# MIS 503 - Final Project

## Hildreth, Evan

### Zillow Home Value Index Analysis

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.0.5

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.5 v dplyr 1.0.7  
## v tidyr 1.1.4 v stringr 1.4.0  
## v readr 2.0.2 v forcats 0.5.1

## Warning: package 'ggplot2' was built under R version 4.0.5

## Warning: package 'tibble' was built under R version 4.0.5

## Warning: package 'tidyr' was built under R version 4.0.5

## Warning: package 'readr' was built under R version 4.0.5

## Warning: package 'dplyr' was built under R version 4.0.5

## Warning: package 'stringr' was built under R version 4.0.5

## Warning: package 'forcats' was built under R version 4.0.5

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(lubridate)

## Warning: package 'lubridate' was built under R version 4.0.5

##   
## Attaching package: 'lubridate'

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

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.

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.

### Exercise 1: Wake County Home Sales

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

-For all of the regions in the Wake county area the general trend is that home values have been going up during this 22 year time frame with a brief dip around 2012/2013 after a previous peak in 2007/2008 but it ends in 2018 with all-time high home values.

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

* The first dip happened aroun 1997 and the second dip happened around 2012/2013 about.

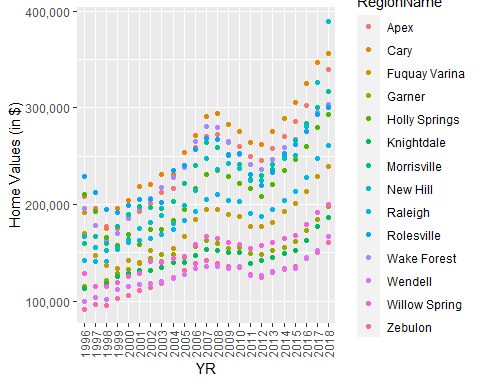
1. Based on the analysis, where would be the least expensive area to purchase home? Most expensive area?

* Wendle and Willow Springs are ususally the least expensive place to buy a home with Kightville also joining that pack in recent years. Apex and Cary tend to be the most expensive place to purchase a home but for the last year, 2018, Raleigh seems to have shot up in home value.

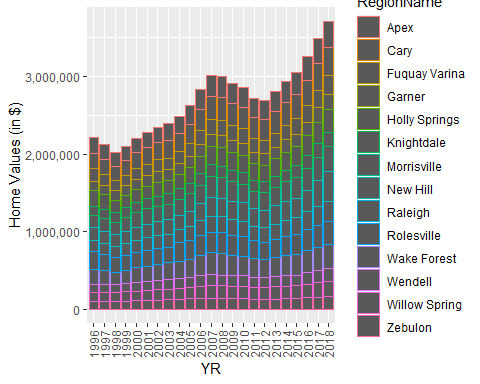
1. Are any area home values trending down? Is there one area that stands out compared to others

* While no home values have trended downwards for the entirity of this study, some have lost ground to some of their comtemporis when this study was first started in 1996 compared to where they are now in 2018. Namely Kightdale and Fuquay Varina.

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')  
  
WakeCountySales <- WakeCountySales %>%   
 pivot\_longer(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)+  
 geom\_point(aes(YR, ZHVI, color= RegionName)) +  
 theme(axis.text.x = element\_text(angle = 90, vjust=0.5)) +  
 scale\_y\_continuous(name="Home Values (in $)", labels = scales::comma)



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



### Exercise 2: North Carolina Rental Market

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. You need to answer the following questions:

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 general trend for rental costs in the major metro areas in NC is to see a steady increase year over year with the major exception being Fayettevile which has seen an over slight decrease in home value prices and if not it has been stagnating. One other note is that Wilmington is seeing a small decrease in 2018 compared to 2017.

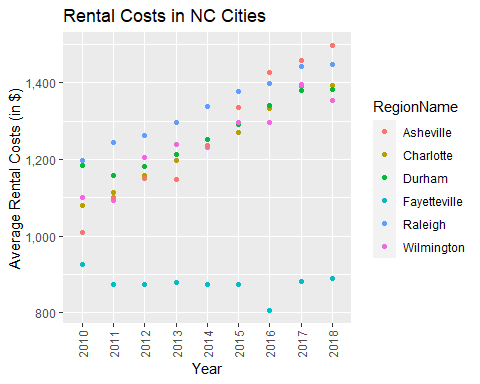
1. Where is the most expensive city to rent in? Least expensive?

* As of 2018 the most expensive city rent is Asheville and the least expensive is Fayetteville.

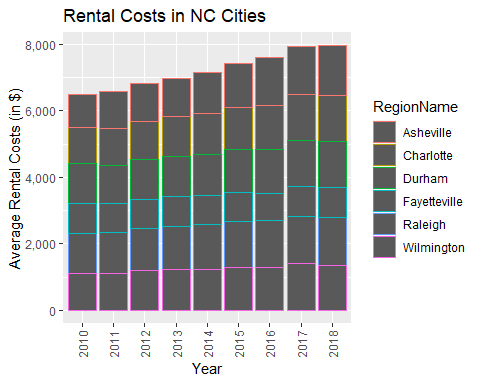
1. You are trying decide between Wilmington and Asheville. Which market has the lowest rent?

* If one is deciding between Wilmington and Asheville and lowest rental is the biggest deciding factor then Wilmington is where you should move. However, if one expands their search to absolute lowest rent in a major NC metro area the answer is, again, Fayetteville.

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")  
  
Rentals <- Rentals %>%  
 pivot\_longer(c('2010', '2011', '2012', '2013', '2014', '2015', '2016', '2017', '2018'), names\_to = 'YR', values\_to = 'ZHVI')  
  
ggplot(Rentals)+  
 geom\_point(aes(YR, ZHVI, color= RegionName)) +  
 theme(axis.text.x = element\_text(angle = 90, vjust=0.5)) +  
 scale\_y\_continuous(name="Average Rental Costs (in $)", labels = scales::comma) +  
 labs(title = "Rental Costs in NC Cities",  
 x= "Year")



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



### Home Values in Select Rental Markets

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. To decide on where to live you need to answer the following questions:

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

* From these results we can see that Charlotte-Concord-Gastonia area has the lowest median home price based off the boxplot within the violin plot.

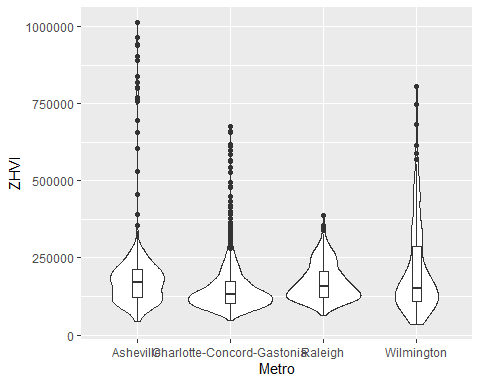
1. The violin plot will show density meaning the wider the plot is, the more observations occur within that area. Which market has the most density around the median value of homes?

* The Charlotte-Gastonia-Concord region has the most density around its mean for home prices.

NCHomeSales <- filter(SingleFamilyResidenceSales, State == "NC" & 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')  
  
NCHomeSales <- NCHomeSales %>%  
 pivot\_longer(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)+  
 geom\_violin(aes(Metro, ZHVI)) +  
 geom\_boxplot(width=0.1, aes(Metro, ZHVI))

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

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



### Relocation to Another City

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). Answer the following questions to help decide on where to move to:

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

* Based off of this analysis, the most affordable housing cost metro area of the 4 being studied is Houston, with a close 2nd being Chicago. New York city has the least affordable housing according to this study.

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)?

* During the last 5 years of this study, Denver has seen the greatest home price increase and Houston has seen very little home price increase over the last 5 years.

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

* New York had the largest downturn in the 2012 housing market correction and subsequently has had the strongest recovery. However, in the time frame since 2012 Denver has had the greatest home price increase but I would not consider it a ‘recovery’ as the dip in home prices in 2012 was fairly mild compared to New York.

NationalHomeSales <- filter(SingleFamilyResidenceSales, RegionName %in% c("Chicago", "Denver", "Houston", "New York"))  
NationalHomeSales <- filter(NationalHomeSales, State %in% c("IL", "CO", "TX", "NY"))  
  
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')  
  
NationalHomeSales <- NationalHomeSales%>%  
 pivot\_longer(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) +  
 geom\_point(aes(YR, ZHVI)) +  
 facet\_wrap(~Metro) +  
 theme(axis.text.x = element\_text(angle = 90, vjust=0.5)) +  
 scale\_y\_continuous(name="Home Value (in $)", labels = scales::comma) +  
 labs(title = "City Home Value Comparison",  
 x= "Year")

