Soundex

• Soundex is a phonetic algorithm used to identify and group words that sound the same.

• Phonetic algorithms help us identify words with different spellings that may be the same or be pronounced similarly, especially names in the US Census.

• Choosing an algorithm for use in social scientific research involves tradeoffs. Choose consciously!
Tradeoffs in phonetic algorithm choice

- Representational Accuracy
- Ease of Analysis
What kind of harm is lack of representation?
Distributive Harms

How should goods or outcomes be distributed?
Distributive Harms

Equality of Opportunity: everyone has the same opportunity to pursue good thing
Distributive Harms

Equality of Outcome:
everyone gets same good things (and same responsibilities)
Distributive Harms

Equality of Welfare: everyone has well-being/happiness, but only some people need strings for their happiness
Representational Harms

Am I represented in this system?

Can I express myself in it?

Does this system recognize me, my culture, and my self-expression?
ASCII

“American Standard Code for Information Interchange” covers 256 symbols

![ASCII Code Chart](image)
Unicode

Covers over 1 million symbols,
Full coverage of 90 languages,
basic coverage for 200 languages

Everyone in the world should be able to use their own language on phones and computers.

LEARN MORE ABOUT UNICODE
Unicode and Representation

Everyone in the world should be able to use their own language on phones and computers.

LEARN MORE ABOUT UNICODE
Language Representation in Unicode

There are at least 7,100 languages spoken today.

Of those, there are at least 24 with 50 million speakers.

Are the languages you speak representable in Unicode?
Katie’s Office Hours

Come say hello any time:

https://calendly.com/kathleencreel