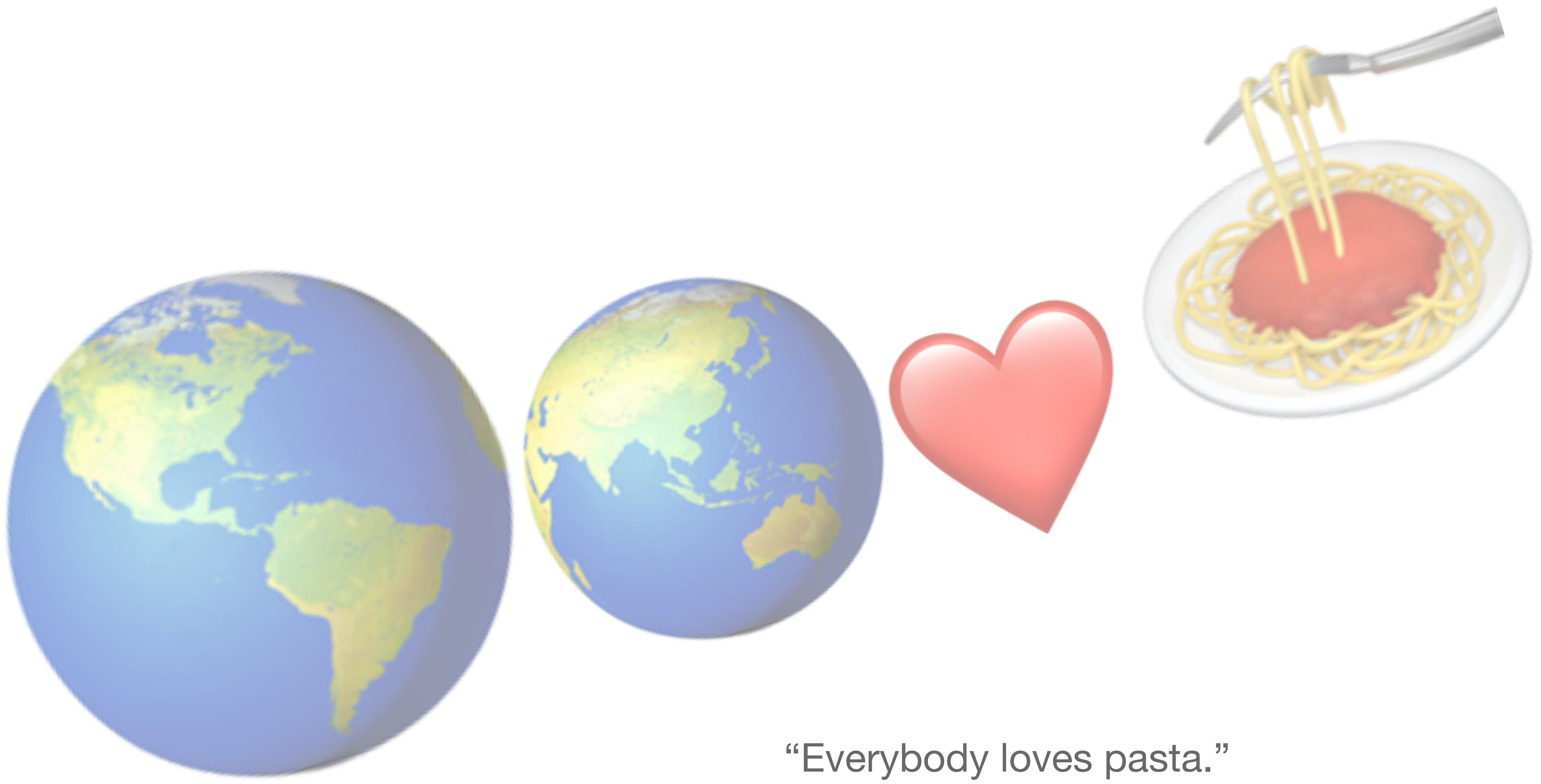
Studying Cultural Differences in Emoji Usage Across the East and the West



"Everybody loves pasta."

Mingyang Li
University of Pennsylvania

Outline

1. **Research Question**:

   Across western countries and eastern countries, what Emojis people associate with various topics? **(Dis)-similarities?**

2. **Method**: Calculate similarities in the vector space between each **Emoji** to each psycholinguistic **category** (LIWC)

3. **Results**

   1. Emoji-Category similarities reflect social & cultural associations

   2. Insights from cross-cultural correlations of similarities

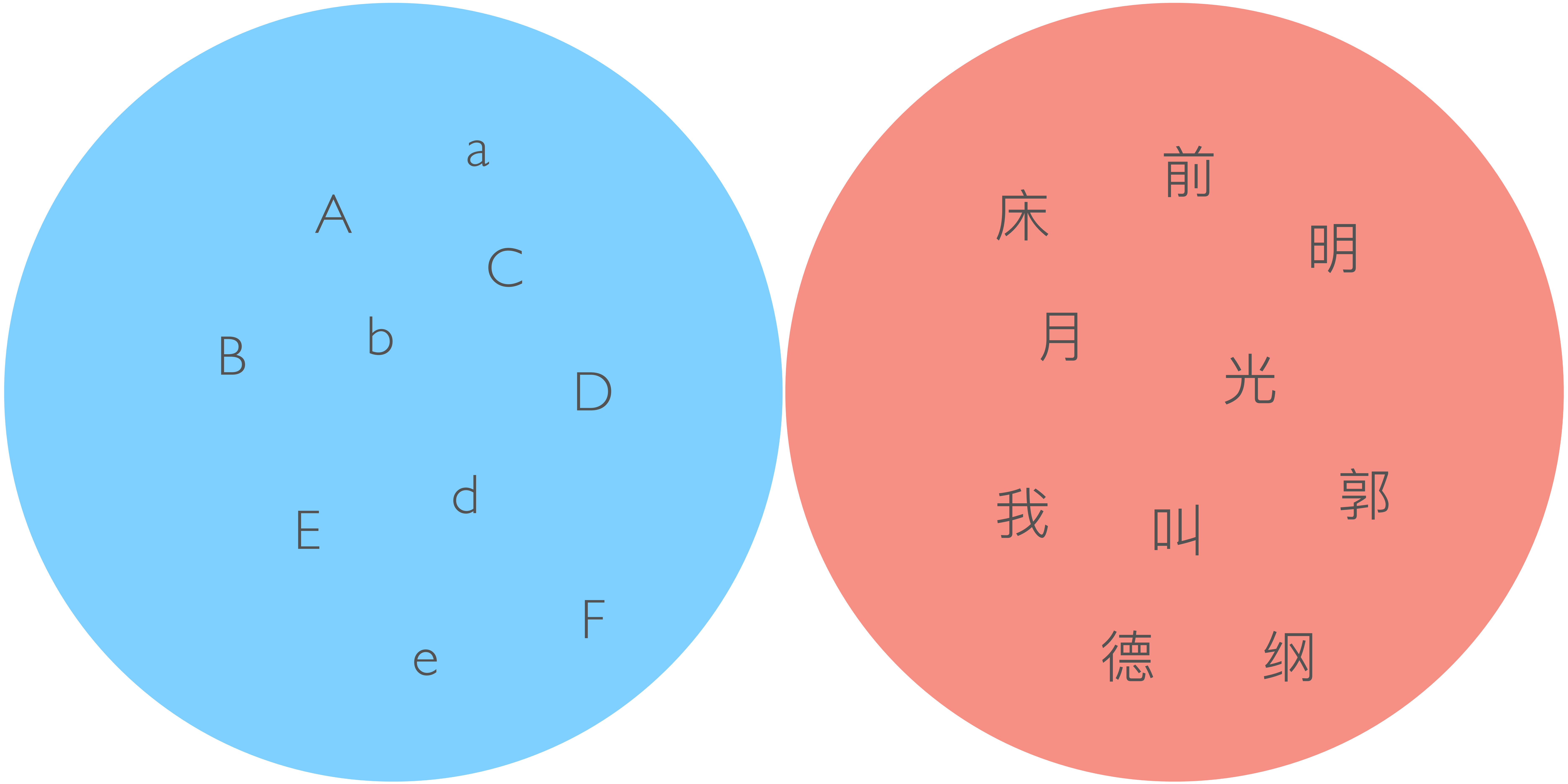
4. **Discussion**: Digging deeper into associations with *emotions*

Research Question

Across the East and the West, what ideograms people associate with various topics?

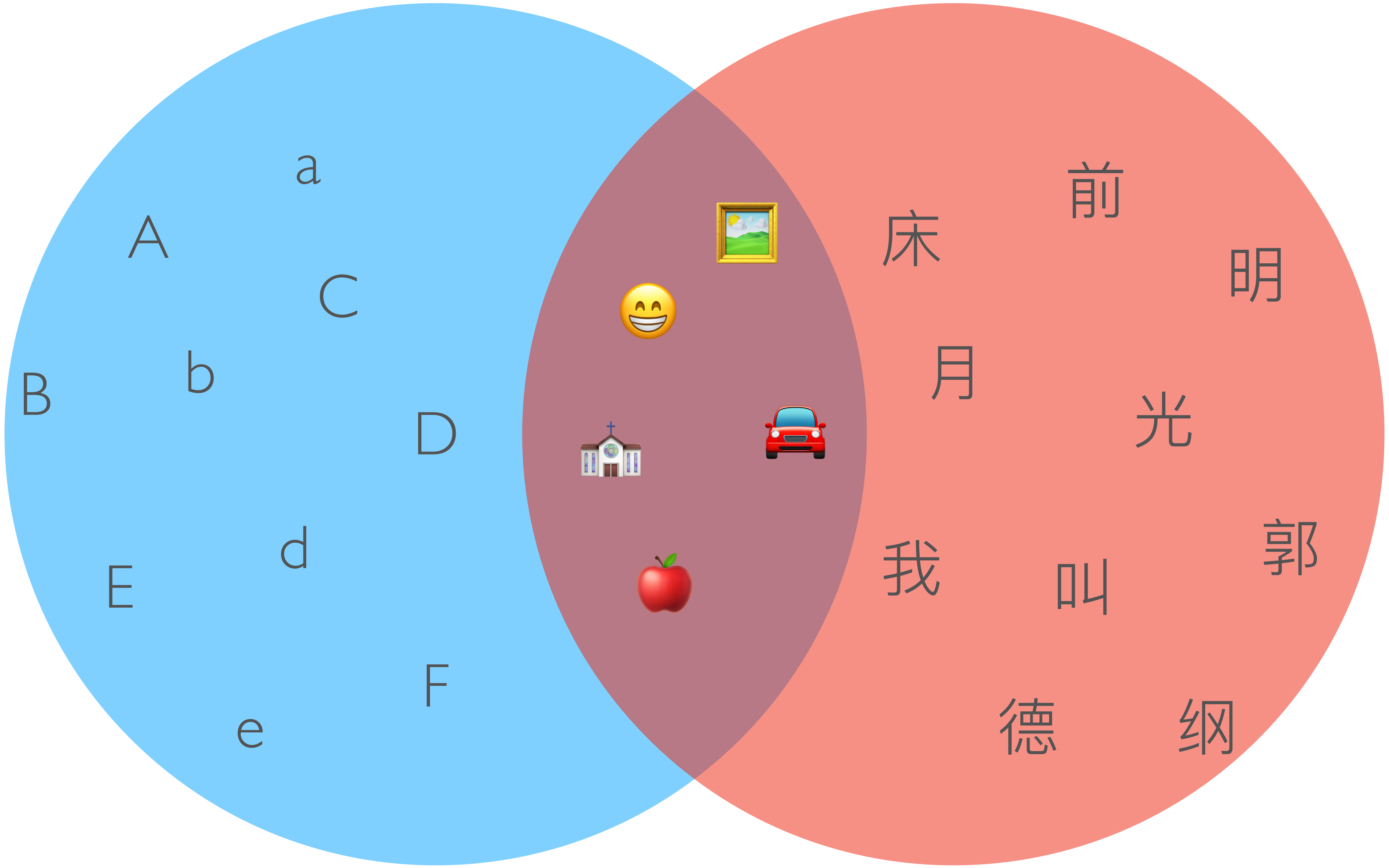
And what existing social/cultural insights do they map onto?

Languages usually share no character...



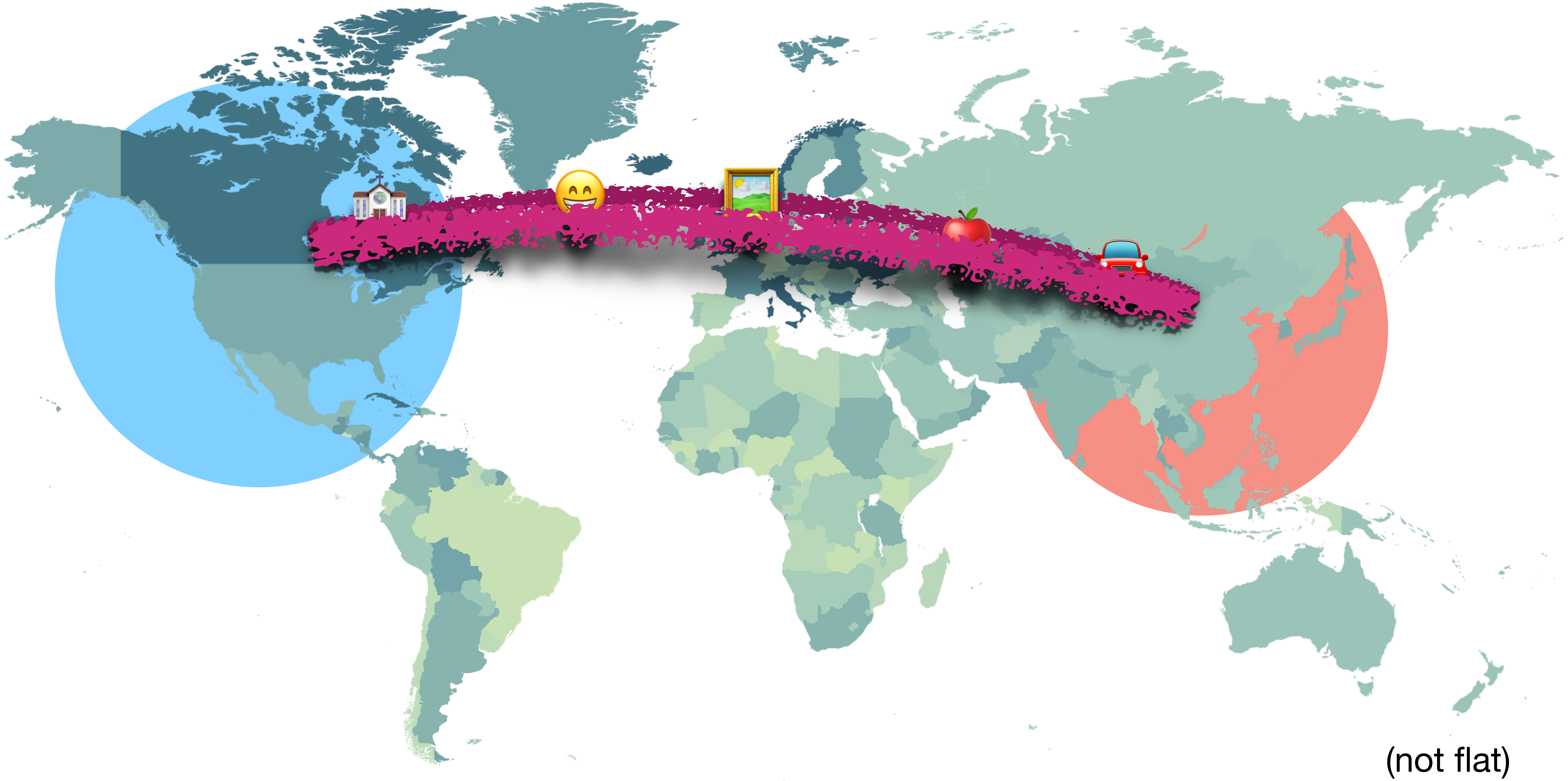
(not sponsored by Flickr)

... until Emojis were invented.



(not sponsored by Mastercard)

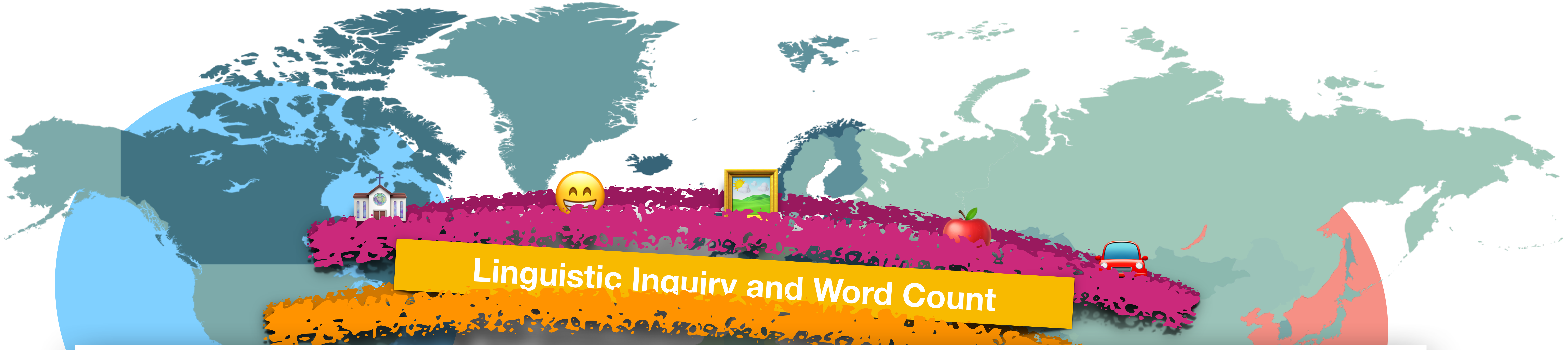
That was our first bridge between East and West.



(not flat)

Another is LIWC.

Linguistic Inquiry and Word Count (LIWC),
James W. Pennebaker and Martha E. Francis, 2015



Category	*Examples*	*Words in category*
Articles	A, an, the	3
Verbs		314
Past tense	Went, ran	145
Present tense	Hear, take	169
Cognitive processes		730
Insight	Think, know	195
Causation	Because, effect	[illegible]
Discrepancy	Should, would	[illegible]

Source: 10.1504/IJWBC.2018.10015402

Method

Calculate similarities from each Emoji to each LIWC category

Linguistic Inquiry and Word Count (LIWC), James W. Pennebaker and Martha E. Francis, 2015

Method

happy mother ❤️ 🏠

bro lol 😁

USA

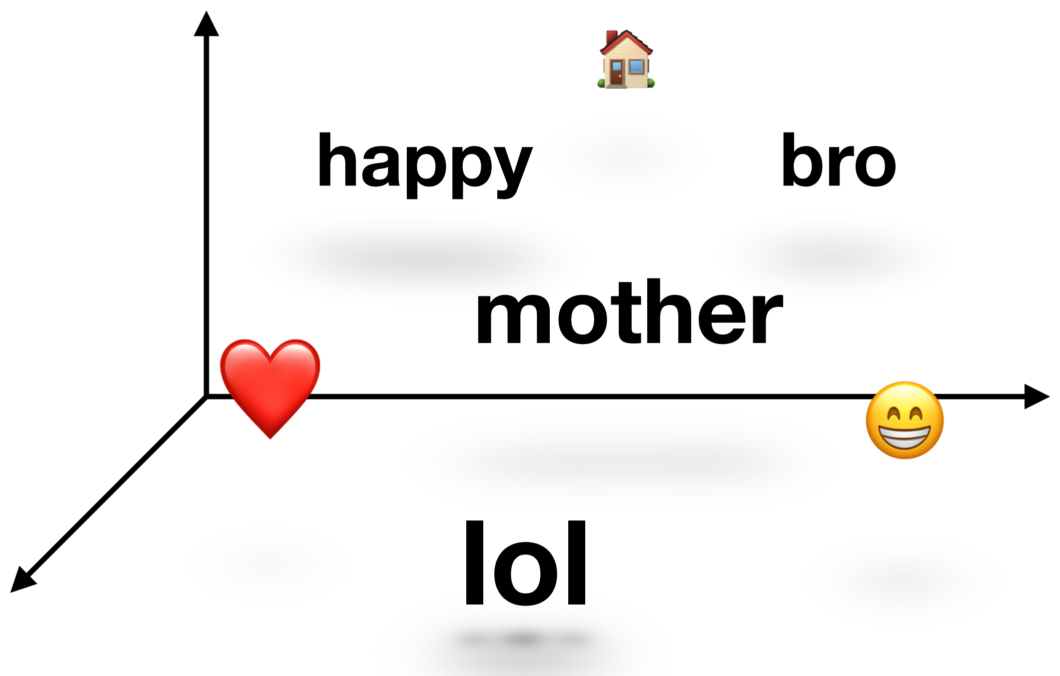
欧耶 哈哈 😁 🏠

家 ❤️ 兄弟

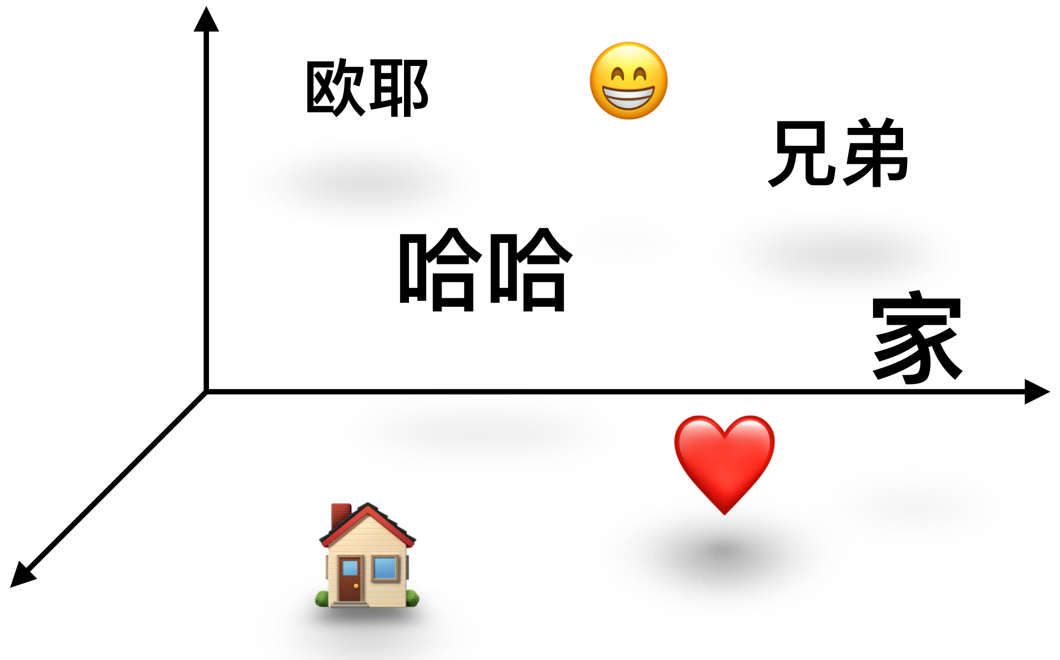
China

take microblog posts as corpora

Method



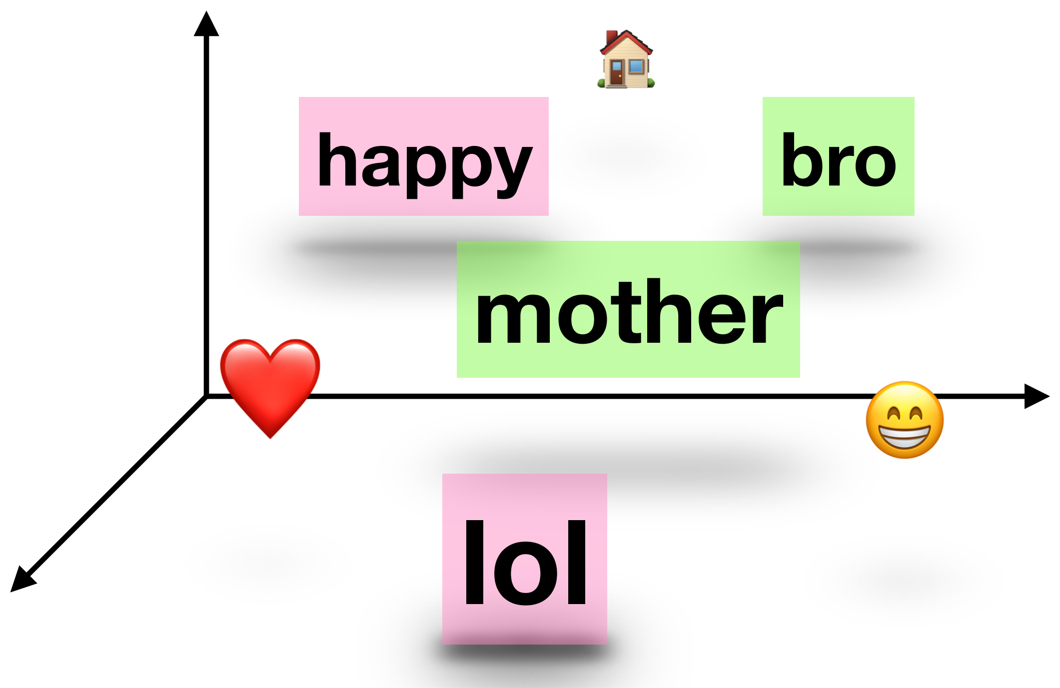
USA



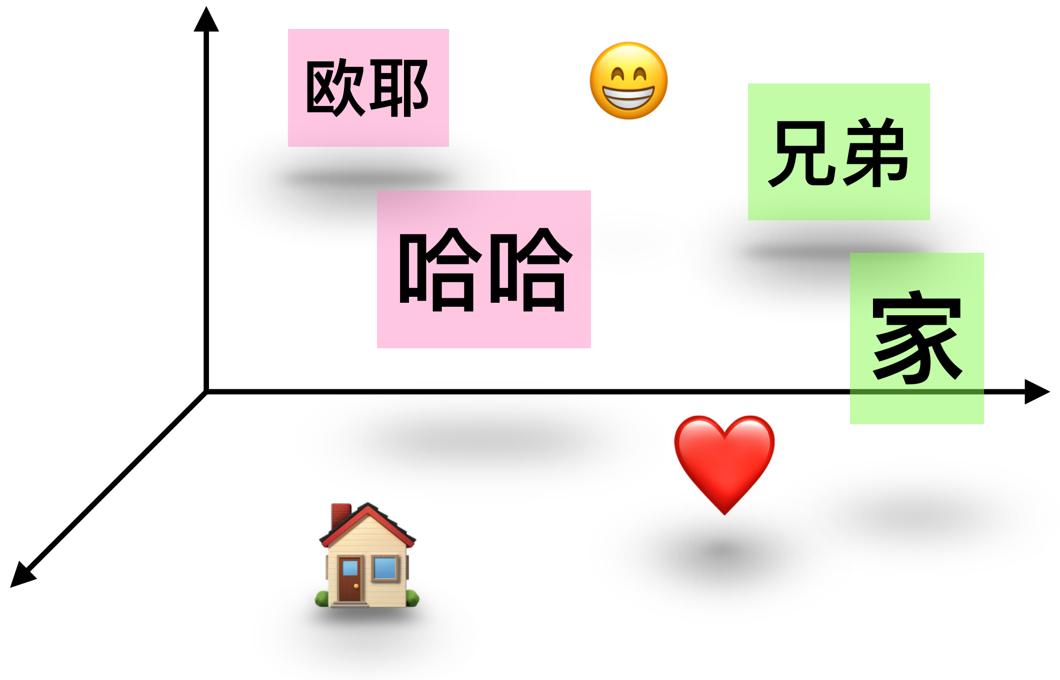
China

learn vectorial representations for all tokens

Method



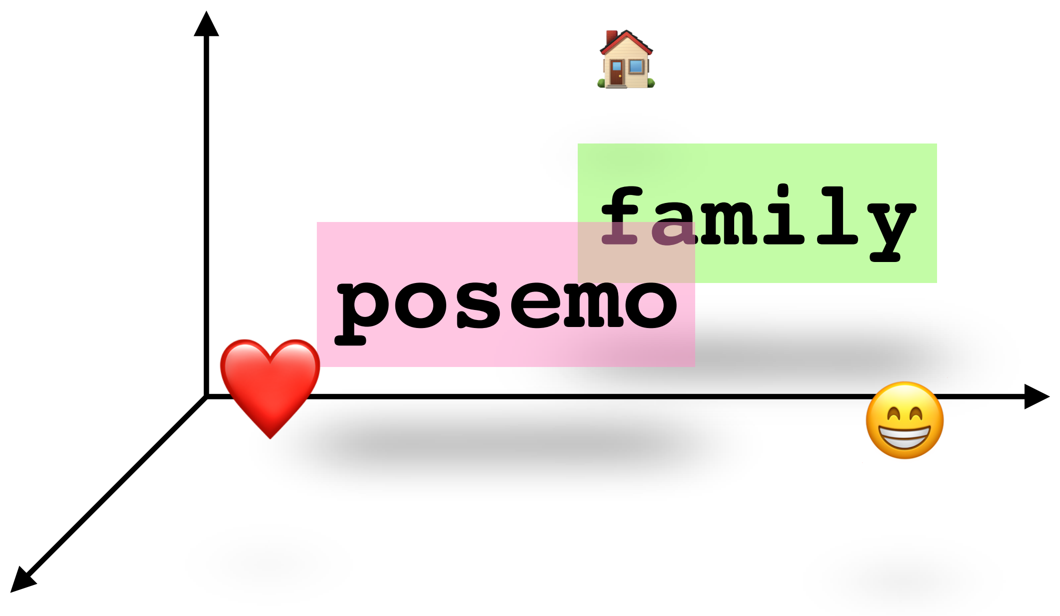
USA



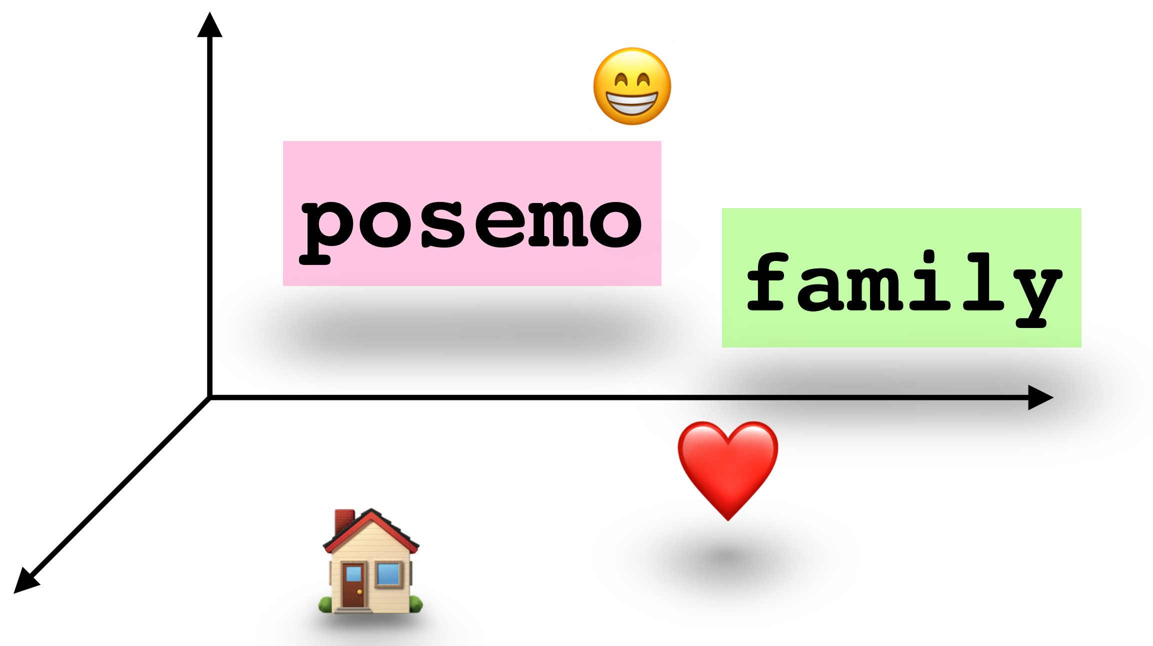
China

group (textual) tokens by LIWC category

Method



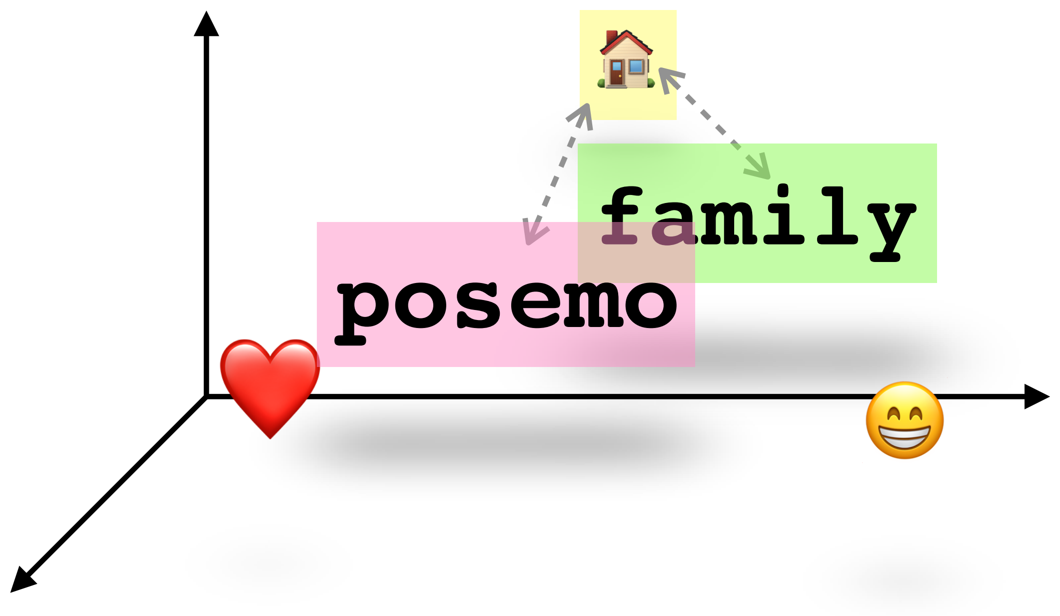
USA



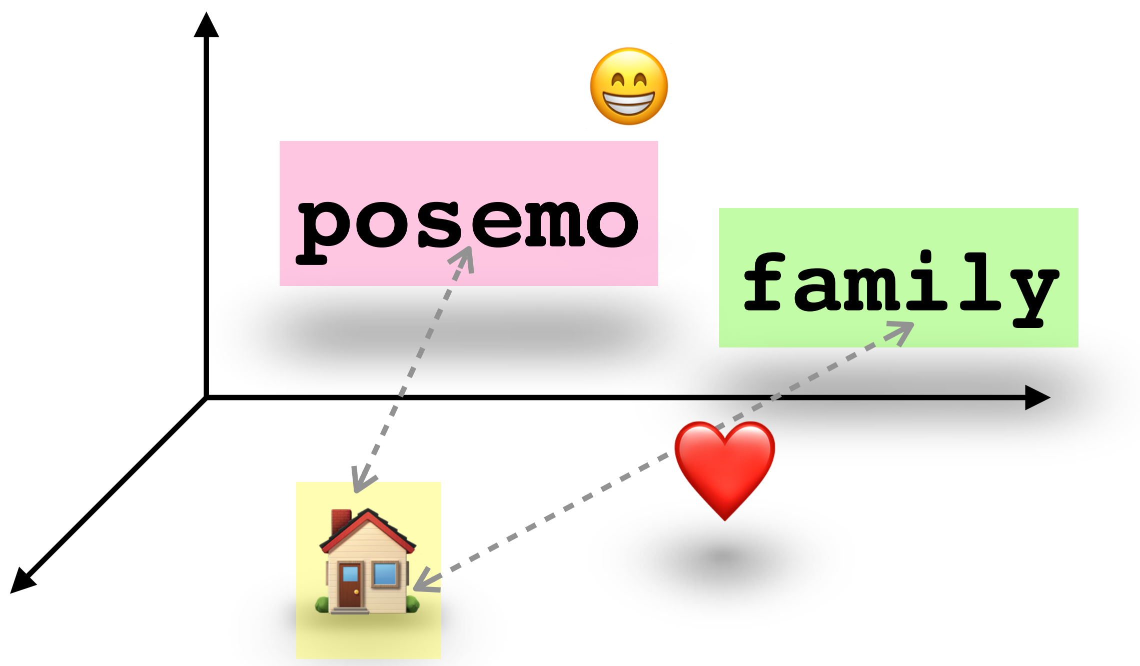
China

take average vector of each category

Method



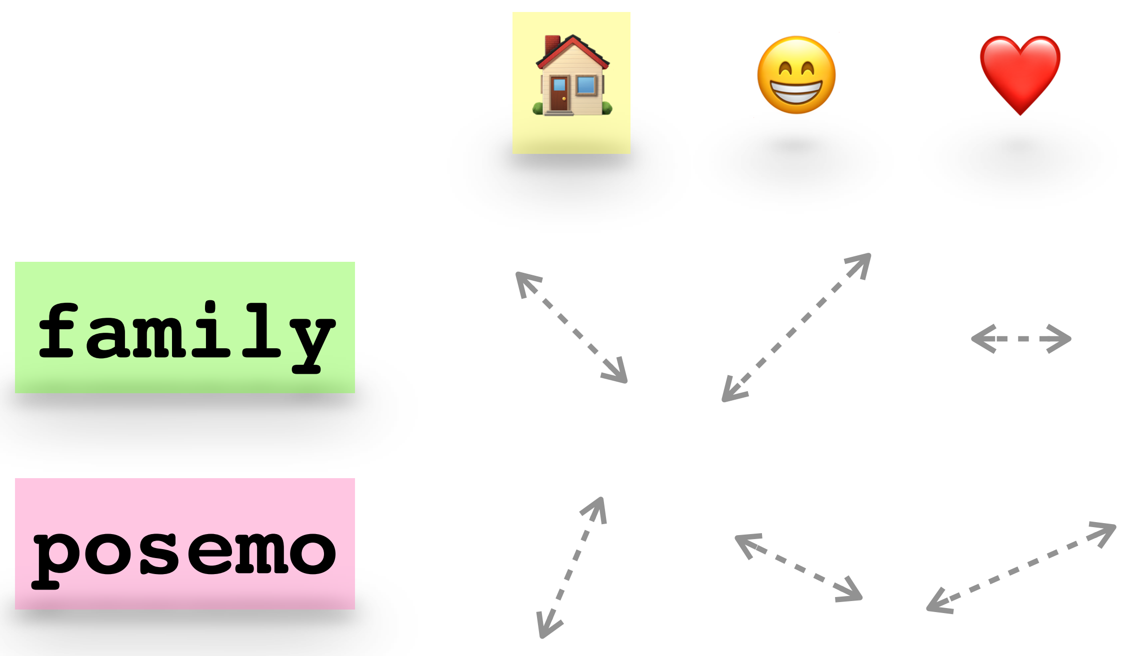
USA



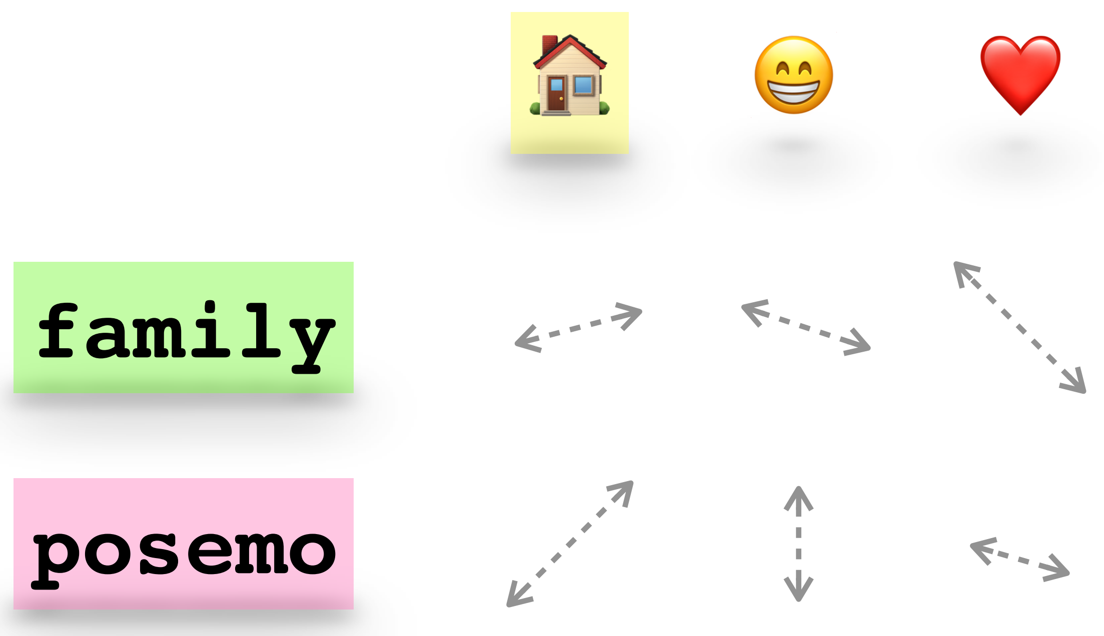
China

compute similarities from each Emoji to each category

Method



USA



China

tabulate and compare

Generalization

- 528 **Emojis**

  1,281 defined in *Emoji 1.0*;
  602 seen across all corpora;
  528 seen >1k times.

- 31 LIWC **categories**

- **Countries**:
  - US🇺🇸 → US🇺🇸, UK🇬🇧, Canada🇨🇦 ≈ West
  - China🇨🇳 → China🇨🇳, Japan🇯🇵 ≈ East
- Group by East/West; take average

Results

Tabulate top-5 most similar Emojis to each LIWC category

LIWC Category	East: China🇨🇳, Japan🇯🇵					West: US🇺🇸, UK🇬🇧, Canada🇨🇦				
	1	2	3	4	5	1	2	3	4	5
Ingest	🍝 80%	🍜 78%	🍛 78%	🍴 78%	🍮 77%	🍝 78%	🍖 76%	🍣 76%	🍗 75%	🍜 75%
Feel	🌂 65%	🍂 64%	🍃 61%	🌿 60%	🌱 60%	💧 62%	🎶 62%	🌞 60%	🎧 60%	🌊 60%
Family	👦 70%	👶 69%	👵 68%	👩 66%	👨 64%	👶 68%	👵 67%	👪 64%	👨 64%	👰 64%
Space	🚉 64%	🗻 61%	🚄 60%	🚃 60%	🗼 60%	🌏 56%	🌎 56%	🌍 56%	🌾 55%	😪 55%
Motion	🏃 67%	🚗 67%	🚲 66%	🚅 66%	🚐 65%	🚐 62%	🚙 61%	🚗 61%	🚶 60%	🏃 60%
Negative Emotion	🚬 61%	😔 60%	😞 60%	😪 59%	🙎 59%	💔 66%	😢 66%	😥 65%	😞 63%	😪 63%

Results

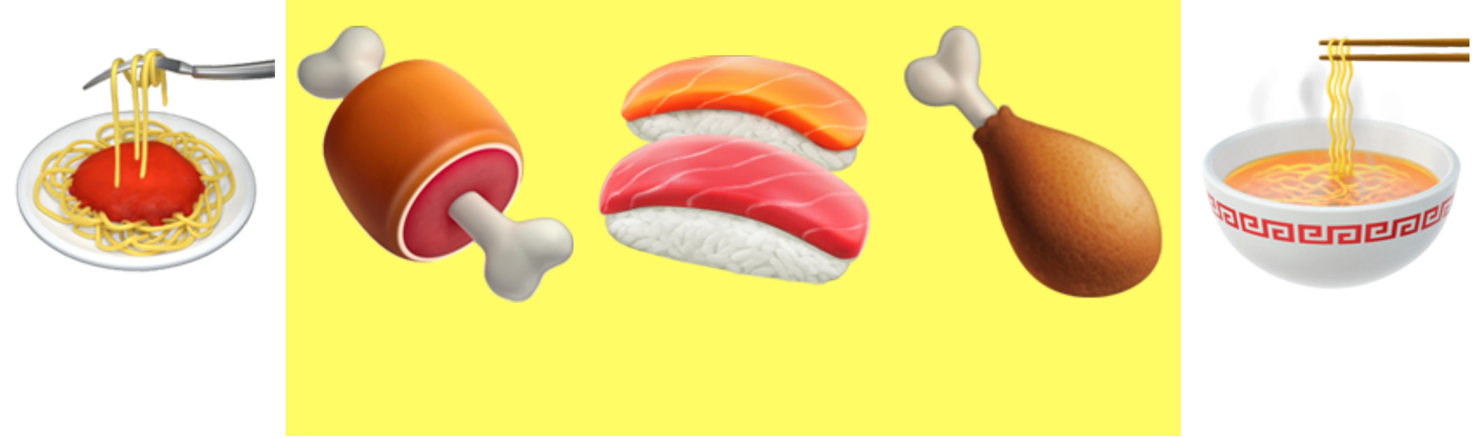
Emoji-Category similarities reflect social & cultural facts

Taking `Ingest`, `Feel`, `Space`, `Family`, and `Money` as examples

`Ingest:` Food Consumptions

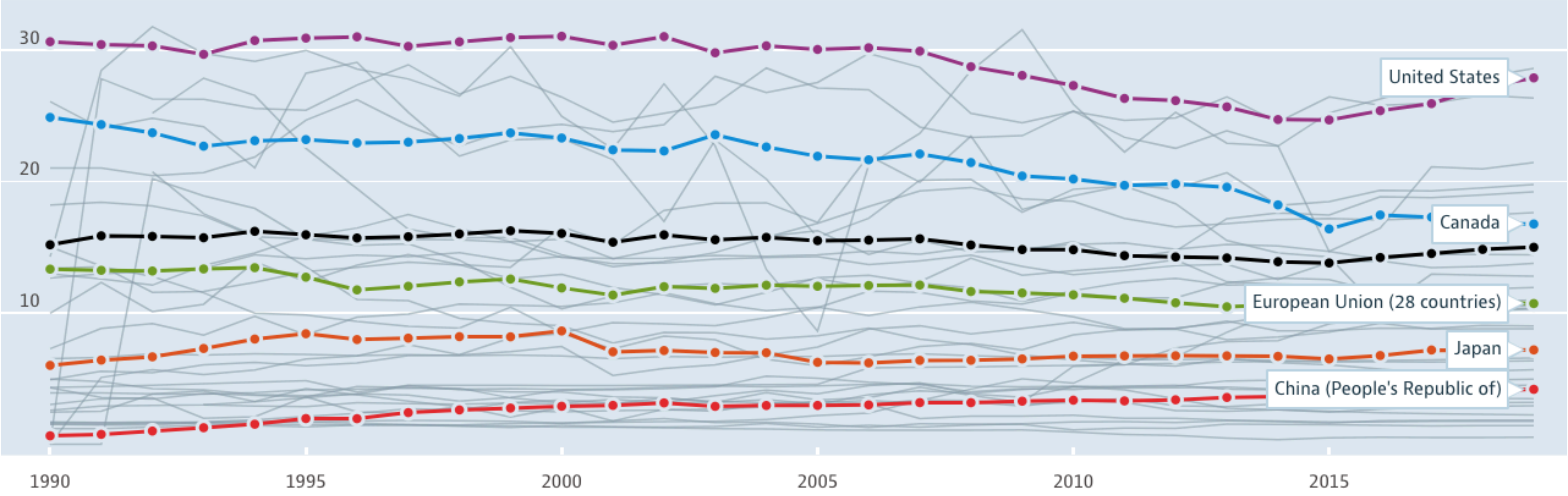
East

West



The West Consumes More Meat

Meat Consumption by Country. https://data.oecd.org/agroutput/meat-consumption.htm



`Feel`: Emotional Feelings

East

West

Seasonal changes on the East

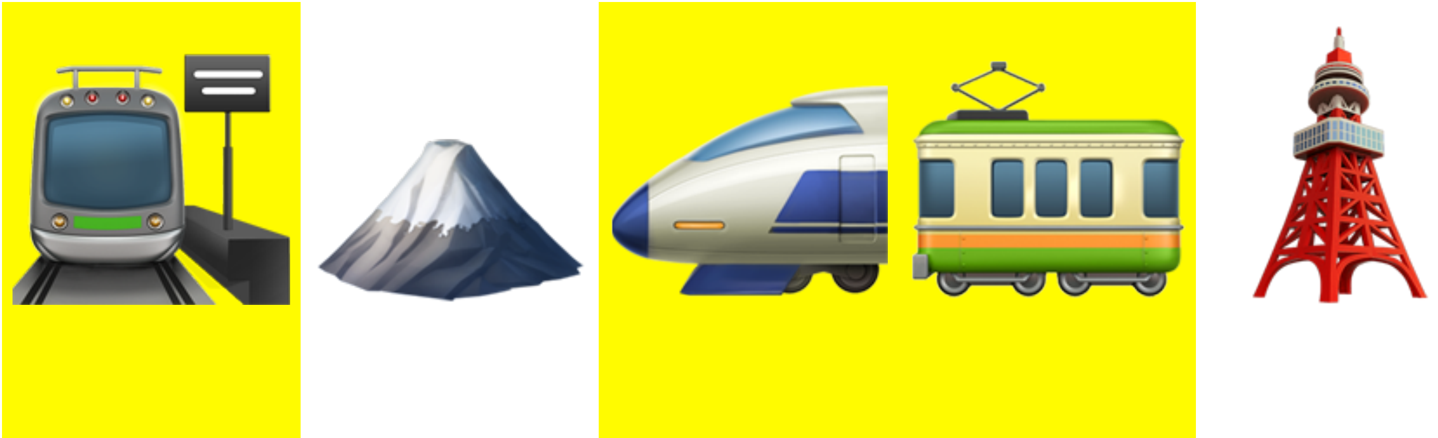
- "[...] **winter depression** being highest in the **Asian** group."[1]
- "[...] a sense of **powerlessness** due to the laws of nature [...] Its association with nature and seasonal changes is **characteristic of Chinese** [...]"[2]

---

1. Suhail, K., & Cochrane, R. (1997). **Seasonal changes in affective state in samples of Asian and white women**. Social Psychiatry and Psychiatric Epidemiology: The International Journal for Research in Social and Genetic Epidemiology and Mental Health Services, 32(3), 149–157.
2. Harkins, J., & Wierzbicka, A. (2001). **Emotions in Crosslinguistic Perspective**. Walter de Gruyter. Page 390.

`Space`: Transportation Methods

East

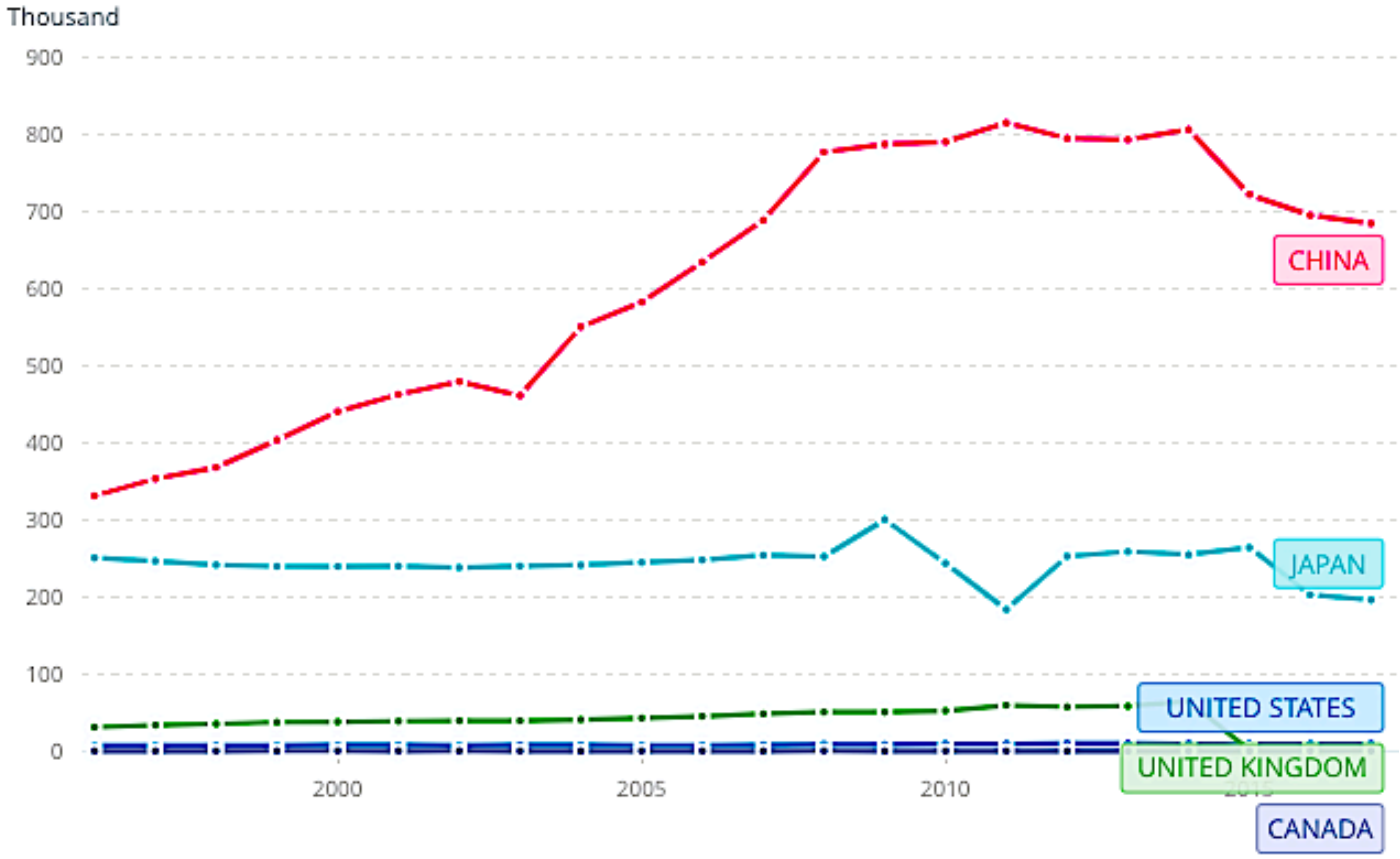


West

The East utilizes more railway

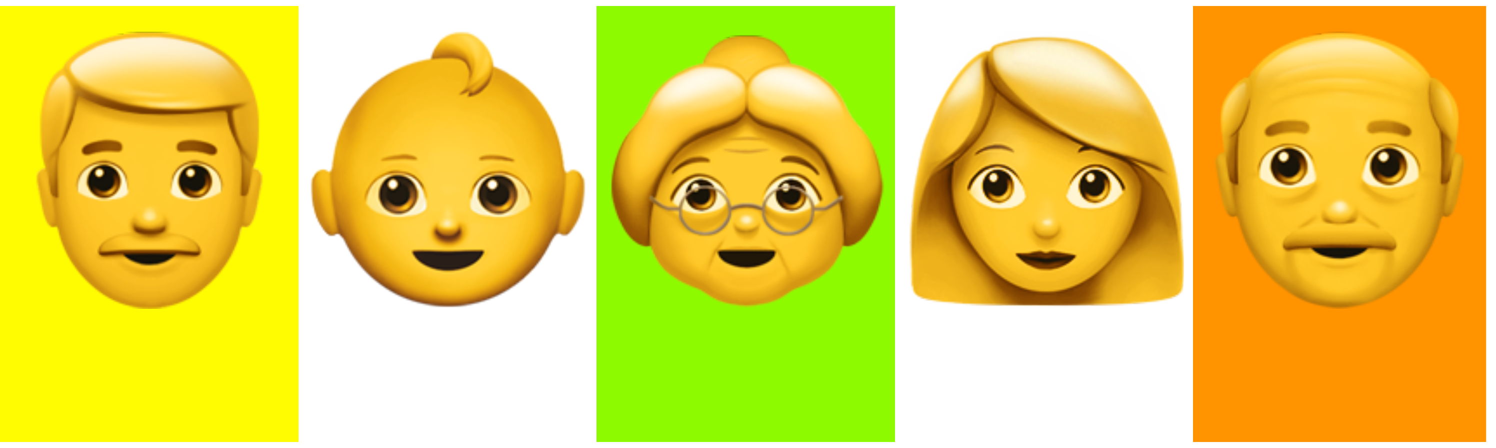
Passengers carried per km of railway.

From the World Bank. Indicator: `IS.RRS.PASG.KM`

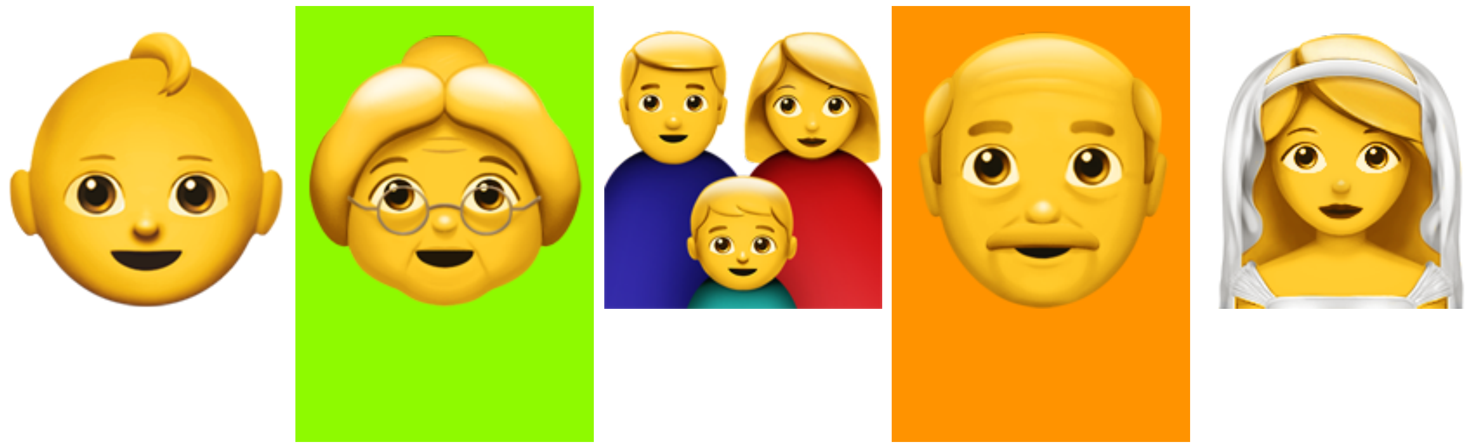


`Family`: Family Compositions

East



West



Father figure on the East

- "[...] Confucian society itself as a large family, in which the **father comes first** [...]"[1]

Grandmothers before grandfathers

- "[...] grandchildren reported more **optimal** relations with **grandmothers**."[2]

1. Kim, K. H. (2007). **Exploring the Interactions between Asian Culture (Confucianism) and Creativity**. The Journal of Creative Behavior, 41(1), 28–53. Page 34.

2. Creasey, G. L., & Koblewski, P. J. (1991). **Adolescent grandchildren’s relationships with maternal and paternal grandmothers and grandfathers**. Journal of Adolescence, 14(4), 373–387. https://doi.org/10.1016/0140-1971(91)90005-C

`Money`: Monetary Matters

East

West

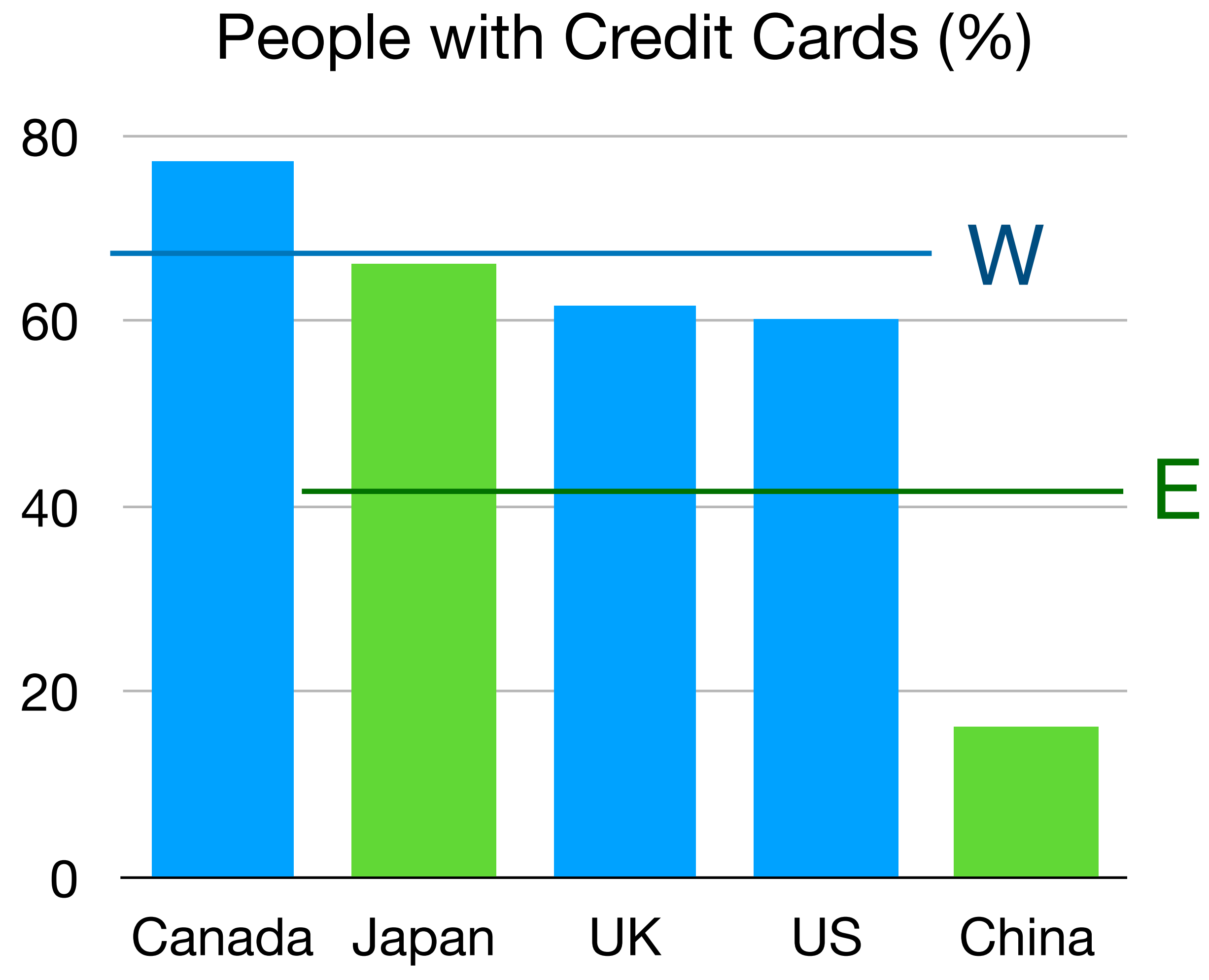
Credit card only on the West[1] -- low popularity of credit cards in Asia:

Yuans/Yens (¥) precedes dollars ($) and pounds (£) on the East:

Japan Yen, ¥

US Dollar, $

UK Pound, £



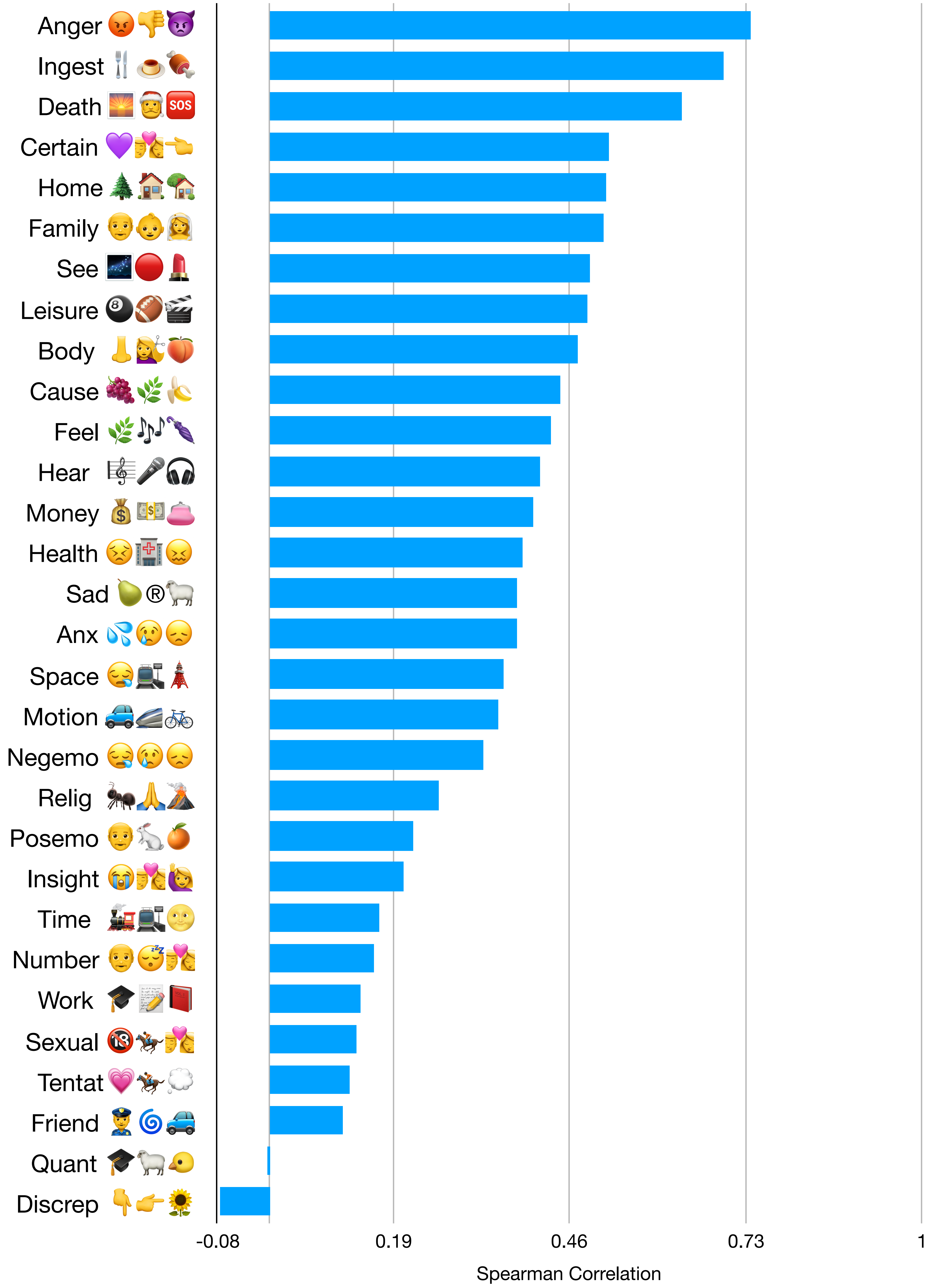
1. https://www.theglobaleconomy.com/rankings/people_with_credit_cards/

Results
Why just top-5?

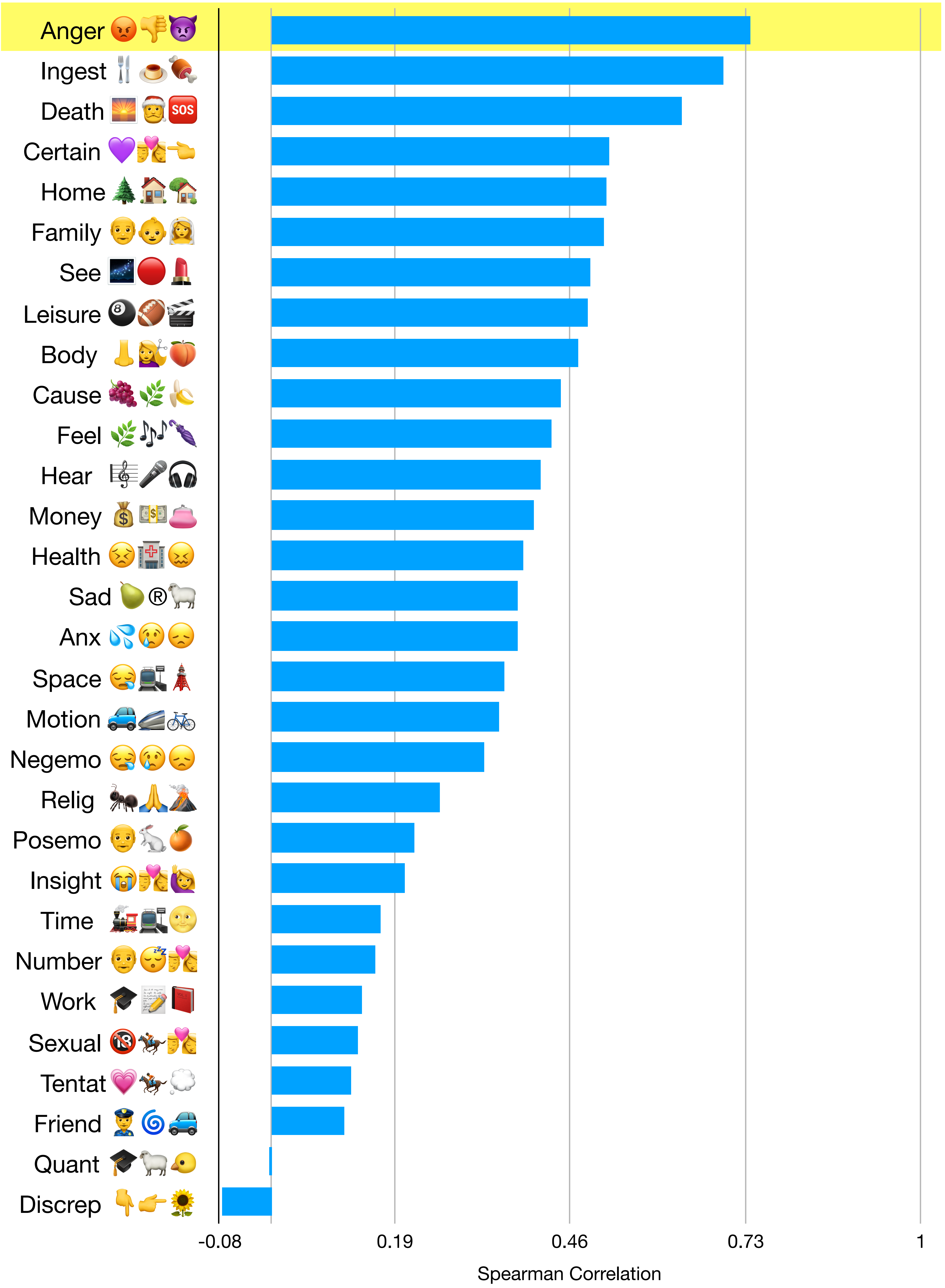
Spearman correlation analysis helps compare usage of **all Emojis** across East and West.

LIWC topics, sorted by **cross-cultural correlations** in Emoji similarity

- **Algorithm:** Spearman's Rank-Order Correlation
  - **Why?** It forgives non-linearity
- **Higher =** more agreement in associations to Emojis across the East and the West.

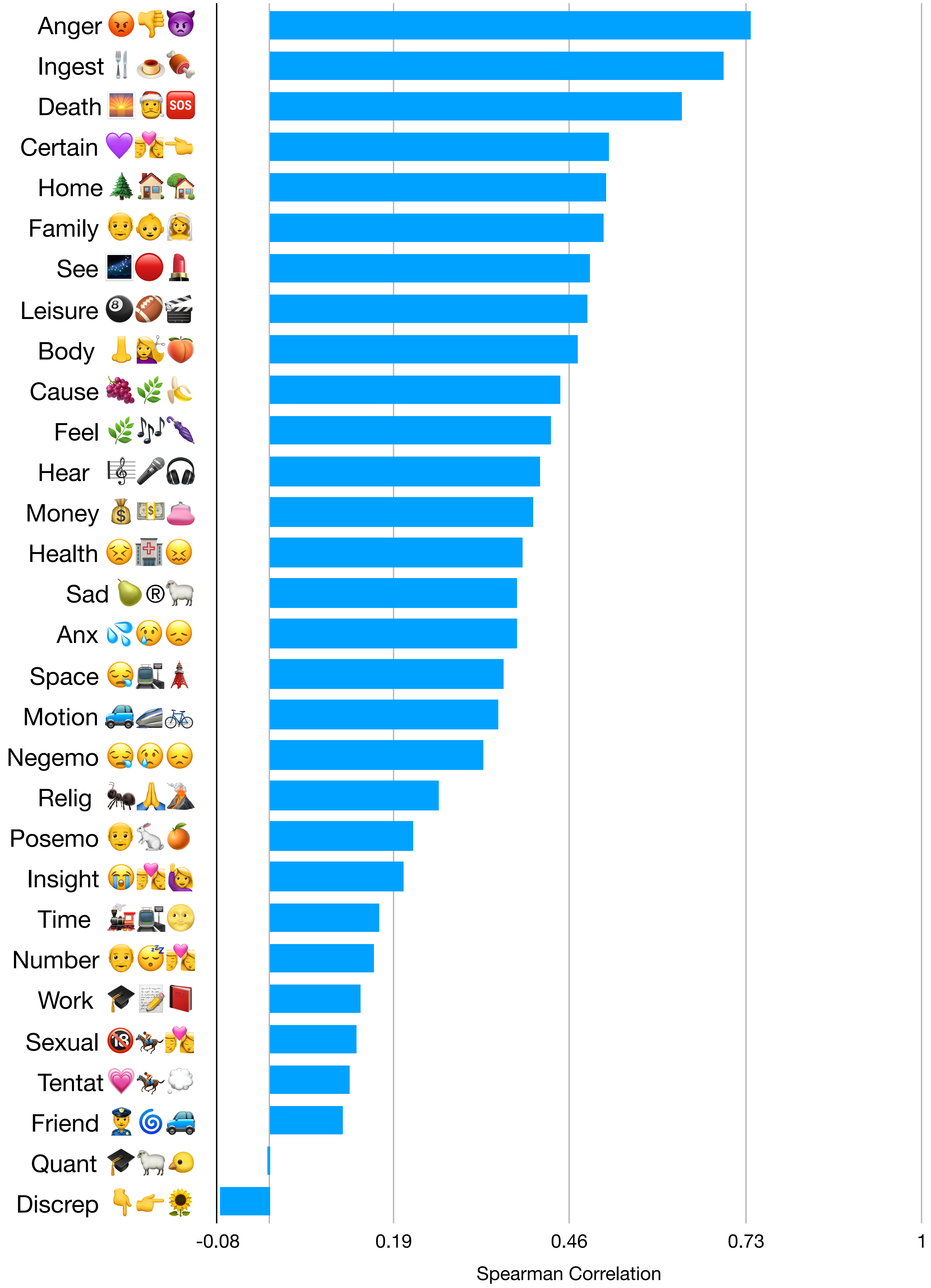


LIWC topics, sorted by **cross-cultural correlations** in Emoji similarity



- **Algorithm:** Spearman's Rank-Order Correlation
  - **Why?** It forgives non-linearity
- **Higher** = more agreement in associations to Emojis across the East and the West.
- **Question:** Among neg. emo., why is anger so universal?
  - anger ~ .7
  - anxiety and sadness ~ .4

LIWC topics, sorted by **cross-cultural correlations** in Emoji similarity



- **Algorithm:** Spearman's Rank-Order Correlation
  - **Why?** It forgives non-linearity
- **Higher =** more agreement in associations to Emojis across the East and the West.
- **Question:** Among neg. emo., why is anger so universal?
  - Better yet, what about pos. emo.?

“Yes; emotions are somewhat universal across cultures, but to different degrees.”

–Ekman, et al., paraphrased.

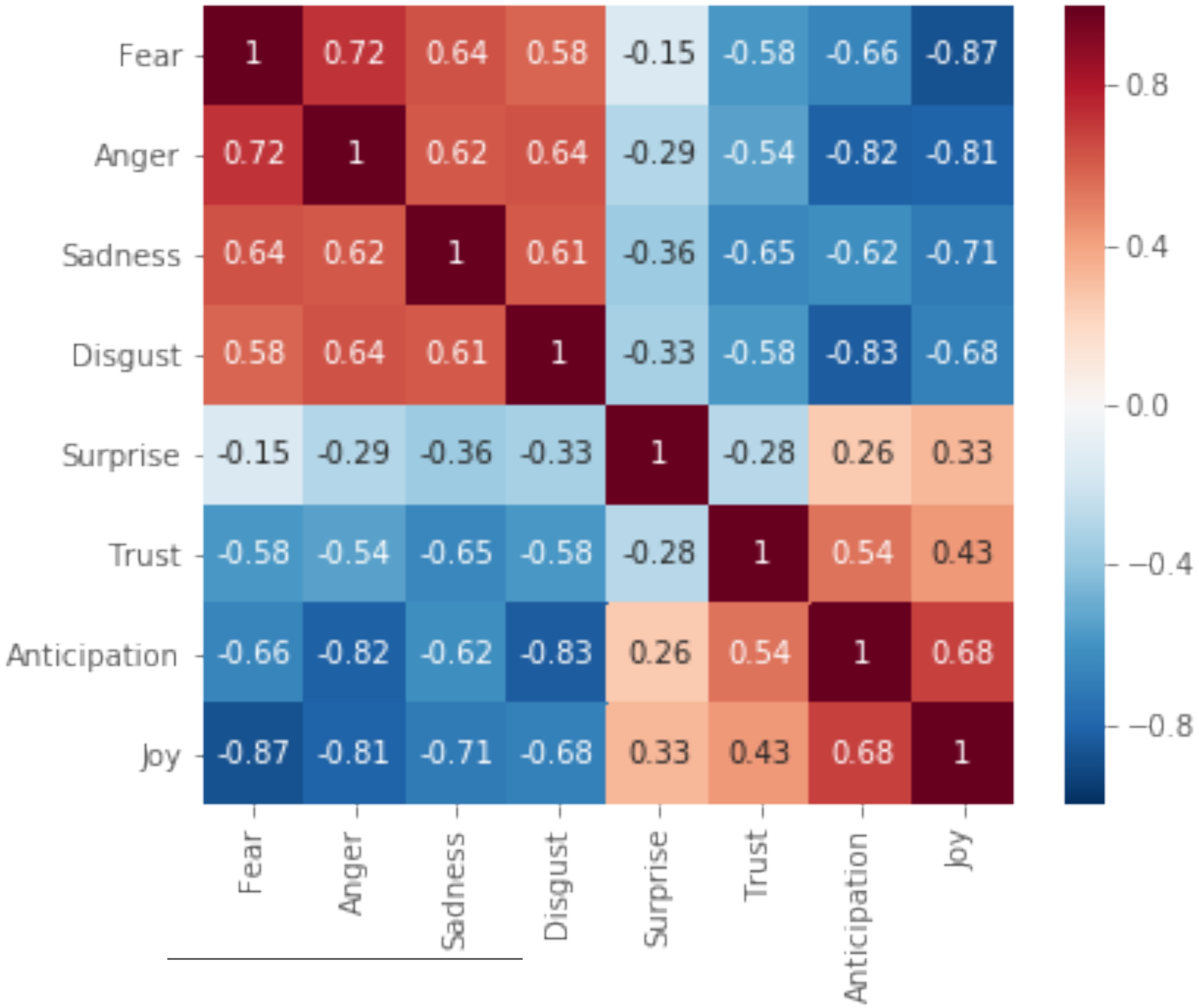
Discussion
Emoji-to-Emotion Similarities

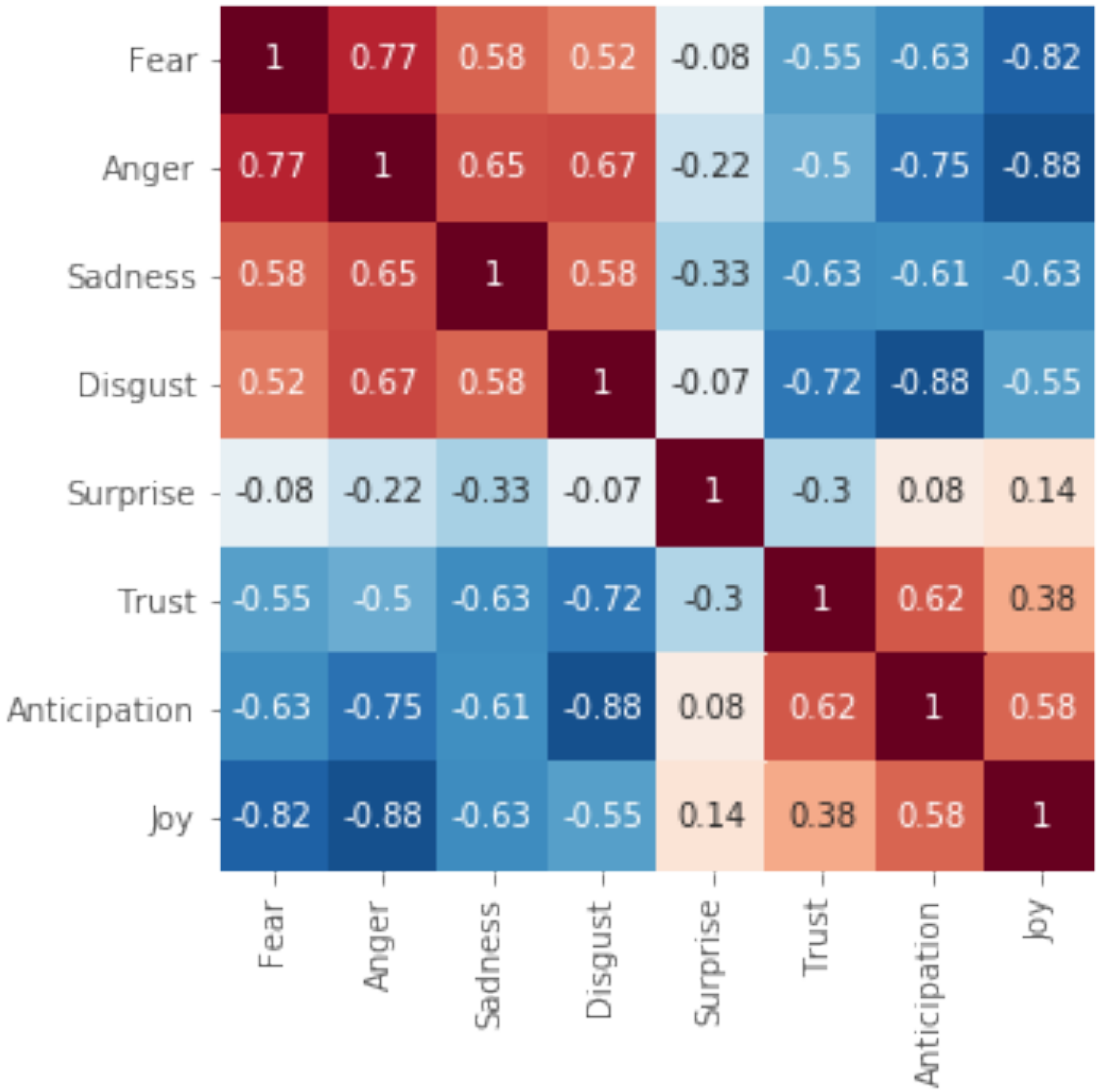
What happens if we replace LIWC with EmoLex[1, 2]?

1. Crowdsourcing a Word-Emotion Association Lexicon, Saif Mohammad and Peter Turney, *Computational Intelligence,* 29 (3), 436-465, 2013.

2. Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon, Saif Mohammad and Peter Turney, In Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, June 2010, LA, California.

Cosine Similarities between Emotion Vectors





1. Li, M., Guntuku, S. C., Jakhetiya, V., & Ungar, L. H. (2019). **Exploring (Dis-) Similarities in Emoji-Emotion Association on Twitter and Weibo.**

Pre-trained word embedding models to share!

- **Language**: Mandarin Chinese
- **Corpora**: Sina Weibo
- **Scope**: 5,000,000 UNIQUE posts every year through 2012~2018
- **Algorithm**: fastText (word2vec improved), 10-fold each year
- A total of 70 (~300GB) binaries!
- Useful for *temporal analysis*.

Let's collaborate!

Thank you!

Special thanks to my co-authors:

Sharath Chandra Guntuku

Louis Tay
(Purdue Univ)

Lyle H. Ungar

Code available at https://github.com/tslmy/ICWSM2019

2019 Mingyang Li, University of Pennsylvania