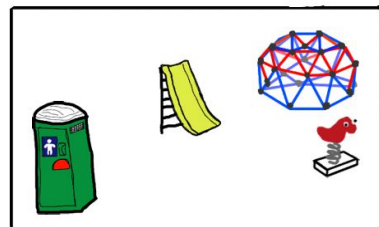
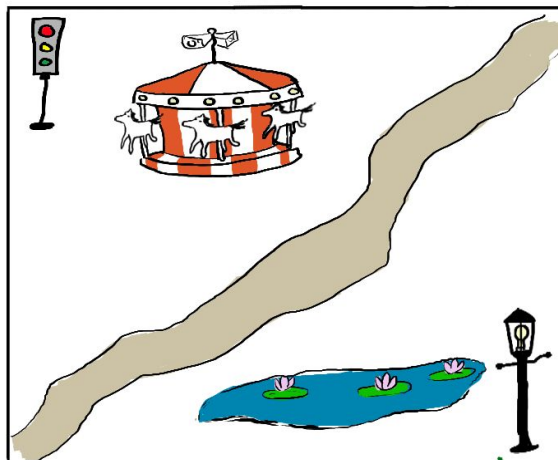
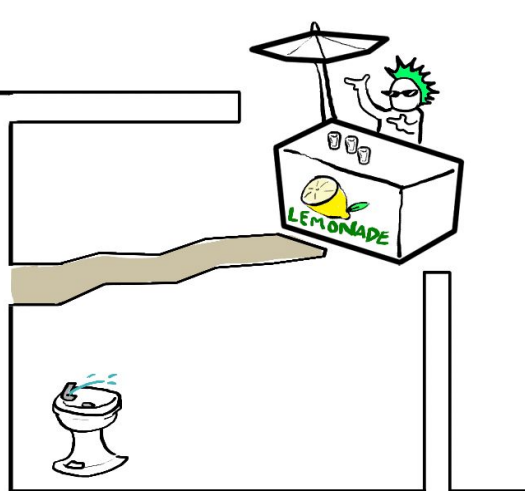


MapLemon: A Text Corpus for Demographic Identification through Writing Style



A project by Theodore Manning in collaboration
with the EVL Lab at Duquesne University with
funding from the CUNY Graduate Center Provost
Digital Initiatives Fund



Background

- Corpus in its infancy; currently in phase 2 with 21k+ words and 185 responses
- Originally created for stylometric demographic identification through language variation using naturally elicited, digital writing
- North American English
- Novel collection method: directions on map and recipe writing
- Participants recruited via Prolific
- Includes responses from 91 transgender and non-binary people

Current Results

1. Has successfully proven it can be used for demographic identification by identifying correctly a naturally occurring unknown nationality
2. Current data gathered from analysis using MapLemon in the Java Graphical Authorship Attribution Program suggests that:
 - a. Trans people write most like other trans people, and
 - b. They secondarily write most like their gender, rather than their sex assigned at birth

Closer to 0, the more similar, with 0 being the same.

1.

Comparison in JGAAP	Mystery Nationality
Canadian	1.25
American	1.5

2.

Comparison in JGAAP	Female to Male	Male to Female
Female to Male	--	1.125
Male to Female	1.125	--
Cisgender Assigned Female at Birth	1.5	1.25
Cisgender Assigned Male at Birth	1.25	1.5

Next Steps

- Gathering more responses from trans women and non-binary people assigned male at birth due to lack of representation in the corpus
- Analysis is being conducted to prevent the binarism of non-binary people within our results.