Data Visualisation

Associate Professor Goh Wooi Boon

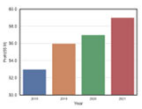
College of Computing and Data Science
Nanyang Technological University

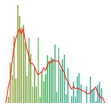
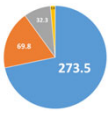
email: *aswbgoh@ntu.edu.sg*

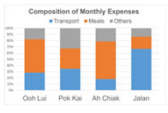
Chapter 5 – Basic Plots and Charts

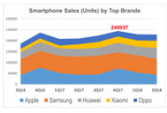
Contents

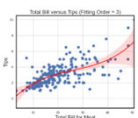
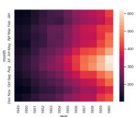
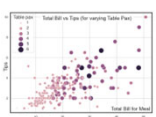
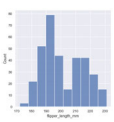
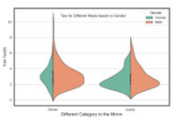
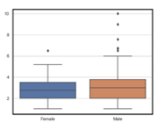
- Comparison Plots
- Composition Plots
- Relation Plots
- Distribution Plots





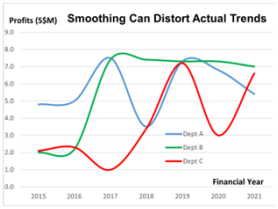


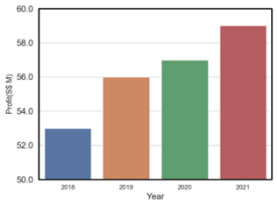







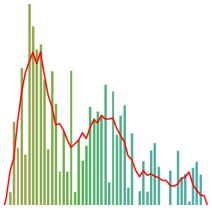
Chapter 5.1 – Visualising Comparison in Data

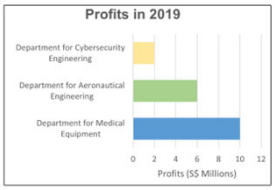
Contents

- Basic Comparison Plots
- Line Charts
- Bar Charts





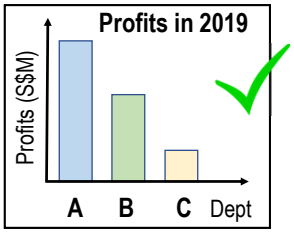


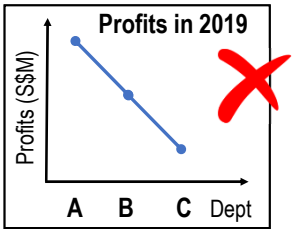


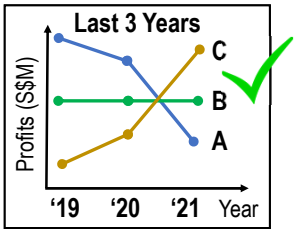
Basic Comparison Plots

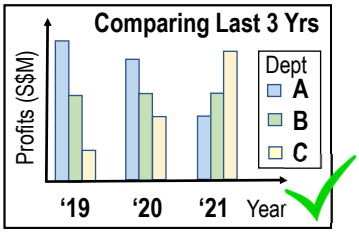
Draw the Line or Raise the Bar?

- Line charts should be use for **time series** (i.e. chronological) or when the x-axis has an **interval scale** (e.g. distances from a point, size of an object). While bar charts should be use when the x-axis have **nominal** (i.e. unordered) scale[1].
- Line charts should be used to show **trends** or pattern changes but **not precise quantitative values**. Bar charts could be used if specific measurements are to be **compared** at each **discrete intervals** along the x-axis.







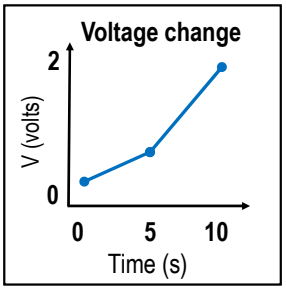


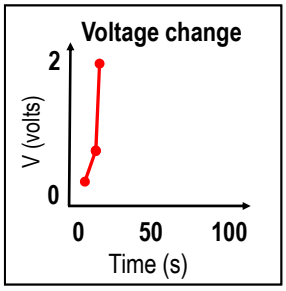
[1] S.M. Kosslyn, Graph Design for the Eye and Mind, Oxford University Press (2006)

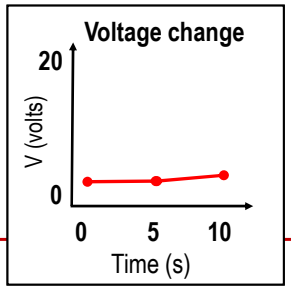
Line Charts

Axes Scaling and Aspect Ratio

- The aspect ratio of a graph is an important factor for judging **rate of change**.
- The "**banking to 45°**" rule[2] can be used to improve the visually encoding accuracy of changing trends.
- The axes scales or aspect ratio are adjusted such that the arithmetic average of all line orientations is nearly +45° or -45° (depending on the the direction of the trend).





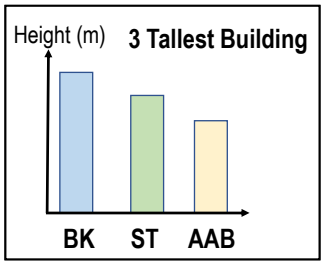


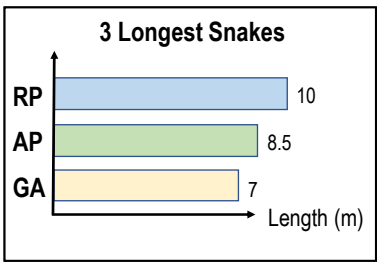
Taken from:
[2] W.S. Cleveland, The Elements of Graphing Data,, Hobart Press (1994)

Bar Charts

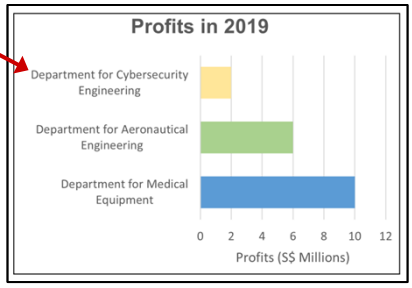
Vertical or Horizontal Bars

- Avoid using vertical text. If the x-axis categories have **long text** descriptions, consider using **horizontal bars**.





Long text descriptions



- The nature of the **conceptual data**[1] can also help decide which bar orientation better **scaffold the interpretation** of the data. (e.g. height - vertical, distance - horizontal) .

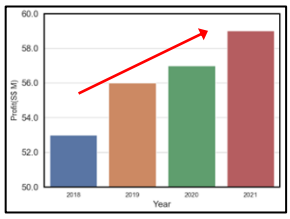
[1] S.M. Kosslyn, Graph Design for the Eye and Mind, Oxford University Press (2006)

Zero Baseline

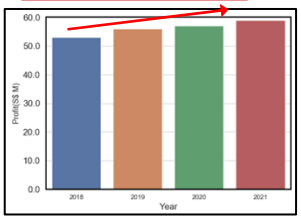
Zero Heroes

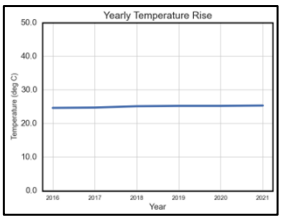
- Some argue that zero baseline must be used for **all plots**[3], whether bar or line charts to maintain visual "truthfulness".
- **Bar charts** should normally have a **zero baseline** as data is encoded using length and truncating the axis is naturally misleading.
- There is **some flexibility** with **line charts**, especially when the changing trends encoded in the **angle cannot be discern** if zero-baseline is imposed.

Visually exaggerated annual profit growth

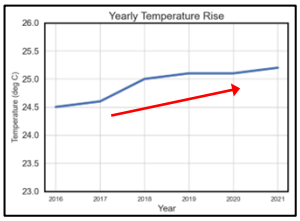


Actual profits plotted with zero base





Up trend not visible when zero base is imposed but becomes apparent when relaxed



7

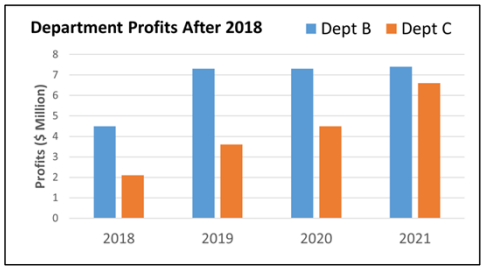
[3] C. Skelton, Bar charts should always start at zero. But what about line charts? (2018), http://www.chadskelton.com/2018/06/bar-charts-should-always-start-at-zero.html

7

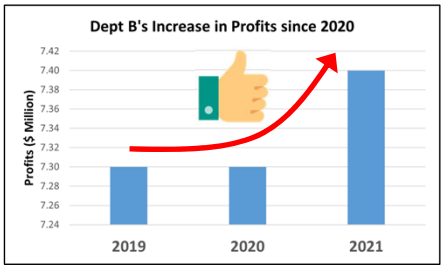
Ethical Visualisation

Don't Exaggerate

- Head of **Dept B** is concerned about the lack of growth in his product line compared to the huge growth in **Dept C**, which is likely to surpass Dept B soon. He has to make a yearly review presentation to the board and started working on his chart...



Changing profits of departments B and C in the last 4 years



The chart that the Head of Dept B presented the board of directors

Has ethical visualisation been violated here?

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

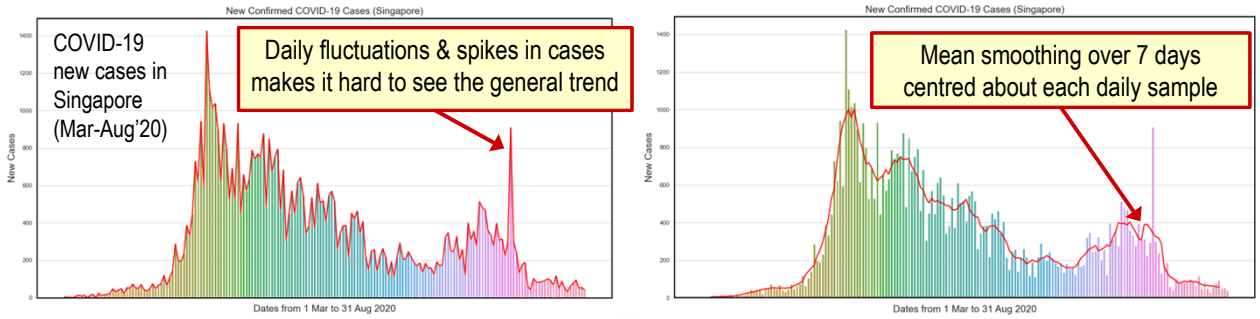
8

8

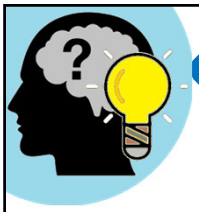
Visualising General Trends

Smoothing the Series

- Smoothing line or bar charts can reveal the more general or long-term trends in the data. Smoothing is particularly useful for noisy time series data.
- However, care must be taken to ensure that such smoothing does not affect the integrity of what the data represents[4].



[4] Jon Peltier, When you use a smoothed line chart, your data is not affected, it's misrepresented! https://www.vizwiz.com/2011/12/when-you-use-smoothed-line-chart-your.html



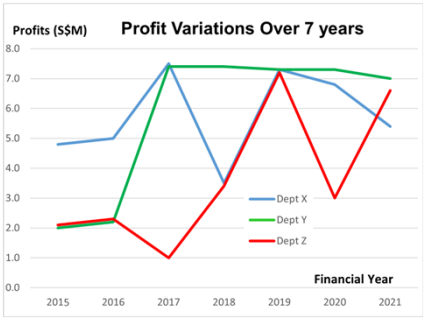
Think and Apply

When Smoothies are a Bad Idea

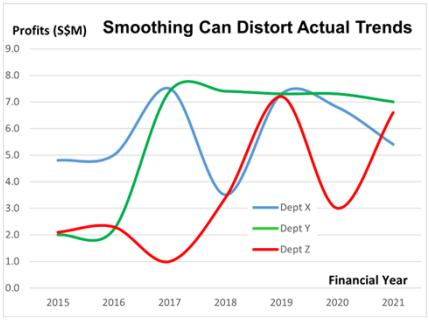
- Smoothing the data series distorts the plot and create misleading visual trends.

	Profits in last 7 years (S$ M)		
Year	Dept X	Dept Y	Dept Z
2015	4.8	2.0	2.1
2016	5.0	2.2	2.3
2017	7.5	7.4	1.0
2018	3.5	7.4	3.4
2019	7.3	7.3	7.2
2020	6.8	7.3	3.0
2021	5.4	7.0	6.6

Profits of three departments over 7 years



Line Chart showing exact values



Line plots smoothed using Excel "smoothed line" feature

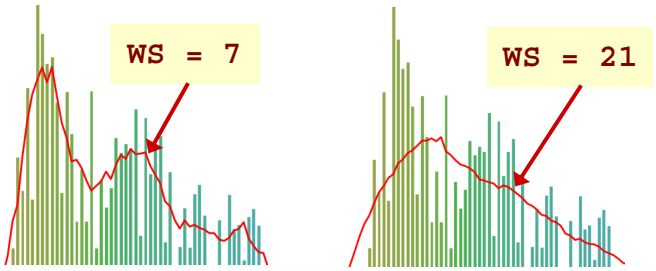
Data Smoothing

Smoothing Data Using Pandas

- Data smoothing[5] is done over a sliding window of a specified **width** . A **wider** window that covers more data points will result in **more smoothing**.

```
Smooth = Data.rolling(
        window = WS
        center = CW,
        win_type = WT).mean()
```

Window size



[5] Pandas, Dataframe.Rolling reference - https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.rolling.html

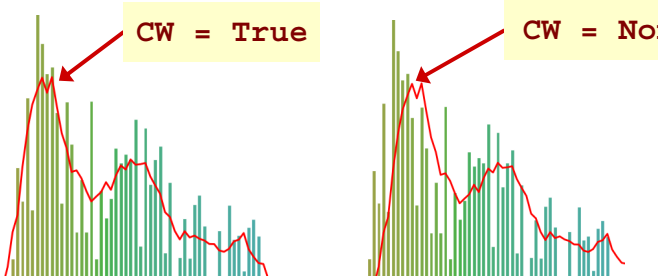
Data Smoothing

Smoothing Data Using Pandas

- Data smoothing[5] is done over a sliding window of a specified **width** . A **wider** window that covers more data points will result in **more smoothing**.
- The window can be **centred** or to the **edge** of the data point during smoothing and can take different "**shapes**"[6] to reduce the influences of data further away in the window when computing each **mean**.

```
Smooth = Data.rolling(
        window = WS
        center = CW,
        win_type = WT).mean()
```

Centering (default is right edge)



[5] Pandas, Dataframe.Rolling reference - https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.rolling.html

[6] Window types for win_type- https://docs.scipy.org/doc/scipy/reference/signal.windows.html#module-scipy.signal.windows

Data Smoothing

Smoothing Data Using Pandas

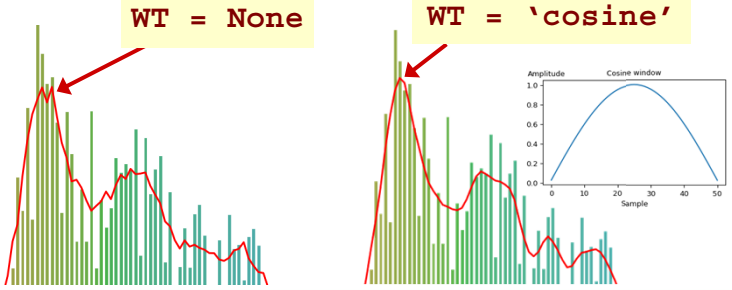
- Data smoothing[5] is done over a sliding window of a specified **width** . A **wider** window that covers more data points will result in **more smoothing**.
- The window can be **centred** or to the **edge** of the data point during smoothing and can take different "**shapes**"[6] to reduce the influences of data further away in the window when computing each **mean**.

```
Smooth = Data.rolling(
        window = WS
        center = CW,
        win_type = WT).mean()
```

Window shape
(default is equal weight)

WT = None

WT = 'cosine'



[5] Pandas, Dataframe.Rolling reference - https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.rolling.html

[6] Window types for win_type- https://docs.scipy.org/doc/scipy/reference/signal.windows.html#module-scipy.signal.windows

Summary

Comparison Plots

- **Line** and **bar** charts are two basic plots for comparing data values.
- **Bar charts** are useful for **comparing values** between variables, especially those having **nominal** scales of measure.
- **Line charts** are effective in observing **changing trends** in the data, especially time series data.
- **Data smoothing** is sometimes needed in order to observe **trend** changes that occur over a **longer time** duration or sampling span.
- However, be careful that smoothing does not **distort** the true interpretation of the data values.

References for Comparison Plots

[1] S.M. Kosslyn, Graph Design for the Eye and Mind, Oxford University Press (2006)

[2] W.S. Cleveland, The Elements of Graphing Data,, Hobart Press (1994)

[3] C. Skelton, Bar charts should always start at zero. But what about line charts? (2018), http://www.chadskelton.com/2018/06/bar-charts-should-always-start-at-zero.html

[4] Jon Peltier, When you use a smoothed line chart, your data is not affected, it's misrepresented! https://www.vizwiz.com/2011/12/when-you-use-smoothed-line-chart-your.html

[5] Pandas, Dataframe.Rolling reference - https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.rolling.html

[6] Window types for win_type- https://docs.scipy.org/doc/scipy/reference/signal.windows.html#module-scipy.signal.windows

Note: All online articles were accessed between May to June 2021