Final Project

2022-08-20

# MIS503 – Final Project Part 3

## Hagen, Tabitha

### Zillow Home Value Index Analysis

### R Setup

knitr::opts\_chunk$set(echo = TRUE)  
library(tidyverse)  
library(readr)  
SingleFamilyResidenceRental <- read\_csv("SingleFamilyResidenceRental.csv") # copy original file  
SingleFamilyResidenceSales <- read\_csv("SingleFamilyResidenceSales.csv") # copy original file

### Home Values in Select Rental Markets

#### You have made the choice that you want to focus on 4 regions (Asheville, Charlotte-Concord-Gastonia, Raleigh and Wilmington) and instead of renting, you would like to purchase a home.

##### a. According to the results, which market has the lowest median price (represented as horizontal bar in box plot)?

The Charlotte-Concord-Gastonia Metro has the lowest median price, however, the other 3 cities are not far from it. Asheville appears to have the highest median price.

##### b. The violin plot will show density meaning the wider the plot is, the more observations occur within that area. Which market has the most density around the median value of homes?

The Charlotte-Concord-Gastonia Metro has the widest violin around the median, however, Raleigh comes in at a close second. Asheville comes in as the one with more higher priced homes. If you added how much homewoner’s insurance cost these homes, I bet Wilmington would have teh highest price.

# create a new dataset that will be used for all commands after the pipe sign  
NCHomeSales <- SingleFamilyResidenceSales %>% select (RegionName,   
 State,   
 CountyName,   
 Metro,   
 ends\_with("-05"))   
  
# Modify the variables names by taking off the suffix "-05" and removing NA's  
NCHomeSales <- NCHomeSales %>% # piping reduces redundant commands   
 filter(!is.na(RegionName)) %>% # don't use the na observations  
 filter(!is.na(State)) %>% # don't use the na observations  
 filter(!is.na(CountyName)) %>% # don't use the na observations  
 filter(!is.na(Metro)) %>% # don't use the na observations  
 rename\_at(vars(matches("-05")), ~str\_remove(., "-05")) %>%  
 # remove suffix from Variables  
 pivot\_longer(c('1996':"2018"),names\_to='YR',values\_to='ZHVI')%>%   
 # rewrite columns to rows  
 filter(!is.na(YR)) %>% # don't use the na observations  
 filter(!is.na(ZHVI)) %>% # don't use the na observations  
 filter(State=="NC") %>% # filter out homes in NC  
 # filter out Raleigh, Wilmington, Asheville & the Charlotte areas  
 filter(Metro=="Raleigh" | Metro=="Asheville" | Metro=="Wilmington" | Metro=="Charlotte-Concord-Gastonia")   
  
NCHomeSales<- NCHomeSales %>% group\_by(Metro)

# violin plot of the Wake County info  
ggplot(NCHomeSales,mapping=aes(Metro,ZHVI, # x axis = Metro, y axis = ZHVI  
 color = Metro  
 )) +   
 geom\_violin() +  
 geom\_boxplot(width=0.1)+  
 labs(title = "Metro vs ZHVI value 2016-2018 for 4 Cities", x = "Metro",   
 y="ZHVI")+ # Graph Labels  
 # changes x-axis labels to display vertical values  
 theme(axis.text.x = element\_text(angle = 90, vjust=0.5))+   
 scale\_y\_continuous(labels = scales::comma) # labels get comma values

