

DATA300_LAB3_HanhPhan

```
mydata <- read.csv("dataset.csv")
```

Part 1 Run through every value in cartype, if the value is blank, replace it with “no car”, otherwise just keep it the same as it was.

```
mydata$cartype <- ifelse(mydata$cartype=="", "no car", mydata$cartype)
```

Part 2 Create a new variable called independent. If the respondent is not Democrat (democrat=0) and not Republican (GOP=0), then set the independent value to 1, otherwise set it to 0

```
mydata$independent <- ifelse(mydata$democrat==0 & mydata$GOP==0, 1, 0)
mydata$independent
```

```
##      [1] 0 0 1 1 1 0 0 0 1 1 0 0 0 1 0 0 1 0 1 0 1 1 0 0 0 0 0 0 0 0 1 0 0 0
0 0 0
##      [38] 0 0 0 1 1 0 0 0 1 0 0 0 0 0 1 0 0 1 1 1 0 1 0 1 0 0 0 0 0 0 0 1 0 1
0 1 1
##      [75] 0 0 1 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 1 0 1 0 1 0 1 0 0 0 0 0 1 1 0 0
0 0 0
##     [112] 0 1 0 1 0 0 1 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 1 0 1 1 1 0 1 0 0 0
1 1 1
##     [149] 0 0 1 0 0 0 1 0 0 0 1 0 0 1 1 0 0 0 1 1 0 1 1 1 0 1 0 0 1 0 0 0 0 1
0 0 1
##     [186] 1 0 0 0 1 0 0 0 0 1 1 1 0 1 0 0 0 1 0 1 1 1 0 1 1 0 0 0 0 0 0 1 0 0
1 0 0
##     [223] 1 0 0 0 0 0 1 1 1 0 0 0 1 1 1 0 0 0 1 0 0 0 1 0 0 0 1 1 1 1 0 0 0 1
1 0 0
##     [260] 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 1 1 0 0 0 1 0 1 0 0 0 0 0 0
0 0 0
##     [297] 0 0 1 1 0 1 1 1 0 0 1 0 1 1 1 0 0 1 0 0 0 1 0 1 0 0 0 0 0 0 0 0 1 0 1
0 1 0
##     [334] 0 0 0 0 0 0 1 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 1 0 0 0 0 1 0 0 1 0 0 1
0 0 0
##     [371] 0 1 0 0 1 0 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 1 1 0 0 0 0 0 0 1
0 1 1
##     [408] 0 0 1 1 1 0 1 0 0 1 0 0 0 0 1 1 0 1 1 1 1 1 0 0 1 0 1 0 1 1 0 0 0 1
0 0 0
##     [445] 1 0 0 0 0 1 1 0 1 0 0 1 1 0 1 0 1 1 0 1 0 1 0 1 0 1 0 0 0 1 0 0 0 1
1 1 1
##     [482] 0 0 1 0 0 0 0 0 1 0 1 0 0 1 0 1 1 0 0 0 1 1 0 0 0 0 0 0 1 1 1 0 0 1
0 0 0
##     [519] 0 0 0 0 1 0 1 1 0 1 1 1 0 0 0 0 1 0 0 1 0 0 1 0 0 0 1 0 1 0 0 1 0 1
0 0 0
##     [556] 0 0 1 0 0 0 0 0 1 1 0 0 1 1 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 1
1 1 0
```

```

## [593] 0 0 1 1 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1
0 0 0
## [630] 1 0 0 0 0 0 1 0 0 0 1 0 0 1 0 1 0 1 1 0 0 0 0 1 1 0 1 0 0 0 0 0 0 0
0 0 0
## [667] 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 1 0 0 1 1 0 0 0
0 0 1
## [704] 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 1 1 0 1
1 1 0
## [741] 0 0 0 1 1 0 0 0 1 0 0 0 1 1 1 0 0 0 0 1 1 0 0 1 1 0 0 0 0 0 1 0 1 0
1 0 0
## [778] 0 1 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0
## [815] 0 0 0 1 0 1 0 0 1 0 0 0 0 0 0 0 1 1 0 1 1 1 0 0 0 0 1 0 0 0 0 1 0 0
1 0 1
## [852] 0 0 0 0 1 1 0 1 0 1 0 0 0 1 0 1 0 1 0 0 1 0 1 1 0 1 1 1 0 0 0 0 0 0
0 0 1
## [889] 0 0 0 0 0 0 0 0 0 0 1 1 0 1 1 0 1 0 0 0 1 0 0 0 1 0 1 1 1 0 0 1 0 1
0 0 0
## [926] 0 0 0 0 0 0 0 0 1 0 1 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 0 1 1 0 0 0
1 1 1
## [963] 1 0 0 1 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 1 1
0 0 1
## [1000] 0 1 0 1 0 1 0 0 1 1 0 1 0 1 0 0 1 1 0 0 0 0 1 1 1 0 0 0 1 1 0 1 0 0
0 0 0
## [1037] 0 0 0 1 0 0 0 0 1 0 0 0 1 0 1 0 0 1 1 0 0 1 0 0 1 1 1 0 0 0 1 0 1 0
0 0 1
## [1074] 1 1 0 0 1 0 0 0 1 0 1 0 1 0 1 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 1
0 0 1
## [1111] 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 1 1 1 0 1 1 1 0 1 0 0 1 0 0 0 0
0 0 1
## [1148] 1 0 1 0 1 0 1 1 0 0 1 0 0 0 0 1 0 0 0 1 0 0 0 0 1 1 0 1 0 1 0 1 1 0
0 1 0
## [1185] 0 0 1 0 0 0 0 0 0 1 0 0 0 1 0 1 0 1 0 1

```

```
summary(mydata)
```

```

##      subjectid      cartype      education      age
## Min.   :  1.0   Length:1204   Min.   :1.00   Min.   :18.00
## 1st Qu.: 301.8   Class :character 1st Qu.:2.00   1st Qu.:36.00
## Median : 602.5   Mode  :character  Median :3.00   Median :47.00
## Mean   : 602.5                Mean  :2.59   Mean  :47.86
## 3rd Qu.: 903.2                3rd Qu.:3.00   3rd Qu.:58.00
## Max.   :1204.0                Max.   :3.00   Max.   :92.00
##                                     NA's   :5      NA's   :24
##      urban      married      kids      black
## Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
## 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000
## Median :0.0000   Median :1.0000   Median :0.0000   Median :0.00000
## Mean   :0.4718   Mean   :0.5905   Mean   :0.3455   Mean   :0.06894
## 3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.00000

```

```
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.00000
##
## female democrat GOP bushapproval
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :1.000
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:1.000
## Median :0.0000 Median :0.0000 Median :0.0000 Median :3.000
## Mean :0.4967 Mean :0.3181 Mean :0.3613 Mean :2.578
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:4.000
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :4.000
## NA's :70
## independent
## Min. :0.0000
## 1st Qu.:0.0000
## Median :0.0000
## Mean :0.3206
## 3rd Qu.:1.0000
## Max. :1.0000
##
```

Part 3 Drop the subjectid column and change the column names to be informative Use stargazer to create a table of summary statistics.

```
library(stargazer)

##
## Please cite as:

## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary
## Statistics Tables.

## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer

drop <- c("subjectid")
df = mydata[,!(names(mydata)%in%drop)]
colnames(df) <-
c("Cartype", "Education", "Age", "Urban", "Married", "Kids", "Black", "Female", "Demo
crat", "GOP", "Bush Approval", "Independent")

stargazer(df, type="html", title="Traffic Survey Data - Summary Statistic
Table", column.sep.width = "15pt")
```

Traffic Survey Data - Summary Statistic Table

Statistic

N

Mean

St. Dev.

Min

Pctl(25)

Pctl(75)

Max

Education

1,199

2.590

0.602

1.000

2.000

3.000

3.000

Age

1,180

47.864

16.119

18.000

36.000

58.000

92.000

Urban

1,204

0.472

0.499

0

0

1

1

Married

1,204

0.591

0.492

0

0

1

1

Kids

1,204

0.346

0.476

0

0

1

1

Black

1,204

0.069

0.253

0

0

0

1

Female

1,204

0.497

0.500

0

0

1

1

Democrat

1,204

0.318

0.466

0

0

1

1

GOP

1,204

0.361

0.481

0

0

1

1

Bush Approval

1,134

2.578

1.308

1.000

1.000

4.000

4.000

Independent

1,204

0.321

0.467

0

0

1

1

Part 4: Create a table of 2 variables: cartype and bushapproval. The table shows that most people who pickup seem to support President Bush (144 supported and 61 didn't). Approximately a half of car, no car, and SUV drivers supported President Bush and the other half didn't.

```
table(mydata$cartype, mydata$bushapproval)
```

```
##  
##           1    2    3    4  
##   Car      252  71 103 213  
##  no car    37  12  15  35  
##  Pickup   45  16  32 112  
##   SUV     62  22  32  75
```