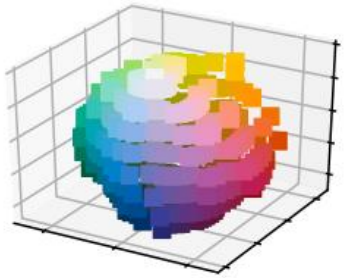
Visualizing the World Color Survey (WCS) Dataset

Language	Speaker	Focus	Term	Chip
1	1	1	LF	A0
1	1	2	WK	D9
1	2	6	WK	D10
...	...	...	...	...



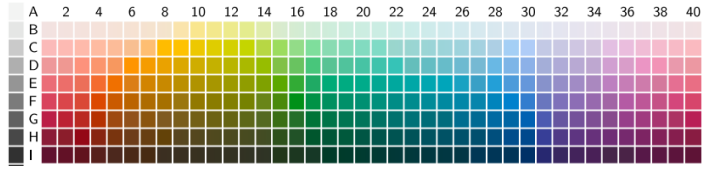
CPSC 547 Final Project
Yi (Joshua) Ren
renyi.joshua@gmail.com

Outlines:

- ***The WCS Project and Our Motivation***
- ***Tasks, Solutions and Vis Designs***
- ***Software Implementation***

The WCS Project and Our Motivation

- WCS Motivation: provide evidences for cognitive science, linguistics, etc.



- WCS task: ask participants (from 110 communities) to name the chips on Munsell card using their first language.

Participants

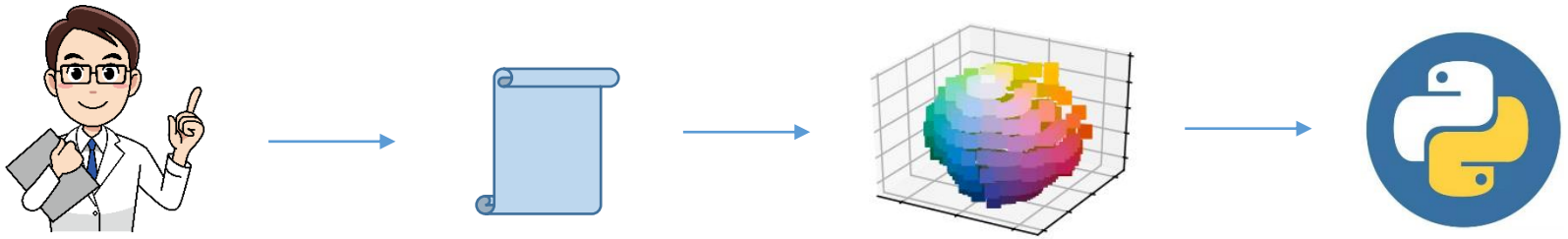
- WCS dataset: in text form, very **hard** to **preprocess** and **get intuition**.

is LF

is WK

Language	Speaker	Focus	Term	Chip
1	1	1	LF	A0
1	1	2	WK	D9
1	2	6	WK	D10
...	...	...	...	...

- Our work: a bunch of designs to help fieldworkers visualizing the WCS dataset.



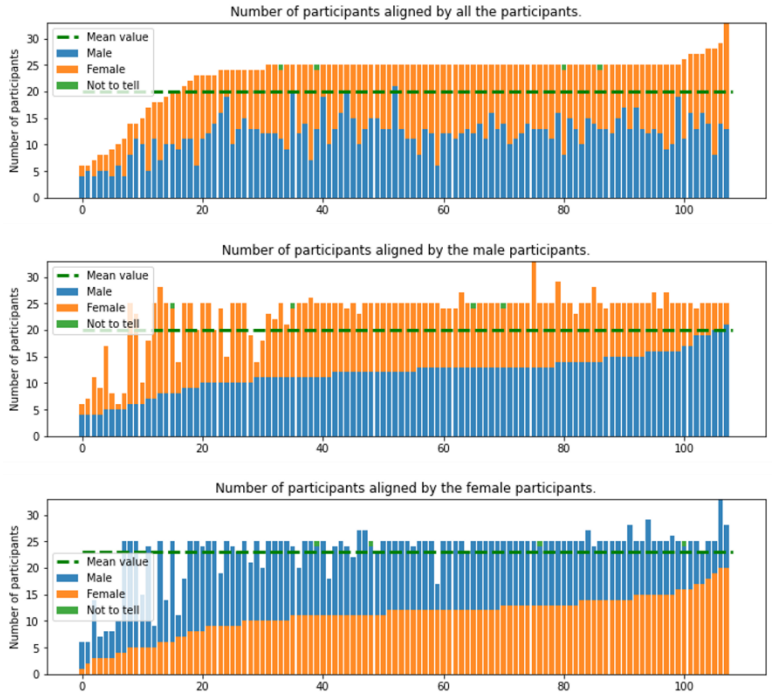
Interview → Tasks → Vis design → Implementation

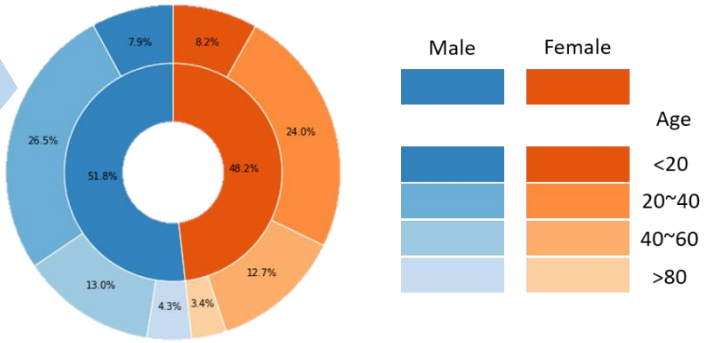
Tasks, Solutions and Vis Designs – Task 1

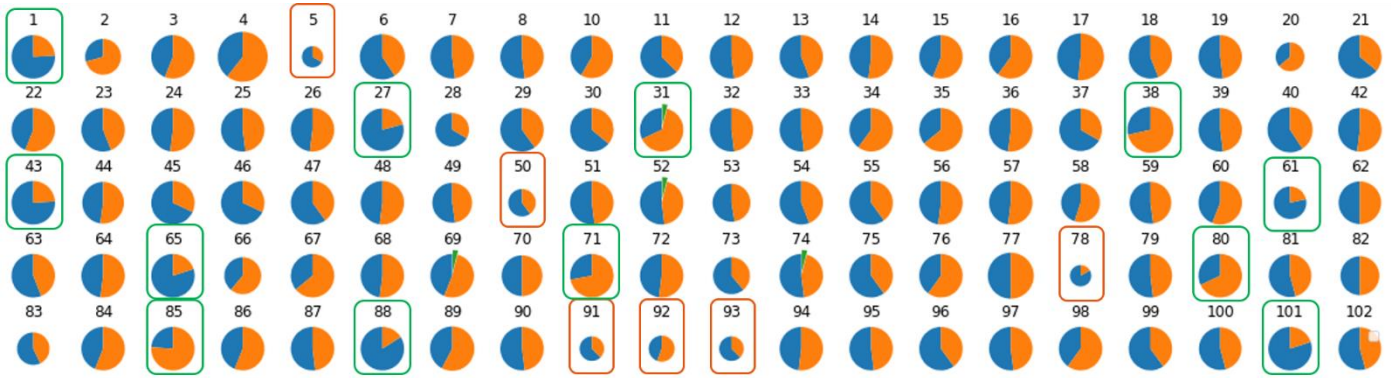
- What: **demographic** information (**gender** and **age**) of all the participants.

- Why: Have a overview of the distribution of the languages they focus.
- How: pie-ring chart design.

- Why: **Rule out** languages with too biased male-to-female ratio.
- How: stacked bar chart and aligned pie charts.





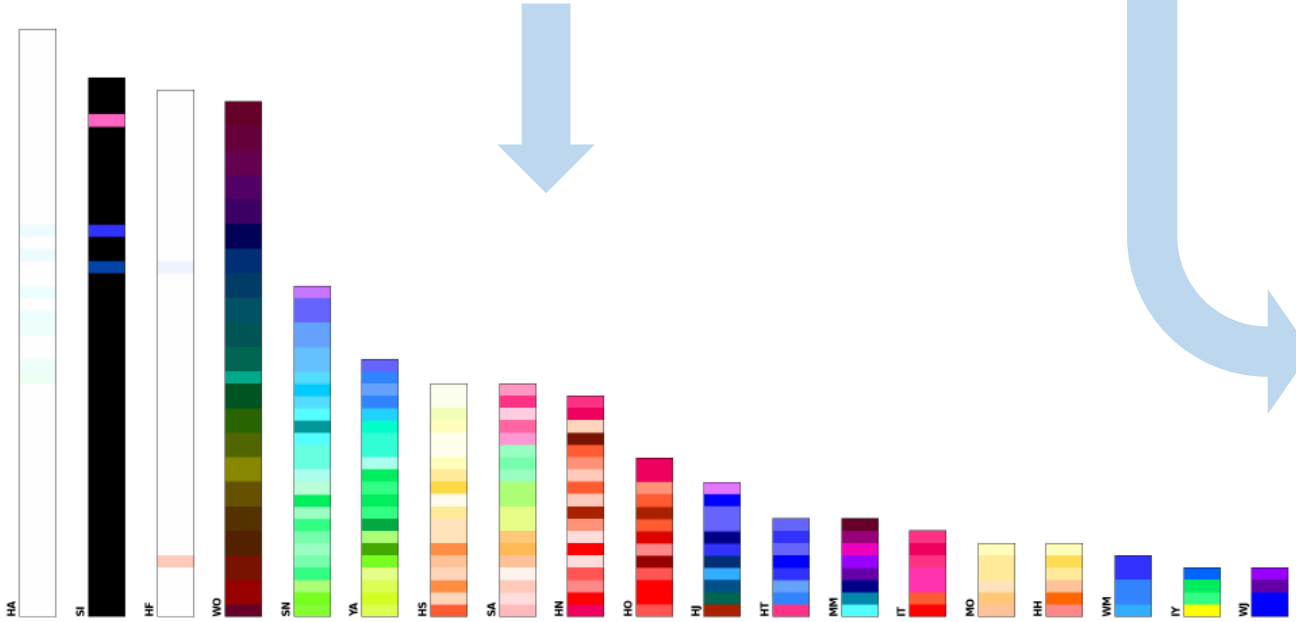


Tasks, Solutions and Vis Designs – Task 2

- What: naming results for each **single** language.

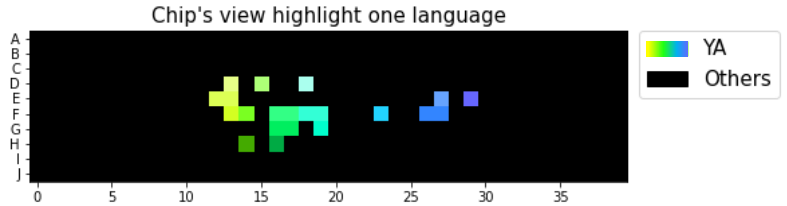
- Why: find language patterns, e.g. **structureness**.
- How: chip's view, show terms on Munsell grid.

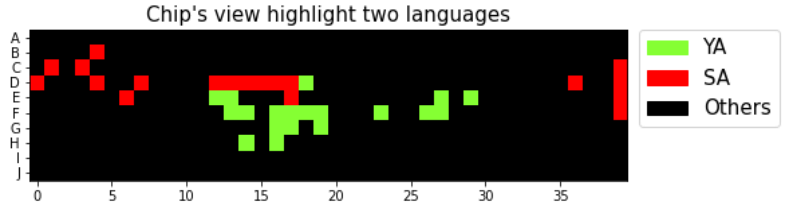
- Why: find term's **expressivity**.
- How: term's view, show chips in a bar.

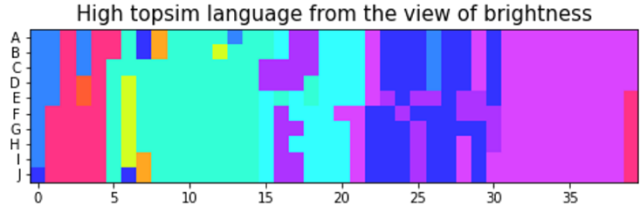


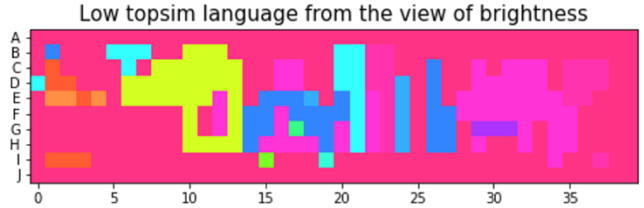
is LF is WK

Language	Speaker	Focus	Term	Chip
1	1	1	LF	A0
1	1	2	WK	D9
1	2	6	WK	D10
...	...	...	...	...







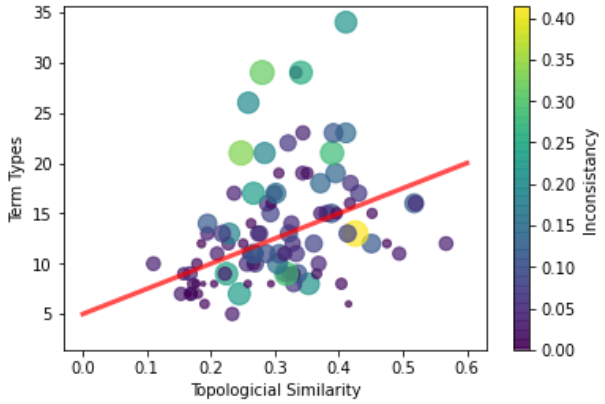


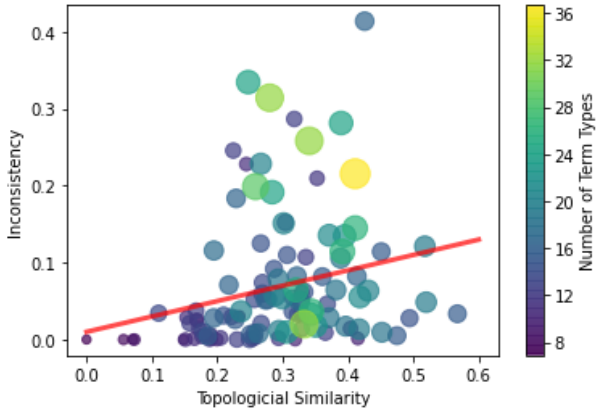
Tasks, Solutions and Vis Designs – Task 3

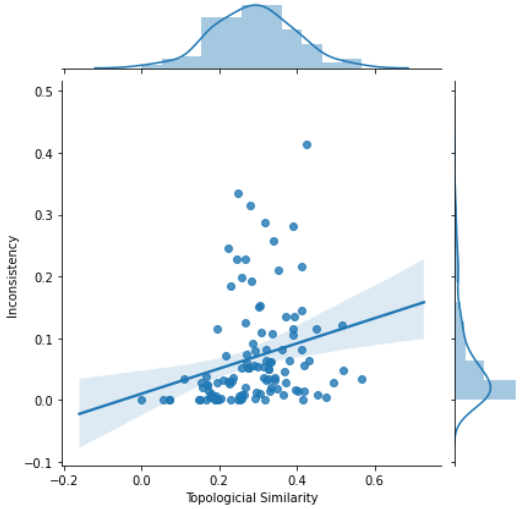
- What: derived **quantitative** metrics.

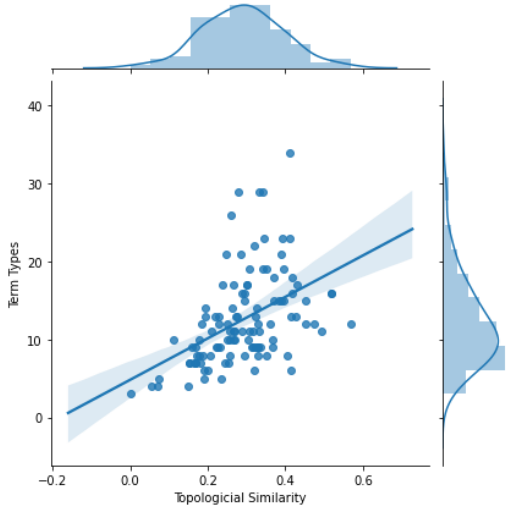
- Why: find correlations between **two** metrics.
- How: scatter plot + linear regression.

- Why: see **more** attributes in the same figure.
- How: use size and color as another channel.









Software Implementation

Python + Pandas + Matplotlib + Jupyter notebook

Munsell.py	Initial	Get Munsell color RGB code
README.md	first commit	
demographic_participants.ipynb	Initial	Code for task 1
foci-exp.txt	Initial	Raw data
lang_stats_yx.txt	Initial	Store the statistics of raw data
munsell.txt	Initial	Munsell color RGB code
quantitative_metric.ipynb	Initial	Code for task 3
single_language.ipynb	Initial	Code for task 2
spkr.txt	Initial	Raw data

https://github.com/Joshua-Ren/CPSC547_YIREN

Thanks for your attention