Analytics of Business Intelligence - Practice Set # 2 Extra

Using the csv file espn.csv answer these questions using data.table, the answers are below.

1. **On average, which Location has the highest Income?**
2. **On average, which Location has the lowest Income?**
3. **What percentage of the data is men? Women?**
4. **What is the highest income for women at each Location (use by)?**
5. **What is the highest income for men at each Location (use by)?**
6. **Between the ages of men from 18 to 64, which age has the highest average income?**
7. **For men older than 65, which age has the highest average income?**
8. **Between the ages of men from 18 to 64, which age has the highest average income, by each location (use by)?**
9. **In which location are there more women living, between the ages of 19 and 29 (use by)?**

Answers:

1. **suburban**

```
espn[,mean(Income),by=Location])
```

2. **urban**

3. **0.796875 0.203125**

```
y = espn[,.N,by=Gender]
y[,N/sum(N)]
```

4. **rural 128, urban 240, suburban 162**

```
espn[Gender=='f',max(Income),by=Location]
```

5. **rural 202, suburban 259, urban 189**

```
 espn[Gender=='m',max(Income),by=Location]
```

6. **age 59, income 139**

```
y = espn[Gender=='m' & Age >= 18 & Age <=64,mean(Income),by=Age]
y = y[order(-V1)]
y[1]
```

7. **age 113, income 203**

```
y = espn[Gender=='m' & Age > 65, mean(Income),by=Age]
y = y[order(-V1)]
y[1]
```

8. **rural 48 115.0000, suburban 53 141.3333, urban 44 115.2500**

```
y = espn[Gender=='m' & Age >= 18 & Age <=64,mean(Income),by=.(Age,Location
    )]
y = y[order(Location,-V1)]
y[,head(.SD,1),by=Location]
```

9. **urban 22, suburban 26, rural 12**

```
 espn[Gender=='f' & Age >= 19 & Age <=29,.N,by=Location]
```