asked to estimate the ratio of the smaller value to the bigger value as a percentage. This question is a frequent requirement in graph research. For instance, all the results of Cleveland's (1985, p. 249) elementary perceptual tasks are based on this question.

Hypotheses

The data in the conventional line and bar graphs all have a common scale, as suggested by Cleveland (1985). The values are estimated according to their position and as such, the two types of presentation do not differ particularly in terms of the elementary task of graphical perception. At the very most, one could contend that the bars convey information about their length, as well as about their position in the graph. Nevertheless, the information about the length is less accurate. It is not clear from Cleveland's theoretical observations which type of presentation can be expected to produce better results. I have put forward the hypothesis on various occasions that bar charts are more suited to this type of question than line graphs. But the justification for this hypothesis was shaken by the results from Experiment 3, as, contrary to all expectations, it was found that both line graphs and bar charts delivered comparable results in terms of accuracy and response times for the estimation of values (point reading). A possible difference between the two types of presentation thus cannot be based on different rates of accuracy for the estimation of individual values, but on the different ways in which the two types of presentation allow two values to be compared. Can I isolate two points on a line better than two bars? This task may possibly be easier using a bar chart.

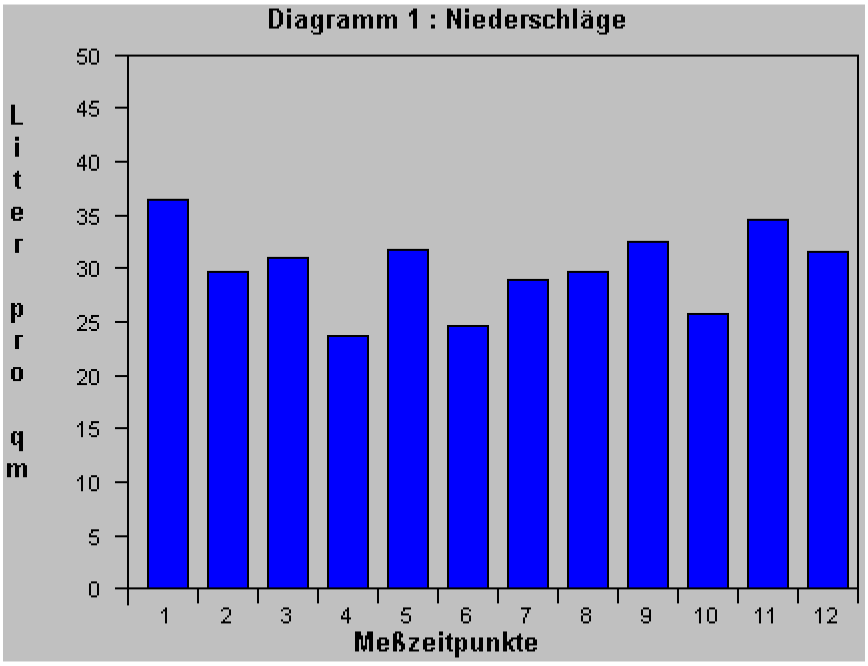
Design of the Experiment

Five presentations were shown each for the graph types bar chart and line graph. Each presentation included 12 levels of X. In two of the five tasks, the levels to be compared were situated directly next to each other; in one of the tasks there was an additional level located between the levels to be compared; in the two remaining tasks, a random number of levels (at least two and not more than ten) were located between the levels to be compared. Each individual task tested the graph types bar chart and line graph using the same levels of X (but not over all the tasks or experimental subjects). For example, subject one might have compared levels 2 and 4 in a line graph and a bar chart in task 1 and levels 6 and 8 in the two graph types in task 2, whereas subject two would have compared different levels, although they remained the same within one task. The levels were selected randomly for each subject from the pool of combinations that fulfilled the given requirements. The values represented in the two levels in each of the tasks were not necessarily the same.

Examples of Tasks

Estimate the ratio of the lower value to the higher value (as a percentage) as fast and as accurately as possible.

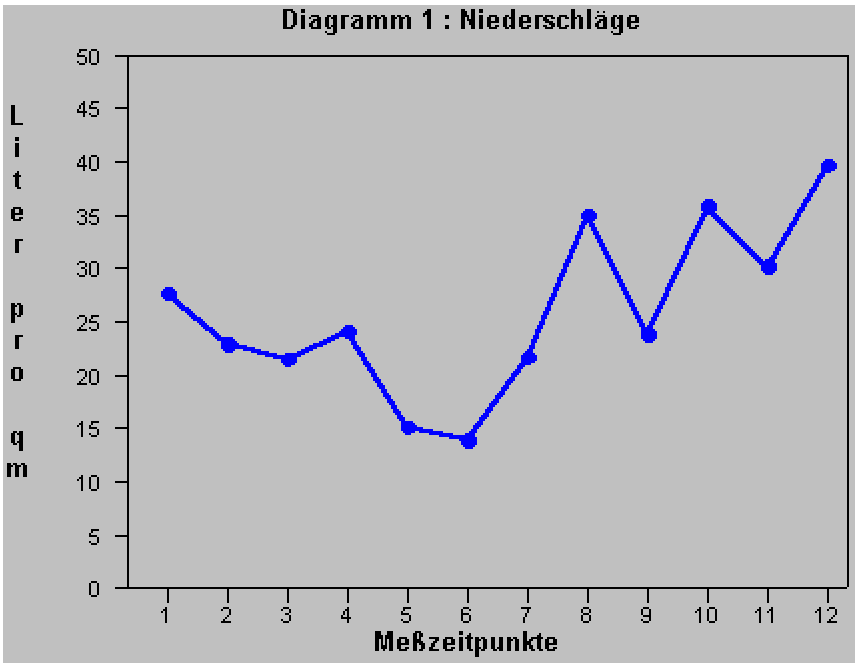
Bar chart: What is the ratio of level 7 to level 9?



correct answer: 89 %

•

Line graph: What is the ratio of level 1 to level 4?



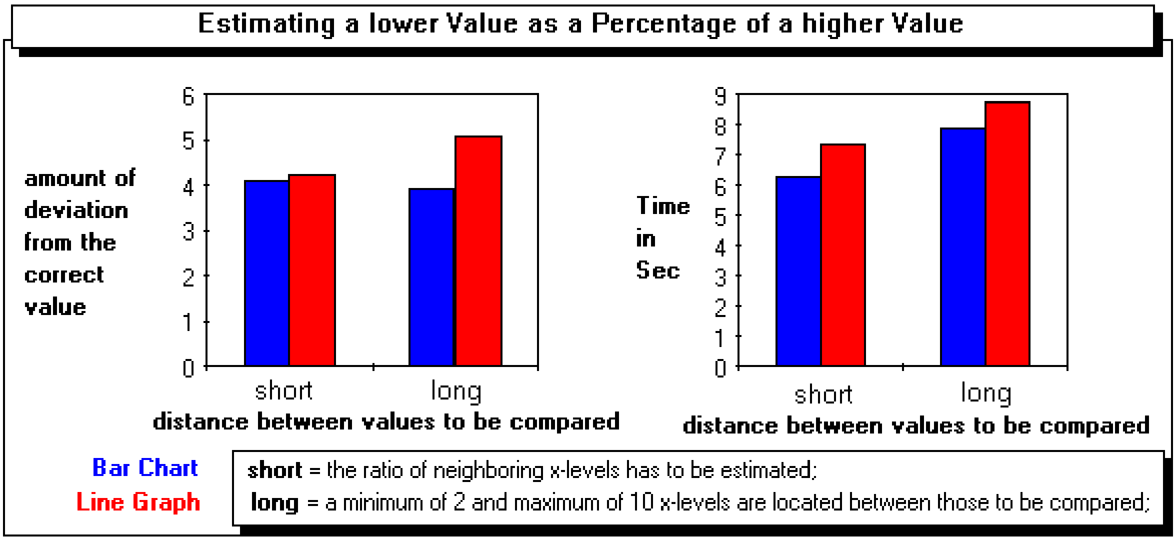
correct answer: 87 %

Results

Means, Standard deviations and t-test for
'Estimating One Value as a Percentage of Another'

	bar chart	line graph
| guess - correct answer |	**3,9**	**4,6**
t(43)=-1,64;p=,108;(ns)	1,7	2,4
Time in Sec.:	**7,0**	**8,0**
t(43)=-3,54;p=,001; (s)	2,6	2,6

The question can be answered faster using bar charts.



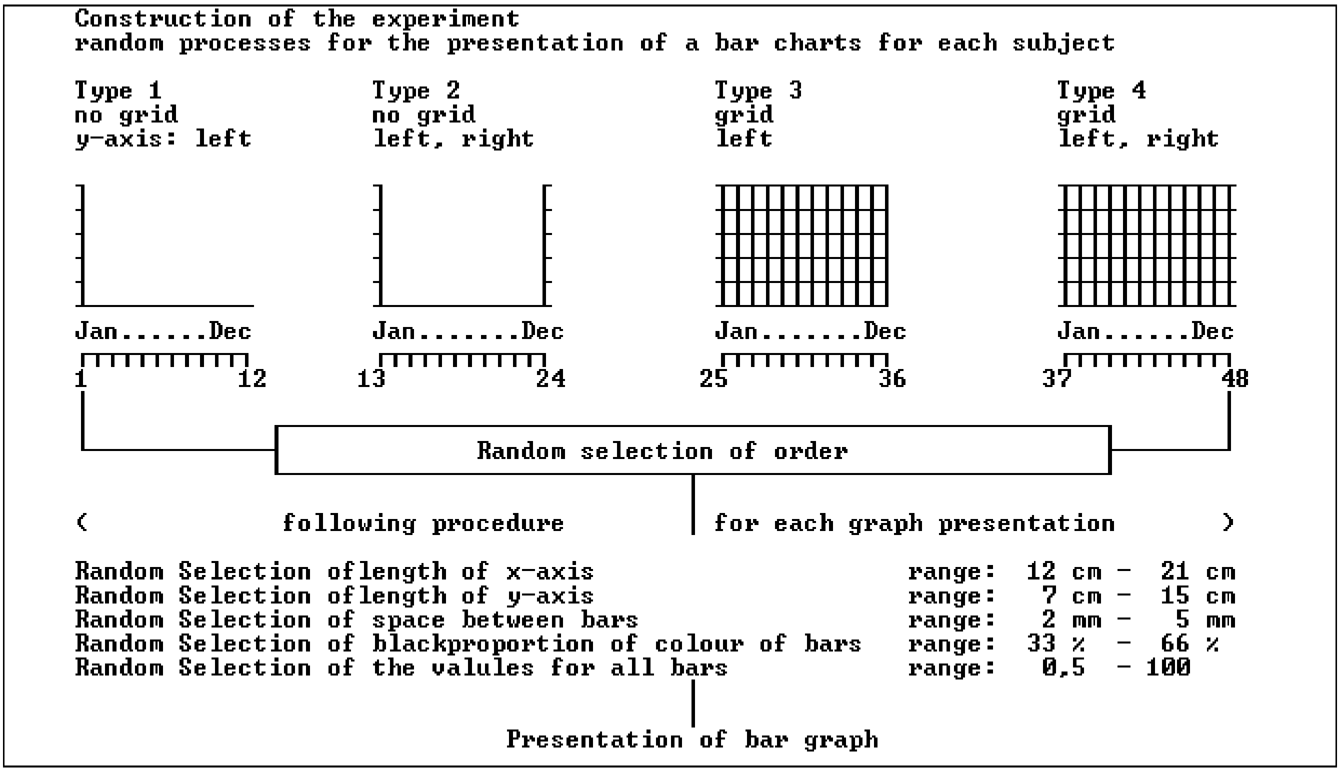
- The further apart the two levels, the more time it takes to estimate the ratio. Contrary to what was expected, no differences in accuracy were found.
- For the time being, the advantage of bar charts can only be seen in statistical terms. The interaction between the graph type and the distance between the levels is insignificant in terms of response times and accuracy.

Summaries about former Works of Jacobs

Point Reading in bar chart (Jacobs 1989)

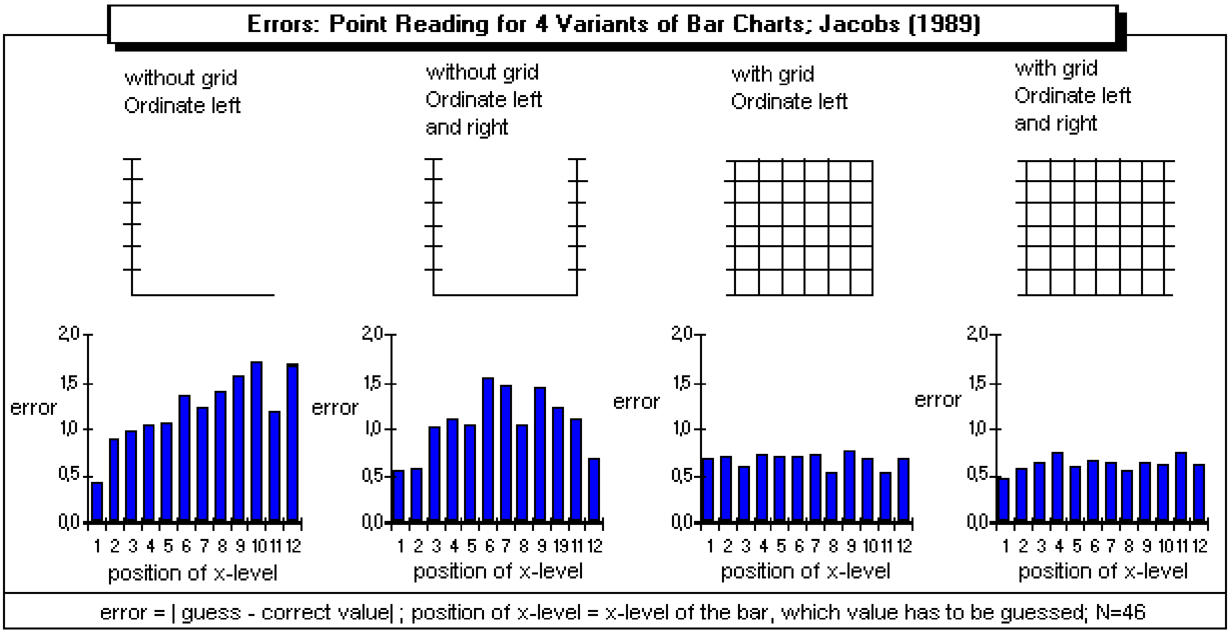
Source: Jacobs, B. (1989). Schnelligkeit und Genauigkeit beim Abschätzen von Größenwerten aus einem Säulendiagramm. Saarbrücken. Arbeitsberichte des Medienzentrums der Universität des Saarlandes, Nr. 2.

48 different bar graphs were shown in the experiment (12 graphs each graph variant). 46 experimental subjects were asked to estimate the values of each category (x-level of the x-axis, here rainfall for a month) with each graph variant (one computer each person). The graphs were generated in the following manner:



The values for each category, the intercepts of y- and x-axis, the distance between the bars and their lightness were determined by chance for each of the 48 bar graphs. The persons were asked to estimate only one value each presentation. The sequence of the presentations was varied by chance. The following data are the means of all the experimental subjects (n=46):

Results



Important Results:

- Both grid variants differ significantly from the bar graphs without grid (this also holds for the times, about 300 msec advantage for the grid variants)
- Errors grow with a rising category position approximately linear with the convential bar graph (see the left graph).
- An additional ordinate reduces significantly the number of errors and the time with the last categories (but only marginally on the whole)..
- The category position loses any influence on the number of errors with a grid.

The results confirm Kosslyn (1994, S.282): "need for a second Y axis I recommend using inner grids, rather than duplicate Y axes, when the reader is to see relatively precise values because I assume that it is easier to trace along a grid line to a single Y axis than visually to track a shorter distance through empty space to a second Y axis."

Additional Results:

During the experiment additional variations (described above) were generated. They lead to the following results:-

- The number of errors decreases with an increasing height of the ordinate, because the resolution turns better.
- There is no significant influence of the length of the x-axis, the distance between the bars, their breadth, their lightness or their length (=value) on accuracy and time.
- Persons differ reliably according to their exactness of estimation and the time needed to estimate.

Comparison: Table versus Graphs (Jacobs 1990)

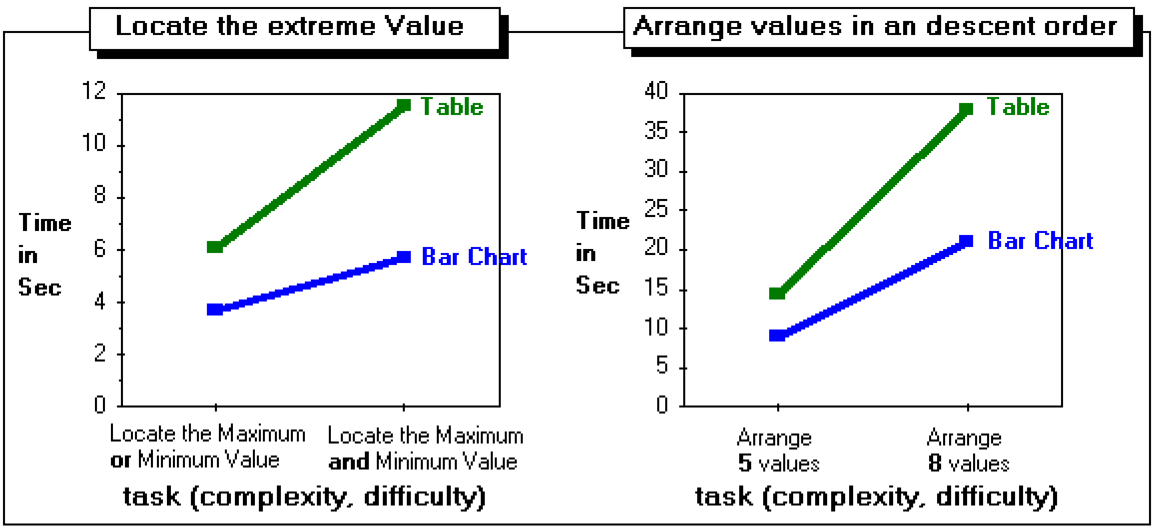
Source: Jacobs , B. (1990). Ein Vergleich der Auswirkungen graphischer und tabellarischer Präsentationsformen auf die Schnelligkeit und Genauigkeit beim Erkennen und Interpretieren statistischer Daten. Saarbrücken. Arbeitsberichte des Medienzentrums der Philosophischen Fakultät der Universität des Saarlandes. Nr. 3

Identifying of Relations between values (Jacobs 1990)

Bigger" and smaller" were the relations between values we were concentrated on in this investigation. 40 experimental subjects were asked to identify the maximum or minimum value in a bar graph or a table (each 12 values). The complexity was also varied. The easier question was to identify either the minimum or the maximum value. The more difficult question was to identify both, the minimum and the maximum value.

The second kind of question was to put 5 values (easy task) or 8 values (hard task) in an descent order. Each value was generated by chance for every experimental subject.

Results: Proved superiority of the bar graph to the table



Important Results:

- The bar graph achieved always a significantly higher percentage of right answers than the table.
- The table needed significantly more time than the bar chart. The times divergences between the kinds of presentation always correspond an effect size of significantly more than 1.
- The interactions between the kinds of presentation and the complexity according to time could be proved statistically on the promille-level.

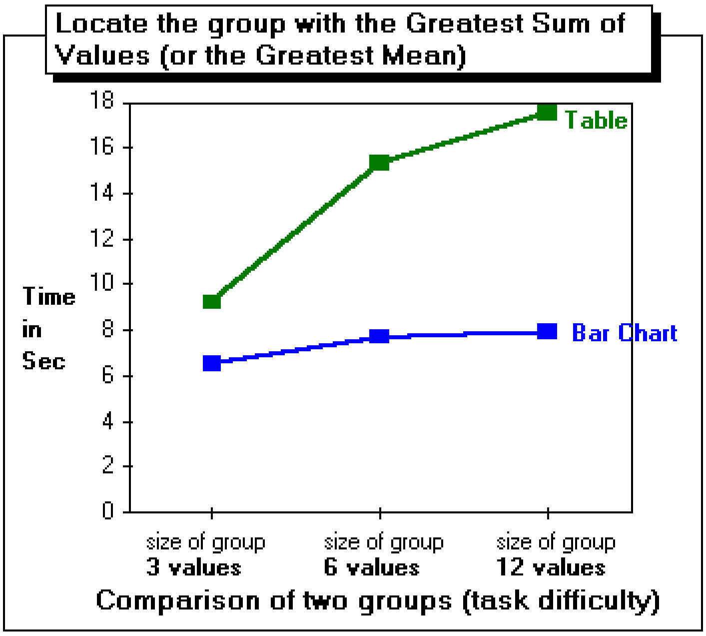
Comparison of Data Groups (Means) within a data serie (Jacobs 1990)

Description of the Experiment:

Data of monthly percipitations of a year were shown. The experimental subjects had to decide, for instance, if there was more percipitation in the first or in the last quarter. To be able to answer this question the experimental subjects had to sum 3 categories for the first and 3 for the last quarter and to compare them with each other. Six months (6 values as a group) and whole years (12 values each group, arranged in juxtaposition) had to be compared later. The Focus of this Experiment: Which kind of presentation achieves the shortest time to answer the question? The basis of this experiment is an design with the factors *kind of presentation* (table, bar graph) and *complexity or difficulty* (3,6 or 12 values each group).

Results

Proved Superiority of the Bar Graph to the Table. (Jacobs 1990)



Important Results:

- Data groups in bar graphs achieve (with similar accuracy) highly significant quicker times (promille-level) than data groups in tables.
- The time divergences between the kinds of presentation grow with rising complexity (in this case with a rising size of groups; significant interaction between the kind of presentation and the complexity on the promille-level).
- The group size 12 was a comparison of 2 dataseries, (for each dataserie a special graph (=juxtaposition). The time divergence between bar graph and table achieved an effect size of .79.

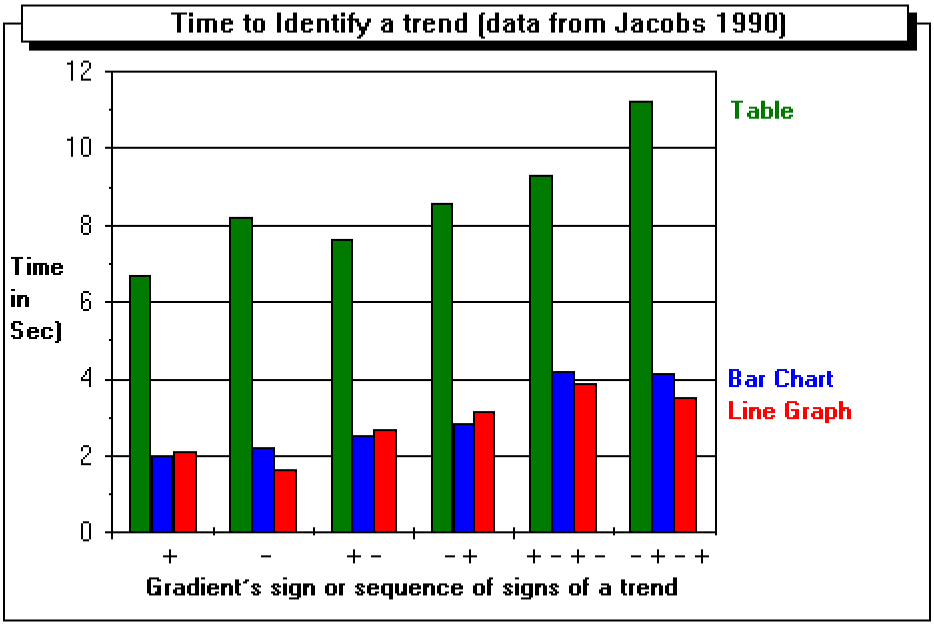
Proved Superiority of the Graphs to the Table (from Jacobs 1990)

The 40 experimental subjects were asked (analogue to the course of Experiment 1) to identify the gradient s sign or sequence of signs of a trend. 21 presentations were shown to each person (3 kinds of presentation x 7 possible trends).

Means concerning the Identifying described above for all trend types (n=40):

	error in %	time in sec
line graph	1	**2.82**
bar chart	2	**2.98**
table	11	**8.61**

Means concerning the Identifying described above for each presentation and each trend types (n=40):



Important Results:

- Both graphical presentations achieve in contrast to the table a significant higher percentage of correct answers (Comparison line graph vs table: Wilcoxon: z =3.2 , p<.01; Comparison bar chart vs table: Wilcoxon: z =3.2 , p<.01).
- The subjects need significantly more time to identify the sign(s) with a table than with graphs. The time divergence between the graphs and the table corresponds an effect size of 2.
- No significant interaction between the kind of presentation and the complexity of the trend s gradient can be proved statistically.
- Bar graph and line graph achieve similar values of accuracy and time.

Literature

Bertin, J. (1974). Graphische Semiologie ((übersetzt von G. Jensch, D. Schade, W. Scharfe).Berlin: Walter de Gruyter.

Carswell, C. M. (1992). Choosing specifiers: An Evaluation of he basic tasks model of graphical perception. Human Factors, 34, 535-554.

Casali, J. G. & Gaylin, K. B. (1988). Selected graph design variables in four interpretation tasks:a microcomputer-based pilot study. Behaviour and Information Technology, 7, 31-49.

Cleveland, W. S. (1985). The Elements of Graphing Data. Monterey, California: Wadsworth advanced Books and Software.

Cleveland, W. S. (1993). Visualizing data. Murray Hill, NJ: AT&T Bell Lab

Coll, R., Thyagaranjan, A. & Chopra, S. (1991). An Experimental Study Comparing the Effectiveness of Computer Graphics Data vers Computer Tabular Data. IEEE Transactions on Systems, Man and Cybernetics, Vol. 21, Nr. 4, pp. 897-900

Coll, J. H. (1992). An experimenal study of the efficacy of tables versus bar graphs with respect to type of task. Information Management 23, pp. 45-51

Coll, R. A., Coll, J. H., Thankur G. (1994). Graphs and Tables A Four-Factor Experiment. Volume 37, Communications of the ACM. No 4, April 1994. 77-86

Croxton, R. E. :. S., H. (1932). Graphic comparison by bars, squares, circles and cubes. Journal of American Statistical Association., 27, 54-60.

Croxton, F. E. :. S., R.E. (1929). Bar charts vs. circle diagrams. Journal of American Statistical Association, 22, 473-482.

Culbertson, H. M. & Powers, R. D. (1959). A study of graph comprehension difficulties. AV Communication Review, 7, 97-100.

Feliciana S.D., Powers, R. D. & Kearl, B. E. (1963). The presentation of statistical information. AV Communication Review, 11, 32-39.

Jacobs, B. (1989). Schnelligkeit und Genauigkeit beim Abschätzen von Größenwerten aus einem Säulendiagramm. Saarbrücken: Arbeitsberichte des Medienzentrums der Universität des Saarlandes, Nr. 2.

Jacobs, B. (1990). Ein Vergleich der Auswirkungen graphischer und tabellarischer Präsentationsformen auf die Schnelligkeit und Genauigkeit beim Erkennen und Interpretieren statistischer Daten. Saarbrücken: Arbeitsberichte des Medienzentrums der Universität des Saarlandes, Nr. 3.

Jacobs, B. (1994). Graphische vs. tabellarische Präsentation von statistischen Daten. Zeitschrift für Pädagogische Psychologie, 8, 73-84.
(Kurzfassung von Jacobs 1989 und 1990)

Jacobs, B. (1994). Der Einfluß von Graphtyp und Graphanordnung auf das Graphverstehen bei der Analyse von Verläufen. Saarbrücken. Arbeitsberichte des Medienzentrums der Universität des Saarlandes, Nr. 13. (Umfangreiche Darstellung von Graphikexperiment 1 **; Postscript (45 KB)** ; **pdf-Datei (103 KB)**)

Jacobs, B. (1995). Experimentelle Analysen zur Wahrnehmung von Kurvenverläufen und Kurvenvergleichen in Säulendiagramm und Liniendiagramm unter Superposition und Juxtaposition. Saarbrücken. Arbeitsberichte des Medienzentrums der Universität des Saarlandes, Nr. 15.

(Umfangreiche Darstellung von Graphikexperiment 2 **; Postscript (152 KB)** ;
pdf-Datei (414 KB))

Jacobs, B. (1995). Globale Vergleiche, lokale Vergleiche und Größenschätzungen in Liniendiagramm und Säulendiagramm unter Superposition und Juxtaposition. Saarbrücken.
Arbeitsberichte des Medienzentrums der Universität des Saarlandes, Nr. 16.
(Umfangreiche Darstellung von Graphikexperiment 3 **; Postscript (120 KB)** ;
pdf-Datei (260 KB))

Jacobs, B. (1995). Die Wahrnehmung besonderer Relationen in Säulendiagramm und Liniendiagramm unter Superposition und Juxtaposition. Saarbrücken.
Arbeitsberichte des Medienzentrums der Universität des Saarlandes, Nr. 17.
(Umfangreiche Darstellung von Graphikexperiment 4 **; Postscript (208 KB)**;
pdf-Datei (625 KB))

Kosslyn, St. M. (1989). Understanding Charts and Graphs. Applied Cognitive Psychology, 3, 185-226.

Kosslyn, St. M. (1994). Elements of Graph Design. New York: W.H. Freeman and Company.

Lewandowsky S. & Spence, I. (1989). Discriminating strata in scatterplots. Journal of American Statistical Association, 84, 682-688.

Lohse, G. L. (1993). A cognitive model for understanding graphical perception, Human-Computer Interaction, 8(4), 353-388.

Lohse, G. L. (1993). Eye Movement-Based Analyses of Graphs and Tables: The next Generation. Proceedings of the Fourteenth International Conference on Information Systems, Orlando, FL. 213-224.

Macdonald-Ross, M. (1977). How numbers are shown: A review of research on the presentation of quantitative data in texts. AV Communication, 25, 359-404.

Maichle, U. (1994). Cognitive Processes in Understanding Line Graphs.
In W. Schnotz & R. W. Kulhavy (Hrsg.), Comprehension of Graphics (S. 207-226). Amsterdam: North-Holland.

Meyer, J. (1996) Performance with tables and graphs. Effects of training and a Visual Model. submitted for publication. paper sent to me with the date 28.Oct 1996

Meyer, J. Shinar, D. & Leiser, D. (1996, submitted for publication in Human Factors) Multiple factors that determine performance with tables und graphs.

Milroy, R. & Poulton E.C. (1978). Labelling Graphs for Improved Reading Speed. Ergonomics, 21, 55-61.

Peterson, L. &. Schramm, W. (1954). How accurately are different kinds of graphs are read ?. AV Communication Review, 2, 178-189.

Pinker, St. (1990). A theory of graph comprehension. In R. Freedle (Hrsg.), Artifical intelligence and the future of testing (S. 73-126). Hillsdale, NJ: Erlbaum.

Poulton, E. C. (1985). Geometric illusions in reading graphs. Perception & Psychophysics, 37, 543-548.

Rinck, M. (1989). Die Strukturierung von Wissen durch statistische Graphen. Marburg: Dissertation.

Schnotz, W. (1994). Wissenserwerb mit logischen Bildern. In B. Weidenmann (Hrsg.), Wissenserwerb mit Bildern (S. 95-147). Bern: Hans Huber.

Schutz, H. G. (1961a). An evaluation of formats for graphic trend displays- Experiment II. Human Factors, 3, 99-107.

Schutz, H. G. (1961b). An Evaluation of Methods for Presentation of Graphic Multiple Trends - Experiment III. Human Factors, 3, 108-119.

Simkin, D. & Hastie, R. (1987). An Information-Processing Analysis of Graph Perception. Journal of the American Statistical Association, Vol. 82, No. 398, 454 - 465.

Spence, I. (1990). Visual Psychophysics of Simple Graphical Elements. Journal of Experimental Psychology: Human Perception and Performance, 19, 683-692.

Spence, I. & Lewandowsky, S. (1991). Displaying Proportions and Percentages:. Applied Cognitive Psychology, 5, 61-77.

Travis, D. (1991). Effective color displays: Theory and Practice. New York: Academic Press.

Tremmel, L. (1992). Untersuchungen zu optimalen Symbolen in graphischen Darstellungen. Hamburg: Verlag Dr. Kovac.

Wainer, H. (1992). Understanding graphs and tables. Educational Researcher, 21, 14-23.

Washburne, J. N. (1927). An experimental study of various graphs,tabular and textual methods of presenting quantitative material. Journal of Educational Psychology, 18, 361-376 und 465-476.

Titel of Project: Experimental Analyses of the Graphical Presentation of Data in Line Graphs and Bar Charts in Superposition and Juxtaposition
Art of Project: Project of the Medienzentrum
period of the Project (german version) ...1993-1995
State of the Project: finished

Coworker of the Project:

Function	Name of coworker	specific task
Project leader:	Bernhard Jacobs	Idea, Conception, Organization and responsibleness
Scientific Assistent:	Bernhard Jacobs	Realizing the idea and the conception. Developing the study reports and the html-Pages
Programmer:	Bernhard Jacobs	Programming of all Experiments
Scientific Helper:	Bernhard Jacobs	Help to install the programs, to organize the Experiments, help by data analysis and so on.
English Translator	Rowan Smith Assistent translator: Patric Beuthen	Translation of the Experiments 1,2,4 and some other parts. Thanks to Jürgen Müller-Ney for helping to organize the translation

Thanks to

- **the developer of the software TOOLBOOK 1.5** (Asymetrix), who had stand by the platform for programming of the Experiments
- **the Toolbook-Mailinglist Toolb-l** for some valuable tips.
- **the experimental Subjects** for their participation without paying.
- **Rowan Smith and Patric Beuthen** for the english translation.
- **the german tax-payer**, who financed my job and so made the project possible.