

데이터 소개

Jinseog Kim

Dongguk University

jskim1986@gmail.com

2018-03-13

Contents

| | | |
|-----|---|----|
| 1 | 데이터 소스 | 3 |
| 1.1 | birth데이터 | 4 |
| 1.2 | contribution데이터 | 7 |
| 2 | cu.summary Automobile Data from 'Consumer Reports' 1990 | 13 |
| 2.1 | 변수설명 | 14 |

1 데이터 소스

1. birth: R in a Nutshell: A Desktop Quick Reference (Adler, 2009)
2. contribution: Midwestern 대학의 기부자 정보 데이터 (Ledolter and Swersey, 2007)
3. Orange juice: bayesm package (P. Rossi, Montgomery 1987).

| data | 자료수 | 변수의 수 |
|--------------|---------|-------|
| birth | 427,323 | 13 |
| contribution | 1,230 | 11 |
| Orange juice | 28,947 | 17 |
| cu.summary | 117 | 5 |

1.1 birth데이터

□ 패키지 설치 및 데이터 로드

```
#library(lattice)
#install.packages("nutshell")
library(nutshell)
data(births2006.smpl)
head(births2006.smpl)
```

```
##          DOB_MM DOB_WK  MAGER  TBO_REC  WTGAIN  SEX  APGAR5
## 591430         9     1    25         2     NA    F     NA
## 1827276        2     6    28         2     26    M     9
## 1705673        2     2    18         2     25    F     9
## 3368269       10     5    21         2     6     M     9
## 2990253        7     7    25         1    36    M    10
## 966967         3     3    28         3    35    M     8
##
##          DMEDUC  UPREVIS  ESTGEST  DMETH_REC  DPLURAL  DBWT
## 591430          NULL     10      99    Vaginal 1 Single 3800
## 1827276  2 years of college  10      37    Vaginal 1 Single 3625
## 1705673          NULL     14      38    Vaginal 1 Single 3650
```

| | | | | | |
|------------|------------------------|----|----|------------------|------|
| ## 3368269 | NULL | 22 | 38 | Vaginal 1 Single | 3045 |
| ## 2990253 | 2 years of high school | 15 | 40 | Vaginal 1 Single | 3827 |
| ## 966967 | NULL | 18 | 39 | Vaginal 1 Single | 3090 |

□ 변수설명

1. DOB_MM: Month of date of birth
2. DOB_WK: Day of week of birth
3. MAGER: Mother's age
4. TBO_REC: Total birth order
5. WTGAIN: Weight gain by mother
6. SEX: a factor with levels F M, representing the sex of the child
7. APGAR5: APGAR score
8. DMEDUC: Mother's education level
9. UPREVIS: Number of prenatal visits
10. ESTGEST: Estimated weeks of gestation
11. DMETH_REC: Delivery Method
12. DPLURAL: "Plural Births;" levels include 1 Single, 2 Twin, 3 Triplet, 4 Quadruplet, and 5 Quintuplet or higher
13. DBWT: Birth weight, in grams

1.1.1 기본 분석

1. 요일별 신생아의 수
2. 요일에 따른 출산방법의 차이 (C-section:제왕절개, Unknown, Vaginal(자연분만))
3. 출산방법에 따른 신생아 체중 차이
4. 엄마의 출산 경험(초산, 두번째, ...)에 따른 신생아의 체중 차이
5. 엄마의 상대 체중(WTGAIN)과 신생아의 체중과의 관계
6. 출산시 엄마의 상태(APGAR score)와 신생아의 체중과의 관계

1.2 contribution데이터

□ <http://www.biz.uiowa.edu/faculty/jledolter/DataMining>

□ 졸업생 기부자 관리 자료

□ 연도별 졸업생들의 기여도

□ 기부 홍보 및 캠페인 전략 수립

□ 데이터 로드

```
don <- read.csv("http://datamining.dongguk.ac.kr/lectures/2018-1/dm/contribution.csv")
don[1:5,]
```

```
##   Gender Class.Year Marital.Status   Major Next.Degree FY04Giving
## 1      M      1957           M History         LLB         2500
## 2      M      1957           M Physics          MS         5000
## 3      F      1957           M   Music         NONE         5000
## 4      M      1957           M History         NONE           0
## 5      M      1957           M Biology          MD         1000
##   FY03Giving FY02Giving FY01Giving FY00Giving AttendanceEvent
## 1         2500         1400         12060         12000             1
```

| | | | | | |
|------|------|------|------|-------|---|
| ## 2 | 5000 | 5000 | 5000 | 10000 | 1 |
| ## 3 | 5000 | 5000 | 5000 | 10000 | 1 |
| ## 4 | 5100 | 200 | 200 | 0 | 1 |
| ## 5 | 1000 | 1000 | 1005 | 1000 | 1 |

`names(don)`

```
## [1] "Gender"           "Class.Year"       "Marital.Status"
## [4] "Major"           "Next.Degree"     "FY04Giving"
## [7] "FY03Giving"     "FY02Giving"     "FY01Giving"
## [10] "FY00Giving"     "AttendanceEvent"
```

□ 변수 설명

1. Gender : 기부자 성별
2. Class.Year : 졸업연도
3. Marital.Status: 결혼상태
4. Major: 전공
5. Next.Degree: 진학 상태
6. FY04Giving: 2004년도 기부액
7. FY03Giving: 2003년도 기부액

8. FY02Giving: 2002년도 기부액
9. FY01Giving: 2001년도 기부액
10. FY00Giving: 2000년도 기부액
11. AttendanceEvent : 이벤트 행사 참여

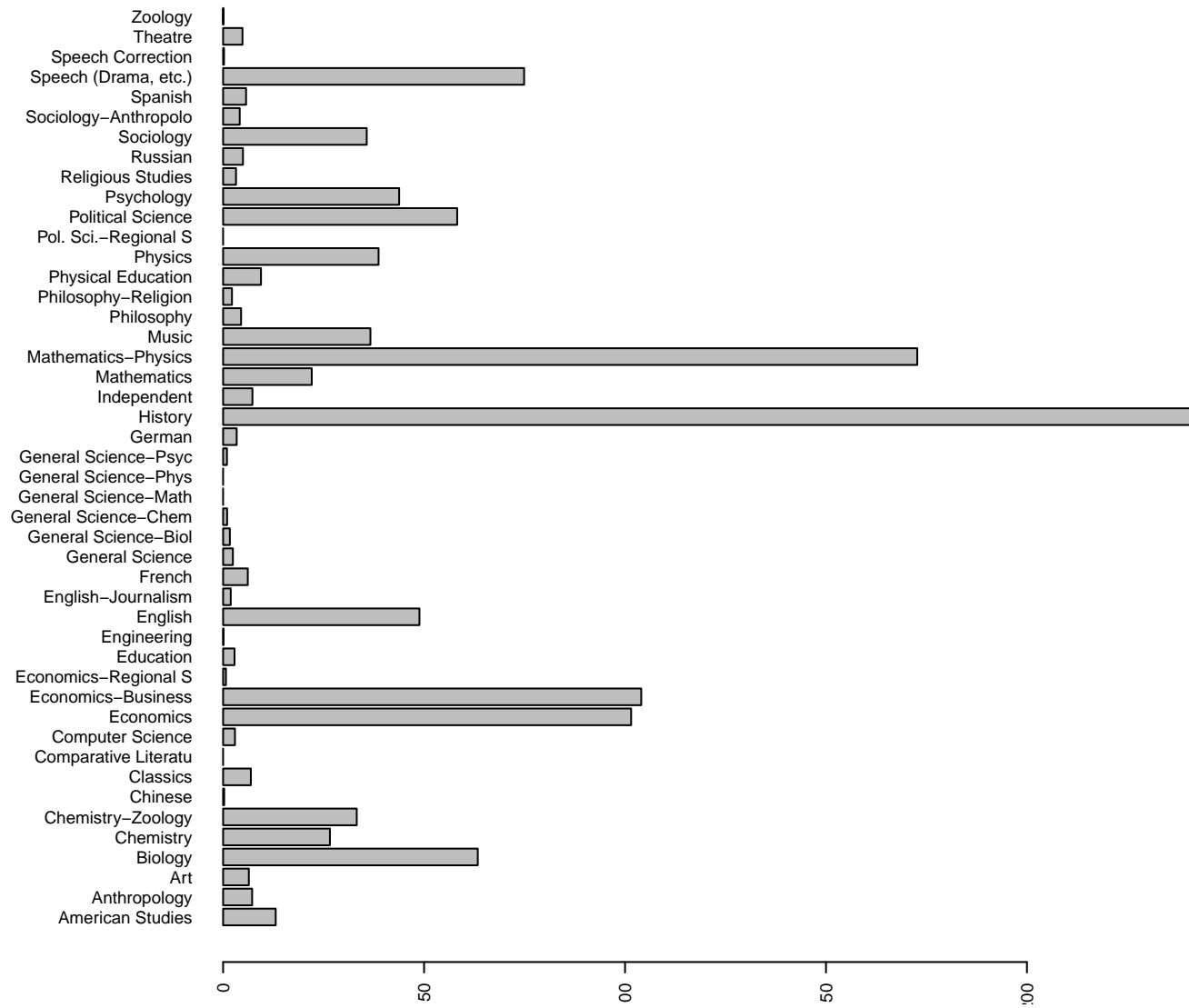
▣ 총 기부금액 계산

```
don$TGiving <- (don$FY00Giving+don$FY01Giving+
                don$FY02Giving+don$FY03Giving+
                don$FY04Giving)/1000
```

1.2.1 기본 분석

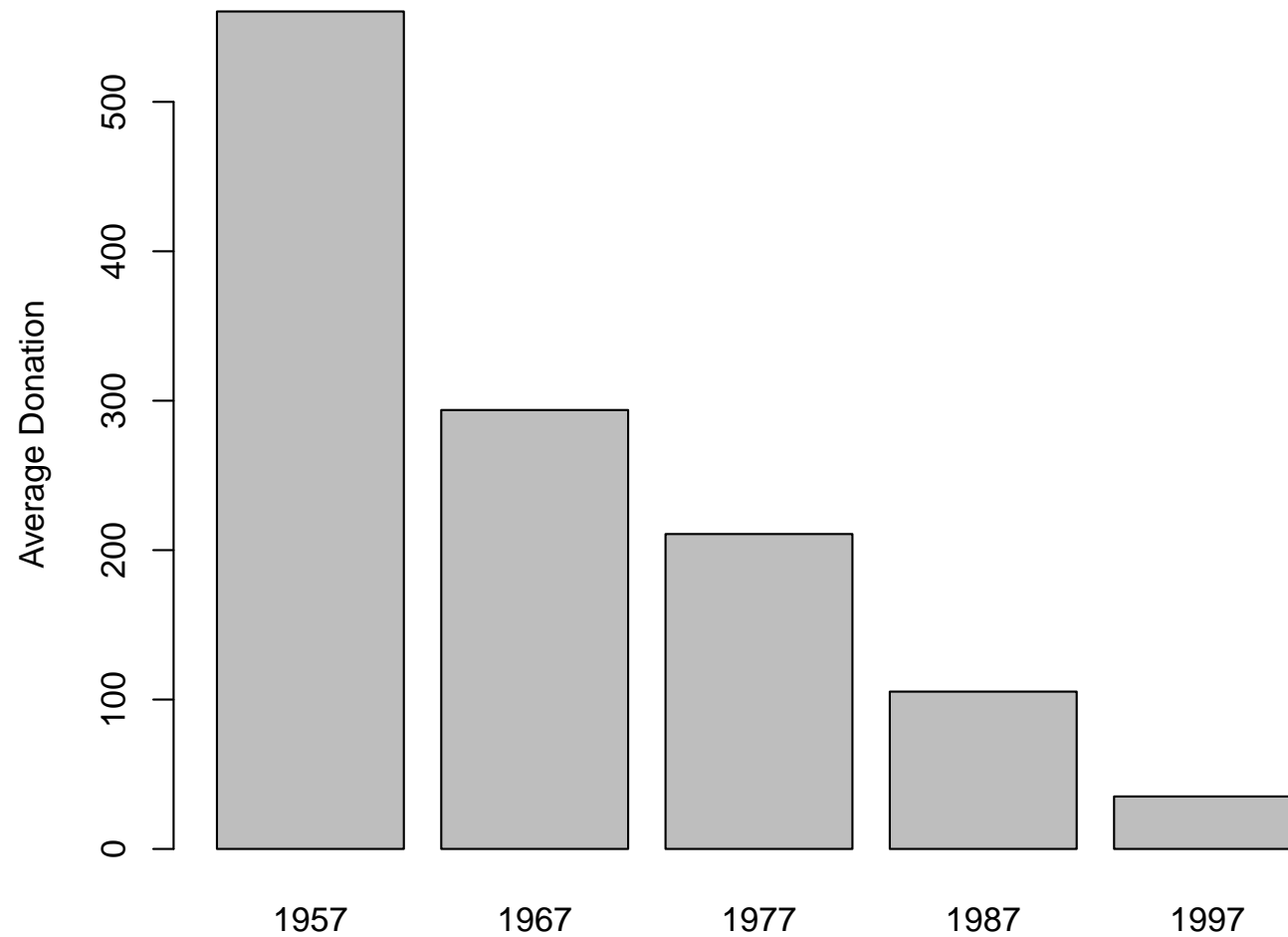
1. 출신학과(전공별) 기부금액(TGiving)

```
t10 <- tapply(don$TGiving,don$Major,FUN=sum,na.rm=TRUE)
par(mar=c(2, 10, 1, 2), cex=0.6)
names(t10) <- substring(names(t10), 1, 20)
barplot(t10, xlab="Average Donation", horiz=T, las=2)
```



2. 연도별 기부금액

```
t11 <- tapply(don$TGiving,don$Class.Year,FUN=sum,na.rm=TRUE)
barplot(t11, ylab="Average Donation")
```



3. 성별 기부금액

4. ...

2 cu.summary Automobile Data from 'Consumer Reports' 1990

□ Data description

- 1990년 4월, 자동차에 관련된 소비자보고서(Consumer Reports)에 수록된 데이터
- rpart 패키지에 내장

2.1 변수설명

| Vaiables | descriptions |
|-------------|---|
| Price | 표준모델에 대한 차량가격(US dollars) |
| Country | 자동차 생산국 'Brazil', 'England', 'France', 'Germany', 'Japan', 'Japan/USA', 'Korea', 'Mexico', 'Sweden' and 'USA' |
| Reliability | 신뢰도(5점; Much worse < worse < average < better < Much better) |
| Mileage | 연비 (연료 소비율; miles/gallon) |
| Type | 외형 (Compact, Large, Medium, Small, Sporty, Van) |