# MIS503 - Final Project

## Kiser, Stephen

### Zillow Home Value Index Analysis

### Wake County Home Sales

library(tidyverse)

## -- Attaching packages -------------------------------------------------------------------------------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.0 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.5  
## v tidyr 1.0.2 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.5.0

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

library(dplyr)  
SingleFamilyResidenceSales <- read\_csv("SingleFamilyResidenceSales.csv")

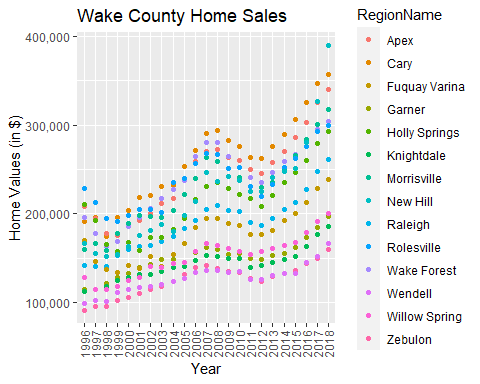
## Parsed with column specification:  
## cols(  
## .default = col\_double(),  
## RegionName = col\_character(),  
## State = col\_character(),  
## Metro = col\_character(),  
## CountyName = col\_character()  
## )

## See spec(...) for full column specifications.

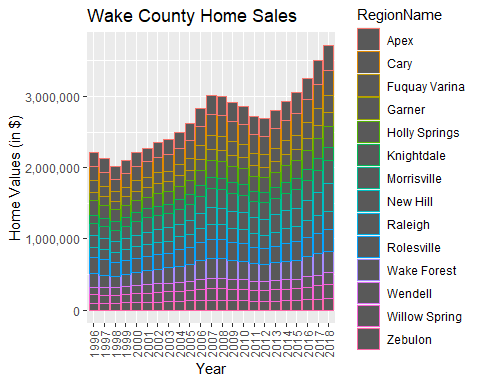
SingleFamilyResidenceRental <- read\_csv("SingleFamilyResidenceRental.csv")

## Parsed with column specification:  
## cols(  
## .default = col\_double(),  
## RegionName = col\_character(),  
## State = col\_character(),  
## Metro = col\_character(),  
## CountyName = col\_character()  
## )  
## See spec(...) for full column specifications.

WakeCountySales <- select(SingleFamilyResidenceSales,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") %>%  
 filter(CountyName == "Wake County")  
  
WakeCountySales <- rename(WakeCountySales, c('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 <-  
 gather(data = WakeCountySales, "1996", "1997", "1998", "1999", "2000", "2001", "2002", "2003", "2004", "2005", "2006", "2007", "2008", "2009", "2010", "2011", "2012", "2013", "2014", "2015", "2016", "2017", "2018", key = "YR", value = "ZHVI")  
  
ggplot(WakeCountySales, aes(YR, 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(YR, 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)



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

The overall trends in Wake County Home Values are an increase in home values.

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

The year 1997-1998 and 2009-2012 are the dips in the trend for Home Values.

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

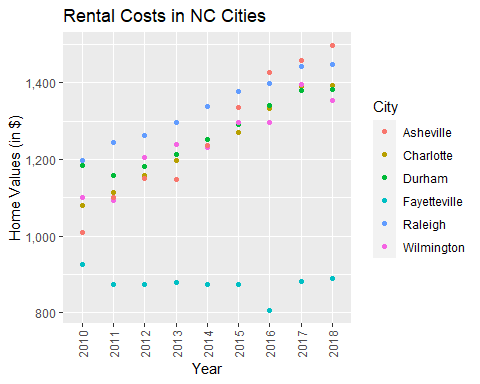
The least expensive area to purchase a home in 2018 is in Zebulon with a home value of around $150,000. The most expensive area to purchase a home is New Hill with about 380,000 dollar home value.

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

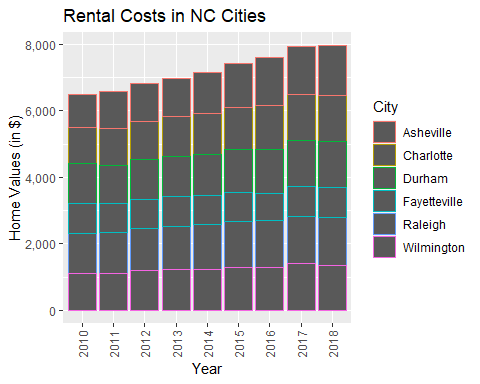
The only area that has a trending down home value is the Garner Area. There really is not any areas trending that stands out. I was surprised Raleigh area was not the highest home value area.

### NC Rental Market

Rentals <- select(SingleFamilyResidenceRental, RegionName, State, "2010-11", "2011-11", "2012-11", "2013-11", "2014-11", "2015-11", "2016-11", "2017-11", "2018-10") %>%  
 filter(RegionName %in% c("Asheville", "Charlotte", "Durham", "Fayetteville", "Raleigh", "Wilmington"), State == "NC")  
  
Rentals <-rename(Rentals, c(City = "RegionName", '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 <- gather(Rentals, "2010", "2011", "2012", "2013", "2014", "2015", "2016", "2017", "2018", key = 'YR', value = 'ZHVI')  
  
ggplot(Rentals, aes(x= YR, y= ZHVI, color = City))+  
 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="Home Values (in $)", labels = scales::comma)



ggplot(Rentals, aes(x= YR, y= ZHVI, color = City))+  
 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="Home Values (in $)", labels = scales::comma)



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 in the rental market is an increase in rental cost throughout the state. Fayetteville is the only city that does not follow this trend.

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

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

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

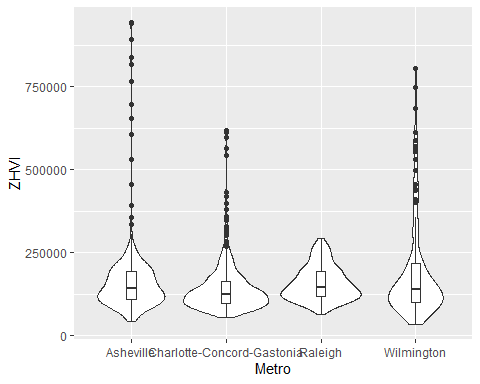
Wilmington has the lower rent than Asheville with less than $1,400 while Asheville is about 1,500 dollars.

### Home Values in Select Rental Markets

NCHomeSales <- select(SingleFamilyResidenceSales, 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") %>%  
 filter(Metro %in% c("Asheville", "Charlotte-Concord-Gastonia", "Raleigh", "Wilmington"), State == "NC")  
  
NCHomeSales <- rename(NCHomeSales, c('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 <- gather(NCHomeSales, "1996", "1997", "1998", "1999", "2000", "2001", "2002", "2003", "2004", "2005", "2006", "2007", "2008", "2009", "2010", key = "YR", value = "ZHVI")  
  
ggplot(NCHomeSales, aes(x= Metro, y= ZHVI))+  
 geom\_violin()+  
 geom\_boxplot(width=0.1)

## Warning: Removed 84 rows containing non-finite values (stat\_ydensity).

## Warning: Removed 84 rows containing non-finite values (stat\_boxplot).



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

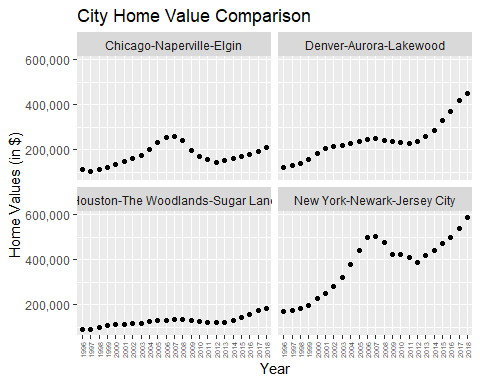
The market with the lowest median price is the Charlotte-Concord-Gastonia area.

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-Concord-Gastonia has the higher density around the median price.

### Relocation Home Value Comparison

NationalHomeSales <- select(SingleFamilyResidenceSales, 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")%>%  
 filter(RegionName %in% c("Chicago", "Denver", "Houston", "New York"), State %in% c("IL", "CO", "TX", "NY"))  
  
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 <- gather(NationalHomeSales, "1996", "1997", "1998", "1999", "2000", "2001", "2002", "2003", "2004", "2005", "2006", "2007", "2008", "2009", "2010", "2011", "2012", "2013", "2014", "2015", "2016", "2017", "2018", key = "YR", value = "ZHVI")  
  
ggplot(NationalHomeSales, aes(x = YR, y= ZHVI))+  
 geom\_point()+  
 facet\_wrap(~Metro)+  
 labs(title = "City Home Value Comparison",  
 x= "Year",  
 y= "Home Value (in $)")+  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, size = 5))+  
 scale\_y\_continuous(name = "Home Values (in $)", labels = scales::comma)



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

As of 2018, the Houston-The Woodlands, Sugar Land are is the most affordable, and the least affordable is New York-Newark-Jersey City area.

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)?

In the past 5 years the New York and Denver area have seen the largest change in home values, and the Houston and Chicago area have remained more consistent.

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

The cities that were most impacted by the market downturn in 2012 were New York and Chicago. The Denver area had a slight decrease but not as high of an impact as New York and Chicago. The New York and Denver areas grew the most and recovered from the market downturn in 2012. Chicago has seen a gradual climb back to where it once was before the downturn.