



Lesson 1

Introduction to Visual Analytics and Applications

AUTHOR

Dr. Kam Tin Seong
Assoc. Professor of Information Systems (Practice)

AFFILIATION

School of Computing and Information Systems,
Singapore Management University

PUBLISHED

04 Jan 2023

What will you learn from this lesson?

- Motivation of Visual Analytics
- What is Visual Analytics
- Learning from the history of visual analytics
- Data Visualisation tools

Welcome to Lesson 1: Introduction to Visual Analytics. In this lesson, I am going to share with you the motivation of this course. This is important because it allows you to understand where I am coming from, as well as getting better understanding to the focus of this course.

This lesson consists of four major sections. First, I will share with you the motivation of this course. This is followed by a discussion on the building block of visual analytics. In the third section, I am going to use four historical data visualisation to differentiate visual analytics from statistical graphics. Lastly, the typology of data visualisation tool kits will be discussed.

Motivation of Visual Analytics

The World is full of charts that lie!



Non-oil domestic exports (Nodx) growth was flat in the second quarter. For the first half this year, Nodx fell 4.5 per cent compared with the same period last year.

Stuttering Nodx performance



We need to get smarter about visual information

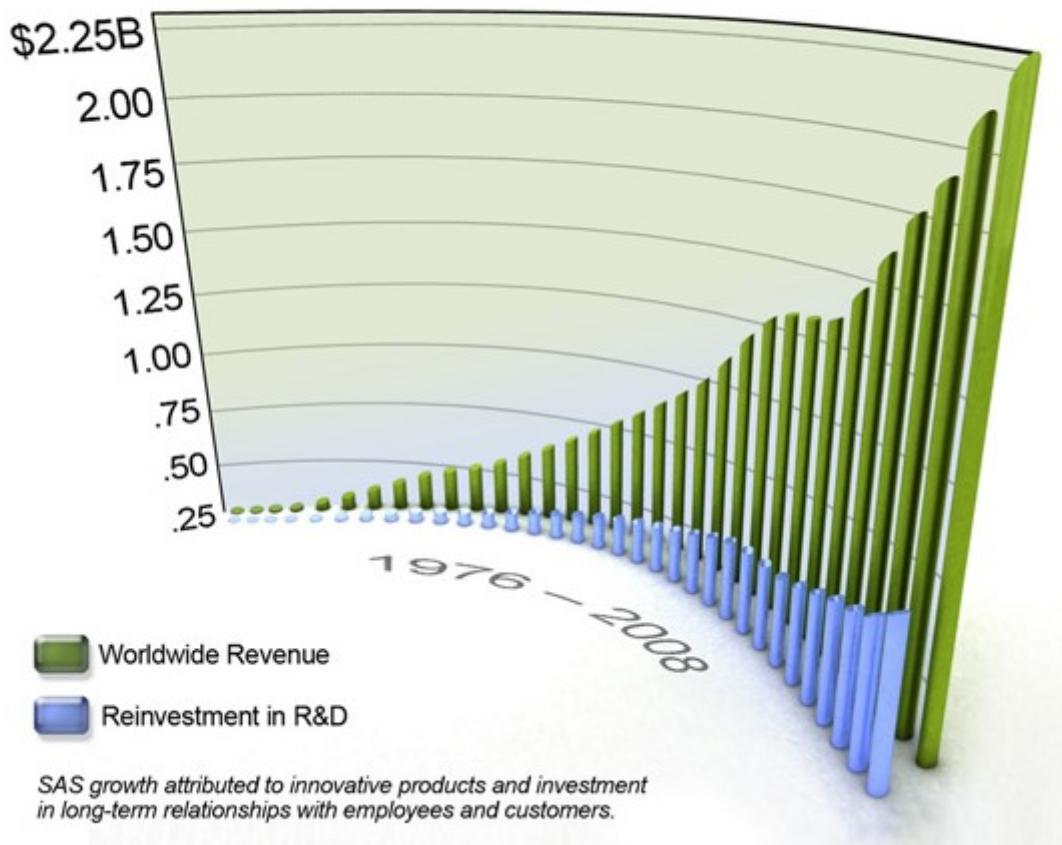
It is very common you see infographic of this kind appears on our daily newspaper. At a first glance, it seems making a lot of sense. But, when we try to take a closer look at the text and graph, you will notice that there are at least two major issues, one is related to truthfulness of the data visualisation and the second one is related to its clarity aspect. For example, the x-axis is showing quarterly time interval. But, the second sentence talks about “For the first half this year, Nodx fell 4.5 per cent compared with the same period last year.” which is not inline with what was shown on the x-axis. Furthermore, the values on the y-axis are in per cent. Hence the correct term should be percentage point instead of per cent because we are comparing the differences between two percentages.

This is what motivate me to design this course. This course is not about using Tableau or R to create beautiful data visualisation only. More importantly, it aims to embark you the skills of getting smarter when reading data visualisation. From today onwards, I would like to encourage you to take a good look at every data visualisation you come across and carefully evaluate them not only from their aestheticness only but also their clarity.

Motivation of Visual Analytics

Junk charts are good for advertising but not for understanding.

SAS Clears \$2.26 Billion



In business, it is very common we see marketing data visualisation similar to the one on this slide. I call this Junk Chart. As we know, junk foods are very tempting but not healthy. Similar, Junk Charts tends to be very appealing but they are not useful to gain data understanding. For example, it is not possible for us to do an effective comparison between global sales revenue and reinvestment in R&D because of the inappropriate use of 3d bar chart.

The important take away from this slide is to avoid creating eye candy data visualisation. We should create data visualisation that are functional instead.

Motivation of Visual Analytics

Data rich, information poor



Source: [Big Data Ocean](#)

In this big data age, data analyst tends to flooded with large amount of data. Before she/he can build any effective model by using advanced AI/ML techniques, she/he need to first understand the data and data visualisation is one of the most commonly used approach.

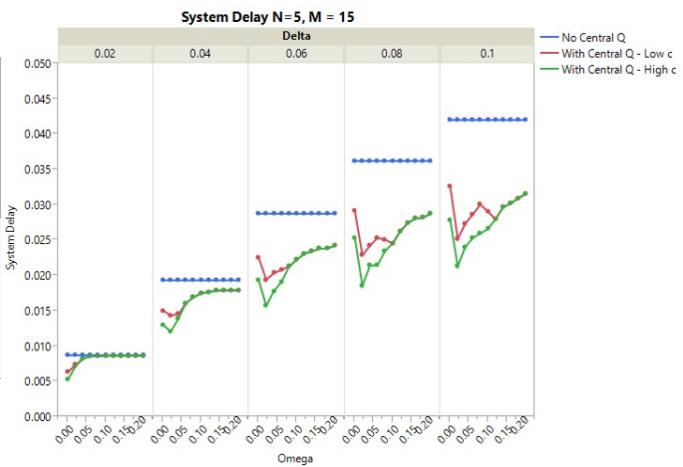
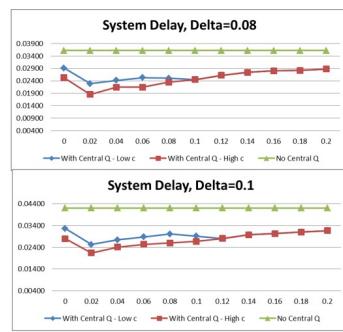
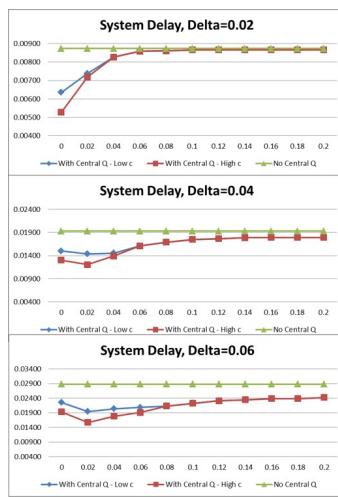
Motivation of Visual Analytics

The World is buried in data,

Just SHOW me the number

| | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | AA | AB | AC | AD | |
|----|---------|----------|------------|-------|----------|-----------------|---------|-------|-------|-------|-------|---------------------|-------------------|-------|--------|--------|---------|---------|---------|---------|--------|---------|---------|---------|---------|---------|---------|--------|--------|-------|
| 1 | Country | Exchange | Name | Forcs | Symbol | ISIN | SEDL | Close | Cap | Cap | Cap | Industry | Supers | 1 Day | 1 Week | 2 Week | 1 Month | 2 Month | 3 Month | 1 Day | Ch | 1 Week | 1 Month | 2 Month | 1 Month | 2 Month | 3 Month | | | |
| 2 | AT | VIE | Erste Gr | EUR | ERST.VI | AT00001288831 | 12.75 | 3E+03 | 3E+03 | 3E+03 | 3E+03 | Financial Banks | 13.75 | 13.63 | 10.27 | 7.75 | 11.84 | 16.2 | -0.073 | -0.075 | -0.063 | -0.085 | 0.245 | 0.2683 | 0.6452 | 0.7131 | 0.0763 | 0.151 | -0.2 | |
| 3 | AT | VIE | Raiffeisn | EUR | RIBK.VI | AT00001207047 | 21.21 | 1E+03 | 1E+03 | 1E+03 | 1E+03 | Financial Banks | 22.65 | 25 | 19.35 | 14.62 | 16 | 19.3 | -0.064 | -0.065 | -0.052 | -0.1638 | 0.0632 | 0.0861 | 0.4508 | 0.5107 | 0.3256 | 0.3734 | 0.05 | |
| 4 | AT | VIE | voestalp | EUR | VOES.VI | AT00001294340 | 9.85 | 1E+03 | 1E+03 | 1E+03 | 1E+03 | Basic Ml Basic Re | 10.11 | 10.47 | 11.19 | 12.4 | 15.22 | 15.04 | -0.026 | -0.028 | -0.053 | -0.0734 | -0.12 | -0.101 | -0.206 | -0.173 | -0.353 | -0.323 | -0.3 | |
| 5 | AT | VIE | Vieorbz | EUR | WBSV.VI | AT00001265931 | 5.93 | 5E+03 | TE+00 | TE+00 | TE+00 | Industri Costrut | 6.15 | 5.6 | 5.5 | 6.57 | 10.06 | 11.3 | -0.056 | -0.053 | 0.0563 | 0.05622 | 0.0752 | 0.1015 | -0.057 | -0.06 | -0.411 | -0.383 | -0.56 | |
| 6 | AT | VIE | Intercell | EUR | ICEL.VI | AT000012067M8 | 23.2 | 5E+03 | 1E+03 | 1E+03 | 1E+03 | Health C Health C | 21.22 | 21.48 | 21.43 | 22.92 | 25.68 | 21.84 | 0.0333 | 0.0303 | 0.0301 | 0.05631 | 0.0736 | 0.1023 | 0.0122 | 0.054 | -0.031 | 0.064 | 0.062 | |
| 7 | AT | VIE | Adtritz | EUR | ANDR.V | AT000012191VPE | 23.16 | 5E+03 | 1E+03 | 1E+03 | 1E+03 | Industri Industri | 21.47 | 21.51 | 19.73 | 24 | 23.3 | 16.16 | 0.0817 | 0.0764 | 0.0167 | 0.05362 | 0.1703 | 0.1655 | -0.035 | 0.0048 | -0.006 | 0.0236 | 0.21 | |
| 8 | AT | VIE | Vistas In | EUR | VIGR.VI | AT000012062KSS | 21.63 | 6E+03 | 1E+03 | 1E+03 | 1E+03 | Financial Insurance | 22.05 | 21.3 | 18.23 | 18.11 | 24.6 | 24.12 | -0.019 | -0.021 | 0.085 | -0.0063 | 0.1865 | 0.2121 | 0.1283 | 0.1743 | -0.121 | -0.089 | -0.1K | |
| 9 | AT | VIE | OMV AC | EUR | OMVVA.VI | AT000012651453 | 25.2 | 4E+03 | 5E+03 | 5E+03 | 5E+03 | Oil & Gas Oil & Gas | 24.23 | 23.84 | 21.23 | 20.72 | 22.33 | 18.72 | 0.0375 | 0.0352 | 0.057 | 0.0438 | 0.1887 | 0.2126 | 0.2162 | 0.2664 | 0.1255 | 0.1661 | 0.34 | |
| 10 | AT | VIE | Telekom | EUR | TELA.VI | AT000012463508 | 11.4 | 4E+03 | 5E+03 | 5E+03 | 5E+03 | Telecom Telecom | 10.86 | 10.43 | 9.35 | 11 | 10.3 | 0.0437 | 0.0474 | 0.0867 | 0.0634 | 0.0867 | 0.1102 | 0.1451 | 0.153 | 0.0364 | 0.0731 | 0.10 | | |
| 11 | AT | VIE | Verbund | EUR | VERB.VI | AT000012661607 | 28.59 | 2E+03 | 3E+03 | 3E+03 | 3E+03 | Utilities Utilities | 28.36 | 29.32 | 26.12 | 25 | 23.87 | 32.56 | -0.019 | -0.005 | -0.044 | -0.0649 | 0.0946 | 0.1863 | 0.1436 | 0.1508 | -0.046 | -0.1 | | |
| 12 | AT | VIE | Atrium E | EUR | ATRV.V | AT000012558684 | 2.27 | 5E+00 | TE+00 | TE+00 | TE+00 | Financial Real Est. | 2.32 | 2.02 | 2.15 | 1.47 | 2.18 | 2.58 | -0.022 | -0.024 | 0.1238 | 0.0366 | 0.0558 | 0.1236 | 0.5442 | 0.6843 | 0.0413 | 0.1301 | -0 | |
| 13 | AU | ASX | Bendigo | AUD | BENAX.AX | AU000012609120 | 8.19 | 2E+03 | 1E+03 | 1E+03 | 1E+03 | Financial Banks | 7.32 | 7.4 | 6.3 | 6.84 | 10 | 11 | 0.0341 | 0.0346 | 0.1058 | 0.0556 | 0.13 | 0.473 | 0.1974 | 0.3737 | -0.161 | -0.035 | -0.21 | |
| 14 | AU | ASX | Syncorp | AUD | SUNAX.AX | AU000012650504 | 6 | 8E+03 | 5E+03 | 5E+03 | 5E+03 | Financial Banks | 6.14 | 6.17 | 5.4 | 5.2 | 7.5 | 8.31 | -0.029 | -0.022 | 0.028 | 0.0241 | 0.1111 | 0.1503 | 0.158 | 0.1578 | -0.2 | 0.237 | -0.2 | |
| 15 | AU | ASX | National | AUD | NAB.AX | AU000012624601 | 20.1 | 4E+10 | 3E+10 | 3E+10 | 3E+10 | Financial Banks | 21.03 | 13.73 | 18.35 | 17.51 | 18.33 | 20.86 | -0.047 | -0.046 | 0.057 | 0.0306 | 0.0354 | 0.1302 | 0.1473 | 0.2522 | 0.0618 | 0.1833 | -0 | |
| 16 | AU | ASX | Bank of I | AUD | BOD.AX | AU000012607624 | 8.44 | 1E+03 | 1E+03 | 1E+03 | 1E+03 | Financial Banks | 8.3 | 7.34 | 7.04 | 6.84 | 7.4 | 7.45 | 0.0163 | 0.0174 | 0.063 | 0.0566 | 0.1883 | 0.2333 | 0.4578 | 0.1405 | 0.3635 | -0.1 | | |
| 17 | AU | ASX | Australia | AUD | ANZ.AX | AU000012605586 | 15.75 | 3E+03 | 2E+10 | 2E+10 | 2E+10 | Financial Banks | 16.55 | 15.25 | 14.05 | 13.05 | 13.27 | 15.25 | -0.048 | -0.048 | 0.0328 | 0.02607 | 0.0281 | 0.2459 | 0.2063 | 0.3562 | 0.1863 | 0.3713 | 0.03 | |
| 18 | AU | ASX | Westpac | AUD | WBC.AX | AU000012076146 | 19.03 | 6E+10 | 4E+10 | 4E+10 | 4E+10 | Financial Banks | 19.14 | 18.35 | 17.12 | 16.65 | 15.64 | 16.63 | -0.003 | -0.002 | 0.0074 | 0.00033 | 0.1144 | 0.2223 | 0.1485 | 0.2753 | 0.2206 | 0.0324 | 0.141 | |
| 19 | AU | ASX | Common | AUD | CBA.AX | AU000012629033 | 34.73 | 3E+10 | 4E+10 | 4E+10 | 4E+10 | Financial Banks | 35 | 33.00 | 31 | 28.8 | 28.8 | 28.3 | -0.008 | -0.007 | -0.07 | -0.0984 | 0.1203 | 0.2417 | 0.1655 | 0.3078 | 0.2391 | 0.4875 | 0.22 | |
| 20 | AU | ASX | Electro | AUD | BLX.AX | AU000012653203 | 2.57 | 2E+03 | 2E+03 | 2E+03 | 2E+03 | Basic Ml Basic Re | 2.73 | 2.55 | 2.25 | 2.92 | 3.35 | 3.45 | -0.073 | -0.0987 | 0.0078 | 0.0123 | 0.1422 | 0.2514 | 0.103 | 0.2235 | -0.272 | -0.171 | -0.2 | |
| 21 | AU | ASX | CSR Ltd | AUD | CSRAX.AX | AU000012623864 | 1.2 | 2E+03 | 1E+03 | 1E+03 | 1E+03 | Basic Ml Basic Re | 1.25 | 1.15 | 1.17 | 1.02 | 1.425 | 1.73 | -0.053 | -0.053 | 0.048 | 0.0412 | 0.0256 | 0.3614 | 0.1765 | 0.5812 | -0.158 | 0.162 | -0.1 | |
| 22 | AU | ASX | Iluka Res | AUD | ILUX.AX | AU000012657515 | 4.03 | 1E+03 | 1E+03 | 1E+03 | 1E+03 | Basic Ml Basic Re | 4.1 | 4.4 | 4.2 | 4.32 | 2.45 | 4.44 | -0.011 | -0.011 | -0.054 | -0.03 | -0.04 | 0.0045 | -0.057 | -0.007 | -0.052 | 0.0361 | -0.05 | |
| 23 | AU | ASX | OneSteel | AUD | OSTAX.AX | AU000012623721 | 2.25 | 2E+03 | 1E+03 | 1E+03 | 1E+03 | Basic Ml Basic Re | 2.38 | 2.15 | 1.65 | 1.8 | 2.43 | 2.33 | -0.055 | -0.054 | 0.0321 | 0.0254 | 0.2465 | 0.3102 | 0.25 | 0.3302 | -0.074 | 0.0107 | -0.01 | |
| 24 | AU | ASX | Alumina I | AUD | AVCA.X | AU000012654585 | 1.28 | 2E+03 | 1E+03 | 1E+03 | 1E+03 | Basic Ml Basic Re | 1.24 | 1.21 | 1 | 1.35 | 1.14 | 1.28 | 0.0328 | 0.0328 | 0.0573 | 0.05037 | 0.28 | 0.3454 | 0.0111 | 0.1333 | 0.1228 | 0.2266 | | |
| 25 | AU | ASX | Neurex | AUD | NCMA.X | AU000012651701 | 32.73 | 2E+10 | 1E+10 | 1E+10 | 1E+10 | Basic Ml Basic Re | 33.54 | 33.66 | 31.71 | 31 | 33.03 | 32.88 | -0.024 | -0.024 | -0.028 | -0.034 | 0.0322 | 0.1518 | 0.091 | 0.2327 | 0.0558 | 0.2248 | -0.1 | |
| 26 | AU | ASX | EHP Billi | AUD | BHPAX.AX | AU000012644650 | 31.91 | 1E+11 | 1E+10 | 1E+10 | 1E+10 | Basic Ml Basic Re | 34.65 | 33.92 | 34 | 30.6 | 30.5 | 29.65 | -0.062 | -0.061 | 0.042 | 0.0485 | 0.0234 | 0.0819 | 0.1517 | 0.1673 | 0.0462 | 0.1431 | 0.07 | |
| 27 | AU | ASX | Lilir Gol | AUD | LGLAX.AX | P00006 | 5101596 | 3.28 | 8E+03 | 5E+03 | 5E+03 | 5E+03 | Basic Ml Basic Re | 3.23 | 3.4 | 2.37 | 3.21 | 3.13 | 3.03 | 0.0185 | 0.016 | -0.035 | 0.0416 | 0.1044 | 0.2597 | 0.0031 | 0.1523 | 0.0473 | 0.2367 | 0.008 |
| 28 | AU | ASX | Fortescue | AUD | FMGAX.AX | AU000012608625 | 2.55 | 5E+03 | 5E+03 | 5E+03 | 5E+03 | Basic Ml Basic Re | 2.6 | 2.3 | 2.31 | 2.55 | 1.77 | 1.68 | -0.015 | 0.015 | 0.1037 | 0.1018 | 0.1033 | 0.1603 | -0.004 | 0.06 | 0.4407 | 0.5741 | 0.55 | |
| 29 | AU | ASX | Poladis F | AUD | PDNAX.AX | AU000012668456 | 3.4 | 2E+03 | 1E+03 | 1E+03 | 1E+03 | Basic Ml Basic Re | 3.57 | 3.47 | 3.16 | 2.75 | 2.41 | 2.36 | 0.045 | 0.045 | -0.047 | -0.02 | -0.0265 | 0.0753 | 0.1309 | 0.2364 | 0.0357 | 0.0968 | 0.184 | |
| 30 | AU | ASX | Rio Tinto | AUD | RIOAX.AX | AU000012622010 | 56.57 | 2E+10 | 1E+10 | 1E+10 | 1E+10 | Basic Ml Basic Re | 56.88 | 50.53 | 50.75 | 46.3 | 42.05 | 37.05 | -0.005 | -0.005 | 0.0115 | 0.1125 | 0.1147 | 0.1776 | 0.2218 | 0.0302 | 0.3421 | 0.4664 | 0.526 | |
| 31 | AU | ASX | Indesic Pl | AUD | IPLEX.AX | AU000012673041 | 2.12 | 3E+03 | 2E+03 | 2E+03 | 2E+03 | Basic Ml Chemical | 2.2 | 2.15 | 2.03 | 2.18 | 2.51 | 2.4 | -0.036 | -0.036 | -0.014 | -0.0204 | -0.113 | -0.048 | -0.028 | 0.0566 | -0.0206 | -0.114 | -0.1 | |
| 32 | AU | ASX | Orica Ltc | AUD | ORIAX.AX | AU0000126458001 | 14.73 | 5E+03 | 4E+03 | 4E+03 | 4E+03 | Basic Ml Chemical | 15.14 | 14.15 | 15.59 | 15.54 | 13.8 | 13.65 | -0.023 | -0.023 | -0.023 | 0.0027 | 0.0352 | 0.0712 | 0.0811 | 0.0505 | 0.0923 | 0.1535 | 0.07 | |
| 33 | AU | ASX | Leighton | AUD | LEIA.X | AU000012651227 | 15.3 | 3E+03 | 2E+03 | 2E+03 | 2E+03 | Industry Construt | 20.16 | 13.25 | 13.24 | 13.03 | 16.16 | 26.23 | -0.043 | -0.042 | 0.0266 | -0.0039 | 0.0031 | 0.0543 | 0.0195 | 0.0764 | 0.1516 | 0.2582 | -0.2 | |
| 34 | AU | ASX | Boral Ltc | AUD | BLDAX.AX | AU000012621670 | 5.61 | 2E+03 | 1E+03 | 1E+03 | 1E+03 | Industry Construt | 3.68 | 3.41 | 2.86 | 2.92 | 3.11 | 4.66 | -0.015 | 0.015 | 0.0587 | 0.0576 | 0.0262 | 0.2367 | 0.3156 | 0.0306 | 0.1917 | -0.2 | | |
| 35 | AU | ASX | James Hc | AUD | JHXAX.AX | AU000012412977 | 4.2 | 2E+03 | 1E+03 | 1E+03 | 1E+03 | Industry Construt | 4.07 | 4.05 | 3.75 | 3.25 | 3.32 | 4.45 | 0.0319 | 0.0325 | 0.0317 | 0.03029 | 0.12 | 0.1772 | 0.2923 | 0.3752 | 0.0714 | 0.1701 | -0.01 | |
| 36 | AU | ASX | Lend Leas | AUD | LLCAX.AX | AU000012612004 | 6.43 | 3E+03 | 2E+03 | 2E+03 | 2E+03 | Industry Construt | 6.43 | 6.25 | 5.82 | 5.06 | 6.15 | 7.03 | 0.0303 | 0.0338 | 0.0384 | 0.0384 | 0.151 | 0.1721 | 0.2326 | 0.3643 | -0.044 | 0.1943 | -0.0 | |
| 37 | AU | ASX | Dowmier I | AUD | DOWAX.AX | AU000012645573 | 4.45 | 1E+03 | 1E+03 | 1E+03 | 1E+03 | Industry Construt | 4.61 | 4.29 | 4.15 | 3.81 | 3.76 | 3.33 | -0.035 | -0.034 | 0.0337 | 0.0326 | 0.0223 | 0.1217 | 0.1453 | 0.2237 | 0.25 | 0.3658 | 0.115 | |
| 38 | AU | ASX | United G | AUD | UGLAX.AX | AU0000126701H1 | 9.16 | 2E+03 | 1E+03 | 1E+03 | 1E+03 | Industry Construt | 3.23 | 7.33 | 7.5 | 7.6 | 6.02 | 8 | -0.005 | 0.005 | | | | | | | | | | |

Spreadsheet: A great hammer but lousy Visual analytics come to the rescue screwdriver



As all of you will notice that the data visualisation on the left were created using Microsoft Excel charting function. It shows the results of five models output. Unfortunately, it is not easy to compare the differences between them. This is partly because the value range of their y-axis are not the same. Further more by stitching the line graph top down and next jump to the right fail for support effective comparison.

The data visualisation on the right is a makeover version. It was created using the trellis method of JMP Pro. In this case, all line graphs maintain the same value range. As a result, the data visualisation reveals the differences between the five models clearly.

Also, by arranging the model results from left to right according to their calibration, allows us to compare the performance of each model effectively.

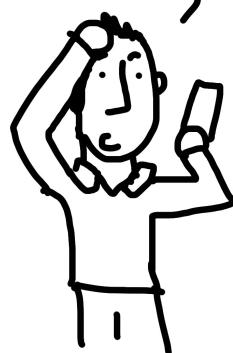
Motivation of Visual Analytics

When statistics failed!

| SALESFORCE A | SALES_REVENUE A | SALESFORCE B | SALES_REVENUE B | SALESFORCE C | SALES_REVENUE C | SALESFORCE D | SALES_REVENUE D |
|-----------------|--------------------|-----------------|--------------------|-----------------|--------------------|-----------------|--------------------|
| 10 | 8.04 | 10 | 9.14 | 10 | 7.46 | 8 | 6.58 |
| 8 | 6.95 | 8 | 8.14 | 8 | 6.77 | 8 | 5.76 |
| 13 | 7.58 | 13 | 8.74 | 13 | 12.74 | 8 | 7.71 |
| 9 | 8.81 | 9 | 8.77 | 9 | 7.11 | 8 | 8.84 |
| 11 | 8.33 | 11 | 9.26 | 11 | 7.81 | 8 | 8.47 |
| 14 | 9.96 | 14 | 8.1 | 14 | 8.84 | 8 | 7.04 |
| 6 | 7.24 | 6 | 6.13 | 6 | 6.08 | 8 | 5.25 |
| 4 | 4.26 | 4 | 3.1 | 4 | 5.39 | 19 | 12.5 |
| 12 | 10.84 | 12 | 9.11 | 12 | 8.15 | 8 | 5.56 |
| 7 | 4.82 | 7 | 7.26 | 7 | 6.42 | 8 | 7.91 |
| 5 | 5.68 | 5 | 4.74 | 5 | 5.73 | 8 | 6.89 |

| Summary Statistics | | Summary Statistics | | Linear Fit | | | |
|--------------------|-----------|--------------------|-----------|--------------------------------------------------|----------|--|--|
| Mean | 9 | Mean | 7.4990909 | SALES_REVENUE = 3.0000909 + 5.0000909*SALESFORCE | | | |
| Std Dev | 3.3166248 | Std Dev | 2.0300614 | Summary of Fit | | | |
| Std Err Mean | 1 | Std Err Mean | 0.6120865 | RSquare | 0.666542 | | |
| Upper 95% Mean | 11.228139 | Upper 95% Mean | 8.8629047 | RSquare Adj | 0.629492 | | |
| Lower 95% Mean | 6.7718611 | Lower 95% Mean | 6.1352771 | Root Mean Square Error | 1.236603 | | |
| N | 11 | N | 11 | Mean of Response | 7.500909 | | |
| | | | | Observations (or Sum Wgts) | 11 | | |

I just can't make sense
of this data.



Have you tried
looking at the
pictures?



fresh spectrum

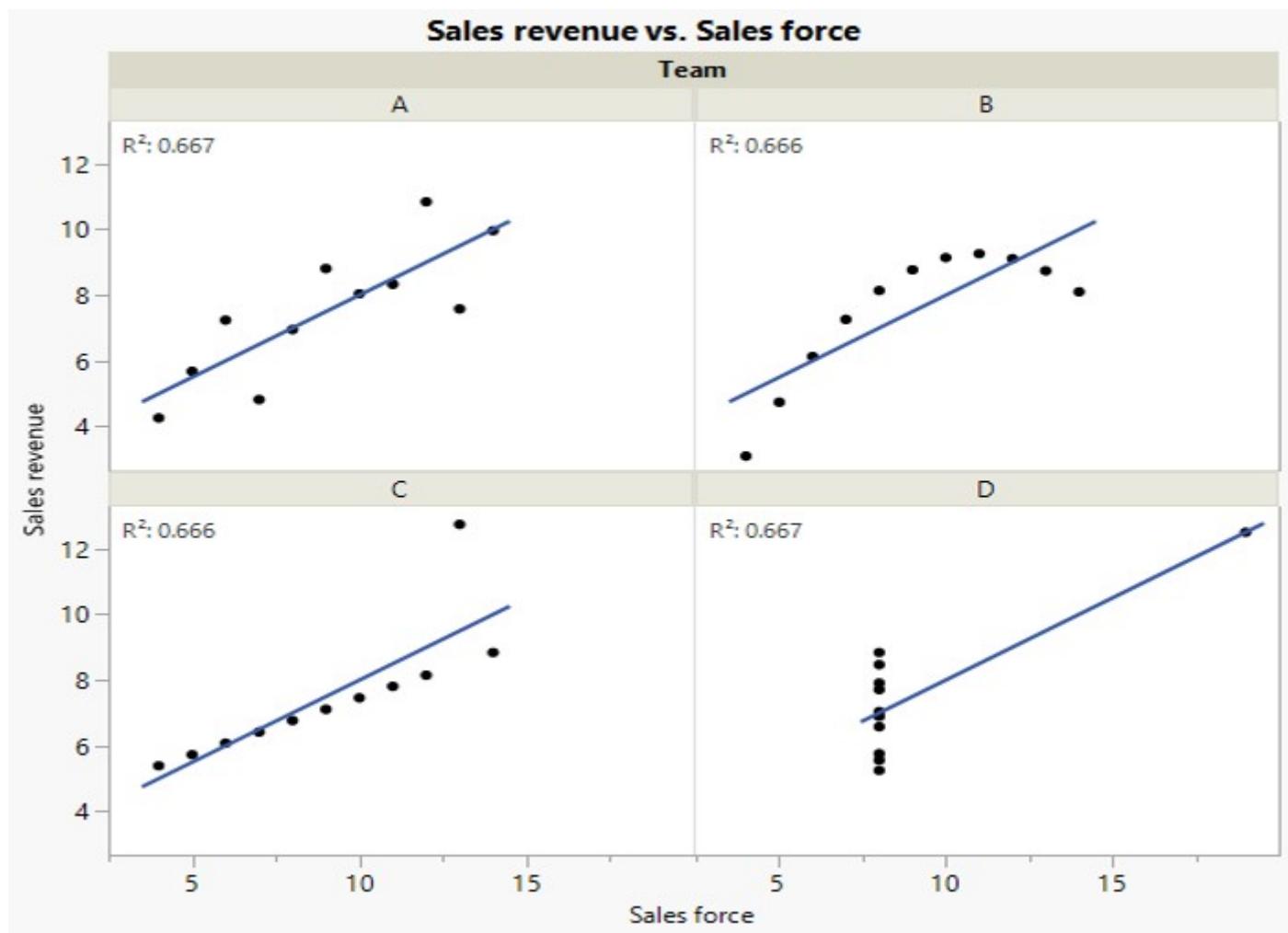
Data analysis without data visualisation is no data analysis

As a data analyst, we tend to use statistics to describe data distribution. However, these might not be enough without data visualisation.

This slide shows a data sets of four groups of observation. When their summary statistics such as mean and standard deviation were computed. It is interesting to note that they are all having the similar values. In fact, even the correlation coefficient and best fit line equations are similar.

Motivation of Visual Analytics

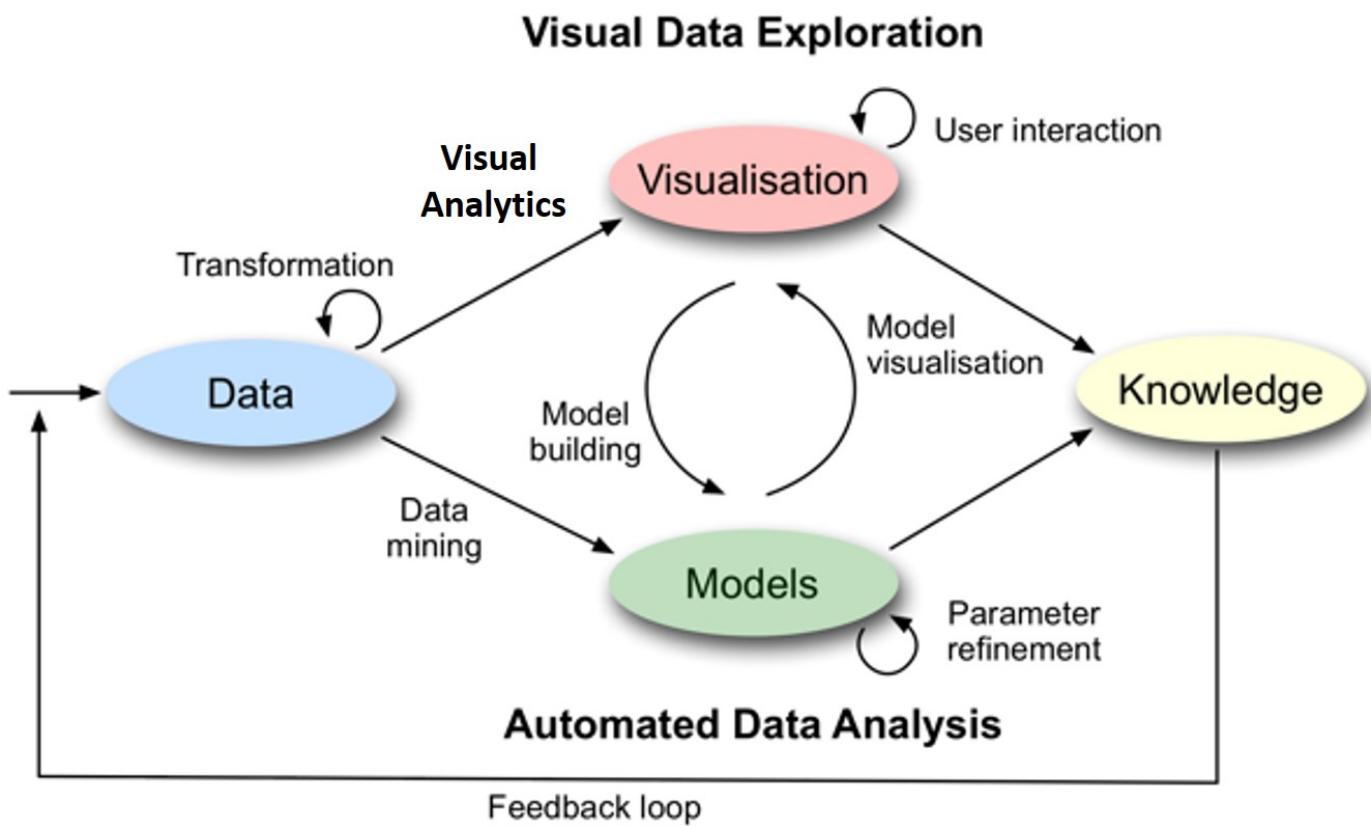
Analytical graphics reveal the truth



However, when we plot these data by using scatter plot method as shown on the slide, it is interesting to note that they reveal very different patterns.

Motivation of Visual Analytics

Is there a place for Visual Analytics in knowledge discovery?



The important take away from this example is that it is always useful to use data visualisation to complement data modelling. As shown in the figure. In a knowledge discovery process, visual analytics provide data analysts useful understanding of the data. These understanding can be used to support effective model building.

Data visualisation also should be used to reveal modelling results. Furthermore, by using data visualisation to communicate complex modelling results to senior management, in general, are more effectively than showing them complex equations.

Motivation of Visual Analytics

Open data need to be unlocked!



Data.gov.sg

Search Singapore's Public Data



View All Datasets



Economy



Education



Environment



Finance



Health



Infrastructure



Society



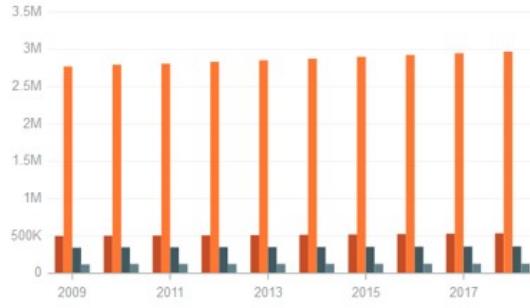
Technology



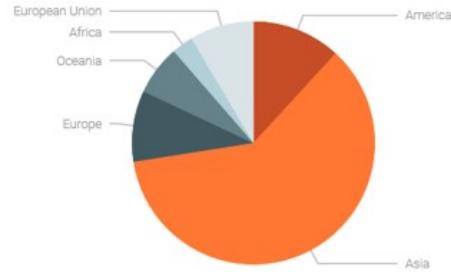
Transport

Singapore at a glance

Singapore Residents By Ethnic Group, End June, Annual (2018)



Domestic Exports By Area (2014)

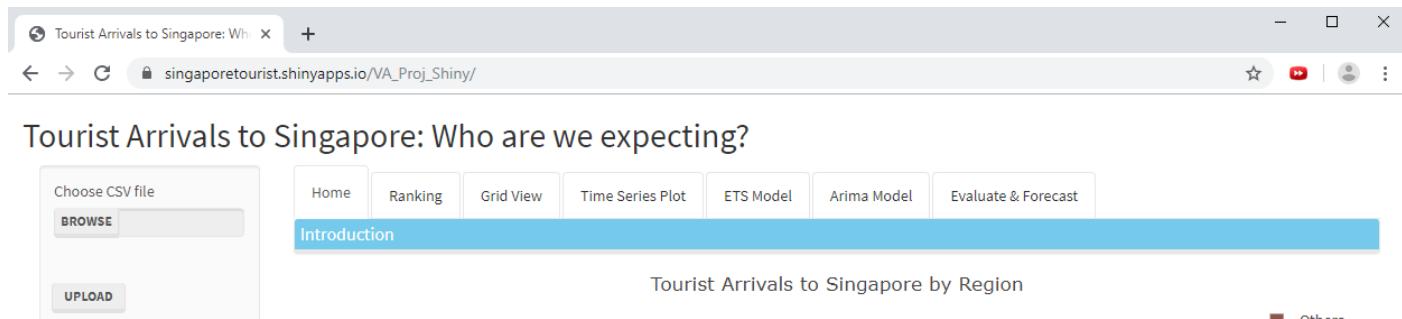


Visit the [link](#)

Recently, there is an increasing available of open source data from the public sector. However, to put these data into good use for informed decision making, we also need appropriate analytics tools. By and large, the COS analytics software are too costly and beyond the reach of SME and casual users. On the other hand, free and open source analytical APIs tend to require certain level of proficiency in programming which is not a common skill possessed by business analysts.

Motivation of Visual Analytics

Democratising Data and Analytics with Visually-driven Web Apps



Visit the [link](#)

In view of this, I strongly believe that it is important for our MITB (Analytics) students master the skill of building visually-driven web-based data analytics applications to democratise data and analytics.

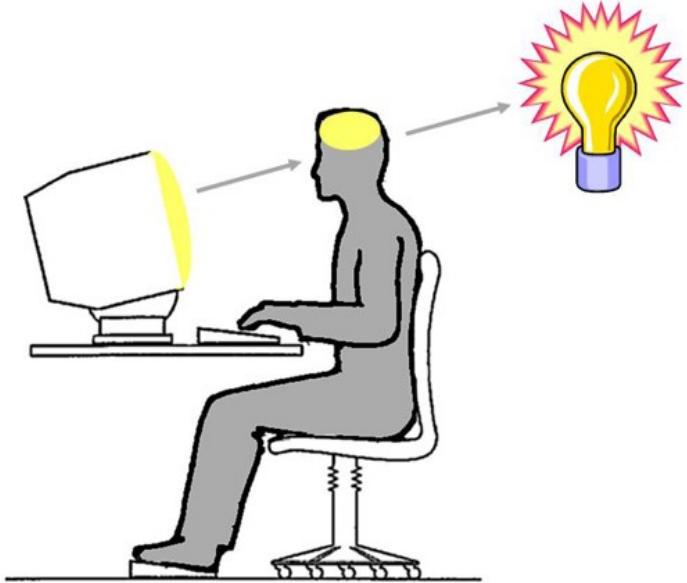
The link on this slide shows an example of a visual analytics web application designed using R packages and Shiny, a framework and API for building web application in R.

The advantage of this approach is that it allows users to perform the analysis without having to learn programming. Furthermore, data visualisation are used extensively in the modelling process. For visual Analytics applications developers, their only need learn one programming language which is R instead of a collection of web development languages such as html, javascripts and css.

What is Visual Analytics?

Visual Analytics is the craft of **analytical reasoning**
supported by:

- Data Science and Analytics
- Data Visualisation
- Human-Computing-Interaction (HCI)
- Interface Design
- Visual Storytelling

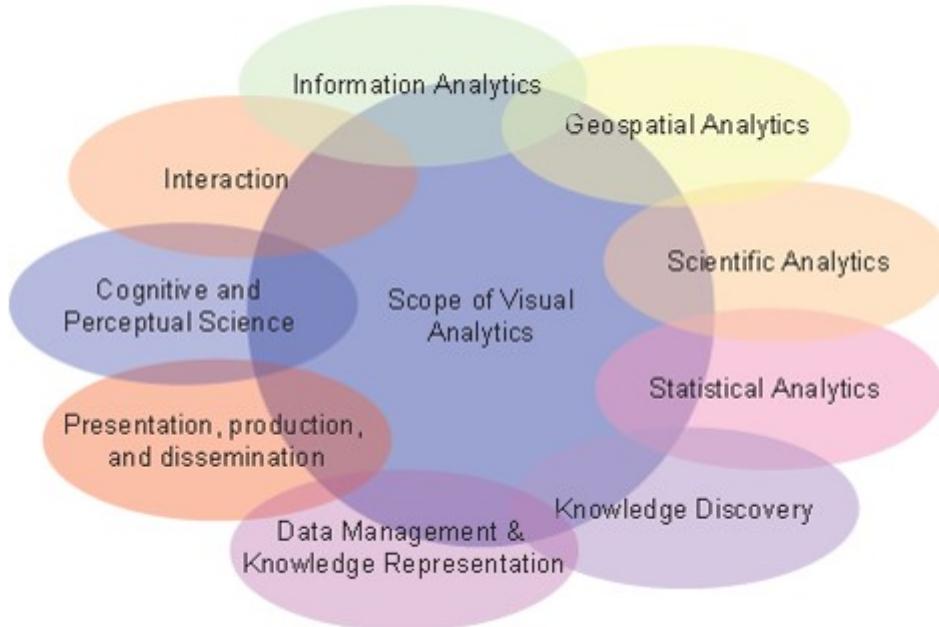


The marriage of data analytics and visualizations

This slide aims to explain the building block of Visual Analytics. It is the craft of

What is Visual Analytics?

Science of Visual Analytics



What is Visual Analytics?

Aims of Visual Analytics

Detecting the expected, discovering the unexpected

Although visual analytics comprises of many scientific disciplines. It is important to note that the goal of a visual analytics is to help data analyst to discover the unexpected from data effectively.

What is Visual Analytics?

Visual Data Sense-making

Visual Analytics' greatest strength is its use for **sense-making**.

*When you have eliminated
the impossible, whatever
remains, however
improbable, must be the
truth.*

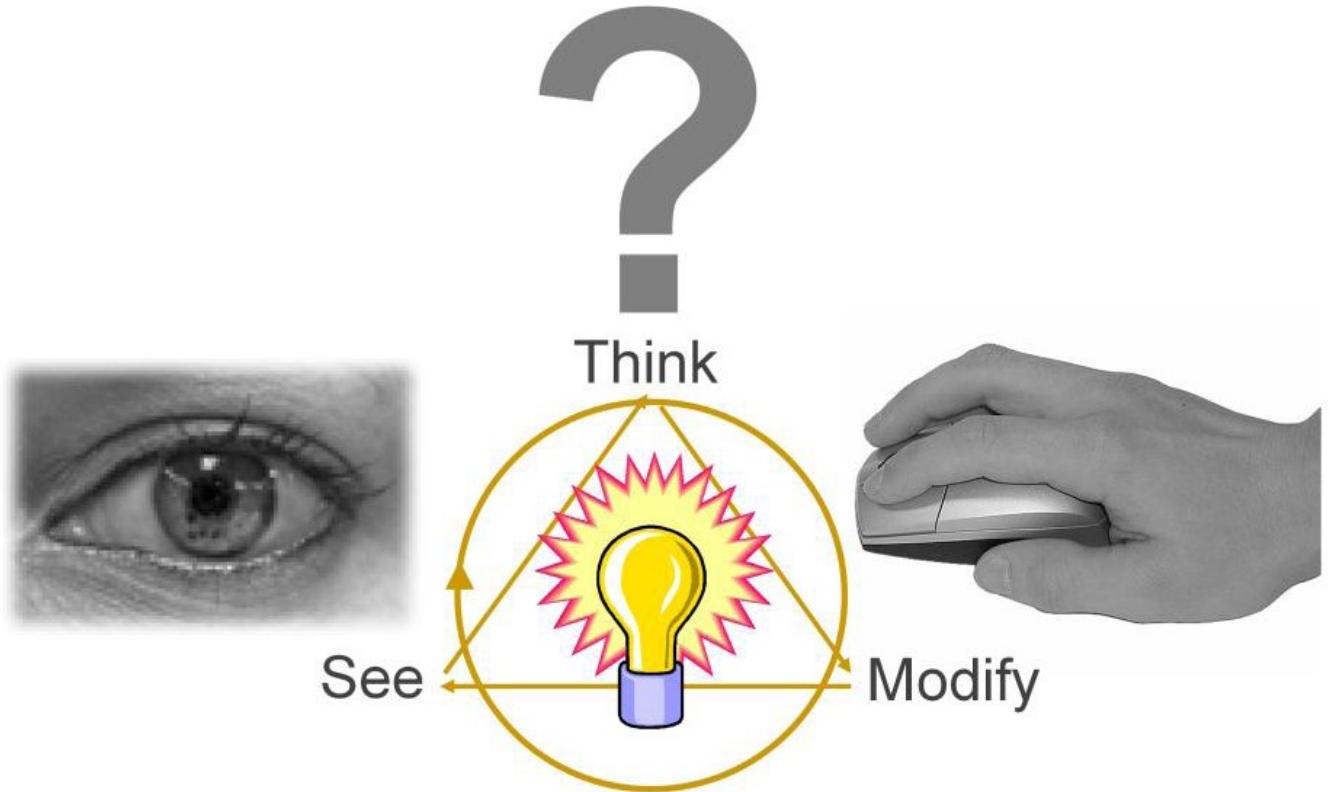
- Sherlock Holmes



Different from statistical analysis, machine learning and AI, Visual analytics emphasise on visual sensing rather than mathematics and statistical skills.

What is Visual Analytics?

Visual Analytics Framework



As a result, when we do Visual Analytics, we need to immerse ourselves in a highly interactive environment. Our eyes will be used to detect the unusual visual pattern with the help of appropriate data visualisation method(s). Our brain will then interpret and synthesise the visual information and send impulses to our hand to change the view in search of new visual patterns

What is Visual Analytics?

Visual Analytics require a shift of mindset

VISUAL THINKING

Exploration

Analysis

VISUAL COMMUNICATION

Synthesis

Presentation

PRIVATE REALM

PUBLIC REALM

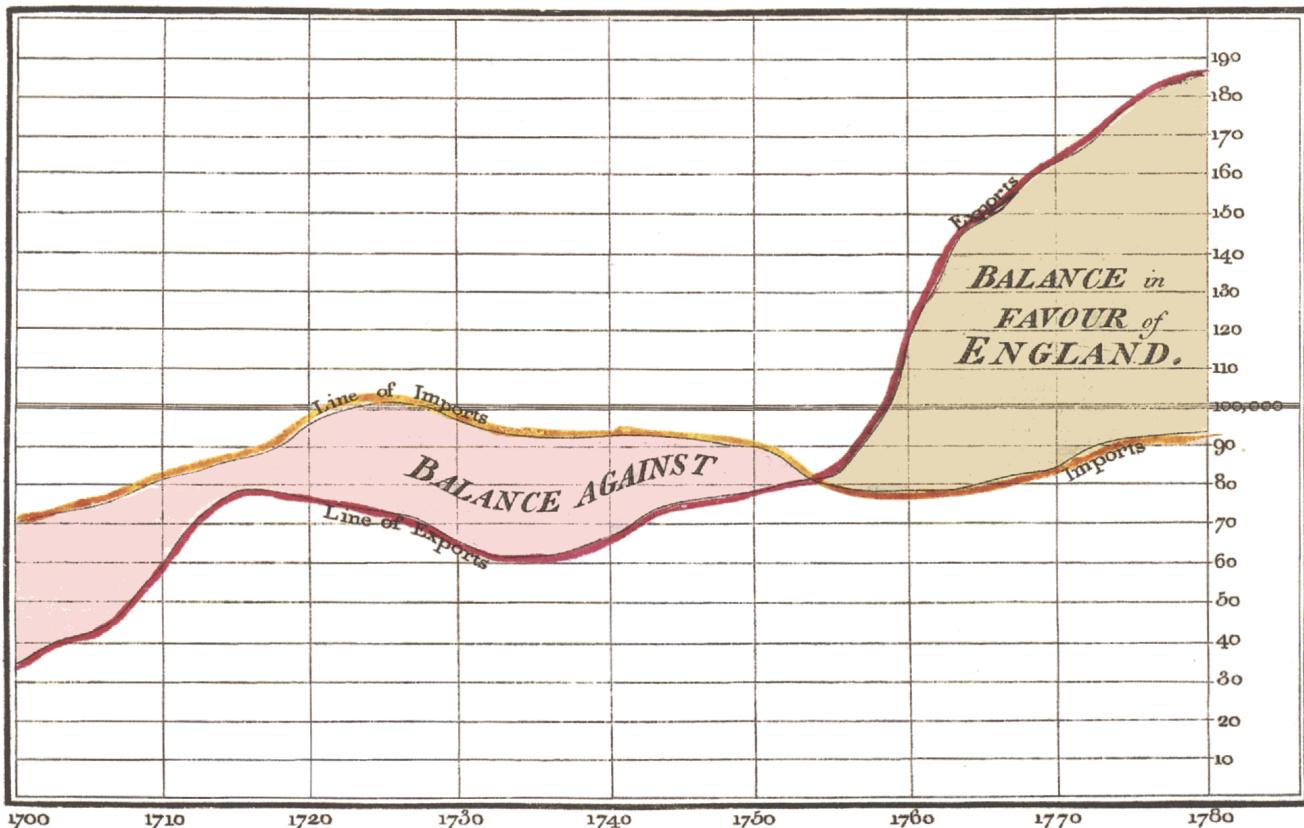
Traditionally, statistical graphics are used for visual communication alone. Visual Analytics require a major paradigm shift from visual communication to visual thinking. Data Visualisation or statistical graphics should be used to support data exploration and analysis instead of visual communication alone. Having said that, this does not mean that visual communication is no longer important. Instead, the new paradigm shift requires us to give equal focus on both visual thinking and visual communication.

What is Visual Analytics?

Learning from the history of Visual Analytics

[Commercial and Political Atlas by William Playfair](#) (22 September 1759 – 11 February 1823)

Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.



The Bottom line is divided into Years, the Right hand line into £10000 each.

Published as the Act directs, 1st May 1786, by W^m Playfair

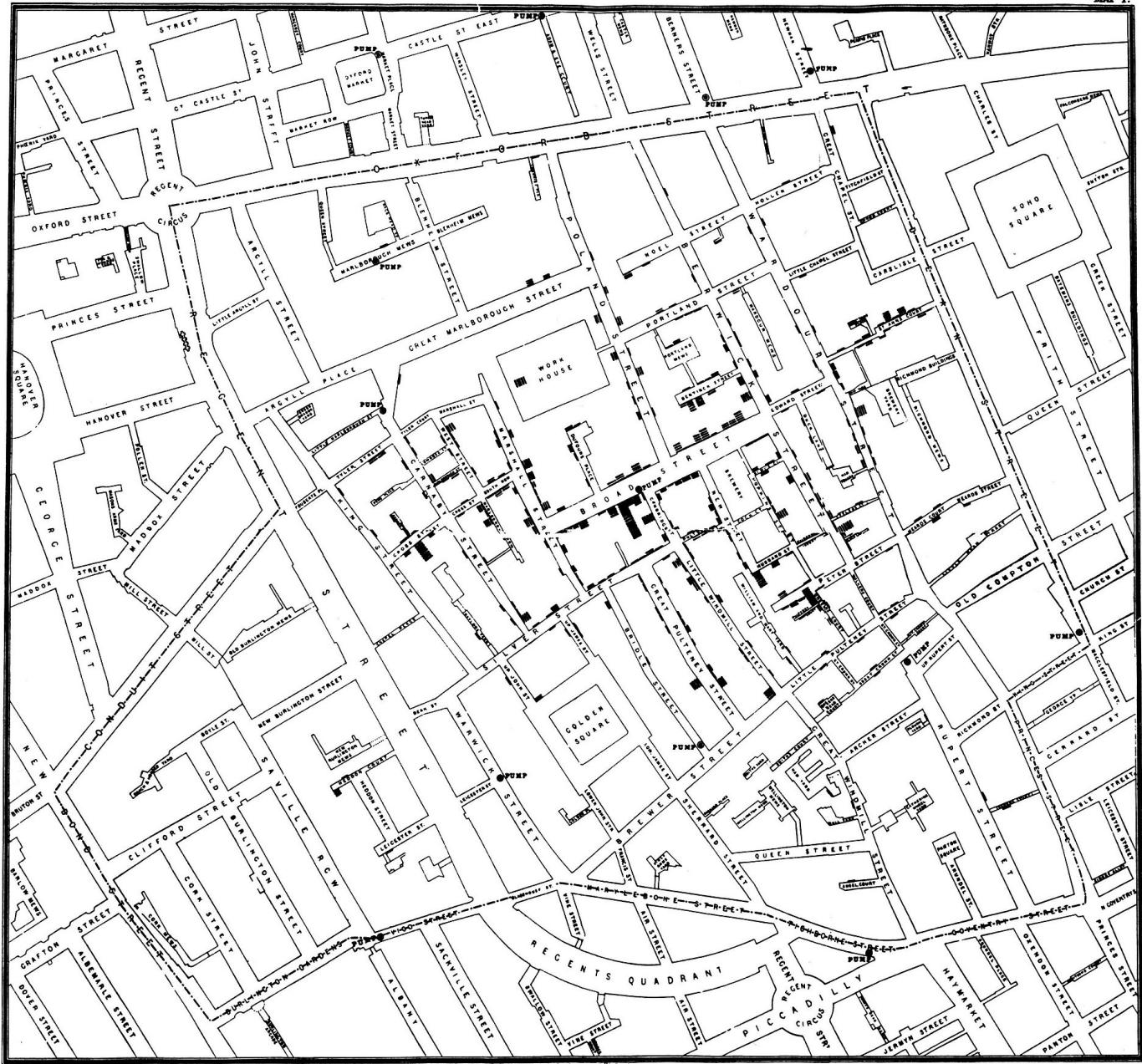
Neale sculpt. 352, Strand, London.

Speaking about Visual Analytics, there are a lot we can learn from history. The line graph on the slide is a good example. This is a trade balance line graph prepared by William Playfair in 1786. You probably have identify this chart as a common line graph. The horizontal axis is years, the vertical axis measures the volume of imports and export. When you take a good look at the graph, you will be very impressed by conscious effort given by the author to shade the space between the import and export lines. When the trade balance was against England, the area was shaded in red and when the trade balance was in favour of England, the area was shaded in brown. With the advancement of graph drawing tools, sadly to say, we hardly see this carefully crafted line graph been used.

What is Visual Analytics?

Learning from the history of Visual Analytics

[Cholera and the Pump on Broad Street](#)(John Snow, 1854).

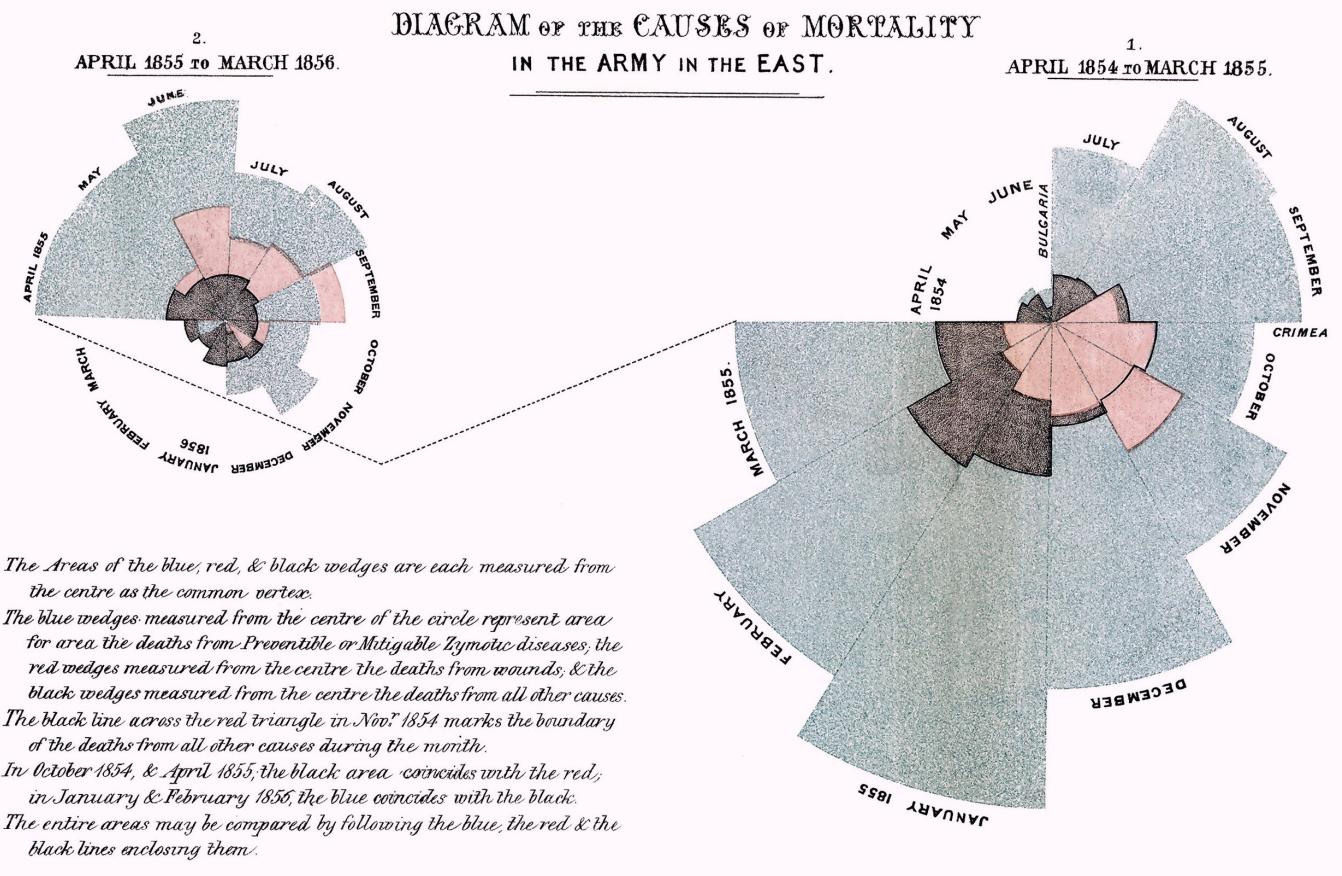


The map on this figure is the ground breaking used of map in medical science. It was one of the many versions prepared by John Snow and published in 1854. If you take a closer look at the map, you will notice that it is not a normal dot map we commonly prepared using Google Map or other data visualisation toolkits like Tableau or PowerBI. In the map, cholera cases are highlighted in black, showing the clusters of cholera cases (indicated by stacked rectangles) in the London epidemic of 1854. The map marks an important part of the development of epidemiology as a field, and of disease mapping as a whole.

What is Visual Analytics?

Learning from the history of Visual Analytics

[Florence Nightingale: The compassionate statistician](#)(12 May 1820 – 13 August 1910)

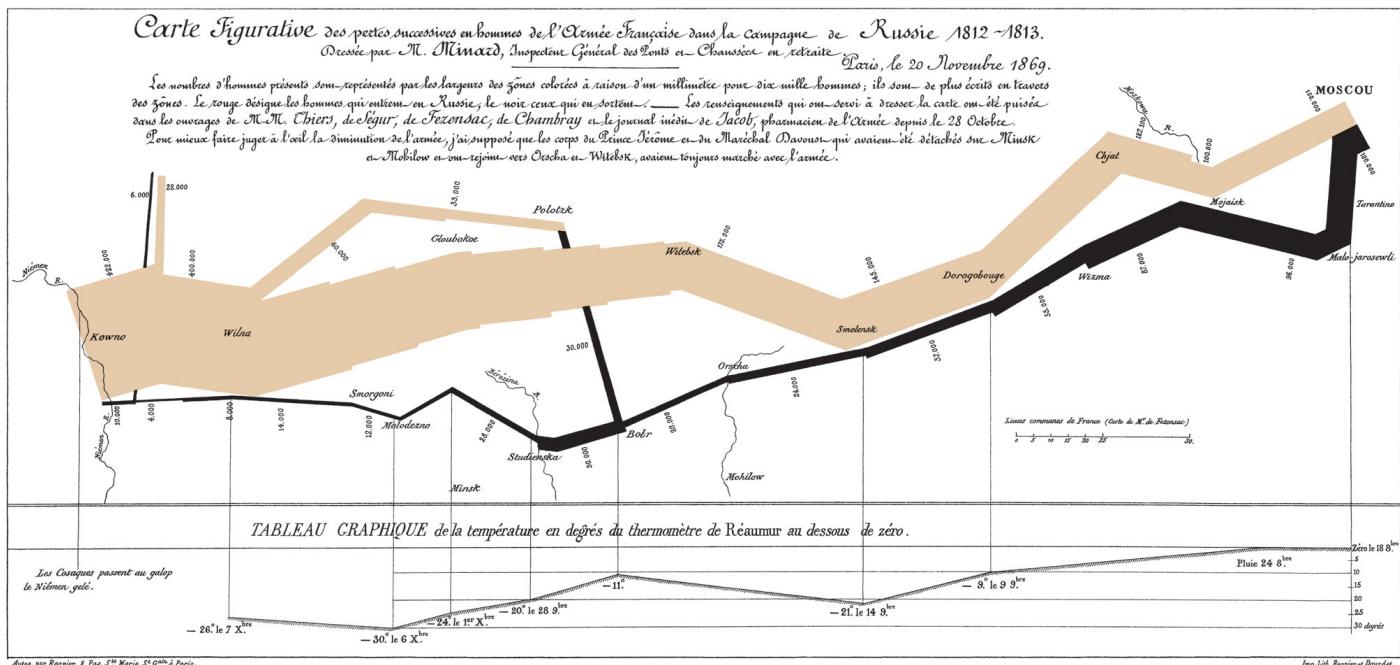


The third example is the famous rose diagram or coxcombs prepared by Florence Nightingale. Many of us know Florence as The Mother of Nursing, but she was also a statistician, and reformer. If we look at the rose diagram, we can see that the area of each coloured wedge, measured from the centre, is in proportion to the statistic it represents. The blue outer wedges represent the deaths from contagious diseases, such as cholera and typhus. The central red wedges show the deaths from wounds. The black wedges in between represent deaths from all other causes. If this rate had continued, and troops had not been replaced frequently, then disease alone would have killed the entire British Army in the Crimea. Nightingale's graph not only dramatised the extent of the needless deaths among the soldiers during the Crimean War, but it was used as a tool to persuade Victorian government and medical profession that deaths were preventable if sanitation reforms were implemented in military and civilian hospitals.

What is Visual Analytics?

Learning from the history of Visual Analytics

[Plot of Napoleon's ill-fated Russian campaign of 1812-1813](#) (Charles Joseph Minard, 1861)



Autog. par Rignier, 8. Par. 5^e Marie 5^e 6^e à Paris.

Imp. L. Rignier et Bourdet.

The last example is the famous Charles Joseph Minard's Napoleon's Russian campaign chart. At the first glance, it resemble the flow map but the map actually reveal more information than a common flow map. The illustration depicts Napoleon's army departing the Polish-Russian border. A thick band illustrates the size of his army at specific geographic points during their advance and retreat. It displays six types of data in two dimensions: the number of Napoleon's troops; the distance traveled; temperature; latitude and longitude; direction of travel; and location relative to specific dates without making mention of Napoleon; Minard's interest lay with the travails and sacrifices of the soldiers. This type of band graph for illustration of flows was later called a Sankey diagram, although Matthew Henry Phineas Riall Sankey (9 November 1853 – 3 October 1926) used this visualisation 30 years later and only for thematic energy flow. Interestingly, If you take a closer look at the lower center of the graph, you will notice that the term Tableau Graphique was used.

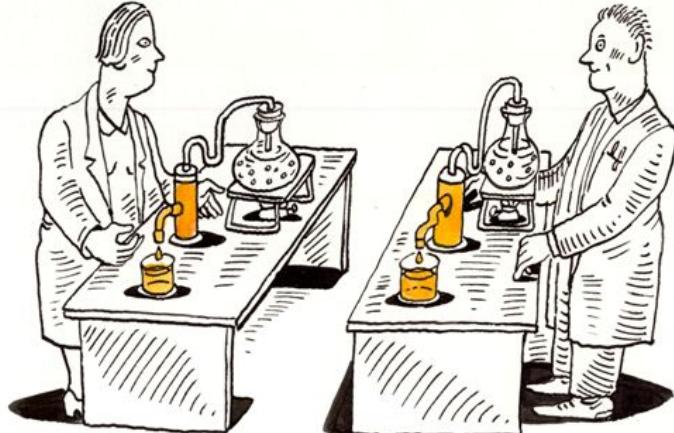
The important learning lesson from these four historical data visualisation is that in order to design a functional and yet enlightening data visualisation, we need to master the science and arts of data visualisation design instead of using default graphing functions of modern data visualisation toolkit. Just like writing, by knowing how to use MS Word, you will not be able to become a great editor.

What is Visual Analytics?

Visual Analytics must be Reproducible

- For every result, keep track of how it was produced.
- Avoid manual data manipulation steps.
- Archive the exact versions of all external programmes used.
- Version control all custom scripts.
- Record all intermediate results, when possible in standard formats.
- For analysis that include randomness, note underlying random seeds.
- Always store raw data behind plots.
- Generate hierarchical analysis output, allowing layers of increasing detail to be inspected.

- Connect textual statements to underlying results.
- Provide public access to scripts, runs, and results.



Reference: [Ten Simple Rules for Reproducible Computational Research](#)

One of the biggest challenges of visual analytics is reproducibility. In this course, you will learn the skill of building reproducible data visualisation by using RMarkdown in order to ensure that all the steps and codes used to build the data visualisation are comprehensively documented and can be reproduced by others.

Data Visualisation Toolkits

- Commercial-off-the-shelf
 - Visual analytics toolkit such as Tableau, JMP, PowerBI, Qlik Sense/View & TIBCO SpotFire.
 - Specialised data visualisation applications such as [TreeMap](#) from [Macrofocus](#)
- Open source
 - Specialised data visualisation applications such as [Gephi](#)
 - Data visualisation API library such as ggplot2, plotly, and [D3.js](#)

In general, we can divide the data visualisation toolkit into two main groups namely, Commercial off the shelf also known as COS and open source. A good example of COS data visualisation toolkits are Tableau, Qlik View/Sense.

COS Data Visualisation Toolkit

Magic Quadrant for BI and Analytics Platforms



COS Data Visualisation Toolkit

Data Visualisation Centric



BI Centric



Tableau: A data visualisation tool for everyone

- Theory to practice
 - The Grammar of Graphics
- Science to product
 - VisQL
 - Show Me
 - Polaris



Programmable Data Visualisation Library

- Javascript
 - [D3.js](#)
- Python
 - [Plotly](#)
 - [Bokeh](#)
 - [Matplotlib](#)

- R, Base Graphics
 - [lattice](#)
 - [ggplot](#)
 - [plotly.R](#)
 - [R Shiny](#)
-

Data Visualisation Toolkit for Data Journalism

- [Flourish](#)
- [Datawrapper](#)

Recently, there are two well developed web-based data visualisation designed specially designed for supporting data journalism. The advantage of these two toolkits are they do not need any installation and can be accessed any way as long as there is internet access.

Highly recommended blogs

- [Perceptual Edge](#)
- [Junk Charts](#)
- [The Functional Art](#)

This slide provides you the links to three highly recommended data visualisation blogs. I strongly encourage students to visit bookmark them and visit them often not only during these course but also when you go our to practice.