MIS503 - Final Project

Grace Williams

2023-04-24

# MIS503 - Final Project

## Williams, Grace

### Zillow Home Value Index Analysis

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.2 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.2 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.1   
## ── 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(readr)  
library(readxl)  
library(lubridate)  
library(dplyr)  
library(tidyr)  
library(ggplot2)  
library(shiny)  
  
SingleFamRental <- read\_csv("SingleFamilyResidenceRental.csv")

## Rows: 13273 Columns: 102  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (4): RegionName, State, Metro, CountyName  
## dbl (98): RegionID, SizeRank, 2010-11, 2010-12, 2011-01, 2011-02, 2011-03, 2...  
##   
## 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.

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

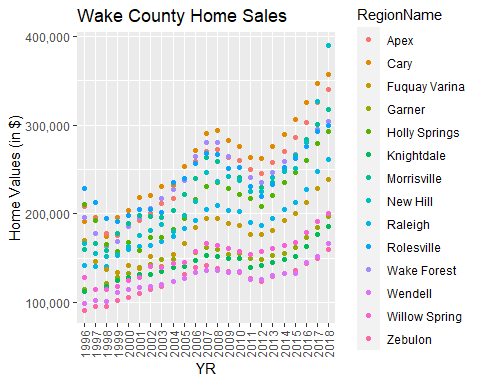
## Rows: 12797 Columns: 277  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (4): RegionName, State, Metro, CountyName  
## dbl (273): RegionID, SizeRank, 1996-04, 1996-05, 1996-06, 1996-07, 1996-08, ...  
##   
## ℹ Use `spec()` to retrieve 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

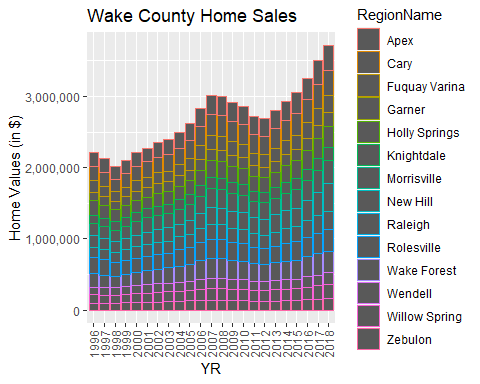
WakeCountySales <- data.frame(SingleFamSales$RegionName, SingleFamSales$State, SingleFamSales$CountyName, SingleFamSales$`1996-05` , SingleFamSales$`1997-05`, SingleFamSales$`1998-05`, SingleFamSales$`1999-05`, SingleFamSales$`2000-05`, SingleFamSales$`2001-05`, SingleFamSales$`2002-05`, SingleFamSales$`2003-05`, SingleFamSales$`2004-05`, SingleFamSales$`2005-05`, SingleFamSales$`2006-05`, SingleFamSales$`2007-05`, SingleFamSales$`2008-05`, SingleFamSales$`2009-05`, SingleFamSales$`2010-05`, SingleFamSales$`2011-05`, SingleFamSales$`2012-05`, SingleFamSales$`2013-05`, SingleFamSales$`2014-05`, SingleFamSales$`2015-05`, SingleFamSales$`2016-05`, SingleFamSales$`2017-05`, SingleFamSales$`2018-05`)%>%  
 filter(SingleFamSales$State == "NC", SingleFamSales$CountyName == "Wake County")  
  
#Rename Columns  
  
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales.RegionName'] <- "RegionName"  
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales.State'] <- "State"  
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales.CountyName'] <- "CountyName"  
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..1996.05.'] <- "1996"   
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..1997.05.'] <- "1997"   
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..1998.05.'] <- "1998"   
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..1999.05.'] <- "1999"   
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..2000.05.'] <- "2000"   
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..2001.05.'] <- "2001"   
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..2002.05.'] <- "2002"   
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..2003.05.'] <- "2003"   
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..2004.05.'] <- "2004"   
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..2005.05.'] <- "2005"   
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..2006.05.'] <- "2006"   
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..2007.05.'] <- "2007"   
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..2008.05.'] <- "2008"   
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..2009.05.'] <- "2009"   
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..2010.05.'] <- "2010"   
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..2011.05.'] <- "2011"   
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..2012.05.'] <- "2012"   
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..2013.05.'] <- "2013"   
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..2014.05.'] <- "2014"   
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..2015.05.'] <- "2015"   
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..2016.05.'] <- "2016"   
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..2017.05.'] <- "2017"   
names(WakeCountySales)[names(WakeCountySales) == 'SingleFamSales..2018.05.'] <- "2018"   
  
#Clean it up  
  
WakeCountySales %>%  
 group\_by(RegionName)

## # A tibble: 14 × 26  
## # Groups: RegionName [14]  
## RegionName State CountyName `1996` `1997` `1998` `1999` `2000` `2001` `2002`  
## <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Raleigh NC Wake Coun… 141500 140900 141000 153800 160600 163200 165000  
## 2 Cary NC Wake Coun… 191900 195700 174900 195900 203600 218500 220700  
## 3 Apex NC Wake Coun… 208700 195000 177500 175400 189000 192700 201000  
## 4 Wake Forest NC Wake Coun… 195500 178000 162900 169500 185400 195900 205700  
## 5 Garner NC Wake Coun… 115200 115000 121700 129000 132800 140300 143700  
## 6 Fuquay Var… NC Wake Coun… 169900 146700 137100 133900 141600 139000 152100  
## 7 Holly Spri… NC Wake Coun… 210500 193000 165600 157300 168400 158700 174000  
## 8 Morrisville NC Wake Coun… 159900 166700 159400 177700 189100 198200 196700  
## 9 Knightdale NC Wake Coun… 113200 114700 118200 125700 128700 131900 131600  
## 10 Zebulon NC Wake Coun… 91700 96100 95500 102500 105800 110900 114500  
## 11 Wendell NC Wake Coun… 99200 103200 101900 112100 115400 117400 117900  
## 12 Willow Spr… NC Wake Coun… 128300 115400 115600 118800 125200 128700 140500  
## 13 Rolesville NC Wake Coun… 228700 212600 194800 191600 198700 205600 204900  
## 14 New Hill NC Wake Coun… 166400 155500 152600 154400 163300 175400 181000  
## # ℹ 16 more variables: `2003` <dbl>, `2004` <dbl>, `2005` <dbl>, `2006` <dbl>,  
## # `2007` <dbl>, `2008` <dbl>, `2009` <dbl>, `2010` <dbl>, `2011` <dbl>,  
## # `2012` <dbl>, `2013` <dbl>, `2014` <dbl>, `2015` <dbl>, `2016` <dbl>,  
## # `2017` <dbl>, `2018` <dbl>

#For my ease of use I simply created another object to prevent confusion.  
  
WakeCountySales2 <- WakeCountySales %>%  
 pivot\_longer(cols=c("1996","1997","1998","1999","2000","2001","2002","2003","2004","2005","2006","2007","2008","2009","2010","2011","2012","2013","2014","2015","2016","2017","2018"),  
 names\_to = "YR",  
 values\_to = "ZHVI",  
 values\_drop\_na = TRUE)  
  
  
#Analysis  
  
WakeCountySales2 %>%  
 ggplot(aes(YR,ZHVI, xlab="Year", ylab="ZHVI"))+  
 geom\_point(aes(color= RegionName))+  
 labs(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)



WakeCountySales2 %>%  
 ggplot(aes(YR,ZHVI, xlab="Year", ylab="ZHVI"))+  
 geom\_col(aes(color= RegionName))+  
 labs(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)



# You are considering a move to the Raleigh area and are interested in understanding trends in home values. Run the analysis in the instructions below and come back to this section of the R Markdown document and address the following questions:

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

## The overall trend is increasing values.

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

## There were two dips, one from 1996 to 1999 then another from 2008 to 2012.

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

## Least expensive appears to be Zebulon and the most expensive is New Hill

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

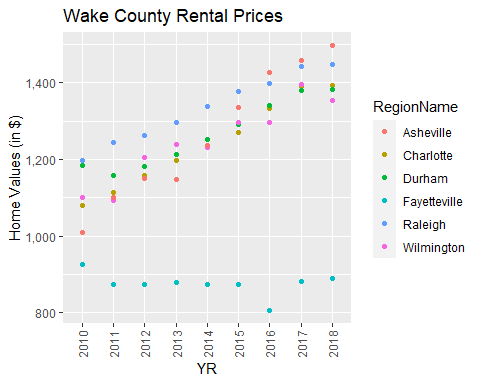
## Zebulon seems to be dipping more than others but the lowest 5 seem to be reaching a peak before dipping. New Hill stands furthest out as the highest value.

### NC Rental Market

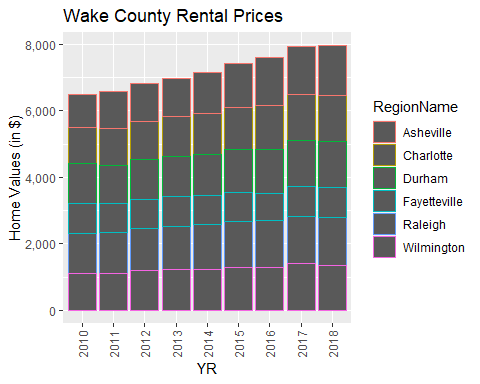
Rentals <- data.frame(SingleFamRental$RegionName, SingleFamRental$State, SingleFamRental$`2010-11`, SingleFamRental$`2011-11`, SingleFamRental$`2012-11`, SingleFamRental$`2013-11`, SingleFamRental$`2014-11`, SingleFamRental$`2015-11`, SingleFamRental$`2016-11`, SingleFamRental$`2017-11`, SingleFamRental$`2018-10`) %>%  
 filter(SingleFamRental$State== "NC" & SingleFamRental$RegionName %in%  
 c("Asheville","Charlotte","Durham","Fayetteville","Raleigh","Wilmington"))  
  
#Rename Columns  
  
names(Rentals)[names(Rentals) == 'SingleFamRental.RegionName'] <- "RegionName"   
names(Rentals)[names(Rentals) == 'SingleFamRental.State'] <- "State"   
names(Rentals)[names(Rentals) == 'SingleFamRental..2010.11.'] <- "2010"  
names(Rentals)[names(Rentals) == 'SingleFamRental..2011.11.'] <- "2011"  
names(Rentals)[names(Rentals) == 'SingleFamRental..2012.11.'] <- "2012"  
names(Rentals)[names(Rentals) == 'SingleFamRental..2013.11.'] <- "2013"  
names(Rentals)[names(Rentals) == 'SingleFamRental..2014.11.'] <- "2014"  
names(Rentals)[names(Rentals) == 'SingleFamRental..2015.11.'] <- "2015"  
names(Rentals)[names(Rentals) == 'SingleFamRental..2016.11.'] <- "2016"  
names(Rentals)[names(Rentals) == 'SingleFamRental..2017.11.'] <- "2017"  
names(Rentals)[names(Rentals) == 'SingleFamRental..2018.10.'] <- "2018"  
  
  
#Clean it up  
  
Rentals %>%  
 group\_by(RegionName)

## # A tibble: 6 × 11  
## # Groups: RegionName [6]  
## RegionName State `2010` `2011` `2012` `2013` `2014` `2015` `2016` `2017`  
## <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Charlotte NC 1080 1113 1157 1196 1231 1271 1332 1390  
## 2 Raleigh NC 1198 1243 1263 1295 1337 1376 1398 1443  
## 3 Durham NC 1185 1157 1182 1213 1253 1291 1341 1379  
## 4 Fayetteville NC 925 875 875 879 875 873 807 882  
## 5 Wilmington NC 1100 1094 1204 1238 1230 1295 1296 1394  
## 6 Asheville NC 1010 1100 1150 1147 1236 1336 1426 1459  
## # ℹ 1 more variable: `2018` <dbl>

#To minimize my confusion I create a second Rentals  
  
Rentals2 <- Rentals %>%  
 pivot\_longer(cols=c("2010","2011","2012","2013","2014","2015","2016","2017","2018"),  
 names\_to = "YR",  
 values\_to = "ZHVI",  
 values\_drop\_na = TRUE)  
  
Rentals2 %>%  
 ggplot(aes(YR,ZHVI, xlab="Year", ylab="ZHVI"))+  
 geom\_point(aes(color= RegionName))+  
 labs(title ="Wake County Rental Prices")+  
 theme(axis.text.x = element\_text(angle = 90, vjust=0.5))+  
 scale\_y\_continuous(name="Home Values (in $)", labels = scales::comma)



Rentals2 %>%  
 ggplot(aes(YR,ZHVI, xlab="Year", ylab="ZHVI"))+  
 geom\_col(aes(color= RegionName))+  
 labs(title ="Wake County Rental Prices")+  
 theme(axis.text.x = element\_text(angle = 90, vjust=0.5))+  
 scale\_y\_continuous(name="Home Values (in $)", labels = scales::comma)



# You are considering relocating to a new city and want to rent. You are looking at some of the larger cities within the state including Asheville, Charlotte, Durham, Fayetteville, Raleigh and Wilmington. Answer the following based upon the analysis above:

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

## Overall the rent prices have increased around the state. Fayetteville is the only city that is not following the trend.

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

## The most expensive city to rent in is Asheville, Fayetteville is the least expensive.

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

## Wilmington has a lower rent than Asheville.

### Home Values in Select Rental Markets

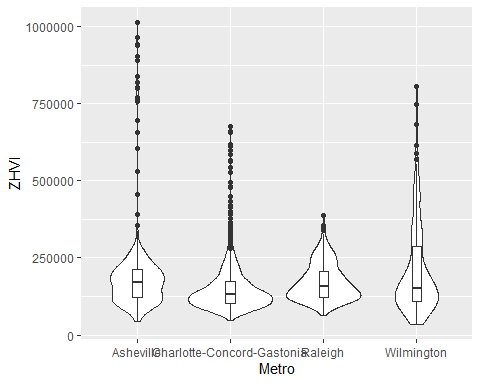
NCHomeSales <- data.frame(SingleFamSales$RegionName, SingleFamSales$State, SingleFamSales$Metro, SingleFamSales$`1996-05` , SingleFamSales$`1997-05`, SingleFamSales$`1998-05`, SingleFamSales$`1999-05`, SingleFamSales$`2000-05`, SingleFamSales$`2001-05`, SingleFamSales$`2002-05`, SingleFamSales$`2003-05`, SingleFamSales$`2004-05`, SingleFamSales$`2005-05`, SingleFamSales$`2006-05`, SingleFamSales$`2007-05`, SingleFamSales$`2008-05`, SingleFamSales$`2009-05`, SingleFamSales$`2010-05`, SingleFamSales$`2011-05`, SingleFamSales$`2012-05`, SingleFamSales$`2013-05`, SingleFamSales$`2014-05`, SingleFamSales$`2015-05`, SingleFamSales$`2016-05`, SingleFamSales$`2017-05`, SingleFamSales$`2018-05`)%>%  
 filter(SingleFamSales$State == "NC", SingleFamSales$Metro %in%  
 c("Asheville","Charlotte-Concord-Gastonia","Raleigh","Wilmington"))  
  
  
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales.RegionName'] <- "RegionName"  
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales.State'] <- "State"  
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales.Metro'] <- "Metro"  
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..1996.05.'] <- "1996"   
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..1997.05.'] <- "1997"   
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..1998.05.'] <- "1998"   
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..1999.05.'] <- "1999"   
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..2000.05.'] <- "2000"   
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..2001.05.'] <- "2001"   
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..2002.05.'] <- "2002"   
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..2003.05.'] <- "2003"   
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..2004.05.'] <- "2004"   
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..2005.05.'] <- "2005"   
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..2006.05.'] <- "2006"   
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..2007.05.'] <- "2007"   
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..2008.05.'] <- "2008"   
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..2009.05.'] <- "2009"   
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..2010.05.'] <- "2010"   
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..2011.05.'] <- "2011"   
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..2012.05.'] <- "2012"   
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..2013.05.'] <- "2013"   
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..2014.05.'] <- "2014"   
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..2015.05.'] <- "2015"   
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..2016.05.'] <- "2016"   
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..2017.05.'] <- "2017"   
names(NCHomeSales)[names(NCHomeSales) == 'SingleFamSales..2018.05.'] <- "2018"  
  
  
NCHomeSales%>%  
 group\_by(Metro)

## # A tibble: 104 × 26  
## # Groups: Metro [4]  
## RegionName State Metro `1996` `1997` `1998` `1999` `2000` `2001` `2002`  
## <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Charlotte NC Charlotte… 116000 121700 125300 130600 136900 139700 141600  
## 2 Raleigh NC Raleigh 141500 140900 141000 153800 160600 163200 165000  
## 3 Wilmington NC Wilmington 99300 106900 122300 128000 141600 140100 144200  
## 4 Cary NC Raleigh 191900 195700 174900 195900 203600 218500 220700  
## 5 Concord NC Charlotte… 109000 106700 113900 121600 120400 125100 133700  
## 6 Asheville NC Asheville 87800 93500 92600 97900 105500 112500 120400  
## 7 Apex NC Raleigh 208700 195000 177500 175400 189000 192700 201000  
## 8 Salisbury NC Charlotte… 66900 73400 82100 83600 87400 96500 101700  
## 9 Mooresville NC Charlotte… 134800 144000 138800 148700 170600 168500 173300  
## 10 Gastonia NC Charlotte… 69400 75300 83300 89800 93500 98500 102500  
## # ℹ 94 more rows  
## # ℹ 16 more variables: `2003` <dbl>, `2004` <dbl>, `2005` <dbl>, `2006` <dbl>,  
## # `2007` <dbl>, `2008` <dbl>, `2009` <dbl>, `2010` <dbl>, `2011` <dbl>,  
## # `2012` <dbl>, `2013` <dbl>, `2014` <dbl>, `2015` <dbl>, `2016` <dbl>,  
## # `2017` <dbl>, `2018` <dbl>

NCHomeSales2 <- NCHomeSales %>%  
 pivot\_longer(cols=c("1996","1997","1998","1999","2000","2001","2002","2003","2004","2005","2006","2007","2008","2009","2010","2011","2012","2013","2014","2015","2016","2017","2018"),  
 names\_to = "YR",  
 values\_to = "ZHVI",  
 values\_drop\_na = TRUE)  
  
NCHomeSales2%>%  
 group\_by(Metro)

## # A tibble: 2,302 × 5  
## # Groups: Metro [4]  
## RegionName State Metro YR ZHVI  
## <chr> <chr> <chr> <chr> <dbl>  
## 1 Charlotte NC Charlotte-Concord-Gastonia 1996 116000  
## 2 Charlotte NC Charlotte-Concord-Gastonia 1997 121700  
## 3 Charlotte NC Charlotte-Concord-Gastonia 1998 125300  
## 4 Charlotte NC Charlotte-Concord-Gastonia 1999 130600  
## 5 Charlotte NC Charlotte-Concord-Gastonia 2000 136900  
## 6 Charlotte NC Charlotte-Concord-Gastonia 2001 139700  
## 7 Charlotte NC Charlotte-Concord-Gastonia 2002 141600  
## 8 Charlotte NC Charlotte-Concord-Gastonia 2003 142100  
## 9 Charlotte NC Charlotte-Concord-Gastonia 2004 144200  
## 10 Charlotte NC Charlotte-Concord-Gastonia 2005 148200  
## # ℹ 2,292 more rows

NCHomeSales2 %>%  
 ggplot(aes(Metro,ZHVI))+  
 geom\_violin()+  
 geom\_boxplot(width=0.1)+  
 scale\_y\_continuous(labels = function(x) format(x, scientific = FALSE))



# 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. Run the analysis above to answer the following:

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

## The Charlotte-Concord-Gastonia metro area.

# 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?

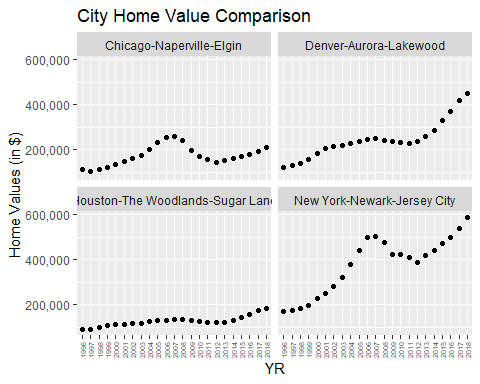
## The Charlotte-Concord-Gastonia metro area, closely followed by the Raleigh area.

### Relocation Home Value Comparison

NationalHomeSales <- data.frame(SingleFamSales$RegionName, SingleFamSales$State, SingleFamSales$Metro, SingleFamSales$`1996-05` , SingleFamSales$`1997-05`, SingleFamSales$`1998-05`, SingleFamSales$`1999-05`, SingleFamSales$`2000-05`, SingleFamSales$`2001-05`, SingleFamSales$`2002-05`, SingleFamSales$`2003-05`, SingleFamSales$`2004-05`, SingleFamSales$`2005-05`, SingleFamSales$`2006-05`, SingleFamSales$`2007-05`, SingleFamSales$`2008-05`, SingleFamSales$`2009-05`, SingleFamSales$`2010-05`, SingleFamSales$`2011-05`, SingleFamSales$`2012-05`, SingleFamSales$`2013-05`, SingleFamSales$`2014-05`, SingleFamSales$`2015-05`, SingleFamSales$`2016-05`, SingleFamSales$`2017-05`, SingleFamSales$`2018-05`)%>%  
 filter(SingleFamSales$State %in% c("IL", "NY", "CO","TX") &  
 SingleFamSales$RegionName %in%  
 c("Chicago","Denver","Houston","New York"))  
  
  
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales.RegionName'] <- "RegionName"  
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales.State'] <- "State"  
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales.Metro'] <- "Metro"  
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..1996.05.'] <- "1996"   
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..1997.05.'] <- "1997"   
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..1998.05.'] <- "1998"   
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..1999.05.'] <- "1999"   
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..2000.05.'] <- "2000"   
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..2001.05.'] <- "2001"   
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..2002.05.'] <- "2002"   
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..2003.05.'] <- "2003"   
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..2004.05.'] <- "2004"   
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..2005.05.'] <- "2005"   
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..2006.05.'] <- "2006"   
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..2007.05.'] <- "2007"   
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..2008.05.'] <- "2008"   
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..2009.05.'] <- "2009"   
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..2010.05.'] <- "2010"   
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..2011.05.'] <- "2011"   
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..2012.05.'] <- "2012"   
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..2013.05.'] <- "2013"   
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..2014.05.'] <- "2014"   
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..2015.05.'] <- "2015"   
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..2016.05.'] <- "2016"   
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..2017.05.'] <- "2017"   
names(NationalHomeSales)[names(NationalHomeSales) == 'SingleFamSales..2018.05.'] <- "2018"   
  
  
NationalHomeSales %>%  
 group\_by(Metro)

## # A tibble: 4 × 26  
## # Groups: Metro [4]  
## RegionName State Metro `1996` `1997` `1998` `1999` `2000` `2001` `2002` `2003`  
## <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 New York NY New … 172100 173000 183500 196400 227200 248500 282400 320900  
## 2 Chicago IL Chic… 113100 103500 110400 123200 134300 146200 161900 175900  
## 3 Houston TX Hous… 89000 90000 100300 106900 111400 112600 115200 118100  
## 4 Denver CO Denv… 119400 128100 138700 155600 181800 204200 216200 220300  
## # ℹ 15 more variables: `2004` <dbl>, `2005` <dbl>, `2006` <dbl>, `2007` <dbl>,  
## # `2008` <dbl>, `2009` <dbl>, `2010` <dbl>, `2011` <dbl>, `2012` <dbl>,  
## # `2013` <dbl>, `2014` <dbl>, `2015` <dbl>, `2016` <dbl>, `2017` <dbl>,  
## # `2018` <dbl>

#For my ease of use I simply created another object to prevent confusion.  
  
NationalHomeSales2 <- NationalHomeSales %>%  
 pivot\_longer(cols=c("1996","1997","1998","1999","2000","2001","2002","2003","2004","2005","2006","2007","2008","2009","2010","2011","2012","2013","2014","2015","2016","2017","2018"),  
 names\_to = "YR",  
 values\_to = "ZHVI",  
 values\_drop\_na = TRUE)  
  
NationalHomeSales2 %>%  
 ggplot(aes(YR,ZHVI, xlab="Year", ylab="ZHVI"))+  
 geom\_point()+  
 facet\_wrap(~Metro)+  
 labs(title ="City Home Value Comparison")+  
 theme(axis.text.x = element\_text(angle = 90, vjust=0.5, size=5))+  
 scale\_y\_continuous(name="Home Values (in $)", labels = scales::comma)



# You have been given a new opportunity to relocate for a new position within your company. They have given you the option of 4 different areas in the country (Chicago, Denver, Houston or New York). Run the analysis in the instructions below and come back to this section of the R Markdown document and address the following questions:

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

## The most affordable is the Houston area. The least affordable is the New York area.

# b. 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 New York and Dever areas saw the greatest change in prices over the last 5 years. Houston has remained the most consistent and Chicago also being consistent.

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

## New York was the most impacted and recovered greater than before the down turn. Chicago also had a down-turn but barely recovered.