지역별 이용객 수

NKH

2020 5 29

setwd("C://Users//CPB06GameN//Desktop//과제=인원")  
library(readxl)  
library(rJava)  
library(xlsx)  
library(ggplot2)  
library(tibble)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(magrittr)  
  
  
pop19 <- read.xlsx2(file = '총이용인원-요일별2019.xlsx', sheetIndex=1, startRow=6, header=T)  
pop19

## X. 월 화 수 목 금  
## 1 전체 11,459,916 11,545,432 11,657,576 11,518,342 11,958,311   
## 2 서울특별시 4,343,223 4,396,946 4,432,399 4,371,111 4,550,668   
## 3 부산광역시 1,040,293 1,051,389 1,057,331 1,068,119 1,088,958   
## 4 대구광역시 494,546 499,173 503,540 500,561 503,901   
## 5 인천광역시 777,999 782,655 791,987 771,217 806,456   
## 6 광주광역시 214,886 216,574 220,360 218,310 223,253   
## 7 대전광역시 250,447 252,601 255,794 255,067 269,584   
## 8 울산광역시 140,918 140,446 141,083 140,443 145,238   
## 9 세종특별자치시 27,320 27,503 28,257 27,890 30,519   
## 10 경기도 2,964,547 2,981,970 3,016,620 2,956,798 3,085,161   
## 11 강원도 93,614 92,875 95,273 91,769 98,044   
## 12 충청북도 116,700 112,965 113,292 113,483 119,853   
## 13 충청남도 167,641 165,039 164,911 167,324 180,395   
## 14 전라북도 120,483 119,865 119,948 119,536 123,531   
## 15 전라남도 123,625 122,781 123,609 123,150 127,764   
## 16 경상북도 165,102 161,013 166,465 163,479 169,260   
## 17 경상남도 336,290 339,614 343,154 346,274 352,694   
## 18 제주특별자치도 82,282 82,023 83,553 83,811 83,032   
## 토 일  
## 1 9,493,453 7,303,182   
## 2 3,613,517 2,848,444   
## 3 926,661 711,771   
## 4 437,091 327,279   
## 5 639,645 490,962   
## 6 167,846 127,180   
## 7 209,078 161,183   
## 8 112,978 82,780   
## 9 20,297 15,614   
## 10 2,403,144 1,806,326   
## 11 74,692 57,670   
## 12 88,858 67,674   
## 13 138,210 109,923   
## 14 87,404 63,644   
## 15 98,443 73,783   
## 16 134,537 99,811   
## 17 276,564 205,894   
## 18 64,488 53,244

str(pop19)

## 'data.frame': 18 obs. of 8 variables:  
## $ X.: chr "전체" "서울특별시" "부산광역시" "대구광역시" ...  
## $ 월: chr "11,459,916 " "4,343,223 " "1,040,293 " "494,546 " ...  
## $ 화: chr "11,545,432 " "4,396,946 " "1,051,389 " "499,173 " ...  
## $ 수: chr "11,657,576 " "4,432,399 " "1,057,331 " "503,540 " ...  
## $ 목: chr "11,518,342 " "4,371,111 " "1,068,119 " "500,561 " ...  
## $ 금: chr "11,958,311 " "4,550,668 " "1,088,958 " "503,901 " ...  
## $ 토: chr "9,493,453 " "3,613,517 " "926,661 " "437,091 " ...  
## $ 일: chr "7,303,182 " "2,848,444 " "711,771 " "327,279 " ...

pop19 <- pop19[2:9,]  
str(pop19)

## 'data.frame': 8 obs. of 8 variables:  
## $ X.: chr "서울특별시" "부산광역시" "대구광역시" "인천광역시" ...  
## $ 월: chr "4,343,223 " "1,040,293 " "494,546 " "777,999 " ...  
## $ 화: chr "4,396,946 " "1,051,389 " "499,173 " "782,655 " ...  
## $ 수: chr "4,432,399 " "1,057,331 " "503,540 " "791,987 " ...  
## $ 목: chr "4,371,111 " "1,068,119 " "500,561 " "771,217 " ...  
## $ 금: chr "4,550,668 " "1,088,958 " "503,901 " "806,456 " ...  
## $ 토: chr "3,613,517 " "926,661 " "437,091 " "639,645 " ...  
## $ 일: chr "2,848,444 " "711,771 " "327,279 " "490,962 " ...

이렇게 해서 불러내었으나, ’,’로 인하여 문자열로 인식되는 문제가 발생하였다. 하여,

mon <- as.numeric(gsub(',', '', pop19[,2]))  
tue <- as.numeric(gsub(',', '', pop19[,3]))  
wed <- as.numeric(gsub(',', '', pop19[,4]))  
thu <- as.numeric(gsub(',', '', pop19[,5]))  
fri <- as.numeric(gsub(',', '', pop19[,6]))  
sat <- as.numeric(gsub(',', '', pop19[,7]))  
sun <- as.numeric(gsub(',', '', pop19[,8]))

요일별로 각각 추출하였다. 이렇게 한 것은 for문 사용법이 미숙하여 선택한 방법이기도 하며, 추가적인 상의를 통해 주중과 주말로 통합하여 비교하기로 결정하였기에 이후 적용을 쉽게하기 위하여 선택한 방법이다. 해당 방법으로 ’,’를 제거하고 숫자열로 바꾼 값들은 이후,

week <- data.frame(mon,tue,wed,thu,fri)  
  
weeked <- data.frame(sat,sun)

데이터 프레임 과정을 거쳤다.

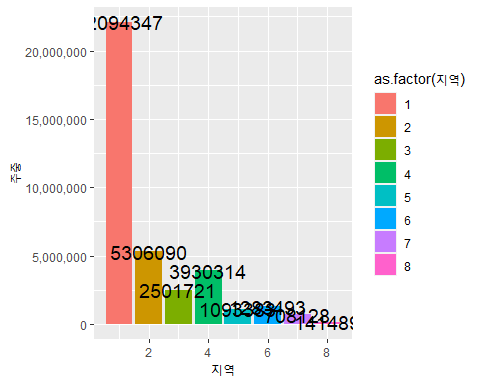
suw1<- sum(week[1,])  
bsw1<- sum(week[2,])  
dgw1<- sum(week[3,])  
icw1<- sum(week[4,])  
gjw1<- sum(week[5,])  
djw1<- sum(week[6,])  
usw1<- sum(week[7,])  
ssw1<- sum(week[8,])  
  
  
suw2<- sum(weeked[1,])  
bsw2<- sum(weeked[2,])  
dgw2<- sum(weeked[3,])  
icw2<- sum(weeked[4,])  
gjw2<- sum(weeked[5,])  
djw2<- sum(weeked[6,])  
usw2<- sum(weeked[7,])  
ssw2<- sum(weeked[8,])  
  
  
low<- data.frame(suw1, bsw1, dgw1, icw1, gjw1,djw1,usw1,ssw1)  
loed <- data.frame(suw2, bsw2, dgw2, icw2, gjw2,djw2,usw2,ssw2)

이후 주중 및 주말 이용객 수의 지역별 총합을 sum을 통하여 정리한 다음, 데이터프레임화를 실시하였다. 그러나 해당 데이터 프레임을 이용한 시각화 시도는 몇 가지 기술적 난관과 시간적인 문제에 봉착, 원하는 형태로 출력하는데 어려움을 겪었다. 때문에 부득이 하게 각 데이터 프레임 값을 r에서 수기로 벡터화한 다음 새로 데이터 프레임으로 만들었다.

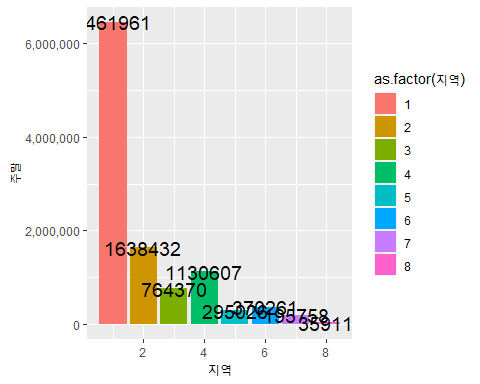
지역 <- c(1,2,3,4,5,6,7,8)  
주중 <- c(22094347, 5306090, 2501721, 3930314, 1093383, 1283493, 708128, 141489)  
주말 <- c( 6461961, 1638432, 764370, 1130607, 295026, 370261, 195758, 35911)  
  
last1 <- data.frame(지역,주중)  
last2 <- data.frame(지역,주말)

이 데이터 프레임을 바탕으로 한 시각화 자료는 다음과 같다.

#지역별 주중 이용객 수 총합  
ggplot(last1, aes(x=지역, y=주중,fill=as.factor(지역)))+  
 geom\_bar(stat = 'identity')+  
 scale\_y\_continuous(labels = scales::comma)+  
 geom\_text(aes(label=주중),  
 size=5,  
 position="stack")



#지역별 주말 이용객 수 총합  
ggplot(last2, aes(x=지역, y=주말,fill=as.factor(지역)))+  
 geom\_bar(stat = 'identity')+  
 scale\_y\_continuous(labels = scales::comma)+  
 geom\_text(aes(label=주말),  
 size=5,  
 position="stack")



#지역별 주중-주말 이용객 수 차이  
차이<- (last1$주중)-(last2$주말)  
지역

## [1] 1 2 3 4 5 6 7 8

last3 <- data.frame(지역, 차이)  
  
ggplot(last3, aes(x=지역, y=차이,fill=as.factor(지역)))+  
 geom\_bar(stat = 'identity')+  
 scale\_y\_continuous(labels = scales::comma)+  
 geom\_text(aes(label=차이),  
 size=5,  
 position="stack")

