apolo_charla

August 26, 2018

1 Flujo de trabajo y exploración de datos con Jupyter Notebook

Material complementario: Este cuaderno de Jupyter es el material complementario y de apoyo usado para la charla realizada el 6 de junio de 2018 en el ciclo de conferencias Apolo.

1.1 Motivación

- Educación: procedimientos y algoritmos para estudiantes.
- Investigación: reproducibilidad y replicación.

1.1.1 Educación

En múltiples disciplinas la enseñanza puede llevar a elementos en los cuales se establecen procedimientos estándar (algoritmos) a los estudiantes, en los cuales la variación de los parámetros iniciales permite la práctica del estudiante pero requiere de un paso a paso para su validación.

1.1.2 Investigación

La reproducibilidad y la replicación son una preocupación actual en los resultados de investigación (se puede decir que atravesamos una crisis de reproducibilidad y de ahí credibilidad). Es posible mitigar parte del problema con herramientas de software en el flujo de trabajo.

1.2 Requisitos

- Portabilidad.
- Control de versiones.
- Facilidad de compartir.
- Colaboración en tiempo real.

1.2.1 Portabilidad

- MyBinder.
- CoCalc.

1.2.2 Control de versiones

- Github.
- nbdime.

```
In [1]: %%bash
        nbdiff apolo_charla.ipynb apolo_charla_inicial.ipynb
```

```
nbdiff apolo_charla.ipynb apolo_charla_inicial.ipynb
--- apolo_charla.ipynb  2018-08-26 21:22:17.546712
+++ apolo_charla_inicial.ipynb  2018-06-06 00:03:24
## modified /cells/5/source:
@@ -1,4 +1,4 @@
 ### Control de versiones

 +  [Github](https://github.com/cosmoscalibur/aula-notebook/commit/d2963b13c36a6b2c5cfe866826c735ea69e18
-+  [nbdime](https://github.com/jupyter/nbdime).

++  [nbdime](https://github.com/jupyter/nbdime).


## inserted before /cells/6:
+  code cell:## deleted /cells/6-22:
-  code cell:-    execution_count: 1-    source:-      %%bash-      nbdiff apolo_charla.ipynb apolo_charla_inicial.ip
-          @@ -1,4 +1,4 @@
-           ### Control de versiones
-
-           +  [Github](https://github.com/cosmoscalibur/aula-notebook/commit/d2963b13c36a6b2c5cfe866826c73
-          -+  [nbdime](https://github.com/jupyter/nbdime).
-
-          ++  [nbdime](https://github.com/jupyter/nbdime).
-
--          ## inserted before /cells/6:
-          +  code cell:--          ## deleted /cells/6-23:
-          -  code cell:-          -    execution_count: 1-          -    source:-          -      %%bash-          -      nbdiff a
-          -          @@ -1,4 +1,4 @@
-          -           ### Control de versiones
-          -
-          -           +  [Github](https://github.com/cosmoscalibur/aula-notebook/commit/d2963b13c36a6b2c5cfe86
-          -          -+  [nbdime](https://github.com/jupyter/nbdime).
-          -
-          -          ++  [nbdime](https://github.com/jupyter/nbdime).
-          -
-          --          -          ## inserted before /cells/6:
-          -          +  code cell:-          --          -          ## deleted /cells/6-23:
-          -          -  code cell:-          -          -    execution_count: 1-          -          -    source:-          -          -
-          -          -          @@ -1,4 +1,4 @@
-          -          -           ### Control de versiones
```

In [2]: ! nbdiff-web apolo_charla.ipynb apolo_charla_inicial.ipynb

[I nbdimeserver:375] Listening on 127.0.0.1, port 34909
[I webutil:29] URL: http://127.0.0.1:34909/diff?base=apolo_charla.ipynb&remote=apolo_charla_inicial.ipynb
[I nbdimeserver:304] Closing server on remote request

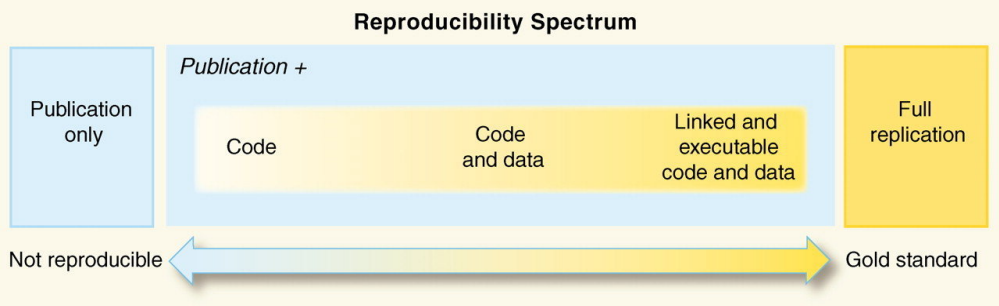


Diagrama de reproducibilidad

1.2.3 Compartir

- NBViewer.
- Github.

1.2.4 Colaboración

- CoCalc.
- Authorea.
- Colaboratory.

1.3 Flujo de trabajo

- Organización.
- Documentación.
- Automatización.
- Difusión y publicación.

1.3.1 Publicación

Generación con nbconvert (línea de comandos o interfaz gráfica).

In [3]:
```
%%bash
jupyter nbconvert apolo_charla.ipynb --to pdf
okular apolo_charla.pdf
```

```
[NbConvertApp] Converting notebook apolo_charla.ipynb to pdf
[NbConvertApp] Writing 297209 bytes to notebook.tex
[NbConvertApp] Building PDF
[NbConvertApp] Running xelatex 3 times: ['xelatex', 'notebook.tex']
[NbConvertApp] Running bibtex 1 time: ['bibtex', 'notebook']
[NbConvertApp] WARNING | bibtex had problems, most likely because there were no citations
```

```
[NbConvertApp] PDF successfully created
[NbConvertApp] Writing 72937 bytes to apolo_charla.pdf
qt5ct: using qt5ct plugin
Invalid Context= "Panel" line for icon theme:  "/usr/share/icons/Numix/16/panel/"
Invalid Context= "Panel" line for icon theme:  "/usr/share/icons/Numix/16@2x/panel/"
Invalid Context= "Panel" line for icon theme:  "/usr/share/icons/Numix/22/panel/"
Invalid Context= "Panel" line for icon theme:  "/usr/share/icons/Numix/22@2x/panel/"
Invalid Context= "Panel" line for icon theme:  "/usr/share/icons/Numix/24/panel/"
Invalid Context= "Panel" line for icon theme:  "/usr/share/icons/Numix/24@2x/panel/"
Invalid Context= "Notifications" line for icon theme:  "/usr/share/icons/Numix/48/notifications/"
Invalid Context= "Notifications" line for icon theme:  "/usr/share/icons/Numix/48@2x/notifications/"
qt5ct: D-Bus global menu: no
Invalid Context= "Panel" line for icon theme:  "/usr/share/icons/Numix/16/panel/"
Invalid Context= "Panel" line for icon theme:  "/usr/share/icons/Numix/16@2x/panel/"
Invalid Context= "Panel" line for icon theme:  "/usr/share/icons/Numix/22/panel/"
Invalid Context= "Panel" line for icon theme:  "/usr/share/icons/Numix/22@2x/panel/"
Invalid Context= "Panel" line for icon theme:  "/usr/share/icons/Numix/24/panel/"
Invalid Context= "Panel" line for icon theme:  "/usr/share/icons/Numix/24@2x/panel/"
Invalid Context= "Notifications" line for icon theme:  "/usr/share/icons/Numix/48/notifications/"
Invalid Context= "Notifications" line for icon theme:  "/usr/share/icons/Numix/48@2x/notifications/"
```

In [4]:
```
%%bash
jupyter nbconvert apolo_charla.ipynb --to slides
```

```
[NbConvertApp] Converting notebook apolo_charla.ipynb to slides
[NbConvertApp] Writing 631767 bytes to apolo_charla.slides.html
```

1.4 Exploración

Algunos paquetes para integrar:

- ipywidgets y widgetsnbextension.
- Matplotlib. Con la misma finalidad, bokeh y plotly.
- 3D: mayavi, pythreejs, ipyvolume.
- Mapas: ipyleaflet.
- Tablas de datos: pandas.

Se recomienda el uso de Anaconda Python para la instalación de dependencias.

conda install -y python=3.6 notebook nbconvert matplotlib ipywidgets widgetsnbextension ipyleaflet ipyvolume num
c conda-forge

In [5]:
```
import ipyvolume as ipv
import numpy as np
x, y, z = np.random.random((3, 10000))
ipv.quickscatter(x, y, z, size=1, marker="sphere")
```

```
VBox(children=(Figure(camera_center=[0.0, 0.0, 0.0], height=500, matrix_projection=[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0
```

In [6]:
```
from ipyleaflet import (
    Map,
    Marker,
    TileLayer, ImageOverlay,
    Polyline, Polygon, Rectangle, Circle, CircleMarker,
    GeoJSON,
    DrawControl
)

center = [ 6.199548, -75.57934]
zoom = 100

m = Map(center=center, zoom=zoom)
m
```

```
/home/cosmoscalibur/miniconda3/envs/flujo/lib/python3.6/importlib/_bootstrap.py:219: RuntimeWarning: num
  return f(*args, **kwds)
/home/cosmoscalibur/miniconda3/envs/flujo/lib/python3.6/importlib/_bootstrap.py:219: RuntimeWarning: num
  return f(*args, **kwds)
```

```
Map(basemap={'url': 'https://{s}.tile.openstreetmap.org/{z}/{x}/{y}.png', 'max_zoom': 19, 'attribution': 'Map
```

In [7]:
```
import pandas as pd
```

In [8]:
```
dispositivos = pd.read_csv('exa__dispositivos_cali.csv')
dispositivos
```

Out[8]:

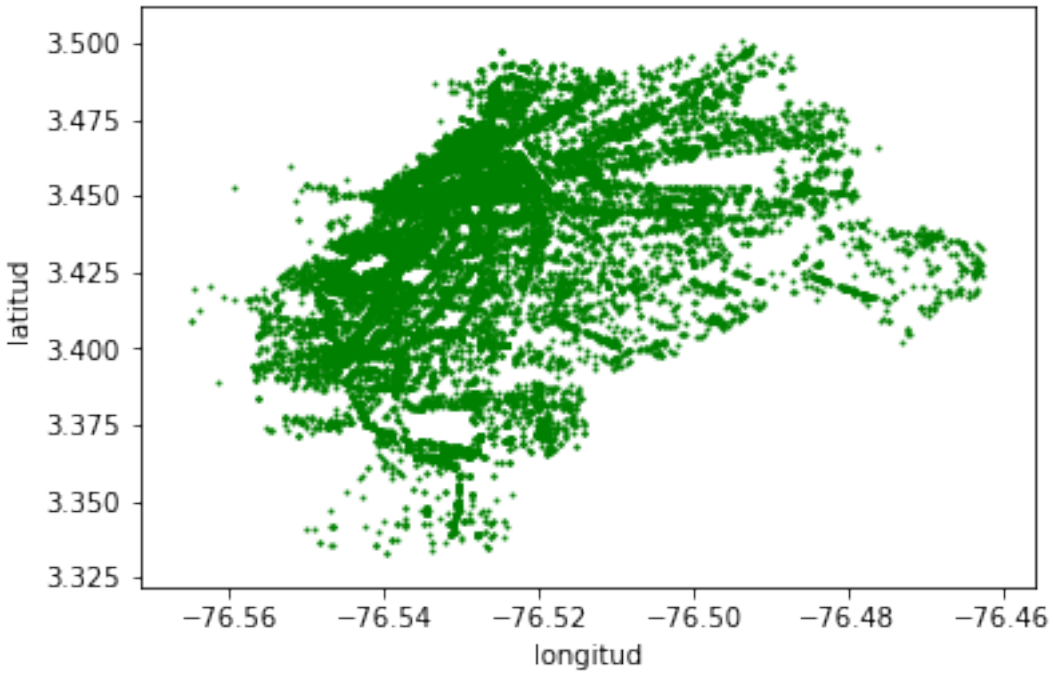
	tipo	codigo	latitud	longitud	id_barrio
0	POS	1024702	3.451135	-76.530893	737998832
1	POS	1076402	3.446585	-76.517672	738000381
2	POS	1077002	3.446585	-76.517672	738000381
3	POS	3342404	3.476909	-76.485286	737998905
4	POS	3378003	3.485575	-76.516418	737998174
5	POS	3445704	3.432185	-76.501997	738000399
6	POS	3464704	3.396439	-76.509977	737996909
7	POS	3511004	3.436057	-76.526317	737999435
8	POS	3655304	3.473362	-76.520498	737998812
9	POS	3680504	3.402724	-76.511339	737996914
10	POS	3803003	3.452645	-76.514656	737998761

11	POS	3831404	3.494200	-76.487790	737998848
12	POS	3906604	3.385157	-76.538646	738000366
13	POS	3920804	3.451854	-76.529130	737998832
14	POS	5013602	3.450511	-76.540392	737998685
15	POS	10004513	3.467342	-76.524312	737998251
16	POS	10004612	3.461620	-76.526902	737998204
17	POS	10005031	3.449755	-76.531958	737998832
18	POS	10295574	3.461991	-76.531079	737998194
19	POS	10295798	3.462113	-76.531073	737998194
20	POS	10296309	3.462169	-76.526489	737998251
21	POS	10300408	3.470768	-76.529413	737998194
22	POS	10301521	3.467718	-76.526981	737998251
23	POS	10301638	3.466495	-76.527735	737998251
24	POS	10302479	3.480543	-76.516562	737998219
25	POS	10304228	3.450075	-76.531170	737998832
26	POS	10304525	3.451482	-76.528353	737998935
27	POS	10304616	3.449040	-76.531869	737998832
28	POS	10305555	3.453316	-76.531538	737998832
29	POS	10306629	3.450474	-76.532197	737998832
...	...	...	...	...	...
37254	POS	15720857	3.402785	-76.541402	738001440
37255	POS	15721194	3.406897	-76.530289	737998942
37256	POS	15765464	3.400436	-76.533783	737998940
37257	POS	15765704	3.386891	-76.539669	737998851
37258	POS	20010202	3.448232	-76.532914	737998195
37259	POS	23072102	3.391888	-76.538562	737998851
37260	POS	23197002	3.432575	-76.542913	737999403
37261	POS	23197003	3.432575	-76.542913	737999403
37262	POS	32030303	3.449459	-76.542771	737998685
37263	POS	32139103	3.450585	-76.532257	737998832
37264	POS	48000202	3.463012	-76.530490	737998194
37265	POS	50013203	3.400229	-76.546695	737996829
37266	POS	50019103	3.461183	-76.531067	737998259
37267	POS	51185603	3.414325	-76.547986	737998095
37268	POS	52438703	3.452426	-76.485687	737999402
37269	POS	52439503	3.477415	-76.526735	737998252
37270	POS	53230903	3.481212	-76.498708	737998838
37271	POS	54807202	3.451628	-76.532053	737998832
37272	POS	60011902	3.451891	-76.532957	737998832
37273	POS	60141103	3.394895	-76.546299	737996834
37274	POS	73329603	3.371983	-76.539279	738000379
37275	POS	73332203	3.435905	-76.519890	737998834
37276	POS	73459503	3.414528	-76.508589	737996735
37277	POS	74193403	3.370904	-76.528167	738000407
37278	POS	74256303	3.483556	-76.499585	737998838
37279	POS	74486603	3.452831	-76.523627	737998935
37280	POS	74632603	3.450660	-76.533361	737998832
37281	POS	4023000120	3.450063	-76.530697	737998832

```
37282  POS  4408002741  3.434776 -76.543102  737999403
37283  POS  4023002563  3.414809 -76.546871  737998095

[37284 rows x 5 columns]
```

```
In [10]: dispositivos.plot.scatter(x='longitud', y='latitud', s=1, c='g')
         import matplotlib.pyplot as plt
         plt.show()
```



1.5 Bibliografía

- The Jupyter Notebook. https://jupyter-notebook.readthedocs.io/en/stable/notebook.html
- Interactive workflows for C++ with Jupyter. https://blog.jupyter.org/interactive-workflows-for-c-with-jupyter-fe9b54227d92
- Reproducibility, replicability, and the two layers of computational science. https://thewinnower.com/papers/reproducibility-replicability-and-the-two-layers-of-computational-science
- Interactive notebooks: Sharing the code. https://www.nature.com/news/interactive-notebooks-sharing-the-code-1.16261?WT.ec_id=NATURE-20141106
- Aula notebook (repositorio). https://github.com/cosmoscalibur/aula-notebook (este notebook se dispondrá en el repositorio y podría actualizarse).