# Demo

Haocheng, Kewei, Zaolin, Chuhan

# Introduction

- blabla
- blabla
  - blabla

# Methodology

- General settings

| | |
|---|---|
| Latent Matches | $M_{ij} \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(\lambda)$ |
| Distance | $\gamma_k(i,j) \mid M_{ij} = m \overset{\text{indep.}}{\sim} \text{Discrete}(\pi_{km})$ |
| Missing Indicator | $\delta_k(i,j) \perp \gamma_k(i,j) \mid M_{ij}$ |
| Probability | $\xi_{ij} := \Pr(M_{ij} = 1 \mid \delta(i,j), \gamma(i,j))$ |

- Quite similar to Fellegi-Sunter[1]
- Capable pf dealing MAP missings

# Methodology

- Probabilistic model, see getPosterior.R

$$\xi_{ij} = \frac{\lambda \prod_{k=1}^{K} \left( \prod_{\ell=0}^{L_k - 1} \pi_{k\ell\ell}^{1\{\gamma_k(i,j)=\ell\}} \right)^{1-\delta_k(i,j)}}{\sum_{m=0}^{1} \lambda^m (1-\lambda)^{1-m} \prod_{k=1}^{K} \left( \prod_{\ell=0}^{L_k - 1} \pi_{km\ell}^{1\{\gamma_\ell(i,j)=\ell\}} \right)^{1-\delta_k(i,j)}}$$

- Nice for evaluation and post-merge analysis

$$X_i^* = \sum_{j=1}^{N_{\mathcal{B}}} \xi_{ij} X_j \Big/ \sum_{j=1}^{N_{\mathcal{B}}} \xi_{ij}$$

# Methodology

- Nice for calculation
  - Likelihood and boosting with EM steps[2]

$$\lambda = \frac{1}{N_A N_B} \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \xi_{ij}$$

$$\pi_{km\ell} = \frac{\sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \mathbf{1}\left\{\gamma_k(i,j) = l\right\} \left(1 - \delta_k(i,j)\right) \xi_{ij}^m \left(1 - \xi_{ij}\right)^{1-m}}{\sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \left(1 - \delta_k(i,j)\right) \xi_{ij}^m \left(1 - \xi_{ij}\right)^{1-m}}$$

  - logemlink.R and emlinkMARmov.R

# Methodology

# Dataset

# Package Implement

```
print(123)
```

```
## [1] 123
```

# References


[1] I. P. Fellegi and A. B. Sunter, "A theory for record linkage," *Journal of the American Statistical Association*, vol. 64, no. 328, pp. 1183–1210, 1969.

[2] W. E. Winkler, *Using the EM algorithm for weight computation in the fellegi-sunter model of record linkage*. US Bureau of the Census Washington, DC, 2000.