Concepts of Visualization

Mk_Data

Two main reasons for Data Visualization:

Exploratory

analysis is done when you are searching for insights.

These visualizations don't need to be perfect. You are using plots to find insights, but they don't need to be aesthetically appealing. You are the consumer of these plots, and you need to be able to find the answer to your questions from these plots.

Explanatory

analysis is done when you are providing your results for others.

These visualizations need to provide you the emphasis necessary to convey your message. They should be accurate, insightful, and visually appealing

Visuals in steps of the data analysis process:

1. **Extract** - Obtain the data from a spreadsheet, SQL, the web, etc.
2. **Clean** - Here, we could use exploratory visuals.
3. **Explore** - Here, we use exploratory visuals.
4. **Analyze** - Here, we might use either exploratory or explanatory visuals.
5. **Share** - Here is where explanatory visuals live.

Mk_Data

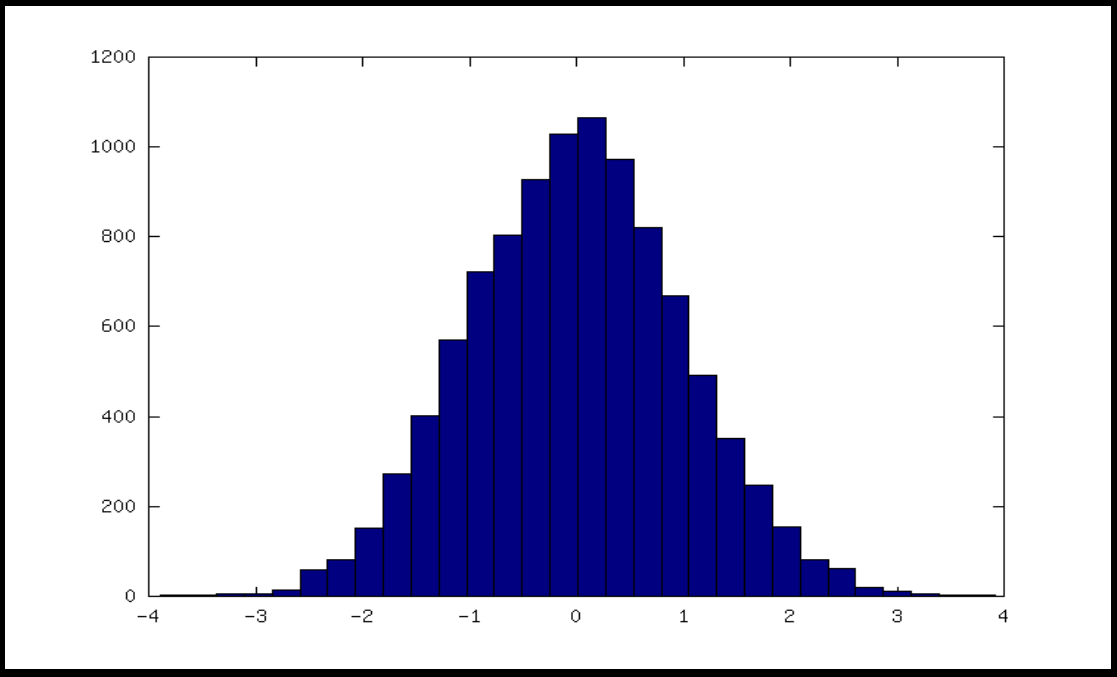
Now let's know:

- How can you choose an appropriate plot?
- What are common types of plots?
- What should you avoid in your visual?

Histogram

When should you use a Histogram?

1. If you have a single continuous variable.
2. You want to ask questions about the shape of its distribution.

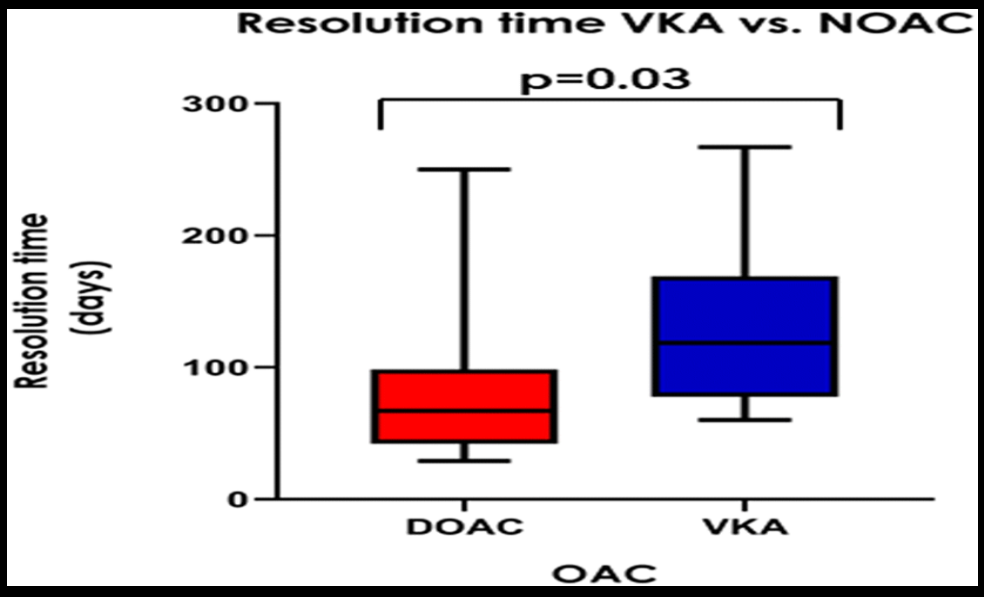


Mk_Data

Box plot

When should you use a boxplot?

1. When you have a continuous variable, split by a categorical variable.
2. When you want to compare the distribution of the continuous variable for each category

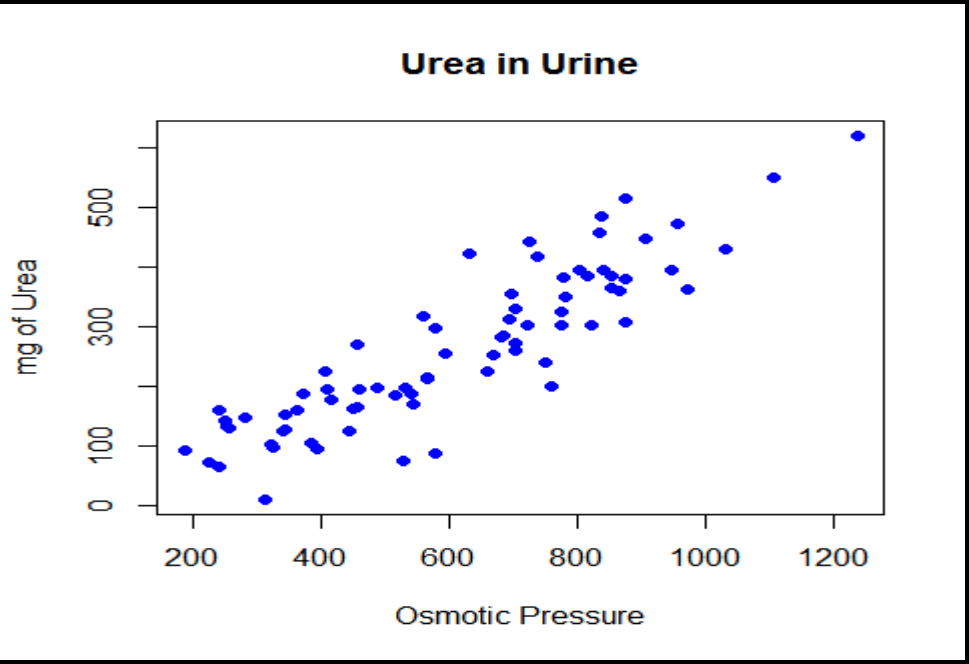


Mk_Data

Scatter plot

When should you use a scatter plot?

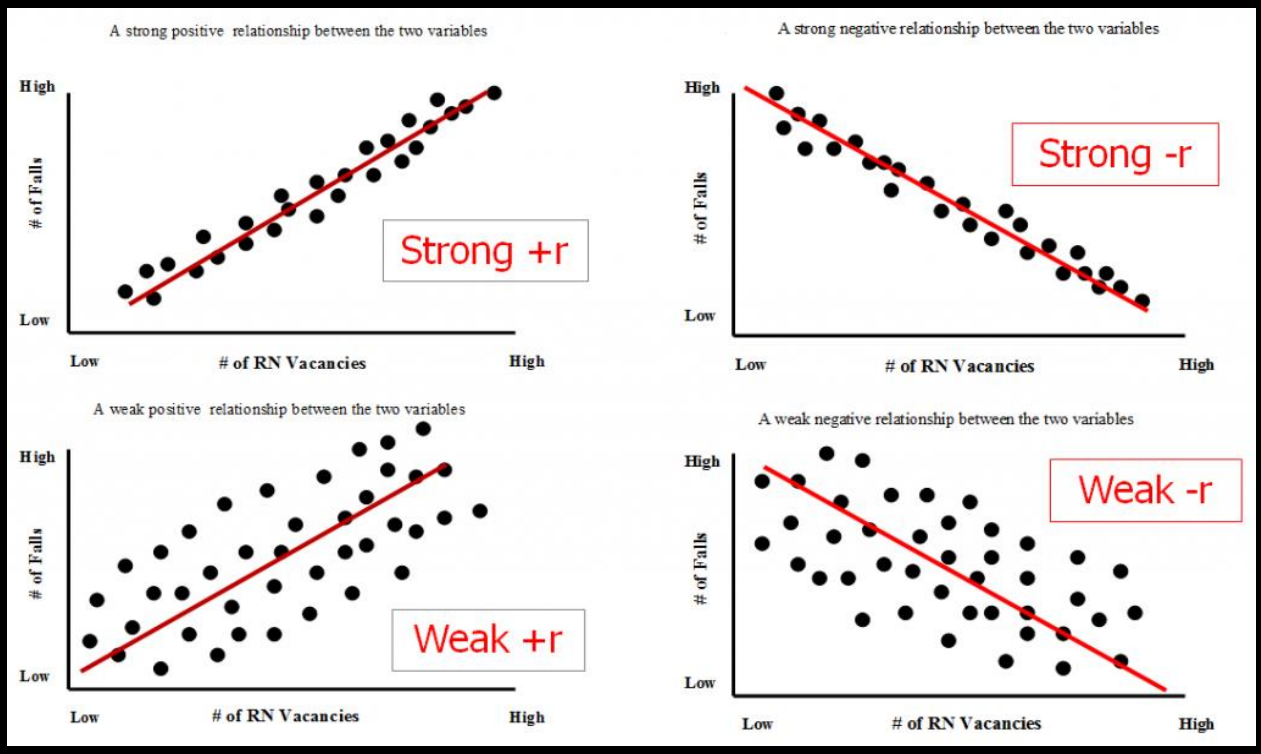
1. You have two continuous variables.
2. You want to answer questions about the relationship between the two variables.



Mk_Data

Correlation

How close are you to being able to fit a straight line through the points?



Mk_Data

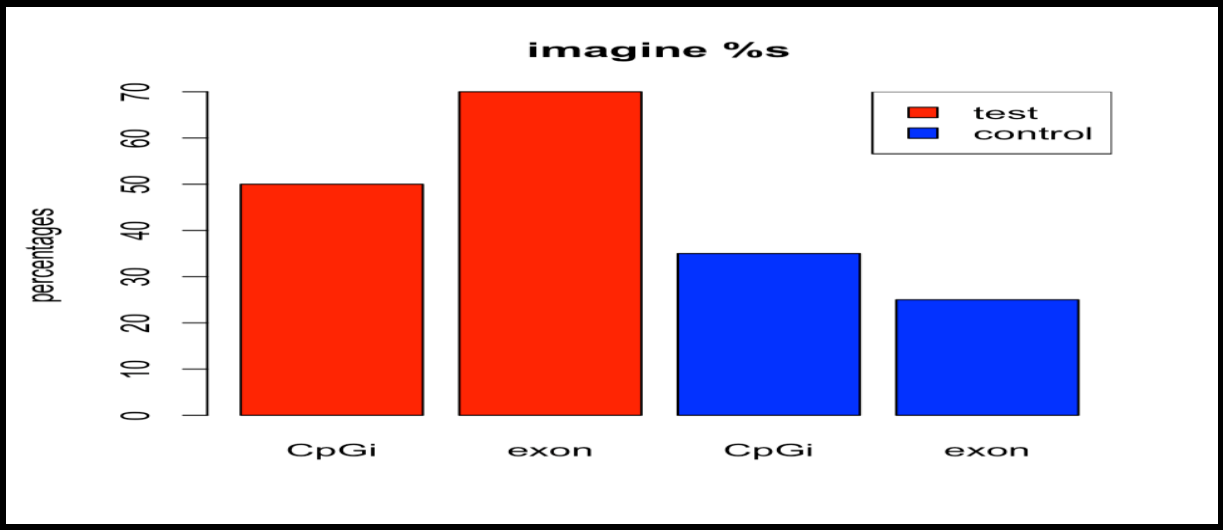
Bar plot

When should you use a bar plot?

1. You have a categorical variable.
2. You want counts or percentages for each c category.

Occasionally:

1.You want another numeric score for each category and need to include zero in the plot.

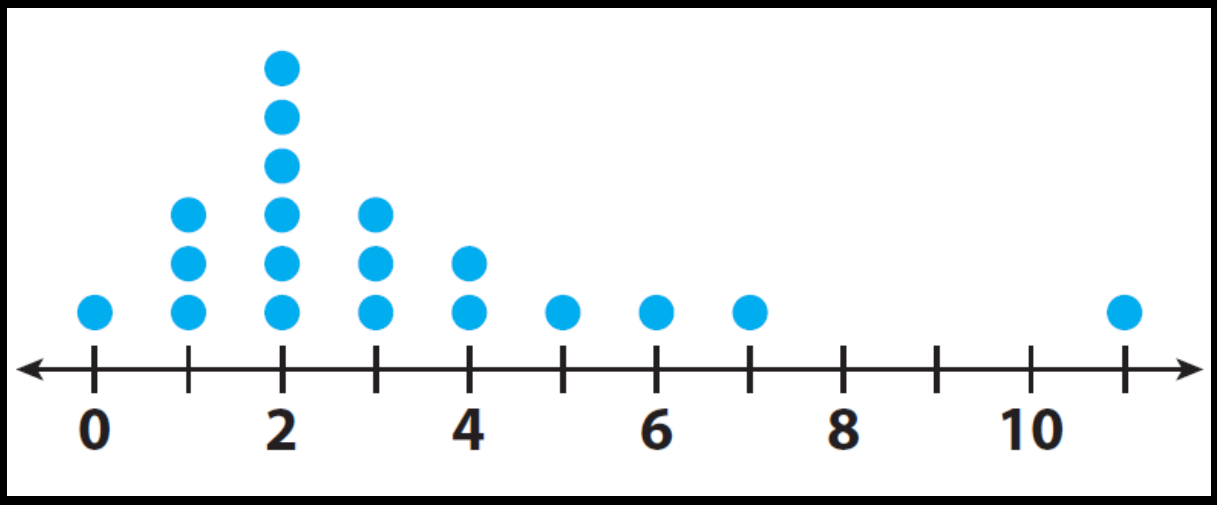


Mk_Data

Dot plot

When should you use a dot plot?

1.You have a categorical variable.
2.You want to display numeric scores for each category on a log scale,
3.You want to display multiple numeric scores for each category.

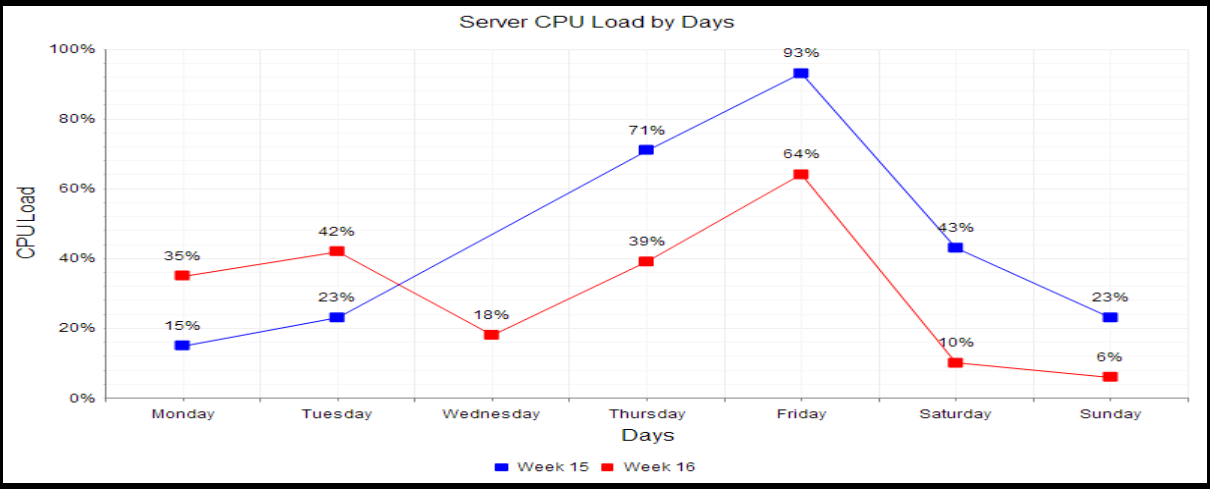


Mk_Data

Line plot

When should you use a line plot?

1.You have two continuous variables.
2.You want to answer questions about their relationship.
3.Consecutive observations are connected somehow
Usually, but not always, the x-axis is dates or times.



Mk_Data

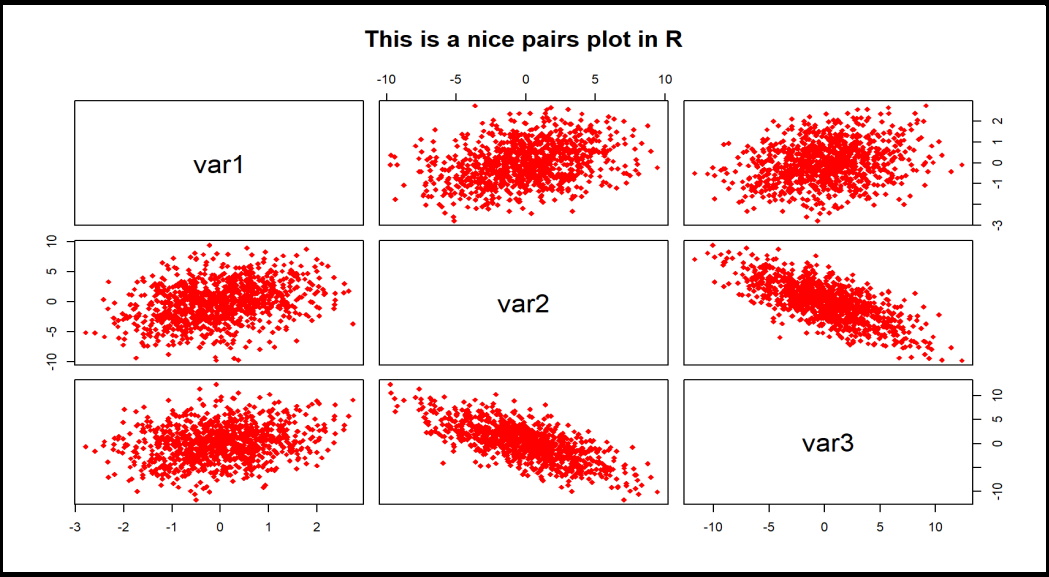
Pair plot

When should you use a pair plot?

You have up to ten variables (either continuous, categorical, or a mix).

You want to see the distribution for each variable.

You want to see the relationship between each pair of variables



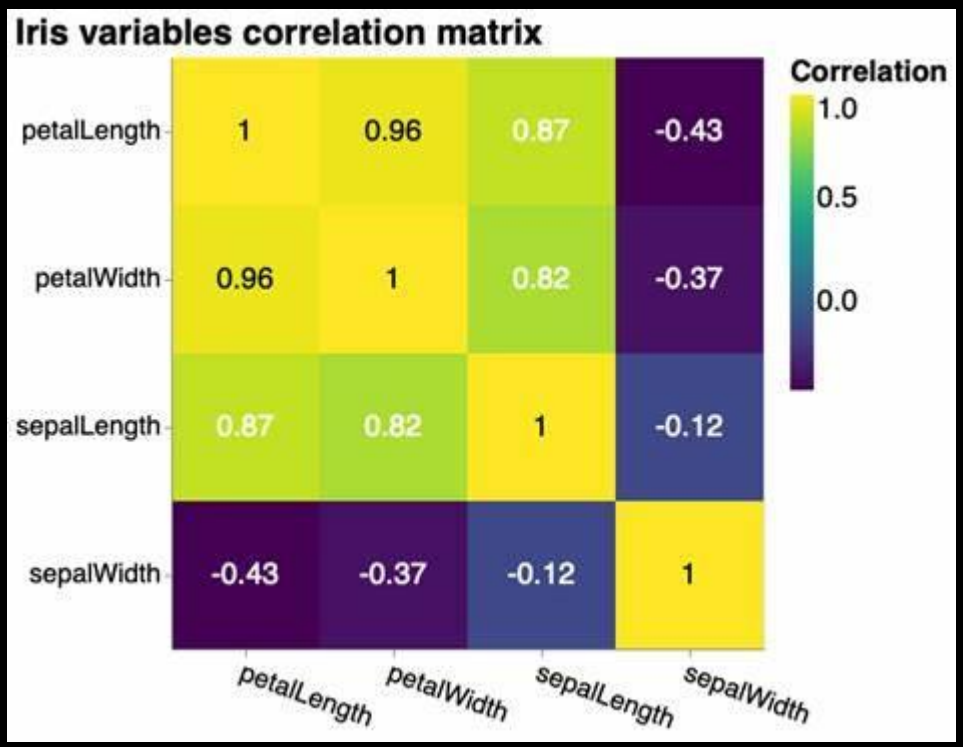
Mk_Data

Heatmap

When use a correlation heatmap?

You have lots of continuous variables.

You want to a simple overview of how each pair of variables is related

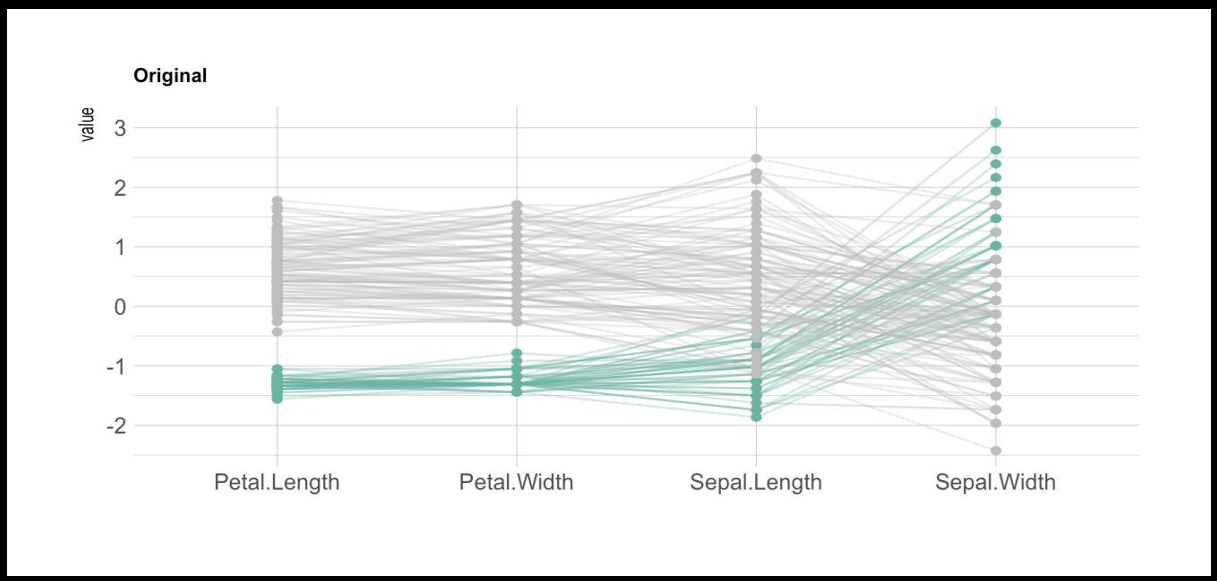


Mk_Data

Parallel coordinates plots

When should you use a parallel coordinates plot?

You have lots of continuous variables.
You want to find patterns across these variables.
You want to visualize clusters of observations.



Mk_Data

Attention!

- A bar plot is almost always easier to interpret than a pie plot
- Dot plots are like bar plots, but you can use log scales and show multiple metrics on each row.

- your visual should be free from Chart junk.

What is a **Chart junk**?

Any element of the plot that distracts from the reader getting insight

- Pictures
- Skeuomorphism: reflections, shadows, etc.
- Extra dimensions
- Ostentatious colors or lines

References:

Data Camp: Visualization for Everyone - Data Literacy Fundamentals track

Udacity: Data Visualization -Advanced Data Analysis Nano degree

Follow me

https://www.linkedin.com/in/maha-koutb/

Mk_Data