

HW1-1 공공데이터포털 데이터 다뤄보기

강현승

2022-10-17

DEBUG = F # 보고용

문제 1

```

# install.packages('xml2')
library(xml2)

serviceKey = 'MlB2k1cGG%2BUDm0OzU5rF8b8ESs7u1y2mmBmt6E%2BOeHdAg5mO%2F7bx1dJDMPqCA2
LgAe2E2H9jaodeKzU8IbFr4g%3D%3D'
numOfRows = 10
pageNo = ''
chrstnType = ''
filterValues = ''
dataType = 'xml'
url_xml = paste0(
  'https://apis.data.go.kr/3740000/suwonEvChrstn/getdatalist?serviceKey=' ,
  serviceKey,
  '&numOfRows=',
  numOfRows,
  '&pageNo=',
  pageNo,
  '&sortKey=',
  chrstnType,
  '&filterKey=',
  chrstnType,
  '&filterValues=',
  filterValues,
  '&type=',
  dataType
)

# urls 링크를 xml로 읽어들이기.
# read_xml(): xml 파일 읽어들이기.
raw_xml <- read_xml(url_xml)

# occrrnc0t. dataVall 변수 모두 찾아서 저장하기
# xml_find_all(): xml 파일에서 원하는 '태그'를 찾아 저장.

xml_1 <- xml_find_all(raw_xml, '//chrstnNm')
xml_2 <- xml_find_all(raw_xml, '//latitude')
xml_3 <- xml_find_all(raw_xml, '//longitude')
xml_4 <- xml_find_all(raw_xml, '//charger_status')

# text로 읽어 저장하기
# xml_text( :) 텍스트 추출.
chrstnNm <- xml_text(xml_1)
latitude <- xml_text(xml_2)
longitude <- xml_text(xml_3)
charger_status <- xml_text(xml_4)

```

문제 2

```
df = data.frame(
  chrstnNm = chrstnNm,
  latitude = latitude,
  longitude = longitude,
  charger_status = charger_status
)
head(df)
```

```
##               chrstnNm latitude longitude charger_status
## 1 SK엔크린 경진셀프주유소 37.308128 126.996455          2
## 2          롯데마트 천천점 37.295945 126.982536          2
## 3          롯데마트 천천점 37.295945 126.982536          2
## 4          수원종합운동장 37.299335 127.007841          2
## 5          수원종합운동장 37.299335 127.007841          2
## 6          수원종합운동장 37.299335 127.007841          2
```

문제 3

```
str(df)
```

```
## 'data.frame':    10 obs. of  4 variables:
## $ chrstnNm      : chr  "SK엔크린 경진셀프주유소" "롯데마트 천천점" "롯데마트 천천점" "수원종합
운동장" ...
## $ latitude      : chr  "37.308128" "37.295945" "37.295945" "37.299335" ...
## $ longitude     : chr  "126.996455" "126.982536" "126.982536" "127.007841" ...
## $ charger_status: chr  "2" "2" "2" "2" ...
```

```
# 문제점) 위도(latitude)와 경도(longitude)를 character형에서 numeric으로 변경하여야.
df$latitude = as.numeric(df$latitude)
df$longitude = as.numeric(df$longitude)
str(df) # df가 고쳐졌는지 확인
```

```
## 'data.frame':    10 obs. of  4 variables:
## $ chrstnNm      : chr  "SK엔크린 경진셀프주유소" "롯데마트 천천점" "롯데마트 천천점" "수원종합
운동장" ...
## $ latitude      : num  37.3 37.3 37.3 37.3 37.3 ...
## $ longitude     : num  127 127 127 127 127 ...
## $ charger_status: chr  "2" "2" "2" "2" ...
```

문제 4

```
df = data.frame()
totalCount = as.numeric(xml_text(xml_find_all(raw_xml, '//totalCount'), trim = T))
totalPage = totalCount %/% 10 - 1
if (totalPage > 30) {
  totalPage = 30
}
```

```

for (pageNo in 0:totalPage) {
  if (sum(ls() == 'serviceKey') == 0) {
    print('NO serviceKey. Arret!')
    break
  }
  if (DEBUG) {
    if (pageNo > 50) {
      break
    }
    print(paste0(c(
      pageNo,
      ' / ',
      totalPage,
      ' [',
      rep('=', times = as.integer(pageNo / totalPage * 20)),
      rep(' ', times = 20 - as.integer(pageNo / totalPage * 20)),
      ']'
    ), collapse = ''))
  }
  numOfRows = 10
  dataType = 'xml'
  url_xml = paste0(
    'https://apis.data.go.kr/3740000/suwonEvChrstn/getdatalist?serviceKey=' ,
    serviceKey,
    '&pageNo=',
    pageNo,
    '&type=',
    dataType,
    '&numOfRows=',
    numOfRows,
    '&sortKey=&filterKey=&filterValues='
  )
  raw_xml <- read_xml(url_xml)
  xml_1 <- xml_find_all(raw_xml, '//chrstnNm')
  xml_2 <- xml_find_all(raw_xml, '//latitude')
  xml_3 <- xml_find_all(raw_xml, '//longitude')
  xml_4 <- xml_find_all(raw_xml, '//charger_status')
  chrstnNm <- xml_text(xml_1)
  latitude <- as.numeric(xml_text(xml_2))
  longitude <- as.numeric(xml_text(xml_3))
  charger_status <- xml_text(xml_4)
  df = rbind(
    df,
    list(
      chrstnNm = chrstnNm,
      latitude = latitude,
      longitude = longitude,
      charger_status = charger_status
    )
  )
}
str(df)

```

```
## 'data.frame':      310 obs. of  4 variables:
## $ chrstnNm      : chr  "SK엔크린 경진셀프주유소" "롯데마트 천천점" "롯데마트 천천점" "수원종합
운동장" ...
## $ latitude      : num  37.3 37.3 37.3 37.3 37.3 ...
## $ longitude     : num  127 127 127 127 127 ...
## $ charger_status: chr  "2" "2" "2" "2" ...
```

```
head(df)
```

```
##               chrstnNm latitude longitude charger_status
## 1 SK엔크린 경진셀프주유소 37.30813  126.9965             2
## 2               롯데마트 천천점 37.29595  126.9825             2
## 3               롯데마트 천천점 37.29595  126.9825             2
## 4               수원종합운동장 37.29933  127.0078             2
## 5               수원종합운동장 37.29933  127.0078             2
## 6               수원종합운동장 37.29933  127.0078             2
```

문제 5

```

fetch_xml_to_df = function(pageNo) {
  numofRows = 10
  dataType = 'xml'
  url_xml = paste0(
    'https://apis.data.go.kr/3740000/suwonEvChrstn/getdatalist?serviceKey=' ,
    serviceKey,
    '&pageNo=',
    pageNo,
    '&type=',
    dataType,
    '&numOfRows=',
    numofRows,
    '&sortKey=&filterKey=&filterValues='
  )
  raw_xml <- read_xml(url_xml)
  xml_1 <- xml_find_all(raw_xml, '//chrstnNm')
  xml_2 <- xml_find_all(raw_xml, '//latitude')
  xml_3 <- xml_find_all(raw_xml, '//longitude')
  xml_4 <- xml_find_all(raw_xml, '//charger_status')
  chrstnNm <- xml_text(xml_1)
  latitude <- as.numeric(xml_text(xml_2))
  longitude <- as.numeric(xml_text(xml_3))
  charger_status <- xml_text(xml_4)
  return(
    data.frame(
      chrstnNm = chrstnNm,
      latitude = latitude,
      longitude = longitude,
      charger_status = charger_status
    )
  )
}

str(fetch_xml_to_df(0))

```

```

## 'data.frame':   10 obs. of  4 variables:
## $ chrstnNm      : chr  "SK엔크린 경진셀프주유소" "롯데마트 천천점" "롯데마트 천천점" "수원종합
운동장" ...
## $ latitude      : num  37.3 37.3 37.3 37.3 37.3 ...
## $ longitude     : num  127 127 127 127 127 ...
## $ charger_status: chr  "2" "2" "2" "2" ...

```

```
head(fetch_xml_to_df(0))
```

##		chrstnNm	latitude	longitude	charger_status
## 1	SK엔크린 경진셀프주유소	37.30813	126.9965		2
## 2	롯데마트 천천점	37.29595	126.9825		2
## 3	롯데마트 천천점	37.29595	126.9825		2
## 4	수원종합운동장	37.29933	127.0078		2
## 5	수원종합운동장	37.29933	127.0078		2
## 6	수원종합운동장	37.29933	127.0078		2

문제 6

```
#install.packages('rjson')
#install.packages('httr')
library(rjson)
library(httr)

fetch_json_to_df = function(pageNo) {
  numOfRows = 10
  dataType = 'json'
  url_json = paste0(
    'https://apis.data.go.kr/3740000/suwonEvChrstn/getdatalist?serviceKey=' ,
    serviceKey,
    '&pageNo=',
    pageNo,
    '&type=',
    dataType,
    '&numOfRows=',
    numOfRows,
    '&sortKey=&filterKey=&filterValues='
  )
  dfQueue = data.frame()
  deserializedItems = fromJSON(content(GET(url_json), 'text'))[['items']]
  for (i in 1:length(deserializedItems)) {
    dfQueue = rbind(
      dfQueue,
      list(
        chrstnNm = deserializedItems[[i]]$chrstnNm,
        latitude = as.numeric(deserializedItems[[i]]$latitude),
        longitude = as.numeric(deserializedItems[[i]]$longitude),
        charger_status = deserializedItems[[i]]$charger_status
      )
    )
  }
  return(dfQueue)
}

df = data.frame()
for (pageNo in 0:13) {
  df = rbind(df, fetch_json_to_df(pageNo))
}
str(df)
```

```
## 'data.frame':    140 obs. of  4 variables:
## $ chrstnNm      : chr  "SK엔크린 경진셀프주유소" "롯데마트 천천점" "롯데마트 천천점" "수원종합운동장" ...
## $ latitude      : num  37.3 37.3 37.3 37.3 37.3 ...
## $ longitude     : num  127 127 127 127 127 ...
## $ charger_status: chr  "2" "2" "2" "2" ...
```

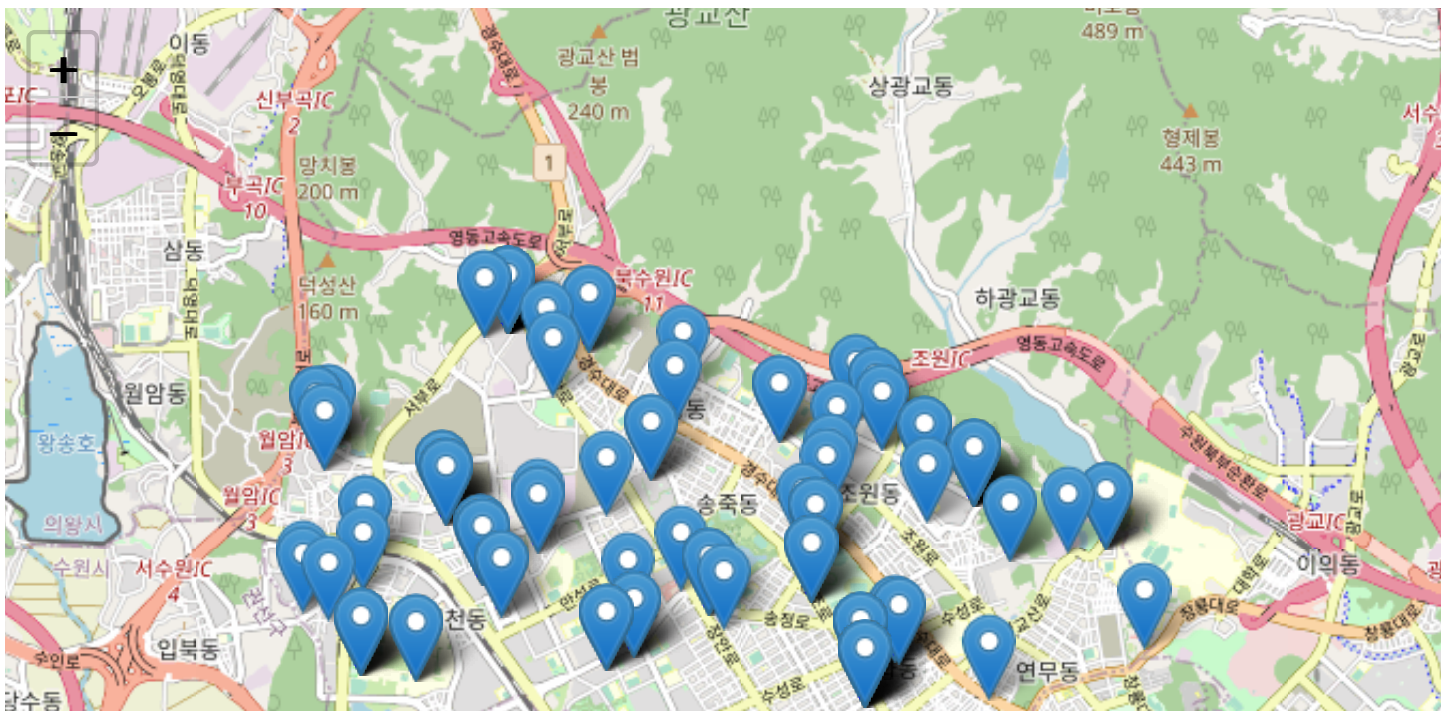
```
head(df)
```

```
##           chrstnNm latitude longitude charger_status
## 1 SK엔크린 경진셀프주유소 37.30813  126.9965          2
## 2 롯데마트 천천점 37.29595  126.9825          2
## 3 롯데마트 천천점 37.29595  126.9825          2
## 4 수원종합운동장 37.29933  127.0078          2
## 5 수원종합운동장 37.29933  127.0078          2
## 6 수원종합운동장 37.29933  127.0078          2
```

문제 7

```
#install.packages('leaflet')
library(leaflet)

my_map = leaflet()
my_map_1 = addTiles(my_map)
my_map_2 = addMarkers(
  my_map_1,
  lng = df$longitude,
  lat = df$latitude,
  popup = df$chrstnNm
)
my_map_2
```





문제 8

기초자치단체 별 음식물쓰레기 종합배출내역 활용

```
waste_api_key = 'M1B2k1cGG%2BUDm0OzU5rF8b8ESS7u1y2mmBmt6E%2BOeHdAg5mO%2F7bx1dJDMPq
CA2LgAe2E2H9jaodeKzU8IbFr4g%3D%3D'

waste_domestic_code_fetch_url = paste0(
  'https://apis.data.go.kr/B552584/RfidFoodWasteServiceNew/getCityList?serviceKey=
',
  waste_api_key,
  '&type=xml&page=1&rowNum=10000'
)
raw_waste_domestic_code_xml <-
  read_xml(waste_domestic_code_fetch_url)
xml_1 <- xml_find_all(raw_waste_domestic_code_xml, '//cityCode')
xml_2 <- xml_find_all(raw_waste_domestic_code_xml, '//citySidoName')
xml_3 <- xml_find_all(raw_waste_domestic_code_xml, '//citySggName')
cityCode <- xml_text(xml_1)
citySidoName <- xml_text(xml_2)
citySggName <- xml_text(xml_3)

waste_df = data.frame(cityCode = cityCode,
                      citySidoName = citySidoName,
                      citySggName = citySggName)

waste_quantities = c()
for (i in 1:length(cityCode)) {
  if (DEBUG) {
    print(paste0(c(
      i,
      ' / ',
      length(cityCode),
      ' [',
      rep('=', times = as.integer(i / length(cityCode) * 20)),
      rep(' ', times = 20 - as.integer(i / length(cityCode) * 20)),
      ']'
    ), collapse = ''))
  }
  disYear = '2022'
  disMonth = '08'
  waste_quantity_fetch_url = paste0(
    'https://apis.data.go.kr/B552584/RfidFoodWasteServiceNew/getCityDateList?servi
ceKey=' ,
```

```

waste_api_key,
'&type=xml&disYear=',
disYear,
'&disMonth=',
disMonth,
'&cityCode=',
cityCode[i],
'&page=1&rowNum=31'
)
raw_waste_quantity_xml <- read_xml(waste_quantity_fetch_url)
xml_1 <- xml_find_all(raw_waste_quantity_xml, '//disQuantity')
waste_quantities = append(waste_quantities, sum(as.numeric(xml_text(xml_1))))
}

waste_df = cbind(waste_df, data.frame(disQuantity = waste_quantities))
str(waste_df)

```

```

## 'data.frame':    178 obs. of  4 variables:
## $ cityCode      : chr  "W01" "W02" "W03" "W04" ...
## $ citySidoName: chr  "서울특별시" "서울특별시" "서울특별시" "서울특별시" ...
## $ citySggName  : chr  "종로구" "중구" "용산구" "성동구" ...
## $ disQuantity  : num  1.24e+08 1.92e+08 5.71e+07 5.12e+08 3.65e+08 ...

```

```
head(waste_df)
```

```

##   cityCode citySidoName citySggName disQuantity
## 1      W01   서울특별시   종로구    124231090
## 2      W02   서울특별시   중구      191763050
## 3      W03   서울특별시   용산구    57088100
## 4      W04   서울특별시   성동구    512324050
## 5      W05   서울특별시   광진구    365494460
## 6      W06   서울특별시   동대문구  530541655

```

문제 9

```

# 2022/08 기초자치단체 별 배출량 평균
mean(waste_df$disQuantity)

```

```
## [1] 471328691
```

```

# 2022/08 한 달 간 배출량이 가장 많은 기초자치단체
waste_df[waste_df$disQuantity == max(waste_df$disQuantity),]

```

```

##   cityCode citySidoName citySggName disQuantity
## 95      W2Y     경기도   용인시   3555578299

```

```
# 2022/08 한 달 간 배출량이 가장 적은 기초자치단체, 0을 제외하고 = RFID를 통한 음식물쓰레기 집계를 하지 않는 기초자치단체를 제외
```

```
waste_df[waste_df$disQuantity == min(waste_df$disQuantity[waste_df$disQuantity != 0]),]
```

```
##      cityCode citySidoName citySggName disQuantity
## 77      W27      경기도 성남시수정구      333200
```

```
# 광역자치단체 별 합계 내역 출력
```

```
disQuantityBySido = data.frame()
```

```
for (sidoName in levels(factor(waste_df$citySidoName))) {
```

```
  disQuantityBySido = rbind(disQuantityBySido,
                             list(
                               citySidoName = sidoName,
                               disQuantity = sum(waste_df[waste_df$citySidoName ==
sidoName, ]$disQuantity)
                             ))
}
```

```
disQuantityBySido
```

```
##      citySidoName disQuantity
## 1      강원도      3207728257
## 2      경기도      21706549842
## 3      경상남도      3925661614
## 4      경상북도      4779333457
## 5      광주광역시      4070512760
## 6      대구광역시      4455239545
## 7      대전광역시      1158779800
## 8      부산광역시      7472361008
## 9      서울특별시      14726506508
## 10     울산광역시      1633917030
## 11     인천광역시      5813083368
## 12     전라남도      1142179990
## 13     전라북도      4878849290
## 14     제주특별자치도      3017661515
## 15     충청남도      194523700
## 16     충청북도      1713619350
```

```
# 2022/08 한 달 간 배출량이 가장 적게 집계된 광역자치단체
```

```
disQuantityBySido[disQuantityBySido$disQuantity == min(disQuantityBySido$disQuantity),]
```

```
##      citySidoName disQuantity
## 15     충청남도      194523700
```