Panel Survey of Income Dynamics

Group D

Executive Summery

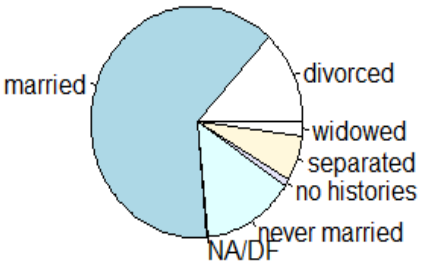
```
     Seq.No          intnum          persnum            age           educatn
Min.   :   1   Min.   :   4   Min.   :  1.00   Min.   :30.00   Min.   : 0.00
1st Qu.:1215   1st Qu.:1905   1st Qu.:  2.00   1st Qu.:34.00   1st Qu.:12.00
Median :2428   Median :5464   Median :  4.00   Median :38.00   Median :12.00
Mean   :2428   Mean   :4598   Mean   : 59.21   Mean   :38.46   Mean   :16.38
3rd Qu.:3642   3rd Qu.:6655   3rd Qu.:170.00   3rd Qu.:43.00   3rd Qu.:14.00
Max.   :4856   Max.   :9306   Max.   :205.00   Max.   :50.00   Max.   :99.00
                                                               NA's   :1
    earnings          hours            kids                 married
Min.   :     0   Min.   :   0   Min.   : 0.000   divorced     : 645
1st Qu.:    85   1st Qu.:  32   1st Qu.: 1.000   married      :3071
Median : 11000   Median :1517   Median : 2.000   NA/DF        :   9
Mean   : 14245   Mean   :1235   Mean   : 4.481   never married: 681
3rd Qu.: 22000   3rd Qu.:2000   3rd Qu.: 3.000   no histories :  43
Max.   :240000   Max.   :5160   Max.   :99.000   separated    : 317
                                                 widowed      :  90
```

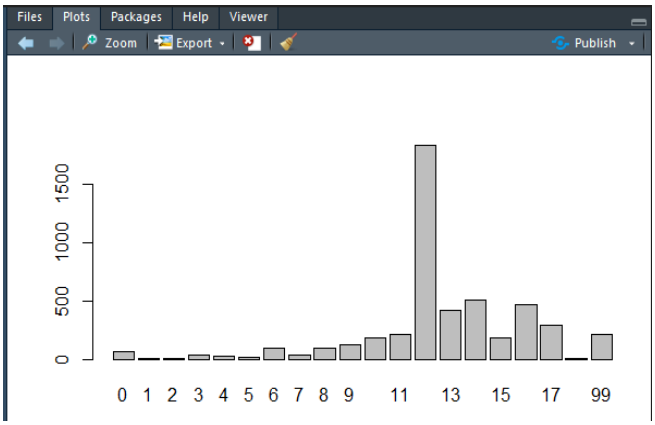
Above summary depicts the statistical attributes related to Panel Survey of Income Dynamics.

For this survey, there were 4856 number of observations were taken, Those breakup as

Status	Count of people
divorced	645
married	3071
NA/DF	9
never married	681
no histories	43
separated	317
widowed	90
Grand Total	**4856**

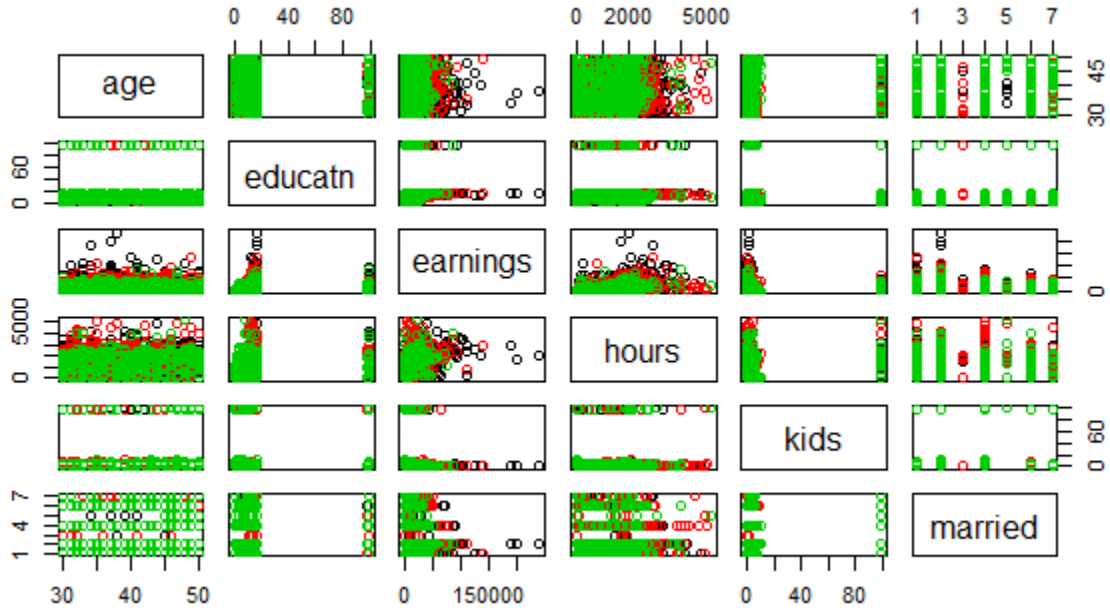
Further, the considered age range is 30 Years ≤ Age ≤ 50 Years. Median age is 38 years old. They were average annual working hrs as 1235.34, based on that, they were able to earn $ 14244.51 per annum. Maximum earning value is $ 240,000 and it is in the 38Y age group. Further analyzing the people education level, there are 216 people in highest education grade completed and majority of 1837 people were completed the 12 education grades. Pls see the below bar chart. The highest divorced age group is 43 years old category there were 46 people.





When we analyzed and applied the K-Means clustering technique to given data set for following columns we were able to get the following chart.

age-age of individual	educatn-highest grade completed	earnings-total labor income	hours-annual work hours	kids	married



Above graph depicts the clustering of the given data set in to 3 clusters and this represents a clear correlation between the hours and age.

Remarks for insight generation:

```
na.omit(psidData)
psidData1 <- na.omit(psidData)
psidData <- subset(psidData, psidData$age >= 40 & psidData$age <= 50)
plot(psidData)
counts <- table(psidData$age)
barplot(counts, main="Highest Education Distribution Vs Age",          xlab="Age")
```

Above code segment would perform the data refinement which means omitting missing values from the data set and import as a new data set.

This histogram depicts the frequency of divorced and no histories frequency as well as the frequency of kids.

s

GIT Link

https://github.com/Team101112/CS5122---Descriptive-and-Predictive-Analytics

Resources allocation Chart.

Group D

Group member	Member Name	Contribution
189105X	I.R.N.P.Gnanaratne	Analyze the data and prepare the document.
189120M	P.L.Sajeewanie	Analyze the data and prepare the document.
189107F	A.A.P.C.Harischandra	Analyze the data and prepare the document.