

지구온난화와 이상기후와의 관계

20205242 장우혁
데이터사이언스 기초[03] 기말프로젝트

목차

01

주제 선정이유

04

강수량 회귀분석

1990~2022까지의 강수빈도 분석

02

온실가스 증가추세

1990~2022까지의 온실가스
증가추세 분석

05

강수빈도 회귀분석

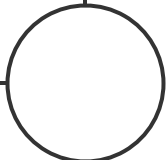
03

강수량과 강수빈도

1990~2022까지의 강수량 및
강수빈도 분석

06

결론

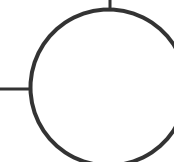


주제선정 이유

최근 인터넷에 떠돌고있는 한 뉴스를 접했다. 바로 7월에는 단 사흘만 맑은 날씨이고, 나머지는 비가 올 것이라는 예보였다. 물론 기상청의 공식 발표는 신빙성이 없는 자료이며, 아무리 슈퍼컴퓨터라고 해도 최대 2주정도의 예측만이 신빙성이 있는 예측이라고 하였다.

하지만 나는 생각이 달랐다. 이미 지구온난화에 대한 문제가 거론된지 오래되었고, 최근 몇년동안 “몇년만의 폭우” “몇년만의 한파”등등의 이상기후가 계속해서 발생하고 있다. 우리는 이를 알면서도 자연환경이 파괴되는것에는 관심이 없고, 오직 우리의 이득을 위해서만 살아가고 있다.

이번 프로젝트를 통해서 이상기후들이 지구온난화와 관련이 있는지와 왜 자꾸 “몇년만의 폭우”라는 타이틀이 붙는지 알아보고싶고, 올해는 얼마나 비가 오는지 예측을 해보고싶다.



가설설정 유의수준 변수설정

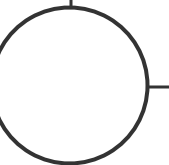
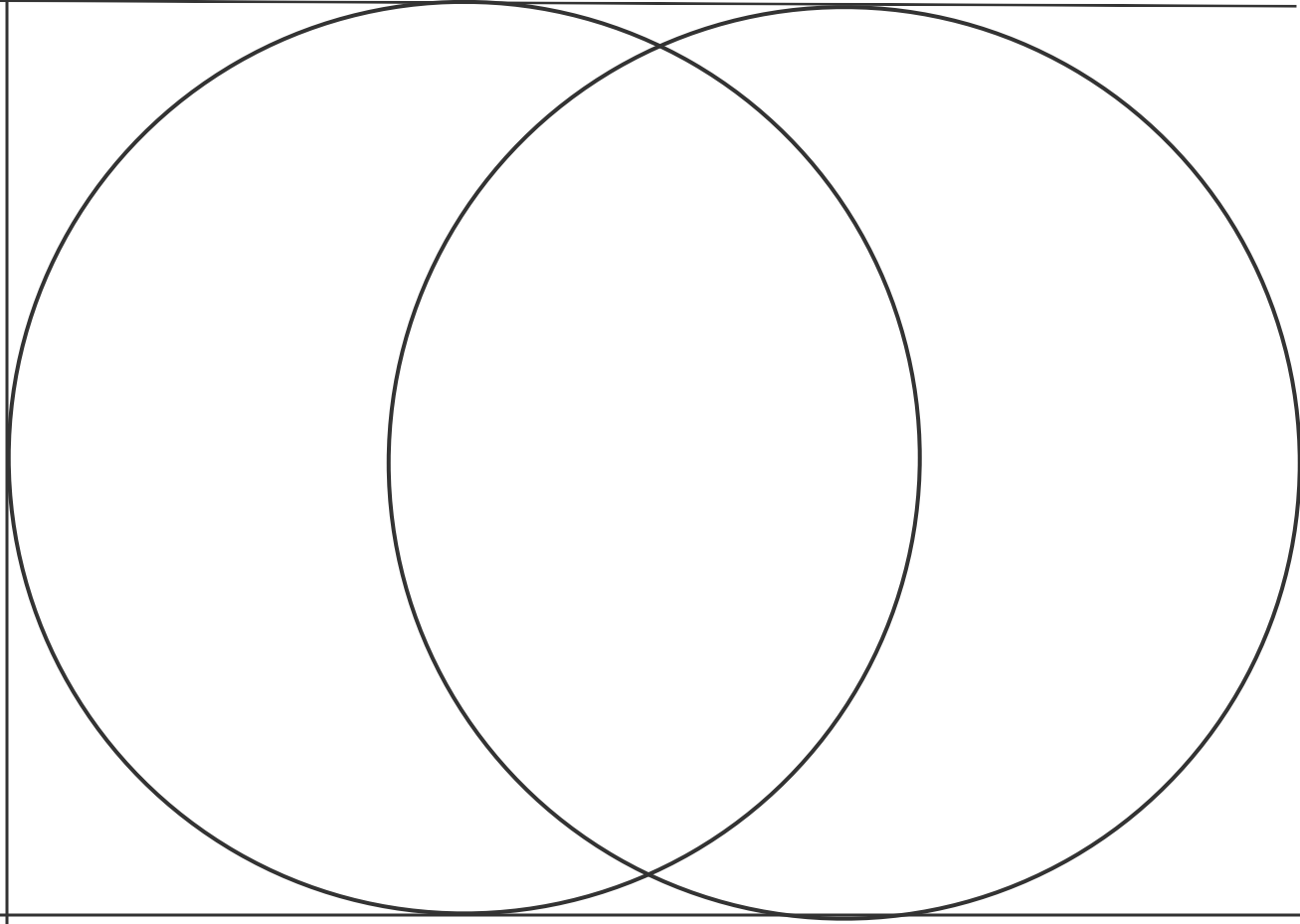
귀무가설 : 강수량은 지구온난화와 관련이 없다.
대립가설 : 강수량은 지구온난화와 관련이 있다.

귀무가설 : 강수빈도는 지구온난화와 관련이 없다.
귀무가설 : 강수빈도는 지구온난화와 관련이 있다.

유의수준 : 0.05

종속변수 : 온실가스 지표
독립변수 : 강수량, 강수빈도

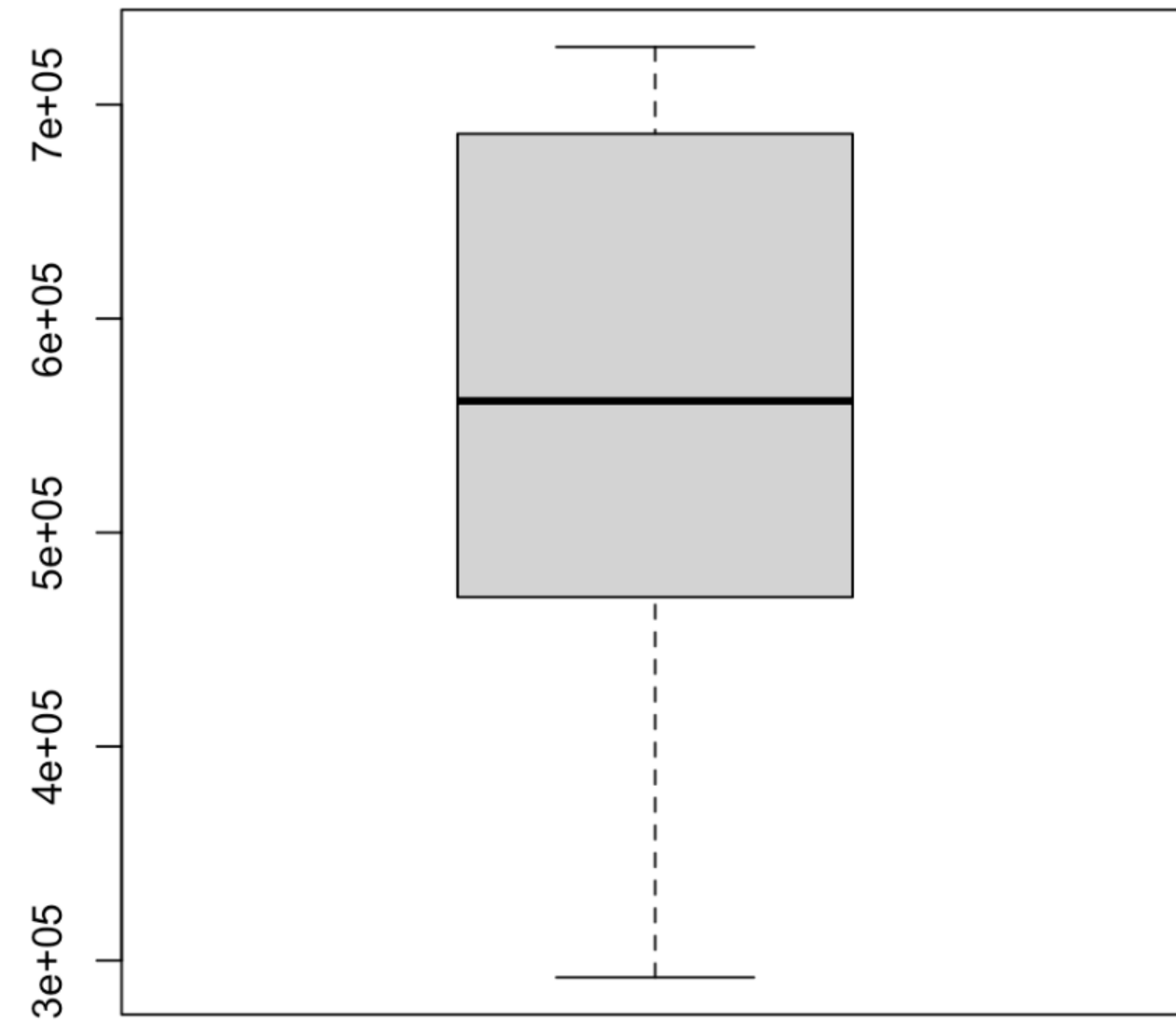
이를 바탕으로 회귀분석을 통해 폭우와 잦은 비가 지구온난화와 관련이 있는지 확인할 수 있다.



온실가스 증가추세

경부 온실가스종합정보센터_국가 온실가스 인벤토리 배출량_20221213

출처 : 공공데이터포털



1990년도부터 2022년까지의 온실가스 증가추세를 알기 위해서 온실가스 데이터 중에서 총 배출량 행을 선택하여 데이터를 가져왔습니다.

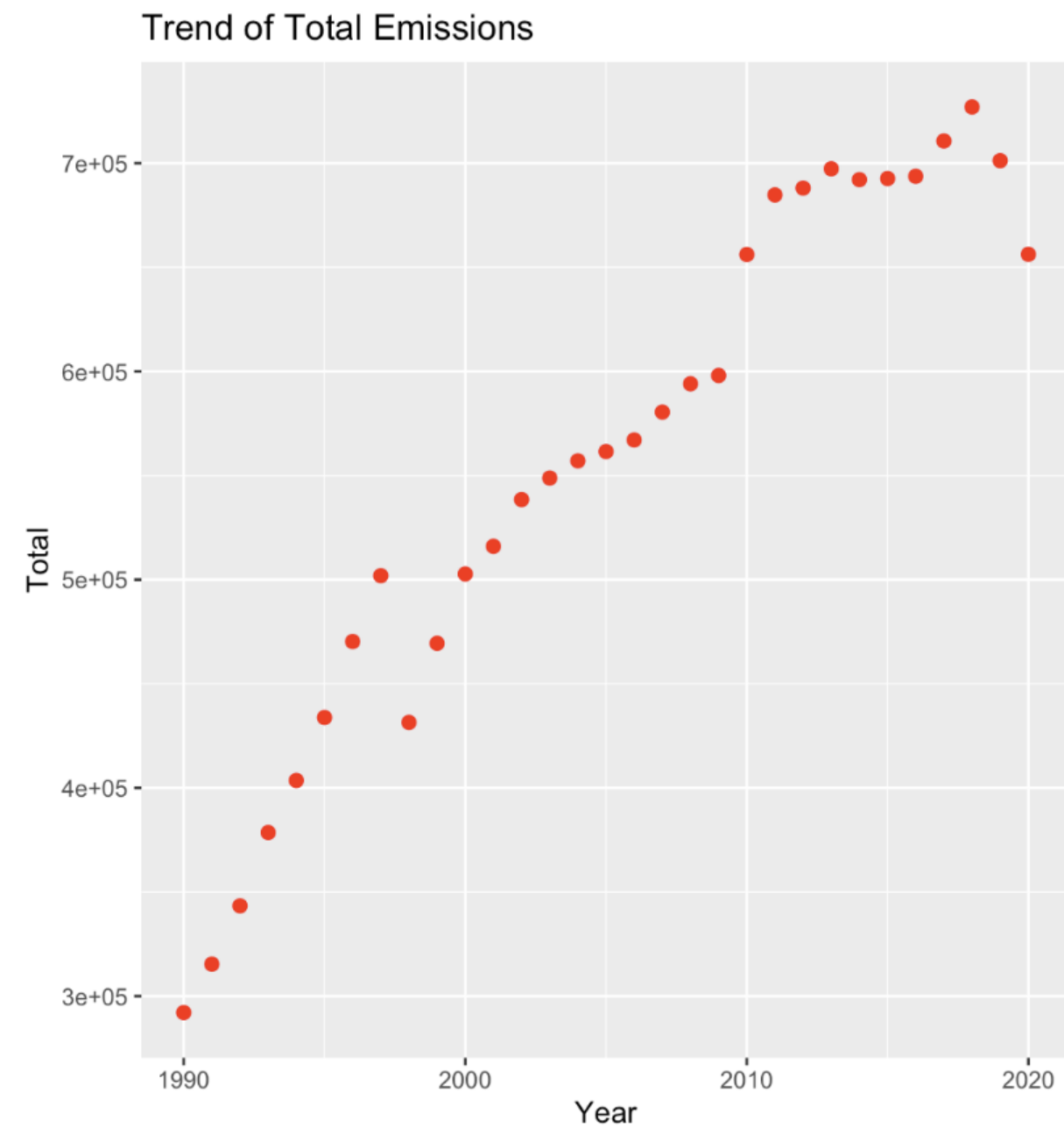
데이터를 가공하는 과정에서 csv파일의 데이터 형태가 불편함을 발견했고, 이를 편하게 고치기 위해 `t()`함수와 `cbind()`함수, `colnames()`함수를 사용하였습니다.

결측값이 있나 확인하기 위해 `boxplot()`을 이용하여 결측값을 확인하였고, `na.omit()`을 이용하여 결측값이 있는 행 전체를 제거하였습니다.

온실가스 증가추세

경부 온실가스종합정보센터_국가 온실가스 인벤토리 배출량_20221213

출처 : 공공데이터포털

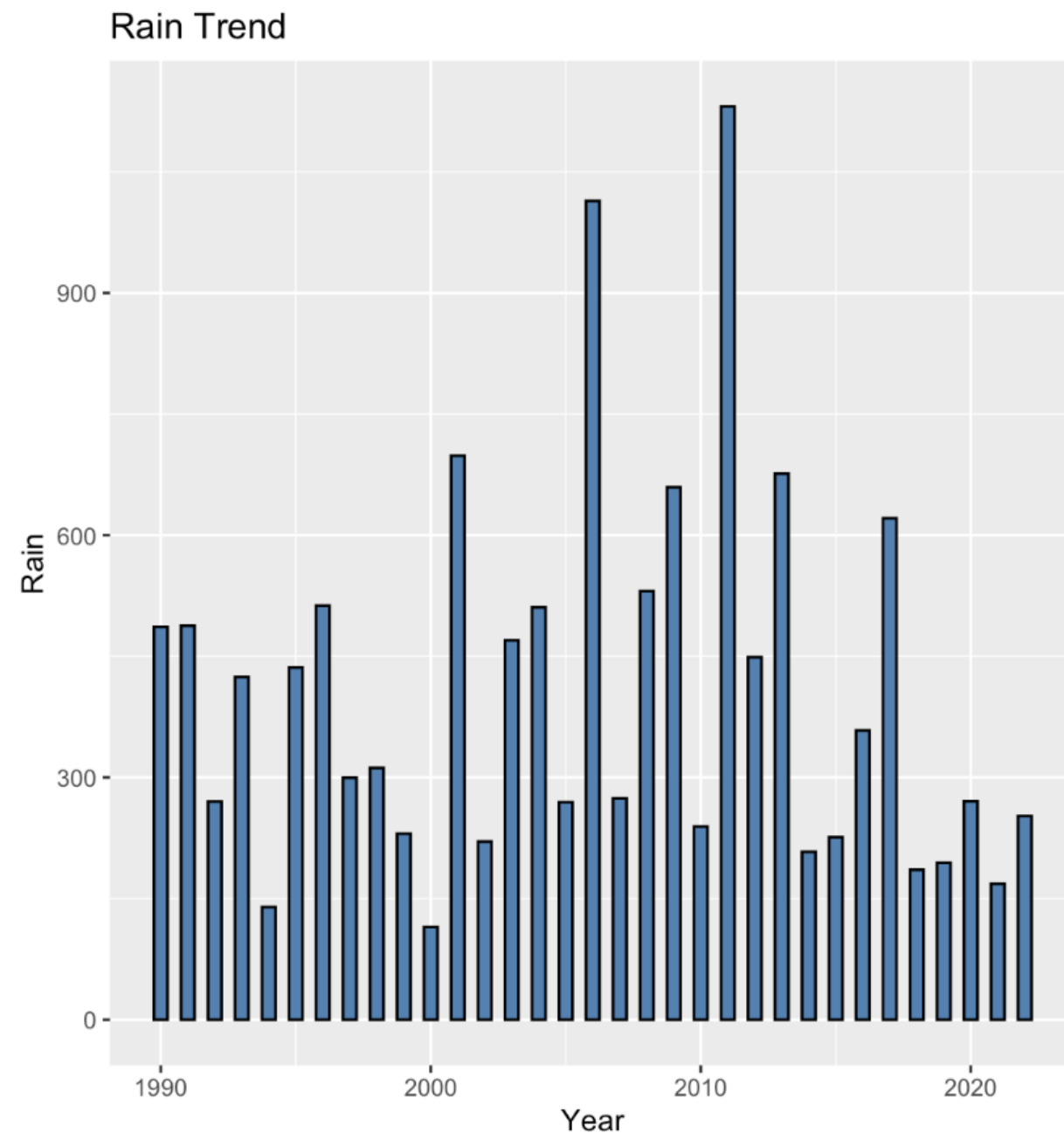


데이터들을 num형으로 바꿔주기 위해 `as.numeric()` 함수를 사용하였고, 이전까지 데이터 가공한것을 토대로 `ggplot()` 함수를 사용하여 그래프를 그렸습니다.

확실히 눈에 띄는것은 그래프의 방향이 우상향을 띄고 있고, 이는 매년 우리가 배출하는 온실가스가 증가하고 있다는 것을 알 수 있습니다.

강수량과 강수빈도

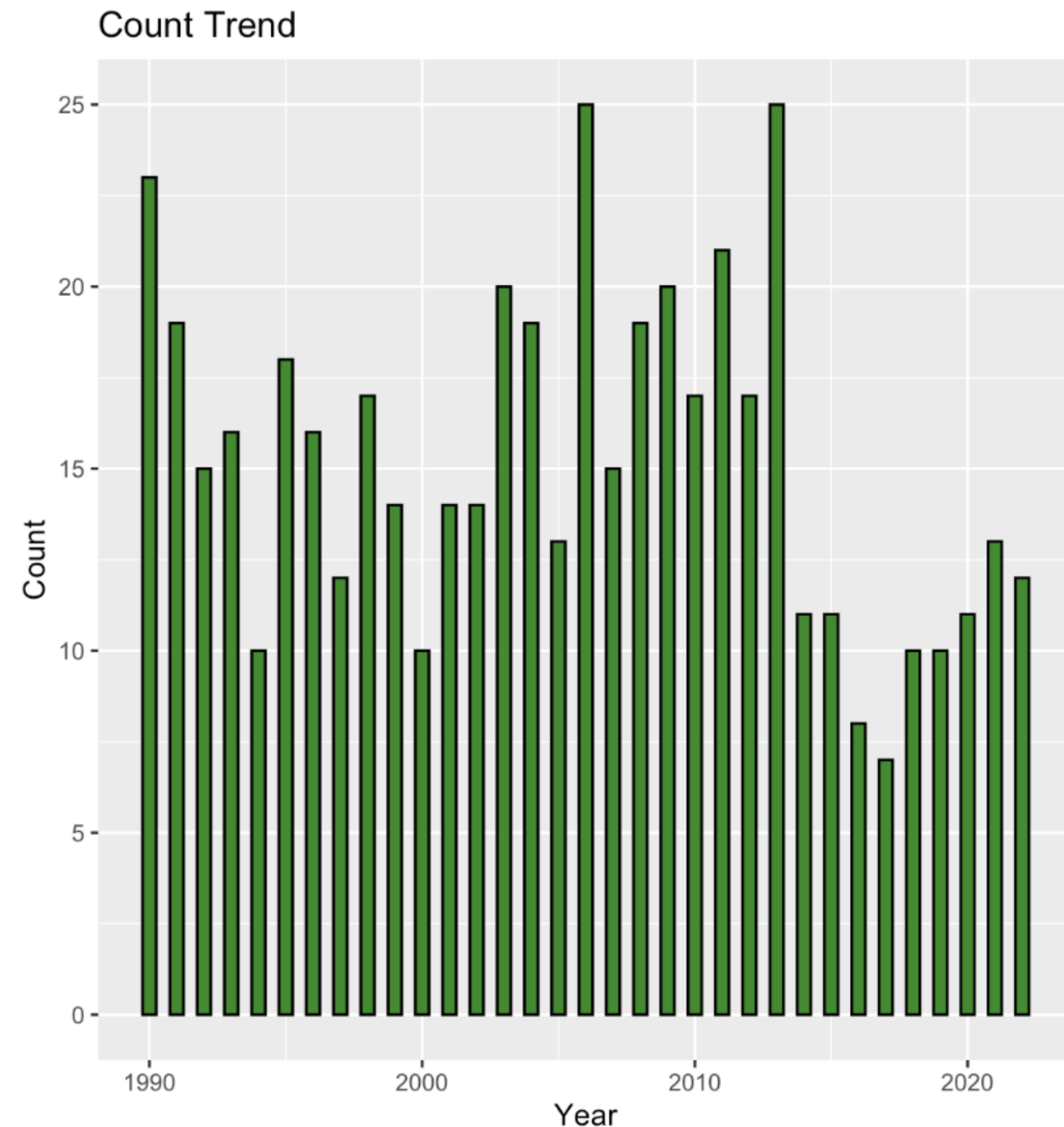
강수량분석, 강수일수
출처 : 기상청 기상자료개방포털



1년중 가장 강수량이 많다고 생각되는 '7월'행의 데이터를 모으기 위해서 `filter()`함수를 사용했고, 필요한 열을 고르기 위해 `select()`함수를 사용했습니다. 이렇게 가공한 데이터는 `data.frame()`을 이용하여 데이터프레임으로 만들어 주었고, 두 데이터(강수량과 강수빈도) 모두 데이터프레임으로 만들어 준 뒤에는 'year'과 'month'를 기준으로 `left_join()` 해주었습니다.

강수량과 강수빈도

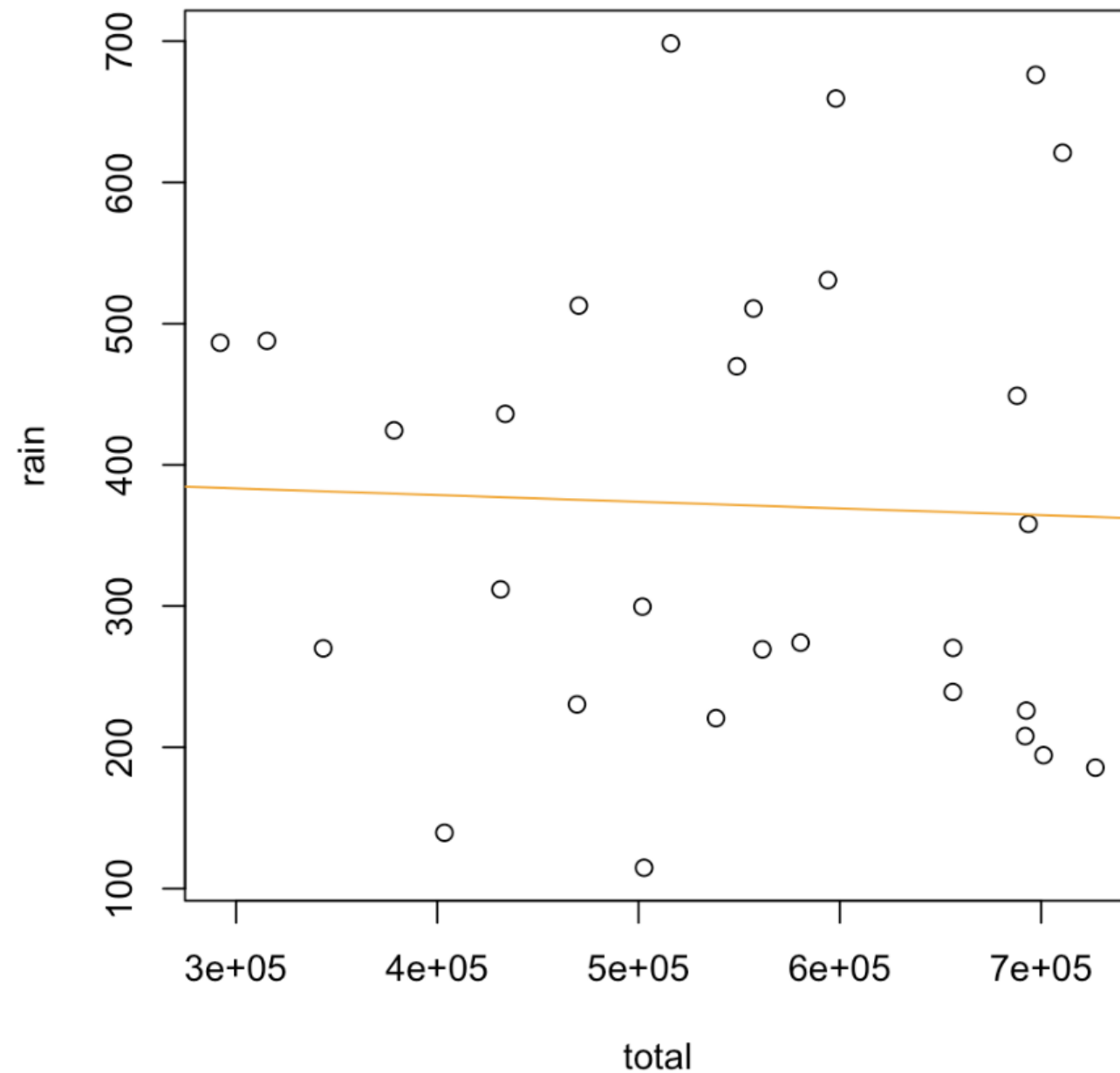
강수량분석, 강수일수
출처 : 기상청 기상자료개방포털



이전에 가공한 데이터를 바탕으로 그래프를 각각 그려주었습니다. 각 데이터들 끼리의 강수량과 강수 빈도를 비교해야하기 때문에 점이나 선그래프 보다는 막대 그래프가 좀 더 한눈에 잘 보일것이라 판단하고 `geom_bar()`를 이용하여 막대그래프를 그렸습니다.

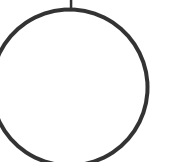
허용치범위 이상으로 비가 많이 오거나 횟수가 많았던 달이 언제인지 궁금해서 `boxplot()`을 이용하여 이상값이 있나 확인해본 결과, 강수빈도에서는 이상값이 없었으나 강수량에서 이상값이 2회 발견되었고 `na.omit()`을 이용하여 이상값이 있는 행을 삭제하였습니다.

강수량 회귀분석

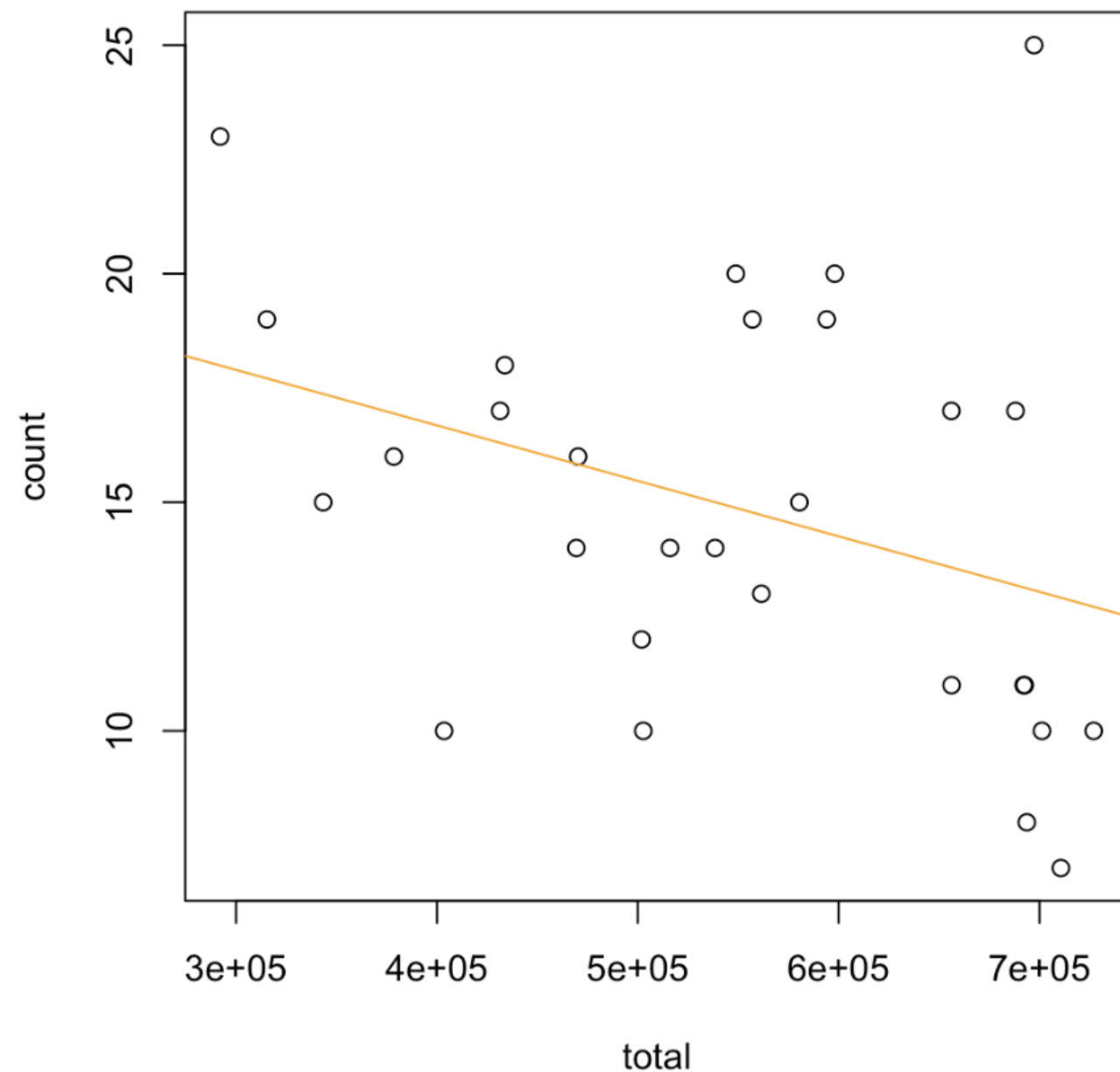


선형회귀분석을 하기 위해 회귀모델을 산출하고 시각화 하였습니다. 종속변수는 지구온난화(total데이터) 이고, 독립 변수는 강수량(rain)으로 설정하였습니다. `lm()`함수를 이용하여 회귀모델을 산출하였고, `plot()`함수와 `abline()`함수를 이용하여 시각화를 하였습니다.

`summary()`함수를 사용한 결과 유의수준이 0.8515로 제가 설정했던 0.05보다 훨씬 높은 값을 보였고, 이로 인하여 대립가설을 기각하고 귀무가설을 채택하였습니다.



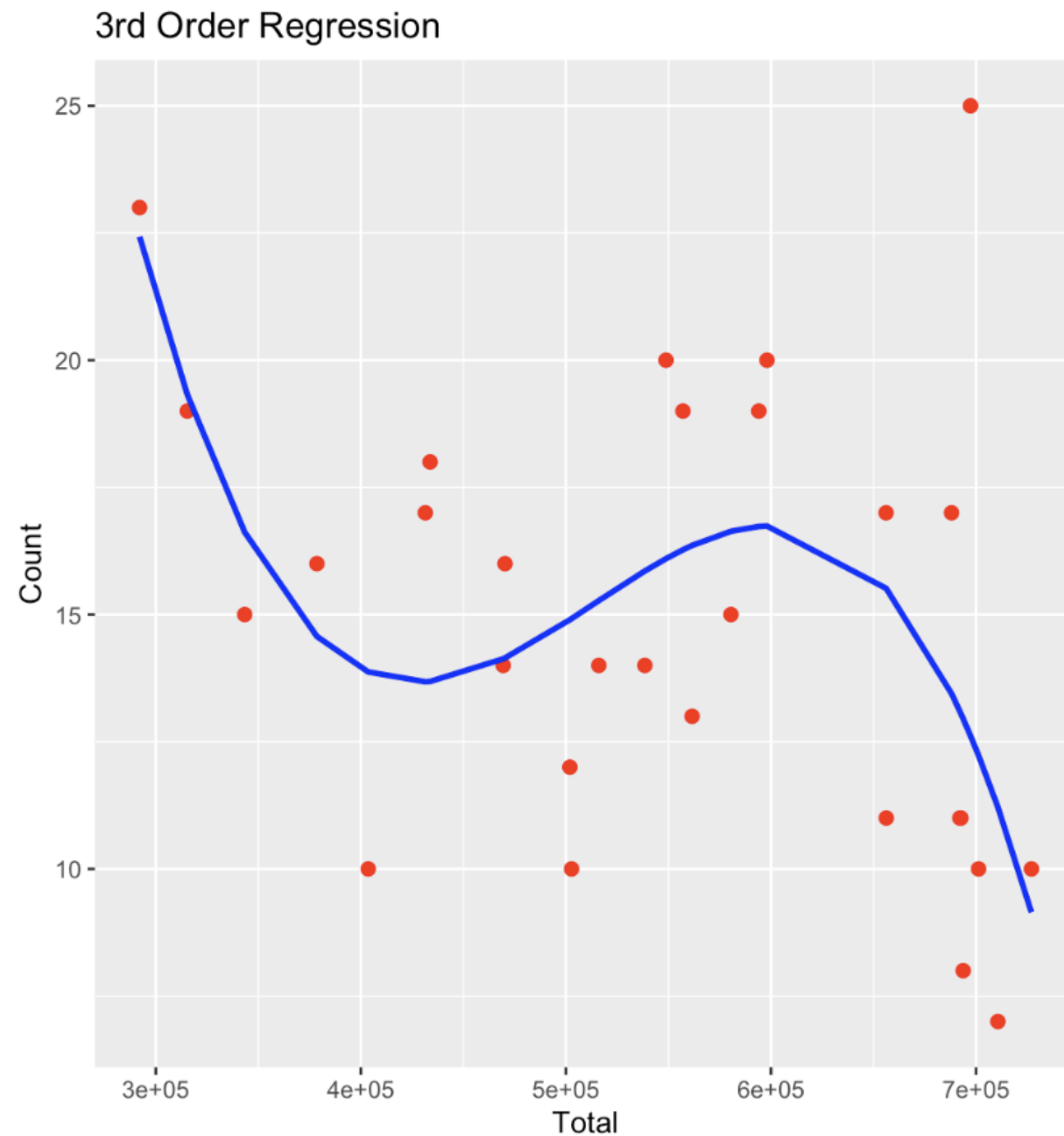
강수빈도 회귀분석



두번째로는 강수빈도의 회귀모델을 산출하였습니다. 종속 변수는 지구온난화(total데이터) 이고, 독립변수는 강수빈도(count)으로 설정하였습니다. lm()함수를 이용하여 회귀모델을 산출하였고, 마찬가지로 plot()함수와 abline() 함수를 이용하여 시각화를 하였습니다.

summary()함수를 사용한 결과 유의수준이 0.06145로 제가 설정했던 0.05보다 미세하게 높은 값을 보였고, 이로 인하여 강수빈도 역시 대립가설을 기각하고 귀무가설을 채택하였습니다.

강수빈도 회귀분석



그나마 조금 더 의미있는 분석을 했다고 생각되는 강수빈도 데이터를 가지고 3차원 모델을 산출하여 회귀분석을 해보았습니다. 역시 점들이 선과 꽤 멀리 떨어져 있는것으로 보아 강수빈도도 지구온난화와 밀접한 연관이 있다고는 보기 어렵다는 결론이 나왔습니다.

3차 회귀분석 모델을 활용하여 2023년의 강수빈도를 예측을 해보았더니 163.5908의 값이 나왔습니다. 올해는 163번의 비가 내린다는 예측을 할 수 있으나, 유의수준이 0.05보다 큰 것으로 보아 163번의 비가 내린다는 말은 신뢰성이 떨어지는 예측이라고 생각합니다.

결론

<두개의 가설 모두 귀무가설 채택 "강수량과 강수빈도는 지구온난화와 관련이 없다."

최근들어서 비가 왔다 하면 "몇년만의 폭우", 조금 더워졌다 하면 "몇년만의 폭염" 이런 뉴스 타 이틀을 더 많이 본 것 같습니다. 저는 그 이유가 당연히 지구온난화라고 생각했고, 지구온난화 의 주된 원인은 온실가스배출이 많아졌다고 생각하였습니다. 온실가스 배출이 매년 증가추세 인것은 그래프를 통해서 확인 하였지만, 강수량과 강수빈도가 온실가스 배출의 증가와 관련이 거의 없다는 것은 의외의 결과였습니다. 당연히 상관관계가 있다고 생각했던 가설들이 전부 기 각되니 회귀분석이 얼마나 중요한지 깨닫게 되었습니다.

귀무가설이 채택된것과는 별개로 온실가스 배출이 매년 증가하고 있다는 것은 매우 심각한 일 이라고 생각합니다. 사실 폭우와 폭염, 초대형 태풍등 이상기후 현상은 이미 병들어버린 지구의 신호라고 생각합니다. 이미 문명이 발전되어 어느 정도의 온실가스 배출은 불가피하지만, 온실 가스 배출을 최대한 줄여야 미래의 내 자식들이 오래오래 건강하고 맑은 지구에서 살 수 있다고 생각합니다.

