

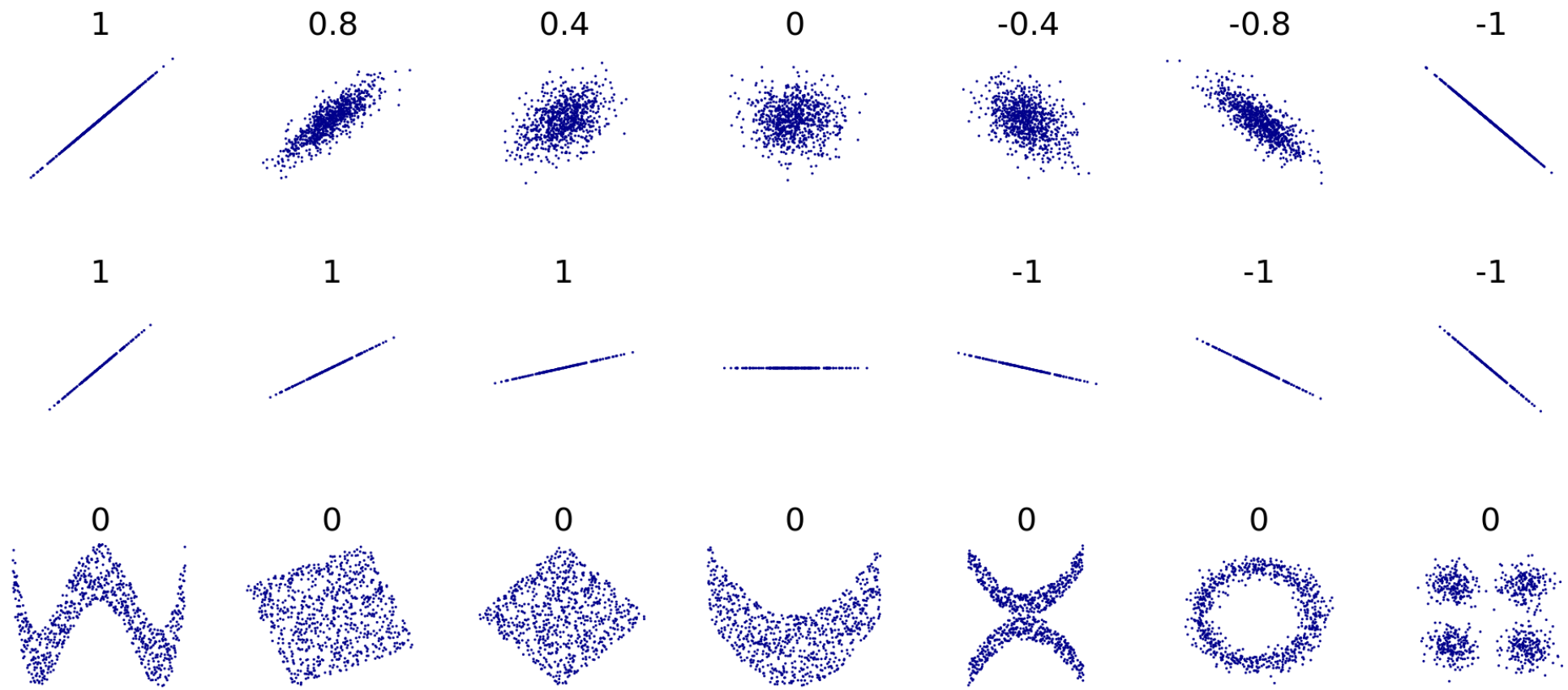
Korelace

# Kde jsme

- Souvislost kategorických proměnných: crosstab
- Souvislost kategorické a kardinální proměnné: srovnání průměrů
- Souvislost kardinálních proměnných: korelační koeficient

# Korelace

- „součinitel souvztažnosti“
- Jak se pozná souvislost:
  - S růstem hodnot jedné proměnné narůstají nebo klesají hodnoty druhé proměnné
- Lineární vztah: nárůst/pokles musí být pro nízké i vysoké hodnoty proměnné stejný



# Typy korelačních koeficientů

- pearson
- Spearman
- kendall

# Pearsonův korelační koeficient

- Jeden ze základních korelačních koeficientů
- Značení - R
- Hodnoty koeficientu:
  - Rozsah od -1 po 1
  - +1 = perfektní kladná souvislost – s růstem jedné proměnné roste druhá
  - -1 = perfektní záporná souvislost – s růstem jedné proměnné klesá druhá
  - 0 = žádná souvislost
- Čím více je hodnota vzdálena od nuly, tím je souvislost silnější
- Existují „tabulky“ k hodnocení síly vztahu

# Kovariance

- Základní mechanismus výpočtu pearsonova korelačního koeficientu
- „sdílený rozptyl“

$$= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N - 1}$$

- Ke smysluplné interpretaci potřebuje standardizaci -> pearson
  - Hodnota kovariance se vydělí součinem obou směrodatných odchylek

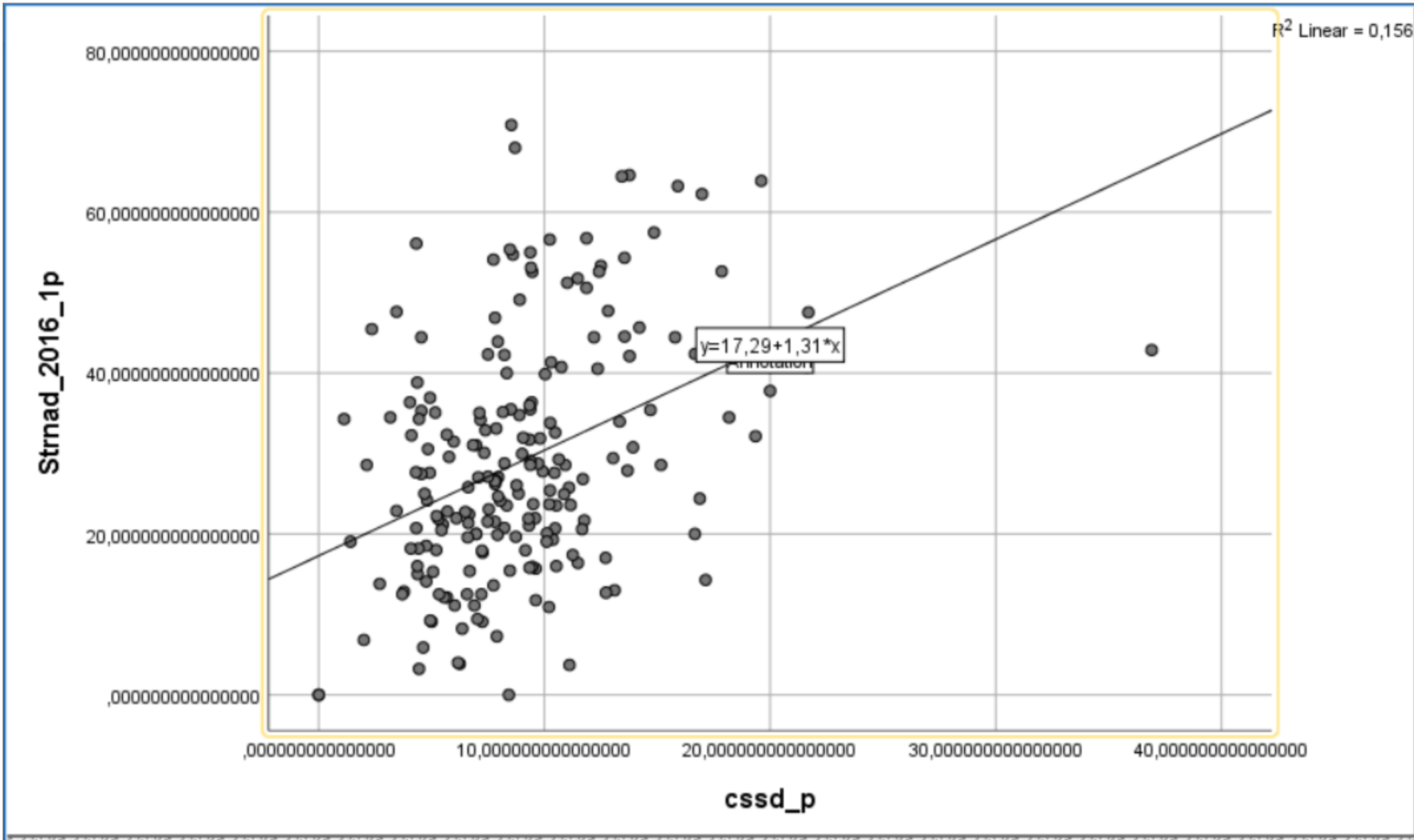
# Pearsonův korelační koeficient

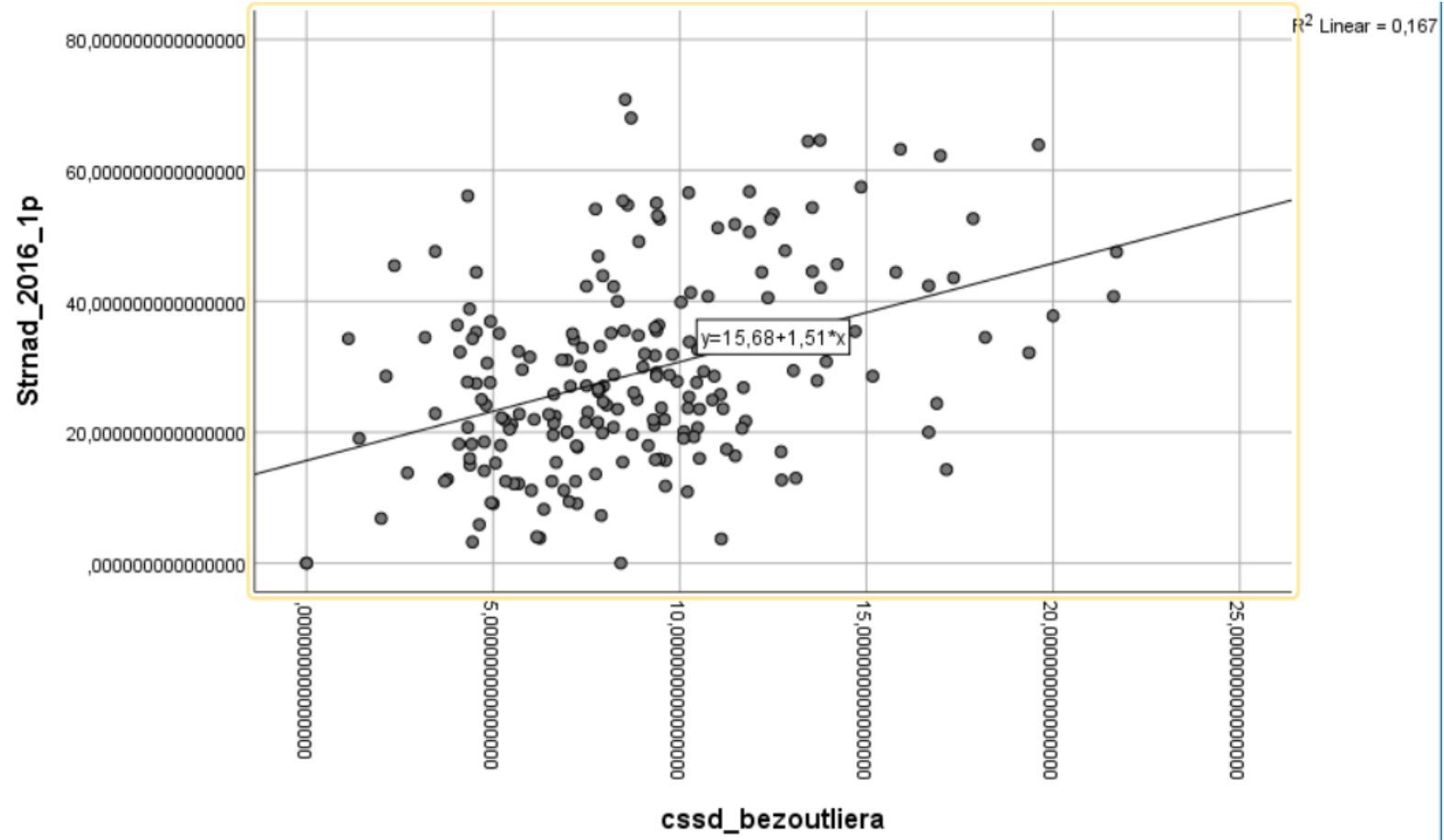
- Předpoklady:
  - Kardinální data (jedna z proměnných může být dichotomická)
  - normální rozložení / dostatečná velikost vzorku (min. 200-500)
- Citlivost na odlehlé případy



# Vsuvka - Graf

- Pokud proměnné nabývají mnoho různých málo se opakujících se hodnot
- Nebo pokud je vzorek malý
- => možnost zobrazit souvislost graficky
- Graphs – legacy dialogs – scatter/dot – simple scatter





# Práce v SPSS

- Analyze → Correlate → Bivariate:
  - Zvolit proměnné
  - Pearsonův koeficient je přednastavený
  - Pro sledování signifikance zvolit *Flag significant correlations*
- Options:
  - Možnost spočítat základní statistiky a kovarianci
  - Vynechání hodnot / případů
    - Listwise – pokud počítáme více korelačních koeficientů, všechny budou založeny na stejných datech
    - pairwise – missing odstraněny zvlášť v každém páru

# Pearsonův korelační koeficient – příklad 1

- Jak spolu souvisí podpora senátního kandidáta v obcích s podporou jeho strany?
- Potřebujeme zohlednit signifikanci?
- Co potřebujeme udělat před spočítáním koeficientu?

# Výpočet koeficientu v SPSS

- Data senat\_jicin\_22
- Analyze → Correlate → Bivariate:
- Všechny proměnné
- Options – missing values – exclude cases listwise
- Flag significant correlations nechat prázdné

### Correlations<sup>a</sup>

		cze22_1p	ded22_1p	cze22_2p	ded22_2p	ANO_PS21p	SPOLU_PS21p
cze22_1p	Pearson Correlation	1	-,493	,696	-,696	-,340	,520
	Sig. (2-tailed)		,000	,000	,000	,000	,000
ded22_1p	Pearson Correlation	-,493	1	-,661	,661	,602	-,408
	Sig. (2-tailed)	,000		,000	,000	,000	,000
cze22_2p	Pearson Correlation	,696	-,661	1	-1,000	-,537	,524
	Sig. (2-tailed)	,000	,000		,000	,000	,000
ded22_2p	Pearson Correlation	-,696	,661	-1,000	1	,537	-,524
	Sig. (2-tailed)	,000	,000	,000		,000	,000
ANO_PS21p	Pearson Correlation	-,340	,602	-,537	,537	1	-,502
	Sig. (2-tailed)	,000	,000	,000	,000		,000
SPOLU_PS21p	Pearson Correlation	,520	-,408	,524	-,524	-,502	1
	Sig. (2-tailed)	,000	,000	,000	,000	,000	

a. Listwise N=175

# Příklad 2 –souvislost mezi hodnocením situace a hodnocením vlády

- Jak spolu souvisí hodnocení vlády a hodnocení ekonomické a politické situace?
- Využitá data: vzorek populace
- Potřebujeme zohlednit signifikanci?
  
- Co potřebujeme udělat před spočítáním koeficientu?



# Data ees9cz

## Correlations<sup>b</sup>

		How satisfied with present state of economy in country	How satisfied with the national government	How satisfied with the way democracy works in country
How satisfied with present state of economy in country	Pearson Correlation	1	,551**	,475**
	Sig. (2-tailed)		,000	,000
How satisfied with the national government	Pearson Correlation	,551**	1	,494**
	Sig. (2-tailed)	,000		,000
How satisfied with the way democracy works in country	Pearson Correlation	,475**	,494**	1
	Sig. (2-tailed)	,000	,000	

\*\* . Correlation is significant at the 0.01 level (2-tailed).

b. Listwise N=2293

# Pearsonův korelační koeficient

- Se zjištěným  $R$  je možné dál pracovat
- Po umocnění získáváme tzv. Index determinace ( $R^2$ )
- $R^2 \times 100$  vymezuje, jaký podíl variability jedné proměnné je sdílený s druhou proměnnou

# Pearsonův korelační koeficient

- Výjimka z kardinálních dat → korelace jedné kardinální proměnné a jedné dichotomické
- Tzv. point-biserial korelace
- Kladné / záporné výsledné hodnoty závisí na kódování dichotomické proměnné
- Mnohem lepší je ale v takovém případě použít srovnání průměrů!

# Korelace pohlaví x příjem

		Correlations <sup>b</sup>	
		Your usual [weekly/monthly/annual] net [pay/pensions /social benefits]	zena
Your usual [weekly/monthly/annual] net [pay/pensions/social benefits]	Pearson Correlation	1	-,214**
	Sig. (2-tailed)		,000
zena	Pearson Correlation	-,214**	1
	Sig. (2-tailed)	,000	

\*\* . Correlation is significant at the 0.01 level (2-tailed).

b. Listwise N=1763

# Spearmanovo rho a Kendallovo tau

- Oboje: neparametrický postup
- RHO:
  - Využíván zejména pro kombinaci ordinálních proměnných
  - V menších vzorcích (do 200-500) při porušení normality
  - pro výpočet využívá pořadí případů, nikoli samotné hodnoty proměnné
  - Výsledné hodnoty jsou ve stejném pásmu jako u PKK (od -1 po 1)
- TAU: (viz příští slide)

# Kendalovo tau

- Neparametrický postup
- Pro malé vzorky
- Menší množství kategorií
  - Volba mezi kendallem a crosstabem
- Některé hodnoty se velice často opakují
  - Pro malé vzorky
  - Menší množství kategorií
    - Volba mezi kendallem a crosstabem
  - Některé hodnoty se velice často opakují

# SPSS

- Analyze → Correlate → Bivariate:
  - Zvolit proměnné
  - Vybrat *Spearman a/nebo Kendall*

### Correlations<sup>b</sup>

			How happy are you	How often socially meet with friends, relatives or colleagues	How many people with whom you can discuss intimate and personal matters
Kendall's tau_b	How happy are you	Correlation Coefficient	1,000	,154**	,072**
		Sig. (2-tailed)	.	,000	,000
	How often socially meet with friends, relatives or colleagues	Correlation Coefficient	,154**	1,000	,245**
		Sig. (2-tailed)	,000	.	,000
	How many people with whom you can discuss intimate and personal matters	Correlation Coefficient	,072**	,245**	1,000
		Sig. (2-tailed)	,000	,000	.
Spearman's rho	How happy are you	Correlation Coefficient	1,000	,193**	,090**
		Sig. (2-tailed)	.	,000	,000
	How often socially meet with friends, relatives or colleagues	Correlation Coefficient	,193**	1,000	,299**
		Sig. (2-tailed)	,000	.	,000
	How many people with whom you can discuss intimate and personal matters	Correlation Coefficient	,090**	,299**	1,000
		Sig. (2-tailed)	,000	,000	.

\*\* . Correlation is significant at the 0.01 level (2-tailed).

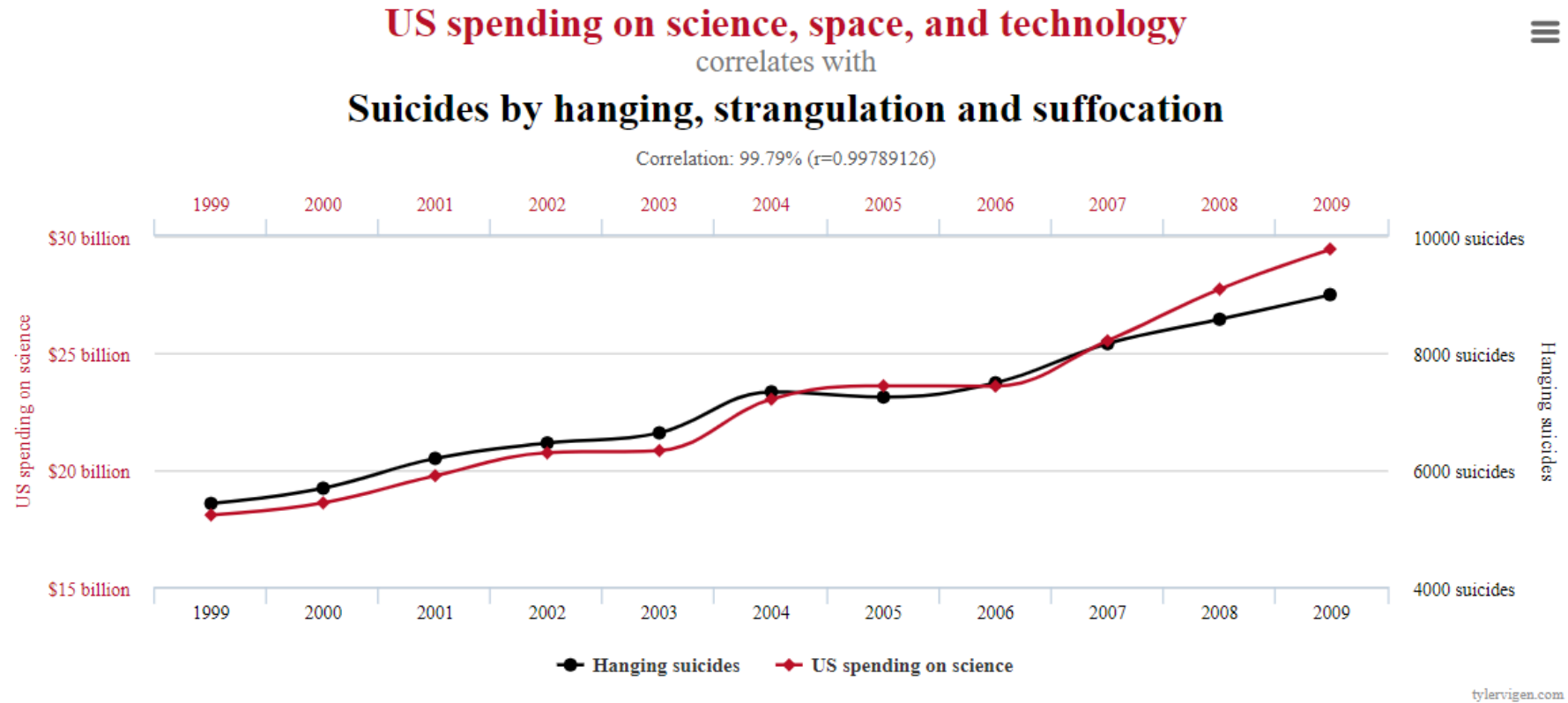
b. Listwise N = 2326



# Interpretace výsledků

- Základní pravidlo – **korelace  $\neq$  kauzalita**
- Korelace vyjadřuje pouze souvislost mezi proměnnými, neukazuje na žádnou příčinu a následek
- Vliv třetích proměnných
- Nemožnost konstatovat kauzalitu i pokud se jeví jako logická
- Statistické zjištění nemá automaticky věcný význam

# <https://www.tylervigen.com/spurious-correlations>



Data sources: U.S. Office of Management and Budget and Centers for Disease Control & Prevention

# Práce s koeficienty

- Když nevíte, který koef. spočítat, spočítejte všechny
  - Všechny vyjdou podobně (kendall má ale obecně nižší hodnoty)
- Co uvádět:
  - Jaký koeficient byl použit, kolik případů bylo v analýze
  - V tabulce: hodnoty koeficientu, hvězdičky a sig jen pokud je potřeba zohlednit signifikanci
  - K hvězdičkám je nutné dodat legendu