



CSS Pitch

"Correlation is not causation" – Exploring & Visualizing Correlations

Nils Düpont CRC 1342, Project INF

∠ duepont@uni-bremen.de

Per capita consumption of margarine correlates with The divorce rate in Maine 8.2 5.0 Pounds of margarine 7.1 4.78 Divorce rate 5.9 4.55 4.32 4.8 2002 2003 2004 2008 2001 2005 2006 2007 2009 ◆ Per capita consumption of margarine in the United States · Source: US Department of Agriculture ■ The divorce rate in Maine · Source: CDC National Vital Statistics 2000-2009, r=0.993, r²=0.985, p<0.01 · tylervigen.com/spurious/correlation/5920

Google searches for 'funny cat videos' correlates with The number of massage therapists in New York 39 6.6K Rel. search volume 5.3K 31 23 15 2.7K 1.4K 2010 2013 2004 2007 2016 2019 2022 ← Relative volume of Google searches for 'funny cat videos' (United States, without quotes) · Source: Google Trends ■ BLS estimate of massage therapists in New York · Source: Bureau of Larbor Statistics 2004-2022, r=0.867, r²=0.751, p<0.01 · tylervigen.com/spurious/correlation/9000

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



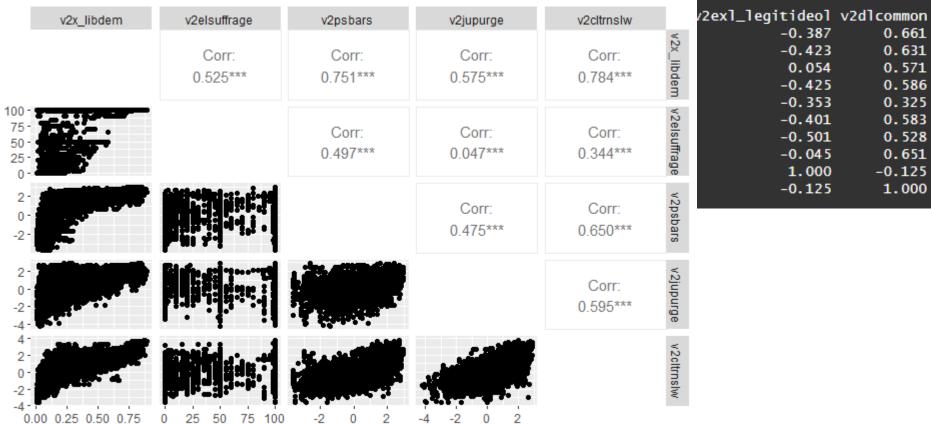


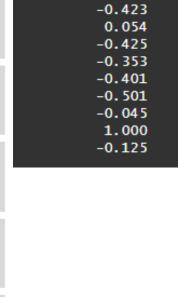
- Correlation is not causation, but...
- ...an essential part in the social sciences
 - in inductive data exploration
 - in proposing new hypotheses
 - in theory building
- Correlation analysis gains importance as the volume of data increases
- Correlation analysis not new to the social sciences
 - usually 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.
v2d1common	0.





-0.387

0.661

0.631

0.571

0.586

0.325

0.583

0.528

0.651

-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)



