Exercise 14: Moving to Canada

ECON 256   Data Analysis and Visualization

Introduction

Attracted by the possibility of obtaining free health care and affordable maple syrup, Mr. and Mrs. Thompson, a young American couple have decided to head north to start a new life in Vancouver, Canada. They need to buy a home in Vancouver, but alas, they know nothing about the Vancouver real estate market. Because of your sophisticated data analysis and visualization skills, they have hired YOU to gather information about the Vancouver real estate market and provide them with a brief report.

Follow these steps to prepare your report. Use a word document to write up your report as you go, noting all the data points you are asked to find.

1 Download/Load the Data

Download data on housing sales in Vancouver (vanhousing.csv) and download a shapefile of census tracts in Vancouver (vantracts). Keep vanhousing.csv and the shapefile folder in your working directory. You'll need to unzip the vantracts folder.

Start a R script. You will need the tidyverse and sf packages for this project. Initialize the packages by using `library(tidyverse)` and `library(sf)`.

Load the vanhousing.csv file with a `read_csv()` function.

This data covers all homes sold in Vancouver from 2005-2015.

Load the vantracts shapefile with a `read_sf()` function, like this:

```
vantracts <- read_sf(dsn = "vantracts", layer = "vantracts")
```

2 Provide Some Summary Statistics

First, provide the couple with some basic summary statistics pertaining to the Vancouver housing market. Prepare a table that looks like this, filling in the blank boxes.

Median home price	
Mean home price	
Standard deviation of home price	
Median number of bedrooms	
Median number of full bathrooms	
Median square footage	

Produce a histogram of housing sale prices using ggplot(). Give the plot a title and label the axes. Include an image of the histogram in your report.

Produce a graph that shows how the average sale price has changed across the years in the data. The year of sale is captured by the variable: `year_of_sale`. Give the plot a title and label the axes. Include an image of the graph in your report.

Mr. Thompson hopes to fill his new home with lots of children. He wants some examples of very large homes. How big is the largest home in the data in terms of number of bedrooms (`Beds`)? What is the address of this

home? How big is the largest home in the data in terms of square footage (`area_sqft`)? What is the address of this home?

3 West vs East Vancouer

The Thompsons are wondering whether they should focus their housing search on the West Side or East Side of Vancouver. Note that the housing sale data includes a variable (`west`) that takes a value of 1 if the home is on the West Side.

What is the median home price of a home on the west side of Vancouver? What is the median price on the east side?

A command like this might be helpful:

```
vanhousing %>%
  group_by(west)%>%
  summarize(medprice = median(price))
```

(Where `vanhousing` is an object containing the vanhousing.csv data).

4 Home Size

The Thompsons are also wondering how large of a house they could afford. Write a similar command to the previous section to find the median home price by number of bedrooms and include a table like the following:

Number of Bedrooms	Median Home Price
0 Bedrooms (studio)	
1 Bedroom	
2 Bedrooms	
3 Bedrooms	
4 Bedrooms	
5 Bedrooms	
6 Bedrooms	
7 Bedrooms	
8 Bedrooms	
9 Bedrooms	
10 Bedrooms	
11 Bedrooms	
12 Bedrooms	
13 Bedrooms	

5 Map the Median Price by Census Tract

The Thompsons want a bit more information about how prices vary by neighbourhood across Vancouver. Let's include a map of median home price by census tract. Note that both the home price data and the shapefile data have a variable called `tract` that can be used to join them together.

Let's calculate the median sales price in each tract, then join this data to the shapefile using the `left_join()` function. Do this with the following code. The first three lines produce a new object that gives the median home price in every census tract in Vancouver. The last line of code joins the shapefile to the new tract level housing data.

```
vanhousingtracts<-vanhousing %>%
```

```
  group_by(tract)%>%
  summarize(medprice = median(price))

combined<-left_join(vantracts,vanhousingtracts,by="tract")
```

(Where `vanhousing` is an object containing the vanhousing.csv data and `vantracts` is an object that contains the vantracts simple feature).

Now use the new simple feature object (`combined`) to produce a map of median home prices by census tract and include it in your report. (Any mapping method is fine, look back at past mapping exercises for examples). Include a title and any other information needed to interpret the map.

6 Analyze the Determinants of Price Differences Across Properties

The Thompson's are trying to weigh the benefits of an additional bedroom against the benefits of other home characteristics. They want to know the cost of various housing characteristics.

Use a regression (the `summary(lm())` function) to estimate the average increase in a home's price that is related to adding one additional bedroom (Hint: `formula= price ~ Beds`).

Perform the same analysis for number of full bathrooms (Hint: `formula= price ~ FullBath`).

Perform a third regression where the regression equation is: `formula= price ~ Beds + FullBath + west`. What do each of the coefficient estimates mean in the result of this regression?

(For help on this section refer to the class notes and exercise from the Chili Cook Off class.)

7 Recommend Some Homes

The Thompson's have mentioned that their ideal home would have the following characteristics:

- at least 6 bedrooms
- at least 5 full bathrooms
- located on the West Side of Vancouver
- cost less that $500,000

Use a filter command to select the properties that have these characteristics and provide the addresses of these properties in your report.

8 Send me Your Report and Code

Ensure your report is clear and readable and that you have explained all of the components.

Save your report as a .pdf.

Name the pdf with your last name, followed by the exercise number.

Name your .R code with your last name, followed by the exercise number.

Submit both files (the pdf and the R code) on Laulima.