BODY ANALYSIS

USING BOX-PLOTS & FREQUENCY TABLE

Libraries used-
1. summarytools
2. ggplot2

KRITIK SETH
J043

Question 1

Divide the dataset on the basis of age in two parts and create a frequency table-

1. <=40

2. >40

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
Age Above 40	159	63.35	63.35	63.35	63.35
Age Under 40	92	36.65	100.00	36.65	100.00
<NA>	0			0.00	100.00
Total	251	100.00	100.00	100.00	100.00

```
BF<-read.csv("/Users/kritik/Desktop/B.Tech Data Science/Semester 3/Managing Uncertainity/Required Excel Files/BODYFAT.csv")
library('summarytools')
BF$AgeGroup<-ifelse(BF$AGE<=40,"Age Under 40","Age Above 40")
freq(BF$AgeGroup)
```

We can infer the following from the above table

- **The number of people in dataset aged above 40 are 159.**
- **63.35 % of the people are more than 40 years old.**
- **The number of people in dataset aged less than or equal to 40 are 92.**
- **36.65 % of the people are less than or equal to 40 years old.**
- **Total number of people are 251.**
- **People who don't fall in either of the categories are 0 (NA).**

Question 2

Create a box plot based on age for each variable

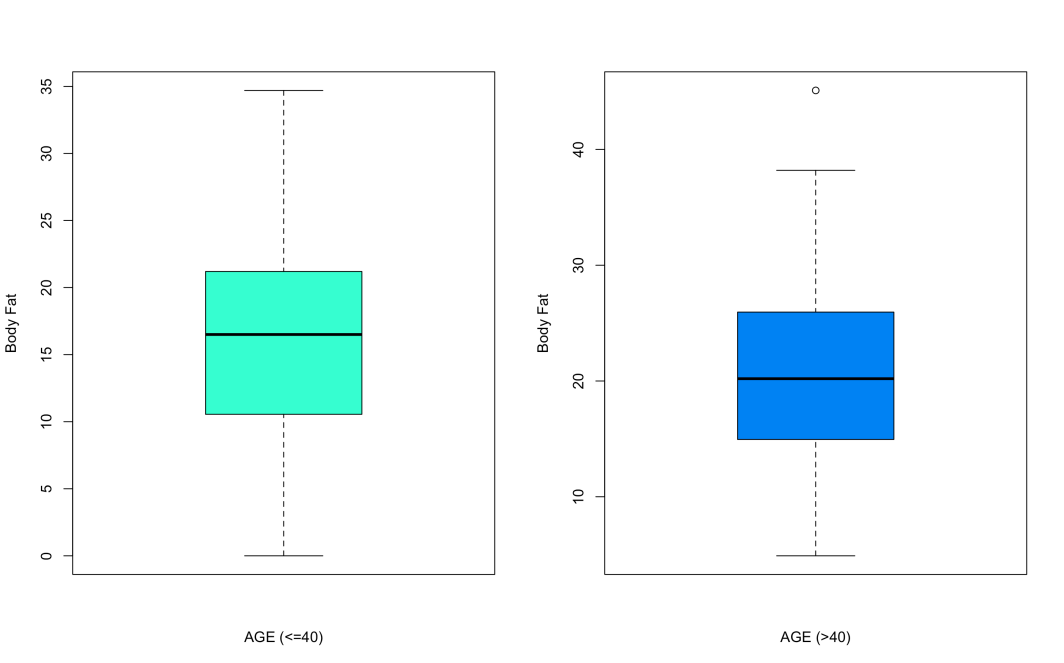
Question 3

Is there any change in age group based on each variable

GENERAL CODE COMMON FOR ALL

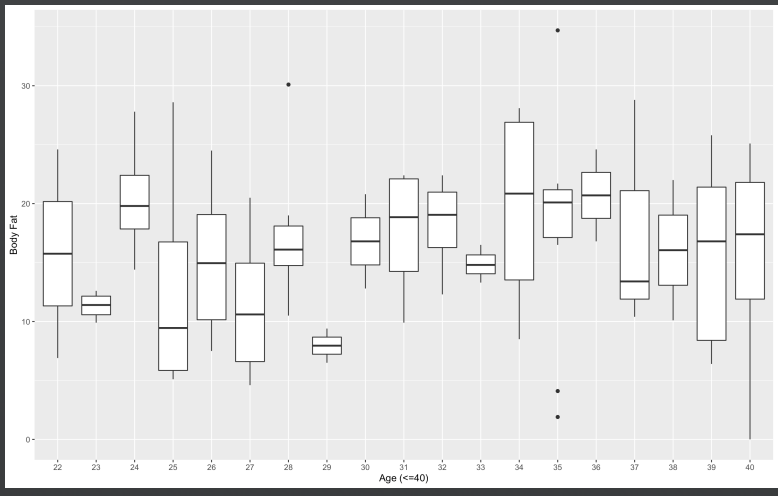
```
file.choose()
BF<-read.csv("/Users/kritik/Desktop/B.Tech Data Science/Semester 3/Managing
Uncertainity/Required Excel Files/BODYFAT.csv")
BFU40<-subset(BF,AGE<=40,select=IDNO:WRIST) #UNDER 40
BFA40<-subset(BF,AGE>40,select=IDNO:WRIST) #ABOVE 40
library('summarytools')
library(ggplot2)
```

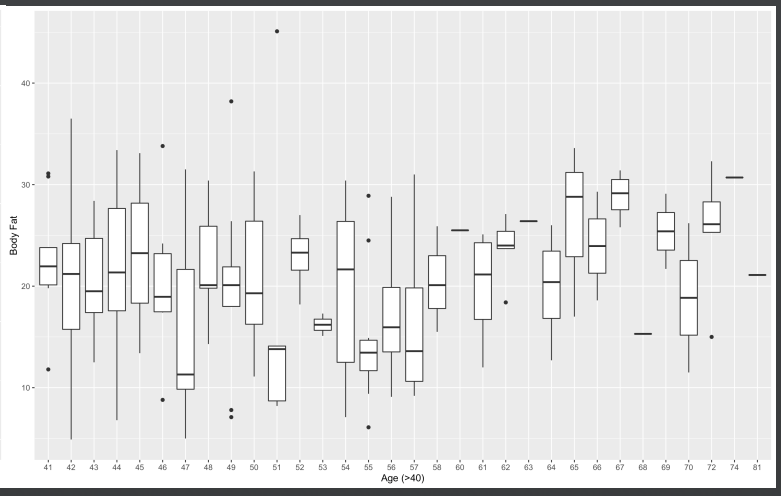
Body Fat




```
par(mfrow=c(1,2))
boxplot(BFU40$BODYFAT,xlab="AGE (<=40)", ylab="Body Fat",col="aquamarine1")
boxplot(BFA40$BODYFAT,xlab="AGE (>40)", ylab="Body Fat", col="dodgerblue2")
```

- **The mean for AGE <= 40 is between 15 and 20 while the mean for AGE > 40 is around 20 which means that people younger than 40 usually have body fat between 15 and 20 while people older than 40 have body fat around 20.**
- **The box plot for AGE <= 40 is a little negatively skewed while the box plot for AGE > 40 is a little positively skewed.**
- **For AGE <=40, 25 % of people (between minimum and first quartile) have body fat between 0 and 10 while other 25 % of people (between third quartile and maximum) have a body fat between 20 and 35.**
- **For AGE > 40, 25 % of the people (between minimum and first quartile) have body fat between 0 and 15 while other 25 % of the people (between third quartile and maximum) have body fat between 25 and 40.**
- **One person having AGE > 40 has abnormally large body fat.**
- **People of AGE <= 40 have a more consistent distribution of body fat amongst different age groups because their Interquartile range roughly lies between 10 and 20 while the full range lies between 0 and 35**
- **People of AGE > 40 have a less consistent distribution of body fat amongst different age groups because their Interquartile range lies roughly between 15 and 25 while the full range lies between 0 and 45 (including the outlier).**






```
ggplot(BFU40,
aes(as.factor(AGE),BODYFAT), fill=AGE) +
geom boxplot() +
```

```
ggplot(BFA40, aes(as.factor(AGE),BODYFAT),
fill=AGE) + geom_boxplot() +
labs(x = "Age (>40)", y = "Body Fat")
```

- **Frome above box-plots-**
- **Body Fats of people of ages 23, 29, 33, 53, 68, 74 & 81 are nearly constant.**
- **Body Fats of people aged 22, 25, 26, 27, 34, 39, 40, 44, 47, 50 & 54 vary a lot**
- **Some people aged 28, 35, 41, 46, 49, 52, 55, 62, 72 are outliers meaning some people aged 28, 35, 41, 46, 49, 55 have abnormally large body fat and some people aged 35, 41, 46, 49, 55, 61, 72 have abnormally low body fat.**

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
BodyFat_AgeUnder_40	0.0	10.575	16.5	16.09674	21.10	34.7
BodyFat_AgeAbove_40	4.9	14.950	20.2	20.50252	25.95	45.1

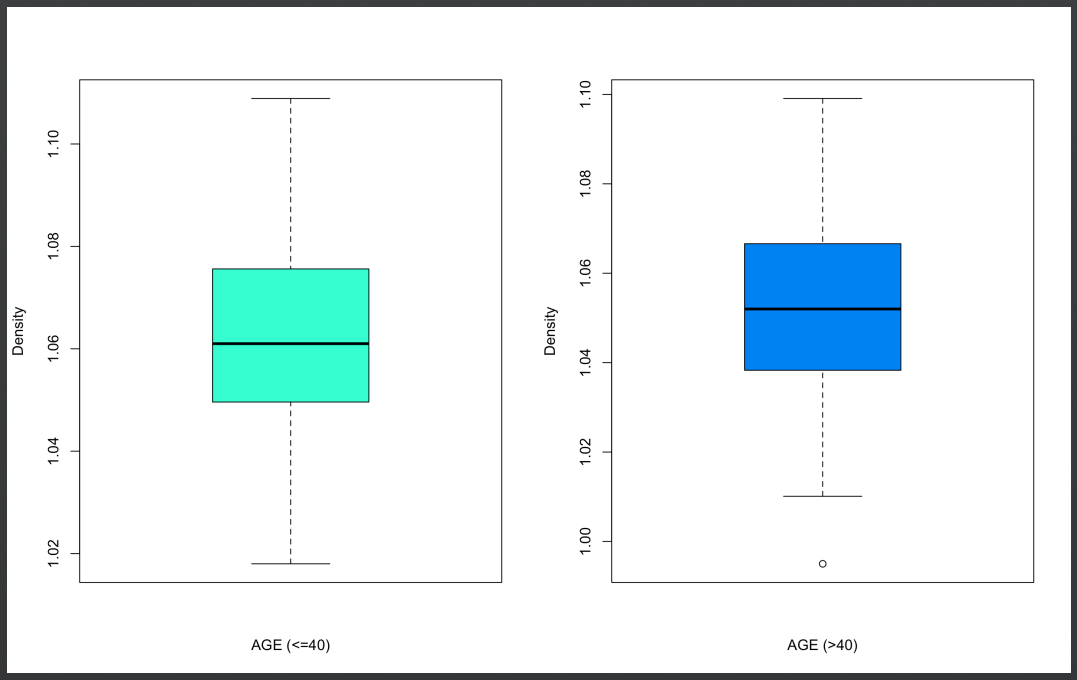
```
BodyFat_AgeUnder_40<-c(summary(BFU40$BODYFAT))
BodyFat_AgeAbove_40<-c(summary(BFA40$BODYFAT))
summary_bodyfat<-data.frame(BodyFat_AgeUnder_40,BodyFat_AgeAbove_40)
```

- **From above table-**
- **In general mean and median of Body Fat of people having AGE <= 40 is less than mean and median of people having AGE > 40.**
- **Body Fat distribution (Max. - Min.) for people of AGE > 40 varies more than Body Fat distribution (Max. - Min.) for people of AGE <=40.**

Body Fat - Conclusion

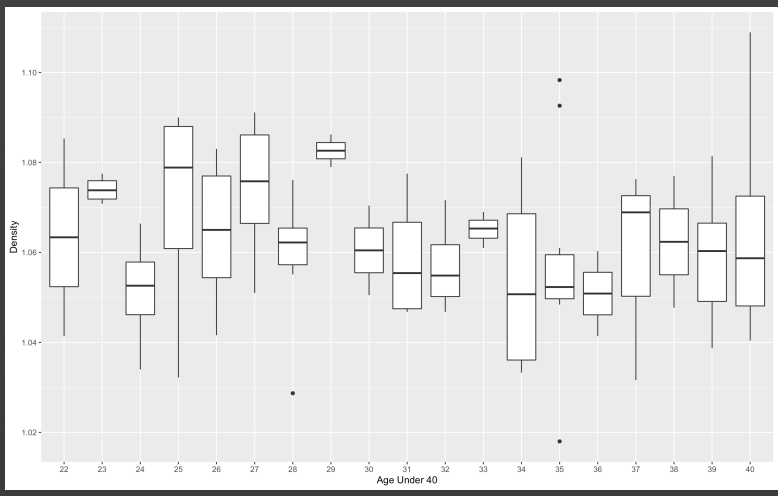
- **In general people of AGE > 40 have more body fat than people of AGE <= 40**
- **Body Fat of people of AGE > 40 is more widely distributed than that of people of AGE <= 40**
- **More people of AGE > 40 have abnormal amount of body fat than people of AGE <= 40.**

Density

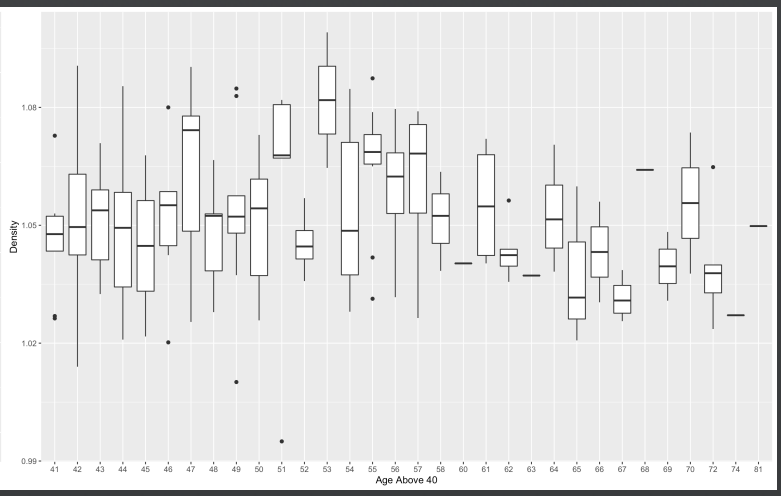



```
par(mfrow=c(1,2))
boxplot(BFU40$DENSITY,xlab="AGE (<=40)", ylab="Density", col="aquamarine1")
boxplot(BFA40$DENSITY,xlab="AGE (>40)", ylab="Density", col="dodgerblue2")
```

- **The mean for AGE <= 40 is around 1.06 while the mean for AGE > 40 is between 1.04 and 1.06 which means that people younger than 40 usually have density around 1.06 while people older than 40 have density between 1.04 and 1.06.**
- **The box plot for AGE <= 40 is a little positively skewed while the box plot for AGE > 40 is normally skewed.**
- **For AGE <=40, 25 % of people (between minimum and first quartile) have density between 1.02 and 1.05 while other 25 % of people (between third quartile and maximum) have a density between 1.07 and 1.11.**
- **For AGE > 40, 25 % of the people (between minimum and first quartile) have density between 1.01 and 1.04 while other 25 % of the people (between third quartile and maximum) have density between 1.65 and 1.10.**
- **One person having AGE > 40 has abnormally less density.**
- **People of AGE <= 40 have a more consistent distribution of density amongst different age groups because their range lies between 1.02 and around 1.11.**
- **People of AGE > 40 have a less consistent distribution of density amongst different age groups because their range lies between 1.00 and 1.10 (including the outlier).**
- **Overall the two box plots are quite similar.**




```
ggplot(BFU40,
aes(as.factor(AGE),DENSITY), fill=AGE) +
geom_boxplot() + labs(x = "Age Under 40",
```




```
ggplot(BFA40, aes(as.factor(AGE),DENSITY),
fill=AGE) + geom_boxplot() + labs(x = "Age
Above 40", y = "Density")
```

- **From above box-plots-**
- **Density of people of ages 23, 29, 33, 60, 62, 63, 68, 74 & 81 is nearly constant.**
- **Density of people aged 24, 25, 34, 44, 45, 47 & 54 varies a lot.**
- **Some people aged 28, 35, 41, 46, 49, 51, 55, 62, 72 are outliers meaning some people aged 35, 41, 46, 49, 55, 62 & 72 have abnormally large density and some people aged 28, 35, 41, 46, 49, 51 & 55 have abnormally low density.**

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Density_AgeUnder_40	1.018	1.0498	1.061	1.062204	1.0755	1.1089
Density_AgeAbove_40	0.995	1.0383	1.052	1.051930	1.0666	1.0991

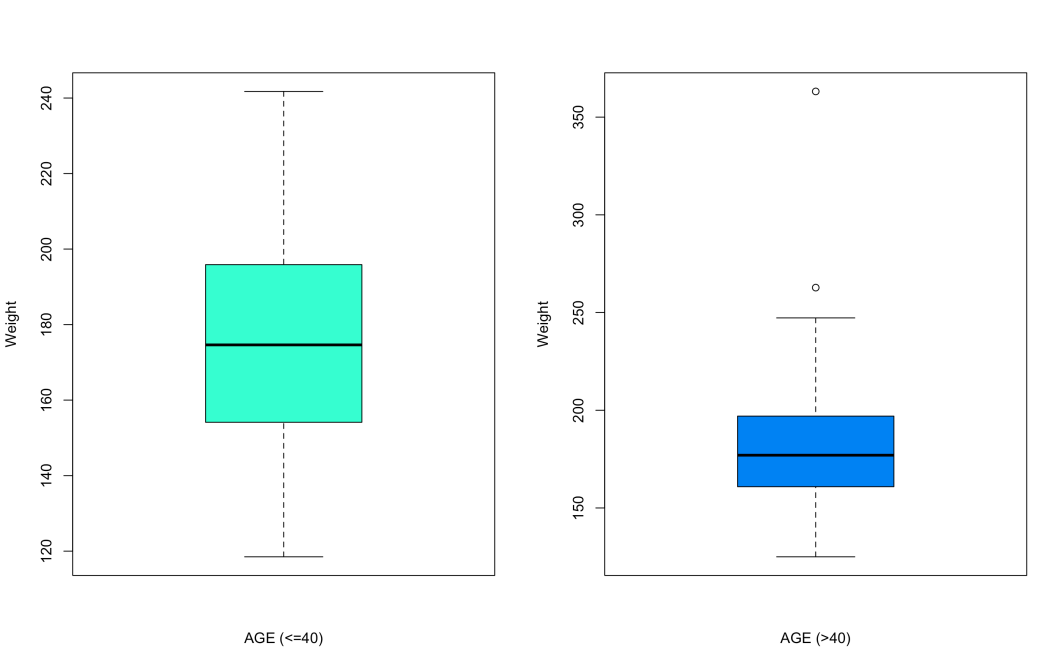
```
Density_AgeUnder_40<-c(summary(BFU40$DENSITY))
Density_AgeAbove_40<-c(summary(BFA40DENSITY))
summary_density<-data.frame(Density_AgeUnder_40,Density_AgeAbove_40)
t(summary_density)
```

- **From above table-**
- **In general mean and median of Density of people having AGE <= 40 is more than mean and median of people having AGE > 40.**

Density - Conclusion

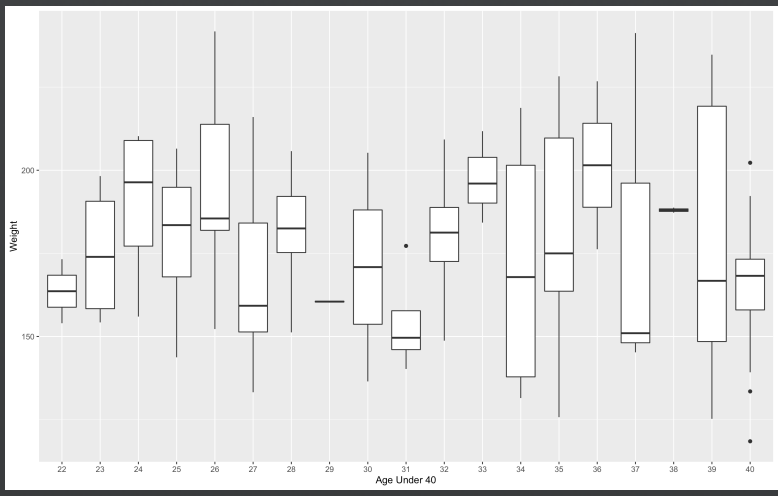
- **In general people of AGE > 40 have less density than people of AGE <= 40**
- **Density of people of AGE > 40 is less widely distributed than that of people of AGE <= 40**
- **More people of AGE > 40 have abnormal densities than people of AGE <= 40 for their respective age as well as their age group.**

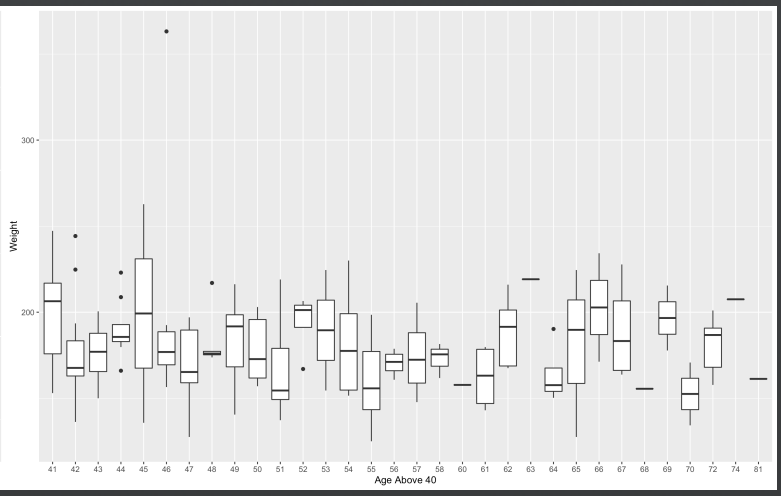
Weight




```
par(mfrow=c(1,2))
boxplot(BFU40$WEIGHT,xlab="AGE (<=40)", ylab="Weight", col="aquamarine1")
boxplot(BFA40$WEIGHT,xlab="AGE (>40)", ylab="Weight", col="dodgerblue2")
```

- **The mean for AGE <= 40 is around 180 pounds while the mean for AGE > 40 is around 170 pounds which means that people younger than 40 usually have weight around 180 while people older than 40 have weight around 170.**
- **The box plot for AGE <= 40 is a normally skewed while the box plot for AGE > 40 is a little positively skewed.**
- **For AGE <=40, 25 % of people (between minimum and first quartile) have weight between 120 and 160 pounds while other 25 % of people (between third quartile and maximum) have a weight between 200 and 240 pounds.**
- **For AGE > 40, 25 % of the people (between minimum and first quartile) have weight between 140 and 160 pounds while other 25 % of the people (between third quartile and maximum) have weight between 200 and 350 pounds,**
- **Two people having AGE > 40 have abnormally large weight.**
- **People of AGE <= 40 have a more consistent distribution of weight amongst different age groups because their range lies between 120 and 240 pounds.**
- **People of AGE > 40 have a less consistent distribution of weight amongst different age groups because their range lies between (around) 150 and 350 pounds (including the**






```
ggplot(BFU40, aes(as.factor(AGE),WEIGHT), fill=AGE) + geom_boxplot() + labs(x = "Age Under 40", y = "Weight")
```

```
ggplot(BFA40, aes(as.factor(AGE),WEIGHT), fill=AGE) + geom_boxplot() + labs(x = "Age Above 40", y = "Weight")
```

- **From above box-plots-**
- **Weight of people of ages 22, 29, 38, 48, 60, 63, 68, 74 & 81 is nearly constant.**
- **Weight of people aged 34, 39, 45, 54 & 65 varies a lot.**
- **Some people aged 31, 40, 42, 44, 46, 48, 52 & 64are outliers meaning some people aged 31, 40, 42, 44, 48, 64 have abnormally large weight and some people aged 40, 44 & 52 have abnormally less weight.**

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Weight_AgeUnder_40	118.5	154.1875	174.625	176.3739	195.8125	241.75
Weight_AgeAbove_40	125.0	160.8750	177.000	180.2362	197.0000	363.15

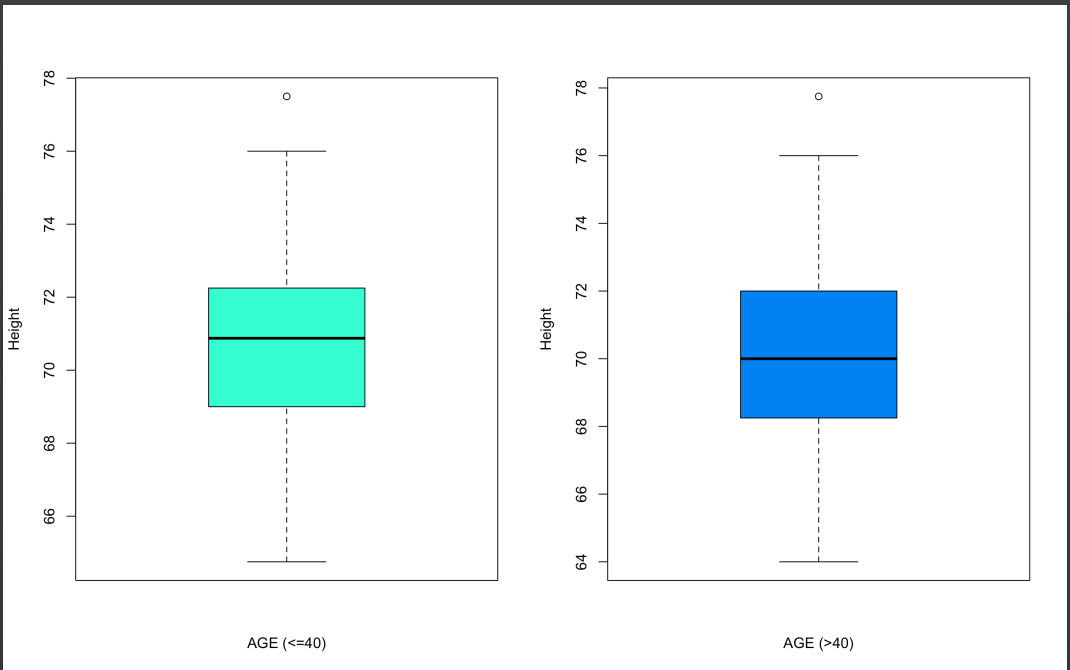
```
Weight_AgeUnder_40<-c(summary(BFU40$WEIGHT))
Weight_AgeAbove_40<-c(summary(BFA40$WEIGHT))
summary_weight<-data.frame(Weight_AgeUnder_40,Weight_AgeAbove_40)
t(summary_weight)
```

- **From above table-**
- **In general mean and median of Weight of people having AGE <= 40 is less than mean and median of people having AGE > 40.**
- **Weight distribution (Max. - Min.) for people of AGE > 40 varies more than Weight distribution (Max. - Min.) for people of AGE <=40.**

Weight - Conclusion

- **In general people of AGE > 40 have more weight than people of AGE <= 40**
- **Weight of people of AGE > 40 is more widely distributed than that of people of AGE <= 40**
- **More people of AGE > 40 have abnormal weight than people of AGE <= 40 for their respective age as well as their age group.**

Height




```
par(mfrow=c(1,2))
boxplot(BFU40$HEIGHT,xlab="AGE (<=40)", ylab="Height", col="aquamarine1")
boxplot(BFA40$HEIGHT,xlab="AGE (>40)", ylab="Height", col="dodgerblue2")
```

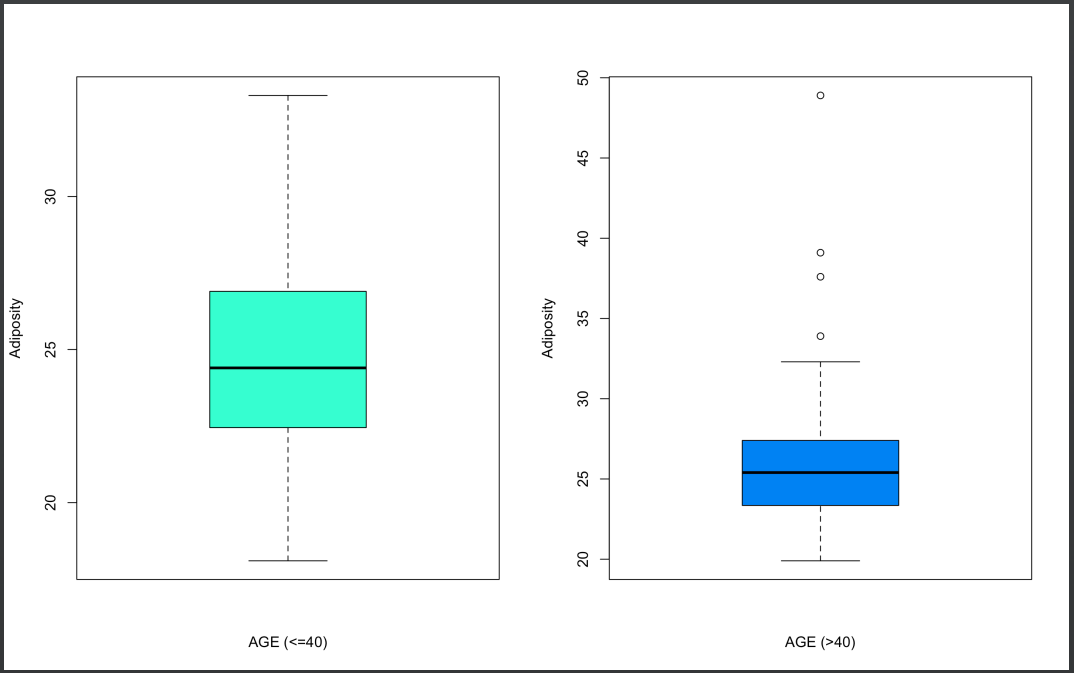
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Height_AgeUnder_40	64.75	69.00	70.875	70.69022	72.25	77.50
Height_AgeAbove_40	64.00	68.25	70.000	70.09119	72.00	77.75

```
Height_AgeUnder_40<-c(summary(BFU40$HEIGHT))
Height_AgeAbove_40<-c(summary(BFA40$HEIGHT))
summary_height<-data.frame(Height_AgeUnder_40,Height_AgeAbove_40)
t(summary_height)
```

Height - Conclusion

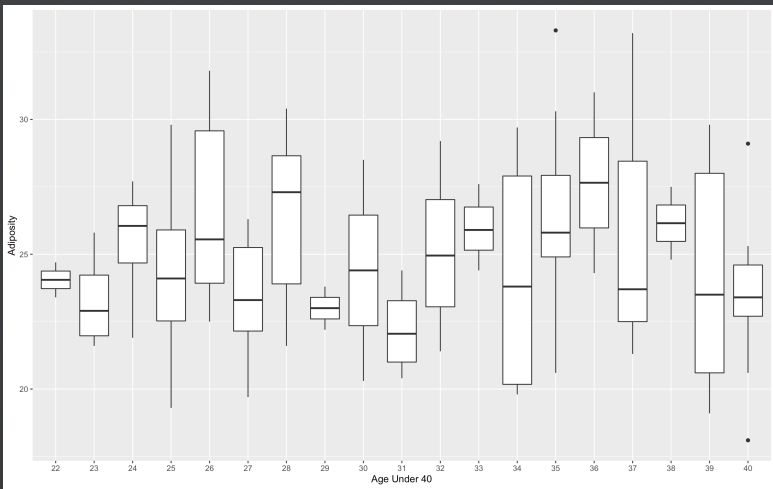
- **From above box-plot-**
- **Height of people for different age group remains almost same.**
- **Skewness and distribution of Height is also nearly same for both the age groups.**
- **Height for AGE <= 40 is a little negatively skewed while height for AGE > 40 is a little positively skewed.**
- **From the table-**
- **Min, 1st Qu, Median, Mean, 3rd Qu, Max values of Height for AGE <= 40 and AGE > 40 are almost same for all practical purposes.**

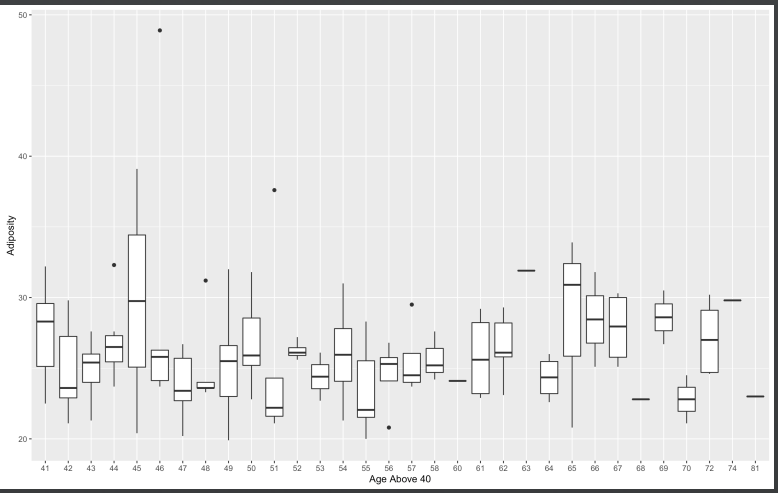
Adiposity




```
par(mfrow=c(1,2))
boxplot(BFU40$ADIPOSITY,xlab="AGE (<=40)", ylab="Adiposity", col="aquamarine1")
boxplot(BFA40$ADIPOSITY,xlab="AGE (>40)", ylab="Adiposity", col="dodgerblue2")
```

- **The mean for AGE <= 40 is less than 25 while the mean for AGE > 40 is more than 25.**
- **The box plot for AGE <= 40 is a little positively skewed while the box plot for AGE > 40 is normally skewed.**
- **For AGE <=40, 25 % of people (between minimum and first quartile) have adiposity between around 20 and around 22.5 while other 25 % of people (between third quartile and maximum) have adiposity between 27.5 and around 32.5.**
- **For AGE > 40, 25 % of the people (between minimum and first quartile) have adiposity between 20 and around 23 while other 25 % of the people (between third quartile and maximum) have adiposity between around 27.5 and 50 (including outlier).**
- **Four people having AGE > 40 have abnormally large adiposity.**
- **People of AGE <= 40 have a more consistent distribution of adiposity amongst different age groups because their range lies between around 20 and around 30.**
- **People of AGE > 40 have a less consistent distribution of adiposity amongst different age groups because their range lies between 20 and 50 (including the outlier).**






```
ggplot(BFU40,
aes(as.factor(AGE),ADIPOSITY), fill=AGE) +
geom_boxplot() + labs(x = "Age Under 40",
y = “Adiposity")
```

```
ggplot(BFA40,
aes(as.factor(AGE),ADIPOSITY), fill=AGE) +
geom_boxplot() + labs(x = "Age Above 40",
y = “Adiposity")
```

- **From above box-plots-**
- **Adiposity of people of ages 22, 29, 48, 52, 60, 63, 68, 74 & 81 is nearly constant.**
- **Adiposity of people aged 26,34, 37, 39, 45 & 65 varies a lot.**
- **Some people aged 35, 40, 44, 46, 48, 51, 56 & 57 are outliers meaning some people aged 35, 40, 44, 46, 48, 51 & 57 have abnormally large adiposity and some people aged have 40 & 56 normally less adiposity.**

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Adiposity_AgeUnder_40	18.1	22.475	24.4	24.75543	26.7	33.3
Adiposity_AgeAbove_40	19.9	23.350	25.4	25.80314	27.4	48.9

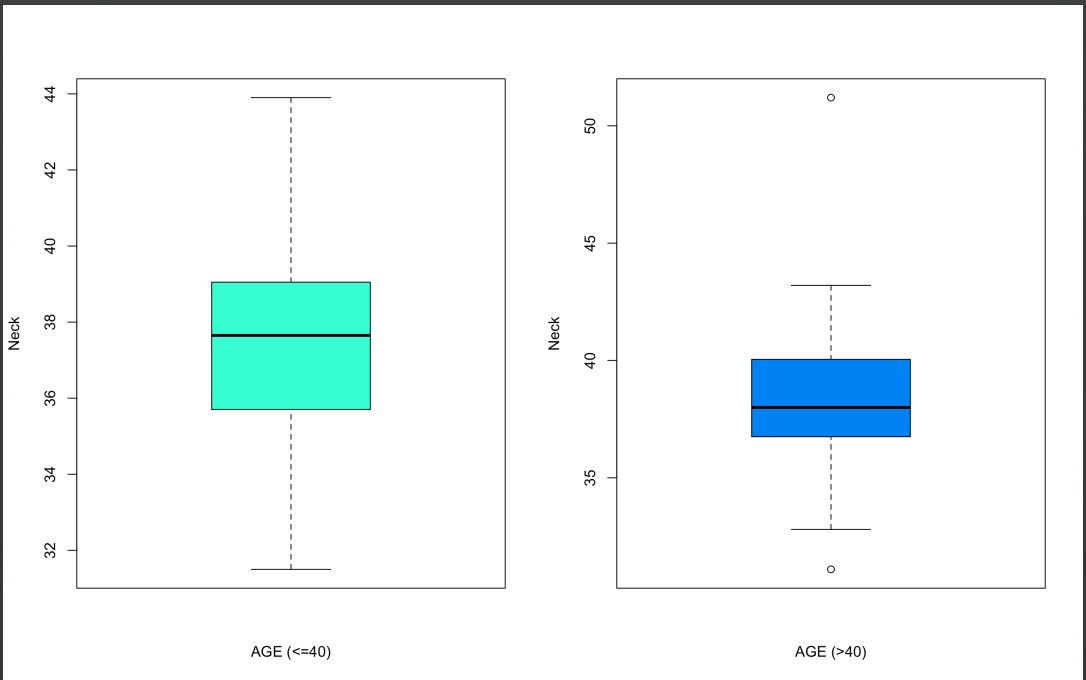
```
Adiposity_AgeUnder_40<-c(summary(BFU40$ADIPOSITY))
Adiposity_AgeAbove_40<-c(summary(BFA40$ADIPOSITY))
summary_adiposity<-data.frame(Adiposity_AgeUnder_40,Adiposity_AgeAbove_40)
```

- **From above table-**
- **In general mean and median of Adiposity of people having AGE <= 40 is less than mean and median of people having AGE > 40.**
- **Adiposity distribution (Max. - Min.) for people of AGE > 40 varies more than Adiposity distribution (Max. - Min.) for people of AGE <=40.**

Adiposity - Conclusion

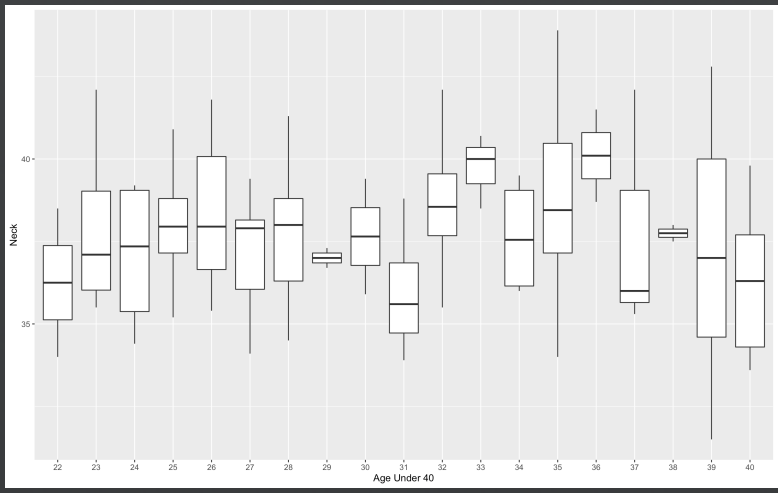
- **In general people of AGE > 40 have more adiposity than people of AGE <= 40**
- **Adiposity of people of AGE > 40 is more widely distributed than that of people of AGE <= 40**
- **More people of AGE > 40 have abnormal adiposity than people of AGE <= 40 for their respective age as well as their age group.**

Neck

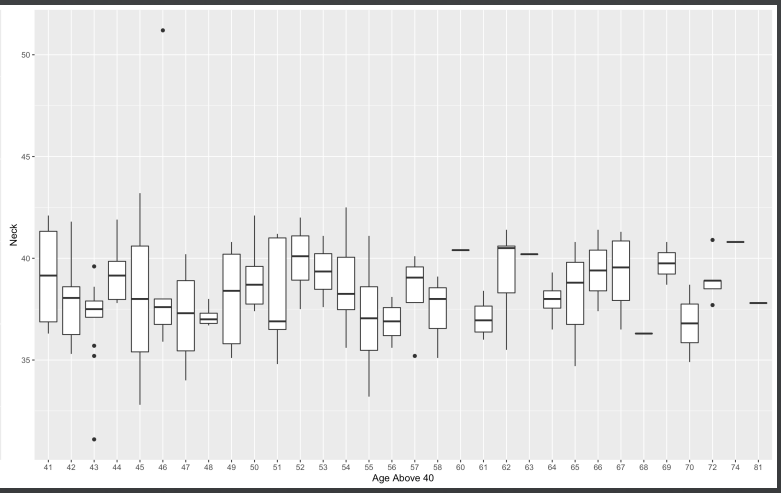



```
par(mfrow=c(1,2))
boxplot(BFU40$NECK,xlab="AGE (<=40)", ylab="Neck", col="aquamarine1")
boxplot(BFA40$NECK,xlab="AGE (>40)", ylab="Neck", col="dodgerblue2")
```

- **The mean for AGE <= 40 is around 38 while the mean for AGE > 40 is between 35 and 40.**
- **The box plot for AGE <= 40 is a little negatively skewed while the box plot for AGE > 40 is a little positively skewed.**
- **For AGE <=40, 25 % of people (between minimum and first quartile) have neck between 32 and 36 while other 25 % of people (between third quartile and maximum) have neck between around 40 and 44.**
- **For AGE > 40, 25 % of the people (between minimum and first quartile) have neck between around 30 and 37 while other 25 % of the people (between third quartile and maximum) have neck between 40 and 50 (including outlier).**
- **One person having AGE > 40 has abnormally large neck (around 50).**
- **People of AGE <= 40 have a more consistent distribution of neck amongst different age groups because their range lies between around 32 and around 44.**
- **People of AGE > 40 have a less consistent distribution of neck amongst different age groups because their range lies between around 35 and 50 (including the outlier).**




```
ggplot(BFU40, aes(as.factor(AGE),NECK), fill=AGE) + geom_boxplot() + labs(x = "Age Under 40", y = “Neck”)
```




```
ggplot(BFA40, aes(as.factor(AGE),NECK), fill=AGE) + geom_boxplot() + labs(x = "Age Above 40", y = “Neck”)
```

- **From above box-plots-**
- **Neck of people of ages 29, 38, 48, 60, 63, 68, 74 & 81 is nearly constant.**
- **Neck of people aged 26, 35, 37, 39, 45, 49, 52 & 65 varies a lot.**
- **Some people aged 43, 46, 57 & 72 are outliers meaning some people aged 43, 46 & 72 have abnormally large neck and some people aged have 43, 57 & 72 abnormally normally less neck.**

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Neck_AgeUnder_40	31.5	35.70	37.65	37.57065	39.025	43.9
Neck_AgeAbove_40	31.1	36.75	38.00	38.24465	40.050	51.2

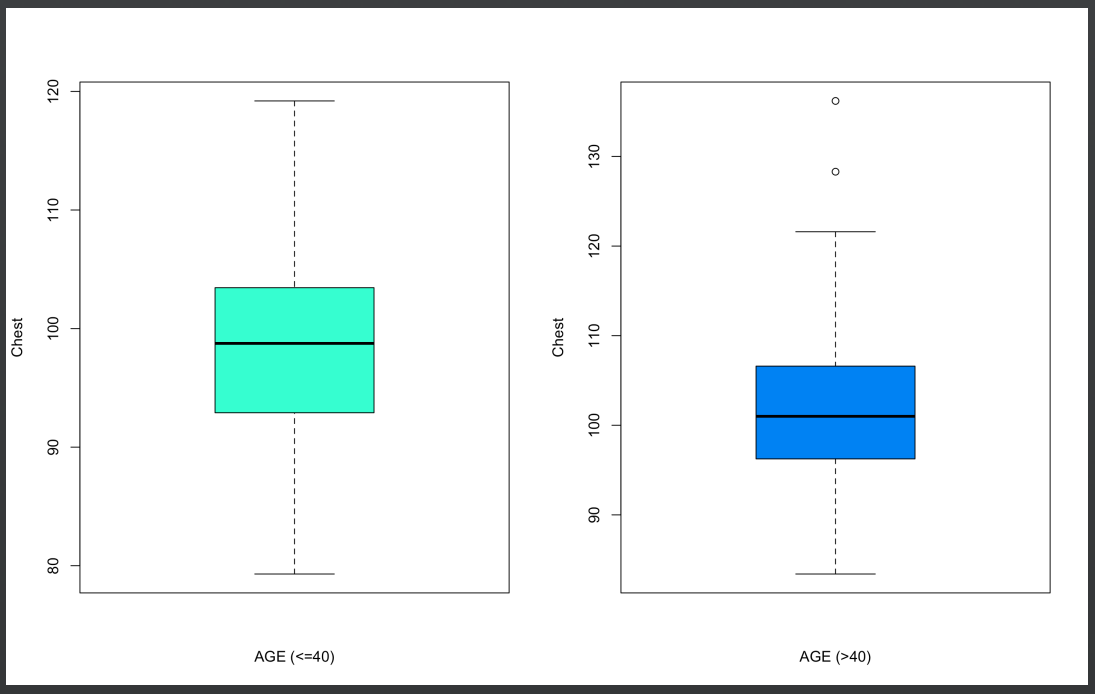
```
Neck_AgeUnder_40<-c(summary(BFU40$NECK))
Neck_AgeAbove_40<-c(summary(BFA40$NECK))
```

- **From above table-**
- **In general mean and median of Neck of people having AGE <= 40 is less than mean and median of Neck of people having AGE > 40.**
- **Neck distribution (Max. - Min.) for people of AGE > 40 varies more than Neck distribution (Max. - Min.) for people of AGE <=40.**

Neck - Conclusion

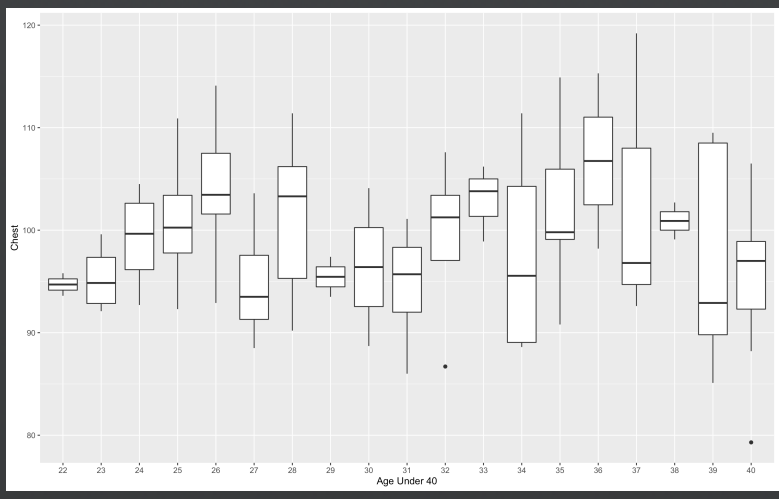
- **In general people of AGE > 40 have more neck size than people of AGE <= 40**
- **Neck size of people of AGE > 40 is more widely distributed than that of people of AGE <= 40**
- **More people of AGE > 40 have abnormal neck than people of AGE <= 40 for their respective age as well as their age group.**

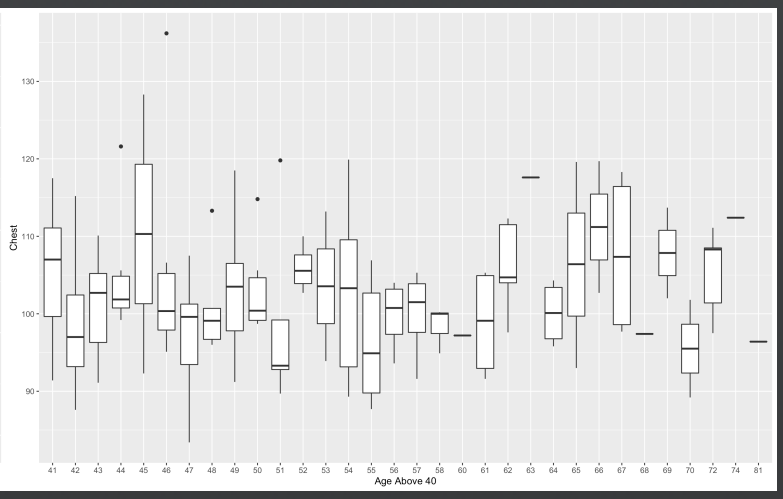
Chest




```
par(mfrow=c(1,2))
boxplot(BFU40$CHEST,xlab="AGE (<=40)", ylab="Chest", col="aquamarine1")
boxplot(BFA40$CHEST,xlab="AGE (>40)", ylab="Chest", col="dodgerblue2")
```

- **The mean for AGE <= 40 is less than 100 while the mean for AGE > 40 is more than 100.**
- **The box plot for AGE <= 40 is a little negatively skewed while the box plot for AGE > 40 is a little positively skewed.**
- **For AGE <=40, 25 % of people (between minimum and first quartile) have chest between 80 and around 90 while other 25 % of people (between third quartile and maximum) have chest between 105 and 120.**
- **For AGE > 40, 25 % of the people (between minimum and first quartile) have chest between around 85 and 95 while other 25 % of the people (between third quartile and maximum) have chest between around 110 and around 130 (including outlier).**
- **Two people having AGE > 40 have abnormally large chest (around 120).**
- **People of AGE <= 40 have a more consistent distribution of chest amongst different age groups because their range lies between around 80 and around 120.**
- **People of AGE > 40 have a less consistent distribution of chest amongst different age groups because their range lies between around 85 and around 135 (including the outlier).**






```
ggplot(BFU40, aes(as.factor(AGE),CHEST), fill=AGE) + geom_boxplot() + labs(x = "Age Under 40", y = “Chest”)
```

```
ggplot(BFA40, aes(as.factor(AGE),CHEST), fill=AGE) + geom_boxplot() + labs(x = "Age Above 40", y = “Chest”)
```

- **From above box-plots-**
- **Chest of people of ages 22, 29, 38, 60, 63, 68, 74 & 81 is nearly constant.**
- **Chest of people aged 28, 34, 37, 39, 45, 49, 54, 65 & 67 varies a lot.**
- **Some people aged 32, 40, 44, 46, 48, 50 & 51 are outliers meaning some people aged 44, 46, 48, 50 & 51 have abnormally large chest and some people aged have 32 & 40**

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Chest_AgeUnder_40	79.3	92.90	98.75	98.59891	103.375	119.2
Chest_AgeAbove_40	83.4	96.25	101.00	102.07925	106.600	136.2

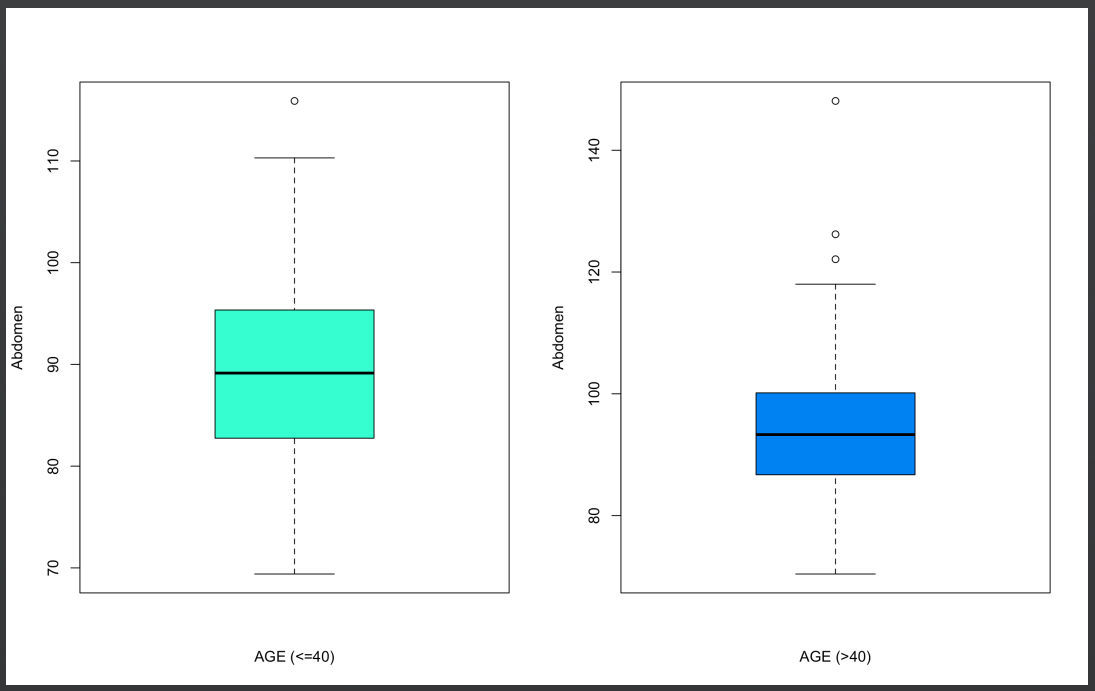
```
Chest_AgeUnder_40<-c(summary(BFU40$CHEST))
Chest_AgeAbove_40<-c(summary(BFA40$CHEST))
summary_chest<-data.frame(Chest_AgeUnder_40,Chest_AgeAbove_40)
t(summary_chest)
```

- **From above table-**
- **In general mean and median of chest of people having AGE <= 40 is less than mean and median of chest of people having AGE > 40.**
- **Chest distribution (Max. - Min.) for people of AGE > 40 varies more than Chest**

Chest - Conclusion

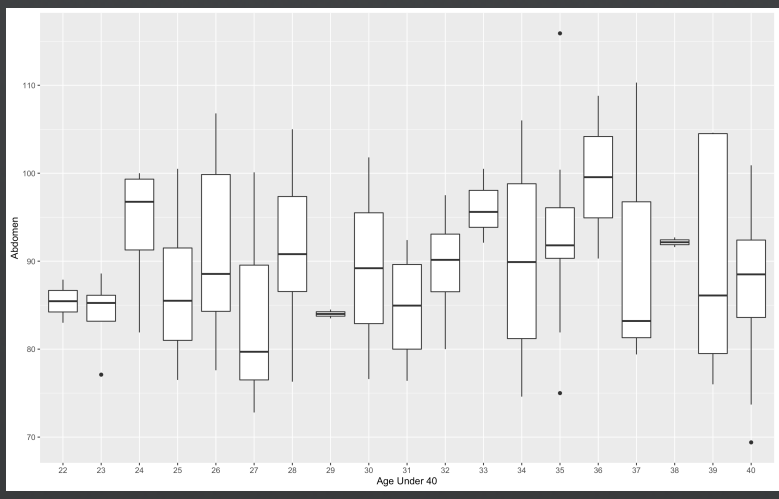
- **In general people of AGE > 40 have more chest size than people of AGE <= 40**
- **Chest size of people of AGE > 40 is more widely distributed than that of people of AGE <= 40**
- **More people of AGE > 40 have abnormal chest than people of AGE <= 40 for their respective age as well as their age group.**

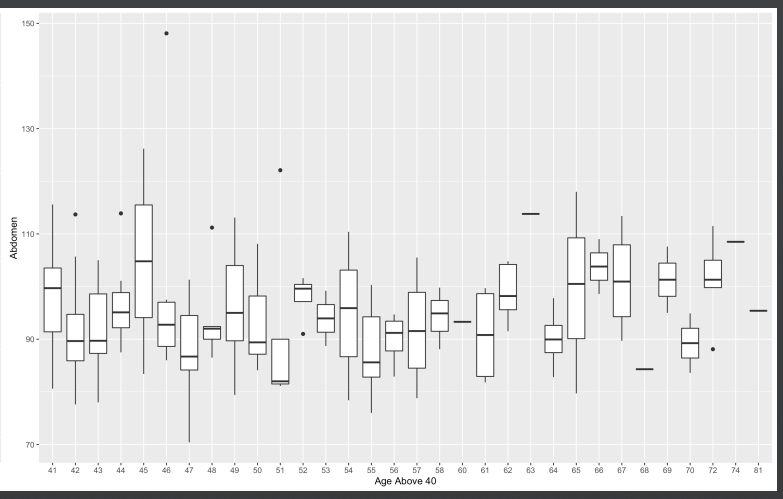
Abdomen




```
par(mfrow=c(1,2))
boxplot(BFU40$ABDOMEN,xlab="AGE (<=40)", ylab="Abdomen", col="aquamarine1")
boxplot(BFA40$ABDOMEN,xlab="AGE (>40)", ylab="Abdmoen", col="dodgerblue2")
```

- **The mean for AGE <= 40 is more than 90 while the mean for AGE > 40 is more than 90.**
- **The box plot for AGE <= 40 is normally skewed while the box plot for AGE > 40 is normally skewed too.**
- **For AGE <=40, 25 % of people (between minimum and first quartile) have abdomen between 70 and around 90 while other 25 % of people (between third quartile and maximum) have abdomen between 95 and 120 (including outlier).**
- **For AGE > 40, 25 % of the people (between minimum and first quartile) have abdomen between around 75 and 90 while other 25 % of the people (between third quartile and maximum) have abdomen between around 100 and around 150 (including outlier).**
- **One person having AGE <= 40 has abnormally large abdomen (around 120) and three people having AGE > 40 have abnormally large abdomen (around 120, 120, 140).**
- **People of AGE <= 40 have a more consistent distribution of abdomen amongst different age groups because their range lies between around 70 and around 120.**
- **People of AGE > 40 have a less consistent distribution of abdomen amongst different age groups because their range lies between around 75 and around 150 (including the outlier).**






```
ggplot(BFU40, aes(as.factor(AGE),ABDOMEN), fill=AGE) + geom_boxplot() + labs(x = "Age Under 40", y = “Abdomen”)
```

```
ggplot(BFA40, aes(as.factor(AGE),ABDOMEN), fill=AGE) + geom_boxplot() + labs(x = "Age Above 40", y = “Abdomen”)
```

- **From above box-plots-**
- **Abdomen of people of ages 29, 38,60, 63, 68, 74 & 81 is nearly constant.**
- **Abdomen of people aged 26, 34, 39, 45, 54, 57 & 65 varies a lot.**
- **Some people aged 23, 35, 40, 42, 45, 46, 48, 51, 52 & 72 are outliers meaning some people aged 35, 42, 44, 46, 48 & 51 have abnormally large abdomen and some people aged have 23, 35, 45, 52, & 72 have abnormally small abdomen.**

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Abdomen_AgeUnder_40	69.4	82.875	89.15	89.23913	95.225	115.9
Abdomen_AgeAbove_40	70.4	86.700	93.30	94.40126	100.150	148.1

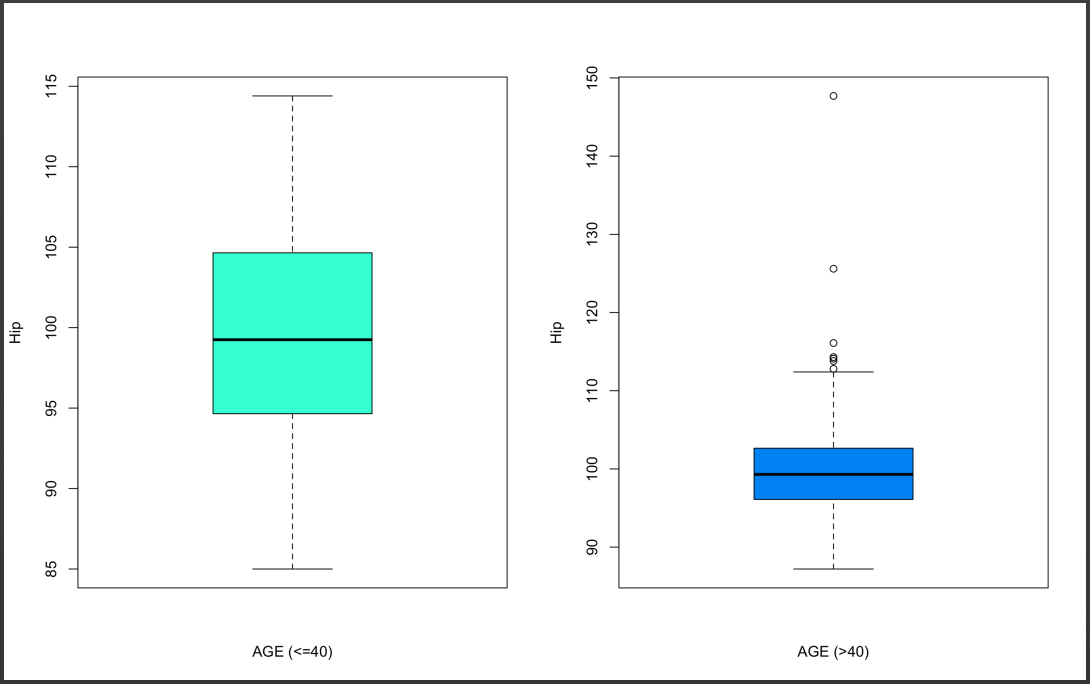
```
Abdomen_AgeUnder_40<-c(summary(BFU40$ABDOMEN))
Abdomen_AgeAbove_40<-c(summary(BFA40$ABDOMEN))
summary_abdomen<-data.frame(Abdomen_AgeUnder_40,Abdomen_AgeAbove_40)
```

- **From above table-**
- **In general mean and median of abdomen of people having AGE <= 40 is less than mean and median of Abdomen of people having AGE > 40.**
- **Abdomen distribution (Max. - Min.) for people of AGE > 40 varies more than Abdomen**

Abdomen - Conclusion

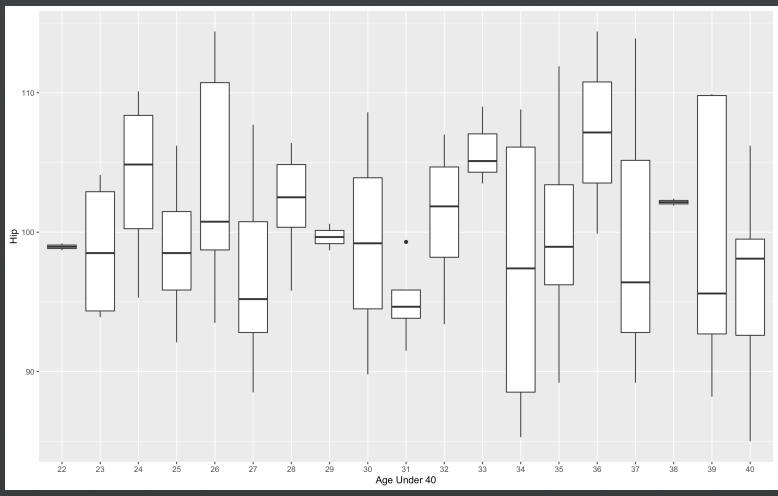
- **In general people of AGE > 40 have more abdomen size than people of AGE <= 40**
- **Abdomen size of people of AGE > 40 is more widely distributed than that of people of AGE <= 40**
- **More people of AGE > 40 have abnormal abdomen than people of AGE <= 40 for their respective age as well as their age group.**

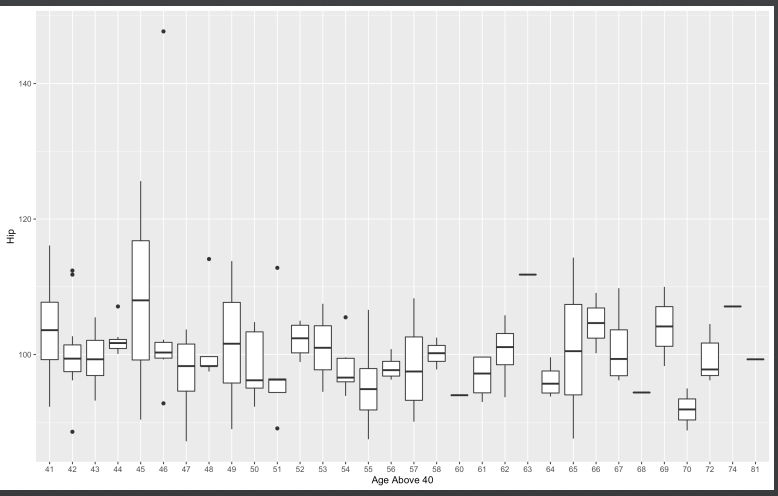
Hip




```
par(mfrow=c(1,2))
boxplot(BFU40$HIP,xlab="AGE (<=40)", ylab="Hip", col="aquamarine1")
boxplot(BFA40$HIP,xlab="AGE (>40)", ylab="Hip", col="dodgerblue2")
```

- **The mean for AGE <= 40 is less than 100 while the mean for AGE > 40 is almost 100.**
- **The box plot for AGE <= 40 is a little positively skewed while the box plot for AGE > 40 is normally skewed.**
- **For AGE <=40, 25 % of people (between minimum and first quartile) have hip between 85 and around 95 while other 25 % of people (between third quartile and maximum) have hip between 105 and 115.**
- **For AGE > 40, 25 % of the people (between minimum and first quartile) have hip between around 90 and 95 while other 25 % of the people (between third quartile and maximum) have hip between around 105 and around 150 (including outlier).**
- **Two people having AGE > 40 have abnormally large hip (around 120).**
- **People of AGE <= 40 have a more consistent distribution of hip amongst different age groups because their range lies between around 85 and around 115.**
- **People of AGE > 40 have a less consistent distribution of hip amongst different age groups because their range lies between around 85 and around 150 (including the outlier).**






```
ggplot(BFU40, aes(as.factor(AGE),HIP),
fill=AGE) + geom_boxplot() + labs(x = "Age
Under 40", y = “Hip”)
```

```
ggplot(BFA40, aes(as.factor(AGE),HIP),
fill=AGE) + geom_boxplot() + labs(x = "Age
Above 40", y = “Hip”)
```

- **From above box-plots-**
- **Hip of people of ages 22, 29, 38, 60, 63, 68, 74 & 81 is nearly constant.**
- **Hip of people aged 28, 34, 37, 39, 45, 49 & 65 varies a lot.**
- **Some people aged 31, 42, 44, 46 & 51 are outliers meaning some people aged 31, 42, 44, 48, 51 & 54 have abnormally large hip and some people aged have 42, 46 & 51 abnormally less hip.**

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Hip_AgeUnder_40	85.0	94.675	99.25	99.59565	104.575	114.4
Hip_AgeAbove_40	87.2	96.100	99.30	99.98553	102.650	147.7

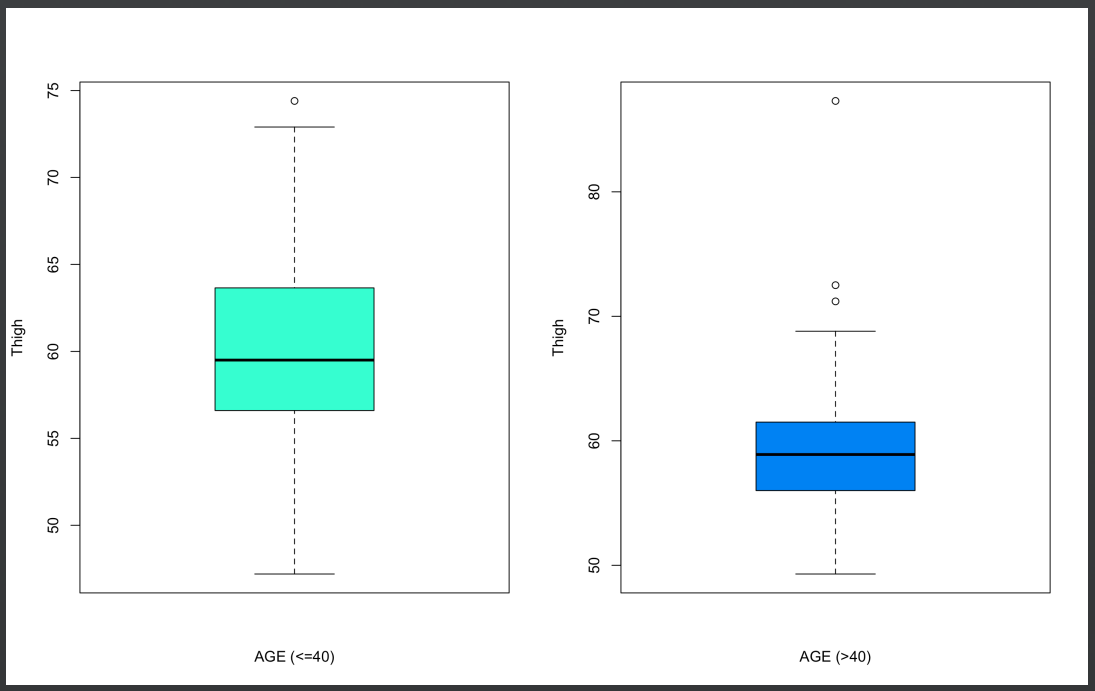
```
Hip_AgeUnder_40<-c(summary(BFU40$HIP))
Hip_AgeAbove_40<-c(summary(BFA40$HIP))
summary_hip<-data.frame(Hip_AgeUnder_40,Hip_AgeAbove_40)
```

- **From above table-**
- **In general mean and median of hip of people having AGE <= 40 is almost same to mean and median of hip of people having AGE > 40.**
- **Hip distribution (Max. - Min.) for people of AGE > 40 varies more than hip distribution**

Hip - Conclusion

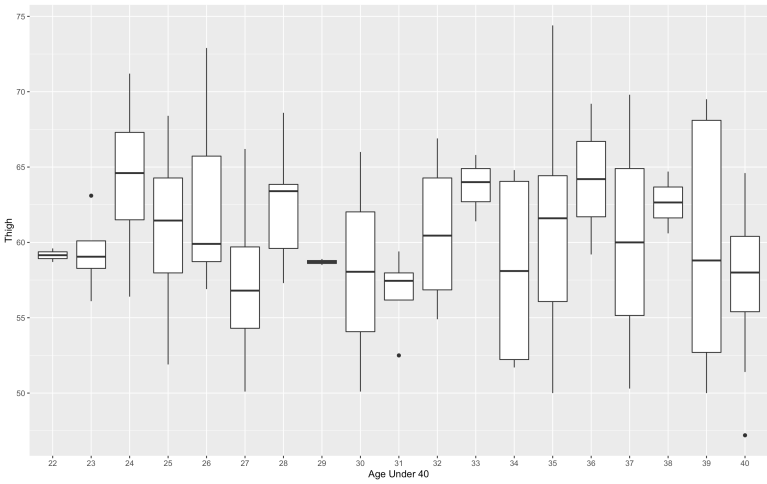
- **In general people of AGE > 40 have same size hip size like that of people of AGE <= 40**
- **Hip size of people of AGE > 40 is more widely distributed than that of people of AGE <= 40**
- **More people of AGE > 40 have abnormal hip than people of AGE <= 40 for their respective age as well as their age group.**

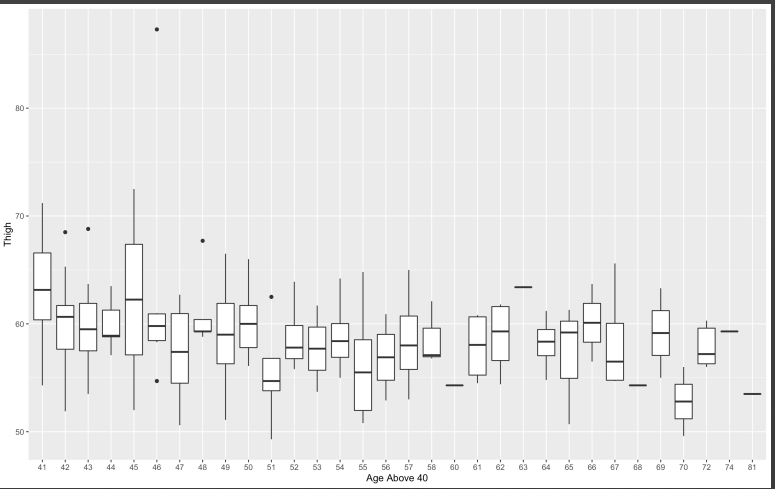
Thigh




```
par(mfrow=c(1,2))
boxplot(BFU40$THIGH,xlab="AGE (<=40)", ylab="Thigh", col="aquamarine1")
boxplot(BFA40$THIGH,xlab="AGE (>40)", ylab="Thigh", col="dodgerblue2")
```

- **The mean for AGE <= 40 and AGE > 40 is less than 60.**
- **The box plot for AGE <= 40 is a little positively skewed while the box plot for AGE > 40 is normally skewed.**
- **For AGE <=40, 25 % of people (between minimum and first quartile) have thigh between 45 and around 55 while other 25 % of people (between third quartile and maximum) have thigh between around 65 and 75.**
- **For AGE > 40, 25 % of the people (between minimum and first quartile) have thigh between around 50 and 55 (including outlier) while other 25 % of the people (between third quartile and maximum) have thigh between around 65 and around 85 (including outlier).**
- **One person having AGE <= 40 has abnormally large thigh (around 75), three people having AGE > 40 have abnormally large thigh (around 85).**
- **People of AGE <= 40 have a more consistent distribution of thigh amongst different age groups because their range lies between around 45 and around 75 (including the outlier).**
- **People of AGE > 40 have a less consistent distribution of thigh amongst different age groups because their range lies between 50 and around 90 (including the outlier).**






```
ggplot(BFU40, aes(as.factor(AGE),THIGH), fill=AGE) + geom_boxplot() + labs(x = "Age Under 40", y = “Thigh”)
```

```
ggplot(BFA40, aes(as.factor(AGE),THIGH), fill=AGE) + geom_boxplot() + labs(x = "Age Above 40", y = “Thigh”)
```

- **From above box-plots-**
- **Thigh of people of ages 22, 29, 60, 63, 68, 74 & 81 is nearly constant.**
- **Some people aged 23, 31, 40, 42, 43, 46, 48 & 51 are outliers meaning some people aged 23, 42, 43, 46, 48 & 51 have abnormally large thigh and some people aged have 31, 40 & 46 abnormally normally less thigh.**

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
THIGH_AgeUnder_40	47.2	56.7	59.5	59.99239	63.575	74.4
THIGH_AgeAbove_40	49.3	56.0	58.9	58.99623	61.500	87.3

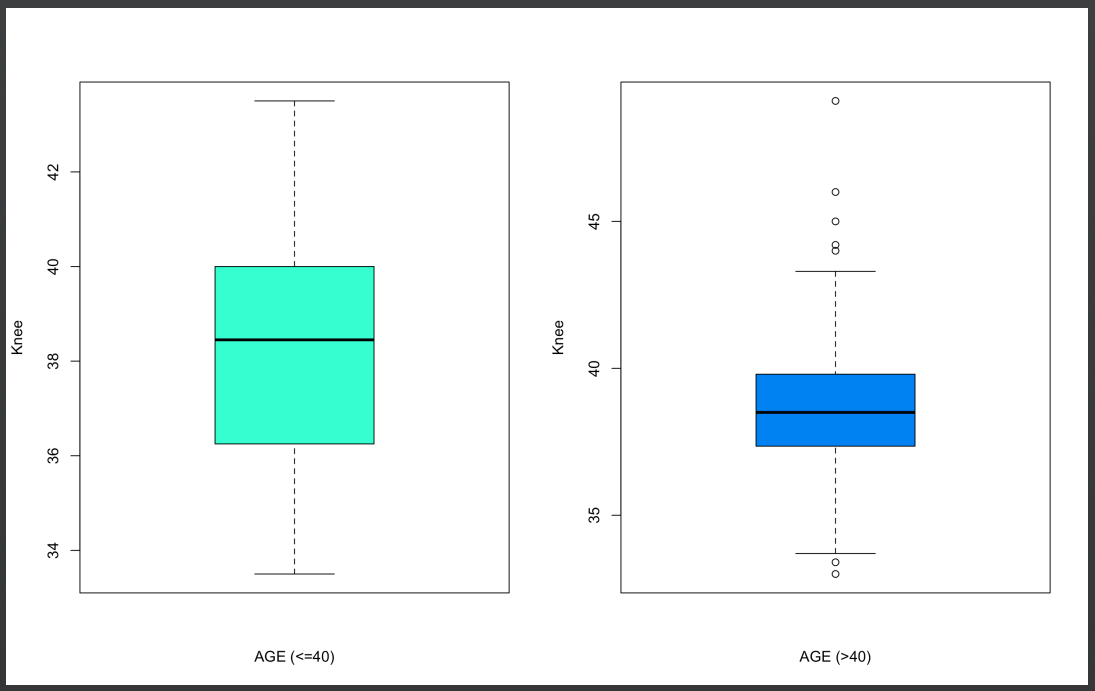
```
THIGH_AgeUnder_40<-c(summary(BFU40$THIGH))
THIGH_AgeAbove_40<-c(summary(BFA40$THIGH))
summary_THIGH<-data.frame(THIGH_AgeUnder_40,THIGH_AgeAbove_40)
t(summary_THIGH)
```

- **From above table-**
- **In general mean and median of thigh of people having AGE <= 40 is more than mean and median of thigh of people having AGE > 40.**
- **Thigh distribution (Max. - Min.) for people of AGE > 40 varies more than thigh distribution (Max. - Min.) for people of AGE <=40.**

Thigh - Conclusion

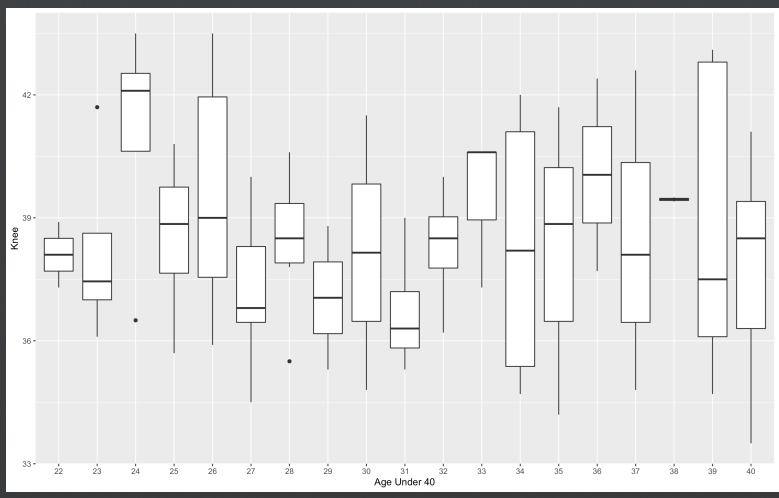
- **In general people of AGE > 40 have less thigh size than people of AGE <= 40**
- **Thigh size of people of AGE > 40 is more widely distributed than that of people of AGE <= 40**
- **More people of AGE > 40 have abnormal thigh than people of AGE <= 40 for their respective age as well as their age group.**

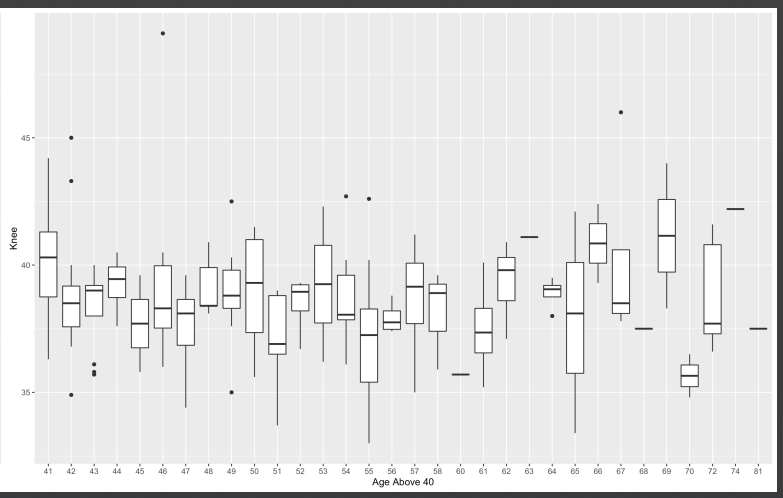
Knee




```
par(mfrow=c(1,2))
boxplot(BFU40$KNEE,xlab="AGE (<=40)", ylab="Knee", col="aquamarine1")
boxplot(BFA40$KNEE,xlab="AGE (>40)", ylab="Knee", col="dodgerblue2")
```

- **The mean for AGE <= 40 is more than 38 while the mean for AGE > 40 is around 38.**
- **The box plot for AGE <= 40 is a little negatively skewed while the box plot for AGE > 40 is almost normally skewed.**
- **For AGE <=40, 25 % of people (between minimum and first quartile) have knee between 34 and around 36 while other 25 % of people (between third quartile and maximum) have knee between 40 and 44.**
- **For AGE > 40, 25 % of the people (between minimum and first quartile) have knee between around 33 and 37 (including outlier) while other 25 % of the people (between third quartile and maximum) have knee between around 40 and around 50 (including outlier).**
- **Five people having AGE > 40 have abnormally large knee (around 45) and Two people have abnormally small knee (around 33).**
- **People of AGE <= 40 have a more consistent distribution of knee amongst different age groups because their range lies between around 34 and around 44.**
- **People of AGE > 40 have a less consistent distribution of knee amongst different age groups because their range lies between around 33 and around 50 (including the outlier).**






```
ggplot(BFU40, aes(as.factor(AGE),KNEE), fill=AGE) + geom_boxplot() + labs(x = "Age Under 40", y = “Knee”)
```

```
ggplot(BFA40, aes(as.factor(AGE),KNEE), fill=AGE) + geom_boxplot() + labs(x = "Age Above 40", y = “Knee”)
```

- **From above box-plots-**
- **Knee of people of ages 22, 29, 38, 60, 63, 68, 74 & 81 is nearly constant.**
- **Some people aged 23, 24, 26, 42, 43, 46, 49, 54, 55, 64 & 67 are outliers.**

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
KNEE_AgeUnder_40	33.5	36.275	38.45	38.39022	40.0	43.5
KNEE_AgeAbove_40	33.0	37.350	38.50	38.68176	39.8	49.1

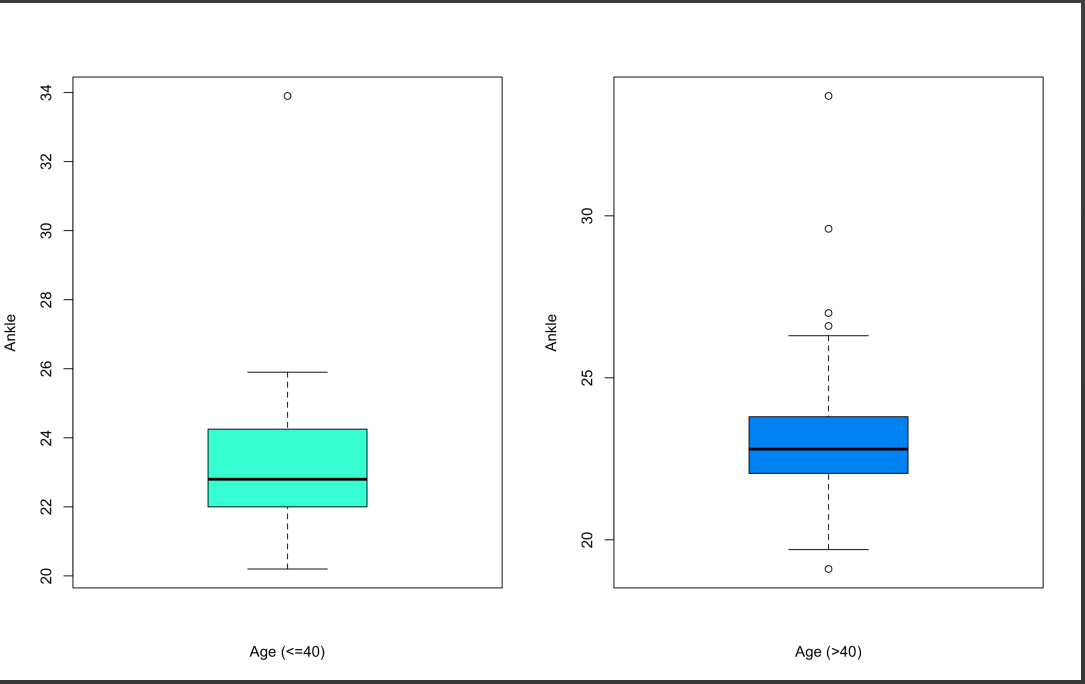
```
KNEE_AgeUnder_40<-c(summary(BFU40$KNEE))
KNEE_AgeAbove_40<-c(summary(BFA40$KNEE))
summary_KNEE<-data.frame(KNEE_AgeUnder_40,KNEE_AgeAbove_40)
t(summary_KNEE)
```

- **From above table-**
- **In general mean and median of knee of people having AGE <= 40 is less than mean and median of knee of people having AGE > 40.**
- **Knee distribution (Max. - Min.) for people of AGE > 40 varies more than knee distribution**

Knee - Conclusion

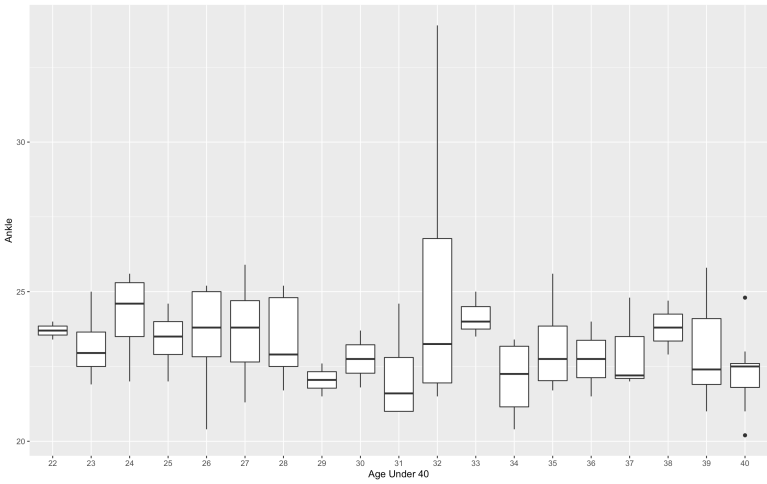
- **In general people of AGE > 40 have more knee size than people of AGE <= 40**
- **Knee size of people of AGE > 40 is more widely distributed than that of people of AGE <= 40**
- **More people of AGE > 40 have abnormal knee than people of AGE <= 40 for their respective age as well as their age group.**

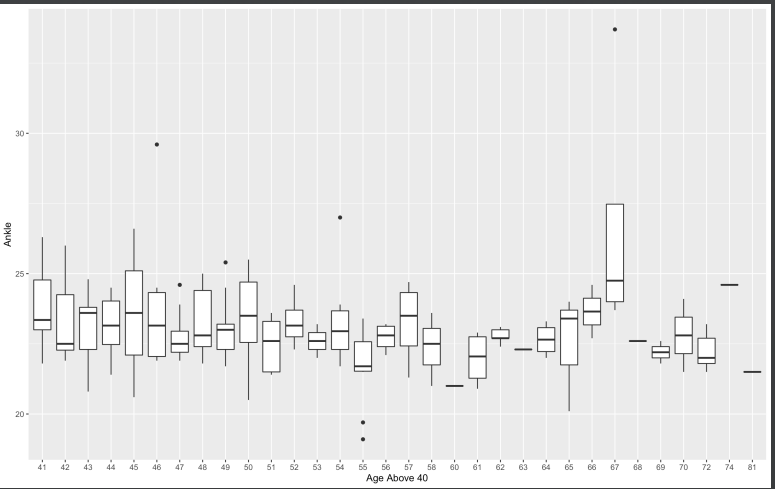
Ankle




```
par(mfrow=c(1,2))
boxplot(BFU40$ANKLE,xlab="AGE (<=40)", ylab="Ankle", col="aquamarine1")
boxplot(BFA40$ANKLE,xlab="AGE (>40)", ylab="Ankle", col="dodgerblue2")
```

- **The mean for AGE <= 40 and AGE > 40 is around 23.**
- **The box plot for AGE <= 40 is a little positively skewed while the box plot for AGE > 40 is a little positively skewed.**
- **For AGE <=40, 25 % of people (between minimum and first quartile) have ankle between 20 and around 22 while other 25 % of people (between third quartile and maximum) have ankle between 24 and 34.**
- **For AGE > 40, 25 % of the people (between minimum and first quartile) have ankle size between around 20 and 22 while other 25 % of the people (between third quartile and maximum) have ankle size between around 24 and around 34 (including outlier).**
- **One person having AGE <= 40 has abnormally large ankle. four people having AGE > 40 has abnormally large ankle (around 30) and one person having AGE > 40 has abnormally small ankle.**
- **Both the age groups have similar consistency because of similar ranges.**






```
ggplot(BFU40, aes(as.factor(AGE),ANKLE), fill=AGE) + geom_boxplot() + labs(x = "Age Under 40", y = “Ankle”)
```

```
ggplot(BFA40, aes(as.factor(AGE),ANKLE), fill=AGE) + geom_boxplot() + labs(x = "Age Above 40", y = “Ankle”)
```

- **From above box-plots-**
- **Ankle of people of ages 22, 29, 38, 60, 63, 68, 74 & 81 is nearly constant.**
- **Some people aged 40, 45, 47, 49, 54 & 55 are outliers.**

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
ANKLE_AgeUnder_40	20.2	22.00	22.8	23.15217	24.175	33.9
ANKLE_AgeAbove_40	19.1	22.05	22.8	23.06981	23.800	33.7

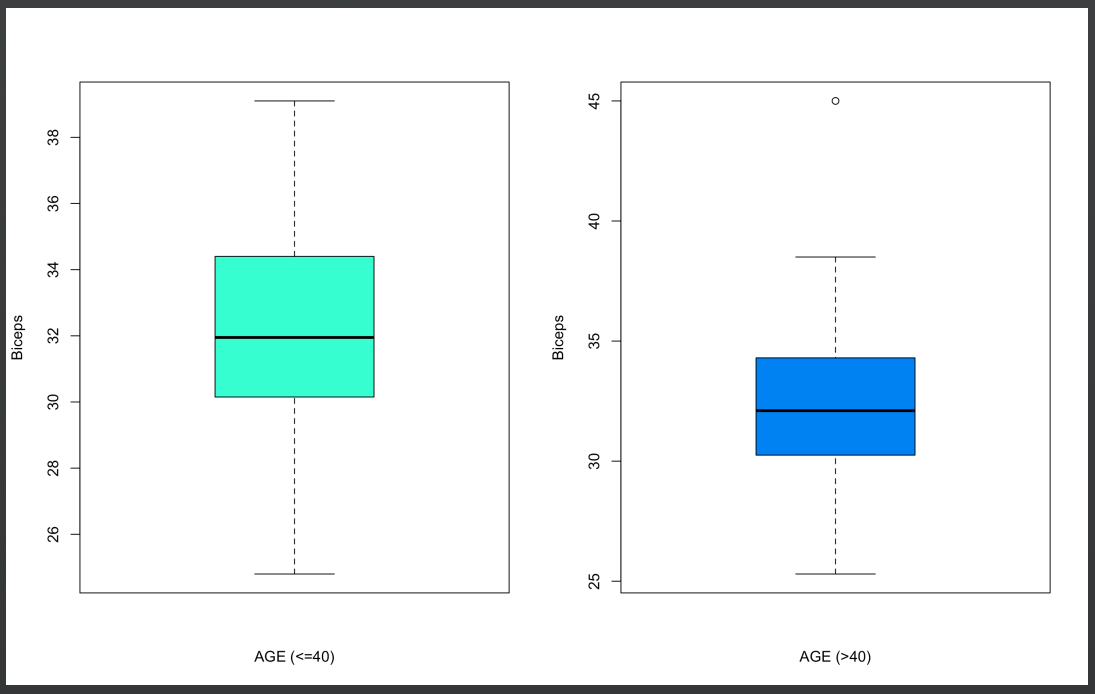
```
ANKLE_AgeUnder_40<-c(summary(BFU40$ANKLE))
ANKLE_AgeAbove_40<-c(summary(BFA40$ANKLE))
summary_ANKLE<-data.frame(ANKLE_AgeUnder_40,ANKLE_AgeAbove_40)
t(summary_ANKLE)
```

- **From above table-**
- **In general mean and median of ankle of people having AGE <= 40 is almost similar to mean and median of Neck of people having AGE > 40.**

Ankle - Conclusion

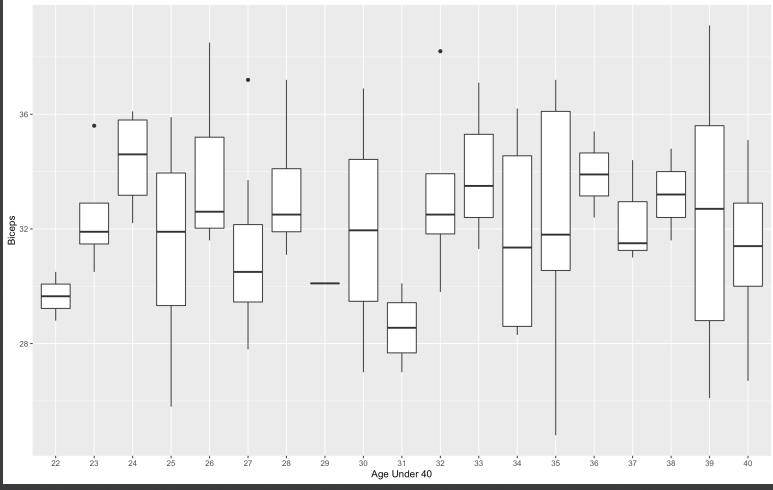
- **In general people of AGE > 40 have similar ankle size to people of AGE <= 40**
- **More people of AGE > 40 have abnormal ankle than people of AGE <= 40 for their respective age as well as their age group.**

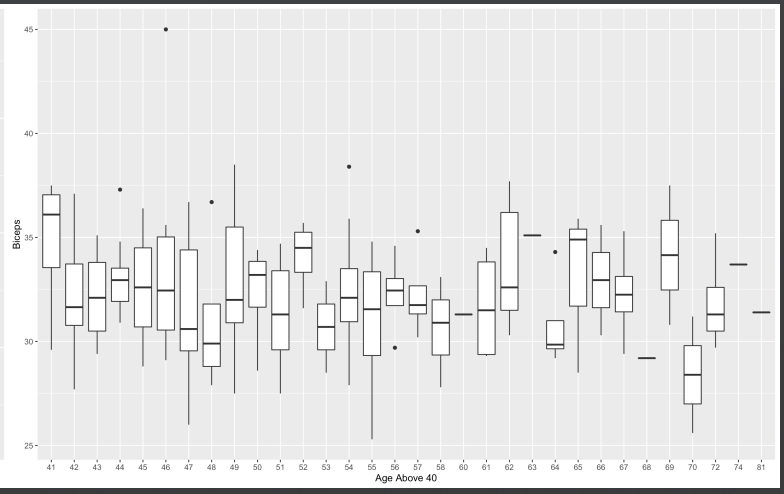
Biceps




```
par(mfrow=c(1,2))
boxplot(BFU40$BICEPS,xlab="AGE (<=40)", ylab="Bieceps", col="aquamarine1")
boxplot(BFA40$BICEPS,xlab="AGE (>40)", ylab="Biceps", col="dodgerblue2")
```

- **The mean for AGE <= 40 is around 32 while the mean for AGE > 40 is more than 32.**
- **The box plot for AGE <= 40 is a little positively skewed while the box plot for AGE > 40 is a little positively skewed.**
- **For AGE <=40, 25 % of people (between minimum and first quartile) have biceps between 25 and around 30 while other 25 % of people (between third quartile and maximum) have biceps between 34 and 40.**
- **For AGE > 40, 25 % of the people (between minimum and first quartile) have biceps between around 25 and 30 while other 25 % of the people (between third quartile and maximum) have biceps between around 35 and around 45 (including outlier).**
- **One person having AGE > 40 has abnormally large bicep (around 45).**






```
ggplot(BFU40, aes(as.factor(AGE),BICEPS), fill=AGE) + geom_boxplot() + labs(x = "Age Under 40", y = “Biceps”)
```

```
ggplot(BFA40, aes(as.factor(AGE),BICEPS), fill=AGE) + geom_boxplot() + labs(x = "Age Above 40", y = “Biceps”)
```

- **From above box-plots-**
- **Biceps of people of ages 22, 29, 38, 60, 63, 68, 74 & 81 is nearly constant.**
- **Some people aged 23, 27, 32, 44, 46, 49, 54, 56 & 57 are outliers.**

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
BICEPS_AgeUnder_40	24.8	30.175	31.95	32.14783	34.4	39.1
BICEPS_AgeAbove_40	25.3	30.250	32.10	32.33774	34.3	45.0

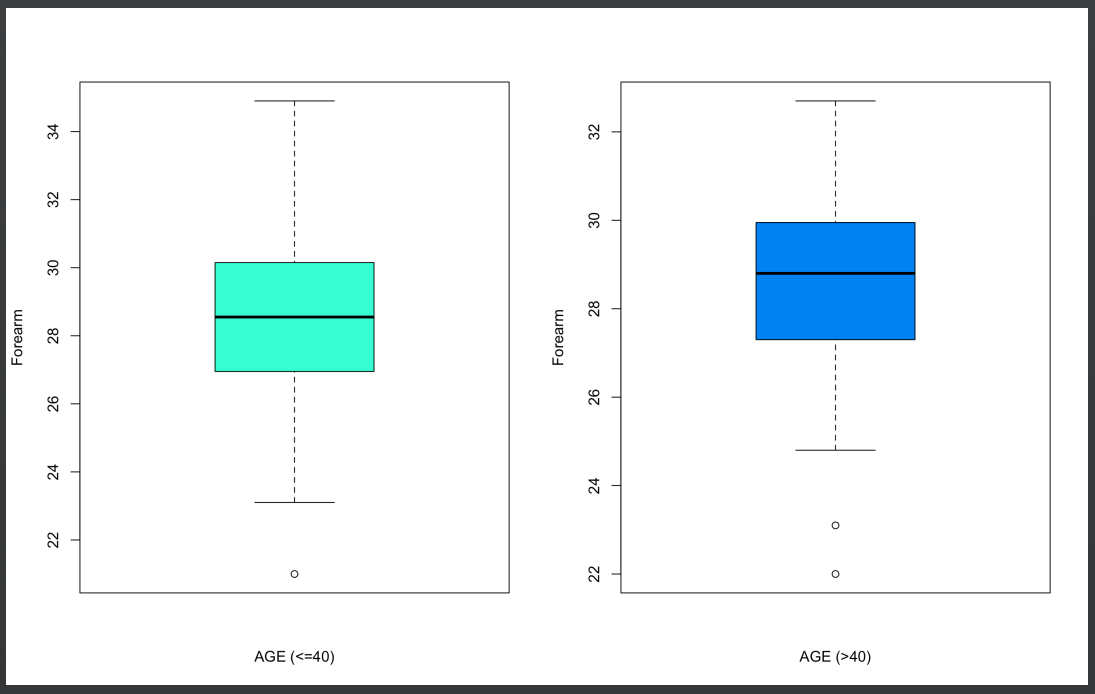
```
BICEPS_AgeUnder_40<-c(summary(BFU40$BICEPS))
BICEPS_AgeAbove_40<-c(summary(BFA40$BICEPS))
summary_BICEPS<-data.frame(BICEPS_AgeUnder_40,BICEPS_AgeAbove_40)
t(summary_BICEPS)
```

- **From above table-**
- **In general mean and median of biceps of people having AGE <= 40 is less than mean and median of biceps of people having AGE > 40.**
- **Bicep distribution (Max. - Min.) for people of AGE > 40 varies more than Chest distribution**

Biceps - Conclusion

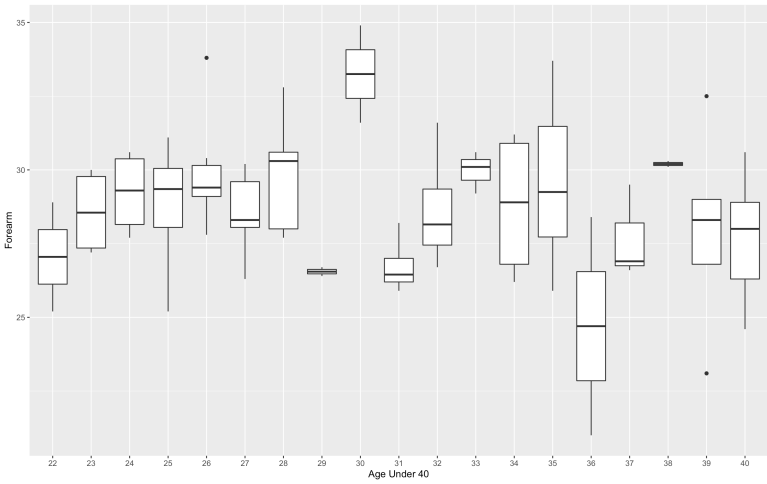
- **In general people of AGE > 40 have more bicep size than people of AGE <= 40**
- **Bicep size of people of AGE > 40 is more widely distributed than that of people of AGE <= 40**
- **More people of AGE > 40 have abnormal biceps than people of AGE <= 40 for their respective age as well as their age group.**

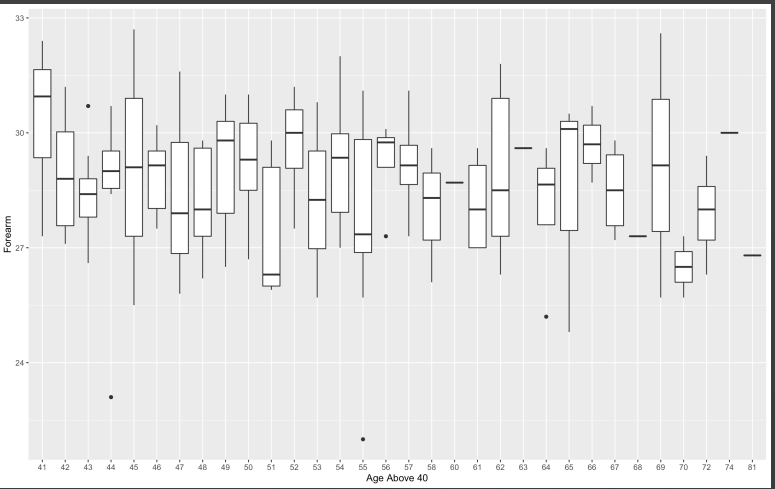
Forearm




```
par(mfrow=c(1,2))
boxplot(BFU40$FOREARM,xlab="AGE (<=40)", ylab="Forearm", col="aquamarine1")
boxplot(BFA40$FOREARM,xlab="AGE (>40)", ylab="Forearm", col="dodgerblue2")
```

- **The mean for AGE <= 40 is almost similar to AGE > 40.**
- **The box plot for AGE <= 40 is normally skewed while the box plot for AGE > 40 is a little negatively skewed.**
- **For AGE <=40, 25 % of people (between minimum and first quartile) have forearm between 22 and around 26 while other 25 % of people (between third quartile and maximum) have chest between 30 and 34.**
- **For AGE > 40, 25 % of the people (between minimum and first quartile) have forearm between around 22 and 28 while other 25 % of the people (between third quartile and maximum) have forearm between around 30 and around 34 (including outlier).**
- **One person having AGE <= 40 has abnormally small forearm (around 22), two people having AGE > 40 has abnormally small forearm (around 22).**






```
ggplot(BFU40, aes(as.factor(AGE),FOREARM), fill=AGE) + geom_boxplot() + labs(x = "Age Under 40", y = “Forearm”)
```

```
ggplot(BFA40, aes(as.factor(AGE),FOREARM), fill=AGE) + geom_boxplot() + labs(x = "Age Above 40", y = “Forearm”)
```

- **From above box-plots-**
- **Forearm of people of ages 22, 29, 38, 60, 63, 68, 74 & 81 is nearly constant.**
- **Some people aged 26, 39, 40, 44, 55, 65 are outliers.**

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
FOREARM_AgeUnder_40	21	26.975	28.55	28.64130	30.125	34.9
FOREARM_AgeAbove_40	22	27.300	28.80	28.67673	29.950	32.7

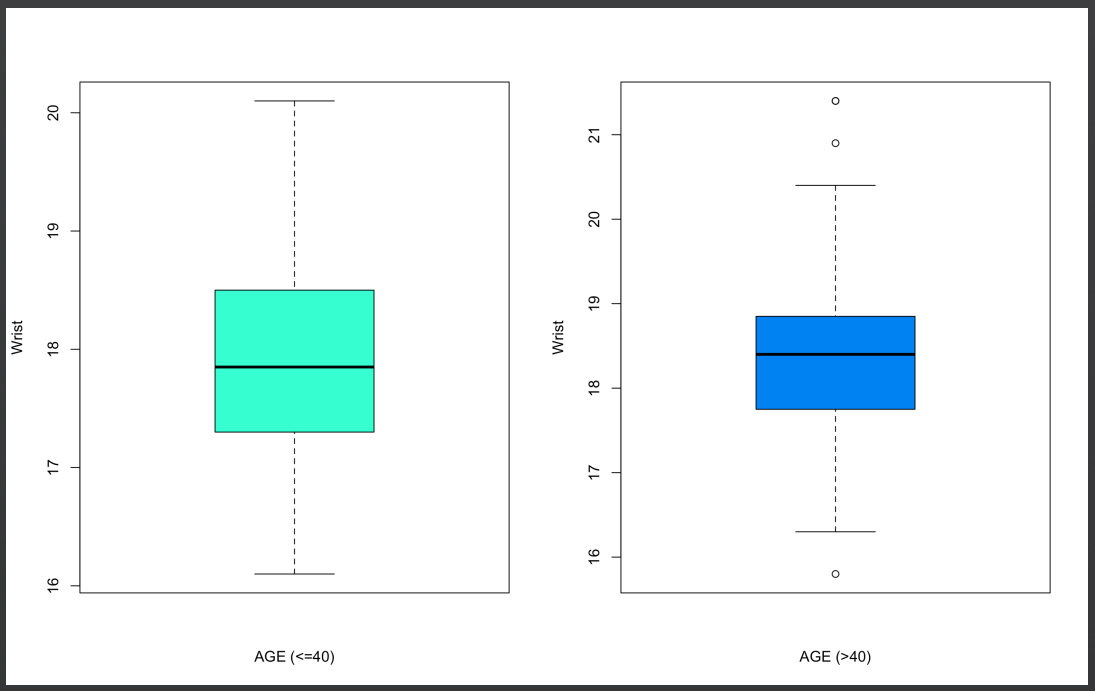
```
FOREARM_AgeUnder_40<-c(summary(BFU40$FOREARM))
FOREARM_AgeAbove_40<-c(summary(BFA40$FOREARM))
summary_FOREARM<-data.frame(FOREARM_AgeUnder_40,FOREARM_AgeAbove_40)
t(summary_FOREARM)
```

- **From above table-**
- **In general mean and median of forearm of people having AGE <= 40 is almost similar to mean and median of forearm of people having AGE > 40.**
- **Forearm distribution (Max. - Min.) for people of AGE > 40 varies less than forearm distribution (Max. - Min.) for people of AGE <=40.**

Forearm - Conclusion

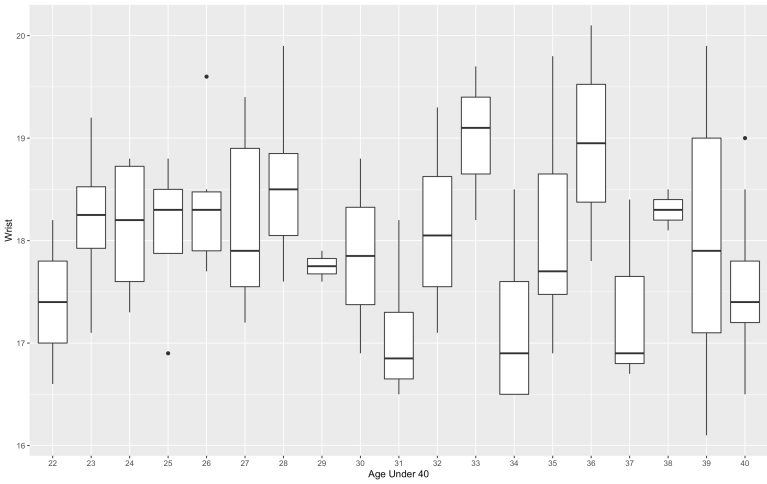
- **In general people of AGE > 40 have similar forearm size than people of AGE <= 40**
- **Forearm size of people of AGE > 40 is less widely distributed than that of people of AGE <= 40**
- **More people of AGE > 40 have abnormal forearm than people of AGE <= 40 for their respective age as well as their age group.**

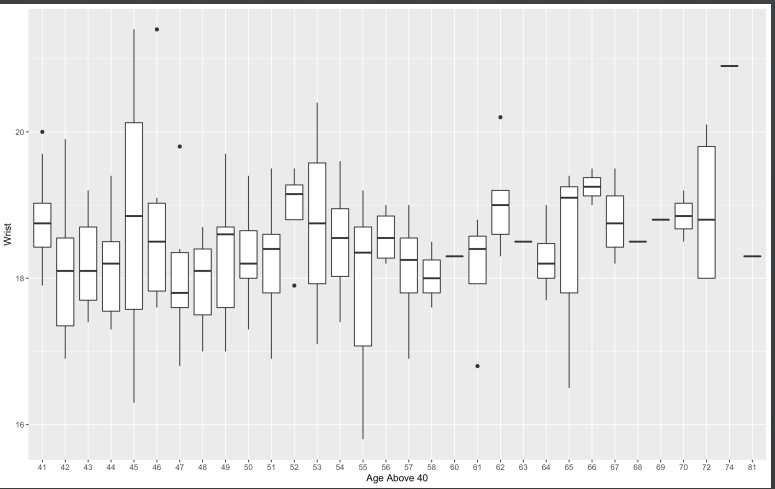
Wrist




```
par(mfrow=c(1,2))
boxplot(BFU40$WRIST,xlab="AGE (<=40)", ylab="Wrist", col="aquamarine1")
boxplot(BFA40$WRIST,xlab="AGE (>40)", ylab="Wrist", col="dodgerblue2")
```

- **The mean for AGE <= 40 is less than 18 while the mean for AGE > 40 is more than 18.**
- **The box plot for AGE <= 40 is a little positively skewed while the box plot for AGE > 40 is a little negatively skewed.**
- **For AGE <=40, 25 % of people (between minimum and first quartile) have wrist between 16 and around 17 while other 25 % of people (between third quartile and maximum) have wrist between around 19 and 20.**
- **For AGE > 40, 25 % of the people (between minimum and first quartile) have wrist between around 16 and 18 while other 25 % of the people (between third quartile and maximum) have wrist between around 19 and around 21 (including outlier).**
- **Two people having AGE > 40 has abnormally large wrist (around 120) and one person in AGE > 40 has abnormally small wrist (around 16).**






```
ggplot(BFU40, aes(as.factor(AGE),WRIST), fill=AGE) + geom_boxplot() + labs(x = "Age Under 40", y = “Wrist”)
```

```
ggplot(BFA40, aes(as.factor(AGE),WRIST), fill=AGE) + geom_boxplot() + labs(x = "Age Above 40", y = “Wrist”)
```

- **From above box-plots-**
- **Chest of people of ages 22, 29, 38, 60, 63, 68, 74 & 81 is nearly constant.**
- **Some people aged 25, 26, 40, 41, 46, 47, 52, 61 & 62 are outliers.**

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
WRIST_AgeUnder_40	16.1	17.30	17.85	17.96848	18.50	20.1
WRIST_AgeAbove_40	15.8	17.75	18.40	18.38616	18.85	21.4

```
WRIST_AgeUnder_40<-c(summary(BFU40$WRIST))
WRIST_AgeAbove_40<-c(summary(BFA40$WRIST))
summary_WRIST<-data.frame(WRIST_AgeUnder_40,WRIST_AgeAbove_40)
t(summary_WRIST)
```

- **From above table-**
- **In general mean and median of wrist of people having AGE <= 40 is less than mean and median of wrist of people having AGE > 40.**
- **Wrist distribution (Max. - Min.) for people of AGE > 40 varies more than wrist distribution (Max. - Min.) for people of AGE <=40.**

Wrist - Conclusion

- **In general people of AGE > 40 have more wrist size than people of AGE <= 40**
- **Wrist size of people of AGE > 40 is more widely distributed than that of people of AGE <= 40**
- **More people of AGE > 40 have abnormal wrist than people of AGE <= 40 for their respective age as well as their age group.**

KRITIK SETH
J043