Mathematical Foundations of Data Science Assignment 1

Trimester 2, 2024

Question 1

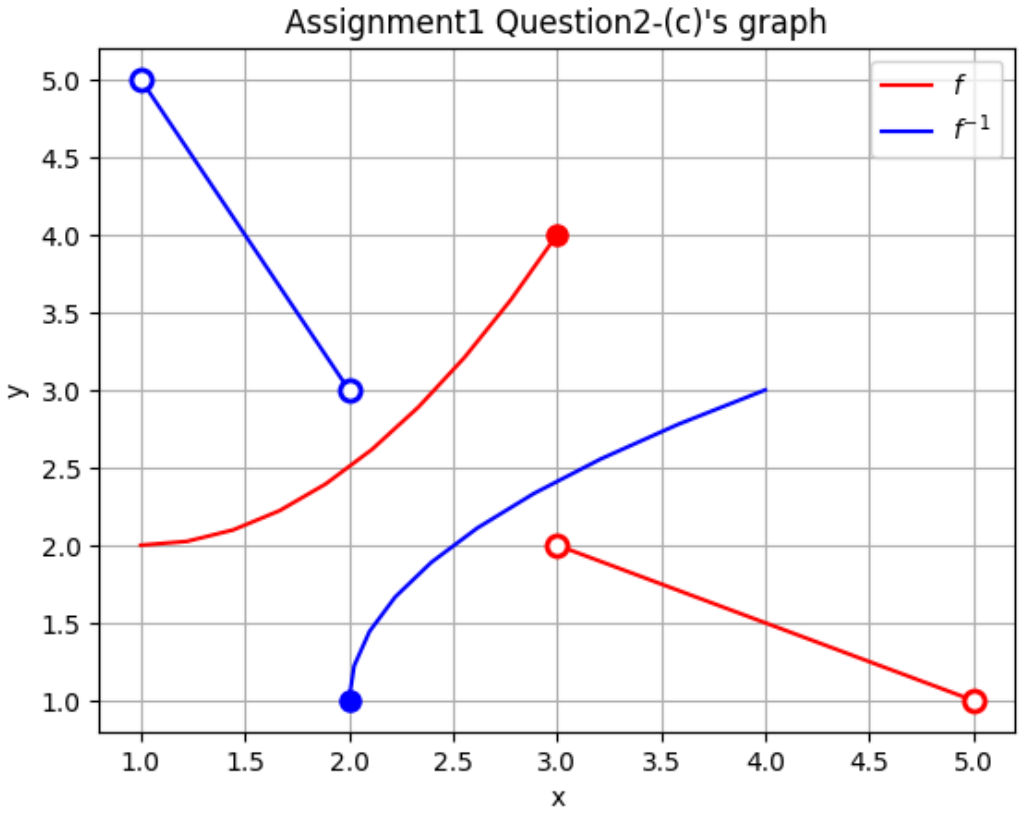
- The statement in set notation means for any y belongs to rational numbers, there is an absolute value of y smaller than y.
- The statement is false. For example there is a positive number like 5 belongs to rational numbers but its absolute value is 5 itself which equals 5 but not smaller.

Question 2

- (a) The domain of *g* is $(-1, 5)$.
- (b) It is. As the graph shows the function has 2 dots at 3 but only the upper dot is solid which means the value of g when $x = 3$ is 4.
- (c) The expression of *g* on domain $[1, 5)$ is

$$f = \begin{cases} \frac{1}{2}(x-1)^2 + 2, & x \in [1, 3] \\ \frac{1}{2}x + \frac{7}{2}, & x \in (3, 5) \end{cases}$$

  Then we channge the places of x and f, then we can plot the graph of f^{-1} in Jupyterbook:



Question 3

- (a) $A \backslash \mathbb{Z} = \{\frac{1}{4}, \frac{3}{2}, \pi\}$
- (b) $A \cap B = \{0, \frac{1}{4}, \frac{3}{2}\}$

Question 4

- To find $h^{-1}(x)$, we just switch all instances of x and $h(x)$, which is

$$x = \frac{h^{-1}(x)+3}{2}$$

$$h^{-1}(x) = 2x - 3$$

In [ ]:
```
import pandas as pd

df1 = pd.read_csv('swimming.csv')
df1.head()
```

Out[ ]:

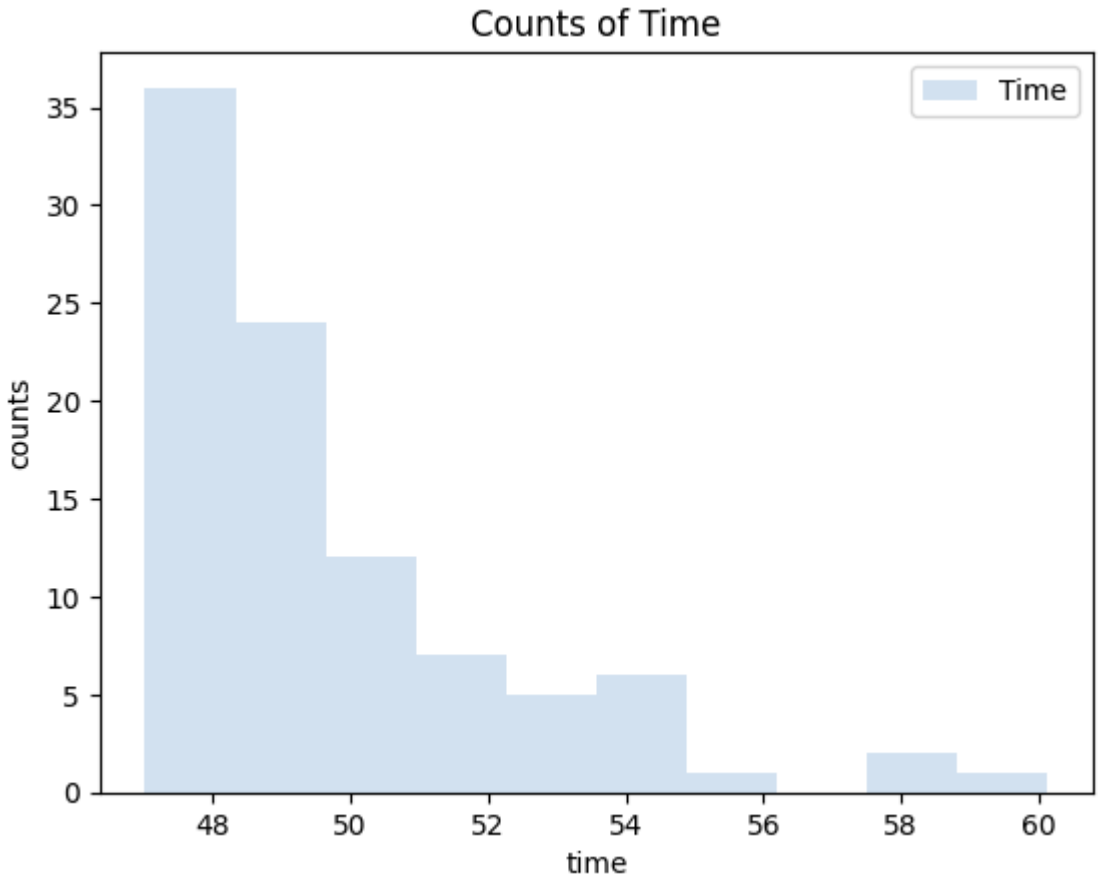
	Rank	Type	Number	Lane	Swimmer	Nation	Time
0	48	Heat	4	1	Matthew Abeysinghe	Sri Lanka	50.62
1	53	Heat	3	1	Issa Al-Adawi	Oman	51.81
2	8	Heat	9	6	David Popovici	Romania	48.03
3	50	Heat	3	6	Yousuf Al-Matrooshi	United Arab Emirates	51.50
4	18	Heat	7	1	Apostolos Christou	Greece	48.50

- (b) We could import `matplotlib` first, then we creat a filter with only times in `df1`. Then we use `plt.hist` to create the histogram of all Times divided into different intervals. Finally we use `plt.xlabel`, `plt.ylabel`, `plt.legend` and `plt.title` to add details.

In [ ]:
```
import matplotlib.pyplot as plt

time = df1['Time']
plt.hist(time,label='Time',alpha = 0.2)
plt.xlabel('time')
plt.ylabel('counts')
plt.legend()
plt.title('Counts of Time')
```

Out[ ]:
```
Text(0.5, 1.0, 'Counts of Time')
```



- (c) The mean time is `49.805`. We could use `.mean()` to calculate it.

In [ ]:
```
m_time = round(df1['Time'].mean(),3)
print(m_time)
```

```
49.805
```

- (d) We could use the code below to find who is the fasteset one in the heats, it's `Thomas Ceccon`.

In [ ]:
```
df2 = df1[df1['Type'] == 'Heat']
df2[df2['Rank']  == 1]
```

Out[ ]:

	Rank	Type	Number	Lane	Swimmer	Nation	Time
92	1	Heat	9	2	Thomas Ceccon	Italy	47.71

- (e) To find who is the slowest, we could use the code below. It's `Roman Mityukov`

In [ ]:
```
df3 = df1[df1['Type'] == 'Semifinal']
df_semi2 = df3[df3['Number'] == 2]
df_semi2[df_semi2['Time'] == df_semi2.max()['Time']]
```

Out[ ]:

	Rank	Type	Number	Lane	Swimmer	Nation	Time
7	16	Semifinal	2	8	Roman Mityukov	Switzerland	48.53