



BERCERITA MELALUI VISUALISASI DATA

Let's get to know each other!

Workshop ini diharapkan dapat membantu anda untuk mengerti dan bisa melakukan operasi dasar visualisasi data menggunakan *package* dari Python.

Pengenalan
Visualisasi Data

Library Visualisasi Data

Chart yang sering
digunakan

Hal-hal yang harus
diperhatikan

Course Outline

Pengenalan pada Visualisasi Data

Kenapa kita memvisualisasikan data?

15975369852359897654614876514635498761567813213549876131687463216876541
46546346541341654316584313547651365749879463131031651060190061650649874
05168746106354984103513169874961005101841810260621621616395213957494844
16544165574165364965648065453310045744487225363679984957130453270246751
48764371536255434764483652147024503724168536768574065674169575460652654

15975369852359897654614876514635498761567813213549876131687463216876541
46546346541341654316584313547651365749879463131031651060190061650649874
05168746106354984103513169874961005101841810260621621616395213957494844
16544165574165364965648065453310045744487225363679984957130453270246751
48764371536255434764483652147024503724168536768574065674169575460652654

Order Date	Ship Date	Customer Name	City	State	Region	Category	Sub-Category	Sales	Profit	Quantity
11/08/2018	11/11/2018	Claire Gute	Henderson	Kentucky	South	Furniture	Bookcases	261.96	41.9136	2
11/08/2018	11/11/2018	Claire Gute	Henderson	Kentucky	South	Furniture	Chairs	731.94	219.582	3
06/12/2018	06/16/2018	Darrin Van Huff	Los Angeles	California	West	Office Supplies	Labels	14.62	6.8714	2
10/11/2017	10/18/2017	Sean O'Donnell	Fort Lauderdale	Florida	South	Furniture	Tables	957.5775	-383.031	5
10/11/2017	10/18/2017	Sean O'Donnell	Fort Lauderdale	Florida	South	Office Supplies	Storage	22.368	2.5164	2
06/09/2016	06/14/2016	Brosina Hoffman	Los Angeles	California	West	Furniture	Furnishings	48.86	14.1694	7
06/09/2016	06/14/2016	Brosina Hoffman	Los Angeles	California	West	Office Supplies	Art	7.28	1.9656	4
06/09/2016	06/14/2016	Brosina Hoffman	Los Angeles	California	West	Technology	Phones	907.152	90.7152	6

1. Berikan saya kota-kota dengan profit tertinggi.

Sample - Superstore - Compatibility Mode - Excel

Fajar Ibnu Fatihan

File Home Insert Page Layout Formulas Data Review View Help Search

Clipboard Font Alignment Number Styles Cells Editing Ideas

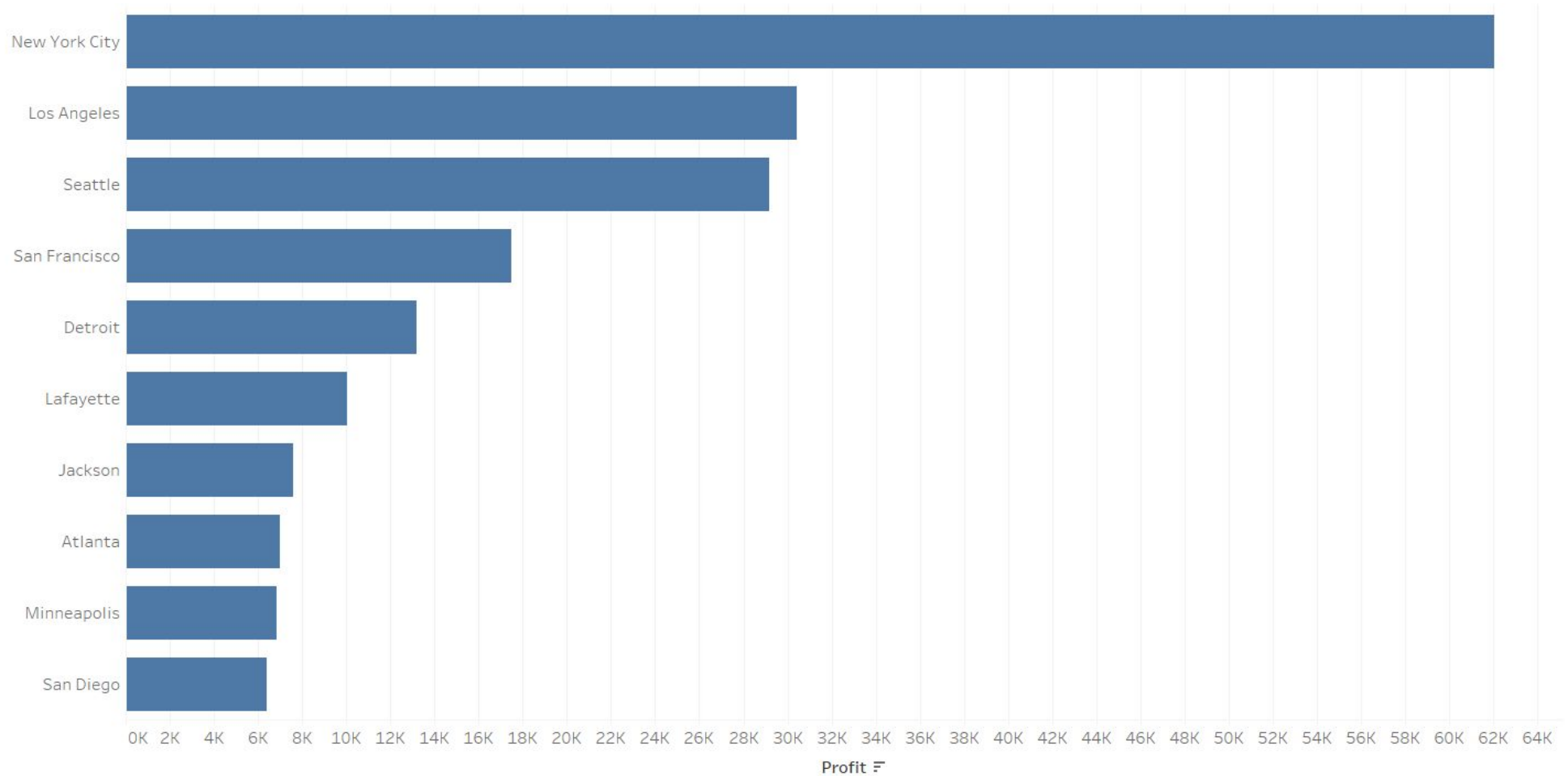
AutoSave Off

O7 X ✓ fx 'Furniture

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country/Region	City	State	Postal Code	Region	Product ID	Category
1	1	CA-2018-152156	11/08/2018	11/11/2018	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-BO-10001798	Furniture
2	2	CA-2018-152156	11/08/2018	11/11/2018	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-CH-10000454	Furniture
3	3	CA-2018-138688	06/12/2018	06/16/2018	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	California	90036	West	OFF-LA-10000240	Office Suppl
4	4	US-2017-108966	10/11/2017	10/18/2017	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	FUR-TA-10000577	Furniture
5	5	US-2017-108966	10/11/2017	10/18/2017	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	OFF-ST-10000760	Office Suppl
6	6	CA-2016-115812	06/09/2016	06/14/2016	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	FUR-FU-10001487	Furniture
7	7	CA-2016-115812	06/09/2016	06/14/2016	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-AR-10002833	Office Suppl
8	8	CA-2016-115812	06/09/2016	06/14/2016	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	TEC-PH-10002275	Technology
9	9	CA-2016-115812	06/09/2016	06/14/2016	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-BI-10003910	Office Suppl
10	10	CA-2016-115812	06/09/2016	06/14/2016	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-AP-10002892	Office Suppl
11	11	CA-2016-115812	06/09/2016	06/14/2016	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	FUR-TA-10001539	Furniture
12	12	CA-2016-115812	06/09/2016	06/14/2016	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	TEC-PH-10002033	Technology
13	13	CA-2019-114412	04/15/2019	04/20/2019	Standard Class	AA-10480	Andrew Allen	Consumer	United States	Concord	North Carolina	28027	South	OFF-PA-10002365	Office Suppl
14	14	CA-2018-161389	12/05/2018	12/10/2018	Standard Class	IM-15070	Irene Maddox	Consumer	United States	Seattle	Washington	98103	West	OFF-BI-10003656	Office Suppl
15	15	US-2017-118983	11/22/2017	11/26/2017	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-AP-10002311	Office Suppl
16	16	US-2017-118983	11/22/2017	11/26/2017	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-BI-10000756	Office Suppl
17	17	CA-2016-105893	11/11/2016	11/18/2016	Standard Class	PK-19075	Pete Kriz	Consumer	United States	Madison	Wisconsin	53711	Central	OFF-ST-10004186	Office Suppl
18	18	CA-2016-167164	05/13/2016	05/15/2016	Second Class	AG-10270	Alejandro Grove	Consumer	United States	West Jordan	Utah	84084	West	OFF-ST-10000107	Office Suppl
19	19	CA-2016-143336	08/27/2016	09/01/2016	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94109	West	OFF-AR-10003056	Office Suppl
20	20	CA-2016-143336	08/27/2016	09/01/2016	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94109	West	TEC-PH-10001949	Technology
21	21	CA-2016-143336	08/27/2016	09/01/2016	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94109	West	OFF-BI-10002215	Office Suppl
22	22	CA-2018-137330	12/09/2018	12/13/2018	Standard Class	KB-16585	Ken Black	Corporate	United States	Fremont	Nebraska	68025	Central	OFF-AR-10000246	Office Suppl
23	23	CA-2018-137330	12/09/2018	12/13/2018	Standard Class	KB-16585	Ken Black	Corporate	United States	Fremont	Nebraska	68025	Central	OFF-AP-10001492	Office Suppl
24	24	US-2019-156909	07/16/2019	07/18/2019	Second Class	SF-20065	Sandra Flanagan	Consumer	United States	Philadelphia	Pennsylvania	19140	East	FUR-CH-10002774	Furniture

Orders People Returns

Jawabannya adalah...



Pentingnya visualisasi data

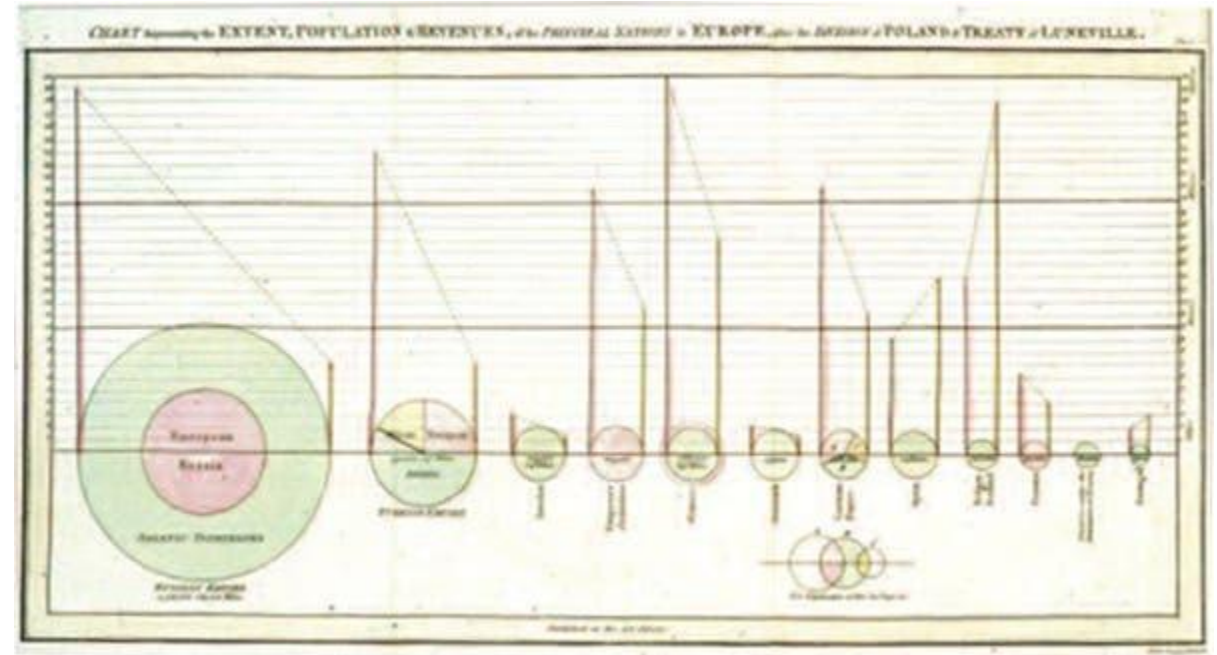
- *Visual representations* merupakan salah satu komponen dasar dalam proses belajar dan pemahaman akan sesuatu.
- Memvisualisasikan sesuatu sudah dilakukan manusia sejak dahulu kala untuk memberi perintah, menyampaikan keinginan dan menceritakan sesuatu.
- Visualisasi data sangat membantu untuk mengidentifikasi pola, relasi, hubungan antar variabel dalam sebuah dataset.

“Human beings are visual creatures. Countless studies show how our brain is wired for the visual and process everything faster when it is through the eye.” –

Badreesh Shetty

Sejarah visualisasi data

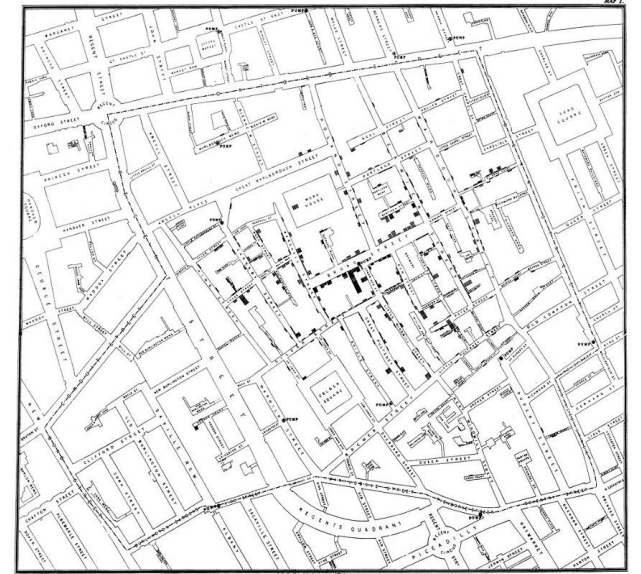
- Salah satu kunci penting dalam sejarah visualisasi data.
- Sebuah pie chart dari William Playfair yang diterbitkan di "Statistical Breviary" pada tahun 1801.



Source: <https://www.nytimes.com/2012/04/22/magazine/who-made-that-pie-chart.html>

Sejarah visualisasi data

- Statistical graphic dari Dr. John Snow yang menunjukkan peta epidemi kolera di London 1855
- Snow menggambarkan setiap sebuah kasus menjadi sebuah dot di peta London.
- Hasilnya menunjukkan bahwa kasus terklastrer di Broad Street
- Ditemukan bahwa ada sebuah pompa air yang digunakan untuk minum terkontaminasi limbah manusia.
- Snow mencabut pompa tersebut dan epidemi pun mereda.

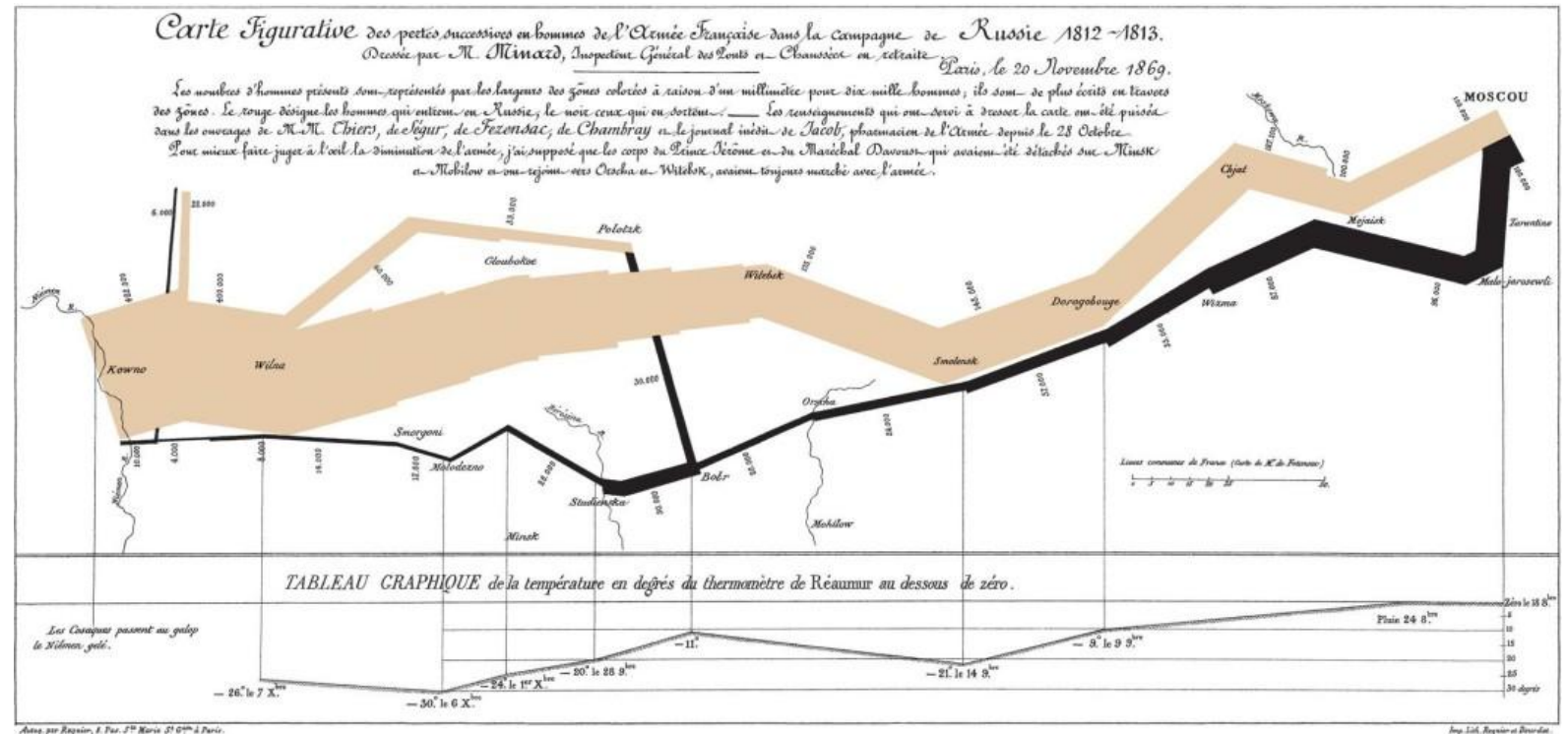


Source: <https://towardsdatascience.com/a-short-history-of-data-visualisation-de2f81ed0b23>

Source: <https://www.theguardian.com/news/datablog/2013/mar/15/john-snow-cholera-map>

Sejarah visualisasi data

- Sebuah chart dari Charles Minard, seorang insinyur teknik sipil pada 1869.
- Ia menggambarkan perjalanan Napoleon dalam menaklukan Eropa dan juga kekalahannya di Moskow.





Library Visualisasi Data

Sekilas mengenai Matplotlib

- Matplotlib merupakan data library yang dibuat diatas NumPy array dan di desain agar cocok dengan SciPy stack.
- Matplotlib mempunyai kesamaan dengan grafik dan visualisasi Matlab.
- Dibuat oleh John Hunter pada 2003.
- Mendapatkan popularitasnya ketika diadopsi sebagai package untuk visualisasi di *Space Telescope Science Institute* yang bertanggung jawab atas Hubble Telescope, yang dari sana mendapatkan dukungan dana untuk pengembangan Matplotlib.
- Bisa bekerja di banyak graphical backend dan operating system.
- Digunakan sebagai dasar dari banyak library python lainnya, seperti Seaborn, ggpy, HoloViews, Altair, bahkan terintegrasi dengan Pandas sekalipun.

Sekilas mengenai Seaborn

- Merupakan library yang dibuat di atas Matplotlib.
- Memberikan interface yang lebih menarik.
- Lebih difungsikan sebagai *statistical plotting library*.
- Terintegrasi dengan Pandas dataframe sangat baik.

Chart yang Sering Digunakan

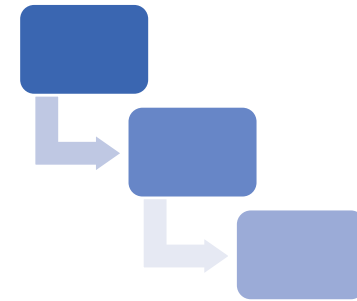
Basic data presentation methods



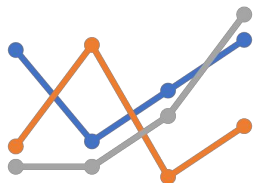
Geo-spatial



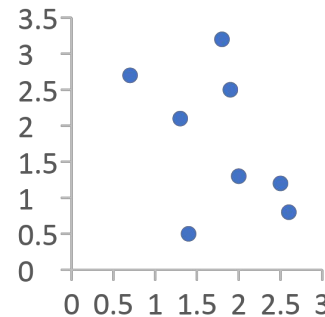
Categorical



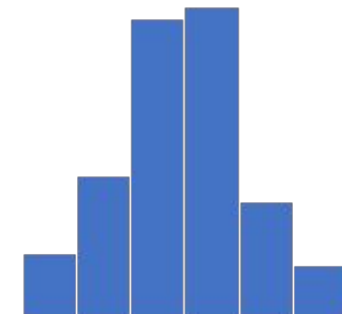
Hierarchical



Time series



Correlation



Distribution

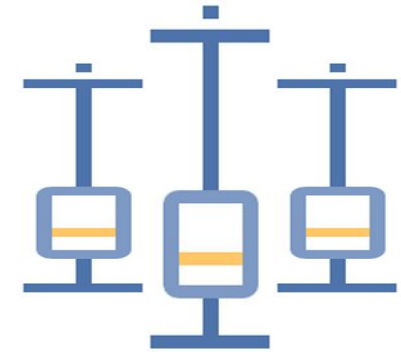
Tipe-tipe dasar grafik



Maps



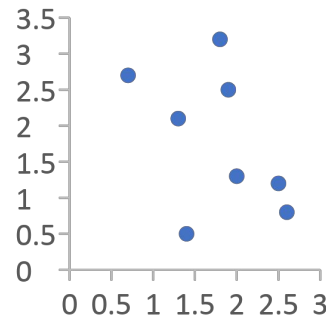
Bar chart



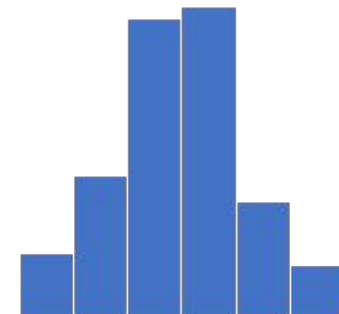
Boxplot



Line chart



Scatterplot



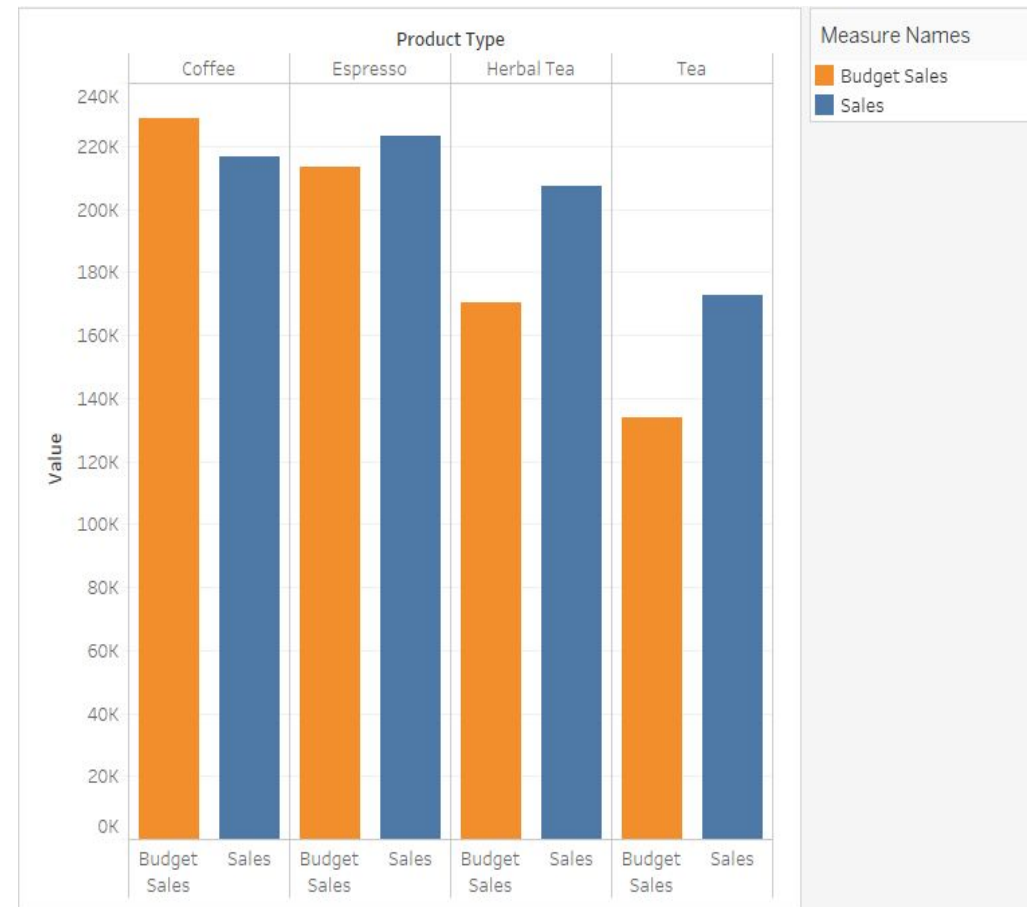
Histograms

Pemilihan chart

Bar chart

Comparing categories:

- Sales by region
- Comparing sales and budget by product type
- Time on site by page
- How many clicks by ads
- Comparing production number by manufacturer

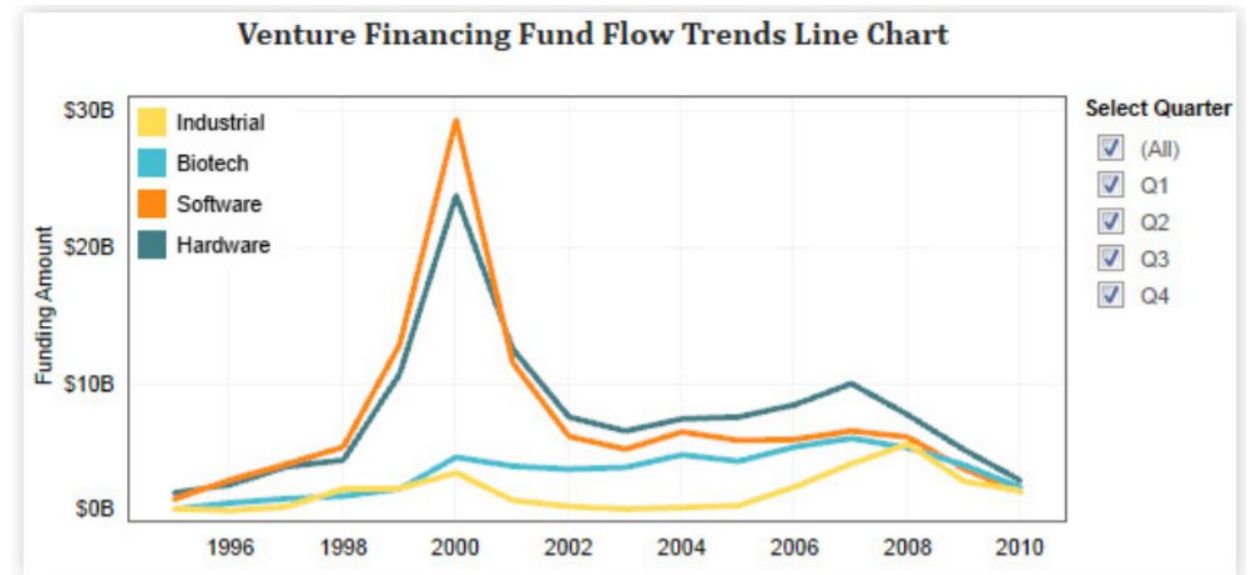


Pemilihan chart (cont.)

Line chart

Trends over time:

- Sales by month
- Profit by year
- Number of store visit per hour
- Year over year growth
- Page visits by day

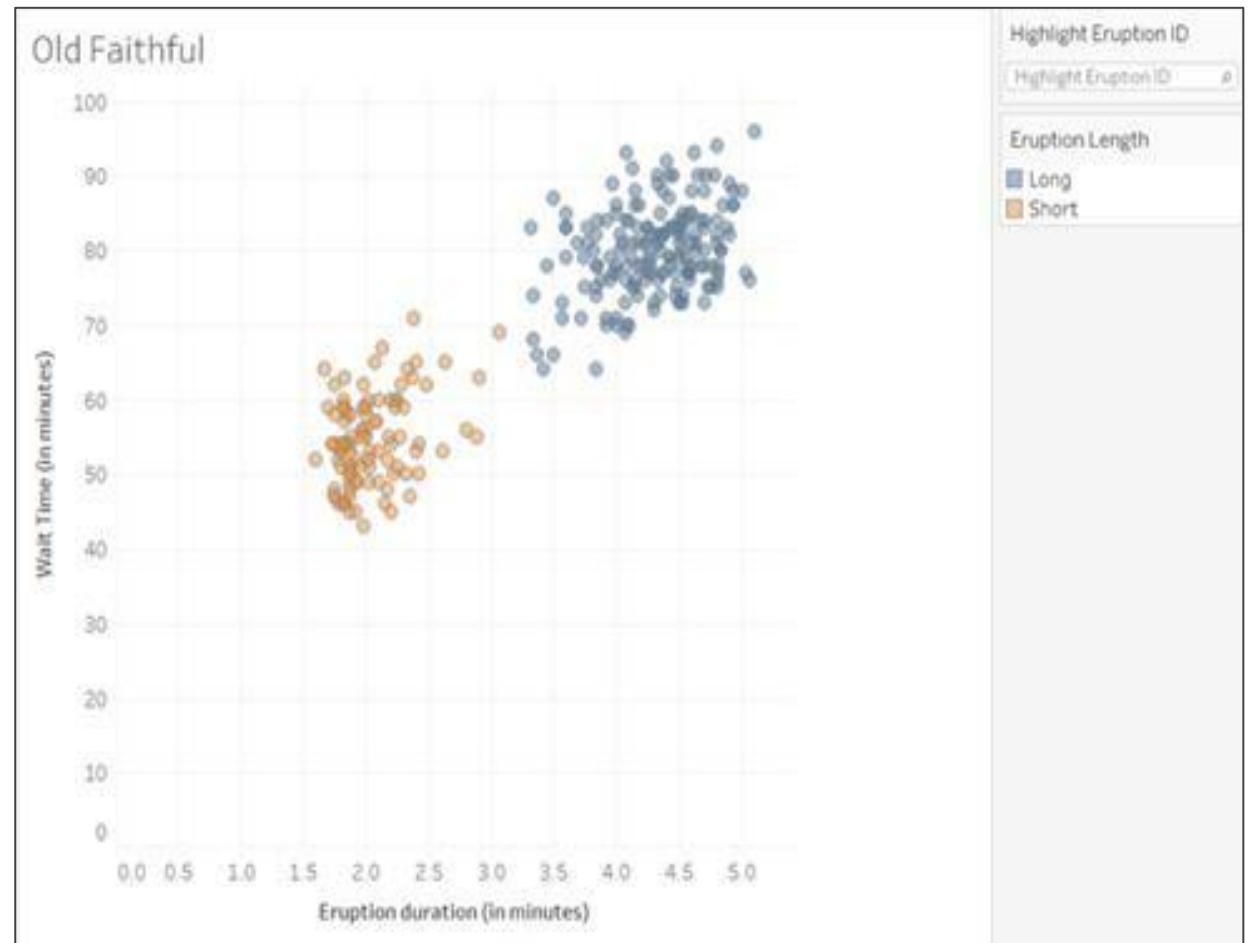


Pemilihan chart (cont.)

Scatterplot

Comparing two numbers
(measures):

- Sales & profit
- Order size & shipping duration
- Customer satisfaction & age
- Calories and body weight
- Cigarettes sales and lung cancer case

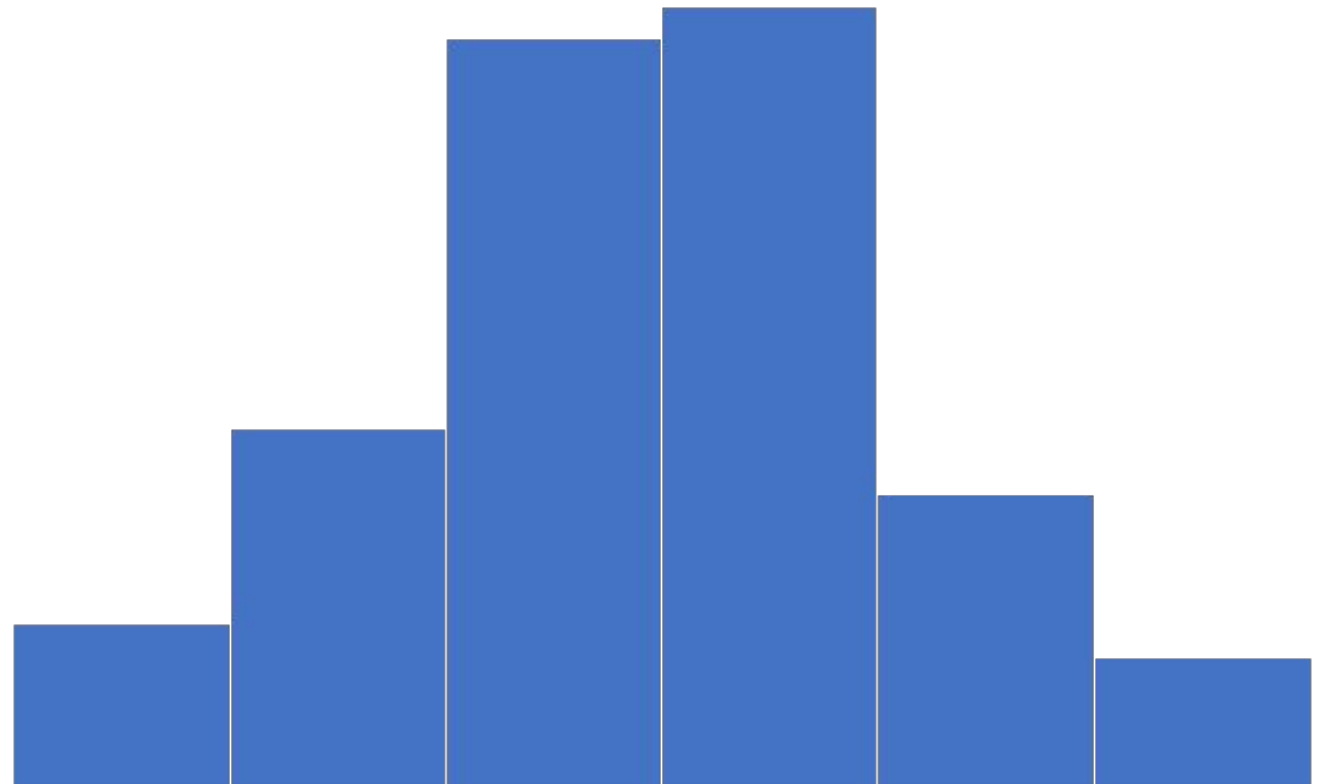


Pemilihan chart (cont.)

Histograms

Looking for distributions:

- Student exam score distributions
- Home price distributions
- Employee salary distributions
- Page load time distributions

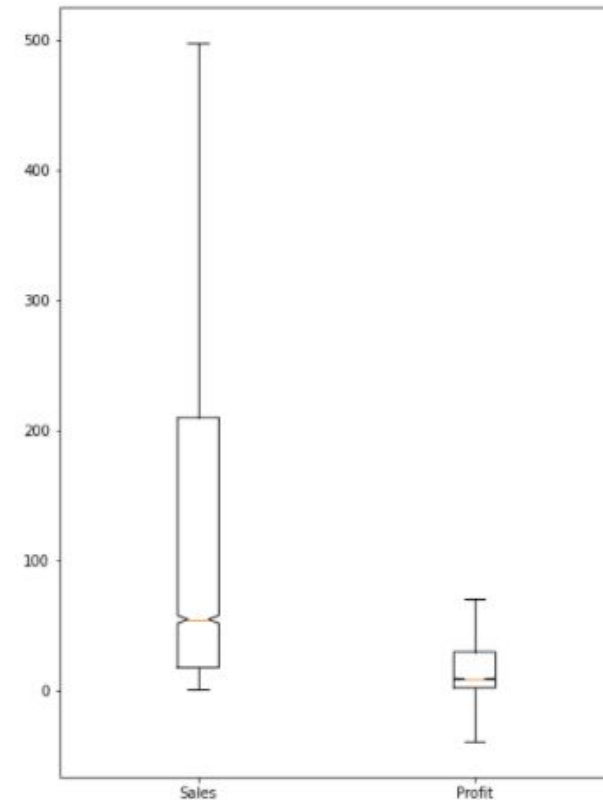


Pemilihan chart (cont.)

Boxplot

Mempunyai fungsi yang mirip dengan *Histogram*. Memberikan kita ringkasan data dengan baik.

- Credit Score of Customer



Pemilihan chart (cont.)

Maps

Regional analysis:

- Answering locational specific questions
- Aid geographical exploration

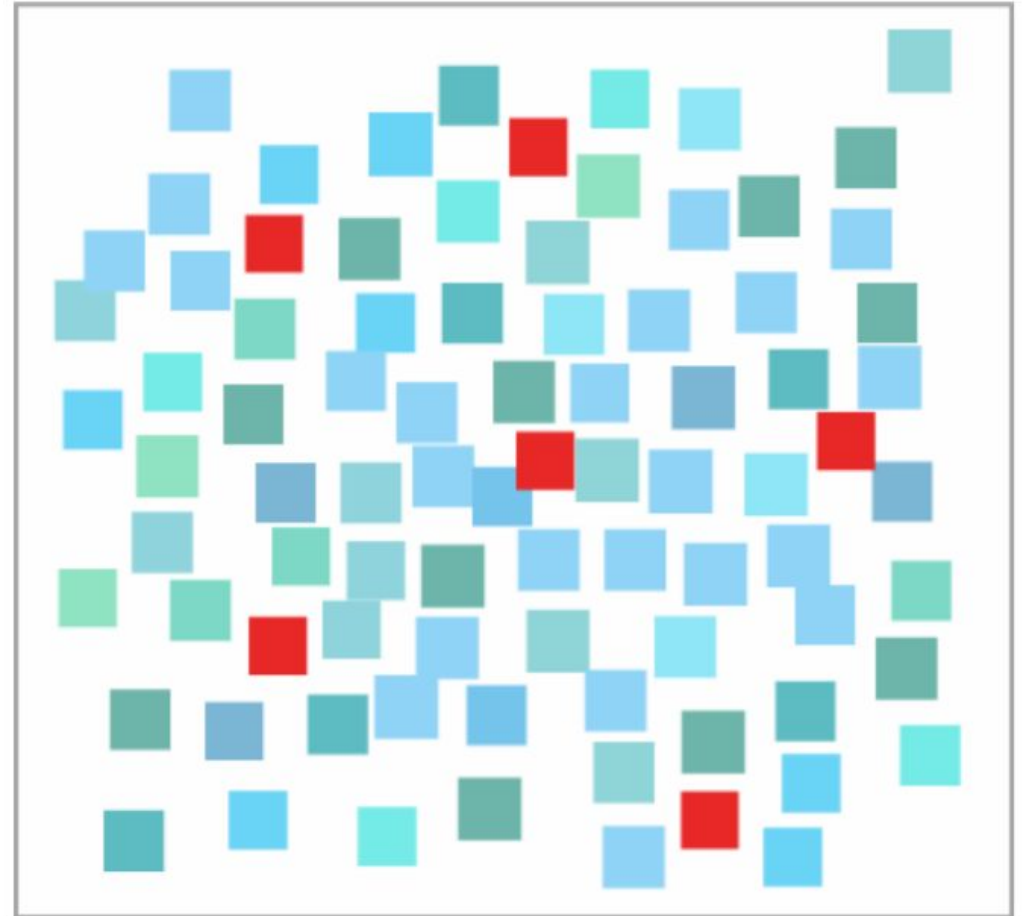


Hal-hal yang harus diperhatikan

Warna

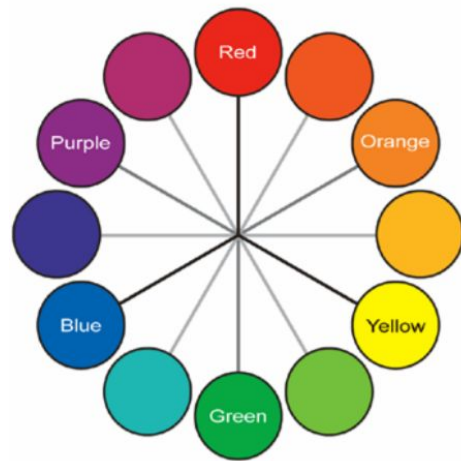
Contrast and Analogy

- Contrast and analogy are the principles that define color design.
- Contrasting colors are different, analogous colors are similar.
- Contrast draws attention, analogy groups.



Source: Choosing Colors for Data Visualization, Maureen Stone, 2006

Warna



Hue

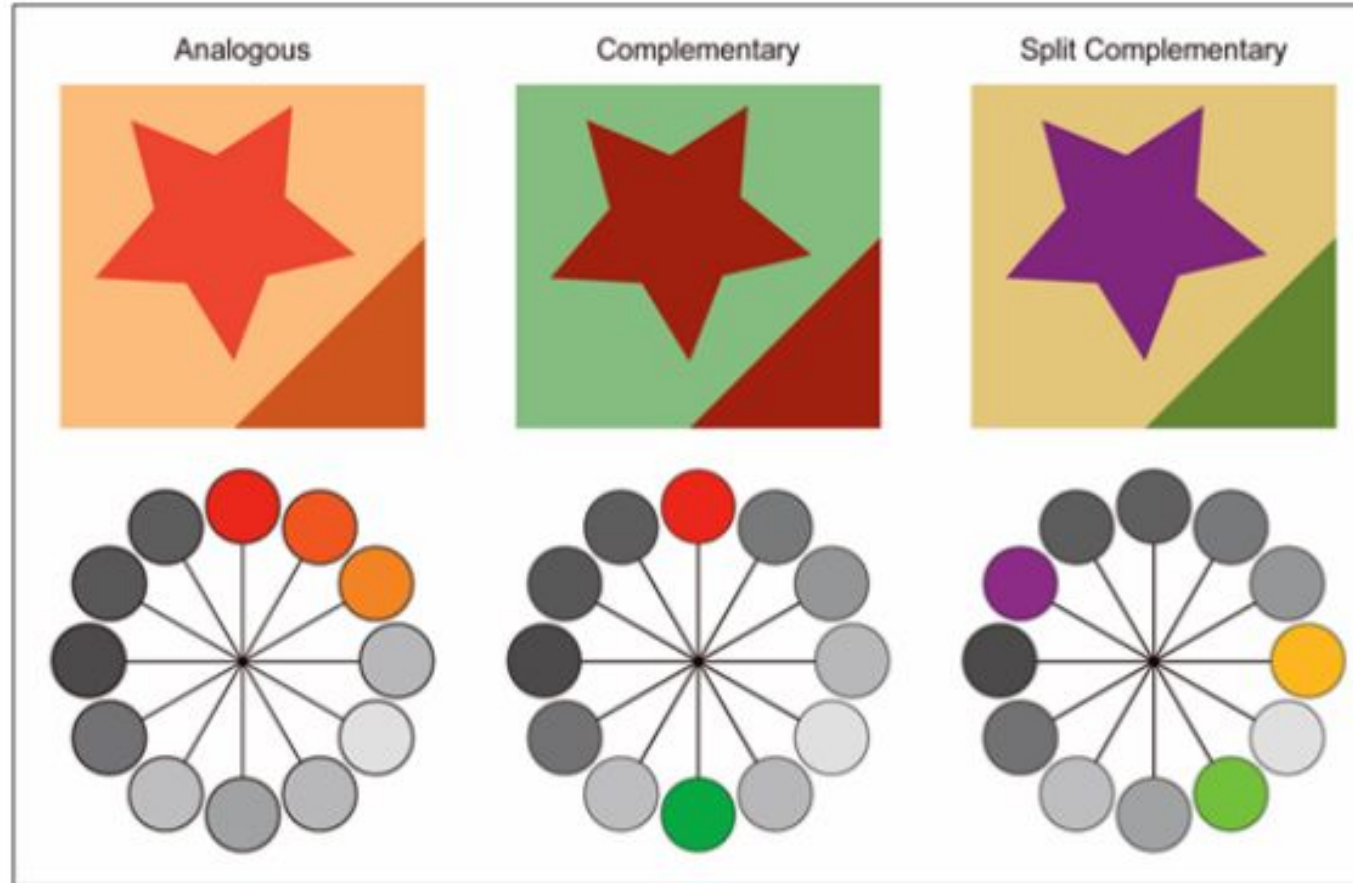


Value
(Lightness)



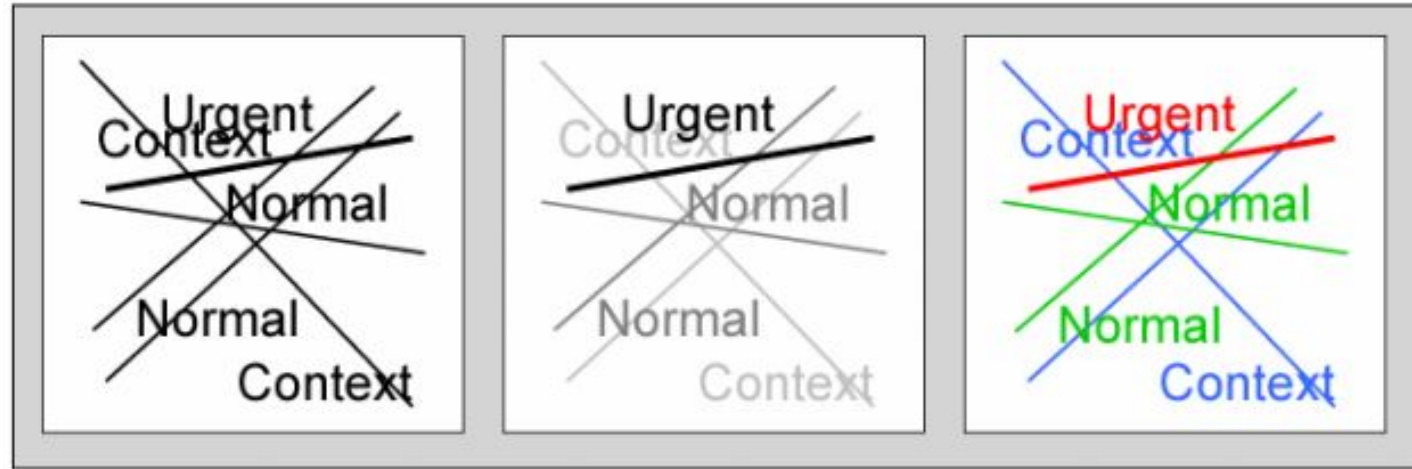
Chroma
(Intensity)

Warna



Source: A Field Guide to Digital Color, Maureen Stone, A. K. Peters, Ltd.

Warna



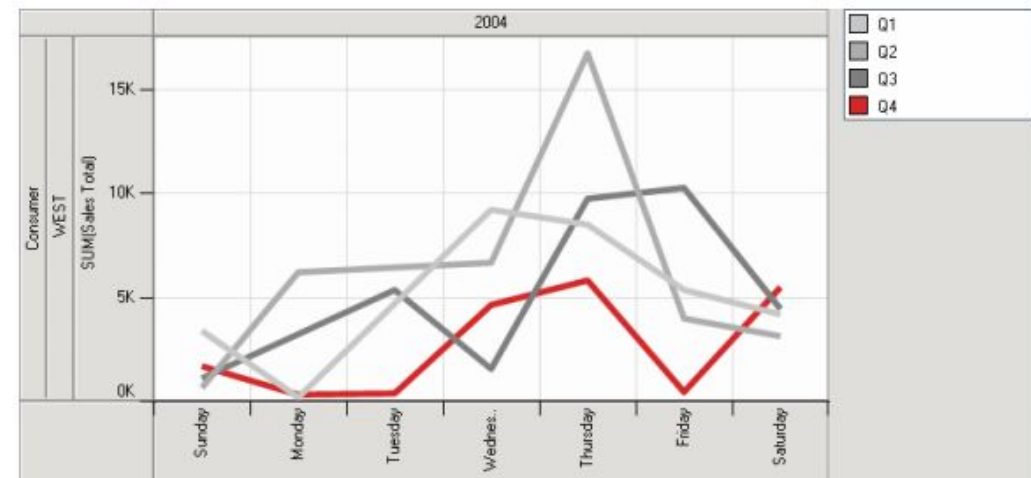
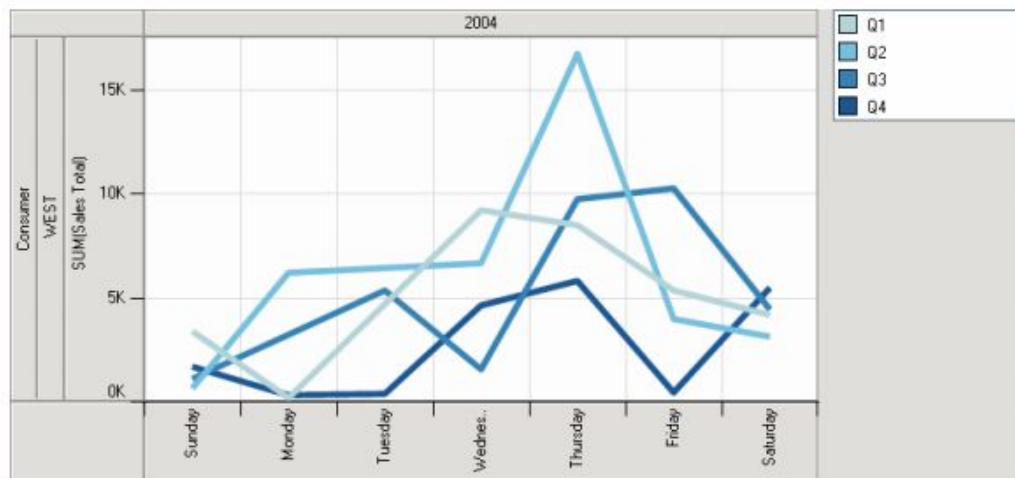
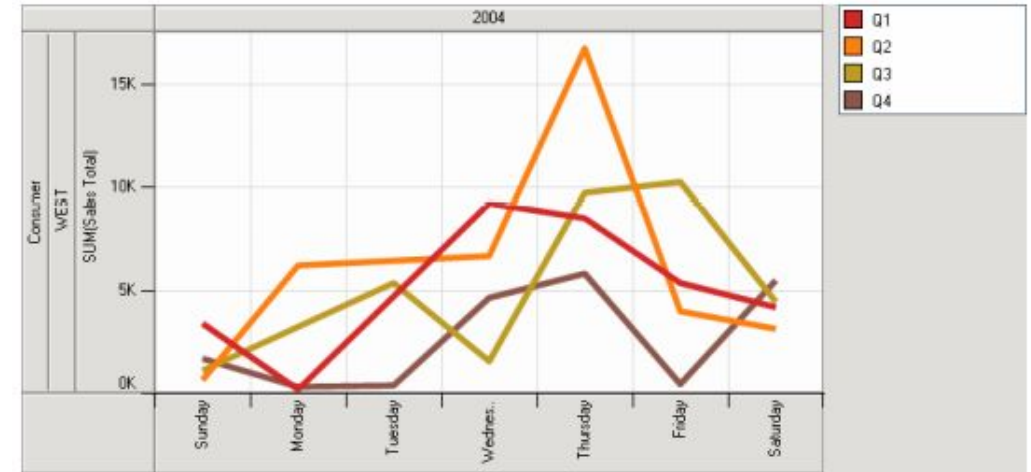
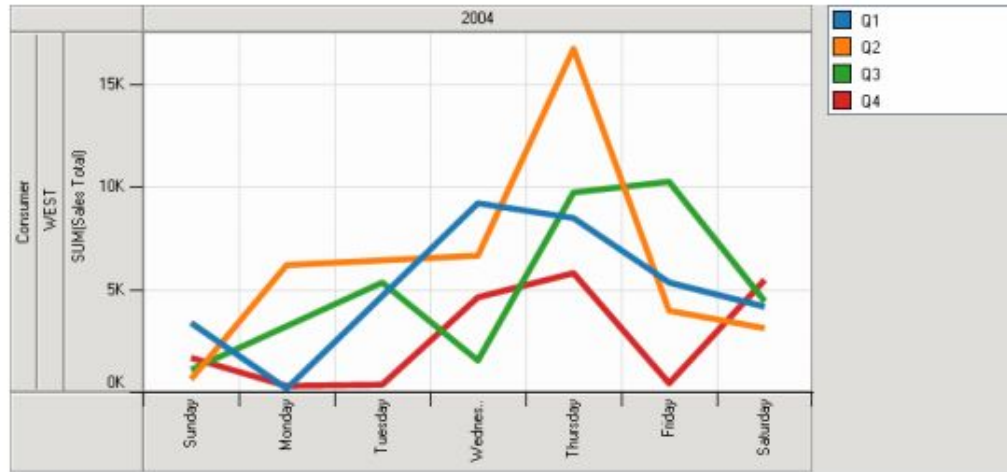
- Color can instantly draw audience attention.
- If we used it correctly can enhance the analysis
- Too many colors creates a visual overload to our audience
- It leads to longer time for analysis

Tips in selecting color:

- Assign color according to function
- Use contrast to highlight
- Control contrast for readability
- Limit color usage to 3 or 4 colors

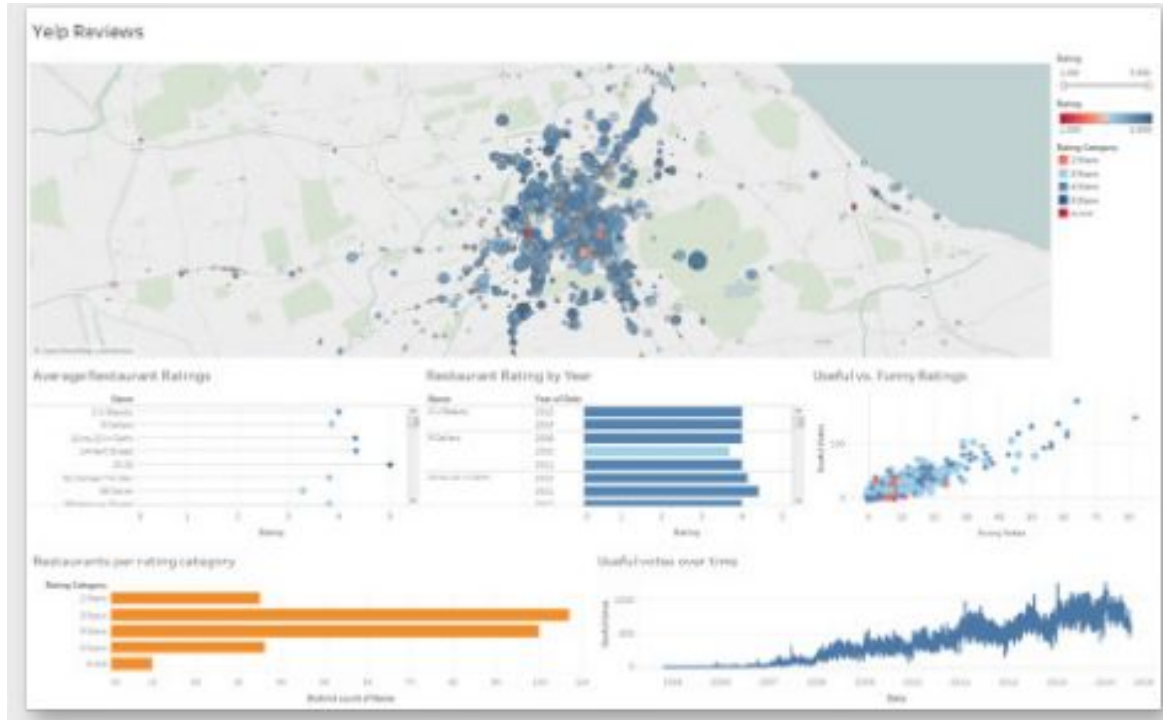
Warna

BERCERITA MELALUI VISUALISASI DATA

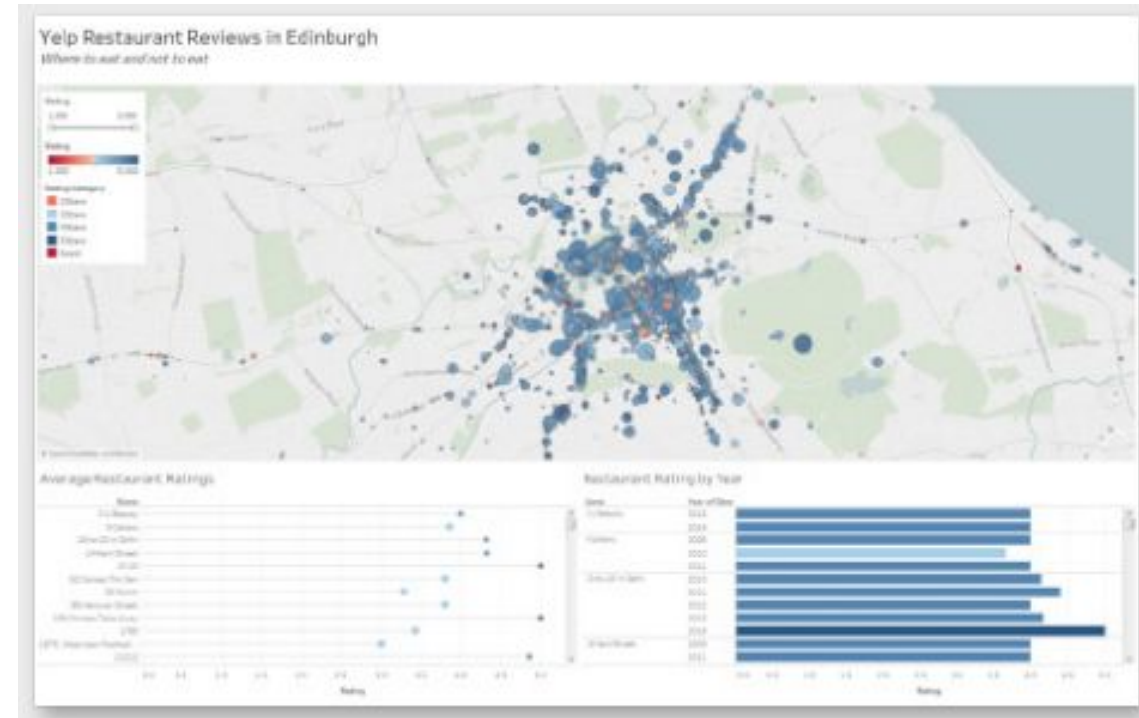


Jumlah chart yang ditampilkan

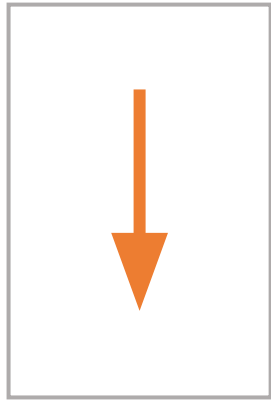
After



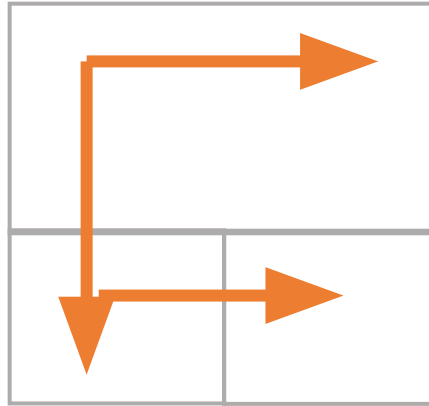
Before



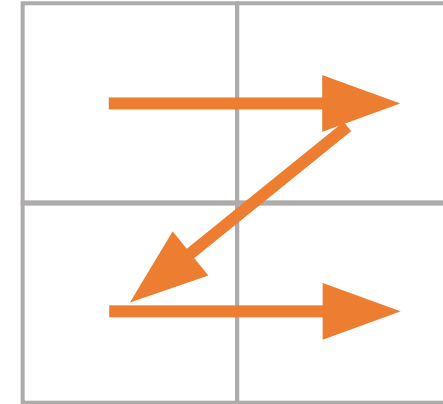
Layout Visualisasi



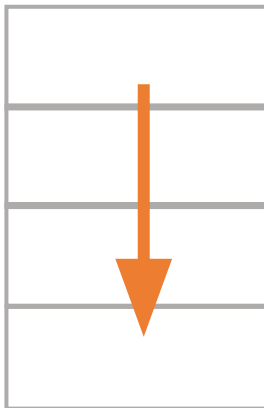
Flow



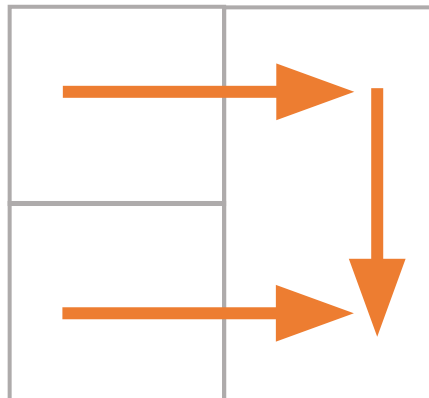
C shape



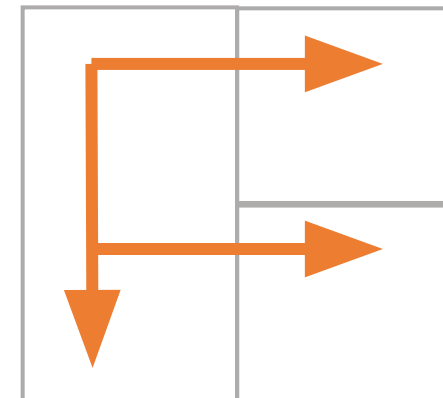
Zig Zag



Chunked



Inverse C



F Natural

Layout Visualisasi

F-shaped natural pattern



Heatmaps from user eye tracking studies of 3 websites. The red areas are areas where users looked the most. Yellow areas indicates fewer views, and blue indicates least-viewed areas.