# MIS503 - Final Project

## Sullivan, Diane

### Zillow Home Value Index Analysis

### Wake County Home Sales

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4  
## ✔ tibble 3.1.7 ✔ dplyr 1.0.9  
## ✔ tidyr 1.2.0 ✔ stringr 1.4.0  
## ✔ readr 2.1.2 ✔ forcats 0.5.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(lubridate)

##   
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':  
##   
## date, intersect, setdiff, union

SingleFamilyResidenceSales <- 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.

message=FALSE

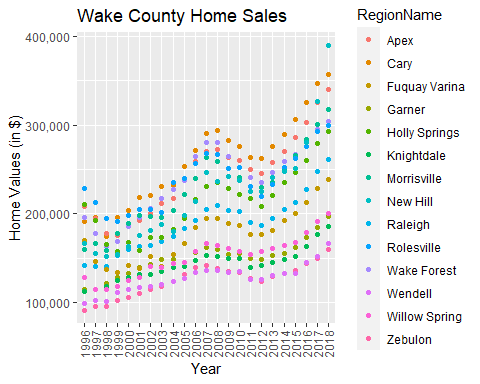
WakeCountySales <- filter(SingleFamilyResidenceSales, State == "NC", CountyName == "Wake County")  
WakeCountySales <- select(WakeCountySales, RegionName, State, CountyName, Metro,"1996-05","1997-05","1998-05","1999-05","2000-05","2001-05", "2002-05","2003-05","2004-05","2005-05","2006-05","2007-05","2008-05","2009-05","2010-05","2011-05","2012-05","2013-05","2014-05","2015-05", "2016-05","2017-05","2018-05")  
WakeCountySales <- rename(WakeCountySales, "1996"="1996-05","1997"="1997-05","1998"="1998-05","1999"="1999-05","2000"="2000-05","2001"="2001-05","2002"="2002-05","2003"="2003-05","2004"="2004-05","2005"="2005-05","2006"="2006-05","2007"="2007-05","2008"="2008-05","2009"="2009-05","2010"="2010-05","2011"="2011-05","2012"="2012-05","2013"="2013-05","2014"="2014-05", "2015"="2015-05", "2016"= "2016-05","2017"="2017-05","2018"="2018-05")

pivot\_longer(WakeCountySales,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")

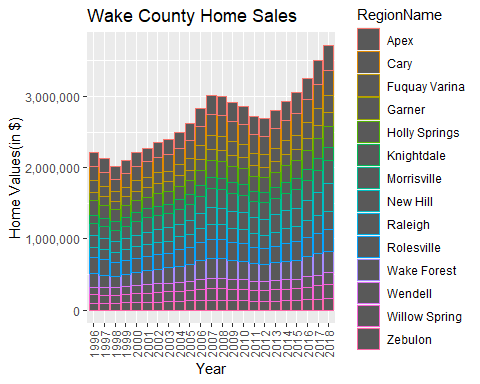
## # A tibble: 322 × 6  
## RegionName State CountyName Metro YR ZHVI  
## <chr> <chr> <chr> <chr> <chr> <dbl>  
## 1 Raleigh NC Wake County Raleigh 1996 141500  
## 2 Raleigh NC Wake County Raleigh 1997 140900  
## 3 Raleigh NC Wake County Raleigh 1998 141000  
## 4 Raleigh NC Wake County Raleigh 1999 153800  
## 5 Raleigh NC Wake County Raleigh 2000 160600  
## 6 Raleigh NC Wake County Raleigh 2001 163200  
## 7 Raleigh NC Wake County Raleigh 2002 165000  
## 8 Raleigh NC Wake County Raleigh 2003 169200  
## 9 Raleigh NC Wake County Raleigh 2004 174600  
## 10 Raleigh NC Wake County Raleigh 2005 183300  
## # … with 312 more rows

WakeCountySales<-pivot\_longer(WakeCountySales,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")

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 = .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 = .5))+  
 scale\_y\_continuous(name="Home Values(in $)", labels = scales::comma)



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

Home prices have trended upward, with the exception of two dips: following 1996 and again following 2008.

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

The years where prices dipped include 1997, 1998 and 2009, 2010, 2011 and 2012.

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

Zebulon

### Most expensive area?

In 2018, New Hill is the most expensive area to purchase a home in Wake County.

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

Apex area was expensive in 1996 and then had a downturn and didn’t regain it’s value until 2003.

### NC Rental Market

SingleFamilyResidenceRental <- 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.

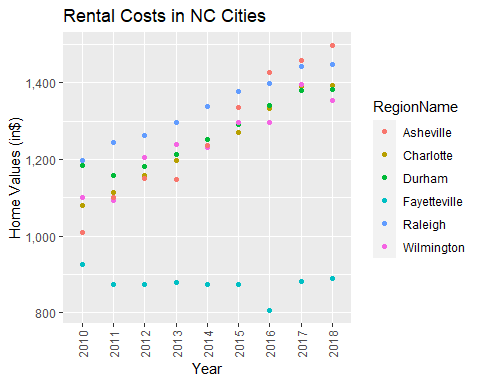
Rentals <- filter(SingleFamilyResidenceRental, State=="NC")  
Rentals <- filter (Rentals, RegionName %in% c("Asheville","Charlotte","Durham","Fayetteville","Raleigh","Wilmington"))  
Rentals <- select(Rentals, "RegionName","State","2010-11","2011-11","2012-11","2013-11","2014-11","2015-11","2016-11","2017-11","2018-10")  
Rentals <- rename(Rentals, "2010"="2010-11","2011"="2011-11","2012"="2012-11","2013"="2013-11","2014"="2014-11","2015"="2015-11","2016"="2016-11","2017"="2017-11","2018"="2018-10")

pivot\_longer(Rentals,cols = c("2010","2011","2012","2013","2014","2015","2016","2017","2018"), names\_to = "YR", values\_to = "ZHVI")

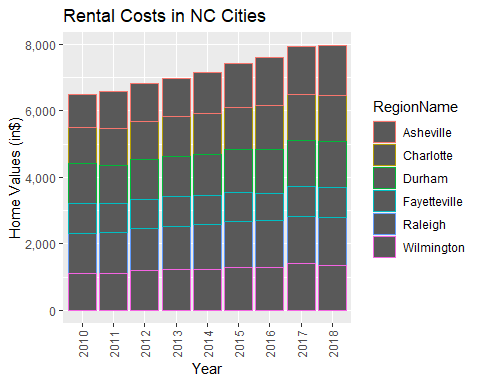
## # A tibble: 54 × 4  
## RegionName State YR ZHVI  
## <chr> <chr> <chr> <dbl>  
## 1 Charlotte NC 2010 1080  
## 2 Charlotte NC 2011 1113  
## 3 Charlotte NC 2012 1157  
## 4 Charlotte NC 2013 1196  
## 5 Charlotte NC 2014 1231  
## 6 Charlotte NC 2015 1271  
## 7 Charlotte NC 2016 1332  
## 8 Charlotte NC 2017 1390  
## 9 Charlotte NC 2018 1393  
## 10 Raleigh NC 2010 1198  
## # … with 44 more rows

Rentals <-pivot\_longer(Rentals,cols = c("2010","2011","2012","2013","2014","2015","2016","2017","2018"), names\_to = "YR", values\_to = "ZHVI")

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 = .5))+  
 scale\_y\_continuous(name = "Home Values (in$)", 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 = .5))+  
 scale\_y\_continuous(name = "Home Values (in$)", labels = scales::comma)



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

Rents are trending upward all around the state. However, from 2017 to 2018 prices were level (no dip and no rise) in the rental markets of Charlotte, Raleigh and Durham. Also from 2017-2018 there was a dip in Wilmington and a small rise in Asheville.

### Where is the most expensive city to rent in?

Asheville

### Least expensive city to rent in?

Fayetteville

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

Wilmington

### Home Values in Select Rental Markets

NCHomeSales <- filter(SingleFamilyResidenceSales,State=="NC")  
NCHomeSales <- filter(NCHomeSales, Metro %in% c("Asheville","Charlotte-Concord-Gastonia","Raleigh", "Wilmington"))  
NCHomeSales <- select(NCHomeSales, RegionName, State, Metro,"1996-05","1997-05","1998-05","1999-05","2000-05","2001-05", "2002-05","2003-05","2004-05","2005-05","2006-05","2007-05","2008-05","2009-05","2010-05","2011-05","2012-05","2013-05","2014-05","2015-05", "2016-05","2017-05","2018-05")  
NCHomeSales <- rename(NCHomeSales,"1996"="1996-05","1997"="1997-05","1998"="1998-05","1999"="1999-05","2000"="2000-05","2001"="2001-05","2002"="2002-05","2003"="2003-05","2004"="2004-05","2005"="2005-05","2006"="2006-05","2007"="2007-05","2008"="2008-05","2009"="2009-05","2010"="2010-05","2011"="2011-05","2012"="2012-05","2013"="2013-05","2014"="2014-05", "2015"="2015-05", "2016"= "2016-05","2017"="2017-05","2018"="2018-05")

pivot\_longer(NCHomeSales, 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")

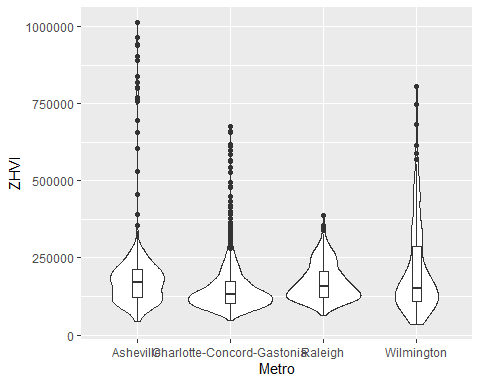
## # A tibble: 2,392 × 5  
## RegionName State Metro YR ZHVI  
## <chr> <chr> <chr> <chr> <dbl>  
## 1 Charlotte NC Charlotte-Concord-Gastonia 1996 116000  
## 2 Charlotte NC Charlotte-Concord-Gastonia 1997 121700  
## 3 Charlotte NC Charlotte-Concord-Gastonia 1998 125300  
## 4 Charlotte NC Charlotte-Concord-Gastonia 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  
## # … with 2,382 more rows

NCHomeSales <-pivot\_longer(NCHomeSales, 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")

ggplot(NCHomeSales, aes(x=Metro, y=ZHVI))+  
 geom\_violin()+  
 geom\_boxplot(width=.1)

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

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



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

The Charlotte-Concord-Gastonia region has the lowest median house price.

### 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 region has the most inventory around the median home value.

### Relocation Home Value Comparison

NationalHomeSales <- filter(SingleFamilyResidenceSales,State%in% c("NY","CO","TX","IL"))  
NationalHomeSales <- filter(NationalHomeSales, RegionName%in% c("Chicago", "Denver", "Houston", "New York"))  
NationalHomeSales <- select(NationalHomeSales, RegionName, State, Metro,"1996-05","1997-05","1998-05","1999-05","2000-05","2001-05", "2002-05","2003-05","2004-05","2005-05","2006-05","2007-05","2008-05","2009-05","2010-05","2011-05","2012-05","2013-05","2014-05","2015-05", "2016-05","2017-05","2018-05")  
NationalHomeSales <- rename(NationalHomeSales,"1996"="1996-05","1997"="1997-05","1998"="1998-05","1999"="1999-05","2000"="2000-05","2001"="2001-05","2002"="2002-05","2003"="2003-05","2004"="2004-05","2005"="2005-05","2006"="2006-05","2007"="2007-05","2008"="2008-05","2009"="2009-05","2010"="2010-05","2011"="2011-05","2012"="2012-05","2013"="2013-05","2014"="2014-05", "2015"="2015-05", "2016"= "2016-05","2017"="2017-05","2018"="2018-05")   
head(NationalHomeSales)

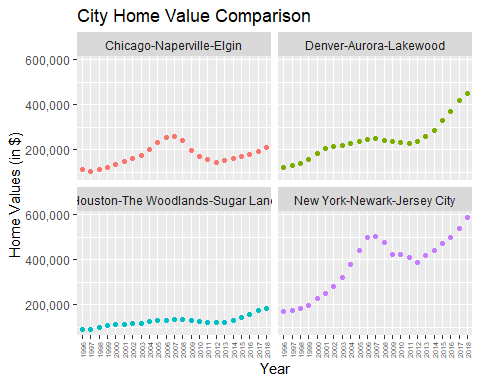
## # A tibble: 4 × 26  
## RegionName State Metro `1996` `1997` `1998` `1999` `2000` `2001` `2002` `2003`  
## <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 New York NY New … 172100 173000 183500 196400 227200 248500 282400 320900  
## 2 Chicago IL Chic… 113100 103500 110400 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  
## # … with 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>

pivot\_longer(NationalHomeSales, 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")

## # A tibble: 92 × 5  
## RegionName State Metro YR ZHVI  
## <chr> <chr> <chr> <chr> <dbl>  
## 1 New York NY New York-Newark-Jersey City 1996 172100  
## 2 New York NY New York-Newark-Jersey City 1997 173000  
## 3 New York NY New York-Newark-Jersey City 1998 183500  
## 4 New York NY New York-Newark-Jersey City 1999 196400  
## 5 New York NY New York-Newark-Jersey City 2000 227200  
## 6 New York NY New York-Newark-Jersey City 2001 248500  
## 7 New York NY New York-Newark-Jersey City 2002 282400  
## 8 New York NY New York-Newark-Jersey City 2003 320900  
## 9 New York NY New York-Newark-Jersey City 2004 377400  
## 10 New York NY New York-Newark-Jersey City 2005 442600  
## # … with 82 more rows

NationalHomeSales <- pivot\_longer(NationalHomeSales, 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")

ggplot(NationalHomeSales, aes(x= YR, y= ZHVI, color = Metro))+  
 geom\_point()+  
 facet\_wrap("Metro")+  
 labs(title = "City Home Value Comparison", x = "Year", y = "Home Values (in $)")+  
 theme(axis.text.x = element\_text(angle = 90, vjust = .5, size = 5))+  
 scale\_y\_continuous(name = "Home Values (in $)", labels = scales::comma)+  
 theme(legend.position = "none")



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

The Houston-The Woodlands-SugarLand areas are most affordable.

### Least affordable?

New York-Newark-Jersey City area is least affordable.

### Which cities saw the largest change in prices over the past 5 years?

From 2013-2018, the ZHVI in the Denver-Aurora-Lakewood area grew from $260,800-$449,300. This change of $188,500 is the largest change of the four regions analyzed.

### Which city has remained more consistent with no huge swings up or down in home values?

The Houston-The Woodlands-SugarLand area is the most consistently valued.

### During the market downturn in 2012, which cities were most impacted?

From 2011-2012, New York-Newark-Jersey City housing was most impacted.

### Which cities have recovered?

All of the market areas have recovered from the 2012 housing market downturn.