Breast Cancer Project Part 2
Question 6: PCA: proportion of variance

2.0/2.0 points (graded)
Perform a principal component analysis of the scaled matrix.

What proportion of variance is explained by the first principal component?

| 0.443 | ✔ **Answer:** 0.443 |

0.443

Explanation
The proportion of variance explained can be determined using the following code:

```
pca <- prcomp(x_scaled)
summary(pca)    # see PC1 Cumulative Proportion
```

How many principal components are required to explain at least 90% of the variance?

| 7 | ✔ **Answer:** 7 |

7

Explanation
The number of principal components can be determined using the following code:

```
pca <- prcomp(x_scaled)
summary(pca)    # first value of Cumulative Proportion that exceeds 0.9: PC7
```

| Submit | You have used 1 of 10 attempts |

ⓘ   Answers are displayed within the problem

Question 7: PCA: plotting PCs

1.0/1.0 point (graded)
Plot the first two principal components with color representing tumor type (benign/malignant).

Which of the following is true?

○ Malignant tumors tend to have smaller values of PC1 than benign tumors.

● Malignant tumors tend to have larger values of PC1 than benign tumors.

○ Malignant tumors tend to have smaller values of PC2 than benign tumors.

○ Malignant tumors tend to have larger values of PC2 than benign tumors.

○ There is no relationship between the first two principal components and tumor type.

✔

Explanation
The plot can be made using the following code:

```
data.frame(pca$x[,1:2], type = brca$y) %>%
  ggplot(aes(PC1, PC2, color = type)) +
  geom_point()
```

From the plot, you can see that the benign tumors tend to have smaller values of PC1 and that the malignant tumors have larger values of PC1. PC2 values have a similar spread for both benign and malignant tumors.

Submit   You have used 1 of 2 attempts

ⓘ   Answers are displayed within the problem

Question 8: PCA: PC boxplot

2.0/2.0 points (graded)
Make a boxplot of the first 10 PCs grouped by tumor type.

Which PCs are significantly different enough by tumor type that there is no overlap in the interquartile ranges (IQRs) for benign and malignant samples?
Select ALL that apply.

- ☑ PC1
- ☐ PC2
- ☐ PC3
- ☐ PC4
- ☐ PC5
- ☐ PC6
- ☐ PC7
- ☐ PC8
- ☐ PC9
- ☐ PC10

✔

Explanation

The boxplot can be generated using the following code:

```
data.frame(type = brca$y, pca$x[,1:10]) %>%
    gather(key = "PC", value = "value", -type) %>%
    ggplot(aes(PC, value, fill = type)) +
    geom_boxplot()
```

When you look at the boxplot, you can see that the IQRs overlap for PCs 2 through 10 but not for PC1.

Submit   You have used 1 of 3 attempts

ⓘ   Answers are displayed within the problem