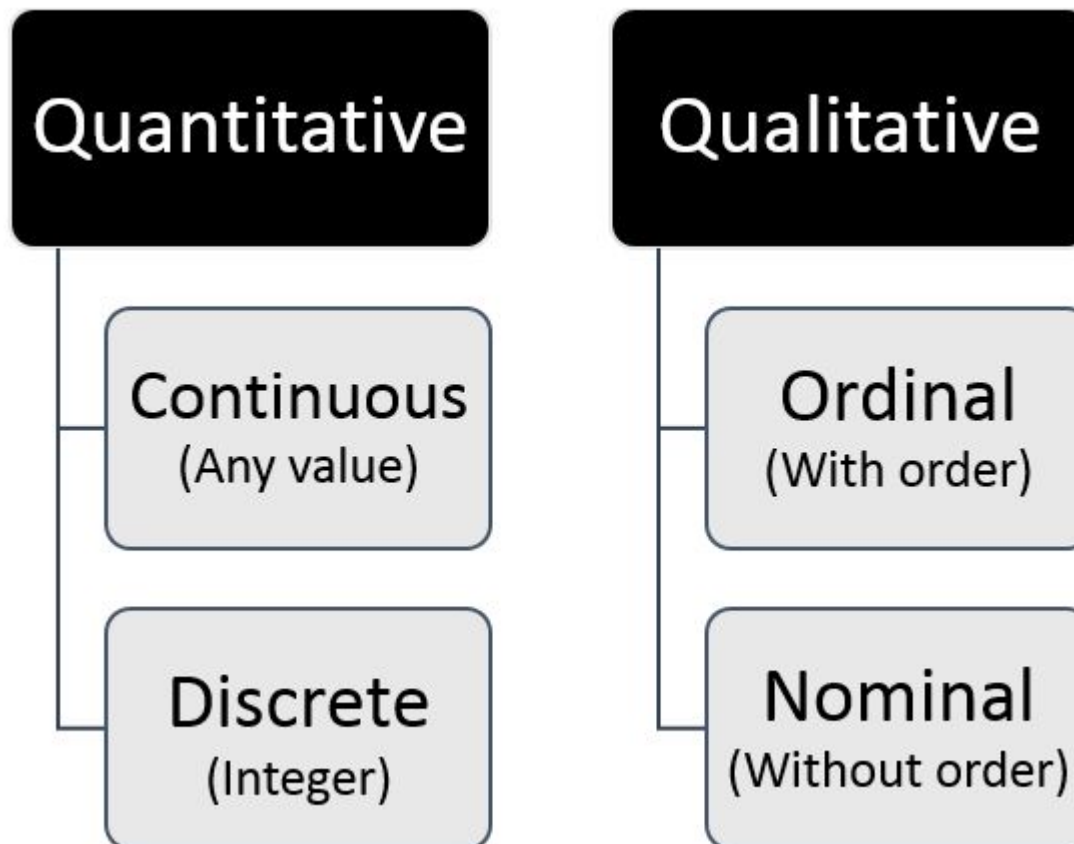


Graphical Representation of Data

Tan Yi Jane T06901105

Liew Shaer Jin T06901103

Data



Data Presentation

Text

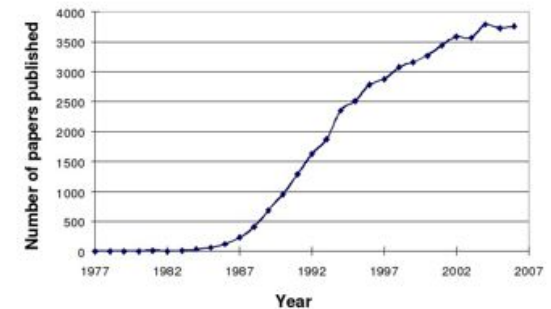
The Philippine Stock Exchange composite index lost 7.19 points to 2,099.12 after trading between 2,095.30 and 2,108.47. Volume was 1.29 billion shares worth 903.15 million pesos (16.7 million dollars). The broader all share index gained 5.21 points to 1,221.34. (From: Freeman dated March 17, 2005)

Table

| Disability status of the civilian noninstitutional population | | | |
|---|-------------|-------------|-------------|
| POPULATION 5 YEARS AND OVER | | | |
| | Both sexes | Male | Female |
| Total | 287,167,627 | 124,636,825 | 132,530,702 |
| With a disability | 49,748,248 | 24,439,531 | 25,308,717 |
| Percent with a disability | 19.3 | 19.6 | 19.1 |

| POPULATION 5 TO 15 YEARS | | | |
|---------------------------|------------|------------|------------|
| | Both sexes | Male | Female |
| Total | 45,133,667 | 23,125,324 | 22,008,343 |
| With a disability | 2,614,919 | 1,666,230 | 948,689 |
| Percent with a disability | 5.8 | 7.2 | 4.3 |
| Sensory | 442,894 | 242,706 | 200,188 |
| Physical | 495,461 | 251,852 | 203,609 |
| Mental | 2,078,502 | 1,387,393 | 691,109 |
| Self-care | 419,016 | 244,824 | 174,194 |

Graph



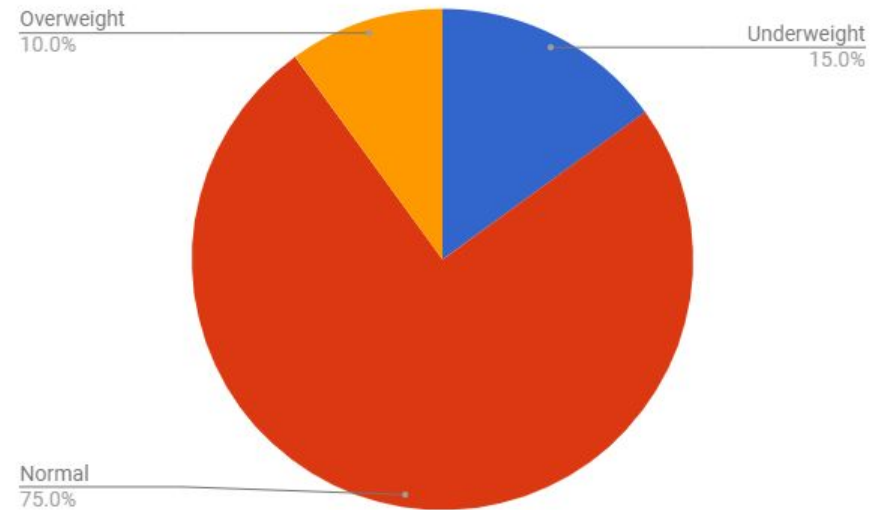
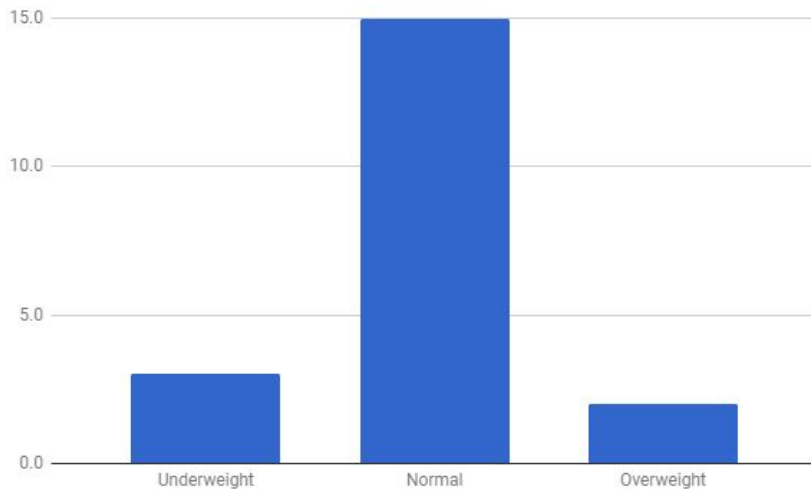
Why?

- In statistics, we often deal with large numerical data.
- A graph is the graphical representation of data using symbols, like lines and bars.
- Graphs can have high information density.
- Graphs help viewers to understand the data quickly.
- For example,

| | | | | | | | |
|-------|-------|-------|-------|-------|-------|-------|-------|
| 22.82 | 19.80 | 23.94 | 24.22 | 28.67 | 19.60 | 16.65 | 22.48 |
| 22.86 | 24.92 | 22.65 | 23.04 | 22.14 | 22.15 | 20.93 | |
| | 21.05 | 21.52 | 20.76 | 20.23 | 20.81 | | |

▲ BMI for 20 patients

Types of graph



In bar chart/line graph/pie chart

There are so many types of graph we can choose from, so which is the most suitable one?

Choosing the wrong graph can be confusing or misleading to your viewers.

Nuts and Bolts of Chart Types

Created by
online-behavior.com



HISTOGRAM

Chart used by responsible analysts who understand the power of segmentation and the sadness that comes from aggregating data.



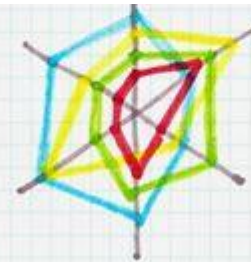
BAR CHART

Safe choice. But make sure you read Stephen Few before you show the chart to your boss, it will increase the probability of getting a raise.



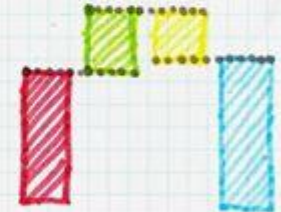
BUBBLE CHART

If you manage to extract insights from this graph your name is Hans Rosling.



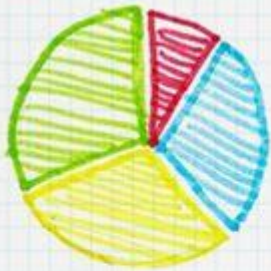
RADAR CHART

If you want to build a complex model around which you have done a ton of research, that's your choice (but only PhDs will understand you).



WATERFALL CHART

Perfect if you want to hide information or misguide other people. Seriously, can you trust a chart that is also known as a "Flying Bricks Chart" or "Mario Chart"? No.



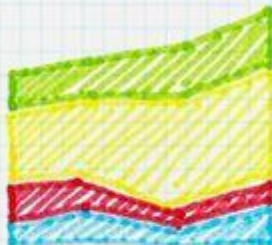
PIE CHART

Extremely useful when creating a well designed document that is intended to people that will not read the data (e.g. management)



LINE CHART

Useful to show trends, especially upwards (for downwards trends people tend to use more obscure charts, like the waterfall chart, see below)



AREA CHART

Please don't use this chart, I beg you! And please buy one of Tufte books.



TREE MAP

I have seen trees and I have seen maps, but how exactly this map is a combination of both? If you use it, good for you.



SCATTERPLOT

Very useful to find outliers, just like the people that commonly create them: human beings that finished their PhD in math by the age of 16.



BOX PLOT

This one is for pros. If you use it successfully, you will get a seat in heaven between Ronald Fisher and Johann Carl Friedrich Gauss.

How to choose the right graph?

First, understand your reasons of having the graph.

What are you showing?

| | |
|--------------|--|
| Comparison | Column, Bar, Circular Area, Line, Scatter plot |
| Composition | Pie, Stacked Bar/Column, Area |
| Distribution | Scatter plot, Line, Column, Bar |
| Trend | Line, Dual-Axis Line, Column |
| Relationship | Scatter Plot, Line |

How to make it more informative?

- A clear title!
 - Provide a description on what the data in the graph refers to.
- Plot more than one line in the graph to compare between different categories
 - Compare BMI for both genders across age group in the same graph
- Dual axis
 - Plot data with two y-axes and a shared x-axis, visualize the correlation between three data sets.
- Use consistent colours throughout the graph and use accent colour to highlight meaningful data point

Disadvantages

- Take up a lot of space
- Hard to read off exact numeric values
- Easy to be trapped by pitfalls
 - I.e Misrepresent data
 - A line may suggest interpolation between data points where none applies

Hence, it should be used moderately, supported with numerical representation of the data.

Common mistakes/ Pitfalls

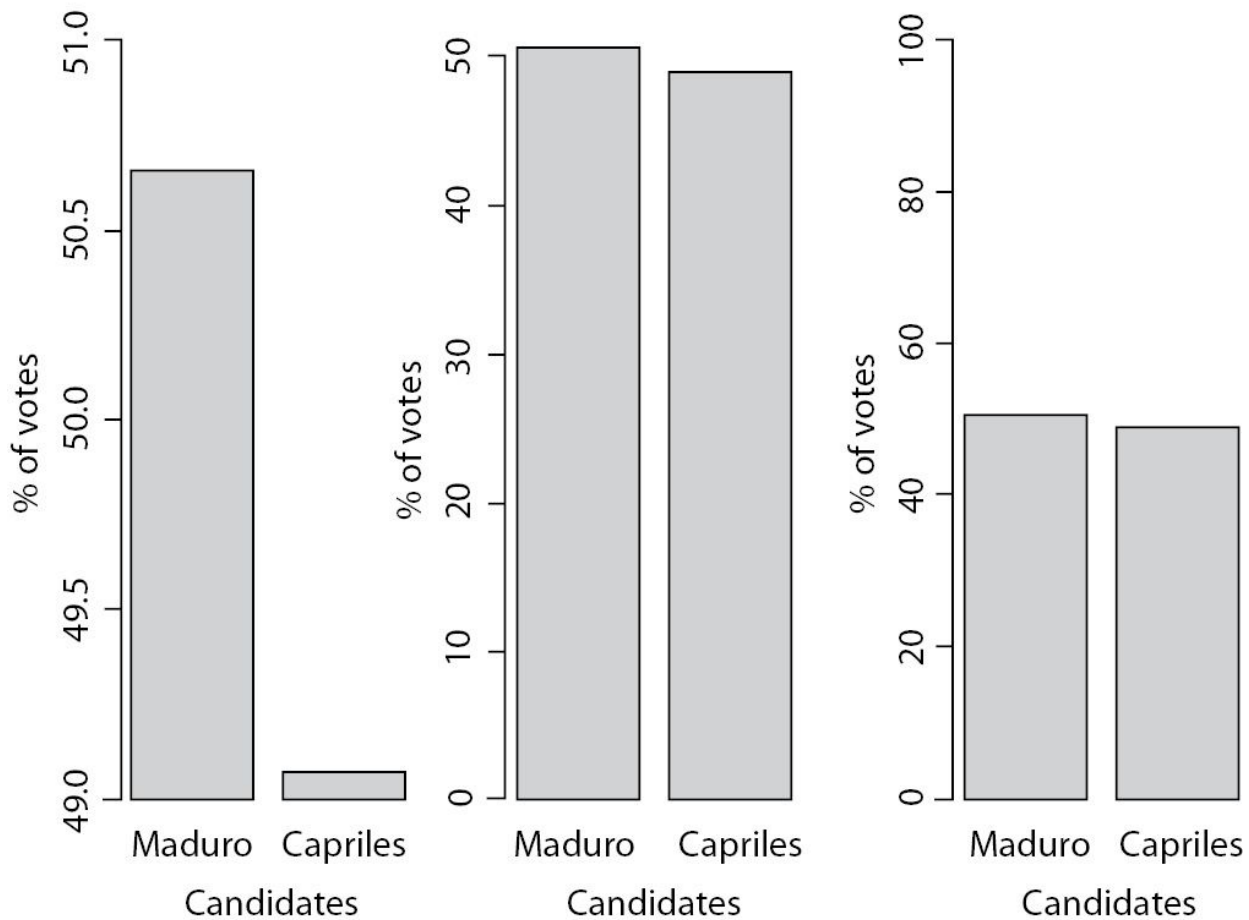
- 1) Wrong titles
- 2) Did not label the axis or label inaccurately
- 3) Too many lines in a certain graph (Can not read)
- 4) Too few lines (Can not compare)
- 5) Playing with axis and scale to enhance/reduce differences when there is none/some.
- 6) Irrelevant (Does not answer any question of interest)
- 7) Font size too small
- 8) Usage of colour but printing with black and white (can not differentiate lines)

Common mistakes

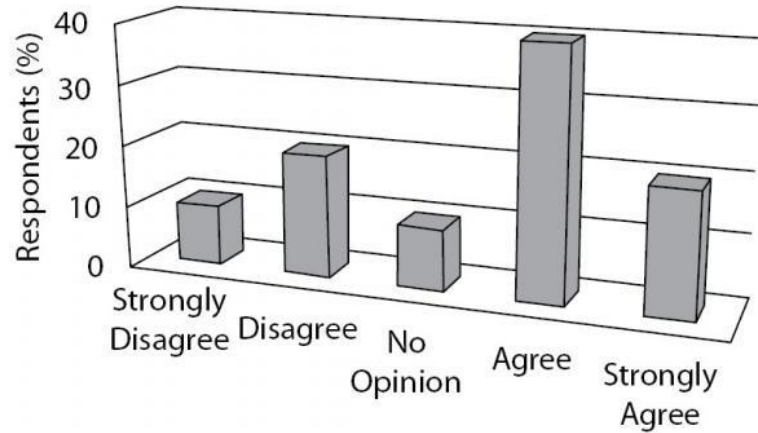
- **Technical** :
 - Does not follow standard method to draw graphs (E.g. 1,2,7,8)
- **Visual** :
 - The graph is technically ok but gives a strange impression (E.g. 3,5)
- **Objective** :
 - The graph is correct technically and visually but it serves no purpose.
 - For example, you want to compare the differences in the changes of heart rate for two different exercise. Best to plot the graphs together but someone plot it on different graphs. (E.g. 4,6)

Examples

Same Data, Different Scales

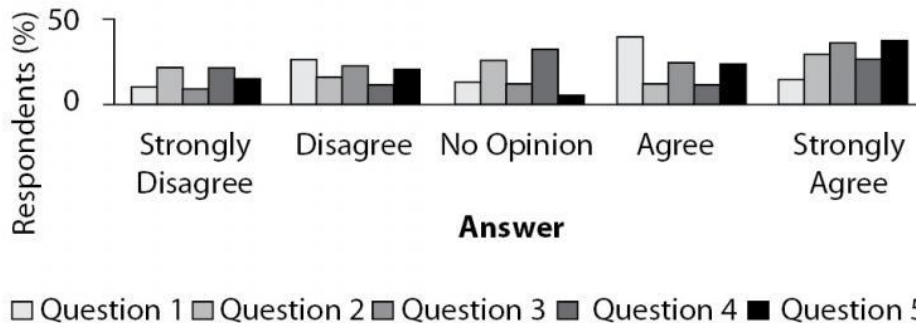


Brazilian team will win the opening match of FIFA World Cup against Croatia



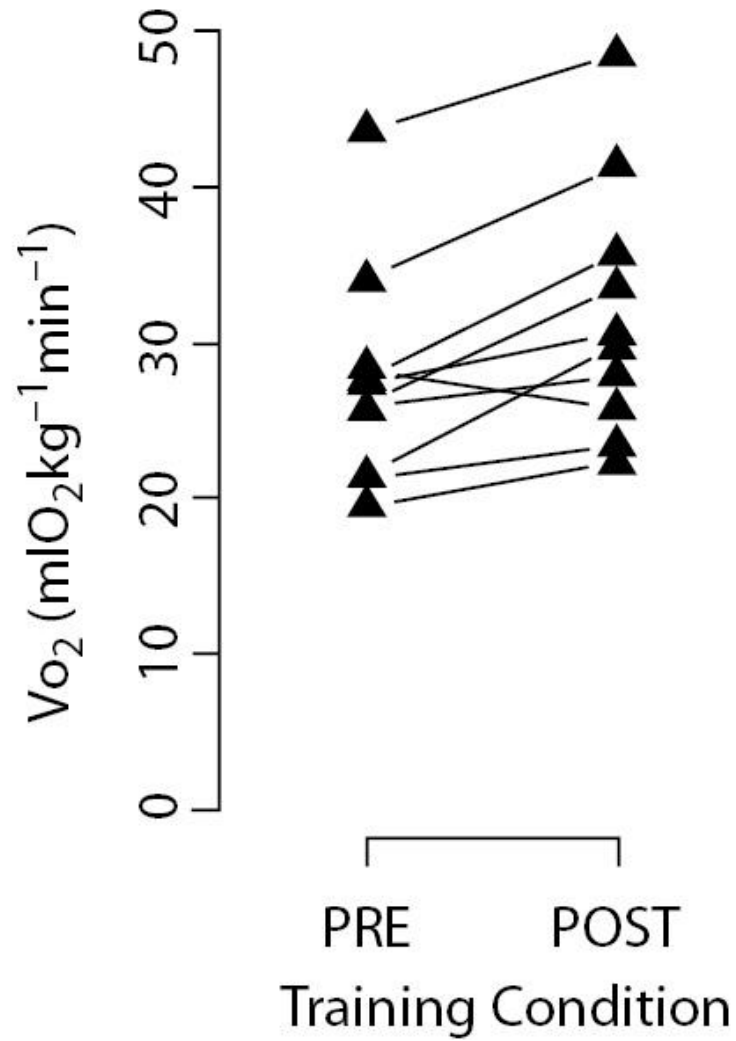
No need to entertain or distract the reader

Likert Scale to Five Questions



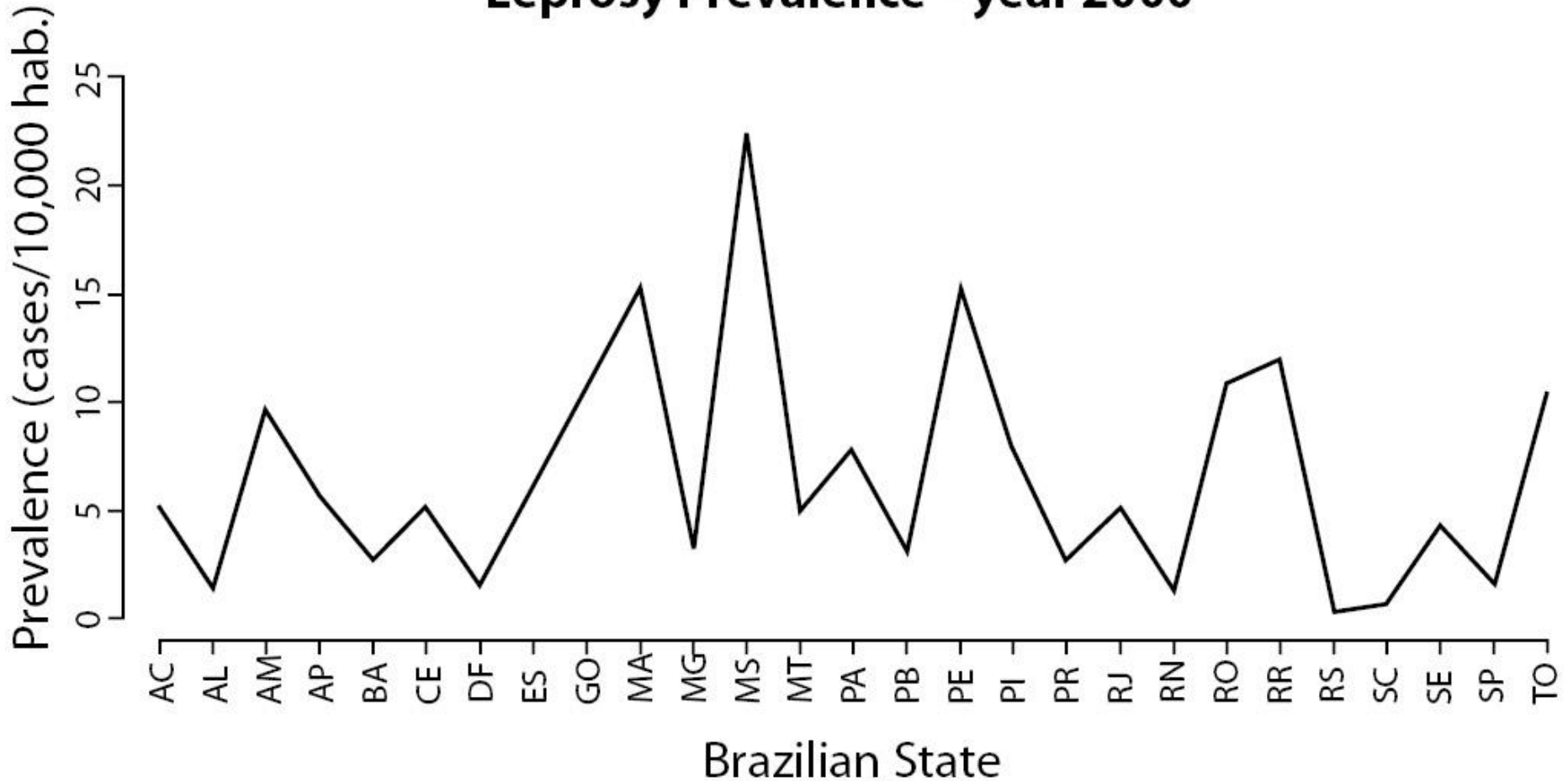
Too many informations in one graphic may be confusing

Line Plot With All Data

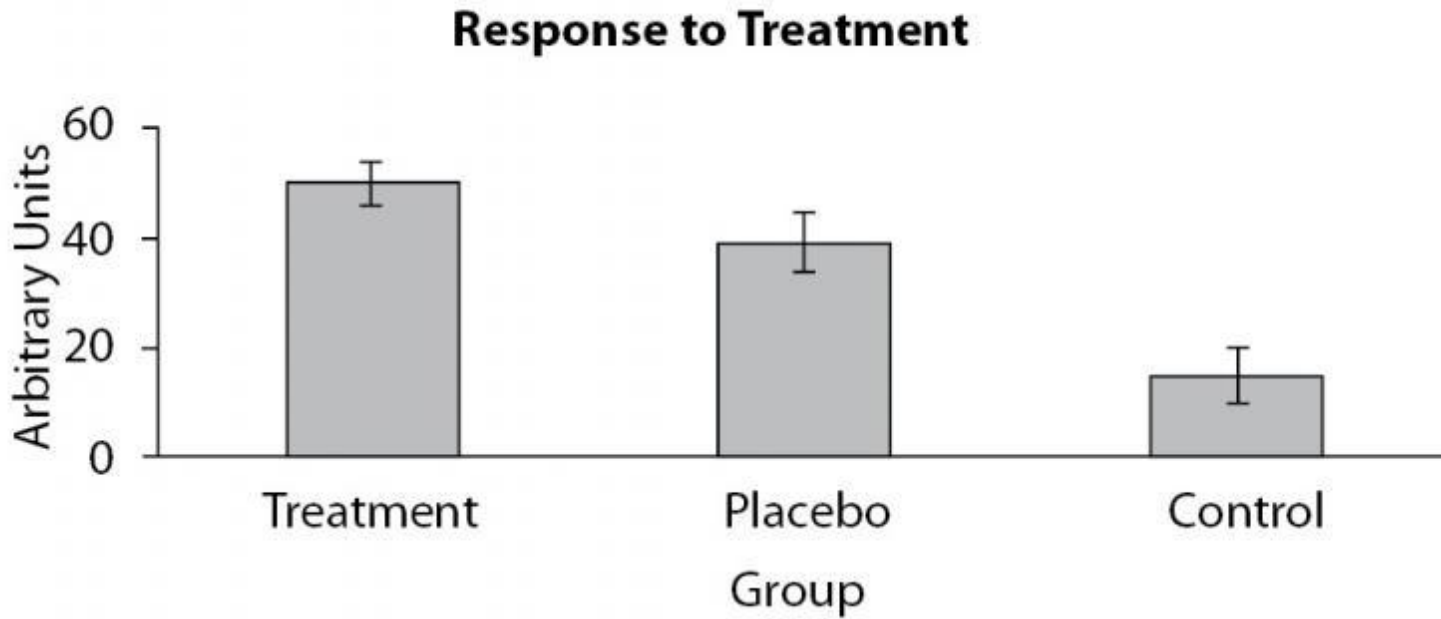


Data superposition would prevent visualization

Leprosy Prevalence – year 2000



For independent or categorical variables, lines might transmit wrong information, the continuity does not show



A good example of the use of bar plots with means and standard errors.

Tips

- **Technical**
 - **: Read more**
 - [Data Jujitsu: The Art of Turning Data into Product](#)
 - [Data Mining Algorithms In R](#)
 - [A Programmer's Guide to Data Mining](#)
 - [Mining of Massive Datasets](#)

- **Visual**
 - **: Books/Online materials show ways to avoid pitfalls**
 - https://www.slideshare.net/qlik_arg/5-data-visualization-pitfalls
 - <http://www.kristensosulski.com/2015/06/ten-common-pitfalls-for-presentations-and-data-visualizations/>
 - <http://www.biochemia-medica.com/2014/24/311>

- **Objectives**
 - **: Revise the objectives to decide graph to/not to draw**

Q&A

Acknowledgement

Informations were adopted from internet while some were provided by a lecturer of University of Malaya, Dr. Ng Siew Cheok.

References:

<http://www.biochemia-medica.com/2014/24/311>

<http://www.doc.govt.nz/Documents/science-and-technical/docts32entire.pdf>

<https://blog.hubspot.com/marketing/data-visualization-choosing-chart>

<http://www2.le.ac.uk/offices/ld/resources/numerical-data/numerical-data>