

## **CLASS 9. 파이썬 – 시각화 및 분석(3)**

## 수업 목차

### Part1. 트래킹데이터(심화)

#1. Release Point

#2. Extension



### Part2. 선수의 능력치

#3. 방사형그래프(Radar Chart)



### Part3. Box Plot

강의자료

sample\_6.xlsx

sample\_13.xlsx

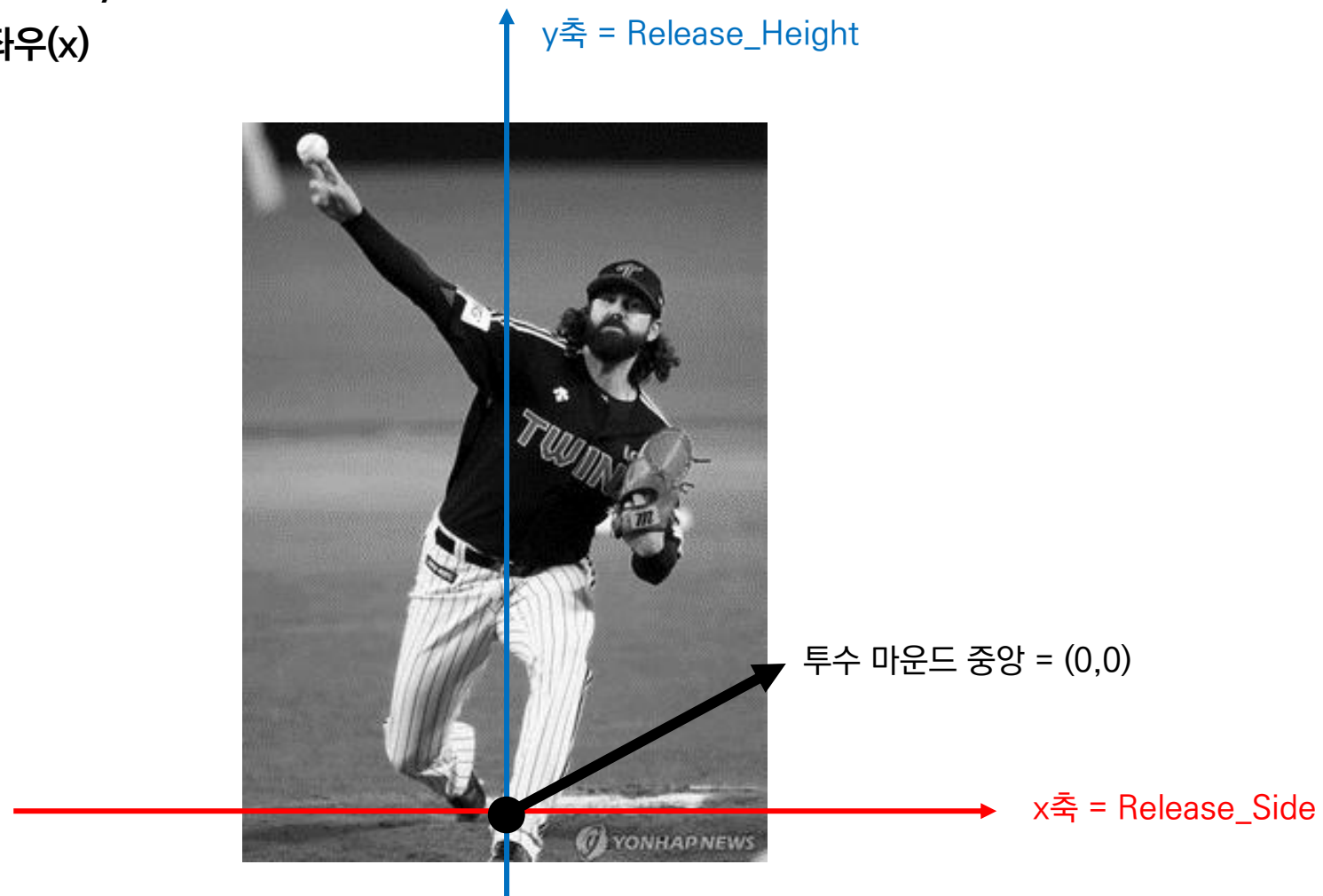
sample\_14.xlsx

## Part1. 트래킹데이터(심화)

## 이론 #1. Release Point : 투수가 공을 던지는 순간의 위치

Release\_Height(m) = 높이(y)

Release\_Side(m) = 좌우(x)



## 실습 #1. 투수 2명의 릴리스 포인트 비교

(예) pitcher1, pitcher2의 릴리스 포인트 비교 산점도(scatter plot)

#1. 환경설정 : 한글 폰트 설치

```
import matplotlib as mpl
import matplotlib.pyplot as plt

mpl.rcParams['axes.unicode_minus'] = False
plt.rcParams['font.family'] = 'NanumGothic'
```

#2. 파일 불러오기 : sample\_13.xlsx(pitcher1), sample\_14.xlsx(pitcher2)

```
import pandas as pd

df1 = pd.read_excel('sample_13.xlsx', engine='openpyxl')
df2 = pd.read_excel('sample_14.xlsx', engine='openpyxl')
```

## 실습 #1. 투수 2명의 릴리스 포인트 비교

(예) pitcher1, pitcher2의 릴리스 포인트 비교 산점도(scatter plot)

```
plt.figure(figsize=(10,10))
plt.xlim(-3,3)
plt.ylim(-3,3)
plt.grid()

x1 = round(df1['Release_Side'].mean(),2)
y1 = round(df1['Release_Height'].mean(),2)
plt.text(x1, y1, '('+str(x1)+' , '+str(y1)+')')

x2 = round(df2['Release_Side'].mean(),2)
y2 = round(df2['Release_Height'].mean(),2)
plt.text(x2, y2, '('+str(x2)+' , '+str(y2)+')')

plt.scatter(x1, y1, s=300, alpha=0.5, label="pitcher1")
plt.scatter(x2, y2, s=300, alpha=0.5, label="pitcher2")
plt.legend()
plt.xlabel('Release_Side (m)')
plt.ylabel('Release_Height (m)')
plt.show()
```

이론 #2. Extension(m) : 마운드 ~ 투수가 공을 던지는 순간의 위치(타자에게 얼마나 가까이에서 투구했는지)



## 실습 #2. 투수 2명의 릴리스하이트-익스텐션 비교

(예) pitcher1, pitcher2의 릴리스하이트-익스텐션 비교 산점도(scatter plot)

#1. 환경설정 : 한글 폰트 설치

```
import matplotlib as mpl
import matplotlib.pyplot as plt

mpl.rcParams['axes.unicode_minus'] = False
plt.rcParams['font.family'] = 'NanumGothic'
```

#2. 파일 불러오기 : sample\_13.xlsx(pitcher1), sample\_14.xlsx(pitcher2)

```
import pandas as pd

df1 = pd.read_excel('sample_13.xlsx', engine='openpyxl')
df2 = pd.read_excel('sample_14.xlsx', engine='openpyxl')
```



## 실습 #2. 투수 2명의 릴리스하이트-익스텐션 비교

(예) pitcher1, pitcher2의 릴리스하이트-익스텐션 비교 산점도(scatter plot)

```
plt.figure(figsize=(10,10))
plt.xlim(0,3)
plt.ylim(0,3)
plt.grid()

x1 = round(df1['Extension'].mean(),2)
y1 = round(df1['Release_Height'].mean(),2)
plt.text(x1, y1, '('+str(x1)+' , '+str(y1)+')')

x2 = round(df2['Extension'].mean(),2)
y2 = round(df2['Release_Height'].mean(),2)
plt.text(x2, y2, '('+str(x2)+' , '+str(y2)+')')

plt.scatter(x1, y1, s=300, alpha=0.5, label="pitcher1")
plt.scatter(x2, y2, s=300, alpha=0.5, label="pitcher2")
plt.legend()
plt.xlabel('Extension (m)')
plt.ylabel('Release_Height (m)')
plt.show()
```

## Part2. 선수의 능력치

### 실습 #3. 축구선수 능력치 그래프 그리기(데이터 출처 : FIFA21)

URL : <https://www.ea.com/ko-kr/games/fifa/fifa-21/ratings/ratings-database>



방사형 차트(Radar Chart)

### 실습 #3. 축구선수 능력치 그래프 그리기(데이터 출처 : FIFA21)

URL : <https://www.ea.com/ko-kr/games/fifa/fifa-21/ratings/ratings-database>

#1. 환경설정 : 한글 폰트 설치

```
import matplotlib as mpl
import matplotlib.pyplot as plt

mpl.rcParams['axes.unicode_minus'] = False
plt.rcParams['font.family'] = 'NanumGothic'
```

#2. 파일 불러오기 : sample\_6.xlsx

```
import pandas as pd

df = pd.read_excel('sample_6.xlsx', engine='openpyxl')
df.head()
```

### 실습 #3. 축구선수 능력치 그래프 그리기(데이터 출처 : FIFA21)

URL : <https://www.ea.com/ko-kr/games/fifa/fifa-21/ratings/ratings-database>

```
import numpy as np

avg = [df['페이스'].mean(), df['슈팅'].mean(), df['패스'].mean(), df['드리블'].mean(), df['수비'].mean(), df['체력'].mean(), df['페이스'].mean()]
son = [df['페이스'][2], df['슈팅'][2], df['패스'][2], df['드리블'][2], df['수비'][2], df['체력'][2], df['페이스'][2]]

theta = np.linspace(start=0, stop=2*np.pi, num=len(son))

plt.figure(figsize=(10,10))
plt.subplot(polar=True)
plt.xticks(theta, labels=['페이스', '슈팅', '패스', '드리블', '수비', '체력', '페이스'], fontsize=11)

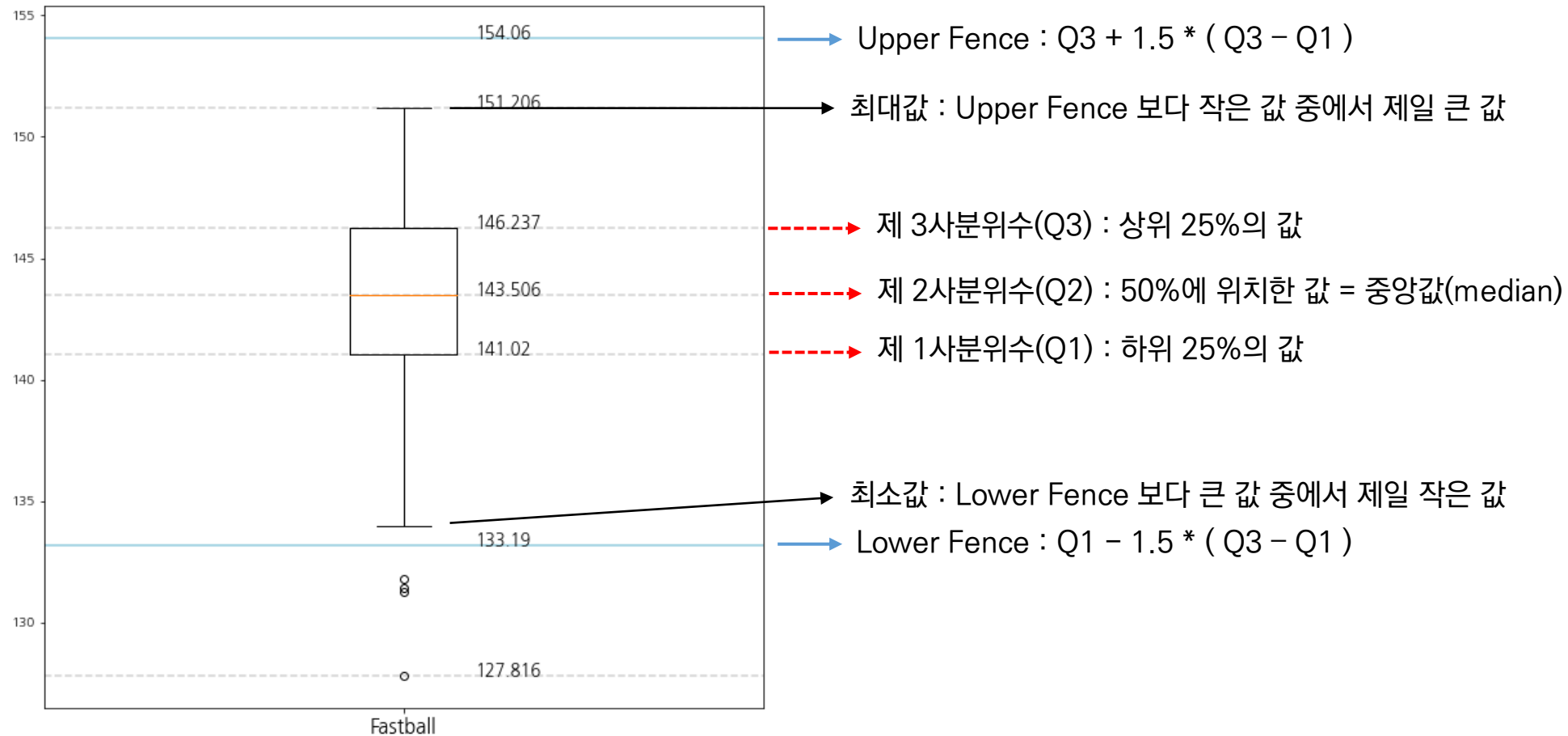
#전체 데이터 평균
plt.plot(theta, avg, color="gray", label="avg", linestyle="--")
plt.fill(theta, avg, color="gray", alpha=0.1)

#손흥민 데이터
plt.plot(theta, son, color="red", label="son")
plt.fill(theta, son, color="red", alpha=0.2)

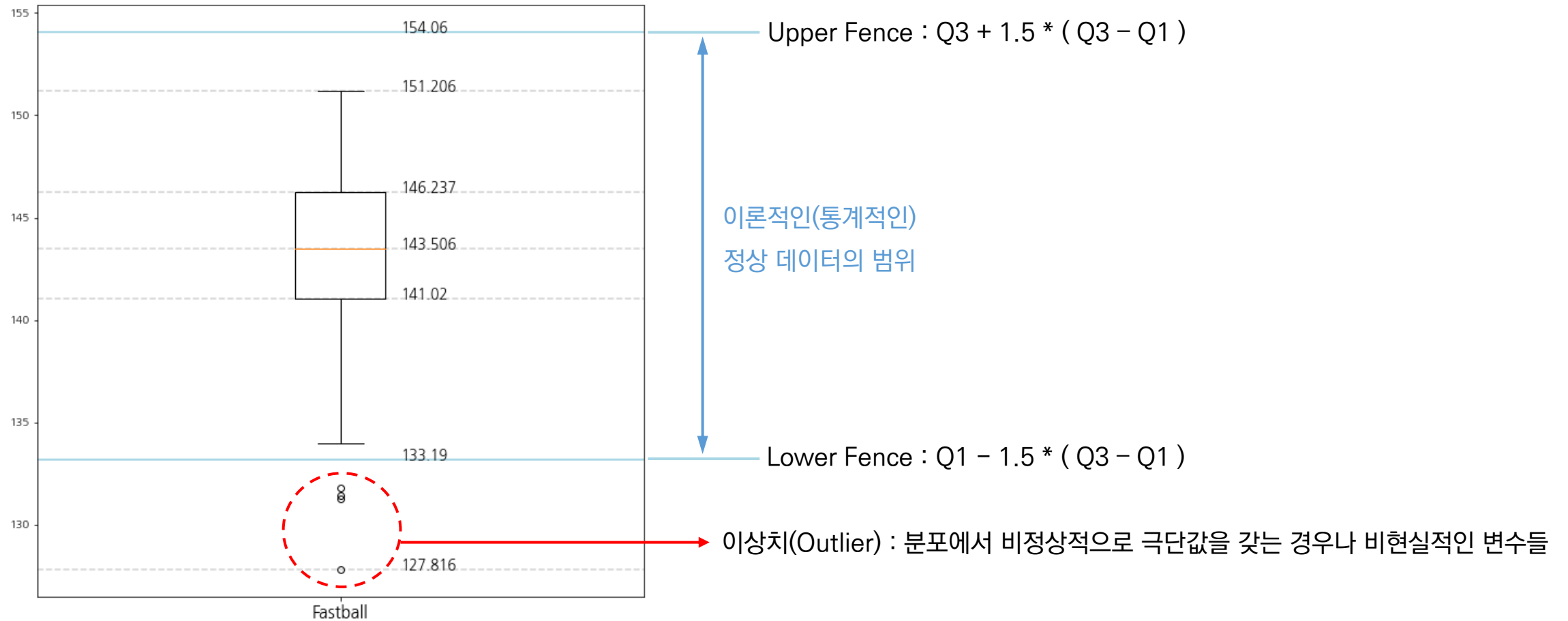
plt.legend(fontsize=13)
plt.show()
```

## Part3. Box Plot

### 이론 #3. Box Plot : 데이터의 분포를 파악하는 그래프



### 이론 #3. Box Plot : 데이터의 분포를 파악하는 그래프





## 실습 #4. 투수 1명의 서로 다른 구종의 분포 비교

(예) pitcher1이라는 투수의 Fastball, Curveball 구속 Box Plot

#1. 환경설정 : 한글 폰트 설치

```
import matplotlib as mpl
import matplotlib.pyplot as plt

mpl.rcParams['axes.unicode_minus'] = False
plt.rcParams['font.family'] = 'NanumGothic'
```

#2. 파일 불러오기 : sample\_13.xlsx

```
import pandas as pd

df = pd.read_excel('sample_13.xlsx', engine='openpyxl')
df.head()
```

## 실습 #4. 투수 1명의 서로 다른 구종의 분포 비교

(예) pitcher1이라는 투수의 Fastball, Curveball 구속 Box Plot

```
filt1 = df['Pitch_Type'] == 'Fastball'  
filt4 = df['Pitch_Type'] == 'Curveball'
```

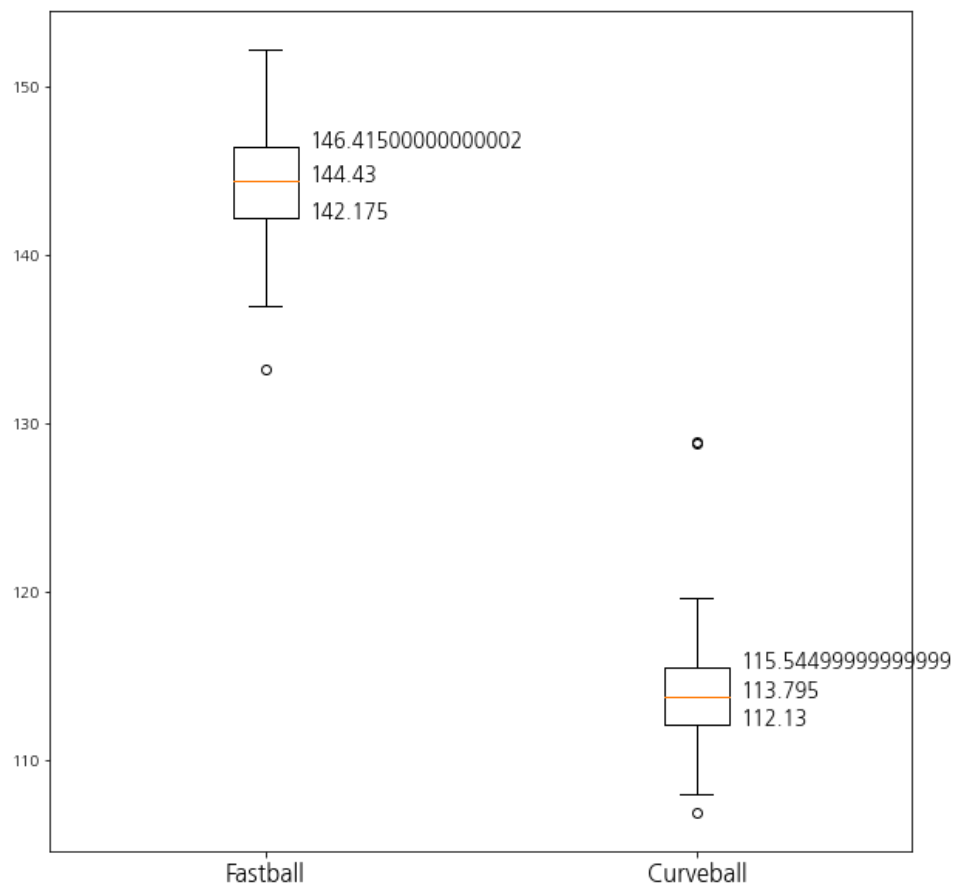
#그래프 옵션

```
plt.figure(figsize=(10,10))  
plt.text(1.1,df[filt1]['Pitch_Speed'].median(), df[filt1]['Pitch_Speed'].median(), fontsize=13)  
plt.text(2.1,df[filt4]['Pitch_Speed'].median(), df[filt4]['Pitch_Speed'].median(), fontsize=13)  
plt.text(1.1,df[filt1]['Pitch_Speed'].quantile(q=0.25), df[filt1]['Pitch_Speed'].quantile(q=0.25), fontsize=13)  
plt.text(1.1,df[filt1]['Pitch_Speed'].quantile(q=0.75), df[filt1]['Pitch_Speed'].quantile(q=0.75), fontsize=13)  
plt.text(2.1,df[filt4]['Pitch_Speed'].quantile(q=0.25), df[filt4]['Pitch_Speed'].quantile(q=0.25), fontsize=13)  
plt.text(2.1,df[filt4]['Pitch_Speed'].quantile(q=0.75), df[filt4]['Pitch_Speed'].quantile(q=0.75), fontsize=13)
```

```
data = [df[filt1]['Pitch_Speed'], df[filt4]['Pitch_Speed']]  
plt.boxplot(data)  
plt.xticks([1,2], ['Fastball','Curveball'], fontsize=15)  
plt.show()
```

## 실습 #4. 투수 1명의 서로 다른 구종의 분포 비교

(예) pitcher1이라는 투수의 Fastball, Curveball 구속 Box Plot



[해석 (예)]

1) Fastball 요약

$Q1 = 142.175$  /  $Q2 = \text{median} = 144.43$  /  $Q3 = 146.42$

2) Curveball 요약

$Q1 = 112.13$  /  $Q2 = \text{median} = 113.795$  /  $Q3 = 115.54$

3) 구종(Fastball, Curveball)에 따른 구속의 분포를 나타내고 있다.

4) 전반적으로 Fastball의 구속이 더 높다.

5) Fastball의 경우 중앙값이 Q3으로 치우쳐 있는 것으로 보아

대부분의 투구가 141km/h ~ 145km/h 정도에 형성되는것을 알 수 있다.

6) Curveball의 경우 이상치 3개가 판단됐다.

7) Fastball의 경우 이상치 1개가 판단됐다.

## 실습 #5. 서로 다른 투수의 같은 구종의 분포 비교

(예) pitcher1과 pitcher2의 Fastball 구속 Box Plot

#1. 환경설정 : 한글 폰트 설치

```
import matplotlib as mpl
import matplotlib.pyplot as plt

mpl.rcParams['axes.unicode_minus'] = False
plt.rcParams['font.family'] = 'NanumGothic'
```

#2. 파일 불러오기 : sample\_13.xlsx(pitcher1), sample\_14.xlsx(pitcher2)

```
import pandas as pd

df1 = pd.read_excel('sample_13.xlsx', engine='openpyxl')
df2 = pd.read_excel('sample_14.xlsx', engine='openpyxl')
```

## 실습 #5. 서로 다른 투수의 같은 구종의 분포 비교

(예) pitcher1과 pitcher2의 Fastball 구속 Box Plot

```
filt1 = df1['Pitch_Type'] == 'Fastball'  
filt2 = df2['Pitch_Type'] == 'Fastball'
```

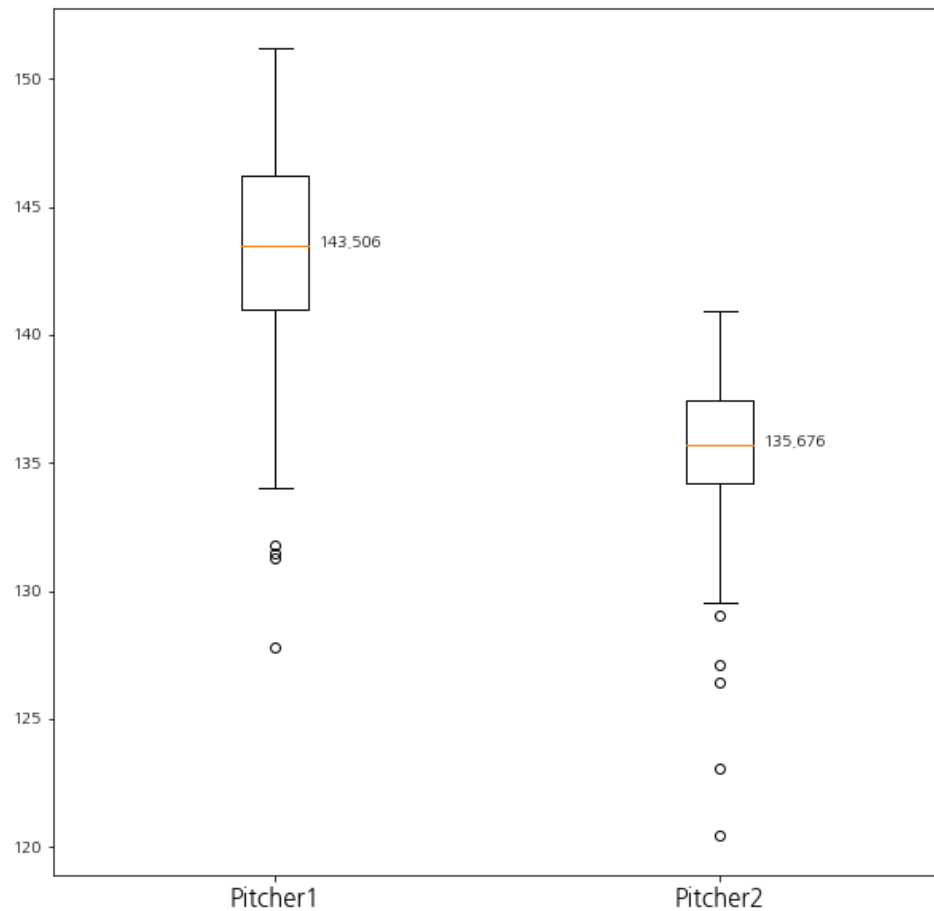
#그래프 옵션

```
plt.figure(figsize=(10,10))  
plt.text(1.1, df1[filt1]['Pitch_Speed'].median(), df1[filt1]['Pitch_Speed'].median())  
plt.text(2.1, df2[filt2]['Pitch_Speed'].median(), df2[filt2]['Pitch_Speed'].median())
```

```
data = [df1[filt1]['Pitch_Speed'], df2[filt2]['Pitch_Speed']]  
plt.boxplot(data)  
plt.xticks([1,2], ['Pitcher1','Pitcher2'], fontsize=15)  
plt.show()
```

## 실습 #5. 서로 다른 투수의 같은 구종의 분포 비교

(예) pitcher1과 pitcher2의 Fastball 구속 Box Plot



[해석 (예)]

- 1) Pitcher 1은 중앙값이 143.506km/h,  
Pitcher 2는 중앙값이 135.676km/h로 나타난다.
- 2) 전반적으로 pitcher 1의 Fastball의 구속이 더 높은 것을 알 수 있다.