

d-wise

Empowering innovation for a healthier world



STRATEGY



DATA



CLOUD



PRIVACY



ANALYTICS

Machine learning in SAS

Cathal Gallagher
Jiangtang Hu



Machine Learning in SAS



Machine Learning in SAS

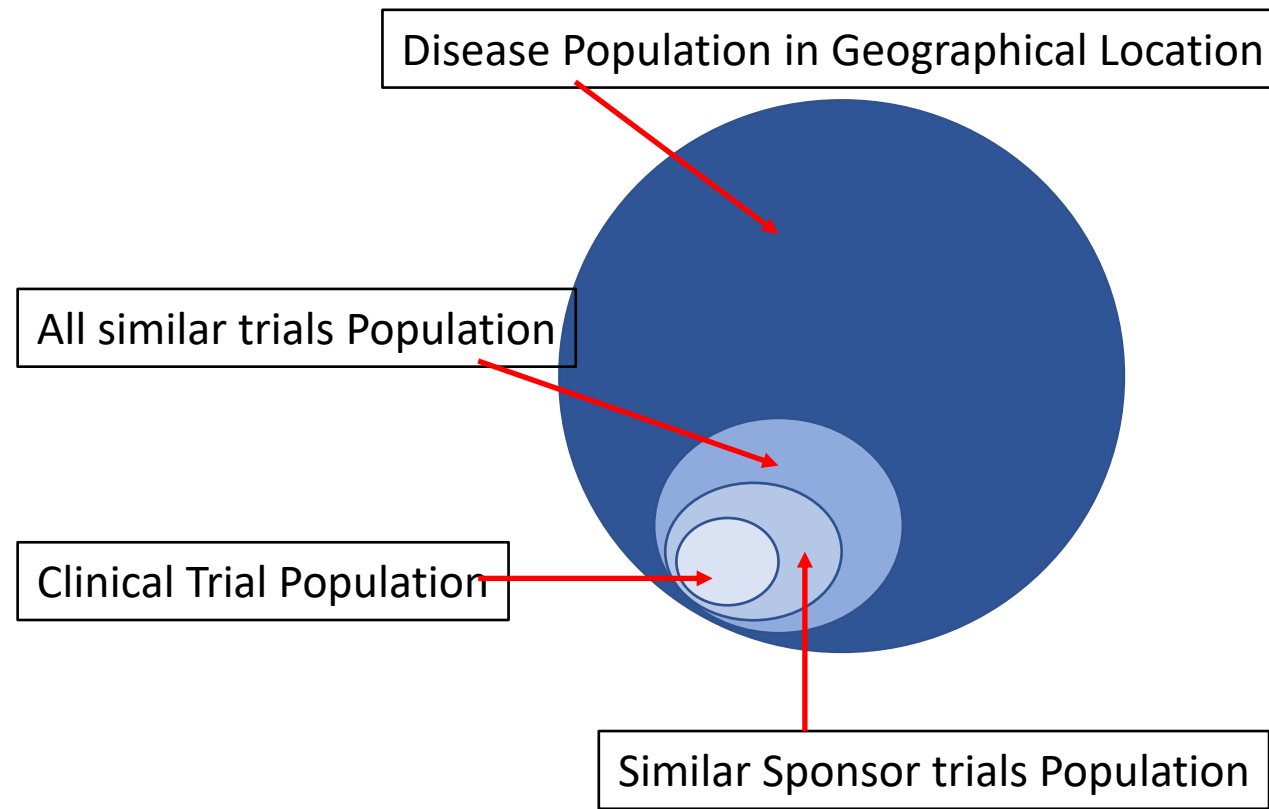
- Naïve Bayes Classifier Algorithm (Supervised Learning - Classification)
- K Means Clustering Algorithm (Unsupervised Learning - Clustering)
- Support Vector Machine Algorithm (Supervised Learning - Classification)
- Linear Regression (Supervised Learning/Regression)
- Logistic Regression (Supervised learning – Classification)
- Artificial Neural Networks (Reinforcement Learning)
- Decision Trees (Supervised Learning – Classification/Regression)
- Random Forests (Supervised Learning – Classification/Regression)
- Nearest Neighbours (Supervised Learning)



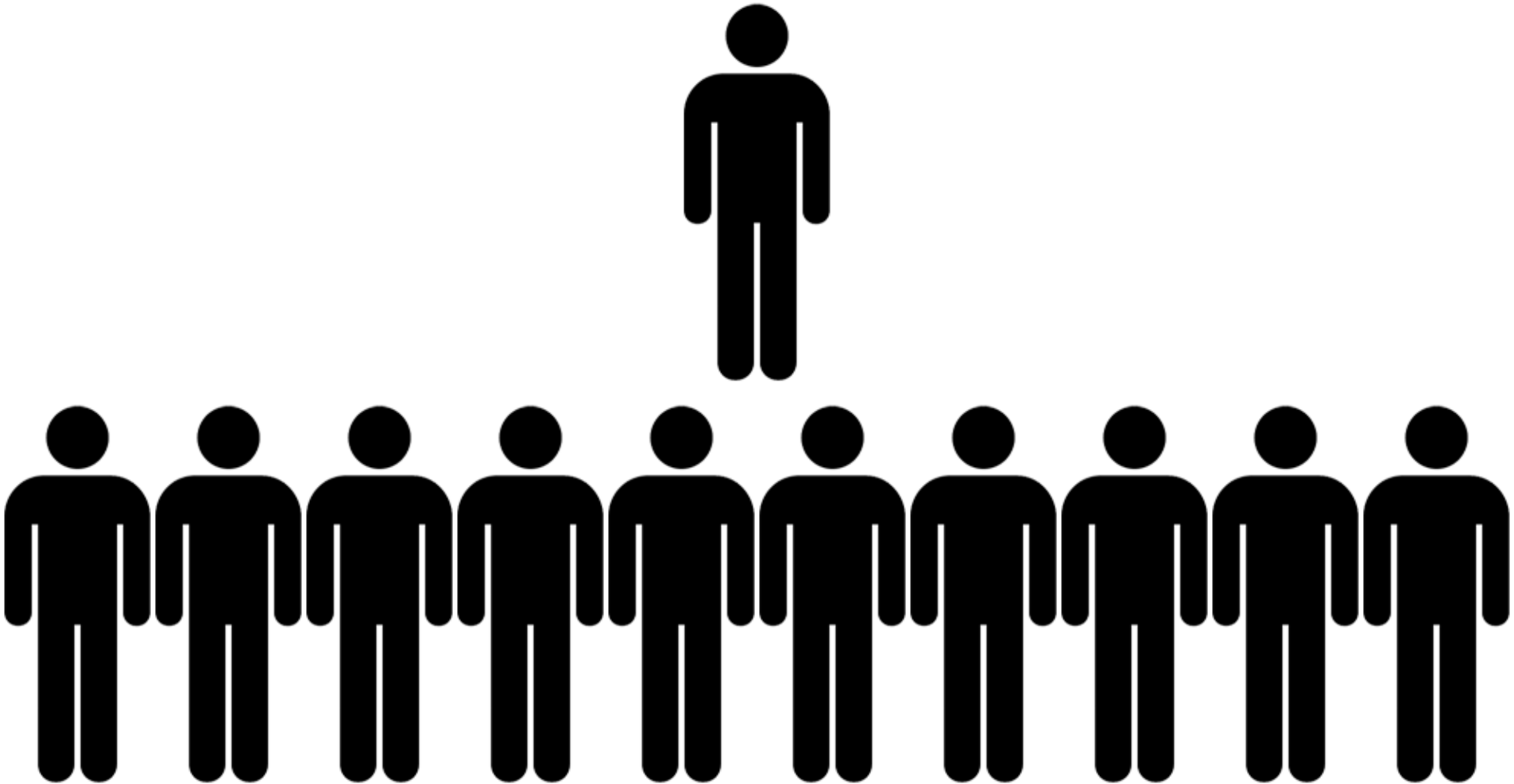
Machine Learning in SAS

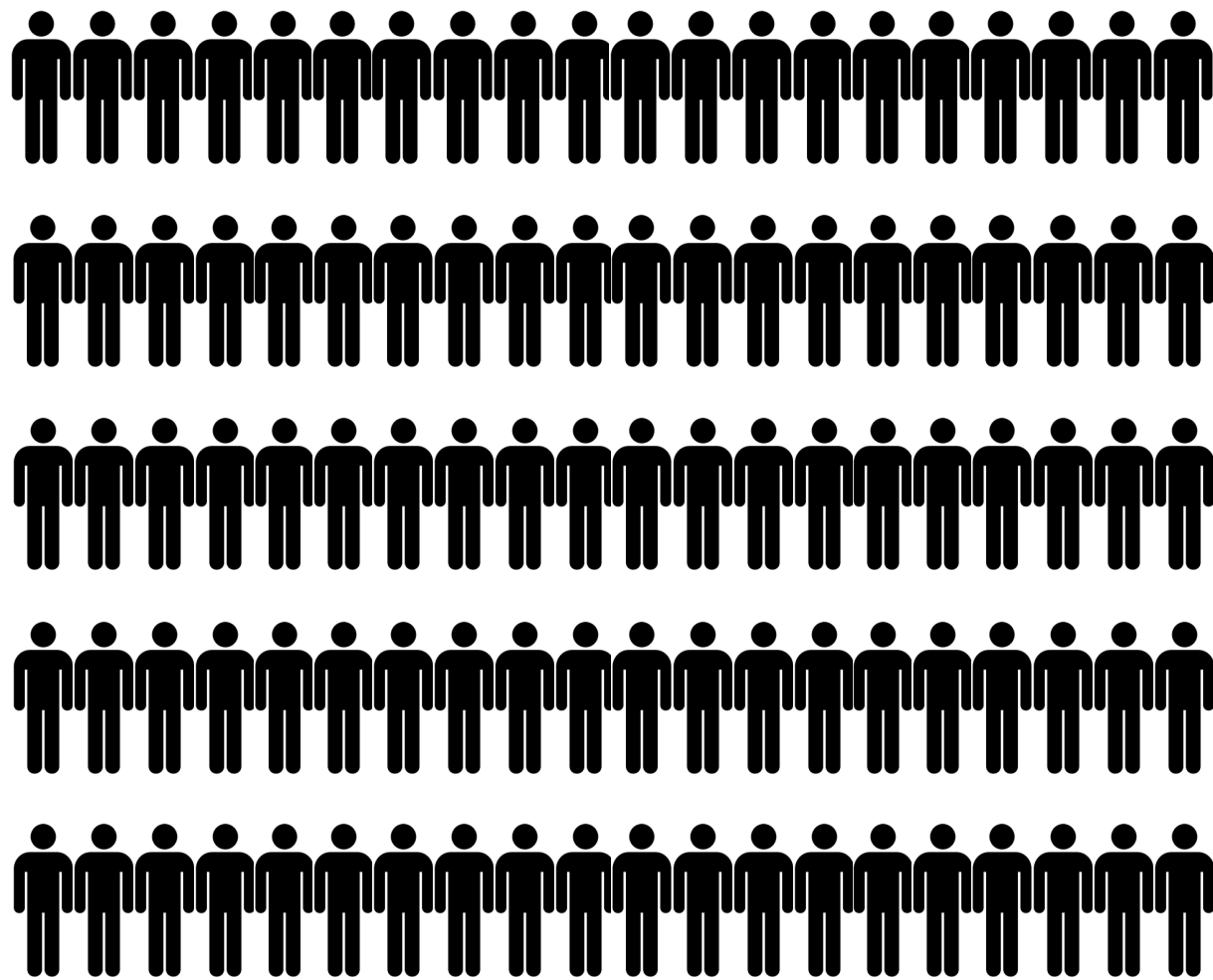
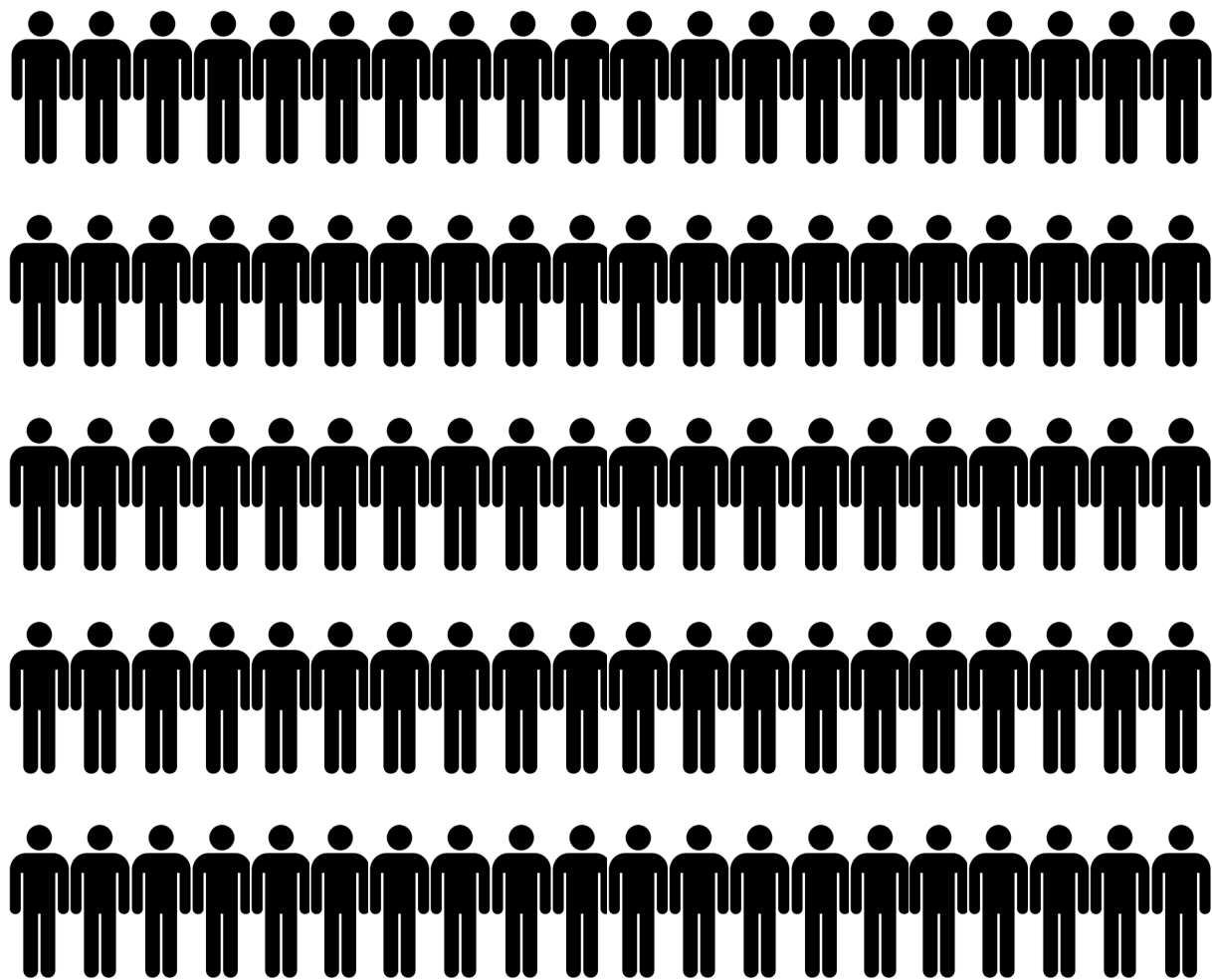
- Mapping of legacy data to CDISC standards
- fraud detection
- candidate selection
- information extraction from free text records
- Generating realistic “fake data”

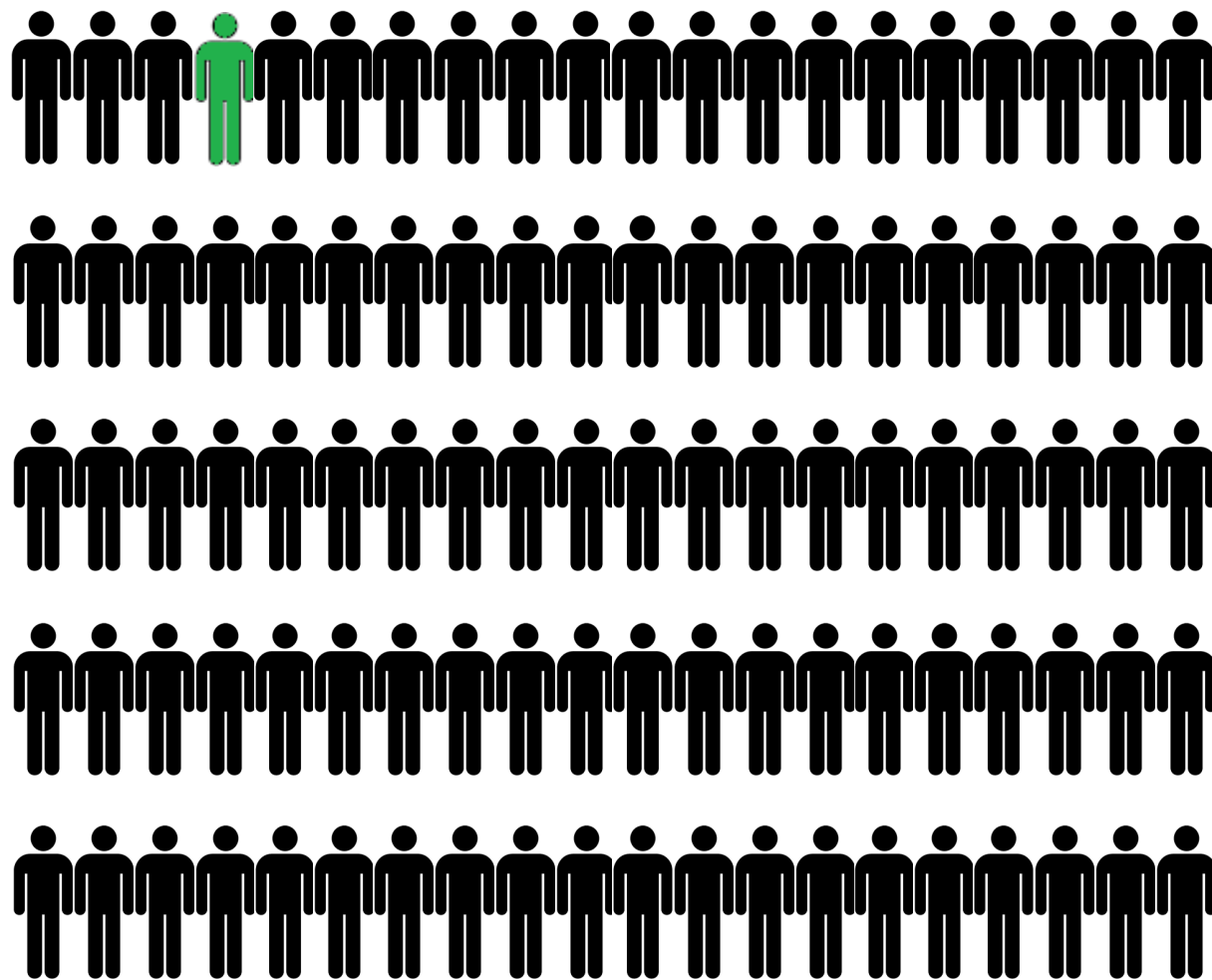
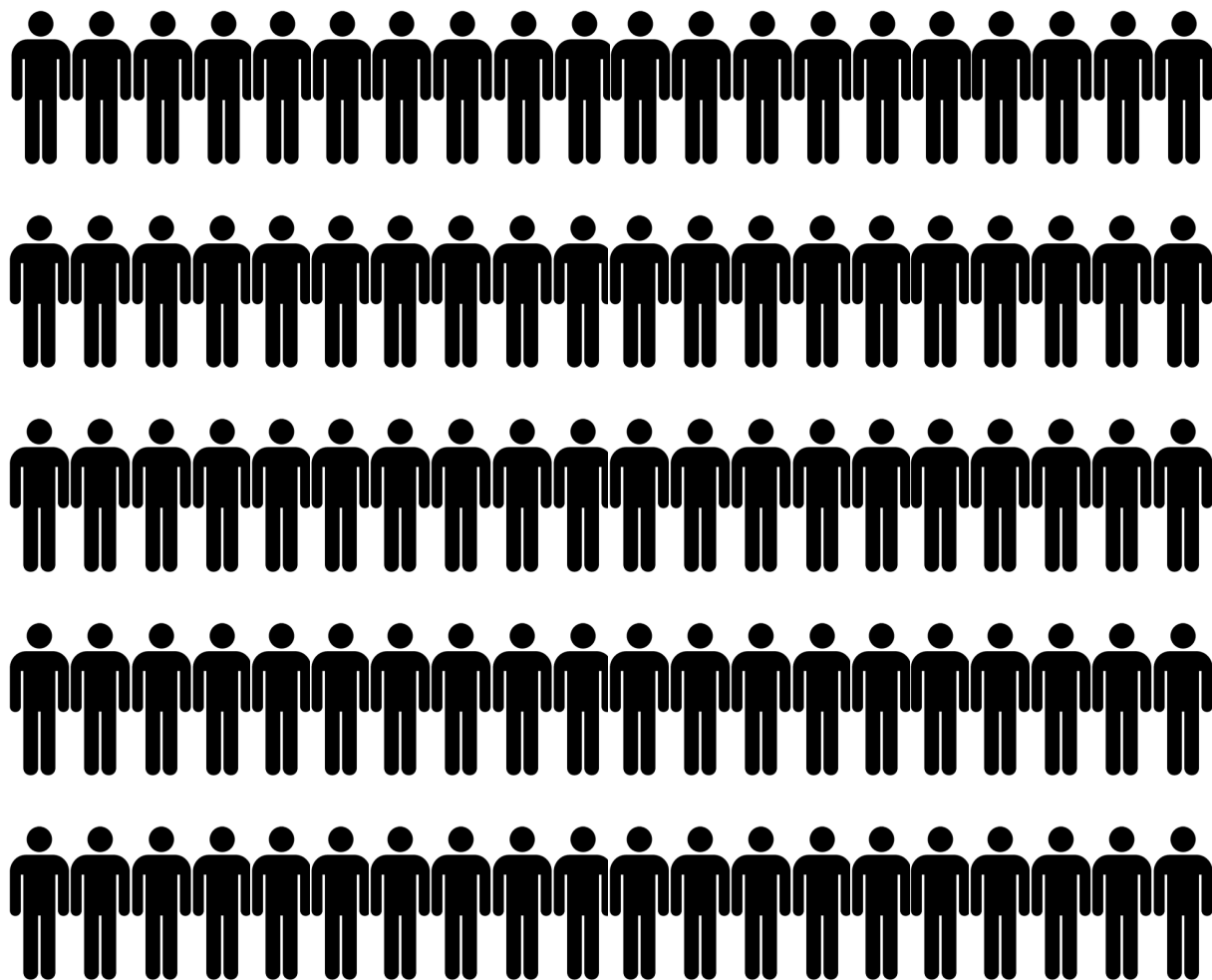
Machine Learning in SAS



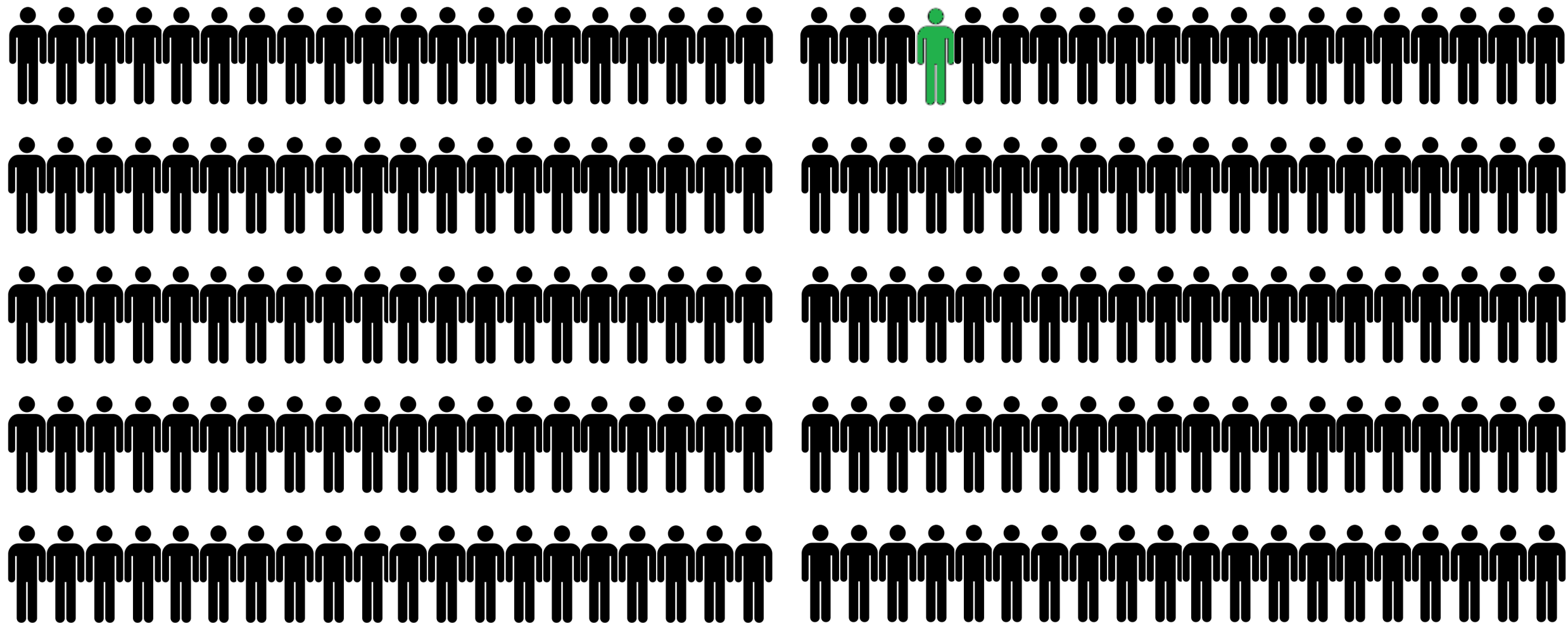
Risk Threshold 0.09



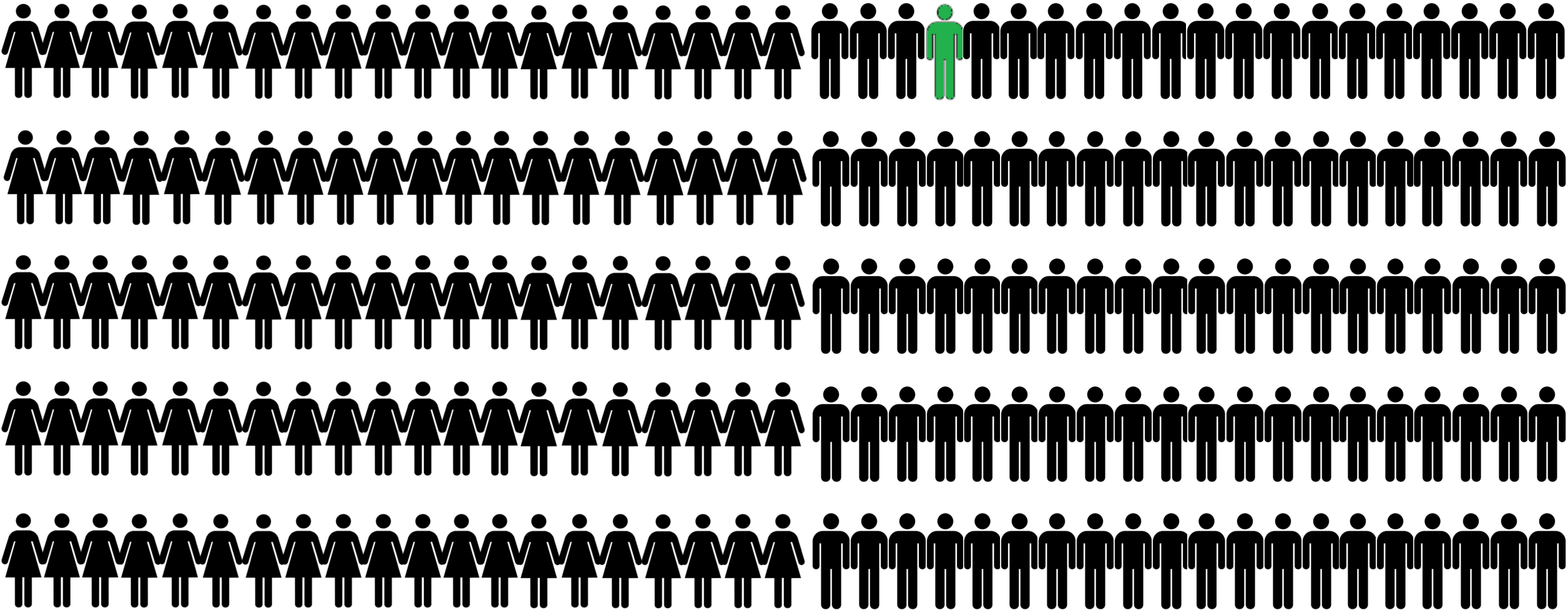




$$\text{POSA} = 1/200 = 0.005$$

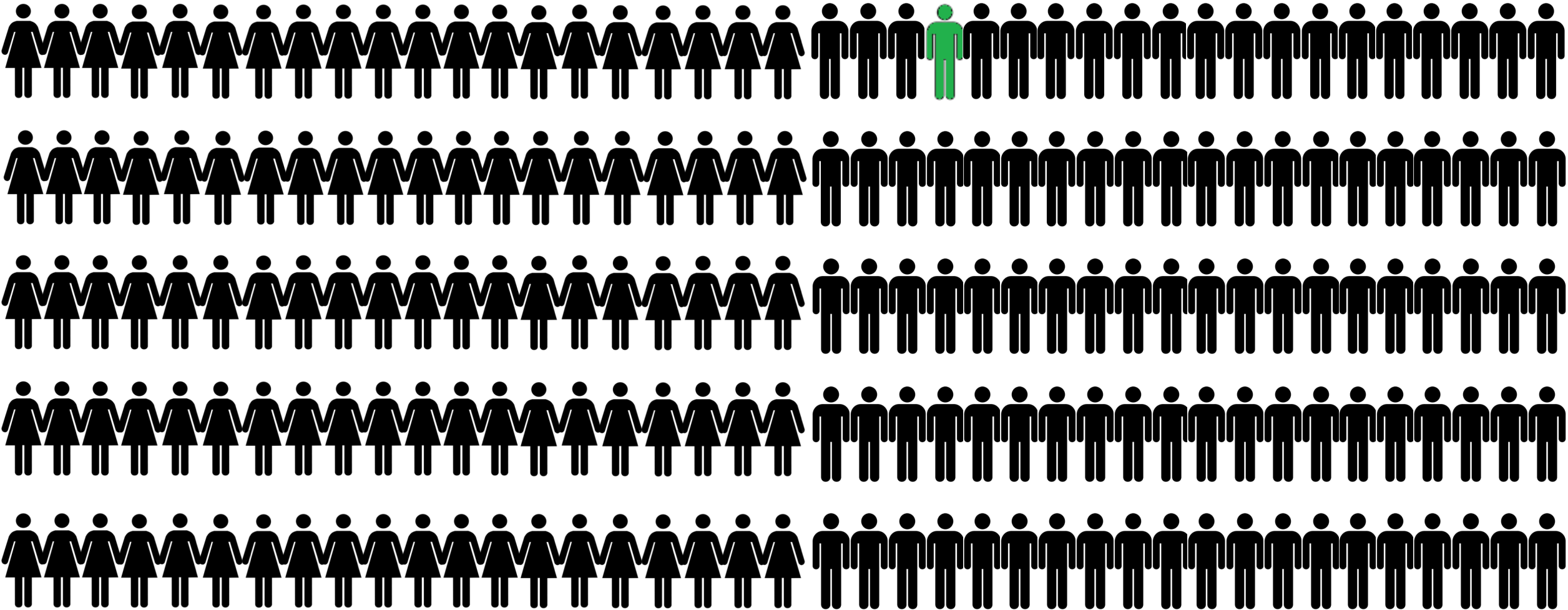


Gender



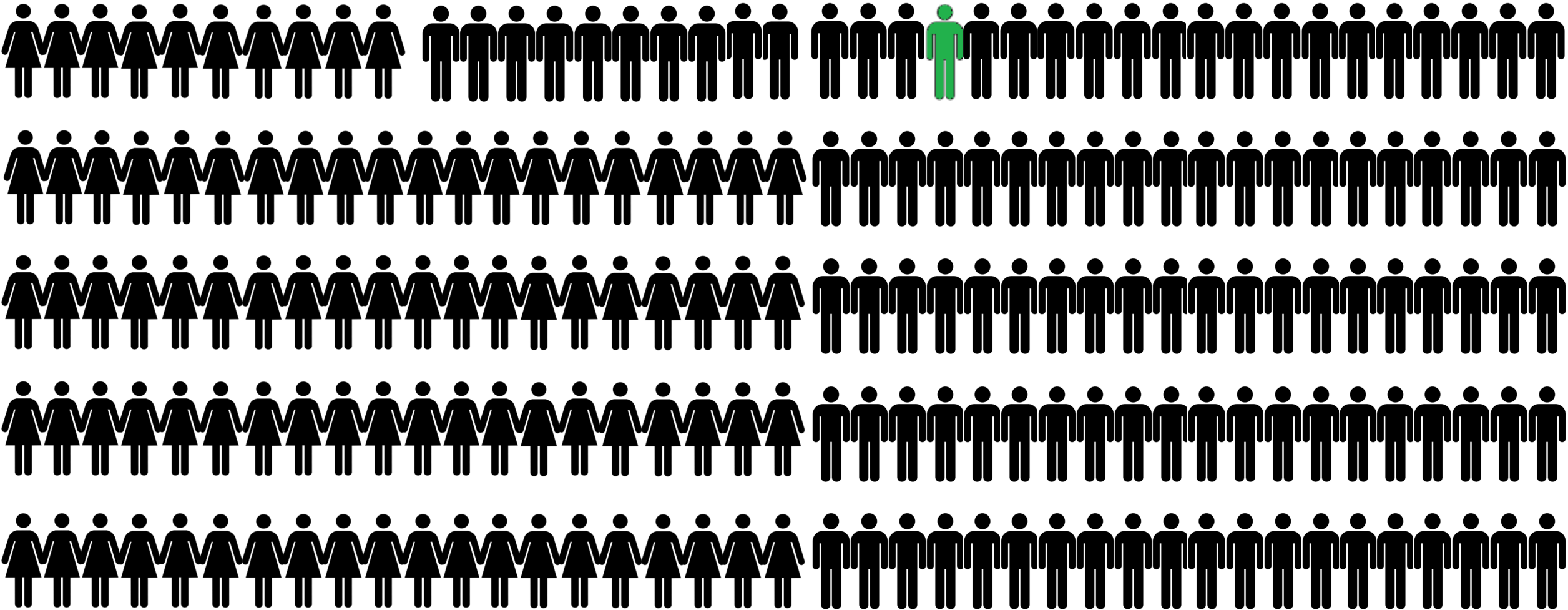
Gender

$$\text{POSA} = 1/100 = 0.010$$



Gender

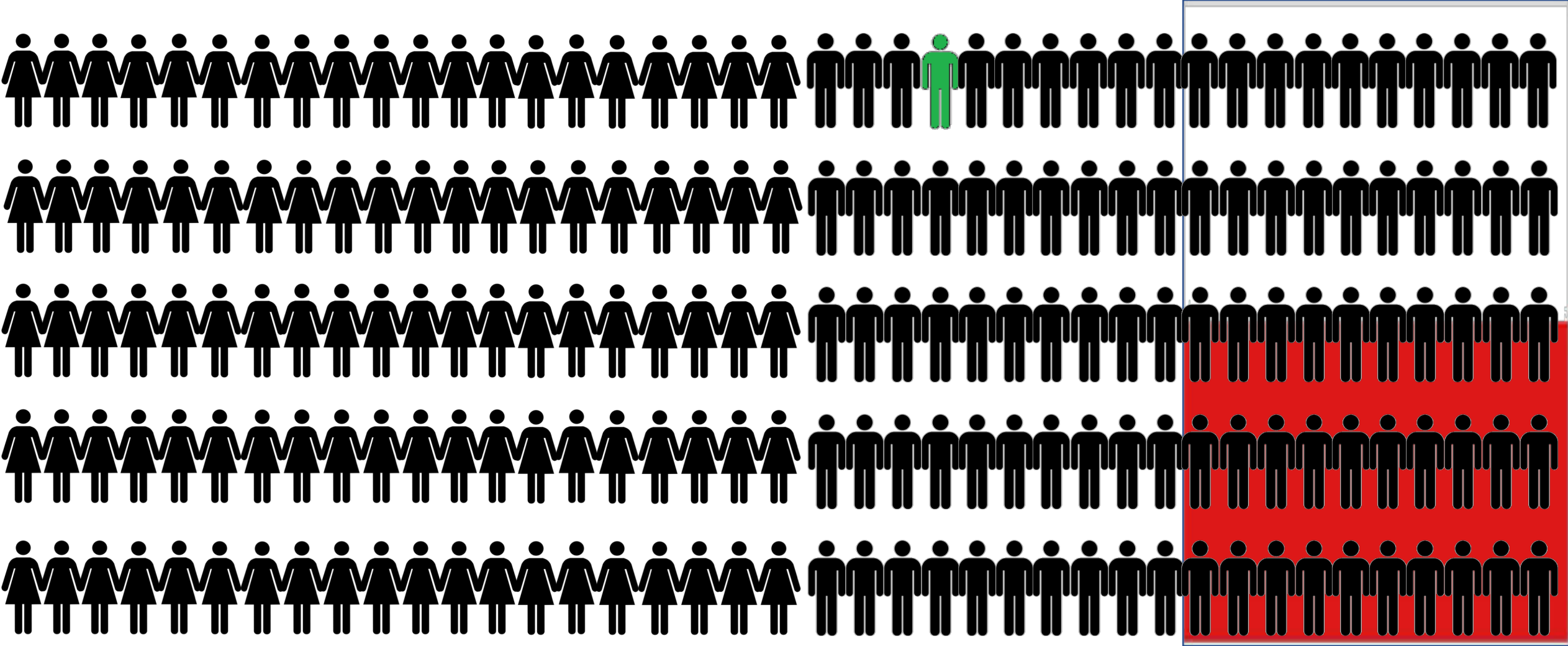
$$\text{POSA} = 1/90 = 0.011$$



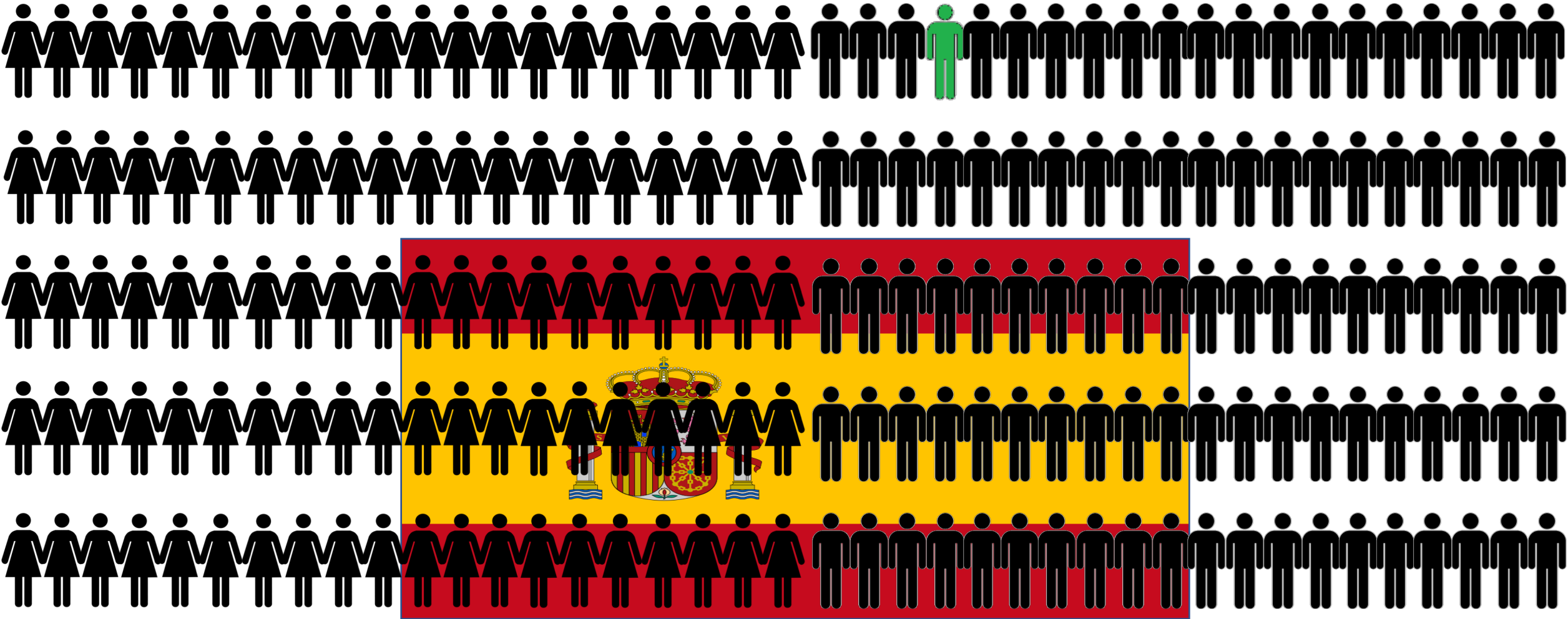
Gender
Country



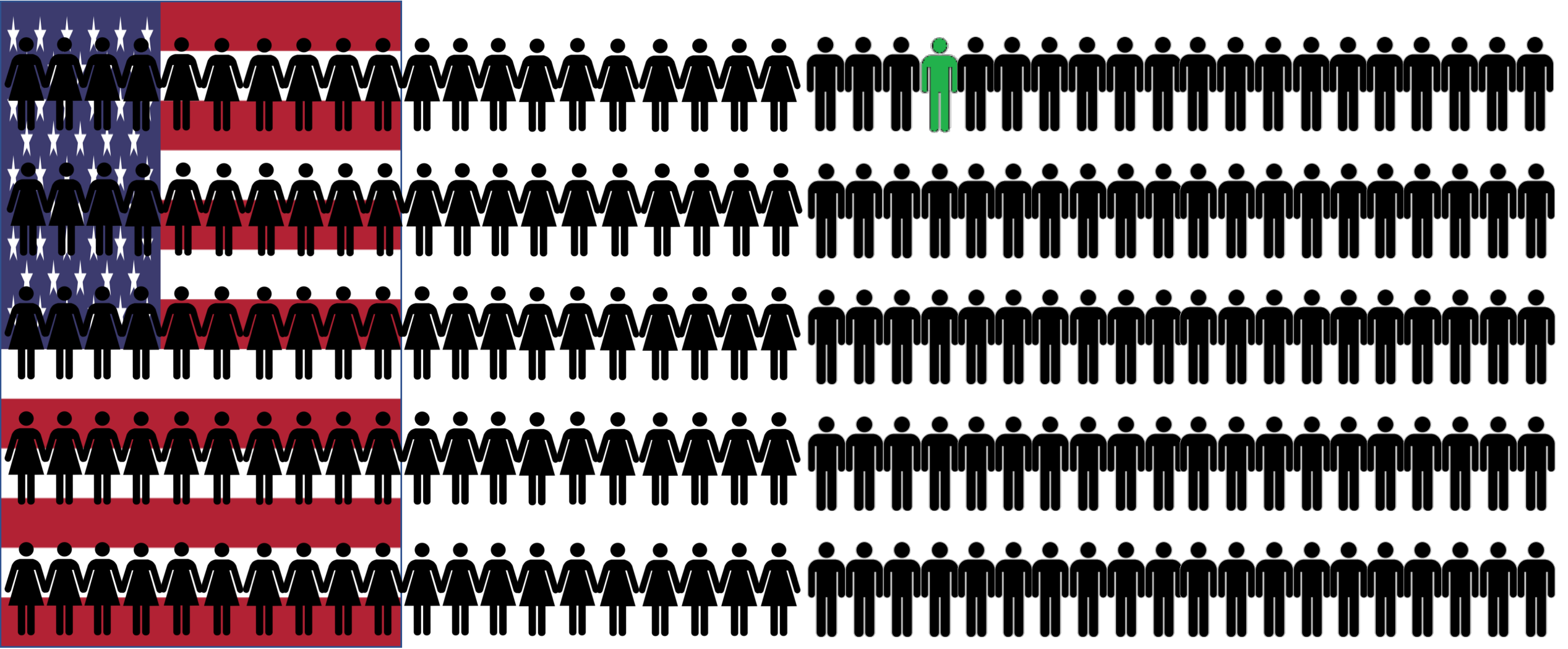
Gender
Country



Gender
Country

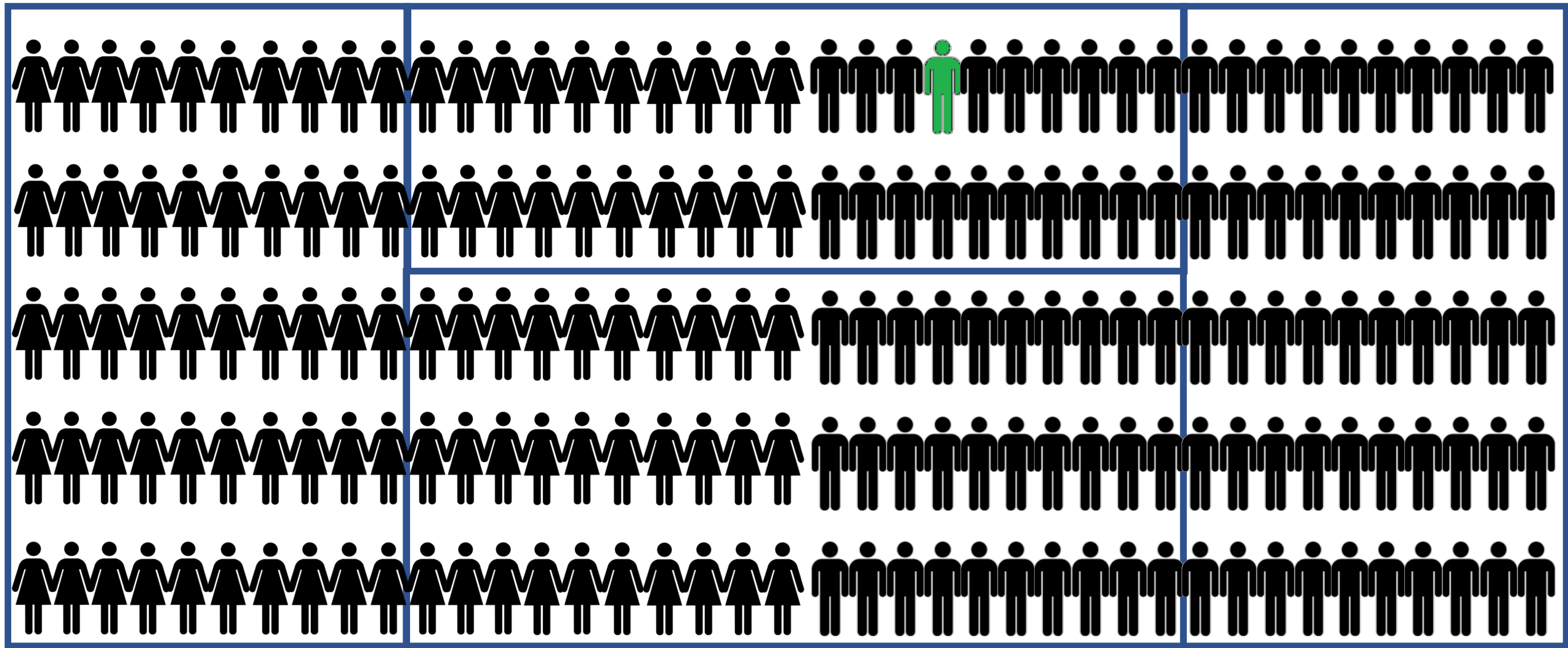


Gender
Country

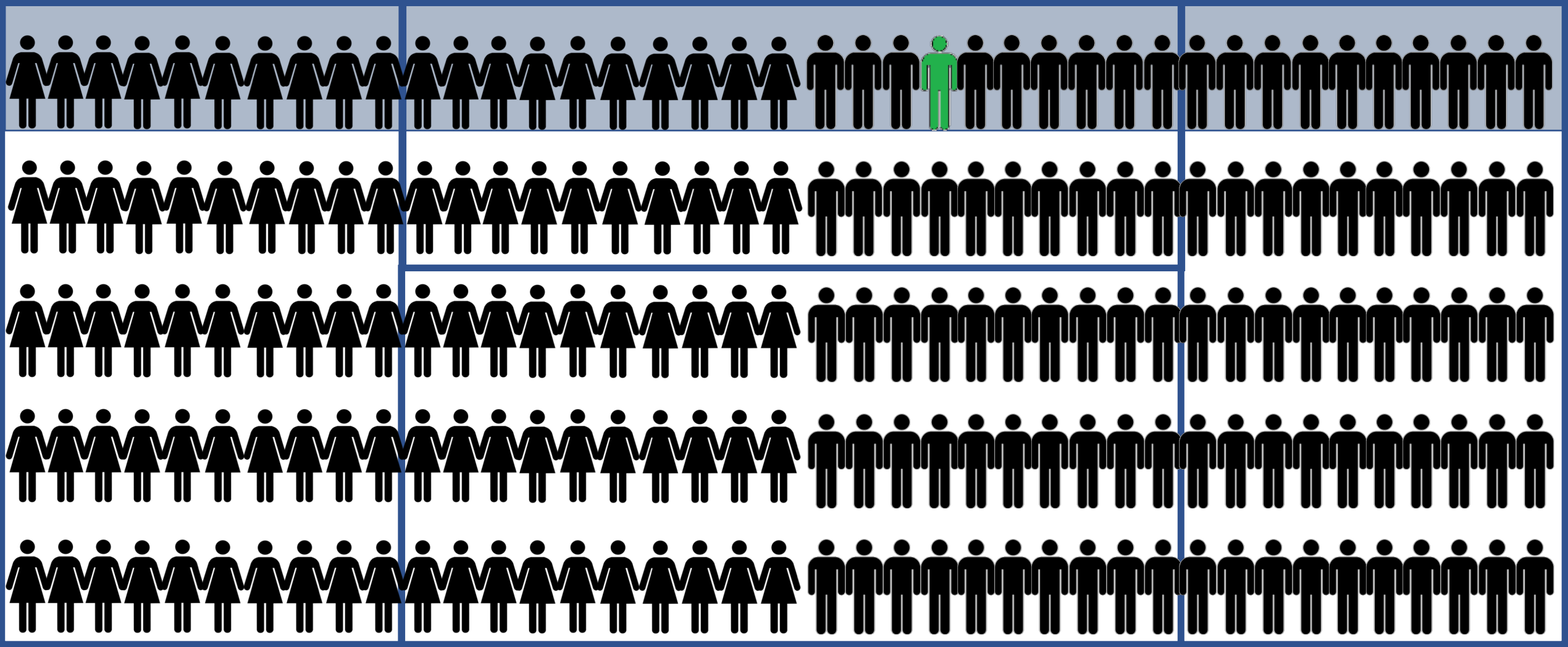


Gender
Country

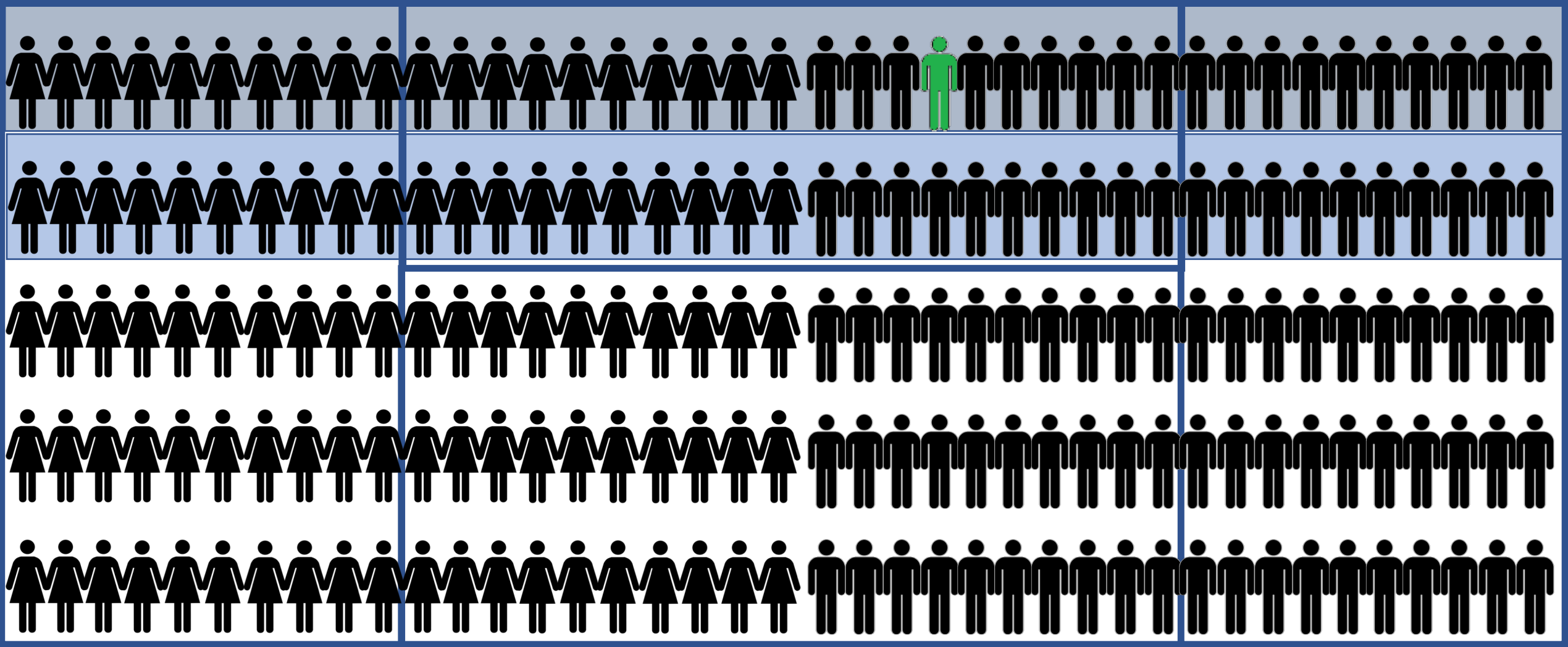
$POSA = 1/20 = 0.05$



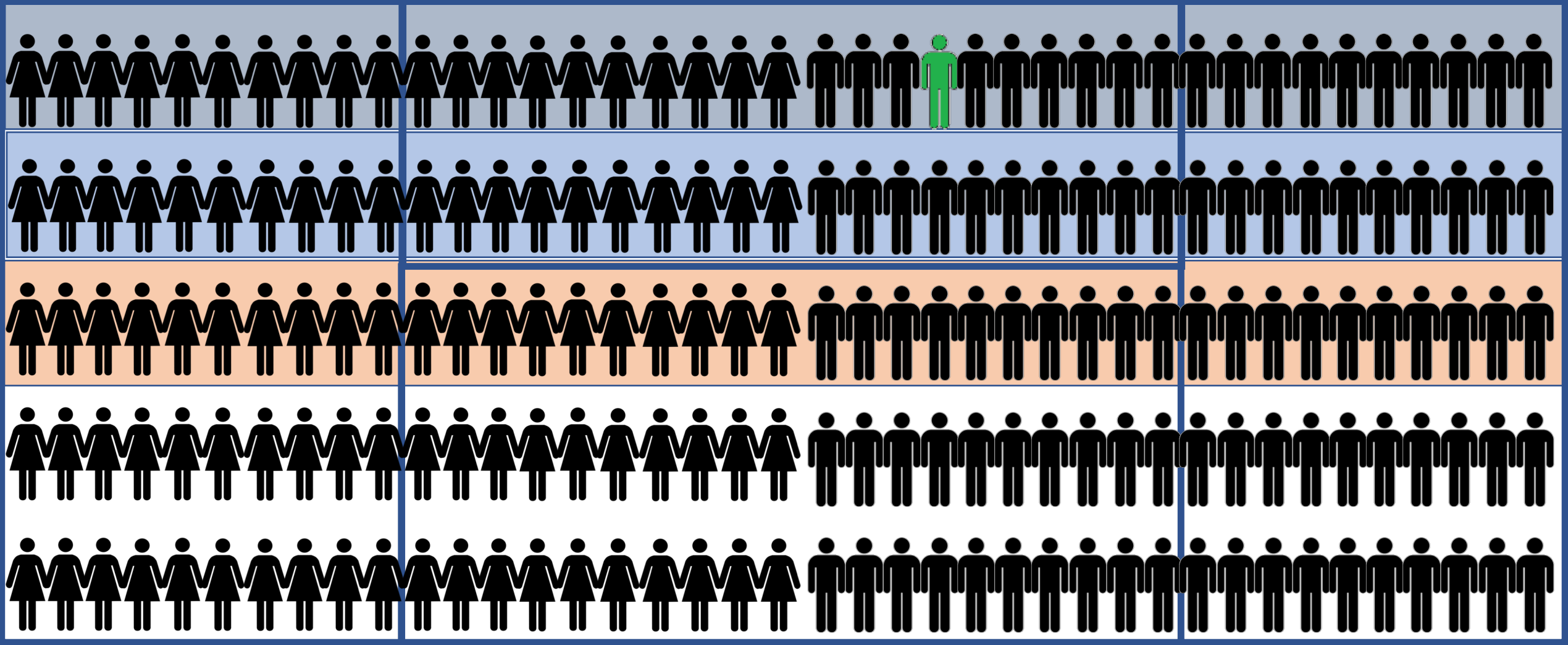
Gender
Country
Age 20 Year Bands



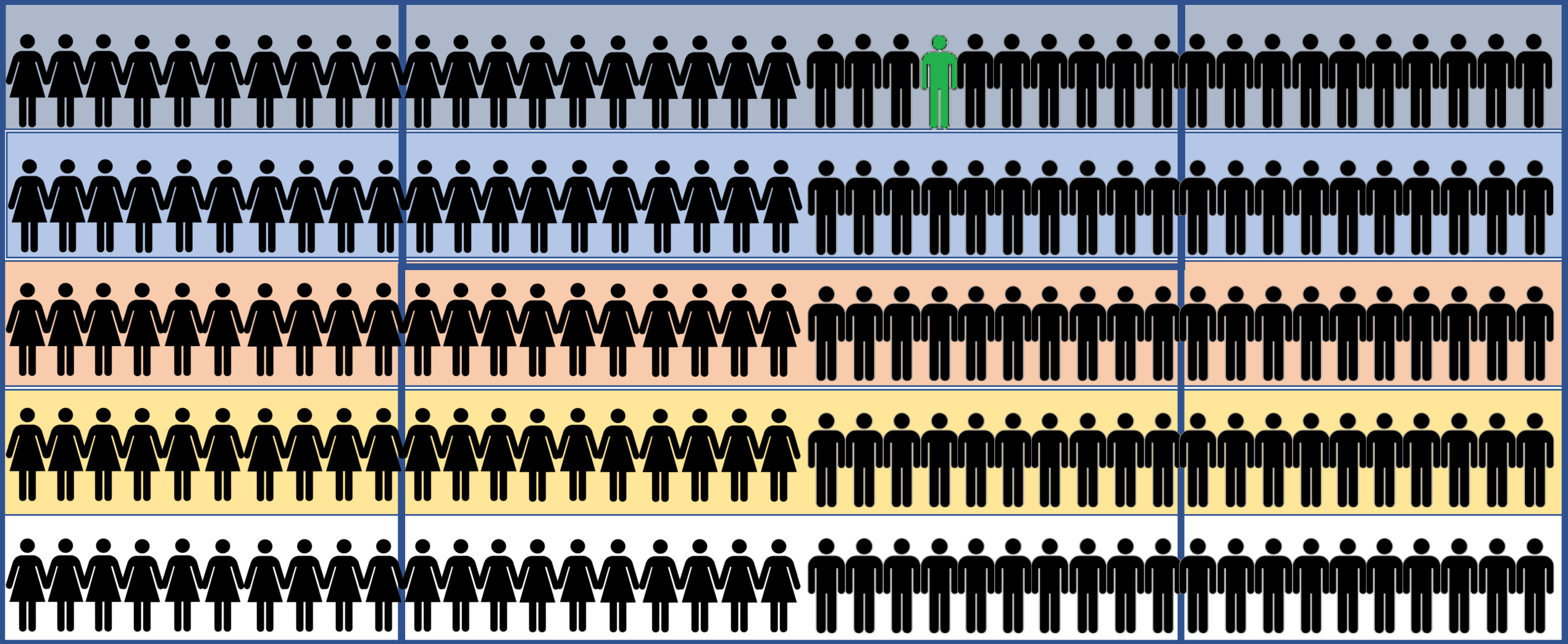
Gender
Country
Age 20 Year Bands



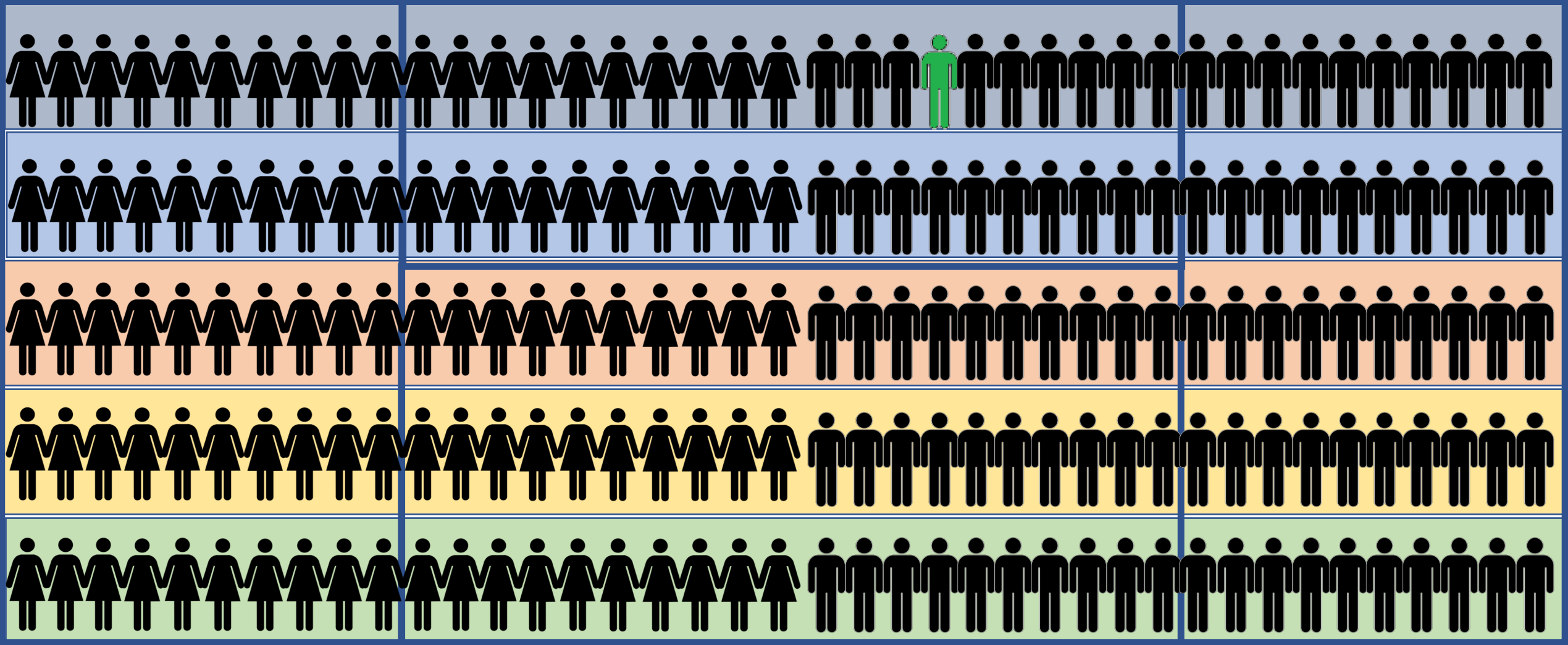
Gender
Country
Age 20 Year Bands



Gender
Country
Age 20 Year Bands



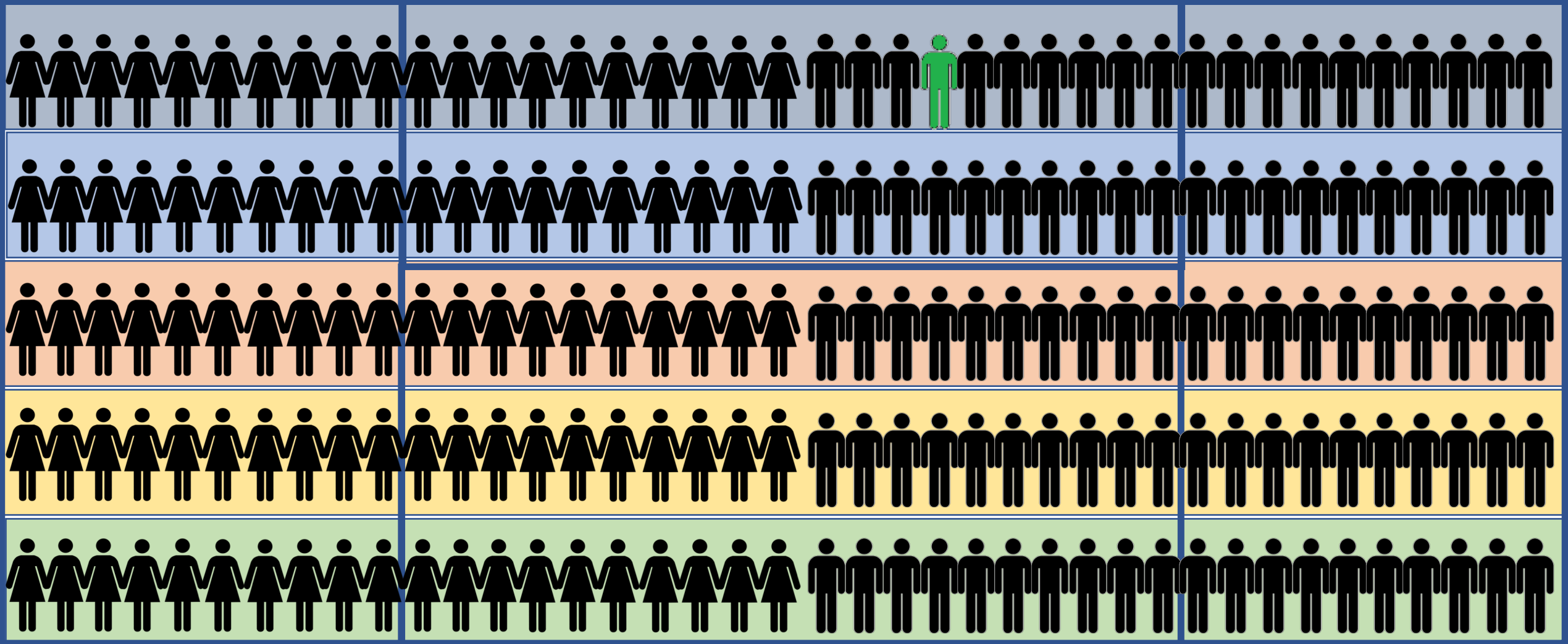
Gender
Country
Age 20 Year Bands



Gender
Country
Age 20 Year Bands

Risk Threshold 0.09

$POSA = 1/10 = 0.10$



Machine Learning in SAS

```
❏ proc surveyselect data=dm2  
    method=srs n=1000 out=dm_train;  
run;
```

```
❏ data dm3;  
    set dm2;  
    drop race;  
run;
```

Machine Learning in SAS

```
❏ proc discrim data = dm_train  
    test=dm3  
    testout=dm_out  
    method=npair k=10  
    testlist;  
  
    class race;  
    var sexn height weight bmi age;  
run;
```



Machine Learning in SAS

- Generation of realistic fake data is possible
- Nearest neighbour attacks on anonymous data is possible
- Quality of training data is key
- Most clinical programmers are familiar with SAS
- Phuse group on ML