## BAN502 Course Project Part 1

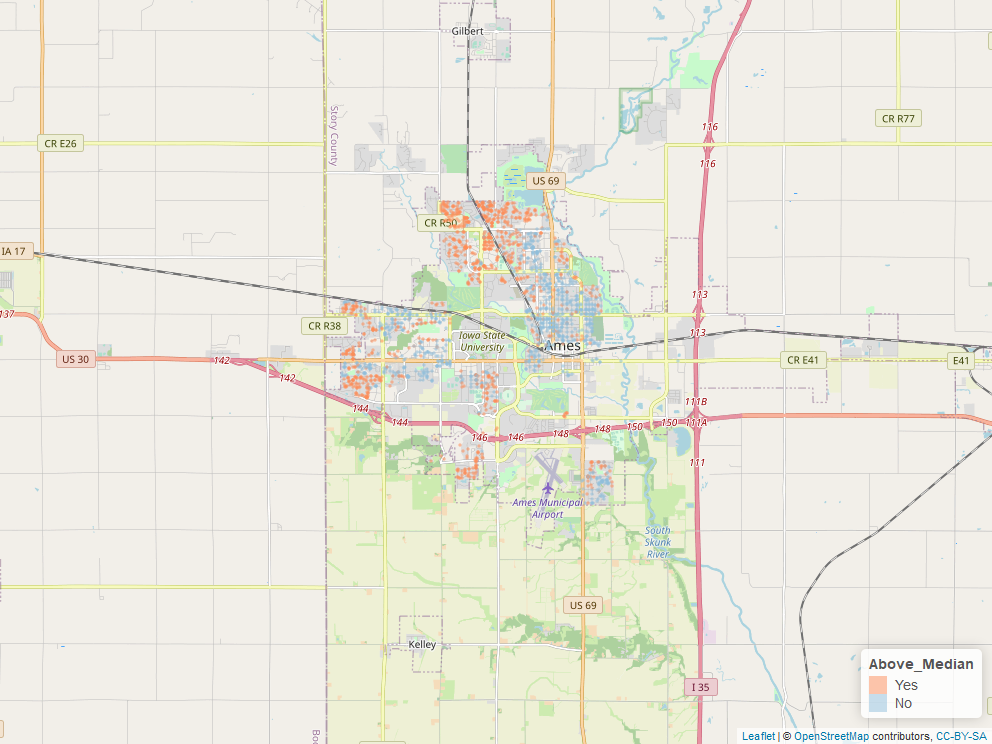
### Kenny Buckenmeyer

ames\_student <- ames\_student %>% mutate\_if(is.character, as\_factor)

colorpal <- colorFactor(palette = "RdYlBu", ames\_student$Above\_Median)

ames\_map <- leaflet() %>% setView(lng = -93.64, lat = 42.02, zoom = 12) %>%  
 addTiles() %>%  
 addCircleMarkers(data = ames\_student, lng = ~Longitude, lat = ~Latitude, weight = 1.5, radius = 1, color = ~colorpal(Above\_Median)) %>%  
 addLegend(data = ames\_student, position = "bottomright", pal = colorpal, values = ~Above\_Median)  
mapshot(ames\_map, file ="~/ames.png")

knitr::include\_graphics("ames.png")



This map provides a visualization comparing the houses that sold above the median price and the houses that sold below the median price. There is noticeable difference among the houses north of Iowa State University. The houses west and northwest of Parkview Heights typically sold above the median price, while the houses east and southeast of Parkview Heights typically sold below the median price.

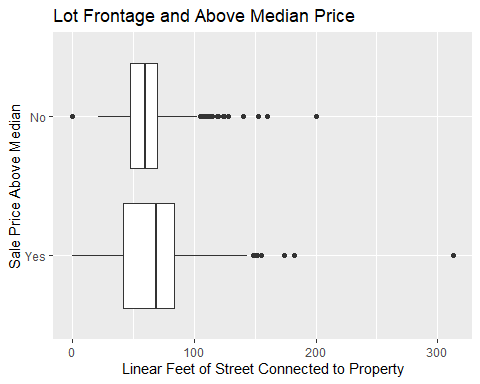
ames\_student$X1 <- NULL  
ames\_student <- ames\_student %>% relocate(Above\_Median, .before = MS\_SubClass)  
head(ames\_student)

## # A tibble: 6 x 81  
## Above\_Median MS\_SubClass MS\_Zoning Lot\_Frontage Lot\_Area Street Alley  
## <fct> <fct> <fct> <dbl> <dbl> <fct> <fct>  
## 1 Yes One\_Story\_~ Resident~ 141 31770 Pave No\_A~  
## 2 No One\_Story\_~ Resident~ 80 11622 Pave No\_A~  
## 3 Yes One\_Story\_~ Resident~ 81 14267 Pave No\_A~  
## 4 Yes One\_Story\_~ Resident~ 93 11160 Pave No\_A~  
## 5 Yes Two\_Story\_~ Resident~ 74 13830 Pave No\_A~  
## 6 Yes Two\_Story\_~ Resident~ 78 9978 Pave No\_A~  
## # ... with 74 more variables: Lot\_Shape <fct>, Land\_Contour <fct>,  
## # Utilities <fct>, Lot\_Config <fct>, Land\_Slope <fct>, Neighborhood <fct>,  
## # Condition\_1 <fct>, Condition\_2 <fct>, Bldg\_Type <fct>, House\_Style <fct>,  
## # Overall\_Qual <fct>, Overall\_Cond <fct>, Year\_Built <dbl>,  
## # Year\_Remod\_Add <dbl>, Roof\_Style <fct>, Roof\_Matl <fct>,  
## # Exterior\_1st <fct>, Exterior\_2nd <fct>, Mas\_Vnr\_Type <fct>,  
## # Mas\_Vnr\_Area <dbl>, Exter\_Qual <fct>, Exter\_Cond <fct>, Foundation <fct>,  
## # Bsmt\_Qual <fct>, Bsmt\_Cond <fct>, Bsmt\_Exposure <fct>,  
## # BsmtFin\_Type\_1 <fct>, BsmtFin\_SF\_1 <dbl>, BsmtFin\_Type\_2 <fct>,  
## # BsmtFin\_SF\_2 <dbl>, Bsmt\_Unf\_SF <dbl>, Total\_Bsmt\_SF <dbl>, Heating <fct>,  
## # Heating\_QC <fct>, Central\_Air <fct>, Electrical <fct>, First\_Flr\_SF <dbl>,  
## # Second\_Flr\_SF <dbl>, Low\_Qual\_Fin\_SF <dbl>, Gr\_Liv\_Area <dbl>,  
## # Bsmt\_Full\_Bath <dbl>, Bsmt\_Half\_Bath <dbl>, Full\_Bath <dbl>,  
## # Half\_Bath <dbl>, Bedroom\_AbvGr <dbl>, Kitchen\_AbvGr <dbl>,  
## # Kitchen\_Qual <fct>, TotRms\_AbvGrd <dbl>, Functional <fct>,  
## # Fireplaces <dbl>, Fireplace\_Qu <fct>, Garage\_Type <fct>,  
## # Garage\_Finish <fct>, Garage\_Cars <dbl>, Garage\_Area <dbl>,  
## # Garage\_Qual <fct>, Garage\_Cond <fct>, Paved\_Drive <fct>,  
## # Wood\_Deck\_SF <dbl>, Open\_Porch\_SF <dbl>, Enclosed\_Porch <dbl>,  
## # Three\_season\_porch <dbl>, Screen\_Porch <dbl>, Pool\_Area <dbl>,  
## # Pool\_QC <fct>, Fence <fct>, Misc\_Feature <fct>, Misc\_Val <dbl>,  
## # Mo\_Sold <dbl>, Year\_Sold <dbl>, Sale\_Type <fct>, Sale\_Condition <fct>,  
## # Longitude <dbl>, Latitude <dbl>

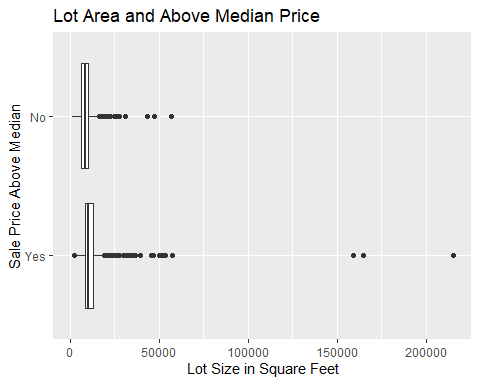
ames\_student %>% count(Above\_Median)

## # A tibble: 2 x 2  
## Above\_Median n  
## <fct> <int>  
## 1 Yes 1043  
## 2 No 1010

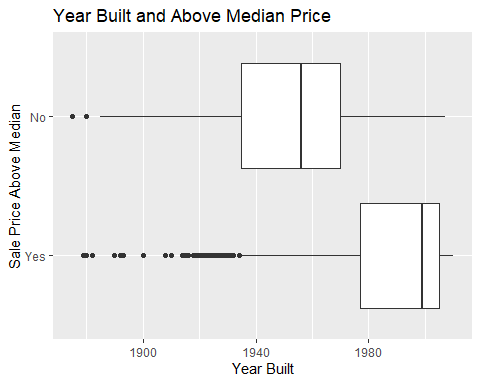
ggplot(ames\_student, aes(x = Lot\_Frontage, y = Above\_Median)) +  
 geom\_boxplot() +  
 labs(title = "Lot Frontage and Above Median Price",  
 x = "Linear Feet of Street Connected to Property",  
 y = "Sale Price Above Median")



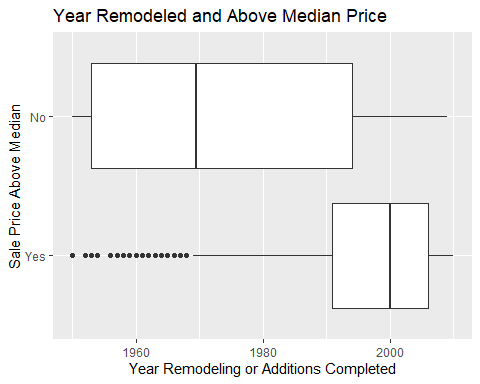
ggplot(ames\_student, aes(x = Lot\_Area, y = Above\_Median)) +  
 geom\_boxplot() +  
 labs(title = "Lot Area and Above Median Price",  
 x = "Lot Size in Square Feet",  
 y = "Sale Price Above Median")



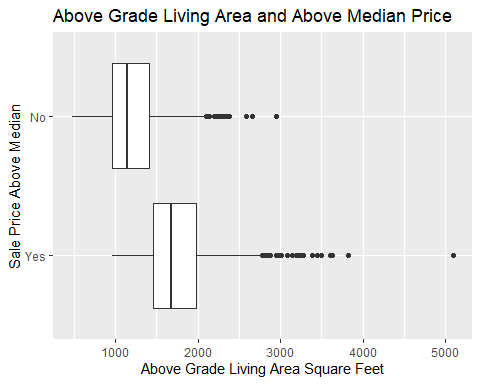
ggplot(ames\_student, aes(x = Year\_Built, y = Above\_Median)) +  
 geom\_boxplot() +  
 labs(title = "Year Built and Above Median Price",  
 x = "Year Built",  
 y = "Sale Price Above Median")



ggplot(ames\_student, aes(x = Year\_Remod\_Add, y = Above\_Median)) +  
 geom\_boxplot() +  
 labs(title = "Year Remodeled and Above Median Price",   
 x= "Year Remodeling or Additions Completed",  
 y = "Sale Price Above Median")

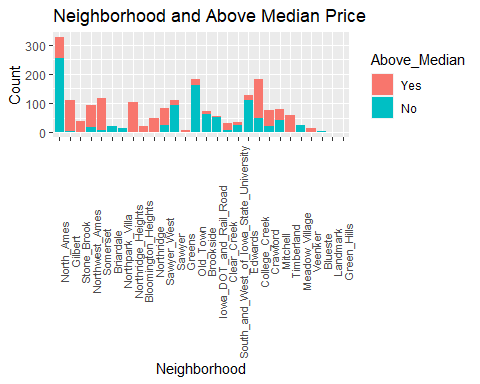


ggplot(ames\_student, aes(x = Gr\_Liv\_Area, y = Above\_Median)) +  
 geom\_boxplot() +  
 labs(title = "Above Grade Living Area and Above Median Price",  
 x = "Above Grade Living Area Square Feet",  
 y = "Sale Price Above Median")



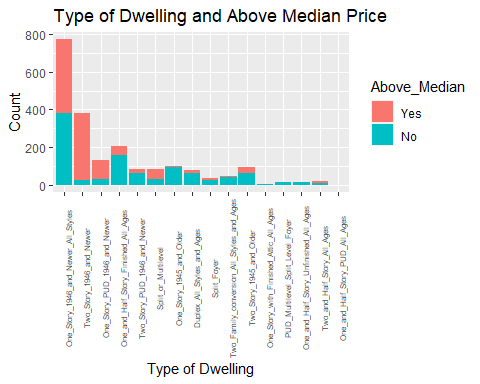
The above plots compared five numeric variables (Lot\_Frontage, Lot\_Area, Year\_Built, Year\_Remod\_Add, Gr\_Liv\_Area). There isn’t a significant difference between the houses that sold above and below the median price when looking at the total linear square feet of street connected to the property. Surprisingly, there isn’t a significant difference between the houses that sold above and below the median price when looking at total lot size. There is a significant difference when looking at the year the house was built. While the houses that sold above the median price have more variance, the mean of these houses is significantly higher than the houses that did not sell above the median price. Houses with more recent remodeling or additions sold above the median price more often than houses with older remodeling or additions. Not surprisingly, houses with more above ground living area sold above the median price.

ggplot(ames\_student, aes(x = Neighborhood, fill = Above\_Median)) +  
 geom\_bar() + theme(axis.text.x = element\_text(size = 8, angle = 90)) +  
 labs(title = "Neighborhood and Above Median Price",  
 y = "Count")



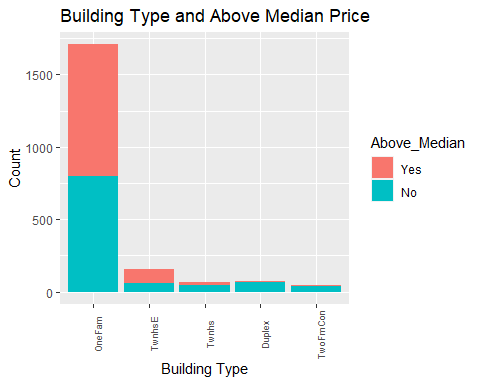
The above visualization compares whether or not houses sold above the median price and the neighborhoods in which the houses were located. A few neighborhoods stand out that have houses selling above the median price, including Gilbert, Stone Brook, Northwest Ames, Somerset, Northridge Heights, Northridge, Sawyer West, College Creek, Crawford, and Timberland.

ggplot(ames\_student, aes(x = MS\_SubClass, fill = Above\_Median)) +  
 geom\_bar() + theme(axis.text.x = element\_text(size = 6, angle = 90)) +  
 labs(title = "Type of Dwelling and Above Median Price",  
 x = "Type of Dwelling",  
 y = "Count")



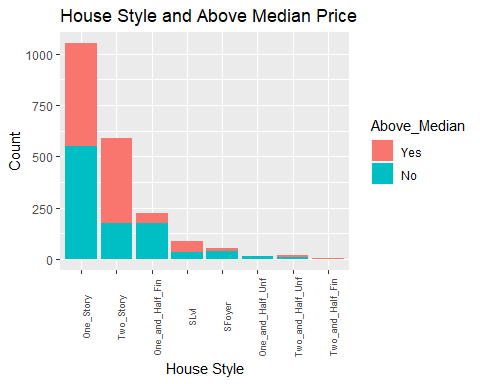
The above visualization compares whether or not houses sold above the median price and the type of dwelling. Two types of dwellings stand out that have houses selling above the median price, including two story houses built after 1945 and one store planned unit developments built after 1945.

ggplot(ames\_student, aes(x = Bldg\_Type, fill = Above\_Median)) +  
 geom\_bar() + theme(axis.text.x = element\_text(size = 7, angle = 90)) +  
 labs(title = "Building Type and Above Median Price",  
 x = "Building Type",  
 y = "Count")



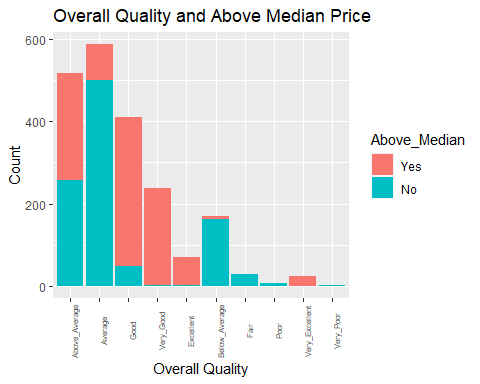
The above visualization compares whether or not houses sold above the median price and the building type. Three building types stand out that have houses not selling above the median price, including townhouses inside units, duplexes, and two family conversions from previous single family housing.

ggplot(ames\_student, aes(x = House\_Style, fill = Above\_Median)) +  
 geom\_bar() + theme(axis.text.x = element\_text(size = 7, angle = 90)) +  
 labs(title = "House Style and Above Median Price",  
 x = "House Style",  
 y = "Count")



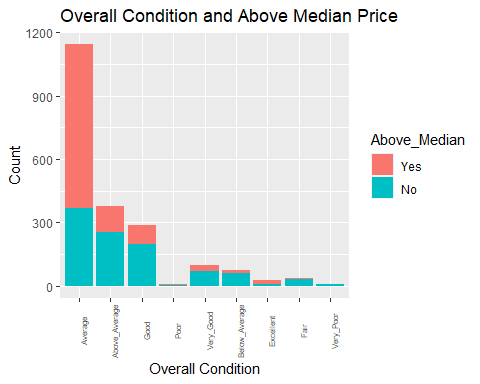
The above visualization compares whether or not houses sold above the median price and the house style. One style stands out that have houses selling above the median price, which is two story houses.

ggplot(ames\_student, aes(x = Overall\_Qual, fill = Above\_Median)) +  
 geom\_bar() + theme(axis.text.x = element\_text(size = 6, angle = 90)) +  
 labs(title = "Overall Quality and Above Median Price",  
 x = "Overall Quality",  
 y = "Count")



The above visualization compares whether or not houses sold above the median price and the overall quality. The overall quality rates the overall material and finish of the house. A few categories stand out that have houses selling above the median price, including Good, Very Good, Excellent, and Very Excellent.

ggplot(ames\_student, aes(x = Overall\_Cond, fill = Above\_Median)) +  
 geom\_bar() + theme(axis.text.x = element\_text(size = 6, angle = 90)) +  
 labs(title = "Overall Condition and Above Median Price",  
 x = "Overall Condition",  
 y = "Count")



The above visualization compares whether or not houses sold above the median price and the overall condition of the house. Surprisingly, only the Average Overall Condition shows a significant increase in the number of houses selling above the median price.