

### Imports

```
import pandas as pd
import matplotlib
import matplotlib.pyplot as plt
plt.ion() setting interactive mode
```

### Load a CSV file

```
pract=pd.read_csv('practice-
dataframe.csv',index_col=0)
```

*index\_col=0 the first column contains the row names*

### Displaying values

#### Example dataframe:

```
age height sex
Ann 22 170 female
Bob 19 182 male
Carla 20 165 dona
```

#### For columns:

```
df.age
df.height
df.sex
```

#### For row names:

```
df.index
```

#### Selecting rows that have a particular value in some column:

```
df[df.age<22]
df[df.height>170]
df[df.sex==female]
```

*These commands return you all the information for the rows.*

#### If we want only the row names rather than all the information:

```
df[df.age==22].index
```

*The second part ('sex') selects the 'sex' column of the dataframe.*

#### If we only want their gender:

```
df[df.age==22].sex
```

```
grad_values=alladjs.gradGlob
```

*we select the column and assign it to a variable (gradGlob). Then, we can compute statistics for this column.*

### If you don't remember...

Operand	Example	Meaning
==	\$variable1 == \$variable2	Has the same value as
!=	\$variable1 != \$variable2	Is NOT the same value as
<	\$variable1 < \$variable2	Less Than
>	\$variable1 > \$variable2	Greater Than
<=	\$variable1 <= \$variable2	Less than or equals to
>=	\$variable1 >= \$variable2	Greater than or equals to

Here you have the comparison operators

### Compute the...

mean	median	standard deviation
variable.me	variable.medi	variable.std()
an()	an()	

### Visualizing data

#### If we want to return the first rows of the data:

```
stud.head()
```

#### If we want to visualize all the data, in a single boxplot:

```
stud.boxplot()
```

```
plt.show()
```

*IMPORTANT: close the graph window with your mouse to continue or use:*

```
plt.close()
```

#### If we want to see one boxplot per variable:

```
stud.boxplot(by='teacher')
```

We can use the same formulas with

**.histogram** command

### Drawing a...

#### Histogram:

```
grad_small.plot(kind='hist')
```

#### Boxplot:

```
grad_small.plot(kind='box')
```

#### Barplot:

```
grad_small.plot(kind='bar')
```

#### Saving a figure:

```
plt.savefig("small-histogram.pdf")
```

*the command .plot can draw different kinds of plots*

### Selecting a...

#### Selecting rows based on +1 condition:

```
sm=small[(small.semantic_class=="qualitative")
| (small.semantic_class=="relational")]
```

#### Selecting columns that we need to be able to do a boxplot:

```
pred_and_dertype=all(['predGlob','derType'])
```

#### Selecting all the participial adjectives of the database:

```
part=all[all.derType=='participi']
```

#### If you want to check what you obtain:

```
pred_and_dertype.head()
```

#### If you want to select a concrete variable, for instance, "participial adjectives":

```
part=all[all.derType=='participi']
```

*symbol "|" means 'or'*

### How to return a value distribution

```
sex_var=pract.sex
```

```
sex_var.value_counts()
```

### Contingency table

#### If we need to compare two categorical variables:

```
first.head()
```

```
pd.crosstab(first.teacher,first.student_passed)
```

*we are cross-tabulating the teacher with whether the student passed the exam or not*

### How to...

sample	get a random sample
first50=adjs.head(50)	random_sample=adjs.sample(50)
first50.head()	random_sample.head()
first50.index	random_sample.index



By [simona20](#)  
[cheatography.com/simona20/](http://cheatography.com/simona20/)

Not published yet.

Last updated 1st March, 2018.

Page 1 of 1.

Sponsored by [CrosswordCheats.com](#)

Learn to solve cryptic crosswords!

<http://crosswordcheats.com>