Wikipedia:

<https://en.wikipedia.org/wiki/Comparative>

In linguistics, the **comparative** is a syntactic construction that serves to express a comparison between two (or more) entities or groups of entities in quality, or degree. See [comparison (grammar)](https://en.wikipedia.org/wiki/Comparison_(grammar)) for an overview of comparison, as well as positive and [superlative](https://en.wikipedia.org/wiki/Superlative) degrees of comparison.

<https://en.wikipedia.org/wiki/Comparison_(grammar)>

**Comparison** is a feature in the [morphology](https://en.wikipedia.org/wiki/Morphology_(linguistics)) or [syntax](https://en.wikipedia.org/wiki/Syntax_(linguistics)) of some [languages](https://en.wikipedia.org/wiki/Language), whereby [adjectives](https://en.wikipedia.org/wiki/Adjective) and [adverbs](https://en.wikipedia.org/wiki/Adverb) are [inflected](https://en.wikipedia.org/wiki/Inflection) or modified to indicate the relative degree of the property defined by the adjective or adverb. The [comparative](https://en.wikipedia.org/wiki/Comparative) expresses a comparison between two (or more) entities or groups of entities in quality, quantity, or degree; the **superlative** is the form of an adverb or adjective that is the greatest degree of a given descriptor.

Morphological comparison:

Morphological comparison uses the [suffixes](https://en.wikipedia.org/wiki/Suffix) *-er* (the "comparative") and *-est* (the "superlative"). These inflections are of [Germanic](https://en.wikipedia.org/wiki/Germanic_languages) origin and are [cognate](https://en.wikipedia.org/wiki/Cognate) with the [Latin](https://en.wikipedia.org/wiki/Latin) suffixes -*ior* and -*issimus* and Ancient Greek -*īōn* and -*istos*. They are typically added to shorter words, words of [Anglo-Saxon](https://en.wikipedia.org/wiki/Old_English_language) origin, and borrowed words which have been fully assimilated into the English vocabulary. Usually the words which take these [inflections](https://en.wikipedia.org/wiki/Inflection) have fewer than three syllables.

This system also contains a number of irregular forms, some of which, like "good", "better", and "best", contain [suppletive](https://en.wikipedia.org/wiki/Suppletion" \o "Suppletion) forms. These irregular forms include:

|  |  |  |
| --- | --- | --- |
| **Positive** | **Comparative** | **Superlative** |
| good | better | best |
| well |
| bad, evil | worse | worst |
| ill |
| far | farther | farthest |
| further | furthest |
| little | less(er) | least |
| many, much | more | most |

**Syntactic comparison**

The second system of comparison in English appends the [grammatical particles](https://en.wikipedia.org/wiki/Grammatical_particle) "more" and "most", themselves the irregular comparatives of "many" and "much", to the adjective or adverb being modified. This series can be compared to a system containing the [diminutives](https://en.wikipedia.org/wiki/Diminutive) "less" and "least".

This system is most commonly used with words of French or Latin derivation; with adjectives and adverbs formed with suffixes other than *-ly* (e.g., "beautiful"); and with longer, technical, or infrequently used words. For example:

|  |  |  |
| --- | --- | --- |
| **Positive** | **Comparative** | **Superlative** |
| beautiful | more beautiful | most beautiful |

Trend analysis

<https://en.wikipedia.org/wiki/Trend_analysis>

**Trend analysis** is the widespread practice of collecting information and attempting to spot a pattern. In some fields of study, the term "trend analysis" has more formally defined meanings.[[1]](https://en.wikipedia.org/wiki/Trend_analysis#cite_note-1)[[2]](https://en.wikipedia.org/wiki/Trend_analysis#cite_note-2)[[3]](https://en.wikipedia.org/wiki/Trend_analysis#cite_note-3)

Although trend analysis is often used to predict future events, it could be used to estimate uncertain events in the past, such as how many ancient kings probably ruled between two dates, based on data such as the average years which other known kings reigned.

## **Statistics[**[**edit**](https://en.wikipedia.org/w/index.php?title=Trend_analysis&action=edit&section=2)**]**

In [statistics](https://en.wikipedia.org/wiki/Statistics), trend analysis often refers to techniques for extracting an underlying pattern of behavior in a [time series](https://en.wikipedia.org/wiki/Time_series) which would otherwise be partly or nearly completely hidden by [noise](https://en.wikipedia.org/wiki/Noise). If the [trend](https://en.wikipedia.org/wiki/Trend_estimation) can be assumed to be linear, trend analysis can be undertaken within a formal [regression analysis](https://en.wikipedia.org/wiki/Regression_analysis), as described in [Trend estimation](https://en.wikipedia.org/wiki/Trend_estimation). If the trends have other shapes than linear, trend testing can be done by non-parametric methods, e.g. Mann-Kendall test, which is a version of [Kendall rank correlation coefficient](https://en.wikipedia.org/wiki/Kendall_rank_correlation_coefficient). For testing and visualization of nonlinear trends also [smoothing](https://en.wikipedia.org/wiki/Smoothing) can be used.

## **Trend analysis in text analysis[**[**edit**](https://en.wikipedia.org/w/index.php?title=Trend_analysis&action=edit&section=3)**]**

Trend analysis can be also used for word usage, how words change in the frequency of use in time (diachronic analysis), in order to find [neologisms](https://en.wikipedia.org/wiki/Neologisms) or [archaisms](https://en.wikipedia.org/wiki/Archaisms). It relates to [diachronic linguistics](https://en.wikipedia.org/wiki/Diachronic_linguistics), a field of linguistics which examines how languages change over time. Google provides tool [Google Trends](https://en.wikipedia.org/wiki/Google_Trends) to explore how particular terms are trending in internet searches. On the other hand, there are tools which provide diachronic analysis for particular texts which compare word usage in each period of the particular text (based on [timestamped](https://en.wikipedia.org/wiki/Timestamp) marks), see e.g. [Sketch Engine](https://en.wikipedia.org/wiki/Sketch_Engine) diachronic analysis (trends).[[6]](https://en.wikipedia.org/wiki/Trend_analysis#cite_note-6)

# **Line chart**

<https://en.wikipedia.org/wiki/Line_chart>

A **line chart** or **line graph** is a type of [chart](https://en.wikipedia.org/wiki/Chart) which displays information as a series of data points called 'markers' connected by straight [line](https://en.wiktionary.org/wiki/line) segments.[[1]](https://en.wikipedia.org/wiki/Line_chart#cite_note-1) It is a basic type of [chart](https://en.wikipedia.org/wiki/Chart) common in many fields. It is similar to a [scatter plot](https://en.wikipedia.org/wiki/Scatter_plot) except that the measurement points are ordered (typically by their x-axis value) and joined with straight line segments. A line chart is often used to visualize a trend in data over intervals of time – a [time series](https://en.wikipedia.org/wiki/Time_series) – thus the line is often drawn chronologically. In these cases they are known as [run charts](https://en.wikipedia.org/wiki/Run_chart).[[2]](https://en.wikipedia.org/wiki/Line_chart#cite_note-2)

# **Time series**

<https://en.wikipedia.org/wiki/Time_series>

A **time series** is a series of [data points](https://en.wikipedia.org/wiki/Data_point) indexed (or listed or graphed) in time order. Most commonly, a time series is a [sequence](https://en.wikipedia.org/wiki/Sequence) taken at successive equally spaced points in time. Thus it is a sequence of [discrete-time](https://en.wikipedia.org/wiki/Discrete-time) data. Examples of time series are heights of ocean [tides](https://en.wikipedia.org/wiki/Tides), counts of [sunspots](https://en.wikipedia.org/wiki/Sunspots), and the daily closing value of the [Dow Jones Industrial Average](https://en.wikipedia.org/wiki/Dow_Jones_Industrial_Average).

Time series are very frequently plotted via [line charts](https://en.wikipedia.org/wiki/Line_chart). Time series are used in [statistics](https://en.wikipedia.org/wiki/Statistics), [signal processing](https://en.wikipedia.org/wiki/Signal_processing), [pattern recognition](https://en.wikipedia.org/wiki/Pattern_recognition), [econometrics](https://en.wikipedia.org/wiki/Econometrics), [mathematical finance](https://en.wikipedia.org/wiki/Mathematical_finance), [weather forecasting](https://en.wikipedia.org/wiki/Weather_forecasting), [earthquake prediction](https://en.wikipedia.org/wiki/Earthquake_prediction), [electroencephalography](https://en.wikipedia.org/wiki/Electroencephalography), [control engineering](https://en.wikipedia.org/wiki/Control_engineering), [astronomy](https://en.wikipedia.org/wiki/Astronomy), [communications engineering](https://en.wikipedia.org/wiki/Communications_engineering), and largely in any domain of applied [science](https://en.wikipedia.org/wiki/Applied_science) and [engineering](https://en.wikipedia.org/wiki/Engineering) which involves [temporal](https://en.wikipedia.org/wiki/Time) measurements.

**Time series *analysis*** comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. **Time series *forecasting*** is the use of a [model](https://en.wikipedia.org/wiki/Model_(abstract)) to predict future values based on previously observed values. While [regression analysis](https://en.wikipedia.org/wiki/Regression_analysis) is often employed in such a way as to test theories that the current values of one or more independent time series affect the current value of another time series, this type of analysis of time series is not called "time series analysis", which focuses on comparing values of a single time series or multiple dependent time series at different points in time.[[1]](https://en.wikipedia.org/wiki/Time_series#cite_note-1) [Interrupted time series](https://en.wikipedia.org/wiki/Interrupted_time_series) analysis is the analysis of interventions on a single time series.

# **Run chart**

<https://en.wikipedia.org/wiki/Run_chart>

A **run chart**, also known as a **run-sequence plot** is a graph that displays observed data in a [time sequence](https://en.wikipedia.org/wiki/Time_sequence). Often, the data displayed represent some aspect of the output or performance of a manufacturing or other business process. It is therefore a form of [line chart](https://en.wikipedia.org/wiki/Line_chart).

Boek Statistics for people who

Line Charts A line chart should be used when you want to show a trend in the data at equal intervals. Here are some examples of when you might want to use a line chart: Number of cases of mononucleosis (mono) per season among college students at three state universities Change in student enrollment over the school year Number of travelers on two different airlines for each quarter.

Uit H4  
Line, to show change over time, A line chart, which tracks changes over time

Bar Charts A bar chart is identical to a column chart, but in this chart, categories are organized on the y-axis and values are shown horizontally on the x-axis.

Uit H4

Bar or column, to display categorical data. A bar chart, which is especially good at displaying categorical or nominal data

Grammar overview:

8 ADJECTIVES AND ADJECTIVE PHRASES

**8.1 Two major functions of adjectives**

Most adjectives can be either **attributive** or **predicative**:

[39]     i         **Attributive**     *a hot day*, *some new DVDs*, *this excellent play*, *lonely people*

ii         **Predicative**   *It's hot. These look new. I found it excellent. They seem lonely.*

Attributive adjectives are pre-head Modifiers in noun phrase structure; predicative adjectives are Predicative Complements in clause structure (see Section5.5).[[6]](http://www.lel.ed.ac.uk/grammar/overview.html" \l "_ftn6" \o ")

There are, however, some adjectives that are restricted to one or other of these functions:

[40]     i         **Attributive-only**       *the main speaker*, *a mere child*, *the only problem*, *my own car*

ii         **Never-attributive**     *I'm afraid. She's asleep. He looks content. It's liable to flood.*

**8.2 Gradability and grade**

The most central adjectives are **gradable**: they denote properties that can apply in varying degrees. As such, they can be modified by adverbs of degree and (under conditions relating to length and form) be inflected for **comparative** (e.g. *taller*) and **superlative**(e.g. *tallest*) **grade**:

[41]     i         **Degree modification**    *very good*, *quite hot*, *rather young*, *too old*, *incredibly bad*

ii         **Inflection for grade**      *hotter*, *younger*, *older*, *better*; *hottest*, *youngest*, *oldest*, *best*

Gradable adjectives that don't inflect mark comparative and superlative degree by means of the adverbs *more* and *most* respectively: *more intelligent*, *most intelligent*.

There are also a good number of adjectives that denote non-scalar properties and hence are **non-gradable**: *alphabetical order*, *the chief difficulty*, *the federal government*, *her right eye*, *third place*. Some adjectives, moreover, can be used in two different senses, one gradable, the other non-gradable (and usually the more basic). In *The door is open*, for example, *open* is non-gradable, but in *You should be more open with us* it is gradable.

**8.3 The structure of adjective phrases**

Adjective phrases consist of an adjective as Head, alone or accompanied by one or more Dependents, which may be Complements or Modifiers:

[42]     i         **Complements***good at chess*, *grateful for your help*, *fond of animals*, *keen on golf*,

*glad that you liked it*, *unsure what had happened*, *eager to help*

ii         **Modifiers***very bad*, *morally wrong*, *this good*, *most useful*, *much better*, *two*

*days long*, *a bit old*, *cautious to excess*, *dangerous in the extreme*

The Complements are preposition phrases or subordinate clauses; in the former case the adjective selects a particular preposition to head the Complement: ***fond*** takes *of*, ***keen*** takes *on*, and so on. The Modifiers are adverbs (e.g. *very*), determinatives (*this*), noun phrases (*two days*) or post-Head prepositional phrases. Adjective phrases containing post-Head Dependents cannot normally be used attributively: *He's good at chess*, but not \**a good at chess schoolboy*.

9 ADVERBS AND ADVERB PHRASES.

**9.1 Adverbs in relation to adjectives**

The majority of adverbs are derived from adjectives by adding the suffix @*ly*: *common* - *commonly*, *rare* - *rarely*, etc. There are a good number of adverbs not formed in this way, some of them very common (e.g. *almost*, *always*, *often*, *quite*, *rather*, *soon*, *too*, *very*), but these are normally recognisable as adverbs by virtue of being replaceable by ones with the @*ly* suffix: compare *It's very good* and *It's extremely good*; *She always wins* and *She frequently wins*; *It'll be over soon*and*It'll be over shortly*, and so on.

The major difference between adverbs and adjectives has to do with their functions. We have seen that adjectives function attributively or predicatively, but adverbs do not normally occur in these functions: compare attributive *a successful meeting*, not \**a successfully meeting*, and predicative *The meeting wassuccessful*, not \**The meeting was successfully*. Adverbs function as Modifier to a wide range of word or phrase classes, as illustrated in [43], where underlining marks the modifying adverb and capitals what it modifies:

[43]               **Adverb modifying:**

  i         **Verb**                    *She SPOKE clearly.                        She PLAYED well.*

ii         **Adjective**             *It's a remarkably GOOD play*            *It looks very GOOD*

iii         **Adverb**                *He spoke virtually INAUDIBLY*.        *They almost NEVER reply.*

iv         **Determinative**     *Nearly ALL copies were sold.*           *Too FEW copies were printed.*

v         **Prep phrase**        *She is completely IN CONTROL.      It's quite BEYOND BELIEF.*

vi         **Rest of clause***Surprisingly EVERYONE AGREED  Frankly, IT'S USELESS*.

In general adverbs that can modify adjectives and other adverbs can also modify verbs, but there are some exceptions, most notably *very* and *too* (in the sense `excessively'). Compare *He's very FOND of her* and \**He very LOVES her* (we need *He loves her very MUCH*).

A few adverbs inflect for grade (*soon*, *sooner*, *soonest*), but for the most part comparatives and superlatives are marked by *more* and *most*: *more carefully*, *most carefully*.

**9.2 The structure of adverb phrases**

The structure of adverb phrases is broadly similar to that of adjective phrases, but simpler: in particular, very few adverbs license complements.

[44]     i         **Complements***Luckily for me, it rained. We handled it similarly to the others.*

ii         **Modifiers**             *She sang very well.* *It won't end that soon.* *We left a bit late.*

*13.1.3 Comparative clauses*

Comparative clauses generally function as Complement to the prepositions *as* and *than*:

[59]     i   a.    *I'm as ready as I ever will be*.                   b.   *As was expected, Sue won easily.*

ii   a.    *More people came than I'd expected.*     b.   *He has more vices than he has virtues.*

The distinctive property of such clauses is that they are structurally incomplete relative to main clauses: there are elements understood but not overtly expressed. In [ia] and [iia] there's a missing Complement and in [ib] a missing Subject. Even in [iib] there's a missing Dependent in the Object noun phrase, for the comparison is between how many vices he has and **how many** virtues he has. The fact that there's some kind of understood quantifier here is reflected in the fact that we can't insert an overt one: \**He has more vices than he has ten virtues.*

Data categories

<https://onlinecourses.science.psu.edu/stat500/node/113/>

Distinguishing between categorical (qualitative) variables and quantitative variables is a basic and intergral part of applied statistics as the methods to analyze these data are very different. Sometimes, when one codes surveys, you would code male as 1 and female as 2. Beware, gender is qualitative: there are two different classes. 1 and 2 just denote two different symbols for gender and there is no ordering between these two symbols when used to denote male and female. Another example is team assignments. For your team project, I will call the teams: Team 1 , Team 2 etc. The team a student belongs to is again qualitative. In statistics, as in most languages, we sometime call the same thing by different names. So qualitative is also called nominal, or categorical.

How can one graph qualitative variables?  Two common choices are pie chart and bar chart. Please pay attention that even though histogram also have bars sticking up, they are used to describe the frequency for quantitative variables; bar chart is reserved to describe graphs that show frequency of categorical variables.

* Graphs for a Categorical Variable
  + - Pie Chart
    - Bar Chart
* Graphs for a Single Quantitative Variable
  + - Dot Plot
    - Frequency Histogram and Relative Frequency Histogram
    - Stem-and-Leaf Diagram
    - Time Plot
    - Boxplot or Box-and-Whisker Plot

**1. Qualitative (Categorical) :** Data that serves the function of a name only. For example, for coding purposes, you may assign Male as 0, Female as 1. The numbers 0 and 1 stand only for the two categories and there is no order between them. Categorical values may be:

* Binary – where there are two choices, e.g. Male and Female;
* Ordinal – where the names imply levels with hierarchy or order of preference, e.g. level of education
* Nominal – where no hierarchy is implied, e.g. political party affiliation.

**2. Quantitative:** Data that takes on numerical values that has a measure of distance between them. Quantitative values can be discrete,  or “counted” as in the number of people in attendance, or continuous or “measured” as in the weight or height of a person.

**Additional examples of both** **include**:

* Number of females in this class (Quantitative, Discrete)
* Nationality (Categorical, nominal)
* Amount of milk in a 1 gallon container (Quantitative, Continuous)
* Sex of students (even if coded as M = 0, F = 1) (Categorical, Binary)

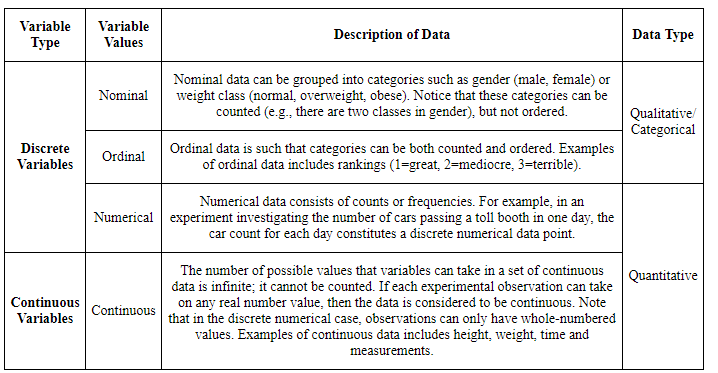
# Displaying Data - Charts and Plots

<http://www4.ncsu.edu/~aelarsen/vet/display/index.html>

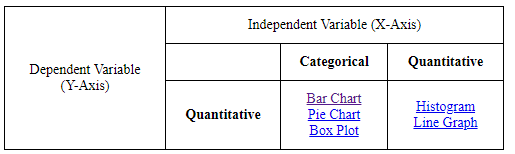
The first step to displaying your data is to identify the independent and dependent variables in your experiment. The values of an *independent variable* are chosen and often controlled by the investigator, who then observes the effect of each independent variable on a *dependent variable*. The dependent variable takes different values in response to the values of the independent variable that are chosen by the investigator.

For example, say a research team wants to evaluate a new anti-anxiety drug. They *choose* four doses of the drug to administer to patients and observe each patients’ blood pressure and heart rate. Based on their observations, they can decide which dosage is most effective for mitigating anxiety. In this scenario, the independent variable is the drug; investigators controlled the values of this variable, or dosage levels, for each patient. The dependent variables are heart rate and blood pressure; the values of each of these variables could not be controlled by the investigators, but instead resulted from different dosages of the drug, or independent variable. This experiment would allow the team to evaluate the effectiveness of the drug by manipulating drug dosages and observing the response of blood pressure and heart rate.

The table below describes the possible types of variables and a description of the values of those variables, or data.



Now that you have identified what types of independent and dependent variables you have in your experiment, you can decide which plots are best for displaying your results using the table below. Click on each plot to see a detailed description and instructions on how to make the display in JMP.



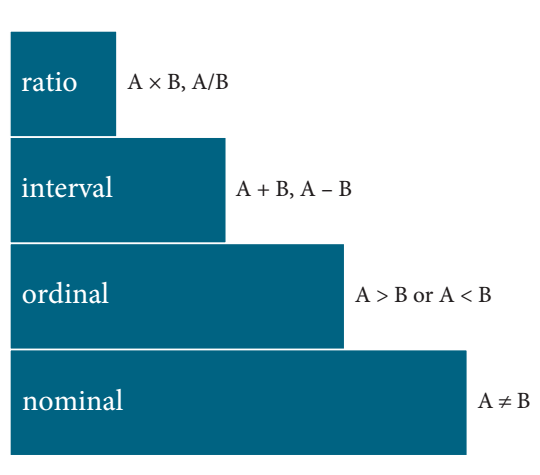
### Bar Charts

Bar charts are useful for comparing frequencies or counts of one category to the next in a qualitative discrete data set. In such a chart, the height of each bar represents the frequency of the corresponding category. Depending on preference, bar charts can be displayed with vertical or horizontal bars. Each bar has equal width.

### Line Graphs

Line graphs are helpful for illustrating trends in continuous data. For example, a line graph could communicate whether or not a patient’s survival time increases or decreases with a given treatment. It is important to be sure that each observation is collected at equal intervals.

Slides hoorcollege week 2 RUG – statistiek 1



# Data: Continuous vs. Categorical

https://eagereyes.org/basics/data-continuous-vs-categorical

Data comes in a number of different types, which determine what kinds of mapping can be used for them. The most basic distinction is that between continuous (or quantitative) and categorical data, which has a profound impact on the types of visualizations that can be used.

The main distinction is quite simple, but it has a lot of important consequences. Quantitative data is data where the values can change continuously, and you cannot count the number of different values. Examples include weight, price, profits, counts, etc. Basically, anything you can measure or count is quantitative.

Categorical data, in contrast, is for those aspects of your data where you make a distinction between different groups, and where you typically can list a small number of categories. This includes product type, gender, age group, etc.

Both quantitative and categorical data have some finer distinctions, but I will ignore those for this posting. What is more important, is: why do those make a difference for visualization?

## **Correlative conjunctions**

**Coordinating conjunctions**, also called **coordinators**, are conjunctions that join, or [coordinate](https://en.wikipedia.org/wiki/Coordination_(linguistics)), two or more items (such as words, main clauses, or sentences) of equal syntactic importance. In English, the [mnemonic](https://en.wikipedia.org/wiki/Mnemonic) acronym *FANBOYS* can be used to remember the coordinators *for*, *and*, *nor*, *but*, *or*, *yet*, and *so*.[[4]](https://en.wikipedia.org/wiki/Conjunction_(grammar)#cite_note-4) These are not the only coordinating conjunctions; various others are used, including[[5]](https://en.wikipedia.org/wiki/Conjunction_(grammar)#cite_note-5):ch. 9[[6]](https://en.wikipedia.org/wiki/Conjunction_(grammar)#cite_note-6):p. 171 "and nor" (British), "but nor" (British), "or nor" (British), "neither" ("They don't gamble; neither do they smoke"), "no more" ("They don't gamble; no more do they smoke"), and "only" ("I would go, only I don't have time"). Types of coordinating conjunctions include cumulative conjunctions, adversative conjunctions, alternative conjunctions, and illative conjunctions.[[7]](https://en.wikipedia.org/wiki/Conjunction_(grammar)#cite_note-7)

Here are some examples of coordinating conjunctions in English and what they do:

* **For** – presents rationale ("They do not gamble or smoke, for they are ascetics.")
* **And** – presents non-contrasting item(s) or idea(s) ("They gamble, and they smoke.")
* **Nor** – presents a non-contrasting negative idea ("They do not gamble, nor do they smoke.")
* **But** – presents a contrast or exception ("They gamble, but they don't smoke.")
* **Or** – presents an alternative item or idea ("Every day they gamble, or they smoke.")
* **Yet** – presents a contrast or exception ("They gamble, yet they don't smoke.")
* **So** – presents a [consequence](https://en.wikipedia.org/wiki/Logical_consequence) ("He gambled well last night, so he smoked a cigar to celebrate.")

## **Correlative conjunctions[**[**edit**](https://en.wikipedia.org/w/index.php?title=Conjunction_(grammar)&action=edit&section=2)**]**

[Correlative](https://en.wikipedia.org/wiki/Correlative) conjunctions work in pairs to join words and groups of words of equal weight in a sentence. There are many different pairs of correlative conjunctions:

* either...or
* not only...but (also)
* neither...nor
* both...and
* whether...or
* just as...so
* the...the
* as...as
* as much...as
* no sooner...than
* rather...than

Examples:

* You *either* do your work *or* prepare for a trip to the office. (Either do or prepare)
* He is *not only* handsome *but also* brilliant. (Not only A but also B)
* *Neither* the basketball team *nor* the football team is doing well.
* *Both* the cross country team *and* the swimming team are doing well.
* You must decide *whether* you stay *or* you go.
* *Just as* many Americans love basketball, *so* many Canadians love ice hockey.
* *The* more you practice dribbling, *the* better you will be at it.
* Football is *as* fast *as* hockey (is (fast)).
* Football is *as much* an addiction *as* it is a sport.
* *No sooner* did she learn to ski *than* the snow began to thaw.
* I would *rather* swim *than* surf.

# **Conjunction (grammar)**

…

Uit codebook bar eerste rij:

|  |  |  |  |
| --- | --- | --- | --- |
|  | (Discrete (variable)) comparison | (<https://en.wikipedia.org/wiki/Comparative>)  In linguistics, the **comparative** is a syntactic construction that serves to express **a comparison between two (or more) entities or groups of entities in quality, degree or quantity (**Since discrete variables can be counted and ordered - [http://www4.ncsu.edu /~aelarsen/vet/display/index.html](http://www4.ncsu.edu/~aelarsen/vet/display/index.html))**.**  Comparisons use terms like more/less, more/fewer, higher/lower, larger/smaller, stronger/weaker; they tend to refer to discrete values. (source: bars and lines article) | X scored 20% of the points, Y scored also 20% of the points and Z scored 25% of the points during the match.  X scored 300 points, as many as Y and Z scored 55 points during the match. |

**https://en.wikipedia.org/wiki/Text\_segmentation**

**Text segmentation** is the process of dividing written text into meaningful units, such as words, [sentences](https://en.wikipedia.org/wiki/Sentence_(linguistics)), or [topics](https://en.wikipedia.org/wiki/Topic_(linguistics)). The term applies both to [mental processes](https://en.wikipedia.org/wiki/Mental_process) used by humans when reading text, and to artificial processes implemented in computers, which are the subject of [natural language processing](https://en.wikipedia.org/wiki/Natural_language_processing). The problem is non-trivial, because while some written languages have explicit word boundary markers, such as the word spaces of written English and the distinctive initial, medial and final letter shapes of [Arabic](https://en.wikipedia.org/wiki/Arabic_language), such signals are sometimes ambiguous and not present in all written languages.

Compare [speech segmentation](https://en.wikipedia.org/wiki/Speech_segmentation), the process of dividing speech into linguistically meaningful portions.

### Topic segmentation[[edit](https://en.wikipedia.org/w/index.php?title=Text_segmentation&action=edit&section=5)]

*Main articles:*[*Topic analysis*](https://en.wikipedia.org/w/index.php?title=Topic_analysis&action=edit&redlink=1)*and*[*Document classification*](https://en.wikipedia.org/wiki/Document_classification)

Topic analysis consists of two main tasks: topic identification and text segmentation. While the first is a simple [classification](https://en.wikipedia.org/wiki/Machine_learning) of a specific text, the latter case implies that a document may contain multiple topics, and the task of computerized text segmentation may be to discover these topics automatically and segment the text accordingly. The topic boundaries may be apparent from section titles and paragraphs. In other cases, one needs to use techniques similar to those used in [document classification](https://en.wikipedia.org/wiki/Document_classification).

Segmenting the text into [topics](https://en.wikipedia.org/wiki/Topic_(linguistics)) or [discourse](https://en.wikipedia.org/wiki/Discourse) turns might be useful in some natural processing tasks: it can improve information retrieval or speech recognition significantly (by indexing/recognizing documents more precisely or by giving the specific part of a document corresponding to the query as a result). It is also needed in [topic detection](https://en.wikipedia.org/w/index.php?title=Topic_detection&action=edit&redlink=1) and tracking systems and [text summarizing](https://en.wikipedia.org/wiki/Text_summarization) problems.

Many different approaches have been tried:[[2]](https://en.wikipedia.org/wiki/Text_segmentation#cite_note-2)[[3]](https://en.wikipedia.org/wiki/Text_segmentation#cite_note-3) e.g. [HMM](https://en.wikipedia.org/wiki/Hidden_Markov_model), [lexical chains](https://en.wikipedia.org/w/index.php?title=Lexical_chains&action=edit&redlink=1), passage similarity using word [co-occurrence](https://en.wikipedia.org/wiki/Co-occurrence), [clustering](https://en.wikipedia.org/wiki/Cluster_analysis), [topic modeling](https://en.wikipedia.org/wiki/Topic_modeling), etc.

It is quite an ambiguous task – people evaluating the text segmentation systems often differ in topic boundaries. Hence, text segment evaluation is also a challenging problem.

# Dependent and independent variables

<http://onlinestatbook.com/2/introduction/variables.html>

Variables are properties or characteristics of some event, object, or person that can take on different values or amounts (as opposed to constants such as π that do not vary). When conducting research, experimenters often manipulate variables. For example, an experimenter might compare the effectiveness of four types of antidepressants. In this case, the variable is "type of antidepressant." When a variable is manipulated by an experimenter, it is called an independent variable. The experiment seeks to determine the effect of the independent variable on relief from depression. In this example, relief from depression is called a dependent variable. In general, the independent variable is manipulated by the experimenter and its effects on the dependent variable are measured.

Boek statistiek:

Variables in a study may have different status, which depends on how you model the relationships between them. The outcome variable, or the one which changes as a function of some other parameters of interest, is called the response, or dependent variable. The variables that influences the outcome are called explanatory, or independent variables.

<https://www.khanacademy.org/math/pre-algebra/pre-algebra-equations-expressions/pre-algebra-dependent-independent/a/dependent-and-independent-variables-review>

Independent variable

An independent variable is a variable that represents a quantity that is being manipulated in an experiment.

x is often the variable used to represent the independent variable in an equation.

Example:

You are doing chores to earn your allowance. For each chore you do, you earn \$3$3dollar sign, 3.

What is the independent variable?

The independent variable is the amount of chores you do because this is the variable you have control over.

Dependent variables

A dependent variable represents a quantity whose value depends on how the independent variable is manipulated.

Y is often the variable used to represent the dependent variable in an equation.

Example:

Let's use the same context.

You are doing chores to earn your allowance. For each chore you do, you earn \$3$3dollar sign, 3.

What is the dependent variable?

The dependent variable is the amount of money you earn because the amount of money you earn depends on how many chores you do.