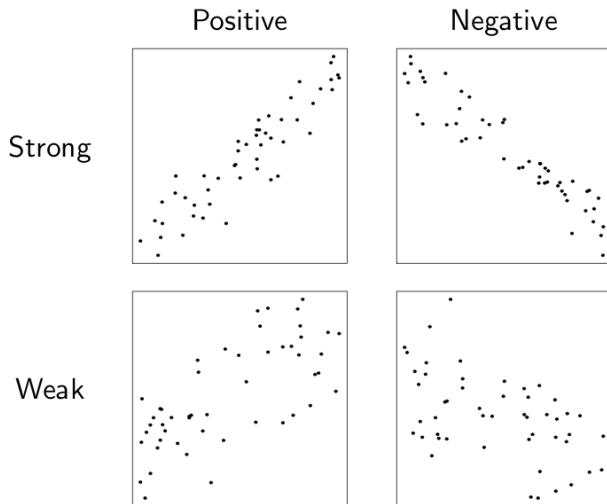


Lecture 2: Tables and Association

Kushal K Dey

04.05.2016

Association between quantitative variables



Categorical variables/ Qualitative variables

- ▶ We move away from quantitative variables to qualitative variables
- ▶ Variables whose values are categories, not numbers
- ▶ Categories may be ordered (ordinal, ex- Agree/Neutral/Disagree) or non-ordered (nominal, - Male/Female).
- ▶ How to analyze such data- Visualization and Inference

Look at some Data

	Dance	Sports	TV
Male	2	10	8
Female	16	6	8

Add Marginals

	Dance	Sports	TV	Total
Male	2	10	8	20
Female	16	6	8	30
Total	18	16	16	50

Joint distribution

	Dance	Sports	TV	Total
Male	0.04	0.2	0.16	0.4
Female	0.32	0.12	0.16	0.6
Total	0.36	0.32	0.32	1

Row marginal distribution

	Dance	Sports	TV	Total
Male	0.04	0.2	0.16	0.4
Female	0.32	0.12	0.16	0.6
Total	0.36	0.32	0.32	1

Column marginal distribution

	Dance	Sports	TV	Total
Male	0.04	0.2	0.16	0.4
Female	0.32	0.12	0.16	0.6
Total	0.36	0.32	0.32	1

Row conditional distribution

	Dance	Sports	TV	Total
Male	0.1	0.5	0.4	1
Female	0.53	0.2	0.27	1

Column conditional distribution

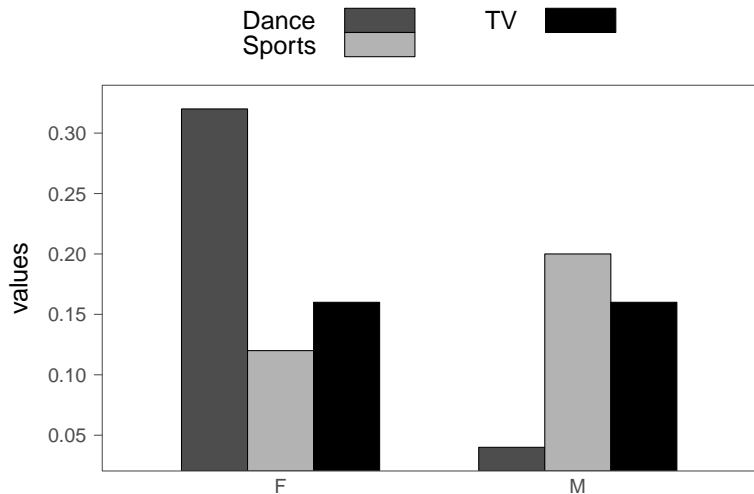
	Dance	Sports	TV
Male	0.11	0.63	0.5
Female	0.89	0.37	0.5
Total	1	1	1

Visualizing Tables I

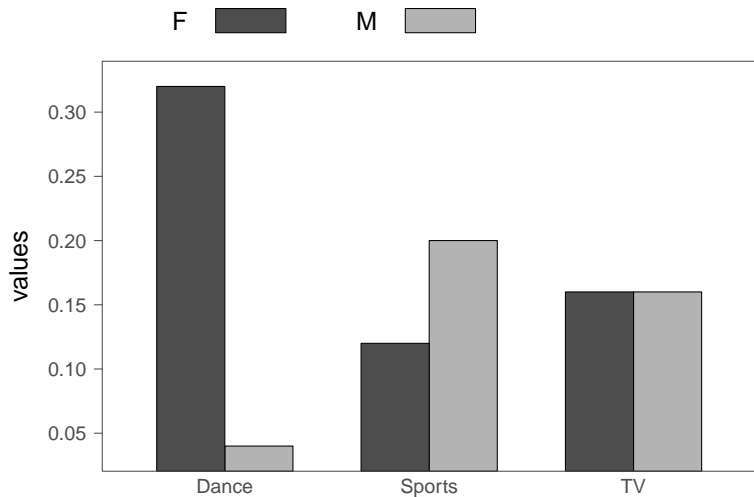
```
d <- data.frame(
  gender_id<- c("M","M","M","F","F","F"),
  entertain_id <- c("Dance", "Sports", "TV",
                   "Dance", "Sports", "TV"),
  values <- c(0.04, 0.2, 0.16, 0.32, 0.12, 0.16))

require(lattice)
barchart(values ~ gender_id, groups=entertain_id,
  data=d, auto.key = list(columns = 2))
```

Visualizing Tables II



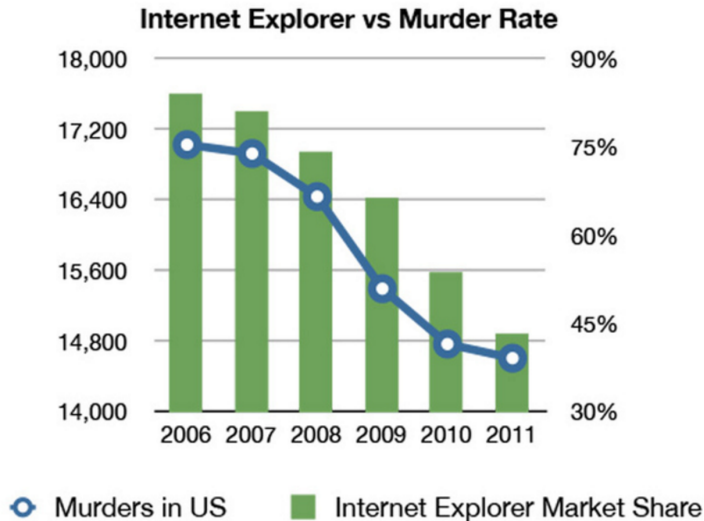
Visualizing Tables III



Confounders I

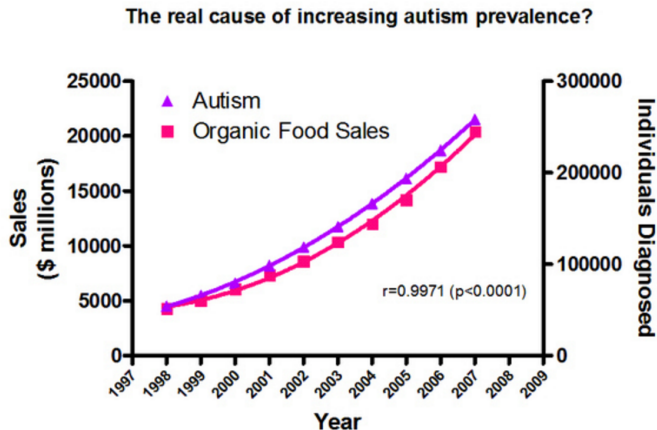
- ▶ From a table or visualization or scatter plot, one may infer two variables are associated (positively or negatively).
- ▶ That does not mean they are CAUSAL!!
- ▶ There may be confounders (hidden variables that may cause the association between observed variables).

When Statistics go Wrong!-I



When Statistics go Wrong! -II

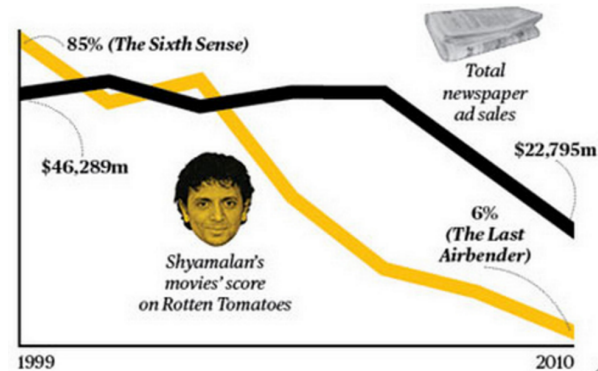
4. Eating organic food causes autism.



Sources: Organic Trade Association, 2011 Organic Industry Survey; U.S. Department of Education, Office of Special Education Programs, Data Analysis System (DANS), OMB# 1820-0043: "Children with Disabilities Receiving Special Education Under Part B of the Individuals with Disabilities Education Act"

When Statistics go Wrong!- III

5. M. Night Shyamalan makes bad movies because people don't buy newspapers.



businessweek.com

When Statistics go Wrong!- IV

10. Facebook also cancelled out the cholesterol-lowering effects of Justin Bieber.

