

# Introducción al análisis cuantitativo de datos lingüísticos

## Bloque 4.1: Correlaciones espúreas y no tanto

Ezequiel Koile (MPI-SSH)  
Carolina Gattai (IFIBA – CONICET)

# Correlación vs causación

Para cada conjunto de dos eventos correlacionados A y B, podemos estar en presencia de los siguientes escenarios:

- ▶ A causa B (causación directa)
- ▶ B causa A (causación inversa)
- ▶ A y B son consecuencia de una causa común, pero ninguna causa la otra
- ▶ A causa B y B causa A (causación bidireccional o cíclica)
- ▶ A causa C, que a su vez causa B (causación indirecta)
- ▶ No hay conexión causal entre A y B: la correlación es una coincidencia

# Correlación vs causación

- Cuanto más rápido giran los molinos, se observa más viento. Por lo tanto, el viento es causado por la rotación de los molinos.
- Edad media europea: los piojos son buenos para la salud, pues casi no se ven en las personas muertas.
- Cuanto más aumentan las ventas de helado, más muertes por ahogamiento se reportan. Por lo tanto, el consumo de helado produce ahogamiento.
- Los niños que duermen con la luz encendida son más proclives a desarrollar miopía en la adultez. Por lo tanto, dormir con la luz encendida causa miopía.
- Cuanto más predadores hay en un ecosistema, más presas habrá. Por lo tanto, los predadores atraen a sus presas.
- Ejemplo yanqui: El resultado del último partido de local de los Washington Redskins anterior a cada elección presidencial predijo el partido ganador en todas las elecciones entre 1936 y 2000.
- Ejemplo ruso: Durante los últimos 200 años, luego de cada mandatario ruso pelado asumió uno con pelo y viceversa.

# Correlación vs causación

Лысый (с плестью) правитель	Портрет	Волосатый правитель	Портрет
Николай I (1825—1855)		Александр II (1855—1881)	
Александр III (1881—1894)		Николай II (1894—1917)	
Георгий Львов (1917)		Александр Керенский (1917)	
Владимир Ленин (1917—1924)		Иосиф Сталин (1924—1953)	

	Никита Хрущёв (1953—1964)		Леонид Брежнев (1964—1982)
	Юрий Андрапов (1982—1984)		Константин Черненко (1984—1985)
	Михаил Горбачёв (1985—1991)		Борис Ельцин (1991—1999)
	Владимир Путин (2000—2008)		Дмитрий Медведев (2008—2012)
	Владимир Путин (с 2012)		

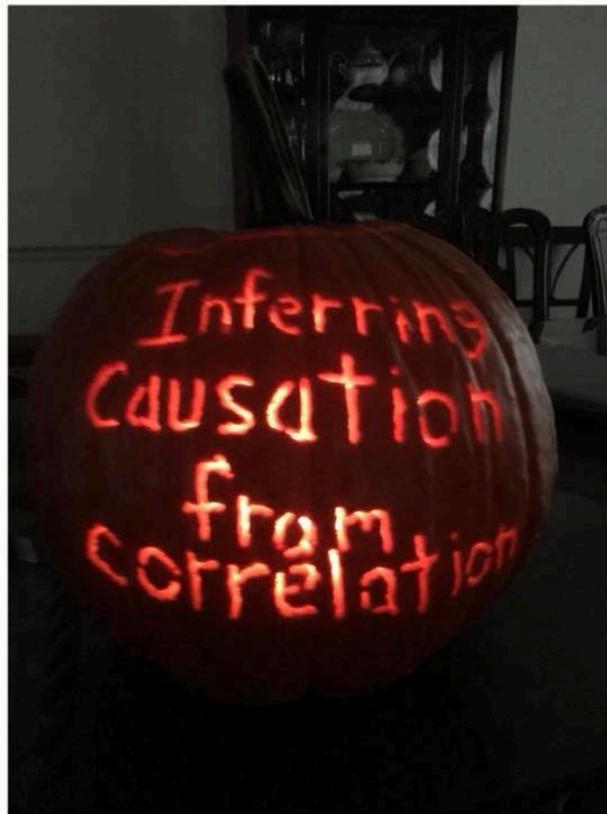
# Correlación vs causación



Dr. John Mohl  
@JohnCMohl

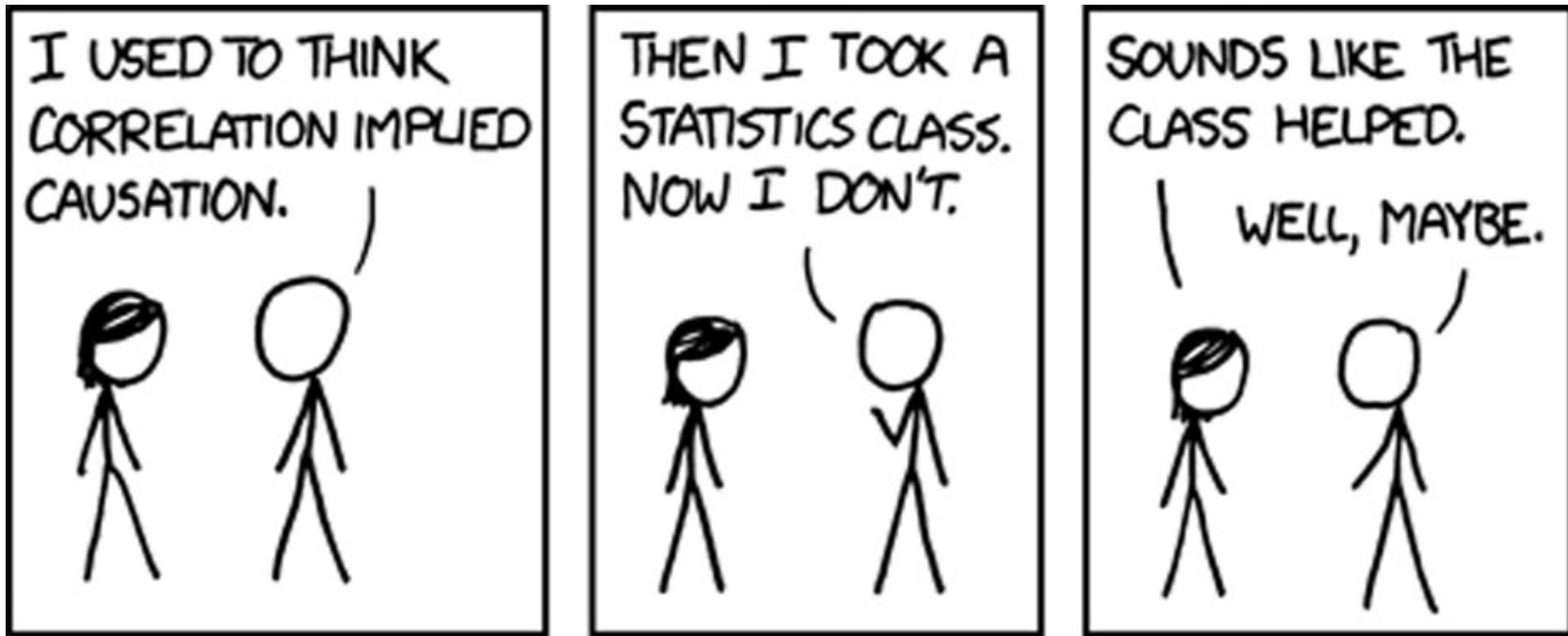


Nothing scares me more...



2017-10-29, 4:03 PM

# Correlación vs causación



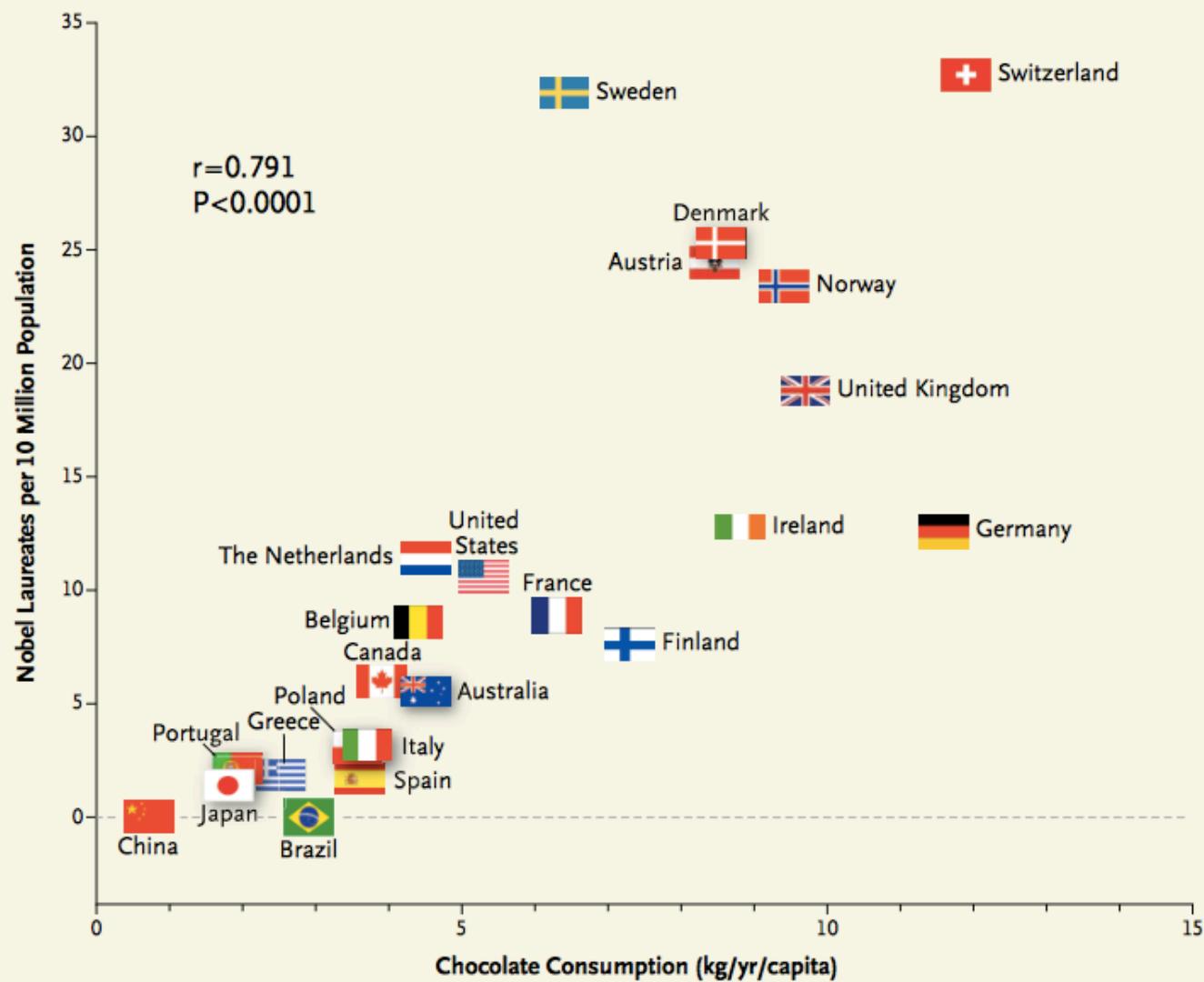
*“Correlation doesn’t imply causation, but it does waggle its eyebrows suggestively and gesture furtively while mouthing ‘look over there’”*

Sin embargo... <https://youtu.be/HUi6vGctQM>

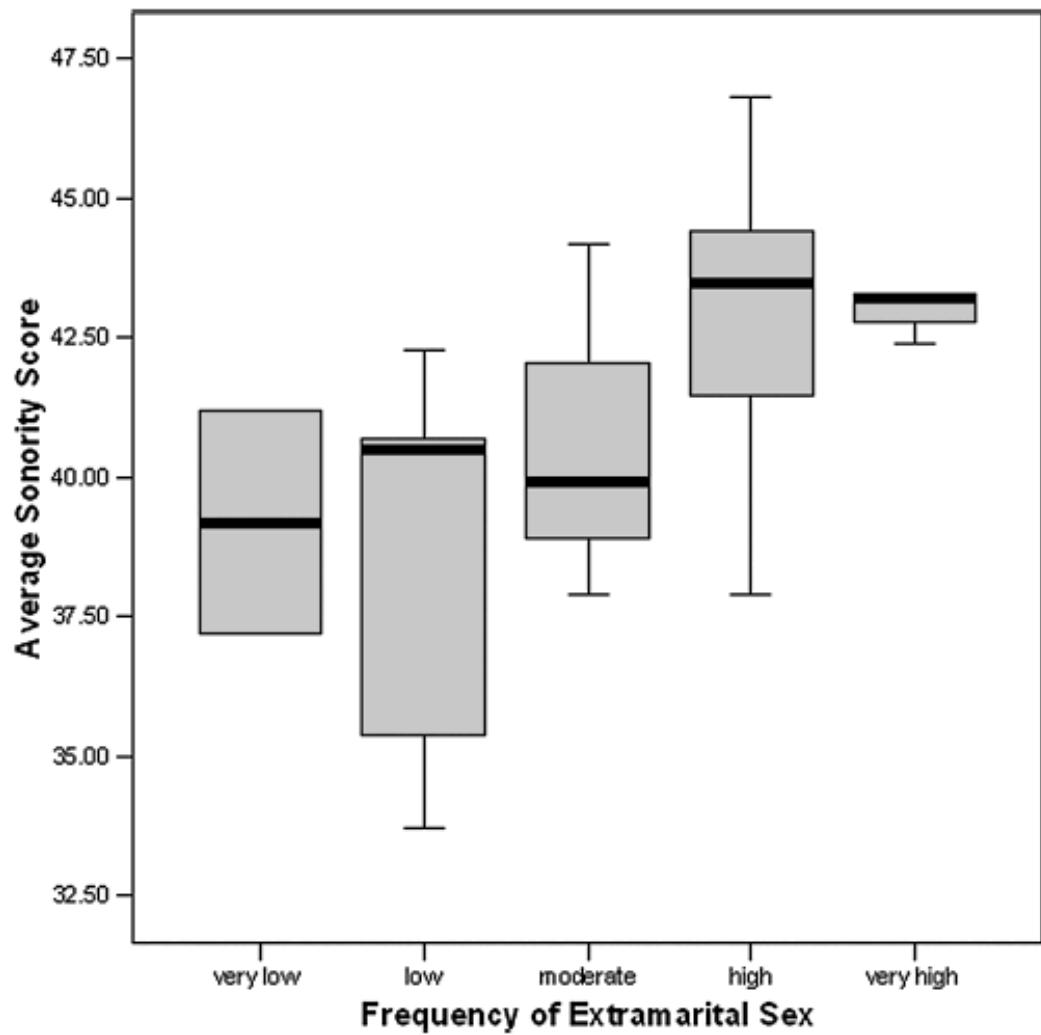
[xkcd.com]

# Correlación vs causación

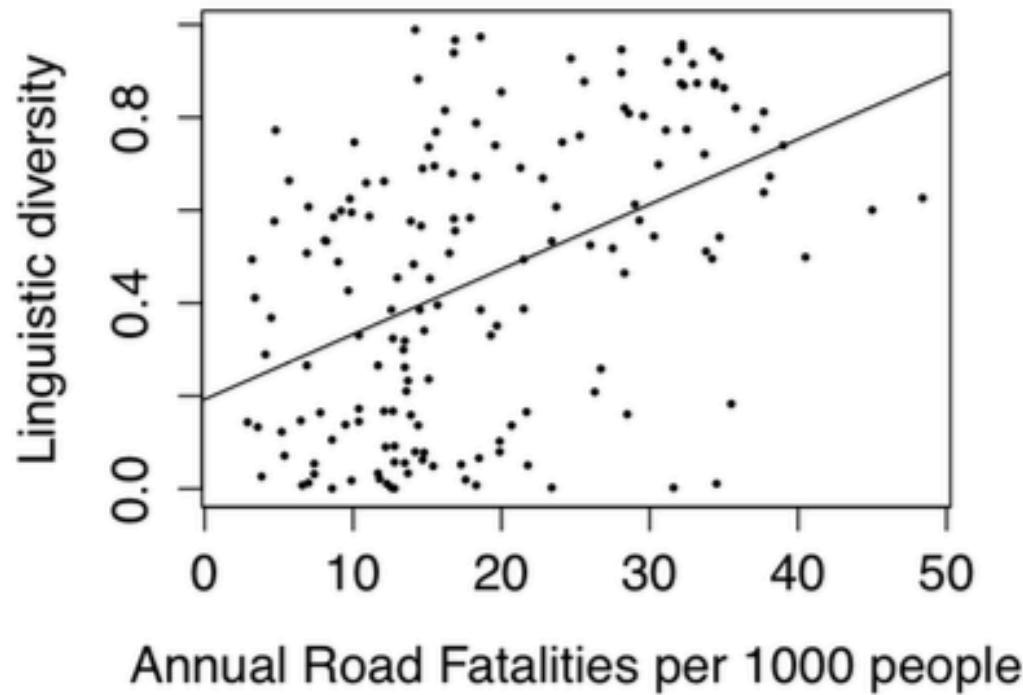
- ▶ Clasificar los enunciados siguientes en las categorías:
  - “Me lo creo”: el resultado tiene sentido, y también se me ocurre un mecanismo causal
  - “Interesante”: el resultado me llama la atención y me parece plausible; me da curiosidad estudiar los mecanismos causales
  - “Huele raro”: se me ocurre un mecanismo causal, pero me parece que hay un problema con en análisis de datos
  - “Fruta”: no me creo ni la correlación ni ningún mecanismo causal
- ▶ Justificar:
  - Cuál es el mecanismo (si lo hubiere)
  - Qué variables estamos “olvidando”



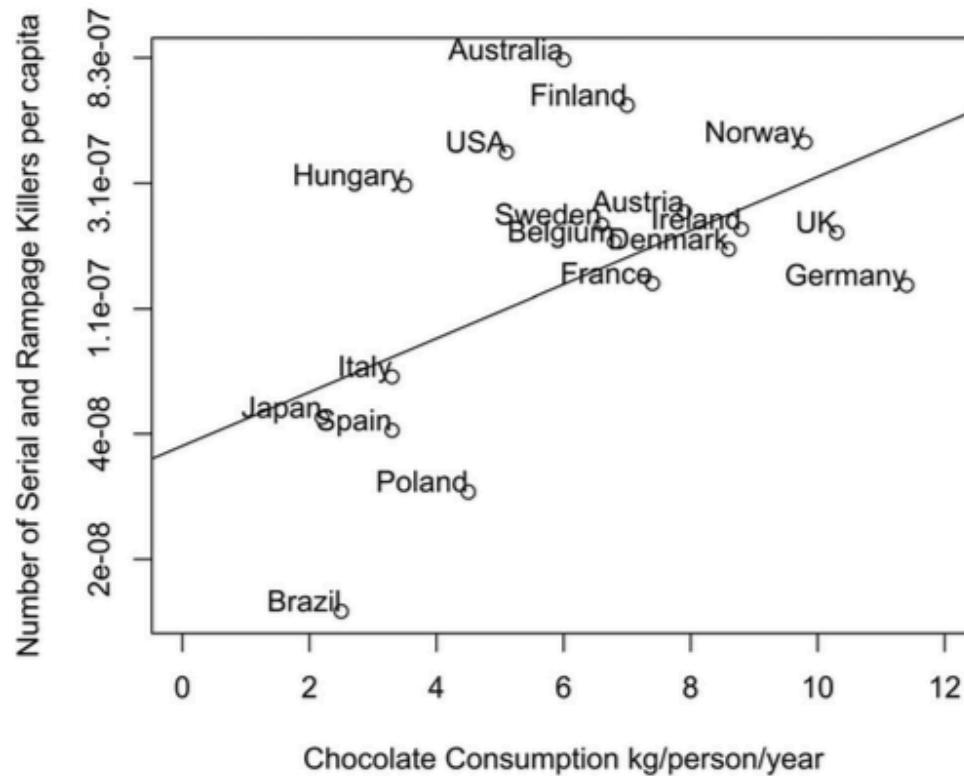
**Figure 1.** Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.



Controlling for  
Weather

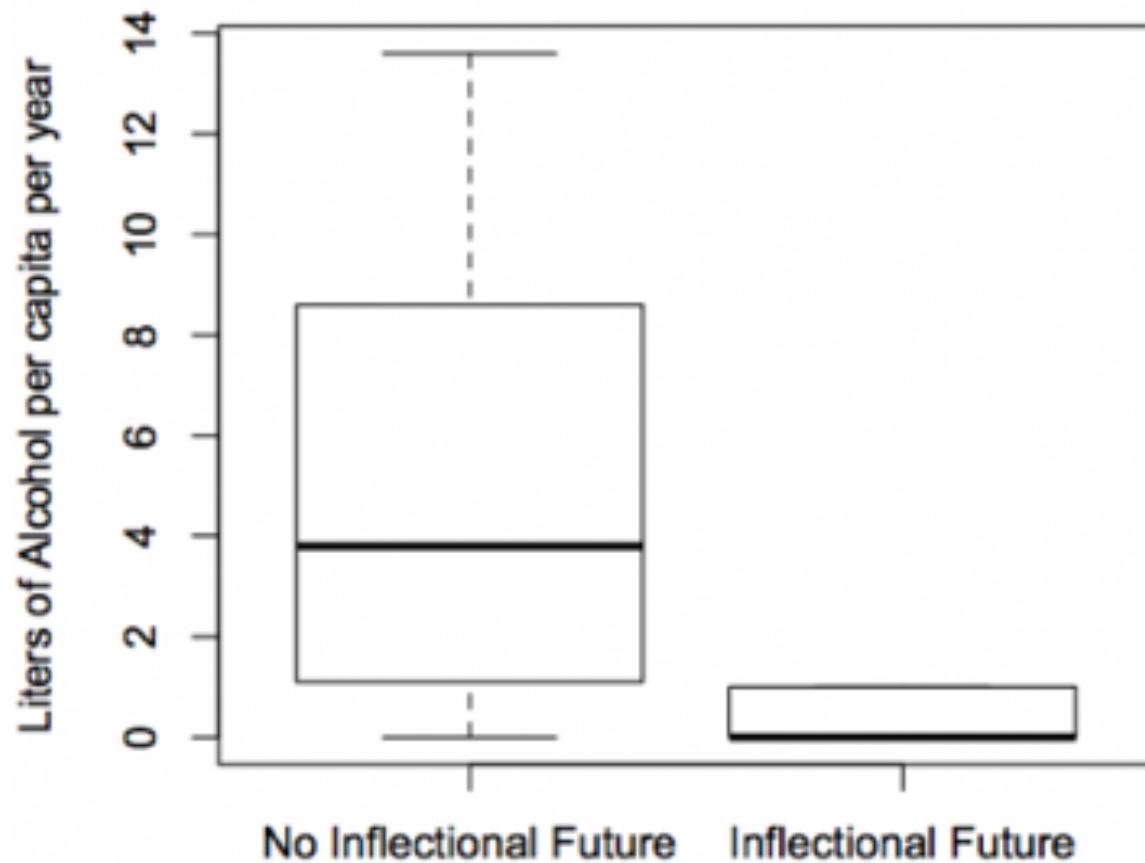


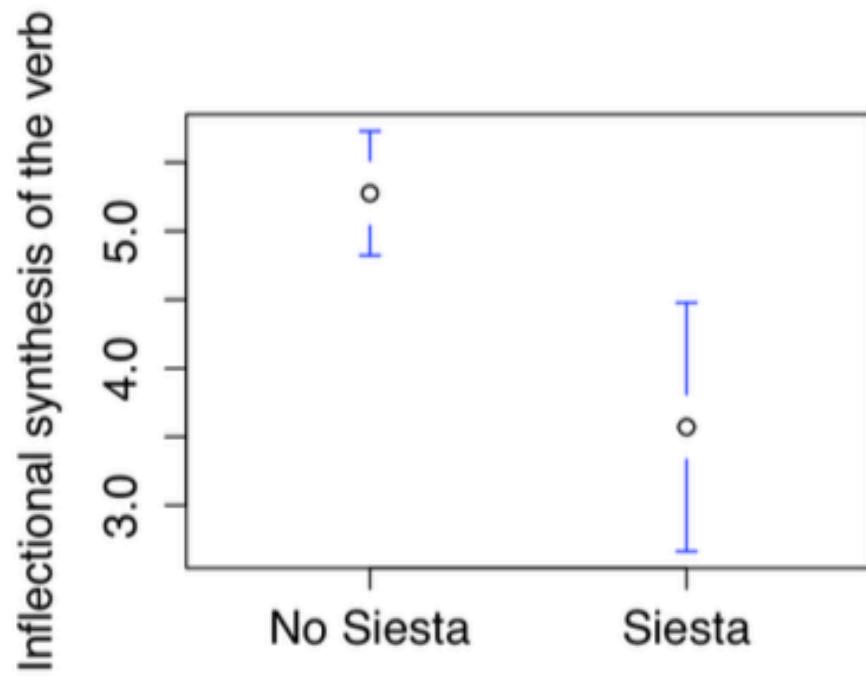
**Figure 2. Linguistic diversity and traffic accidents.** Countries' linguistic diversity (Greenberg diversity index) as a function of the annual traffic fatalities per 1000 people, with linear regression line.



**Figure 10. Chocolate consumption per capita (kg) as a function of the log number of serial and rampage killers per capita since 1900.**

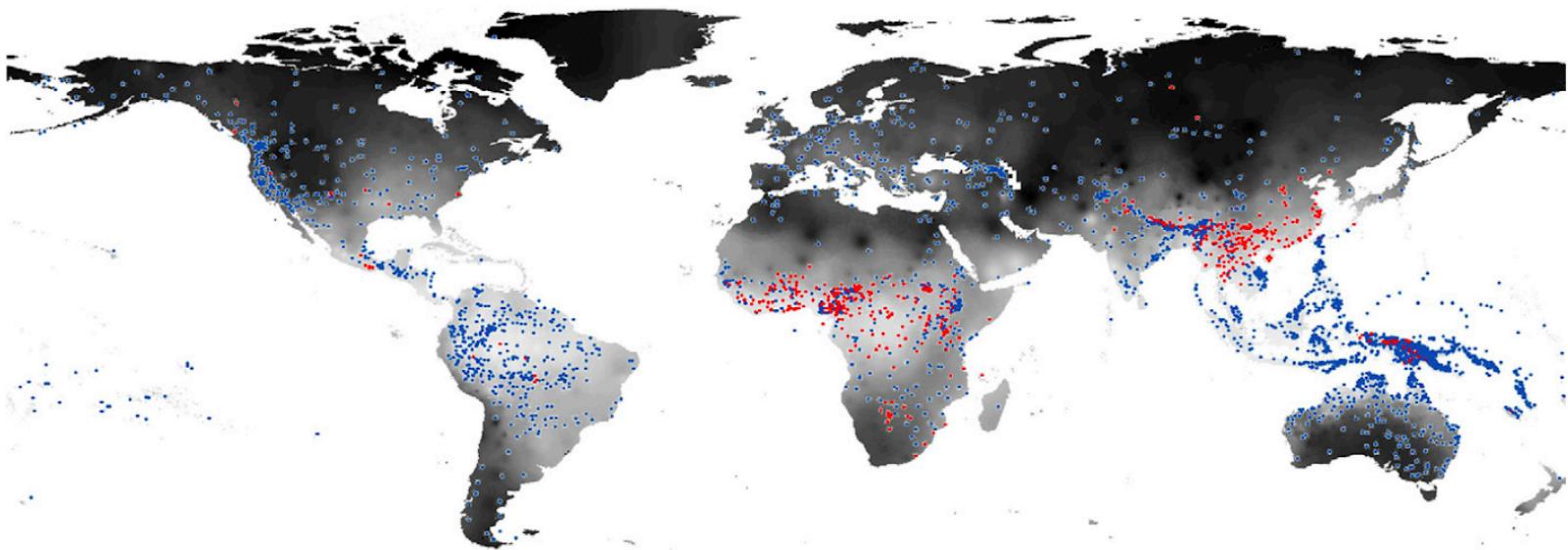
doi:10.1371/journal.pone.0070902.g010



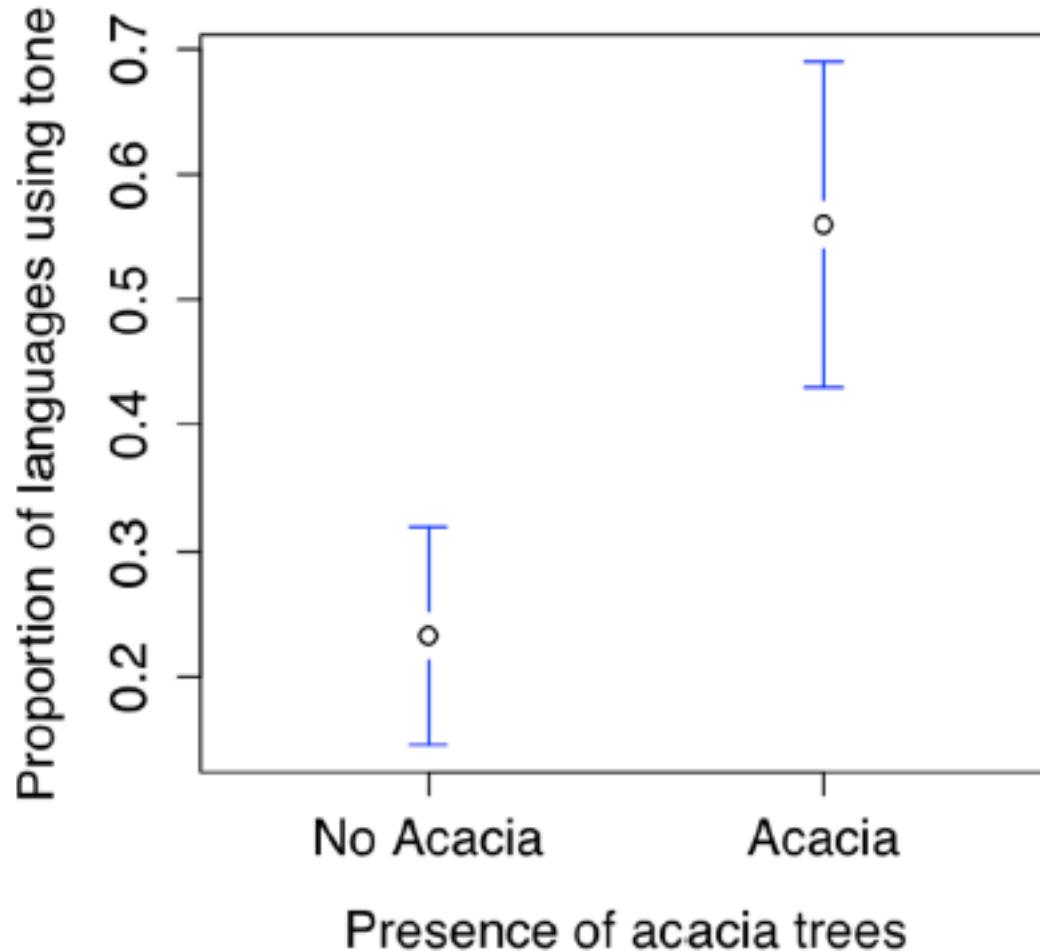


**Figure 5. Relationship between siestas and number of grammatical categories a verb can take.**

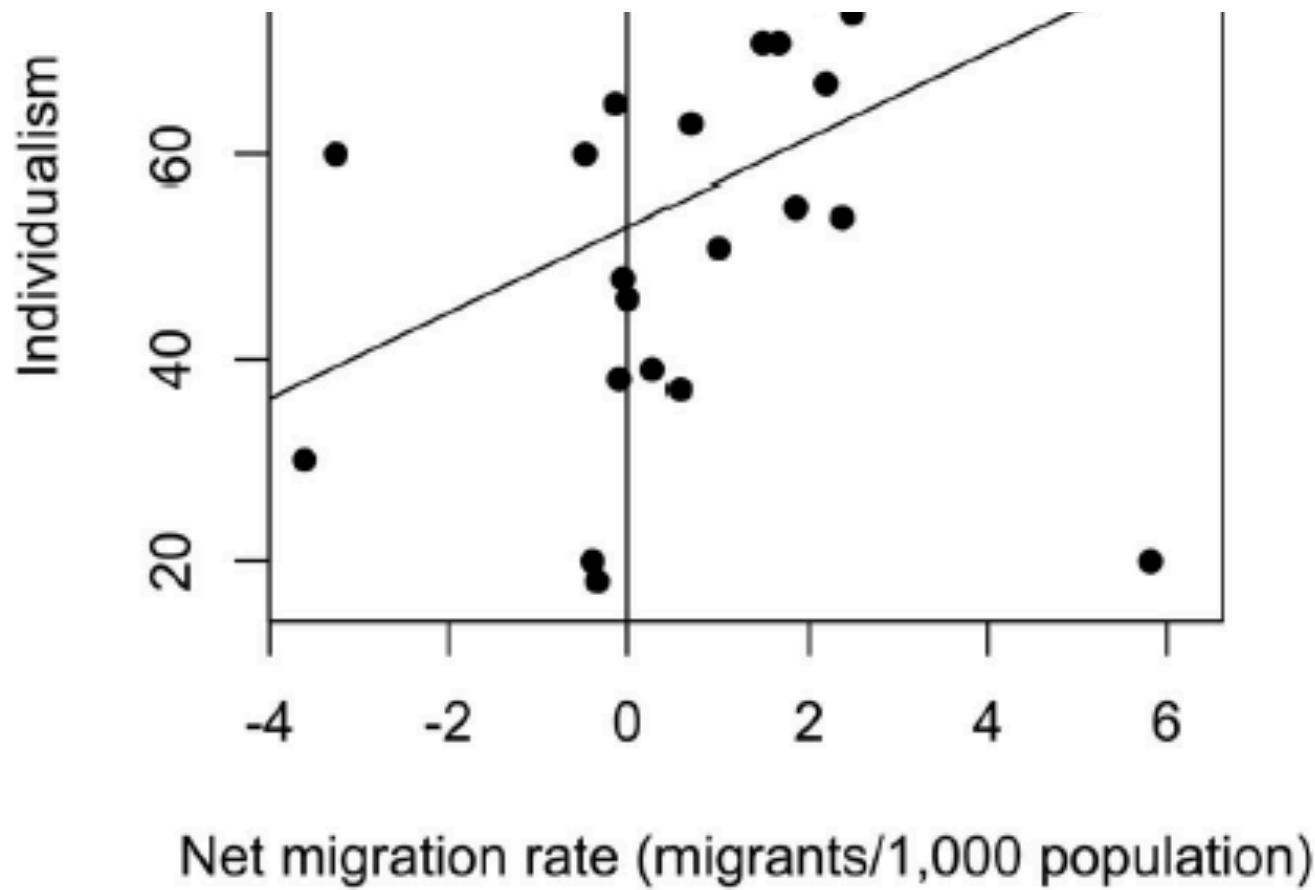
doi:10.1371/journal.pone.0070902.g005

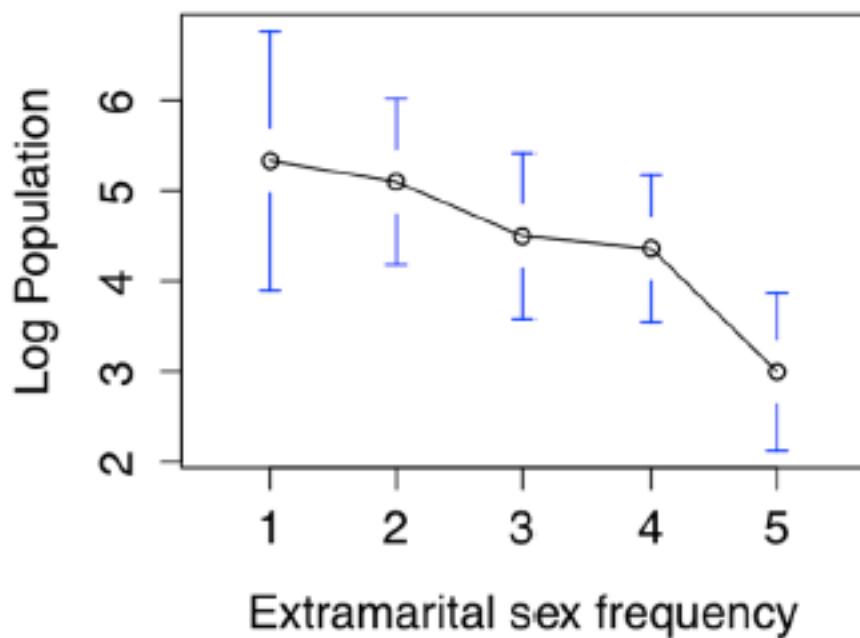


**Fig. 1.** Distribution of languages with complex tone (red dots) and without complex tone (blue dots) in the ANU database. Darker shading on map corresponds to lower MH.



**Figure 8. The proportion of tone languages within a country as a function of the presence of *Acacia nilotica*.**  
doi:10.1371/journal.pone.0070902.g008





**Figure 6. Relationship between population size and frequency of extramarital sex in a society.**  
doi:10.1371/journal.pone.0070902.g006