# Final Project

## Strickland, Marie

### BAN 502 - Predictive Analytics

library(readr)  
shark\_student <- read\_csv("shark\_student.csv")

## Warning: Missing column names filled in: 'X1' [1]

##   
## ── Column specification ────────────────────────────────────────────────────────  
## cols(  
## .default = col\_double(),  
## Company = col\_character(),  
## SeasonEpisode = col\_character(),  
## CompanyState = col\_character()  
## )  
## ℹ Use `spec()` for the full column specifications.

summary(shark\_student)

## X1 Company SeasonEpisode ReceiveOffer   
## Min. : 1.0 Length:551 Length:551 Min. :0.0000   
## 1st Qu.:138.5 Class :character Class :character 1st Qu.:0.0000   
## Median :276.0 Mode :character Mode :character Median :1.0000   
## Mean :276.0 Mean :0.6225   
## 3rd Qu.:413.5 3rd Qu.:1.0000   
## Max. :551.0 Max. :1.0000   
## RejectOffer Deal\_Yes Deal\_No Number of Presenters  
## Min. :0.00000 Min. :0.0000 Min. :0.0000 Min. :1.000   
## 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:1.000   
## Median :0.00000 Median :1.0000 Median :0.0000 Median :1.000   
## Mean :0.07078 Mean :0.5463 Mean :0.4537 Mean :1.539   
## 3rd Qu.:0.00000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:2.000   
## Max. :1.00000 Max. :1.0000 Max. :1.0000 Max. :4.000   
## Eth1 Eth2 Eth3 Eth4   
## Min. :0.000 Min. :0.000 Min. :0.0000 Min. :0.00000   
## 1st Qu.:2.000 1st Qu.:0.000 1st Qu.:0.0000 1st Qu.:0.00000   
## Median :2.000 Median :0.000 Median :0.0000 Median :0.00000   
## Mean :2.163 Mean :1.134 Mean :0.1143 Mean :0.00726   
## 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:0.0000 3rd Qu.:0.00000   
## Max. :4.000 Max. :4.000 Max. :4.0000 Max. :2.00000   
## Eth5 Male1 Male2 Male3   
## Min. :0.00000 Min. :0.0000 Min. :0.0000 Min. :0.00000   
## 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.00000   
## Median :0.00000 Median :1.0000 Median :0.0000 Median :0.00000   
## Mean :0.00363 Mean :0.7405 Mean :0.2613 Mean :0.03267   
## 3rd Qu.:0.00000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.00000   
## Max. :2.00000 Max. :1.0000 Max. :1.0000 Max. :1.00000   
## Male4 Female1 Female2 Female3   
## Min. :0 Min. :0.0000 Min. :0.0000 Min. :0.00000   
## 1st Qu.:0 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.00000   
## Median :0 Median :0.0000 Median :0.0000 Median :0.00000   
## Mean :0 Mean :0.2595 Mean :0.2214 Mean :0.01996   
## 3rd Qu.:0 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:0.00000   
## Max. :0 Max. :1.0000 Max. :1.0000 Max. :1.00000   
## Female4 Novelties Health / Wellness Food and Beverage  
## Min. :0.00000 Min. :0.00000 Min. :0.00000 Min. :0.0000   
## 1st Qu.:0.00000 1st Qu.:0.00000 1st Qu.:0.00000 1st Qu.:0.0000   
## Median :0.00000 Median :0.00000 Median :0.00000 Median :0.0000   
## Mean :0.00363 Mean :0.03811 Mean :0.02178 Mean :0.1742   
## 3rd Qu.:0.00000 3rd Qu.:0.00000 3rd Qu.:0.00000 3rd Qu.:0.0000   
## Max. :1.00000 Max. :1.00000 Max. :1.00000 Max. :1.0000   
## Business Services Lifestyle / Home Software / Tech Children / Education  
## Min. :0.00000 Min. :0.0000 Min. :0.00000 Min. :0.000   
## 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:0.000   
## Median :0.00000 Median :0.0000 Median :0.00000 Median :0.000   
## Mean :0.05445 Mean :0.1688 Mean :0.04356 Mean :0.098   
## 3rd Qu.:0.00000 3rd Qu.:0.0000 3rd Qu.:0.00000 3rd Qu.:0.000   
## Max. :1.00000 Max. :1.0000 Max. :1.00000 Max. :1.000   
## Automotive Fashion / Beauty Media / Entertainment  
## Min. :0.0000 Min. :0.0000 Min. :0.00000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.00000   
## Median :0.0000 Median :0.0000 Median :0.00000   
## Mean :0.0127 Mean :0.1724 Mean :0.06715   
## 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:0.00000   
## Max. :1.0000 Max. :1.0000 Max. :1.00000   
## Fitness / Sports / Outdoors Pet Products Travel   
## Min. :0.00000 Min. :0.00000 Min. :0.000000   
## 1st Qu.:0.00000 1st Qu.:0.00000 1st Qu.:0.000000   
## Median :0.00000 Median :0.00000 Median :0.000000   
## Mean :0.09982 Mean :0.03267 Mean :0.005445   
## 3rd Qu.:0.00000 3rd Qu.:0.00000 3rd Qu.:0.000000   
## Max. :1.00000 Max. :1.00000 Max. :1.000000   
## Green/CleanTech Uncertain / Other MalePresenter FemalePresenter   
## Min. :0.000000 Min. :0.000000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.000000 1st Qu.:0.000000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.000000 Median :0.000000 Median :1.0000 Median :0.0000   
## Mean :0.001815 Mean :0.009074 Mean :0.6062 Mean :0.2359   
## 3rd Qu.:0.000000 3rd Qu.:0.000000 3rd Qu.:1.0000 3rd Qu.:0.0000   
## Max. :1.000000 Max. :1.000000 Max. :1.0000 Max. :1.0000   
## MixedGenderPresenters CompanyState AmountRequested EquityRequested   
## Min. :0.0000 Length:551 Min. : 10000 Min. :0.0150   
## 1st Qu.:0.0000 Class :character 1st Qu.: 100000 1st Qu.:0.1000   
## Median :0.0000 Mode :character Median : 150000 Median :0.1500   
## Mean :0.1579 Mean : 274205 Mean :0.1576   
## 3rd Qu.:0.0000 3rd Qu.: 300000 3rd Qu.:0.2000   
## Max. :1.0000 Max. :5000000 Max. :1.0000   
## ImpliedValuationRequested BarbaraCorcoran MarkCuban LoriGreiner   
## Min. : 40000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.: 500000 1st Qu.:0.0000 1st Qu.:1.0000 1st Qu.:0.0000   
## Median : 1000000 Median :1.0000 Median :1.0000 Median :1.0000   
## Mean : 2829148 Mean :0.5971 Mean :0.8984 Mean :0.7187   
## 3rd Qu.: 2500000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :40000000 Max. :1.0000 Max. :1.0000 Max. :1.0000   
## RobertHerjavec DaymondJohn KevinOLeary KevinHarrington   
## Min. :0.000 Min. :0.0000 Min. :0.0000 Min. :0.00000   
## 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.:1.0000 1st Qu.:0.00000   
## Median :1.000 Median :1.0000 Median :1.0000 Median :0.00000   
## Mean :0.931 Mean :0.6534 Mean :0.9601 Mean :0.08711   
## 3rd Qu.:1.000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.00000   
## Max. :1.000 Max. :1.0000 Max. :1.0000 Max. :1.00000   
## Guest   
## Min. :0.0000   
## 1st Qu.:0.0000   
## Median :0.0000   
## Mean :0.1633   
## 3rd Qu.:0.0000   
## Max. :1.0000

#str(shark\_student)

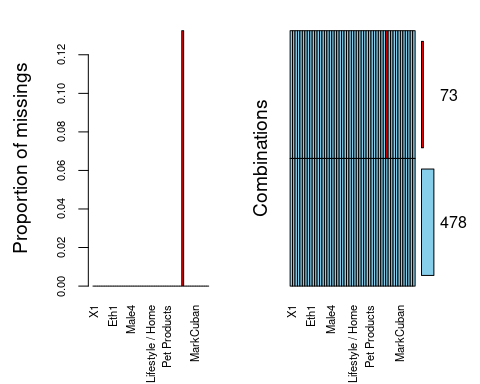
Based off this summary we see 3 variables we need to convert to factors. 62.25% of the companys received an offer in this dataset and 54.63% made a deal with the investor. Green/CleanTech is the smallest represented product category, and Food/Beverage is highest with 17.42%, and Fashion/Beauty following close behind representing 17.24%. The average funding requested by the companys is $274,205 with average equity being 15.76%.

shark\_student <- shark\_student %>%  
 mutate\_if(is.character, as\_factor) %>%  
 mutate(ReceiveOffer = factor(ReceiveOffer), RejectOffer = factor(RejectOffer), Deal\_Yes = factor(Deal\_Yes), Deal\_No = factor(Deal\_No), Male1 = factor(Male1), Male2 = factor(Male2), Male3 = factor(Male3), Male4 = factor(Male4), Female1 = factor(Female1), Female2 = factor(Female2), Female3 = factor(Female3), Female4 = factor(Female4), Novelties = factor(Novelties), `Health / Wellness` = factor(`Health / Wellness`), `Food and Beverage` = factor(`Food and Beverage`), `Business Services` = factor(`Business Services`), `Lifestyle / Home` = factor(`Lifestyle / Home`), `Software / Tech` = factor(`Software / Tech`), `Children / Education` = factor(`Children / Education`), Automotive = factor(Automotive), `Fashion / Beauty` = factor(`Fashion / Beauty`), `Media / Entertainment` = factor(`Media / Entertainment`), `Fitness / Sports / Outdoors` = factor(`Fitness / Sports / Outdoors`), `Pet Products` = factor(`Pet Products`), Travel = factor(Travel), `Green/CleanTech` = factor(`Green/CleanTech`), `Uncertain / Other` = factor(`Uncertain / Other`), MalePresenter = factor(MalePresenter), FemalePresenter = factor(FemalePresenter), MixedGenderPresenters = factor(MixedGenderPresenters), BarbaraCorcoran = factor(BarbaraCorcoran), MarkCuban = factor(MarkCuban), LoriGreiner = factor(LoriGreiner), RobertHerjavec = factor(RobertHerjavec), DaymondJohn = factor(DaymondJohn), KevinHarrington = factor(KevinHarrington), KevinOLeary = factor(KevinOLeary), Guest = factor(Guest), Eth1 = factor(Eth1), Eth2 = factor(Eth2), Eth3 = factor(Eth3), Eth4 = factor(Eth4), Eth5 = factor(Eth5))  
#str(shark\_student)

shark\_student <- shark\_student %>%  
 mutate(  
 Eth1 = fct\_recode(Eth1,  
 'No Presenter' = '0',  
 'Black' = '1',   
 'White' = '2',   
 'Asian' = '3',   
 'Latino' = '4'),  
 Eth2 = fct\_recode(Eth2,  
 'No Presenter' = '0',  
 'Black' = '1',   
 'White' = '2',   
 'Asian' = '3',   
 'Latino' = '4'),  
 Eth3 = fct\_recode(Eth3,  
 'No Presenter' = '0',  
 'Black' = '1',   
 'White' = '2',   
 'Asian' = '3',   
 'Latino' = '4'),   
 Eth4 = fct\_recode(Eth4,  
 'No Presenter' = '0',  
 'Black' = '1',   
 'White' = '2',   
 'Asian' = '3',   
 'Latino' = '4'),  
 Eth5 = fct\_recode(Eth5,  
 'No Presenter' = '0',  
 'Black' = '1',   
 'White' = '2',   
 'Asian' = '3',   
 'Latino' = '4')  
 )

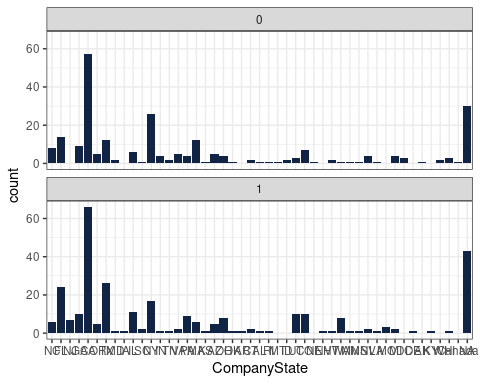
## Warning: Unknown levels in `f`: 1, 3, 4  
  
## Warning: Unknown levels in `f`: 1, 3, 4

aggr(shark\_student, numbers = TRUE, prop = c(TRUE, FALSE),cex.axis=.7)



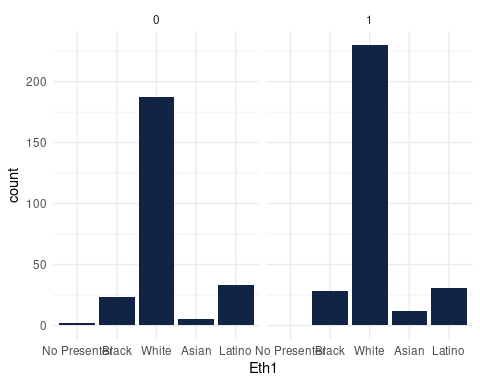
73 rows contain missing data in the CompanyState column. Is this data necessary to keep or can we drop all rows with missing data?

ggplot(shark\_student) +  
 aes(x = CompanyState) +  
 geom\_bar(fill = "#112446") +  
 theme\_bw() +  
 facet\_wrap(vars(Deal\_Yes),   
 nrow = 10L)



Based on this plot, state does not seem to be a huge predictor if the company receives an offer or not. We can tell this because the plots are very similar between receiving and not receiving an offer. I will choose to keep these rows as they contain other information more valuable than missing state.

ggplot(shark\_student) +  
 aes(x = Eth1) +  
 geom\_bar(fill = "#112446") +  
 theme\_minimal() +  
 facet\_wrap(vars(Deal\_Yes))

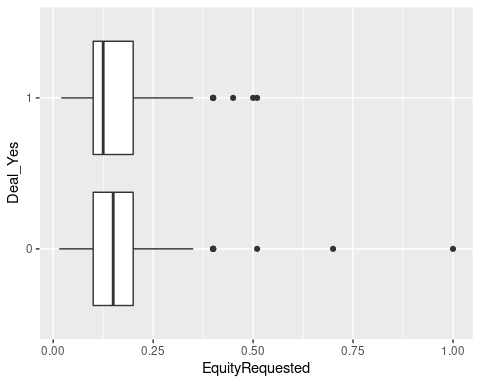


Eth1.Asian may have a little impact on Deal\_Yes

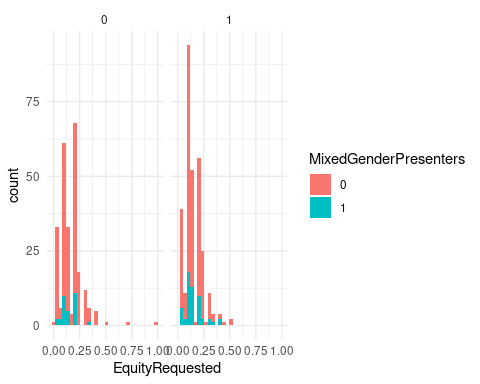
#ggplot(shark\_student, aes(x = AmountRequested, y = Deal\_Yes)) +  
# geom\_boxplot()

AmountRequested does not seem to be an important variable to predict Deal\_Yes because the boxplots are VERY similar, except for the outlier for level ‘0’ of Deal\_Yes.

ggplot(shark\_student, aes(x = EquityRequested, y = Deal\_Yes)) +  
 geom\_boxplot()

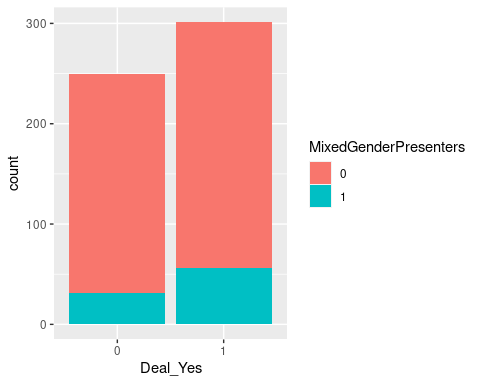


ggplot(shark\_student) +  
 aes(x = EquityRequested, fill = MixedGenderPresenters) +  
 geom\_histogram(bins = 30L) +  
 scale\_fill\_hue(direction = 1) +  
 theme\_minimal() +  
 facet\_wrap(vars(Deal\_Yes))

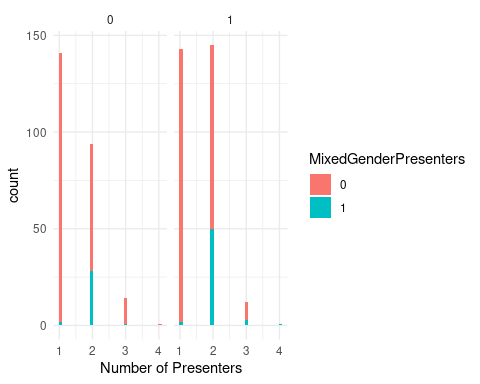


EquityRequested seems to have potential to be a predictor, as the mean is lower for those who made a deal despite Q1 and Q3 being similar between the levels. The range is smaller for those who made a deal as well.

ggplot(shark\_student) +  
 aes(x = Deal\_Yes, fill = MixedGenderPresenters) +  
 geom\_bar() +  
 scale\_fill\_hue(direction = 1) +  
 theme\_gray()



ggplot(shark\_student) +  
 aes(  
 x = `Number of Presenters`,  
 fill = MixedGenderPresenters  
 ) +  
 geom\_histogram(bins = 30L) +  
 scale\_fill\_hue(direction = 1) +  
 theme\_minimal() +  
 facet\_wrap(vars(Deal\_Yes))



It seems when mixed genders are present when number of presenters are 2+ is a relatively strong predictor of Deal\_Yes

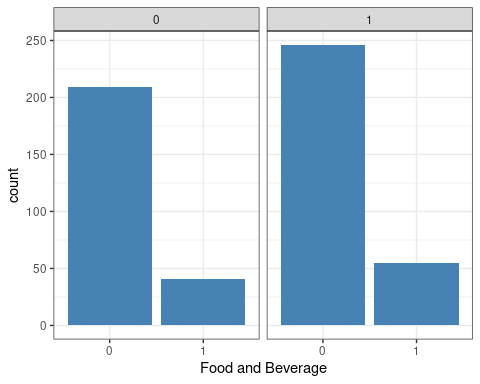
#ggplot(shark\_student, aes(x = ImpliedValuationRequested, y = Deal\_Yes)) +  
# geom\_boxplot()

ImpliedValuationRequested does not appear to be a great predictor either. The means are very close, and ranges are closer together as well. The Q1 and Q3 vary, but not enough to see huge importance in this variable

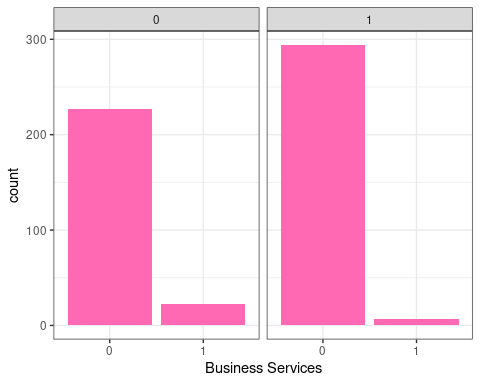
#mosaicplot(Deal\_Yes~Novelties, shark\_student)

#mosaicplot(Deal\_Yes~shark\_student$`Health / Wellness`, shark\_student)

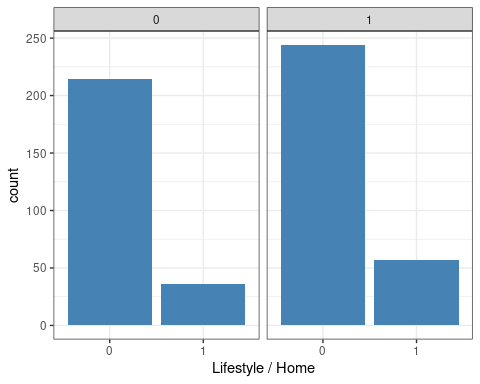
ggplot(shark\_student) +  
 aes(x = `Food and Beverage`) +  
 geom\_bar(fill = "#4682B4") +  
 theme\_bw() +  
 facet\_wrap(vars(Deal\_Yes))



ggplot(shark\_student) +  
 aes(x = `Business Services`) +  
 geom\_bar(fill = "#FF69B4") +  
 theme\_bw() +  
 facet\_wrap(vars(Deal\_Yes))

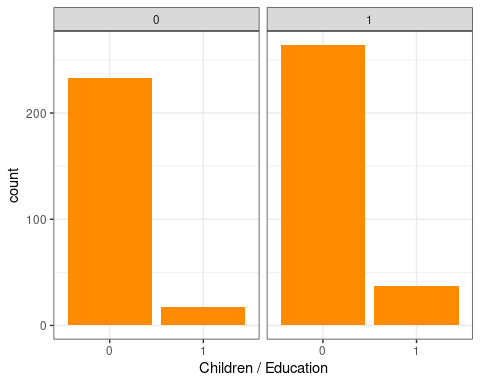


ggplot(shark\_student) +  
 aes(x = `Lifestyle / Home`) +  
 geom\_bar(fill = "#4682B4") +  
 theme\_bw() +  
 facet\_wrap(vars(Deal\_Yes))

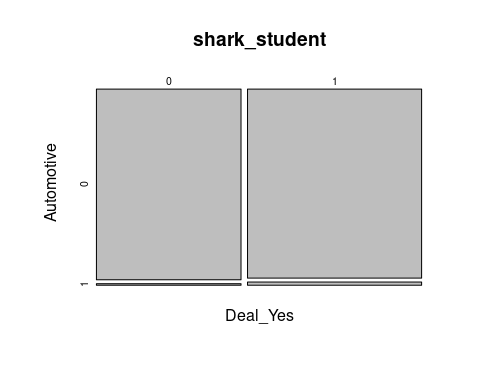


#mosaicplot(Deal\_Yes~shark\_student$`Software / Tech`, shark\_student)

ggplot(shark\_student) +  
 aes(x = `Children / Education`) +  
 geom\_bar(fill = "#FF8C00") +  
 theme\_bw() +  
 facet\_wrap(vars(Deal\_Yes))



mosaicplot(Deal\_Yes~Automotive, shark\_student)



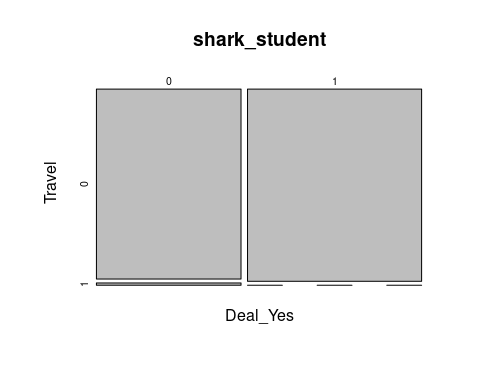
#mosaicplot(Deal\_Yes~shark\_student$`Fashion / Beauty`, shark\_student)

#mosaicplot(Deal\_Yes~shark\_student$`Media / Entertainment`, shark\_student)

#mosaicplot(Deal\_Yes~shark\_student$`Fitness / Sports / Outdoors`, shark\_student)

#mosaicplot(Deal\_Yes~shark\_student$`Pet Products`, shark\_student)

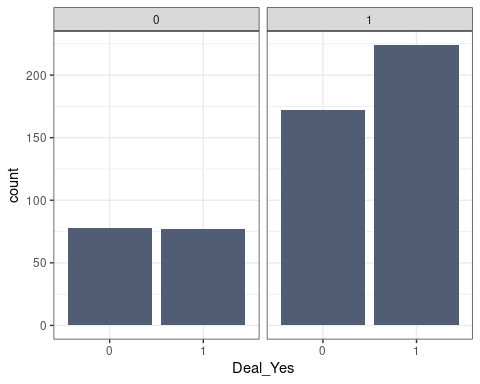
mosaicplot(Deal\_Yes~Travel, shark\_student)



#mosaicplot(Deal\_Yes~shark\_student$`Green/CleanTech`, shark\_student)

Out of the categories of product type, Food and Beverage, Business Service, Lifestyle / Home, Children / Education, Automotive, and Travel seem to be possible predictors for Deal\_Yes. Not surprised Green/CleanTech is not a strong influencer because there is only one product under this category.

ggplot(shark\_student) +  
 aes(x = Deal\_Yes) +  
 geom\_bar(fill = "#505D75") +  
 theme\_bw() +  
 facet\_wrap(vars(LoriGreiner))



Lori Greiner has an impact on Deals if she is there. It seems when she is there more people accept their offers.