

---

# **pandas-plink Documentation**

*Release 1.1.6*

**Danilo Horta**

April 14, 2017



<b>1</b>	<b>Install</b>	<b>3</b>
<b>2</b>	<b>Usage</b>	<b>5</b>
<b>3</b>	<b>Functions</b>	<b>7</b>
	<b>Python Module Index</b>	<b>9</b>



You can get the source and open issues on [Github](#).



---

## Install

---

The recommended way of installing it is via `conda`:

```
conda install -c conda-forge pandas-plink
```

An alternative way would be via `pip`:

```
pip install pandas-plink
```





---

## Usage

---

It is as simple as:

```
from pandas_plink import read_plink
(bim, fam, G) = read_plink('/path/to/data')
```

assuming that you have the files

- */path/to/data.bim*
- */path/to/data.fam*
- */path/to/data.bed*

The returned matrix G contains 0, 1, 2, or 3s:

- 0 Homozygous for first allele in .bim file
- 1 Heterozygous
- 2 Homozygous for second allele in .bim file
- 3 Missing genotype



---

## Functions

---

Read PLINK files into Pandas data frames.

`pandas_plink.read_plink(file_prefix, verbose=True)`

Convert PLINK files into Pandas data frames.

### Parameters

- **file\_prefix** (*str*) – Path prefix to the set of PLINK files.
- **verbose** (*bool*) – *True* for progress information; *False* otherwise.

### Returns

parsed data containing:

- `pandas.DataFrame`: alleles.
- `pandas.DataFrame`: samples.
- `numpy.ndarray`: genotype.

**Return type** `tuple`

### Examples

We have shipped this package with an example so can load and inspect by doing

```
>>> from pandas_plink import read_plink
>>> from pandas_plink import example_file_prefix
>>> (bim, fam, bed) = read_plink(example_file_prefix(), verbose=False)
>>> print(bim.head())
  chrom      snp      cm      pos a0 a1  i
0     1  rs10399749  0.0   45162  G  C  0
1     1  rs2949420  0.0   45257  C  T  1
2     1  rs2949421  0.0   45413  0  0  2
3     1  rs2691310  0.0   46844  A  T  3
4     1  rs4030303  0.0   72434  0  G  4
>>> print(fam.head())
   fid      iid      father      mother gender trait  i
0  Sample_1  Sample_1         0         0      1     -9  0
1  Sample_2  Sample_2         0         0      2     -9  1
2  Sample_3  Sample_3  Sample_1  Sample_2      2     -9  2
>>> print(bed.compute())
[[2 2 1]
 [2 1 2]]
```

```
[3 3 3]
[3 3 1]
[2 2 2]
[2 2 2]
[2 1 0]
[2 2 2]
[1 2 2]
[2 1 2]]
```

Notice the *i* column in bim and fam data frames. It maps to the corresponding position of the bed matrix:

```
>>> from pandas_plink import read_plink
>>> from pandas_plink import example_file_prefix
>>> (bim, fam, bed) = read_plink(example_file_prefix(), verbose=False)
>>> chrom1 = bim.query("chrom=='1'")
>>> X = bed[chrom1.i,:].compute()
>>> print(X)
[[2 2 1]
 [2 1 2]
 [3 3 3]
 [3 3 1]
 [2 2 2]
 [2 2 2]
 [2 1 0]
 [2 2 2]
 [1 2 2]
 [2 1 2]]
```

`pandas_plink.test()`

Tests this package.

You will need *pytest* installed in order to use this function.

`pandas_plink.example_file_prefix()`

Data files prefix.

**p**

`pandas_plink`, 7



## E

`example_file_prefix()` (in module `pandas_plink`), 8

## P

`pandas_plink` (module), 7

## R

`read_plink()` (in module `pandas_plink`), 7

## T

`test()` (in module `pandas_plink`), 8