

## **DATA ACTIVITY 5 – HEALTH DATA**

### **Task 5.1 – Find out the mean, median, and mode of variables sbp, dbp, and income.**

```
#Calculate mean, median, and mode of sbp, dbp, income
#sbp
mean(Health_Data$sbp)
median(Health_Data$sbp)
mode_sbp <- as.numeric(names(sort(table(Health_Data$sbp), decreasing=TRUE)[1]))
show(mode_sbp)

#dbp
mean(Health_Data$dbp)
median(Health_Data$dbp)
mode_dbp<-as.numeric(names(sort(table(Health_Data$dbp), decreasing=TRUE)[1]))
show(mode_dbp)

#income
mean(Health_Data$income)
median(Health_Data$income)
mode_income<-as.numeric(names(sort(table(Health_Data$income),decreasing=TRUE)[1]))
show(mode_income)
```

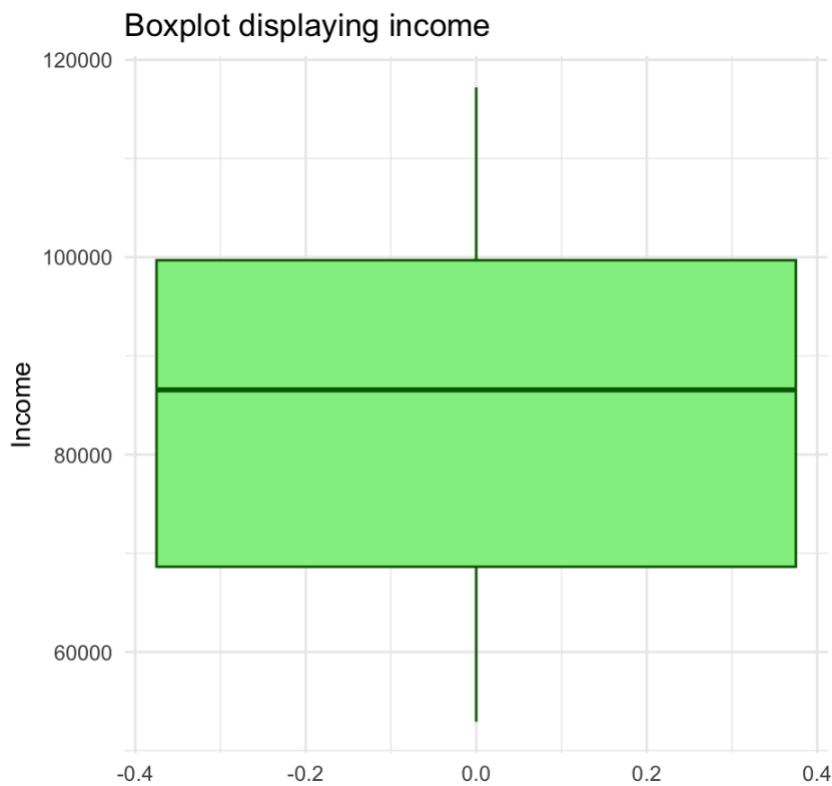
```
> #Calculate mean, median, and mode of sbp, dbp, income
> #sbp
> mean(Health_Data$sbp)
[1] 127.7333
> median(Health_Data$sbp)
[1] 123
> mode_sbp <- as.numeric(names(sort(table(Health_Data$sbp), decreasing=TRUE) [1]))
> show(mode_sbp)
[1] 120
> mean(Health_Data$dbp)
[1] 82.76667
> median(Health_Data$dbp)
[1] 82
> mode_dbp<-as.numeric(names(sort(table(Health_Data$dbp), decreasing=TRUE)[1]))
> show(mode_dbp)
[1] 74
> mean(Health_Data$income)
[1] 85194.49
> median(Health_Data$income)
[1] 86560.5
> mode_income<-as.numeric(names(sort(table(Health_Data$income),decreasing=TRUE)[1]))
> show(mode_income)
[1] 52933
```

### **Task 5.2 – Find out the five-figure summary of income and present it as a boxplot.**

```
#5-figure summary of income & boxplot
summary(Health_Data$income)

ggplot(Health_Data, aes(y=income))+
  geom_boxplot(fill="lightgreen",color="darkgreen")+
  labs(title="Boxplot displaying income",y="Income")+
  theme_minimal()

> summary(Health_Data$income)
   Min. 1st Qu. Median      Mean 3rd Qu.      Max.
52933    68636    86560    85194    99696   117210
```



**Task 5.3 – Run a suitable hypothesis test to see if there is any association between systolic blood pressure and presence & absence of peptic ulcer.**

```
#Run independent samples t-test between sbp & peptic ulcer
t.test(sbp~pepticulcer,data=Health_Data)
```

```
data: sbp by pepticulcer
t = 1.2142, df = 57.562, p-value = 0.2296
alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
95 percent confidence interval:
-2.889367 11.795703
sample estimates:
mean in group 1 mean in group 2
131.3171      126.8639
```