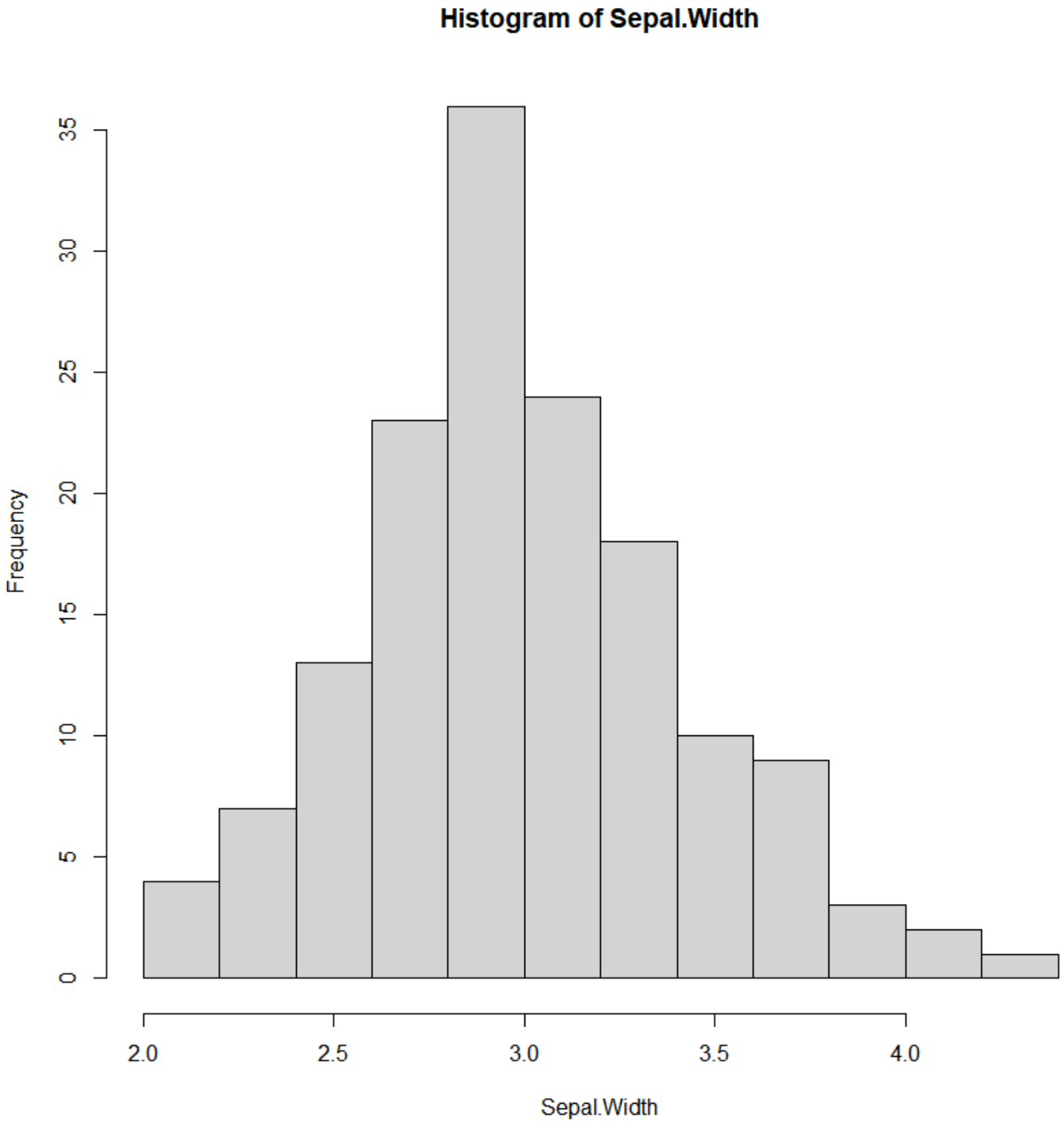
Veronika Titarchuk HW3

1. Using the iris dataset…

   a. Make a histogram of the variable Sepal.Width.

attach(iris)

hist(Sepal.Width)



b. Based on the histogram from #1a, which would you expect to be higher, the mean or the median? Why?

Since it is right-skewed, we can expect the mean to be bigger than the median.

c. Confirm your answer to #1b by actually finding these values.

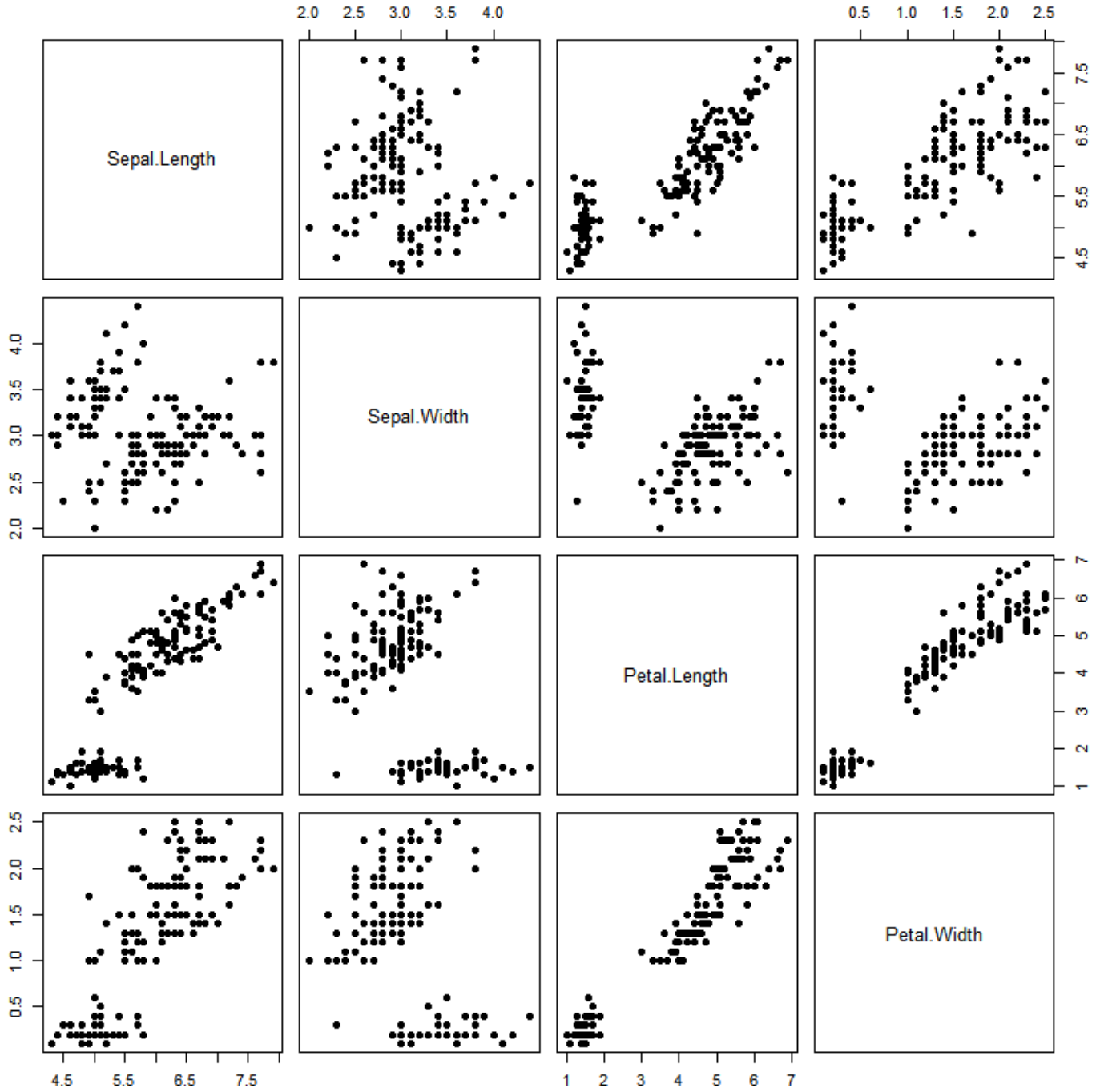
```
> mean(Sepal.Width)
[1] 3.057333
> median(Sepal.Width)
[1] 3
>
```

d. Only 27% of the flowers have a Sepal.Width higher than __3.3_____ cm.

```
> quantile(Sepal.Width,c(1-.27))
73%
3.3
```

e. Make scatterplots of each pair of the numerical variables in iris (There should be 6 pairs/plots).

```
#1e
pairs(iris[,c(1:4)],pch=16)
```



f. Based on #1e, which two variables appear to have the strongest relationship? And which two appear to have the weakest relationship?

Strongest relationship: Petal.Length with Petal.Width since we can see strong linear relationship between them

Weakest relationship: Sepal.Length with Sepal.Width since we can't see any kind of relationship between them.

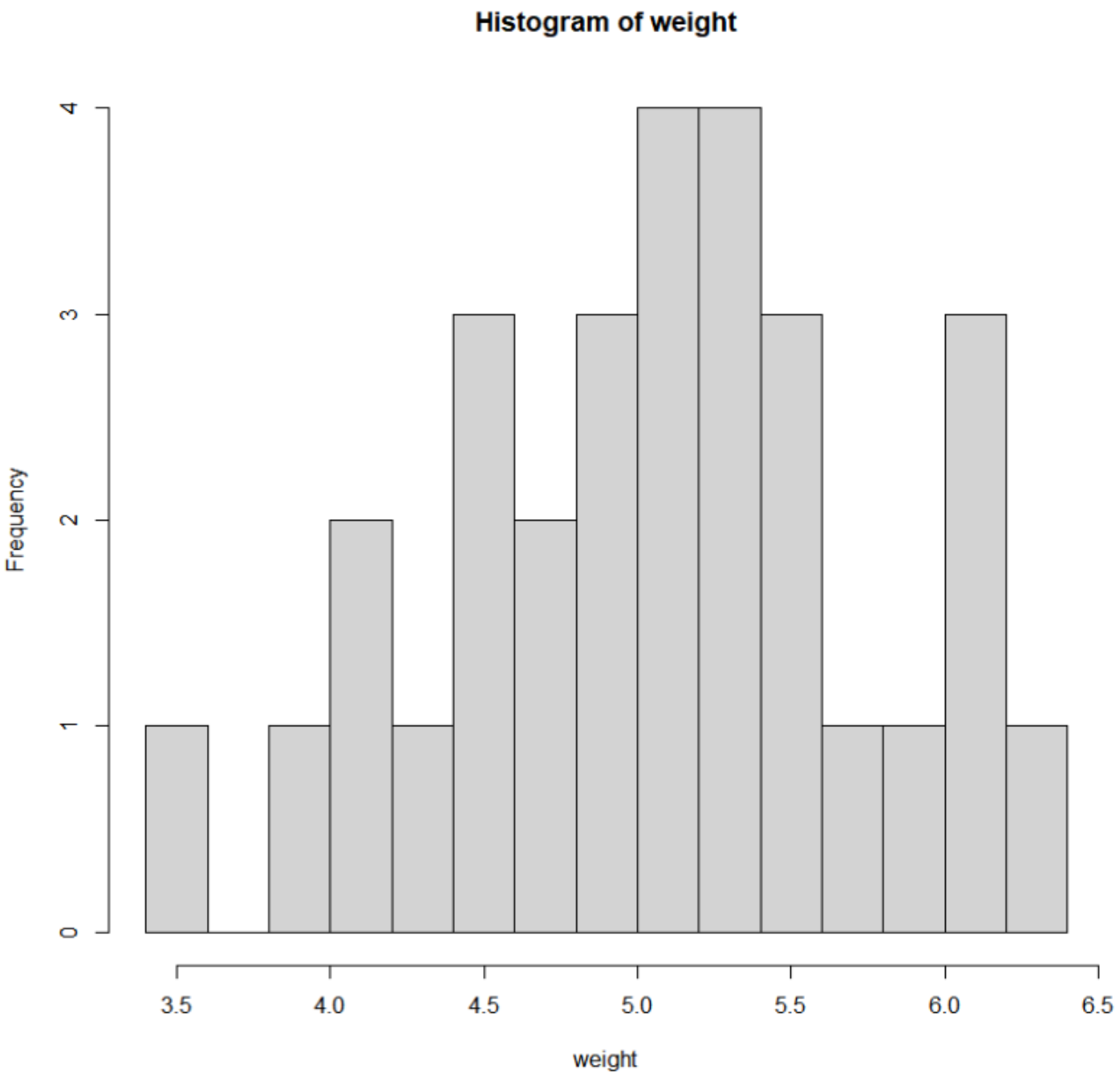
2. Using the PlantGrowth dataset…

a. Make a histogram of the variable weight with breakpoints (bin edges) at every 0.3 units, starting at 3.3.

Since we have 3 unit long x axis, in order for bin edges to be 0.3, we need to divide 3 by 10, so breaks=10:

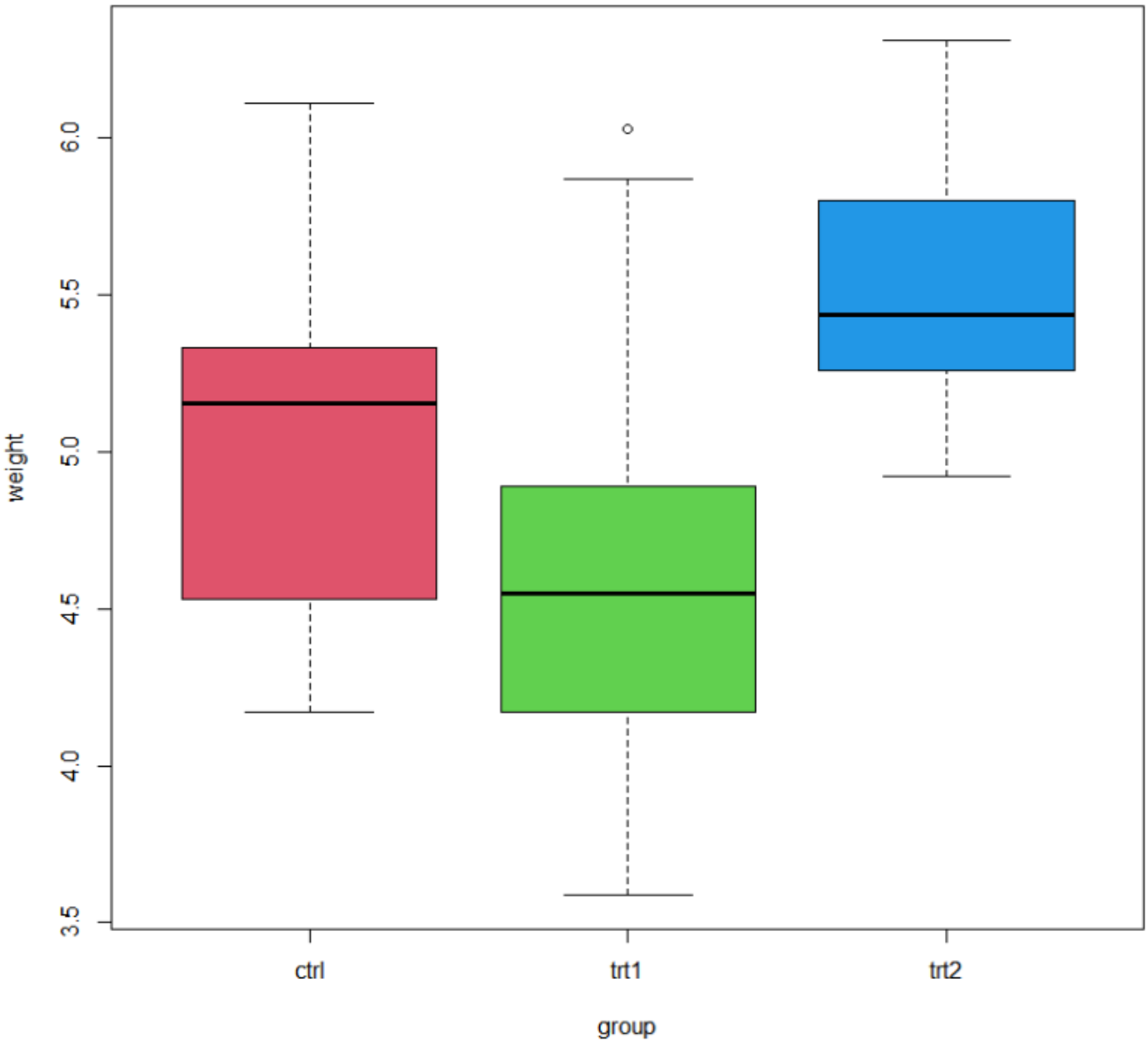
```
#2a
attach(PlantGrowth)
head(PlantGrowth)

hist(weight,breaks=10)
?hist
```



b. Make boxplots of weight separated by group in a single graph.

```
#2b
boxplot(weight~group,col=2:4)
```



c. Based on the boxplots in #2b, approximately what percentage of the "trt1" weights are below the minimum "trt2" weight?

Since Q3 of trt1 is on the same level as minimum of trt2, we can conclude that 75% of the "trt1" weights are below the minimum "trt2" weight.

d. Find the exact percentage of the "trt1" weights that are below the minimum "trt2" weight.

```
> #2d
> min(weight[group=="trt2"])
[1] 4.92
> sort(weight[group=="trt1"])
 [1] 3.59 3.83 4.17 4.32 4.41 4.69 4.81 4.89 5.87 6.03
>
```

There are 8 observations of trt1 that are less than a minimum of trt2. Since the total amount of observations int trt1 is 10, then we get 80%. We can make a conclusion that visual representation in #2c gave us a little wrong idea since trt2 min apparently are on the same level as Q3 of trt1.

e. Only including plants with a weight above 5.5, make a barplot of the variable group. Make the barplot colorful using some color palette (in R, try running ?heat.colors and/or check out https://www.r-bloggers.com/palettes-in-r/).

```
32
33  #2e
34  tgroup <- table(group[weight>5.5])
35  colors <-  brewer.pal(3, "Purples")
36  barplot(tgroup, main = "Barplot for Plant Growth dataset", xlab = "Group", ylab = "Number of plants with the weight >5.5", col=colors)
37  ?brewer.pal
```