# **Mis 503 - Final Project**

## **Bradley, Zackery**

### **Zillow Home Value Index Analysis**

library(tidyverse)

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

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.2 v dplyr 1.0.7  
## v tidyr 1.1.3 v stringr 1.4.0  
## v readr 1.4.0 v forcats 0.5.1

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

library(readr)  
library(dplyr)  
library(lubridate)

##   
## Attaching package: 'lubridate'

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

library(ggplot2)  
SingleFamilyResidenceRental <- read\_csv("C:/Users/bradl/OneDrive/Desktop/MODULE 2/module 7/SingleFamilyResidenceRental.csv")

##   
## -- Column specification --------------------------------------------------------  
## cols(  
## .default = col\_double(),  
## RegionName = col\_character(),  
## State = col\_character(),  
## Metro = col\_character(),  
## CountyName = col\_character()  
## )  
## i Use `spec()` for the full column specifications.

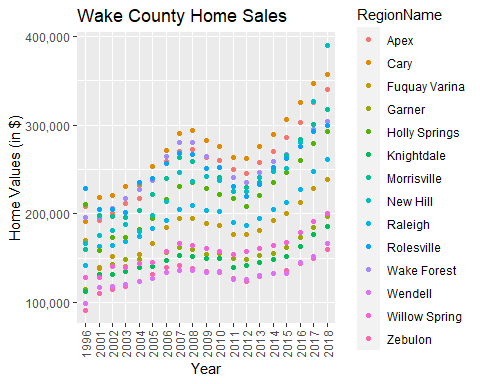
SingleFamilyResidenceSales <- read\_csv("C:/Users/bradl/OneDrive/Desktop/MODULE 2/module 7/SingleFamilyResidenceSales.csv")

##   
## -- Column specification --------------------------------------------------------  
## cols(  
## .default = col\_double(),  
## RegionName = col\_character(),  
## State = col\_character(),  
## Metro = col\_character(),  
## CountyName = col\_character()  
## )  
## i Use `spec()` for the full column specifications.

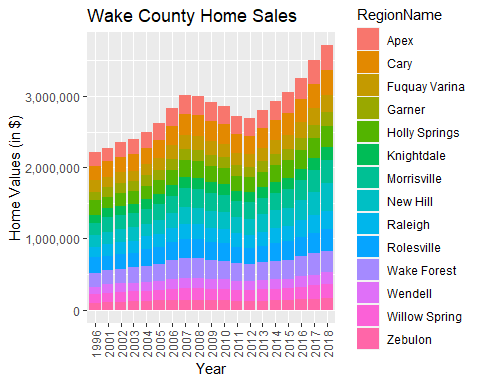
#exercise 1

### **Wake County Home Sales**

WakeCountySales <- select(SingleFamilyResidenceSales, RegionName, State, CountyName, Metro, "1996-05","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", "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 <- pivot\_longer(WakeCountySales, c('1996','2001','2002','2003','2004','2005','2006','2007','2008','2009','2010','2011','2012','2013','2014','2015','2016', '2017', '2018'),names\_to = 'YR', values\_to = 'ZHVI')  
  
WakeCountySales <- filter(WakeCountySales, CountyName=="Wake County" & State== "NC")  
  
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,fill=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)

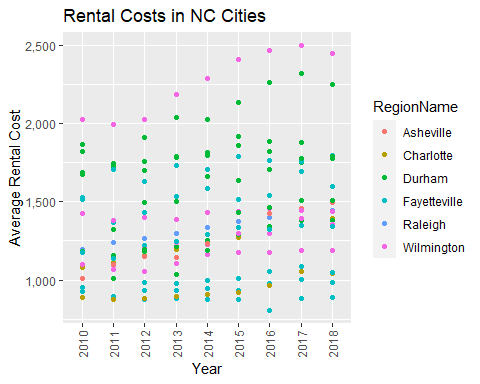
 #a. What have been the overall trends in Wake County Home Values? Homes have been steadily increasing in price since 2012. We can see there is an exception to this in the years 2007-2008, which are famously known as the “housing market crash”. #b. There were dips in home values in the past 20 years. What years did these occur? These dips were between 2008-2012. #c. Based on the analysis, where would be the least expensive area to purchase home? Most expensive area? Wendell would be the least expensive place to purchase a home while Apex is consistently the highest, however there is an outlier in 2018. #d. Are any area home values trending down? Is there one area that stands out compared to others? All areas seem to be following a similiar market trend.

#Exercise 2: North Carolina Rental Market

### **NC Rental Market**

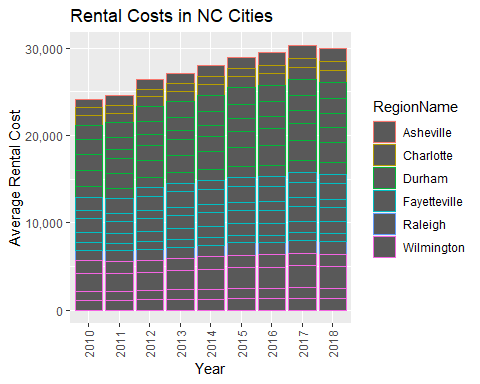
Rentals <- SingleFamilyResidenceRental %>%  
select(RegionName, State, "2010-11","2011-11","2012-11", "2013-11", "2014-11", "2015-11", "2016-11", "2017-11", "2018-10") %>%  
   
filter(RegionName %in% c("Asheville", "Charlotte", "Durham", "Fayetteville", "Raleigh", "Wilmington",State== "NC"))  
  
Rentals<- 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<- pivot\_longer(Rentals, c('2010','2011','2012','2013','2014','2015','2016', '2017', '2018'),names\_to = 'YR', values\_to = 'ZHVI')  
  
ggplot(Rentals, aes(x=YR, y=ZHVI, color=RegionName))+  
geom\_point()+  
labs(title = "Rental Costs in NC Cities", x="Year", y="Average renta Costl")+  
theme(axis.text.x = element\_text(angle = 90, vjust=0.5))+  
scale\_y\_continuous(name="Average Rental Cost", labels = scales::comma)

## Warning: Removed 3 rows containing missing values (geom\_point).



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

## Warning: Removed 3 rows containing missing values (position\_stack).

 #a. What has been the overall trend in the rental market around the state? Are there any cities that have not followed this trend? Rent has increased around the state over the years. All cities seem to follow this same trend, although some cities have higher fluctuation than others. #b. Where is the most expensive city to rent in? Least expensive? Wilmington is the most expensive place to rent, while Fayetteville is the cheapest. #c. You are trying decide between Wilmington and Asheville. Which market has the lowest rent? Asheville

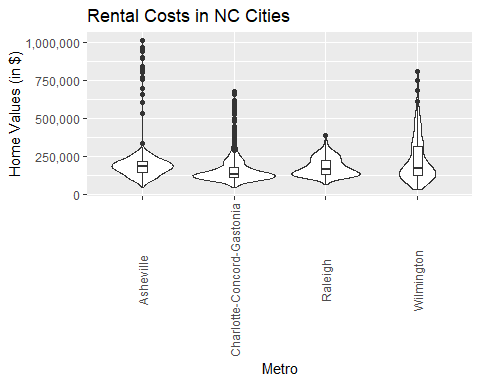
#Exercise 3 **Home Values in Select North Carolina Markets**

### **Home Values in Select North Carolina Markets**

NCHomeSales <- select(SingleFamilyResidenceSales, RegionName, State, Metro, "1996-05","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,"1996"="1996-05", "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 <- pivot\_longer(NCHomeSales, c('1996','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, aes(x=Metro, y=ZHVI))+  
geom\_violin( )+  
geom\_boxplot(width=0.1) +  
labs(title = "Rental Costs in NC Cities", x="Metro", y="ZHVI")+  
theme(axis.text.x = element\_text(angle = 90, vjust=0.5))+  
scale\_y\_continuous(name="Home Values (in $)", labels = scales::comma)

## Warning: Removed 69 rows containing non-finite values (stat\_ydensity).

## Warning: Removed 69 rows containing non-finite values (stat\_boxplot).



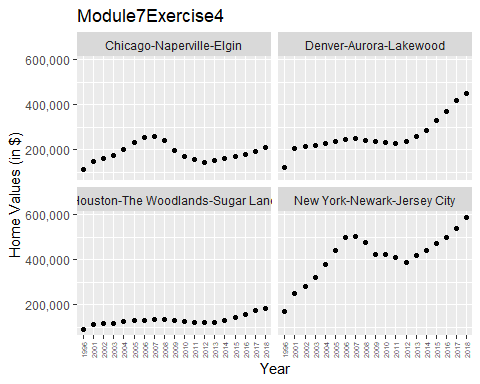
#a. According to the results, which market has the lowest median price (represented as horizontal bar in box plot)? Wilmington has the lowest median price.

#b. 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 have the most density around the median value.

#Exercise 4: **Relocation to Another City**

### **Relocation Home Value Comparison**

NationalHomeSales <- select(SingleFamilyResidenceSales, RegionName, State, Metro, "1996-05","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", "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 <- pivot\_longer(NationalHomeSales, c('1996','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 = NationalHomeSales, aes(x=YR, y=ZHVI))+  
geom\_point()+  
facet\_wrap(~Metro)+  
labs(title = "Module7Exercise4", 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)



#a. Based on your analysis, which city’s housing is most affordable? Least affordable? The most affordable housing would be Houston-the woodlands-sugar land area, while the least affordable would be the New-York-Newark-New Jersey area.

#b. 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)? Both the Denver-Aurora-Lakewood area, in addition to the New York-Newark-New Jersey area have experienced major price upswings in the last 5 years. The Houston-The Woodlands-Sugar Land area has maintained relatively modest with price change the last 5 years.

#c. During the market downturn in 2012, which cities were most impacted? Which cities have recovered?While all of our subject matters for analysis experienced economic downturn in 2012, the data points to New York-Newark-Jersey City being the most economically impacted.However, since 2012 all cities have recovered, with the New York and Denver areas showing the most progression.