Population of blocks

---

20 marks

In this question, you will examine different possible attributes of interest for this population.

a. Simple numerical attributes.

i. *(1 mark)* Summarize this population by the following attributes on the variates `weight` and `perimeter`: the population median, mean, and standard deviation (here computed using `sd()` with denominator $N-1$).

My answer follows

You would answer this by inserting some R code

```
popAttributesWeight <- list(median = median(blocks[,'weight']), mean = mean(blocks[,'weight'])
popAttributesWeight
```

```
## $median
## [1] 30
##
## $mean
## [1] 32.4
##
## $sd
## [1] 16.04098
```

ii. *(2 marks)* Repeat the above summaries but now conditional on the group to which each block belongs. Now include the number in each group.

Answer

```
groupA <- with(blocks, group =='A')
group_A_median <- with(blocks[groupA,],median(weight, na.rm = TRUE))
group_A_mean <- with(blocks[groupA,],mean(weight, na.rm = TRUE))
group_A_sd <- with(blocks[groupA,],sd(weight, na.rm = TRUE))
groupB <- with(blocks, group =='B')
group_B_median <- with(blocks[groupB,],median(weight, na.rm = TRUE))
group_B_mean <- with(blocks[groupB,],mean(weight, na.rm = TRUE))
group_B_sd <- with(blocks[groupB,],sd(weight, na.rm = TRUE))
group_A_median
```

```
## [1] 20
```

```
group_A_mean
```

```
## [1] 21.1
```

```
group_A_sd
```

```
## [1] 7.304597
```

```
group_B_median
```

```
## [1] 40
```

```
group_B_mean
```

```
## [1] 43.7
```

```
group_B_sd
```

```
## [1] 14.35021
```

iii. *(3 marks)* On the basis of the above computed attributes, describe how each group differs from the whole population and from each other.

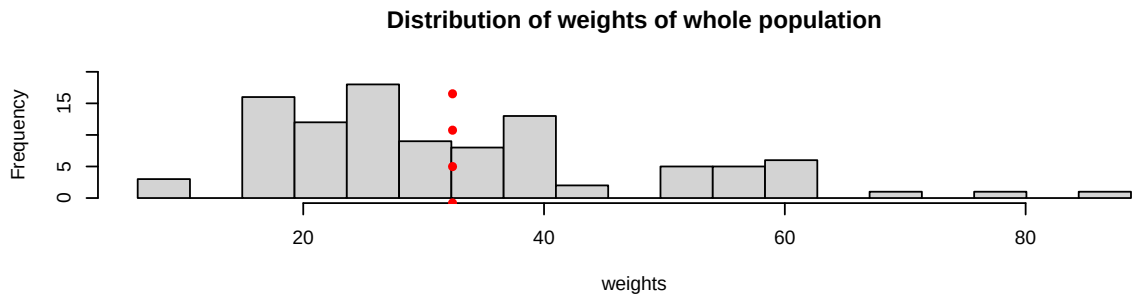
Answer

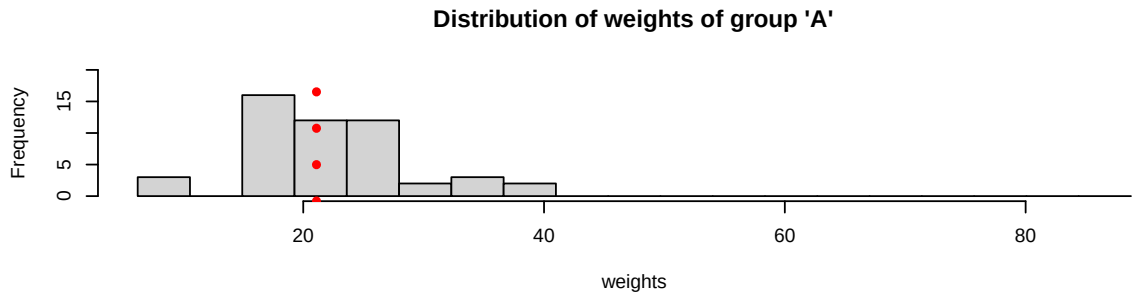
From the above computed attributes, groupA's mean and median is only half the groupB's. However, groupB has a standard deviation that is around twice of groupA's. While the whole population's mean and median is lied between groupA and groupB. The standard deviation is greater than both groupA's and groupB's

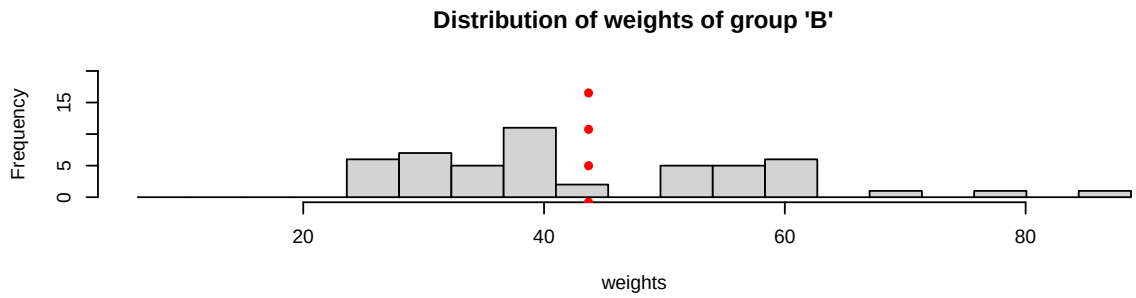
b. Simple graphical attributes.

i. *(8 marks)* Draw (suitably labelled) histograms of the weight for the whole population, only the blocks in group `A`, and only the blocks in group `B`. **Answer**

```
A = blocks[blocks$group == "A",]
B = blocks[blocks$group == "B",]
xlim <- extendrange(blocks$weight)
ylim <- c(0,20)
breaks <- seq(min(xlim), max(xlim), length.out = 20)
savePar <- par(mfrow = c(3,1))
hist(blocks$weight, xlim = xlim,ylim = ylim, breaks = breaks, main = "Distribution of weights
abline(v = mean(blocks[,'weight']), col = 'red', lty = 3, lwd = 5)
hist(A$weight, xlim = xlim,ylim = ylim, breaks = breaks, main = "Distribution of weights of
abline(v = group_A_mean, col = 'red', lty = 3, lwd = 5)
hist(B$weight, xlim = xlim,ylim = ylim, breaks = breaks, main = "Distribution of weights of
abline(v = group_B_mean, col = 'red', lty = 3, lwd = 5)
```








```
par(savePar)
```

ii. *(6 marks)* Using formula notation (e.g., `boxplot(weight ~ group, data = blocks, ...)`) draw pairs of (suitably labelled) boxplots comparing the two groups:

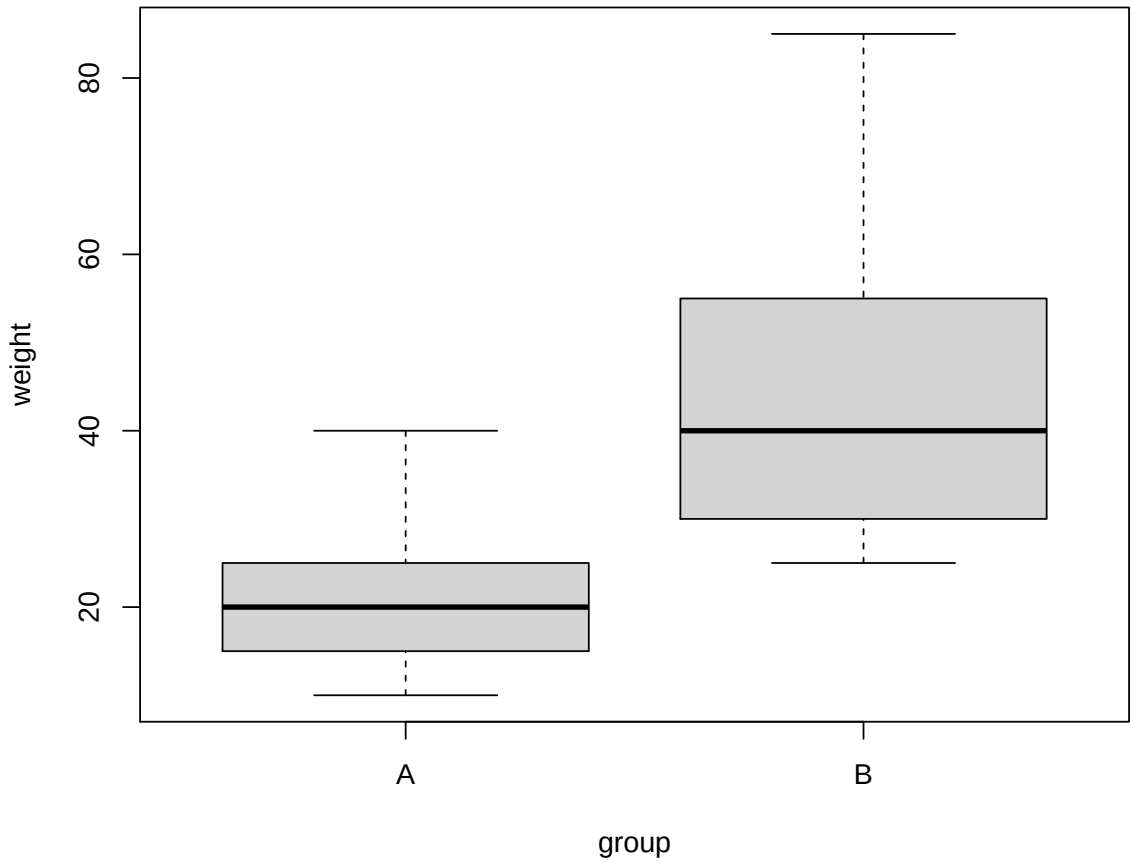
- first with respect to block weights and
- then with respect to the perimeters.

Comment on how the two groups compare.

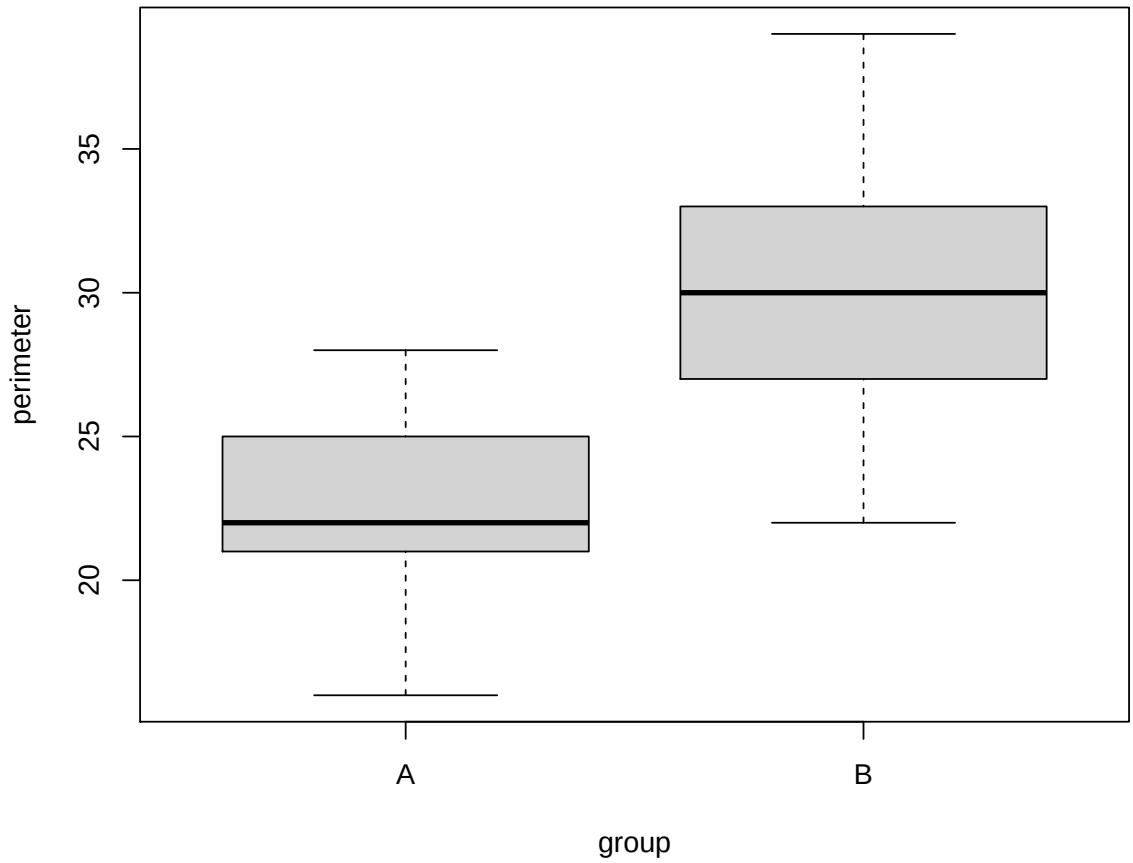
Show your code.

Answer

```
boxplot(weight~group, data = blocks)
```




```
boxplot(perimeter~group, data=blocks)
```



The boxplot shows that the median weight of group B is around 40 and is much higher than group A's median which is around 20. Also, the range of group B is from about mid 20s to high 80s while group A's range is from around 10 to 40. When we look into the boxplot of perimeter, the median perimeter of group A is around 23 and is lower than group B's median which is around 30. And the range of group A is from around 15 to 28 while group B's range is from low 20's to almost 40. These results indicates group B is higher than group A in both weight and perimeter, which means mostly group B's blocks are heavier and bigger than group A's.