

How to Lie with Statistics



Ali Akbar Septiandri

Universitas Al Azhar Indonesia

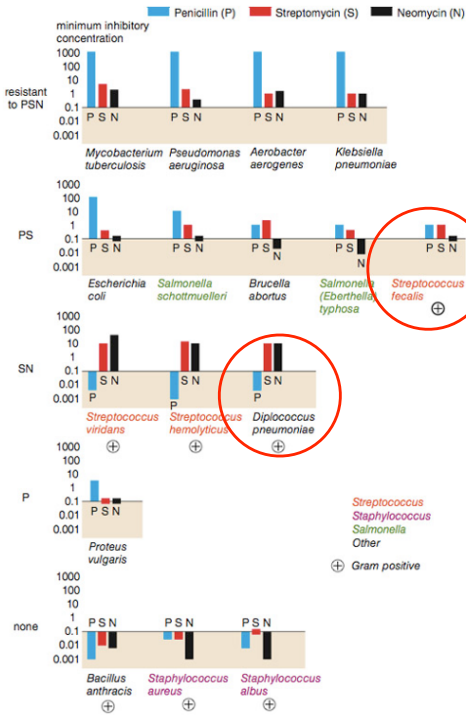
March 19, 2019

“There are three kinds of lies:
lies, damned lies, and **statistics.**”

Visualisasi

Sebagai analis data yang baik,
penting sekali untuk memvisualisasikan data.

Mengapa?

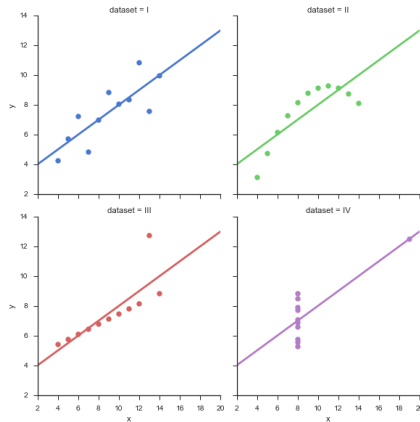


How do the bacteria compare?

Not a streptococcus!
(realized ~30 years later)

Really a streptococcus!
(realized ~20 years later)

Visualisasi vs Summary Statistics

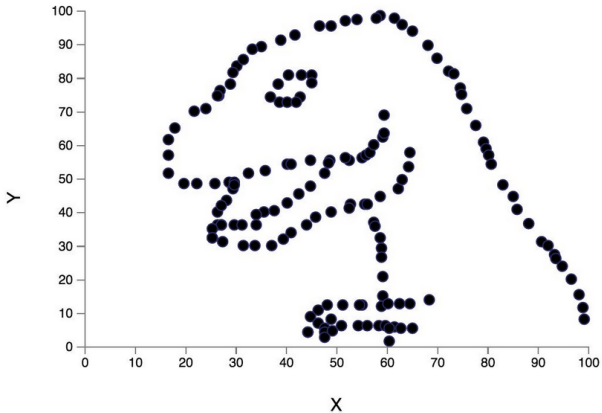


Gambar: *The infamous Anscombe's quartet*

Keempat data tersebut punya
 $\mu_x, \mu_y, \sigma_x, \sigma_y, N, \rho$ yang sama!

*Don't trust summary statistics.
Always **visualize** your data!*

N = 157 ; X mean = 50.7333 ; X SD = 19.5661 ; Y mean = 46.495 ; Y SD = 27.2828 ;
Pearson correlation = -0.1772



Alberto Cairo @albertocairo · 15 Aug 2016

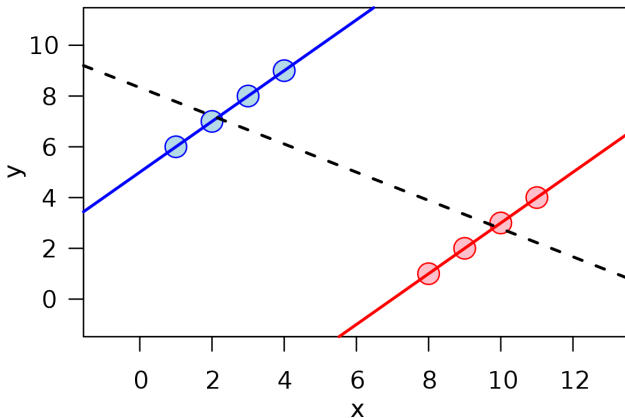
Don't trust summary statistics. Always visualize your data first robertgrantstats.co.uk/drawmydata.html pic.twitter.com/5j94Dw9UAf

13

895

984

Simpson's Paradox



Gambar: Kesalahan dalam interpretasi pola karena tidak dibagi per grup

Dengan visualisasi, bukan berarti bebas masalah.

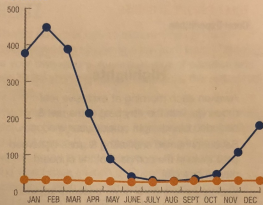
Kita tetap bisa berbohong!

WEDDING BUDGET BREAKDOWN

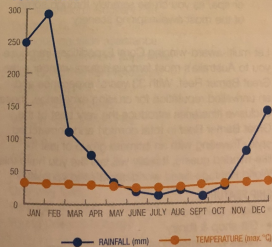


TEMPERATURE & RAIN CHART

CAIRNS

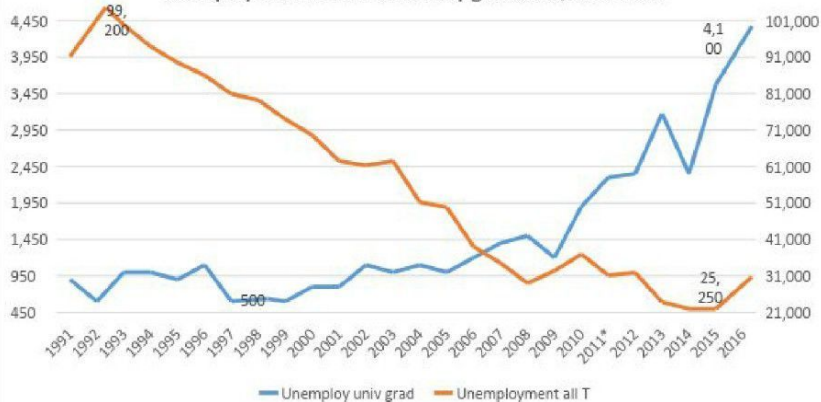


TOWNSVILLE



● RAINFALL (mm) ● TEMPERATURE (max. °C)

Figure 1: Trends in the national unemployment level and the unemployment level of university graduates, 1991-2016





Ainun Najib
@ainunnajib

Following



A good example of “How to Lie with Statistics”. Maaf mas dokter [@Gamal_Albinsaid](#) , we’re good friends, namun kali ini anda keliru dan mengelirukan masyarakat.



dr. Gamal Albinsaid [@Gamal_Albinsaid](#)

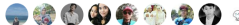
3. Ini bukan soal angka, ini soal luka, Luka Indonesiaku.

Saya katakan ini karena saya menyadari betapa beratnya tugas yang menunggu kita semua di depan sana....

Show this thread

3:02 PM - 7 Jan 2019 from [Central Region, Singapore](#)

5,253 Retweets 3,948 Likes

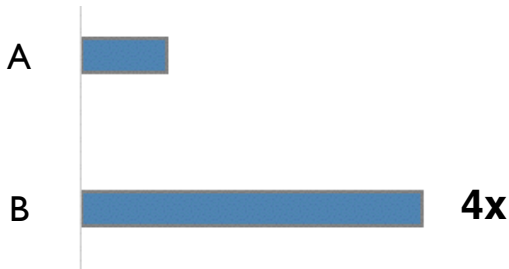


253 5.3K 3.9K

Gambar: Dapat digunakan untuk keuntungan politik. Sumber: @ainunnajib

Visualisasi Efektif

How much longer?



How much steeper slope?



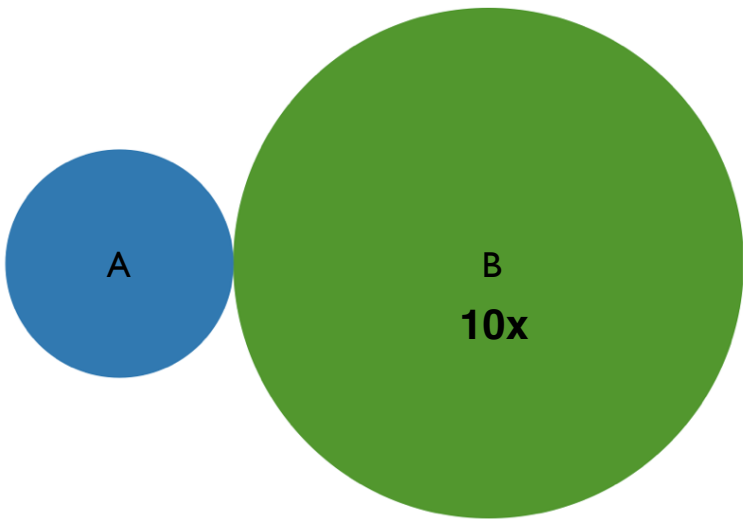
A

4x



B

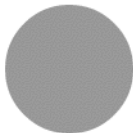
How much larger area?



How much darker?



A



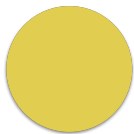
B

2x

How much bigger value?



A



B

4x



Most
Efficient



Least
Efficient

Position



Length



Slope



Angle



Area



Intensity



Color



Shape

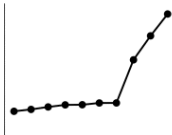


Quantitative

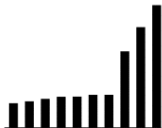
Ordered

Categories

Most Effective



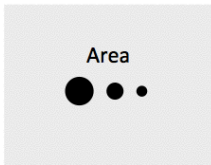
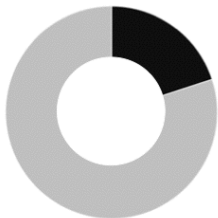
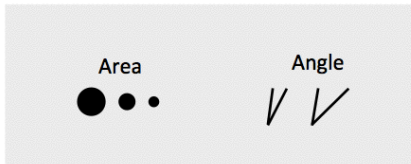
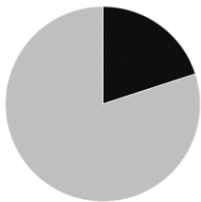
Position



Length

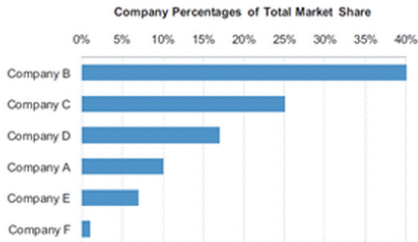
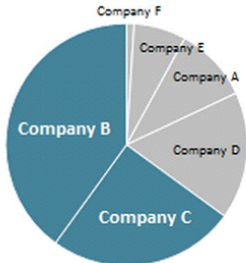


Less Effective



Pie vs. Bar Charts

65% of the market is controlled by companies B and C



Least Effective

SANFORD AND SELNICK

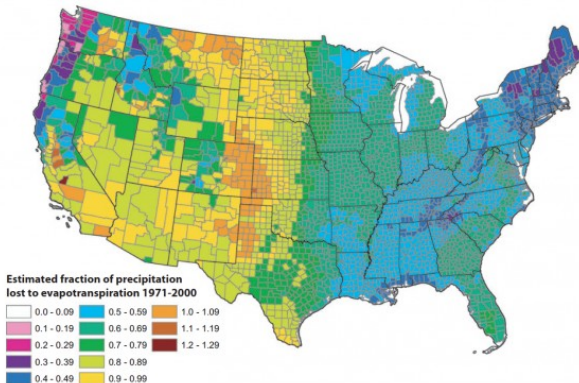


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (P) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/P were made first at the 800-m resolution of the PRISM climate data. The mean values for the counties (shown) were then calculated by averaging the 800-m values within each county. Areas with fractions >1 are agricultural counties that either import surface water or mine deep groundwater.

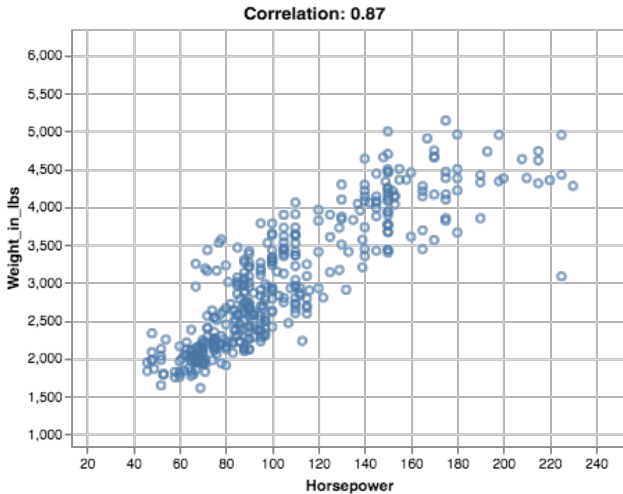
CS109 Data Science

Lecture 3: Exploratory Data Analysis

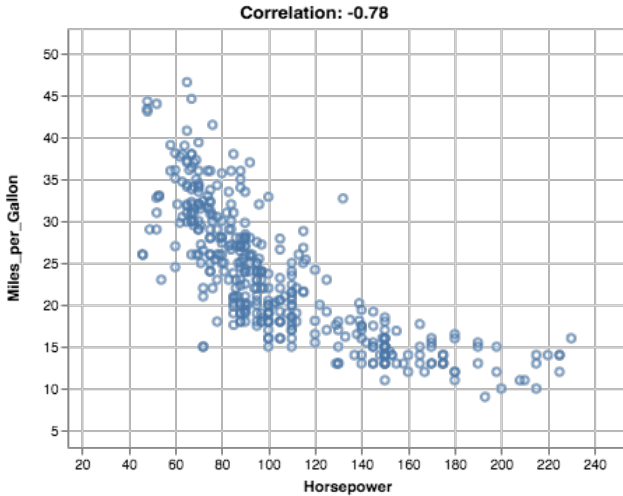
Korelasi

Korelasi

- Korelasi mengukur hubungan linear antarvariabel
- Nilai korelasi ada di rentang $[-1, 1]$
- Korelasi = 1 berarti berbanding lurus sempurna
- *Correlation doesn't imply causation!*

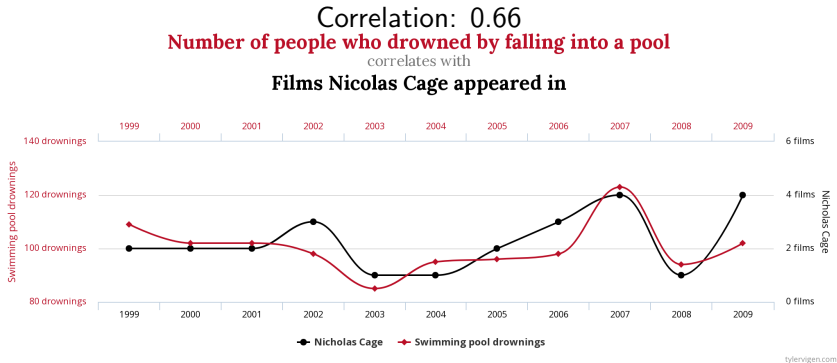


Gambar: Semakin besar tenaga kuda, mobil juga semakin berat



Gambar: Semakin besar tenaga kuda, semakin boros

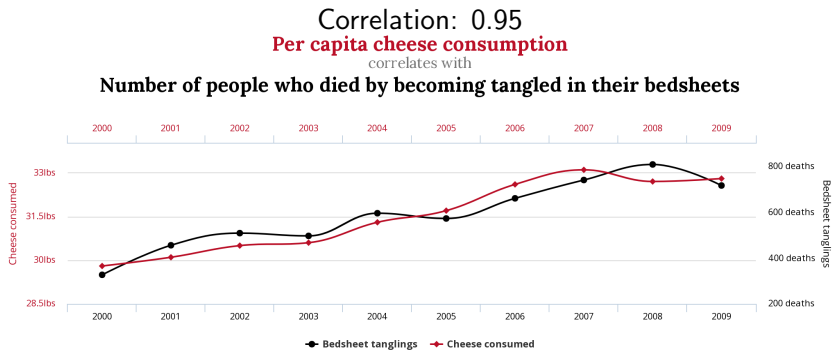
Spurious Correlation 1



Gambar: Apakah kita harus meminta Nicolas Cage berhenti main film?



Spurious Correlation 2



tylervigen.com

<http://www.tylervigen.com/spurious-correlations>



© marketoonist.com

Pekan depan:

Kombinatorika

Terima kasih