**팀 07**

이상명: 데이터 분석 및 시각화 담당, 보고서 작성, 상관관계 분석, PPT 제작 및 13주차 발표

김성수: 데이터 분석 및 시각화 담당, 보고서 작성, 회귀분석

정상철: 데이터 분석 및 시각화 담당, 보고서 작성, 회귀분석

정현수: 데이터 분석 및 시각화 담당, 보고서 작성, 상관관계 분석 ,PPT 제작 및 14주차 발표

**제목 : 축구 국가별 포지션 분석을 통한 상위 팀 하위 팀 비교 (남녀구분)**

# 1 분석 주제와 문제 설정

제공받은 데이터는 남자 2018년, 여자 2019년의 월드컵 자료로 결과 순위에 대한 외부 데이터를 가져와 활용하였다. 그러면 ‘우승팀의 전략은 무엇일까?’라는 의문점을 가지게 되었고, Keyword를 ‘포지션’으로 설정하고 브레인스토밍을 진행했다.

첫 번째 질문으로 포지션별 나이대이다. 상위 팀 포지션별 나이대와 하위 팀 포지션별 나이대를 분석해서 경험은 부족하지만 젊은 20대를 많이 기용했는지, 나이는 많지만, 경험이 많은 30대를 많이 기용했는지를 비교하여 ‘공격수는 젊은 20대보다 경험이 많은 30대를 쓰는 것이 더 좋은 결과를 내지 않을까?’를 알아볼 수 있어 흥미롭다.

두 번째 질문으로 포지션별 선수의 수이다. 상위 팀 포지션별 선수의 수와 하위 팀 포지션별 선수의 수를 분석해서 ‘남자 우승팀인 프랑스는 공격수를 많이 기용해서 우승하지 않았을까?’를 파악할 수 있어 흥미롭다.

세 번째 질문으로 포지션별 선수의 출전 횟수이다. 상위 팀 포지션별 선수의 출전 횟수와 하위 팀 포지션별 선수의 출전 횟수를 분석해서 ‘상위 팀이 된 이유 중 하나가 출전 횟수가 많은 선수를 많이 기용해서이지 않을까?’를 확인할 수 있어 흥미롭다.

네 번째 질문으로 포지션별 골 수이다. 상위 팀 포지션별 골 수와 하위 팀 포지션별 골 수를 분석해서 ‘상위 팀이 하위 팀보다 골 수가 많아서 상위 팀이 되었는지’, ‘포지션별 골 수가 우승에 어떠한 영향을 미쳤는지’에 대한 상관관계를 분석할 수 있다. 또한 위에서 분석한 포지션별 출전 횟수와 포지션별 골 수를 관련지어 생각해 볼 수 있어 흥미롭다.

이러한 질문에 대한 답으로 상위 팀과 하위 팀의 차이를 비교해보고, 비교분석과 상관관계 분석을 통해 답을 찾을 계획이다. 마지막에는 남녀까지도 비교할 계획이다.

# 2 탐색적 분석

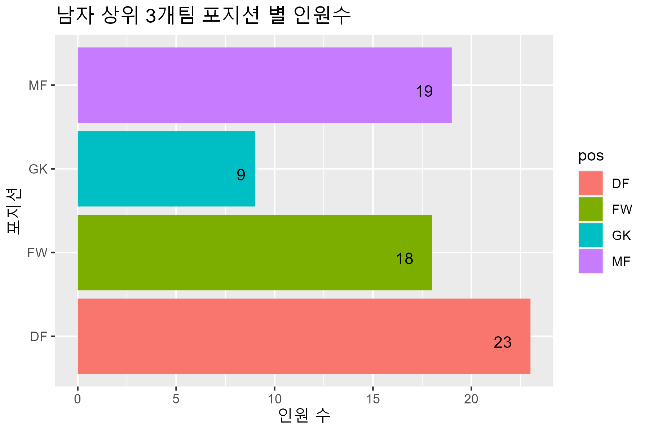
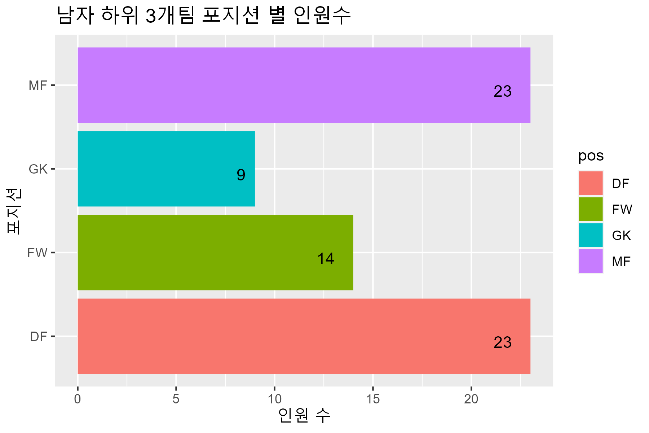
위에서 언급한 아이디어들은 범주형 데이터로 통계의 양을 한눈에 알아보기 쉽게 하기 위해 주로 Barplot과 Boxplot을 이용하여 데이터들을 시각화 하였다. 또한 문제에 대한 답을 찾기 위해 남녀구분, 상위 팀 3개, 하위 팀 3개로 나눠 비교하였다. [남자]

상위 팀 3개=프랑스 크로아티아 벨기에 / 하위 팀 3개=자메이카 이집트 아이슬란드

2-1) 포지션별 나이대

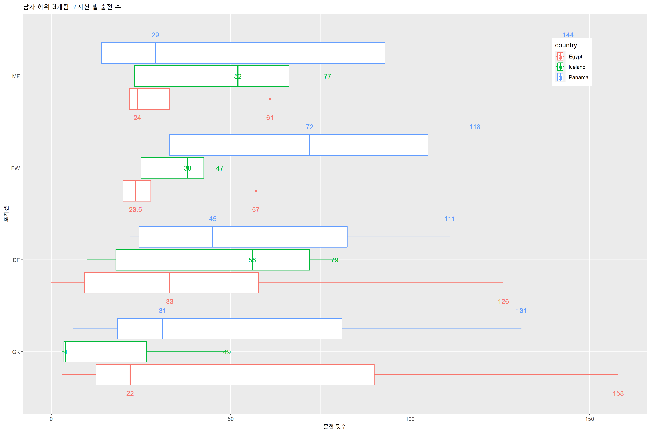
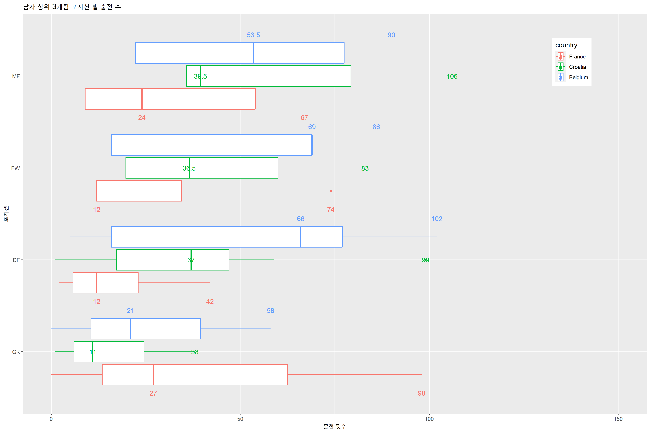
분석 결과 유의미한 결과를 내지 못했다.

2-2) 포지션별 선수의 수

 역시 통계의 양을 한눈에 알아보기 쉽게 하기 위해서 Barplot을 사용하였다.

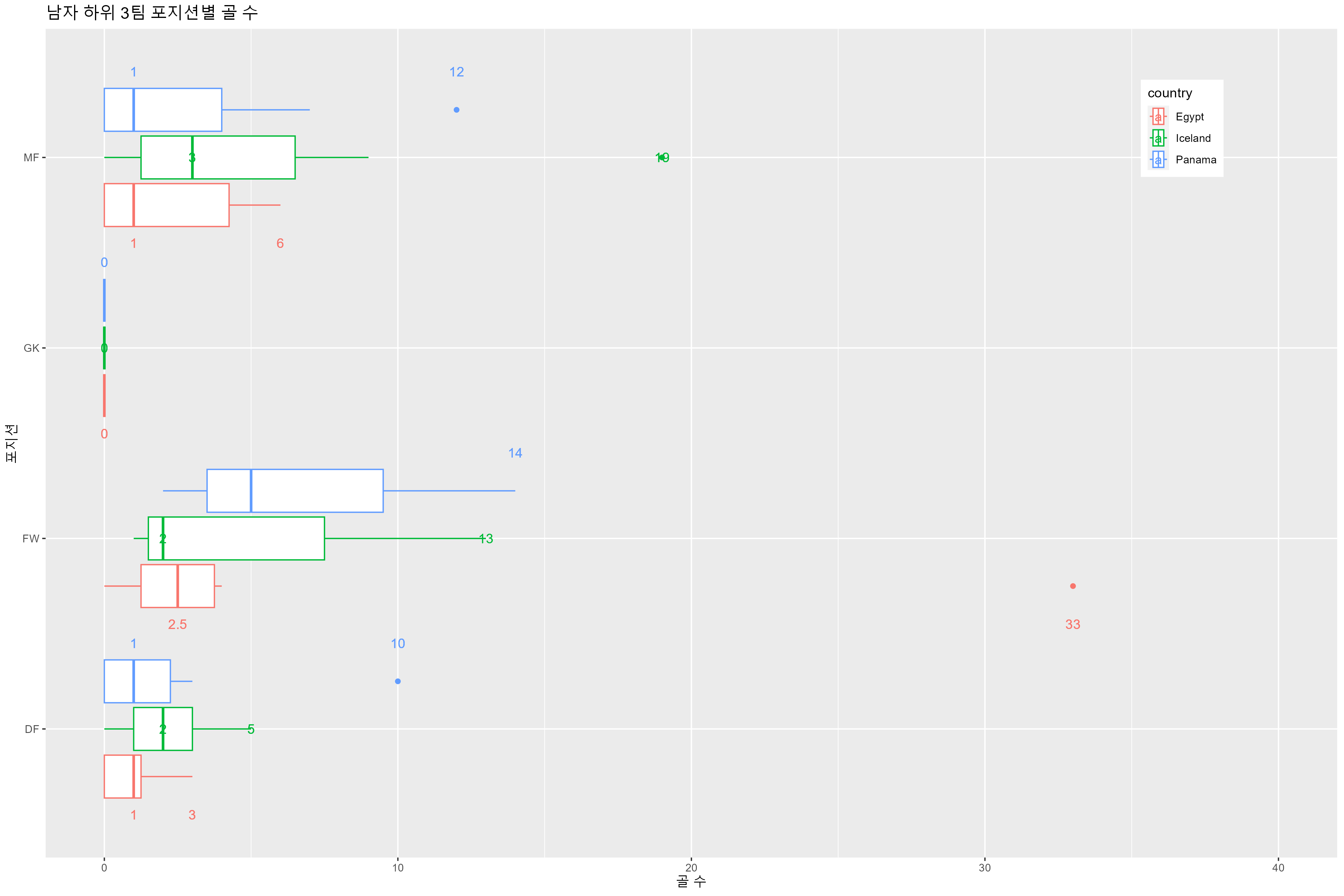
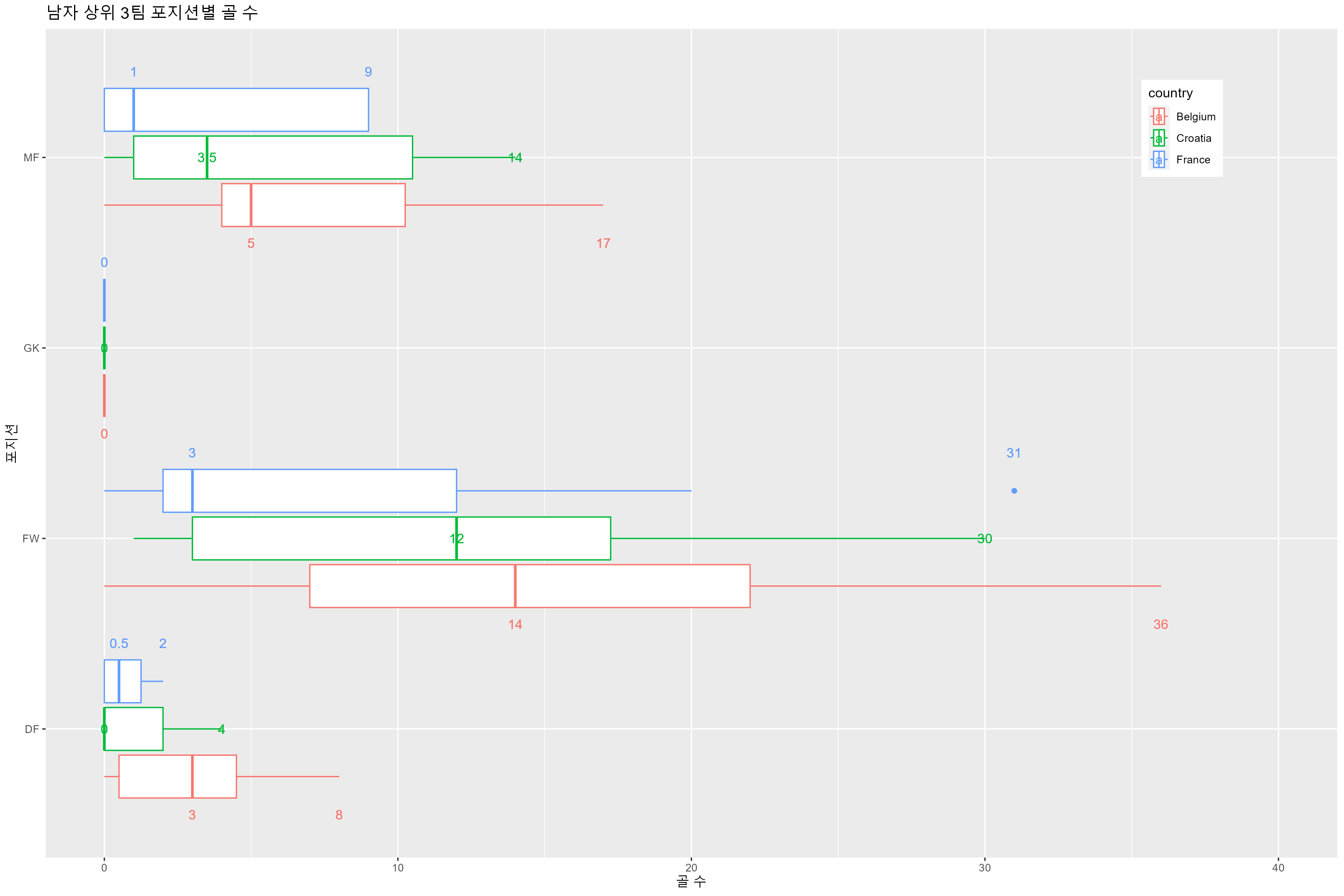
2-3) 포지션별 출전 횟수

출전 횟수는 데이터 집합의 범위와, 중앙값을 빠르게 확인하고 최솟값과 최대값을 넘어가는 위치에 있는 이상치 또한 확인하려는 목적으로 Boxplot을 사용하였다.



2-4) 포지션별 골 수

포지션별 골 수 역시 세 나라의 기술 통계량인 최댓값, 최솟값과 중앙값을 표현하여 비교 분석 하는 게 좋을 것 같아 Boxplot을 사용하였다.

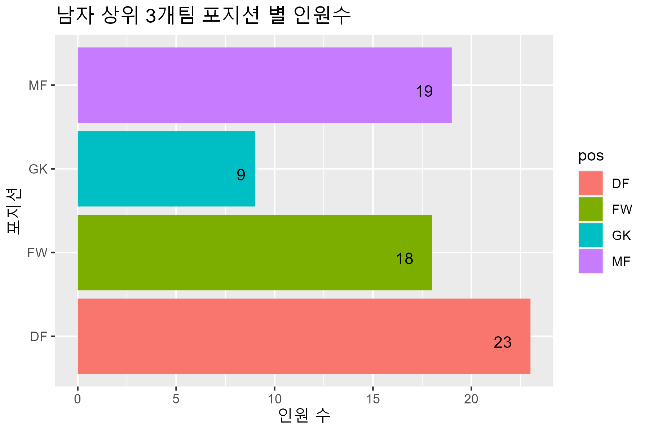
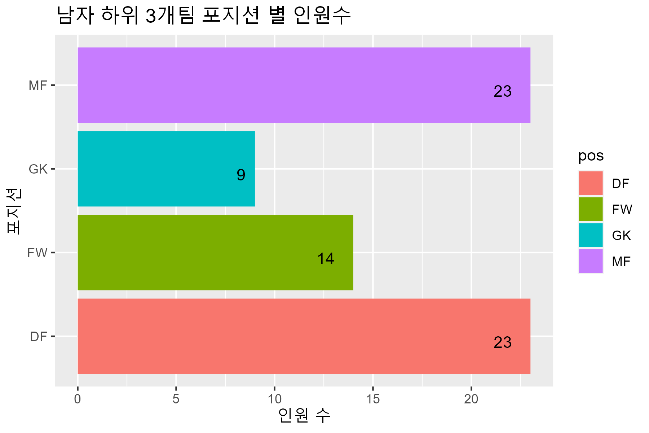


[여자]

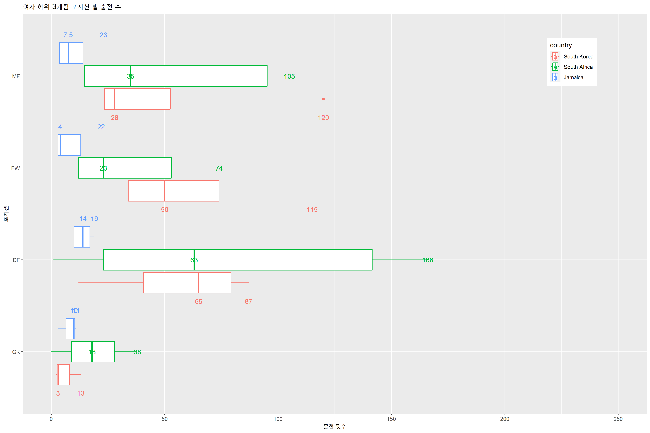
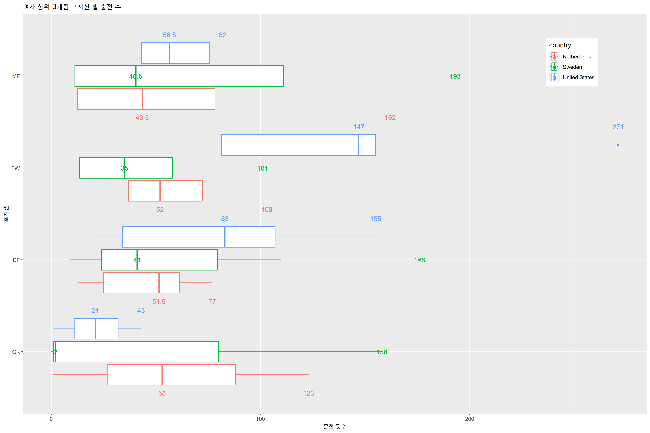
상위 팀 3개=미국 네덜란드 스웨덴 / 하위 팀 3개=자메이카 대한민국 남아공

2-1) 포지션별 나이대

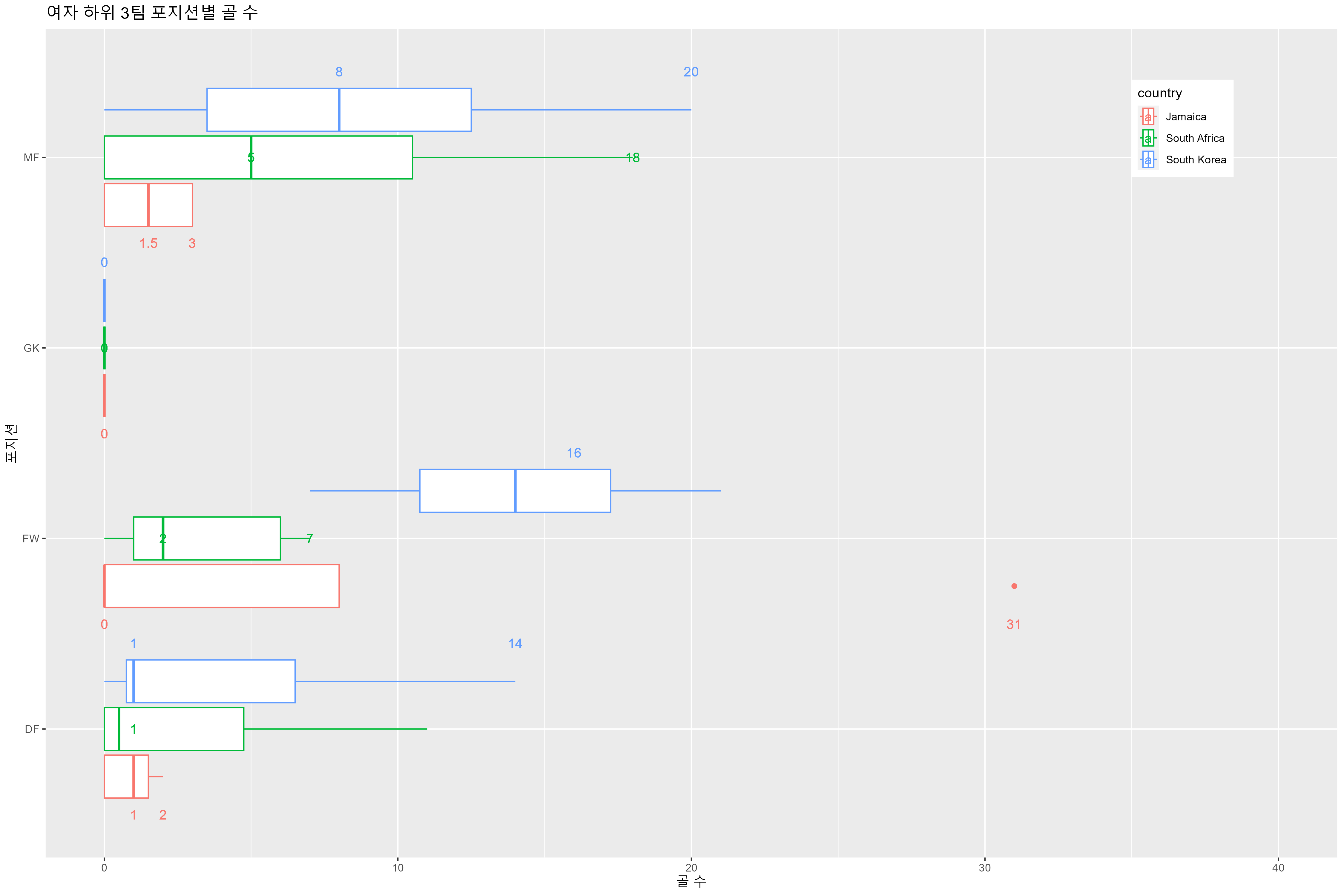
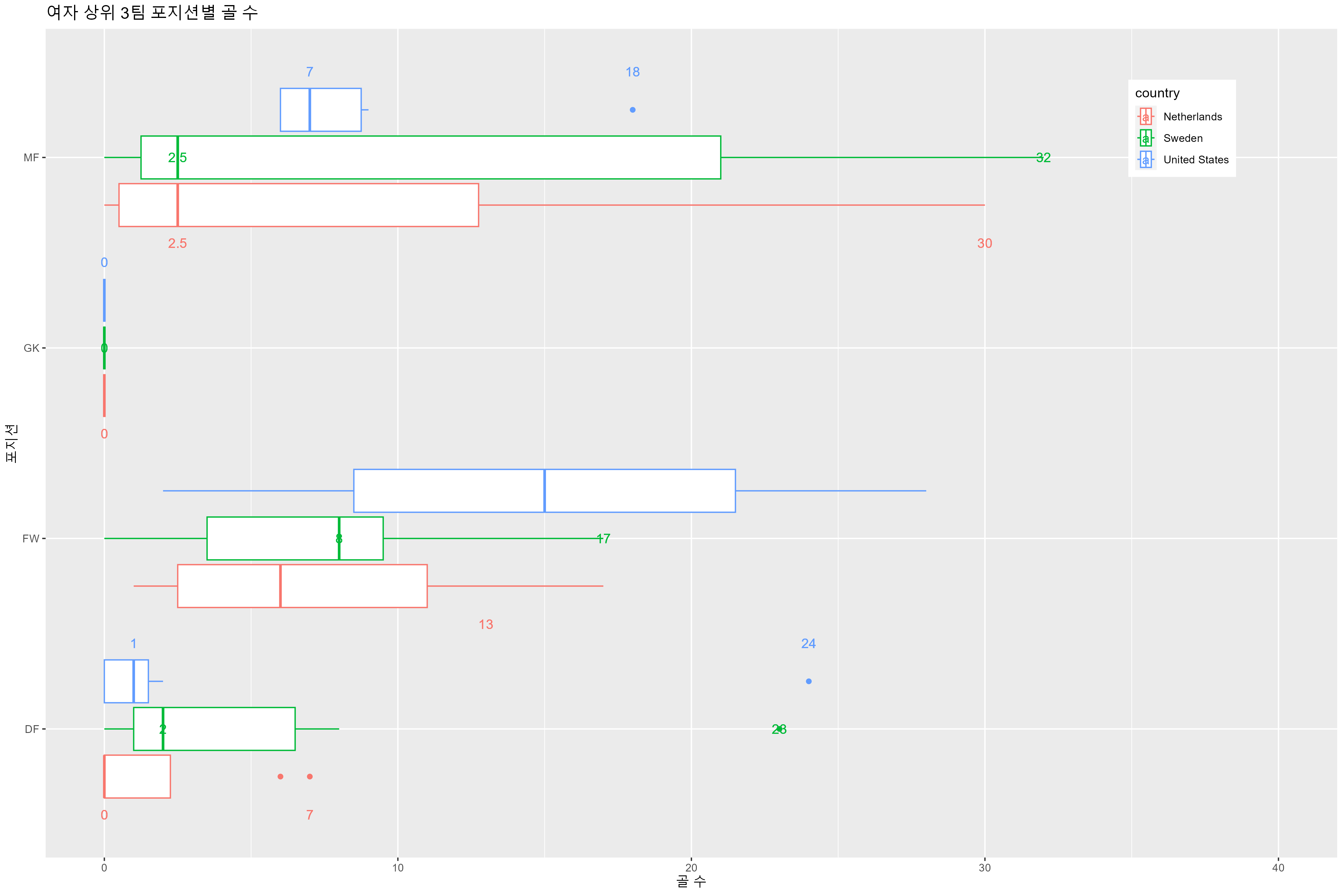
분석 결과 유의미한 결과를 내지 못했다.

2-2) 포지션별 선수의 수

2-3) 포지션별 출전 횟수



2-4) 포지션별 골 수



# 3 주요 결과 제시

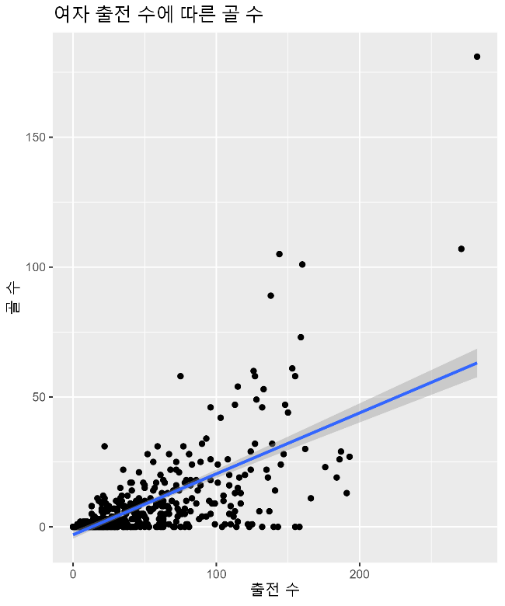
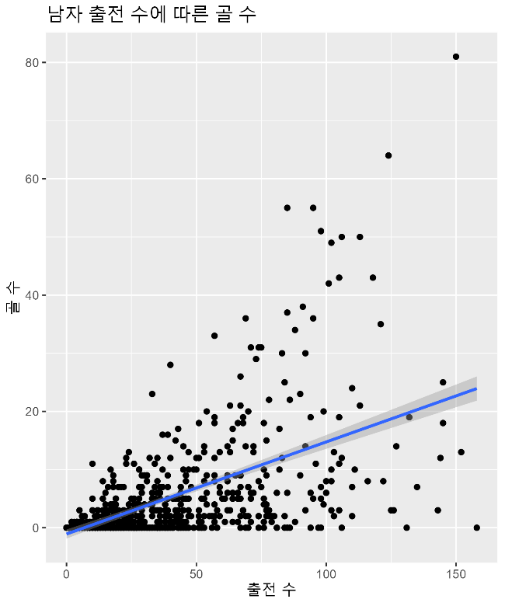
포지션별 나이대: 주어진 데이터를 한국식 나이로 분석했을 때 남자 상위팀은 20대, 여자 상위팀은 30대 공격수를 많이 뽑는다.

포지션별 선수의 수: 남녀 모두 상위팀은 하위팀에 비해서 공격수를 많이 뽑고 미드필더를 적게 뽑는다.

포지션별 출전 횟수: 남자는 상위 팀과 하위 팀 모두 제각각이고, 여자는 그래프를 통해서 하위 팀 중 하나인 자메이카가 포지션별 모두 출전 횟수가 현저히 적다.

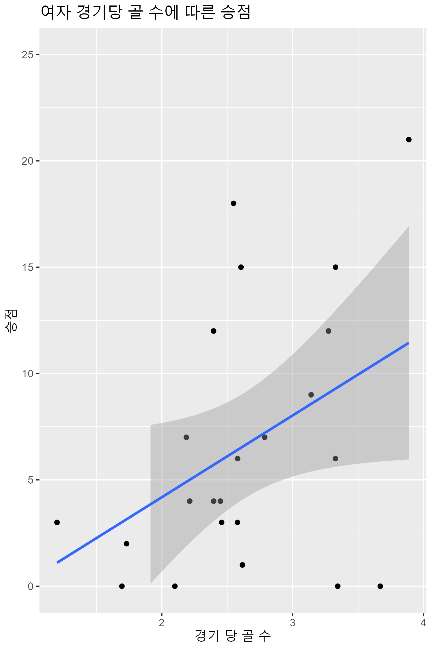
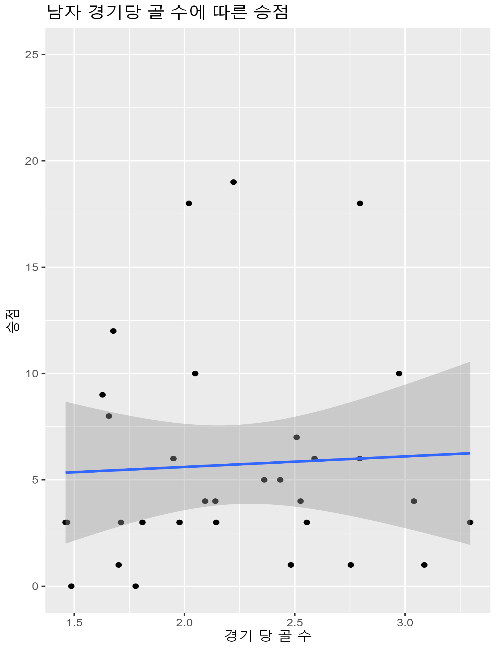
포지션별 골 수: 남자 우승팀은 프랑스는 상위 팀 중에서 포지션별 모두 골 수가 제일 적고, 여자 우승팀은 미국은 상위 팀 중에서 공격수에 골 수가 압도적으로 많다.

# 4 데이터를 통해 얻은 인사이트 논의

출전 횟수에 따른 골 수

그래프와 같이 출전 횟수가 많으면 많을수록 골 수도 많아진다. 여기서 남자는 상관계수가 0.57, 여자 상관계수는 0.67로 높은 관계 형성하는 것을 알 수 있다.

남녀 경기당 골 수에 따른 승점



그래프와 같이 골 결정력과 승점에 상관계수는 남자가 0.05, 여자가 0.4이다.

<결론>

남녀별 인원수를 통하여 상위 팀들의 공격수가 많은 것을 확인하였고, 공격수들이 많은 것이 골 결정력과 승리와 관련이 있는지 알아보기 위해 추가로 상관관계 분석을 진행하였다. 여기서 골 결정력을 '골 수/ 출전 수' 라고 할 때, 남자 골 수 와 출전 수와의 상관계수는 0.57로 높은 관계를 형성한다. 이후 2018년 월드컵 승점 데이터를 가져와 추가로 분석하였을 때 골 결정력과 승점(등수)과의 관계는 0.05로 굉장히 낮은 것으로 보아 골 결정력이랑 등수는 골 수로 판단할 수 없다는 것을 알 수 있었다. 여자도 같은 방식으로 골 수와 출전 수와의 상관계수는 0.67, 골 결정력과 승점과의 관계는 0.4로 관계가 없지는 않지만, 여자 또한 골 결정력으로 순위에 영향을 줬는지 판단할 수 없다는 것을 알 수 있었다.