Information Visualization

Final Paper on "Line Chart visualization"

Department of Computer Science, Kent State University

CS 67302: Information Visualization

Bill Reed

Submitted by: Shiva Kumar Peddapuram

811235874 (speddpau@kent.edu)

Abstract:

The main intention of this Final paper is to summarise and give detailed information about the Line Chart Visualization. Line Chart Visualisation is the strongest visualization tool which depicts and shows the patterns and trends in each dataset over a certain time. In this paper we can explore the uses of line charts in data visualization which focuses on the strengths limitation about it. Even In this paper I am going to produce some strong evidence like strengths and weakness of this visualisation over the different datasets. This paper describes how the line charts can be used to convey useful summaries and support research conclusions.

Introduction:

Basically, The Line Chart is used to represent a set of information in a form of series of data points that are represented on two axes (can be of with any variable but mostly we use x and y axis) connected by a straight line. It depicts the changes in data over the time and to identify trends, patterns in each dataset. Line plots are used in plotting a wide range of the data, like stock Prices, Population growth, university students over a given period and many more forms.

Though they look simply but gives powerful insights and are widely used in the domains of finance, economics, sales etc.

In a line chart the horizontal axis represents mainly time or in some cases some other data from the datasets, and on the vertical axis represents the data that is needed to be plotted. Each data point can be represented either in dots or any other symbol (based on user interest). These set of points provides a clear visualization trend over the given time frame in the dataset.

History of the Visualisation:

To begin with the history of line charts, Line charts has been started back in the **18th century** where they first used to plot time graphs of series data, such as financial market data, Wheat Price changes etc.

Different sources says that Line chart was invented by Scottish Engineer and political economist **"William Playfair".** He is also the author of "**The commercial and political Atlas."** Which is written in 1786. Initially he used 43-line charts to understand time-series data about economic and political matters.

William Playfair line chart showed the wheat price in England from 1660 to 1782 which played a crucial role in the data visualization in that time. Before the introduction of line chart data is showed in different forms like tables or simple graphs when they started using this line chart it made easy in terms of comparison for the data over different series of time and enabled users to identify different trends and patterns.

The use of line-chart drastically increased in the 19th century mainly in the field of economics.

Next, the French economist Adolphe Quetelet used line charts to analyse demographics data, while the English Economist and Statistician William Stanley Jevons used line charts to study economics fluctuations.

As the 20th century started the usage of line charts became even more widespread with this technology era. Even different software's came into to picture where it became easy to visualize charts using them. Like: MS-Excel, Tableau, Powerbi etc

Today line charts are commonly used in a variety of field like finance (to analyse the trends of a stock over the time), science, and engineering to visualise the data over the given period and identify the trends.

Strengths of Line Chart Visualization:

1.) Line charts are "**Easy to read and summarize**" as we use simple and intuitive visual elements such as dots and lines to represent the data.

2.) Line charts are the best for showing different trends in a data over the time. Making them useful foe tracking changes in data over extended periods.

3.) We can use multiple datasets in a single chart making it easy to compare and analyse different trends in different datasets in a single chart.

4.) The best thing about the line chart is that we can identify outliers in the given dataset by highlighting data points which are deviating drastically from the trend lines.

5.) Line charts can be used to represent the carious data types of data like continuous categorical data making flexible tool for data visualization.

6.) Line charts are effective for presentations and reports. As they are effective in communicating trends.

7.) Line charts can help decision-makers identify trends and make informed decisions based on data-driven insights.

8.) Line charts can be used for larger data sets which make them useful for analysing larger trends over extended periods.

Weakness of the Visualization or Limitations:

1.) Line charts are limited to one dimension that means they can display the data along one dimensions. If we want to compare multiple variables, we need to use multiple line charts.

2.) Line charts are limited to continuous data.

3.) If we have multiple lines on a chart, it gets difficult to compare them accurately. Mainly when they have different scales or units of measurements.

4.) If the datasets have a large set of data, then it may hide the data. This can be rectified by increasing or managing the size of the chart.

Visualization:

Dataset from Kaggle: https://www.kaggle.com/code/mrwolfgang/google-playstore-eda

Example of the Line chart Visualization using Google Play store dataset:

Process for performing the Line chart visualization:

1) Initially I have downloaded the dataset from the Kaggle (which an open source where we get lot of datasets for free) the link for the dataset has been provided above.

2) I have opened the file in tableau desktop (cleaning the data and making the raw dataset to the usable dataset) and then cleared all the null values. After cleaning I have saved the file in my localhost.

3) Then I have loaded the data into MS-Excel by opening the file in it.

4) To perform the visualisation, I have taken some parameters (objectives).

5) I have taken 6 objectives and then performed different operations on it then I got different line chart visualizations.

6) The 6 objectives are as follows:

   a.) Top 10 apps based on ratings on each category,

   b.) Number of Apps based on content Rating (paid or free versions),

   c.) Top Rankings of Apps based on Installs in each category,

   d.) Number of Apps based on rating,

   e.) Number of Apps that are most recently updated,

   f.) Top Maximum size of apps category wise.

Whenever we download a new application, we consider lot of parameters into consideration, so I have taken some of them as my objectives and then worked on it

These objectives help a software developer to focus on those aspects when developing a particular application.
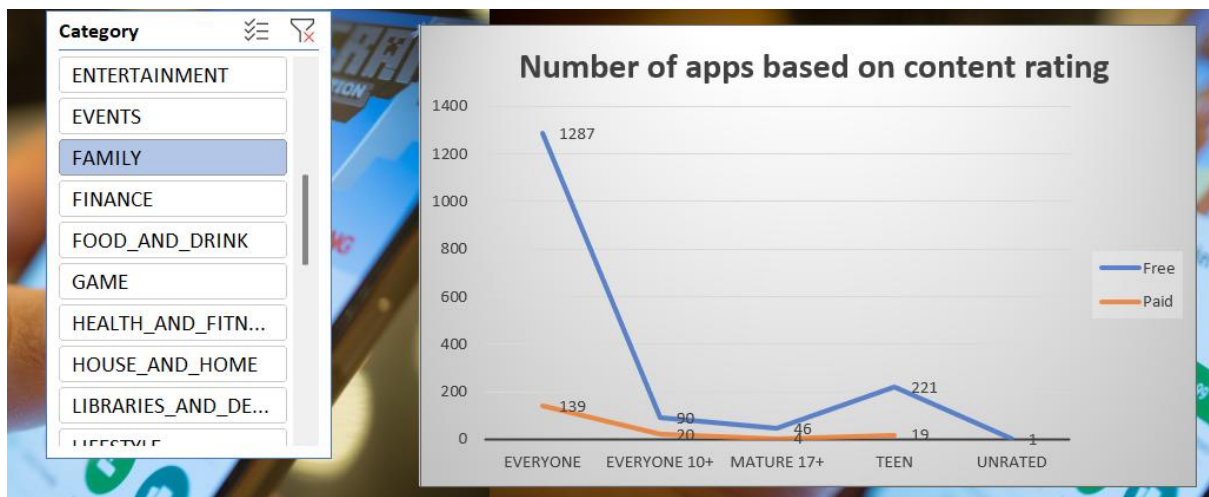
Visualization-01:

In the below visualization we can see the top 10 apps with high review count. The number on the point shows the number of reviews for app that means for angry birds epic RPG it has a review count of 2634605 in this way it follows for Top 10 Apps

Top 10 Reviews of apps based on catogry
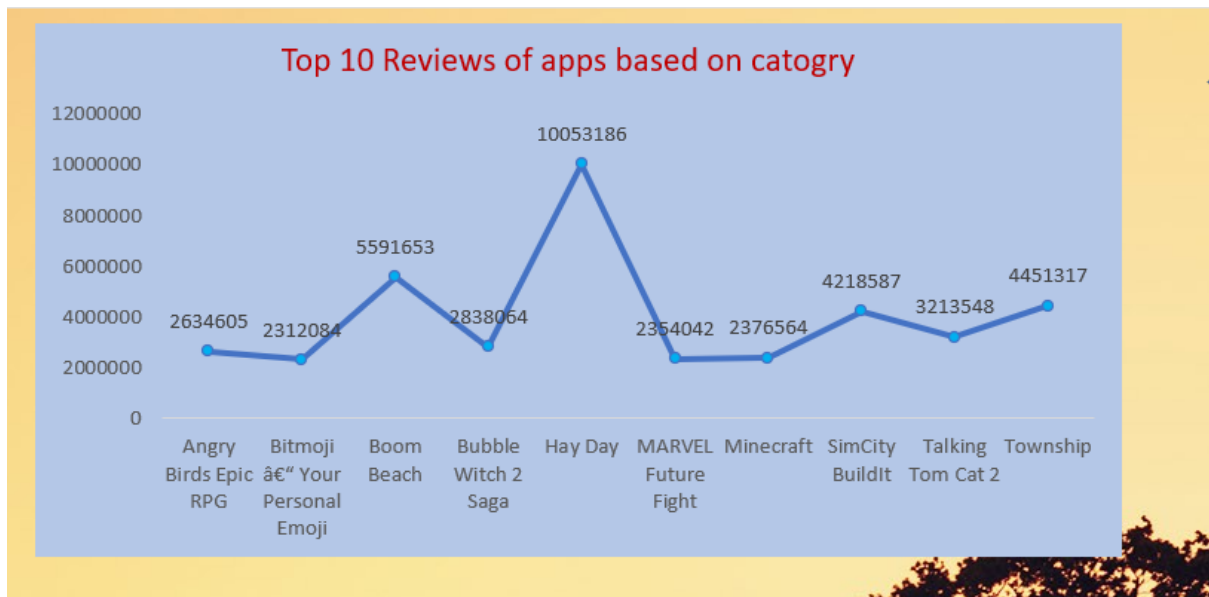
Visualization-02: (line chart with multiple parameters)

In the following visualization I have made a line chart visualization and I have compared number of apps based on content rating like is a particular app suitable for everyone or else is it restricted to any age group and here we have two points that is free version and paid version of the app.


So, from the above visualisation the learning outcome is for the family Category there are 1287(free Apps which doesn't require any subscription) and 139 (apps which are paid) apps that can be used by everyone.

Visualization-03 (number of apps that are most recently updated in finance category)
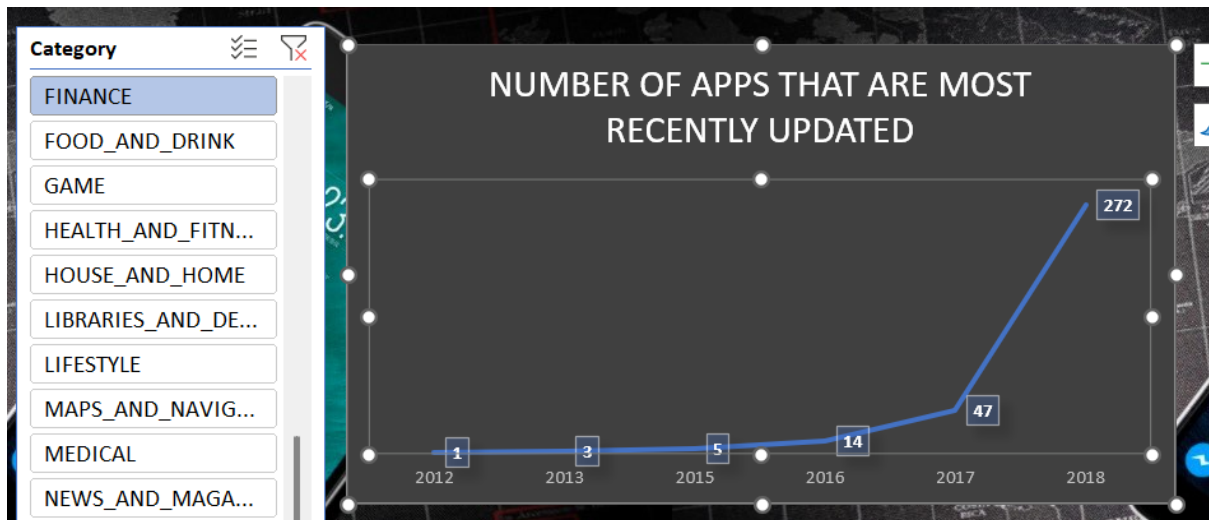
Visualization-04: Number of apps based on rating from Travel and Local Category
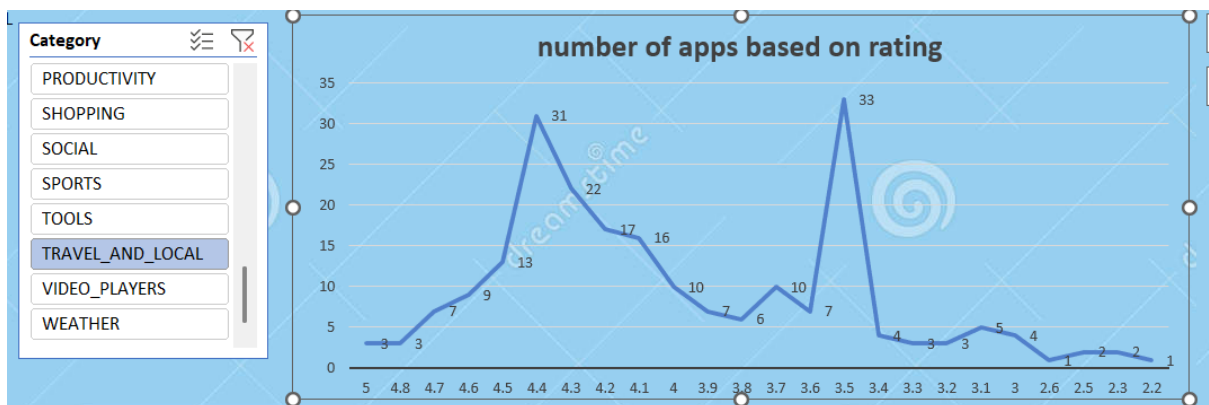
Learning Outcome: From the above figure the learning outcome is 33 apps from travel and local category has 3.5/5.0 rating and 31 apps has 4.4/5.0 ratings I have also added numbers on the top so that it becomes easy for us to determine the apps based with rating.

specific situations in which one would use this line chart visualizations:

1.) Line charts are best suitable to showcase the price fluctuations of a stock over a given period.

2.) Temperature changes. Line charts helps us to show temperature change in terms of Day Wise, week wise, month wise, year wise. So that we can study the exact change in the temperature over the time.

3.) Revenue comparison of a company over the period.

4.) Sales trends for a particular product in a company.

5.) Line chart is not only restricted to above situations but also it can be used for those type of datasets we can also use it in different cases.

Reference:

1.) The Line Chart - How and When to Use It (blog Post)

https://infogram.com/blog/the-line-chart-how-and-when-to-use it/#:~:text=Many%20sources%20give%20credit%20for,about%20political%20and%20economic%20matters. (History)

2.) Parth Sharma (2018) Google Play store EDA dataset

https://www.kaggle.com/code/mrwolfgang/google-playstore-eda (for the dataset)

3.) Mike Y. A Complete Guide to Line Charts. https://chartio.com/learn/charts/line-chart-complete-guide/

4.) Katelyn Peters (2022). Line Chart: Definition, Types, Examples, How to Make in Excel

https://www.investopedia.com/terms/l/linechart.asp