Statistics

Week 2 Recitation

ESD, SUTD

Term 5, 2017

Question 1 (Forecasting)

Download the *Excel* file on *triple jump* data from eDimension.

Make *1. Moving Average*, and *2. EWMA* forecasts.

- 'Predict' the winning distance for 2016
- Plot the actual data, moving average, and EWMA forecasts
- Find the MAPE for each one
- Use *Excel Solver* to find the α that minimize the MAPE of EWMA

Reminder:

- Exponentially weighted moving average:

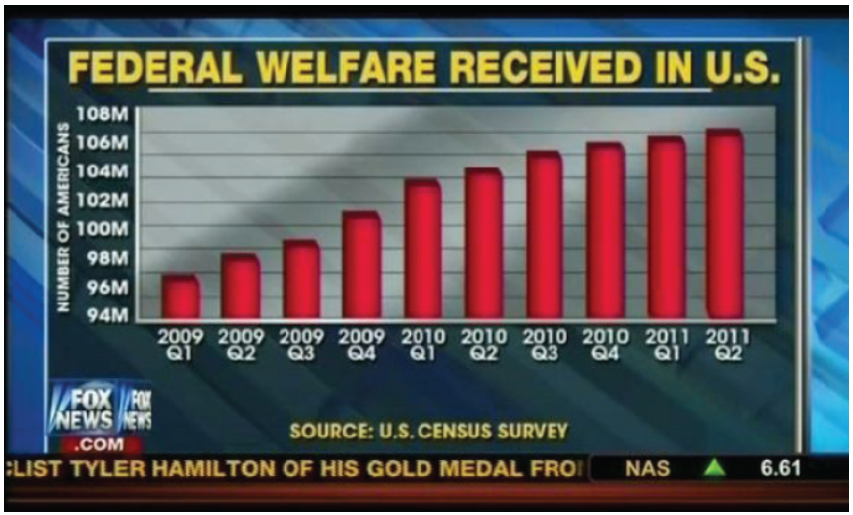
$$F_{t+1} = EWMA_t = \alpha x_t + (1-\alpha) EWMA_{t-1} \tag{1}$$

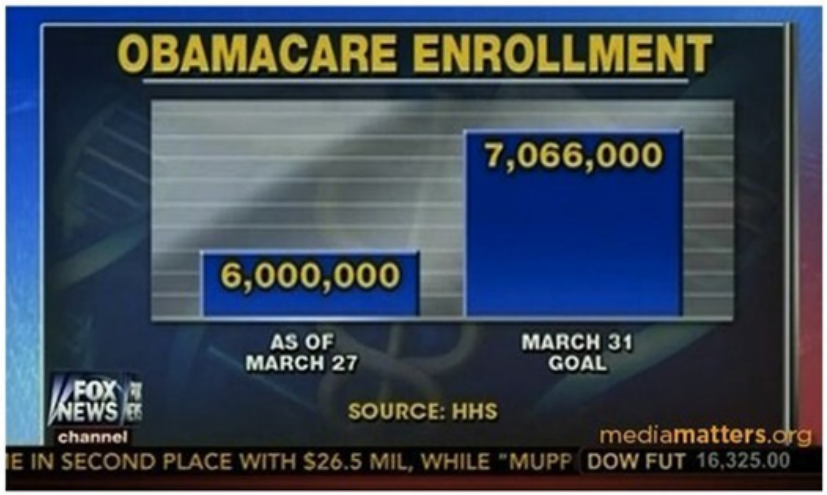
- Mean absolute percent error:

$$MAPE = \frac{1}{T-1} \sum_{t=2}^{T} \left| \frac{e_t}{x_t} \right| \times 100\% \tag{2}$$

Question 2 (misleading graphs)

I. What are some of the problems with the following graphs?

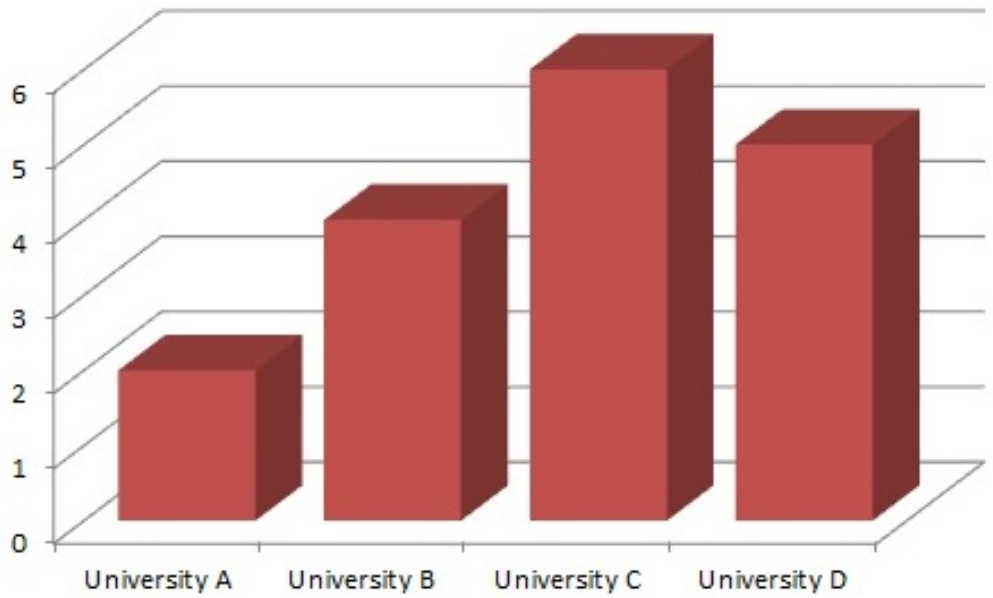




How can we fix them?

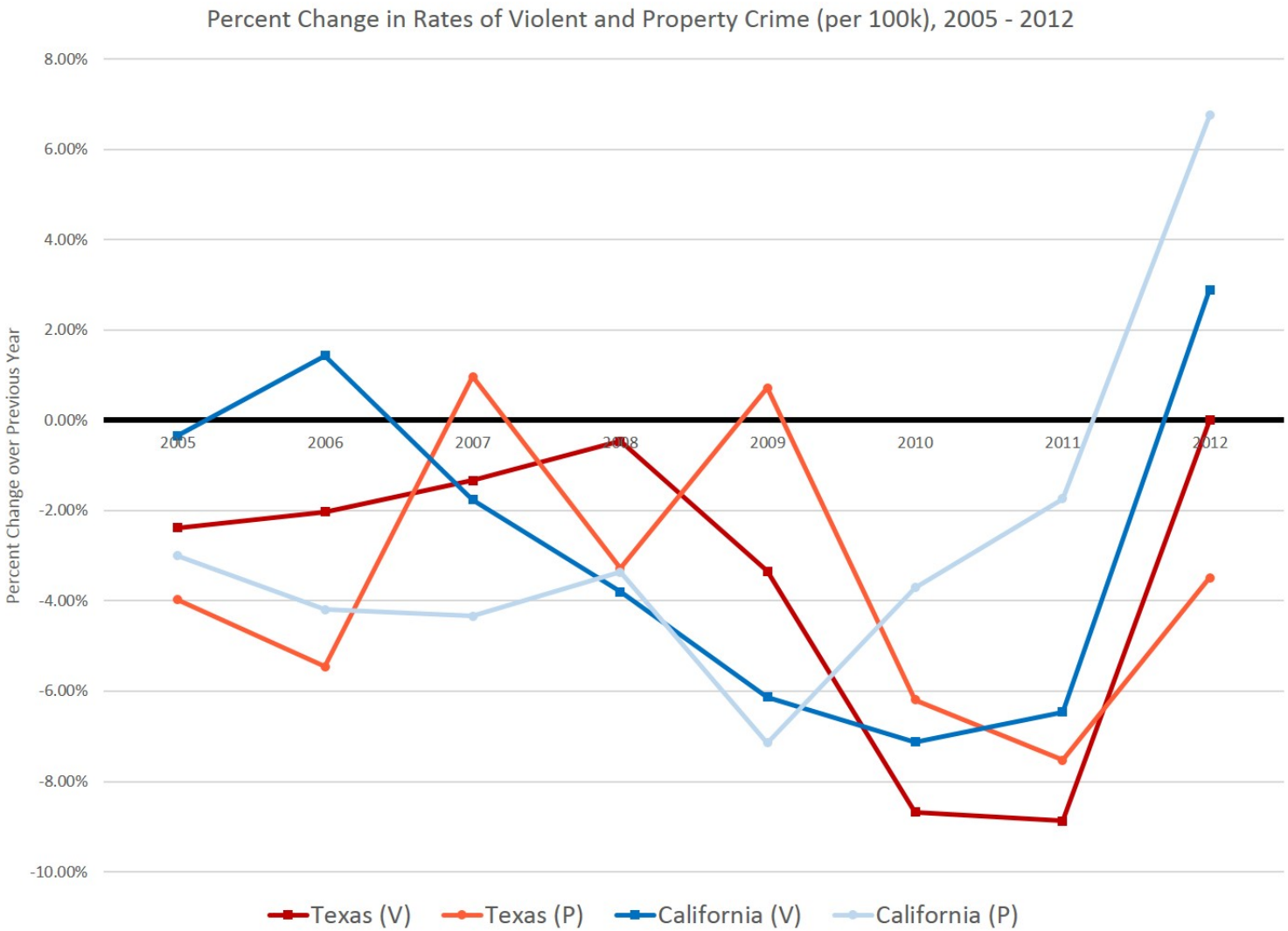
Question 2 (misleading graphs)

II. How can this graph be misleading?



Question 2 (misleading graphs)

III. Why is this graph not very informative?



Question 2 solutions

I. The y-axis does not start at 0; in fact the scale of the axis is chosen such that small differences are magnified. Also, for the second graph, the y-axis is not even labeled.

A correctly drawn graph should either start the y-axis at 0, or clearly indicate (for instance, using a zigzag line) that it does not start at 0.

II. 3D graphs should generally be avoided. In this case, making the bars 3D has visually the same effect as adding a fixed height on top of each bar.

So for instance, even though the value for University C (6) is 3 times the value for University A (2), this effect can make the ratio seem less than 3.

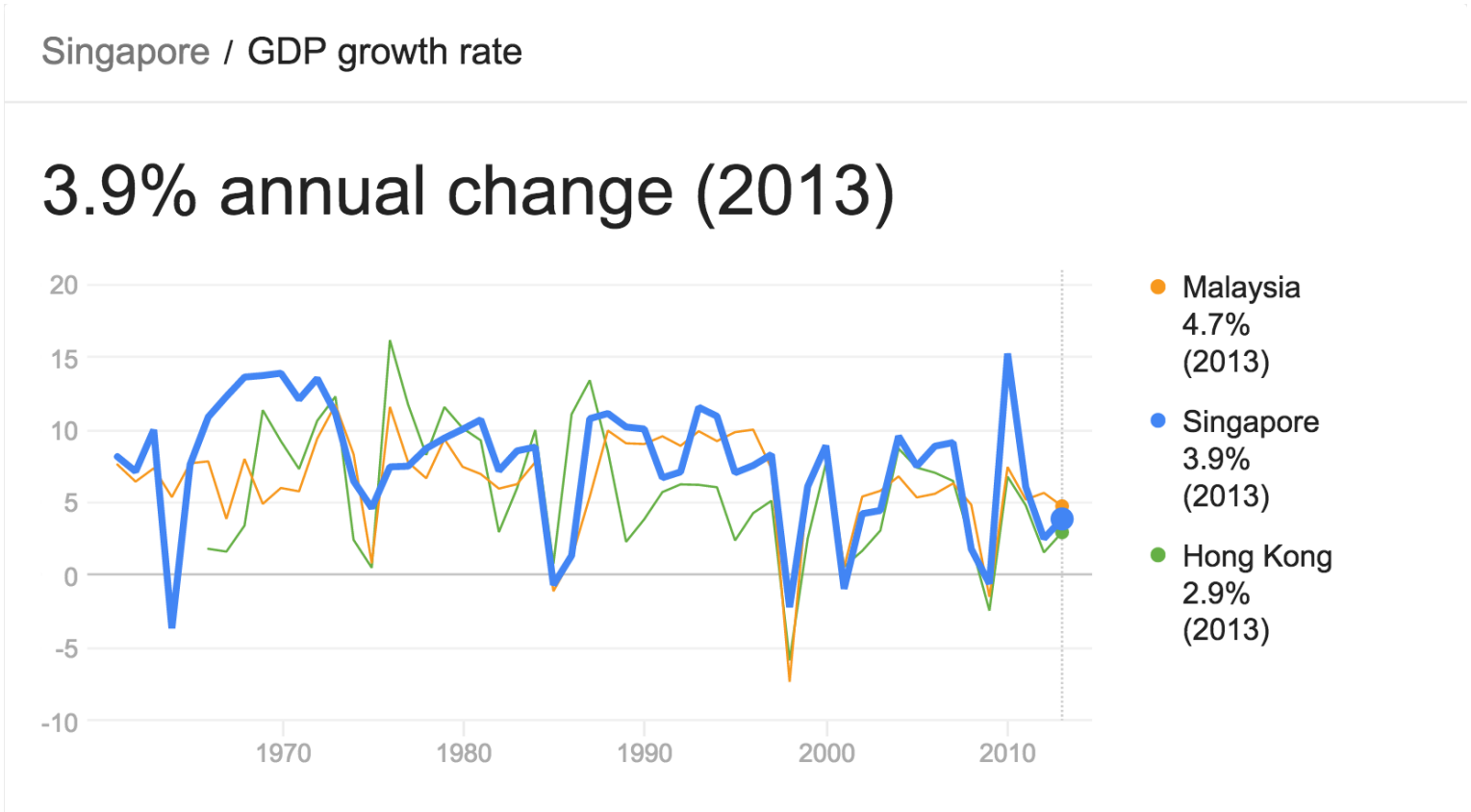
Question 2 solutions

III. Usually, graphs showing percentage change from the previous year are not very informative, since a generally upward trend may still show 0 or negative change from year to year, and vice versa. It is therefore very difficult to spot any trend from the graph.

In this case, it is much better to show the raw crime rates for each year.

Question 2 (Good graph examples)

An exception of graph showing percentage change



Question 2 (Good graph examples)

Rules of thumb for good data visualization

- Pick an appropriate graph type
- Provide explanation as needed - graph title, axis label, legend
- Focus on the data, avoid distracting details
- Think about the purpose of visualizing data - to find patterns, to present your findings ...

RStudio and R

Install R
https://cran.rstudio.com/

Install RStudio
https://www.rstudio.com/products/rstudio/download/

Install the softwares before next recitation (Feb.9th).
Step-by-step tutorial available on eDimension.