

A.S.A Lab Assignment

6

Name: Swanand Garge

Div : D(D2)

Roll no :42

SRN: 202201589

Q.To create two way cross tabulations to explore the relationship between several variables and to use the Chart Builder to visualize the relationship Using the preexisting Census.csv data file

```
data <- read.csv(file.choose())
View(data)

class(data)

View(data)

str(data)

tab1 <- table(data$workclass, data$income)
tab1
```

```
tab2 <- table(data$sex, data$income)
tab2

margin.table(tab1, 1)
margin.table(tab1, 2)

prop.table(tab1)

prop.table(tab1, 1)

prop.table(tab1, 2)

margin.table(tab2, 1)
margin.table(tab2, 2)

prop.table(tab2)

prop.table(tab2, 1)

prop.table(tab2, 2)

library(ggplot2)

df1 <- data.frame(tab1)
colnames(df1) <- c("WorkClass", "Income", "Freq")
df1

ggplot(df1, aes(x = WorkClass, y = Freq, fill = Income)) +
  geom_col()

ggplot(df1, aes(x = WorkClass, y = Freq, color = Income, group = Income)) +
  geom_line()

df2 <- data.frame(tab2)
colnames(df2) <- c("Gender", "Income", "Freq")
df2

ggplot(df2, aes(x = Gender, y = Freq, fill = Income)) +
```

```
geom_col()
```

Output :

```
> class(data)
[1] "data.frame"
>
> View(data)
>
> str(data)
'data.frame':   32561 obs. of  15 variables:
 $ age       : int   39 50 38 53 28 37 49 52 31 42 ...
 $ workclass : chr   " State-gov" " Self-emp-not-inc" " Private" " Private" ...
 $ final.weight : int  77516 83311 215646 234721 338409 284582 160187 209642 45781 159449 ...
 $ education : chr   " Bachelors" " Bachelors" " HS-grad" " 11th" ...
 $ education.num : int   13 13 9 7 13 14 5 9 14 13 ...
 $ marital.status: chr   " Never-married" " Married-civ-spouse" " Divorced" " Married-civ-spouse" ...
 $ occupation   : chr   " Adm-clerical" " Exec-managerial" " Handlers-cleaners" " Handlers-cleaners" ...
 $ relationship : chr   " Not-in-family" " Husband" " Not-in-family" " Husband" ...
 $ race         : chr   " White" " White" " White" " Black" ...
 $ sex         : chr   " Male" " Male" " Male" " Male" ...
 $ capital.gain : int   2174 0 0 0 0 0 0 0 0 14084 5178 ...
 $ capital.loos : int    0 0 0 0 0 0 0 0 0 ...
 $ hour.per.week : int   40 13 40 40 40 40 16 45 50 40 ...
 $ native.country: chr   " United-States" " United-States" " United-States" " United-States" ...
 $ income       : chr   " <=50K" " <=50K" " <=50K" " <=50K" ...
>
```

```
> tab1 <- table(data$workclass, data$income)
> tab1

      <=50K >50K
?          1645 191
Federal-gov 589 371
Local-gov   1476 617
Never-worked 7 0
Private     17733 4963
Self-emp-inc 494 622
Self-emp-not-inc 1817 724
State-gov   945 353
Without-pay 14 0
>
> tab2 <- table(data$sex, data$income)
> tab2

      <=50K >50K
Female  9592 1179
Male    15128 6662
>
> margin.table(tab1, 1)

      ?      Federal-gov      Local-gov      Never-worked      Private      Self-emp-inc      Self-emp-not-inc
      1836      960      2093      7      22696      1116      2541
      State-gov      Without-pay
      1298      14
> margin.table(tab1, 2)

<=50K >50K
24720 7841
>
```

```
> prop.table(tab1)
      <=50K      >50K
?      0.0505205614 0.0058659132
Federal-gov 0.0180891250 0.0113939990
Local-gov   0.0453303031 0.0189490495
Never-worked 0.0002149811 0.0000000000
Private     0.5446085808 0.1524216087
Self-emp-inc 0.0151715242 0.0191026074
Self-emp-not-inc 0.0558029545 0.0222351893
State-gov   0.0290224502 0.0108411904
Without-pay 0.0004299622 0.0000000000
>
>
> prop.table(tab1, 1)
      <=50K      >50K
?      0.8959695 0.1040305
Federal-gov 0.6135417 0.3864583
Local-gov   0.7052078 0.2947922
Never-worked 1.0000000 0.0000000
Private     0.7813271 0.2186729
Self-emp-inc 0.4426523 0.5573477
Self-emp-not-inc 0.7150728 0.2849272
State-gov   0.7280431 0.2719569
Without-pay 1.0000000 0.0000000
>
>
```

```
> prop.table(tab1, 2)
      <=50K      >50K
?      0.0665453074 0.0243591379
Federal-gov 0.0238268608 0.0473153934
Local-gov   0.0597087379 0.0786889427
Never-worked 0.0002831715 0.0000000000
Private     0.7173543689 0.6329549802
Self-emp-inc 0.0199838188 0.0793266165
Self-emp-not-inc 0.0735032362 0.0923351613
State-gov   0.0382281553 0.0450197679
Without-pay 0.0005663430 0.0000000000
>
>
> margin.table(tab2, 1)
Female      Male
10771      21790
> margin.table(tab2, 2)
      <=50K      >50K
24720      7841
>
>
> prop.table(tab2)
      <=50K      >50K
Female 0.29458555 0.03620896
Male   0.46460490 0.20460060
>
>
```

```
> prop.table(tab2, 1)
      <=50K      >50K
Female 0.8905394 0.1094606
Male   0.6942634 0.3057366
>
>
> prop.table(tab2, 2)
      <=50K      >50K
Female 0.3880259 0.1503635
Male   0.6119741 0.8496365
>
>
```

```

> library(ggplot2)
>
>
> df1 <- data.frame(tab1)
> colnames(df1) <- c("WorkClass", "Income", "Freq")
> df1

```

	WorkClass	Income	Freq
1	?	<=50K	1645
2	Federal-gov	<=50K	589
3	Local-gov	<=50K	1476
4	Never-worked	<=50K	7
5	Private	<=50K	17733
6	Self-emp-inc	<=50K	494
7	Self-emp-not-inc	<=50K	1817
8	State-gov	<=50K	945
9	Without-pay	<=50K	14
10	?	>50K	191
11	Federal-gov	>50K	371
12	Local-gov	>50K	617
13	Never-worked	>50K	0
14	Private	>50K	4963
15	Self-emp-inc	>50K	622
16	Self-emp-not-inc	>50K	724
17	State-gov	>50K	353
18	Without-pay	>50K	0

```

> ggplot(df1, aes(x = WorkClass, y = Freq, fill = Income)) +
+   geom_col()
>
>
> ggplot(df1, aes(x = WorkClass, y = Freq, color = Income, group = Income)) +
+   geom_line()
>
>
> df2 <- data.frame(tab2)
> colnames(df2) <- c("Gender", "Income", "Freq")
> df2

```

	Gender	Income	Freq
1	Female	<=50K	9592
2	Male	<=50K	15128
3	Female	>50K	1179
4	Male	>50K	6662

```

>
> ggplot(df2, aes(x = Gender, y = Freq, fill = Income)) +
+   geom_col()
> |

```

Final Graph:

