# MIS503 – Final Project

## Warrene, Audrey

### Zillow Home Value Index Analysis

Exercise 1 —-

### Wake County Home Sales

1. What have been the overall trends in Wake County Home Values? – Increasing, substantially in 2022-2023
2. There were dips in home values in the past 10 years. What years did these occur? – The most obvious dip in prices is from 2022-2023, which we can see on both the bar chart and the scatterplot. Otherwise, there is not an obvious drop in prices on the charts.
3. Based on the analysis, where would be the least expensive area to purchase home? Most expensive area? – Least Expensive: Zebulon. Most Expensive: Cary.
4. What has happened to the overall property values in Apex and Cary in 2023? – It has slightly gone down since 2022, but is still consistently staying high on the chart. Cary is still more expensive than Apex as well.

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 3.5.0 ✔ 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.

WakeCountySales <- select(SingleFamilyResidenceSales, 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")  
WakeCountySales <- filter(WakeCountySales, State=="NC", CountyName=="Wake County")  
  
WakeCountySales

## # A tibble: 12 × 15  
## RegionName State CountyName Metro `5/31/2013` `5/31/2014` `5/31/2015`  
## <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 Raleigh NC Wake County Raleigh-… 370453. 392624. 410637.  
## 2 Cary NC Wake County Raleigh-… 471561. 500631. 521536.  
## 3 Apex NC Wake County Raleigh-… 432532. 457372. 480114.  
## 4 Wake Forest NC Wake County Raleigh-… 401073. 420710. 432643.  
## 5 Garner NC Wake County Raleigh-… 258592. 266658. 275585.  
## 6 Holly Springs NC Wake County Raleigh-… 405887. 432719. 454887.  
## 7 Morrisville NC Wake County Raleigh-… 349518. 373493. 387268.  
## 8 Knightdale NC Wake County Raleigh-… 243654. 258379. 270110.  
## 9 Zebulon NC Wake County Raleigh-… 229737. 240266. 250907.  
## 10 Wendell NC Wake County Raleigh-… 239614. 249043. 258299.  
## 11 Willow Spring NC Wake County Raleigh-… 263189. 271888. 283073.  
## 12 Rolesville NC Wake County Raleigh-… 370686. 390404. 400318.  
## # ℹ 8 more variables: `5/31/2016` <dbl>, `5/31/2017` <dbl>, `5/31/2018` <dbl>,  
## # `5/31/2019` <dbl>, `5/31/2020` <dbl>, `5/31/2021` <dbl>, `5/31/2022` <dbl>,  
## # `5/31/2023` <dbl>

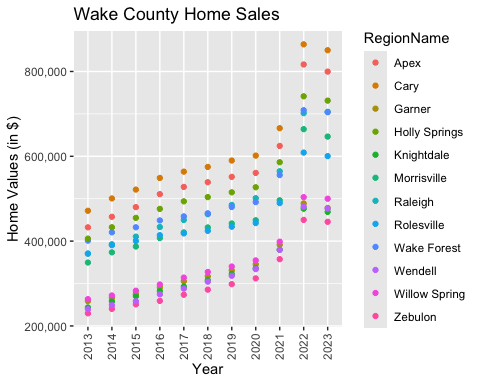
WakeCountySales <- rename(WakeCountySales, "2013" = "5/31/2013")  
WakeCountySales <- rename(WakeCountySales, "2014" = "5/31/2014")  
WakeCountySales <- rename(WakeCountySales, "2015" = "5/31/2015")  
WakeCountySales <- rename(WakeCountySales, "2016" = "5/31/2016")  
WakeCountySales <- rename(WakeCountySales, "2017" = "5/31/2017")  
WakeCountySales <- rename(WakeCountySales, "2018" = "5/31/2018")  
WakeCountySales <- rename(WakeCountySales, "2019" = "5/31/2019")  
WakeCountySales <- rename(WakeCountySales, "2020" = "5/31/2020")  
WakeCountySales <- rename(WakeCountySales, "2021" = "5/31/2021")  
WakeCountySales <- rename(WakeCountySales, "2022" = "5/31/2022")  
WakeCountySales <- rename(WakeCountySales, "2023" = "5/31/2023")  
  
WakeCountySales

## # A tibble: 12 × 15  
## RegionName State CountyName Metro `2013` `2014` `2015` `2016` `2017` `2018`  
## <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Raleigh NC Wake Coun… Rale… 3.70e5 3.93e5 4.11e5 4.33e5 4.50e5 4.63e5  
## 2 Cary NC Wake Coun… Rale… 4.72e5 5.01e5 5.22e5 5.49e5 5.64e5 5.75e5  
## 3 Apex NC Wake Coun… Rale… 4.33e5 4.57e5 4.80e5 5.11e5 5.28e5 5.39e5  
## 4 Wake Forest NC Wake Coun… Rale… 4.01e5 4.21e5 4.33e5 4.49e5 4.58e5 4.66e5  
## 5 Garner NC Wake Coun… Rale… 2.59e5 2.67e5 2.76e5 2.91e5 3.05e5 3.17e5  
## 6 Holly Sprin… NC Wake Coun… Rale… 4.06e5 4.33e5 4.55e5 4.76e5 4.94e5 5.04e5  
## 7 Morrisville NC Wake Coun… Rale… 3.50e5 3.73e5 3.87e5 4.07e5 4.21e5 4.32e5  
## 8 Knightdale NC Wake Coun… Rale… 2.44e5 2.58e5 2.70e5 2.82e5 2.93e5 3.08e5  
## 9 Zebulon NC Wake Coun… Rale… 2.30e5 2.40e5 2.51e5 2.59e5 2.74e5 2.86e5  
## 10 Wendell NC Wake Coun… Rale… 2.40e5 2.49e5 2.58e5 2.75e5 2.88e5 3.04e5  
## 11 Willow Spri… NC Wake Coun… Rale… 2.63e5 2.72e5 2.83e5 2.98e5 3.14e5 3.27e5  
## 12 Rolesville NC Wake Coun… Rale… 3.71e5 3.90e5 4.00e5 4.14e5 4.18e5 4.24e5  
## # ℹ 5 more variables: `2019` <dbl>, `2020` <dbl>, `2021` <dbl>, `2022` <dbl>,  
## # `2023` <dbl>

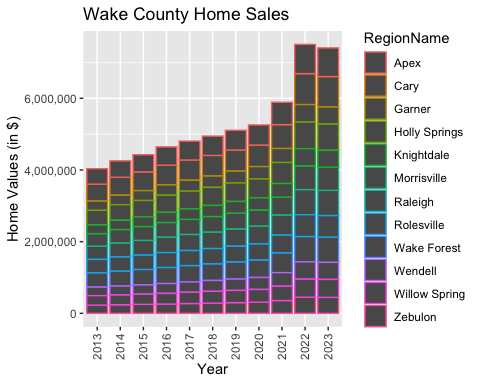
WakeCountySales <- WakeCountySales %>% pivot\_longer(c('2013','2014', '2015', '2016', '2017', '2018', '2019', '2020', '2021', '2022', '2023'),names\_to='YR',values\_to='ZHVI')  
  
WakeCountySales

## # A tibble: 132 × 6  
## RegionName State CountyName Metro YR ZHVI  
## <chr> <chr> <chr> <chr> <chr> <dbl>  
## 1 Raleigh NC Wake County Raleigh-Cary 2013 370453.  
## 2 Raleigh NC Wake County Raleigh-Cary 2014 392624.  
## 3 Raleigh NC Wake County Raleigh-Cary 2015 410637.  
## 4 Raleigh NC Wake County Raleigh-Cary 2016 433227.  
## 5 Raleigh NC Wake County Raleigh-Cary 2017 449954.  
## 6 Raleigh NC Wake County Raleigh-Cary 2018 463451.  
## 7 Raleigh NC Wake County Raleigh-Cary 2019 485307.  
## 8 Raleigh NC Wake County Raleigh-Cary 2020 501162.  
## 9 Raleigh NC Wake County Raleigh-Cary 2021 564715.  
## 10 Raleigh NC Wake County Raleigh-Cary 2022 701662.  
## # ℹ 122 more rows

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



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



Exercise 2 —-

### 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 is an increase in rental price over the years, especially in 2022 and 2023. All cities have followed this trend, but Fayetteville is following it at a much smaller pace.
2. Where is the most expensive city to rent in? Least expensive? – Most Expensive: Asheville. Least Expensive: Fayetteville.
3. You are trying to decide between Wilmington and Asheville. Which market has the lowest rent? – Wilmington has the lower rent between the two.

Rentals <- select(SingleFamilyResidenceRental, 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")  
Rentals <- filter(Rentals, State=="NC", RegionName %in% c("Asheville", "Charlotte", "Durham", "Fayetteville", "Raleigh", "Wilmington"))  
  
Rentals

## # A tibble: 6 × 11  
## RegionName State `1/31/2015` `1/31/2016` `1/31/2017` `1/31/2018` `1/31/2019`  
## <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Charlotte NC 1086. 1149. 1204. 1272. 1355.  
## 2 Raleigh NC 1075. 1125. 1160. 1212. 1260.  
## 3 Durham NC 1077. 1119. 1167. 1231. 1278.  
## 4 Fayetteville NC 798. 807. 820. 840. 875.  
## 5 Wilmington NC 945. 985. 1043. 1079. 1184.  
## 6 Asheville NC 1093. 1183. 1235. 1322. 1381.  
## # ℹ 4 more variables: `1/31/2020` <dbl>, `1/31/2021` <dbl>, `1/31/2022` <dbl>,  
## # `1/31/2023` <dbl>

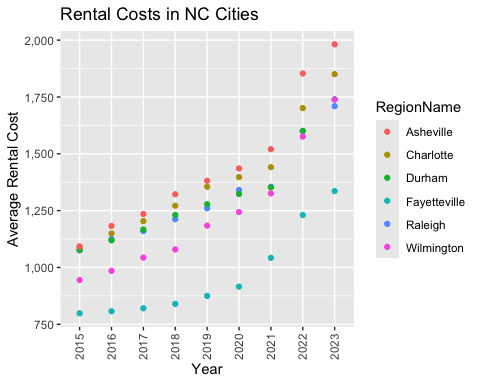
Rentals <- rename(Rentals, "2015" = "1/31/2015")  
Rentals <- rename(Rentals, "2016" = "1/31/2016")  
Rentals <- rename(Rentals, "2017" = "1/31/2017")  
Rentals <- rename(Rentals, "2018" = "1/31/2018")  
Rentals <- rename(Rentals, "2019" = "1/31/2019")  
Rentals <- rename(Rentals, "2020" = "1/31/2020")  
Rentals <- rename(Rentals, "2021" = "1/31/2021")  
Rentals <- rename(Rentals, "2022" = "1/31/2022")  
Rentals <- rename(Rentals, "2023" = "1/31/2023")  
  
Rentals

## # A tibble: 6 × 11  
## RegionName State `2015` `2016` `2017` `2018` `2019` `2020` `2021` `2022`  
## <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Charlotte NC 1086. 1149. 1204. 1272. 1355. 1397. 1441. 1701.  
## 2 Raleigh NC 1075. 1125. 1160. 1212. 1260. 1341. 1355. 1600.  
## 3 Durham NC 1077. 1119. 1167. 1231. 1278. 1323. 1352. 1601.  
## 4 Fayetteville NC 798. 807. 820. 840. 875. 916. 1042. 1231.  
## 5 Wilmington NC 945. 985. 1043. 1079. 1184. 1244. 1326. 1576.  
## 6 Asheville NC 1093. 1183. 1235. 1322. 1381. 1436. 1520. 1853.  
## # ℹ 1 more variable: `2023` <dbl>

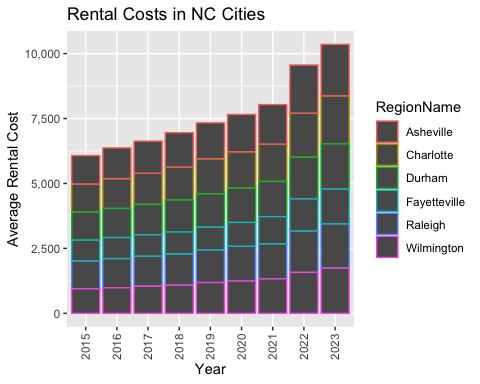
Rentals <- Rentals %>% pivot\_longer(c('2015', '2016', '2017', '2018', '2019', '2020', '2021', '2022', '2023'),names\_to='YR',values\_to='ZHVI')  
  
Rentals

## # A tibble: 54 × 4  
## RegionName State YR ZHVI  
## <chr> <chr> <chr> <dbl>  
## 1 Charlotte NC 2015 1086.  
## 2 Charlotte NC 2016 1149.  
## 3 Charlotte NC 2017 1204.  
## 4 Charlotte NC 2018 1272.  
## 5 Charlotte NC 2019 1355.  
## 6 Charlotte NC 2020 1397.  
## 7 Charlotte NC 2021 1441.  
## 8 Charlotte NC 2022 1701.  
## 9 Charlotte NC 2023 1850.  
## 10 Raleigh NC 2015 1075.  
## # ℹ 44 more rows

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



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



Exercise 3 —-

### 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
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? – Asheville
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

NCHomeSales <- select(SingleFamilyResidenceSales, 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")  
  
NCHomeSales <- filter(NCHomeSales, State=="NC", Metro %in% c("Asheville", "Charlotte-Concord-Gastonia", "Raleigh-Cary", "Wilmington"))  
  
NCHomeSales

## # A tibble: 144 × 14  
## RegionName State Metro `5/31/2013` `5/31/2014` `5/31/2015` `5/31/2016`  
## <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Charlotte NC Charlotte-… 294616. 320763. 339470. 364399.  
## 2 Raleigh NC Raleigh-Ca… 370453. 392624. 410637. 433227.  
## 3 Wilmington NC Wilmington 320159. 339868. 348600. 363770.  
## 4 Cary NC Raleigh-Ca… 471561. 500631. 521536. 548965.  
## 5 Concord NC Charlotte-… 249646. 262916. 277254. 294959.  
## 6 Asheville NC Asheville 339119. 361407. 392763. 420754.  
## 7 Gastonia NC Charlotte-… 184917. 191386. 195051. 205114.  
## 8 Apex NC Raleigh-Ca… 432532. 457372. 480114. 510721.  
## 9 Mooresville NC Charlotte-… 405861. 429771. 452840. 474414.  
## 10 Wake Forest NC Raleigh-Ca… 401073. 420710. 432643. 448840.  
## # ℹ 134 more rows  
## # ℹ 7 more variables: `5/31/2017` <dbl>, `5/31/2018` <dbl>, `5/31/2019` <dbl>,  
## # `5/31/2020` <dbl>, `5/31/2021` <dbl>, `5/31/2022` <dbl>, `5/31/2023` <dbl>

NCHomeSales <- rename(NCHomeSales, "2013" = "5/31/2013")  
NCHomeSales <- rename(NCHomeSales, "2014" = "5/31/2014")  
NCHomeSales <- rename(NCHomeSales, "2015" = "5/31/2015")  
NCHomeSales <- rename(NCHomeSales, "2016" = "5/31/2016")  
NCHomeSales <- rename(NCHomeSales, "2017" = "5/31/2017")  
NCHomeSales <- rename(NCHomeSales, "2018" = "5/31/2018")  
NCHomeSales <- rename(NCHomeSales, "2019" = "5/31/2019")  
NCHomeSales <- rename(NCHomeSales, "2020" = "5/31/2020")  
NCHomeSales <- rename(NCHomeSales, "2021" = "5/31/2021")  
NCHomeSales <- rename(NCHomeSales, "2022" = "5/31/2022")  
NCHomeSales <- rename(NCHomeSales, "2023" = "5/31/2023")  
  
NCHomeSales

## # A tibble: 144 × 14  
## RegionName State Metro `2013` `2014` `2015` `2016` `2017` `2018` `2019`  
## <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Charlotte NC Charlotte… 2.95e5 3.21e5 3.39e5 3.64e5 3.87e5 4.08e5 4.22e5  
## 2 Raleigh NC Raleigh-C… 3.70e5 3.93e5 4.11e5 4.33e5 4.50e5 4.63e5 4.85e5  
## 3 Wilmington NC Wilmington 3.20e5 3.40e5 3.49e5 3.64e5 3.76e5 3.89e5 4.06e5  
## 4 Cary NC Raleigh-C… 4.72e5 5.01e5 5.22e5 5.49e5 5.64e5 5.75e5 5.90e5  
## 5 Concord NC Charlotte… 2.50e5 2.63e5 2.77e5 2.95e5 3.13e5 3.28e5 3.40e5  
## 6 Asheville NC Asheville 3.39e5 3.61e5 3.93e5 4.21e5 4.42e5 4.72e5 4.90e5  
## 7 Gastonia NC Charlotte… 1.85e5 1.91e5 1.95e5 2.05e5 2.22e5 2.37e5 2.55e5  
## 8 Apex NC Raleigh-C… 4.33e5 4.57e5 4.80e5 5.11e5 5.28e5 5.39e5 5.52e5  
## 9 Mooresville NC Charlotte… 4.06e5 4.30e5 4.53e5 4.74e5 4.97e5 5.10e5 5.26e5  
## 10 Wake Forest NC Raleigh-C… 4.01e5 4.21e5 4.33e5 4.49e5 4.58e5 4.66e5 4.81e5  
## # ℹ 134 more rows  
## # ℹ 4 more variables: `2020` <dbl>, `2021` <dbl>, `2022` <dbl>, `2023` <dbl>

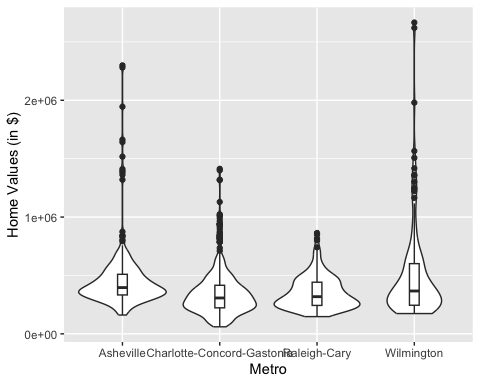
NCHomeSales <- NCHomeSales %>% pivot\_longer(c('2013','2014', '2015', '2016', '2017', '2018', '2019', '2020', '2021', '2022', '2023'),names\_to='YR',values\_to='ZHVI')  
  
NCHomeSales <- arrange(NCHomeSales, Metro)  
  
NCHomeSales

## # A tibble: 1,584 × 5  
## RegionName State Metro YR ZHVI  
## <chr> <chr> <chr> <chr> <dbl>  
## 1 Asheville NC Asheville 2013 339119.  
## 2 Asheville NC Asheville 2014 361407.  
## 3 Asheville NC Asheville 2015 392763.  
## 4 Asheville NC Asheville 2016 420754.  
## 5 Asheville NC Asheville 2017 441739.  
## 6 Asheville NC Asheville 2018 472270.  
## 7 Asheville NC Asheville 2019 489722.  
## 8 Asheville NC Asheville 2020 508237.  
## 9 Asheville NC Asheville 2021 584840.  
## 10 Asheville NC Asheville 2022 697533.  
## # ℹ 1,574 more rows

ggplot(NCHomeSales, aes(x = Metro, y = ZHVI)) +  
 geom\_violin() +  
 geom\_boxplot(width = 0.1) +  
 labs(x = "Metro",   
 y = "Home Values (in $)")

## 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()`).



Exercise 4 —-

### Relocation Home Value Comparison

1. Based on your analysis, which city’s housing is most affordable? Least affordable? – Most Affordable: Houston. Least Affordable: 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)? – The city with the largest changes in prices is Denver. The city that has remained more consistent is Chicago, because it has consistently remained at 500,000 dollars.
3. Which cities saw a decline in value during 2023 and which cities remained consistent? – Decline: New York, Denver, Technically Chicago. The cities remaining most consistent is Houston and technically Chicago, because even though the value did drop, it BARELY dropped.

NationalHomeSales <- select(SingleFamilyResidenceSales, 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")  
  
cities <- c("Chicago", "Denver", "Houston", "New York")  
counties <- c("Cook County", "Denver County", "Harris County", "Queens County")  
  
NationalHomeSales <- filter(NationalHomeSales, RegionName %in% cities, CountyName %in% counties)  
  
NationalHomeSales

## # A tibble: 4 × 15  
## RegionName State CountyName Metro `5/31/2013` `5/31/2014` `5/31/2015`  
## <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 New York NY Queens County New York-N… 723160. 828522. 917431.  
## 2 Houston TX Harris County Houston-Th… 264062. 302672. 344415.  
## 3 Chicago IL Cook County Chicago-Na… 409180. 455543. 470035.  
## 4 Denver CO Denver County Denver-Aur… 468904. 512596. 567297.  
## # ℹ 8 more variables: `5/31/2016` <dbl>, `5/31/2017` <dbl>, `5/31/2018` <dbl>,  
## # `5/31/2019` <dbl>, `5/31/2020` <dbl>, `5/31/2021` <dbl>, `5/31/2022` <dbl>,  
## # `5/31/2023` <dbl>

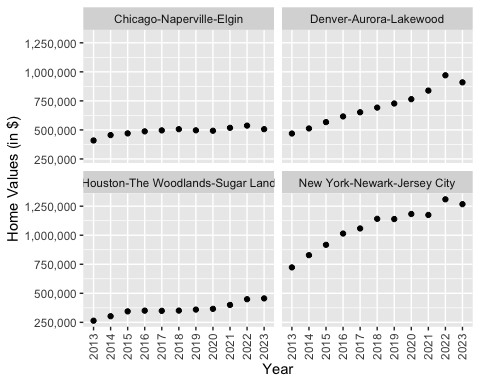
NationalHomeSales <- rename(NationalHomeSales, "2013" = "5/31/2013")  
NationalHomeSales <- rename(NationalHomeSales, "2014" = "5/31/2014")  
NationalHomeSales <- rename(NationalHomeSales, "2015" = "5/31/2015")  
NationalHomeSales <- rename(NationalHomeSales, "2016" = "5/31/2016")  
NationalHomeSales <- rename(NationalHomeSales, "2017" = "5/31/2017")  
NationalHomeSales <- rename(NationalHomeSales, "2018" = "5/31/2018")  
NationalHomeSales <- rename(NationalHomeSales, "2019" = "5/31/2019")  
NationalHomeSales <- rename(NationalHomeSales, "2020" = "5/31/2020")  
NationalHomeSales <- rename(NationalHomeSales, "2021" = "5/31/2021")  
NationalHomeSales <- rename(NationalHomeSales, "2022" = "5/31/2022")  
NationalHomeSales <- rename(NationalHomeSales, "2023" = "5/31/2023")  
  
NationalHomeSales

## # A tibble: 4 × 15  
## RegionName State CountyName Metro `2013` `2014` `2015` `2016` `2017` `2018`  
## <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 New York NY Queens County New … 7.23e5 8.29e5 9.17e5 1.01e6 1.06e6 1.14e6  
## 2 Houston TX Harris County Hous… 2.64e5 3.03e5 3.44e5 3.51e5 3.49e5 3.51e5  
## 3 Chicago IL Cook County Chic… 4.09e5 4.56e5 4.70e5 4.88e5 4.96e5 5.06e5  
## 4 Denver CO Denver County Denv… 4.69e5 5.13e5 5.67e5 6.16e5 6.52e5 6.92e5  
## # ℹ 5 more variables: `2019` <dbl>, `2020` <dbl>, `2021` <dbl>, `2022` <dbl>,  
## # `2023` <dbl>

NationalHomeSales <- NationalHomeSales %>% pivot\_longer(c('2013','2014', '2015', '2016', '2017', '2018', '2019', '2020', '2021', '2022', '2023'),names\_to='YR',values\_to='ZHVI')  
  
NationalHomeSales

## # A tibble: 44 × 6  
## RegionName State CountyName Metro YR ZHVI  
## <chr> <chr> <chr> <chr> <chr> <dbl>  
## 1 New York NY Queens County New York-Newark-Jersey City 2013 723160.  
## 2 New York NY Queens County New York-Newark-Jersey City 2014 828522.  
## 3 New York NY Queens County New York-Newark-Jersey City 2015 917431.  
## 4 New York NY Queens County New York-Newark-Jersey City 2016 1014488.  
## 5 New York NY Queens County New York-Newark-Jersey City 2017 1058360.  
## 6 New York NY Queens County New York-Newark-Jersey City 2018 1141261.  
## 7 New York NY Queens County New York-Newark-Jersey City 2019 1139390.  
## 8 New York NY Queens County New York-Newark-Jersey City 2020 1182898.  
## 9 New York NY Queens County New York-Newark-Jersey City 2021 1174606.  
## 10 New York NY Queens County New York-Newark-Jersey City 2022 1309176.  
## # ℹ 34 more rows

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



Exercise 5 —-

### Future Home Values

1. Which is the only city that is projected to have a decrease in home values in the next 3 months? – None of them really dropped in home value, but Hoston, TX maybe dropped? (or just remains the same)
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? – Chicaco, IL has the best percentage after 12 months.

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

## Rows: 895 Columns: 9  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (3): RegionName, RegionType, StateName  
## dbl (5): RegionID, SizeRank, 2024-04-30, 2024-06-30, 2025-03-31  
## 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 <- select(Projections, RegionName, "2024-04-30", "2024-06-30", "2025-03-31")  
  
FutureHomeValues

## # A tibble: 895 × 4  
## RegionName `2024-04-30` `2024-06-30` `2025-03-31`  
## <chr> <dbl> <dbl> <dbl>  
## 1 United States 0.4 0.7 1.4  
## 2 New York, NY 0.7 1.1 0   
## 3 Los Angeles, CA 0.3 0.7 1.2  
## 4 Chicago, IL 0.7 1.3 0.4  
## 5 Dallas, TX 0.2 0.5 1.3  
## 6 Houston, TX 0.2 0.2 0.2  
## 7 Washington, DC 0.6 1 -0.1  
## 8 Philadelphia, PA 0.6 1.1 1.6  
## 9 Miami, FL 0.3 0.6 2.9  
## 10 Atlanta, GA 0.4 0.9 3   
## # ℹ 885 more rows

cities <- c("Chicago, IL", "Denver, CO", "Houston, TX", "New York, NY")  
  
FutureHomeValues <- filter(FutureHomeValues, RegionName %in% cities)  
  
FutureHomeValues

## # A tibble: 4 × 4  
## RegionName `2024-04-30` `2024-06-30` `2025-03-31`  
## <chr> <dbl> <dbl> <dbl>  
## 1 New York, NY 0.7 1.1 0   
## 2 Chicago, IL 0.7 1.3 0.4  
## 3 Houston, TX 0.2 0.2 0.2  
## 4 Denver, CO 0.2 0.4 0.1

FutureHomeValues <- rename(FutureHomeValues, "Current" = "2024-04-30")  
FutureHomeValues <- rename(FutureHomeValues, "ThreeMonths" = "2024-06-30")  
FutureHomeValues <- rename(FutureHomeValues, "TwelveMonths" = "2025-03-31")  
  
FutureHomeValues

## # A tibble: 4 × 4  
## RegionName Current ThreeMonths TwelveMonths  
## <chr> <dbl> <dbl> <dbl>  
## 1 New York, NY 0.7 1.1 0   
## 2 Chicago, IL 0.7 1.3 0.4  
## 3 Houston, TX 0.2 0.2 0.2  
## 4 Denver, CO 0.2 0.4 0.1

FutureHomeValues <- FutureHomeValues %>% pivot\_longer(c('Current', 'ThreeMonths', 'TwelveMonths'),names\_to='Time',values\_to='PercentageChange')  
  
FutureHomeValues

## # A tibble: 12 × 3  
## RegionName Time PercentageChange  
## <chr> <chr> <dbl>  
## 1 New York, NY Current 0.7  
## 2 New York, NY ThreeMonths 1.1  
## 3 New York, NY TwelveMonths 0   
## 4 Chicago, IL Current 0.7  
## 5 Chicago, IL ThreeMonths 1.3  
## 6 Chicago, IL TwelveMonths 0.4  
## 7 Houston, TX Current 0.2  
## 8 Houston, TX ThreeMonths 0.2  
## 9 Houston, TX TwelveMonths 0.2  
## 10 Denver, CO Current 0.2  
## 11 Denver, CO ThreeMonths 0.4  
## 12 Denver, CO TwelveMonths 0.1

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

