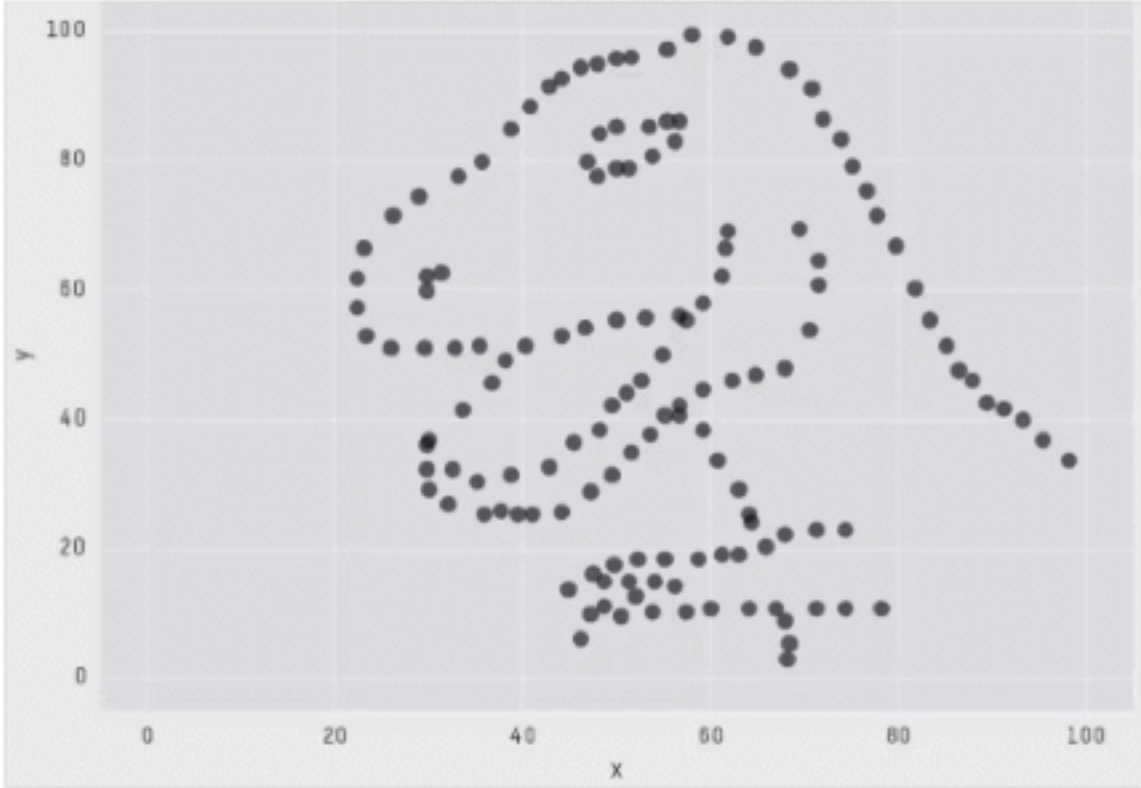
IN·NOMINE·PATRIS

Radboud University Nijmegen

Behavioural Science Institute

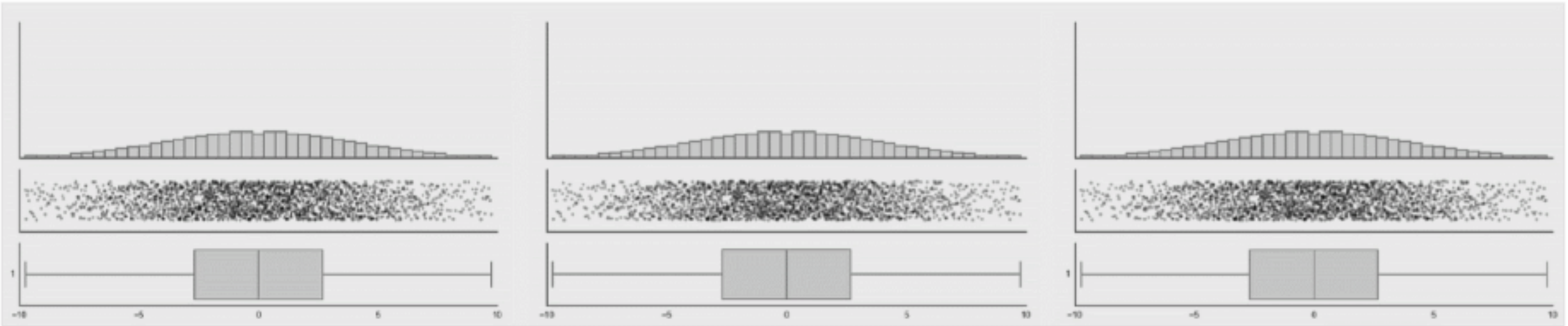

2

“Analyse then Aggregate!” same stats - different patterns

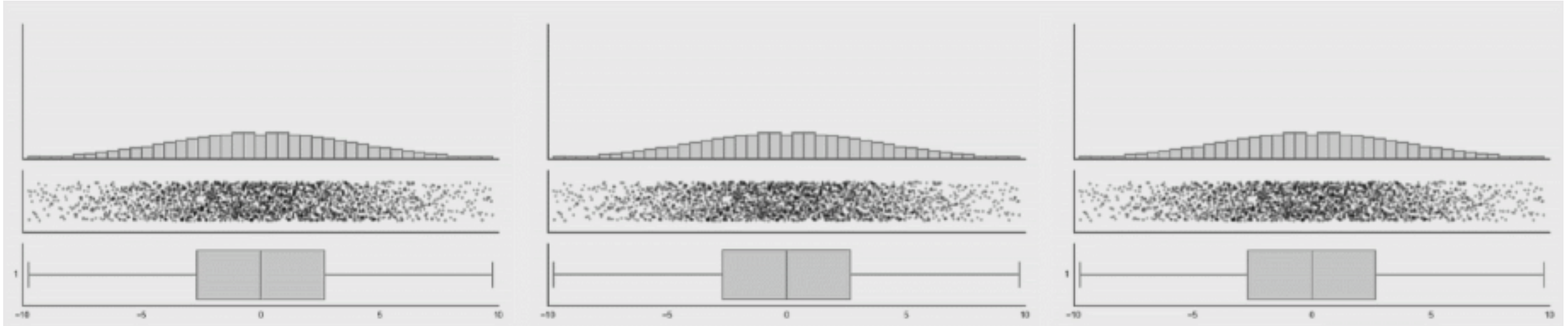


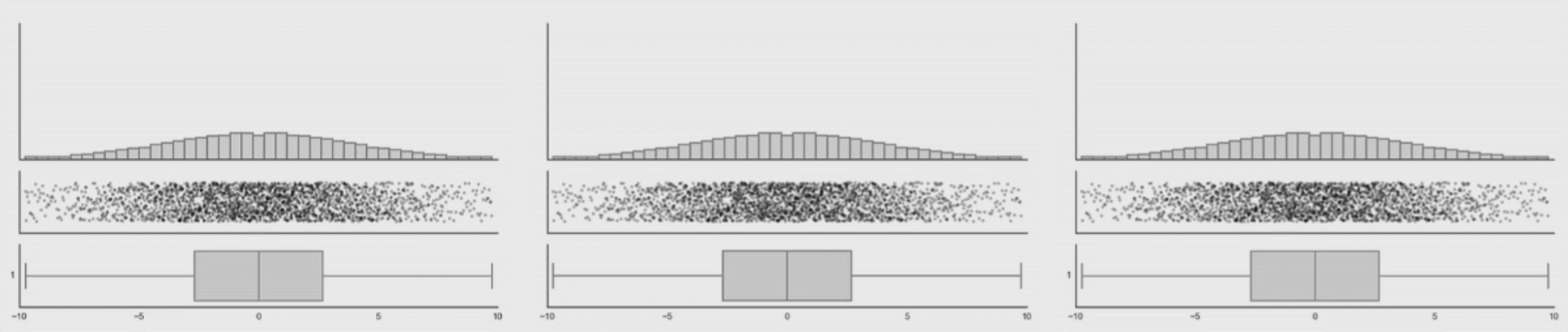
X Mean: 54.2659224
Y Mean: 47.8313999
X SD : 16.7649829
Y SD : 26.9342120
Corr. : -0.0642526

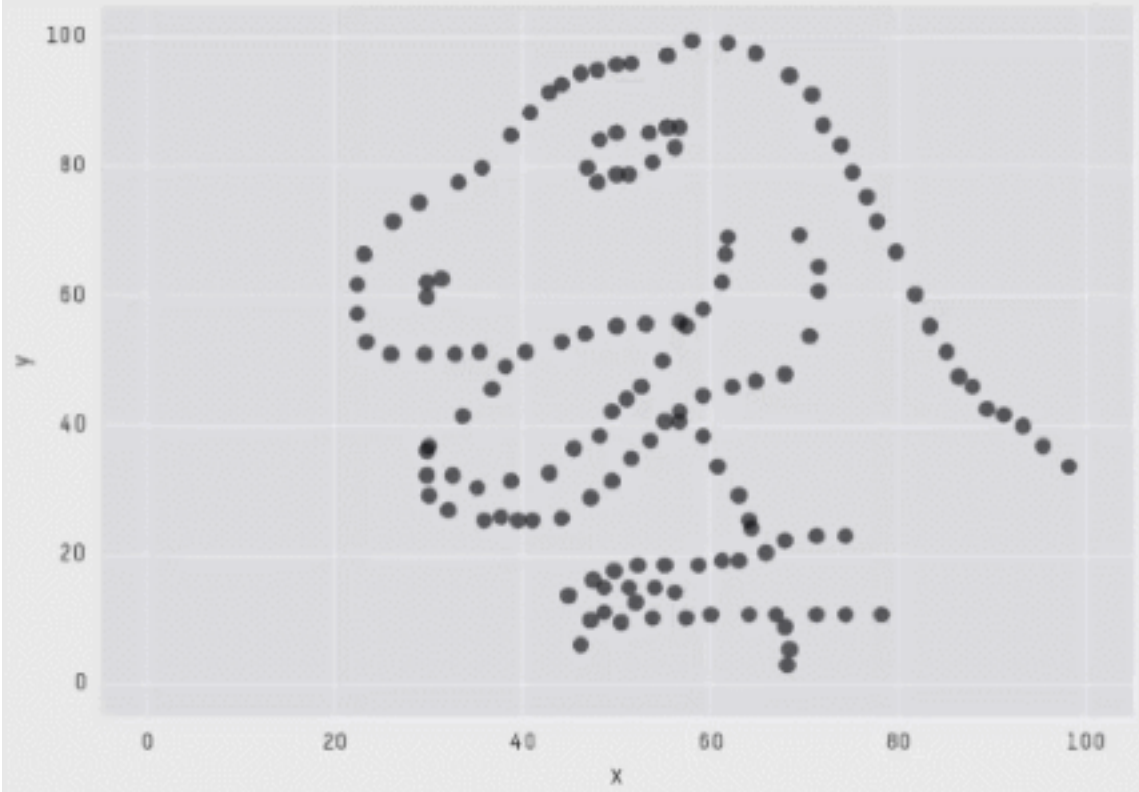
https://www.autodeskresearch.com/publications/samestat


1
−10
−5
0
5
10
1
−10
−5
0
5
10
1
−10
−5
0
5
10

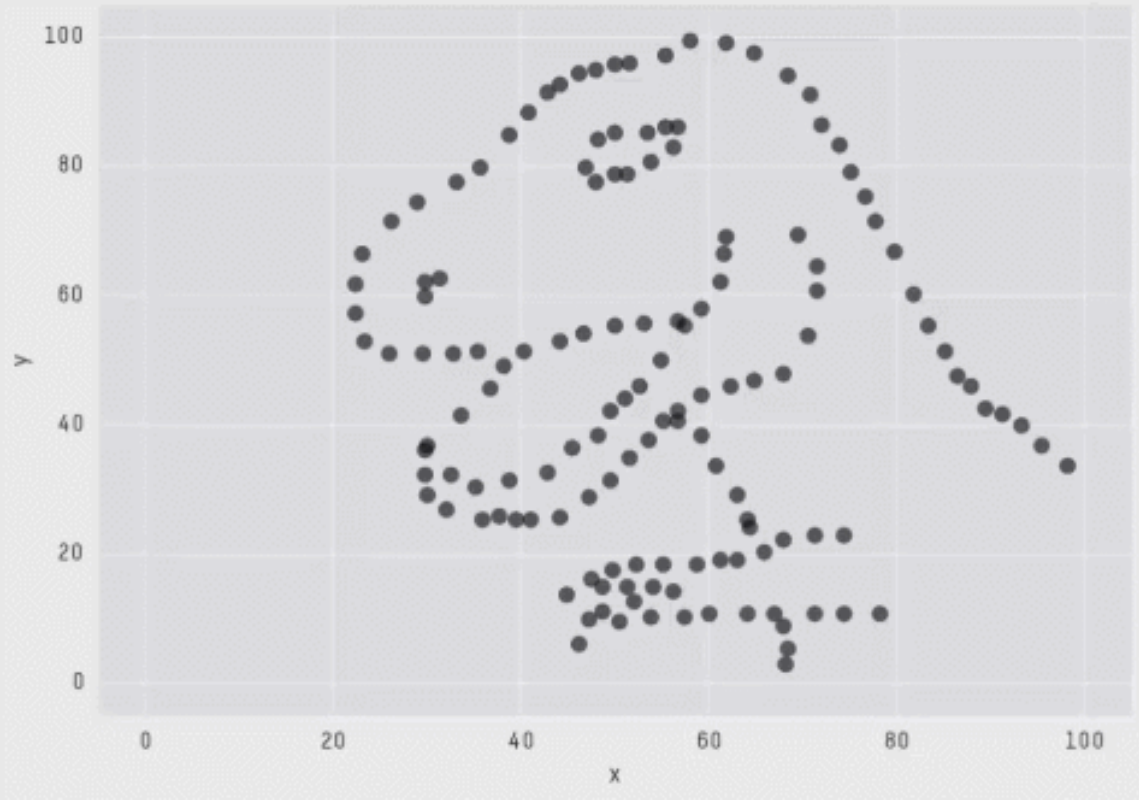
Matejka, J., & Fitzmaurice, G. (2017, May). Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Sys*


1
−10
−5
0
5
10
1
−10
−5
0
5
10
1
−10
−5
0
5
10


1
−10
−5
0
5
10



X Mean: 54.2659224
Y Mean: 47.8313999
X SD : 16.7649829
Y SD : 26.9342120
Corr. : -0.0642526



X Mean: 54.2659224
Y Mean: 47.8313999
X SD : 16.7649829
Y SD : 26.9342120
Corr. : -0.0642526

Fundamental problems for main-stream Social & Life Sciences

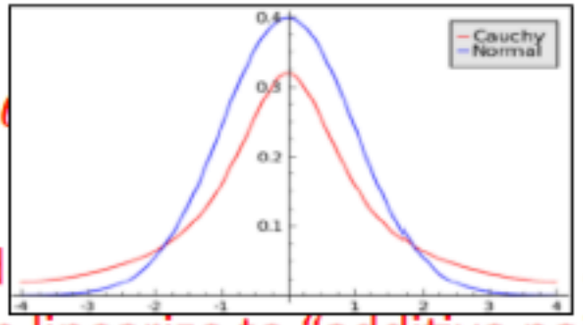

3

Press, W. H. (2014). Reproducibility Now at Risk? Paper presented at the Symposium on Evidence in the Natural Sciences New York, NY. https://www.simonsfoundation.org/event/symposium-on-evidence-in-the-natural-sciences/

Most of this talk is about human frailties, but some deeper foundational issues are also worth mentioning.

$$y = f(x)$$ "Discover" f by controlling x, measuring y

$$y = f(x;\theta)$$ But f also depends on unknown parameters (θ) which must be determined from the data.



$$y = f(x;\theta,R)$$ Of course the result also depends on rand[om] R in an arbitrarily nonlinear way – which we often linearize to "additive noise".

$$\langle y \rangle \approx \langle f(x;\theta,R) \rangle$$ So we are now measuring relations between expectations – if they exist (cf. Cauchy distribution).

$$\langle y(S) \rangle \approx \langle f(x;\theta,R,S) \rangle$$ Systematic errors are additional long-term random variables that don't average away.

$$P_{Y(S)}(y(S)) = \langle f(y,x;\theta,R,S) \rangle$$

Finally, y may itself be intrinsically probabilistic, as in quantum measurement or classical chaos (e.g., turbulence).