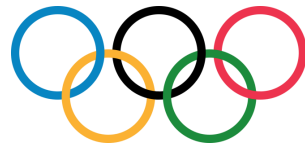


# 120 Years History Of Olympic Games

**5조** 1602104 전해린 1685102 양혜지  
1685128 이유진 1829008 김민영  
1829017 김하영



# Olympic Game Dataset

120년간 시행된 올림픽 출전 국가, 선수, 메달 수 등에 대한 자료



## <Character>

Name - 선수명

Team - 팀명

NOC - 국가코드

City - 개최 도시

Sport - 종목

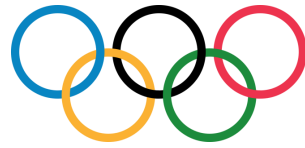
Event - 세부종목

Games - 연도와 계절

## <Numeric>

Height - 키 (단위 cm)

Weight - 몸무게 (단위 kg)



# Olympic Game Dataset

120년간 시행된 올림픽 출전 국가, 선수, 메달 수 등에 대한 자료



## <Integer>

ID - 선수 식별 번호

Age - 나이

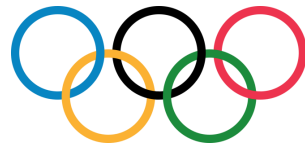
Year - 연도

## <Categorical>

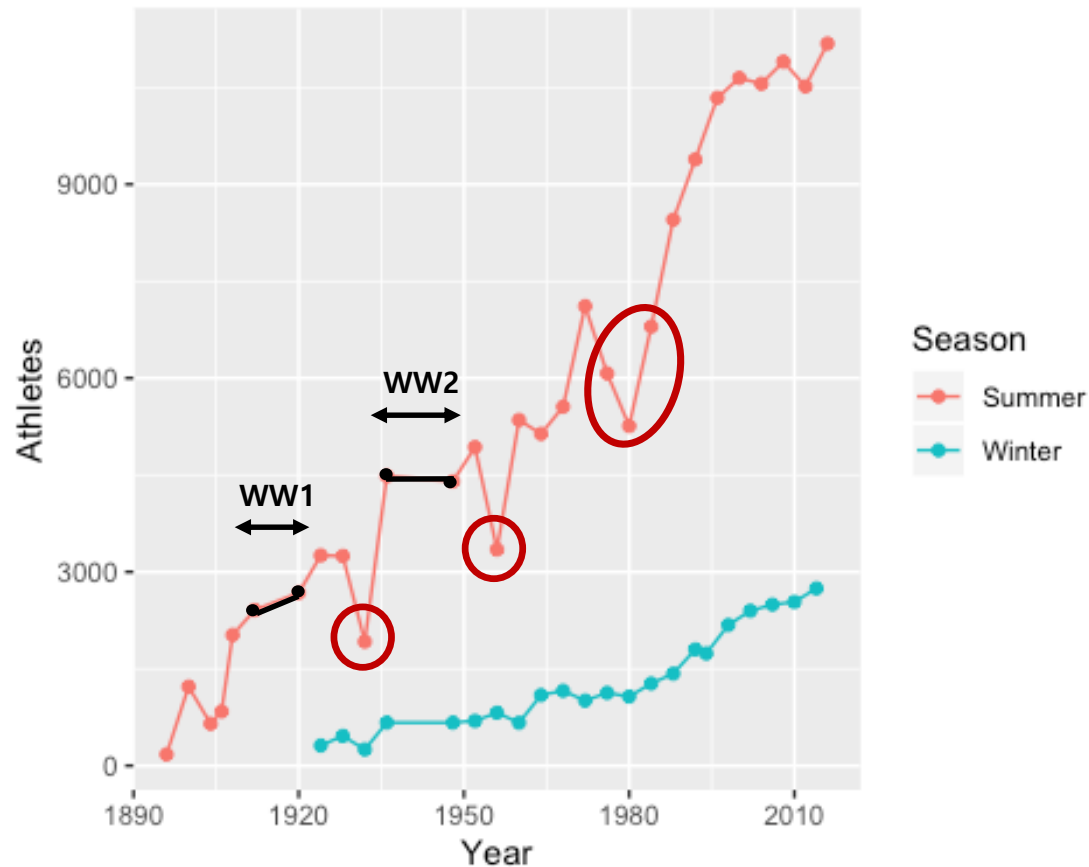
Sex - 성별 (M or F)

Season - 계절 (Summer or Winter)

Medal - 메달 (Gold, Silver, Bronze, or NA)



## 연도별 출전 선수



1935년, 1955년과 1980년경

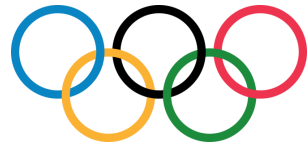
세 점에서 대폭 감소하는 것이 눈에 띈다.

- 1932년 - 미국 경제 대공황
- 1956년 - 제 2차 중동 전쟁, 헝가리 혁명, 중국의 보이콧
- 1976년, 1980년, 1984년 - 차례대로 아프리카 28개국의 보이콧, 냉전 당시 서방 진영의 보이콧, 동방 진영의 보복 보이콧

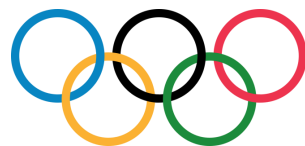
1920년 직전과 1935~1950년경

연속한 두 점을 잇는 선의 길이가 유독 길다는 것을 확인할 수 있다.

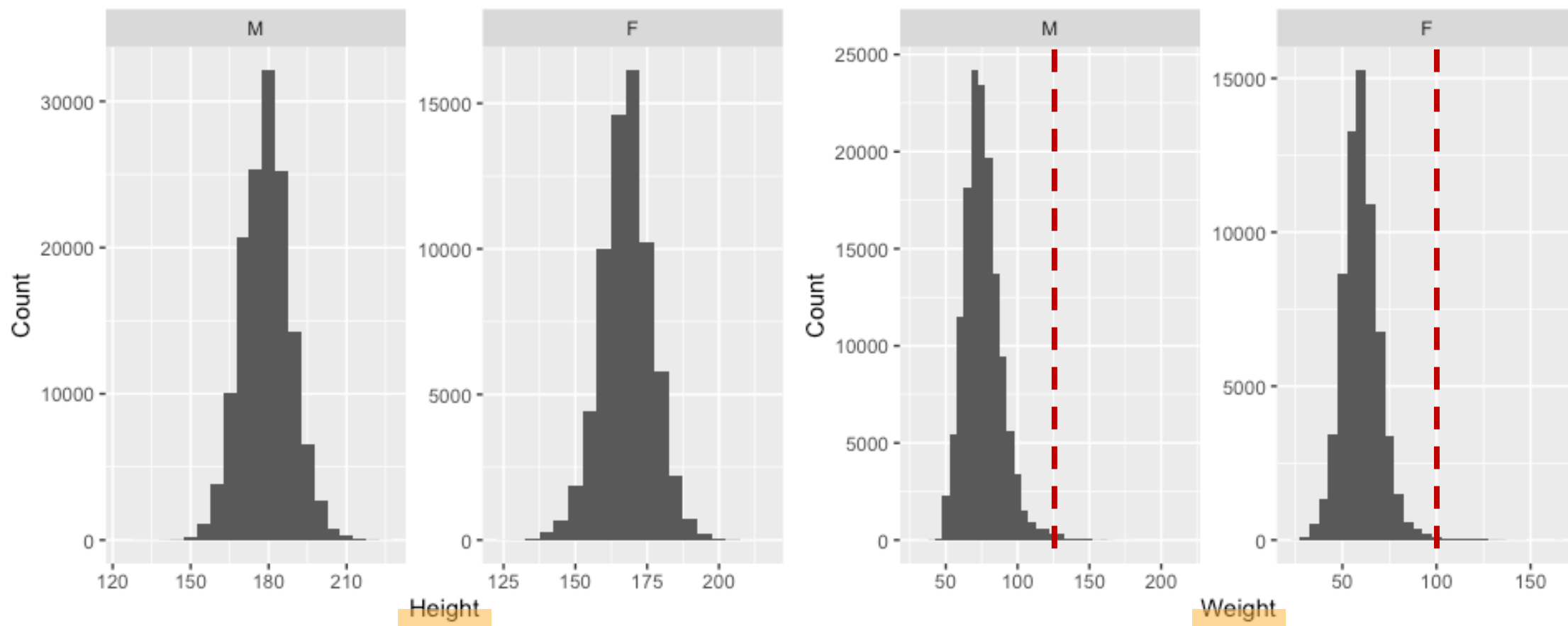
이는 이 기간에 한동안 올림픽이 중단되었음을 의미한다.

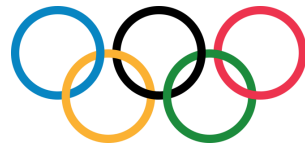


# 1. Height & Weight

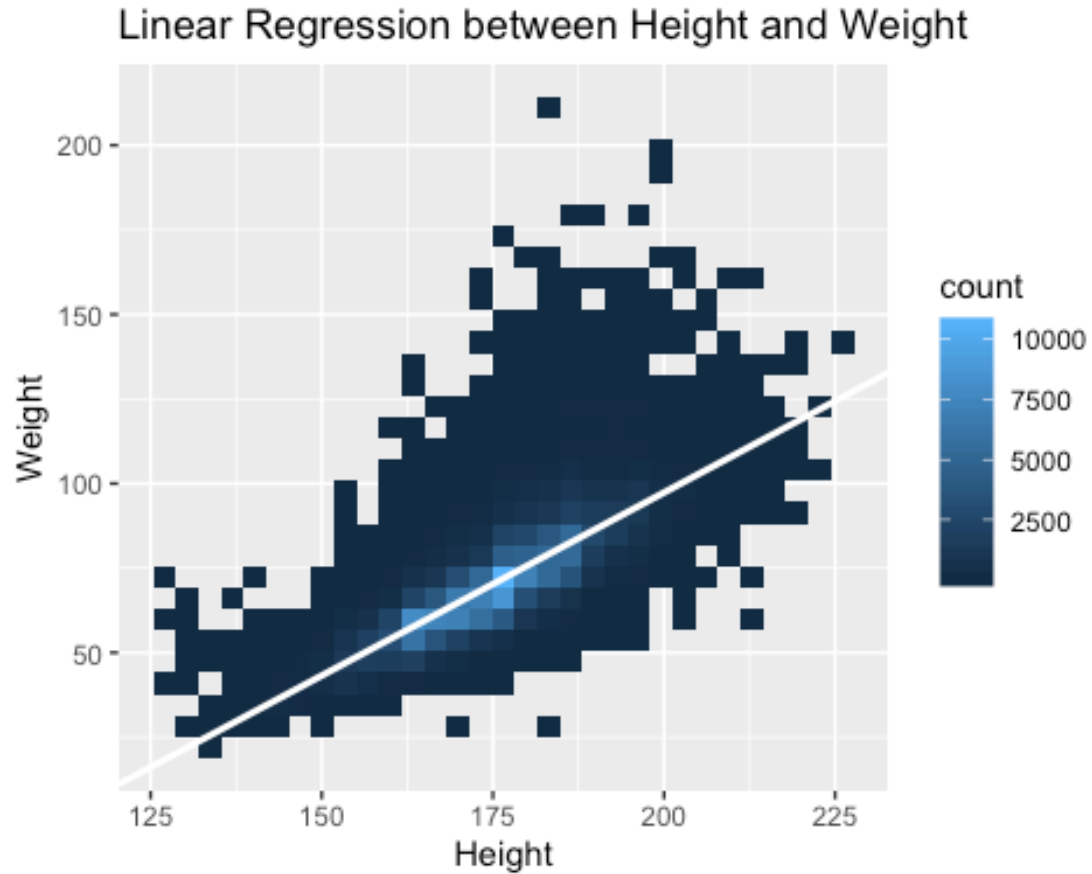


## 키와 몸무게 히스토그램



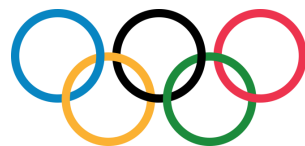


## 키와 몸무게 단일선형회귀 모형

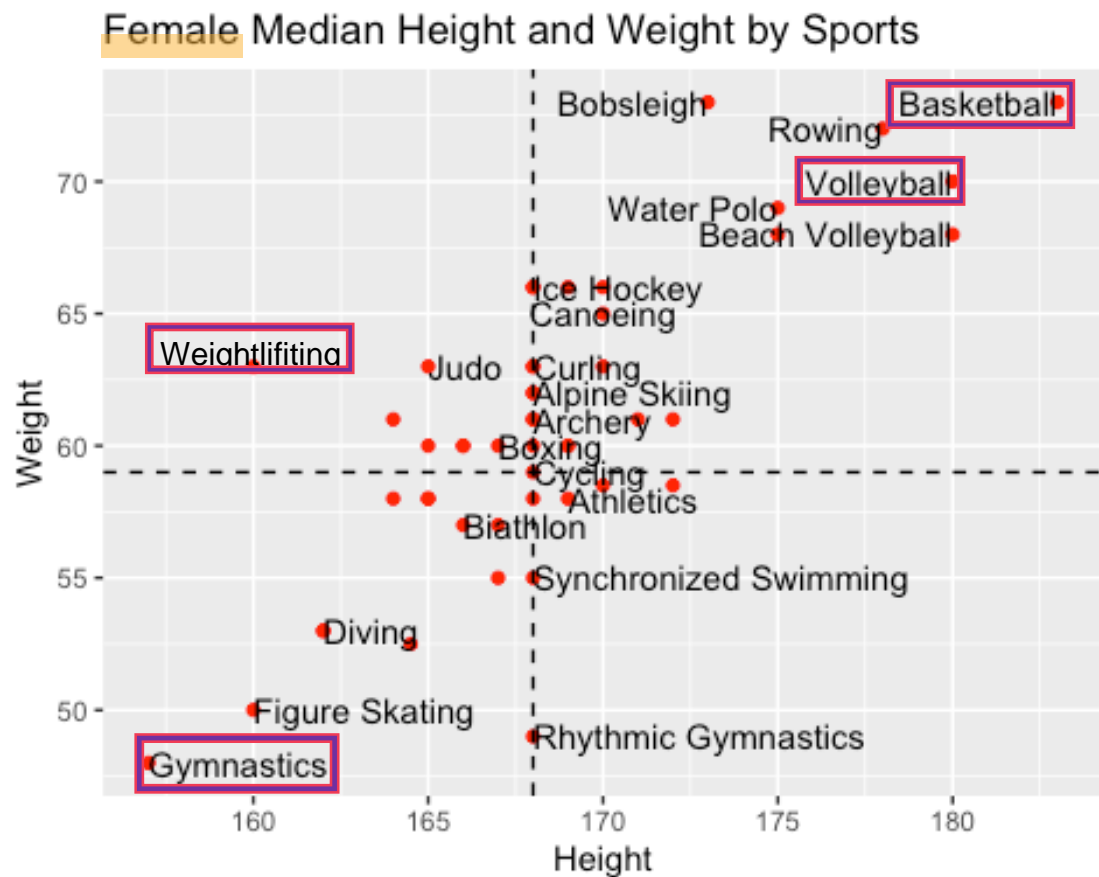
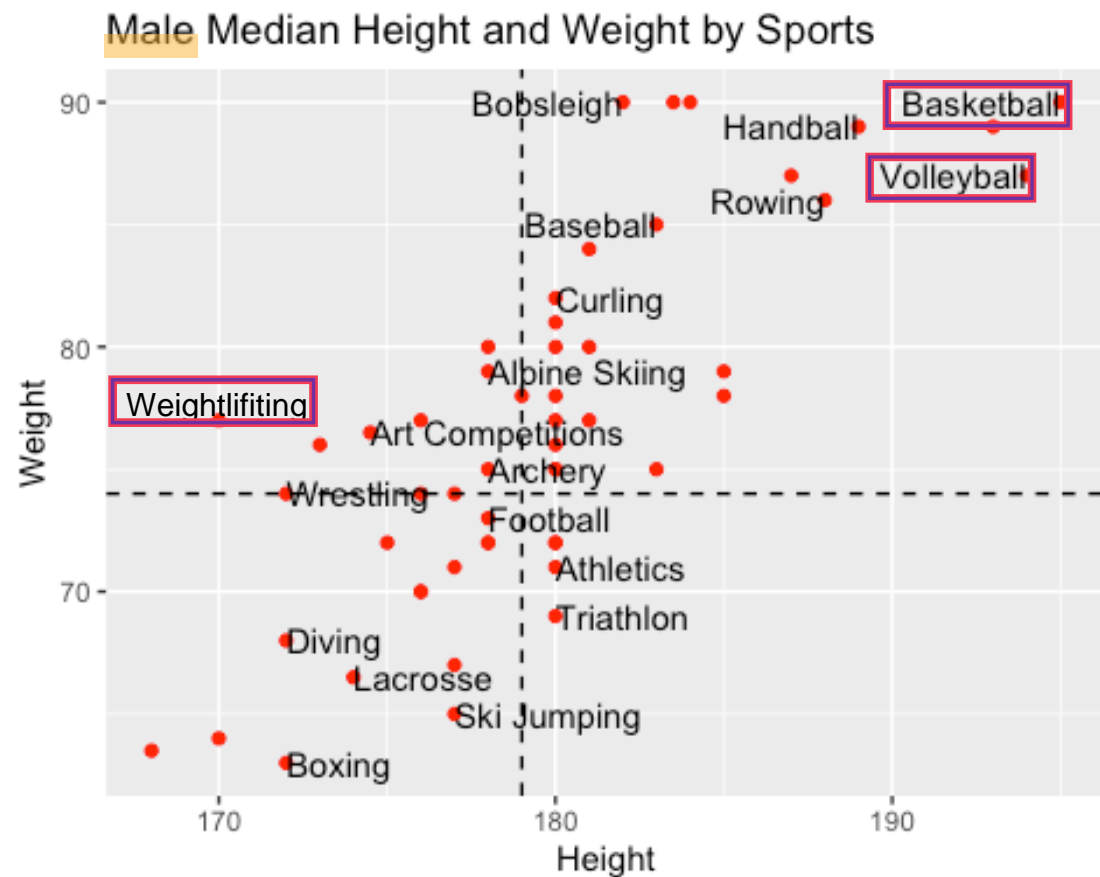


키와 몸무게 사이의 선형 회귀 직선을 보면  
두 변수 사이에 **양의 상관관계**가 존재함을 알 수 있다.

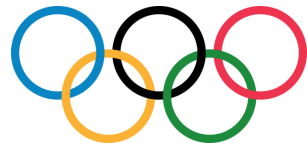
자료의 분포가 직선의 아래쪽보다 위쪽에 더 많은 것으로 보아  
**등분산이 아님**을 유추할 수 있다.



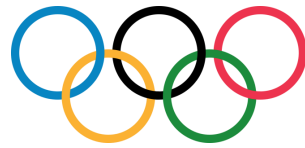
## 종목별 키와 몸무게 중간값 분포



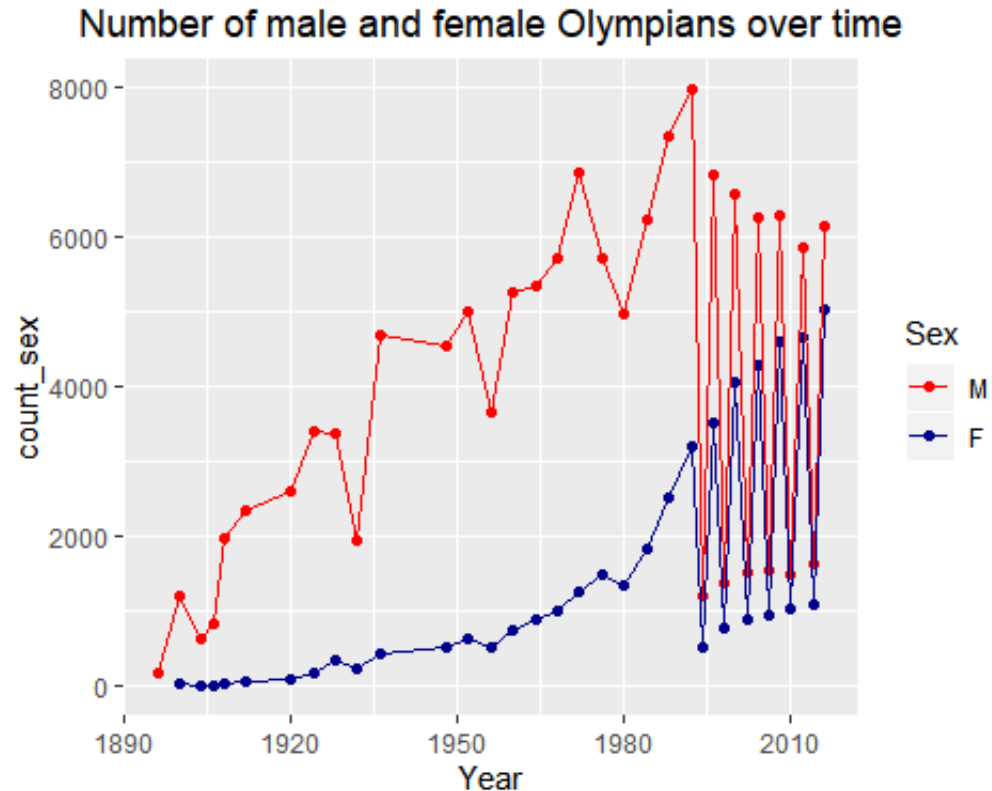




## 2. Gender



## 여성 선수 참가



남성과 여성 올림픽 참여 수를 비교하는 그래프를 그리면

### 1. 여성 참가 허용 시기

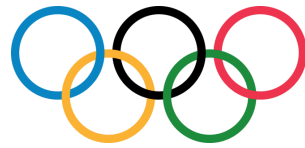
여성 선수 출전 허용 시기는 1900년 이후인 것을 확인할 수 있다.

(조사결과 실제로도 1900년 이후 여성 참가 허용)

### 2. 여성 선수 참가 수 또한 점차 늘고 있는 것을 확인할 수 있다.

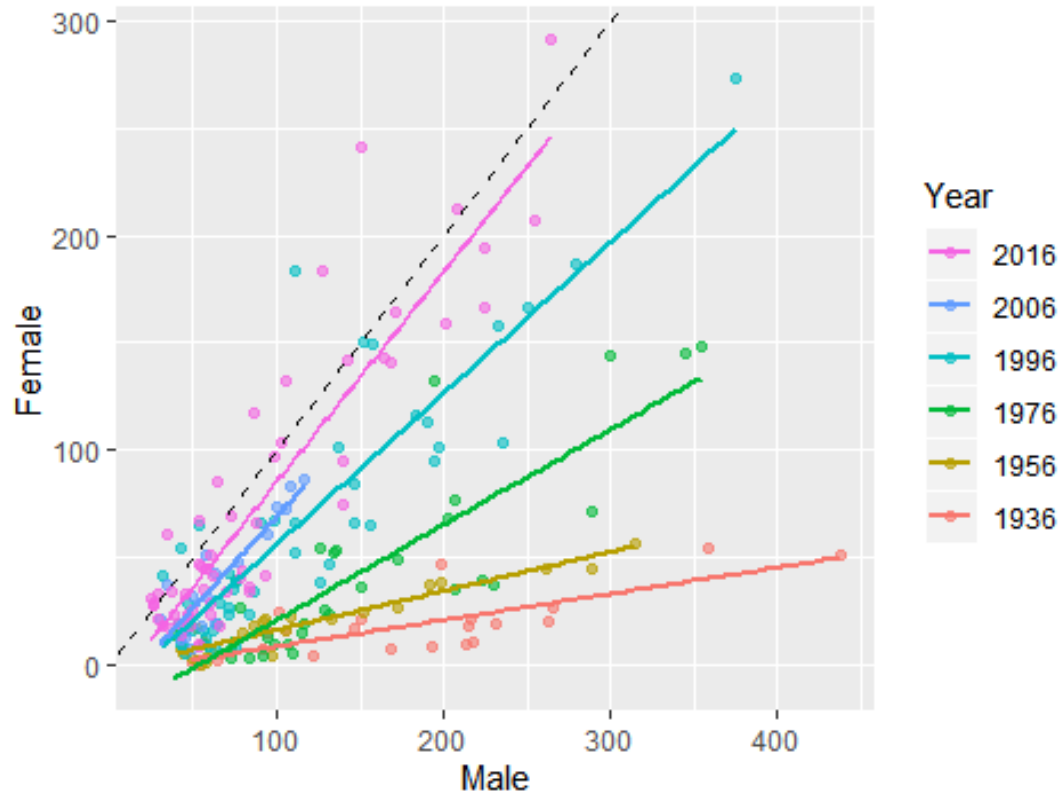
남성 선수 참가 수가 점차 줄고 있는 것 또한 확인된다.

여성 선수 참가 비율을 확인하기 위해 추가적으로 확인하여 보자.



## 여성 선수 참가

Female vs. Male Olympians from participating NOCs



10년을 주기로 데이터를 나눠

나라별 남녀의 출전 수에 대한 산점도와 함수를 확인해보았다.

여성의 참여율이 점차 높아지기 때문에

직선 그래프가 시간이 지날수록 기울기가 증가하는 것을 확인할 수 있다.

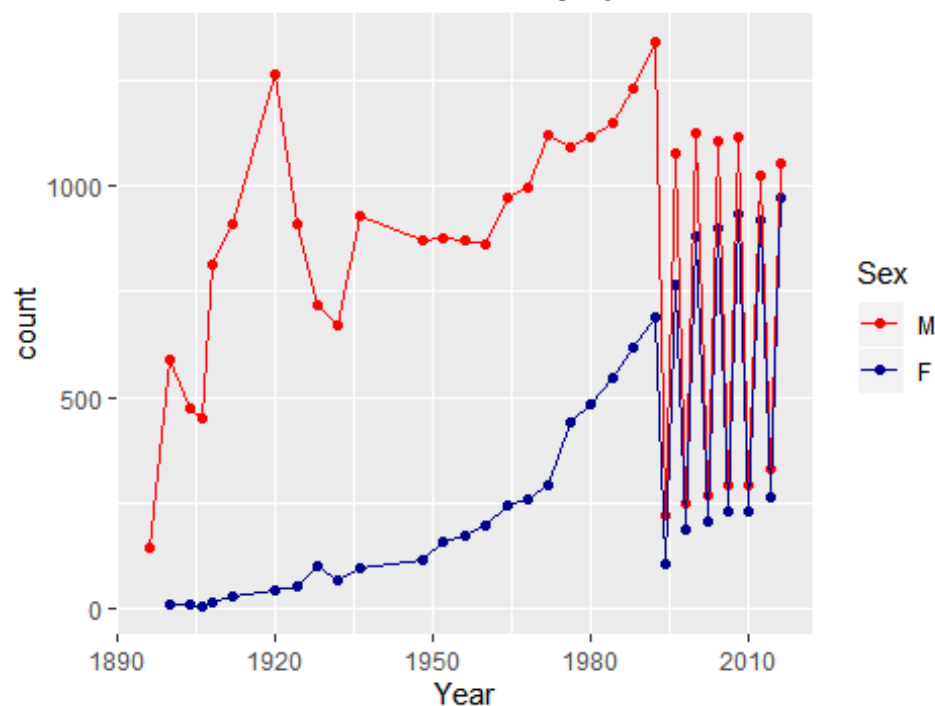
(1946년은 세계 2차 대전으로 인해 중단되었고,  
1966년, 1986년은 올림픽 게임 개최 시기가 현재와 달라  
게임이 개최되지 않았다.)



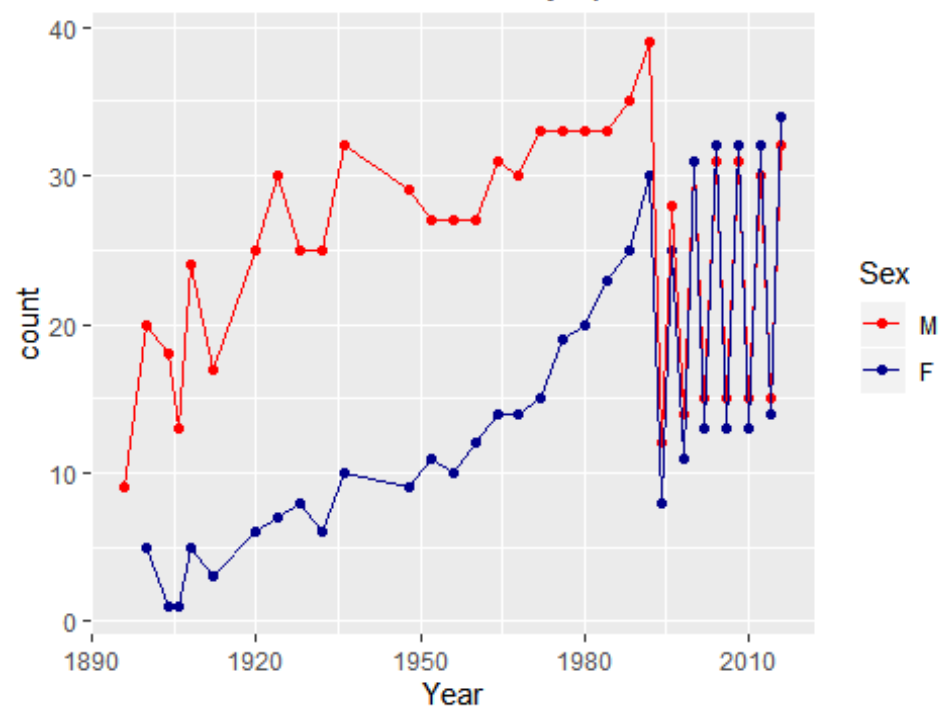
## 여성 선수 참가

여성 메달 수 또한 **올림픽 내 여성 종목 개설, 참가율이 높아짐에 따라** 더 많아지고 있다는 것을 확인할 수 있다.  
이를 확인하기 위해 종목 수 그래프와도 비교, 분석해보았다. 그 결과 대체적으로 유사한 패턴을 보이고 있음을 확인할 수 있었다.

Medals of male and female Olympians over time

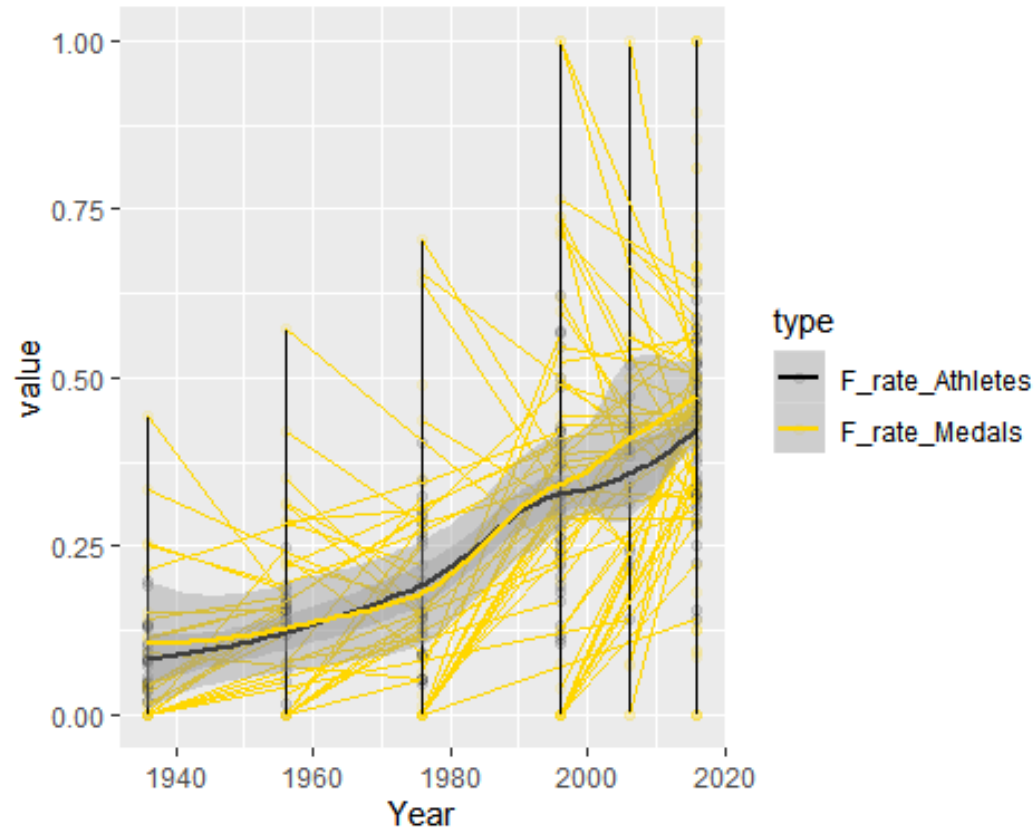


Games of male and female Olympians over time





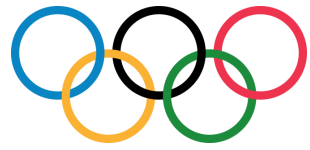
## 여성의 연도별 참가율과 메달 비율



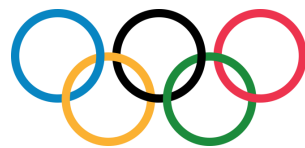
10년 주기로 연도별 여성 선수 참가율과 메달 획득율을 비교해보자.  
(1946년과 1966년, 1986년은 올림픽이 개최되지 않았다.)

1936년에는 전체 **18%**가량의 선수만이 여성 선수였다면,  
2016년에는 **42%**가량의 선수가 여성 선수로 참여율이 증가하였다.

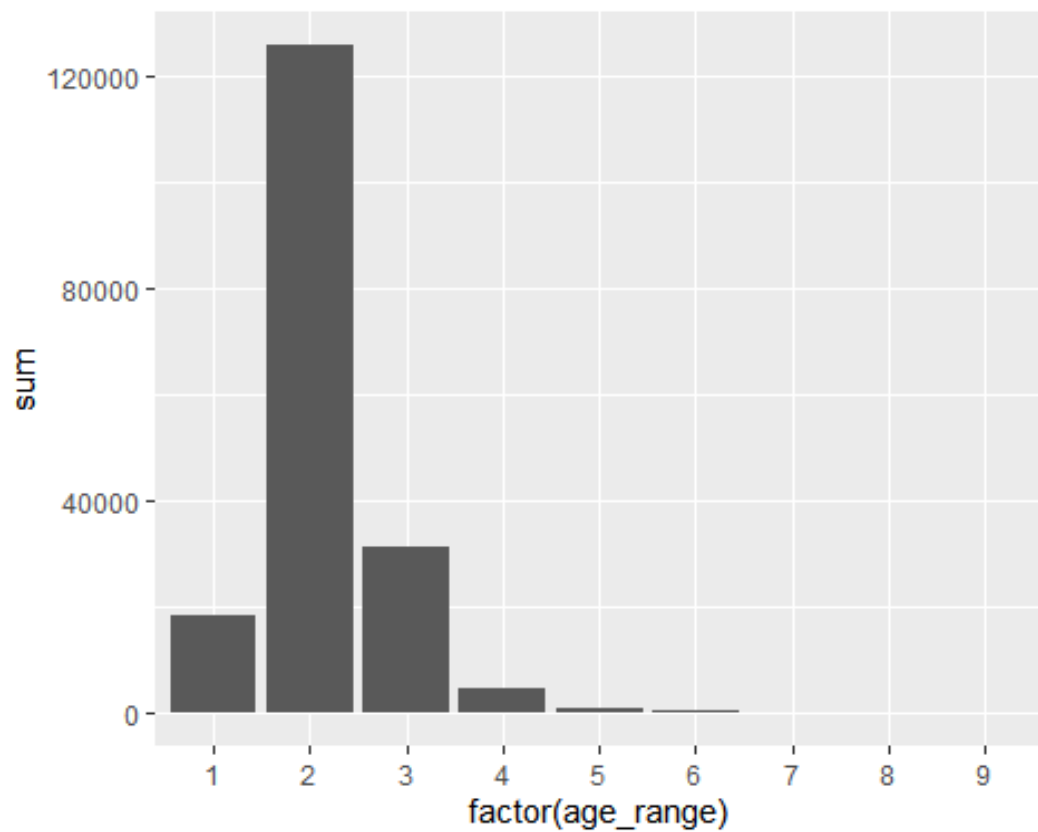
메달 획득 비율 또한 1936년 약 **19%** 에서  
2016년 약 **48%** 가량으로 크게 증가한 것을 확인할 수 있다.



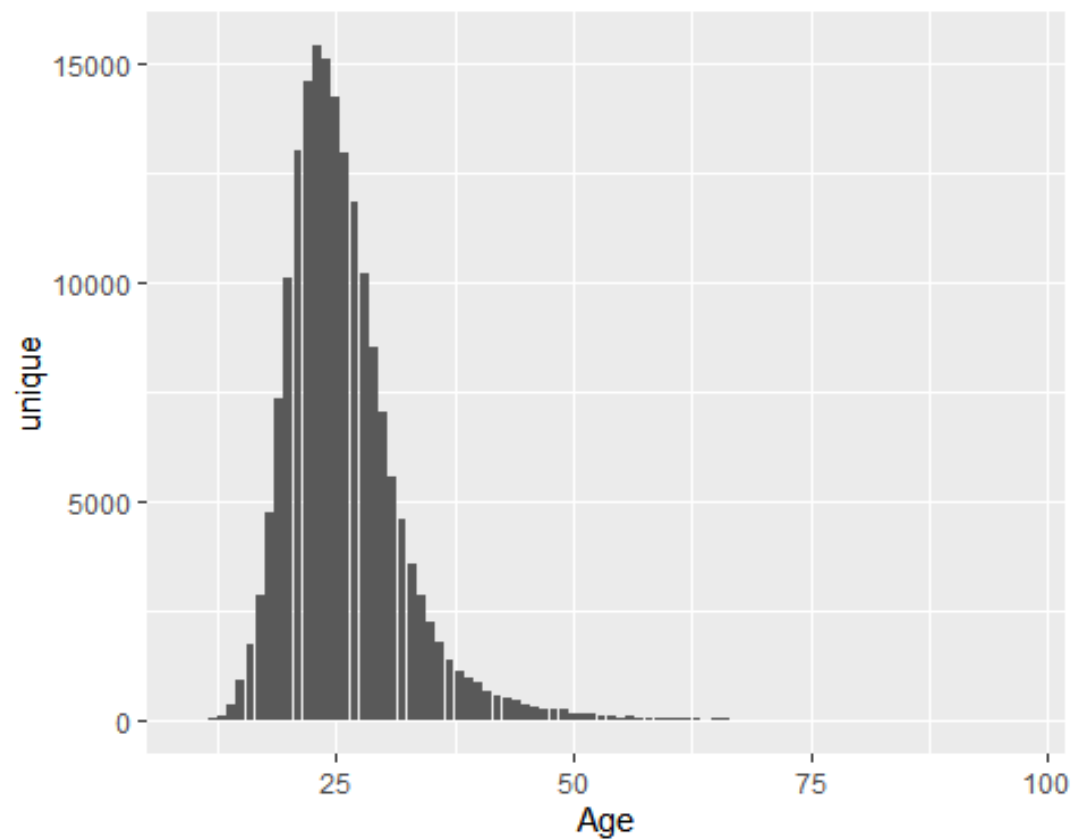
### 3. Age



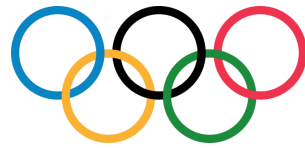
## 출전 선수들의 나이 분포



연령대 Distribution

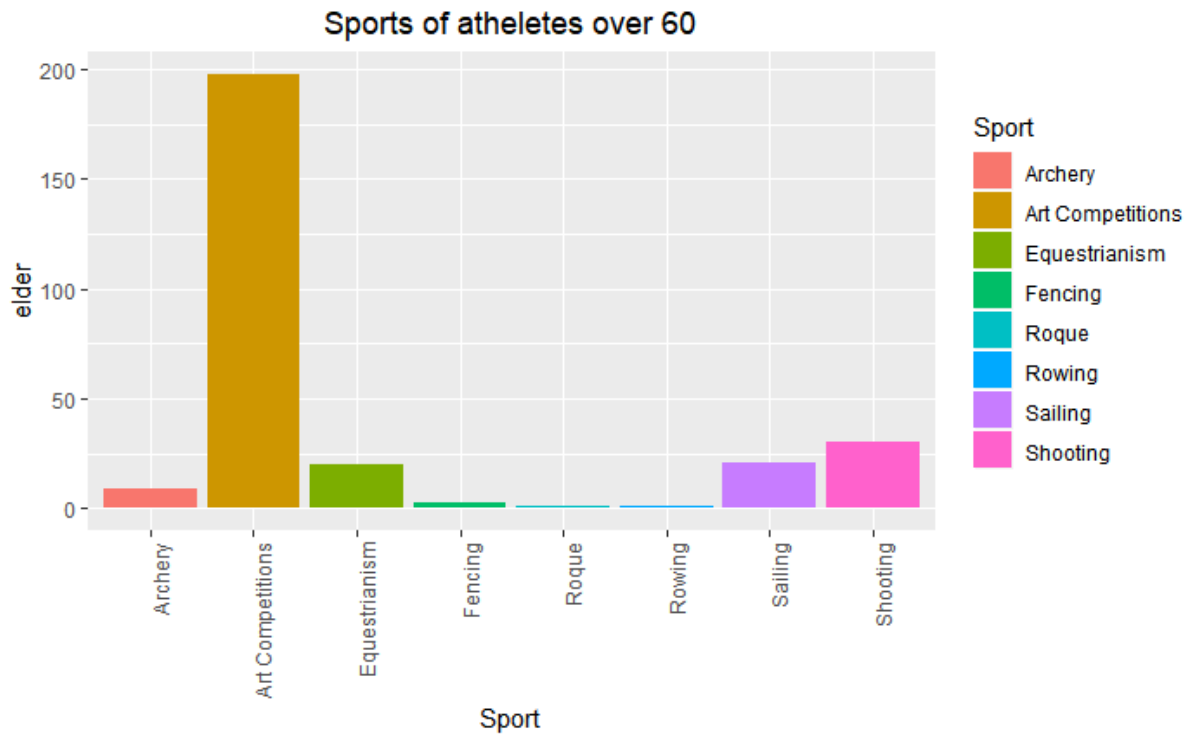


Age Distribution



## 고령 선수들의 종목

60세 이상의 출전 선수들의 종목에 대한 bar chart

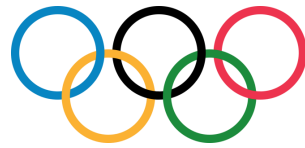


양궁, 미술, 승마, 펜싱, 로크, 조정, 세일링, 사격 (8종목)  
올림픽 종목이 총 66개임을 감안했을 때 많은 수는 아니다.

미술 > 사격 > 승마 순으로

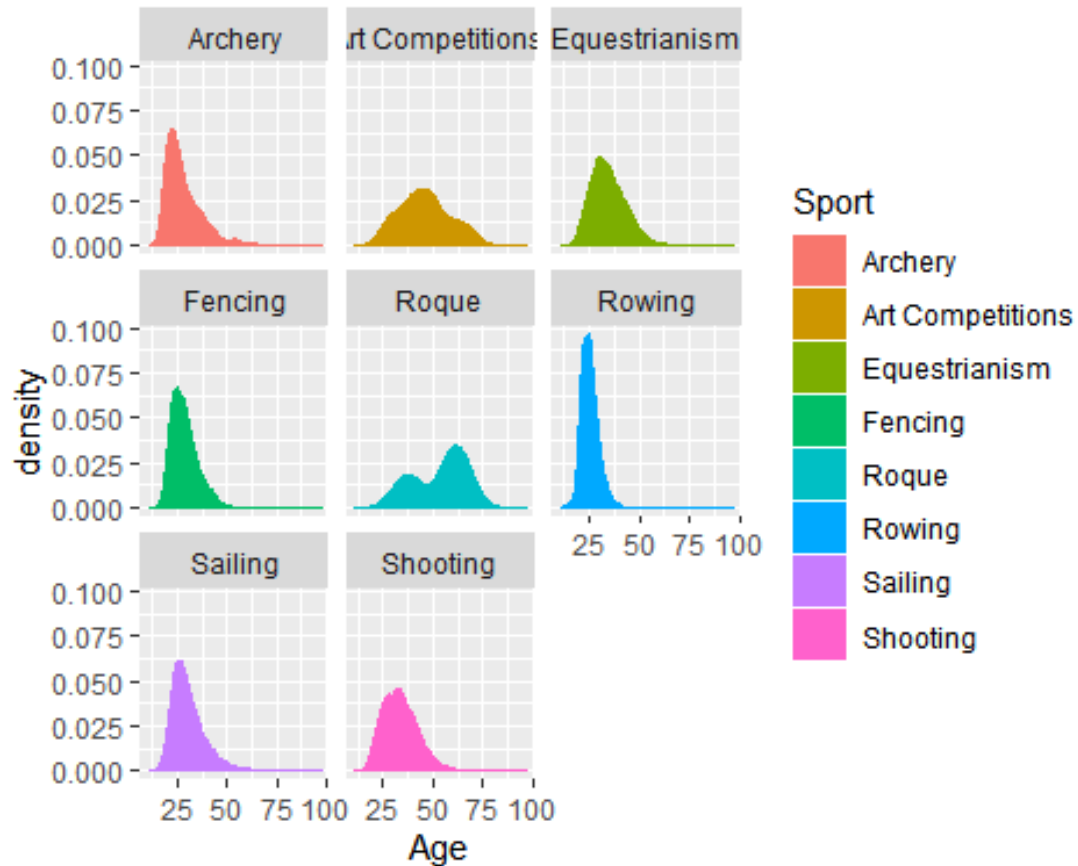
고령 선수들이 많이 참여하였음을 확인할 수 있다.





## 고령 선수들의 종목 (ctd.)

각 종목들의 Age에 대한 density plot

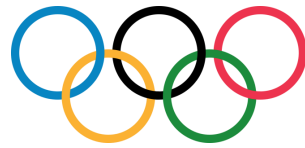


대체적으로 선수 연령대가

20대에 중점적으로 분포하는 것을 확인할 수 있다.

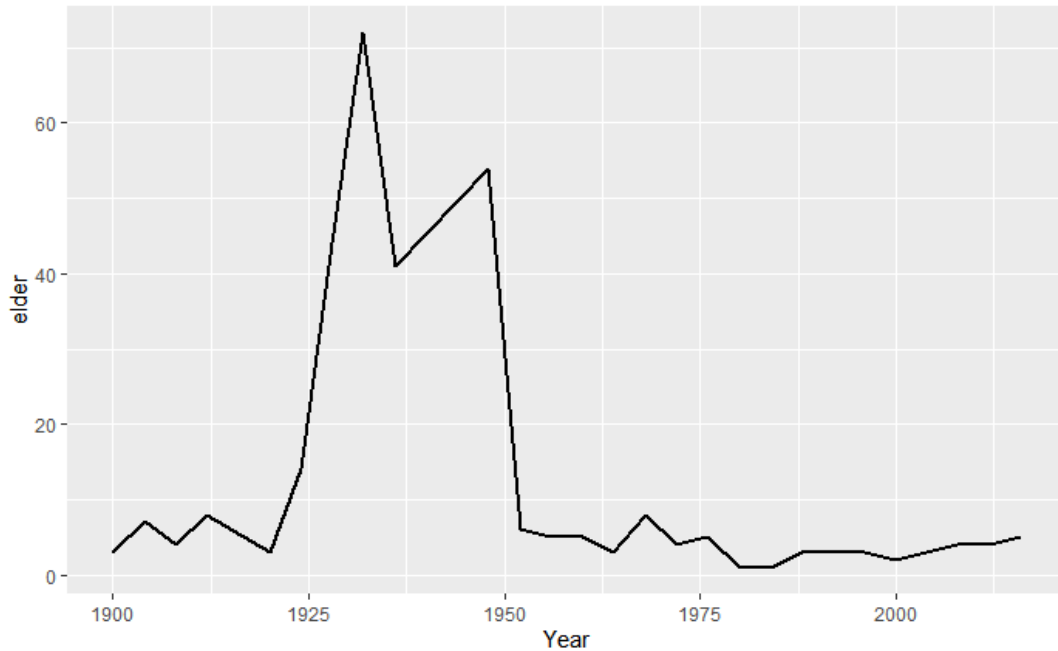
단, 로크, 미술의 경우,

출전 선수 연령대가 좀 더 다양한 분포를 보인다.



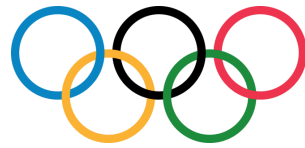
## 고령 선수 출전에 대한 Trend

60세 이상의 고령 선수들의 출전이 올림픽 전반적으로 나타나는 트렌드인가?



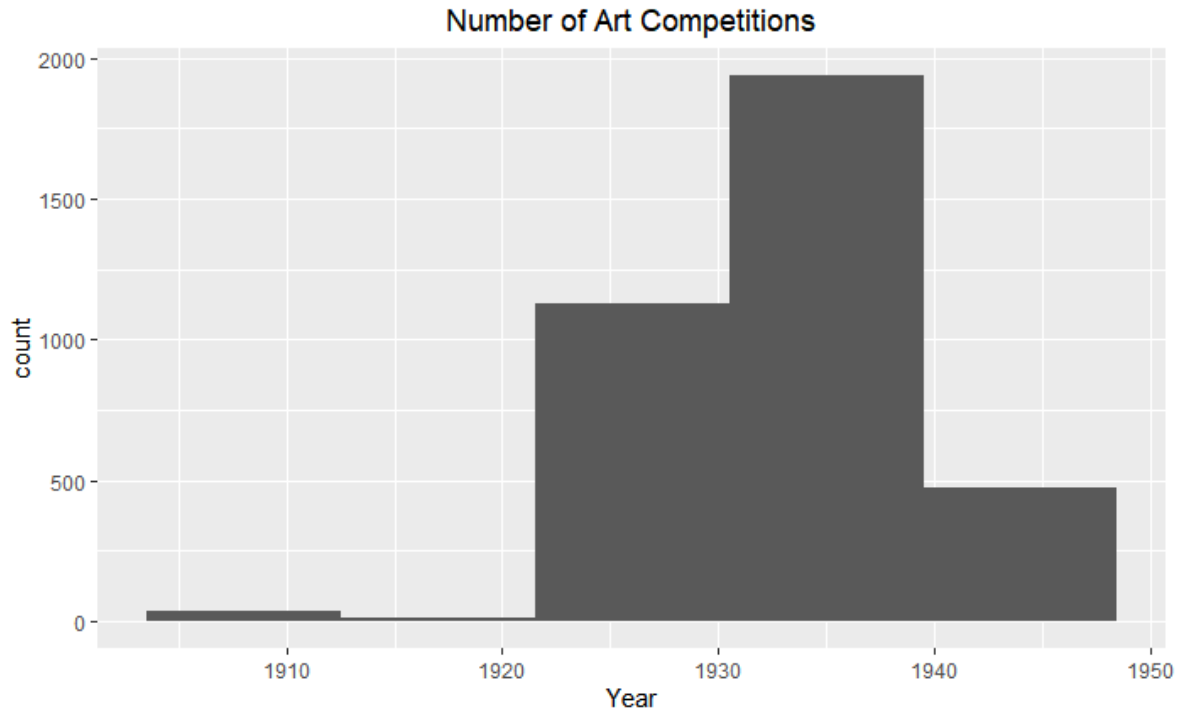
1925년에서 1950년 사이에 고령 선수의 출전이 많으나,  
그 전후에는 매우 적은 것을 확인할 수 있다.

아마도 출전 종목의 변화에 영향을 받았을 것임을 유추할 수 있다.  
고령 선수의 최다 출전 종목을 살펴보자.



## 고령 선수 출전에 대한 Trend

고령 선수들이 가장 많이 출전했던 미술 종목의 변화



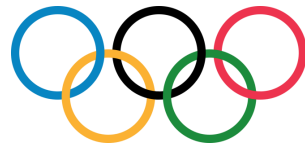
1920년과 1950년 사이에

미술 종목경기가 몰려 있음을 확인할 수 있다.

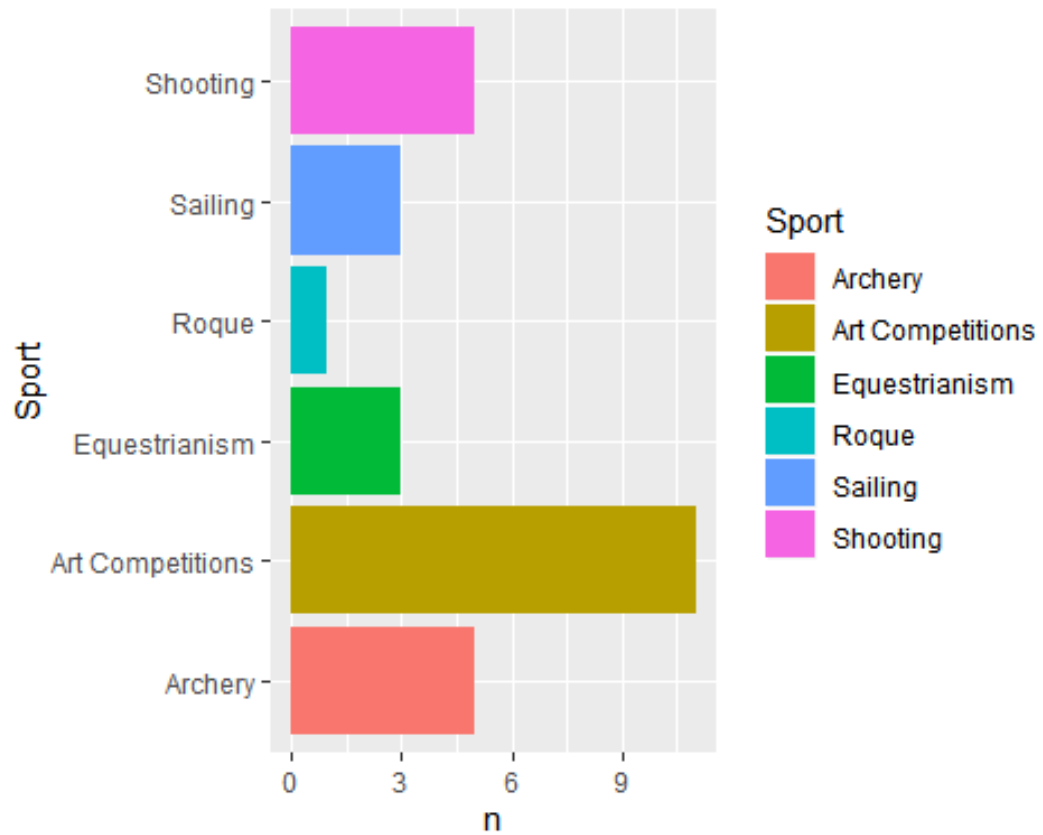
따라서 고령 선수 출전은

올림픽 전반적으로 나타나는 패턴이 아니라

특정 종목 (미술)의 영향이었음을 확인할 수 있다.



## 고령 선수들의 메달 획득 종목



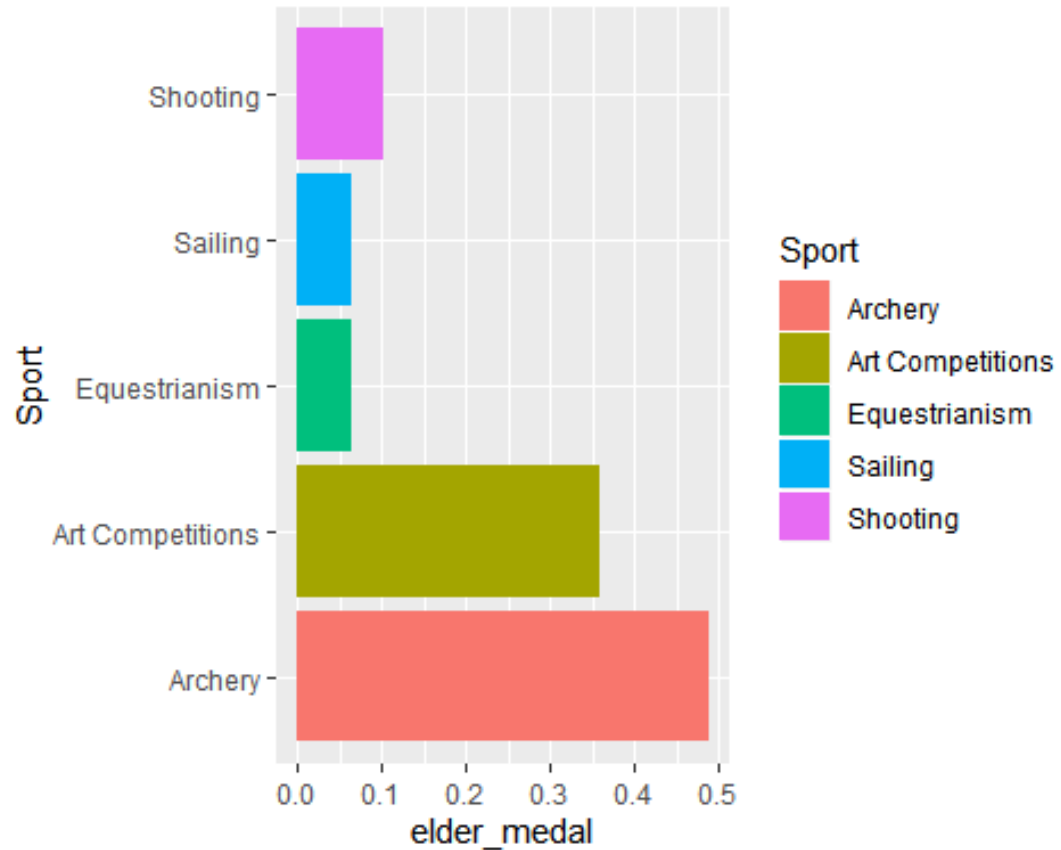
출전 종목 중 **조정과 펜싱을 제외**하고  
모든 종목에서 메달 획득한 사실을 확인할 수 있다.

**미술 > 양궁 > 사격**의 순서로  
많은 메달을 획득한 것을 확인할 수 있다.



## 고령 선수들의 메달 획득 비율

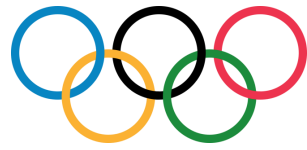
60세 이상 선수들이 메달을 획득한 경기 비율(% , 종목별)



(로크 종목의 경우 게임 수가 하나라 제외하였다.)

비율 자체가 높은 것은 아니나

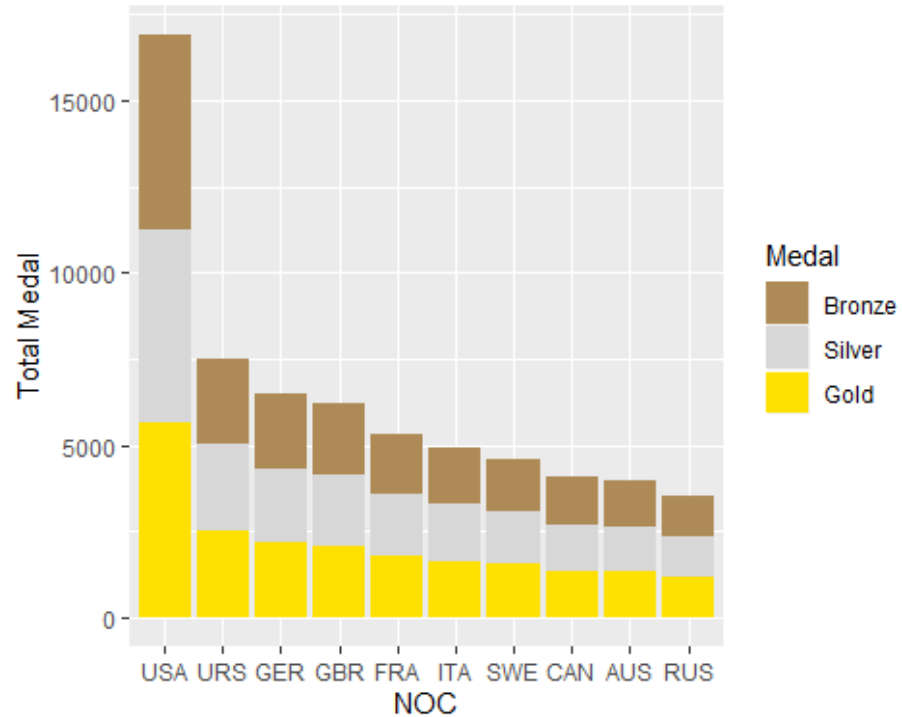
양궁, 미술 종목에서의 메달 획득율이 상대적으로 두드러짐.



## 4. Nations



## 메달 수와 GDP 변화량의 연관성



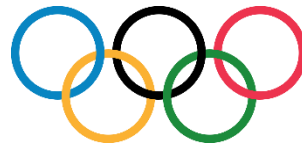
메달 수 많은 국가 top10 추출 후,  
국가 내 메달 비율을 나타낸 막대 그래프로 나타내었다.

메달 수 많은 국가는

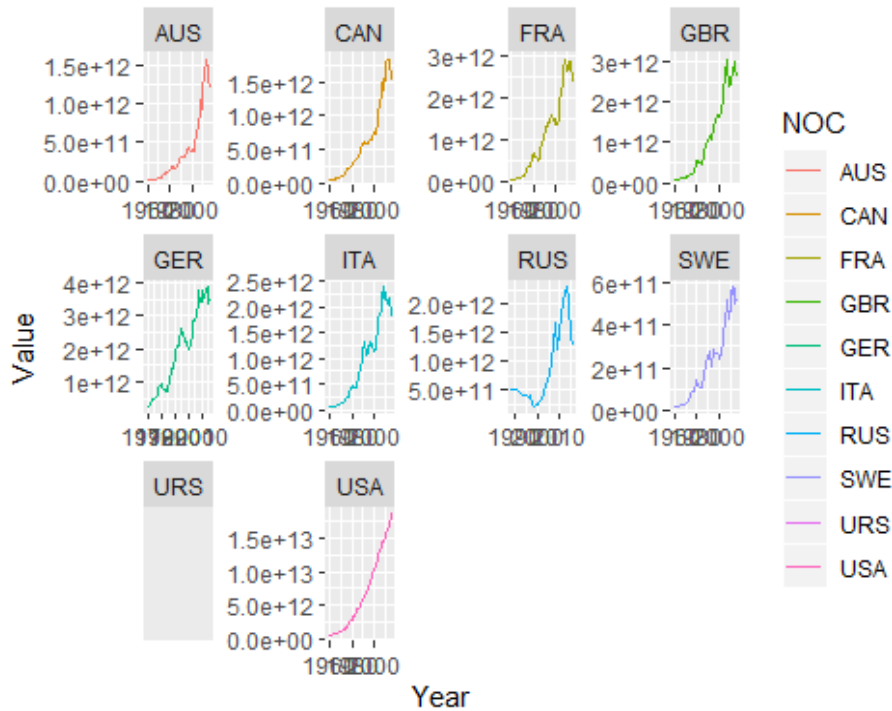
미국 > 소련 > 독일 > 영국 > 프랑스 > 이탈리아 > 스웨덴 >  
캐나다 > 오스트레일리아 > 러시아 순으로 나타났다.

국가 간 금, 은, 동 메달의 비율은 비슷한 것을 확인할 수 있다.

USA	URS	GER	GBR	FRA	ITA	SWE	CAN	AUS	RUS
미국	소련	독일	영국	프랑스	이탈리아	스웨덴	캐나다	오스트레일리아	러시아



## 메달 수와 GDP 변화량의 연관성



추출한 국가 별 GDP 변화 그래프를 확인하여 보자.

GDP 변화량이 큰 국가는

미국 > 독일 > 영국 > 프랑스 > 이탈리아 > 캐나다 >  
오스트레일리아 > 러시아 > 스웨덴 순으로 나타난다.

이는 앞선 메달 비율

(미국 > 소련 > 독일 > 영국 > 프랑스 > 이탈리아 > 스웨덴 >  
캐나다 > 오스트레일리아 > 러시아) 과도 유사한 순서이다.

USA	URS	GER	GBR	FRA	ITA	SWE	CAN	AUS	RUS
미국	소련	독일	영국	프랑스	이탈리아	스웨덴	캐나다	오스트레일리아	러시아

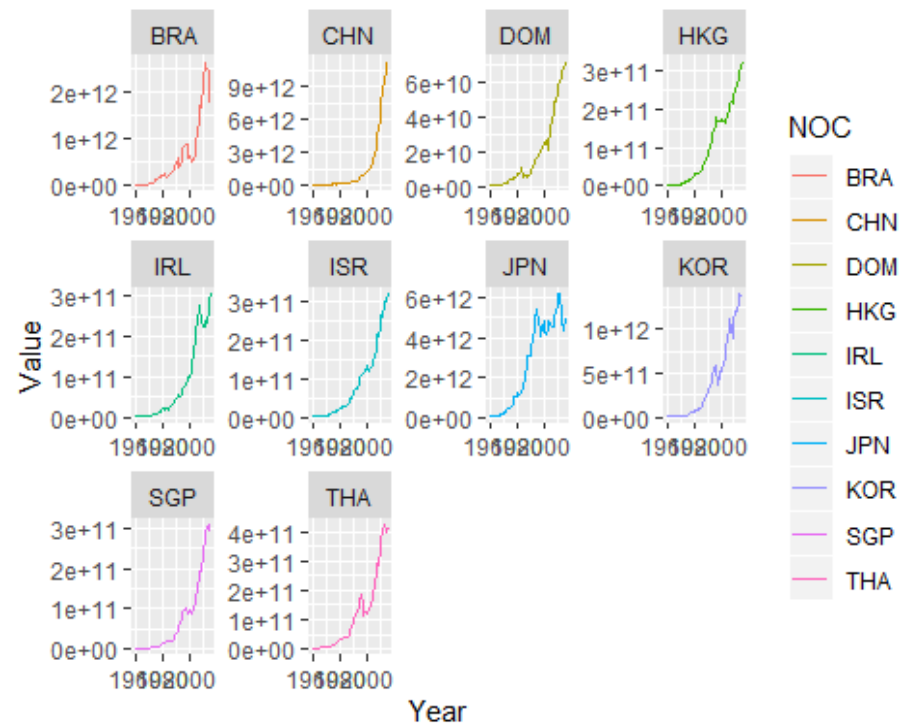
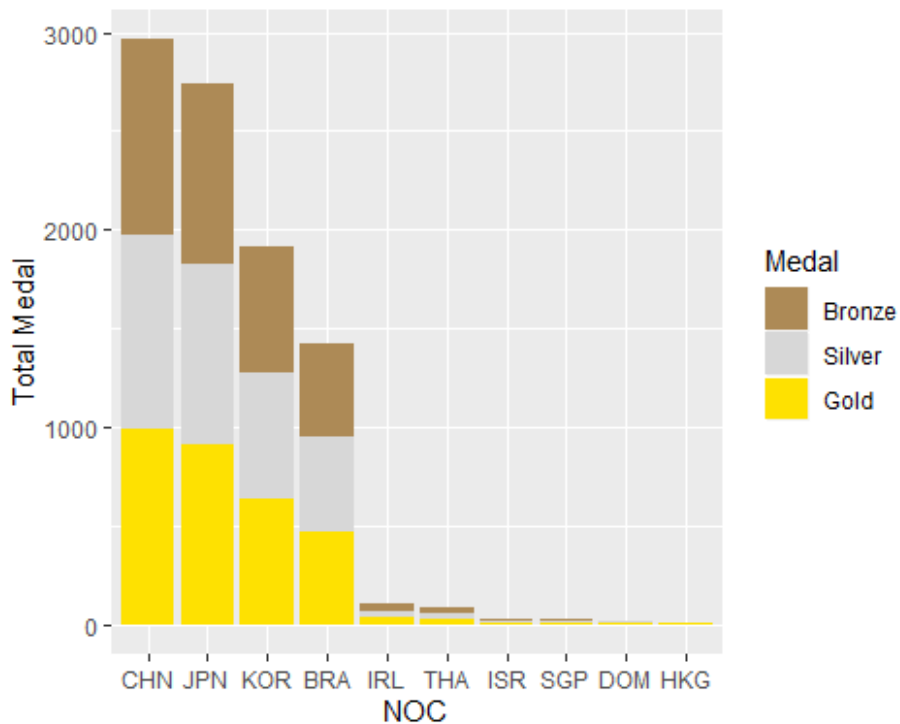
즉, 메달 수가 큰 국가는 GDP 변화량도 크다.





## 메달 수의 변화량과 GDP 변화량의 연관성

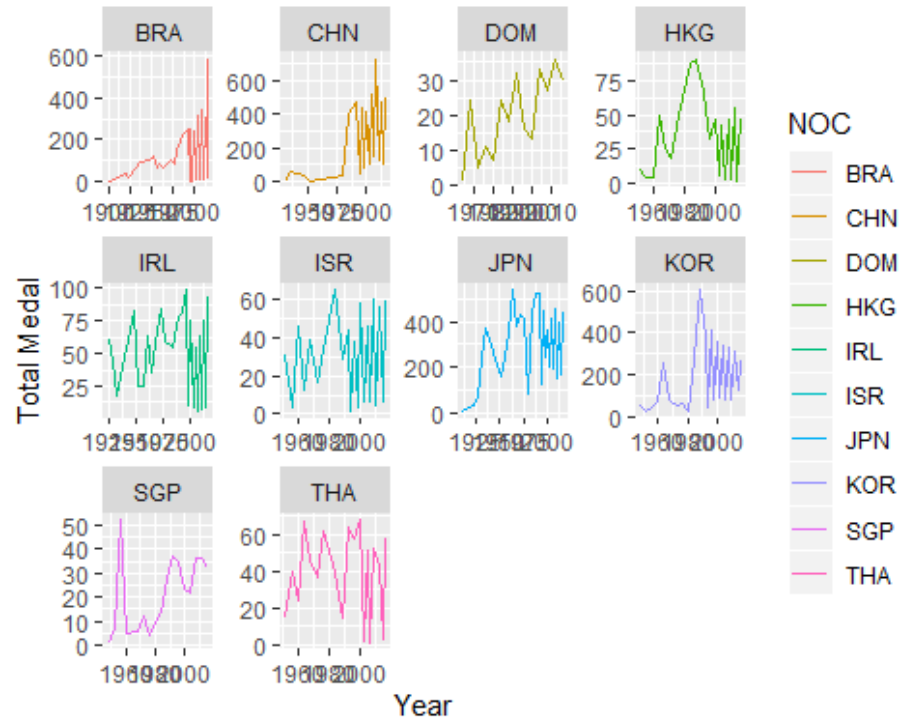
GDP 변화가 큰 10개 국가의 메달 비율을 나타낸 막대 그래프와 국가 별 GDP 변화 그래프



CHN	JPN	KOR	BRA	IRL	THA	ISR	SGP	DOM	HKG
중국	일본	한국	브라질	아일랜드	태국	이스라엘	싱가포르	도미니카	홍콩



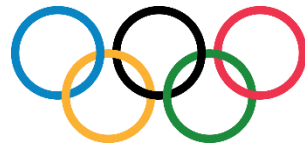
## 메달 수의 변화량과 GDP 변화량의 연관성



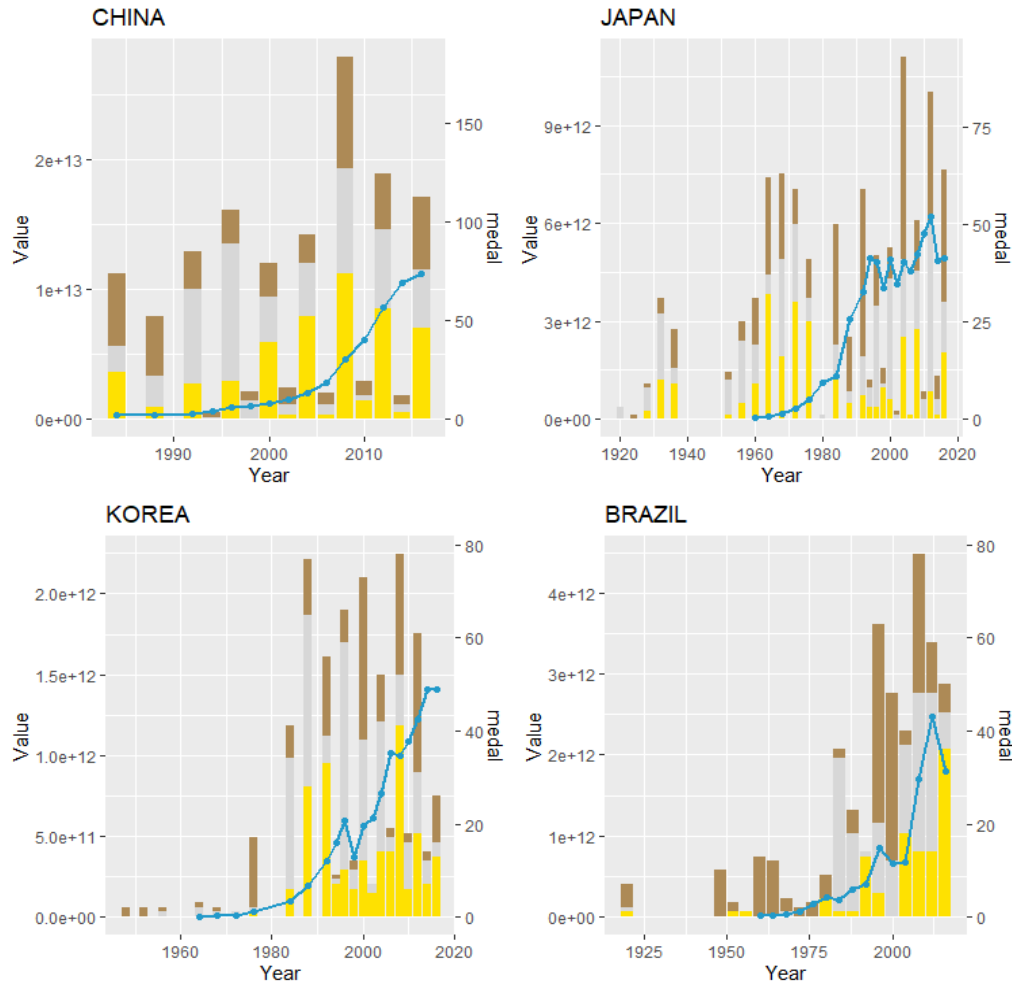
추출한 국가의 메달 수 변화 그래프를 살펴보면,  
모든 국가에서 일부 피크를 제외하고  
전반적으로 **메달 수가 증가하는 추세**를 보인다.

여기에서 **브라질, 중국, 일본, 한국의** 경우가 특히 더 유의미한 결과를  
얻을 것이라 보여 더 자세한 그림을 그려보았다.

CHN	JPN	KOR	BRA	IRL	THA	ISR	SGP	DOM	HKG
중국	일본	한국	브라질	아일랜드	태국	이스라엘	싱가포르	도미니카	홍콩



## 메달 수의 변화량과 GDP 변화량의 연관성



네 국가 (중국, 일본, 한국, 브라질)

모두 전반적으로 메달 수와 GDP가 증가한다.

즉, GDP의 변화가 큰 국가는 메달 수의 변화도 크다.

따라서 올림픽 메달 수와 GDP 사이에는 연관이 있다고 할 수 있다.

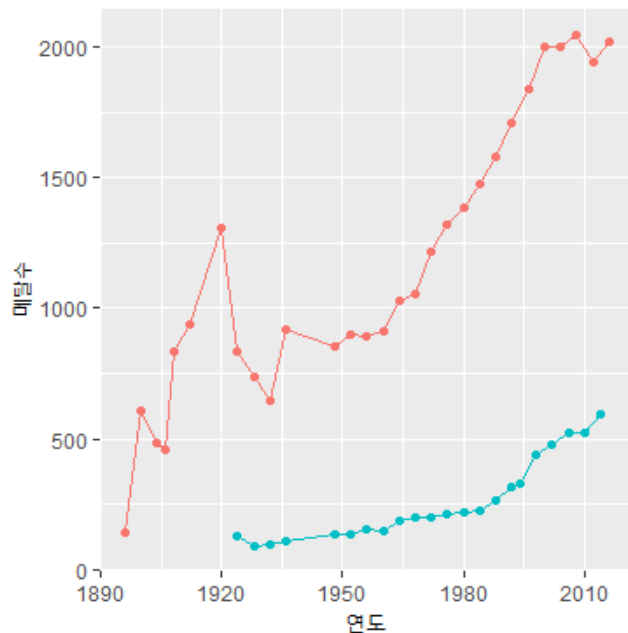


## 5. Season

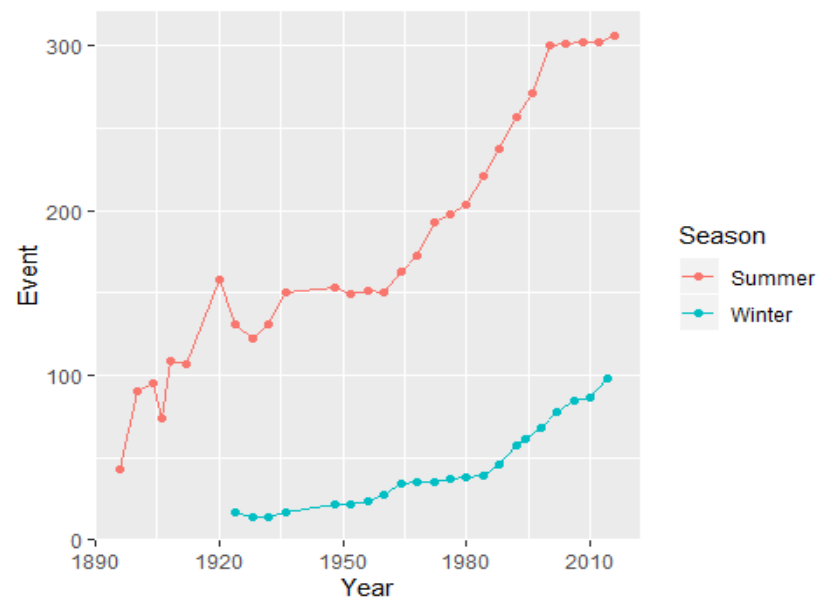


## 동계 올림픽 vs 하계 올림픽

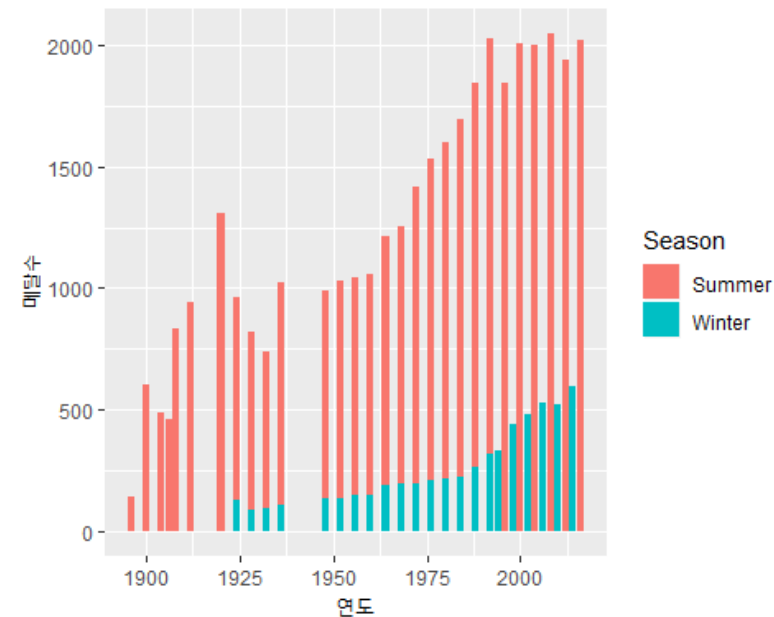
종목 수



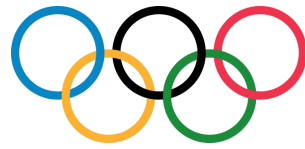
세부 종목 수



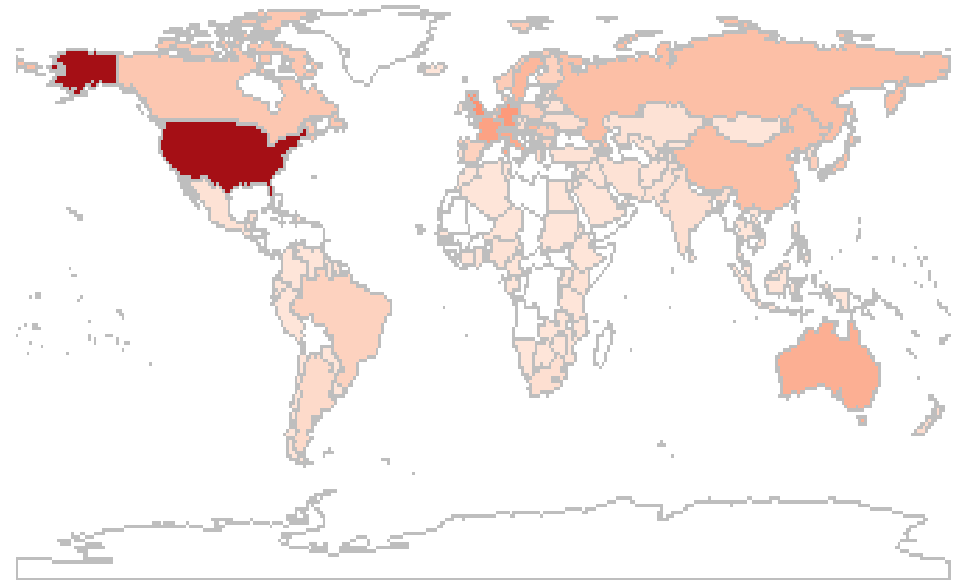
메달 수



하계올림픽이 동계올림픽보다 메달 수, 종목 수, 세부 종목 수가 월등히 많다.  
따라서 하계올림픽이 동계올림픽보다 더 규모가 크고 인기있다고 볼 수 있다.

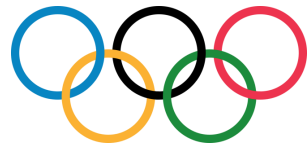


## 하계 올림픽 메달 수와 나라 기후 간의 상관관계

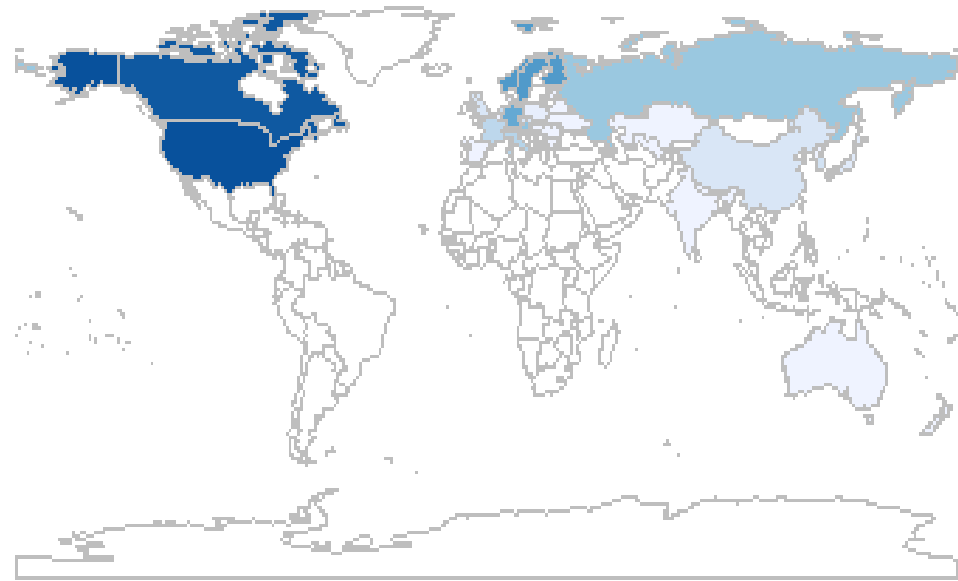


미국 > 소련 > 독일 > 프랑스 > 이탈리아 > 호주 > 헝가리 순으로 메달 수가 높았다.

하계올림픽에서 더운 나라가 메달을 많이 딸 것이라는 가설에서 벗어난 결과를 보였다.

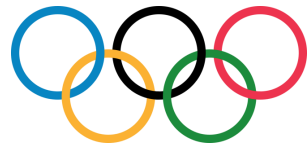


## 동계 올림픽 메달 수와 나라 기후 간의 상관관계



반면, 미국 > 캐나다 > 노르웨이 > 소련 > 스웨덴 > 핀란드 > 독일 > 호주 > 스위스 순으로 메달 수가 높았다.

동계올림픽에서 추운 나라가 메달을 많이 딸 것이라는 가설이 증명되었다.



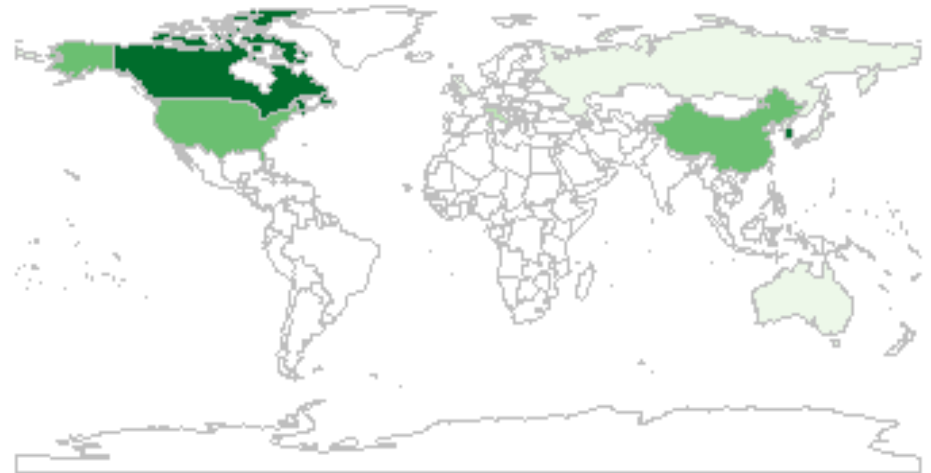
## 동계 올림픽 메달 수와 나라 기후 간의 상관관계 종목별 비교

스키종목(설상종목)



노르웨이 > 스웨덴 > 핀란드

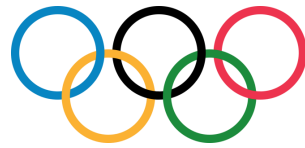
쇼트트랙 종목(실내종목)



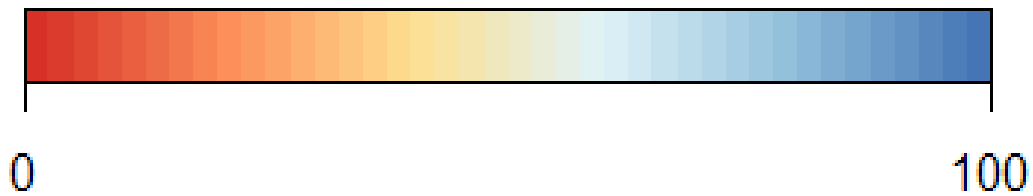
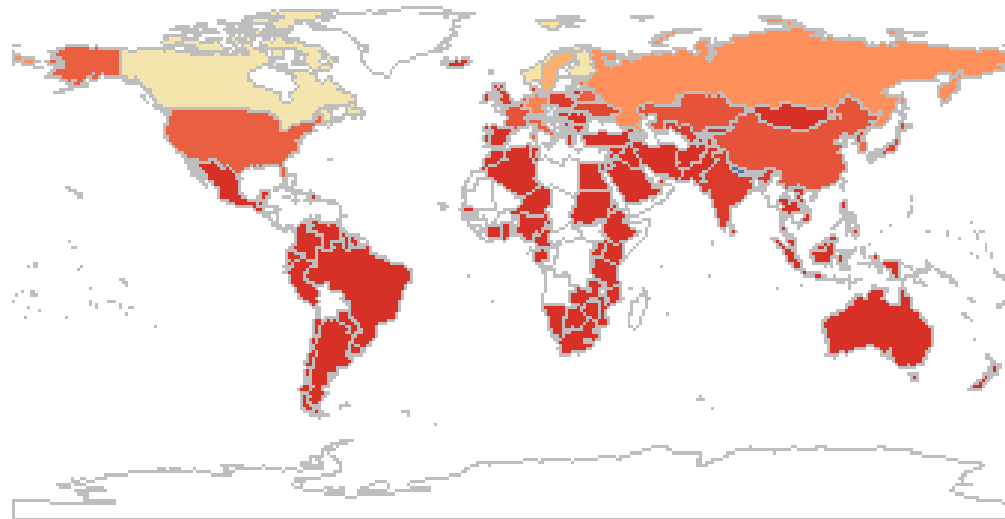
한국, 중국 등 동아시아 국가

동계올림픽 종목 내에서도 설상종목/실내종목 간 차이가 있다.





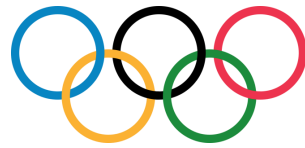
## 동/하계 메달 비율과 나라의 기후 간의 상관관계



더운 지역(저위도 지역, 아프리카 등)의 국가는  
하계올림픽 메달 수보다 동계올림픽 메달 수가 많다.

반면, 추운지역 (고위도 지역, 히말라야 근처 산지 등)의 국가는  
동계 올림픽 메달 비율이 다른 지역에 비해 높다.

다만, 하계올림픽 메달수가 많기 때문에  
0 ~ 50 사이의 값을 보이는 국가가 많은 것을 확인할 수 있다.



## 결론

### 1. Height & Weight

: 종목에 따라 출전선수들의 전반적인 키와 몸무게가 상이하다

### 2. Gender

: 1900년 여성 출전 허가 이후 점차 여성 출전 비율과 메달 획득율이 높아지고 있다.

### 3. Age

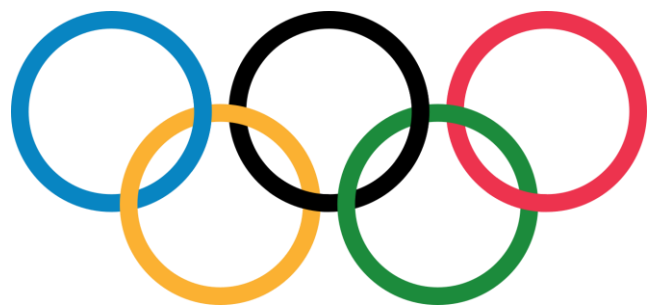
: 출전 선수들의 나이는 20대 위주였으나, 97세 등의 고령 선수들의 존재도 돋보임. 특히 미술, 양궁 등에서 성과를 보인다.

### 4. Nations

: GDP와 메달 수는 어느정도 연관성이 있다.

### 5. Season

: 하계 올림픽은 동계 올림픽과 달리 나라 기후와 상관 관계가 적다.



감사합니다