# MIS503 – Final Project

## Chandler, Natalie

### Zillow Home Value Index Analysis

library("tidyverse")

## ── Attaching packages ──────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────── tidyverse 1.2.1 ──

## ✔ ggplot2 3.2.1 ✔ purrr 0.3.3  
## ✔ tibble 2.1.3 ✔ dplyr 0.8.3  
## ✔ tidyr 1.0.0 ✔ stringr 1.4.0  
## ✔ readr 1.3.1 ✔ forcats 0.4.0

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

library(readr)  
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.

View(SingleFamilyResidenceSales)  
library(readr)  
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.

View(SingleFamilyResidenceRental)

### Wake County Home Sales

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

Home values in Wake County have increased over the past 20 years.

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

Dips in home values occured between 1996 to 1998 and 2007 to 2013.

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

The least expensive areas are Willow Spring and Zebulun. The most expensive are Apex and Cary.

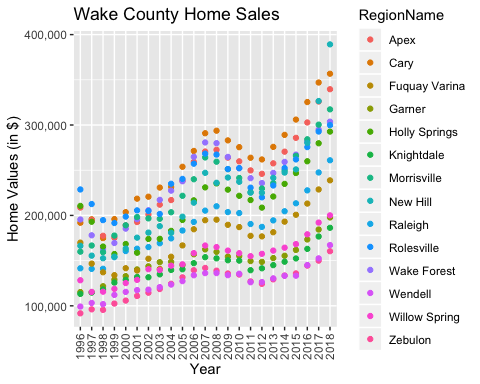
1. **Are any area home values trending down? Is there one area that stands out compared to others?**

Overall, the regions and their decline in home value seem to be connected to economic fluctuations during those time periods (i.e., the 2007 recession). However, the rise in home price seems to be increasing at a faster rate in areas such as Cary and Apex.

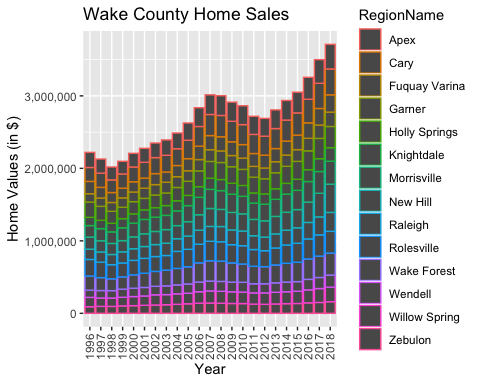
#The select function was used for selecting items to be included within the new tibble (WakeCountySales)  
WakeCountySales <- select(SingleFamilyResidenceSales,RegionName,State,CountyName,Metro,'1996-05','1997-05','1998-05','1999-05','2000-05','2001-05','2002-05','2003-05',  
'2004-05','2005-05','2006-05','2007-05','2008-05','2009-05','2010-05','2011-05','2012-05','2013-05','2014-05','2015-05','2016-05','2017-05','2018-05')  
#The rename function is used to rename columns within the tibble.   
WakeCountySales <- rename(WakeCountySales, '1996' = '1996-05')  
WakeCountySales <- rename(WakeCountySales, '1997' = '1997-05', '1998' = '1998-05')  
WakeCountySales <- rename(WakeCountySales, '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')  
#The tibble was filtered by State and County to include only NC and Wake County.   
WakeCountySales <- filter(WakeCountySales,State=="NC", CountyName == "Wake County")  
#Tidy data mandates that each variable has its own colunn, each observation has its own row, and each value must have its own cell. The gather() function is used to move the years from columns to rows.   
WakeCountySales <- WakeCountySales %>%  
 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')  
#as\_tibble is used to display the tibble itself.   
as\_tibble(WakeCountySales)

## # A tibble: 322 x 6  
## RegionName State CountyName Metro YR ZHVI  
## <chr> <chr> <chr> <chr> <chr> <dbl>  
## 1 Raleigh NC Wake County Raleigh 1996 141500  
## 2 Cary NC Wake County Raleigh 1996 191900  
## 3 Apex NC Wake County Raleigh 1996 208700  
## 4 Wake Forest NC Wake County Raleigh 1996 195500  
## 5 Garner NC Wake County Raleigh 1996 115200  
## 6 Fuquay Varina NC Wake County Raleigh 1996 169900  
## 7 Holly Springs NC Wake County Raleigh 1996 210500  
## 8 Morrisville NC Wake County Raleigh 1996 159900  
## 9 Knightdale NC Wake County Raleigh 1996 113200  
## 10 Zebulon NC Wake County Raleigh 1996 91700  
## # … with 312 more rows

#Use ggplot to create a scatter plot. In this case, YR was placed on the x-axis and ZHVI on the y-axis. theme() was used to change the x-axis to display vertically and scale\_y\_continuous changes the values on the axis to include commas.   
ggplot(WakeCountySales,aes(YR,ZHVI))+labs(title="Wake County Home Sales")+labs(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)+  
geom\_point(aes(color=RegionName))



#Similar code is used to create a bar chart. Alternatively, fill=Region name may be used to remove dark background.   
ggplot(WakeCountySales,aes(YR,ZHVI, color=RegionName))+labs(title = "Wake County Home Sales")+labs(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)+  
geom\_col()



### 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?**

Overall, the cost of rentals has increased in the last decade. However, Fayetteville has remained relatively flat with little change.

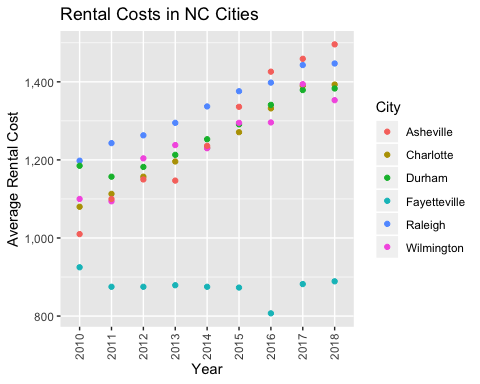
1. **Where is the most expensive city to rent in? Least expensive?**

The most expensive cities are Asheville and Raleigh and the least are Fayetteville and Wilmington.

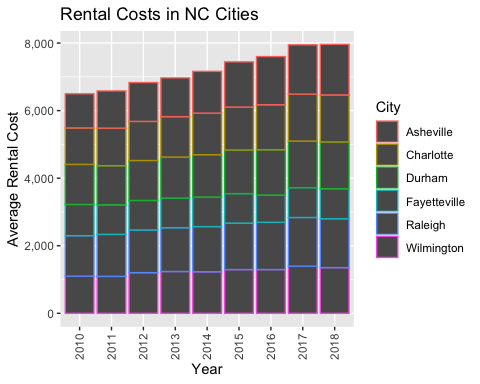
1. **You are trying decide between Wilmington and Asheville. Which market has the lowest rent?**

Wilmington has the lowest rent.

Rentals <- select(SingleFamilyResidenceRental,State,RegionName,'2010-11','2011-11','2012-11','2013-11','2014-11','2015-11','2016-11','2017-11','2018-10')  
Rentals <- rename(Rentals,'2010' = '2010-11')  
Rentals <- rename(Rentals,'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 <- rename (Rentals, 'City' = 'RegionName')  
Rentals <- filter(Rentals,State== "NC")  
# %in% was used to filter within specific variables.   
Rentals <- filter(Rentals, City %in% c("Asheville","Charlotte","Durham","Fayetteville","Raleigh","Wilmington"))  
Rentals <- Rentals %>%  
 gather('2010','2011','2012','2013','2014','2015','2016','2017','2018',key = 'YR', value = 'ZHVI')  
  
ggplot(Rentals,aes(YR,ZHVI))+labs(title="Rental Costs in NC Cities")+labs(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)+  
geom\_point(aes(color=City))



ggplot(Rentals,aes(YR,ZHVI, color=City))+labs(title = "Rental Costs in NC Cities")+labs(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)+  
geom\_col()



### 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 has the lowest median price.

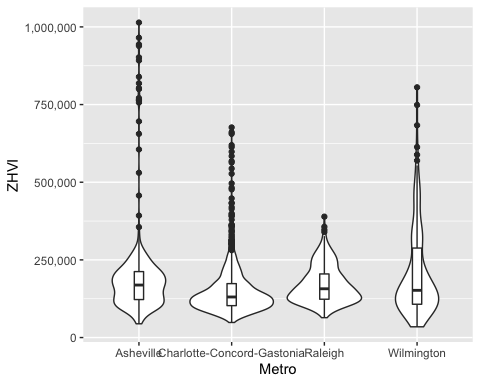
1. **The violin plot will show density meaning the wider the plot is, the more observations occur within that area. Which market has the most density around the median value of homes?**

Charlotte-Concord-Gastonia has the most density around the median. However, Raleigh is comparable.

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')  
NCHomeSales <- rename(NCHomeSales, '1996' = '1996-05')  
NCHomeSales <- rename(NCHomeSales, '1997' = '1997-05', '1998' = '1998-05')  
NCHomeSales <- rename(NCHomeSales, '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 %>%  
 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')  
NCHomeSales <- filter(NCHomeSales,State== "NC")  
NCHomeSales <- filter(NCHomeSales, Metro %in% c("Asheville","Charlotte-Concord-Gastonia","Raleigh","Wilmington"))  
  
ggplot(NCHomeSales,aes(Metro,ZHVI))+labs(x="Metro",y="ZHVI")+scale\_y\_continuous(name="ZHVI", labels = scales::comma)+  
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?**

Houston is most affordable, and New York is the Least affordable.

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)?**

New York and Denver saw the largest change in price over the past 5 years. Houston has remained consistent.

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

New York and Chicago were impacted most. New York has recovered.

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')  
NationalHomeSales <- rename(NationalHomeSales, '1996' = '1996-05')  
NationalHomeSales <- rename(NationalHomeSales, '1997' = '1997-05', '1998' = '1998-05')  
NationalHomeSales <- rename(NationalHomeSales, '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 %>%  
 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')  
NationalHomeSales <- filter(NationalHomeSales,RegionName %in% c("Chicago","Denver","Houston","New York"))  
NationalHomeSales <- filter(NationalHomeSales,Metro %in% c("Chicago-Naperville-Elgin","Denver-Aurora-Lakewood","Houston-The Woodlands-Sugar Land","New York-Newark-Jersey City"))  
  
ggplot(NationalHomeSales,aes(YR,ZHVI))+labs(title="City Home Value Comparison")+labs(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)+facet\_wrap(~Metro)+  
geom\_point()

