DataWitzz

Learning: We made it Easy

DataWitzz
Learning:we made it Easy

Ultimate guide to choose correct type of charts

DataWitzz
Learning:we made it Easy

About Me

My name is Ijtaba and I am a Data Analyst having more than 15 years of experience in diversified fields like **Business Analysis, Work force Management(WFM) , Data Analysis, Process Improvement and Machine Learning**.

My first encountered with excel was in 2008 when I was told to enter few paper receipt in excel and it was a pathetic experience. After that I took it as challenge to learn excel. Though I faced a lot obstacle while accumulating the knowledge, but what I also learned that I should prepare an army of excel skilled professional. I have personally trained many subordinates, colleagues, friends and professionals in the journey.

As the time passes interest in **SQL,R & Python** has also grown for data slicing & dicing along with Power **BI, Tableau & OBIEE** for visualization. I also feel that I should be sharing my knowledge with as many people as possible. So I started my **YouTube** Channel for voice and **Blog** for text & visuals to spread my knowledge. It helps me to pursue my passion as the desire to help other get more out of it.

Preface

I will be sharing use of different charts basis different criteria. Usually I found people uses charts randomly and this doesn't help in presenting data in a good manner and sometimes these chart got interpreted wrongly as well. I am a very big fond of visualizing data and using it in a very good way.

Learning we made it Easy

My hope for this e-book is that you share it with as many people as possible, and by sharing the knowledge many more people will start excelling.

Charts are very important tool to visualize our data in such a way that it can be understood and help in taking decision making. Charts compares different sets of data, can show trends, can show data for timeline etc. It is very important to choose correct type of charts to visualize our data. In absence of correct type of chart it could be read in a wrong way.

We can select chart basis data type, basis relationship, composition etc. I am trying to explain below how to choose correct chart type basis data type of these criteria.

What kind of data we have ?

1)If data is **numeric** and it is **univariate**(One numeric variable) then use:

Histogram

- Variable breaks into bins
- The number of observation per bin is represented by the height of the bin
- Don't use more than 3-4 plots in a chart.

Both represents **distribution** of numerical variable

Density Plot

- Uses kernel smoothing to plot values
- Allowing for smoother distributions by smoothing out noise.
- Don't use more than 3-4 plots in a chart.

2) If data is **numeric** and it is **bivariate**(two numeric variable) and it is in order then use:

Connected Scatter Plot

- Show evolution of the numeric variable
- Individual observation are highlighted

Area Plot

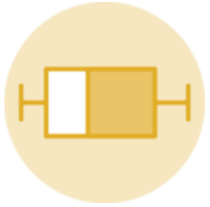
- Show evolution of the numeric variable
- Like a line chart but area under is filled with colour

Line Plot

- Show evolution of the numeric variable
- Data points connected by straight line
- Use dots with line if data point are less

3) If data is **numeric** and it is **bivariate**(two numeric variable) and it is not **ordered** and data points are **few** then use:

Box Plot



- Provides 5 summary data along with outliers
- If amount of data is high then add jitter
- While summarizing some information can be lost.

Histogram

- Variable breaks into bins
- The number of observation per bin is represented by the height of the bin
- Don't use more than 3-4 plots in a chart.

Scatter Plot

- Display relationship between two variable
- Provide correlation coefficient to measure linear relationship.

4) If data is **numeric** and it is **bivariate**(two numeric variable) and it is not **ordered** and data points are **large** then use:

Violin

- The shape represents the density estimate of the variable
- It is close to a box plot but a deeper understanding of the distribution.
- Provides info about ranking and distribution

Density Plot

- Uses kernel smoothing to plot values
- Allowing for smoother distributions by smoothing out noise.
- Don't use more than 3-4 plots in a chart.

Scatter Plot

- Display relationship between two variable
- Provide correlation coefficient to measure linear relationship.

5) If data is **numeric** and it is **trivariate** (three numeric variable) and it is **ordered** then use:

Line Plot

- Show evolution of the numeric variable
- Data points connected by a straight line
- Use dots with line if data point are less

Area Plot

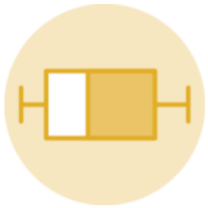
- Show evolution of the numeric variable
- Like a line chart but area under is filled with colour

Stacked Area Plot

- Extension of area chart
- Appropriate to study the evolution of the whole and the relative portions of each group

6) If data is **numeric** and it is **trivariate** (three numeric variable) and it is **not ordered** then use:

Box plot



- Provides 5 summary data along with outliers
- If amount of data is high then add jitter
- While summarizing some information can be lost.

Violin

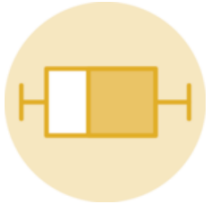
- The shape represents the density estimate of the variable
- It is close to a box plot but a deeper understanding of the distribution.
- Provides info about ranking and distribution

Bubble plot

- Like an scatter plot where third dimension is added and represented by the size of dot.
- Little difficult to interpret.

7) If data is **numeric** and it is **multivariate** (many numeric variable) and it is not **ordered** then use:

Box plot



- Provides 5 summary data along with outliers
- If amount of data is high then add jitter
- While summarizing some information can be lost.

Violin

- The shape represents the density estimate of the variable
- It is close to a box plot but a deeper understanding of the distribution.
- Provides info about ranking and distribution

Dendrogram

- Is a network structure
- Root node that gives birth to several nodes connected by edges or branches
- The last node of the hierarchy are called leaves.

What kind of data we have ?

DataWitzz
Learning:we made it Easy

1)If data is **Categoric** and total variable is **One** then use:

Bar Plot

- Shows relationship between a numeric and categorical variable.
- Size of the bar represents its numeric variable

Doughnut

- *A pie plot with an area of the center cut out.*
- It shows proportion of whole

Pie

- *A circular graphic divided into slices to illustrate numerical proportion.*

Tree map

- Displays hierarchical data as set of nested rectangle.
- Each rectangle represents proportions.

2)If data is **Categoric** and total variable is **two or more** and it **is nested** then use:

Sunburst

- Also called as multilevel pie chart
- It is very similar to Treemap except it uses a radial layout

Bar Plot

- Shows relationship between a numeric and categorical variable.
- Size of the bar represents its numeric variable

Tree Map

- Displays hierarchical data as set of nested rectangle.
- Each rectangle represents proportions.

Dendrogram

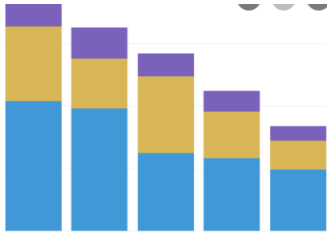
- Is a network structure
- The last node of the hierarchy are called leaves.

3)If data is **Categoric** and total variable is **two or more** and it **is subgroup** then use:

Spider Plot

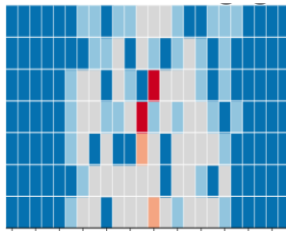
- To plot one or more series of values over multiple quantitative variable
- Each variable has its own axis and meet at centre

Stacked Bar Plot



- Showing proportion in the whole
- Use when categories has no order

Heatmap



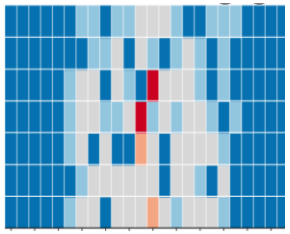
- Individual values contained in the matrix are represented as colours.
- Useful to display the result of hierarchical clustering

Dendrogram

- Is a network structure
- The last node of the hierarchy are called leaves.

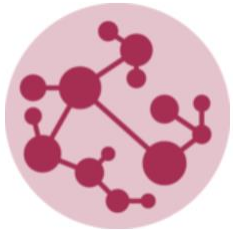
4)If data is **Categoric** and total variable is **two or more** and it **is adjacent** then use:

Heat Map



- Individual values contained in the matrix are represented as colours.
- Useful to display the result of hierarchical clustering

Network



- A set of nodes and links
- Shows interconnections between a set of variable. Each entity is represented by Node.

Arc

- A set of nodes and links.
- It is constituted by nodes that represent entities and by link that show relationship between entities.

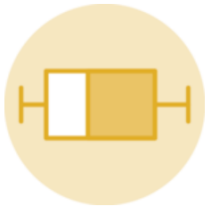
Chord

- Allows to visualize flows and connections between several entities.
- Each entity is represented by a fragment on the outer part of the circular layout.

What kind of data we have ?

1)If data is in **Time Series** and total series is **One** then use:

Box Plot



- Provides 5 summary data along with outliers
- If amount of data is high then add jitter
- While summarizing some information can be lost.

Violin

- The shape represents the density estimate of the variable
- It is close to a box plot but a deeper understanding of the distribution.

Area

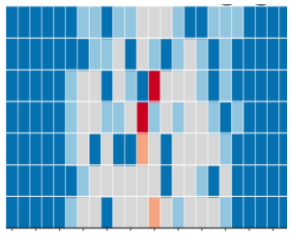
- Show evolution of the numeric variable
- Like a line chart but area under is filled with colour

Line Plot

- Show evolution of the numeric variable
- Data points connected by a straight line
- Use dots with line if data point are less

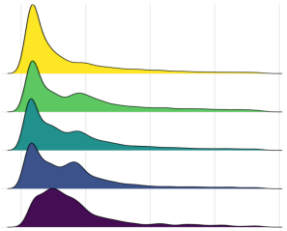
2)If data is in **Time Series** and total series is **more than One** then use:

Heatmap



- Individual values contained in the matrix are represented as colours.
- Useful to display the result of hierarchical clustering

Ridgeline

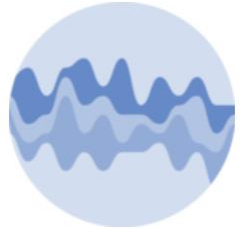


- Shows distribution of numerical value for several group.
- Works well when there is no ranking in groups.

Stacked Area

- Extension of area chart
- Appropriate to study the evolution of the whole and the relative portions of each group

Stream

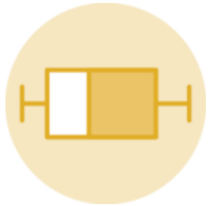


- A smooth version of the stacked area
- Displays evolution of numerical value following another numeric variable

What kind of data we have ?

1)If data is **one number and one categorical** then use:

Box Plot

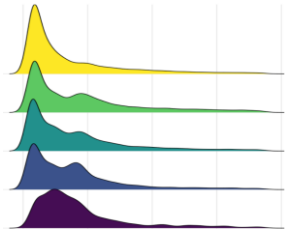


- Provides 5 summary data along with outliers
- If amount of data is high then add jitter
- While summarizing some information can be lost.

Violin

- The shape represents the density estimate of the variable
- It is close to a box plot but a deeper understanding
- of the distribution.
- Provides info about ranking and distribution

Ridge



- Shows distribution of numerical value for several group.
- Works well when there is no ranking in groups.

Histogram

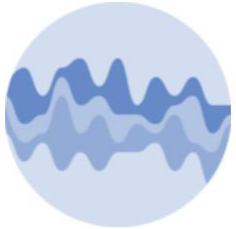
- Variable breaks into bins
- The number of observation per bin is represented by the height of the bin
- Don't use more than 3-4 plots

2)If data is **multiple number and one categorical** then use:

Stacked Area

- Extension of area chart
- **Appropriate to study the evolution of the whole and the relative portions of each group**

Stream



- A smooth version of the stacked area
- Displays evolution of numerical value following another numeric variable

Line

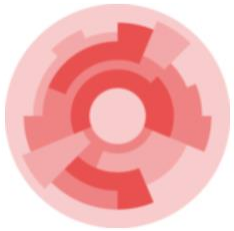
- Show evolution of the numeric variable
- **Data points connected by straight line**
- Use dots with line if data point are less

Connected Scatter

- Show evolution of the numeric variable
- **Individual observation are highlighted**

3)If data is **One number and Multiple category** then use:

Sunburst



- Also called as multilevel pie chart
- It is very similar to Treemap except it uses a radial layout

Treemap

- Displays hierarchical data as set of nested rectangle.
- Each rectangle represents proportions.

Chord

- Allows to visualize flows and connections between several entities.
- Each entity is represented by a fragment on the outer part of the circular layout.

Dendrogram

- Is a network structure
- The last node of the hierarchy are called leaves.

Thank You for reading this content. Don’t forget to get connected on our social media platforms for more details..