EE3001 Foundations of Data Engineering

http://www.cityu.edu.hk/ug/current/course/EE3001.htm

Lecturer: Dr. S.Y. Yuen, Kelvin

WWW: http://www.ee.cityu.edu.hk/~syyuen

Course Aim

- This is a new course to be offered in Sem. B 2016/17

- The aim of this course is to provide students with an understanding of the basic techniques used in data engineering and give introductory examples to how these techniques are applied to more advanced data engineering applications.

Data helps us to identify problems
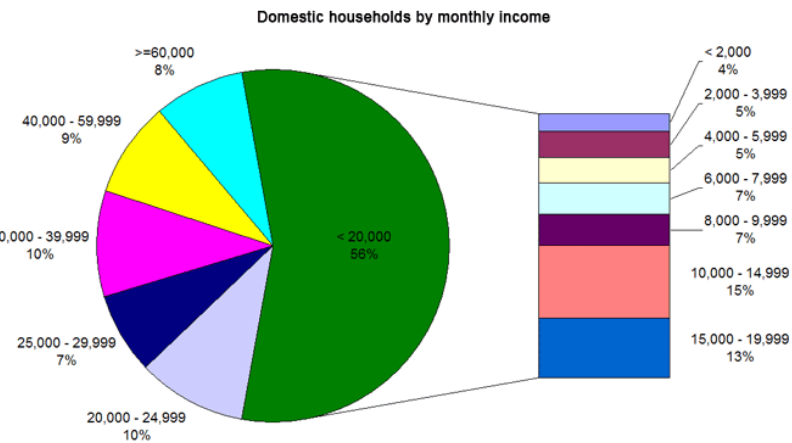
Data helps us to identify correlations and trends

Data helps us to make predictions
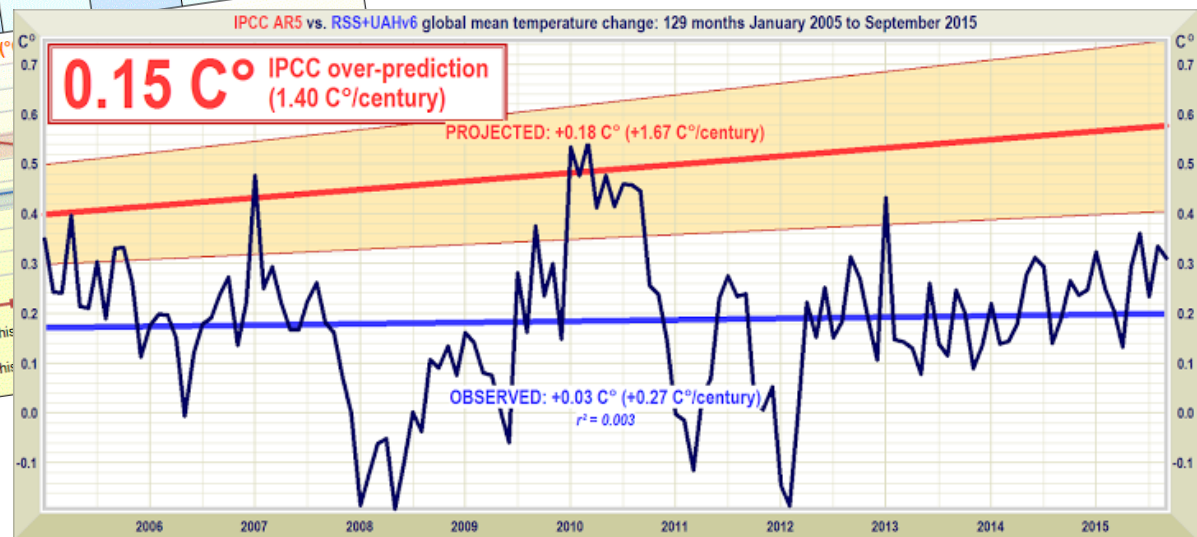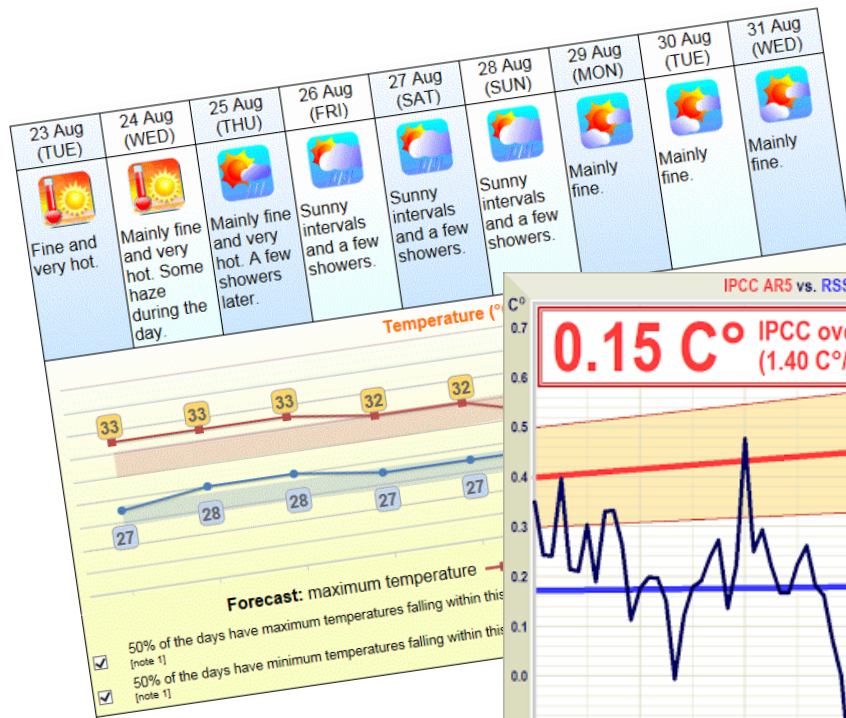
Beware of incorrect use/pitfalls/limitations in using data

Birthdays are good for your health.

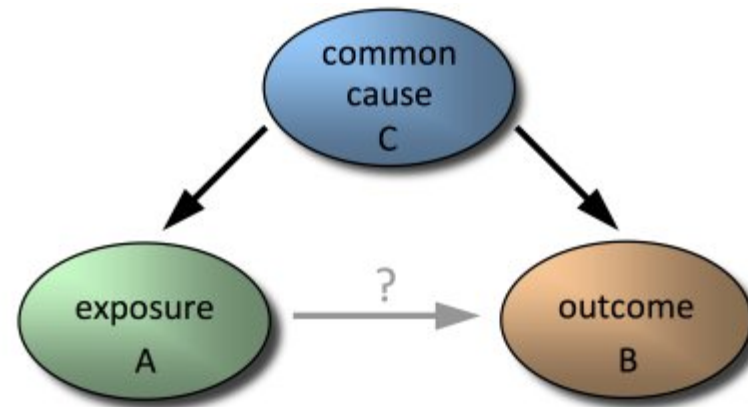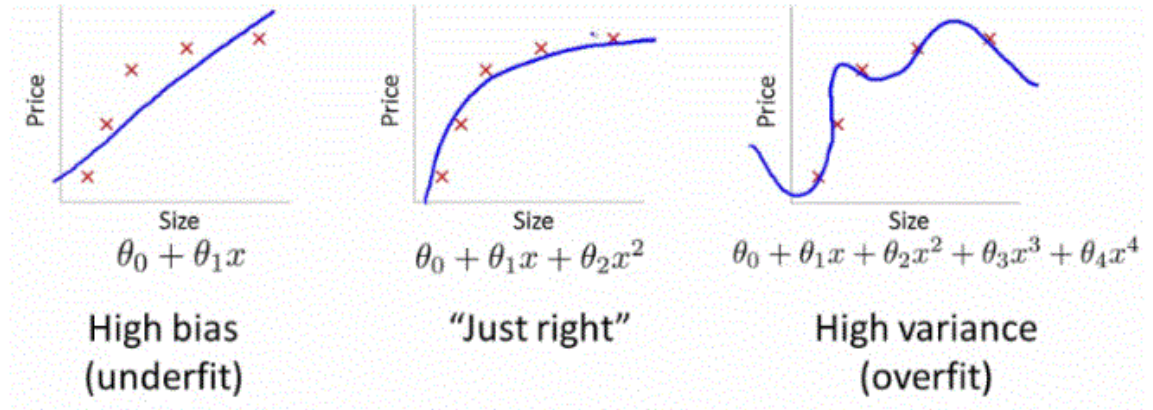**Studies have shown that people who have the most birthdays live the longest.**

High bias (underfit) — $\theta_0 + \theta_1 x$

"Just right" — $\theta_0 + \theta_1 x + \theta_2 x^2$

High variance (overfit) — $\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$

common cause C

exposure A

outcome B

?
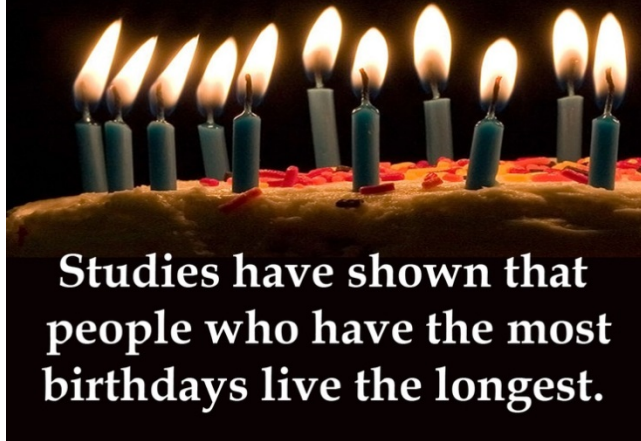
(taken from
All models are wrong, 7 sources of model risk (accessed 24 Aug. 2016)

Course Topics (Provisional)

- Fundamentals of Statistics
- Descriptive Statistics
- Probability
- Probability Distributions
- Sampling Distributions and Confidence Intervals
- Hypothesis Testing
- Regression
- Predictive and Descriptive Analytics
- How to correctly use data analysis tools and their limitations

Assessment Methods

- Continuous Assessment 40%
  - Quizzes. There may also be assignments and small projects

- Examination 60%
  - 2 hour examination

- Students need to obtain at least 35 (out of 100) in both the Continuous Assessment and Examination to pass the course

References

- D.M. Levine and D.F. Stephan, *Even you can learn statistics and analytics*, 3rd Ed., Pearson, 2015.

- Walpole, Myers, Myers, and Ye, Probability and statistics for Engineers and Scientists, 9th Ed., Pearson, 2014.

- D. Wackerly, W. Mendenhall, and R.L. Scheaffer, *Mathematical statistics with applications*, 7th Ed., 2007.

- M. Mitzenmacher and E. Upfal, *Probability and computing*, Cambridge University Press, 2005.

- All images in this presentation, except explicitly stated, are publicly available in Google Images