Cheatography

B1_zorianflowers_week08_cheatsheet Cheat Sheet

by zflow via cheatography.com/127367/cs/24779/

Pandas

import pandas as pd	pd.series([values])
	ad = {}
	area = pd.series(ad)

Retrieving Values	area["a"]
	To see all keys:
	area.keys()
	data.items()

Dataframe as dictionary	area = pd.series({...})
	data = pd.Dataframe({"area":area,})

Opening data	import pandas as pd import numpy as np dat = np.genfromtxt('phoneBook.csv',delimiter=',',skip_header=1,dtype='<U16')

Grouping	index = pd.MultiIndex.from_tuples(index)
	index
	
	pop = pop.reindex(index) pop
	
	pop[:, 2010]

Merging and Joining

Merging	pd.merge()
	df = pd.merge()
Many to one	Duplicate entries display('df3', 'df4', 'pd.merge(df3, df4)')
Merge Key	Add on = "key column name"
Drop	.drop('name', axis=1)

Aggregation and Grouping

Aggregation Functions	count() | Total number of items first(), last() | First and last item mean(), median() | Mean and median min(), max() | Minimum and maximum std(), var() | Standard deviation and variance mad() | Mean absolute deviation prod() | Product of all items sum() | Sum of all items
Grouping	name.groupy("key")

Pivot Tables

Pivot Tables by Hand	Require groupby
Pivot	name.pivot_table("what is taking the action", index = "groupby row" , columns = "groupbycol")
Aggregation Functions	name.pivot_table(index = "groupby row" , columns = "groupbycol" , aggfunc={'taking action':sum, 'taking action':'mean'})

Matplotlib

Line Plots	Set linspace x = np.linspace(0, 10, 100)

Creating figure and axis	fig = plt.figure()
	ax = plt.axes()

Add graph and x,y	x = np.linspace(0, 10, 1000)
	y = np.sin(x)
	plt.plot(x,y)
	plt.show()

Changing linestyle and color	plt.plot(x,y,linestyle='--', color='c')

Multile curves and a legend	plt.plot(x,np.sin(x-.5),color='g',label="sin(x-0.5)")
	
	plt.plot(x,np.sin(x-1),color='pink', label = "sin(x-1)")
	
	plt.plot(x,np.cos(x-0.5),color='c',linestyle='--',label = "cos(x-0.5)") plt.legend() plt.show()

Adding limits	plt.xlim(-5,12)
	plt.ylim(-2,2)

Scatter Plot	x = np.random.randint(-1000,1000,150)
	y = np.random.randint(-1000,1000,150) plt.scatter(x,y)
	
	or plt.plot(x,y,'o');

Matplotlib (cont)

Other ways for plt	plt.xlabel() → ax.set_xlabel()
	plt.ylabel() → ax.set_ylabel()
	lt.xlim() → ax.set_xlim()
	plt.ylim() → ax.set_ylim()
	plt.title() → ax.set_title()

Histograms	fig = plt.figure()
	ax = plt.axes()
	ax.hist(data);