I'm not a robot

Image by Author | Ideogram Pareto charts are simple yet powerful visualization artifacts based on statistical concepts. They are frequently used in data analysis, decision-making processes, and quality control. Interested in understanding what they are, what they are used for, and how to visually interpret them? This article has you covered. What are Pareto Charts and What are They Used For? Pareto charts are the combination of a bar chart and a line graph, and they are primarily used to describe the relative importance of different categories or groups in your data. They are named after the well-known Pareto Principle or “80/20 rule”, according to which a small number of causes (about 20%) usually account for the majority of effects (about 80%). In the original scenario that led to the formulation of this principle, Vilfredo Pareto observed that 80% of the land in Italy was owned by 20% of the country’s population. In terms of sales, the Pareto principle would state that about 80% of sales come from approximately 20% of customers, and so on. Aligning with this principle that focuses on the “vital few” rather than the “trivial many”, Pareto charts shows absolute frequencies of data categories in bars ranked in descending order, as well as total cumulative frequencies in the data displayed by a line. This combined view provides an invaluable approach for identifying and prioritizing opportunities, problems, or priorities in several domains, such as: Quality control: Pareto charts help pinpoint defects contributing to most complaints in products. Customer service: they are used to analyze customer complaints and shift efforts to the most common “pain points”. Inventory management: they help identify items that incur the majority of costs. Let’s see a first visual example of a Pareto chart. The following chart displays a dataset of customer complaints categorized by type. Pareto chart of customer complaints categorized by type Here’s how to interpret this chart: Just like in plain bar charts, the vertical bars represent frequencies of data by categories. In this example, bars show the frequency of each complaint type, with late delivery being the most common cause for customer complaints, followed by damaged items. The cumulative percentage line shows the cumulative percentage of complaints as we move from left to right across the previously listed categories. The 80% threshold represented by a horizontal dashed line is used to easily identify the causes 80% of complaints come from. In this scenario, by looking at this intersection between the cumulative line and this threshold, we can state that roughly 80% of the complaints come from the two most common categories: late delivery and damaged items. The key insight is, therefore, that by addressing just two of the existing issues -late delivery and damaged items- we can resolve nearly 80% of customer complaints. This insight is used to leverage resource prioritization or make more informed decisions upon data, with the ultimate goal of improving processes or solving key problem areas or pain points. Another example Pareto chart used for quality control processes shows the frequencies and cumulative counts of defect types in manufactured components. Pareto chart of defect types for quality control in a manufacturing plant In this case, the frequency distribution is slightly more even, such that a moderately larger range of possible causes (defect types) account for 80% of the effects (the faults). Scratches, cracks, color issues, misalignments, and rough edges, would all be the major issues to further investigate in order to address the 80% of faults in manufactured components. Best Practices and Limitations of Pareto Charts To build a useful Pareto chart, it is important to keep in mind these key guidelines: Ordering categories in descending order of frequencies is vital to easily identify the most significant contributor causes. Calculate and use cumulative percentages -rather than cumulate frequencies- in the line to clearly illustrate the contribution of categories (issues) from most to least frequent ones. Stick to the ‘80’ in the 80/20 rule, but use the 80% threshold to identify categories contributing to 80% of the effects to guide decisions. It is normal if sometimes the contributing causes are a bit more than 20% of existing types, as seen in the second example above. Your compass will be the 80% threshold in the last instance. Regarding the limitations of Pareto charts, while they are very useful for identifying major contributors to effects, they do not take into account the context, failing to capture why certain issues occur. They also oversimplify the data, which might be an issue when dealing with complex datasets where some relationships or interactions matter between factors. Wrapping Up This article provided an introductory guide to understanding and leveraging Pareto charts as an easy-to-interpret yet useful data visualization tool in processes and scenarios like quality control, customer management, and other contexts where problem-solving and decision-making assets are needed to identify causes for common problems. Despite having existed for a while, they are still a must-tool for data analysts to assist them in where to focus their attention on the most critical areas. Probability distribution Pareto Type I Probability density functionPareto Type I probability density functions for various α {\displaystyle \alpha } with x m = 1. {\displaystyle x_{\mathrm {m} }=1.} As α → ∞ , {\displaystyle \alpha \rightarrow \infty ,} the distribution approaches δ (x − x m) , {\displaystyle \delta (x-x_{\mathrm {m} }),} where δ {\displaystyle \delta } is the Dirac delta function. Cumulative distribution functionPareto Type I cumulative distribution functions for various α {\displaystyle \alpha } with x m = 1. {\displaystyle x_{\mathrm {m} }=1.} Parameters x m > 0 {\displaystyle x_{\mathrm {m} }>0} scale (real) α > 0 {\displaystyle \alpha >0} shape (real)Support x ∈ [x m , ∞) {\displaystyle x\in [x_{\mathrm {m} },\infty)} PDF α x m α x α + 1 {\displaystyle {\frac {\alpha x_{\mathrm {m} }^{\alpha }}{x^{\alpha +1}}}} CDF 1 − (x m x) α {\displaystyle 1-\left({\frac {x_{\mathrm {m} }}{x}}\right)^{\alpha }} Quantile x m (1 − p) − 1 α {\displaystyle x_{\mathrm {m} }(1-p)}^{-{\frac {1}{\alpha }}}} Mean { ∞ for α ≤ 1 α x m α − 1 for α > 1 {\displaystyle {\begin{cases}\infty &{\text{for }}\alpha \leq 1\\{\dfrac {\alpha x_{\mathrm {m} }}{\alpha -1}}&{\text{for }}\alpha >1\end{cases}}} Median x m 2 α {\displaystyle x_{\mathrm {m} }{\sqrt[{\alpha }]{2}}} Mode x m {\displaystyle x_{\mathrm {m} }} Variance { ∞ for α ≤ 2 x m 2 α (α − 1) 2 (α − 2) for α > 2 {\displaystyle {\begin{cases}\infty &{\text{for }}\alpha \leq 2\\{\dfrac {x_{\mathrm {m} }^{2}\alpha }{(\alpha -1)^{2}(\alpha -2)}}&{\text{for }}\alpha >2\end{cases}}} Skewness 2 (1 + α) α − 3 α − 2 α for α > 3 {\displaystyle {\frac {2(1+\alpha)}{\alpha -3}}{\sqrt {\frac {\alpha -2}{\alpha }}}{\text{ for }}\alpha >3} Excess kurtosis 6 (α 3 + α 2 − 6 α − 2) α (α − 3) (α − 4) for α > 4 {\displaystyle {\frac {6(\alpha ^{3}+\alpha ^{2}-6\alpha -2)}{\alpha (\alpha -3)(\alpha -4)}}{\text{ for }}\alpha >4} Entropy log ((x m α) e 1 + 1 α) {\displaystyle \log \left(\left({\frac {x_{\mathrm {m} }}{\alpha }}\right)\,e^{1+{\tfrac {1}{\alpha }}}\right)} MGF does not existCF α (− i x m t) α Γ (− α , − i x m t) {\displaystyle \alpha (-ix_{\mathrm {m} }t)^{\alpha }\Gamma (-\alpha ,-ix_{\mathrm {m} }t)} Fisher information I (x m , α) = [α 2 x m 2 0 0 1 α 2] {\displaystyle {\mathcal {I}}(x_{\mathrm {m} },\alpha)={\begin{bmatrix}{\dfrac {\alpha ^{2}}{x_{\mathrm {m} }^{2}}}&0\\0&{\dfrac {1}{\alpha ^{2}}}\end{bmatrix}}} Expected shortfall x m α (1 − p) 1 α (α − 1) {\displaystyle {\frac {x_{m}\alpha }{(1-p)^{\frac {1}{\alpha }}(\alpha -1)}}} [1] The Pareto distribution, named after the Italian civil engineer, economist, and sociologist Vilfredo Pareto,[2] is a power-law probability distribution that is used in description of social, quality control, scientific, geophysical, actuarial, and many other types of observable phenomena; the principle originally applied to describing the distribution of wealth in a society, fitting the trend that a large portion of wealth is held by a small fraction of the population.[3][4] The Pareto principle or "80:20 rule" stating that 80% of outcomes are due to 20% of causes was named in honour of Pareto, but the concepts are distinct, and only Pareto distributions with shape value (α) of log 4 5 ≈ 1.16 precisely reflect it. Empirical observation has shown that this 80:20 distribution fits a wide range of cases, including natural phenomena[5] and human activities.[6][7] If X is a random variable with a Pareto (Type I) distribution,[8] then the probability that X is greater than some number x, i.e., the survival function (also called tail function), is given by F ¯ (x) = Pr (X > x) = { (x m x) α x ≥ x m , 1 x < x m , {\displaystyle {\overline {F}}(x)=\Pr(X>x)={\begin{cases}\left({\frac {x_{\mathrm {m} }}{x}}\right)^{\alpha }&x\geq x_{\mathrm {m} },\\1&x 2. {\displaystyle \operatorname {Var} (X)={\begin{cases}\infty &\alpha \in (1,2],\\\left({\frac {x_{\mathrm {m} }}{\alpha -1}}\right)^{2}{\frac {\alpha }{\alpha -2}}&\alpha >2.\end{cases}}} (If α ≤ 2, the variance does not exist.) The raw moments are μ n ′ = { ∞ α ≤ n , α x m n α − n α > n . {\displaystyle \mu _{n}'={\begin{cases}\infty &\alpha \leq n,\\{\frac {\alpha x_{\mathrm {m} }^{n}}{\alpha -n}}&\alpha >n.\end{cases}}} The moment generating function is only defined for non-positive values t ≤ 0 as M (t ; α , x m) = E [e t X] = α (− x m t) α Γ (− α , − x m t) {\displaystyle M\left(t;\alpha ,x_{\mathrm {m} }\right)=\operatorname {E} \left[e^{tX}\right]=\alpha (-x_{\mathrm {m} }t)^{\alpha }\Gamma (-\alpha ,-x_{\mathrm {m} }t)} M (0 , α , x m) = 1. {\displaystyle M\left(0,\alpha ,x_{\mathrm {m} }\right)=1.} Thus, since the expectation does not converge on an open interval containing t = 0 {\displaystyle t=0} we say that the moment generating function does not exist. The characteristic function is given by φ (t ; α , x m) = α (− i x m t) α Γ (− α , − i x m t) , {\displaystyle \varphi (t;\alpha ,x_{\mathrm {m} })=\alpha (-ix_{\mathrm {m} }t)^{\alpha }\Gamma (-\alpha ,-ix_{\mathrm {m} }t),} where Γ(a, x) is the incomplete gamma function. The parameters may be solved for using the method of moments.[9] The conditional probability distribution of a Pareto-distributed random variable, given the event that it is greater than or equal to a particular number x 1 {\displaystyle x_{1}} exceeding x m {\displaystyle x_{\text{m}}} , is a Pareto distribution with the same Pareto index α {\displaystyle \alpha } but with minimum x 1 {\displaystyle x_{1}} instead of x m {\displaystyle x_{\text{m}}} : Pr (X ≥ x | X ≥ x 1) = { (x 1 x) α x ≥ x 1 , 1 x < x 1 . {\displaystyle {\text{Pr}}(X\geq x|X\geq x_{1})={\begin{cases}\left({\frac {x_{1}}{x}}\right)^{\alpha }&x\geq x_{1},\\1&x 1 {\displaystyle \alpha >1} } is proportional to x 1 {\displaystyle x_{1}} : E (X | X ≥ x 1) ∝ x 1 . {\displaystyle {\text{E}}(X|X\geq x_{1})\propto x_{1}.} In case of random variables that describe the lifetime of an object, this means that life expectancy is proportional to age, and is called the Lindy effect or Lindy's Law.[10] Suppose X 1 , X 2 , X 3 , ... {\displaystyle X_{1},X_{2},X_{3},\dotsc } are independent identically distributed random variables whose probability distribution is supported on the interval [x m , ∞) {\displaystyle [x_{\text{m}},\infty)} for some x m > 0 {\displaystyle x_{\text{m}}>0} . Suppose that for all n {\displaystyle n} , the two random variables min { X 1 , ... , X n } {\displaystyle \min\{X_{1},\dotsc ,X_{n}\}} and (X 1 + ⋯ + X n) / min { X 1 , ... , X n } {\displaystyle (X_{1}+\dotsb +X_{n})/\min\{X_{1},\dotsc ,X_{n}\}} are independent. Then the common distribution is a Pareto distribution.[citation needed] The geometric mean (G) is[11] G = x m exp (1 α) . {\displaystyle G=x_{\text{m}}\exp \left({\frac {1}{\alpha }}\right).} The harmonic mean (H) is[11] H = x m (1 + 1 α) . {\displaystyle H=x_{\text{m}}\left(1+{\frac {1}{\alpha }}\right).} The characteristic curved 'long tail' distribution, when plotted on a linear scale, masks the underlying simplicity of the function when plotted on a log-log graph, which then takes the form of a straight line with negative gradient: It follows from the formula for the probability density function that for x ≥ xm, log f X (x) = log (α x m α x α + 1) = log (α x m α) − (α + 1) log x . {\displaystyle \log f_{X}(x)=\log \left(\alpha {\frac {x_{\mathrm {m} }^{\alpha }}{x^{\alpha +1}}}\right)=\log(\alpha x_{\mathrm {m} }^{\alpha })-(\alpha +1)\log x.} Since α is positive, the gradient −(α + 1) is negative. See also: Generalized Pareto distribution There is a hierarchy [8][12] of Pareto distributions known as Pareto Type I, II, III, IV, and Feller–Pareto distributions.[8][12][13] Pareto Type IV contains Pareto Type I–III as special cases. The Feller–Pareto[12][14] distribution generalizes Pareto Type IV. The Pareto distribution hierarchy is summarized in the next table comparing the survival functions (complementary CDF). When μ = 0, the Pareto distribution Type II is also known as the Lomax distribution.[15] In this section, the symbol xm, used before to indicate the minimum value of x, is replaced by σ. Pareto distributions F ¯ (x) = 1 − F (x) {\displaystyle {\overline {F}}(x)=1-F(x)} Support Parameters Type I [x σ] − α {\displaystyle \left[{\frac {x}{\sigma }}\right]^{-\alpha }} x ≥ σ {\displaystyle x\geq \sigma } σ > 0 , α {\displaystyle \sigma >0,\alpha } Type II [1 + x − μ σ] − α {\displaystyle \left[1+{\frac {x-\mu }{\sigma }}\right]^{-\alpha }} x ≥ μ {\displaystyle x\geq \mu } μ ∈ R , σ > 0 , α {\displaystyle \mu \in \mathbb {R} ,\sigma >0,\alpha } Lomax [1 + x σ] − α {\displaystyle \left[1+{\frac {x}{\sigma }}\right]^{-\alpha }} x ≥ 0 {\displaystyle x\geq 0} σ > 0 , α {\displaystyle \sigma >0,\alpha } Type III [1 + (x − μ σ) 1 / γ] − 1 {\displaystyle \left[1+\left({\frac {x-\mu }{\sigma }}\right)^{1/\gamma }\right]^{-1}} x ≥ μ {\displaystyle x\geq \mu } μ ∈ R , σ , γ > 0 , α {\displaystyle \mu \in \mathbb {R} ,\sigma ,\gamma >0,\alpha } The shape parameter α is the tail index, μ is location, σ is scale, γ is an inequality parameter. Some special cases of Pareto Type (IV) are P (I V) (σ , σ , 1 , α) = P (I) (σ , α) , {\displaystyle P(IV)(\sigma ,\sigma ,1,\alpha)=P(I)(\sigma ,\alpha),} P (I V) (μ , σ , 1 , α) = P (I I) (μ , σ , α) , {\displaystyle P(IV)(\mu ,\sigma ,1,\alpha)=P(II)(\mu ,\sigma ,\alpha),} P (I V) (μ , σ , γ , 1) = P (I I I) (μ , σ , γ) . {\displaystyle P(IV)(\mu ,\sigma ,\gamma ,1)=P(III)(\mu ,\sigma ,\gamma).} The finiteness of the mean, and the existence and the finiteness of the variance depend on the tail index α (inequality index γ). In particular, fractional δ-moments are finite for some δ > 0, as shown in the table below, where δ is not necessarily an integer. Moments of Pareto I–IV distributions (case μ = 0) E [X] {\displaystyle \operatorname {E} [X]} Condition E [X δ] {\displaystyle \operatorname {E} [X^{\delta }]} Condition Type I σ α α − 1 {\displaystyle {\frac {\sigma \alpha }{\alpha -1}}} α > 1 {\displaystyle \alpha >1} σ δ α α − δ {\displaystyle {\frac {\sigma ^{\delta }\alpha }{\alpha -\delta }}} δ < α {\displaystyle \delta 1 {\displaystyle \alpha >1} σ δ Γ (α − δ) Γ (1 + δ) Γ (α) {\displaystyle {\frac {\sigma ^{\delta }\Gamma (\alpha -\delta)\Gamma (1+\delta)}{\Gamma (\alpha)}}} 0 < δ < α {\displaystyle 0