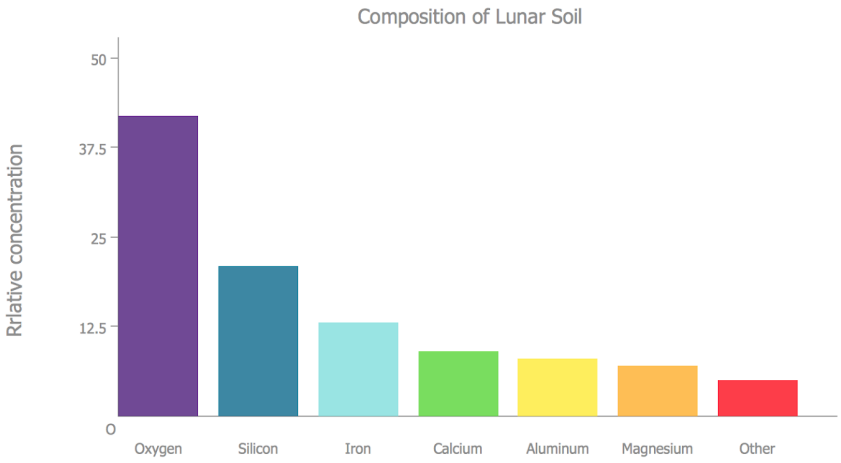
# Different Kinds of Plots

Sijal Dhakal
Knox College
sdhakal@knox.edu

Badri Adhikari
University of Missouri-St. Louis
adhikarib@umsl.edu

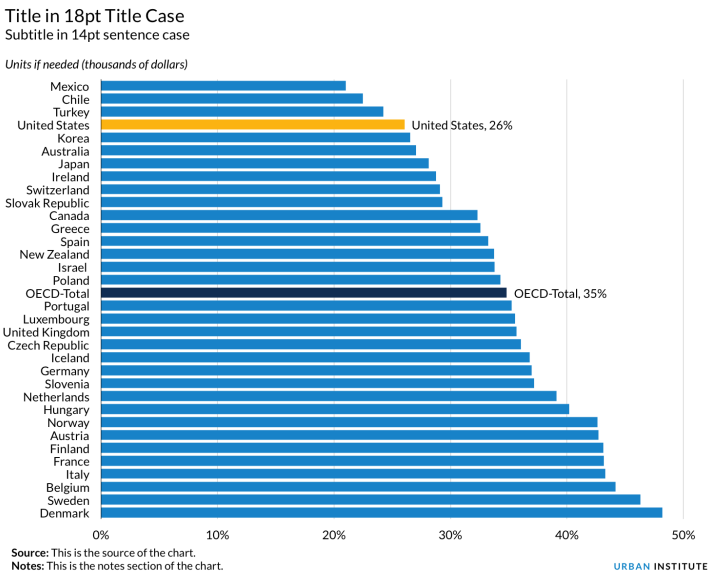
1. Vertical Bar-chart

- This kind of chart can be used when one of the variables is discrete quantitative or when it is qualitative; can be used in both ordinal and nominal (For nominal, see the ex. below).
- Vertical bar charts are often used when we are comparing categories.
- The categories to be compared must be small in number.



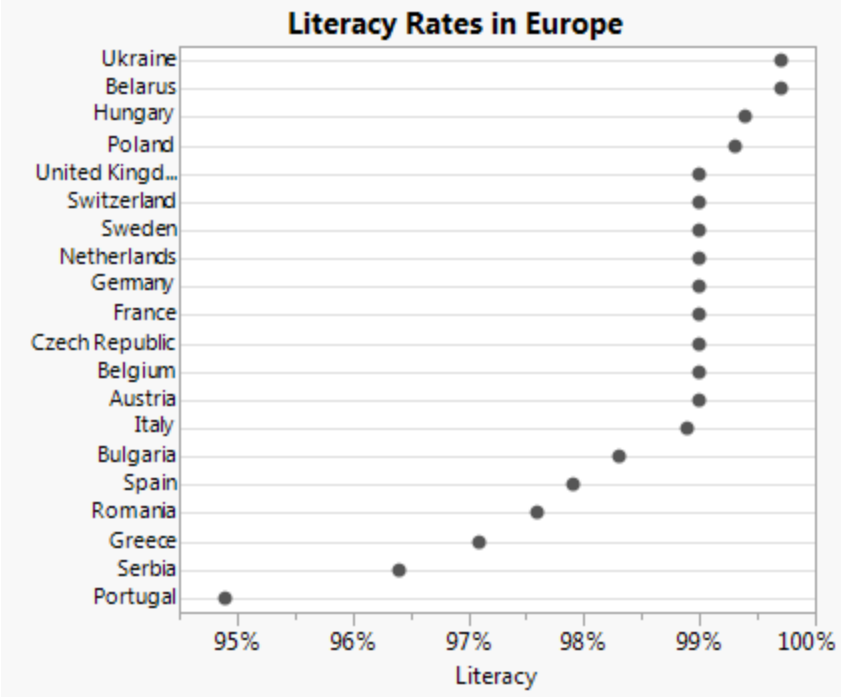

2. Horizontal Bar-chart

- This chart can be used exactly like the vertical bar-chart, except this chart is often used when there are more categories.



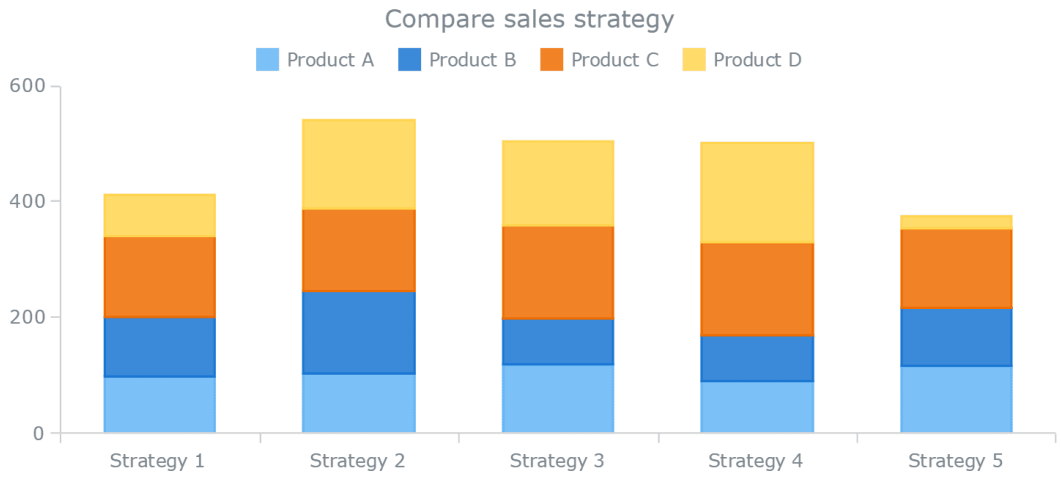

## 3. Dot Plot

- Again , very similar to horizontal and vertical bar charts. Use this chart instead of those two if you have a lot of categories.
- When the difference between the two subcategories is very low, it would be easier for the audience to see the difference if we use a dot plot



## 4. Stacked Bar Chart

- Horizontal Stacked Bar chart is often used when there is a primary variable and a subcategory and the purpose is to compare between primary variables and also between subcategories.
- Use this chart when there are more than two sub categories.
- This bar chart is especially useful if the primary variables all have a fixed sum. For instance, if one wants to find a percentage composition of males and females in different divisions of a workplace. The total percentage in each variable=100%. *What if not a fixed sum?
- Use this chart if you want your audience to focus on the comparison of primary variables before the sub categories, because visually the first thing one notices in this chart is the difference in heights of the main bars (primary variables).
  *If you want your audience to focus on the comparison of the sub-categories instead, use clustered bar charts
- Use different colors to differentiate the subcategories for more clarity




## 5. Clustered Bar Chart

- Qualitative nominal and ordinal variables.
- Clustered Bar chart is also used when there is a primary variable and a subcategory and the purpose is to focus on the comparison between subcategories and to judge their trends (in case of ordinal variables).

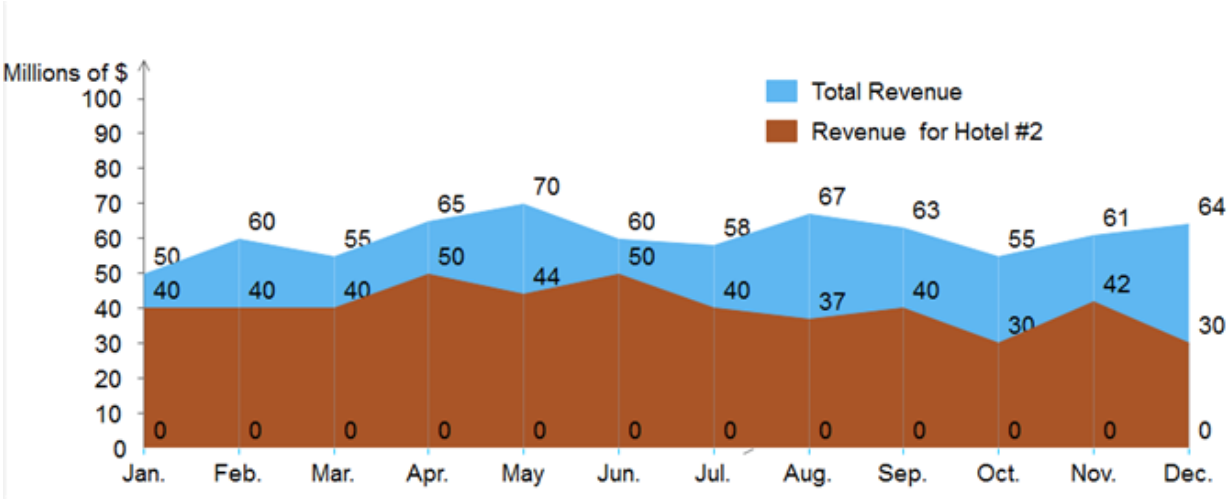
- Choose your primary variables and sub category based on your priority of comparison. If the comparison of subcategories is the main focus/ is more important, make that your subcategory.
- Use different colors to differentiate the subcategories for more clarity.
- Use this when there are few categories.


## 6. Line Chart

- Line charts are often used to compare trends of a variable over time.
- It can also be used to compare between the changes in categorical variables over the same period of time but if the variables have more than one group.
- The data concerned should be continuous for the line chart to be intuitive.

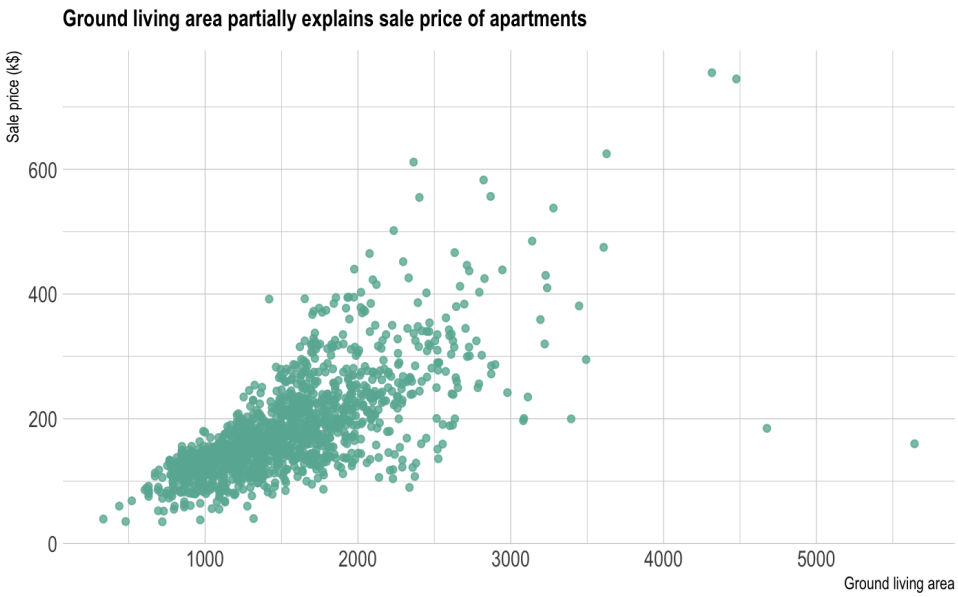
## 7. Area Line Chart

- Area chart is very similar to a line chart but their aptness differs.
- Area charts are especially useful if we want to show the composition of multiple subgroups in a single umbrella group, and the individual change in trends over time.
- The data concerned should be continuous for the area chart to be intuitive.



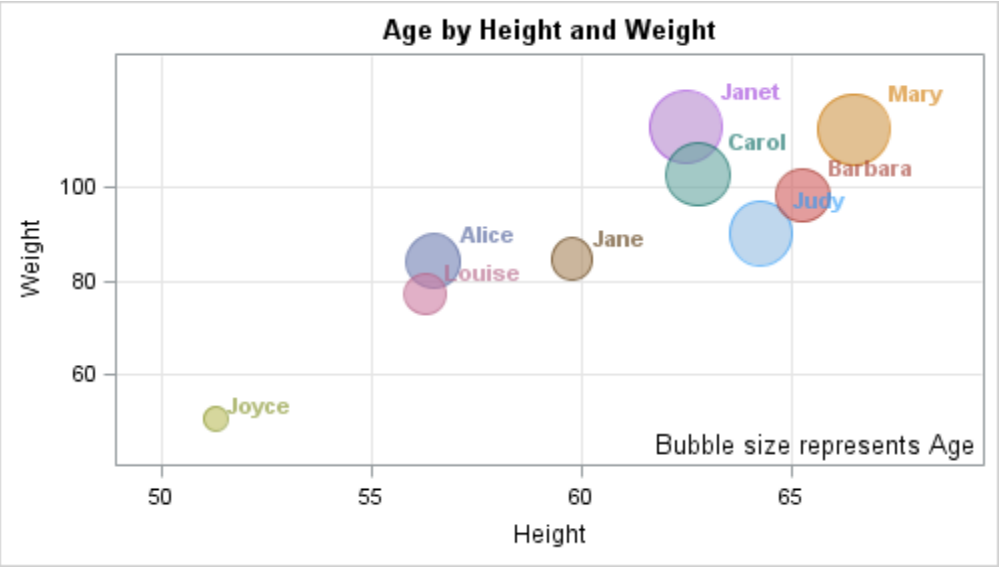

## 8. Scatterplot

- Scatter plot looks similar to line charts but scatter plot is widely used to show the relationship or correlation between two variables.
- The sample (dots) should be discrete and should represent individual pieces of data.
- When looking at a scatter plot, one should be able to distinguish the relationship between the variables concerned, if there is any. (positively or negatively correlated or not correlated at all)
- If one of the two variables is already known to be independent, put that variable in the x-axis.
- Often when there are a lot of data points, overplotting occurs and the data plots seem blotched and unclear. One of the best ways to deal with overlapping is to make the data points partially transparent.



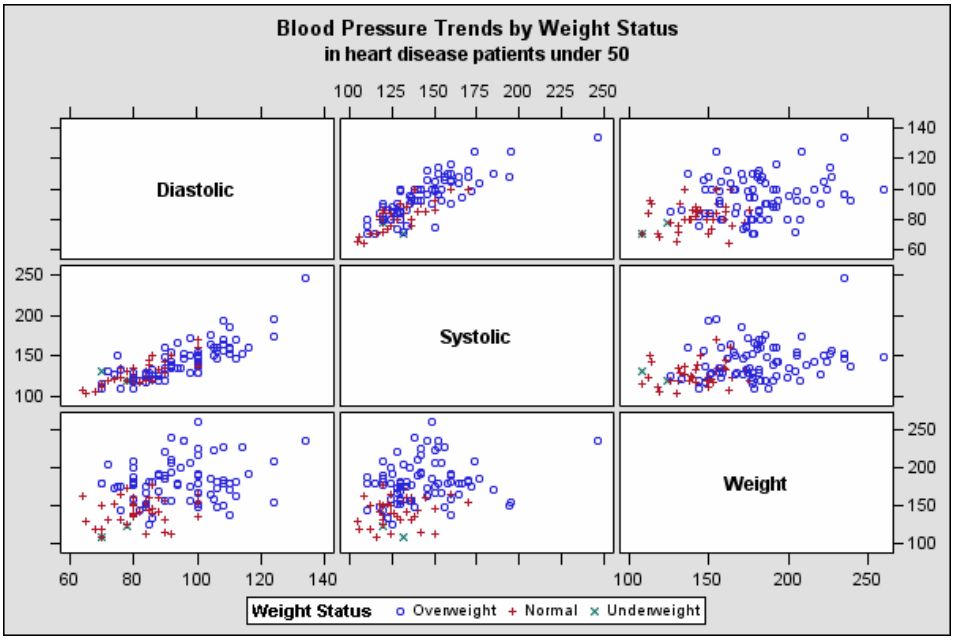

## 9. Bubble Graph

- Bubble graph is very similar to a scatter plot, but it is used to show the relationship between three variables.
- The size of the bubbles depend on the third variable i.e the largest third variable will have the biggest bubble.
- If there are a large number of data points, then the bubble chart may not be effective as the bubbles may overlap which might obscure its accuracy
- If the overlapping occurs, make sure to keep the smaller bubble on top of the large bubble so it does not get obscured by the larger bubble.
- Bubble sizes become more clear with different colors +



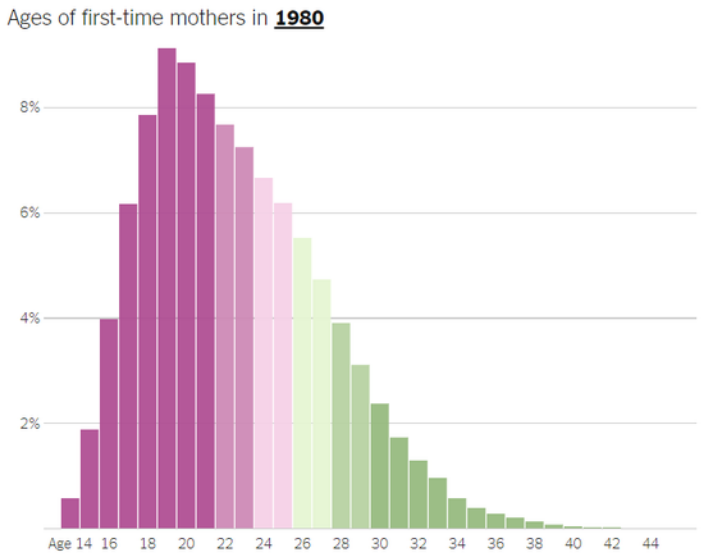

## 10. Scatterplot Matrices

- This plot is very useful if you want to show the relationship between any two variables and among combinations of those variables.
- The scatterplot matrix helps visualize the bivariate relationships, and allows us to analyze many bivariate relationships in one chart.



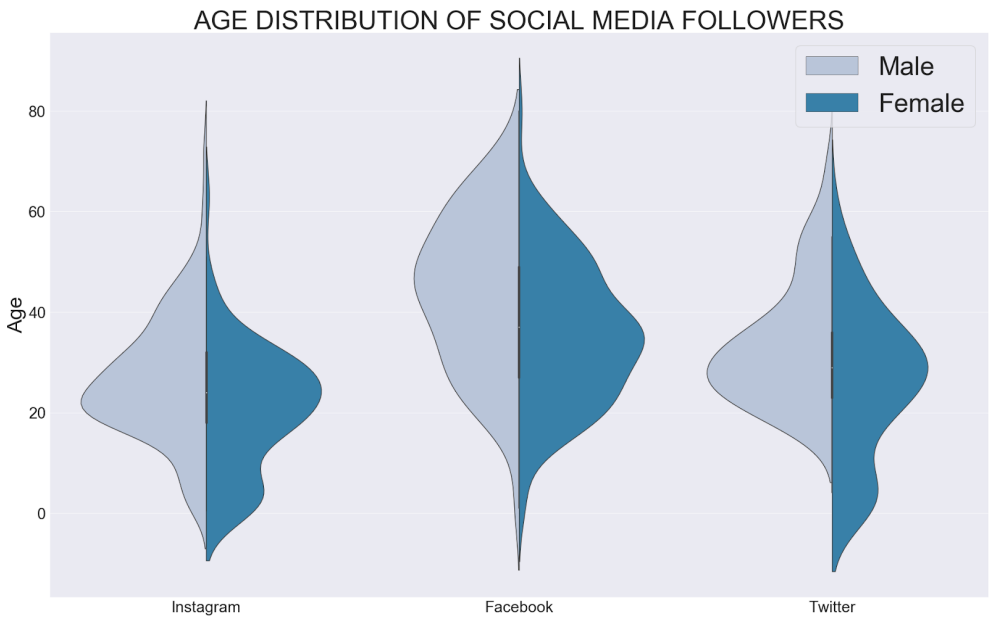
## 11. Histogram

- Looks similar to a bar graph, but this chart can only be used when the data distribution is continuous.
- Histograms are widely used to interpret the frequency distribution of a continuous data and to inspect the skewness, outliers etc



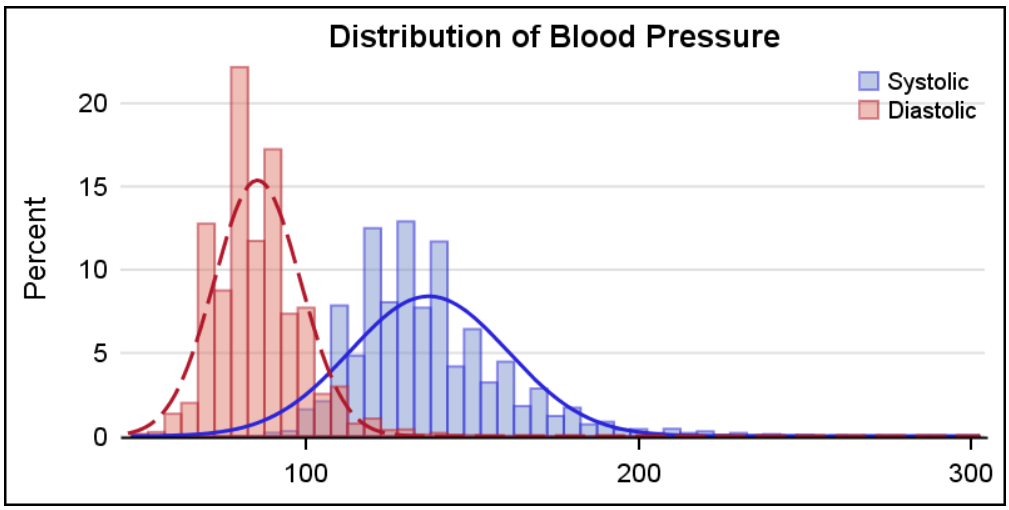

## 12. Violin Chart

- This chart is often used to compare the distribution of two subcategories of qualitative variables. It can also sometimes be used to compare the distribution of two charts over time.
- The chart is also used to show the probability distribution.
- Use this also if you want the audience to know the medians, interquartile range etc.
- Violin charts are especially useful if the data distribution is quantitative and continuous.



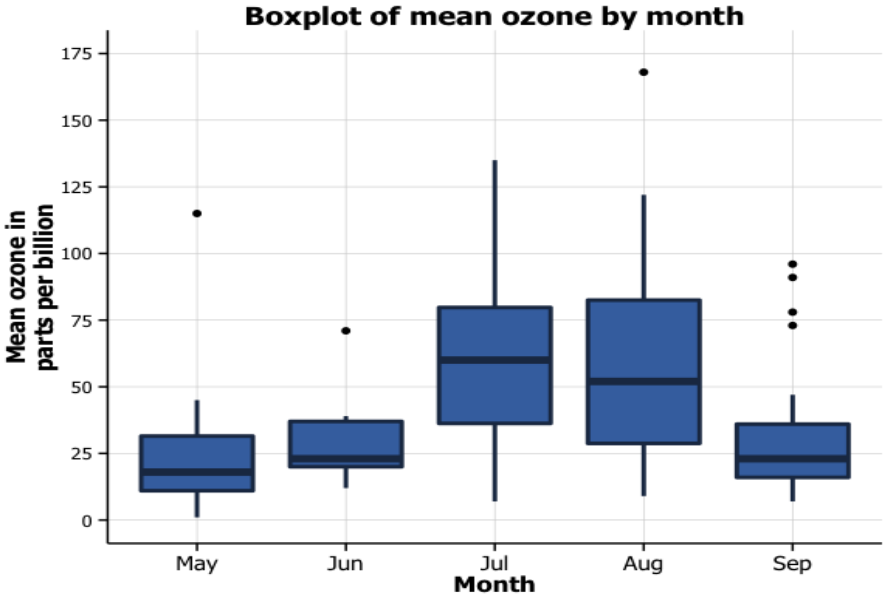

## 11. Stacked line Histogram

- Stacked line histogram is used to compare the sample distribution when the variable has more than one group.
- This chart can be used when there are more than two categories
- The data distribution has to be continuous.



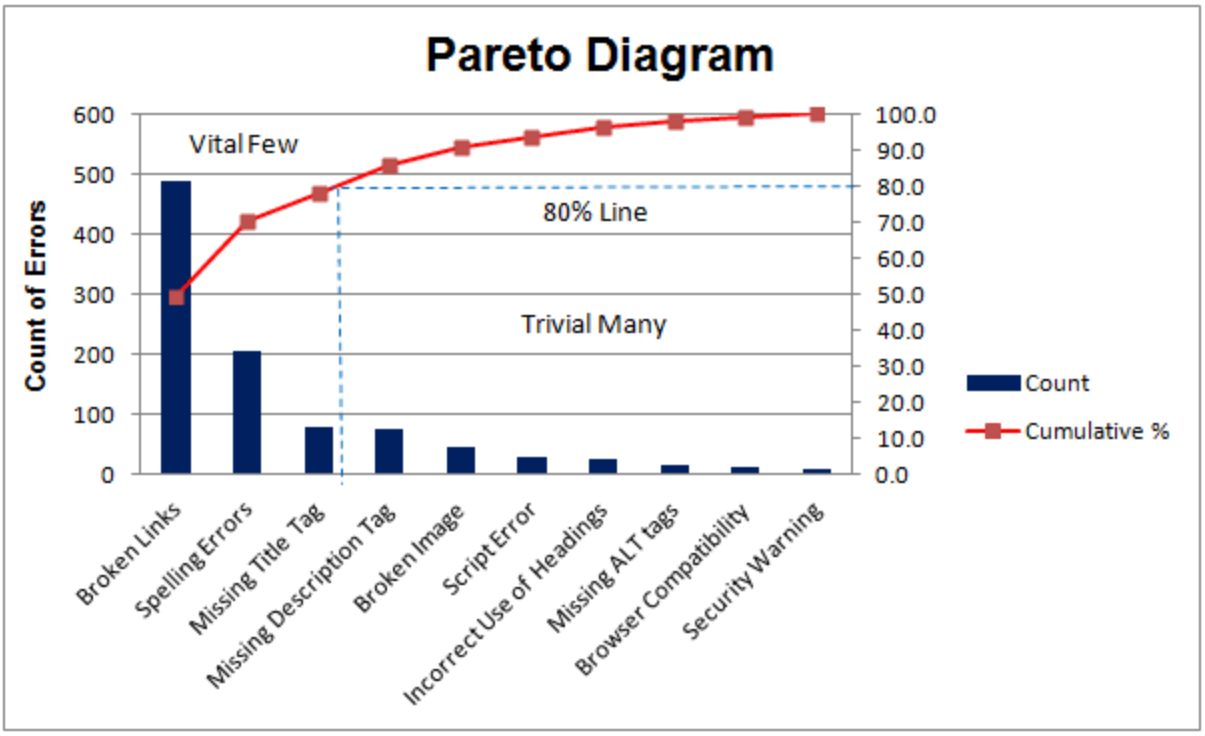
## 13. Box plot

- Box plots are especially useful to compare the central tendency and interquartile ranges across many categories.
- Threshold for determining outliers
- It can also be used to compare the overall range and dispersion of the variables.
- It can also be used to evaluate how far outliers are from its central value.



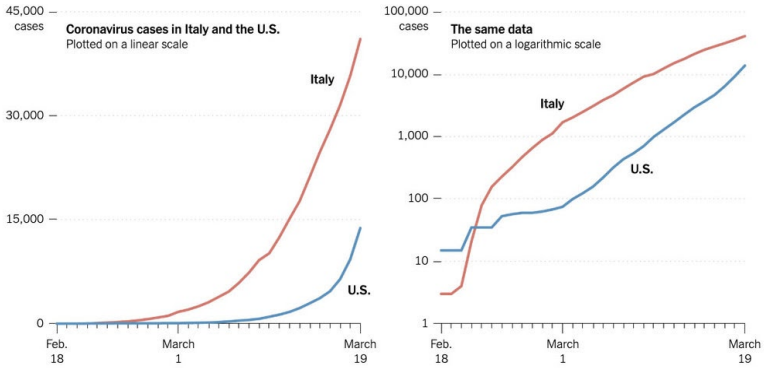

## 14. Pareto Chart

- Pareto Chart has both bar graphs and a line graph. The bar graph shows the individual values and the line graph shows the the cumulative total of the sample
- It is used to depict the relative importance of a defect/problem/error.
- Pareto Chart is especially used by a firm or a company to identify the few problems that have the greatest impact. It uses the Pareto principle (80:20 rule) which contends that for many events, roughly 80% of the effects come from 20% of the causes.
- To amplify this effect, one must always arrange the bar graphs in descending order. And the 80% cutoff should be properly shown.




## When to use LOG

Sometimes charts will be more meaningful if we use logarithmic scales, instead of regular scales. It is most often used when the data is skewed i.e when a majority of the data is populated in either side. Using log in such cases will allow the audience to interpret the graph properly as the large range in the chart will be displayed without small values being compressed down into the bottom of the graph.



Another time when using logarithmic scales are advantageous is when the article talks about the percent change. For example, the distance between $10 and $14 would be equal to the distance between $20 and $28 because they both represent a price increase of 40%. In such cases, using logarithmic scales will turn out to be more accurate and illustrative. Remember: Log transforms result in percentage changes.

However, regardless of how useful log transformation can be, it is important to understand that not everybody can decipher log transformations. So, know your audience.