1. **Problem**

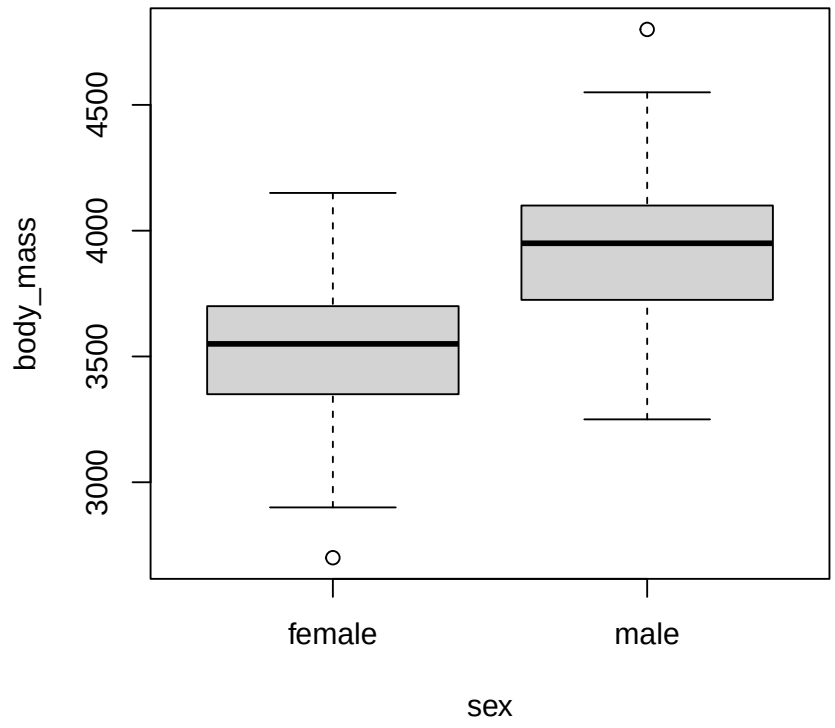
   The `penguins` data in base R provides various measurements of adult penguins from three different species. See `?penguins` for more details. Originally, the data was used to study sex dimorphism separately for the three species.

   The first three rows of the data can be inspected as follows. Employ `summary()` to obtain a first overview.

```
data("penguins", package = "datasets")
head(penguins, 3)
```

```
##   species    island bill_len bill_dep flipper_len body_mass    sex year
## 1  Adelie Torgersen     39.1     18.7         181      3750   male 2007
## 2  Adelie Torgersen     39.5     17.4         186      3800 female 2007
## 3  Adelie Torgersen     40.3     18.0         195      3250 female 2007
```

   Explore the sex differences with respect to body mass (weight, in grams) of the penguins. Create parallel boxplots of weight by sex, such as the one below, separately for the three species.



   Which species does this plot pertain to?

   To complement the plot complete the corresponding table of groupwise statistics:

	median	mean	std. deviation
female			
male			

The average weight difference of is thus slightly higher / lower than the median weight difference of .

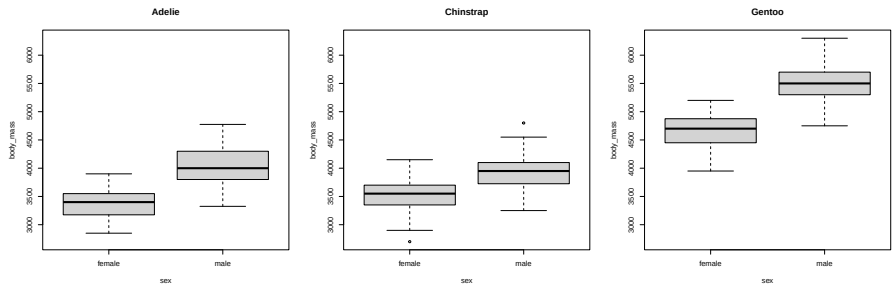
Compute the full `summary()` of weight by sex for this species and select the correct statements in the following list.

Less than half of the male penguins weigh more than 3731.25 grams. / None of the penguins weighs less than 2700 grams. / The standard deviation of weight is lower for males compared to females.

Solution

One way to obtain the exploratory boxplots separately for the three species is:

```
par(mfrow = c(1, 3))
for(i in levels(penguins$species)) plot(body_mass ~ sex, data = penguins,
  subset = species == i, main = i, ylim = range(penguins$body_mass, na.rm = TRUE))
```



The question shows the parallel boxplots for the Chinstrap species.

Groupwise statistics of body mass by sex and species (including mean, median, and standard deviation) can be obtained by aggregating the data with the combined `summary()` and `sd()` functions.

```
aggregate(body_mass ~ sex + species, data = penguins,
  FUN = function(x) c(summary(x), 'Std. dev.' = sd(x)))
```

```
##      sex   species body_mass.Min. body_mass.1st Qu. body_mass.Median
## 1 female    Adelie      2850.0000         3175.0000        3400.0000
## 2   male    Adelie      3325.0000         3800.0000        4000.0000
## 3 female Chinstrap      2700.0000         3362.5000        3550.0000
## 4   male Chinstrap      3250.0000         3731.2500        3950.0000
## 5 female    Gentoo      3950.0000         4462.5000        4700.0000
## 6   male    Gentoo      4750.0000         5300.0000        5500.0000
##   body_mass.Mean body_mass.3rd Qu. body_mass.Max. body_mass.Std. dev.
## 1      3368.8356         3550.0000      3900.0000            269.3801
## 2      4043.4932         4300.0000      4775.0000            346.8116
## 3      3527.2059         3693.7500      4150.0000            285.3339
## 4      3938.9706         4100.0000      4800.0000            362.1376
## 5      4679.7414         4875.0000      5200.0000            281.5783
## 6      5484.8361         5700.0000      6300.0000            313.1586
```

Based on this the remaining elements of the question can be answered.