# MIS 503 Final Project

## Graybill, Gil

### Zillow Home Value Index Analysis

### Wake County Home Sales

library(tidyverse)

## -- Attaching packages -------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.2.1 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.3  
## v tidyr 1.0.0 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

## -- Conflicts ----------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(lubridate)

##   
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':  
##   
## date

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

## Parsed with column specification:  
## cols(  
## .default = col\_double(),  
## RegionName = col\_character(),  
## State = col\_character(),  
## Metro = col\_character(),  
## CountyName = col\_character()  
## )

## See spec(...) for full column specifications.

1. What have been the overall trends in Wake County Home Values?

ANSWER: For the most part, it’s been steadily increasing.

1. There were dips in home values in the past 20 years. What years did these occur?

ANSWER: The dip started in 2008, and ended in 2011. This corresponds to the nationwide mortgage crisis.

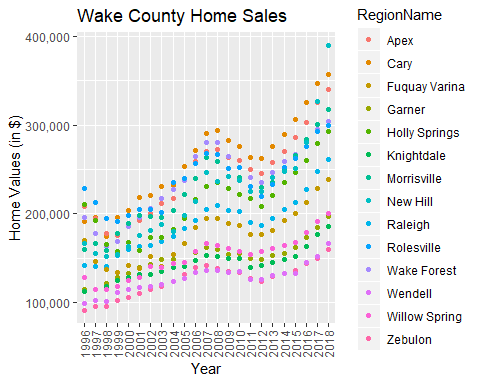
1. Based on the analysis, where would be the least expensive area to purchase home? Most expensive area?

ANSWER: It looks like Zebulon and Wendell would be the least expensive areas. Can we factor in loss of dignity? Most expensive areas look like Cary and Apex. I hope you like beige.

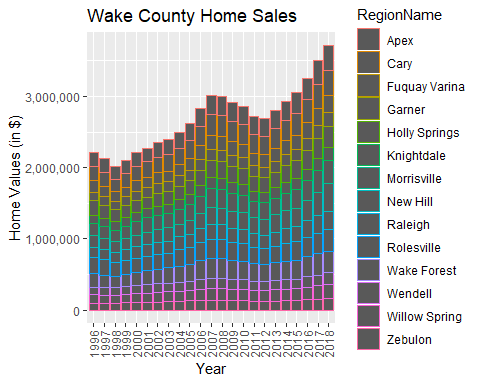
1. Are any area home values trending down? Is there one area that stands out compared to others?

ANSWER: They all seem to be going up, although Zebulon, Wendel, Garner and Knightdale seem to be growing at a slower rate.

WakeCountySales = filter(SingleFamilyResidenceSales, State == "NC", CountyName == "Wake County") %>%   
 select(RegionName, State, CountyName, Metro, ends\_with("-05"))  
colnames(WakeCountySales) <- sub("\\-05", "", colnames(WakeCountySales))   
WakeCountySales <- gather(WakeCountySales, `1996`, `1997`, `1998`, `1999`, `2000`, `2001`, `2002`, `2003`, `2004`, `2005`, `2006`, `2007`, `2008`, `2009`, `2010`, `2011`, `2012`, `2012`, `2013`, `2014`, `2015`, `2016`, `2017`, `2018`, key = "YR", value = "ZHVI")  
  
p1 = ggplot(data = WakeCountySales, aes(x = YR, y = ZHVI, color = RegionName))+  
 geom\_point()+  
 labs(title = "Wake County Home Sales", # plot title  
 x = "Year")+  
 theme(axis.text.x = element\_text(angle = 90, vjust=0.5)) +  
 scale\_y\_continuous(name="Home Values (in $)", labels = scales::comma)  
p1



p2 = ggplot(data = WakeCountySales, aes(x = YR, y = ZHVI, color = RegionName))+  
 geom\_col()+  
 labs(title = "Wake County Home Sales", # plot title  
 x = "Year")+  
 theme(axis.text.x = element\_text(angle = 90, vjust=0.5)) +  
 scale\_y\_continuous(name="Home Values (in $)", labels = scales::comma)  
p2



### NC Rental Market

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

## Parsed with column specification:  
## cols(  
## .default = col\_double(),  
## RegionName = col\_character(),  
## State = col\_character(),  
## Metro = col\_character(),  
## CountyName = col\_character()  
## )

## See spec(...) for full column specifications.

1. What has been the overall trend in the rental market around the state? Are there any cities that have not followed this trend?

ANSWER: The overall trend has been a steady increase for the past 9 years. The exception is Fayetteville, which is mostly steady at around $900.

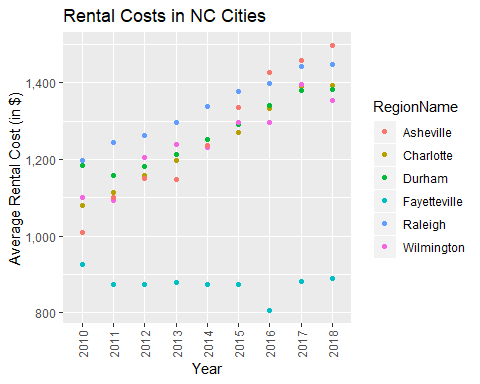
1. Where is the most expensive city to rent in? Least expensive?

ANSWER: For 2018, the most expensive city to rent in is Asheville at $1496. The least expensive Fayetteville at $889

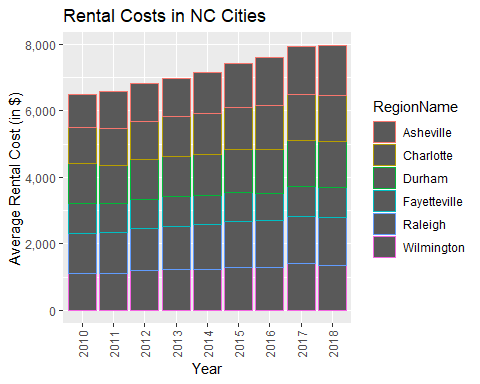
1. You are trying decide between Wilmington and Asheville. Which market has the lowest rent?

ANSWER: First of all, is there even a choice between Wilmington and Asheville, at any price? There’s an awesome school called UNCW there, and in Asheville they have vegans. But to the point, Wilmington’s average rent is around $1350, and Asheville is around $1500. So, to no one’s surprise, Wilmington comes out the winner.

Rentals = filter(SingleFamilyResidenceRental, State == "NC", RegionName %in% c("Asheville", "Charlotte", "Durham", "Fayetteville", "Raleigh" , "Wilmington")) %>%  
 select(RegionName, State, ends\_with("-11"), '2018-10')   
colnames(Rentals) <- sub("\\-11", "", colnames(Rentals))   
colnames(Rentals) <- sub("\\-10", "", colnames(Rentals))  
Rentals <- gather(Rentals, `2010`, `2011`, `2012`, `2012`, `2013`, `2014`, `2015`, `2016`, `2017`, `2018`, key = "YR", value = "ZHVI")  
  
p3 = ggplot(data = Rentals, aes(x = YR, y = ZHVI, color = RegionName))+  
 geom\_point()+  
 labs(title = "Rental Costs in NC Cities", # plot title  
 x = "Year")+  
 theme(axis.text.x = element\_text(angle = 90, vjust=0.5)) +  
 scale\_y\_continuous(name="Average Rental Cost (in $)", labels = scales::comma)  
p3



p4 = ggplot(data = Rentals, aes(x = YR, y = ZHVI, color = RegionName))+  
 geom\_col()+  
 labs(title = "Rental Costs in NC Cities", # plot title  
 x = "Year")+  
 theme(axis.text.x = element\_text(angle = 90, vjust=0.5)) +  
 scale\_y\_continuous(name="Average Rental Cost (in $)", labels = scales::comma)  
p4



### Home Values in Select Rental Markets

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

ANSWER: Charlotte-Concord-Gastonia has the lowest median price, right around $130,000.

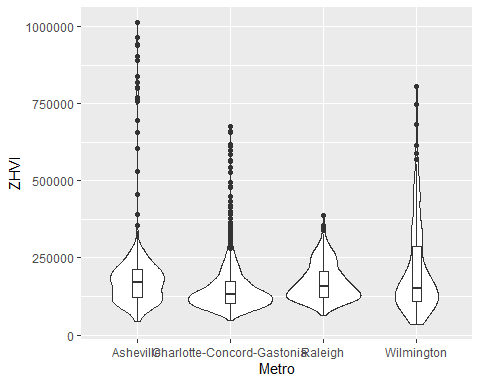
1. The violin plot will show density meaning the wider the plot is, the more observations occur within that area. Which market has the most density around the median value of homes?

ANSWER: Charlotte-Concord-Gastonia is the widest one around the median, so there should be a lot of houses available around the median value.

NCHomeSales = filter(SingleFamilyResidenceSales, State == "NC", Metro %in% c("Asheville", "Charlotte-Concord-Gastonia", "Raleigh" , "Wilmington")) %>%   
 select(RegionName, State, Metro, ends\_with("-05"))  
colnames(NCHomeSales) <- sub("\\-05", "", colnames(NCHomeSales))   
NCHomeSales <- gather(NCHomeSales, `1996`, `1997`, `1998`, `1999`, `2000`, `2001`, `2002`, `2003`, `2004`, `2005`, `2006`, `2007`, `2008`, `2009`, `2010`, `2011`, `2012`, `2012`, `2013`, `2014`, `2015`, `2016`, `2017`, `2018`, key = "YR", value = "ZHVI")  
NCHomeSales <- NCHomeSales %>% group\_by(Metro)  
  
p5 = ggplot(data = NCHomeSales, aes(x = Metro, y = ZHVI))+  
 geom\_violin()+  
 geom\_boxplot(width=0.1)  
   
p5

## Warning: Removed 90 rows containing non-finite values (stat\_ydensity).

## Warning: Removed 90 rows containing non-finite values (stat\_boxplot).



### Relocation Home Value Comparison

1. Based on your analysis, which city’s housing is most affordable? Least affordable?

ANSWER: In 2018, Chicago and Houston look to be the most affordable at around $200,000. New York tops it off at around $600,000

1. Which cities saw the largest change in prices over the past 5 years? Which city has remained more consistent (i.e., no huge swings up or down in home values)?

ANSWER: In the last 5 years, Denver has gone from about $260K to about $450K, or a change of about $190K, or a 73% increase. New York has risen from $417K to $590K, or a change of about $173K, or a 41% increase. The Houston Area looks pretty consistent over the time period.

1. During the market downturn in 2012, which cities were most impacted? Which cities have recovered?

New York and Chicago seemed to feel the affects of the downturn. New York has recovered quite well, whereas Chicago is still trying to gain that all back.

NationalHomeSales = filter(SingleFamilyResidenceSales, (RegionName == "Chicago" & State == "IL") | (RegionName == "Denver" & State == "CO") | (RegionName == "Houston" & State == "TX") |(RegionName == "New York"& State == "NY")) %>%   
 select(RegionName, State, Metro, ends\_with("-05"))  
colnames(NationalHomeSales) <- sub("\\-05", "", colnames(NationalHomeSales))   
NationalHomeSales <- gather(NationalHomeSales, `1996`, `1997`, `1998`, `1999`, `2000`, `2001`, `2002`, `2003`, `2004`, `2005`, `2006`, `2007`, `2008`, `2009`, `2010`, `2011`, `2012`, `2012`, `2013`, `2014`, `2015`, `2016`, `2017`, `2018`, key = "YR", value = "ZHVI")  
  
p6 = ggplot(data = NationalHomeSales, aes(x = YR, y = ZHVI))+  
 geom\_point()+  
 facet\_wrap(~Metro)+  
 labs(title = "City Home Value Comparison", # plot title  
 x = "Year")+  
 theme(axis.text.x = element\_text(angle = 90, vjust=0.5)) +  
 scale\_y\_continuous(name="Home Values (in $)", labels = scales::comma)  
  
p6

