

Adatvizualizáció Pythonnal

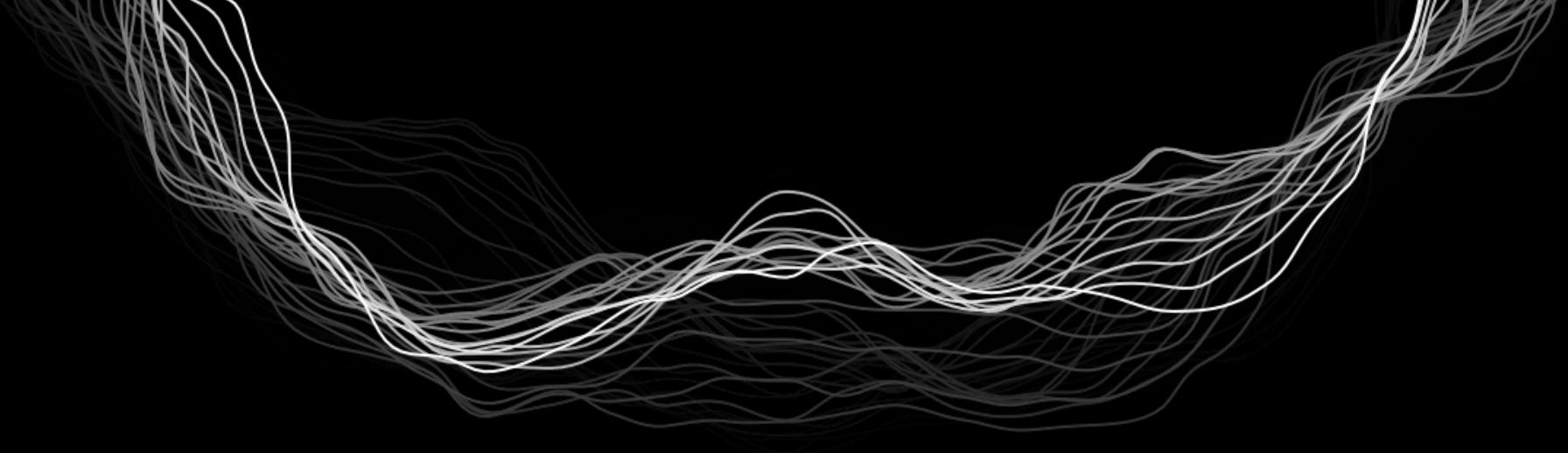
MOME Open
Adatvizualizáció, 2020.

Dévai Nándor
mail@nandordevai.net



Bemutatkozás

Bemutatkozás



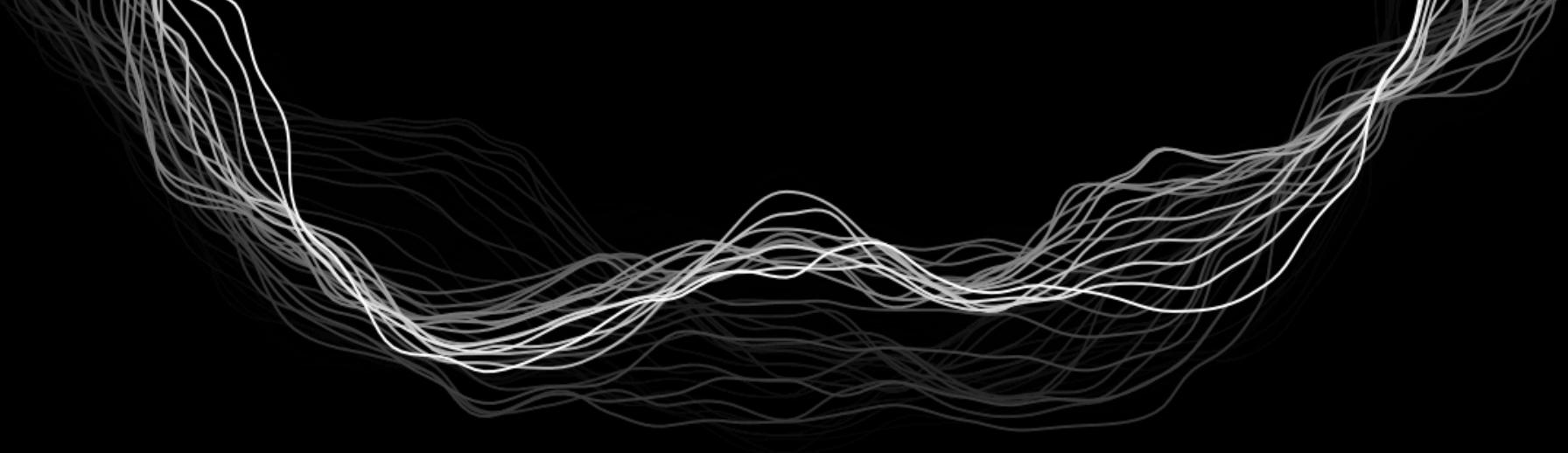
_MOME Kreatív Technológia Hub

_20 éves webes fejlesztői múlt

_frontend fejlesztés, interfészek

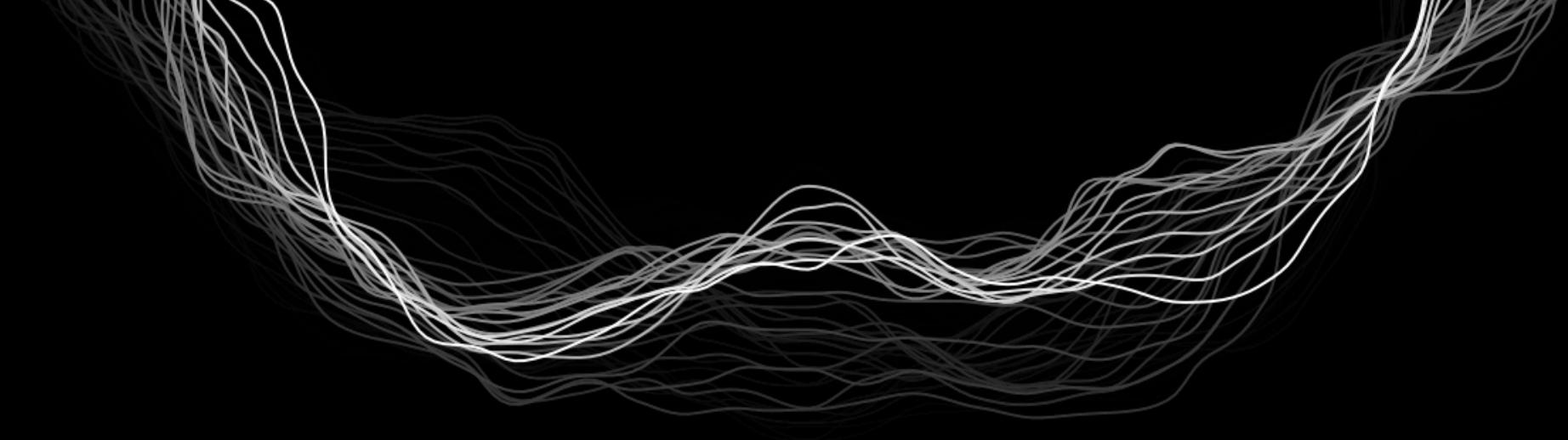
_adatvizualizációk készítése

_kreatív kód, elektronikus zene, generatív algoritmusok,
3D nyomtatás, Arduino...

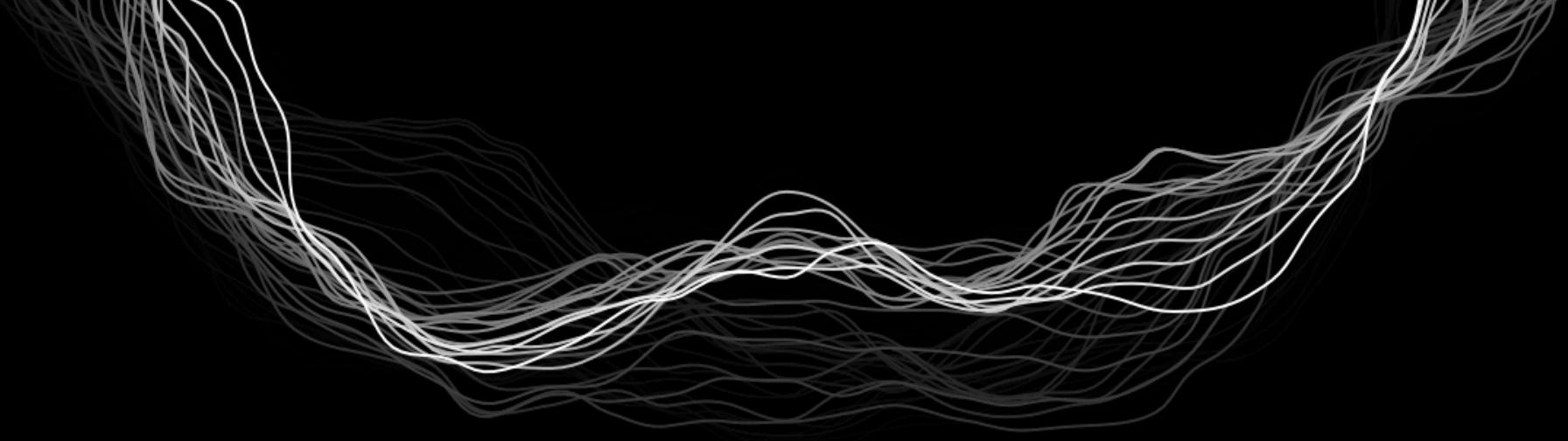


Eszközök

Eszközök

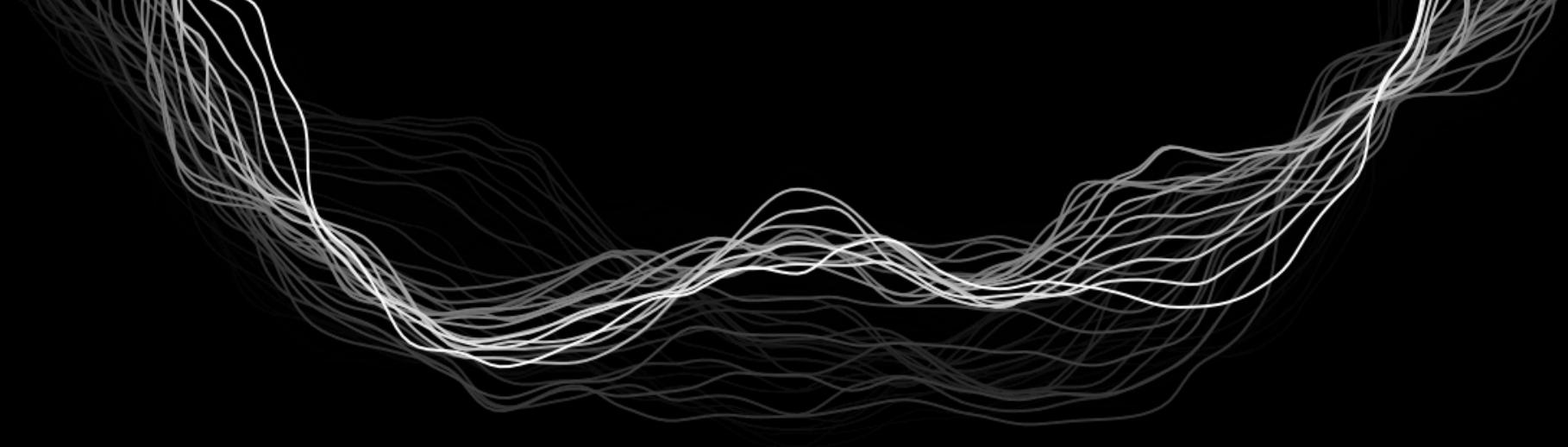


- _ Python (www.python.org)
- _ Jupyter (jupyter.org)
- _ Pandas (pandas.pydata.org)
- _ Altair (altair-viz.github.io)



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

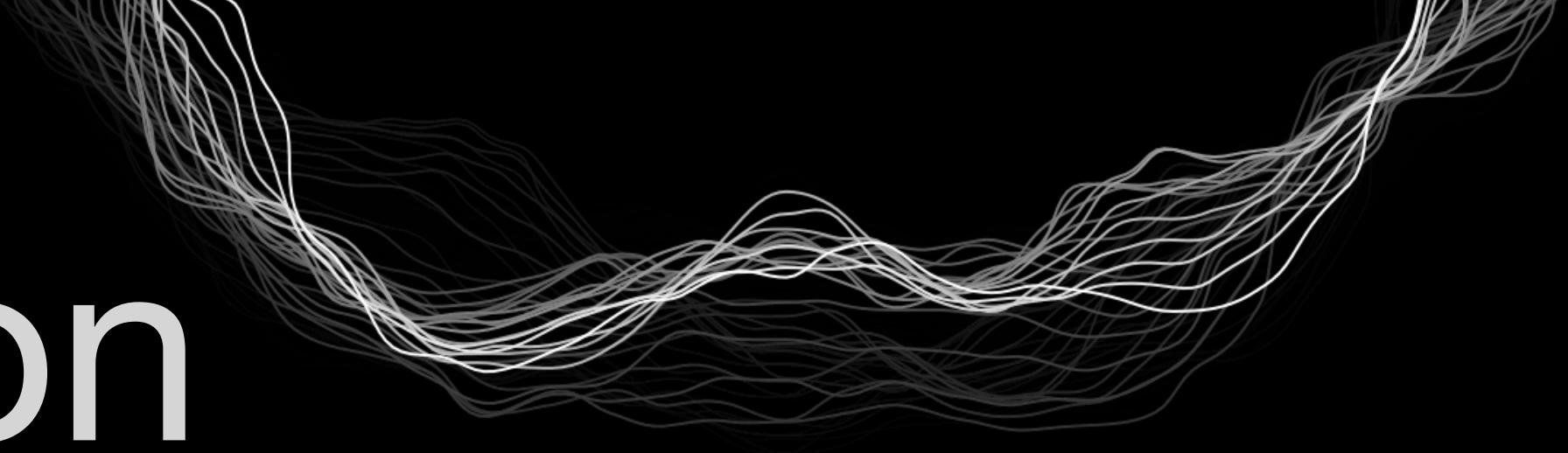
_bool: logikai (True, False)

_string: karaktersor

Alapvető típusok

- _None: érték hiányának kifejezésére szolgál
- _list: lista (tömb)
- _tuple: hasonló a listához, de nem módosítható
- _dictionary: kulcs-érték párok

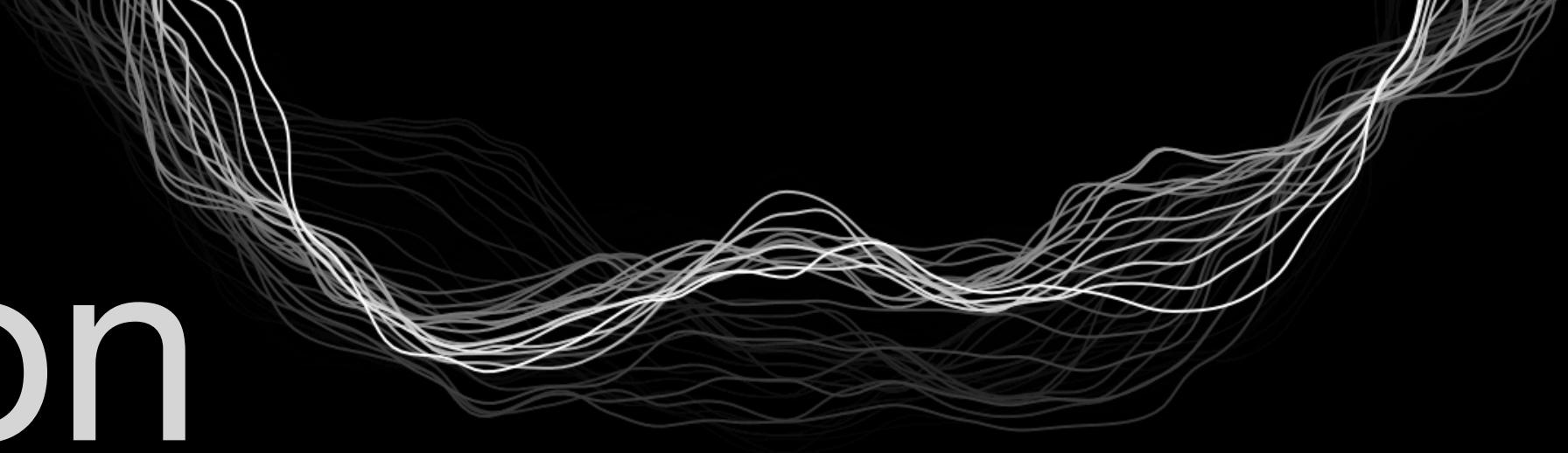
List comprehension



Új lista létrehozása létező alapján, ciklussal:

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

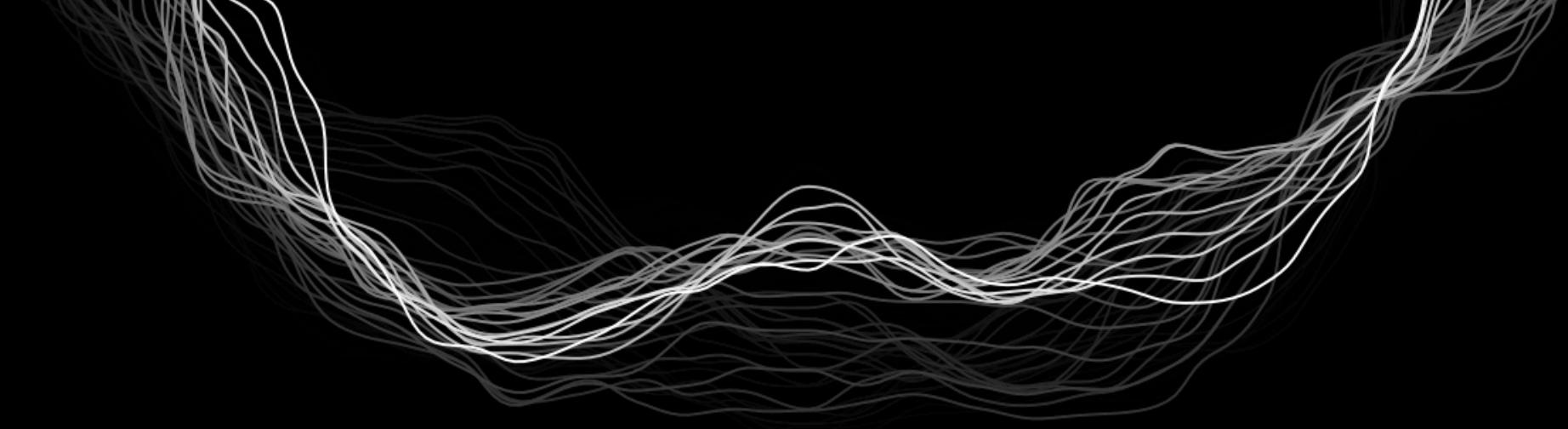
List comprehension



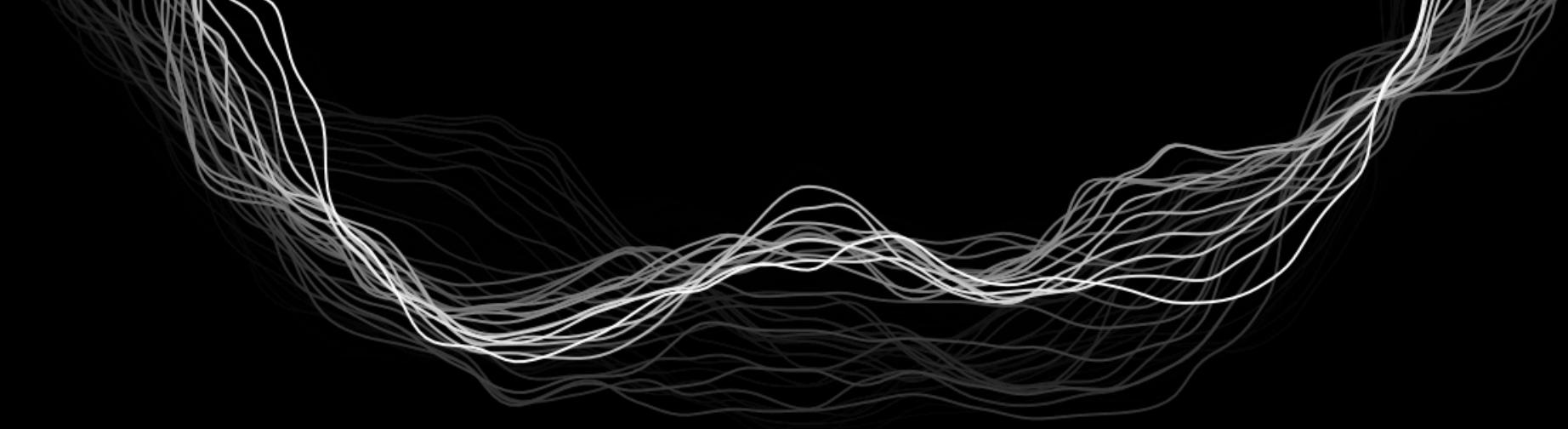
Idiomatikus Python:

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

Jupyter



Jupyter

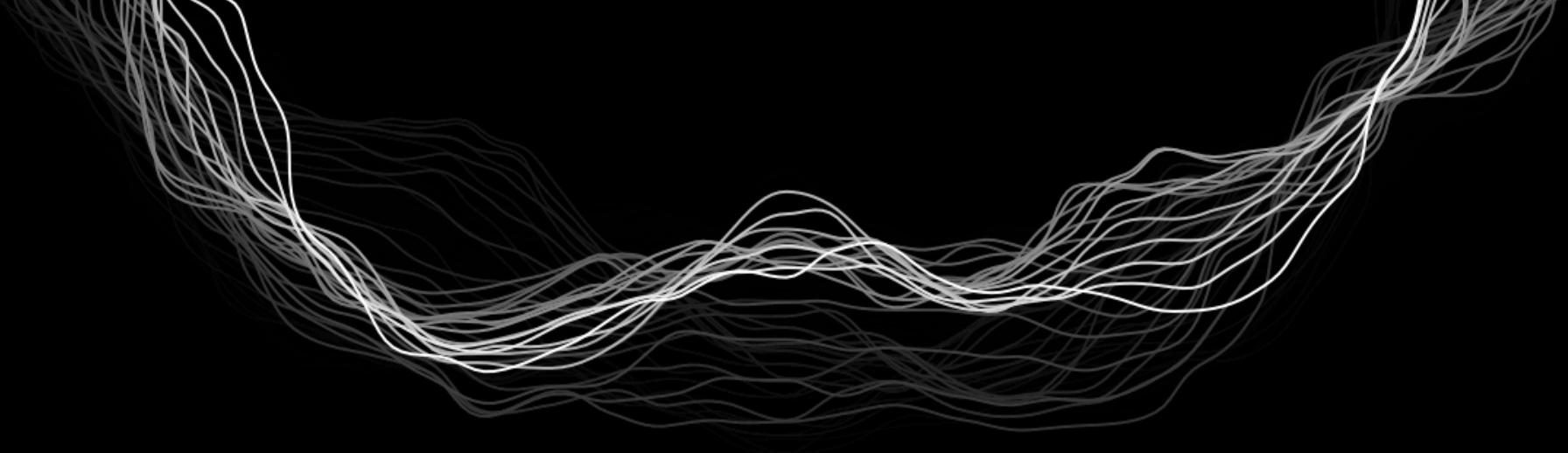


_ <https://jupyter.org>

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

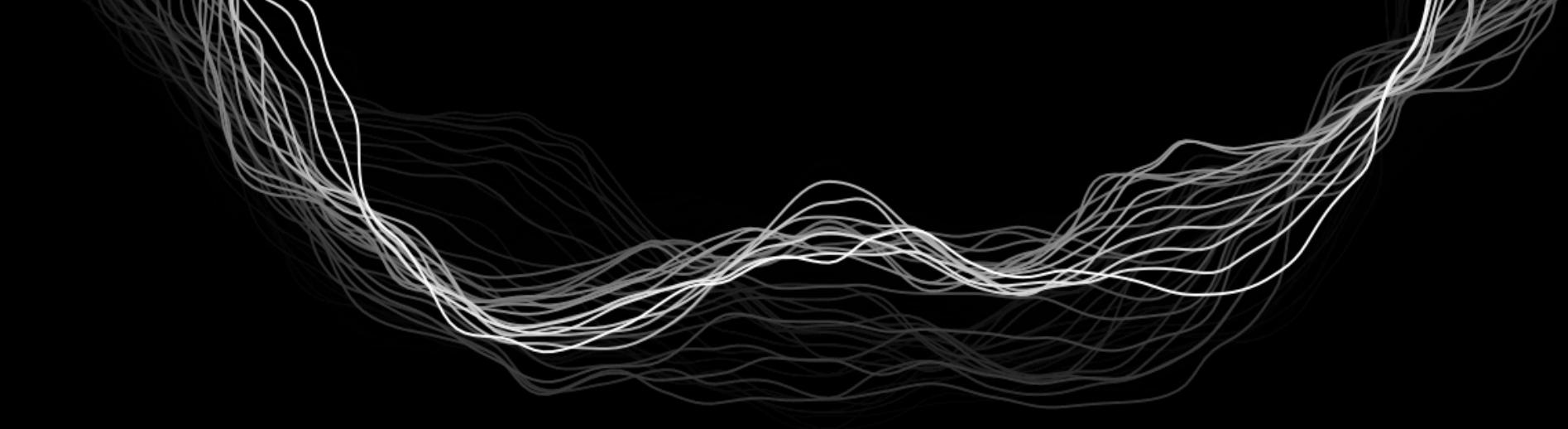
_ használható weben, lokálisan browserben vagy egyéb szoftverrel (pl. VS Code)

_ 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

_fájltípusok: 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

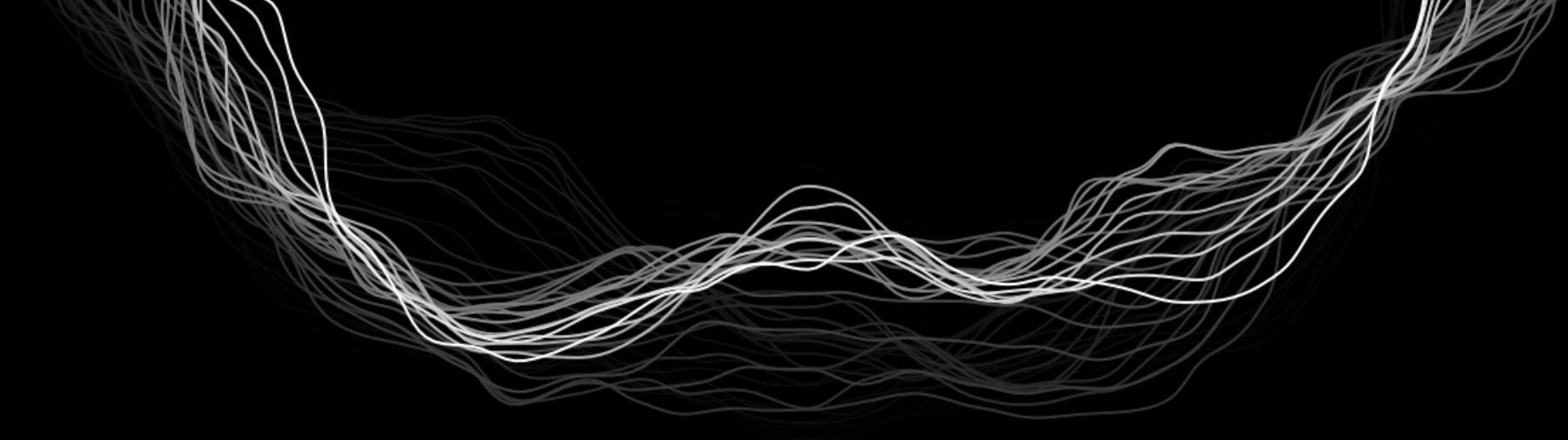
_statisztikai funkciók

_szűrés, aggregálás

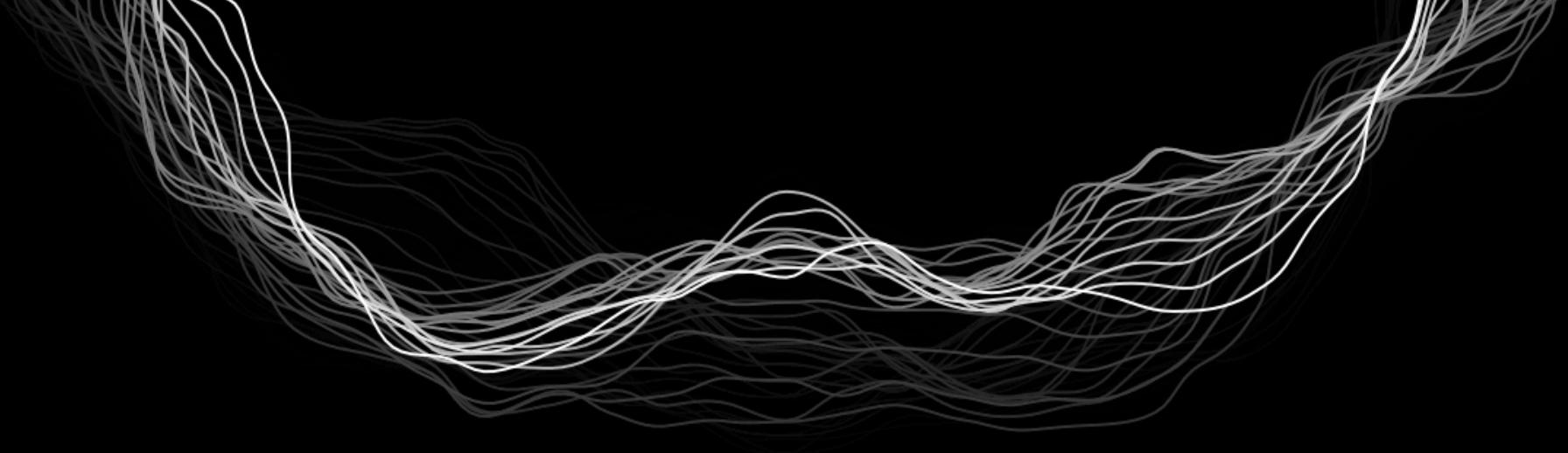
_rengeteg egyéb, ld.

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

Műveletek adatokkal

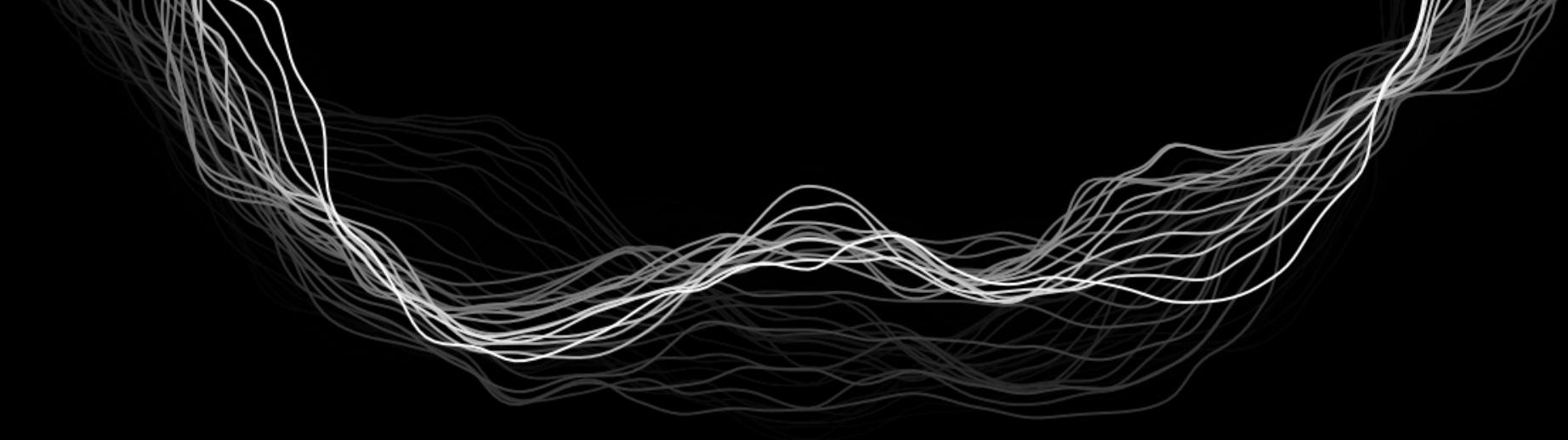


<https://mybinder.org/v2/gh/nandordevai/python-dataviz.git/main>



Altair

Altair



Vega-Lite alapú

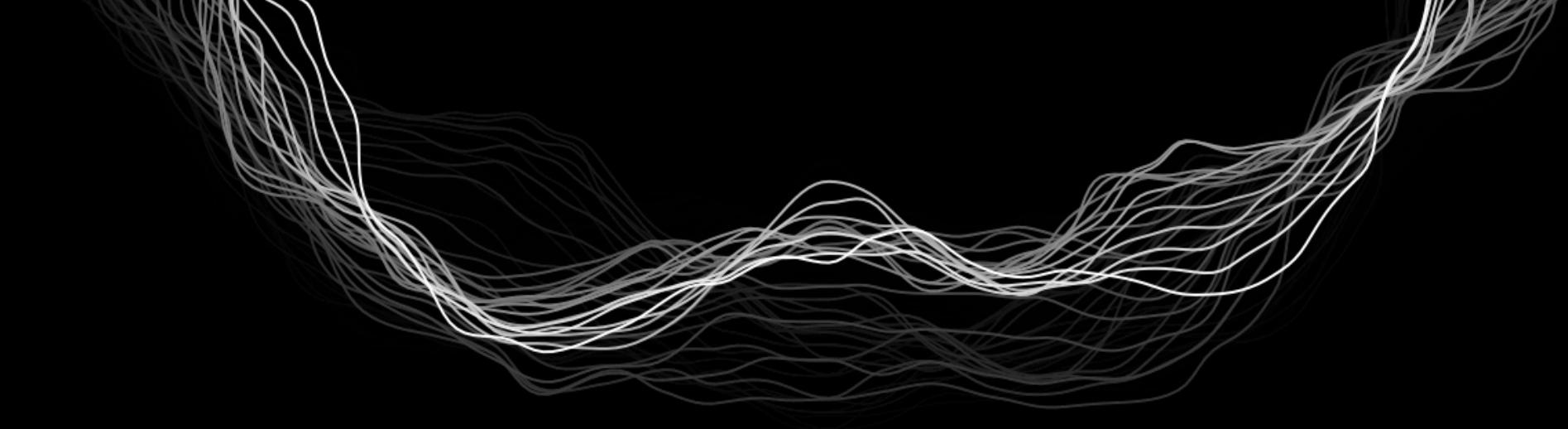
_tiszta, expresszív API

_UW Interactive Data Lab fejlesztése

_nyelvfüggetlen, JSON alapú

_Altair → Vega-Lite → Vega → D3

Altair



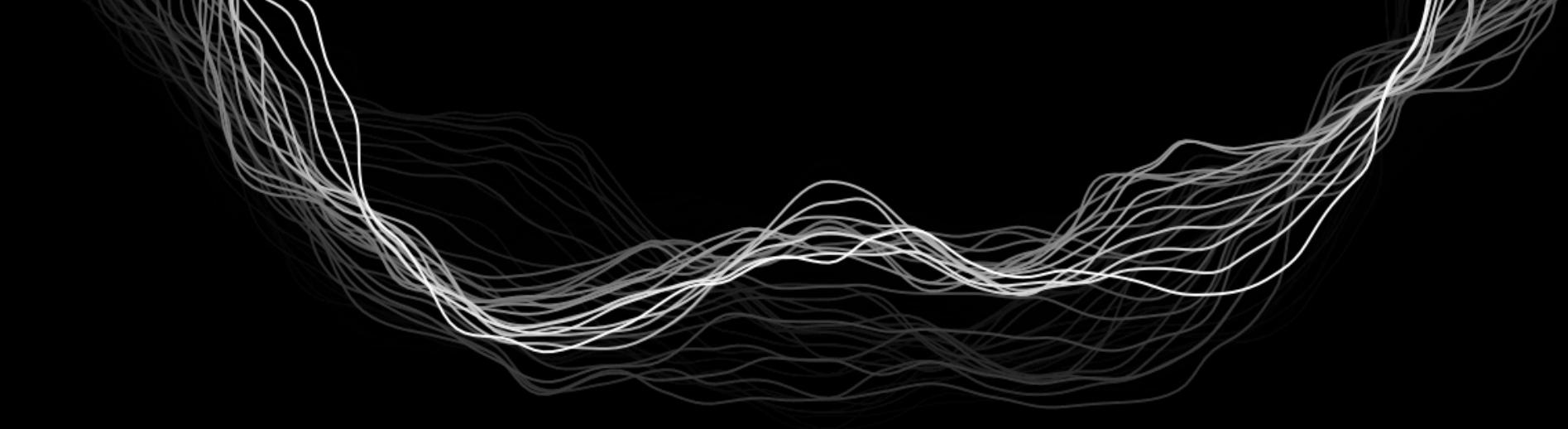
Marks & encodings

_mark: hogyan jelenjenek meg a vizuális elemek
(diagramtípus)

_encoding: adataelemek összekapcsolása vizuális
tulajdonságokkal

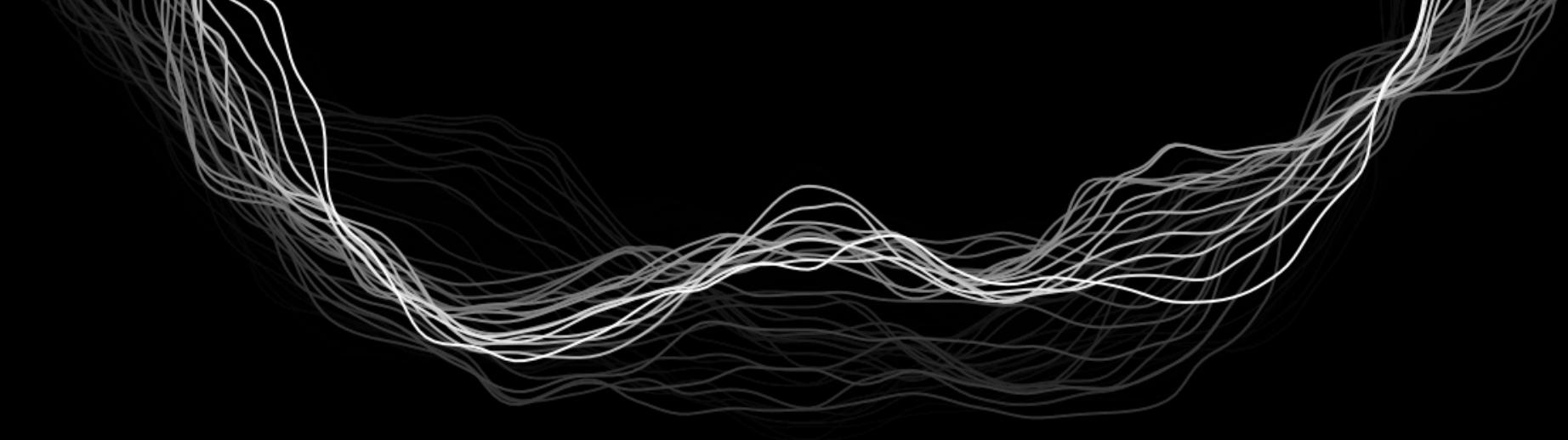
_encoding channels: pozíció, szín, méret, forma, stb.

Altair



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

Exportálás



_teljes notebook: HTML, PDF

_diagram: HTML, SVG, PNG