

Statistics - Task 4

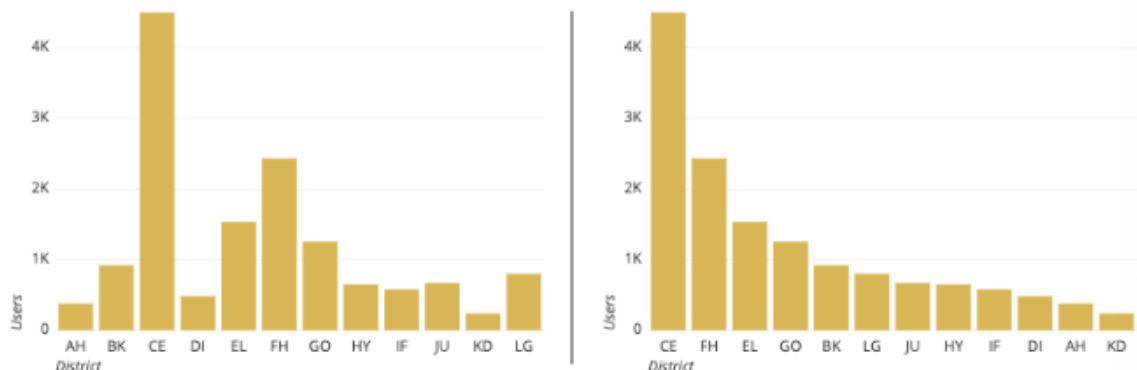
By Menna Jaheen

"Effective data visualization can mean the difference between success and failure when it comes to communicating the findings of your study, raising money for your nonprofit, presenting to your board, or simply getting your point across to your audience."

Types of Graphs and when to use them:

1. Bar Chart :

- A bar chart (aka bar graph, column chart) **plots numeric values for levels of a categorical feature as bars.**
- Levels are plotted on one chart axis, and values are plotted on the other axis. Each categorical value claims one bar, and the length of each bar corresponds to the bar's value.
- A bar chart is used when you want to show a distribution of data points or perform a comparison of metric values across different subgroups of your data.

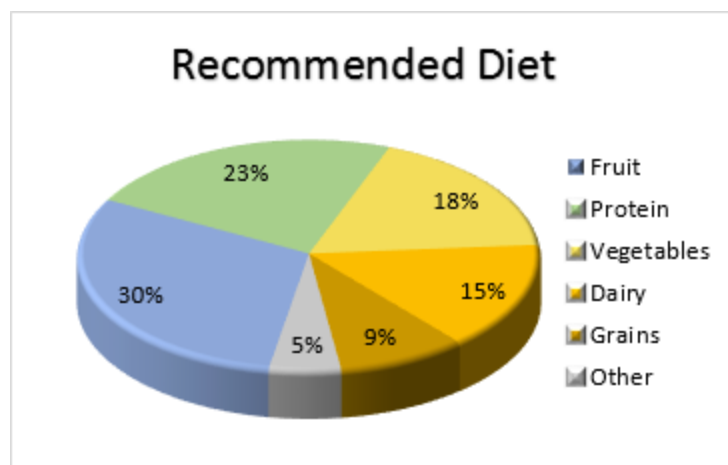


- From a bar chart, we can see which groups are highest or most common, and how other groups compare against the others. Since this is a fairly common

task, bar charts are a fairly ubiquitous chart type.

2. Pie Chart :

- A pie chart is a circular statistical graphic divided into slices to illustrate numerical proportions.
- Each slice's size corresponds to that category's percentage of the whole.
- Pie charts are best used when displaying data for categories that are parts of a whole and typically don't exceed 5-7 categories.



- They are effective for showing percentage or proportional data and can be easily understood at a glance.
 - However, pie charts can become difficult to read when there are too many slices or when the values are very similar.
 - Alternatives to pie charts include bar charts, stacked bar charts, or treemaps, especially when dealing with many categories or similar values.
-

3. Heat map:

- A heat map is a graphical representation of data where individual values are represented as colors.

- It's used to visualize complex data sets and identify patterns, trends, or anomalies at a glance.
- In a heat map, higher values are typically represented by warmer colors (e.g., red, orange) and lower values by cooler colors (e.g., blue, green).
- Heat maps are particularly useful for showing the correlation between two variables in a table format.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Sales in Thousands (USD)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	
2	2010	45	57	59	64	70	81	90	83	75	68	59	50	
3	2011	60	58	81	71	92	82	72	48	47	58	60	64	
4	2012	72	84	99	96	94	75	61	82	81	90	65	34	
5	2013	40	74	59	84	99	58	60	59	62	94	83	28	
6	2014	68	64	53	68	45	85	49	41	76	91	87	39	
7	2015	86	67	46	51	98	69	93	84	67	62	45	20	
8	2016	47	91	60	82	93	75	58	79	95	70	85	32	
9	2017	77	73	100	82	97	77	59	52	80	52	94	42	
10	2018	71	50	91	51	81	97	49	80	68	69	75	25	
11	2019	69	54	47	100	89	97	98	75	69	52	79	27	
12	2020	86	49	59	75	79	44	70	48	91	68	71	31	
13	2021	40	49	65	75	80	74	82	92	73	60	40	30	
14														

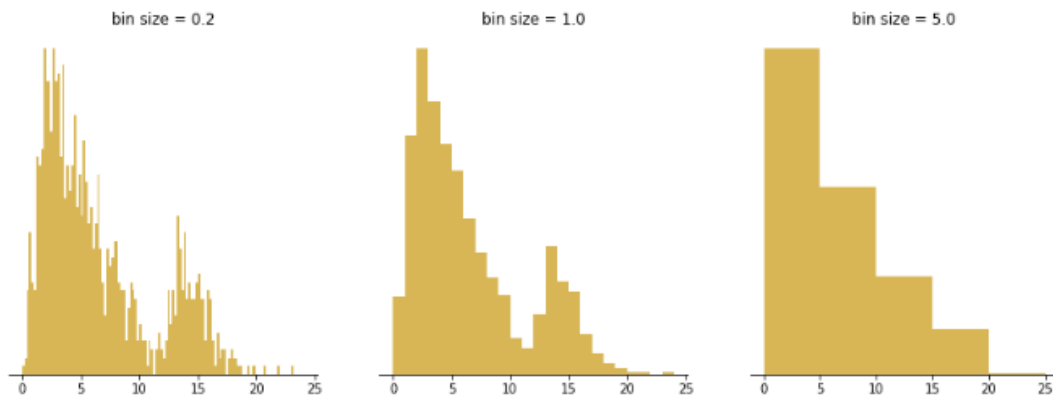
- They can be applied in various fields, including website analytics (to show user interaction), financial markets (to display stock performance), and scientific research (to visualize large datasets).
- Heat maps are effective when you need to display a large amount of multivariable data in a compact form.
- They can help in identifying outliers, recognizing patterns, and comparing categories within the data.

4.Histogram:

- A histogram is a chart that plots the distribution of a numeric variable's values as a series of bars.
- Each bar typically covers a range of numeric values called a bin or class; a bar's height indicates the frequency of data points with a value within the

corresponding bin.

- Histograms are good for showing general distributional features of dataset variables.



In summary, histograms are useful for:

- Visualizing the distribution of a continuous variable
- Identifying the central tendency, spread, and shape of the data
- Detecting outliers and gaps in the data
- Comparing distributions across different groups or datasets

Histograms are particularly valuable in exploratory data analysis, helping researchers and analysts understand the underlying structure of their data before applying more complex statistical techniques.
