Analysis of the Healthcare cost and Utilization in Wisconsin hospitals
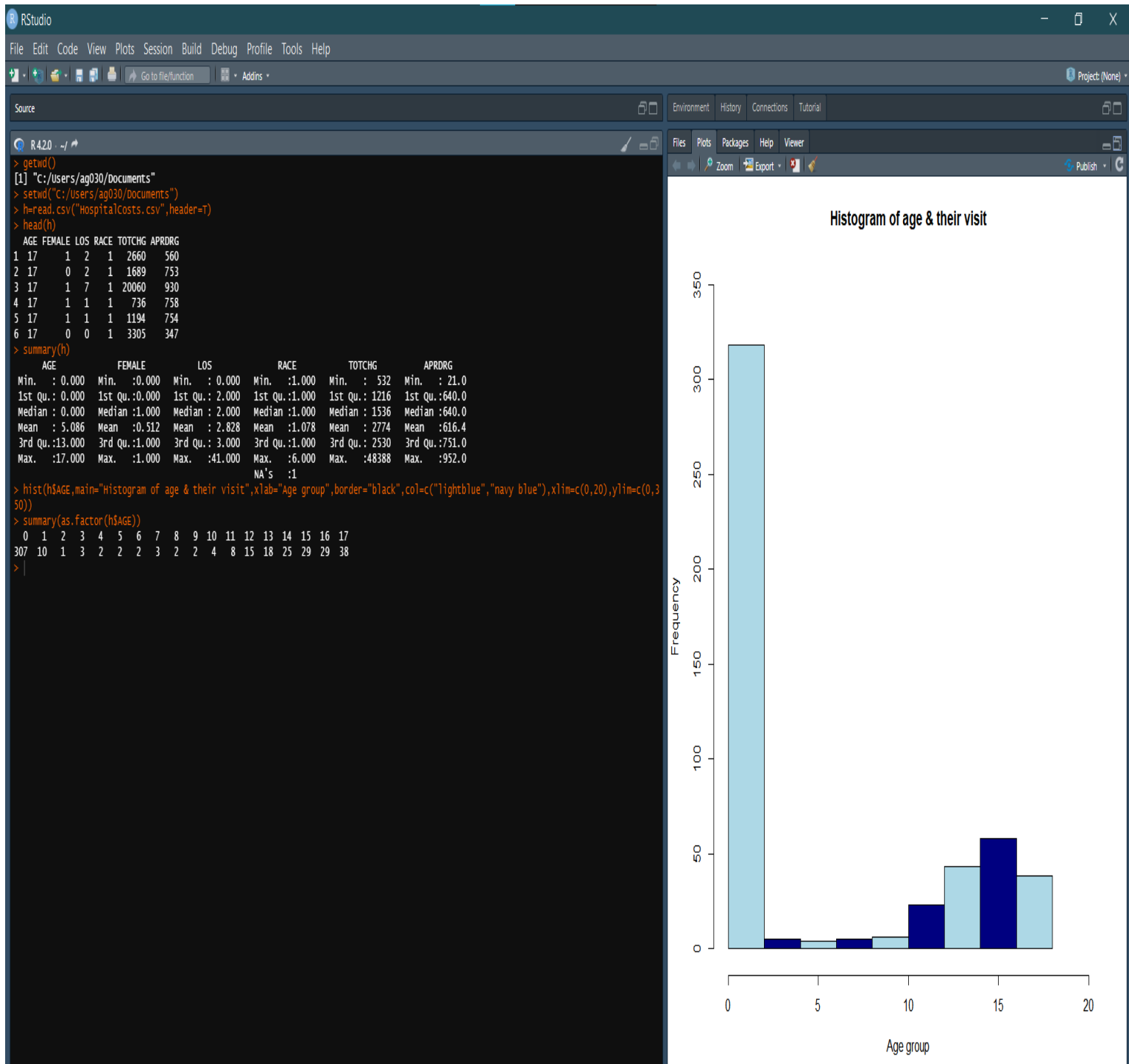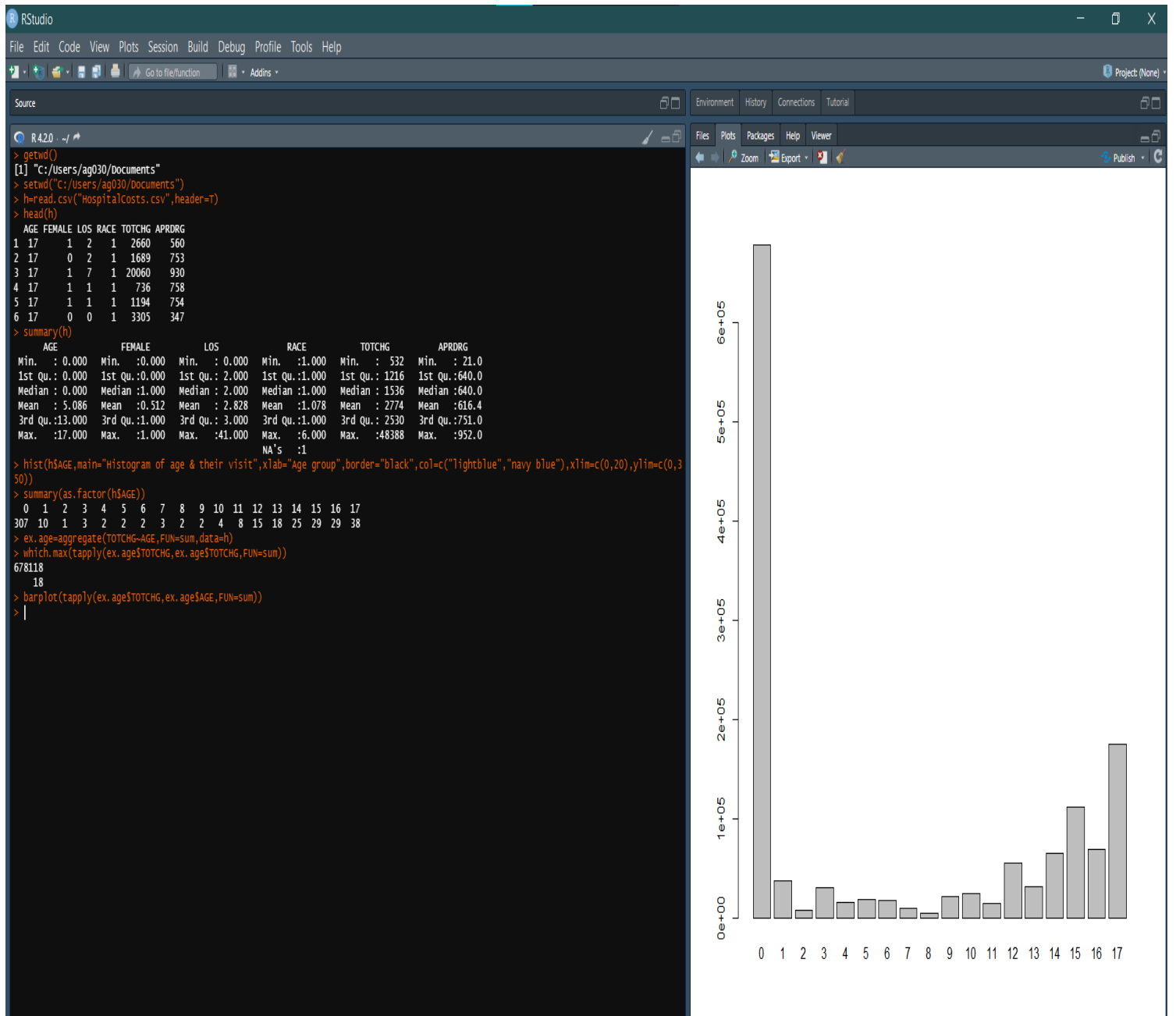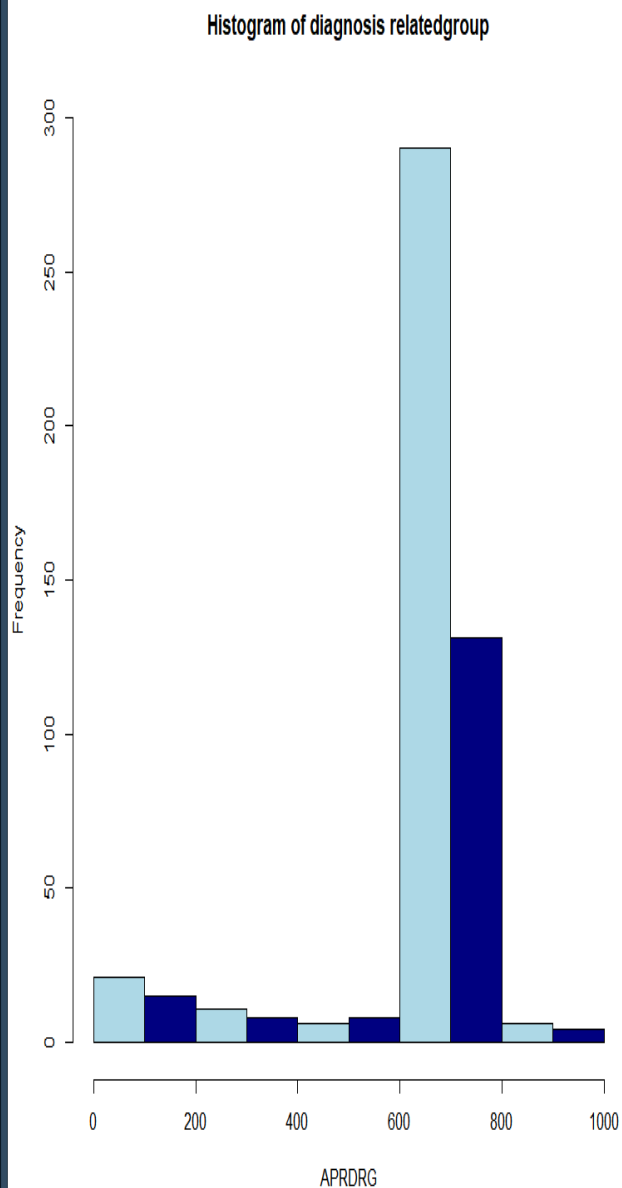
RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function   Addins ▾   Project: (None) ▾

Source   Environment History Connections Tutorial

R 4.2.0 · ~/   Files Plots Packages Help Viewer

```
> getwd()
[1] "C:/Users/ag030/Documents"
> setwd("C:/Users/ag030/Documents")
> h=read.csv("HospitalCosts.csv",header=T)
> head(h)
  AGE FEMALE LOS RACE TOTCHG APRDRG
1  17      1   2    1   2660    560
2  17      0   2    1   1689    753
3  17      1   7    1  20060    930
4  17      1   1    1    736    758
5  17      1   1    1   1194    754
6  17      0   0    1   3305    347
> summary(h)
      AGE            FEMALE           LOS             RACE           TOTCHG          APRDRG
 Min.   : 0.000   Min.   :0.000   Min.   : 0.000   Min.   :1.000   Min.   :  532   Min.   : 21.0
 1st Qu.: 0.000   1st Qu.:0.000   1st Qu.: 2.000   1st Qu.:1.000   1st Qu.: 1216   1st Qu.:640.0
 Median : 0.000   Median :1.000   Median : 2.000   Median :1.000   Median : 1536   Median :640.0
 Mean   : 5.086   Mean   :0.512   Mean   : 2.828   Mean   :1.078   Mean   : 2774   Mean   :616.4
 3rd Qu.:13.000   3rd Qu.:1.000   3rd Qu.: 3.000   3rd Qu.:1.000   3rd Qu.: 2530   3rd Qu.:751.0
 Max.   :17.000   Max.   :1.000   Max.   :41.000   Max.   :6.000   Max.   :48388   Max.   :952.0
                                                   NA's   :1
> hist(h$AGE,main="Histogram of age & their visit",xlab="Age group",border="black",col=c("lightblue","navy blue"),xlim=c(0,20),ylim=c(0,3
50))
> summary(as.factor(h$AGE))
  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17
307  10   1   3   2   2   2   3   2   2   4   8  15  18  25  29  29  38
> |
```

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins ▾

Project: (None) ▾

Source

Environment History Connections Tutorial

R 4.2.0 · ~/

Files Plots Packages Help Viewer

Zoom | Export ▾ | Publish ▾

```
> getwd()
[1] "C:/Users/ag030/Documents"
> setwd("C:/Users/ag030/Documents")
> h=read.csv("HospitalCosts.csv",header=T)
> head(h)
  AGE FEMALE LOS RACE TOTCHG APRDRG
1  17      1   2    1   2660    560
2  17      0   2    1   1689    753
3  17      1   7    1  20060    930
4  17      1   1    1    736    758
5  17      1   1    1   1194    754
6  17      0   0    1   3305    347
> summary(h)
      AGE            FEMALE           LOS             RACE           TOTCHG          APRDRG
 Min.   : 0.000   Min.   :0.000   Min.   : 0.000   Min.   :1.000   Min.   :  532   Min.   : 21.0
 1st Qu.: 0.000   1st Qu.:0.000   1st Qu.: 2.000   1st Qu.:1.000   1st Qu.: 1216   1st Qu.:640.0
 Median : 0.000   Median :1.000   Median : 2.000   Median :1.000   Median : 1536   Median :640.0
 Mean   : 5.086   Mean   :0.512   Mean   : 2.828   Mean   :1.078   Mean   : 2774   Mean   :616.4
 3rd Qu.:13.000   3rd Qu.:1.000   3rd Qu.: 3.000   3rd Qu.:1.000   3rd Qu.: 2530   3rd Qu.:751.0
 Max.   :17.000   Max.   :1.000   Max.   :41.000   Max.   :6.000   Max.   :48388   Max.   :952.0
                                                   NA's   :1
> hist(h$AGE,main="Histogram of age & their visit",xlab="Age group",border="black",col=c("lightblue","navy blue"),xlim=c(0,20),ylim=c(0,3
50))
> summary(as.factor(h$AGE))
  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17
307  10   1   3   2   2   2   3   2   2   4   8  15  18  25  29  29  38
> ex.age=aggregate(TOTCHG~AGE,FUN=sum,data=h)
> which.max(tapply(ex.age$TOTCHG,ex.age$TOTCHG,FUN=sum))
678118
    18
> barplot(tapply(ex.age$TOTCHG,ex.age$AGE,FUN=sum))
>
```

```
> getwd()
[1] "c:/Users/ag030/Documents"
> setwd("C:/Users/ag030/Documents")
> h=read.csv("HospitalCosts.csv",header=T)
> head(h)
  AGE FEMALE LOS RACE TOTCHG APRDRG
1  17      1   2    1   2660    560
2  17      0   2    1   1689    753
3  17      1   7    1  20060    930
4  17      1   1    1    736    758
5  17      1   1    1   1194    754
6  17      0   0    1   3305    347
> summary(h)
      AGE            FEMALE           LOS             RACE           TOTCHG          APRDRG
 Min.   : 0.000   Min.   :0.000   Min.   : 0.000   Min.   :1.000   Min.   :  532   Min.   : 21.0
 1st Qu.: 0.000   1st Qu.:0.000   1st Qu.: 2.000   1st Qu.:1.000   1st Qu.: 1216   1st Qu.:640.0
 Median : 0.000   Median :1.000   Median : 2.000   Median :1.000   Median : 1536   Median :640.0
 Mean   : 5.086   Mean   :0.512   Mean   : 2.828   Mean   :1.078   Mean   : 2774   Mean   :616.4
 3rd Qu.:13.000   3rd Qu.:1.000   3rd Qu.: 3.000   3rd Qu.:1.000   3rd Qu.: 2530   3rd Qu.:751.0
 Max.   :17.000   Max.   :1.000   Max.   :41.000   Max.   :6.000   Max.   :48388   Max.   :952.0
                                                   NA's   :1
> hist(h$AGE,main="Histogram of age & their visit",xlab="Age group",border="black",col=c("lightblue","navy blue"),xlim=c(0,20),ylim=c(0,3
50))
> summary(as.factor(h$AGE))
  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17
307  10   1   3   2   2   2   2   4   8  15  18  25  29  29  38
> ex.age=aggregate(TOTCHG~AGE,FUN=sum,data=h)
> which.max(tapply(ex.age$TOTCHG,ex.age$TOTCHG,FUN=sum))
678118
    18
> barplot(tapply(ex.age$TOTCHG,ex.age$AGE,FUN=sum))
> hist(h$APRDRG,main="Histogram of diagnosis relatedgroup",xlab="APRDRG",border="black",col=c("light blue","navy blue"),xlim=c(0,1000),yl
im=c(0,300))
> summary(as.factor(h$APRDRG))
 21  23  49  50  51  53  54  57  58  92  97 114 115 137 138 139 141 143 204 206 225 249 254 308 313 317 344 347 420 421 422 560
  1   1   1   1   1  10   1   2   1   1   1   1   2   1   4   5   1   1   1   2   6   1   1   1   1   2   3   2   1   3   2
561 566 580 581 602 614 626 633 634 636 639 640 710 720 723 740 750 751 753 754 755 756 758 760 776 811 812 863 911 930 952
  1   1   1   1   3   6   4   2   3   4 267   1   1   2   1   1  14  36  37  13   2  20   2   1   2   3   1   1   2   1
> diag.cost=aggregate(TOTCHG~APRDRG,FUN=sum,data=h)
> diag.cost[which.max(diag.cost$TOTCHG),]
   APRDRG TOTCHG
44    640 437978
> |
```

Histogram of diagnosis relatedgroup

```
> summary(h)
      AGE             FEMALE           LOS             RACE           TOTCHG          APRDRG
 Min.   : 0.000   Min.   :0.000   Min.   : 0.000   Min.   :1.000   Min.   :  532   Min.   : 21.0
 1st Qu.: 0.000   1st Qu.:0.000   1st Qu.: 2.000   1st Qu.:1.000   1st Qu.: 1216   1st Qu.:640.0
 Median : 0.000   Median :1.000   Median : 2.000   Median :1.000   Median : 1536   Median :640.0
 Mean   : 5.086   Mean   :0.512   Mean   : 2.828   Mean   :1.078   Mean   : 2774   Mean   :616.4
 3rd Qu.:13.000   3rd Qu.:1.000   3rd Qu.: 3.000   3rd Qu.:1.000   3rd Qu.: 2530   3rd Qu.:751.0
 Max.   :17.000   Max.   :1.000   Max.   :41.000   Max.   :6.000   Max.   :48388   Max.   :952.0
                                                   NA's   :1
> hist(h$AGE,main="Histogram of age & their visit",xlab="Age group",border="black",col=c("lightblue","navy blue"),xlim=c(0,20),ylim=c
(0,350))
> summary(as.factor(h$AGE))
  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17
307  10   1   3   2   2   2   3   2   4   8  15  18  25  29  29  38
> ex.age=aggregate(TOTCHG~AGE,FUN=sum,data=h)
> which.max(tapply(ex.age$TOTCHG,ex.age$TOTCHG,FUN=sum))
678118
    18
> barplot(tapply(ex.age$TOTCHG,ex.age$AGE,FUN=sum))
> hist(h$APRDRG,main="Histogram of diagnosis relatedgroup",xlab="APRDRG",border="black",col=c("light blue","navy blue"),xlim=c(0,1000),
ylim=c(0,300))
> summary(as.factor(h$APRDRG))
 21  23  49  50  51  53  54  57  58  92  97 114 115 137 138 139 141 143 204 206 225 249 254 308 313 317 344 347 420 421 422 560
  1   1   1   1   1  10   1   2   1   1   1   1   2   1   4   5   1   1   1   1   2   6   1   1   1   1   2   3   2   1   3   2
561 566 580 581 602 614 626 633 634 636 639 640 710 720 723 740 750 751 753 754 755 756 758 760 776 811 812 863 911 930 952
  1   1   1   3   1   3   6   4   2   3   4 267   1   1   2   1   1  14  36  37  13   2  20   2   1   2   3   1   1   2   1
> diag.cost=aggregate(TOTCHG~APRDRG,FUN=sum,data=h)
> diag.cost[which.max(diag.cost$TOTCHG),]
   APRDRG TOTCHG
44    640 437978
> summary(as.factor(h$RACE))
  1   2   3   4   5   6 NA's
484   6   1   3   3   2   1
> h=na.omit(h)
> summary(as.factor(h$RACE))
  1   2   3   4   5   6
484   6   1   3   3   2
> reg1=lm(TOTCHG~RACE,data=h)
> summary(reg1)

Call:
lm(formula = TOTCHG ~ RACE, data = h)

Residuals:
   Min     1Q Median     3Q    Max
 -2256  -1560  -1227   -258  45600

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2925.7      405.0   7.224 1.92e-12 ***
RACE          -137.3      339.1  -0.405    0.686
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3895 on 497 degrees of freedom
Multiple R-squared:  0.0003299,	Adjusted R-squared:  -0.001681
F-statistic: 0.164 on 1 and 497 DF,  p-value: 0.6856
```

```
R 4.2.0 · ~/

> reg1=lm(TOTCHG~RACE,data=h)
> summary(reg1)

Call:
lm(formula = TOTCHG ~ RACE, data = h)

Residuals:
    Min     1Q Median     3Q    Max
 -2256  -1560  -1227   -258  45600

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    2925.7      405.0   7.224 1.92e-12 ***
RACE           -137.3      339.1  -0.405    0.686
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3895 on 497 degrees of freedom
Multiple R-squared:  0.0003299,  Adjusted R-squared:  -0.001681
F-statistic: 0.164 on 1 and 497 DF,  p-value: 0.6856

> anova1=aov(TOTCHG~RACE,data=h)
> summary(anova1)
             Df    Sum Sq  Mean Sq F value Pr(>F)
RACE          1 2.488e+06  2488459   0.164  0.686
Residuals   497 7.540e+09 15170268
>
```

```
306  10  1  3  2  2  2  3  2  2  4  8  15  18  25  29  29  38
> summary(as.factor(h$FEMALE))
  0   1
244 255
> reg2=lm(TOTCHG~AGE+FEMALE,data=h)
> summary(reg2)

Call:
lm(formula = TOTCHG ~ AGE + FEMALE, data = h)

Residuals:
   Min    1Q Median    3Q    Max
 -3403  -1444   -873   -156  44950

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2719.45     261.42  10.403  < 2e-16 ***
AGE            86.04      25.53   3.371 0.000808 ***
FEMALE       -744.21     354.67  -2.098 0.036382 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3849 on 496 degrees of freedom
Multiple R-squared:  0.02585,   Adjusted R-squared:  0.02192
F-statistic: 6.581 on 2 and 496 DF,  p-value: 0.001511

>
```

```
Multiple R-squared:  0.02585,    Adjusted R-squared:  0.02192
F-statistic: 6.581 on 2 and 496 DF,  p-value: 0.001511

> reg3=lm(LOS~AGE+FEMALE+RACE,data=h)
> summary(reg3)

Call:
lm(formula = LOS ~ AGE + FEMALE + RACE, data = h)

Residuals:
   Min    1Q Median    3Q    Max
 -3.22  -1.22  -0.85   0.15  37.78

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.94377    0.39318   7.487 3.25e-13 ***
AGE         -0.03960    0.02231  -1.775   0.0766 .
FEMALE       0.37011    0.31024   1.193   0.2334
RACE        -0.09408    0.29312  -0.321   0.7484
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.363 on 495 degrees of freedom
Multiple R-squared:  0.007898,   Adjusted R-squared:  0.001886
F-statistic: 1.314 on 3 and 495 DF,  p-value: 0.2692

>
```

```
R 4.2.0 · ~/

> model=lm(TOTCHG~.,data=h)
> summary(model)

Call:
lm(formula = TOTCHG ~ ., data = h)

Residuals:
   Min      1Q  Median      3Q     Max
 -6377    -700    -174     122   43378

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 5218.6769   507.6475  10.280  < 2e-16 ***
AGE          134.6949    17.4711   7.710 7.02e-14 ***
FEMALE      -390.6924   247.7390  -1.577    0.115
LOS          743.1521    34.9225  21.280  < 2e-16 ***
RACE        -212.4291   227.9326  -0.932    0.352
APRDRG        -7.7909     0.6816 -11.430  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2613 on 493 degrees of freedom
Multiple R-squared:  0.5536,    Adjusted R-squared:  0.5491
F-statistic: 122.3 on 5 and 493 DF,  p-value: < 2.2e-16

>
```

```
F-statistic: 122.3 on 5 and 493 DF,  p-value: < 2.2e-16

> hcm=lm(TOTCHG~AGE+FEMALE+LOS+APRDRG,data=h)
> summary(hcm)

Call:
lm(formula = TOTCHG ~ AGE + FEMALE + LOS + APRDRG, data = h)

Residuals:
   Min    1Q Median    3Q    Max
 -6344   -687   -168    132  43387

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 4971.980    433.116  11.480  < 2e-16 ***
AGE          134.241     17.462   7.688 8.16e-14 ***
FEMALE      -383.082    247.571  -1.547    0.122
LOS          743.618     34.914  21.298  < 2e-16 ***
APRDRG        -7.767      0.681 -11.405  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2613 on 494 degrees of freedom
Multiple R-squared:  0.5528,    Adjusted R-squared:  0.5492
F-statistic: 152.7 on 4 and 494 DF,  p-value: < 2.2e-16

>
```

```
Multiple R-squared:  0.5528,    Adjusted R-squared:  0.5492
F-statistic: 152.7 on 4 and 494 DF,  p-value: < 2.2e-16


> hcm1=lm(TOTCHG~AGE+LOS+APRDRG,data=h)
> summary(hcm1)

Call:
lm(formula = TOTCHG ~ AGE + LOS + APRDRG, data = h)

Residuals:
   Min     1Q Median     3Q    Max
 -6603   -719   -169    124  43350

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 4960.1705   433.6579   11.44  < 2e-16 ***
AGE          128.5519    17.0946    7.52 2.59e-13 ***
LOS          740.8057    34.9161   21.22  < 2e-16 ***
APRDRG        -8.0055     0.6643  -12.05  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2617 on 495 degrees of freedom
Multiple R-squared:  0.5506,    Adjusted R-squared:  0.5479
F-statistic: 202.2 on 3 and 495 DF,  p-value: < 2.2e-16

>
```

```
R 4.2.0 · ~/
```

Residual standard error: 2617 on 495 degrees of freedom
Multiple R-squared: 0.5506,   Adjusted R-squared: 0.5479
F-statistic: 202.2 on 3 and 495 DF, p-value: < 2.2e-16

```
> hcm2=lm(TOTCHG~AGE+LOS,data=h)
> summary(hcm2)
```

Call:
lm(formula = TOTCHG ~ AGE + LOS, data = h)

Residuals:
   Min   1Q Median   3Q   Max
 -4783   -1103   -458   -133   41382

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   200.66   203.48   0.986   0.325
AGE   97.96   19.21   5.101 4.83e-07 ***
LOS   734.27   39.66 18.512 < 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2973 on 496 degrees of freedom
Multiple R-squared: 0.4188,   Adjusted R-squared: 0.4164
F-statistic: 178.7 on 2 and 496 DF, p-value: < 2.2e-16

> |