

DATA SCIENCE FUNDAMENTALS

LESSON 10

Hay Kranen
Wednesday October 23rd, 2018

TODAY'S PROGRAMME

The next two weeks
Assignments 4 + 5
Functions

Week

Monday

Wednesday

Friday

43

~~22/10~~
Functions & scraping

24/10
More functions

26/10
Datalab

44

29/10
Work in class

31/10
Work in class

2/11
Int. assignment
Poster presentation

45

5/11
12.00 Deadline DSF
17.00 Deadline int. assignment



ASSIGNMENT 4

What your program should do:

Your program must:

- Import the **pandas** and **praw** libraries
- Create a new praw API instance with your client id, client secret and user agent.
- Get the 500 **most recent** submissions from a subreddit of your choice. A 'submission' is a Reddit post that can contain a link, text or image and has comments attached to it.
- Transform the submissions in a new list with a dictionary per submission. Every dictionary should contain these bits of information about the submission:
 - The name of the author (e.g. 'aron_124409')
 - Score (e.g. '20')
 - Number of comments (e.g. '100')
 - Title (e.g. 'Programmers needed!!!!')
 - Time of creation (e.g. 'Tue Oct 23 15:29:18 2018')
Note that the Reddit API gives back timestamps, not dates, so you need to do a little conversion (I would suggest the `ctime` method).
- Create a new Pandas dataframe from this list and show the first five items.
- Show the names of the 10 authors in the dataframe that made the most submissions, and how many submissions they made.
- Plot **two** histograms with ten bins: one for the comment counts and one for the scores of all submissions. Both plots should have a title indicating their contents.
- Make a line plot showing the number of posts per hour/day/month, depending on the frequency of posting in your chosen subreddit. You'll need to convert the time of creation to a new Pandas series and resample that to hour/day/month.
- In your Dataframe, find the single submission with the **most** comments. **Optionally**, show the **title**, **number of comments** and **score** for this submission.
Hint: either use the original data you got from the `praw` library, or do a new request for a single submission by ID.
- For the same submission, show the **top five** comments and their score, name of the commenter and the first 100 characters of the comment text.
Note that you need to take into account the 'MoreComments' instances (look in the examples notebook) and that some comments might not have an author.
- A single user might have multiple comments in this submission. Calculate the **summed score** for every commenter in the submission. Then, show the **top five** commenters and their **summed score**.
- Display a horizontal bar chart with the **top 10** commenters with the highest summed scores.



ASSIGNMENT 5

Our Employees

More About Us

[OUR ORGANISATION](#)[OUR EMPLOYEES](#)[OUR FACILITIES](#)[PARTNERSHIPS AND COOPERATIONS](#)[SUBSIDIARIES AND AFFILIATES](#)[QUALITY POLICY AND ETHICAL
GUIDELINES](#)[ACCREDITATION](#)

Our Diverse Team Makes Us Who We Are

Akvaplan-niva has a multidisciplinary and international staff, including 50 environmental and aquaculture specialists and a number of Ph.D students.

The following alphabetical listing shows their names, fields of specialisation, titles and direct office phone lines. Click on the name to retrieve more information (CV, projects worked on, publications, etc).

SORT BY DEPARTMENT

[Management & Administration](#) | [Aquaculture](#) | [Aquaculture R&D](#) | [Arctic R&D](#) | [Coastal & Freshwater](#)
[Marine](#) | [Sensitive Environments Decision Support - SensE](#) | [Russia and Eastern Europe](#) | [Arctic Frontiers](#)
[Laboratory](#)



Anne Tårand
Aasen
Advisor



Hector Andrade
Advisor



Ana Sofia Aniceto
Ph.D. candidate



Starlight
Augustine
Scientist



Magnus Aune
Advisor

[Home](#) » [About Us](#) » [Our Employees](#)

Our Employees

More About Us

[OUR ORGANISATION](#)[OUR EMPLOYEES](#)[OUR FACILITIES](#)[PARTNERSHIPS AND COOPERATIONS](#)[SUBSIDIARIES AND AFFILIATES](#)[QUALITY POLICY AND ETHICAL
GUIDELINES](#)[ACCREDITATION](#)

Anne Tårånd Aasen

Position: Advisor

Department: Coastal & Freshwater

Email: anne.aasen@akvaplan.niva.no

Phone: 90 67 47 94

Location: Tromsø, Norway

[« Go back to our Employees List.](#)

	name	department	function	href
0	Anne Tårand Aasen	Coastal & Freshwater	Advisor	http://www.akvaplan.niva.no
1	Hector Andrade	Marine	Advisor	http://www.akvaplan.niva.no
2	Ana Sofia Aniceto	Arctic R&D	Ph.D. candidate	http://www.akvaplan.niva.no
3	Starrlight Augustine	Arctic R&D	Scientist	http://www.akvaplan.niva.no
4	Magnus Aune	Marine	Advisor	http://www.akvaplan.niva.no

FUNCTIONS (AGAIN)

5m



?m²

5m

Calculate the area of a square room

$$\text{AREA} = \text{SIDE} * \text{SIDE}$$

```
side = 5
area = side * side
print(area) # 25
```

```
area = side * side
```

$$Y = X^2$$

```
y = x * x
```

```
def square(x):  
    y = x * x  
    return y  
  
side = 5  
area = square(side)
```



5

$5 * 5$

25

10

$10 * 10$

100

12

$12 * 12$

144

Celsius to Fahrenheit

Write a function that converts Celsius to Fahrenheit.

Then test your function with the values -2.5, 5 and 30.

Fahrenheit = Celsius * 1.8 + 32

E.g:

12 C = 53.6 F

12 * 1.8 = 21.6 + 32 = 53.6

```
# A function looks like this  
  
def function_name(arg):  
    # Do something with arg  
    return ....
```

```
f = convert(-2.5)  
print(f) # Should return 27.5
```

```
def convert(c):
    return c * 1.8 + 32

f = convert(-2.5)
print(f)
```

```
def convert(c):
    return c * 1.8 + 32

f = convert(-2.5)
print(f)
```

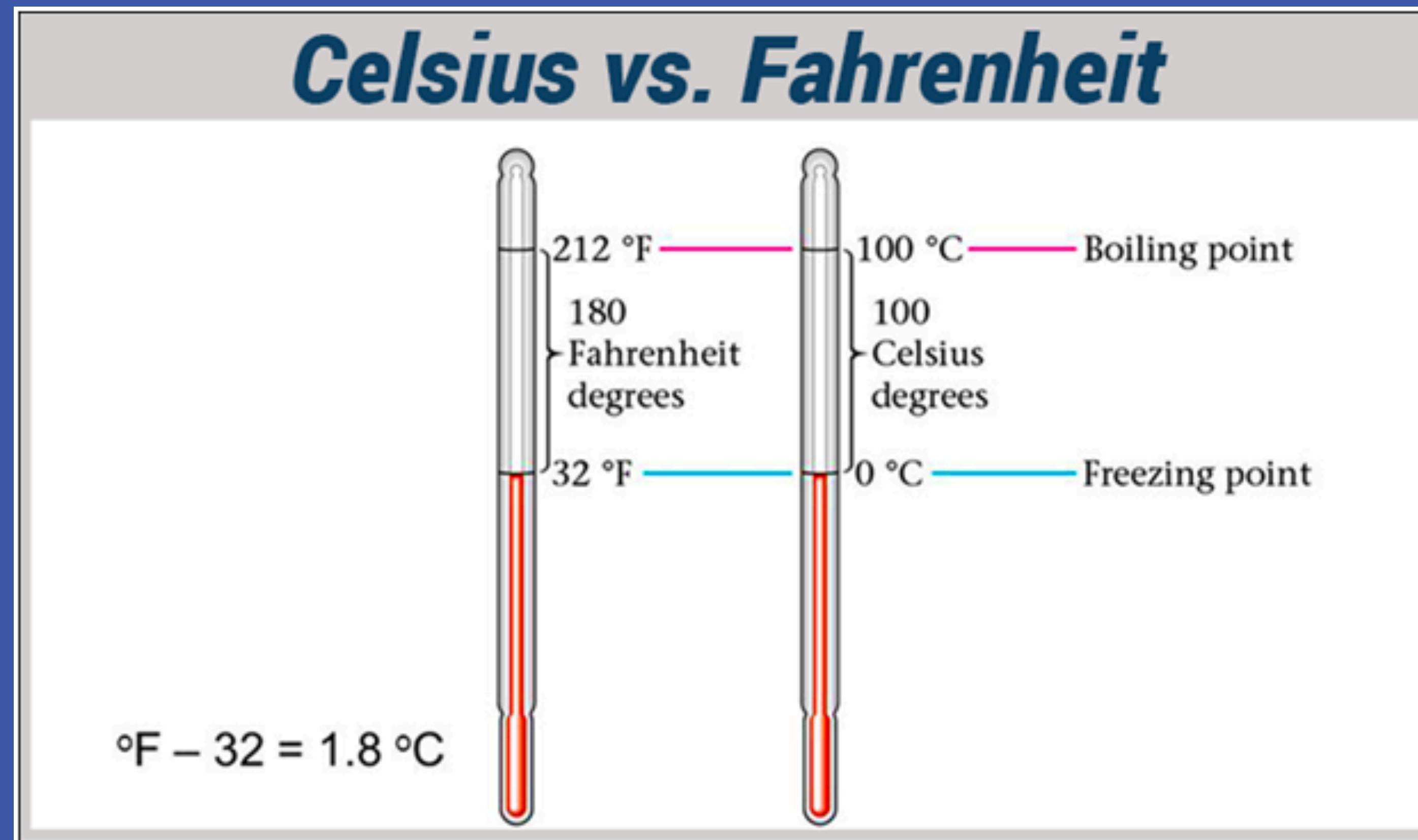


-2.5

$-2.5 * 1.8 + 32$

27.5

Celsius → Fahrenheit



How about the other way around?

```
def c_to_f(c):
    return c * 1.8 + 32

def f_to_c(f):
    return (f - 32) * (5/9)

f = c_to_f(-2.5)
print(f)

c = f_to_c(f)
print(c)
```

```
def convert(temp, input_type):
    if input_type == "c":
        return temp * 1.8 + 32
    elif input_type == "f":
        return (temp - 32) * (5/9)

f = convert(-2.5, "c")
print(f)

c = convert(f, "f")
print(c)
```

Money converter



Convert a currency

PRICE = INPUT * CONVERSION RATE



```
def eur_to_usd(eur):
    EUR_TO_USD = 1.14616
    usd = eur * EUR_TO_USD
    return round(usd, 2)

usd = eur_to_usd(30)
print(usd) # 34.38
```

Input

Transformation

Output



Currency converter

Write a function that converts Euros to US Dollars, British Pounds or Japanese Yen. If the input is invalid, print an error message.

Then test your function with these combinations:

12 Euros to British Pounds

43 Euros to Japanese Yen

9.32 Euros to US Dollars

5 Euros to Oreo cookies

Use the conversion rates on this slide.

Inverse	<input checked="" type="checkbox"/>
	EUR
	JPY
	USD
	GBP
	1 EUR
1.00000	128.487
1.14590	0.88311

Celsius to Fahrenheit example

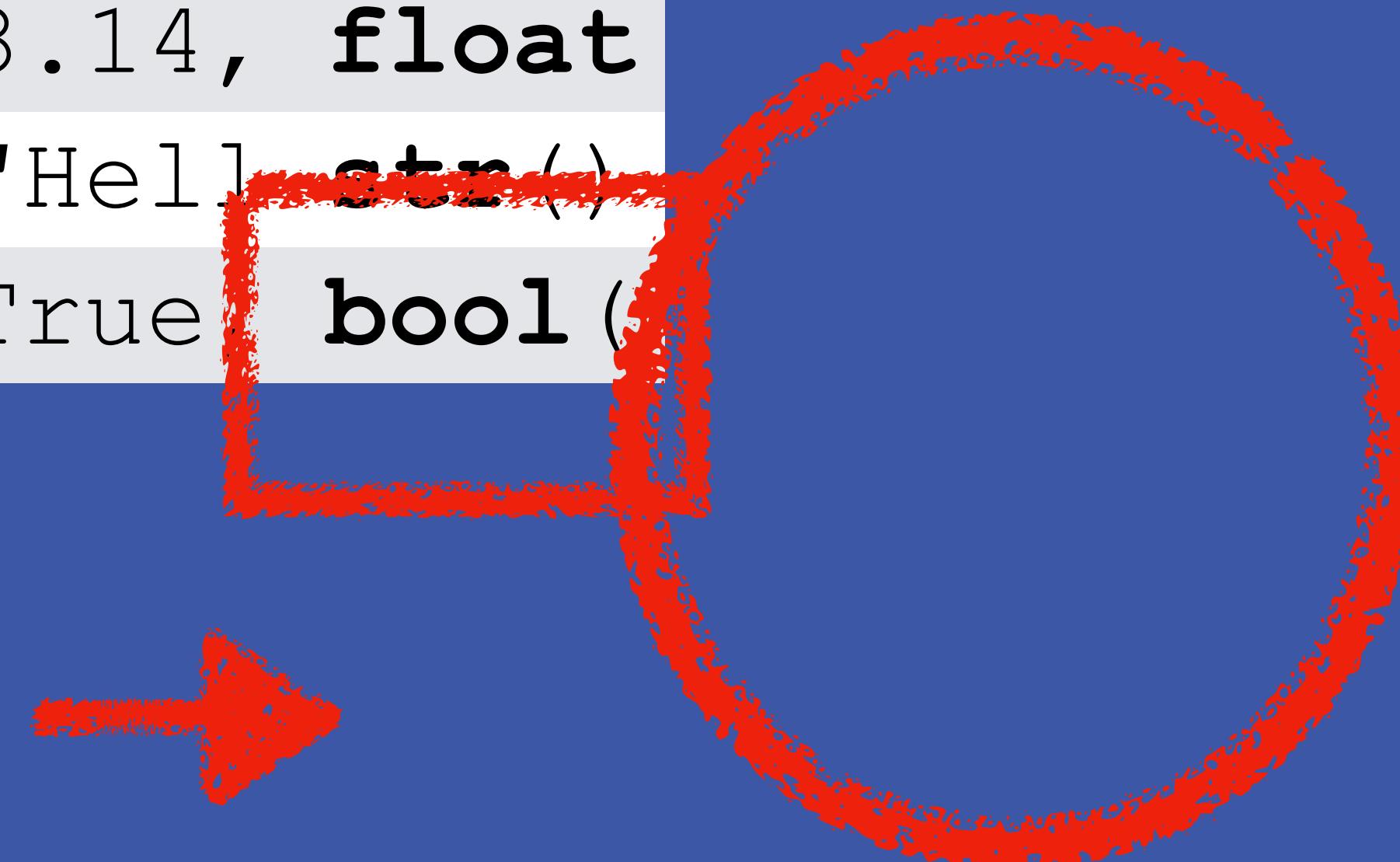
```
def convert(temp, input_type):  
    if input_type == "c":  
        return temp * 1.8 + 32  
    elif input_type == "f":  
        return (temp - 32) * (5/9)  
  
f = convert(-2.5, "c")  
print(f)  
  
c = convert(f, "f")  
print(c)
```

Function use example

```
print(convert_euros(12, "pound"))  
print(convert_euros(43, "yen"))  
print(convert_euros(9.32, "dollar"))  
print(convert_euros(5, "oreo"))
```


Type Example Conversion

Integ	42, int()
Float	3.14, float
String	"Hello", str()
Boolean	True, bool()



```
age = 20

if age < 20:
    print("option 1")
elif age <= 20 and age > 20:
    print("option 2")
else:
    print("option 3")
```

Compilation
Interpretation