# MIS503 – Final Project

# Subramanian, Subbu

# Zillow Home Value Index Analysis

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ 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()

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")

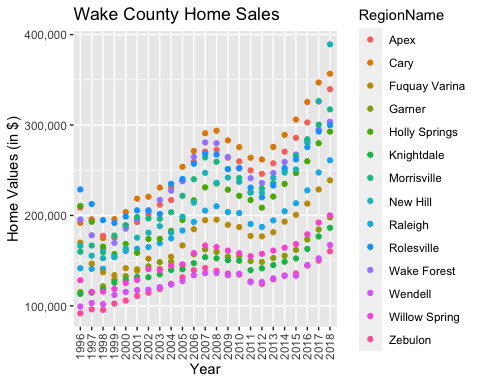
## Rows: 12797 Columns: 277  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (4): RegionName, State, Metro, CountyName  
## dbl (273): RegionID, SizeRank, 1996-04, 1996-05, 1996-06, 1996-07, 1996-08, ...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Wake County Home Sales

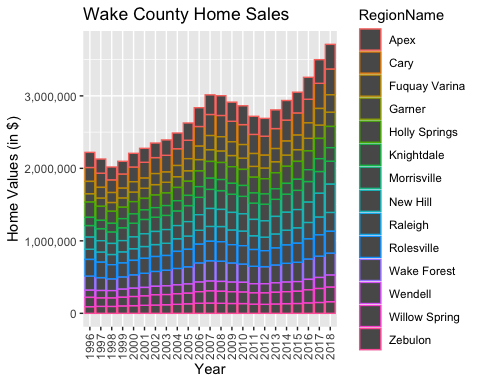
WakeCountySales<- SingleFamilyResidenceSales %>%  
 select(RegionName,State,CountyName,Metro,`1996-05`,`1997-05`,`1998-05`,`1999-05`,`2000-05`,`2001-05`,`2002-05`,`2003-05`,`2004-05`,`2005-05`,`2006-05`,`2007-05`,`2008-05`,`2009-05`,`2010-05`,`2011-05`,`2012-05`,`2013-05`,`2014-05`,`2015-05`,`2016-05`,`2017-05`,`2018-05`)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

## # A tibble: 14 × 27  
## RegionName State CountyName Metro `1996` `1997` `1998` `1999` `2000` `2001`  
## <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Raleigh NC Wake Coun… Rale… 141500 140900 141000 153800 160600 163200  
## 2 Cary NC Wake Coun… Rale… 191900 195700 174900 195900 203600 218500  
## 3 Apex NC Wake Coun… Rale… 208700 195000 177500 175400 189000 192700  
## 4 Wake Forest NC Wake Coun… Rale… 195500 178000 162900 169500 185400 195900  
## 5 Garner NC Wake Coun… Rale… 115200 115000 121700 129000 132800 140300  
## 6 Fuquay Vari… NC Wake Coun… Rale… 169900 146700 137100 133900 141600 139000  
## 7 Holly Sprin… NC Wake Coun… Rale… 210500 193000 165600 157300 168400 158700  
## 8 Morrisville NC Wake Coun… Rale… 159900 166700 159400 177700 189100 198200  
## 9 Knightdale NC Wake Coun… Rale… 113200 114700 118200 125700 128700 131900  
## 10 Zebulon NC Wake Coun… Rale… 91700 96100 95500 102500 105800 110900  
## 11 Wendell NC Wake Coun… Rale… 99200 103200 101900 112100 115400 117400  
## 12 Willow Spri… NC Wake Coun… Rale… 128300 115400 115600 118800 125200 128700  
## 13 Rolesville NC Wake Coun… Rale… 228700 212600 194800 191600 198700 205600  
## 14 New Hill NC Wake Coun… Rale… 166400 155500 152600 154400 163300 175400  
## # … with 17 more variables: `2002` <dbl>, `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>

WakeCountySales<- pivot\_longer(WakeCountySales,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(data=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(data=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)



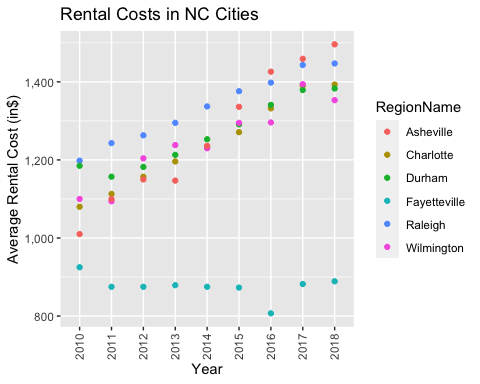
* What have been the overall trends in Wake County Home Values? Values have been consistently rising except for a small dip in the years 1998, 2009, 2010, 2011 and 2012.
* There were dips in home values in the past 20 years. What years did these occur? 1998, 2009, 2010, 2011 and 2012. The reason for dip between 2009 to 20102 could be due to the recession that happened during 2008 to 2009.
* Based on the analysis, where would be the least expensive area to purchase home? Most expensive area? Least expensive area is Zebulon, most expensive are is Raleigh.
* Are any area home values trending down? Is there one area that stands out compared to others? New Hill seems to have dropped a bit between 2017 to 2018

## NC Rental Market

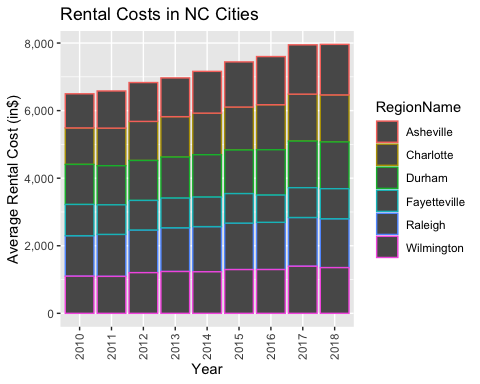
Rentals<- SingleFamilyResidenceRental %>%  
 select(RegionName,State,`2010-11`,`2011-11`,`2012-11`,`2013-11`,`2014-11`,`2015-11`,`2016-11`,`2017-11`,`2018-10`) %>%  
 filter(State %in% c("NC"), RegionName %in% c("Asheville", "Charlotte", "Durham", "Fayetteville", "Raleigh", "Wilmington")) %>%  
 rename("2010"="2010-11","2011"="2011-11","2012"="2012-11","2013"="2013-11","2014"="2014-11","2015"="2015-11","2016"="2016-11","2017"="2017-11","2018"="2018-10") %>%  
 pivot\_longer(c('2010','2011','2012','2013','2014','2015','2016','2017','2018'),names\_to="YR",values\_to="ZHVI")  
  
as\_tibble(Rentals)

## # 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

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



ggplot(data=Rentals,aes(x=YR,y=ZHVI,color=RegionName)) +  
 geom\_col() +  
 labs(title = "Rental Costs in NC Cities", x = "Year", y="Average Rental Cost (in$)") +  
 theme(axis.text.x = element\_text(angle = 90, vjust=0.5)) +  
 scale\_y\_continuous(name="Average Rental Cost (in$)", 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? Rental has been on the increase overall, except for Fayetteville where it is almost around the same pricing and even had a dip in 2016.
* Where is the most expensive city to rent in? Least expensive? Asheville is the most expensive city rent in whereas Fayetteville is the lease expensive.
* You are trying decide between Wilmington and Asheville. Which market has the lowest rent? Wilmington.

## Home Values in Select Rental Markets

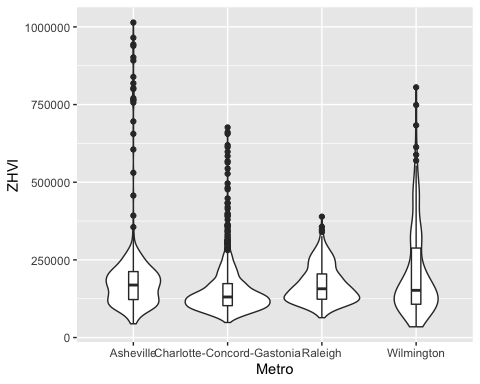
NCHomeSales<- SingleFamilyResidenceSales %>%  
 select(RegionName,State,Metro,`1996-05`,`1997-05`,`1998-05`,`1999-05`,`2000-05`,`2001-05`,`2002-05`,`2003-05`,`2004-05`,`2005-05`,`2006-05`,`2007-05`,`2008-05`,`2009-05`,`2010-05`,`2011-05`,`2012-05`,`2013-05`,`2014-05`,`2015-05`,`2016-05`,`2017-05`,`2018-05`) %>%  
 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") %>%  
 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") %>%  
 group\_by(Metro)  
  
as\_tibble(NCHomeSales)

## # 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

ggplot(data=NCHomeSales,aes(x=Metro,y=ZHVI)) +  
 geom\_violin() +  
 geom\_boxplot(width=0.1)

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

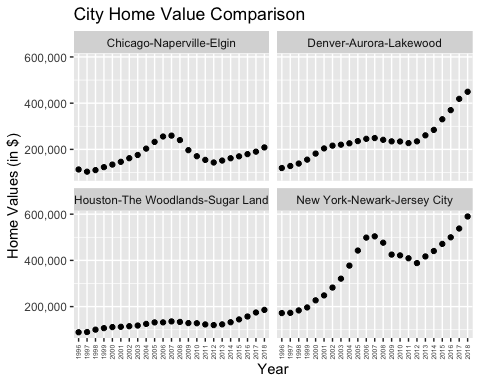
## Warning: Removed 90 rows containing non-finite values (stat\_boxplot).



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

## Relocation Home Value Comparison

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") %>%  
 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") %>%  
 group\_by(Metro)  
  
ggplot(data=NationalHomeSales,aes(x=YR,y=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)



* Based on your analysis, which city’s housing is most affordable? Least affordable? Houston
* 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)? Largest Change = NewYork-Newark-Jersey City, Consistent = Houston-The Woodlands-Sugar Land
* During the market downturn in 2012, which cities were most impacted? Which cities have recovered? Most Impacted = Chicago-Napeville-Elgin, Recovered Cities = NewYork-Newark-Jersey City, Denver-Aurora-Lakewood