Lab Worksheet: Project 6

Suppose we want to diagnose whether or not a given patient's tumor is malignant (cancerous) or benign (not cancerous). We have a group of 10 patients with known diagnoses, and the predictions of two arbitrary classification models: model A and model B. Their results are recorded below:

```
# A tibble: 10 × 3
   reality   model_a   model_b
   <chr>     <chr>     <chr>
 1 benign    benign    benign
 2 benign    benign    benign
 3 benign    benign    benign
 4 benign    benign    benign
 5 benign    malignant benign
 6 benign    benign    benign
 7 benign    benign    benign
 8 benign    malignant benign
 9 benign    malignant benign
10 malignant malignant benign
```

These models are being used by a hospital to help diagnose patients. You have been tasked with evaluating the performance of these two models.

1) Create functions that find the number of true positives, true negatives, false positives, and false negatives for each model. Then, use these functions to calculate these values for both models. (In this problem, we're treating "malignant" as the positive class).

```
library(tidyverse)

# your code here
```

2) Create functions that calculate the accuracy, precision, and recall for each model. These functions should be able to take in two vectors and return a single number (between 0 and 1). Use these functions to calculate these metrics for both models.

Hint:

$$\text{Accuracy} \quad = \quad \frac{TP + TN}{TP + TN + FP + FN} \quad = \quad \mathbb{P}[\text{correct prediction}],$$

$$\text{Precision} \quad = \quad \frac{TP}{TP + FP} \quad = \quad \mathbb{P}[(+ \text{ in reality}) \text{ given } (\text{test } +)]$$

$$\text{Recall} \quad = \quad \frac{TP}{TP + FN} \quad = \quad \mathbb{P}[(\text{test } +) \text{ given } (+ \text{ in reality})]$$

```
# your code here
```

3) Based on the metrics you calculated, which model would you recommend the hospital use? Why?