

REPORT

[2차 과제_14조]

과 목 명 : R로배우는데이터분석입문

이 름 : 2014720 정치외교학과 김성은

1912392 문화관광학부 이서연

1912168 문화관광학부 이호림

1716862 경영학부 조은아

1. 데이터 소개

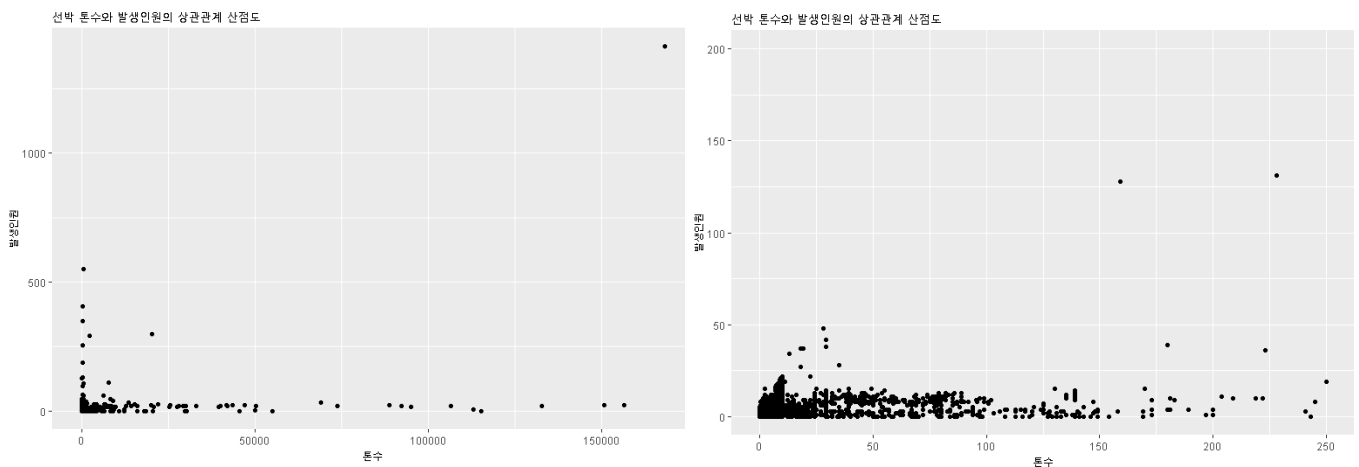
데이터 이름: 해양경찰청_해상조난사고 상세데이터 현황

데이터 내용: 2020.1.1 ~ 2020.12.31 동안 대한민국 해상에서 발생한 해양사고 해양경찰청에 신고, 접수된 총 3778건의 해양 사고 상세 데이터를 담고 있다.¹

- 날짜 척도(1): 발생일시(date로 이름 변경)
- 범주형 척도(8): 월별/시간대별/관할해경서/발생해역/기상상태/발생원인/발생유형/선종
- 계량형 척도(7): 사고선박수/발생인원/구조/부상/사망/실종/톤수

2. 산점도

2-1) 선박 톤수와 발생인원의 상관관계 산점도



```
ggplot(sea, aes(톤수, 발생인원)) + geom_point() + labs(title = "선박 톤수와 발생인원의 상관관계 산점도") + xlim(0,250) + ylim(0,200)
```

- ➔ 선박 톤수와 발생인원 간의 뚜렷한 상관관계는 없었으며, 우측 산점도를 통해 대체로 하루 기준 평균 사고 발생 인원은 5명, 그에 따른 선박 톤수의 대부분이 100톤 이하로 나타났다는 사실을 파악했다. 더불어, 전체 관측치를 다룬 좌측 산점도의 우측 상단에 위치한 이상점을 통해 2020년 내에 하루 약 1500명 이상의 사고 인원이 발생한 해양 사고가 있었음을 알 수 있었다.

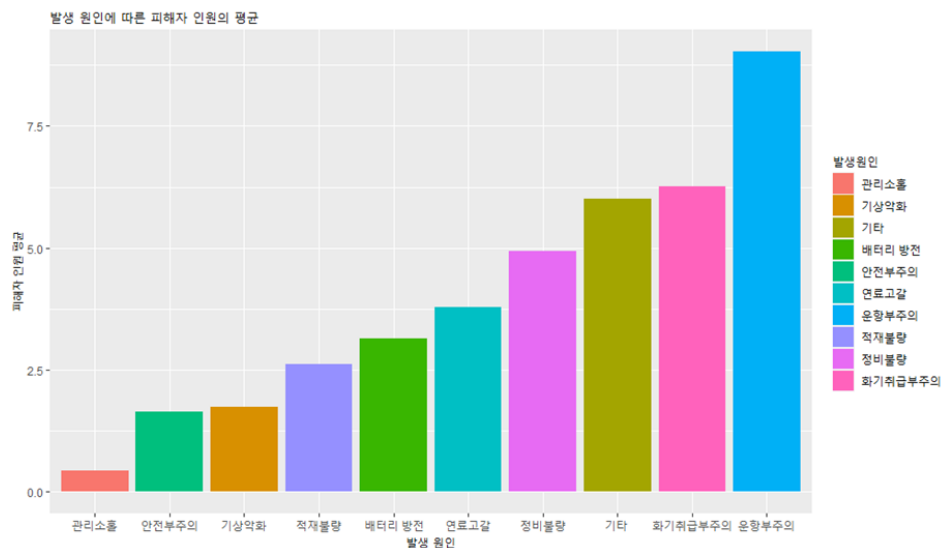
¹ 공공데이터포털, “공공데이터포털”, <https://www.data.go.kr/data/3043630/fileData.do> (2021.11.21)

월	고무보트	관공선	기타	낚시어선	모터보트	어선	여객선	예부선	요트	유도선	유조선	화물선
1	10	10	10	10	10	170	150	10	10	10	10	10
2	10	10	10	10	10	120	120	10	10	10	10	10
3	10	10	10	10	10	180	120	10	10	10	10	10
4	10	10	10	10	10	120	120	10	10	10	10	10
5	10	10	10	10	10	200	100	10	10	10	10	10
6	10	10	10	10	10	160	160	10	10	10	10	10
7	10	10	10	10	10	160	200	10	10	10	10	10
8	10	10	10	10	10	220	120	10	10	10	10	10
9	10	10	10	10	10	260	100	10	10	10	10	10
10	10	10	10	10	10	180	180	10	10	10	10	10
11	10	10	10	10	10	140	160	10	10	10	10	10
12	10	10	10	10	10	180	160	10	10	10	10	10

```
ggplot(sea,aes(월별, fill = 선.종)) + geom_bar() + labs(title = "선종에 따른 월별 해양사고 발생 빈도 막대그래프", x = "월별", y = "사고 발생 빈도수")
```

- ➔ 해양 사고 발생 빈도수가 가장 높은 시기는 9월이었다. 9월의 결과값을 선박의 종류로 구분해보면, 다른 범주들보다 특히 고무보트, 모터보트, 낚시어선의 비중이 늘어난 것을 확인할 수 있다. 고무보트와 모터보트는 주로 사람들이 해양 레저 스포츠를 즐길 때 많이 이용하는 선박이고, 낚시어선은 국가에 어업신고를 하고 전문적으로 어업활동을 하는 어부가 아닌 낚시를 취미로 즐기는 일반인들이 고기를 잡기 위해 주로 사용하는 선박이다. 기상청 통계에 따르면 2020년 9월 평균 기온은 20.3도였고, 9~10월에는 수온이 적당해 물고기의 움직임이 활발해진다는 사실²을 고려해 봤을 때, 해양 레저스포츠와 취미낚시를 즐기는 사람들이 늘어남에 따라 9~10월의 해양 사고 발생 빈도 또한 함께 증가했다는 가설을 추측할 수 있다.

4-2) 발생 원인에 따른 피해자 인원의 평균



```
sea1 <- sea %>% group_by(발생원인) %>% summarise(발생인원평균=mean(발생인원))
ggplot(sea1, aes(reorder(발생원인, 발생인원평균), 발생인원평균, fill=발생원인)) +
geom_bar(stat = "identity")+labs(title="발생 원인에 따른 피해자 인원의 평균", x="발생 원인", y="피해자 인원 평균")
```

- ➔ 운항부주의로 인한 해양 사고의 하루 평균 피해자 인원이 가장 많았다는 사실을 알

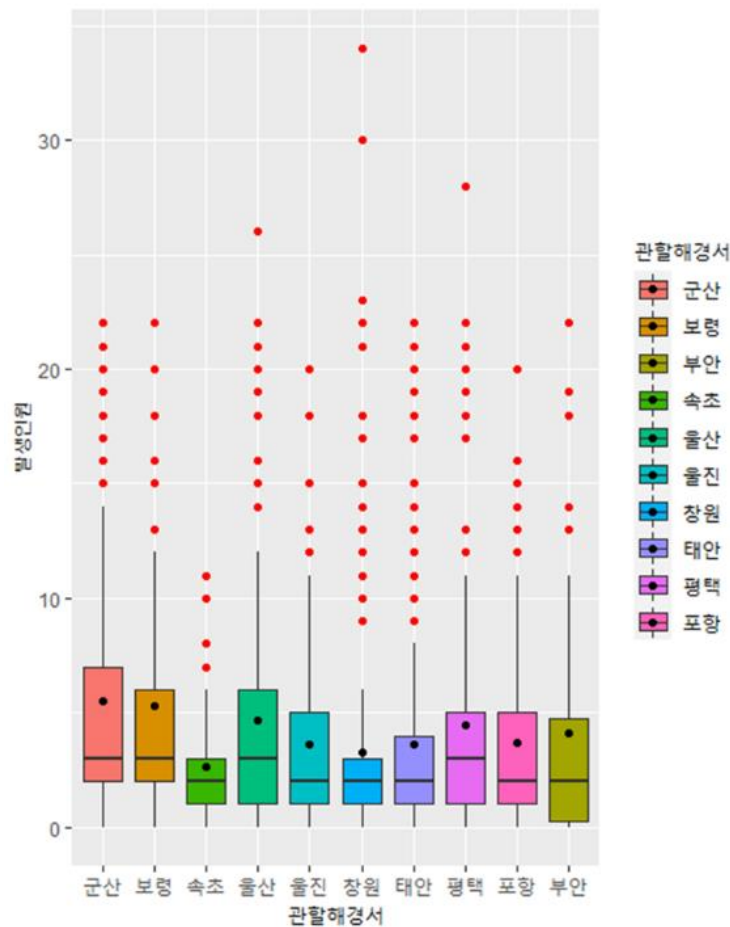
² 인어교주해적단, “가을에 고기가 잘 잡히는 이유는?”,

<https://www.tpirates.com/%EC%BD%98%ED%85%90%EC%B8%A0/2389/%EC%B4%88%EB%B3%B4%EC%9E%90%EB%A5%BC-%EC%9C%84%ED%95%9C-%EB%B0%94%EB%8B%A4%EB%82%9A%EC%8B%9C-%ED%8A%B9%EC%A7%91-2>, 2021.11.21

수 있었다. 앞선 그래프를 통해 사고 발생 빈도수가 가장 높은 시기는 9월이라는 것을 파악했다. 여름~초가을의 날씨 특성을 고려해 봤을 때 해양 사고의 주된 원인으로 태풍 등의 기상악화가 가장 비중이 크지 않을까 예상했지만, 오히려 운항부주의와 화기취급부주의가 대부분의 해양 사고의 가장 큰 요인이었다는 사실을 파악할 수 있었다.

5. 상자 그래프

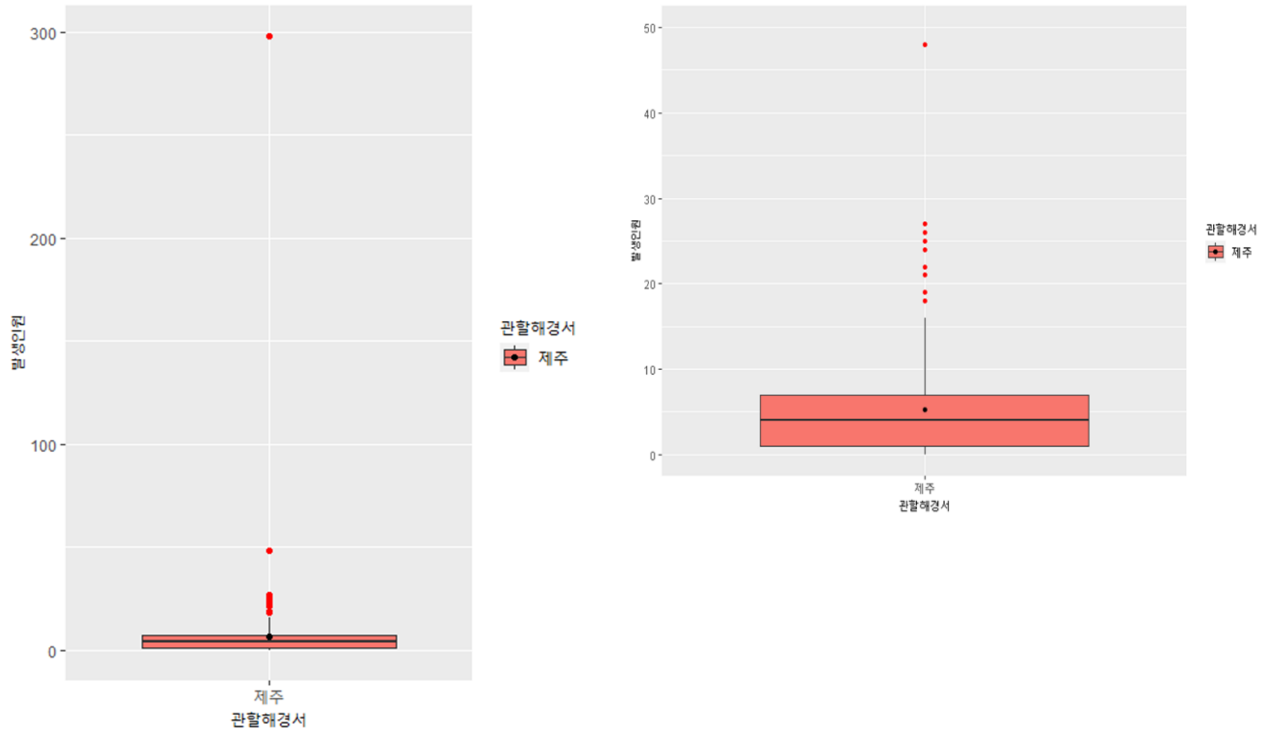
5-1) 관할 해경서별 사고 발생인원 상자그래프 (발생인원 척도 최대 35명 이내 지역들)



```
ggplot(sea, aes(관할해경서, 발생인원, fill = 관할해경서)) + geom_boxplot(outlier.colour = "red")+stat_summary(fun="mean", geom = "point")+xlim(c("군산", "보령", "속초", "울산", "울진", "창원", "태안", "평택", "포항", "부산"))
```

- ➔ 데이터가 가장 넓게 분포된 지역은 군산이었다. 그에 반해 창원의 관측값들의 분포는 나머지 범주들보다 좁게 나타났지만, 가장 높은 이상치를 지닌 지역 또한 창원이었다. 전체 지역의 중앙값의 위치는 대체로 비슷했으며, 하루 평균 발생인원은 군산이 가장 많고, 속초가 가장 적었다.

5-2) 제주시 발생인원 분포 상자그래프 (큰 값의 이상치가 존재하는 지역)



```
ggplot(sea, aes(관할해경서, 발생인원, fill = 관할해경서)) + geom_boxplot(outlier.colour = "red")+stat_summary(fun="mean", geom = "point")+xlim(c("제주")) + ylim(0,50)
```

- ➔ 제주의 하루 평균 피해 발생인원은 대략 6명이었다. 중앙값이 가운데에 위치하기 때문에 관측값들의 전체적인 분포는 고른 편이나, 평균 피해 발생인원과 크게 벗어난 약 300명 가량의 높은 이상치 값을 확인할 수 있었다. 제주 지역의 상자그래프와 비슷하게 부산, 서귀포, 통영에서도 각각 1413명, 254명, 60명의 높은 이상치가 존재했으며, 세 지역의 상자 그래프 또한 제주 지역의 상자그래프와 비슷한 형태로 나타났다.