

Human dataset analysis details

Irineo Cabreros and John D. Storey

4/18/2019

R Markdown

In this document, we detail the procedures used to analyze the human datasets studied in Cabreros and Storey, 2018 (<https://www.biorxiv.org/content/early/2019/03/27/240812.full.pdf>).

Data acquisition

There are four human datasets analyzed in this work:

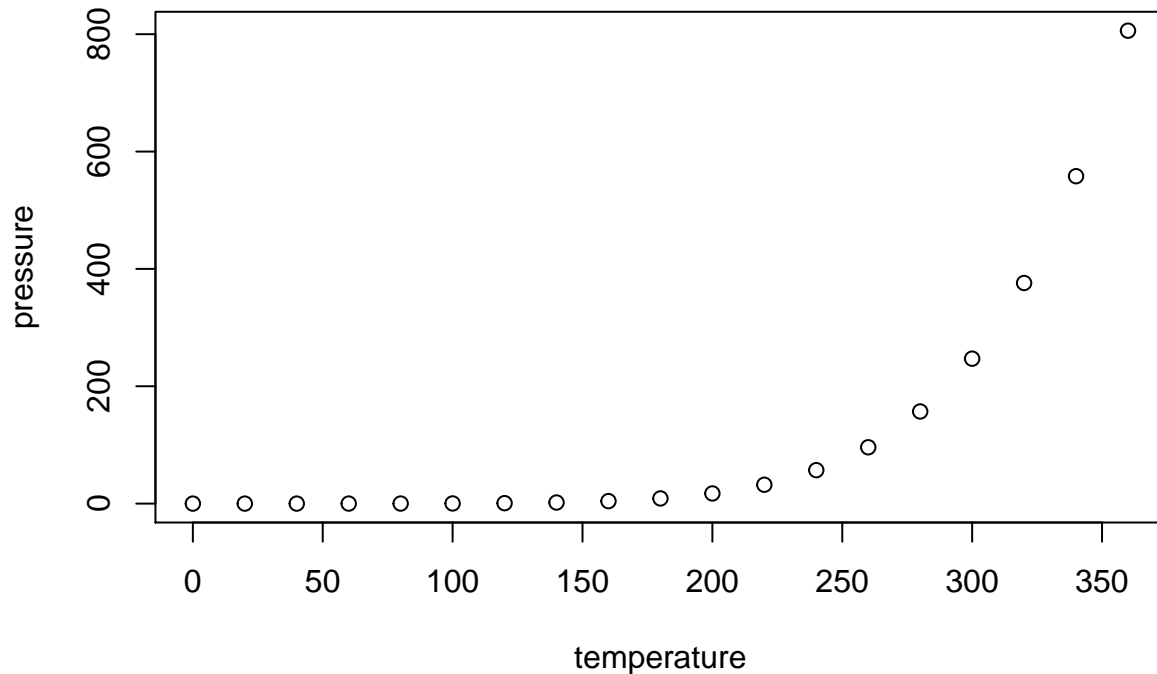
1. TGP (1000 Genomes Project) (The 1000 Genomes Project Consortium 2015)
2. HGDP (Human Genome Diversity Project) (Cavalli-Sforza 2005)
3. HO (Human Origins) (Lazaridis and others 2014)
4. INDIA (Basu, Sarkar-Roy, and Majumder 2016)

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   : 2.00
## 1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##   Mean  :15.4    Mean   : 42.98
## 3rd Qu.:19.0    3rd Qu.: 56.00
##   Max.  :25.0    Max.    :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Basu, Analabha, Neeta Sarkar-Roy, and Partha P. Majumder. 2016. “Genomic Reconstruction of the History of Extant Populations of India Reveals Five Distinct Ancestral Components and a Complex Structure.” *Proceedings of the National Academy of Sciences* 113 (6). National Academy of Sciences: 1594–9. doi:10.1073/pnas.1513197113.

Cavalli-Sforza, L. L. 2005. “The Human Genome Diversity Project: Past, Present and Future.” *Nature Reviews Genetics* 6: 333–40.

Lazaridis, I., and others. 2014. “Ancient Human Genomes Suggest Three Ancestral Populations for Present-Day Europeans.” *Nature* 513: 409–13.

The 1000 Genomes Project Consortium. 2015. “A Global Reference for Human Genetic Variation.” *Nature* 526: 68–74.