

# Ancestry

Deepika Dokuru

1. Examine the dataset:

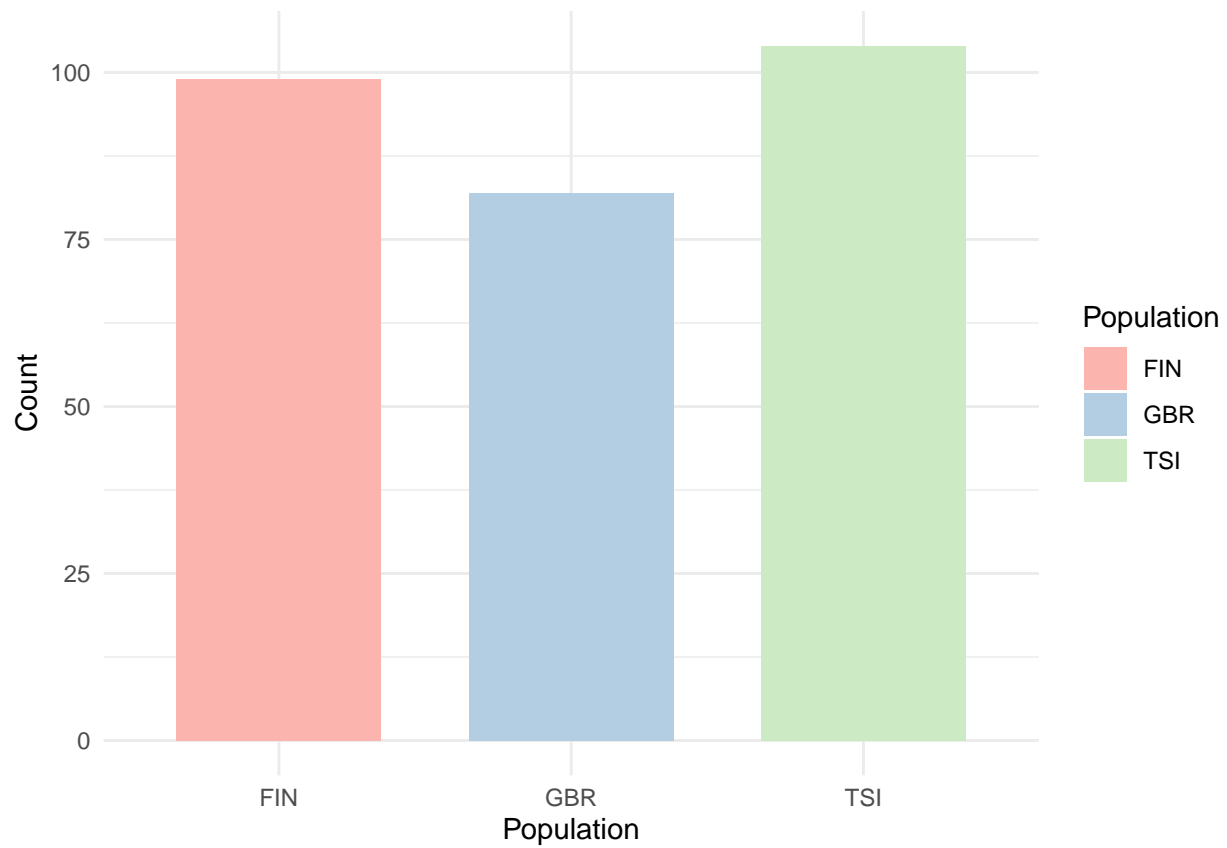
- How many samples are present?

```
## [1] 285
```

- How many SNPs?

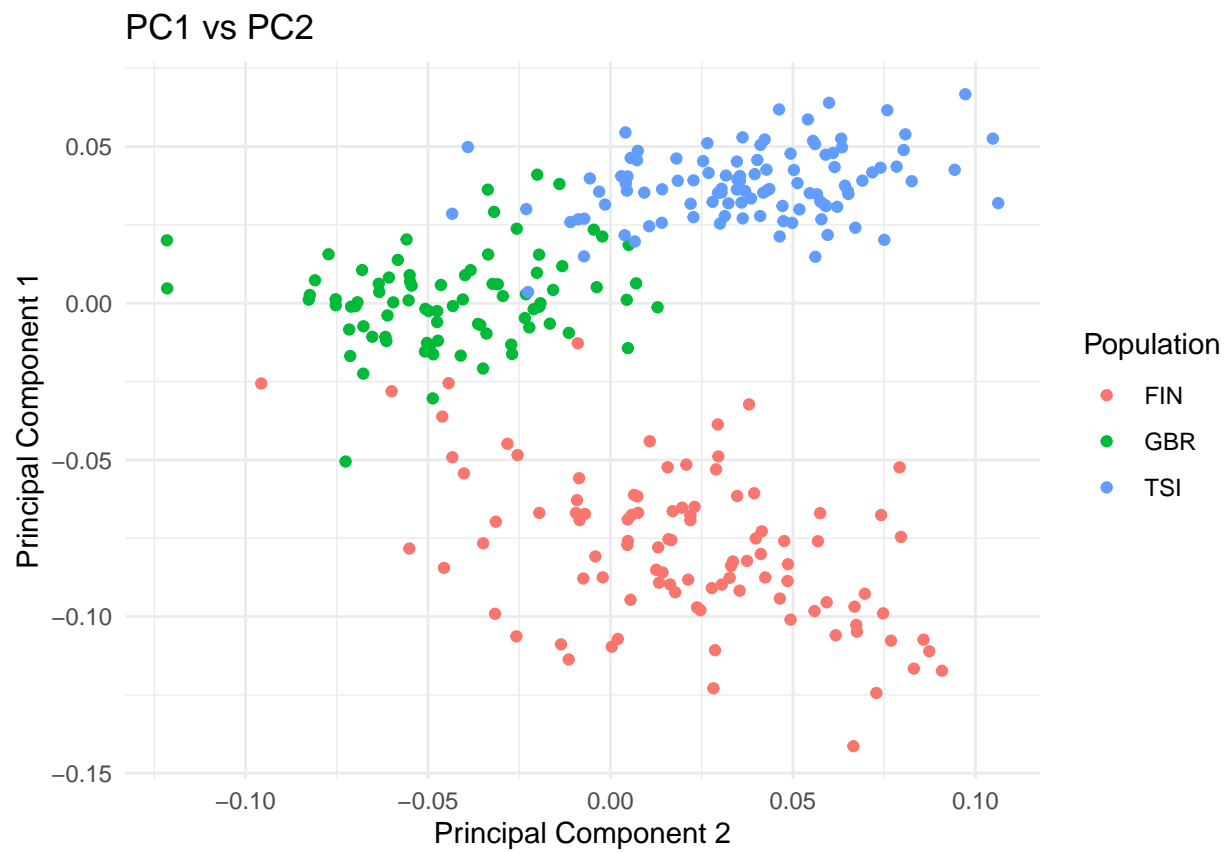
```
## [1] 141123
```

- What is the number of samples in each population?



2. Get the first 10 principal components (PCs) in PLINK using all SNPs. The basic command would look like

- Make a scatterplot of the first two PCs with each point colored by population membership.



- Project new individuals onto PCs

