MAS 627

Due Monday 9/28 by Midnight

NAME: _____

Read These Instructions

- Complete the exam using R Markdown (you may use an R script if needed, for a 10% deduction).
  - All data cleaning/manipulation should be done within a `dplyr` chain and connected directly to the `read.csv()` function.
  - All plots should be created using the `ggplot2` package.
  - All code and output should be visible in your knitted document.
  - If your document will not knit due to a mistake in your code, add `eval=F` to knit as-is.
  - Format your document appropriately: do not leave any irrelevant code or output in your final document.

Submit both your knitted document and .Rmd file to Blackboard

Part 1

The data for this part comes from the Inc. 5000 2018 list. It represents an annual ranking of the fastest growing privately held companies in America.

Note that two companies were removed for having mostly missing values (thus, you have the Inc. 4,998 list).

- The data can be read in from the following location:
  - https://s3.amazonaws.com/douglas2/data/inc.csv

Questions

1. Remove the 'X_...' from the column names.

2. Return the name of the company on the list that has between \$5 billion and \$6 billion in revenue?

3. What is the average growth rate among Inc. 5000 companies that are on the list for the first time?

4. For companies that remain on the Inc. 5000 list, how do their growth rates change through time? Build a barplot that displays the average growth rate by number of years on the list.

   - Format axes, axis breaks, and labels appropriately
   - Make it look nice

Part 2

The data from this section comes from data.world ("National Farmers Markets List"). The Farmers Market Directory lists markets that feature two or more farm vendors selling agricultural products directly to customers at a common, recurrent physical location. Maintained by the Agricultural Marketing Service, the Directory is designed to provide customers with convenient access to information about farmers market listings to include: market locations, directions, operating times, product offerings, accepted forms of payment, and more. The data is sourced from the USDA website if you are interested.

- The data can be read in from the following location:
    - https://s3.amazonaws.com/douglas2/data/farmersMarkets.csv

Questions

1. What percent of markets sell coffee?

2. Among markets that sell coffee, what percent sell wine?

3. How many markets in Florida sell coffee?

4. Construct a barplot that shows the number of markets open by day of the week (which can be found in the Hours variable).

    - Format as needed and make it look nice.

Part 3

The data for this section comes from data.world ("Atlanta Open Checkbook"). It contains information related to purchasing in the city of Atlanta.

- Data can be read in from the following location:
  - https://s3.amazonaws.com/douglas2/data/ledger.csv

Please recreate the following bar plot using the **ggplot2** package. Your colors do not have to match mine, but you should change the color defaults to something you like.
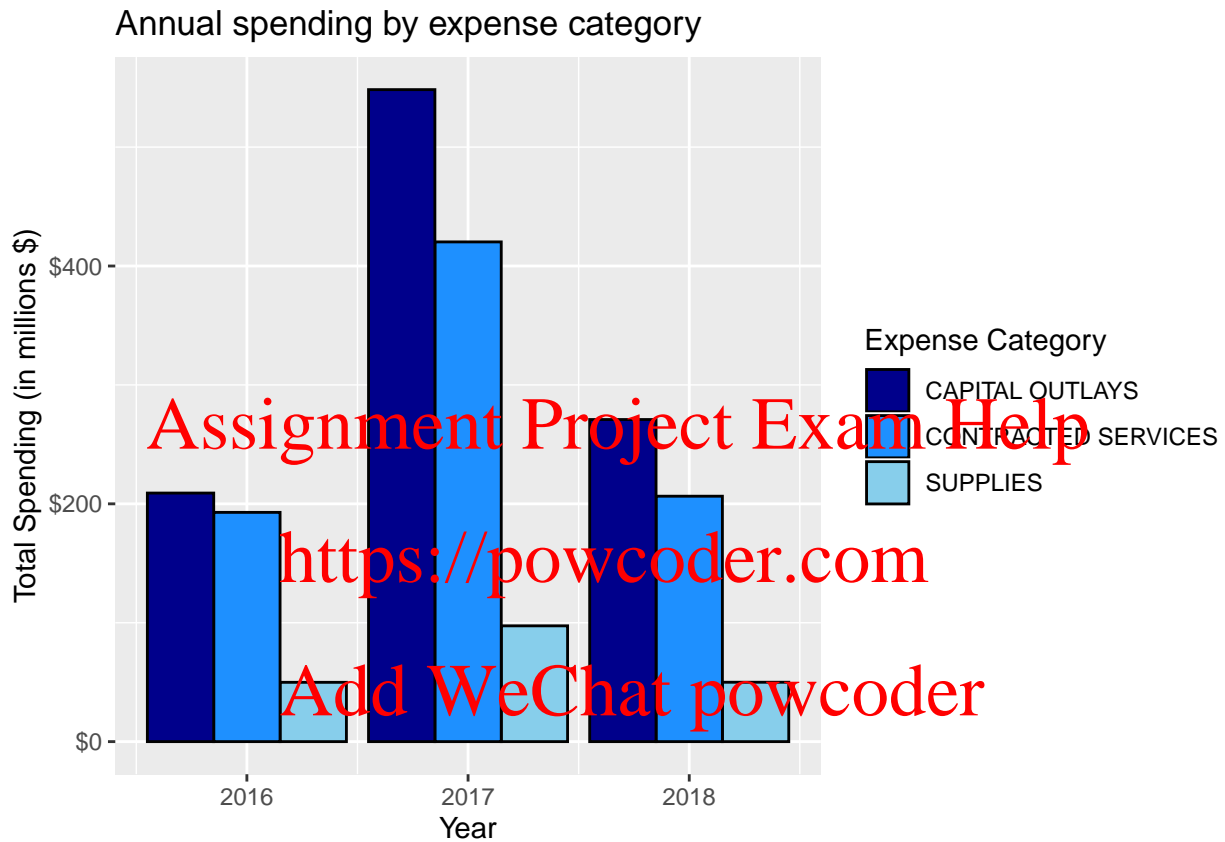
Part 4 - Zillow Observed Rent Index

The data for this section comes from the Zillow Research data repository. It contains the typical observed market rate rent for a large number of metropolitan areas around the country.

- Data can be read in from the following location:
  - https://douglas2.s3.amazonaws.com/data/zillowZORI.csv

Please replicate the following time series visual using the `ggplot2` package. Note that grey lines each represent a metropolitan area, while the red line represents a state-wide average. You are welcome to choose other or more states, as long as they have at least two metro areas represented in the data!

Zillow Observed Rent Index by Metropolitan Area
*red line represents the state–wide average