

A Large-Scale Multi-Lingual Color Thesaurus

Albrecht Lindner¹, Bryan Zhi Li¹,
Nicolas Bonnier², and Sabine Süsstrunk¹

1: School of Computer and Communication Sciences, EPFL, Switzerland
2: Océ Print Logic Technologies, Créteil, France

СВЕТЛО ОЛИВКОВЫЙ

gialloverde

파스텔 핑크

violeta profundo

Rasei green

用本列茶色

viridian

lavanda azulado claro

暗蓝宝石色

cerise rouge

R G B ?

Светло оливковый

gialloverde

파스텔 핑크

violeta profundo

Rasen grün

明るい茶色

lavanda azulado claro

viridian

暗蓝宝石色

cerise rouge

Outline

- State of the art.
- 9000+ Color names.
- Statistical color estimation.
- Accuracy.
- Demo.

Color Databases

Psychophysical experiment:

- XKCD color survey
- N. Moroney

No experiment:

- Perbang.dk
- XII
- W3C – HTML

- Time consuming or no psychophysical foundation.
- Only English color names.
- Data-mining approach.

9000+ Color Names

- XKCD color survey, psychophysical experiment.

9000+ Color Names

- XKCD color survey, psychophysical experiment.
- 950 English **color names + color values**.

9000+ Color Names

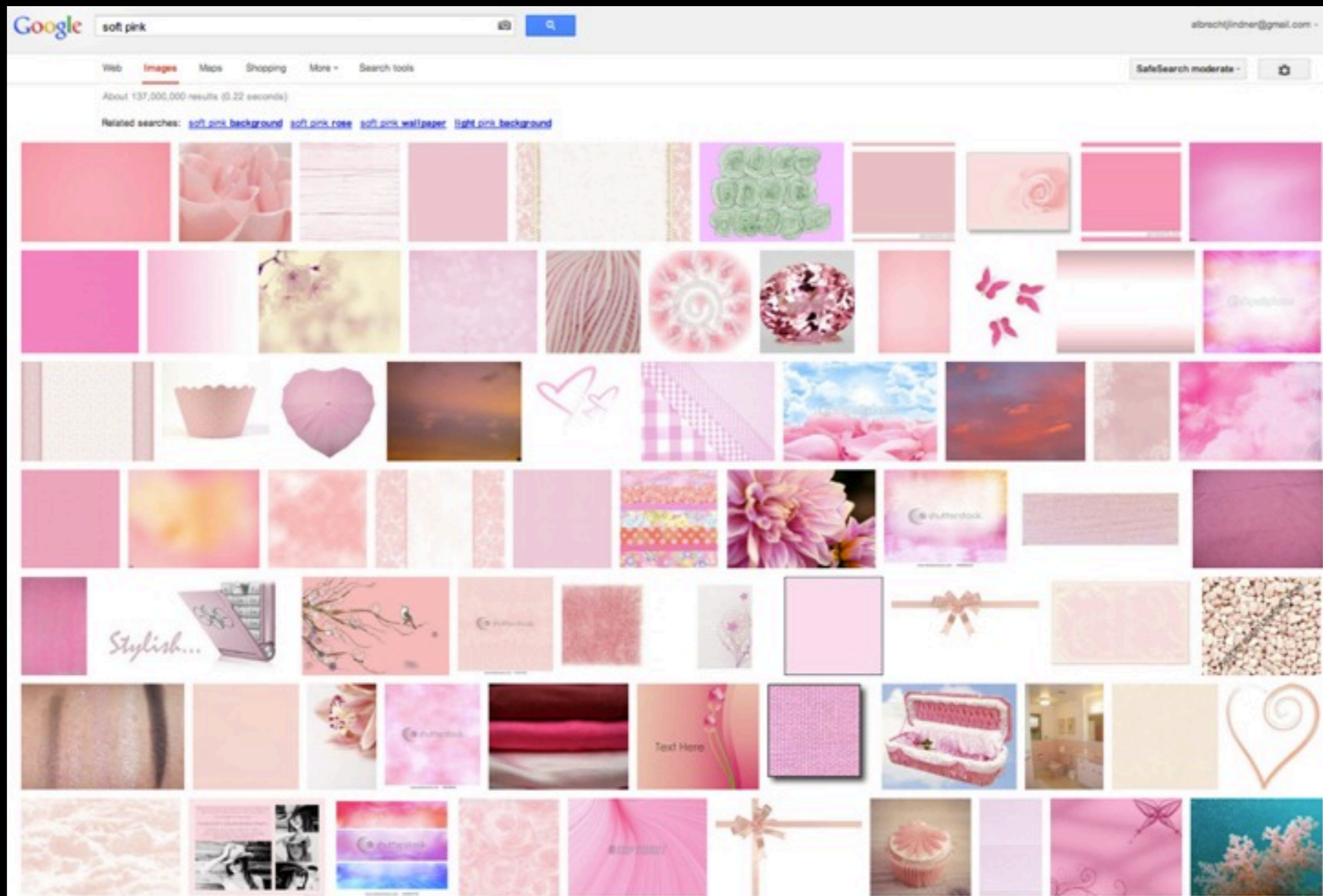
- XKCD color survey, psychophysical experiment.
- 950 English color names + color values.
- Translate to **9 other languages:**
Chinese, French, German, Italian, Japanese, Korean,
Portuguese, Russian, and Spanish.

9000+ Color Names

- XKCD color survey, psychophysical experiment.
- 950 English color names + color values.
- Translate to 9 other languages:
Chinese, French, German, Italian, Japanese, Korean, Portuguese, Russian, and Spanish.
- Example: 柔和的粉红色, *soft pink*, *rose tendre*, *sanftes pink*, *rosa tenue*, ソフトピンク, 부드러운 녹색, *rosa suave*, *нежно розовый*, *rosa suave*.

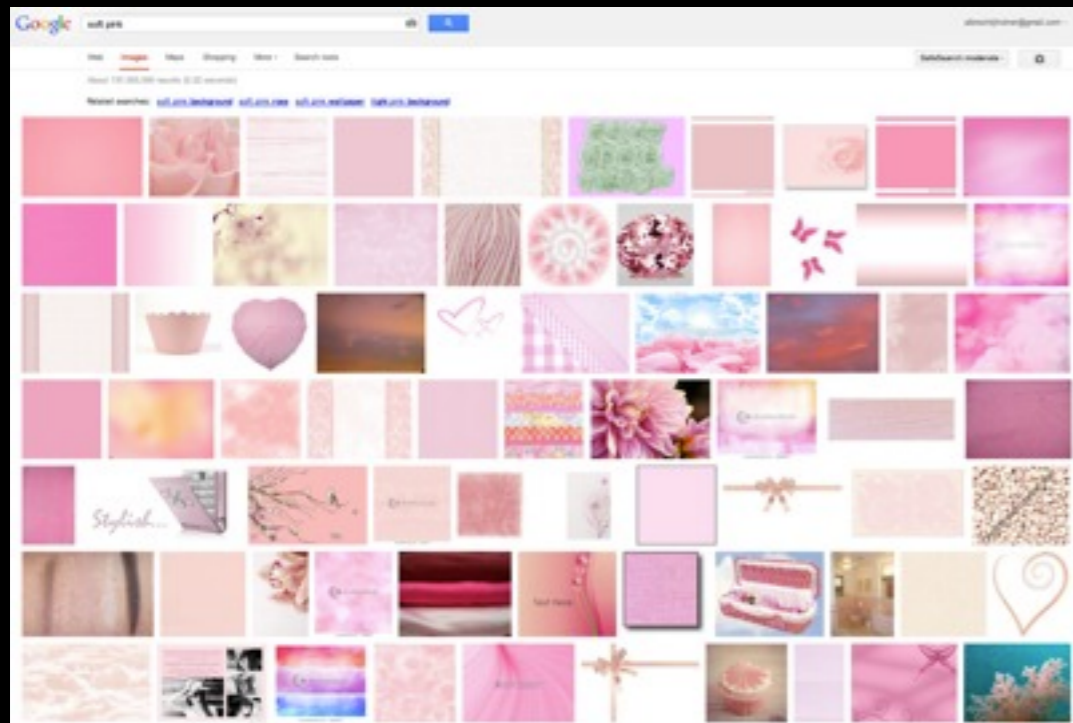
Data Acquisition

Google Image: *soft pink*



Data Acquisition

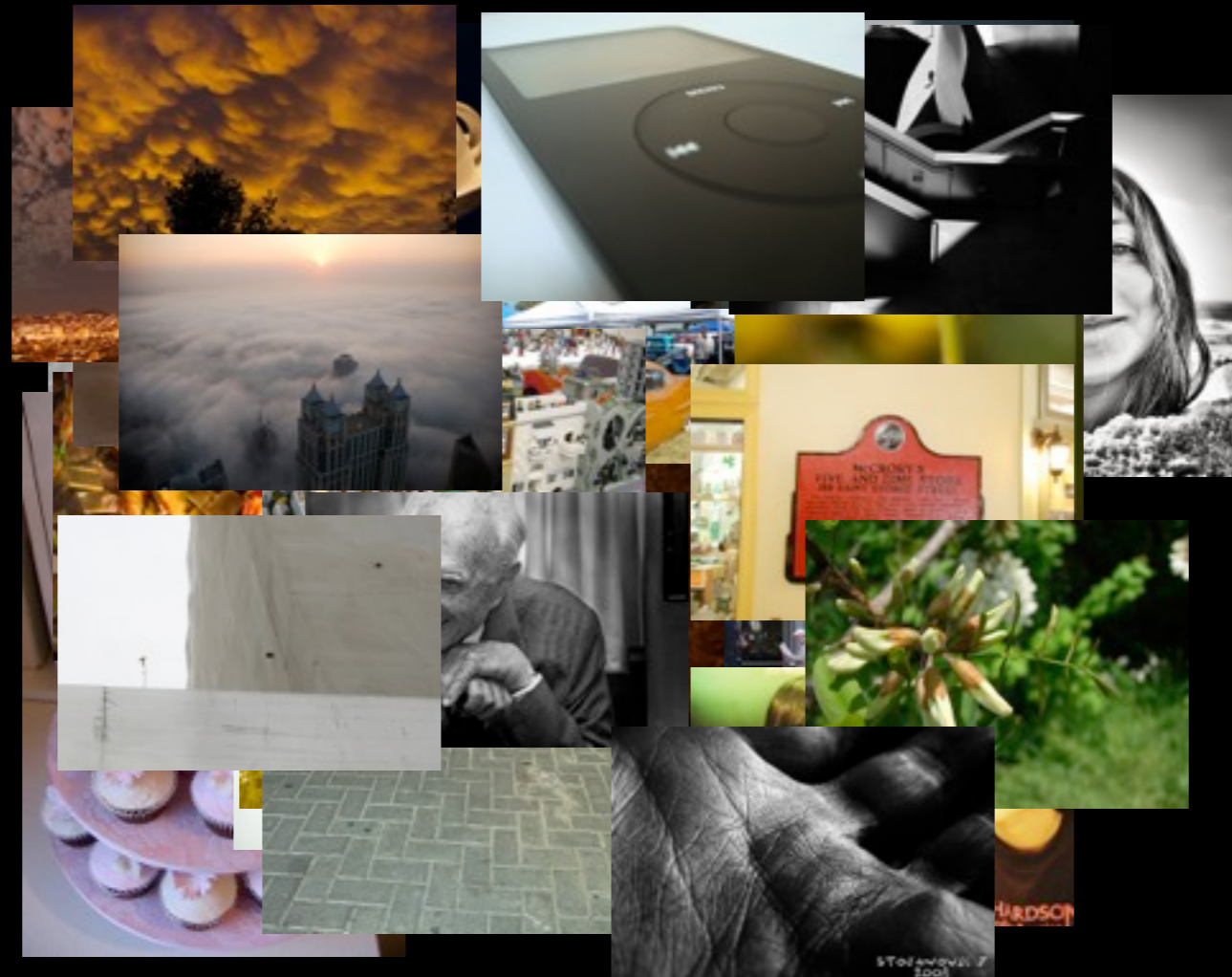
Google Image: *soft pink*



- 100 images per color name.
- Language and country restrict.
- Assume sRGB encoding.
- Almost 1M images.

Statistical Framework

soft pink



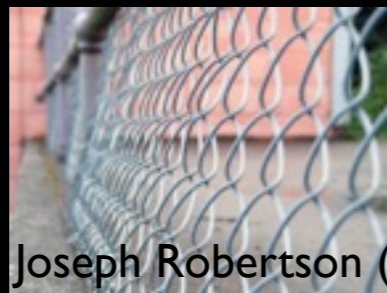
IM images

Statistical Framework

soft pink



EmPREzZz (cc)



Joseph Robertson (cc)



Dorocia (cc)



lacquerbelle (cc)

4

soft pink



TW Collins (cc)



raketentim (cc)



Dunechaser (cc)



golfride (cc)



Tal Bright (cc)



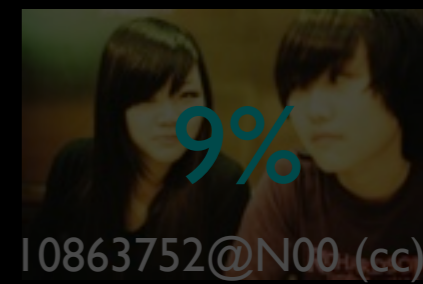
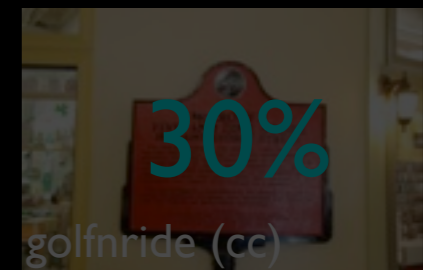
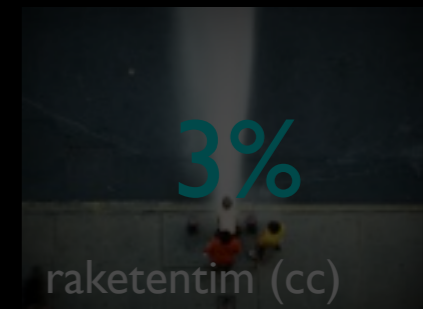
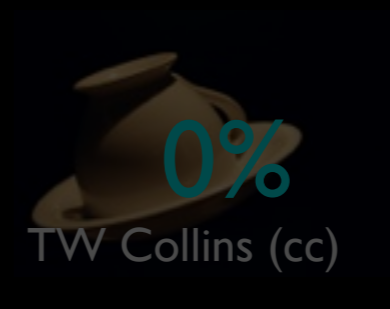
10863752@N00 (cc)

6

Statistical Framework

soft pink

soft pink



percentage of pinkish pixels

Statistical Framework

sorted list: 0% 2% 3% 5% 8% 9% 25% 30% 40% 90%
rank index: 1 2 3 4 5 6 7 8 9 10
ranksum: $T = 4 + 7 + 9 + 10 = 30$

Mann-Whitney-Wilcoxon ranksum test

$$\mu_T = \frac{n_1(n_1 + n_2 + 1)}{2}$$

$$\sigma_T^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

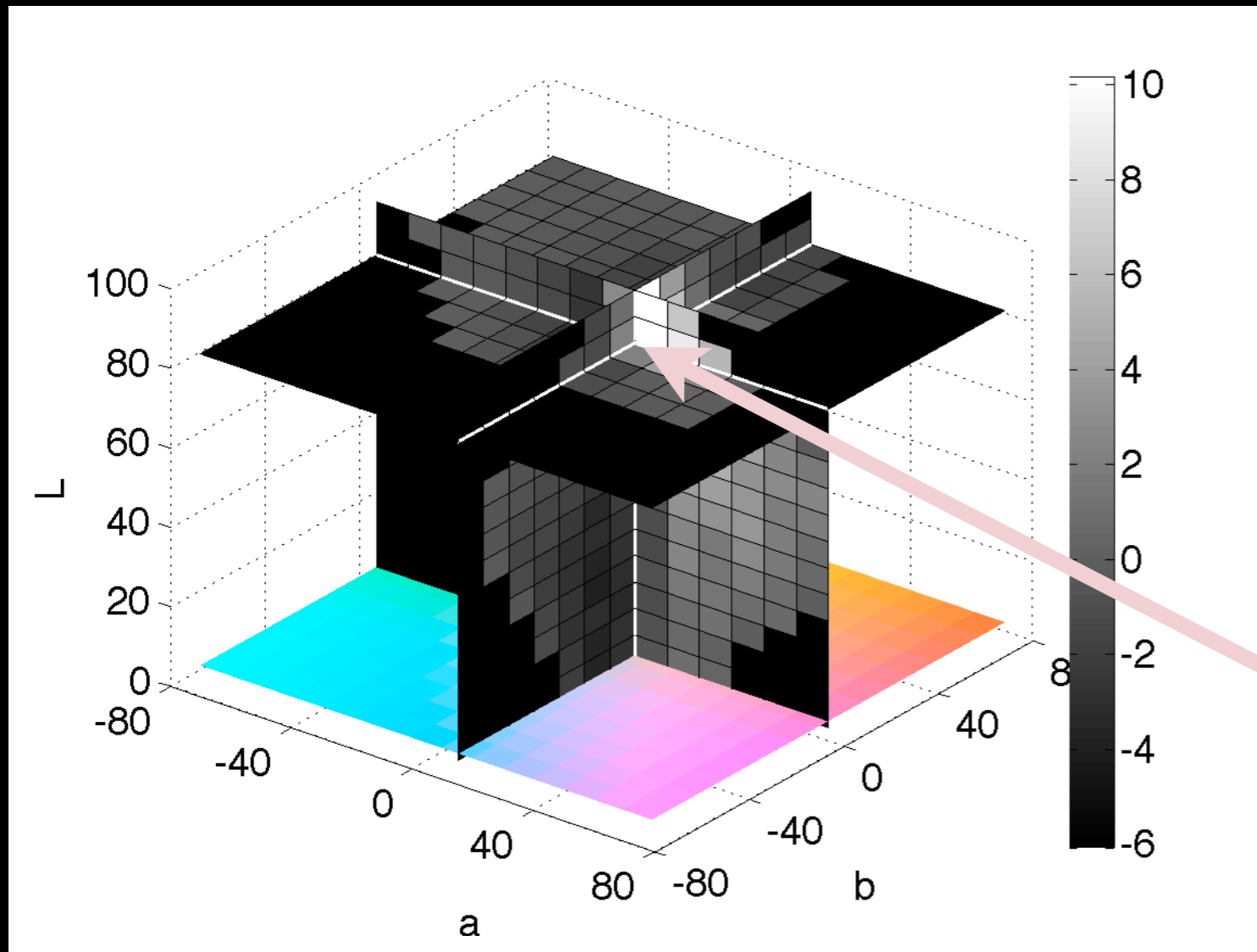
n_1, n_2 cardinalities
of both sets

$$z = \frac{T - \mu_T}{\sigma_T} = \frac{30 - 22}{4.69} \approx 1.71$$

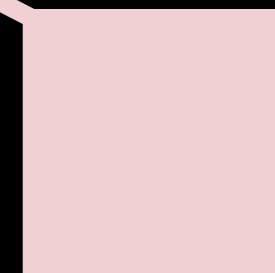
$z > 0 \rightarrow$ significantly more pink pixels in *soft pink* images.

ζ Distribution

soft pink, English

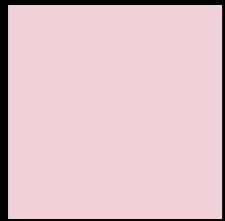


- CIELAB histogram 15x15x15 bins.
- ζ values indicate the most significant color value for a color name.

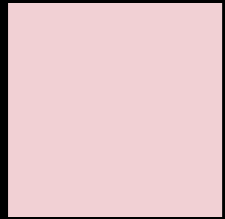


sRGB: 238, 197, 203

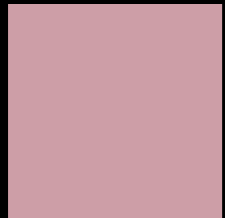
Soft Pink



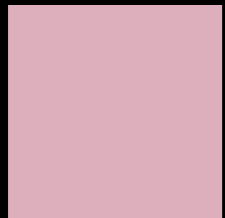
柔和的粉红色, cn



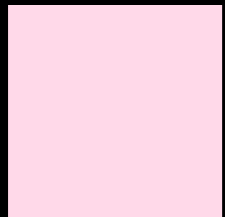
soft pink, en



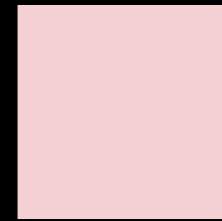
rose tendre, fr



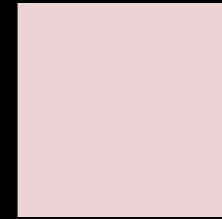
sanftes pink, de



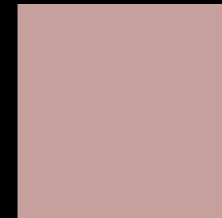
rosa tenue, it



ソフトピンク, jp



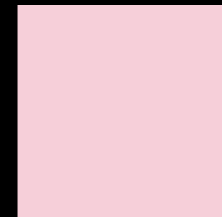
부드러운 녹색, ko



rosa suave, pt

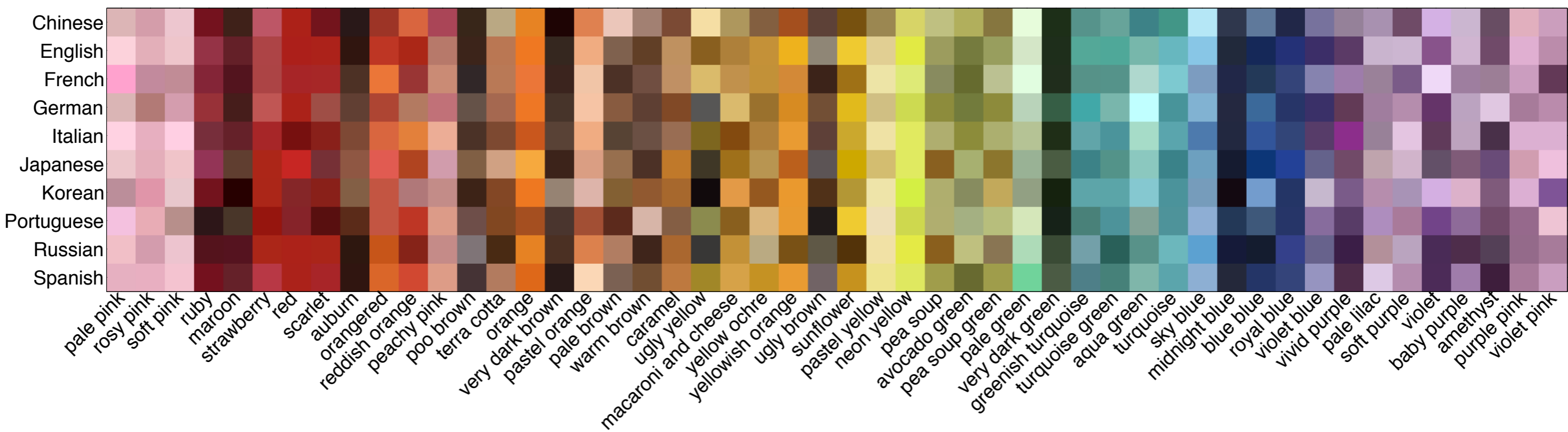


нежно розовый, ru



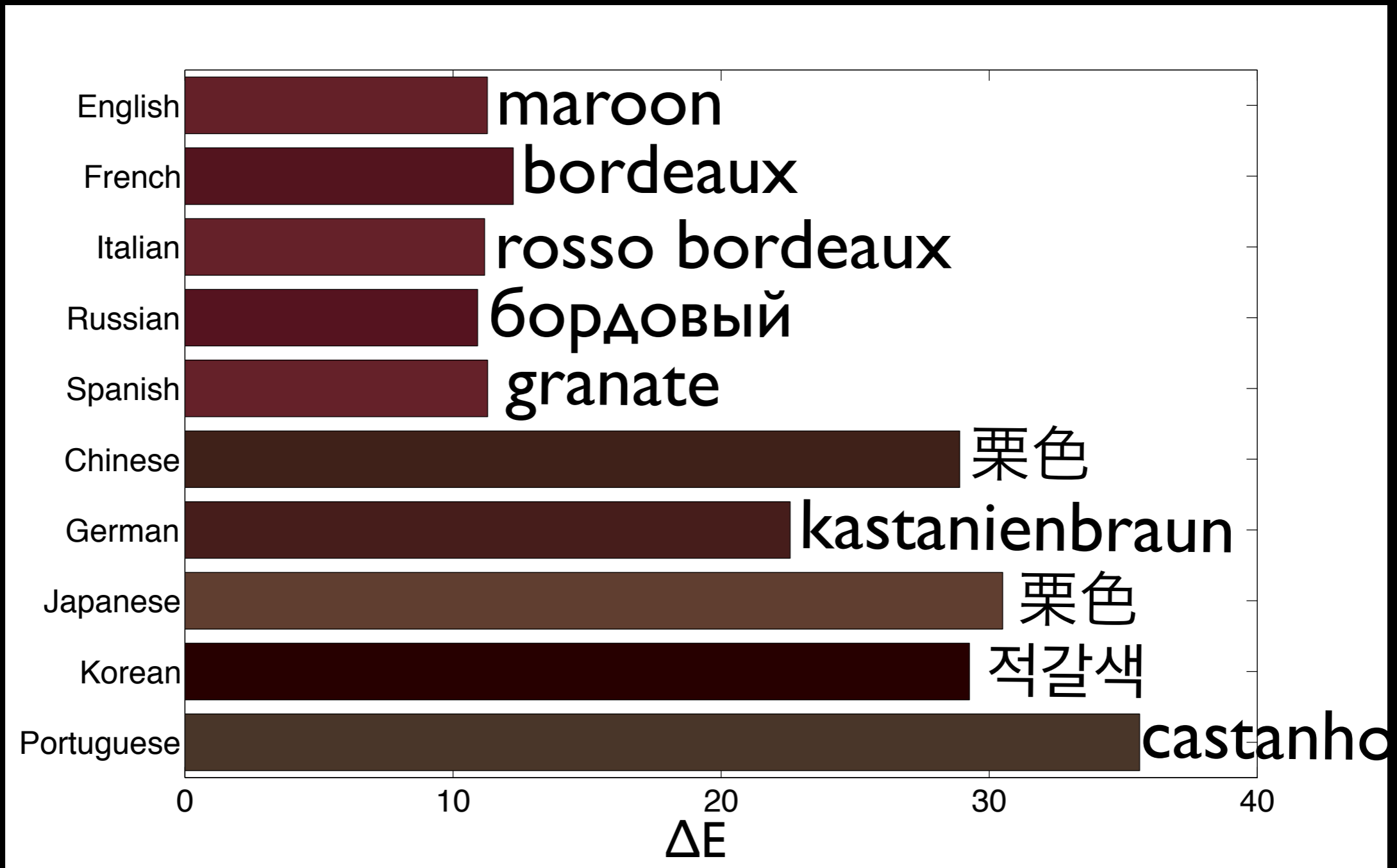
rosa suave, es

Color Estimations



Accuracy for *maroon*

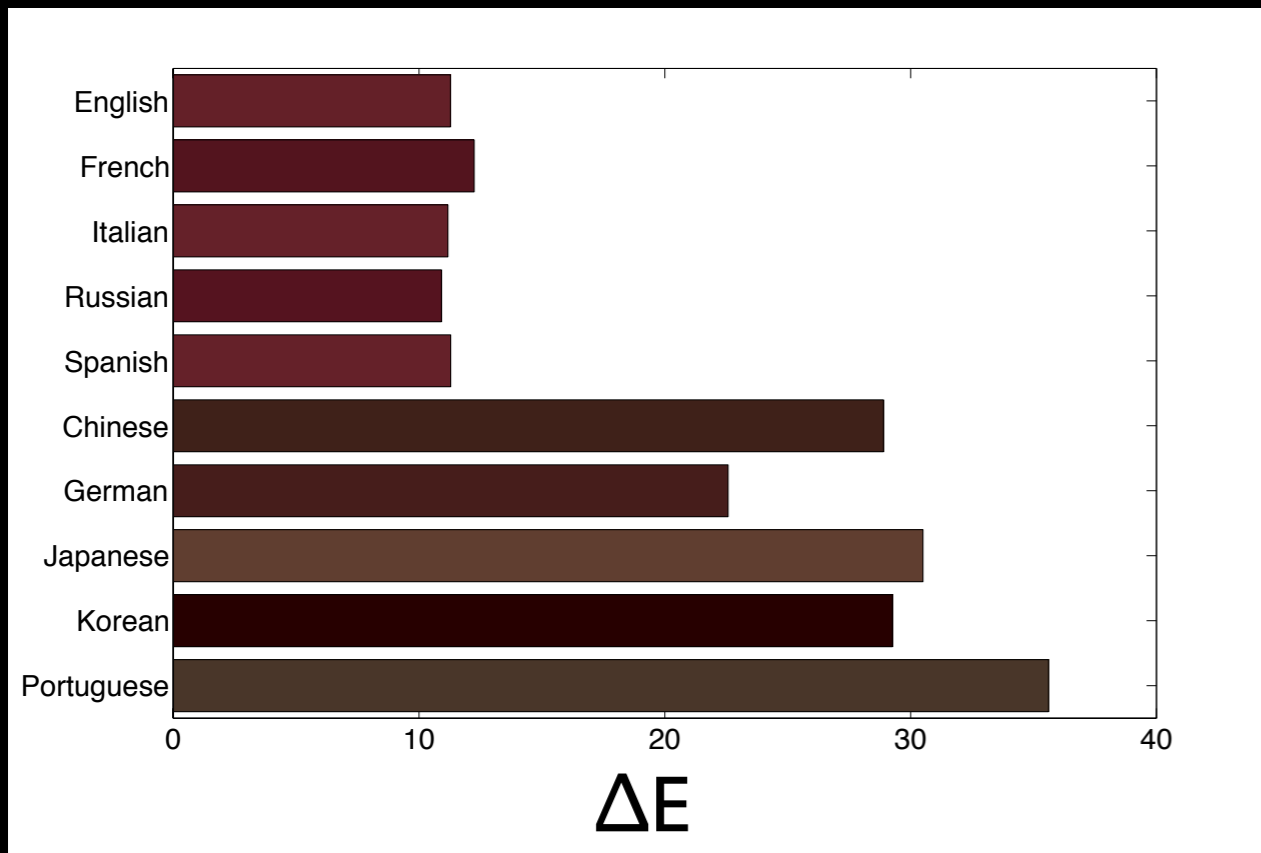
ΔE distances to **English** XKCD ground truth.



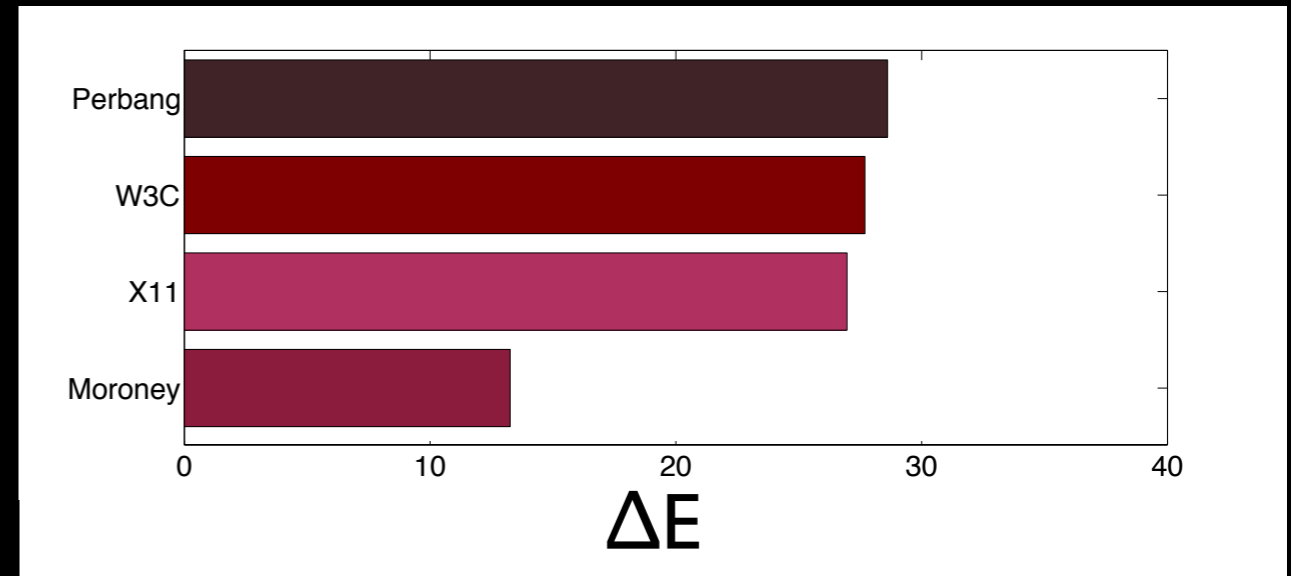
Accuracy for *maroon*

ΔE distances to **English** XKCD ground truth.

maroon, ours

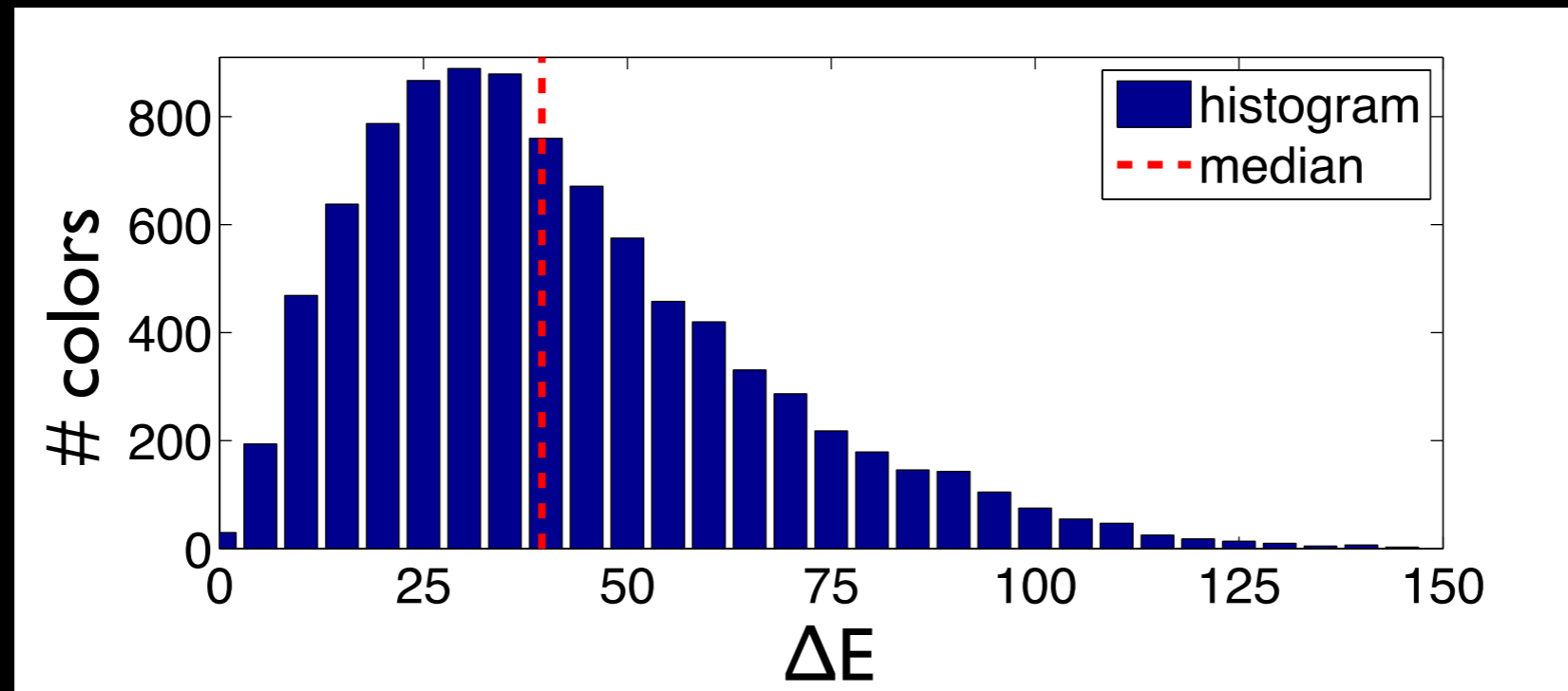


maroon, others



Accuracy

ΔE distances to **English** XKCD ground truth.



Our estimations are reasonably accurate considering:

- Disagreements between other databases.
- Language translations.

DEMO

Conclusions

- Statistical framework for automatic color value estimations.
- 9000+ color names in 10 languages.
- Interactive Color Thesaurus.
- <http://ivrg.epfl.ch/ColorThesaurus>
- International color **BYNGO** at CIC30.

Thank you
Q&A