

2018.4.15

숙제 0324 로 낸 dt8 데이터를 다운 받아서 경로 상에 한글이 들어가지 않는 폴더에 두시오. r 에서 dt8 을 불러오시오.

위의 yaml(hyphen 3 개 — 사이에 있는 부분. 영어로는 재미 있게 “YAML Ain’t Markup Language”의 약자라고 함.)을 자신에게 맞게 고치시오.

니팅해서 rmd 파일과 html 파일을 함께 제출하시오.

1 문제 0

다음의 코드를 그대로 실행하시오

```
library(data.table)
## Warning: package 'data.table' was built under R version 3.4.4
```

```
# 다음에서 12345 대신에 자기 학번을 넣으시오.
set.seed(201428027)
```

```
# 다음에서 컴퓨터에 맞게 경로를 고치세요.
load(file="C:/201428027/2018/kao/dt8.RData")
```

```
sp = sample(1:nrow(dt8), size = 1000, replace = F)
dt8 = data.table(dt8)
dt9 = dt8[sp, ][order(V1)]
dt8 자료의 변수가 의미하는 바는 다음과 같다.
V1 가구번호
```

V2 조사년도

V3 조사부문

V4 가구주성별

V5 가구주만나이

V6 가구주동거여부

V7 가구주교육정도_학교

V8 가구주교육정도_수료여부

V9 가구주혼인상태

V10 가구주종사상지위

V11 가구주산업

V12 가구주직업

V13 가구원수

V14 가구유형

V15 현재살고있는주택의종류

V16 전용면적

V17 현재살고있는주택의입주형태

V18 자산총액

V19 금융자산

V20 저축액

V21 적립식저축

V22 예치식저축

V23 기타저축

V24 현거주지전월세보증금

V25 실물자산

V26 부동산

V27 거주주택

V28 거주주택이외부동산

V29 계약금및중도금납입액

V30 기타실물자산

V31 자동차

V32 자동차이외

V33 부채총액

V34 금융부채

V35 담보대출

V36 부동산담보대출

V37 거주주택담보대출

V38 신용대출

V39 신용카드관련대출

V40 외상및할부미상환액

V41 켓탄후불입금액

V42 임대보증금

V43 순자산액

V44 경상소득

V45 근로소득

V46 사업소득

V47 재산소득

V48 이전소득

V49 공적이전소득

V50 사적이전소득

V51 가처분소득

V52 균등화가처분소득_분배

V53 비소비지출

V54 세금및과징금

V55 공적연금및사회보험료

V56 가구간이전

V57 비영리단체이전

V58 연간지급이자

V59 연간상환액

V60 연간지급이자및상환액

V61 담보대출용도별(담보_대출잔액)

V62 거주주택마련(담보_대출잔액)

V63 거주주택이외부동산마련(담보_대출잔액)

V64 전(월)세보증금마련(담보_대출잔액)

V65 증권투자금마련(담보_대출잔액)

V66 부채상환(담보_대출잔액)

V67 사업자금(영농자금포함)마련(담보_대출잔액)

V68 결혼자금마련(담보_대출잔액)

V69 의료비마련(담보_대출잔액)

V70 교육비마련(담보_대출잔액)

V71 생활비마련(담보_대출잔액)

V72 기타용도(담보_대출잔액)

V73 담보형태(담보_대출잔액)

V74 거주주택담보(담보_대출잔액)

V75 거주주택이외부동산담보(담보_대출잔액)

V76 예금,적금,보험담보(담보_대출잔액)

V77 기타담보형태(담보_대출잔액)

V78 신용대출용도(신용_대출잔액)

V79 거주주택마련(신용_대출잔액)

V80 거주주택이외부동산마련(신용_대출잔액)

V81 전(월)세보증금마련(신용_대출잔액)

V82 증권투자금마련(신용_대출잔액)

V83 부채상환(신용_대출잔액)

V84 사업자금마련(신용_대출잔액)

V85 결혼자금마련(신용_대출잔액)

V86 의료비마련(신용_대출잔액)

V87 교육비마련(신용_대출잔액)

V88 생활비마련(신용_대출잔액)

V89 기타용도(신용_대출잔액)

V90 부채보유여부

V91 수도권여부

V92 한부모가구

V93 조손가구

V94 노인가구

V95 가구주연령 10 세계층

V96 소득 5 분위

V97 자산총액 5 분위

V98 순자산액 5 분위

V99 소득계층구간별

V100 입주형태통합별

V101 가구주종사상지위

V102 가구주종사상지위(보도용)

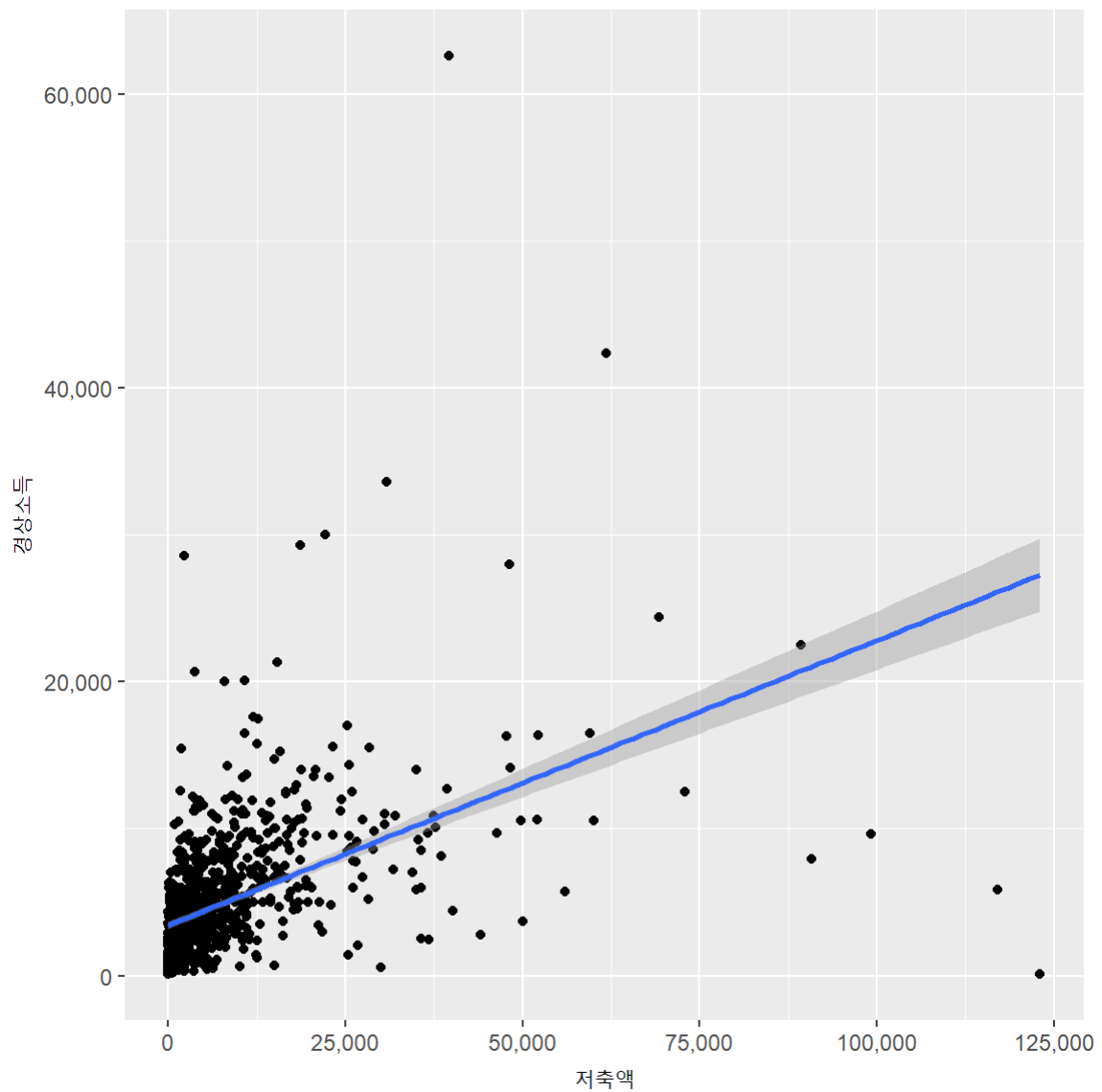
V103 가구주교육정도통합별

V104 가중치

2 문제 1

dt9 에서 하나의 종속변수(반응변수)와 하나의 설명변수(독립변수)를 선택하여 산점도를 그리고, 회귀 직선을 추가하시오.

```
library(ggplot2)
library(scales)
#종속변수-v20 저축액
#설명변수-v44 경상소득
ggplot(dt9,aes(x=V20,y=V44))+geom_point(stat='identity')+scale_x_continuous(labels=comma)+scale_y_continuous(labels=comma)+ geom_smooth(method='lm')+xlab("저축액")+ylab("경상소득")
```



3 문제 2

위에서 선택한 변수를 가지고 회귀분석을 하고, 그 결과를 해석하시오.

```
dt9.lm = lm(V20 ~ V44, data=dt9)
summary(dt9.lm)
##
## Call:
## lm(formula = V20 ~ V44, data = dt9)
##
## Residuals:
```

##	Min	1Q	Median	3Q	Max
##	-39928	-3776	-1767	508	122308

```
##
## Coefficients:
```

##		Estimate	Std. Error	t value	Pr(> t)
##					

```
## (Intercept) 646.04522 463.52004 1.394 0.164
## V44 1.26099 0.07022 17.958 <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10240 on 998 degrees of freedom
## Multiple R-squared: 0.2442, Adjusted R-squared: 0.2435
## F-statistic: 322.5 on 1 and 998 DF, p-value: < 2.2e-16
경상소득이 1 단위 증가하면 저축액이 1.26 증가한다.
```

귀무가설이란 경상소득의 계수가 0 이라는 것이다. 귀무가설이 참일때 현재와 같은 결과가 나올 확률이 p 값인데, 그값이 $2e-16$ 으로 거의 0 이다. 따라서 귀무가설을 기각한다. 설명변수들이 종속변수의 24%를 설명해준다.

4 문제 3

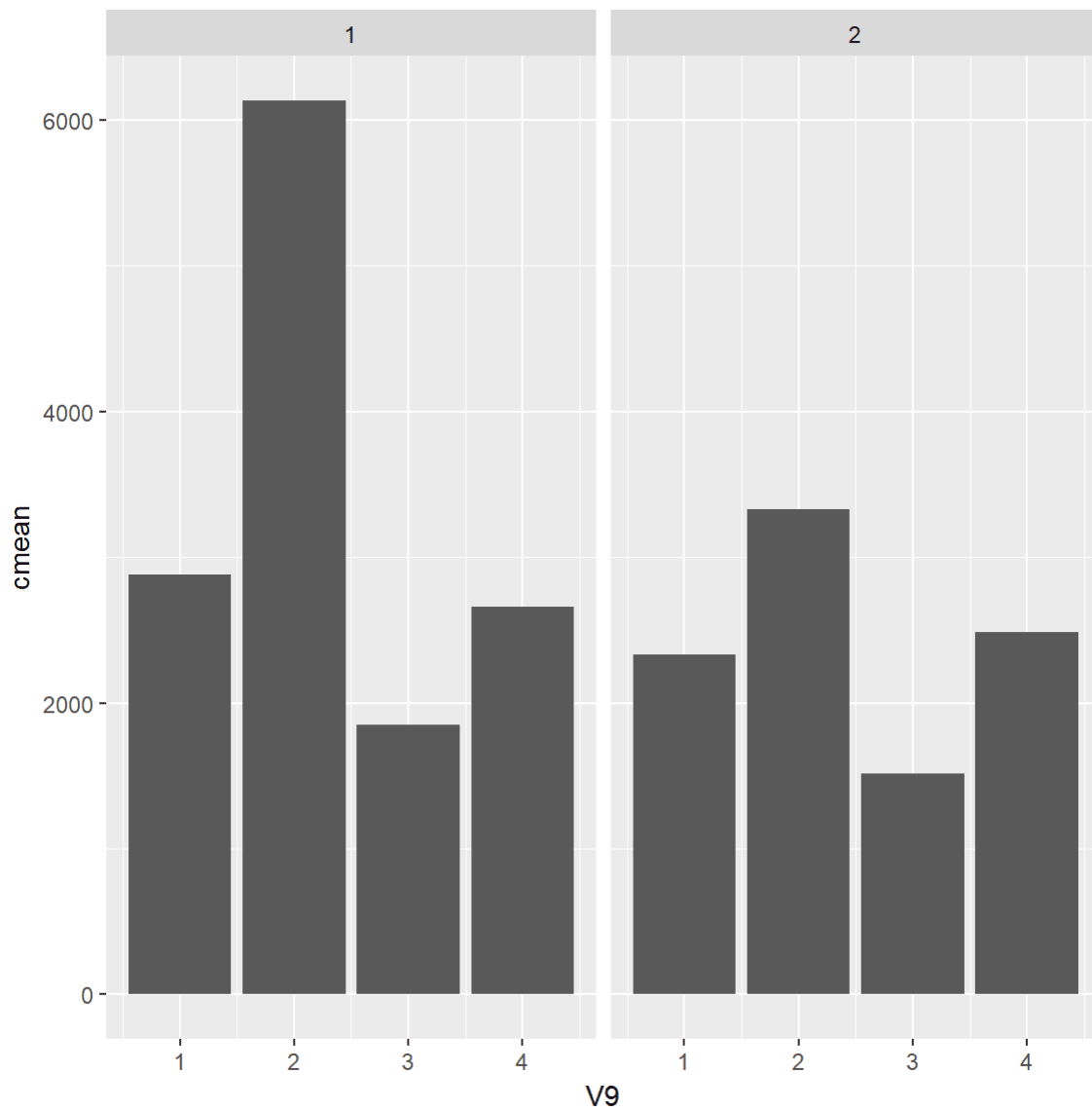
질적 변수 2 개를 선택하고, 양적 변수 1 개를 선택하시오. 양적변수의 평균을 2 개의 질적 변수별로 구하시오. 힌트: `data.table` 에서 `by` 활용

```
#질적변수-v4 성별, v9 가구주혼인상태
#양적변수-v44 경상소득
dt9=data.table(dt9)
q1=dt9[,.(cmean=(mean(V44))),by=.(V4,V9)][order(V4,V9,cmean)]
q1
##      V4 V9      cmean
## 1:   1  1 2879.660
## 2:   1  2 6132.387
## 3:   1  3 1849.526
## 4:   1  4 2656.093
## 5:   2  1 2333.071
## 6:   2  2 3329.171
## 7:   2  3 1517.163
## 8:   2  4 2487.531
```

5 문제 4

문제 3을 막대 그래프로 그리시오.

```
ggplot(q1) + geom_bar(aes(x=V9,y=cmean),stat='identity') +
facet_wrap(~V4)
```

6 문제 5

질적 변수 3 개를 선택하시오. 2 개의 질적변수 별로 나머지 질적변수를 가진 가구가 몇 개나 있는지 구하시오.

힌트: `data.table` 에서 `.N` 활용

```
#질적변수-v4 성별, v9 가구주혼인상태, v15 현재살고있는 주택종류
dt9[V4==1, .(N), by=(V9, V15)]
##      V9 V15    N
##  1:   1   3    14
##  2:   2   3    71
##  3:   2   1   195
##  4:   2   2   355
##  5:   2   4    14
```

```
## 6: 3 1 11
## 7: 4 2 13
## 8: 3 2 7
## 9: 4 1 18
## 10: 1 1 23
## 11: 4 3 11
## 12: 1 2 15
## 13: 1 4 1
## 14: 4 4 1
## 15: 3 3 1
```

```
dt9[V4==2,.(.N),by=(V9,V15)]
##      V9 V15  N
## 1: 1 4 4
## 2: 1 1 11
## 3: 3 3 15
## 4: 2 3 5
## 5: 4 2 32
## 6: 3 2 43
## 7: 3 1 61
## 8: 2 2 13
## 9: 1 2 9
## 10: 2 1 16
## 11: 3 4 4
## 12: 4 3 7
## 13: 4 1 23
## 14: 4 4 2
## 15: 1 3 4
## 16: 2 4 1
```

7 문제 6

양적 변수 1 개를 선택하시오. 양적 변수를 5 개의 집단으로 구분하는 질적 변수를 만드시오. 힌트: **cut** 명령어 활용

```
#V43 순자산액
cut1 <- cut(dt9$V43, breaks=5, include.lowest=T)
cut1=as.character(cut1)
cut1=data.table(cut1)
cut1[cut1=="[-2.84e+04,1.1e+05]",v43n=="1"]
cut1[cut1=="(1.1e+05,2.47e+05]",v43n=="2"]
cut1[cut1=="(2.47e+05,3.84e+05]",v43n=="3"]
cut1[cut1=="(3.84e+05,5.22e+05]",v43n=="4"]
cut1[cut1=="(5.22e+05,6.6e+05]",v43n=="5"]
unique(cut1)[order(v43n)]
##      cut1 v43n
## 1: [-2.84e+04,1.1e+05] 1
## 2: (1.1e+05,2.47e+05] 2
## 3: (2.47e+05,3.84e+05] 3
## 4: (3.84e+05,5.22e+05] 4
## 5: (5.22e+05,6.6e+05] 5
```

8 문제 7

질적 변수 하나와 양적 변수 하나를 선택하시오. 양적 변수의 평균을 질적변수별로 구하시오.

```
#질적변수-v4 가구주 성별
#양적변수-v44 경상소득
q7=dt9[,.(cmean=mean(V44)),by=.(sex=V4)]
q7
##      sex      cmean
## 1:    1 5594.721
## 2:    2 2110.640
```

9 문제 8

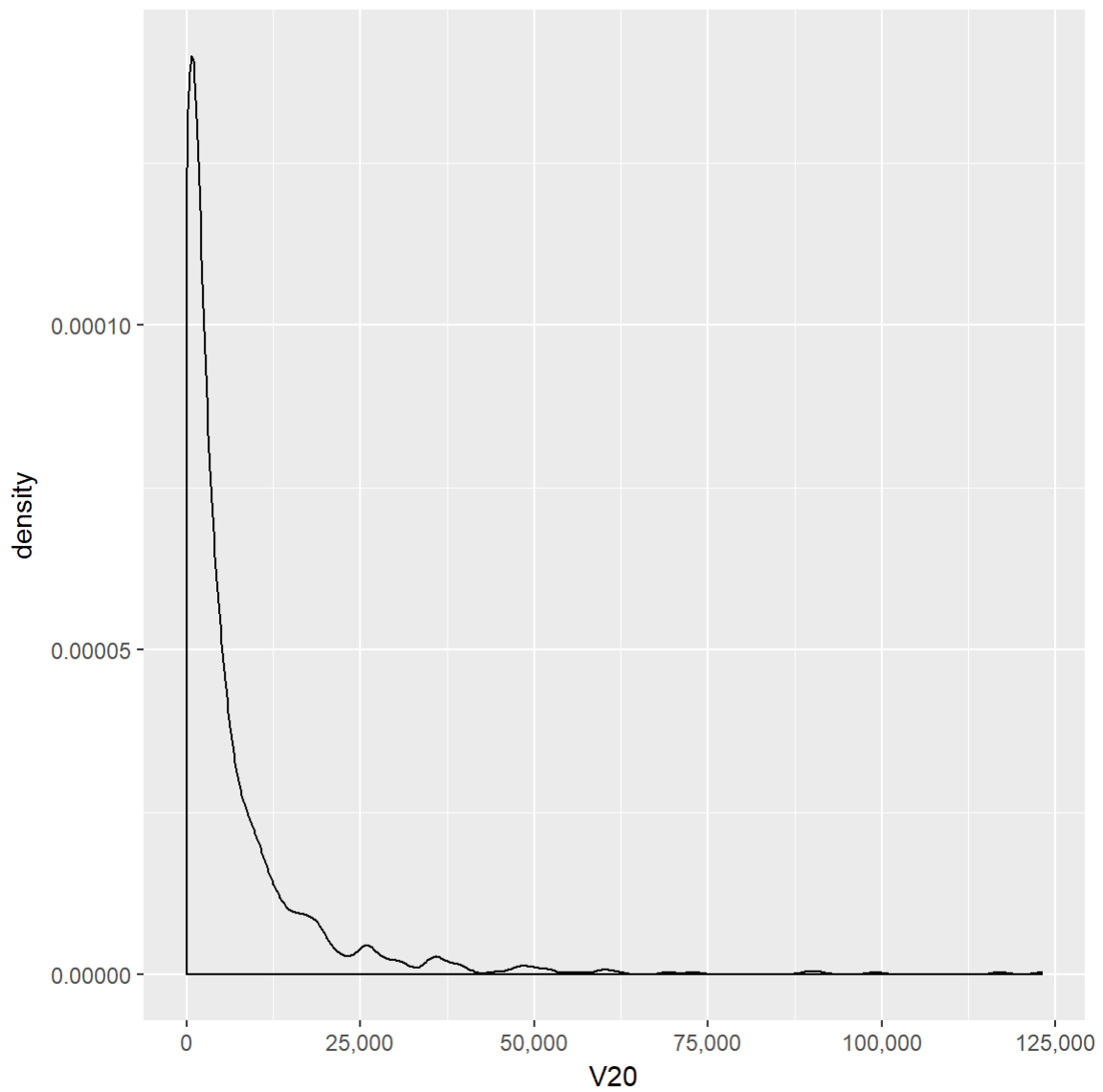
문제 7에서 질적 변수를 입력하면 그 집단의 양적 변수 평균을 되돌려주는 함수를 만드시오. 예: 직업을 입력하면 그 직업의 평균 부채를 보여주는 함수.

```
myf=function(x){
a=q7[sex==(x),cmean]
  rst=paste('성별',x,'경상소득',a)
  rst
}
myf(2)
## [1] "성별 2 경상소득 2110.64"
```

10 문제 9

양적 변수 하나를 선택해서 그 분포를 보여주는 밀도함수를 그리시오.

```
#양적변수-v20 저축액
ggplot(dt9)+geom_density(aes(x=V20))+scale_x_continuous(labels=comma)
)+scale_y_continuous(labels=comma)
```



11 문제 10

부채총액(V33)이 자산총액(V18)보다 많은 가구가 전체의 몇 %나 되는가?

```
mean(dt9$V33>dt9$V18)
## [1] 0.023
```