# MIS503 - Final Project

## Broach, Stuart

### Zillow Home Value Index Analysis

### Wake County Home Sales

1. What have been the overall trends in Wake County Home Values? Overall the price has trended up with the exception of the housing bubble crashing, which has since rebounded.
2. There were dips in home values in the past 20 years. What years did these occur? 1997,1998,2008,2009,2010, 2011, 2012
3. Based on the analysis, where would be the least expensive area to purchase home? Most expensive area? The least expensive area would be Zebulon and the most expensive would be Apex.
4. Are any area home values trending down? Is there one area that stands out compared to others? None are really tending down at the moment. Apex and Cary are tending up faster than others, though.

library(tidyverse)

## -- Attaching packages --------------------------------------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.1.0 v purrr 0.2.5  
## v tibble 1.4.2 v dplyr 0.7.7  
## v tidyr 0.8.2 v stringr 1.3.1  
## v readr 1.1.1 v forcats 0.3.0

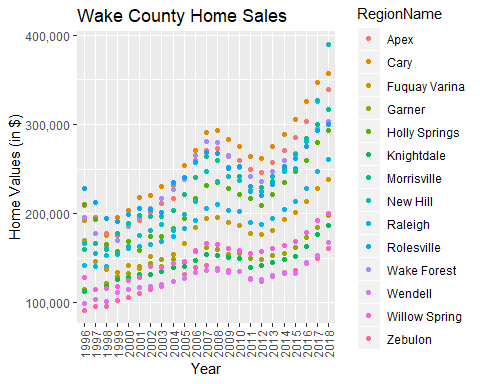
## -- Conflicts ------------------------------------------------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

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

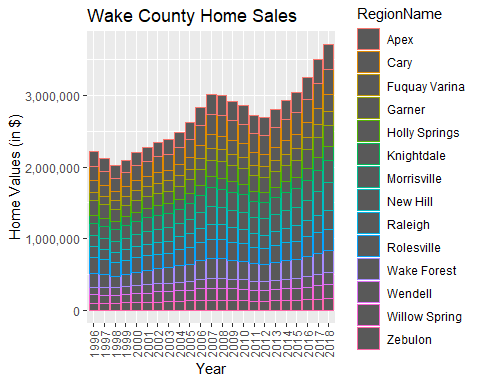
## Parsed with column specification:  
## cols(  
## .default = col\_integer(),  
## RegionName = col\_character(),  
## State = col\_character(),  
## Metro = col\_character(),  
## CountyName = col\_character()  
## )

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

WakeCountySales <- SingleFamilyResidenceSales %>%  
 select(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(State %in% c("NC"), CountyName %in% c("Wake County")) %>%   
 rename("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") %>%  
 gather('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(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)



### NC Rental Market

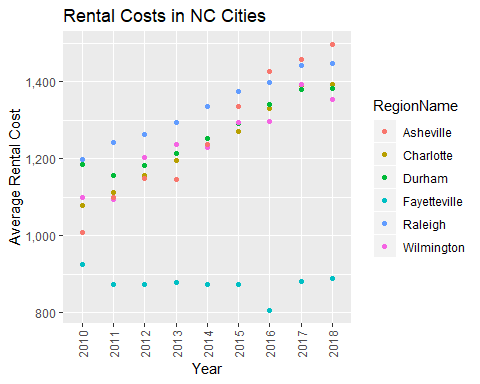
1. What has been the overall trend in the rental market around the state? Are there any cities that have not followed this trend? The overall trend is that rent is going up across the state. The only city to not follow this trend is Fayetteville.
2. Where is the most expensive city to rent in? Least expensive? The most expensive city to rent in is Raleigh. The least expensive is Fayetteville.
3. You are trying decide between Wilmington and Asheville. Which market has the lowest rent? Wilmington has the cheaper rent as of 2018.

library(tidyverse)  
library(readr)  
SingleFamilyResidenceRental <- read\_csv("SingleFamilyResidenceRental.csv")

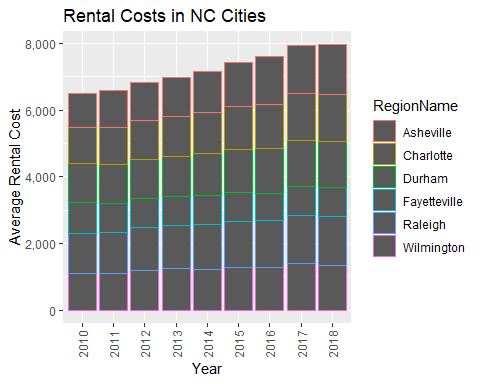
## Parsed with column specification:  
## cols(  
## .default = col\_integer(),  
## RegionName = col\_character(),  
## State = col\_character(),  
## Metro = col\_character(),  
## CountyName = col\_character()  
## )

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

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



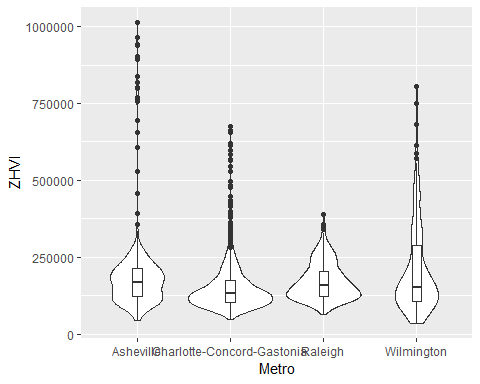
### Home Values in Select Rental Markets

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? Charlotte-Concord-Gastonia

library(tidyverse)  
NCHomeSales <- SingleFamilyResidenceSales %>%  
 select(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(State %in% c("NC"), Metro %in% c("Asheville", "Charlotte-Concord-Gastonia",  
"Raleigh", "Wilmington")) %>%   
 rename("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") %>%  
 gather('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') %>%  
 group\_by(Metro)  
 ggplot(NCHomeSales, aes(Metro, ZHVI)) +  
 geom\_violin() +  
 geom\_boxplot(width=0.1)

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

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



### Relocation Home Value Comparison

1. Based on your analysis, which city’s housing is most affordable? Least affordable? The most afordable housing would be in Houston. Least affordable would be 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)? Denver saw the largest increase while Houston remained consistent.
3. During the market downturn in 2012, which cities were most impacted? Which cities have recovered? New York and Chicago took the biggest downturn. All cities have recovered.

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

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

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

NationalHomeSales <- SingleFamilyResidenceSales %>%  
 select(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(State %in% c("IL", "CO", "TX", "NY"), RegionName %in% c("Chicago", "Denver", "Houston", "New York")) %>%  
 rename("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") %>%  
 gather('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') %>%  
 group\_by(Metro)  
 ggplot(NationalHomeSales, aes(YR, ZHVI)) +  
 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=0.5, size=5)) +  
 scale\_y\_continuous(name="Home Values (in $)", labels = scales::comma)

