Quiz6 - MA478

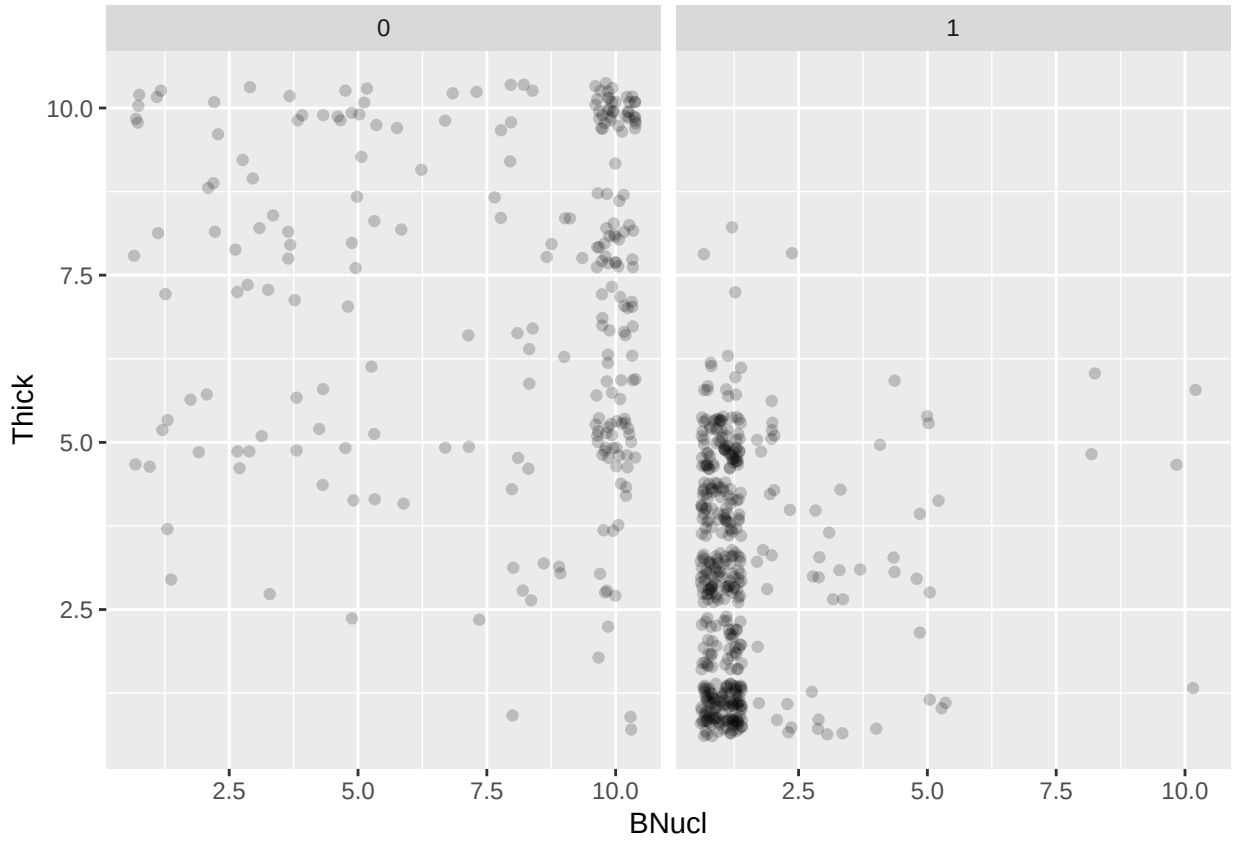
Clark

The dataset `wbca` comes from a study of breat cancer in Wisconsin. There are 681 cases of potentially cancerous tumors of which 238 are actually malignant. Determining whether a tumor is really malignant is traditionally determined by an invasive surgical procedure. The purpose of this study was to determine whether a new procedure called fine needle aspiration which draws only a small sample of tissue could be effective in determining tumor status.

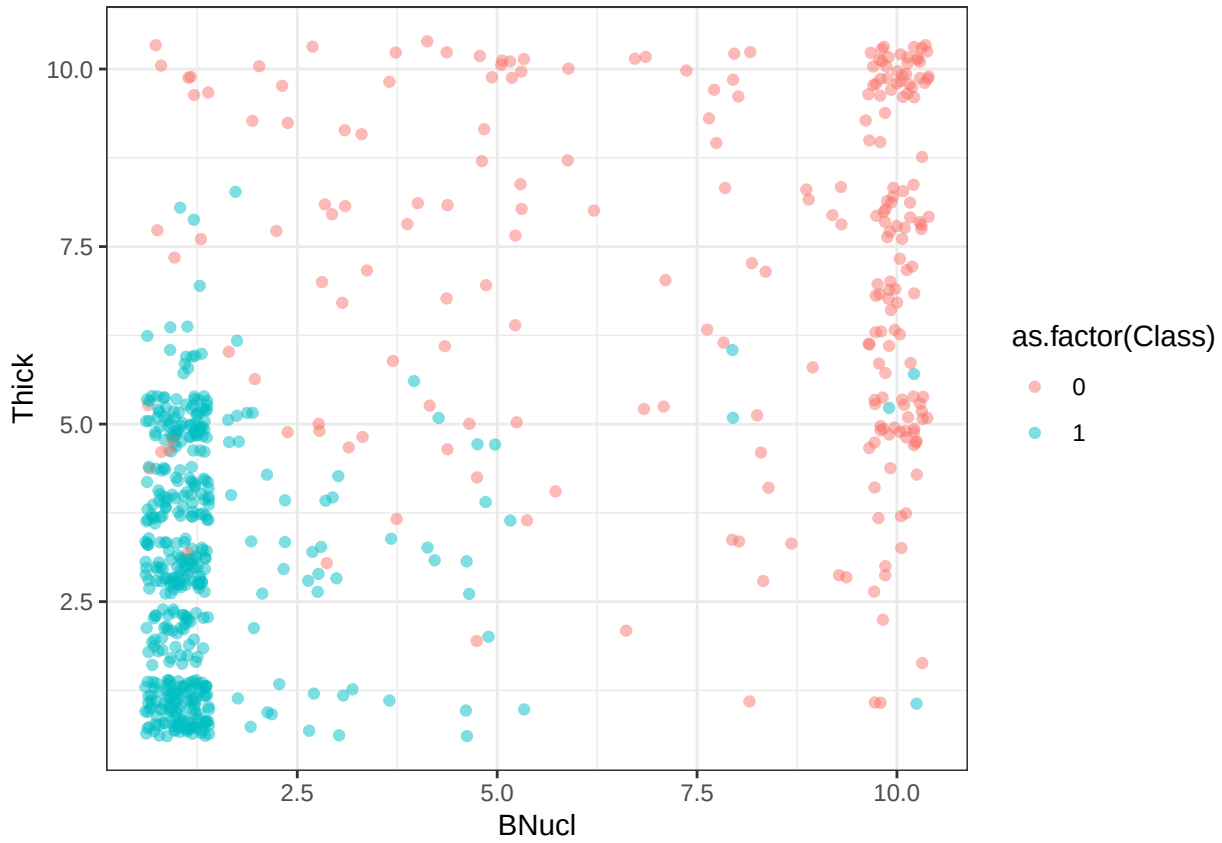
Produce a version of Figure 2.3 from Faraway for the predictors `BNucl` and `Thick`. Produce an alternative version with only one panel but where the two types are plotted differently. Compare the two plots and describe what they say about the ability to distinguish the two types using these two predictors.

```
library(faraway)
data("wbca")
```

```
library(faraway)
library(tidyverse)
data("wbca")
wbca %>% ggplot(aes(x=BNucl,y=Thick))+geom_point(alpha=0.2,position=
                                                   position_jitter())+
  facet_grid(~Class)
```




```
wbca %>% ggplot(aes(x=BNucl,y=Thick,color=as.factor(Class)))+geom_point(alpha=0.5,position=
                                                        position_jitter()) +
  theme_bw()
```



From both plots we can see that there clearly are differences between `BNucl` and `Thick` between those who have benign tumors and those who have malignant tumors.

Build a binary regression model with `Adhes`, `BNucl`, `Thick`, `Mitos` as predictors. Use an appropriate test to determine if, conditional on `Adhes`, `Thick`, and `Mitos`, if tumor presence is related to `BNucl`. Give the appropriate test, statistic, P-Value, and conclusion.

```
model_1 <- glm(Class~Adhes+BNucl+Thick+Mitos,data=wbca,family="binomial")
model_2 <- glm(Class~Adhes+Thick+Mitos,data=wbca,family="binomial")


anova(model_2,model_1,test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: Class ~ Adhes + Thick + Mitos
## Model 2: Class ~ Adhes + BNucl + Thick + Mitos
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1       677     228.19
## 2       676     132.82  1   95.364 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As the models are nested we can compute a χ^2 statistic from the difference in deviance. Here our statistic is 95.6 which has a χ^2_1 distribution under H_0. Our P value is close to zero suggesting the more complex model is appropriate, meaning `BNucl` is a meaningful covariate to add to our model.

Regardless of whether you found an effect due to `BNucl`, report the impact of change in odds of a patient having a malignant tumor comparing a patient with a normal `BNucl` to a patient with the most abnormal `BNucl` possible. (The help file for `wbca` may assist you with this). Note the coding of `Class` here. You may want to change this around to make your model more interpretable.

```
wbca_mod <- wbca %>% mutate(Malig=ifelse(Class==1,0,1))



model_3 <- glm(Malig ~ Adhes+BNucl+Thick+Mitos,data=wbca_mod,family="binomial")
coef(model_3)
```

```
## (Intercept)       Adhes       BNucl       Thick       Mitos
## -10.0956982   0.6337358   0.6723158   0.8886429   0.7685932
```

```
exp(9*0.6723158)
```

```
## [1] 424.4704
```

Here we see that the odds ratio between those with a Bnucl of 1 compared to those with a Bnucl of 10 is 424.5. Or, in other words the odds that you have cancer if you have a Bnucl of is 424.5 times the odds that you have cancer if you have a Bnucl of 1.