

Data Visualization

Effective Visualization

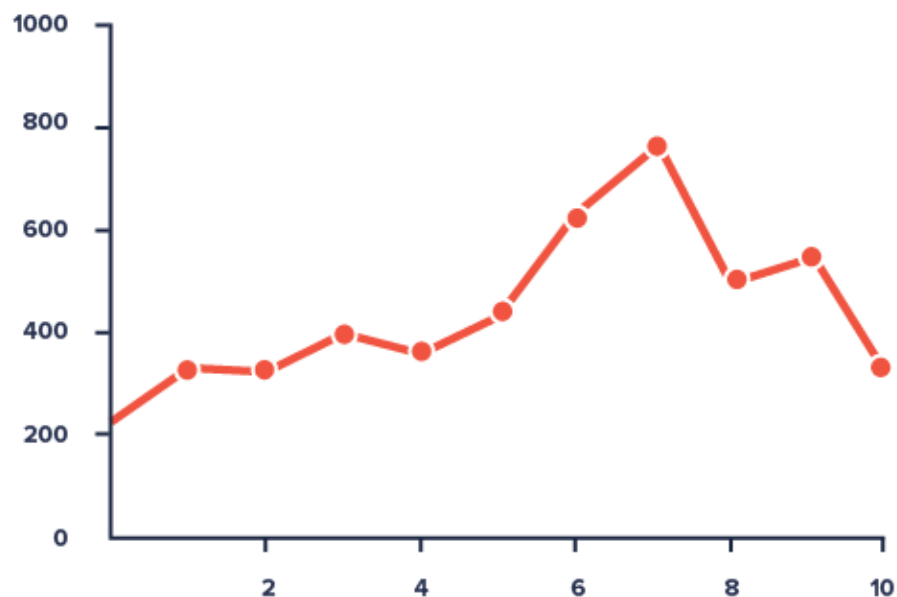
Made Satria Wibawa

Visualisasi untuk Menggambarkan Trend

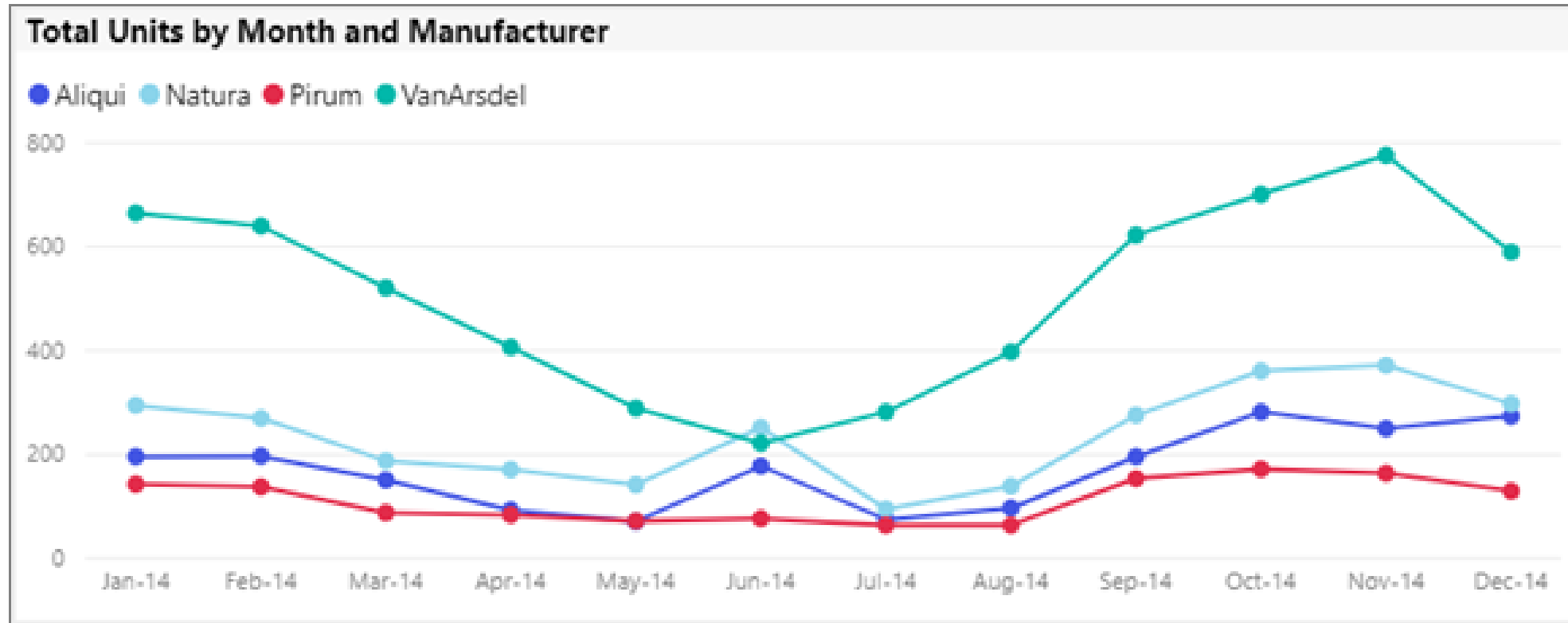
- Line Chart
- Sparklines Chart
- Stacked Area Chart
- Candlestick Chart

Line Chart

Line chart atau grafik garis merupakan tipe visualisasi yang memberikan informasi dari sekumpulan data poin (marker) yang dihubungkan dengan garis lurus. Axis lainnya (biasanya axis x) menggambarkan interval (biasanya waktu) perubahan nilai data poin tersebut.

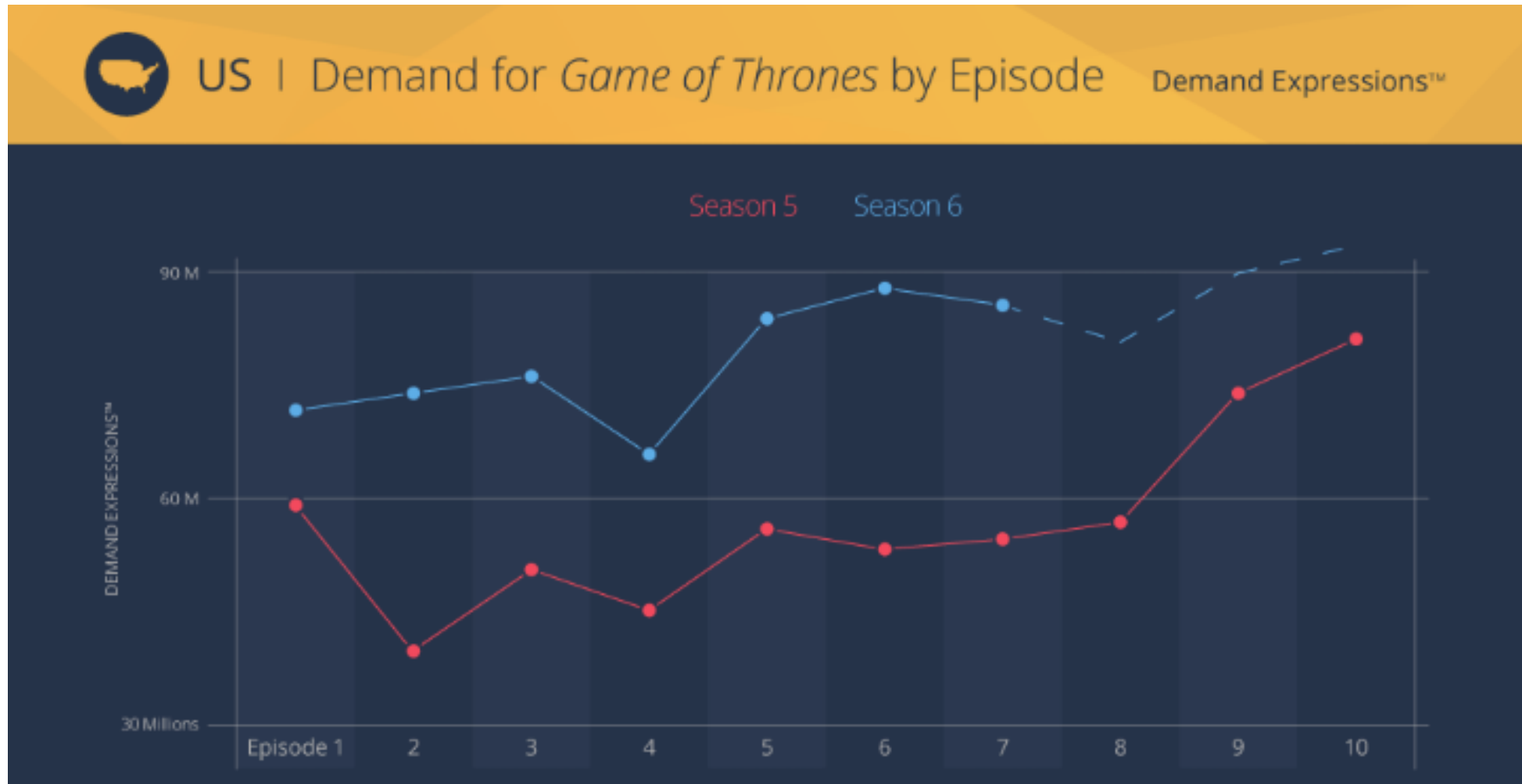


Contoh Line Chart



line chart dengan axis x adalah waktu

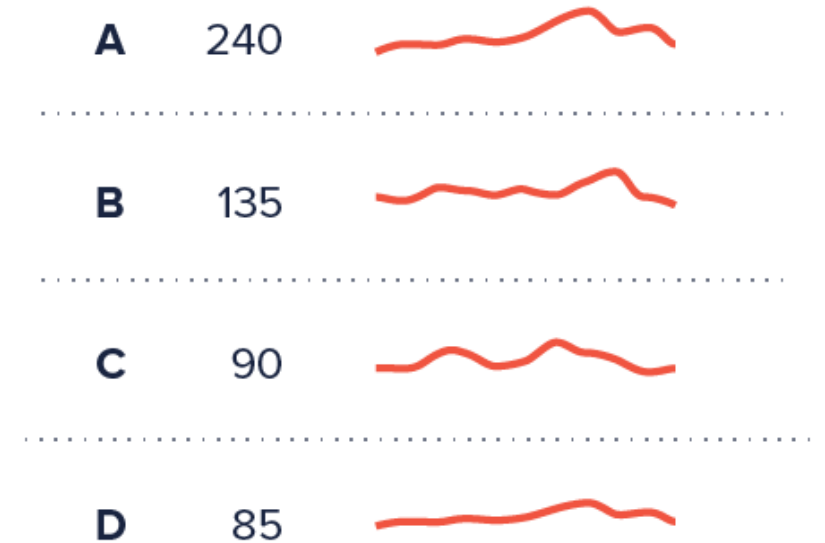
Contoh Line Chart



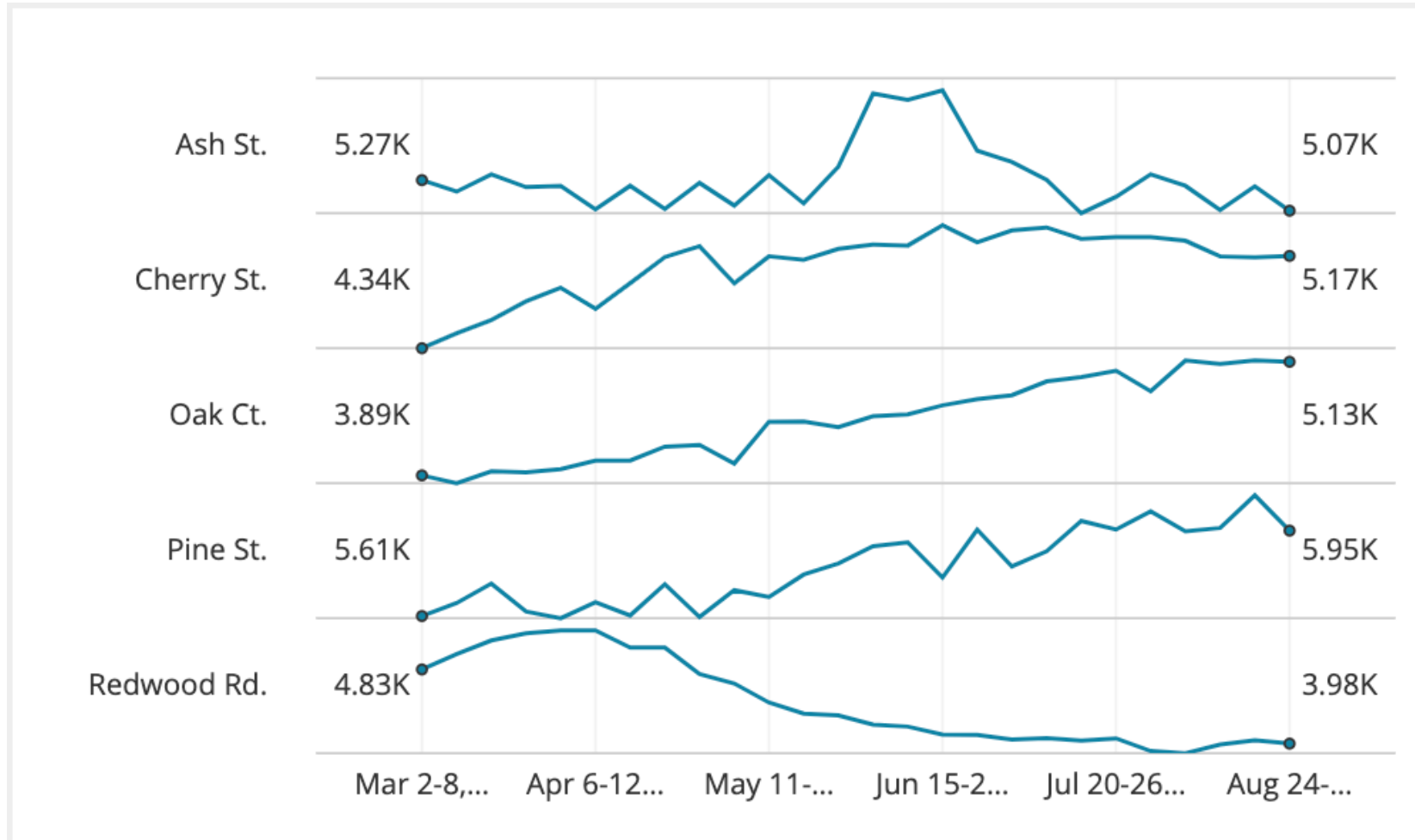
line chart dengan axis x adalah urutan sekuen

Sparklines Chart

Sparkline chart merupakan sekumpulan line chart yang ditumpuk secara vertical. Axis yang menunjukkan interval memiliki rentang dan nilai yang sama. Axis yang menunjukan nilai data tidak memiliki rentang yang sama. Chart ini menekankan penggambaran perubahan pada kategori yang berbeda.



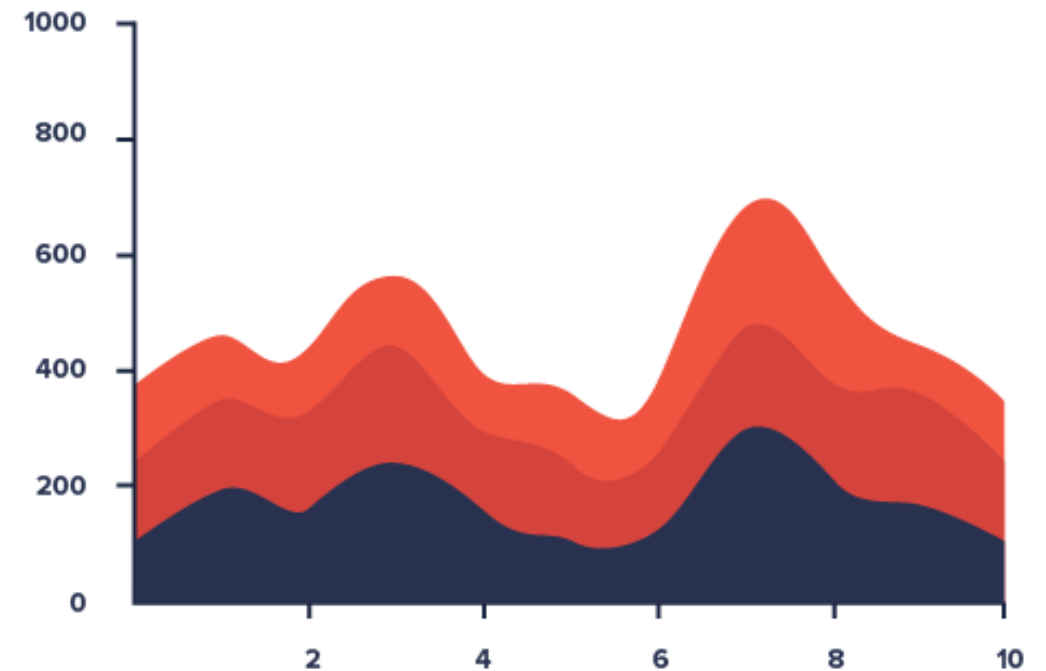
Contoh Sparklines Chart



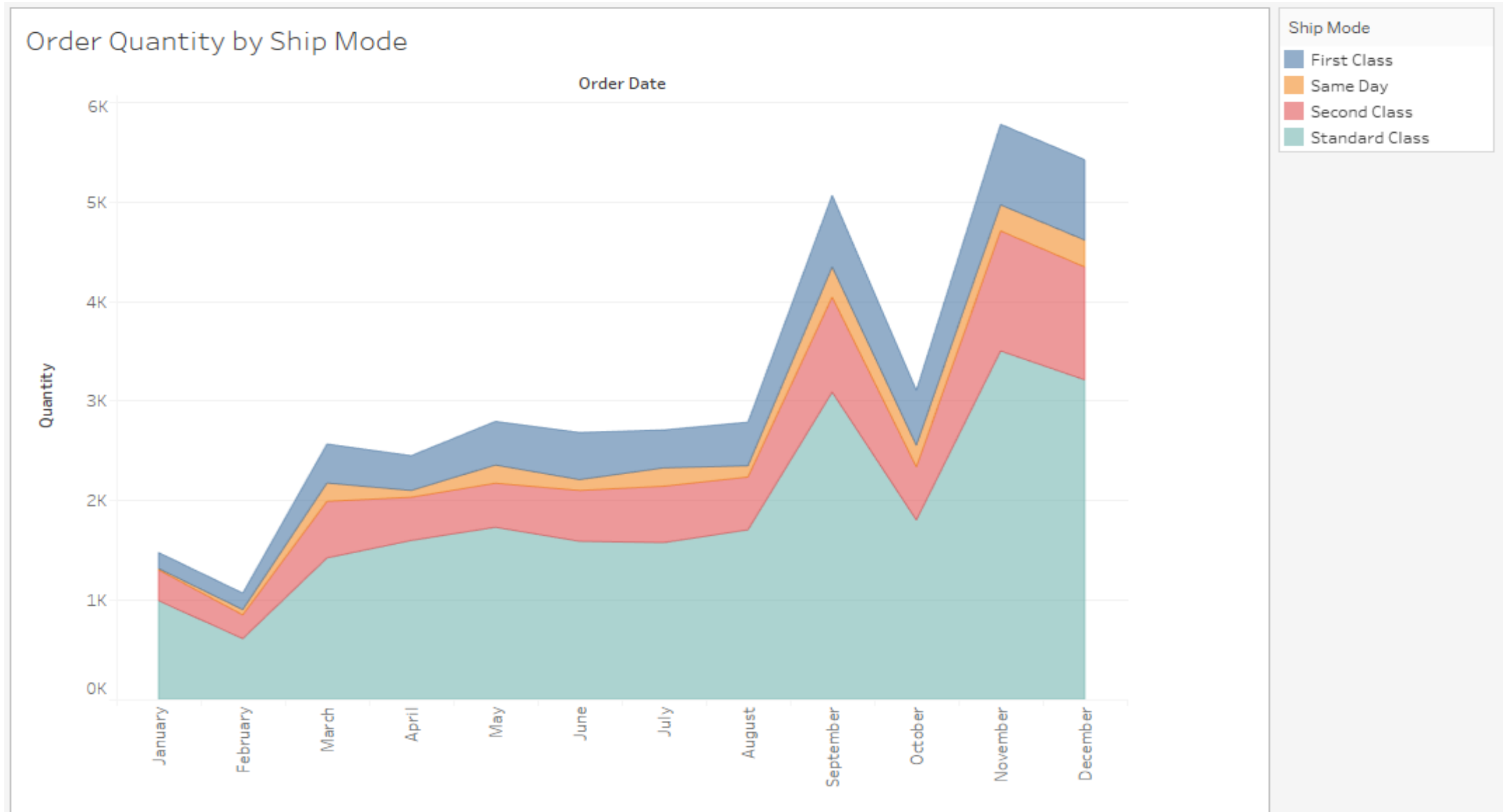
Stacked Area Chart

Stacked area chart berfungsi untuk membandingkan perubahan nilai beberapa variabel seiring perubahan waktu.

Nilai variabel pada stacked area chart dilihat dari lebar area pada chart.



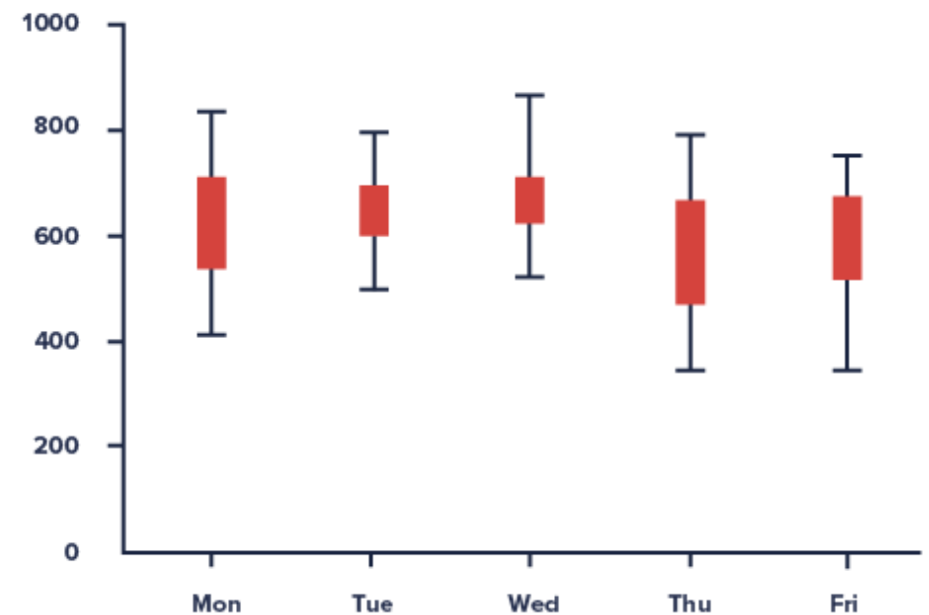
Contoh Stacked Area Chart



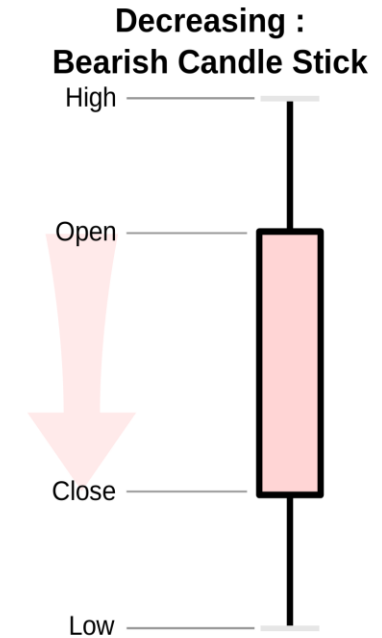
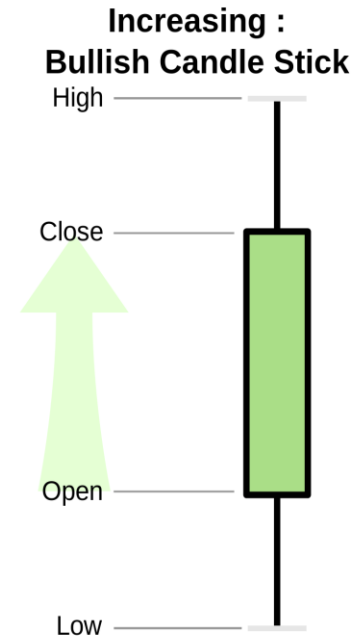
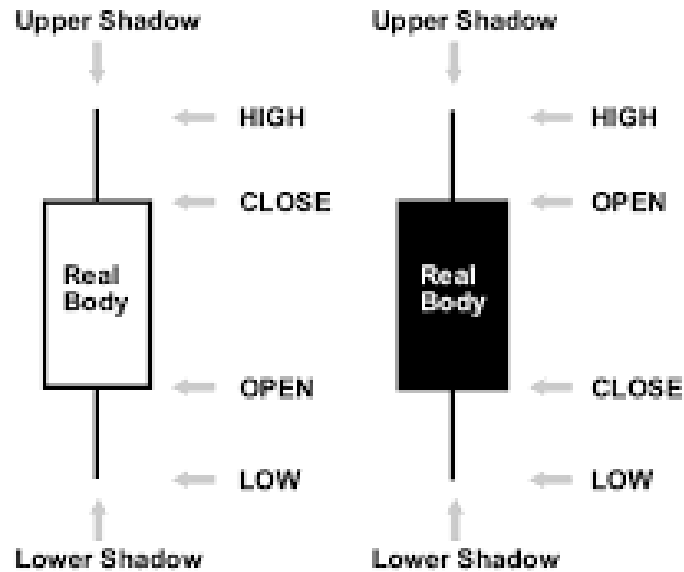
Candlestick Chart

Candlestick chart memberikan informasi untuk perubahan harga. Setiap bar merepresentasikan perubahan harga dari suatu interval waktu tertentu.

Dataset dari candlestick memiliki informasi harga terendah (low), tertinggi (high), harga pembukaan (open) dan harga penutupan (close).



Candlestick Chart



Nilai low dan high ditunjukkan pada bawah dan atas dari bar.
Nilai open dan close ditunjukkan dari warna bar.

Korelasi

- Korelasi adalah **hubungan statistik** antara **dua** variabel.
- Korelasi menunjukkan seberapa besar kekuatan hubungan **linear** antar variabel.
- Korelasi memiliki nilai dari **-1 hingga 1**.
- Jika korelasi bernilai **negatif**, maka hubungan dikatakan **berbanding terbalik**. Artinya jika salah satu variabel nilainya semakin besar, maka variabel lain semakin kecil dan sebaliknya.
- Jika korelasi bernilai **positif**, maka hubungan dikatakan **berbanding lurus**. Artinya jika salah satu variabel nilainya semakin besar, maka variabel lain semakin besar dan sebaliknya.

Korelasi

- Untuk menghitung korelasi biasanya dihitung menggunakan persamaan **Pearson**

- Pearson Correlation : $\rho_{X,Y} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$

- Korelasi **!=** Kausalitas

Visualisasi untuk Korelasi

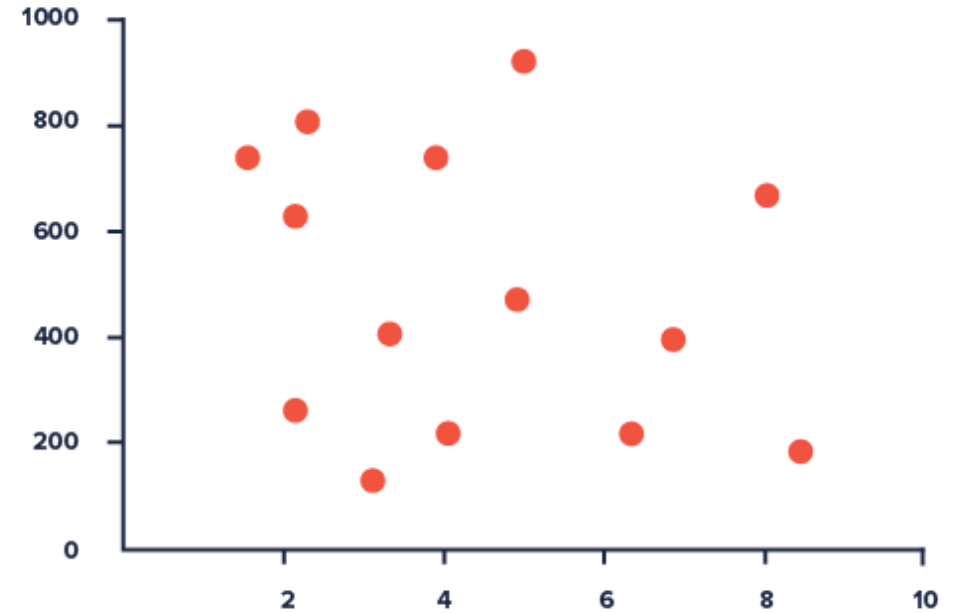
- Scatter Plot
- Bubble Plot
- Scatter Plot Matrix
- Heatmap
- Network Graph
- Radial Network

Scatter Plot

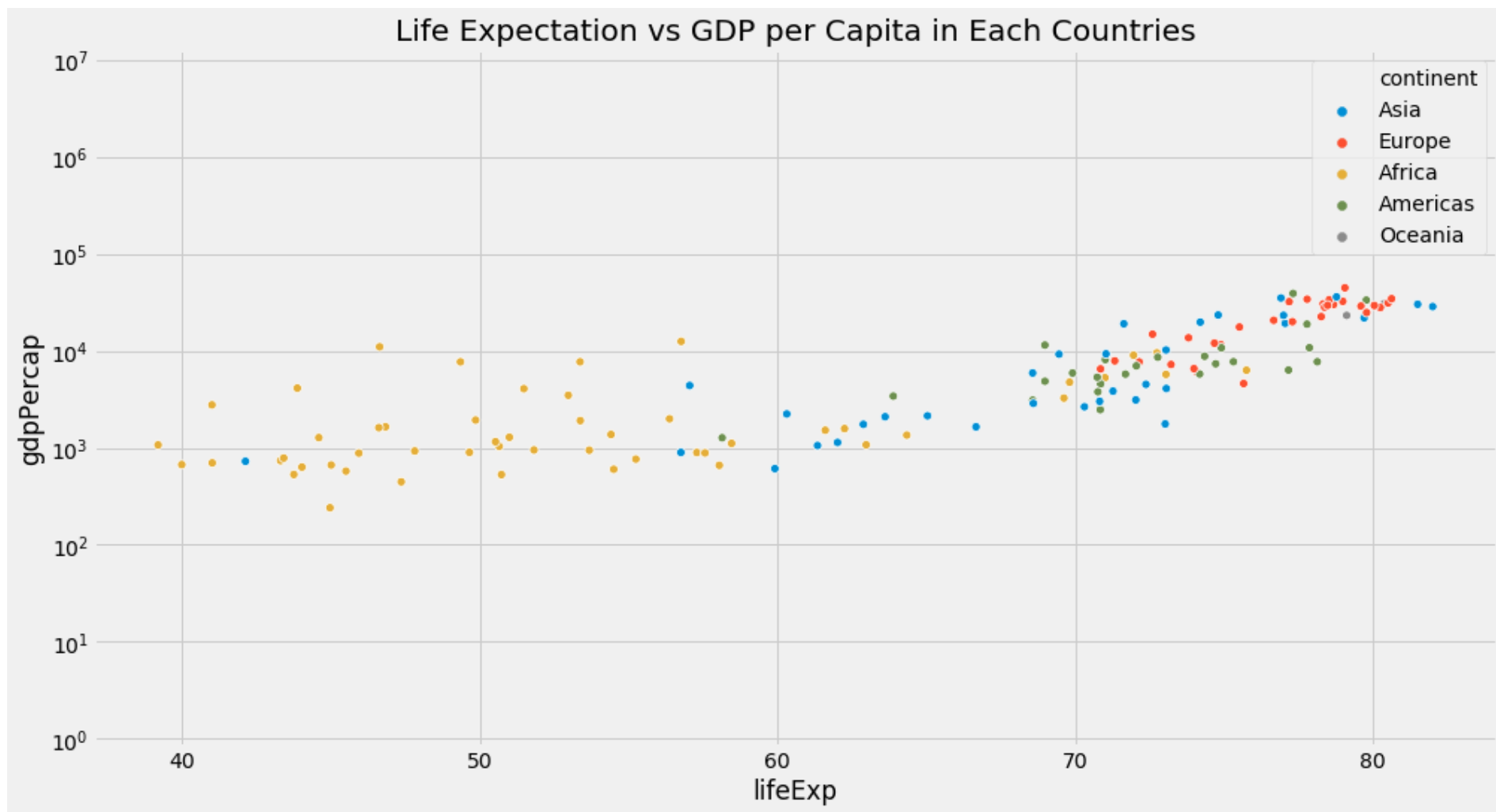
Scatter plot menggambarkan hubungan antar dua buah variabel. Variabel direpresentasikan dengan axis x dan y.

Data poin/instance dalam dataset digambarkan dalam bentuk titik/lingkaran kecil dengan ukuran yang sama.

Posisi dari titik tersebut, menggambarkan nilai dari instance.



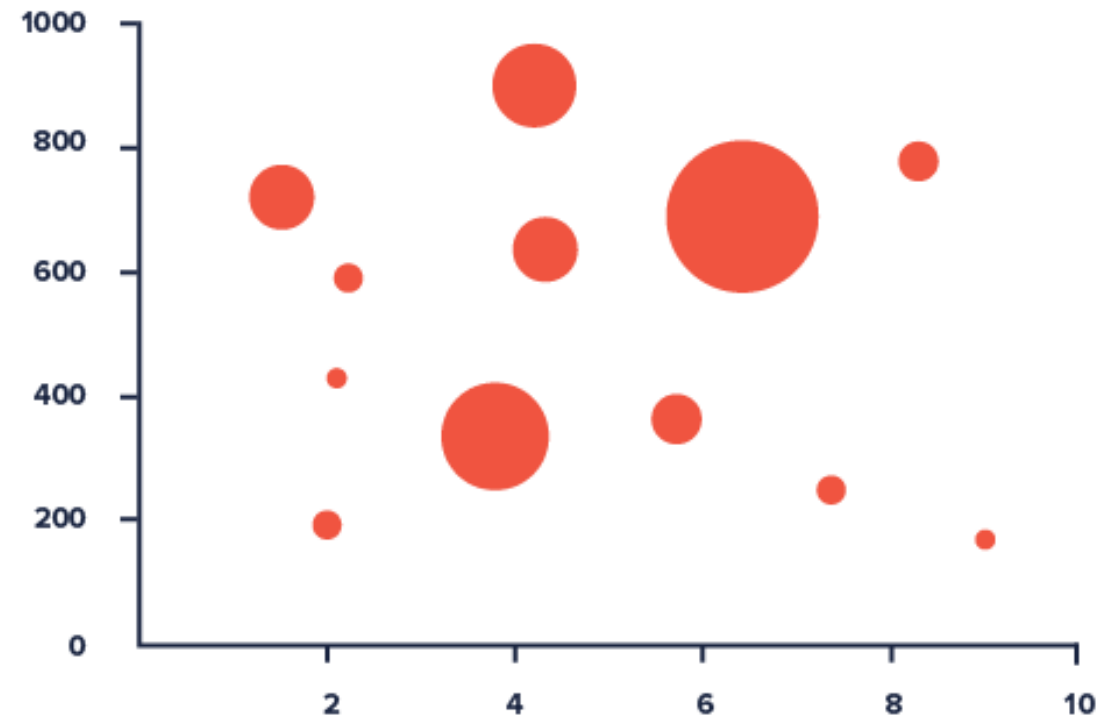
Contoh Scatter Plot



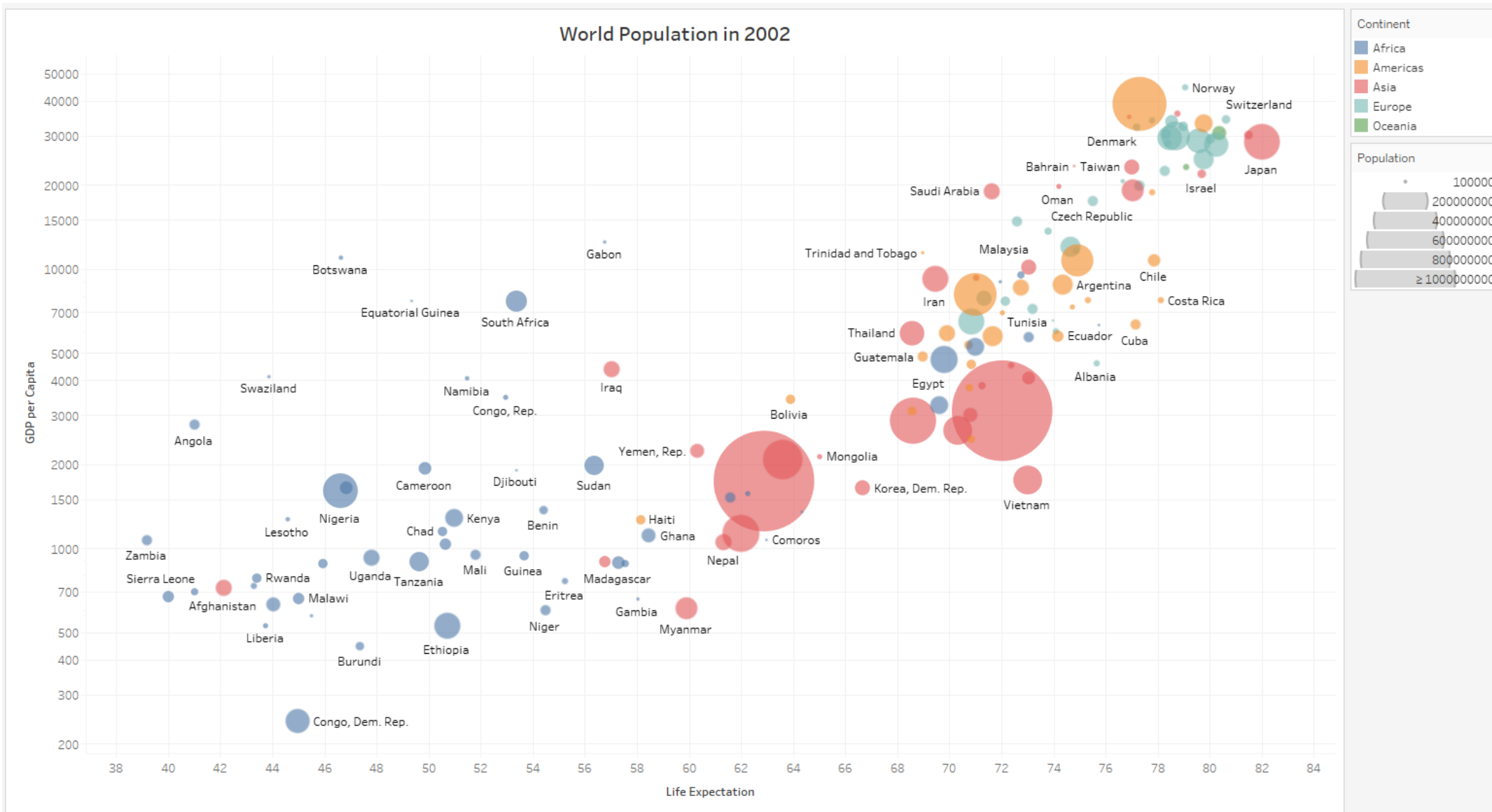
Bubble Plot

Bubble plot adalah grafik yang mirip dengan scatter plot. Hanya saja bubble plot dapat menggambarkan korelasi antar 3 variabel sekaligus.

Nilai data poin dari variabel ketiga digambarkan dengan ukuran titik/lingkaran dalam plot.



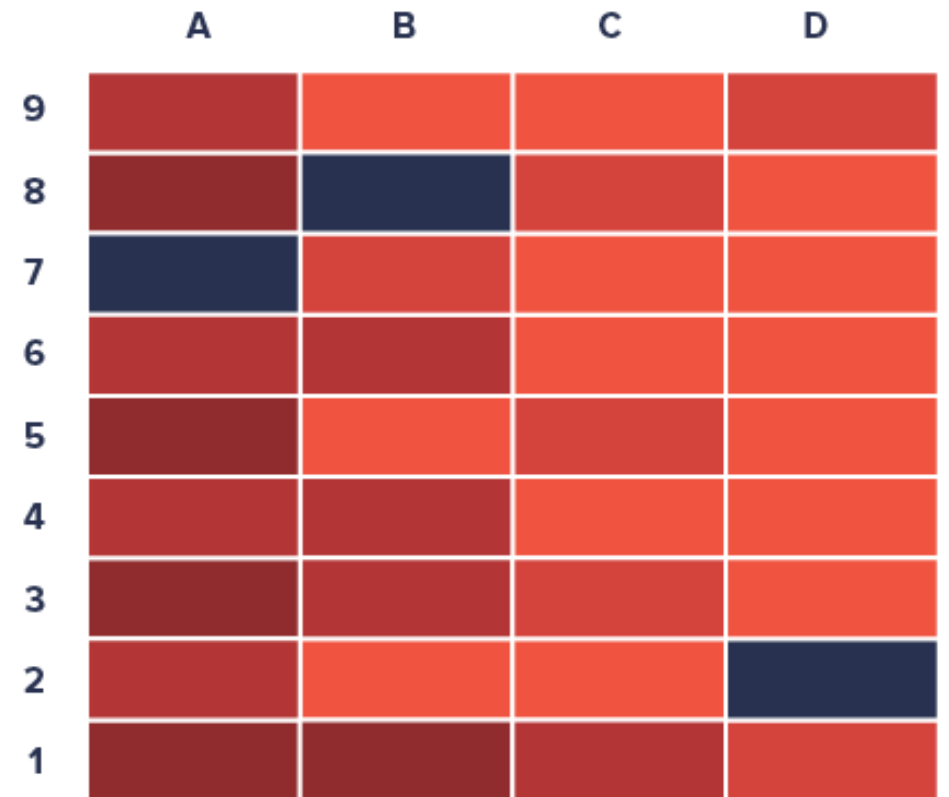
Contoh Bubble Plot



Heatmap

Heatmap menggambarkan nilai dalam bentuk matriks, dimana nilai direpresentasikan melalui warna.

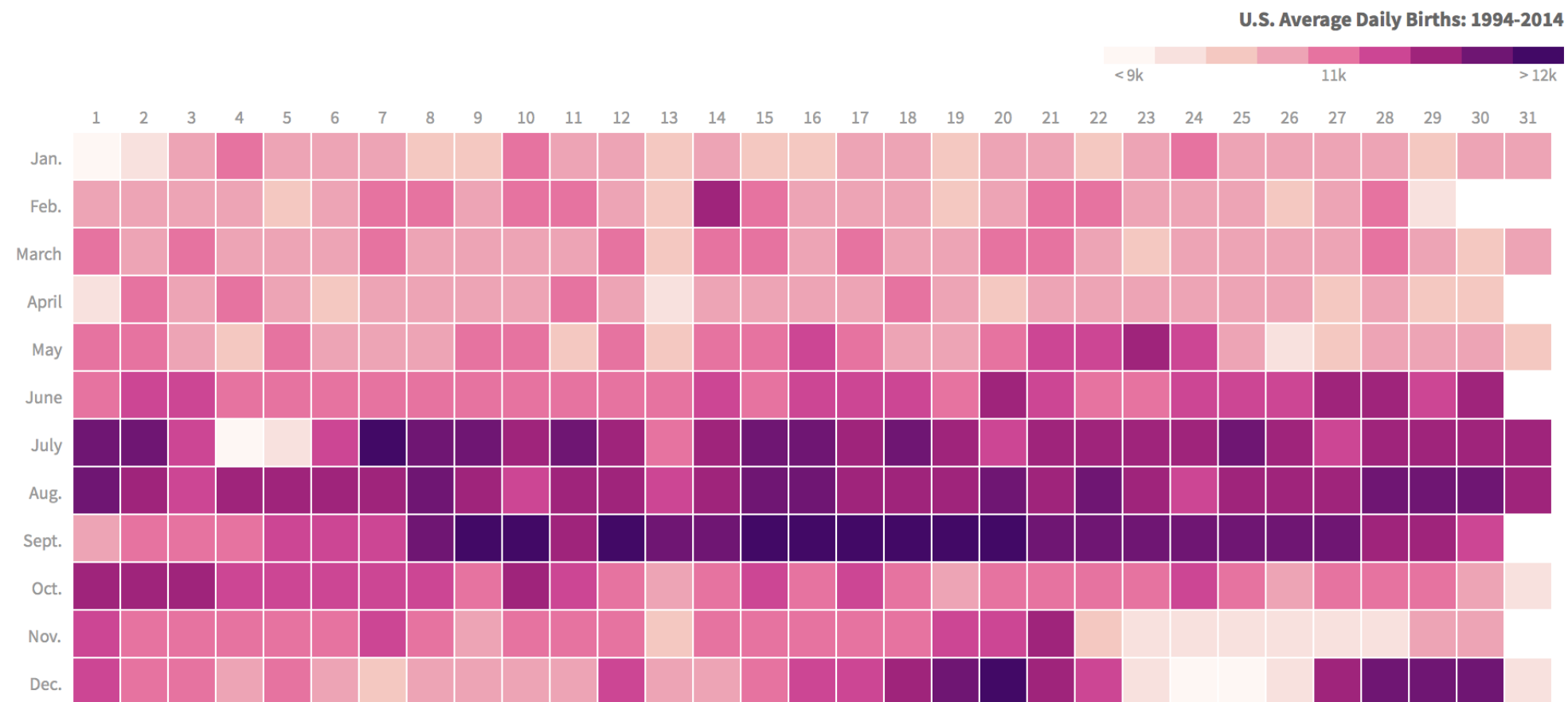
Warna pada heatmap berguna untuk melihat nilai kontinyu di variabel yang jamak dalam waktu singkat. Selain itu, heatmap juga dapat memperlihatkan pola dalam korelasi.



Contoh Heatmap

How Popular Is Your Birthday?

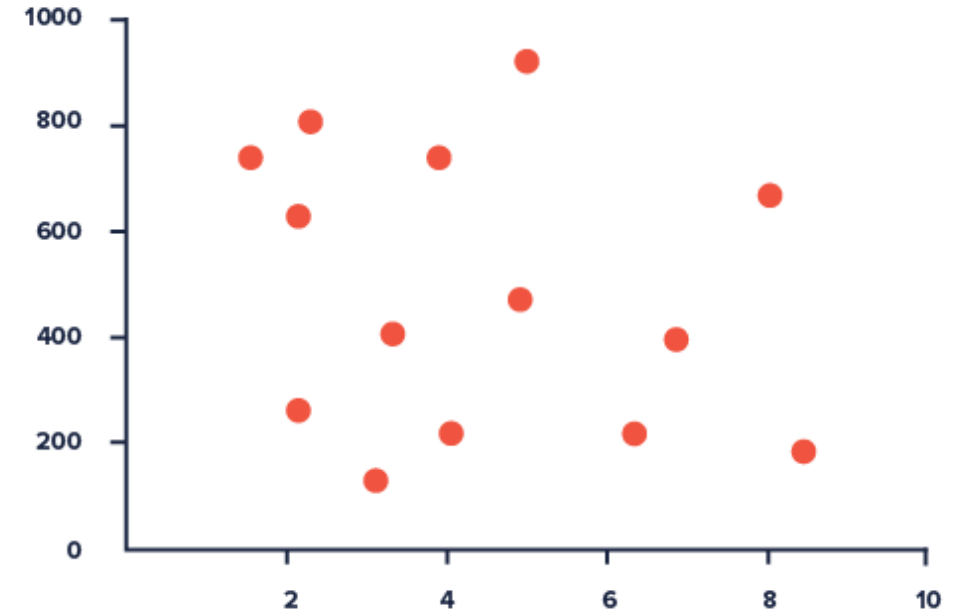
Two decades of American birthdays, averaged by month and day.



Scatter Plot Matrix

Scatter plot matrix atau istilah lainnya pairplot merupakan scatter plot yang disusun dalam bentuk matriks untuk menggambarkan korelasi bivariate (2 variabel).

Pada chart diagonal, biasanya digambarkan distribusi univariate (1 variabel).



Contoh Scatter Plot Matrix



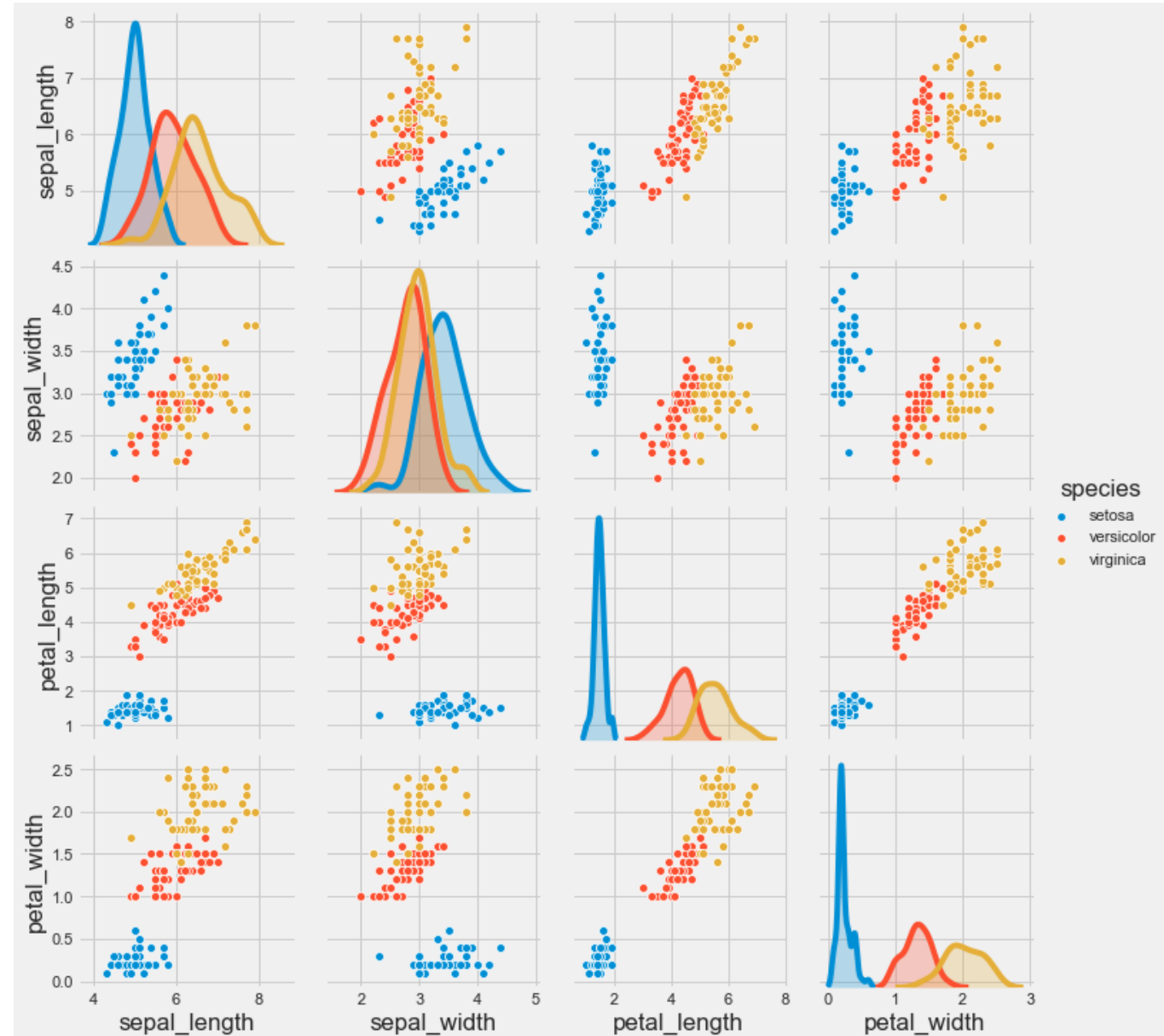
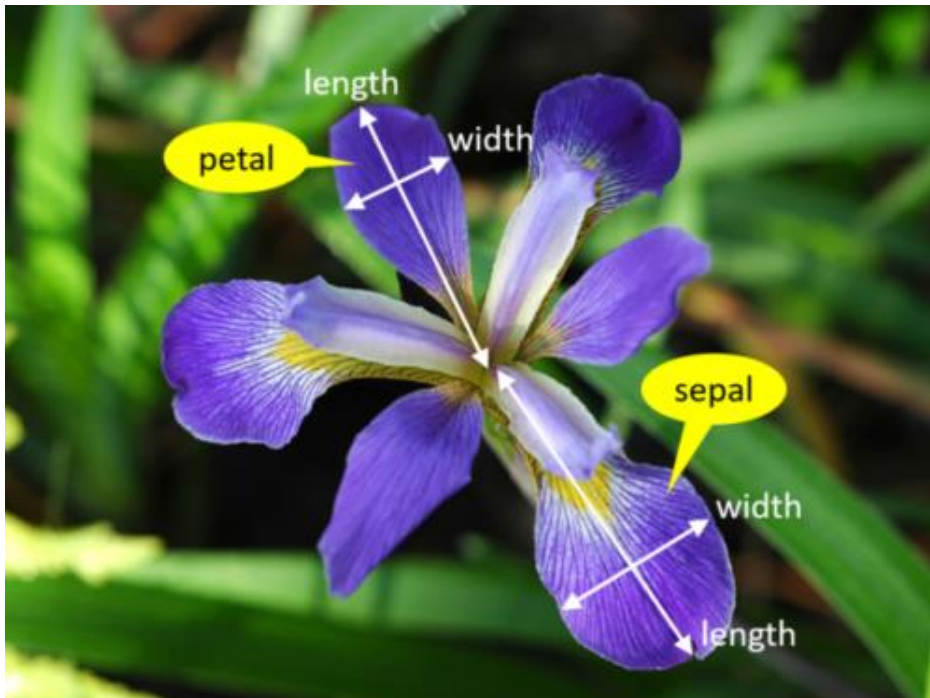
Iris setosa



Iris versicolor



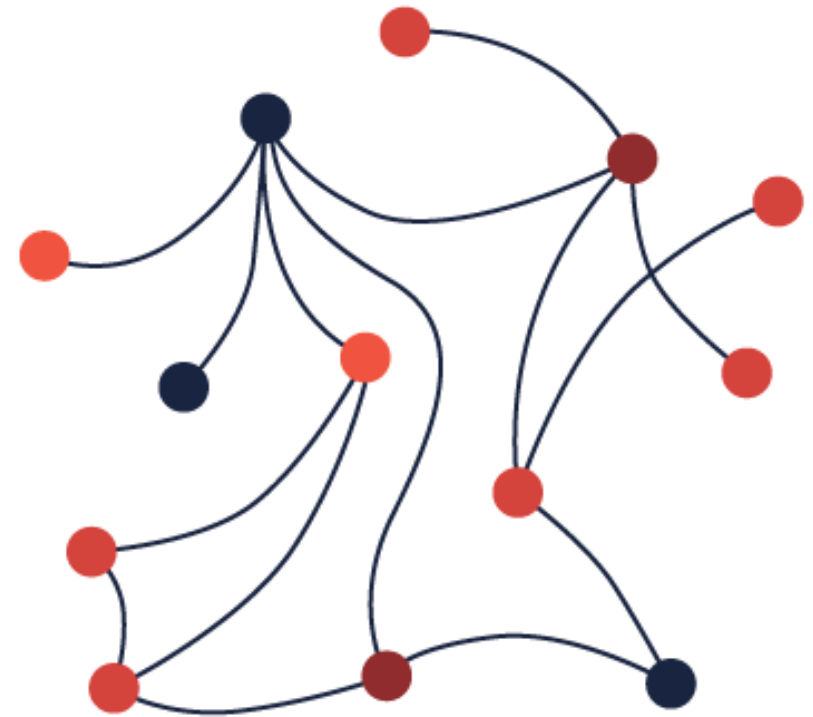
Iris virginica



Network Graph

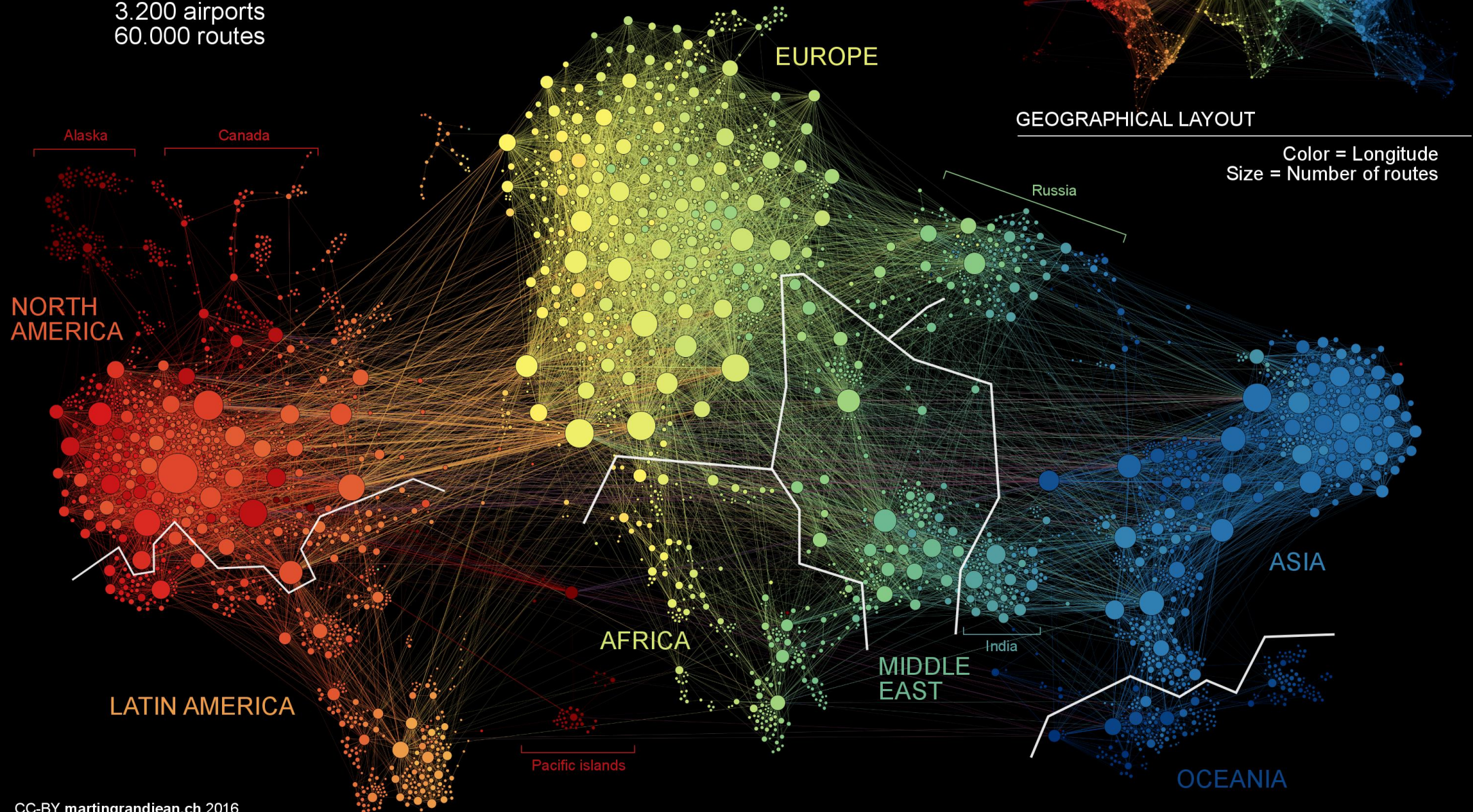
Network graph atau biasa disebut juga dengan network visualization digunakan untuk menggambarkan hubungan yang kompleks diantara elemen dengan jumlah yang sangat banyak.

Network graph dapat menggunakan stuktur undirected graph atau directe graph.



TRANSPORTATION CLUSTERS

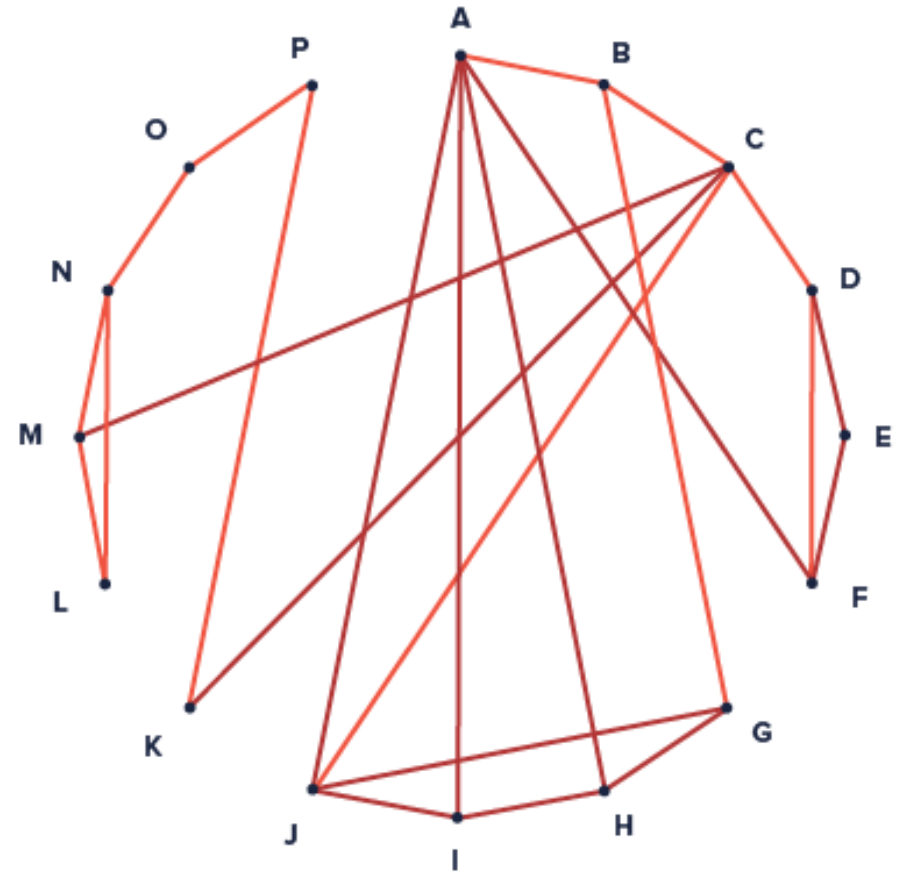
3.200 airports
60.000 routes

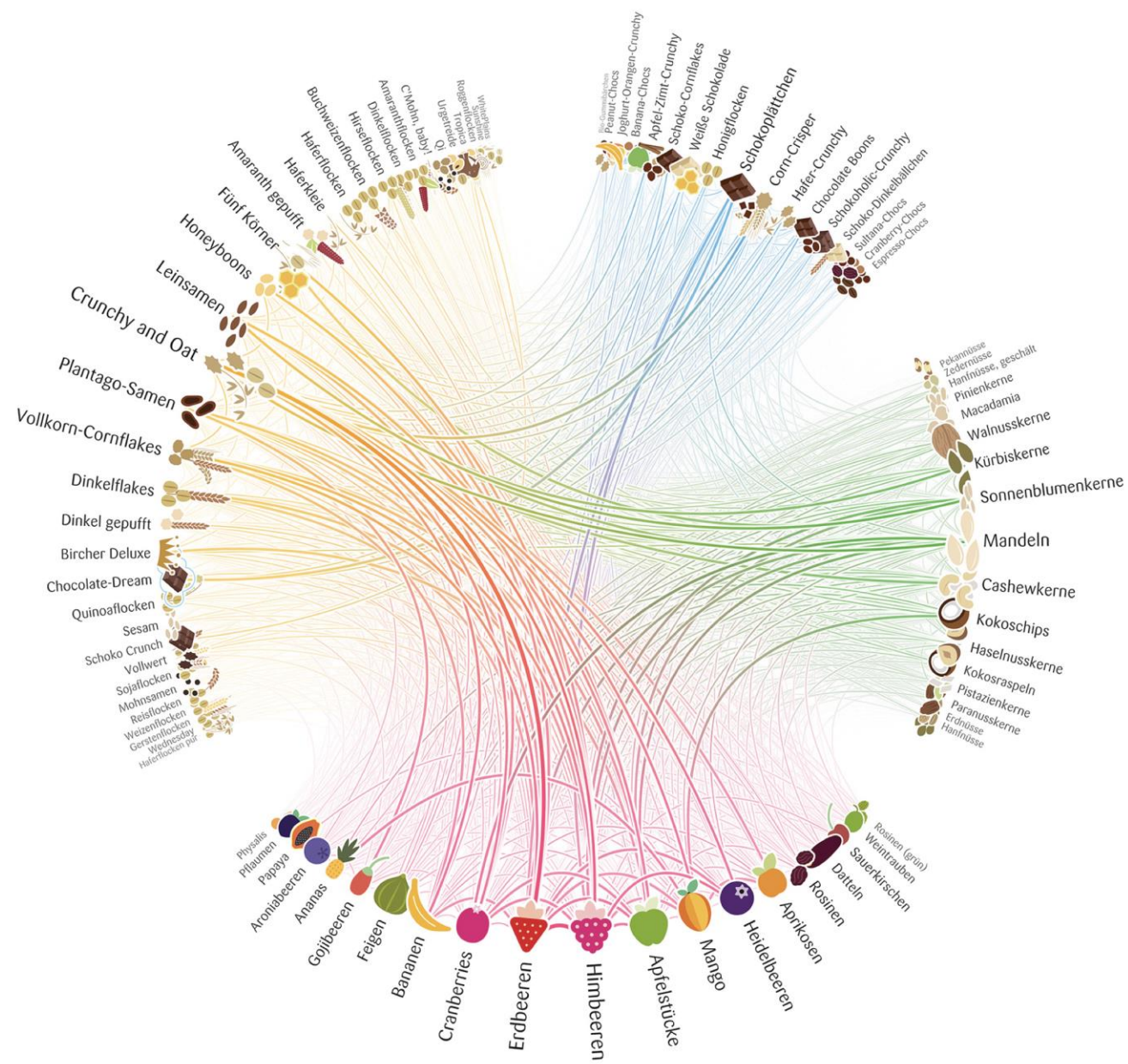


Radial Network

Pada dasarnya radial network adalah network graph yang setiap elemennya disusun secara melingkar.

Garis pada radial network memperlihatkan hubungan antara elemen.







Question ?