

빅데이터 실습

10주차 2차시

왕좌의 게임 데이터 분석 [1]

데이터 시각화



학습개요

- 1/ 왕좌의 게임 데이터 소개
- 2/ 왕좌의 게임 데이터 분석 [1]
 - 책 권별 사망자 수 추이 시각화
- 3/ 왕좌의 게임 데이터 분석 [2]
 - 전쟁의 공격/수비 진영 규모 시각화

01

왕좌의 게임 데이터터 소개





1 왕좌의 게임 데이터 소개





왕좌의 게임 데이터 소개

➤ 왕좌의 게임 소개

- ✓ HBO에서 방영한 판타지 TV 시리즈
- ✓ 2011년 시즌1을 시작으로 2019년 8시즌을 마지막으로 종료
- ✓ 원작 소설 : 조지 R.R 마틴의 원작 '얼음과 불의 노래'
- ✓ 내용
 - ◎ 판타지 세계인 웨스테로스 대륙의 여러 국가들의 연맹 국가인 7왕국의 통치권인 '철 왕좌'를 차지하기 위한 싸움을 그려낸 드라마
 - ◎ 왕좌를 차지하기 위한 가문 간의 전쟁 외에도 북부의 '백귀'의 위협으로부터 웨스테로스 대륙을 지켜내는 이야기도 포함



왕좌의 게임 데이터 소개

➤ 왕좌의 게임 데이터

✓ Kaggle.com에서 제공하는 데이터

✓ Kaggle

◎ 2010년 설립된 데이터 과학자와 머신러닝 개발자들을 위한 온라인 커뮤니티

Compete

분석 경진 대회

Datasets

데이터셋 제공

Notebooks

분석 소스코드 제공

Courses

교육 콘텐츠 제공

1 왕좌의 게임 데이터 소개

➤ 왕좌의 게임 데이터

- ✓ Kaggle.com에서 제공하는 데이터
- ✓ Kaggle

◎ 2010년 설립된 데이터 과학자와 머신러닝 개발자들을 위한 온라인 커뮤니티





왕좌의 게임 데이터 소개



〈출처〉 (1) https://en.wikipedia.org/wiki/Facebook_Zero

(2) <https://imagescdn.gettyimagesbank.com/500/19/204/707/0/1193728668.jpg>

(3) <https://imagescdn.gettyimagesbank.com/500/18/683/409/0/961875454.jpg>

왕좌의 게임 데이터

© Battles.csv (전쟁 정보)

	name	year	battle_number	attacker_king	defender_king	attacker_1	attacker_2	attacker_3	attacker_4	defender_1	defender_2	defender_3
0	Battle of the Golden Tooth	298	1	Joffrey/Tommen Baratheon	Robb Stark	Lannister	NaN	NaN	NaN	Tully	NaN	NaN
1	Battle at the Mummer's Ford	298	2	Joffrey/Tommen Baratheon	Robb Stark	Lannister	NaN	NaN	NaN	Baratheon	NaN	NaN
2	Battle of Riverrun	298	3	Joffrey/Tommen Baratheon	Robb Stark	Lannister	NaN	NaN	NaN	Tully	NaN	NaN
3	Battle of the Green Fork	298	4	Robb Stark	Joffrey/Tommen Baratheon	Stark	NaN	NaN	NaN	Lannister	NaN	NaN
4	Battle of the Whispering Wood	298	5	Robb Stark	Joffrey/Tommen Baratheon	Stark	Tully	NaN	NaN	Lannister	NaN	NaN
5	Battle of the Camps	298	6	Robb Stark	Joffrey/Tommen Baratheon	Stark	Tully	NaN	NaN	Lannister	NaN	NaN
6	Sack of Darry	298	7	Joffrey/Tommen Baratheon	Robb Stark	Lannister	NaN	NaN	NaN	Darry	NaN	NaN
7	Battle of Moat Cailin	299	8	Balon/Euron Greyjoy	Robb Stark	Greyjoy	NaN	NaN	NaN	Stark	NaN	NaN
8	Battle of Deepwood Motte	299	9	Eddard Stark	Eddard Stark	Greyjoy	NaN	NaN	NaN	Stark	NaN	NaN
9	Battle of the Stony Shore	299	10	Eddard Stark	Eddard Stark	Greyjoy	NaN	NaN	NaN	Stark	NaN	NaN

총 25개 컬럼

전투의 이름, 전투가 벌어진 연도, 공격 진영과 수비 진영 정보 등

왕좌의 게임 데이터

© Battles.csv (전쟁 정보)

defender_4	attacker_outcome	battle_type	major_death	major_capture	attacker_size	defender_size	attacker_commander	defender_commander	summer	location	region	note
NaN	win	pitched battle	1.0	0.0	15000.0	4000.0	Jaime Lannister	Clement Piper, Vance	1.0	Golden Tooth	The Westerlands	NaN
NaN	win	ambush	1.0	0.0	NaN	120.0	Gregor Clegane	Beric Dondarrion	1.0	Mummer's Ford	The Riverlands	NaN
NaN	win	pitched battle	0.0	1.0	15000.0	10000.0	Jaime Lannister, Andros Brax	Edmure Tully, Tytos Blackwood	1.0	Riverrun	The Riverlands	NaN
NaN	loss	pitched battle	1.0	1.0	18000.0	20000.0	Roose Bolton, Wyllis Manderly, Medger Cerwyn, H...	Tywin Lannister, Gregor Clegane, Kevan Lannist...	1.0	Green Fork	The Riverlands	NaN
NaN	win	ambush	1.0	1.0	1875.0	6000.0	Robb Stark, Brynden Tully	Jaime Lannister	1.0	Whispering Wood	The Riverlands	NaN
NaN	win	ambush	0.0	0.0	6000.0	12625.0	Robb Stark, Tytos Blackwood, Brynden Tully	Lord Andros Brax, Forley Prester	1.0	Riverrun	The Riverlands	NaN
NaN	win	pitched battle	0.0	0.0	NaN	NaN	Gregor Clegane	Lyman Darry	1.0	Darry	The Riverlands	NaN
NaN	win	pitched battle	0.0	0.0	NaN	NaN	Victarion Greyjoy	NaN	1.0	Moat Cailin	The North	NaN
NaN	win		0.0	0.0	1000.0	NaN	Asha Greyjoy	NaN	1.0	Deepwood Motte	The North	NaN
												Greyjoy's troop

총 25개 컬럼

공격/수비 진영 병사 숫자, 어느 지역에서 전쟁이 이루어졌는지 등의 정보 제공

왕좌의 게임 데이터

© Death.csv (사망한 캐릭터 정보)

	Name	Allegiances	Death Year	Book of Death	Death Chapter	Book Intro Chapter	Gender	Nobility	GoT	CoK	SoS	FfC	DwD
0	Addam Marbrand	Lannister	NaN	NaN	NaN	56.0	1	1	1	1	1	1	0
1	Aegon Frey (Jinglebell)	None	299.0	3.0	51.0	49.0	1	1	0	0	1	0	0
2	Aegon Targaryen	House Targaryen	NaN	NaN	NaN	5.0	1	1	0	0	0	0	1
3		House Greyjoy	300.0	5.0	20.0	20.0	1	1	0	0	0	0	1
4		Lannister	NaN	NaN	NaN	NaN	1	1	0	0	1	0	0
5			NaN	NaN	NaN	NaN	1	1	0	1	1	0	0
6			300.0	4.0	35.0	21.0	1	1	1	0	1	1	0
7			300.0	5.0	NaN	59.0	0	1	1	1	1	0	1
8			NaN	NaN	NaN	11.0	1	1	0	1	0	1	0
9			NaN	NaN	NaN	0.0	1	0	0	0	1	0	0

Death.csv

사망한 캐릭터 정보 제공

- 이름, 가문, 사망 연도
- 성별, 귀족 여부
- 1~5권에서 나왔는지 여부

02

왕좌의 게임 데이터 분석 [1]





책 권별 사망자 수 추이 시각화

왕좌의 게임 데이터

◎ 왕좌의 게임 데이터 분석 소개

✓ Death.csv (사망한 캐릭터 정보)

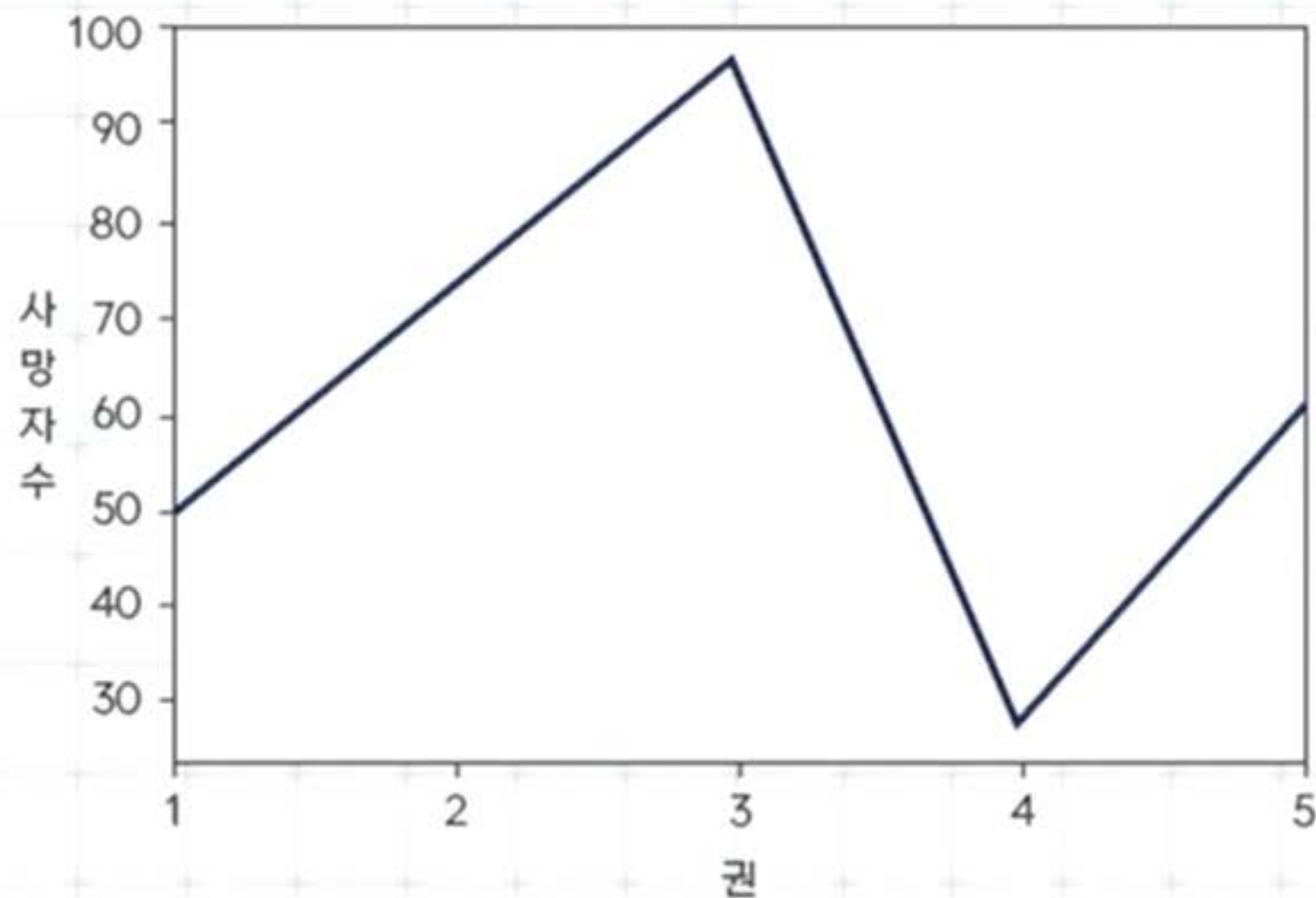
	Name	Allegiances	Death Year	Book of Death	Death Chapter	Book Intro Chapter	Gender	Nobility	GoT	CoK	SoS	FfC	DwD
0	Addam Marbrand	Lannister	NaN	NaN	NaN	56.0	1	1	1	1	1	1	0
1	Aegon Frey (Jinglebell)	None	299.0	3.0	51.0	49.0	1	1	0	0	1	0	0
2	Aegon Targaryen	House Targaryen	NaN	NaN	NaN	5.0	1	1	0	0	0	0	1
3	Adrack Humble	House Greyjoy	300.0	5.0	20.0	20.0	1	1	0	0	0	0	1
4	Aemon Costayne	Lannister	NaN	NaN	NaN	NaN	1	1	0	0	1	0	0
5	Aemon Estermont	Baratheon	NaN	NaN	NaN	NaN	1	1	0	1	1	0	0
6	Aemon Targaryen (son of Maekar I)	Night's Watch	300.0	4.0	35.0	21.0	1	1	1	0	1	1	0
7	Aenys Frey	None	300.0	5.0	NaN	59.0	0	1	1	1	1	0	1
8	Aeron Greyjoy	House Greyjoy	NaN	NaN	NaN	11.0	1	1	0	1	0	1	0
9	Aethan	Night's Watch	NaN	NaN	NaN	0.0	1	0	0	0	1	0	0

왕좌의 게임 데이터

◎ 왕좌의 게임 데이터 분석 소개

✓ 책 권별 사망자 수 추이 시각화

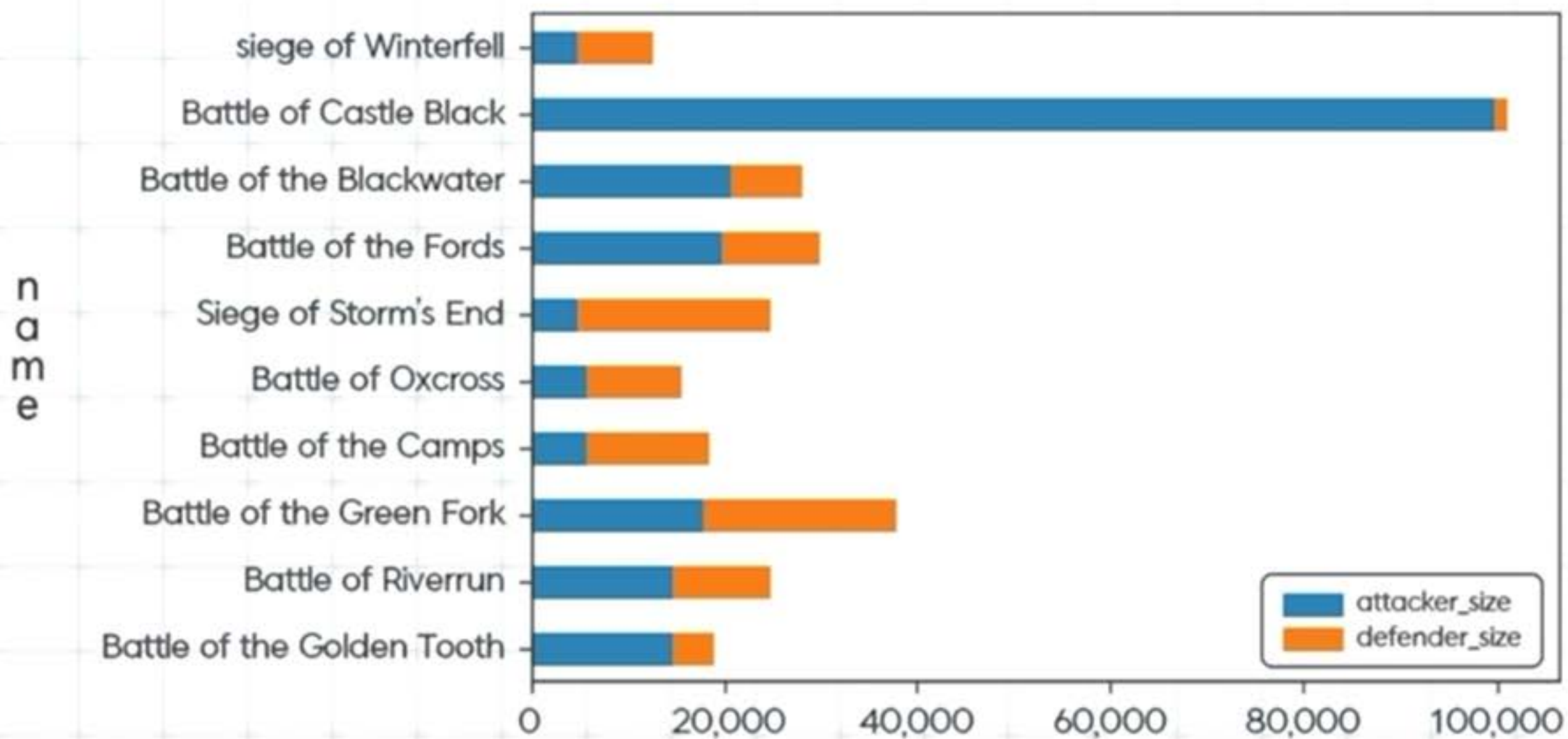
〈 권별 사망자 수 추이 〉



왕좌의 게임 데이터

◎ 왕좌의 게임 데이터 분석 소개

✓ 전쟁의 공격/수비 진영 규모 시각화



왕좌의 게임 데이터 분석 및 시각화 실습

◎ 데이터 적재

✓ battles.csv / deaths.csv

3. 왕좌의 게임 데이터 분석 및 시각화 실습

1) 데이터 적재

- battles.csv : 전투의 이름, 발생년도, 공격/수비 정보(가문, 왕, 군대 규모 등)을 저장
- deaths.csv : 캐릭터들의 이름, 성별 등 기본 정보와 사망 정보

```
In [ ]: battles = pd.read_csv('data/data-society-game-of-thrones/battles.csv')
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

battles.csv

전투의 이름, 발생 년도, 공격/수비 정보 저장

파일 경로는 상대 경로이므로 주피터 파일에 있는 자신들의 경로에 맞춰 작성하세요.

왕좌의 게임 데이터 분석 및 시각화 실습

◎ 데이터 적재

✓ battles.csv / deaths.csv

공격 진영 왕 정보

총 25개 컬럼

	name	year	battle_number	attacker_king	defender_king	attacker_1	attacker_2	attacker_3
0	Battle of the Golden Tooth	298	1	Joffrey/Tommen Baratheon	Robb Stark	Lannister	NaN	↑
1	Battle at the Mummer's Ford	298	2	Joffrey/Tommen Baratheon	Robb Stark	Lannister	NaN	↑
2	Battle of Riverrun	298	3	Joffrey/Tommen Baratheon	Robb Stark	Lannister	NaN	↑
3	Battle of the Green Fork	298	4	Robb Stark	Joffrey/Tommen Baratheon	Stark	NaN	↑
4	Battle of the Whispering Wood	298	5	Robb Stark	Joffrey/Tommen Baratheon	Stark	Tully	↑

전투 이름

아이디값

왕좌의 게임 데이터 분석 및 시각화 실습

◎ 데이터 적재

✓ battles.csv / deaths.csv

가문에 대한 정보

g	defender_king	attacker_1	attacker_2	attacker_3	attacker_4	defender_1	...	major_death	majc
n	Robb Stark	Lannister	NaN	NaN	NaN	Tully	...	1.0	
n	Robb Stark	Lannister	NaN	NaN	NaN	Baratheon	...	1.0	
n	Robb Stark	Lannister	NaN	NaN	NaN	Tully	...	0.0	
k	Joffrey/Tommen Baratheon	Stark	NaN	NaN	NaN	Lannister	...	1.0	
k	Joffrey/Tommen Baratheon	Stark	Tully	NaN	NaN	Lannister	...	1.0	

attacker 및 defender 1~4까지 있는 이유는 가문이 연합해서 공격 or 수비하는 경우도 있기 때문입니다.

왕좌의 게임 데이터 분석 및 시각화 실습

◎ 데이터 적재

✓ battles.csv / deaths.csv

가문에 대한 정보

g	defender_king	attacker_1	attacker_2	attacker_3	attacker_4	defender_1	...	major_death	majc
n	Robb Stark	Lannister	NaN	NaN	NaN	Tully	...	1.0	
n	Robb Stark	Lannister	NaN	NaN	NaN	Baratheon	...	1.0	
n	Robb Stark	Lannister	NaN	NaN	NaN	Tully	...	0.0	
k	Joffrey/Tommen Baratheon	Stark	NaN	NaN	NaN	Lannister	...	1.0	
k	Joffrey/Tommen Baratheon	Stark	Tully	NaN	NaN	Lannister	...	1.0	

연합하지 않고 하나의 가문끼리 싸운 경우에는 attacker_2~4, defender_2~4는 NaN값으로 저장됩니다.

왕좌의 게임 데이터 분석 및 시각화 실습

◎ 데이터 적재

✓ battles.csv / deaths.csv

attacker_3	attacker_4	defender_1	...	major_death	major_capture	attacker_size	defender_size	attacker
NaN	NaN	Tully	...	1.0	0.0	15000.0	4000.0	
NaN	NaN	Baratheon	...	1.0	0.0	NaN	120.0	
NaN	NaN	Tully	...	0.0	1.0	15000.0	10000.0	
NaN	NaN	Lannister	...	1.0	1.0	18000.0	20000.0	Rob
NaN	NaN	Lannister	...	1.0	1.0	1875.0	6000.0	Rob

주요 캐릭터 사망 및 포로 여부

공격/수비 진행 병사 숫자

왕좌의 게임 데이터 분석 및 시각화 실습

◎ 데이터 적재

✓ battles.csv / deaths.csv

전투의 이름, 발생년도, 공격/수비 정보(가문, 왕, 군대 규모 등)을 저장

• deaths.csv : 캐릭터들의 이름, 성별 등 기본 정보와 사망 정보

```
In [101]: battles = pd.read_csv('data/data-society-game-of-thrones/battles.csv')
battles.head()
```

...

```
In [103]: death = pd.read_csv('data/data-society-game-of-thrones/death.csv')
death.head(2)
```

FileNotFoundError

<ipython-input-103-910

```
----> 1 death = pd.read_csv('data/data-society-game-of-thrones/death.csv')
      2 death.head(2)
```

deaths.csv

Traceback (most recent call last)

캐릭터들의 이름, 성별 등 기본 정보와 사망 정보

```
C:\ProgramData\Anaconda3\lib\site-packages\pandas\io\parsers.py in read_csv(filepath, sep, delimiter, header, names, index_col, usecols, squeeze, prefix, mangle_dupe_cols, dtype, engine, converters, true_values, false_values, skipinitialspace, skiprows, skipfooter, nrows, na_values, keep_default_na, na_filter, verbose, skip_blank_lines, parse_dates, infer_datetime_format, keep_date_col, date_parser, dayfirst, cache_dates, iterator, chunksize, compression, thousands, decimal, linet
```


왕좌의 게임 데이터 분석 및 시각화 실습

◎ 데이터 적재

✓ battles.csv / deaths.csv

전투의 이름, 발생년도, 공격/수비 정보(가문, 왕, 군대 규모 등)을 저장
• deaths.csv : 캐릭터들의 이름, 성별 등 기본 정보와 사망 정보

```
In [101]: battles = pd.read_csv('data/data-society-game-of-thrones/battles.csv')
battles.head()
```

...

```
In [104]: death = pd.read_csv('data/data-society-game-of-thrones/character-deaths.csv')
death.head(2)
```

Out [104]:

	Name	Allegiances	Death Year	Book of Death	Death Chapter	Book Intro Chapter	Gender	Nobility	GoT	CoK	SoS	F
0	Addam Marbrand	Lannister	NaN	NaN	NaN	56.0	1	1	1	1	1	
1	Aegon Frey (Jinglebell)	None	299.0	3.0	51.0	49.0	1	1	0	0	1	

In []:

왕좌의 게임 데이터 분석 및 시각화 실습

◎ 데이터 적재

✓ battles.csv / deaths.csv

• battles.csv : 전투의 이름, 발생년도, 공격/수비 정보(가문, 왕, 군대 규모 등)을 저장
• deaths.csv : 캐릭터들의 이름, 성별 등 기본 정보와 사망 정보

```
In [101]: battles = pd.read_csv('data/data-society-game-of-thrones/battles.csv')
battles.head()
```

...

```
In [104]: death = pd.read_csv('data/data-society-game-of-thrones/character-deaths.csv')
death.head(2)
```

귀족 여부(귀족1, 평민0)

Out [104]:

Name	Allegiances	Death Year	Book of Death	Death Chapter	Book Intro Chapter	Gender	Nobility	GoT	CoK	SoS	FfC	DwD
Stannis Baratheon	Lannister	NaN	NaN	NaN	56.0	1	1	1	1	1	1	0
Raeys Frey (lebell)	None	299.0	3.0	51.0	49.0	1	1	0	0	1	0	0

In []:

왕좌의 게임 데이터 분석 및 시각화 실습

◎ 책 권별 사망자 수 추이 시각화

✓ 사망자 수 구하기 : death

In []:

2) 책의 권(Book of Death)별로 사망자 수 추이 시각화

In [105]: death

Out [105]:

	Name	Allegiances	Death Year	Book of Death	Death Chapter	Book Intro Chapter	Gender	Nobility	GoT	CoK	SoS
0	Addam Marbrand	Lannister	NaN	NaN	NaN	56.0	1	1	1	1	1
1	Aegon Frey (Jinglebell)	None	299.0	3.0	51.0	49.0	1	1	0	0	1
2	Aegon Targaryen	House Targaryen	NaN	NaN	NaN	5.0	1	1	0	0	0
3	Adrack Humble	House Greyjoy	300.0	5.0	20.0	20.0	1	1	0	0	0

왕좌의 게임 데이터 분석 및 시각화 실습

◎ 책 권별 사망자 수 추이 시각화

① value_counts() 함수 활용

In []:

2) 책의 권(Book of Death)별로 사망자 수 추이

```
In [106]: # 1) Series.value_counts() 함수 활용
death.value_counts()
```

```
Out[106]:
```

Name	ro	Chapter	Gender	Allegiances	Death Year	Book of Death	Death Chapter	Book Int			
				Nobility	GoT	CoK	SoS	FfC	DwD		
Yurkhaz zo Yunzak				None		300.0		5.0		59.0	47.0
1		0		0	0	0	1	1			
Hazzea				None		300.0		5.0		11.0	11.0
0		0		0	0	0	1	1			
Grubbs				Night's Watch		300.0		5.0		4.0	0.0
1		0		0	0	1	0	1	1		
Gunnar Sungegg				Baratheon		300.0		2.0		10.0	0.0

value_counts()

Series 객체가 제공하는 함수

Book of Death 컬럼을 선택하여 Series 객체를 얻은 후에 value_counts()를 실행합니다.

왕좌의 게임 데이터 분석 및 시각화 실습

◎ 책 권별 사망자 수 추이 시각화

② pivot_table()

책의 권(Book of Death)별로 사망자 수 추이 시각화

```
In [108]: # 1) Series.value_counts() 함수 활용
#death['Book of Death'].value_counts()
# 2) pivot_table()
death.pivot_table(index = 'Book of Death', values = |)
```

```
Out [108]: 3.0    97
2.0    73
5.0    61
1.0    49
4.0    27
Name: Book of Death, dtype: int64
```

pivot_table() 그룹 집계

- 집계 컬럼 : Book of Death
- values : NaN이 없는 컬럼

집계 컬럼이 하나인 경우, index나 values 중 하나로 지정합니다.

왕좌의 게임 데이터 분석 및 시각화 실습

◎ 책 권별 사망자 수 추이 시각화

② pivot_table()

2) 책의 권(Book of Death)별로 사망자 수 추이 시각화

```
In [108]: # 1) Series.value_counts() 함수 활용
#death['Book of Death'].value_counts()
# 2) pivot_table()
death.pivot_table(index = 'Book of Death', values = 'Name', aggfunc = 'count')
```

Out[108]:

3.0	97
2.0	73
5.0	61

death.info()를 실행하여 각 컬럼 값에 대한 정보를 알 수 있습니다.

NaN(= Null) 값 유무가 중요한 이유는 NaN 값인 경우 카운트 수에서 배제되기 때문입니다.

왕좌의 게임 데이터 분석 및 시각화 실습

◎ 책 권별 사망자 수 추이 시각화

✓ value_counts 활용 시각화

death)별로 사망자 수 추이 시각화

```
In [113]: # 1) Series.value_counts() 함수 활용  
death['Book of Death'].value_counts().sort_index()  
# 2) pivot_table()  
#death.pivot_table(index = 'Book of Death', values = 'Name', aggfunc = 'count')
```

```
Out[113]: 3.0    97  
          2.0    73  
          5.0    61  
          1.0    49  
          4.0    27  
          Name: Book of Death, dtype: int64
```

sort_index()

인덱스 라벨 순으로 오름차순 정렬

In []:

In []:

In []:

In []:

왕좌의 게임 데이터 분석 및 시각화 실습

◎ 책 권별 사망자 수 추이 시각화

✓ value_counts 활용 시각화

```
(index = 'Book of Death', values = 'Name', aggfunc = 'count')
```

```
In [119]: bookOfDeath.plot(figsize = (5,3),  
                        xticks = bookOfDeath.index,  
                        ylim = [0, 120])
```

Out[119]: <AxesSubplot:>



xticks를 book of death의 인덱스, 즉, [1,2,3,4,5]로 지정합니다.

ylim 값을 [0, 120]으로 지정합니다.

왕좌의 게임 데이터 분석 및 시각화 실습

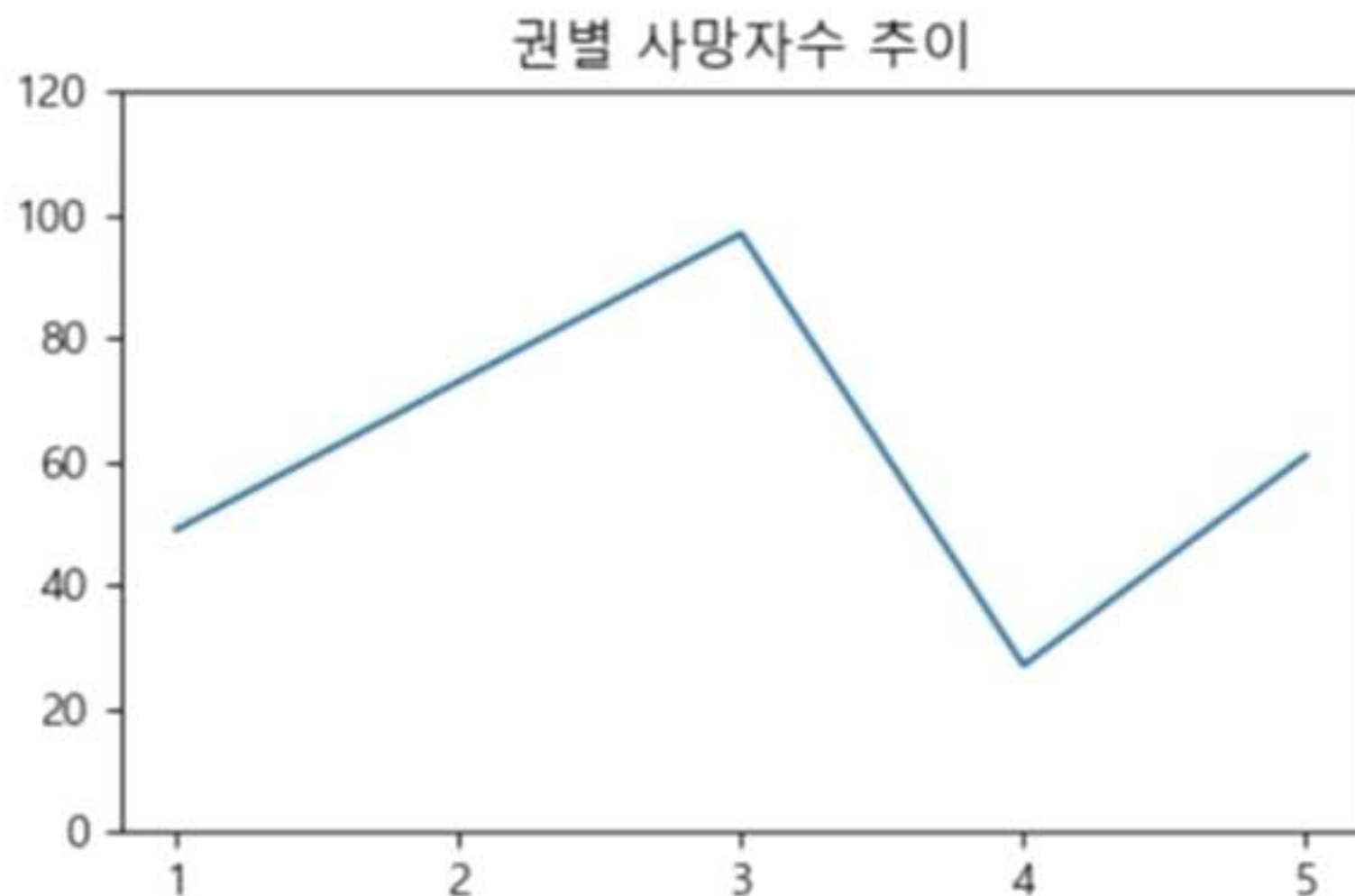
◎ 책 권별 사망자 수 추이 시각화

✓ value_counts 활용 시각화

```
figsize = (5,3),  
ticks = bookOfDeath.index,  
ylim = [0, 120],  
title = '권별 사망자수 추이')
```

Out[121]: <AxesSubplot:title={'center': '권별 사망자수 추이'}>

title 제목 지정



왕좌의 게임 데이터 분석 및 시각화 실습

◎ 책 권별 사망자 수 추이 시각화

✓ Figure와 Axes 간의 관계

```
In [121]: bookOfDeath.plot(figsize = (5,3),  
                        xticks = bookOfDeath.index,  
                        ylim = [0, 120],  
                        title = '권별 사망자수 추이')
```

```
Out[121]: <AxesSubplot:title={'center':'권별 사망자수 추이'}>
```

권별 사망자수 추이

AxesSubplot

plot 함수 실행되고 나서 리턴되는 리턴된 객체

Axes 객체와 동일

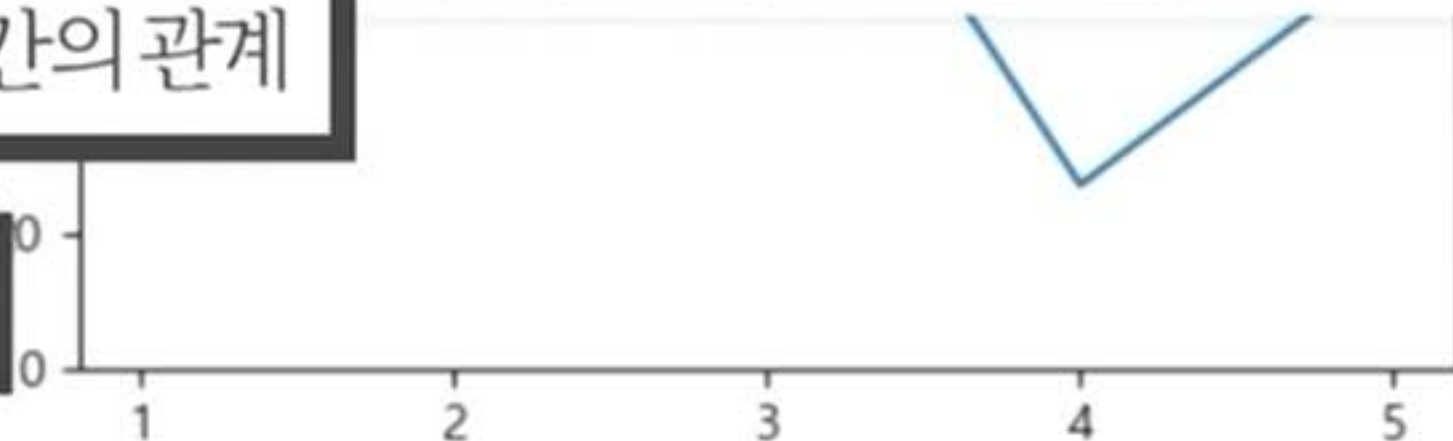
plot() 함수 실행하면 AxesSubplot가 나타납니다.

왕좌의 게임 데이터 분석 및 시각화 실습

◎ 책 권별 사망자 수 추이 시각화

✓ Figure와 Axes 간의 관계

✓ 저장 및 출력



```
In [123]: ax
```

```
Out[123]: <AxesSubplot:title={'center': '권별 사망자수 추이'}>
```

```
In [ ]:
```

3) 대규모 전투의 공격진영과 수비 진영의 숫자 시각화

대규모 전투 = 공격과 수비 모두 합쳐서 10,000명 이상 참가한 전쟁

axes만 출력하면 axes에 대한 정보만 나오고, figure를 출력해야 그래프가 보이게 됩니다.

```
In [ ]:
```

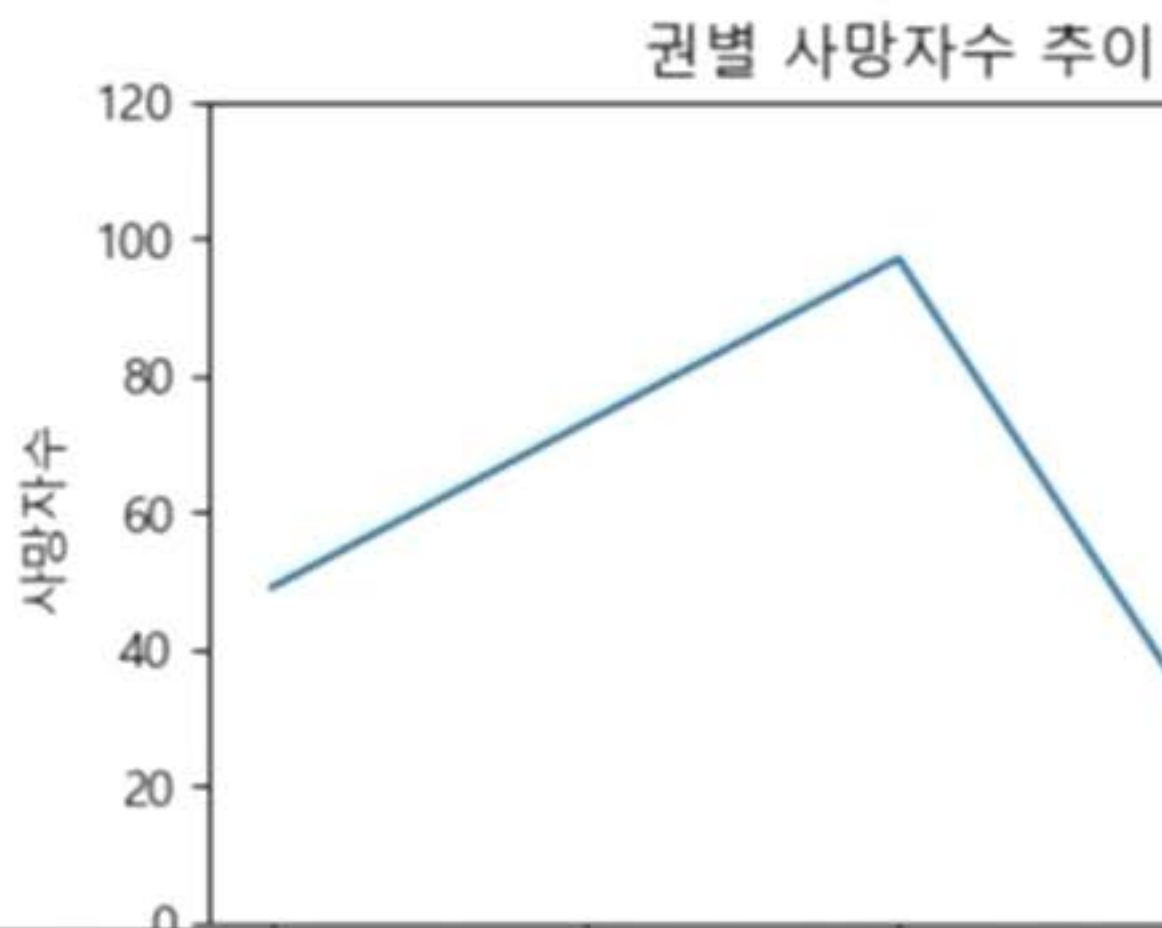
왕좌의 게임 데이터 분석 및 시각화 실습

◎ 책 권별 사망자 수 추이 시각화

✓ Figure와 Axes 간의 관계

```
In [126]: ax.figure
```

```
Out[126]:
```



set_xlabel

x축 제목 지정

set_ylabel

y축 제목 지정

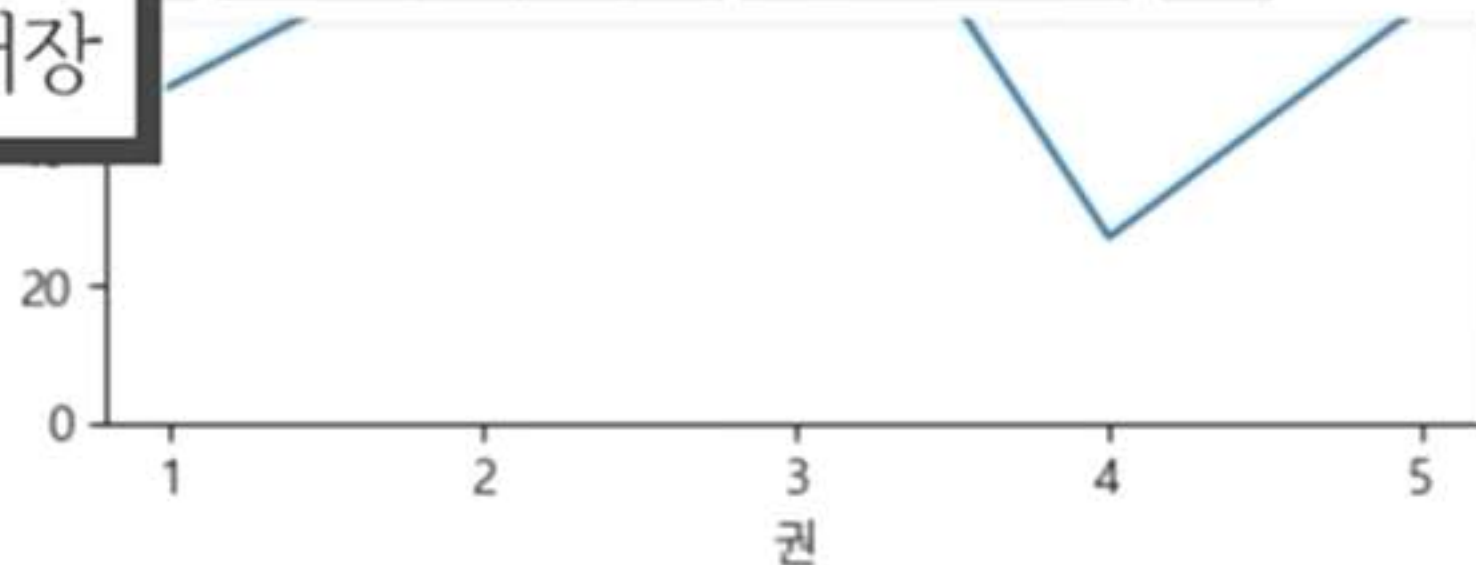
Axes의 함수

axes와 figure의 관계를 잘 이해해야 합니다.

왕좌의 게임 데이터 분석 및 시각화 실습

◎ 책 권별 사망자 수 추이 시각화

✓ 그래프를 파일로 저장



```
In [127]: # 그래프를 파일로 저장  
ax.savefig('test.jpg')
```

```
AttributeError                                Traceback (most recent call last)  
<ipython-input-127-b9f1a5f440a2> in <module>  
    1 # 그래프를 파일로 저장  
----> 2 ax.savefig('test.jpg')
```

```
AttributeError: 'AxesSubplot' object has no attribute 'savefig'
```

ax.savefig로 실행하면 에러가 발생합니다.

no attribute

```
In [ ]:
```

왕좌의 게임 데이터 분석 및 시각화 실습

◎ 책 권별 사망자 수 추이 시각화

✓ 그래프를 파일로 저장

주의하기

✓ axes에서 제공하는 함수와 figur에서 제공하는 함수들은 다르므로, axes와 figure의 차이를 잘 알고 있어야 합니다.

```
In [126]: ax.figure
```

```
Out[126]:
```

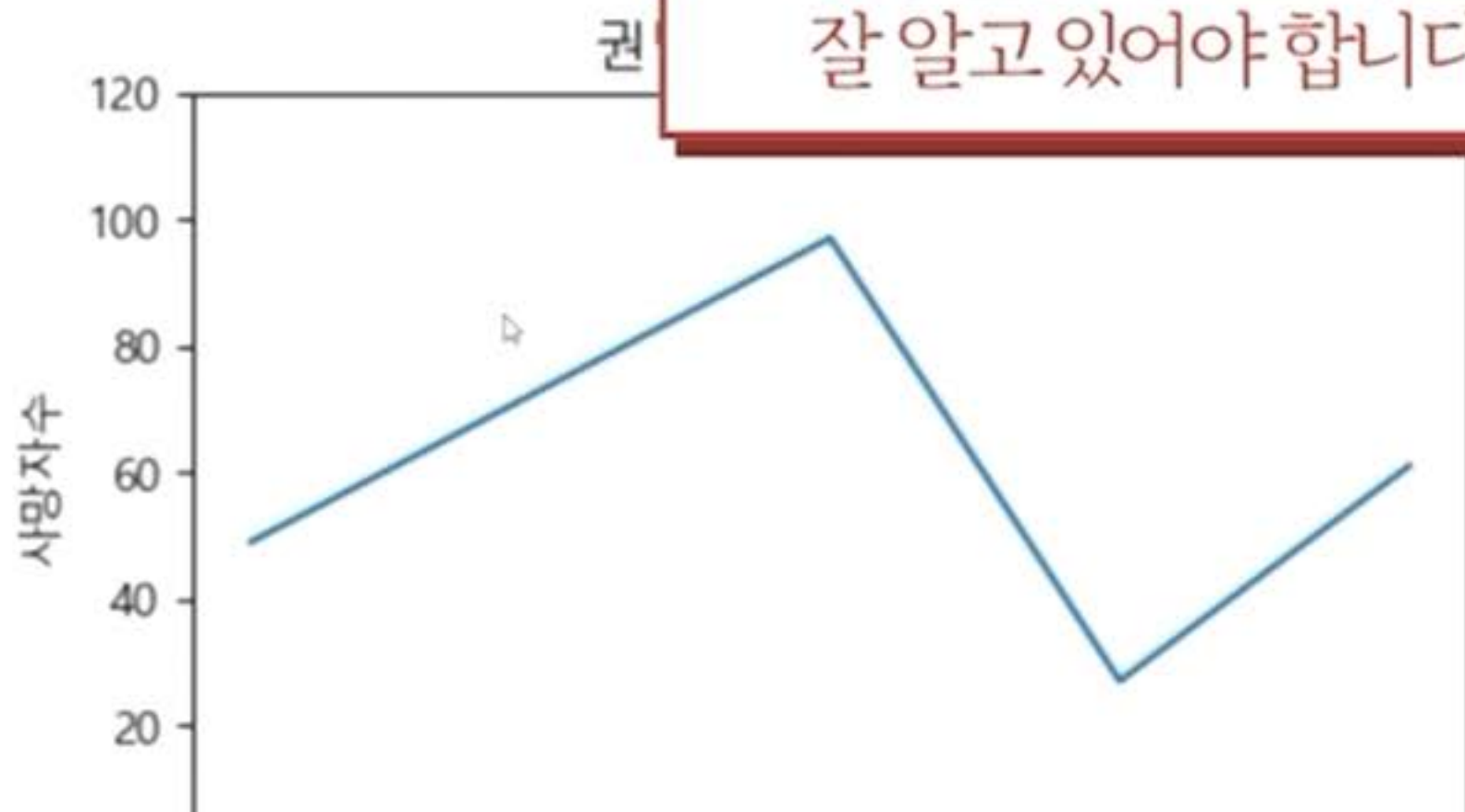


figure 선택 후, savefig 실행해야 정상적으로 파일이 생성됩니다.

```
→ 2 ax.savefig('test.jpg')
```

03

왕좌의 게임 데이터 분석 [2]



왕좌의 게임 데이터 분석 및 시각화 실습

◎ 대규모 전투의 공격 진영과 수비 진영의 숫자 시각화

✓ 대규모 전투만 선택

```
In [128]: # 그래프를 파일로 저장  
ax.figure.savefig('test.jpg')
```

```
In [ ]:
```

battles 조건 설정

$\text{attacker_size} + \text{defender_size} \geq 10,000$

3) 대규모 전투의 공격진영과 수비 진영의 숫자 시각화

대규모 전투 = 공격과 수비 모두 합쳐서 10,000명 이상 참가한 전쟁

```
In [130]: #1. 대규모 전투만 선택  
battles.attacker_size + battles.defender_size >= 10000
```

```
Out [130]: 0    True  
          1    False  
          2    True  
          3    True  
          4    False  
          5    True  
          6    False  
          7    False  
          8    False
```

대규모 전투

공격과 수비 모두 합쳐서 10,000명 이상 참가한 전쟁

왕좌의 게임 데이터 분석 및 시각화 실습

◎ 대규모 전투의 공격 진영과 수비 진영의 숫자 시각화

✓ 필요한 컬럼만 선택

파일로 저장
fig.savefig('test.jpg')

In []:

3) 대규모 전투의 공격진영과 수비 진영의 숫자 시각화

대규모 전투 = 공격과 수비 모두 합쳐서 10,000명 이상 참가한

In [132]: #1. 대규모 전투만 선택
big_battles = battles[battles.attacker_size + battles.defender_size >= 10000]

In []: #2. 필요한 컬럼만 선택
big_battles[['name', 'attacker_size', 'defender_size']]

In []:

In []:

In []:

필요한 컬럼

- ✓ name
- ✓ attacker_size
- ✓ defender_size

왕좌의 게임 데이터 분석 및 시각화 실습

◎ 대규모 전투의 공격 진영과 수비 진영의 숫자 시각화

✓ 시각화

3) 대규모 전투의 공격진영과 수비 진영의 숫자 시각화

대규모 전투 = 공격과 수비 모두 합쳐서 10,000명 이상 참가한 전쟁

```
In [132]: #1. 대규모 전투만 선택  
big_battles = battles[battles.at
```

plot함수 bar 그래프 실행

```
In [134]: #2. 필요한 컬럼만 선택  
big_battles = big_battles[['name', 'attacker_size', 'defender_size']]
```

```
In [ ]: # 3. 시각화  
big_battles.plot(kind = 'bar')
```

```
In [ ]:
```

```
In [ ]:
```

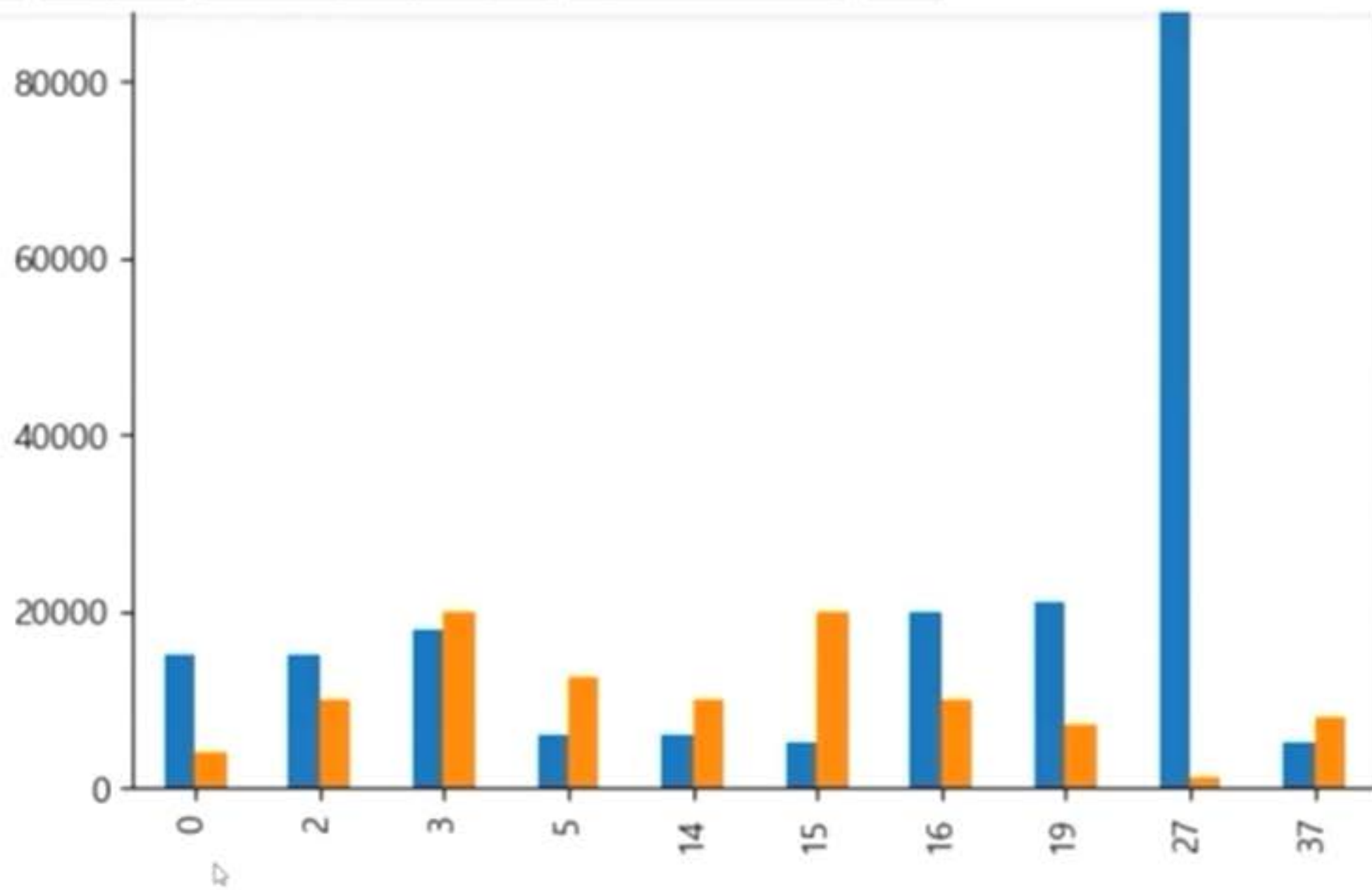
```
In [ ]:
```

```
In [ ]:
```


왕좌의 게임 데이터 분석 및 시각화 실습

◎ 대규모 전투의 공격 진영과 수비 진영의 숫자 시각화

✓ 시각화



x축은 기본적으로 로우 인덱스가 지정됩니다.

왕좌의 게임 데이터 분석 및 시각화 실습

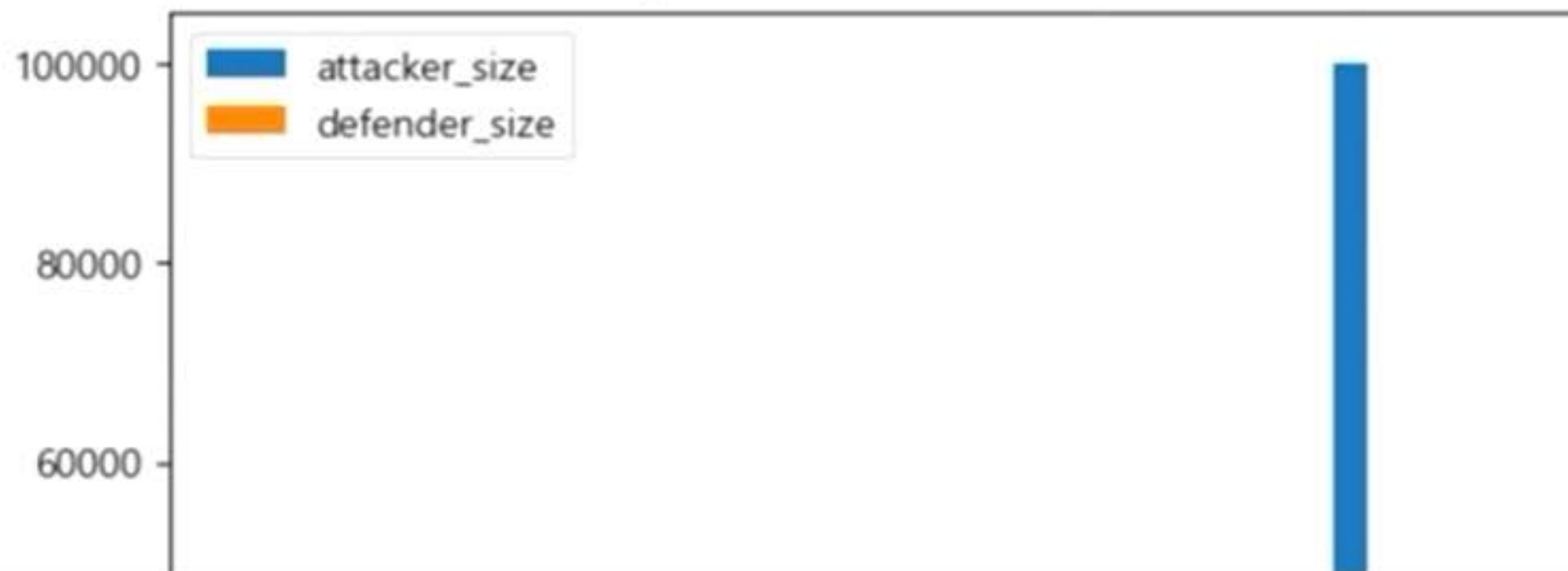
◎ 대규모 전투의 공격 진영과 수비 진영의 숫자 시각화

✓ 시각화

```
[134]: #2. 필요한 컬럼만 선택  
big_battles = big_battles[['name', 'attacker_size', 'defender_size']]
```

```
In [135]: # 3. 시각화  
big_battles.plot(kind = 'bar',  
                  x = 'name')
```

Out [135]: <AxesSubplot:>



name 컬럼을 로우 인덱스로 지정하거나, x인자로 지정할 수 있습니다.

왕좌의 게임 데이터 분석 및 시각화 실습

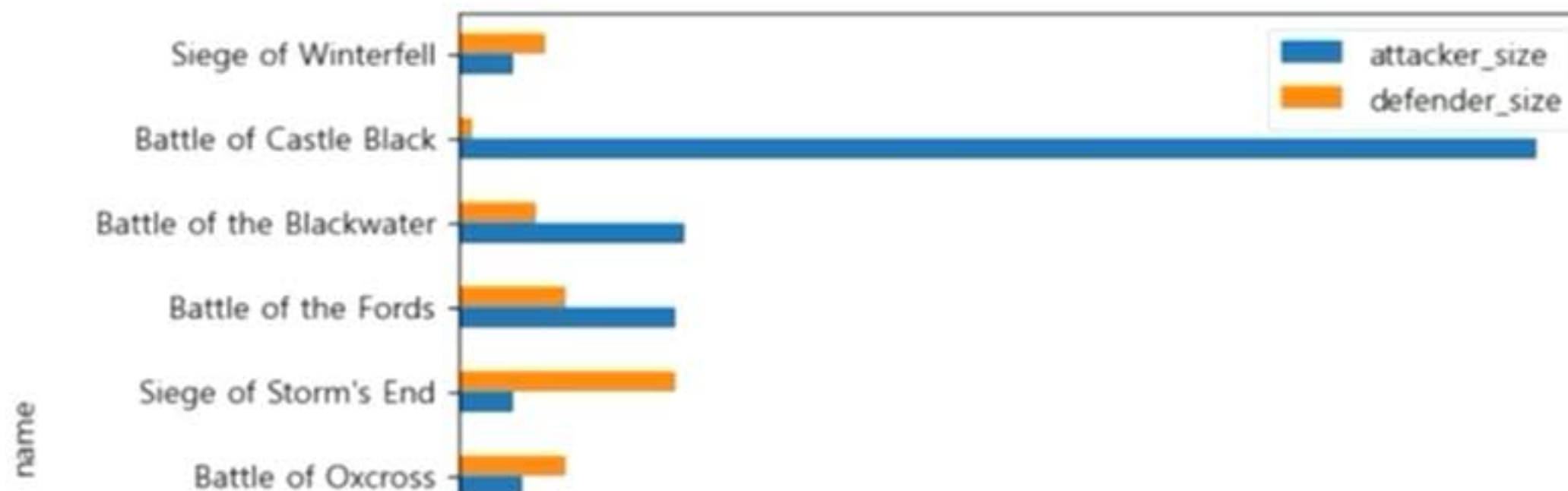
◎ 대규모 전투의 공격 진영과 수비 진영의 숫자 시각화

✓ 시각화

```
big_battles = big_battles[['name', 'attacker_size', 'defender_size']]
```

```
In [137]: # 3. 시각화  
big_battles.plot(kind = 'barh',  
                 x = 'name')
```

```
Out[137]: <AxesSubplot:ylabel='name'>
```



attacker와 defender의 격차를 한 눈에 볼 수 있으나,
가장 많은 병사가 투입된 전쟁은 한 눈에 보기 어렵습니다.

Battle of Riverrun

왕좌의 게임 데이터 분석 및 시각화 실습

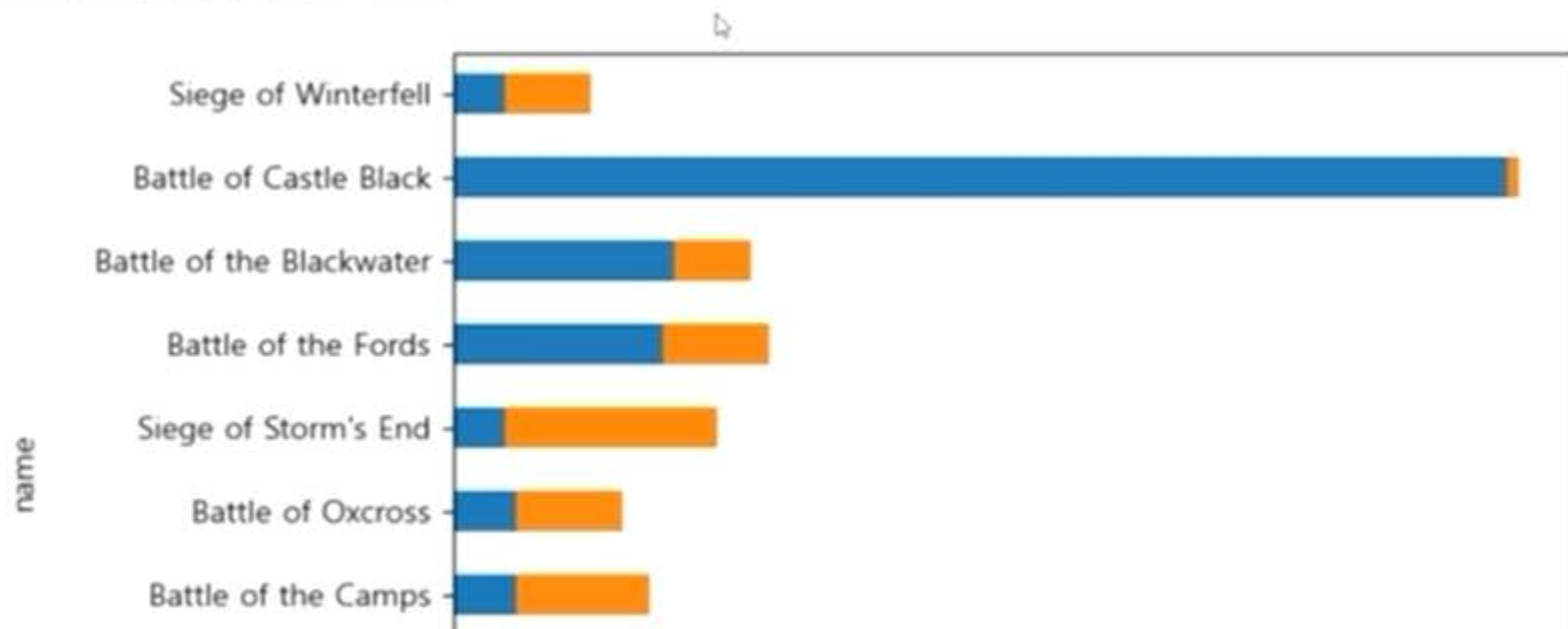
◎ 대규모 전투의 공격 진영과 수비 진영의 숫자 시각화

✓ 시각화

```
[134]: #2. 필요한 컬럼만 선택  
big_battles = big_battles[['name', 'attacker_size', 'defender_size']]
```

```
In [138]: # 3. 시각화  
big_battles.plot(kind = 'barh',  
                 x = 'name',  
                 stacked = True)
```

```
Out [138]: <AxesSubplot: ylabel='name'>
```



stacked

기본값 = False

True로 변경하여 누적 막대 표시

데이터 시각화



학습완료

- 1/ 왕좌의 게임 데이터 소개
- 2/ 왕좌의 게임 데이터 분석 [1]
 - 책 권별 사망자 수 추이 시각화
- 3/ 왕좌의 게임 데이터 분석 [2]
 - 전쟁의 공격/수비 진영 규모 시각화



수고하셨습니다!