

# TEST\_05

- Guide
  - 1. Data Wrangling Problem
  - 2. Visualization Problem(Using ggplot2)
- 

- 필요 패키지 : dplyr , ggplot2 , tidyr
  - 필요 데이터 : co2.csv
- 

## Guide

- 시험시간 : 13:30 ~ 15:00 (총90분)
- R markdown을 이용하여 생성한 html파일 모두 제출
- Dropbox에 제출시 html 파일과 rmd 파일 모두 업로드 해야함
  - 공유된 폴더안에 시험날짜 폴더를 만들고 (예 : 20211125 )
  - 관련 데이터는 Data폴더 를 만들어 제출
  - html,rmd파일은 Rmd폴더 에 학번\_이름으로 작성 제출 (예 : 202101234\_홍길동)
- 문제에서 주어진 결과와 같게 작성
  - ex) 작성방식 : 1.1번 문제라고 적고 R Chunk를 생성하여 그 안에 해당 code 기입
- 업로드 및 수정시간이 모두 기록 됨에 유의
- 코드공유 적발시 0점 처리
- 문제에서 출력한대로 출력(전체데이터 출력은 지양)
- 위의 유의사항을 어길 시 감점
- 시험시간을 넘겨서 제출하면 0점 처리



## 0. Packages

```
library(dplyr)
library(tidyr)
library(ggplot2)
```

# 1. Data Wrangling Problem

## 1.1 co2 를 불러오고 속성을 확인하시오.[10점]

- head()로 출력 한 뒤 str()을 확인할 것

```
##      Year      Country Total Solid.Fuel Liquid.Fuel Gas.Fuel Cement G
as.Flaring
## 1 1751 UNITED KINGDOM 2552      2552      0      0      0
0
## 2 1752 UNITED KINGDOM 2553      2553      0      0      0
0
## 3 1753 UNITED KINGDOM 2553      2553      0      0      0
0
## 4 1754 UNITED KINGDOM 2554      2554      0      0      0
0
## 5 1755 UNITED KINGDOM 2555      2555      0      0      0
0
## 6 1756 UNITED KINGDOM 2731      2731      0      0      0
0
##      Per.Capita Bunker.fuels..Not.in.Total.
## 1      0      0
## 2      0      0
## 3      0      0
## 4      0      0
## 5      0      0
## 6      0      0
```

```
## 'data.frame':    17232 obs. of  10 variables:
## $ Year      : int  1751 1752 1753 1754 1755 1756 1
757 1758 1759 1760 ...
## $ Country   : chr   "UNITED KINGDOM" "UNITED KINGDO
M" "UNITED KINGDOM" "UNITED KINGDOM" ...
## $ Total     : int  2552 2553 2553 2554 2555 2731 2
732 2733 2734 2734 ...
## $ Solid.Fuel : int  2552 2553 2553 2554 2555 2731 2
732 2733 2734 2734 ...
## $ Liquid.Fuel : int  0 0 0 0 0 0 0 0 0 0 0 ...
## $ Gas.Fuel   : int  0 0 0 0 0 0 0 0 0 0 0 ...
## $ Cement     : int  0 0 0 0 0 0 0 0 0 0 0 ...
## $ Gas.Flaring : int  0 0 0 0 0 0 0 0 0 0 0 ...
## $ Per.Capita : num  0 0 0 0 0 0 0 0 0 0 0 ...
## $ Bunker.fuels..Not.in.Total.: int  0 0 0 0 0 0 0 0 0 0 0 ...
```

**1.2 다음과 같이 years 변수를 생성 후 str()을 통해서 확인하시오. [15점]**

- 1960년 이후(1960년(포함) 이후) 데이터만 filtering 하시오.[5점]
- years 열을 다음과 같이 생성하시오.
  - 힌트: case\_when 함수와 mutate 함수를 활용하여 years 열을 생성[5점]
    - Year를 토대로 1960s, 1970s, 1980s, 1990s, 2000s, 2010s 와 같이 데이터를 변경하시오
    - ex) Year가 1960~1969에 해당되는 경우 years 변수는 1960s로 생성
  - 최종 생성된 years 변수를 factor를 활용하여 변경하시오[5점]
    - levels는 1960s, 1970s, ..., 2010s가 되도록 하시오.
- str()을 통해서 확인하시오

```
## 'data.frame':    10880 obs. of  11 variables:
## $ Year          : int  1960 1960 1960 1960 1960 1960 1
960 1960 1960 1960 ...
## $ Country       : chr  "AFGHANISTAN" "ALBANIA" "ALGERI
A" "ANGOLA" ...
## $ Total         : int  113 552 1680 150 10 13312 24053
8405 112 157 ...
## $ Solid.Fuel    : int  35 89 343 34 0 1066 16083 4926
0 0 ...
## $ Liquid.Fuel   : int  74 430 1193 95 10 11242 7590 22
97 112 157 ...
## $ Gas.Fuel      : int  0 23 0 0 0 645 0 797 0 0 ...
## $ Cement       : int  5 10 144 22 0 359 380 385 0 0 .
..
## $ Gas.Flaring   : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Per.Capita    : num  0.01 0.34 0.16 0.03 0.18 0.65 2
.34 1.19 1.02 0.97 ...
## $ Bunker.fuels..Not.in.Total.: int  0 0 438 41 0 211 1219 1 468 605
...
## $ years         : Factor w/ 6 levels "1960s","1970s",.
.: 1 1 1 1 1 1 1 1 1 1 ...
```

### 1.3 years별 총 co2 배출량이 많은 국가를 구하시오. [15점]

- years, 국가 별 통계량 [5점]
  - 힌트: summarise 활용 시 Total 변수의 평균을 활용하시오
- 연도별, 배출 평균을 정렬하시오[5점]
  - 힌트: 위에서(1.2 번) factor가 제대로 적용되어야 정렬이 됨
- 연도별 상위 3개국만 출력하시오[5점]

```
## # A tibble: 18 x 3
## # Groups:   years [6]
##   years Country          emission
##   <fct> <chr>              <dbl>
## 1 1960s UNITED STATES OF AMERICA  917323.
## 2 1960s RUSSIAN FEDERATION       501016.
## 3 1960s FEDERAL REPUBLIC OF GERMANY 172014.
## 4 1970s UNITED STATES OF AMERICA 1259168
## 5 1970s RUSSIAN FEDERATION       794073.
## 6 1970s CHINA (MAINLAND)         303941.
## 7 1980s UNITED STATES OF AMERICA 1251891.
## 8 1980s RUSSIAN FEDERATION       1021060.
## 9 1980s CHINA (MAINLAND)         518248.
## 10 1990s UNITED STATES OF AMERICA 1399952
## 11 1990s CHINA (MAINLAND)         832580.
## 12 1990s RUSSIAN FEDERATION       571814.
## 13 2000s UNITED STATES OF AMERICA 1541212.
## 14 2000s CHINA (MAINLAND)         1514265
## 15 2000s RUSSIAN FEDERATION       439572.
## 16 2010s CHINA (MAINLAND)         2677289.
## 17 2010s UNITED STATES OF AMERICA 1429948.
## 18 2010s INDIA                   537393
```

## 1.4 다음과 같이 long format으로 변형하시오[10점]

- Solid.Fuel, Liquid.Fuel, Gas.Fuel, Cement, Gas.Flaring을 long format으로 변형하시오 [5점]
  - key 변수명은 factor로 value 변수명은 emission으로 하시오 [5점]
- head()로 다음과 같이 일부만 출력할 것

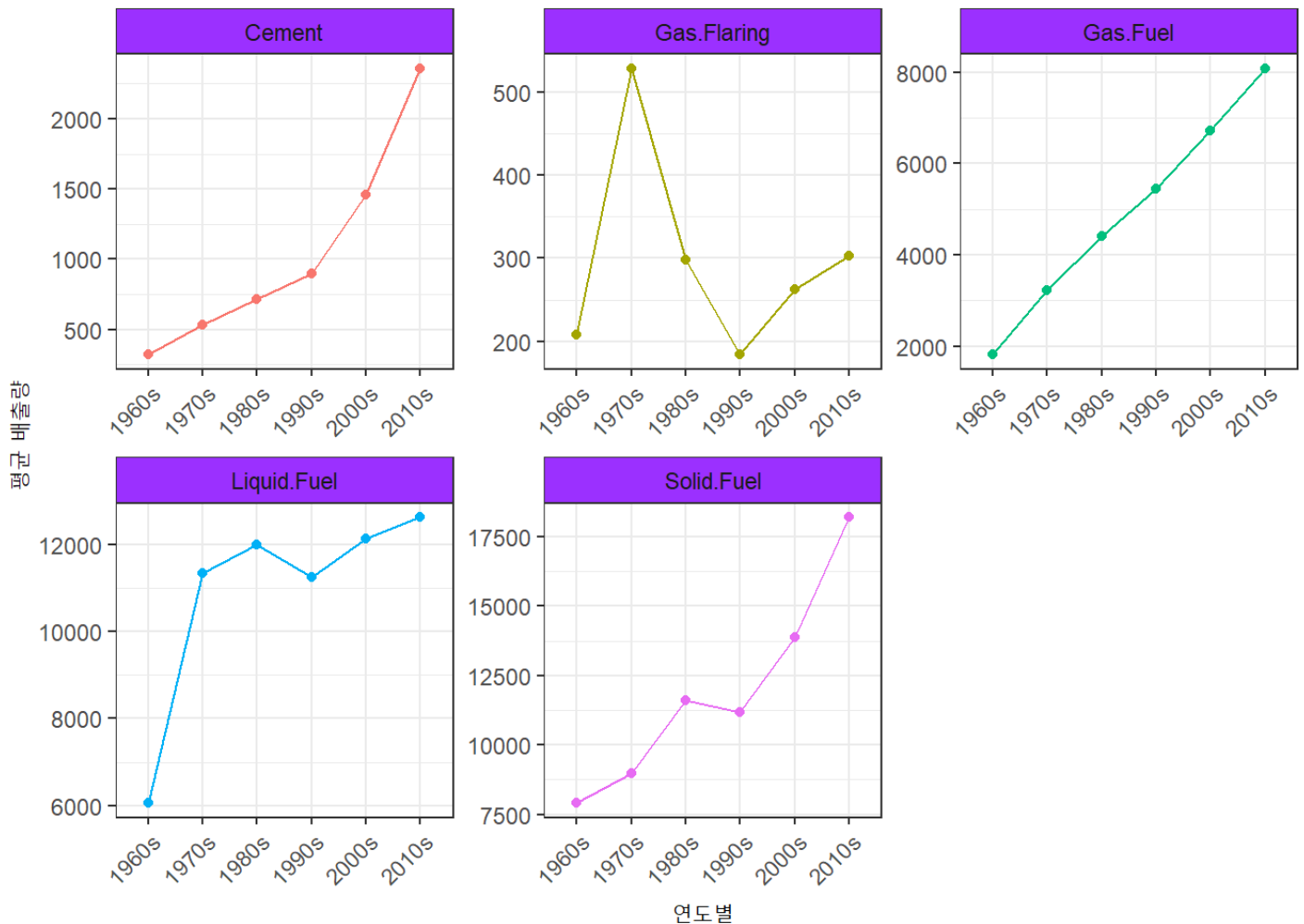
```
##      Year                Country Total Per.Capita Bunker.fuels..Not.in.Total
. years
## 1 1960      AFGHANISTAN    113      0.01
0 1960s
## 2 1960      ALBANIA      552      0.34
0 1960s
## 3 1960      ALGERIA     1680      0.16                                43
8 1960s
## 4 1960      ANGOLA       150      0.03                                4
1 1960s
## 5 1960 ANTIGUA & BARBUDA    10      0.18
0 1960s
## 6 1960      ARGENTINA  13312      0.65                                21
1 1960s
##      factor emission
## 1 Solid.Fuel      35
## 2 Solid.Fuel      89
## 3 Solid.Fuel     343
## 4 Solid.Fuel      34
## 5 Solid.Fuel       0
## 6 Solid.Fuel    1066
```

## 2. Visualization Problem(Using ggplot2)

### 2.1 long format데이터를 이용해 다음과 같이 그리시오 [25점]

- years, factor 별 emission 평균을 구하시오[5점]
- geom\_point와 geom\_line을 활용하여 그림을 그리시오.[5점]
- strip 배경 색을 “purple1”으로 바꾸시오.[5점]
- factor 별로 그림을 따로 그리시오.[5점]
- x축 제목과 y축 제목을 바꾸고 x축 label은 45도 기울이시오.[5점]

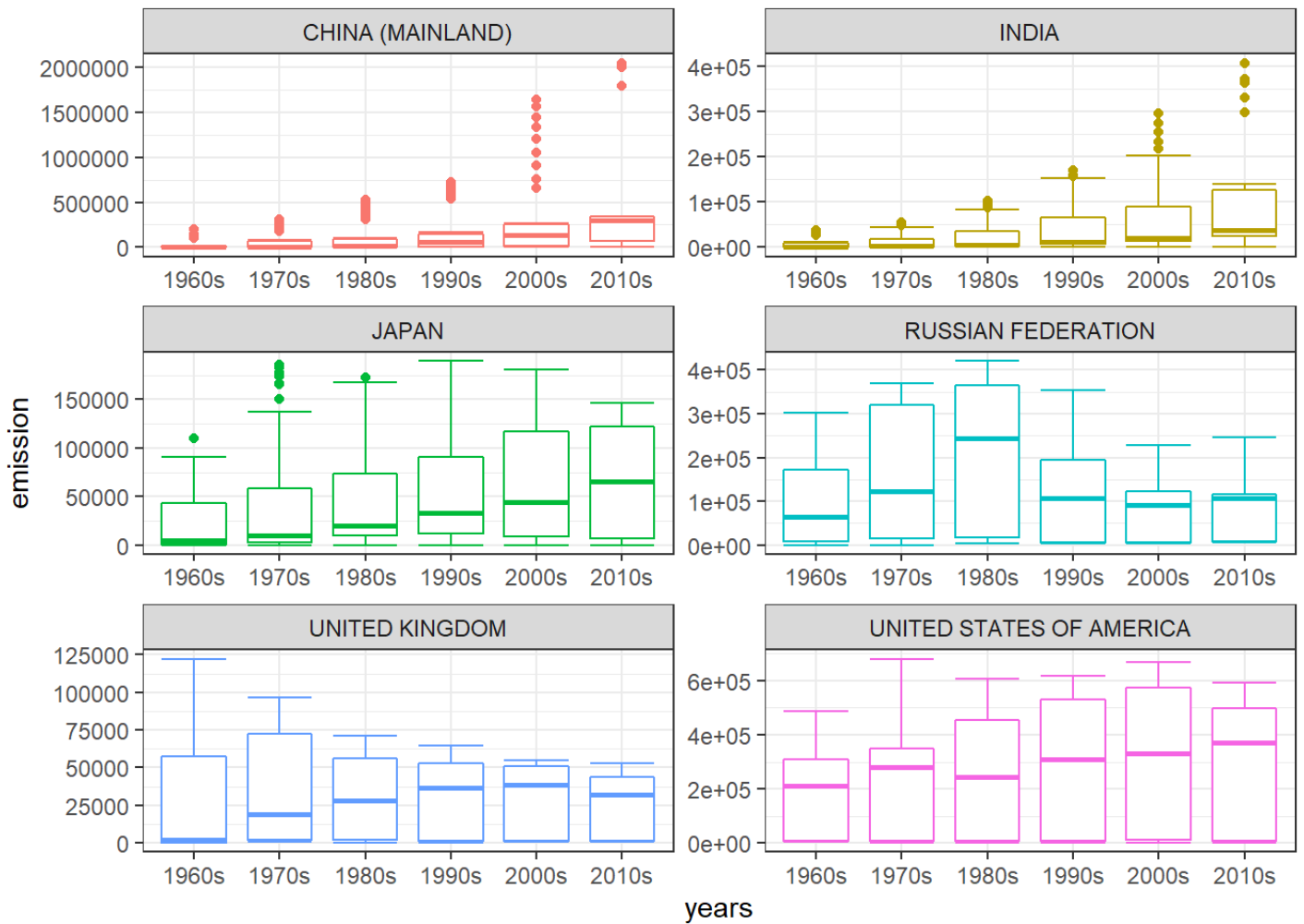
```
## # A tibble: 30 x 3
## # Groups:   years [6]
##   years factor      value
##   <fct> <chr>      <dbl>
## 1 1960s Cement        319.
## 2 1960s Gas.Flaring   208.
## 3 1960s Gas.Fuel    1816.
## 4 1960s Liquid.Fuel  6052.
## 5 1960s Solid.Fuel   7900.
## 6 1970s Cement        529.
## 7 1970s Gas.Flaring   529.
## 8 1970s Gas.Fuel    3225.
## 9 1970s Liquid.Fuel 11334.
## 10 1970s Solid.Fuel   8974.
## # ... with 20 more rows
```



## 2.2 주요 국가별 배출량을 그리시오 [25점]

- 기본그림[5점]
- 다음과 같이 특정 국가만 filtering 하시오.[5점]

- boxplot을 다음과 같은 모양으로 그리시오.[5점]
- 국가별로 따로 그리되 그림과 같이 나타나게 그리시오.[5점]
- legend가 표시되지 않게 하시오.[5점]



© 2021 Advice, All Rights Reserved

No part of this contents may be reproduced, copied, modified or adapted, without the prior written consent of the author, unless otherwise indicated for stand-alone materials.