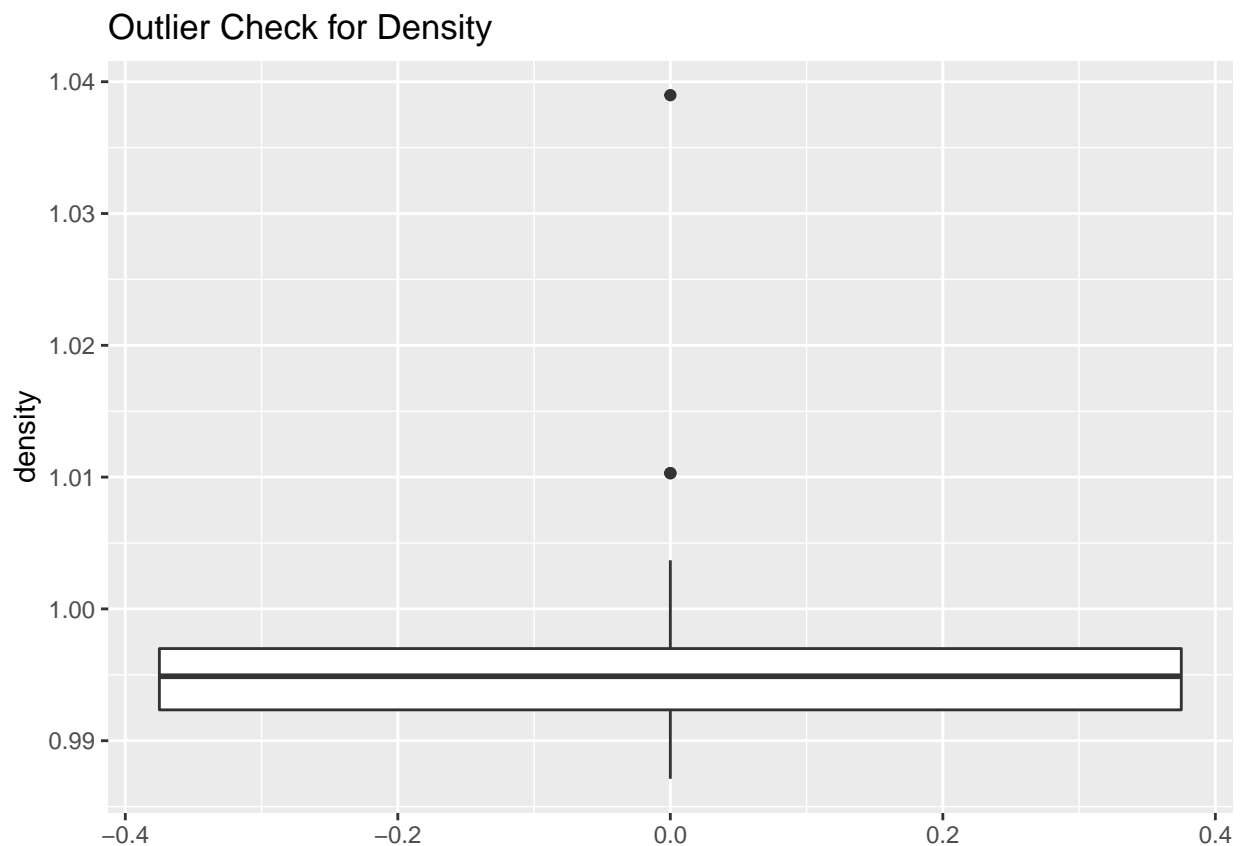


```
rm(list=ls())
library(ucidata)
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr  0.3.4
## v tibble  3.1.8      v dplyr  1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

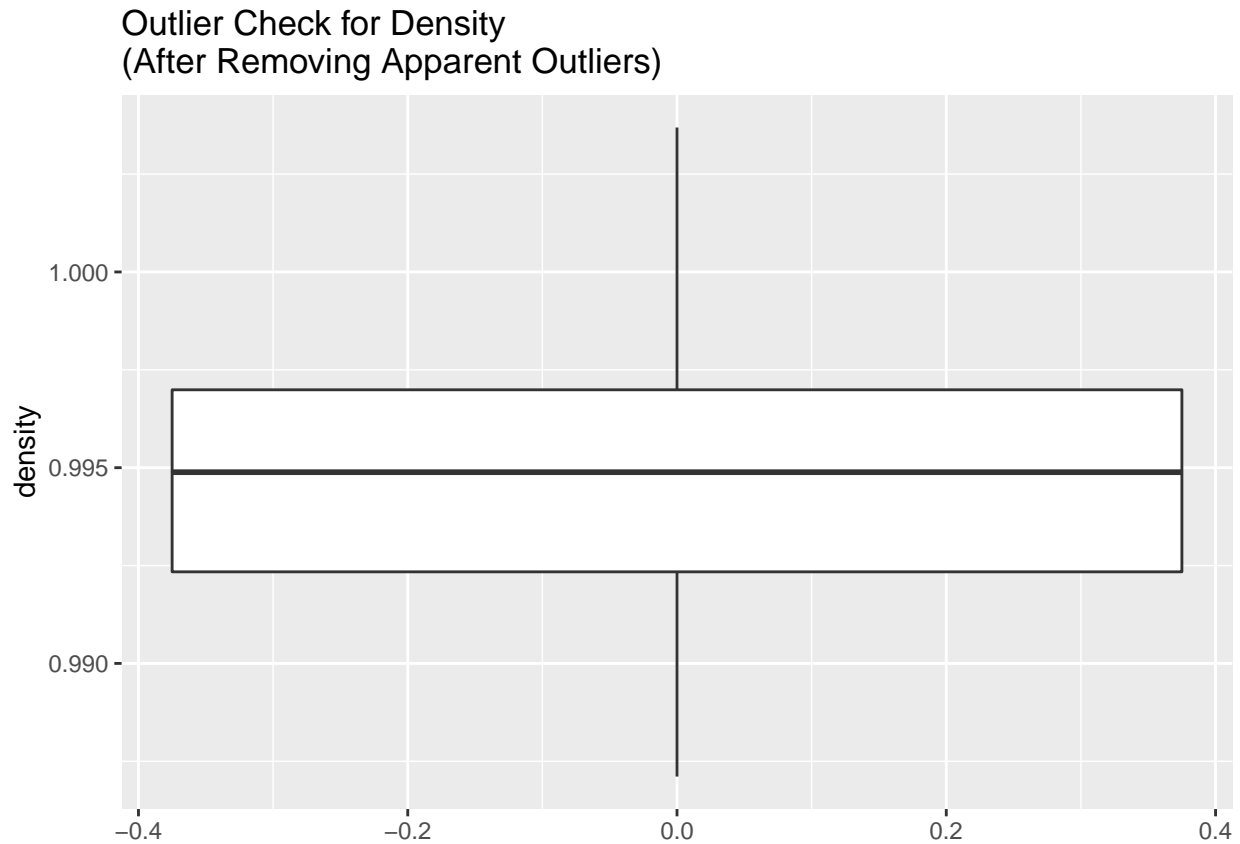
```
data(wine)
wine <- wine |>
  select(total_sulfur_dioxide, density, color)
```

```
## 1.
ggplot(data=wine, aes(y=density)) +
  geom_boxplot() +
  ggtitle("Outlier Check for Density")
```



```
wine <- wine |> filter(density < 1.01)

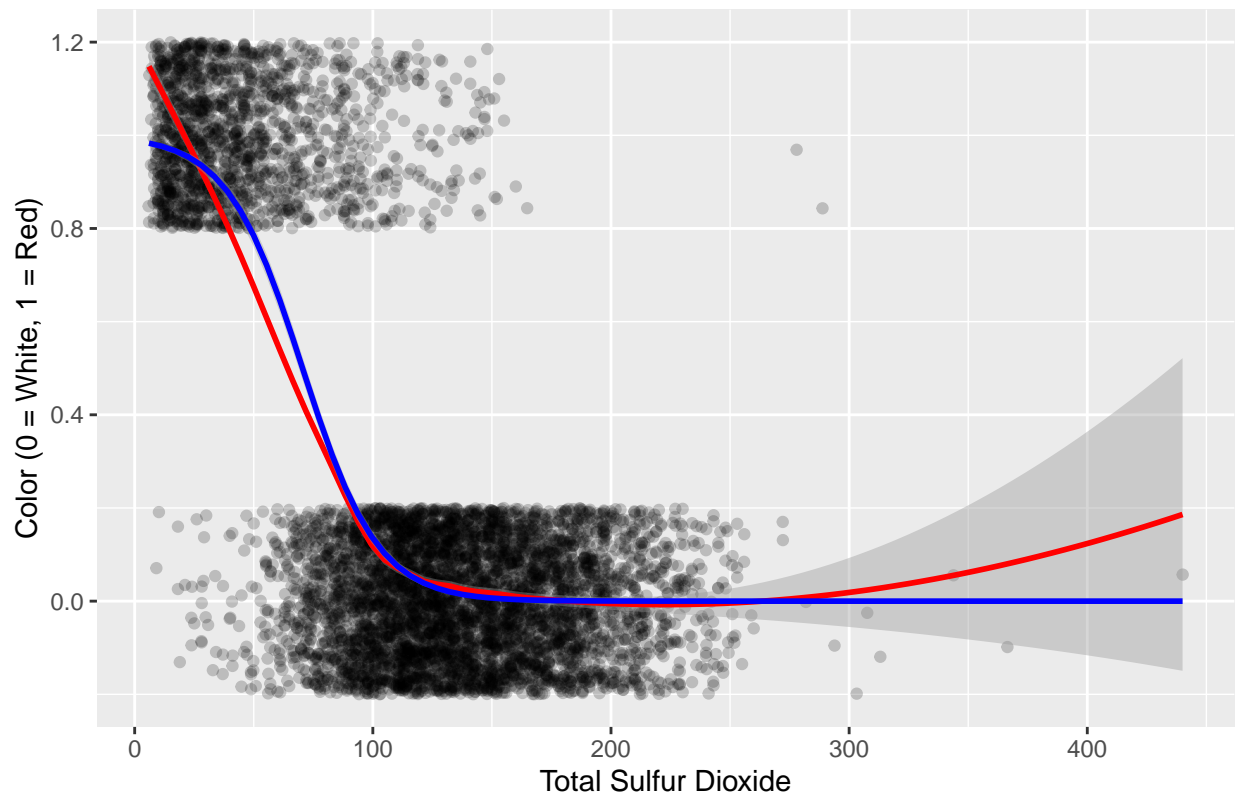
ggplot(data=wine, aes(y=density)) +
  geom_boxplot() +
  ggtitle("Outlier Check for Density\n(After Removing Apparent Outliers)")
```



```
## 2.
wine$color <- as.character(wine$color)
wine$color[wine$color == "White"] <- 0
wine$color[wine$color == "Red"] <- 1
wine$color <- factor(wine$color, levels=c(0,1))

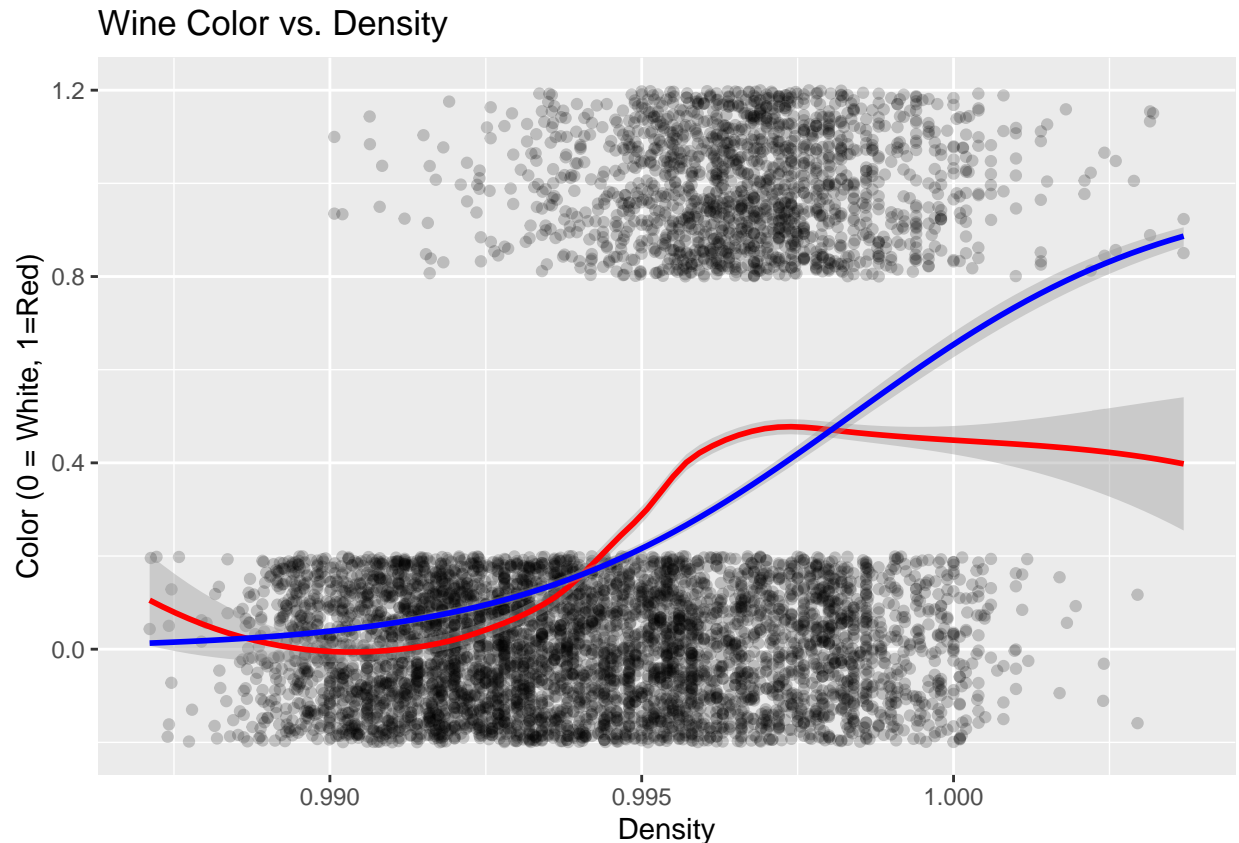
ggplot(data=wine, aes(x=total_sulfur_dioxide, y=(as.numeric(color) - 1))) +
  geom_jitter(height=0.2, alpha=0.2) +
  geom_smooth(method="loess", formula="y~x", color="red") +
  geom_smooth(method="glm", method.args=list(family="binomial"), formula="y~x", col="blue") +
  labs(title="Wine Color vs. Total Sulfur Dioxide",
       y="Color (0 = White, 1 = Red)",
       x="Total Sulfur Dioxide")
```

Wine Color vs. Total Sulfur Dioxide



Interpretation: Overall, the LOESS smoother (in red) and the logistic regression smoother (in blue) are similar. Both smoothers start off at values of red wine for lower values of total sulfur dioxide. Additionally, both smoothers drop sharply into values for white wine at higher values of total sulfur dioxide. However, the LOESS smoother drops sooner while the logistic smoother drops in more of an “S” shape. Finally, at the highest values of total sulfur dioxide, the LOESS smoother ticks up slightly to account for two outlier red wine points, while the logistic regression smoother does not.

```
## 3.
ggplot(data=wine, aes(x=density, y=(as.numeric(color) - 1))) +
  geom_jitter(height=0.2, alpha=0.2) +
  geom_smooth(method="loess", formula="y~x", color="red") +
  geom_smooth(method="glm", method.args=list(family="binomial"), formula="y~x", col="blue") +
  labs(title="Wine Color vs. Density",
       y="Color (0 = White, 1=Red)",
       x="Density")
```



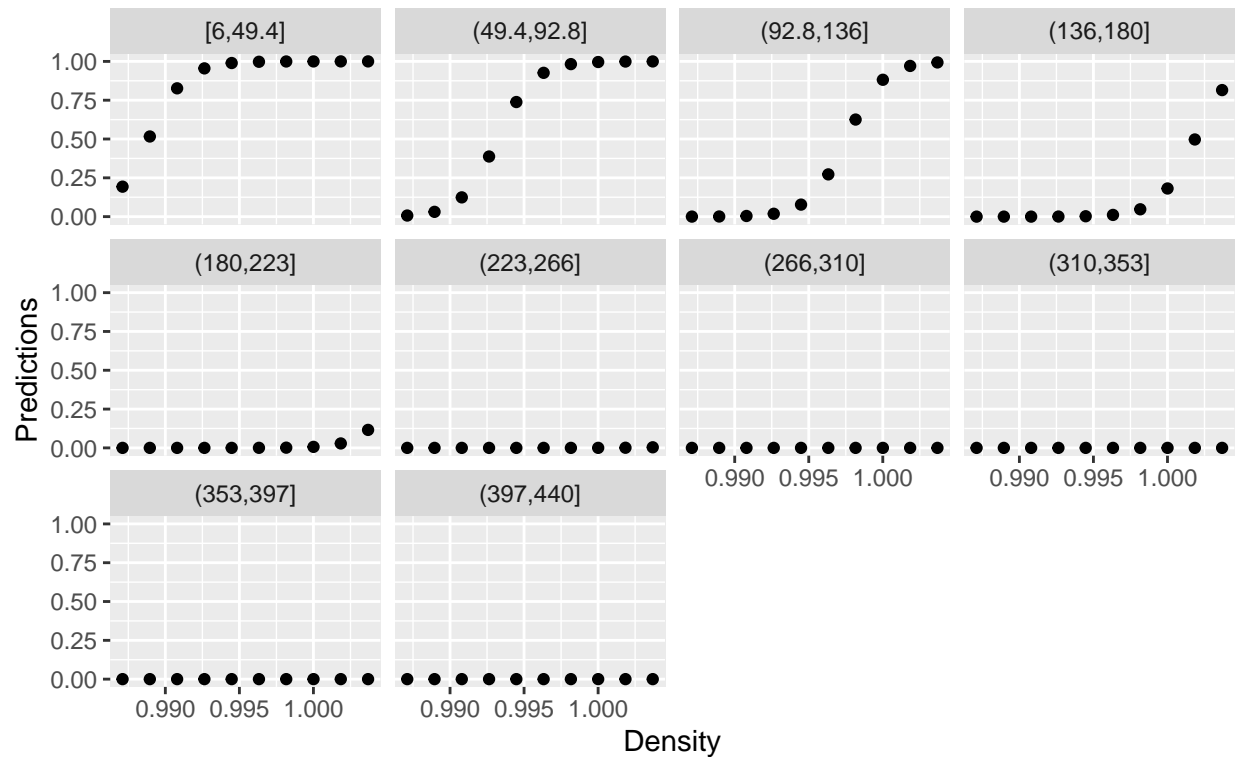
Interpretation: When plotting color vs. density, the two smoothers differ more drastically. Both the LOESS smoother (in red) and the logistic smoother (in blue) start at values for white wine at the lowest values of density. As density increases, the LOESS smoother first decreases, then increases, and finally stabilizes between the two factor levels. The logistic smoother, on the other hand, monotonically increases towards red wine as density increases. Ultimately, this plot demonstrates that LOESS is more flexible than logistic regression, which makes sense as LOESS is non-parametric while logistic regression is parametric.

4.

```
no_interaction_model <- glm(color ~ density + total_sulfur_dioxide, data=wine, family="binomial")
newdata <- expand_grid(density = seq(min(wine$density), max(wine$density), length.out = 10),
                      total_sulfur_dioxide = seq(min(wine$total_sulfur_dioxide), max(wine$total_sulfur_dioxide), length.out = 10))
predictions <- predict(no_interaction_model, newdata=newdata, type="response")
predictions <- data.frame(predictions, newdata)

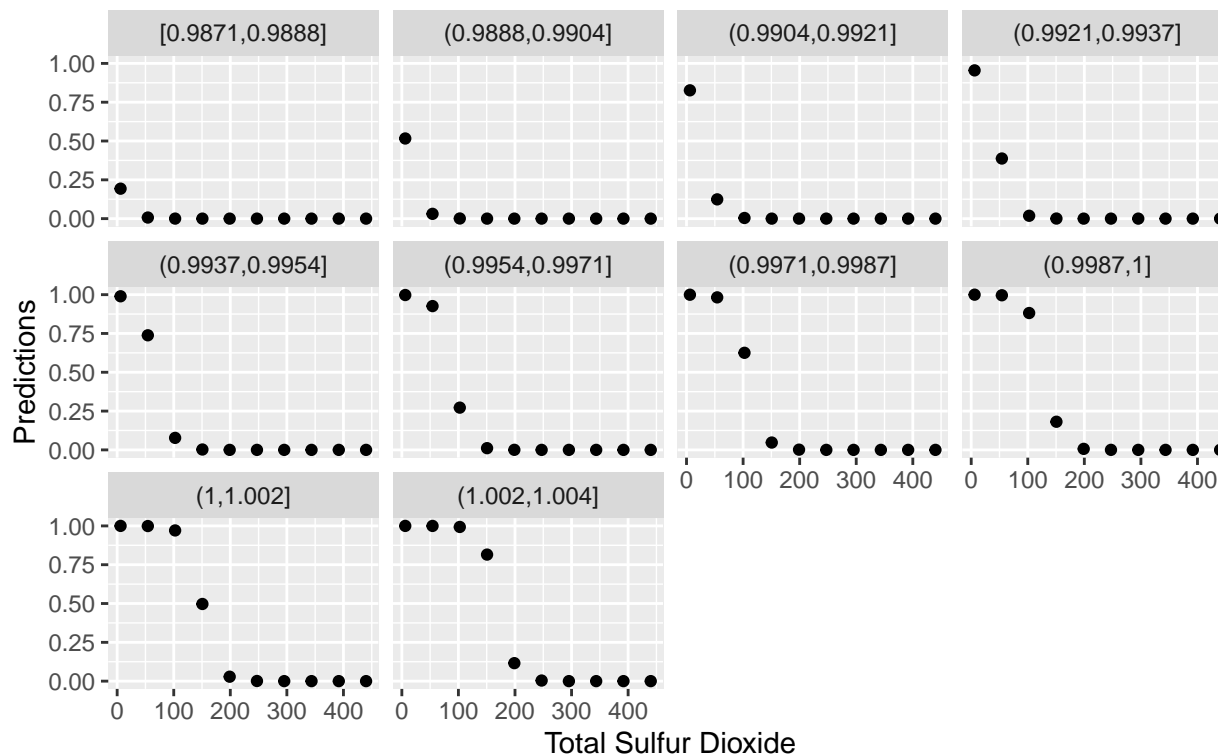
ggplot(data=predictions, aes(x=density, y=predictions)) +
  geom_point() +
  facet_wrap(vars(cut_number(total_sulfur_dioxide, 10))) +
  labs(title="No Interaction Model Predictions vs. Density\nFaceted by Total Sulfur Dioxide",
       y="Predictions",
       x="Density")
```

No Interaction Model Predictions vs. Density Faceted by Total Sulfur Dioxide



```
ggplot(data=predictions, aes(x=total_sulfur_dioxide, y=predictions)) +
  geom_point() +
  facet_wrap(vars(cut_number(density, 10))) +
  labs(title="No Interaction Model Predictions vs. Total Sulfur Dioxide\nFaceted by Density",
        y="Predictions",
        x="Total Sulfur Dioxide")
```

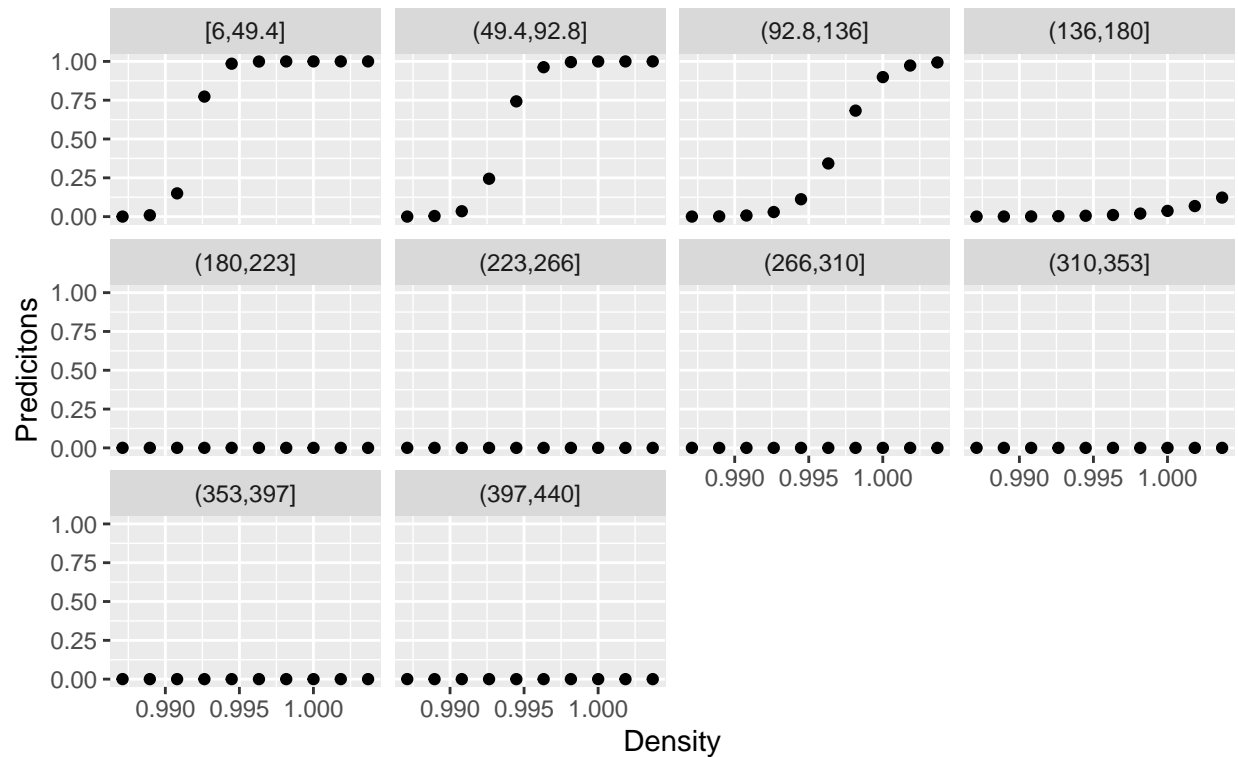
No Interaction Model Predictions vs. Total Sulfur Dioxide Faceted by Density



```
interaction_model <- glm(color ~ density*total_sulfur_dioxide, data=wine, family="binomial")
newdata <- expand_grid(density = seq(min(wine$density), max(wine$density), length.out=10),
                      total_sulfur_dioxide = seq(min(wine$total_sulfur_dioxide), max(wine$total_sulfur_dioxide), length.out=10))
predictions <- predict(interaction_model, newdata=newdata, type="response")
predictions <- data.frame(predictions, newdata)

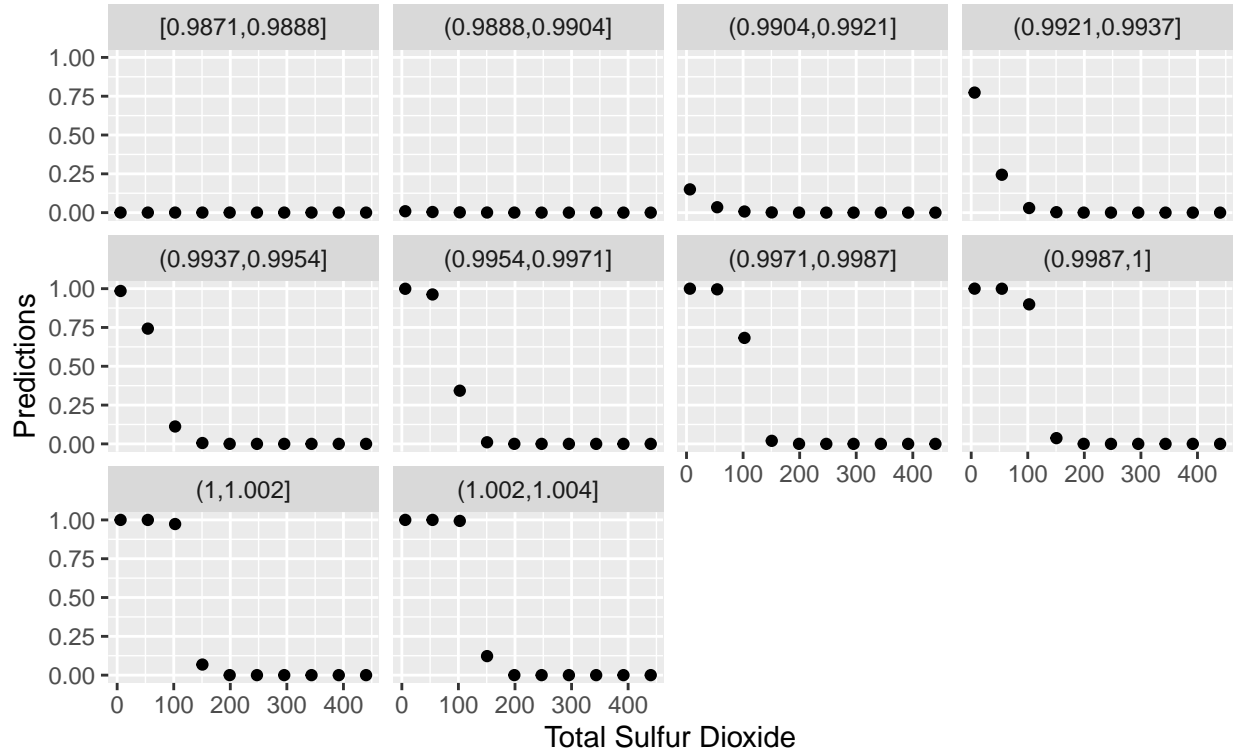
ggplot(data=predictions, aes(x=density, y=predictions)) +
  geom_point() +
  facet_wrap(vars(cut_number(total_sulfur_dioxide, 10))) +
  labs(title="Interaction Model Predictions vs. Density\nFaceted by Total Sulfur Dioxide",
       y="Predictitons",
       x="Density")
```

Interaction Model Predictions vs. Density Faceted by Total Sulfur Dioxide



```
ggplot(data=predictions, aes(x=total_sulfur_dioxide, y=predictions)) +
  geom_point() +
  facet_wrap(vars(cut_number(density, 10))) +
  labs(title="Interaction Model Predictions vs. Total Sulfur Dioxide\nFaceted by Density",
        y="Predictions",
        x="Total Sulfur Dioxide")
```

Interaction Model Predictions vs. Total Sulfur Dioxide Faceted by Density



5.

Interpretation: At low values of density, the models predict the wine type to be white. As density increases, the relationship with wine type is contingent on total sulfur dioxide. At low values of total sulfur dioxide, the predicted response is red wine as density increases. However, at moderate to high values of sulfur dioxide, the predicted response is always white wine regardless of the density level. At high values of sulfur dioxide, the predicted response is white wine, regardless of density level. However, at low levels of sulfur dioxide, the predicted response is more likely to be red wine as density level increases. Overall, there is not much difference between the interaction model and the no interaction model coplots. This is because coplots allow the visualization of interactive effects.