**MIS503 - FINAL PROJECT**

**Cardella,April**

**Zillow Home Value Index Analysis**

**Wake County Homes Sales**

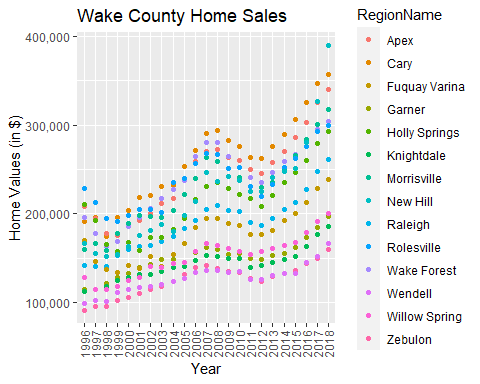
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.0 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.1 ✔ tibble 3.2.0  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.1   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force all conflicts to become errors

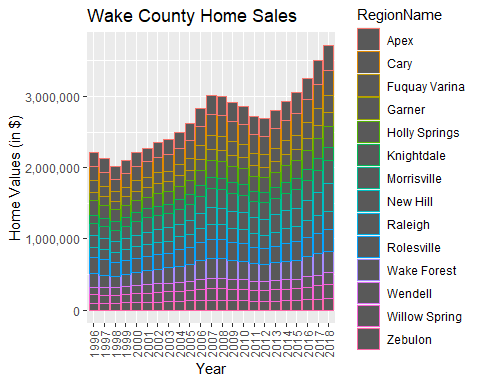
library(readr)  
SingleFamilyResidenceSales <- read\_csv("C:/Users/april/OneDrive - UNC-Wilmington/MIS 503 Programming for Analytics/Module 7/SingleFamilyResidenceSales.csv")

## 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.

WakeCountySales <- filter(SingleFamilyResidenceSales,State=="NC",CountyName=="Wake County")  
  
WakeCountySales<- select(WakeCountySales, "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")  
  
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")   
  
WakeCountySales <- pivot\_longer(WakeCountySales,c('1996','1997','1998','1999','2000','2001','2002','2003','2004','2005','2005','2006','2007','2008','2009','2010','2011','2012','2013','2014','2015','2016','2017','2018'),names\_to='Year',values\_to='ZHVI')  
  
  
ggplot( WakeCountySales, aes(x=Year, y=ZHVI, color=RegionName))+  
 geom\_point()+  
 ggtitle("Wake County Home Sales") +  
 ylab("Home Values in ($)") +  
 xlab("Year") +   
 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=Year, y=ZHVI, color=RegionName))+  
 geom\_col()+  
 ggtitle("Wake County Home Sales") +  
 ylab("Home Values in ($)") +  
 xlab("Year") +   
 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? Over all home prices in Wake County took a dip in 1998 & 1999 as well as in 2011 & 2012. With the highest values being in 2008, 2017 & 2018. The prices have overall gone up and down but have consistently risen after falling.   
   
# There were dips in home values in the past 20 years. What years did these occur? 1998 & 1999. 2011 & 2012  
  
# Based on the analysis, where would be the least expensive area to purchase home? Most expensive area? Wendell for the least and Cary for the most.   
  
#Are any area home values trending down? Is there one area that stands out compared to others? I dont see any that stand out... maybe Zebulon?

**NC Rental Market**

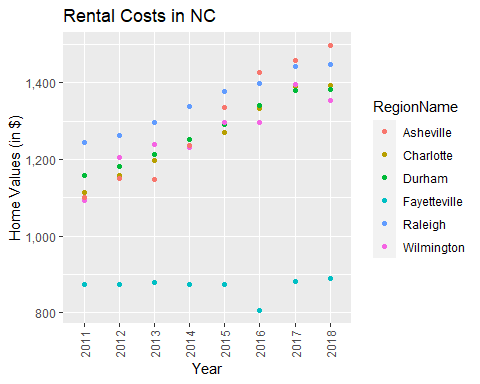
library(readr)  
SingleFamilyResidenceRental <- read\_csv("C:/Users/april/OneDrive - UNC-Wilmington/MIS 503 Programming for Analytics/Module 7/SingleFamilyResidenceRental.csv")

## 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.

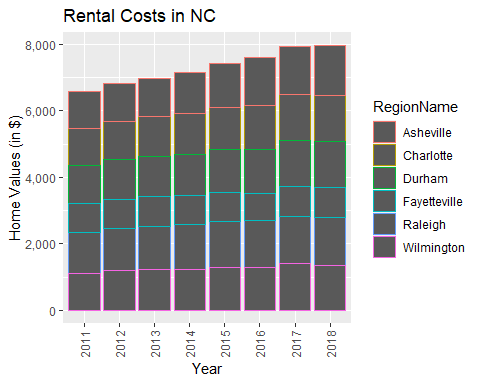
Rentals <- filter( SingleFamilyResidenceRental, State=="NC",RegionName =="Charlotte" | RegionName=="Asheville"|RegionName=="Durham" | RegionName=="Fayetteville"| RegionName=="Raleigh"| RegionName=="Wilmington")  
  
Rentals <- select(Rentals,"State","RegionName","2010-11","2011-11","2012-11","2013-11","2014-11","2015-11","2016-11","2017-11","2018-10")  
  
Rentals <- rename(Rentals, "2010"="2010-11","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(Rentals,c('2011','2012','2013','2014','2015','2016','2017','2018'),names\_to='Year',values\_to='ZHVI')

## # A tibble: 48 × 5  
## State RegionName `2010` Year ZHVI  
## <chr> <chr> <dbl> <chr> <dbl>  
## 1 NC Charlotte 1080 2011 1113  
## 2 NC Charlotte 1080 2012 1157  
## 3 NC Charlotte 1080 2013 1196  
## 4 NC Charlotte 1080 2014 1231  
## 5 NC Charlotte 1080 2015 1271  
## 6 NC Charlotte 1080 2016 1332  
## 7 NC Charlotte 1080 2017 1390  
## 8 NC Charlotte 1080 2018 1393  
## 9 NC Raleigh 1198 2011 1243  
## 10 NC Raleigh 1198 2012 1263  
## # … with 38 more rows

Rentals <-pivot\_longer(Rentals,c('2011','2012','2013','2014','2015','2016','2017','2018'),names\_to='Year',values\_to='ZHVI')  
  
ggplot( Rentals, aes(x=Year, y=ZHVI, color=RegionName))+  
 geom\_point()+  
 ggtitle("Rental Costs in NC") +  
 ylab("Average Rental Costs") +  
 xlab("Year") +   
 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=Year, y=ZHVI, color=RegionName))+  
 geom\_col()+  
 ggtitle("Rental Costs in NC") +  
 ylab("Average Rental Costs") +  
 xlab("Year") +   
 theme(axis.text.x = element\_text(angle = 90, vjust=0.5)) +  
 scale\_y\_continuous(name="Home Values (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? Over all most of the market has risen, with the exception of Fayetteville, but just by taking a slight slip in 2016.  
  
#Where is the most expensive city to rent in? Least expensive? Asheville is the highest, Fayetteville the lowest.   
  
#You are trying decide between Wilmington and Asheville. Which market has the lowest rent? Most recently in 2018, Wilimington is cheaper than Asheville.

**Home Values in Select Markets**

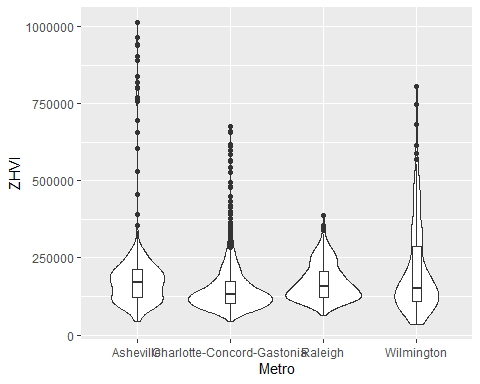
NCHomeSales <- filter(SingleFamilyResidenceSales, Metro == "Asheville"| Metro=="Charlotte-Concord-Gastonia"| Metro=="Raleigh"|Metro=="Wilmington")  
  
NCHomeSales <- select(NCHomeSales,"State","RegionName","Metro","1996"="1996-05","1997"="1997-05","1998"="1998-05", "1999"="1999-05","2000"="2000-05","2001"="2001-05","2002"="2002-05","2003"="2003-05","2004"="2004-05","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 <-pivot\_longer(NCHomeSales,c('1996','1997','1998','1999','2000','2001','2002','2003','2004','2005','2005','2006','2007','2008','2009','2010','2011','2012','2013','2014','2015','2016','2017','2018'),names\_to='Year',values\_to='ZHVI')  
  
NCHomeSales %>%  
 group\_by(Metro)

## # A tibble: 2,553 × 5  
## # Groups: Metro [4]  
## State RegionName Metro Year ZHVI  
## <chr> <chr> <chr> <chr> <dbl>  
## 1 NC Charlotte Charlotte-Concord-Gastonia 1996 116000  
## 2 NC Charlotte Charlotte-Concord-Gastonia 1997 121700  
## 3 NC Charlotte Charlotte-Concord-Gastonia 1998 125300  
## 4 NC Charlotte Charlotte-Concord-Gastonia 1999 130600  
## 5 NC Charlotte Charlotte-Concord-Gastonia 2000 136900  
## 6 NC Charlotte Charlotte-Concord-Gastonia 2001 139700  
## 7 NC Charlotte Charlotte-Concord-Gastonia 2002 141600  
## 8 NC Charlotte Charlotte-Concord-Gastonia 2003 142100  
## 9 NC Charlotte Charlotte-Concord-Gastonia 2004 144200  
## 10 NC Charlotte Charlotte-Concord-Gastonia 2005 148200  
## # … with 2,543 more rows

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

## Warning: Removed 105 rows containing non-finite values (`stat\_ydensity()`).

## Warning: Removed 105 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)? Charlotte Concord Gastonia  
   
#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**

#Based on your analysis, which city’s housing is most affordable? Least affordable?  
#Houston - The Woodlands- Sugar Land is the most affordable, Least is New York - Newark - Jersy City.   
  
#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 - Newark - Jersy City has seen the most changes, Houston has stayed the same  
  
#During the market downturn in 2012, which cities were most impacted? Which cities have recovered? Chicago-Naperville-Elgin, and New York - Newark - Jersy City were most impacted, all have recovered.  
  
NationalHomeSales<- filter(SingleFamilyResidenceSales, RegionName == "Chicago"| RegionName== " Denver"| RegionName== "Houston"| RegionName == "New York")  
  
NationalHomeSales <- select(NationalHomeSales,"State","RegionName","Metro","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 <-pivot\_longer(NationalHomeSales,c('1996','1997','1998','1999','2000','2001','2002','2003','2004','2005','2005','2006','2007','2008','2009','2010','2011','2012','2013','2014','2015','2016','2017','2018',),names\_to='Year',values\_to='ZHVI')  
  
ggplot(NCHomeSales, aes(x=Year, y=ZHVI,))+  
 geom\_point()+  
 ggtitle ("City Home Value Comparison")+  
 ylab("Home Values (in $)") +  
 xlab("Year") +  
 facet\_wrap(~Metro)+  
theme(axis.text.x = element\_text(angle = 90, vjust=0.5, size=5))+  
scale\_y\_continuous(name="Home Values (in $)", labels = scales::comma)

## Warning: Removed 105 rows containing missing values (`geom\_point()`).

