# MIS503 – Final Project

## Black, Tyler

### Zillow Home Value Index Analysis

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

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

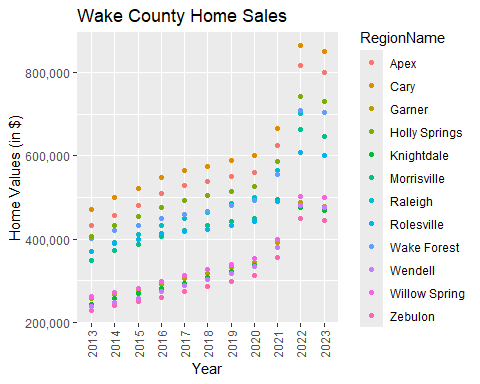
## Rows: 3503 Columns: 107  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (4): RegionName, State, Metro, CountyName  
## dbl (103): RegionID, 1/31/2015, 2/28/2015, 3/31/2015, 4/30/2015, 5/31/2015, ...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

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

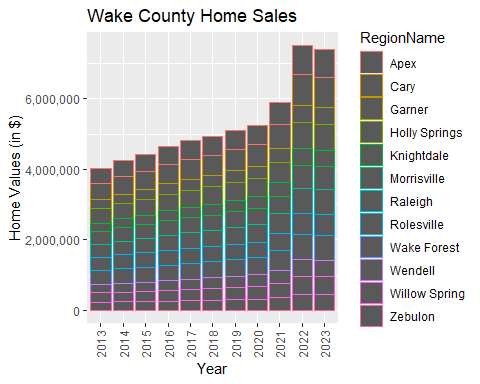
## Rows: 22275 Columns: 287  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (4): RegionName, State, Metro, CountyName  
## dbl (283): RegionID, 1/31/2000, 2/29/2000, 3/31/2000, 4/30/2000, 5/31/2000, ...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

### Wake County Home Sales

WakeCountySales <- SingleFamilyResidenceSales %>%  
 filter( CountyName == 'Wake County' & State == 'NC') %>%  
 select(RegionID,RegionName,CountyName, Metro, '5/31/2013', '5/31/2014', '5/31/2015',  
 '5/31/2016', '5/31/2017', '5/31/2018', '5/31/2019', '5/31/2020', '5/31/2021',  
 '5/31/2022', '5/31/2023') %>%  
 rename('2013' = '5/31/2013', '2014' = '5/31/2014', '2015' = '5/31/2015',  
 '2016' = '5/31/2016', '2017' = '5/31/2017', '2018' = '5/31/2018',   
 '2019' ='5/31/2019', '2020' = '5/31/2020', '2021' ='5/31/2021',  
 '2022' = '5/31/2022', '2023' = '5/31/2023') %>%  
 pivot\_longer(cols = '2013':'2023', names\_to = 'YR', values\_to = 'ZHVI')  
  
ggplot(data = WakeCountySales, mapping = aes( x=YR ,y = ZHVI, color = RegionName)) +  
 geom\_point() +  
 theme(axis.text.x = element\_text(angle = 90, vjust=0.5)) +  
 scale\_y\_continuous(name="Home Values (in $)", labels = scales::comma) +  
 labs(title = 'Wake County Home Sales', x = "Year")



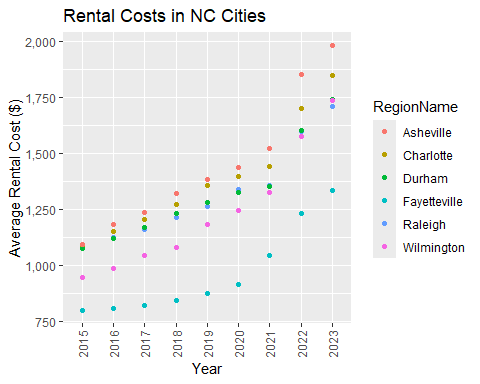
ggplot(data = WakeCountySales, mapping = aes( x=YR ,y = ZHVI, color = RegionName)) +  
 geom\_col() +  
 theme(axis.text.x = element\_text(angle = 90, vjust=0.5)) +  
 scale\_y\_continuous(name="Home Values (in $)", labels = scales::comma) +  
 labs(title = 'Wake County Home Sales', x = "Year")



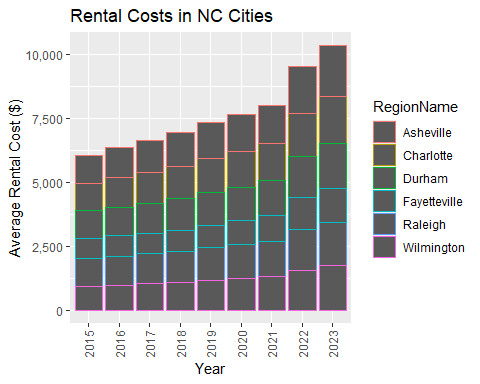
1. What have been the overall trends in Wake County Home Values?  
   **They have increased over the years, with a slight decrease in 2023**
2. There were dips in home values in the past 10 years. What years did these occur?  
   **Looks like the only dip is from 2022 to 2023**
3. Based on the analysis, where would be the least expensive area to purchase home? Most expensive area?  
   **Most expensive is Cary and the least expensive is Zebulon**
4. What has happened to the overall property values in Apex and Cary in 2023?  
   **They have decreased**

### NC Rental Market

Rentals <- SingleFamilyResidenceRental %>%  
 filter(RegionName %in% c('Asheville', 'Charlotte', 'Durham',   
 'Fayetteville', 'Raleigh', 'Wilmington') &   
 State == 'NC') %>%  
 select(RegionName, State, '1/31/2015', '1/31/2016', '1/31/2017', '1/31/2018',  
 '1/31/2019', '1/31/2020', '1/31/2021', '1/31/2022', '1/31/2023') %>%  
 rename('2015' = '1/31/2015', '2016' = '1/31/2016', '2017' = '1/31/2017',   
 '2018' = '1/31/2018', '2019' = '1/31/2019', '2020' = '1/31/2020',   
 '2021' = '1/31/2021', '2022' = '1/31/2022', '2023' = '1/31/2023') %>%  
 pivot\_longer(cols = '2015':'2023', names\_to = 'YR', values\_to = 'ZHVI')  
  
  
ggplot(data = Rentals, mapping = aes( x=YR ,y = ZHVI, color = RegionName)) +  
 geom\_point() +  
 theme(axis.text.x = element\_text(angle = 90, vjust=0.5)) +  
 scale\_y\_continuous(name="Average Rental Cost ($)", labels = scales::comma) +  
 labs(title = 'Rental Costs in NC Cities', x = "Year")



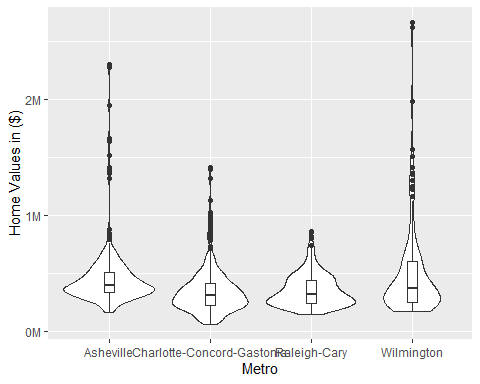
ggplot(data = Rentals, mapping = aes( x=YR ,y = ZHVI, color = RegionName)) +  
 geom\_col() +  
 theme(axis.text.x = element\_text(angle = 90, vjust=0.5)) +  
 scale\_y\_continuous(name="Average Rental Cost ($)", labels = scales::comma) +  
 labs(title = 'Rental Costs in NC Cities', x = "Year")



1. What has been the overall trend in the rental market around the state? Are there any cities that have not followed this trend?  
   **The prices have increased over the years, doesn’t seem like any city hasn’t followed this trend**
2. Where is the most expensive city to rent in? Least expensive?  
   **The most expensive is Asheville and the least expensive is Fayetteville**
3. You are trying to decide between Wilmington and Asheville. Which market has the lowest rent?  
   **Wilmington has a lower rent than Asheville**

### Home Values in Select Markets

NCHomeSales <- SingleFamilyResidenceSales %>%  
 filter( Metro %in% c('Asheville', 'Charlotte-Concord-Gastonia', 'Raleigh-Cary', 'Wilmington')   
 & State == 'NC') %>%  
 select(RegionName,State, Metro, '5/31/2013', '5/31/2014', '5/31/2015',  
 '5/31/2016', '5/31/2017', '5/31/2018', '5/31/2019', '5/31/2020', '5/31/2021',  
 '5/31/2022', '5/31/2023') %>%  
 rename('2013' = '5/31/2013', '2014' = '5/31/2014', '2015' = '5/31/2015',  
 '2016' = '5/31/2016', '2017' = '5/31/2017', '2018' = '5/31/2018',   
 '2019' ='5/31/2019', '2020' = '5/31/2020', '2021' ='5/31/2021',  
 '2022' = '5/31/2022', '2023' = '5/31/2023') %>%  
 pivot\_longer(cols = '2013':'2023', names\_to = 'YR', values\_to = 'ZHVI') %>%  
 filter(!is.na(ZHVI)) %>%  
 group\_by(Metro)   
   
  
ggplot(data = NCHomeSales, mapping = aes(x=Metro, y= ZHVI)) +  
 geom\_violin() +   
 geom\_boxplot(width = .1) +  
 scale\_y\_continuous(labels = scales::label\_number(scale = 1e-6, suffix = 'M')) +  
 labs(y = 'Home Values in ($)')

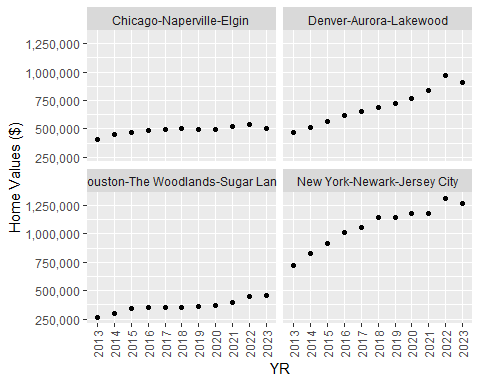


NCHomeSales2 <- NCHomeSales %>%  
 group\_by(Metro) %>%  
 summarise(median = median(ZHVI))

1. According to the results, which market has the lowest median price (represented as horizontal bar in box plot)?  
   **Looks like Charlotte-Concord-Gastonia, but Raleigh-Cary is very close**
2. 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?  
   **I would say Asheville, but someone could argue Raleigh-Cary**
3. The box plot will also show outliers in the various markets. Which metro area had the largest outlier (i.e., the highest value home sold in the past 10 years)?  
   **Wilmington had the largest outlier**

### Relocation Home Value Comparison

NationalHomeSales <- SingleFamilyResidenceSales %>%  
 filter(RegionName %in% c('Chicago', 'New York', 'Denver', 'Houston')   
 & CountyName %in% c('Queens County', 'Denver County', 'Cook County', 'Harris County')) %>%  
 select(RegionName,State, CountyName, Metro, '5/31/2013', '5/31/2014', '5/31/2015',  
 '5/31/2016', '5/31/2017', '5/31/2018', '5/31/2019', '5/31/2020', '5/31/2021',  
 '5/31/2022', '5/31/2023') %>%  
 rename('2013' = '5/31/2013', '2014' = '5/31/2014', '2015' = '5/31/2015',  
 '2016' = '5/31/2016', '2017' = '5/31/2017', '2018' = '5/31/2018',   
 '2019' ='5/31/2019', '2020' = '5/31/2020', '2021' ='5/31/2021',  
 '2022' = '5/31/2022', '2023' = '5/31/2023') %>%  
 pivot\_longer(cols = '2013':'2023', names\_to = 'YR', values\_to = 'ZHVI') %>%  
 filter(!is.na(ZHVI))  
  
ggplot(data = NationalHomeSales, mapping = aes( x=YR ,y = ZHVI)) +  
 geom\_point() +  
 theme(axis.text.x = element\_text(angle = 90, vjust=0.5)) +  
 scale\_y\_continuous(name="Home Values ($)", labels = scales::comma) +   
 facet\_wrap(~Metro)



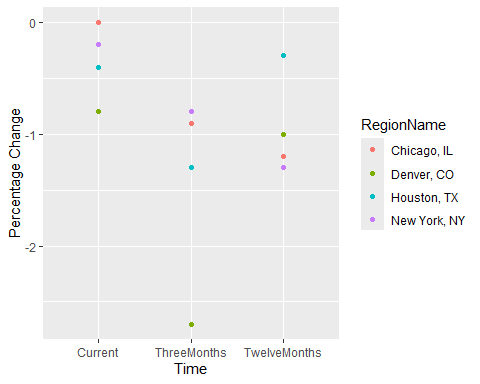
1. Based on your analysis, which city’s housing is most affordable? Least affordable?  
   **The most affordable city is Houston and the least affordable is New York**
2. Which cities saw the largest change in prices over the past 5 years? Which city has remained more consistent (i.e., no huge swings up or down in home values)?  
   **You could argue that either New York or Denver have had the largest change in prices over the past the five years, but Denver seems to have the most within the the more recent years. Chicago seems to be the most consistent in home prices, but Houston is not too far behind**
3. Which cities saw a decline in value during 2023 and which cities remained consistent?  
   **Every city but Houston saw a decline in value during 2023**

### Future Home Values

Projections <- read\_csv("Projections.csv")

## Rows: 895 Columns: 9  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (3): RegionName, RegionType, StateName  
## dbl (5): RegionID, SizeRank, 2024-07-31, 2024-09-30, 2025-06-30  
## date (1): BaseDate  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

FutureHomeValues <- Projections %>%  
 filter(RegionName %in% c('Chicago, IL', 'Denver, CO', 'Houston, TX', 'New York, NY')) %>%  
 select(RegionName, '2024-07-31', '2024-09-30', '2025-06-30') %>%  
 rename('Current' = '2024-07-31', 'ThreeMonths' = '2024-09-30', 'TwelveMonths' = '2025-06-30') %>%  
 pivot\_longer(cols = 'Current':'TwelveMonths', names\_to = 'Time', values\_to = 'Percentage Change')  
  
ggplot(data=FutureHomeValues, mapping = aes(x=Time,y=`Percentage Change`, color = RegionName)) +  
 geom\_point()



1. Which is the only city that is projected to have a decrease in home values in the next 3 months?  
   **Every city is expected to have a decrease in home values in the next 3 months**
2. If you are only concerned about the largest home value increase (by percentage) in the next 12 months, which city would you choose to relocate to?  
   **New York is predicted to have the largest home value decrease in the next twelve months, so I would move there if I was concerned with a home value increase. However, if for some reason, I wanted to move to a city with the “largest’ home value ‘increase’, I would move to Houston since it is seeing the lowest home value decrease in the next twelve months.**