

경제학부 응용통계전공 통계경진대회



201411098 강도형
201411102 강재성
201411149 이준서

목차

1. 프로젝트 주제
2. 작업환경 및 Dataset
3. 프로젝트 내용

The logo consists of the text "GO GO STUDY" in a bold, italicized, sans-serif font. The words "GO GO" are on the top line and "STUDY" is on the bottom line. The text is white and is set against a dark gray square background.

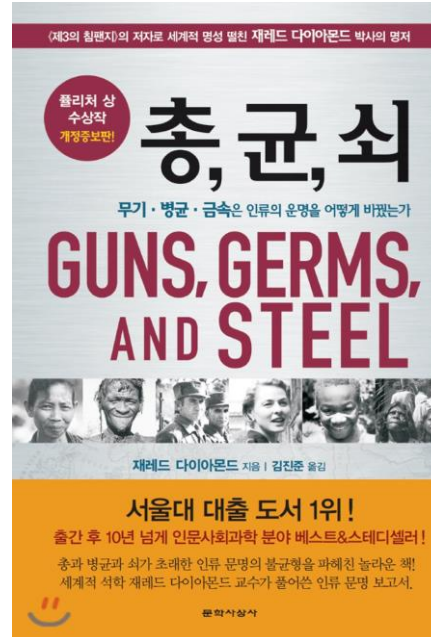
***GO GO
STUDY***

01. 프로젝트 소개

IDEA



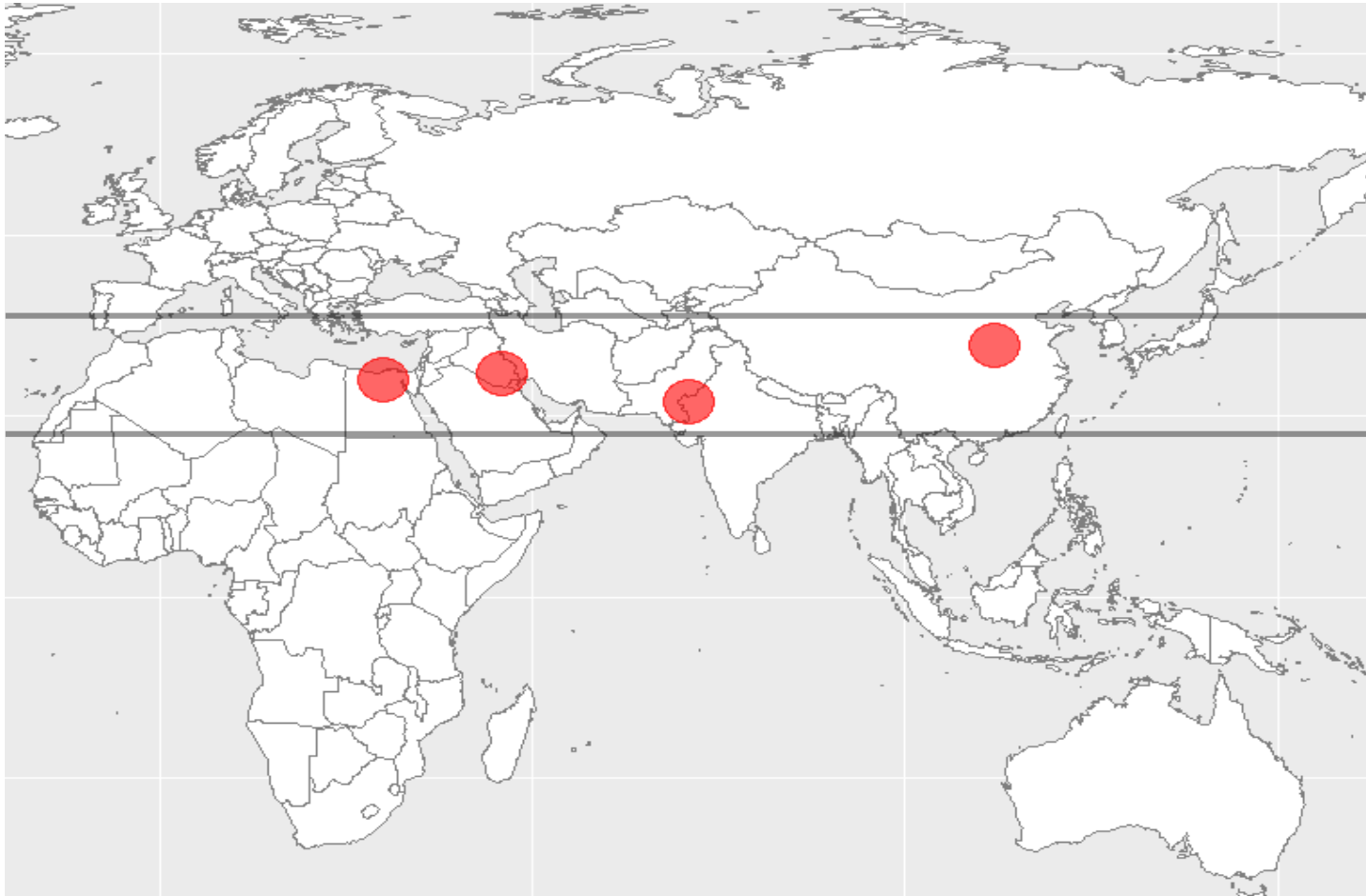
재레드 다이아몬드 (1937 ~)



문명이 발생하기 필요한 위도와 경도가 있다.

그렇다면 국가가 가지는 다양한 지표와
위도와 경도는 상관관계를 가지지 않을까?

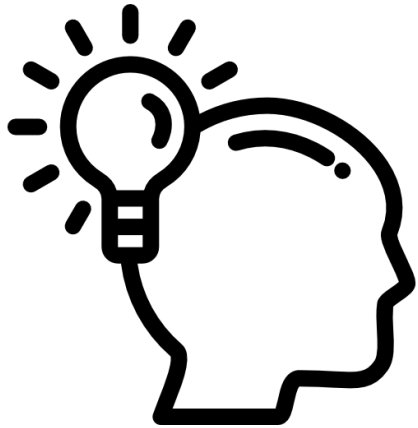
IDEA



실제로 4대 문명 발상지는
비슷한 위도를 공유하고 있다.

황하문명 (35/112)
이집트문명 (29/30)
인더스문명 (27/71)
메소포타미아문명 (31/46)

IDEA



문맹율, 범죄발생율, GDP, 문맹률 등 다양한 사회지표를
선정한 뒤 표준화하여 Scoring을 한다.

그 다음 Score를 공간 좌표에 Mapping하면 Score와 위도, 경도와의
상관관계를 파악할 수 있지 않을까??

프로젝트 진행과정

주제 설정

자료 탐색

자료 전처리

분석

시각화

의미 도출

02. 작업 환경 및 Dataset

작업 환경

R Studio



- ✓ 데이터 전처리
- ✓ 상관관계분석
- ✓ 데이터 시각화

Python



- ✓ numpy, pandas 를 이용한 데이터 전처리

Markdown

- ✓ 스크립트를 도시하여 보고서 작성

사용한 R 패키지

ggplot, ggmap, maps

- ✓ 세계지도에 데이터 mapping

caret

- ✓ featureplot을 이용한 상관관계 도시

DMwR

- ✓ knnimputation을 NA값 수정

Dataset

182개 국가의 위도와 경도를 조사하여

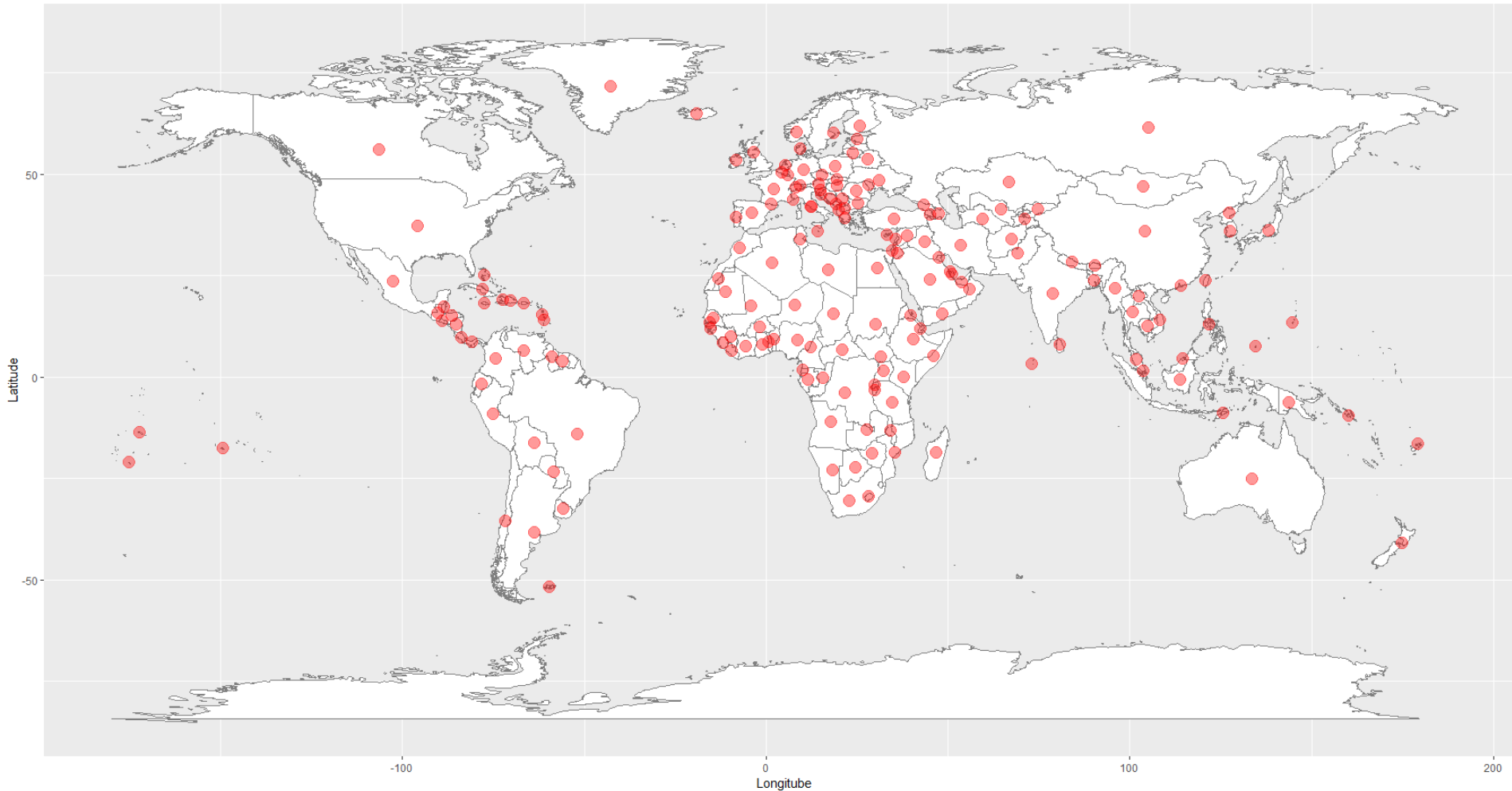
Dataset의 뼈대를 생성하고

다양한 지표들을 국가에 추가

✓ 사용한 지표: 인구
1960' GDP
2017' GDP
인구 10만명당 자살인구
부패인식지수
인구 10만명당 유아 사망 수
인구 10만명당 범죄 발생 수
문해율

기대수명
뉴욕 대비 소비자 물가지수
뉴욕 대비 주거지 임대료 지수
뉴욕 대비 식료품가격 지수
뉴욕 대비 소비자 구매력 지수
농경지 비율

Dataset

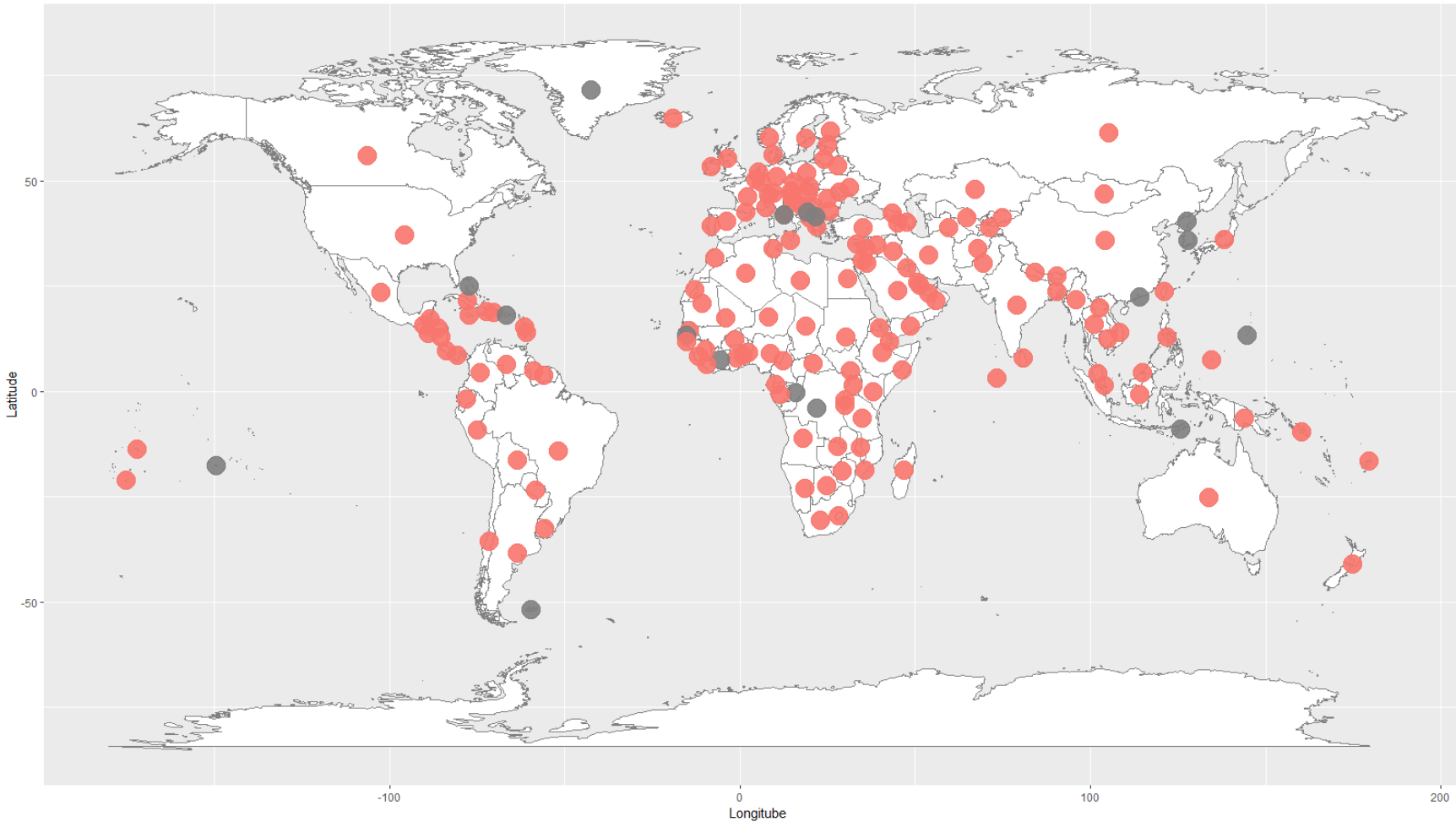


182개 국가의 좌표

Dataset

결측 값 처리 :

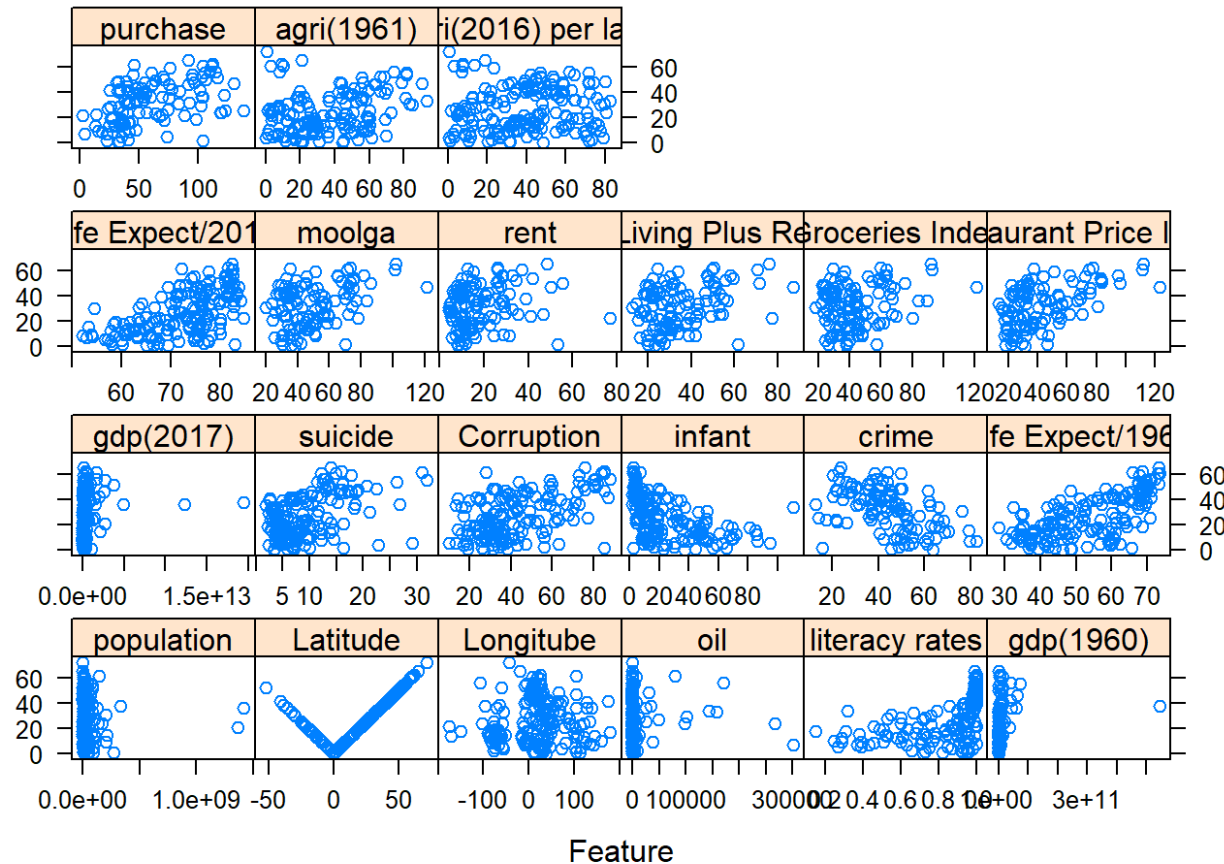
knnimputation를 통해
비슷한 위도와 경도를 가진
국가의 지표들의 평균값으로 대체



검정색으로 표시된
인구 10만명당 유아사망수 자료의 결측값

03. 프로젝트 내용

featureplot을 통해 모형 파악



X축에 지표, Y축에 위도를 넣고

featureplot 도시 결과

상관관계가 느껴지는 것도 있지만

대부분의 지표가 없는 것으로 보임

Scoring

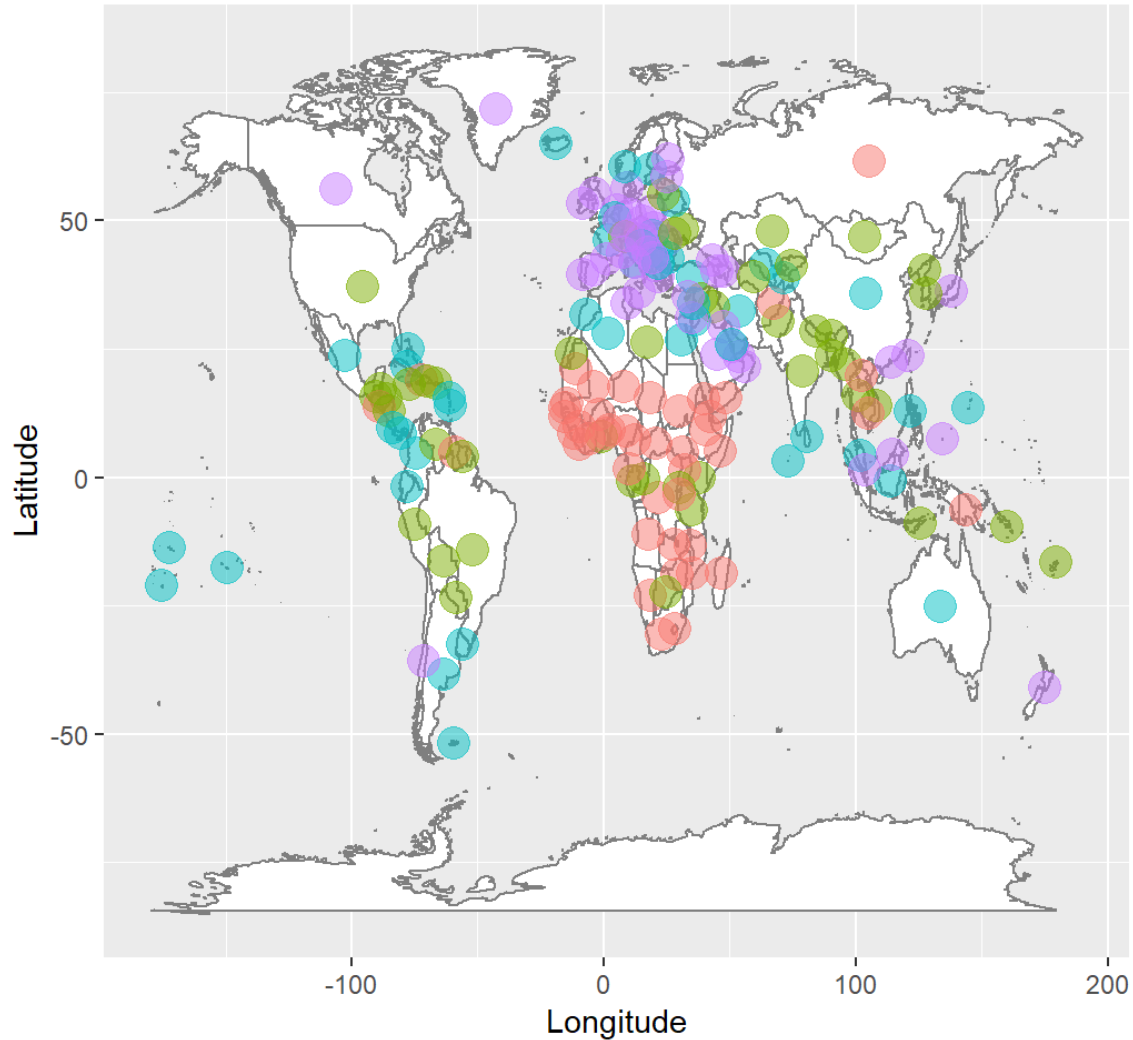
지표들을 표준화한 뒤
지표들 중 위도, 경도와 상관관계가
높다고 판단되는 지표를 선정함

긍정적인 지표는 +, 부정적인 지표는 - 하여 Scoring 진행

✓ 사용한 지표: 문해율 (+)
부패인식지수 (+)
기대수명 (+)

뉴욕 대비 식료품가격 지수 (-)
2017' 1인당 GDP (-)
인구 10만명당 유아사망수 (-)
인구 10만명당 자살인구 (-)
인구 10만명당 범죄발생수 (-)

Scoring



factor(new\$score2)

- 1
- 2
- 3
- 4

Score의 fivenum을 기준으로
범주형 자료로 변경

세계지도에 mapping

Score와 경도의 상관관계

```
##  
## Pearson's product-moment correlation  
##  
## data: new$Longitude and new$Score  
## t = 0.26315, df = 180, p-value = 0.7927  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.1262054 0.1645961  
## sample estimates:  
## cor  
## 0.01961009
```

적합한 Score와 경도의 상관계수는 0.019

상관관계를 가진다고 볼 수 없다.

Score와 위도의 상관관계

```
##  
## Pearson's product-moment correlation  
##  
## data: abs(new$Latitude) and new$Score  
## t = 8.4808, df = 180, p-value = 7.949e-15  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.4216361 0.6307552  
## sample estimates:  
## cor  
## 0.534322
```

적합한 Score와 위도의 상관계수는 0.534

약한 상관관계를 보인다.

결론

위도, 경도를 따지기에 상관성은 있으나 그 정도에 있어서 약하다.

특히 경도에 있어서는 상관관계를 찾을 수 없다.

국가별 상황에 영향을 끼치는 요소는 종교나 문화, 혹은 경제정책 등으로 보인다.

And... Q & A

Dataset 자료출처

범죄지수 / 2018 mid-year (https://www.numbeo.com/crime/rankings_by_country.jsp?title=2018-mid)

부패인식지수 (https://en.wikipedia.org/wiki/Corruption_Perceptions_Index)

위도/경도 (https://developers.google.com/public-data/docs/canonical/countries_csv)

석유매장량 / U.S. EIA (start of 2017) (https://en.wikipedia.org/wiki/List_of_countries_by_proven_oil_reserves)

문맹률 (<https://www.worldatlas.com/articles/the-highest-literacy-rates-in-the-world.html>)

GDP (<https://data.worldbank.org/indicator/NY.GDP.MKTP.CD>)

농지비율 (<https://data.worldbank.org/indicator/AG.LND.AGRI.ZS>)

기대수명 (<https://data.worldbank.org/indicator/SP.DYN.LE00.IN>)

100,000명당 영아 사망률 (https://en.wikipedia.org/wiki/List_of_countries_by_infant_and_under-five_mortality_rates)

100,000명당 자살률(https://es.wikipedia.org/wiki/Anexo:Pa%C3%ADses_por_tasa_de_suicidio)

끝!

감사합니다.