



DeepPlay – Insights into competitive freediving

Fabien Plisson, Feb 3 2017

AIDA website

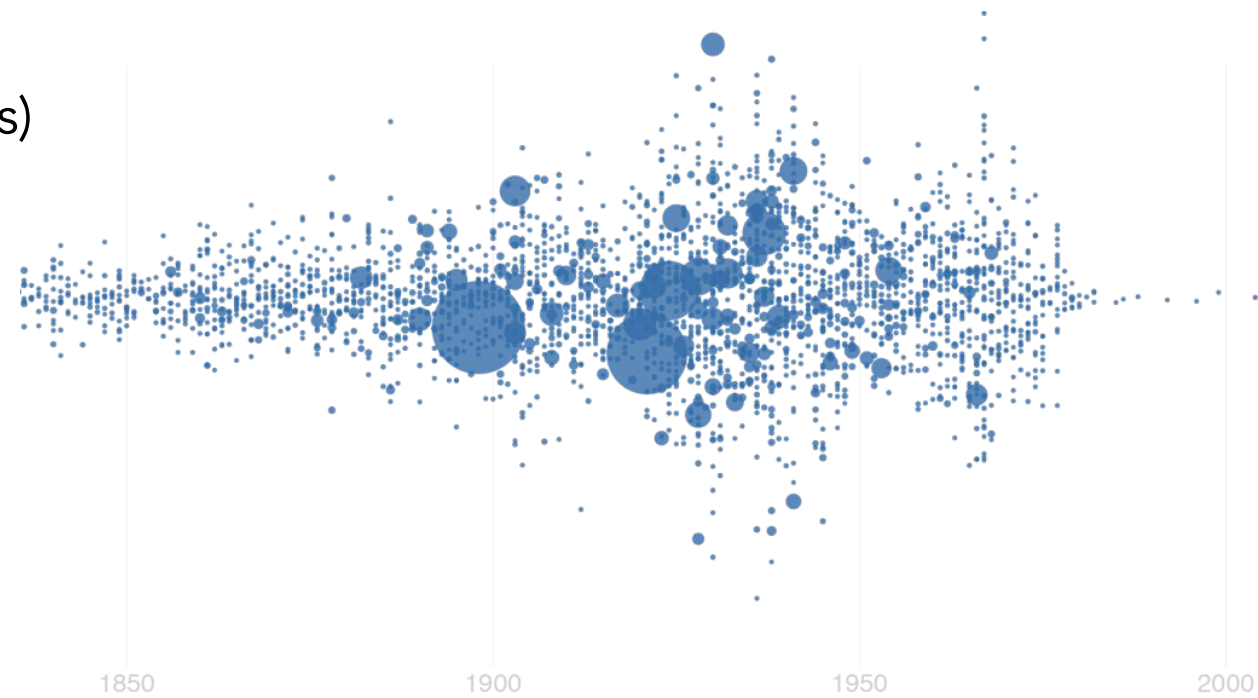
49960 observations (individual records)

-1068 duplicated observations

7875 athletes

72 countries

6 disciplines:
STA, CWT, CNF, DYN, DNF and FIM



PROBLEM - Did you say tables?

CWT		Select country		Select Gender		Select Year		APPLY
#	Name	Result	Announced	PTS	PEN	Date	Place	
1.	Alexey MOLCHANOV (Russian Federation)	128 m	128	127	1	11/20/2012	Suunto Vertical Blue 2012	
2.	Alexey MOLCHANOV (Russian Federation)	128 m	128	127	1	11/20/2012	Suunto Vertical Blue 2012	
3.	Alexey MOLCHANOV (Russian Federation)	128 m	128	127	1	11/20/2012	Suunto Vertical Blue 2012	
4.	Alexey MOLCHANOV (Russian Federation)	128 m	128	128	0	9/15/2013	AIDA INDIVIDUAL DEPHT WORLD CHAMPIONSHIP 20	
5.	Alexey MOLCHANOV (Russian Federation)	128 m	128	0	0	9/9/2013	2013 Kalamata Open	
6.	Guillaume NÉRY (France)	126 m	126	126	0	9/5/2015	2015 AIDA Depth World Championships Pre competition	
7.	Alexey MOLCHANOV (Russian Federation)	126 m	126	126	0	11/20/2012	Suunto Vertical Blue 2012	

INSIGHTS

Athlete with longest recorded career in competitive freediving?

Progress in national records over time?



DeepPlay – A Visual Solution



almost

50%

of your brain
is involved in

**visual
processing** ^[5]



70%

of all your

**sensory
receptors**

are in your eyes ^[4]



we can get
the sense of a

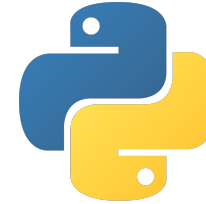
**visual
scene**

in less than
1/10 of a second ^[6]

WORKFLOW – Web Scraping & Insights

get_discipline_value()

Gets discipline ID
Checks it exists



cleanser()

Saves data into
a CLEAN dataframe

scraper()

Scrapes all URLs / discipline
Parses html pages
Aggregates all data
Calls cleanser()

SCRIPT – scraping on demand

<https://aidainternational.org/Ranking/Rankings?page=1&disciplineId=3>

1... max_pages

```
def scraper(key):  
    ''' This function crawls through an entire freediving discipline, regardless how many pages it consists of. '''  
  
    #Obtain html code for url and Parse the page  
    base_url = 'https://www.aidainternational.org/Ranking/Rankings?page='  
    url = '{}1{}'.format(base_url, get_discipline_value(key))  
  
    page = rq.get(url)  
    soup = BeautifulSoup(page.content, "lxml")  
  
    #Use regex to identify the maximum number of pages for the discipline of interest  
    page_count = soup.findAll(text=re.compile(r"Page .+ of .+"))  
    max_pages = str(page_count).split(' ')[3].split('\\')[0]
```

SCRIPT – scraping them all

<https://aidainternational.org/Ranking/Rankings?page=1&disciplineId=3>

value

```
data = []
for p in range(1, int(max_pages)+1):

    #For each page, create corresponding url, request the library, obtain html code and parse the page
    url = '{}{}{}'.format(base_url, p, get_discipline_value(key))

    #The break plays the role of safety guard if dictionary key is wrong (not spelled properly or non-existent) then
    #for library is not executed (and not going through the for loop to generate the data), an empty dataframe is saved
    if url == '{}{}None'.format(base_url, p):
        break
    else:
        new_page = rq.get(url)
        new_soup = BeautifulSoup(new_page.content, "lxml")

        #For each page, each parsed page is saved into the list named "data"
        rows = new_soup.table.tbody.findAll('tr')
        for row in rows:
            cols = row.find_all('td')
            cols = [ele.text.strip() for ele in cols]
            data.append([ele for ele in cols if ele])

        p += 1
```


RESULTS

```
scraper( 'CWT' )
```

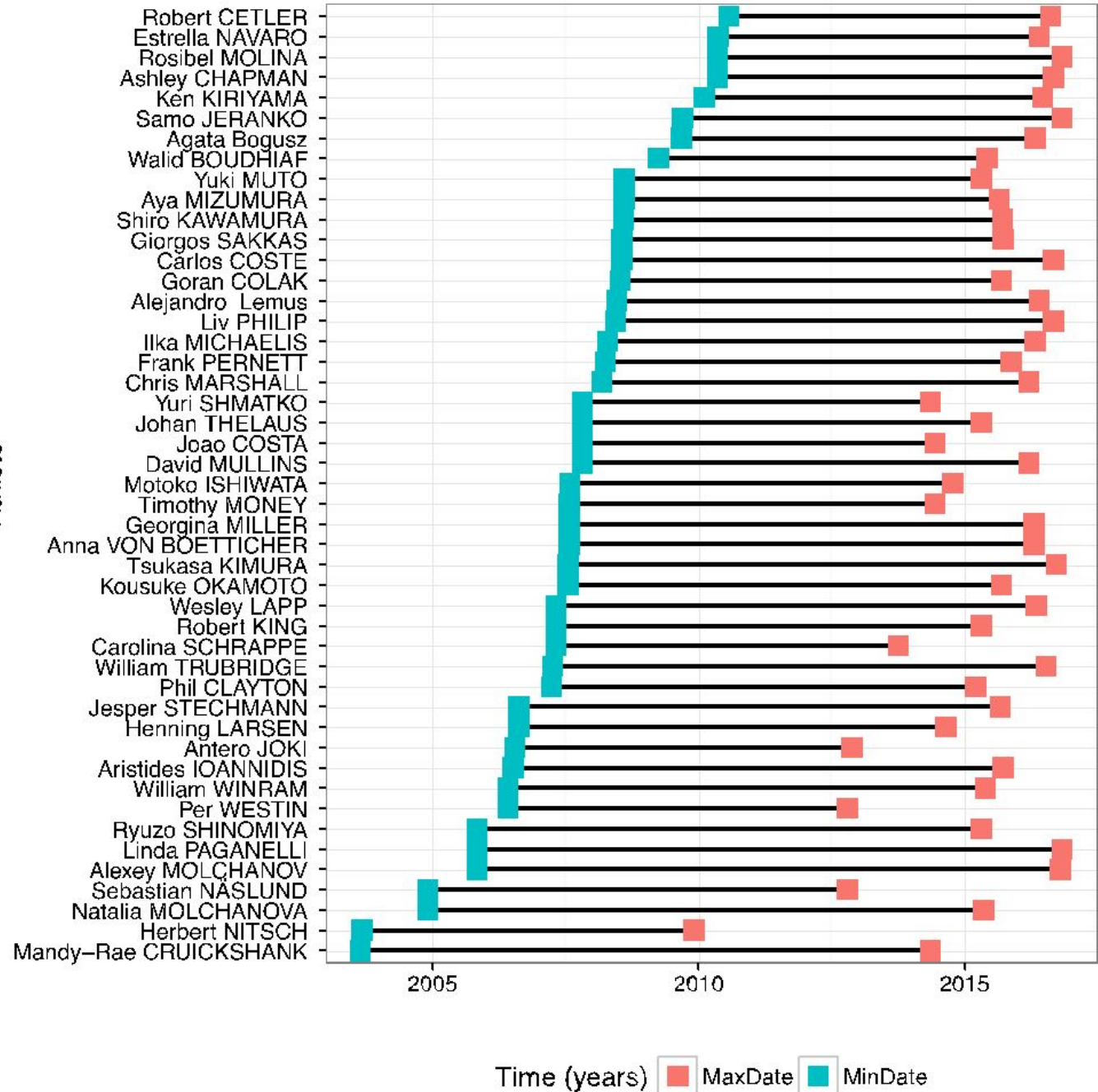
[illegible]

INSIGHTS

With the longest career
in competitive freediving?



Athlete

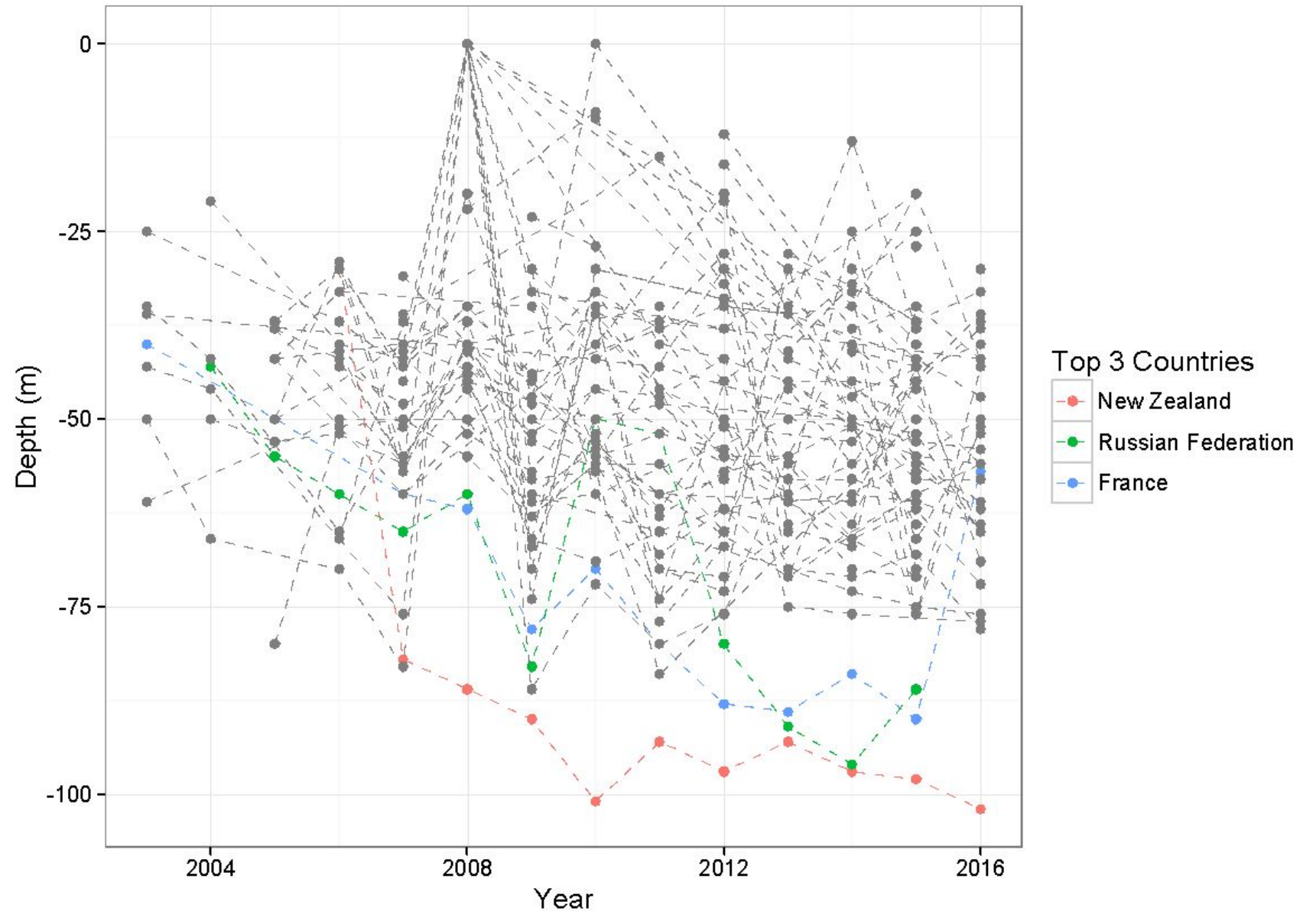


INSIGHTS

National Records
over time?



National Records in Constant Weight (CWT) over Time



ABOUT ME

PhD Medicinal Chemistry
The University of Queensland, Australia

10 Years Research in Drug Discovery



US



Australia



Spain



Pierre Fabre
France

