

Adatvizualizáció Pythonnal



Eszközök

Eszközök

_Python (www.python.org)

_Jupyter (jupyter.org)

_Pandas (pandas.pydata.org)

_Altair (altair-viz.github.io)



A Python nyelv

A Python nyelv

_egyszerű és barátságos

_az olvashatóság fontos szempont

_data science-ben elterjedt, felváltani látszik az R-t

_nem a kígyóról, hanem a Monty Pythonról kapta a nevét

Python: szintaxis

```
import sys
# this is a comment

title = 'Sir'
knights = ['Galahad', 'Lancelot', 'Robin']

def dub(name):
    if name == 'Robin':
        prefix = 'Brave '
    else:
        prefix = ''
    return '{}{} {}'.format(prefix, title, name)

for k in knights:
    print(dub(k)) # inline comment

sys.exit(0)
```


Python: szintaxis



```
import sys
# this is a comment

title = 'Sir'
knights = ['Galahad', 'Lancelot', 'Robin']

def dub(name):
    prefix = 'Brave ' if name == 'Robin' else ''
    return '{}{} {}'.format(prefix, title, name)

for k in knights:
    print(dub(k)) # inline comment

sys.exit(0)
```

Significant whitespace



- _a behúzás a szintaxis része

- _mindegy, hogy tab vagy space karaktert használunk...

- _...de nem lehet keverni!

- _egy szinten pontosan ugyanannyi legyen a behúzás

- _az ajánlás 4 space karakter szintenként

Alapvető típusok

`_int`: egész szám

`_float`: lebegőpontos szám

`_bool`: logikai (True, False)

`_string`: karaktersor

Alapvető típusok



_None: érték hiányának kifejezésére szolgál

_list: lista (tömb)

_tuple: hasonló a listához, de nem módosítható

_dictionary: kulcs-érték párok

List comprehension



Új lista létrehozása létező alapján, ciklussal:

```
numbers = [1, 2, 3]
doubles = []
for n in numbers:
    doubles.append(n * 2)
```

List comprehension



Idiomatikus Python:

```
numbers = [1, 2, 3]  
doubles = [n * 2 for n in numbers]
```



Jupyter

Jupyter



<https://jupyter.org>

_interaktív programozási környezet, amely egyszerre tudja megjeleníteni a kódot és annak eredményét

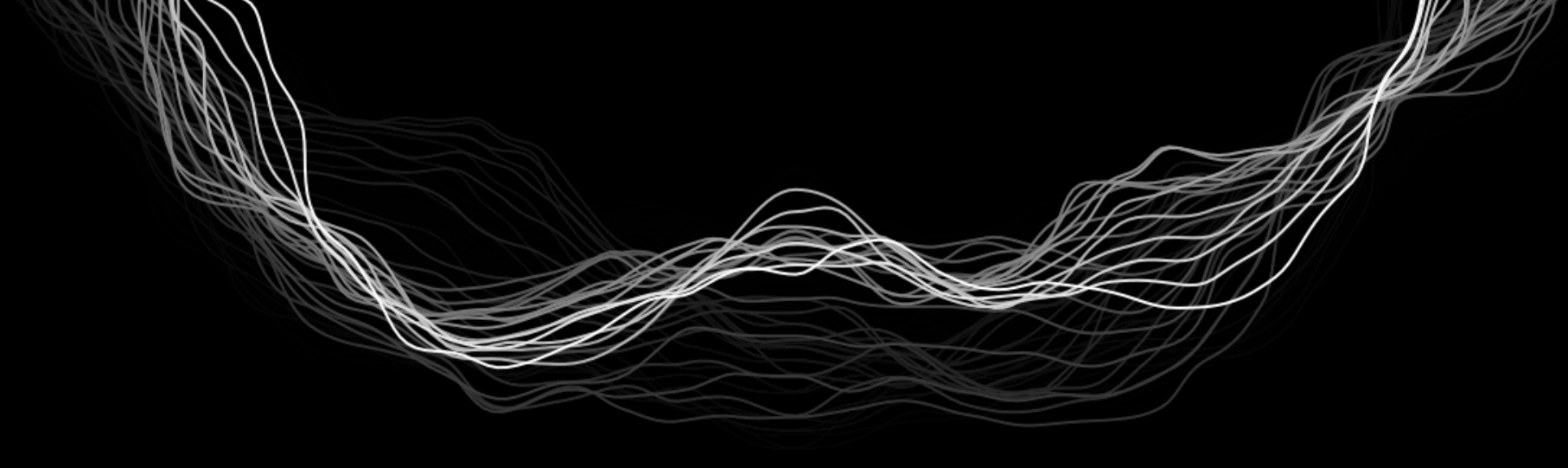
_használható weben, lokálisan browserben vagy egyéb szoftverrel (pl. VS Code)

_a notebookok megoszthatók (pl. GitHub, Binder)



Pandas

Pandas



_adatelemző és -manipulációs eszköz

_különböző forrásokból származó adattáblák olvasására,
szűrésére, feldolgozására, kombinálására használható

_„Excel Pythonban”

Adatok beolvasása

_fajltípusok: CSV, Excel, JSON, SPSS, HTML...

_elérés: lokális és távoli – HTTP, FTP, S3, SQL

_read_csv, read_excel, stb. függvények

Adatstruktúrák



_ **Series**: egydimenziós indexelt tömb

_ **DataFrame**: kétdimenziós indexelt tömb, tetszőleges típusú oszlopokkal

_ a DataFrame felfogható táblázatként, a Series pedig ennek egy sora vagy oszlopa.

Műveletek adatokkal

_kiválasztás sor, oszlop, logikai feltétel alapján

_adatbázis stílusú műveletek: merge, join

_statisztikai funkciók

_szűrés, aggregálás

_rengeteg egyéb, ld.

https://pandas.pydata.org/docs/user_guide/

Műveletek adatokkal

Notebook:

<https://mybinder.org/v2/gh/nandordevai/python-dataviz.git/main>



Altair

Altair

Vega-Lite alapú

_tisztá, expresszív API

_UW Interactive Data Lab fejlesztése

_nyelvfüggetlen, JSON alapú

_Altair → Vega-Lite → Vega → D3

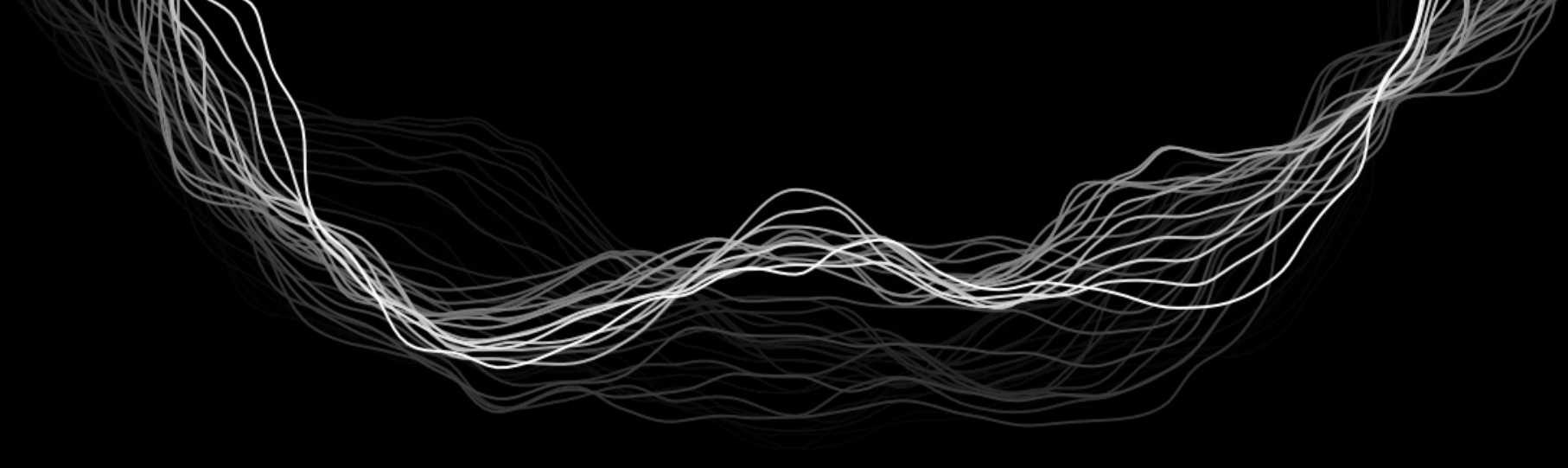
Altair



Marks & encodings

- _mark: hogyan jelenjenek meg a vizuális elemek (diagramtípus)
- _encoding: adatalemek összekapcsolása vizuális tulajdonságokkal
- _encoding channels: pozíció, szín, méret, forma, stb.

Altair



Testreszabási lehetőségek

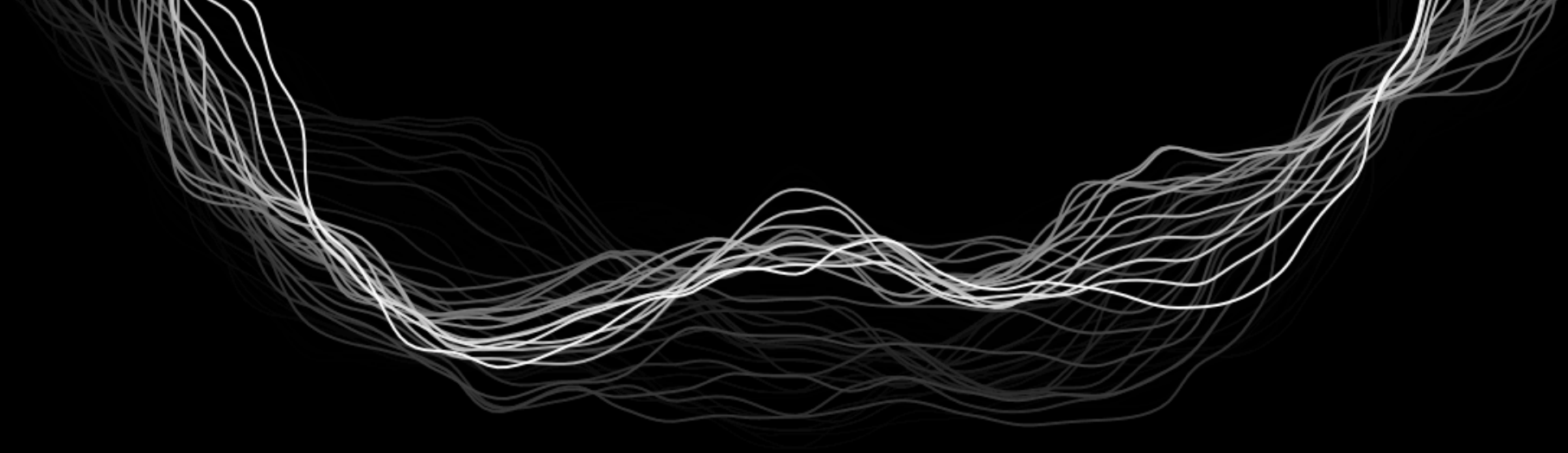
_globális: diagramra vonatkozó

_lokális: markra vonatkozó

_encoding channel

_érdemes először a legspecifikusabbat használni

Altair



Összetett diagramok

_átfedő

_összefűzött

_ismétlődő

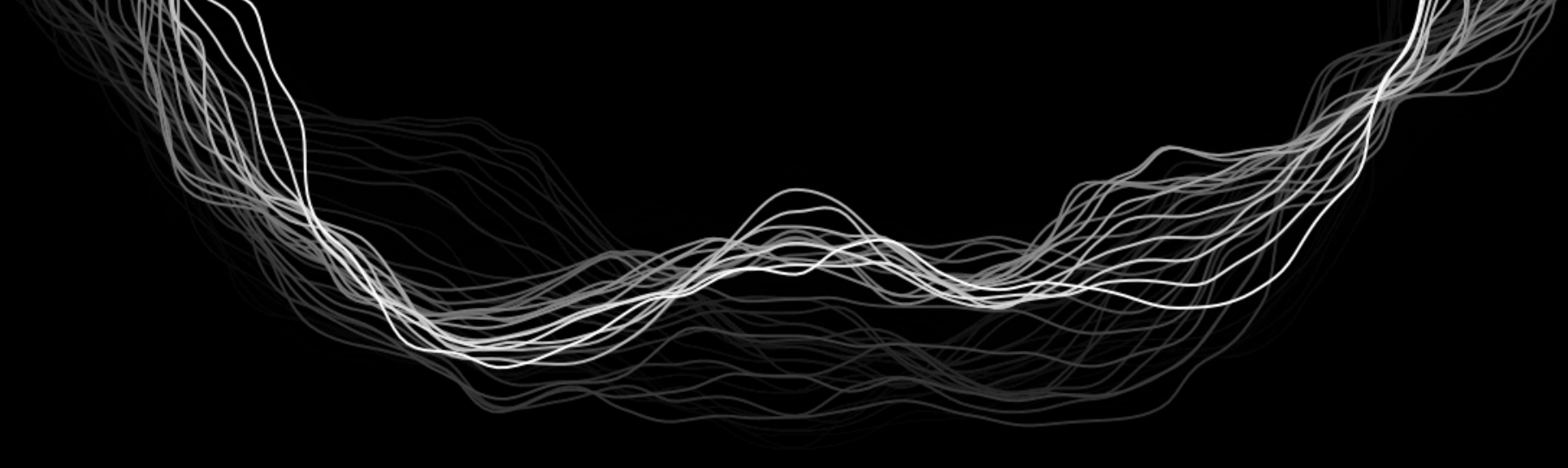
Altair

Interaktivitás

`_tooltipek`

`_kiválasztás`

`_brush` (kapcsolt diagramok)



Altair



Notebook:

<https://mybinder.org/v2/gh/nandordevai/python-dataviz.git/main>

Exportálás

_teljes notebook: HTML, PDF

_diagram: HTML, SVG, PNG

Források

_dokumentáció: <https://altair-viz.github.io>

_Vega-Lite: <https://vega.github.io/vega-lite/>

_példák: https://github.com/altair-viz/altair_notebooks