

Spatial Maps

Charles Hwang

12/17/2022

Data Cleaning

```
rm(list=ls())
load("~/Desktop/Notes/Graduate/STAT 401 - Statistical Consulting/consulting_data.RData")
load("~/Desktop/Notes/Graduate/STAT 401 - Statistical Consulting/location_data.RData")
location_data$XCoordinate<-as.numeric(location_data$XCoordinate)
location_data$YCoordinate<-as.numeric(location_data$YCoordinate)
location_data[which(is.na(location_data$NeighborhoodCluster)), "Neighborhood"]<-"0040"
location_data[which(is.na(location_data$NeighborhoodCluster)), "NeighborhoodCluster"]<-"1"
residential_data<-residential_data[residential_data$xrImprovedStatusID==2,]
residential_data<-residential_data[residential_data$ParcelID!="3480765000",]
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(tidyr)
residential_data<-residential_data %>% inner_join(location_data,by=c("ParcelID"="ParcelID"))
residential_data<-residential_data[-which(is.na(residential_data$XCoordinate)),]
residential_data$X<-residential_data$XCoordinate # Shortening variable names
residential_data$Y<-residential_data$YCoordinate
residential_data$Cluster<-residential_data$NeighborhoodCluster
c<-c("X","Y","Cluster","BldType","PhysicalCondition","Quality","YearBuilt","FullBath","HalfBath","Kitchen")
res<-residential_data[,c]
rm(c,location_data,property_values,residential_data,sales_data,sfyi_data)
```

Spatial Maps

```
library(ggplot2)
library(ggrepel)
# Neighborhood Cluster
res$Cluster<-factor(res$Cluster,labels=1:16)
ggplot(map_data("county","wisconsin,milwaukee"),aes(long,lat,col=res$Cluster))+ 
  geom_polygon(aes(long,lat,group=group),fill="white",color="black")+
```

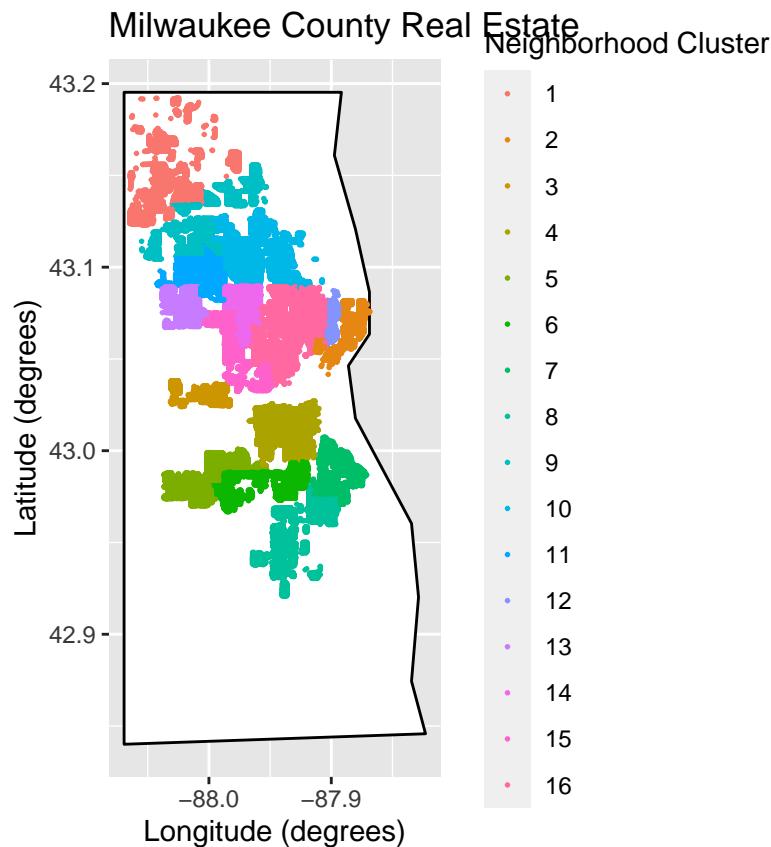
```

coord_fixed(1.5)+  

labs(color="Neighborhood Cluster",x="Longitude (degrees)",y="Latitude (degrees)",title="Milwaukee County Real Estate Neighborhood Cluster")  

geom_point(data=data.frame(long=res$X,lat=res$Y,stringsAsFactors=FALSE),aes(long,lat),size=0.3)

```



```

# Building Type  

res$BldType[which(res$BldType=="")]<-NA  

res$BldType[res$BldType=="Duplex 0/S"]<-"Duplex"  

res$BldType[res$BldType=="Res 0/S A & 1/2"]<-"Residential 0/S"  

res$BldType[res$BldType!="Ranch" & res$BldType!="Cape Cod" & res$BldType!="Duplex" & res$BldType!="Residential 0/S"]<-"Other"  

res$BldType<-factor(res$BldType,levels=c("Ranch","Cape Cod","Duplex","Residential 0/S","Milwaukee Bungalow"))  

ggplot(map_data("county","wisconsin,milwaukee"),aes(long,lat,col=res$BldType))+  

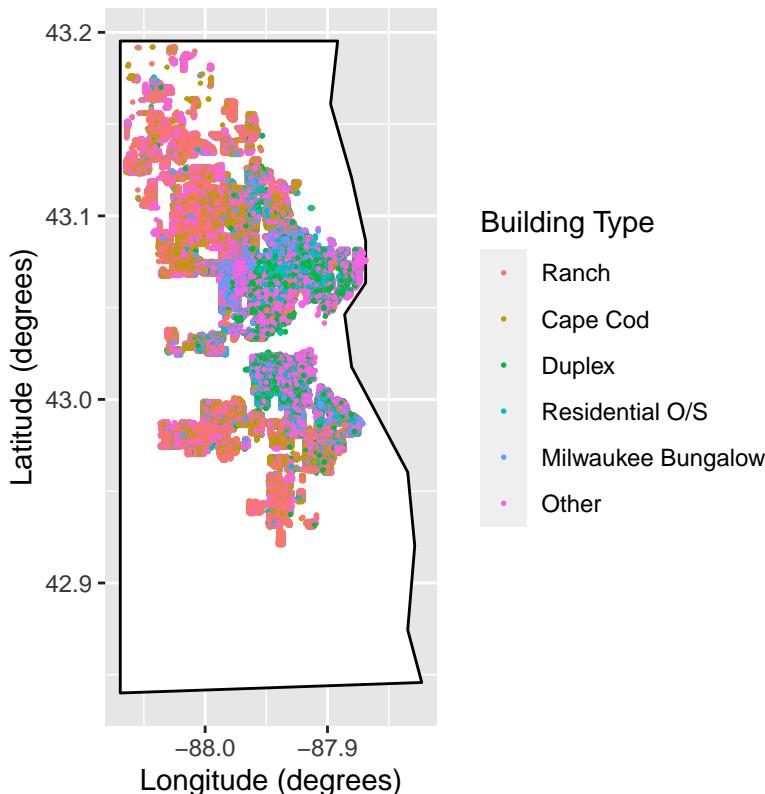
  geom_polygon(aes(long,lat,group=group),fill="white",color="black")+
  coord_fixed(1.5)+  

  labs(color="Building Type",x="Longitude (degrees)",y="Latitude (degrees)",title="Milwaukee County Real Estate Building Type")  

  geom_point(data=data.frame(long=res$X,lat=res$Y,stringsAsFactors=FALSE),aes(long,lat),size=0.3)

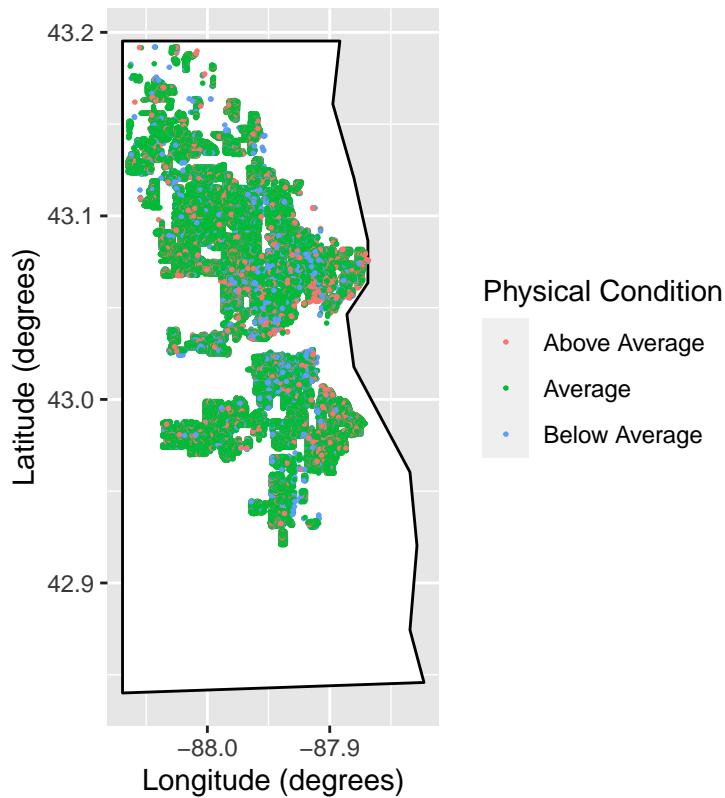
```

Milwaukee County Real Estate



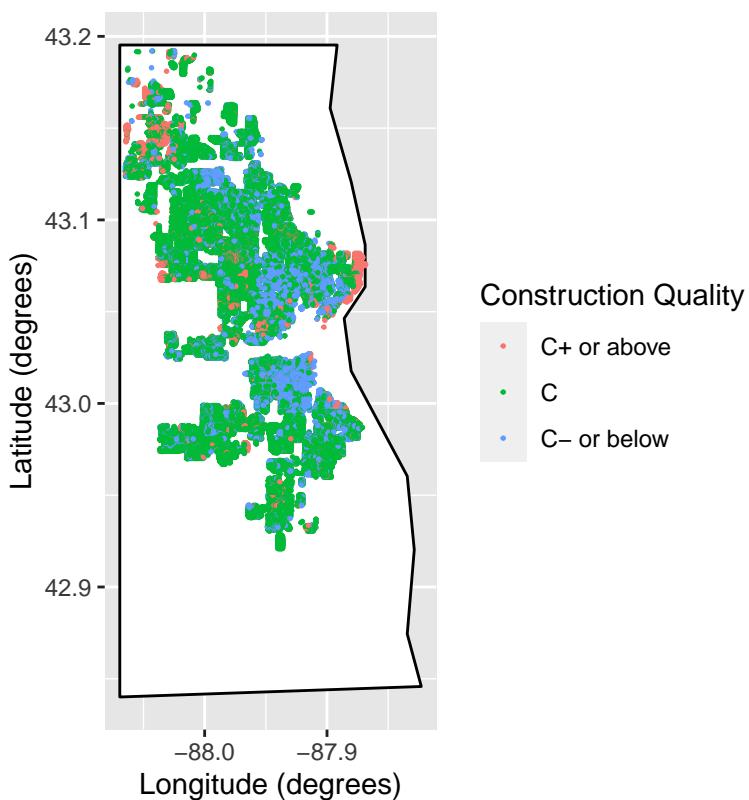
```
# Physical Condition
res$PhysicalCondition[res$PhysicalCondition=="EX" | res$PhysicalCondition=="VG" | res$PhysicalCondition=="P"]
res$PhysicalCondition[res$PhysicalCondition=="AV"] <- "Average"
res$PhysicalCondition[res$PhysicalCondition=="FR" | res$PhysicalCondition=="PR" | res$PhysicalCondition=="F"]
res$PhysicalCondition<-factor(res$PhysicalCondition)
ggplot(map_data("county","wisconsin,milwaukee"),aes(long,lat,col=res$PhysicalCondition))+  
  geom_polygon(aes(long,lat,group=group),fill="white",color="black")+
  coord_fixed(1.5)+  
  labs(color="Physical Condition",x="Longitude (degrees)",y="Latitude (degrees)",title="Milwaukee County Real Estate")+
  geom_point(data=data.frame(long=res$X,lat=res$Y,stringsAsFactors=FALSE),aes(long,lat),size=0.3)
```

Milwaukee County Real Estate

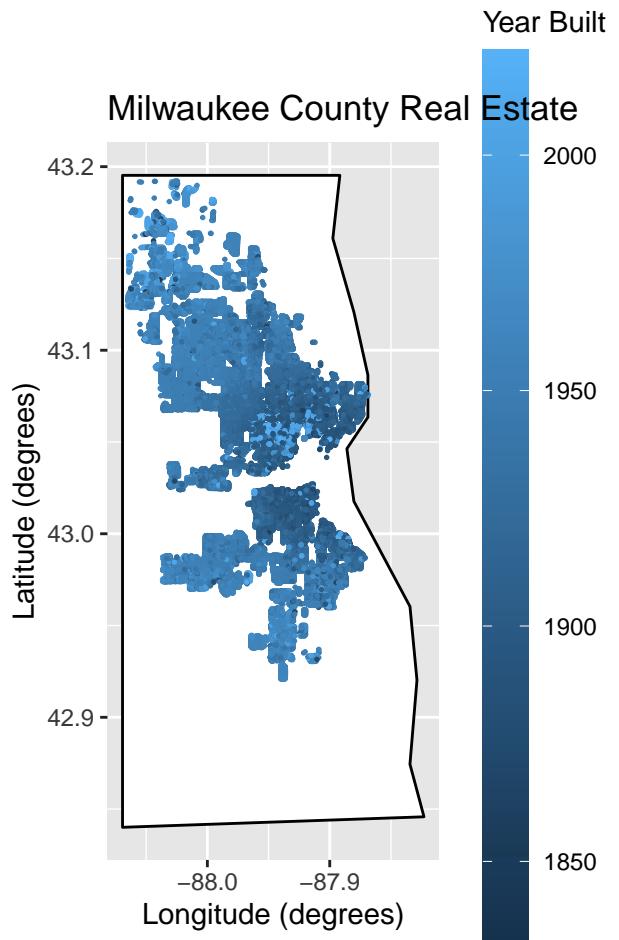


```
# Quality
res$Quality[res$Quality=="AA+" | res$Quality=="AA" | res$Quality=="AA-" | res$Quality=="A+" | res$Quality=="A" | res$Quality=="A-"] <- factor(res$Quality, levels=c("AA+", "AA", "AA-", "A+", "A", "A-"))
res$Quality[res$Quality=="C+" | res$Quality=="D+" | res$Quality=="D" | res$Quality=="D-" | res$Quality=="C-"] <- factor(res$Quality, levels=c("C+ or above", "C", "C- or below"))
ggplot(map_data("county", "wisconsin_milwaukee"), aes(long, lat, col=res$Quality)) +
  geom_polygon(aes(long, lat, group=group), fill="white", color="black") +
  coord_fixed(1.5) +
  labs(color="Construction Quality", x="Longitude (degrees)", y="Latitude (degrees)", title="Milwaukee County Real Estate") +
  geom_point(data=data.frame(long=res$X, lat=res$Y, stringsAsFactors=FALSE), aes(long, lat), size=0.3)
```

Milwaukee County Real Estate

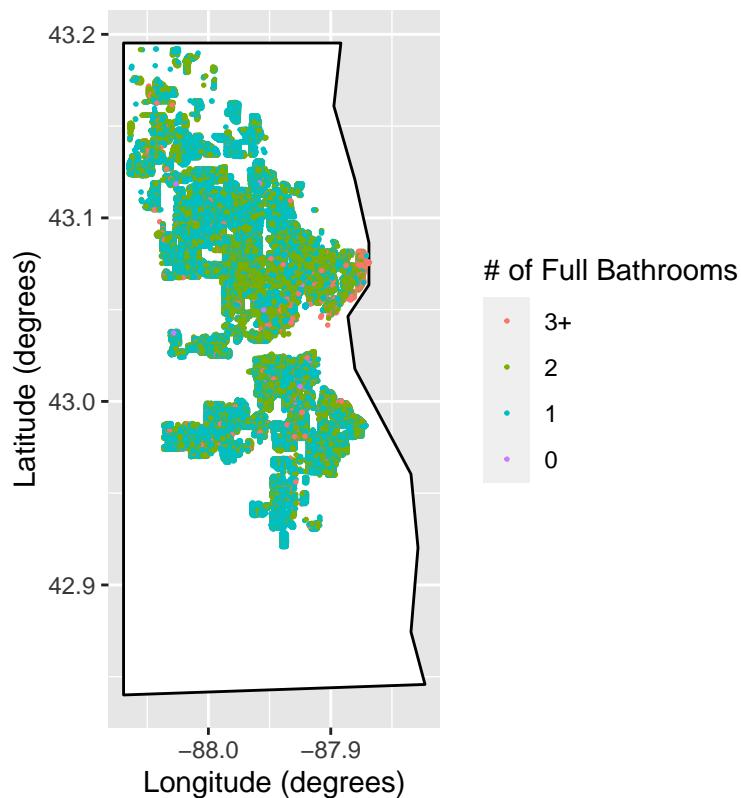


```
# Year Built
res$YearBuilt[res$YearBuilt==0] <- NA
ggplot(map_data("county", "wisconsin_milwaukee"), aes(long, lat, col=res$YearBuilt)) +
  geom_polygon(aes(long, lat, group=group), fill="white", color="black") +
  coord_fixed(1.5) +
  labs(color="Year Built", x="Longitude (degrees)", y="Latitude (degrees)", title="Milwaukee County Real Estate") +
  geom_point(data=data.frame(long=res$X, lat=res$Y, stringsAsFactors=FALSE), aes(long, lat), size=0.3) +
  theme(legend.key.height=unit(2.5, "cm"))
```



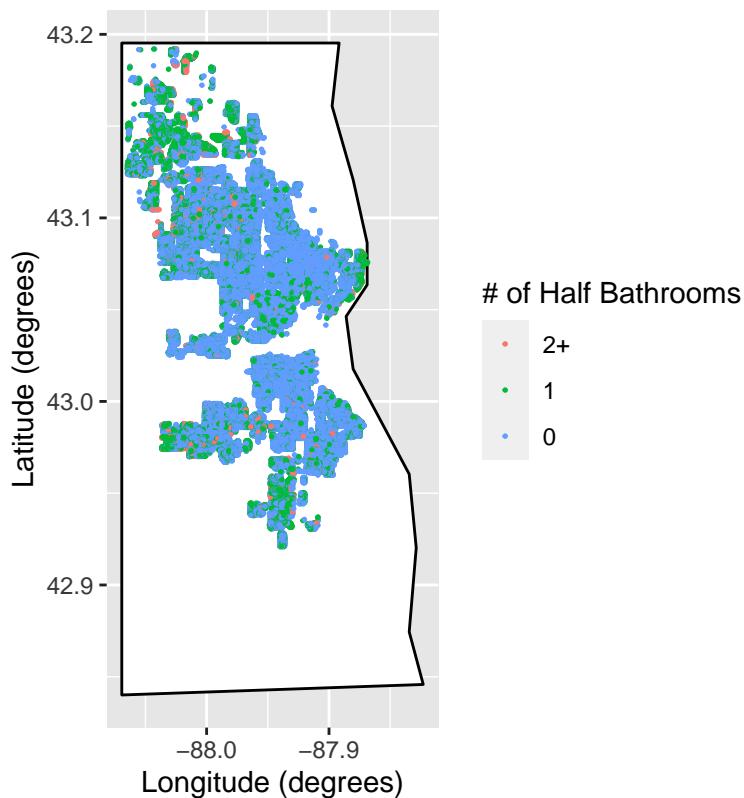
```
# Full Bathrooms
res$FullBath[is.na(res$FullBath)] <- 0
res$FullBath[res$FullBath>=3] <- "3+"
res$FullBath <- factor(res$FullBath, levels=c("3+","2","1","0"))
ggplot(map_data("county", "wisconsin_milwaukee"), aes(long, lat, col=res$FullBath)) +
  geom_polygon(aes(long, lat, group=group), fill="white", color="black") +
  coord_fixed(1.5) +
  labs(color="# of Full Bathrooms", x="Longitude (degrees)", y="Latitude (degrees)", title="Milwaukee County Real Estate")
  geom_point(data=data.frame(long=res$x, lat=res$y, stringsAsFactors=FALSE), aes(long, lat), size=0.3)
```

Milwaukee County Real Estate



```
# Half Bathrooms
res$HalfBath[is.na(res$HalfBath)] <- 0
res$HalfBath[res$HalfBath>=2] <- "2+"
res$HalfBath <- factor(res$HalfBath, levels=c("2+","1","0"))
ggplot(map_data("county", "wisconsin_milwaukee"), aes(long, lat, col=res$HalfBath)) +
  geom_polygon(aes(long, lat, group=group), fill="white", color="black") +
  coord_fixed(1.5) +
  labs(color="# of Half Bathrooms", x="Longitude (degrees)", y="Latitude (degrees)", title="Milwaukee County Real Estate")
  geom_point(data=data.frame(long=res$x, lat=res$y, stringsAsFactors=FALSE), aes(long, lat), size=0.3)
```

Milwaukee County Real Estate



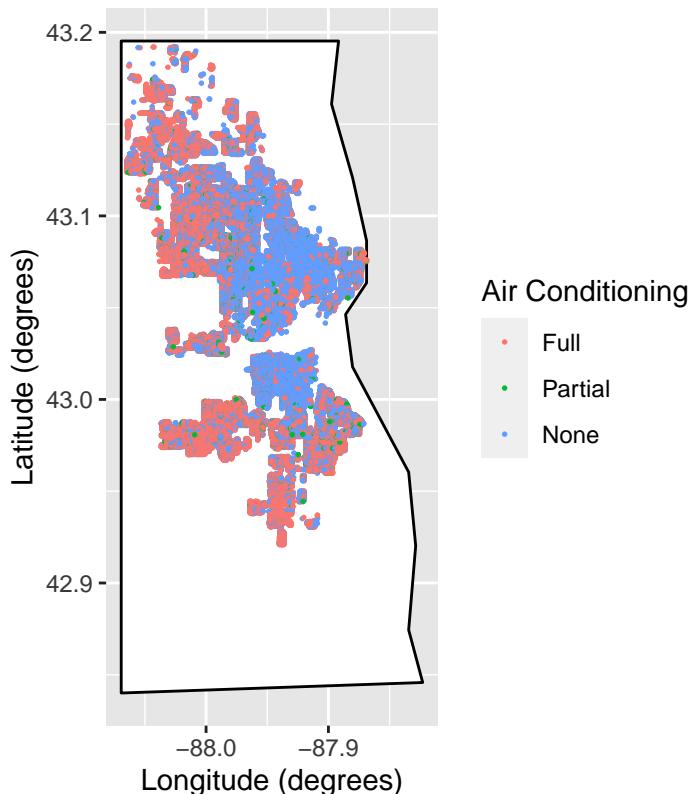
```
# Kitchens
res$Kitchen[is.na(res$Kitchen)]<-0
#res$Kitchen[res$Kitchen<=1]<- "0-1"
res$Kitchen[res$Kitchen>=3]<- "3+"
res$Kitchen<-factor(res$Kitchen,levels=c("3+","2","1","0"))
ggplot(map_data("county","wisconsin,milwaukee"),aes(long,lat,col=res$Kitchen))+  
  geom_polygon(aes(long,lat,group=group),fill="white",color="black")+
  coord_fixed(1.5)+  
  labs(color="# of Kitchens",x="Longitude (degrees)",y="Latitude (degrees)",title="Milwaukee County Real  
estate")  
  geom_point(data=data.frame(long=res$X,lat=res$Y,stringsAsFactors=FALSE),aes(long,lat),size=0.3)
```

Milwaukee County Real Estate



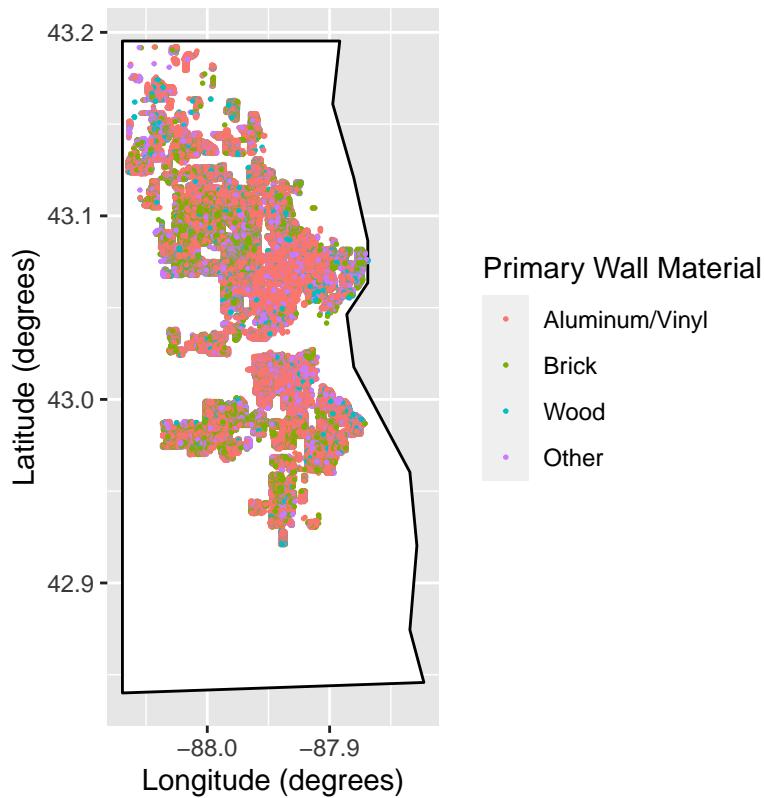
```
# Air Conditioning Status
res$PercentAirConditioned[which(res$PercentAirConditioned==1)] <- 100
res$PercentAirConditioned[which(res$PercentAirConditioned==200)] <- 100
res$AC[res$PercentAirConditioned==100] <- "Full"
res$AC[res$PercentAirConditioned!=0 & res$PercentAirConditioned!=100] <- "Partial"
res$AC[res$PercentAirConditioned==0] <- "None"
res$AC[is.na(res$AC)] <- "Partial"
res$AC<-factor(res$AC,levels=c("Full","Partial","None"))
ggplot(map_data("county","wisconsin,milwaukee"),aes(long,lat,col=res$AC))+
  geom_polygon(aes(long,lat,group=group),fill="white",color="black")+
  coord_fixed(1.5)+
  labs(color="Air Conditioning",x="Longitude (degrees)",y="Latitude (degrees)",title="Milwaukee County Real Estate")
  geom_point(data=data.frame(long=res$x,lat=res$y,stringsAsFactors=FALSE),aes(long,lat),size=0.3)
```

Milwaukee County Real Estate



```
# Primary Wall
res$PrimaryWall[res$PrimaryWall!="Aluminum/Vinyl" & res$PrimaryWall!="Brick" & res$PrimaryWall!="Wood"]
res$PrimaryWall<-factor(res$PrimaryWall,levels=c("Aluminum/Vinyl","Brick","Wood","Other"))
ggplot(map_data("county","wisconsin,milwaukee"),aes(long,lat,col=res$PrimaryWall))+  
  geom_polygon(aes(long,lat,group=group),fill="white",color="black")+
  coord_fixed(1.5)+  
  labs(color="Primary Wall Material",x="Longitude (degrees)",y="Latitude (degrees)",title="Milwaukee Co  
  geom_point(data=data.frame(long=res$X,lat=res$Y,stringsAsFactors=FALSE),aes(long,lat),size=0.3)
```

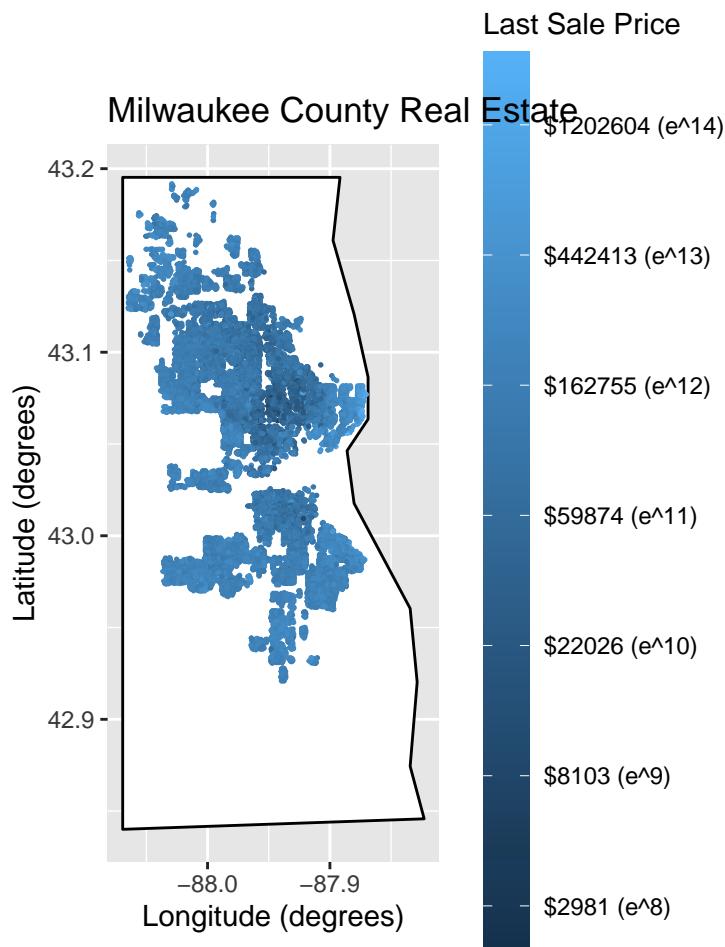
Milwaukee County Real Estate



```
# Last Sale Price
sales<-read.csv("~/Desktop/Notes/Graduate/STAT 401 - Statistical Consulting/Fixed Data for Modeling.csv")
ggplot(map_data("county","wisconsin,milwaukee"),aes(long,lat,col=log(sales$LastSalePrice)))+
  geom_polygon(aes(long,lat,group=group),fill="white",color="black")+
  coord_fixed(1.5)+
  labs(color="Last Sale Price",x="Longitude (degrees)",y="Latitude (degrees)",title="Milwaukee County Real Estate Sales Map")+
  geom_point(data=data.frame(long=sales$XCoordinate,lat=sales$YCoordinate,stringsAsFactors=FALSE),aes(long,lat),size=1,shape=16)+
```

scale_color_continuous(breaks=seq(8,14,1),labels=paste("\$",round(exp(seq(8,14,1)))), " (e^",seq(8,14,1)))

```
theme(legend.key.height=unit(2.5,"cm"))
```



```
# Total Area (square feet)
res$TotalFinishedArea[which(res$TotalFinishedArea==0)]<-NA
ggplot(map_data("county","wisconsin,milwaukee"),aes(long,lat,col=log(res$TotalFinishedArea)))+
  geom_polygon(aes(long,lat,group=group),fill="white",color="black")+
  coord_fixed(1.5)+  

  labs(color="Total Area (square feet)",x="Longitude (degrees)",y="Latitude (degrees)",title="Milwaukee")
  geom_point(data=data.frame(long=res$X,lat=res$Y,stringsAsFactors=FALSE),aes(long,lat),size=0.3)+  

  scale_color_continuous(breaks=seq(6,10,0.5),labels=paste(round(exp(seq(6,10,0.5))),"  

  theme(legend.key.height=unit(2.5,"cm")))
```

