

Taller de Programación

Numpy & Pandas

Daniela Opitz
dopitz@udd.cl



Basada en presentaciones oficiales de libro Introduction to Programming in Python (Sedgewick, Wayne, Dondero).

Disponible en <https://introcs.cs.princeton.edu/python>






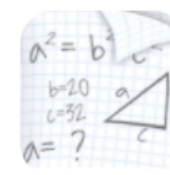





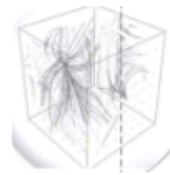



Outline

- Introducción a Numpy
- Actividades

Numpy



- NumPy es el paquete fundamental para la computación científica en Python.
- Es una biblioteca de Python que proporciona un objeto de matriz multidimensional, varios objetos derivados (como matrices y matrices enmascaradas).
- Contiene una variedad de rutinas para operaciones rápidas en matrices, que incluyen manipulación matemática, lógica, cambio de dimensiones, clasificación, selección, transformadas discretas de Fourier, álgebra lineal básica, operaciones estadísticas básicas, simulación aleatoria y mucho más.

Quantum Computing  QuTiP PyQuil Qiskit	Statistical Computing  Pandas statsmodels Seaborn	Signal Processing  SciPy PyWavelets	Image Processing  Scikit-image OpenCV	3-D Visualization  Mayavi Napari	Symbolic Computing  SymPy	Astronomy Processes  AstroPy SunPy SpacePy	Cognitive Psychology  PsychoPy
Bioinformatics  BioPython Scikit-Bio PyEnsembl	Bayesian Inference  PyStan PyMC3	Mathematical Analysis  SciPy SymPy cvxpy FEniCS	Simulation Modeling  PyDSTool	Multi-variate Analysis  PyChem	Geographic Processing  Shapely GeoPandas Folium	Interactive Computing  Jupyter IPython Binder	

Arrays

Un arreglo o array es la estructura de datos central de la biblioteca NumPy. Consiste de una cuadrícula de valores y contiene información sobre los datos sin procesar, cómo ubicar un elemento y cómo interpretar un elemento.

- Una dimension

```
a = np.array([1, 2, 3, 4, 5, 6])
```



```
a[0]: 1
```

Todos los elementos de un arreglo (array) deben ser homogéneos (del mismo tipo)!

- Dos dimensiones (3 filas x 4 columnas)

```
a = np.array([[1, 2, 3, 4], [5, 6, 7, 8], [9, 10, 11, 12]])
```



```
a[0]: [1, 2, 3, 4]
```

Operaciones Básicas con Numpy

Zeros y Unos

```
>>> np.zeros(2)
array([0., 0.]
```

```
>>> np.ones(2)
array([1., 1.]
```

Rango de Elementos

```
>>> np.arange(4)
array([0, 1, 2, 3])
```

```
>>> np.arange(2, 9, 2)
array([2, 4, 6, 8])
```

```
>>> np.linspace(0, 10, num=5)
array([ 0. ,  2.5,  5. ,  7.5, 10. ])
```

Ordenar

```
>>> arr = np.array([2, 1, 5, 3, 7, 4, 6, 8])
```

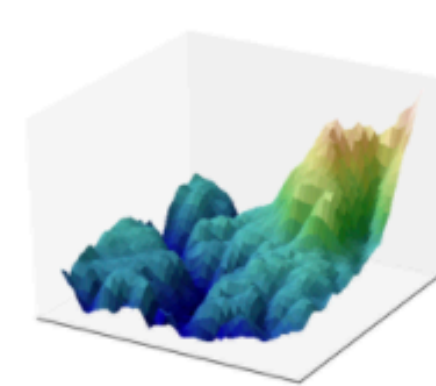
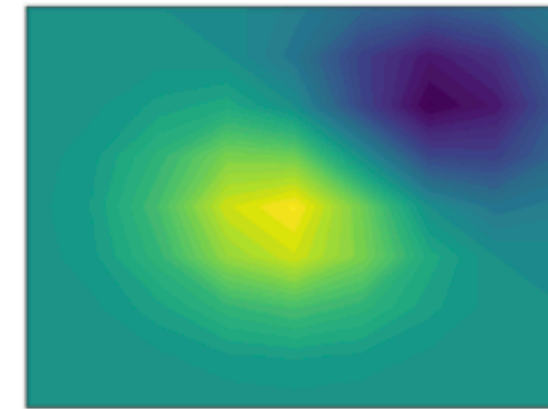
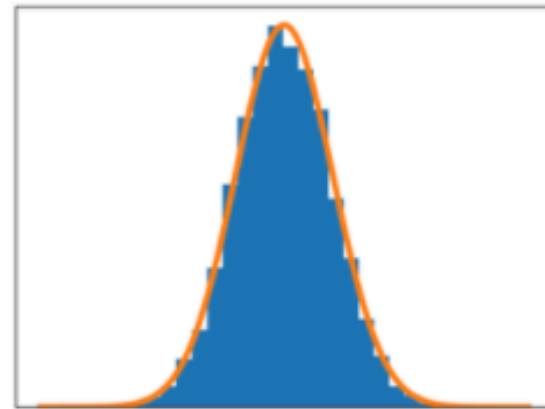
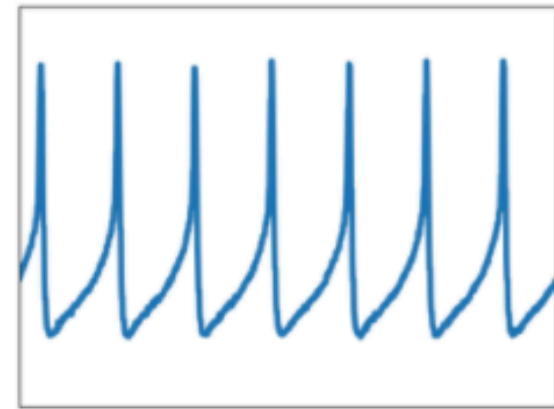
```
>>> np.sort(arr)
array([1, 2, 3, 4, 5, 6, 7, 8])
```

Concatenar

```
>>> a = np.array([1, 2, 3, 4])
>>> b = np.array([5, 6, 7, 8])
```

```
>>> np.concatenate((a, b))
array([1, 2, 3, 4, 5, 6, 7, 8])
```

Matplotlib



Leer instrucciones de instalación aquí:

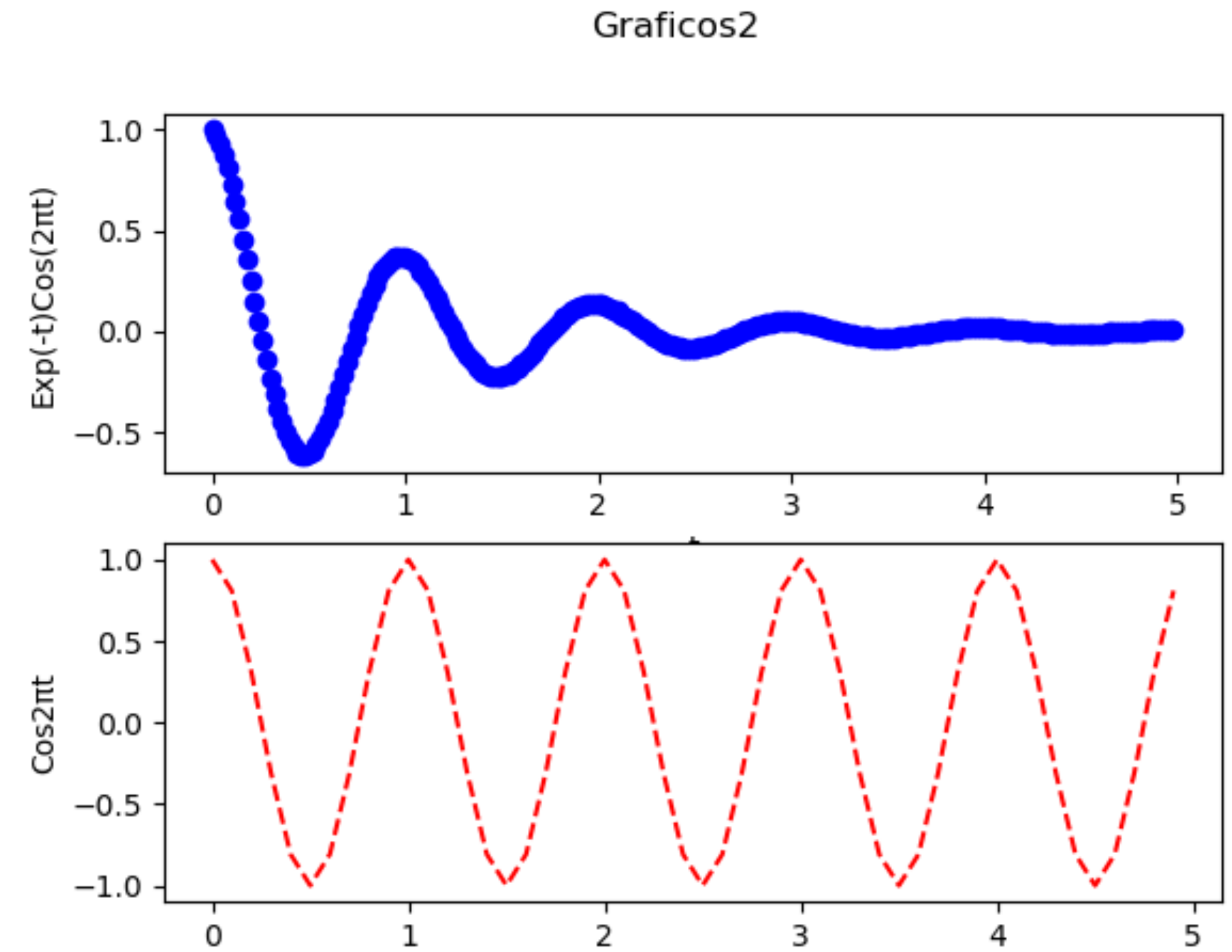
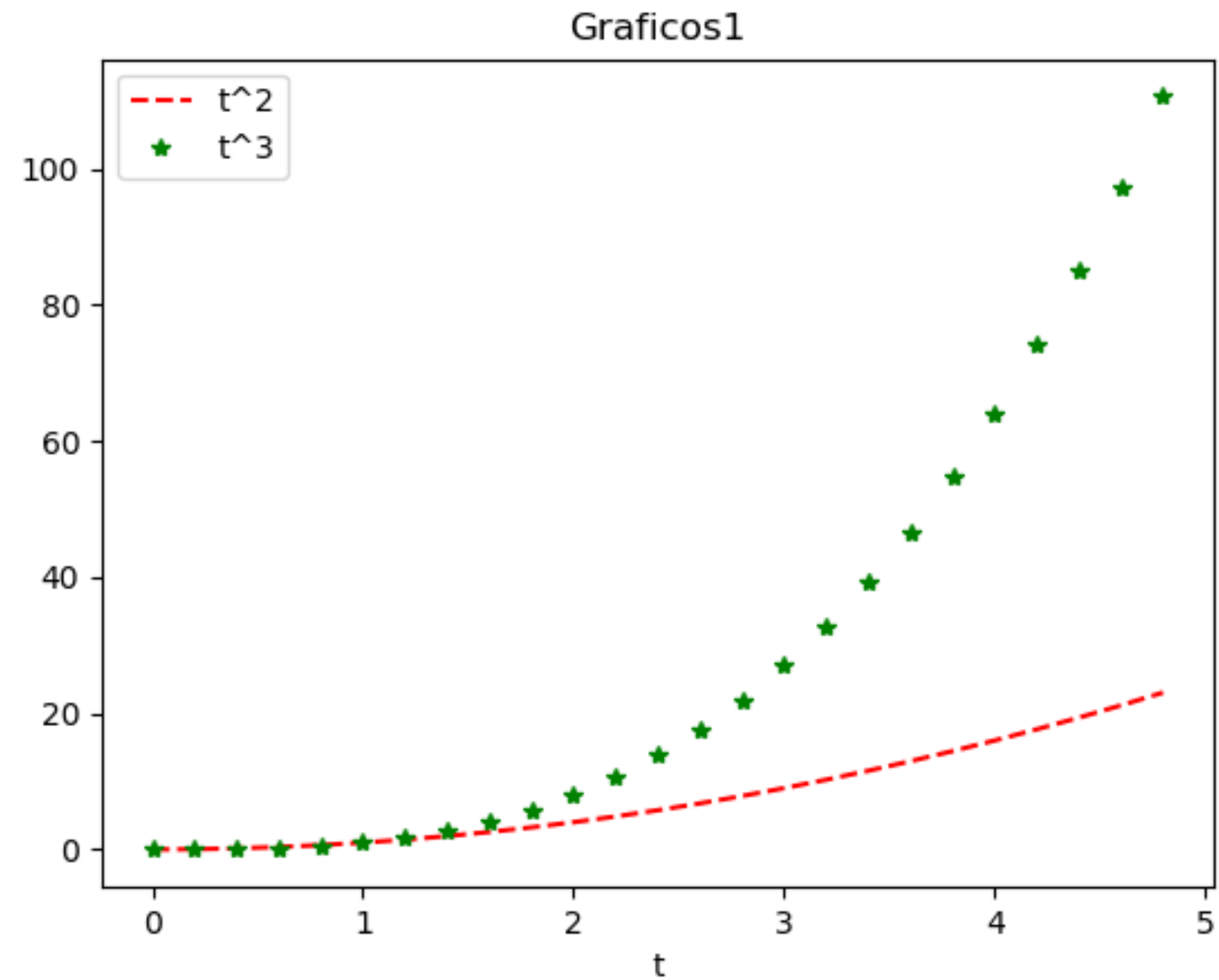
<https://matplotlib.org/users/installing.html#installing-an-official-release>

Instalar con conda:

```
conda install -c conda-forge matplotlib
```

Guia de uso : <https://matplotlib.org/tutorials/introductory/usage.html#sphx-glr-tutorials-introductory-usage-py>

Numpy & Matplotlib

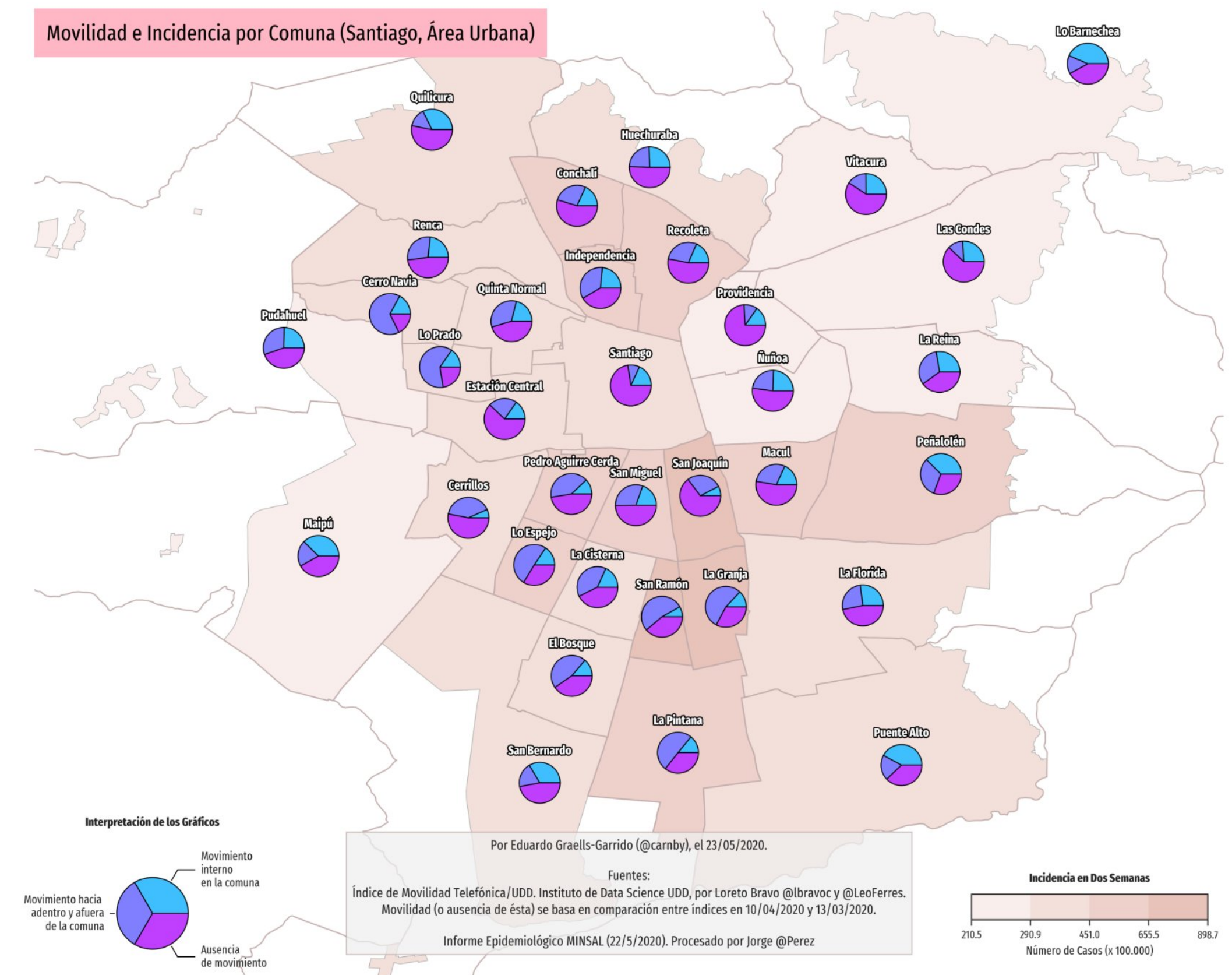


Actividad

El archivo `IndiceDeMovilidad-IM.csv` contiene la evolución del índice de movilidad para cada comuna, donde la movilidad es una medida de los viajes que ocurren al interior y al exterior de dicha unidad administrativa. Este índice se obtiene considerando el movimiento de los teléfonos móviles conectados a la red de Telefónica en el territorio nacional, de manera agrupada y anónima.

Grafique el índice de movilidad versus el tiempo para estudiar su evolución para la comuna de Las Condes. Compare el gráfico con la evolución de otra comuna, por ejemplo, la comuna de La Granja

Nota: para más detalles revisar <https://github.com/MinCiencia/Datos-COVID19/tree/master/output/producto33>.



Fuente: Eduardo Graells <https://twitter.com/carnby/status/1264228522165403654>

Pandas



INSTALAR



```
conda install pandas
```