# MIS503 - Final Project

## Fry, Colby

### Zillow Home Value Index Analysis

### Wake County Home Sales

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

**Home values have increased every year throughout the past decade, with a considerable spike from 2021 and 2022, 2023 is slightly down from the prior year**

1. There were dips in home values in the past 10 years. What years did these occur?

**The only home value dip for the years in this dataset occured from 2022 to 2023**

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

**The most expensive area would be Cary, the least expensive would be Zebulon**

1. What has happened to the overall property values in Apex and Cary in 2023?

**The overall property values of Apex and Cary in 2023 have remained mostly the same as the previous year, though slightly lower. Not much has changed for those two regions, both are still around the $800,000 mark.**

library(tidyverse)

## Warning: package 'purrr' was built under R version 4.3.3

## ── 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 3.5.0 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.4   
## ── 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

library(lubridate)  
library(scales)

## Warning: package 'scales' was built under R version 4.3.3

##   
## Attaching package: 'scales'  
##   
## The following object is masked from 'package:purrr':  
##   
## discard  
##   
## The following object is masked from 'package:readr':  
##   
## col\_factor

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

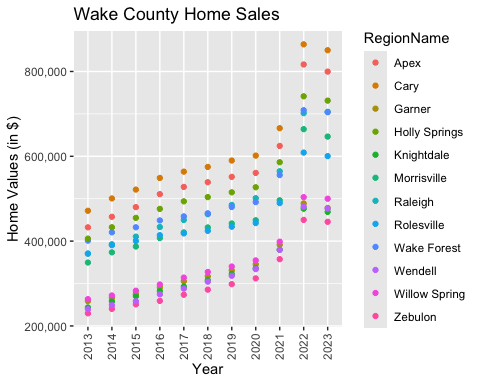
## Rows: 3503 Columns: 107  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (4): RegionName, State, Metro, CountyName  
## dbl (103): RegionID, 1/31/2015, 2/28/2015, 3/31/2015, 4/30/2015, 5/31/2015, ...  
##   
## ℹ Use `spec()` to retrieve 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")

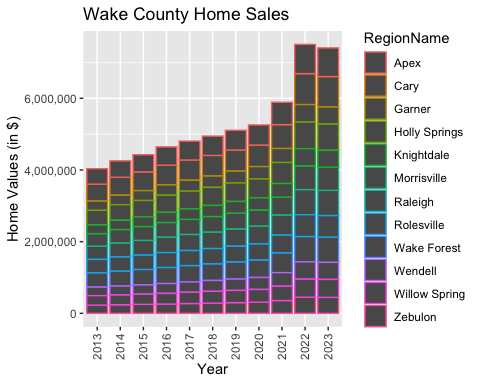
## Rows: 22275 Columns: 287  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (4): RegionName, State, Metro, CountyName  
## dbl (283): RegionID, 1/31/2000, 2/29/2000, 3/31/2000, 4/30/2000, 5/31/2000, ...  
##   
## ℹ Use `spec()` to retrieve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

WakeCountySales <- SingleFamilyResidenceSales %>%  
 filter(State == "NC", CountyName == "Wake 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")

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



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



### NC Rental Market

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 overall trend for rental costs in NC is that the cost to rent a home/apartment has consistently risen every year, from 2015 to 2023**

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

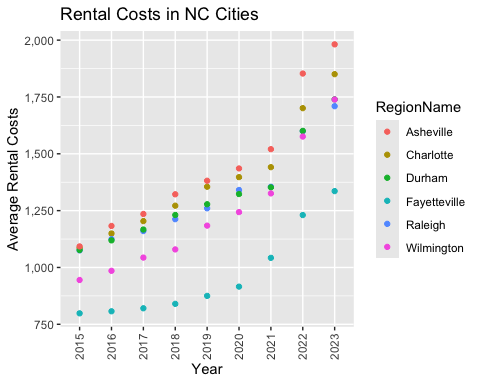
**The most expensive city to rent in NC would be Asheville, with the most affordable city being Fayetteville**

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

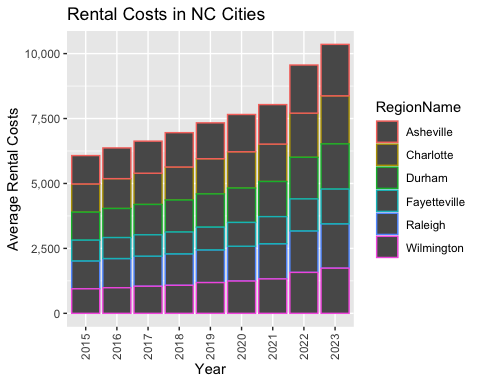
**Wilmington’s rent prices are overall cheaper than Asheville’s, considerably.**

needed\_cities <- c("Asheville", "Charlotte", "Durham", "Fayetteville", "Raleigh", "Wilmington")  
  
Rentals <- SingleFamilyResidenceRental %>%  
 filter(RegionName %in% needed\_cities, 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(col = `2015`:`2023`, names\_to = "YR", values\_to = "ZHVI")

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



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



### Home Values in Select Markets

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

**Charlotte-Concord-Gastonia**

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?

**Asheville (possibly Charlotte-Concord-Gastonia?)**

1. 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)?

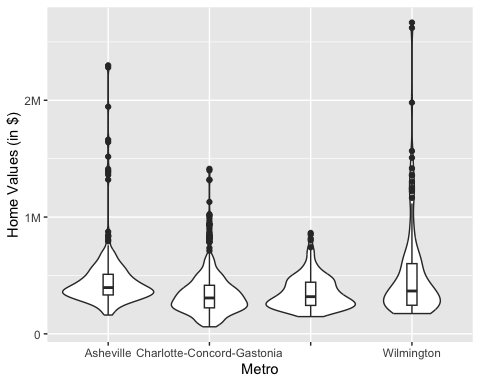
**Wilmington**

four\_metro <- c("Asheville", "Charlotte-Concord-Gastonia", "Raleigh-Cary", "Wilmington")  
NCHomeSales <- SingleFamilyResidenceSales %>%  
 filter(Metro %in% four\_metro, 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(col = `2013`:`2023`, names\_to = "YR", values\_to = "ZHVI") %>%  
 group\_by(Metro)

metro\_violin <- ggplot(NCHomeSales, aes(x = Metro, y = ZHVI)) +  
 geom\_violin() +  
 geom\_boxplot(width = 0.1) +  
 labs(  
 x = "Metro",  
 y = "Home Values (in $)"  
 ) +  
 scale\_y\_continuous(labels = label\_number(scale\_cut = cut\_short\_scale())) +  
 scale\_x\_discrete(guide = guide\_axis(check.overlap = TRUE))  
  
metro\_violin

## Warning: Removed 18 rows containing non-finite outside the scale range  
## (`stat\_ydensity()`).

## Warning: Removed 18 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).



### Relocation Home Value Comparison

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

**Houston, TX has the most affordable homes, with the least affordable homes residing in New York City**

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)?

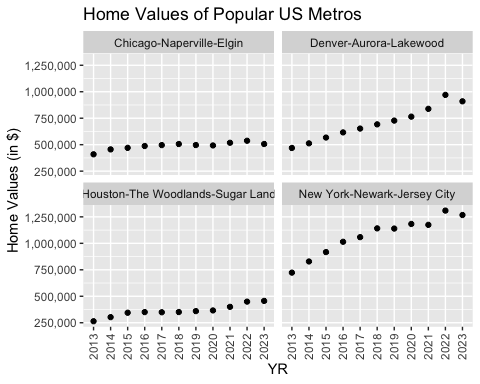
**Denver, CO appears to have seen the more significant changes in prices over the past five years, Chicago, IL has remained the most consistent of the four cities**

1. Which cities saw a decline in value during 2023 and which cities remained consistent?

**Chicago, New York, and Denver all saw declines from 2022 to 2023. The only city that stayed consistent was Houston.**

relocated\_cities <- c("Chicago", "Denver", "Houston", "New York")  
relocated\_counties <- c("Harris County", "Denver County", "Cook County", "Bronx County", "Kings County", "Richmond County", "Queens County", "New York County")  
NationalHomeSales <- SingleFamilyResidenceSales %>%  
 filter(RegionName %in% relocated\_cities, CountyName %in% relocated\_counties) %>%  
 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(col = `2013`:`2023`, names\_to = "YR", values\_to = "ZHVI") %>%  
 filter(!is.na(Metro))

(relocation\_scatter <- ggplot(NationalHomeSales, aes(x = YR, y = ZHVI)) +  
 geom\_point() +  
 facet\_wrap(~Metro)) +  
 labs( x = "YR",  
 y = "Home Values (in $)",  
 title = "Home Values of Popular US Metros") +  
 theme(axis.text.x = element\_text(angle = 90, vjust=0.5)) +  
 scale\_y\_continuous(name="Home Values (in $)", labels = scales::comma)

 ### Future Home Values

1. Which is the only city that is projected to have a decrease in home values in the next 3 months?

**Based on the dataset we were requested to retrieve on Zillow, every city is expected to have a decrease in the next 3 months. However, Chicago has the largest decrease out of the cities.**

1. 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?

**Every city is expected to decrease in the next 12 months, not increase. However, Houston is expected to decrease in value the least out of the four cities, with a little under a 2 percent decrease**

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

## Rows: 895 Columns: 9  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (4): RegionName, RegionType, StateName, BaseDate  
## dbl (5): RegionID, SizeRank, 11/30/23, 1/31/24, 10/31/24  
##   
## 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.

forecast\_states <- c("IL", "CO", "TX", "NY")  
forecast\_cities <- c("Chicago, IL", "Denver, CO", "Houston, TX", "New York, NY")  
FutureHomeValues <- Projections %>%  
 filter(RegionName %in% forecast\_cities, StateName %in% forecast\_states) %>%  
 select(RegionName, "11/30/23", "1/31/24", "10/31/24") %>%  
 rename("Current" = "11/30/23", "ThreeMonths" = "1/31/24", "TwelveMonths" = "10/31/24") %>%  
 pivot\_longer(col = c(`Current`, `ThreeMonths`, `TwelveMonths`), names\_to = "Time", values\_to = "PercentageChange")

(forecast\_scatter <- ggplot(FutureHomeValues, aes(x = Time, y = PercentageChange , color = RegionName)) +  
 geom\_point())

