University of Bremen

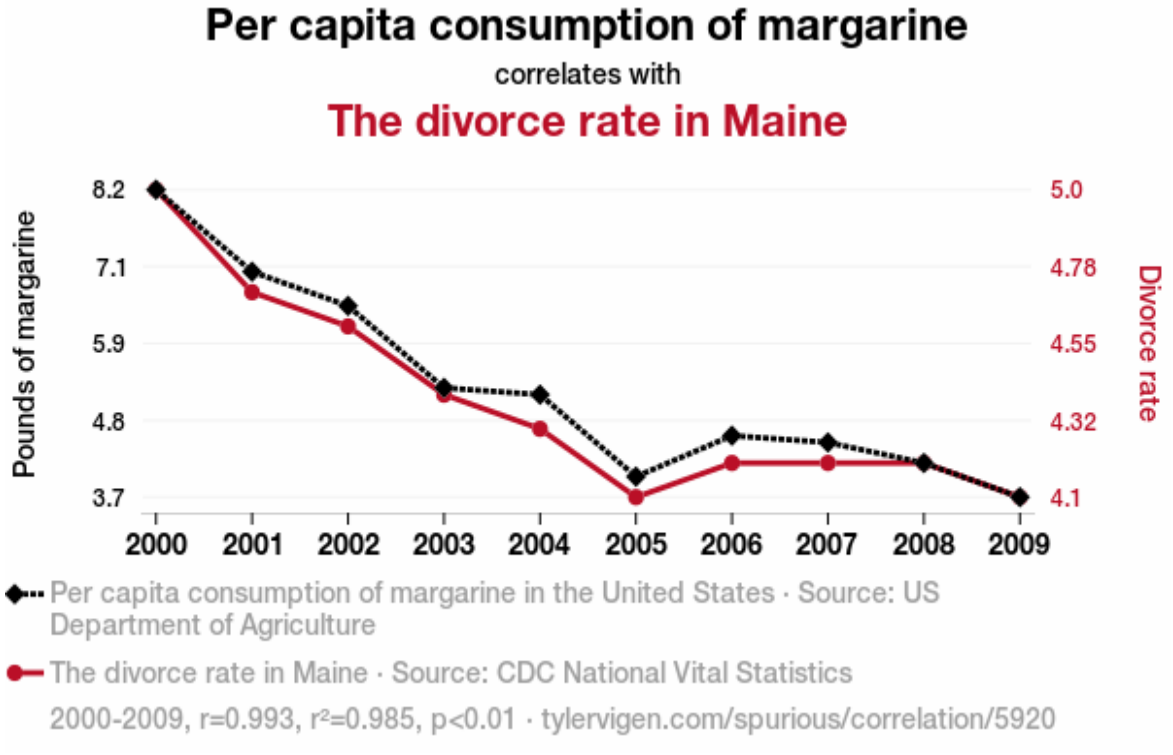
Global Dynamics of Social Policy
CRC 1342

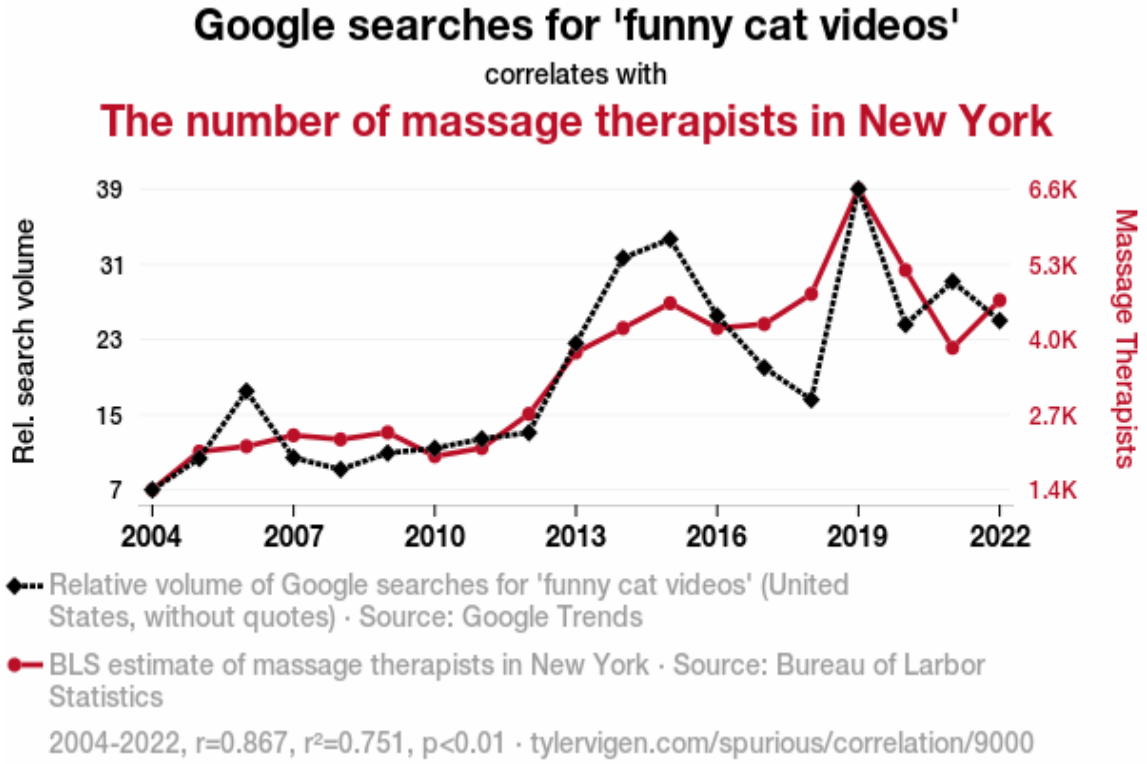
CSS Pitch

“Correlation is not causation” – Exploring & Visualizing Correlations

Nils Düpont
CRC 1342, Project INF

✉ duepont@uni-bremen.de



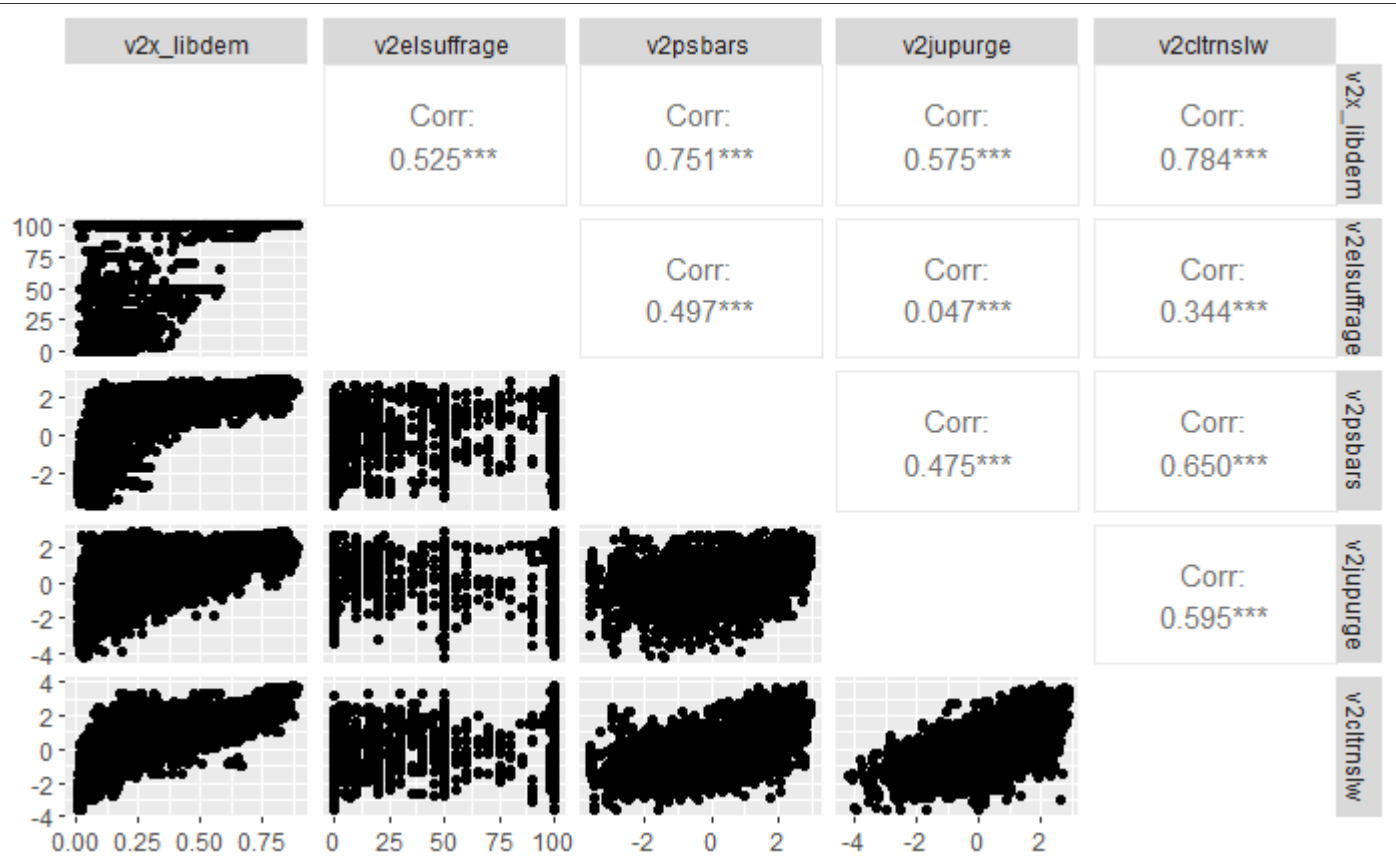


For more spurious correlations check: https://tylervigen.com/spurious-correlations

- Correlation is not causation, but…
- …an essential part in the social sciences
  - in inductive data exploration
  - in proposing new hypotheses
  - in theory building

- Correlation analysis gains importance as the volume of data increases
- Correlation analysis not new to the social sciences
  - usually conducted in statistical software like R or Stata
  - becomes “cumbersome” as the number of variables increases

	v2x_polyar
v2x_polyarchy	1.
v2x_libdem	0.
v2elsuffrage	0.
v2psbars	0.
v2jupurge	0.
v2cltrnslw	0.
v2mecenefm	0.
v2pepwrses	0.
v2exl_legitideol	-0.
v2dlcommon	0.



v2exl_legitideol	v2dlcommon
-0.387	0.661
-0.423	0.631
0.054	0.571
-0.425	0.586
-0.353	0.325
-0.401	0.583
-0.501	0.528
-0.045	0.651
1.000	-0.125
-0.125	1.000

- “Correlation is not causation” project pitch
  - Up to ten variables
  - Interactive → dashboard?
  - Different ways of visualizing correlations → new ones?
  - Verbalize correlations → integrate generative AI models?

- The data set: “Varieties of Democracy” (https://www.v-dem.net/)
  - V-Dem captures the multidimensionality of *democracy*
    - high-level: electoral, liberal, participatory, deliberative, and egalitarian democracy
  - 4000+ variables, 27.000+ country-year observations
  - almost complete data (← quite rare in social sciences)