Interactive Data Kung Fu with Shaolin

Guillem Duran Ballester


@Miau_DB
Guillem-db

what is this about?

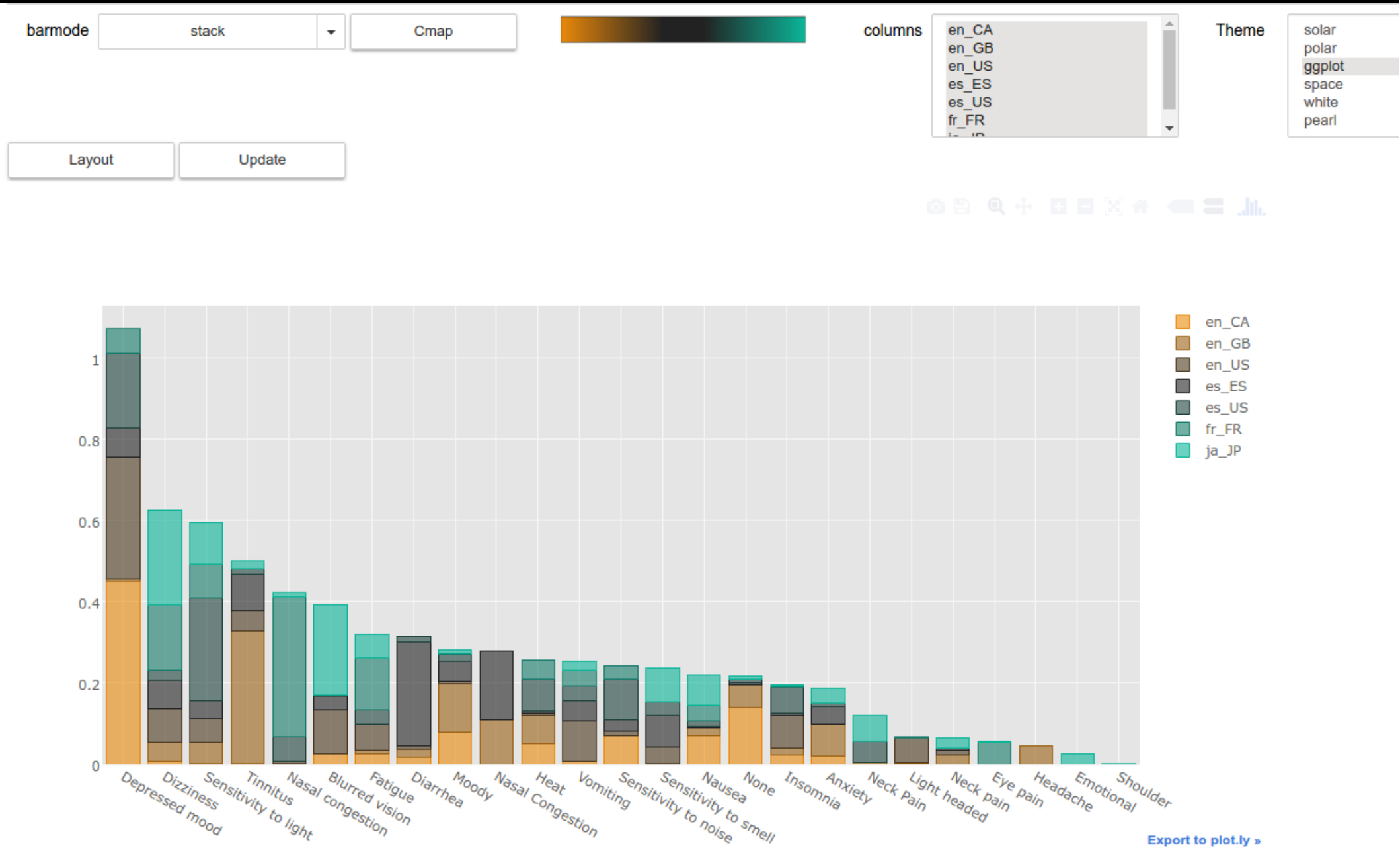
- Shaolin
- Not a Tutorial
- Introduction
  - Core capabilities
  - Demo videos
  - Call for help

How am I presenting it?

The Ways of Data Kung Fu!

1. Science
2. Martial arts
3. Hacking
4. Teaching

And dashboards!



SHAOLIN

Structure Helper for dAshbOard LINking


Helps you perform some
nice Data Kung Fu!

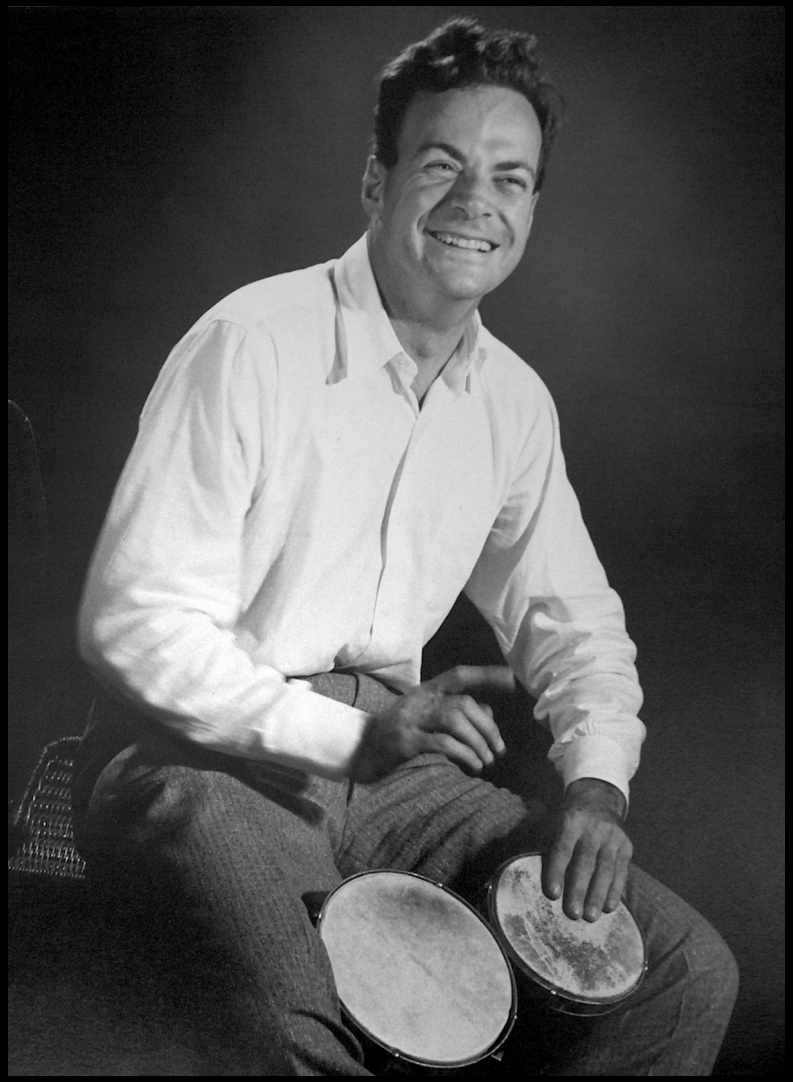

Data Kung(Man) Fu (Work)

What is it?

Just our way of working

Why?

Because Feynman!



WWFD?...

- If he had python
- If he had the Internet

"Study hard what interests you the most in the most undisciplined, irreverent and original manner possible."

Chapter 1: Science

“Science is the belief in the ignorance of experts.”

Think Science

“Religion is a culture of faith; science is a culture of doubt.”

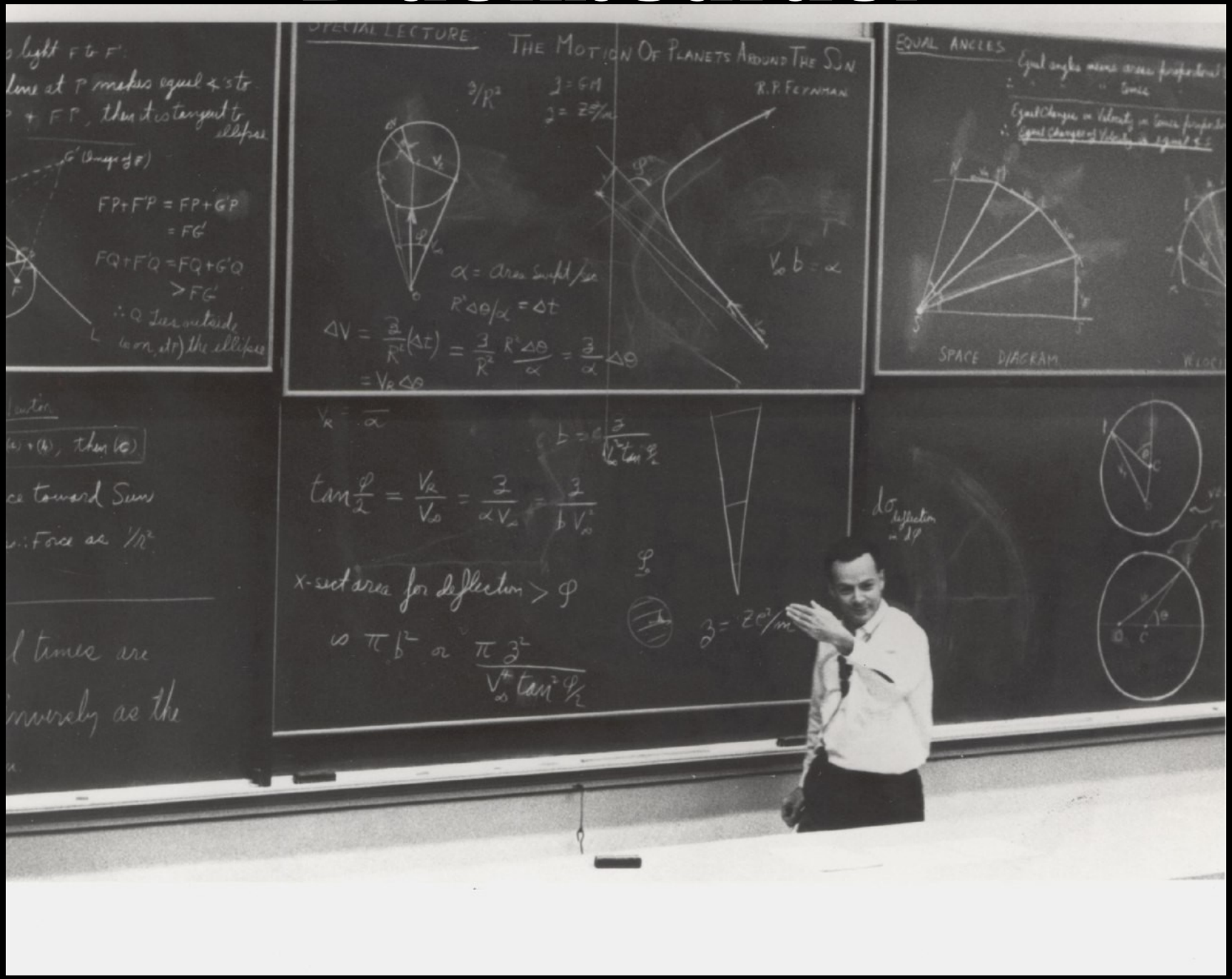
Doubt + experiment $\rightarrow$ Model

Pure math is outdated*

Python is a more clear abstraction to explain language + reasoning

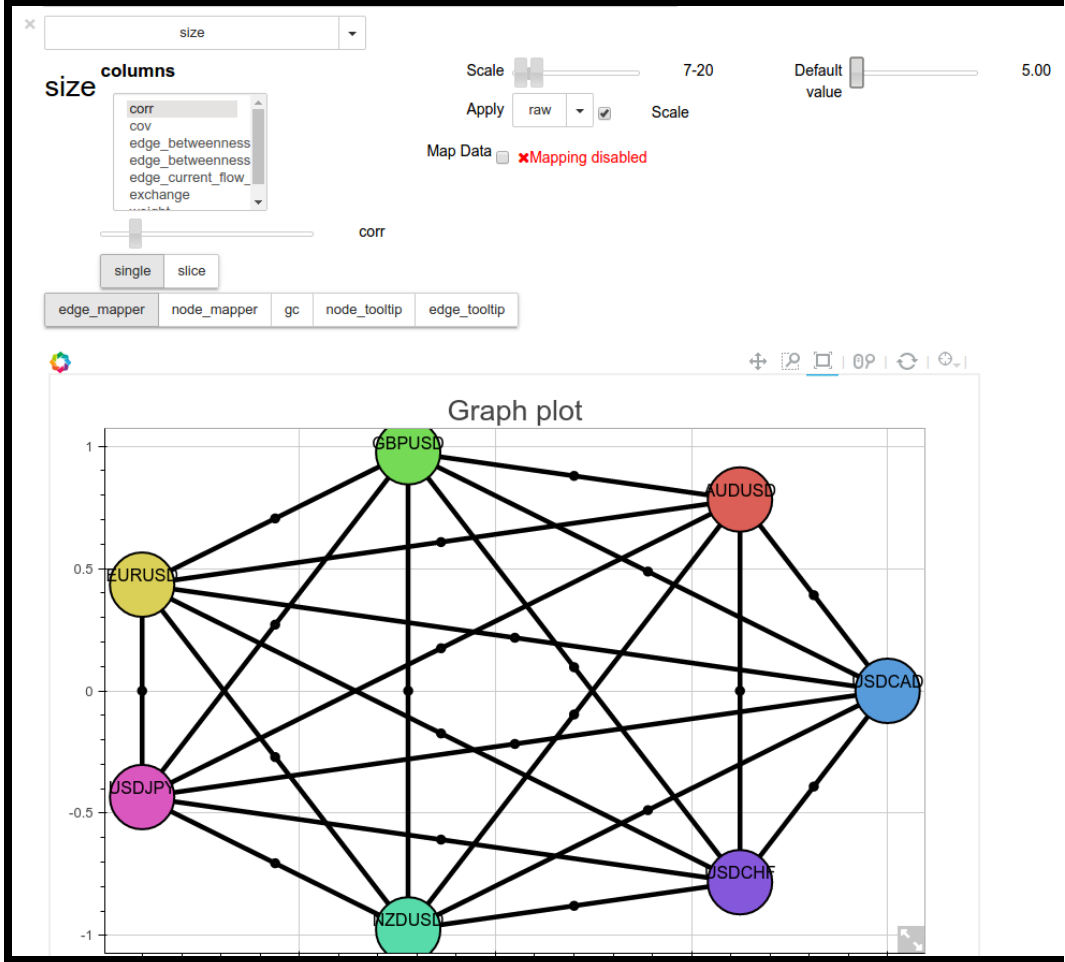
Write your models using objects

Blackboards?



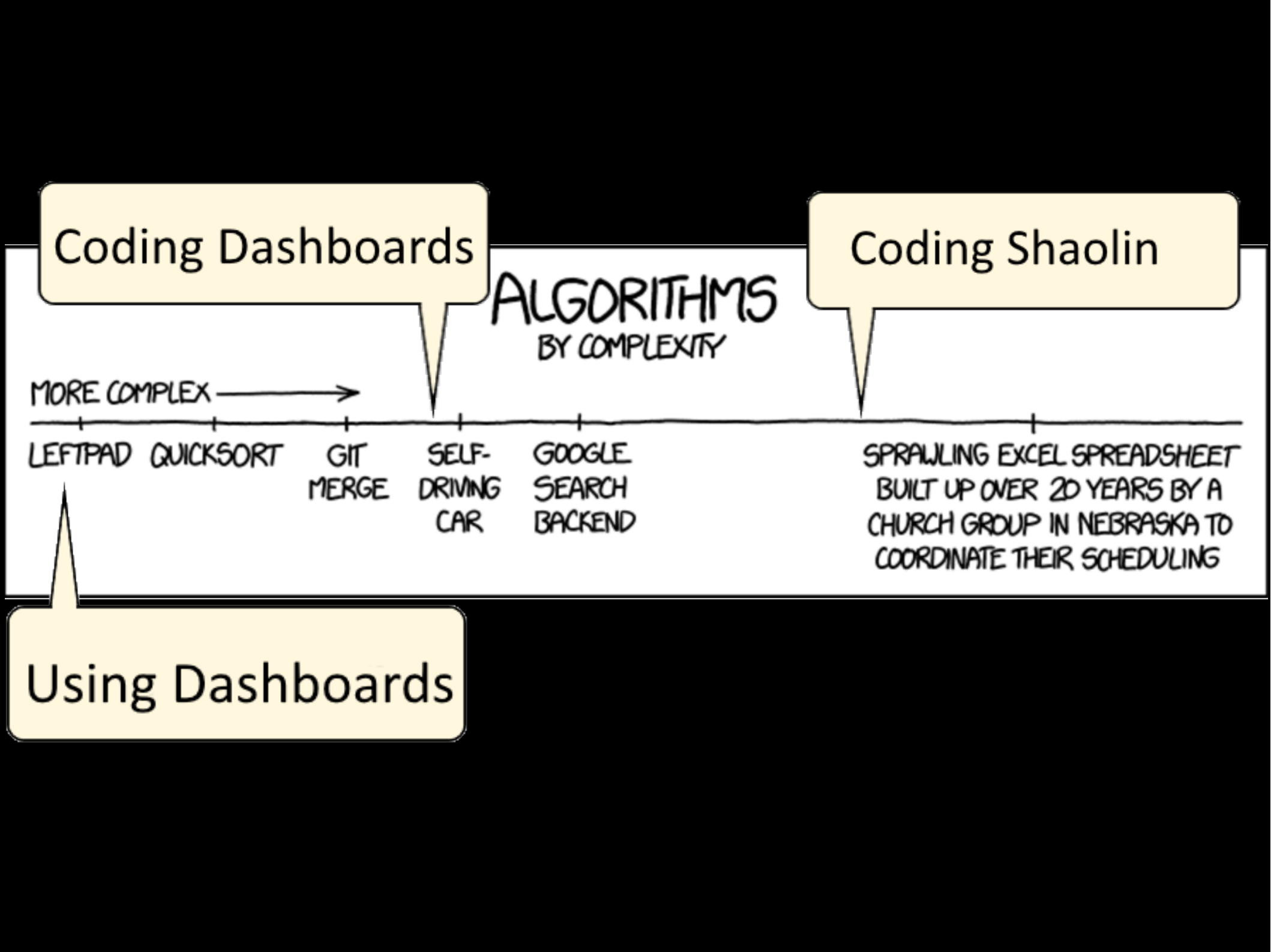
Dashboards!


IPs
IP
192.168.1.1
Ip_range
0-255
Single
Range
Add
delete
update
['192.168.1.0', '192.168.1.1', '192.168.1.2' , '...', '192.168.1.252', '192.168.1.253', '192.168.1.254']
Load results
Process data
barmode
stack
Cmap
Ports
port
443
Single
Range
Add
delete
[22, 80, 23, 443]
File name
trap.html
Notebook
Export
Scan
Html mode
columns
ip
max
min
mean
std
alive
Theme
solar
ggplot
polar
white
space
pearl
Layout
Title
MAGICS!
Paper bgcolor
#fff
height
650
width
1450
Plot bgcolor
#f0f0fc
layout_params
xaxis
yaxis
margin
Layout
Update
Data
Results
By IP
Visualize
Bars
Graph
Title font
Family
Ume Gothic S5
KacstFarsi
Webdings
Ume P Gothic
KacstDecorative
Impact
Size
26
Color
#249e94
MAGICS!

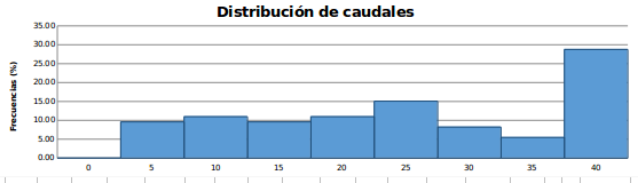

size
size
columns
corr
cov
edge_betweenness
edge_betweenness
edge_current_flow_
exchange
Scale
7-20
Default value
5.00
Apply
raw
Scale
Map Data
Mapping disabled
corr
single
slice
edge_mapper
node_mapper
gc
node_tooltip
edge_tooltip
Graph plot
GBPUSD
AUDUSD
EURUSD
USDCAD
USDJPY
USDCHF
NZDUSD
1
0.5
0
-0.5
-1

Symplify

“Nature has a great simplicity and therefore a great beauty”


Coding Dashboards
Coding Shaolin
ALGORITHMS
BY COMPLEXITY
MORE COMPLEX
LEFTPAD
QUICKSORT
GIT MERGE
SELF-DRIVING CAR
GOOGLE SEARCH BACKEND
SPRAWLING EXCEL SPREADSHEET BUILT UP OVER 20 YEARS BY A CHURCH GROUP IN NEBRASKA TO COORDINATE THEIR SCHEDULING
Using Dashboards

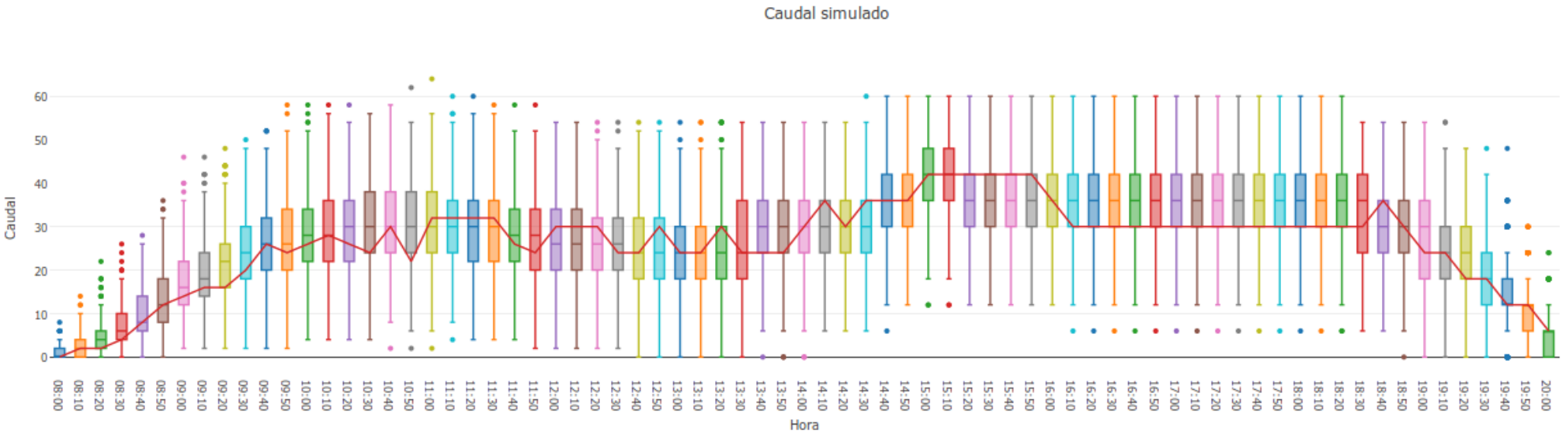
Please...


Distribución de caudales
Frecuencias (%)
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
0
5
10
15
20
25
30
35
40

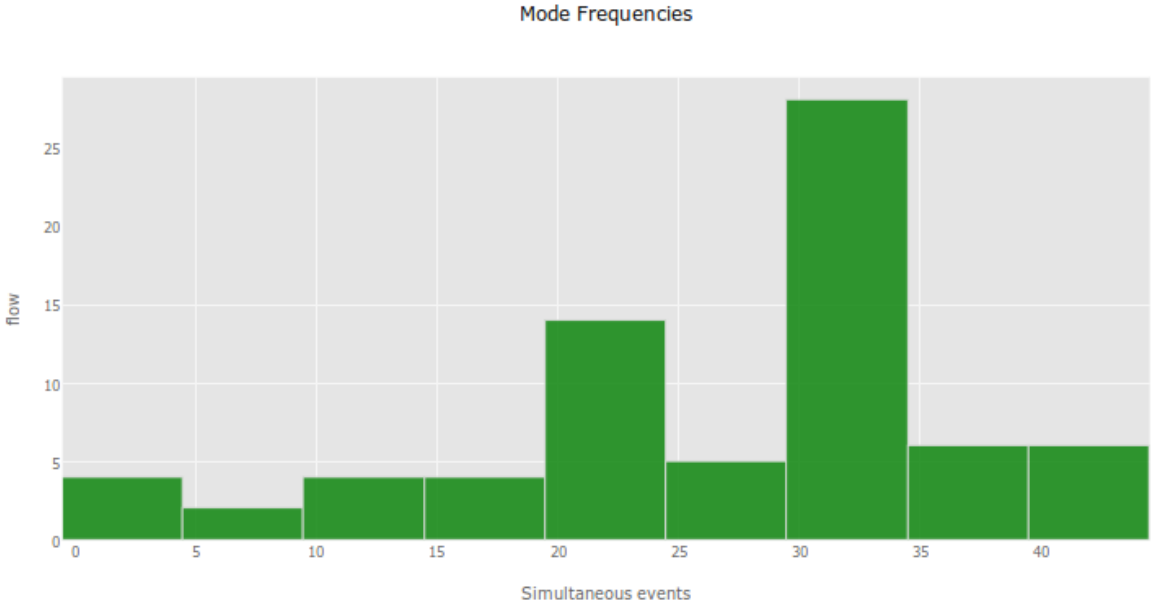
In [18]:
```
sc.widget
```

Numero a generar: 518 | Caudal: 2 | Num horas: 1 | **csv guardado correctamente**

Guardar | Nombre csv: comunidad_regantes | Dist prob: personalizada



In [20]:
```
sc.resultados.mode(axis=1)[0].iplot(kind='hist',bins=20, color='green', title='Mode Frequencies', xTitle="Simultaneous events", yTitle="flow")
```



Shaolin merges OOP with ipywidgets

```python
from shaolin.core.dashboard import Dashboard
import numpy as np
from IPython.core.display import display
import pandas as pd
class ArrayScaler(Dashboard):

    def __init__(self,
                 data,
                 funcs=None,
                 min=-100.,
                 max=100.,
                 step=None,
                 low=None,
                 high=None,
                 mode='interactive',
                 **kwargs):
        if funcs is None:
            self.funcs = {'raw':lambda x: x,
                          'zscore': lambda x: (x-np.mean(x))/np.std(x),
                          'log': np.log,
                          'rank':lambda x: pd.DataFrame(x).rank().values.flatten(),
                          'abs': np.abs
                         }
        else:
            self.funcs  = funcs
        self._data = data
        if min is None:
            min = self._data.min().values[0]
        if max is None:
            max = self._data.max().values[0]
        if step is None:
            step = (max-min)/100.
        if low is None:
            low = min
        if high is None:
            high = max

        dash = ['c$N=array_scaler',
                ['@('+str(min)+', '+str(max)+', '+str(step)+', ('+str(low)+', '+str(high)+'))$N=scale_slider&d=Scale',
                 ['r$N=main_row',['@dd$d=Apply&N=dd_sel&val=raw&o='+str(list(self.funcs.keys())),'@False$N=scale_chk&d=Scale']],
                ]
               ]
        Dashboard.__init__(self, dash, state='array_scaler.pkl',mode=mode, **kwargs)
        self.observe(self.update)
        self.update()

    @property
    def data(self):
        return self._data

    @data.setter
    def data(self, val):
        self._data = val
        self.update()

    def scale_func(self, data):
        Ma = np.max(data)
        mi = np.min(data)
        score = ((data-mi)/(Ma-mi))#to 0-1 interval
        if mi == Ma:
               return np.ones(len(score)) *0.5
        scale_h = self.scale_slider.value[1]
        scale_l = self.scale_slider.value[0]
        return score*(scale_h-scale_l)+scale_l

    def update(self, _=None):
        if self.scale_chk.value:
            self.scale_slider.visible = True
            self.output = self.funcs[self.dd_sel.value](self.scale_func(self._data))
        else:
            self.scale_slider.visible = False
            self.output = self.funcs[self.dd_sel.value](self._data)

```

Widgets

‘widget$param1=val1&param2=val2’

In [2]:
```
from IPython.display import Image #this is for displaying the widgets in the web version of the notebook
from shaolin.core.shaoscript import shaoscript
wid = shaoscript('fs') #this creates a FloatSlider widget
type(wid.widget), type(wid.target)
```

Out[2]:
```
(ipywidgets.widgets.widget_box.Box,
 ipywidgets.widgets.widget_float.FloatSlider)
```

In [3]:
```
wid.target.layout.border = 'red solid'
wid.widget.layout.border = 'blue solid'
#wid.widget
Image(filename='shaolin_syntax_data/img_1.png')
```

Out[3]:

Shaolin Dashboard

```
dash = ['c$N=array_scaler',
        ['@('+str(min)+', '+str(max)+', '+str(step)+', ('+str(low)+', '+str(high)+'))$N=scale_slider&d=Scale',
         ['r$N=main_row',['@dd$d=Apply&N=dd_sel&val=raw&o='+str(list(self.funcs.keys())),'@False$N=scale_chk&d=Scale']],
        ]
       ]
Dashboard.__init__(self, dash, state='array_scaler.pkl',mode=mode, **kwargs)
self.observe(self.update)
self.update()
```

- Strings for defining widgets.
- Nested list for layout boxes.
- Inherit from the Dashboard class.

Out[18]:
```
['c$N=array_scaler',
 ['@(-100.0, 100.0, 1, (-100, 100))$N=scale_slider&d=Scale',
  ['r$N=main_row',
   ["@dd$d=Apply&N=dd_sel&val=raw&o=['abs', 'raw', 'rank', 'zscore', 'log']",
    '@False$N=scale_chk&d=Scale']]]]
```

Science recap

- Think science → Write your models in Python.
- Simplify → Write a GUI using lists and strings.

> “Scientific knowledge is a body of statements of varying degrees of certainty -- some most unsure, some nearly sure, none absolutely certain.”

Chapter 2: Judo!

Gentle or Flexible way

“The highest forms of understanding we can achieve are laughter and human compassion.”


Be flexible

Take advantage of your opponent’s strength.

Judo Recap

- Stack as many widgets as you want.
- Data agnostic dashboards.
- Widgets for css.

Chapter 3: Be a hacker

Find your bugs

"It is not unscientific to make a guess, although many people who are not in science think it is"

Jupyter notebook → Like a charm.
Jupyter_dashboards → Awesome.
Embed, Save → Don’t even try.
Nb_present → Trolling css.

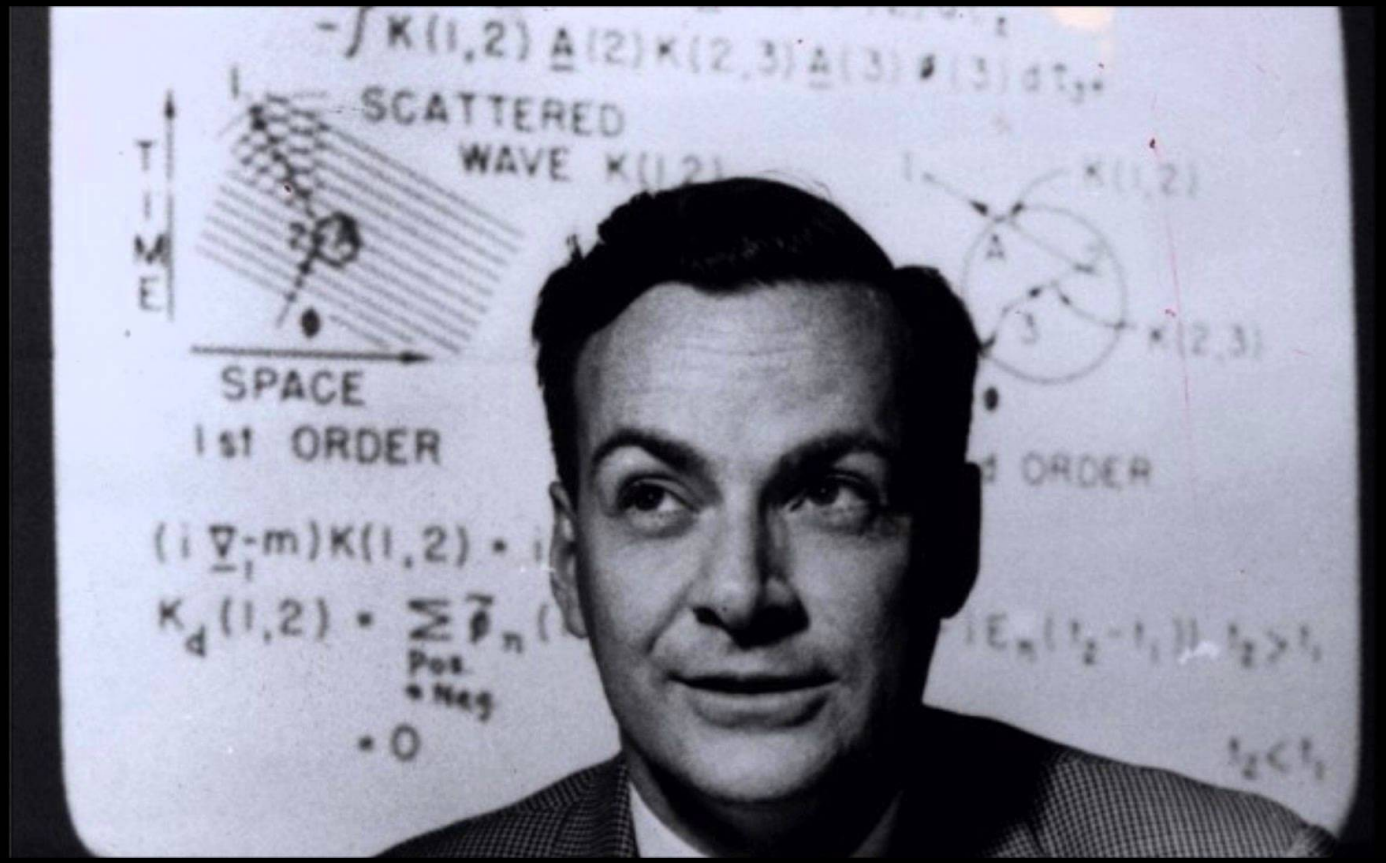
Shaolin vs Interact vs Caravel

Information should be free

- Free software for the win!
- Jupyter notebook as the ultimate science-sharing platform.
- Show me the code.

https://github.com/HCsoft-RD/shaolin

Use it for something different



"What I cannot create I do not understand."

Be a hacker recap

- Load/Save using Shaolin's own functionality.
- Nerd oriented framework.
- Free software.
- Snowball effect.

Chapter 4: Teach & Play

“I don't know anything, but I do know that everything is interesting if you go into it deeply enough.”

Have fun

“Science is like sex:

sometimes something useful comes out,but that is not the reason we are doing it. ”

Be accessible

“We are trying to prove ourselves wrong as quickly as possible, because only in that way can we find progress.”

Final recap

- I hope you had fun
- Thank you
- Feedback time!