Home and Rental Price Evaluation in North Carolina

Create the R Markdown Document

1.) In RStudio, select *File -> New File -> Text File*. This will create a blank text file in the same area that scripts were created in previous assignments (upper left panel). Save this file to your project as **Mod7FinalAssignAnswer.rmd** (it is important to save with the .rmd extension as this saves the text file as an R Markdown file).

2.) Create a Header 1 with the title: **MIS503 – Final Project**

3.) Create a Header 2 with the title: **Last Name, First Name** (replace with your name)

4.) Create a Header 3 with the title: **Zillow Home Value Index Analysis**

5.) Click on the dropdown arrow next to the Knit icon 

at the top of the R Markdown Pane in RStudio and select Knit to Word.

6.) Notice that you now have a document in your files for the project named **Mod7FinalAssignAnswer.docx**. This is the file you will be uploading later to Canvas.

7.) For this assignment, you will need to download the following files from Canvas which should be saved in the same folder as your project:

   a. SingleFamilyResidenceSales.csv
   b. SingleFamilyResidenceRental.csv

8.) The goal of this assignment is for you to apply all of the skills you have learned over the past 6 weeks to answer research questions posed through a case study on residential home sales and rentals in North Carolina. The files you downloaded are from Zillow who publishes the Zillow Home Value Index (ZHVI) that includes list prices, sales prices and other statistics related to home sales and rentals throughout the country (https://www.zillow.com/research/zhvi-methodology-6032/).

Exercise 1: Trends in home values in Wake County, North Carolina

1.) Create a Header 3 in R Markdown titled: **Wake County Home Sales**

2.) You are considering a move to the Raleigh area and are interested in understanding trends in home values. Run the analysis in the instructions below and come back to this section of the R Markdown document and address the following questions:

   a. What have been the overall trends in Wake County Home Values?
   b. There were dips in home values in the past 20 years. What years did these occur?
   c. Based on the analysis, where would be the least expensive area to purchase home? Most expensive area?
   d. Are any area home values trending down? Is there one area that stands out compared to others?

3.) For this analysis, you will need to import the **SingleFamilyResidenceSales.csv** file from Canvas. Remember to include the correct code in the R Markdown document. Also, remember that you must also load the appropriate packages that will be used for this analysis.

4.) Create a new *tibble* called **WakeCountySales** and use *dplyr* to only include the cities in Wake County. This will require you to filter by State and CountyName. In addition to filtering by the above, you will want to only include the following columns of data in your new tibble (we are taking a snapshot of prices during the month of May each year):

a. RegionName	j. 2001-05	s. 2010-05
b. State	k. 2002-05	t. 2011-05
c. CountyName	l. 2003-05	u. 2012-05
d. Metro	m. 2004-05	v. 2013-05
e. 1996-05	n. 2005-05	w. 2014-05
f. 1997-05	o. 2006-05	x. 2015-05
g. 1998-05	p. 2007-05	y. 2016-05
h. 1999-05	q. 2008-05	z. 2017-05
i. 2000-05	r. 2009-05	aa. 2018-05

5.) Once you have created **WakeCountySales**, you will need to rename the fields above to only include the year (e.g., 1996-05 should be 1996 in the tibble).

6.) If you take a closer look at the data, you will notice it is not tidy. Use *tidyr* to tidy the data so the year is no longer a column but is a row. You will need to use the key ='YR' and value ='ZHVI'.

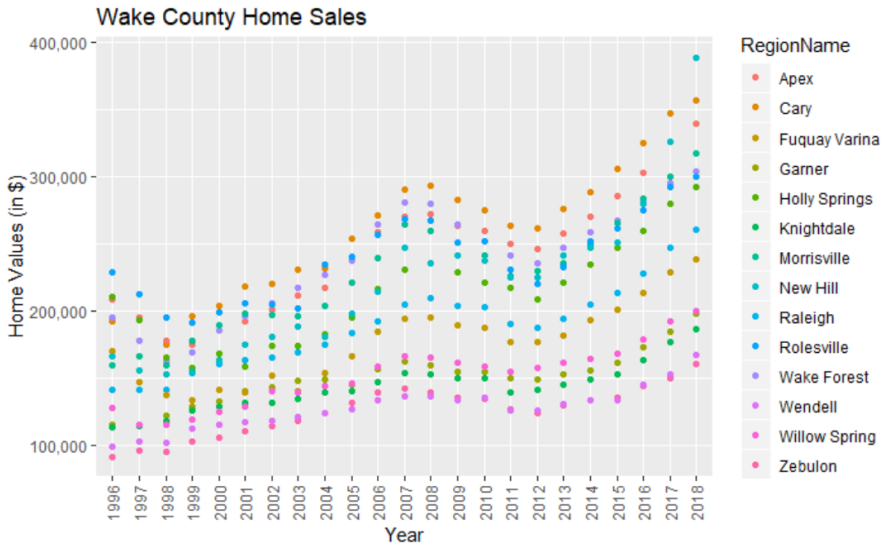
7.) You should now have a tidy set of data that can be analyzed using *ggplot2*.

8.) For the first analysis, we will be creating a scatter plot. You will want to have year on the x-axis and home value (ZHVI) on the y-axis. Also, include the RegionName as a color on your plot. Finally, you will need to include code to give the plot a title, an x-axis name and a y-axis name.

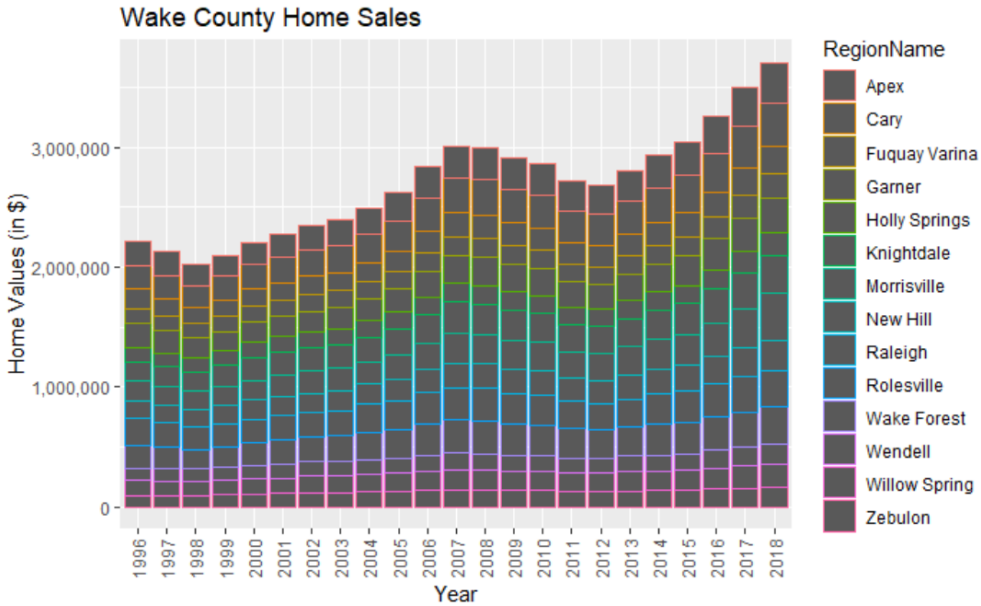
9.) The years will run together and the y-axis values will not show up correctly. To fix this, include the following two lines of code in your *ggplot* function. The first will change the x-axis to display vertically. The second line of code will replace the label for the y-axis and changes the values on the axis to include commas.

   a. theme(axis.text.x = element_text(angle = 90, vjust=0.5))
   b. scale_y_continuous(name="Home Values (in $)", labels = scales::comma)

10.)This analysis should result in the following plot that can be used to evaluate the previous questions:



11.)Before moving to the next exercise, you should also create a stacked bar graph. You can use much of the same code as you did in the scatterplot but will need to use geom_col() to create the below chart.



12.)The last step in this exercise is to go back to the beginning to answer the questions posed.

Exercise 2: North Carolina Rental Market

1.) Create a Header 3 in R Markdown titled: **NC Rental Market**

2.) You are considering relocating to a new city and want to rent. You are looking at some of the larger cities within the state including Asheville, Charlotte, Durham, Fayetteville, Raleigh and Wilmington. Run the analysis in the instructions below and come back to this section of the R Markdown document and address the following questions:

   a. What has been the overall trend in the rental market around the state? Are there any cities that have not followed this trend?
   b. Where is the most expensive city to rent in? Least expensive?
   c. You are trying decide between Wilmington and Asheville. Which market has the lowest rent?

3.) For this analysis, you will need to import the SingleFamilyResidenceRental.csv file from Canvas. Remember to include the correct code in the R Markdown document. Also, remember that you must also load the appropriate packages that will be used for this analysis.

4.) Create a new tibble called **Rentals.** Use *dplyr* to only include the cities listed in instruction 2 above. Remember that there may be cities with the same name in different states. In addition to filtering by the above, you will want to include the following columns of data in your new tibble (we are taking a snapshot of rent during the month of November each year):

   a. RegionName (this is the field that contains the City)
   b. State
   c. 2010-11
   d. 2011-11
   e. 2012-11
   f. 2013-11
   g. 2014-11
   h. 2015-11
   i. 2016-11
   j. 2017-11
   k. 2018-10*****

   ***Note that this set does not contain 2018-11 data so just use 2018-10**

5.) Once you have created **Rentals**, you will need to rename the fields above to only include the year (e.g., 2010-10 should be 2010 in the tibble).

6.) If you take a closer look at the data, you will notice it is not tidy. Use *tidyr* to tidy the data so the year is no longer a column but is a row. You will need to use the key ='YR' and value ='ZHVI'.

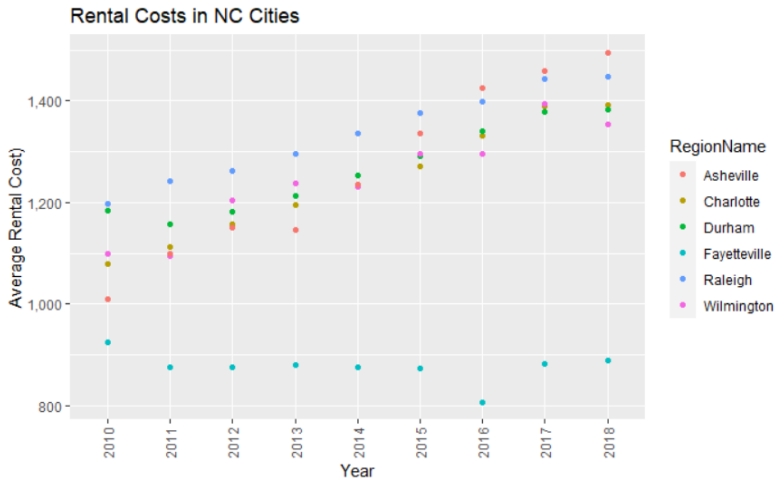
7.) You should now have a tidy set of data that can be analyzed using *ggplot*.

8.) For the first analysis, we will be creating a scatter plot similar to the previous exercise. You will want to have year on the x-axis and home value (ZHVI) on the y-axis. Also, include the City as a color on your plot. Finally, you will need to include code to give the plot a title, an x-axis name and a y-axis name.

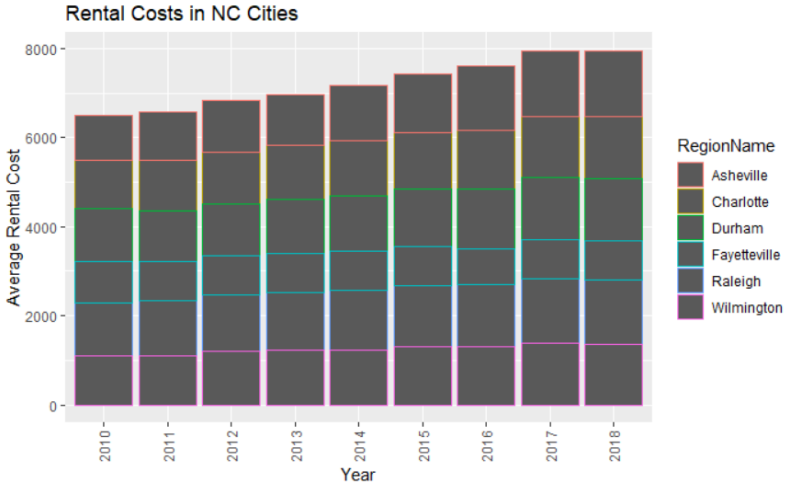
9.) The years will run together and the values will not show up correctly. To fix this, include the following two lines of code in your ggplot function. The first will change the x-axis to display vertically. The second line of code will replace the label for the y-axis and changes the values on the axis to include commas.

   a. theme(axis.text.x = element_text(angle = 90, vjust=0.5))
   b. scale_y_continuous(name="Home Values (in $)", labels = scales::comma)

10.)This analysis should result in the following plot that can be used to evaluate the previous questions:



11.)Before moving to the next exercise, you should also create a stacked bar graph. You can use much of the same code as you did in the scatterplot but will need to use geom_col() to create the below chart.



12.)The last step in this exercise is to go back to the beginning to answer the questions posed.

Exercise 3: Home Values in Select North Carolina Markets

1.) Create a Header 3 in R Markdown titled: **Home Values in Select Rental Markets**

2.) You have made the choice that you want to focus on 4 regions (Asheville, Charlotte-Concord-Gastonia, Raleigh and Wilmington) and instead of renting, you would like to purchase a home. Run the analysis in the instructions below and come back to this section of the R Markdown document and address the following questions (you may need to do some research on reading a violin plot https://blog.modeanalytics.com/violin-plot-examples/):

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

3.) For this analysis, we will be using the SingleFamilyResidenceSales file again.

4.) Create a new tibble called **NCHomeSales** and use dplyr to only include the cities (using the Metro field in the NCHomeSales file) listed above in instruction 2. This will require you to filter by State and Metro. In addition to filtering by the above, you will want to only include the following columns of data in your new tibble (we are taking a snapshot of prices during the month of May each year):

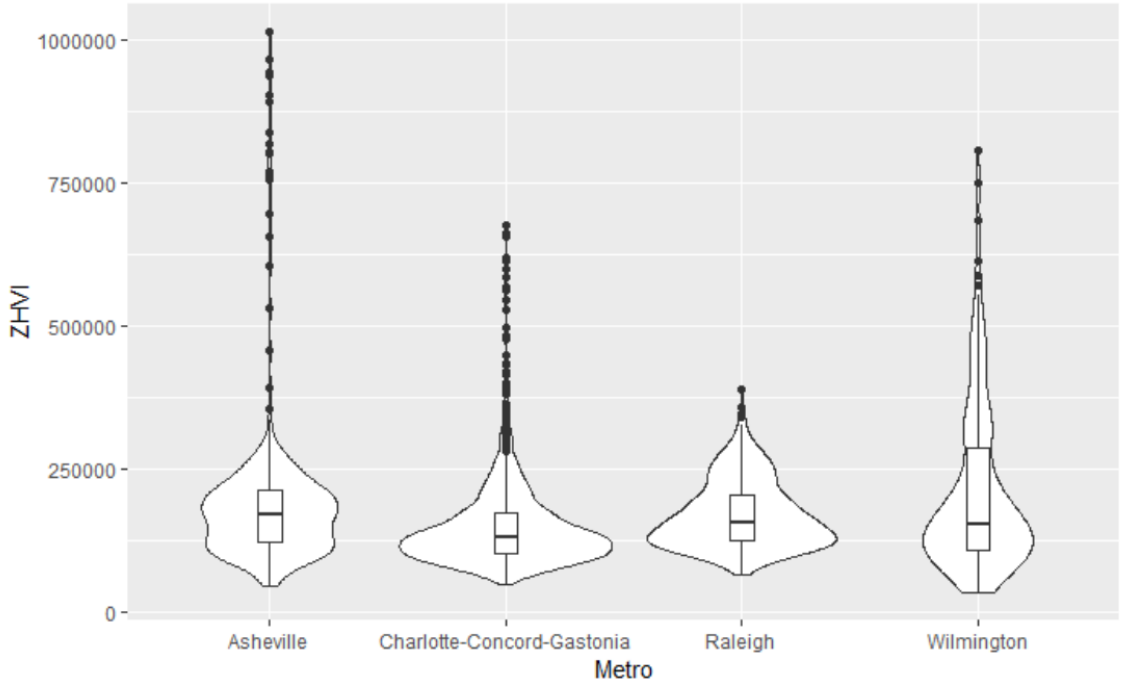
   a. RegionName
   b. State
   c. Metro
   d. 1996-05
   e. 1997-05
   f. 1998-05
   g. 1999-05
   h. 2000-05
   i. 2001-05
   j. 2002-05
   k. 2003-05
   l. 2004-05
   m. 2005-05
   n. 2006-05
   o. 2007-05
   p. 2008-05
   q. 2009-05
   r. 2010-05
   s. 2011-05
   t. 2012-05
   u. 2013-05
   v. 2014-05
   w. 2015-05
   x. 2016-05
   y. 2017-05
   z. 2018-05

5.) Once you have created **NCHomeSales**, you will need to rename the fields above to only include the year (e.g., 2010-10 should be 2010 in the tibble).

6.) If you take a closer look at the data, you will notice it is not tidy. Use *tidyr* to tidy the data so the year is no longer a column but is a row. You will need to use the key ='YR' and value ='ZHVI'. Finally, to run our analysis you will need to group the **NCHomeSales** by **Metro** to get all of the data in the correct order for our plots.

7.) You should now have a tidy set of data that can be analyzed using ggplot.

8.) For this analysis, we will be creating a violin and boxplot that appears on the same plot. You will want to have Metro on the x-axis and home value (ZHVI) on the y-axis. To run this analysis, you will use the geom_violin( ) and geom_boxplot(width=0.1) commands.



9.) The last step in this exercise is to go back to the beginning to answer the questions posed.

Exercise 4: Relocation to Another City

1.) Create a Header 3 in R Markdown titled: **Relocation Home Value Comparison**

2.) You have been given a new opportunity to relocate for a new position within your company. They have given you the option of 4 different areas in the country (Chicago, Denver, Houston or New York). Run the analysis in the instructions below and come back to this section of the R Markdown document and address the following questions:

   a. Based on your analysis, which city's housing is most affordable? Least affordable?
   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)?
   c. During the market downturn in 2012, which cities were most impacted? Which cities have recovered?

3.) For this analysis, you will need to import the SingleFamilyResidenceSales.csv file from Canvas. Remember to include the correct code in the R Markdown document. Also, remember that you must also load the appropriate packages that will be used for this analysis.

4.) Create a new tibble called **NationalHomeSales** and use dplyr to only include the cities listed above in instruction 2 of Exercise 4. This will require you to filter by RegionName using the %in% code in the command and remember the possibility of multiple states having the same city name. In addition to filtering by the above, you will want to only include the following columns of data in your new tibble (we are taking a snapshot of prices during the month of May each year):

   a. RegionName
   b. State
   c. Metro
   d. 1996-05
   e. 1997-05
   f. 1998-05
   g. 1999-05
   h. 2000-05
   i. 2001-05
   j. 2002-05
   k. 2003-05
   l. 2004-05
   m. 2005-05
   n. 2006-05
   o. 2007-05
   p. 2008-05
   q. 2009-05
   r. 2010-05
   s. 2011-05
   t. 2012-05
   u. 2013-05
   v. 2014-05
   w. 2015-05
   x. 2016-05
   y. 2017-05
   z. 2018-05

5.) Once you have created **NationalHomeSales**, you will need to rename the fields above to only include the year (e.g., 1996-05 should be 1996 in the tibble).

6.) If you take a closer look at the data, you will notice it is not tidy. Use *tidyr* to tidy the data so the year is no longer a column but is a row. You will need to use the key ='YR' and value ='ZHVI'. Finally, to run our analysis you will need to group the tibble by Metro.

7.) You should now have a tidy set of data that can be analyzed using *ggplot*.

8.) For the first analysis, we will be creating a scatter plot. You will want to have year on the x-axis and home value (ZHVI) on the y-axis. However, instead of setting the city to be different colors, we would like to create a facet of the information so each city appears in a separate scatterplot. To do this you will need to add the following code to your function:

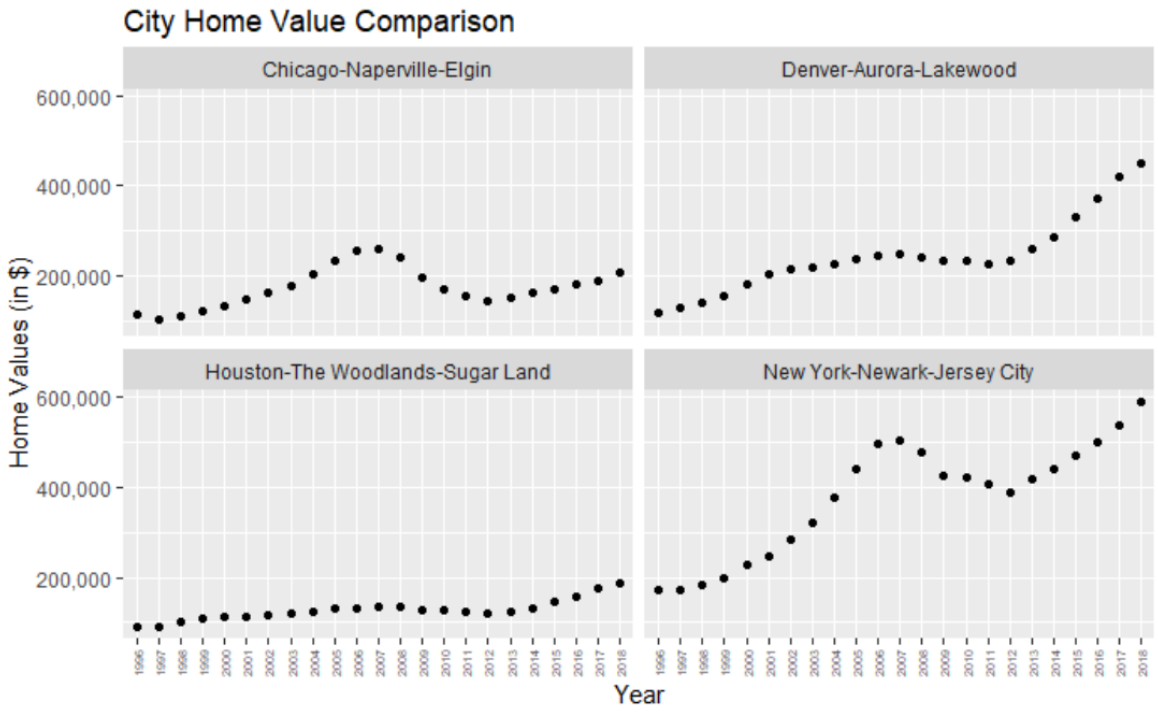
```
facet_wrap(~Metro)
```

9.) Finally, you will need to include code to give the plot a title, an x-axis name and a y-axis name.

10.) The years will run together and the values will not show up correctly. To fix this, include the following two lines of code in your ggplot function. The first will change the x-axis to display vertically and adjust the size of the text to 5pt. The second line of code will replace the label for the y-axis and changes the values on the axis to include commas.

a. theme(axis.text.x = element_text(angle = 90, vjust=0.5, size=5))
b. scale_y_continuous(name="Home Values (in $)", labels = scales::comma)

11.) This analysis should result in the following plot that can be used to evaluate the previous questions:



10.) The last step in this exercise is to go back to the beginning to answer the questions posed.

11.) Finally, knit your R Markdown to Word and upload the document to Canvas. Before submitting, open the word document and adjust the size of the visuals to make them a little larger in the document.