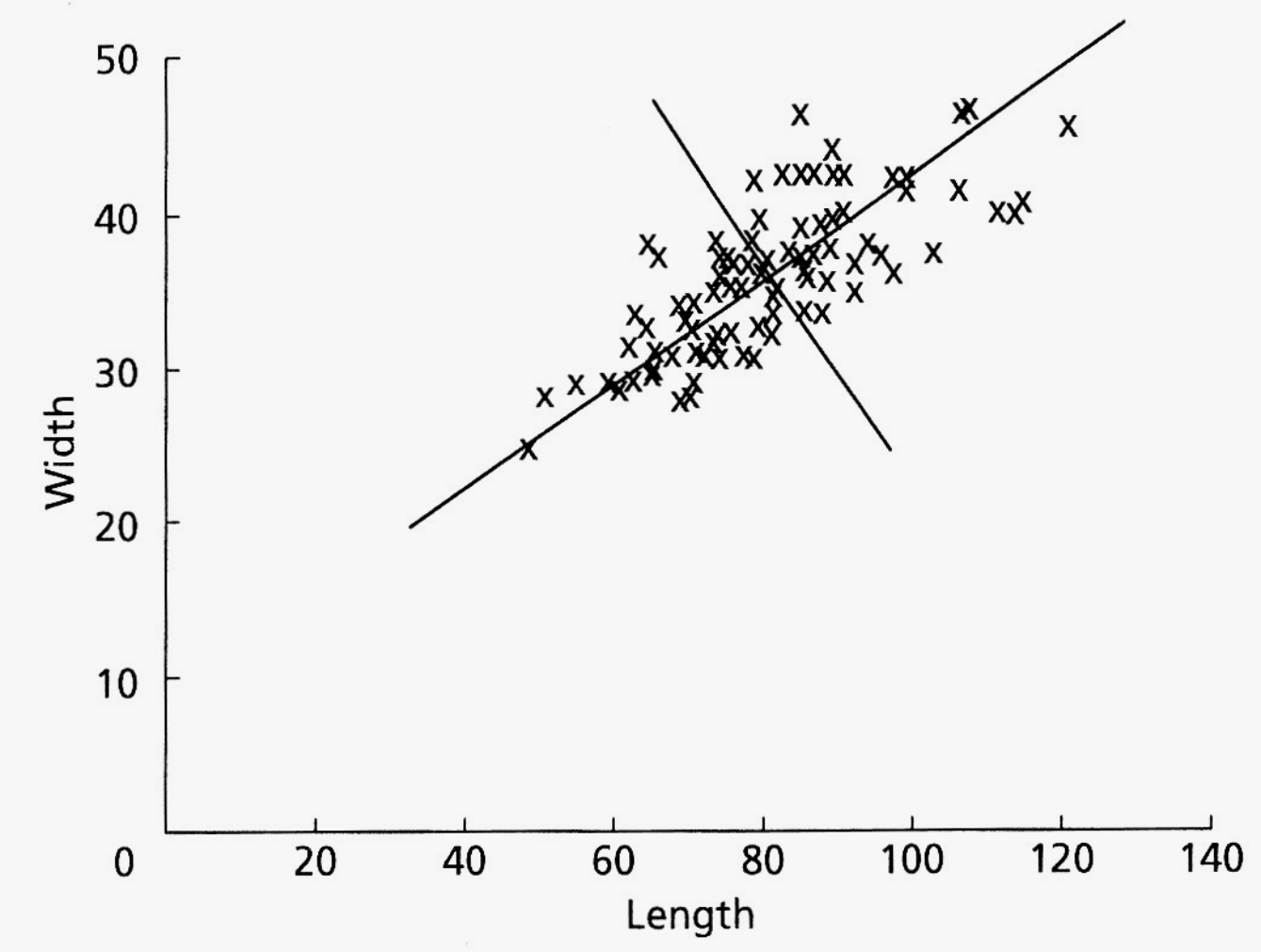
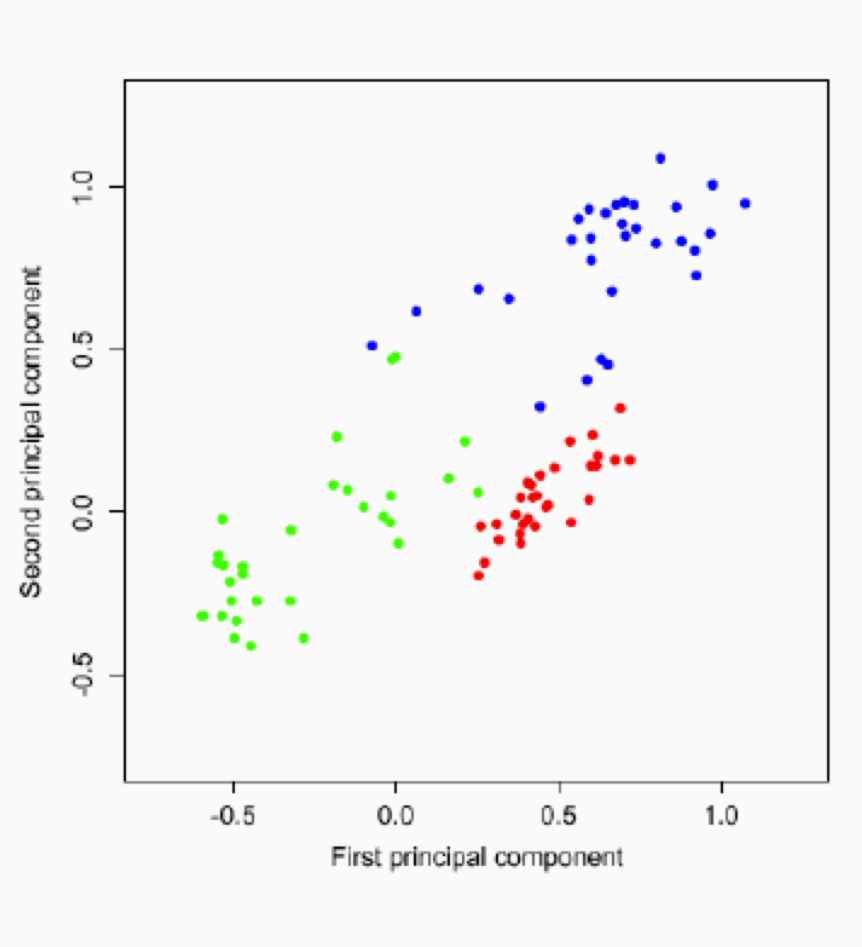
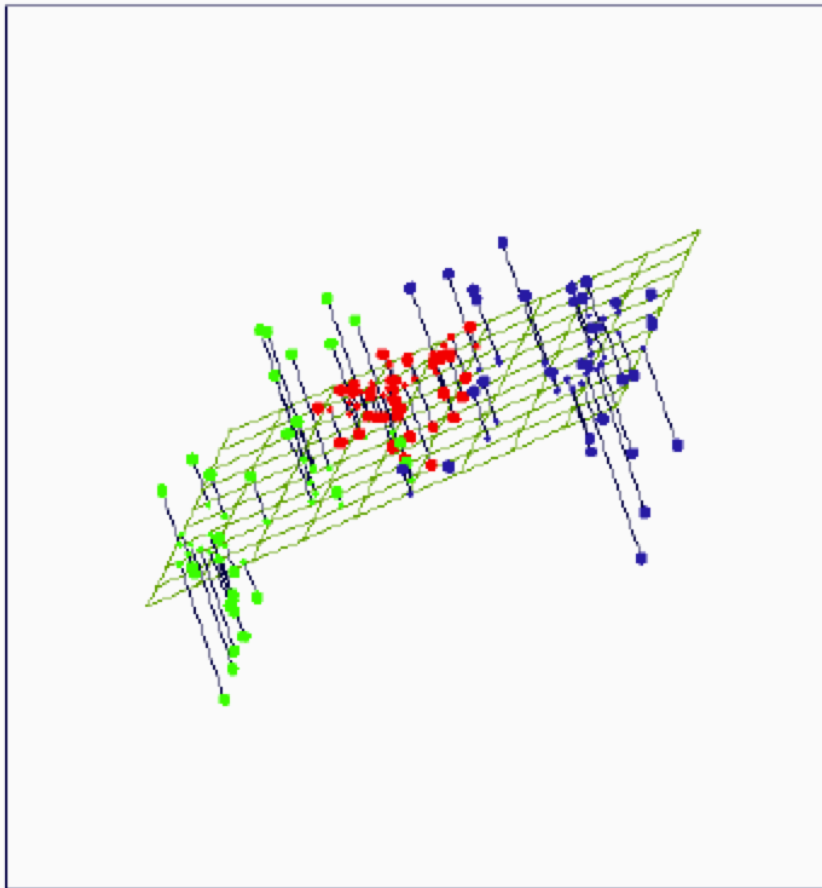
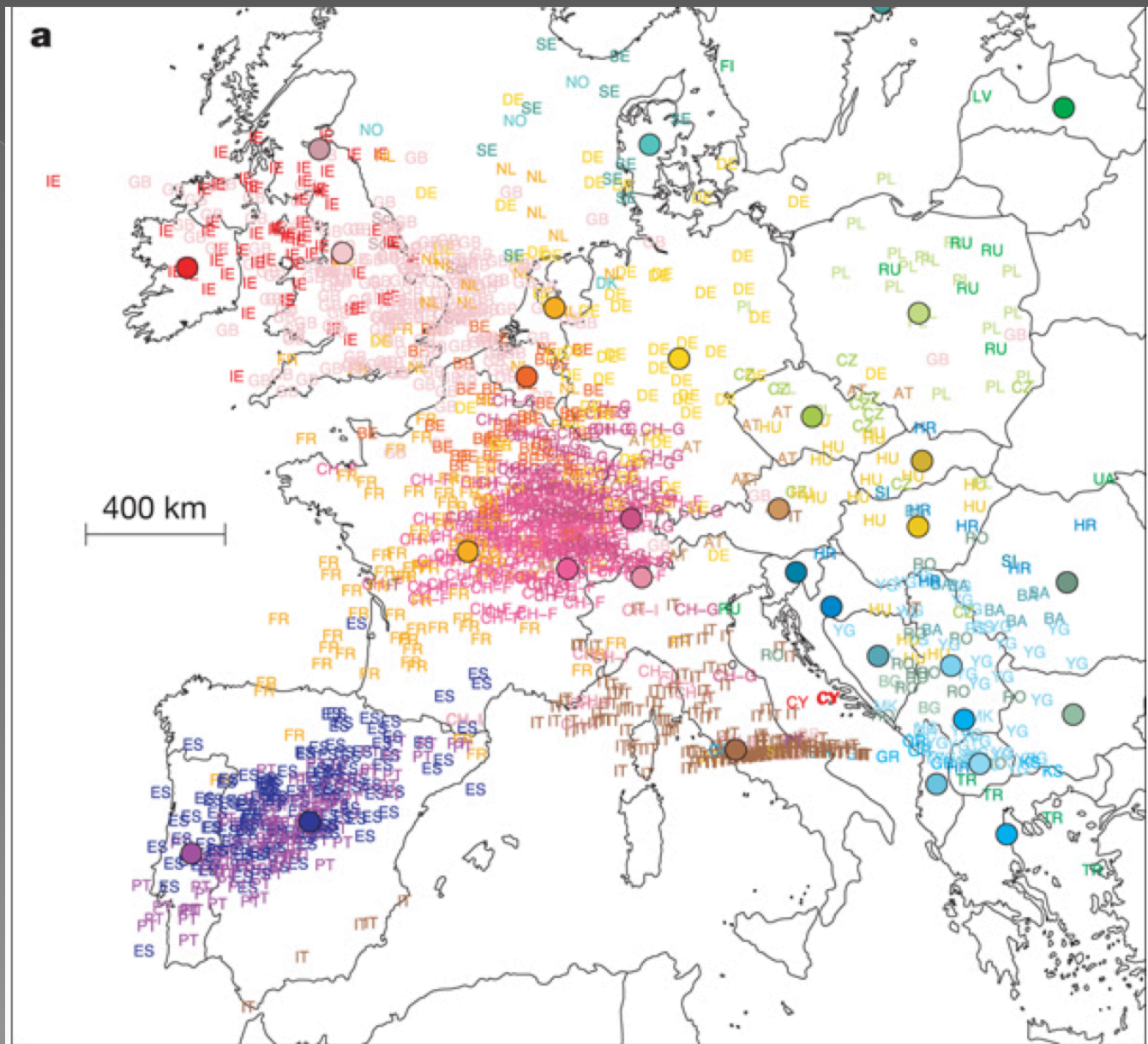


Principal Component Analysis on Ancient Genomes

Tizard







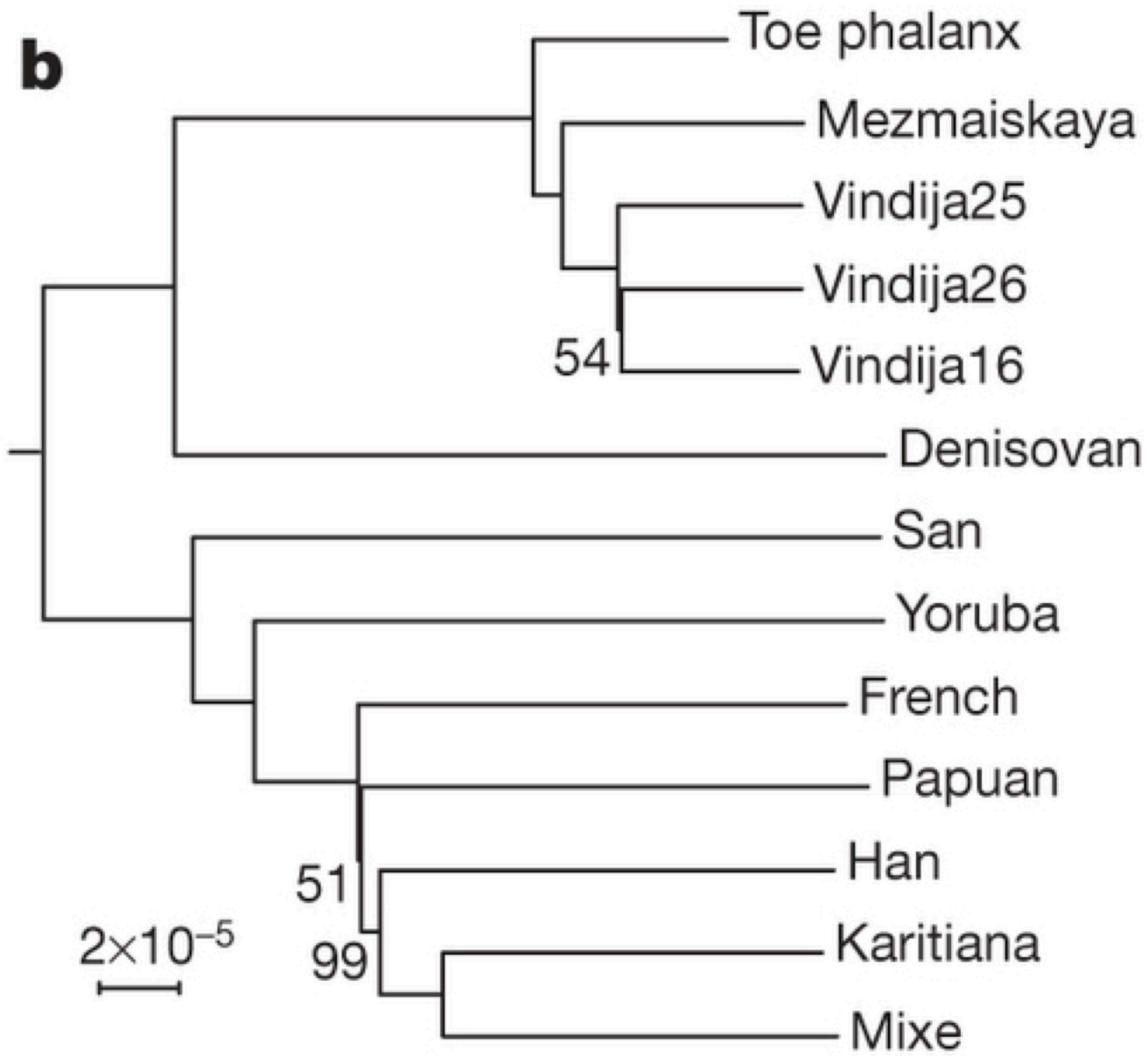
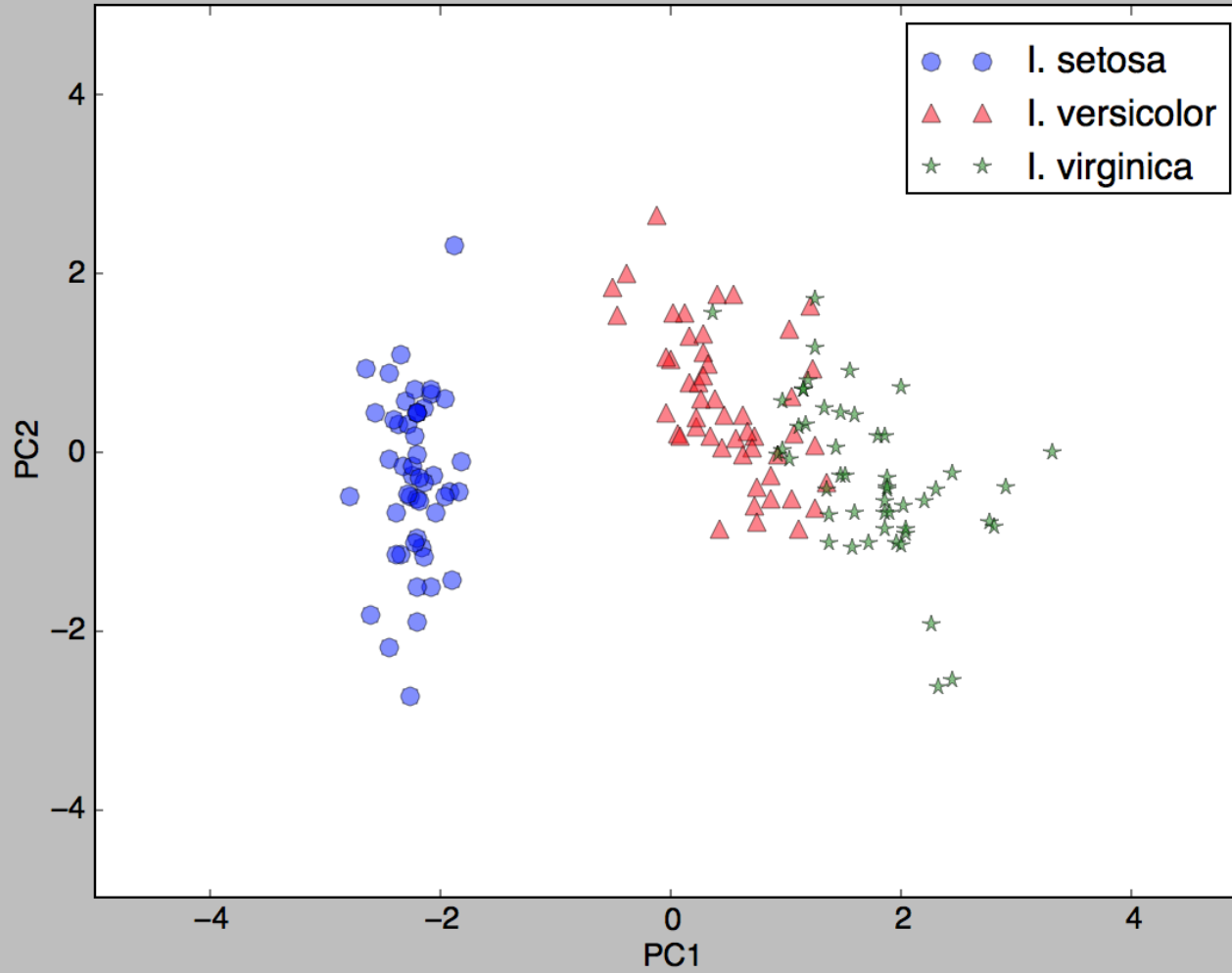


Figure 1

PCA on the iris dataset



```
import numpy as np
from matplotlib.mlab import PCA
import matplotlib.pyplot as plt

f = open('iris.data.txt', 'r')

sepal_len = []
sepal_wid = []
petal_len = []
petal_wid = []

fields = ["sepal length", "sepal width", "petal length", "petal width"]

for line in f:
    line = line.split(',')
    sepal_len.append(line[0])
    sepal_wid.append(line[1])
    petal_len.append(line[2])
    petal_wid.append(line[3])

data = np.zeros(shape=(150, 4))

for t in range(len(data)):
    data[t] = [sepal_len[t], sepal_wid[t], petal_len[t], petal_wid[t]]

for i in range(20):
    print(data[i])

myPCA = PCA(data) # new PCA object

###
# .Y is 'data' put into PCA space
plt.plot(myPCA.Y[0:50, 0],myPCA.Y[0:50,1], 'o', markersize=7, color='blue', alpha=0.5, label="I. setosa")
plt.plot(myPCA.Y[50:100, 0],myPCA.Y[50:100,1], '^', markersize=7, color='red', alpha=0.5, label="I. versicolor")
plt.plot(myPCA.Y[100:150, 0],myPCA.Y[100:150,1], '*', markersize=7, color='green', alpha=0.5, label="I. virginica")
plt.xlim([-5,5])
plt.ylim([-5,5])
plt.xlabel('PC1')
plt.ylabel('PC2')
plt.legend()
plt.title('PCA on the iris dataset')
plt.show()
```

iris.data.txt

```
4.5,2.3,1.3,0.3,Iris-setosa
4.4,3.2,1.3,0.2,Iris-setosa
5.0,3.5,1.6,0.6,Iris-setosa
5.1,3.8,1.9,0.4,Iris-setosa
4.8,3.0,1.4,0.3,Iris-setosa
5.1,3.8,1.6,0.2,Iris-setosa
4.6,3.2,1.4,0.2,Iris-setosa
5.3,3.7,1.5,0.2,Iris-setosa
5.0,3.3,1.4,0.2,Iris-setosa
7.0,3.2,4.7,1.4,Iris-versicolor
6.4,3.2,4.5,1.5,Iris-versicolor
6.9,3.1,4.9,1.5,Iris-versicolor
5.5,2.3,4.0,1.3,Iris-versicolor
6.5,2.8,4.6,1.5,Iris-versicolor
5.7,2.8,4.5,1.3,Iris-versicolor
6.3,3.3,4.7,1.6,Iris-versicolor
4.9,2.4,3.3,1.0,Iris-versicolor
6.6,2.9,4.6,1.3,Iris-versicolor
5.2,2.7,3.9,1.4,Iris-versicolor
5.0,2.0,3.5,1.0,Iris-versicolor
```



```
[ 5.1  3.5  1.4  0.2]
[ 4.9  3.   1.4  0.2]
[ 4.7  3.2  1.3  0.2]
[ 4.6  3.1  1.5  0.2]
[ 5.   3.6  1.4  0.2]
[ 5.4  3.9  1.7  0.4]
[ 4.6  3.4  1.4  0.3]
[ 5.   3.4  1.5  0.2]
[ 4.4  2.9  1.4  0.2]
[ 4.9  3.1  1.5  0.1]
[ 5.4  3.7  1.5  0.2]
[ 4.8  3.4  1.6  0.2]
[ 4.8  3.   1.4  0.1]
[ 4.3  3.   1.1  0.1]
[ 5.8  4.   1.2  0.2]
[ 5.7  4.4  1.5  0.4]
[ 5.4  3.9  1.3  0.4]
[ 5.1  3.5  1.4  0.3]
[ 5.7  3.8  1.7  0.3]
[ 5.1  3.8  1.5  0.3]
```


RSID, CHROMOSOME, POSITION, RESULT

"rs58108140", "1", "10583", "GG"
"rs189107123", "1", "10611", "CC"
"rs140337953", "1", "30923", "TT"
"rs187298206", "1", "51476", "TT"
"rs116400033", "1", "51479", "TT"
"rs190452223", "1", "51914", "TT"
"rs181754315", "1", "51935", "CC"
"rs185832753", "1", "51954", "GG"
"rs62637813", "1", "52058", "GG"
"rs190291950", "1", "52144", "TT"
"rs140052487", "1", "54353", "CC"
"rs146477069", "1", "54421", "AA"
"rs141149254", "1", "54490", "GG"
"rs2462492", "1", "54676", "CC"
"rs143174675", "1", "54753", "TT"
"rs10399749", "1", "55299", "CT"
"rs182462964", "1", "55313", "AA"
"rs3107975", "1", "55326", "TT"
"rs185215913", "1", "55330", "GG"
"rs190850374", "1", "55367", "GG"
"rs182711216", "1", "55388", "CC"
"rs2949420", "1", "55394", "TT"
"rs193242050", "1", "55416", "GG"
"rs183189405", "1", "55427", "TT"
"rs187434873", "1", "55816", "GG"
"rs191890754", "1", "55850", "CC"
"rs184233019", "1", "55852", "GG"
"rs143342222", "1", "56644", "AA"
"rs114420996", "1", "58814", "GA"

RSID, CHROMOSOME, POSITION, RESULT

"rs58108140", "1", "10583", "GG"
"rs189107123", "1", "10611", "CC"
"rs140337953", "1", "30923", "TT"
"rs187298206", "1", "51476", "TT"
"rs116400033", "1", "51479", "TT"
"rs190452223", "1", "51914", "TT"
"rs181754315", "1", "51935", "CC"
"rs185832753", "1", "51954", "GG"
"rs62637813", "1", "52058", "GG"
"rs190291950", "1", "52144", "TT"
"rs140052487", "1", "54353", "CC"
"rs146477069", "1", "54421", "AA"
"rs141149254", "1", "54490", "GG"
"rs2462492", "1", "54676", "CC"
"rs143174675", "1", "54753", "TT"
"rs3091274", "1", "55164", "CA"
"rs10399749", "1", "55299", "CC"
"rs182462964", "1", "55313", "AA"
"rs3107975", "1", "55326", "TT"
"rs185215913", "1", "55330", "GG"
"rs190850374", "1", "55367", "GG"
"rs182711216", "1", "55388", "CC"
"rs2949420", "1", "55394", "TT"
"rs193242050", "1", "55416", "GG"
"rs183189405", "1", "55427", "TT"
"rs187434873", "1", "55816", "GG"
"rs191890754", "1", "55850", "CC"
"rs184233019", "1", "55852", "GG"
"rs143342222", "1", "56644", "AA"

- "rs147215883","1","79772","CC" → 0
- "rs147215883","1","79772","CG" → 1
- "rstotallyfake1","1","79772","GG" → 2

RSID, CHROMOSOME, POSITION, RESULT

"rs58108140", "1", "10583", "GG"
"rs189107123", "1", "10611", "CC"
"rs140337953", "1", "30923", "TT"
"rs187298206", "1", "51476", "TT"
"rs116400033", "1", "51479", "TT"
"rs190452223", "1", "51914", "TT"
"rs181754315", "1", "51935", "CC"
"rs185832753", "1", "51954", "GG"
"rs62637813", "1", "52058", "GG"
"rs190291950", "1", "52144", "TT"
"rs140052487", "1", "54353", "CC"
"rs146477069", "1", "54421", "AA"
"rs141149254", "1", "54490", "GG"
"rs2462492", "1", "54676", "CC"
"rs143174675", "1", "54753", "TT"
"rs10399749", "1", "55299", "CT"
"rs182462964", "1", "55313", "AA"
"rs3107975", "1", "55326", "TT"
"rs185215913", "1", "55330", "GG"
"rs190850374", "1", "55367", "GG"
"rs182711216", "1", "55388", "CC"
"rs2949420", "1", "55394", "TT"
"rs193242050", "1", "55416", "GG"
"rs183189405", "1", "55427", "TT"
"rs187434873", "1", "55816", "GG"
"rs191890754", "1", "55850", "CC"
"rs184233019", "1", "55852", "GG"
"rs143342222", "1", "56644", "AA"
"rs114420996", "1", "58814", "GA"

RSID, CHROMOSOME, POSITION, RESULT

"rs58108140", "1", "10583", "GG"
"rs189107123", "1", "10611", "CC"
"rs140337953", "1", "30923", "TT"
"rs187298206", "1", "51476", "TT"
"rs116400033", "1", "51479", "TT"
"rs190452223", "1", "51914", "TT"
"rs181754315", "1", "51935", "CC"
"rs185832753", "1", "51954", "GG"
"rs62637813", "1", "52058", "GG"
"rs190291950", "1", "52144", "TT"
"rs140052487", "1", "54353", "CC"
"rs146477069", "1", "54421", "AA"
"rs141149254", "1", "54490", "GG"
"rs2462492", "1", "54676", "CC"
"rs143174675", "1", "54753", "TT"
"rs3091274", "1", "55164", "CA"
"rs10399749", "1", "55299", "CC"
"rs182462964", "1", "55313", "AA"
"rs3107975", "1", "55326", "TT"
"rs185215913", "1", "55330", "GG"
"rs190850374", "1", "55367", "GG"
"rs182711216", "1", "55388", "CC"
"rs2949420", "1", "55394", "TT"
"rs193242050", "1", "55416", "GG"
"rs183189405", "1", "55427", "TT"
"rs187434873", "1", "55816", "GG"
"rs191890754", "1", "55850", "CC"
"rs184233019", "1", "55852", "GG"
"rs143342222", "1", "56644", "AA"

