CORRELATION

1. Correlation
2. Correlation matrixes
3. Practice

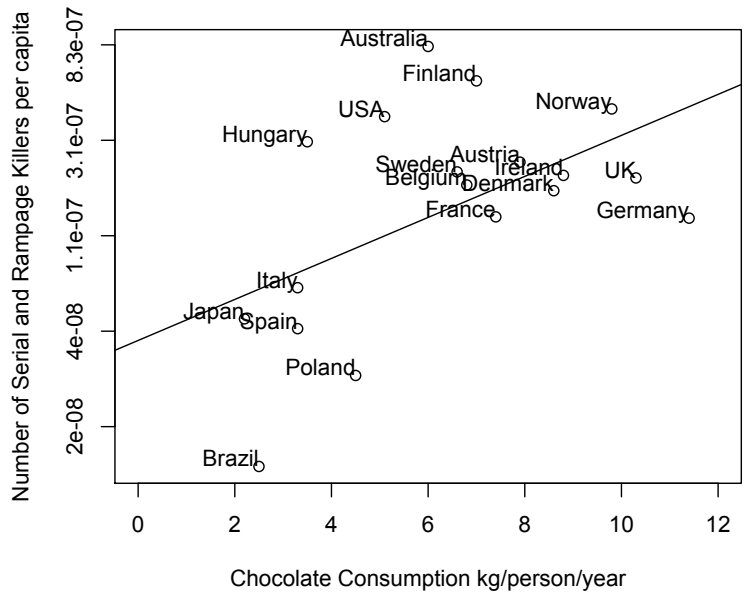
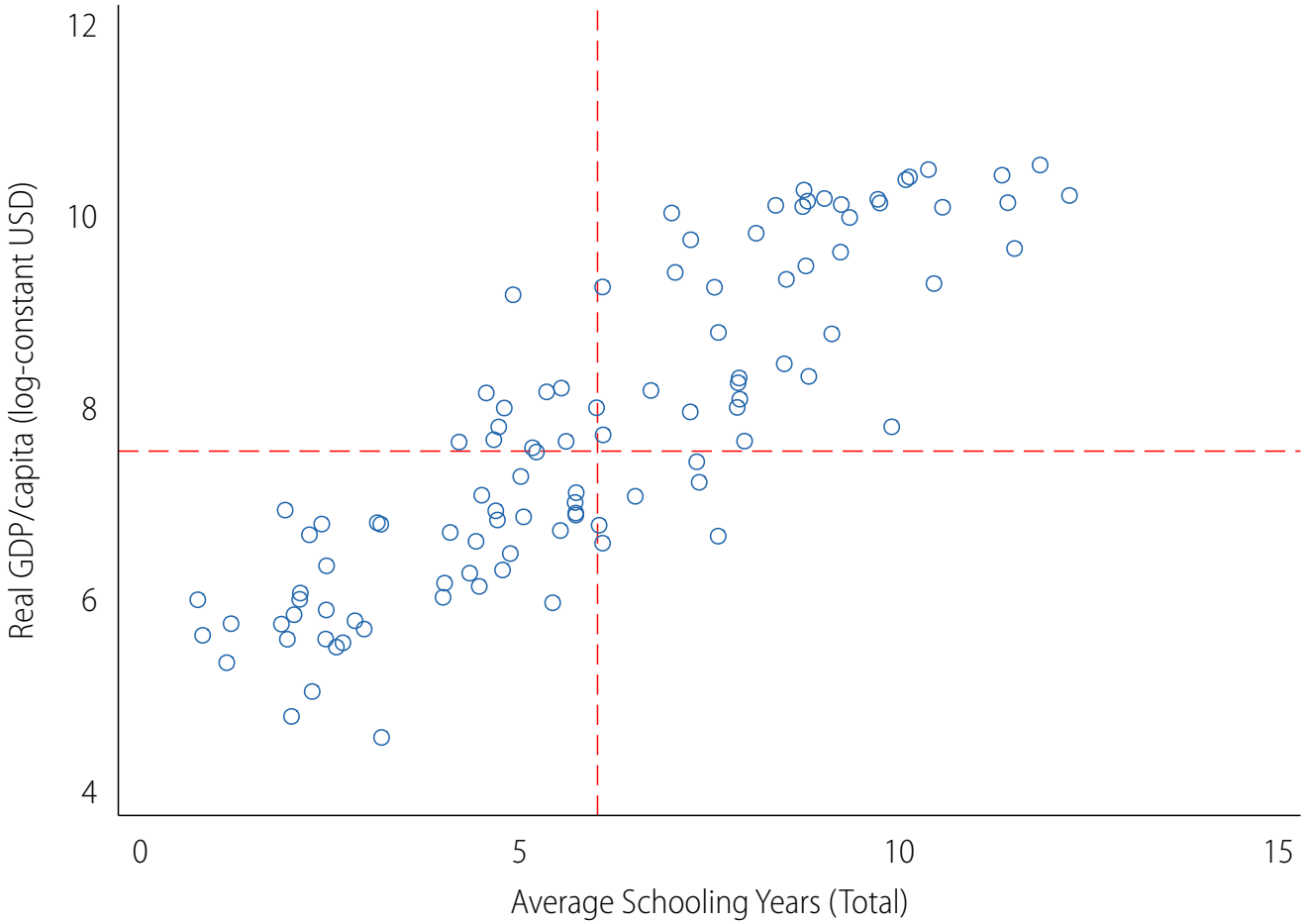


Figure 1: Correlation between countries' annual per capita chocolate consumption and the serial and rampage killers per capita since 1900.

1. Correlation



Pearson correlation coefficient

Measuring association as the linear dependence of two variables:

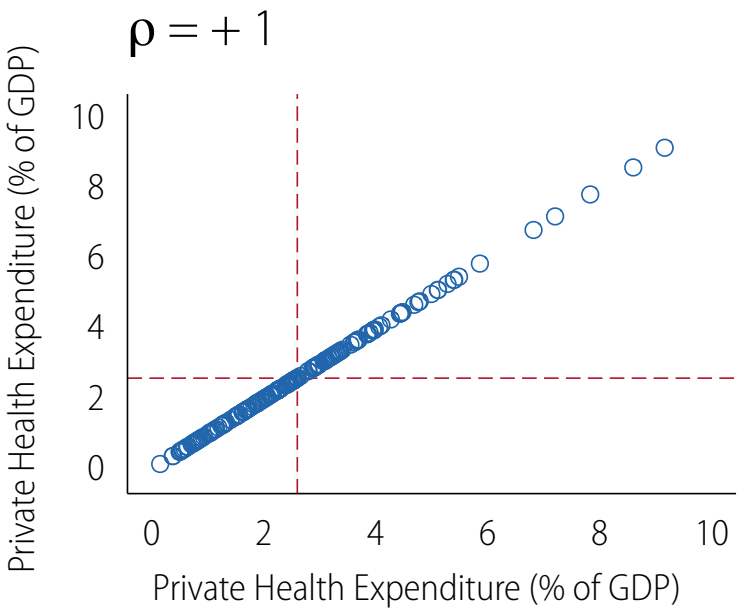
Population notation $\rho = \frac{\text{Cov}(X, Y)}{\text{Var}_X \text{Var}_Y}, \quad -1 \leq \rho \leq 1$

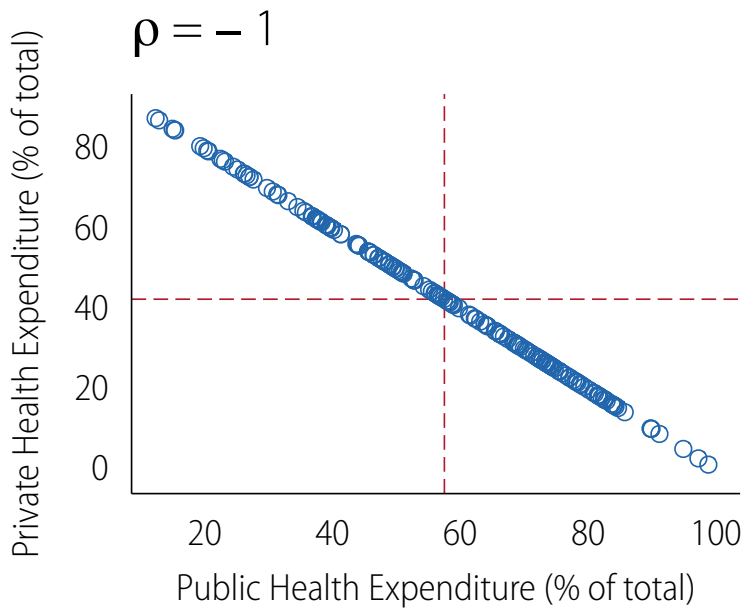
Sample notation $r = \frac{1}{n-1} \sum_{i=1}^{n} (\frac{X_i - \bar{X}}{s_X})(\frac{Y_i - \bar{Y}}{s_Y})$

Detects linear correlation

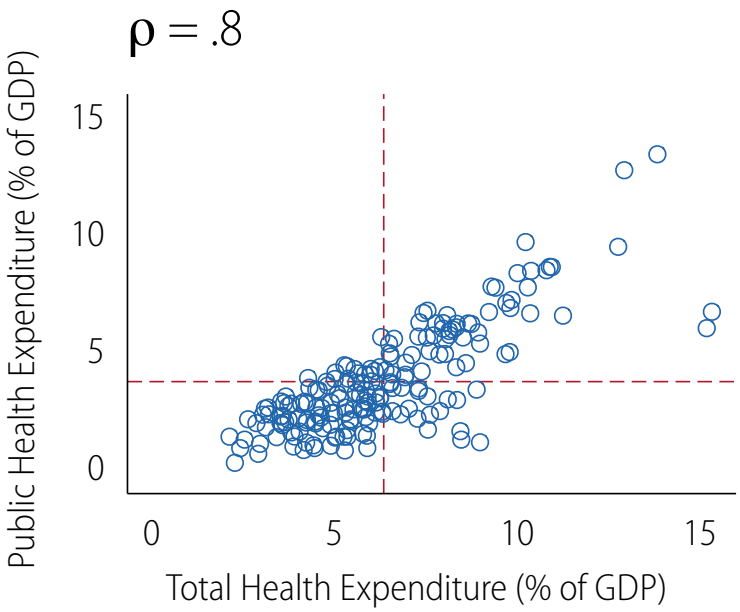
- Uncorrelated $\neq$ unrelated
- Correlated $\neq$ unconfounded

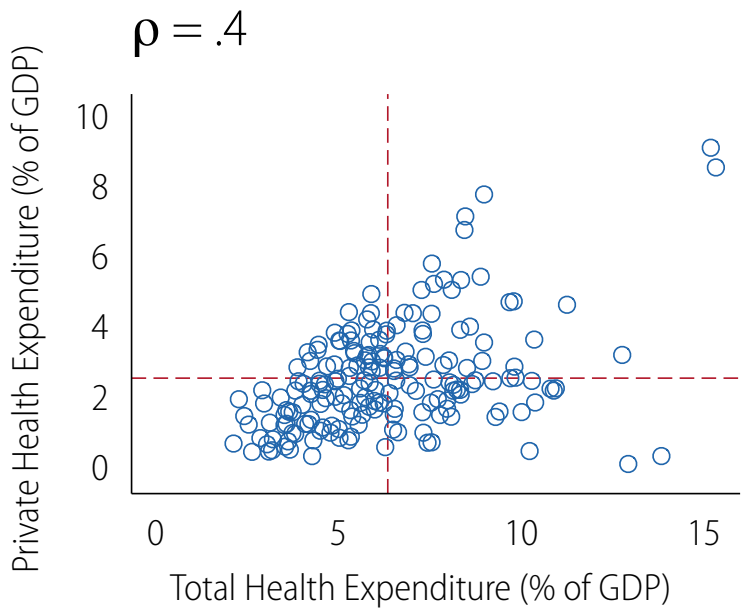
Perfect (positive, negative) correlation



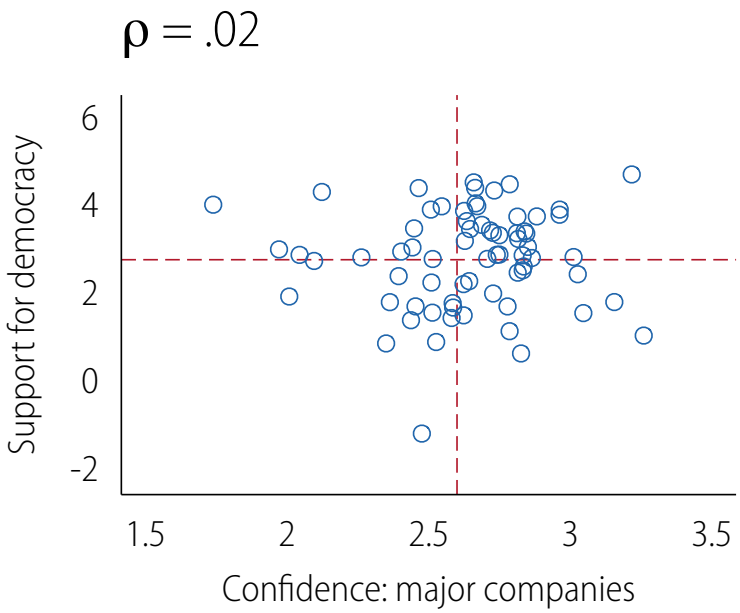


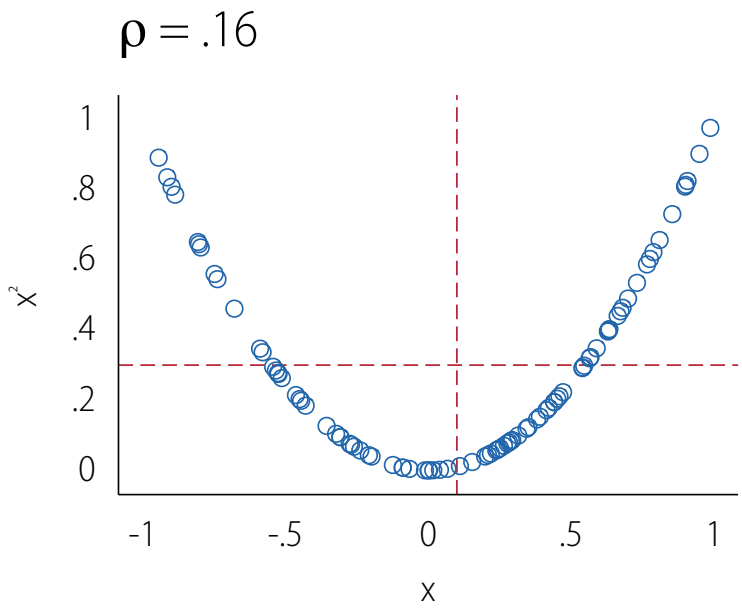
Significant (moderate, strong) correlation





Insignificant (weak, non-linear) correlation





Pearson correlation coefficient

Significance test:

$$\text{Null hypothesis } H_0 \quad r = 0$$

$$\text{Test statistic} \quad T = r\sqrt{\frac{n-2}{1-r^2}}$$

Sanity check

- Uncorrelated $\neq$ independent
- Correlated $\neq$ causally related

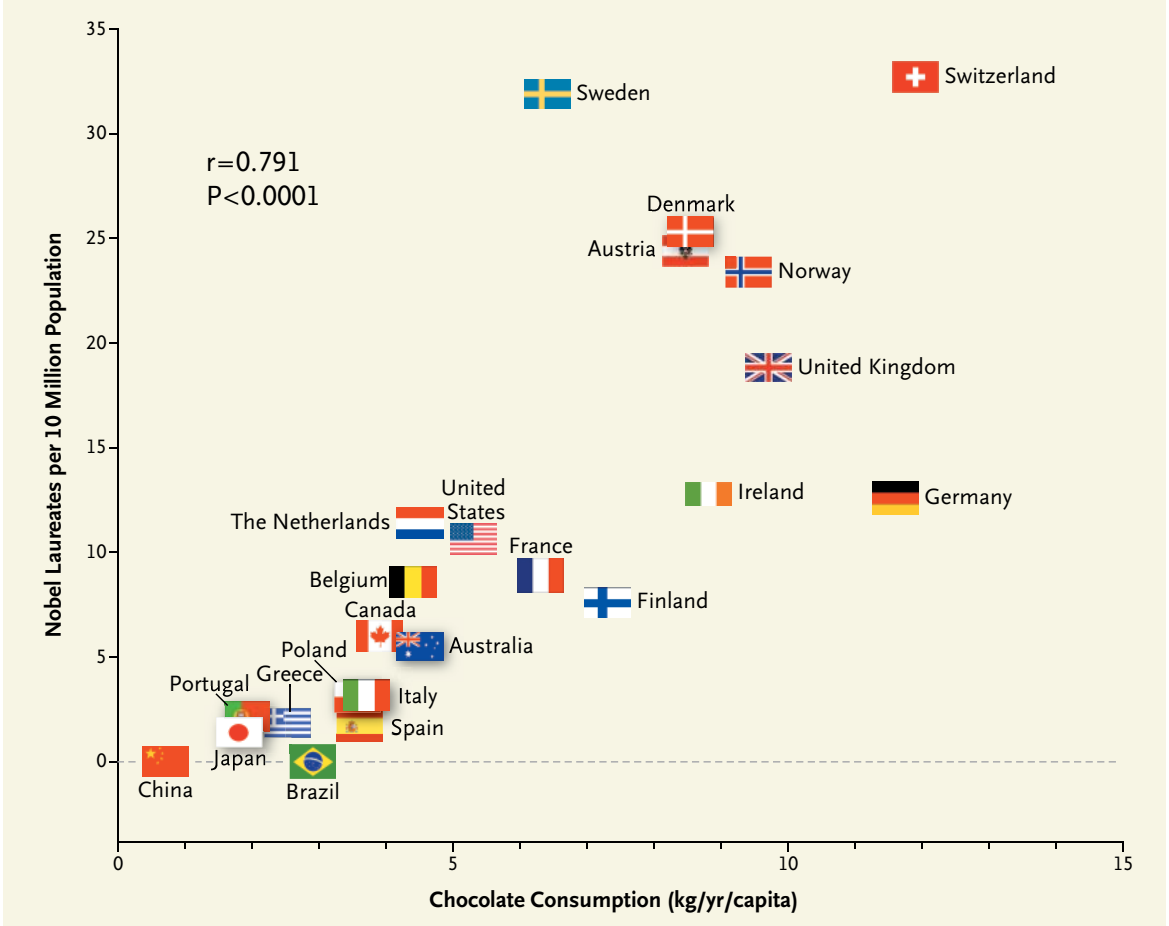


Figure 1. **Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.**

Source: Messerli, "Chocolate Consumption, Cognitive Function, and Nobel Laureates", *New England Journal of Medicine*, 2012.

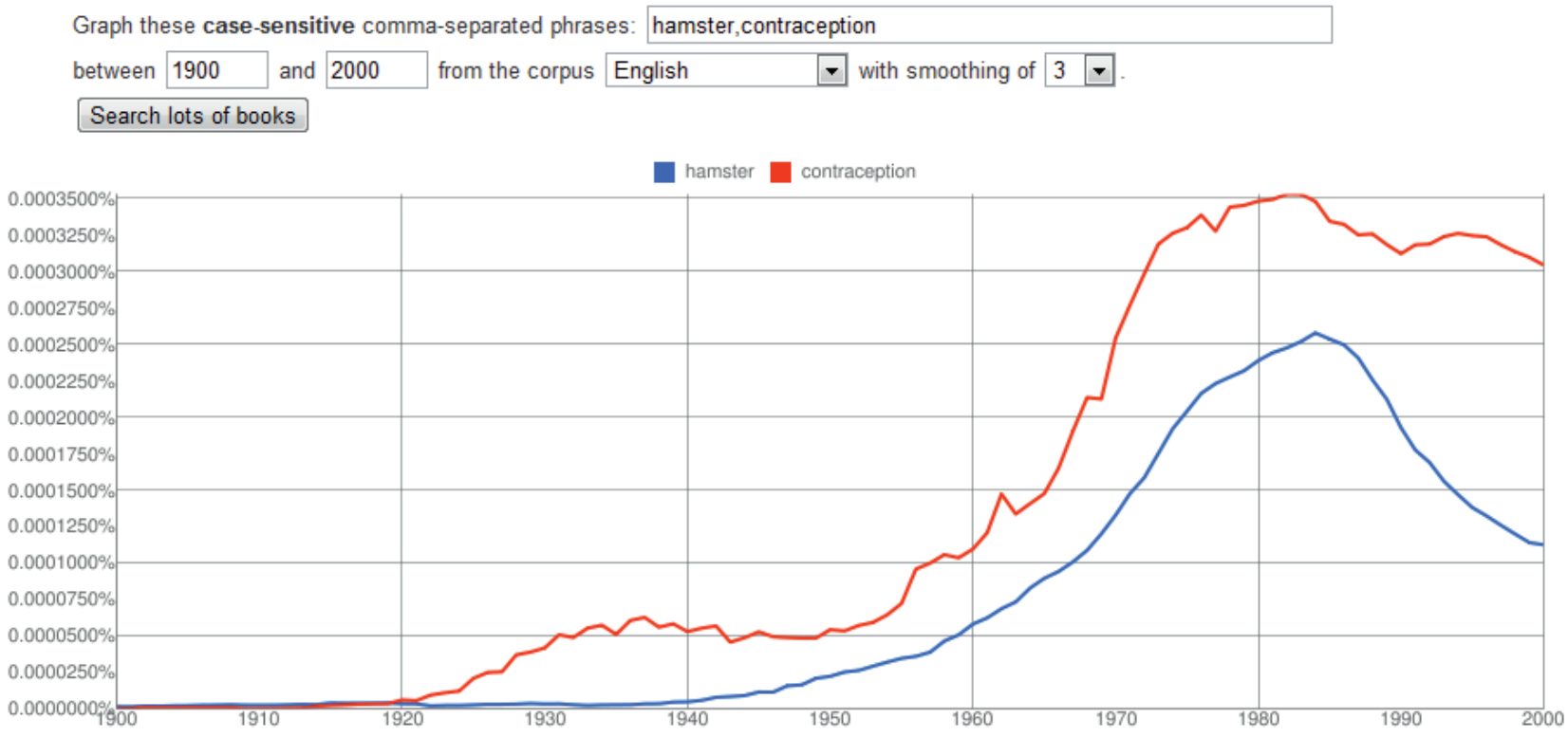


Figure 1: Frequencies of the words “hamster” and “contraception” in Google Books, 1900–2000

Source: Harkness, “Seduced by Stats?”, *Significance*, 2012.

2. Correlation matrixes

`pwcorr [varlist], [obs sig]`

- `obs` shows the number of observations
- `sig` shows the coefficient's *p*-value

`gr mat [varlist], [half etc.]`

- `half` plots only half of all graphs (quicker)
- accepts scatterplot options (`jitter`, `mlab`, etc.)

Correlation matrixes

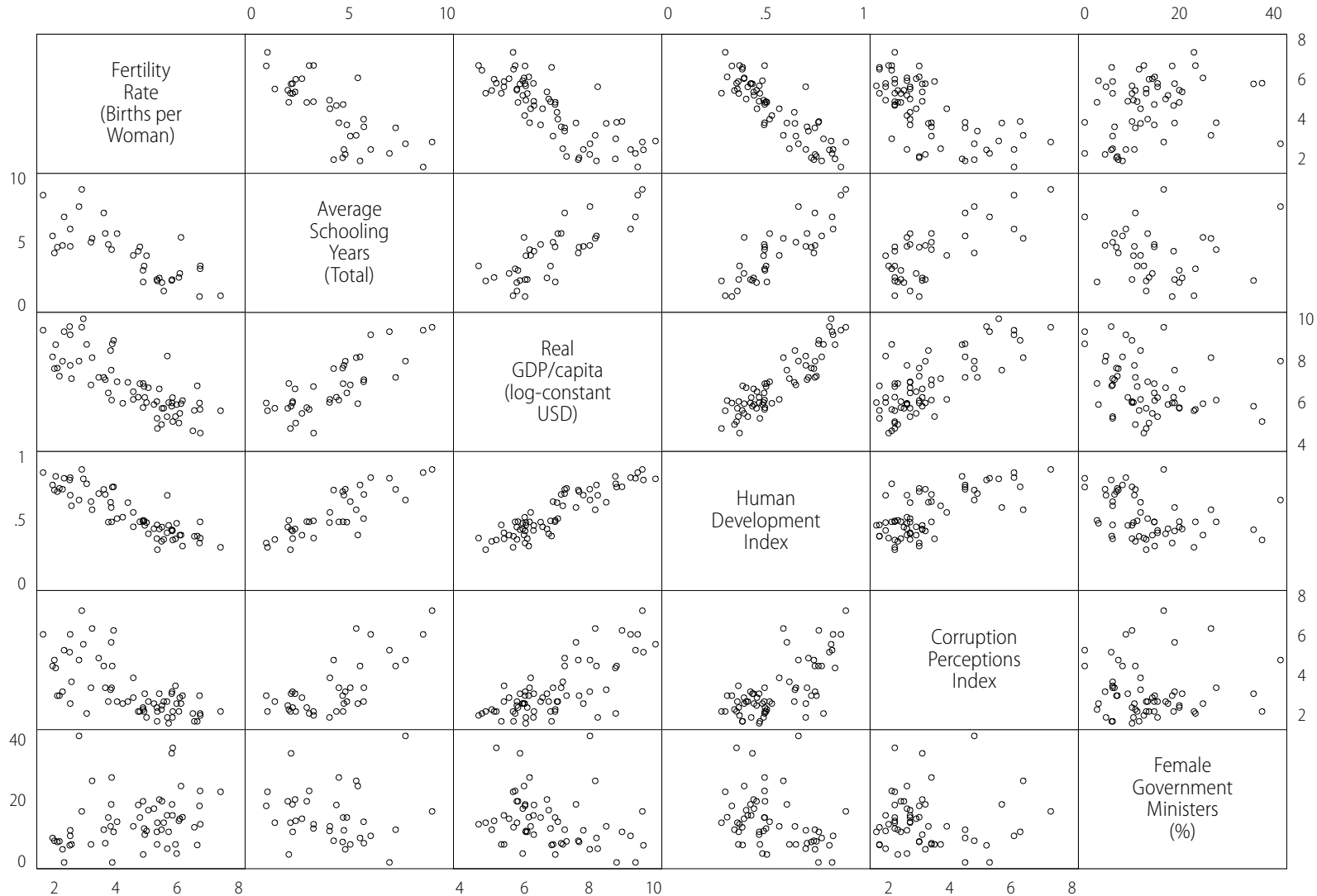
`mkcorr [varlist], lab num sig log(file.txt) replace`

- `ssc install mkcorr` to install
- `help mkcorr` to understand the options

Computer skills

- Import as a table in a spreadsheet editor.
- Convert from text to table in a rich text editor.

gr mat



Showing only Africa and the Middle East ($N = 68$).

From Stata output...

```
. pwcorr wdi_hiv wdi_hec wdi_prhe wdi_puhegdp, obs sig star(.05)
```

	wdi_hiv	wdi_hec	wdi_prhe	wdi_pu~p
wdi_hiv	1.0000			
	141			
wdi_hec	−0.1953*	1.0000		
	0.0207			
	140	187		
wdi_prhe	0.0979	−0.0555	1.0000	
	0.2497	0.4509		
	140	187	188	
wdi_puhegdp	−0.0607	0.5490*	−0.2099*	1.0000
	0.4759	0.0000	0.0038	
	140	187	188	188

$r = -.2$

$p < .02$

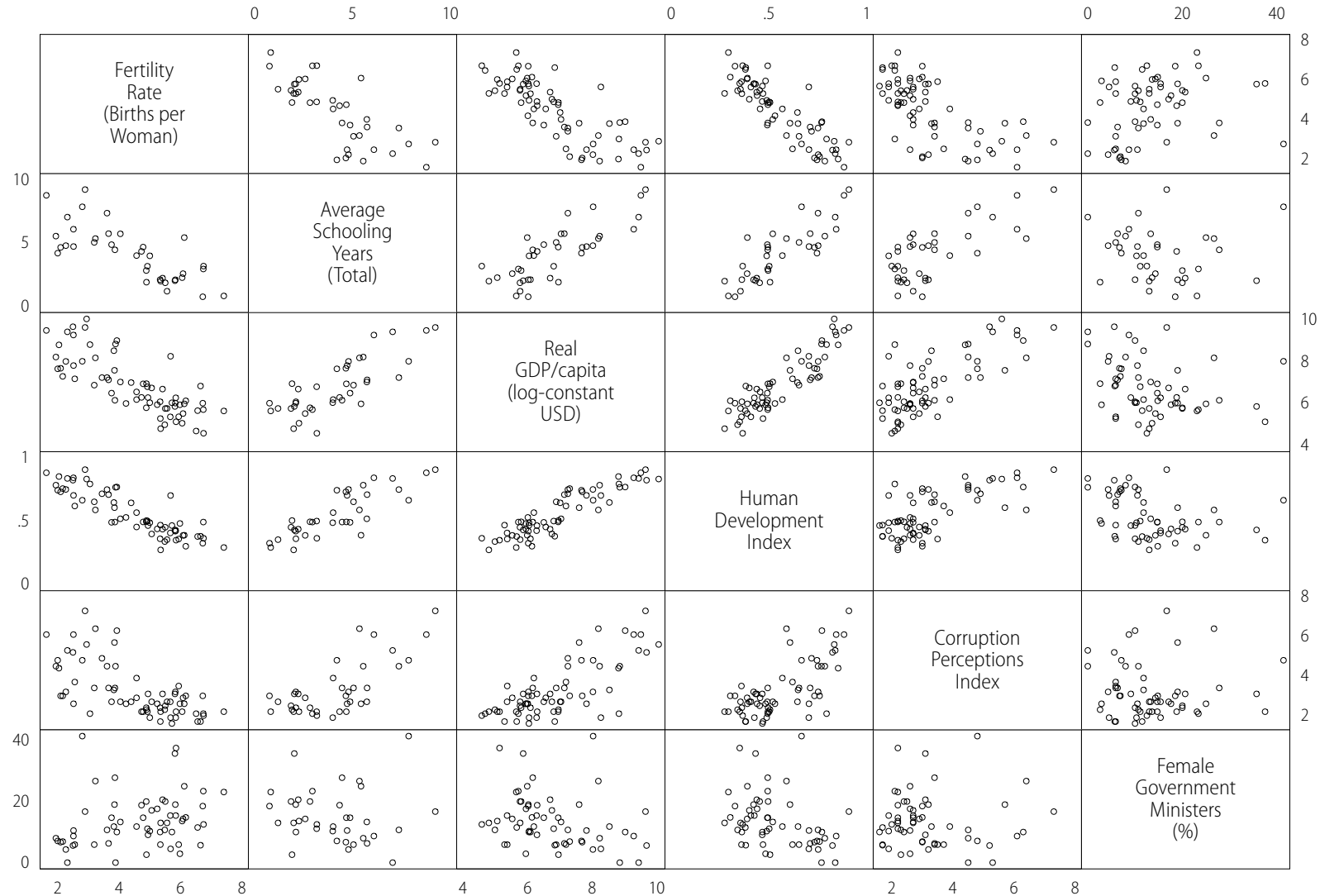
$N = 140$

coefficient

p-value

observations

… to publishing standard



Showing only Africa and the Middle East ($N = 68$).

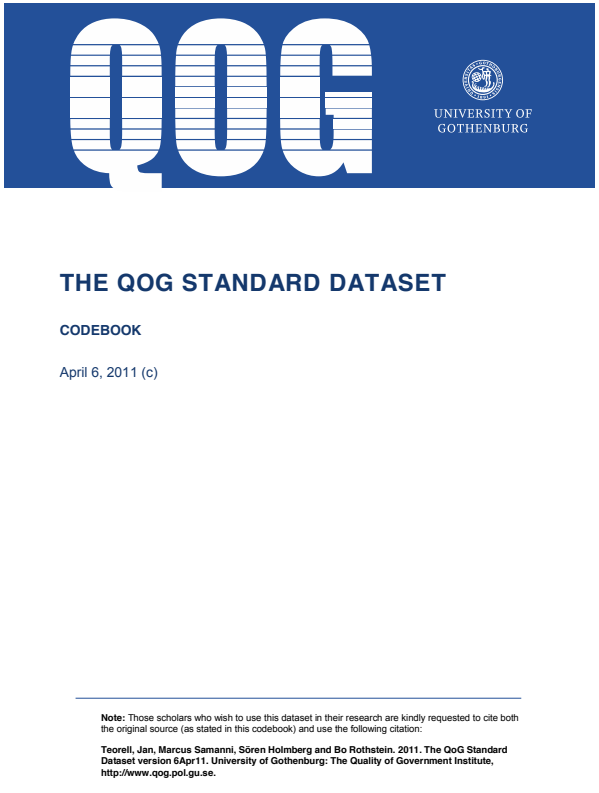
Practice: QOG dataset

Data:

- Quality of Government (QOG)
- Sample: countries, c. 2002

Variables:

- Fertility rate
- Education years
- Corruption Perceptions Index
- Human Development Index
- Female ministers



Practice session

Class

```
* Get the do-file for this week.
srqm fetch week7.do
* Open to read and replicate.
doedit code/week7
```

Coursework

- Finish the do-file and read all comments at home.
- Correct your do-file and add significance tests.
- Correct your paper and substantiate its hypotheses.

Exercise

Ex 7.1. Quality of Government 2011

- Variables: `d wdi_brd wdi_mege wdi_pb2 wdi_the`
- Inspect and plot the correlation matrix.

Ex 7.2. Quality of Government 2011

- Variables: `d wdi_puhegdp wdi_the wdi_prhe`
- Visualize and export the correlations and scatterplots.