Visual Analytics

Communicating data-driven insights through data visualization techniques and useful dashboards

Miren Berasategi
miren.berasategi@deusto.es

Deusto

Introduction

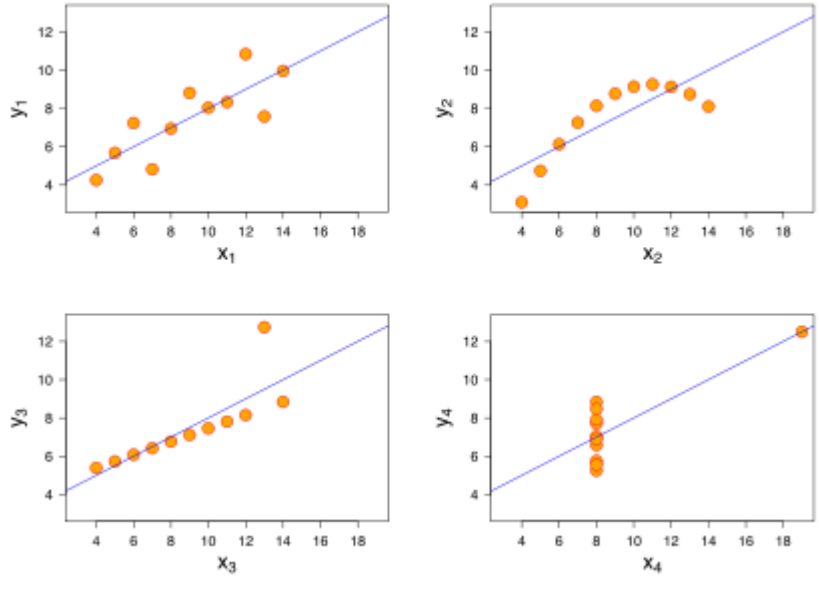
Definition(s)

Why use visualizations

Anscombe's quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Property	Value	Accuracy
Mean of *x*	9	exact
Sample variance of *x*	11	exact
Mean of *y*	7.50	to 2 decimal places
Sample variance of *y*	4.125	plus/minus 0.003
Correlation between *x* and *y*	0.816	to 3 decimal places
Linear regression line	$y = 3.00 + 0.500x$	to 2 and 3 decimal places, respectively



Anscombe's Quartet

What to use visualizations for

Munzner 2015, p.42

Graphics

Reminder: variable types

- Quantitative
  - Discrete
  - Continuous
- Qualitative
  - Ordinal
  - Nominal

Reminder: variable types
A question of time

Spatial and time/hour variables are special variable types.
Time variables are specially complex:

- are there 365 days in every year? 30 days in every month? 24 hours in every day?
- *timezones* make it even more complex to use hours or time of day

Reminder: variable types
A question of time

Time may be used as a continuous or as a qualitative variable.

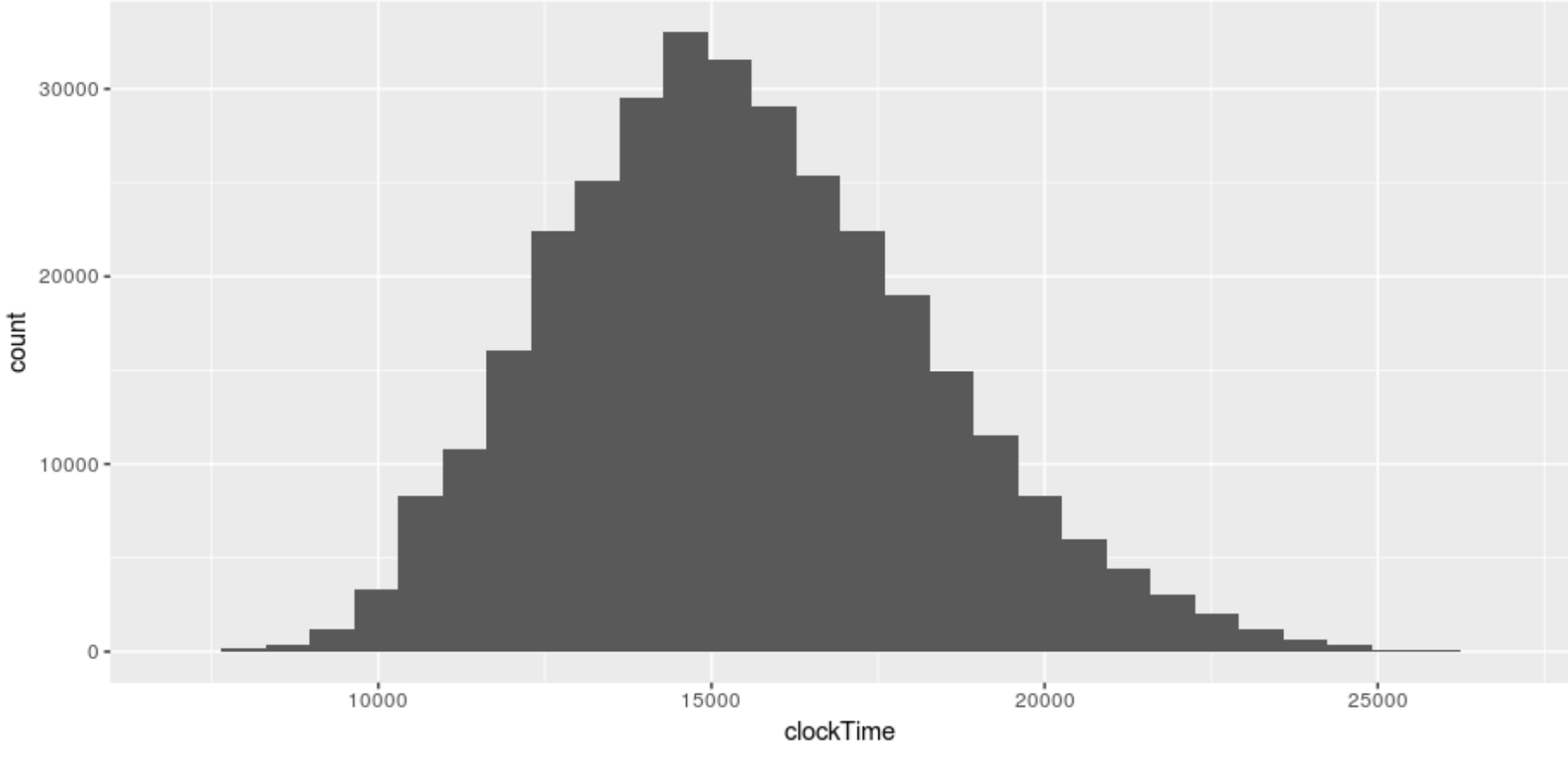
- as a qualitative variable, it has a hierarchy: year > month > (week >) day > hour > minute
- but different hierarchies may be necessary: bimonthly publications, multiple work shifts in a day...

Mapping variables to graphics

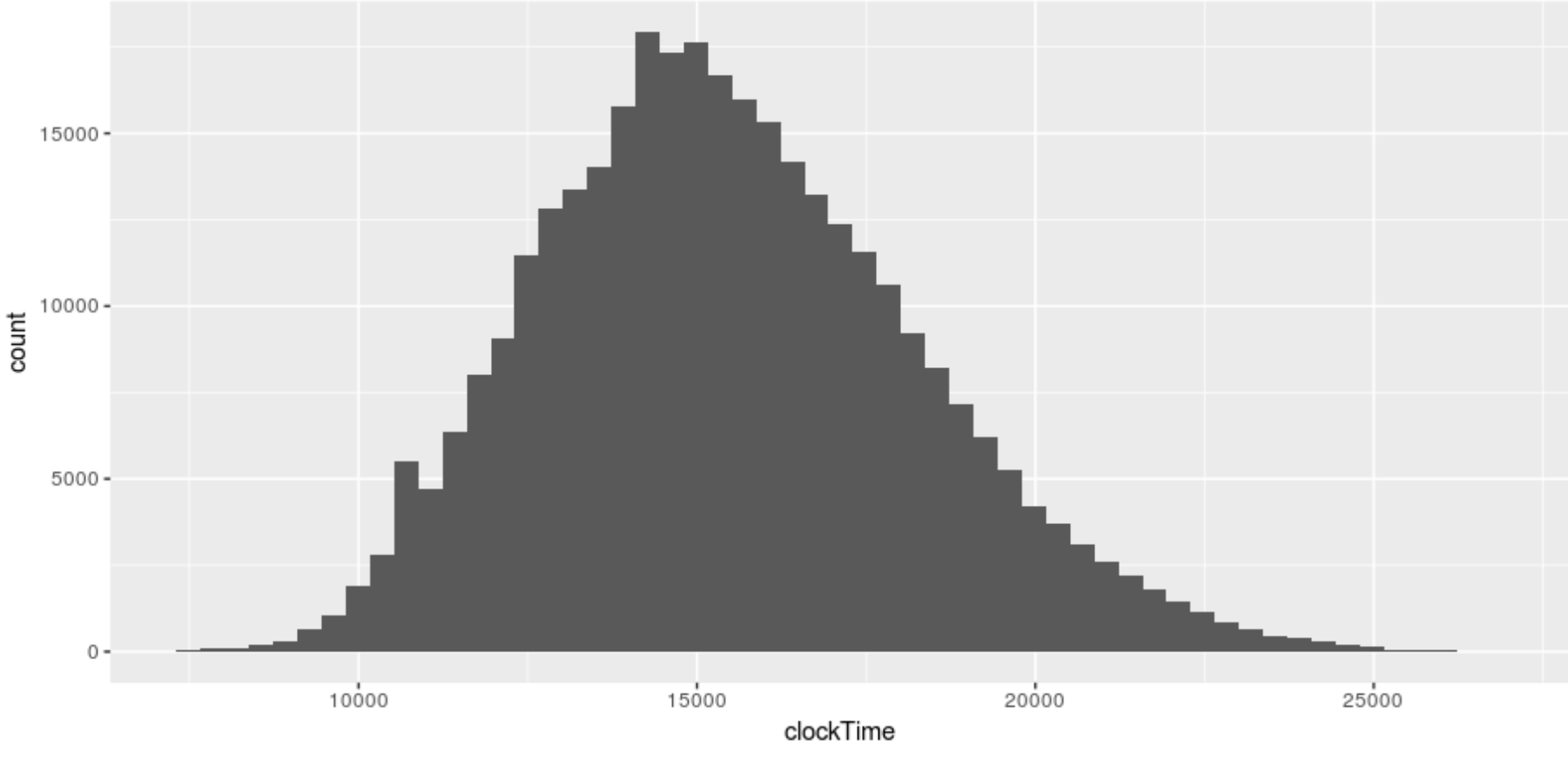
(Kirk 153-153)

Provide easier analysis

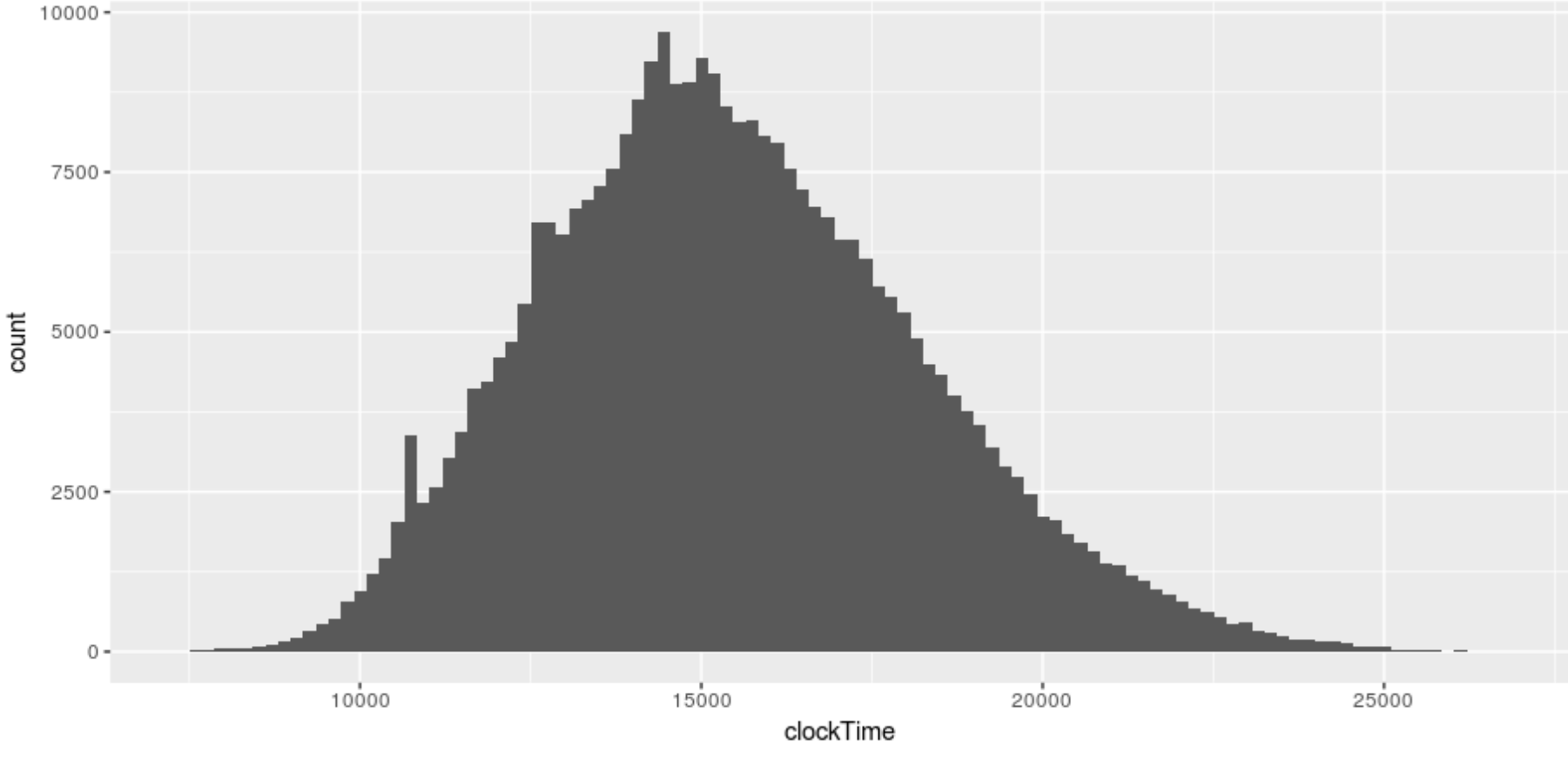
Change default settings



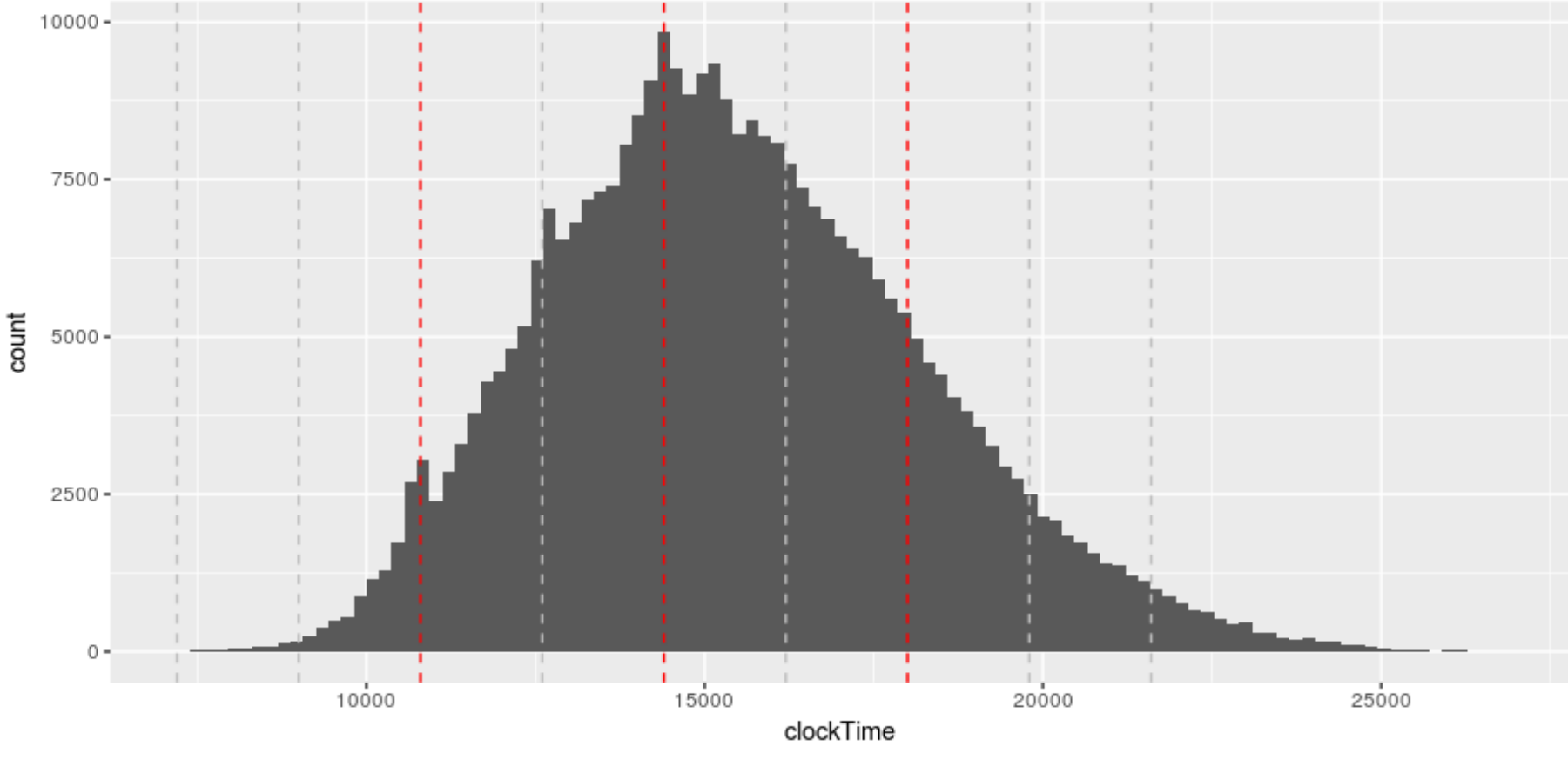
Change default settings



Change default settings



Change default settings

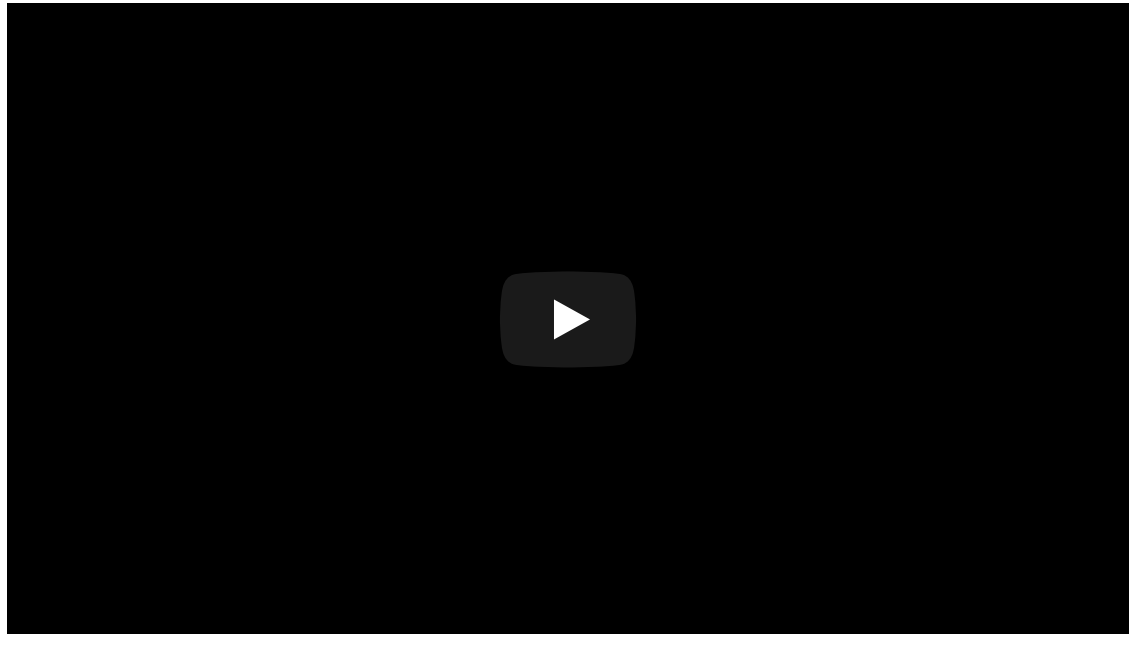


Make simpler graphs

> *Data-ink is the non-erasable core of the graphic, the non-redundant ink arranged in response to variation in the numbers represented. (Tufte 1983)*

Make simpler graphs

A step-by-step example: Data looks better naked



Nussbaumer, Declutter Your Data Visualizations

Highlight observations

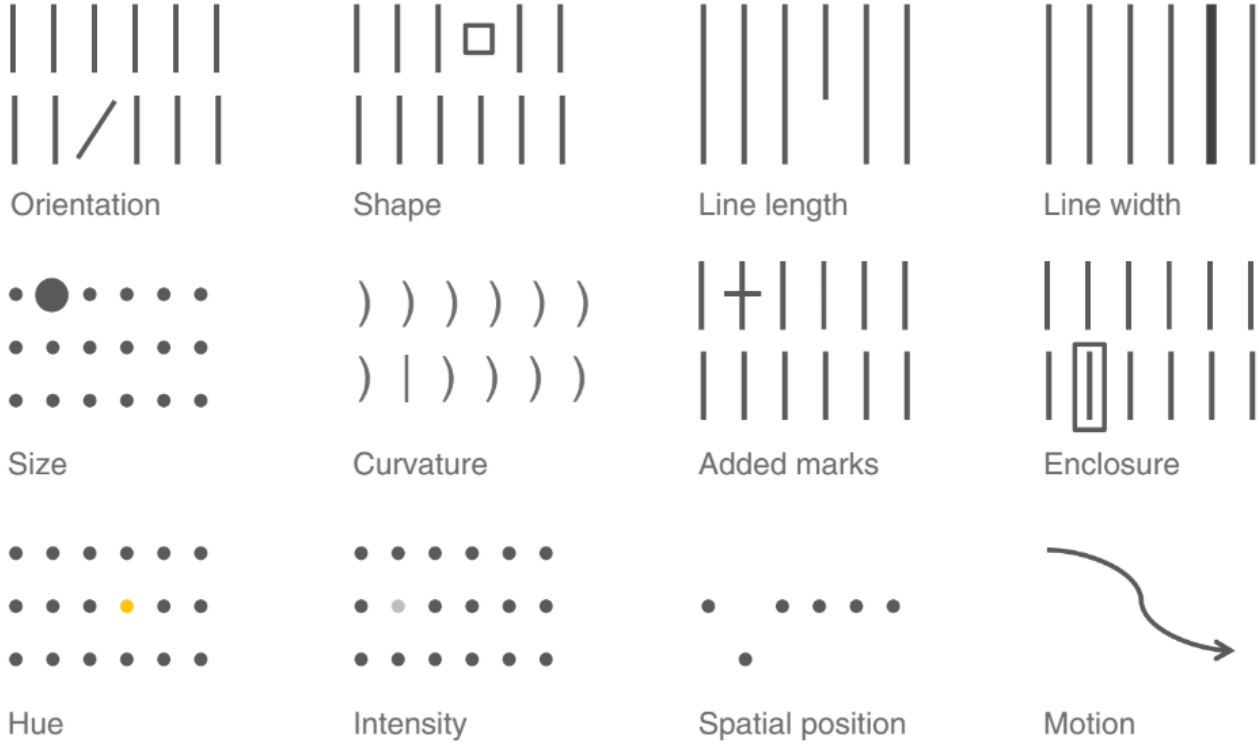


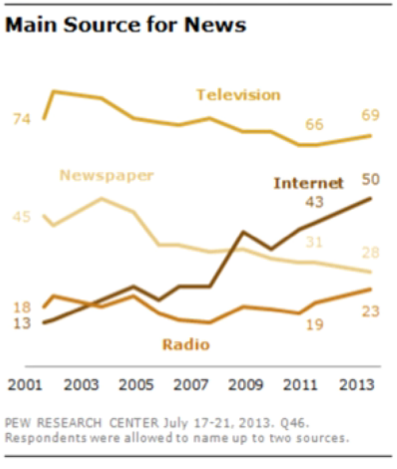
FIGURE 4.4 Preattentive attributes

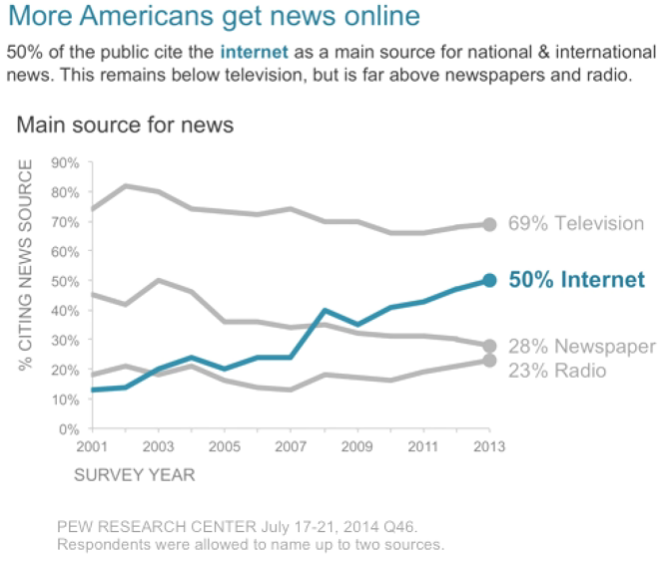
Source: Adapted from Stephen Few's *Show Me the Numbers*, 2004.

Nussbaumber xxxx, p.105

Highlight observations

1. More Americans get news online... 50% of the public now cites source for national and international news, still below television, newspapers and radio. (Report)





Source: http://www.pewresearch.org/fact-tank/2013/10/16/12-trends-shaping-digital-news/

storytelling with data

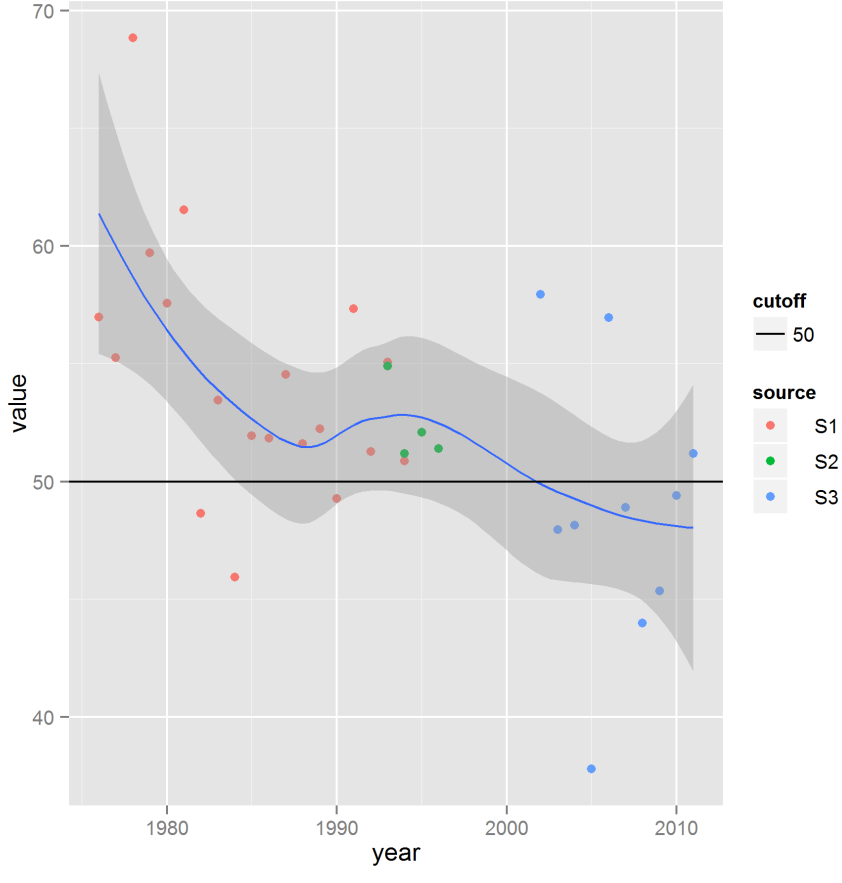
Nussbaumber, Do you see it? The importance of contrast when communicating with data [video]

Add variables (as context)

- Adding preexisting variables (con **mesura**)
- Creating conditional variables from preexisting variables
  - binaries or with few levels are best
  - example of calculated field or variable: weekend date

Add statistical information

- statistical summaries
- models



source

Test slides

First one

2nd level heading

Testing background images

3rd level heading

Second one with content