



LAB 5

DATA 101 - MAKING PREDICTIONS USING DATA

IF YOU WANT TO FOLLOW ALONG

- If you want to follow along for the barplot part of this lab, you can grab the data on canvas via the link in chat. You can use the code below to import it into R.

```
motor <- read.table("motor_vibrations.txt",header=FALSE, sep=",")
```

```
motor
```

ASSIGNMENT FEEDBACK

- Using levels

LEVELS

chickwts\$feed

levels(chickwts\$feed)

```
> chickwts$feed
 [1] horsebean horsebean horsebean horsebean horsebean horsebean horsebean
 [8] horsebean horsebean horsebean linseed   linseed   linseed   linseed
[15] linseed   linseed   linseed   linseed   linseed   linseed   linseed
[22] linseed   soybean   soybean   soybean   soybean   soybean   soybean
[29] soybean   soybean   soybean   soybean   soybean   soybean   soybean
[36] soybean   sunflower sunflower sunflower sunflower sunflower sunflower
[43] sunflower sunflower sunflower sunflower sunflower sunflower sunflower
[50] meatmeal  meatmeal  meatmeal  meatmeal  meatmeal  meatmeal  meatmeal
[57] meatmeal  meatmeal  meatmeal  casein    casein    casein    casein
[64] casein    casein    casein    casein    casein    casein    casein
[71] casein
Levels: casein horsebean linseed meatmeal soybean sunflower
> levels(chickwts$feed)
[1] "casein"      "horsebean" "linseed"    "meatmeal"  "soybean"   "sunflower"
```

LEVELS

```
chickwts$Feed <- chickwts$feed
```

```
levels(chickwts$Feed)<- c("Casein", "Horsebean", "Linseed", "Meatmeal","Soybean","Sunflower")
```

```
chickwts$Feed
```

```
> chickwts$Feed
```

```
[1] Horsebean Horsebean Horsebean Horsebean Horsebean Horsebean Horsebean  
[8] Horsebean Horsebean Horsebean Linseed Linseed Linseed Linseed  
[15] Linseed Linseed Linseed Linseed Linseed Linseed Linseed  
[22] Linseed Soybean Soybean Soybean Soybean Soybean Soybean  
[29] Soybean Soybean Soybean Soybean Soybean Soybean Soybean  
[36] Soybean Sunflower Sunflower Sunflower Sunflower Sunflower Sunflower  
[43] Sunflower Sunflower Sunflower Sunflower Sunflower Sunflower Sunflower  
[50] Meatmeal Meatmeal Meatmeal Meatmeal Meatmeal Meatmeal Meatmeal  
[57] Meatmeal Meatmeal Meatmeal Casein Casein Casein Casein  
[64] Casein Casein Casein Casein Casein Casein Casein  
[71] Casein
```

```
Levels: Casein Horsebean Linseed Meatmeal Soybean Sunflower
```

```
> levels(chickwts$Feed)
```

```
[1] "Casein" "Horsebean" "Linseed" "Meatmeal" "Soybean" "Sunflower"
```

NOTE: the new level vector has to be the same length as the original vector, even if there are duplicates!

LEVELS

cuckoos\$species

```
[1] meadow.pipit meadow.pipit meadow.pipit meadow.pipit meadow.pipit
[6] meadow.pipit meadow.pipit meadow.pipit meadow.pipit meadow.pipit
[11] meadow.pipit meadow.pipit meadow.pipit meadow.pipit meadow.pipit
[16] meadow.pipit meadow.pipit meadow.pipit meadow.pipit meadow.pipit
[21] meadow.pipit meadow.pipit meadow.pipit meadow.pipit meadow.pipit
[26] meadow.pipit meadow.pipit meadow.pipit meadow.pipit meadow.pipit
[31] meadow.pipit meadow.pipit meadow.pipit meadow.pipit meadow.pipit
[36] meadow.pipit meadow.pipit meadow.pipit meadow.pipit meadow.pipit
[41] meadow.pipit meadow.pipit meadow.pipit meadow.pipit meadow.pipit
[46] tree.pipit tree.pipit tree.pipit tree.pipit tree.pipit
[51] tree.pipit tree.pipit tree.pipit tree.pipit tree.pipit
[56] tree.pipit tree.pipit tree.pipit tree.pipit tree.pipit
[61] hedge.sparrow hedge.sparrow hedge.sparrow hedge.sparrow hedge.sparrow
[66] hedge.sparrow hedge.sparrow hedge.sparrow hedge.sparrow hedge.sparrow
[71] hedge.sparrow hedge.sparrow hedge.sparrow hedge.sparrow robin
[76] robin robin robin robin robin
[81] robin robin robin robin robin
[86] robin robin robin robin robin
[91] pied.wagtail pied.wagtail pied.wagtail pied.wagtail pied.wagtail
[96] pied.wagtail pied.wagtail pied.wagtail pied.wagtail pied.wagtail
[101] pied.wagtail pied.wagtail pied.wagtail pied.wagtail pied.wagtail
[106] wren wren wren wren wren
[111] wren wren wren wren wren
[116] wren wren wren wren wren
Levels: hedge.sparrow meadow.pipit pied.wagtail robin tree.pipit wren
```


LEVELS

```
levels(cuckoos$m.pipitFactor) <- c("other", "meadow.pipit", "other", "other", "other", "other")
```

```
# OR using rep()
```

```
# levels(cuckoos$m.pipitFactor) <- c("other", "meadow.pipit", rep("other",4))
```

R will automatically remove the duplicates from levels.

```
[1] meadow.pipit meadow.pipit meadow.pipit meadow.pipit meadow.pipit
[6] meadow.pipit meadow.pipit meadow.pipit meadow.pipit meadow.pipit
[11] meadow.pipit meadow.pipit meadow.pipit meadow.pipit meadow.pipit
[16] meadow.pipit meadow.pipit meadow.pipit meadow.pipit meadow.pipit
[21] meadow.pipit meadow.pipit meadow.pipit meadow.pipit meadow.pipit
[26] meadow.pipit meadow.pipit meadow.pipit meadow.pipit meadow.pipit
[31] meadow.pipit meadow.pipit meadow.pipit meadow.pipit meadow.pipit
[36] meadow.pipit meadow.pipit meadow.pipit meadow.pipit meadow.pipit
[41] meadow.pipit meadow.pipit meadow.pipit meadow.pipit meadow.pipit
[46] other        other        other        other        other
[51] other        other        other        other        other
[56] other        other        other        other        other
[61] other        other        other        other        other
[66] other        other        other        other        other
[71] other        other        other        other        other
[76] other        other        other        other        other
[81] other        other        other        other        other
[86] other        other        other        other        other
[91] other        other        other        other        other
[96] other        other        other        other        other
[101] other        other        other        other        other
[106] other        other        other        other        other
[111] other        other        other        other        other
[116] other        other        other        other        other
Levels: other meadow.pipit
```

ASSIGNMENT FEEDBACK

- Using levels
- Using subset

SUBSET

Subset format

Subset(dataframe, Boolean)

- If the boolean statement is true, the row will be part of the result of the subset.
- The boolean can be anything, ==, !=, <=, >=, <, >
- Subset will return a dataframe.

examples

```
cuckoosMPipit <- subset(cuckoos, m.pipitFactor == "meadow.pipit")
```

```
cuckoosOther <- subset(cuckoos, m.pipitFactor == "other")
```

```
cuckoosLongLength <- subset(cuckoos, length>23)
```

```
chick240<-subset(chickwts, weight)
```

```
chickwtsCasein<-subset(chickwts, feed=="casein")
```

ASSIGNMENT FEEDBACK

- Using levels
- Using subset
- \$ operator

\$ OPERATOR

The \$ operator is the same thing as indexing for a column, only it's easier for the coder to read what the code is actually doing.

For example, `chickwts$feed` is easy to understand, `chickwts[,2]` is less useful to the reader of your code what you're trying to do. You should use this over indexing if your dataframe is labeled.

ASSIGNMENT FEEDBACK

- Using levels
- Using subset
- \$ operator
- Specific code

SPECIFIC CODE

When possible, try to write code that returns only the answer you're asked for.

For example, if you're asked how many observations there are, you should return a function like `nrow(cuckoos)` or `dim(cuckoos)[1]` that returns the exact number, rather than `summary(cuckoos)` or simply printing `cuckoos` and scrolling to the end.

Same goes for the `mean()` function. Avoid using the summary function if all you're asked for is the mean.

ASSIGNMENT FEEDBACK

- Using levels
- Using subset
- \$ operator
- Specific code

If you're ever unclear about a question or aren't sure where or how to get started, message me. You can also send me your completed assignment and I can give it a look over and point out anything you should take another look at.

ASSIGNMENT FEEDBACK

- Using levels
- Using subset
- \$ operator
- Specific code

If you're ever unclear about a question or aren't sure where or how to get started, message me. You can also send me your completed assignment and I can give it a look over and point out anything you should take another look at.

Expect assignment 3 marks starting Friday and over the weekend.

LAB 5 - BOXPLOTS

BOTH THIS LAB AND ASSIGNMENT WILL BE ABOUT
CREATING AND FORMATTING BOXPLOTS.

DATA

You can download the data on canvas, link in the chat.

```
motor <- read.table("motor_vibrations.txt",header=FALSE, sep=",")
```

```
motor
```

	V1	V2	V3	V4	V5
1	13.1	16.3	13.7	15.7	13.5
2	15.0	15.7	13.9	13.7	13.4
3	14.0	17.2	12.4	14.4	13.2
4	14.4	14.9	13.8	16.0	12.7
5	14.0	14.4	14.9	13.9	13.4
6	11.6	17.2	13.3	14.7	12.3

DATA

You can download the data on canvas, link in the chat.

```
motor <- read.table("motor_vibrations.txt",header=FALSE, sep=",")
```

```
motor
```

	V1	V2	V3	V4	V5
1	13.1	16.3	13.7	15.7	13.5
2	15.0	15.7	13.9	13.7	13.4
3	14.0	17.2	12.4	14.4	13.2
4	14.4	14.9	13.8	16.0	12.7
5	14.0	14.4	14.9	13.9	13.4
6	11.6	17.2	13.3	14.7	12.3

- The columns are a sample of 5 different motors.
- The rows are a different brand of ball bearing.

DATA

You can download the data on canvas, link in the chat.

```
motor <- read.table("motor_vibrations.txt",header=FALSE, sep=",")
```

```
motor
```

	V1	V2	V3	V4	V5
1	13.1	16.3	13.7	15.7	13.5
2	15.0	15.7	13.9	13.7	13.4
3	14.0	17.2	12.4	14.4	13.2
4	14.4	14.9	13.8	16.0	12.7
5	14.0	14.4	14.9	13.9	13.4
6	11.6	17.2	13.3	14.7	12.3

- The columns are a sample of 5 different motors.
- The rows are a different brand of ball bearing.
- Each sample then is the vibration of each engine using a particular ball bearing.
- We want to know if there are differences in the mean vibration between engine brands.

DATA

Naming the columns.

```
names(motor) <- c("Ford", "Nissan", "BMW", "Honda", "Audi")
```

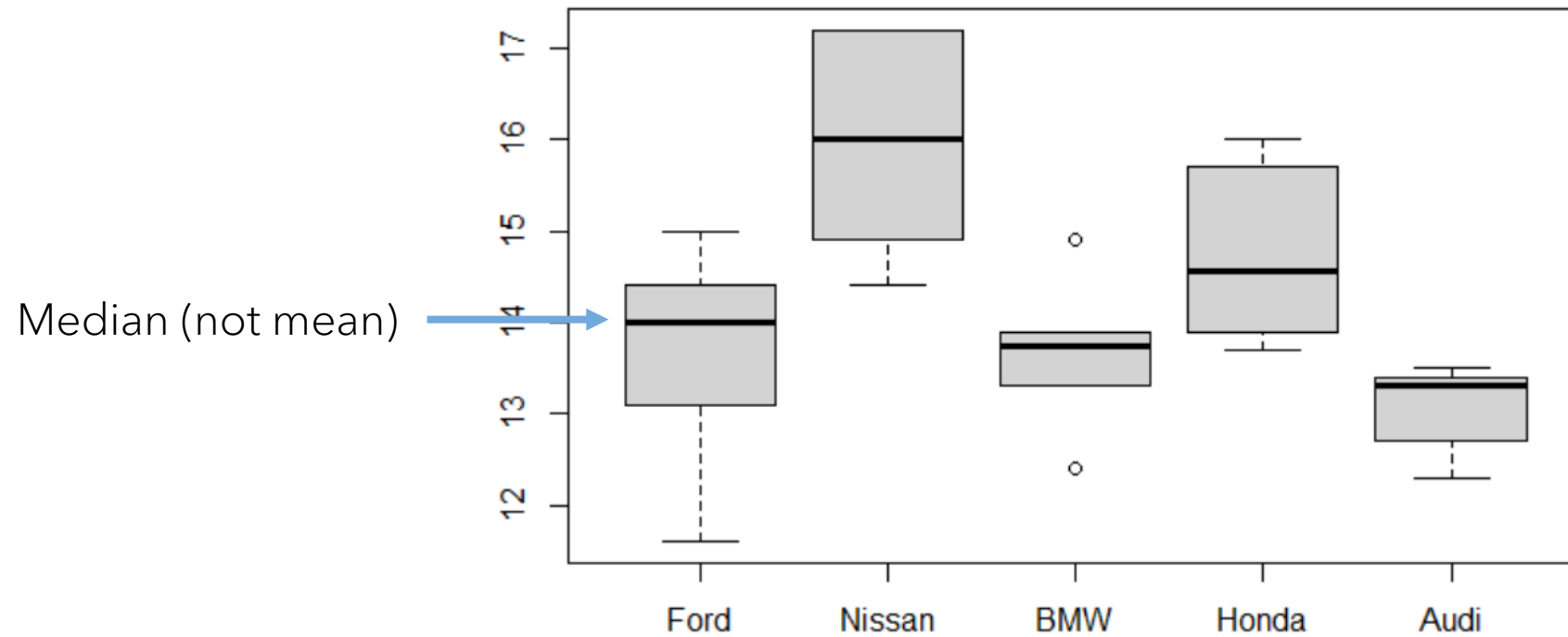
```
motor
```

	Ford	Nissan	BMW	Honda	Audi
1	13.1	16.3	13.7	15.7	13.5
2	15.0	15.7	13.9	13.7	13.4
3	14.0	17.2	12.4	14.4	13.2
4	14.4	14.9	13.8	16.0	12.7
5	14.0	14.4	14.9	13.9	13.4
6	11.6	17.2	13.3	14.7	12.3

DATA

A boxplot.

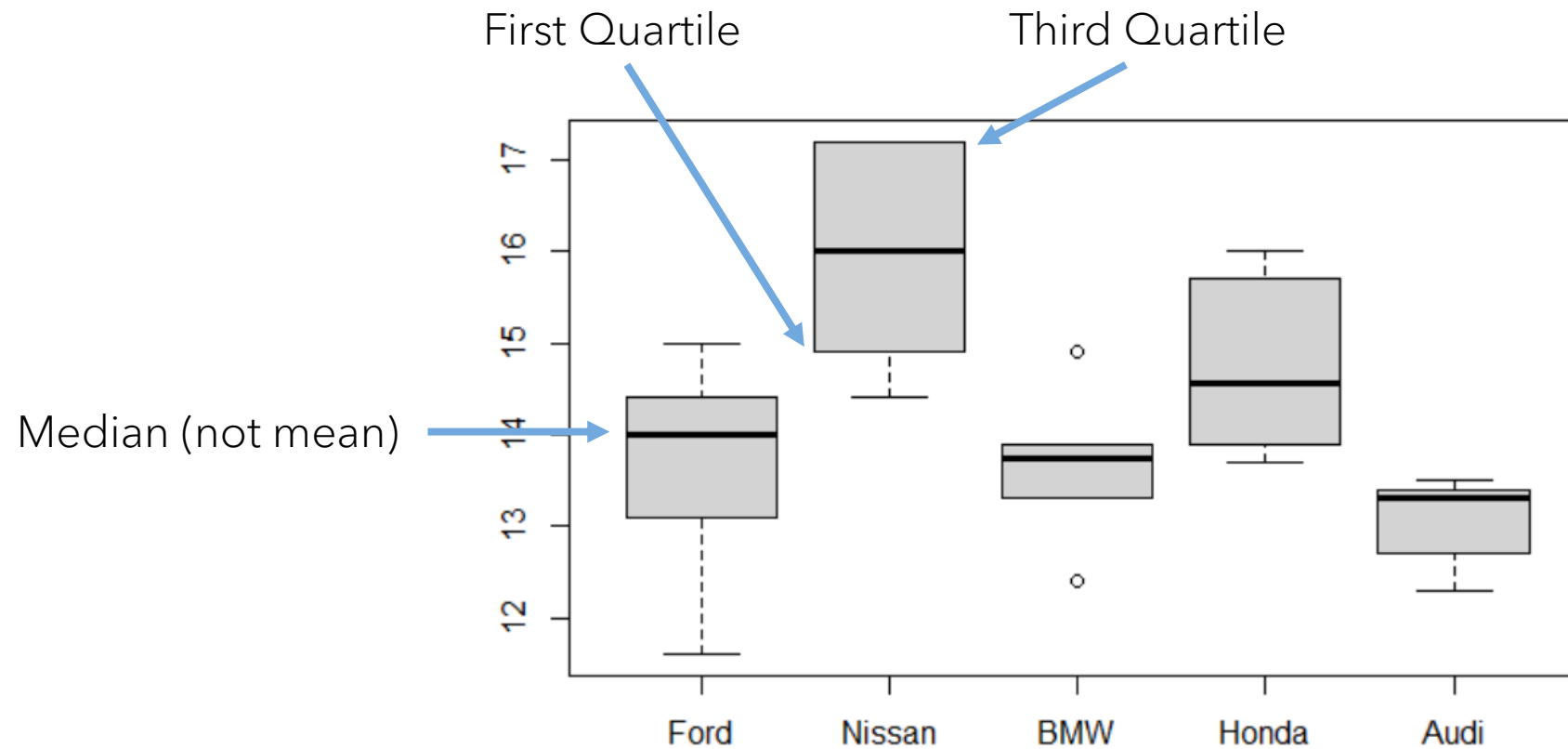
```
boxplot(motor)
```



DATA

A boxplot.

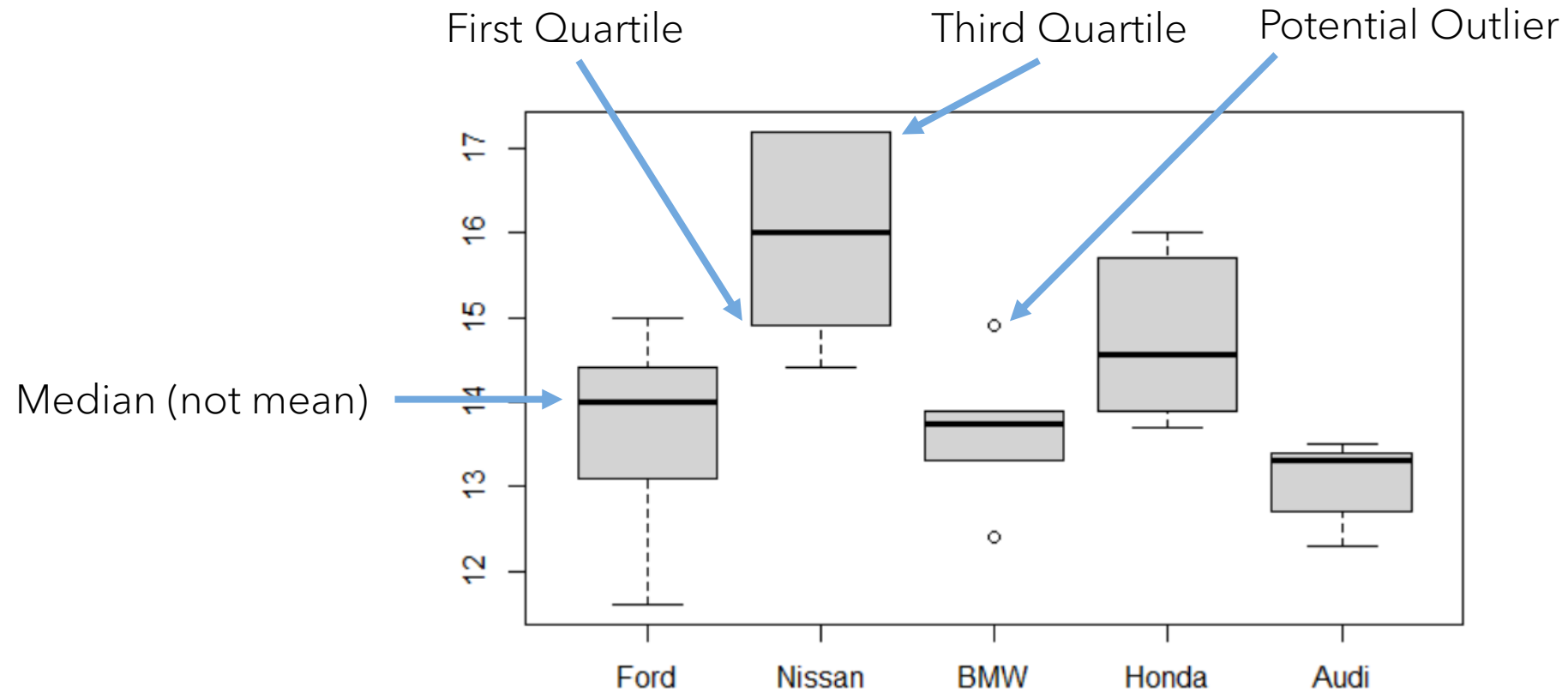
```
boxplot(motor)
```



DATA

A boxplot.

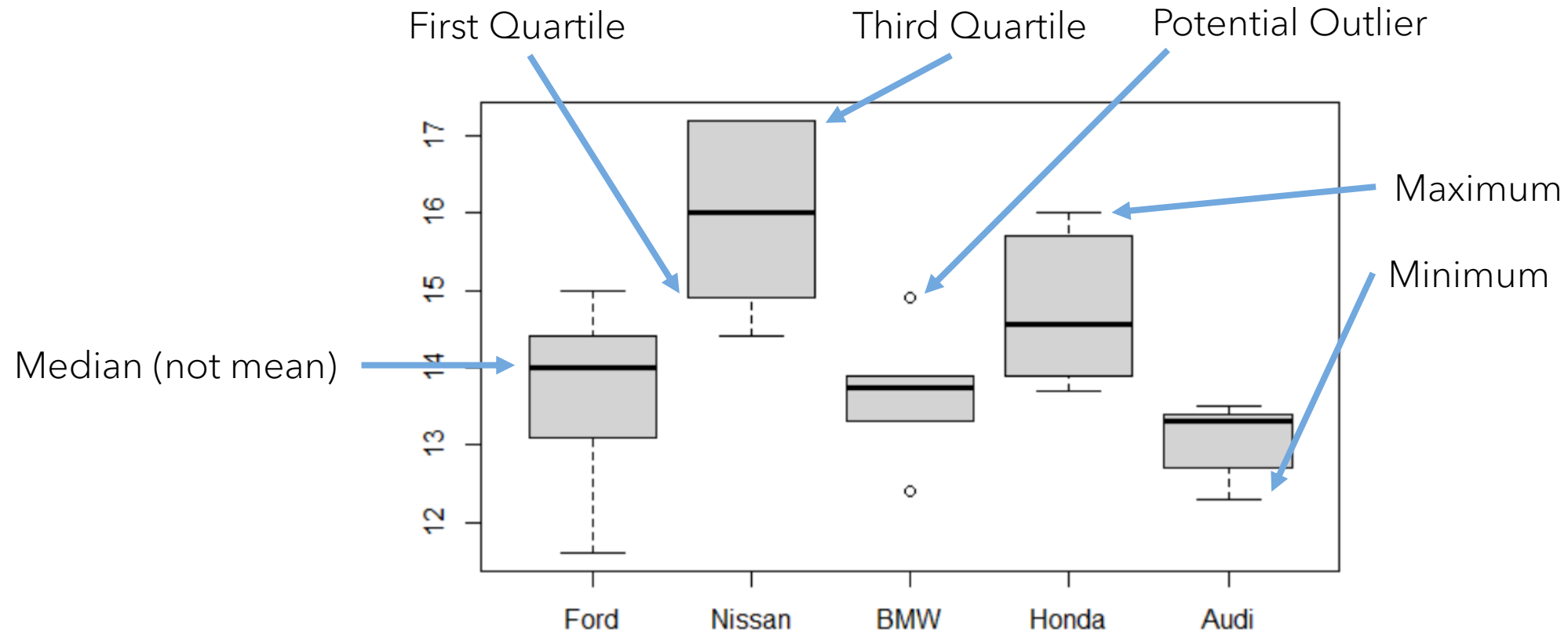
```
boxplot(motor)
```



BOXPLOT

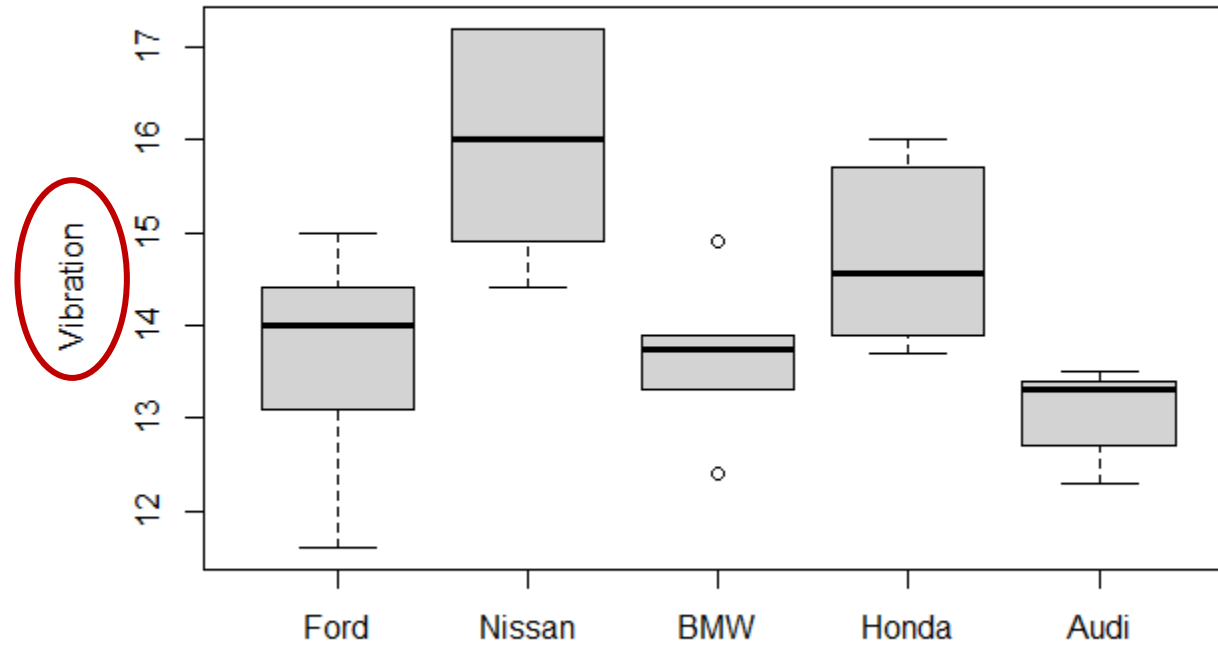
A boxplot.

boxplot(motor)



Y-AXIS LABEL

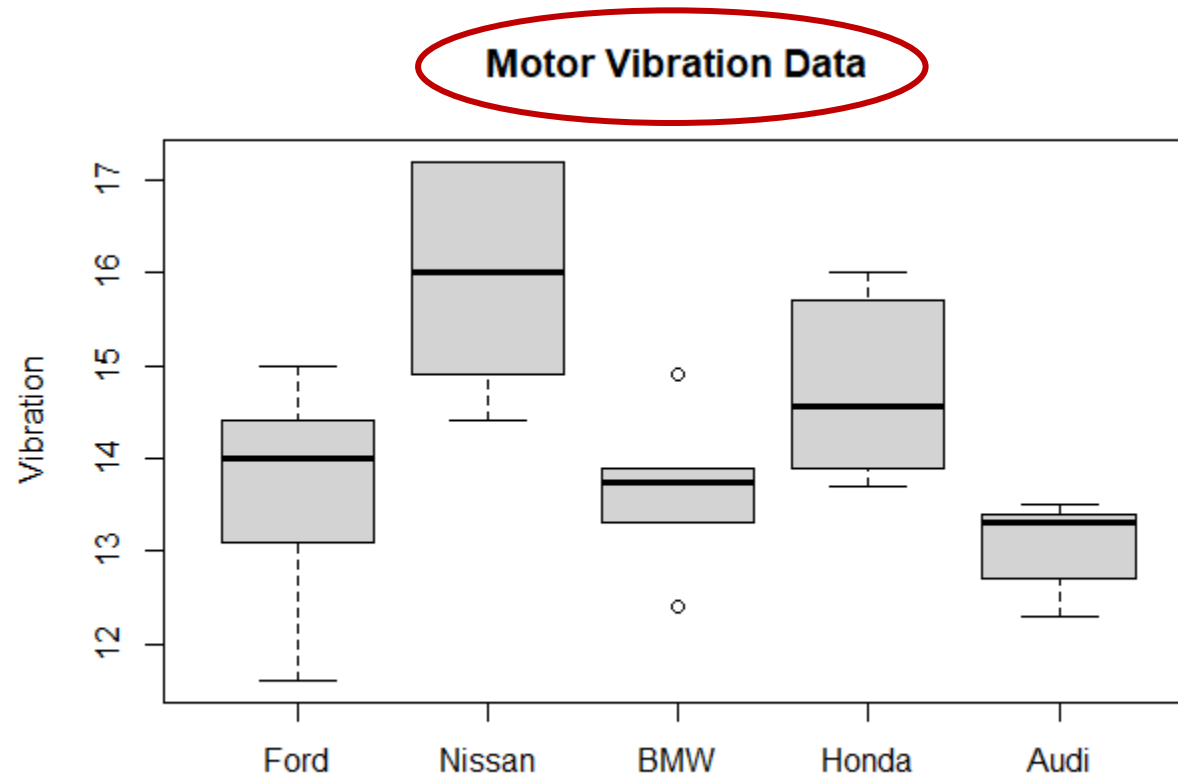
```
boxplot(motor, ylab="Vibration")
```



TITLES

```
boxplot(motor, ylab="Vibration")
```

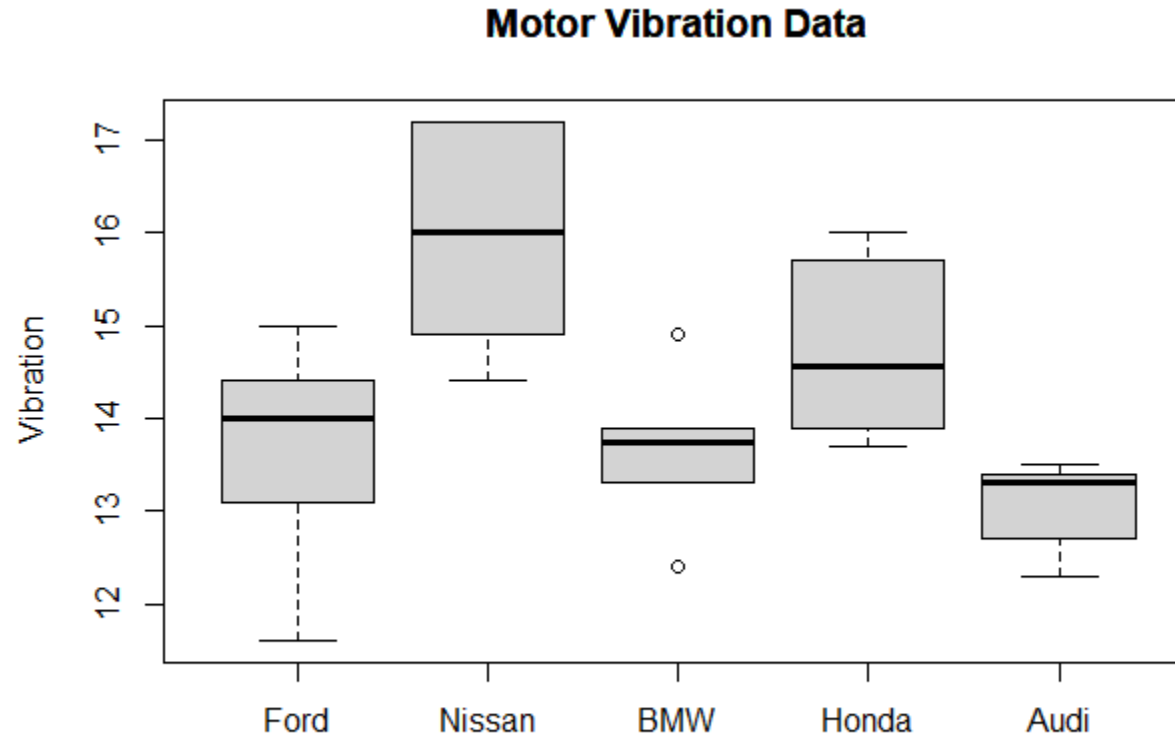
```
title("Motor Vibration Data")
```



SUBTITLES (BOTTOM)

```
boxplot(motor, ylab="Vibration")
```

```
title(main="Motor Vibration Data", sub="The Effects of Different Bearings")
```



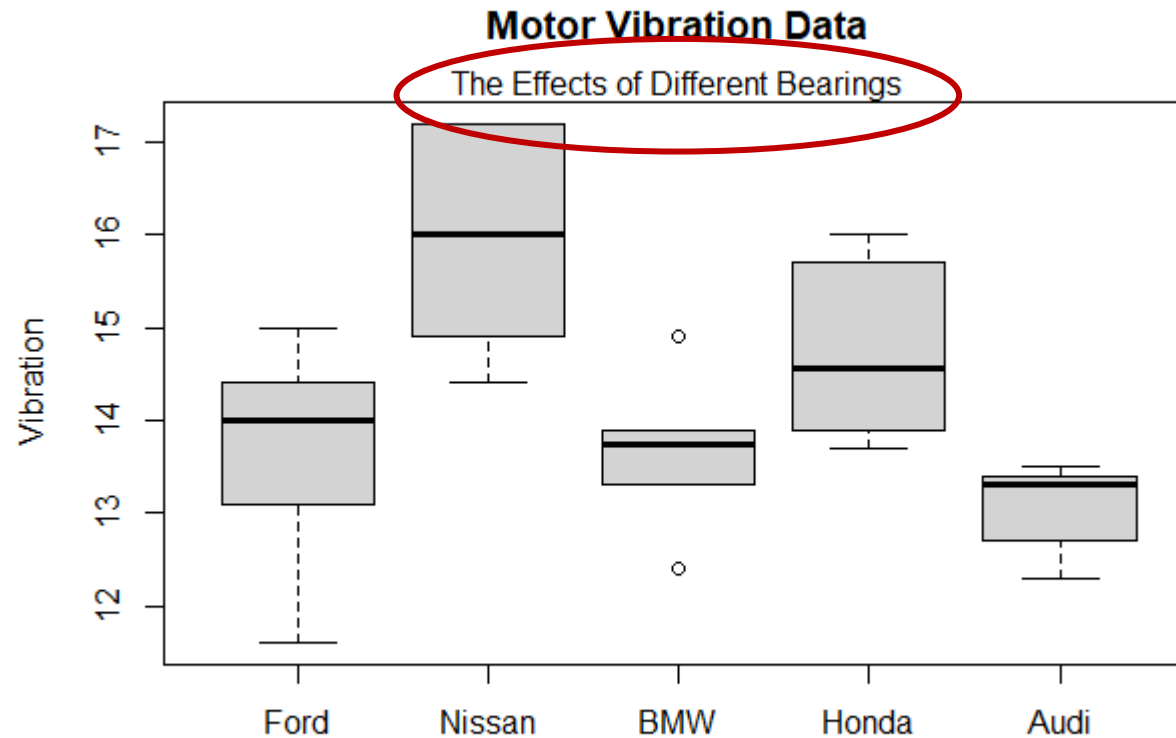
The Effects of Different Bearings

SUBTITLES (TOP)

```
boxplot(motor, ylab="Vibration")
```

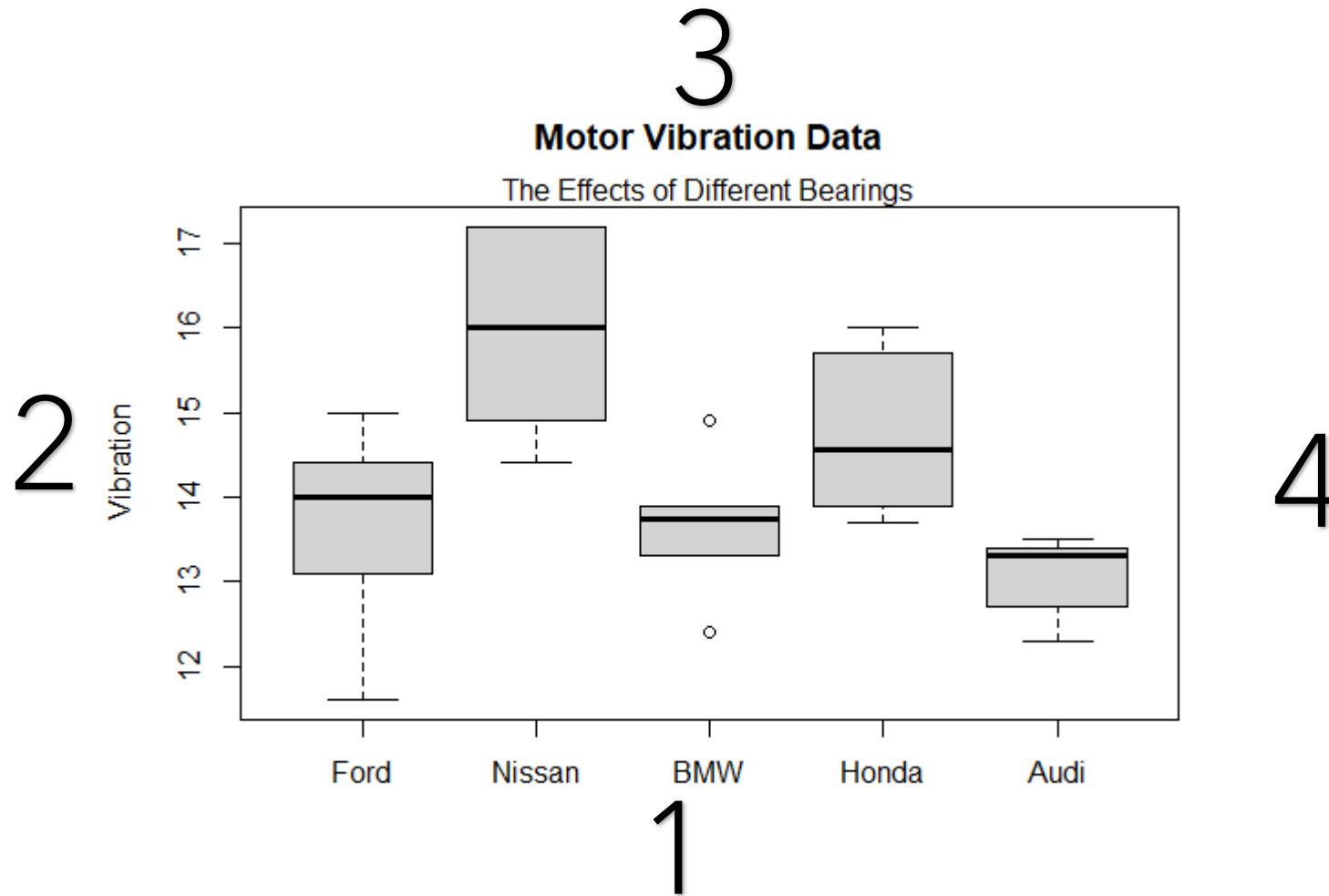
```
title(main="Motor Vibration Data")
```

```
mtext("The Effects of Different Bearings", side=3, line=0)
```



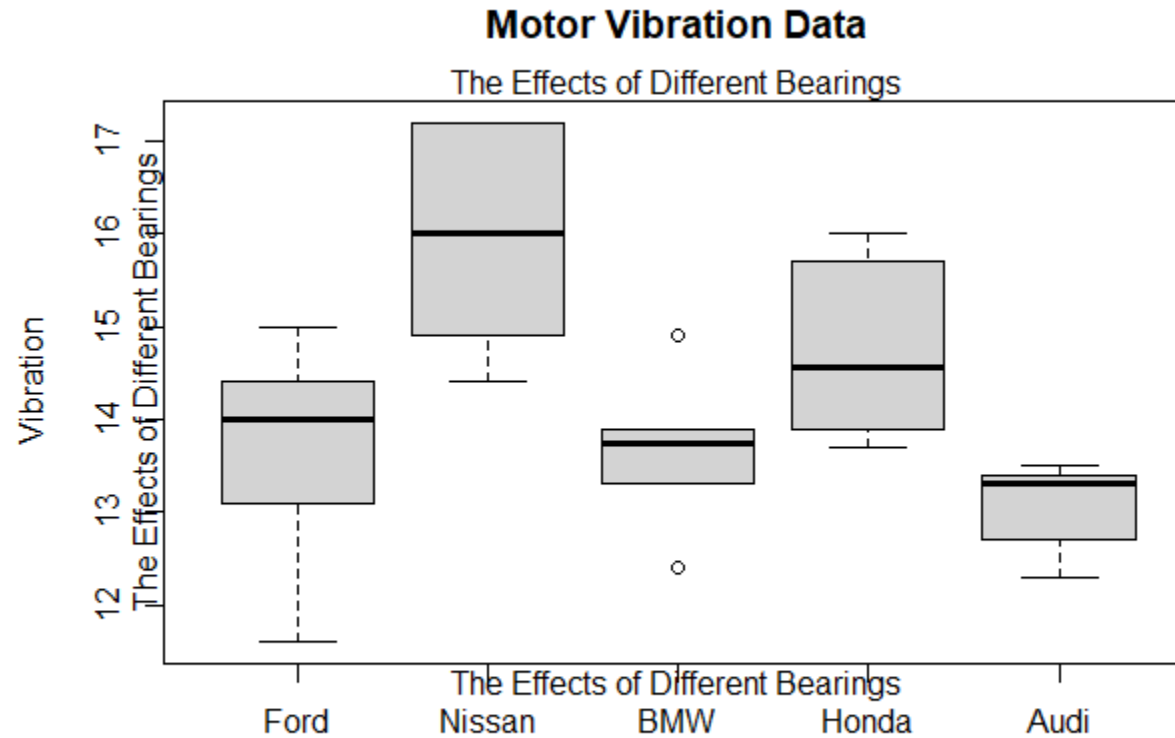
SIDES

```
mtext("The Effects of Different Bearings", side=3, line=0)
```



SIDES

```
mtext("The Effects of Different Bearings", side=1, line=0)
```

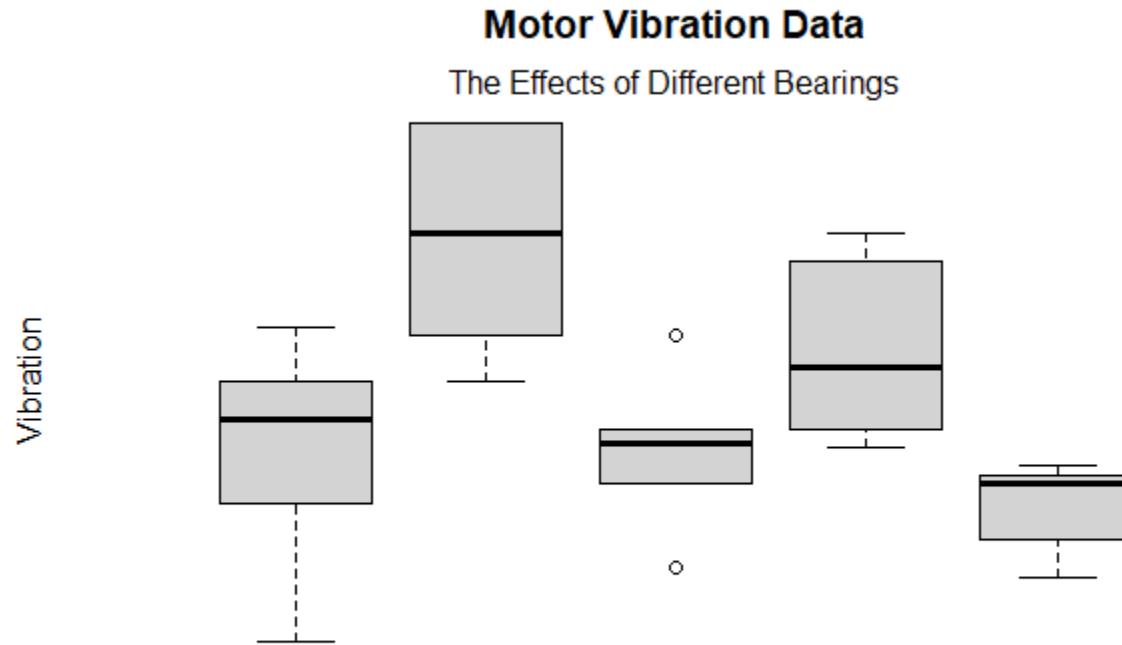


NO AXES

```
boxplot(motor, ylab="Vibration", axes=FALSE)
```

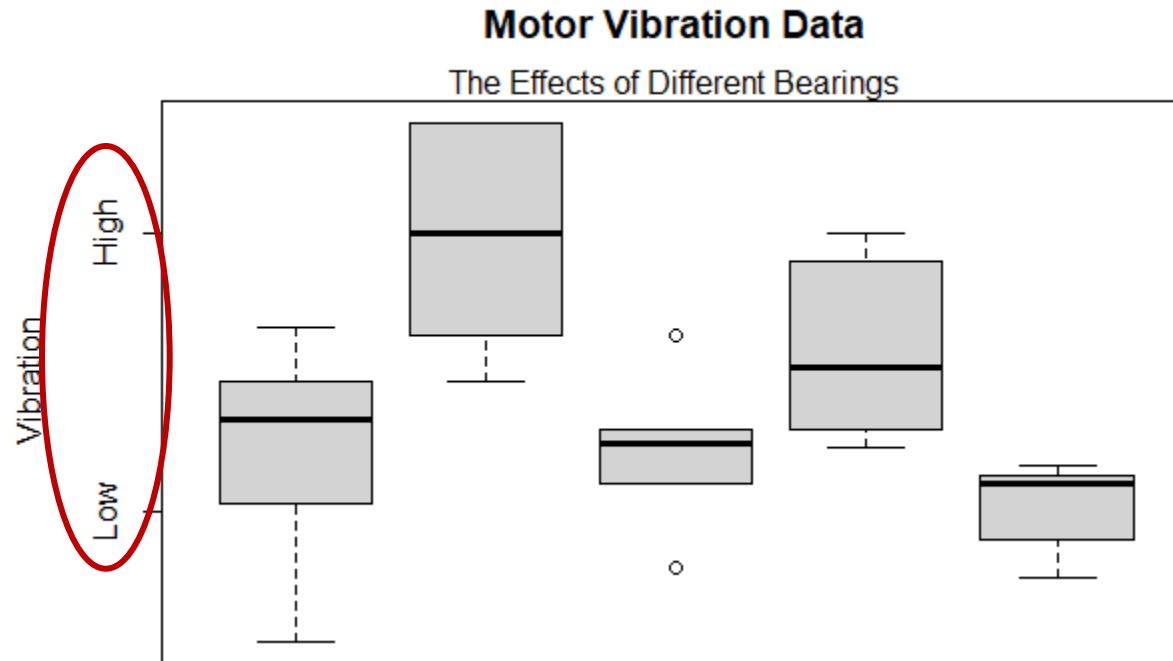
```
title(main="Motor Vibration Data")
```

```
mtext("The Effects of Different Bearings", side=3, line=0)
```



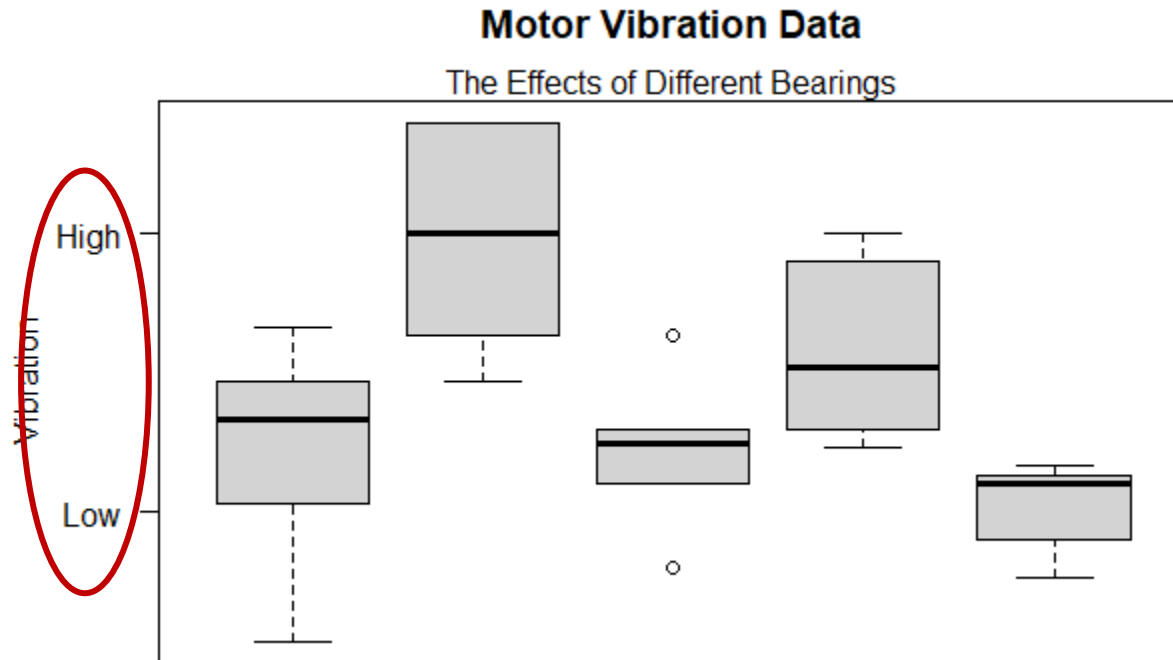
Y-AXIS TICKS

```
boxplot(motor, ylab="Vibration", axes=FALSE)  
title(main="Motor Vibration Data")  
mtext("The Effects of Different Bearings", side=3, line=0)  
box()  
axis(side=2, at=c(13, 16), label=c("Low", "High"))
```



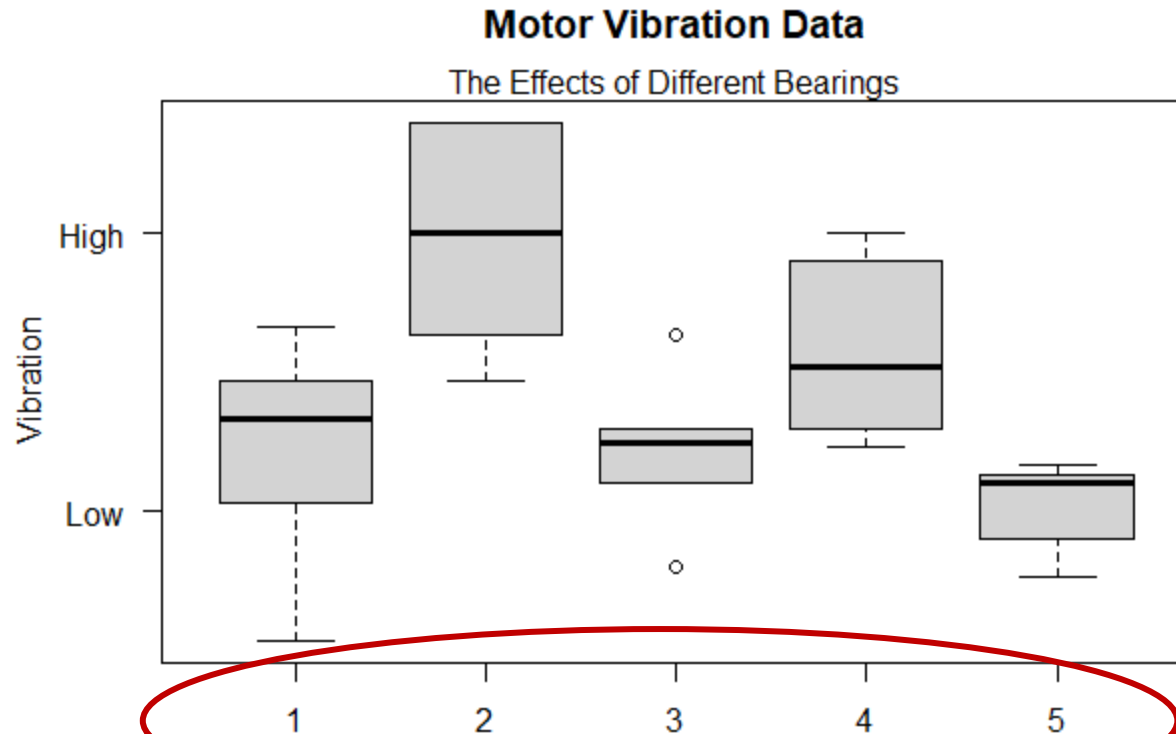
Y-AXIS TICKS

```
boxplot(motor, ylab="Vibration", axes=FALSE)  
title(main="Motor Vibration Data")  
mtext("The Effects of Different Bearings", side=3, line=0)  
box()  
axis(side=2, at=c(13, 16), label=c("Low", "High"), las=2)
```



X-AXIS TICKS

```
boxplot(motor, ylab="Vibration", axes=FALSE)  
title(main="Motor Vibration Data")  
mtext("The Effects of Different Bearings", side=3, line=0)  
box()  
axis(side=2, at=c(13, 16), label=c("Low", "High"), las=2)  
axis(side=1)
```



X-AXIS TICKS

```
boxplot(motor, ylab="Vibration", axes=FALSE)
title(main="Motor Vibration Data")
mtext("The Effects of Different Bearings", side=3, line=0)
box()
axis(side=2, at=c(13, 16), label=c("Low", "High"), las=2)
axis(side=1, at=1:5, label=names(motor))
```

